

Integrating Physics-Based Modeling and Machine Learning for Degradation Diagnostics of Lithium-Ion Batteries

Adam Thelen^{1,†}, Yu Hui Lui^{1,†}, Sheng Shen¹, Simon Laflamme², Shan Hu¹, Hui Ye³, and Chao Hu^{1,*}

¹ Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA

² Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA 50011, USA

³ Medtronic Energy and Component Center, Brooklyn Center, MN 55430 USA

[†] The authors contributed equally to this work.

* Indicates corresponding author (Email address: chaohu@iastate.edu).

ABSTRACT

Traditional lithium-ion (Li-ion) battery state of health (SOH) estimation methodologies that focused on estimating present cell capacity do not provide sufficient information to determine the cell's lifecycle stage or value in second-life use. Quantifying the underlying degradation modes that cause capacity fade can give further insight into the electrochemical state of the cell and provide more detailed health information such as the remaining active materials and lithium inventory. However, current physics-based methods for degradation diagnostics require long-term cycling data and are computationally expensive to deploy locally on a device. To improve upon current methods, we propose and extensively test two light-weight physics-informed machine learning methods for online estimating the capacity of a battery cell and diagnosing its primary degradation modes using only limited early-life experimental degradation data. To enable late-life prediction (e.g. > 1.5 years) without the use of late-life experimental data, each of the methods is trained using simulation data from a physics-based half-cell model and early-life (e.g. < 3 months) degradation data obtained from cycling tests. The proposed methods are comprehensively evaluated using data from a long-term (3.5 years) cycling experiment of 16 implantable-grade Li-ion cells cycled under two temperatures and C-rates. Results from a four-fold cross-validation study show that the proposed physics-informed machine learning models are capable of improving the estimation accuracy of cell capacity and the state of three primary degradation modes by over 50% compared to a purely data-driven approach. Additionally, this work provides insights into the role of temperature and C-rate in cell degradation.

Keywords: Lithium-ion battery; State of health estimation; Degradation diagnostics; Physics-informed machine learning; Half-cell model

1 INTRODUCTION

Lithium-ion (Li-ion) batteries are an attractive mobile energy storage device due to their high energy density, long cycle life, and continuously falling cost [1]–[3]. Despite the advantages, Li-ion battery cells degrade over time due to irreversible internal electrochemical reactions during operation. As a cell ages, it exhibits a loss of capacity and an increase in impedance where the cell's operating conditions influence the rate of degradation. To ensure safe and reliable operation within a battery-powered system, it is important to continuously monitor and track the cell's state of health (SOH). SOH is a metric to evaluate the aging level of batteries and is commonly defined as a cell's present capacity or impedance relative to its initial capacity or impedance. This definition works well because as a cell ages, its capacity decreases and its impedance increases, often proportionally to the degradation of the cell's internal components [4]. This definition has been the case for some time because obtaining accurate direct measurements of the chemical processes taking place inside a Li-ion battery deployed in the field is still largely infeasible [4]. As a result, many battery health diagnostic methods have been developed to estimate cell capacity based on readily available measurements (i.e., cell voltage, current, and temperature) for the purpose of providing advanced notice of cell replacement or failure.

Capacity estimation methods can generally be classified into two categories: (1) model-based methods [5]–[14] and (2) data-driven methods [15]–[22]. Model-based capacity estimation methods require the use of underlying models such as empirical models [5], [8], [10], [11], equivalent circuit models (ECM) [9], [12], [13], and electrochemical models [6], [7], [14]. Empirical models are mathematical models created to estimate future measurements based on historical measurements. In contrast to ECM and electrochemical models, empirical models typically have no physical meaning. ECM models are higher fidelity than empirical models because they consider a cell’s thermodynamic and kinetic effects in the circuit model. Among the three types of model-based methods, electrochemical models that incorporate chemical/electrochemical kinetics and transport phenomena attain the highest accuracy levels. However, the run time for electrochemical models is also the greatest among the three model-based methods, making electrochemical models mostly infeasible for online/real-time SOH estimation. Typically, the model-based methods are combined with advanced filtering techniques such as a Kalman filter and particle filter. The filtering methods are first used to estimate the battery model parameters by tracking the voltage (V) vs. capacity (Q) curve (VQ curve). Then, the battery model is used to estimate the present capacity. Filtering methods achieve acceptable accuracy but are limited in their ability to share information between cells. The lack of shared information between cells limits filtering methods’ applicability to large datasets, and as a result, these methods may become less favorable as modern datasets grow larger. Furthermore, the ECM and electrochemical require domain knowledge to construct the models and/or experimental data under carefully designed and well-controlled conditions, which is often not available or costly to acquire.

Recently, data-driven methods have begun to emerge as an appealing alternative to model-based methods due to advances in data generation and the increasing number of large publicly available battery degradation datasets [23]. Data-driven methods are often much quicker to construct and deploy on larger datasets compared to model-based methods. As a result, a growing number of studies in the literature have employed data-driven machine learning techniques to estimate cell capacity by leveraging correlations in features extracted from cell voltage and current measurements. In particular, popular machine learning methods such as linear regressors [18], support vector machines [18], relevance vector machines [19], Gaussian processes [20], [22], and neural networks [16], [17], [21] have been successfully applied to estimate the capacity of Li-ion batteries. These methods train a machine learning model to learn the correlation between features extracted from cell measurements (voltage, current, temperature) and the cell’s SOH. The main advantage of data-driven methods is that they can be used on battery cycling data, which do not require special equipment for additional tests, e.g., electrochemical impedance spectroscopy (EIS) tests, and they are chemistry/physics agnostic, which means the method can apply to battery cells with different chemistry. These methods perform their best with access to data from many cells operating under similar conditions.

While there have been many successes using both model-based and data-driven methods to estimate cell capacity, these methods fall short in providing insight on the underlying degradation modes which cause capacity fade. Limiting the definition of cell SOH to be the available cell capacity or internal impedance does not fully reflect battery health because capacity and resistance are high-level measurements driven by low-level internal degradation mechanisms. Diagnosing the underlying degradation modes can provide more insight into battery health, which can facilitate more accurate prediction of future battery health [24], [25] and further optimization of the usage of battery cells [26]. To meet these needs, researchers have attempted to extend the traditional definition of SOH estimation to include the internal cell degradation modes that drive capacity fade.

Three degradation modes in a Li-ion battery cell are commonly reported in the literature: loss of lithium inventory (LLI) and loss of active materials in the positive and negative electrodes (LAM_{PE} and LAM_{NE}), respectively [26]–[29]. These three modes have been extensively studied as popular ways to further quantify the SOH of Li-ion cells. Estimating the state of the degradation modes provide more insight into a cell’s SOH and may shed light on the future rate of degradation, e.g. a cell which has a faster rate of negative active material loss due to lithium plating or other mechanisms may soon begin more rapid capacity fade.

Several degradation diagnostics methods to estimate the state of these modes have been proposed in the past. Han et al. proposed using membership functions to quantify the areas under the peak locations of the differential capacity (dQ/dV) curve and correlated these capacities to LLI and LAM_{NE} degradation modes [26]. However, the authors noted that this method was too computationally intensive to be deployed on a battery management system (BMS) because of the intensive genetic algorithm used to analyze cell degradation. Birkel et al. proposed a diagnostic algorithm for estimating the degradation modes (LLI, LAM_{PE} , and LAM_{NE}) of a battery cell [27]. The proposed algorithm was experimentally validated by reconstructing the pseudo-OCV curve of coin-cells with known amounts of LLI, LAM_{PE} , and LAM_{NE} . Dubarry et al. proposed an SOH estimation method by comparing features extracted from the differential voltage (dV/dQ) and dQ/dV curves of an online cell to features in a simulated offline degradation database [28]. However, generation of the offline path degradation database is extremely computationally intensive. Tian et al. trained a convolutional neural network (CNN) to learn the relationship between a partial IC charge curve and the electrode aging parameters (obtained from a pseudo-OCV curve) [29]. In [29], the predicted electrode aging parameters could be used to quantify the degradation modes of the battery cell.

Despite the promising results, many of the methods proposed for advanced degradation diagnostics have a few drawbacks. First, a handful of the methods used long-term degradation data when constructing their estimators, thereby greatly increasing the time and resources required to collect enough data to train the estimator before deployment. Second, a few of the methods

demonstrated required the use of computationally intensive algorithms. In these cases, the authors recognized that more work needs to be done to reduce computational costs and enable on-board BMS deployment. Third, some methods exhibited high errors when estimating heavily aged cells. It is of great significance to develop methods that can not only enable accurate degradation diagnostics at a late aging stage but also do so using only early aging data. Such methods can alleviate the time and costs of collecting long-term aging data.

To address the aforementioned drawbacks and needs, we propose two different methodologies (dataset augmentation and delta learning) that use light-weight physics-informed machine learning models to online estimate the state of a battery cell's capacity and the three primary degradation modes (LLI, LAM_{PE}, and LAM_{NE}). Each of the proposed methods exploits the well-known correlations between cell dQ/dV curves and cell aging modes to accurately estimate the state of each degradation mode. Drawing inspiration from [28], both methodologies leverage early-life experimental aging data (<14 months test time) and abundant simulation data from a physics-based model to enable late-life estimation (2-4 years) of cell capacity and degradation modes. The two estimation methodologies differ in how they leverage the cheap and readily available physics-based simulation data:

1. Methodology 1, a data augmentation method, combines both the early-life experimental aging data and the simulation data into an augmented dataset which is used to train a machine learning model to estimate the capacity and the state of the three degradation modes.
2. Methodology 2, a delta learning method, first trains an estimator model using the simulation data to estimate the capacity and state of the three degradation modes. Then, the predictions from the estimator model are corrected for bias by the corrector model which is trained using the early-life experimental aging data.

The data augmentation methodology was first shown in our earlier work [30]. This paper expands on our previous work by investigating the delta learning methodology and considering an additional year of newly collected aging data from the long-term cycling tests. We perform extensive parametric studies to quantify whether the use of physics-based simulation data improves SOH estimation model accuracy over simply using the available experimental data alone.

This work advances knowledge in the area of online battery degradation diagnostics by showing how data augmentation and delta learning are effective methods to utilize data from a physics-based simulation model. Additionally, we investigate the long-term evolution of both the full- and half-cell VQ curves for each of the 16 cells for all 3.5 years of data. A high precision charger is used to accurately measure cell voltage and current, and half-cell curve analysis is used to compare the observed degradation behavior across the different cycling conditions [31].

The rest of the paper is arranged in the following manner. Section 2 presents the experimental setup for aging data acquisition. Section 3 covers the physics-based half-cell model and degradation mode quantification. Section 4 outlines the different machine learning models used, their configurations, training datasets, and the evaluation criteria used to evaluate model performance. Section 5 shows and discusses the results in estimating the health parameters (i.e., degradation parameters and cell capacity). Section 6 summarizes the key ideas of this study and offers concluding remarks.

2 EXPERIMENTAL SETUP

We conducted a long-term cycle aging test on 16 fresh implantable-grade Li-ion battery cells. The high-quality prismatic cells with a hermetic encasement are of lithium cobalt oxide/graphite (LCO) chemistry. The implantable-grade cells are used in implantable neuromodulation devices with 10 or more years of targeted longevity. For confidentiality of the technology, the capacity values were normalized throughout this study. The cells were cycled at two ambient temperatures and two discharge rates, a total of four test configurations (see Table 1). A temperature of 37 °C was chosen to simulate the normal working temperature of the implantable-grade battery cells, and 55 °C was chosen to accelerate the capacity fade while maintaining the stability of the battery materials [32]. For each temperature, a charge C-rate of C/3 was used, and two discharge C-rates were considered, i.e., C/24 and C/3. Charging was carried out via a constant-current (CC), constant voltage (CV) step, where the cells were charged at C/3 to an upper cutoff voltage of 4.075 V, at which point charging continued at a constant voltage of 4.075 V until either the charge current was C/50 or the charge time reached 30 min. Following the CC-CV charging step, the cells were CC discharged at C/3 or C/24 until the voltage reached a lower cutoff of 3.4 V. Compared to the nominal operating temperature and discharge rate in an implantable application, 55 °C and C/3 conditions highly accelerated the aging of the cells [11].

During cycling, a characterization test was conducted at an ambient temperature of 40 °C every 2 weeks during the first 3 months and every 4 weeks thereafter. To mitigate the effect of kinetics on the voltage curve, we characterize a battery cell at an extremely low C-rate (i.e., C/50). At such a low C-rate, the voltage curves are approaching the thermodynamic voltage of the battery cell. The main advantage of performing a low-rate characterization test in a lab setting is that the obtained voltage curves better capture the degradation present in the cell with minimal noise or interference from other effects. In turn, this treatment makes differential capacity and differential voltage analysis easier, as there is less noise in the signals. As a result of

lower signal noise, we can expect the proposed physics-informed machine learning models to have an easier time learning to estimate the SOH of the cells.

The characterization test included four sequential steps: 1) CC and CV charge to 4.075 V at C/3 with a cutoff current of C/50, 2) CC discharge to 3.4 V at C/50 and rest for 30 min, 3) CC charge to 4.075 V at C/50 and rest for 30 min, and 4) repeated CC discharge to a voltage corresponding to 10% of the state of charge reduction at C/10 followed by a 1-hour rest period until the cell voltage reached 3.4 V. The temperature (40 °C) and charge rate (C/50) were selected to be consistent with the test condition of the half-cell data used for half-cell analysis [25].

The capacity evolution of the 16 battery cells is shown in Fig. 1. Cell C1 and C2 from group G1 and G3, and cell C3 and C4 from group G2 and G4 were removed roughly halfway through the test for destructive analysis. The destructive analysis was used to confirm that the fitted half-cell model degradation parameters matched the experimental results. Briefly, the group G1 cells (C/24, 37 °C) exhibited the least capacity fade because of the mild ambient temperature and lower C-rate cycling conditions. In contrast, the high ambient temperature used on the group G2 and G4 cells rapidly accelerated the capacity fade.

The CC charge curve at step 3 of the characterization test was used to quantify the degradation parameters with half-cell curve analysis. Both the cycling and characterization tests were carried out using a high precision battery cyclers from NOVONIX company [33]. The high precision cycler has both a lower noise floor and a higher measurement resolution than most other cyclers. As a result, the recorded voltage curves at low C-rates (during characterization) have very little noise, improving the accuracy of our differential voltage and differential capacity analysis described in the following sections. However, the work in this study could likely be carried out using high precision cyclers offered by other companies or even more standard testing equipment.

Table 1 Test matrix of the experimental data.

Group	Charge rate	Discharge rate	Temperature	Number of cells
G1	C/3	C/24	37 °C	4
G2	C/3	C/24	55 °C	4
G3	C/3	C/3	37 °C	4
G4	C/3	C/3	55 °C	4

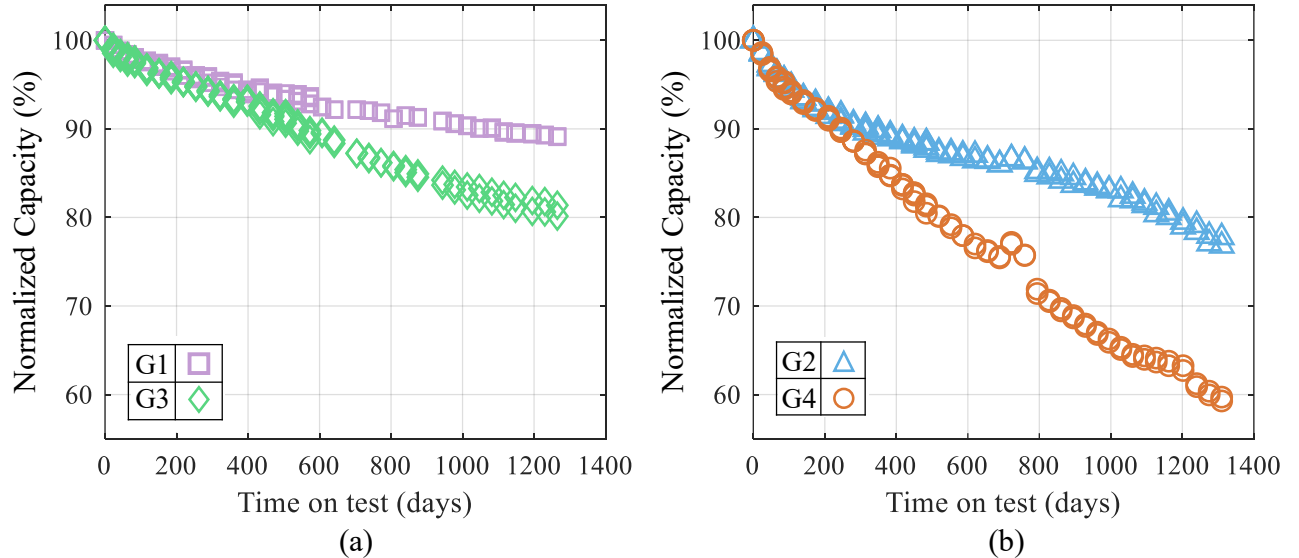


Fig. 1 Capacity fade plots of cells tested at (a) 37°C and (b) 55°C. Each cell has anywhere between 17 and 40 data points, depending on whether or not the cell has been removed for destructive analysis [25].

3 DEGRADATION MODES ANALYSIS AND SIMULATION MODEL

3.1 Degradation Modes

SOH is directly related to the aging of a Li-ion battery cell and can be interpreted on multiple levels (see Fig. 2). On the first level, a cell's aging can be observed as a reduction of cyclable capacity and power capability. Traditional SOH estimation methods mainly focus on estimating SOH at Level 1. Changes in both capacity and power can be measured directly, making quantification of battery SOH using Level 1 metrics straightforward. Level 2 comprises of the three main degradation modes that contribute to capacity and power degradation. The three modes, LAM_{PE} , LAM_{NE} , and LLI, cannot be quantified based on direct measurement from a cell. To quantify the degradation modes, further analysis is required, e.g., half-cell curve analysis and post-mortem analysis. The SOH estimation methods proposed in this study are able to estimate both the cell capacity (Level 1) and the state of the three degradation modes (Level 2). The third level includes different undesirable chemical and physical processes that lead to the degradation modes. These chemical processes are caused/accelerated by different stress factors (Level 4), such as operational duration, temperature, charge and discharge current, and cutoff voltage.

A Li-ion cell's degradation trajectory is highly dependent on its usage and environment. For example, a cell operating at high temperatures may experience more rapid electrolyte decomposition than it would at room temperature. Combine high temperature with long operating times, and a substantial loss of the cell's available lithium inventory is likely. Consequently, the loss of lithium inventory will cause a decrease in the cell's available capacity. The LLI degradation mode includes mechanisms which lead to the consumption of lithium ions, like lithium plating and SEI growth. These mechanisms consume lithium ions, making them no longer available for cycling between the electrodes, ultimately leading to capacity fade.

Shown in Fig. 2, the LAM_{PE} and LAM_{NE} degradation modes describe electrode degradation mechanisms such as particle cracking and loss of electric contact. These mechanisms create small areas of dead active material on the electrodes which are no longer available for lithium insertion. The loss of active materials, both on the positive and negative electrodes, drive cell capacity and power fade. Altogether, quantifying the state of each of the three degradation mechanisms over the life of a cell can provides much more rich information about cell health, which can further improve technology for cell development, second life use, and preventative maintenance.

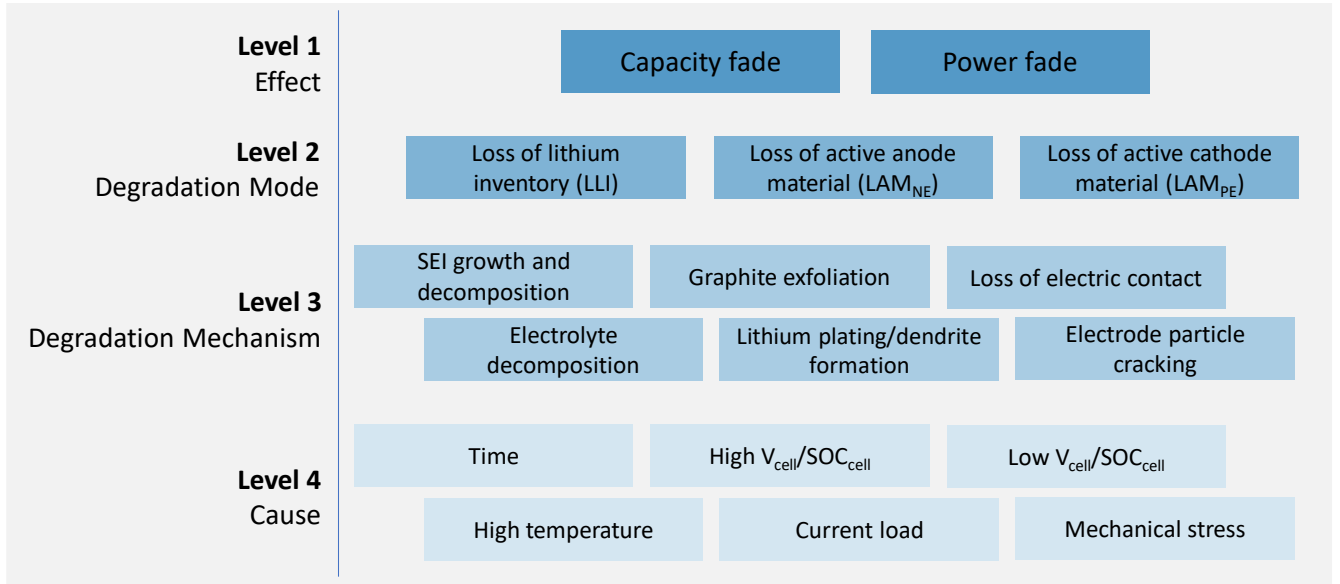


Fig. 2 Simplified tree diagram from [27] showing the relationship between cell use/environment (Level 4), the corresponding degradation mechanisms (Level 3), their connections to the degradation modes (Level 2), and the resulting capacity/power fade (Level 1).

3.2 Half-Cell Model

To quantify the state of the three degradation modes over the life of a cell, we use a physics-based half-cell model like the one reported in [34]–[36]. The half-cell model is a non-destructive degradation analysis method that estimates the state of the three commonly reported degradation modes (i.e., LAM_{PE} , LAM_{NE} , and LLI) in a Li-ion battery cell by reconstructing the measured full-cell V (dV/dQ) vs. Q curve using positive and negative half-cell V (dV/dQ) vs. Q curves. The half-cell model is based on the assumption that at low C-rates, where kinetic effects in the battery cell are negligible, the full-cell curve can be

estimated as the difference between the positive and negative half-cell curves. Figure Fig. 3 (a and c) illustrates the half-cell model where the full-cell curve (black solid line) is calculated by taking the difference between the positive (blue dash line) and negative (red dash-dot line) half-cell curves. It has been reported in the literature that changes in the shape and length of the positive and negative half-cell V (dV/dQ) vs. Q curves relative to their initial measured curves are highly correlated with the three main degradation modes [34], [37], [38]. To quantitatively measure and track the changes in the full- and half-cell curves over the life of a cell, we define three “degradation parameters” used to describe the relative changes in shape and length of the VQ and dV/dQ curves. We discuss the new degradation parameters after briefly introducing VQ and dV/dQ curve analysis. The equations to reconstruct the full-cell VQ and dV/dQ curves from the half-cell curves are shown in equations (1) and (2), respectively:

$$V_c(Q)|_{Q=Q_c} \approx V_p(q_p)|_{q_p=\frac{Q_c-\delta_p}{m_p}} - V_n(q_n)|_{q_n=\frac{Q_c-\delta_n}{m_n}} \quad (1)$$

$$\frac{dV(Q)}{dQ}|_{Q=Q_c} = \frac{1}{m_p} \frac{dV_p(q_p)}{dq_p}|_{q_p=\frac{Q_c-\delta_p}{m_p}} - \frac{1}{m_n} \frac{dV_n(q_n)}{dq_n}|_{q_n=\frac{Q_c-\delta_n}{m_n}} \quad (2)$$

where $V_c(Q_c)$ is the full-cell VQ curve with Q_c denoting the cell capacity, $V(q)$ is the half-cell curve, q is the specific capacity (mAh/g), m is the active mass (g), and δ is the half-cell curve slippage (mAh). The subscripts p and n correspond to the positive and negative electrodes (PE and NE), respectively. Slippage δ_p/δ_n quantifies the horizontal distance the left endpoint of the positive/negative half-cell curve with respect to $Q_c = 0$ mAh [24], [39].

To effectively use equations (1) and (2) for VQ and dV/dQ analysis, a solid understanding of the effect each model parameter has on the full- and half-cell curves is important. Increasing (decreasing) δ , will shift the half-cell curve to the left (right). The active masses in the PE and NE (m_p and m_n) control the capacity ($Q_p = m_p q_p$ and $Q_n = m_n q_n$) of the corresponding half-cell curves, which shrink when the active masses decrease. To ensure the upper and lower cutoff voltages of the simulated curves match those of the measured curves, we primarily look at the VQ curve and adjust the parameters so the endpoints of the simulated curve are close to those of the measured curve. Likewise, to ensure the shape of the fitted full-cell curve matches that of the measured curve, we adjust m_p and m_n while primarily looking at the dV/dQ curve because the phase transition peaks in the dV/dQ curve are sensitive to the values of m_p and m_n . The higher sensitivity of the dV/dQ curve to the values of m_p and m_n makes fitting the shape of the measured curve easier. Figure Fig. 3 shows an example of the half-cell VQ and dV/dQ curve analysis where each parameter has been annotated.

As mentioned earlier, the state of the three degradation modes cannot be determined via direct measurement. In order to estimate the state of these modes, we first need to quantify them. To quantify the state of the three degradation modes in a Li-ion cell, we identify three important “degradation parameters” from the half-cell model. The parameters provide a quantitative link between a cell’s measurable VQ and dV/dQ curves, and its SOH (defined as the state of LAM_{PE} , LAM_{NE} , LLI, and capacity). The degradation parameters are m_p and m_n , which are used to quantify LAM_{PE} and LAM_{NE} , respectively, and the lithium inventory indicator (LII), which is defined as $LII = Q_p - (\delta_p - \delta_n)$, used to quantify LLI [25]. The active mass parameters, m_p and m_n , quantify both the loss of lithiated and delithiated active masses on the electrodes. Any lithium contained in the lithiated LAM is included in the estimate of the total LLI. For instance, if there is a 10% loss in lithiated active material in one of the electrodes ($LAM_{PE/NE,li}$) and 10% loss of lithium due to the solid electrolyte interface (SEI) growth (pure LLI), the total loss of lithium inventory quantified by the LII parameter is 20%, and the total $LAM_{PE/NE}$ is 10%. The LAM parameter used throughout this study is used to quantify delithiated LAM, whereas lithiated LAM can be quantified by a linear combination of both LAM and LII. Practically, the increase of LLI results from the accumulation of parasitic reactions in the cell that contribute to lithium inventory loss (e.g., SEI growth, electrolyte decomposition, and delamination of lithiated electrode materials).

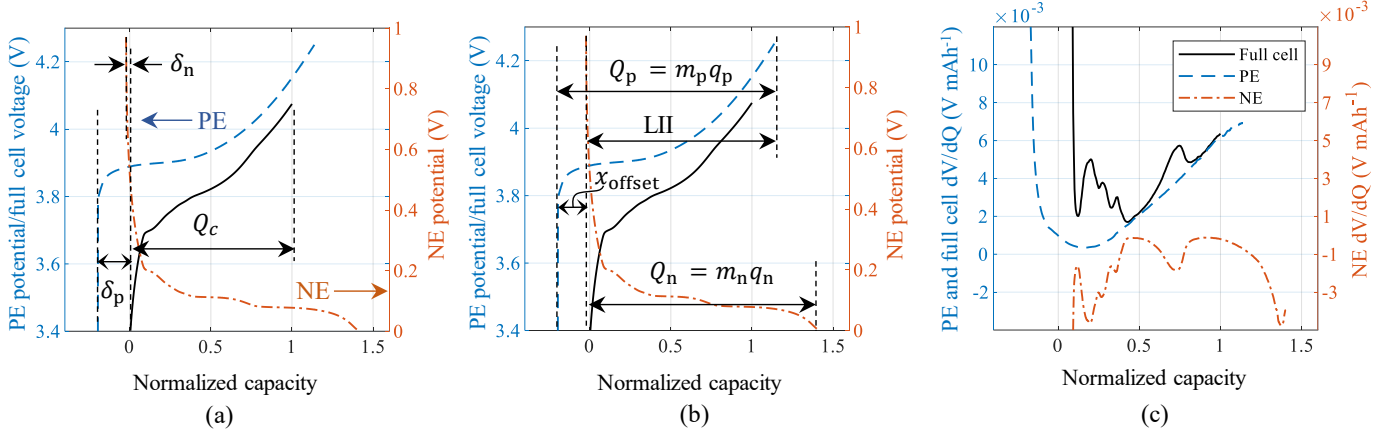


Fig. 3 Half-cell model parameters and degradation parameters visualized and annotated on two separate half-cell VQ curve analysis plots in (a) and (b). Plot (c) shows the corresponding dV/dQ curve derived from the fitted VQ curve in (a) and (b).

3.3 Degradation Mode Quantification and dQ/dV Curves

To quantify the degradation modes of the implantable-grade Li-ion cells over the entirety of the long-term aging experiments, we fit the fresh half-cell VQ and dV/dQ curves to the curves measured from the cell characterization tests by adjusting the four tunable parameters, m_p , m_n , δ_p , and δ_n , in the half-cell model. Manually fitting the half-cell model to these two curves establishes the “true” values for m_p , m_n , δ_p , and δ_n which define the three degradation parameters m_p , m_n , and $LII = Q_p - (\delta_p - \delta_n)$, and ultimately quantify the state of the three degradation modes. Figure Fig. 3 shows an example of VQ and dV/dQ curve analysis with annotated half-cell model parameters. To accurately fit the half-cell model to the experimental data, we adjust the four model parameters as follows.

1. Adjust the values of m_n and δ_n according to the locations of selected dV/dQ features (peaks and valleys) and the voltage reading in the VQ curve at the beginning of charging. We mainly focus on the first and the third peaks as they are consistently visible throughout the long-term aging test. The VQ and dV/dQ curves are shown in Figs. A1–A4 and the two peaks of interest are labeled ① and ②, respectively.
2. Adjust the value of m_p and δ_p according to the magnitude of the dV/dQ features (① and ②) and the voltage reading in the VQ curve at the end of charging.
3. Repeat steps 1 and 2 as necessary until the fitted dV/dQ and VQ curves match as close as possible to the experimentally measured curves.

We chose to manually adjust the parameters in the half-cell model using the VQ and dV/dQ curves because we found that automatic fitting via least squares of the VQ curves produced results inconsistent with underlying physics interpretations of the dV/dQ features’ evolution [25]. Furthermore, automatic fitting of the dV/dQ curves may be inaccurate because many of the dV/dQ features (peaks and valleys) disappeared as the cells aged. We checked the quality of our parameter fit by closely matching the voltage cutoffs on the VQ curve and refined the values of the half-cell model parameters by leveraging the larger feature in the dV/dQ curve which did not disappear with aging, i.e., the first valley near the beginning of charging.

Using the method outlined above, we can determine the degradation parameters for each cell at every characterization test in the last 3.5 years, with the exception of group G4 cells. The cells in group G4 were cycled with a high C-rate and high temperature and therefore exhibited faster degradation, which caused the dV/dQ features to disappear at a certain time on test and prevented any reasonable half-cell model fitting after that. Therefore, to keep the fitting as accurate as possible, we only consider data from group G4 cells up to day 929.

As stated earlier, cells C1 and C2 from groups G1 and G3, and cells C3 and C4 from group G2 and G4 were removed for destructive analysis at days 573 and 484, respectively. Destructive analysis is a common way to verify the fitted degradation parameters from the half-cell model are consistent with open-cell results. The fitted degradation parameters are labeled as “fitted” in Figure A10, A11, A12, and A13 in Appendix A.8 to represent that they are determined via the half-cell model. In Section 5.1, we also present an in-depth analysis on the half-cell model fitting results to show the rationale and accuracy of the fitting that was used to produce the fitted degradation parameters. Our earlier work [25] conducted the destructive analysis to obtain measurements of the remaining active mass from each aged electrode. The measurements of remaining active materials were compared to the fitted degradation parameter values of m_p and m_n to check that the half-cell model was estimating values which were close to the true measurements. In the destructive analysis, an aged cell was first disassembled to obtain the two

electrode materials. These electrode materials were then used to construct a pair of coin cells. To construct each coin cell, the electrode material (either the positive or negative) extracted from the full battery cell was used as the working electrode, and lithium metal was used as the counter and reference electrode. Then, we measured the capacities of these coin cells at the same conditions as the characterization test (see Section 2), that is, by charging and discharging the coin cells with a C-rate of C/50 and a temperature of 40 °C. To derive the remaining positive/negative active mass of the aged full-cell, we divided the measured capacity of a given coin cell by its electrode’s specific capacity. We have shown in [25] that the fitted positive active masses were consistent with the open-cell results and thus verified our degradation modes quantification process that manually fit the half-cell model.

Several studies have shown that changes in a cell’s dQ/dV curve are highly correlated to the capacity fade and the underlying degradation modes LLI, LAM_{PE}, and LAM_{NE} [18], [28], [40]–[42]. That is, features extracted from the dQ/dV curves may be predictive of cell degradation and can potentially be learned by a regression model. These studies proposed using manually extracted dQ/dV features as input to a model which estimated cell capacity and the three degradation parameters. Manual feature selection has the benefit of finding the best feature or a few better features that are highly correlated to the capacity fade of the battery cell from the dQ/dV curves. Common extracted features are the voltage and incremental capacity value of the maximum intensity peak in the dQ/dV charge curve, likewise, the voltage and incremental capacity value of the minimum intensity peak from the dQ/dV discharge curve, and for many chemistries, tracking the values of the dQ/dV curves at a fixed voltage (often 4.0 V) is of interest because it is highly correlated to loss of positive active material [18].

Despite previous literature showing strong correlations between manually extracted features from dQ/dV curves and the degradation modes, we opt to avoid this process altogether and use the full dQ/dV curve as a vector of input features to the machine learning models. Using the entire dQ/dV curve as an input ensures our methodology applies to other battery chemistries without the need for advanced domain knowledge. In some cases, dQ/dV curves can evolve quite heavily at high degradation and cause some manually selected features to fade out. Likewise, cell degradation is highly complex, and some manually extracted features may fail to fully capture the unique changes which appear in the dQ/dV curve as a cell ages. This is best explained with Fig. 4, where we simulate using the half-cell model the dQ/dV curves of a cell which has undergone a 20% decrease from the initial value (loss) for a single degradation parameter while the other two are held constant at a 0% decrease. Depending on the extent to which each degradation mechanism contributes to total cell degradation, the shape of the degraded dQ/dV curve relative to a healthy dQ/dV curve will appear differently. Altogether, using the full dQ/dV curve as input to the machine learning models avoids many problems associated with manual feature selection and improves our methodologies’ applicability to other battery chemistries without extensive changes to the models.

Briefly, some limitations of this work are that it requires using characterization tests under low C-rates and specific cycling (charge/discharge) conditions. That is, the method does not account for cells which may age under different charge/discharge profiles. Future work could extend the method to consider regular cycling data, higher C-rates, and different charge/discharge profiles.

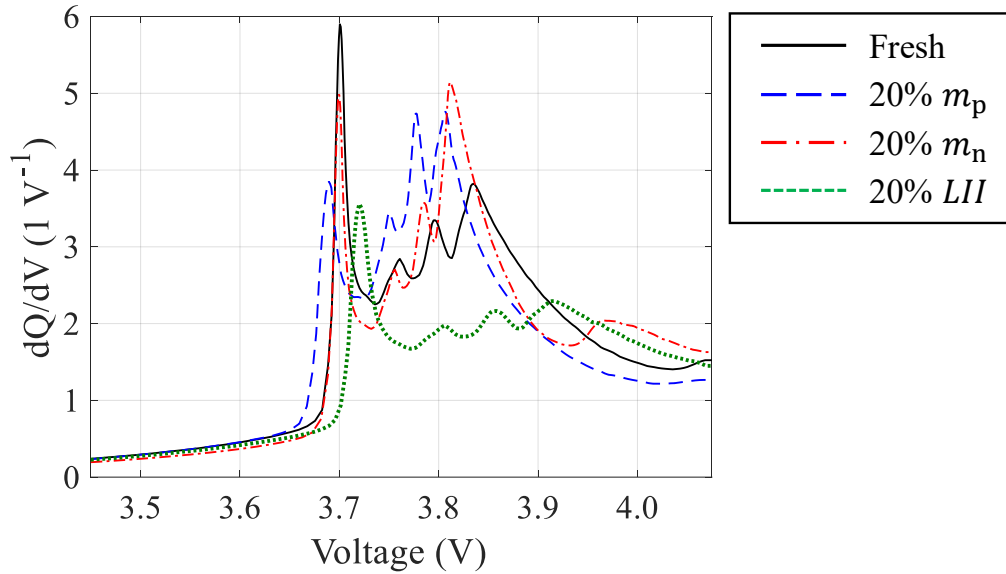


Fig. 4 Simulated dQ/dV curves for a fresh cell and three aged cells, each with a 20% decrease from the initial value of a single degradation parameter.

3.4 Half-Cell Model for Data Generation

Just as the half-cell model can infer the values of the degradation parameters by fitting the measured full-cell VQ and dV/dQ curves, it can also be used in reverse to generate a full-cell VQ curve from a set of degradation parameter inputs. In turn, dV/dQ and dQ/dV curves can be obtained from the generated VQ curve. To create “simulated” input/output training data pairs of dQ/dV curves and degradation parameter tuples, we first define a design space of the three degradation parameters (m_p , m_n , and LII), where the upper and lower limits are selected to be slightly larger and smaller than those observed in the experimental cells, respectively. Any sample in the three-dimensional design space represents a combination of the three degradation parameters. Next, we use Latin hypercube sampling to create tuples of degradation parameter values sampled from the predefined design space. Then, one by one, the tuples of degradation parameter values are input to the half-cell model, and the corresponding VQ curve is obtained for the input. In this way, a single degradation parameter tuple only has one unique VQ (dQ/dV) curve. These generated input/output pairs of dQ/dV curves and degradation parameter tuples are referred to as the physics-based “simulation data” throughout this paper. In the next section, we discuss the SOH estimation problem: estimating the values of these degradation parameter tuples given a dQ/dV curve as input.

The half-cell model equations (see Section 3.2) are implemented in a hand-crafted MATLAB script for simplicity. In the script, the many randomly generated degradation parameter values are recursively evaluated in the set of equations to generate the corresponding VQ, dV/dQ, and dQ/dV curves used in our machine learning pipeline.

4 SOH ESTIMATION PROBLEM FORMULATION

4.1 Machine Learning Models for SOH Estimation

Consider the SOH estimation problem is to learn the relationship between the inputs and outputs of N distinct training samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^L$ and $\mathbf{y}_i \in \mathbb{R}^T$. Here, L is the number of input features, where each feature is a dQ/dV value sampled at an evenly spaced voltage interval across the entire dQ/dV curve. Likewise, T is the number of tasks, equal to four in this study, as we aim to estimate cell capacity Q and the values of the three degradation parameters m_p , m_n , and LII , corresponding to the dQ/dV curve input. Then, the t th output can be represented by an output function $y_t = f_t(\mathbf{x})$, where $t = 1, \dots, T$.

In this work, we consider four different machine learning models, two parametric and two non-parametric. To keep the algorithms lightweight and easily embeddable on a BMS, we opt to use traditional machine learning models instead of more complex deep learning ones. More complex deep learning models can sometimes have better predictive accuracy, but they require far more training data than we have available (only 16 cells, roughly 450 datapoints altogether). The two parametric models employed for the SOH estimation task are multi-task versions of the popular lasso and elastic net regression models. These two models were selected because they are easy to implement, have built-in variable selection and regularization, and have a small number of parameters, making them computationally efficient and easily embeddable on a BMS. By simultaneously considering the fitting of T tasks, these models may learn a more general representation of the data, which may, in some cases, help improve the model’s performance on new, unseen test data. Likewise, the linear nature of these models allows them to extrapolate beyond the training data distribution more easily, making them excellent candidates to accurately predict late-life cell degradation. Detailed mathematical explanations of the lasso and elastic net models can be found in Appendix A.1. In each model, there exists a parameter which controls the extent that weight regularization is applied during training. Since elastic net has two regularization terms, there is an additional parameter which balances the ratio of the two applied weight regularization terms. The weight regularization hyperparameters help prevent overfitting of the training data, which can lead to poor performance on training data but better performance on unseen test data. In both cases, these hyperparameters need to be optimized prior to testing. For each of the methods outlined in Sections 4.3 and 4.4, a unique cross-validation study was run 50 times using a dataset which was carefully selected to be a fair combination of all the data and methods we aim to test in this study. After determining the hyperparameters for the two methodologies, they remain fixed for all the tests.

The first of the non-parametric models employed for the SOH estimation task is a multi-output Gaussian process (MOGP) regression model. Analogous to the multi-task lasso and elastic net models, the MOGP was chosen because it considers the covariance between tasks, potentially leading to a more generalizable model. In this study, the commonly used radial basis function (RBF) kernel is used for all MOGP models. During training of the MOGP models, the kernel parameters are learned from the data, and as a result, no manual hyperparameter optimization is needed. To enable the multi-output structure of the Gaussian process, a symmetric intrinsic model of coregionalization (ICM) is used to learn the relationships between tasks. The ICM method uses another GP model to learn the covariance between tasks, enabling multiple outputs. For a brief introduction to this concept, and related Gaussian process mathematics, see Appendix A.2. For a more detailed description of the intrinsic model of coregionalization, see [43].

The second of the non-parametric models employed is an extreme learning machine (ELM). An ELM is a single hidden layer feedforward neural network with random input weights and biases that do not change throughout the training process. The ELM was selected because its underlying neural network structure allows it to approximate any complex non-linear mapping directly from input samples without having to perform the traditional lengthy backpropagation training process of typical neural networks. Neural networks like ELM are not known for their extrapolation, so it will be useful to examine them for use as SOH estimation models, especially in a scenario with limited training data. A detailed description of the mathematical formulation for an ELM can be found in Appendix A.3. For an ELM model, the main tunable hyperparameter is the number of hidden neurons in the model. To select the optimum number of hidden neurons, a cross-validation study was run 50 times using a selected training and validation dataset similar to lasso and elastic net. The optimization was done independently for each of the two different methods proposed in Sections 4.3 and 4.4. Once the optimum number of neurons is found, it is fixed for all training configurations and tests.

4.2 Input and Output Variables

This study aims to estimate the three degradation parameters and capacity of a cell based on its dQ/dV curve by using machine learning models to learn the correlation between the two. The input features of the machine learning models are the dQ/dV readings, calculated by differentiating the capacity over the voltage. A sampling interval, ΔV , is defined to calculate the dQ/dV value at the specified voltage, $[V_{LC}, V_{LC} + \Delta V, V_{LC} + 2\Delta V, \dots, V_{LC} + (L - 1)\Delta V]$, within the lower and upper cutoff voltages (V_{LC} and V_{UC} , respectively), where $(L - 1) = \left\lfloor \frac{V_{UC} - V_{LC}}{\Delta V} \right\rfloor$. Here, dQ/dV analysis possesses several benefits compared to VQ analysis and dV/dQ analysis. Compared to VQ analysis, dQ/dV analysis transforms the phase equilibrium of active electrode materials to identifiable peaks in the dQ/dV curve which are sensitive to small changes in the materials. Machine learning models can learn these small changes to improve diagnostic accuracy. Unlike dV/dQ and VQ curves, the dQ/dV curve is dependent on a fixed, directly measurable voltage range (e.g., 3.4 V – 4.075 V considered in this study) instead of a capacity range. Using a voltage basis instead of a capacity basis is ideal because the capacity could vary over the course of cell aging and be unmeasurable due to an unknown initial capacity in a partial charge cycle. The dQ/dV values are sampled at fixed evenly spaced voltages between 3.4 V – 4.075 V, and the L sampled points are considered the input features to be fed into a machine learning model. The larger the value L , the higher resolution the dQ/dV curve possesses.

Two datasets are considered in this study: 1) a simulation dataset generated using the half-cell model and 2) an experimental dataset collected from the 16-cell aging test described earlier in Section 2. Table 2 shows the overall size and origin of each dataset. Use of the datasets is described in the following sections.

Table 2 Summary of the two datasets used in this study.

Dataset	Simulation data	Experimental data
Number of cells	NA	16
Data size	10,000	434 (up to 3.5 years of aging test)
Source	Half-cell model	Implantable-grade Li-ion cells (see Table 1)

4.3 SOH Estimation Methodology 1 – Augmentation

Data augmentation refers to a number of methods which aim to improve machine learning model accuracy by increasing the number of training data available [44]–[46]. One popular method of performing data augmentation is to perturb the existing dataset with noise, quickly doubling the number of samples. Another method, and the one used in this work, is to use a simple and computationally efficient physics-based simulation model to generate many new input/output samples that are similar to the original dataset.

In the context of this study, if we train a machine learning model with only the early-life battery degradation data from the cycling experiment (roughly 60 data points), we do not expect the model to provide an accurate estimation of the battery cell’s SOH at late life because the degradation trends change over time. The limited data combined with the complex and non-linear degradation trends observed in Li-ion cells makes this SOH estimation task challenging. With this in mind, we augment the training dataset by including simulation data from the physics-based half-cell model. The simulation data, sampled from the design space described in Section 3.4, will expand both the observed range of degradation and the number of samples seen by the machine learning model. The augmented dataset will contain more information about the future degradation path of the experimental cells than was learnable from only the early-life experimental data. The simulation data from the half-cell model can indirectly incorporate degradation physics into the data-driven models, thus facilitating more accurate estimation of the late-life health parameters. In this way, the augmented machine learning model directly leverages the experimental data and simulation data to learn the observed degradation trend. A high-level overview of this methodology is shown in Fig. 5.

To extensively evaluate the data augmentation methodology, we create and test a number of training dataset configurations to study the effect of dataset composition on model performance. We label each configuration as $\text{EXP}_{\text{N}_{\text{EXP}}} \text{Sim}_{\text{N}_{\text{Sim}}}$, where

N_{EXP} is the number of experimental data points (early-life data) from each cell included in the training dataset and N_{Sim} is the number of randomly selected simulation data included in the training dataset. To thoroughly understand the performance of the model with different combinations of simulated and experimental training data, we implemented the datasets described below.

In the first few dataset configurations, we explore baseline models trained using simulation data only ($N_{\text{EXP}} = 0$). These configurations explore the accuracy of the half-cell model in its ability to mirror the true degradation observed in the experimental cells. Next, we investigate a second set of baseline models trained using only early-life experimental data ($N_{\text{Sim}} = 0$). These tests help expose the models' extrapolation limits in estimating late-life degradation. To form a standard across the baseline models and the two proposed methodologies, we fix the early-life experimental data to be the first five data points from each cell ($N_{\text{EXP}} = 5$). For the experimental data used in this study, five data points are equivalent to approximately three months of lab testing. Using only five experimental data points from each cell mimics a scenario where resources are scarce and testing time is limited.

In some cases, simulation data from a physics-based model may not accurately describe all phenomenon observed in the experimental data. In this case, combining dissimilar simulation data with accurate experimental data only serves to dilute the concentration of information from the experimental data. Consequently, a machine learning model trained on the combined dataset will not learn the correct trend of the experimental data because of the "noise" introduced by the simulation data. To understand the extent of this effect on the SOH estimation problem, in the next few dataset combinations, we limit the amount and type of simulation data augmented to the training dataset. The large 10,000 datapoint simulation dataset is filtered to only contain data points corresponding to the highest 20% degradation in m_p , m_n , and LII . This new subset of high-degradation data is labeled as HiDeg (short for High Degradation). It is believed that by combining the early-life experimental data with simulated HiDeg data, we may be able to coerce the machine learning models into learning a more accurate degradation trend, further increasing their ability to extrapolate and accurately estimate late-life SOH. By choosing to add a limited number of HiDeg simulation data to form an augmented dataset, the information in the experimental dataset may not be diluted as severely as some of the other dataset combinations mentioned earlier. In turn, this may improve model prediction accuracy.

4.4 SOH Estimation Methodology 2 – Delta Learning

Delta learning is a bias-correction method that is often used with physics-based models. In delta learning, the inaccuracies and bias of a physics-based model are learned by a secondary machine learning model, so that they can be corrected, and the overall delta learning model is then more accurate [47], [48]. In a more general sense, delta learning involves an estimation model and a corrector model, where the two models are used in series to yield more accurate predictions.

In our specific implementation of delta learning, the estimation model will be a machine learning model trained using simulation data to predict the late-life capacity and degradation parameters of the experimental cells. However, the simulation is largely imperfect and does not accurately represent the heavily aged dQ/dV curves of the experimental cells, so the estimation model alone cannot accurately estimate the experimental data. To account for this, a corrector machine learning model is trained using light-degradation data to learn the "delta" or prediction error of the estimation model on the light-degradation data. The corrector model will take advantage of the available early-life experimental data and will use it to learn the estimation model's prediction bias. In testing, the two models predict in tandem, where the final prediction is the addition of the estimation and corrector models. This methodology is illustrated in Fig. 5. In the delta learning methodology, the early-life experimental data is leveraged to train the corrector model to learn the difference between the estimator predictions and the observed experimental data. This method of using the available experimental data is different from data augmentation in that the two datasets remain separate. This difference will make for an interesting comparison.

In this implementation, the input to both the estimation and corrector machine learning models is solely a dQ/dV curve. In some instances of delta learning, it may be beneficial to include the estimation model's prediction as input to the corrector model. However, in our preliminary testing, it was found that the many dQ/dV values provided more than enough information, and introducing the estimation model's output as an additional input feature had little effect on accuracy or generalization performance.

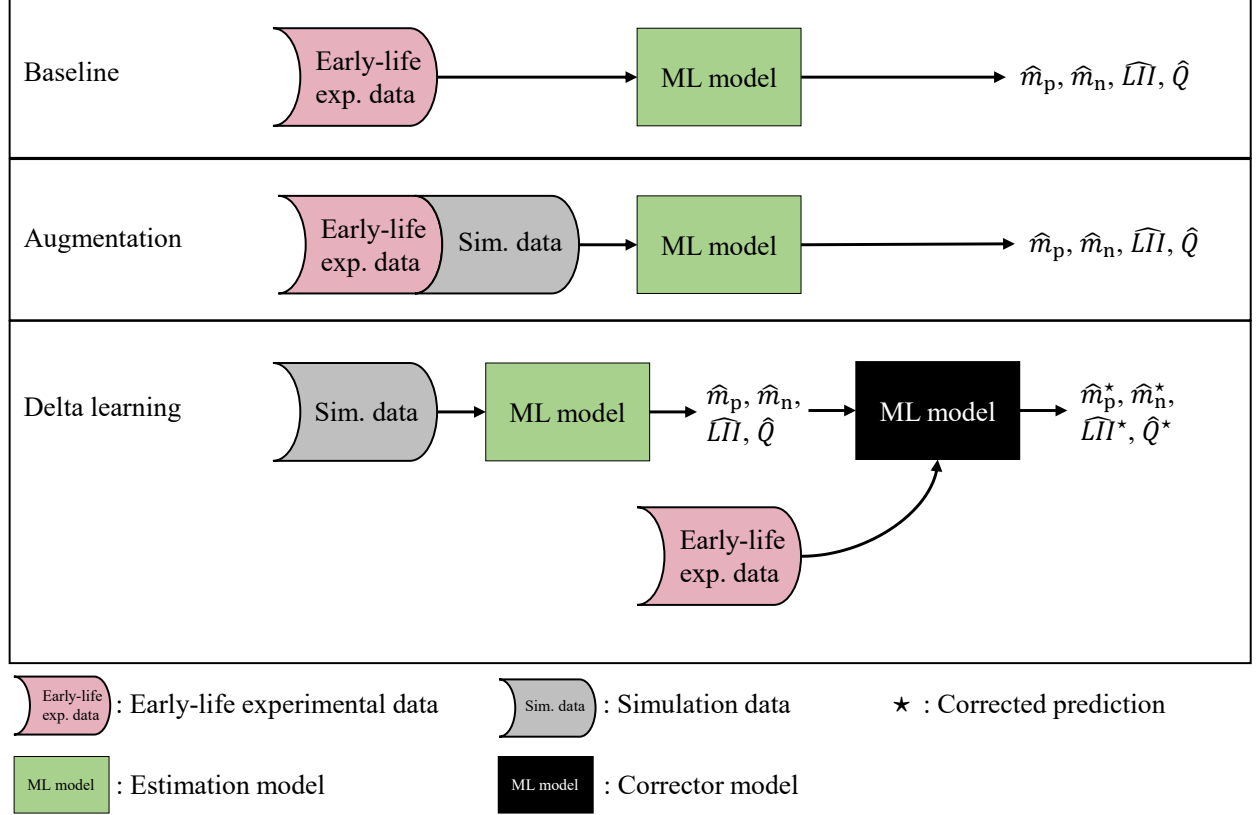


Fig. 5 Block diagram describing the proposed baseline, augmentation, and delta learning methodologies for SOH estimation.

5 RESULTS AND DISCUSSION

5.1 Half-Cell Model Fitting and Degradation Analysis

Here, we discuss the half-cell model fitting results at three characterization tests of interest: the first characterization test (EXP1), the fifth characterization test (EXP5), and the characterization test just before a few of the battery cells underwent the destructive analysis. In Appendix A.4, we show plots of both the VQ and dV/dQ curves from the half-cell fitting process for cell C1 from each of the four groups. We do not show the dQ/dV curves here as they were not used in the half-cell fitting process. For brevity, and because the implantable-grade Li-ion cells are extremely consistent cell-to-cell in the observed capacity fade plot in Fig. 1, we only plot cell C1 from each group, and our analysis hold for all the cells.

Figure A1 in Appendix A.4 shows the VQ and dV/dQ fittings of battery cell G1C1. The first characterization test shows the fitting results when the battery cell is fresh and healthy (see Fig. A1 (a) and (d)). Next, we show the fifth characterization test which is representative of roughly 3 months of cycling (see Fig. A1 (b) and (e)). Third, the final set of plots shows the fitting of the half-cell curves to the experimental data just before the destructive analysis of a few cells (Fig. A1 (c) and (f)). For each plot, we included the experimental data, the simulated curve from the half-cell model, and both the positive and negative half-cell curves used in the half-cell model. The plotting structure is consistent across every group.

For group G1 cells, the experimental and fitted curves agree with each other very well. This cell group has the least degradation among all the four test configurations shown in Section 2 because these cells were cycled at the lower temperature (37 °C) and lower discharge C-rate (C/24). As shown in Figs. A1 (d–f), the dV/dQ curves of the experimental cells at test 19 (day 573) look very similar to those at test 1. Likewise, the fitted VQ curves (labeled as “simulation” in the legend) almost completely overlap with the experimental curves. This observation indicates the half-cell curve has an excellent fit. Furthermore, the two peaks used to gauge the fitting accuracy in the dV/dQ curve (labeled ① and ②) match excellently between the simulated and experimental curves. Turning attention to the capacity fade, we observe a roughly six percent decrease in available capacity of group G1 cells at day 573. Most of the observed capacity-fade is attributed to an increase in

LLI because we observe minimal shift between the positive and negative half-cell curves. Looking at the capacity-fade plot in Fig. 1 (a), the capacity-fade trend of group G1 cells first exhibit a high rate of degradation which quickly slows down at a continuously decreasing fade-rate. This consistently decreasing fade-rate is best described as following a square-root of time model which is known to describe SEI growth [49].

The cells in group G2 were tested at the same discharge C-rate as group G1 (C/24) but were subjected to a higher ambient temperature of 55 °C. Overall, the fitted curves look like the experimental ones with the exception of the VQ curves at higher degradation (Figure A2 (b) and (c)). Specifically, we can see in Fig. A2 (b) and (c), the middle of the fitted VQ curve has more curvature downwards compared to the experimental one. This larger curvature was likely caused by faster positive active material loss, which caused the positive half-cell curve to shrink and thereby changed the shape of the simulated full-cell VQ curve in this region. The disagreement between the experimental VQ curves and the half-cell model's VQ curves is due to electrode aging. The positive half-cell curve is measured from a fresh electrode which is unlike the aged electrodes of the experimental cells. Our previous study showed that using an aged positive half-cell curve in the half-cell model significantly reduced the disagreement between the experimentally measured VQ curve and the half-cell model's VQ curve [25]. Looking at the capacity-fade of group G2 cells, we observe a roughly 11% decrease in available capacity at day 484 (see Fig. 1 (b)). Like the group G1 cells, the capacity-fade of group G2 cells was also found to follow the square-root of time model. This indicates that the main contributor to the observed capacity-fade was once again LLI. However, in contrast to group G1 cells, the capacity-fade of group G2 cells is faster due to the higher ambient temperature.

For the cells in group G3, the simulated VQ curves look similar to the experimental VQ curves. However, the simulated dV/dQ curves deviate significantly from the experimental dV/dQ curves. This disagreement in the evolution of the dV/dQ curves is likely a result of the cells undergoing faster positive active material loss, which caused the left endpoint of the positive half-cell curve to shift to the right, and the left endpoint of the negative half-cell curve to shift to the left. The first valley of the dV/dQ curve has a higher dV/dQ value than cells in other groups, even at the first characterization test. Therefore, the rising of the first valley's value was in part caused by the shifting of both the positive and negative half-cell curves, and some other internal modes that the half-cell model cannot explain. For the battery cells in group G4, neither the fitted VQ curves nor the fitted dV/dQ curves match the experimental data well. Group G4 cells exhibit similar behavior to group G3 cells in that there is significant positive active material loss. Our analysis attributes the large differences in the fitted and experimental curves to the disappearance of characteristic features caused by the relatively heavy aging of the cells in these two groups. The rising and disappearance of the first valley (before peak ①) of the dV/dQ curve is largely due to the leftward shift of the negative half-cell curve and the rightward shift of the positive half-cell curve. Note that the shifting of the positive half-cell curve is more obvious in group G4 cells, indicating a faster positive active mass loss than group G3 cells. Looking at the capacity-fade trends of group G3 and G4 at days 573 and 484 in Fig. 1, respectively, we observe there are two distinct fade trends. In group G3 cells, the initial fade-rate is higher, and then proceeds to decrease with time. This capacity-fade trend is similar to groups G1 and G2 where the capacity-fade rate is largely found to follow the square-root of time SEI growth model. However, group G4 cells were found to have a fast initial fade-rate which did not decrease much over time. This sustained faster fade-rate is likely caused by the combination effect of LLI and LAM [25].

We further analyze the long-term degradation data collected in this study by plotting the evolution of the fitted full- and half-cell VQ curves over time. Once again, due to the very small cell-to-cell variation observed in the cycling tests, we only plot cell C1 from each group, and our analysis holds for all cells. Figure A5 shows complete voltage curve results from our fitting of cell C1 from groups G1 and G3 (C/24 @ 37 °C and C/3 @ 37 °C respectively). At the top, the fitted voltage curves are shown from test 1 to test 19 where a lighter color indicates the voltage curves of later characterization tests. Both the full-cell and the positive half-cell voltage curves from each group are similar in structure and evolution. However, we observe a larger negative active mass loss when the battery is cycled at a higher C-rate (see figure A5 (b)). The faster negative active mass loss at higher C-rates could be attributed to more material degradation (e.g., graphite particle cracking) caused by a larger number of charge-discharge cycles. During the battery charge-discharge process, the electrode materials undergo intercalation and deintercalation processes, which cause slight expansion and contraction of the material. The continuous expansion and contraction may induce cracks in the material and could, in some cases, cause local delamination [27], [50].

Figure A6 shows the VQ curve evolution for cell C1 from groups G2 and G4 (C/24 @ 55 °C and C/3 @ 55 °C respectively). The voltage curves of group G2 look similar to those of group G1, albeit with group G2 experiencing more capacity fade driven by the higher ambient temperature. The higher aging factors (both higher temperature and C-rate) caused group G4 cells to exhibit a higher degree of degradation, manifested through the obvious positive half-cell curve shrinkage and shift throughout the aging test. However, the increase in temperature did not necessarily increase the rate of negative active mass loss. This is evident when comparing the shrinkage of the negative half-cell curves of group G4 and group G3. The negative half-cell curves of both groups look similar. These results may indicate that cycle number, instead of temperature, is the key driver to negative active mass loss.

Altogether, we find that during normal working conditions (37 °C with a discharge C-rate less than C/24), the implantable-grade Li-ion cells tested in this study could last for a very long time with negligible capacity fade (see Fig. A5 (a) and Fig. 1).

5.2 Error Metrics

To evaluate the performance of the machine learning models, we perform a four-fold cross-validation study. In particular, the complete experimental dataset consisting of 16 cells is divided into four mutually exclusive folds. The test data in a single fold consists of one battery cell (see Table 3) from each group in the test matrix (Table 1). The overall test errors $\varepsilon_{t, \text{RMSE}\%}^{\text{All}}$ of the machine learning models are calculated by taking the average of the individual health parameter test errors across the four folds:

$$\varepsilon_{t, \text{RMSE}\%}^{\text{All}} = \sqrt{\frac{1}{\sum_{k=1}^4 N_k} \sum_{k=1}^4 \sum_{i=1}^{N_k} \left[\left(\frac{\hat{y}_{ti} - y_{ti}}{y_{ti}} \right) \times 100\% \right]^2} \quad (3)$$

where the subscript t denotes the t th health parameter, $\hat{y}_{ti}$ and y_{ti} denote the predicted value and true value for the t th degradation parameter at the i th test point, respectively, and N_k denotes the number of test samples. Using a normalized error metric is important when the outputs have different magnitudes and comparison among them is desired.

Table 3 Summary of online cells (test data) in each cross-validation fold.

Discharge rate	C/24		C/3	
Temperature	37 °C	55 °C	37 °C	55 °C
Fold 1	G1C1	G2C1	G3C1	G4C1
Fold 2	G1C2	G2C2	G3C2	G4C2
Fold 3	G1C3	G2C3	G3C3	G4C3
Fold 4	G1C4	G2C4	G3C4	G4C4

For every model configuration shown in the next sections, the machine learning models are run 50 times, and the errors are averaged over the 50 runs. Computing an average RMSE% over 50 runs ensures we examine more stable behavior of the models. This method will better account for the random selection of simulation data points that are included in the model’s training procedure.

5.3 Baseline SOH Estimation Models

For both the data augmentation and delta learning methodologies, we are most interested in determining whether the inclusion of simulation data from the physics-based half-cell model improves machine learning model accuracy at the SOH estimation task. Therefore, we first establish a set of baseline models and benchmark their performance for later comparison. Figure Fig. 6 shows the health parameters estimation accuracy for models trained on either simulation data only or experimental data only. These training data configurations represent naïve approaches to the SOH estimation task. The models trained with simulation data only (SimX EXP0) roughly represent the maximum accuracy of the half-cell model in its ability to accurately estimate the degradation trends observed in the experimental cells.

Briefly, the simulation dataset is generated using a hand-crafted MATLAB script where the 10,000 degradation parameter values are drawn randomly using Latin hypercube sampling (see Section 3.4) and are input into the half-cell model equations to generate the corresponding dQ/dV curves used for training the machine learning models. The simulation is extremely quick, as the half-cell model is not overly complicated and can be considered a simple and computationally efficient physics-based model.

As described in Section 4.2, the simulation dataset encompasses a much wider range of degradation scenarios than the early-life experimental dataset. However, the models trained only with simulation data performed worse when compared to those trained only with early-life experimental data. This lower accuracy can be attributed to the fact that the half-cell model, being a simplified, largely imperfect physics-based model, cannot accurately reconstruct the dQ/dV curves of the experimental cells, especially when the cells have aged heavily (see Fig. 7). In Fig. 7, the locations of features in the simulated dQ/dV curves roughly match those in the experimental dQ/dV curves; however, the magnitudes are vastly different. With inaccurate dQ/dV curves as input, the machine learning models have difficulty accurately estimating the degradation parameters. We have discussed such observation in our previous paper [25]. Due to the disagreement between the simulation and the experimental data, the health parameters estimation errors in Fig. 6 (a-d) do not possess clear decreasing trends as the amount of simulation data increases.

In contrast, the estimation errors of the models trained with only early-life experimental data (EXPX Sim0) show obvious decreasing trends as more experimental data are used for training. However, using 12 or more experimental data points (an approximately 14-month or longer test time) is often not feasible because of the extensive time, money, and labor required to perform the tests. In response to this issue, and as previously mentioned, we select EXP5 as the standard amount of experimental data used for both data augmentation and delta learning. As noted before, the first five data points from each cell account for approximately three months of test time.

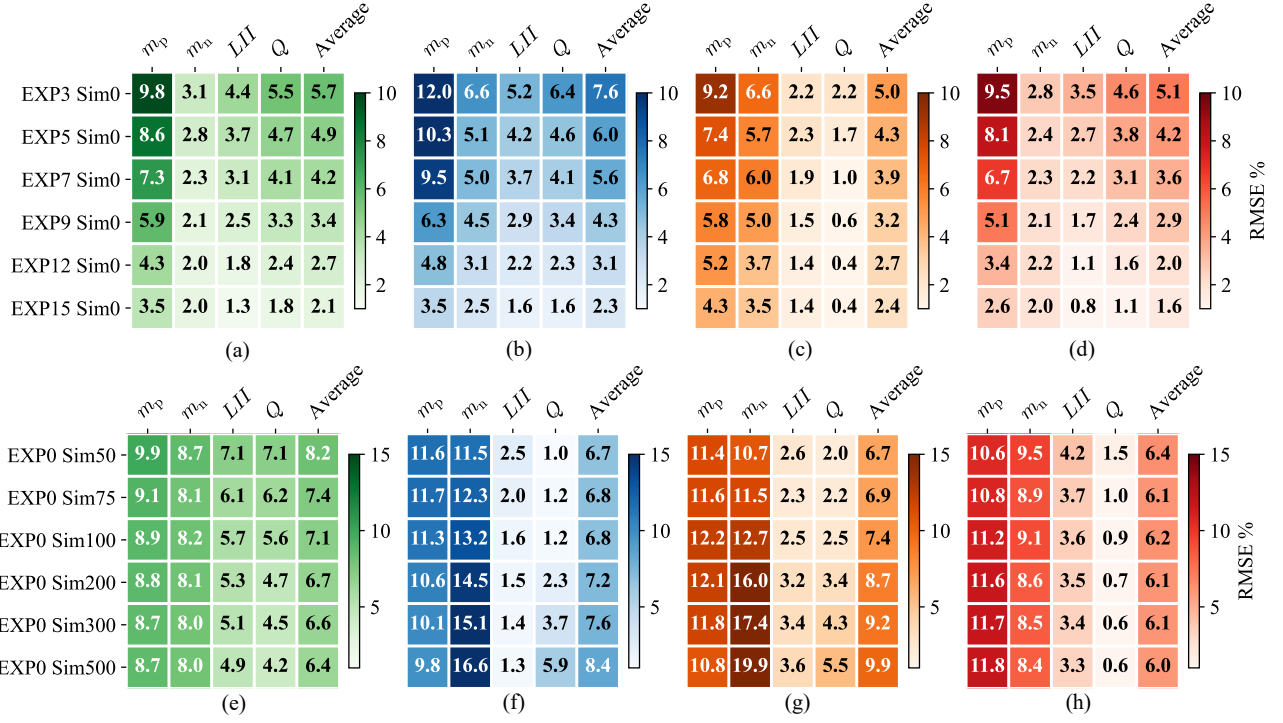


Fig. 6 RMSE % of models when trained exclusively on half-cell model simulation data for (a) Lasso, (b) MOGP, (c) ELM, and (d) Elastic net. RMSE % of models when trained exclusively on experimental data for (e) Lasso, (f) MOGP, (g) ELM, and (h) Elastic net.

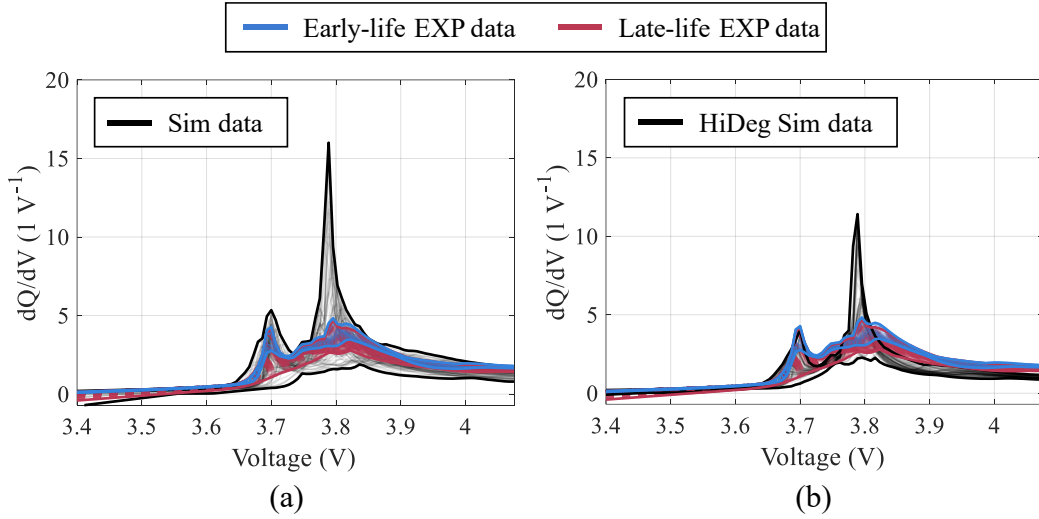


Fig. 7 Comparison of experimental and simulated dQ/dV curves where the simulated curves are randomly selected from either (a) the entire simulation dataset or (b) the high-degradation dataset. For clarity, the early-life EXP data are plotted in a different color so that the evolution between early- and late-life experimental dQ/dV curves is clearer.

5.4 Physics-Informed SOH Estimation Models

Table 4 summarizes results from extensive parametric studies on the data augmentation and delta learning methodologies. The parametric studies investigate using different amounts of simulation and experimental data to train the machine learning

models. Detailed plots and explanations of the studies performed can be found in Appendices A.5–A.7. In Table 4, we summarize the best performing dataset combinations for each machine learning model and methodology. From Table 4, it is evident that both data augmentation and delta learning are successful methods for improving SOH estimation accuracy. Every machine learning model tested (except ELM delta learning) benefits from including the physics-based simulation data into the training procedure, evident by the lower RMSE% values compared to the baseline models.

The minor change in RMSE% of ELM and MOGP with the delta learning methodology is due to the limited extrapolation ability of the individual models. The MOGP and the ELM models are known to not extrapolate well because of their mathematical formulation. Furthermore, in the delta learning method, the two datasets are split up, and the estimation and correction models can only train on either the simulation data or the early-life experimental data. This treatment creates a *separation of information*, where each model in the delta learning method is limited in what data it sees, making extrapolation difficult. The MOGP and ELM estimation models trained using simulation data only struggle to accurately and consistently predict the late-life degradation parameters. As a result, the MOGP and ELM corrector models have difficulty improving the estimator models’ predictions. Altogether, the two-step process with unshared information between datasets made this method far less effective for the MOGP and ELM.

In contrast, the lasso and elastic net methods perform better with the delta learning method because they already extrapolate well using the simulation data. In addition to this, they can also take advantage of the early-life experimental data with the corrector model. Both lasso and elastic net can extrapolate the estimation model error learned from the early-life data to late-life, yielding lower prediction errors.

Table 4 Estimation accuracy of best performing machine learning models.

Model	Methodology	Training Data	m_p	m_n	LII	Q	Mean
Lasso	Baseline	EXP5 Sim0	8.55	2.75	3.68	4.74	4.93
Lasso	Augmentation	EXP5 Sim60	8.49	5.30	1.39	1.68	4.21
Lasso	Augmentation	EXP5 HiDeg Sim20	5.43	4.05	1.63	2.12	3.31
Lasso	Delta Learning	EXP5 Sim500	7.50	3.88	1.26	0.76	3.35
MOGP	Baseline	EXP5 Sim0	10.3	5.05	4.18	4.58	6.03
MOGP	Augmentation	EXP5 Sim80	5.51	4.59	1.37	0.49	2.99
MOGP	Augmentation	EXP5 HiDeg Sim60	5.06	4.72	1.43	0.68	2.97
MOGP	Delta Learning	EXP5 Sim300	8.15	7.14	1.98	0.82	4.52
ELM	Baseline	EXP5 Sim0	7.41	5.71	2.33	1.71	4.29
ELM	Augmentation	EXP5 Sim80	5.39	4.73	1.40	0.53	3.01
ELM	Augmentation	EXP5 HiDeg Sim60	4.44	4.91	1.58	0.59	2.88
ELM	Delta Learning	EXP5 Sim100	6.10	6.88	2.28	2.35	4.40
Elastic net	Baseline	EXP5 Sim0	8.05	2.44	2.73	3.76	4.25
Elastic net	Augmentation	EXP5 Sim20	8.55	3.99	1.15	1.19	3.72
Elastic net	Augmentation	EXP5 HiDeg Sim10	4.98	3.40	0.74	1.05	2.54
Elastic net	Delta Learning	EXP5 Sim500	3.66	5.73	1.17	0.40	2.74

Data augmentation with high-degradation simulation data proves to be the most effective method to improve SOH estimation accuracy for every machine learning model tested. The most significant improvement is observed in MOGP, with over a 50% reduction in RMSE% compared to the baseline. For lasso, elastic net, and ELM, the addition of high-degradation simulation data generally helps decrease estimation error, but the performance gains are not as large as MOGP. The large improvement in performance for MOGP is attributed to the increased size and concentration of the dataset the model is trained on, namely HiDeg simulation data. This phenomenon is best described with Fig. 8. In Fig. 8 (a), we see a visualization of the simulation data included in the training dataset of any given data augmentation model. The data are scattered throughout the degradation parameter space and act more like noise instead of helping the machine learning models learn the correct degradation trends. In contrast, the HiDeg simulation data pictured in Fig. 8 (b) are all clustered in the same region. As a result, the machine learning models are able to interpolate between the early-life experimental data and the high-degradation simulation data. Compared to normal simulation data, the HiDeg models learn a degradation trend that more closely resembles the experimental data, making their predictions more accurate. Augmenting each model’s training dataset with HiDeg simulation data incorporates physical knowledge of cell degradation, enabling more accurate estimation of the late-life health parameters.

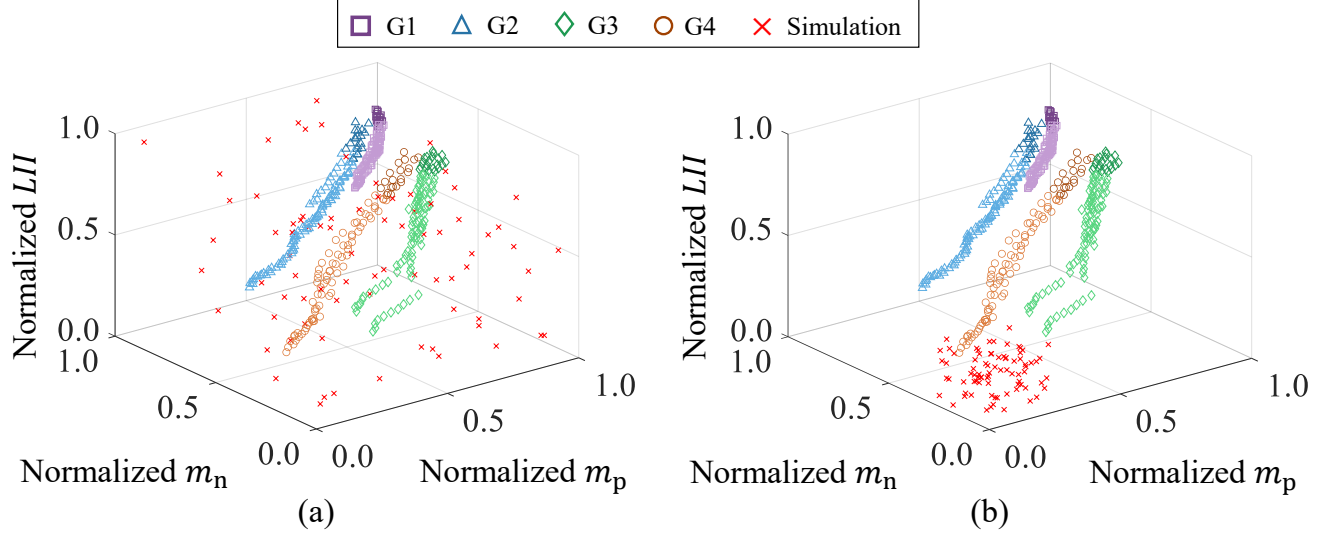


Fig. 8 (a) A single realization of the randomly selected simulation training data, and (b) a single realization of the randomly selected high-degradation (HiDeg) simulation data, both plotted in the degradation parameter space alongside the experimental data.

In selecting the best model, we chose to prioritize all the four SOH estimation tasks equally and selected the best model based on average RMSE% across the four estimation tasks. With these criteria, the best model between the two proposed methodologies is an augmented dataset elastic net model trained with five early-life experimental data and ten high-degradation simulation data. This model achieves an average prediction error of 2.54 RMSE% over the four estimation tasks. The estimation error on late-life LII and Q is excellent, with RMSE% of 0.74 and 1.05, respectively. Below in Fig. 9, we visualize a single instance of the model's parameter predictions. The normalized degradation parameter and capacity predictions vs. time are plotted for every cell and group in Appendix A.8. We can see that the capacity and LII estimation tasks are rather trivial when the model is given even very limited experimental data. This excellent performance may be due to the increased degradation trend information given to the model by including the high-degradation simulation data. Furthermore, there may be a stronger, more easily learned correlation between the high-degradation simulation dQ/dV curves and the cell capacity, which leads to an increase in model performance. More importantly, the best elastic net model did well in estimating the loss of active material on the positive and negative electrodes. These two estimation tasks are not trivial, as the two degradation parameters (m_p and m_n) are not directly correlated to the cell's capacity and exhibit much more complex degradation trends.

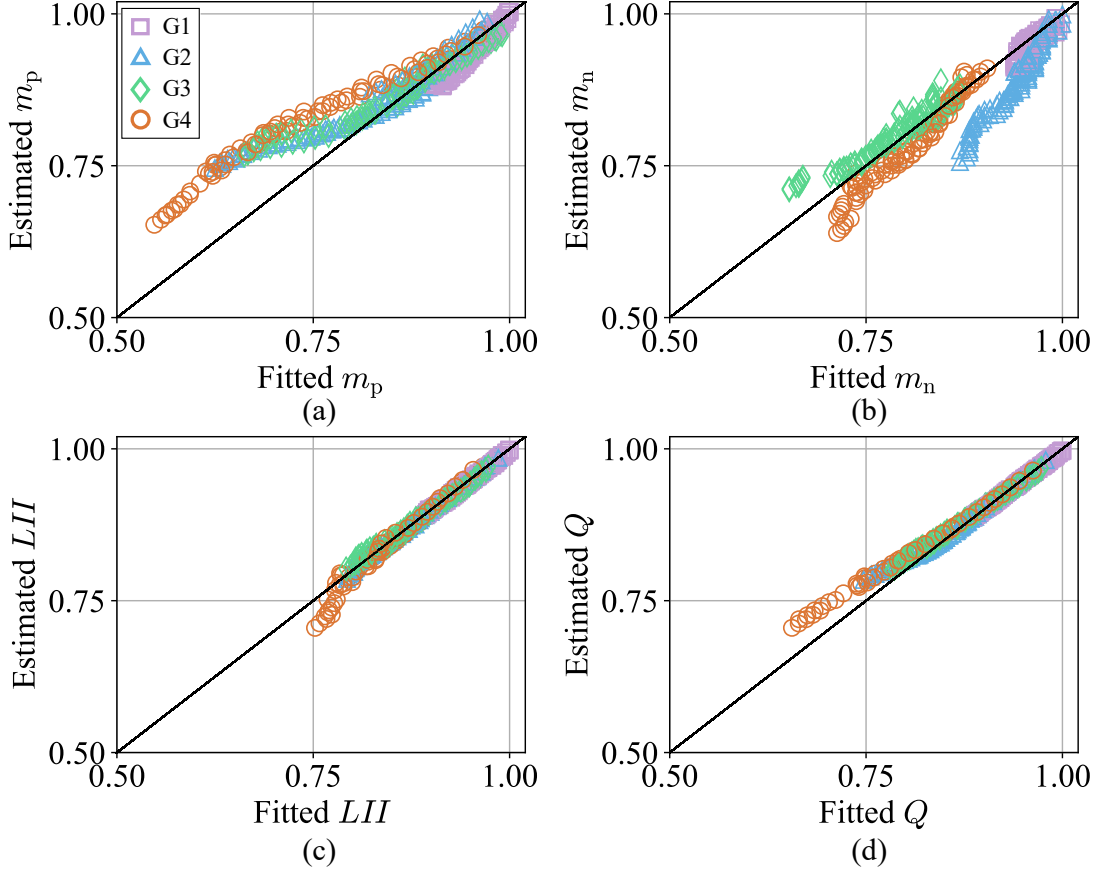


Fig. 9 Normalized estimated vs fitted (true) health parameters predictions for: (a) m_p , (b) m_n , (c) LII , and (d) Q , from the best performing model (elastic net Delta Learning when trained with only the first five experimental data points and ten high-degradation simulation data, EXP5 HiDeg Sim10). In each plot, the predicted health parameters are divided into the four battery groups as shown in Table 1 and Fig. 1.

We also highlight the high accuracy at which all the models and methods were able to estimate capacity. Looking back at Fig. 6(e,f,g,h), the models trained using only simulation data were still able to estimate experimental cell capacity with remarkable accuracy. The highest recorded RMSE% of 7.1 from the lasso model EXP0 Sim50 is reasonable considering the model trained on limited simulation data. The overall accuracy of the capacity estimation task for models trained only on simulation data highlights the usefulness of the half-cell model simulation data. While not perfect, the half-cell model was still able to accurately simulate many VQ and dV/dQ curves under many specific combinations of degradation parameter values which correspond to specific capacity fade trajectories. The machine learning models easily learned the relationship between the simulated curves and the state of cell capacity. Perhaps even more useful is the ability of the GP model to predict the mean and standard deviation of a Gaussian distribution for a given capacity measurement. Below in Fig. 10, we visualize the GP-predicted Gaussian distributions (probability density functions) of capacity for Cell 1 from each group. We plot every fourth capacity measurement and prediction for clarity. The GP model is the best performing, trained using data augmentation with dataset formulation EXP5 HiDeg Sim60. Since the GP model was trained on the early life data, we observe that the first few predictions have very narrow Gaussian probability density functions, indicating little uncertainty about the capacity values. The further away in time we go from the training data, the wider the Gaussian probability density functions become, indicating higher uncertainty in the predicted capacity measurements.

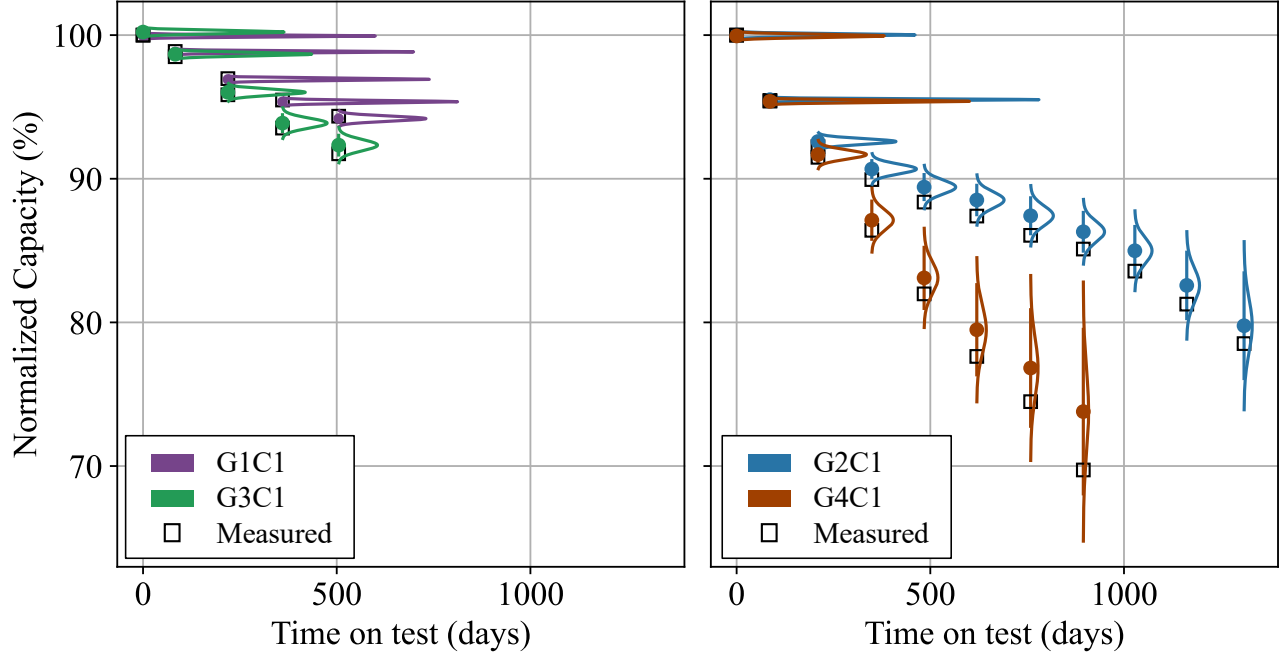


Fig. 10 Every fourth capacity measurement (for distribution clarity) and predicted capacity distribution from the GP model for CV Fold 1. The GP model outputs a mean and standard deviation for every prediction, which are visualized as a Gaussian probability distribution, indicating the uncertainty in the capacity prediction.

Last, we briefly mention that in testing, using the multi-output version of each machine learning model (as opposed to training four individual models) did not provide a substantial improvement in SOH estimation accuracy. This is likely because the information relating two tasks is uninformative [43]. For example, cell capacity decreases almost one-to-one with the LII degradation parameter, but the relationship is mostly uninformative, because both are accurately predicted regardless of training style. The machine learning models were unable to extract any meaningful information to improve accuracy by training the multiple tasks simultaneously. On the other hand, a multi-output model is more computationally efficient to train as it can be done a single time, instead of four separate times for each individual model. In general, multi-task learning is expected to produce a more global model with better generalization capability to new, unseen test data.

6 CONCLUSION

This study has demonstrated the possibility of accurately estimating Li-ion battery capacity and the state of three primary degradation modes by training a machine learning model to leverage both limited early-life experimental data obtained through cycling tests and simulation data from a half-cell model. The resulting light-weight physics-informed machine learning models exhibit improved accuracy compared to models trained strictly on early-life experimental data. Both data augmentation and delta learning were explored as methods to combine the early-life experimental data and simulation data in an intelligent machine learning framework. Each methodology was implemented using four different machine learning models to assess the sensitivity of the methods to the chosen model type and were compared based on their ability to accurately predict the late-life cell degradation parameters and capacity. The first method, data augmentation using high-degradation simulation data, showed to be the most effective method to improve SOH estimation accuracy for every machine learning model tested. The second method, delta learning, successfully improved the SOH estimation accuracy of the Lasso and elastic net linear models because they already extrapolated accurately and benefitted more from having an additional correction model. However, delta learning was less effective for MOGP and ELM because of their inability to extrapolate well. Altogether, both proposed methods were effective at improving SOH estimation accuracy over baseline models without requiring the use of more experimental data.

Overall, our work highlights the value of leveraging inexpensive simulation data from a physics-based model to improve machine learning models' SOH estimation accuracy. The proposed methods successfully increase the SOH estimation models' accuracy at a late aging stage by leveraging available early-life experimental data and physics-based simulation data from a half-cell model. The results in this study suggest that integrating physics-based modeling and data-driven machine learning may enable quick, accurate, and automated online degradation diagnostics of Li-ion batteries when implemented in a BMS. Furthermore, the proposed methodology can significantly reduce the amount of experimental degradation data required for

accurate late-life degradation estimation. In turn, this can reduce the expenses, labor, and time required to characterize cell degradation in a laboratory setting for the purpose of online degradation diagnostics over a cell's lifetime.

ACKNOWLEDGEMENTS

This research was in part supported by the US National Science Foundation (NSF) Grant Nos. ECCS-1611333 and ECCS-2015710. Any opinions, findings, or conclusions in this paper are those of the authors and do not necessarily reflect the views of the sponsoring agency.

REFERENCES

- [1] B. Nykvist and M. Nilsson, "Rapidly falling costs of battery packs for electric vehicles," *Nat. Clim. Chang.*, vol. 5, no. 4, pp. 329–332, Mar. 2015, doi: 10.1038/nclimate2564.
- [2] X. Zeng *et al.*, "Commercialization of Lithium Battery Technologies for Electric Vehicles," *Advanced Energy Materials*, vol. 9, no. 27. Wiley-VCH Verlag, Jul. 19, 2019, doi: 10.1002/aenm.201900161.
- [3] B. Dunn, H. Kamath, and J. M. Tarascon, "Electrical energy storage for the grid: A battery of choices," *Science (80-.)*, vol. 334, no. 6058, pp. 928–935, 2011, doi: 10.1126/science.1212741.
- [4] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, "A review on lithium-ion battery ageing mechanisms and estimations for automotive applications," *J. Power Sources*, vol. 241, pp. 680–689, Nov. 2013, doi: 10.1016/J.JPOWSOUR.2013.05.040.
- [5] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs - Part 3. State and parameter estimation," *J. Power Sources*, vol. 134, no. 2, pp. 277–292, 2004, doi: 10.1016/j.jpowsour.2004.02.033.
- [6] S. J. Moura, N. A. Chaturvedi, and M. Krstić, "Adaptive partial differential equation observer for battery state-of-charge/state-of-health estimation via an electrochemical model," *J. Dyn. Syst. Meas. Control. Trans. ASME*, vol. 136, no. 1, pp. 1–11, 2014, doi: 10.1115/1.4024801.
- [7] R. Xiong, L. Li, Z. Li, Q. Yu, and H. Mu, "An electrochemical model based degradation state identification method of Lithium-ion battery for all-climate electric vehicles application," *Appl. Energy*, vol. 219, no. 5, pp. 264–275, 2018, doi: 10.1016/j.apenergy.2018.03.053.
- [8] C. Hu, B. D. Youn, and J. Chung, "A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation," *Appl. Energy*, vol. 92, pp. 694–704, 2012, doi: 10.1016/j.apenergy.2011.08.002.
- [9] T. Feng, L. Yang, X. Zhao, H. Zhang, and J. Qiang, "Online identification of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction," *J. Power Sources*, vol. 281, pp. 192–203, 2015, doi: 10.1016/j.jpowsour.2015.01.154.
- [10] B. Saha, K. Goebel, S. Poll, and J. Christophersen, "Prognostics methods for battery health monitoring using a Bayesian framework," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 2, pp. 291–296, 2009, doi: 10.1109/TIM.2008.2005965.
- [11] C. Hu, H. Ye, G. Jain, and C. Schmidt, "Remaining useful life assessment of lithium-ion batteries in implantable medical devices," *J. Power Sources*, vol. 375, pp. 118–130, Jan. 2018, doi: 10.1016/j.jpowsour.2017.11.056.
- [12] D. Andre, M. Meiler, K. Steiner, H. Walz, T. Soczka-Guth, and D. U. Sauer, "Characterization of high-power lithium-ion batteries by electrochemical impedance spectroscopy. II: Modelling," *J. Power Sources*, vol. 196, no. 12, pp. 5349–5356, Jun. 2011, doi: 10.1016/J.JPOWSOUR.2010.07.071.
- [13] X. Hu, S. Li, and H. Peng, "A comparative study of equivalent circuit models for Li-ion batteries," *J. Power Sources*, vol. 198, pp. 359–367, Jan. 2012, doi: 10.1016/J.JPOWSOUR.2011.10.013.
- [14] B. Bole, C. S. Kulkarni, and M. Daigle, "Adaptation of an Electrochemistry-based Li-Ion Battery Model to Account for Deterioration Observed Under Randomized Use." Oct. 02, 2014, Accessed: Oct. 24, 2021. [Online]. Available: <https://apps.dtic.mil/sti/citations/AD1002781>.
- [15] A. Braun *et al.*, "The electronic, chemical and electrocatalytic processes and intermediates on iron oxide surfaces during photoelectrochemical water splitting," *Catal. Today*, 2015, doi: 10.1016/j.cattod.2015.07.024.
- [16] S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 260, no. July 2019, p. 114296, 2020, doi: 10.1016/j.apenergy.2019.114296.
- [17] S. Shen, M. Sadoughi, X. Chen, M. Hong, and C. Hu, "A deep learning method for online capacity estimation of lithium-ion batteries," *J. Energy Storage*, vol. 25, no. March, p. 100817, 2019, doi: 10.1016/j.est.2019.100817.
- [18] M. Bercebar *et al.*, "Online state of health estimation on NMC cells based on predictive analytics," *J. Power Sources*, vol. 320, pp. 239–250, 2016, doi: 10.1016/j.jpowsour.2016.04.109.
- [19] C. Hu, G. Jain, C. Schmidt, C. Strief, and M. Sullivan, "Online estimation of lithium-ion battery capacity using sparse Bayesian learning," *J. Power Sources*, vol. 289, pp. 105–113, 2015, doi: 10.1016/j.jpowsour.2015.04.166.
- [20] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *J. Power Sources*, vol. 384, pp. 387–395, Apr. 2018, doi: 10.1016/j.jpowsour.2018.03.015.
- [21] D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn, "Machine learning pipeline for battery state-of-health estimation," *Nat. Mach. Intell.* 2021 35, vol. 3, no. 5, pp. 447–456, Apr. 2021, doi: 10.1038/s42256-021-00312-3.
- [22] R. R. Richardson, C. R. Birkel, M. A. Osborne, and D. A. Howey, "Gaussian Process Regression for in Situ Capacity Estimation of Lithium-Ion Batteries," *IEEE Trans. Ind. Informatics*, vol. 15, no. 1, pp. 127–138, Jan. 2019, doi: 10.1109/TII.2018.2794997.
- [23] G. dos Reis, C. Strange, M. Yadav, and S. Li, "Lithium-ion battery data and where to find it," *Energy AI*, vol. 5, p.

100081, Sep. 2021, doi: 10.1016/J.EGYAI.2021.100081.

- [24] A. Downey, Y.-H. Lui, C. Hu, S. Laflamme, and S. Hu, "Physics-based prognostics of lithium-ion battery using non-linear least squares with dynamic bounds," *Reliab. Eng. Syst. Saf.*, vol. 182, 2019, doi: 10.1016/j.res.2018.09.018.
- [25] Y. H. Lui *et al.*, "Physics-based prognostics of implantable-grade lithium-ion battery for remaining useful life prediction," *J. Power Sources*, vol. 485, p. 229327, Feb. 2021, doi: 10.1016/j.jpowsour.2020.229327.
- [26] X. Han, M. Ouyang, L. Lu, J. Li, Y. Zheng, and Z. Li, "A comparative study of commercial lithium ion battery cycle life in electrical vehicle: Aging mechanism identification," *J. Power Sources*, vol. 251, pp. 38–54, 2014, doi: 10.1016/j.jpowsour.2013.11.029.
- [27] C. R. Birkl, M. R. Roberts, E. McTurk, P. G. Bruce, and D. A. Howey, "Degradation diagnostics for lithium ion cells," *J. Power Sources*, vol. 341, pp. 373–386, 2017, doi: 10.1016/j.jpowsour.2016.12.011.
- [28] M. Dubarry, M. Bercebar, A. Devie, D. Anseán, N. Omar, and I. Villarreal, "State of health battery estimator enabling degradation diagnosis: Model and algorithm description," *J. Power Sources*, vol. 360, pp. 59–69, 2017, doi: 10.1016/j.jpowsour.2017.05.121.
- [29] J. Tian, R. Xiong, W. Shen, and F. Sun, "Electrode ageing estimation and open circuit voltage reconstruction for lithium ion batteries," *Energy Storage Mater.*, 2021, doi: 10.1016/j.ensm.2021.02.018.
- [30] A. Thelen, Y. H. Lui, S. Shen, S. Laflamme, S. Hu, and C. Hu, "Physics-Informed Machine Learning for Degradation Diagnostics of Lithium-Ion Batteries," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2021.
- [31] A. J. Smith, H. M. Dahn, J. C. Burns, and J. R. Dahn, "Long-term low-rate cycling of LiCoO₂/graphite li-ion cells at 55°C," *J. Electrochem. Soc.*, vol. 159, no. 6, pp. A705–A710, Jan. 2012, doi: 10.1149/2.056206jes.
- [32] L. Lu, X. Han, J. Li, J. Hua, and M. Ouyang, "A review on the key issues for lithium-ion battery management in electric vehicles," *J. Power Sources*, vol. 226, pp. 272–288, 2013, doi: 10.1016/j.jpowsour.2012.10.060.
- [33] 2020, "Cyclers." <http://www.novonix.ca/cyclers> (accessed Apr. 06, 2020).
- [34] H. M. Dahn, a. J. Smith, J. C. Burns, D. a. Stevens, and J. R. Dahn, "User-Friendly Differential Voltage Analysis Freeware for the Analysis of Degradation Mechanisms in Li-Ion Batteries," *J. Electrochem. Soc.*, vol. 159, no. 9, pp. A1405–A1409, 2012, doi: 10.1149/2.013209jes.
- [35] I. Bloom *et al.*, "Differential voltage analyses of high-power, lithium-ion cells 1. Technique and application," *J. Power Sources*, vol. 139, no. 1–2, pp. 295–303, 2005, doi: 10.1016/j.jpowsour.2004.07.021.
- [36] R. Fathi *et al.*, "Ultra High-Precision Studies of Degradation Mechanisms in Aged LiCoO₂/Graphite Li-Ion Cells," *J. Electrochem. Soc.*, vol. 161, no. 10, pp. A1572–A1579, 2014, doi: 10.1149/2.0321410jes.
- [37] M. Dubarry, C. Truchot, and B. Y. Liaw, "Synthesize battery degradation modes via a diagnostic and prognostic model," *J. Power Sources*, vol. 219, pp. 204–216, 2012, doi: 10.1016/j.jpowsour.2012.07.016.
- [38] A. Marongiu, N. Nlandi, Y. Rong, and D. U. Sauer, "On-board capacity estimation of lithium iron phosphate batteries by means of half-cell curves," *J. Power Sources*, vol. 324, pp. 158–169, Aug. 2016, doi: 10.1016/J.JPOWSOUR.2016.05.041.
- [39] K. Honkura, H. Honbo, Y. Koishikawa, and T. Horiba, "State Analysis of Lithium-Ion Batteries Using Discharge Curves," *ECS Trans.*, vol. 13, no. 19, pp. 61–73, 2008, doi: 10.1149/1.3018750.
- [40] X. Li, J. Jiang, L. Y. Wang, D. Chen, Y. Zhang, and C. Zhang, "A capacity model based on charging process for state of health estimation of lithium ion batteries," *Appl. Energy*, vol. 177, pp. 537–543, 2016, doi: 10.1016/j.apenergy.2016.05.109.
- [41] Y. Li *et al.*, "A quick on-line state of health estimation method for Li-ion battery with incremental capacity curves processed by Gaussian filter," *Journal of Power Sources*, vol. 373, pp. 40–53, 2018, doi: 10.1016/j.jpowsour.2017.10.092.
- [42] C. Weng, X. Feng, J. Sun, and H. Peng, "State-of-health monitoring of lithium-ion battery modules and packs via incremental capacity peak tracking," *Appl. Energy*, vol. 180, pp. 360–368, 2016, doi: 10.1016/j.apenergy.2016.07.126.
- [43] H. Liu, J. Cai, and Y. S. Ong, "Remarks on multi-output Gaussian process regression," *Knowledge-Based Syst.*, vol. 144, pp. 102–121, 2018, doi: 10.1016/j.knosys.2017.12.034.
- [44] S. Shen, V. Nemani, J. Liu, C. Hu, and Z. Wang, "A hybrid machine learning model for battery cycle life prediction with early cycle data," *2020 IEEE Transp. Electr. Conf. Expo, ITEC 2020*, pp. 181–184, Jun. 2020, doi: 10.1109/ITEC48692.2020.9161647.
- [45] D. A. van Dyk and X.-L. Meng, "The Art of Data Augmentation," <http://dx.doi.org/10.1198/10618600152418584>, vol. 10, no. 1, pp. 1–50, 2012, doi: 10.1198/10618600152418584.
- [46] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," Dec. 2017, Accessed: Oct. 30, 2021. [Online]. Available: <https://arxiv.org/abs/1712.04621v1>.
- [47] M. Aykol *et al.*, "Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime," *J. Electrochem. Soc.*, vol. 168, no. 3, p. 030525, Mar. 2021, doi: 10.1149/1945-7111/ABEC55.

- [48] C. Kunlong, J. Jiuchun, Z. Fangdan, S. Bingxiang, and Z. Yanru, "SOH estimation for lithium-ion batteries: A cointegration and error correction approach," *2016 IEEE Int. Conf. Progn. Heal. Manag. ICPHM 2016*, Aug. 2016, doi: 10.1109/ICPHM.2016.7542828.
- [49] P. M. Attia, W. C. Chueh, and S. J. Harris, "Revisiting the $t^{0.5}$ Dependence of SEI Growth," *J. Electrochem. Soc.*, vol. 167, no. 9, p. 090535, May 2020, doi: 10.1149/1945-7111/ab8ce4.
- [50] J. Vetter *et al.*, "Ageing mechanisms in lithium-ion batteries," *J. Power Sources*, vol. 147, no. 1–2, pp. 269–281, 2005, doi: 10.1016/j.jpowsour.2005.01.006.
- [51] G. Bin Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006, doi: 10.1016/j.neucom.2005.12.126.

APPENDIX

A.1 Multi-Task Lasso and Elastic Net

Lasso and elastic net are two regression methods with built-in variable selection and regularization. Let x_{ij} denote the j th reading (or discretized measurement) in the i th dQ/dV charge curve where $i = 1:N$ and $j = 1:L$. In a multi-task model with T tasks, y_{ti} with $t = 1, \dots, T$ represents an instance of a single task output. With the prior knowledge that the T tasks are related to one another, we define a multi-task regression problem to map from a dQ/dV curve to the four tasks. The multi-task regression model consists of a single design matrix X and multiple task-specific weights β_t :

$$\mathbf{y}_t = X\beta_t + \varepsilon_t \quad (\text{A1})$$

where ε_t is normalized random noise. To improve the model performance and reduce overfitting, the $\mathbf{x}_j$'s are first standardized to have zero mean and unit variance. To find the optimal weights β_t for the T tasks, multi-task lasso solves the following:

$$\min_{\beta} J_1(\beta) = \frac{1}{2} \sum_{t=1}^T \|\mathbf{y}_t - X\beta_t\|_2^2 + \alpha \sum_{j=1}^L \|\beta_{tj}\|_2 \quad (\text{A2})$$

where the term $\sum_{j=1}^L \|\beta_{tj}\|_2$ is a mixed ℓ_1/ℓ_2 -norm which encourages sparsity across related tasks and α is a hyperparameter which controls the extent to which the mixed ℓ_1/ℓ_2 -norm is enforced in the loss function. After iterative testing, it was found that 0.1 is an acceptable value for α . By including another quadratic penalty $\sum_{j=1}^L \|\beta_{tj}\|_2^2$ in the loss function along with an additional hyperparameter γ for adjusting the ratio of the two loss penalties, we arrive at the objective function for a multi-task elastic net model with the following formulation:

$$\min_{\beta} J_2(\beta) = \frac{1}{2} \sum_{t=1}^T \|\mathbf{y}_t - X\beta_t\|_2^2 + \alpha\gamma \sum_{j=1}^L \|\beta_{tj}\|_2 + \alpha(1-\gamma) \sum_{j=1}^L \|\beta_{tj}\|_2^2 \quad (\text{A3})$$

Through iterative testing, it was determined that optimal values for α and γ are 0.05 and 0.05, respectively.

A.2 Multi-Output Gaussian Process

Multi-output Gaussian process (MOGP) extends the Gaussian process (GP) to jointly model a vector of outputs [43]. Let us first consider a single-output GP defined by its zero mean trend function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$. We are interested in the target output $f_t(\mathbf{x})$ which can be expressed as

$$f_t(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (\text{A4})$$

For simplicity, we only consider the squared exponential covariance function in this study. This kernel function takes the following form

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^\top P^{-1}(\mathbf{x} - \mathbf{x}')\right) \quad (\text{A5})$$

where the input variance is σ_f^2 and the length scales are encoded in $P \in \mathbb{R}^{L \times L}$.

Now we consider an MOGP with T outputs. We aim to model the isotopic training set which is now $\mathbf{x}_{t,i} = \mathbf{x}_i$ and $X_1 = \dots = X_T = \bar{X}$. Just like the single-output GP, the T outputs denoted as $\mathbf{f} = [f_1, \dots, f_T]^\top$ also follow a GP model

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k_M(\mathbf{x}, \mathbf{x}')) \quad (\text{A6})$$

where the new multi-output covariance function $k_M(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{T \times T}$ is defined to be

$$k_M(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} k_{11}(\mathbf{x}, \mathbf{x}') & \dots & k_{1T}(\mathbf{x}, \mathbf{x}') \\ \vdots & \ddots & \vdots \\ k_{T1}(\mathbf{x}, \mathbf{x}') & \dots & k_{TT}(\mathbf{x}, \mathbf{x}') \end{bmatrix}. \quad (\text{A7})$$

Each element $k_{tt'}(\mathbf{x}, \mathbf{x}')$ represents the covariance between outputs, or in other words, the similarity between the tasks. Just like a single-output GP, we define the MOGP relationship to be

$$y_t(\mathbf{x}) = f_t(\mathbf{x}) + \epsilon_t \quad (\text{A8})$$

where the term ϵ_t is iid Gaussian noise for each of the T outputs such that $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The corresponding multi-output likelihood function is written as

$$p(\mathbf{y}|\mathbf{f}, \mathbf{x}, \Sigma_s) = \mathcal{N}(\mathbf{f}(\mathbf{x}), \Sigma_s) \quad (\text{A9})$$

where Σ_s is a diagonal matrix which contains the noise information for each output, i.e., $\Sigma_s = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$. Given the original multi-output training dataset, the posterior distribution at a new point $\mathbf{x}_*$ can be written

$$\mathbf{f}(\mathbf{x}_*)|\mathbf{y}, \mathbf{x}, \mathbf{x}_* \sim \mathcal{N}(\hat{\mathbf{f}}(\mathbf{x}_*), \Sigma_*) \quad (\text{A10})$$

The corresponding prediction mean and covariance are, respectively, given as

$$\hat{\mathbf{f}}(\mathbf{x}_*) = K_{M*}^\top [K_M(\bar{X}, \bar{X}) + \Sigma_M]^{-1} \mathbf{y} \quad (\text{A11})$$

$$\Sigma_* = K_M(\mathbf{x}_*, \mathbf{x}_*) - K_{M*}^\top [K_M(\bar{X}, \bar{X}) + \Sigma_M]^{-1} K_{M*} \quad (\text{A12})$$

where $K_{M*} = K_M(\bar{X}, \mathbf{x}_*)$ and has blocks $K_{tt'}(\bar{X}, \mathbf{x}_*) = [k_{tt'}(\mathbf{x}_i, \mathbf{x}_*)]$ for $t, t' = 1, \dots, T$ and $i = 1, \dots, n$.

A.3 Extreme Learning Machine

ELM is a single hidden layer feedforward neural network (SLFN) with random input weights and biases that do not change throughout the training process. Mathematically, the SLFN with N distinct training samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^L$ and $\mathbf{y}_i \in \mathbb{R}^T$, can be formulated as

$$\mathbf{o}_i = \sum_{j=1}^D \beta_j g(\mathbf{w}_j^\top \cdot \mathbf{x}_i + b_j) = \mathbf{y}_i + \epsilon_i, \quad i = 1:N \quad (\text{A13})$$

where $\mathbf{o}_i$ is the output of the network, $\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jL}]^\top$ is the hidden layer weight vector connecting the input neurons and the j th hidden neuron, b_j is the bias for j th hidden neuron, D is the total number of hidden neurons, $g(\cdot)$ is the activation function that is used to introduce nonlinearity to the SLFN, $\beta_j = [\beta_{1j}, \beta_{2j}, \dots, \beta_{Tj}]^\top$ is the output bias vector that connects the j th hidden neuron to output neurons, and ϵ is noise that includes both random noise and noise that depends on variables other than the inputs [51].

The above N equations can be written in the following matrix form for compact representation,

$$\mathbf{H}\beta = \mathbf{O} \quad (\text{A14})$$

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_D \cdot \mathbf{x}_1 + b_D) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_D \cdot \mathbf{x}_N + b_D) \end{bmatrix}_{N \times D} \quad (\text{A15})$$

$$\beta = \begin{bmatrix} \beta_1^\top \\ \vdots \\ \beta_D^\top \end{bmatrix}_{D \times T} \quad \text{and} \quad \mathbf{O} = \begin{bmatrix} \mathbf{o}_1^\top \\ \vdots \\ \mathbf{o}_N^\top \end{bmatrix}_{N \times T} \quad (\text{A16})$$

where $\mathbf{H}$ is called the hidden layer output matrix.

During the training process, ELM uses the Moore-Penrose inverse (pseudoinverse) to find a solution to the linear system shown in equation (4). For many machine learning problems, the number of training instances (i.e., N) is much larger than the number of hidden neurons (i.e., D). Therefore, the linear system is an overdetermined system, and a unique solution is obtained by minimizing the ℓ_2 norm of the vector of training errors:

$$\hat{\beta} = \mathbf{H}^\dagger \mathbf{Y} \quad (\text{A17})$$

$$\mathbf{H}^\dagger = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \quad (\text{A18})$$

where $\mathbf{H}^\dagger$ is the pseudoinverse of matrix $\mathbf{H}$, and $\hat{\beta}$ is the solution of the overdetermined system of linear equations $\mathbf{H}\beta = \mathbf{Y}$. Compared to the model training with back propagation which requires the iterative updating of input weights and biases, the model training with ELM is much faster as it requires only a single process, pseudoinverse, to find the optimum solution.

A.4 Fitted VQ and dV/dQ Curves

Here, we show the experimental VQ and dV/dQ curves, the positive and negative half-cell curves, and the fitted VQ and dV/dQ curves, for three datapoints of significance for cell C1 from each group. The plots show the evolution of the VQ and dV/dQ curves over the course of approximately half the data. The last datapoint collected for any given cell corresponds to the point in time at which two of the cells from each group were removed for destructive evaluation.

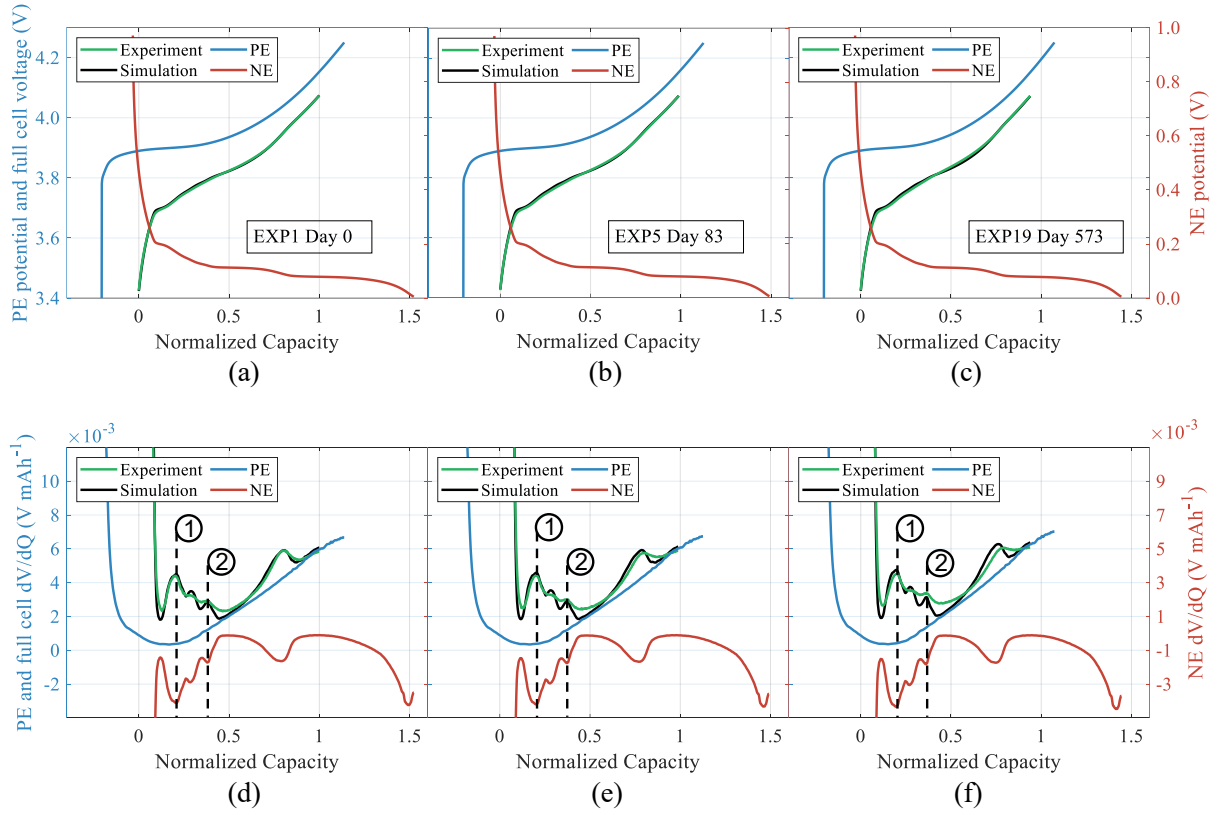


Fig. A1 Group G1 cell C1 (C/24, 37° C) V/Q and dV/dQ experimental and fitted curves (shown in legend as simulation) at (a) EXP1 Day 0, (b) EXP5 Day 83, and (c) EXP19 Day 573. The dashed lines indicate the two peaks used for fitting the dV/dQ curve.

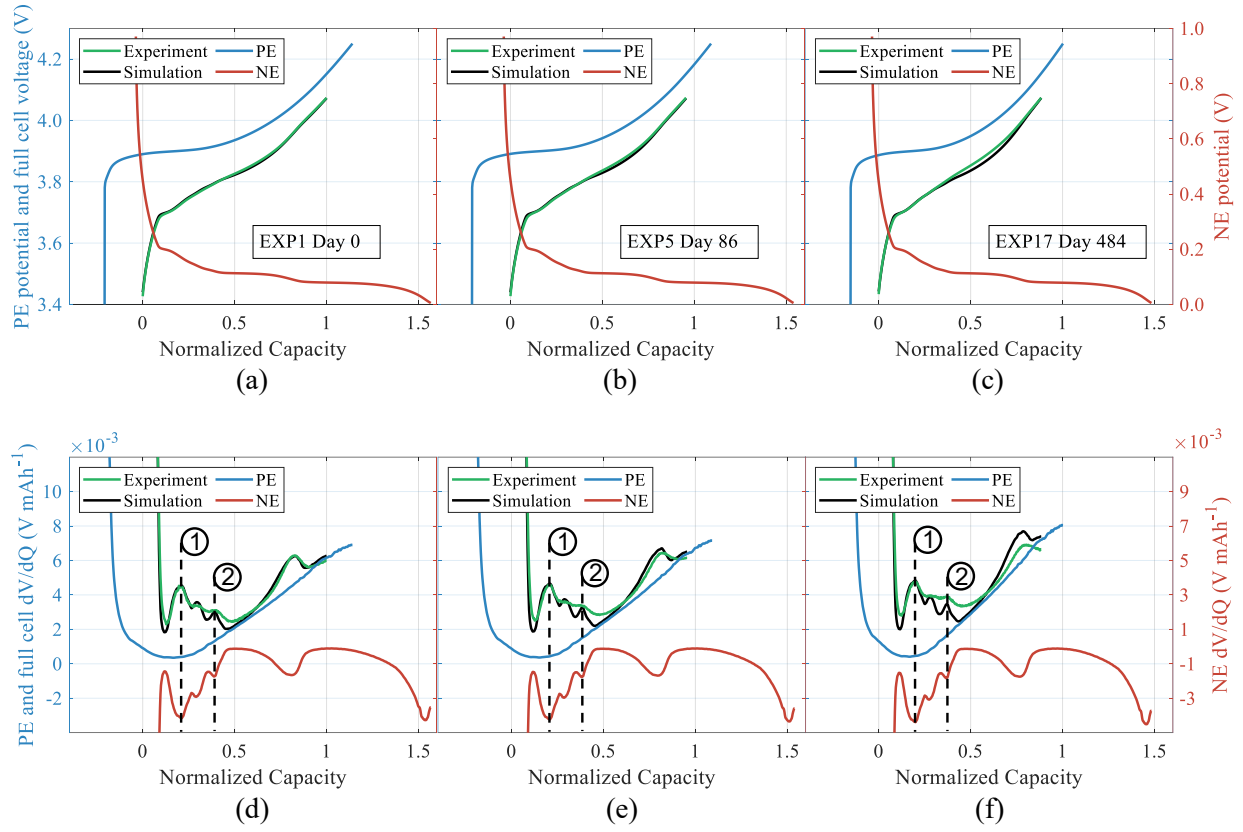


Fig. A2 Group G2 cell C1 (C/24, 37° C) V/Q and dV/dQ experimental and fitted curves (shown in legend as simulation) at (a) EXP1 Day 0, (b) EXP5 Day 86, and (c) EXP17 Day 484. The dashed lines indicate the two peaks used for fitting the dV/dQ curve.

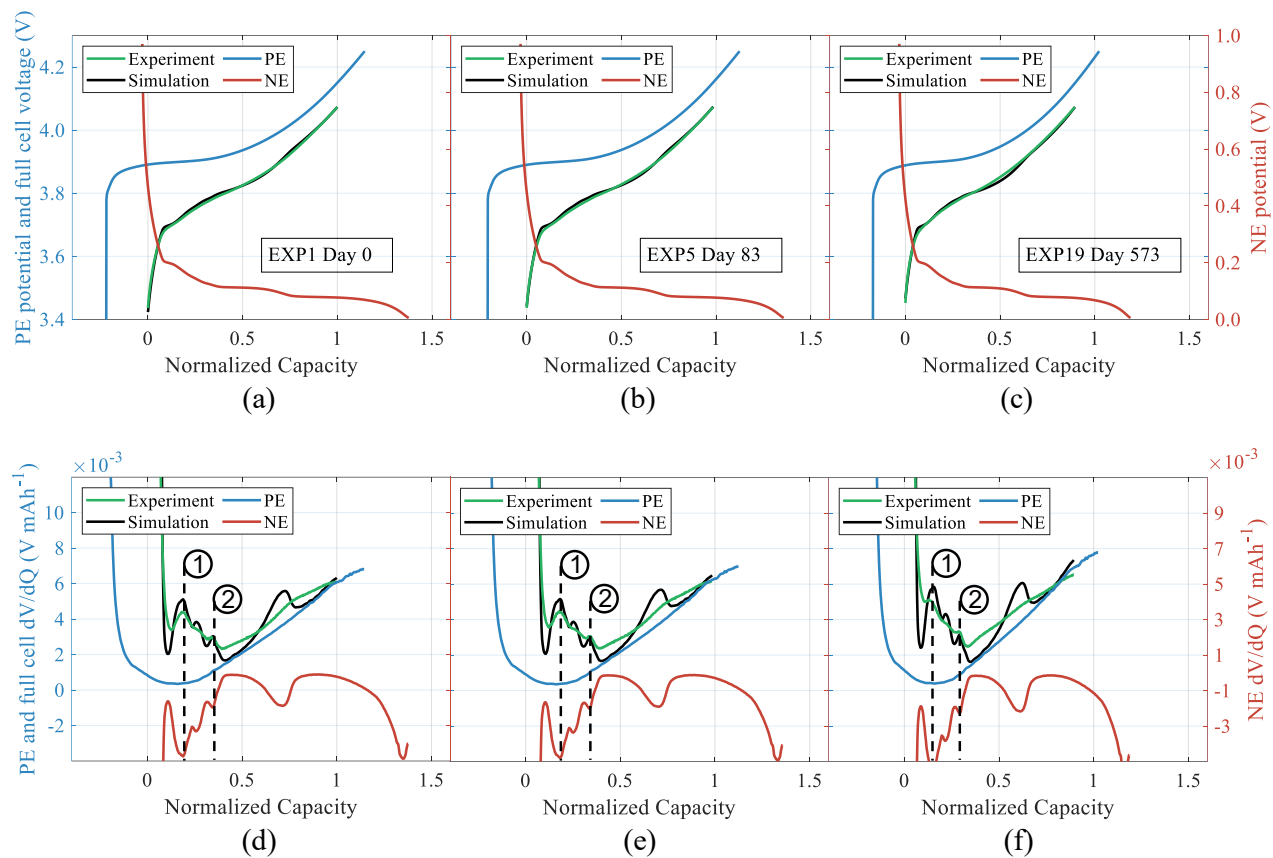


Fig. A3 Group G3 cell C1 (C/24, 37° C) V/Q and dV/dQ experimental and fitted curves (shown in legend as simulation) at (a) EXP1 Day 0, (b) EXP5 Day 83, and (c) EXP19 Day 573. The dashed lines indicate the two peaks used for fitting the dV/dQ curve.

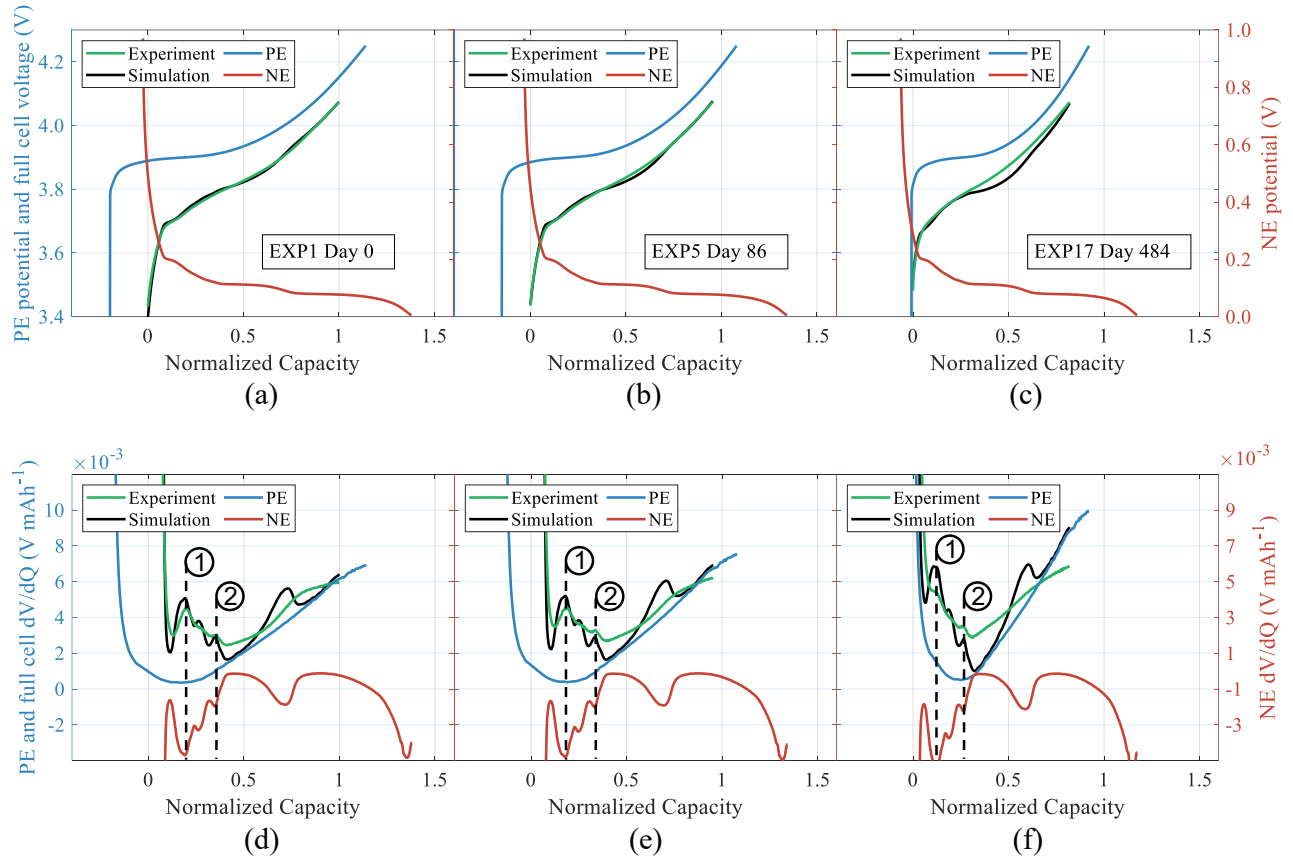


Fig. A4 Group G4 cell C1 (C/24, 37° C) V/Q and dV/dQ experimental and fitted curves (shown in legend as simulation) at (a) EXP1 Day 0, (b) EXP5 Day 86, and (c) EXP17 Day 484. The dashed lines indicate the two peaks used for fitting the dV/dQ curve.

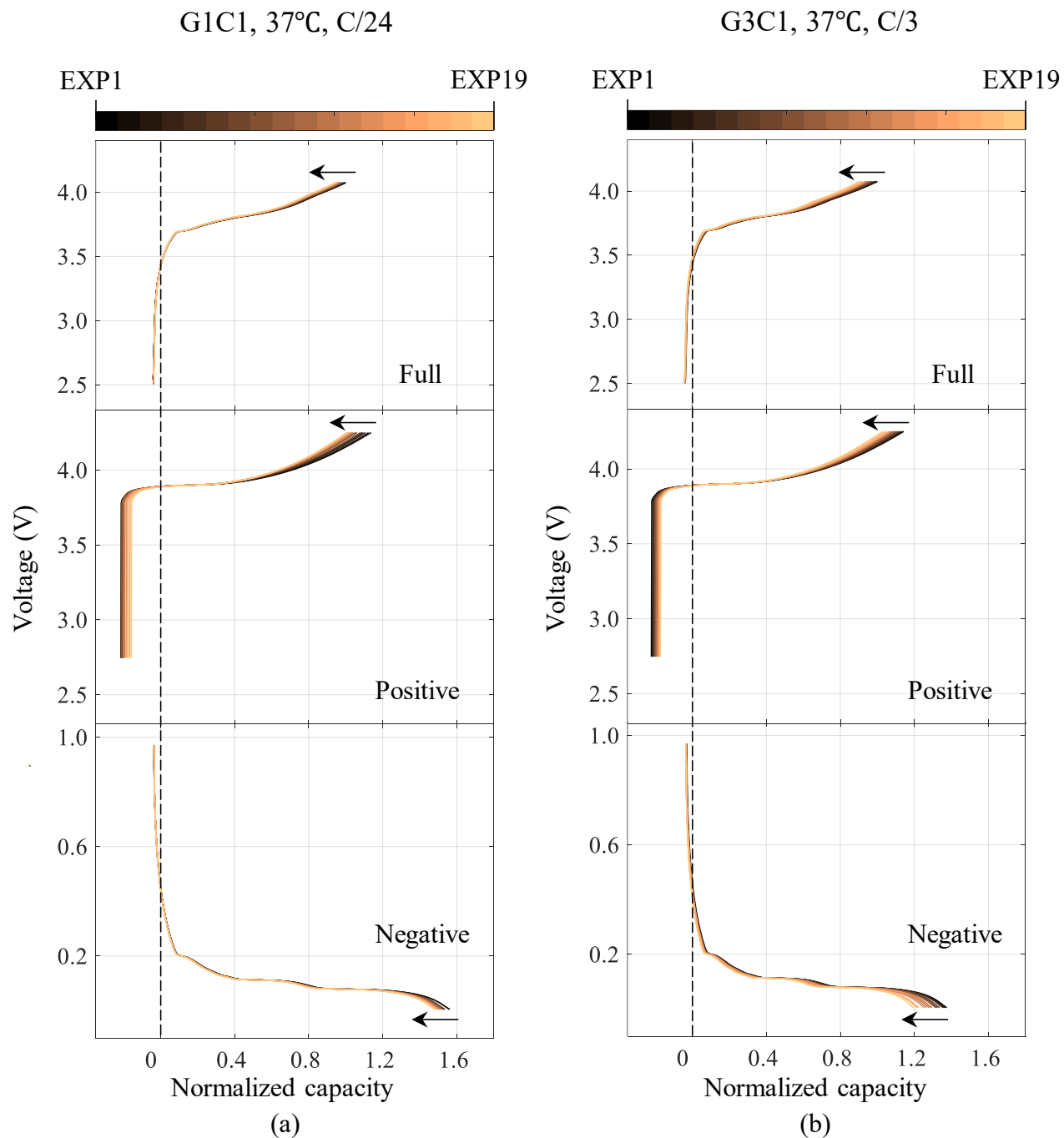


Fig. A5 Evolution of full-cell, positive half-cell, and negative half-cell VQ curves for cell (a) G1C1, (b) G3C1. The arrow in the full-cell VQ curves shows the direction of the curve evolution at the end of charge point.

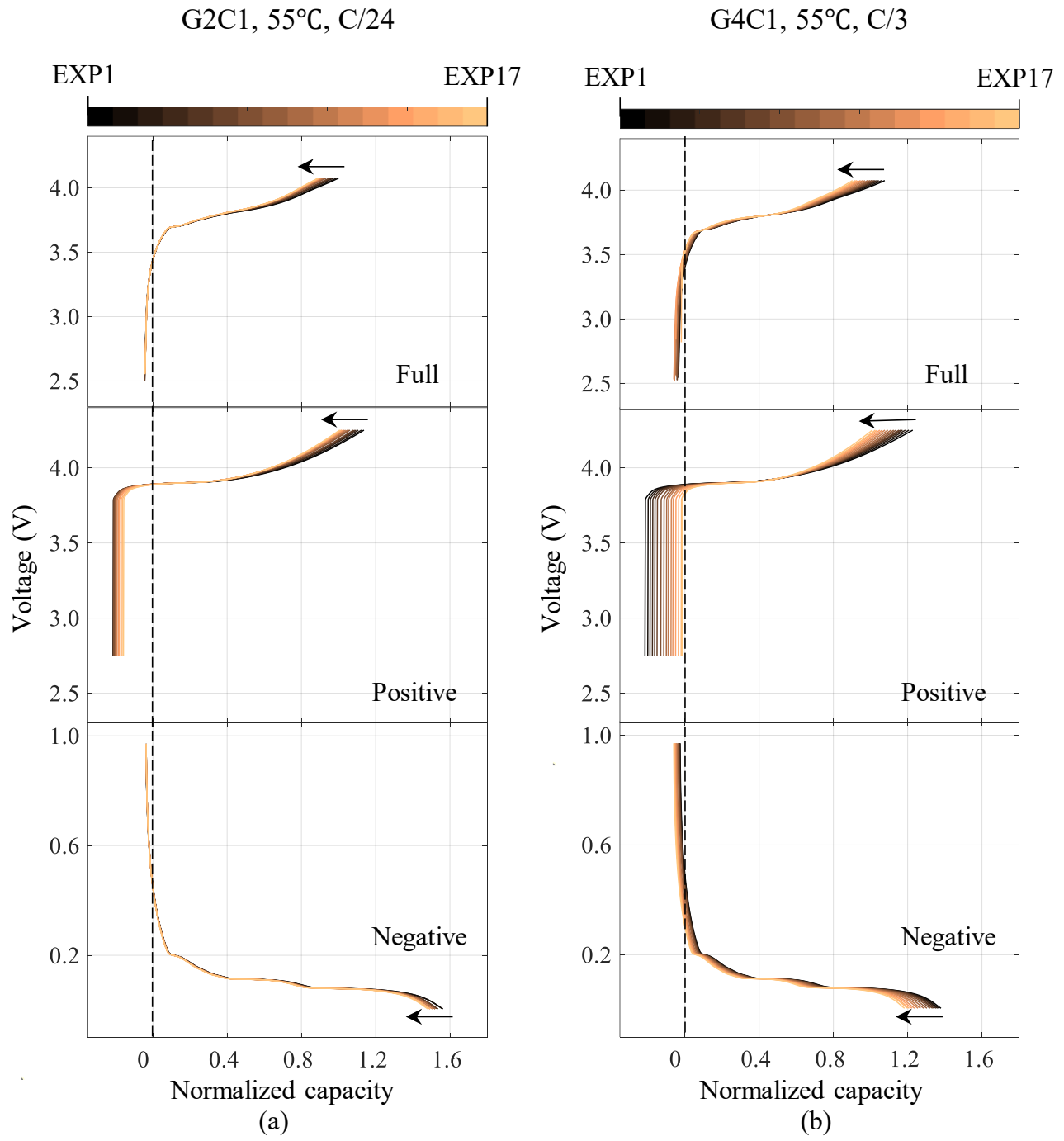


Fig. A6 Evolution of full-cell, positive half-cell, and negative half-cell VQ curves for cell (a) G2C1, (b) G4C1. The arrow in the full-cell VQ curves shows the direction of the curve evolution at the end of charge point.

A.5 Data Augmentation Parametric Study

In the data augmentation methodology, we are most interested in determining whether including simulation data from the physics-based half-cell model improves machine learning model accuracy at the SOH estimation task. Therefore, a suitable baseline to compare the data augmentation methodology to is a model without simulation data. We have previously discussed that collecting five data points from our long-term cell aging experiment is reasonable because it took only three months of test time. Therefore, we establish the baseline model to be the results obtained from EXP5 Sim0. The data augmentation methodology results are displayed in Fig. A7. Here, Fig. A7 (a) shows the SOH estimation errors of different machine learning models trained with different numbers of augmented simulation data included in the training dataset. The error values displayed are the average over the four estimation tasks (m_p , m_n , LII , and Q). When we report dataset compositions as $EXP N_{EXP} Sim N_{Sim}$, it is important to keep in mind that the first five data points from each cell will be 60 total data points. This is because, in each CV fold, twelve cells are used for training while the remaining four are used for testing. From Fig. A7 (c), we observe that the inclusion of simulation data in the training dataset has significantly improved the MOGP and ELM accuracies. However, lasso and elastic net do not benefit nearly as much from this data augmentation strategy. This reduced benefit is because both lasso and elastic net are linear models, and by nature, can easily extrapolate beyond the range of the training dataset. The early-life experimental data (EXP5) proved to be enough data for both lasso and elastic net to learn the degradation trend and provide acceptable accuracy at later aging stages. Conversely, MOGP does not extrapolate well because it is a kernel-based learning algorithm where the chosen kernel, RBF, decays to its mean value, zero, the further we move from the mean of the training data distribution. Likewise, ELM is a simple neural network which is known for only being intelligent within the training data distribution and provides inconsistent results when asked to generalize beyond. We also observe that the deterioration of elastic net accuracy when more simulation data are included is caused by the inclusion of too many non-representative training data points. This result is best explained by the degradation parameter space visualization in Fig. 8. The randomly chosen simulation data covers a very large degradation parameter space. So, there is no guarantee the chosen points will reflect the degradation trends observed in the experimental cells. In effect, the random simulation data prove to corrupt the experimental data, making it more difficult for the elastic net model to learn the correct mapping from dQ/dV features to late-life degradation parameters. Likewise, the experimental dQ/dV curves do not exactly match the dQ/dV curves (see Fig. 7) simulated by the half-cell model and introduce uncertainty into the models through the input feature values.

Inspired by the way simulation data caused the lasso and elastic net to learn an incorrect degradation trend, we investigate whether the models can be coerced into learning a trend more like that observed in the experimental data. We filter the large simulation dataset to only data points corresponding to the highest 20% degradation in m_p , m_n , and LII . Then, this new subset of simulation data, labeled as HiDeg (short for high degradation), is added in small amounts to the early-life experimental data. The results are shown in Fig. A7 (b) and (d). The inclusion of HiDeg simulation data into the models' training dataset produces consistent accuracy improvements across all models. However, the magnitudes of the improvements are not the same. Once again, this can be attributed to a model's ability to extrapolate beyond its training data distribution. For lasso, elastic net, and ELM, the addition of high-degradation simulation data generally helps decrease estimation error, but the performance gains are not as large as MOGP. The large improvement in performance for MOGP can be attributed to the increased size of the degradation parameter space the model is trained on. Likewise, all models benefitted more from the HiDeg simulation data than they did from the regular simulation data. This observation further confirms the notion that regular simulation data is often not representative of the observed experimental degradation trends, and by selectively choosing HiDeg simulation data, the models are not overrun with noise but instead learn a degradation trend that more closely resembles the experimental data, making their predictions more accurate. Altogether, the data augmentation methodology for improving the SOH estimation accuracy of a machine learning model proves effective.

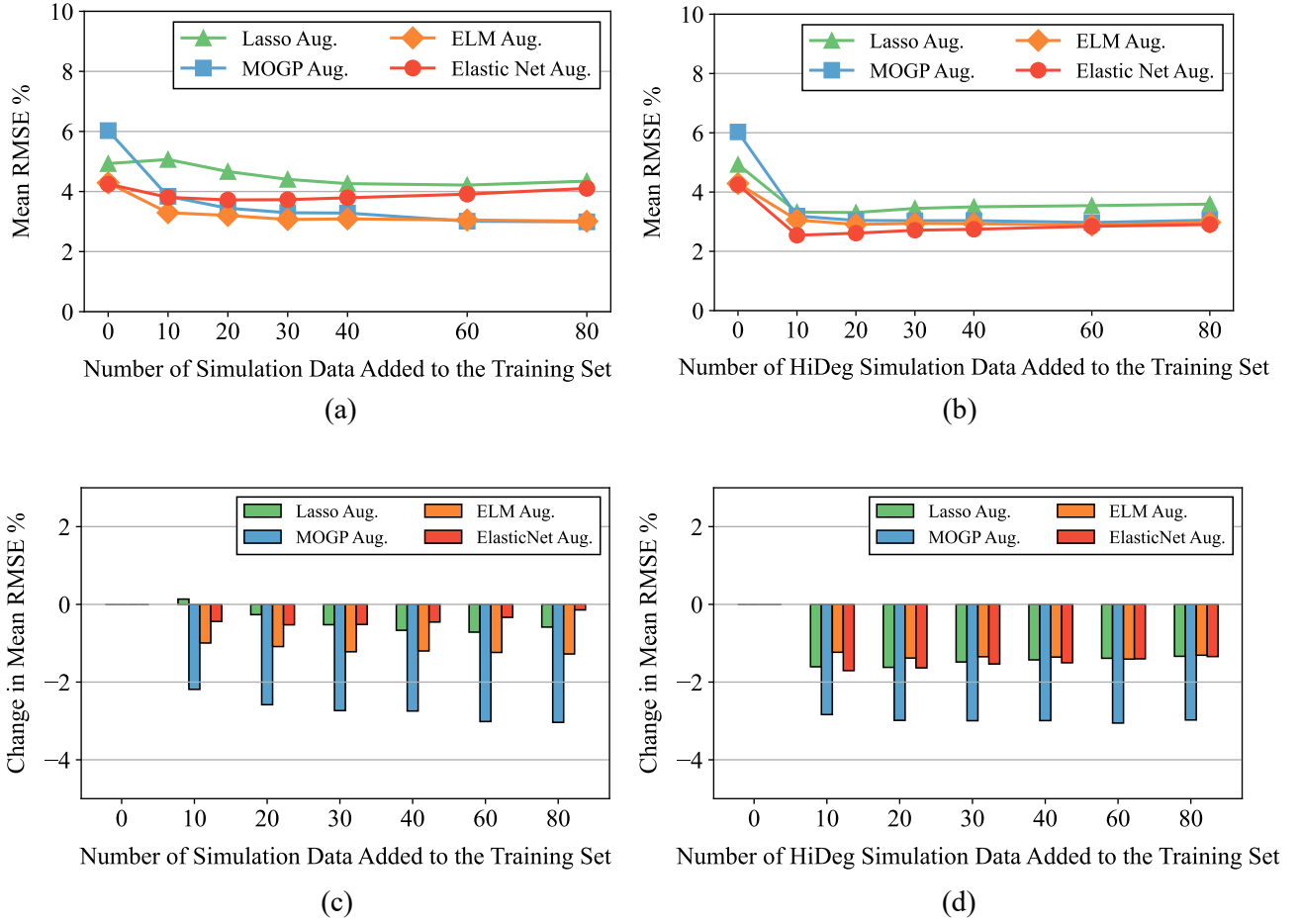


Fig. A7 (a) SOH estimation RMSE % averaged over four tasks of models trained with a combination of early-life experimental data and a different number of simulation data. (b) SOH estimation RMSE % averaged over four tasks of models trained with a combination of early-life experimental data and a different number of high-degradation simulation data. (c) Change in SOH estimation RMSE% with respect to the baseline single-model EXP5 Sim0. (d) Change in SOH estimation RMSE% with respect to the baseline single-model EXP5 Sim0.

A.6 Delta Learning Parametric Study

With the delta learning methodology, the goal is to improve the accuracy of the estimation model trained with only simulation data by training a second model (corrector model) to learn its estimation error. The results for the delta learning model parametric studies are displayed in Fig. A8. First, we investigate the role of the estimation model in the delta learning method by fixing the data for the corrector (the first five experimental data points) and varying the amount of simulation data used to train the estimation model. As shown in Fig. A8 (a), delta learning model error did not decrease with an increase in the number of simulation data used to train the estimation model. This indicates that as few as 50 simulation data points are sufficient for training a machine learning model to learn the mapping between the simulated dQ/dV curves and the health parameters. Knowing this, we then fix the estimation model to use 100 simulation data points and vary the number of early-life experimental data used to train the corrector model (see Figs. A8 (b) and (d)). With this test, the reference models are a single model trained with only the experimental data (from Fig. 6 (e–h)). Traditionally, a proper comparison would be to compare only the estimation model to the two-step delta learning model; however, we have already shown that models trained with only simulation data (the estimation model) are largely inaccurate at the SOH estimation task. Considering this, a more challenging baseline comparison is a model trained with the early-life experimental data because the data is readily available, and the models are more accurate than the simulation-data-only models.

From the tests in Figs. A8 (b) and (d), we find that all delta learning models are consistently more accurate than their baseline counterparts, especially with limited early-life experimental data (three to seven experimental data points).

Furthermore, we can also compare the results from the delta learning method (see Fig. A8 (a)) to those of the augmentation method (see Figs. A7 (a) and (b)). With the same number of available early-life experimental data (first five data points), the MOGP and ELM perform better with the data augmentation method than the delta learning method. The data augmented MOGP and ELM achieved best results of 2.97 and 2.88 average RMSE % across the four estimation tasks compared to 4.52 and 4.40 RMSE % with delta learning method, respectively. This happened because the individual estimation models in the augmentation approach can learn from both the early-life experimental data and the simulation data at the same time. In contrast, the two datasets in the delta learning method are split up, and each model can only train on either the simulation data or the early-life experimental data. The MOGP and ELM, known to not extrapolate well, benefit more from seeing the augmented dataset because they can learn from a larger degradation space and do not need to extrapolate beyond their training dataset. This explains the better performance of MOGP and ELM with the data augmentation method.

Similarly, the lasso and elastic net methods performed better with the delta learning method because they already extrapolated well using the simulation data, and in addition to this, they were also able to take advantage of the early-life experimental data with the corrector model. Both lasso and elastic net were able to extrapolate the estimation model error learned from the early-life data to late-life, yielding lower errors prediction errors.

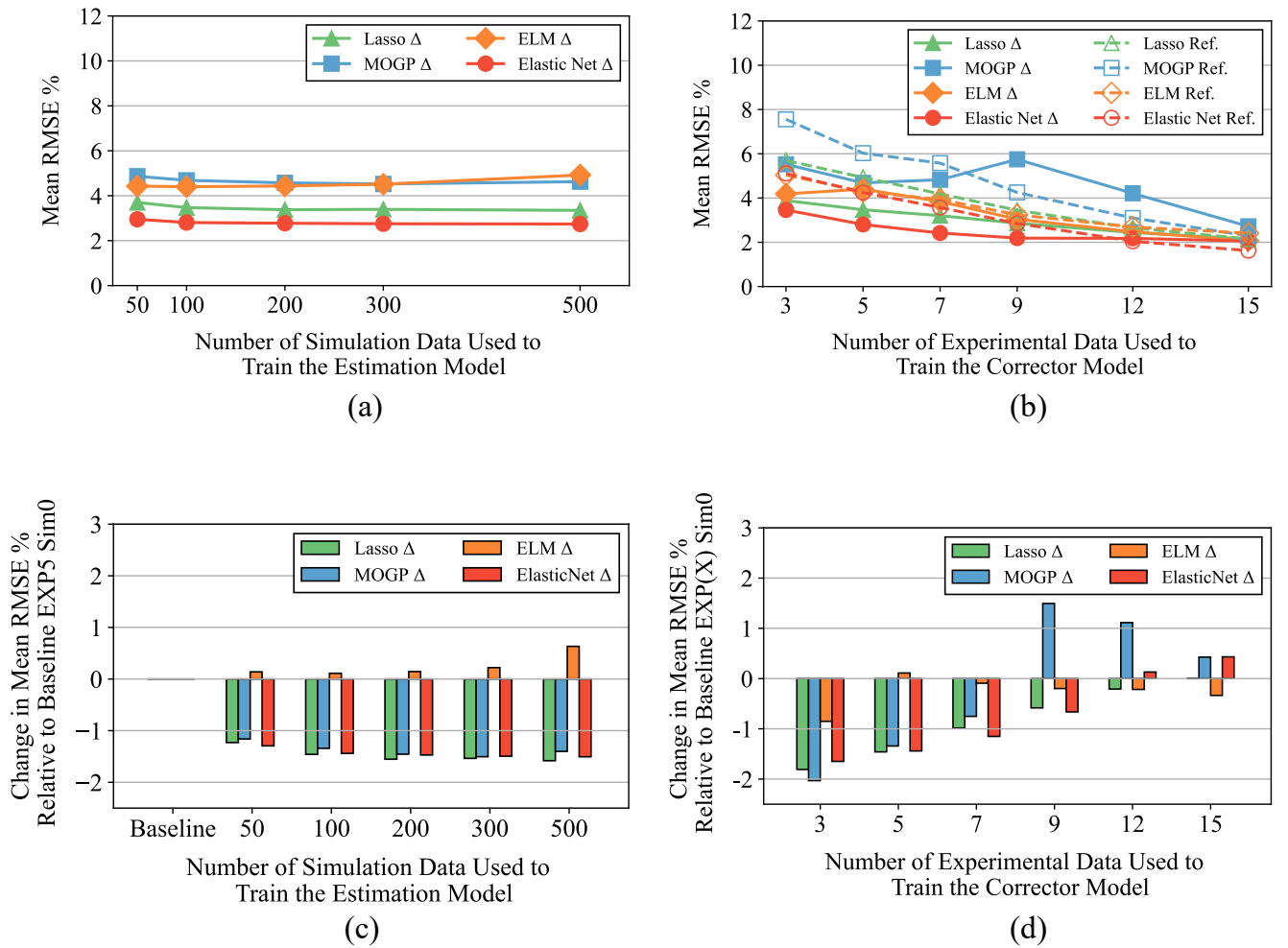


Fig. A8 (a) SOH estimation RMSE % averaged over four tasks of delta learning models where the estimator is trained with a different number of simulation data while the corrector is fixed to EXP5. (b) SOH estimation RMSE % averaged over four tasks of delta learning models where the estimator is trained with 100 simulation data and corrector is trained with different amount of EXP early-life data. (c) Change in SOH estimation RMSE% with respect to the baseline single-model EXP5 Sim0 in (d) Change in SOH estimation RMSE% with respect to each of the baseline models EXP(X) Sim0.

A.7 Feature Reduction Study

A key parameter in the generation of half-cell data is the voltage increment between successive dQ/dV samples. In the previous tests, we had selected an input feature size of 100. To determine whether 100 dQ/dV samples are sufficient for optimal model performance, we varied the input feature size of each machine learning model and recorded the average error across the four outputs. The tests are run using an augmented dataset selected from the middle of all the combinations we tested. In this way, the model will be a good representation of the other models and methods tested in this study. Each model was four-fold cross-validated 50 times with a mixed dataset, namely EXP5 Sim20 HiDeg (33% of the training dataset consisted of simulation data). The results are shown in Fig. A9 (a), and the dQ/dV feature vectors of different lengths are visualized in Fig. A9 (b). Relative to a feature vector length of 100, there is minimal effect on model performance until the number of features is reduced to 25 or less. This indicates that 100 dQ/dV samples are enough to represent the curve's major features (peaks and valleys) and those that are useful for degradation diagnostics. Likewise, using too many dQ/dV samples proves to be counterproductive, and the models have difficulty determining the most important features.

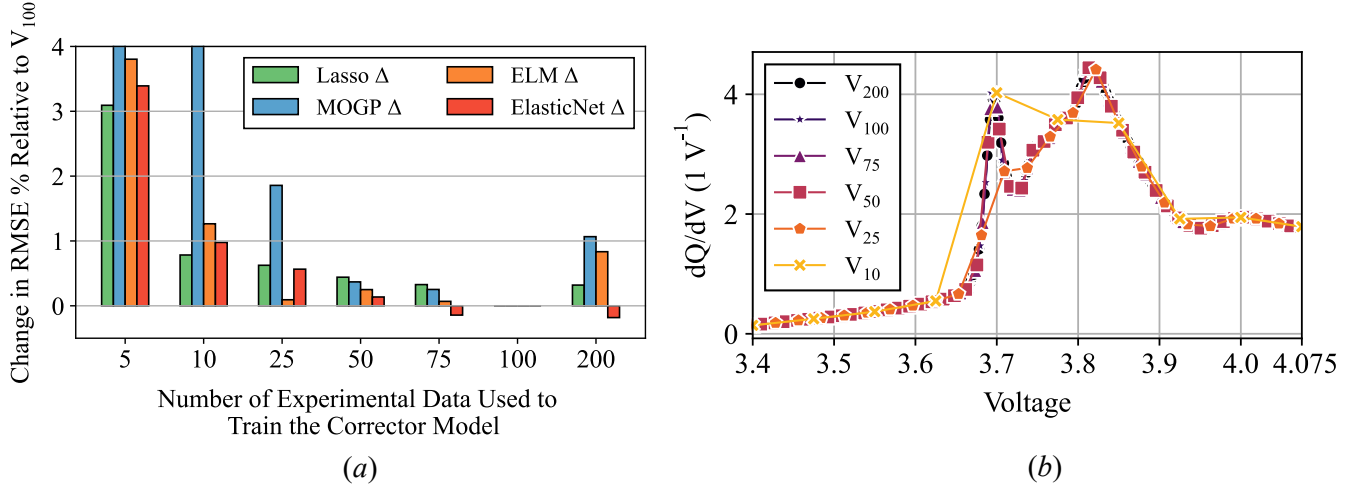


Fig. A9 (a) Effect of dQ/dV discretization length on model accuracy. (b) Visualizations of dQ/dV curves sampled to different lengths.

A.8 Predicted Health Parameter Values

Here, we show the fitted (true) health parameter values compared to the predictions for the baseline model (Elastic Net EXP5 Sim0) and the best model (Elastic Net w/ Data Augmentation EXP5 HiDeg Sim10) for every cell in the dataset. Error bars are plotted for the best performing model to show how its predicted health parameter values vary from run to run. Most of the variation is due to the random selection of the few HiDeg simulation data from the larger pool of HiDeg simulation data. The error bars represent three standard deviations where the standard deviation is determined from the 50 cross-validation runs. As stated in section 3.3, we only consider data from group G4 cells up to day 929 as the dV/dQ features disappeared on the test later and prevented any reasonable half-cell model fitting after that.

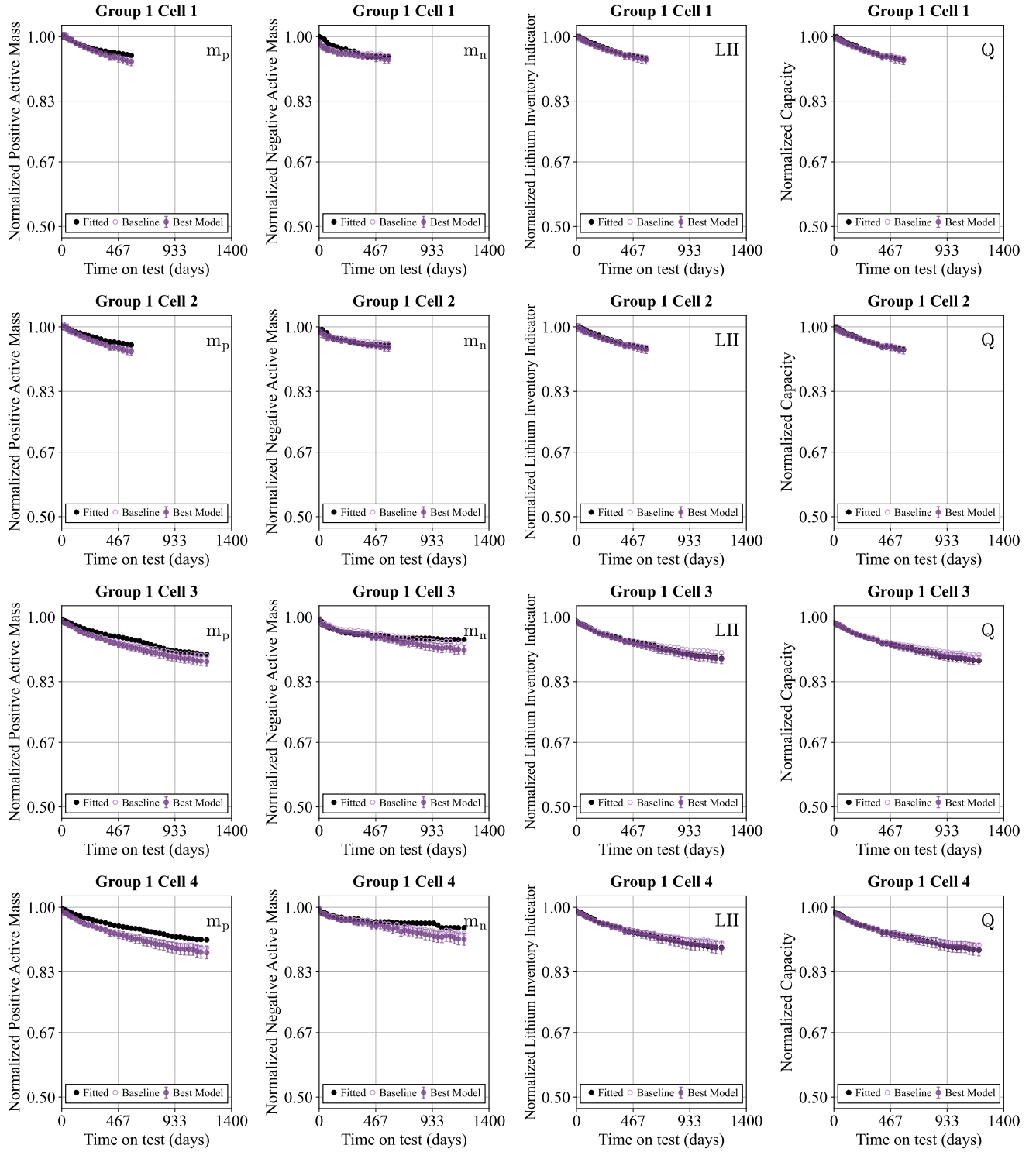


Fig. A10 Normalized group G1 health parameter predictions. Error bars represent three standard deviations where the standard deviation is determined from 50 cross-validation runs.

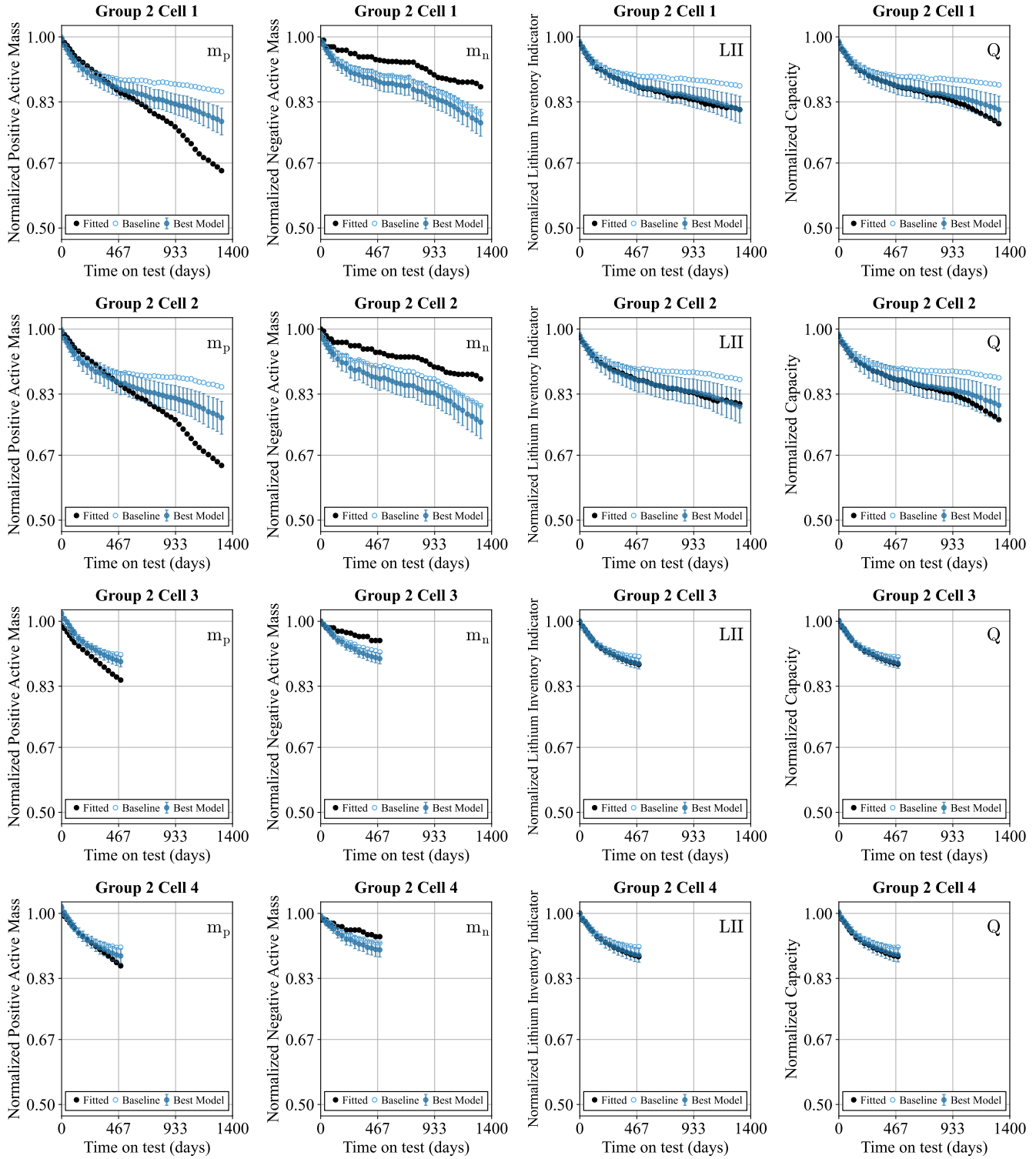


Fig. A11 Normalized group G2 health parameter predictions. Error bars represent three standard deviations where the standard deviation is determined from 50 cross-validation runs.

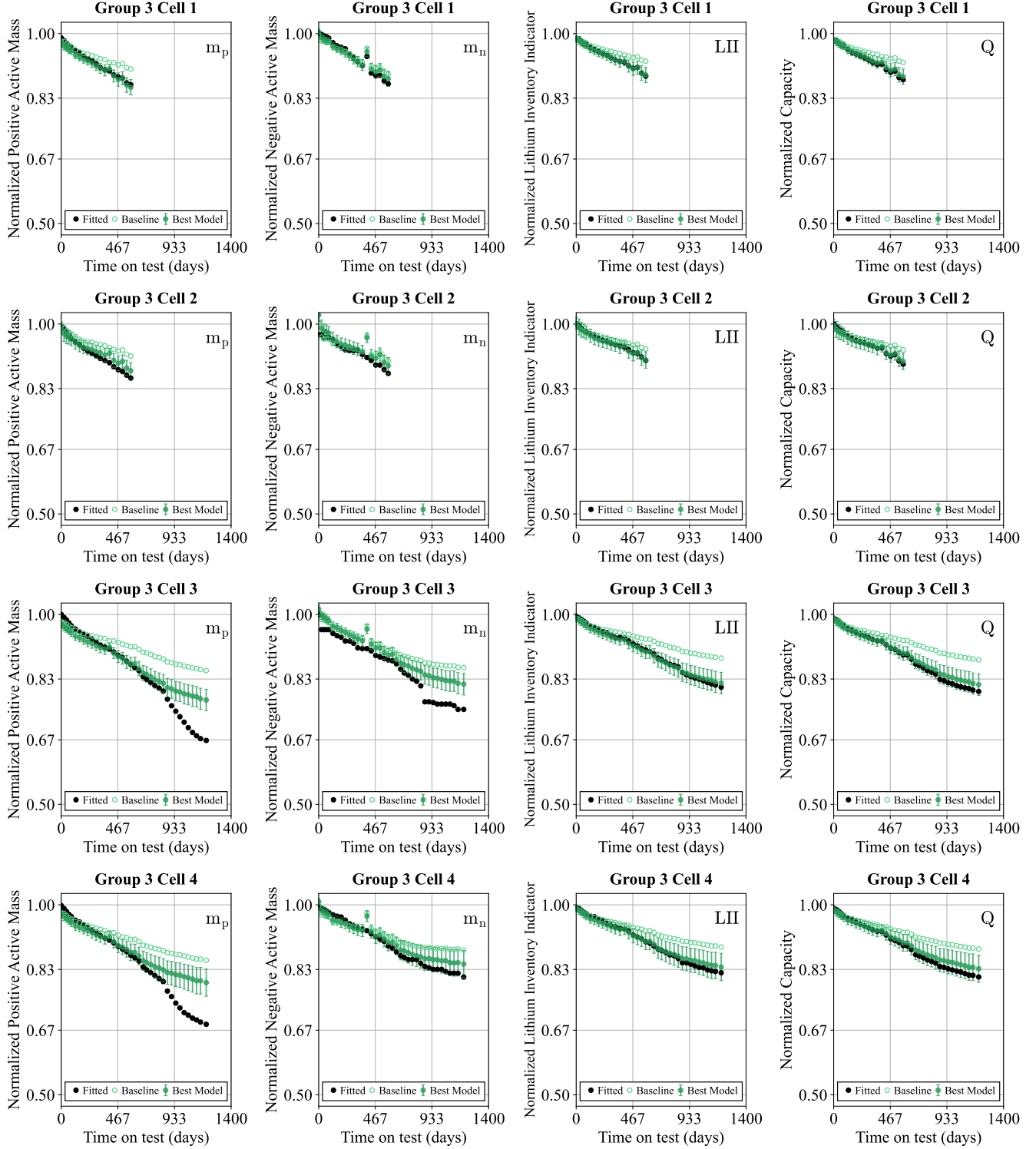


Fig. A12 Normalized group G3 health parameter predictions. Error bars represent three standard deviations where the standard deviation is determined from 50 cross-validation runs.

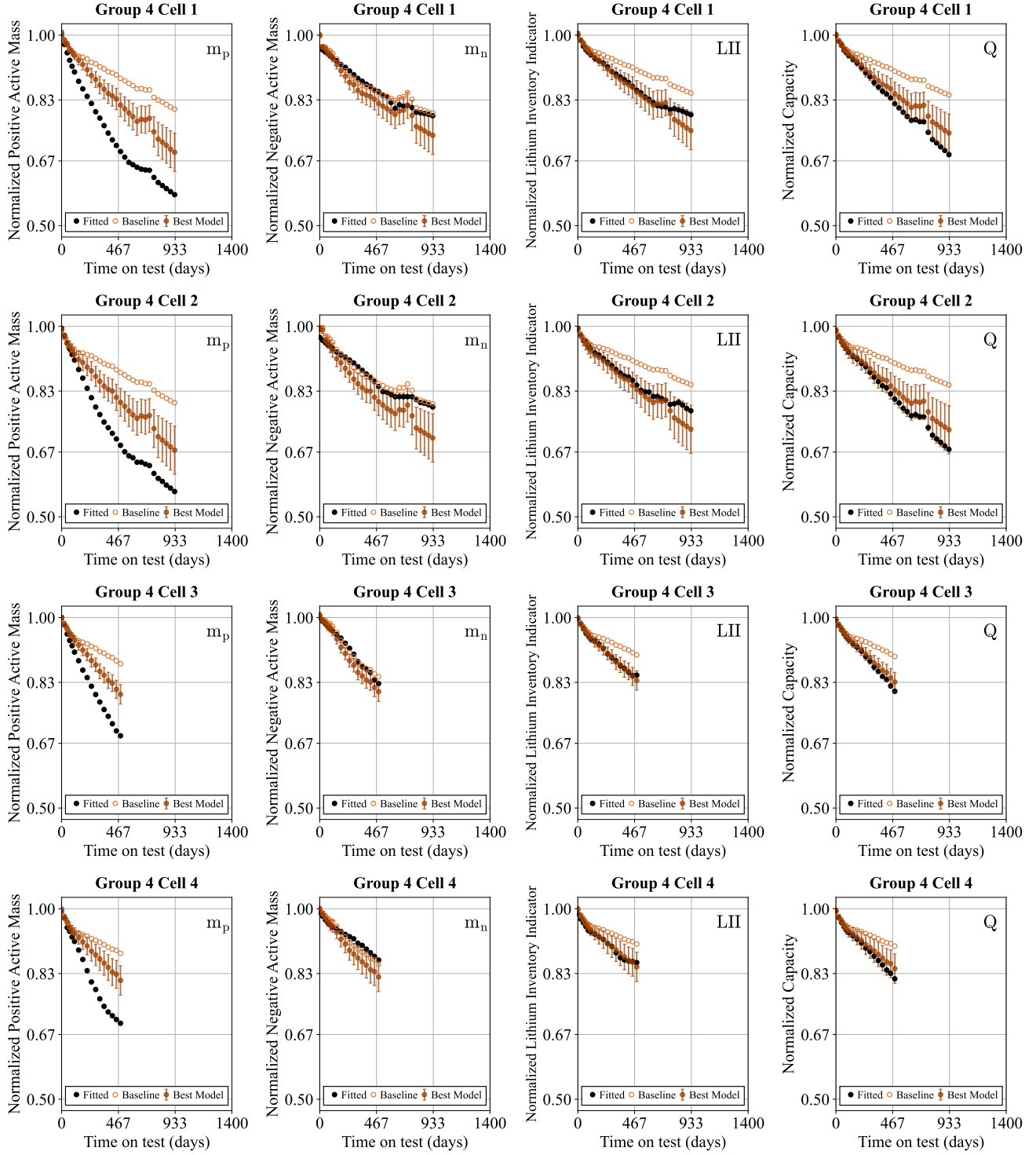


Fig. A13 Normalized group G4 health parameter predictions. Error bars represent three standard deviations where the standard deviation is determined from 50 cross-validation runs.