



A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction

Peiyao Guo*, Ze Cheng, Lei Yang

School of Electrical and Information Engineering, Tianjin University, Tianjin, 300072, China



HIGHLIGHTS

- Health features extracted from charging voltage, current and temperature curves.
- Feature optimization based on grey relational and principal component analysis.
- Remaining capacity estimation with relevance vector machine.
- Validations with battery data in various operating conditions.

ARTICLE INFO

Keywords:

Lithium-ion battery
Health factor
Capacity estimation
Relevance vector machine

ABSTRACT

Capacity degradation monitoring of lithium batteries is necessary to ensure the reliability and safety of electric vehicles. However, capacity of cell is related to its complex internal physicochemical reactions and thermal effects and cannot be measured directly. A data-driven remaining capacity estimation approach for lithium-ion batteries based on charging health feature extraction is presented in this work. The proposed method utilizes rational analysis and principal component analysis to extract and optimize health features of charging stage which adapt to various working conditions of battery. The remaining capacity estimation is realized by relevance vector machine and validations of different working conditions are made with six battery data sets provided by NASA Prognostics Center of Excellence. The results show high efficiency and robustness of the proposed method.

1. Introduction

With growing energy and environment crisis worldwide, electric vehicles (EVs) technologies have received much attention and developed rapidly. Lithium-ion batteries have been widely used in EVs due to its merits of high operating voltage, high energy density, low self-discharge rate and no memory effect. Battery management system (BMS) plays a critical role in ensuring a long mileage, safety and reliability of EVs [1,2]. The accurate battery state estimation, as a key part of BMS, is desired to provide information for safety management and charging/discharging optimal control. Remaining useful life (RUL) is a measure of the change of the ability to store and release electrical energy of a battery compared with a fresh new one, essentially reflecting the aging and damage conditions of the battery [3,4]. It is necessary to replace batteries before the failure of battery make the whole system crash. Generally, the end of life (EoL) of battery reach when its actual capacity has decreased to 70% or 80% of its nominal value.

Capacity degradation of lithium-ion battery is closely related to its

internal physicochemical reaction and thermal effects. Over the repeated charging/discharging cycles, side reactions occur between electrode and electrolyte continuously yielding the growth of solid electrolyte interface (SEI) with poor conductivity and the loss of cyclable lithium ion [5,6]. In addition, cycling cause morphological damage of electrodes (i.e., porosity decrease and particle crack) and active electrode material loss. Extreme operating conditions such as overcharging, overdischarging, high voltage, both low and high temperatures, would accelerate battery aging process [7,8].

Over recent years, extensive research on RUL and SOH estimation has been conducted. In general, the methodologies adopted can be divided into data-driven and model-based (e.g., equivalent circuit model (ECM), electrochemical model, etc.) methods. ECMS made up of various circuit elements neglect the complex internal physicochemical aging mechanisms and mimic the output dynamics of the battery. Allafi et al. [9] established a modified Wiener battery model, consisting of a linear ECM and a new static sigmoid block and the parameters of the model were identified at different temperatures and SOC. Fleischer et al.

* Corresponding author.

E-mail address: guopeiyao@tju.edu.cn (P. Guo).

[10,11] proposed a weighted recursive least quadratic squares estimator for online parameter identification of an ECM composed of RC-circuit, Warburg element and charge-transfer resistance for capacity estimation of the battery. Waag et al. [12] proposed a new approach to estimate electromotive force (EMF) instead of OCV of the ECM saving rest time. Methods based on ECMs have relatively low computational cost, however are only applicable to limited operating conditions in which parameters have been identified. In addition, the attempt of extending the application range of ECMs and reducing the work of parameterizing both would improve the model complexity [5,12–15]. The main drawback of ECMs is little physical and chemical meanings of parameters causing inaccuracy at various operating situations. Alternatively, the electrochemical models describe the thermodynamics, reaction kinetics and ion diffusion inside the cells and therefore have potential of more accurate state estimation and a broader range of application. The so-called pseudo two-dimensional (P2D) model is based on the first principles in porous electrode theory [16] and relies on partial differential equations (PDEs) to describe the internal dynamics of lithium-ion batteries for which solutions are quite computationally expensive. Moreover, the uncertainty of P2D parameter can limit the actual accuracy of state estimation based on P2D model. There have been a few attempts to use simplified P2D models for RUL and SOH estimation, in which the computational efficiency has been improved at cost of accuracy [17–20].

Data-driven methods can reflect the inherent relationships of cells without expert knowledge on aging mechanisms. The data-driven methods have been widely used for RUL estimation including Neural Networks (NNs) [21–23], Particle Filter (PF) [24,25], Autoregressive model [26,27], Bayesian Networks [28,29], etc. NNs have strong nonlinear approximation ability, however, need a large amount of data for training and suffer from local optimality. In contrast, SVM also can handle nonlinear systems particularly with small set of training sample and find one global solution. Nevertheless as a decision machine, SVM only provide single-point estimation. The relevance vector machine (RVM) based on a Bayesian formulation shares many characteristics of the SVM whilst avoiding its principal limitations. RVM offers probabilistic outputs as well as much sparser solutions than SVM. Moreover, it has faster performance on test data whilst maintaining comparable generalization error [30].

It should be noted that, apart from improving estimation accuracy and efficiency of algorithms, some of the recent research is focused on health feature extraction. Widodo et al. [31] used Sample entropy (Sampen) of discharge voltage for battery health monitoring. Sampen of surface temperature in charging process served as an indicator for remaining capacity estimation in Ref. [32]. Liu et al. [33] used the discharging voltage difference of equal time interval to model battery degradation. Patil et al. [34] chose energy and fluctuation index of discharging voltage curve as HFs for RUL estimation. Ref. [35] extracted four geometrical features: length of constant voltage (CV) stage of the charging current curve, maximum radius of curvature of the CV curve, the area under the CV curve and the maximum slope of the early stage of a discharging voltage curve. Ref. [36] utilized five features extracted from current, voltage and temperature curve separately during charging and discharging process. HF extraction is indeed of great significance in characterizing the underlying degradation process and is more comprehensible and easier to be captured compared to electrochemical mechanisms [32,33]. The literature related to HF extraction above has greatly contributed to accurately modeling the degradation of cells. However, some issues of these methods still should be noticed: (1) Being unsuitable for various operating conditions of batteries as in Refs. [31,33–36]. The most of HFs extracted from discharging voltage curve in the references above limited by their own characteristics, are more applicable and only tested in constant current (CC) discharge conditions. (2) Neglecting thermal factors which are not only indicators of thermal stability of battery materials but also have important influence on the physical and chemical reactions inside the

battery as in Refs. [31,33–35]. (3) Lack of reasonable and clear extraction procedures. The HFs in Refs. [31–33,35] are given directly and the HF extraction and selection process in Ref. [34] is not very clear.

To overcome the aforementioned issues, a novel approach for HF extraction based on the charging voltage, current and temperature curves is proposed. Grey relational analysis is used to numerically analyze the relevance of HFs and capacity. Moreover, Principal component analysis (PCA) is applied to eliminate the redundant information of original HFs extracted and reduce computational complexity. The goal is to achieve a simple and reliable framework for lithium-ion battery remaining capacity estimation. An adaptive RVM based on particle swarm optimization (PSO) is utilized to ensure capacity prediction and robustness.

This paper is organized as follows. The related algorithms including RVM algorithm and principal component analysis are introduced in Section 2. Section 3 details the HF extraction method based on relevance analysis. Section 4 describes the overall procedure of RUL estimation for lithium-ion battery. The experimental results are presented in Section 5. Section 6 shows the conclusions.

2. Related algorithms

2.1. Adaptive relevance vector machine

2.1.1. Relevance vector regression

For battery capacity estimation, a set of training data $\{\mathbf{x}_i, t_i\}_{i=1}^D$, $\mathbf{x}_i \in R^m$, $t_i \in R$ is given, where D is the sample size, \mathbf{x}_i is an input HF vector and t_i is the targeted output capacity value. The relationship between input and targeted vectors can be expressed as follows:

$$t_i = y(\mathbf{x}) + \varepsilon_i \quad (1)$$

where ε_i is the independent noise following $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. The nonlinear function $y(\cdot)$ generally can be described in the form of:

$$y(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^D w_i K(\mathbf{x}, \mathbf{x}_i) + w_0 \quad (2)$$

where $K(\mathbf{x}, \mathbf{x}_i)$ is a kernel function and $w_i = (w_1, w_2, \dots, w_D)^T$ is the corresponding weight vector, w_0 is the bias parameter. Thus, the conditional distribution of a real capacity output is given by the form:

$$p(t_i | \mathbf{x}_i, \mathbf{w}, \sigma^2) = \mathcal{N}(t_i | y(\mathbf{x}), \sigma^2) \quad (3)$$

Assuming the t_i is independent, the likelihood of the whole given data set can be expressed as:

$$p(\mathbf{t} | \mathbf{x}_i, \mathbf{w}, \sigma^2) = \prod_{i=1}^D p(t_i | \mathbf{x}_i, \mathbf{w}, \sigma^2) \quad (4)$$

Because the parameter number is the same as the size of the training data set, the maximum likelihood estimation of \mathbf{w} and σ^2 would cause over-fitting. To avoid this problem, the likelihood function is complemented by introducing a Gaussian prior over \mathbf{w} :

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{i=1}^D \mathcal{N}(w_i | 0, \alpha_i^{-1}) \quad (5)$$

where α_i is the hyperparameter for corresponding weight w_i , $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_D)$.

Next, the posterior distribution over the weights can be obtained as follows:

$$\begin{aligned} p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) &= \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2)p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)} \\ &= (2\pi)^{-\frac{N+1}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ \frac{(\mathbf{w} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{w} - \boldsymbol{\mu})}{2} \right\} \end{aligned} \quad (6)$$

where the mean and covariance are given by:

$$\boldsymbol{\mu} = \sigma^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{t} \quad (7)$$

$$\boldsymbol{\Sigma} = \mathbf{t}(\mathbf{A} + \sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \quad (8)$$

where $\boldsymbol{\Phi}$ is a $D \times (D+1)$ design matrix with $\boldsymbol{\Phi} = [\boldsymbol{\Phi}(x_1), \boldsymbol{\Phi}(x_1), \dots, \boldsymbol{\Phi}(x_D)]^T$ and $\boldsymbol{\Phi}(x_i) = [1, K(x_i, x_1), \dots, K(x_i, x_D)]$, and $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_D)$. To maximize the posterior distribution of the hyperparameters, the marginal likelihood function should be maximized which is obtained by integrating the weights as follows:

$$\begin{aligned} p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2) &= \int p(\mathbf{t}|\mathbf{w}, \sigma^2) p(\mathbf{w}|\boldsymbol{\alpha}) d\mathbf{w} \\ &= (2\pi)^{-\frac{N}{2}} |\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \mathbf{t}^T (\sigma^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T)^{-1} \mathbf{t}\right\} \end{aligned} \quad (9)$$

The re-estimation formulas to update $\boldsymbol{\alpha}$ and σ^2 can be given by setting the derivatives of the log marginal likelihood:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2} \quad (10)$$

$$(\sigma^2)^{new} = \frac{\|\mathbf{t} - \boldsymbol{\Phi} \boldsymbol{\mu}\|^2}{N - \sum_{i=0}^D \gamma_i} \quad (11)$$

where $\gamma_i = 1 - \alpha_i \Sigma_{ii}$, Σ_{ii} is the i th component of $\boldsymbol{\Sigma}$ and μ_i is the i th component of the $\boldsymbol{\mu}$. α_{MP} and σ_{MP}^2 respectively denote the most probable values of $\boldsymbol{\alpha}$ and σ^2 , which can be obtained at the end of iteration.

Given new data \mathbf{x}_* , we can get targeted output \mathbf{t}_* following $p(\mathbf{t}_*|\mathbf{t}) \sim \mathcal{N}(\boldsymbol{\mu}_*, \sigma_*^2)$, for which

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}^T \boldsymbol{\Phi}(\mathbf{x}_*) \quad (12)$$

$$\sigma_*^2 = \sigma_{MP}^2 + \boldsymbol{\Phi}(\mathbf{x}_*)^T \boldsymbol{\Sigma} \boldsymbol{\Phi}(\mathbf{x}_*) \quad (13)$$

2.1.2. Adaptive kernel based on PSO

In RVM, the natural of kernel function is to map the nonlinear data to the high-dimensional space and make it linearly separable in high-dimensional space. Therefore, it is necessary to select a appropriate kernel function and optimize its parameter. The kernel function in Eq. (2) is chosen as radial basic kernel(RBF) function as Eq. (14) which is written as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{r^2}\right) \quad (14)$$

In this work a optimized PSO algorithm [37] is utilized to search for optimal parameter of RBF kernel function in this paper. To avoid overfitting and improve the generalization of ARVM, the logarithm of Eq. (9) is adopted as the fitness value [36]:

$$\text{Maximize: } \log(p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2)) \quad (15)$$

2.2. PCA algorithm

Principal component analysis(PCA) is a dimensionality reduction technique widely used for feature extraction. PCA can transform a large set of data to a small set of data meanwhile obtaining most of the original information.

Consider a data set $\mathbf{X}(D \times m)$ consisting of D samples of m original variables, our goal is to convert the data into a space with lower dimensionality while maximizing the variance of the data. To begin with, standardize the original data and then compute the covariance matrix \mathbf{S} :

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^{*T} \mathbf{X}^* \quad (16)$$

where \mathbf{X}^* is the standardized data matrix. The eigenvectors and eigenvalues of \mathbf{X}^* denoted respectively as \mathbf{u}_i and λ_i , $i = 1, \dots, m$ are obtained using following mathematical formula:

$$\mathbf{S} \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (17)$$

Table 1
Batteries with their operating parameters.

Battery ID	Discharge current(A)	End voltage(V)	Ambient temperature	Rated capacity	End-of-life criteria
B0005	2	2.7	24°C	2Ahr	1.4Ahr
B0006		2.5			
B0007		2.2			
B0018		2.5			
B0025	0.05 Hz	2.0	24°C	–	–
B0026	square wave	2.2			
B0027	of 4A and	2.5			
B0028	50% duty	2.7			
B0029	4	2.0	43°C	–	–
B0030		2.2			
B0031		2.5			
B0032		2.7			
B0033	33,34-4A	2.0	24°C	2Ahr	1.6Ahr
B0034		2.2			
B0036	36-2A	2.7			
B0045	1	2.0	4°C	2Ahr	1.4Ahr
B0046		2.2			
B0047		2.5			
B0048		2.7			
B0053	2	2.0	4°C	2Ahr	1.4Ahr
B0054		2.2			
B0055		2.5			
B0056		2.7			

where \mathbf{S} is an $n \times n$ matrix, \mathbf{u}_i is an $n \times 1$ vector, and λ_i is a scalar. The principal component scores can be computed as:

$$\mathbf{Z} = \mathbf{X}^* \times \mathbf{S} \quad (18)$$

To determine the number of principal components, the variance contribution rate can be calculated as:

$$r_i = \frac{\lambda_i}{\sum_{i=1}^m \lambda_i} \times 100\% \quad (19)$$

Next, determine the number of principal components according to the predefined minimum accumulated variance contribution rate and select the factors which have the highest eigenvalue. The dimensionality decreased feature matrix can be obtained.

3. HF extraction and optimization

3.1. HF extraction

In this work, cycling data of lithium-ion batteries were sourced from Prognostics Center of Excellence(PCoE) at Ames Research Center, NASA. Table 1 shows the batteries, as well as the operating conditions and parameters used in this work. The obvious anomaly cycle data are eliminated like discharge capacity turning to zero when battery is far from the EoL and invalid data. As mentioned above, to adapt to different operating conditions, the HFs are extracted from charging profiles(current, voltage and temperature) which are shown in Fig. 1. The lithium batteries were charged at 1.5A until the battery voltage reached 4.2 V. Then a CV mode continued until the charge current fell to 20 mA. The charging process consists of two phases: CC phase and CV phase. It is observed that the changes of voltage, current and temperature charging profiles are regular with battery aging. Charge temperature increase and reach its peak during the CC phase and then drop during the CV phase. The measured voltage, current and temperature are acquired with non-uniform sampling rate [34]. And the large amount of data would also make it difficult to process and the precision of sensors is limited. Hence, it is critical to extract HFs can describe the change trend of these curves. In this work, 14 HFs are extracted from the charging process which can be divided into four groups. The four

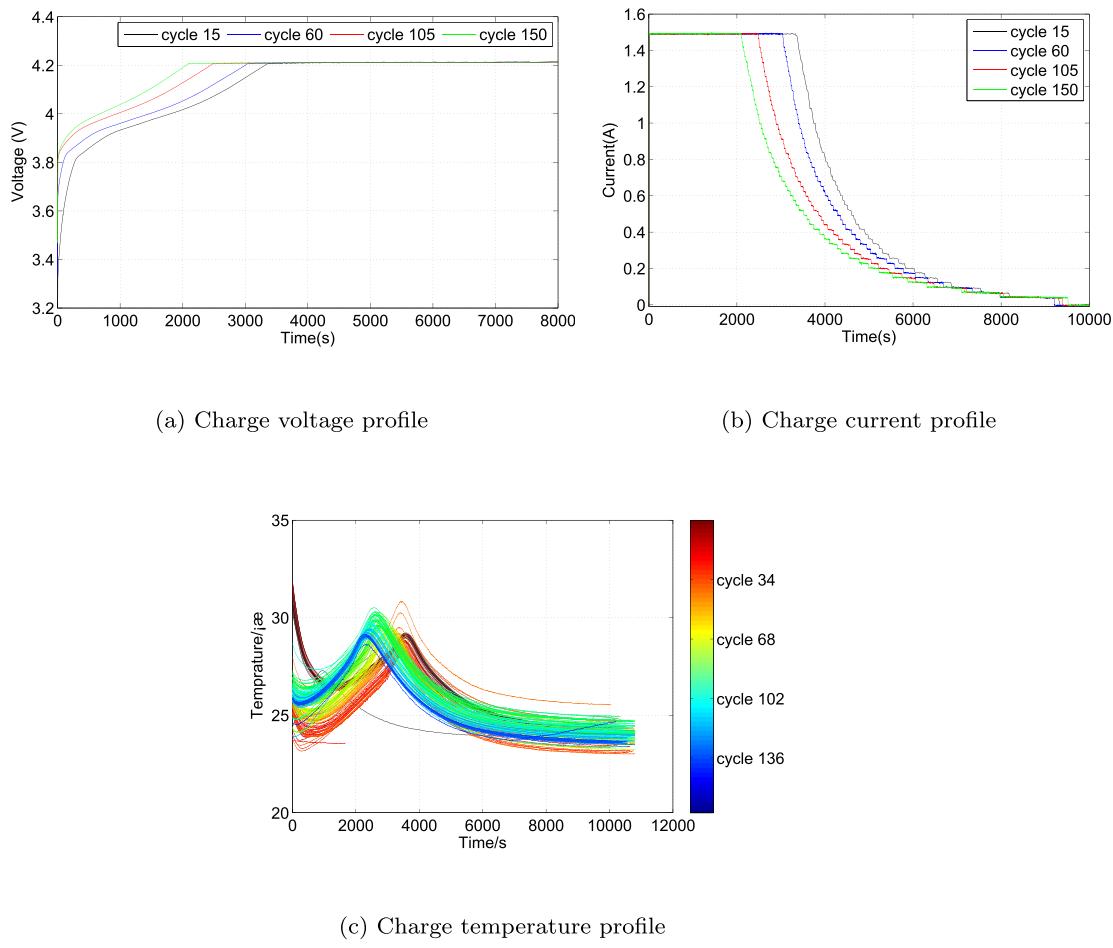


Fig. 1. Charge profiles of B0007.

groups were respectively related to capacity, charge time, temperature and current/voltage drop at CV and CC stage. Group1: A_1 , A_2 , A . A_1 is the area under the current curve of CC phase. A_2 and A denote the corresponding area in CV phase and the whole charge process. These three HFs can reflect the change trend of charge capacity. Group 2: L_1 , L_2 , L_1/L_2 . L_1 and L_2 indicate the CC time and CV time respectively. L_1/L_2 is the ratio of CC time to CV time. Group 3: T_1 , T_2 , T , T_1/A_1 , T_2/A_2 , T/A . T_1 , T_2 and T are the area under the temperature curve of CC phase, CV phase and the whole charge process. T_1/A_1 , T_2/A_2 and T/A are the ratio of area under temperature curve to the corresponding area under the current curve, which can indicate the temperature change per unit capacity. Group 4: K_1 , K_2 . K_1 is the maximum slope of voltage curve in CC phase. K_2 is the maximum slope of current curve in CV phase. Fig. 2 depicts variation of all these HFs for battery B0007.

3.2. HF optimization

Fig. 2 directly shows the qualitative relevance between HFs and capacity. To further verify and select highly related HFs, it is paramount to make quantitative analyze. The grey relational analysis can be applied to quantitatively analyze uncertain relations between reference factor and other factors in a given system. Consider capacity sequence is reference sequence $X_0 = \{x_0(k)\}$, and HF sequences serve as comparison sequences denoted as $X_i = \{x_i(k)\}$. The grey relation coefficient for the i th factor can be expressed as:

$$\xi_i(k) = \frac{\min_{\forall i} \max_{\forall k} |x_0(k) - x_i(k)| + \rho \max_{\forall i} \max_{\forall k} |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_{\forall i} \max_{\forall k} |x_0(k) - x_i(k)|} \quad (20)$$

where ρ is identification coefficient, $\rho \in [0,1]$. ρ is taken as 0.5 here. The grey relational grade is usually taken by averaging the value of the grey relational coefficients $\xi_i(k)$ as follows:

$$\gamma_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \quad (21)$$

The grey relational grade γ is used to measure the relevance degree between the reference and the comparability sequence. The larger γ_i is, the level of relevance is higher. If γ is equal to 1, the two sequences are identical. Thus, we can verify and select the HF_s closely related to the capacity degradation according to grey relational grades. Table 2 list the grey relational analysis results of 14 HF_s in 3.1. The HF_s in Table 2 are numbered from K_1 in the left to A_2 in the right starting with 1. The relevance degree sort result is shown in Table 3.

It is easy to find that despite the grey relational grades vary with different cells in different experimental conditions, some HFs obviously still have higher relevance. To guarantee the robustness and accuracy of HFs to different batteries and conditions, eight HFs, i.e. A_1 , A , L_1 , L_1/L_2 , T , T_1 , T_2/A_2 , K_2 are chosen to estimate capacity. And to eliminate the information redundancy which may exist among these HFs, PCA is applied to reduce dimensionality of data which can also be seen as a de-noising process.

4. Proposed methodology for capacity estimation

This section illustrates the proposed methodology for capacity estimation for lithium battery. Fig. 3 shows the whole procedure of capacity estimation which can be divided into two parts: offline model

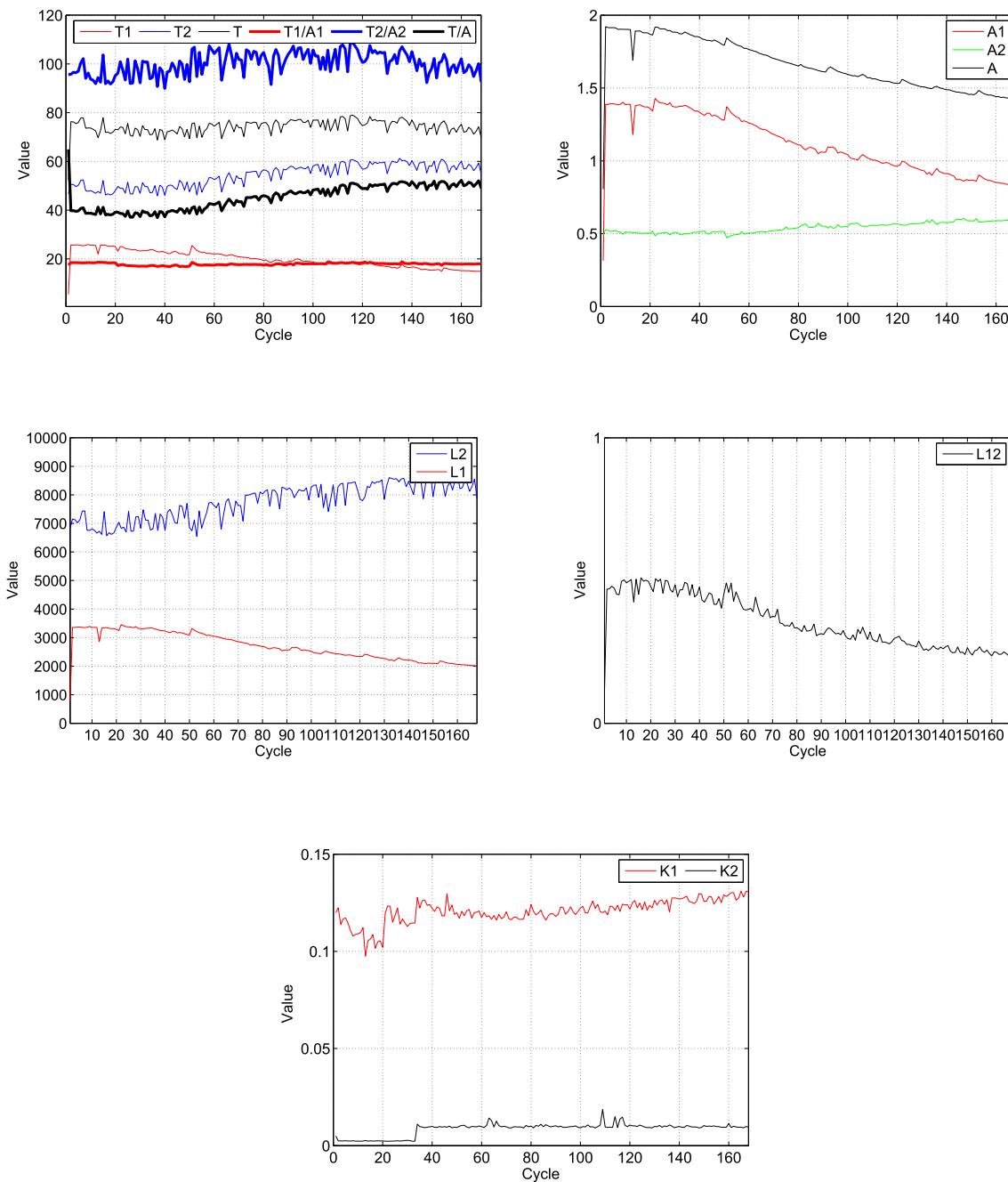


Fig. 2. 14 HFs extracted from the charge profiles of B0007.

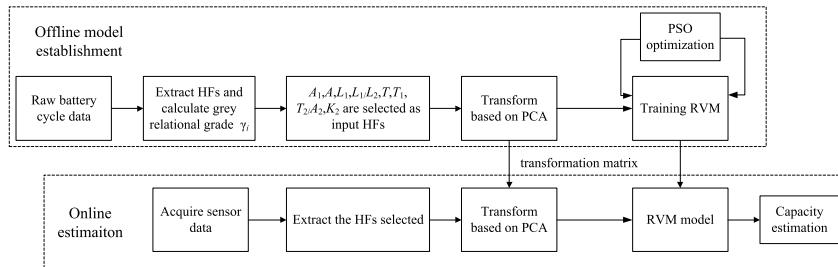
Table 2
Grey relational grades of HFs.

	K_1	K_2	L_1	L_2	L_1/L_2	T	T_1	T_2	T/A	T_1/A_1	T_2/A_2	A	A_1	A_2
B0005	0.5454	0.5930	0.7705	0.5120	0.8698	0.5987	0.7742	0.5025	0.5494	0.5729	0.6740	0.7130	0.7698	0.5112
B0006	0.5277	0.6626	0.9403	0.5163	0.8632	0.5931	0.9434	0.5061	0.5362	0.6386	0.7386	0.7783	0.9395	0.4982
B0025	0.5864	0.6214	0.7056	0.5809	0.7402	0.6001	0.7159	0.5795	0.6060	0.5671	0.5751	0.6353	0.7051	0.5805
B0026	0.4452	0.5261	0.4995	0.6173	0.5285	0.6190	0.5000	0.5988	0.5200	0.5807	0.5834	0.5671	0.4985	0.5452
B0029	0.6797	0.8073	0.9269	0.6915	0.7698	0.7858	0.7989	0.6938	0.5776	0.5803	0.7095	0.9251	0.9270	0.6076
B0030	0.7156	0.8217	0.9132	0.7158	0.7605	0.8323	0.7798	0.7177	0.5797	0.5652	0.6399	0.9329	0.9137	0.7519
B0045	0.5917	0.5838	0.8654	0.5478	0.8675	0.7368	0.8197	0.7294	0.6702	0.7475	0.6712	0.7101	0.8651	0.6938
B0046	0.7038	0.7203	0.8193	0.5366	0.8206	0.6174	0.7248	0.5626	0.6640	0.5638	0.6616	0.5898	0.8192	0.5784
B0053	0.7006	0.7483	0.6253	0.6975	0.6291	0.5429	0.5466	0.6734	0.5689	0.5429	0.5812	0.7977	0.6252	0.7038
B0054	0.7200	0.8058	0.6238	0.4764	0.7463	0.5256	0.6253	0.4829	0.5731	0.6011	0.5575	0.7297	0.6237	0.7670

Table 3

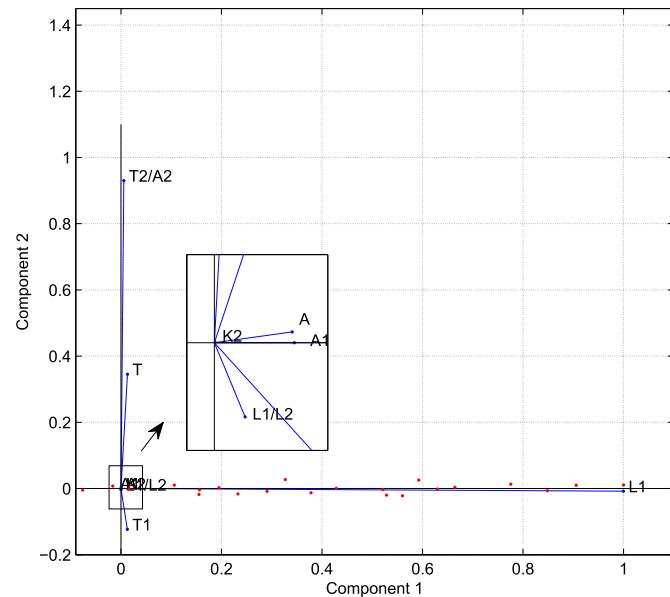
Relevance degree sort result of HFs.

	Relevance:	High															Low
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Low
B0005	5	7	3	13	12	11	6	2	10	9	1	4	14	8			
B0006	7	3	13	5	12	11	2	10	16	9	1	4	18	14			
B0025	5	7	3	13	12	2	9	6	1	4	14	8	11	10			
B0026	6	4	8	11	10	12	14	5	2	9	7	3	13	1			
B0029	13	3	12	2	7	6	5	11	8	4	1	14	10	9			
B0030	12	13	3	6	2	7	4	14	8	4	1	11	9	10			
B0045	5	3	13	7	10	6	8	12	14	1	9	1	2	4			
B0046	5	3	13	7	2	1	9	11	6	12	13	1	8	4			
B0053	12	2	14	1	4	8	5	3	13	1	9	7	10	6			
B0054	2	14	5	12	1	7	3	13	10	9	11	6	8	4			

**Fig. 3.** Flow chart for capacity estimation.**Table 4**

Cumulative contribution rate of components.

	Cumulative contribution rate(%)					
	Component1	Component2	Component3	Component4	Component5	Component6
B0005	9.999418e-01	5.057113e-05	7.233026e-06	4.325682e-07	1.591656e-10	1.777094e-11
B0006	9.999765e-01	1.696692e-05	6.255165e-06	3.176149e-07	1.703607e-10	1.358012e-11
B0025	9.940654e-01	4.529538e-03	1.365394e-03	3.970701e-05	1.745994e-09	2.576266e-10
B0026	9.945076e-01	4.219488e-03	1.248549e-03	2.439599e-05	1.318236e-09	3.983660e-10
B0029	9.995899e-01	3.831405e-04	2.239424e-05	4.544619e-06	1.951158e-11	7.097724e-12
B0030	9.994873e-01	4.432214e-04	5.442527e-05	1.502983e-05	2.532585e-11	1.900708e-11
B0045	9.956755e-01	4.174163e-03	1.498810e-04	3.606494e-07	4.873250e-08	1.768220e-10
B0046	9.999582e-01	3.812034e-05	3.365123e-06	2.959782e-07	1.777405e-09	1.125733e-11
B0053	9.999628e-01	3.583579e-05	1.097943e-06	3.062481e-07	1.507342e-10	1.118272e-11
B0054	9.999176e-01	6.568645e-05	1.531322e-05	1.440661e-06	3.445600e-09	4.156193e-11

**Fig. 4.** Visualized result of PCA.

establishment and online estimation. In offline stage, the cycling data of cells are acquired and preprocessed to serve as training data set for RVM. A series of HFs are extracted from charging voltage current and temperature curves. Then, through grey relational analysis, the relevance between the HFs and capacity is evaluated and the highly related HF, i.e. A_1 , A , L_1 , L_1/L_2 , T , T_1 , T_2/A_2 , K_2 are selected in this work. The HF selected are further optimized with PCA of which results are utilized to train the RVM regression model. Finally, the RVM model is built. In online stage, the data are preprocessed in the identical way with offline stage and serve as the model input. The well-trained model provides accurate capacity estimation.

5. Experimental results and discussion

5.1. Principal component analysis

To illustrate the performance of HF optimization based on PCA, **Table 4** lists the cumulative contribution rate of components produced by PCA of five sets of batteries. It can be seen that the processed data are all reduced to 6-dimension and the contribution rate of Component1 of all the batteries already can reach above 99%. Thus, in this paper, the component1 is used in capacity estimation. **Fig. 4** gives the visualized result of PCA. makes the greatest contribution to component1, then the T , T_1 , A_1 , A , L_1/L_2 , T_2/A_2 also make little contributions to component1.

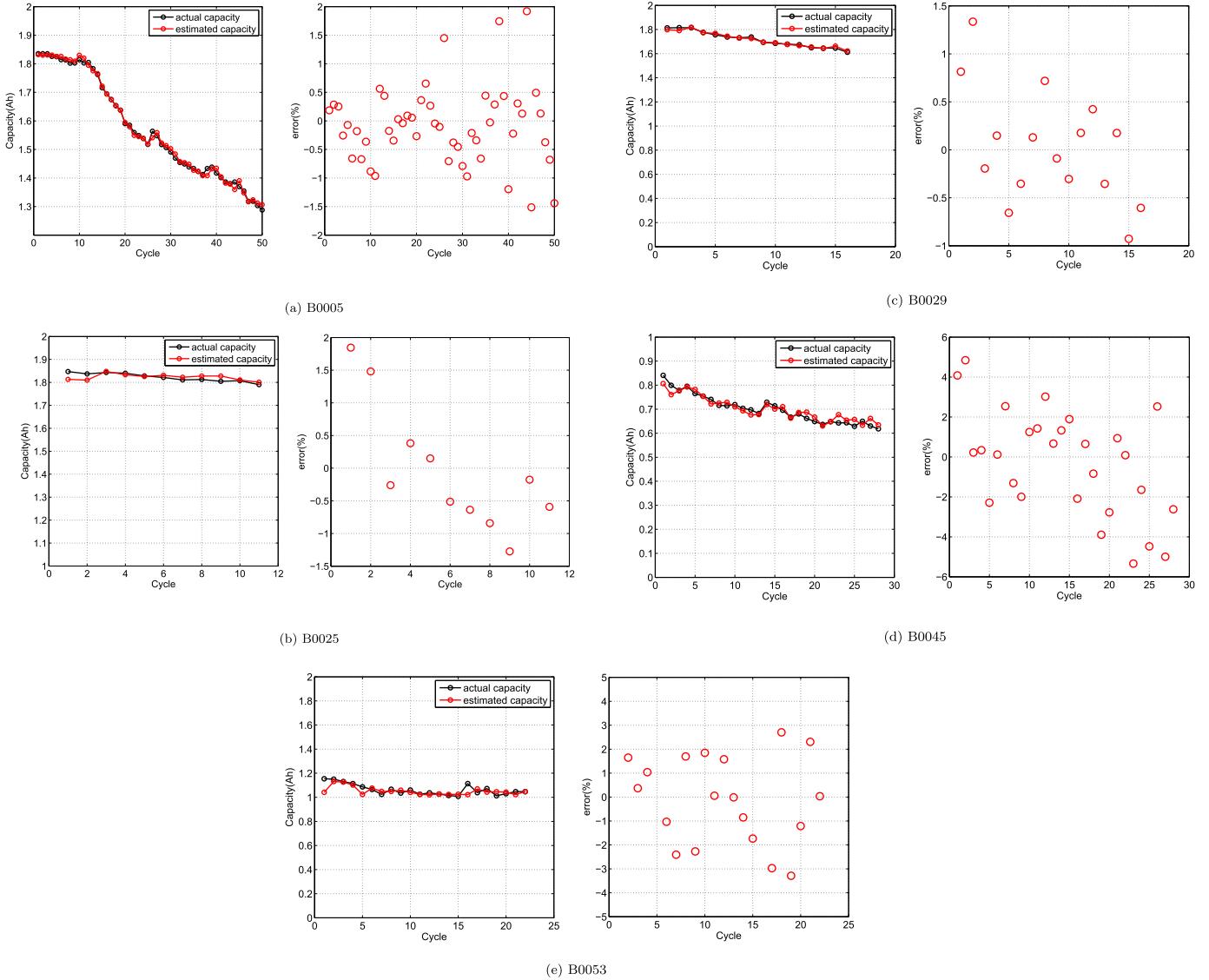


Fig. 5. Single battery experiments.

Table 5
Estimation errors of single battery experiments.

	B0005	B0025	B0029	B0045	B0053
R_2	0.999958	0.999916	0.999966	0.999317	0.998693
ξ_{RMSE}	0.010222	0.016749	0.010046	0.018376	0.038024

The contribution of K_2 can be neglected.

5.2. Results of estimation

In this section, the six battery data sets listed in Table 1 are implemented to verify the accuracy and robustness of proposed methodology. Here, the root mean squared error (RMSE) and the Coefficient of determination R_2 are utilized to evaluating the performance of proposed ARVM model:

$$\xi_{RMSE} = \sqrt{\frac{1}{N} \left(\sum_{i=1}^N (t_i - \hat{t})^2 \right)} \quad (22)$$

$$R_2 = 1 - \frac{\sum_{i=1}^N (t_i - \hat{t})^2}{\sum_{i=1}^N (t_i - \bar{t})^2} \quad (23)$$

The first two and fourth battery data sets (from B0005 to B0028 and from B0033 to B0036) are utilized to investigate the adaptability of the proposed model to different discharge current and the third and last two data sets (from B0029 to B0032 and from B0045 to B0056) can test the adaptability to different temperatures.

5.2.1. Single battery test

The first battery of five battery data sets (B0005, B0025, B0029, B0045, B0053) are chosen to verify the accuracy of the proposed method for capacity estimation. The cycle data of B0005 are randomly divided to 70% training set and 30% testing set. The cycle data of B0025, B0029, B0045 and B0053 are randomly divided to 60% training set and 40% testing set. Fig. 5 provides the comparison of estimated capacity with data of the testing set and the error of estimation by regression model with testing data. Table 5 gives the R_2 and RMSE. The relative error of three battery at 24°C and 43°C are all less than 2% and the relative error of B0045 and B0053 at 4°C is little higher but still within the range of 5%.

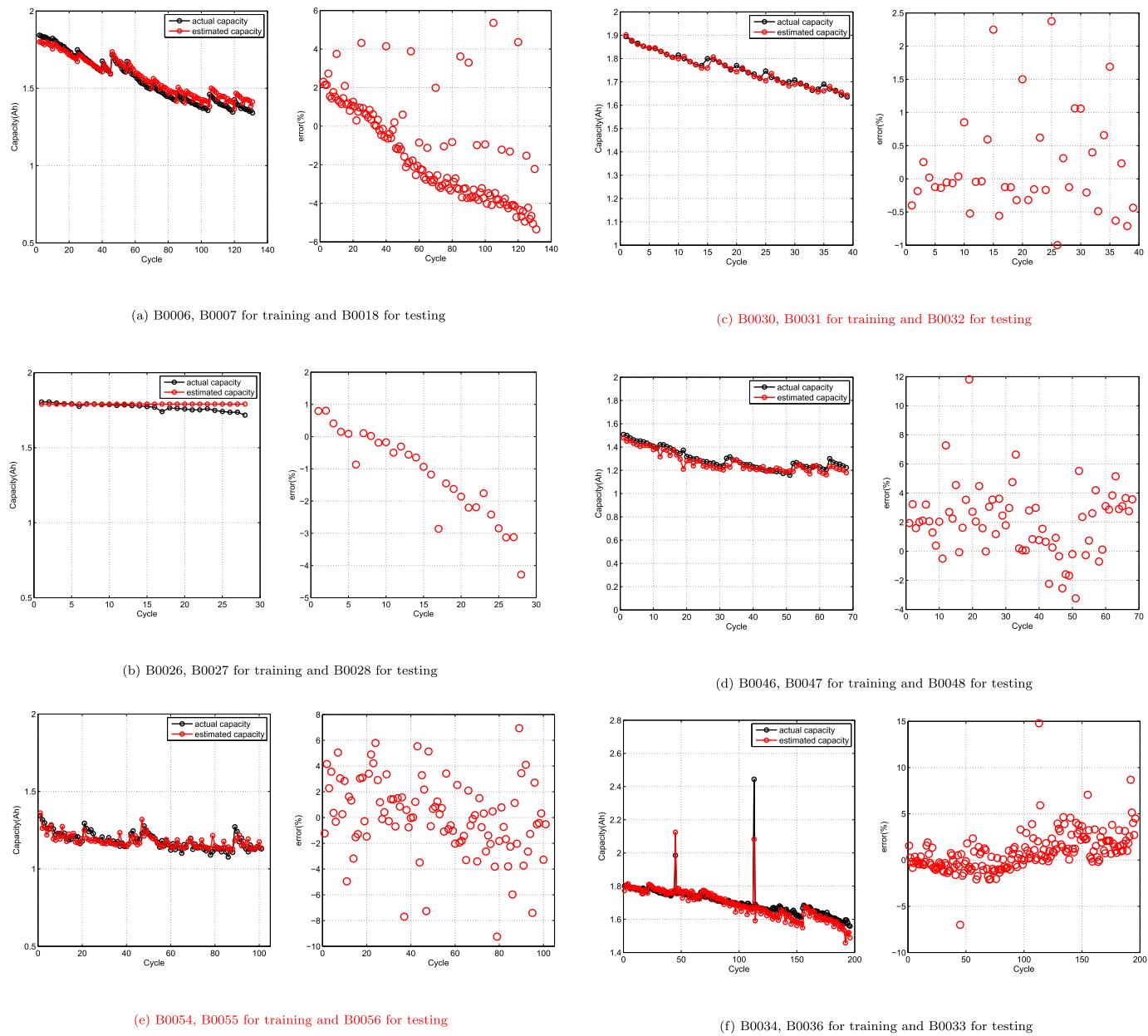
**Fig. 6.** Multiple battery experiments.

Table 6
Estimation errors of multiple battery experiments.

	B0018	B0028	B0032	B0048	B0056	B0033
R^2	0.999320	0.999710	0.999939	0.998944	0.999112	0.999376
ξ_{RMSE}	0.041133	0.030476	0.013651	0.041288	0.035195	0.042244

5.2.2. Multiple battery data

The experiments in this section use the middle two battery data of the data sets in 5.2.1 as the training data and the last battery data as the testing data. Apart from this, B0034 with 4A constant discharge current and B0036 with 2A are used as training set and B0033 with 4A is tested, which verified the performance of model in hybrid operating condition. Fig. 6 gives the estimated results and Table 6 lists the estimation errors of multiple battery data experiments. Although the error of multiple battery experiments is higher than the single battery experiments generally, the relative errors are still within the range of 10% despite several isolated points and the capacity degradation trend is well traced

by the proposed method. The experiment of B0033, B0034 and B0036 shows the robustness to training data set of model.

6. Conclusions

This work presents a data-driven capacity estimation approach for lithium-ion batteries based on charging HF extraction and optimization. The charge HFs with high relevance are picked out from 14 HFs extracted from charge current, voltage and temperature profiles through grey rational analysis and PCA is performed for data dimensionality reduction. With the advantages of not being influenced by the battery discharge conditions and considering the temperature factor, the processed HFs are used as input of RVM model. Validations are conducted based on the NASA battery data obtained in different operating conditions and the test results demonstrate the HFs extracted well reflect battery internal aging process and proposed method have good accuracy and robustness on capacity estimation. It can be concluded this approach have potential for applications in EVs. However, there is still

considerable space for further improvements. Due to lack of battery data, the validations of FUDS, DST conditions are not made. The precision of RVM model still should be improved in the future.

Acknowledgments

This study is supported by the National Natural Science Foundation of China, China (Grant No. 61873180).

References

- [1] H. Rahimieichi, U. Ojha, F. Baronti, M. Chow, Battery Management System: an Overview of its Application in the Smart Grid and Electric Vehicles, *Electrets*, 2013, pp. 441–445, <https://doi.org/10.1109/MIE.2013.2250351>.
- [2] K.W.E. Cheng, B.P. Divakar, H. Wu, K. Ding, H.F. Ho, Battery-management system (BMS) and soc development for electrical vehicles, *IEEE Trans. Veh. Technol.* 60 (1) (2011) 76–88, <https://doi.org/10.1109/TVT.2010.2089647>.
- [3] X.S. Si, W. Wangbde, D.H. Zhouc, Remaining useful life estimation - a review on the statistical data driven approaches, *Eur. J. Oper. Res.* 213 (1) (2011) 1–14, <https://doi.org/10.1016/j.ejor.2010.11.018>.
- [4] B. Saha, K. Goebel, J. Christoffersen, Comparison of prognostic algorithms for estimating remaining useful life of batteries, *Trans. Inst. Meas. Contr.* 31 (3–4) (2009) 293–308, <https://doi.org/10.1177/0142331208092030>.
- [5] A. Farmann, W. Waag, A. Marongiu, D.U. Sauer, Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles, *J. Power Sources* 281 (2015) 114–130, <https://doi.org/10.1016/j.jpowsour.2015.01.129>.
- [6] L. Yang, X. Cheng, Y. Ma, Changing of SEI film and electrochemical properties about MCMB electrodes during long-term charge/discharge cycles, *J. Electrochem. Soc.* 160 (11) (2013) A2093–A2099, <https://doi.org/10.1149/2.064311jes>.
- [7] M. Klett, T.G. Zavalis, M.H. Kjell, R.W. Lindström, M. Behm, G. Lindbergh, Altered electrode degradation with temperature in LiFePO₄/mesocarbon microbead graphite cells diagnosed with impedance spectroscopy, *Electrochim. Acta* 141 (28) (2014) 173–181, <https://doi.org/10.1016/j.electacta.2014.06.081>.
- [8] J. Vazquez-Arenas, M. Fowler, X. Mao, S.K. Chen, Modeling of combined capacity fade with thermal effects for a cycled Li_xC₆-Li_yMn₂O₄ cell, *J. Power Sources* 215 (4) (2012) 28–35, <https://doi.org/10.1016/j.jpowsour.2012.04.092>.
- [9] W. Allafi, K. Uddin, C. Zhang, R.M.R.A. Sha, J. Marco, On-line scheme for parameter estimation of nonlinear lithium ion battery equivalent circuit models using the simplified refined instrumental variable method for a modified wiener continuous-time model, *Appl. Energy* 204 (2017) 497–508, <https://doi.org/10.1016/j.apenergy.2017.07.030>.
- [10] C. Fleischer, W. Waag, H.-M. Heyn, D.U. Sauer, On-line adaptive battery impedance parameter and state estimation considering physical principles in reduced order equivalent circuit battery models: Part 1. requirements, critical review of methods and modeling, *J. Power Sources* 260 (2014) 276–291, <https://doi.org/10.1016/j.apenergy.2017.07.030>.
- [11] C. Fleischer, W. Waag, H.-M. Heyn, D.U. Sauer, On-line adaptive battery impedance parameter and state estimation considering physical principles in reduced order equivalent circuit battery models part 2. parameter and state estimation, *J. Power Sources* 262 (2014) 457–482, <https://doi.org/10.1016/j.jpowsour.2014.03.046>.
- [12] W. Waag, D.U. Sauer, Adaptive estimation of the electromotive force of the lithium-ion battery after current interruption for an accurate state-of-charge and capacity determination, *Appl. Energy* 111 (4) (2013) 416–427, <https://doi.org/10.1016/j.apenergy.2013.05.001>.
- [13] Y. Hu, S. Yurkovich, Y. Guezennec, B.J. Yurkovich, Electro-thermal battery model identification for automotive applications, *J. Power Sources* 196 (1) (2011) 449–457, <https://doi.org/10.1016/j.jpowsour.2010.06.037>.
- [14] X. Lin, H.E. Perez, S. Mohan, J.B. Siegel, A.G. Stefanopoulou, Y. Ding, M.P. Castanier, A lumped-parameter electro-thermal model for cylindrical batteries, *J. Power Sources* 257 (257) (2014) 1–11, <https://doi.org/10.1016/j.jpowsour.2014.01.097>.
- [15] N. Damay, C. Forgez, M.P. Bichat, G. Friedrich, Thermal modeling of large prismatic LiFePO₄/graphite battery, coupled thermal and heat generation models for characterization and simulation, *J. Power Sources* 283 (72) (2015) 37–45, <https://doi.org/10.1016/j.jpowsour.2015.02.091>.
- [16] M. Doyle, T.F. Fuller, J. Newman, Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell, *J. Electrochem. Soc.* 140 (6) (1993) 1526–1533, <https://doi.org/10.1149/2.1181811jes>.
- [17] S.J. Moura, N.A. Chaturvedi, M. Krstic, PDE estimation techniques for advanced battery management systems-part 1: SOC estimation, *American Control Conference*, 2012, pp. 559–565, <https://doi.org/10.1109/ACC.2012.6315019>.
- [18] S.J. Moura, M. Krstic, N.A. Chaturvedi, Adaptive PDE observer for battery SOC/ SOH estimation, *ASME 2012 5th Annual Dynamic Systems and Control Conference Joint with the JSME 2012 11th Motion and Vibration Conference*, 2012, pp. 101–110, <https://doi.org/10.1115/DSCC2012-MOVIC2012-8800>.
- [19] A.P. Schmidt, M. Bitzer, A.W. Imre, L. Guzzella, Model-based distinction and quantification of capacity loss and rate capability fade in li-ion batteries, *J. Power Sources* 195 (22) (2010) 7634–7638, <https://doi.org/10.1016/j.jpowsour.2010.06.011>.
- [20] A. Bartlett, J. Marcicki, S. Onori, G. Rizzoni, X.G. Yang, T. Miller, Electrochemical model-based state of charge and capacity estimation for a composite electrode lithium-ion battery, *IEEE Trans. Contr. Syst. Technol.* 24 (2) (2016) 384–399, <https://doi.org/10.1109/TCST.2015.2446947>.
- [21] M. Rezvani, S. Lee, J. Lee, A Comparative Analysis of Techniques for Electric Vehicle Battery Prognostics and Health Management (Phm), *Tech. rep SAE Technical Paper*, 2011, <https://doi.org/10.1109/TCST.2015.2446947>.
- [22] D. Andre, A. Nuhic, T. Soczka-Guth, D.U. Sauer, Comparative study of a structured neural network and an extended kalman filter for state of health determination of lithium-ion batteries in hybrid electric vehicles, *Eng. Appl. Artif. Intell.* 26 (3) (2013) 951–961, <https://doi.org/10.1016/j.engappai.2012.09.013>.
- [23] H. Chaoui, State of charge and state of health estimation for lithium batteries using recurrent neural networks, *IEEE Trans. Veh. Technol.* 99 (99) (2017), <https://doi.org/10.1109/TVT.2017.2715333> 1–1.
- [24] A. Guha, A. Patra, State of health estimation of lithium-ion batteries using capacity fade and internal resistance growth models, *IEEE Transactions on Transportation Electrification* (99) (2017), <https://doi.org/10.1109/TTE.2017.2776558> 1–1.
- [25] Y. Song, D. Liu, C. Yang, Y. Peng, Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery, *Microelectron. Reliab.* 75 (2017) 142–153, <https://doi.org/10.1109/TTE.2017.2776558>.
- [26] D. Liu, Y. Luo, Y. Peng, Lithium-ion Battery Remaining Useful Life Estimation Based on Nonlinear Ar Model Combined with Degradation Featuredoi:10.1007/s00521-013-1520-x.
- [27] B. Long, W. Xian, L. Jiang, Z. Liu, An improved autoregressive model by particle swarm optimization for prognostics of lithium-ion batteries, *Microelectron. Reliab.* 53 (6) (2013) 821–831, <https://doi.org/10.1016/j.microrel.2013.01.006>.
- [28] S.S.Y. Ng, Y. Xing, K.L. Tsui, A naive bayes model for robust remaining useful life prediction of lithium-ion battery, *Appl. Energy* 118 (4) (2014) 114–123, <https://doi.org/10.1016/j.apenergy.2013.12.020>.
- [29] Z. He, M. Gao, G. Ma, Y. Liu, S. Chen, Online state-of-health estimation of lithium-ion batteries using dynamic bayesian networks, *J. Power Sources* 267 (3) (2014) 576–583, <https://doi.org/10.1016/j.jpowsour.2014.05.100>.
- [30] N. Abramson, D. Braverman, G. Sebestyen, *Pattern Recognition and Machine Learning*, Academic Press. doi:10.1016/j.jpowsour.2014.05.100.
- [31] A. Widodo, M.C. Shim, W. Caesarendra, B.S. Yang, Intelligent prognostics for battery health monitoring based on sample entropy, *Expert Systems with Applications An International Journal* 38 (9) (2011) 11763–11769, <https://doi.org/10.1016/j.eswa.2011.03.063>.
- [32] J. Li, L. Chao, L. Wang, L. Zhang, C. Li, Remaining capacity estimation of li-ion batteries based on temperature sample entropy and particle filter, *J. Power Sources* 268 (268) (2014) 895–903, <https://doi.org/10.1016/j.jpowsour.2014.06.133>.
- [33] D. Liu, J. Zhou, H. Liao, Y. Peng, X. Peng, A health indicator extraction and optimization framework for lithium-ion battery degradation modeling and prognostics, *IEEE Transactions on Systems Man & Cybernetics Systems* 45 (6) (2015) 915–928, <https://doi.org/10.1109/TSMC.2015.2389757>.
- [34] M.A. Patil, P. Tagade, K.S. Hariharan, S.M. Kolake, T. Song, T. Yeo, S. Doo, A novel multistage support vector machine based approach for li ion battery remaining useful life estimation, *Appl. Energy* 159 (2015) 285–297, <https://doi.org/10.1016/j.apenergy.2015.08.119>.
- [35] C. Lu, L. Tao, H. Fan, Li-ion battery capacity estimation: a geometrical approach, *J. Power Sources* 261 (261) (2014) 141–147, <https://doi.org/10.1016/j.jpowsour.2014.03.058>.
- [36] Y. Zhang, B. Guo, Online capacity estimation of lithium-ion batteries based on novel feature extraction and adaptive multi-kernel relevance vector machine, *Energies* 8 (11) (2015) 12439–12457, <https://doi.org/10.3390/en81112320>.
- [37] K. Tang, B. Liu, J. Yang, A. Sun, Double center particle swarm optimization algorithm, *J. Comput. Res. Dev.* 41 (2012) 951–957.