Battery health diagnosis and prognosis using physics-informed data-driven methods

Zihao Zhou

Department of Engineering Science

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy

St Anne's college

July 2025

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and the acknowledgements and publications pages. Parts of the work have been published in journal papers and presented at conferences and seminars. These are specified in the text and referenced as appropriate.

Zihao Zhou *July 2025*

Acknowledgements

Spending the last four years completing my DPhil at Oxford, working with wonderful people, has been a life-changing experience. I am deeply grateful to everyone who helped my research, social activities, as well as the funding that made this journey possible.

I owe thanks to my fiancée, Jingyu, whose understanding and support over the years were invaluable. Without her accompanying, this work would undoubtedly not have been completed. I am also deeply grateful to my family-Ying Han and Huimin Zhou-and to all my friends for their steadfast encouragement and support.

I was lucky to do my research with excellent people in the Battery Intelligence Lab. First and foremost, I would like to thank Professor David Howey for his support, guidance and mentorship throughout my DPhil. I would also like to give special thanks to Dr. Antti Aitio, whose technical knowledge and willingness to answer my endless questions is greatly appreciated. Also, these four years would not have been the same without my fellow students and colleagues, Malgorzata Wojtala, Volkan Kumtepeli, Nicola Courtier, Brady Planden, Dibyendu Ghosh, Adam Lewis-Douglas, Masaki Adachi, Rebecca Perriment, Joe Ross, Emmanuelle Hagopian, Blanka Gaál, and Tihana Stefanic. With them, the hours spent in the office were an absolute delight.

I am also deeply grateful to the Chinese Scholarship Council-University of Oxford Scholarship, and Russell studentship supported by Rimac Automobili. These financial supports enabled me to pursue my studies in Oxford and engage in academic activities.

Abstract

Estimation and prediction of battery state of health (SOH) are critical to ensuring operational safety and reliability. Due to complex degradation mechanisms and the limitations of measurement techniques in real-world environments, accurate SOH estimation and prediction remain challenging. This thesis begins with a review of health diagnosis and prognosis approaches in the battery literature, focusing on two problem settings: battery lifetime prediction and SOH estimation.

For battery lifetime prediction, a hierarchical Bayesian regression model is proposed and thoroughly evaluated. The approach is motivated by the observed variation in the relationship between extracted health features and lifetime labels under different aging conditions. Health features are categorized into two groups: individual cell-level features (reflecting intrinsic variability across cells) and population-level features (capturing the influence of cycling conditions on the average behaviour of the population). It is shown how this relationship can be explicitly modeled through a hierarchical dependency between these two types of features. The proposed method is validated using a public dataset consisting of cells subjected to fast-aging experiments, as well as a self-tested dataset that covers more realistic and diverse cycling conditions.

For state of health estimation, an aging-aware equivalent circuit model is proposed that combines the flexibility of data-driven methods with a model-based framework. Gaussian process regression is used to include parameter dependencies on operating conditions and lifetime for the equivalent circuit model. Both capacity and resistance are estimated from operational data without requiring ground-truth labels. The close relationship between the estimated resistance function and aging-induced changes of the open circuit voltage curve allows further estimation of degradation modes. Results from 114 battery packs deployed in off-grid solar systems indicate resting at high voltage will accelerate battery aging. Robust state of heath estimation for large fleets of field-deployed batteries enables a deeper understanding about how customer usage profiles influence battery degradation, which in turn has the potential to inform optimized operational strategies to prolong battery lifetime.

Publications

The work presented in this thesis is based on the following publications, and * denotes the equal-contribution as co-first author:

[1] **Zhou Zihao**, and David A. Howey. "Bayesian hierarchical modelling for battery lifetime early prediction." *IFAC-PapersOnLine* 56.2 (2023): 6117-6123.

[2] Li Tingkai*, **Zhou Zihao***, Adam Thelen, David A. Howey and Hu Chao. "Predicting battery lifetime under varying usage conditions from early aging data." *Cell Reports Physical Science* 5.4 (2024).

[3] **Zhou Zihao**, Antti Aitio, and David Howey. "Learning Li-ion battery health and degradation modes from data with aging-aware circuit models." *Applied Energy* 397 (2025): 126375.

[4] **Zhou Zihao** and David Howey. "Battery capacity and degradation modes diagnosis for solar off-grid systems in Africa" (**To be submitted**)

For all the above publications [1, 2, 3, 4], I was the primary author. I contributed the theory, code, analysis and writing for papers [1, 3, 4], all of which was guided, advised, reviewed and edited by David Howey. For the equal-contributed publication [2] with Tingkai Li, I contributed to the hierarchical Bayesian modeling part, as well as corresponding writing and editing. Tingkai was mainly responsible for battery testing, data pre-processing, feature extraction and paper writing. Hu Chao and David Howey were the principle investigators for this study and contributed supervision and review.

Table of Contents

Nomenclature

Acronyms and Abbreviations

LIB . . . . . Lithium-ion battery

LFP . . . . . lithium iron phosphate graphite

NMC . . . . . Nickel manganese cobalt oxides

BESS . . . . . Battery energy storage system

BMS . . . . . Battery management system

EV . . . . . Electric vehicle

OCV . . . . . Open circuit voltage

SOC . . . . . State of charge

SOH . . . . . State of health

SOE . . . . . State of energy

SOP . . . . . State of power

BoL . . . . . Beginning of life

EoL . . . . . End of life

RUL . . . . . Remaining useful lifetime

EIS . . . . . Electrochemical impedance spectroscopy

HPPC . . . . . Hybrid pulse power characterization

GITT . . . . . Galvanostatic intermittent titration technique

CCCV . . . . . constant current constant voltage

RPT . . . . . Reference performance test

DoD . . . . . Depth of discharge

EoC . . . . . End of charge

EoD . . . . . End of discharge

EFC . . . . . Equivalent full cycle

ECM . . . . . Equivalent circuit model

ESC . . . . . Enhanced self-correcting cell

EM . . . . . Electrochemical model

P2D Pseudo-2-dimensional

SPM(e) . . . . . Single particle model (with electrolyte)

ENR . . . . . Elastic net regression

RF . . . . . Random forest

SVM . . . . . Support vector machine

SVR . . . . . Support vector regression

GPR . . . . . Gaussian process regression

ARD . . . . . Automatic relevance determination

WV . . . . . Wiener velocity

NN . . . . . Neural network

RLS . . . . . Recursive least squares

LSTM . . . . . Long short-term memory

HBM . . . . . Hierarchical Bayesian regression model

MCMC . . . . . Markov chain Monte Carlo

DV . . . . . Differential voltage

IC . . . . . Incremental capacity

LLI . . . . . Loss of lithium inventory

LAM . . . . . Loss of active material

SEI . . . . . Solid electrolyte interphase

RMSE . . . . . Root mean square error

MAPE . . . . . Mean percentage absolute error

VPC . . . . . Variance partition coefficient

NLML . . . . . Negative log maximum likelihood

LTI . . . . . linear time invariant

SPDE . . . . . Stochastic partial differential equation

RTS . . . . . Rauch-Tung-Striebel

MAP . . . . . Maximum a posteriori

Equivalent circuits

V_0 . . . . . Open circuit voltage

$V_{\mathbf{term}}$ . . . . . Terminal voltage

z . . . . . State of charge

$Q^{-1}(.)$ . . . . . Inverse capacity

I . . . . . Applied current

Gaussian process

F . . . . . Continuous-time state transition matrix

K . . . . . Gramian matrix

P . . . . . State covariance matrix

W . . . . . Process noise covariance matrix

$\mathbf{Q_c}$ . . . . . Spectral density matrix

X . . . . . Training input

$k(.)$ . . . . . Kernel/covariance function

1 | Introduction

1.1 Battery state of health

Lithium-ion batteries (LIBs) are playing an increasingly important role in the modern energy industry. On one hand, LIBs are widely used in decentralized solar-battery systems across the developing world to provide local families with affordable energy [5]. On the other hand, battery energy storage systems (BESSs) are essential for grid stability and frequency regulation, especially with the growing integration of renewable energy sources such as solar and wind [6, 7]. Additionally, LIBs with high energy density and long cycle life have enabled the rapid development of the electric vehicle (EV) industry [8], contributing to the reduction of fossil fuel consumption and associated CO_2 emissions.

The deployment rate of LIBs across all applications depends on their cost-effectiveness, safety, and reliability. These factors heavily rely on accurate knowledge of the battery's state of health (SOH). Accurate SOH estimation allows for preventative maintenance, which in turn improves customer experience and supply chain management. Furthermore, with reliable SOH estimation, the battery management system (BMS) can implement health-aware control strategies to prolong battery life [9], potentially reducing over-engineering and lowering overall costs. Moreover, second-life applications for LIBs—such as second-hand EV pricing—become more viable, as SOH estimates serve as key indicators of remaining battery value. Finally, SOH estimation using real-world data enables population-level battery health analyses that map usage conditions to corresponding battery degradation patterns, providing valuable insights for next-generation battery design.

The SOH of a battery describes its ability to deliver energy and power relative to either its beginning-of-life (BoL) condition or manufacturer specifications. Two commonly used metrics for SOH are discharge capacity and internal resistance [10]. Battery degradation over time and usage leads to capacity fade and increased resistance. Generally, for applications where available energy is critical—such as electric vehicles (EVs)—capacity

is used to characterize SOH [11]. In contrast, in applications where power performance is more important—such as hybrid electric vehicles (HEVs) and grid frequency regulation—resistance is commonly used as the SOH metric [12]. Battery energy and power performance are also influenced by operating conditions. To tackle the needs for different applications, SOH metrics can be categorized into charge-based, energy-based and power-based SOH [13, 14, 15]. The most intuitive one is charge-based SOH, which measures how much charge (in ampere-hours, Ah) the battery can store relative to its nominal capacity.

$$\mathrm{SOH}_{\mathrm{charge}} = \frac{Q_{\mathrm{measured}}}{Q_{\mathrm{nominal}}} \times 100\% \tag{1.1}$$

The energy-based SOH considers not just the charge, but total energy output (in the unit of watt-hour). Sometimes, even the battery still holds nearly same charge, a lower discharge voltage curve can reduce total energy output. It is defined as,

$$\mathrm{SOH}_{\mathrm{energy}} = \frac{E_{\mathrm{measured}}}{E_{\mathrm{nominal}}} \times 100\%, \quad E = \int_T V(t) \cdot I(t) dt \tag{1.2}$$

where T means the whole discharging time period between given lower and upper voltage limits. The power-based SOH measures the maximum deliverable power relative to the original power capability. It is often used to describe how fast the battery can be chagred or discharged safely.

$$\mathrm{SOH}_{\mathrm{power}} = \frac{P_{\mathrm{max\,,\,measured}}}{P_{\mathrm{max\,,\,nominal}}} \times 100\%, \quad P_{\mathrm{max}} = \frac{V^2}{R_{\mathrm{internal}}} \tag{1.3}$$

Since the internal resistance R_{internal} may varies under different operation condition, P_{max} can also be experimentally measured under standardized load conditions. In this thesis, if without special explanation, SOH refers to the charge-based SOH definition.

An important concept related to SOH is end of life (EoL), which is typically defined by specific thresholds, such as capacity decreasing to 80% of the BoL value or resistance doubling. The remaining useful lifetime (RUL) is then defined as the number of

charge/discharge cycles (or full equivalent cycles) remaining before reaching EoL.

Both capacity and resistance can be relatively easy to measure under laboratory conditions [16]. Capacity is typically measured using the Coulomb-counting method during a galvanostatic-discharge experiment. Internal resistance can be measured through various methods, including electrochemical impedance spectroscopy (EIS) and hybrid pulse power characterization (HPPC). Resistance is also known to be highly dependent on the battery operating conditions, such as SOC level and temperature [17]. Unlike routine cycling, these experiments require specific test profiles or specialized equipment and are often regarded as standard 'reference performance tests' (RPTs).

Many high-quality, publicly available datasets exist that contain laboratory-generated aging data for lithium-ion batteries (LIBs) [18]. Periodic RPTs are conducted in these datasets to obtain ground-truth values for capacity and resistance. However, in real-world applications, RPTs are rarely feasible due to the need to interrupt normal operation and the time-consuming nature of such tests. Therefore, capacity and resistance typically need to be estimated or predicted rather than directly measured.

This thesis discusses two major problems related to battery SOH. First: given historical usage data and assuming a fixed future usage condition, what is the EoL or RUL point? This corresponds to the topic of battery lifetime prediction. When the historical usage data covers only the initial 5–10% of the battery's lifetime, the problem is often referred to as lifetime early prediction. Second, given present cycling data, what is the current SOH? This pertains to the topic of battery SOH estimation. When cycling data is available across the entire lifetime, it is also possible to estimate the full SOH trajectory. These two problems are closely related, as historical SOH estimates can be used to extrapolate and predict future EoL or RUL. In the remainder of this chapter, state-of-the-art methods for both lifetime prediction and SOH estimation are reviewed.

1.2 Methods for lifetime prediction

The problem of battery lifetime prediction can be framed as a regression task, where selected health indicators (features extracted from cycling data) are mapped to corre-

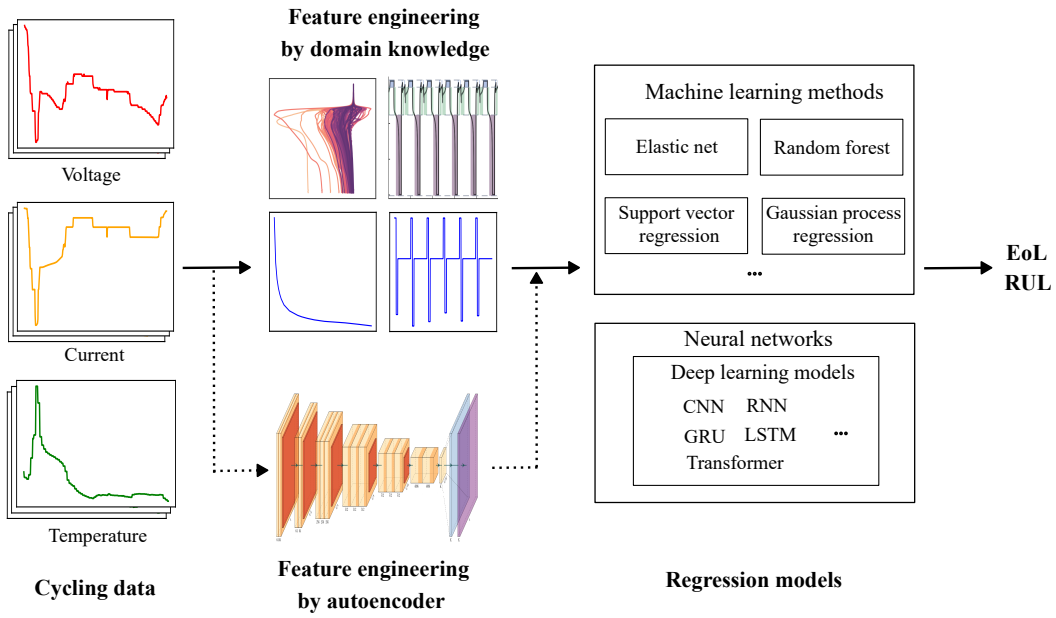


Figure 1.1: Schematic of methods for lifetime prediction. Health features are extracted from cycling data either by domain knowledge or with a neural-network autoencoder. These features serve as inputs to regression models to get EoL/RUL predictions.

sponding EoL or RUL labels. Therefore, health feature extraction and the development of regression models are the two main research directions. Several reviews [19, 20, 21, 22] have summarized the challenges and existing methods for EoL/RUL prediction. As previously mentioned, EoL/RUL predictions can also be obtained by extrapolating SOH estimates into the future. In such cases, techniques for SOH *estimation* (rather than prediction) become more important, and these are primarily discussed in Section 1.3. To avoid redundancy, this section focuses only on methods that directly map health features to EoL/RUL. Fig. 1.1 briefly outlines the logic flow for battery lifetime prediction. The feature engineering and regression models parts are discussed in detail below.

1.2.1 Health features

The most intuitive input features are the raw cycling signals, i.e., the applied current, terminal voltage, and temperature. Raw data is usually recorded at a much higher frequency than the timescale over which battery degradation occurs. In addition, a considerable amount of noise may exist in the raw signals. Thus, it is relatively difficult

to directly map raw cycling signals to corresponding EoL/RUL points. These signals are often passed through encoders [23] (e.g., neural network models [24]), where the inputs are transformed and optimized for EoL/RUL prediction.

Alternatively, representative features can be extracted from raw signals using domain knowledge of battery degradation. From the terminal voltage curve, multiple features have been proposed [19, 21], such as the voltage drop from pulse tests, relaxation time during the constant voltage stage, and the slope of the voltage curve under constant current. Research by Baumhöfer [25] describes a lifetime prediction model for 48 cells cycled under identical conditions. This model uses 24 early-life features (extracted from impedance spectra, pulse tests, and standard capacity tests), and accurate predictions were observed within just 16 cycles. However, further analysis showed that the model's accuracy was highly dependent on the number of features used, suggesting overfitting on the small dataset (i.e., 48 cells).

More recently, the discharge capacity vs. voltage ($Q(V)$) curve has gained increasing attention. Severson et al. [26] proposed an early-life prediction model using features derived from the $Q(V)$ curve. These features quantify the degradation rate by tracking the variation of the $Q(V)$ curve between cycles 10 and 100. Alongside the paper, a fast-aging battery dataset consisting of 124 lithium iron phosphate/graphite (LFP) cells was published. A follow-up study by Attia et al. [27] expanded the dataset to include 169 LFP cells cycled under 61 different fast-aging conditions. Many subsequent works have been based on this dataset, aiming to improve predictive performance and feature extraction techniques [28, 29, 30]. For example, in the work by Greenbank et al. [28], the time ratio spent in certain voltage windows is shown to correlate strongly with aging, enabling accurate EoL predictions.

Features not directly related to voltage—such as the number of cycles (or full equivalent cycles), Ah throughput, average surface temperature, and the rate of capacity drop—are also reported in the literature. Paulson et al. [31] provided a systematic feature engineering study that extracted features ranging from simple (e.g., changes in capacity and Coulombic efficiency) to more sophisticated (e.g., locations and magnitudes of dQ/dV and dV/dQ peaks), as well as parameters of equivalent circuit models (ECMs) fitted to

cycling data.

Alexis et al. [32] highlight the importance of classifying features based on whether they encode the influence of cycling conditions. Using the dataset published by Severson et al. [26], they show that features capturing cycling conditions significantly boost model performance. However, they also caution that such models may not generalize well to batteries undergoing identical cycling, as they primarily capture protocol-to-protocol variability rather than cell-to-cell variability. Similar arguments are presented by Zhou et al. [1], where health features are categorized as either individual-level or condition-level. The dependency between these two groups of features is explicitly modeled using a hierarchical Bayesian regression approach, enabling improved out-of-distribution prediction.

1.2.2 Regression methods

Various kinds of regression models have been proposed for lifetime prediction, including elastic net model [26], random forest [33], support vector regression [34], Gaussian process regression [35], and neural network models [36, 37].

Elastic net regression (ENR) is a regularized method that linearly combines penalties from both Lasso (L1) and Ridge (L2) techniques. It is particularly useful for handling high-dimensional data (i.e., many features) with multicollinearity. Severson et al. [26] used an elastic net model to map health features from early-life cycling (the first 100 cycles) into EoL cycle numbers. The model performance was evaluated using three different feature sets with increasing numbers of features. Using only a single powerful feature (the log variance of $Q(V)$ curve differences between cycles 10 and 100), the model achieved an average percentage error of 11% on test samples. With additional features related to capacity fade and temperature, the performance improved to an average percentage error of 9.1%.

Random forest (RF) models are a typical ensemble approach that aggregates the predictions from multiple decision trees. A decision tree is itself a non-parametric supervised learning method, which recursively split the dataset into subsets based on feature values,

forming a hierarchy of decision nodes that ultimately lead to prediction outcomes at the leaf nodes. In an RF, each tree is trained on a different random subset of features and input samples. By introducing such randomness, the RF is able to increase the diversity of decision trees and thus capture more patterns in the data with more robustness to overfitting. Niankai et al. [33] investigates the RUL prediction under partial charge and discharge with a limited degradation history. The RF is adopted to help prioritize different measurements and identify the least amount of operating data required for accurate RUL prediction. It is found the voltage signal between SOC range of [0.1, 0.2] contains the richest RUL-related information.

Support vector regression (SVR) adapts the principles of support vector machines (SVM) to regression tasks. It performs nonlinear regression by transforming the problem from a low-dimensional space to a high-dimensional space where it becomes linear. SVR approximates target values within an acceptable error margin. Meru et al. proposed a two-stage approach for RUL prediction [34]. In the first stage, SVM is used to achieve a coarse RUL classification based on features from discharge cycles. Depending on the number of completed cycles, this classification divides cells into 4 groups. In the second stage, SVR is adopted to achieve accurate prediction of RUL for those cells belonging to the group D with a relatively large number of cycles.

Gaussian process regression (GPR) is a Bayesian non-parametric approach. Instead of assuming a specific function form, GPR defines a prior over functions, and updates it to a predictive posterior after observing data. The kernel function measures similarity between inputs and defines the shape and smoothness of the model outputs [38]. Different choices of kernels for lifetime prediction are analyzed in the work of Richardson [35]. An additive kernel consisting of Matérn 3/2 and Matérn 5/2 kernels was reported with the best performance on SOH prediction. GPR has also been used to build the mapping between EIS spectra and RUL in the work of Yunwei et al [39]. Specifically, 20,000 EIS spectra of commercial Li-ion batteries were collected at different SOH, SOC and temperatures. GPR took the entire spectrum as input, and automatic relevance determination (ARD) enabled the finding of two important frequencies in the dataset out of 120 possibilities in the range of 0.02 Hz-20 kHz.

Thanks to the universal fitting ability of neural networks, as well as the fast development of model training techniques, neural network (NN) models are widely used for battery lifetime prediction. Ji Wu et al [36] implemented a fully-connected NN model to learn the relationship between the charge curve and corresponding RUL. In the work of Ning He et al. [37], graphical features were extracted from the $Q(V)$ curve, and a CNN model used to build the mapping between graphical features and corresponding RUL values. NN can also be used as an autoencoder for dimension reduction of features, generating more representative features [40].

All of the reviewed models in this section are 'supervised learning' methods, and this requires a sufficient amount of labeled data for model training. However, the model performance may deteriorate when predicting on batteries that undergo aging conditions different from those seen in the training data [20].

1.3 Methods for SOH estimation

The objective of SOH estimation is to provide a reliable diagnosis of the current battery health state from available measured data, i.e. terminal voltage, applied current and surface temperature. The granularity of these measurements may vary depending on sensors and applications. Reliable SOH estimation throughout battery usage is often a prerequisite for accurate RUL prediction.

Several reviews [41, 42] have summarized existing methods for SOH estimation. Most of them classify these methods into two categories: model-based and data-driven approaches. Fig. 1.2 briefly summarizes these methods following this common categorization. Both types of methods are discussed in detail below:

1.3.1 Model-based SOH estimation

Model-based battery SOH estimation employs a model to describe the input–output behaviour of a battery. The parameters of the model reflect the current SOH of the battery. Therefore, SOH estimation becomes a parameter identification problem. Repeated identification of model parameters enables tracking of battery degradation over its lifetime.

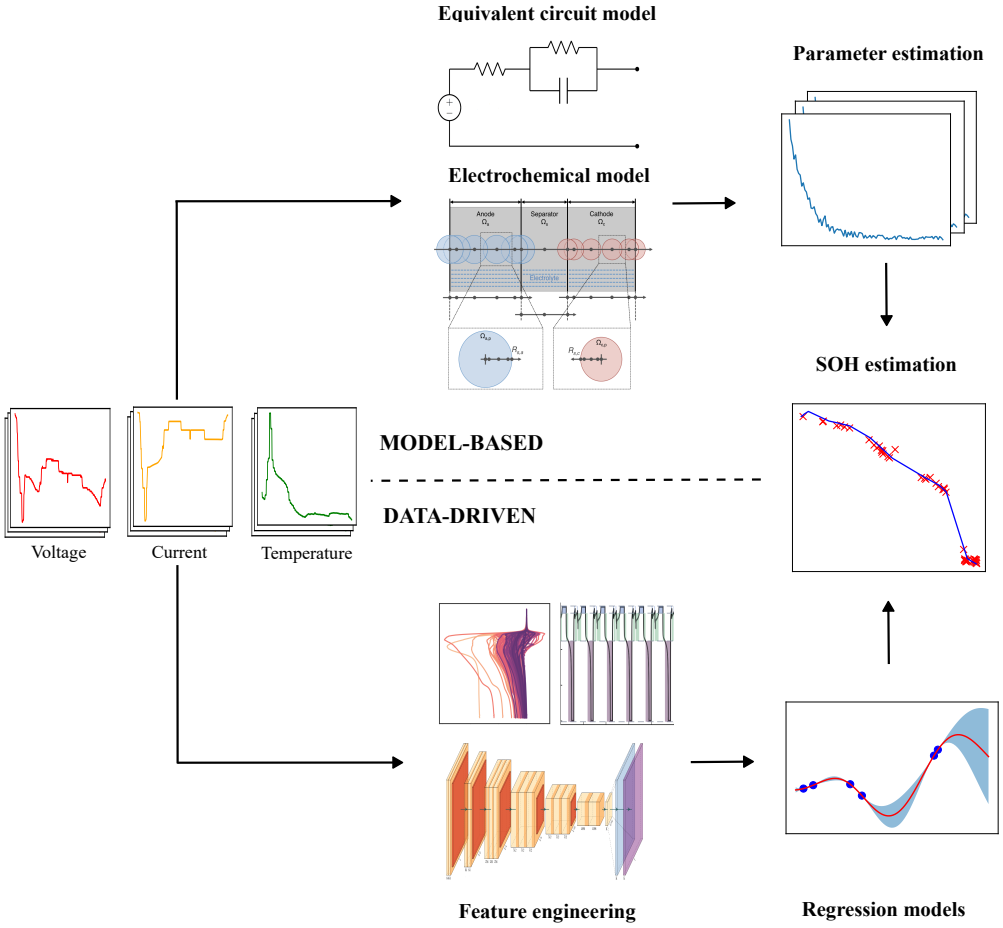


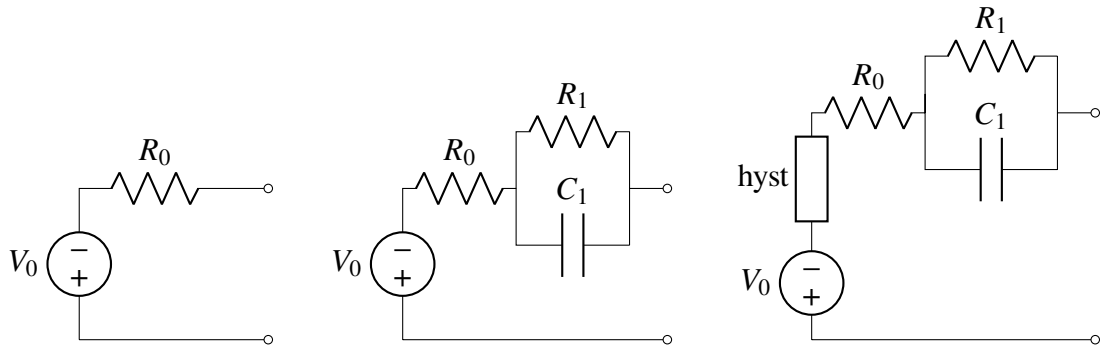
Figure 1.2: Schematic of SOH estimation approaches. The literature commonly divides these methods into model-based and data-driven categories. The key difference between the two is that model-based approaches employ explicit models for battery dynamics, whereas a data-driven framework uses regression models to map extracted features directly to SOH estimates. When full cycling data is available, both approaches can be used to reconstruct the entire SOH trajectory across the battery's lifetime.

Two main types of battery models are commonly used in the literature: equivalent circuit models (ECMs) and electrochemical models (EMs).

Due to their simplicity and computational efficiency, ECMs are the most commonly used battery models in the literature [43, 44, 45, 46, 47]. Ideal circuit elements—such as voltage sources, resistors, inductors, and capacitors—are used to approximate the electrical behaviour of the battery. Coupled thermal models have also been incorporated into ECMs to describe heat generation during cycling [48]. Depending on the number of included circuit elements, 2-3 ordinary differential equations are typically formulated to describe battery dynamics. Fig. 1.3 shows the model structure for three commonly used ECMs. The simplest ECM is the 'Rint' model, consisting of only a voltage source and a pure Ohmic resistance, which captures only instantaneous voltage drops. By including one RC pair to describe transient voltage behaviour, the Thevenin model has become one of the most widely adopted ECMs. Two or more RC pairs can be introduced to capture multiple electrochemical processes occurring at different timescales. The enhanced self-correcting cell (ESC) model proposed by Plett [49] further includes a hysteresis voltage term to model the voltage lagging behaviour under current direction switches.

Twelve types of ECMs were summarized and analyzed in the comparative study by Xiaosong et al. [50]. Although ECMs are relatively easy to parameterise, they often perform poorly when applied to operating conditions different from those under which they were calibrated. Gómez [51] highlighted that the temperature and SOC dependence of circuit parameters may vary by an order of magnitude, which can cause significant performance degradation if not correctly accounted for. Smiley [52] similarly pointed out that ECMs do not extrapolate well to operating conditions outside their training range.

Derived from first principles, EMs consider the actual electrochemical reaction processes in batteries, including solid-phase diffusion, liquid-phase diffusion, migration, and more. These models consist of coupled sets of partial differential equations that describe thermodynamics, reaction kinetics, and transport processes in electrodes, the electrolyte, and their interfaces. The pseudo-2-dimensional (P2D) model framework, proposed by Doyle, Fuller and Newman [53, 54], is the current gold standard EM for battery modeling.



(a): Rint (b): Thevenin (c): ESC

Figure 1.3: Schematic of different kinds of ECMs. (a) Rint (b) Thevenin (c) ESC.

It models battery geometry in two dimensions: one representing the macro-scale across battery thickness (x-direction) and another representing the micro-scale within porous electrodes (radial direction). The fully spatially resolved P2D model for Li-ion batteries consists of a system of non-linear partial differential equations, enabling detailed insights into the physical processes occurring during cycling. Since the parameters of the P2D model have direct physical interpretations, it can accurately predict voltage responses over a broad range of operating conditions.

However, the complexity and high computational cost of parameterising the P2D model are major limitations [55], which has driven significant research into simplified versions that retain sufficient accuracy [56]. The single-particle model with electrolyte (SPMe) is a widely used reduced-order version of the P2D model [57]. Assuming isotropy in the electrodes, it represents each porous electrode as a single representative particle. If electrolyte dynamics are further neglected, the SPMe reduces to the single-particle model (SPM) [58], which is the simplest EM for LIBs. A 4000-fold reduction in computational time for the SPM compared with the full P2D model has been reported in [59]. However, because the SPM ignores the Li^+ concentration and potential distributions in the electrolyte phase, it is generally only valid for low charge/discharge rates up to 1 C (1 C corresponds to fully charging or discharging a battery in one hour). For high C-rate charging/discharging, the model accuracy may become unacceptable. Nevertheless, studies have shown that the SPM can provide accurate voltage responses at rates up to about 4 C–6 C [60].

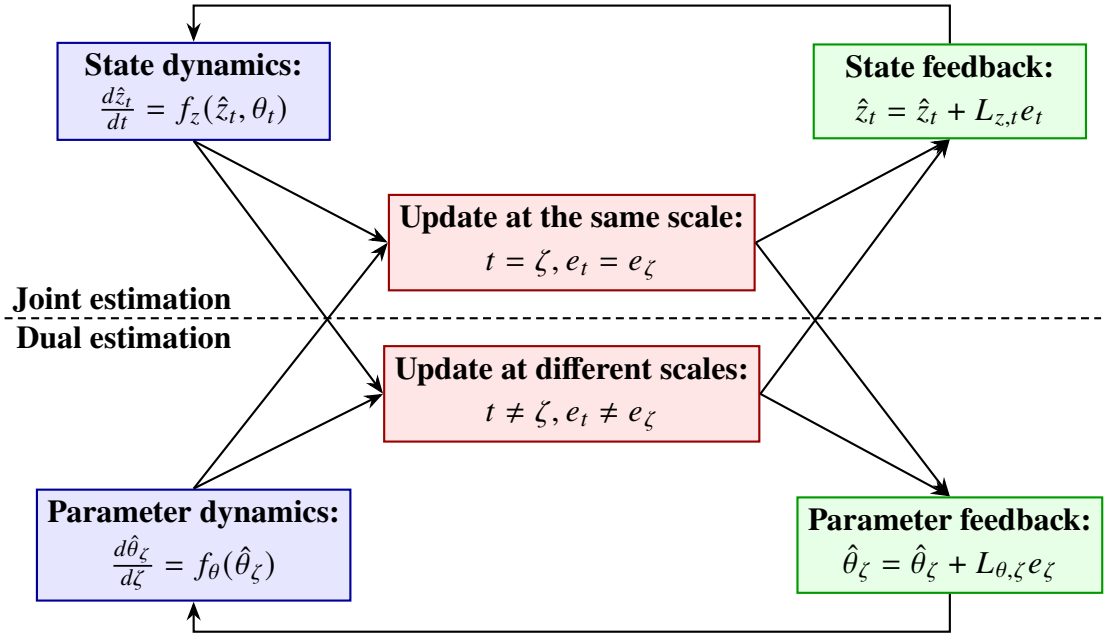


Figure 1.4: Schematic for recursive co-estimation of model states and parameters. The literature commonly divides these methods into joint and dual estimation. For joint estimation, states and parameters are propagated and update at the same scale. In dual estimation, states and parameters are estimated separately under different timescales.

Parameter identification for battery models involves finding an optimized parameter set that minimizes a cost function (e.g., least squares) against measurements, typically terminal voltage. With known SOC values, the recursive least squares (RLS) method and its variants are classic online parameter identification techniques for ECMs [61, 45]. Plett [62] proposed a recursive weighted total least squares method for capacity estimation that accounts for noise in both SOC and voltage measurements. However, in real-world applications, SOC, as a state of the ECM, also needs to be estimated from data. In this case, co-estimation of both model states and parameters becomes necessary.

The workflow for co-estimation involves recursively propagating a state-space model——defined by the battery model——forward through time-series data, while using error feedback to adaptively update both the states and parameters. Figure 1.4 schematically summarizes this process. Depending on whether updates of states and parameters are under the same time scale, two estimation paradigms are commonly used to achieve co-estimation of model states and parameters. One is joint estimation [63, 45], which combines states and parameters into an augmented state vector. A single nonlinear Kalman filter is then employed to estimate both states and parameters simultaneously.

Although joint estimation is straightforward to implement, it suffers from poor numerical conditioning due to the vastly different timescales of states and parameters [49].

The other is dual estimation [43, 64], which estimates states and parameters separately. Different timescales can be applied for states and parameters. However, dual estimation may be more numerically stable by time scale separation. Any cross-correlations between states and parameters are lost due to the decoupling of their dynamics [49]. In both joint and dual estimation methods, parameter dynamics are often modeled as a random walk process. There is no built-in guarantee that parameter estimates will converge to physically meaningful values, especially with poor initial guesses [64]. The limitations of field data—such as low resolution, missing data, and high sensor noise—further threaten the convergence and accuracy of SOH estimation using model-based methods.

1.3.2 Data-driven SOH estimation

Without explicitly modeling battery dynamics, data-driven methods regards SOH estimation as a regression task. The aim is to build a mapping between operational data (or features extracted from operational data) and SOH labels (often capacity). As battery degradation is a complex path-dependent process [65, 66], data-driven approaches have become increasingly popular due to the flexibility and ease of model parameterization—especially since model-based methods can be difficult to parameterize and may become overly complex when incorporating multiple interacting mechanisms [67].

Note that data-driven SOH estimation shares a very similar regression problem setup to the topic of battery lifetime prediction discussed in Section 1.2. The only distinction is that features are mapped to SOH labels rather than EoL/RUL labels. Therefore, the regression models discussed in Section 1.2.2 can also be used for data-driven SOH estimation. The discussion of feature extraction, which has already been covered, is not repeated here.

The first category of regression methods consists of empirical models. Usually, specific function forms, such as exponential [68] and polynomial functions [69], are used to build empirical aging models. These predefined functions map health features to SOH

metrics, while the coefficients of the function are learned from data. In the work of Wang et al. [70], a power-law function is used to describe the relationship between capacity loss and cycling time. The model is trained on cycling data across a range of temperatures (-30 to 60 °C), depths of discharge (10-90 %) and discharge rates (C/2 to 10C). Both calendar and cycling aging are modeled with exponential functions in the work of Petit et al. [71].

Machine learning regression models—such as the previously mentioned ENR, RF, GPR, and NN—are also used for SOH estimation. Each technique has its strengths and weaknesses in terms of robustness, computational cost, and model storage requirements, which are important considerations when models need to be adaptively retrained over battery lifetime [72]. Viewing feature evolution and capacity fade as time series, more sophisticated deep learning models have also been implemented [73, 74]. Li et al. [73] proposed an online capacity estimation approach based on a recurrent neural network with long short-term memory (LSTM) cells. Voltage–time sensor data from the partial constant-current phase of the charging curve is used as input, reflecting the availability of real-world signals. Jingyuan et al. [74] developed a specialized deep learning model for SOH estimation using transfer learning. Specifically, a transformer model with a multi-head ProbSparse self-attention mechanism is employed.

Data-driven models have shown strong performance on laboratory-generated datasets. However, their generalization ability to other datasets remains uncertain, particularly those involving batteries under previously unseen aging conditions. The mapping between health features and corresponding SOH may vary significantly across different usage scenarios. Since models are often trained on data from accelerated aging tests, their performance may degrade when applied to real-world observations. To improve model reliability, extensive sets of degradation experiments are needed to train data-driven models, which is often impractical. Moreover, the path-dependent nature of battery aging [66] further exacerbates this problem.

1.4 Thesis outline

This thesis is presented in an integrated format, consisting of five chapters that together provide a coherent narrative. Chapter 1 serves as the overall introduction. It includes a background description of the thesis topic through a comprehensive review of existing literature and remaining challenges.

Chapters 2–4 are research chapters, each categorized by a different topic. These chapters are based on papers published (or prepared) during the doctorate. Each chapter includes one or two self-contained papers grouped under the same research theme. A preface section is included at the beginning of each research chapter to better integrate the papers into the overall thesis structure.

In Chapter 2, the problem of battery lifetime early prediction is addressed. A hierarchical Bayesian regression model (HBM) is proposed in my first work [1] and validated on a publicly available fast-aging battery dataset. The core idea involves categorizing early-life features into two groups: individual cell-level features (reflecting intrinsic variability across cells) and population-level features (capturing the influence of cycling conditions on the average behaviour of the population). We extracted the 'SOC-average' charging C-rate as the population-level feature, which is related to the total ohmic heat generation during charging period. Individual-level features were extracted from the $Q(V)$ curve variation between cycles 10 and 100. The mapping between individual cell features and target labels (i.e. capacity) can vary under different cycling conditions, as determined by population-level features. The HBM is employed to capture this within-features dependency and enable accurate lifetime early prediction. The novel contributions of Chapter 2 are listed below:

- We proposed a new idea to categorize health features into two groups: cell-level and population-level features, reflecting inherent hierarchical structure within battery aging data.
- We first built a HBM for the problem of battery lifetime early prediction, which address the hierarchical nature of battery aging data and quantify the uncertainty

in lifetime predictions.

- We generated and publicly shared a large battery aging dataset consisting of 225 NMC cells cycled under a wide range of operating conditions. The proposed HBM is validated on this dataset in my second work [2].

Moving beyond point prediction in lifetime modeling, Chapters 3 and 4 address the problem of SOH trajectory estimation. In Chapter 3, my third work [3] introduces an aging-aware equivalent circuit model for health estimation that combines the flexibility of data-driven methods with a model-based framework. A simplified ECM is combined with Gaussian process regression to learn capacity fade over lifetime, as well as the dependence of resistance on operating conditions and lifetime. By expressing the GP in state-space form, a recursive co-estimation framework is established, where ECM states (e.g., SOC) and model parameters (represented by GPs) are jointly estimated from operational data. The novel contributions of Chapter 3 are listed below:

- We proposed an aging-aware battery ECM incorporating GPs to describe the dependency of circuit-model parameters on SOC and lifetime. This aging-aware ECM can be used to jointly estimate ECM states and model parameters from operational data.
- Incorporating the SOC and lifetime dependency of ECM parameters via GP kernels makes parameter estimates more stable over battery life compared to traditional random walk based methods.
- The estimated resistance function is found to be strongly influenced by aging-induced shape changes in OCV-SOC function, which leads to a way to further estimate differential voltage curves.

Unlike previous chapters, Chapter 4 presents SOH estimation results on real-world battery data. My fourth work [4] implements the aging-aware ECM on 114 LFP battery packs deployed in off-grid solar systems in Africa. Both capacity fade curves and corresponding degradation modes are estimated from field-collected cycling data. It is found that long time resting at high voltages can cause severe aging effects on loss of lithium inventory (LLI) and loss of active material for the negative electrode (LAM_{ne}).

The novel contributions of Chapter 4 are listed below:

- We analysed data from 114 commercial LFP battery packs deployed in sub-Saharan Africa for up to 1200 days of operation. Using the proposed aging-aware ECM, we estimate time-varying capacity, resistance, and degradation modes directly from operational data.
- We show that time spent resting at high voltage, reflecting periods of limited user demand, strongly accelerates degradation, whereas regular cycling mitigates rapid aging. This suggests practical steps to extend battery life, for example by adjusting control algorithms to avoid prolonged high-voltage rests.

Chapter 5 concludes the thesis by summarizing the research findings presented in the previous chapters and outlining directions for future work.

2 | Data-driven battery lifetime early prediction

2.1 Preface

This chapter is based on the following publications:

> [1] **Zhou Zihao**, and David A. Howey. "Bayesian hierarchical modelling for battery lifetime early prediction." *IFAC-PapersOnLine* 56.2 (2023): 6117-6123.
>
> [2] Li Tingkai*, **Zhou Zihao***, Adam Thelen, David A. Howey and Hu Chao. "Predicting battery lifetime under varying usage conditions from early aging data." *Cell Reports Physical Science* 5.4 (2024).
>
> *Equal contribution

In Chapter 2, our goal is to build a regression model for the problem of battery lifetime early prediction, which considers aging effects from two distinct causes: usage variability and intrinsic cell variability.

The influence of usage variability on battery aging can be clearly observed in the work of Severson et al. [26]. Their dataset was generated through accelerated aging experiments, where all batteries underwent identical discharge cycles at 4 C, but were subjected to different fast-charging protocols. The EoL times ranged from 5 to 80 days (150–2300 cycles), with an average lifetime of 26 days (797 cycles). Good prediction accuracy was achieved by using features extracted from shape changes in the $Q(V)$ curve. However, under such extreme usage conditions, a natural hypothesis is that usage-induced degradation is the dominant aging factor. The influence of intrinsic cell differences is largely suppressed, as all cells quickly failed due to the extremely high C-rate cycling. Alexis et al. [32] further showed that features encoding cycling conditions significantly improve model accuracy by capturing protocol-to-protocol variability.

On the other hand, lifetime differences are also observed among batteries cycled under identical protocols [25, 75, 76]. Although initial manufacturing variability among new cells is typically small, no significant correlation has been found experimentally between initial health and subsequent aging rates later in life [77, 78]. However, when coupled

with the effects of cycling aging, cell-to-cell variability increases. The shape change in the $Q(V)$ curve between cycles 10 and 100—identified as the source of the most effective health features [26]—captures a mixture of usage-induced degradation and intrinsic cell variability.

The absence of explicit modeling for these two aging factors (usage variability and intrinsic cell variability) will reduce a model's ability to generalize to batteries under unseen usage conditions. The learned mapping between health features and EoL/RUL labels may only hold under the specific usage conditions included in training set. In my first publication [1], we present a hierarchical Bayesian model (HBM) that explicitly models the changing relationship between features and labels across different cycling conditions. The model performance is validated on the public dataset introduced by Severson et al. [26]. The second publication [2] further tests the model's generalization ability under a wide range of aging conditions using a self-collected dataset of 225 NMC batteries and includes a detailed discussion of more sophisticated early-life features.

2.2 Manuscripts

For convenience, the used benchmark methods in both papers of Chapter 2 are summarized here for reference. Details can also be found within each paper.

In the first paper [1], the baseline model is a ridge regression model with two different feature settings. One with 6 features used by Severson [26], it is denoted as baseline (original). Another with 4 proposed features (both conditional level and individual cell level features), it is denoted as baseline (with feature g). In the second paper [41]. Though ridge regression is a traditional method, it is commonly chosen as one of baselines in the topic of battery lifetime prediction.

For the second paper [2], multiple baseline models are implemented under different input feature settings. The first baseline model is called a "Dummy" model, which does not use any input features or have any trainable parameters, instead it predicts the mean cell lifetime of the training set for all cells. We use it as a way to determine if a more complex model is truly learning new information from the input data or instead only

appears to be learning because of similar training/test dataset distributions that lead to similar error metrics. The second baseline is called "Cycling conditions" model, which is a ridge regression model using only the cycling condition parameters as input features. It is used to show the contribution of condition level features on lifetime prediction. The third baseline is called a "Discharge" model, which corresponds to the elastic net model used in Severson's work [26] using the same feature set. The fourth baseline is called "Degradation-informed" model, which is also a elastic net model but built on the newly proposed features (both individual cell level and condition level features).

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 6117–6123

Bayesian hierarchical modelling for battery lifetime early prediction

Zihao Zhou * **David A. Howey** *

* *University of Oxford, Parks Road, Oxford OX1 3PJ, UK (e-mail: david.howey@eng.ox.ac.uk).*

Abstract: Accurate prediction of battery health is essential for real-world system management and lab-based experiment design. However, building a life-prediction model from different cycling conditions is still a challenge. Large lifetime variability results from both cycling conditions and initial manufacturing variability, and this—along with the limited experimental resources usually available for each cycling condition—makes data-driven lifetime prediction challenging. Here, a hierarchical Bayesian linear model is proposed for battery life prediction, combining both individual cell features (reflecting manufacturing variability) with population-wide features (reflecting the impact of cycling conditions on the population average). The individual features were collected from the first 100 cycles of data, which is around 5-10% of lifetime. The model is able to predict end of life with a root mean square error of 3.2 days and mean absolute percentage error of 8.6%, measured through 5-fold cross-validation, overperforming the baseline (non-hierarchical) model by around 12-13%.

Keywords: Bayesian; lithium-ion battery; hierarchical model; lifetime variability

1. INTRODUCTION

Lithium-ion batteries are ubiquitous due to their relatively long lifetime and high energy density (Cano et al. (2018); Schmuch et al. (2018)), but their performance degrades with time and usage. Battery aging behaviours have been widely explored (Birkl et al. (2017)), but remain difficult to understand and quantify since batteries are complex electrochemical devices that operate in widely varying conditions.

Existing battery lifetime prediction models can be roughly divided into either physics-based (Reniers et al. (2019)) or data-driven approaches (Sulzer et al. (2021a)). Physics-based models are built from first principles, accounting for diverse aging mechanisms such as growth of the solid-electrolyte interphase (Liu et al. (2020)), lithium plating (Liu et al. (2016)), and particle cracking (Ai et al. (2019)). While these are good at explaining underlying aging mechanisms, developing models that predict behaviour under different cycling conditions is challenging, since the influential degradation mechanisms may be vary across cycling conditions (Su et al. (2016); Raj et al. (2020)).

Data-driven methods directly map from chosen health indicators (features) to battery lifetime (labels) without requiring underlying domain knowledge. For this purpose, (Severson et al. (2019)) proposed an input feature based on discharging curve differences in early cycles, called $\Delta Q(V)$, giving accurate predictions of lifetime using only early cycling data. Subsequently, other aging features have been proposed based on charging or discharging curve differences (Paulson et al. (2022)). However, the effectiveness of these may rely on the specific cycling conditions (Sulzer et al. (2021b)). Alternatively, instead of generating features from the raw signals (voltage/current/temperature) and then using these as inputs for a model, other works implemented sophisticated machine learning methods to directly build models from raw measurements, such as ensemble learning (Li et al. (2019)) and long-short memory networks (Zhang et al. (2018)).

Although data-driven models achieve satisfactory results on their datasets, their ability to generalize is still unproven. Most of these models are built at a population level, i.e. all battery cells within a dataset follow the same feature-label mapping function, or to put it another way, the model that is learnt is an *average* model for the whole cell population. As a result, these models often lack the ability to give good performance on *individual* cells going through unseen cycling conditions. Recent work (Dechent et al. (2021); Strange et al. (2022)) also shows that cell-to-cell intrinsic manufacturing variability may not be properly addressed by a single population model. To address this, (Deng et al. (2022)) clustered cells under similar cycling conditions into subgroups, so that the usage variability could be reduced within each subgroup. Then, separate models were built for each subgroup. However, this separate modelling strategy needs to reach a balance between reducing within-group usage variability (more groups) versus group sample size (fewer groups). In reality, separate subgroup models often suffer from a lack of available data and tend to over-fit because of limited samples.

Different from either a single general population model or separate individual models, there is a reasonable middle ground in the form of hierarchical modelling (also called multi-level modelling). Hierarchical models are statistical models that have parameters at more than one level; they have proven to be a useful tool to deal with naturally structured data. For example, in a two-layer model there are pa-

10.1016/j.ifacol.2023.10.708

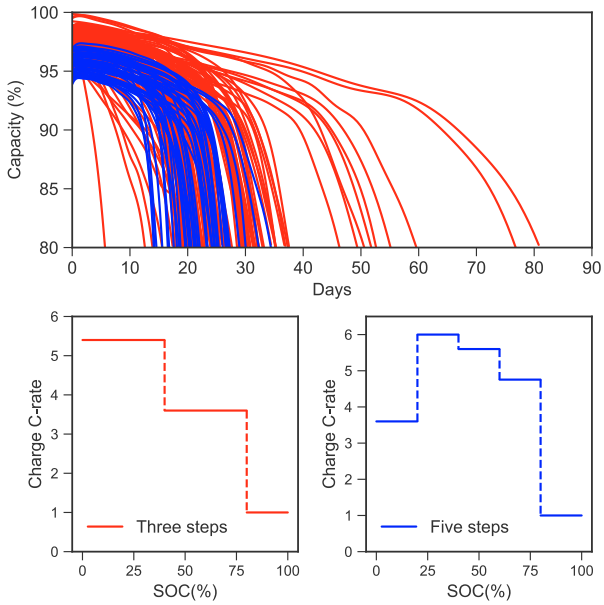


Fig. 1. Datasets of (Severson et al. (2019)) (red, three steps) and (Attia et al. (2020)) (blue, five steps) used here; top shows measured ageing trajectories, bottom shows exemplary individual charging protocols from each dataset respectively; in both cases the final step is a 1C constant current charge from 80–100% SOC.

rameters at the lower level which are assumed to be drawn from distributions, and these distributions are themselves parameterised by higher level parameters (Gelman and Hill (2006)). Hierarchical models have been widely used in different areas including ecology, psychology, sociology and computer vision (Lake et al. (2015); Pedersen et al. (2019)).

In this work, a hierarchical Bayesian model (HBM) is presented to achieve battery life prediction from early life measurements under varying cycling conditions. The individual cell feature-label relationships can vary across different cycling conditions according to cycling condition variables. Here we combine the MIT battery ageing dataset (Severson et al. (2019)) with extra samples from (Attia et al. (2020)) to generate a dataset consisting of 169 cells from 61 different cycling conditions. The lifetimes of these 169 cells range from 5 to 80 days. The individual battery health indicators (features) were calculated from the first 100 cycles of cycling data, which is only around 5-10% of a cell's average lifetime. We show that the HBM overperforms a baseline ridge regression model by 12% relative root-mean-square error (RMSE) and 13% mean-percentage-error (MPE).

2. DATA SOURCES AND FEATURE EXTRACTION

2.1 Data sources

In order to explore the impact of usage variability, different cycling conditions need to be included in the selected dataset. Here, the open source battery cycling datasets of (Severson et al. (2019)) and (Attia et al. (2020)) are combined together to give a dataset of 169 cells in total. The first dataset (Severson et al. (2019)) consists of 124 lithium iron phosphate/graphite 18650 Li-ion cells (A123) that were cycled at 30 °C. The second dataset (Attia et al. (2020)) is a follow up experiment that contains 45 cells of the same type as before. All these cells underwent identical discharge cycles at 4C but had different fast charging protocols. (In this paper C-rate has the conventional definition, i.e. charge or discharge current over nominal capacity.) The cells had a 1.1 Ah nominal capacity and the end of life (EoL) is defined as 80% of nominal capacity remaining. Cycling data (voltage/current/temperature) with time resolution of 1 second are included. More details can be found in (Severson et al. (2019)) As shown in Fig 1, the EoL time covers a wide range from 5 to 80 days (150–2300 cycles), with an average lifetime of 26 days (797 cycles). Cells from the first dataset experienced charging protocols with three steps, while cells from the second set experienced five-step charging protocols.

2.2 Feature extraction

Feature engineering is not the main focus of this work, hence we chose a small number of features at both the group and individual levels in the hierarchical model, guided by existing literature and our prior understanding of battery ageing. In the datasets considered, the discharging protocol is uniform across all cells, while the charging protocols are quite different from cell to cell. The main difference between them is the sequence of C-rates used during charging, so one way to compare them is to calculate some kind of average charging C-rate for each protocol. Here we used the 'SOC-average' charging C-rate as a metric to define differences between protocols—this is related to the total ohmic heat generation (I^2RT, where R is resistance) in the cell during each charging period, normalised by capacity, since

$$\begin{aligned} \frac{d(\mathrm{SOC})}{dt} &= \frac{I}{Q_0}, \quad d(\mathrm{SOC}) = \frac{I}{Q_0}dt \\ \text{hence } \int_0^1 I d(\mathrm{SOC}) &= \int_0^T I^2 \frac{1}{Q_0} dt, \end{aligned} \tag{1}$$

where the Q_0 is the nominal capacity, T the total charge time, and I indicates current. Taking the three steps protocol shown in Fig 1 as an example, the cell is first charged to 40% SOC at 5.4C, and then charged from 40% to 80% SOC at 3.6C. So, the average charging C-rate for this charging protocol is

$$\begin{aligned} &q_1 = 40\%, \quad q_2 = 40\%, \quad q_3 = 20\% \\ &5.4 \times q_1 + 3.6 \times q_2 + 1 \times q_3 = 3.8\mathrm{C}. \end{aligned} \tag{2}$$

Note that the final 80% to 100% SOC stage is always at 1C for all cells within the dataset.

As shown in Fig 2, the average charging C-rate has a strong predictive relationship with the end-of-life (EoL) time in this combined dataset. Multiple cells may have similar or the same average charging C-rate, and these cells will be clustered together into groups later. We denote this average charging C-rate metric as a group level feature g, as shown in Table 1.

At the individual cell level, three features which are widely used in existing literature (Severson et al. (2019); Paulson et al. (2022); Fei et al. (2021)) were calculated from the first 100 cycles of data for every cell. These are

Table 1. Input features used in his work

Symbol	Description	Level
g	Average charging C-rate	Group
F1	Variance of discharge $\Delta Q(V)$ curve between 10th and 100th cycles	Individual
F2	Minimum of discharge $\Delta Q(V)$ curve between 10th and 100th cycles	Individual
F3	Discharge capacity at cycle 2	Individual

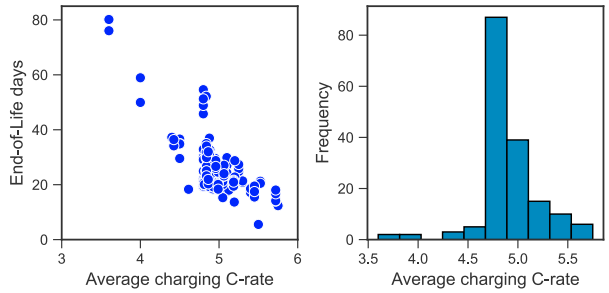


Fig. 2. Left shows the relationship between the average charging C-rate and EoL time in days; the absolute linear correlation coefficient is 0.71. Right shows histogram of calculated average charging C-rates.

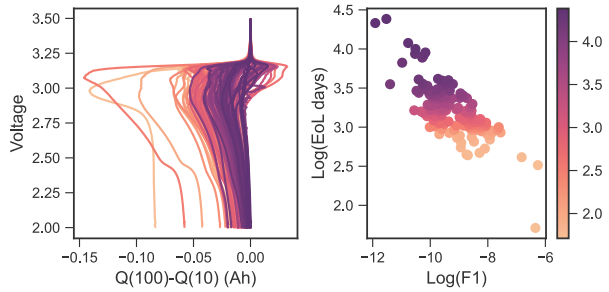


Fig. 3. Left shows the feature 'F1', the $\Delta Q(V)$ curve between the 10th and 100th discharge cycles; right shows the log-log relationship between the most predictive feature (F1) and the EoL in days. As already reported (Severson et al. (2019)), the correlation is strong, in this case the absolute value of Pearson's linear coefficient is 0.76.

denoted as individual-level features (F1-F3) in Table 1. The most predictive feature (the one having the largest linear coefficient with log of EoL time), and its relationship with log EoL time, is shown in Fig. 3.

3. METHODOLOGY

The overall lifetime variability consists of variability caused by usage differences, and variability from manufacturing differences. Ideally there would be many samples for every usage condition so that the cell-to-cell differences under usage could be estimated. However, in the dataset used, there are only 1-2 cells tested for each different protocol—a sample size too small to give reliable conclusions. To address this, we used constrained K-means clustering to gather cells into usage-related groups each having similar (although not necessarily identical) average charging C-rates.

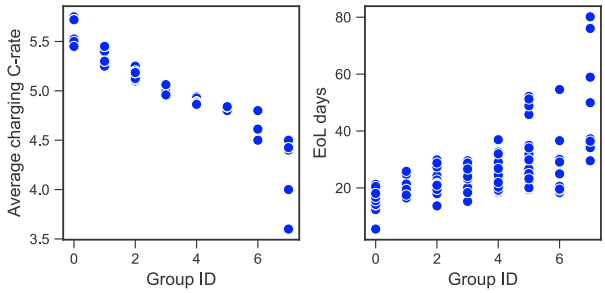


Fig. 4. Left shows average charging C-rate for each group with cells clustered into eight groups; right shows associated EoL days for each group. There is relatively large usage variability within the final group due to lack of cells at low (<3.5C) charging C-rates.

3.1 Cycling conditions clustering

A commonly used technique to address the balance between individual cell behaviour vs. population model accuracy is clustering, i.e. grouping of cells into 'similar' behavioural subgroups that each contain a relatively large number of cells. However, most previous works (Deng et al. (2022); Jiang et al. (2021)) cluster cells based on features from individual cell voltage signals (or cell EoL lifetime) rather than features representing usage. As a result, cells undergoing quite different cycling conditions may be grouped together. In this work, clustering uses the cycling condition level features (g), which enables cells with similar cycling conditions to be grouped together.

The tool used is constrained K-means clustering (Bhattacharya et al. (2018)). It is an improved version of K-means that avoids local solutions containing empty clusters, or clusters having very few samples. The goal is to have sufficient sample size (at least 10 cells) in each group, while also maintaining small usage variability within each group (at most 100 cells).

There are 169 cells with 61 different average charging C-rates in the dataset, so a reasonable number of clusters is $K \in [4, 16]$. Here, $K = 8$ is used as an example (further investigation of the influence of cluster number would be interesting future work). The corresponding clustering result is shown in Fig 4.

3.2 Bayesian hierarchical linear model

We now discuss the construction of the hierarchical model for life prediction. To motivate the need for a hierarchical model, one should first check whether the group level effect is significant in the target dataset. The variance partition coefficient (VPC), Equation 3, is a metric which represents

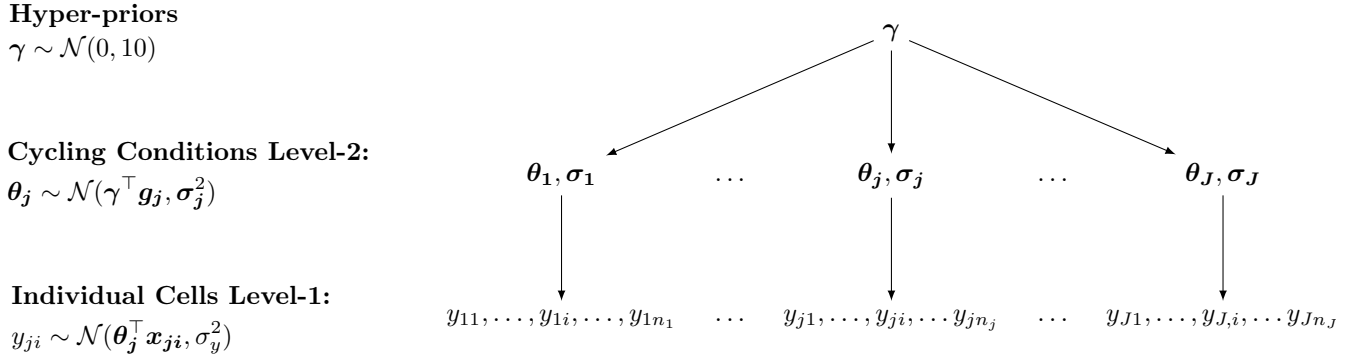


Fig. 5. Hierarchical model structure: parameters can be divided into individual cell level parameters $(\boldsymbol{\theta_j}, \boldsymbol{\sigma_j})$ and cycling condition level hyper-parameters $(\boldsymbol{\gamma})$. While j represents cycling condition group index, i represents individual cell index, y_{ji} represents lifetime of ith cell in jth cycling group. This two level structure allows the individual cell level feature-label $(x_{ji} - y_{ji})$ relationship to vary across different cycling conditions based on cycling condition level features $(\boldsymbol{g_j})$.

the percentage of label variance explained by including a grouping effect.

$$\begin{aligned} \text{VPC} &= \frac{\sigma^2_{\text{group}}}{\sigma^2_{\text{individual}} + \sigma^2_{\text{group}}} \\ \sigma^2_{\text{group}} &= \text{var}\left(\bar{u}_j - \bar{u}\right), \bar{u} = \frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{n_j} y_{ji} \\ \sigma^2_{\text{individual}} &= \text{var}\left(y_{ji} - \bar{u}_j\right), \bar{u}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ji} \end{aligned} \tag{3}$$

Here, σ_{group} defines the variance between groups, and $\sigma_{\text{individual}}$ defines the variance of the whole population where each cell is compared to the mean of its group, $\bar{u}$ is the average lifetime for all cells in the population, $\bar{\mu}_j$ is the average lifetime of the jth group, var has the usual definition of sample variance, y_{ji} is all individual lifetimes for the ith cell within the jth group, n_j is the total number of cells within the jth group and N is the total number of cells in the population. When the group variance $\sigma^2_{\text{group}} = 0$, there is no difference between groups, that is, every group has same mean lifetime compared to the overall population battery mean lifetime. When the $\sigma^2_{\text{individual}} = 0$, there is no differences within a group, that is, within each group, samples share the same lifetime value. For our dataset, VPC = 74.5%, which denotes that 74.5% of the lifetime variance is caused by the cycling condition group effect.

Given the large influence from cycling condition group, it is reasonable to apply a hierarchical model. The generative model structure is described by the schematic in Fig. 3.2. In this framework, within the jth group, the first (lower) level inference is done for individual cell lifetime regression parameters $(\boldsymbol{\theta_j})$ where θ_{j0} is a scalar representing the mean lifetime within the jth group and $\boldsymbol{\theta_j}$ is a 4×1 vector because the number of features is three. In this case because we only have three features, we could have explicitly written the lower level forward model for the jth group and ith cell within that group as

$$y_{ij} \sim \mathcal{N}(\theta_{j0} + \theta_{j1}x^j_{i1} + \theta_{j2}x^j_{i2} + \theta_{j3}x^j_{i3}, \sigma_y^2) \tag{4}$$

Since the EoL time y_{ij} is accurately measured in the dataset, the sample noise variance σ_y was fixed to 1 here (it can also be assigned as a distribution and learnt from the dataset). For the sake of flexibility and compactness we use vector notion. The second (upper) level inference is undertaken for cycling condition parameters $(\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\gamma_3})$, where the $\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\gamma_3}$ parameters are each 2×1 vectors (because each group has one group level feature and an intercept parameter), defined as

$$\begin{aligned} &\boldsymbol{\theta_j} = \begin{bmatrix} \theta_{j0} \\ \theta_{j1} \\ \theta_{j2} \\ \theta_{j3} \end{bmatrix} \boldsymbol{g_j} = \begin{bmatrix} 1 \\ \frac{1}{n_j}\sum_{i=1}^{n_j} g_{ji} \end{bmatrix} \boldsymbol{x_{ji}} = \begin{bmatrix} 1 \\ x^j_{i1} \\ x^j_{i2} \\ x^j_{i3} \end{bmatrix} \\ &\boldsymbol{\gamma_0} = \begin{bmatrix} \gamma_{0,0} \\ \gamma_{0,g} \end{bmatrix} \boldsymbol{\gamma_1} = \begin{bmatrix} \gamma_{1,0} \\ \gamma_{1,g} \end{bmatrix} \boldsymbol{\gamma_2} = \begin{bmatrix} \gamma_{2,0} \\ \gamma_{2,g} \end{bmatrix} \boldsymbol{\gamma_3} = \begin{bmatrix} \gamma_{3,0} \\ \gamma_{3,g} \end{bmatrix} \\ &\gamma = [\,\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \boldsymbol{\gamma_3}\,], Y_j = \left(y_{j1}, \ldots, y_{jn_j}\right) \\ &\{Y\} = \{Y_1, \ldots Y_j\}, \end{aligned} \tag{5}$$

where Y_j is an output vector of lifetimes for all cells within the jth group.

The individual slopes of each group model, $\boldsymbol{\theta_j}$, are assumed to be drawn from normal distributions centered at $\boldsymbol{\gamma^\top g_j}$,

$$\boldsymbol{\theta_j} \sim \mathcal{N}\left(\boldsymbol{\gamma^\top g_j}, \boldsymbol{\sigma_j^2}\right), \tag{6}$$

where $\boldsymbol{g_j}$ is a 2×1 vector of known specific mean features for the jth group (i.e., the mean of the individual average charging C-rate for each group). In the top layer, weakly informative (i.e., wide) priors $\mathcal{N}(0, 10)$ are assigned for all hyper-parameters. And for setting the priors for θ_j, a non-centered reparameterization trick, proposed by (Betancourt and Girolami (2015)), is applied to improve sampler efficiency.

Our objective is to infer posterior distributions for both the group models and the population model, $P\left(\boldsymbol{\theta_j} \mid Y_j\right)$ and $P\left(\boldsymbol{\gamma} \mid \{Y\}\right)$ respectively. The overall process is

(1) Calculate level-2 posterior distribution $P\left(\boldsymbol{\gamma} \mid \{Y\}\right)$;
(2) Use level-2 posterior distribution as prior for level-1 parameters, calculate level-1 posterior distribution $P\left(\boldsymbol{\theta_j} \mid \boldsymbol{\gamma}, Y_j\right)$;
(3) Use level-1 posterior distribution to make prediction on individual labels.

Using Bayes' rule, the level-2 posterior can be estimated by multiplying the likelihood function with the hyper-prior:

$$\begin{aligned} P(\boldsymbol{\gamma} \mid \{Y\}) &\propto \text{ Likelihood } \times \text{Prior} \\ \text{Likelihood } &= \prod_{j=1}^{J} P(Y_j \mid \boldsymbol{\gamma}) \\ \text{Prior} &= \mathcal{N}(\mathbf{0}, S_0), S_0 = 10 \cdot I \end{aligned} \tag{7}$$

Since the level-2 hyperparameters $\boldsymbol{\gamma}$ are not directly related to individual observations, the likelihood function can be calculated by integrating out the level-1 parameters, giving

$$\begin{aligned} P(Y_j \mid \boldsymbol{\gamma}) &= \int_{\boldsymbol{\theta_j}} P(Y_j \mid \boldsymbol{\theta_j}) \cdot P(\boldsymbol{\theta_j} \mid \boldsymbol{\gamma})\, d\boldsymbol{\theta_j} \\ P(Y_j \mid \boldsymbol{\theta_j}) &= \prod_{i=1}^{n_j} \mathcal{N}(\mathrm{err}_{\mathrm{i}}(\boldsymbol{\theta_j}); 0, 1) \\ \mathrm{err}_{\mathrm{i}}(\boldsymbol{\theta_j}) &= \left\{ \boldsymbol{\theta_j}^{\top} x_{ij} - y_{ij} \right\}. \end{aligned} \tag{8}$$

Notice that the level-2 likelihood component $P(Y_j \mid \boldsymbol{\gamma})$ is the model evidence (marginal likelihood) for the level-1 model of jth subgroup, which indicates there is an averaging effect for level-1 model selection of each subgroup in our proposed two level structure.

Given the level-2 likelihood expression (8), the posterior distribution for hyper-parameters $P(\boldsymbol{\gamma} \mid \{Y\})$ can be estimated by Markov chain Monte Carlo (MCMC). After that, the posterior distributions for level-1 parameters $P(\boldsymbol{\theta_j} \mid \boldsymbol{\gamma}, Y_j)$ can be estimated by applying Bayes' rule,

$$P(\boldsymbol{\theta_j} \mid \boldsymbol{\gamma}, Y_j) = \frac{P(Y_j \mid \boldsymbol{\theta_j}) \cdot P(\boldsymbol{\theta_j} \mid \boldsymbol{\gamma}) \cdot P(\boldsymbol{\gamma})}{P(Y_j \mid \boldsymbol{\gamma})}. \tag{9}$$

The whole inference process was performed in Python using the PYMC3 package Salvatier et al. (2016).

3.3 Baseline comparison and evaluation metrics

The performance of the approach was evaluated by comparing against a baseline ridge regression model under two different error metrics, defined below. To give a richer comparison, the ridge model was built on two different feature sets: one has 6 features (denoted as the "discharge" model in Severson et al. (2019)), and the other has 4 features (including g and F1-F3, the same features as the HBM). Notice that the 6 features used in the "discharge" model also include F1-F3 used in this paper.

The first error metric is the root mean square error (RMSE) of the EoL predictions. This is a commonly used metric for predictive performance, however, it is significantly influenced by outliers. Therefore, the second metric is mean absolute percentage error (MAPE), which measures the percentage difference in EoL time between the predicted y_{pred} and observed end of life y_{real}. The two are defined by

$$\begin{aligned} \text{RMSE } &= \sqrt{\frac{\sum_{i=1}^{N} \left(y^i_{\text{true}} - y^i_{\text{pred}} \right)^2}{N}}, \\ \text{MAPE } &= \frac{100}{N} \sum_{i=1}^{N} \left| \frac{y^i_{\text{true}} - y^i_{\text{pred}}}{y^i_{\text{true}}} \right|. \end{aligned} \tag{10}$$

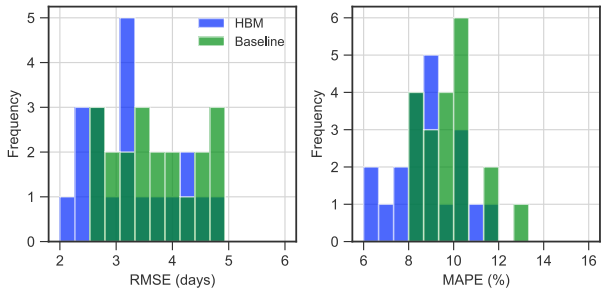


Fig. 6. Histograms of all results, as RMSE and MAPE

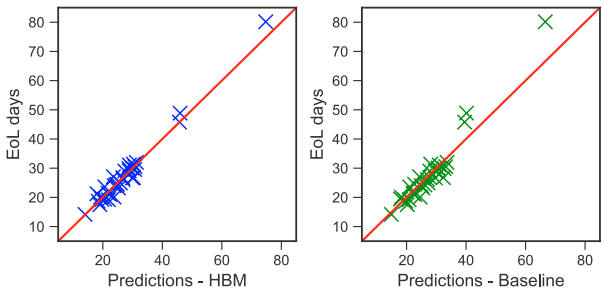


Fig. 7. Scatter plot of predicted versus actual EoL (days) for HBM vs. baseline model

K-fold cross validation is a commonly used paradigm to evaluate predictive ability (Arlot and Celisse (2010)). Here, the train and test sets were selected with 5-fold cross-validation, resulting in a 135 cells training set and a 34 cells test set. Then, 5-fold cross validation was performed 4 times (totalling 20 RMSE values) to observe the general predictive ability of the proposed model.

Table 2 summarises the results. The HBM gives 3.20 days median RMSE and 8.6% median MAPE, surpassing the baseline model on both feature sets. Notice that, by including usage feature g, the performance of the baseline model is also improved, which indicates the importance of cycling condition as a feature. More visualization of the results is shown in Fig. 6. For the HBM, the majority (70%) of the RMSE values on EoL time are below 3.5 days. For the baseline model, 50% of trials give RMSE values larger than 3.5 weeks. The case for MAPE is also similar. The HBM generally gives around 12-13% improvement in performance compared to the original baseline model in both RMSE and MAPE.

4. RESULTS AND DISCUSSION

Fig. 7 shows a scatter plot of predictions versus actual EoL time from one trial for the HBM and the baseline model. Both give relatively good predictions in the mid-life range (around 30 days). However, the baseline model fails to give good predictions for long-life (>40 days) cells. The HBM has a reasonably consistent predictive performance across the whole lifetime range.

The superior performance of the proposed HBM can be explained by its ability to account for relationships between different levels of features, or in other words, the inherent structure within the dataset. The cycling condition level feature influences the relationships between

Table 2. Summary of model predictive performance (RMSE is in days)

		Baseline (original)	Baseline (with feature g)	HBM	Improvement
RMSE	Median	3.67	3.38	3.20	12.8%
	Mean	3.72	3.49	3.26	12.4%
MPE	Median	9.9%	9.1%	8.6%	13.1%
	Mean	9.9%	9.2%	8.7%	12.1%

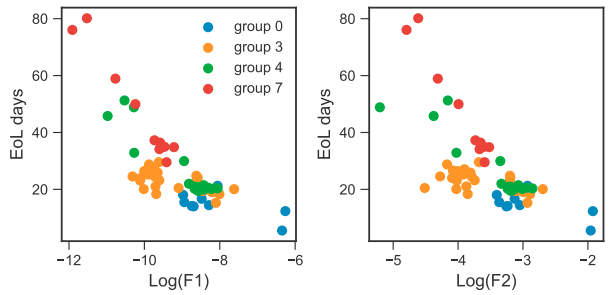


Fig. 8. Relationships between EoL vs. individual cell features and group labels based on different cycling conditions. Groups $0, 3, 4, 7$ are selected to show the influence of cycling conditions. See also Fig 4.

individual cell level features. Fig. 8 shows cells from four different cycling condition clusters—while all of them share negative correlations between features and labels, their slopes and input feature ranges are quite different. As the average charging C-rate decreases from group 0 to group 7, the slope become steeper and the range of log F1 values decreases.

In a conventional regression setting, the cycling condition features and individual cell features are treated in the same way, assuming they are independent. As a result, the inherent relationships between these two different categories of features are ignored. In contrast, in hierarchical approaches, this kind of inherent relationship is explicitly modelled, which enables better model performance when facing structured data.

5. CONCLUSION

A hierarchical Bayesian linear model was proposed to address the problem of battery early life prediction under varying usage conditions and using only the first 100 cycles data (only 5-10% of the entire life) as inputs. The proposed HBM was tested on a open data set consisting of 169 cells experiencing 61 different cycling protocols. In a 5-fold cross-validation experiment, the HBM gives a 3.20 days median RMSE and a 8.6% median MAPE, which overperforms the baseline model by 12%. These results show the effectiveness and further potential of the hierarchical models for battery early life prediction.

There are two directions for future work. First, the influence of cluster number and cluster size needs to be further investigated, as these two parameters are known to be important in many other fields when using hierarchical models (Gelman and Hill (2006)). In our dataset, eight clusters is a reasonable number to balance within-group usage variability versus sample size. However, this was an empirical choice based mainly on trial and error. Second, more different usage conditions need to be included. In the used dataset, the usage differences only locate in charging period. The influence from discharge C-rate, depth of discharge and rest periods are not included. More useful group level features may be found by including more usage conditions. Third, nonlinear functions, or even non-parametric models such as Gaussian processes, could be implemented hierarchically, and may give better performance compared to linear regression models. The codes for this work is available online (Zhou (2023)). And a further collaboration work with Prof. Hu Chao's group aims to address above challenges is under preparation.

ACKNOWLEDGEMENTS

This work was supported in part by the Chinese Scholarship Council and the Engineering Science Department at the University of Oxford. Thanks to Tingkai Li, Adam Thelen and Prof. Hu Chao for the helpful discussion about model details. Thanks to Dr Nicola Courtier for suggestions about feature extraction. Thanks also to Masaki Adachi for proofreading.

REFERENCES

Ai, W., Kraft, L., Sturm, J., Jossen, A., and Wu, B. (2019). Electrochemical thermal-mechanical modelling of stress inhomogeneity in lithium-ion pouch cells. *Journal of The Electrochemical Society*, 167(1), 013512.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4, 40–79.

Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(7795), 397–402.

Betancourt, M. and Girolami, M. (2015). Hamiltonian monte carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, 79(30), 2–4.

Bhattacharya, A., Jaiswal, R., and Kumar, A. (2018). Faster algorithms for the constrained k-means problem. *Theory of computing systems*, 62(1), 93–115.

Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., and Howey, D.A. (2017). Degradation diagnostics for lithium ion cells. *Journal of Power Sources*, 341, 373–386.

Cano, Z.P., Banham, D., Ye, S., Hintennach, A., Lu, J., Fowler, M., and Chen, Z. (2018). Batteries and fuel cells for emerging electric vehicle markets. *Nature Energy*, 3(4), 279–289.

Dechent, P., Greenbank, S., Hildenbrand, F., Jbabdi, S., Sauer, D.U., and Howey, D.A. (2021). Estimation of li-ion degradation test sample sizes required to understand cell-to-cell variability. *Batteries & Supercaps*, 4(12), 1821–1829.

Deng, Z., Lin, X., Cai, J., and Hu, X. (2022). Battery health estimation with degradation pattern recognition

and transfer learning. *Journal of Power Sources*, 525, 231027.

Fei, Z., Yang, F., Tsui, K.L., Li, L., and Zhang, Z. (2021). Early prediction of battery lifetime via a machine learning based framework. *Energy*, 225, 120205.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Jiang, B., Gent, W.E., Mohr, F., Das, S., Berliner, M.D., Forsuelo, M., Zhao, H., Attia, P.M., Grover, A., Herring, P.K., et al. (2021). Bayesian learning for rapid prediction of lithium-ion battery-cycling protocols. *Joule*, 5(12), 3187–3203.

Lake, B.M., Salakhutdinov, R., and Tenenbaum, J.B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332–1338.

Li, Y., Zhong, S., Zhong, Q., and Shi, K. (2019). Lithium-ion battery state of health monitoring based on ensemble learning. *IEEE access*, 7, 8754–8762.

Liu, Q., Du, C., Shen, B., Zuo, P., Cheng, X., Ma, Y., Yin, G., and Gao, Y. (2016). Understanding undesirable anode lithium plating issues in lithium-ion batteries. *RSC advances*, 6(91), 88683–88700.

Liu, W., Liu, P., and Mitlin, D. (2020). Review of emerging concepts in sei analysis and artificial sei membranes for lithium, sodium, and potassium metal battery anodes. *Advanced Energy Materials*, 10(43), 2002297.

Paulson, N.H., Kubal, J., Ward, L., Saxena, S., Lu, W., and Babinec, S.J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. *Journal of Power Sources*, 527, 231127.

Pedersen, E.J., Miller, D.L., Simpson, G.L., and Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*, 7, e6876.

Raj, T., Wang, A.A., Monroe, C.W., and Howey, D.A. (2020). Investigation of path-dependent degradation in lithium-ion batteries. *Batteries & Supercaps*, 3(12), 1377–1385.

Reniers, J.M., Mulder, G., and Howey, D.A. (2019). Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries. *Journal of The Electrochemical Society*, 166(14), A3189.

Salvatier, J., Wiecki, T.V., and Fonnesbeck, C. (2016). Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2, e55.

Schmuch, R., Wagner, R., Hörpel, G., Placke, T., and Winter, M. (2018). Performance and cost of materials for lithium-based rechargeable automotive batteries. *Nature Energy*, 3(4), 267–278.

Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5), 383–391.

Strange, C., Allerhand, M., Dechent, P., and dos Reis, G. (2022). Automatic method for the estimation of li-ion degradation test sample sizes required to understand cell-to-cell variability. *Energy and AI*, 100174.

Su, L., Zhang, J., Huang, J., Ge, H., Li, Z., Xie, F., and Liaw, B.Y. (2016). Path dependence of lithium ion cells aging under storage conditions. *Journal of Power Sources*, 315, 35–46.

Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y.T., Steinbacher, F., Khan, M.U., Lee, J.W., Siegel, J.B., Stefanopoulou, A.G., et al. (2021a). The challenge and opportunity of battery lifetime prediction from field data. *Joule*, 5(8), 1934–1955.

Sulzer, V., Mohtat, P., Lee, S., Siegel, J.B., and Stefanopoulou, A.G. (2021b). Promise and challenges of a data-driven approach for battery lifetime prognostics. In *2021 American Control Conference (ACC)*, 4427–4433. IEEE.

Zhang, Y., Xiong, R., He, H., and Pecht, M.G. (2018). Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Transactions on Vehicular Technology*, 67(7), 5695–5705.

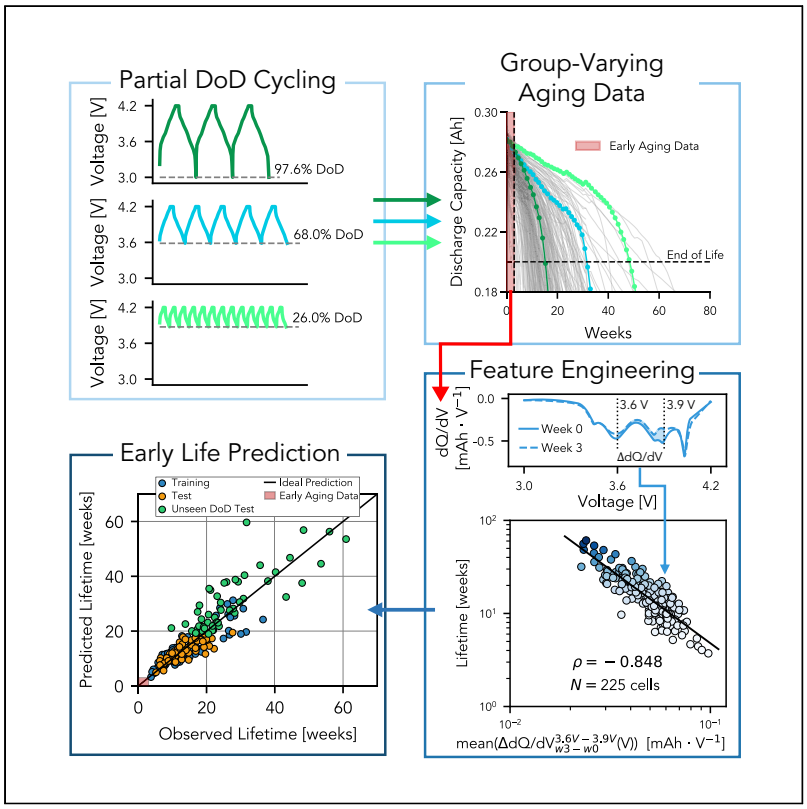
Zhou, Z. (2023). https://github.com/battery-intelligence-lab/hbm_ifac.git.

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Predicting battery lifetime under varying usage conditions from early aging data



Li and Zhou et al. demonstrate a method for predicting the lifetime of cells under widely varying cycling conditions using early-life measurements. This method utilizes degradation-informed features from early-life data and captures the hierarchical structure of battery aging data, showing potential for extension to different chemistries.

Tingkai Li, Zihao Zhou, Adam Thelen, David A. Howey, Chao Hu

chao.hu@uconn.edu

Highlights

A new early prediction problem, considering multiple stress factors, is defined

Predictive features with degradation-mode information are introduced

The hierarchical structure of battery aging data is captured in lifetime prediction

An aging dataset of 225 cells subjected to diverse cycling conditions is shared

Li et al., Cell Reports Physical Science 5, 101891
April 17, 2024
https://doi.org/10.1016/j.xcrp.2024.101891

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

Predicting battery lifetime under varying usage conditions from early aging data

Tingkai Li,[1,4] Zihao Zhou,[2,4] Adam Thelen,[3] David A. Howey,[2] and Chao Hu[1,5,*]

SUMMARY

Accurate battery lifetime prediction is important for maintenance, warranties, and cell design. However, manufacturing variability and usage-dependent degradation make life prediction challenging. Here, we investigate new features derived from capacity-voltage data in early life to predict the lifetime of cells cycled under varying charge rates, discharge rates, and depths of discharge. The early-life features capture a cell's state of health and the change rate of component-level degradation modes. Using a newly generated dataset from 225 nickel-manganese-cobalt/graphite lithium-ion cells aged under a wide range of conditions, we demonstrate a lifetime prediction of in-distribution cells with 15.1% mean absolute percentage error (MAPE). A hierarchical Bayesian model shows improved performance on extrapolation, achieving 21.8% MAPE for out-of-distribution cells. Our approach highlights the importance of using domain knowledge of battery degradation to inform feature engineering and model construction. Further, a new publicly available battery lifelong aging dataset is provided.

INTRODUCTION

Understanding the long-term degradation of lithium-ion (Li-ion) batteries is crucial for their optimal manufacturing, design, and control.[1,2] However, repeatedly assessing cell performance via aging experiments is a time- and cost-intensive task.[3] Manufacturers and researchers need quick and accurate methods to screen long-term performance and to quantify the impact of new designs and control changes without having to cycle cells to the end of life (EOL) each time a new question arises. Models using data from early life could significantly shorten the time needed to make accurate predictions of long-term degradation,[4] and this could lead to rapid screening of new battery performance and optimization of charging protocols.[5–7]

The idea that lifetime can be predicted using measurements from the early stages of battery aging experiments has its roots in research from over a decade ago by J. Dahn and researchers at Dalhousie University, who were investigating the impact of new electrolyte additives and electrode designs on battery performance. In late 2009, they published a paper describing how high-precision measurements of Coulombic efficiency during the first few cycles could be used to predict cell lifetime and rank it qualitatively against other cells.[8] Coulombic efficiency is an important performance metric, and it is calculated as the discharge-to-charge capacity ratio, where an ideal value of unity indicates perfect cyclic efficiency. Measuring cell Coulombic efficiency with an error of < 0.01% can indicate cell-to-cell differences caused by different rates of undesirable side reactions that lead to capacity fade. Using purpose-built high-precision equipment, the Dalhousie team published a paper in 2011 that compared long-term cycling data (> 750 cycles) with predicted lifetimes

[1]School of Mechanical, Aerospace, and Manufacturing Engineering, University of Connecticut, Storrs, CT 06269, USA

[2]Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

[3]Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA

[4]These authors contributed equally

[5]Lead contact

*Correspondence: chao.hu@uconn.edu

https://doi.org/10.1016/j.xcrp.2024.101891

extrapolated from short-term (<500 h) high-precision Coulombic efficiency measurements.[9]

Since this work, many new studies have been published on "early-life prediction." In 2013, the Dalhousie University group published another paper demonstrating the lifetime ranking of 160 Li-ion cells with various electrolyte additives, using high-precision Coulombic efficiency measurements from the first 50 cycles of data.[10] The Coulombic efficiency measurements strongly correlated with the cells' lifetimes. However, many researchers and industry professionals do not have access to high-precision machines for testing. Furthermore, it would be even more useful to predict lifetimes using early-life measurements made during faster cycling experiments and under a broader range of operating conditions, enabling the technology to be deployed in more research areas and even for cells operating in the field.

Research by Baumhöfer et al.[11] and Harris et al.[12] investigated alternative approaches not requiring the use of a high-precision cycler. Baumhöfer et al. developed a lifetime prediction model on 48 cells cycled under identical conditions.[11] Hundreds of early-life features extracted from impedance spectra, pulse characterization tests at different states of charge, and standard capacity tests were reduced to a set of 24 features and used for prediction. The model using 24 features was accurate within 16 cycles; however, further analysis showed that model accuracy was highly dependent on the number of features used, with more features generally being better, suggesting the model may possibly be overfitting the small dataset ($N = 48$). Harris et al. examined the failure statistics of 24 cells cycled under identical conditions and established a weak correlation between the cells' capacity at cycle 80 and the capacity at cycle 500.[12] These works suggest simpler and more easily obtainable early-life features might be found to correlate with eventual lifetime.

Severson et al.[5] in 2019 demonstrated an early-life prediction model using features extracted from the discharge capacity vs. voltage ($Q(V)$) curves during regular cycling. The feature extraction method was unique, quantifying the cells' degradation rates by tracking the early-life variation of their $Q(V)$ curves between cycles 10 and 100, referred to as $\Delta Q_{100-10}(V)$. The approach was also used in follow-up work by Attia et al.[6] to accelerate an experimental campaign to optimize the constant current portion of a fast charging protocol. The researchers in these papers generated a large battery aging dataset from 169 lithium-iron-phosphate/graphite (LFP) cells cycled under various fast charging protocols. This was made publicly available, and many other researchers have investigated methods of further improving predictive performance and feature extraction techniques using these data.[13–22] Notably, Paulson et al.[22] demonstrated accurate early-life prediction on six different metal oxide cathode chemistries. Fermin-Cueto et al.[20] investigated predicting the knee point (when capacity begins to decrease rapidly) in a cell's capacity degradation curve using early-life features. Similarly, Li et al.[21] demonstrated a prediction model capable of projecting the entire capacity degradation trajectory from early-life features.

Despite this growing body of research, many fundamental questions about battery life modeling remain unanswered. One fundamental issue is that, in order to train machine learning models to predict lifetime from early-life cycles, data from the *entire* lifetime are required. Therefore, these approaches are best suited to applications such as screening cells after manufacturing, or relative comparisons, rather than quantitatively absolute predictions. A second issue is a lack of publicly available battery lifetime data that covers a wide range of conditions. The dataset published

by Severson et al. and Attia et al.[5,6] was specifically generated to study high-rate fast charging protocols for LFP cells, leaving the discharge rate and depth of discharge fixed. Even though the dataset is relatively large compared to existing publicly available datasets (N = 169 cells), the limited range of operating conditions, in this case, induced a single dominant degradation mode (loss of active material at the anode or negative electrode, "LAM_{NE}"), causing all of the capacity degradation trajectories to have very similar shapes, and perhaps making lifetime prediction easier.[23] While the relationships between cell operating conditions and the corresponding degradation modes are well understood,[1,3,24,25] it remains unclear how the $\Delta Q(V)$ feature transfers to cells of different chemistries and to situations where multiple interacting degradation modes are present. This is especially the case for cells that experience milder degradation resulting in less obvious changes in the $Q(V)$ curve. Furthermore, all cells in the dataset from Severson et al. and Attia et al.[5,6] were cycled under a fixed depth of discharge, making it easy to extract features from any cycle along the cell's degradation trajectory. However, in practice, the depth of discharge of cells depends on applications and user preferences, and cells may not be subjected to full depth-of-discharge cycles in many cases. So, there is a need to explore alternative methods of collecting early-life feature data and validating results using periodic reference performance tests or other means.

In this work, we investigate new early-life features derived from capacity-voltage data that can be used to predict the lifetimes of cells cycled under a wide range of charge rates, discharge rates, and depths of discharge. To study this, we generated a new battery aging dataset from 225 nickel-manganese-cobalt/graphite (NMC/Gr) cells, cycled in groups of four per condition, under a much wider range of operating conditions than existing publicly available datasets.[26] The cells in our dataset exhibit larger variations in their capacity degradation trajectories than previous open datasets, driven by the interactions and accumulations of various component-level degradation mechanisms.[1,23] To predict the lifetimes of cells experiencing different degradation pathways accurately, we introduce new early-life features extracted from the differential voltage (dV/dQ vs. Q) and incremental capacity (dQ/dV vs. V) data gathered during regular weekly reference performance tests (RPTs). The RPTs, two complete cycles at full depth of discharge, enable consistent feature extraction and lifetime prediction for cells that normally cycle at fractional depths of discharge, some as low as 4.0%. Using as little as the first 5% of the aging data, we achieve a mean absolute percentage error (MAPE) of 22% on the lifetime. Including up to 15% of the entire cell lifetime data, we achieve an average prediction error of 15.1% MAPE and a root-mean-square error (RMSE) of 2.8 weeks on in-distribution test sets when testing the new features in traditional machine learning models built with regularized linear regression. Given that our dataset has a hierarchical structure (i.e., the "group" level and the "cell" level) in nature, we also explore the possibility of applying hierarchical Bayesian linear modeling to predict lifetime, which achieves better extrapolation performance on out-of-distribution samples, viz., 7.3 weeks RMSE and 21.8% MAPE lifetime prediction error.

The major contributions of this work are 4-fold:

(1) proposing a new idea to categorize early-life features into two hierarchical levels, the condition (upper) and the cell (lower) level, in order to capture an inherent hierarchical structure in the battery aging data and enable greater generalization especially to out-of-distribution data;
(2) creating a hierarchical Bayesian model (HBM) to address the hierarchical nature of the aging data and quantify the uncertainty in lifetime predictions;

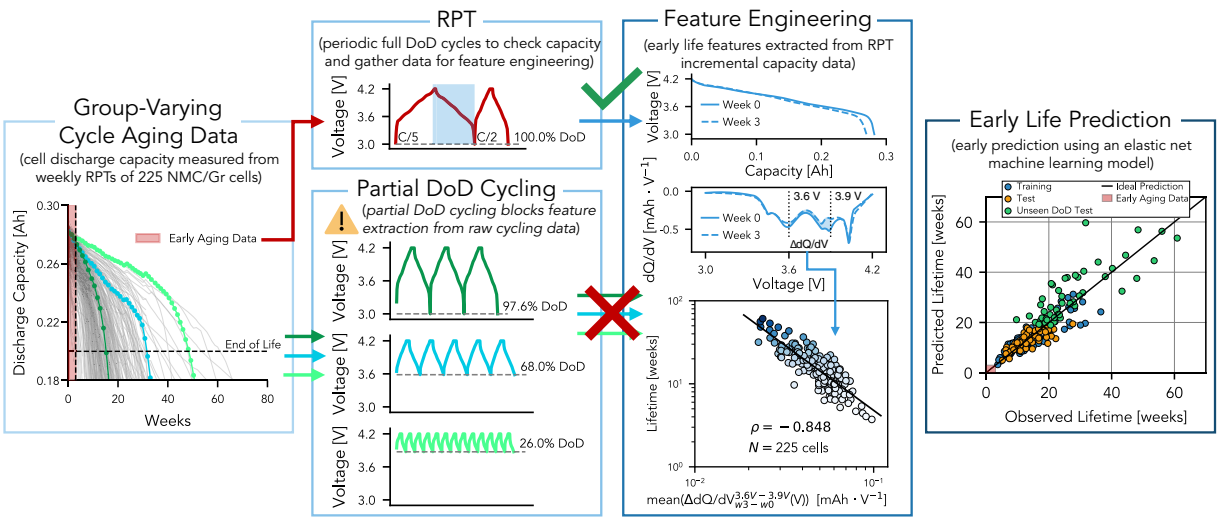


Figure 1. High-level overview of our approach

Unlike existing approaches for early prediction, we extract features from periodic reference performance tests instead of regular cycling data. In this example, we extract a feature from a partial voltage window of incremental capacity that is highly correlated with lifetime. From this and other features, we build a machine learning model to predict the lifetimes of new unseen cells.

(3) demonstrating a new method for extracting predictive features from incremental capacity curves, incorporating optimization of voltage window to improve correlations with lifetime;

(4) generating and publicly sharing a large battery aging dataset consisting of 225 NMC cells cycled under a wide range of operating conditions, enabling researchers without access to battery testing equipment to study lifetime modeling and other related topics.

RESULTS AND DISCUSSION

Approach for predicting lifetime from early-life data

Prediction of lifetime from early data is more challenging when there are multiple varying stress factors, because this leads to diverging capacity trajectories. Our approach, outlined in Figure 1, differs from the prior art[5,8–10] in several ways.

First, to apply early prediction to cells cycled under different depths of discharge, we extract features from periodic RPTs instead of regular cycling data. This means that the discharge voltage curves obtained from periodic RPTs are complete and consistent for every cell, making feature extraction more consistent. Second, we develop new features based on partial voltage windows of $Q(V)$ curves and their derivatives (differential voltage and incremental capacity data). Using a new feature extraction method (see details in "extracting features from incremental capacity data"), we find features that better correlate with cell lifetime for our dataset than existing features reported in the literature.[5,15,19] Additionally, we explore using cycling protocol information (C_{chg}, C_{dchg}, DoD) as features to predict lifetime, establishing a link between the two. All extracted features are reduced to a highly predictive subset using a feature selection method (see "feature selection"). Then, the selected features are used as input to a machine learning model to predict cell lifetime. In what follows, we outline our approach to feature engineering for early-life prediction and discuss the challenges of applying existing feature engineering methodologies proven on

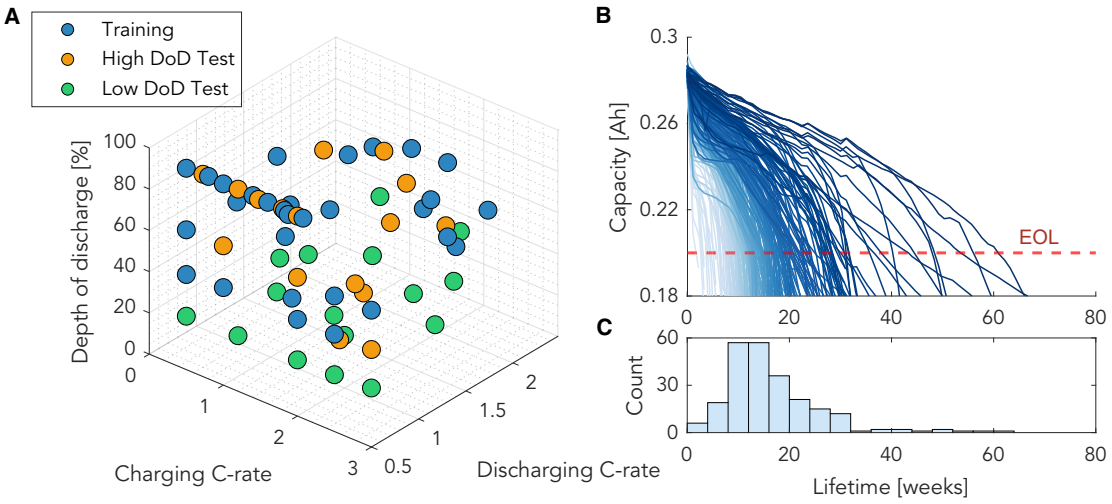


Figure 2. Overview of battery aging test conditions and capacity data

(A) 3D scatterplot showing train-test split and cycling conditions used—each point represents conditions for a group of four cells, and marker color indicates a data subset used to generate prediction results in "predicting lifetime using machine learning models."

(B) Discharge capacity fade curves for all 225 NMC/graphite cells plotted past 80% their rated capacity (250 mAh); color of each curve is scaled by cell lifetime.

(C) Histogram of the cell lifetimes at end of life (EOL) using 80% of rated capacity as threshold.

LFP/Gr to our NMC/Gr cells that are cycled under a wider range of operating conditions. Last, we introduce HBMs for early-life prediction.

Lithium-ion battery dataset under varying usage conditions

Publicly available datasets, such as those from NASA's Prognostics Center of Excellence,[27,28] the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland, College Park,[29,30] and Sandia National Laboratories,[31] contain cells of different chemistries cycled under a range of charge rates, discharge rates, and temperatures. These datasets are frequently used in research studies for battery modeling since aging commercial-grade Li-ion cells is slow and expensive. Also, these datasets report measurements including capacity, internal resistance (NASA and CALCE), voltage, current, and temperature, enabling researchers to study different aspects of battery modeling. However, the relatively small size of these datasets (roughly 30 cells per group) makes investigating machine learning-based approaches to early-life prediction challenging. On the other hand, in recent years, datasets such as those from the Toyota Research Institute[5,6] and Argonne National Laboratory[22] contain many more cells (>150 cells). However, they focus on a limited range of operating conditions—fast charging and symmetric C/2 cycling, respectively—making it difficult to build machine learning models that generalize across cycling conditions.

In light of this, we designed our battery aging dataset to study more cells under a broader range of operating conditions than current publicly available datasets.[26] Our dataset comprises 225 cells cycled in groups of four to capture some of the intrinsic cell-to-cell aging variability.[32] A unique feature of our dataset is the many capacity degradation trajectories that reflect different accumulated degradation modes induced by the various operating conditions. These trajectories, shown in Figure 2, exhibit different one-, two-, and three-stage degradation trends driven by the interaction and accumulation of hidden, threshold, and snowballing

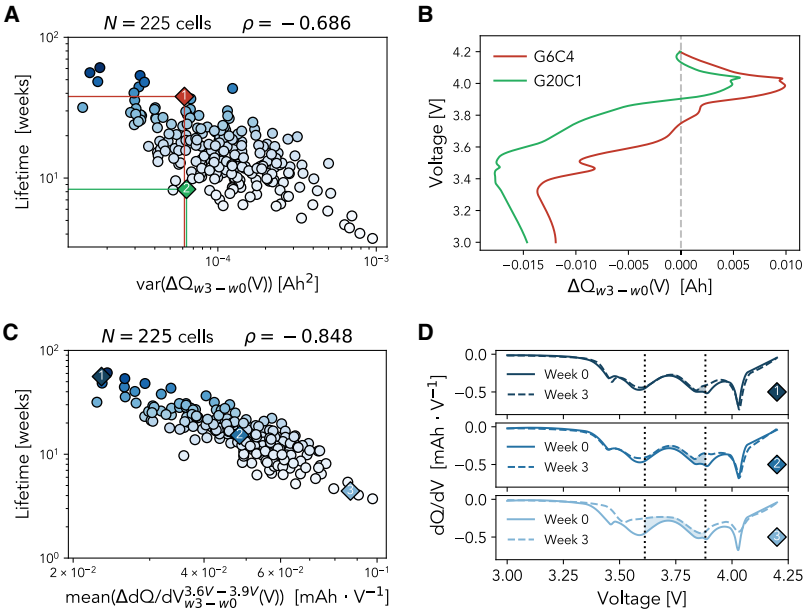


Figure 3. Feature extraction from early-life data

Well-known early-life features do not explain the variance in our dataset, and a newly extracted feature from incremental capacity curves correlates better with lifetime.

(A) Cell lifetime for 225 NMC cells plotted as a function of $\mathrm{var}(\Delta Q_{w3-w0}(V))$; Pearson correlation coefficient −0.686. The two cells highlighted have similar values of $\mathrm{var}(\Delta Q_{w3-w0}(V))$ but very different lifetimes.

(B) Difference between discharge capacity curves as a function of voltage between week 3 and 0 for the two cells highlighted in (A).

(C) Cell lifetime plotted as a function of optimized feature $\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.60V-3.90V}(V))$, Pearson correlation coefficient − 0.848.

(D) Incremental capacity curves from weeks 3 and 0 for three representative cells (G1C2, G16C1, and G53C2, respectively); the shaded areas indicate the change of curves between the voltage bound (3.60 − 3.90 V) after 3 weeks of cycling.

degradation modes.[23] These varying trends produce cell lifetimes from 1.5 to 60.9 weeks. Experimental details and testing procedures used to generate the dataset can be found in "battery aging test design" and Note S1 and Note S2 of the supplemental information.

Extracting predictive features from early usage data

Initially, we extracted features previously reported to correlate strongly with cell lifetime.[5,15,19] We adopt the notation $\Delta Q_{w3-w0}(V)$ to describe the features, where the subscripts w3 and w0 correspond to data obtained from the RPTs from weeks 3 and 0, respectively. Preliminary testing of these well-established early-life features reveals that they do not fully explain the variance in our dataset. This is illustrated in Figure 3A, where we extract the $\mathrm{var}(\Delta Q(V))$ feature reported by Severson et al.[5] using discharge data from RPTs $\mathrm{var}(\Delta Q_{w3-w0}(V))$ and plot it against lifetime, revealing a large unexplained variance in the predicted lifetimes.

To understand why this occurs, consider two cells (G6C4 and G20C1) that have similar feature values but vastly different lifetimes. In this case, even though the $\Delta Q(V)$ curves have the same variance, they do not have the same shape and location (Figure 3B). It can be seen that the group twenty cell (G20C1) experienced more

significant capacity loss during this time, evident by the endpoint of $\Delta Q(V)$ at 3.0 V. Other noticeable changes exist in the $dV/dQ(Q)$ curves that differ between the cells (shown in Note S7 of the supplemental information), indicating additional but more subtle degradation modes are present. However, these differences in the evolution of the $Q(V)$ curve during early life are not captured by the var($\Delta Q_{w3-w0}(V)$) feature, causing the unexplained variance in the dataset. A further analysis compares the $\Delta Q_{w3-w0}(V)$ curves from our dataset with the $\Delta Q_{100-10}(V)$ curves from the dataset of Severson et al.,[5] shown in Figure S8 of the supplemental information. From this analysis, we find that the var($\Delta Q_{w3-w0}(V)$) feature only captures the capacity fade within the first 3 weeks in our dataset, which can be attributed to the differences in the trend of the $Q(V)$ curve evolution. Both the differences in cell chemistries between these two datasets and different discharging profiles from which the variance feature is extracted could contribute to the trend differences. The former impacts the overall discrepancy of voltage responses due to different phase change mechanisms involved during discharging. The latter, especially for the difference in C-rates, would cause different levels of heat generation during discharging. The discharge curve for which Severson et al.[5] introduced their feature is at 4 C, and the effect of self-heating on the voltage response cannot be neglected under this high C-rate, especially toward the end of discharge. In contrast, we extracted the equivalent var($\Delta Q_{w3-w0}(V)$) feature from measurements at C/5, where self-heating is much less of a concern.

While we only showed an example in Figure 3 for this particular feature, var($\Delta Q_{w3-w0}(V)$), the unexplained variance in the data persists using most other early-life features we tested. Typically, it is not a requirement that all model input features exhibit a strong correlation with cell lifetime, but finding a few features that do correlate well is generally advantageous because it can improve model fit and accuracy. Considering this, we explored extracting features from differential voltage and incremental capacity curves using partial voltage intervals in order to capture the diverse cell-specific degradation trends observed in our dataset more accurately.

Differential voltage (dV/dQ vs. Q) and incremental capacity (dQ/dV vs. V) curves have been widely adopted in battery aging diagnostics because certain features (e.g., peaks and valleys) on these curves are closely associated with phase transitions of electrodes and allow us to investigate electrode-specific aging modes. For example, simply plotting dQ/dV vs. V curves over successive cycles within a long-term aging test and observing changes in positions, amplitudes, and widths of certain peaks over cycling can help detect the underlying degradation modes (e.g., loss of active materials on the negative and positive electrodes and loss of lithium inventory associated with capacity fade) that drive these changes. This diagnostic property of differential voltage and incremental capacity curves could even allow reasonably accurate quantification of degradation modes.[33] Also, several battery lifetime prediction studies include features describing the locations and magnitudes of peaks from differential voltage and incremental capacity curves.[15,19,22] Here, we present a method to find an optimal incremental capacity feature, which is backed up by our feature selection process in "feature selection" and Note S11 and Note S12 of the supplemental information.

Rather than only using locations and magnitudes of peaks from incremental capacity ($dQ/dV(V)$) curves, we examine how the incremental capacity curve evolves over selected voltage intervals. Specifically, segments of the incremental capacity curve within a certain voltage interval are extracted from two RPT tests (the week 0 RPT and

the week 3 RPT, respectively). Then, two major summary statistics—the mean and the variance—of the difference between these two extracted segments are used as features to reflect the evolution of the incremental capacity curve within this voltage interval during the 3 weeks. Rather than manually identifying the voltage interval, a grid search method is employed to find an optimal voltage interval with the highest correlation between feature and lifetime (see "extracting features from incremental capacity data" for more details). The voltage interval search result using the mean summary statistics is shown in Figure S9 of the supplemental information.

We find the voltage interval that produces the highest linear correlation with cell lifetime is a mid-range where the upper and lower voltage limits are centered around prominent peaks in the incremental capacity curves at 3.60 V and 3.90 V. Figure 3C shows that the change in incremental capacity in this range is inversely proportional to lifetime. The new feature captures the change in incremental capacity intensity, calculated as the mean change in mAh/V over the middle voltage interval, $\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.60V-3.90V}(V)) = \mathrm{mean}(dQ/dV_{w3}^{3.60V-3.90V}(V) - dQ/dV_{w0}^{3.60V-3.90V}(V))$; see Figure 3D. This new feature shows a much stronger correlation with cell lifetime and better explains the variance in our dataset compared with the traditional feature $\mathrm{var}(\Delta Q_{w3-w0}(V))$.

The new feature derived from the incremental capacity curve likely captures the rate of active material loss during early life. This idea is supported by degradation diagnostics literature, which shows that changes in the intensity of the incremental capacity (mAh/V) curve at constant voltage correspond to a loss of active material.[1,33–35] Also, a close relationship between this feature and average cycling stress, with a Pearson correlation coefficient of 0.89, can be found in Figures S11 and S14 of the supplemental information, where an increase in our proposed feature is accompanied by an increase in the Stress_{avg} feature introduced in "extracting features from cycling conditions." Based on the assumption that higher stresses induced during cycling are likely to cause more loss of active materials, the found relationship helps back up our hypothesis that this incremental capacity feature captures the loss of active materials in general. Additional analysis to understand this feature regarding degradation information is included in Note S8 of the supplemental information. Additionally, we use the upper and lower voltage limits imposed during cycling to create two more intervals, 3.00 – 3.60 V and 3.90 – 4.20 V. We then extract two features from each voltage interval using the mean and variance summary statistics. In total, we extracted six features from $\Delta dQ/dV(V)$, two from each of the three voltage intervals using the mean and variance summary statistics.

Lifetime modeling work on NMC/Gr cells by Smith et al.[36] showed that the capacity fade rate due to cycling tracked nearly linearly with the square-root-of-cycling throughput, calculated as $(C_{chg}DoD)^{0.5}$, where C_{chg} is charging C-rate, and DoD is depth of discharge for the experiments. This metric is described as tracking the concentration gradient of lithium ions in the cathode active material and is a proxy for diffusion-induced stress.[36–38] We further investigate this feature as a model input for early-life prediction ("extracting features from cycling conditions") and as a condition-level grouping variable for our hierarchical Bayesian modeling approach ("hierarchical Bayesian models for early prediction").

In addition to the incremental capacity and stress-related features, the processes of extracting features from other data sources (e.g., differential voltage, constant-voltage

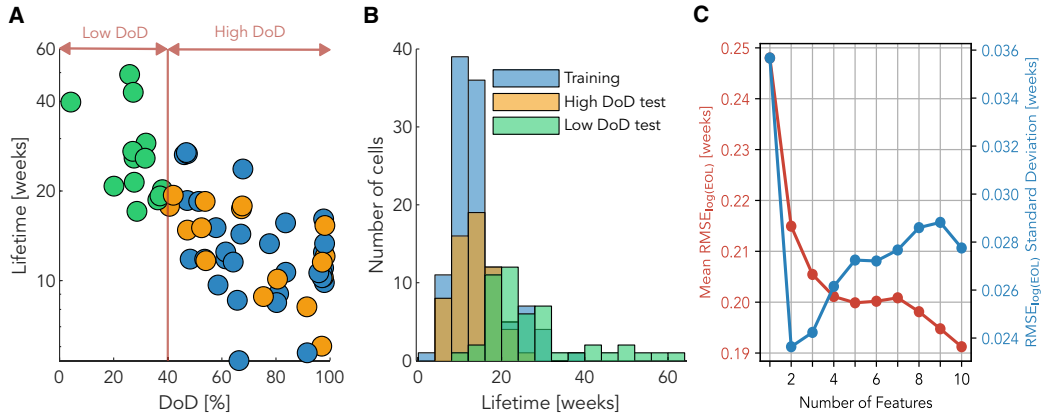


Figure 4. Overview of data partition and feature selection

(A) Scatterplot of mean group lifetime vs. DoD; marker color indicates train/test subset.

(B) Histogram showing each subset's distribution of cell lifetimes.

(C) Mean and standard deviation of $RMSE_{log(EOL)}$ for 5-fold repeated cross-validation on the ten candidate models.

charging segment, and capacity fade) are reported in Note S9 and Note S10 of the supplemental information. By extracting features from a total of 6 sources, we obtained a set of 29 candidate features, outlined in Table S2 of the supplemental information. All the early-life features in this work were extracted from C/5 cycles in RPTs.

Partitioning data for machine learning and feature selection

Dataset partitioning was done at the group rather than the cell level, for three reasons. First, practical battery aging tests for product validation typically cycle multiple cells under the same conditions to capture the aging variability due to manufacturing. Second, it is desirable to build an early prediction model to predict the lifetimes of cells cycled under previously untested conditions. Finally, although building an early prediction model with cells tested under rapidly accelerated aging conditions is useful in minimizing the time and costs of collecting aging data, one cannot preemptively know the lifetime (before tests), so grouping must be done using an alternative indicator of cell lifetime. Since the depth of discharge is the dominant cycling stress factor impacting the battery lifetimes in our aging dataset (Figure 4A), this was used to determine the dataset partitioning.

We first separate our dataset into a high-DoD region and a low-DoD region, with a boundary at 40% depth of discharge (Figure 4A). In the high-DoD region, we further divide the data into a training set and an in-distribution high-DoD test set. The high-DoD test set is used to evaluate the model's prediction accuracy for cells with conditions similar to the ones the model was trained on. Last, we assign all data in the low-DoD region ($<40\%$) to a second test set used to test the model's ability to extrapolate to unseen test conditions. The dataset split is also visualized in Figure 2A, where each axis is one of the three cycle aging stress factors (C_{chg}, C_{dchg}, DoD), and the marker color indicates the data subset that the group belongs to. The training set contains cells with lifetimes ranging from 3.7 to 36.6 weeks, and the high-DoD test set has cells with lifetimes between 5.2 and 31.6 weeks. On the other hand, the low-DoD test set is more diverse, with lifetimes ranging from 9.7 to 60.9 weeks. Histograms of cell lifetimes for each data subset are visualized in Figure 4B.

After extracting a total of 29 features, down-selecting a smaller feature subset is required before training machine learning models for a couple of reasons. First,

Table 1. Stepwise forward search results

Step number	Selected feature	Description
1	$\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$	best incremental capacity feature from "extracting features from incremental capacity data" (shown in Figure 3C)
2	$\log(\lvert\Delta \mathrm{CV\ Time}_{w3-w0}\rvert)$	change in CV hold time (see Note S10 of the supplemental information)
3	DoD	depth of discharge
4	ΔQ_{w3-w0}^{1}	change in DVA-based capacity $Q^{DVA,1}$ (see Note S9 of the supplemental information)
5	$C_{chg}^{0.5}\mathrm{DoD}^{0.5}$	charge-induced stress (see "extracting features from cycling conditions")
6	C_{chg}	charging C-rate
7	$\log(\mathrm{var}(\Delta dQ/dV_{w3-w0}^{3.0V-3.6V}(V))$	variance of low-voltage incremental capacity feature (see "extracting features from incremental capacity data")
8	ΔQ_{w3-w0}^{3}	change in DVA-based capacity $Q^{DVA,3}$ (see Note S9 of the supplemental information)
9	$\log(\lvert\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.0V-3.6V}(V)\rvert)$	mean of low-voltage incremental capacity feature (see "extracting features from incremental capacity data")
10	$\log(\lvert\Delta \mathrm{CV\ Time}_{w0}\rvert)$	CV hold time of the initial RPT (see Note S10 of the supplemental information)

some features are strongly correlated with one another, known as multicollinearity. A model trained with collinear features can be sensitive to minor changes in the feature values and may extrapolate poorly.[39] Second, while the new dataset is large (225 cells) relative to existing publicly available aging datasets, it is still considered small for training a machine learning model. Small datasets require special care to avoid overfitting and to improve generalization performance on unseen test data. This is especially the case when the number of training points is not significantly larger than the number of features ($N_{data} \gg N_{features}$). Therefore, it is crucial to select a subset of highly predictive features before training.[40,41] We performed stepwise forward feature selection on the training dataset following the method described in "feature selection." To avoid poor performance on the test sets due to overfitting, we performed a 5-fold cross-validation study five times, using up to 10 features. With the cross-validated errors, including the mean and standard deviation for each candidate model, one can evaluate the candidate feature subsets from two aspects. As the mean of $RMSE_{EOL}$ decreases, the feature subset fits this training set better in general. As the standard deviation increases, the feature subset yields larger run-to-run variation for different folds, which indicates unstable performance on prediction. So, when selecting features, one must consider both aspects simultaneously to ensure good performance, especially for extrapolating to the low-DoD test set. The trends of the mean and standard deviation of cross-validation $RMSE_{log(EOL)}$ of this trial are reported in Figure 4C, and the selected feature in each step is listed in Table 1. The model with two features, namely $\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$ and $\log(\lvert\Delta \mathrm{CV\ Time}_{w3-w0}\rvert)$, has the lowest run-to-run variance and relatively low mean error $RMSE_{EOL}$. Adding a third feature to the set, DoD, produces a model with lower mean $RMSE_{EOL}$ but increases the run-to-run variation. For a more comprehensive evaluation, we compare the results of models trained using both two and three features.

Predicting lifetime using machine learning models

There are two types of machine learning models used in this study: one is elastic net regression, and the other one is the HBM. Except for the dummy model and HBM, the rest of the models compared in this section are trained with elastic net models

Table 2. Prediction errors for selected models tested using the high- and low-DoD test datasets

		MAPE (%)			RMSE (weeks)		
Model	$N_{feature}$	Training	High DoD	Low DoD	Training	High DoD	Low DoD
Dummy model	0	35.0	31.5	47.5	6.5	4.8	18.5
Cycling conditions	3	24.8	19.0	23.7	4.0	3.3	9.8
Discharge model[5]	5[a]	23.9	28.0	24.8	4.6	4.7	11.5
Degradation-informed	2	17.3	16.0	24.4	3.2	3.0	7.8
Degradation-informed	3	16.5	15.1	33.0	3.1	2.8	9.7
HBM	2[b]	18.6	16.9	21.8	3.3	3.1	7.3
HBM	3[b]	17.4	15.8	24.1	3.1	2.9	7.5

[a]The discharge model[5] contains six features, with one of them being the difference between the maximum capacity and capacity at cycle two, ΔQ_{max-2}. However, this feature cannot be calculated for our dataset due to the partial depth of discharge cycling and the continuously decreasing capacity-fade curves for all cells and has thus been omitted.

[b]The number of features listed refers to the number of cell-level input features. For both HBMs, a single cycling condition-level feature is used for grouping cells, and, as indicated in the table, either two or three cell-level features are used for regression.

on different feature sets. To predict the lifetime, we initially establish a pair of baseline models. The first baseline model is a dummy model that does not use any input features or have any trainable parameters, and instead, it predicts the mean cell lifetime of the training set for all cells. This is a good way to determine if a more complex model is truly learning new information from the input data or instead only appears to be learning because of similar training/test dataset distributions that lead to similar error metrics. When tested on the two test datasets, the dummy model achieves $MAPE_{EOL}$ of 31.5% and 47.5% on the high-DoD and low-DoD test sets, respectively. The error metrics for all models tested are shown in Table 2.

The second baseline model is built using only the cycling condition parameters as input features. This model predicts lifetimes without using cell-specific aging measurements. This model achieves an $MAPE_{EOL}$ of 19.0% and 23.7% on the high-DoD and low-DoD test sets, respectively. The substantial decrease in prediction error over the dummy model shows that the usage parameters convey a significant amount of information that can be used to predict lifetime accurately. This result is expected, as a great deal of battery lifetime modeling work[36,42,43] has already explored the strong connection between usage and degradation. However, only using condition-level cycling features does not account for intrinsic cell-to-cell variability. Hence, the next set of models we tested included cell-level features extracted from the early aging data.

The first cell-level feature model is the "discharge model" described in Severson et al.[5] and "extracting predictive features from early usage data." This model, and all other models built on cell-level inputs, use features extracted from the RPTs of weeks 0 and 3, which is just under 18% of the average lifetime. The original "discharge model" has six features, with five of them from statistics of $\Delta Q(V)$ curves and one feature ΔQ_{max-2} capturing the capacity recovery at initial cycles. Due to the monotonic decreasing capacity trajectories observed in our dataset, an analogous feature for ΔQ_{max-2} cannot be extracted and thus is omitted. The main feature included is $var(\Delta Q_{w3-w0}(V))$; however, we found that this did not completely describe the variance in our dataset. When tested on the high- and low-DoD test datasets, the discharge model achieved 28.0% and 24.8% $MAPE_{EOL}$, respectively. The performance on the two test datasets is slightly worse than the cycling condition

model, yet still better than the dummy model, indicating that the features used in the discharge model do carry useful information but are not optimal for our dataset (see Table 2).

The remaining models we compare are the degradation-informed model and HBM. We refer to our elastic net models as *degradation-informed* in Table 2 because of the newly developed degradation-based features used as model inputs. Both the degradation-informed model and HBM use the same sets of input features, and for thoroughness, we compare models built using two and three features each. Compared to the cycling condition baseline, the two-feature elastic net model shows decreased $MAPE_{EOL}$ on the high-DoD test of 16.0% and a slight increase in error on the low-DoD test set to 24.4%. However, the $RMSE_{EOL}$ of the low-DoD test set drops considerably from 9.8 to 7.8 weeks. For the HBMs, we observe small increases in the training and the high-DoD test errors with a noticeable improvement in the low-DoD test errors over the degradation-informed models using the same set of features.

For both the degradation-informed and hierarchical models, we observe that including the third feature decreases model prediction error on the training and high-DoD test datasets but increases error for the low-DoD test dataset. When the third feature is added, both models overfit the training dataset and exhibit poor extrapolation capability to the low-DoD test dataset where the cells have longer lifetimes. Regardless, the HBM trained with three features still performs better when predicting the low-DoD test set compared with its elastic net counterpart. Generally, by comparing the evaluation metrics of the two models (degradation-informed model and HBM), we find that the HBM has better generalizability to the low-DoD test set but at the cost of slightly higher training and high-DoD test errors.

The large improvement in performance observed for models using cell-level (as opposed to only using cycling condition features) features prompts us to further investigate why the feature $\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$ explains cell-to-cell variability better than other features. Firstly, it is more accurate to use measured health metrics from individual cells in operation to predict their lifetime. This reveals the intrinsic cell-to-cell variability that could cause different aging behaviors under identical cycling conditions. Secondly, this optimized feature, which likely captures how much loss of active material happens during early life, has a balanced representation of the variability within the group and among the entire dataset.

In summary, we find that the best feature $\log(\mathrm{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V))$ explains the cell-to-cell variability well for a majority of cells. The remaining variance in the feature-lifetime correlation may be contributed jointly by measurement inaccuracy and unexplained manufacturing variability. Hence, our analysis of the results suggests that a predictive early-life feature should capture the variability introduced by the difference in cycling conditions and information about intrinsic cell-to-cell variation that causes different performances under identical loads. Also, our feature engineering methodology ("extracting features from incremental capacity data") can be extended to find good features for other cell chemistries. Additional analyses on benchmarking different feature subsets and machine learning models are included in Note S16 of the supplemental information.

Analysis of HBM results

The probabilistic nature of HBMs enables us to extract a deeper understanding by considering both the mean and the uncertainty of lifetime predictions. Assuming

individual cluster fitting parameters and noise variance, $\boldsymbol{\theta}_j$ and σ_j respectively, are independent, the posterior predictive distribution can be written as

$$p\left(y_j^*\middle|Y_j\right) = \iint p(\sigma_j|Y_j)p(\boldsymbol{\theta}_j|Y_j)p\left(y_j^*\middle|\boldsymbol{\theta}_j,\sigma_j\right)d\boldsymbol{\theta}_j d\sigma_j, \quad \text{(Equation 1)}$$

where y_j^* is the predicted lifetime of a new cell belonging to the j th group, and Y_j represents all lifetime observations associated with the j th group. For a point-wise prediction, one can estimate the mean value of $p(y_j^*|Y_j)$. Table 2 lists the performance of the HBM built using two different feature sets. The first uses two cell-level features, $\log\left(\left|\text{mean}\left(\Delta dQ/dV_{\text{w3}-\text{w0}}^{3.6V-3.9V}(V)\right)\right|\right)$ and $\log(\Delta\text{CV Time}_{\text{w3}-\text{w0}})$, and achieves 3.1 weeks RMSE and 16.9% MAPE for the high-DoD test set, which is almost the same as the performance of the degradation-informed model using the same feature set, while for the low-DoD test set, the HBM achieves 7.3 weeks RMSE and 21.8% MAPE, which outperforms the degradation-informed model by 7% and 10% for RMSE and MAPE, respectively.

Similar to the degradation-informed model, we observe that the HBM model overfits the training dataset when the third feature (DoD) is added. This is evident by the increased performance on the training and high-DoD test set but worse performance on the low-DoD test set. Specifically, under the high-DoD test set, RMSE improved from 3.1 to 2.9 weeks, and MAPE improved from 16.9% to 15.8%. However, for the low-DoD test set, RMSE increased from 7.3 to 7.5 weeks, and MAPE increased from 21.8% to 24.1%. Notably, the HBM shows more resistance to overfitting than the degradation-informed model, whose performance decreased substantially more than the HBM when the third feature was included in the feature set. Similar results can also be found in Tables S3 and S4 of the supplemental information, where the HBM consistently shows better robustness on unseen test samples (i.e., the low-DoD test set) compared to elastic net models on different feature sets or different machine learning algorithms on the same feature set.

Figure 5B shows the uncertainty (2 standard deviations) of $p(y_j^*|Y_j)$ for posterior lifetime predictions of each cluster. The uncertainty levels for clusters 0 and 1 are around $\pm$ 4.5 weeks (at 2 SD), whereas for clusters 2 and 3, the uncertainty levels are around $\pm$ 9.5 and 10.5 weeks, respectively, which reflects the model's uncertainty when predicting cells from unseen cycling conditions. According to Table 3, there are only 12 cells from cluster 3 in the training set, while there are 23 cells from cluster 3 in the low-DoD test set. Due to the lack of data, the uncertainty for all regression parameters $(\boldsymbol{\theta}_3, \sigma_3)$ for cluster 3 is much larger than that of clusters 0 and 1. On the other hand, as the prediction uncertainty becomes large for long-life cells, uncertainty itself can be used as an indicator to denote whether one should include more early-life data for feature calculation. For example, when running HBM in a forward mode (using the trained model to give predictions), for test samples in cluster 3, large prediction uncertainty is observed (>10 weeks). One may consider including the 4th or 5th week of training data to retrain the model so that the prediction uncertainty on cluster 3 test samples can be reduced. Since the used 3 weeks of training data only take up 7% of the average lifetime for cluster 3 samples, using 1–2 more weeks train data still only covers the very early stage of these long-life cells.

Further analysis of uncertainty for model parameters can be found in Note S15 of the supplemental information. This uncertainty on both lifetime predictions and model parameters can be more beneficial to real-world applications compared to only a point-wise prediction. For example, instead of knowing the exact EOL lifetime,

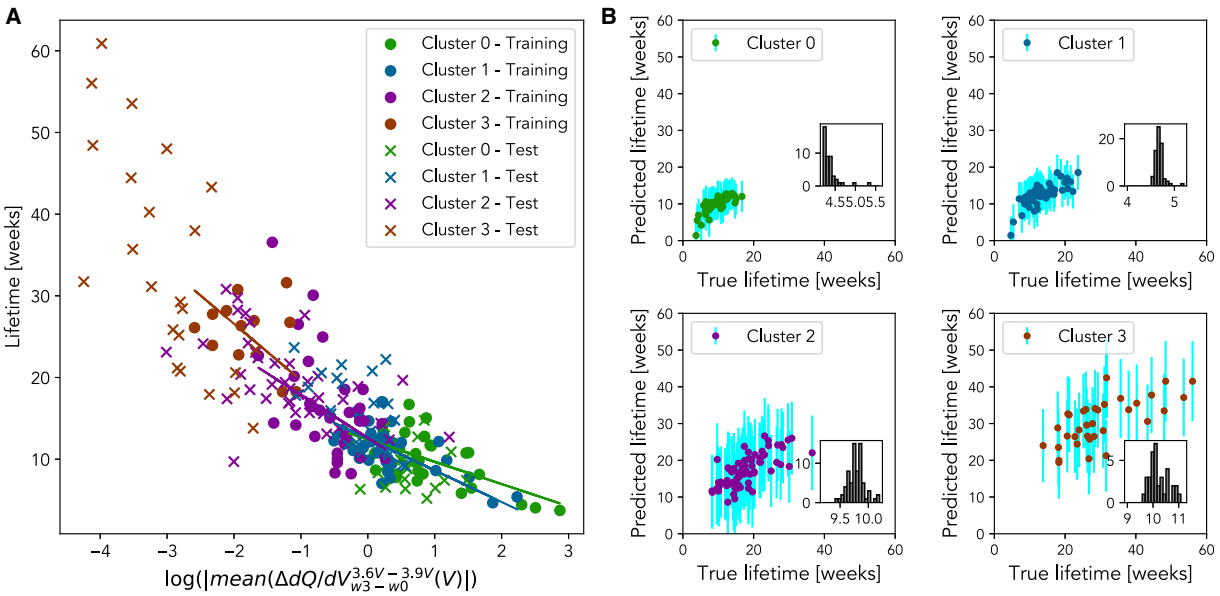


Figure 5. Overview of HBM prediction results

(A) Relationship between $\log(|\text{mean}(\Delta dQ/dV_{w3-w0}^{3.6V-3.9V}(V)|)$ and true lifetime across different clusters and train-test split ("Test" denotes samples from both high- and low-DoD sets). Fits, corresponding to mean parameter values, are plotted for each cluster.

(B) Predictions for each cluster with 2 standard deviations as the corresponding error bar for each sample. The embedded histograms show a summary of error bars.

customers care more about a warranty for the worst-case lifetime, which can be satisfied by using the standard deviation of prediction distributions.

In this study, we have developed two data-driven models to tackle the problem of battery early-life prediction on a large and unique aging dataset, which consists of 225 NMC cells cycled under a wide range of charge and discharge C-rates (0.5 C–3 C) and DoDs (4%–100%). Our feature engineering process identifies a new predictive feature, $\text{mean}(\Delta dQ/dV_{w3-w0}^{3.60V-3.90V}(V))$, derived from incremental capacity curves and closely related to the degradation induced by loss of active materials. Also, our analysis shows that the widely used $\Delta Q(V)$ features in the existing early prediction literature may not explain cell-to-cell lifetime variability within our dataset.

In terms of results, two distinct machine learning models are trained to predict the lifetime. Our degradation-informed model, trained using elastic net regression, yields 3.0 and 7.8 weeks RMSE and 15.1% and 33.0% MAPE on the high- and low-DoD test sets, respectively. The HBM produces 3.1 and 7.3 weeks RMSE and 16.9% and 21.8%

Table 3. Summary of train-test split for each cluster

		n samples		
Cluster ID	Stress_{avg}	Training	High-DoD test	Low-DoD test
0	1.12	30	18	0
1	0.95	41	24	4
2	0.76	33	18	22
3	0.51	12	0	23
Total	0.86	116	60	49

MAPE for the high- and low-DoD test sets, respectively. While the HBM shows performance improvement for point-wise predictions on the low-DoD test set, it also gives uncertainty information for its predictions, which can be used in applications like the cell lifetime warranty. And we found that the uncertainty grows across groups with the decrease of cycling stress factor $Stress_{avg}$, which indicates the lack of observability for cell-to-cell differences from early-life features, and thus more cycling time range may need to be included for cells under mild cycling conditions.

A limitation of this work is that the models are demonstrated on battery aging data collected in a well-controlled laboratory setting under constant cycling conditions over the life of the cells. However, depending on the applications, battery data from real-world applications may be more variable and noisy, posing a challenge to feature extraction and lifetime prediction. To investigate this further, we will expand the dataset by aging cells using electric grid duty cycle protocols (e.g., protocols simulating peak shaving and frequency regulation). This duty cycle dataset is currently being collected in the Reliability Engineering and Informatics Laboratory at the University of Connecticut. It will add a valuable addition to the ISU-ILCC dataset, particularly in evaluating lifetime models on cells with dynamic cycling profiles.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Requests for any additional information about this work should be directed to the lead contact, Chao Hu (chao.hu@uconn.edu).

Materials availability

No new material was generated in this study.

Data and code availability

- The battery aging dataset collected and used for this work is available for download from the open-access data repository of Iowa State University (Iowa State University DataShare) at https://doi.org/10.25380/iastate.22582234.[44] Please refer to the dataset as the *ISU-ILCC NMC/Gr battery aging dataset*. A sample code for preprocessing the data is included.
- The code for feature extraction and early prediction modeling is available at https://doi.org/10.5281/zenodo.10648587.

Cell and tester specifications

The Li-ion cells used in this study were commercial 502030 size Li-polymer cells with NMC as the positive electrode and graphite as the negative electrode, manufactured by Honghaosheng Electronics in Shenzhen, China. The rated capacity is 250 mAh (giving 1 C as 250 mA), and the operating voltage ranges from 3.0 to 4.2 V. All cells were tested on two 64-channel Neware BTS4000 battery testers, in thermal chambers set at 30°C.

Battery aging test design

The aging experiments were designed around three main stress factors that impact battery lifetime: charge rate (C_{chg}), discharge rate (C_{dchg}), and depth of discharge (DoD). To make the scope of our aging campaign manageable and work within the limitations of our equipment, we decided not to introduce temperature as an additional variable stress factor. Therefore, as also mentioned in "cell and tester specifications," all cells were placed in thermal chambers set at 30°C over the entire span of the aging experiments. The design of experiments for our aging campaign

involved deliberately subjecting various groups of cells to different stress levels between RPTs, achieved by randomly sampling the three stress factors within a wide design space. This treatment contributed to a wide distribution of the EOL in this dataset. However, this aging dataset is limited in reflecting realistic usage profiles, particularly dynamic charging/discharging profiles where the charge and discharge rates vary rapidly over time.

To track the full discharge capacity of cells with partial depths of discharge cycling, we periodically ran RPTs that measured cell capacity and gathered complete $Q(V)$ data for feature engineering. Each RPT consisted of two cycles performed at slow rates (C/2 and C/5) to capture cell voltage response while minimizing the impact of the cell kinetics. Before beginning the aging tests, an initial RPT was conducted to determine the beginning-of-life health. Aging tests consisted of 1 week of cycling followed by an RPT, and they were repeated until cell capacity decreased below 200 mAh (80% of the rated capacity). In real field operations, the necessity of incorporating RPTs to obtain consistent measurements (e.g., capacity-voltage data) from cells varies. Taking electric vehicles (EVs) as an example, the onboard charge controller could regulate the charging process, ensuring all EVs of the same model operate under certain constant-current conditions until sufficient data are collected. Then, data-driven diagnostics or prognostics can be achieved solely based on these data.[45] However, in grid storage applications, both the charging and discharging profiles can be random and noisy in operation, especially the profiles for frequency regulation.[46] Extracting clear predictive features can be practically infeasible in these applications. Therefore, periodic RPT cycles can help collect consistent measurements to estimate the state of health and predict lifetime. Since the grid may have multiple power-generation sources (e.g., power plants, wind farms, and solar farms) and large battery storage systems, the downtime of battery storage systems during diagnostic cycles, ideally executed in alternating batches, would have less impact on the overall operation of the grid or storage system. Furthermore, the frequency of RPTs can be less often, i.e., monthly or quarterly, to provide flexibility and avoid disruption in real applications—the weekly frequency in our aging campaign is to capture the rapid degradation due to the nature of the accelerated aging test.

As previously mentioned, four cells were cycled at each test condition. We refer to a specific cell using its group number and cell identifier, e.g., G7C3, where the numbers following each letter indicate the group and cell, respectively. Initially, we aimed to study two stress factors: DoD and C_{chg}. Conditions were selected using a grid search, with the discharge rate fixed at 0.5 C for all cells. Later, we expanded the dataset to study the third stress factor, C_{dchg}. Additional conditions were then selected using random sampling. The charge/discharge rates and depths of discharge were sampled evenly from the ranges 0.5 C–3 C and 25%–100%, respectively.

The cycling conditions for all cell groups can be found in Table S1 of the supplemental information. However, the depth of discharge design values do not exactly match the measured depths of discharge from the cycling experiments. When we programmed the cycling protocols, we determined the cutoff voltages using a reference discharge capacity vs. voltage curve from a cell cycled at C/2. Unfortunately, the voltage hysteresis that the cells experience under C/2 discharge causes the cells to reach the cutoff voltage quicker than expected, thus causing the difference between the measured and designed depth of discharge. In this paper, we present and discuss the depth of discharge using the actual

measured values since they more accurately represent the test conditions the cells experienced.

Extracting features from incremental capacity data

Extracting features from incremental capacity curves is a natural extension to using the $Q(V)$ discharge curve since it is defined over the same fixed voltage range for every cell. After fitting a spline and downsampling each cell's $Q(V)$ curve to 1,000 points, we calculated incremental capacity ($dQ/dV(V)$) as a finite difference approximation (difference quotient) of the first derivative of $Q(V)$ based on measurements of the Q and V time series.[5] It is well documented that incremental capacity analysis is an effective method for cell degradation diagnostics.[1,33,47] Measuring changes to the incremental capacity curve over the lifetime enables the diagnosis of different degradation modes, specifically loss of lithium inventory, and loss of active material in each electrode. Hence, we calculate core summary statistics of $\Delta dQ/dV(V)$ over a partial voltage interval so as to focus the feature extraction on specific areas that may correspond to specific degradation modes. This approach is inspired by work in Greenbank and Howey,[13] where the authors showed a strong correlation between the time a cell spends in a specific voltage interval and its capacity loss, although here the incremental capacity curve is a result of degradation rather than a cause. Instead of manually specifying the voltage interval to calculate the summary statistics, we exhaustively searched the entire $3.0 - 4.2$ V range in increments of 0.01 V, with a minimum window size of 0.02 V searching for the maximum Pearson correlation coefficient.

Literature in this field reports a wide range of possible features for lifetime prediction, either derived from capacity, voltage, and temperature measurements during cycling or other measurements such as impedance spectra from electrochemical impedance spectroscopy.[5,11,13,19,22] Most of these features are hand-crafted based on direct mathematical manipulation and are not always optimized for maximum correlation with lifetime. As a distinction from other feature extraction processes, the method we used to extract the $\Delta dQ/dV(V)$ feature optimizes the correlation coefficient by an exhaustive grid search. So, even though the optimal voltage interval identified for our aging dataset, i.e., $3.60 - 3.90$ V, may not directly apply to other aging datasets, the general methodology of optimizing a voltage interval when identifying early-life features allows researchers the possibility of extracting features of higher predictive power from their own datasets.

Extracting features from cycling conditions

As briefly mentioned in "extracting predictive features from early usage data," we consider a set of stress-related features for early-life prediction, which is $\text{Stress}_{\text{chg}} = C_{\text{chg}}{}^{0.5}\text{DoD}^{0.5}$. This feature captures the square-root-of-cycling charge throughput and is a proxy for diffusion-induced stress in the electrode active materials.[36–38] In addition to the charge-based feature, we also calculate a discharge feature, $\text{Stress}_{\text{dchg}} = C_{\text{dchg}}{}^{0.5}\text{DoD}^{0.5}$. Further, to capture the effects of different charge and discharge rates in a single feature, we calculate an average stress feature as $\text{Stress}_{\text{avg}} = (\text{Stress}_{\text{chg}} + \text{Stress}_{\text{dchg}})/2$ and also calculate a multiplicative stress feature as $\text{Stress}_{\text{mult}} = \text{Stress}_{\text{chg}} \cdot \text{Stress}_{\text{dchg}}$. For all features, we use the measured DoD from the first week of cycling in the calculation. A unique characteristic of these features is that they require no cell-specific measurements, assuming the calculation of DoD is accurate and accounts for voltage hysteresis. For this reason, these features are excellent candidates as condition-level grouping variables in our hierarchical Bayesian modeling approach to early prediction (see "hierarchical Bayesian models for early prediction").

Feature selection

To minimize collinearity and the risk of overfitting, we perform stepwise forward selection using a linear model and repeated cross-validation with $RMSE_{EOL}$ as the evaluation metric. Starting with a null model, one feature is added to the model for each step until the number of selected features reaches a preset threshold ($N_{feature} = 10$). During each step, all features are tested in the model, and the feature that reduces the mean of the cross-validation $RMSE_{EOL}$ the most is selected and added to the model for the next step. Simultaneously, we evaluated the selected model at each step using the standard deviation of the cross-validation $RMSE_{EOL}$. We then select the features to use corresponding to the set with a balance between low mean and small standard deviation of cross-validated $RMSE_{EOL}$. In practice, we tend toward selecting fewer features so that the resulting model will be less complex and extrapolate better.

Elastic net regression for lifetime prediction

To predict cell lifetime, we formulate a regression problem with the extracted early-life features $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m]$ as inputs and the measured cell lifetimes $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]^T$ in logarithmic scale as outputs, where m is the number of early-life features, and n is the number of cells. Each element of $\boldsymbol{X}$ is a column vector containing the specific features selected through the technique introduced in "feature selection." We assume that the lifetime is a linear function of the early-life features, giving

$$\widehat{\boldsymbol{y}} = f(\boldsymbol{X}) = \boldsymbol{\beta}_0 + \boldsymbol{X}\boldsymbol{\beta}_1, \qquad \text{(Equation 2)}$$

where $\boldsymbol{\beta}_0$ is an $n \times 1$ column vector of the intercept, and $\boldsymbol{\beta}_1$ is a vector of coefficients, one for each feature, $\boldsymbol{\beta}_1 = [\beta_1, \beta_2, \ldots, \beta_m]^T$.

To find the coefficients of this equation, we formulate an optimization problem with elastic net regularization, which is a combination of L_1 and L_2 penalization. The objective function is

$$\widehat{\boldsymbol{\beta}} = \underset{\beta_0, \boldsymbol{\beta}_1}{\operatorname{argmin}} \left(\|\boldsymbol{y} - \boldsymbol{\beta}_0 - \boldsymbol{X}\boldsymbol{\beta}_1)\|_2^2 + \lambda \left(\frac{1-\alpha}{2} \|\boldsymbol{\beta}\|_2^2 + \alpha \|\boldsymbol{\beta}\|_1 \right) \right), \qquad \text{(Equation 3)}$$

where α and λ are hyperparameters that control the balance between the L_1 and L_2 penalties and the magnitude of regularization, respectively. To select optimal values of α and λ, we perform repeated cross-validation using randomized dataset splits.

Hierarchical Bayesian models for early prediction

As a comparison and contrast to the method in the previous section, we also consider HBMs for lifetime prediction. These have a layered structure that can model changes in the feature-target relationship throughout the dataset. HBMs have been applied to model naturally structured data in various research fields, from ecology to sociology, psychology, and computer vision.[48,49] Several studies in the battery field explored the concept of HBM to solve different problems, such as estimating the state of health against Ah throughput[50] and identifying parameters of an equivalent circuit model.[51] However, their applications were not early-life prediction, and the model structures adopted in their studies differed from those used in this work.

Clustering for hierarchical modeling

For our problem of early-life prediction, features can be viewed as coming from two levels: the "cycling condition" level and the "individual cell" level. Condition-level features relate to user-defined test protocols rather than measured data. For our dataset, the charge/discharge C-rates and depth of discharge (C_{chg}, C_{dchg}, DoD), and any mathematical combination of these are all condition-level features. In contrast,

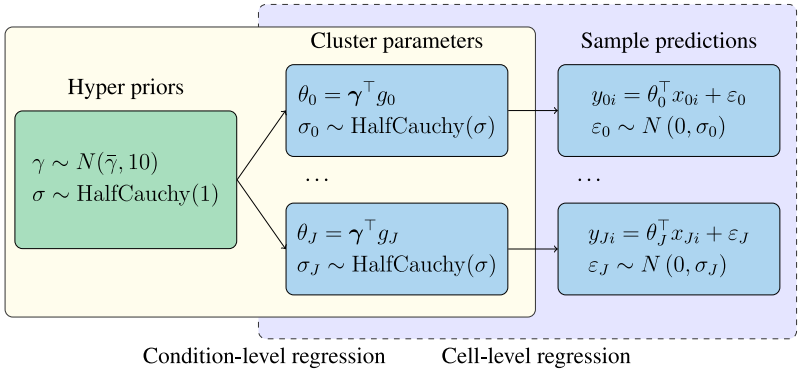


Figure 6. Overview of HBM structure

Model parameters can be classified as either individual-level ($\boldsymbol{\theta}_j, \boldsymbol{\sigma}_j$) or conditional-level ($\boldsymbol{\gamma}, \boldsymbol{\sigma}$); j represents cycling condition group index, i represents individual cell index, y_{ji} represents lifetime of i th cell in j th cycling group. The two-level structure allows the individual cell-level feature-label ($x_{ji} - y_{ji}$) relationship to vary with cycling condition based on cycling condition-level features ($\boldsymbol{g}_j$). The hyper-priors are assumed to be $\gamma \sim N(\bar{\gamma} = 0, 10)$ and $\sigma \sim \text{HalfCauchy}(\beta = 1)$.

features that require specific cell measurements during cycling are considered cell-level features. Features such as mean($\Delta dQ/dV_{w3-w0}^{3.60V-3.90V}(V)$) and var($\Delta Q_{w3-w0}(V)$) are examples of cell-level features that are unique to each cell.

To validate the hypothesis that conditional-level features have a strong impact on the relationship between cell-level features and lifetime, we calculate the condition-level feature $\text{Stress}_{avg} = (C_{chg}{}^{0.5}\text{DoD}^{0.5} + C_{dchg}{}^{0.5}\text{DoD}^{0.5})/2$ described in "extracting features from cycling conditions." This represents the average diffusion-induced stress that a cell experiences.[36] To take advantage of an HBM's ability to model the change in feature-target relationship across different levels, we investigate clustering cell data based on cycling conditions, quantified by average stress (Stress_{avg}). In general, we expect cells with similar average stress levels to share the same feature-lifetime relationship, enabling the HBM to better fit the dataset. We adopt a constrained K-means clustering algorithm,[52] which is an improved version of the traditional K-means algorithm that imposes minimum and maximum cluster size limits. An optimal cluster number $K = 4$ is used in later analysis, where details can be found in Note S15 of the supplemental information.

Bayesian hierarchical linear model

Similar to the HBM used in our preliminary study,[53] the model in this study also has a bi-level structure, as shown in Figure 6. The first level considers the cycling condition parameters. As mentioned previously, cells are first divided into four clusters (indexed from 0) based on their average stress Stress_{avg}, calculated using the cycling condition parameters.

At this level, we aim to find the mapping (parameterized by $\boldsymbol{\gamma}, \sigma$) between condition-level features ($\boldsymbol{g}_j$) and the cell-level regression parameters ($\boldsymbol{\theta}_j, \sigma_j$). Notice that, different from Zhou and Howey,[53] the noise terms σ_j are assumed to come from the same hyper-prior distribution HalfCauchy(σ):

$$\begin{aligned} \boldsymbol{\theta}_j &= \boldsymbol{\gamma}^\top \boldsymbol{g}_j \\ \sigma_j &\sim \text{HalfCauchy}(\sigma) \end{aligned} \qquad \text{(Equation 4)}$$

After the coefficients ($\boldsymbol{\theta}_j, \sigma_j$) are decided for each cluster, the individual cell-level regression is built as the second level of the HBM. The cell-level regression uses

individual health features (x_{ji}) and coefficients (θ_j, σ_j) to give lifetime predictions (y_{ji}) for individual cells.

$$y_{ji} \sim N\left(\theta_j{}^{\top} x_{ji}, \sigma_j^2\right) \quad \text{(Equation 5)}$$

The overall training objective is to infer posterior distributions for both the condition-level model and the individual cell-level models, $P(\theta_j|Y_j)$ and $P(\gamma|\{Y\})$ respectively, where Y_j represents lifetimes from only the j th group but $\{Y\}$ represents data from all lifetimes. More details about the training procedure are included in Note S14 of the supplemental information.

Model evaluation metrics

We use two standard error metrics to evaluate the lifetime prediction accuracy of our approaches, namely, mean absolute percentage error ($MAPE_{EOL}$) and root mean squared error ($RMSE_{EOL}$), both calculated using the measured and predicted values of cell lifetime on a linear scale. The metrics are

$$MAPE_{EOL} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad \text{(Equation 6)}$$

$$RMSE_{EOL} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad \text{(Equation 7)}$$

where y are the measured cell lifetimes, $\hat{y}$ are the predicted cell lifetimes, and n is the number of cells. These two error metrics are commonly used to measure the prediction accuracy in regression problems. MAPE measures the overall prediction error on a percentage scale, while RMSE penalizes predictions with larger errors, making this error metric more sensitive to outliers.[54] Lower values are more desirable for both error metrics, indicating the model has better prediction accuracy.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2024.101891.

ACKNOWLEDGMENTS

We acknowledge the hard work of Jinqiang Liu from Iowa State University and Chad Tischer and Reuben D. Schooley from Iowa Lakes Community College for executing and maintaining the battery aging tests. We also want to acknowledge Murtaza Zohair for assembling the half-cells used in this study.

The work at Iowa State University and the University of Connecticut was partly supported by the Iowa Economic Development Authority under the Iowa Energy Center grant no. 20-IEC-018 and partly by the US National Science Foundation under grant no. ECCS-2015710. The China Scholarship Council and the Department of Engineering Science supported the work at the University of Oxford. Any opinions, findings, or conclusions in this paper are those of the authors and do not necessarily reflect the sponsors' views.

AUTHOR CONTRIBUTIONS

Conceptualization, T.L., A.T., Z.Z., C.H., and D.H.; data collection, data management, and raw data processing, T.L.; investigation, methodology, visualization,

software, formal analysis, and writing – original draft, T.L., A.T., and Z.Z.; writing – review and editing, T.L., A.T., Z.Z., C.H., and D.H.

DECLARATION OF INTERESTS

D.H. is a co-founder of Brill Power Ltd.

Received: August 25, 2023
Revised: November 22, 2023
Accepted: February 29, 2024
Published: March 22, 2024

REFERENCES

1. Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., and Howey, D.A. (2017). Degradation diagnostics for lithium ion cells. J. Power Sources *341*, 373–386. https://doi.org/10.1016/j.jpowsour.2016.12.011.

2. Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y.T., Steinbacher, F., Khan, M.U., Lee, J.W., Siegel, J.B., Stefanopoulou, A.G., and Howey, D.A. (2021). The challenge and opportunity of battery lifetime prediction from field data. Joule *5*, 1934–1955. https://doi.org/10.1016/j.joule.2021.06.005.

3. Thelen, A., Lui, Y.H., Shen, S., Laflamme, S., Hu, S., Ye, H., and Hu, C. (2022). Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries. Energy Storage Mater. *50*, 668–695. https://doi.org/10.1016/j.ensm.2022.05.047.

4. Kunz, M.R., Dufek, E.J., Yi, Z., Gering, K.L., Shirk, M.G., Smith, K., Chen, B.-R., Wang, Q., Gasper, P., Bewley, R.L., and Tanim, T.R. (2021). Early battery performance prediction for mixed use charging profiles using hierarchal machine learning. Batter. Supercaps *4*, 1186–1196. https://doi.org/10.1002/batt.202100079.

5. Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy *4*, 383–391. https://doi.org/10.1038/s41560-019-0356-8.

6. Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.-H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature *578*, 397–402. https://doi.org/10.1038/s41586-020-1994-5.

7. Dave, A., Mitchell, J., Kandasamy, K., Wang, H., Burke, S., Paria, B., Póczos, B., Whitacre, J., and Viswanathan, V. (2020). Autonomous discovery of battery electrolytes with robotic experimentation and machine learning. Cell Reports Physical Science *1*, 100264. https://doi.org/10.1016/j.xcrp.2020.100264.

8. Smith, A.J., Burns, J.C., Trussler, S., and Dahn, J.R. (2010). Precision measurements of the coulombic efficiency of lithium-ion batteries and of electrode materials for lithium-ion batteries. J. Electrochem. Soc. *157*, A196. https://doi.org/10.1149/1.3268129.

9. Burns, J.K., Nashed, J.Y., Blohm, G., Eberman, K., Scott, E., Gardner, J., and Dahn, J. (2011). Evaluation of effects of additives in wound Li-ion cells through high precision coulometry. J. Vis. *11*, 3. https://doi.org/10.1149/1.3531997.

10. Burns, J.C., Kassam, A., Sinha, N.N., Downie, L.E., Solnickova, L., Way, B., and Dahn, J.R. (2013). Predicting and extending the lifetime of Li-ion batteries. J. Electrochem. Soc. *160*, A1451–A1456. https://doi.org/10.1149/2.060309jes.

11. Baumhöfer, T., Brühl, M., Rothgang, S., and Sauer, D.U. (2014). Production caused variation in capacity aging trend and correlation to initial cell performance. J. Power Sources *247*, 332–338. https://doi.org/10.1016/j.jpowsour.2013. 08.108.

12. Harris, S.J., Harris, D.J., and Li, C. (2017). Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells. J. Power Sources *342*, 589–597. https://doi.org/10.1016/j.jpowsour.2016.12.083.

13. Greenbank, S., and Howey, D. (2022). Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life. IEEE Trans. Ind. Inf. *18*, 2965–2973. https://doi.org/10.1109/TII. 2021.3106593.

14. Zhang, Y., Peng, Z., Guan, Y., and Wu, L. (2021). Prognostics of battery cycle life in the early-cycle stage based on hybrid model. Energy *221*, 119901. https://doi.org/10.1016/j.energy.2021.119901.

15. Yang, F., Wang, D., Xu, F., Huang, Z., and Tsui, K.-L. (2020). Lifespan prediction of lithium-ion batteries based on various extracted features and gradient boosting regression tree model. J. Power Sources *476*, 228654. https://doi.org/10.1016/j.jpowsour.2020.228654.

16. Weng, A., Mohtat, P., Attia, P.M., Sulzer, V., Lee, S., Less, G., and Stefanopoulou, A. (2021). Predicting the impact of formation protocols on battery lifetime immediately after manufacturing. Joule *5*, 2971–2992. https://doi.org/10.1016/j.joule.2021.09.015.

17. Saxena, S., Ward, L., Kubal, J., Lu, W., Babinec, S., and Paulson, N. (2022). A convolutional neural network model for battery capacity fade curve prediction using early life data. J. Power Sources *542*, 231736. https://doi.org/10.1016/j.jpowsour.2022.231736.

18. Herring, P., Balaji Gopal, C., Aykol, M., Montoya, J.H., Anapolsky, A., Attia, P.M., Gent, W., Hummelshøj, J.S., Hung, L., Kwon, H.-K., et al. (2020). BEEP: A python library for battery evaluation and early prediction. SoftwareX *11*, 100506. https://doi.org/10.1016/j.softx.2020.100506.

19. Fei, Z., Yang, F., Tsui, K.-L., Li, L., and Zhang, Z. (2021). Early prediction of battery lifetime via a machine learning based framework. Energy *225*, 120205. https://doi.org/10.1016/j.energy.2021.120205.

20. Fermín-Cueto, P., McTurk, E., Allerhand, M., Medina-Lopez, E., Anjos, M.F., Sylvester, J., and Dos Reis, G. (2020). Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. Energy and AI *1*, 100006. https://doi.org/10.1016/j.egyai.2020.100006.

21. Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D.U. (2021). One-shot battery degradation trajectory prediction with deep learning. J. Power Sources *506*, 230024. https://doi.org/10.1016/j.jpowsour.2021. 230024.

22. Paulson, N.H., Kubal, J., Ward, L., Saxena, S., Lu, W., and Babinec, S.J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. J. Power Sources *527*, 231127. https://doi.org/10.1016/j.jpowsour. 2022.231127.

23. Attia, P.M., Bills, A., Brosa Planella, F., Dechent, P., Dos Reis, G., Dubarry, M., Gasper, P., Gilchrist, R., Greenbank, S., Howey, D., et al. (2022). "Knees" in lithium-ion battery aging trajectories. J. Electrochem. Soc. *169*, 060517. https://doi.org/10.1149/1945-7111/ac6d13.

24. Waldmann, T., Hogg, B.-I., and Wohlfahrt-Mehrens, M. (2018). Li plating as unwanted side reaction in commercial Li-ion cells–A review. J. Power Sources *384*, 107–124. https://doi.org/10.1016/j.jpowsour.2018.02.063.

25. Han, X., Lu, L., Zheng, Y., Feng, X., Li, Z., Li, J., and Ouyang, M. (2019). A review on the key issues of the lithium ion battery degradation among the whole life cycle. ETransportation *1*, 100005. https://doi.org/10.1016/j.etran.2019.100005.

26. Dos Reis, G., Strange, C., Yadav, M., and Li, S. (2021). Lithium-ion battery data and where to find it. Energy and AI *5*, 100081. https://doi.org/10.1016/j.egyai. 2021.100081.

27. Bole, B., Kulkarni, C.S., and Daigle, M. (2014). Adaptation of an electrochemistry-based li-ion battery model to account for deterioration

observed under randomized use. In Annual Conference of the PHM Society. https://doi.org/10.36001/phmconf.2014.v6i1.2490.

28. Saha, B., Goebel, K., Poll, S., and Christophersen, J. (2009). Prognostics methods for battery health monitoring using a Bayesian framework. IEEE Trans. Instrum. Meas. *58*, 291–296. https://doi.org/10.1109/TIM.2008.2005965.

29. He, W., Williard, N., Osterman, M., and Pecht, M. (2011). Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. J. Power Sources *196*, 10314–10321. https://doi.org/10.1016/j.jpowsour.2011.08.040.

30. Xing, Y., Ma, E.W., Tsui, K.-L., and Pecht, M. (2013). An ensemble model for predicting the remaining useful performance of lithium-ion batteries. Microelectron. Reliab. *53*, 811–820. https://doi.org/10.1016/j.microrel.2012.12.003.

31. Preger, Y., Barkholtz, H.M., Fresquez, A., Campbell, D.L., Juba, B.W., Romàn-Kustas, J., Ferreira, S.R., and Chalamala, B. (2020). Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. J. Electrochem. Soc. *167*, 120532. https://doi.org/10.1149/1945- 7111/abae37.

32. Dechent, P., Greenbank, S., Hildenbrand, F., Jbabdi, S., Sauer, D.U., and Howey, D.A. (2021). Estimation of Li-Ion Degradation Test Sample Sizes Required to Understand Cell-to-Cell Variability. Batter. Supercaps *4*, 1821–1829. https://doi.org/10.1002/batt.202100148.

33. Pastor-Fernández, C., Uddin, K., Chouchelamane, G.H., Widanage, W.D., and Marco, J. (2017). A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. J. Power Sources *360*, 301–318. https://doi.org/10.1016/j. jpowsour.2017.03.042.

34. Berecibar, M., Dubarry, M., Omar, N., Villarreal, I., and Van Mierlo, J. (2016a). Degradation mechanism detection for NMC batteries based on Incremental Capacity curves. World Electric Vehicle Journal *8*, 350–361. https://doi.org/10.3390/wevj8020350.

35. Berecibar, M., Dubarry, M., Villarreal, I., Omar, N., and Van Mierlo, J. (2016b). "Degradation mechanisms detection for HP and HE NMC cells based on incremental capacity curves". In 2016 IEEE Vehicle Power and Propulsion Conference (VPPC) (IEEE), pp. 1–5. https://doi.org/10.1109/VPPC.2016.7791648.

36. Smith, K., Gasper, P., Colclasure, A.M., Shimonishi, Y., and Yoshida, S. (2021). Lithium-Ion Battery Life Model with Electrode Cracking and Early-Life Break-in Processes. J. Electrochem. Soc. *168*, 100530. https://doi.org/10.1149/1945-7111/ac2ebd.

37. Reniers, J.M., Mulder, G., and Howey, D.A. (2019). Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries. J. Electrochem. Soc. *166*, A3189–A3200. https://doi.org/10.1149/2.0281914jes.

38. Smith, K., and Wang, C.-Y. (2006). Solid-state diffusion limitations on pulse operation of a lithium ion cell for hybrid electric vehicles. J. Power Sources *161*, 628–639. https://doi.org/10.1016/j.jpowsour.2006.03.050.

39. Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography *36*, 27–46. https://doi.org/10.1111/j.1600-0587.2012. 07348.x.

40. Cai, J., Luo, J., Wang, S., and Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing *300*, 70–79. https://doi.org/10.1016/j.neucom.2017.11.077.

41. Hastie, T.J., Tibshirani, R.J., and Friedman, J.H. (2013). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. Eng, 2nd. (Springer).

42. Gasper, P., Collath, N., Hesse, H.C., Jossen, A., and Smith, K. (2022). Machine-Learning Assisted Identification of Accurate Battery Lifetime Models with Uncertainty. J. Electrochem. Soc. *169*, 080518. https://doi.org/10.1149/1945-7111/ac86a8.

43. Jiang, B., Gent, W.E., Mohr, F., Das, S., Berliner, M.D., Forsuelo, M., Zhao, H., Attia, P.M., Grover, A., Herring, P.K., et al. (2021). Bayesian learning for rapid prediction of lithium-ion battery-cycling protocols. Joule *5*, 3187–3203. https://doi.org/10.1016/j.joule.2021.10.010.

44. Li, T., Thelen, A., Liu, J., Tischer, C., and Hu, C. (2023) (ISU-ILCC Battery Aging Dataset). https://doi.org/10.25380/iastate.22582234.

45. Richardson, R.R., Birkl, C.R., Osborne, M.A., and Howey, D.A. (2019). Gaussian Process Regression for In Situ Capacity Estimation of Lithium-Ion Batteries. IEEE Trans. Ind. Inf. *15*, 127–138. https://doi.org/10.1109/TII.2018.2794997.

46. Kim, N., Shamim, N., Crawford, A., Viswanathan, V.V., Sivakumar, B.M., Huang, Q., Reed, D., Sprenkle, V., and Choi, D. (2022). "Comparison of Li-ion battery chemistries under grid duty cycles". en. J. Power Sources *546*, 231949. https://doi.org/10.1016/j.jpowsour.2022.231949.

47. Dubarry, M., and Anseán, D. (2022). Best practices for incremental capacity analysis. Front. Energy Res. *10*. https://doi.org/10.3389/fenrg.2022.1023555.

48. Lake, B.M., Salakhutdinov, R., and Tenenbaum, J.B. (2015). Human-level concept learning through probabilistic program induction. Science *350*, 1332–1338. https://doi.org/10.1126/science.aab3050.

49. Pedersen, E.J., Miller, D.L., Simpson, G.L., and Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ *7*, e6876. https://doi.org/10.7717/peerj.6876.

50. Jafari, M., Brown, L.E., and Gauchia, L. (2019). Hierarchical Bayesian model for probabilistic analysis of electric vehicle battery degradation. IEEE Trans. Transp. Electrific. *5*, 1254–1267. https://doi.org/10.1109/TTE.2019.2956350.

51. Mishra, M., Martinsson, J., Rantatalo, M., and Goebel, K. (2018). Bayesian hierarchical model-based prognostics for lithium-ion batteries. Reliab. Eng. Syst. Saf. *172*, 25–35. https://doi.org/10.1016/j.ress.2017. 11.020.

52. Bhattacharya, A., Jaiswal, R., and Kumar, A. (2018). Faster algorithms for the constrained k-means problem. Theor. Comput. Syst. *62*, 93–115. https://doi.org/10.1007/s00224-017-9820-7.

53. Zhou, Z., and Howey, D.A. (2023). Bayesian hierarchical modelling for battery lifetime early prediction. IFAC-PapersOnLine *56*, 6117–6123. https://doi.org/10.1016/j.ifacol.2023.10.708.

54. Naser, M.Z., and Alavi, A.H. (Dec. 2023). Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences. Archit. Struct. Constr. *3*, 499–517. https://doi.org/10.1007/s44150-021-00015-8.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Bayesian hierarchical modelling for battery lifetime early prediction
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	Zhou, Zihao, and David A. Howey. "Bayesian hierarchical modelling for battery lifetime early prediction." IFAC-PapersOnLine 56.2 (2023): 6117-6123.

Student Confirmation

Student Name:	Zihao Zhou		
Contribution to the Paper	First author I developed the core idea of using hierarchical Bayesian linear regression model to address the problem of battery lifetime early prediction. I implemented the corresponding code, conducted all experiments for the paper, performed additional analyses, and authored the whole manuscript.		
Signature	Zihao Zhou	Date	09 July 2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: David A. Howey, Professor of Engineering Science			
Supervisor comments Zihao led this work and the description given above is accurate.			
Signature	David Howey	Date	15 Aug 2025

This completed form should be included in the thesis, at the end of the relevant chapter.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Predicting battery lifetime under varying usage conditions from early aging data
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	Li Tingkai, Zihao Zhou, Adam Thelen, David A. Howey, and Chao Hu. "Predicting battery lifetime under varying usage conditions from early aging data." Cell Reports Physical Science 5, no. 4 (2024).

Student Confirmation

Student Name:	Zihao Zhou		
Contribution to the Paper	Co-first author I developed the core idea of using hierarchical Bayesian model (HBM) to address the problem of battery lifetime early prediction. I helped with health feature extraction. I implemented the corresponding code for HBM, conducted corresponding experiments for the paper, performed additional analyses. I authored the manuscript, focusing primarily on the section of Results and discussion, as well as model description for HBM.		
Signature	Zihao Zhou	Date	09 July 2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: David A. Howey, Professor of Engineering Science			
Supervisor comments Zihao led this work and the description given above is accurate.			
Signature	David Howey	Date	15 Aug 2025

This completed form should be included in the thesis, at the end of the relevant chapter.

3 | Physics-informed data-driven models for SOH estimation

3.1 Preface

This chapter is based on the following publication:

> [3] **Zhou Zihao**, Antti Aitio, and David Howey. "Learning Li-ion battery health and degradation modes from data with aging-aware circuit models." *Applied Energy* 397 (2025): 126375.

In the previous chapter we showed that data-driven methods, when combined with carefully extracted health features, can achieve good lifetime prediction accuracy. However, in reality, a point-wise prediction of the lifetime is not sufficient. For example, BMSs require accurate SOH estimates throughout the battery lifetime to support more reliable daily state-of-charge (SOC) estimation and enable preventative maintenance. Moreover, historical SOH estimates can be used for future extrapolation to get lifetime predictions.

As discussed in Section 1.3, both model-based and data-driven methods have been proposed for SOH estimation. Pure data-driven approaches are relatively easy to parameterize and can achieve high estimation accuracy on lab-test datasets. However, three difficulties limit the usage of proposed HBM for SOH estimation: First, the health features may not be available across battery operation lifetime. Uniform individual level features are normally extracted from controlled lab aging tests. For cells used in field, limited depth of discharge, and highly dynamic cycling profiles makes feature extraction challenging. Second, the mapping between health features and current SOH may vary across battery operation lifetime, which means the HBM needs to be frequently retrained or fine-tuned. This is different from EoL/RUL prediction, where the mapping between health features and EoL/RUL labels only need to be learnt once on training data. Third, at least dozens of training samples are required to capture feature-label mapping, which makes it unable to address individual cells when only limited training data is available.

Together with the varying feature-label relationship across lifetime, it is very challenging for pure regression models like HBM to deal with lifelong SOH estimation problem.

On the other hand, model-based SOH estimation involves repeatedly parameterizing a battery model—either equivalent circuit models (ECMs) or electrochemical models (EMs)—to reflect SOH changes over time. As prior knowledge about battery dynamics are encoded in these battery models, and SOH is often one of their model parameters, it can be estimated for each individual cell only from its own cycling data, without requiring other training samples. The major limitation arises from poor parameter identifiability based solely on standard terminal voltage and current measurements. Due to their relative simplicity compared to EMs, ECMs are more suitable for real-world SOH estimation. However, their accuracy can degrade significantly when operating conditions differ from those under which the model parameters were initially identified. This is reflected in the variation of circuit parameters with changes in temperature and SOC [51]. As a result, SOH estimation using ECMs may fail to converge, resulting in an unstable aging trajectory estimation over battery lifetime.

In Chapter 3, to address these issues, we introduce an aging-aware ECM that incorporates a Bayesian nonparametric approach (Gaussian process regression, GPR) to model ECM parameters as functions of both operating conditions and lifetime. A state space formulation of a GP was used to convert the GPs into time-stepping models, enabling inference with linear time complexity. Also, it enables the aging dynamics represented by GPR being coupled with battery operation dynamics represented by ECM. Here, the GPR is not built between health features and current SOH, it is built between discretized aging times and life-long SOH labels, while the SOH labels are estimated through ECM. The GPR is effectively used to capture and smoothing the long-term SOH decaying trend, not transient mapping between health features and current SOH. Thus, it avoids the difficulties mentioned previously for HBM. The model was built on approaches developed previously [79, 80] but with a simpler ECM emphasizing the importance of the estimated resistance. It was found aging-induced shape changes of the OCV-SOC curve largely influence estimates of the resistance function, which enables further estimation for differential voltage curves. This approach effectively combines model-based

and data-driven SOH estimation: the model-based component comes from the battery dynamics described by the ECM, while the data-driven component uses GPR to capture the dependency of the ECM functional parameters on operating conditions and aging.

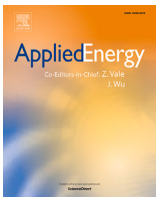
3.2 Manuscript

Applied Energy 397 (2025) 126375

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apen



Learning Li-ion battery health and degradation modes from data with aging-aware circuit models

Zihao Zhou [a], Antti Aitio [a], David Howey [a,b,*]

[a] Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, Oxford, UK
[b] The Faraday Institution, Didcot, OX11 0RA, UK

HIGHLIGHTS

- An aging-aware equivalent circuit model is proposed for battery health estimation.
- It estimates capacity and resistance only from operational data without labels.
- Changes of open-circuit voltage are strongly reflected in resistance estimation.
- It may enable an in operando estimation of differential voltage curves.

ARTICLE INFO

Keywords:
Battery
Health
Estimation
Gaussian process
State space
Equivalent circuit

ABSTRACT

Non-invasive estimation of Li-ion battery state-of-health from operational data is valuable for battery applications, but remains challenging. Pure model-based methods may suffer from inaccuracy and long-term instability of parameter estimates, whereas pure data-driven methods rely heavily on training data quality and quantity, causing lack of generality when extrapolating to unseen cases. We apply an aging-aware equivalent circuit model for health estimation, combining the flexibility of data-driven techniques within a model-based approach. A simplified electrical model with voltage source and resistor incorporates Gaussian process regression to learn capacity fade over time and also the dependence of resistance on operating conditions and time. The approach was validated against two datasets and shown to give accurate performance with less than 1 % relative root mean square error (RMSE) in capacity and less than 2 % mean absolute percentage error (MAPE). Critically, we show that changes from the open circuit voltage versus state-of-charge function will strongly influence the learnt resistance. We use this feature to further estimate *in operando* differential voltage curves from operational data.

1. Introduction

Lithium-ion batteries play an important role in the transition to clean energy. Demand is increasing rapidly for grid storage, electric vehicle and home energy storage applications [38]. One of the most crucial challenges for batteries is estimating their state of health (SOH). Accurate estimation of SOH improves lifetime, system safety, and costs of warranties and maintenance. Many studies [14,33,46] also discuss the importance of dealing with retired batteries, the number of which will increase drastically in the near future. Estimating the evolution of SOH from historical usage data is crucial for enabling possible 'second-life' applications that rely on accurate knowledge of battery residual value.

A commonly used SOH indicator, discharge capacity, normally requires the battery to undergo a low-rate constant-current discharge from 100 % to 0 % state of charge (SOC). This is time consuming, and therefore often not undertaken in real-world applications. Therefore, extensive studies have been done on faster capacity estimation methods [4,40]. These are usually categorised into model-driven and data-driven approaches. Model-driven methods use a specific battery model, such as an electrochemical model or electrical equivalent circuit model (ECM), and update the parameters by repeated fitting as the battery ages [10,25]. Electrochemical models usually need half-cell test data for parameterization and tend to suffer from parameter identifiability problems [23,24]. Equivalent-circuit models, the focus in this paper, have

* Corresponding author at: Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, Oxford, UK
Email addresses: zihao.zhou@eng.ox.ac.uk (Z. Zhou), antti.aitio@gmail.com (A. Aitio), david.howey@eng.ox.ac.uk (D. Howey).

https://doi.org/10.1016/j.apenergy.2025.126375
Received 7 March 2025; Received in revised form 28 May 2025; Accepted 23 June 2025
Available online 2 July 2025

fewer parameters but are less connected to physical processes. Recursive state estimation techniques, such as extended Kalman filters [41] and particle filters [42], are often adopted for co-estimation of model states and parameters.

On the other hand, data-driven methods for capacity estimation try to map from carefully selected features to corresponding capacity values [19]. Several features indicative of battery usage and aging may be extracted from cell voltage, current, and temperature measurements [12,17,32]. Then, machine learning models such as support vector machines [8], random forests [20], neural networks [18], and Gaussian processes [28], have been used to capture the relationship between these features and capacity. Depending on the choice of inputs, data-driven models can be used either for estimating present SOH, or predicting the future SOH. If the learnt mapping is between input features and future capacity values, these models can be directly used for forecasting. The performance of machine learning models for battery health is largely determined by the breadth, quantity and accuracy of the data used for model training.

Moreover, capacity fade only represents one aspect of battery degradation—increasing impedance (often just represented by resistance) also plays an essential role in battery health [9], leading to power fade, which is often correlated with capacity fade. Impedance values may be easier to estimate from cycling data and field data, where capacity measurements are very challenging [27,36]. Many SOH estimation works treat impedance as a scalar or rely on a look-up table that needs to be calibrated in standard reference tests [22,25]. However, many experimental works show that impedance does change with operating conditions [31,39].

In this work, we use an aging-aware electrical equivalent circuit model to estimate both capacity and operando battery impedance (resistance) from cycling data. The model parameters are described as functions of aging time and operating conditions (SOC and applied current in this case) and these functions are learned from data using Gaussian process regression. The model builds on former work [2] but with a simpler ECM emphasizing the importance of the estimated resistance. Results show that the learned resistance function is strongly impacted by the assumptions about the battery open circuit voltage (OCV) versus SOC function. Specifically, aging-induced shape changes in OCV-SOC function are revealed indirectly through the learned resistance function. This leads to a method to estimate differential voltage (dV/dQ) curves indirectly from impedance data. Open-source datasets from two batteries with different cycling conditions were used to validate model performance. Our method not only gives accurate performance with less than 1 % relative RMSE and less than 2 % MAPE for capacity estimation, but also enables dV/dQ estimation. Differential voltage analysis enables deeper understanding of battery aging by quantifying metrics such as loss of lithium inventory (LLI) and loss of active material in each electrode ($\mathrm{LAM_n}$, $\mathrm{LAM_p}$) [6,7].

2. Methodology

2.1. State space formulation of Gaussian processes

In this work Gaussian processes (GPs) are used to fit functions to data. Full details of GPs are documented in several books and papers, for example Williams and Rasmussen [44]—we give only a brief introduction here. A Gaussian process (GP) is defined as a distribution of functions over input $\mathbf{x}$, characterised by a mean and covariance, with the covariance defined by a kernel function $k(\mathbf{x}, \mathbf{x}')$,

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}'\right)\right). \tag{1}$$

When this input includes both locations in space and time $\mathbf{x} = (s, t)$, it is often called a spatial-temporal Gaussian process [34],

$$f(s,t) \sim \mathcal{GP}\left(m(s,t), k\left((s,t),(s',t')\right)\right). \tag{2}$$

Without loss of generality, the prior mean function $m(\mathbf{x})$ may be set to zero. By assuming a measurement noise term ϵ, the mapping between $\mathbf{x}$ and $\mathbf{y}$ can be expressed as

$$\mathrm{y} = f(\mathbf{x}) + \epsilon, \epsilon \sim \mathcal{N}\left(0, \sigma_n^2 \mathbf{I}\right). \tag{3}$$

To make predictions, we require the posterior predictive distribution at test samples $\mathbf{X}^*$ based on observations from a labeled training set of input-output data $\mathcal{D} = \left\{\left(\mathbf{x}_i, y_i\right)\right\}_{i=1}^{N}$. In this case, the analytical solution is given by [44],

$$p\left(\mathbf{y}^* \mid X^*, X, \mathbf{y}\right) = \mathcal{N}\left(\mathbf{y}^* \mid \mathbf{m}^*, \Sigma^*\right), \tag{4}$$

where

$$\begin{aligned} \mathbf{m}^* &= \mathbf{K}_{X,*}^T \left[\mathbf{K}_X + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{y} \\ \Sigma^* &= \mathbf{K}_{*,*} - \mathbf{K}_{X,*}^T \left[\mathbf{K}_X + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{K}_{X,*}, \end{aligned} \tag{5}$$

and the kernel matrices are simplified as $\mathbf{K}_X = k(\mathbf{X}, \mathbf{X})$, $\mathbf{K}_{X,*} = k(\mathbf{X}, \mathbf{X}^*)$. For a zero-mean GP, the model is defined by its training data and the selected kernel function $k(\mathbf{x}, \mathbf{x}')$ including fitted hyperparameters. The negative log maximum likelihood (NLML) estimates of kernel hyperparameters θ are given by

$$\begin{aligned} \mathrm{NLML} = -\log p(\mathbf{y} \mid X, \theta) = &-\frac{1}{2}\mathbf{y}^{\mathrm{T}} \left[\mathbf{K}_X + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{y} \\ &- \frac{1}{2}\log\left|\mathbf{K}_X + \sigma_n^2\right| - \frac{n}{2}\log 2\pi. \end{aligned} \tag{6}$$

Large datasets pose challenges for GPs due to the requirement for inversion of an $N \times N$ matrix (Eqs. 5 and 6) which scales computationally at $\mathcal{O}\left(n^3\right)$. To address this, we adopt a state-space formulation of a GP that instead scales linearly with the 'time' dimension, $\mathcal{O}(n)$.

Proposed by Särkkä et al. [29] and Solin [34], the state-space formulation interprets a spatial-temporal Gaussian process $f(s,t)$ as the solution to a linear time-invariant stochastic partial differential equation (LTI-SPDE),

$$\begin{aligned} \frac{\partial \mathbf{x}(s,t)}{\partial t} &= \mathcal{F}\mathbf{x}(s,t) + \mathbf{L}\omega(s,t) \\ \mathrm{y_t} &= \mathcal{H}_\mathrm{t}\mathbf{x}(s,t) + \epsilon_\mathrm{t}, \quad \epsilon \sim \mathcal{N}\left(0, \sigma_n^2\right). \end{aligned} \tag{7}$$

We can evaluate this system of equations over a finite collection of spatial points of interest $\{s_i\}_1^{n_\mathrm{s}}$. This results in n_s sets of state vectors, $\left(\mathbf{x}\left(s_1, t\right), \mathbf{x}\left(s_2, t\right), \ldots, \mathbf{x}\left(s_{n_s}, t\right)\right)$, where each is

$$\mathbf{x}\left(s_i, t\right) = \left[f\left(s_i, t\right), \frac{df\left(s_i, t\right)}{dt}, \ldots, \frac{df\left(s_i, t\right)^{p-1}}{dt^{p-1}}\right]^\top. \tag{8}$$

Here, $\mathcal{F}$ and $\mathcal{H}$ are linear operators, $\mathbf{L}$ is a dispersion matrix, and $\omega(\mathrm{x}, t)$ is a spatially resolved white noise process decided by the spatial kernel. The mathematical foundation comes from the Wiener-Khinchin theorem [5], which indicates that the GP kernel covariance matrix and the spectral density of a solution process from a LTI-SPDE form a Fourier-transform pair. At a specific spatial point s_i, all of the time gradient information up to order $p-1$ is maintained and propagated into future, and this propagation occurs for all n_s spatial points. The exact value for p is decided by the selected time kernel. Intuitively, one can view this approach as constructing a LTI system that takes a white noise process as input and generates an output that is the desired Gaussian process.

To further simplify the problem, we assume kernel separability,

$$k\left((s,t),\left(s',t'\right)\right) = k\left(s,s'\right)k\left(t,t'\right). \tag{9}$$

Under this assumption, $\mathcal{F}$ and $\mathcal{H}$ become constant matrices $\mathbf{F}$ and $\mathbf{H}$. As a result, a GP kernel function may be directly mapped to $\mathbf{F}$, $\mathbf{H}$ and $\omega(s,t)$,

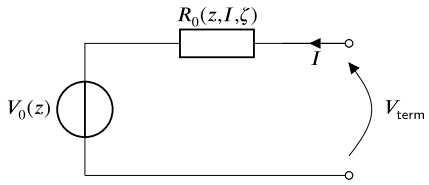


Fig. 1. Equivalent circuit model for Li-ion battery.

defining a corresponding LTI system. The practical implementation of Eq. (7) requires it to have a discretized form,

$$\begin{aligned}\mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}\\ \mathrm{y}_k &= \mathbf{H}_k\mathbf{x}_k + v_k, \quad v_k \sim \mathcal{N}\left(0, \sigma_n^2\right),\end{aligned} \tag{10}$$

where $\mathbf{x}_k = \left[\mathbf{x}_{s_1,k}, \mathbf{x}_{s_2,k}, \ldots, \mathbf{x}_{s_{n_s},k}\right]^\top$ represents a state vector along all n_s space locations at time step t_k, and $\mathbf{A}_k = \exp\left(\Delta t_k \mathbf{I}_{n_s} \otimes \mathbf{F}\right)$ is a matrix exponential on $\mathbf{F}$ and time interval $\Delta t_k = t_{k+1} - t_k$. The Kronecker product here comes from the space discretization. Each space location is propagated in time individually, and correlations in space are captured by the process noise term $\mathbf{w}_{k-1}$ which is a zero-mean multivariate GP over the n_s space locations with covariance function expressed as,

$$\begin{aligned}\mathbf{w}_{k-1} &\sim \mathcal{GP}\left(\mathbf{0}, \mathbf{K}_s \otimes w_{k-1}\right)\\ w_{k-1} &= \int_0^{\Delta t_k} \mathbf{A}_{k-1}\mathbf{L}\mathrm{Q_c}\mathbf{L}^\top\mathbf{A}_{k-1}^\top \mathrm{d}\tau.\end{aligned} \tag{11}$$

Here $\mathbf{K}_s$ is the $n_s \times n_s$ covariance matrix of the spatial kernel $k(s, s')$ at n_s locations, and spectral density matrix $\mathrm{Q_c}$ is decided by the temporal kernel $k(t, t')$ as a mathematical combination of kernel hyperparameters.

Once the GP is formulated as a discrete-time LTI system, both the estimation of posterior predictive distribution in Eq. (5) and the hyperparameter NLML in Eq. (6) can be achieved recursively using a Kalman filter and Rauch-Tung-Striebel (RTS) smoother [30], enabling linear time complexity in forward model runs [2].

2.2. Gaussian processes for equivalent circuit models

As shown in Fig. 1, a simple ECM is used for modelling the battery dynamics in this work. This approach is similar to the method we used in Aitio et al. [2] but simplified. The corresponding mathematical formulation is given as

$$\begin{aligned}V_{\text{term}} &= V_0(z) + R_0(z, I, \zeta)\, I\\ \frac{\mathrm{d}z}{\mathrm{d}t} &= I Q^{-1}(\zeta)\end{aligned} \tag{12}$$

where z is the SOC, I is the applied current, and $V_0(z)$ represents the OCV-SOC curve, which is often measured at beginning of life and assumed to be fixed over life. Two timescales are included in this ECM, one is the operating timescale (t), which is often at the seconds level. Another is the life-long aging timescale (ζ), which can vary from days to months depending on usage. Although this circuit model is simple, for a high quality commercial cell operated at low rates, the internal overpotentials should be small, and therefore this model produces reasonable results. Even in highly dynamic applications, there are usually still some periods of low-rate operation that could be used for health estimation.

The variables I, z, V_{term} are all functions of t, unless written in their discrete-time versions that are demarcated with subscripts (e.g. z_i). For clarity the time-dependence of the continuous-time variables is assumed, rather than written out in every equation. For simplicity, we use a vector to denote operating conditions,

$$\mathbf{s} = [z, I]^\top. \tag{13}$$

The inverse capacity term Q^{-1} is modeled as a GP over ζ, while the resistance term R_0 is modeled as a GP over $\mathbf{s}$ and ζ,

$$\begin{aligned}Q^{-1}(\zeta) &\sim \mathcal{GP}\left(m(\zeta), k\left(\zeta, \zeta'\right)\right)\\ R_0(\mathbf{s}, \zeta) &\sim \mathcal{GP}\left(m(\mathbf{s}, \zeta), k\left((\mathbf{s}, \zeta), (\mathbf{s}', \zeta')\right)\right).\end{aligned} \tag{14}$$

Temperature was tightly controlled and consistent for battery samples used in our work, so had little impact on the results. Temperature effects can however be included as another dimension of the operating-condition dependency of $R_0(s, \zeta)$, where $s = [z, I, T]$ [1]. Given the nonzero nature of impedance and capacity, instead of directly modelling R_0 and Q, we chose to model them each as affine transformations of a zero-mean GP, which is equivalent to setting a nonzero prior mean for impedance and capacity [2]. Specifically, Eqs. (14) are written as,

$$\begin{aligned}Q^{-1}(\zeta) &= q_0(1 + q(\zeta))\\ R_0(\mathbf{s}, \zeta) &= r_0(1 + r(\mathbf{s}, \zeta))\\ q(\zeta) &\sim \mathcal{GP}\left(0, k\left(\zeta, \zeta'\right)\right)\\ r(\mathbf{s}, \zeta) &\sim \mathcal{GP}\left(0, k\left((\mathbf{s}, \zeta), (\mathbf{s}', \zeta')\right)\right)\end{aligned} \tag{15}$$

where constants r_0, q_0 reflect prior assumptions for the values of R_0, Q^{-1}, respectively. In practice, these can be set to the corresponding beginning-of-life (BoL) values. As a Gaussian distribution remains Gaussian under arbitrary affine transformation, R_0 and Q^{-1} are still GPs.

The operating-condition dependency on $\mathbf{s}$ and the life-long aging dependency on ζ are modelled with different kernels, since we expect that degradation is irreversible, whereas a more general smooth function describes operating-condition dependency,

$$k\left((\mathbf{s}, \zeta), (\mathbf{s}', \zeta')\right) = k_{\text{Matern32}}\left(\mathbf{s}, \mathbf{s}'\right) k_{\text{WV}}\left(\zeta, \zeta'\right) \tag{16}$$

where k_{WV} and k_{Matern32} represent a Wiener velocity (WV) kernel and a Matern ($\nu = 3/2$) kernel, respectively, so

$$\begin{aligned}k_{\text{WV}}\left(\zeta, \zeta'\right) &= \sigma_\zeta^2\left(\frac{\min^3\left(\zeta, \zeta'\right)}{3} + \left|\zeta - \zeta'\right| \frac{\min^2\left(\zeta, \zeta'\right)}{2}\right)\\ k_{\text{Matern32}}\left(\mathbf{s}, \mathbf{s}'\right) &= \sigma_s^2(1 + \sqrt{3}d)\exp(-\sqrt{3}d)\\ d^2 &= \left|\mathbf{s} - \mathbf{s}'\right|^\top \begin{bmatrix} l_z^{-2} & 0\\ 0 & l_I^{-2}\end{bmatrix} \left|\mathbf{s} - \mathbf{s}'\right|.\end{aligned} \tag{17}$$

While stationary kernels eventually converge back to the mean when extrapolated, the WV kernel is a non-stationary kernel that is more suitable for extrapolating aging trends into the future.

As shown in Fig. 2, the battery operating dynamics, described by the ECM, and aging dynamics, described by the GPs, are evaluated at two different timescales. While the ECM state (z) changes over seconds during cycling (t), the GPs representing Q^{-1} and R_0 may change more slowly, over weeks or months, as a function of cell-aging timescale ζ. In subsequent sections we will use discrete-time rather than continuous-time dynamics, with the indexes i and j denoting the discrete-time variables associated with t and ζ, respectively, with the faster discrete time step size defined as Δt_i and the slower step size as $\Delta\zeta_j$.

The discrete-time state propagation for the ECM is a simple Coulomb counter that can be expressed as

$$z_i = z_{i-1} + Q^{-1} I_i \Delta t_i. \tag{18}$$

The discrete-time propagation for the inverse capacity and resistance parameters, which are GPs, can be written as

$$\begin{aligned}\mathbf{x}_{Q^{-1},j} &= \exp\left(\Delta\zeta_j \mathbf{F}_{\text{WV}}\right)\mathbf{x}_{Q^{-1},j-1} + w_{j-1}\\ \mathbf{x}_{R_0,j} &= \exp\left(\Delta\zeta_j \mathbf{I}_{n_s} \otimes \mathbf{F}_{\text{WV}}\right)\mathbf{x}_{R_0,j-1} + \mathbf{w}_{j-1},\end{aligned} \tag{19}$$

where, for the WV kernel, $\mathbf{F}_{\text{WV}} = \begin{bmatrix} 0 & 1\\ 0 & 0\end{bmatrix}$ represents the discrete-time interval on the lifelong aging scale. Because of the operating condition

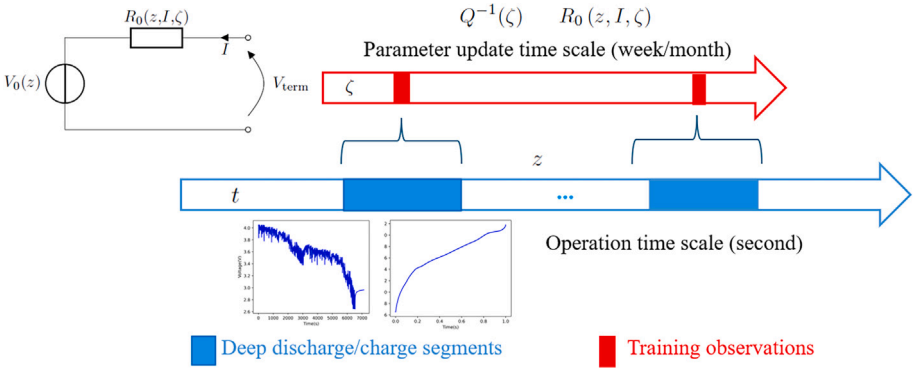


Fig. 2. Modelling the separation of timescales with battery operating dynamics that change quickly (as a function of t) vs. aging dynamics that change slowly (as a function of ζ).

dependency of R_0, Kronecker products are used to describe the corresponding discretisation in $\mathbf{I}_{n_s} \otimes \mathbf{F}_{\text{WV}}$ and $\mathbf{w}_{j-1}$. In this work, $n_s = n_z n_I$. Specifically, the 0 to 1 SOC range is evenly divided into n_z discrete SOC levels, while the applied current range (in absolute scale) is evenly divided into n_I discrete current levels. The process noise terms w_{j-1} and $\mathbf{w}_{j-1}$ can be calculated following Eq. (11), where Q_c for the WV kernel is σ_ζ^2.

2.3. Joint estimation of battery states and GPs

The state covariance initialisation and process noise matrix of the joint system and GP subsystem are summarized as follows:

$$\begin{aligned}\mathbf{P}_{\text{joint},0} &= \begin{bmatrix} \mathrm{P}_{z,0} & 0 \\ 0 & \mathbf{P}_{\text{GP},0} \end{bmatrix}, \mathbf{P}_{\text{GP},0} = \begin{bmatrix} \mathbf{P}_{Q^{-1},0} & 0 \\ 0 & \mathbf{P}_{R_0,0} \end{bmatrix} \\ \mathbf{W}_{\text{joint}} &= \begin{bmatrix} \mathrm{W}_z & 0 \\ 0 & \mathbf{W}_{\text{GP}} \end{bmatrix}, \mathbf{W}_{\text{GP}} = \begin{bmatrix} \mathbf{W}_{Q^{-1}} & 0 \\ 0 & \mathbf{W}_{R_0}, \end{bmatrix}\end{aligned} \tag{20}$$

where

$$\begin{aligned}&\mathbf{P}_{R_0,0} = \mathbf{K}_{\text{Mat}} \otimes \mathbf{P}_{\zeta_0,\text{WV}}, \quad \mathbf{P}_{Q^{-1},0} = \mathbf{P}_{\zeta_0,\text{WV}} \\ &\mathbf{W}_{R_0} = \mathbf{K}_{\text{Mat}} \otimes \mathbf{W}_{\text{WV}}, \quad \mathbf{W}_{Q^{-1}} = \mathbf{W}_{\text{WV}}.\end{aligned} \tag{21}$$

In these equations, ζ_0 represents the initial aging time step and $\mathbf{K}_{\text{Mat}}$ is the covariance matrix of the Matern kernel at n_s discrete operating conditions. For the selected WV kernel, the initial covariance and noise matrix are

$$\mathbf{P}_{\zeta_0,\text{WV}} = \sigma_\zeta^2 \begin{bmatrix} \frac{1}{3}\zeta_0^3 & \frac{1}{2}\zeta_0^2 \\ \frac{1}{2}\zeta_0^2 & \zeta_0 \end{bmatrix}, \mathbf{W}_{\text{WV}} = \sigma_\zeta^2 \begin{bmatrix} \frac{1}{3}\Delta\zeta^3 & \frac{1}{2}\Delta\zeta^2 \\ \frac{1}{2}\Delta\zeta^2 & \Delta\zeta \end{bmatrix}. \tag{22}$$

In literature, several works [13,41] use a dual estimation framework to propagate these two subsystems separately, with the estimated SOC (z) as the 'observation' for parameter updates. By decoupling states from parameters, however, any cross-correlations between the two are lost and this may lead to poor estimation accuracy [26]. For example, the uncertainty information from the SOC estimation, which is important for parameter updates, is discarded. As an alternative, other works [21,45] adopt a joint estimation framework with a single timescale to update both states and parameters. This may suffer from poor numerical conditioning, though, due to the vastly different timescales of the state and parameter dynamics (seconds versus months).

To address these issues, we adopted our previously proposed co-estimation framework [1]. Within charge or discharge segments, the GP states (i.e. circuit model parameters) and the ECM state are propagated jointly. This means the GP dynamics described in Eq. (19) also evolve within each segment over time steps of Δt. However, since $\Delta t \ll \Delta\zeta$, the changes in Q and R_0 over small timesteps will be small since the GP is relatively smooth on the timescale of Δt. Between segments, only the GP states are preserved; the ECM state is reinitialized at the beginning of every new discharge segment based on the measured initial SOC conditions. In this way, the GPs describing capacity and resistance and the ECM describing state of charge dynamics each share the same observations (Eq. (12)), but their propagation is at different timescales.

Within a segment, the overall joint state vector for both ECM and GPs is given by

$$\mathbf{x}_{\text{joint},i} = \begin{bmatrix} z_i \\ \mathbf{x}_{Q^{-1},i} \\ \mathbf{x}_{R_0,i} \end{bmatrix}. \tag{23}$$

Between segments, the GP subsystem is linear and the GP state vector is given by

$$\mathbf{x}_{\text{GP},j} = \begin{bmatrix} \mathbf{x}_{Q^{-1},j} \\ \mathbf{x}_{R_0,j} \end{bmatrix}. \tag{24}$$

Note that within a segment the joint system is nonlinear because Eq. (12) is nonlinear in SOC (z) due to the $V_0(z)$ and $R_0(z)$ functions. For simplicity, we write the nonlinear joint system dynamics as

$$\begin{aligned}\mathbf{x}_{\text{joint},i} &= g\left(\mathbf{x}_{\text{joint},i-1}, I_{i-1}, \mathbf{W}_{\text{joint},i-1}\right) \\ V_i &= h\left(\mathbf{x}_{\text{joint},i}, I_{i-1}, v_i\right).\end{aligned} \tag{25}$$

where v_i represents the measurement noise with variance σ_v^2. An extended Kalman filter is used to solve the system dynamics with local linearisation via Jacobian matrices

$$\mathbf{G}_i = \frac{\mathrm{d}g}{\mathrm{d}\mathbf{x}_{\text{joint},i}}, \quad \mathbf{H}_i = \frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}_{\text{joint},i}}. \tag{26}$$

The full co-estimation framework is given in Algorithm 1. Because R_0 is modelled with a GP, its predictive variance needs to be included when calculating the noise covariance of output (the $I_i^2\mathbf{\Sigma}_{R_0,i}$ term in $\mathbf{S}_i$). Detailed derivations can be found in Appendices A and B. To provide a baseline comparison against our method, we also implemented a random-walk dual-estimation approach for parameter updates, see Appendix D.

The co-estimation in Algorithm 1 can be viewed as a single forward run of the proposed model under a specific GP hyperparameter set and measurement noise variance (i.e., $\sigma_\zeta^2, \sigma_s^2, l_z^{-1}, l_I^{-1}, \sigma_v^2$). To find the optimal values of these, we use the accumulated NLML (Φ) as our cost function. Because of the additive nature of the NLML, Φ can be updated during the state and parameter co-estimation process. An outer optimisation loop using a gradient-based L-BFGS-B algorithm is then used to obtain *maximum a posteriori* (MAP) estimates of the hyperparameters for GPs, running the forward model many times, with each forward run giving one evaluation of Φ. The hyperparameter optimisations were solved on a virtualized Linux platform with two physical Intel Xeon Silver 5216 CPUs at 2.1 GHz. Finding the optimal hyperparameters took approximately 60 and 90 minutes for cells A and B, respectively.

Algorithm 1 Co-estimation of battery state and GPs using extended Kalman filter. The NLML (Φ) is calculated recursively on every terminal voltage measurement.

Initialisation of co-estimation at $\zeta = \zeta_0$
$\mathbf{x}^+_{\text{GP}} = \mathbf{x}_{\text{GP},0}, \mathbf{P}^+_{\text{GP}} = \mathbf{P}_{\text{GP},0}$
$\Phi = 0$
for $\text{seg}_j \in$ segments **do**
  GP propagation
  $\mathbf{x}^-_{\text{GP},j} = \exp(\mathbf{F}\Delta\zeta_j)\mathbf{x}^+_{\text{GP},j-1}$
  $\mathbf{P}^-_{\text{GP},j} = \exp(\mathbf{F}\Delta\zeta_j)\mathbf{P}^+_{\text{GP}}\exp(\mathbf{F}\Delta\zeta_j)^{\text{T}} + \mathbf{W}_{\text{GP}}(\Delta\zeta_j)$
  Initialisation at start of j**th segment**
  $\mathbf{x}^+_{\text{joint}} = \mathbf{x}_{\text{joint},0} = [z_0, \mathbf{x}^-_{\text{GP},j}]^\top$
  $\mathbf{P}^+_{\text{joint}} = \mathbf{P}_{\text{joint},0} = \begin{bmatrix} \mathbf{P}_{z,0} & 0 \\ 0 & \mathbf{P}^-_{\text{GP},j} \end{bmatrix}$
  for $V_i \in \text{seg}_j$ **do**
    Joint states propagation
    $\mathbf{x}^-_{\text{joint},i} = g\left(\mathbf{x}^+_{\text{joint},i-1}, I_{i-1}, \mathbf{W}_{\text{joint},i-1}\right)$
    $\mathbf{P}^-_{\text{joint},i} = \mathbf{G}_{i-1}\mathbf{P}^+_{\text{joint},i-1}\mathbf{G}^{\text{T}}_{i-1} + \mathbf{W}_{\text{joint},i-1}$
    Joint states update
    $\mathbf{e}_i = V_i - h\left(\mathbf{x}^-_{\text{joint},i}, I_{i-1}, v_i\right)$
    $\mathbf{S}_i = \mathbf{H}_i\mathbf{P}^-_{\text{joint},i}\mathbf{H}^\top_i + I_i^2\mathbf{\Sigma}_{R_0,i} + \sigma_v^2$
    $\mathbf{L}_i = \mathbf{P}^-_{\text{joint},i}\mathbf{H}^\top_i\mathbf{S}^{-1}_i$
    $\mathbf{x}^+_{\text{joint},i} = \mathbf{x}^-_{\text{joint},i} + \mathbf{L}_i\mathbf{e}_i$
    $\mathbf{P}^+_{\text{joint},i} = (\mathbf{I} - \mathbf{L}_i\mathbf{H}_i)^\top\mathbf{P}^-_{\text{joint},i}(\mathbf{I} - \mathbf{L}_i\mathbf{H}_i) + \mathbf{L}_i\mathbf{e}_i\mathbf{L}^\top_i$
    $\Phi = \Phi + \frac{1}{2}\mathbf{e}^\top_i\mathbf{S}^{-1}_i\mathbf{e}_i + \frac{1}{2}\log\left|2\pi\mathbf{S}_i\right|$
  end for
end for

3. Data preprocessing and model setup

3.1. Data description

3.1.1. Cell A

Cell A is a nickel-manganese-cobalt vs. graphite Li-ion battery from an open test dataset [37]. It is a commercial 502030 cell with nominal capacity 250 mAh and operating voltage from 3.0 to 4.2 V. The true initial capacity was measured as around 280 mAh and the beginning-of-life DC resistance around 130 mΩ. This battery was aged with repeated charge/discharge cycles and checked through a weekly reference performance test (RPT). Cycling aging was conducted with a constant 1.825 C charging C-rate and a constant 0.5 C discharging C-rate. The depth of discharge (DoD) was 100 % and the environmental temperature was kept at 30 °C. The RPT consisted of low rate (0.2 C) constant-current constant-voltage (CC-CV) charging and constant-current discharging, yielding ground truth discharge capacity measurements and an up-to-date pseudo OCV-SOC curve. The cell number is G37C2 in the original dataset.

3.1.2. Cell B

Cell B is a nickel-cobalt-aluminium oxide vs. graphite-silicon battery from another open dataset [15], with cell ID 015. It is a commercial 18,650 cell with nominal capacity 3400 mAh and BoL DC resistance around 25 mΩ. This battery was charged with a CC-CV regime and discharged using a scaled drive cycle profile from 0 to 9.7 A (absolute value). Cycle aging was conducted at 25 °C between 10 % and 90 % SOC. Regular RPTs were taken every 30 cycles, at 0.1 C.

3.2. Extracting discharge segments from cycling data

The sampling frequency for cycling was 0.2 Hz and 1 Hz for cells A and B, respectively. The corresponding full cycling datasets (over all of life) are 1,840,632 and 41,346,756 rows of current and voltage measurements, respectively. Using the entire cycle life data to optimize GP hyperparameters requires substantial computational effort. However, it is reasonable to assume that the aging between two adjacent cycles is small, and therefore we only selected a small number of cycling segments from which to train the model, with relatively large between-segment distances $\Delta\zeta_j$.

In this work, the process of down-selecting discharge segments was as follows. Discharge segments were selected evenly over the life of each battery. In the case of cell A, 15 discharge segments in total were chosen, and in the case of cell B, 18 segments were chosen. This reduced the data volume to 21,881 and 120,801 rows, respectively. In practice, the time interval $\Delta\zeta_j$ between segments can be in the unit of days, weeks or months depending on usage and test conditions. The method would also work using units such as equivalent full cycles or cumulative charge throughput. One reason for using days is that segment extraction can be done in a convenient manner. The first 10 segments were used to train the model (i.e. to optimize the kernel hyperparameters for the GPs). We chose 10 segments for model training since it was computationally efficient but sufficient to demonstrate the efficacy of our method.

The RPT data at BoL was used to calibrate the OCV-SOC relationship, while other RPT data were not used in the proposed model. The extracted discharge segments for cells A and B are shown in Fig. 3. The aging effect can be clearly seen from the shrinkage of discharge curves on the time axis.

Considering the computational cost and data range, we set the operating condition discretization for cell A to $n_z = 25$ and $n_I = 1$, and for cell B to $n_z = 25$ and $n_I = 5$.

4. Results and discussion

4.1. Health estimation and prediction

Figs. 4 and 5 show projections of the GP estimates of discharge capacity Q and impedance R_0 over battery lifetime for cells A and B,

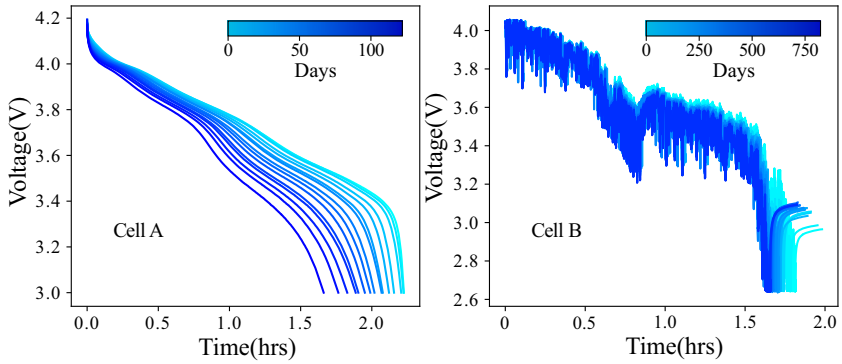


Fig. 3. All extracted discharge segments, cells A and B.

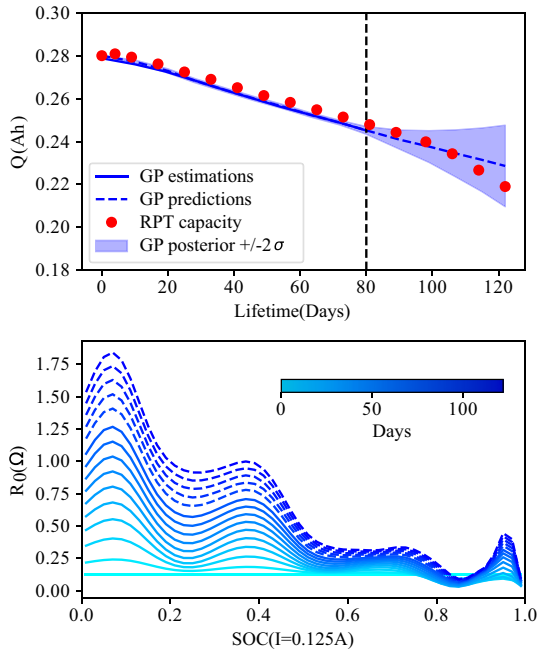


Fig. 4. Cell A projections of GP posteriors for Q and R_0. Ground truth RPTs are only used for capacity validation (red dots). RMSE and MAPE for capacity estimates are 0.0026 Ah and 0.86 %, and for predictions are 0.0038 Ah and 1.53 %.

respectively. Comparisons against the baseline random-walk method (Appendix D) show that GP estimates are much more accurate and smoother than the random-walk approach. The train-test split is shown by the vertical black line. Here, 'training' refers to the data used to fit GP hyper-parameters, whereas 'testing' refers to short-term future predictions that are obtained by running the parameter dynamics forward in time without seeing new data. Such short-term future SOH predictions are useful for many applications, e.g., battery early maintenance. For cell A, the discharge current is constant, thus R_0 is only plotted against SOC. For cell B, with dynamic discharging profile, R_0 is plotted against both SOC and applied current.

It is clear that the GP captures the aging trend for capacity fade well based only on cycling data without RPTs. As the non-stationary WV kernel is used for modelling the time dependency of $Q^{-1}(\zeta)$, the GP extrapolation maintains the last known aging trend from the training stage. This linear extrapolation fits well for cell B, while it appears slightly worse for cell A where the ground truth aging speed accelerates in later life. The uncertainty for estimates is shown by the shaded area that grows larger when extrapolating further into the future.

For the estimation results of R_0, a clear upwards shift can be found for cell A during aging, indicating a systematic increase of battery resistance over all SOC levels. For cell B, R_0 changes in a more complex way, with both increasing and decreasing trends observed against different SOC levels. The scale of R_0 dependency on applied current is much smaller than the SOC dependency. Thus, we mainly discuss the SOC dependency in this work, and more details about the applied current dependency can be found in Appendix C.

All estimations are shown in solid lines, extrapolations are shown as dotted lines. The blue gradient colors denote the corresponding aging times. The extrapolation for R_0 also maintains the increasing trend of training stage. To keep the figure simple, the uncertainties for R_0 estimates are included in Fig. 6.

In order to ensure smoothness in the estimated R_0 functions that are displayed, the discretization of SOC was further increased to 100 evenly spaced locations. The trained model was then rerun in forward mode, with the previously learnt GP hyperparameters, but using the extended SOC discretization.

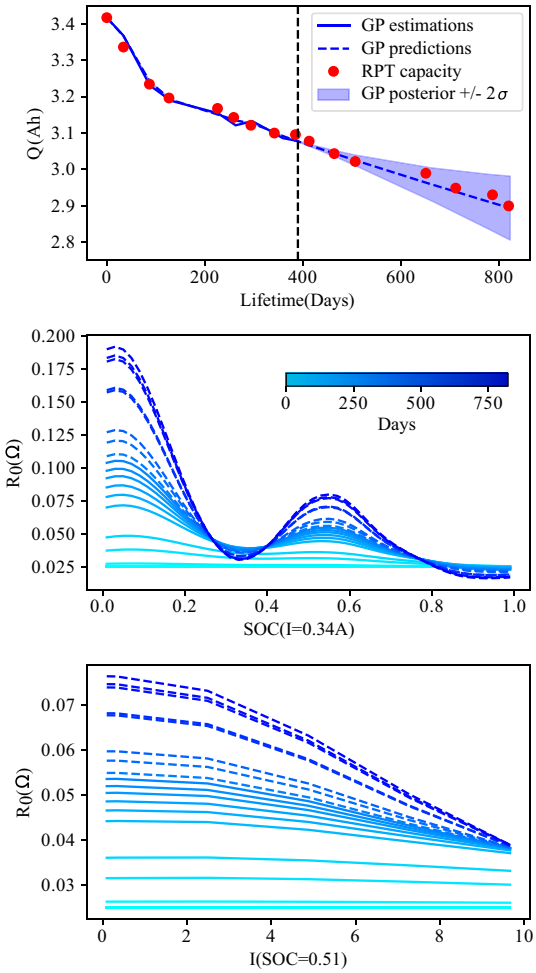


Fig. 5. Cell B projections of GP posteriors for Q and R_0. Ground truth RPTs are only used as capacity validations (red dots). RMSE and MAPE for capacity estimates are 0.0122 Ah and 0.31 %, and for predictions are 0.009 Ah and 0.27 %.

4.2. Dependence of R_0 on SOC

Considering the results of the previous section, there are two abnormal features to note in the evolution of $R_0(z,\zeta)$. First, the increasing scale of R_0 is much larger than expected. Given that the BoL DC resistance calculated from a pulse test is around 0.13 Ω and 0.025 Ω for cells A and B respectively, the increase in R_0 by nearly five times at low SOC level as the cell ages seems unrealistic. Second, there is a wave-like curve shape and a corresponding uneven increase of resistance against SOC. This SOC dependency is also observed by others [31,39]. Fig. 6 shows impedance increases at different SOC levels. In the training set, the increase at high SOC (> 0.9) is almost negligible for both cells. However, at low SOC (< 0.1), the increase of R_0 is nearly 0.88 Ω and 0.07 Ω for cell A and B respectively.

We propose that these unexpected behaviours are not caused by 'true' changes in the underlying resistance, but instead hypothesize that a significant reason for increases in the estimated $R_0(z,\zeta)$ is shape changes in the battery OCV-SOC curve as it ages. Specifically, the battery dynamics in Eq. (12) assume that the OCV-SOC functional relationship is kept the same across the entire life, and only scaled by Q. A more realistic formulation could be expressed as

$$V_{\text{term}} = V_0(z,\zeta) + \hat{R}_0(z,\zeta)I, \tag{27}$$

where the OCV-SOC relationship $V_0(z,\zeta)$ is itself a function of both SOC (z) and aging lifetime (ζ). In this case, $\hat{R}_0$ is the 'pure' impedance term

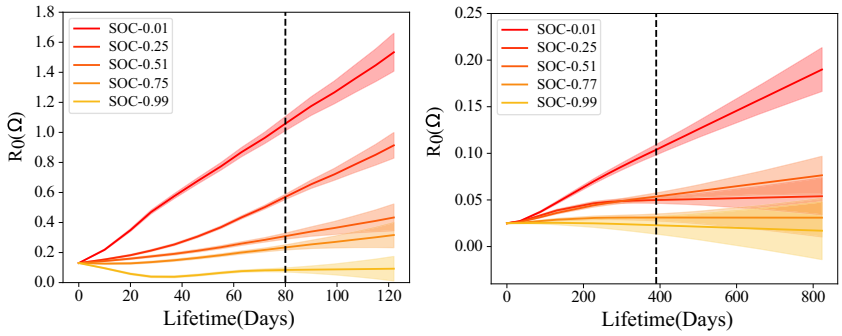


Fig. 6. GP estimation of resistance increase at different SOC levels.

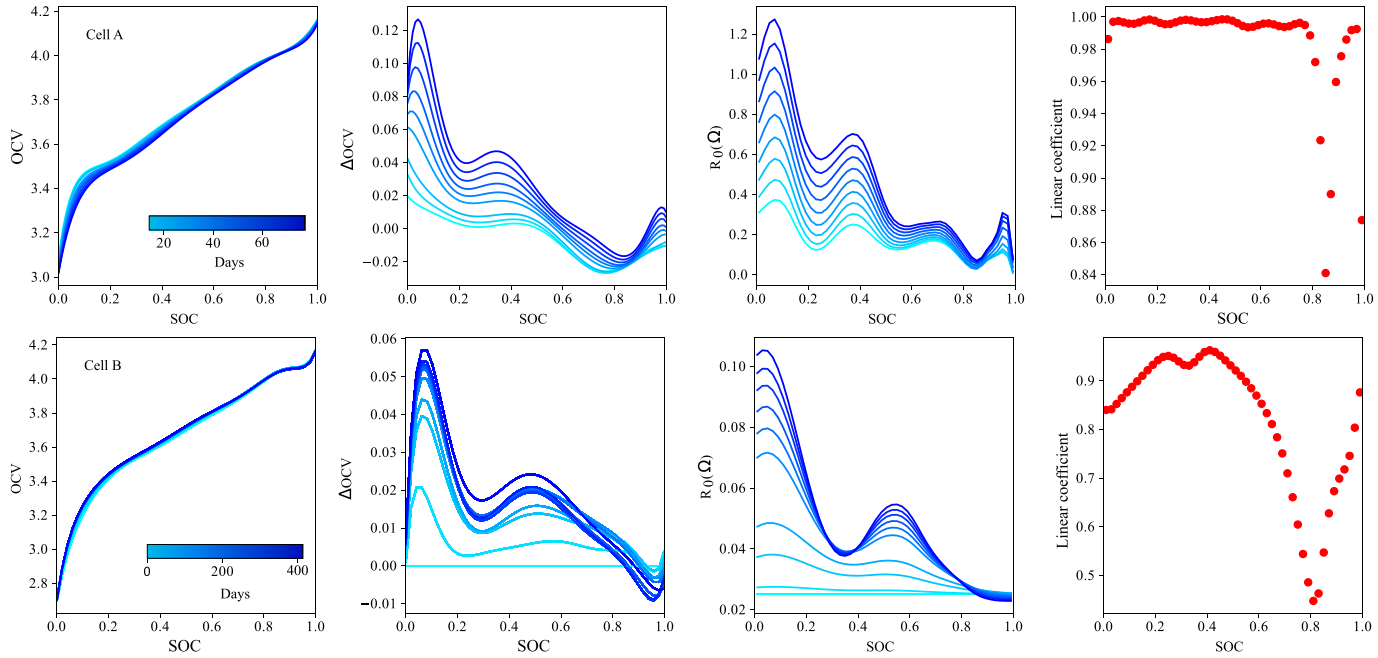


Fig. 7. Comparison between ground-truth OCV-SOC and estimated R_0 function for cell A and B.

(unaffected by OCV). By further separating the OCV term into a fixed part $\bar{V}_0(z,\zeta_0)$ (representing BoL OCV-SOC), and an age-sensitive variable part $\Delta V_0(z,\zeta)$, one can rewrite this equation as

$$V_{\text{term}} = \bar{V}_0\left(z,\zeta_0\right) + \Delta V_0(z,\zeta) + \hat{R}_0(z,\zeta)I$$

$$V_{\text{term}} = \bar{V}_0(z) + I\underbrace{\left[\frac{\Delta V_0(z,\zeta)}{I} + \hat{R}_0(z,\zeta)\right]}_{R_0(z,\zeta)}. \tag{28}$$

It is now clear that, rather than obtaining the pure resistance $\hat{R}_0$, the learnt function $R_0(z,\zeta)$ is influenced both by changes in the resistance and changes in the OCV-SOC functional form as the battery ages.

Thanks to the regular RPT data, we are able to validate this hypothesis using ground-truth OCV-SOC changes. Fig. 7 compares the learnt $R_0(z,\zeta)$ functions and ground-truth OCV-SOC curves. The applied current was 0.125 A and 0.34 A for cells A and B, respectively. Here, ΔOCV represents the difference between 'up-to-date' OCVs (i.e., incorporating changes as the cell ages) and BoL OCVs across all SOC levels. A strong consistency can be observed in the changing trends of R_0 and ΔOCV. This can be quantified using a Pearson linear correlation coefficient—specifically, for each discrete SOC value, the linear coefficient between R_0 and ΔOCV as a function of age was calculated. For cell A, a linear coefficient of more than 0.8 is observed over all SOC levels, decreasing slightly around 0.8 SOC. Similarly, for cell B, a coefficient of more than 0.7 is seen for most SOC levels, decreasing around 0.8 SOC.

One remaining issue concerns the (un)identifiability of $\Delta V_0(z,\zeta)$ and $\hat{R}_0(z,\zeta)$. Given the additive nature of these, it is impossible to distinguish them individually from terminal voltage measurements. However, the differences in their influences on terminal voltage can be compared. If one assumes, for example, a 50 % increase in resistance compared to BoL DC resistance, this would result in overpotential changes of 8 mV and 4 mV for cells A and B at 0.5 C and 0.1 C, respectively. However, the average changes in the OCV-SOC function over life are 30 mV and 20 mV, respectively—on the order of ten times higher.

4.3. Differential voltage analysis

Capacity decrease and impedance increase reflect battery SOH changes at the full cell level, but are insufficient to elucidate different degradation modes. One commonly used quantitative method for

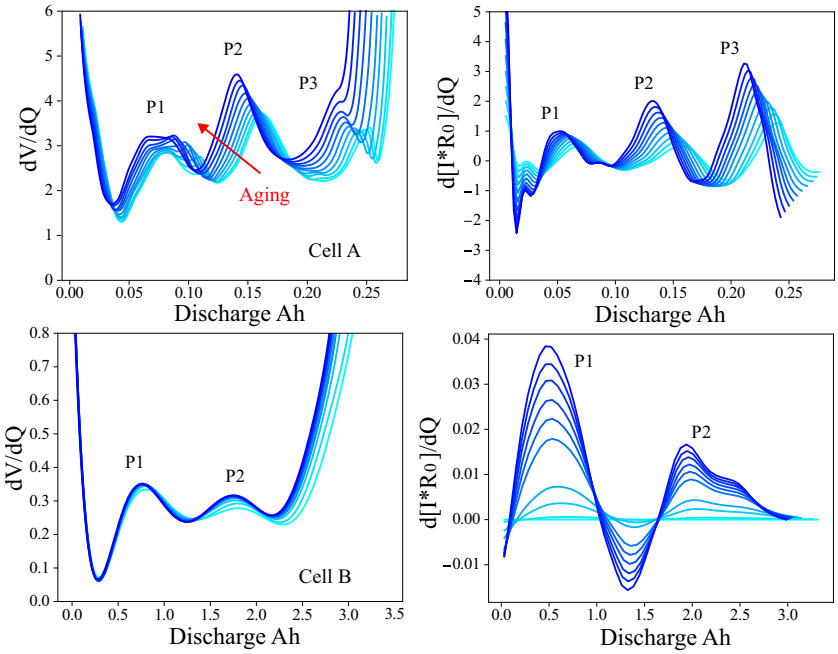


Fig. 8. Comparison between RPT dV/dQ and estimated $d[IR_0]/dQ$.

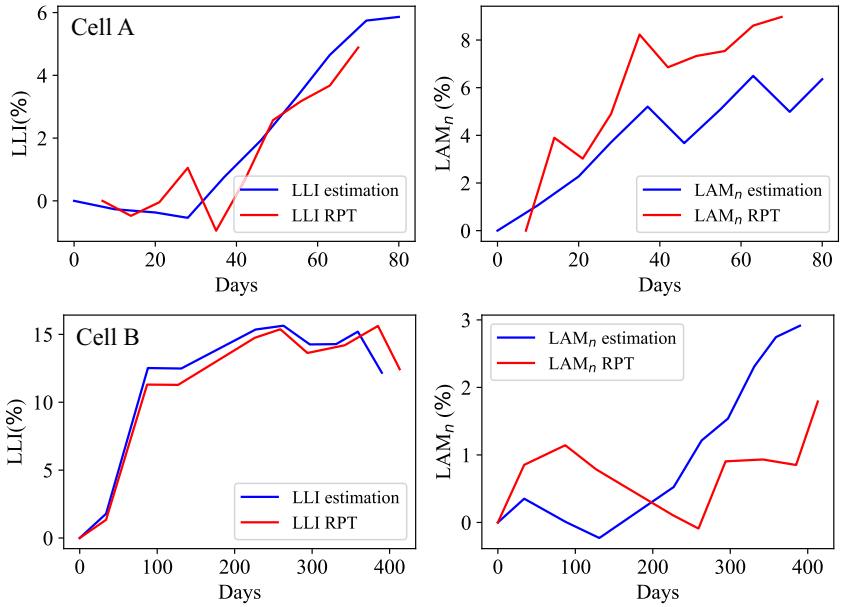


Fig. 9. Comparison of estimated LLI and LAM_n for dV/dQ from RPTs and $d[IR_0]/dQ$ estimations.

the latter is differential voltage analysis (dV/dQ vs. discharge Ah). The dV/dQ curve is obtained by differentiating discharge voltage with respect to cumulative discharge Ampere-hours. The peaks on the dV/dQ curve can be assigned to electrode material phase transitions. The peak shifts and the changing between-peak distances are used to indicate different degradation modes, i.e. lost lithium inventory LLI, and loss of active material in the negative and positive electrodes, LAM_n and LAM_p [6]. One major limitation of the dV/dQ method is the requirement for a low C-rate (<C/20) deep (dis)charge, which is generally unavailable except in a carefully controlled lab environment. In our case, because of experiment limits, the RPT C-rates for cells A and B were higher, at C/5 and C/10, respectively.

Given the observed close similarity between $\Delta\text{OCV}(z,\zeta)$ and $R_0(z,\zeta)$ as discussed in the previous section, a natural extension is to compare the dV/dQ information from the RPTs with the learnt $d[IR_0]/dQ$ vs. discharge curve. As shown in Fig. 8, if one multiplies the SOC with the learnt capacity $Q(\zeta)$, one can derive a corresponding discharge Ampere-hour estimate and thus observe similar peak shifts to differential voltage functions.

Due to the lack of half cell test data, we are not able to attribute each dV/dQ peak to a corresponding electrode, However, we can speculate (on the basis of literature [16,47] and our understanding for these cell chemistries) about what the peak shifts mean, and thus estimate LLI, LAM_n for cells A and B respectively (Fig. 9). (Positive electrode loss of

active material is less clear so we omit estimation of LAM_p.) For cell A with a CC discharging profile and NMC cathode, a 6–8 % increase is observed for both LLI and LAM_n. This is consistent with the relatively uniform capacity drop across life in Fig. 4. For cell B, with a dynamic driving discharging profile and NCA cathode, there is a sharp increase of LLI within the first 100 days, which is likely related to diffusion-limited solid electrolyte interphase (SEI) layer formation [3,35]. This is also consistent with the early capacity drop in Fig. 5 which then levels off.

5. Conclusions

We have proposed an aging-aware battery ECM incorporating GPs to describe the dependency of circuit-model parameters on SOC and lifetime. A state-space formulation was used to transform the GPs into time-stepping models enabling solution in linear time complexity with a Kalman filter and RTS smoother. This also enables co-estimation of both parameters and states in a unified framework. Compared to parameter random walks or deterministic models, using GPs to model parameter evolution has several advantages:

- The state-space formulation of a GP maintains gradient information from the states whilst random walk or empirical models do not.
- Incorporating the SOC and aging-time dependency of parameters via GP kernels makes ECM parameter estimates more stable over battery life, i.e. gives smoother interpolation and more reasonable extrapolation compared to random walk or deterministic models.
- A GP prior helps to mitigate numerical ill-conditioning as it has a regularisation effect when parameters are not easily identifiable (e.g. due to measurement noise or extrapolation to unseen conditions).

We found that estimates of capacity Q exhibited good consistency compared with independent ground-truth measurements from RPTs, whereas estimates of the resistance function R_0 showed a close relationship with shape changes in the OCV-SOC curve as a battery ages. A major part of the learnt $R_0(z,\zeta)$ function comes from OCV changes, $\Delta V_0(z,\zeta)$. High linear coefficients (> 0.8) were observed between $R_0(z,\zeta)$ and $\Delta V_0(z,\zeta)$ over a wide range of SOCs. By using the estimates of both $Q(\zeta)$ and $R_0(z,\zeta)$, we further calculated $d[IR_0]/dQ$ as an approximation to a differential voltage dV/dQ function. The estimated peak shifts and degradation modes (LLI, LAM_n) in $d[IR_0]/dQ$ were found to show similar behaviours compared with ground-truth from dV/dQ.

This work highlights both a challenge, to ensure that shape changes to the OCV curve do not distort battery SOH estimates, and an opportunity—to estimate degradation modes (lost lithium inventory, lost electrode material) directly from operational data. Further tests are required on other cell form factors, duty cycles (e.g. shallow cycling), chemistries, temperature variations, and cell sizes to explore the generalisation of these results.

CRediT authorship contribution statement

Zihao Zhou: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Antti Aitio:** Writing – review & editing, Methodology. **David Howey:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Zihao Zhou reports financial support was provided by Rimac Automobili. Zihao Zhou reports financial support was provided by China Scholarship Council. David Howey reports a relationship with Brill Power that includes: equity or stocks. David Howey and Antti Aitio have patent #GB Application no. 2105995.1 pending via Oxford University Innovation Ltd. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Department of Engineering Science, University of Oxford, the China Scholarship Council, and Rimac Automobili.

Appendix

A. Gaussian process with input uncertainty

In the joint system dynamics (Eq. 25), R_0 needs to be evaluated at every time step. The SOC at step i is itself a Gaussian distribution expressed as

$$z_i \sim \mathcal{N}(z_i^-, \mathbf{P}_{z_i}^-). \tag{29}$$

This means that R_0 is a GP with input uncertainty from z_i. To incorporate this, the predictive distribution in Eq. (5) needs to be marginalised (averaged) over all z_i,

$$p\left(R_0\right) = \int p\left(R_0 \mid z_i\right) p\left(z_i\right) \mathrm{d}z_i. \tag{30}$$

This integral is usually intractable but can be approximated using Taylor expansions [11]. A first-order expansion can be written as

$$\begin{aligned} m_{R_0} &= K_{Z,z_i}^\top \left[K_Z + \sigma_{\mathrm{GP}}^2 I\right]^{-1} z_i \\ \mathbf{\Sigma}_{R_0} &= K_{z_i,z_i} - K_{Z,z_i}^\top \left[K_Z + \sigma_{\mathrm{GP}}^2 I\right]^{-1} K_{Z,z_i} + \left.\frac{\partial R_0}{\partial z}\right|_{z_i}^\top \mathbf{P}_{z_i}^- \left.\frac{\partial R_0}{\partial z}\right|_{z_i}. \end{aligned} \tag{31}$$

where K_Z and K_{Z,z_i} are a covariance matrix and vector following the same definition as Eq. (5).

B. Predictive variances for terminal voltage

In the extended Kalman filter used for the joint GP-ECM system, an extra variance term $I_i^2 \mathbf{\Sigma}_{R_0,i}$ is included in the calculation of the innovation sequence $\mathbf{S}_i$. This corresponds to the contribution of the R_0 uncertainty when calculating the predictive distribution of terminal voltage V_i:

$$\begin{aligned} V_i &= V_0\left(z_i\right) + I_i R_{0,i} + v_i, \quad v_i \sim N\left(0, \sigma_v^2\right) \\ \mathrm{E}\left[V_i\right] &= V_0\left(z_i\right) + I_i m_{R_{0i}} \\ \mathrm{cov}\left(V_i\right) &= \mathrm{E}\left[\left(V_i - \mathrm{E}\left[V_i\right]\right)\left(V_i - \mathrm{E}\left[V_i\right]\right)^\top\right] \\ \mathrm{cov}\left(V_i\right) &= I_i^2 \Sigma_{R_{0i}} + \sigma_v^2. \end{aligned} \tag{32}$$

C. Impedance dependency on applied current

Fig. 10 gives a full set of resistance estimates at several SOCs and currents, for cell A.

D. The random-walk baseline model

To benchmark the proposed method, a random-walk model was used to track aging with the commonly used dual-estimation approach [25,41,43]. The same ECM in Fig. 11 was used, without the dependency of R_0 on the operating condition.

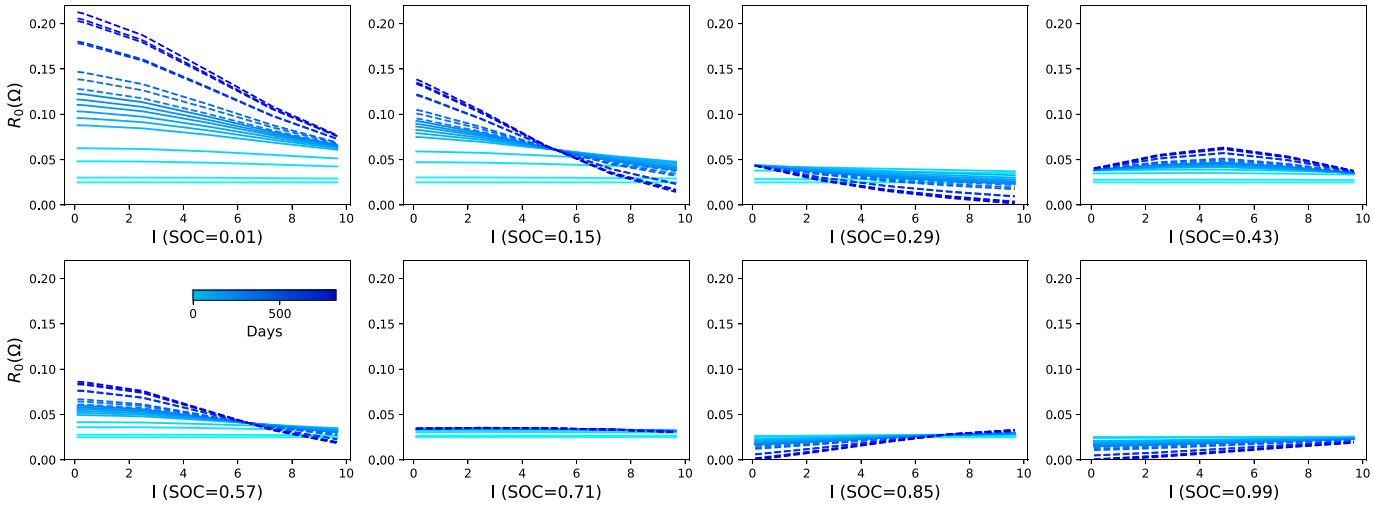


Fig. 10. Dependence of resistance on applied current at several SOCs, for cell A.

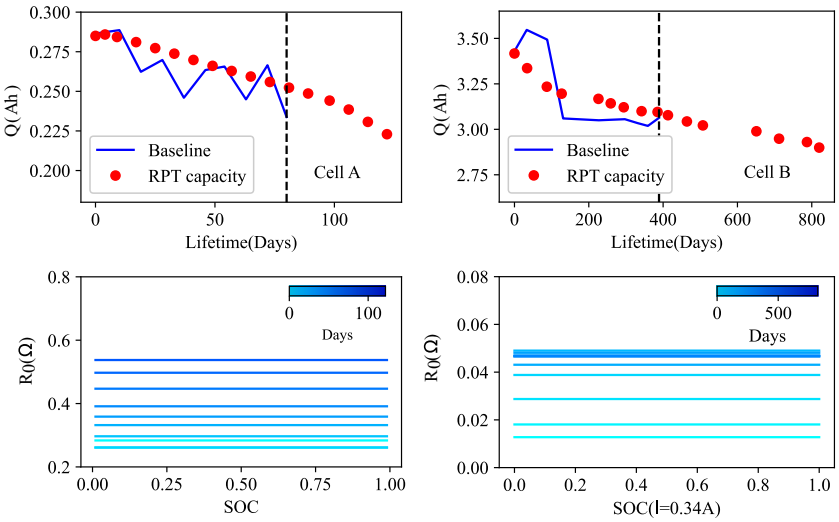


Fig. 11. Results for the random-walk baseline model: cell A estimates have 0.129 Ah RMSE and 3.27 % MAPE, Cell B estimates have 0.019 Ah RMSE and 6.74 % MAPE.

Data availability

Datasets used in this paper are already open source.

References

[1] Aitio A, Howey DA. Predicting battery end of life from solar off-grid system field data using machine learning. Joule 2021;5(12):3204–20.

[2] Aitio A, Jöst D, Sauer DU, Howey D. Learning battery model parameter dynamics from data with recursive Gaussian process regression. J Dyn Syst Meas Control 2023:1–13.

[3] An SJ, Li J, Daniel C, Mohanty D, Nagpure S, Wood DL. The state of understanding of the lithium-ion-battery graphite solid electrolyte interphase (SEI) and its relationship to formation cycling. Carbon 2016;105:52–76.

[4] Berecibar M, Gandiaga I, Villarreal I, Omar N, Van Mierlo J, Van den Bossche P. Critical review of state of health estimation methods of li-ion batteries for real applications. Renew Sustain Energy Rev 2016;56:572–87.

[5] Cohen L. The generalization of the Wiener–Khinchin theorem. In: Proceedings of the 1998 IEEE international conference on acoustics, speech and signal processing, ICASSP'98 (Cat. No. 98CH36181). IEEE; 1998. p. 1577–80.

[6] Dubarry M, Anseán D. Best practices for incremental capacity analysis. Front Energy Res 2022;10:1023555.

[7] Dubarry M, Truchot C, Liaw BY. Synthesize battery degradation modes via a diagnostic and prognostic model. J Power Sour 2012;219:204–16.

[8] Feng X, Weng C, He X, Han X, Lu L, Ren D, et al. Online state-of-health estimation for li-ion battery using partial charging segment based on support vector machine. IEEE Trans Veh Technol 2019;68(9):8583–92.

[9] Fu Y, Xu J, Shi M, Mei X. A fast impedance calculation-based battery state-of-health estimation method. IEEE Trans Ind Electron 2021;69(7):7019–28.

[10] Gao Y, Liu K, Zhu C, Zhang X, Zhang D. Co-estimation of state-of-charge and state-of-health for lithium-ion batteries using an enhanced electrochemical model. IEEE Trans Ind Electron 2021;69(3):2684–96.

[11] Girard A, Rasmussen C, Candela JQ, Murray-Smith R. Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. Adv Neural Inf Process Syst 2002;15.

[12] Greenbank S, Howey D. Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life. IEEE Trans Ind Inf 2021;18(5):2965–73.

[13] Hu C, Youn BD, Chung J. A multiscale framework with extended Kalman filter for lithium-ion battery soc and capacity estimation. Appl Energy 2012;92:694–704.

[14] Hua Y, Liu X, Zhou S, Huang Y, Ling H, Yang S. Toward sustainable reuse of retired lithium-ion batteries from electric vehicles. Resour Conserv Recycl 2021;168:105249.

[15] Jöst D, Blömeke A, Sauer DU, Ringbeck F. Timeseries data of a drive cycle aging test of 28 high energy NCA/C+ Si round cells of type 18650. Technical Report. Institut für Stromrichtertechnik und Elektrische Antriebe; 2021.

[16] Keil P, Schuster SF, Wilhelm J, Travi J, Hauser A, Karl RC, et al. Calendar aging of lithium-ion batteries. J Electrochem Soc 2016;163(9):A1872.

[17] Li T, Zhou Z, Thelen A, Howey DA, Hu C. Predicting battery lifetime under varying usage conditions from early aging data. Cell Rep Phys Sci 2024;5(4):101891.
[18] Li W, Sengupta N, Dechent P, Howey D, Annaswamy A, Sauer DU. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. J Power Sour 2021;482:228863.
[19] Li Y, Liu K, Foley AM, Zülke A, Berecibar M, Nanini-Maury E, et al. Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review. Renew Sustain Energy Rev 2019;113:109254.
[20] Li Y, Zou C, Berecibar M, Nanini-Maury E, Chan JCW, Van den Bossche P, et al. Random forest regression for online capacity estimation of lithium-ion batteries. Appl Energy 2018;232:197–210.
[21] Liu B, Tang X, Gao F. Joint estimation of battery state-of-charge and state-of-health based on a simplified pseudo-two-dimensional model. Electrochim Acta 2020;344:136098.
[22] Locorotondo E, Cultrera V, Pugi L, Berzi L, Pierini M, Lutzemberger G. Development of a battery real-time state of health diagnosis based on fast impedance measurements. J Energy Storage 2021;38:102566.
[23] Lu D, Trimboli MS, Fan G, Zhang R, Plett GL. Implementation of a physics-based model for half-cell open-circuit potential and full-cell open-circuit voltage estimates: part I. processing half-cell data. J Electrochem Soc 2021;168:070532.
[24] Miguel E, Plett GL, Trimboli MS, Oca L, Iraola U, Bekaert E. Review of computational parameter estimation methods for electrochemical models. J Energy Storage 2021;44:103388.
[25] Plett GL. Extended Kalman filtering for battery management systems of LIPB-based HEV battery packs: part 3. state and parameter estimation. J Power Sour 2004;134(2):277–92.
[26] Plett GL. Battery management systems, volume II: equivalent-circuit methods. Artech House; 2015.
[27] Pozzato G, Allam A, Pulvirenti L, Negoita GA, Paxton WA, Onori S. Analysis and key findings from real-world electric vehicle field data. Joule 2023;7(9):2035–53.
[28] Richardson RR, Birkl CR, Osborne MA, Howey DA. Gaussian process regression for in situ capacity estimation of lithium-ion batteries. IEEE Trans Ind Inf 2018;15(1):127–38.
[29] Sarkka S, Solin A, Hartikainen J. Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: a look at Gaussian process regression through Kalman filtering. IEEE Signal Process Mag 2013;30(4):51–61.
[30] Särkkä S, Svensson L. Bayesian filtering and smoothing. vol. 17. Cambridge university press; 2023.
[31] Schönleber M, Uhlmann C, Braun P, Weber A, Ivers-Tiffée E. A consistent derivation of the impedance of a lithium-ion battery electrode and its dependency on the state-of-charge. Electrochim Acta 2017;243:250–59.
[32] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. Nat Energy 2019;4(5):383–91.
[33] Shahjalal M, Roy PK, Shams T, Fly A, Chowdhury JI, Ahmed MR, et al. A review on second-life of li-ion batteries: prospects, challenges, and issues. Energy 2022;241:122881.
[34] Solin A, et al. Stochastic differential equation methods for spatio-temporal Gaussian process regression. Ph.D. thesis. Aalto University; 2016.
[35] Spotte-Smith EWC, Kam RL, Barter D, Xie X, Hou T, Dwaraknath S, et al. Toward a mechanistic model of solid–electrolyte interphase formation and evolution in lithium-ion batteries. ACS Energy Lett 2022;7(4):1446–53.
[36] Sulzer V, Mohtat P, Aitio A, Lee S, Yeh YT, Steinbacher F, et al. The challenge and opportunity of battery lifetime prediction from field data. Joule 2021;5(8):1934–55.
[37] Thelen A, Li T, Liu J, Tischer C, Hu C. ISU-ILCC battery aging dataset; 2023. https://doi.org/10.25380/iastate.22582234.v2.
[38] Tsiropoulos I, Tarvydas D, Lebedeva N, et al. Li-ion batteries for mobility and stationary storage applications. Publications Office of the European Union; 2018.
[39] Waag W, Käbitz S, Sauer DU. Experimental investigation of the lithium-ion battery impedance characteristic at various conditions and aging states and its influence on the application. Appl Energy 2013;102:885–97.
[40] Wang Y, Tian J, Sun Z, Wang L, Xu R, Li M, et al. A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. Renew Sustain Energy Rev 2020;131:110015.
[41] Wassiliadis N, Adermann J, Frericks A, Pak M, Reiter C, Lohmann B, et al. Revisiting the dual extended Kalman filter for battery state-of-charge and state-of-health estimation: a use-case life cycle analysis. J Energy Storage 2018;19:73–87.
[42] Wei J, Dong G, Chen Z. Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression. IEEE Trans Ind Electron 2017a;65(7):5634–43.
[43] Wei Z, Zhao J, Ji D, Tseng KJ. A multi-timescale estimator for battery state of charge and capacity dual estimation based on an online identified model. Appl Energy 2017b;204:1264–74.
[44] Williams CK, Rasmussen CE. Gaussian processes for machine learning. MIT press Cambridge, MA; 2006.
[45] Yu Q, Xiong R, Lin C, Shen W, Deng J. Lithium-ion battery parameters and state-of-charge joint estimation based on H-infinity and unscented Kalman filters. IEEE Trans Veh Technol 2017;66(10):8693–701.
[46] Zhou Z, Ran A, Chen S, Zhang X, Wei G, Li B, et al. A fast screening framework for second-life batteries based on an improved bisecting k-means algorithm combined with fast pulse test. J Energy Storage 2020;31:101739.
[47] Zhu J, Darma MSD, Knapp M, Sørensen DR, Heere M, Fang Q, et al. Investigation of lithium-ion battery degradation mechanisms by combining differential voltage analysis and alternating current impedance. J Power Sour 2020;448:227575.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Learning Li-ion battery health and degradation modes from data with aging-aware circuit models
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	Zhou Zihao, Antti Aitio, and David Howey. "Learning Li-ion battery health and degradation modes from data with aging-aware circuit models." Applied Energy 397 (2025): 126375.

Student Confirmation

Student Name:	Zihao Zhou		
Contribution to the Paper	First author I developed the core aging-aware equivalent circuit model to address the problem of battery SOH estimation. I implemented the corresponding code, conducted all experiments for the paper, performed additional analyses, and authored the whole manuscript.		
Signature	Zihao Zhou	Date	11 July 2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: David A. Howey, Professor of Engineering Science			
Supervisor comments Zihao led this work and the description given above is accurate.			
Signature	David Howey	Date	15 Aug 2025

This completed form should be included in the thesis, at the end of the relevant chapter.

4 | Estimating SOH from field data

4.1 Preface

This chapter is based on the following publication:

> [4] **Zhou Zihao** and David Howey. "Battery capacity and degradation mode diagnosis for solar off-grid systems in Africa." 2025. (To be submitted)

The previous chapter introduced the aging-aware ECM designed for battery lifelong SOH estimation. It was shown that the learned resistance function closely reflects the shape change of the OCV-SOC curve, which enables a reconstruction of the differential voltage (DV) curve and facilitates the estimation of degradation modes such as LLI and LAM. However, model performance was only validated on lab aging data, which consists of high-resolution measurements collected under a uniform lifelong cycling profile. The ultimate goal of SOH estimation should be to achieve reliable performance on batteries deployed in field.

Several challenges arise when working with field data. First, complex and variable usage profiles, combined with sensor noise, further compromise the identifiability of model parameters and the observability of battery states from standard terminal voltage and current measurements. For example, due to the flat OCV characteristic of LFP batteries, it can be difficult to estimate capacity from shallow and dynamic discharging profiles. Second, the lack of ground-truth capacity measurements makes it hard to validate estimation performance. Capacity is typically measured through a low C-rate constant-current discharge from 100% to 0% SOC, a process that is too time-consuming to be performed regularly in field applications. Third, the sampling rate in field data is often much lower than in lab settings. Due to storage and transmission constraints, field signals are often recorded at a rate of 1/60 to 1/600 Hz (i.e., one data point every 1 to 10 minutes). This low frequency further challenges Coulomb-counting methods, especially when paired with highly dynamic cycling profiles.

In Chapter 4, we apply the aging-aware ECM to study 114 LFP battery packs from solar off-grid systems operating in Africa. So-called 'pseudo-labels' for capacity are computed for 56 of these packs to validate the model's capacity estimates. Furthermore, population-analysis across all 114 packs shows that time spent below 6.7 V is strongly correlated with aging rate. We observe that lower value for time spent below 6.7 V over the operational life of a pack is caused by limited customer discharging and long rest periods at high voltage. Also, DV curves are reconstructed from the estimated resistance function (as in the preceding chapter), enabling inference of degradation modes (LLI and $\mathrm{LAM_{ne}}$). Once again, we find that resting at high voltage accelerates degradation through LLI and $\mathrm{LAM_{ne}}$. Based on these findings, we conclude that avoiding or reducing prolonged rest periods at high voltage is an effective strategy to extend battery life in off-grid solar-battery systems.

4.2 Manuscript

Battery capacity and degradation mode diagnosis for solar off-grid systems in Africa

Zihao Zhou[1] and David Howey[1*]

[1*]Department of Engineering Science, University of Oxford, Woodstock Road, Oxford, OX1 3PJ, UK.

*Corresponding author(s). E-mail(s): david.howey@eng.ox.ac.uk;
Contributing authors: zihao.zhou@eng.ox.ac.uk;

Abstract

Aging studies of lithium-ion batteries using field data face two major difficulties: the complex degradation behaviour of battery and the limitations of field measurements. With years of extensive field usage, large-scale battery aging data becomes available. An important question is how much we can know about battery health from field data? In this work, we address this by analysing data from 114 commercial lithium ion phosphate (LFP) battery packs used in sub-Saharan Africa, each running for 100 to 1200 days. A hybrid method, combining Gaussian processes with equivalent circuit models within an extended Kalman filter based co-estimation framework, was used to track battery health directly from daily operational data. To validate the method we also calculated health labels using a separate, independent approach for comparison. Comparing the two methods, we achieve 1.4% mean absolute percentage error and 0.2 Ah mean square error (on 9.9 Ah nominal capacity) on average for capacity estimation. Population-level analysis shows time spent below 6.7 V is the dominant factor determining pack aging rate. Additionally, degradation mode estimation was achieved by reconstructing differential voltage curves from field data. It is found that long rest times at high voltages could causing severe aging. This work highlights the opportunity for multidimensional health diagnosis and life prediction of LFP cells from field data.

Keywords: Battery, field data, degradation, health, estimation, differential voltage

1 Introduction

Due to energy poverty and lack of grid infrastructure in Africa, 600 million people do not have access to electricity [1]. In recent years, decentralized solar-battery systems (especially lithium-ion) have begun to be widely used in many countries to provide local families with affordable energy. While the demand for energy services in Africa is set to grow rapidly, whether solar-battery systems are a viable long-term economical solution is still questionable. One of the major challenges is the uncertainty of battery aging [2]. Accurate state-of-health (SOH) diagnosis and life prediction in real-world operating conditions are essential for safety, warranties, and timely maintenance [3]. Also, SOH estimation from field data helps to understand the influence of varying customer usage on battery aging, improving future battery design and selection [4]. Unlike high resolution multimodal data from lab testing, field-collected signals often only contain terminal voltage, temperature and applied current at a relatively low sampling rate [5]. Further challenges arise because of the lack of ground truth health labels, which

are available through low C-rate lab tests, or a on-site reference test requiring a certain system downtime [6]. Also, difficulties such as data absence and low-quality sensors may make the problem even more difficult.

Many studies have been undertaken on the topic of battery health estimation [7]. Generally, one may categorize them into model-based [8–11] and data-driven approaches [12–16]. For model-based SOH estimators, equivalent circuit models (ECMs) and electrochemical models (EMs) are typically used. State-of-charge (SOC) is one of the model states, while SOH metrics are regarded as model parameters. Terminal voltage is often the only measurement or model prediction. Battery capacity, as one of the major SOH metrics, is indirectly connected to terminal voltage via SOC.

Two estimation paradigms are widely used to achieve co-estimation of model states and parameters. One is joint estimation [9, 10], which combines states and parameters into an augmented state vector. A nonlinear Kalman filter is then employed to estimate both states and parameters at the same time scale. Although joint estimation is straightforward to implement, it suffers from poor numerical conditioning due to the vastly different time scales of the state and parameter dynamics [17]. Another approach is dual estimation [8, 11], which separately estimates states and parameters, and different time scales can be applied for states and parameters. Dual estimation may be more numerically stable by enforcing time scale separation. Any cross-correlations between states and parameters are lost by decoupling their dynamics [17]. For both joint and dual estimation, the parameter dynamic is often described by a random walk process. There is no built-in guarantee that the parameter estimation will converge to anything with physical meaning, especially with a poor initial guess [11]. The limitations of field data (low resolution, data gaps, sensor noise) will further threaten the SOH estimation convergence of model-based methods.

As an alternative, data-driven methods regard SOH estimation as a regression task, without requiring an underlying physically-motivated model to describe input-output behaviour. These methods aim to find the mapping directly from extracted aging-related features, as inputs, to SOH labels, as outputs. Various kind of features are reported to be closely related to aging, such as time spent in certain operating conditions [14] and capacity-voltage curve changes [13]. While these features works well on lab datasets, for field data with uncontrollable cycling, these features may not always be available. A wide range of algorithms can be adopted for regression, such as support vector machines [18], Gaussian process (GP) regression [12] and various neural network approaches [15, 19]. As most of these methods are supervised learning techniques, a certain amount of feature-label pairs are needed to train the model, and the learnt mapping may change across different datasets.

While most studies discuss SOH estimation from lab aging data, there are several that now include real field aging data [20–22]. Large numbers of solar-connected lead-acid batteries are analyzed in Aitio et al. [20], where resistance is regarded as the key SOH metric. A 73% accuracy for end-of-life (EOL) prediction was achieved at 8 weeks in advance. Unlike in lead-acid systems, for Li-ion batteries, capacity is often used to denote health. Because of a lack of ground-truth capacity measurements from battery systems in the field, extensive experiment efforts were made in Figgener et al. [22], with regular check-up tests operated for 21 private battery management systems for up to 8 years. Coulomb counting was used to estimate capacity by down-selecting cycling segments between end of charge (EOC) and end of discharge (EOD).

This method is, however, fragile for LFP systems because accurate SOC is challenging to estimate due to the flat OCV curve, and long resting times to achieve equilibrium may not be available. Although capacity and resistance are very common SOH metrics, they do not provide information on underlying aging mechanisms. Degradation modes go some way to filling the gap; lost lithium-ion inventory (LLI) and lost active material for the positive ($\mathrm{LAM_{pe}}$) and negative electrode ($\mathrm{LAM_{ne}}$) are widely accepted degradation mode metrics. These may be estimated from differential voltage (DV) [23] or incremental capacity (IC) analysis [24], using OCV measurements.

Experimental works [25, 26] show that different usage profiles cause different evolution in degradation modes. An ideal health estimation scenario would be to estimate both capacity

drop and the contribution of degradation modes from field usage data. However, a constant current discharge at low C-rate (normally < 0.1C) is often required to obtain useable DV and IC curves, and this is usually not available in dynamic field cycling.

To conclude, the challenges for SOH estimation of battery systems in field can be summarized into three topics: First, the indirect connection between SOH and terminal voltage, as well as the limitations of field-collected signals, make parameter estimation with model-based methods difficult. Second, varying operating conditions and lack of ground truth labels (e.g., independent capacity measurements) make it difficult to build pure data-driven models. Third, capacity itself is insufficient to describe the link between field cycling and corresponding aging effects. The estimation of degradation modes is also needed to provide insight into underlying causes of degradation.

To tackle these challenges, this paper analyzes 114 commercial LFP packs used in sub-Saharan Africa, each running 100 to 1200 days. A hybrid method combining an ECM and GP was employed to achieve multidimensional SOH estimation from field data without requiring labels. Specifically, a GP is used to model the aging effects on ECM parameters (capacity and resistance). The resistance was estimated as a function of both aging time and operating conditions. So-called 'pseudo-labels' of capacity for 56 packs were estimated to provide a separate comparison and validate the approach. Using this method, comparing the GP-ECM method against the direct Coulomb counting method, on average 2% MAPE and 0.2 Ah RMSE were achieved for capacity estimates, and 13% MAPE and 0.72 Ah RMSE were achieved for predictions. Population analysis on all 114 packs shows that the time spent below 6.7 V over a system lifetime has a close relationship with capacity drop per day. A low number of days spent below 6.7 V is likely to be caused by long rest periods at high voltage, due to lack of customer use, for these systems. Moreover, DV curves were constructed from the estimated resistance functions, enabling estimation of degradation modes (LLI and $\mathrm{LAM_{ne}}$). Compared to normal customer use, accelerated aging (increased LLI and $\mathrm{LAM_{ne}}$) was found when packs were rested at high voltage. Thus, we conclude that decreasing or avoiding long rest periods at high voltage is a good strategy to prolong battery life in this application of off-grid solar-batter systems. Along with this paper, we publish the dataset consisting of operational data from the 114 LFP packs covering over 2500 months of field cycling, which to our knowledge will be the largest open field dataset for LFP batteries.

2 Results and discussion

2.1 Dataset

The dataset used in this work was provided by Bboxx, a company that manufactures and operates solar-battery energy systems. The product considered here is a small LFP-based system with nominal voltage 6.4 V and nominal capacity 9.9 Ah, consisting of 3 cells in parallel and 2 in series, and a warranty of 5 years (details in Supplementary Information). The nominal capacity was calibrated by the manufacturer at 0.2 C and 25 °C. The measured quantities available from these field packs are pack-level terminal voltage, applied current, and temperature. Due to memory limits, data are only collected with a sample rate of 2 minutes. The operation time for the 56 packs covers from 100 to 1200 days.

Fig. 1(a) shows an overview of the geographical distribution of the packs considered in this work. All are located near the equator, giving plenty of solar radiation and a 20-40 °C yearly temperature variation, Fig. 1(b,f). The overall voltage usage histogram is summarized in Fig. 1(c). After a battery pack is charged to its upper cut-off voltage (7.2 V) from daytime solar power, it is disconnected from the solar panel and then has a long resting period, resulting in the first peak around 6.7 V. The second and third voltage peaks in Fig. 1(c) represent rest voltage values after customer discharging—they are less clear because of shorter rest periods after discharging. For applied current, Fig. 1(d) reveals that the solar charging C-rate ranges from 0.05 to 0.3 C and the discharging C-rate is lower (<0.1 C), with clear peaks. These represent the customer usage, which for these systems is quite uniform and primarily dictated by the number of appliances connected (e.g., lights, phone charging, radio and TV). The operational data for these packs can be downloaded from a public repository.

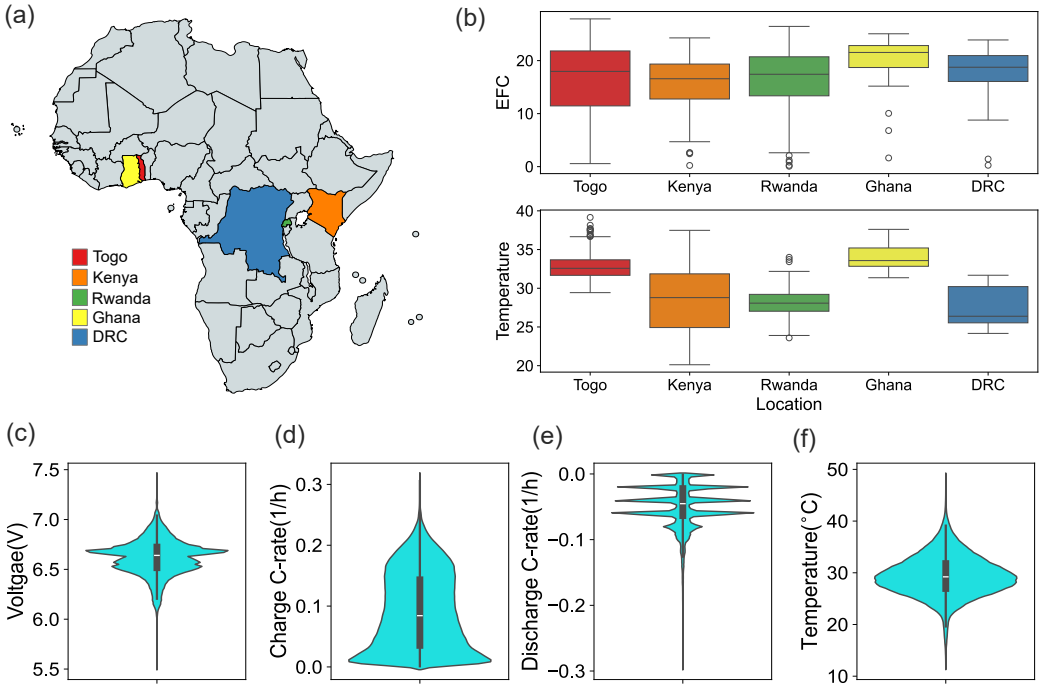


Fig. 1 Overview of the battery dataset used in this study. (a) Geographical locations of LFP packs. Pack number summary: 62 in Rwanda, 17 in Togo, 12 in Kenya, 12 in DRC, and 11 in Ghana. (b) Monthly average equivalent full cycles (EFCs) and temperatures, per country. (c)–(f) Butterfly plots of measured (c) terminal voltages, (d) charging and (e) discharging currents (normalised by nominal capacity), (f) temperatures, for all systems.

2.2 Baseline capacity estimates

Unlike lab battery health data, field measurements do not usually include regular check-up tests. This presents a challenge for validating health estimation methods. To address this issue and provide comparison data for the estimation algorithms we discuss later, we used a simple coulomb-counting method to calculate 'pseudo-labels' for discharge capacity. This provides baseline capacity estimates as long as sufficiently deep, low-rate discharge segments can be found within the data [22].

Two typical 24-hour usage profiles are shown in Fig. 2. Solar-battery systems often have clearly separated charging and discharging segments, in contrast to applications such as electric vehicles. There is an excess of solar during the day, allowing both battery charging and supply to loads. Each battery is typically fully charged during the daytime, and then discharged to customer loads from late afternoon until the next morning. By focusing only on discharge segments when estimating capacity, the impacts of voltage hysteresis in LFP cells [27] can be avoided. To identify suitable discharge periods where simple coulomb-counting can be used, we searched for end-of-charge (EOC) and end-of-discharge (EOD) events, defined as points after a lengthy rest period (more than 2 hours) following a charge or discharge, respectively. Fig. 2(a) shows a clearly defined discharge segment with both EOC and EOD demarcations, whereas Fig. 2(b) shows a segment with EOC but no clear EOD event.

Where usable discharge segments are available, such as in Fig. 2(a), we can regard the beginning and end voltages as OCV measurements. Using these measurements, and a separate beginning-of-life lab measurement of the OCV-SOC relationship (see Supplementary Information), one can estimate SOC at beginning and end of the segment, then calculate capacity

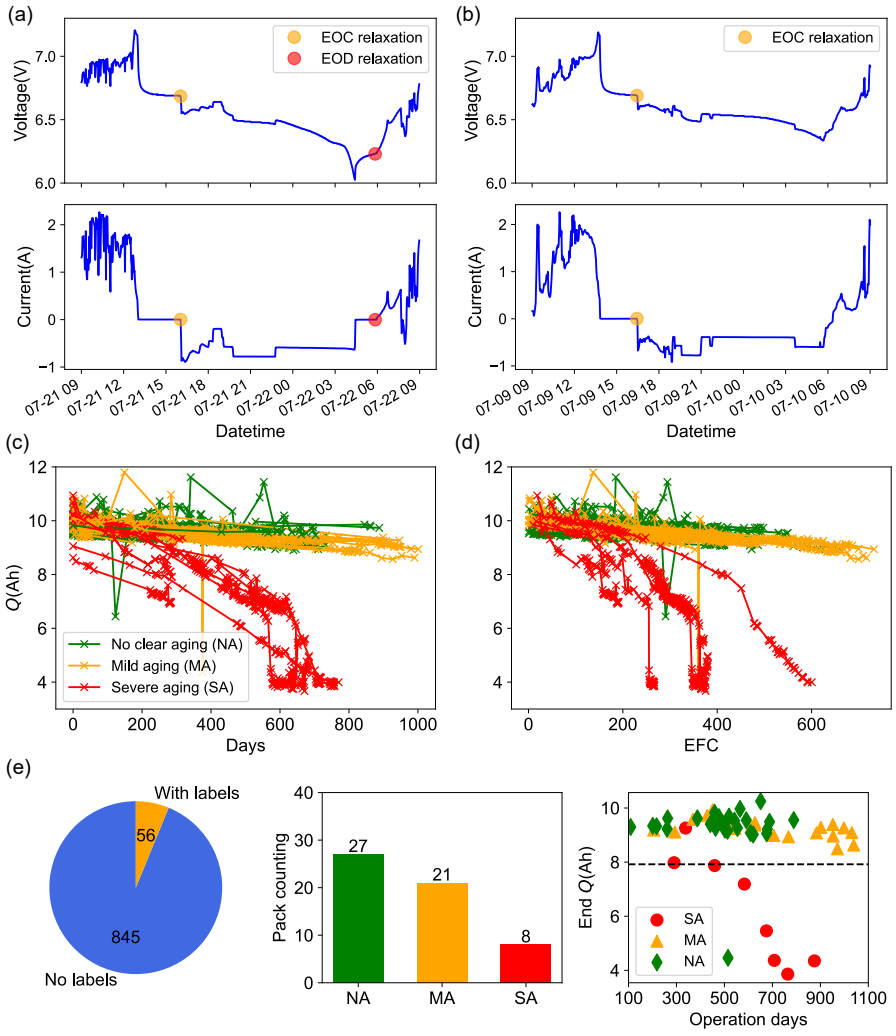


Fig. 2 Baseline capacity estimates using Coulomb counting between relaxation points. (a) Example discharge profile with two adjacent EOC and EOD relaxation periods, suitable for 'pseudo-label' capacity estimation. (b) Example profile without EOD relaxation that cannot be used for capacity estimation. (c) Extracted capacity pseudo-labels vs. operational days for all packs. (d) Extracted capacity pseudo-labels vs. equivalent full cycles for all packs. (e) Summary of numbers of packs with and without pseudo-labels, as well as numbers in each aging group. (f) Capacity values at the end of each operational period.

'pseudo-label' estimates $\hat{Q}$ by rearranging the usual equations for Coulomb counting,

$$\begin{aligned} \mathrm{SOC}_1 - \mathrm{SOC}_0 &= \frac{\int_0^T I(t)dt}{Q}, \\ \hat{Q} &= \frac{\sum I_k \cdot \Delta t_k}{\mathrm{SOC}_1 - \mathrm{SOC}_0}, \end{aligned} \tag{1}$$

where I_k are current measurements at time step size Δt_k, and $\mathrm{SOC}_{\{0,1\}}$ indicate known SOCs (from OCVs) at beginning and end of discharge, respectively.

Have established a simple method to estimate capacity, we can now examine the lifetimes of different products by looking at capacity fade trajectories over time. We categorized systems roughly into three groups according to the capacity drop between the first and the last identified pseudo-labels: (i) no clear aging ('NA'), with less than 0.5 Ah capacity drop; (ii) mild aging ('MA'), with capacity drop 0.5 Ah–1.5 Ah; (iii) severe aging ('SA'), with capacity drop greater than 1.5 Ah. Figs. 2(c,d) show the calculated pseudo-labels for all three groups plotted against operational days and equivalent full cycles (EFC), respectively. Aging trends are

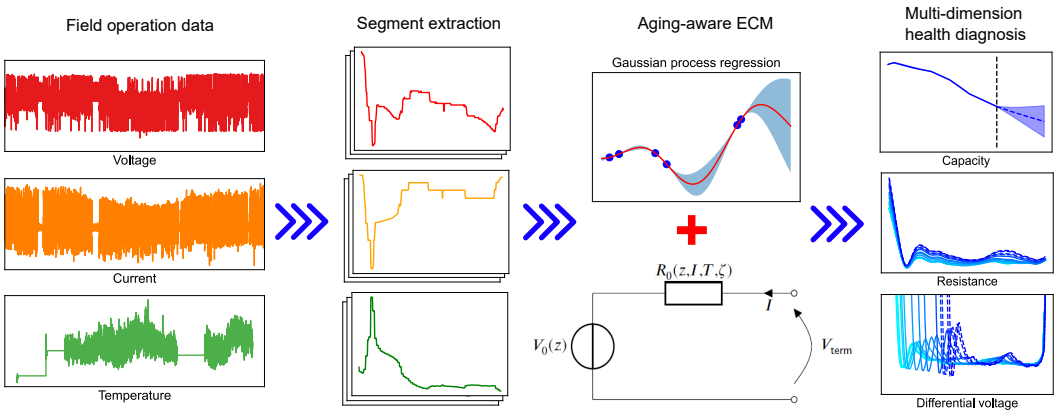


Fig. 3 Overall workflow for health estimation. L–R: Field operational data from LFP battery packs measured; 8–20 discharging segments extracted across life; equivalent circuit model with functional parameters mapped by GPs fitted, resulting in capacity, resistance and differential voltage analysis.

relatively easy to identify for the severe and moderate aging groups. Because of the flatness of OCV curve within the middle SOC range, however, this simple method of capacity estimation is sensitive to voltage noise especially under shallow discharge cycles, and thus considerable noise can be observed in Figs. 2(c,d).

A major limitation of this naive estimation method is the absence of EOD relaxation conditions. As shown in Fig. 2(b), usage profiles may only have EOC relaxation points without corresponding EOD points. As the discharging period usually continues until the following morning, there is often no rest period between discharging and the subsequent next charging period. As shown in Fig. 2(e), amongst all 903 packs considered, only 56 packs had more than 10 pseudo-labels throughout their operational time. Specifically, there were found to be 27 packs with no aging, 21 with moderate aging, and 8 with severe aging. At the data collection date (October 1, 2024), all packs with no or moderate aging still maintained more than 80% nominal capacity, a point that is often considered to be 'end-of-life' (EOL), although this is an arbitrary convention. Most of the severely aged packs, in contrast, had capacity well below 80% and some had dropped to as low as 4 Ah by this point.

2.3 Multi-dimensional health diagnostics approach

To overcome the limitations of the simple approach described in the preceding section, we developed a probabilistic machine-learning based approach that will now be outlined. Shown in Fig. 3, the overall modelling and inference process includes four steps. First, recorded voltage, current, and temperature data at a sampling rate of 2 mins are obtained from a central database. Discharge segments are selected (see details in Supplementary Information) for model fitting. Overall, 8–20 segments were extracted for each pack depending on the operational time for each pack. Then, an aging-aware ECM is used to estimate health by fitting both capacity (Q) and resistance (R_0), the first as a function of time, and the second as a function of time and operational conditions (SOC, current, temperature). Differential voltage curves are estimated indirectly using the learned Q and R_0 (as per [28]), allowing estimation of degradation modes (LLI and $\mathrm{LAM_{ne}}$). Finally, the customer usage profile is linked to the capacity drop as well as corresponding degradation modes to build an population usage-driven lifetime model. Further details on the implementation can be found in the Methods section.

2.4 Health diagnostics and prediction from field data

We now present and discuss several health diagnostics results from LFP field data. Fig. 4(a,b) show estimated capacity fade trajectories for two example packs taken from the severe aging (SA) and moderate aging (MA) groups, respectively. The trajectories include the mean and uncertainty from the GP posterior, and are also compared to the 'pseudo-label' capacity

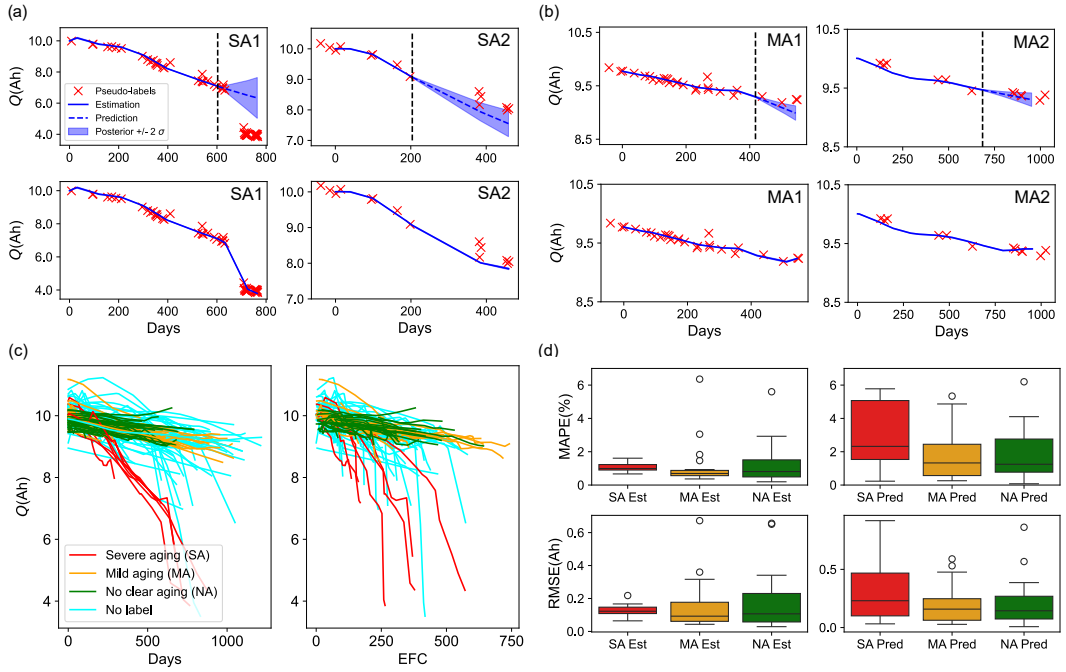


Fig. 4 Results of discharge capacity estimation and prediction. (a) Two example packs from SA group. Capacity estimates under 80%/100% number of segments for training. Predictions into near future is included to show the extrapolation ability of proposed model. (b) Two example packs from MA group. (c) Estimated capacity curves for packs from all three groups. (d) Statistic evaluation of model performance.

estimates described in Section 2.2. To demonstrate our method's ability to make short-term capacity forecasts, the top row of plots shows fits from data up to the black dashed vertical lines, but the trajectories to the right of the vertical lines are created by running the GP forward into the future without seeing new data; in these regions the trajectory follows the most recently learned linear trajectory in the data.

For both moderate and severely aged battery packs, good estimation accuracy can be observed—the estimated capacity fade coincides well with pseudo-label baseline estimates. In terms of short-term predictions, the property of the nonstationary kernel used [29] just tracks the most recent path, and this may or may not fit subsequent baseline estimates well. When there is no sudden change in aging rate, the short-term forecasts can give good results compared to baseline estimates, such as in the SA2, MA1, and MA2 examples. However, an example where there is a significant deviation is in SA1, Fig. 4(a), which has a large capacity drop after 600 days—this leads to large prediction errors. However, with more up-to-date segments of newer data included, the overall aging trend is captured again, as shown by the second row of plots. Capacity estimation results for all 114 packs, fitting all data (i.e., no short-term predictions) are given in Fig. 4(c). There are 56 packs (belonging to the SA, MA, NA groups) with pseudo-labels, and another 58 packs without labels are also included.

To quantify the accuracy of the ECM-GP estimation approach compared with the baseline estimates, the mean absolute percentage error (MAPE) and the root mean square error (RMSE) are used, defined as

$$\begin{aligned} \text{MAPE} &= 100 \cdot \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \end{aligned} \tag{2}$$

where y_t represents true value of capacity (here assumed to be the baseline pseudo-label) and $\hat{y}_t$ is the model estimation. Fig. 4(d) summarizes the MAPE and RMSE results for all packs in the SA, MA and NA groups when using 80% of segments for fitting (i.e. results in the second row of Fig. 4(a)(b) are not summarized here). Less than 2% MAPE and less than 0.3 Ah RMSE are achieved for most packs, with the worst performance up to 6.4% MAPE and 0.67 Ah RMSE. The moderately aged packs have the best group median results of 0.7% MAPE and 0.09 Ah. When making short-term predictions compared against unseen data, less than 5% MAPE and less than 0.5 Ah RMSE are achieved for most packs. The overall predictive performance is worse than the estimation performance, which is unsurprising given that the trajectory just follows the most recently learned linear path into the future, and cannot predict sudden accelerated capacity fade. The method is fundamentally a data-driven tracking and smoothing approach, not a mechanistic model of aging (however, we will address this later). Nonetheless, these results demonstrate that the estimates are accurate and useful for short-term prediction of capacity fade. For better visualization, three SA packs with very large prediction errors were excluded; details can be found in the Appendix E. Additionally, a random walk-based dual estimation model is included as a benchmark comparison in Appendix D; this shows that our method outperforms the random-walk model, achieving more accurate and smoother capacity estimates for SA1, SA2, MA1 and MA2.

2.5 Resistance function estimation and rationalization

In addition to capacity estimation, the ECM-GP hybrid method also estimates battery internal resistance. The estimates are a multi-dimensional function of time as well as operating conditions (temperature, SOC, current), and the mappings are learnt from the data. Fig. 5(b) shows the estimated internal resistance as a function of SOC for the four packs (SA1, SA2, MA1, MA2) shown before. The influence of temperature and current variation is small compared to the influence of SOC. The temperature dependence of R_0 can be found in Appendix F. For the dependence on applied current, Fig. 1(e) shows that discharge C-rate is generally small (<0.2 C), which makes its influence on resistance negligible. Aging causes a general increase in resistance R_0 across all SOCs. We also note that the shape of the R_0-SOC curve shows a rapid increase at low SOC, as well as smaller fluctuations typically around 25% and 70% SOC.

This learnt functional form of the resistance vs. SOC is consistent with expected underlying Li-ion battery electrochemical behaviour. First, like most ECMs used in literature [30], a fixed OCV-SOC relationship from beginning of life is assumed, and this is used to estimate subsequent SOC values from OCV estimates. However, in reality, aging changes the OCV-SOC function in more complex ways than just scaling a fixed function with diminishing cell capacity. Also, OCV-SOC changes may arise from other factors, for example cell-to-cell variability. Thus, a more realistic expression of the learnt $\hat{R}_0$ function includes changes to the OCV-SOC curve, and can be written

$$\begin{aligned} V &= V_0(\zeta) + IR_0(C,\zeta), \quad C = (z, I, T) \\ V &= \left[\, \overline{V}_0 + \Delta V_0(\zeta)\right] + IR_0(C,\zeta) \\ V &= \overline{V}_0 + I \cdot \underbrace{\left[\frac{\Delta V_0(\zeta)}{I} + R_0\right]}_{\hat{R}_0(C,\zeta)}, \end{aligned} \tag{3}$$

where $\Delta V_0(\zeta)$ is the change in the OCV-SOC relationship at some later point in time, and $\overline{V}_0$ is the fixed beginning of life relationship.

To further understand the behaviour of the resistance, we compare the output equations of the single particle model (SPM) and the simple ECM used here, it is apparent that the learnt resistance $\hat{R}_0$ includes contributions from three factors: OCV changes, charge transfer

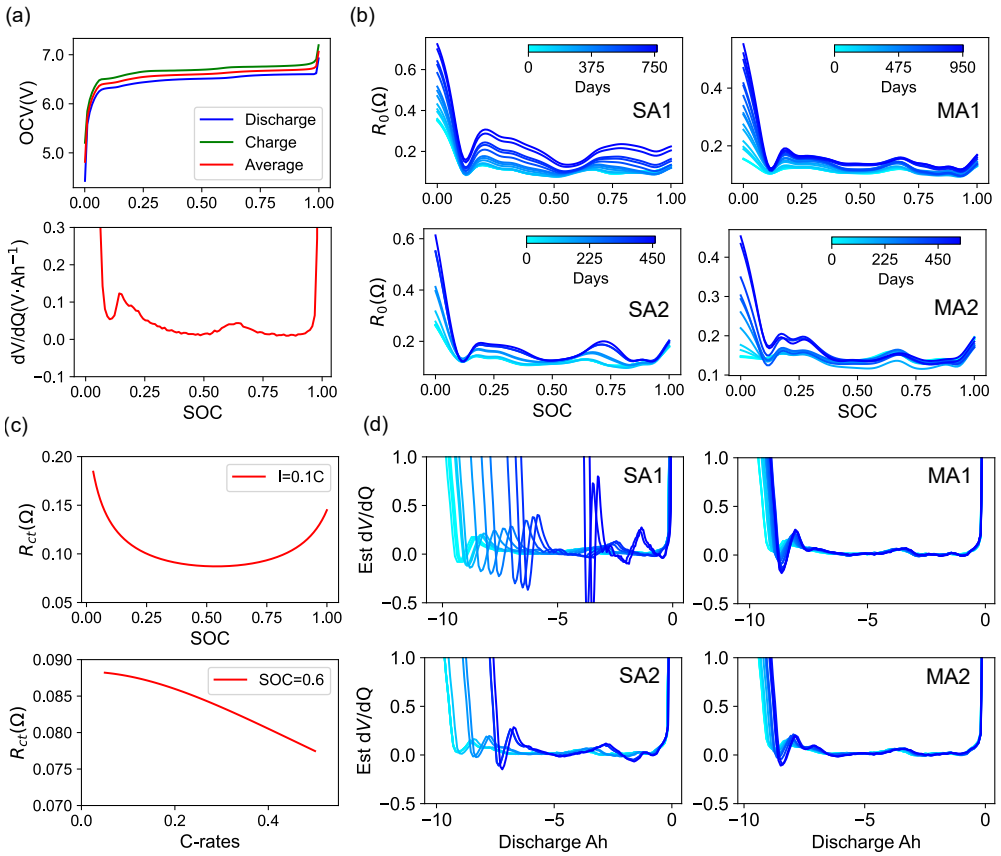


Fig. 5 Resistance estimates from field data reflect changes in underlying OCV curves and can be used for differential voltage analysis. (a) Measured beginning-of-life OCV-SOC curves for discharging, charging and average of them. Large hysterisis effect can be observed between charging and discharging OCV-SOC curves The dV/dQ curve for average OCV-SOC is also included. (b) Estimate of R_0 as a function of SOC for four example packs (SA1, SA2, MA1, MA2). (c) Simulation of charge transfer resistance $R_{\rm ct}$ against SOC and against applied C-rate. (d) Estimates of dV/dQ curves across operating life for the four example packs.

resistance ($R_{\rm ct}$), and ohmic resistance from the electrolyte, electrodes, tabs etc.:

$$V = \underbrace{\mathrm{OCP}_{\rm pe} - \mathrm{OCP}_{\rm ne}}_{\mathrm{OCV},\, V_0} + \underbrace{\eta_{\rm pe} - \eta_{\rm ne}}_{\substack{\text{reaction}\\ \text{overpotential}}} + \underbrace{I \cdot R_{\rm s}}_{\substack{\text{ohmic and ionic}\\ \text{resistance}}} \tag{4}$$

$$\hat{R}_0 = \frac{\Delta V_0}{I} + R_{\rm ct} + R_{\rm s}, \quad R_{\rm ct} = \frac{\eta_{\rm pe} - \eta_{\rm ne}}{I}.$$

Fig. 5(a) shows the lab-measured beginning-of-life OCV-SOC curve and corresponding dV/dQ-SOC curve (for the same cell used in Bboxx packs, details in Supplementary Information). As the overall OCV shape changes relate to the relative shift in positions of the OCPs of the individual electrodes, as well as shrinking of individual electrode OCPs [31], differences between the estimated OCV and the beginning-of-life OCV cause peaks to grow and shift in the dV/dQ-SOC curve (and changes to the OCV-SOC curve). This then leads to the measured fluctuations seen in the learnt $\hat{R}_0$ function. Fig. 5(c) shows a simulation of R_{ct} against SOC generated according to the Bulter-Volmer equation [32]. This simulation is done through Pybamm with the built-in 'Prada2013' parameter set.This explains the overall U-shape component in the learnt $\hat{R}_0$ function.

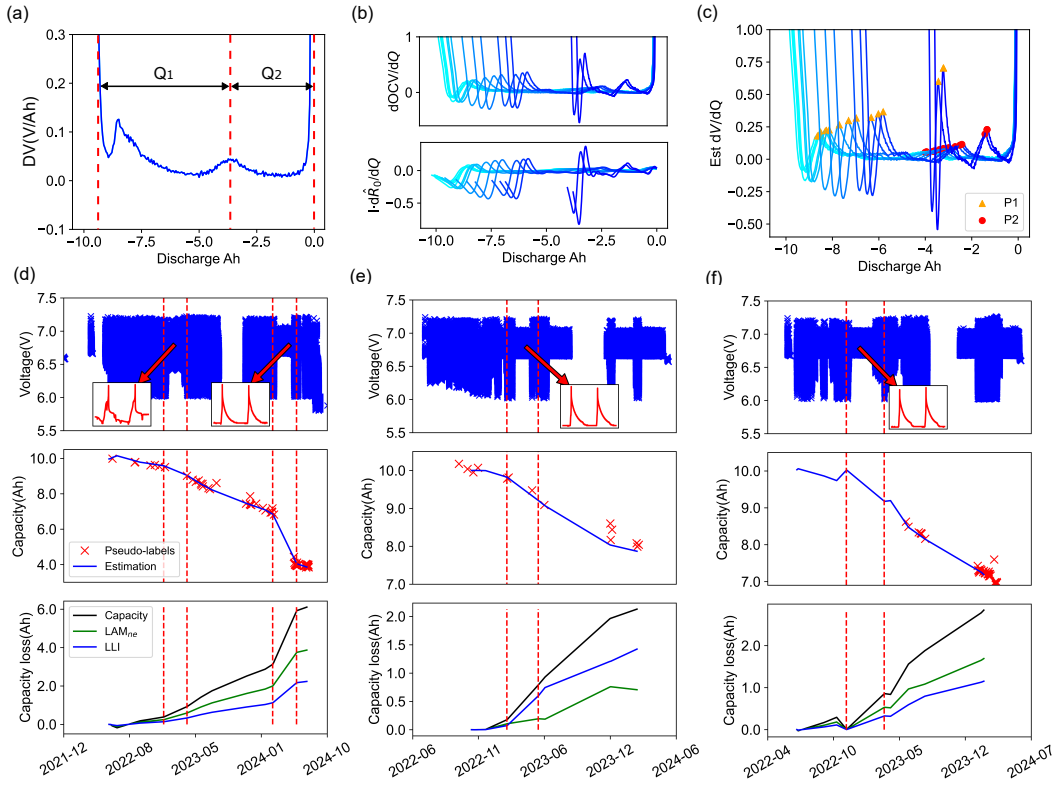


Fig. 6 Examples for linking field usage to degradation modes. (a) The DV curve with two characteristic capacity Q_1 and Q_2 representing LAM_n and LLI respectively. (b) Reconstructed $\frac{d\overline{V}_0}{dQ}$ and $I \cdot \frac{d\hat{R}_0}{dQ}$ in eqn. 5. (c) The estimated DV curve across operation lifetime. Peaks (P1, P2) are highlighted. (d-f) three example packs experiencing accelerated aging due to long rest time at high SOC level. Special voltage profiles are highlighted through red dot lines, and enlarged in inset windows

.

2.6 Differential voltage analysis from field data

Differential voltage analysis (DVA) is a commonly used tool for understanding battery degradation modes, enabling quantification of lost lithium inventory (LLI) and lost active material (LAM) in each electrode [33, 34]. The DV curve is usually obtained by differentiating low-rate measured discharge voltage with respect to cumulative discharge Ampere-hours. Based on half-cell data, peaks on the DV curve can be assigned to electrode phase transitions [35, 36]. Peak shifts and inter-peak distances can be used to calculate LLI, LAM_{pe} and LAM_{ne}.

As mentioned in the preceding section, and in Zhou et al. [28], reconstruction of a DV curve from field data is possible if one assumes that the majority of changes in estimated internal resistance are actually caused by changes in OCV, rather than resistance. Substituting for $\hat{R}_0$ from Eqn. 4, the DV curve can be further expressed as,

$$\text{DV}(z) = \frac{1}{Q}\frac{dV}{dz} = \frac{1}{Q}\left[\frac{d\overline{V}_0}{dz} + I \cdot \frac{d\hat{R}_0}{dz}\right] \tag{5}$$

Typically, a low rate (e.g., C/20) is required to obtain a DV curve; this aims to achieve near electrochemical equilibrium conditions and can be interpreted as maximizing the influence of the ΔV_0 term and minimizing the influence from the R_0 term (see Eqn. 3). For the Bboxx packs analyzed, the small and relatively uniform discharging C-rates (Fig. 1(e)) enable reconstruction of DV curves with relatively small influence from R_0. Thus, given the fixed beginning-of-life OCV ($\overline{V}_0$), and combining our estimates for Q and $\hat{R}_0$, DV curves can be reconstructed across the operational time for each pack. Fig 5(d) shows reconstructed DV curves for the

four example packs (SA1, SA2, MA1, MA2). In the severely aged packs, clear peak shifts can be observed, whereas in the moderately aged packs, noticeable changes in peak locations are minimal.

Although we do not have half-cell data for the cells used here, we can attribute peaks in the DV curve to individual electrodes on the basis of literature [37, 38]. Since the LFP positive-electrode material has a very flat voltage plateau across a wide range of SOC levels [34], the DV curve for LFP cells is mainly dominated by the graphite negative electrode voltage behaviour. Thus, only LLI and $\mathrm{LAM_{ne}}$ can reliably be estimated from reconstructed DV curves. Fig. 6(a) shows that $\mathrm{LAM_{ne}}$ is represented by changes in $\mathrm{Q_1}$, and LLI is represented by changes in $\mathrm{Q_2}$. Fig. 6(b) shows an example of dV_0/dQ and $I \cdot \hat{R}_0/dQ$, which correspond to two terms in the definition of the reconstructed DV curve, Eqn. 5. Combining these, we obtain estimated DV curves across operational lifetime in Fig. 6(c).

Degradation mode estimates from field data allow us to observe how customer usage profiles affect battery aging, beyond just using capacity fade trajectories. Fig. 6(d,e,f) show three severely aged packs as examples, demonstrating that resting at high voltage can cause accelerated increase in both LLI and $\mathrm{LAM_{ne}}$. For most of their use, these packs were cycled within a relatively wide voltage window (6.0 V to 7.2 V), representing a relatively intense daily usage profile. However, as shown in the right-hand zoomed-in section of Fig. 6(d), there is significant time period without any customer discharging. This causes the pack to be charged to the upper cut-off voltage and left to rest there, every day, for several months. Under this extreme usage profile, SA1, SA2 and SA3 all show accelerated increase in LLI. There are several possible explanations for this. First, as the graphite electrode is almost fully intercalated with lithium, it becomes more reactive with the electrolyte solvent, making it susceptible to faster solid-electrolyte interphase (SEI) growth, even at rest [39, 40]. Continued SEI formation consumes Li-ions permanently, causing increase in LLI. Second, the positive electrode is also more likely to oxidize electrolyte at fully de-intercalated state (i.e., high cell SOC), resulting in oxidation products and triggering further parasitic reactions [41]. Third, the increased temperature (35–40 °C) during daytime will exponentially accelerate these effects. On the other hand, in the left zoomed-in window of Fig. 6(d), a shallow discharging behaviour prevents the pack being rested at fully charged state, leading to more modest aging.

Another point worth noting is that $\mathrm{LAM_{ne}}$ also shows accelerated increase during fully charged long rest stages for SA1 and SA3. Compared to LLI, $\mathrm{LAM_{ne}}$ is normally not listed as major degradation factor in the literature for LFP cells. The potential reasons may includes particle isolation and fracture caused by mechanical stress. Specifically, the previously mentioned SEI growth may isolate graphite particles, making them electrochemically inactive [40], which result in $\mathrm{LAM_{ne}}$ increase. Also, deposition of Fe on the graphite electrode has been experimentally observed in literature and causes increasing graphite inaccessibility [42]. On the other hand, full lithiation of graphite induces lattice expansion, and when the cell is held in this state for a long time, stress can be built at particle interfaces, leading to structural changes [43, 44] and causing particles to be unable to participate in charge transfer.

2.7 Population lifetime model

Capacity trajectory estimates from all 114 packs provide a basis for examining the factors that govern aging at the population level. To compare packs with different operating times and aging profiles, we defined an aging rate metric instead of using the absolute capacity fade over the entire operating period. This metric is the capacity drop (ΔQ) per day before the pack degrades to 90% of its initial capacity. Fig. 7(a) shows that, across all 114 packs, the distribution of ΔQ per day is right-skewed with a long tail. We therefore applied a log transformation to this metric to give an aging-rate metric with a more symmetric distribution, i.e. log(ΔQ per day). This can be considered the 'output' of the life model.

The 'inputs' of the model should reflect the factors driving aging, and ideally these should be controllable by the user or designer. For example, if we learn that high temperature accelerates aging, then we might try to improve the thermal management approach. To summarize usage in each system, we considered multiple features extracted from the voltage, current, and temperature data of each pack. A full list of features is given in the Appendix G. To avoid

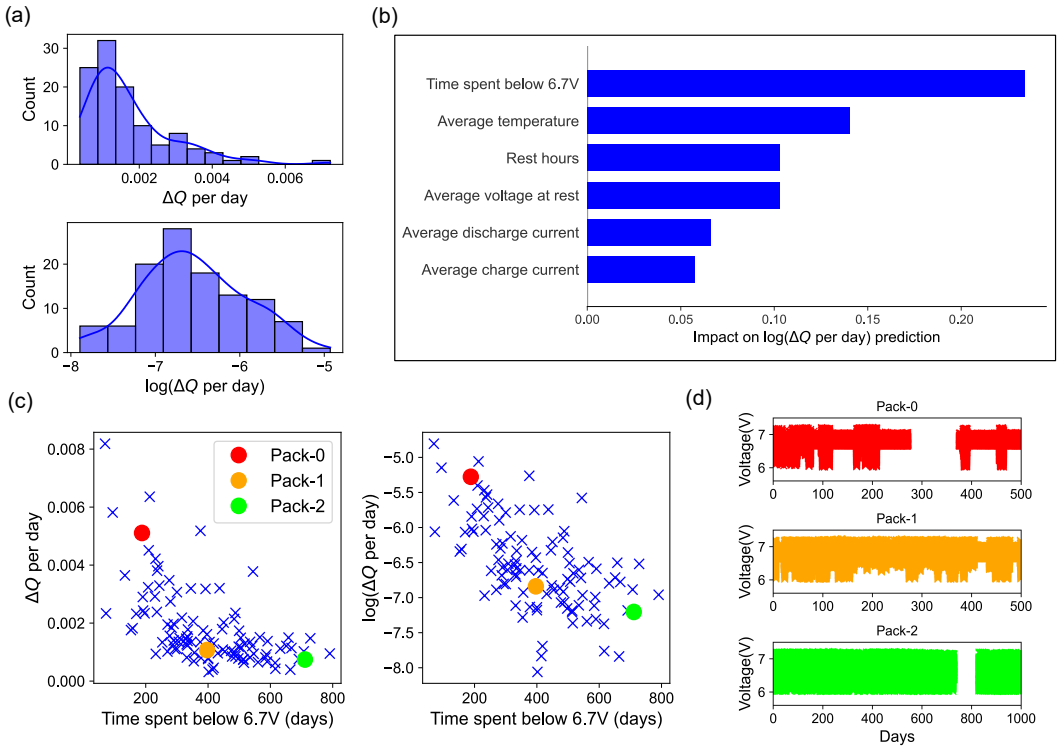


Fig. 7 Population life modelling shows that aging is driven by usage: systems left at high voltage degrade faster. (a) Distribution of capacity drop per day and its log-scale transformation. (b) SHAP analysis showing importance of usage-related features when predicting log(ΔQ per day) labels. (c) Scatter plots of the dominant feature, time spent below 6.7 V, vs. labels, including three example packs (colours). (d) Plots of the terminal voltage for the example packs.

overfitting, highly correlated input features were discarded, retaining only those with relatively strong linear relationships with the output label, log(ΔQ per day). Fig. 7(b) presents the results of a SHAP (SHapley Additive exPlanations) analysis that indicates the relative importance of the selected features. After examining many different possible features we found that a key factor is time spent in certain voltage regions [14]. The dominant feature found here, days spent below 6.7 V, reflects the relative usage intensity of systems. Below this voltage level, customers discharge their systems during the day rather than leaving them charging but not supplying loads.

The relationship between this dominant feature, time spent below 6.7 V, and the output label (log(ΔQ per day)) is directly plotted in Fig. 7(c), where three example packs having high, median, and low values for time spent below 6.7 V are highlighted. Their detailed terminal voltage profiles are also shown in Fig. 7(d). It can be seen that aging is accelerated in systems that do not spend much time below this voltage threshold (i.e., spend a lot of time at high voltage). The correlation is reasonable but not perfect—there is also impact from temperature, the second most important feature, and cell-to-cell variability. However, to reiterate, differences in time spent below 6.7 V across packs mainly arise from the intensity of customer discharging. As the packs are connected to and charged from solar panels, in the absence of customer usage (i.e., discharging to loads), they are charged to the upper cut-off voltage and rest at full SOC on a daily basis, with a high rest voltage (>6.7 V). More rest periods mean less customer discharging, resulting in a smaller value for time spent below 6.7 V. From Pack-0 to Pack-2, the terminal voltage profiles show increasing customer discharging (and correspondingly fewer resting periods). We conclude that, for these LFP packs used in solar off-grid systems, prolonged resting at high voltage accelerates pack aging and should be avoided.

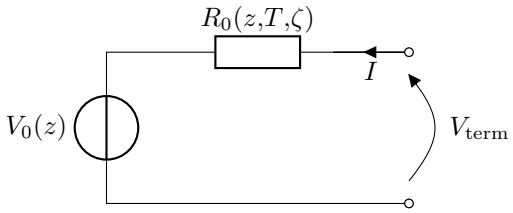


Fig. 8 Equivalent circuit model for Li-ion battery

3 Conclusions

Real-world operating data from 114 commercial LFP packs has been analyzed using a probablistic machine learning approach combined with equivalent circuit models to diagnose health and understand the factors driving life. Accurate estimates for capacity, and reconstruction of degradation modes (LLI and LAM_{ne}), have been demonstrated for these packs. To validate our method, we also generated baseline capacity estimates using Coulomb counting from a limited number of deep discharges between relaxation periods. The resulting comparison shows that we can achieve 1.4% MAPE and 0.2 Ah RMSE on average for capacity estimation across the packs considered. In addition, all 114 packs were used in a population aging analysis where it was found that time spent below 6.7 V is closely related to pack aging rate. This is largely influenced by battery resting behaviour. The less a customer discharges the pack, the more time it will rest at high voltage. This indicates that resting at high voltage (nearly full SOC) will accelerate pack aging in the LFP cells considered here.

Beyond capacity estimation and life modelling, we also discuss degradation mode analysis from field data, using estimated resistance as a proxy for OCV-SOC functional changes. The estimated LLI and LAM_{ne} quantities enable a direct link to be established between usage profile and electrode-specific degradation. High voltage rests accelerate both modes.

Alongside this paper, we also publish the dataset consisting of operational data from 114 packs, covering more than 2500 months of field cycling, which to our knowledge will be the largest available field dataset for LFP batteries. Broadly, this work highlights the opportunity to use machine learning combined with conventional circuit models to estimate multidimensional battery health from live operational data, and then gain insights into the factors driving aging. There are several key impacts: Capacity estimates and predictions help to inform cell replacement logistics and total cost of ownership accounting. Understanding the link between usage patterns and aging can lead to opportunities to extent battery life for example by reducing time spent at high voltage. Finally, rolling out these methods across different types of cells (e.g., from different suppliers) can enable comparative performance assessment and inform future cell selection.

4 Methods

4.1 Aging-aware equivalent circuit model

As shown in Fig. 8, a simple ECM is used to model battery dynamics. This approach is based on our former work [28] and the method in Aitio et al. [45]. The corresponding mathematical formulation is given as

$$\begin{aligned} V_{\text{term}} &= V_0\left(z\right) + R_0\left(z, T, \zeta\right) I \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= IQ^{-1}\left(\zeta\right) \end{aligned} \tag{6}$$

where z is the SOC, I is the applied current, and $V_0(z)$ represents the OCV-SOC curve, which is often measured at beginning of life and assumed to be fixed over life. In this work, a new LFP cell, the same type used in Bboxx packs, was tested in our lab for a initial-of-life $V_0(z)$. Two timescales are involved in the model, one is the operating timescale (t), which is often at the seconds level. Another is the life-long aging timescale (ζ), which can vary from days to months depending on usage. The variables I, z, V_{term} are all functions of t, unless written in their discrete-time versions that are demarcated with subscripts (e.g. z_i). For clarity the

time-dependence of the continuous-time variables is assumed, rather than written out in every equation. For simplicity, we use a vector to denote operating conditions,

$$\mathbf{s} = [z, T]^\top. \tag{7}$$

The inverse capacity term Q^{-1} is modeled as a GP over ζ, while the resistance term R_0 is modeled as a GP over $\mathbf{s}$ and ζ,

$$\begin{aligned} Q^{-1}(\zeta) &\sim \mathcal{GP}\left(m(\zeta), k\left(\zeta, \zeta'\right)\right) \\ R_0\left(\mathbf{s}, \zeta\right) &\sim \mathcal{GP}\left(m\left(\mathbf{s}, \zeta\right), k\left(\left(\mathbf{s}, \zeta\right), \left(\mathbf{s}', \zeta'\right)\right)\right). \end{aligned} \tag{8}$$

where k is the covariance/kernel function used in GPs. Different kernels are used for capacity Q and resistance R_0. For Q, life-long aging dependency is modeled with a wiener velocity(WV) kernel,

$$k_{\mathrm{WV}}\left(\zeta, \zeta'\right) = \sigma_\zeta^2 \left(\frac{\min^3\left(\zeta, \zeta'\right)}{3} + \left|\zeta - \zeta'\right| \frac{\min^2\left(\zeta, \zeta'\right)}{2} \right) \tag{9}$$

The WV kernel is a non-stationary kernel that is more suitable for extrapolating aging trends into the future. For R_0, the dependency on operating condition and life-long aging are described by a multiplication of two kernels,

$$\begin{aligned} &k\left(\left(\mathbf{s}, \zeta\right), \left(\mathbf{s}', \zeta'\right)\right) = k_{\text{Matern32}}\ \left(\mathbf{s}, \mathbf{s}'\right) k_{\mathrm{WV}}\left(\zeta, \zeta'\right) \\ &k_{\text{Matern32}}\left(\mathbf{s}, \mathbf{s}'\right) = \sigma_s^2 (1 + \sqrt{3}d) \exp(-\sqrt{3}d) \\ &d^2 = \left|\mathbf{s} - \mathbf{s}'\right|^\top \begin{bmatrix} l_z^{-2} & 0 \\ 0 & l_I^{-2} \end{bmatrix} \left|\mathbf{s} - \mathbf{s}'\right|. \end{aligned} \tag{10}$$

The goal is to get maximum-a-posteriori(MAP) estimates for kernel hyper-parameters θ by minimize the negative log maximum likelihood(NLML).

$$\begin{aligned} &\text{NLML} = -\log p(\mathbf{y} \mid X, \theta) = \\ &-\frac{1}{2}\mathbf{y}^{\mathrm{T}} \left[\mathbf{K}_X + \sigma_n^2 \mathbf{I}\right]^{-1} \mathbf{y} - \frac{1}{2}\log\left|\mathbf{K}_X + \sigma_n^2 \mathbf{I}\right| - \frac{n}{2}\log 2\pi. \end{aligned} \tag{11}$$

where, the kernel matrices are simplified as $\mathbf{K}_X = k(\mathbf{X}, \mathbf{X})$, $\mathbf{K}_{X,*} = k(\mathbf{X}, \mathbf{X}^*)$. $\mathbf{X}$ and $\mathbf{X}^*$ represent train and test samples, respectively. However, as the matrix inversion scales computationally at $\mathcal{O}\left(n^3\right)$, long time series from battery field operation pose challenges for training GPs in a standard way. To address this, a stata-space formulation [46] is employed to interprets the Gaussian process $Q^{-1}(\zeta)$ and $R_0(s, \zeta)$ as the solution to a linear time-invariant stochastic partial differential equation (LTI-SPDE),

$$\begin{aligned} \frac{\partial Q(\zeta)}{\partial \zeta} &= \mathbf{F} Q^{-1}(\zeta) + \mathbf{L}\omega(\zeta) \\ \frac{\partial R_0(\mathbf{s}, \zeta)}{\partial \zeta} &= (\mathbf{I}_{n_\mathrm{s}} \otimes \mathbf{F}) R_0(\mathbf{s}, \zeta) + \mathbf{L}\omega(\mathbf{s}, \zeta). \\ \omega(\mathbf{s}, \zeta) &\sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_\mathrm{s} \otimes \omega(\zeta)) \end{aligned} \tag{12}$$

where, the transition matrix $\mathbf{F}$, dispersion matrix $\mathbf{L}$, and the the white noise process $\omega(\zeta)$ are determined by the life-aging time kernel k_{WV}. $R_0(\mathbf{s}, \zeta)$ represents the values of R_0 on a set of discretized operational points $\mathbf{s} = (s_0, s_1, ..., s_{ns})$ at a specific aging time ζ. $\mathbf{K}_\mathrm{s}$ is the covariance matrix for operation condition kernel k_{Matern32} . Intuitively, one can view this approach as constructing a LTI system which takes a white noise process as input and generates an output that is the desired Gaussian process. For $R_0(\mathbf{s}, \zeta)$, the dependency of the operating condition is maintained through the constructed noise process. Once the GP is formulated as a discrete-time LTI system, the hyperparameter NLML in eqn. 11 can be achieved recursively using a Kalman filter and Rauch-Tung-Striebel (RTS) smoother [47], enabling linear time complexity $\mathcal{O}\left(n\right)$ in forward model runs [46].

References

[1] Citaristi, I.: International energy agency—iea. In: The Europa Directory of International Organizations 2022, pp. 701–702. Routledge, London (2022)

[2] Shell Foundation: Enabling batteries to die another day: Improving battery economics through the circular economy for africa. Technical report, Shell Foundation (2021). Accessed May 2025. https://shellfoundation.org/wp-content/uploads/2022/07/ENABLI1.pdf

[3] Hu, X., Xu, L., Lin, X., Pecht, M.: Battery lifetime prognostics. Joule **4**(2), 310–346 (2020)

[4] Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y.-H., Chen, M.H., Cheong, B., Perkins, N., Yang, Z., *et al.*: Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature **578**(7795), 397–402 (2020)

[5] Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y.T., Steinbacher, F., Khan, M.U., Lee, J.W., Siegel, J.B., Stefanopoulou, A.G., *et al.*: The challenge and opportunity of battery lifetime prediction from field data. Joule **5**(8), 1934–1955 (2021)

[6] Dubarry, M., Tun, M., Baure, G., Matsuura, M., Rocheleau, R.E.: Battery durability and reliability under electric utility grid operations: analysis of on-site reference tests. Electronics **10**(13), 1593 (2021)

[7] Ge, M.-F., Liu, Y., Jiang, X., Liu, J.: A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries. Measurement **174**, 109057 (2021)

[8] Plett, G.L.: Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 3. state and parameter estimation. Journal of Power sources **134**(2), 277–292 (2004)

[9] Wang, Y., Zhang, C., Chen, Z.: A method for joint estimation of state-of-charge and available energy of lifepo4 batteries. Applied energy **135**, 81–87 (2014)

[10] Yu, Q., Xiong, R., Yang, R., Pecht, M.G.: Online capacity estimation for lithium-ion batteries through joint estimation method. Applied Energy **255**, 113817 (2019)

[11] Wassiliadis, N., Adermann, J., Frericks, A., Pak, M., Reiter, C., Lohmann, B., Lienkamp, M.: Revisiting the dual extended kalman filter for battery state-of-charge and state-of-health estimation: A use-case life cycle analysis. Journal of Energy Storage **19**, 73–87 (2018)

[12] Richardson, R.R., Osborne, M.A., Howey, D.A.: Gaussian process regression for forecasting battery state of health. Journal of Power Sources **357**, 209–219 (2017)

[13] Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., *et al.*: Data-driven prediction of battery cycle life before capacity degradation. Nature Energy **4**(5), 383–391 (2019)

[14] Greenbank, S., Howey, D.: Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life. IEEE Transactions on Industrial Informatics **18**(5), 2965–2973 (2021)

[15] Ma, Y., Shan, C., Gao, J., Chen, H.: A novel method for state of health estimation of lithium-ion batteries based on improved lstm and health indicators extraction. Energy **251**, 123973 (2022)

[16] Li, T., Zhou, Z., Thelen, A., Howey, D.A., Hu, C.: Predicting battery lifetime under varying usage conditions from early aging data. Cell Reports Physical Science **5**(4) (2024)

[17] Plett, G.L.: Battery Management Systems, Volume II: Equivalent-circuit Methods. Artech House, Boston (2015)

[18] Li, H., Pan, D., Chen, C.P.: Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine. IEEE Transactions on Systems, Man, and Cybernetics: Systems **44**(7), 851–862 (2014)

[19] Gu, X., See, K.W., Li, P., Shan, K., Wang, Y., Zhao, L., Lim, K.C., Zhang, N.: A novel state-of-health estimation for the lithium-ion battery using a convolutional neural network and transformer model. Energy **262**, 125501 (2023)

[20] Aitio, A., Howey, D.A.: Predicting battery end of life from solar off-grid system field data using machine learning. Joule **5**(12), 3204–3220 (2021)

[21] Pozzato, G., Allam, A., Pulvirenti, L., Negoita, G.A., Paxton, W.A., Onori, S.: Analysis and key findings from real-world electric vehicle field data. Joule **7**(9), 2035–2053 (2023)

[22] Figgener, J., Ouwerkerk, J., Haberschusz, D., Bors, J., Woerner, P., Mennekes, M., Hildenbrand, F., Hecht, C., Kairies, K.-P., Wessels, O., *et al.*: Multi-year field measurements of home storage systems and their use in capacity estimation. Nature Energy **9**(11), 1438–1447 (2024)

[23] Dubarry, M., Truchot, C., Liaw, B.Y.: Synthesize battery degradation modes via a diagnostic and prognostic model. Journal of power sources **219**, 204–216 (2012)

[24] Dubarry, M., Anseán, D.: Best practices for incremental capacity analysis. Frontiers in Energy Research **10**, 1023555 (2022)

[25] Keil, P., Jossen, A.: Impact of dynamic driving loads and regenerative braking on the aging of lithium-ion batteries in electric vehicles. Journal of the electrochemical society **164**(13), 3081 (2017)

[26] Qu, J., Jiang, Z., Zhang, J.: Investigation on lithium-ion battery degradation induced by combined effect of current rate and operating temperature during fast charging. Journal of Energy Storage **52**, 104811 (2022)

[27] Ven, A., See, K.A., Pilon, L.: Hysteresis in electrochemical systems. Battery Energy **1**(2), 20210017 (2022)

[28] Zhou, Z., Aitio, A., Howey, D.: Learning operando impedance function for battery health with aging-aware equivalent circuit model. arXiv preprint arXiv:2407.06639 (2024)

[29] Williams, C.K., Rasmussen, C.E.: Gaussian Processes for Machine Learning vol. 2. MIT press, Cambridge, MA (2006)

[30] Hu, X., Li, S., Peng, H.: A comparative study of equivalent circuit models for li-ion batteries. Journal of Power Sources **198**, 359–367 (2012)

[31] Birkl, C.R., Roberts, M.R., McTurk, E., Bruce, P.G., Howey, D.A.: Degradation diagnostics for lithium ion cells. Journal of Power Sources **341**, 373–386 (2017)

[32] Newman, J., Balsara, N.P.: Electrochemical Systems. John Wiley & Sons, Hoboken, New Jersey (2021)

[33] Safari, M., Delacourt, C.: Aging of a commercial graphite/lifepo4 cell. Journal of The

Electrochemical Society **158**(10), 1123 (2011)

[34] Lewerenz, M., Marongiu, A., Warnecke, A., Sauer, D.U.: Differential voltage analysis as a tool for analyzing inhomogeneous aging: A case study for lifepo4— graphite cylindrical cells. Journal of Power Sources **368**, 57–67 (2017)

[35] Bloom, I., Jansen, A.N., Abraham, D.P., Knuth, J., Jones, S.A., Battaglia, V.S., Henriksen, G.L.: Differential voltage analyses of high-power, lithium-ion cells: 1. technique and application. Journal of Power Sources **139**(1-2), 295–303 (2005)

[36] Bloom, I., Christophersen, J., Gering, K.: Differential voltage analyses of high-power lithium-ion cells: 2. applications. Journal of Power Sources **139**(1-2), 304–313 (2005)

[37] Keil, P., Schuster, S.F., Wilhelm, J., Travi, J., Hauser, A., Karl, R.C., Jossen, A.: Calendar aging of lithium-ion batteries. Journal of The Electrochemical Society **163**(9), 1872 (2016)

[38] Zhu, J., Darma, M.S.D., Knapp, M., Sørensen, D.R., Heere, M., Fang, Q., Wang, X., Dai, H., Mereacre, L., Senyshyn, A., *et al.*: Investigation of lithium-ion battery degradation mechanisms by combining differential voltage analysis and alternating current impedance. Journal of Power Sources **448**, 227575 (2020)

[39] Dubarry, M., Liaw, B.Y.: Identify capacity fading mechanism in a commercial lifepo4 cell. Journal of power sources **194**(1), 541–549 (2009)

[40] Kassem, M., Bernard, J., Revel, R., Pelissier, S., Duclaud, F., Delacourt, C.: Calendar aging of a graphite/lifepo4 cell. Journal of Power Sources **208**, 296–305 (2012)

[41] Qiao, X., Mu, Y., Peng, J., Pei, B., Wang, S.: Deciphering the calendar aging degradation mechanism of lifepo4-graphite pouch cells. ACS Applied Energy Materials (2025)

[42] Li, D., Danilov, D.L., Xie, J., Raijmakers, L., Gao, L., Yang, Y., Notten, P.H.: Degradation mechanisms of c6/lifepo4 batteries: experimental analyses of calendar aging. Electrochimica Acta **190**, 1124–1133 (2016)

[43] Haik, O., Ganin, S., Gershinsky, G., Zinigrad, E., Markovsky, B., Aurbach, D., Halalay, I.: On the thermal behavior of lithium intercalated graphites. Journal of the Electrochemical Society **158**(8), 913 (2011)

[44] Hölderle, T., Monchak, M., Baran, V., Kriele, A., Mühlbauer, M.J., Dyadkin, V., Rabenbauer, A., Schökel, A., Ehrenberg, H., Müller-Buschbaum, P., *et al.*: Thermal structural behavior of electrochemicallylithiated graphite (lixc6) anodes in li-ion batteries. Batteries & supercaps **7**(3), 202300499 (2024)

[45] Aitio, A., Jöst, D., Sauer, D.U., Howey, D.A.: Learning battery model parameter dynamics from data with recursive gaussian process regression. Journal of Dynamic Systems, Measurement, and Control **147**(3), 031010 (2025)

[46] Sarkka, S., Solin, A., Hartikainen, J.: Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering. IEEE Signal Processing Magazine **30**(4), 51–61 (2013)

[47] Särkkä, S., Svensson, L.: Bayesian Filtering and Smoothing vol. 17. Cambridge university press, Cambridge, England (2023)

[48] Hu, C., Youn, B.D., Chung, J.: A multiscale framework with extended kalman filter for lithium-ion battery soc and capacity estimation. Applied Energy **92**, 694–704 (2012)

[49] Xiong, R., Sun, F., Chen, Z., He, H.: A data-driven multi-scale extended kalman filtering

based parameter and state estimation approach of lithium-ion polymer battery in electric vehicles. Applied energy **113**, 463–476 (2014)

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Battery capacity and degradation modes diagnosis for solar off-grid systems in Africa
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	Zhou Zihao, Brady Planden, and David Howey. "Battery capacity and degradation modes diagnosis for solar off-grid systems in Africa." (To be submitted).

Student Confirmation

Student Name:	Zihao Zhou		
Contribution to the Paper	First author I developed both the segment extraction and SOH estimation model for all used Bboxx samples. I implemented the corresponding code, conducted all experiments for the paper, performed additional analyses, and authored the whole manuscript.		
Signature	Zihao Zhou	Date	21 July 2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: David A. Howey, Professor of Engineering Science			
Supervisor comments Zihao led this work and the description given above is accurate.			
Signature	David Howey	Date	15 Aug 2025

This completed form should be included in the thesis, at the end of the relevant chapter.

5 | Conclusion

5.1 Contributions

5.1.1 Battery lifetime early prediction

The aim of the Chapter 2 was to achieve accurate battery EoL prediction based on cycling data from early-life usage. We divided health features into two categories, individual cell-level features (reflecting cell intrinsic variability) and population-wide features (reflecting impact of cycling conditions on a population average). This categorization was motivated by commonly observed battery aging behaviour. For a group of identical batteries, cycling under different usage conditions often leads to significant differences in aging behaviour, primarily reflecting usage-induced degradation. On the other hand, lifetime differences are also observed among batteries cycled under the same usage profile, which are attributed to intrinsic cell variability. These initially small cell-to-cell differences tend to magnify as aging progresses.

In the first paper of Chapter 2, the proposed population-wide feature—SOC-averaged charging C-rate, shows a strong linear relationship (0.71 Pearson coefficient) with EoL labels. This feature defines differences between protocols in the sense of total ohmic heat generation in the cell during each charging period. Given that the open dataset used [26, 27] was designed for fast-charging applications, this result suggests that a significant portion of the aging is driven by variations in the charging profile, rather than intrinsic manufacturing variability. Also, it is likely the cell heating effect is the key factor for aging under this fast-charging test. This finding also agrees with the analysis of Alexis's work [32], in which they found features encoding cycling conditions boost model accuracy mainly because they predict protocol-to-protocol variability. We also recognised features derived from the shift in the $Q(V)$ curve to be individual-level features, representing intrinsic cell variability.

In the second paper of Chapter 2, we expanded the range of usage conditions (in terms of charging C-rate, discharging C-rate, and depth of discharge, DoD) using a self-

collected dataset (generated by collaborators in University of Connecticut). A feature was proposed encoding average cycling 'stress', defined $\text{Stress}_{\text{avg}} = (C_{\text{dchg}}^{0.5} \cdot \text{DoD}^{0.5} + C_{\text{chg}}^{0.5} \cdot \text{DoD}^{0.5})/2$. Individual-level features were collected from the incremental capacity curves. Specifically, segments of the incremental capacity curve within a certain voltage interval (3.6 V to 3.9 V) are extracted from two RPT tests (the week 0 RPT and the week 3 RPT, respectively). Then, the mean and variance of the difference between these two extracted segments are used as individual-level features, which reflects the evolution of the incremental capacity curve during early life usage (i.e. first 3 weeks).

The datasets used to validate HBM performance are different in chemistry. The first paper [1] of Chapter 2 analysed an open dataset consists of 169 LFP cells underwent fast aging experiments, while the second paper [2] analysed and published aging data from 225 NMC cells covering 61 more realistic cycling conditions. In general, LFP cells tend to show more cycle life under moderate operation conditions compared to NMC cells [81, 82]. NMC cells often have higher energy density which makes them more sensitive to certain degradation mechanisms, such as high-voltage stress, structural degradation of cathode, and Li-plating at the anode. However, in Chapter 2, NMC cells show much longer lifetime compared to LFP cells. This is due to the vast differences in the aging environments they experienced. Each fast-aging cycling condition operated on LFP cells consist of a step-wise 3-8C charging profile as well as a 4C discharging profile. And the depth of discharge is set to 100% (though, the high discharging C-rate will result in the cell reaching lower cut-off voltage before fully discharged). While for NMC cells, comparably moderate aging conditions are operated with charging and discharging C-rates between 0.5-3C, and DoD between 5-100%. Under such extreme-rate operation for LFP and moderate rates for NMC—the influence of intrinsic cell chemistry on aging becomes secondary to that of operating conditions. In this case, degradation trends are governed primarily by kinetic and thermal stresses imposed by the cycling protocol rather than by material-specific reactions [83].

We observed that the relationship between individual-level features and EoL labels changes under different usage conditions. This can be interpreted as a hierarchical dependency between the two categories of features. To model this dependency explicitly

and quantify uncertainty in predictions, we proposed a hierarchical Bayesian model (HBM). Compared to baseline methods, the HBM achieved superior prediction accuracy, particularly on out-of-distribution samples.

5.1.2 Battery SOH estimation

The purpose of Chapters 3 and 4 was to estimate the SOH degradation trajectory from battery cycling data. In this context, the main contribution of Chapter 3 was the proposed aging-aware ECM that integrates Gaussian process regression (GPR) into the ECM framework. The dependence of ECM parameters on aging and operating conditions was modeled using GPR, enhancing the model's flexibility and effectively framing SOH estimation as a learning task of this dependency. A state-space formulation of a Gaussian process was employed to transform the GPs into a time-stepping state space model, enabling a solution with linear time complexity. This formulation also allows us to combine GP-represented parameter dynamics and ECM state (i.e. SOC) dynamics into a unified nonlinear system. A recursive co-estimation framework was then proposed to jointly estimate both model parameters and battery states from operational data.

The proposed aging-aware ECM builds upon previous works [79, 80] but with a simpler ECM emphasizing the importance of the estimated resistance function. We observed the estimated resistance function is significantly affected by the aging-induced shape changes in the OCV-SOC relationship. This insight enables us to reconstruct differential voltage (DV) curves, thereby allowing SOH estimation not only in terms of capacity but also in terms of degradation modes. One remaining issue concerns the identifiability of changes in OCV-SOC curve versus estimates of the resistance function. Given their additive nature, it is impossible to distinguish them individually from terminal voltage measurements. However, the differences in their influences on terminal voltage can be compared. When the applied current is relatively small ($< 0.5\,\mathrm{C}$), the influence from OCV changes dominant, which allows us to reconstruct reliable DV curves and get estimations for degradation modes.

In Chapter 4, the goal was to estimate SOH from real-world operating data of LFP packs and derive population-level aging insights based on the estimated SOH trajectories.

Operational data from 114 LFP packs used in solar off-grid systems was carefully analyzed. Compared to lab-tested data, field data presents greater challenges. This work addressed these challenges systematically:

- First, the combination of the flat OCV profile of LFP batteries and shallow cycling usage reduces ECM parameter identifiability. To address this, we developed a segment extraction process, in which only deep and continuous discharging segments are selected across the packs' operational lifetimes.
- Second, field data often lacks ground-truth SOH values for model validation. Based on the definition of discharge capacity, we derived 'pseudo-labels' for 56 packs, which validated the performance of our proposed aging-aware ECM by providing a comparative SOH estimate.
- Third, the low sampling rate typical of field data can lead to significant errors in Coulomb counting. However, we observed that the distribution of discharge C-rates in these off-grid systems has distinct peaks, suggesting limited variability in customer usage patterns.

A population-level aging analysis revealed that time spent below 6.7 V is closely related to capacity drop per day. Low value for time spent below 6.7 V over a pack's operational life is often caused by limited customer discharging, which leads to long rest periods at high voltage. This effect is also reflected in the estimated degradation modes—LLI and LAM_{ne}. Compared to regular discharging usage, these degradation modes are more pronounced when packs are left at high voltage for long durations. Therefore, we recommend reducing or avoiding extended rest periods at high voltage as a strategy to extend battery life in off-grid solar-battery system applications.

5.2 Future work

Over the course of this doctoral research, there has been an ever-increasing quantity of research papers on implementing machine learning methods on the problem of battery lifetime prediction and SOH estimation. Every innovation in the machine learning domain fosters several new application papers in batteries. While powerful new methods

often report outstanding accuracy on test cases, several concerns still remain. First, the generalizability of these models beyond their specific training datasets is often unclear. Many studies report high accuracy on well-curated lab datasets, but their performance in real-world, noisy, and variable conditions is rarely tested. This raises the question of whether the models are truly learning generalised underlying degradation mechanisms or simply capturing dataset-specific patterns. Second, there's the issue of data availability. High-quality, diverse, and representative datasets are essential for training reliable battery health machine learning models, but such datasets are still scarce—especially for field-deployed batteries in diverse applications. This limits the scalability and reliability of many proposed methods in broader industrial settings. Finally, the interpretability of complex models is often limited, especially deep learning-based approaches. While they may offer high prediction accuracy, they usually function as black boxes, making it difficult to extract physical insights or understand why a battery is predicted to degrade in a certain way. For engineers and practitioners seeking to optimize battery usage or design better systems, such insights are often much more valuable than marginal gains in accuracy.

5.2.1 Limitations and potential improvements

The work presented in this thesis has some limitations. For the topic of battery lifetime early prediction, the proposed HBM explicitly models the hierarchical dependency between usage condition level and cell-level features. However, extracting informative features that adequately summarize differences among usage profiles remains challenging. For lab aging datasets, cycling profiles are typically well controlled, making it relatively straightforward to construct meaningful usage-condition features using variables such as charging/discharging C-rates and DoD. In contrast, field data exhibit high-dimensional and heterogeneous usage conditions, with limited samples spanning this space. As a result, reducing customer usage profiles to a low-dimensional set of features becomes a substantially more difficult task. Another limitation arises from the scarcity of training data. In principle, robust model performance relies on having sufficient samples that cover the full usage-condition space. This requirement is difficult

to meet because obtaining labels (e.g., discharge capacity) necessitates time-consuming low-rate constant-current tests. Dataset expansion through data augmentation may offer a promising direction to mitigate this issue.

For the topic of SOH estimation, the proposed aging-aware ECM can effectively estimate both discharge capacity and the resistance function using discharging segments extracted from daily usage. Although several principles have been proposed to guide the segment extraction process, there is still a lack of quantitative metrics for comparing the effectiveness of different segments in battery SOH estimation. Another limitation arises in pack-level implementation. Daily operational data are often recorded only at the pack or module level, providing very limited visibility into individual cell behaviour—typically only the maximum and minimum values across the pack or module are available. Cell-to-cell variability within a pack or module is a critical factor in assessing pack health, yet it is not explicitly captured in the current aging-aware ECM. Finally, a further potential improvement lies in adopting a more sophisticated ECM that incorporates voltage relaxation and hysteresis effects. However, this would substantially increase the difficulty of model training, as the aging behaviour of all these circuit elements would need to be learned from data.

The work presented in this thesis also gives rise to several future research directions. Broadly, there are two key objectives when considering battery health. First, we aim to obtain accurate SOH estimates for individual systems (either cells, modules, or packs) based on field operational data. Second, we seek to derive population-level aging insights and lifetime models that can inform practical strategies to extend battery lifetime. Future works for both these two directions are discussed in the next two subsections.

5.2.2 Accuracy of state and parameter co-estimation under real-world applications

The co-estimation of states and parameters remains a significant challenge, particularly in real-world applications. In the lab, with sufficient time and resources, ECM states and parameters can be calibrated separately. Since the terminal voltage after a long rest period can be considered a proxy for the OCV, the battery state (i.e., SOC) can be

inferred using a pre-tested OCV–SOC lookup table. With this calibrated SOC value, carefully designed experiments, such as HPPC and GITT, can be used to separately calibrate resistance, RC time constants, and capacity.

However, in the field, one must estimate both parameters and states jointly from operational data. The data required to identify capacity and resistance are different. Internal resistance, as a function of SOC and temperature, can be identified using short current pulses. Capacity, on the other hand, requires deep discharge periods at low rates so that its influence on terminal voltage becomes observable. Moreover, accurate estimation of both resistance and capacity depends heavily on SOC estimation accuracy. The key question becomes: what types of operational segments enable better co-estimation results?

An intuitive segment extraction process was introduced in Chapter 4. However, a clear criterion for what kinds of field usage profiles are optimal for co-estimation of state and parameter is still lacking. Some earlier work [84] uses Fisher information and the Cramér–Rao bound to provide analytic limits on the estimation accuracy of SOC, capacity, and resistance. However, the influence of sampling rate is not included in these analyses, and the methods have only been validated on limited lab-test current profiles, not on real-world cycling data.

5.2.3 Population-level mapping from usage profiles to aging effects

One major expectation for the study of battery SOH estimation is to get practical advice about in-field operating strategy so that one can prolong battery lifetime. To achieve this goal, the first step will be build the mapping between usage profiles and corresponding aging effects. In Chapter 4, we did a preliminary population-level aging analysis for 114 LFP packs, which revealed accelerated aging effects caused by long rest periods at high voltage.

However, it is difficult to build such a mapping for two main reasons. First, we notice that summarizing usage profiles into several statistic values (or namely features) is a very difficult task. In lab-test profiles, cycling conditions are largely controlled, often varying

only one aspect at a time (such as charge C-rate, discharge C-rate, depth of discharge, or ambient temperature). In contrast, for batteries used in the field, usage profiles can vary in all of these aspects simultaneously, and may also involve additional factors that are difficult to quantify —— for example, the dynamic nature of the applied current. Moreover, for LFP batteries, the effectiveness of features extracted from voltage signals is significantly reduced due to the flatness of the OCV-SOC curve. Cycling behaviors within the middle SOC range (20–80%) produce only limited variation in the terminal voltage signal. The hysteresis effect further worsens this problem.

Second, a reliable mapping that can guide daily operation requires a sufficiently large number of feature-label pairs (i.e., different usage profiles and their corresponding aging behaviors) in the training data. From a regression modeling perspective, more training data helps the true underlying mapping to emerge more clearly, overriding noise and spurious patterns. In population-level aging analysis, noise arises from two primary sources: first, the accuracy of SOH trajectory estimation for individual batteries; and second, the inherent cell-to-cell variability. Even batteries of the same type, produced by the same manufacturer and cycled under identical protocols, can exhibit different aging trajectories.

In the off-grid solar systems analyzed in Chapter 4, variability in usage profiles is limited, as users typically operate them for common applications such as home lighting and phone charging. We anticipate that more meaningful and informative mappings between usage patterns and corresponding aging effects will emerge in applications like electric vehicles (EVs) and energy trading, where usage profiles are far more diverse.

References

[1] Z. Zhou and D. A. Howey, "Bayesian hierarchical modelling for battery lifetime early prediction," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 6117–6123, 2023.

[2] T. Li, Z. Zhou, A. Thelen, D. A. Howey, and C. Hu, "Predicting battery lifetime under varying usage conditions from early aging data," *Cell Reports Physical Science*, vol. 5, no. 4, 2024.

[3] Z. Zhou, A. Aitio, and D. Howey, "Learning li-ion battery health and degradation modes from data with aging-aware circuit models," *Applied Energy*, vol. 397, p. 126375, 2025.

[4] Z. Zhou, P. Brady, and D. Howey, "Battery capacity and degradation modes diagnosis for solar off-grid systems in africa," *submitted for peer review in July 2025*, 2025.

[5] E. S. M. A. Programme, *Mini Grids for Half a Billion People: Market Outlook and Handbook for Decision Makers*. World Bank, 2019.

[6] E. S. G. Challenge, "Energy storage market report," *US Department of Energy: Washington, DC, USA*, 2020.

[7] D. S. Mallapragada, N. A. Sepulveda, and J. D. Jenkins, "Long-run system value of battery energy storage in future grids with increasing wind and solar generation," *Applied Energy*, vol. 275, p. 115390, 2020.

[8] E. M. Bibra, E. Connelly, S. Dhir, M. Drtil, P. Henriot, I. Hwang, J.-B. Le Marois, S. McBain, L. Paoli, and J. Teter, "Global ev outlook 2022: Securing supplies for an electric future," 2022.

[9] J. M. Reniers, G. Mulder, S. Ober-Blöbaum, and D. A. Howey, "Improving optimal control of grid-connected lithium-ion batteries through more accurate battery and degradation modelling," *Journal of Power Sources*, vol. 379, pp. 91–102, 2018.

[10] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 1. background," *Journal of Power sources*, vol. 134, no. 2, pp. 252–261, 2004.

[11] S. M. Rezvanizaniani, Z. Liu, Y. Chen, and J. Lee, "Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (ev) safety and mobility," *Journal of power sources*, vol. 256, pp. 110–124, 2014.

[12] H. C. Hesse, M. Schimpe, D. Kucevic, and A. Jossen, "Lithium-ion battery storage for the grid—a review of stationary battery storage system design tailored for applications in modern power grids," *Energies*, vol. 10, no. 12, p. 2107, 2017.

[13] S. B. Sarmah, P. Kalita, A. Garg, X.-d. Niu, X.-W. Zhang, X. Peng, and D. Bhattacharjee, "A review of state of health estimation of energy storage systems: Challenges and possible solutions for futuristic applications of li-ion battery packs in electric vehicles," *Journal of Electrochemical Energy Conversion and Storage*, vol. 16, no. 4, p. 040801, 2019.

[14] X. Hu, F. Feng, K. Liu, L. Zhang, J. Xie, and B. Liu, "State estimation for advanced battery management: Key challenges and future trends," *Renewable and Sustainable Energy Reviews*, vol. 114, p. 109334, 2019.

[15] E. Vanem, C. B. Salucci, A. Bakdi, and Ø. Å. sheim Alnes, "Data-driven state of health modelling—a review of state of the art and reflections on applications for maritime battery systems," *Journal of Energy Storage*, vol. 43, p. 103158, 2021.

[16] M. Dubarry, G. Baure, and D. Anseán, "Perspective on state-of-health determination in lithium-ion batteries," *Journal of Electrochemical Energy Conversion and Storage*, vol. 17, no. 4, p. 044701, 2020.

[17] M. Schönleber, C. Uhlmann, P. Braun, A. Weber, and E. Ivers-Tiffée, "A consistent derivation of the impedance of a lithium-ion battery electrode and its dependency on the state-of-charge," *Electrochimica Acta*, vol. 243, pp. 250–259, 2017.

[18] G. Dos Reis, C. Strange, M. Yadav, and S. Li, "Lithium-ion battery data and where to find it," *Energy and AI*, vol. 5, p. 100081, 2021.

[19] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," *Joule*, vol. 4, no. 2, pp. 310–346, 2020.

[20] V. Sulzer, P. Mohtat, A. Aitio, S. Lee, Y. T. Yeh, F. Steinbacher, M. U. Khan, J. W. Lee, J. B. Siegel, A. G. Stefanopoulou, *et al.*, "The challenge and opportunity of battery lifetime prediction from field data," *Joule*, vol. 5, no. 8, pp. 1934–1955, 2021.

[21] M.-F. Ge, Y. Liu, X. Jiang, and J. Liu, "A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries," *Measurement*, vol. 174, p. 109057, 2021.

[22] X. Qu, D. Shi, J. Zhao, M.-K. Tran, Z. Wang, M. Fowler, Y. Lian, and A. F. Burke, "Insights and reviews on battery lifetime prediction from research to practice," *Journal of Energy Chemistry*, 2024.

[23] W. Dang, S. Liao, B. Yang, Z. Yin, M. Liu, L. Yin, and W. Zheng, "An encoder-decoder fusion battery life prediction method based on gaussian process regression and improvement," *Journal of Energy Storage*, vol. 59, p. 106469, 2023.

[24] D. Chen, W. Hong, and X. Zhou, "Transformer network for remaining useful life prediction of lithium-ion batteries," *Ieee Access*, vol. 10, pp. 19621–19628, 2022.

[25] T. Baumhöfer, M. Brühl, S. Rothgang, and D. U. Sauer, "Production caused variation in capacity aging trend and correlation to initial cell performance," *Journal of Power Sources*, vol. 247, pp. 332–338, 2014.

[26] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, *et al.*, "Data-driven prediction of battery cycle life before capacity degradation," *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.

[27] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, *et al.*, "Closed-loop optimization of fast-charging protocols for batteries with machine learning," *Nature*, vol. 578, no. 7795, pp. 397–402, 2020.

[28] S. Greenbank and D. Howey, "Automated feature extraction and selection for data-driven models of rapid battery capacity fade and end of life," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 2965–2973, 2021.

[29] A. Weng, P. Mohtat, P. M. Attia, V. Sulzer, S. Lee, G. Less, and A. Stefanopoulou, "Predicting the impact of formation protocols on battery lifetime immediately after manufacturing," *Joule*, vol. 5, no. 11, pp. 2971–2992, 2021.

[30] S. Saxena, L. Ward, J. Kubal, W. Lu, S. Babinec, and N. Paulson, "A convolutional neural network model for battery capacity fade curve prediction using early life data," *Journal of Power Sources*, vol. 542, p. 231736, 2022.

[31] N. H. Paulson, J. Kubal, L. Ward, S. Saxena, W. Lu, and S. J. Babinec, "Feature engineering for machine learning enabled early prediction of battery lifetime," *Journal of Power Sources*, vol. 527, p. 231127, 2022.

[32] A. Geslin, B. Van Vlijmen, X. Cui, A. Bhargava, P. A. Asinger, R. D. Braatz, and W. C. Chueh, "Selecting the appropriate features in battery lifetime predictions," *Joule*, vol. 7, no. 9, pp. 1956–1965, 2023.

[33] N. Yang, H. Hofmann, J. Sun, and Z. Song, "Remaining useful life prediction of lithium-ion batteries with limited degradation history using random forest," *IEEE Transactions on Transportation Electrification*, vol. 10, no. 3, pp. 5049–5060, 2023.

[34] M. A. Patil, P. Tagade, K. S. Hariharan, S. M. Kolake, T. Song, T. Yeo, and S. Doo, "A novel multistage support vector machine based approach for li ion battery remaining useful life estimation," *Applied energy*, vol. 159, pp. 285–297, 2015.

[35] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health," *Journal of Power Sources*, vol. 357, pp. 209–219, 2017.

[36] J. Wu, C. Zhang, and Z. Chen, "An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks," *Applied*

energy, vol. 173, pp. 134–140, 2016.

[37] N. He, Q. Wang, Z. Lu, Y. Chai, and F. Yang, "Early prediction of battery lifetime based on graphical features and convolutional neural networks," *Applied Energy*, vol. 353, p. 122048, 2024.

[38] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.

[39] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee, "Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning," *Nature communications*, vol. 11, no. 1, p. 1706, 2020.

[40] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A data-driven auto-cnn-lstm prediction model for lithium-ion battery remaining useful life," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3478–3487, 2020.

[41] 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*, vol. 113, p. 109254, 2019.

[42] O. Demirci, S. Taskin, E. Schaltz, and B. A. Demirci, "Review of battery state estimation methods for electric vehicles-part ii: Soh estimation," *Journal of Energy Storage*, vol. 96, p. 112703, 2024.

[43] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 3. state and parameter estimation," *Journal of Power sources*, vol. 134, no. 2, pp. 277–292, 2004.

[44] C. Hu, B. D. Youn, and J. Chung, "A multiscale framework with extended kalman filter for lithium-ion battery soc and capacity estimation," *Applied Energy*, vol. 92, pp. 694–704, 2012.

[45] Q. Yu, R. Xiong, R. Yang, and M. G. Pecht, "Online capacity estimation for lithium-ion batteries through joint estimation method," *Applied Energy*, vol. 255, p. 113817, 2019.

[46] A. Chu, A. Allam, A. C. Arenas, G. Rizzoni, and S. Onori, "Stochastic capacity loss and remaining useful life models for lithium-ion batteries in plug-in hybrid electric vehicles," *Journal of Power Sources*, vol. 478, p. 228991, 2020.

[47] D. Natella, S. Onori, and F. Vasca, "A co-estimation framework for state of charge and parameters of lithium-ion battery with robustness to aging and usage conditions," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 6, pp. 5760–5770, 2022.

[48] D. Zhang, S. Dey, H. E. Perez, and S. J. Moura, "Real-time capacity estimation of lithium-ion batteries utilizing thermal dynamics," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 3, pp. 992–1000, 2019.

[49] G. L. Plett, *Battery management systems, Volume II: Equivalent-circuit methods*. Artech House, 2015.

[50] X. Hu, S. Li, and H. Peng, "A comparative study of equivalent circuit models for li-ion batteries," *Journal of Power Sources*, vol. 198, pp. 359–367, 2012.

[51] J. Gomez, R. Nelson, E. E. Kalu, M. H. Weatherspoon, and J. P. Zheng, "Equivalent circuit model parameters of a high-power li-ion battery: Thermal and state of charge effects," *Journal of Power Sources*, vol. 196, no. 10, pp. 4826–4831, 2011.

[52] A. Smiley and G. L. Plett, "An adaptive physics-based reduced-order model of an aged lithium-ion cell, selected using an interacting multiple-model kalman filter," *Journal of Energy Storage*, vol. 19, pp. 120–134, 2018.

[53] M. Doyle, T. F. Fuller, and J. Newman, "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell," *Journal of the Electrochemical society*, vol. 140, no. 6, p. 1526, 1993.

[54] T. F. Fuller, M. Doyle, and J. Newman, "Simulation and optimization of the dual lithium ion insertion cell," *Journal of the electrochemical society*, vol. 141, no. 1, p. 1, 1994.

[55] J. C. Forman, S. J. Moura, J. L. Stein, and H. K. Fathy, "Genetic identification and fisher identifiability analysis of the doyle–fuller–newman model from experimental

cycling of a lifepo4 cell," *Journal of Power Sources*, vol. 210, pp. 263–275, 2012.

[56] A. Jokar, B. Rajabloo, M. Désilets, and M. Lacroix, "Review of simplified pseudo-two-dimensional models of lithium-ion batteries," *Journal of Power Sources*, vol. 327, pp. 44–55, 2016.

[57] S. K. Rahimian, S. Rayman, and R. E. White, "Extension of physics-based single particle model for higher charge–discharge rates," *Journal of Power Sources*, vol. 224, pp. 180–194, 2013.

[58] M. Guo, G. Sikha, and R. E. White, "Single-particle model for a lithium-ion cell: Thermal behavior," *Journal of The Electrochemical Society*, vol. 158, no. 2, p. A122, 2010.

[59] S. Santhanagopalan, Q. Guo, P. Ramadass, and R. E. White, "Review of models for predicting the cycling performance of lithium ion batteries," *Journal of power sources*, vol. 156, no. 2, pp. 620–628, 2006.

[60] A. M. Bizeray, J.-H. Kim, S. R. Duncan, and D. A. Howey, "Identifiability and parameter estimation of the single particle lithium-ion battery model," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 5, pp. 1862–1877, 2018.

[61] X. Hu, F. Sun, Y. Zou, and H. Peng, "Online estimation of an electric vehicle lithium-ion battery using recursive least squares with forgetting," in *Proceedings of the 2011 American Control Conference*, pp. 935–940, IEEE, 2011.

[62] G. L. Plett, "Recursive approximate weighted total least squares estimation of battery cell total capacity," *Journal of Power Sources*, vol. 196, no. 4, pp. 2319–2331, 2011.

[63] Y. Wang, C. Zhang, and Z. Chen, "A method for joint estimation of state-of-charge and available energy of lifepo4 batteries," *Applied energy*, vol. 135, pp. 81–87, 2014.

[64] N. Wassiliadis, J. Adermann, A. Frericks, M. Pak, C. Reiter, B. Lohmann, and M. Lienkamp, "Revisiting the dual extended kalman filter for battery state-of-

charge and state-of-health estimation: A use-case life cycle analysis," *Journal of Energy Storage*, vol. 19, pp. 73–87, 2018.

[65] C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diagnostics for lithium ion cells," *Journal of Power Sources*, vol. 341, pp. 373–386, 2017.

[66] T. Raj, A. A. Wang, C. W. Monroe, and D. A. Howey, "Investigation of path-dependent degradation in lithium-ion batteries," *Batteries & Supercaps*, vol. 3, no. 12, pp. 1377–1385, 2020.

[67] J. M. Reniers, G. Mulder, and D. A. Howey, "Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries," *Journal of The Electrochemical Society*, vol. 166, no. 14, pp. A3189–A3200, 2019.

[68] L. Zhang, Z. Mu, and C. Sun, "Remaining useful life prediction for lithium-ion batteries based on exponential model and particle filter," *Ieee Access*, vol. 6, pp. 17729–17740, 2018.

[69] Y. Xing, E. W. Ma, K.-L. Tsui, and M. Pecht, "A case study on battery life prediction using particle filtering," in *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing)*, pp. 1–6, IEEE, 2012.

[70] J. Wang, P. Liu, J. Hicks-Garner, E. Sherman, S. Soukiazian, M. Verbrugge, H. Tataria, J. Musser, and P. Finamore, "Cycle-life model for graphite-lifepo4 cells," *Journal of power sources*, vol. 196, no. 8, pp. 3942–3948, 2011.

[71] M. Petit, E. Prada, and V. Sauvant-Moynot, "Development of an empirical aging model for li-ion batteries and application to assess the impact of vehicle-to-grid strategies on battery lifetime," *Applied energy*, vol. 172, pp. 398–407, 2016.

[72] M. Lucu, E. Martinez-Laserna, I. Gandiaga, and H. Camblong, "A critical review on self-adaptive li-ion battery ageing models," *Journal of Power Sources*, vol. 401, pp. 85–101, 2018.

[73] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, "One-shot battery degradation trajectory prediction with deep learning," *Journal*

of Power Sources, vol. 506, p. 230024, 2021.

[74] J. Zhao and Z. Wang, "Specialized convolutional transformer networks for estimating battery health via transfer learning," *Energy Storage Materials*, vol. 71, p. 103668, 2024.

[75] S. J. Harris, D. J. Harris, and C. Li, "Failure statistics for commercial lithium ion batteries: A study of 24 pouch cells," *Journal of Power Sources*, vol. 342, pp. 589–597, 2017.

[76] X. Liu, W. Ai, M. N. Marlow, Y. Patel, and B. Wu, "The effect of cell-to-cell variations and thermal gradients on the performance and degradation of lithium-ion battery packs," *Applied Energy*, vol. 248, pp. 489–499, 2019.

[77] S. Rothgang, T. Baumhofer, and D. U. Sauer, "Diversion of aging of battery cells in automotive systems," in *2014 IEEE vehicle power and propulsion conference (VPPC)*, pp. 1–6, IEEE, 2014.

[78] M. Lucu, E. Martinez-Laserna, I. Gandiaga, K. Liu, H. Camblong, W. D. Widanage, and J. Marco, "Data-driven nonparametric li-ion battery ageing model aiming at learning from real operation data-part b: Cycling operation," *Journal of Energy Storage*, vol. 30, p. 101410, 2020.

[79] A. Aitio and D. A. Howey, "Predicting battery end of life from solar off-grid system field data using machine learning," *Joule*, vol. 5, no. 12, pp. 3204–3220, 2021.

[80] A. Aitio, D. Jöst, D. U. Sauer, and D. A. Howey, "Learning battery model parameter dynamics from data with recursive gaussian process regression," *Journal of Dynamic Systems, Measurement, and Control*, vol. 147, no. 3, p. 031010, 2025.

[81] S. Evro, A. Ajumobi, D. Mayon, and O. S. Tomomewo, "Navigating battery choices: A comparative study of lithium iron phosphate and nickel manganese cobalt battery technologies," *Future Batteries*, vol. 4, p. 100007, 2024.

[82] J. Nájera, J. Arribas, R. De Castro, and C. Núñez, "Semi-empirical ageing model for lfp and nmc li-ion battery chemistries," *Journal of Energy Storage*, vol. 72, p. 108016, 2023.

[83] M. Weiss, R. Ruess, J. Kasnatscheew, Y. Levartovsky, N. R. Levy, P. Minnmann, L. Stolz, T. Waldmann, M. Wohlfahrt-Mehrens, D. Aurbach, *et al.*, "Fast charging of lithium-ion batteries: a review of materials aspects," *Advanced Energy Materials*, vol. 11, no. 33, p. 2101126, 2021.

[84] X. Lin and A. G. Stefanopoulou, "Analytic bound on accuracy of battery state and parameter estimation," *Journal of The Electrochemical Society*, vol. 162, no. 9, p. A1879, 2015.

A

Appendix of Chapter 2

Cell Reports Physical Science, Volume *5*

Supplemental information

Predicting battery lifetime under varying usage conditions from early aging data

Tingkai Li, Zihao Zhou, Adam Thelen, David A. Howey, and Chao Hu

Note S1 Cell Specifications

The Li-ion cells used in this study were 502030-size Li-polymer cells manufactured by Hong-haosheng (HHS) Electronics in Shenzhen, China. The design and dimensions of these cells are shown in Fig. S1. The specifications provided by the vendor are shown below.

- Nominal capacity: 250 mAh
- Nominal voltage: 3.7 V
- Fully charged voltage: 4.2 V
- Fully discharged voltage: 3.0 V

Figure S1: 502030 pouch cell used in this study, annotated with its dimension.

Note S2 Detailed Battery Test Conditions

After completing the aging tests on a small batch of cells, we found the capacity loss during the first week of cycling was extremely large. To better capture the capacity fade trend in this region during future tests, we added an extra RPT after 0.5 weeks of cycling. Starting from G20, we performed an additional RPT at week 0.5. This note is important since it affects feature extraction methods that use numerical indexing to select the appropriate data.

By the time this manuscript was prepared, G49 and G57 hadn't had any cells reach the end of life. So, these two groups of cells were not included in this study, but the data will be available with other groups in this dataset. G15 is omitted from the dataset because of severe degradation within the first week of cycling and cells were removed for a safety concern. For the G15 condition listed below, the DoD value is the designed (ideal) value.

Table S1: Cycling conditions of all groups

Group #	Charging C-rate	Discharging C-rate	Mean DoD	Mean Lifetime [weeks]
1	0.500	0.500	25.9%	49.27
2	0.500	0.500	46.1%	26.49
3	0.500	0.500	67.8%	23.71
4	1.000	0.500	47.3%	26.71

Continued on next page

Table S1: Cycling conditions of all groups (Continued)

Group #	Charging C-rate	Discharging C-rate	Mean DoD	Mean Lifetime [weeks]
5	1.000	0.500	97.5%	12.45
6	2.000	0.500	27.2%	42.92
7	2.000	0.500	46.9%	26.93
8	2.000	0.500	67.2%	17.43
9	2.000	0.500	96.9%	6.02
10	2.500	0.500	47.2%	18.58
12	3.000	0.500	28.6%	17.09
13	3.000	0.500	47.3%	14.78
14	3.000	0.500	66.3%	5.39
15	3.000	0.500	100.0%	
16	0.500	0.500	97.6%	16.11
17	1.000	0.500	67.5%	17.82
18	2.500	0.500	27.5%	25.79
19	2.500	0.500	65.7%	8.59
20	0.800	0.500	98.2%	12.07
21	1.200	0.500	98.0%	13.32
22	1.400	0.500	97.9%	9.85
23	1.600	0.500	97.6%	11.03
24	1.800	0.500	97.5%	10.27
25	1.800	0.600	51.4%	18.45
26	1.400	2.200	4.2%	39.79
27	0.600	2.400	57.8%	15.06
28	2.400	1.600	81.0%	9.03
29	1.600	1.800	52.5%	15.05
30	0.800	0.800	77.6%	13.31
31	1.200	1.000	61.4%	11.83
32	1.000	1.400	38.9%	18.72
33	2.000	1.200	40.6%	17.77
34	2.200	2.000	27.6%	21.40
35	1.825	0.500	97.0%	10.67
36	2.075	0.500	97.1%	10.23
37	0.725	0.500	98.1%	15.32
38	1.875	0.500	95.8%	10.62
39	1.475	0.500	97.1%	11.53
40	1.825	1.025	31.9%	28.97
41	2.075	1.775	67.0%	14.33

Continued on next page

Table S1: Cycling conditions of all groups (Continued)

Group #	Charging C-rate	Discharging C-rate	Mean DoD	Mean Lifetime [weeks]
42	0.725	2.375	36.3%	18.67
43	1.875	2.325	38.0%	20.21
44	0.775	1.275	37.1%	19.25
45	1.125	1.725	80.3%	8.44
46	1.225	2.025	75.4%	8.84
47	2.325	1.925	48.3%	11.81
48	2.375	2.225	58.6%	9.66
49	0.975	0.675	18.8%	
50	2.425	1.625	20.1%	20.77
51	2.275	1.875	53.6%	11.80
52	1.425	0.875	83.5%	15.63
53	2.025	0.825	91.5%	5.72
54	0.925	1.125	27.1%	27.19
55	1.025	2.475	61.4%	12.48
56	2.175	0.975	53.8%	18.46
57	1.775	1.175	17.4%	
58	2.475	0.575	42.0%	19.36
59	1.325	1.825	31.6%	25.78
60	0.675	1.325	83.6%	10.67
61	2.125	1.975	54.1%	11.67
62	1.575	2.425	64.2%	11.54
63	1.975	1.675	80.6%	10.12
64	1.175	1.425	91.4%	8.16

Note S3 Cycling Parameters vs. Lifetime

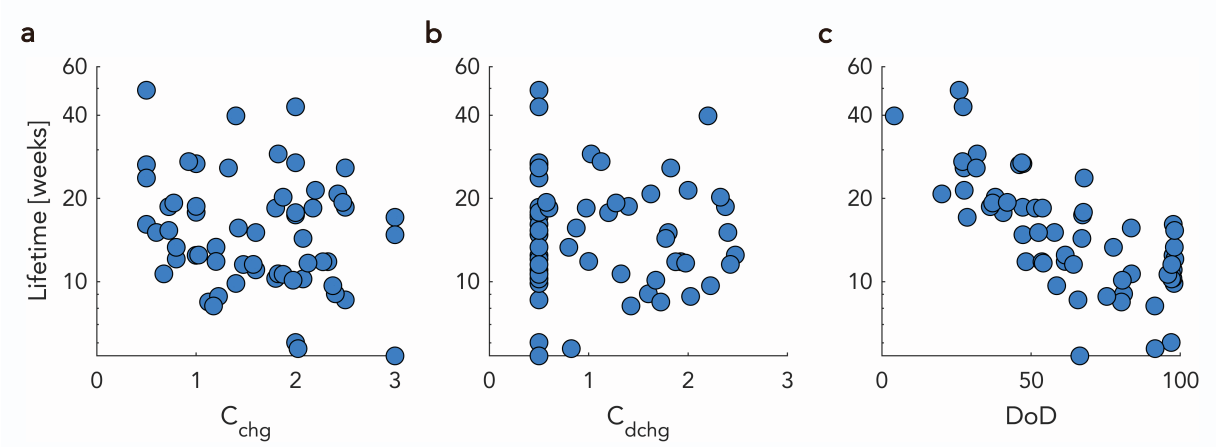


Figure S2: Group mean lifetime plotted as a function of **a**, C_{chg}; **b**, C_{dchg}; **c**, DoD.

Note S4 Initial Capacity vs. Lifetime

An analysis is done to check if initial capacity affects the lifetime. Fig. S3a shows the distribution of the initial capacity of our dataset. To investigate if the initial capacity has an impact on the lifetime, we plot the lifetime vs. the initial capacity, as shown in Fig. S3b. There is no clear trend between the initial capacity and lifetime, even though some extremely long-lifetime cells have relatively low initial capacities.

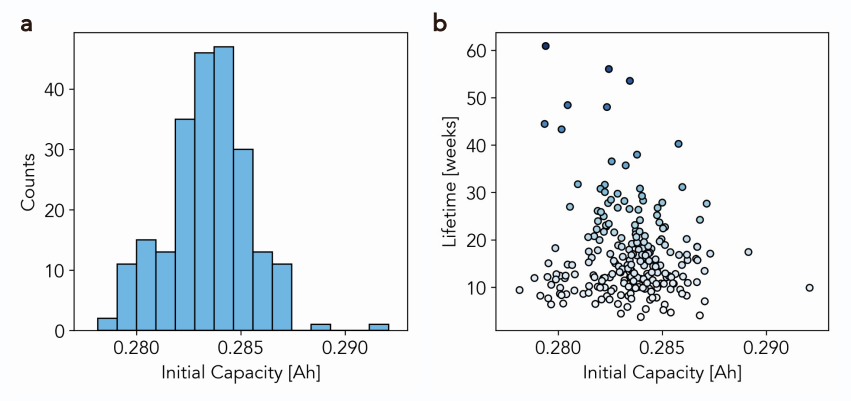


Figure S3: Analysis of the initial capacity. **a,** Histogram of the initial capacity of the dataset. **b,** Lifetime plotted as a function of initial capacity.

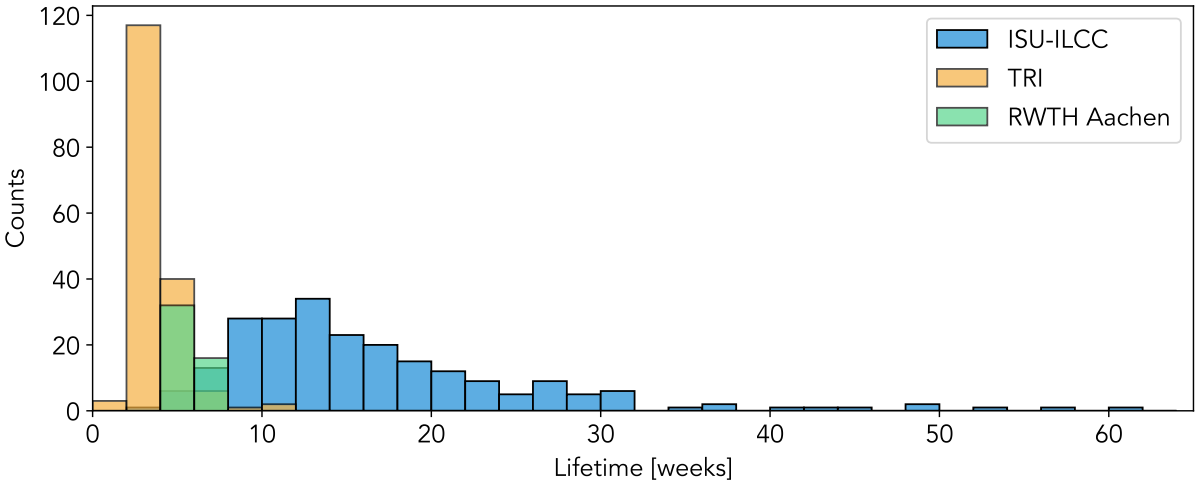


Figure S4: Histograms of cell lifetime from the ISU-ILCC dataset of this work, the 169 LFP dataset from Toyota Research Institute[1,2], and the 48 NMC dataset from RWTH Aachen[3,4].

Note S5 Comparison of Lifetime Distributions with Existing Datasets

To showcase the uniqueness of this new ISU-ILCC dataset, especially for lifetime modeling studies, we compare the lifetime distribution of this dataset with those of several other datasets widely used in the battery modeling literature. The histograms compared in Fig. S4 are from the ISU-ILCC dataset in this work, the dataset with 169 LFP/Gr cells from Toyota Research Institute[1,2], and the dataset with 48 NMC/Gr cells from RWTH Aachen University[3]. To better compare across all datasets, we calculate the lifetime of all datasets in the unit of weeks. Note that the 48 NMC/Gr dataset from RWTH Aachen was collected under identical aging conditions across all 48 cells, as the original study focused on cell-to-cell manufacturing variability[3]. As a result, the 48 cells exhibited very small differences in their lifetime.

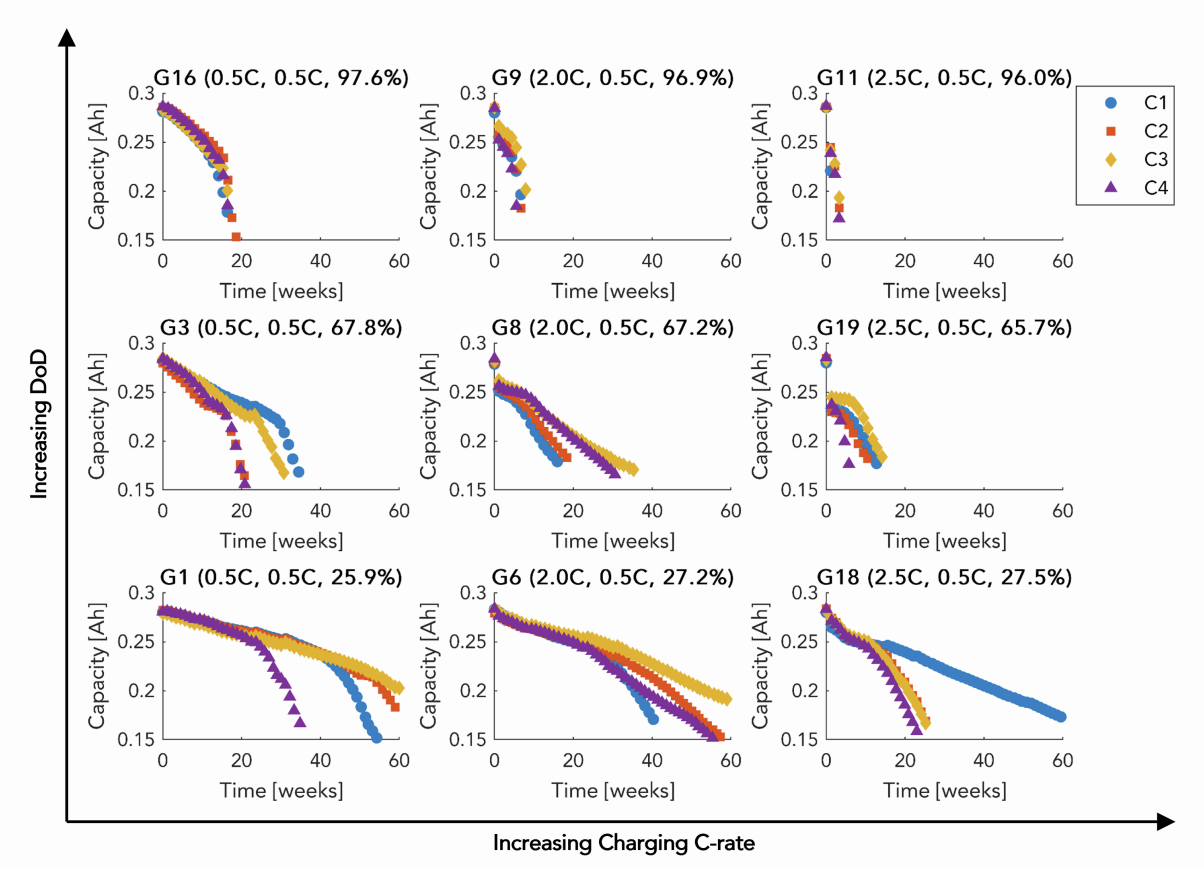


Figure S5: Example capacity fade trajectories for groups cycled under different charging C-rates and DoDs. The values inside parentheses indicate charging C-rate, discharging C-rate, and mean DoD, respectively.

Note S6 Overview of Battery Aging under Varying Conditions

To showcase the many unique capacity-fade trajectories present in the dataset, we plot capacity-fade curves from groups of cells whose cycling conditions make up a complete grid spanning a range of charge C-rates and depths of discharge. This subset of 9 groups of cells, shown in Fig. S5, were cycled with different charging rates and depths of discharge but a constant discharge C-rate of 0.5C in all cases. We observe that groups with high charging rates and moderate-to-low depth of discharge (e.g., G8, G18, G19) experienced three-stage capacity fade. Their capacity initially decreases quickly, then stabilizes into a slower linear fade, and then accelerates again towards the end of life. More frequently, we observe a two-stage capacity fade trend from cells in some groups (e.g. G1, G3, G6, G16). However, in a few cases, we also observe a one-stage capacity trend for cells cycled at full depth of discharge and high charging C-rates (e.g., G9, G11). Our dataset's diverse capacity degradation trajectories make early-life feature engineering challenging because cells experiencing rapid capacity fade during the first few weeks of aging can sometimes end up having moderately long lifetimes. For example, G18 cells in Fig. S5 show rapid capacity fade during the first few weeks of cycling but eventually had lifetimes greater than 20 weeks. On the other hand, G16 cells show much slower capacity fade during the first few weeks but have lifetimes of less than 20 weeks.

Additionally, we observe considerable in-group lifetime variation. Groups G1, G6, and G18 in Fig. S5 show a large variation in lifetime for cells operating under the same test conditions. Cell aging

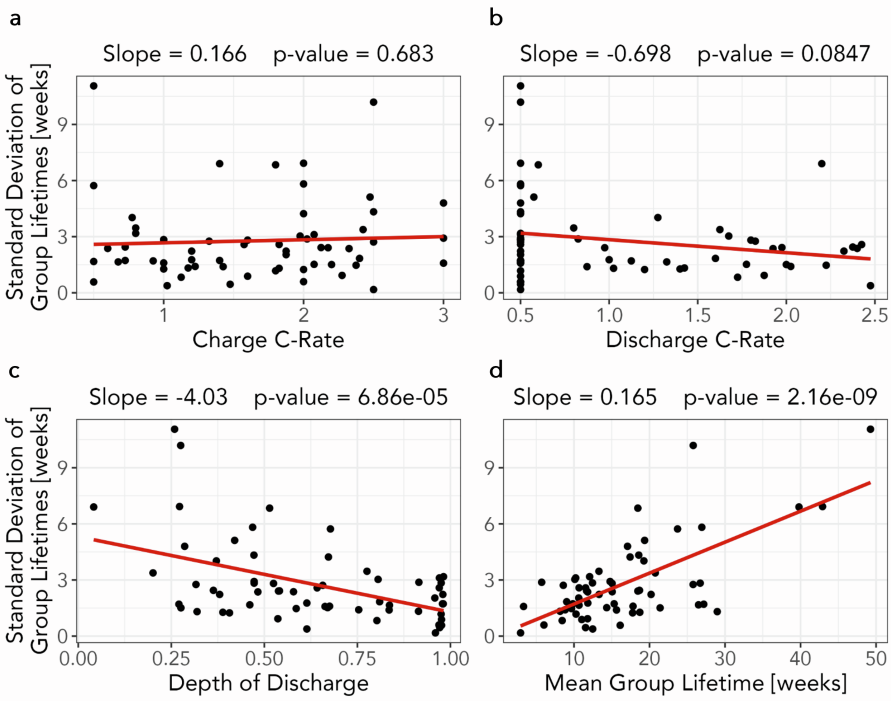


Figure S6: Depth of discharge has a strong impact on lifetime variability. Here, the standard deviation of group lifetimes is plotted vs. **a,** charging C-rate, **b,** discharging C-rate, **c,** depth of discharge, **d,** the mean group lifetime. Smaller p-values indicate greater statistical significance of the fitted value of the slope term in the regression fit.

variability can be caused by testing equipment inaccuracies, manufacturing variations, and even internal defects, and is highly undesirable when designing battery-powered products. We conducted a statistical analysis to elucidate the relationship between the three varying aging stress factors (i.e., C_{chg}, C_{dchg}, and DoD) and lifetime variability. We calculated the in-group standard deviation of cell lifetime as a function of each aging stress factor, as well as the mean group lifetime, Fig. S6. This reveals that only the depth of discharge (Fig. S6c) has a statistically significant relationship with the observed cell-to-cell lifetime variability. However, we also observe that cells with longer lifetimes have higher lifetime variability (Fig. S6d). These two results make it difficult to determine the true source of lifetime variability—it might result either from shallow depths of discharge or increased cell lifetime.

Note S7 Validation of $\Delta Q(V)$ Features

Fig. S7 shows the $dV/dQ(Q)$ curve evolution of two cells highlighted in the main text for comparison, both having very similar $\mathrm{var}(\Delta Q(V)_{\mathrm{w3-w0}})$ values.

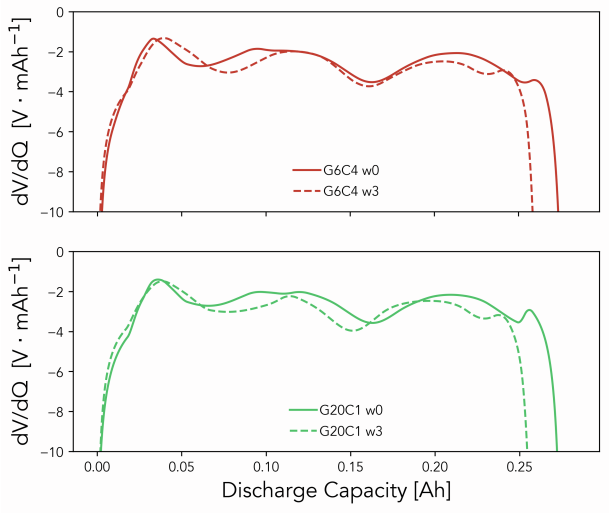


Figure S7: $dQ/dV(V)$ curve changes between weeks 3 and 0 of two highlighted cells in the main text, which have similar $\mathrm{var}(\Delta Q(V)_{\mathrm{w3-w0}})$ values.

To further understand the reason behind the unsatisfactory performance of the well-known feature extracted from $\Delta Q(V)$ curves, we plot the $\Delta Q(V)_{\mathrm{w3-wo}}$ curves of the ISU-ILCC dataset (see Fig. S8a), along with the $\Delta Q(V)_{100-10}$ from the TRI dataset[1,2] (see Fig. S8b). Apparently, the maximum difference exists at the lower cutoff voltage ($V = 3.0\,\mathrm{V}$), which also reflects the total capacity fade between weeks 0 and 3. So, as shown in Fig. S8c, the $\mathrm{var}(\Delta Q(V)_{\mathrm{w3-w0}})$ correlates with the capacity fade. Both phenomena indicate that the adapted feature on our dataset mainly captures the capacity loss instead of the information related to early-stage degradation mechanisms.

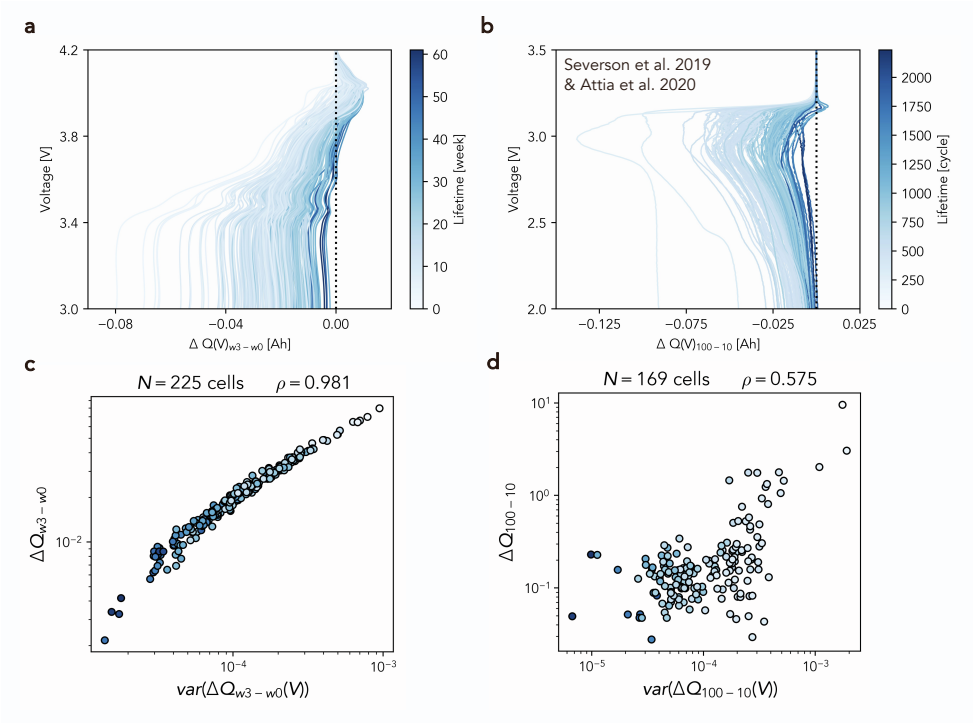


Figure S8: An overview of $\Delta Q(V)$ curves from cells in two aging datasets. **a**, the ISU-ILCC battery aging dataset. **b**, the dataset of Severson et al. and Attia et al.[1,2].**c**, $\mathrm{var}\left(\Delta Q(V)_{\mathrm{w3-w0}}\right)$ of the ISU-ILCC dataset. **d**, $\mathrm{var}\left(\Delta Q(V)_{100-10}\right)$ of the dataset of Severson et al. and Attia et al.[1,2].

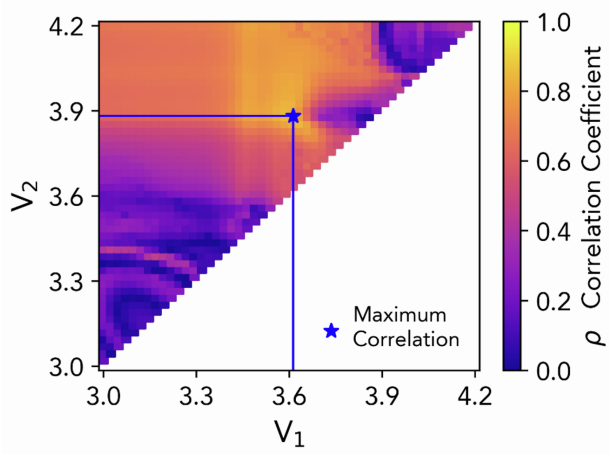


Figure S9: The grid search result for the optimal voltage interval to extract feature from $dQ/dV(V)$.

Note S8 Analysis of Incremental Capacity Feature

By using an exhaustive grid search with a minimum window size of 0.02 V, the result of finding the optimal voltage interval is shown in Fig. S9. From this heatmap, we can clearly observe that the feature extracted from voltage intervals around the major peak of the $dQ/dV(V)$ curve has noticeably stronger correlations with lifetime.

To clarify the relationship between the peaks in the differential voltage curve and cell health, we constructed half-cells from electrode materials obtained from disassembling a fresh cell. We cycled the half-cells at a slow rate (C/20) and reconstructed a full-cell pseudo-open circuit voltage curve. However, the negative electrode data is poor because half-cell assembly was challenging. During assembly, we had to remove a water-soluble coating covering the negative electrode material by scratching it off, as using solvents would have damaged it. This process is inexact, and it produced poor electrode material, which then yielded poor results during cycling. Thus, we were unable to attribute peaks on incremental capacity curves to the specific side of electrodes. Future work on exploring degradation mechanisms in this dataset and the relationship between the early life features and the dominant degradation modes will help us determine which electrode this feature corresponds to.

The remaining unexplained variance in the new feature-lifetime correlation is likely due to the unavoidable influence of a decreasing lithium inventory on the shape of the $dQ/dV(V)$ curves. Decreases in lithium inventory can cause shifts in the voltages where peaks occur[5]. This causes a small misalignment between the curves at weeks three and zero that varies cell-to-cell and introduces variation in the incremental capacity feature extraction. Destructively analyzing specific cells from the dataset would help to determine more concretely what the new feature $\text{mean}\left(\Delta dQ/dV_{\text{w3}-\text{w0}}^{3.60\text{V}-3.90\text{V}}(V)\right)$ is capturing, but this was outside the current scope.

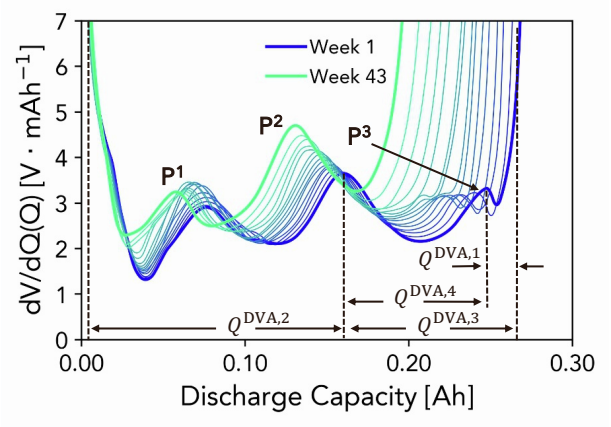


Figure S10: An overview of the peak identification and tracking method for differential voltage feature engineering.

Note S9 Differential Voltage Features

Like incremental capacity, differential voltage $(dV/dQ(Q))$ analysis can effectively diagnose different component-level degradation modes in Li-ion cells[6,7,8]. However, differential voltage analysis has yet to be widely used as part of automated feature extraction methods because curve manipulation and automatic peak detection are challenging. Unlike incremental capacity, the differential voltage is defined as a function of cell capacity, which can change cycle-to-cycle. The changing capacity makes curve manipulation and feature extraction via vector operations more difficult, as any two curves will not be the same length. Furthermore, the peaks and valleys of cells experiencing fast degradation often merge, confusing maxima and minima detection algorithms.

Despite these challenges, we investigated extracting four capacity-based features from differential voltage curves. The four features, $Q^{\mathrm{DVA},1}$ to $Q^{\mathrm{DVA},4}$ in Fig. S10, are designed to capture the evolution of the differential voltage curve during early life and are derived from the locations of peaks. The features capture the rate of change of different capacities and the relative shifts in the differential voltage curves, calculated as $\Delta Q^{\mathrm{DVA},1}_{\mathrm{w3-w0}} = Q^{\mathrm{DVA},1}_{\mathrm{w3}} - Q^{\mathrm{DVA},1}_{\mathrm{w0}}$.

The four differential voltage features are designed to quantify capacity losses attributed to each electrode and capture shifts in the relative electrode balancing[9]. Keil et al.[9] suggest certain capacities can be estimated to determine the change in electrode balancing and the loss of active materials at the positive and negative electrodes ($\mathrm{LAM_{PE}}$ and $\mathrm{LAM_{NE}}$, respectively). A change in the cathode capacity is captured through $Q^{\mathrm{DVA},2}$ since all the features of interest in this range are cathode specific. Similarly, the anode capacity is captured through $Q^{\mathrm{DVA},3}$. The different balances of the two electrodes are captured through $Q^{\mathrm{DVA},1}$ and $Q^{\mathrm{DVA},4}$, tracking the anode and cathode peaks, respectively. Each of the four features is included in the feature selection process.

Note S10 Capacity Fade and Constant Voltage Charging Time Features

In addition to features extracted from capacity-voltage curves and their derivatives, we derive a set of features from direct cell measurements of time and capacity. A benefit of these features is that they can be achieved using lower sampling frequency, measurement precision, and less data processing than the aforementioned curve difference features, making them suitable for implementation on battery health monitoring devices. The first feature extracted is the time spent in the constant-voltage (CV) charging step during each RPT, denoted $\mathrm{CV\ Time}_{wi}$. We also calculate the difference between two weeks' constant-voltage charging times, denoted $\Delta\mathrm{CV\ Time}_{\mathrm{w3-w0}}$. A panel plot illustrating the extraction of these features and their correlation with cell lifetimes is included in the Supplementary Information. During charging, the constant-voltage charging step occurs as the final stage in charging. Data collected from CV charging steps have successfully been used to estimate the state of health of Li-ion batteries in recent literature[10,11,12]. The extracted CV features reflect the interaction between capacity loss (decreasing the overall charging time) and increasing resistance to intercalation due to the degradation of the active electrode material.

Additionally, we extracted features from the discharge capacity in RPTs, such as the cells' initial capacity Q_{w0} and the capacity fade between weeks three and zero $\Delta Q_{\mathrm{w3-w0}}$, capturing the initial state of the cell and its relative change during early life, respectively.

Note S11 Feature Extraction Details

The following table contains all 29 extracted features and their Pearson (linear) correlation coefficients with $\log(lifetime)$. Additionally, we mark the specific features selected for use in the degradation-informed models using the values "2" and "3", corresponding to the elastic net and HBM models built using two and three features, respectively. Figure S11 shows a correlation matrix containing the correlation coefficients between the extracted features, color-coded by the direction and strength of the correlation. This heatmap clearly shows that many features have strong collinearity, suggesting feature selection is necessary.

Table S2: All extracted early-life features. Includes both condition-level and cell-level features.

Feature	ρ with log(lifetime)	Selected for Elastic Net	Selected for HBM
C_{chg}	-0.178		
C_{dchg}	-0.086		
DoD	-0.682	3	Cell-level: 3
${C_{\mathrm{chg}}}^{0.5}DoD^{0.5}$	-0.689		
${C_{\mathrm{dchg}}}^{0.5}DoD^{0.5}$	-0.543		
$({C_{\mathrm{chg}}}^{0.5}DoD^{0.5} + {C_{\mathrm{dchg}}}^{0.5}DoD^{0.5})/2$	-0.784		Clustering Variable: 2, 3
${C_{\mathrm{chg}}}^{0.5}DoD^{0.5} \times {C_{\mathrm{dchg}}}^{0.5}DoD^{0.5}$	-0.733		
$\log(\lvert\mathrm{mean}(\Delta dQ/dV_{\mathrm{w3-w0}}(V)\rvert)$	-0.668		
$\log(\lvert\mathrm{var}(\Delta dQ/dV_{\mathrm{w3-w0}}(V)\rvert)$	-0.634		
$\log(\lvert\mathrm{mean}(\Delta dQ/dV_{\mathrm{w3-w0}}^{3.0V-3.6V}(V)\rvert)$	-0.143		
$\log(\lvert\mathrm{mean}(\Delta dQ/dV_{\mathrm{w3-w0}}^{3.6V-3.9V}(V)\rvert)$	-0.848	2, 3	Cell-level: 2, 3
$\log(\lvert\mathrm{mean}(\Delta dQ/dV_{\mathrm{w3-w0}}^{3.9V-4.2V}(V)\rvert)$	-0.097		
$\log(\lvert\mathrm{var}(\Delta dQ/dV_{\mathrm{w3-w0}}^{3.0V-3.6V}(V)\rvert)$	-0.367		
$\log(\lvert\mathrm{var}(\Delta dQ/dV_{\mathrm{w3-w0}}^{3.6V-3.9V}(V)\rvert)$	-0.315		
$\log(\lvert\mathrm{var}(\Delta dQ/dV_{\mathrm{w3-w0}}^{3.9V-4.2V}(V)\rvert)$	-0.482		
$\log(\lvert\mathrm{mean}(\Delta Q_{\mathrm{w3-w0}}(V)\rvert)$	-0.716		
$\log(\lvert\mathrm{var}(\Delta Q_{\mathrm{w3-w0}}(V)\rvert)$	-0.686		
$\log(\mathrm{CV\ Time}_{\mathrm{w0}})$	0.223		
$\log(\mathrm{CV\ Time}_{\mathrm{w3}})$	-0.141		
$\log(\lvert\Delta\mathrm{CV\ Time}_{\mathrm{w3-w0}}\rvert)$	0.369	2, 3	Cell-level: 2, 3
$\log(Q_{\mathrm{w0}})$	-0.048		
$\log(Q_{\mathrm{w0}}^{3.0V-3.6V})$	0.241		
$\log(Q_{\mathrm{w0}}^{3.6V-3.9V})$	0.103		
$\log(Q_{\mathrm{w0}}^{3.9V-4.2V})$	-0.278		
$\log(\Delta Q_{\mathrm{w3-w0}})$	-0.668		
$\Delta Q_{\mathrm{w3-w0}}^{1}$	0.102		
$\Delta Q_{\mathrm{w3-w0}}^{2}$	-0.603		
$\Delta Q_{\mathrm{w3-w0}}^{3}$	-0.470		
$\Delta Q_{\mathrm{w3-w0}}^{4}$	-0.409		

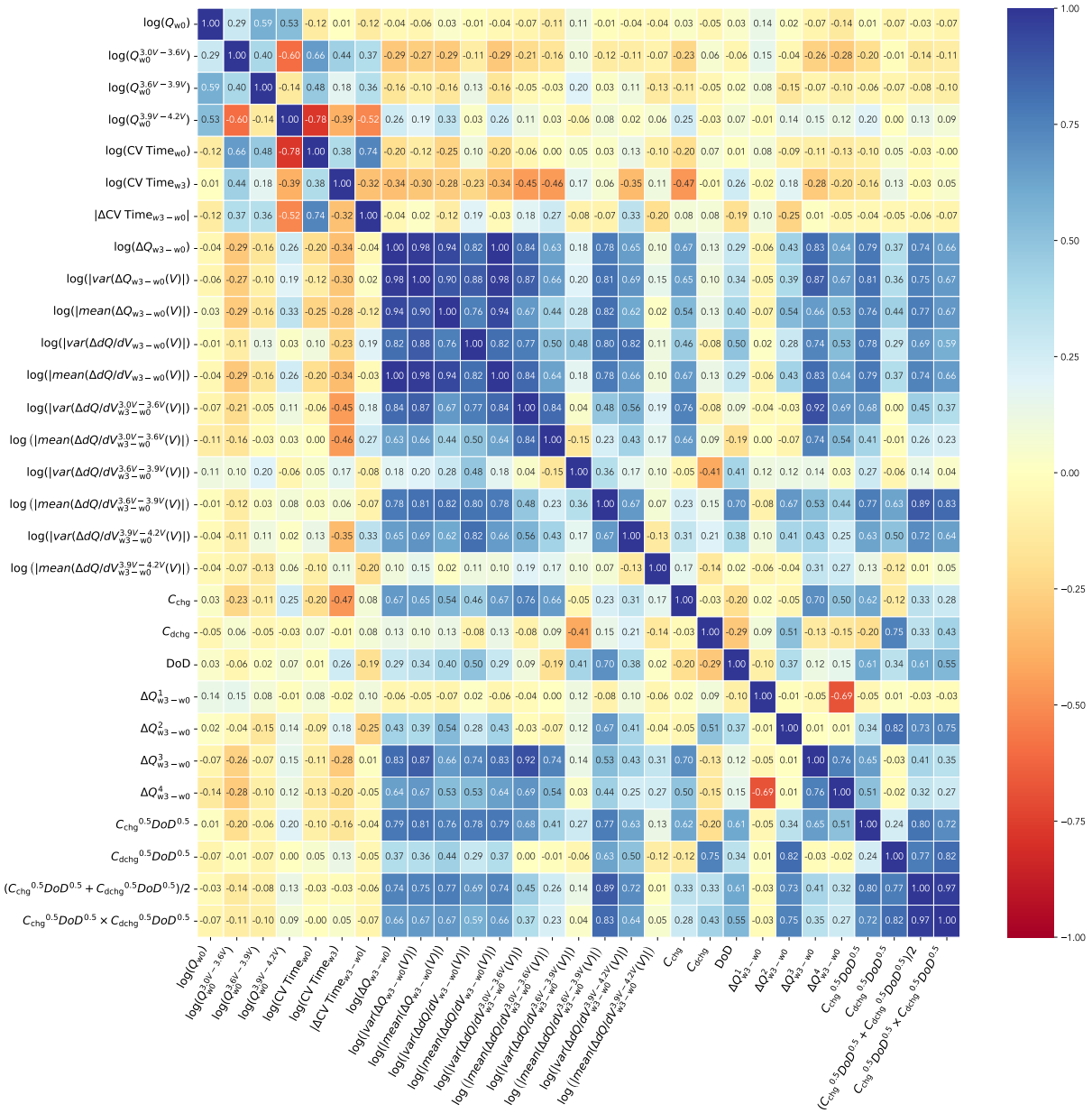


Figure S11: Corrlation matrix of all 29 extracted early-life features.

Note S12 Selected Feature Details

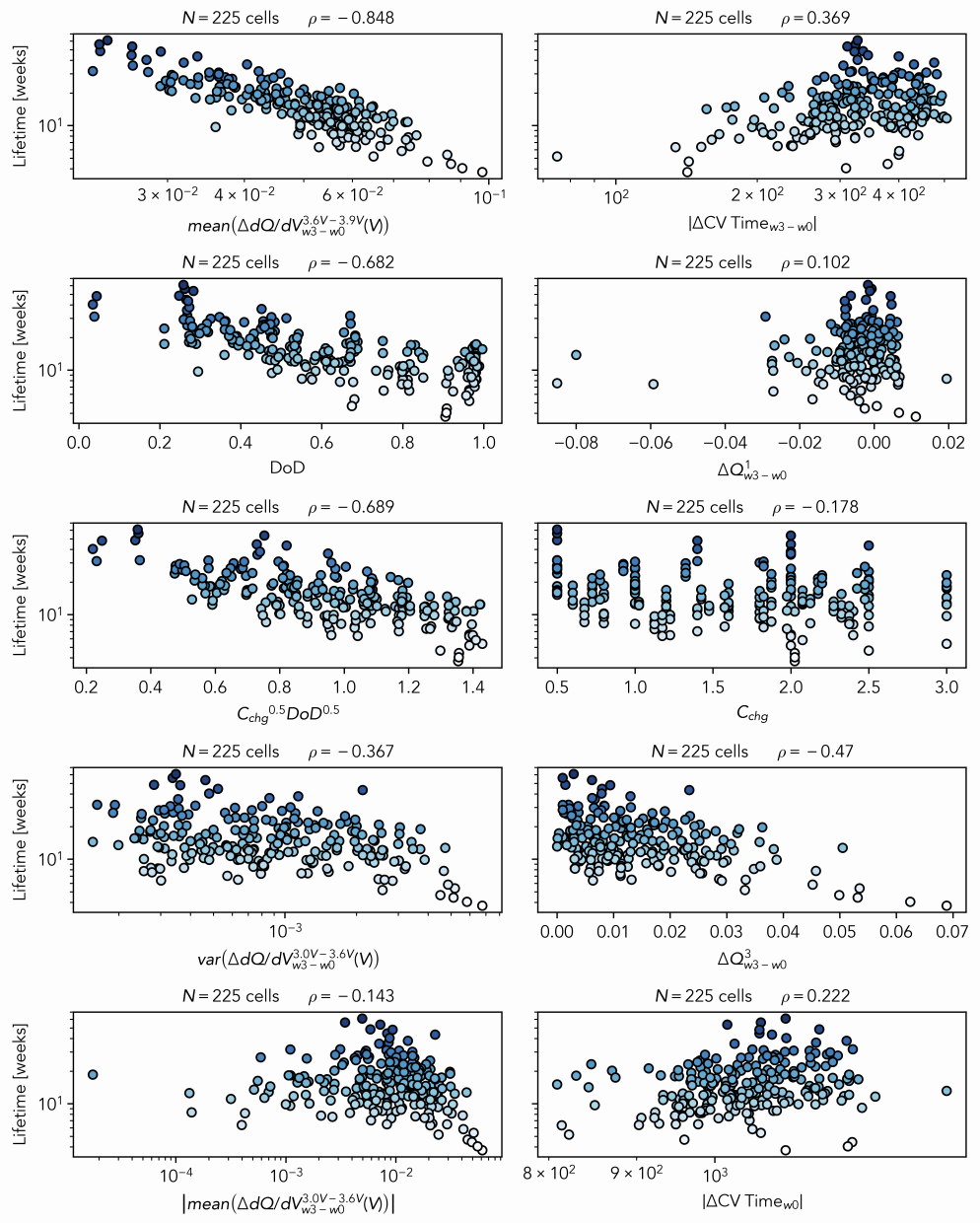


Figure S12: An overview of the 10 selected features based on the results from the step-wise feature selection technique.

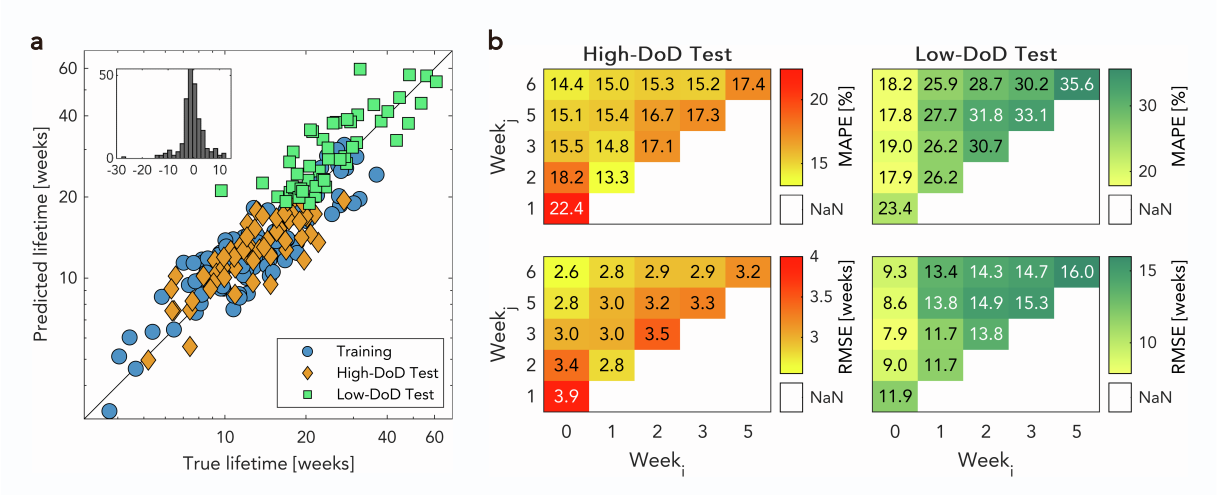


Figure S13: Overview of prediction results for *degradation-informed* elastic net model with two input features. **a**, True and predicted lifetimes using features extracted from weeks three and zero $(\mathrm{w}_3 - \mathrm{w}_0)$ with embedded histogram showing prediction residuals. **b**, The $\mathrm{RMSE}_{\mathrm{EOL}}$ and $\mathrm{MAPE}_{\mathrm{EOL}}$ error metrics as a function of week numbers from which the early-life features, denoted $(w_j - w_i)$. Week four data is omitted from the study due to a testing error.

Note S13 Time Dependence of Input Features

Motivated by the need to predict cell lifetime as early as possible, we performed a study varying the time frame from which the features are extracted to understand the impact on model accuracy. Unfortunately, a testing error during week four caused irreversible data loss for a large batch of cells, so week four data is omitted from this study. Using the degradation-informed elastic net model with two features, we vary the RPTs from which the features are extracted and record the test errors for the high and low-DoD test datasets. The results are shown in Fig. S13b.

First, we analyze the accuracy trend with the starting week fixed to week zero (i.e., $w_i = 0$). Under this setting, the prediction errors on the high-DoD dataset consistently decrease as the time between RPTs increases. However, the prediction errors for the low-DoD test set are found to slightly increase with increasing time-frame around weeks five and six $w_j = 5, 6$. This is likely because many cells experience rapid degradation after week five/six, which alters the feature-lifetime relationship for cells with short lives. This causes the model to change its fit, decreasing its prediction accuracy on long-lifetime cells.

Second, we analyze the accuracy trend by looking at the time between any two RPTs. Along the diagonal, the delta between any two RPTs is one week. Under these conditions, we observe a substantial increase in model prediction error on both the high- and low-DoD test sets compared with counterparts toward the upper left corner (i.e., models with features extracted with intervals longer than one week). This suggests a minimum time interval of $(w_j - w_i) \geq 2$ is required to accurately estimate the rate of degradation inside the cell from early-life features.

Finally, we observe that the model prediction error on the low-DoD test set continuously increases with increasing starting week w_i. This could be an effect of optimizing the incremental capacity feature using data from weeks three and zero. The optimal voltage interval for this feature may change with the RPTs used and was not accounted for in this study.

Note S14 Training Process of Hierarchical Bayesian Model

The overall training process follows:

- Estimate level-2 posterior distribution by Bayes's rule

$$P\left(\boldsymbol{\gamma}, \sigma \mid \{Y\}\right) \propto \prod_{j=1}^{J} P\left(Y_j \mid \boldsymbol{\gamma}, \sigma\right) \cdot P(\boldsymbol{\gamma}, \sigma)$$

- Use level-2 posterior distribution as prior for level-1 parameters, calculate level-1 posterior distribution

$$P\left(\boldsymbol{\theta_j}, \sigma_j \mid \boldsymbol{\gamma}, \sigma, Y_j\right) = \frac{P\left(Y_j \mid \boldsymbol{\theta_j}, \sigma_j\right) \cdot P\left(\boldsymbol{\theta_j}, \sigma_j \mid \boldsymbol{\gamma}, \sigma\right) \cdot P(\boldsymbol{\gamma}, \sigma \mid \{Y\})}{P\left(Y_j \mid \boldsymbol{\gamma}, \sigma\right)}$$

- Use level-1 posterior distribution to make predictions on individual labels

$$P\left(y_j^*\right) = \int_{\theta_j, \sigma_j} P\left(y_j^* \mid \boldsymbol{\theta_j}, \sigma_j\right) P\left(\boldsymbol{\theta_j}, \sigma_j \mid \boldsymbol{\gamma}, Y_j\right) d\boldsymbol{\theta_j} d\sigma_j$$

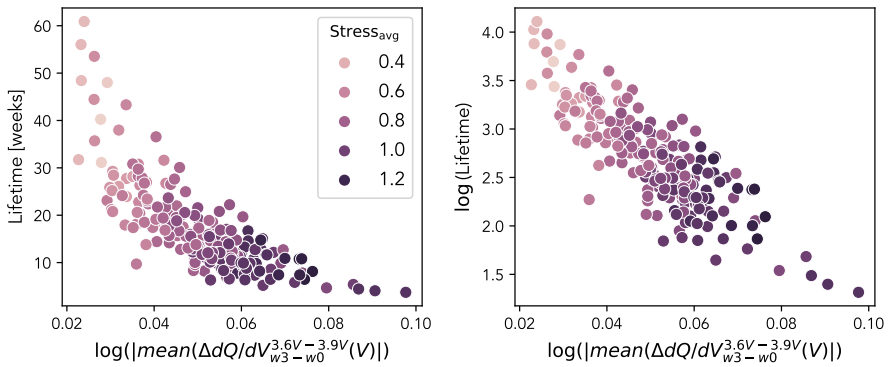


Figure S14: Relationship between cell-level feature mean $\left(\Delta dQ/dV^{3.60\mathrm{V}-3.90\mathrm{V}}_{\mathrm{w3-w0}}(V)\right)$ and cell lifetime, colored by $\mathrm{Stress}_{\mathrm{avg}}$.

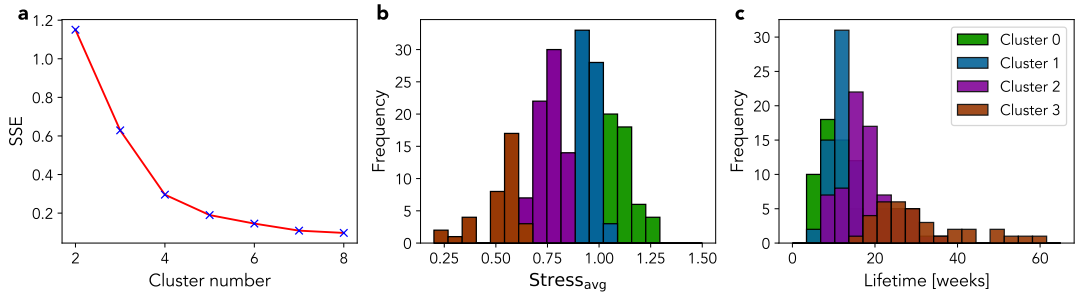


Figure S15: Overview of clustering results. **a**, Influence of number of clusters on clustering score SSE. **b**, Histogram of stress factor $\mathrm{Stress}_{\mathrm{avg}}$ colored by cluster. **c**, Corresponding lifetime distribution for each cluster.

Note S15 HBM Related Analysis

Fig. S14 shows a scatter plot of the cell-level feature mean $\left(\Delta dQ/dV^{3.60\mathrm{V}-3.90\mathrm{V}}_{\mathrm{w3-w0}}(V)\right)$ vs. lifetime, colored by $\mathrm{Stress}_{\mathrm{avg}}$.

As $\mathrm{Stress}_{\mathrm{avg}}$ decreases, the slope of the cell-level feature-lifetime relationship becomes steeper. However, the changing trend is not as clear when analyzing the data on a log-log plot. Normally, the reason for using the log transform on both the feature and the target is to increase the Pearson linear correlation coefficient, as a higher linear correlation will generally improve model prediction performance. This pursuit of a one-for-all linear relationship between the feature and the target hides the data's differences and hierarchical structure caused by the various cycling conditions.

The clustering score $\mathrm{SSE} = \sum_{i=0}^{N} (x_i - c_i)^2$, which describes the sum of squared distances between sample points and their assigned centroid, is used to evaluate the influence of the number of clusters on the clustering results.

Fig. S15a shows the SSE as a function of the number of clusters. According to the empirical elbow rule[13], we select $K = 4$ clusters. From Figs. S15b and S15c, we observe two sources of variability that affect lifetimes. The first is the cross-cluster lifetime variability, which arises from differences in usage, and is measured as a difference in $\mathrm{Stress}_{\mathrm{avg}}$. The other source of lifetime variability arises from in-cluster differences due to manufacturing variability and cycling tester variability.

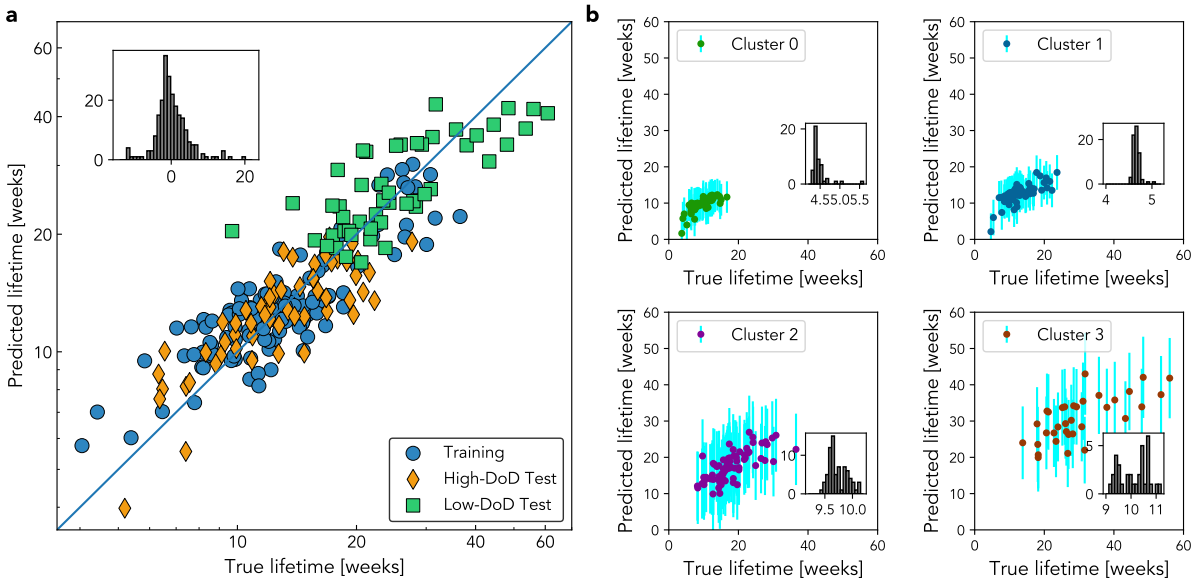


Figure S16: Overview of HBM results. **a**, True vs. predicted lifetimes using the optimal two features extracted from weeks three and zero $(\mathrm{w}_3 - \mathrm{w}_0)$, with embedded histogram showing prediction residuals. **b**, Predictions for each cluster with 2 standard deviations as the corresponding error bar for each sample. The embedded histograms show a summary of error bars

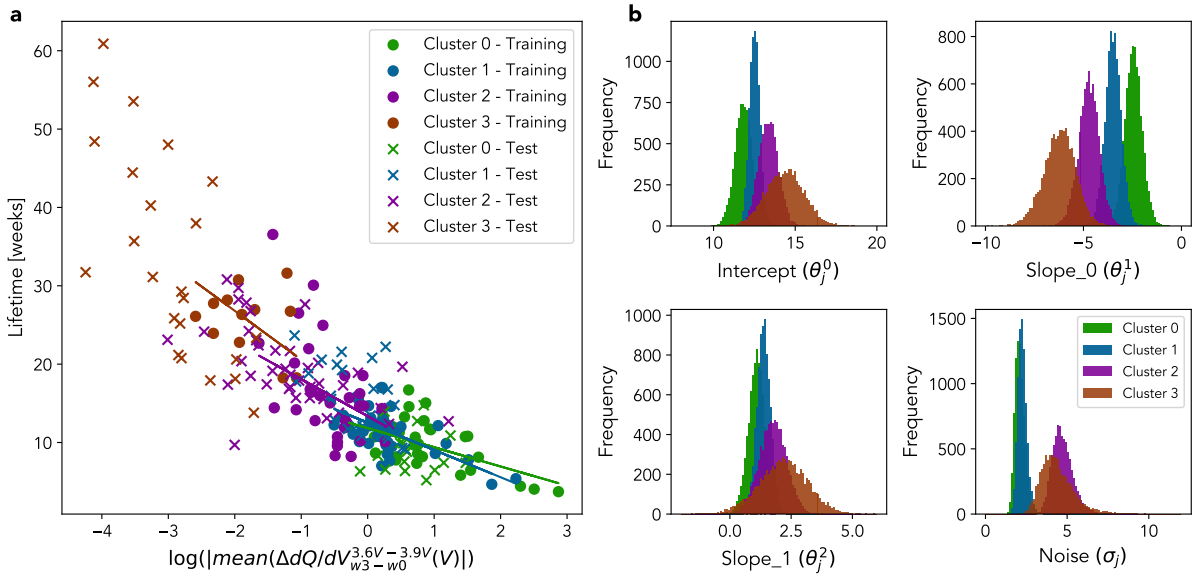


Figure S17: Visualization of HBM regression parameter uncertainties. **a**, Relationship between $\log(\left|\mathrm{mean}(\Delta dQ/dV_{\mathrm{w3-w0}}^{3.6V-3.9V}(V)\right|)$ and true lifetime across different clusters and train-test split ("Test" denotes samples from both high- and low-DoD sets). Fits, corresponding to mean parameter values, are plotted for each cluster. **b**, Uncertainty histograms for regression parameters $(\theta_j^0, \theta_j^1, \theta_j^2, \sigma_j)$ of each cluster.

Further analysis of uncertainty is shown in Fig. S17. The HBM successfully captures the changing slope describing the relationship between $\log(\left|\text{mean}(\Delta dQ/dV_{\text{w3-w0}}^{3.6V-3.9V}(V)\right|)$ and true lifetime in Fig. S17a. By exploiting the assumption that cell-level regression coefficients are decided by cycling stress cluster-level features, the HBM gives a reasonable fit for Cluster 3 samples (46) based on a very limited training set (12). Considering the posterior predictive distribution expression $p\left(y_j^* \mid Y_j\right)$, the uncertainty on predictions is influenced by both the uncertainly from the regression intercepts and slopes θ_j, and the uncertainty due to measurement noises σ_j. Fig. S17b shows these two kinds of uncertainties across all clusters. The posterior probability distributions for $\boldsymbol{\theta_j}$ and σ_j are much wider for cluster 3 than for any other clusters.

There are a couple of potential improvements regarding the performance of HBM on unseen data. First, as discussed in Sec. 2.4, additional training samples or longer time intervals for extracting features could reduce the overfitting issue. Second, since the feature selection process in this work is based on the performance of a regular linear model, marginal likelihood-related criteria, such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC), could be incorporated in deciding the optimal feature subsets for HBM at both the condition and cell level.

Table S3: Prediction results under different feature sets

	Feature Set	Model	MAPE [%]			RMSE [weeks]		
			Training	High DoD	Low DoD	Training	High DoD	Low DoD
Single-Feature Models	$\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$	Elastic net	20.1	21.8	24.1	3.8	3.8	7.5
		HBM	20.5	22.4	23.8	3.7	3.7	7.9
	$\log(\lvert\text{var}(\Delta Q_{\text{w3-w0}}(V)\rvert)$	Elastic net	25.7	29.5	35.4	5.1	4.9	14.1
		HBM	25.3	29.5	29.9	4.6	5.2	11.4
Multi-Feature Models	$\log(\lvert\text{var}(\Delta Q_{\text{w3-w0}}(V)\rvert)$, $\log(\lvert\text{mean}(\Delta Q_{\text{w3-w0}}(V)\rvert)$	Elastic net	25.5	28.4	30.5	5.0	4.8	12.1
		HBM	25.1	28.2	25.8	4.6	5.0	9.7
	$\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$, $\log(\lvert\text{var}(\Delta Q_{\text{w3-w0}}(V)\rvert)$, C_{chg}, C_{dchg}, DoD	Elastic net	19.9	20.0	25.8	3.7	3.5	7.7
		HBM	19.5	20.5	23.2	3.5	3.3	8.1
	$\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$, $\log(\lvert\Delta\text{CV Time}_{\text{w3-w0}}\rvert)$, C_{chg}, C_{dchg}, DoD	Elastic net	16.5	15.0	29.2	3.1	2.8	8.6
		HBM	17.3	15.3	21.0	3.0	2.8	7.8

Note S16 Comparisons with Algorithms and Features in Early Life Prediction

Control experiments have been done to examine the prediction accuracy of the two machine learning models, elastic net and HBM, when trained with different feature subsets. The prediction results of these experiments are summarized in Table S3, and the feature subsets are the following:

- Feature subset 1: $\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$
- Feature subset 2: $\log(\lvert\text{var}(\Delta Q_{\text{w3-w0}}(V)\rvert)$, known as "variance model" in[1].
- Feature subset 3: $\log(\lvert\text{var}(\Delta Q_{\text{w3-w0}}(V)\rvert)$, $\log(\lvert\text{mean}(\Delta Q_{\text{w3-w0}}(V)\rvert)$
- Feature subset 4: $\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$, $\log(\lvert\text{var}(\Delta Q_{\text{w3-w0}}(V)\rvert)$, C_{chg}, C_{dchg}, DoD
- Feature subset 5: $\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$, $\log(\lvert\Delta\text{CV Time}_{\text{w3-w0}}\rvert)$, C_{chg}, C_{dchg}, DoD

There are several findings from these comparative results.

- In general, the HBM outperforms the elastic net model in extrapolation when there are multiple features or when the feature has a relatively low linear correlation with lifetime across the whole dataset.
- When we only consider a single feature, the proposed $\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$ feature outperforms $\log(\lvert\text{var}(\Delta Q_{\text{w3-w0}}(V)\rvert)$, attributable to the fact that the former feature has a higher linear correlation with lifetime.
- A comparison of the prediction results by the "variance model" and "discharge model" in[1] shows that the discharge model, which has five input features, outperforms the variance model, which has only one input feature.
- For feature subset 5, by combining cycling conditions (C_{chg}, C_{dchg}, DoD) with the two optimal early-life features, both the training error and the prediction error on the high-DoD test set (in-distribution samples) using a simple elastic net model decrease, and the prediction accuracy on the low-DoD test set (out-of-distribution samples) worsens. However, the results from HBM show its robustness to over-fitting. This finding is similar to that discussed in Sec. 2.4, where we compared the results for the optimal 2- and 3-feature models.

To further illustrate the advantage of HBM in tackling the early life prediction problem formulated in this paper, we perform some additional machine learning experiments. For this comparison study, we use the feature set with two optimal features, $\log(\lvert\text{mean}(\Delta dQ/dV^{3.6V-3.9V}_{\text{w3-w0}}(V)\rvert)$ and $\log(\lvert\Delta\text{CV Time}_{\text{w3-w0}}\rvert)$. The four candidate algorithms are elastic net, HBM, NuSVR, and Extra-Trees. The latter two are proven effective in early life prediction in a large dataset with various chemistries[14]. The results of this comparative study are summarized in Table S4.

Table S4: Prediction results of different candidate algorithms

Model	**MAPE [%]**			**RMSE [weeks]**		
	Training	High DoD	Low DoD	Training	High DoD	Low DoD
Elastic net	17.3	16.0	24.4	3.2	3.0	7.8
HBM	18.6	16.9	21.8	3.3	3.1	7.3
NuSVR	17.6	16.2	28.4	3.2	3.0	8.9
ExtraTrees	8.6	20.0	28.3	1.7	3.6	11.3

In general, HBM is advantageous in solving the problem defined in this work from two aspects. First, the HBM considers the hierarchical structure of the dataset by learning the difference in feature-lifetime relations resulting from varying aging conditions. In contrast, the other three models consider all training samples equally, thus neglecting the effect of aging conditions. Second, the problem setting in this work prevents NuSVR and Extratrees models from performing at their full capability. NuSVR has a similar performance with elastic net because the optimal kernel selected by cross-validated hyperparameter grid search, in this case, is a linear kernel. One advantage of NuSVR or a general SVR is to tackle non-linear regression problems by selecting different kernels. The ExtraTrees model fits the training set well, with the lowest training error, while producing higher test errors on both test sets. This observation can be attributed to two potential issues. First, the size of the training set is considered small for a decision-tree-based model in terms of both the number of input features and the number of training points. Second, the prediction performance of ExtraTrees on a test set is largely influenced by the discrepancy between the training and test data distributions. Such a data distribution shift may likely be larger for the low-DoD test set, requiring the model to extrapolate more when making predictions on this test set.

References

1. Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nature Energy *4*, 383–391. https://doi.org/10.1038/s41560-019-0356-8.

2. Attia, P. M., Grover, A., Jin, N., Severson, K. A., Markov, T. M., Liao, Y.-H., Chen, M. H., Cheong, B., Perkins, N., Yang, Z., et al. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature *578*, 397–402. https://doi.org/10.1038/s41586-020-1994-5.

3. Baumhöfer, T., Brühl, M., Rothgang, S., and Sauer, D. U. (2014). Production caused variation in capacity aging trend and correlation to initial cell performance. Journal of Power Sources *247*, 332–338. https://doi.org/10.1016/j.jpowsour.2013.08.108.

4. Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A., and Sauer, D. U. (2021). One-shot battery degradation trajectory prediction with deep learning. Journal of Power Sources *506*, 230024. https://doi.org/10.1016/j.jpowsour.2021.230024.

5. Dubarry, M. and Anseán, D. (2022). Best practices for incremental capacity analysis. Frontiers in Energy Research *10*. https://doi.org/10.3389/fenrg.2022.1023555.

6. Raj, T., Wang, A. A., Monroe, C. W., and Howey, D. A. (2020). Investigation of Path-Dependent Degradation in Lithium-Ion Batteries. Batteries & Supercaps *3*, 1377–1385. https://doi.org/10.1002/batt.202000160.

7. Dahn, H. M., Smith, A., Burns, J., Stevens, D., and Dahn, J. (2012). User-friendly differential voltage analysis freeware for the analysis of degradation mechanisms in Li-ion batteries. Journal of The Electrochemical Society *159*, A1405. https://doi.org/10.1149/2.013209jes.

8. Li, X., Colclasure, A. M., Finegan, D. P., Ren, D., Shi, Y., Feng, X., Cao, L., Yang, Y., and Smith, K. (2019). Degradation mechanisms of high capacity 18650 cells containing Si-graphite anode and nickel-rich NMC cathode. Electrochimica Acta *297*, 1109–1120. https://doi.org/10.1016/j.electacta.2018.11.194.

9. Keil, P. and Jossen, A. (2016). Calendar aging of NCA lithium-ion batteries investigated by differential voltage analysis and coulomb tracking. Journal of The Electrochemical Society *164*, A6066. https://doi.org/10.1149/2.0091701jes.

10. Eddahech, A., Briat, O., and Vinassa, J.-M. (2014). Determination of lithium-ion battery state-of-health based on constant-voltage charge phase. Journal of Power Sources *258*, 218–227. https://doi.org/10.1016/j.jpowsour.2014.02.020.

11. Yang, J., Xia, B., Huang, W., Fu, Y., and Mi, C. (2018). Online state-of-health estimation for lithium-ion batteries using constant-voltage charging current analysis. Applied Energy *212*, 1589–1600. https://doi.org/10.1016/j.apenergy.2018.01.010.

12. Wang, Z., Zeng, S., Guo, J., and Qin, T. (2019). State of health estimation of lithium-ion batteries based on the constant voltage charging curve. Energy *167*, 661–669. https://doi.org/10.1016/j.energy.2018.11.008.

13. Yuan, C. and Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. J *2*, 226–235. https://doi.org/10.3390/j2020016.

14. Paulson, N. H., Kubal, J., Ward, L., Saxena, S., Lu, W., and Babinec, S. J. (2022). Feature engineering for machine learning enabled early prediction of battery lifetime. Journal of Power Sources *527*, 231127. https://doi.org/10.1016/j.jpowsour.2022.231127.

B

Appendix of Chapter 3

Supplymentary: Proof of state space transformation of Gaussian process

We provide the proof of state space transformation of Gaussian process. The core idea is to represent the Gaussian process $f(t)$ having the covariance function $k(t,t')$ as output of a linear time invariant (LTI) stochastic differential equation (SDE).

1 Link Gaussian process and state space model via covariance function

In particular, we consider m-th order scalar LTI SDEs as following:

$$a_0 f(t) + a_1 \frac{df(t)}{dt} + \ldots + a_m \frac{d^m f(t)}{dt^m} = w(t), \tag{1}$$

where $a_0, a_1, ..., a_{m-1}$ are known constants and $w(t)$ is a white noise process. This can be further written as first order vector form Markov process:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}\mathbf{x}(t) + \mathbf{L}w(t), \tag{2}$$

where $\mathbf{x}(t) = [f(t), \frac{df(t)}{dt}, ..., \frac{d^{m-1}f(t)}{dt^{m-1}}]^\top$, which contains the derivatives of $f(t)$ up to order $m-1$. The matrices $\mathbf{F}$ and $\mathbf{L}$ can be written as

$$\mathbf{F} = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & \cdots & -a_{m-2} & -a_{m-1} \end{pmatrix}, \mathbf{L} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{3}$$

Now, operate Fourier transform on both sides of eqn. 2, we get

$$(i\omega)X(i\omega) = \mathbf{F} \cdot X(i\omega) + \mathbf{L} \cdot W(i\omega) \tag{4}$$

where $X(i\omega), W(i\omega)$ are the Fourier transform of $\mathbf{x}(t), w(t)$ respectively. Further calculate the spectral density $S_x(\omega)$ of $\mathbf{x}(t)$,

$$\begin{aligned} S_x(\omega) &= (i\omega I - \mathbf{F})^{-1}\mathbf{L} \cdot \underbrace{W(i\omega)W(i\omega)^\top}_{S_w(\omega)=q} \mathbf{L}^\top \left[(i\omega I - \mathbf{F})^{-1}\right]^\top, \\ S_x(\omega) &= (i\omega I - \mathbf{F})^{-1}\mathbf{L}q\mathbf{L}^\top \left[(i\omega I - \mathbf{F})^{-1}\right]^\top. \end{aligned} \tag{5}$$

Notice that, as a white noise process, $w(t)$ has a spectral density $S_w(\omega) = q$. By defining $\mathbf{H} = [1, 0, ..., 0]^\top$ we can extract the desired Gaussian process $f(t)$ from $\mathbf{x}(\mathbf{t})$ as $f(t) = \mathbf{H}\mathbf{x}(t)$. Thus, we can now get the spectral density $S_f(\omega)$ of $f(t)$ as

$$S_f(\omega) = \mathbf{H}(i\omega I - \mathbf{F})^{-1}\mathbf{L}q\mathbf{L}^\top[(i\omega\mathbf{I} - \mathbf{F})^{-1}]^\top \mathbf{H}^\top. \tag{6}$$

According to the Wiener-Khinchin theorem, spectral density is the Fourier transform of the covariance function. For a zero-mean Gaussian process, it is defined by its covariance matrix calculating from the chosen kernel function $k(t, t')$. Thus, for a Gaussian process $f(t)$, in theory, one can design a LTI system through $\mathbf{F}, \mathbf{L}, \mathbf{H}$ and q, so that a white noise process with spectral density q is fed as system input, while the system output is effectively the desired Gaussian process $f(t)$.

However, not every Gaussian process can be written as LTI SDE as eqn. 1. One premise needs to be satisfied, that is, the spectral density $S_f(\omega)$ can be written as a special form:

$$S(\omega) = \frac{(\text{ constant })}{(\text{ polynomial in } \omega^2)}, \tag{7}$$

Then, by applying spectral factorization [1], we can write the spectral density as

$$S(\omega) = H(\mathrm{i}\omega)qH(-\mathrm{i}\omega), \tag{8}$$

where the transfer functions $H(\mathrm{i}\omega)$ and $H(-\mathrm{i}\omega)$ have their poles in upper and lower planes, respectively. This transfer function enables us to get a frequency domain representation of the process $f(t)$:

$$F(\omega) = H(i\omega)W(\omega) \tag{9}$$

By comparing it with eqn. 1,

$$\begin{aligned} &(i\omega)^m F(\omega) + a_{m-1}(i\omega)^{m-1}F(\omega)\cdots + a_0 F(\omega) = W(\omega) \\ &\underbrace{\left[(i\omega)^m + a_{m-1}(i\omega)^{m-1} + \ldots + a_0\right]}_{H(i\omega)^{-1}} \cdot F(\omega) = W(\omega). \end{aligned} \tag{10}$$

where $a_0, ..., a_{m-1}$ are the coefficients of polynomial in the denominator of $H(\mathrm{i}\omega)$.

2 Matern32 kernel example

After proving the equivalence of Gaussian process $f(t)$ and the corresponding LTI SDE, the remaining question is, given a known kernel function $k(t, t')$, how exactly one can formulate such a LTI SDE system to represent corresponding Gaussian process $f(t)$. Here, we give an concrete example for a Matern32 kernel, which can be written as,

$$k_{\text{Mat}}(t, t') = \sigma^2 \left(1 + \frac{\sqrt{3}(t - t')}{\ell}\right) \exp\left(-\frac{\sqrt{3}(t - t')}{\ell}\right) \tag{11}$$

By set $\lambda = \sqrt{3}/\ell$, the spectral density of this Martern32 covariance function can be expressed as [2]

$$S(\omega) \propto \frac{4\lambda^3\sigma^2}{(\lambda + i\omega)^{-2}(\lambda - i\omega)^{-2}} \tag{12}$$

Notice that, the transfer function $H(i\omega) = (\lambda + i\omega)^{-2}$ and the spectral density of white noise process $q = 4\lambda^3\sigma^2$. According to eqn. 10, we can get

$$a_0 = 0, a_1 = 2, a_2 = 1 \tag{13}$$

Thus, the corresponding LTI SDE for Matern32 kernel can be represented by:

$$\frac{d\mathbf{x}(t)}{dt} = \begin{bmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t) \tag{14}$$

References

[1] Wiener, N.: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. The MIT press, Cambridge, Massachusetts (1964)

[2] Sarkka, S., Solin, A., Hartikainen, J.: Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering. IEEE Signal Processing Magazine **30**(4), 51–61 (2013)

C

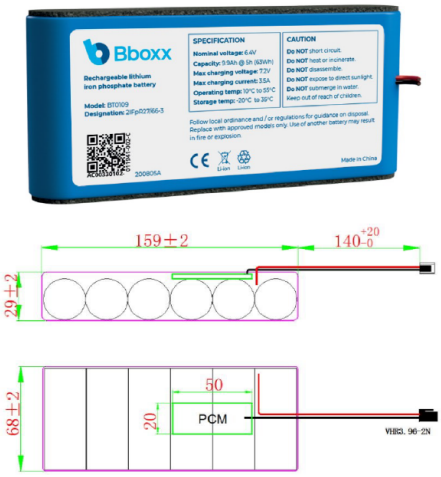
Appendix of Chapter 4

Supplymentary: Battery capacity and degradation modes diagnosis for solar off-grid systems in Africa

Appendix A Bboxx LFP pack details

Basic performance	
Nominal voltage	6.4 V
Nominal capacity	9.9 Ah
Nominal energy	63.36 Wh
Expected cycle life	2000 cycles
Impedance (1kHz)	100 mΩ
Operation condition	
Maximum charge voltage	7.2 V
Discharge cut-off voltage	5.0 V
Maximum charge current	3.5 A
Minimum charge current life	3.5 A
Charge temperature	10-55 °C
Discharge temperature	-20-60 °C
Storage temperature[c]	-10-35 °C
Self discharge (at 50% SOC)	$\leq$ 3% per month

Table A1 Student Database



Appendix B Segment extraction

An intuitive question about battery field data is whether all of them are necessary to achieve a good estimation accuracy for the desired battery state and health parameters. Values of health parameters, such as capacity and resistance, are expected to not change dramatically during short periods of time. On the other hand, for an average LFP pack used in this work, 3 years usage data consist of 788400 recorded pairs for terminal voltage (V), applied current (I) and ambient temperature (T). It would be a time-consuming and yet unnecessary task to build a health estimation model based on all these usage data.

Table 1. summarizes the criterion we used to select discharge segments throughout the pack operation time. The idea behind this criterion is to select the deepest discharge segment in a monthly manner. The minimum voltage and discharge Ahs are set to mitigate the flat OCV property of LFP, which guarantee a noticeable voltage drop within the discharge segment. The EOC relaxation before head enables us to recalibrate SOC estimation at the beginning of each discharge segment. Notice that, for some packs with less than 1 year operation time, the one month sampling rate will be changed into half month or even 1 week. For the two-stage training procedure, special care needs to be take for the first selected segment. In order to learn a good representative of operation condition dependency of R_0 and mitigate the problem of overfitting to one single segment, multiple segments (2-5) are extracted in the first operation month.

Appendix C Learning process of Q^{-1} and R_0

The discrete-time state propagation for the ECM is a simple Coulomb counter that can be expressed as

$$z_i = z_{i-1} - Q^{-1} I_i \Delta t_i. \tag{C1}$$

Table B2 Discharge segments selection

Criterion	Accepted range
Segment duration	>7200 s
Discharge current	<-0.05 A
Minimum voltage	<6.3 V
Minimum discharge Ahs	> 3 Ah
Max data gap	< 600 s
One deepest segment per month	Yes
Continuous discharge	Yes
EOC relaxation before head	Yes
Minimum segments number across life	8

The discrete-time propagation for the inverse capacity and resistance parameters, which are GPs, can be written as

$$\begin{aligned} \mathbf{x}_{Q^{-1},j} &= \exp\left(\Delta\zeta_j \mathbf{F}_{\mathrm{WV}}\right) \mathbf{x}_{Q^{-1},j-1} + w_{j-1} \\ \mathbf{x}_{R_0,j} &= \exp\left(\Delta\zeta_j \mathbf{I}_{n_s} \otimes \mathbf{F}_{\mathrm{WV}}\right) \mathbf{x}_{R_0,j-1} + \mathbf{w}_{j-1}, \end{aligned} \tag{C2}$$

where, for the WV kernel, $\mathbf{F}_{\mathrm{WV}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ represents the discrete-time interval on the life-long aging scale. Because of the operating condition dependency of R_0, Kronecker products are used to describe the corresponding discretisation in $\mathbf{I}_{n_s} \otimes \mathbf{F}_{\mathrm{WV}}$ and $\mathbf{w}_{j-1}$. One needs to balance the number of discrete operating conditions and the calculation cost. For SOC dependency, the 0 to 1 SOC range is evenly divided into $n_z = 20$ discrete SOC levels. For temperature influence, because all these packs locate in area near equator and most of discharging happen at night, the temperature variation for selected training segments is small (within 10 °C, see in Fig. F3). Thus, the temperature range (between max and min value) is evenly divided into $n_T = 5$ discrete points. In summary, we set $n_z = 20, n_T = 5$, resulting in $n_s = 100$ discrete operating points. The process noise terms w_{j-1} and $\mathbf{w}_{j-1}$ can be calculated by,

$$\begin{aligned} \mathbf{w}_{k-1} &\sim \mathcal{GP}\left(\mathbf{0}, \mathbf{K}_{\mathrm{s}} \otimes w_{k-1}\right) \\ w_{k-1} &= \int_0^{\Delta t_k} \mathbf{A}_{k-1} \mathbf{L} \mathrm{Q}_{\mathrm{c}} \mathbf{L}^{\top} \mathbf{A}_{k-1}^{\top} \mathrm{d}\tau. \end{aligned} \tag{C3}$$

where Q_c for the WV kernel is σ_ζ^2. The state covariance initialization and process noise matrix of the joint system and GP subsystem are summarized as follows.

$$\begin{aligned} \mathbf{P}_{\mathrm{joint},0} &= \begin{bmatrix} \mathrm{P}_{z,0} & 0 \\ 0 & \mathbf{P}_{\mathrm{GP},0} \end{bmatrix}, \mathbf{P}_{\mathrm{GP},0} = \begin{bmatrix} \mathbf{P}_{Q^{-1},0} & 0 \\ 0 & \mathbf{P}_{R_0,0} \end{bmatrix} \\ \mathbf{W}_{\mathrm{joint}} &= \begin{bmatrix} \mathrm{W}_z & 0 \\ 0 & \mathbf{W}_{\mathrm{GP}} \end{bmatrix}, \mathbf{W}_{\mathrm{GP}} = \begin{bmatrix} \mathbf{W}_{Q^{-1}} & 0 \\ 0 & \mathbf{W}_{R_0}, \end{bmatrix} \end{aligned} \tag{C4}$$

where

$$\begin{aligned} \mathbf{P}_{R_0,0} &= \mathbf{K}_{\mathrm{Mat}} \otimes \mathbf{P}_{\zeta_0,\mathrm{WV}}, \quad \mathbf{P}_{Q^{-1},0} = \mathbf{P}_{\zeta_0,\mathrm{WV}} \\ \mathbf{W}_{R_0} &= \mathbf{K}_{\mathrm{Mat}} \otimes \mathbf{W}_{\mathrm{WV}}, \quad \mathbf{W}_{Q^{-1}} = \mathbf{W}_{\mathrm{WV}}. \end{aligned} \tag{C5}$$

In these equations, ζ_0 represents the initial aging time step and $\mathbf{K}_{\mathrm{Mat}}$ is the covariance matrix of the Matern kernel at n_z discrete SOC levels. For the selected WV kernel, the initial covariance and noise matrix are

$$\mathbf{P}_{\zeta_0,\mathrm{WV}} = \sigma_z^2 \begin{bmatrix} \frac{1}{3}\zeta_0^3 & \frac{1}{2}\zeta_0^2 \\ \frac{1}{2}\zeta_0^2 & \zeta_0 \end{bmatrix}, \mathbf{W}_{\mathrm{WV}} = \sigma_z^2 \begin{bmatrix} \frac{1}{3}\Delta\zeta^3 & \frac{1}{2}\Delta\zeta^2 \\ \frac{1}{2}\Delta\zeta^2 & \Delta\zeta \end{bmatrix}. \tag{C6}$$

we adopted our previously proposed co-estimation framework [20]. Within discharge segments, the GP states (i.e. circuit model parameters) and the ECM state are propagated jointly. This means the GP dynamics described in eqn. C2 also evolve within each segment over time steps of Δt. However, since $\Delta t \ll \Delta\zeta$, the changes in Q and R_0 over small time steps will be small since the GP is relatively smooth on the timescale of Δt.

Between segments, however, only the GP states are preserved; the ECM state is reinitialized at the beginning of every new discharge segment by inverse the terminal voltage via initial-of-life OCV-SOC relationship (as long rest is guaranteed before every discharge segment). In this way, the GPs describing capacity and resistance, as well as the ECM describing state of charge dynamics each share the same observations (i.e. terminal voltage), but their propagation is at different timescales.

Within a segment, the overall joint state vector for both ECM and GPs is given by

$$\mathbf{x}_{\text{joint},i} = \begin{bmatrix} z_i \\ \mathbf{x}_{Q^{-1},i} \\ \mathbf{x}_{R_0,i} \end{bmatrix}. \tag{C7}$$

Between segments, the GP subsystem is linear and the GP state vector is given by

$$\mathbf{x}_{\text{GP},j} = \begin{bmatrix} \mathbf{x}_{Q^{-1},j} \\ \mathbf{x}_{R_0,j} \end{bmatrix}. \tag{C8}$$

Note that within a segment the joint system is nonlinear because Eqn. 6 is nonlinear in SOC (z) due to the $V_0(z)$ and $R_0(z)$ functions. For simplicity, we write the nonlinear joint system dynamics as

$$\begin{aligned} \mathbf{x}_{\text{joint},i} &= g\left(\mathbf{x}_{\text{joint},i-1}, I_{i-1}, \mathbf{W}_{\text{joint},i-1}\right) \\ V_i &= h\left(\mathbf{x}_{\text{joint},i}, I_{i-1}, v_i\right). \end{aligned} \tag{C9}$$

An extended Kalman filter is used to solve the system dynamics with local linearisation via Jacobian matrices

$$\mathbf{G}_i = \frac{\mathrm{d}g}{\mathrm{d}\mathbf{x}_{\text{joint},i}}, \quad \mathbf{H}_i = \frac{\mathrm{d}h}{\mathrm{d}\mathbf{x}_{\text{joint},i}}. \tag{C10}$$

The full co-estimation framework is given in Algorithm 1. Because R_0 is modelled with a GP, its predictive variance needs to be included when calculating the noise covariance of output (the $I_i^2 \mathbf{\Sigma}_{R_0,i}$ term in $\mathbf{S}_i$).

$$\begin{aligned} V_i &= V_0(z_i) + I_i R_{0,i} + \nu_i, \quad \nu_i \sim N\left(0, \sigma_v^2\right) \\ \mathrm{E}[V_i] &= V_0(z_i) + I_i m_{R_{0i}} \\ \mathrm{cov}(V_i) &= \mathrm{E}\left[\left(V_i - \mathrm{E}[V_i]\right)\left(V_i - \mathrm{E}[V_i]\right)^\top\right] \\ \mathrm{cov}(V_i) &= I_i^2 \Sigma_{R_{0i}} + \sigma_\nu^2. \end{aligned} \tag{C11}$$

The co-estimation in Algorithm 1 can be viewed as a single forward run of the proposed model under a specific GP hyperparameter set and measurement noise variance (i.e., $\sigma_\zeta^2, \sigma_s^2, l_z^{-1}, l_T^{-1}, \sigma_n^2$). To find the optimal values of these, we use the accumulated NLML (Φ) as our cost function. Because of the additive nature of the NLML, Φ can be updated during the state and parameter co-estimation process. An outer optimisation loop using a gradient-based L-BFGS-B algorithm is then used to obtain *maximum a posteriori* (MAP) estimates of the hyperparameters for GPs, running the forward model many times, with each forward run giving one evaluation of Φ.

The hyperparameter optimisation was solved in a parallel way on a virtualized Linux platform with two physical Intel Xeon Silver 5216 CPUs at 2.1 GHz. Specifically, optimization was performed for 8 packs in parallel at once, which took approximately 120 minutes to finish.

Appendix D Benchmark comparison

First proposed by Plett [8], dual extended Kalman filter are often used for the problem of battery SOC and SOH co-estimation. By separating time scales for state and parameter dynamics, multi-scale dual estimation methods are proposed [48, 49]. Here, a multi-scale dual EKF estimation model is used as a benchmark for model comparison. The parameter dynamics (i.e. capacity and resistance) is described by a random walk process. Two extended Kalman filters (EKF) are used for state (SOC) estimation and parameter estimates. Following the setting of

Algorithm 1 Co-estimation of battery state and GPs using extended Kalman filter. The NLML (Φ) is calculated recursively on every terminal voltage measurement.

Input: Operational segments across life, a hyper-parameter set ($\hat{\theta}$)
Output: estimates for Q and R_0, corresponding cost function value $\Phi(\hat{\theta})$
Initialisation for co-estimation at $\zeta = \zeta_0$
$\mathbf{x}^{+}_{\mathrm{GP}} = \mathbf{x}_{\mathrm{GP},0}, \mathbf{P}^{+}_{\mathrm{GP}} = \mathbf{P}_{\mathrm{GP},0}$
$\Phi = 0$
for $\mathrm{seg}_j \in$ segments **do**
  $\mathbf{x}^{-}_{\mathrm{GP},j} = \exp(\mathbf{F}\Delta\zeta_j)\mathbf{x}^{+}_{\mathrm{GP},j-1}$ ▷ GP propagation
  $\mathbf{P}^{-}_{\mathrm{GP},j} = \exp(\mathbf{F}\Delta\zeta_j)\mathbf{P}^{+}_{\mathrm{GP}}\exp(\mathbf{F}\Delta\zeta_j)^{\mathrm{T}} + \mathbf{W}_{\mathrm{GP}}(\Delta\zeta_j)$
  $\mathbf{x}^{+}_{\mathrm{joint}} = \mathbf{x}_{\mathrm{joint},0} = [z_0, \mathbf{x}^{-}_{\mathrm{GP},j}]^{\top}$ ▷ Joint system initialization
  $\mathbf{P}^{+}_{\mathrm{joint}} = \mathbf{P}_{\mathrm{joint},0} = \begin{bmatrix} \mathbf{P}_{z,0} & 0 \\ 0 & \mathbf{P}^{-}_{\mathrm{GP},j} \end{bmatrix}$
  for $V_i \in \mathrm{seg}_j$ **do**
    $\mathbf{x}^{-}_{\mathrm{joint},i} = g\left(\mathbf{x}^{+}_{\mathrm{joint},i-1}, I_{i-1}, \mathbf{W}_{\mathrm{joint},i-1}\right)$ ▷ Joint system propagation
    $\mathbf{P}^{-}_{\mathrm{joint},i} = \mathbf{G}_{i-1}\mathbf{P}^{+}_{\mathrm{joint},i-1}\mathbf{G}^{\mathrm{T}}_{i-1} + \mathbf{W}_{\mathrm{joint},i-1}$
    $\mathbf{e}_i = V_i - h\left(\mathbf{x}^{-}_{\mathrm{joint},i}, I_{i-1}, v_i\right)$
    $\mathbf{S}_i = \mathbf{H}_i\mathbf{P}^{-}_{\mathrm{joint},i}\mathbf{H}^{\top}_i + I^2_i\mathbf{\Sigma}_{R_0,i} + \sigma^2_v$
    $\mathbf{L}_i = \mathbf{P}^{-}_{\mathrm{joint},i}\mathbf{H}^{\top}_i\mathbf{S}^{-1}_i$
    $\mathbf{x}^{+}_{\mathrm{joint},i} = \mathbf{x}^{-}_{\mathrm{joint},i} + \mathbf{L}_i\mathbf{e}_i$ ▷ Joint system update
    $\mathbf{P}^{+}_{\mathrm{joint},i} = (\mathbf{I} - \mathbf{L}_i\mathbf{H}_i)^{\top}\mathbf{P}^{-}_{\mathrm{joint},i}(\mathbf{I} - \mathbf{L}_i\mathbf{H}_i) + \mathbf{L}_i\mathbf{e}_i\mathbf{L}^{\top}_i$
    $\Phi = \Phi + \frac{1}{2}\mathbf{e}^{\top}_i\mathbf{S}^{-1}_i\mathbf{e}_i + \frac{1}{2}\log|2\pi\mathbf{S}_i|$ ▷ Cost function update
  end for
end for

previous study [49], parameters are estimated with macro scale (i.e., in a time step of 1200 seconds in this work), while the state is estimated with micro scale (i.e. in a time step of 120 seconds).

Adjustments are made to make this dual EKFs capable for long term capacity tracking. Specifically, this dual EKFs model is operated from scratch for every single discharge segment. For the first discharge segment, capacity and resistance are initialized with norminal values (9.9 Ah and 0.1 Ω). For later segments, capacity and resistance are initialized with estimation values from last segment. While the variance of capacity and resistance are always initialized with 1 and 0.01 to enable noticeable updates. The capacity estimation result is shown in Fig. D1. Overall, the benchmark model can somehow capture the decay trend of capacity; however, the estimated values are quite noisy. There are two reasons: first, there is no theoretical guarantee for the convergence of dual estimation [11]. And a bad estimation will be used as the next initialization value, which further exacerbate this convergence problem. Second, the random walk assumption for parameter dynamic allows arbitrary changes for parameters without considering lifelong consistency with other estimated capacity values.

Appendix E Extreme capacity predictions

Fig. E2 shows packs with extremely large short-term prediction errors caused by a rapid 'knee' just beyond the prediction horizon.

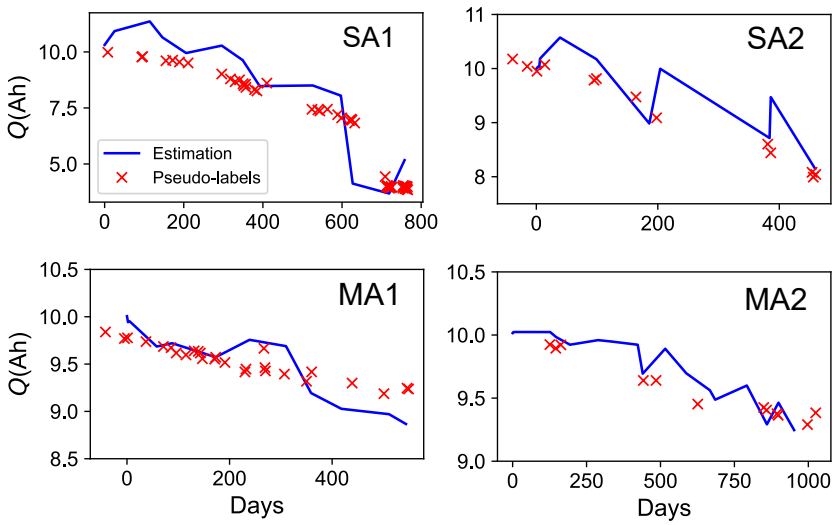


Fig. D1 Capacity estimation results from benchmark model for SA1, SA2, MA1 and MA2 packs. Noisy estimates comes from the random walk assumption for parameter dynamics.

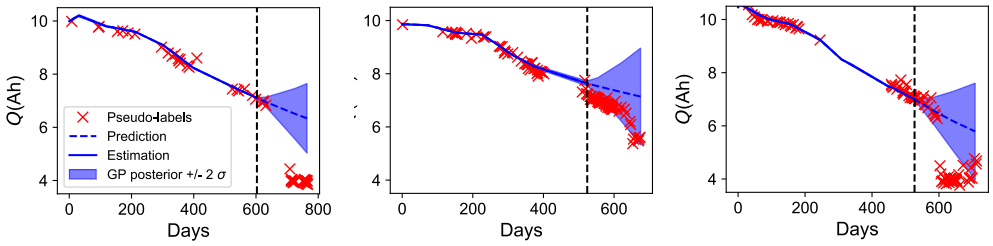


Fig. E2 Three SA packs excluded in Fig. 4 with extreme large prediction errors. From left to right, RMSEs for capacity predictions are 2.13, 0.74 and 1.71. The MAPEs are 47.21%,10.12%,34.56%

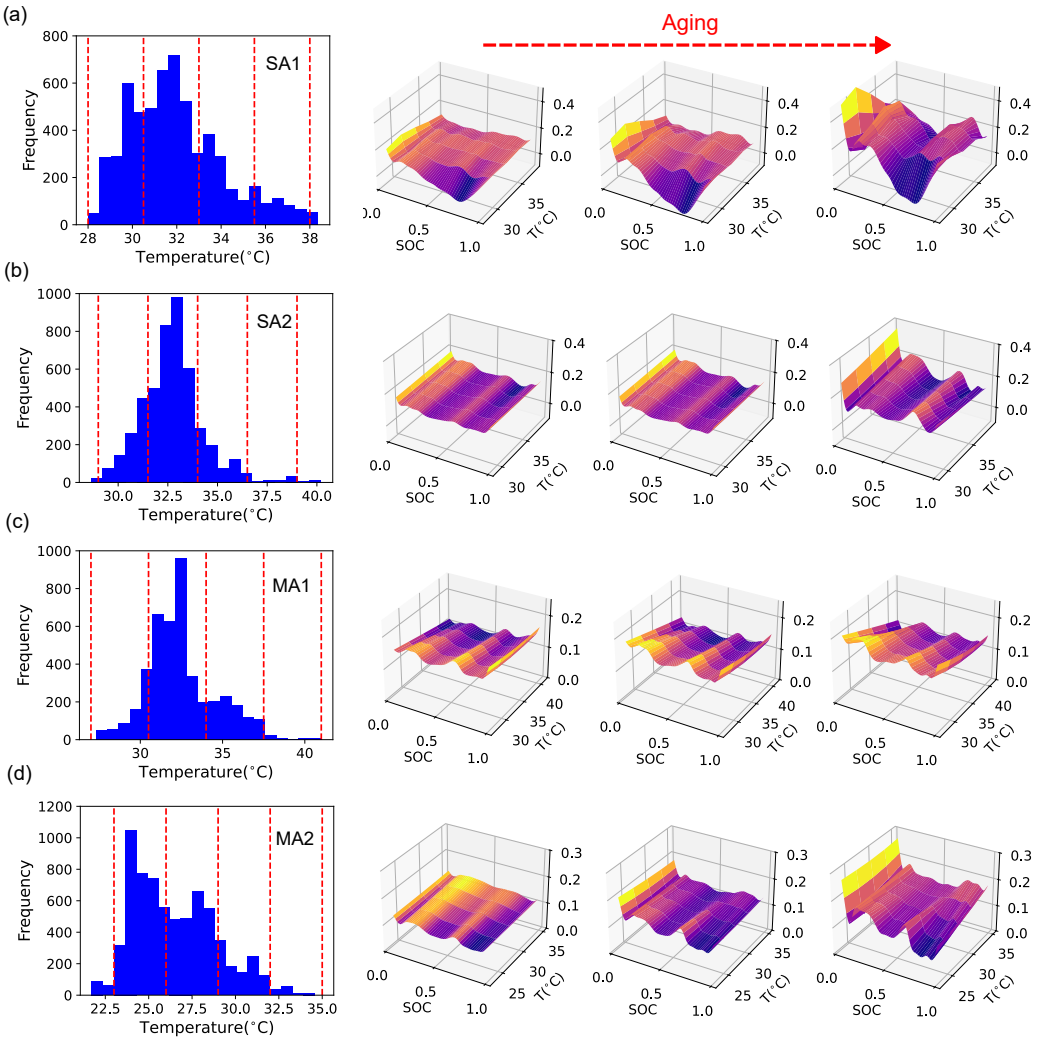


Fig. F3 Temperature dependency of R_0 for the four example packs (SA1,SA2,MA1,MA2), respectively. The temperature variation for selected discharge segments are summarized in histogram plots. And the aging effects on temperature dependency are visualized through 3D plots.

Appendix F Temperature dependency of R_0

As mentioned in main text, R_0 is modeled as a function on SOC and temperature. Different from SOC dependency, in which the R_0 shows large variations, R_0 has only small changes against different temperatures. Fig. F3 shows the temperature dependency of R_0 for four major example packs (SA1, SA2, MA1, MA2).

Appendix G Features for population aging analysis

Table G3 Full set of features extracted for population aging analysis

Feature column 1	Feature column 2
Time spent below 6.7 V	Time spent at rest
Time spent above 6.7 V	Time ratio of rest
Time ratio spent below 6.7 V	Average temperature at rest
Time ratio spent above 6.7 V	Average voltage at rest
Overall average voltage	Overall average temperature
Variance of overall voltage	Variance of voltage at charging
Variance of overall voltage at discharging	Overall EFCs
Average EFC per day	Average temperature at charging
Average temperature at discharging	Average current at charging
Average current at discharging	