Reliable GPR-based SOH Estimation using Partial Charging Data

1st Kibum Cheon
Dept. of Mechanical Engineering
Chungbuk National University
Cheongju, Korea
cjsrlqja99@chungbuk.ac.kr

2nd Haejun Kim
Dept. of Mechanical Engineering
Chungbuk National University
Cheongju, Korea
skh1111@chungbuk.ac.kr

3rd Jongho Shin
Dept. of Mechanical Engineering
Chungbuk National University
Cheongju, Korea
jshin@chungbuk.ac.kr

***Abstract*—This paper proposes a robust methodology for estimating the State of Health (SOH) of Lithium-ion batteries(LIBs) using Gaussian Process Regression (GPR). "To overcome the limitation of prior studies that rely on full charge–discharge cycles, we leverage partial-charging data, motivated by the fact that partial charging is commonplace in real-world application. To capture electrochemical degradation, we add Incremental Capacity Analysis (ICA) peak features and tune kernel hyperparameters via Bayesian optimization for calibrated uncertainty. Using Samsung INR 18650-25R cells, the proposed model achieves an RMSE of 1.6347% and an EICP of 94.34% (nominal 95%), outperforming baselines without ICA or optimization. Results show that ICA features improve accuracy and Bayesian optimization corrects over-conservative intervals, yielding reliable uncertainty alongside strong point estimates.**

***Index Terms*—Lithium-Ion Battery, SOH Estimation, Bayesian Optimization, Incremental Capacity Analysis, Gaussian Process Regression**

I. Introduction

Amid the shift from fossil fuels to sustainable energy systems, Lithium-ion batteries (LIBs) have become a core enabler across sectors—powering electric vehicles, supporting grid-scale storage that smooths intermittent renewables, and driving portable electronics [1], [2]. Their pervasive role makes LIBs a foundational technology in today's energy transition.

As battery reliance deepens, management technologies that ensure safety, lifetime, and performance have become indispensable. Because cell condition directly affects system reliability and operational integrity, real-time assessment and accurate SOH estimation are required [3], [4]. The SOH is a key indicator that quantitatively represents the degree of a battery's aging. It is defined as the ratio of the maximum capacity a battery can currently store to its initial nominal capacity, expressed as a percentage, by the following equation.

$$\mathrm{SOH}(\%) = \frac{Q_{\text{current}}}{Q_{\text{nominal}}} \times 100 \tag{1}$$

This research was supported by Unmanned Vehicles Core Technology Research and Development Program through the National Research Foundation of Korea (NRF), Unmanned Vehicle Advanced Research Center (UVARC) funded by the Ministry of Science and ICT, Republic of Korea (2020M3C1C1A01083162), and by the Regional Innovation System & Education(RISE) program through the (Chungbuk Regional Innovation System & Education Center), funded by the Ministry of Education(MOE) and the (Chungcheongbuk-do), Republic of Korea(2025-RISE-11-014-03).

Here, Q_{current} denotes the maximum discharge capacity measured in the current cycle and Q_{nominal} the initial nominal capacity at manufacture, so SOH serves as an intuitive indicator of the battery's remaining lifespan.

Research on SOH estimation is broadly classified into circuit-based models, electrochemical models. Circuit-based approaches, such as the Equivalent Circuit Model (ECM), represent battery behavior using combinations of electrical components, while physics-based Electrochemical Models (EMs) mathematically describe the internal electrochemical dynamics. ECMs are simple and suitable for real-time applications but suffer from limited accuracy. In contrast, EMs offer higher fidelity at the cost of severe model complexity and computational burden, which hinders their deployment in practical Battery Management Systems [5], [6].

As an alternative, data-driven machine learning methods have been actively studied. Models ranging from SVMs and ANN to Recurrent Neural Networks can achieve high estimation accuracy by directly learning degradation patterns from data [5], [7], [8]. However, most of these approaches assume access to full-cycle data and suffer from the black-box problem: the internal reasoning process is opaque, making it difficult to guarantee the reliability of the predictions [9], [10].

To mitigate these limitations, this study proposes a Gaussian Process Regression (GPR)–based SOH estimation model. GPR produces both point predictions and quantitative uncertainty, providing a basis for assessing the credibility of the estimates and partially alleviating the black-box nature of conventional deep learning models. Since the reliability of this uncertainty crucially depends on the kernel hyperparameters [11], we employ Bayesian optimization to systematically tune them and obtain well-calibrated predictive distributions.

Most prior work relies on full charge–discharge data for SOH estimation, which is difficult to obtain under realistic operating conditions. In contrast, we use readily collectible partial-charging data and their statistical descriptors as inputs. To further enhance estimation performance, we additionally incorporate ICA peak features, which are sensitive to changes in the battery's electrochemical state.

In this work, we validate the proposed methodology by comparing three models: (1) a baseline GPR model using only statistical features, (2) a GPR model augmented with ICA peak

features, and (3) the final proposed model in which the kernel hyperparameters are tuned via Bayesian optimization.

The remainder of this paper is organized as follows. Section II describes the experimental dataset, preprocessing procedures, and feature extraction from voltage–current data. Section III introduces the structure of the GPR-based SOH estimation model and details the Bayesian hyperparameter optimization process. Section IV presents and analyzes the training and validation results through comparison of the three models. Finally, Section V concludes the paper.

II. Data for LIB SOH Estimation

This section details the preprocessing procedures for extracting features from experimental data and preparing them in a suitable format for the SOH estimation model.

A. Battery Dataset and Experimental Environment

The dataset used in this study was constructed through cycling experiments on five Samsung INR 18650-25R Li-ion battery cells. The nominal capacity and voltage of the battery are 2.5Ah and 3.7V, respectively. All experiments were conducted in a thermo-hygrostatic chamber maintained at a temperature of 25°C and 60% humidity. The cycling protocol consisted of a 0.5C Constant Current-Constant Voltage (CC-CV) charge and a 0.5C Constant Current (CC) discharge, repeated 957 times. The time-series data of voltage and current measured during each cycle were used for modeling.

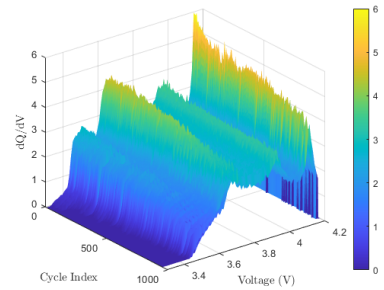
B. Input Data Preprocessing and Feature Extraction

The following preprocessing steps were performed to extract input features that effectively reflect SOH changes from the collected raw time-series data. First, the analysis was confined to the CC charging phase of the full charging profile, where voltage behavior changes due to SOH are prominent.

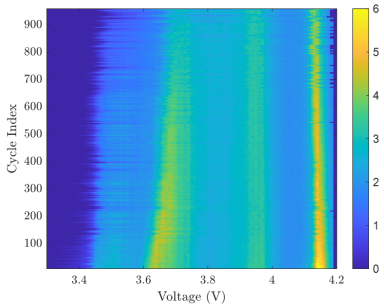
To emulate the partial charging patterns of real world usage, the CC charging section was divided into multiple segments, termed partial charging data in this study. Specifically, a sliding window approach was applied to the voltage range of 3.60V to 4.19V. A voltage window with a width of 0.4V was moved at 0.01V intervals, generating a total of 20 partial charging data segments from a single full charge cycle. This structured data serves as the fundamental unit for extracting the statistical and physical features used for model training [12].

C. Configuration of LIB SOH Estimation Data

This section describes the process of constructing the final input feature vector and output data for the model from the previously generated partial charging data. The input feature vector consists of key indicators that best describe the current SOH of the Li-ion battery. In this study, the final input vector was designed by extracting two groups of features: statistical features of charge capacity and ICA peak-related features.



(a) ICA Surface Plot



(b) ICA Heatmap

Fig. 1: Visualization of ICA curve evolution with surface plot and heatmap

1) Statistical Features of Partial Charge Capacity (ΔQ): The first feature group uses statistical properties of partial charge capacity, calculated by integrating current over time for each partial charge segment.

$$\Delta Q = \int_{t_{start}}^{t_{end}} I(t)dt \tag{2}$$

Here, t_{start} and t_{end} represent the start and end times of each segment. As the battery degrades, its internal resistance increases and effective capacity decreases, altering the time and charge (ΔQ) required to charge through the same voltage window (ΔV). To capture this pattern, two statistical features were extracted from the ΔQ values of all segments within a single charge cycle: the mean of partial charge capacity ($\mu_{\Delta Q}$), representing the overall charging characteristic, and the standard deviation ($\sigma_{\Delta Q}$), indicating the uniformity of capacity change across different voltage ranges.

2) Incremental Capacity Analysis based Peak Features: The second feature group utilizes peak information extracted via the Incremental Capacity Analysis (ICA) technique. ICA is an effective method for diagnosing the internal electrochemical state and degradation mechanisms of a battery by analyzing the derivative of the charge curve ($\frac{dQ}{dV}$ versus voltage V) [13].

Under Constant Current (CC) conditions, the terminal voltage V of a battery is generally expressed as:

$$V(Q) = E_{eq}(Q) + I \times R(Q) + \eta(Q, I, T) \tag{3}$$

where E_{eq} is the equilibrium voltage, R is the internal resistance, and η is the overpotential. The ICA curve, represented

by $\frac{dQ}{dV}$, forms peaks in regions where phase transitions, such as lithium-ion intercalation and de-intercalation, occur actively.

As the battery degrades, the shape of the ICA curve changes, with the peak's amplitude and position showing a strong correlation with degradation. Generally, as SOH decreases, the peak amplitude decreases, and its position shifts to a higher voltage. This amplitude decrease is visible in Fig. 1(a), and the position shift is observed as the distinctly colored region moves with cycling in Fig. 1(b). In this study, the peak value P_j and peak position V_j for each partial charging data segment j were extracted as features, defined as follows, where i is the discrete data sample index within the segment.

$$P_j = \max_i \frac{\Delta Q_{j,i}}{\Delta V_{j,i}}, \quad V_j = \arg\max_i \frac{\Delta Q_{j,i}}{\Delta V_{j,i}} \tag{4}$$

The primary cause of peak amplitude decrease is the Loss of Active Material and SEI growth, which reduces the available lithium-ions to react, consequently lowering the $\frac{dQ}{dV}$ value [14].

The rightward shift is mainly due to increased internal resistance (R). This requires a higher terminal voltage V to drive the same current I, shifting the phase transition voltage range upwards [15]–[17].

In this study, the peak value (P_j) and position (V_j), which reflect this physical degradation, were used as key input features. If multiple peaks existed within a segment, the one with the largest amplitude was selected.

3) Final Input Vector Construction: The four features described above—mean partial charge capacity ($\mu_{\Delta Q}$), standard deviation of partial charge capacity ($\sigma_{\Delta Q}$), ICA peak value (P), and peak position (V)—are integrated to form the final 4-dimensional input feature vector. The dataset for model training and validation was constructed by randomly sampling data from the entire lifecycle of the battery cells.

D. Output Data Definition

The output data for the model developed in this study is the SOH. SOH is a key indicator that intuitively represents the degree of aging and remaining lifespan based on the current available capacity, making it one of the most important parameters in a BMS.

Generally, when SOH drops below 80% of the initial capacity, it is considered the End of Life (EOL), after which battery replacement is recommended. Therefore, the SOH value estimated by this model aims to provide users with crucial information for making decisions about battery replacement timing and safety management.

III. GPR-based LIB SOH Estimation Model

This section describes the structure of the GPR model for SOH estimation and the Bayesian optimization-based method for setting its hyperparameters.

A. GPR-based Learning Model

GPR is a nonparametric method that models the relationship between inputs and outputs as a distribution over functions. It can effectively capture complex nonlinearities and provides quantitative uncertainty along with its estimates, allowing for an assessment of the result's reliability [11], [18].

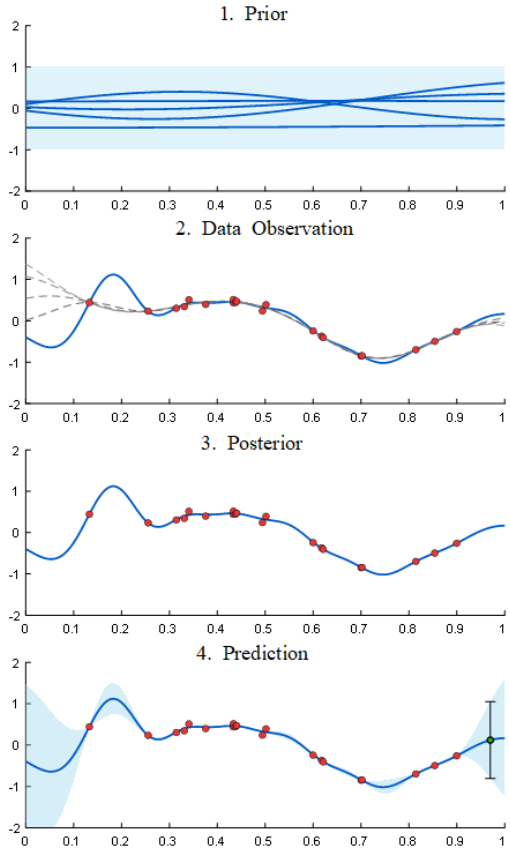


Fig. 2: Gaussian Process Regression Workflow

Fig. 2 illustrates the GPR inference process. It begins by assuming a set of possible functions from a prior distribution defined by a kernel function. This distribution is then updated based on observed data to form a posterior distribution that is consistent with the data. Finally, predictions, including an estimate and a confidence interval, are made for new inputs using this posterior. The estimated mean $\bar{f}_*$ and covariance $\text{cov}(f_*)$ are derived as follows.

$$\bar{f}_* = K(x_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \tag{5}$$

$$\text{cov}(f_*) = K(x_*, x_*) - K_*[K + \sigma_n^2 I]^{-1} K_*^T \tag{6}$$

Here, $K_* = K(x_*, X)$ and $K = K(X, X)$. X and y are the training inputs and outputs, respectively, K is the covariance matrix defined by the kernel function, and σ_n^2 is the noise variance. The estimated mean serves as the point estimate for SOH, while the estimated covariance is used to calculate the confidence interval, quantifying the estimation uncertainty. In this study, the Matérn 5/2 kernel function was used to effectively reflect the local non-linearity of the data.

B. Hyperparameter Tuning using Bayesian Optimization

The performance of a GPR model is highly dependent on its hyperparameters, such as the length-scale and signal variance, which define the shape of the kernel. To find the optimal

combination that reflects the data's unique characteristics, this study employed Bayesian optimization. This technique is more efficient than grid search or random search, as it can approach the optimal solution with fewer iterations [19], [20].

Bayesian optimization sequentially locates the optimum by approximating the objective with a surrogate model and selecting the next evaluation via an acquisition function. In our case, the surrogate model is a Gaussian process regression over the hyperparameter space that models the loss $L(\theta)$ with mean $\mu(\theta)$ and variance $\sigma^2(\theta)$. The acquisition function leverages this information to select the most promising candidate for the next evaluation. In this study, we employ the Expected Improvement (EI) acquisition function. EI is designed to take larger values when the expected loss is lower (exploitation) and when the predictive uncertainty is higher (exploration), thereby effectively balancing the two strategies. This iterative process—updating the surrogate model and searching for the optimum—continues until convergence.

The objective function for Bayesian optimization—the loss $L(\theta)$—must reliably estimate the generalization error for a given hyperparameter setting θ. Accordingly, we set the final loss $L(\theta)$ to the cross-validated mean squared error (CVMSE), i.e., the arithmetic mean of the mean squared error (MSE) over five validation folds:

$$L(\theta) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{MSE}_{(k)}(\theta) \quad (\text{where } K = 5) \tag{7}$$

$$\mathrm{MSE}_{(k)}(\theta) = \frac{1}{n_k}\sum_{i\in\mathcal{V}_k} \big(y_i(\theta) - \hat{y}_i(\theta)\big)^2 \tag{8}$$

where $\mathcal{V}_k$ denotes the k-th validation fold and n_k its size. Bayesian optimization then searches for the hyperparameter combination that minimizes $L(\theta)$, thereby ensuring stable estimation performance on the validation set. To prevent overfitting and data leakage during the search, the test data used for final evaluation are excluded from the loss computation.

C. Model Performance Metrics

This study used Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Estimation Interval Coverage Probability (EICP) to comprehensively evaluate the proposed model's estimation accuracy and reliability.

1) Estimation Accuracy Metrics: We evaluate predictive accuracy using the mean absolute error (MAE) and root mean squared error (RMSE), defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{10}$$

where n is the total number of samples, y_i is the actual observed value, and $\hat{y}_i$ is the model's estimated value. MAE reports the mean absolute deviation and RMSE emphasizes larger errors due to squaring.

2) Estimation Uncertainty Metric: A unique advantage of GPR is its ability to provide quantitative uncertainty. This study used the 95% Confidence Interval (CI) and the Estimation Interval Coverage Probability (EICP) to evaluate the reliability of this uncertainty estimation.

The 95% CI is calculated using the GPR's estimated mean ($\bar{f}_*$) and standard deviation (σ_*) as follows:

$$95\%\ \mathrm{CI} = [\bar{f}_* - 1.96\sigma_*, \quad \bar{f}_* + 1.96\sigma_*] \tag{11}$$

EICP represents the proportion of samples in test data where the actual value (y_i) falls within the 95% CI. An EICP value close to the nominal CI of 95% indicates that the model represents its estimation uncertainty accurately.

IV. Model Training and Validation Results

This section describes the training of the GPR-based SOH estimation models and analyzes their results.

A. Experimental Setup

The SOH estimation model was developed using MATLAB and its Statistics and Machine Learning Toolbox. The dataset used in this study was collected through charge-discharge tests on five Samsung INR 18650-25R cells; data from four cells were used as training data, while data from one cell was used as validation data.

B. Model Training

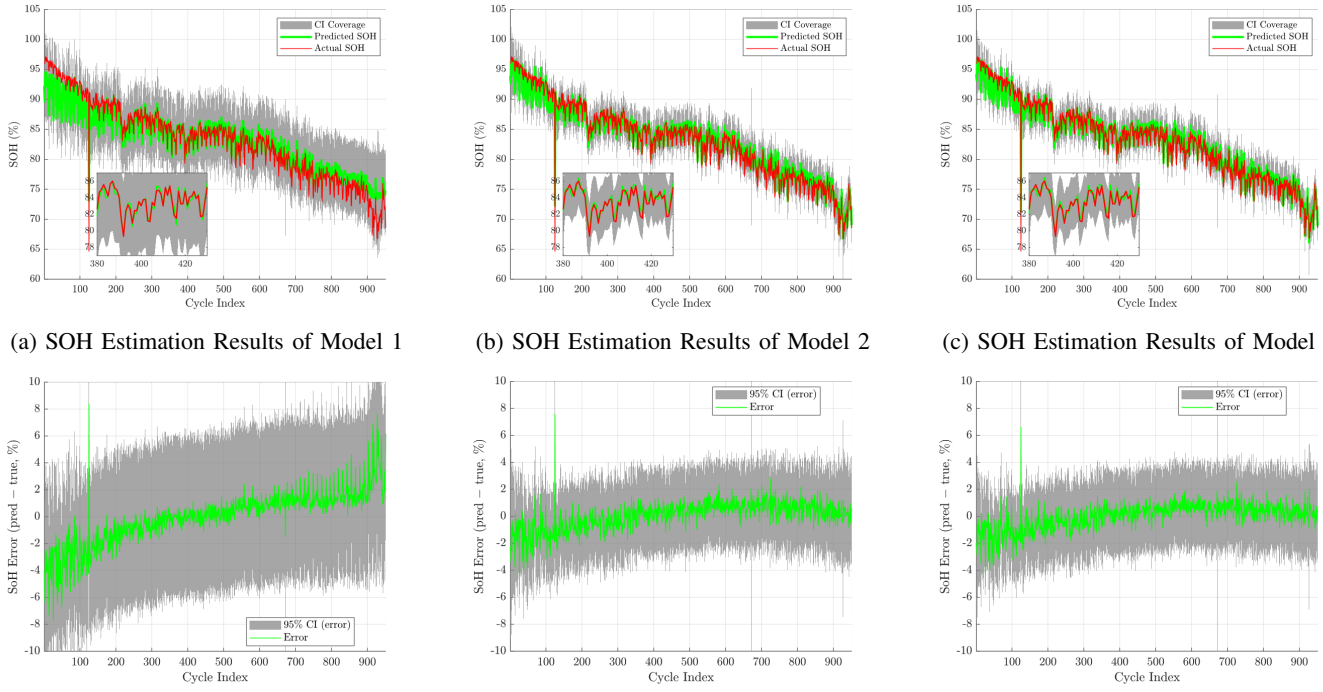
To validate the proposed SOH estimation methodology, three GPR models with different input data configurations and hyperparameter optimization methods were designed and compared.

1) Dataset Configuration: The dataset for model training and validation was constructed through the preprocessing steps described in Section II. From the data acquired from all battery cells, 2,000 samples were randomly extracted for model training. The entire lifecycle data of one specific cell, consisting of 957 samples, was used as the validation set. This was done to evaluate the model's generalization performance on unseen data and to prevent overfitting to the degradation patterns of specific cells.

2) Comparative Model Design: The specifications of the three models designed for the comparative experiment are as follows:

Model 1: Baseline GPR The first model is a Gaussian Process Regression (GPR) model whose hyperparameters are optimized via random search. The input features consist only of the statistical characteristics of the partial charging capacity ($\mu_{\Delta Q}, \sigma_{\Delta Q}$). The model uses the Matérn 5/2 kernel with heuristic initial values, and serves as a baseline to assess the effectiveness of the proposed input features and hyperparameter optimization strategies.

Model 2: GPR with ICA Features The second model extends the baseline GPR by incorporating ICA peak features (P, V) in addition to the statistical features, resulting in a total of four input features. As in Model 1, the hyperparameters are tuned using random search. This model is designed to verify whether



(a) SOH Estimation Results of Model 1 (b) SOH Estimation Results of Model 2 (c) SOH Estimation Results of Model 3

(d) SOH Error of Model 1 (e) SOH Error of Model 2 (f) SOH Error of Model 3

Fig. 3: Comparison of SOH Estimation Results and Relative Error Rates by Model

the ICA-based features provide tangible improvements in SOH estimation performance.

Model 3: GPR with Bayesian Optimization and ICA Features The third model is the final proposed model in this study. It uses the full set of four input features, including both statistical and ICA peak features, and applies the Bayesian optimization scheme described in Section 3 to optimize the kernel hyperparameters. The objective function is the 5-fold cross-validated mean squared error (CVMSE), so that the model is trained to enhance generalization performance

C. Model Validation

This section validates and compares the performance of the three trained GPR models using a separate validation dataset. The models' performance in terms of accuracy and reliability on the data is quantitatively evaluated based on MAE, RMSE, and 95% EICP, with the detailed results presented in Table I.

D. Results Analysis

This section quantitatively analyzes the effectiveness of the ICA features and the Bayesian optimization based on the validation results of the three GPR models.

TABLE I: SOH Estimation Models Performance Comparison

Model	Accuracy		Uncertainty
	MAE(%)	RMSE(%)	EICP(%)
Model 1	2.3278	3.1902	93.20
Model 2	1.2473	1.7426	93.72
Model 3 (Proposed)	**1.1393**	**1.6347**	**94.34**

In this section, the performance of the three GPR models is validated and compared. A separate validation dataset, which is not used during training, is employed to quantitatively evaluate how accurately and reliably each model estimates the SOH for unseen data. The models are evaluated in terms of MAE, RMSE, and 95% EICP, and the detailed quantitative performance indicators, computed on a partial charging data–wise basis, are summarized in Table I.

For visual comparison, Fig. 3 presents the prediction results aggregated at the cycle level. Since all partial charging data within the same cycle share an identical ground-truth SOH, the partial charging data-wise SOH estimations within each cycle are averaged, and this cycle-averaged value is plotted as a single representative point for that cycle.

E. Result Analysis

In this subsection, the validation results of the three GPR models are analyzed to quantitatively assess the effectiveness of the ICA-based features and the impact of Bayesian hyperparameter optimization.

To examine the contribution of ICA peak features, the performance of Model 1 and Model 2 is compared. As shown in Table I, Model 2 achieves reductions in MAE and RMSE of approximately 1.08% and 1.45% points, respectively, indicating a substantial improvement in estimation accuracy. This trend is also clearly observed in the error distributions in Fig. 3(d)–(e), where Model 2 exhibits a more concentrated and reduced error distribution compared to Model 1, which uses only statistical features. These results suggest that the

ICA peak features effectively capture local degradation characteristics of the battery.

The effect of Bayesian optimization is assessed by comparing Model 2 with the proposed Model 3. According to Table I, Model 3 attains the best overall accuracy, with an MAE of 1.1393% and an RMSE of 1.6347%, thereby achieving the most stable prediction performance among the three models.

From the perspective of uncertainty quantification, the final proposed Model 3 also shows the most reliable behavior. Its EICP is 94.34%, which is the closest to the nominal confidence level of 95% among all models, indicating that the predicted confidence intervals are well calibrated.

The results of the three models lead to two key conclusions. First, the ICA peak features play a decisive role in improving SOH estimation accuracy. Second, Bayesian optimization enhances the reliability of uncertainty estimation with high estimation accuracy. Therefore, Model 3, which incorporates both ICA-based features and Bayesian hyperparameter optimization, is confirmed to be the best-performing model in terms of both point estimation accuracy and calibrated uncertainty under partial charging conditions.

V. Conclusion

In this study, we proposed a GPR-based methodology to estimate the SOH of Li-ion batteries using partial charging data and ICA peak features. By focusing on the stable CC charging phase and training the model exclusively on partial charging segments, the proposed approach is well aligned with practical BMS operation. In addition, the kernel hyperparameters of the GPR model were tuned via Bayesian optimization to enhance generalization performance and uncertainty reliability.

Experimental results confirmed that incorporating ICA peak features improves SOH estimation accuracy compared to using only statistical descriptors of partial charge capacity. The final proposed model (Model 3) achieved the best performance with an MAE of 1.1393% and an RMSE of 1.6347%. In terms of uncertainty, Model 1 and Model 2, which determine hyperparameters via random search under the same evaluation budget, exhibited undercoverage with EICP values of 93.72% and 93.20%, respectively. In contrast, the Bayesian-optimized Model 3 attained an EICP of 94.34%, yielding empirical coverage that is closer to the nominal 95% level. This demonstrates that, unlike naive random sampling, Bayesian optimization exploits information from previous evaluations to concentrate on promising regions of the hyperparameter space and obtain more consistent estimates of the kernel and noise parameters.

Finally, we verified that accuracy and uncertainty calibration can be maintained even when the model is trained from randomly sampled partial charging data, which is advantageous in practical environments where full charge histories are difficult to obtain. Nonetheless, this study is limited to a single Li-ion cell type under fixed environmental conditions. Future work will extend the proposed framework to cell modules, and investigate additional features such as the temporal evolution of ICA peaks and higher-order distributional statistics to further improve robustness.

References

[1] J.-M. Tarascon and M. Armand, "Issues and challenges facing rechargeable lithium batteries," *nature*, vol. 414, no. 6861, pp. 359–367, 2001.

[2] M. Yilmaz and P. T. Krein, "Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles," *IEEE Transactions on Power Electronics*, vol. 28, no. 5, pp. 2151–2169, 2013.

[3] M. Hannan, M. Lipu, A. Hussain, and A. Mohamed, "A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations," *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 834–854, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032117306275

[4] H. Meng and Y. Li, "A review on prognostics and health management (phm) methods of lithium-ion batteries," *Renewable and Sustainable Energy Reviews*, vol. 116, Dec. 2019, publisher Copyright: © 2019.

[5] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche, "Critical review of state of health estimation methods of li-ion batteries for real applications," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 572–587, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032115013076

[6] Y. Wang, J. Tian, Z. Sun, L. Wang, R. Xu, M. Li, and Z. Chen, "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems," *Renewable and Sustainable Energy Reviews*, vol. 131, p. 110015, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032120303063

[7] Y.-F. Luo and K.-Y. Lu, "An online state of health estimation technique for lithium-ion battery using artificial neural network and linear interpolation," *Journal of Energy Storage*, vol. 52, p. 105062, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352152X22010647

[8] E. Chemali, P. J. Kollmeyer, M. Preindl, Y. Fahmy, and A. Emadi, "A convolutional neural network approach for estimation of li-ion battery state of health from charge profiles," *Energies*, vol. 15, no. 3, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/3/1185

[9] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," *Mathematical Problems in Engineering*, vol. 2015, no. 1, p. 793161, 2015.

[10] M. Zhang, D. Yang, J. Du, H. Sun, L. Li, L. Wang, and K. Wang, "A review of soh prediction of li-ion batteries based on data-driven algorithms," *Energies*, vol. 16, no. 7, 2023. [Online]. Available: https://www.mdpi.com/1996-1073/16/7/3167

[11] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[12] Z. Deng, X. Hu, P. Li, X. Lin, and X. Bian, "Data-driven battery state of health estimation based on random partial charging data," *IEEE Transactions on Power Electronics*, vol. 37, no. 5, pp. 5021–5031, 2022.

[13] D.-I. Stroe and E. Schaltz, "Lithium-ion battery state-of-health estimation using the incremental capacity analysis technique," *IEEE Transactions on Industry Applications*, vol. 56, no. 1, pp. 678–685, 2019.

[14] M. Dubarry, C. Truchot, and B. Y. Liaw, "Synthesize battery degradation modes via a diagnostic and prognostic model," *Journal of power sources*, vol. 219, pp. 204–216, 2012.

[15] Z. Wang, J. Ma, and L. Zhang, "State-of-health estimation for lithium-ion batteries based on the multi-island genetic algorithm and the gaussian process regression," *Ieee Access*, vol. 5, pp. 21 286–21 295, 2017.

[16] S. Duan, Z. Yu, J. Li, Z. Zhao, and H. Liu, "Rapid screening for retired batteries based on lithium-ion battery ic curve prediction," *World Electric Vehicle Journal*, vol. 15, no. 10, 2024. [Online]. Available: https://www.mdpi.com/2032-6653/15/10/451

[17] A. Fly and R. Chen, "Rate dependency of incremental capacity analysis (dq/dv) as a diagnostic tool for lithium-ion batteries," *Journal of Energy Storage*, vol. 29, p. 101329, 2020.

[18] L. Wu, X. Fu, and Y. Guan, "Review of the remaining useful life prognostics of vehicle lithium-ion batteries using data-driven methodologies," *Applied Sciences*, vol. 6, no. 6, 2016. [Online]. Available: https://www.mdpi.com/2076-3417/6/6/166

[19] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.

[20] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The journal of machine learning research*, vol. 13, no. 1, pp. 281–305, 2012.