IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 31, 2017, accepted September 24, 2017, date of publication October 4, 2017, date of current version October 25, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2759094

State-of-Health Estimation for Lithium-Ion Batteries Based on the Multi-Island Genetic Algorithm and the Gaussian Process Regression

ZHENPO WANG, JUN MA, AND LEI ZHANG, (Member, IEEE)
National Engineering Laboratory for Electric Vehicles, Collaborative Innovation Center for Electric Vehicles, Beijing Institute of Technology, Beijing 100081, China

Corresponding author: Lei Zhang (lei_zhang@bit.edu.cn)

The work was supported by the State Key Program of National Natural Science Foundation of China under Grant U1564206.

ABSTRACT Battery State-of-Health (SOH) estimation is of utmost importance for the performance and cost-effectiveness of electric vehicles. Incremental capacity analysis (ICA) has been ubiquitously used for battery SOH estimation. However, challenges remain with regard to the characteristic parameter selection, estimation viability and feasibility for practical implementation. In this paper, a novel ICA-based method for battery SOH estimation is proposed, with the goals to identify the most effective characteristic parameters of IC curves, optimize the SOH model parameters for better prediction accuracy and enhance its applicability in realistic battery management systems. To this end, the IC curve is first derived and filtered using the wavelet filtering, with the peak value and position extracted as health factors (HFs). Then, the correlations between SOH and HFs are explored through the grey correlation analysis. The SOH model is further established based on the Gaussian process regression (GPR), in which the optimal hyper parameters are calculated through the conjugate gradient method and the multi-island genetic algorithm (MIGA). The effects of different HFs and kernel functions are also analyzed. The effectiveness of the proposed MIGA-GPR SOH model is validated by experimentation.

INDEX TERMS Batteries, incremental capacity analysis, state of health, Gaussian process regression, multi-island genetic algorithm.

I. INTRODUCTION

Electric vehicles (EVs) are widely recognized as a viable solution to deal with the problems of oil depletion and environmental pollution [1], [2]. Battery systems are key components of EVs, and strongly influence their driving performance and driving range per charge [3]. Tremendous progress has been made in the pursuit of affordable and durable rechargeable batteries in recent decades. So far, lithium-ion batteries represent a leading candidate in this regard, thanks to their intrinsic advantages including high energy density and long lifespan along with no memory effect compared with other battery chemistries [4]–[6]. Battery management systems (BMSs) are always necessarily positioned to ensure safe, efficient and reliable operation of battery systems, in which State-of-Health (SOH) estimation is a crucial but challenging task [7]. Accurate SOH monitoring would contribute to effective battery health management, which helps avoid catastrophic hazards and premature failure [8]. However, a comprehensive and clear statement of the underlying mechanisms of battery degradation is still absent at present due to complexity and coupled effects of the involved electrochemical reactions inside. Besides, different types of lithium-ion batteries involve with distinct aging mechanisms that are dependent on battery structural design and used materials.

In view of the significance of accurate SOH estimation, many attempts have been directed to revealing battery degradation mechanisms and correspondingly developing SOH estimation schemes, giving rise to a rich library of related literature. The state-of-art SOH estimation methods have been systematically summarized in several review papers [9]–[11]. Generally, these methods can be sorted into

three categories, i.e., empirical or semi-empirical models, physical-based and data-driven approaches, each with advantages and limitations.

Empirical or semi-empirical models are derived by fitting experimental data. They have satisfactory prediction accuracy under certain conditions for a specific battery type but lack physical meanings. Nevertheless, the volatility and unpredictability of working conditions for battery systems during practical vehicular operation may make these laboratory-calibrated models incapable of providing precise SOH estimation. In contract, physical-based models are developed based on the first principle to account for realistic side reactions that lead to battery degradation. For example, the growth of solid electrolyte interface (SEI) film has widely been considered as the main source of capacity loss and loss of active material (LAM), resulting in a flotilla of electrochemical models [12], [13]. These physical-based models can provide direct interpretations for some battery degradation phenomena, and enable the understanding of electrochemical dynamics. However, due to the use of coupled partial differential equations (PDEs), the extremely high computational requirements impede their practical applicability in realistic BMSs. Data-driven approaches for battery SOH estimation have emerged, owing to their flexibility and model-free characteristics. A variety of data-driven approaches have been put forward for battery SOH assessment, including support vector machine (SVM) [14]–[16], Bayesian network [17], [18], Autoregressive model [19], [20], particle filter (PF) [21] and Gaussian process regression [22], [23]. For instance, the SVM approaches can depict the non-linear relationship between measureable features and battery SOH using kernel functions, and are considered to outperform ordinary regression methods due to its insensitivity to data uncertainty. Recently, signal processing methodologies are also employed for modeling battery degradation, including incremental capacity analysis (ICA) and differential voltage analysis (DVA) [24]–[27]. Both ICA and DVA have been used as enabling in-situ techniques to explicitly identify battery degradation mechanisms for a variety of battery types [28]. For instance, Dubarry et al. [29] utilized the ICA method to analyze battery degradation process for large format $LiFePO_4$ (LFP) batteries and their path dependence in plug-in hybrid electric vehicle (PHEV) application. Weng *et al*. [30] and Feng *et al.* [31] applied ICA associated with support vector regression and probability density function to SOH monitoring, respectively. Berecibar *et al.* compared three different regression models for battery SOH estimation based on the use of ICA and/or DVA. Initially developed for investigating the intercalation-based battery cells, ICA can transform voltage plateaus of a charging/discharging voltage curve into peaks of the corresponding incremental capacity (IC) curve by differentiating the charged battery capacity verse the terminal voltage. It has the advantages of insensitivity to battery types and efficacy in identifying capacity loss mechanisms. Additionally, it has huge potential for online battery degradation diagnosis since only a low-rate constant charge regime is required to derive the IC curve. Despite the widespread use of ICA-based methods for battery SOH estimation, there are still substantial challenges in terms of effective characteristic feature extraction of IC curves, and estimation viability and applicability in realistic BMSs.

In this paper, the IC curve is derived and filtered using the wavelet filtering, with the peak value and position extracted as health factors (HFs). Then, the correlations between SOH and HFs are explored through the grey correlation analysis. The Gaussian process regression is applied to estimating SOH with mean and confidence intervals that are considered as the uncertainty representations of SOH. To optimize the hyper parameters, an improved GPR method combining the multi-island genetic algorithm is utilized to capture the actual degradation of Lithium-ion battery with HFs extracted from the ICA. The effectiveness of the proposed method has been validated through experimentation.

The remainder of this paper is structured as follows: Section II gives a brief introduction for the ICA method. Section III reveals details of experimentation for battery charging dataset collection. Section IV presents the proposed SOH model and its validation, whose parameters are identified using the Gaussian process regression based on the conjugate gradient method and the Multi-Island Genetic Algorithm (MIGA). The key conclusions are summarized in Section V.

II. INCREMENTAL CAPACITY ANALYSIS

The ICA method is capable of providing SOH estimation with high accuracy for lithium-ion batteries using basic measurable parameters such as voltage and current. The charging curves under the constant-current-constant-voltage (CC-CV) regime are usually used to generate the IC curves through differentiation. It has been demonstrated in literature that the incremental capacity $\Delta Q/\Delta V$ is more indicative of the underlying electrochemical changes during battery aging than the conventional charging/discharging curves themselves. For example, during a constant current charging process, the following relationship holds:

$$dQ/dV = I \cdot dt/dV \tag{1}$$

where Q is the charged capacity, V is the terminal voltage, and I is the loading current.

It is obvious that the incremental capacity is inversely proportional to dV/dt. The constant current charging curve can be translated into the incremental capacity curve according to Eq. (2).

$$Q = It, \quad V = f(Q), \; Q = f^{-1}(V) \tag{2}$$

The slope dQ/dV can be re-written as a function of V according to Eq. (3)

$$(f^{-1})' = \frac{dQ}{dV} = \frac{d(It)}{dV} = \frac{I}{dV/dt} = g(V) \tag{3}$$

In the data processing, dV and dQ are replaced by δV and δQ in the discrete form, respectively. Thus, when $\delta V \to 0$,

$$\frac{\mathrm{d}Q}{dV} \approx \frac{\delta Q}{\delta V} \tag{4}$$

A detailed calculation will be provided in the following parts. The voltage sequence V and current sequence I correspond to the voltage and current at the time sequence T, respectively.

$$\begin{aligned} V &= (V_1, V_2, V_3 \cdots, V_n) \\ I &= (I_1, I_2, I_3, \cdots, I_n) \\ T &= (t_1, t_2, t_3, \cdots, t_n) \end{aligned} \tag{5}$$

Therefore, the incremental capacity from V_k to V_{k+1} can be expressed as

$$Q_{V_k \to V_{k+1}} = I_k(t_{k+1} - t_k) \tag{6}$$

The voltage range is set as $[V_{min}, V_{max}]$, which covers the entire voltage change in the charging process. Based on the method of integral forward, when $m_k = m_{k+1} = \ldots = m_{k+s} = m$, the incremental capacity at V_m can be derived as

$$\frac{\mathrm{d}Q}{dV} \approx \left.\frac{\delta Q}{\delta V}\right|_{V_m} = \sum_{h=k}^{k+s} Q_{V_h \to V_{h+1}} = \sum_{h=k}^{k+s} I_h(t_{h+1} - t_h) \tag{7}$$

where $h = k, k+1, \ldots, k+s, \delta V = 1\mathrm{mV}$.

It is difficult to directly derive the peaks on the incremental capacity curve since their corresponding points on the charging/discharging curve always locate in the voltage plateau region that is flat and sensitive to measurement noise. Hence, effective and robust algorithms are needed to filter the charging/discharging curve and extract the signature features. To this end, the machine learning methods such as support vector machine [30] and the probability density function [31] have been proposed to extract the signature onboard. In this study, the wavelet transform (WT) filtering is used to smooth the incremental capacity curve due to its predominance over other traditional filtering methods. Different from the traditional filtering techniques that always base on linear methods, the WT uses nonlinear methods to process the discrete signals, and has a good ability at localizing signal in both time and frequency planes. It can analyze and decompose an original signal into different localized contributions, with each representing a part of the signal covering a different frequency range. The success of WT filtering method is mainly due to the features of wavelet transform such as low entropy, multiresolution, decorrelation and flexibility of base selection, etc. The mother wavelet function is critical since different wavelet basis functions give rise to different results. The most commonly used wavelet basis functions include the Symlets wavelet and the Daubechies wavelet. Here, the Daubechies wavelet is used to smooth the incremental capacity curve with five wavelet decomposition levels in this study. A comparison of the original and filtered ICA curves is shown in Fig. 1.

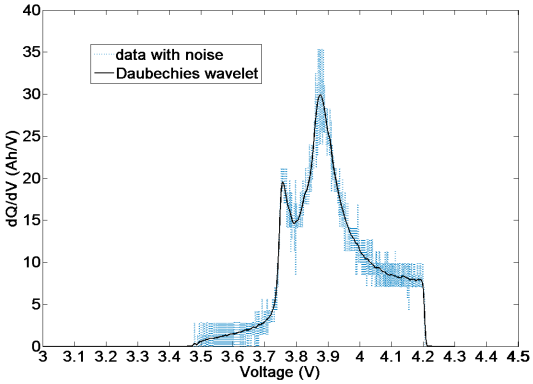

FIGURE 1. The incremental capacity curve filtered by the Daubechies wavelet.

It can be seen that the WT method can effectively filter out the noise and derive a smooth ICA curve.

III. EXPERIMENTAL

In this section, experimental work is elaborated, together with an analysis of the influence of the charging rate, temperature, and aging. A test rig was established in order to obtain the charging/discharging data. It comprises of a battery cycler, a thermal chamber, a host computer, and the test Li-ion batteries. The battery cycler is used to charge or discharge the test batteries in accordance with pre-defined loading profiles. It has the capability of recording various parameters, including terminal voltage, loading current and accumulated capacity. The accuracy of the voltage and current measurement is up to 1 mV and 1 mA, respectively. The pouch NMC batteries with a capacity of 10 Ah were used in this study. The detailed specifications of the tested cells are listed in Table 1.

TABLE 1. The detailed specification of the tested cells.

Specification	
Anode	Graphite
Cathode	$LiNi_{1/3}Mn_{1/3}Co_{1/3}O_2$ (NMC)
Internal resistance	≤18mΩ
Weight	≤290g
Size	85mm × 56 mm × 24 mm
Charging/discharging rate	0.05C, 0.25C, 0.5C, 0.75C,1C
End of discharge voltage	3.0 V
Max. charge voltage	4.2 V

The CC-CV charging regime was used to charge the battery with a cut-off voltage of 4.2V. As the performance of lithium-ion batteries is highly sensitive to charging rates and environmental temperatures, the incremental capacity curves under different charging rates (0.05C, 0.25C, 0.5C, 0.75C and 1C) under 25°C are derived and shown in Fig. 2. It is obvious that all the incremental capacity curves have similar patterns with two peak values under different charging rates. With the increasing charging rate, both the related peak value

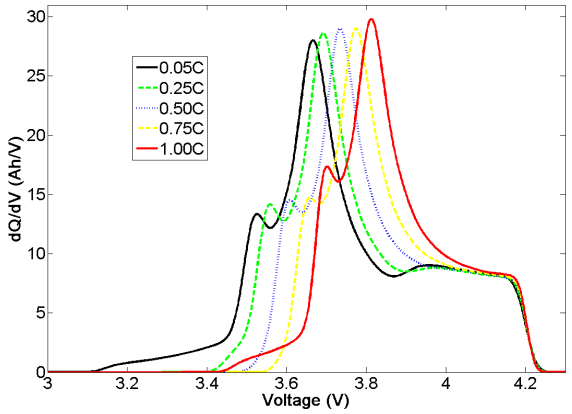

FIGURE 2. The incremental capacity curves under different charging C-rates.

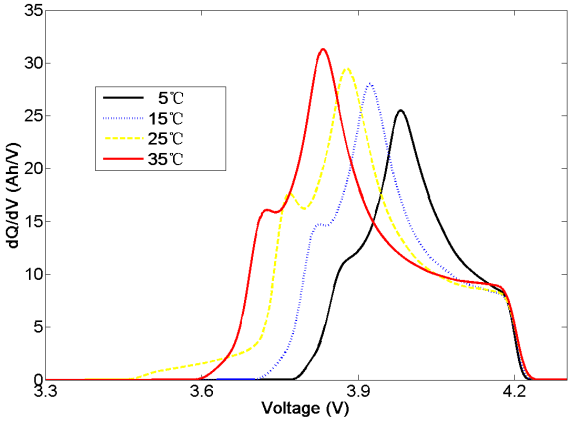

FIGURE 3. The incremental capacity curves under different temperatures.

and the corresponding voltage position move right to higher voltages. In order to manifest the impact of temperature, the incremental capacity curves under different temperatures (5°C, 15°C, 25°C and 35°C) with 1 C-rate charging are illustrated in Fig. 3. Also, all the incremental capacity curves have the similar shapes. However, along with the increasing temperature, the related peak value increases while the corresponding voltage position moves left to lower voltages.

In order to obtain the incremental capacity curves under a variation of aging status, the accelerated aging tests were performed to simulate the real battery aging process. The upper and lower cut-off voltages are set as 4.2V and 3.0V, respectively. It is worth mentioning that the data during the constant voltage charging is excluded for analysis for the sake of simplification. The incremental capacity curves under different cycles (Cycles of 1, 60, 120, 180, 240, and 300) are sketched in Fig. 4. It can be seen that there are two peaks at the early stage of accelerated aging tests, which gradually fade away with the increasing number of cycles. Meanwhile, the peak value at high voltage exhibits an obvious decline trend while the peak position moves right to higher voltage with aging. It is worth noted that three repeated discharging at 1C-rate were conducted, the capacities of which were

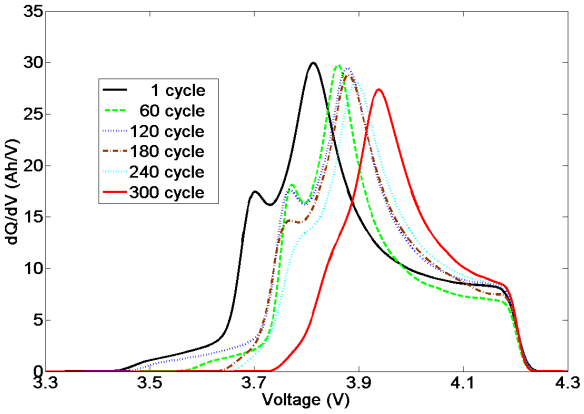

FIGURE 4. The incremental capacity curves under different cycles.

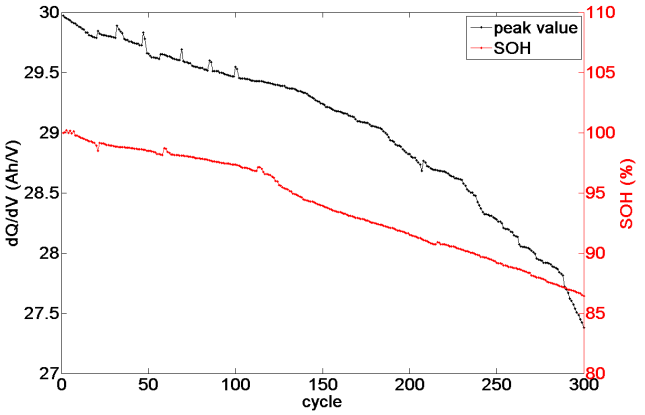

FIGURE 5. The relationship between the peak value and SOH.

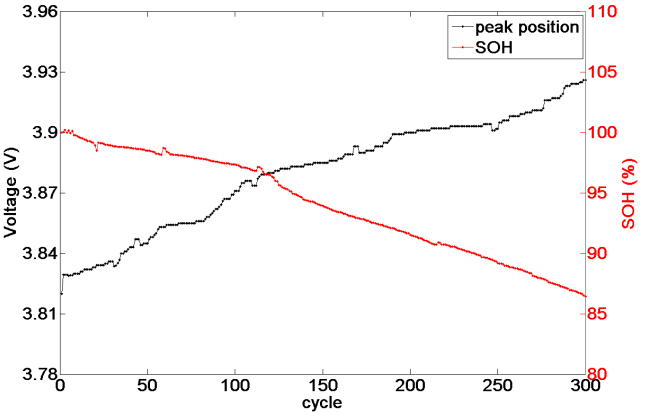

FIGURE 6. The relationship between the peak position and SOH.

9.83Ah, 9.81Ah and 9.816Ah, respectively. Due to the consistency of three repeated capacity tests, the average capacity of 9.818Ah was selected as the benchmark battery capacity for SOH assessment. The battery SOH degrades with the increasing number of cycles, resulting in changed peak values and position. The relationship between the peak value/position and SOH is shown in Fig. 5 and Fig. 6, respectively. Therefore, the peak value and peak position voltage can be extracted as indicative HFs for battery SOH estimation.

Based on the grey correlation analysis presented in Ref. [32], the SOH can be selected as Y reference sequence, while the peak value and peak position voltage can be regarded as X_i comparative sequences as shown below.

$$X_i = \{x_i(k) | \, k = 1, 2, \cdots, n\} \tag{8}$$

$$Y = \{y(k) | \, k = 1, 2, \cdots, n\} \tag{9}$$

where $i = 1, 2, \ldots, m$, m is the number of comparative sequence, n is the length of comparative sequence or reference sequence.

Correlation coefficient $\xi_i(k)$ can be calculated,

$$\xi_i(k) = \frac{\min\limits_i \min\limits_k |y(k) - x_i(k)| + \rho \max\limits_i \max\limits_k |y(k) - x_i(k)|}{|y(k) - x_i(k)| + \rho \max\limits_i \max\limits_k |y(k) - x_i(k)|} \tag{10}$$

where $\rho \in [0, 1]$ is a resolution factor. So correlation degree of X_i comparative sequence and Y reference sequence shows

$$r_i = \frac{1}{n} \sum_{k=1}^{n} \xi_i(k), \quad k = 1, 2, \cdots, n \tag{11}$$

where $r_i \in [0, 1]$, when r_i approximates to 1, the relevance of X_i and Y is high.

Calculation results show that the relevance of peak value and SOH is 0.6383, and the relevance of peak position and SOH is 0.6086. Thus, the relevance of peak value and SOH is higher, namely, peak value as HF is better than peak position which is also demonstrated in the following section.

IV. MODEL BASED ON GAUSSIAN PROCESS REGRESSION

The Gaussian process is defined as a collection of random variables,

$$\mathbf{S} = \{(\mathbf{x}_i, y_i) | \, i = 1, 2, \cdots, n\} = (\mathbf{X}, \mathbf{y}) \tag{12}$$

where $\mathbf{x}_i \in \boldsymbol{R}^d$ is the input matrix, and $y_i \in \boldsymbol{R}$ is the output indexed by $\mathbf{x}_i$. The stochastic process is characterized by the probability distribution for each finite subset of variables $f(\cdot)$: $\boldsymbol{R}^d \mapsto \boldsymbol{R}$ in a consistent manner. A Gaussian process $f(\mathbf{x})$ can be fully described by its mean μ and covariance matiax Σ, which is defined as

$$f(\mathbf{x}) \sim N\left(\mu, \Sigma^2\right) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left\{-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right\} \tag{13}$$

In the Gaussian process regression model, the mean value is zero, and the covariance function can be the squared exponential covariance function, the Mattern class of covariance function or the neural network covariance function.

The squared exponential covariance function (SEISO):

$$\mathrm{k}(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \exp\left(-\frac{1}{2} (\mathbf{x}_p - \mathbf{x}_q)^T M (\mathbf{x}_p - \mathbf{x}_q)\right) \tag{14}$$

where $\boldsymbol{M}$ is the covariance scale matrix.

Generally, $M = \begin{bmatrix} \lambda^{-2} & & \\ & \ddots & \\ & & \lambda^{-2} \end{bmatrix}$,

Thus, Eq. (14) can be further reformulated as

$$\mathrm{k}(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \exp\left(-\frac{1}{2\lambda^2} |\mathbf{x}_p - \mathbf{x}_q|^2\right) \tag{15}$$

The squared exponential kernel is a stationary kernel since the correlation between points is purely a function of the difference in their inputs, $\mathbf{x}_p$-$\mathbf{x}_q$.

The Mattern class of covariance function has many forms, and the common form (M3ISO) is

$$k(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \left(1 + \frac{\sqrt{3} |\mathbf{x}_p - \mathbf{x}_q|}{\lambda}\right) \exp\left(-\frac{\sqrt{3} |\mathbf{x}_p - \mathbf{x}_q|}{\lambda}\right) \tag{16}$$

The neural network covariance function (NN) is

$$k(\mathbf{x}_p, \mathbf{x}_q) = \sigma_f^2 \sin^{-1}\left(\frac{\bar{\mathbf{x}}_p^T \cdot \lambda^{-2} \cdot \bar{\mathbf{x}}_q}{\sqrt{\left(1 + \bar{\mathbf{x}}_p^T \cdot \lambda^{-2} \cdot \bar{\mathbf{x}}_p\right)\left(1 + \bar{\mathbf{x}}_q^T \cdot \lambda^{-2} \cdot \bar{\mathbf{x}}_q\right)}}\right) \tag{17}$$

where $\bar{\mathbf{x}}_p = \begin{bmatrix} 1 \\ \mathbf{x}_p \end{bmatrix}$, $\bar{\mathbf{x}}_q = \begin{bmatrix} 1 \\ \mathbf{x}_q \end{bmatrix}$. $\mathbf{x}_p$ and $\mathbf{x}_q$ represent two different inputs, and the covariance function parameters σ_f and λ control the y-scaling and x-scaling, respectively.

There are some free parameters in the covariance functions, that is, $\theta = [\sigma_f, \sigma_n, \lambda]$. Generally, the hyper parameters are needed to be optimized by marginal likelihood function $p(y|X, \theta)$ as shown in Eq. (14).

$$p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\alpha|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2} \mathbf{y}^T \beta\right\}$$
$$\alpha = \mathbf{cov}(\mathbf{x}) + \sigma_n^2 \boldsymbol{I}, \quad \beta = (\mathbf{cov}(\mathbf{x}) + \sigma_n^2 \boldsymbol{I})^{-1} \mathbf{y} \tag{18}$$

where $\boldsymbol{I}$ is the unit matrix with d dimensions, d is the number of training data sets, and $\mathbf{cov}(\mathbf{x})$ represents the covariance of $\mathbf{x}$.

Based on the conjugate gradient method, the hyper parameters $\boldsymbol{\theta}$ can be obtained with the maximization of the log-likelihood function given by

$$L(\boldsymbol{\theta}) = \log p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{2} \mathbf{y}^T \beta - \frac{1}{2} \log(|\alpha|) - \frac{n}{2} \log(2\pi) \tag{19}$$

Given the effect of noise, the target y can be described as $y = f(\mathbf{x}) + \varepsilon$, where ε is the white Guassian noise, and $\varepsilon \sim N(0, \sigma_n^2)$. If $f(\mathbf{x})$ is a Gaussian process, the joint distribution of limited observations from y is also a Gaussian process. Once a posterior distribution is derived, it can be used to estimate predictive values for the test data points. The following equations describe the predictive distribution for GPR.

Prior:

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{K}(\mathbf{x}) + \sigma_n^2\mathbf{I} & \mathbf{k}(\mathbf{x}^*) \\ \mathbf{k}(\mathbf{x}^*)^{\mathrm{T}} & \mathrm{cov}(y^*) \end{bmatrix}\right) \tag{20}$$

$$\mathbf{k}(\mathbf{x}^*) = [k(\mathbf{x}^*, \mathbf{x}_1), k(\mathbf{x}^*, \mathbf{x}_2), \cdots, k(\mathbf{x}^*, \mathbf{x_n})]^{\mathrm{T}} \tag{21}$$

Posterior:

$$\overline{y^*} = \mathrm{k}(\mathbf{x}^*)^T[\mathrm{K}(\mathbf{x}) + \sigma_n^2\mathbf{I}]^{-1}\mathbf{y} \tag{22}$$

$$\mathrm{cov}(y^*) = k(\mathbf{x}^*, \mathbf{x}^*) + \sigma_n^2 - \mathbf{k}(\mathbf{x}^*)^{\mathrm{T}}[\mathrm{K}(\mathbf{x}) + \sigma_n^2\mathbf{I}]^{-1}\mathbf{k}(\mathbf{x}^*) \tag{23}$$

$$y^* \sim N\left(\bar{y}^*, \mathrm{cov}(y^*)\right) \tag{24}$$

where **(x,y)** represents the training points, $\mathbf{x}^*$ and y^* stands for the test inputs and prediction values, and $\overline{y^*}$ is the mean prediction value.

When $Z \sim N(u, \sigma_n^2)$, the confidence interval with $(1-\alpha)$ confidence level is written as $[u - z_{\alpha/2}\sigma, u + z_{\alpha/2}\sigma]$. Generally, $z_{\alpha/2} = 1.96$ when $(1-\alpha) = 0.95$. So in the GPR model, the confidence interval with 95% confidence level is described as follow:

$$\left[\overline{y^*} - 1.96\sqrt{\mathrm{cov}(y^*)}, \overline{y^*} + 1.96\sqrt{\mathrm{cov}(y^*)}\right] \tag{25}$$

A. CONJUGATE GRADIENT METHOD

The Gaussian process regression is used to establish the SOH model with the HFs as the inputs and the SOH estimation as the output. The hyper parameters are calculated by the conjugate gradient method, which possesses good convergence property. A set of conjugate directions are constructed by the gradient of known points, and the hyper parameters are optimized with the maximization of the objective function. The flowchart of the proposed GPR model is shown in Fig. 7.

(1) Two characteristic parameters, namely the peak value and peak position, are extracted as the health factors from the incremental capacity curves. Combining with battery SOH, two sample collections are made up, and respectively are (peak value, SOH) and (peak position, SOH).

(2) The training data $(\mathbf{x},\mathbf{y})$ and test data $(\mathbf{x}^*, y^*)$ are screened out from the two sample collections as mentioned above.

(3) For the training, the GPR model is proposed and the SOH estimation is given by

$$\mathbf{y} \sim N(0, \mathbf{K}(\mathbf{x}) + \sigma_n^2\mathbf{I}) \tag{26}$$

which conforms to the Gaussian distribution.

(4) Based on the conjugate gradient method, the hyper parameters $\theta_\mathbf{0} = [\sigma_f, \sigma_n, \lambda]$ can be obtained with the maximization of the log-likelihood function as shown in Eq. (15). The proper search direction is chose as

$$\mathbf{d}_k = \begin{cases} -\nabla L(\boldsymbol{\theta}_k) + \beta_{k-1}\mathbf{d}_{k-1} & k \geq 1 \\ -\nabla L(\boldsymbol{\theta}_k) & k = 0 \end{cases}$$

$$\beta_{k-1} = \frac{\nabla L(\boldsymbol{\theta}_k)^{\mathrm{T}} \cdot \nabla L(\boldsymbol{\theta}_k)}{\nabla L(\boldsymbol{\theta}_{k-1})^{\mathrm{T}} \cdot \nabla L(\boldsymbol{\theta}_{k-1})} \tag{27}$$

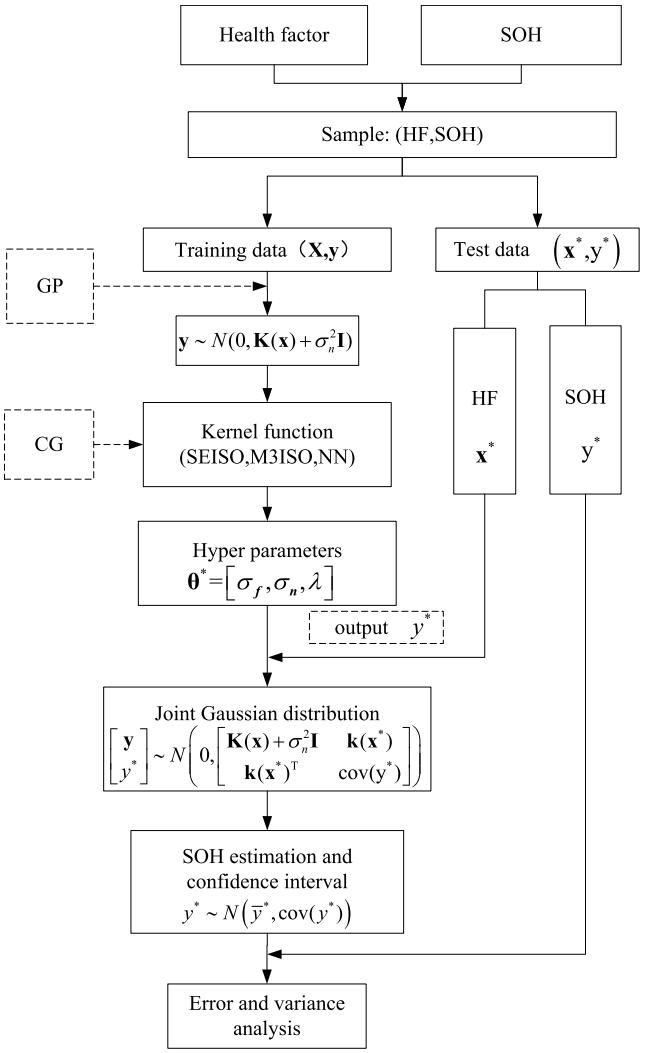

FIGURE 7. The flow chart of Gaussian Process Regression model for battery SOH estimation.

The searching step size is calculated with the minimization of $L(\theta_\mathrm{k} + \alpha_k\mathbf{d}_\mathrm{k})$

$$L(\boldsymbol{\theta}_k + \alpha_k\mathbf{d}_k) = \min_{\alpha \geq 0} L(\boldsymbol{\theta}_k + \alpha\mathbf{d}_k) \tag{28}$$

until meeting the stop condition.

$$\|\nabla L(\boldsymbol{\theta}_k)\| \leq \varepsilon \tag{29}$$

(5) SOH estimation is completed by new input $\mathbf{x}^*$ and output y^* with the confidence interval given by

$$y^* \sim N\left(\bar{y}^*, \mathrm{cov}(y^*)\right) \tag{30}$$

B. MULTI-ISLAND GENETIC ALGORITHM

Genetic Algorithms have been successfully applied to a variety of optimizaiton problems [33], [34]. The multi-island genetic algorithm (MIGA) represents an iterative procedure evolving a population of individuals that correspond to candidate solutions. During each generation, the individuals in

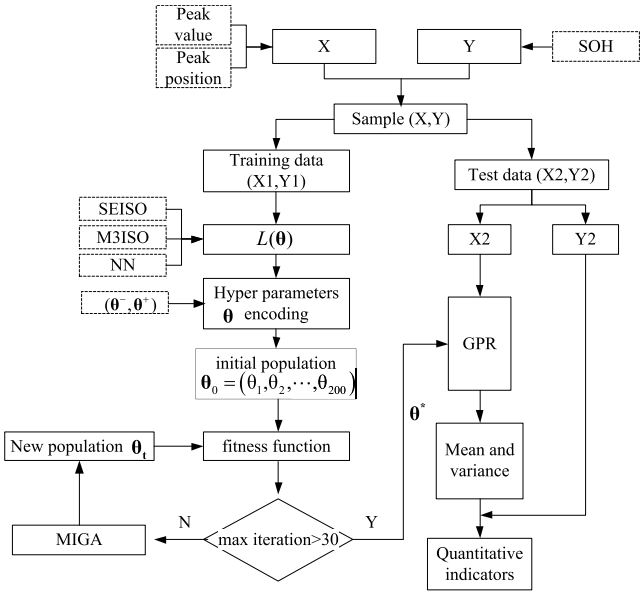

FIGURE 8. The flow chart of MIGA-GPR model for battery SOH estimation.

the current population are rated based on the fitness function, and a new population of candidate solutions is formed through reproduction operators, such as selection, crossover, mutation, and migration. The MIGA can keep the diversity of population through migration and avoid local premature converge, and has an obvious advantage in searching the global optimal solution relative to the conventional GA. Therefore, the MIGA is introduced here to optimize the hyper parameters of the GPR model. The resultant MIGA-GPR model for battery SOH estimation is presented in Fig. 8.

Three kinds of kernel functions (i.e., SEISO, M3ISO and NN) are analyzed here, and the hyper parameters are encoded using the grey code with an initial population size of 200, a islands number of 5, a crossover probability factor of 0.7, a mutation probability factor of 0.01 and a migration probability factor of 0.1. The maximum iteration is set 30. The log-likelihood function comparing CG and post-optimization by the MIGA is shown in Table 2. The value of $L(\theta)$ has a descent trend after the hyper parameters optimization by the MIGA. The SOH estimation results with SEISO, M3ISO and NN are shown in Fig. 9, Fig. 10 and Fig. 11, respectively.

TABLE 2. The log-likelihood function comparing between CG and MIGA.

Kernel function	Health factors	Log-likelihood function		descent rate / %
		CG	MIGA	
SEISO	Peak value	76.93	55.14	28.32
	Peak position	71.59	63.21	11.70
M3ISO	Peak value	71.24	68.73	3.52
	Peak position	82.05	70.72	13.80
NN	Peak value	68.69	62.99	8.29
	Peak position	70.15	65.93	6.01

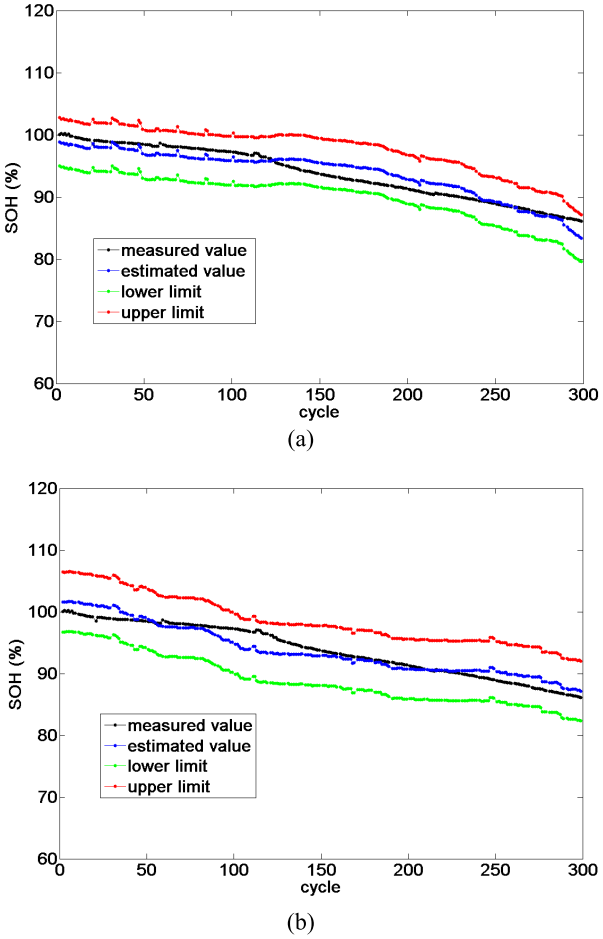

FIGURE 9. The SOH estimation results with SEISO. (a) Peak value as health factor. (b) Peak position as health factor.

C. RESULTS AND ANALYSIS

Quantitative indicators are employed to assess the SOH prediction accuray. They are the max error ε_1, the mean error ε_2, the max relative error ε_3, the mean relative error ε_4, the maximum width of confidence interval ε_5, the average width of confidence interval ε_6, the maximum width ratio of confidence interval ε_7, and the average width ratio of confidence interval ε_8.

$$\varepsilon_1 = \max\{|\bar{y}_i - y_i|\}, \quad \varepsilon_2 = \frac{1}{n}\sum_{i=1}^{n}|\bar{y}_i - y_i|$$

$$\varepsilon_3 = \max\left\{\left|\frac{\bar{y}_i - y_i}{y_i}\right|\right\} \times 100\%$$

$$\varepsilon_4 = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\bar{y}_i - y_i}{y_i}\right| \times 100\%$$

$$\varepsilon_5 = \max\left|\bar{y}_i^+ - \bar{y}_i^-\right|, \quad \varepsilon_6 = \frac{1}{n}\sum_{i=1}^{n}\left|\bar{y}_i^+ - \bar{y}_i^-\right|$$

$$\varepsilon_7 = \max\left|\frac{\bar{y}_i^+ - \bar{y}_i^-}{y_i}\right| \times 100\%$$

TABLE 3. Quantitative indicators of SOH estimation results with SEISO, M3ISO and NN based on MIGA-GPR model.

Quantitative indicators	SEISO		M3ISO		NN	
	Peak value	Peak position	Peak value	Peak position	Peak value	Peak position
ε_1	2.89	3.72	3.21	3.19	2.76	3.53
ε_2	1.10	1.20	1.37	1.18	1.27	1.17
ε_3	2.97	3.86	3.70	3.24	3.21	3.64
ε_4	1.16	1.28	1.48	1.27	1.36	1.24
ε_5	7.86	9.67	16.62	13.29	13.52	12.41
ε_6	7.80	9.59	13.48	11.94	11.90	11.91
ε_7	8.85	11.23	19.31	15.44	13.57	14.41
ε_8	8.34	10.34	14.41	12.76	12.73	12.72

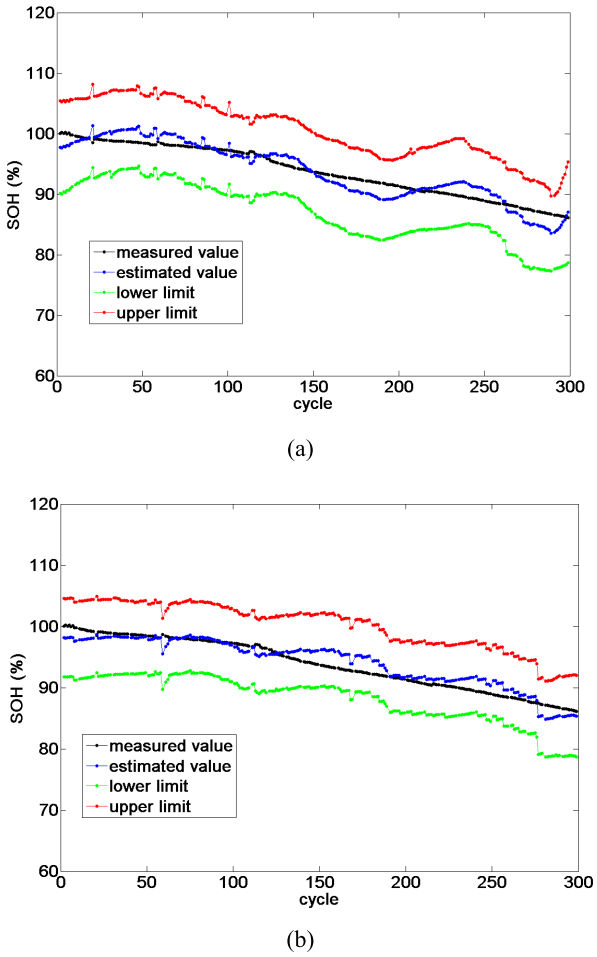

FIGURE 10. The SOH estimation results with M3ISO. (a) Peak value as health factor. (b) Peak position as health factor.

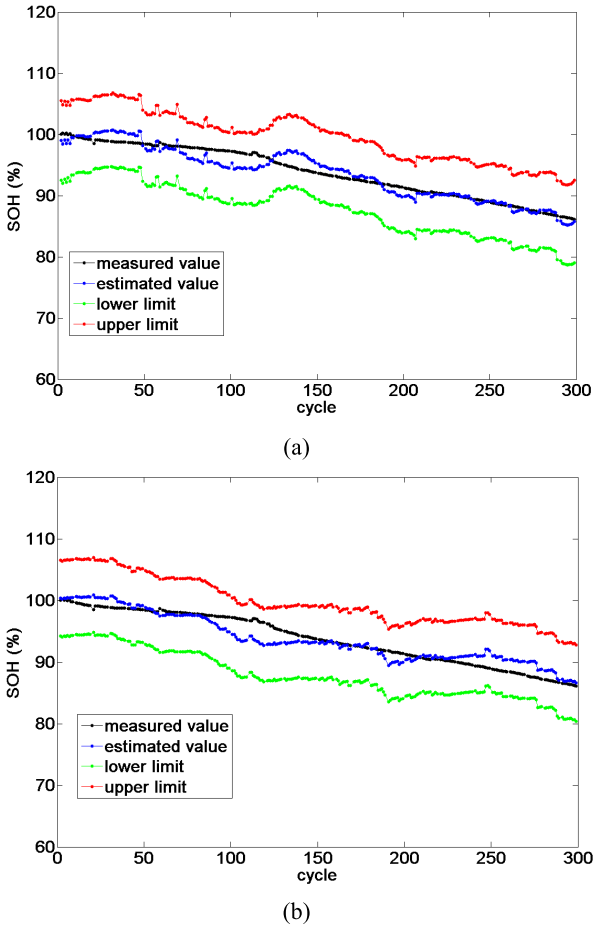

FIGURE 11. The SOH estimation results with NN. (a) Peak value as health factor. (b) Peak position as health factor.

$$\varepsilon_8 = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\bar{y}_i^+ - \bar{y}_i^-}{y_i}\right| \times 100\%$$
$$\bar{y}_i^+ = \overline{y_i^*} + 1.96\sqrt{\mathrm{cov}(y_i^*)}$$
$$\bar{y}_i^- = \overline{y_i^*} - 1.96\sqrt{\mathrm{cov}(y_i^*)} \quad (31)$$

where $\bar{y}_i^+$, $\bar{y}_i^-$, $\bar{y}_i$ and y_i are the upper limit, the lower limit, the predicted value and the true value, respectively.

The first four quantitative indicators (ε_1, ε_2, ε_3, ε_4) are used to evaluate the SOH estimation error. And, the other quantitative indicators (ε_5, ε_6, ε_7, ε_8) are used to assess the reliability and stability of the proposed SOH model based on the confidence interval. The SOH estimation results under the SEISO, M3ISO and NN are shown in Table 3. The smaller the width of confidence interval is, the more stable the SOH estimation would be. It can be seen that the stability property

TABLE 4. Six battery cells with different SOH.

Battery	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6
SOH (%)	102.5	99.5	99.0	96.99	92.3	87.17
Peak value (Ah/V)	29.97	29.87	29.786	29.426	29.233	27.813

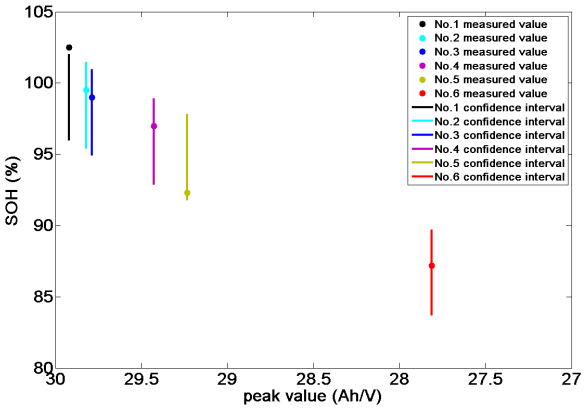

FIGURE 12. The SOH estimation results of six battery cells with different aging levels.

of SEISO is better than that of M3ISO and NN due to smaller width of the confidence interval. Considering the estimation error and confidence interval, the SEISO was selected as the kernel function. Comparing the peak value with the peak position, it is clear that the peak value as the HF yields better performance. This is also validated by the grey correlation analysis in Section III.

Additionally, six battery cells with different aging levels were used to further evaluate the proposed SOH estimaiton model. Their actual capacities were measured to calculate the true SOH values. The measrued capacities and the corresponding peak values of IC curves are listed in Table 4. It is worth noted that No. 1 cell is a fresh cell with a measured capacity larger than its rated one, which renders a SOH metering larger than 100%. The proposed MIGA-GPR model was then used to estimate the SOH based on the partial charging data with the peak value as the HF. As shown in Fig. 12, it can be seen that the MIGA-GPR model exhibits a good accuracy of SOH prediction with a maximum error of 3.5%.

V. CONCLUSION

This paper presents a battery SOH estimation method based on the Gaussian process regression (GPR) and the multi-island genetic algorithm (MIGA). Firstly, the incremental capacity analysis (ICA) method is used to extract the characteristic parameters of constant-current charging process as health factors (HFs) for SOH estimation, together with the wavelet transform for filtering. Then ,the Gaussian process regression is leveraged to approximate the complex relationship between SOH and HFs (peak value and peak position), and the multi-island genetic algorithm is used to globally optimize the hyper parameters. The established GPR model has the property of presenting confidence interval since it can estimate SOH with the mean and variance values as the uncertainty representations. The effects of different HFs and kernel functions are also analyzed, and the results show that the SEISO as kernel function with the peak value as the HF would yield the best estimation performance. Finally, the effectiveness of the proposed SOH estimation scheme is verfied through the accelerated battery life test.

REFERENCES

[1] L. Zhang, X. Hu, Z. Wang, F. Sun, and D. G. Dorrell, "A review of supercapacitor modeling, estimation, and applications: A control/management perspective," *Renew. Sustain. Energy Rev.*, to be published. [Online]. Available: http://dx.doi.org/10.1016/j.rser.2017.05.283

[2] Z. Wang, J. Hong, P. Liu, and L. Zhang, "Voltage fault diagnosis and prognosis of battery systems based on entropy and *Z*-score for electric vehicles," *Appl. Energy*, vol. 196, pp. 289–302, Jun. 2017.

[3] L. Zheng, L. Zhang, J. Zhu, G. Wang, and J. Jiang, "Co-estimation of state-of-charge, capacity and resistance for lithium-ion batteries based on a high-fidelity electrochemical model," *Appl. Energy*, vol. 180, pp. 424–434, Oct. 2016.

[4] C. Zhao, H. Yin, and C. Ma, "Quantitative efficiency and temperature analysis of battery-ultracapacitor hybrid energy storage systems," *IEEE Trans. Sustain. Energy*, vol. 7, no. 4, pp. 1791–1802, Oct. 2016.

[5] C. Zou, C. Manzie, and D. Nešić, "A framework for simplification of PDE-based lithium-ion battery models," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 5, pp. 1594–1609, Sep. 2016.

[6] Z. Wang, J. Ma, and L. Zhang, "Finite element thermal model and simulation for a cylindrical Li-ion battery," *IEEE Access*, vol. 5, pp. 15372–15379, 2017.

[7] X. Hu, S. E. Li, Z. Jia, and B. Egardt, "Enhanced sample entropy-based health management of Li-ion battery for electrified vehicles," *Energy*, vol. 64, pp. 953–960, Jan. 2014.

[8] X. Hu, J. Jiang, D. Cao, and B. Egardt, "Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2645–2656, Apr. 2016.

[9] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, "A review on lithium-ion battery ageing mechanisms and estimations for automotive applications," *J. Power Sour.*, vol. 241, pp. 680–689, Nov. 2013.

[10] S. M. Rezvanizaniani, Z. Liu, Y. Chen, and J. Lee, "Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility," *J. Power Sour.*, vol. 256, pp. 110–124, Jun. 2014.

[11] L. Lu, X. Han, J. Li, J. Hua, and M. Ouyang, "A review on the key issues for lithium-ion battery management in electric vehicles," *J. Power Sour.*, vol. 226, pp. 272–288, Mar. 2013.

[12] M. Doyle, T. F. Fuller, and J. Newman, "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell," *J. Electrochem. Soc.*, vol. 140, no. 6, pp. 1526–1533, Jun. 1993.

[13] V. R. Subramanian, V. Boovaragavan, V. Ramadesigan, and M. Arabandi, "Mathematical model reformulation for lithium-ion battery simulations: Galvanostatic boundary conditions," *J. Electrochem. Soc.*, vol. 156, no. 4, pp. A260–A271, Jan. 2009.

[14] V. Klass, M. R. Behm, and G. R. Lindbergh, "A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation," *J. Power Sour.*, vol. 270, no. 3, pp. 262–272, Dec. 2014.

[15] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods," *J. Power Sour.*, vol. 239, pp. 680–688, Oct. 2013.

[16] M. Berecibar *et al.*, "Online state of health estimation on NMC cells based on predictive analytics," *J. Power Sour.*, vol. 320, pp. 239–250, Jul. 2016.

[17] S. S. Y. Ng, Y. Xing, and K. L. Tsui, "A naive Bayes model for robust remaining useful life prediction of lithium-ion battery," *Appl. Energy*, vol. 118, pp. 114–123, Apr. 2014.

[18] Z. He, M. Gao, G. Ma, Y. Liu, and S. Chen, "Online state-of-health estimation of lithium-ion batteries using Dynamic Bayesian Networks," *J. Power Sour.*, vol. 267, no. 3, pp. 576–583, Dec. 2014.
[19] D. Liu, Y. Luo, J. Liu, Y. Peng, L. Guo, and M. Pecht, "Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm," *Neural Comput. Appl.*, vol. 25, nos. 3–4, pp. 557–572, Sep. 2014.
[20] B. Long, W. Xian, L. Jiang, and Z. Liu, "An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries," *Microelectron. Rel.*, vol. 53, no. 6, pp. 821–831, Jun. 2013.
[21] Y. Song, D. Liu, C. Yang, and Y. Peng, "Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery," *Microelectron. Rel.*, vol. 75, pp. 142–153, Aug. 2017.
[22] D. Liu, J. Pang, J. Zhou, Y. Peng, and M. Pecht, "Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression," *Microelectron. Rel.*, vol. 53, no. 6, pp. 832–839, Jun. 2013.
[23] Y. Zhang, D. Liu, J. Yu, Y. Peng, and X. Peng, "EMA remaining useful life prediction with weighted bagging GPR algorithm," *Microelectron. Rel.*, vol. 75, pp. 253–263, Apr. 2017.
[24] X. Han, M. Ouyang, L. Lu, J. Li, Y. Zheng, and Z. Li, "A comparative study of commercial lithium ion battery cycle life in electrical vehicle: Aging mechanism identification," *J. Power Sour.*, vol. 251, pp. 38–54, Apr. 2014.
[25] Z. Ma, J. Jiang, W. Shi, W. Zhang, and C. C. Mi, "Investigation of path dependence in commercial lithium-ion cells for pure electric bus applications: Aging mechanism identification," *J. Power Sour.*, vol. 274, no. 3, pp. 29–40, Jan. 2015.
[26] I. Bloom *et al.*, "Differential voltage analyses of high-power, lithium-ion cells: 1. Technique and application," *J. Power Sour.*, vol. 139, nos. 1–2, pp. 295–303, Jan. 2005.
[27] I. Bloom, J. Christophersen, and K. Gering, "Differential voltage analyses of high-power lithium-ion cells: 2. Applications," *J. Power Sour.*, vol. 139, nos. 1–2, pp. 304–313, Jan. 2005.
[28] M. Dubarry *et al.*, "Evaluation of commercial lithium-ion cells based on composite positive electrode for plug-in hybrid electric vehicle applications. Part II. Degradation mechanism under 2 C cycle aging," *J. Power Sour.*, vol. 196, no. 23, pp. 10328–10335, Dec. 2011.
[29] M. Dubarry *et al.*, "Identifying battery aging mechanisms in large format Li ion cells," *J. Power Sour.*, vol. 196, no. 7, pp. 3420–3425, Apr. 2011.
[30] C. Weng, Y. Cui, J. Sun, and H. Peng, "On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression," *J. Power Sour.*, vol. 235, no. 4, pp. 36–44, Aug. 2013.
[31] X. Feng, J. Li, M. Ouyang, L. Lu, J. Li, and X. He, "Using probability density function to evaluate the state of health of lithium-ion batteries," *J. Power Sour.*, vol. 232, no. 18, pp. 209–218, Jun. 2013.
[32] B. Sun *et al.*, "Practical state of health estimation of power batteries based on Delphi method and grey relational grade analysis," *J. Power Sour.*, vol. 282, pp. 146–157, May 2015.
[33] L. Zhang and D. G. Dorrell, "Genetic algorithm based optimal component sizing for an electric vehicle," in *Proc. 39th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2013, pp. 7331–7336.
[34] X. Han *et al.*, "An efficient genetic algorithm for optimization problems with time-consuming fitness evaluation," *Int. J. Comput. Methods*, vol. 12, pp. 1350106–1350130, Feb. 2015.

ZHENPO WANG received the Ph.D. degree in automotive engineering from the Beijing Institute of Technology, Beijing, China, in 2005.

He is currently a Professor with the Beijing Institute of Technology, and an Associate Director with the Collaborative Innovation Center for Electric Vehicles in Beijing and the National Engineering Laboratory for Electric Vehicles. His current research interests include pure electric vehicle integration, packaging and energy management of battery system and charging station design. He has published 4 monographs and translated books and over 60 technical papers. He also holds over 10 patents.

Dr. Wang was the recipient of numerous awards including the second National Prize for Progress in Science and Technology and the first Prize for Progress in Science and Technology from the Ministry of Education, China and the second Prize for Progress in Science and Technology from Beijing Municipal, China.

JUN MA received the B.S. degree with science applied physics from the Beijing Institute of Technology, China in 2011. He is currently pursuing the Ph.D. degree in mechanical engineering with the National Engineering Laboratory for Electric Vehicles from the Beijing Institute of Technology, China. His main research interests are in the areas of the battery system management and diagnosis.

LEI ZHANG (S'12–M'16) received the Ph.D. degree in mechanical engineering from the Beijing Institute of Technology, Beijing, China, in 2016. He is currently an Assistance Professor with the School of Mechanical Engineering, Beijing Institute of Technology.

His research interest includes ultracapacitor and battery modeling and state estimation, and energy management development for hybrid energy storage system for electric vehicle application.