Forecasting battery capacity for second-life applications using physics-informed recurrent neural networks

Navidi, Sina; Bajarunas, Kristupas; Chao, Manuel Arias; Hu, Chao

DOI
10.1016/j.etran.2025.100432

Publication date
2025

Document Version
Final published version

Published in
eTransportation

eTransportation 25 (2025) 100432

Contents lists available at ScienceDirect

eTransportation

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

Forecasting battery capacity for second-life applications using physics-informed recurrent neural networks

Sina Navidi [a], Kristupas Bajarunas [b,c], Manuel Arias Chao [b,c], Chao Hu [a,*]

[a] School of Mechanical, Aerospace, and Manufacturing Engineering, University of Connecticut, Storrs, 06269, CT, USA
[b] Faculty of Aerospace Engineering, Delft University of Technology, Delft, HS 2926, Netherlands
[c] Institute of Data Analysis and Process Design, Zurich University of Applied Science, Winterthur, 8401, Switzerland

ARTICLE INFO

Dataset link: https://github.com/REIL-UConn/Forecasting-Battery-Capacity-for-Second-Life-Applications-Using-PI-RNN

Keywords:
Lithium-ion battery
Physics-informed machine learning
Capacity forecasting
Recurrent neural network
Second-life application

ABSTRACT

Accurately forecasting lithium-ion battery capacity degradation is crucial for optimizing the second-life utilization of these batteries, enabling reliable operation, reduced maintenance costs, and extended life cycle performance. However, achieving consistent forecasting accuracy across cells and over time remains challenging due to significant cell-to-cell variability and substantial changes in real-world usage conditions during the transition from first to second life. In this study, we propose a new physics-informed machine learning method that integrates an aging-aware electrochemical model with a recurrent neural network, creating a physics-informed recurrent neural network (PI-RNN). This hybrid model leverages both physics-based insights and data-driven learning to predict capacity fade under diverse usage conditions, including transitions from first- to second-life applications. We evaluate PI-RNN using two datasets: an open-source NASA dataset comprising 28 lithium cobalt oxide/graphite cells, and a newly collected dataset of 39 commercial lithium iron phosphate/graphite cells, where cells were initially cycled to 80% capacity in their first life before undergoing milder cycling in their second life. While PI-RNN performs comparably to data-driven models in the first-life phase, it demonstrates a clear advantage in second-life forecasting, reducing root mean squared error by approximately 40%–70% compared to baseline models when forecasting periods span the transition from first to second life, even when trained on as few as two cells. Parametric studies highlight the advantages of incorporating physics-based modeling, and uncertainty quantification ensures the reliability of long-term capacity forecasting. In addition, we conducted benchmarking studies to systematically assess the advantages and limitations of the proposed model, thus identifying the scenarios where this approach excels.

1. Introduction

Lithium-ion batteries have experienced rapid market growth in recent years, driven by their advantageous qualities, such as low self-discharge rates, high power densities, and long lifetimes [1,2]. This growth has expanded their application across various fields, including transportation systems, smart grids, and renewable energy sources [3]. However, continuous usage of lithium-ion batteries inevitably leads to degradation over time due to irreversible side reactions and changes in material morphology. This degradation typically manifests as a decrease in capacity and an increase in internal resistance, both of which can lead to operational inefficiencies, unplanned downtime, and in severe cases, safety risks if the degradation reaches a critical threshold [4]. Internal resistance is commonly measured using techniques such as the hybrid pulse power characterization (HPPC) test [5] or standard cycling and pulsing protocols for equivalent circuit model (ECM) parameter identification [6]. HPPC tests are designed to assess battery behavior under high current pulses. However, these tests often require precisely controlled protocols to ensure reliable data, making standardization and interpretation a significant challenge [7]. Furthermore, traditional HPPC tests lack a strong physical basis for protocol design, limiting their ability to diagnose specific degradation mechanisms accurately. ECM parameters, which are derived from these tests, are highly sensitive to state of charge (SOC), state of health (SOH), and temperature. This variability necessitates extensive testing under diverse conditions to produce robust models, which can be resource-intensive and time-consuming [8]. Capacity measurements, another critical metric for assessing SOH, typically involve discharging the battery fully between predefined upper and lower voltage cut-offs and integrating the discharge curve to calculate the ampere-hour capacity. However, in practical applications, full discharge cycles are

* Corresponding author.
E-mail address: chao.hu@uconn.edu (C. Hu).

https://doi.org/10.1016/j.etran.2025.100432
Received 13 January 2025; Received in revised form 19 May 2025; Accepted 19 May 2025
Available online 9 June 2025

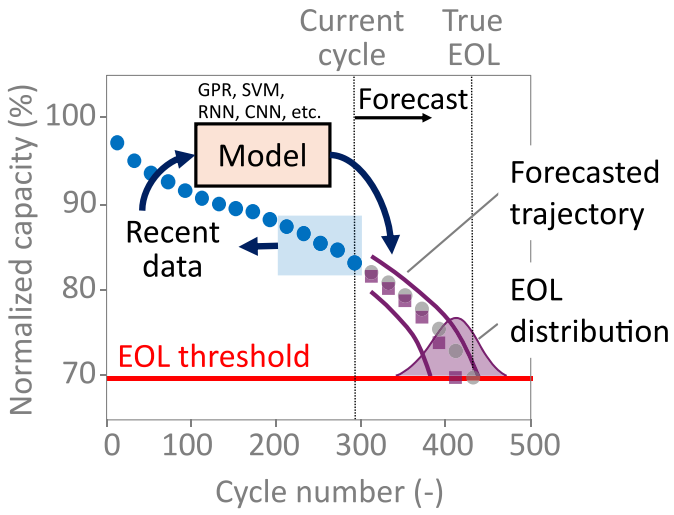

Fig. 1. Schematic representation of the battery capacity forecasting problem adapted from [12].

often avoided due to operational constraints and the risk of accelerated degradation [9]. For instance, lithium-ion batteries are frequently operated in partial charge/discharge cycles to extend their lifespan, as full discharges can induce stress and damage to the cells [10]. Additionally, external factors such as temperature, current rates (C-rates), and battery age significantly influence voltage cut-offs, leading to variability in capacity measurements [11]. In practical settings, greater focus is placed on tracking capacity loss, as it directly impacts the energy a battery can store and deliver, making it a key indicator of SOH.

Accurate and computationally efficient prediction of a battery's capacity degradation is essential for battery management systems or other battery health management platforms to recommend/implement suitable management strategies throughout the battery's service life. It also aids manufacturers in promptly assessing battery lifetime, optimizing cell and battery pack design, and strategizing for recycling and repurposing efforts [13].

1.1. Background

Numerous studies have investigated lithium-ion battery health prognostics [14], with a particular focus on predicting capacity degradation. These approaches can be broadly categorized into model-based and data-driven methods. Capacity forecasting traditionally involves using past and current cycle data as inputs to a machine learning model, which generates probabilistic capacity estimates (e.g., mean and standard deviation) for future cycles. These estimates form a capacity degradation trajectory, allowing an end-of-life (EOL) estimate based on when the trajectory crosses a predefined capacity threshold, typically 80% of the nominal capacity for automotive applications [12] as illustrated in Fig. 1. This threshold defines the transition from a battery's first-life, during which it is used in its primary application (e.g., electric vehicles), to its second life, where it may still be repurposed for less demanding applications such as stationary energy storage. More recently, researchers have begun integrating model-based and data-driven approaches, leading to the development of physics-informed machine learning (PIML) frameworks [15].

1.1.1. Model-based methods

Model-based methods (category I in Fig. 2) for battery health prognostics involve creating mathematical models, which can be electrochemical models [16,17], ECMs [18] and empirical models [19,20]. Electrochemical models (subcategory I.1 in Fig. 2) detail internal battery degradation processes at a microscopic level, focusing on mechanisms such as solid electrolyte interface (SEI) growth [21] and lithium plating [22]. Including these mechanisms in the model allows simulations to forecast future capacity degradation. However, capacity loss can arise from various failure mechanisms, making it challenging to develop comprehensive electrochemical models that account for different loading conditions over the battery's lifespan. This is because, as batteries age, the initial model parameters can vary significantly, leading to substantial errors in predicting capacity degradation. ECMs (subcategory I.2 in Fig. 2) are used to model the dynamic characteristics between the current and terminal voltages of a battery using circuit elements such as capacitors, resistors, and voltage sources. The sum of the equivalent impedances is then identified using electrochemical impedance spectroscopy (EIS) data to quantify the battery's SOH. However, acquiring EIS data requires additional equipment and extensive testing time [23], which limits the practical use of ECM models. Empirical models (subcategory I.3 in Fig. 2), on the other hand, fit capacity fading trends using functions like exponential [24], polynomial [25], or ensemble methods [26] to extrapolate capacity over time, offering the advantage of directly predicting degradation levels. However, these models often require extensive experiments to account for various factors, making them unsuitable for long-term forecasting in the early stages.

1.1.2. Data-driven methods

Unlike model-based methods, data-driven approaches (category II in Fig. 2) can predict battery SOH without understanding complex physics, making them more adaptable to different loading conditions in practical applications [27]. However, their accuracy heavily depends on available historical data. Both lightweight machine learning and deep learning algorithms (subcategories II.1 and II.2 in Fig. 2) have been successfully applied in battery health prognostics. When features are well extracted, lightweight algorithms such as linear regression [28], support vector machines (SVM) [29,30], relevance vector machines (RVM) [31], and Gaussian process regression (GPR) models [32,33] are used to develop prognostics models. In contrast, deep learning algorithms, which can automatically extract features, are able to directly use raw time-series data as input. Typical examples include long short-term memory (LSTM) networks [34,35], convolutional neural networks (CNNs) [36], and transformers [37,38]. Practical challenges for data-driven methods include incomplete battery usage histories and difficulties in obtaining full charging/discharging curves in real-world applications, which limit data availability for training the machine learning models. Moreover, purely data-driven approaches lack an inherent understanding of underlying physical principles, posing challenges in improving their performance, particularly for scenarios beyond their training data distribution [39].

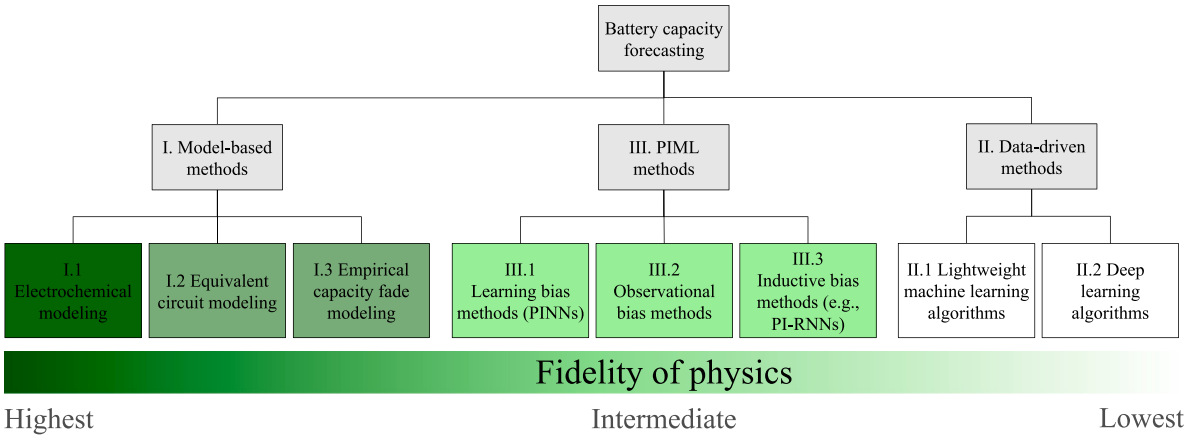

Fig. 2. A broad classification of battery capacity forecasting approaches, categorized according to the extent of physics-based information incorporated, from purely data-driven methods to fully physics-based models, with hybrid PIML strategies occupying the intermediate spectrum.

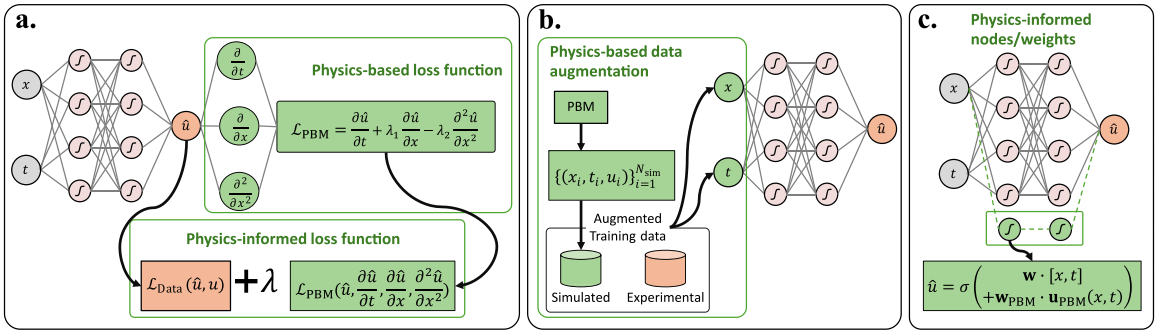

Fig. 3. Categories of PIML approaches: **a**. learning bias, **b**. observational bias, and **c**. inductive bias. $\mathcal{L}_{\text{Data}}$ is data-driven loss and $\mathcal{L}_{\text{PBM}}$ is physics-based loss. In **b**, N_{sim} denotes the number of input–output pairs generated by physics-based simulations. In **c**, $\mathbf{w}_{\text{PBM}}$ denotes physics-informed weights and $\mathbf{u}_{\text{PBM}}$ represents the PBM output vector function.

1.1.3. PIML methods

To address the challenges in model-based and data-driven methods, PIML methods have recently gained popularity in the field of prognostics and health management due to their ability to combine the strengths of physics-based and data-driven models. Due to incorporating physical information, PIML methods can show enhanced generalization capabilities and require fewer high-fidelity data for training [40]. Karniadakis et al. [41] conducted a comprehensive review of PIML methods, highlighting biases introduced by prior physical knowledge. They defined PIML as leveraging prior observational, empirical, physical, or mathematical understanding to enhance learning algorithms, categorizing PIML approaches into observational, learning, and inductive biases as shown in Fig. 3. Learning biases arise from selecting appropriate loss functions and inference algorithms to guide machine learning models toward solutions aligned with physical principles (Fig. 3a). Observational biases in PIML are introduced through data that reflect underlying physics or through well-designed data augmentation techniques (Fig. 3b). Inductive biases are prior assumptions or constraints integrated into the architecture of machine learning models, particularly within the neural operations of neural networks. These biases help ensure that predictions conform to specific physical laws and principles, which are often represented as mathematical constraints (Fig. 3c).

In battery research literature, Aykol et al. [42] proposed a classification of PIML methods into five categories: three "sequential integration" methods (A1–A3) and two "hybrid" methods (B1–B2). The "sequential integration" methods feature standalone physics-based and machine learning models, akin to observational bias techniques. In contrast, "hybrid" methods integrate both models, with the B1 architecture incorporating physics-based elements into the machine learning framework, potentially leveraging learning or inductive biases. The observational bias (subcategory III.2 in Fig. 2) method is relatively straightforward to implement because the physics-based and machine learning models are standalone, making it a practical near-term strategy for battery prognostics modeling [43–45]. However, "hybrid" models that integrate both learning bias and inductive bias (subcategories III.1 and III.3 in Fig. 2) have been implemented far less frequently, primarily due to the complexity of embedding physics-based equations into neural networks. To develop a "hybrid" PIML framework effectively, it is important to choose neural network architectures that are tailored to the specific problem.

1.1.3.1. Learning bias in feed-forward neural networks. Classical feed-forward neural networks are effective in capturing nonlinear relationships between inputs and outputs through parameter adjustment using backpropagation and gradient descent. However, these models face limitations in lithium-ion battery modeling because they process only the current input at a given time step, making it difficult to account for long-term effects, such as polarization, which depend on the previous load history. Accurately capturing these effects requires access to the complete load history [46]. For example, You et al. [47] developed a neural network using raw voltage, current, and temperature signals to estimate the SOH. However, this approach struggles in prognostic applications due to its inability to capture the temporal characteristics necessary for accurate long-term predictions. Recent studies by Wang et al. [9] and Navidi et al. [15] introduced physics-informed neural network (PINN) models for estimating battery SOH, each employing distinct strategies to incorporate domain-specific knowledge. Wang et al. utilized state-space equations in a partial differential equations format in the loss function, leveraging neural networks to model underlying degradation dynamics. Navidi et al. on the other hand, integrated a half-cell empirical degradation model directly into the

PINN framework. Both approaches adopt a learning bias strategy to guide the optimization process and enhance predictive accuracy in capturing battery degradation. However, these methods are not tailored for long-term prognostics, limiting their applicability in this domain. In a more recent study, Bajarunas et al. proposed an unsupervised learning bias method for SOH estimation, embedding general knowledge about degradation trends at different stages of a battery's life into a convolutional autoencoder's architecture [48]. While this method has shown promise in predicting remaining useful life (RUL) in a prognostics setting, its ability to generalize to real-world applications, where degradation trends may not follow predefined patterns, is yet to be demonstrated.

1.1.3.2. Inductive bias in recurrent neural networks. Recurrent neural networks (RNNs) are able to incorporate an internal memory mechanism that allows them to consider past time steps without requiring the entire load history as explicit input [49]. This feature facilitates time-series prediction, as RNNs can process current inputs while inherently accounting for prior data. Additionally, RNNs can predict time series of varying lengths sequentially, making them more flexible than traditional neural networks that demand a predefined input format. This adaptability enables RNNs to model long-term dependencies more effectively [50]. For example, Liu et al. [51] developed an adaptive RNN that captures temporal characteristics and dynamic states during battery degradation. Although this model adapts well to abrupt state changes, it faces challenges in long-term predictions due to the gradient vanishing problem associated with RNNs [52]. To mitigate this issue, Zhang et al. [34] investigated the use of LSTM, a variant of RNNs, for on-board SOH and remaining useful life (RUL) estimation. Their model achieved accurate SOH predictions with or without offline training data, while maintaining reasonable training times. However, its validation was limited to a narrow range of experimental conditions. Lu et al. [35] introduced an RNN-based deep learning framework designed to predict long-term degradation trajectories for NMC/Gr cells under both constant and variable usage conditions. Their model employs a gated recurrent unit to forecast charge and discharge capacities using future current load data along with early capacity-voltage data as input.

Nascimento and Viana [53] were the first to introduce an inductive bias approach in an RNN structure by embedding a physics-based model into the RNN cells. Inspired by cumulative damage modeling [54], they used a multi-layer perceptron within the recurrent cell to model stress intensity and integrated it with the Paris law layer, accurately simulating fatigue crack propagation in wind turbine main bearings. Following the same approach, Kim et al. [55] presented an inductive bias approach by using a hybrid recurrent unit. In their approach, they incorporated an empirical model within the recurrent unit to compute the damage increment, representing the capacity degradation for each charge–discharge cycle. Nascimento et al. [56] applied the same inductive bias approach to a battery prognostics problem, developing a hybrid model that integrates physics-based and empirical equations within an RNN unit. Each unit includes a principle-based node using a simplified electrochemical model (Butler–Volmer–Nernst equations) for electrode overpotential, a data-driven node for internal voltage, and a variational Bayesian node for uncertainty. This model leveraged fleet-wide data to predict future capacity drops, merging information on battery aging parameters with existing aging data. However, it is important to note that their work primarily demonstrated the potential of this approach, rather than providing a thorough validation or a clear pathway for practical deployment. The approach was validated using two subsets from the NASA prognostics data repository, one with constant loading and one with randomized loading conditions, each including 4 cells. While the results demonstrated effectiveness in predicting voltage discharge cycles and estimating the residual capacity of batteries under randomized loading conditions, it is important to note that the validation on randomized conditions was conducted on only 4 cells, which limits its statistical significance.

1.2. Knowledge gaps and our contributions

Despite the development of numerous battery degradation prognostic models demonstrating excellent performance for specific applications, several areas require improvement:

- **Limited real-world applicability:** Many existing studies are constrained in their applicability to real-world scenarios due to their reliance on historical state information and the assumption that cells operate under constant conditions. This neglects the varying future usage conditions that batteries experience in practical settings. Battery capacity degradation is closely tied to diverse loading conditions throughout the service life. Therefore, a more practical approach would involve health predictions as an evolving function of changing usage conditions. Two of the studies that tackled this gap were conducted by Richardson et al. [33] and Lu et al. [35]. Both relied on the assumption of access to the cell's response, either in the form of temperature or voltage data. Richardson et al. utilized a GPR model to predict short-term capacity degradation while accounting for varying usage conditions, using temperature data as a key input. Lu et al. on the other hand, predicted capacity degradation trajectories under random future usage conditions by integrating future current load patterns with early capacity-voltage response data. While these studies provided valuable insights, their dependence on specific cell responses highlighted a key limitation, as such information was often unavailable in practical applications.
- **Complexity of modeling approaches:** The majority of existing prognostic studies rely on either physics-based or data-driven approaches, both of which present significant challenges for practical deployment for online inference (e.g., embedded on battery management systems). Physics-based models require extensive parameterization, as noted by Chen et al. [57]. This process is time-consuming and requires specialized characterization methods, making it impractical for many applications. On the other hand, data-driven models depend on large training datasets and highly predictive features derived from temperature or voltage measurements, which are not always available in real-world settings. For example, Deng et al. [58] employed an LSTM model with a GPR residual, relying on a large dataset from 20 vehicles and temperature and voltage data. Additionally, the complexity of deep learning models, such as LSTM networks, further complicates their deployment on battery management systems or other edge computing platforms that require lightweight and computationally efficient models.
- **Lack of second-life capacity forecasting studies:** To the best of our knowledge, there have been no studies focused on second-life capacity forecasting following the end of a battery's first life, particularly at the critical transition point when cycling conditions might change. Most research has concentrated on fixed protocols until the end of the first cycling life of batteries. However, second-life applications are becoming increasingly important as a sustainable solution for managing retired batteries from electric vehicles and other applications. These batteries often retain significant capacity and can be repurposed for less demanding applications.
- **Lack of hybrid PIML implementations:** While there have been studies applying PIML for battery health prognostics, the literature lacks straightforward and reliable implementations of inductive bias methods for health prognostics that are suitable for deployment on edge-based battery health management platforms such as battery management systems [42].

To address the gaps highlighted above, we propose a PIML approach for capacity forecasting of lithium-ion batteries, capable of handling time-varying usage conditions and second-life usage. This new PIML

approach builds a new class of RNNs, which we call physics-informed RNNs (PI-RNNs), to forecast capacity drops based on features extracted from future load patterns and a recent history of capacity degradation. It incorporates an initial estimate of degradation from a physics-based model as a new input layer into standard RNNs, introducing a physics-informed inductive bias. The resulting PI-RNNs are equipped with a more informative starting point than features from raw data that are often empirically defined. The key contributions of this paper include:

- **Dynamic capacity loss forecasting:** Our approach predicts capacity loss dynamically as conditions evolve, moving away from models that rely on static historical data and fixed cycling protocols.
- **Independence from historical state information:** Our method exclusively utilizes features extracted from current versus time data during future load cycles, eliminating the need for temperature or voltage response data during inference.
- **Data efficiency:** We reduce the amount of battery cell data required for training while maintaining predictive accuracy. The proposed method also demonstrates excellent out-of-distribution prediction accuracy.
- **Computational efficiency:** Once trained, PI-RNNs are computationally efficient during online inference, making them well-suited for deployment in on-the-edge battery health analytics applications (e.g., battery management systems). Moreover, the implementation is inherently scalable, capable of being extended with additional resources as needed. This approach offers a straightforward implementation of inductive bias techniques, which are not yet widely adopted in the battery field.
- **Performance benchmarking:** We benchmarked the performance of the proposed PI-RNN method against baseline state-of-the-art methods in the literature for capacity forecasting using two comprehensive datasets, highlighting the advantages and use cases where it may be preferable to other methods.

1.3. Article organization

The remaining parts of this paper are organized as follows: Section 2 provides the problem definition. Section 3 introduces the battery cycling experiments and the datasets. Section 4 details the proposed PI-RNN methodology. In Section 5, we elaborate on the results, validation, and discussions of the proposed method. Finally, Section 6 presents the conclusions drawn from this study.

2. Problem definition

This study tackles the challenge of forecasting capacity degradation across successive load patterns using a novel PIML health transition model. Unlike traditional battery cycle approaches, a "load pattern" is defined as a sequence of load periods, each characterized by a battery current vs. time profile, between two consecutive reference performance tests (RPTs) [33], Q_t at RPT_t and Q_{t+1} at RPT_{t+1}. These load patterns can vary in both duration and complexity, often including multiple charge–discharge events. We assume that only current versus time data (i.e., time series measurements of battery current) is available for inference, representing practical scenarios where detailed cell state and response information are unavailable, and only the loading conditions are known.

The goal is to predict the capacity drop for an upcoming load pattern based on the current capacity and features derived from the future loading conditions. By accurately predicting the capacity drop for each load pattern, this approach allows for long-term forecasting of cumulative capacity degradation over multiple load patterns. To achieve this, we seek to learn the mapping from an input feature vector $\mathbf{x}$, which captures the loading conditions over a load pattern, to the output y, representing the level of capacity degradation associated with that load pattern. This mapping is learned using a labeled training set composed of N_{train} input–output pairs $\{(\mathbf{x}_t, y_t)\}_{t=1}^{N_{\text{train}}}$ collected from one or multiple cells. Here, N_{train} corresponds to the number of load patterns, equivalent to the number of RPTs, across all training cells. Each load pattern $\mathbf{L}_t$ is characterized by a set of features extracted from the time series of battery current. The input vectors $\mathbf{x}_{t+1}, \ldots, \mathbf{x}_{t+M}$ used for predicting capacity degradation over M future load patterns are the feature vectors extracted from the corresponding load patterns $\mathbf{L}_{t+1}, \ldots, \mathbf{L}_{t+M}$. The outputs $y_{t+1}, \ldots, y_{t+M}$ represent the incremental capacity drops measured over the forecasting horizon of M load patterns.

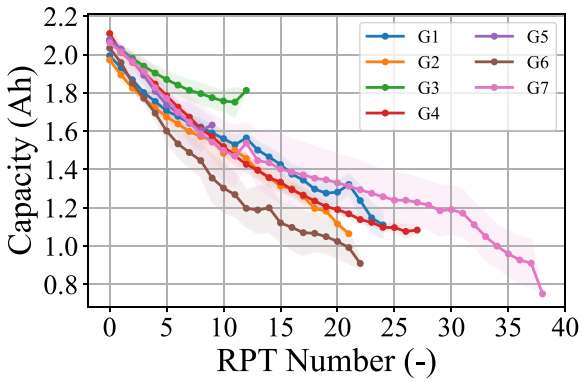

Fig. 4. Capacity fade plot of cells from the NASA random walk dataset. Groups G1, G2, G3, G5, and G7 were cycled at room temperature, while groups G4 and G6 were cycled at 40 °C. For all groups except G7, the discharge current was randomly selected from the range of 0.5 A to 4 A. The charge current for groups G1-G5 was fixed at 2 A, with the charge duration in G1 randomly selected from 0.5 to 3 h. All charge and discharge cycles, except for G7, were performed until the cells were fully charged and discharged. In group G7, both charge and discharge currents were randomly selected from the range of −4.5 A to 4.5 A.

3. Dataset description

3.1. NASA randomized battery usage repository (LCO random walk dataset [59])

This first dataset was originally used in [60] to study capacity degradation under randomized load profiles, providing a more realistic representation of practical battery usage conditions. It consists of 28 experimental cells manufactured by LG Chem, with an 18650 form factor and lithium-cobalt-oxide/graphite (LCO/Gr) chemistry. The cells have a rated capacity of 2.1 Ah, offering relatively high energy density but shorter cycle life. The dataset is divided into seven groups, each containing four cells, with each group subjected to a distinct randomized load profile as outlined in Table A.4. Groups 1, 2, 3, 5, and 7 were cycled at room temperature, while groups 4 and 6 were cycled at 40 °C. Periodically, an RPT was conducted, involving a 2 A (approximately 1C) charge–discharge cycle between the cell voltage cut-offs. The capacity was evaluated from the discharge curves obtained during these RPTs, integrating the current over time to assess the total capacity. Key experimental variables in this dataset include discharge current and ambient temperature as shown in Fig. 6a. Fig. 4 displays the plotted capacities of cells across different groups against RPT number, illustrating how varying discharge operational conditions lead to distinct capacity trajectories within each group. Each trajectory represents the mean and standard deviation of the capacities for groups of four cells.

3.2. Battery aging dataset for second-life SOH tracking (LFP second-life dataset)

This dataset consists of 39 lithium-iron-phosphate/graphite (LFP/Gr) cells, manufactured by LithiumWerks, with an 18650 form

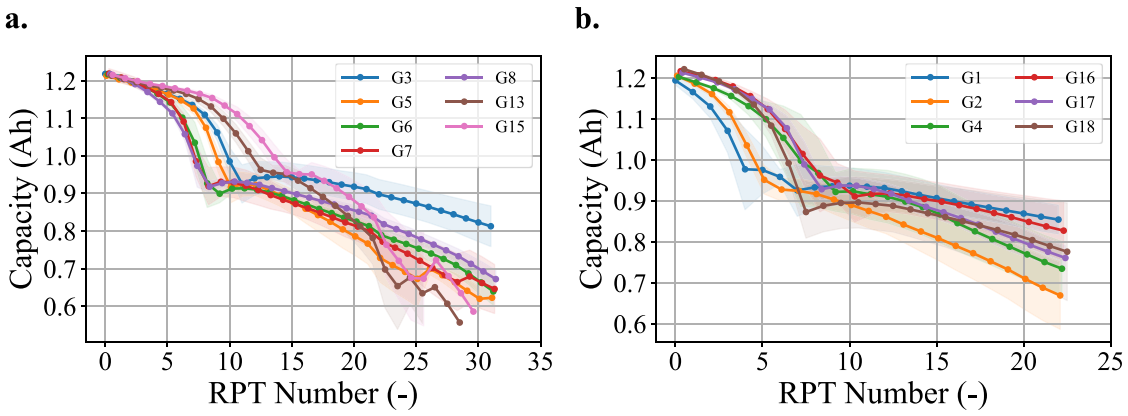

Fig. 5. Capacity fade plots of the cells in **a.** batch 1 and **b.** batch 2 of case study 2 (LFP second-life dataset). Groups G1-G5 and G16-G18 followed the one-step fast charging without derating protocol with charge C-rates of 3.8C-5.7C, discharge C-rates of 2.1C-3.6C, and DODs of 70%–97%. Groups G6-G10 used the one-step fast charging with SOH-based derating protocol, with initial charge C-rates of 4.2C-5.8C, discharge C-rates of 2.1C-3.9C, and DODs of 71.9%–95.7%. Groups G11-G15 followed the two-step fast charging protocol with CC charging at 4.1C-5.7C up to a switching SOC of 71.1%–86.8%, followed by one step of CCCV at 1C. Discharge C-rates were 2.1C-3.9C, with DODs of 74.7%–95.3%. After reaching the 80% of initial capacity, the second-life cycling protocol consisted of a CCCV charging step and a CC discharging step, both performed at a C-rate of 1C and full DOD.

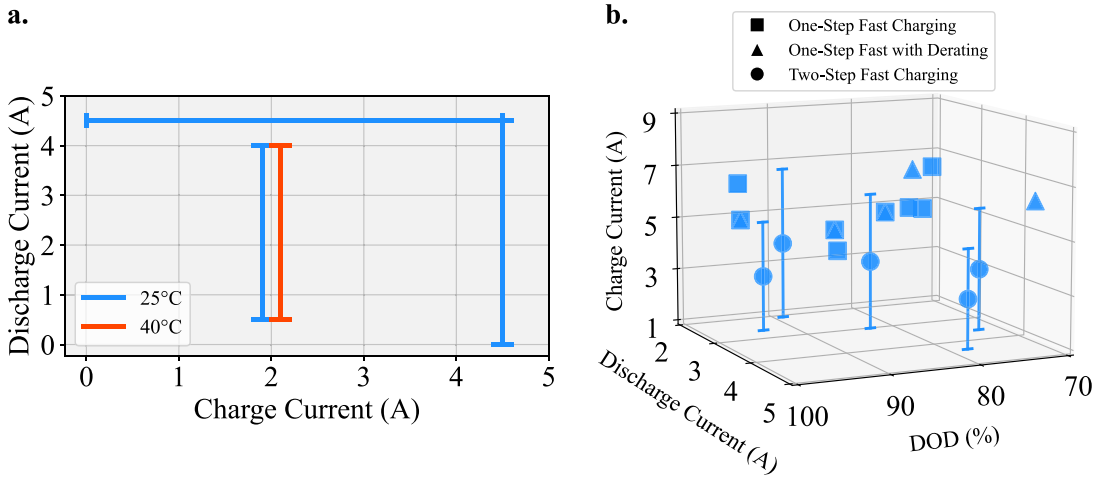

Fig. 6. **a.** Experimental design variables for case study 1, where the primary variable is the discharge current, with limited variability in charge current. **b.** Experimental design variables for case study 2, showcasing a more complex scenario with significantly broader ranges for both charge and discharge currents and DOD, highlighting the increased variability in cycling conditions. For detailed cycling conditions, refer to Tables A.4 and A.5.

factor, tested under various cycling conditions. The exact electrolyte composition is not disclosed by the manufacturer, but it is expected to be a typical lithium-ion formulation such as $LiPF_6$ in organic carbonate solvents. The cells, with a nominal capacity of 1.2 Ah, were initially aged to 80% of their remaining capacity under high-rate cycling conditions (first-life application). Afterward, they were subjected to milder cycling conditions simulating second-life usage. The dataset consisted of 39 cells, divided into 13 groups, with each group containing 3 cells. These groups were further categorized into two batches based on the range of design variables. Batch 2 was subjected to slightly more aggressive conditions compared to batch 1. The first-life testing involved fast charging protocols with changing design variables such as C-rate, derating ratio, switching state of charge (SOC), and depth of discharge (DOD) as illustrated in Fig. 6b.

The first fast charging protocol, termed “one-step fast charging”, utilized a single constant-current consant-voltage (CCCV) charge step, with the charging C-rate varied between 4C and 6C. This C-rate remained constant throughout the entire first-life cycle. The second fast charging protocol, referred to as “one-step fast charging with SOH-based derating”, also implemented a CCCV charging step; however, the charging C-rate decreased by a derating ratio of 5% to 10% for every 5% reduction in remaining capacity. The initial C-rate for a fresh cell similarly ranged from 4C to 6C, matching the one-step fast charging protocol. The third fast charging protocol, termed “two-step fast charging”, involved a constant-current (CC) charging step up to a predetermined switching SOC level, with the C-rate maintained between 4C and 6C. Subsequently, the cell continued charging using a CCCV step at a rate of 1C, with the switching SOC level sampled between 70% and 90%. After every 5 days of cycling, an RPT was conducted to measure the remaining capacity and internal resistances. Once the cells reached 80% capacity, they were “retired” from first-life usage and subjected to second-life testing. The second-life cycling protocol consisted of a CCCV charging step and a CC discharging step, both performed at a C-rate of 1C and full DOD. A Latin hypercube sampling strategy was used to explore the design space during first-life cycling, determining variables such as DOD, C-rates, and switching SOC levels (or derating ratio when applicable).

The detailed cycling conditions are summarized in Table A.5. To ensure consistent ambient conditions, all cells were cycled and tested at 25 °C. This uniform temperature setting was maintained during both load patterns and RPTs. Fig. 5 shows the capacity fade plots of the cells from both batches in this dataset. Each trajectory represents the mean and standard deviation of the capacity fade for groups of three cells. Please note that groups 11, 12, and 14 are excluded since they have not reached the end of their first life. A noticeable transition point from first life to second life is present in all trajectories, where some capacity retention occurs due to the shift to milder usage conditions.

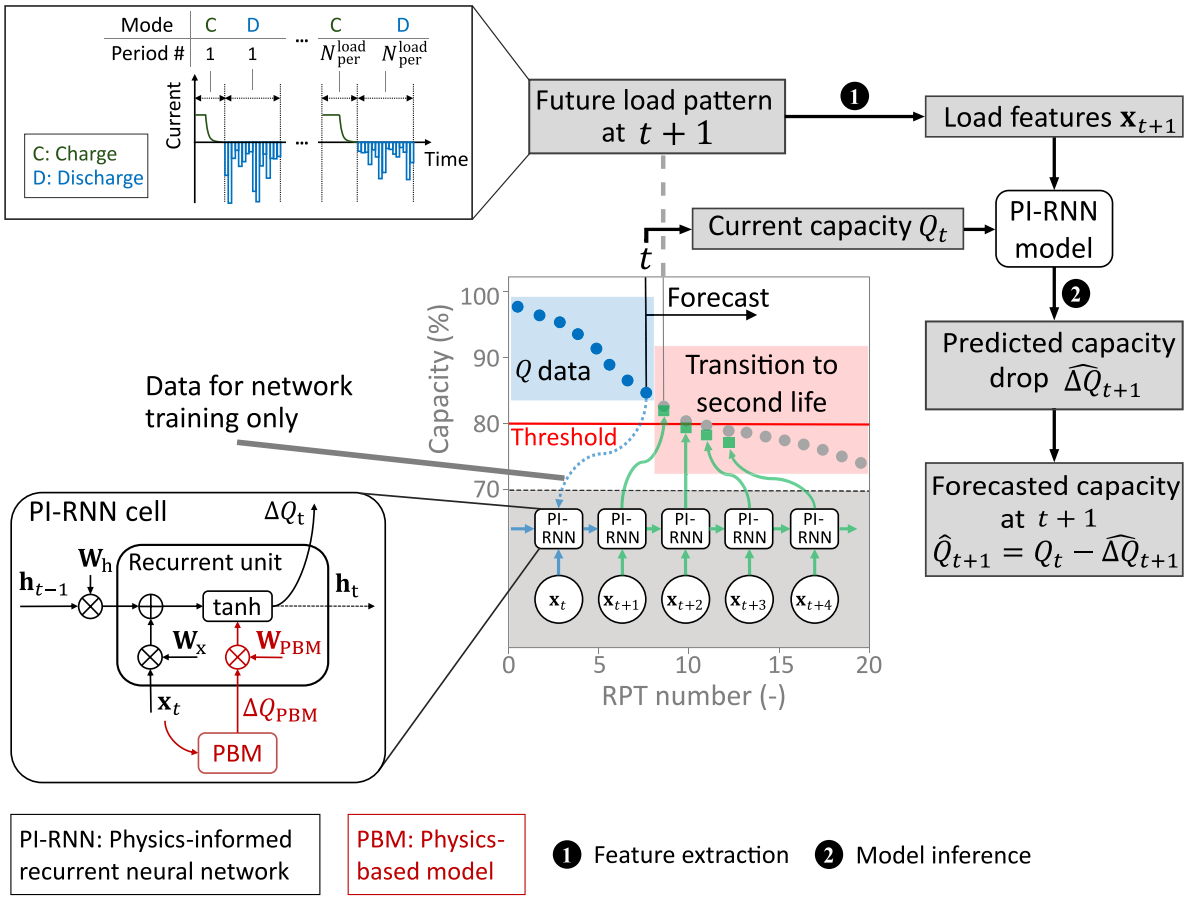

Fig. 7. Overview of the proposed PI-RNN method. The schematic illustrates the mapping from input features of a future load pattern (e.g., at RPT $t+1$, consisting of $N_{\text{per}}^{\text{load}}$ charge/discharge periods) to the predicted capacity drop $\Delta\hat{Q}_{t+1}$. The PI-RNN model shown on the middle right is composed of sequential PI-RNN cells, each incorporating outputs from the PBM to guide learning. The PI-RNN cell architecture on the bottom left integrates PBM-informed degradation into the recurrent update.

4. Methodology

Fig. 7 illustrates the proposed framework for forecasting capacity degradation in a future load pattern. The framework involves extracting input features from the load pattern at RPT $t+1$. These features are fed into a PI-RNN model, which predicts the capacity drop from RPT t to $t+1$. The same process is repeated iteratively for subsequent steps, where each step is forecasted using a single PI-RNN unit. The figure also highlights the structure of a PI-RNN. For each time step, the input features are processed through a physics-based model (PBM), which outputs an estimated capacity drop. This output is then refined by the recurrent unit of the PI-RNN to account for time-dependent corrections and provide the final capacity drop prediction. The sequential nature of the framework ensures that the capacity degradation for each future load pattern is predicted step by step.

4.1. Input feature extraction

The inputs to the PIML model are features extracted from the current time-series data. Inspired by Richardson et al. [33], we adopted a straightforward histogram-based method for feature extraction. Since capacity measurements are only available per load pattern, the first step involves mapping the time-series data to a fixed-size input vector for each load pattern. Specifically, if a load pattern $\mathbf{L}$ consists of $N_{\text{meas}}^{\text{load}}$ measurement time steps, the $N_{\text{meas}}^{\text{load}}$ battery current measurements $\mathbf{I} \in \mathbb{R}^{N_{\text{meas}}^{\text{load}}}$ are transformed into an $n_{\mathbf{x}}$-dimensional input vector $\mathbf{x}$, where $n_{\mathbf{x}}$ represents the number of selected features. Regardless of the number of time steps in $\mathbf{L}$, the input vector $\mathbf{x}$ remains uniform in size. The extracted features for each load pattern $\mathbf{L}$ are based on predefined assumptions regarding stress factors. It is worth mentioning that accurate results can be achieved with a relatively small number of factors that contribute to battery aging, such as the time under specific load conditions and the intensity of the load over a given time period. Our approach mirrors that of Richardson et al. [33], excluding temperature-based features.

The assumption that future current load patterns are available is reasonable, as many real-world applications benefit from insights into expected loads derived from typical usage scenarios. In many industrial settings, historical data enable the derivation of a representative or baseline usage profile, where key load features, such as ampere-hour throughput and duration under specific load conditions, remain relatively stable over time. In contrast, using temperature-based features from the future is not realistic, as the thermal conditions of the cell can vary unpredictably and are difficult to anticipate in advance. A detailed list of the selected input features is provided in Table 1. Please note that the values of I_{low} and I_{high} are determined based on the dataset. For the first dataset, I_{low} is 2 A and I_{high} is 3 A, corresponding to applied C-rates of approximately 0.91C and 1.36C, respectively. These initial bounds were selected following the approach taken in Richardson et al. [33]. In contrast, for the second dataset, I_{low} and I_{high} are 3 A and 4 A, respectively, resulting in applied C-rates of 2.5C and 3.33C, given that the nominal capacity of these cells is 1.2 Ah. The bounds for this dataset were adjusted to reflect the harsher usage conditions, allowing us to define a feature that distinctly characterizes these different scenarios.

Fig. 8 displays joint distribution plots that illustrate the relationship between selected input features and the corresponding capacity changes for each batch in the two aging datasets. Joint distribution

plots reveal detailed characteristics of the data, such as density, skewness, and clustering, which help identify out-of-distribution (OOD) conditions. This insight is particularly valuable for our PIML approach, where robust performance in OOD scenarios is critical. The kernel density contours provide a smooth estimate of the probability density function, revealing the density of observations within the feature space. This technique overlays a kernel on each data point and sums them to visualize the underlying distribution. We categorize the NASA dataset into two batches for consistency with our other case study: batch 2, consisting of cell numbers divisible by 3, and batch 1, which includes the remaining cells according to Table C.7. These designations allow us to maintain uniform terminology across our analyses. While the kernel density contour plots show some overlap between the distributions of batch 1 and batch 2, batch 2 occupies a smaller region of the feature space. Therefore, we chose batch 2 as the primary training set and batch 1 as the test set, as this configuration introduces OOD scenarios in the test data and presents a more challenging learning problem. To further assess the robustness of our proposed method, we plan to gradually reduce the size of the training set (batch 2), thereby limiting feature space coverage and rigorously evaluating model generalization.

In this study, we use SHAP (SHapley Additive exPlanations) values to assess feature importance, providing a unified and interpretable measure of each feature's contribution to the model's output. SHAP values are grounded in cooperative game theory and specifically use Shapley values, which allocate the "credit" for the model's predictions to individual features by considering all possible combinations of features [61]. Mathematically, the Shapley value for a feature x_j is computed by considering all possible subsets of features S that exclude x_j, and calculating the average marginal contribution of x_j across these subsets. The formula for the Shapley value is expressed as:

$$\phi_{x_j} = \sum_{S \subseteq \mathbf{x} \setminus \{x_j\}} \frac{|S|!(|\mathbf{x}| - |S| - 1)!}{|\mathbf{x}|!} \left[\hat{y}(S \cup \{x_j\}) - \hat{y}(S)\right] \tag{1}$$

where ϕ_{x_j} is the Shapley value for feature x_j, $\mathbf{x}$ is the vector of all features in the model (i.e., $\mathbf{x} = \{x_1, x_2, \ldots, x_{n_\mathbf{x}}\}$), S is a subset of features excluding x_j, $\hat{y}(S)$ is the model's prediction given the feature subset S, $\hat{y}(S \cup \{x_j\}) - \hat{y}(S)$ represents the change in the prediction when feature x_j is added to the subset S.

In this study, we use the light gradient boosting machine (LightGBM) model to perform feature importance analysis. LightGBM is a gradient boosting framework that uses decision trees as base learners. By training a LightGBM model, we obtain predictions that are then explained using SHAP values to identify and quantify the influence of each feature on capacity degradation. Our motivation for this analysis is to elucidate the relative impact of different operational metrics, such as ampere-hour throughput and durations under specific load conditions, on battery degradation, thereby linking measurable load features to capacity fade. The feature importance plot in Fig. 9 presents the mean absolute SHAP values for each feature, summarizing their overall contribution to the model's predictions. Each bar represents a feature, with its length indicating the magnitude of its mean absolute SHAP value. Features at the top of the chart have the highest importance, contributing more significantly to the model's output. From this bar chart, we observe that total time elapsed from start, ampere-hour throughput, and the time duration between the two defined loading thresholds are important predictors of battery capacity degradation. In case study 1, where the loading conditions are relatively consistent, the SHAP values are distributed more uniformly across features, indicating that the model relies on a broad set of inputs. In contrast, in case study 2, characterized by a more complex and variable experimental design, the time duration between the two thresholds emerges as a more effective differentiator. This suggests that under more complex loading conditions, specific load-duration metrics provide additional and better insight into capacity fade for cases that include transitions to milder second-life operations.

4.2. Inductive bias via PI-RNNs

Inductive bias approaches involve the design of specialized network architectures that incorporate prior knowledge related to a specific predictive task. This method assigns physics-based information to neurons or layers, as illustrated in Fig. 3c. In our proposed framework, the predicted capacity drop provided by the physics-based model is incorporated into the hidden state update of the RNN through a learnable weight. This integration of physics-based priors serves as an inductive bias. Although widely used in other domains, this aspect of PIML has not been extensively applied to battery research.

RNNs have been successfully applied to model time series data [62]. Fundamentally, RNNs process sequences by applying transformations to a hidden state, described by the recurrence relation:

$$\mathbf{h}_{t+1} = \mathbf{f}(\mathbf{h}_t, \mathbf{x}_{t+1}) \tag{2}$$

Here, $t \in \{0, 1, \ldots, T\}$ represents the time steps (RPT numbers in our case), where T is the total number of time steps. The hidden state $\mathbf{h}_{t+1} \in \mathbb{R}^{n_\mathbf{h}}$ at RPT $t+1$ encodes the temporal dynamics and memory of the sequence. The input vector $\mathbf{x}_{t+1} \in \mathbb{R}^{n_\mathbf{x}}$, which includes the capacity at RPT $t+1$ and six other features extracted from the next time step (as detailed in Table 1), has a dimensionality of $n_\mathbf{x}$. The function $\mathbf{f}(\cdot)$ transforms the previous hidden state $\mathbf{h}_t$ and the current input $\mathbf{x}_{t+1}$ into the new hidden state $\mathbf{h}_{t+1}$, with $n_\mathbf{h}$ representing the dimensionality of the hidden state space. A fully connected linear layer then maps $\mathbf{h}_{t+1}$ to the output, which in this case is the capacity drop.

Fig. 10a illustrates the baseline RNN architecture. In this structure, a fully connected dense layer combines the input at RPT $t+1$ and the hidden state at t to generate the hidden state at RPT $t+1$. Specifically, the hidden state is updated as follows for the baseline RNN model:

$$\mathbf{h}_{t+1} = \tanh(\mathbf{W}_\mathrm{h}\mathbf{h}_t + \mathbf{W}_\mathbf{x}\mathbf{x}_{t+1}) \tag{3}$$

where $\mathbf{x}_{t+1}$ represents the input feature vector at RPT $t+1$, $\mathbf{h}_t$ is the hidden state from the previous time step, $\mathbf{W}_h$ and $\mathbf{W}_\mathbf{x}$ are weight matrices for the hidden state and input, respectively, and $\tanh(\cdot)$ is the hyperbolic tangent activation function, introducing non-linearity to the model. This setup enables the RNN to capture temporal dependencies by updating the hidden state through a combination of past states and current inputs. To forecast capacity degradation for the next M load patterns, we utilize a sequence of M RNN units. Each unit corresponds to RPT $t+k$ for $k = 1, \ldots, M$, and predicts the capacity drop ΔQ_{t+k}. This capacity drop results from the degradation accumulated during the load pattern between RPT $t+k-1$ and RPT $t+k$ and is expressed as:

$$\Delta Q_{t+k} = Q_{t+k-1} - Q_{t+k} \tag{4}$$

The cumulative degradation over M load patterns is computed by summing the predicted capacity drops $(\widehat{\Delta Q}_{t+1}, \ldots, \widehat{\Delta Q}_{t+M})$ from the M RNN units as follows:

$$\hat{Q}_M = Q_t - \sum_{k=t+1}^{t+M} \widehat{\Delta Q}_k \tag{5}$$

In lithium-ion batteries, polarization effects can create long-term dependencies, meaning that distant time steps are correlated. During backpropagation, the gradients used to update the model can gradually shrink with each time step, eventually becoming too small to reach the relevant time steps. This issue, known as the vanishing gradient problem [52], led to the development of more advanced RNN models like LSTM [63] and GRU [64], which help address this challenge. While LSTM and GRU cells effectively mitigate the vanishing gradient problem and capture long-term dependencies, they come with certain drawbacks. Both models are computationally expensive due to their complex gating mechanisms, which involve multiple weight matrices and non-linear transformations at each time step. This increased complexity results in longer training times and higher memory usage. Additionally, for some tasks with short-term dependencies or when

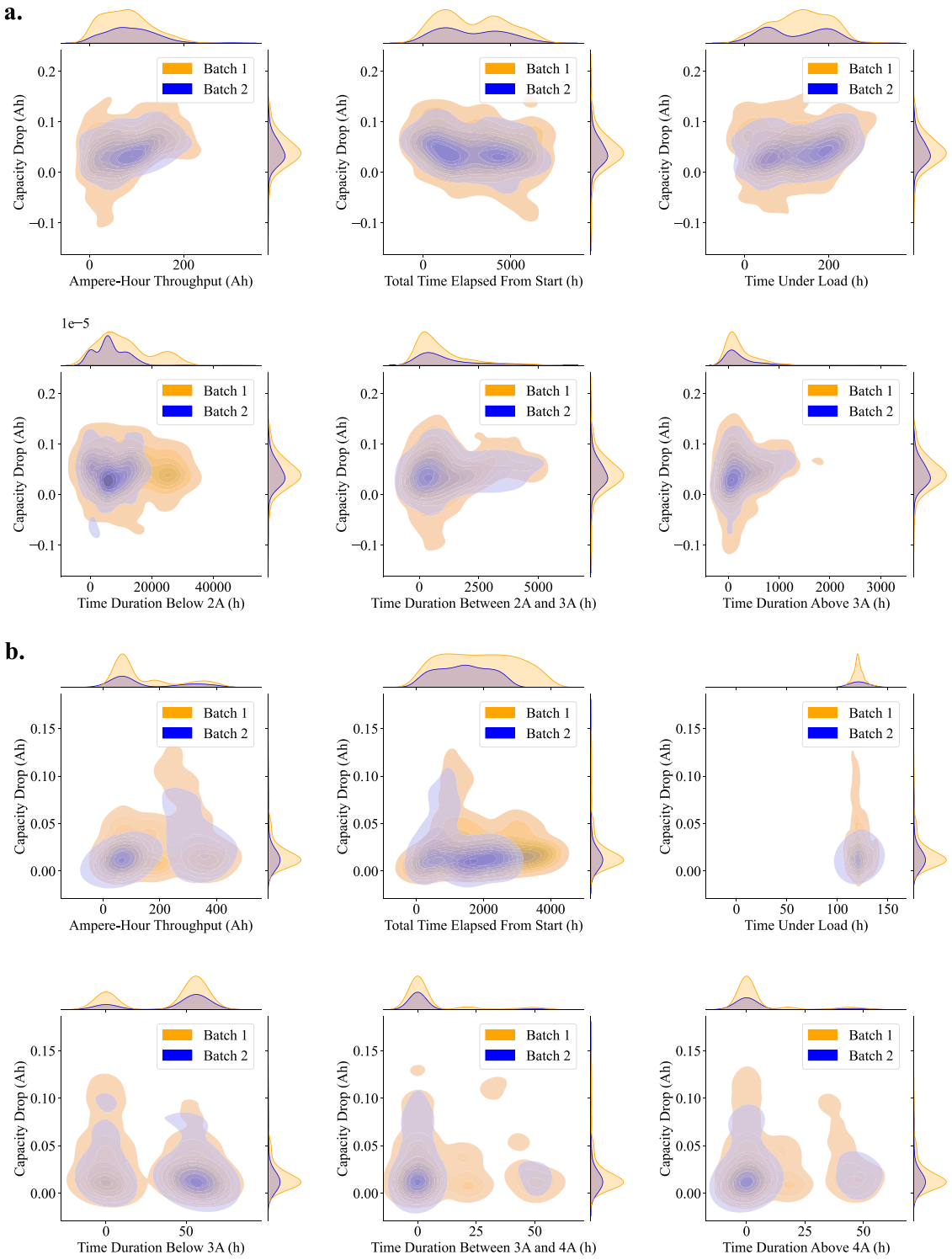

Fig. 8. Joint distribution plots illustrating the relationship between selected input features and the corresponding capacity drop for each batch in **a.** case study 1 and **b.** case study 2.

limited training data is available, simpler architectures may outperform LSTM and GRU due to overfitting risks and unnecessary overhead.

We propose a physics-informed architecture, illustrated in Fig. 10b. In this architecture, capacity drop predictions from a PBM are integrated into an RNN unit. The PBM, trained on simulation data from a P2D model (see Section 4.3), maps input features to the capacity drop for a given load pattern. While not perfect, the PBM aims to assist the RNN in extrapolation by introducing inductive bias. In addition to feeding the input features into the RNN, they are processed by the PBM surrogate, which independently predicts the capacity drop for the upcoming load pattern. The hidden state update is then computed as:

$$\mathbf{h}_{t+1} = \tanh(\mathbf{W}_{\mathrm{h}}\mathbf{h}_t + \mathbf{W}_{\mathbf{x}}\mathbf{x}_{t+1} + \mathbf{h}_{\mathrm{pbm}}) \tag{6}$$

where

$$\mathbf{h}_{\mathrm{pbm}} = \mathbf{W}_{\mathrm{pbm}}\Delta Q_{\mathrm{pbm}}(\mathbf{x}_{t+1}) \tag{7}$$

Here, $\mathbf{h}_{\mathrm{pbm}}$ is adjusted to match the dimensionality of the hidden state ($\mathbf{h}_t$) via a transformation layer. The physics-based prediction is

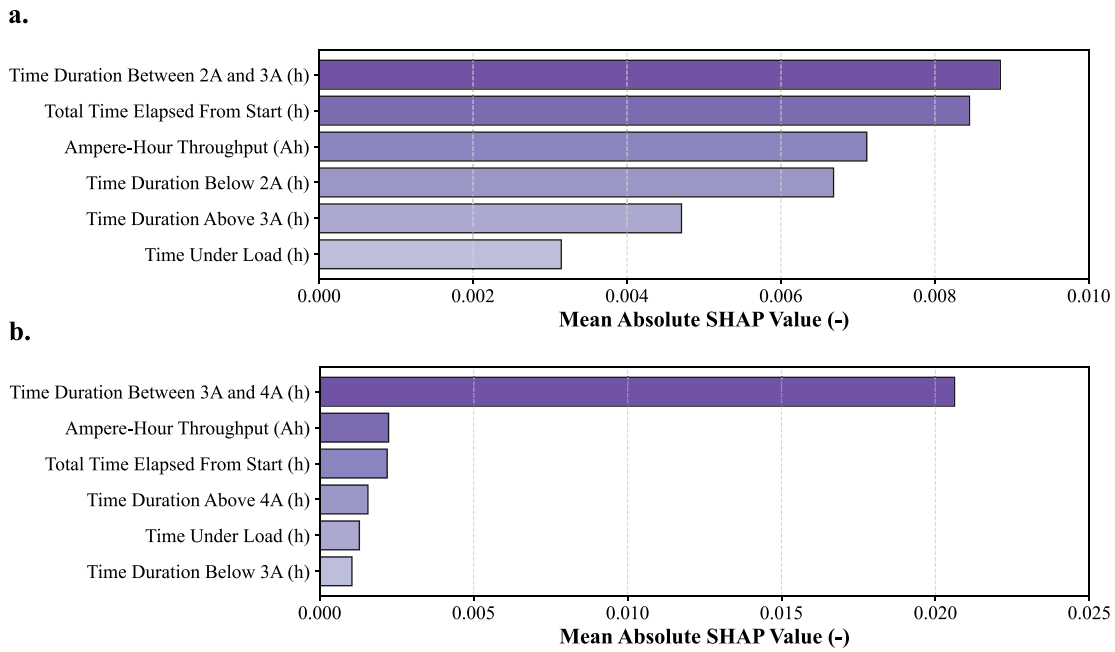

Fig. 9. Feature importance analysis for **a.** case study 1 and **b.** case study 2 using SHAP values. The bar charts display the mean absolute SHAP values for each feature, summarizing their overall contribution to the model's predictions and highlighting the most influential features in each case study.

Table 1
Input features and their descriptions for a single load pattern. Note that T represents time durations and τ represents time steps within each load pattern used for summation or integration, and $I(\tau)$ denotes the current at time τ. I_{low} and I_{high} are predefined thresholds used to classify current ranges in each case study.

Feature	Formulation	Description
RPT number [–]	t	Index representing the position of the RPT cycle in the overall experiment timeline. Used for temporal ordering of load patterns.
Elapsed time during a load pattern [s]	$T_{\text{L},t} = T_{t+1} - T_t$	The total duration of load pattern applied at t or time between two consecutive RPTs (t and $t+1$).
Ampere-hour throughput [Ah]	$Q_{\text{thr}} = \int_{T_t}^{T_{t+1}} \lvert I(\tau)\rvert \, d\tau$	The total charge passed through the cell during the load pattern corresponding to RPT t.
Remaining capacity [Ah]	$Q_t = \int_{\tau_{\text{dchg}_0}}^{\tau_{\text{dchg}_1}} I(\tau)\, d\tau$	Capacity measured at RPT t, calculated as the integral of current $I(\tau)$ over the discharge period. Here, τ_{dchg_0} and τ_{dchg_1} denote the start and end times of the low C-rate discharge step, respectively.
Total elapsed time since start [s]	$T_{\text{total}} = \sum_{k=1}^{t} T_{\text{L},k}$	The cumulative time elapsed from the beginning of the experiment up to RPT t.
Duration with current magnitude $< I_{\text{low}}$ [s]	$T_{\text{low}} = \sum_{\lvert I(\tau)\rvert < I_{\text{low}}} \Delta\tau$	The total time duration within the load pattern when the current magnitude is below I_{low}.
Duration with $I_{\text{low}} \leq$ current magnitude $\leq I_{\text{high}}$ [s]	$T_{\text{mid}} = \sum_{I_{\text{low}} \leq \lvert I(\tau)\rvert \leq I_{\text{high}}} \Delta\tau$	The total time duration within the load pattern when the current magnitude is between I_{low} and I_{high}.
Duration with current magnitude $> I_{\text{high}}$ [s]	$T_{\text{high}} = \sum_{\lvert I(\tau)\rvert > I_{\text{high}}} \Delta\tau$	The total time duration within the load pattern when the current magnitude is above I_{high}.

incorporated with the previous hidden state through the activation function, with a learnable weight parameter for the PBM component. This parameter allows the network to dynamically control the contribution of the PBM output. Ultimately, the RNN unit uses the hidden state update to correct the PBM prediction, leveraging both data-driven and physics-based contributions.

By integrating physics-based predictions from the PBM, PI-RNNs leverage the PBM's ability to capture general trends in the input–output relationships, even though the PBM predictions may lack accuracy. The recurrent units refine these predictions by adjusting them based on the available data, reducing the need for the RNNs to learn the underlying trends from scratch. This guided approach ensures a more efficient learning process and improved generalization, particularly in data-scarce or noisy scenarios. Learnable weight parameters dynamically balance the contributions of physics-based predictions, enabling PI-RNNs to be adaptable to diverse battery health prognostic tasks.

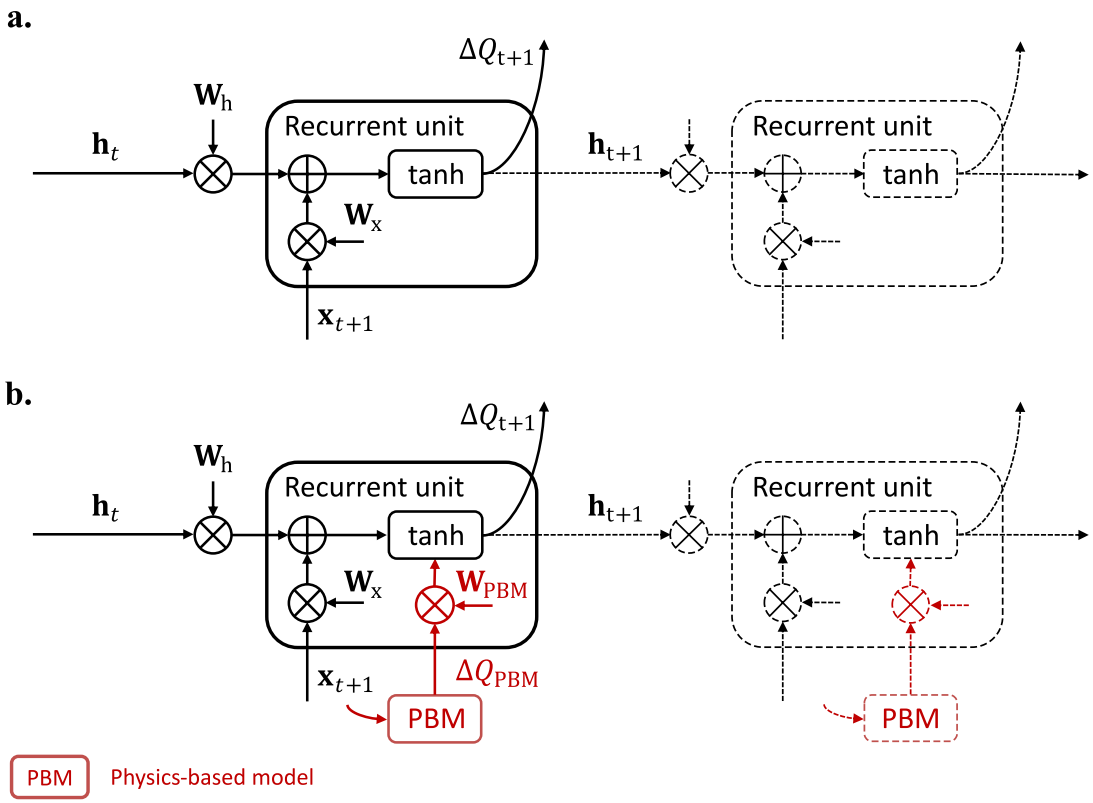

Fig. 10. Structures of the **a**. baseline RNN and **b**. physics-informed RNN.

4.3. Physics-based P2D model

The most prevalent physics-based lithium-ion battery model, introduced by Newman and his colleagues, describes the behavior of porous electrode batteries and is commonly referred to as the P2D model [65]. This model operates under the assumption that each point within the electrode is represented by a spherical particle of active material, incorporating one spatial dimension for the electrode thickness and another for the particle radius as shown in Fig. 11 [66]. The P2D model has been extensively researched and serves as a robust tool for estimating, optimizing, and predicting battery performance.

To fully characterize the physical, chemical, and electrochemical properties of a cell using the P2D model, more than 30 parameters are typically required, with the exact number varying based on the model's specifications. Obtaining accurate parameter values can be challenging and time-consuming, often demanding various characterization techniques. This challenge is compounded for commercial cell chemistries, as manufacturers do not disclose critical information regarding parameters, compositions, or components in their specifications. Consequently, modeling practitioners often resort to parameter sets derived from existing literature [57].

In this work, we focus on using a physics-based model to simulate degradation trends across different cell groups rather than developing a high-fidelity model. Achieving high accuracy would require extensive experimental parameterization, which is impractical for this study. Instead, we employ the Python Battery Mathematical Modeling (PyBaMM) package [67] and its P2D model to simulate discharge capacity under cycling conditions aligned with our case studies. The P2D model serves as a low-fidelity tool to estimate battery behavior, capturing general degradation trends rather than providing highly accurate predictions. In case study 2, it highlights aggressive capacity fade during the first-life phase and a slower degradation rate after the transition point at 80% SOH. This approach prioritizes computational efficiency over detailed accuracy, making it suitable as a starting point for further refinement.

To address the limitations of the P2D model, particularly its inability to capture long-term degradation trends accurately, we integrate a customized RNN to correct capacity fade predictions. This hybrid approach effectively captures complex degradation patterns while maintaining computational efficiency, avoiding the extensive parameterization required for a high-fidelity physics-based model. The following assumptions are made for the P2D model:

- A fixed number of cycles is assumed between each RPT for each group of cells. However, the number of cycles varies relatively across different groups to reflect varying usage conditions.
- In case study 2, we assumed that we have access to an estimate of the number of load patterns required to reach the end of first life for each group of cells.
- We followed the framework by O'Kane et al. [68], which modeled four coupled degradation mechanisms: SEI growth, lithium plating, particle cracking, and stress-driven loss of active material. These mechanisms are considered in our simulations, with SEI growth and particle cracking coupled through the "SEI on cracks" option in PyBaMM, and particle cracking and stress-driven loss of active material inherently linked via stress–strain relationships within the particles. For lithium plating, we considered partially reversible plating, where the SEI influences both reversible and irreversible components. SEI growth in our simulations follows the solvent-diffusion-limited mechanism available in PyBaMM, which considers transport limitations but does not explicitly model solvent consumption or electrolyte depletion. These effects have been addressed in more recent works [69,70]. However, we did not include them here, as our goal was not to improve standalone PBM accuracy.

To fit the P2D model for each of the two case studies, which feature different cell chemistries, we used consistent electrode characteristics. For each case, a separate electrochemical P2D model was used. Rather than measuring specific parameters, we used values from

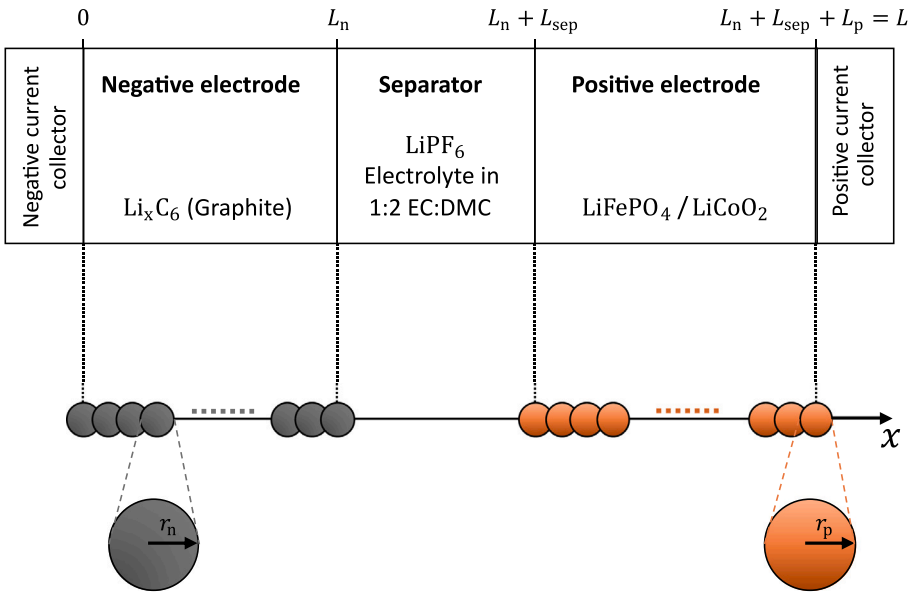

Fig. 11. Schematic of P2D model for a lithium-ion battery cell.

the literature that match commercial cells with similar chemistries or adjusted assumed values to achieve the best fit to the experimental low current voltage response at room temperature (Fig. B.24). The final fitted parameters are shown in Table B.6. Essential intercalation parameters were set based on the specific chemistries of the electrodes and electrolyte in the experimental cells, while geometric parameters, like electrode and separator dimensions, were measured for case study 2 and taken from studies of comparable chemistries for case study 1. Where literature data was unavailable, we calibrated remaining parameters to achieve an optimal fit. The purpose of using the P2D model here is not to develop a perfectly accurate physics-based model but to leverage its ability to integrate multiple degradation mechanisms. Rather than calibrating the model to experimental data across different aging conditions, we intentionally employ it as a low-fidelity approximation to provide meaningful inductive bias when integrated into the proposed PI-RNN architecture. This allows us to explore how fundamental degradation processes, such as SEI growth, lithium plating, and active material loss, impact capacity fade trends under the same conditions as our experimental protocols, and to enhance the RNN's ability to generalize and extrapolate in data-limited regimes. While these mechanisms are physically grounded and appropriate for modeling generic degradation pathways, they may not reflect the dominant aging phenomena in our specific cells and use cases. Nevertheless, the PBM successfully introduces a physics-based prior and inductive bias into the learning process. Accepting its limited standalone accuracy is a deliberate design choice, made to prioritize improved generalization within the PI-RNN over precise model fidelity.

We solve the P2D model as outlined in Table 2, with four key variables: the concentration of lithium in the electrode particles $c_{s,k}$, the ion concentration in the electrolyte $c_{e,k}$, the electrode potential $\phi_{s,k}$, and the electrolyte potential $\phi_{e,k}$. The governing equations for these variables are derived from the principles of mass and charge conservation. Lithium transport within the electrode particles is governed by Fickian diffusion, while charge transport follows Ohm's law. The electrolyte dynamics are modeled using concentrated solution theory. These conservation laws are interconnected through the intercalation reaction, which is described by the Butler–Volmer equation, assuming symmetry in the reaction. For further details on these equations, readers are referred to standard battery modeling handbooks [66]. The equations in Table 2 are defined separately for each region: the positive electrode, separator, and negative electrode. Thus, the second subscript k in the variables can take one of three values: p, s, or n, corresponding to the positive electrode, separator, and negative electrode, respectively. Since the electrolyte spans all three regions, we must enforce continuity conditions at the interfaces between the electrodes and the separator (at $x = L_n$ and $x = L_p$, respectively). These conditions ensure the continuity of concentration, potential, molar flux, and current. In our case, we define the reference potential as $\phi_{s,n} = 0$ at $x = 0$. In the equations, F represents the Faraday constant, and R is the universal gas constant. The model assumes isothermal conditions, meaning that the system's temperature T is held constant.

Once fitted, the model was used to simulate load patterns for the conditions of each group of cells listed in Tables A.4 and A.5. To simulate the complete version of the experiments, we defined the RPTs in the same manner as in the experimental datasets between each pair of load patterns. We then ran the simulations for a fixed number of cycles and load patterns in each group.

5. Results and discussion

This section presents the main results. We begin with physics-based simulations to inform model design. We then benchmark PI-RNN against baseline methods across various scenarios and perform sensitivity analyses. Finally, we conduct predictive uncertainty quantification for PI-RNN to be able to convey model confidence in long-term forecasts.

5.1. Physics-based simulation of capacity fade

Here, we present the capacity fade plots generated by simulating cycling conditions with the developed P2D model for each case study. As outlined earlier, we first created a P2D model specific to each case study, then simulated the cycling conditions using the same cycling and RPT protocols as in the experimental setup. Fig. 12 displays the resulting capacity fade plots for each group within the case studies, fixed at a set number of load patterns. Groups with SOH-based derating have been excluded from these plots, as tracking capacity after each RPT is computationally impractical for simulations. Additionally, we assume access to the relative cycle counts and transition points from the first life to the second life regime, as the simulation data is designed to mimic the limited experimental data available for training purposes.

The results presented herein offer an initial estimate of the capacity fade rate and associated trends derived from the physics-based model. Although these estimates are not accurate, they differentiate between

Table 2
P2D model equations. The table provides the governing equations, boundary conditions, and initial conditions for a P2D model. Key variables include: (i) $c_{s,k}$: the lithium concentration in the electrode particles, (ii) c_e: the lithium concentration in the electrolyte, (iii) $\phi_{s,k}$: the electrode potential, and (iv) ϕ_e: the electrolyte potential.

Description	Equation	Boundary conditions
	Electrodes	
Mass conservation	$\frac{\partial c_{s,k}}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 D_{s,k}\frac{\partial c_{s,k}}{\partial r}\right)$	$\left.\frac{\partial c_{s,k}}{\partial r}\right\|_{r=0} = 0, \quad -D_{s,k}\left.\frac{\partial c_{s,k}}{\partial r}\right\|_{r=r_k} = \frac{J_k}{a_k F}$
Charge conservation	$\frac{\partial}{\partial x}\left(\sigma_k \frac{\partial \phi_s}{\partial x}\right) = J_k$	$\left.\frac{\partial \phi_s}{\partial x}\right\|_{x=0} = 0, \quad \left.\frac{\partial \phi_s}{\partial x}\right\|_{x=L_k} = -i_{\text{app}}$
	Electrolyte	
Mass conservation	$\frac{\partial c_e}{\partial t} = \frac{\partial}{\partial x}\left(D_e \frac{\partial c_e}{\partial x}\right) + \frac{1-t_+^0}{F} J_k$	$c_e\|_{x=0} = c_e\|_{x=L}$
Charge conservation	$\frac{\partial}{\partial x}\left(\kappa_e \frac{\partial \phi_e}{\partial x}\right) - \frac{2RT}{F}\frac{1-t_+^0}{F}\frac{\partial}{\partial x}\log c_e = -J_k$	$\phi_e\|_{x=0} = \phi_e\|_{x=L}$
	Reaction kinetics	
Butler–Volmer equation	$J_k = a_k i_0 \left[\sinh\left(\frac{\alpha_a F \eta_k}{RT}\right)\right]$	$\eta_k = \phi_s - \phi_e - U(c_{s,k})$
Exchange current density	$i_0 = k_0 (c_e)^{\alpha_c} (c_{s,k}^{\max} - c_{s,k})^{\alpha_a} (c_{s,k})^{\alpha_c}$	
	Initial Conditions and terminal voltage	
Initial conditions	$c_{s,k} = c_{s,k,0}, \quad c_e = c_{e,0}$	
Terminal voltage	$V = \phi_s\|_{x=L} - \phi_s\|_{x=0}$	

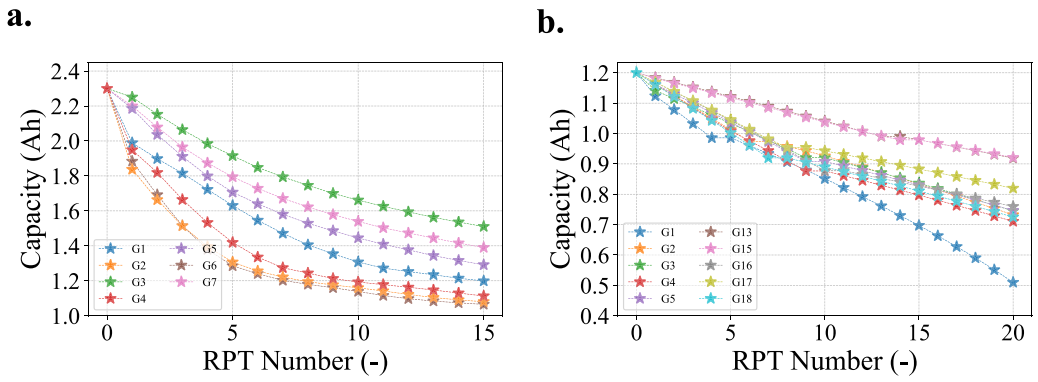

Fig. 12. Simulated capacity fade plots for **a.** case study 1 and **b.** case study 2. For both case studies, each group is simulated using the same cycling conditions as the corresponding experiments.

the groups based on their respective cycling conditions. We extracted the same features from the simulation data (see Section 4.1) as those utilized in the experimental datasets. To further reduce computational complexity and allow for efficient training, we construct a surrogate model that maps the input features to the corresponding capacity drop estimated by the PBM over each simulated load pattern. This surrogate model is used only during the training phase, as the full P2D model operates as an offline data generator. The P2D model's computationally intensive operations are confined to generating training data, and it is not iteratively updated during prediction. While the surrogate greatly speeds up model training and deployment, it introduces an additional layer of approximation, leading to further reductions in predictive accuracy. Additionally, the PBM parameters were primarily fitted under low C-rate conditions and were not recalibrated for more aggressive cycling protocols, which contributes to the discrepancies observed between simulated and experimental results. Although the PBM surrogate is not perfectly accurate, it provides a meaningful initial estimate and embeds a physics-based bias into the model. This initial estimator is then further refined and incorporated through the recurrent units within the PI-RNN architecture, therefore enhancing overall prediction quality.

5.2. Comparative results

In this section, we evaluate the performance of the proposed PI-RNN method alongside a few baseline methods, including a standalone PBM, a standard RNN, and a GPR. These methods were tested across both case studies to identify scenarios in which the proposed PI-RNN method demonstrates effectiveness. The PBM model, used independently here, is identical to the one integrated within the PI-RNN, serving as a baseline for comparison. RNN and GPR models were employed as the purely data-driven baselines. The baseline RNN model shares the same architecture and input features as the PI-RNN, with the primary difference being that it directly maps input features to the capacity drop (illustrated in Fig. 10a), whereas the model-based GPR uses a fixed number of available capacity points as its input. The proposed PI-RNN first uses a PBM surrogate to produce an initial estimate of the capacity drop, which is then incorporated and refined by the RNN units (Fig. 10b). Table 3 summarizes the hyperparameters and training setup used in both the baseline RNN and PI-RNN models, showcasing a consistent and straightforward setup. Hyperparameters were optimized with Optuna [71] for the baseline RNN (details provided in the Appendix E). Optimizing for different training setups is beyond the scope of the present study. The models were designed with the goal of achieving accurate and consistent predictions using the simplest possible structure. By maintaining identical hyperparameters across both models, we ensure a fair comparison, isolating the specific contributions of the PBM within the PI-RNN method. Furthermore, the networks were initialized identically to minimize variations due to the randomness of initialization, enabling a focused evaluation of the added value provided by the PBM-informed structure. Also, both models are trained using a maximum-horizon approach with a sequence length of 10 steps and evaluated on sequences of length 1 to 10. Additionally, comparative results between the baseline RNN, LSTM, their augmented versions, and the proposed PI-RNN model are provided in Appendix D.

To benchmark the proposed PI-RNN method, we also compare it against a state-of-the-art prognostics approach: a Gaussian process (GP)

Table 3
Hyperparameters and training settings shared between the proposed PI-RNN and baseline RNN models.

Hyperparameter	Setting
Hidden layer size	50
Learning rate	0.001
Optimizer	Adam
Loss function	MSE
Early stopping patience	25–100
Number of epochs	1500–2500
Evaluation metric	MAE, RMSE

model-based method, referred to as GPR in this section. This method integrates an empirical capacity fade model as its trend function. Often known as explicit mean functions (EMFs) or semi-parametric GPRs, this approach enables the incorporation of prior knowledge regarding the expected functional form of the degradation process [32]. An empirical degradation model predicting capacity fade can be expressed in the form

$$Q = m(t; \boldsymbol{\theta}_{\text{emp}}) \tag{8}$$

where $m(t; \boldsymbol{\theta}_{\text{emp}})$ represents the empirical capacity fade model with parameters $\boldsymbol{\theta}_{\text{emp}}$. If additional contributions to capacity fade, not captured by the empirical model, are suspected, the formulation can be extended by combining the empirical model with a GPR model:

$$Q = m(t; \boldsymbol{\theta}_{\text{emp}}) + f(t; \boldsymbol{\theta}_{\text{ker}}) + \epsilon \tag{9}$$

where $m(t; \boldsymbol{\theta}_{\text{emp}})$ provides the trend function, $f(t; \boldsymbol{\theta}_{\text{ker}})$ captures the residual contributions, and ϵ is the noise term. The GPR model is defined as:

$$f(t; \boldsymbol{\theta}_{\text{ker}}) \sim \mathcal{GP}(m(t), k(t, t'; \boldsymbol{\theta}_{\text{ker}})) \tag{10}$$

where $m(t) = m(t; \boldsymbol{\theta}_{\text{emp}})$ is the trend function derived from the empirical model, and $k(t, t'; \boldsymbol{\theta}_{\text{ker}})$ is the covariance kernel function, which governs the similarity between data points. In this work, we used the radial basis function (RBF) kernel, also known as the squared exponential kernel, expressed as:

$$k(t, t'; \boldsymbol{\theta}_{\text{ker}}) = \sigma_s^2 \exp\left(-\frac{(t - t')^2}{2\ell^2}\right) \tag{11}$$

where σ_s^2 is the signal variance, controlling how much the function can vary vertically in the output dimension, and ℓ is the length scale, controlling how quickly correlations decay with distance. Therefore, $\boldsymbol{\theta}_{\text{ker}} = [\sigma_s^2, \ell]$. The RBF kernel is particularly effective for modeling smooth functions and is a common choice for capturing residual trends in capacity degradation.

The empirical model, widely employed in the literature and specifically designed for the first-life phase of batteries [72], is initially fitted to a fixed number of load patterns (6 to 12 RPTs depending on the group in case study 1) and to all first-life data in case study 2. This approach aligns with the common practice of using capacity data up to the current cycle for training, with varying definitions of the current cycle employed in the literature. We utilize an explicit mean function of the form $m(t; \boldsymbol{\theta}_{\text{emp}}) = a_1 + a_2 \exp(a_3 t)$, where the model parameters are defined as $\boldsymbol{\theta}_{\text{emp}} = [a_1, a_2, a_3]$. This function is equivalent to the degradation model used by Goebel et al. [73] and Wang et al. [74]. This model captures the primary trend of capacity degradation. Residuals, calculated as the difference between the true capacity values and the predictions from the empirical model, are then used to train a GPR model. During online prediction, the empirical model provides an initial estimate of the capacity trend, while the GPR predicts the residuals. The final prediction is obtained by combining the empirical model's estimate with the GPR's residual prediction.

All baseline models were tuned with their hyperparameters empirically selected to achieve the best possible performance, ensuring a robust and fair comparison. The purely data-driven baselines were selected to serve two distinct purposes: the model-based GPR, which represents a state-of-the-art approach in capacity prognostics using different input features, and the baseline RNN, which shares the same inputs as the PI-RNN and is intended to isolate and highlight the contributions of the physics-informed architecture in the proposed model. The forecasting accuracy of each model was evaluated using the root mean square error (RMSE) metric at each forecasting horizon M, calculated as follows:

$$\text{RMSE}(M) = \sqrt{\frac{1}{M\, N_{\text{test}}} \sum_{n=1}^{N_{\text{test}}} \sum_{t=1}^{M} \left(\hat{y}_{n,t} - y_{n,t}\right)^2} \tag{12}$$

where $y_{n,t}$ denotes the true value for the tth forecast of the nth test sample, $\hat{y}_{n,t}$ is the corresponding prediction.

5.2.1. Forecasting strategies

To assess and compare forecasting strategies, we examine three distinct training and evaluation approaches: fixed-horizon, recursive, and maximum-horizon forecasting. These strategies differ in their training objectives, prediction horizons, and evaluation protocols.

In strategy S1 (fixed-horizon forecasting), the model is trained to predict capacity over a fixed forecast length M_{fh}. The training loss is defined as

$$\mathcal{L}^{\text{S1}}_{\text{train}}(\boldsymbol{\theta}_{\text{RNN}}) = \frac{1}{M_{\text{fh}} N_{\text{train}}} \sum_{n=1}^{N_{\text{train}}} \sum_{t=1}^{M_{\text{fh}}} \left(\hat{y}_{n,t} - y_{n,t}\right)^2 \tag{13}$$

At test time, the same model is used to forecast M_{fh} steps using the learned parameters $\boldsymbol{\theta}^*_{\text{RNN}}$, with the evaluation loss given by

$$\mathcal{L}^{\text{S1}}_{\text{test}}(\boldsymbol{\theta}^*_{\text{RNN}}) = \frac{1}{M_{\text{fh}} N_{\text{test}}} \sum_{n=1}^{N_{\text{test}}} \sum_{t=1}^{M_{\text{fh}}} \left(\hat{y}_{n,t} - y_{n,t}\right)^2. \tag{14}$$

While effective for short-term prediction, this approach requires training separate models for each desired forecast length M_{fh}, increasing deployment complexity.

In strategy S2 (recursive forecasting), the model is trained for one-step prediction (i.e., $M_{\text{rec}} = 1$) using

$$\mathcal{L}^{\text{S2}}_{\text{train}}(\boldsymbol{\theta}_{\text{RNN}}) = \frac{1}{N_{\text{train}}} \sum_{n=1}^{N_{\text{train}}} \left(\hat{y}_{n,1} - y_{n,1}\right)^2. \tag{15}$$

At inference time, the model is recursively applied to generate a forecast over M_{rec} steps. The test loss is

$$\mathcal{L}^{\text{S2}}_{\text{test}}(\boldsymbol{\theta}^*_{\text{RNN}}) = \frac{1}{M_{\text{rec}} N_{\text{test}}} \sum_{n=1}^{N_{\text{test}}} \sum_{t=1}^{M_{\text{rec}}} \left(\hat{y}^{(\text{rec})}_{n,t} - y_{n,t}\right)^2. \tag{16}$$

Here, $\hat{y}^{(\text{rec})}_{n,t}$ is the recursively generated forecast at time t, where $\hat{y}^{(\text{rec})}_{n,t-1}$ is fed into the model to produce the next prediction. This approach often suffers from compounding errors.

In strategy S3 (maximum-horizon forecasting), the model is trained on full sequences of length M_{max}. The training loss is

$$\mathcal{L}^{\text{S3}}_{\text{train}}(\boldsymbol{\theta}_{\text{RNN}}) = \frac{1}{M_{\text{max}} N_{\text{train}}} \sum_{n=1}^{N_{\text{train}}} \sum_{t=1}^{M_{\text{max}}} \left(\hat{y}_{n,t} - y_{n,t}\right)^2. \tag{17}$$

During evaluation, the model can be used to predict up to M_{max} steps, but we evaluate only the first $M_{\text{eval}} \leq M_{\text{max}}$ steps to maintain consistency

$$\mathcal{L}^{\text{S3}}_{\text{test}}(\boldsymbol{\theta}^*_{\text{RNN}}) = \frac{1}{M_{\text{eval}} N_{\text{test}}} \sum_{n=1}^{N_{\text{test}}} \sum_{t=1}^{M_{\text{eval}}} \left(\hat{y}_{n,t} - y_{n,t}\right)^2. \tag{18}$$

Although it may require a larger hidden size to model long sequences, this approach enables forecasting at arbitrary horizons with a single model, offering superior flexibility and scalability.

As illustrated in Fig. 13, each subplot presents the PI-RNN and baseline RNN predictions under the three forecasting strategies. The

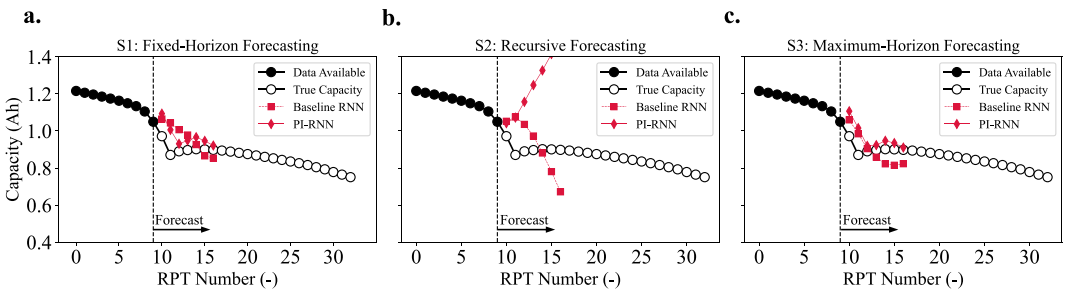

Fig. 13. Trajectory predictions for a test cell (G3C1) under three forecasting strategies: **a.** S1: fixed-horizon, **b.** S2: recursive, and **c.** S3: maximum-horizon.

recursive approach suffers from noticeable drift over time due to error accumulation. Although the fixed-horizon strategy offers better short-term accuracy, it requires training a separate model for each forecast horizon, which limits scalability and flexibility. The maximum-horizon forecasting strategy strikes a balance between accuracy and efficiency, delivering competitive performance with a single model capable of handling variable forecasting horizons. Therefore, we adopt this approach to train baseline RNN and PI-RNN models in all the following evaluations.

5.2.2. Single-step forecasting

We assess the performance of the proposed PI-RNN and the baseline models in single-step forecasting, where the model predicts the capacity for the next step based on the current capacity and features extracted from the subsequent load pattern. For case study 1, the primary train/test split involves using cell numbers divisible by 3 for training (batch 2), while the remaining cells are designated for testing (batch 1) as described in Section 4.1. In case study 2, batch 1 and batch 2 are defined according to Fig. 5, with batch 2 used for training and batch 1 for testing.

The results for true versus predicted capacity in single-step forecasting are shown in Fig. 14 for both case studies. In case study 1, the trend of capacity fade follows a predictable pattern, which is effectively captured by an empirical model within the model-based GPR framework. Thus, while the proposed PI-RNN demonstrates an advantage over the baseline models, both the GPR and baseline RNN models perform well. This performance can be attributed to the consistent cycling protocols in this case study, which remained unchanged even after reaching the 80% SOH threshold in some of the cells. As a result, the capacity trajectories exhibit smooth, trend-like behavior with minimal abrupt variations, making them more predictable and easier to model. In case study 2, however, the scenario is more complex, featuring a transition in capacity degradation behavior after reaching the end of the first-life (i.e., 80% of nominal capacity), due to a shift to milder conditions. Here, the benefits of the PI-RNN are more pronounced, as its error rate is lower than that of the baseline models. This demonstrates the PI-RNN's enhanced effectiveness in handling complex, non-linear forecasting scenarios where traditional models struggle to capture abrupt changes in degradation patterns.

5.2.3. Multi-step forecasting

To evaluate the performance of a PI-RNN model in long-term forecasting, we assess its accuracy alongside baseline models in a multi-step forecasting scenario. The primary objective is to predict the remaining capacity after a series of upcoming load patterns, given the current capacity and features extracted from future load patterns. A PI-RNN incorporates sequential dependencies by leveraging an evolving hidden state, which captures temporal dynamics over multiple forecast steps. The input to a PI-RNN consists of the current capacity and features corresponding to a sequence of M subsequent load patterns, enabling the direct prediction of the capacity trajectory across M steps. This implementation mitigates the error accumulation typically associated with recursive methods by simultaneously predicting the entire trajectory for the specified forecast horizon. At each step, a PI-RNN dynamically refines a PBM's initial estimate of capacity drop by utilizing the hidden state, which encodes information from all prior steps, in conjunction with learned weights and input features. The hybrid architecture combines physics-informed priors from a PBM with RNNs' ability to learn sequential and interdependent relationships between inputs and capacity degradation. The proposed PI-RNN forecasting method proves particularly advantageous in scenarios like case study 2, where the dataset exhibits an abrupt transition from the first to the second life. These transitions introduce non-linear, horizon-specific behaviors in capacity degradation that are effectively captured by the trained PI-RNN models. This is demonstrated in Fig. 15 for cell C1 in group G3. The baseline GPR model (referred to as GPR in the figure) follows the same trend observed in the first-life stage, failing to adapt to the second-life behavior. A low-fidelity PBM provides an initial estimate that captures the general degradation pattern: a faster degradation rate in the first-life stage followed by a milder rate after the transition point. This initial estimate from the PBM is further refined by RNN components in the proposed PI-RNN structure. As a result, the PI-RNN achieves a more accurate prediction of capacity degradation, capturing both the first-life and second-life regions.

The GPR and PBM baseline models treat forecasting as independent tasks for each horizon, lacking mechanisms to capture sequential relationships or cumulative degradation effects. In the GPR baseline, an empirical model provides a general trend of capacity degradation, while a GPR model refines the predictions by modeling the residuals. However, this approach assumes that residuals are independent across horizons, failing to incorporate temporal dependencies or interactions between successive steps. This limitation restricts its ability to accurately model long-term degradation dynamics, particularly during the transition from first to second life. Furthermore, the reliance on a fixed empirical trend constrains the GPR model's adaptability, making it less effective in scenarios where degradation behaviors deviate. In PBM simulations, capacity degradation is modeled based on a set of predefined degradation mechanisms that update select electrochemical parameters. Although these updates incorporate important physics-based insights, they only partially capture the full dynamic evolution of the system. Over extended horizons, this limited parameter adaptation can lead to drift, particularly in scenarios where degradation accelerates or undergoes non-linear transitions. Therefore, while the PBM simulations provide valuable insights into degradation trends, their long-term forecasting reliability may be constrained by the incomplete representation of the evolving electrochemical state.

Multi-step forecasting accuracy across varying forecasting steps for both case studies is illustrated in Fig. 16. The GPR model performs well in case study 1, where there is no transition from first to second life and the cycling conditions remain consistent throughout the entire degradation trajectory. This success can be attributed to the empirical model's ability to capture the degradation trend under these stable operating conditions. Although the cycling conditions in case study 1 are dynamic, they are applied within a relatively narrow range (up to 2.1C), resulting in degradation trends that are both consistent and predictable. This allows the empirical model, fitted on early degradation

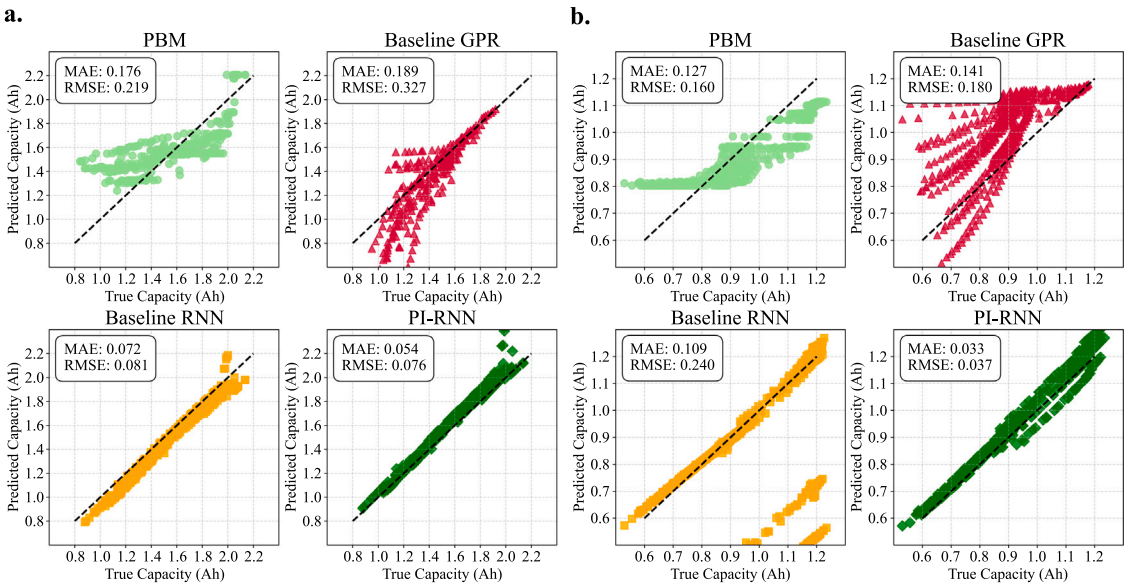

Fig. 14. Single-step forecasting results of the proposed PI-RNN and baseline models for **a.** case study 1 and **b.** case study 2.

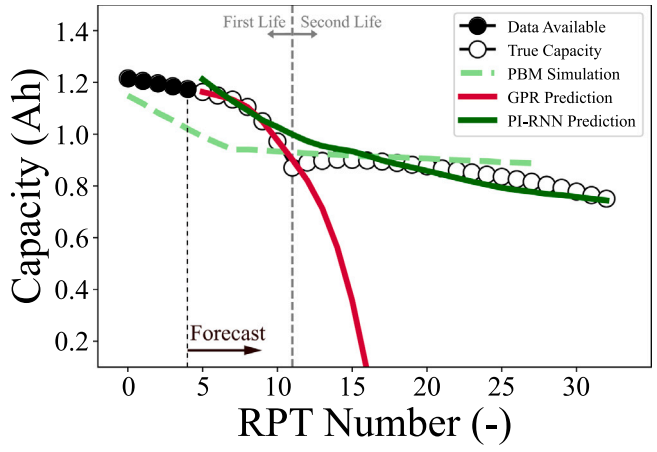

Fig. 15. Capacity degradation trajectories for a test cell (G3C1), highlighting the distinct behaviors of predictive models during the transition from first- to second-life phases.

behavior, to extrapolate the trend over the full degradation trajectory. Furthermore, for shorter-term forecasting (forecasting steps ≤ 5), baseline GPR outperforms PI-RNN, reflecting the empirical model's strong alignment with short-term degradation trends. However, as the forecast horizon increases (forecasting steps ≥ 6), PI-RNN's recurrent updates yield lower errors in these longer-term forecasting. In contrast, the performance of the GPR model declines in case study 2, where abrupt transitions occur between the first-life and second-life phases, and the fixed empirical trend is less effective at capturing the evolving degradation behavior.

The PBM baseline underperforms in case study 1 largely because we only had access to generic cell parameters and geometric details (e.g., electrode thicknesses) were unavailable. Additionally, other degradation mechanisms not represented in the model may become dominant in these older cells. However, in case study 2, PBM performs particularly well for longer-term forecasting horizons (≥ 6 steps), capturing the abrupt transition to second life and outperforming purely data-driven baselines in that phase. The error bars indicate one standard deviation of the RMSE computed from three independent runs (for PBM, baseline RNN, and PI-RNN) initialized with different random seeds, or the standard deviation of per-cell RMSE across all test cells at each forecasting horizon (for baseline GPR).

The proposed PI-RNN method demonstrates consistently strong performance across both case studies and longer forecasting horizons. In case study 1, while the GPR method performs well under stationary usage conditions, the PI-RNN method still yields competitive results. In case study 1, RMSE values for all four methods lie within a narrow band (roughly 0–0.4 Ah) with largely overlapping error bars (see Fig. 16a), demonstrating that purely data-driven approaches can provide reliable short- to medium-horizon forecasts under stationary cycling conditions. However, in case study 2, characterized by a wider operating range and abrupt transitions from first to second life, PI-RNN sustains low RMSEs at extended horizons, whereas PBM, GPR, and baseline RNN produce sharply increased errors. This superior performance stems from PI-RNN's integration of temporal sequence modeling with physics-informed inputs, enabling it to better adapt to complex, non-stationary degradation dynamics.

To illustrate the comparative performance of the models in trajectory forecasting, a selection of test cells is used to showcase each model's ability to follow the capacity degradation trajectory at various stages of aging. This approach provides a closer look at how well each model adapts to the nonlinear capacity degradation patterns over time, highlighting strengths and limitations in their predictive capabilities throughout different stages of the cells' lifetime. Fig. 17 presents capacity forecasts by the proposed PI-RNN and baseline models for three test cells from groups G3, G8, and G13 in case study 2, spanning three distinct life stages. The first stage corresponds to the cells' first-life region, the second stage captures the transition from first to second life, and the third stage focuses on the cells' second life. Performance across the degradation stages reveals distinct strengths for each model. In the first-life phase, where capacity fade follows a predictable trend, the GPR model delivers the best accuracy, thanks to its empirical degradation framework. However, during the transition phase, the proposed PI-RNN exhibits a pronounced benefit over both the GPR model and the baseline RNN. In the first-life and second-life phases, the baseline RNN performs relatively close to the PI-RNN, with the latter holding only a slight advantage; nevertheless, the baseline RNN still outperforms the GPR model in the transition and second-life phases.

5.2.4. *Parametric study: impact of training set size on model performance*

One anticipated advantage of PIML models is their robustness in data-limited scenarios and their capacity for extrapolation. To evaluate

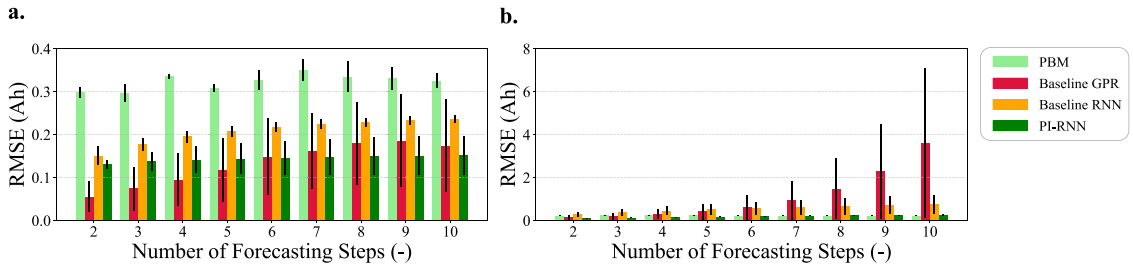

Fig. 16. Multi-step forecasting results of baseline and PI-RNN models for **a.** case study 1 and **b.** case study 2.

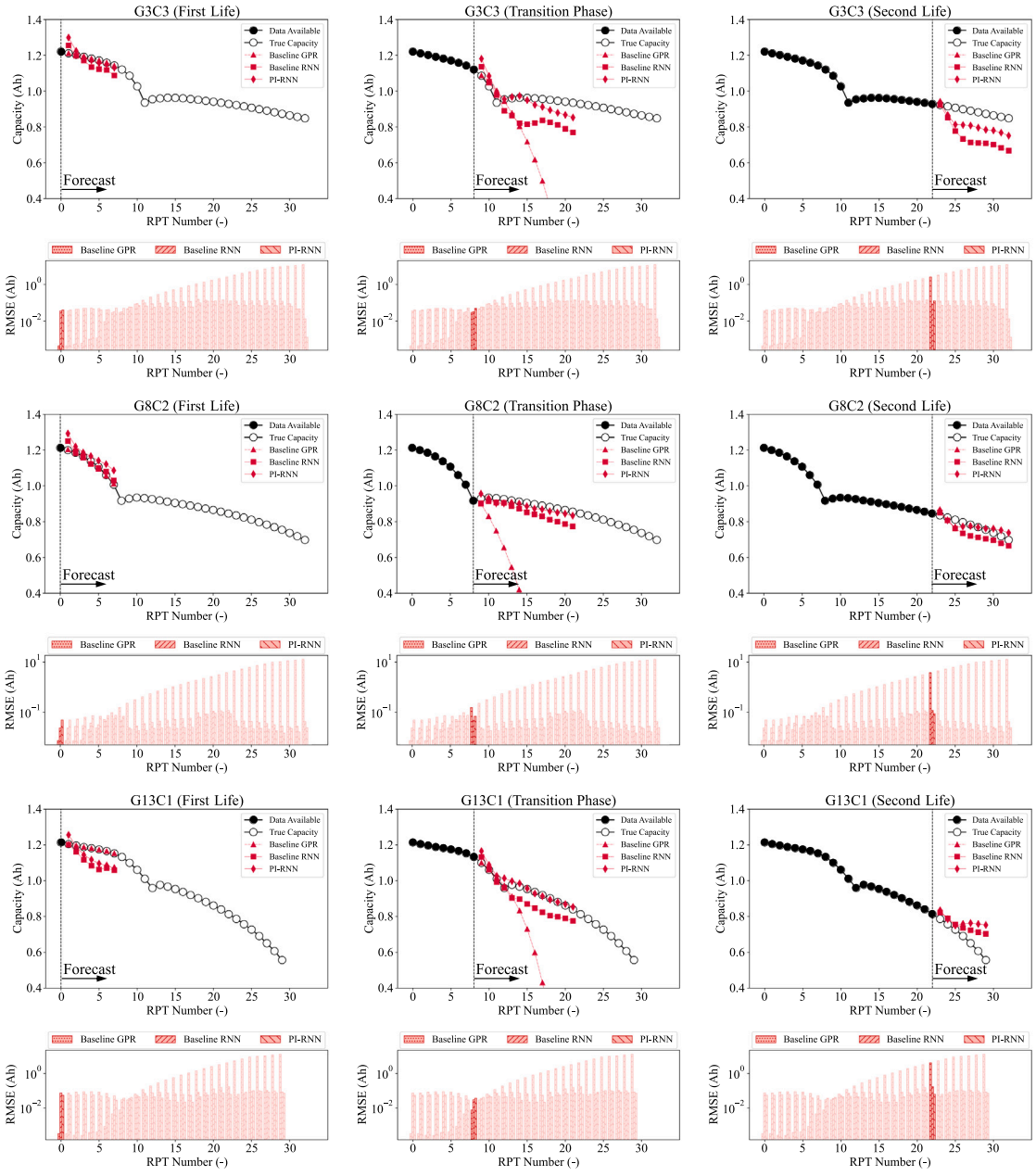

Fig. 17. Multi-step forecasting for test cells in groups G3, G8, and G13 of case study 2, illustrating capacity trajectory forecasts in different aging phases by the proposed PI-RNN and baseline models. The RMSE plot shows a 5-step forecasting error at each RPT. The GPR predictions extend beyond the visible range during the second-life phase.

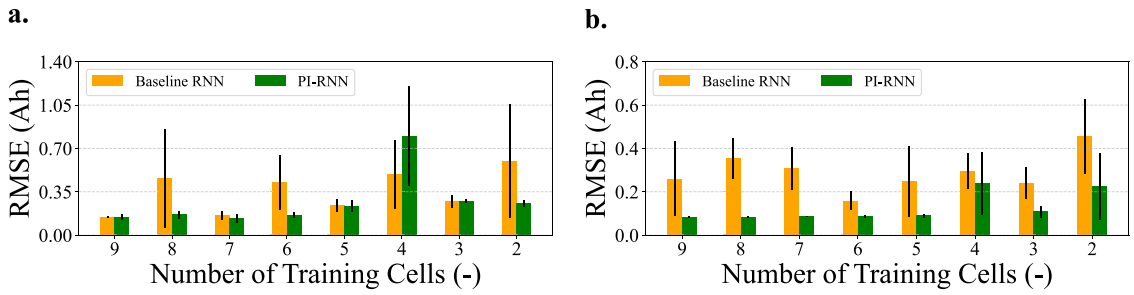

Fig. 18. Impact of the number of training cells on the performance of the proposed PI-RNN and the baseline models in **a.** case study 1 and **b.** case study 2.

this, we conducted a parametric study by progressively reducing the number of cells used for training in the original train-test split across each case study for single-step forecasting.

In both case studies, we conducted a parametric analysis in which we began with 9 training cells and systematically removed one cell at a time until only 1 remained. The test set remained fixed throughout (see Appendix C). All models were trained using the max-horizon approach with a 20-step forecasting horizon and evaluated on a 10-step horizon. As shown in Fig. 18, the RMSE does not consistently increase as the number of training cells decreases. This may be explained by the fact that individual cells differ in the quality of the information they provide; some cells contribute more to generalization even with fewer samples, while the removal of key cells can degrade performance due to reduced data diversity. Furthermore, because the dataset is relatively small, the RMSE trends are subject to higher variability. To address this variability, each training process with each set of selected cells was repeated over 3 trials and the mean RMSE along with the standard deviation are reported.

In case study 1, where both training and test cells were selected based on divisibility by 3, the proposed PI-RNN method exhibited more consistent performance as the training cells were reduced, maintaining stable accuracy down to 5 cells. When the training set was reduced to 4 cells, the PI-RNN performance became inferior to that of the baseline RNN; however, both models performed poorly, likely due to the influence of a noise-dominant cell. In case study 2, where models were trained on 9 cells from batch 2 and tested on 21 cells from batch 1, the PI-RNN demonstrated good performance up to 5 training cells, with errors increasing when fewer cells were used. In contrast, the baseline RNN showed fluctuating results and consistently higher RMSE compared to the PI-RNN.

To better understand the extent of out-of-distribution input–output coverage in the reduced training set scenarios, we visualized the capacity drop as a function of input features in the same manner as shown in Fig. 8. For case study 1, this corresponds to using one cell for training, and for case study 2, two cells for training, representing the final setups illustrated in Fig. 18. Similar to Fig. 8, Fig. 19 illustrates the input–output coverage for these reduced training sets. Compared to the initial configurations in Fig. 8, this visualization highlights a significant reduction in alignment between the training set (batch 2) and the test set (batch 1).

5.2.5. Impact of neural network architecture on model performance

To understand the role of the recurrent units and hidden states in the proposed PI-RNN method, we conducted an architecture substitution. Specifically, we replaced the PI-RNNs' recurrent units with a simple feedforward neural network of comparable structure, where each input is processed independently by a multi-layer perceptron as shown in Fig. 20. This modified model, termed "PI-NN", also integrates the surrogate model's output but removes the recurrent connections and hidden state mechanism of the original PI-RNNs.

By comparing the results of this non-recurrent architecture to those of PI-RNNs, we can assess the contribution of recurrence and hidden states to the model's performance on temporal dependencies. The comparison results for multi-step forecasting between a PI-NN and the proposed PI-RNN, with the same hidden layer size, are illustrated in Fig. 21, highlighting the performance differences across various forecasting steps. The proposed PI-RNN demonstrates an advantage across different numbers of forecasting steps. This trend suggests that the recurrent structure of PI-RNNs better captures temporal dependencies, leading to more robust performance over extended forecasting horizons.

5.3. Uncertainty quantification via Monte Carlo (MC) dropout

In this section, we focus on uncertainty quantification, an essential aspect of our prognostics approach. We adopted the MC dropout approach to estimate predictive uncertainty of the PI-RNN model, trained on the original train–test split in case study 2, for multi-step forecasting of a battery cell's remaining capacity. MC dropout is a technique that enables quantifying uncertainty in model predictions by keeping dropout layers active during inference, rather than only during training [55,75]. By running multiple forward passes on the same input, each with a different dropout mask, we introduce randomness into the network's predictions. This allows us to simulate an ensemble of models without needing to retrain, which is efficient for scenarios like ours where computational cost is a concern.

To quantify and refine forecast uncertainty, we implemented MC dropout combined with a recalibration step using isotonic regression. Specifically, we performed 40 forward passes with MC dropout enabled at each forecast step and then recalibrated the resulting predictions to produce a refined mean prediction (expected capacity) and standard deviation (model confidence). Fig. 22 illustrates the predicted degradation trajectories along with the uncertainty bounds for the same cells (from groups 3, 8, and 13) used in the trajectory forecasting results. Here, the error bars represent two standard deviations (95% prediction interval) around the recalibrated mean prediction, with recalibration ensuring that uncertainty grows monotonically over the forecast horizon. The figure highlights the predicted capacity and uncertainty across three key stages of the cell's life.

To assess the quality of uncertainty quantification, calibration curves were generated to compare observed and expected confidence levels. On a calibration curve, the observed confidence level is estimated as the fraction of true values that fall within prediction intervals constructed at the expected conference level [76]. This is achieved by computing a series of confidence levels, $c \in [0,1]$, and determining the corresponding two-sided bounds using the predicted means and standard deviations. The critical value z is obtained from the standard normal distribution as $z = \Phi^{-1}\left(\frac{1+c}{2}\right)$, where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution.

The observed proportion for each confidence level c is calculated as the fraction of true values falling within the corresponding prediction interval. The expected calibration error (ECE) quantifies the discrepancy between the observed confidence, $\hat{c}$, and the expected confidence [77], c, and is defined as:

$$\mathrm{ECE} = \frac{1}{N_{\mathrm{conf}}} \sum_{i=1}^{N_{\mathrm{conf}}} \left| \hat{c}^{(i)} - c^{(i)} \right|, \tag{19}$$

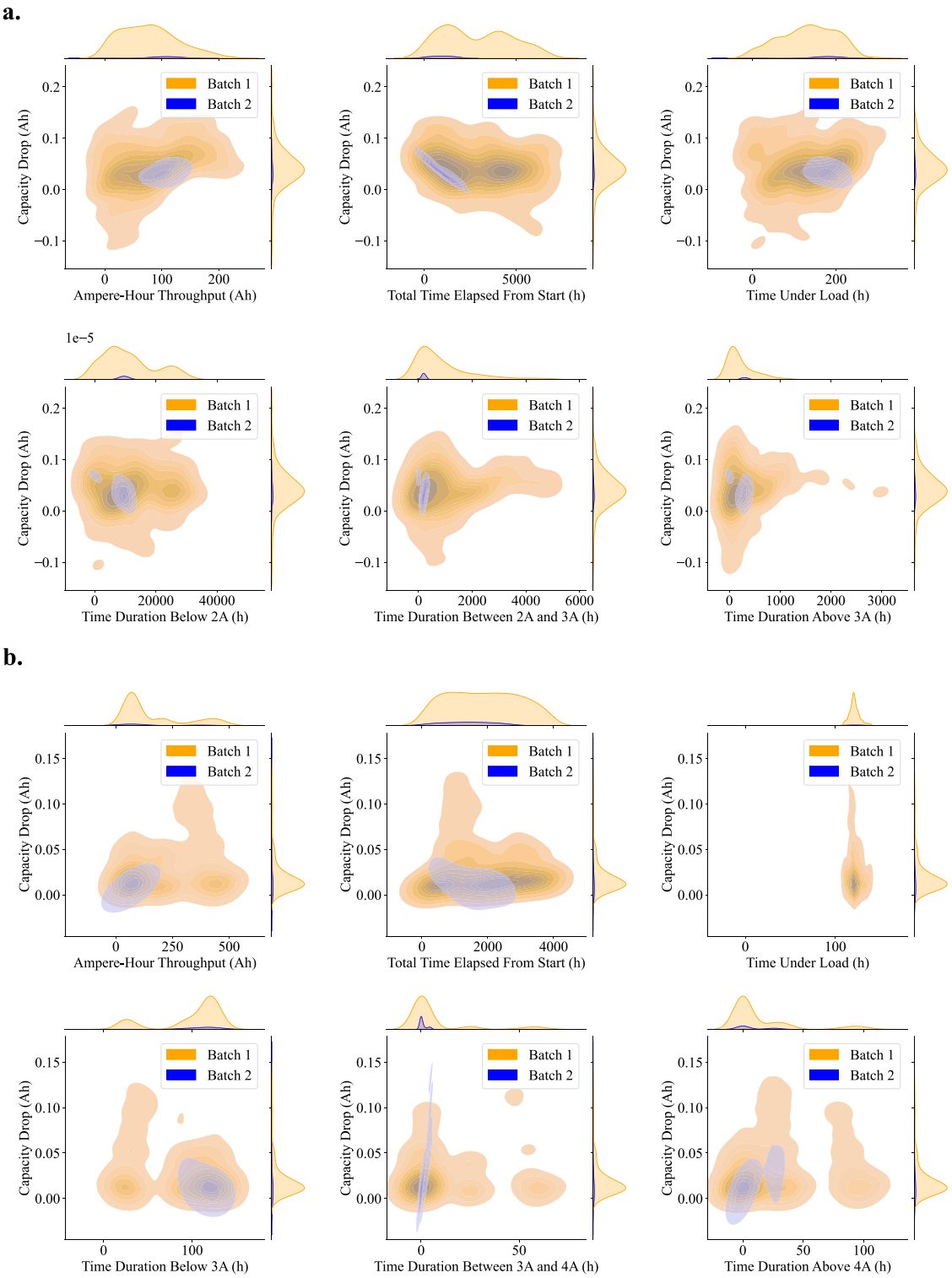

Fig. 19. Joint distribution plots showing the relationship between selected input features and the output capacity drop for each case study in reduced training set scenarios. For **a.** case study 1, one cell is used for training, and for **b.** case study 2, two cells are used.

where N_{conf} is the number of confidence levels. The calibration curves are visualized by plotting $\hat{c}$ against c, with the diagonal line $y = x$ representing perfect calibration. The ECE value is also reported to summarize the calibration quality, with lower ECE values indicating better uncertainty quantification. To address potential miscalibration in the model's predicted uncertainties, recalibration was performed using isotonic regression [77]. Isotonic regression is a non-parametric method that ensures the recalibrated predictions maintain a monotonic relationship between predicted means and true values. During recalibration, predicted means are adjusted to improve alignment with observed values, while predicted standard deviations are scaled to better reflect the actual spread of the data. A scaling factor is applied to the standard deviations to ensure that the recalibrated prediction intervals more accurately represent the true coverage probabilities. This recalibration process improves the reliability of the model's uncertainty

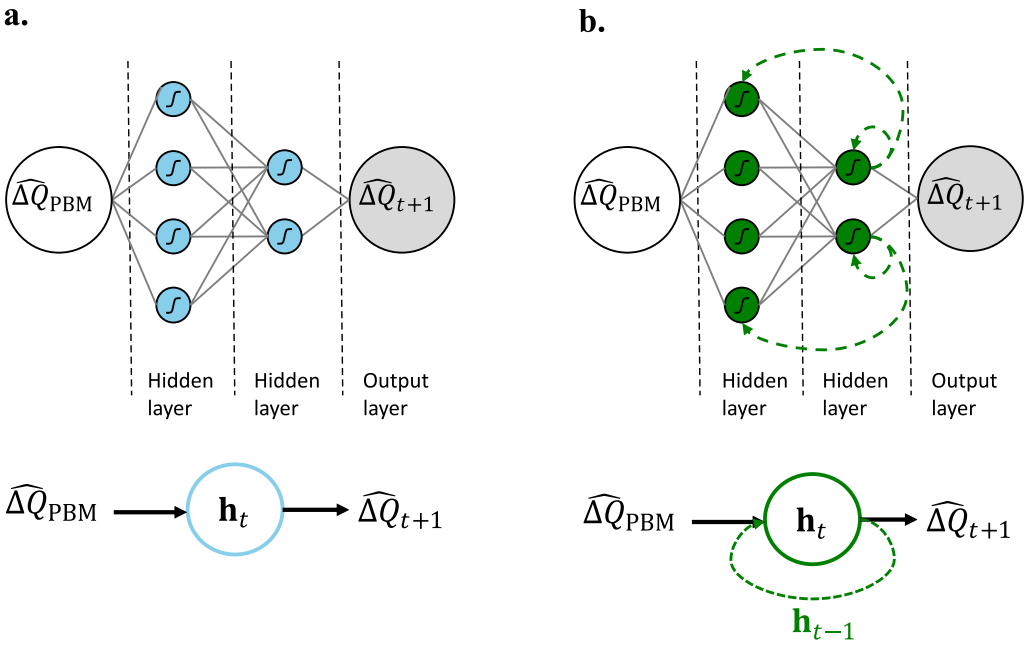

Fig. 20. Comparison of a PI-NN and a PI-RNN architecture. **a.** PI-NN processes data in a single direction, predicting capacity fade without accounting for temporal dependencies. **b.** PI-RNN incorporates feedback loops through hidden states h_t, enabling memory of past computations and improved sequential modeling for long-term predictions.

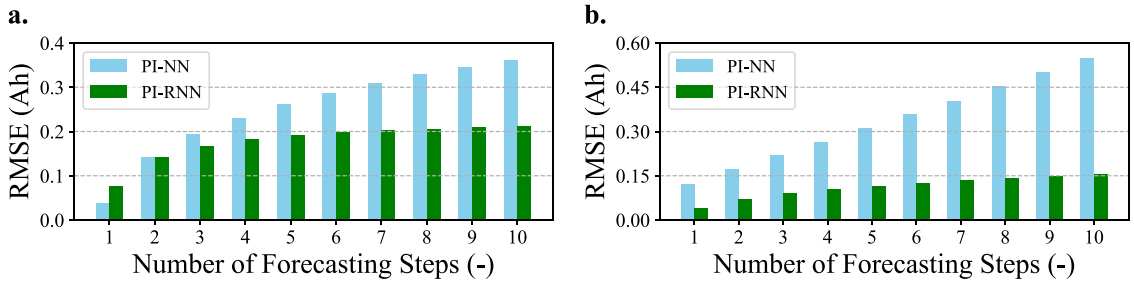

Fig. 21. Comparison of the proposed PI-RNN and a PI-NN, both with the same hidden layer size, to evaluate the impact of recurrent units on model performance in **a.** case study 1 and **b.** case study 2.

estimates and ensures that the prediction intervals correspond more closely to the observed data.

Fig. 23 presents calibration curves after applying recalibration for 1-step, 2-step, 4-step, 6-step, 8-step, and 10-step forecasts, shown for different numbers of MC dropout samples (5, 10, 20, and 40). The average ECE for each forecast horizon is reported as a measure of miscalibration. For recalibration, data from C2 cells from the training groups were used (6 cells in total). During recalibration, the predicted mean values were adjusted using isotonic regression to better align with the true values, and the predicted uncertainties were scaled by a factor of 0.45 to account for observed variability. The calibration curves for shorter horizons (1-step, 2-step) align closely with the diagonal, demonstrating well-calibrated uncertainty estimates with low ECE values. For intermediate horizons (4-step, 6-step), the curves show a mild upward deviation, reflecting slightly under-confident estimates. Longer horizons (8-step, 10-step) exhibit more pronounced deviations, with ECE values rising to ≈0.15, indicating the growing challenge of accurately quantifying uncertainty over extended forecasts. The longer forecast horizons introduce compounding errors due to the sequential dependency of predictions, which makes uncertainty quantification increasingly challenging. Overall, the results align with expectations, as uncertainty naturally increases with longer forecasting horizons due to the accumulation of errors and the sequential dependencies inherent in multi-step predictions. This is reflected in the higher ECE values and greater deviations observed in the calibration curves for extended horizons. The consistent behavior across different MC dropout sample sizes further validates the robustness of the recalibration process, demonstrating its effectiveness in adjusting uncertainty estimates to better reflect the inherent challenges of forecasting over multiple steps.

6. Conclusions

In this study, we introduced a novel hybrid approach that combines a physics-based P2D model with an RNN to create a PIML method. We termed the proposed PIML method "PI-RNN", representing a specific type of RNN that leverages both physics-based insights and data-driven learning to forecast capacity fade accurately across diverse usage conditions. We evaluated the proposed PI-RNN method on two datasets: an open-source LCO random walk dataset from NASA and an LFP second-life dataset collected for this study, where cells were initially cycled to 80% capacity before transitioning to a milder protocol for second-life operation. The PI-RNN proved most advantageous on the second-life dataset and in scenarios involving abrupt transitions in usage conditions from first-life to second-life, as well as in data-limited settings over extended forecasting horizons. In contrast, during the first-life phase of case study 2 and in case study 1, where the experimental variables were less complex and abrupt changes were absent, the PI-RNN performed comparably to data-driven approaches. When the training dataset is limited, the PI-RNN tends to outperform purely data-driven methods, whereas with ample training data, conventional data-driven approaches may be more suitable.

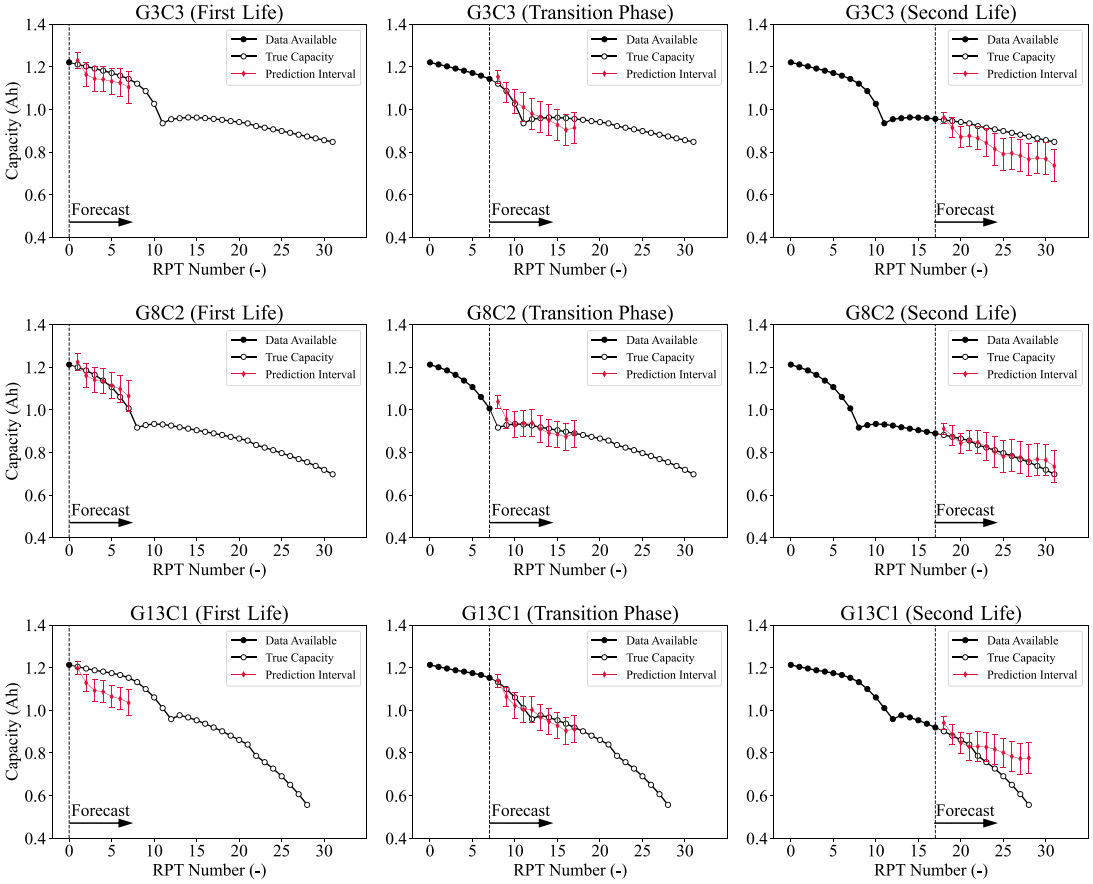

Fig. 22. Multi-step capacity degradation forecasts for test cells from groups 3, 8, and G13 in case study 2. Forecasts are presented for three distinct life stages: early first-life phase (left), transition phase (middle), and second-life phase (right). The plots display the recalibrated predicted degradation trajectories and corresponding uncertainty bounds (95% prediction interval), obtained from 40 MC dropout forward passes and isotonic regression calibration.

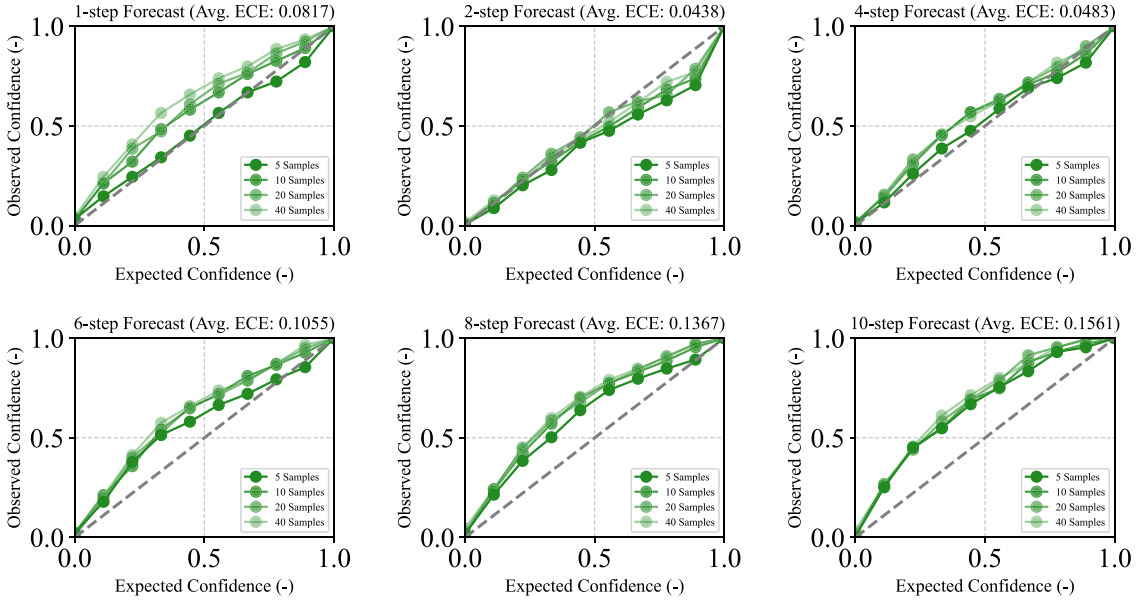

Fig. 23. Calibration curves for uncertainty quantification of the PI-RNN model trained using the original train–test split in case study 2 across different forecast horizons.

Nomenclature

Symbols

Symbol	Definition
α_a	Anodic transfer coefficient.
α_c	Cathodic transfer coefficient.
a_k	Specific interfacial area for electrode k.
c_e	Lithium concentration in the electrolyte.
$c_{e,0}$	Initial lithium concentration in the electrolyte.
$c_{s,k}^{\max}$	Maximum lithium concentration in electrode particles for electrode k.
$c_{s,k,0}$	Initial lithium concentration in electrode particles for electrode k.
D_e	Diffusion coefficient in the electrolyte.
$D_{s,k}$	Lithium diffusion coefficient in electrode particles for electrode k.
F	Faraday's constant.
$\boldsymbol{\theta}$	Model parameters.
$\boldsymbol{\theta}_{\text{emp}}$	Parameters of the empirical capacity fade model.
$\boldsymbol{\theta}_{\text{ker}}$	Kernel parameter vector for the GPR model, $[\sigma_s^2, \ell]$.
$\mathbf{h}_t$	Hidden state at time t.
$\mathbf{h}_{\text{pbm}}$	Physics-based model contribution term to the hidden state update.
$\mathbf{I}$	Battery current measurements.
i_{app}	Applied current.
J_k	Reaction current density for electrode k.
$k(t, t'; \boldsymbol{\theta}_{\text{ker}})$	Covariance function in the GPR model describing similarity between times t and t'.
k_0	Reaction rate constant.
$\mathbf{L}_t$	Load pattern applied between RPT $t-1$ and t.
L_k	Thickness of electrode k.
ℓ	Length scale in the RBF kernel of the GPR model.
M	Forecasting horizon.
$n_{\mathbf{x}}$	Number of selected features (size of the input vector $\mathbf{x}$).
$\mathbf{x}_t$	Features extracted from the battery current load pattern at time t.
N_{train}	Number of training samples.
N_{test}	Number of test samples.
$N_{\text{meas}}^{\text{load}}$	Number of measurement time steps in each load pattern.
$N_{\text{per}}^{\text{load}}$	Number of charge and discharge load periods in each load pattern.
I_{high}	Predefined higher current threshold.
I_{low}	Predefined lower current threshold.
$\mathbf{W}_{\mathbf{x}}$	Weight matrix associated with the input component.
$\mathbf{W}_{\text{h}}$	Weight matrix associated with the hidden state component.
$\mathbf{W}_{\text{pbm}}$	Weight matrix for the PBM contribution.
$\Delta Q_{\text{pbm}}(\mathbf{x}_t)$	Capacity drop predicted by the PBM based on the input at time t.
Q_{thr}	Ampere-hour throughput.
Q_t	Capacity measured at RPT t.
τ	Time steps within each load pattern (for summation or integration).
τ_{dchg}	Low C-rate discharge step for capacity calculation.
T	Temperature.
$T_{\mathbf{L},t}$	Duration of load pattern $\mathbf{L}$ at time t or time between consecutive RPTs.
T_{total}	Total elapsed time up to RPT t.
t	RPT cycle index in the experiment timeline.
t_+^0	Transference number of the electrolyte.
$U(c_{s,k})$	Open circuit potential as a function of lithium concentration in electrode k.
V	Terminal voltage.
r	Radial coordinate in electrode particles.
r_k	Particle radius in electrode k.
σ_k	Electronic conductivity of electrode k.
σ_s^2	Signal variance in the RBF kernel of the GPR model.
κ_e	Ionic conductivity of the electrolyte.
R	Universal gas constant.
y_t	Battery capacity at RPT t.
$\hat{y}_t$	Forecasted capacity at RPT t.

Another novel aspect of this work is the incorporation of future load patterns as input features, which enables the proposed method to more effectively adapt to forthcoming operating conditions. This approach shows promise in capturing complex degradation trends across a variety of usage profiles. For future research, rather than relying solely on experimental cycling protocols to generate load patterns, one could simulate a broader spectrum of conditions and use the resulting data to train a physics-based surrogate model. Such an approach would permit

Abbreviations

Abbreviation	Definition
CC	Constant-Current
CCCV	Constant-Current Constant-Voltage
CNN	Convolutional Neural Network
DOD	Depth of Discharge
ECM	Equivalent Circuit Model
EOL	End-of-Life
GPR	Gaussian Process Regression
HPPC	Hybrid Pulse Power Characterization
LCO	Lithium Cobalt Oxide
LFP	Lithium Iron Phosphate
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Square Error
MC	Monte Carlo
MNC	Nickel Manganese Cobalt
OOD	Out-of-Distribution
PBM	Physics-Based Model
PI-NN	Physics-Informed (feedforward) Neural Network
PI-RNN	Physics-Informed Recurrent Neural Network
PIML	Physics-Informed Machine Learning
RBF	Radial Basis Function
RNN	Recurrent Neural Network
RPT	Reference Performance Test
RMSE	Root Mean Square Error
SEI	Solid Electrolyte Interface
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machines
RVM	Relevance Vector Machines
SOC	State of Charge
SOH	State of Health
EMF	Explicit Mean Function

more randomized testing that mirrors real-world vehicle or energy storage operations, though it would introduce additional challenges in terms of computational cost, calibration, and numerical stability. While this strategy was computationally prohibitive for the current study, it has the potential to further enhance the robustness and flexibility of the PI-RNN methodology. Moreover, we envision extending this framework to field data, which might be facilitated by leveraging extra computational resources and exploring more sophisticated RNN architectures to improve adaptability to real-world battery usage patterns. In addition, future extensions of the PBM could incorporate additional degradation mechanisms, such as explicit modeling of solvent consumption, to improve the PBM's representational accuracy and further strengthen the hybrid framework's predictive capabilities.

CRediT authorship contribution statement

Sina Navidi: Writing – original draft, Validation, Software, Methodology, Conceptualization. **Kristupas Bajarunas:** Writing – review & editing, Methodology. **Manuel Arias Chao:** Writing – review & editing, Supervision, Methodology. **Chao Hu:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This work was supported in part by the US National Science Foundation under Grant ECCS-2015710 and in part by a MathWorks Development Collaborative Research Grant (DCRG), USA. The authors would also like to thank Tingkai Li and Aidan Lawlor at the University of Connecticut for collecting the LFP second-life dataset used in this study. Additionally, we express our gratitude to Benjamin Nowacki and Tingkai Li for their assistance in proofreading the manuscript. Any opinions, findings, or conclusions in this paper are those of the authors and do not necessarily reflect the sponsors' views.

Appendix A. Experimental protocols

See Tables A.4 and A.5.

Appendix B. P2D model

See Fig. B.24 and Table B.6

Appendix C. Train-test split

See Tables C.7 and C.8.

Table A.4
Experimental NASA (LCO random walk) dataset cell groups and protocols.

Group	Charge protocol	Discharge protocol	RPT period	Ambient temperature	Details
1	Charge to 4.2 V with random duration (0.5 h–3 h)	Discharge to 3.2 V with random currents (0.5 A–4 A)	50 cycles	25 °C	–
2	Charge to 4.2 V with 2 A	Discharge to 3.2 V with random currents (0.5 A–4 A)	50 cycles	25 °C	–
3	Charge to 4.2 V with 2 A	Discharge to 3.2 V with random currents (0.5 A–4 A)	50 cycles	40 °C	Probability distribution skewed towards higher currents, new setpoint every 1 min during discharging
4	Charge to 4.2 V with 2 A	Discharge to 3.2 V with random currents (0.5 A–4 A)	50 cycles	25 °C	Probability distribution skewed towards higher currents, new setpoint every 1 min during discharging
5	Charge to 4.2 V with 2 A	Discharge to 3.2 V with random currents (0.5 A–4 A)	50 cycles	40 °C	Probability distribution skewed towards lower currents, new setpoint every 1 min during discharging
6	Charge to 4.2 V with 2 A	Discharge to 3.2 V with random currents (0.5 A–4 A)	50 cycles	25 °C	Probability distribution skewed towards lower currents, new setpoint every 1 min during discharging
7	Random charging and discharging currents (−4.5 A to 4.5 A)	Random charging and discharging currents (−4.5 A to 4.5 A)	1500 cycles (5 days)	25 °C	–

Table A.5
Cycling and RPT conditions for LFP second-life dataset at 25 °C with an RPT interval of 5 days.

Group	Charge protocol	Discharge protocol	Details
1	Charge to 3.6 V with 6.647 A	Discharge to 97% DOD with 2.798 A	One-step fast-charging without derating
2	Charge to 3.6 V with 6.884 A	Discharge to 75.3% DOD with 2.992 A	One-step fast-charging without derating
3	Charge to 3.6 V with 4.897 A	Discharge to 92.5% DOD with 4.336 A	One-step fast-charging without derating
4	Charge to 3.6 V with 6.145 A	Discharge to 83.7% DOD with 4.158 A	One-step fast-charging without derating
5	Charge to 3.6 V with 5.722 A	Discharge to 79.6% DOD with 3.641 A	One-step fast-charging without derating
6	Charge to 3.6 V with 5.099 A	Discharge to 95.7% DOD with 2.543 A and 7.1% derating	One-step fast-charging with SOH-based derating
7	Charge to 3.6 V with 5.810 A	Discharge to 84.7% DOD with 3.802 A and 5.5% deratin	One-step fast-charging with SOH-based derating
8	Charge to 3.6 V with 5.494 A	Discharge to 91.7% DOD with 4.090 A and 8.9% derating	One-step fast-charging with SOH-based derating
9	Charge to 3.6 V with 7.024 A	Discharge to 79.2% DOD with 3.314 A and 9.4% derating	One-step fast-charging with SOH-based derating
10	Charge to 3.6 V with 6.367 A	Discharge to 71.9% DOD with 4.680 A and 6.3% derating	One-step fast-charging with SOH-based derating
11	Charge to 3.6 V with 6.278 A	Discharge to 84.7% DOD with 3.440 A and switching SOC of 86.8%	Two-step fast-charging
12	Charge to 3.6 V with 5.294 A	Discharge to 95.3% DOD with 3.062 A and switching SOC of 79.7%	Two-step fast-charging
13	Charge to 3.6 V with 4.887 A	Discharge to 80.0% DOD with 4.766 A and switching SOC of 82.6%	Two-step fast-charging
14	Charge to 3.6 V with 6.917 A	Discharge to 90.8% DOD with 2.530 A and switching SOC of 71.1%	Two-step fast-charging
15	Charge to 3.6 V with 5.798 A	Discharge to 74.7% DOD with 3.990 A and switching SOC of 77.7%	Two-step fast-charging
16	Charge to 3.6 V with 5.099 A	Discharge to 95.7% DOD with 2.543 A	One-step fast-charging without derating
17	Charge to 3.6 V with 5.810 A	Discharge to 84.7% DOD with 3.802 A	One-step fast-charging without derating
18	Charge to 3.6 V with 4.578 A	Discharge to 91.7% DOD with 4.090 A	One-step fast-charging without derating

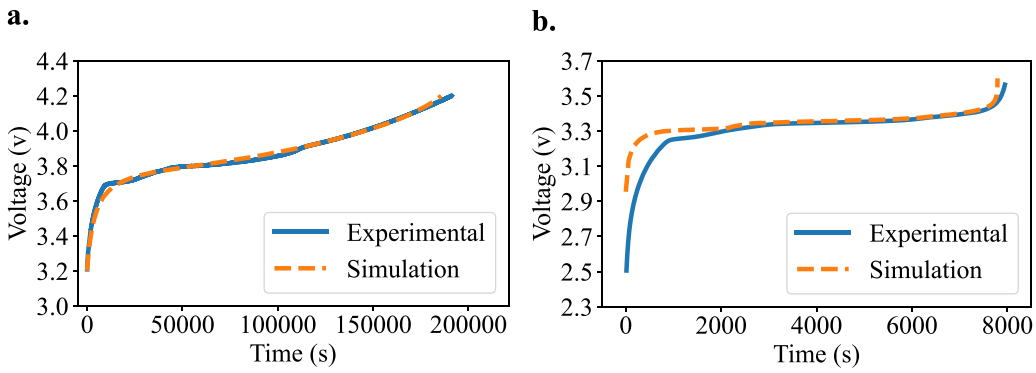

Fig. B.24. Fitted simulated voltage response to experimental voltage response at a low C-rate (0.04 A) for **a.** case study 1 and **b.** case study 2. The open circuit potential (OCP) values for each electrode were taken from Ramadass et al. [16], Prada et al. [78], and O'Kane et al. [68].

Table B.6
Parameter values for the P2D model adjusted to case study 1 and case study 2.

Parameter	Symbol	Unit	Case study 1	Case study 2
Positive electrode thickness	L_p	m	10.5×10^{-5} [b]	5.8×10^{-5} [a]
Negative electrode thickness	L_n	m	13×10^{-5} [b]	3.6×10^{-5} [a]
Separator thickness	L_{sep}	m	1.2×10^{-5} [b]	1.9×10^{-5} [a]
Electrode height	H	m	0.065 [b]	0.055 [a]
Electrode width	W	m	1.58 [b]	0.85 [a]
Positive particle radius	r_p	m	2×10^{-6} [16]	5.2×10^{-6}
Negative particle radius	r_n	m	2×10^{-6} [16]	5.8×10^{-6}
Solid phase volume fraction (positive)	$\varepsilon_{s,p}$	–	0.68	0.54
Solid phase volume fraction (negative)	$\varepsilon_{s,n}$	–	0.70	0.75
Separator porosity	ε_{sep}	–	0.50 [16]	0.45 [78]
Initial salt concentration in solution	c_e^0	mol m^{-3}	1000	1200 [78]
Initial concentration in positive electrode	$c_{s,p}^0$	mol m^{-3}	20 440	29 440
Initial concentration in negative electrode	$c_{s,n}^0$	mol m^{-3}	17 500	20 100
Maximum concentration in positive electrode	$c_{s,p}^{max}$	mol m^{-3}	23 260	30 370
Maximum concentration in negative electrode	$c_{s,n}^{max}$	mol m^{-3}	27 850	45 850
Solid phase diffusivity (positive)	$D_{s,p}$	m^2 s^{-1}	5×10^{-15}	6×10^{-15}
Solid phase diffusivity (negative)	$D_{s,n}$	m^2 s^{-1}	5×10^{-14}	5×10^{-14}
Electrolyte ionic diffusivity	D_e	m^2 s^{-1}	2×10^{-10}	2×10^{-10}
Bruggeman exponent	β	–	1.5	1.5
Positive electrode charge transfer coefficient	α_p	–	0.5	0.5
Positive electrode conductivity	σ_p	S m^{-1}	100	0.34
Positive electrode porosity	ε_p	–	0.385 [16]	0.43
Reference temperature	T_{ref}	K	298	298
Negative electrode charge transfer coefficient	α_n	–	0.5	0.5
Negative electrode conductivity	σ_n	S m^{-1}	100	215
Negative electrode porosity	ε_n	–	0.48	0.36
Lithium plating kinetic rate constant	$k_{Li,plating}$	m s^{-1}	1×10^{-9}	1×10^{-9}
Lithium plating transfer coefficient	$\alpha_{Li,plating}$	–	0.65	0.65
Dead lithium decay constant	$k_{Li,decay}$	s^{-1}	1×10^{-6}	1×10^{-6}
Electrode LAM constant exponential term	n_{LAM}	–	2.0	2.0
Electrode LAM constant proportional term	k_{LAM}	s^{-1}	2.7778×10^{-7}	2.7778×10^{-7}
SEI growth activation energy	E_{SEI}	J mol^{-1}	38 000.0	38 000.0
SEI kinetic rate constant	k_{SEI}	m s^{-1}	1×10^{-12}	1×10^{-12}
SEI reaction exchange current density	$i_{0,SEI}$	A m^{-2}	1.5×10^{-7}	1.5×10^{-7}
SEI resistivity	ρ_{SEI}	Ω m	2×10^5	2×10^5

[a] Indicates measured values.

[b] Indicates assumed values.

Table C.7
Original train-test split of case study 1 (LCO random-walk dataset).

Batch	Cells	Dataset	Number of cells
Batch 1	RW1, RW2, RW4, RW5, RW7, RW8, RW10, RW13, RW14, RW16, RW17, RW19, RW20, RW23, RW25, RW26, RW28	Testing	17
Batch 2	RW3, RW6, RW9, RW11, RW12, RW15, RW18, RW21, RW22, RW24, RW27	Training and validation	11

Table C.8
Original train-test split of case study 2 (LFP second-life dataset).

Batch	Groups	Cells	Dataset	Number of cells
Batch 1	G3, G5, G6, G7, G8, G13, G15	C1, C2, C3	Testing	21
Batch 2	G1, G2, G4, G16, G17, G18	C1, C3	Training and validation	12
Batch 2	G1, G2, G4, G16, G17, G18	C2	Calibration	6

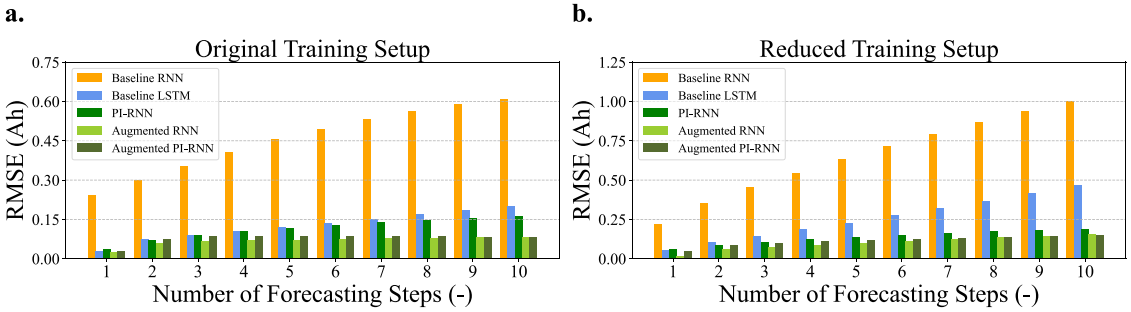

Fig. D.25. Comparison of forecasting RMSEs by LSTM, baseline and augmented RNN, PI-RNN, and augmented PI-RNN models using **a.** original training setup and **b.** reduced training setup.

Appendix D. Benchmarking against LSTM and data augmentation

We conducted a benchmarking analysis to compare the proposed PI-RNN method with state-of-the-art methods. In particular, we compared PI-RNN against LSTM and an alternative PIML approach based on observational bias through data augmentation. The evaluation was carried out under two training setups as described in Section 5.2.4.

- **Original training setup:** Using the full training set, LSTM performed very well, almost on par with PI-RNN and outperforming baseline RNN.
- **Reduced training setup:** When the training set was reduced to only two cells, the performance of LSTM declined relative to PI-RNN, especially in longer forecasting horizons.

These results suggest that while advanced methods like LSTM can perform comparably when more data and future loading conditions are available, PI-RNN consistently delivers more robust long-term forecasts, especially under data-limited scenarios.

For completeness, we also added another PIML approach based on observational bias via data augmentation. This augmented RNN is trained on both experimental and simulated data from the PBM, providing a straightforward means to integrate observational bias. The results indicate that for long-term forecasting in data-constrained scenarios, leveraging both inductive and observational biases, as in the augmented PI-RNN method, is particularly beneficial. Fig. D.25 displays the RMSE values over forecasting horizons from 1 to 10 steps for the compared models.

Appendix E. Hyperparameter selection

Hyperparameter optimization using Optuna [71] was conducted on the baseline RNN model to verify the suitability of the selected hyperparameters (see Table 3). Optimal values obtained from this analysis (summarized in Table E.9) support our original parameter selection.

Table E.9
Summary of Optuna hyperparameter optimization settings and optimal results for baseline RNN model.

Optuna Hyperparameter Settings	Value
Sampler	TPE (Tree-structured Parzen Estimator)
Direction	Minimize validation loss (MSE)
Trials	50
Epochs	2000
Early stopping patience	50
Hidden size range	[10, 60], step=5
Learning rate range	$[10^{-4}, 10^{-2}]$
Optimal hyperparameters	
Case study 1: Hidden size	45
Case study 1: Learning rate	0.00726
Case study 2: Hidden size	50
Case study 2: Learning rate	0.00038

Appendix F. GPR baseline model results on LCO random walk dataset (case study 1)

The trajectory forecasting results for GPR baseline model as the best-performing model in case study 1 is shown in Fig. F.26. The degradation trend in this case study remains relatively consistent throughout each cell's lifetime, allowing the GPR baseline to capture the underlying trend. By fitting an empirical model to the initial portion of the capacity fade and then applying GPR to the residuals, the forecast remains close to the true capacity data for multiple steps ahead, even with a limited amount of training points. This illustrates that when usage conditions are relatively stable and degradation trends are consistent, the GPR baseline offers reliable long-term predictions with comparatively low forecasting errors.

Data and code availability

The processed LFP second-life data and the code supporting the findings of this study are publicly available at: https://github.com/REIL-UConn/Forecasting-Battery-Capacity-for-Second-Life-Applications-Using-PI-RNN. A link to the raw data of the LFP second-life dataset, including the first two batches used in this study and additional batches to be released, will be added when available.

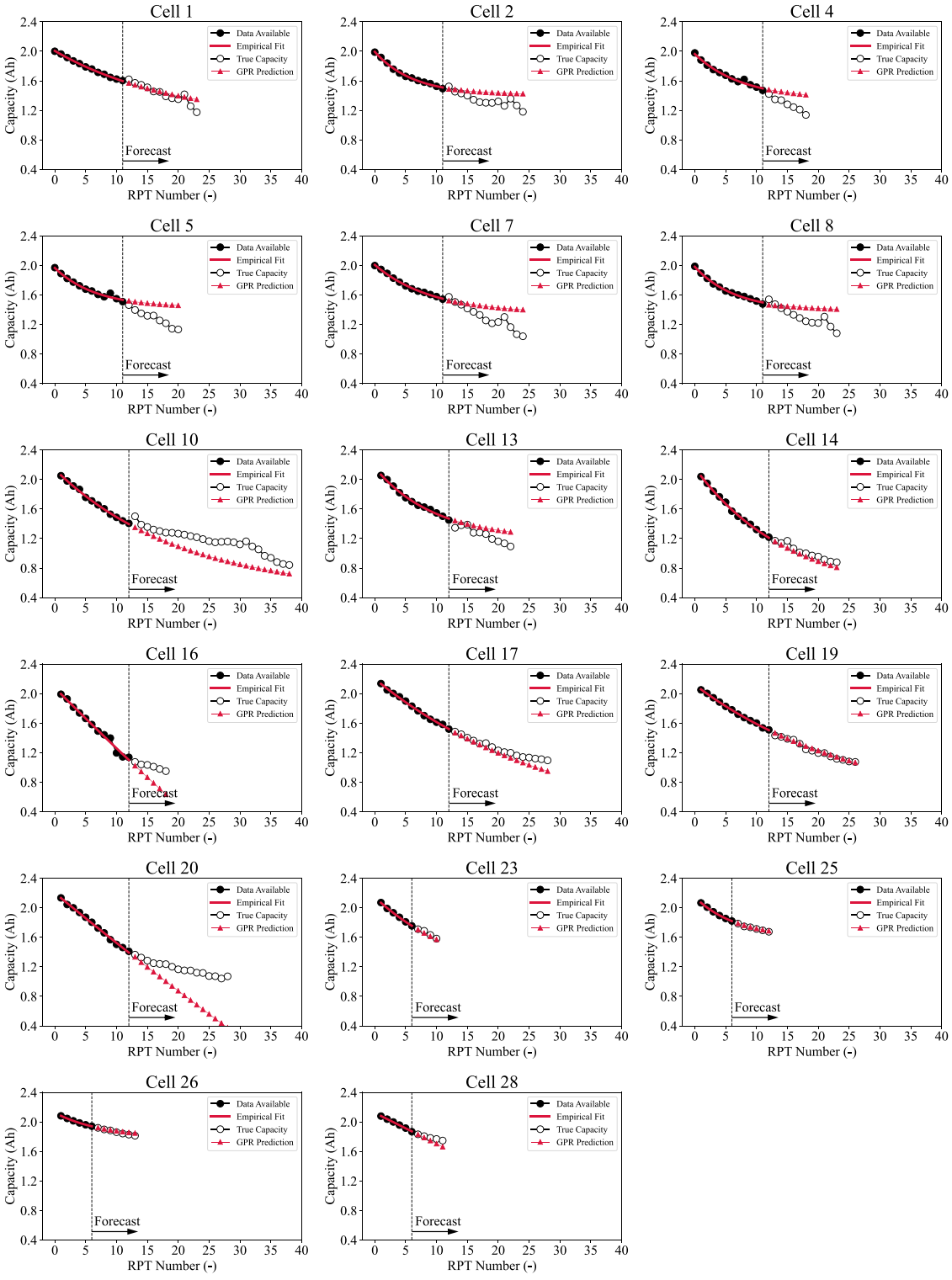

Fig. F.26. Capacity trajectories for case study 1 forecasted using GPR baseline model.

References

[1] Horiba T, Maeshima T, Matsumura T, Koseki M, Arai J, Muranaka Y. Applications of high power density lithium ion batteries. J Power Sources 2005;146(1–2):107–10.

[2] Qian C, Xu B, Chang L, Sun B, Feng Q, Yang D, et al. Convolutional neural network based capacity estimation using random segments of the charging curves for lithium-ion batteries. Energy 2021;227:120333.

[3] Hu X, Xu L, Lin X, Pecht M. Battery lifetime prognostics. Joule 2020;4(2):310–46.

[4] Li X, Zhang L, Wang Z, Dong P. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and

elman neural networks. J Energy Storage 2019;21:510–8.
[5] Ungurean L, Cârstoiu G, Micea MV, Groza V. Battery state of health estimation: a structured review of models, methods and commercial devices. Int J Energy Res 2017;41(2):151–81.
[6] Deng Z, Yang L, Cai Y, Deng H, Sun L. Online available capacity prediction and state of charge estimation based on advanced data-driven algorithms for lithium iron phosphate battery. Energy 2016;112:469–80.
[7] Zhuang D, Li ML, Lam VN, Braatz RD, Chueh WC, Bazant MZ. Physics-informed design of hybrid pulse power characterization tests for rechargeable batteries. J Electrochem Soc 2024;171(5):050510.
[8] Zhang C, Allafi W, Dinh Q, Ascencio P, Marco J. Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique. Energy 2018;142:678–88.
[9] Wang F, Zhai Z, Zhao Z, Di Y, Chen X. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. Nat Commun 2024;15(1):4332.
[10] Sun Y, Tian H, Hu F, Du J. Method for evaluating degradation of battery capacity based on partial charging segments for multi-type batteries. Batteries 2024;10(6):187.
[11] Liu H, Deng Z, Yang Y, Lu C, Li B, Liu C, Cheng D. Capacity evaluation and degradation analysis of lithium-ion battery packs for on-road electric vehicles. J Energy Storage 2023;65:107270.
[12] Thelen A, Huan X, Paulson N, Onori S, Hu Z, Hu C. Probabilistic machine learning for battery health diagnostics and prognostics—review and perspectives. Npj Mater Sustain 2024;2(1):14.
[13] Harper G, Sommerville R, Kendrick E, Driscoll L, Slater P, Stolkin R, et al. Recycling lithium-ion batteries from electric vehicles. Nature 2019;575(7781):75–86.
[14] Sulzer V, Mohtat P, Aitio A, Lee S, Yeh YT, Steinbacher F, et al. The challenge and opportunity of battery lifetime prediction from field data. Joule 2021;5(8):1934–55.
[15] Navidi S, Thelen A, Li T, Hu C. Physics-informed machine learning for battery degradation diagnostics: A comparison of state-of-the-art methods. Energy Storage Mater 2024;68:103343.
[16] Ramadass P, Haran B, Gomadam PM, White R, Popov BN. Development of first principles capacity fade model for li-ion cells. J Electrochem Soc 2004;151(2):A196.
[17] Li J, Wang L, Lyu C, Wang H, Liu X. New method for parameter estimation of an electrochemical-thermal coupling model for LiCoO2 battery. J Power Sources 2016;307:220–30.
[18] Guha A, Patra A. Online estimation of the electrochemical impedance spectrum and remaining useful life of lithium-ion batteries. IEEE Trans Instrum Meas 2018;67(8):1836–49.
[19] Chang Y, Fang H, Zhang Y. A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. Appl Energy 2017;206:1564–78.
[20] He W, Williard N, Osterman M, Pecht M. Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. J Power Sources 2011;196(23):10314–21.
[21] Randall AV, Perkins RD, Zhang X, Plett GL. Controls oriented reduced order modeling of solid-electrolyte interphase layer growth. J Power Sources 2012;209:282–8.
[22] Han X, Lu L, Zheng Y, Feng X, Li Z, Li J, Ouyang M. A review on the key issues of the lithium ion battery degradation among the whole life cycle. ETransportation 2019;1:100005.
[23] Galavotti L, Sebastiani F, Simi M. Experiments on the use of feature selection and negative evidence in automated text categorization. In: Research and advanced technology for digital libraries: 4th European conference, ECDL 2000 Lisbon, Portugal, September 18–20, 2000 proceedings 4. Springer; 2000, p. 59–68.
[24] El Mejdoubi A, Chaoui H, Gualous H, Van Den Bossche P, Omar N, Van Mierlo J. Lithium-ion batteries health prognosis considering aging conditions. IEEE Trans Power Electron 2018;34(7):6834–44.
[25] Xing Y, Ma EW, Tsui K-L, Pecht M. A case study on battery life prediction using particle filtering. In: Proceedings of the IEEE 2012 prognostics and system health management conference (PHM-2012 Beijing). IEEE; 2012, p. 1–6.
[26] Su X, Wang S, Pecht M, Zhao L, Ye Z. Interacting multiple model particle filter for prognostics of lithium-ion batteries. Microelectron Reliab 2017;70:59–69.
[27] Li Y, Liu K, Foley AM, Zülke A, Berecibar M, Nanini-Maury E, et al. Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review. Renew Sustain Energy Rev 2019;113:109254.
[28] Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. Nat Energy 2019;4(5):383–91.
[29] Zhao Q, Qin X, Zhao H, Feng W. A novel prediction method based on the support vector regression for the remaining useful life of lithium-ion batteries. Microelectron Reliab 2018;85:99–108.
[30] Patil MA, Tagade P, Hariharan KS, Kolake SM, Song T, Yeo T, et al. A novel multistage support vector machine based approach for li ion battery remaining useful life estimation. Appl Energy 2015;159:285–97.
[31] Chen Z, Shi N, Ji Y, Niu M, Wang Y. Lithium-ion batteries remaining useful life prediction based on BLS-RVM. Energy 2021;234:121269.
[32] Richardson RR, Osborne MA, Howey DA. Gaussian process regression for forecasting battery state of health. J Power Sources 2017;357:209–19.
[33] Richardson RR, Osborne MA, Howey DA. Battery health prediction under generalized conditions using a Gaussian process transition model. J Energy Storage 2019;23:320–8.
[34] Zhang Y, Xiong R, He H, Pecht MG. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. IEEE Trans Veh Technol 2018;67(7):5695–705.
[35] Lu J, Xiong R, Tian J, Wang C, Hsu C-W, Tsou N-T, et al. Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning. Energy Storage Mater 2022;50:139–51.
[36] Tian J, Xiong R, Shen W, Lu J, Yang X-G. Deep neural network battery charging curve prediction using 30 points collected in 10 min. Joule 2021;5(6):1521–34.
[37] Biggio L, Bendinelli T, Kulkarni C, Fink O. Dynaformer: A deep learning model for ageing-aware battery discharge prediction. 2022, arXiv preprint arXiv:2206.02555.
[38] Gomez W, Wang F-K, Chou J-H. Li-ion battery capacity prediction using improved temporal fusion transformer model. Energy 2024;296:131114.
[39] Cheng S, Azarian MH, Pecht MG. Sensor systems for prognostics and health management. Sensors 2010;10(6):5774–97.
[40] Chen Z, Liu Y, Sun H. Physics-informed learning of governing equations from scarce data. Nat Commun 2021;12(1):6136.
[41] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nat Rev Phys 2021;3(6):422–40.
[42] Aykol M, Gopal CB, Anapolsky A, Herring PK, van Vlijmen B, Berliner MD, et al. Perspective—combining physics and machine learning to predict battery lifetime. J Electrochem Soc 2021;168(3):030525.
[43] Hofmann T, Hamar J, Rogge M, Zoerr C, Erhard S, Schmidt JP. Physics-informed neural networks for state of health estimation in lithium-ion batteries. J Electrochem Soc 2023;170(9):090524.
[44] Shi J, Rivera A, Wu D. Battery health management using physics-informed machine learning: Online degradation modeling and remaining useful life prediction. Mech Syst Signal Process 2022;179:109347.
[45] Kohtz S, Xu Y, Zheng Z, Wang P. Physics-informed machine learning model for battery state of health prognostics using partial charging segments. Mech Syst Signal Process 2022;172:109002.
[46] Schmitt J, Horstkötter I, Bäker B. Electrical lithium-ion battery models based on recurrent neural networks: A holistic approach. J Energy Storage 2023;58:106461.
[47] You G-w, Park S, Oh D. Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach. Appl Energy 2016;176:92–103.
[48] Bajarunas K, Baptista ML, Goebel K, Chao MA. Health index estimation through integration of general knowledge with unsupervised learning. Reliab Eng Syst Saf 2024;251:110352.
[49] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
[50] Li C, Xiao F, Fan Y. An approach to state of charge estimation of lithium-ion batteries based on recurrent neural networks with gated recurrent unit. Energies 2019;12(9):1592.
[51] Liu J, Saxena A, Goebel K, Saha B, Wang W. An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries. In: Annual conference of the PHM society, vol. 2, (no. 1). 2010.
[52] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 1994;5(2):157–66.
[53] Nascimento RG, Viana FA. Fleet prognosis with physics-informed recurrent neural networks. 2019, arXiv preprint arXiv:1901.05512.
[54] Fatemi A, Yang L. Cumulative fatigue damage and life prediction theories: a survey of the state of the art for homogeneous materials. Int J Fatigue 1998;20(1):9–34.
[55] Kim SW, Oh K-Y, Lee S. Novel informed deep learning-based prognostics framework for on-board health monitoring of lithium-ion batteries. Appl Energy 2022;315:119011.
[56] Nascimento RG, Viana FA, Corbetta M, Kulkarni CS. A framework for li-ion battery prognosis based on hybrid Bayesian physics-informed neural networks. Sci Rep 2023;13(1):13856.
[57] Chen C-H, Planella FB, O'regan K, Gastol D, Widanage WD, Kendrick E. Development of experimental techniques for parameterization of multi-scale lithium-ion battery models. J Electrochem Soc 2020;167(8):080534.
[58] Deng Z, Xu L, Liu H, Hu X, Duan Z, Xu Y. Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles. Appl Energy 2023;339:120954.
[59] Bole B, Kulkarni C, Daigle M. Randomized battery usage data set. NASA AMES Progn Data Repos 2014;70.
[60] Bole B, Kulkarni CS, Daigle M. Adaptation of an electrochemistry-based li-ion battery model to account for deterioration observed under randomized use. In: Annual conference of the PHM society, vol. 6, (no. 1). 2014.
[61] Rozemberczki B, Watson L, Bayer P, Yang H-T, Kiss O, Nilsson S, et al. The shapley value in machine learning. 2022, arXiv preprint arXiv:2202.05594.
[62] Connor JT, Martin RD, Atlas LE. Recurrent neural networks and robust time series prediction. IEEE Trans Neural Netw 1994;5(2):240–54.
[63] Hochreiter S. Long short-term memory. 1997, Neural Computation MIT-Press.
[64] Cho K. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014, arXiv preprint arXiv:1406.1078.

[65] Doyle M, Newman J, Gozdz AS, Schmutz CN, Tarascon J-M. Comparison of modeling predictions with experimental data from plastic lithium ion cells. J Electrochem Soc 1996;143(6):1890.
[66] Plett GL. Battery management systems, Volume I: Battery modeling. Artech House; 2015.
[67] Sulzer V, Marquis SG, Timms R, Robinson M, Chapman SJ. Python battery mathematical modelling (PyBaMM). J Open Res Softw 2021;9(1).
[68] O'Kane SE, Ai W, Madabattula G, Alonso-Alvarez D, Timms R, Sulzer V, et al. Lithium-ion battery degradation: how to model it. Phys Chem Chem Phys 2022;24(13):7909–22.
[69] Li R, O'Kane S, Marinescu M, Offer GJ. Modelling solvent consumption from SEI layer growth in lithium-ion batteries. J Electrochem Soc 2022;169(6):060516.
[70] Li R, O'Kane S, Huang J, Marinescu M, Offer GJ. A million cycles in a day: enabling high-throughput computing of lithium-ion battery degradation with physics-based models. J Power Sources 2024;598:234184.
[71] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019, p. 2623–31.
[72] Thelen A, Li M, Hu C, Bekyarova E, Kalinin S, Sanghadasa M. Augmented model-based framework for battery remaining useful life prediction. Appl Energy 2022;324:119624.
[73] Goebel K, Saha B, Saxena A, Celaya JR, Christophersen JP. Prognostics in battery health management. IEEE Instrum Meas Mag 2008;11(4):33–40.
[74] Wang D, Miao Q, Pecht M. Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model. J Power Sources 2013;239:253–64.
[75] Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: International conference on machine learning. PMLR; 2016, p. 1050–9.
[76] DeGroot MH, Fienberg SE. The comparison and evaluation of forecasters. J R Stat Soc: Ser D (Statistician) 1983;32(1–2):12–22.
[77] Kuleshov V, Fenner N, Ermon S. Accurate uncertainties for deep learning using calibrated regression. In: International conference on machine learning. PMLR; 2018, p. 2796–804.
[78] Prada E, Di Domenico D, Creff Y, Bernard J, Sauvant-Moynot V, Huet F. Simplified electrochemical and thermal model of LiFePO4-graphite li-ion batteries for fast charge applications. J Electrochem Soc 2012;159(9):A1508.