



Cite this: DOI: 10.1039/d6eb00042h

A multi-stage modelling framework for accurate early-life SoH and EOL prediction

Ahmed Abdelrahman and Jie Liu *

Accurate estimation of the State of Health (SOH) of lithium-ion batteries is essential to ensure their reliable performance and operational safety. This study integrates functional principal component analysis (FPCA) of incremental capacity (ICA) curves and then propagates the extracted functional features through a two-stage learning architecture, where a linear regression model captures the dominant degradation trend and a Gaussian process regression (GPR) layer adaptively corrects the nonlinear residuals. This multi-stage framework enables highly accurate prediction of both state-of-health (SoH) evolution and end-of-life (EOL) timing from early-life data. Initially, Incremental Capacity Analysis (ICA) is employed to extract characteristic voltage–capacity features that encapsulate the electrochemical degradation behavior of the cells. These features are subsequently transformed through functional decomposition using FPCA, enabling the identification of dominant degradation modes and temporal evolution patterns within the cycling data. The resulting functional components are then mapped to the SOH through a hybrid regression framework, combining the linear interpretability of traditional regression techniques with the nonlinear learning capability of GPR. The proposed approach demonstrates strong predictive performance, achieving a mean absolute percentage error (MAPE) of 2.99% and an average end-of-life (EOL) prediction error of 4.95%, underscoring its effectiveness and robustness in accurately characterizing lithium-ion battery degradation.

Received 22nd February 2026,
Accepted 5th May 2026

DOI: 10.1039/d6eb00042h

rsc.li/EESBatteries

Broader context

Accurate early-life estimation of the State of Health (SoH) and End of Life (EOL) of lithium-ion batteries is critical for ensuring safety, reliability, and cost-effective operation in electric vehicles (EVs) and energy storage systems (ESS). Conventional model-based and data-driven approaches often rely on predefined degradation models, which limits their ability to generalize across varying operating conditions and aging behaviors, which led to the growing interest in non-parametric degradation analysis techniques. At the same time, Incremental Capacity Analysis (ICA) provides valuable insight into electrochemical degradation mechanisms, however, its high dimensionality and sensitivity to noise hinder its direct application in prognostic frameworks. To address these challenges, this work adopts a functional data analysis approach in which ICA curves are treated as continuous functions and decomposed using Functional Principal Component Analysis (FPCA) to extract dominant degradation modes and their evolution across the battery's life cycle. The resulting low-dimensional functional features are then propagated through a two-stage regression model, where linear regression captures the dominant global aging trend and Gaussian Process Regression (GPR) adaptively models nonlinear residual behavior. This hybrid and interpretable framework enables highly accurate prediction of SoH evolution and EOL timing from early-life data, demonstrating strong robustness and predictive performance and highlighting the potential of FPCA-based functional representations for next-generation battery prognostics.

1 Introduction

The use of electric vehicles has grown much faster than expected over the past few years. In 2024, more than 17 million EVs were sold worldwide,¹ which means that about one in five new cars was electric for the first time. China is leading this growth, selling over 11 million EVs in 2024, almost as many as the whole world sold just two years earlier. Other regions are

also showing strong increases. Emerging markets in Asia and Latin America are becoming new centres of growth, with electric car sales jumping by over 60% in 2024 to almost 600 000, about the size of the European market 5 years earlier. In Canada, new registrations of battery electric vehicles went up by about 57% in 2024 compared to the previous year, and plug-in hybrids rose by 75% in the same period.² 1 in 4 new cars sold in 2025 will be electric, new report says, and China is the undisputed EV leader.³

Such growth in the adoption of electric vehicles has led to a much greater reliance on lithium-ion batteries as the main source of power.⁴ With this widespread usage comes the chal-

Department of Mechanical and Aerospace Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada. E-mail: jie.liu@carleton.ca



lence of dealing with how these batteries age and lose capacity over time. As drivers depend on their vehicles for both performance and range, battery state of health (SOH) estimation and prediction have become increasingly important.⁵

SOH estimation for lithium-ion batteries (LIBs) have been searched in various ways.⁶ The most common ways are model-based methods and data-driven methods. Model-based methods involve parameter estimation of battery equivalent circuit models or electrochemical models.⁷

Model-based methods, including equivalent-circuit,⁸ electrochemical,⁹ and degradation-mechanism-based formulations, are attractive because they are physically interpretable and can embed known battery behaviour. However, their practical use is often constrained by parameter identifiability,¹⁰ sensitivity to operating conditions, chemistry dependence, and the computational burden associated with online implementation of electrochemical models. Recent reviews have emphasized that no single physics-based model can fully capture the coupled and evolving degradation pathways of lithium-ion batteries across varying temperatures, current rates, and usage histories, especially under realistic field conditions.¹¹ These issues become even more pronounced when transferring calibrated models across cells, chemistries.

Data-driven methods, by contrast, are well suited to learning nonlinear degradation patterns directly from measured voltage, current, temperature, or diagnostic features, and they have achieved strong predictive accuracy in many controlled datasets. Nevertheless, their limitations are now well documented. First, they are often heavily dependent on large, well-labeled, and distributionally representative datasets,¹² which are difficult to obtain in early life or under real-world EV operation. Second, many models suffer from weak generalization when the testing domain differs from the training domain in terms of operating window, charging protocol, ambient temperature, or cell-to-cell variability. Third, many deep learning models remain difficult to interpret physically, which limits trust and deployment in safety-critical battery management settings.¹³

A further limitation shared by both model-based and data-driven methods is their difficulty in delivering reliable early-life prediction. Early degradation trajectories are typically subtle, noisy, and partially confounded by formation effects, hysteresis, and operating-condition variability. As a result, models built only on cycle index or simple scalar health indicators may miss higher-order structure in the degradation trajectory. This limitation has motivated the move toward richer signal representations and multi-source features. For example, internal-resistance-informed learning has been proposed as a way to bridge equivalent-circuit information and data-driven mapping, improving robustness and interpretability.

In response to these limitations, recent research has increasingly shifted toward hybrid physics-informed machine learning methods.^{14,15} These approaches aim to combine the extrapolation capability and interpretability of physics-based models with the flexibility and pattern-recognition strength of data-driven techniques. Aykol *et al.*¹⁶ argued that neither purely physics-based nor purely data-driven approaches are

sufficient in isolation, and proposed several integration strategies that explicitly fuse mechanistic understanding with statistical learning for battery lifetime prediction. Building on this perspective, subsequent studies have operationalized hybrid frameworks in practical settings. For instance, Nascimento *et al.*¹⁷ developed hybrid physics-informed neural networks for battery modelling and prognosis, while Kohtz *et al.*¹⁸ introduced a physics-informed learning framework leveraging partial charging segments with embedded physical constraints. More recently, Wang *et al.*¹⁹ proposed a physics-informed neural network capable of stable SOH estimation across varying battery chemistries and operating conditions. Collectively, these studies indicate that hybrid approaches can improve generalization, reduce data requirements, and enforce physically consistent predictions, particularly when full-cycle labelled datasets are unavailable.

Another important emerging direction is uncertainty-aware prognostics. Conventional SOH estimation methods typically produce deterministic point predictions, which can be misleading given the stochastic nature of degradation and the significant variability observed across battery populations. To address this, probabilistic and uncertainty-aware frameworks have been developed to quantify prediction confidence and support more reliable decision-making in battery management systems. Thelen *et al.*²⁰ provided a comprehensive review of probabilistic machine learning approaches for battery diagnostics and emphasized uncertainty quantification as a critical requirement for real-world deployment. In a similar vein, Zou *et al.*²¹ introduced a Bayesian model averaging approach for SOH estimation that captures both model-form and parameter uncertainty, enabling probabilistic outputs alongside point estimates. Earlier work by Richardson *et al.*²² also highlighted the importance of uncertainty quantification in battery lifetime prediction using Gaussian process regression. These developments are particularly important for early-life prediction, where limited degradation information and higher noise levels inherently increase uncertainty, and deterministic models may otherwise lead to overconfident and potentially misleading predictions.

Despite these advancements, a fundamental limitation remains: most existing approaches, whether physics-based, data-driven, or hybrid, still rely on compressed representations of battery degradation, typically in the form of scalar indicators such as capacity and internal resistance or a limited set of engineered features extracted from voltage and current signals. While hybrid frameworks improve robustness and predictive reliability, they generally operate on reduced-dimensional representations that may not fully capture the underlying structure of degradation processes.

In particular, degradation in lithium-ion batteries is reflected in the continuous evolution of electrochemical signals, such as voltage profiles, incremental capacity curves,²³ and differential voltage characteristics. Prior studies have shown that these signals contain rich information about electrode-level aging mechanisms, phase transitions, and kinetic limitations that are not adequately preserved when reduced to



scalar features or sparse descriptors. Even in advanced machine learning pipelines, feature extraction is often performed as a preprocessing step, resulting in the loss of important shape-based information and temporal dependencies embedded in the full signal.

This creates a critical gap in existing SOH estimation and prognostics frameworks: the lack of methodologies that explicitly model degradation as a functional object rather than a collection of discrete features. As a result, subtle yet systematic changes in curve morphology, particularly in early life, may remain undetected or underutilized.

To address this limitation, this paper introduces a multi-stage functional framework for early and accurate prediction of lithium-ion battery (LIB) degradation.

The key contributions are summarized as follows:

1. Functional decomposition and multi-stage predictive learning: this work establishes a functional-learning approach in which ICA curves are represented in a compact FPCA subspace that captures the dominant electrochemical ageing modes, including peak migration and compression. Instead of relying on a single regression model, we propose a multi-stage hybrid architecture that first extracts the deterministic ageing trajectory through linear regression and then models the remaining nonlinear, cycle-dependent dynamics using Gaussian process regression (GPR). This layered structure delivers higher predictive accuracy, improved robustness to noise, and enhanced physical interpretability, outperforming conventional single-stage ICA-based approaches.

2. Early-life SoH and EOL forecasting with an early-cycle observation window: the proposed framework delivers high-accuracy prediction of SoH evolution and end-of-life EOL using only the first 33% of the battery's operational life, offering a level of early-cycle sensitivity that is rarely achieved in existing prognostic methods. By exploiting the functional structure embedded in ICA curves and refining the remaining nonlinear dynamics through an uncertainty-aware GPR layer, the model provides reliable long-horizon forecasts at the battery's optimal operating temperature.

Such early and accurate degradation prediction is pivotal for a wide range of applications. For EVs, early EOL prediction is crucial for anticipating when a battery pack will no longer meet performance requirements, supporting informed decisions on vehicle resale value, second-life deployment, and long-term fleet planning. In smartphones, laptops, and other portable electronics, it helps prevent unexpected failures and enables proactive battery replacement. The framework therefore, enhances reliability and health-aware management across both large-scale and consumer battery systems.

The graphical abstract below shown in Fig. 1 illustrates the proposed multi-stage functional framework for early prediction of lithium-ion battery degradation and SoH. The process begins with the extraction of ICA features from voltage-capacity profiles, which are subsequently smoothed, aligned, and normalized to form comparable functional representations. FPCA is then applied to capture the dominant degradation modes as a set of low-dimensional functional scores.

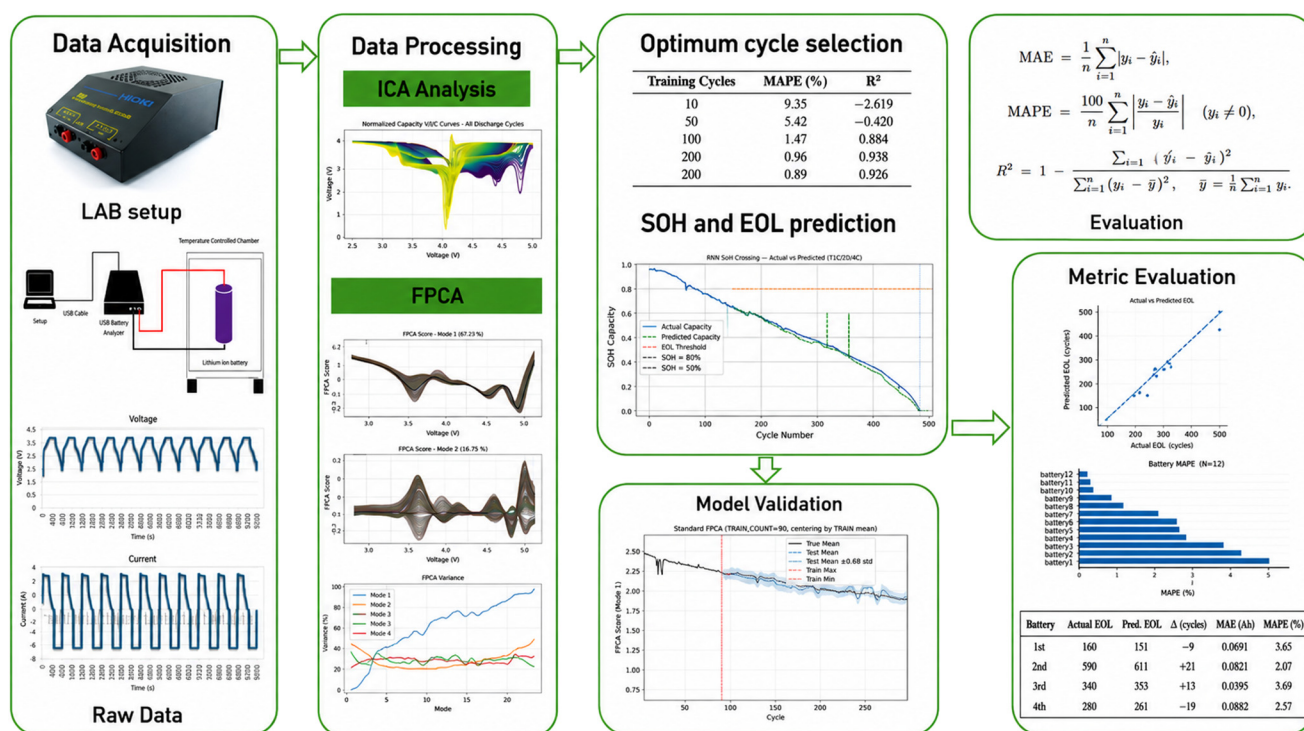


Fig. 1 Unified hybrid functional framework integrating ICA feature extraction, FPCA-based functional representation, and hybrid linear-GPR prediction for early estimation of lithium-ion battery degradation and end-of-life.



These scores are fed into a two-stage regression model, combining the global linear trend with nonlinear residual learning *via* GPR, to estimate both SoH and EOL at different operating temperatures. The integrated workflow bridges electrochemical interpretability with statistical learning, enabling accurate and early EOL prediction after observing only one-third of the total cycle life.

2 Literature review

FPCA extends standard PCA to functional data, enabling curves to be expressed in terms of mean functions and eigenfunctions.²⁴

FPCA was first brought into use for MRI by Viviani *et al.*,²⁵ mainly to find and summarize the main patterns in brain activity data.

FPCA has been applied alongside several types of battery diagnostic curves to capture degradation behaviour compactly. One well-known diagnostic tool is the ICA curve, which can reveal distinct signatures of aging and has been widely used in SoH estimation. FPCA has been adopted in battery prognostics to reduce high-dimensional cycle data into a few informative modes of variation and then regress health indicators on those mode scores. Early work modeled discharge-voltage trajectories with FPCA and a Bayesian updating scheme to predict residual life, demonstrating that nonparametric functional representations can capture degradation dynamics effectively.²⁶ Subsequent studies applied FPCA to multi-sensor monitoring streams (voltage, current, temperature) and used sparse regression to predict capacity/SOH from functional scores, showing competitive accuracy on the NASA datasets.²⁷ More recently, nonparametric degradation frameworks have incorporated FPCA within reliability settings to handle incomplete/truncated signals for remaining useful life (RUL) prediction.²⁸ Beyond operational datasets, FPCA has also been used to compress and forecast time-resolved outputs in battery manufacturing simulations, highlighting its utility as a general functional-data compressor before downstream prediction.²⁹ While FPCA has demonstrated strong capability in extracting low-dimensional degradation features across diverse battery signals, accurate early-life EOL prediction remains challenging, with errors generally exceeding 6–10% or tens of cycles MAE. This motivates the development of a more structure-aware functional framework capable of leveraging derivative-based diagnostics such as ICA while maintaining robustness to noise and misalignment.

In parallel, the ICA literature has largely focused on extracting peak/shape features from incremental-capacity curves and then learning SOH/RUL with statistical or machine-learning models. Recent works pair ICA-derived indicators with Gaussian Process Regression (GPR) to forecast RUL in high-capacity cells.³⁰ Separately, GPR has been extensively employed for capacity/SOH estimation directly from voltage–time snippets or operating-condition inputs without ICA.^{31,32} To the best of our knowledge, no existing work has jointly leveraged

ICA, FPCA, and a two-stage modelling architecture combining linear regression with Gaussian process residual learning for SOH estimation and EOL prediction. Our approach is the first to formalize ICA curves as functional representations of ageing behaviour, extract degradation modes *via* FPCA, and then model capacity evolution by using a linear model for the global trend followed by a GPR model that captures nonlinear residual behaviour. This integrated, mode-aware formulation enables earlier and more accurate EOL estimation than traditional scalar-feature or single-model approaches.

However, ICA curves present practical challenges: they are often noisy, highly sensitive to testing conditions, and produce large volumes of data that are difficult to process directly. To address these limitations, this study applies FPCA to extracted ICA curves, capturing the dominant degradation patterns and transforming them into a compact set of representative features. These FPCA scores are then tracked across cycles, enabling a direct connection between their evolution and the battery's degradation behavior. Subsequently, the extracted functional features are used as inputs to a multi-stage regression framework that combines linear regression with GPR to predict the battery's SoH and end-of-life EOL.

3 Proposed framework

This section outlines the methodological framework for ICA processing and SoH estimation using FPCA. The dataset used in this methodology was generated in our laboratory through controlled cycling tests. An Efest 18650 LIB with a nominal voltage of 3.7 V was employed. The cell was continuously cycled using the UBA5 battery analyzer until it reached its EOL condition, defined as 80% of the initial capacity, thereby capturing the full degradation trajectory required for state-of-health analysis and prediction. The charge and discharge were performed at a 1C rate under controlled conditions at a temperature of 25 °C.

Before going into detail about the ICA processing and the structure of predictive model, it is important to recognize the challenges inherent in using ICA for functional learning. ICA curves are highly sensitive to noise, easily lose monotonicity, and exhibit nonlinear peak shifts with ageing, all of which complicate their direct use in functional decomposition. Applying FPCA to such data is nontrivial: misalignment, noise amplification, and mixed electrochemical effects can produce principal components that are difficult to interpret and unstable across cycles. Even when meaningful components are extracted, their relationship with capacity is only approximately linear, leaving residual nonlinear structure unmodeled and necessitating an additional modeling stage to capture the remaining degradation dynamics.

These challenges motivate the need for a carefully designed preprocessing pipeline and a multi-stage modelling strategy. In the next subsection, we outline how ICA monotonic reconstruction, voltage-axis alignment, functional centering, and staged linear–nonlinear regression are integrated to overcome



these limitations and yield stable, interpretable, and cycle-sensitive prognostic features.

3.1 Incremental capacity analysis

ICA is performed on each cycle, allowing evolution of the voltage–capacity relationship to be visualized over time, making it possible to clearly observe how the battery degrades across cycling. By tracking the shifts and changes in ICA curves, early signs of capacity fade and internal resistance growth can be identified,³³ providing the foundation for further analysis and SOH estimation.

The ICA curve is obtained by differentiating discharge/charge capacity with respect to voltage. Given time-sampled data ($V(t)$, $I(t)$), the accumulated capacity is:

$$Q(t) = \int_0^t I(\tau) d\tau. \quad (1)$$

The ICA curve is then estimated as:

$$\frac{dQ}{dV} = \frac{\Delta Q}{\Delta V}. \quad (2)$$

3.2 ICA preprocessing

ICA preprocessing ensures comparability and noise reduction across cycles. The main steps include enforcing the monotonicity of $Q(V)$, voltage alignment, and smoothing, each described mathematically below.

1. Monotonic construction of $Q(V)$

During constant-current discharge, the instantaneous capacity is obtained by integrating the current over time as shown in eqn (1):

$$Q(t) = \int_0^t I(\tau) d\tau. \quad (1)$$

where $I(\tau)$ is the discharge current. Because voltage V generally decreases as discharge progresses, Q should be a monotonic function of V . However, voltage measurement noise and transient effects may cause small reversals in $V(Q)$. Therefore, samples are first sorted by voltage and median-aggregated at duplicate V values:

$$Q(V_j) = \text{median}\{Q_k | V_k = V_j\} \quad (3)$$

This step guarantees that $Q(V)$ is single-valued and monotonic, enabling a stable numerical derivative $\frac{dQ}{dV}$ for ICA computation.

2. Voltage alignment to a common axis

To enable point-wise comparison across cycles, each capacity curve $Q(V)$ is interpolated to a common voltage axis $V_i \in [2.5, 4.25]$ V using:

$$Q_i = \text{interp}(Q(V_k), V_k, V_i) \quad (4)$$

This alignment ensures that all curves share identical voltage coordinates, thereby allowing element-wise operations such as mean subtraction and covariance computation across

cycles. Such consistency is essential for building the covariance matrix used in functional principal component analysis (FPCA), defined as:

$$\Sigma = \frac{1}{N-1} \sum_{n=1}^N (Q_n - \bar{Q})(Q_n - \bar{Q})^T \quad (5)$$

where Q_n is the n -th aligned and smoothed incremental capacity curve and \bar{Q} is their mean function.

3. Smoothing and filtering

Differentiation amplifies noise; hence smoothing is applied to $Q(V)$ or directly to dQ/dV .

The ICA smoothing strategy was selected based on a sensitivity analysis that quantified the impact of preprocessing on the downstream functional principal component analysis (fPCA). Several filtering techniques were evaluated, including moving average, Gaussian smoothing, Savitzky–Golay (SG), and weighted Savitzky–Golay (WSG). The evaluation was performed relative to the raw ICA baseline by measuring changes in the mean ICA curve, explained variance structure, and fPCA score stability.

The smoothing methods are defined as follows. The moving-average filter is

$$\tilde{x}_i = \frac{1}{M} \sum_{k=-m}^m x_{i+k}, \quad M = 2m + 1 \quad (6)$$

The Gaussian filter is

$$\tilde{x}_i = \sum_{k=-m}^m w_k x_{i+k}, \quad w_k = \frac{\exp\left(-\frac{k^2}{2\sigma^2}\right)}{\sum_{j=-m}^m \exp\left(-\frac{j^2}{2\sigma^2}\right)} \quad (7)$$

The Savitzky–Golay filter fits a local polynomial:

$$\tilde{x}_i = \hat{a}_0, \quad \{\hat{a}_r\} = \text{argmin} \sum_{k=-m}^m \left(x_{i+k} - \sum_{r=0}^p a_r k^r \right)^2 \quad (8)$$

The weighted variant becomes:

$$\tilde{x}_i = \hat{a}_0, \quad \{\hat{a}_r\} = \text{argmin} \sum_{k=-m}^m \omega_k \left(x_{i+k} - \sum_{r=0}^p a_r k^r \right)^2 \quad (9)$$

Table 1 summarizes the most relevant metrics from the sensitivity analysis. The weighted Savitzky–Golay filter achieved the lowest overall instability among the filtered cases, with

Table 1 Comparison of ICA filtering methods based on fPCA stability

Filter	RMSE	EV shift	fPC corr.	Score RMSE
WSG (11,2)	0.0068	0.310	0.572	0.658
Moving (5)	0.0107	0.370	0.567	0.784
SG (11,2)	0.0100	0.467	0.399	1.024
Gaussian (2)	0.0154	0.242	0.339	—

Lower RMSE and variance shifts, and higher correlations indicate better performance.



improved score stability and strong preservation of the mean ICA shape. Although Gaussian smoothing resulted in smaller variance shifts, it significantly degraded the fPCA structure. Therefore, the Savitzky–Golay family was selected due to its balance between smoothing and structural preservation.

To further justify the preprocessing choice, a dedicated sensitivity analysis was conducted on the Savitzky–Golay parameters, namely the window size M and polynomial order p . The results are summarized in Table 2. It is observed that smaller window sizes with higher polynomial orders provide the most stable fPCA representation.

In particular, the configuration $(M,p) = (5,4)$ produced nearly identical results to the raw baseline, indicating minimal distortion. Among practical smoothing configurations, (7,4) and (9,4) offered the best trade-off between noise reduction and preservation of the underlying functional structure. Larger window sizes consistently increased score RMSE and reduced fPC correlation, confirming that excessive smoothing distorts the extracted principal modes.

Overall, this analysis demonstrates that the fPCA results are sensitive to ICA preprocessing, and that Savitzky–Golay filtering with carefully selected parameters provides a robust and interpretable solution. The final choice is therefore guided by quantitative stability criteria rather than arbitrary smoothing assumptions.

The Savitzky–Golay filter is commonly used, which performs a local polynomial fit and computes a smoothed value or derivative analytically:

$$Q_s(V_i) = \sum_{j=-m}^m c_j Q(V_{i+j}) \quad (10)$$

where the coefficients c_j correspond to a least-squares polynomial of order p over a window of $2m + 1$ points. When the derivative is required, the same filter yields:

$$\frac{dQ_s}{dV}(V_i) = \sum_{j=-m}^m c'_j Q(V_{i+j}) \quad (11)$$

3.3 FPCA for SoH detection

When plotting all ICA curves together across hundreds of cycles, the result is often overcrowded and difficult to interpret. The curves overlap heavily, and the noise introduced by measurement variations makes it challenging to clearly see the underlying patterns of degradation. This creates a problem

Table 2 Savitzky–Golay parameter sensitivity analysis (top-performing configurations)

(M,p)	RMSE	EV shift	fPC corr.	Score RMSE
(5,4)	≈0	≈0	1.000	≈0
(7,4)	0.0017	0.052	0.541	0.541
(9,4)	0.0033	0.176	0.678	0.504
(5,3)	0.0021	0.049	0.497	0.559
(5,2)	0.0023	0.049	0.497	0.559

when trying to extract meaningful features for SOH estimation, as important trends can easily be hidden within the clutter. To address this, we applied FPCA, which reduces the dimensionality of the ICA data while preserving its key modes of variation. By transforming the curves into a smaller set of principal components and corresponding scores, FPCA provides a clearer and more manageable way to capture how the battery evolves over time, making degradation trends easier to track and relate directly to SOH. The resulting coefficients (scores) are used as predictors for SoH estimation.

Let $x_i(v)$ denote the ICA curve (e.g., dQ/dV) for cycle $i \in 1, \dots, N$, defined on a common voltage domain $v \in [a, b]$ (e.g., [2.5, 4.2] V). This represents the studied cells' normal working range. FPCA seeks a mean function $\mu(v)$ and orthonormal eigenfunctions $\{\phi_k(v)\}_{k \geq 1}$ such that

$$x_i(v) = \mu(v) + \sum_{k=1}^{\infty} \xi_{i,k} \phi_k(v), \quad (12)$$

where $\int_a^b \phi_k(v) \phi_\ell(v) dv = \delta_{k\ell}$.

The FPCA mean is the pointwise average over cycles. The mean ICA curve gives the average behavior, and centering shows how each cycle differs from that average. By looking at how these differences progress from one cycle to the next, we can see the signs of degradation taking shape:

$$\mu(v) = \frac{1}{N} \sum_{i=1}^N x_i(v), \quad \tilde{x}_i(v) = x_i(v) - \mu(v). \quad (13)$$

Covariance shows how different parts of the ICA curve change together across cycles. In batteries, this means that if one peak shifts or shrinks, covariance reveals which other regions of the curve tend to change at the same time, pointing to linked degradation processes.

$$C(v_1, v_2) = \frac{1}{N} \sum_{i=1}^N \tilde{x}_i(v_1) \tilde{x}_i(v_2). \quad (14)$$

After covariance, we extract eigenfunctions. Eigenfunctions are the main patterns of how the ICA curves change. Each one highlights a different way the battery ages, like peaks shrinking or shifting, and their scores tell us how much each cycle follows that pattern.

The main functional patterns of degradation are obtained as eigenfunctions of the covariance operator, solving

$$\int_a^b C(v, u) \phi_k(u) du = \lambda_k \phi_k(v), \quad \lambda_1 \geq \lambda_2 \geq \dots \geq 0, \quad (15)$$

with orthonormality condition

$$\int_a^b \phi_k(v) \phi_\ell(v) dv = \delta_{k\ell}. \quad (16)$$

Each eigenfunction $\phi_k(v)$ represents a dominant mode of variation, and its eigenvalue λ_k quantifies the variance explained by that mode.



Once the eigenfunctions are obtained, each ICA curve can be represented as a combination of these orthogonal basis functions:

$$x_i(v) = \mu(v) + \sum_{k=1}^{\infty} \xi_{i,k} \phi_k(v), \quad (17)$$

where the functional principal component (FPC) scores are computed as

$$\xi_{i,k} = \int_a^b \tilde{x}_i(v) \phi_k(v) dv. \quad (18)$$

The score $\xi_{i,k}$ measures how strongly cycle i exhibits the degradation pattern described by eigenfunction $\phi_k(v)$.

The total functional variance is given by

$$\sum_{k \geq 1} \lambda_k, \quad (19)$$

and the fraction of this variance captured by component k is expressed as the explained variance ratio (EVR):

$$\text{EVR}_k = \frac{\lambda_k}{\sum_{r=1}^P \lambda_r}. \quad (20)$$

The cumulative explained variance ratio (CEVR) for the first K components is

$$\text{CEVR}_K = \sum_{k=1}^K \text{EVR}_k. \quad (21)$$

In this study, five principal components were retained, which together captured more than 95% of the total functional variance.

The FPCA scores derived in this section serve as the input features to the proposed hybrid prediction framework, which combines linear regression and GPR to estimate the battery's SOH. The following section 3.4 presents the structure and formulation of this predictive model in detail.

3.4 Hierarchical linear-GPR predictive model

This section describes the hierarchical modelling framework used to predict battery capacity from functional data. The model operates in two stages: a linear component first captures the dominant, interpretable aging trend extracted through FPCA, and a Gaussian process (GP) model is then applied to learn the residual nonlinear structure that the linear model cannot explain. The approach is semi-parametric: it combines the interpretability and simplicity of a linear model with the flexibility of a non-parametric GP.

We retain only the first few (typically up to five) fPCA scores for each cycle, as these capture most of the functional variability while avoiding overfitting. These scores form the input feature vector for the regression model.

These compact, information-rich components serve as the predictive features for the regression model.

3.4.1 Linear mean model. The dominant trend between the FPCA features and the target variable (capacity) is approximately linear. Thus, we first fit a simple linear regression model:

$$\hat{y}_{\text{lin}} = \beta_0 + \sum_{j=1}^K \beta_j z_j,$$

where z_j are the standardized FPCA scores. This component captures the large-scale, global relationship between the functional features and the capacity. The residuals from this linear fit contain the finer, potentially nonlinear structure that the linear model cannot explain.

3.4.2 Gaussian process for residuals. The residuals from the linear model are modelled using a GP with a squared-exponential (RBF) kernel. The GP is a flexible, non-parametric model that defines a distribution over smooth functions. It can learn complex, localized nonlinear dependencies between the input features and the residuals. Modelling residuals instead of full outputs allows the GP to focus on subtle deviations, improving both accuracy and computational efficiency.

The chosen kernel has the form:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{j=1}^K \frac{(x_j - x'_j)^2}{\ell_j^2}\right),$$

where σ_f^2 is the signal variance and ℓ_j are per-feature length-scales controlling sensitivity to each input dimension. These hyperparameters are optimized by maximizing the marginal likelihood of the residual data.

The RBF kernel was chosen because battery degradation and capacity fade happen gradually, not suddenly. Cycles that are close to each other in time or have similar FPCA features are expected to show similar capacity values. The RBF kernel captures this smooth and continuous behavior by making nearby points strongly correlated and distant points weakly correlated. After the linear regression step, the remaining residuals are small and smooth, with no large jumps or oscillations, so the RBF kernel is well suited to model these +non-linear corrections.

The final prediction is obtained by combining the linear mean estimate and the GP-predicted residual mean:

$$\hat{y} = \hat{y}_{\text{lin}} + \hat{r}_{\text{GP}}.$$

The GP also provides a posterior variance that quantifies the predictive uncertainty of the residual correction, offering confidence intervals for each prediction. This hybrid model therefore provides both interpretability and flexibility while maintaining calibrated uncertainty estimates.

The metrics used for model evaluation were the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), defined as follows:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (22)$$



$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (24)$$

where y_i is the actual value, \hat{y}_i is the predicted value, \bar{y} is the mean of the actual values, and n is the total number of data points.

4 Experimental results

The cycling experiments in this study were conducted under highly controlled laboratory conditions to ensure accuracy and repeatability. Fig. 2 illustrates the complete experimental setup used for data acquisition. This controlled environment enabled precise monitoring of voltage, current, and temperature parameters, ensuring reliable and consistent data for model development and validation.

In this work, the input features x are taken as the principal component (FPCA) scores extracted from ICA curves, which effectively reduce the dimensionality of the degradation signals. The outputs y are the measured SOH values. The hybrid model utilized in this study learns the nonlinear mapping from FPCA scores to SOH, while also quantifying the uncertainty of predictions. This property is particularly valuable in battery prognostics, where reliability and confidence in predictions are crucial (Fig. 3).

To investigate the influence of training size on prediction accuracy, the model was trained using progressively larger portions of early-life cycling data, as summarized in Table 3. Prediction performance consistently improves as additional cycles are included, indicating that longer training histories provide richer information about the underlying degradation process. Nonetheless, using approximately 33% of the total dataset (around 160 cycles) is already sufficient to capture the dominant aging behavior and to support reliable SoH prediction. While further extending the training window continues to yield incremental improvements, the 33% training horizon was chosen because it enables accurate and stable prediction

while allowing SoH estimation to be performed as early as possible in the battery lifetime.

As discussed earlier, the proposed hybrid linear + GPR model integrates a linear regression layer to capture the global degradation trend and a GPR layer to model the nonlinear residuals. Building on this framework, the FPCA scores extracted from the incremental capacity curves were employed as the primary input features for model training.

For comparison, a standalone GPR model was also tested to evaluate the benefit of hybridization. While the GPR-only approach captured the general degradation behavior, its performance was limited, achieving a mean absolute percentage error (MAPE) of 2.18% and a coefficient of determination (R^2) of 0.531 at 155 training cycles. In contrast, the hybrid model demonstrated superior accuracy, with a MAPE of 0.78% and an R^2 of 0.936.

Furthermore, the hybrid model was able to forecast the 80% state-of-health (SOH) threshold at cycle 476, closely matching the experimentally observed end-of-life (EOL) at cycle 480. This strong agreement highlights the model's reliability and robustness across the entire degradation range.

As illustrated in Fig. 4, the hybrid approach provides a smoother and more accurate trajectory than the standalone GPR model, especially in the later stages of cycling, where degradation tends to accelerate. The improved alignment with experimental data validates the effectiveness of combining linear and nonlinear components in a single predictive framework.

The FPCA scores were used as inputs to the two-stage model, which resulted in a mean absolute percentage error (MAPE) of 0.89% and a coefficient of determination (R^2) of 0.926. Based on these features, the model was able to forecast the 80% SOH threshold at cycle 476, which is in close agreement with the actual end-of-life observed at cycle 480.

4.1 Model validation

The laboratory dataset used for model development was obtained from a single cell tested under optimal and well-controlled conditions. While this ensured high consistency and repeatability, it also meant that the data reflected ideal laboratory behavior rather than real-world variability. Therefore, it was essential to further validate the proposed method on multiple cells with different cycling histories and degradation profiles to confirm its robustness and general applicability beyond the controlled lab environment.

The dataset used in this study is the NASA Randomized Battery Usage Dataset (Table 4),³⁴ which contains cycling experiments on commercial 18 650 lithium-ion cells. For consistency, the regular battery subset was selected, as these cells underwent repeatable charge–discharge cycles under controlled conditions. This makes them particularly suitable for ICA and FPCA, as the data provide clear degradation trajectories without additional variability introduced by randomized stress tests. In this dataset, the batteries were cycled through full discharge. This ensured that the complete usable charge of the cell was measured in every cycle.

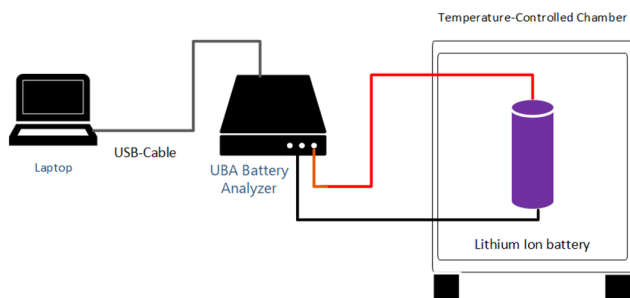


Fig. 2 Experimental setup.



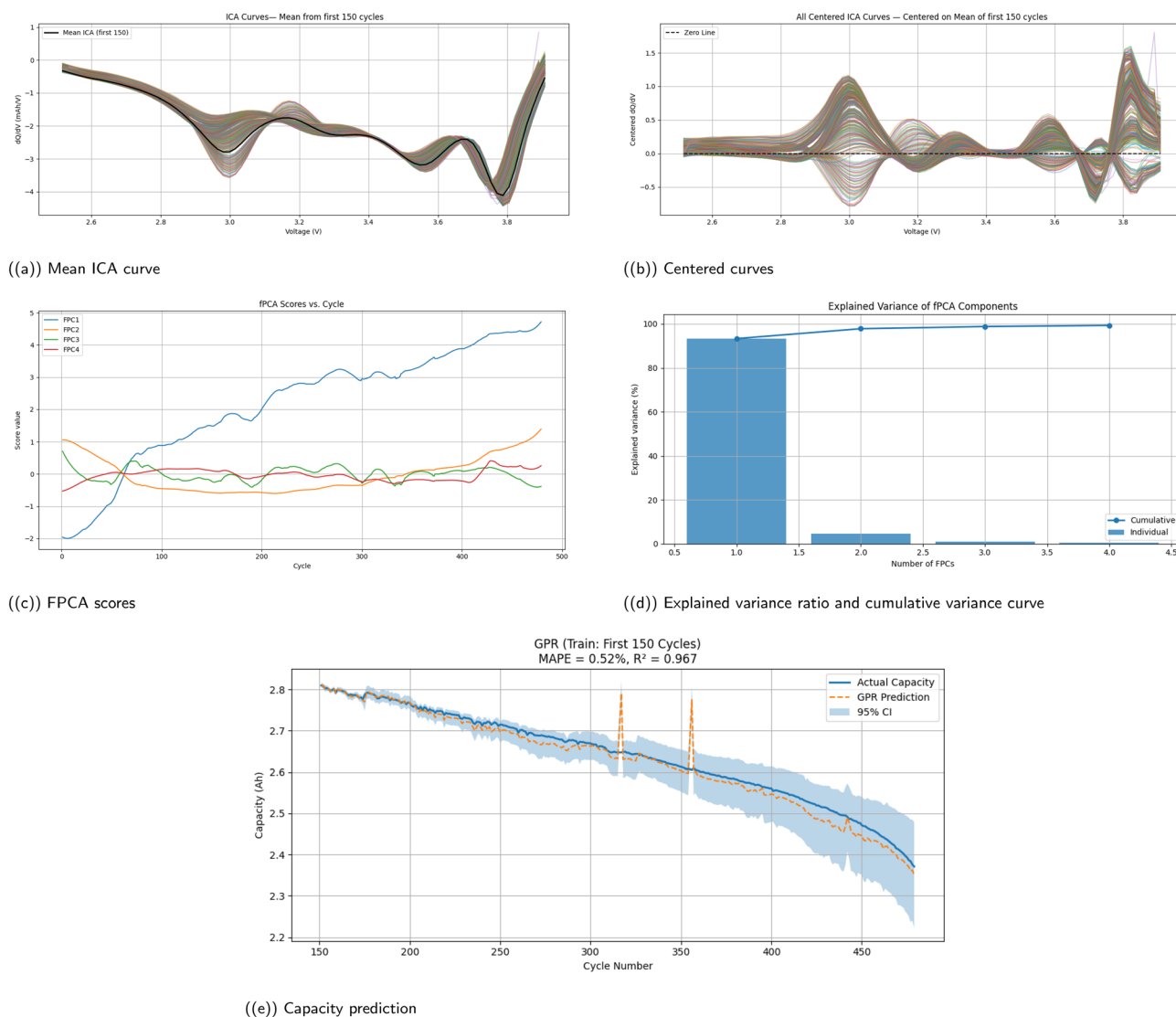


Fig. 3 Overall FPCA-ICA pipeline showing (a) mean ICA curve, (b) centered curves, (c) FPCA scores, (d) explained variance ratio and cumulative variance curve, and (e) capacity prediction.

Table 3 Linear + GP(res) performance at different training sizes

Training cycles	MAPE (%)	R^2
1	9.35	-2.619
10	5.42	-0.420
50	1.47	0.884
100	0.96	0.938
200	0.89	0.926

Moreover, the same preprocessing pipeline applied to the lab dataset was also consistently implemented on the NASA dataset to ensure a unified and fixed methodology across all analyses.

To obtain the capacity, Coulomb counting was applied, where the discharge current is integrated over time. This method makes it possible to quantify the capacity for each

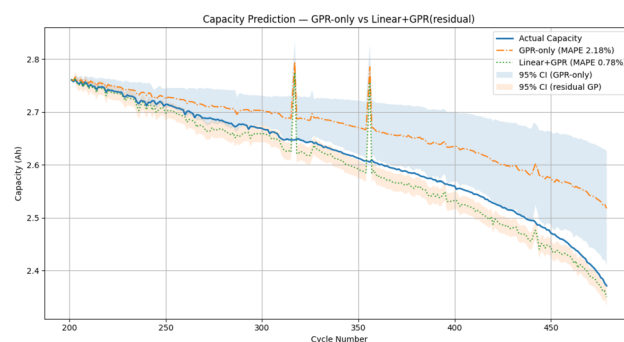


Fig. 4 Comparison between the standalone GPR and the proposed hybrid linear + GPR model. The hybrid approach better captures the actual degradation trajectory and provides a more accurate prediction of the end-of-life (EOL) cycle.



Table 4 Description of the NASA regular battery dataset fields

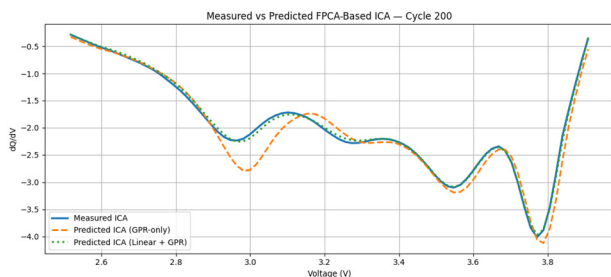
Field (units)	Symbol	Description
Time (s)	t	Elapsed time since experiment start.
Mode (—)	—	0 = rest, 1 = charge, -1 = discharge.
Voltage (V)	V	Terminal cell voltage.
Current (A)	I	Applied current (discharge +, charge -).
Capacity (Ah)	Q	Cumulative transferred charge.
Temperature (°C)	T	Cell surface temperature (if available).
File index (—)	—	One file per battery (e.g., battery01.csv).

cycle and track its reduction over time, thus allowing the degradation of the battery to be observed clearly.

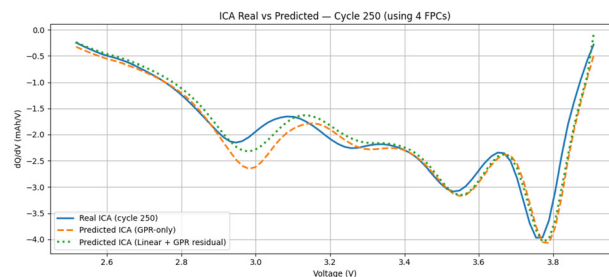
$$Q = \int_{t_0}^{t_f} I(t) dt$$

where Q is the discharged capacity (in Ah), $I(t)$ is the discharge current (in A), and t_0 to t_f represent the start and end of the discharge process.

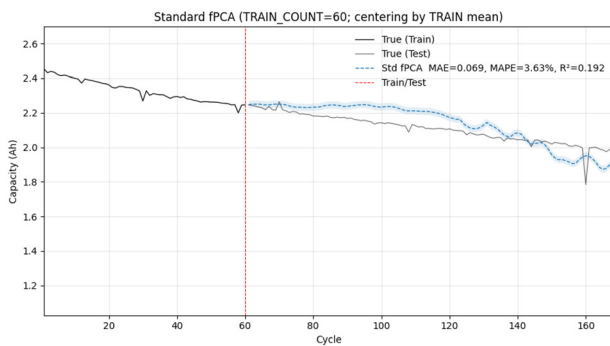
ICA curves were generated for all charge–discharge cycles to capture the evolution of the battery's electrochemical behavior over time. To evaluate the model's predictive capability, the first 33% of the available cycles were used for training, while



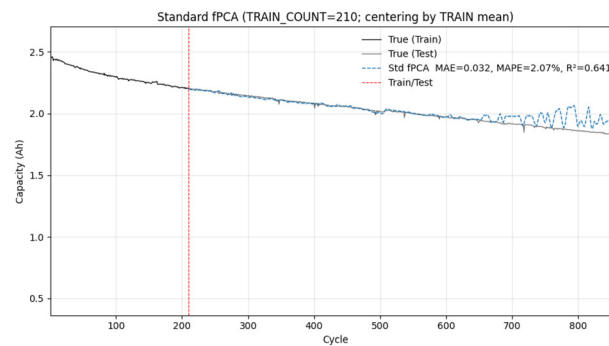
((a)) cycle 200



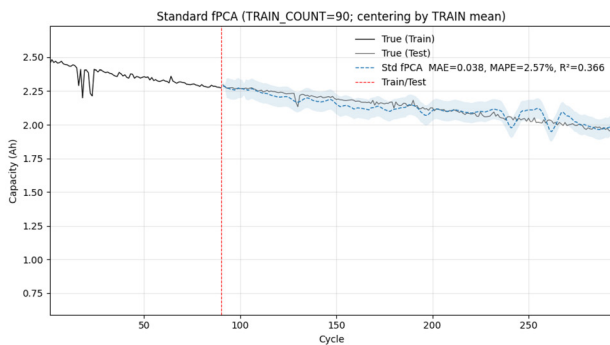
((b)) Cycle 250

Fig. 5 Experimental vs. predicted FPCA-reconstructed ICA.

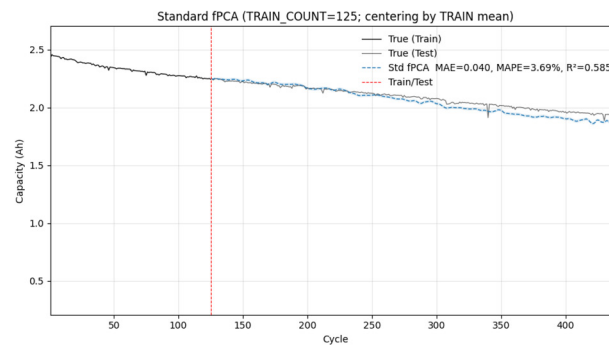
((a)) Battery00



((b)) Battery11



((c)) Battery51



((d)) Battery22

Fig. 6 Overall FPCA pipeline.

Table 5 MAPE of ICA reconstruction using predicted FPCA scores

Cycle	Model	MAPE (%)
200	GPR-only	8.15
	Linear + GPR	1.97
250	GPR-only	13.40
	Linear + GPR	10.55

the remaining cycles were reserved for testing. This approach ensures that the model learns only from the early degradation phase and is then challenged to predict the long-term performance, reflecting a realistic and forward-looking scenario for battery health estimation.

As presented in Fig. 6, the model effectively reproduces the actual degradation trajectory of the battery. The quantitative results summarized in Table 7 confirm this performance, achieving an MAPE of 2.99% in estimating the SoH and an average EOL prediction error of 4.95%. These results demonstrate that the model maintains strong predictive accuracy even when trained on a limited portion of the data, highlighting its ability to generalize degradation trends beyond the training range.

Further comparison in Fig. 7 shows a close correspondence between the predicted and actual EOL cycles. The predicted curve follows the measured degradation path with high consistency, with minor deviations appearing only near the end-of-life region, where degradation dynamics become more non-linear and difficult to model. Despite this, the multi-stage functional framework effectively captures the underlying progression of degradation, illustrating the robustness of combining ICA-based feature extraction with FPCA-GPR modeling.

Following the evaluation of SoH estimation and EOL prediction accuracy, the predictive models were further examined in the functional domain through ICA reconstruction. Specifically, ICA curves were reconstructed from predicted FPCA scores using two approaches: a GPR-only score model and a hybrid linear + GPR residual model, and compared against the experimentally derived ICA curves at representative mid-to-late life cycles. This analysis provides insight beyond capacity-level metrics by assessing how well each model preserves the underlying electrochemical degradation signatures encoded in the ICA profiles.

Table 7 Comparison of actual vs. FPCA predicted EOL and performance metrics

Battery	Actual EOL	Pred. EOL	Δ (cycles)	MAE (Ah)	MAPE (%)
1st	160	151	-9	0.0691	3.63
2nd	590	611	+21	0.0321	2.07
3rd	340	353	+13	0.0395	3.69
4th	280	261	-19	0.0382	2.57

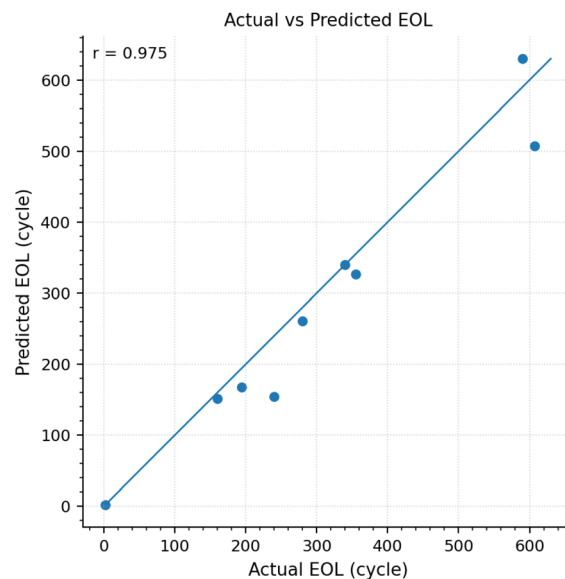
**Fig. 7** Actual vs. predicted EOL.

Fig. 5 shows that both models are able to capture the overall evolution of the ICA profiles; however, the hybrid linear + GPR approach demonstrates slightly more consistent reconstructions than the GPR-only model. As summarized in Table 5, the MAPE between the true ICA curves and the FPCA-based reconstructed ICA (using predicted scores) remains consistently low across all cycles.

To further demonstrate the competitiveness and robustness of the proposed methodology, a comparison is conducted against representative state-of-the-art studies from the literature focusing on early-life prediction of lithium-ion battery

Table 6 Comparison of end-of-life (EOL) prediction errors from recent literature and the proposed method

Paper	Methodology	Cycles used	EOL error
Online lifetime prediction for lithium-ion batteries ³⁵	Data-driven prediction with cycle-based updates and ensembling	200	6.1% (MAPE)
Early prediction of lithium-ion degradation trajectories ³⁶	Voltage curve feature extraction with regression-based prediction	100	10.0% (MAPE)
Aging trajectory and end-of-life prediction <i>via</i> similar fragment extraction ³⁷	Capacity degradation trajectory analysis using fragment extraction	Not explicitly specified	3.23% (abs. rel. error)
Lithium-ion battery end of life prediction based on the decelerating aging point ³⁸	Feature-based prediction using degradation characteristics	Not explicitly specified	8.7% (MAPE, max)
Proposed FPCA + linear + GPR method	ICA-based FPCA feature extraction with hybrid regression	33% of cycles	4.95% (EOL error)



degradation. As shown in Table 6, the proposed hybrid FPCA-based framework achieves competitive performance in both SOH estimation and EOL prediction, while maintaining interpretability and reduced model complexity compared to deep learning approaches.

It should be noted that the reported EOL error metrics across the literature are not fully standardized, with some studies reporting mean absolute percentage error (MAPE), while others report absolute relative error or maximum error. Despite these differences, the comparison provides a general benchmark for evaluating prediction accuracy.

Although certain methods in the literature report lower EOL prediction errors, such as,³⁷ these approaches often rely on a larger or unspecified portion of the degradation trajectory. In contrast, the proposed method achieves competitive accuracy while utilizing only early-life data (33% of cycles), highlighting its effectiveness for early prediction applications.

Overall, these findings indicate that early-cycle data can be used to accurately predict future battery health and lifespan. The proposed approach offers a practical and data-efficient solution for real-time health forecasting, reducing the need for extensive long-term testing while maintaining a high level of predictive reliability.

5 Conclusion

This work demonstrates that integrating Incremental Capacity Analysis (ICA) with Functional Principal Component Analysis (FPCA), combined with a multi-stage framework of linear regression and Gaussian Process Regression (GPR), provides a robust and accurate approach for battery State of Health (SoH) estimation and end of life (EOL) prediction. The proposed model achieved a mean absolute percentage error (MAPE) of 0.89% and a coefficient of determination ($R^2 = 0.926$) on laboratory test data. Furthermore, validation using the NASA dataset confirmed the model's generalization capability, successfully predicting SoH with a MAPE of 2.99% and an average end-of-life (EOL) prediction error of 4.95%. These results highlight the effectiveness of the ICA-FPCA-GPR framework as a reliable tool for advanced battery prognostics.

One limitation of this work is that it is developed and validated on a relatively limited dataset, which may not fully represent the variability seen across larger battery populations and real operating conditions. In addition, the framework relies on full charge-discharge data, whereas in practical EV applications, batteries are often only partially cycled, making such data difficult to obtain. This could limit how directly the approach can be used in real systems. Finally, although the method performs well under controlled conditions, applying it within real-world battery management systems may present challenges due to measurement noise, changing operating conditions, and practical implementation constraints.

Future work will focus on extending the framework to operate effectively with partial charge-discharge data, enabling its use in more realistic scenarios where full cycling is not

available. This will make the approach more applicable to real-world battery management systems, particularly in EVs where operating conditions are highly dynamic. In parallel, a hybrid fPCA framework will be developed by integrating a weighted covariance matrix, allowing greater emphasis on voltage regions more strongly correlated with capacity degradation. This enhancement is expected to improve the model's ability to capture critical degradation signatures and enhance the accuracy of SoH prediction. Further validation will also be carried out using real-world EV datasets to assess the robustness and practical applicability of the proposed approach.

Conflicts of interest

The authors declare no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

The authors sincerely thank the funding support from the Natural Sciences and Engineering Research Council of Canada.

References

- 1 International Energy Agency, *Global EV Outlook 2025*, IEA, Paris, 2025. Licence: CC BY 4.0. Available from: <https://www.iea.org/reports/global-ev-outlook-2025>.
- 2 S. Brinley, Market share of EVs rises more quickly in Canada than in US; 2024. Accessed: YYYY-MM-DD. Blog post, S&P Global Mobility. Available from: <https://www.spglobal.com/mobility/en/research-analysis/electric-vehicle-growth-canada-usa.html>.
- 3 CBC News. EV sales in China: Global growth driven by strong demand, says IEA. CBC News. 2025; accessed: YYYY-MM-DD. Available from: <https://www.cbc.ca/news/science/ev-sales-china-iea-1.7535190>.
- 4 K. R. Ngoy, V. T. Lukong, K. O. Yoro, J. B. Makambo, N. C. Chukwuati, C. Ibegbulam, *et al.* Lithium-ion batteries and the future of sustainable energy: A comprehensive review, *Renewable Sustainable Energy Rev.*, 2025, **223**, 115971. Available from: <https://www.sciencedirect.com/science/article/pii/S1364032125006446>.
- 5 S. Rout, S. K. Samal, D. J. Gelmecha and S. Mishra, Estimation of state of health for lithium-ion batteries using advanced data-driven techniques, *Sci. Rep.*, 2025, **15**(1), 30438, DOI: [10.1038/s41598-025-93775-y](https://doi.org/10.1038/s41598-025-93775-y).
- 6 Q. Li, R. Song and Y. Wei, A review of state-of-health estimation for lithium-ion battery packs, *J. Energy Storage*,



- 2025, **118**, 116078. Available from: <https://www.sciencedirect.com/science/article/pii/S2352152X25007911>.
- 7 Q. Yu, Y. Nie, S. Liu, J. Li and A. Tang, State of health estimation method for lithium-ion batteries based on multiple dynamic operating conditions, *J. Power Sources*, 2023, **582**, 233541. Available from: <https://www.sciencedirect.com/science/article/pii/S0378775323009175>.
 - 8 H. Thomas and M. H. Weatherspoon, Capacity and State-of-Health Prediction of Lithium-Ion Batteries Using Reduced Equivalent Circuit Models, *Batteries*, 2025, **11**(4), 162. Available from: <https://www.mdpi.com/2313-0105/11/4/162>.
 - 9 F. Guo, L. D. Couto, G. Mulder, K. Trad, G. Hu, O. Capron, *et al.*, A systematic review of electrochemical model-based lithium-ion battery state estimation in battery management systems, *J. Energy Storage*, 2024, **101**, 113850. Available from: <https://www.sciencedirect.com/science/article/pii/S2352152X24034364>.
 - 10 J. Sun, Y. Liu and J. Kainz, A comparative study of parameter identification methods for equivalent circuit models for lithium-ion batteries and their application to state of health estimation, *J. Energy Storage*, 2025, **114**, 115707. Available from: <https://www.sciencedirect.com/science/article/pii/S2352152X25004207>.
 - 11 J. S. Edge, S. O'Kane, R. Prosser, N. P. Brandon, S. J. Harris and G. J. Offer, Lithium ion battery degradation: what you need to know, *Phys. Chem. Chem. Phys.*, 2021, **23**(14), 8200–8221.
 - 12 K. Q. Zhou, Y. Qin and C. Yuen, Graph neural network-based lithium-ion battery state of health estimation using partial discharging curve, *J. Energy Storage*, 2024, **100**, 113502. Available from: <https://www.sciencedirect.com/science/article/pii/S2352152X24030883>.
 - 13 K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, *et al.*, Data-driven prediction of battery cycle life before capacity degradation, *Nat. Energy*, 2019, **4**(5), 383–391.
 - 14 P. Wen, Z. S. Ye, Y. Li, S. Chen, P. Xie and S. Zhao, Physics-Informed Neural Networks for Prognostics and Health Management of Lithium-Ion Batteries, *IEEE Trans. Intell. Veh.*, 2024, **9**(1), 2276–2289.
 - 15 L. Yang, M. He, Y. Ren, B. Gao and H. Qi, Physics-informed neural network for co-estimation of state of health, remaining useful life, and short-term degradation path in Lithium-ion batteries, *Appl. Energy*, 2025, **398**, 126427. Available from: <https://www.sciencedirect.com/science/article/pii/S0306261925011572>.
 - 16 M. Aykol, C. B. Gopal, A. Anapolsky, P. K. Herring, B. van Vlijmen, M. D. Berliner, *et al.*, Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime, *J. Electrochem. Soc.*, 2021, **168**(3), 030525.
 - 17 R. G. Nascimento, M. Corbetta, C. S. Kulkarni and F. A. C. Viana, Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis, *J. Power Sources*, 2021, **513**, 230526.
 - 18 S. Kohtz, Y. Xu, Z. Zheng and P. Wang, Physics-informed machine learning model for battery state of health prognostics using partial charging segments, *Mech. Syst. Signal Process.*, 2022, **172**, 109002.
 - 19 F. Wang, Z. Zhai, Z. Zhao, Y. Di and X. Chen, Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis, *Nat. Commun.*, 2024, **15**(1), 4332.
 - 20 A. Thelen, X. Huan, N. H. Paulson, S. Onori, Z. Hu and C. Hu, Probabilistic Machine Learning for Battery Health Diagnostics and Prognostics—Review and Perspectives, *npj Mater. Sustainability*, 2024, **2**(1), 14.
 - 21 Q. Zou and J. Wen, Battery State-of-Health Estimation Incorporating Model Uncertainty Based on Bayesian Model Averaging, *Energy*, 2024, **308**, 132884.
 - 22 R. R. Richardson, M. A. Osborne and D. A. Howey, Battery Health Prediction Under Generalized Conditions Using a Gaussian Process Transition Model, *J. Energy Storage*, 2019, **23**, 320–328.
 - 23 D. Anseán, V. M. Garcia, M. Gonzalez, C. Blanco, J. C. Viera, Y. Fernández Pulido, *et al.*, Lithium-Ion Battery Degradation Indicators Via Incremental Capacity Analysis, *IEEE Trans. Ind. Appl.*, 2019, **55**(3), 2992–3002.
 - 24 Science TD., Functional Principal Component Analysis and Functional Data, Towards Data Science, 2019, accessed: 2025-10-06.
 - 25 R. Viviani, G. Grön and M. Spitzer, Functional principal component analysis of fMRI data, *Hum. Brain Mapp.*, 2005, **24**(2), 109–129.
 - 26 Y. Cheng, C. Lu, T. Li and L. Tao, Residual lifetime prediction for lithium-ion battery based on functional principal component analysis and Bayesian approach, *Energy*, 2015, **90**, 1983–1993.
 - 27 M. Shoriat Ullah and K. Seo, Prediction of Lithium-Ion Battery Capacity by Functional Principal Component Analysis of Monitoring Data, *Appl. Sci.*, 2022, **12**(9), 4296.
 - 28 N. Li, M. Wang, Y. Lei, Z. Lu, X. Tan, F. Yang, *et al.*, Remaining useful life prediction of lithium-ion battery with nonparametric degradation modeling and incomplete data, *Reliab. Eng. Syst. Saf.*, 2025, **256**, 110721.
 - 29 M. Duquesnoy, *et al.*, Functional data-driven framework for fast forecasting of electrode slurry rheology simulated by molecular dynamics, *npj Comput. Mater.*, 2022, **8**, 161.
 - 30 Y. Tang, *et al.*, Remaining useful life prediction of high-capacity lithium-ion batteries based on incremental capacity analysis and Gaussian process regression, *Sci. Rep.*, 2024, **14**, 23524.
 - 31 R. Richardson, C. Birkl, M. Osborne and D. Howey, Gaussian Process Regression for *In situ* Capacity Estimation of Lithium-Ion Batteries, *IEEE Trans. Ind. Inf.*, 2019, **15**(1), 127–138.
 - 32 S. Park, J. Lee and S. Heo, Gaussian Process Regression-Based Lithium-Ion Battery End-of-Life Prediction Model under Various Operating Conditions, *arXiv*, 2024, preprint, arXiv:241019886, <https://arxiv.org/abs/2410.19886>.
 - 33 X. Feng, C. Weng, X. He, L. Wang, D. Ren, L. Lu, *et al.*, Incremental Capacity Analysis on Commercial Lithium-Ion Batteries using Support Vector Regression: A Parametric



- Study, *Energies*, 2018, **11**(9), 2323, available from: <https://www.mdpi.com/1996-1073/11/9/2323>.
- 34 K. Fricke, R. G. Nascimento and F. A. C. Viana, An accelerated Life Testing Dataset for Lithium-Ion Batteries with Constant and Variable Loading Conditions, *Int. J. Progn. Health Manag.*, 2023, **14**, DOI: [10.36001/ijphm.2023.v14i2.3587](https://doi.org/10.36001/ijphm.2023.v14i2.3587).
- 35 C. Strange, R. Ibraheem and G. dos Reis, Online Lifetime Prediction for Lithium-Ion Batteries with Cycle-by-Cycle Updates, Variance Reduction, and Model Ensembling, *Appl. Energy*, 2023, **334**, 120680.
- 36 R. Ibraheem, C. Strange and G. dos Reis, Early Prediction of Lithium-Ion Cell Degradation Trajectories Using Signatures of Voltage Curves up to 4-Minute Sub-Sampling Rates, *Appl. Energy*, 2023, **344**, 121215.
- 37 S. Zhang, *et al.*, Aging Trajectory and End-of-Life Prediction for Lithium-Ion Battery via Similar Fragment Extraction of Capacity Degradation Curves, *J. Cleaner Prod.*, 2024, **433**, 140119.
- 38 X. Zhu, *et al.*, Lithium-ion battery end of life prediction based on the decelerating aging point, *Appl. Energy*, 2025, **368**, 121875.

