CHALMERS
UNIVERSITY OF TECHNOLOGY

State of health estimation for lithium-ion batteries under arbitrary usage using data-driven multi-model fusion

Citation for the original published paper (version of record):
Zhang, Y., Wik, T., Bergström, J. et al (2024). State of health estimation for lithium-ion batteries under arbitrary usage using data-driven
multi-model fusion. IEEE Transactions on Transportation Electrification, 10(1): 1494-1507.
http://dx.doi.org/10.1109/TTE.2023.3267124

(article starts on next page)

State of health estimation for lithium-ion batteries under arbitrary usage using data-driven multi-model fusion

Yizhou Zhang, Torsten Wik, *Member, IEEE*, John Bergström, Changfu Zou, *Senior Member, IEEE*

***Abstract*—Accurately estimating the state of health (SoH) of batteries is indispensable for the safety, reliability, and optimal energy and power management of electric vehicles. However, from a data-driven perspective, complications, such as dynamic vehicle operating conditions, stochastic user behaviors, and cell-to-cell variations, make the estimation task challenging. This work develops a data-driven multi-model fusion method for SoH estimation under arbitrary usage profiles. All possible operating conditions are categorized into six scenarios. For each scenario, an appropriate feature set is extracted to indicate the SoH. Based on the obtained features, four machine learning algorithms are applied individually to train SoH estimation models using time-series data. In addition to the estimates at the current time step, a histogram data-based and online adaptive model is taken from previous work for predicting the next-step SoH. Then, a Kalman filter is applied to systematically fuse the results of all the estimation and prediction models. Experimental data collected from different types of batteries operated under diverse profiles verify the effectiveness and practicability of the developed method, as well as its superiority over individual models.**

***Index Terms*—Battery capacity estimation, SoH estimation, Machine learning, Multi-model fusion, Kalman filter.**

I. Introduction

Massive electrification in the transportation sector has become an unstoppable global trend thanks to the ever-growing environmental consciousness of the public, stringent regulations on vehicle emissions, and advancements in electric propulsion and energy storage technology [1]. Automakers are increasingly shifting from the combustion engine-centered powertrain architecture toward an electrified solution. Many of them have made bold commitments to electrify all their product portfolios by 2025 and stop selling combustion engine vehicles by as early as 2030 [2]. Lithium-ion batteries play an essential role in this transition because of their high energy density, relatively low cost, and long lifetime [3]–[5]. As one of the most critical and expensive vehicle components, the battery attracts tremendous attention from the industry and academia. Among different properties, aging is pivotal as it weakens the battery's performance and reliability, deteriorates the state of health (SoH) and increases the risk of safety hazards. Furthermore, such a process can be easily accelerated if the battery is used inappropriately. Therefore, for optimal energy management, timely maintenance, and accurate residual value prediction, it is vital to understand the aging process and monitor SoH in real-time.

Generally speaking, the existing methods for battery SoH estimation can broadly be categorized into empirical, model-based, and data-driven methods [6]–[10]. Based on extensive laboratory cycling data, empirical methods typically use polynomial, exponential, or quadratic functions to fit the battery aging trend. These methods commonly rely on a hypothesis of regular cycling profiles. However, batteries deployed in electrified vehicles (EVs) encounter irregular and complicated operating conditions, which inevitably undermines the accuracy and reliability of these estimation methods [11].

In contrast to an open-loop estimation of the empirical method, model-based methods adopt a closed-loop solution. Two different kinds of models are commonly employed, i.e., equivalent circuit models (ECMs) developed from Kirchhoff's laws and electrochemical models (EMs) derived from porous electrode theory [12]. Because of their structural simplicity and ease of implementation, ECMs have been extensively studied and applied to battery aging prognostics. However, the fidelity of such models degrades when the profiles experienced in real-world applications differ from those of laboratory characterization or when some model parameters deviate from their true values due to the lack of regular reference performance tests [13]. EMs have the capability of modeling local dynamics within a battery, including aging mechanisms, e.g., solid-electrolyte interface growth, lithium plating, and particle cracking [14]–[16]. However, the requirement for high computation and the difficulty of model parameterization make this kind of model hard to use in online applications [8], [11], [17].

Many research attempts have recently been made to apply data-driven methods to battery SoH estimation because they are flexible, mechanism-agnostic, and have the capability of recognizing patterns and trends under complex dynamic situations [18], [19]. Feature construction is one of the critical steps for such methods, as the performance of ML algorithms will heavily depend on if the selected features contain enough information to indicate the aging state of the battery or not [20], [21]. A typical battery management system (BMS) only

This work was funded by the Swedish Energy Agency under the Vehicle Strategic Research and Innovation Program (Grant No. 50187-1). Corresponding author: Changfu Zou.

Y. Zhang is with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden, and also with the China Euro Vehicle Technology AB, Gothenburg, 41755, Sweden (e-mail: yizhou@chalmers.se)

T. Wik and C. Zou are with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden (e-mail: tw@chalmers.se, changfu.zou@chalmers.se).

J. Bergström is with the China Euro Vehicle Technology AB, Gothenburg, 41755, Sweden (e-mail: john.bergstrom@cevt.se).

measures the individual cell voltage, and the pack current, together with a few temperature sensors for each module in the pack [22]. Hence, selecting features based on the information available in real-world battery systems is essential. Richardson *et al.* [23] chose the accumulated cycling time, and Ah throughput between two neighboring reference performance tests of the laboratory cycled cells as features to estimate battery capacity. However, the uncontrolled operating conditions of end-users, especially the discharging phase with intermittent regenerative braking and potentially long parking time causing calendar aging, can pose a significant challenge to the applicability of such a method. Deng *et al.* [24] constructed features from the discharge capacity and voltage curve under pulse discharging conditions and showed a clear correlation between the features and the cell capacity. Compared to discharging processes that are unpredictable and stochastic, the charging profiles are easy to control, e.g., the constant current-constant voltage (CC-CV) scheme that is adopted widely in commercial BMSs. This has attracted great research interest in using CC-CV charging curves for SoH estimation. Li *et al.* [25] applied the raw voltage curve extracted under the repetitive laboratory CC-CV charging cycles to estimate the cell capacity, where long short-term memory (LSTM) neural networks were employed for model development. However, the variations in battery temperature and the initial state of charge (SoC) were not considered in their study, which essentially limits the method's practicability for vehicle battery systems.

Incremental capacity (IC) and differential voltage (DV) curves that are usually used for the analysis of battery aging mechanisms can also be used to estimate battery SoH [26], [27]. To do so, characteristic values associated with the peaks or valleys of IC/DV curves, e.g., the height, are typically selected as features. However, due to random user behaviors, the SoC range may frequently vary over different charging cycles in actual EV usage. Since the peaks and valleys only appear at specific voltage windows, the corresponding estimator becomes useless when such a window does not show up in the charging realization. This solicits research endeavors to develop reliable estimators from partial charging curves. Wei *et al.* [28] extracted morphological IC and voltage entropy from the partial charging curve, and by using them as features in ML, they achieved comparable estimation results as using the complete charging curves. Tian *et al.* [29] used a regional voltage curve to reconstruct the whole charging curve, after which the IC method can easily be used to estimate battery SoH. Nevertheless, in these studies, the battery cells were cycled originally under the complete CC-CV charging, and then some of the charging profiles were removed manually during data processing to form partial charging curves. These synthetic curves will differ from real-world battery cells' natural partial charging profiles due to the voltage polarization effect and the internal resistance caused by the initial voltage rise [30]. Such a model-plant mismatch may expose these designed estimators to a serious pitfall and result in poor performance for vehicle applications. To the best of our knowledge, none of the existing IC/DV feature-based SoH estimators has thoroughly, systematically, and statistically avoided this pitfall.

To cope with the wide operating range of EVs and always deliver reliable SOH estimation, there exist some early works to formulate the estimation task based on different usage profiles. Mitici et al. [31] constructed the features based on the charge and discharge process and conducted a feature importance analysis. However, the highly dynamic discharging process and irregular charging profiles may hinder the performance of such a method in real-world EV applications. According to the SoC level during driving cycles, Rahimian et al. [32] divided the operation conditions into three consecutive SoC windows to extract relevant features. But such a way of dividing the operating conditions can be affected by the SoC estimation error. Additionally, due to the rather large SoC interval, many operating conditions cannot be covered. In [33], Zhang et al. improved the way of categorizing the operating conditions by only using vehicle charging profiles and dividing them into three consecutive categories according to the voltage level. They assigned all scenarios that do not contain any IC peak area to one category, for which only the historical SoH estimation data were used as input to train a linear regression algorithm. However, such an algorithm will inevitably result in limited accuracy and robustness for SoH estimation.

When learning the output behavior of a system from a given input-output dataset, an ensemble of ML models has the potential to improve the prediction performance from individual models, as reviewed and summarized in [34]. The advantages of the combined model over its individual counterparts become even greater when the presumptions for the individual models do not hold strictly for certain cases [35]. Because of such merits, model fusion has recently been introduced to address the problem of battery aging diagnostics. Lin *et al.* [36] trained three sub-models based on different ML algorithms and used a random forest regression (RFR) to fuse their results to form a final battery SoH estimation. Similar work was conducted in [21] by using a modified RFR. Shen *et al.* [37] constructed several deep convolutional neural networks and added another fully connected layer to ensemble individual models to estimate the battery capacity. Such methods increase the accuracy, but usually, all the models need to be trained offline and directly employed during online usage. As a result, the individual models may not be optimally combined during the real deployment. Apart from that, the estimated SoH values are often used for later-on decision-making and corrective action planning in BMSs. To this end, a rigorous propagating of uncertainty through the estimation algorithm becomes indispensable [38]. Bayesian ML algorithms, such as Gaussian process regression (GPR), naturally enable uncertainty quantification and are accordingly a good candidate to apply in battery state estimation and prediction [39].

To bridge the identified research gaps, this work develops an efficient, practical, and easy-to-implement method for battery SoH estimation. First, we categorize all the possible charging processes of EVs into six feasible and mutually exclusive scenarios. For each scenario, the most relevant set of features is extracted. Secondly, two Bayesian ML algorithms and two frequentist ML algorithms are separately applied to develop

SoH estimation models, each of which is able to provide a quantitative estimation of the confidence interval. Thirdly, we develop a Kalman filter (KF)-based fusion algorithm to systematically combine all these models. The model fusion can be implemented in real-time with minor extra computation efforts. Eventually, illustrative results from a large number of batteries with different chemistry verify the efficacy of the proposed approach. Specifically, relative to individual models, the online fusion algorithm achieves highly accurate and robust results along with a significantly tightened confidence interval.

NOMENCLATURE

$(\frac{dQ}{dV})_{\max}$	The peak value on the IC curve.
t_{cc}	The time spent during CC charging stage.
E_U	The energy of the voltage signal.
ΔI_{CV}	The slope of the current during CV charging phase.
ΔV	The slope of the voltage in the defined voltage window.
Δt	Time spent in the defined voltage window.
$U_{\frac{dQ}{dV}\max}$	The voltage corresponding to when IC has peak value.
$T_{\min}$	Minimum temperature during charging.
$T_{\max}$	Maximum temperature during charging.
SoC_0	Initial SoC level when charging starts.

II. DATASET INTRODUCTION

This work aims to tackle the problem of battery SoH estimation under arbitrary vehicle operating conditions. In this regard, the dataset for model development and validation should be as close to real-world battery usage as possible. The dataset used here was acquired from Sandia National Laboratories (SNL) [40]. The experiments were initially conducted to investigate the effect of different stress factors, e.g., the discharge rate, depth of discharge, and environment temperature, on the degradation performance of commercial cells of different types of chemistry. The cells with Nickle Manganese Cobalt (NMC) and Nickle Cobalt Aluminum (NCA) as the positive electrode were selected. The detailed cell specifications are listed in Table I. Additionally, Fig. 1a and b illustrate

TABLE I
BATTERY CELL SPECIFICATIONS

Battery	NMC	NCA
Manufacture	LG Chem	Panasonic
Part number	18650 HG2	NCR18650B
Nominal capacity (Ah)	3	3.2
Nominal voltage (V)	3.6	3.6
Voltage range (V)	2–4.2	2.5–4.2
Maximum discharge current (A)	20	6
Operating temperature(°C)	-5–50	0-45

the capacity retention trend of the NMC and NCA battery cells, respectively. As all the cells were charged under 0.5C during the reference performance tests, we accumulated the capacity during the charging phase and used the calculated result as the ground truth of battery capacity. This work aims to estimate such capacity value in real-time, along with battery usage. Fig. 1c, d, and e exemplify the current, voltage, and temperature curves of a typical charge and discharge cycle, respectively. The test campaign setups are shown in Table II. As can be seen, the cells were exposed to different partial charging levels and cycled under three different environment temperatures.

TABLE II
BATTERY CELL TEST CAMPAIGN

Depth of discharge	Temperature	Discharge rate
20%–80%	25°C	0.5C
40%–60%	25°C	0.5C
0%–100%	15°C	1C
	15°C	2C
	15°C	3C
	25°C	0.5C
	25°C	1C
	25°C	2C
	25°C	3C
	35°C	1C
	35°C	2C
	35°C	3C

III. FEATURE CONSTRUCTION AND ENGINEERING

As stated earlier, nowadays, the CC-CV charging strategy is commonly used by automotive companies. Following the mainstream, we consider it the first choice to charge our studied battery cells but keep the doors open for any other charging strategy. Furthermore, all the measurements were generated at a sampling rate of 1 Hz. When the charging realization is CC-CV, the raw current and voltage data were used to extract smooth and consistent IC curves. The charging current and voltage profiles over different cycles of one typical battery under study are shown in Fig. 2a and b, respectively. The corresponding IC curves are illustrated in Fig. 2c.

A. Charging scenarios

Unlike cycling battery cells in the laboratory, with repeatable profiles under a controlled environment, practical battery usage is highly dynamic with frequent changes in the operating conditions [11]. The usage profile can become even more complex when an EV is shared by different drivers, and, in general, the driving profiles are also unpredictable. Additionally, seasonal and local temperature variations will have a large impact on the battery's performance and internal aging state. Compared to the discharging profiles, charging profiles are easy to control, and the same control law is often repeatedly used. In this regard, we use only the charging curves to estimate the capacity values. Here we eliminate the assumption commonly applied in laboratory tests that all charging cycles start from the same initial SoC level and are contained at a fixed ambient temperature. In other words, this study is not limited to rigorously repetitive CC-CV charging cycles. Instead, we consider arbitrary CC-CV charging profiles and any other charging strategies.

Based on the above discussion, we categorize all the existing charging profiles into six different and mutually exclusive scenarios, illustrated in Fig. 2a.

S1: Complete CC-CV charging from 0% SoC to 100%.

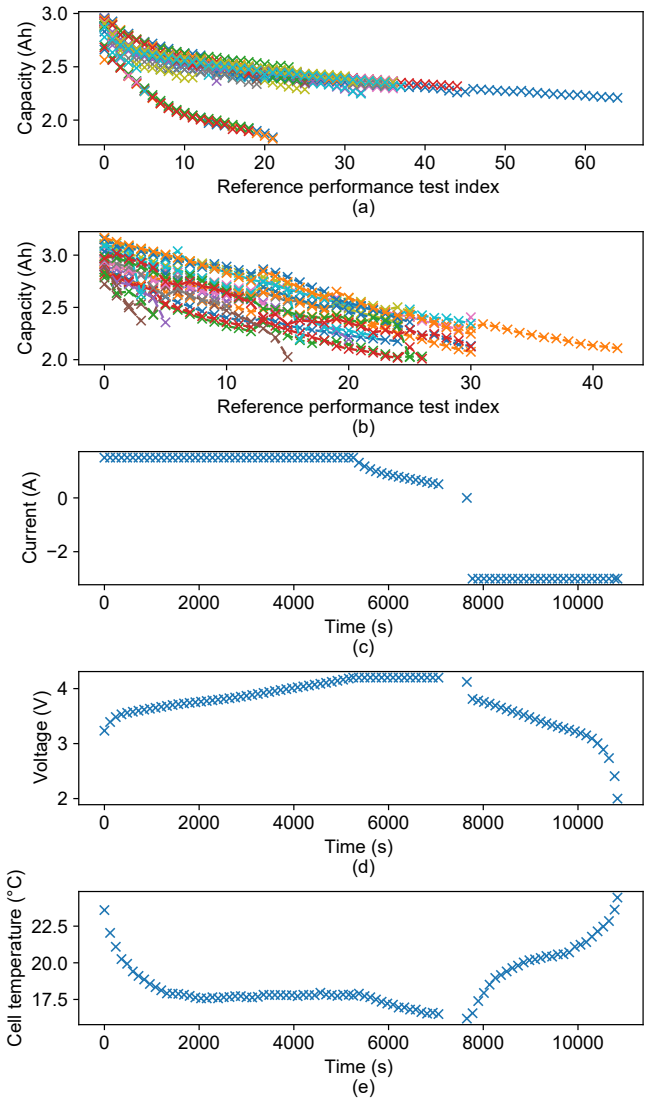

Fig. 1. Illustration of the battery dataset used in this work. (a) shows the capacity degradation curve of 31 NMC cells, and (b) represents the capacity degradation curve of 24 NCA cells. (d)–(e) exemplify a typical cycling profile of the studied battery.

S2: Partial CC-CV charging that starts before the IC peak value (see Fig. 2c) and ends with the complete CV phase (see Fig. 2a).
S3: Partial CC charging that starts before the IC peak value and ends without the CV phase.
S4: Partial CC-CV charging that starts after the IC peak value and ends with the complete CV phase.
S5: CC charging that starts after the IC peak value and ends without the CV phase.
S6: All the remaining scenarios not covered in S1–S5.

It is worth mentioning that S1–S5 covers all the scenarios associated with CC-CV charging. In addition to CC-CV charging, various charging scenarios have been proposed and tested in the literature, e.g., multistage CC charging, pulse charging, and dynamic charging [41], [42]. To enable SoH monitoring under any charging, S6 has been introduced. Note that temperature is not mentioned in any of the above scenarios, implying that common constraints imposed on the ambient temperature for SoH estimator design are removed.

B. Feature construction

For ML algorithms to work accurately and efficiently, it is of great importance to select relevant and informative features. At the same time, for the algorithms to take effect, the chosen features must be available under practical charging cycles. For the defined charging scenarios, different features will be added to the feature pool according to their availability.

Specific voltage window-related features. The voltage and current curves during the charging phase will gradually change as the battery degrades, as shown in Fig 2a–b. A predefined sub-range of the raw charging voltage data is first selected as the base to construct features. First, the time interval between a predefined voltage window is chosen since the aged cell will gradually need less time for charging than the newer cell, as can be seen in Fig. 2a. Second, the signal strength over time, often regarded as the energy of the signal, can also be used as a feature, which is calculated as

$$E = \int_0^\infty s(t)^2 \, dt, \tag{1}$$

where $s(t)$ is the signal. Third, the area under the current and voltage curves could also represent the aging state of the batteries and have previously been used by She *et al.* [21] to successfully estimate battery capacity. Last, by observing the voltage curve changes over the battery lifetime, it is noticed that the voltage curve in the CC phase becomes steeper as the battery ages. Based on this, it is natural to also select the slope of the voltage curve as one of the feature candidates. Similarly, the slope of the current curve in the CV phase could serve as a feature if there is a CV phase in the charging profile.

IC curve-related features. IC is defined as the ratio between the change in the charging capacity and the corresponding voltage change in a predefined time interval. As a promising way for non-destructive battery aging mechanism identification and battery characteristics method, IC analysis is also a source to extract informative health indicators since the underlying pattern of the IC curves changes during battery degradation. However, the specified time interval, the division calculation, and the measurement noise from current and voltage sensors will greatly impact the final IC curve. To make the obtained IC curves over the battery's entire lifetime comparable, we adopt a time interval of ten seconds and apply a KF to smooth the calculated IC curve [30]. It is noteworthy that when applying IC-related features for battery capacity estimation in real-world applications, the initial SoC level and the cell temperature can dramatically change the character of the constructed IC curve due to the cell polarization and internal resistance changes. To cope with such changing effects, we propose to select the initial charging SoC and charging temperature-related features into the feature pool. For the features extracted from the IC curve, we choose the peak value and its corresponding voltage level since their changes during the battery aging process are quite prominent, as illustrated in Fig. 2c.

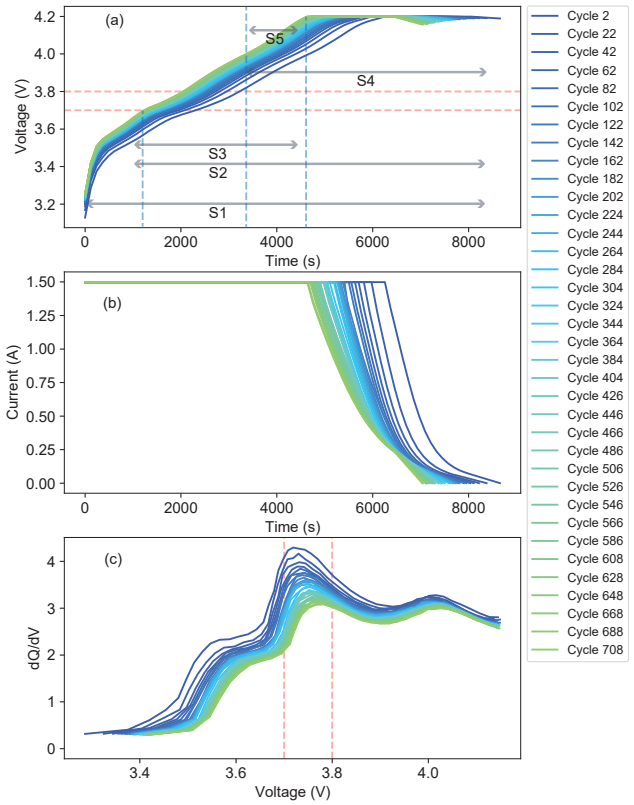

Fig. 2. Illustration of the varying health indicator as the battery gradually aged. (a) shows the change of the voltage curve along with an illustration of different charging scenarios. (b) represents the change in the current curve. (c) shows the change in the IC curve.

C. Feature engineering

Since we adopt a manual feature selection process to incorporate domain knowledge, there is a risk of selecting uncorrelated or redundant features into the feature pool. Therefore, a Spearman correlation analysis was conducted to rank the features according to their correlation with the target variable, i.e., the capacity value [43]. The Spearman correlation coefficient, ρ, is calculated by

$$d_i = R_{x_i} - R_{y_i}, \tag{2}$$

$$\rho = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \tag{3}$$

where N is the number of data samples, and R_{x_i} is the rank of x_i compare to other samples. For example, $R_{x_i} = 1$ if x_1 is the smallest value in the set $\{x_1, x_2, \cdots, x_N\}$. After that, the feature check-and-control scheme proposed in [44] is applied. The key idea is to select the features most correlated to the system output, namely the battery capacity, but remove the redundant features for developing ML models. By doing so, we expect the obtained model to be accurate, robust, and have low requirements in computation and memory.

IV. Estimator Design

The proposed estimator contains two types of internal models, with one being the capacity estimation model developed from time-series data and the other being the capacity prediction model developed from histogram-based data. Then, the KF is adopted to optimally fuse all the model results and provide the final capacity estimate. The proposed SoH estimation pipeline is illustrated in Fig. 3.

Three evaluation matrices are applied to quantify the estimation performance, namely the mean absolute percentage error (MAPE), the root mean square percentage error (RMSPE), and the 95% percentile confidence interval of the estimation result. They are mathematically defined as

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|Q_i - \hat{Q}_i|}{Q_i} \times 100\%, \tag{4}$$

$$\text{RMSPE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{Q_i - \hat{Q}_i}{Q_i}\right)^2} \times 100\%, \tag{5}$$

$$\text{CI} = \hat{Q}_i \pm 2 \times \sqrt{\sigma(\hat{Q}_i)}, \tag{6}$$

where Q_i is the measured capacity value, $\hat{Q}_i$ is the estimated capacity value, N is the total sample number in the test set, and CI is a 95.4% probability for Gaussian distribution with the covariance denoted by $\sigma(\hat{Q}_i)$.

A. Time-series data-based model for capacity estimation

As the estimation result can imply a decision, corrective actions, or carry forward to the prognostic step, the estimation uncertainty evaluation becomes indispensable. Four ML algorithms, two of which are probabilistic and two frequentist-based, are selected to develop models for capacity estimation. All of them are able to quantitatively propagate their estimation uncertainty to provide a confidence interval for their results. Henceforth, the ML algorithm output, i.e., the battery capacity Q, will be referred to as y, and the corresponding features will be represented by x. A random-search hyperparameter tuning is applied together with 5-fold cross-validation to find the optimal hyperparameters for each ML algorithm.

1) Gaussian process regression: As a non-parametric and probabilistic model, GPR attempts to learn a posterior distribution rather than a single value of the model parameter. Furthermore, instead of treating the model parameters as random variables, as in Bayesian ridge regression (BRR), GPR treats the model function $f(x)$ as a Gaussian process and uses it to compute the posterior probability $P(f(x)|y)$. Here, $f(x)$ is defined as [45]

$$f(x) \sim \mathcal{GP}(\mu(x), \kappa(x, x')),$$

where x and x' are two arbitrary feature samples, $\mu(x)$ is the mean function, and $\kappa(x, x')$ is the covariance function or, equivalently, the kernel function. Intuitively, the kernel function determines how strong the correlation is between the two samples, x and x'. Given a series of training pairs $\{\mathcal{X}, \mathcal{Y}\}$ with $\mathcal{X} = \{x_1, x_2, \cdots, x_n\}$ and $\mathcal{Y} = \{y_1, y_2, \cdots, y_n\}$, GPR predicts a new test sample x^* by conditioning $f(x)$ on the training data to find the posterior prediction distribution. GPR has high prediction accuracy, even though there is only a relatively small dataset, and is able to quantify the uncertainty

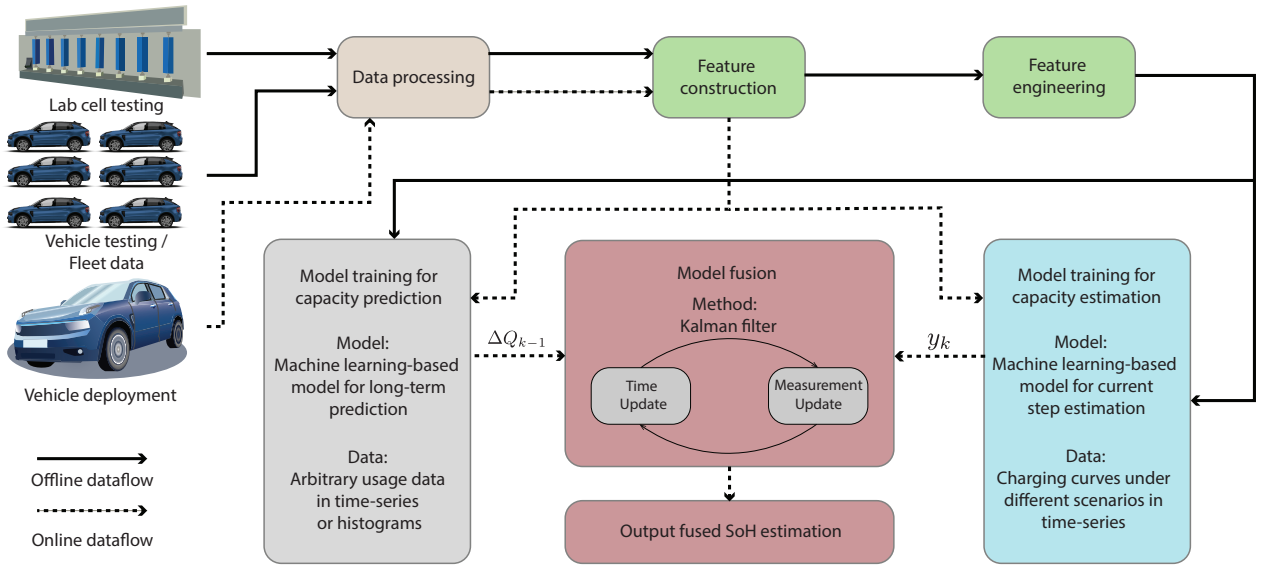

Fig. 3. The overall battery SoH estimation pipeline.

of predictions. These merits make it a popular method for battery diagnostics. However, with a large dataset or high feature dimensions, GPR becomes prohibitively computationally expensive due to matrix inversions.

2) Bayesian ridge regression: Applying a Bayesian approach to a linear regression model leads to the so-called BRR. Instead of treating the coefficient θ as a single variable, BRR assumes θ as a spherical Gaussian distribution, defined as $P(\theta) = \mathcal{N}(\theta; 0, \Sigma_0)$, having zero mean and the covariance Σ_0. Here we choose $\Sigma_0 = \mathcal{I}\alpha$ to simplify the calculation, where $\mathcal{I}$ is the identity matrix and α is a positive hyperparameter. β is defined as the noise variance of the output y and can be viewed as a regularization parameter. We assume α and β are uninformative priors subject to gamma distributions [46], [47]. By maximizing the marginal likelihood, these two parameters can be optimized during training. With α, β, and the conjugate prior, the posterior can be explicitly calculated. For detailed handling of the algorithm, the readers are directed to [48], [49]. The simplicity of training and testing and the good interpretability of the estimation result make it a preferable ML algorithm. However, the linear assumption may limit its applicability, especially for complex and nonlinear systems.

3) Random forest regression: RFR consists of a collection of different decision trees and can achieve a good bias-variance trade-off by aggregating each tree's output. Through bootstrap replicates that are randomly selected with replication of a subset of the training dataset and features to each individual tree, we can adapt the infinitesimal jackknife variance estimation method [50] to quantify the estimation confidence interval of the RFR result.

4) Deep ensemble neural network (DeNN): Deep neural networks (DNN) have recently gained great interest from both academia and industry due to their impressive performance in pattern recognition and trend identification on a wide spectrum of dynamic and complex tasks. We adopt an ensemble-based method to combine the results of each DNN model and quantify the estimation uncertainty as in [51]. We construct the final layer with two outputs, one being the predicted mean value $\mu(x)$ and the other being the predicted variance $\sigma(x)^2$. Furthermore, we select the negative log-likelihood function as the loss function.

B. Histogram data-based model for capacity prediction

For a given battery charging profile, the predefined scenarios S1–S5 may not be realized due to abnormal usage behavior (e.g., extremely shallow discharge/charge), data corruption, communication delay/faults, etc. Under such circumstances (i.e., S6), the selected features cannot be extracted, rendering all the ML models introduced in Section IV-A infeasible. To overcome the issue, we treat the capacity monitoring task at the current time step k as a prediction problem instead of estimation, using historical capacity estimates up to $k-1$ and features extracted from usage data. To enable online prediction of battery aging trajectory and lifetime with any format of raw data collected under various operating conditions, we have recently proposed a histogram data-based machine learning framework [44]. This framework fits well with the capacity prediction task and is therefore implemented recursively whenever S6 is triggered. The key idea is introduced here for completeness, while the readers are directed to the original work for details.

The end-user usage data, e.g., the accumulated discharge/charge energy throughput, discharge range, cell temperature, charging current, discharging current, voltage, and vehicle parking time, no matter in time series or histograms of any dimension, are transformed into one-dimensional (1D) histogram data. Then, a set of statistic properties of the constructed 1D histograms are extracted and used as feature candidates. After feature engineering, the selected features, denoted by x_{hd}, are employed to learn battery aging behavior. A global model is first developed offline from N labeled

input-output pairs in the available dataset, generated from a statistically significant number of batteries

$$\Delta\hat{Q}_n = f_{\text{global}}(x_{hd,n}),\ n \in \{1, \cdots, N\}, \tag{7}$$

where $\Delta\hat{Q}_n$ represents the capacity loss between two consecutive samples for any battery in the dataset. For the offline regression problem, the RFR is employed to develop the function $f_{\text{global}}()$. -

After obtaining a reasonable global model in (7), the second step is to adapt it online to any considered individual battery, indexed by $m \in \{1, \cdots, M\}$. This is achieved by expressing the degradation of cell m as $\lambda_{m,k} f_{\text{global}}(x_{hd,m,t})$, where $\lambda_{m,k}$ is a correction coefficient. With the historical capacity estimation values (e.g., obtained when any scenario of S1–S5 is active) saved in the BMS memory, a cell-individualized model for the faded capacity of battery m, is then determined by making a trade-off between the cell-corrected model, i.e., $\lambda^*_{m,k} f_{\text{global}}(x_{hd,m,t})$ and the global model, i.e., $f_{\text{global}}(x_{hd,m,t})$, where $\lambda^*_{m,k}$ is optimized online to get the best possible fit to historical data, i.e., $\{\hat{Q}_{m,1}, \cdots, \hat{Q}_{m,k-1}\}$. The capacity fade from step $k-1$ to any future time step $t \geq k$ is then expressed as

$$\begin{aligned}\Delta\hat{Q}_{m,t} = &(1 - w^*_{k,t}) f_{\text{global}}(x_{hd,m,t}) \\ &+ w^*_{k,t} \lambda^*_{m,k} f_{\text{global}}(x_{hd,m,t}),\end{aligned} \tag{8}$$

where w^*_{k-1} is a weight coefficient calculated offline to optimally trade-off between the global estimate $f_{\text{global}}(x_{hd,m,t})$ and the individually corrected estimate $\lambda^*_{m,k} f_{\text{global}}(x_{hd,m,t})$ [44]. By assigning t to $k+1$ in (8), the following one-step capacity prediction model is obtained

$$\begin{aligned}\Delta\hat{Q}_k = &(1 - w^*_{k-1,k}) f_{\text{global}}(x_{hd,k}) \\ &+ w^*_{k-1,k} \lambda^*_{k-1} f_{\text{global}}(x_{hd,k}),\end{aligned} \tag{9}$$

$$\hat{Q}_k = \hat{Q}_{k-1} + \Delta\hat{Q}_k, \tag{10}$$

where the subscript m is dropped for any arbitrary battery during online deployment.

C. Optimal model fusion

The ML models introduced in the above two subsections have their advantages and disadvantages. The best-performing algorithm may vary depending on different datasets and operating conditions. Fusing the results of all algorithms may therefore give a more accurate and reliable estimation of battery capacity. We use a KF to optimally combine the estimation models in Section IV-A and the prediction model (9)-(10) in Section IV-B. The overall estimation pipeline is shown in Fig. 4.

By defining the process noise as w and the measurement noise as v, the dynamic system of battery capacity is formulated as

$$Q_k = Q_{k-1} + \Delta Q_{k-1} + w_{k-1}, \tag{11}$$

$$\mathbf{y}_k = C Q_k + v_k. \tag{12}$$

When the true capacity is not measured during operations, the system output $\mathbf{y}$ is a vector of the estimation results from the ML models introduced in Section IV-A and is defined by

$$\mathbf{y} = \begin{bmatrix}\hat{Q}_{\text{GPR}} & \hat{Q}_{\text{BRR}} & \hat{Q}_{\text{RFR}} & \hat{Q}_{\text{DeNN}}\end{bmatrix}^T \tag{13}$$

implying that $C = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^T$. Σ_w and Σ_v are covariances of the process noise and the measurement noise, respectively. The covariance of the measurement noise is composed of the uncertainty quantifications of the estimates using four ML models.

Assumption 1: The noise w and v are uncorrelated [52].

Justification 1: The estimation model (12) is developed from time-series features (extracted in Section III) using the probabilistic or frequentist-based ML algorithms. In contrast, the prediction model (11) is obtained from usage-related histogram-based features using RFR. Therefore, with different inputs, ML models, and training processes, the corresponding noise terms w and v are naturally uncorrelated.

Assumption 2: w and v are zero-mean, white, and Gaussian noise [52].

Justification 2: All the estimations conducted to obtain the system output $\mathbf{y}$ are independent and random, so as the process to obtain the system input ΔQ_k. Therefore we can assume v_k and w_k are at least close to white noise. However, for the highly nonlinear battery system and ML-based estimation/prediction models, analytically proving the zero-mean, whiteness, and Gaussian properties of the two noise terms is difficult and not pursued here. Instead, numerical justification will be conducted, with the result presented in Section V.

Based on the nominal model of the dynamic system (11)–(12), a standard KF is designed, i.e.,

$$\hat{Q}^-_k = \hat{Q}^+_{k-1} + \Delta Q_{k-1}, \tag{14}$$

$$P^-_k = P^+_{k-1} + \Sigma_{w,k}, \tag{15}$$

$$K_k = P^-_k C^T_k (C_k P^-_k C^T_k + \Sigma_{v,k})^{-1}, \tag{16}$$

$$\hat{Q}^+_k = \hat{Q}^-_k + K_k(\mathbf{y}_k - C_k \hat{Q}^-_k), \tag{17}$$

$$P^+_k = (\mathcal{I} - K_k C_k) P^-_k, \tag{18}$$

where P is the state covariance matrix, the superscripts $-$ and $+$ signify the prior and posterior, respectively, and K_k is the Kalman gain.

Remark 1: Although each of the used estimation and prediction models can deliver an estimation of the capacity, based on *Assumptions* 1 and 2, the KF fuses each model's results optimally, resulting in increased estimation accuracy and robustness to noise.

Remark 2: During online deployment, the ground truth of battery capacity is seldom available from today's BMS. Then, the KF will, on most occasions, work as an open-loop filter to fuse the estimation results, i.e., $\mathbf{y}_k$ defined in (13), and the prediction result $\hat{Q}^-_k$ according to their uncertainties.

The overall procedure of the SoH estimation pipeline is shown in Algorithm 1. It is worth mentioning that the estimation model, (12)–(13), cannot be executed under S6 due to the lack of available time-series-based feature input. On the contrary, the prediction model (11), using histogram-usage-related features, can be running all the time. Therefore, the

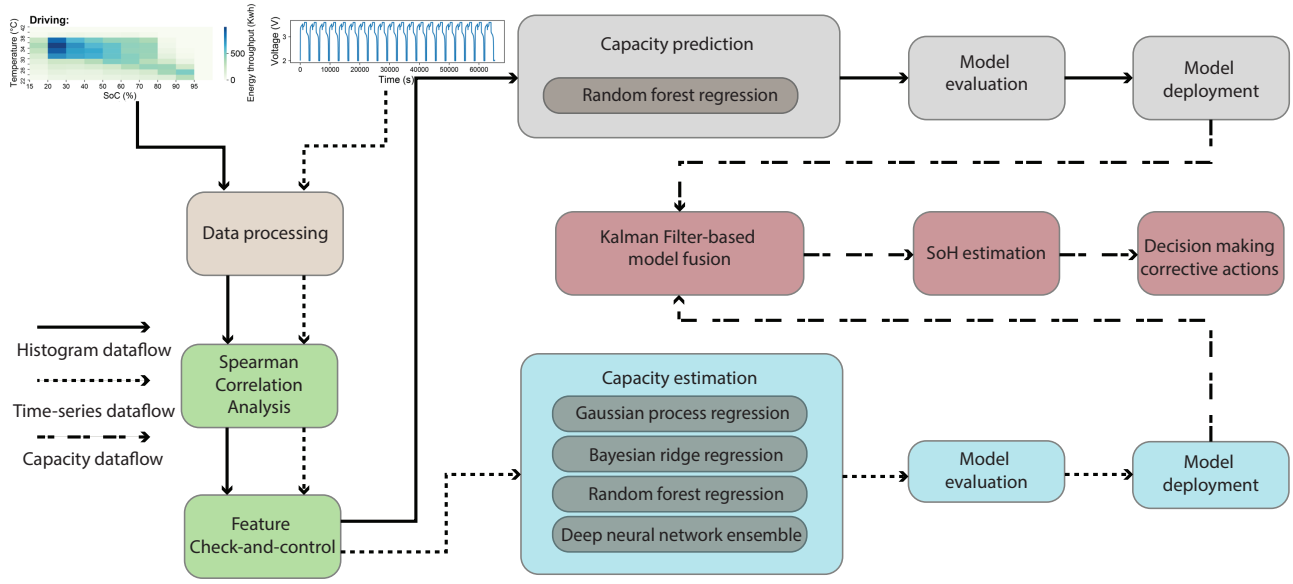

Fig. 4. Illustration of the capacity estimation method.

implementation of the KF-based fusion algorithm will be adjusted according to different charging scenarios.

Algorithm 1 KF-based fusion algorithm for battery SoH estimation

Require: Measured I, V, T
$C \leftarrow \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^T$
if $s \in \{S1, S2, S3, S4, S5\}$ **then** ▷ According to the measured I, V, T to decide which scenario s belongs to.
$y_k \leftarrow \begin{bmatrix}\hat{Q}_{\text{GPR}} & \hat{Q}_{\text{BRR}} & \hat{Q}_{\text{RFR}} & \hat{Q}_{\text{DeNN}}\end{bmatrix}^T$ ▷ estimation model
$\Delta Q_{k-1} \leftarrow (1 - w^*_{k-1,k}) f_{\text{global}}(x_{hd,k}) + w^*_{k-1,k} \lambda^*_{k-1} f_{\text{global}}(x_{hd,k})$
$\hat{Q}_k^- \leftarrow \hat{Q}_{k-1}^+ + \Delta Q_{k-1}$ ▷ prediction model
$P_k^- \leftarrow P_{k-1}^+ + \Sigma_{w,k}$
$K_k \leftarrow P_k^- C_k^T (C_k P_k^- C_k^T + \Sigma_{v,k})^{-1}$
$\hat{Q}_k^+ \leftarrow \hat{Q}_k^- + K_k(\mathbf{y}_k - C_k \hat{Q}_k^-)$
$P_k^+ \leftarrow (\mathcal{I} - K_k C_k) P_k^-$
else if $s \in \{S6\}$ **then**
$\hat{Q}_k^- \leftarrow \hat{Q}_{k-1}^+ + \Delta Q_{k-1}$ ▷ prediction model
$\hat{Q}_k^+ \leftarrow \hat{Q}_k^-$
end if

V. Results and Discussion

A. Hyperparameter values

We have adopted random search hyperparameter tuning for all the employed ML algorithms. The corresponding hyperparameter ranges are reported in Table III. The coefficient of determination (R^2) is adopted here to evaluate the performance of each model hyperparameter and is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{N}(Q_i - \bar{Q})^2}, \tag{19}$$

where $\bar{Q} = \frac{1}{N}\sum_{i=1}^{N}\hat{Q}_i$.

TABLE III
Hyperparameter value range used in random search hyperparameter tunning

GPR	Kernel function: RBF kernel, 2/3 Matérn kernel, 2/5 Matérn kernel, rational quadratic kernel, periodic kernel, and the dot product kernel
BRR	α_1: $[10^{-6}, 10^{-3}]$ α_2: $[10^{-6}, 10^{-3}]$ β_1: $[10^{-6}, 10^{-3}]$ β_2: $[10^{-6}, 10^{-3}]$
RFR	Maximum features: $[3, 5]$ Maximum tree depth: $[3, 10]$ Maximum sample split: $[3, 9]$ Maximum sample leaf: $[5, 15]$
DeNN	Hidden unit number: $[8, 32]$ Drop out rate: $[0.05, 0.1]$ Learning rate: $[10^{-4}, 10^{-1}]$

B. Result of feature engineering

According to the feature selection and engineering procedure presented in Section III, the selected features for each charging scenario of the NMC cells are shown in Fig. 5. Noticeably, when the IC peak value, denoted as $(\frac{dQ}{dV})_{\max}$, exists in any given charging profile, it is always selected and ranked the highest, as seen for S1, S2, and S3. This means $(\frac{dQ}{dV})_{\max}$ is the most representative feature among others for the cell aging state. This finding is in line with the previous ICA studies that demonstrated $(\frac{dQ}{dV})_{\max}$ being a good indicator for battery health diagnostic [21], [53]. Another finding is that when partial charging profiles are used, the initial SoC level is selected as one of the critical features (see the sub-figures for S3 and S5). This is expected since varying the initial SoC dramatically changes the corresponding voltage behavior because of the polarization effect, even if the batteries experience the same current and operating temperature. It also verifies the aforementioned argument about maintaining the practicability

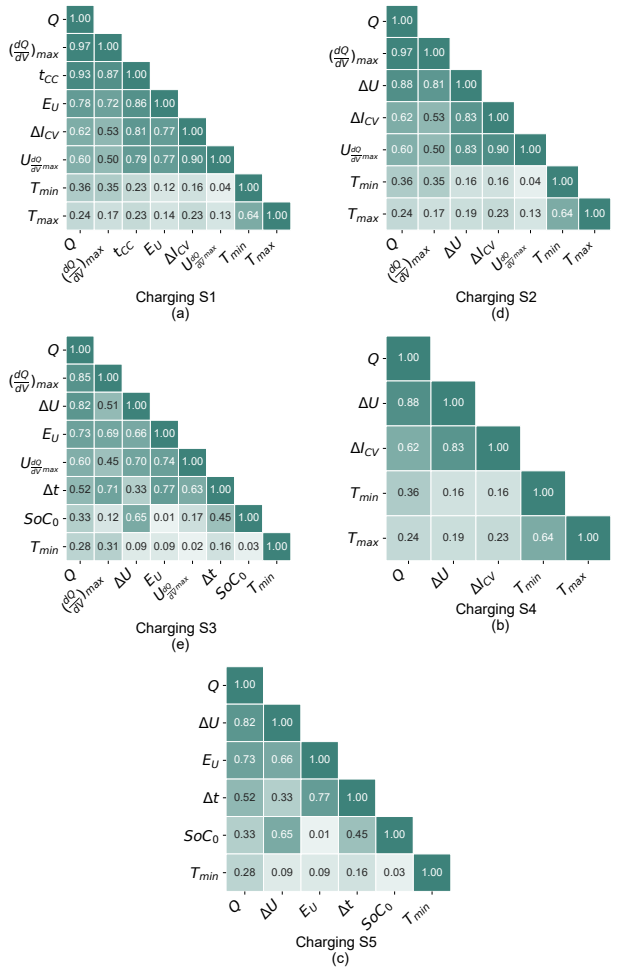

Fig. 5. Feature engineering results for NMC-type battery cells.

of SoH estimators developed upon partial charging profiles. Temperature-related features are also selected, and they are heavily overlapped among different charging scenarios. This result echoes the fact that temperature is an important stress factor for battery degradation.

Sorting from S1 to S6, the available charging profiles become narrower, and correspondingly, the available features in the pool satisfying our specified selection criteria become less, implying a harsher situation for the ML models to learn battery characteristics.

C. SoH estimation results under lab tests

To examine the efficacy of the developed models, we first demonstrate their performance for each charging scenario, i.e., applying them to the data where the cells have been receiving the same charging scenario throughout their entire lifetime. The partition of training and test datasets is different for the six considered charging scenarios. For example, for S1, partial charging data cannot be used to train the ML models.

The estimation errors of different SoH estimation algorithms are first quantitatively studied for NMC-type batteries, with the results listed in Table IV. It can be seen that from S1 to S5, all the estimation ML models derived, and the KF, achieve reasonable estimates, with a MAPE of 0.629% for the best case and 2.27% for the worst case. When the batteries are operated under S6 over their entire lifespan, albeit a very rare case for EV applications, the proposed model fusion method can estimate the capacity trajectory at a MAPE of 3.899%. It is worth mentioning that under S6, the estimation models (12) derived by GPR, BRR, RFR, and DeNN are infeasible, and then the capacity estimation can only be performed by the prediction model (11). The relatively poor estimation result is simply attributed to the fact that a very small and sparse set of training data has been used. From the table, it can also be observed that for the six scenarios, when there are more qualified features in the pool, the estimation tends to become more accurate. This means one should always conduct the capacity estimation under the feature-richest scenario.

The results in Table IV also verify the superiority of the proposed KF-based fusion method—it generally performs better than or as well as the best-performing individual model. To look more closely, Fig. 6 illustrates that under the first five scenarios, the KF rarely exceeds $\pm 5\%$ error, while under S1–2, all the estimation results are within $\pm 2.5\%$ error; From the estimated capacity trajectories of Fig. 7 an instantiated for a randomly selected cell under S1, KF follows the measured capacity better than the best individual ML model, i.e., RFR in this case. The obtained numerical results are, to a large extent, consistent with *Remark* 1. However, the global optimality is not achieved by the KF with the employed battery dataset. This is because the zero-mean part for the measurement noise v in *Assumption* 2 does not hold. As demonstrated in Fig. 7c, the mean values slightly deviate from zero and are located in a range of $[0.4\%, 0.6\%]$ for the four estimation models in (12). The remaining part of *Assumption* 2 is valid, as both v and w have a (nearly) Gaussian distribution which can be seen from Fig. 7c and [44] (i.e., Fig. 7e therein), respectively.

All the models we developed here can provide a 95 percentile estimation confidence interval, as demonstrated in Fig. 8, therefore providing valuable information for predictive battery maintenance and usage optimization. Moreover, by fusing the estimation results from all the individual models using KF, the confidence interval is considerably tightened, in other words reducing the uncertainty significantly. As shown in Fig. 9, the standard deviation of the estimates from KF, σ_{KF}, is always smaller than that of any individual model. After the seventh index, the standard deviations of the best individual model and the worst model are 2 and 5 times larger than σ_{KF}, respectively.

Furthermore, the proposed KF-fused algorithm was compared with three state-of-the-art fusion algorithms for battery SoH estimation, including RFR-based fusion [36], NN-based fusion [37], and ELM-based fusion algorithms [54]. Noteworthy, to have a fair and meaningful comparison, for the algorithms using RFR and NN to fuse the estimation results, all the individual ML models to be fused are kept the same as the one in the Kalman filter. To implement the ELM-based fusion algorithm, 10 one-layer NN models are used and fused according to the method introduced in [54]. The detailed numerical results are presented in Table V.

It can be seen that the developed KF-based fusion algo-

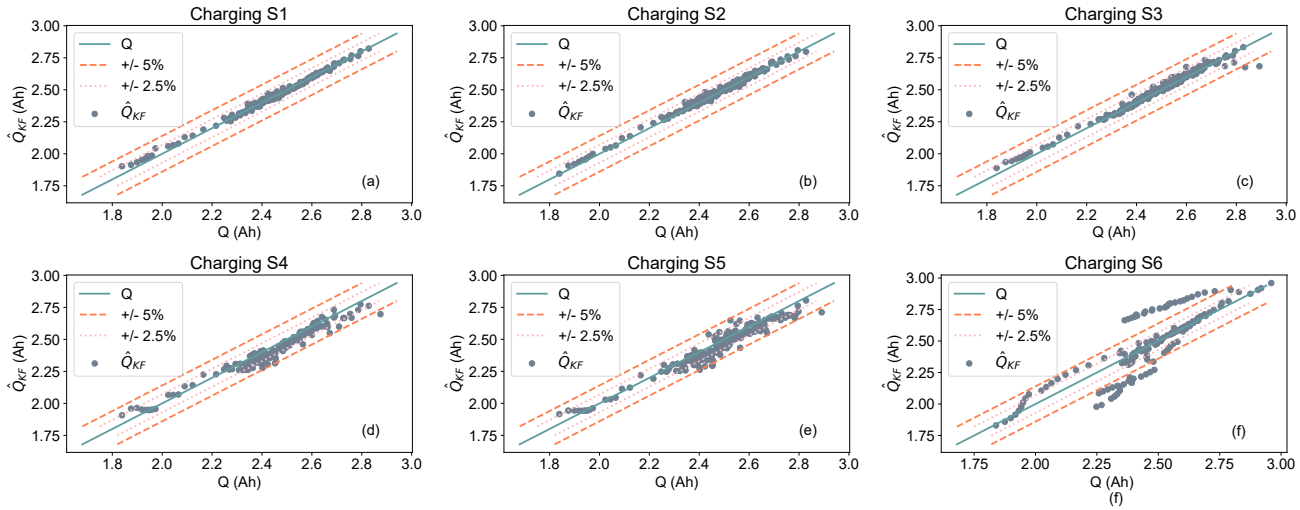

Fig. 6. The capacity estimation results for the NMC-type battery under different charging scenarios.

TABLE IV
RESULTS OF DIFFERENT SOH ESTIMATION ALGORITHMS FOR NMC BATTERIES UNDER CHARGING SCENARIOS S1–S6

Algorithms	S1		S2		S3		S4		S5		S6	
	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE
GPR	0.693	1.073	0.887	1.323	0.915	1.491	**1.532**	**1.999**	2.110	2.757	-	-
BRR	0.772	1.182	0.799	1.009	1.025	1.527	1.984	2.598	2.270	2.910	-	-
RFR	0.632	**0.857**	0.874	1.202	0.877	1.231	1.817	2.312	**1.725**	2.262	-	-
DeNN	0.919	1.418	0.816	1.114	1.027	1.471	1.576	2.132	1.860	2.334	-	-
KF	**0.629**	0.861	**0.714**	**0.880**	**0.751**	**1.178**	1.662	2.110	1.731	**2.229**	**3.899**	**5.611**

TABLE V
RESULTS OF DIFFERENT FUSION ALGORITHMS FOR NMC BATTERIES UNDER CHARGING SCENARIOS S1–S5

Algorithms	S1		S2		S3		S4		S5	
	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE
RFR fusion	0.836	1.193	0.920	1.273	1.004	1.444	**1.515**	**2.003**	2.200	2.761
NN fusion	1.078	1.811	1.011	1.456	1.059	1.631	1.644	2.091	2.002	2.472
ELM fusion	2.196	2.852	2.796	3.864	1.553	2.148	2.271	2.862	2.396	3.092
KF	**0.629**	**0.861**	**0.714**	**0.880**	**0.751**	**1.178**	1.662	2.110	**1.731**	**2.229**

rithm clearly outperforms all its alternatives in most cases. This advantage is attributed to the fact that the KF fuses the prediction model and a library of individual estimation models at each time step based on their real-time estimation confidences. In comparison, the available fusion algorithms for SoH estimation are usually trained offline and then deployed in real-world usage without online adaptation in accordance with the battery's actual behavior. Under certain cases, e.g., S4, where *Assumption* 2 cannot hold, the RFR-based fusion algorithm is slightly superior to the KF-based fusion algorithm by 0.107% in terms of RMSPE.

D. SoH estimation results under vehicle usage

In practice, it is rather rare that a battery only undergoes one charging profile during its whole lifetime. Therefore, it is worthwhile to explore the estimation performance when various charging profiles are exhibited on one specific battery. Given no cell in the database was cycled using multiple charging profiles, we selected the cells experienced with the full CC-CV charging and then manually cut their profiles to mimic partial charging. By doing so, we can demonstrate and compare the effectiveness of different models under practical charging cases. Indeed, the corresponding results will not reflect the real-world usage precisely, but the focus here is more placed on the qualitative results. The charging profiles of each cell are supposed to rotate among the six scenarios periodically during its whole lifetime. The detailed rotating protocols, as well as the results, are given in Table VI. The corresponding features are extracted according to each scenario's availability. Specifically, for S6, since no time-series feature is available, it is not feasible for the individual estimation models (12) to estimate the capacity value. In such a case, whenever S6 is triggered, the capacity remains unchanged from its previous time-step estimate.

It can be seen from Table VI that within each protocol, the KF always provides better estimation results than any individual ML model. When S6 is activated more frequently, the

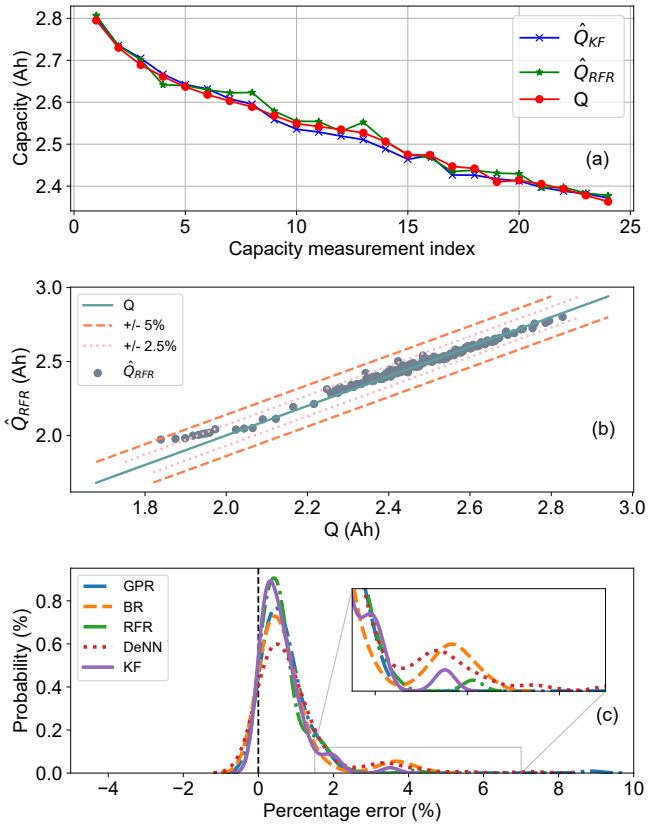

Fig. 7. Estimation results of a randomly selected NMC-type cell under scenario S1.

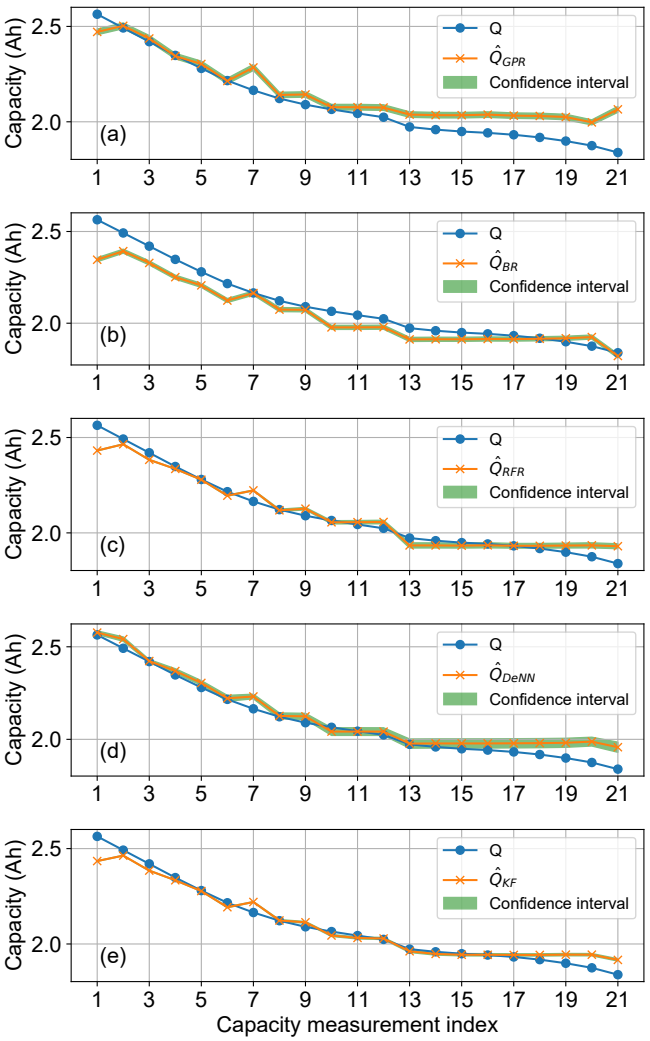

Fig. 8. Estimation results of a randomly selected NMC-type cell.

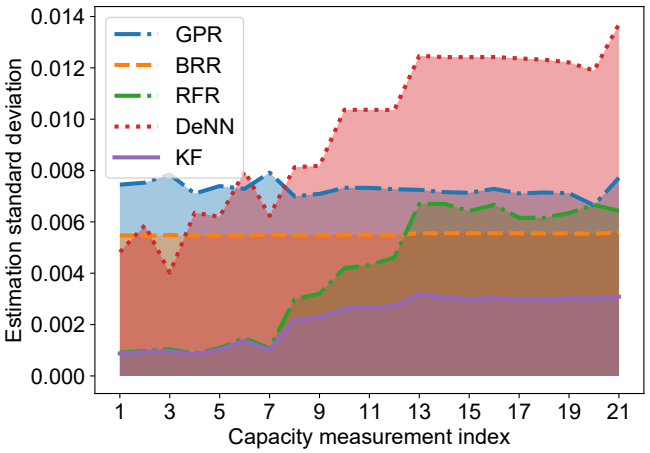

Fig. 9. The standard deviation of the results from the individual ML model and the model fusion for a randomly selected cell.

estimation results of the individual models generally become worse. Most of their estimates deviate from the measurements with a RMSPE greater than 2%, which is unacceptable for vehicle applications. On the contrary, the KF is still very reliable and continuously follows the ground truth at around 1% RMSPE. By comparing different protocols and comparing the results under lab tests to the vehicle usage case, it can be concluded that the advantage of the KF becomes more significant, particularly when S6 appears more times. This is because the KF incorporates a capacity prediction model in (11) and four capacity estimation models in (12). Even though the estimation models lose efficacy under S6, the prediction model developed from histogram data under arbitrary operating conditions is always able to gauge the capacity, and so does the KF that, in such a case, works as an open-loop model-based predictor based on battery usage information. These results corroborate the necessity to use the KF in real-world vehicle battery usage.

By looking into the detailed estimation result shown in Fig. 10b and c, the best-performing individual model, i.e., GPR in this case, still has a few estimation points violating the $\pm 2.5\%$ error bounds, whereas the KF manages to contain all estimation points within such error bounds. Under-estimating the capacity will lead to conservative usage, while over-estimation can cause abuse and, in extreme conditions, trigger thermal safety issues. As illustrated in Fig. 10d, the KF has a narrow probability distribution of the estimation errors, closest to zero among all the estimators. This means the KF is much less likely to under- or over-estimate the capacity than all the individual methods.

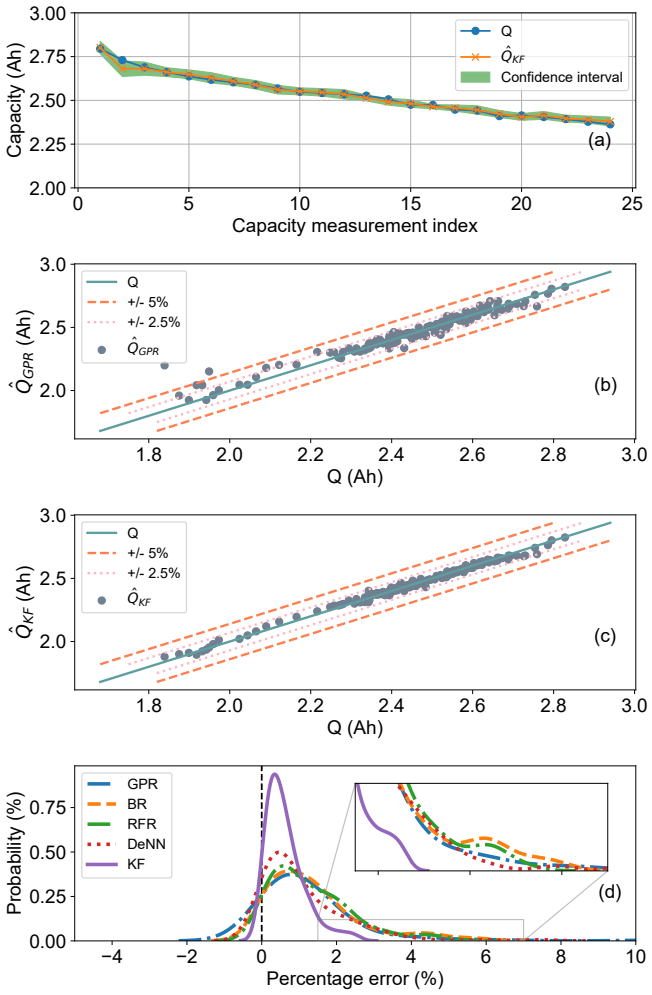

Fig. 10. The results of a randomly selected NMC-type battery capacity estimation under practical charging cycles.

Since voltage window plays an important role in the developed method, especially when dealing with partial charging scenarios in which it is impossible to select all the voltage windows for feature extraction. With this consideration, the effects of different voltage windows on the estimation performance are investigated. To perform a fair comparison, only voltage window-related features and temperature-related features are included. The detailed numerical results are shown in Table VII, where only the cells that have undergone S1 are used in order to access any of the voltage windows. Overall, with a 100 mV voltage interval, the KF's results are all acceptable and meet the average industrial standard of a 2% MAPE, demonstrating the practicability of the proposed model fusion method. It is worth mentioning that the results acquired in Table IV for different scenarios have adopted the voltage window of 3.65–3.75 V for the voltage window-related features (see SectionIII-B). As can be seen from Table VII, the designed estimators, including both the KF and individual models, will perform much better when the voltage window of 4.05–4.15 V is selected. One hypothesis of such superiority is because this voltage window coincidentally covers an IC peak value, as can be seen from Fig.2c, and meanwhile, it is also close to the charging profile transition from the CC to CV phase. Nevertheless, a lot of vehicle usage may not fall into such a voltage window, which means that fewer data samples are available.

TABLE VI
SOH ESTIMATION RESULTS UNDER VARIOUS PRACTICAL CHARGING SCENARIOS

Algorithms	Protocol 1: Periodically repeat $\{S1, \cdots, S5, S6\}$	
	MAPE	RMSPE
GPR	1.337	2.408
BR	1.373	1.847
RFR	1.314	1.706
DeNN	1.092	1.58
KF	**0.631**	**0.813**
	Protocol 2: Periodically repeat $\{S1, \cdots, S5, S6, S6\}$	
GPR	1.486	2.278
BR	1.723	2.245
RFR	1.710	2.224
DeNN	1.327	1.820
KF	**0.648**	**0.914**
	Protocol 3: Periodically repeat $\{S1, \cdots, S5, S6, S6, S6\}$	
GPR	1.678	2.598
BR	1.843	2.366
RFR	1.810	2.297
DeNN	1.310	1.806
KF	**0.667**	**0.846**
	Protocol 4: Periodically repeat $\{S1, \cdots, S5, S6, S6, S6, S6, S6\}$	
GPR	2.043	2.999
BR	2.271	3.045
RFR	2.199	2.820
DeNN	1.706	2.387
KF	**0.797**	**1.031**

E. SoH estimation results for NCA batteries

To verify the applicability of the developed method to different types of cell chemistry, we also applied it to the NCA-type batteries, and the results are shown in Table VIII. In general, the conclusions that we drew on the NMC-type batteries are also valid here. This indicates that the developed method is sufficiently general to cover batteries with different chemistry and cycling conditions. Comparatively, the estimation accuracy is slightly worse than the result we achieved for the NMC-type batteries, especially for the partial charging cases. One hypothesis of this is that the number of data samples in the training set for the NCA-type battery dataset is less than the ones used for the NMC-type batteries. The other possible reason could be that the variance between cells undergoing similar cycling profiles is generally larger than for the NMC-type battery dataset, which naturally increases the estimation challenge.

F. Discussion on real-world applications

Similar to other data-driven methods, a sufficiently large and rich dataset is required to train the proposed SoH estimation algorithm, where the data richness means that as many real-world usage scenarios as possible are covered for learning the battery's characteristics. Otherwise, if using a small dataset, the calculated confidence interval by the individual ML algorithm may be underestimated and not cover the true

TABLE VII
SOH ESTIMATION RESULTS USING DIFFERENT VOLTAGE WINDOWS FOR NMC-TYPE BATTERY

Algorithms	3.55–3.65 (V)		3.65–3.75 (V)		3.75–3.85 (V)		3.85–3.95 (V)		3.95–4.05 (V)		4.05–4.15 (V)	
	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE
GPR	2.176	2.601	1.840	2.374	0.708	0.857	**0.889**	**1.184**	0.657	0.791	0.536	0.673
BR	2.062	2.642	2.118	2.776	0.810	1.032	1.846	2.413	0.789	0.971	0.642	0.764
RFR	**1.763**	**2.308**	**1.751**	**2.321**	0.830	1.046	0.970	1.264	0.766	0.986	0.672	0.884
DeNN	3.028	4.379	2.397	2.794	1.037	1.405	1.242	1.805	1.062	1.487	0.822	1.296
KF	1.98	2.405	1.853	2.344	**0.701**	**0.856**	0.894	1.195	**0.625**	**0.759**	**0.532**	**0.672**

TABLE VIII
RESULT OF DIFFERENT SOH ESTIMATION ALGORITHMS FOR NCA BATTERIES

Algorithms	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6	
	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE
GPR	0.493	0.672	1.247	1.542	2.234	3.097	1.139	1.478	2.100	3.038	-	-
BR	1.014	1.277	1.154	1.431	3.295	6.308	1.316	1.594	3.197	4.684	-	-
RFR	0.884	1.135	1.131	1.496	2.533	4.152	1.013	1.331	2.518	3.707	-	-
DeNN	0.845	1.083	1.475	1.946	**1.952**	**2.662**	1.248	1.687	**1.962**	**2.809**	-	-
KF	**0.493**	**0.671**	**1.024**	**1.332**	2.399	3.702	**0.997**	**1.321**	2.274	3.375	**4.124**	**5.392**

value. Additionally, when the assumptions underpinning the KF do not hold, the proposed SoH estimation algorithm cannot guarantee global optimality and, thus, may not always be better than the best individual ML model. Moreover, when S6 is triggered repeatedly over the entire battery lifespan, the designed individual estimation models cannot work, while the KF's error may be too large for some applications. Under such circumstances, new features should be constructed carefully from S6. Indeed, the proposed estimation algorithm is generic and can incorporate any new features for improved estimation.

VI. CONCLUSION

This paper developed a practical method to estimate battery SoH accurately under arbitrary usage profiles. The technical contributions arise from three aspects. First, all possible charging profiles were categorized into six feasible and mutually excluded scenarios, for each of which relevant features were extracted from time series or histogram data. Second, for each scenario, Bayesian and frequentist-based ML algorithms were used individually to derive SoH estimation models using time-series data. Last, based on an SoH prediction model previously derived from histogram data, a KF was applied to systematically fuse all the models' results.

Experimental data from batteries of different types illustrated that the proposed model fusion method is able to considerably increase the estimation accuracy and robustness while significantly tightening the confidence interval of the estimation result. For example, the estimates obtained for the entire lifespan of NMC battery cells had a MAPE of 0.631% under a practical operating protocol (Protocol 1), and the error was less than 0.8% for all the studied protocols.

REFERENCES

[1] "Road transport: Reducing CO2 emissions from vehicles," https://ec.europa.eu/clima/policies/transport/vehicles_en, 2017, accessed: 14-April-2021.

[2] "The Future is Electric," https://group.volvocars.com/company/innovation/electrification, 2019, accessed: 14-April-2021.

[3] B. Nykvist and M. Nilsson, "Rapidly falling costs of battery packs for electric vehicles," *Nat. Clim. Change*, vol. 5, no. 4, pp. 329–332, Apr. 2015.

[4] R. Schmuch, R. Wagner, G. Hörpel, T. Placke, and M. Winter, "Performance and cost of materials for lithium-based rechargeable automotive batteries," *Nat. Energy*, vol. 3, no. 4, pp. 267–278, Apr. 2018.

[5] Z. P. Cano, D. Banham, S. Ye, A. Hintennach, J. Lu, M. Fowler, and Z. Chen, "Batteries and fuel cells for emerging electric vehicle markets," *Nat. Energy*, vol. 3, no. 4, pp. 279–289, Apr. 2018.

[6] 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," *Renew. Sustain. Energy Reviews*, vol. 56, pp. 572–587, Apr. 2016.

[7] Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, and H. E. Hoster, "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," *Renew. Sustain. Energy Reviews*, vol. 113, p. 109254, Oct. 2019.

[8] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," *Joule*, vol. 4, no. 2, pp. 310–346, Feb. 2020.

[9] W. Vermeer, G. R. Chandra Mouli, and P. Bauer, "A comprehensive review on the characteristics and modeling of lithium-ion battery aging," *IEEE Trans. Transp. Electrific.*, vol. 8, no. 2, pp. 2205–2232, 2022.

[10] X. Hu, H. Yuan, C. Zou, Z. Li, and L. Zhang, "Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10 319–10 329, 2018.

[11] V. Sulzer, P. Mohtat, A. Aitio, S. Lee, Y. T. Yeh, F. Steinbacher, M. U. Khan, J. W. Lee, J. B. Siegel, A. G. Stefanopoulou, and D. A. Howey, "The challenge and opportunity of battery lifetime prediction from field data," *Joule*, p. S2542435121002932, Jul. 2021.

[12] R. Xiong, L. Li, and J. Tian, "Towards a smarter battery management system: A critical review on battery state of health monitoring methods," *J. Power Sources*, vol. 405, pp. 18–29, Nov. 2018.

[13] M. Lucu, E. Martinez-Laserna, I. Gandiaga, and H. Camblong, "A critical review on self-adaptive Li-ion battery ageing models," *J. Power Sources*, vol. 401, pp. 85–101, Oct. 2018.

[14] T. R. Tanim and C. D. Rahn, "Aging formula for lithium ion batteries with solid electrolyte interphase layer growth," *J. Power Sources*, vol. 294, pp. 239–247, Oct. 2015.

[15] X.-G. Yang, Y. Leng, G. Zhang, S. Ge, and C.-Y. Wang, "Modeling of lithium plating induced aging of lithium-ion batteries: Transition from linear to nonlinear aging," *J. Power Sources*, vol. 360, pp. 28–40, Aug. 2017.

[16] I. Laresgoiti, S. Käbitz, M. Ecker, and D. U. Sauer, "Modeling mechanical degradation in lithium ion batteries during cycling: Solid electrolyte interphase fracture," *J. Power Sources*, vol. 300, pp. 112–122, Dec. 2015.

[17] R. Gu, P. Malysz, H. Yang, and A. Emadi, "On the suitability of electrochemical-based modeling for lithium-ion batteries," *IEEE Trans. Transp. Electrific.*, vol. 2, no. 4, pp. 417–431, 2016.

[18] X. Sui, S. He, S. B. Vilsen, J. Meng, R. Teodorescu, and D.-I. Stroe, "A review of non-probabilistic machine learning-based state of health estimation techniques for lithium-ion battery," *Appl. Energy*, vol. 300, p. 117346, Oct. 2021.

[19] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," *Nat. Mach. Intell.*, vol. 2, no. 3, pp. 161–170, Mar. 2020.

[20] Y. Li, D.-I. Stroe, Y. Cheng, H. Sheng, X. Sui, and R. Teodorescu, "On the feature selection for battery state of health estimation based on charging–discharging profiles," *J. Energy Storage*, vol. 33, p. 102122, Jan. 2021.

[21] C. She, Y. Li, C. Zou, T. Wik, Z. Wang, and F. Sun, "Offline and online blended machine learning for lithium-ion battery health state estimation," *IEEE Trans. Transp. Electrific.*, vol. 8, no. 2, pp. 1604–1618, 2022.

[22] M. Lelie, T. Braun, M. Knips, H. Nordmann, F. Ringbeck, H. Zappen, and D. Sauer, "Battery Management System Hardware Concepts: An Overview," *Appl. Sci.*, vol. 8, no. 4, p. 534, Mar. 2018.

[23] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Battery health prediction under generalized conditions using a Gaussian process transition model," *J. Energy Storage*, vol. 23, pp. 320–328, Jun. 2019.

[24] Z. Deng, X. Hu, X. Lin, L. Xu, Y. Che, and L. Hu, "General Discharge Voltage Information Enabled Health Evaluation for Lithium-Ion Batteries," *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 3, pp. 1295–1306, Jun. 2021.

[25] W. Li, N. Sengupta, P. Dechent, D. Howey, A. Annaswamy, and D. U. Sauer, "Online capacity estimation of lithium-ion batteries with deep long short-term memory networks," *J. Power Sources*, vol. 482, p. 228863, Jan. 2021.

[26] C. She, Z. Wang, F. Sun, P. Liu, and L. Zhang, "Battery Aging Assessment for Real-World Electric Buses Based on Incremental Capacity Analysis and Radial Basis Function Neural Network," *IEEE Trans. Ind. Inf.*, vol. 16, no. 5, pp. 3345–3354, May 2020.

[27] M. Berecibar, M. Garmendia, I. Gandiaga, J. Crego, and I. Villarreal, "State of health estimation algorithm of $LiFePO_4$ battery packs based on differential voltage curves for battery management system application," *Energy*, vol. 103, pp. 784–796, 2016.

[28] Z. Wei, H. Ruan, Y. Li, J. Li, C. Zhang, and H. He, "Multi-Stage State of Health Estimation of Lithium-ion Battery with High Tolerance to Heavily-Partial Charging," *IEEE Trans. Power Electron.*, pp. 1–1, 2022.

[29] J. Tian, R. Xiong, W. Shen, J. Lu, and X.-G. Yang, "Deep neural network battery charging curve prediction using 30 points collected in 10 min," *Joule*, vol. 5, no. 6, pp. 1521–1534, Jun. 2021.

[30] B. Jiang, H. Dai, and X. Wei, "Incremental capacity analysis based adaptive capacity estimation for lithium-ion battery considering charging condition," *Appl. Energy*, vol. 269, p. 115074, Jul. 2020.

[31] M. Mitici, B. Hennink, M. Pavel, and J. Dong, "Prognostics for lithium-ion batteries for electric vertical take-off and landing aircraft using data-driven machine learning," *Energy and AI*, vol. 12, p. 100233, 2023.

[32] S. K. Rahimian and Y. Tang, "A Practical Data Driven Battery State of Health Estimation for Electric Vehicles," *IEEE Trans. Ind. Electron.*, p. 10.

[33] C. Zhang, Y. Kang, B. Duan, Z. Zhou, Q. Zhang, Y. Shang, and A. Chen, "An adaptive battery capacity estimation method suitable for random charging voltage range in electric vehicles," *IEEE Trans. Ind. Electron.*, pp. 1–1, 2021.

[34] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15.

[35] B. Clarke, "Comparing Bayes Model Averaging and Stacking When Model Approximation Error Cannot be Ignored," p. 30.

[36] M. Lin, D. Wu, J. Meng, J. Wu, and H. Wu, "A multi-feature-based multi-model fusion method for state of health estimation of lithium-ion batteries," *J. Power Sources*, vol. 518, p. 230774, Jan. 2022.

[37] S. Shen, M. Sadoughi, M. Li, Z. Wang, and C. Hu, "Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries," *Appl. Energy*, vol. 260, p. 114296, Feb. 2020.

[38] D. Roman, S. Saxena, V. Robu, M. Pecht, and D. Flynn, "Machine learning pipeline for battery state-of-health estimation," *Nat. Mach. Intell.*, Apr. 2021.

[39] Z. Wang, C. Yuan, and X. Li, "Lithium battery state-of-health estimation via differential thermal voltammetry with gaussian process regression," *IEEE Trans. Transp. Electrific.*, vol. 7, no. 1, pp. 16–25, 2021.

[40] Y. Preger, H. M. Barkholtz, A. Fresquez, D. L. Campbell, B. W. Juba, J. Romàn-Kustas, S. R. Ferreira, and B. Chalamala, "Degradation of Commercial Lithium-Ion Cells as a Function of Chemistry and Cycling Conditions," *J. Electrochem. Soc.*, vol. 167, no. 12, p. 120532, Sep. 2020.

[41] C. Zou, X. Hu, Z. Wei, T. Wik, and B. Egardt, "Electrochemical estimation and control for lithium-ion battery health-aware fast charging," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6635–6645, 2018.

[42] Y. Li, D. M. Vilathgamuwa, E. Wikner, Z. Wei, X. Zhang, T. Thiringer, T. Wik, and C. Zou, "Electrochemical model-based fast charging: Physical constraint-triggered PI control," *IEEE Trans. Energy Conver.*, vol. 36, no. 4, pp. 3208–3220, 2021.

[43] *Spearman Rank Correlation Coefficient*. New York, NY: Springer New York, 2008, pp. 502–505.

[44] Y. Zhang, T. Wik, J. Bergström, M. Pecht, and C. Zou, "A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data," *J. Power Sources*, vol. 526, p. 231110, Apr. 2022.

[45] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, ser. Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2006.

[46] D. J. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992.

[47] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[48] C. M. Bishop, *Pattern recognition and machine learning*, ser. Information science and statistics. New York: Springer, 2006.

[49] A. Lindholm, N. Wahlström, F. Lindsten, and T. B. Schön, *Machine Learning - A First Course for Engineers and Scientists*. Cambridge University Press, 2022.

[50] S. Wager, T. Hastie, and B. Efron, "Confidence intervals for random forests: The jackknife and the infinitesimal jackknife," *J. Machine Learning Research*, vol. 15, no. 1, pp. 1625–1651, 2014.

[51] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[52] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[53] E. Schaltz, D.-I. Stroe, K. Norregaard, L. S. Ingvardsen, and A. Christensen, "Incremental capacity analysis applied on electric vehicles for battery state-of-health estimation," *IEEE Trans. Ind. Appl.*, vol. 57, no. 2, pp. 1810–1817, Mar. 2021.

[54] B. Gou, Y. Xu, and X. Feng, "An Ensemble Learning-Based Data-Driven Method for Online State-of-Health Estimation of Lithium-Ion Batteries," *IEEE Trans. Transp. Electrific.*, vol. 7, no. 2, pp. 422–436, Jun. 2021.