Multiobjective Optimization of Data-Driven Model for Lithium-Ion Battery SOH Estimation with Short-term Feature

Cai, Lei; Meng, Jinhao; Stroe, Daniel-Ioan; Peng, Jichang; Guangzhao, Luo; Teodorescu, Remus

Published in:
I E E E Transactions on Power Electronics

DOI (link to publication from Publisher):
10.1109/TPEL.2020.2987383

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Multi-objective Optimization of Data-driven Model for Lithium-ion Battery SOH estimation with Short-term Feature

Lei Cai, Jinhao Meng, Daniel-Ioan Stroe, *Member, IEEE*, Jichang Peng, Guangzhao Luo, *Member, IEEE*, and Remus Teodorescu, *Fellow, IEEE*

***Abstract*—As a favorable energy storage component, Lithium-ion (Li-ion) battery has been widely used in the Battery Energy Storage Systems (BESS) and Electric Vehicles (EV). Data driven methods estimate the battery State of Health (SOH) with the features extracted from the measurement. However, excessive features may reduce the estimation accuracy and also increases the human labor in the lab. By proposing an optimization process with Non-dominated Sorting Genetic Algorithm II (NSGA-II), this paper is able to establish a more efficient SOH estimator with Support Vector Regression (SVR) and the short-term features from the current pulse test. NSGA-II optimizes the entire process of establishing a SOH estimator considering both the measurement cost of the feature and the estimation accuracy. A series of non-dominated solutions are obtained by solving the multi-objective optimization problem, which also provides more flexibility to establish the SOH estimator at various conditions. The degradation features in this paper are the knee points at the transfer instants of the voltage in the short-term current pulse test, which is fairly convenient and easy to be obtained in real applications. The proposed method is validated on the measurement from two $LiFePO_4$/C batteries aged with the mission profile providing the Primary Frequency Regulation (PFR) service to the grid.[1]**

***Index Terms*—state of health estimation; multi-objective optimization; feature selection; lithium-ion battery.**

I. Introduction

With an increasing amount of renewable energy sources integrated into the grid, improving the flexibility of the power system becomes an urgent requirement [1]. However, the intrinsic nature of the renewable energy (i.e., intermittent, variable, and not 100% predictable) has a negative effect on the stability of the power system. Battery Energy Storage Systems (BESS) are proposed as a suitable solution for enhancing the grid stability because they are able to perform multiple tasks, i.e., frequency regulation, peaking shaving, voltage support, .etc [2]–[4]. BESS has the advantages of fast response and good scalability. The superior performance and the continuously decreasing price have made Lithium-ion (Li-ion) battery technology the prime candidate for BESS [4]–[7]. As a core component, more attentions have been paid in order to enhance the performance of the Li-ion batteries in the BESS. Overusing the battery cell will accelerate the degradation and cause reduction of the battery pack lifetime. Battery lifespan management, which is closely related to the cost and reliability of the BESS, is one of the key functions in the Battery Management System (BMS). In this condition, the State-of-Health (SOH) of each cell in the battery pack needs to be known [8]. However, the ageing process of battery is triggered by different stress factors and their interactions, and it's hardly to clarify those reactions in theory. The main reason for the battery degradation is the side effects of the electrochemistry reactions. For example, the metallic lithium plating causes the passivation layer growth in the anode, the insertion and extraction of Li-ion changes the volume of the active materials in the cathode, etc. Additionally, various battery macroscopic parameters (e.g., current rate, State-of-Charge (SOC) and temperature) can also accelerate the degradation [9]. High temperature and SOC speed up the loss of lithium, and the high current rate and Depth of Discharge (DoD) result in the loss of active material.

SOH indicates the current health status of a battery during its lifespan, which can be expressed in two forms: the capacity fade and the power fade [10], [11]. The capacity fade means the amount of energy stored in a battery is decreased with the degradation, while the power fade indicates that the internal resistance increases with ageing [12]. The rates of the capacity fade and the power fade are varied with battery degradation, which means that the SOH has to be estimated after a certain period. Measuring the capacity and the internal resistance are definitely the most straightforward ways to predict the battery SOH. The capacity measurement needs a long testing time to fully charge or discharge the battery. In addition, since the internal resistance is usually a very small value, the internal resistance measurement of a battery may not be accurate enough considering the external resistance of the test bench. Online resistance estimation is easily affected by the measurement noises from the sensors [13], while the capacity estimation generally relies on an accurate SOC as a prerequisite. For example, in Kalman filter based method [11], SOC is often

[1] This work was supported by Natural Science Foundation of Shaanxi Provincial Department of Education under grant 19JK0575, the Fundamental Research Funds for the Central Universities under grant YJ202013 and National Natural Science Foundation of China under grant 61973042. (*Corresponding author: Jinhao Meng.*)

L. Cai is both with the Faculty of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China, and Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an 710048, China (e-mail: caileid@gmail.com).

J. Meng is with the College of Electrical Engineering, Sichuan University, Chengdu 610065, China (e-mail: scmjh2008@163.com).

D. Stroe and R. Teodorescu are with the Department of Energy Technology, Aalborg University, Aalborg 9220, Denmark (e-mails: dis@et.aau.dk; ret@et.aau.dk).

J. Peng is with the Smart Grid Research Institute, Nanjing Institute of Technology, Nanjing 211167, China (email: linkpjc@gmail.com).

G. Luo is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China (email: guangzhao.luo@nwpu.edu.cn).

used as the input of the state space equation for capacity estimation. However, the accuracy of the SOC estimation is still questionable [14], [15]. An improve Ampere-Count method is proposed in [16] to estimate the SOH estimation of Li-ion battery, which uses Incremental Capacity Analysis (ICA) from a specific voltage range. However, it should be noted that incremental capacity (IC) curve and Differential Voltage (DV) curve are expected to charge or discharge the battery with extremely low current (1/25C), which may not always realistic in real applications.

Empirical models are proposed to predict the SOH by describing the correlation between the ageing stress factors and the SOH with the mathematical functions, such as, the exponential, polynomial, or power functions [17]–[20]. The parameters in the empirical model are generally extracted from the long-term lifetime test of a specific battery, which may not always ensure an accurate SOH estimation of other batteries. Electrochemical models, such as the ones presented in [21] and [22], explain the degradation mechanism of the battery based on complex electrochemical parameters, which are not easy to be implemented in the real applications. Hence, semi-empirical models are proposed to combine the previous electrochemical knowledge of the battery with the mathematical model obtained from the lifetime tests, in order to reach a tradeoff on the estimation complexity [18], [19].

Data driven models are another category of battery SOH estimation method, which essentially utilizes machine learning algorithms to establish a mapping from features to SOH. Compared with the above mentioned methods, various features extracted from the measurement can be integrated to estimate SOH by data driven models. From example, in [23], the velocity and arc-length curvature extracted from the terminal voltage during the battery charge process are utilized to establish a neural network based estimator. Features form Electrochemical Impedance Spectroscopy (EIS) [24] , are also possible to be used for battery SOH estimation [25]. The features for SOH estimation should be easily obtained in real applications. However, the features from [23] needs a complex calculation process, and EIS measurement often show high sensitive to the noise from sensors. Currently, the robust of data-driven methods rely mainly on the supervised learning method they chosen, since this is the most direct way to enhance the general performance of a learning process. For example, extreme learning machine [26] and neural network [27] have been applied to establish the connections between the health indicator and the battery SOH. In addition, Support Vector Regression (SVR) [28], [29], Gaussian process regression [30]–[32], Recurrent Neural Networks (RNN) [33], etc, are also chosen to create the regression models for SOH estimation. The above-mentioned algorithms have archived great success in handling the training process. Unfortunately, the optimization of the feature and the training process have not been fully discussed in these methods.

The process of establishing a data driven model includes three steps: original measurement collection, feature extraction, and training the model. Despite the fact that data driven methods have already attracted certain interests in the literature, existing methods have to face following challenges.

The first major challenge is that the features used for SOH estimation have a significant impact on the performance of a data driven model. In an estimation task, we always extract more features hoping that these features may not missing any possible information. Unfortunately, a large number of possible features may cause a “curse of dimensionality” and the redundant features will even decrease the performance of an estimator.

The second challenge occurs in the phase of training the model. In practice, the estimating results suffer from lacking of the knowledge of the selection of the hyperparameters. For example, to ensure the performance of SVR for SOH estimation [28], [29], [34], tuning hyperparameters is a key. In order to obtain better results, a natural way is to iteratively adjust the parameters for those estimators, e.g., for SVR, with different hyperparameters C, ε and γ, each iteration for one result. However, this is obviously deficient, especially for a large-scale problem.

The third challenge lies in the fact that the two major aims of feature extraction are sometimes conflicted, i.e, to maximize the accuracy of the estimator and to minimize the number of selected features for the purpose of alleviating the curse of dimensionality. How to balance these two goals is not an easy task.

To our best knowledge, there is no method designed to simultaneously address all the aforementioned challenges. In this work, we aim to devise a reliable model to solve the issues of feature extraction and training model simultaneously. This approach picks up the features from the raw measurement of the current pulse test for the purpose of establishing an accurate and efficient SOH estimator. The short-term current pulse tests performed at different SOCs are used to capture the degradation information of the battery. The features are extracted from the knee points in the voltage response of the current pulse test. In addition, the proposed model is expected to get a trade-off between different goals of the feature extraction. The final optimization problem is solved by Non-dominated Sorting Genetic Algorithm II (NSGA-II) considering the minimization of the number of features and improving the estimation accuracy simultaneously. The main contributions of this paper are summarized as follows:

(1) The knee points in the voltage response of the current pulse test are easily extracted in real applications. The short-term features, which lasts only few seconds, is very convenient to be obtained.

(2) The feature extraction and the hyperparameter in SVR are optimized simultaneously. In addition, the decision space and the corresponding encoding method are made.

(3) A series of non-dominated solutions are obtained from NSGA-II, which provides more freedom to train an efficient SOH estimator. NSGA-II optimizes the process of establishing the estimator according to two aspects: less indispensable measurement and higher estimation accuracy.

(4) 5-fold cross-validation is used to verify the actual accuracy of the SOH estimators. The dataset is collected from two $LiFePO_4$/C batteries aged with mission profile offering the Primary Frequency Regulation (PFR) service to the grid [35].

The rest of the paper is organized as follows. Section II introduces the features for SOH estimation. The SVR based

SOH estimator and the optimized feature selection with NSGA-II are detailed in Section III. Section IV shows the experimental validation of the proposed approach. Conclusions are given in Section V.

II. FEATURES FOR SOH ESTIMATION

Since the current pulse test can detect the power fade of the Li-ion battery, it is used to generate the features for SOH estimation. The schematic of a discharging current pulse and the corresponding voltage response is shown in Fig. 1. It is seen from Fig. 1 that the voltage gradually recovers to a stable value when a sudden variation of the current is applied to a battery. The voltage responses under the same current pulse test are varied with the battery ageing stages. Hence, the four knee points (i.e., *A*, *B*, *C*, *D* in Fig. 1) in the voltage response curve are selected as the features for SOH estimation.

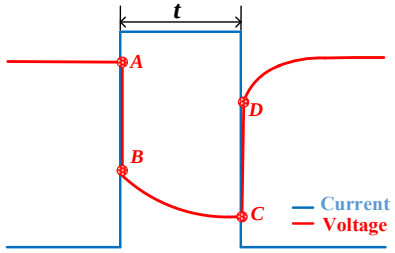


Fig.1. The current and voltage in one current pulse

Let's take the voltage measurement from a $LiFePO_4$/C battery at SOC=20% and I=10A as an example. The voltage responses at the discharge current pulse are shown in Fig. 2 from Week 1 to Week 33. It should be noted that the voltage responses of the battery are not exactly fixed to 18 seconds because of the voltage hysteresis and the physical limitations of the test bench. The voltage curve is varied with battery degradation from week to week.

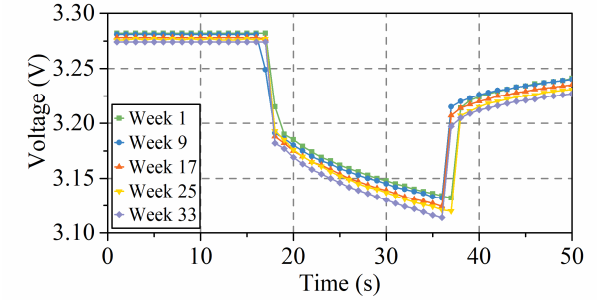


Fig. 2. The voltage response of a $LiFePO_4$/C battery

Therefore, the knee points (*A*, *B*, *C*, *D*) extracted from the voltage measurement in Fig. 2 are shown as follows.

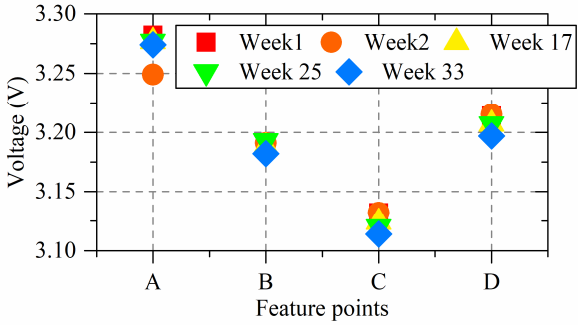


Fig. 3. The extracted features from the voltage measurement

In Fig. 3, the features at point *B* are very close to each other, which indicates that the knee point at *B* may not make a significant contribution to the SOH estimation. The extracted features have to be selected to exclude those invalid interferences. The current pulse test is convenient to be performed on various SOCs and current rates. It is noted that the current pulse tests are not exactly the same on the high and low SOC area. In addition, the current rates also make a difference on the voltage measurement in the current pulse test. As a result, the current pulse tests are performed at various SOCs (SOC=20%, 50%, 80%) and current rates (4C, 2C, 1C) to obtain sufficient information for the battery degradation. More details about the current pulse test will be introduced in Section IV. It is easily realized that the proposed features are conveniently obtained because the current transfer time has already known from the controller of the power converter in real application. Moreover, the duration of the current pulse only lasts 18 seconds. In order to avoid the effect of the measurement noise, the features can be obtained by calculating the average voltage of several sampling points near the transfer time. Additionally, since the current pulse test can be easily performed in real application, the voltage measurement noise could also be removed by repeating the current pulse several times and calculating the average value.

III. SOH ESTIMATOR WITH OPTIMIZED FEATURE

Before describing the proposed approach in detail, we provide a diagram in Fig. 4 to illustrate the main procedures. First, the feature extraction block utilizes the data of current pulse test which has been described in Section II. It can provide sufficient information for the battery degradation. To dealt with redundant and corrupted features, the feature selection step is employed for selecting a proper combination of the features. Then, the selected set of features are sent to an estimator. The proposed method searches proper choice of features and hyperparameters of estimator simultaneously by using NSGA-II.

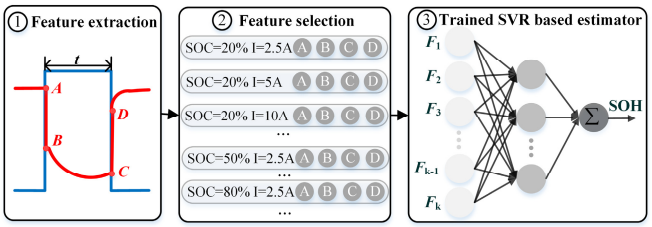


Fig. 4. The main steps of the proposed approach

A. NSGA-II

This section introduces the multi-objective optimization algorithm, i.e., NSGA-II. Since the current pulses are measured at various SOC and current rates, it's better to use the features at the same SOC to reduce the workload. Meanwhile, the SOH estimator should also be as accurate as possible. Generally, selecting more measurement does not mean that the estimation accuracy will definitely be increased or decreased. As a result, a two-objective optimization problem is formed to extract the features.

Multi-objective optimization usually has more than one solution, which altogether forms the Pareto front. All the non-dominated solutions in the Pareto front cannot be replaced

without degrading the values of the objective functions. Therefore, the purpose of the multi-objective optimization is to obtain all the non-dominated solutions.

The evolutionary algorithms are able to deal with a group of solutions simultaneously. Thus, it is popular to be used to forecast the Pareto front with high efficiency. NSGA-II is a fast and elitist algorithm for searching all the non-dominated solutions of a multi-objective optimization [36]. The flowchart of NSGA-II is illustrated as follows.

As shown in Fig. 5, the non-dominated sorting approach can sort the population of each new generation at the beginning. The crowding distance, which indicates the density of the solution inside the cuboid formed by the nearest neighbors, is assigned to each solution in NSGA-II. Then, the crowded comparison operator is performed to select the individuals acted as the parents in the current generation, while the crossover and mutation are applied to generate the offspring. After combining the parents and offspring, the elitism selection can choose the solutions for the next generation. If the terminated condition is not reach, the NSGA-II will continue with the next term.

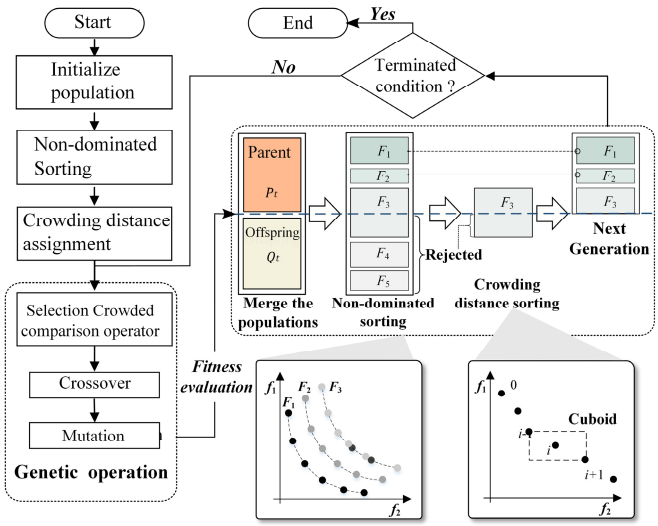


Fig. 5. The flowchart of NSGA-II

Non-dominated sorting, crowding distance and elitism endow the NSGA-II with the efficiency to solve the multi-objective optimization. The fast non-dominated sorting approach is capable of sorting the population members according to their non-domination level. At the same non-domination level, the crowding distance is calculated to ensure the diversity of the population. Afterwards, the elitism can help to find the best solutions for the next generation.

The crowding distance means the density of the solution in one population, which actually can be measured by the absolute normalized difference of the two adjacent solutions. The crowding distance is calculated by in NSGA-II [36],

$$D[i]_{distance} = \frac{\left(D[i+1]_{distance}.m - D[i-1]_{distance}.m\right)}{\left(f_{m,\max} - f_{m,\min}\right)} \tag{1}$$

where the $D[i]_{distance}$ is the crowding distance of the i-th solution, $D[i]_{distance}.m$ is the m-th value of the objective function, $f_{m,max}$ and $f_{m,min}$ are the maximum and minimum of the m-th objective function.

B. SVR based estimator

This paper utilizes the SVR to establish the regression between the features and the battery SOH. Taking the advantage of the Vapnik-Chervonenkis (VC) and statistical theory, SVR is able to form the regression between the dependent variables and the predictors with a good generalization to the unknown dataset. The kernel function maps the input to a higher dimensional space, which can easily solve the nonlinear regression by using a hyperplane in a new space. On the foundation of the kernel function, the nonlinear regression can be generated by SVR without increasing much computational burden.

The expression of SVR is,

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \cdot \varphi(\boldsymbol{x}) + b \tag{2}$$

where $\boldsymbol{w}$ is the combination weights of the support vectors, $\varphi(\boldsymbol{x})$ maps the input to a new feature space. After introducing the slack variables ξ_i and ξ_i^*, the following optimization problem is formed [37].

$$minimize\ \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \tag{3}$$

subject to the constraints,

$$\begin{cases} y_i - \boldsymbol{w}^T \cdot \varphi(\boldsymbol{x}_i) - b \le \varepsilon + \xi_i \\ \boldsymbol{w}^T \cdot \varphi(\boldsymbol{x}_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

where C is a positive constant determining the trade-off between the flatness of $f(\boldsymbol{x})$ and the selection of the "soft margin". SVR can find the largest ε deviation from the training dataset as well as ensure the flatness of the Eq. (3).

In most conditions, Eq. (3) can be solved in a dual formulation. Hence, $\boldsymbol{w}$ is calculated by using the Lagrange multipliers as follows,

$$\boldsymbol{w} = \sum_{i=1}^{N}\left(\beta_i^* - \beta_i\right)\cdot \boldsymbol{x}_i \tag{4}$$

where β_i^* and β_i are Lagrangian multipliers.

Afterwards, Eq. (2) is rewritten as,

$$f(\boldsymbol{x}) = \sum_{i=1}^{N}\left(\beta_i^* - \beta_i\right)\cdot K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{5}$$

where $K(\boldsymbol{x}_i, \boldsymbol{x}) = \left\langle \varphi(\boldsymbol{x}_i)\cdot\varphi(\boldsymbol{x}_j)\right\rangle$ is the kernel function that meets the Mercer's condition [38].

The kernel functions are able to convert the nonlinear space into a linear one. Among the kernel functions, Radial Basis Function (RBF) is used to construct the SVR in this paper. The expression of the RBF is,

$$K(\boldsymbol{x}_i, \boldsymbol{x}) = e^{-\frac{1}{2}\|\boldsymbol{x}_i - \boldsymbol{x}\|^2 / \gamma} \tag{6}$$

where γ is the parameter of the kernel function. One advantage of the RBF kernel function is that it can straightforwardly approximate other kernel functions by tuning the parameter γ.

From the descriptions in this section, it is known that the parameters C, ε, and γ are critical for the training process of SVR.

C. Optimized feature extraction with NSGA-II

This subsection introduces how the NSGA-II is used to select the features from the current pulse test. In this work, we model feature extraction and hyperparameters selection as one optimization problem and try to solve it by simultaneously considering the estimation accuracy as well as the length of features. Before that, the individual representation and the fitness evaluation should be designed. A chromosome-like structure is generally used to represent each individual in the evolutionary algorithm. Hence, the selection of the candidate features is encoded into a chromosome-like form. The binary is applied to represent the selection of each candidate feature. For example, 1 means the candidate feature is selected, while 0 indicates it is excluded. In addition, the parameters C, ε, and γ should also be optimized to guarantee an excellent performance of the SVR. Therefore, those parameters are encoded in a 40 bits binary format. As a result, the individual representation is as follows,

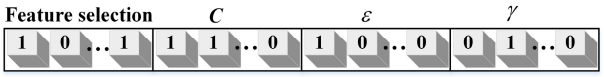


Fig. 6. The individual representation in the NSGA-II

After encoding the feature extraction and the hyperparameters in the SVR, the genetic operations (crossover and mutation) are applied to generate the offspring. Since the hybrid encoding is used to represent each individual by a chromosome-like structure as shown in Fig. 6, the crossover is operated in two parts separately according to the feature extraction and the parameter optimization. In this way, the genetic operation can create the new individuals for the two parts at the same time. For the same reason, the mutation is also operated in two parts independently.

The fitness evaluation guides the search of the optimal solutions in the NSGA-II. For the purpose of reducing the labor and excluding the invalid features, less measured features are preferred. The accuracy of the SOH estimation anyway needs to be guaranteed. Thus, the two objective functions are defined as,

$$f_1 = N_{feature} \tag{7}$$

$$f_2 = \frac{1}{n}\sum_{x_i, y_i \in \mathcal{D}} \left(\mathcal{F}_{SVR}\left(\mathcal{D} \setminus \mathcal{D}_i, x_i\right) - y_i\right)^2, i \in [1, n] \tag{8}$$

where $N_{feature}$ is the number of the selected features, $\mathcal{D}$ represents the training dataset including all the candidate features, $\mathcal{D}_i$ is the testing dataset to evaluate the accuracy of the SOH estimator.

Since less features in different SOCs can save the measurement labor, a smaller f_1 means a better result. The SOH estimator needs to be accurate enough, and a smaller f_2 also means a better estimation accuracy. It should be noted that the accuracy of the SOH estimation in this paper is based on the 5-fold cross-validation. It means that the training dataset is divided into 5 groups randomly each time, and the validation repeats 5 times to traverse all the subsets. In each term, 4 groups are used to train the SVR estimator and the remaining one is the dataset for validation. It can be seen that the cross-validation has the ability to verify the performance of the estimator on an independent dataset. The MSE with 5-fold cross-validation is closer to the performance of the estimator in real application.

IV. Experimental Validation

The cycling ageing test on two $LiFePO_4$/C batteries is used to verify the proposed method. The detailed parameters of the $LiFePO_4$/C battery are list in TABLE I.

TABLE I Parameters of The Tested Battery

Parameter	Value
Shape	cylindrical
Weight [g]	76
Dimensions [mm]	ø 26 x 65
Nominal capacity [Ah]	2.5
Nominal voltage [V]	3.3
Maximum voltage [V]	3.6
Minimum voltage [V]	2.0
Maximum continuous charge current [A]	10
Maximum continuous discharge current [A]	50
Operating temperature	-30ºC ~ +55ºC

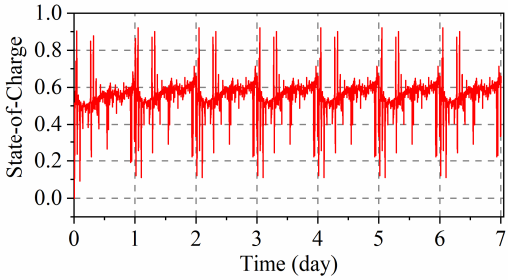


Fig. 7. The mission profile for cycling the battery

A mission profile from BESS, which provides PFR service to the grid, is used to cycle the two $LiFePO_4$/C batteries for one week [35]. The SOC variation of the mission profile is shown in Fig. 7, and the average SOC is 50 %. The battery SOC ranges from 10% to 90% during the cycling profile, the degradation process is thus accelerated. During the test, the batteries are cycled in the chamber of the FuelCon test station. The ambient temperature in the test chamber is set to 25 ºC, in order to alleviate the effect of temperature.

The entire procedure of the cycling ageing test in this paper contains three main steps. In the first step, the Li-ion battery is cycling with the profile in Fig. 7. Then, the battery is fully charged and discharge for twice to obtain the reference capacity as shown in Fig. 8. Thirdly, the current pulse test in Fig. 9 is performed to collect the features for SOH estimation. A testing period of 56 weeks in total is spent on the two batteries. The capacity degradation of the batteries during the test is shown in Fig. 8.

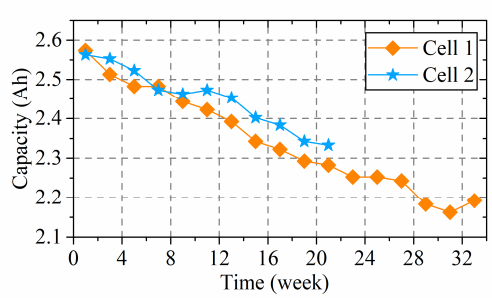


Fig. 8. The capacity degradation of the battery during the cycling test

Since 4 points are received from 1 current pulse, 18 current pulses generate the feature with 72 dimensions. NSGA-II is then used to select the features from various measurement conditions, in order to obtain an accurate SOH estimation with less measurement labor.

The current pulses are performed at several SOCs and current amplitudes to capture the features. The schematic of the current pulse test is shown in Fig. 9. The current pulses are performed at three different SOCs (20%, 50%, 80%) and current rates (1C, 2C, 4C). Each current pulse lasts for 18 seconds, and then rests 15 minutes before performing the next one. The knee points in the voltage response curve are extracted as the features, which are $f_1, f_2, \ldots, f_{18}$.

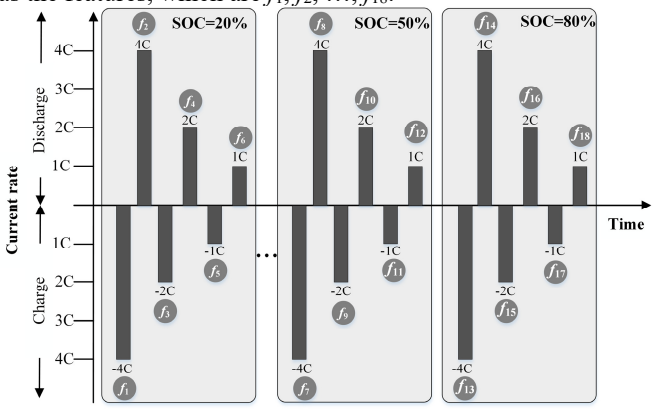


Fig. 9. The current pulse test

In order to clarify how the features affect the performance of the data-driven model, the estimation process with different dimension of the feature matrices is repeated for 4 times as shown in Fig. 10. The dimensions of the feature are gradually increased, and the subsets of features are randomly extracted from dataset. We can find that the estimation shows unstable performance if very limited number of dimensions is selected. The instability will fade when the scale of dimensions is increased, yet the performance may not be always improved with the consistently increased features.

The features are selected according to the measurement at different SOCs. In order to clarify the performance of the SOH estimation with feature selection, seven different conditions are optimized by NSGA-II. The feature selection of Cell 1 with NSGA-II is shown in Fig. 11. In the legend, SOC=20% & 50 % & 80% means that at least one feature should be selected at 20%, 50% and 80% SOC respectively; SOC=20% represents the feature can only be selected from 20% SOC, and so on.

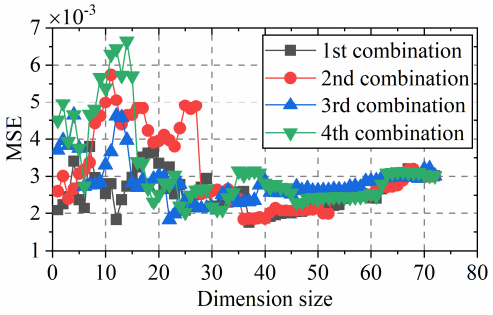


Fig. 10. The trajectories of the estimation process with different dimension features

Hence, all the points in Fig. 11 are the optimal solutions for the feature selection of Cell 1. Since none of the non-dominated solutions in Fig. 10 can be replaced by the others without degrading the values of f_1 and f_2, all the solutions can be regarded as the best one from a specific point of view. However, the best solution can still be chosen from those solutions according to the requirement of a specific application. As previously described, f_1 is the number of the selected features, f_2 is the MSE of the SOH estimation. If an application can easily perform the corresponding current pulse test at various SOCs, the solutions with smaller f_2 should be chosen for SOH estimation. Meanwhile, if the features can only be measured at one certain SOC, the solutions, which have a smaller f_1, should be considered with a higher priority.

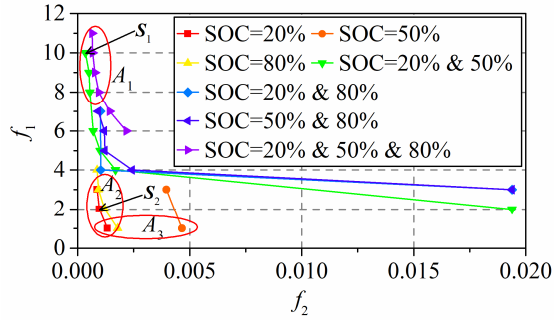
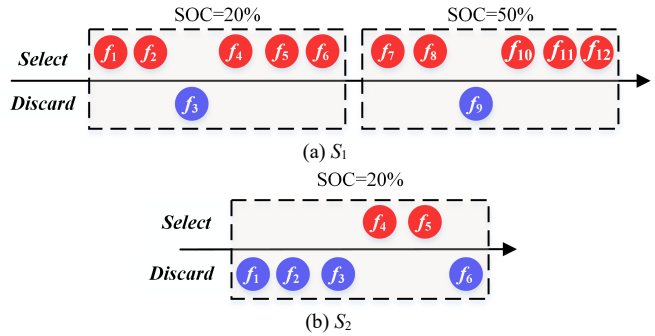


Fig. 11. The optimized feature selection of Cell 1 by NSGA-II

Three regions, A_1, A_2 and A_3, are picked up from Fig. 11 for the discussion of how to choose a suitable solution. In general, the solutions close to the coordinate origin are a better trade-off of the two objectives. Therefore, the solutions in A_2 are a better trade-off than the other solutions. Nonetheless, the solutions in A_1 and A_3 can also be regard as the best solution from a specific point of view. For example, the MSE of the solutions in A_1 are usually less than the others, and the required number of the current pulses is fewer in A_3. All the solutions close to the ordinate are possibly used to guide the feature selection in real application. The SOH estimation results of two typical solutions S_1 and S_2 from A_1 and A_2 are illustrated in Fig. 12. In S_1, two current pulses at SOC = 20% and 50% are excluded. The rest is used as the features to train the SVR. The selected features in S_2 include only two current pulses as shown in Fig. 12(b). The performance of the proposed method is compared with three other methods to prove its advantages. Method 1 uses all the features and the Extreme Learning Machine (ELM) [26]. Method 2 optimizes both the hyperparameters of SVR and the feature extraction simultaneously as shown in [34]. Method

3 uses all the features and only optimizes the hyperparameters of SVR by grid search algorithm. In order to validate the performance of the proposed method in reality, all the estimation results are obtained through 5-fold cross-validation.

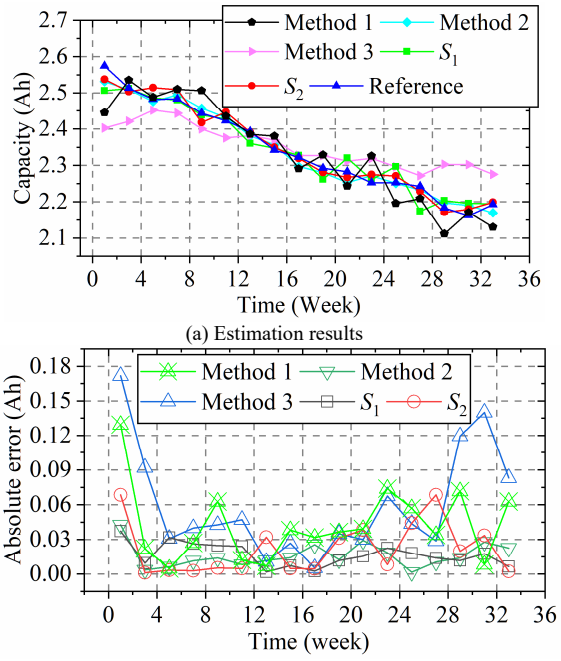
The estimation results and the absolute error in Fig. 13 prove the effectiveness of the proposed SOH estimators regarding to the two solutions. The MSE of S_1 solution is 3.7269×10^{-4}, while it is 9.7017×10^{-4} for S_2. The MSE of Method 1 is 0.0027 and the MSE is 0.0056 for Method 3, which indicates the advantage of simultaneously optimizing the hyperparameters and the feature extraction. Although the MSE of Method 2 is 3.6562×10^{-4}, Method 2 can only provide one solution for SOH estimation. If the real application cannot meet the specific requirement of Method 2, the SOH is impossible to be estimated. It can be found that the proposed method provides sufficient non-dominated solutions for various conditions, which improves the flexibility of the SOH estimation using features from the current pulse test. It is noted that the proposed method also provides the solutions with a comparable accuracy compared with Method 2.



(a) S_1

(b) S_2

Fig. 12. The solutions of feature selection: S_1 and S_2



(a) Estimation results

(b) Absolute error

Fig. 13. Estimation results for Cell 1

The optimized solutions from NSGA-II of Cell 2 are shown in Fig. 14. The solutions in A_4 has a smaller f_2, which means the feature selections in A_4 are able to train a more accurate SVR based estimator. Meanwhile, the solutions in A_5 rely on fewer current pulses, and the features in that area are convenient to be extracted in real applications. In order to clarify the performance of the SOH estimation, three solutions (S_3, S_4, and S_5) are chosen from Fig. 14.

The SOH estimation results of Cell 2 are shown in Fig. 15. As shown in Fig. 15(a), all the estimation results follow the reference during the entire degradation process of Cell 2. The maximum absolute error of the three chosen solutions is less than 0.07 Ah. Especially, in Fig. 15(b), the estimation results of S_5 receives the absolute errors less than 0.04 Ah. The MSE of S_3 is 9.5286×10^{-4}, and it is 0.0011 for S_4. As for S_5, the MSE is only 2.8337×10^{-4}. However, the MSE of Method 1 is 0.0038 and the MSE is 0.0051 for Method 3, which proves the necessity of the optimization process. The MSE of Method 2 is 1.7279×10^{-4}. Compared with Method 2, the proposed method provides more freedom to train an efficient SOH estimator considering both the estimation accuracy and the current pulses measurement complexity.

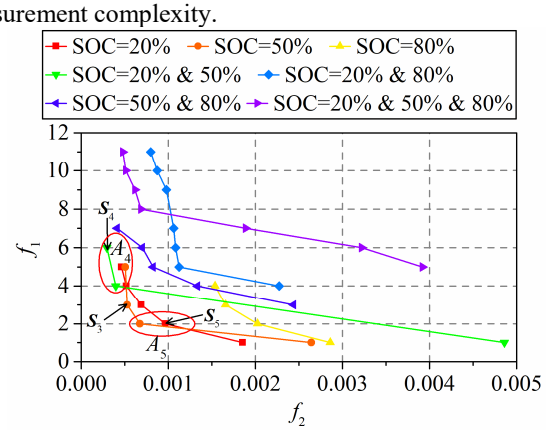
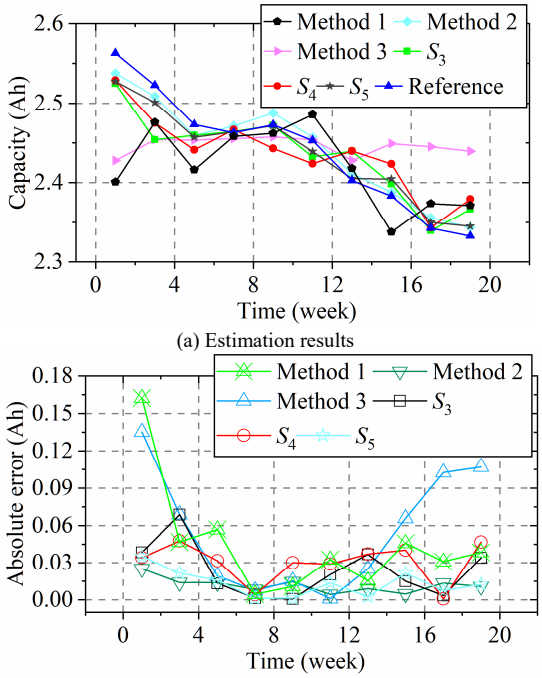


Fig. 14 The optimized feature selection of Cell 2 by NSGA-II



(a) Estimation results

(b) Absolute error

Fig. 15 Estimation results of SVR in Cell 2

From the above experimental results, we can find that measuring the current pulses at SOC=20% and SOC=20% & 50% is a better choice for the SOH estimation of both Cell 1 and Cell 2. According to the solutions in Figs. 11 and 14, more features does not guarantee a more accurate SOH estimator. The value of the feature selection process is then proved.

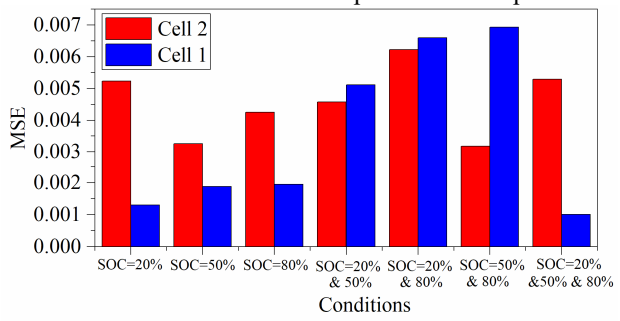


Fig. 16. Estimation results of Cell 2 using optimized solutions from Cell 1

In order to further verify the generalization of the proposed method, the optimized solution in Cell 1 is directly used to establish the SOH estimator for Cell 2. The results are shown in Fig. 16. Since the average MSE of Cell 2 in all conditions is 0.0046 and it is 0.0035 using the results from Cell 1, the good generalization of the proposed method is thus proved.

V. Conclusion

This paper establishes an efficient SOH estimator with features selected from the short-term current pulse test. The current pulse tests, lasting only few seconds, are convenient to be performed at various working conditions of the Li-ion battery. In order to train an accurate SOH estimator with less measurement labor, the features from the current pulse test are selected by NSGA-II. The optimized results in Figs. 11 and 14 have already proved the necessity of the feature selection process. According to the parameters optimized by NSGA-II, SVR is used to establish the SOH estimator. The estimator is validated on two $LiFePO_4$/C batteries with mission profile providing the PFR service to the grid. 5-fold cross validation is used to verify the accuracy of the SOH estimators. Five typical solutions are picked up from the optimized feature selection solutions to show the performance of the proposed SOH estimator. The maximum MSE of the $S_{1\text{-}5}$ is less than 0.0011, which proves the validation of the SOH estimator. It should be noted that all the solutions from NSGA-II can be regarded as the optimal solution from a certain point of view. The best solution is chosen according to the requirement of a specific application. Thus, more freedom is given to design an efficient SOH estimator by using the proposed method. A comparison with three other methods also demonstrates the advantages of the proposed method.

Temperature will affect the performance of the battery cell in reality. Thus, future works are focusing on investigating the performance of the proposed method when the temperature of each cell is varied in the battery pack. In addition, some online learning methods should be investigated in the future such that the proposed method could adaptively learn online whenever new data is available.

References

[1] P. Fairley, "Energy storage: Power revolution," *Nature*, vol. 526, p. S102, Oct. 2015.

[2] B. M. Gundogdu, S. Nejad, D. T. Gladwin, M. P. Foster, and D. A. Stone, "A Battery Energy Management Strategy for U.K. Enhanced Frequency Response and Triad Avoidance," *IEEE Trans. Ind. Electron.*, vol. 65, no. 12, pp. 9509–9517, 2018.

[3] M. Dubarry, A. Devie, K. Stein, M. Tun, M. Matsuura, and R. Rocheleau, "Battery Energy Storage System battery durability and reliability under electric utility grid operations: Analysis of 3 years of real usage," *J. Power Sources*, vol. 338, pp. 65–73, 2017.

[4] A. Saez-de-Ibarra, E. Martinez-Laserna, D. Stroe, M. Swierczynski, and P. Rodriguez, "Sizing Study of Second Life Li-ion Batteries for Enhancing Renewable Energy Grid Integration," *IEEE Trans. Ind. Appl.*, vol. 52, no. 6, pp. 4999–5008, 2016.

[5] B. Diouf and R. Pode, "Potential of Lithium-ion Batteries in Renewable Energy," *Renew. Energy*, vol. 76, pp. 375–380, 2015.

[6] D. Larcher and J.-M. Tarascon, "Towards greener and more sustainable batteries for electrical energy storage," *Nat. Chem.*, vol. 7, p. 19, Nov. 2014.

[7] H. Liu, L. Chen, Q. Zhang, and K. Deb, "Adaptively Allocating Search Effort in Challenging Many-Objective Optimization Problems," *IEEE Trans. Evol. Comput.*, vol. 22, no. 3, pp. 433–448, 2018.

[8] A. El Mejdoubi, H. Chaoui, H. Gualous, P. Van Den Bossche, N. Omar, and J. Van Mierlo, "Lithium-Ion Batteries Health Prognosis Considering Aging Conditions," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 6834–6844, 2019.

[9] J. Vetter *et al.*, "Ageing mechanisms in lithium-ion batteries," *J. Power Sources*, vol. 147, no. 1–2, pp. 269–281, Sep. 2005.

[10] M. Cacciato, G. Nobile, G. Scarcella, and G. Scelba, "Real-Time Model-Based Estimation of SOC and SOH for Energy Storage Systems," *IEEE Trans. Power Electron.*, vol. 32, no. 1, pp. 794–803, 2017.

[11] C. Chen, R. Xiong, and W. Shen, "A Lithium-Ion Battery-in-the-Loop Approach to Test and Validate Multiscale Dual H Infinity Filters for State-of-Charge and Capacity Estimation," *IEEE Trans. Power Electron.*, vol. 33, no. 1, pp. 332–342, 2018.

[12] D. I. Stroe, M. Swierczynski, S. K. Kær, and R. Teodorescu, "Degradation Behavior of Lithium-Ion Batteries During Calendar Ageing - The Case of the Internal Resistance Increase," in *IEEE Transactions on Industry Applications*, 2018, vol. 54, no. 1, pp. 517–525.

[13] Z. Song, X. Wu, X. Li, J. Sun, H. F. Hofmann, and J. Hou, "Current Profile Optimization for Combined State of Charge and State of Health Estimation of Lithium Ion Battery Based on Cramer–Rao Bound Analysis," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 7067–7078, 2019.

[14] J. Meng *et al.*, "An Overview and Comparison of Online Implementable SOC Estimation Methods for Lithium-Ion Battery," in *IEEE Transactions on Industry Applications*, 2018, vol. 54, no. 2, pp. 1583–1591.

[15] J. Meng, D. Stroe, M. Ricco, G. Luo, and R. Teodorescu, "A Simplified Model based State-of-Charge Estimation Approach for Lithium-ion Battery with Dynamic Linear Model," *IEEE Trans. Ind. Electron.*, p. 1, 2018.

[16] E. Locorotondo, L. Pugi, L. Berzi, M. Pierini, and A. Pretto, "Online State of Health Estimation of Lithium-Ion Batteries Based on Improved Ampere-Count Method," in *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 2018, pp. 1–6.

[17] D. I. Stroe, M. Swierczynski, A. I. Stroe, R. Laerke, P. C. Kjaer, and R. Teodorescu, "Degradation Behavior of Lithium-Ion Batteries Based on Lifetime Models and Field Measured Frequency Regulation Mission Profile," in *IEEE Transactions on Industry Applications*, 2016, vol. 52, no. 6, pp. 5009–5018.

[18] D.-I. Stroe, M. Swierczynski, A.-I. Stroe, S. K. Kaer, and R. Teodorescu, "Lithium-ion battery power degradation modelling by electrochemical impedance spectroscopy," *IET Renew. Power Gener.*, vol. 11, no. 9, pp. 1136–1141, 2017.

[19] X. Jin *et al.*, "Comparison of Li-ion battery degradation models for system design and control algorithm development," in *2017 American Control Conference (ACC)*, 2017, pp. 74–79.

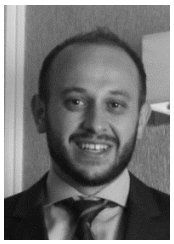
[20] L. Lam and P. Bauer, "Practical Capacity Fading Model for Li-Ion Battery Cells in Electric Vehicles," *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5910–5918, 2013.
[21] E. Prada, D. Di Domenico, Y. Creff, J. Bernard, V. Sauvant-Moynot, and F. Huet, "A Simplified Electrochemical and Thermal Aging Model of LiFePO4-Graphite Li-ion Batteries: Power and Capacity Fade Simulations," *J. Electrochem. Soc.* , vol. 160, no. 4, pp. A616–A628, Jan. 2013.
[22] 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, 2015.
[23] J. Wu, Y. Wang, X. Zhang, and Z. Chen, "A novel state of health estimation method of Li-ion battery using group method of data handling," *J. Power Sources*, 2016.
[24] E. Locorotondc *et al.*, "Modeling and simulation of Constant Phase Element for battery Electrochemical Impedance Spectroscopy," in *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*, 2019, pp. 225–230.
[25] M. Galeotti, L. Cinà, C. Giammanco, S. Cordiner, and A. Di Carlo, "Performance analysis and SOH (state of health) evaluation of lithium polymer batteries through electrochemical impedance spectroscopy," *Energy*, vol. 89, pp. 678–686, 2015.
[26] H. Pan, Z. Lü, H. Wang, H. Wei, and L. Chen, "Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine," *Energy*, 2018.
[27] D. Yang, Y. Wang, R. Pan, R. Chen, and Z. Chen, "A Neural Network Based State-of-Health Estimation of Lithium-ion Battery in Electric Vehicles," *Energy Procedia*, vol. 105, pp. 2059–2064, 2017.
[28] D. Yang, Y. Wang, R. Pan, R. Chen, and Z. Chen, "State-of-health estimation for the lithium-ion battery based on support vector regression," *Appl. Energy*, vol. 227, pp. 273–283, 2018.
[29] D. Liu, Y. Song, L. Li, H. Liao, and Y. Peng, "On-line life cycle health assessment for lithium-ion battery in electric vehicles," *J. Clean. Prod.*, vol. 199, pp. 1050–1065, 2018.
[30] R. R. Richardson, M. A. Osborne, and D. A. Howey, "Gaussian process regression for forecasting battery state of health," *J. Power Sources*, 2017.
[31] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, "A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve," *J. Power Sources*, 2018.
[32] K. Liu, Y. Li, X. Hu, M. Lucu, and D. Widanalage, "Gaussian Process Regression with Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-ion Batteries," *IEEE Trans. Ind. Informatics*, p. 1, 2019.
[33] G.-W. You, S. Park, and D. Oh, "Diagnosis of Electric Vehicle Batteries Using Recurrent Neural Networks," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4885–4893, 2017.
[34] L. Cai, J. Meng, D.-I. Stroe, G. Luo, and R. Teodorescu, "An evolutionary framework for lithium-ion battery state of health estimation," *J. Power Sources*, vol. 412, pp. 615–622, 2019.
[35] D.-I. Stroe, "Lifetime Models for Lithium-ion Batteries used in Virtual Power Plant Applications," Aalborg University, 2014.
[36] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, 2002.
[37] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
[38] R. Soentpiet, *Advances in kernel methods: support vector learning*. MIT press, 1999.

Lei Cai received the Ph.D degree from Northwestern Polytechnical University (NPU), Xi'an, China, in 2017. He was supported by the China Scholarship Council as a joint Ph.D. student with the School of Computer Science, University of Birmingham, Birmingham, U.K. He is currently a lecturer with the Faculty of Computer Science and Engineering, Xi'an University of Technology, Xi' an, China.

His current research interests include evolutionary computation, optimization, data-driven battery modeling, battery states estimation, and energy management.

Jinhao Meng received the M.S. degree in control theory and control engineering and the Ph.D. degree in electrical engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 2013 and 2019, respectively. He was supported by the China Scholarship Council as a joint Ph.D. student with the Department of Energy Technology, Aalborg University, Aalborg, Denmark. He is currently an associate researcher in Sichuan University, Chengdu, China.

His research interests include battery modeling, battery states estimation, and energy management of battery energy storage system.



Daniel-Ioan Stroe (M'11) received the Dipl.Ing. degree in automatics from the Transilvania University of Brasov, Brasov, Romania, in 2008, and the M.Sc. degree in wind power systems and the Ph.D. degree in lifetime modelling of Lithium-ion batteries from Aalborg University, Aalborg, Denmark, in 2010 and 2014, respectively. He is currently an Assistant Professor with the Department of Energy Technology, Aalborg University. He was a Visiting Researcher at RWTH Aachen, Germany, in 2013. He has co-authored more than 70 journals and conference papers. His current research interests include energy storage systems for grid and e-mobility, Lithium-based batteries testing and modelling, and lifetime estimation of Lithium-ion batteries.

Jichang Peng received the B.S. degree in electrical engineering from North China University of Water Resources and Electric Power, Zhengzhou, China, in 2010, and the M.S. degree in control theory and control engineering and the Ph.D. degree in electrical engineering, both from Northwestern Polytechnical University, Xi'an, China, in 2013 and 2019, respectively. He is currently a lecturer in Nanjing Institute of Technology, Nanjing, China.

His research interests include battery modeling, energy management, and aircraft starter/generator and sensorless control of brushless synchronous machines.

Guangzhao Luo (SM'19) received the M.S. and Ph.D. degrees in electrical engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 1998 and 2003, respectively. From 2003 to 2004, he was a Postdoctoral Research at the University of Federal Defense, Munich, Germany. He is currently a Professor with NPU. He is the Vice Director of the Rare Earth Permanent Magnet (REPM) Electric Machine and Control Engineering Center, Shaanxi Province. His research interests include advance control theory of permanent magnet electrical machine, high performance control technology of permanent magnet synchronous motor for electric traction and electric vehicle, real-time simulation technology for electrical drive system, and intelligence control of new energy conversion. Dr. Luo received the Second Prize from the China National Defense Science and Technology Progress Award in 1995 and 2011.

Remus Teodorescu (F'12) received the Dipl.Ing. degree in electrical engineering from the Polytechnical University of Bucharest, Bucharest, Romania, in 1989, and the Ph.D. degree in power electronics from the University of Galati, Galati, Romania, in 1994. In 1998, he joined the Power Electronics Section, Department of Energy Technology, Aalborg University, Aalborg, Denmark, where he is currently a Full Professor. Since 2013, he has been a Visiting Professor with Chalmers University. His research interests include design and control of grid-connected converters for photovoltaic and wind power systems, high voltage dc/flexible ac transmission systems based on modular multilevel converters, and storage systems based on Li-ion battery technology including modular converters and active battery management systems.