



## Data-driven battery health prognosis with partial-discharge information

Zhao, Chunyang; Andersen, Peter Bach; Træholt, Chresten; Hashemi, Seyedmostafa

*Published in:*  
Journal of Energy Storage

*Link to article, DOI:*  
[10.1016/j.est.2023.107151](https://doi.org/10.1016/j.est.2023.107151)

*Publication date:*  
2023

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Zhao, C., Andersen, P. B., Træholt, C., & Hashemi, S. (2023). Data-driven battery health prognosis with partial-discharge information. *Journal of Energy Storage*, 65, Article 107151. <https://doi.org/10.1016/j.est.2023.107151>

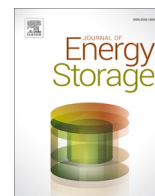
---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



## Research papers

## Data-driven battery health prognosis with partial-discharge information

Chunyang Zhao<sup>\*</sup>, Peter Bach Andersen, Chresten Træholt, Seyedmostafa Hashemi

Division for Power and Energy Systems, Department of Wind and Energy Systems, Technical University of Denmark, Kongens Lyngby 2800, Denmark

## ARTICLE INFO

## Keywords:

Battery degradation  
 Partial discharge  
 Data-driven model  
 State of health  
 Gaussian process regression

## ABSTRACT

The unpredictability of battery degradation behavior is a challenging issue impeding the development of battery applications, due to the complexity of the degradation and the limitation of state measurement methods. Nowadays, with accessible battery aging datasets and machine learning algorithms, there are opportunities for data-driven battery health prognosis. However, most of the previous work is restricted in the scope of full-discharge capacity records extrapolation, which has insufficient prospects in real-life applications. In this work, we propose using partial discharge information for degradation estimation and prediction. Our Gaussian process regression model achieves good performance by limited partial discharge information without requirements of feature selection. The accurate battery health prognosis in 300 cycles can be carried out by one partial-discharge cycle at any degradation stage. The capacity estimation gives around 1 % root mean square error (RMSE) when using 30 % information on the discharge process. As full-cycle discharge is not required, the proposed model can diagnose the battery state of health (SOH) with a limited portion of battery operation information extracted during the discharge process and reduce the effort of capacity tests. Further development of this method brings opportunities for battery state evaluation and prediction in real applications with better applicability and accuracy.

## 1. Introduction

As one of the key components of energy storage systems, the rechargeable battery plays an important role in promoting electrification and carbon neutrality [1]. Nowadays, the battery aging represents one of the main concerns of battery usage, and a substantial number of studies have been carried out to investigate the degradation mechanism and behavior [2]. The battery lifetime prognostic has been an emerging field of research that ranges from the investigation of fundamental chemical reactions, battery testing, and modeling, to the model implementations in real cases [3]. The degradation modes, such as loss of lithium inventory and loss of active electrode materials, are proposed to explain the power and capacity fade during battery cell usage [4]. Investigating the intrinsic causes, dozens of degradation mechanisms are recognized and demonstrated, including solid electrolyte interphase growth and decomposition, lithium plating, structural disordering, etc. Furthermore, accelerated battery aging tests are carried out to investigate the degradation performance related to the usage history directly [5].

The battery aging prognostic technologies are categorized as model-based, data-driven, and hybrid methods [6]. The model-based

approaches, such as physics-based models, equivalent circuit models, etc., reveal the degradation performance by the theoretical mechanisms [7]. As battery degradation is a complex multifactorial process, it is hard to be fully captured through model-based approaches. However, data-driven modeling is more flexible and adaptive, correlating the battery aging performance directly with the operation records. The difference between the empirical model and the data-driven model is minor, however, the latter emphasizes the improvement of the data size and mathematical methodologies. With the gradual enrichment of battery aging datasets, data-driven models are increasingly promising for accurate battery lifetime prognostic modeling [8].

The objective of battery lifetime prognostic is to estimate the aging development of battery performance over time [9]. In the scope of the data-driven approach, the battery aging datasets are mainly gathered by accelerated experimental tests and a few by real operation [10,11]. Multiple statistical metrics are used to evaluate the performance of the data-driven models. In the early research based on battery aging datasets from NASA, the artificial neural network and polynomial regression exhibit a larger error of remaining useful life (RUL) prediction compared to the particle filter [12,13]. With the aging dataset of 110 lithium cobalt oxide batteries, 0.28 % average error and 1.15 % standard deviation of

<sup>\*</sup> Corresponding author.

E-mail address: [chuzh@dtu.dk](mailto:chuzh@dtu.dk) (C. Zhao).

<https://doi.org/10.1016/j.est.2023.107151>

Received 19 December 2022; Received in revised form 17 February 2023; Accepted 14 March 2023

Available online 6 April 2023

2352-152X/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

state of health (SOH) prediction are obtained with the probabilistic neural network method [14]. Utilizing the aging dataset of three batteries under randomized duty profiles, multi-output Gaussian process regression models secured battery end-of-life (EOL) prediction within 5 days over the average EOL of around 150 days [15]. Using the empirical mode decomposition method, the long short-term memory model was combined with Gaussian process regression, which achieved 0.0032 Ah RMSE and 0.6 % maximum error [16]. Recently, the Lasso and elastic net regression are implemented for early prediction of cycle life before degradation, which achieves a 9.1 % test error using the information from the first 100 cycles [17]. Based on the previous observations of capacity records, the correlation between different cells is used to improve prediction accuracy [15]. Moreover, pattern analysis methods are used to further process the capacity degradation trajectory such as the empirical mode decomposition [16]. Deep learning has been used extensively in recent research works. For instance, the degradation trajectories are predicted by the model trained by 100 cycles with improved accuracy and calculation speed [18]. Integrating the initial capacity test and the further usage plan, the recurrent neural network (RNN) enables the aging prognostic for upcoming random cycles [19]. Limiting training information for deep neural networks, one-cycle information is used to achieve battery life prediction, and data-driven features are proved to be more informative [20].

However, the most of previous models are built on full-cycle capacity measurement, which limits the battery degradation research in the scope of the extrapolation of aging test results [21]. Therefore, exploring battery degradation modeling with less information, better adaptation, and longer projection is an imperative need for contemporary degradation research. There are a few recent works exploring the possibility of degradation prognosis by partial or limited data of battery operation. The differential voltage curve, which is the derivative of the capacity-voltage relationship, is used for SOH estimation [22]. The performance of the differential voltage curve at the different stages can be interpreted by degradation mechanisms, such as solid electrolyte interphase formation, therefore, partial charging or discharging can already be used to analyze the battery SOH. However, a very low current rate (C-rate) is required for high-quality differential voltage curve preparation, which is time-consuming and interrupts the normal operation of the battery. Based on the 15-minute partial charging test results of 2 lithium cobalt oxide batteries, online SOH diagnosis is achieved by the support vector machine for SOH estimation with a 2 % error [23]. The voltage, current, and temperature records extracted from the battery operation in selected SOC ranges of 15 % are used for SOH estimation, and 0.9 % of RMSE accuracy is achieved regardless of the difference in battery usage [24]. However, SOC is already a processed indicator by Column counting or voltage-SOC relation, which contains uncertainty along with the battery aging over usage and time. Using the measurement within partial voltage ranges is an alternative reliable way for data acquisition, while the previous works are normally based on the early battery aging datasets with short-life batteries, which does not give sufficient insight into the modern battery with long-term aging performance under complex duty profiles [25]. Recently, the deep Gaussian process regression is applied to the battery dataset with a cycle life of around 150 cycles, claiming the time-series records inputs can be used for partial discharge information without feature engineering [26]. The empirical degradation model of lithium cobalt oxide batteries up to 1000 cycles is proposed in [27]. However, instead of diagnosis directly from the battery partial cycle performance, the equivalent full cycles are used for battery degradation extrapolation.

There are two categories of battery health prognosis approaches: one is to extrapolate the SOH trajectory from previous cycles to the future, and another is to extract the features from the battery operation records, such as voltage, current, and temperature, to assess and predict the state of the batteries directly [6]. Previously, the signal processing techniques are implemented for noise filtering or mode decomposition on SOH trajectories, which improves the prediction accuracy [15,16]. From the

scope of degradation trajectories, it brings insights into the degradation speed at different stages, which are highlighted by knee point when the degradation speed has a significant change [28]. The knee point detection can be carried out from curve fitting techniques and relate to the degradation mechanism in the battery cell, and as an early alarm of severe degradation [29]. Most of the improvement for battery degradation modeling tries to predict the degradation trajectory from the early stage, which is why the features of the trajectory like knee points are important [6]. However, a versatile model to estimate and predict battery health at any stage is more practical in industrial applications than predicting multiple degradation indicators from the early beginning of battery life [30].

With the imperative need of improving the applicability of the degradation models, there are increasing amounts of research to model the battery degradation to further enhance the degradation model performance beyond accuracy, targeting less information, longer-term prediction, and higher calculation efficiency. For example, the first 100 cycles of battery tests are proven to have enough information for cycle life prediction with a 9.1 % testing error, and the classification can be done with only 5 cycle information [17]. Battery end-of-life and knee points in the cycle aging process are early predicted with the input of 100 cycles [18]. Recently, the hierarchical Bayesian model is used to predict the battery cycle life with only a 3-cycle battery testing protocol [31]. However, most of the existing studies extract the feature from the full-cycle testing results, which requires a dedicated testing condition. With the increasing adequacy and fidelity of battery aging data, looking into the detailed information inside each cycle to diagnose the battery degradation gives more insights than traditional capacity record extrapolation [32]. Instead of further exploiting the full-discharge cycles, using the partial information of the battery cycling process appears to be the opportunity to enhance the degradation research [33].

This work promotes the data-driven degradation model built with partial information on the discharge process and investigates how much information is needed for battery health prognosis. The partial information from the battery discharge process is used for present-cycle full-discharge capacity estimation noted as capacity estimation, and future-cycle full-discharge capacity prediction noted as capacity prediction. Instead of extrapolating the degradation records of a specific battery, the Gaussian process regression models are built by features extracted from the discharge process of a mixture of more than 100 battery cells under various duty profiles, and the feature extraction is applied at different aging stages. The proposed modeling framework gives better applicability and reproducibility for degradation prognosis and offers great potential for industrial application.

The paper is organized as follows, an overview of the recent development of battery degradation prognosis is given as an introduction, underlining the necessity of improving the model's applicability. In Section 2, the battery aging dataset is introduced, and the degradation behavior is demonstrated. In Section 3, The Gaussian process regression and other terminologies are defined for modeling and evaluation. The novel idea of using partial discharge information for battery degradation prognosis is proposed in Section 4, emphasizing the feature extraction design. The model performance of the capacity estimation and prediction is demonstrated in Section 5. The opportunities and limitations are discussed in Section 6. In the end, the paper concludes in Section 7.

## 2. Material

It is essential to equip with sufficient battery operation records for data-driven degradation prognosis. The battery dataset used in this work consists of 124 commercial lithium iron phosphate (LFP) batteries under fast-charging conditions from the previous work [17]. Different charge policies have been implemented on the A123 System APR18650M1A batteries, and a standard discharge process is carried out after each charging cycle for battery state characterization. The cycle number, internal resistance, temperature, charge time, charge capacity,

discharge capacity, etc. are recorded from the beginning to the EOL, which is from 100 % usable capacity to 80 %. Conventionally, “calendar life” and “cycle life” are used to indicate how much burden that battery could bear under designated operating conditions until the EOL. In the selected battery aging dataset, the batteries are under high C-rate full-cycle operation constantly, therefore, the cycle life is used to evaluate the battery health.

It requires a time-consuming process to test the battery aging performance. For instance, the battery capacity record of the first 10 cycles from the dataset is shown in Fig. 1. Each of the charging and discharging cycle takes around 60 min to finish, and the accumulated charging/discharging capacity of the battery is recorded, which is noted as the available capacity of that cycle. At the end of the capacity test, the trajectories of the remaining battery capacity with cycle count are obtained, as shown in Fig. 2. The batteries show diversiform patterns of the degradation trajectories in various operating conditions, including the initial capacity difference at an early stage, different deterioration speed, various EOL reaching 80 % of nominal capacity, etc., which are hard to track by the conventional curve-fitting approach. As the battery degradation is influenced by multiple parameters, the focus of contemporary battery testing is shifting from the generic full-cycle battery cycling test to specific usage of the battery, e.g., the C-rate, state of charge (SOC) range, stochasticity of battery usage, and so on [3].

On the left of Fig. 3, we detail the charging and discharging process of one cycle in the selected battery aging dataset. The battery is under the three-phase charging and two-phase discharging process with the combination of constant current (CC) and constant voltage (CV). For instance, the battery starts to be charged by 6C to 40 % capacity and by 3C to 80 % capacity. The battery finishes charging with CC: 1C and CV: 3.6 V. Since the original purpose of this aging dataset is to investigate the influence of the fast-charging process on battery degradation, various C-rate combinations of the first two phases are used, which are from 1C to 8C. The discharge process is controlled in the standard CC-CV process with 4C-2 V for every cycle of all the batteries tested, which gives stable capacity measurement for degradation reference. The battery cells are placed in the forced convection chambers with the temperature set to 30 °C, and the sensors are attached to the exposed cell to measure the temperature records as there is heat released during the operation.

As shown in Fig. 4, the capacity, temperature, and terminal voltage of the batteries are recorded constantly during the cycling process, which can be used for further modeling [17]. Instead of open circuit voltage, the terminal voltage is commonly used for degradation prognosis, which is easier to acquire and better represents the battery operating states in real applications [11,31,32]. Despite our research proposal to drive the degradation modeling toward the real application, the well-recorded aging test is a great resource for model building and validation. As we mentioned, the discharge capacity is commonly used for capacity record extrapolation in most of the prior works. It is

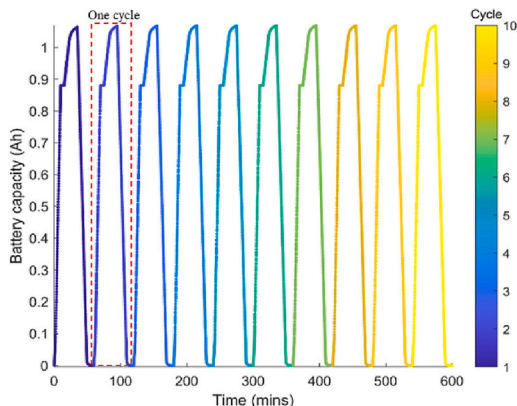


Fig. 1. Battery full-cycle capacity test: charging-discharging behavior.

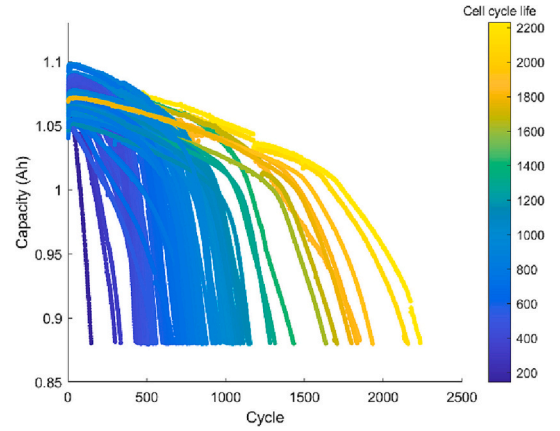


Fig. 2. Battery full-cycle capacity test: degradation trajectories.

normally a process using Coulomb counting integrating the active flowing current during the discharging process. However, the discharge capacity is a condensed value from a long process, the information during the discharging process is not used during the capacity record extrapolation approach, which wastes the great potential of the battery operating information. Another limitation of capacity record extrapolation is that it requires cycling the battery thoroughly in a standard process, which is hard to achieve during the provision of industrial applications. Furthermore, the dedicated capacity test for battery systems requires a long testing period interrupting the original system operation. In this work, by limiting the modeling inputs from full-cycle capacity measurement to partial-cycle discharging performance, the limited early information on discharging is explored for battery health prognosis, and elaborate aging testing datasets are used for model accuracy validation.

### 3. Methods

#### 3.1. Terminology of battery

Conventionally, the electric current is the primary way to describe battery usage. The C-rate is used to quantify and generalize the intensity of battery charging and discharge. The electric current which can fully discharge the battery in 1 h is defined as 1C, in other words, the C-rate is 1C. No matter whether the battery is charging or discharging, the battery is operating under XC when an electric current is numerically X time of 1C. For example, for a 1 Ah battery, if the charging current is 2A, the battery is operating under 2C. If the same battery is discharged under 0.5A, it is 0.5C. The C-rate is calculated by

$$C\text{-rate} = \frac{I}{I_{1C}} = I^* \frac{1 [h]}{Q_{\text{specification}} [Ah]} \quad (1)$$

where the  $I_{1C}$  represents the current when charging/discharging by 1C, and  $Q_{\text{specification}}$  represents the capacity of the battery in the specification. SOC to describe the present state of stored electric charge compared to the maximum storage capacity of electric charge, which equals the full-discharge capacity. It is a common approach to use Coulomb counting to integrate the current overtime for calculating the current state of stored electric charge. The SOC is calculated by

$$SOC [\%] = \frac{Q_{\text{present}} [Ah]}{Q_{\text{full}} [Ah]} = \frac{Q_0 + \int Idt}{Q_{\text{full}}} \quad (2)$$

where the  $Q_0$ ,  $Q_{\text{present}}$ , and  $Q_{\text{full}}$  are the battery capacity at the beginning of the usage, present state, and full-charge state, respectively.

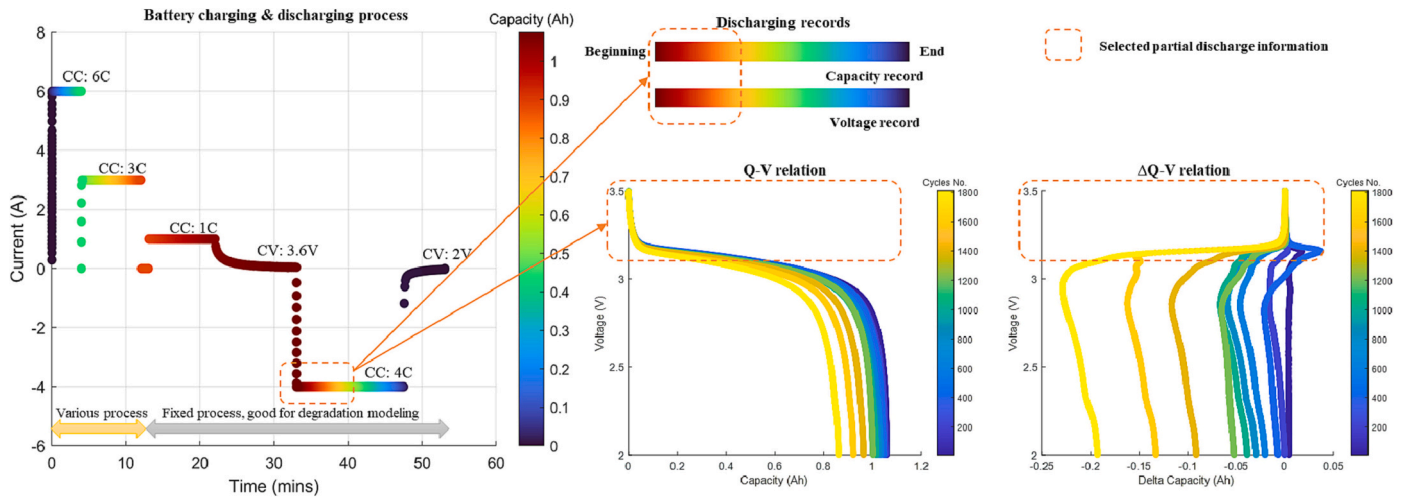


Fig. 3. Battery partial discharge information of the discharge process. The selected partial discharge information is a portion of early discharge information in the constant current discharge process, which is an array of capacity records.

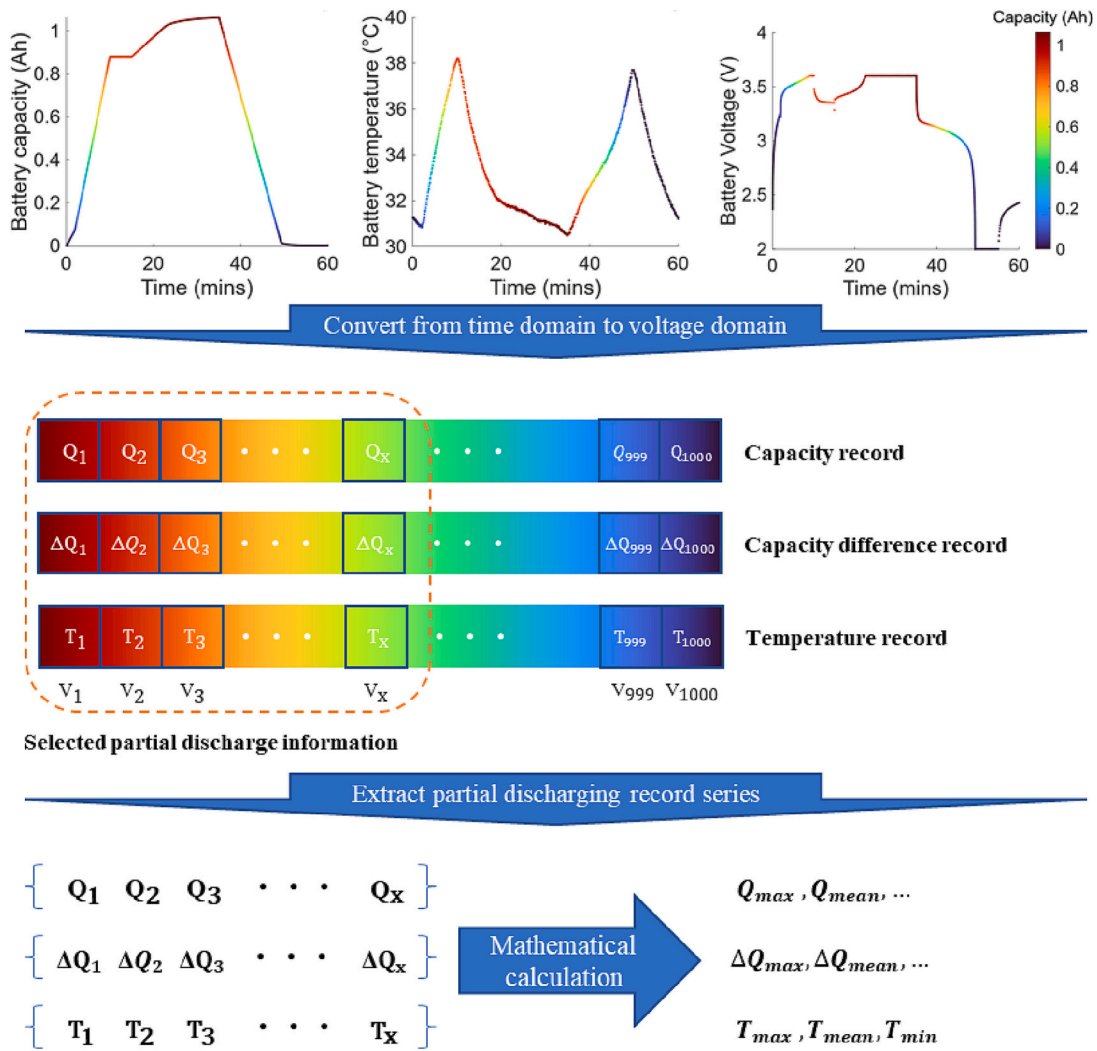


Fig. 4. Data extraction and feature processing for selected partial discharge information.

### 3.2. Gaussian process regression

Gaussian process regression is a nonparametric kernel-based probabilistic method, and it allows us to make predictions for data by incorporating prior knowledge [34]. The strengths of Gaussian process regression include the capability to estimate the mean and confidence of the regression by using the probabilistic approach and to achieve good regression performance with a limited amount of training data [35]. One prior assumption is the underlying model has the form of  $y = f(x) + \varepsilon$ , where  $f(x)$  represents the latent function and  $\varepsilon$  follows the Gaussian distribution, i.e.,  $\varepsilon \sim N(0, \sigma^2)$ . A Gaussian process is a set of random variables following a joint Gaussian distribution, which is defined by its mean function  $m(x)$ , and covariance function  $\kappa(x, x')$ , where

$$m(x) = E(f(x)) \quad (3)$$

$$\kappa(x, x') = E((f(x) - m(x))(f(x') - m(x'))^T) \quad (4)$$

where  $\sigma_l$  is the characteristic length scale and  $\sigma_f$  is a scale factor, which determines the average distance of the function away from its mean value. Similarly, the exponential kernel is defined as

$$\kappa_E(x, x') = \sigma_f^2 \exp\left(-\frac{\sqrt{(x-x')^T(x-x')}}{\sigma_l}\right) \quad (5)$$

where the  $\sqrt{(x-x')^T(x-x')}$  represents the Euclidean distance between  $x$  and  $x'$ , and the definition of  $\sigma_l$  and  $\sigma_f$  is the same as mentioned above.

With the training dataset extracted from battery discharge performance as the independent variable, predicting the current and further battery capacity falls into the scope of the regression problem. As introduced in the data selection part, the inputs of this regression model are a set of measurements and feature extraction result  $X$ , and the output is the battery capacity  $y$ . The training data is given as

$$X, y = \{(X_i, y_i)\}_{i=1}^N \quad (6)$$

where the  $(X_i, y_i)$  are the pairs of inputs ( $X_i \in \mathbb{R}^d$ ) and output ( $y \in \mathbb{R}$ ), and  $N$  is the number of training pairs. Conventionally, to predict the test dataset  $X^*$ , the  $m(x)$  is set to 0 for simplicity. By conditioning Gaussian distribution,

$$y^* | X^*, X, y \sim N(y^* | m^*, \sigma^*) \quad (7)$$

where  $m^* = K(X, X^*)^T K(X, X)^{-1} y$ , and  $\sigma^* = K(X^*, X^*) - K(X, X^*)^T K(X, X)^{-1} K(X, X^*)$ . As the result of maximizing  $p(y^* | X^*, X, y)$ ,

$$y^* = K(X, X^*)^T K(X, X)^{-1} y. \quad (8)$$

### 3.3. Model evaluation

The root-mean-square error (RMSE) is used to measure the performance of our model, comparing the difference between the predicted value with the measured value. It is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (9)$$

where the  $\hat{y}_i$  is the estimation and  $y_i$  is the real value, and  $N$  is the number of samples. In our case, the  $\hat{y}_i$  is the full-discharge capacity of the prediction, and the  $y_i$  is the measurement full-discharge value at the end of each cycle. Besides RMSE, the standard deviation with a similar definition is used for model evaluation. However, the RMSE is used to evaluate the results of the model prediction, comparing the capacity measurement with the prediction. The standard deviation is one of the outputs of the trained Gaussian process regression model, representing

the confidence of the prediction.

## 4. Partial information for degradation modeling

Conventional degradation prediction approaches exhibit reproducibility issues since the model is trained to predict the performance of the same battery cell, and sequential full-discharge capacity records are required as the input of the training dataset [36]. The battery health diagnosis approaches normally require special treatments like electrochemical impedance spectroscopy, pulse charging, very low C-rate operation, etc., which are time-consuming and difficult to be obtained in real applications [37]. Therefore, our model aims to assess the battery health across different usage with the information extracted from the normal operation of the high C-rate discharging process without selection, which saves time and experimental effort significantly.

Since the time length of discharging varies during the different stages of the battery degradation, converting the records to time-independent voltage-capacity relation facilitates training and stabilizes the model performance. Along with the battery test, various battery state parameters are recorded as time series, including current, voltage, and temperature. Since the time length of the battery discharging process is getting shorter because of the capacity degradation, the data length of degradation is decreasing, which creates inconsistency for modeling. Therefore, using the time-independent records by pairing the voltage and discharge capacity is superior.

Our novel idea of using partial information on the battery degradation prognosis is depicted in Figs. 3 and 4. On the left of Fig. 3, the partial early discharge information is taken from the early stage of the discharge record, specifically, when each time the battery discharges from 100 % SOC to a certain level. Similar approaches can be implemented in the charging process and other stages of the battery cycle to adapt the battery usage to ensure a consistent usage profile for aging investigation. As the selected dataset gives adequate discharging information for every cycle from 100 % SOC to the end of the CC discharging process, the partial discharge information is extracted here.

On the right side of Fig. 3, the development of capacity and capacity difference with voltage along the battery aging process is shown with a resolution of 200 cycles. Capacity-voltage relation (Q-V relation) in the constant current discharge process shows the discharge capacity accumulation when the battery discharges from 3.5 V to 2 V. The corresponding cycle number during the battery aging process is indicated by the color. At the end of each discharge process, the accumulated full-cycle capacity decreases along with the battery aging, as shown at the crossing point at the x-axis when the voltage equals 2 V. Along with the battery degradation, the crossing point at the x-axis moves left, and the area under the voltage-capacity curve shrinks. Most of the full-discharge capacity is developing from 1.08 Ah to 0.88 Ah, which is around 80 % of the initial capacity.

The capacity difference-voltage relation ( $\Delta Q$ -V relation) is calculated by subtracting the first-cycle capacity record from the present cycle. It shows the difference in accumulated discharging capacity at different voltage stages along with the aging process, addressing the inherent disparity of the battery cells. At the beginning of the discharging process starting from 3.5 V, there is no significant change in the  $\Delta Q$ -V relation. The discrepancy appears around 3.1 V, where the smaller capacity difference is observed along battery aging in the early discharge stage of each cycle. However, the performance of capacity difference growth is not linear. For instance, the capacity difference of cycle No. 400 shows a positive value at the early stage of discharge. After a general decrease of the capacity difference around 3 V, the battery capacity difference increases steadily until the end of discharge.

As shown in the Q-V relation and  $\Delta Q$ -V relation of Fig. 3, there is a significant difference in battery discharging behavior at different aging stages, especially the  $\Delta Q$ -V relation at the high voltage range, where the battery does not fully discharge. The divergence at the high battery voltage is a suitable feature to extract for battery modeling even without

finishing the discharging process. The phenomenon may be induced by the loss of active material of the delithiated negative electrode, which shifts the voltage curve at the early discharging stage [4,17]. The sign of degradation is already shown in the early stage of each cycle for the simulation cases of comparing the effect between loss of active material and loss of lithium inventory, where fully discharging the battery may not bring a significant increase of information [28]. However, the research to quantify the battery capacity-voltage performance at different voltage levels along the degradation progress is limited [38–40].

In Fig. 4, the process of converting the selected partial discharge information to the features for machine learning models is detailed. Besides capacity and voltage records, the temperature record is also collected along the cycling process. Limiting the influence of the inconsistency of the discharging time length, the time domain battery operation records are converted to the voltage domain. Referring to the battery dataset, there are 1000 recording points from 3.5 V to 2 V [17]. After converting the time-series records to the time-independent capacity-voltage discharge performance, the capacity-voltage series, capacity difference-voltage series, and temperature series are prepared as the main inputs in our prognosis model.

The selected partial discharge information is a portion of early discharge information in the constant current discharge process, which is an array of capacity records or temperature records. Since the sign of degradation already exists at the beginning of discharge, using partial discharge information to estimate the degradation shows great potential compared to the conventional full-cycle capacity test. To reduce the computational requirement, mathematical calculations are applied to the data series. The measures of the central tendency, such as mean, median, and mode, and the measures of variability, such as the range, variance, skewness, and kurtosis are taken. Since the aforementioned measures are based on well-established mathematic definitions, the calculation is not detailed in this work and can be found in [41]. The advantage of descriptive statistics is to get condensed values to describe a series of data, which fits the further Gaussian process regression framework.

As shown in Table 1, there are three categories of features, which are capacity records (noted as  $Q$ ), capacity difference records (noted as  $\Delta Q$ ), and temperature records (noted as  $T$ ). Mathematical calculations are applied to the selected partial discharge information, and the results of the calculation are used for modeling. In summary, the maximum, mean, and minimum values of temperature information are added in each cycle to investigate the correlation between the temperatures to battery full-discharge capacity. The full-discharge capacity is noted as  $Q_{full}$ , which is the response to the proposed prediction algorithm. With limited discharge information in the present cycle, the capacity estimation concept is to estimate the full-discharge capacity of the present cycle, and the capacity prediction concept is to predict the full-discharge capacity of the future, for example, in 100 cycles, 200 cycles, etc.

The correlation coefficients of 30 % discharge and full information are presented in Fig. 5. Revealing the inconsistency of the battery aging performance in each use case, the correlation test is carried out between

each feature with the present-cycle full-discharge capacity for each battery in the training dataset. As suggested in the data source, the training dataset and two testing datasets contain 41, 43, and 40 battery cells, respectively [17]. Besides 41 battery use cases used in the training, which are tested individually in the dataset, the correlation test is carried out on the aggregation of all the training use cases with and without the outliers, noted as No. 42. and No. 43. The outlier of the battery use case is defined as the average absolute correlation factor is more than 3 scaled median absolute deviation from the median, regarding the mean value of the features that have the absolute correlation factor of more than 0.85 in each battery use case. The correlation is significantly higher when the dataset is without outliers. Although 100 % discharge information gives the exact information for battery full-discharge capacity, some features in the partial information are already highly correlated to the full-discharge capacity.

From the comprehensive results of the correlation test, the discrepancy in the degradation performance of battery use cases is observed. For instance, there are batteries not following the common relation regarding features and response, such as cases No. 2 and No. 9. There are use cases that show a high correlation between temperature and degradation, but other cases give a very low correlation. The cleaning of outliers gives a great improvement in the feature correlation to the response for some features; however, the temperature feature is proved to be not generalized enough and the correlation factor is around zero. With the heatmap showing the correlation test of full discharge information, there are features with strong correlation outstanding. For example, the features from capacity records, already contain the capacity information. Reasonably, the temperature feature is not directly improved in this situation, and the correlation factors remain unchanged.

The results of feature selection are illustrated in Fig. 6, where 30 % of partial discharge information is used. To investigate the best feature combination, a greedy algorithm for feature selection is implemented. In each round of calculation, it goes through all the features for modeling and adds the next best feature to the model, which achieves the lowest RMSE. However, adding more features does not always result in better model performance. There is a significant drop in RMSE for training rather than testing when increasing the number of features. With 5 and more features, the training dataset gives an RMSE near 0, however, the model performance still improves slightly, and the full features model does not show overfitting on testing datasets. From our observation, Gaussian process regression works very well with the redundant features, and the feature selection is not necessary as the model performance on the testing dataset keeps improving with extra features. Therefore, further modeling work is implemented with all 16 features.

## 5. Results

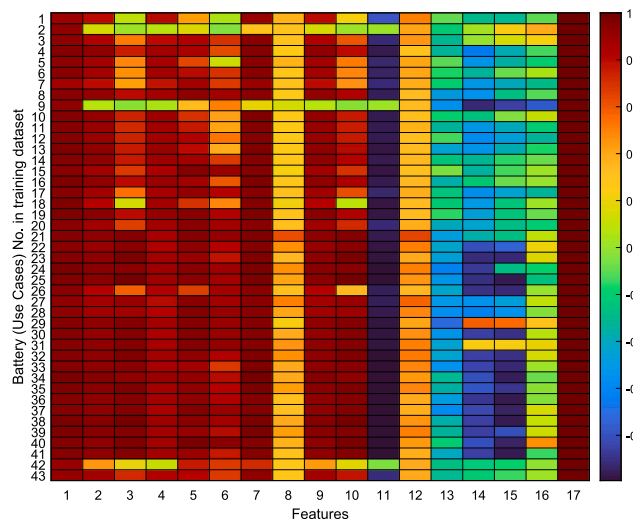
### 5.1. Model performance overview

In this section, the results of capacity estimation and capacity prediction are demonstrated by the Gaussian process regression model with

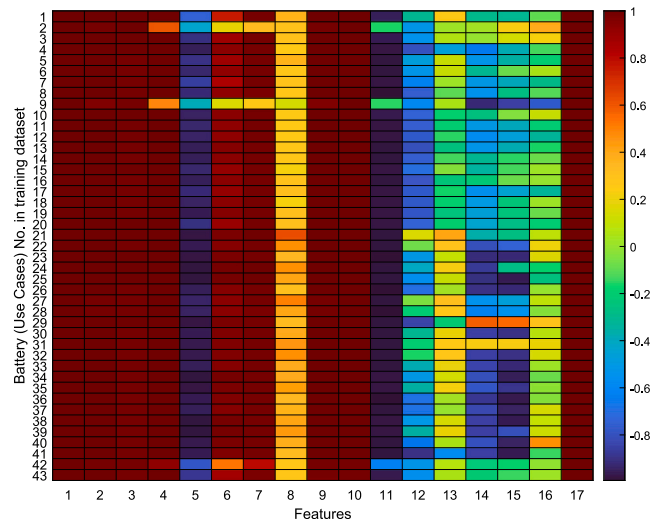
**Table 1**  
Features extracted from the partial discharge information series.

Predictors			Response
Capacity records	Capacity-difference records	Temperature records	Full-discharge capacity (present or future cycle)
Feature No. 1 to 6	Feature No. 7 to 13	Feature No. 14 to 16	Feature No. 17
$Q_{max}$	$\Delta Q_{min}$	$T_{max}$	$Q_{full}$ in $n$ cycle
$Q_{mean}$	$\Delta Q_{max}$	$T_{mean}$	
$Q_{median}$	$\Delta Q_{mean}$	$T_{min}$	
$Q_{variance}$	$\Delta Q_{median}$		
$Q_{skewness}$	$\Delta Q_{variance}$		
$Q_{kurtosis}$	$\Delta Q_{skewness}$		
	$\Delta Q_{kurtosis}$		

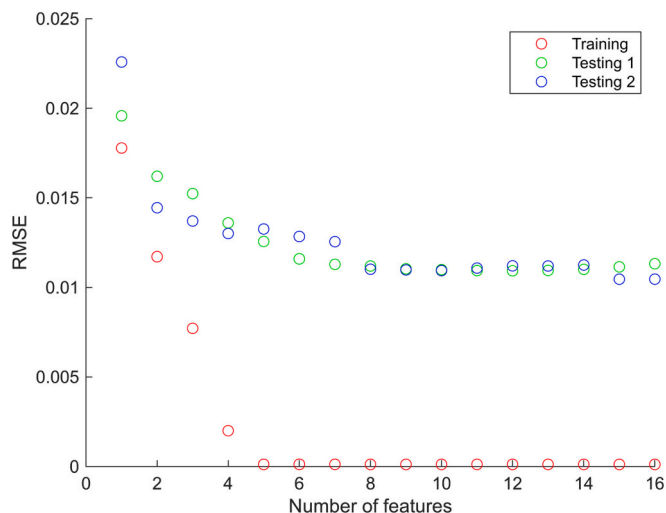
A:



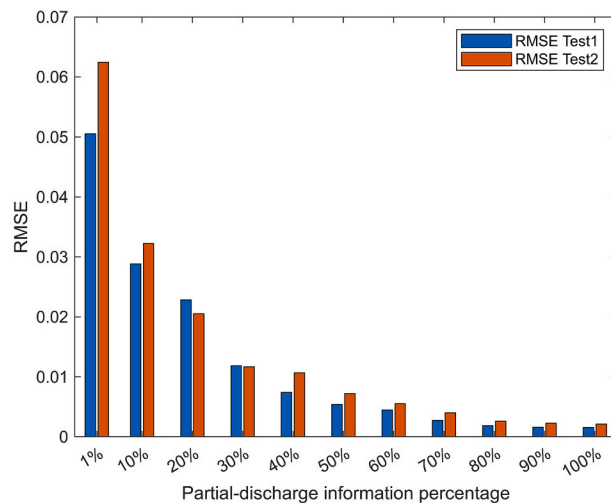
B:



**Fig. 5.** Correlation test between features and response: from the performance of the individual cell to the performance of training datasets. A: 30 % discharge information. B: 100 % discharge information. The use cases from No.1 to No.41 are the 41 battery test samples in the training dataset, No. 42 is calculated by combining all the battery samples in the training dataset, and No. 43 is combining all the samples in the training dataset excluding the battery use cases which are outliers. Features No. 1 to No. 17 is defined in Table 1, and Feature No. 17 represents the full-discharge capacity in the present cycle.



**Fig. 6.** Model performance with increasing feature numbers.



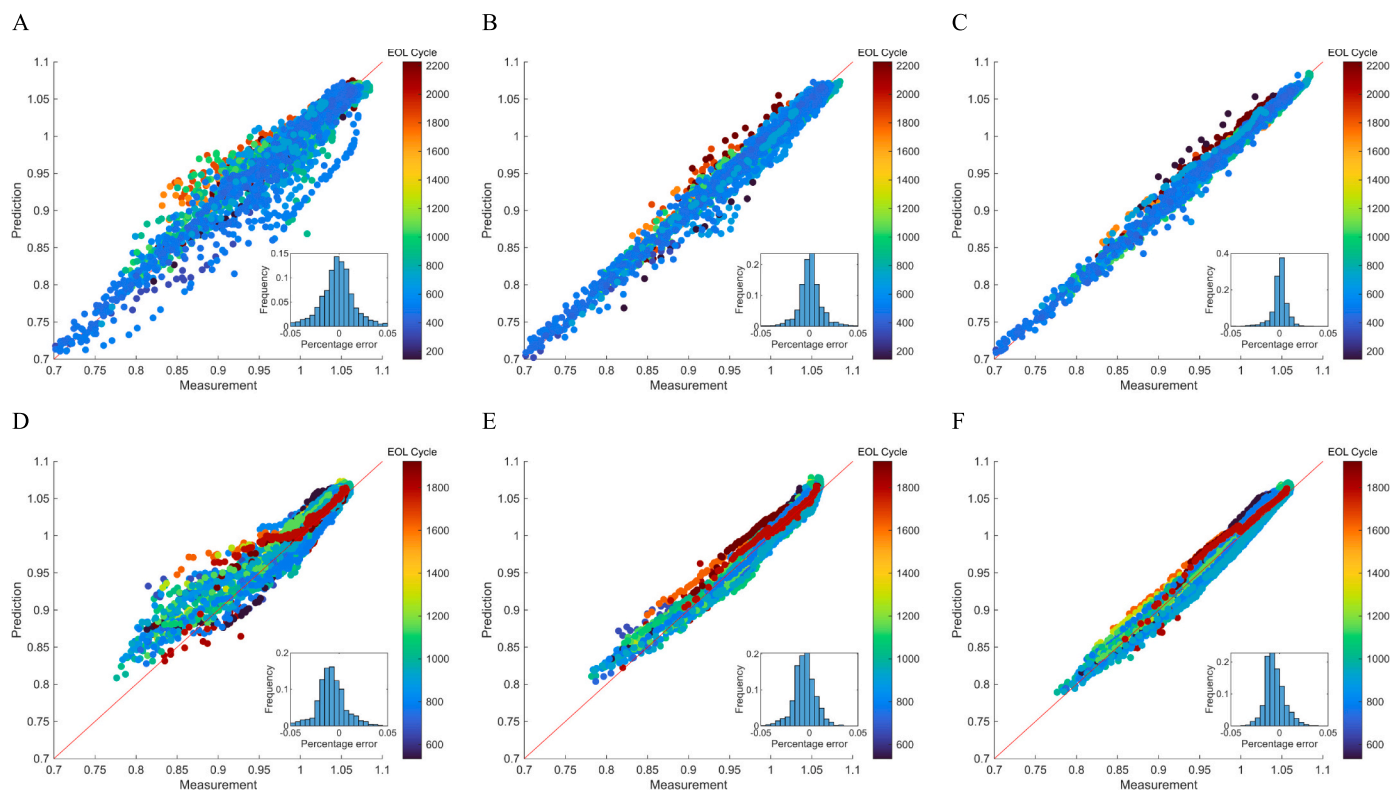
**Fig. 7.** Capacity estimation accuracy for models with different partial discharge information.

partial discharge information. The models are trained with the training dataset and tested by the trained model on testing datasets individually for evaluation. Inherently, the battery performance is different even among the same batch of production, and it is important to cover the initial capacity test to differentiate the initial inconsistency of the batteries. From the industrial application point of view, it is also beneficial to have an initial battery capacity test as calibration before battery usage. Therefore, the capacity difference records are used, which subtract the discharge record of the first cycle to track the performance with the baseline of the initial performance of each battery. Based on the three groups of features including the capacity record, capacity difference record, and temperature records, 16 features are selected as the predictors. As we observed the limited model improvement of the model performance by further feature selection, all the introduced features are put into the model.

The present capacity estimation model estimates the full-discharge

capacity of the present cycle without fully discharging the batteries. The most important tradeoff is the estimation accuracy and amount of information needed. Therefore, a sensitivity analysis is carried out with different percentages of partial discharge information, as shown in Fig. 7. Various percentages of discharge information selection are gathered from the discharging process, and the ranges are from 1 % to 100 % of the voltage discharge range. In the case of capacity measurement, the battery operates from 3.5 V to 2 V during the discharge process, the corresponding discharge capacity and temperature at each voltage are measured and recorded with a step of 0.0015 V, which gives recording 1000 points for each full-discharge process. For example, a 10 % voltage discharge range means using the discharge information from 3.5 V to 3.35 V for modeling, which is a 0.15-V range. As a reference, the partial discharge information with very limited information around 1 % is also modeled. The model already gives insights into the present



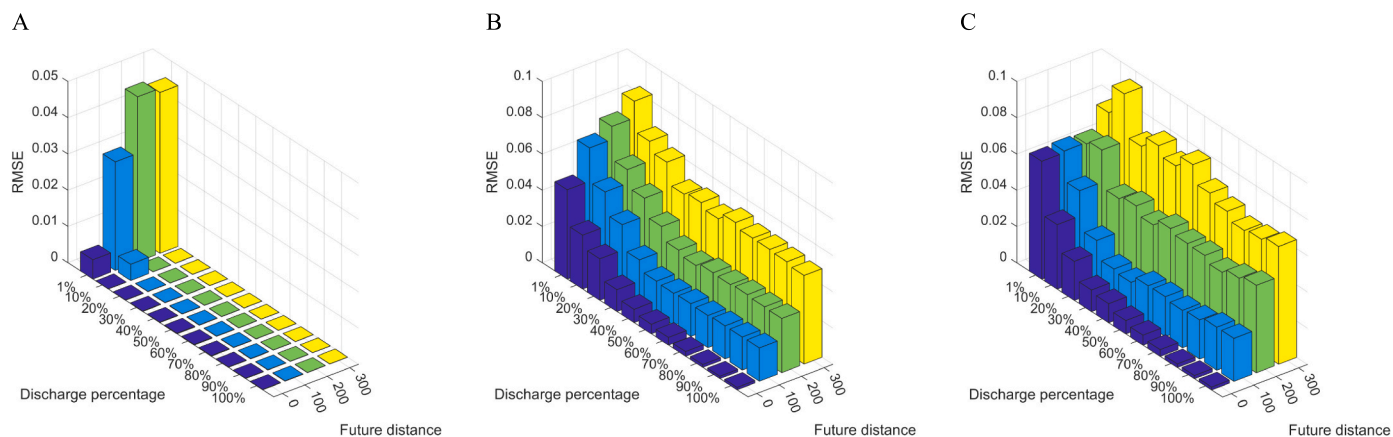


**Fig. 8.** Model performance for capacity estimation. The testing dataset 1: A, B, C; testing dataset 2: D, E, F; The amount of partial information includes 20 %: A, D; 30 % B, E; and 40 %: C, F. The red dash line crossing the figure represents the cases the prediction matches the measurement, and the points are colored by the battery EOL cycle. The inset shows the model performance by the frequency of entities by percentage error in each testing case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

capacity with an RMSE less than 0.07 with only 1 % of partial discharge information, and in the case of a very high percentage of partial-discharge information given, the testing EMSR is close to 0.

As shown in Fig. 8, the model performance is presented with a selected ratio of partial information, including 20 %, 30 %, and 40 %. The scatter diagrams give a comparison of each capacity prediction and measurement during the whole lifetime of the battery use cases. The model performance with 20 % partial discharge information already gives an insight into the full-discharge capacity, and the increasing discharge information brings better model performance. The model

testing performance keeps improving with an increasing amount of partial discharge information. A significant improvement in model performance when increasing the partial discharge information from 20 % to 30 % is observed. For instance, the percentage error of the model with higher discharge formation gets smaller and more concentrated. Potential reasons include that the capacity record is getting closer to full-discharge capacity and the discharge records contain more aging information in this range. As shown from the outliers of prediction, the extremely long-life and short-life batteries are hard to predict. It is challenging to predict EOF performance, especially when there is no



**Fig. 9.** Model performance for capacity prediction. The model performance on the training dataset, test dataset 1, and test dataset 2 is shown in A, B, and C, respectively.

**Table 2**  
Model performance of capacity prediction with the 30 % partial discharge information.

Future distance	RMSE				Average standard deviation			
	0	100	200	300	0	100	200	300
Training	0.0001	0.0001	0	0	0.0014	0.0014	0.0014	0.0014
Testing 1	0.0119	0.0237	0.0366	0.0506	0.0084	0.0163	0.0253	0.0379
Testing 2	0.0117	0.0191	0.0464	0.0723	0.0105	0.0203	0.0309	0.0466

sufficient partial discharge information. In Fig. 8 B and C, the battery use cases with extremely long EOL fall far from the middle crossing line.

To quantitatively describe the model performance, the histograms of percentage error are presented in the insets of Fig. 8 with the bin width of 0.005 and x range from  $-0.5$  to  $0.5$ . In testing dataset 1, as the percentage of discharge information for modeling increases, the highest point of the frequency distribution increases from 0.15 to 0.4, and the distribution is more concentrated around zero, which aligns with the scatter plots. As shown in D, E, and F in Fig. 8, the error distributions of the testing dataset 2 skew toward the left side based on a normal distribution shape. The reason might be that this batch of batteries has some inherent differences compared to the training dataset. The peak value of the error histograms of test case 2 is lower than test case 1, showing there are fewer extremely low error cases in test case 2. With more partial information given for modeling, the model performance increment is less significant in testing case 2, as shown by the peak values of the histogram increasing slightly from 1.7 to 2.2. The scatter plots together with the error histograms of the model performance also reveal the inherent discrepancy of battery use cases. Noted by the color of the scatter plots, the model performance in testing case 2 shows a significant correlation to battery EOL cycle count, which is not obvious in the case of testing case 1. Specifically, the scattering points are more correlated and connected to form a curve of each battery usage case in testing case 2, indicating further possibilities for feature engineering and predictive modeling.

The partial discharge record also contains information that could be used as a better resource to extract features for further capacity prediction. With a similar input as capacity estimation, we simply replace the model response to the battery full-discharge capacity of the future. For example, the future distance of 0, 100, 200, and 300 are selected. To verify the features extracted to be used for future capacity prediction, an illustration is made by various combinations of inputs and prediction targets, which are shown in Fig. 9. Like the prior sections, various amounts of discharge information are used for capacity prediction from 1 % to 100 %. In Table 2, the model performance of capacity prediction is demonstrated with 30 % partial discharge information, as 30 % gives a good balance between information adequacy and model performance. Together with RMSE, the average standard deviations are provided since Gaussian process regression gives an insight into the confidence of the prediction. The average standard deviation is calculated by the average of all the capacity predictions in each selected dataset. Increasing RMSE and average standard deviation are observed with larger future distance for prediction.

As indicated by RMSE, the model performs very well on the cycle capacity prediction on training and testing datasets. Around 1 %, 2 %, and 3 % errors of capacity prediction are achieved on testing for the future distances of 0, 100, and 200, respectively. As the inputs of the model are very limited, it is hard to make a long-distance prediction. Besides RMSE, the average standard deviation is generated by the Gaussian process regression model, which shows the confidence of the model prediction. The results of the two indicators align with each other.

## 5.2. Capacity estimation at all aging stages

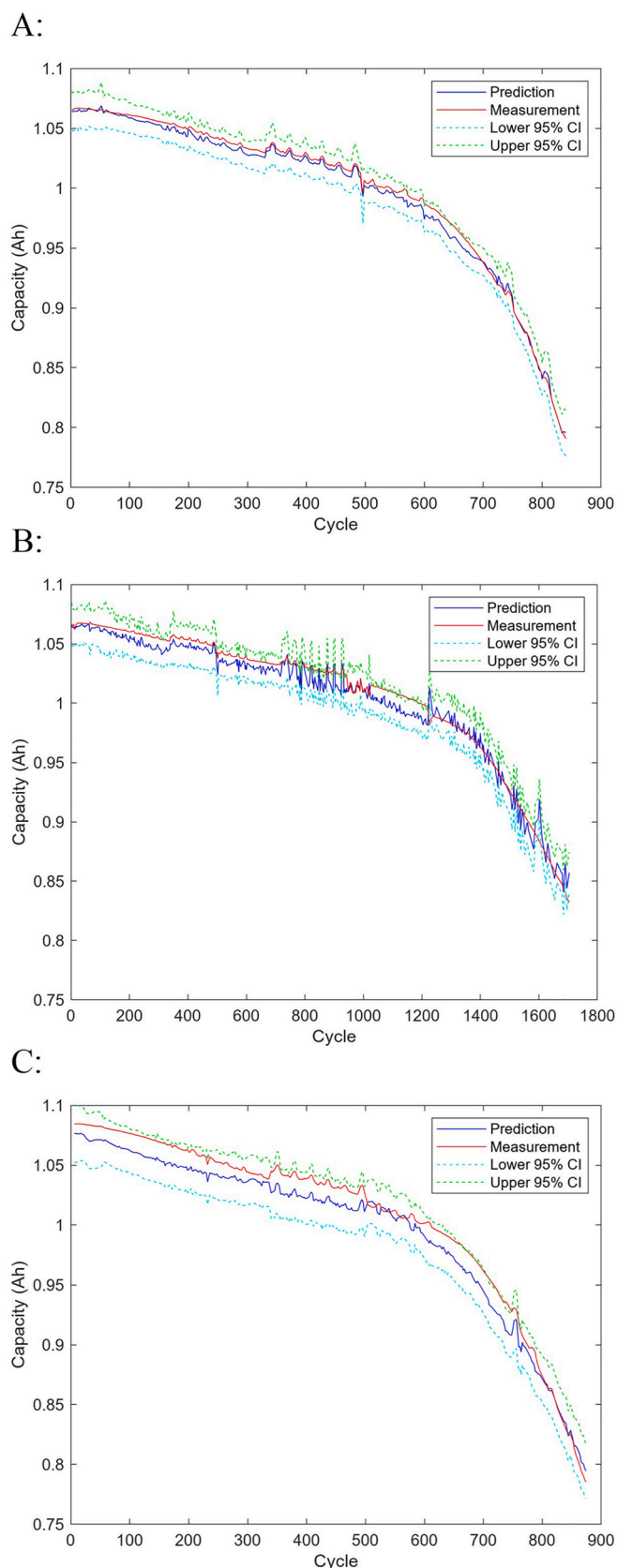
As shown in Fig. 10, three cases from testing datasets are presented for comparison of model prediction and capacity measurement. Case A, B, and C are selected with various RMSEs, which represent the different capacity estimation accuracy. The RMSEs of 0.0044, 0.0086, and 0.0130 are acquired by implementing the trained model on three different testing datasets. Besides the capacity estimation, the lower 95 % confidence interval (CI) and upper 95 % CI are provided by the trained Gaussian process regression model. Overall, the battery degradation trajectory is well captured in most situations, which overcomes the bias of conventional extrapolation methods.

In Case A, the prediction follows the measurement nicely even during the middle period when the capacity increment occurs around 400 cycles, which may be caused by the temperature fluctuation in the testing chamber. The measurement is within the 95 % CI range at most of the aging stages. In Case B, there is more fluctuation in the battery measurement records and the battery has an extremely long lifetime under the fast-charging testing process. Specifically, the CI has a larger range and the prediction is off-track in the range from 1000 to 1200 cycles. However, our model provides an individual assessment of the SOH at each cycle, which is very robust when parts of the records are abnormal. Case C shows a battery with an extra-large capacity from the very beginning. Since there is an inherent difference in the battery from the other samples, it is hard to capture the capacity value at the early stage, however, when the battery is aged, the prediction accuracy is improved significantly. The possible reason might be there is extra active material in this battery cell, which is an exceptional case that has not been observed in the training dataset. Also, the CI in case C is significantly large than A and B, indicating that the model knows it is dealing with an abnormal battery case. Overall, a single model trained by our proposed method captures the battery SOH at all stages along the capacity degradation trajectory, including the capacity increment at the early stage, the knee point when the degradation pace changes, fluctuations caused by external factors, and the EOL performance.

## 5.3. Capacity trajectory prediction

As shown in Fig. 11, the model performance of capacity prediction up to 300 cycles is presented. There are 4 Gaussian process regression models trained on the training dataset with the future distance of 0, 100, 200, and 300 cycles. The trained models are implemented on a battery aging record from the testing dataset, which is the same battery of Case A in Fig. 10. The present-cycle full-discharge capacity is denoted by blue dots as "Measurement", the model prediction is denoted by hollow circles in the scatter plot, and the 95 % CI is denoted by dash lines. Overall, all the models follow the trend of degradation trajectory well, and the model with a short prediction distance gives better accuracy and CI.

As shown in the detailed performance of the early age of the cycle in Fig. 11. B, the model predicts the degradation performance within a very narrow band near the measurement. With increasing prediction distance



**Fig. 10.** Model performance of capacity estimation on three different cases in testing datasets. A: RMSE = 0.0044 Ah; B: RMSE = 0.0086 Ah; C: RMSE = 0.0130 Ah.

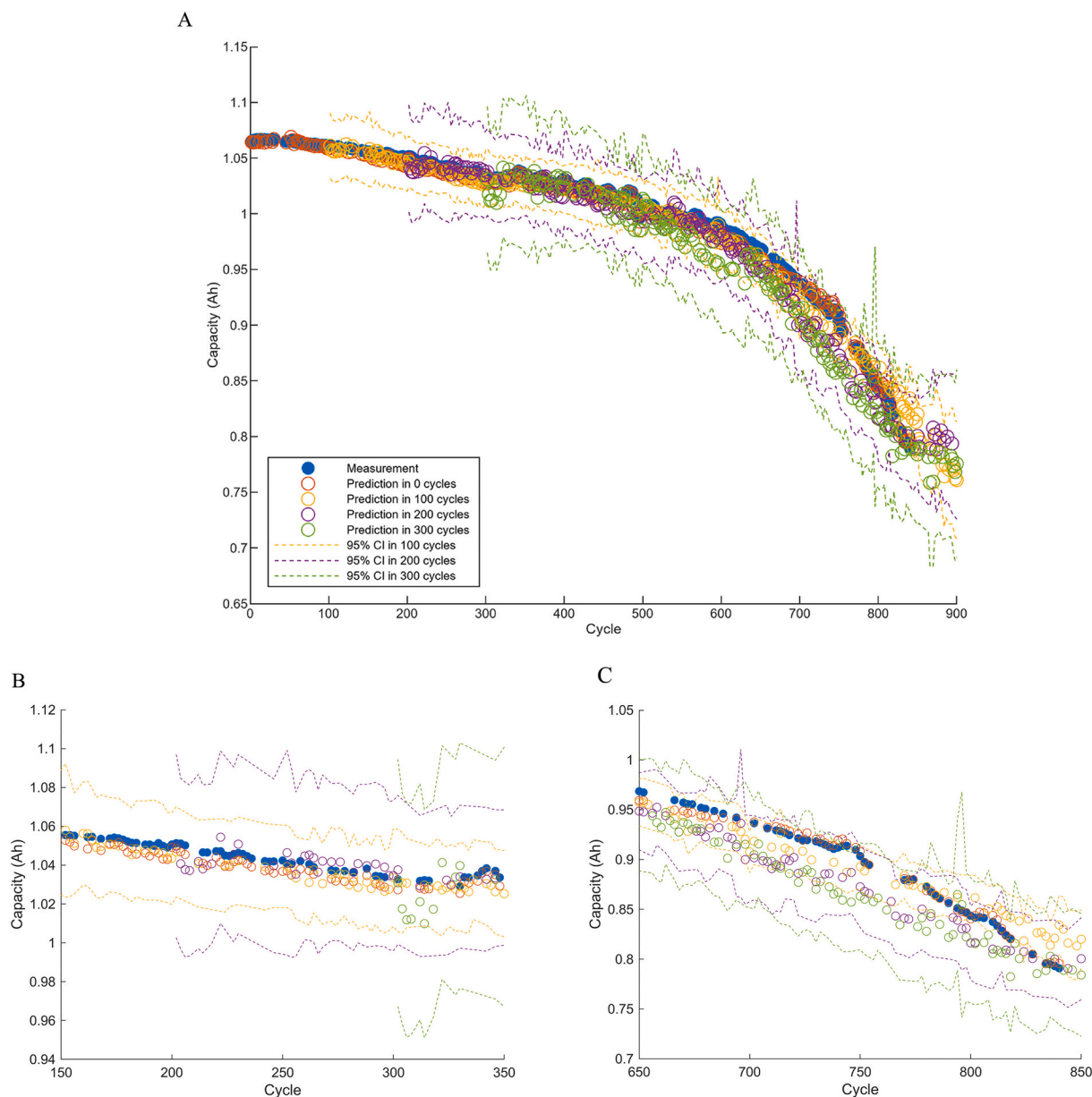
from 100 to 300, the CI increases gradually. Since our trained model only requires information from one cycle, the prediction of cycle No. 300 can be made with the information from the first cycle. As the battery performance is not stable at the very early usage period, the prediction results fluctuate, which can be observed around 200 cycles and 300 cycles. Considering it is the result of the pre-trained model requiring the input of only one partial-discharge cycle of the battery, the model shows great competence in degradation prognosis.

In Fig. 11. C, the model performs well at the end of battery life and even gives degradation prediction when the battery capacity is not measured. The reason is that there are some batteries tested further than 0.8 Ah in the training dataset, and the degradation trend is learned by the long-term degradation model with a longer cycle distance. Though the speed of degradation is much faster compared to the early stage, the trend of degradation is still captured. There are some measurements are excluded because of the noise in data records as shown in the breaking points, however, the trained model still shows some estimated values of full-discharge capacity. In summary, our model gives great degradation performance at different stages of battery degradation with limited partial discharge information.

## 6. Discussion

As the aging dataset used in this work is based on the LFP/graphite battery, modification is needed for implementing the model to the other type of batteries. However, the modeling methodologies and even the trained model are promising to be used directly or indirectly on other types of batteries with the development of machine learning. For example, the hierarchical Bayesian model for rapid prediction of degradation is implemented for both LFP/graphite and lithium-manganese-cobalt-oxide (NMC)/graphite cells, which proves the feasibility of feature selection and data-driven modeling for different battery chemistries [31]. Some mechanisms support the generalizability of data-driven models on various batteries. For example, if the batteries share the same kinds of material that dominate the degradation process, for example, the cathode material, the degradation mechanism might be similar and the degradation performance could be similar [17,42]. Furthermore, there is an increasing amount of research in using transfer learning to fine-tune or adapt the model between different battery chemistries, battery formats, and different usage conditions, which boosts the applicability of data-driven degradation prognosis [43].

The improvement of the data science and machine learning toolbox brings opportunities for battery health prognosis. On one hand, finding the features which can achieve the best accuracy is of vital importance. On the other hand, to improve the model's applicability for different feature combinations and different use cases is a need for further implementation. With further feature selection, the model may achieve a slightly better RMSE for degradation prediction. However, the best feature combination for different lengths of selected partial discharge information is different, which means some of the features are more related to the long-term degradation performance and some are not. It might not be necessary to fine-tune the feature combination to achieve a similar accuracy between training and testing datasets, because we observe that the redundant features will not overfit the model and still gives good results on testing datasets. In addition, our work uses the statistical characteristics of the features in the discharging process. It uses condensed values for less computational cost but also loses some detailed information. Using other machine learning techniques can utilize a bigger amount of data, which may take a better usage of large battery operating datasets. Regarding capacity estimation, our model gives good results at different stages of battery aging. However, it is challenging to correlate the battery discharge performance of one cycle to the degradation performance in a few hundred cycles in the future. Future development of our work aims at integrating accessible partial discharging information of battery degradation at different stages toward degradation prediction with higher accuracy and applicability.



**Fig. 11.** Model performance of degradation prognosis by 30 % partial-discharge information on one battery use case in testing datasets, including 0-cycle, 100-cycle, 200-cycle, and 300-cycle ahead prediction: A: overall performance; B: performance of early aging; C: performance of EOL.

## 7. Conclusion

In this work, we propose the model utilizing partial discharge information for present full-discharge capacity estimation and future full-discharge capacity prediction. The proposed method achieves high prediction accuracy using a very limited amount of data, which is from the discharging record of the initial cycle and the present cycle. Instead of capacity record extrapolation of a single cell, our model carries out the degradation prognosis crossing different battery use causes. Since our approach does not require the full discharge process to measure the current battery capacity, it is promising in industrial applications for real-time degradation estimation. Using Gaussian process regression improves the performance of the model and gives good flexibility in feature engineering. With 30 % of partial discharge information, around 1 % RMSE and less than 1 % average standard deviation of prediction are achieved regardless of the stage of battery aging. The proposed method also gives insight into the battery health prognosis for the next 300 cycles with the input of only one partial discharge cycle at any stage

of the battery degradation. Since the information required in this model is easy to be acquired during battery real-life operation, a foreseeable great amount of valuable information can be collected and predictions can be made during the battery operation, which will provide great improvement for battery prognosis by breaking through the insufficiency of battery operation information.

### CRedit authorship contribution statement

**Chunyang Zhao:** Conceptualization, Methodology, Investigation, Writing – original draft. **Peter Bach Andersen:** Conceptualization, Writing – review & editing, Supervision. **Chresten Træholt:** Writing – review & editing, Supervision. **Seyedmostafa Hashemi:** Methodology, Writing – review & editing, Supervision, Funding acquisition.

### Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The work is based on the published datasets.

## Acknowledgment

This work is supported by the Danish project “BOSS: Bornholm smartgrid secured by grid-connected battery systems” co-founded by Danish Energy Technology Development and Demonstration Program (EUDP) contract no. 64018-0618.

## References

- [1] C. Zhao, S. Hashemi, P.B. Andersen, C. Traholt, Data-driven State of health modeling of battery energy storage systems providing grid services, in: 2021 11th International Conference on Power, Energy and Electrical Engineering, CPEEE 2021, 2021, pp. 43–49, <https://doi.org/10.1109/CPEEE51686.2021.9383356>.
- [2] X. Han, L. Lu, Y. Zheng, X. Feng, Z. Li, J. Li, M. Ouyang, A review on the key issues of the lithium ion battery degradation among the whole life cycle, *ETransportation* 1 (2019), 100005, <https://doi.org/10.1016/j.etrans.2019.100005>.
- [3] X. Hu, L. Xu, X. Lin, M. Pecht, Battery lifetime prognostics, *Joule* 4 (2020) 310–346, <https://doi.org/10.1016/j.joule.2019.11.018>.
- [4] C.R. Birkel, M.R. Roberts, E. McTurk, P.G. Bruce, D.A. Howey, Degradation diagnostics for lithium ion cells, *J. Power Sources* 341 (2017) 373–386, <https://doi.org/10.1016/j.jpowsour.2016.12.011>.
- [5] M. Ecker, J.B. Gerschler, J. Vogel, S. Käbitz, F. Hust, P. Dechent, D.U. Sauer, Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data, *J. Power Sources* 215 (2012) 248–257, <https://doi.org/10.1016/j.jpowsour.2012.05.012>.
- [6] X. Hu, L. Xu, X. Lin, M. Pecht, Battery lifetime prognostics, *Joule* 4 (2020) 310–346, <https://doi.org/10.1016/j.joule.2019.11.018>.
- [7] J.M. Reniers, G. Mulder, D.A. Howey, Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries, *J. Electrochem. Soc.* 166 (2019) A3189–A3200, <https://doi.org/10.1149/2.0281914JES/XML>.
- [8] Y. Li, K. Liu, A.M. Foley, A. Zülke, M. Berceibar, E. Nanini-Maury, J. Van Mierlo, H. E. Hoster, Data-driven health estimation and lifetime prediction of lithium-ion batteries: a review, *Renew. Sust. Energ. Rev.* 113 (2019), 109254, <https://doi.org/10.1016/j.rser.2019.109254>.
- [9] X. Hu, Y. Che, X. Lin, S. Onori, Battery health prediction using fusion-based feature selection and machine learning, *IEEE Trans. Transp. Electrification* 7 (2021) 382–398, <https://doi.org/10.1109/TTE.2020.3017090>.
- [10] J.C. Forman S.J. Moura J.L. Stein H.K. Fathy Optimal experimental design for modeling battery degradation ASME 2012 5th Annual Dynamic Systems and Control Conference Joint With the JSME 2012 11th Motion and Vibration Conference, (n.d.). doi:10.0/Lin-x86.64.
- [11] G. dos Reis, C. Strange, M. Yadav, S. Li, Lithium-ion battery data and where to find it, *Energy and AI* 5 (2021), 100081, <https://doi.org/10.1016/j.egyai.2021.100081>.
- [12] B. Saha, K. Goebel, Battery Data Set, NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007. <http://ti.arc.nasa.gov/project/prognostic-data-repository>.
- [13] A. Saxena, J.R. Celaya, I. Roychoudhury, B. Saha, S. Saha, K. Goebel, Designing data-driven battery prognostic approaches for variable loading profiles: some lessons learned, in: First European conference of the prognostics and health management society 2012, 2012, p. 10.
- [14] H.T. Lin, T.J. Liang, S.M. Chen, Estimation of battery state of health using probabilistic neural network, *IEEE Trans Industr Inform* 9 (2013) 679–685, <https://doi.org/10.1109/TII.2012.2222650>.
- [15] R.R. Richardson, M.A. Osborne, D.A. Howey, Gaussian process regression for forecasting battery state of health, *J. Power Sources* 357 (2017) 209–219, <https://doi.org/10.1016/j.jpowsour.2017.05.004>.
- [16] K. Liu, Y. Shang, Q. Ouyang, W.D. Widanage, A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of Lithium-ion battery, *IEEE Trans. Ind. Electron.* 68 (2021) 3170–3180, <https://doi.org/10.1109/TIE.2020.2973876>.
- [17] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, M.Z. Bazant, S.J. Harris, W.C. Chueh, R. D. Braatz, Data-driven prediction of battery cycle life before capacity degradation, *Nat. Energy* 4 (2019) 383–391, <https://doi.org/10.1038/s41560-019-0356-8>.
- [18] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, D.U. Sauer, One-shot battery degradation trajectory prediction with deep learning, *J. Power Sources* 506 (2021), 230024, <https://doi.org/10.1016/j.jpowsour.2021.230024>.
- [19] J. Lu, R. Xiong, J. Tian, C. Wang, C.-W. Hsu, N.-T. Tsou, F. Sun, J. Li, Battery degradation prediction against uncertain future conditions with recurrent neural network enabled deep learning, *Energy Storage Mater.* 50 (2022) 139–151, <https://doi.org/10.1016/j.ensm.2022.05.007>.
- [20] C.W. Hsu, R. Xiong, N.Y. Chen, J. Li, N.T. Tsou, Deep neural network battery life and voltage prediction by using data of one cycle only, *Appl. Energy* 306 (2022), <https://doi.org/10.1016/j.apenergy.2021.118134>.
- [21] W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte, D.U. Sauer, Digital twin for battery systems: cloud battery management system with online state-of-charge and state-of-health estimation, *J. Energy Storage* 30 (2020), <https://doi.org/10.1016/j.est.2020.101557>.
- [22] M. Berceibar, M. Garmendia, I. Gandiaga, J. Crego, I. Villarreal, State of health estimation algorithm of LiFePO<sub>4</sub> battery packs based on differential voltage curves for battery management system application, *Energy* 103 (2016) 784–796, <https://doi.org/10.1016/J.ENERGY.2016.02.163>.
- [23] X. Feng, C. Weng, X. He, X. Han, L. Lu, D. Ren, M. Ouyang, Online state-of-health estimation for li-ion battery using partial charging segment based on support vector machine, *IEEE Trans. Veh. Technol.* 68 (2019) 8583–8592, <https://doi.org/10.1109/TVT.2019.2927120>.
- [24] S. Khaleghi Rahimian, Y. Tang, A practical data-driven battery state-of-health estimation for electric vehicles, *IEEE Trans. Ind. Electron.* 70 (2023) 1973–1982, <https://doi.org/10.1109/TIE.2022.3165295>.
- [25] C. Zhao, P.B. Andersen, C. Traeholt, S. Hashemi, Data-driven cycle-calendar combined battery degradation modeling for grid applications, in: 2022 IEEE Power & Energy Society General Meeting (PESGM), IEEE, 2022, pp. 1–5, <https://doi.org/10.1109/PESGM48719.2022.9917143>.
- [26] P. Tagade, K.S. Hariharan, S. Ramachandran, A. Khandelwal, A. Naha, S.M. Kolake, S.H. Han, Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis, *J. Power Sources* 445 (2020), <https://doi.org/10.1016/J.JPOWSOUR.2019.227281>.
- [27] S. Saxena, C. Hendricks, M. Pecht, Cycle life testing and modeling of graphite/LiCoO<sub>2</sub> cells under different state of charge ranges, *J. Power Sources* 327 (2016) 394–400, <https://doi.org/10.1016/J.JPOWSOUR.2016.07.057>.
- [28] P.M. Attia, A. Bills, F.B. Planella, P. Dechent, G. dos Reis, M. Dubarry, P. Gasper, R. Gilchrist, S. Greenbank, D. Howey, O. Liu, E. Khoo, Y. Preger, A. Soni, S. Sripad, A.G. Stefanopoulou, V. Sulzer, Review—“knees” in lithium-ion battery aging trajectories, *J. Electrochem. Soc.* 169 (2022), 060517, <https://doi.org/10.1149/1945-7111/AC6D13>.
- [29] P. Fermín-Cueto, E. McTurk, M. Allerhand, E. Medina-Lopez, M.F. Anjos, J. Sylvester, G. dos Reis, Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells, *Energy and AI* 1 (2020), 100006, <https://doi.org/10.1016/J.EGYAI.2020.100006>.
- [30] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, D.U. Sauer, Online capacity estimation of lithium-ion batteries with deep long short-term memory networks, *J. Power Sources* 482 (2021), 228863, <https://doi.org/10.1016/j.jpowsour.2020.228863>.
- [31] B. Jiang, W.E. Gent, F. Mohr, S. Das, M.D. Berliner, M. Forsuelo, H. Zhao, P. M. Attia, A. Grover, P.K. Herring, M.Z. Bazant, S.J. Harris, S. Ermon, W.C. Chueh, R.D. Braatz, Bayesian learning for rapid prediction of lithium-ion battery-cycling protocols, *Joule* 5 (2021) 3187–3203, <https://doi.org/10.1016/J.JOULE.2021.10.010>.
- [32] R.R. Richardson, M.A. Osborne, D.A. Howey, Battery health prediction under generalized conditions using a Gaussian process transition model, *J. Energy Storage* 23 (2019) 320–328, <https://doi.org/10.1016/j.est.2019.03.022>.
- [33] E. Schaltz, D.I. Stroe, K. Norregaard, B. Johnsen, A. Christensen, Partial charging method for lithium-ion battery state-of-health estimation, in: 2019 14th International Conference on Ecological Vehicles and Renewable Energies, EVER 2019, 2019, <https://doi.org/10.1109/EVER.2019.8813645>.
- [34] C.E.R., C.K.I. Williams, *Gaussian Processes for Machine Learning* 69, 2006.
- [35] J. Görtler, R. Kehlbeck, O. Deussen, A visual exploration of Gaussian processes, *Distill.* 4 (2019), e17, <https://doi.org/10.23915/distill.00017>.
- [36] R.R. Richardson M.A. Osborne D.A. Howey Battery health prediction under generalized conditions using a Gaussian process transition model *J. Energy Storage*. 23 2019 320 328 10.1016/j.est.2019.03.022.
- [37] J. Peng, J. Meng, X. Du, L. Cai, D.-I. Stroe, A fast impedance measurement method for Lithium-ion battery using power spectrum property, *IEEE trans industrinform* (2022) 1–9, <https://doi.org/10.1109/TII.2022.3217474>.
- [38] A. Tomaszewska, Z. Chu, X. Feng, S. O’Kane, X. Liu, J. Chen, C. Ji, E. Endler, R. Li, L. Liu, Y. Li, S. Zheng, S. Vetterlein, M. Gao, J. Du, M. Parkes, M. Ouyang, M. Marinescu, G. Offer, B. Wu, Lithium-ion battery fast charging: a review, *ETransportation* 1 (2019), 100011, <https://doi.org/10.1016/J.ETRAN.2019.100011>.
- [39] S.E.J. O’Kane, W. Ai, G. Madabattula, D. Alonso-Alvarez, R. Timms, V. Sulzer, J. S. Edge, B. Wu, G.J. Offer, M. Marinescu, Lithium-ion battery degradation: how to model it, *Phys. Chem. Chem. Phys.* 24 (2022) 7909–7922, <https://doi.org/10.1039/D2CP00417H>.
- [40] M. Jiang, D.L. Danilov, R.A. Eichel, P.H.L. Notten, A review of degradation mechanisms and recent achievements for Ni-rich cathode-based Li-ion batteries, *Adv. Energy Mater.* 11 (2021), <https://doi.org/10.1002/aem.202103005>.
- [41] M. Ezekiel, F.K. August, *Methods of Correlation and Regression Analysis*, 1963.
- [42] D. Anseán, M. Dubarry, A. Devie, B.Y. Liaw, V.M. García, J.C. Viera, M. González, Fast charging technique for high power LiFePO<sub>4</sub> batteries: a mechanistic analysis of aging, *J. Power Sources* 321 (2016) 201–209, <https://doi.org/10.1016/j.jpowsour.2016.04.140>.
- [43] K. Liu, Q. Peng, Y. Che, Y. Zheng, K. Li, R. Teodorescu, D. Widanage, A. Barai, Transfer learning for battery smarter state estimation and ageing prognostics: recent progress, challenges, and prospects, *Adv. Appl. Energy* 9 (2023), 100117, <https://doi.org/10.1016/J.ADAPEN.2022.100117>.