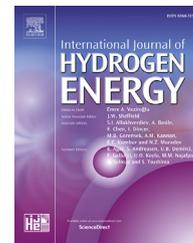




ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

journal homepage: [www.elsevier.com/locate/ijhe](http://www.elsevier.com/locate/ijhe)

# A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell

Yujie Cheng <sup>a,b,c</sup>, Nouredine Zerhouni <sup>b</sup>, Chen Lu <sup>c,d,\*</sup>

<sup>a</sup> School of Aeronautic Science and Engineering, Beihang University, Beijing, 100191, China

<sup>b</sup> FEMTO-ST Institute, UMR CNRS 6174 - UBFC/UFC/ENSMM/UTBM, FC-LAB Research, FR CNRS 3539, 24 rue Alain Savary, 25000, Besançon, France

<sup>c</sup> Science & Technology on Reliability and Environmental Engineering Laboratory, China

<sup>d</sup> School of Reliability and Systems Engineering, Beihang University, China

## ARTICLE INFO

### Article history:

Received 7 January 2018

Received in revised form

17 April 2018

Accepted 19 April 2018

Available online xxx

### Keywords:

Proton exchange membrane fuel cell

Prognostic

Residual useful life

Least squares support vector machine

Regularized particle filter

Hybrid method

## ABSTRACT

Proton Exchange Membrane Fuel Cell (PEMFC) has become a promising power source with wide applications to many electronic and electrical devices. However, even if it is a competitive energy converter, PEMFC still suffers from its limited lifespan. Prognostics appear to be a good solution to helping take actions to extend its lifetime. Considering both advantage and disadvantage of model-based and data-driven based prognostic methods, this study proposes a hybrid prognostic method for PEMFC based on a data-driven method, least square support vector machine (LSSVM) and a model-based method, regularized particle filter (RPF). The main contributions of the proposed method include: 1) It can provide not only an estimated value but also an uncertainty characterization of RUL with a probability distribution; 2) It has a better capability to capture the nonlinearities in degradation data and a lower reliance on PEMFC degradation model; 3) The RPF method improves the standard particle filter algorithm by reducing the degeneration phenomenon and loss of diversity among the particles. Effectiveness of the proposed method is verified based on PEMFC dataset provided by FCLAB Research Federation. The results indicate that the proposed hybrid method can effectively combine both advantages of data-driven and model-based methods, providing a higher accuracy of RUL prediction for PEMFC.

© 2018 Hydrogen Energy Publications LLC. Published by Elsevier Ltd. All rights reserved.

## Introduction

Proton Exchange Membrane Fuel Cell (PEMFC) has been considered as one of the most promising power source which can be widely applied in military, transportation, and combined heat and power systems by virtue of its high power density, environmental friendliness, light weight, and abundant resources [1–3]. However, its limited lifespan, long-term

performances, and maintenance costs is a big obstacle for the deployment and commercialization of PEMFC [4,5]. Prognostics and health management (PHM), and particularly prognostics, has attracted increasing attention in recent years, which offers a good solution to extend the lifetime of PEMFC. PHM for PEMFC aims at utilizing real monitoring data to predict the health degradation of PEMFC and estimate the residual useful life (RUL) of PEMFC, so as to help making good decisions to take adequate actions at the right time. In this

\* Corresponding author. School of Reliability and Systems Engineering, Beihang University, China

E-mail address: [luchen@buaa.edu.cn](mailto:luchen@buaa.edu.cn) (C. Lu).

<https://doi.org/10.1016/j.ijhydene.2018.04.160>

0360-3199/© 2018 Hydrogen Energy Publications LLC. Published by Elsevier Ltd. All rights reserved.

### Nomenclature

$x_i$	The $i$ th input data
$y_i$	The $i$ th output data
$d$	The dimension of $x_i$
$\varphi(\cdot)$	Nonlinear function
$w$	Weight vector
$b$	Bias term
$C$	Adjusting factor
$\xi$	Error variable
$L(w, b, \xi, \alpha)$	Lagrange function
$\alpha_i$	Lagrange multipliers
$K$	Kernel matrix
$K_{ij}$	RBF kernel
$x_k$	System state
$z_k$	System observation
$f(x_{k-1}, v_{k-1})$	State transition function
$h(x_k, n_k)$	Measurement function
$v_{k-1}$	System noise
$n_k$	Measurement noise
$p(x_k z_{1:k-1})$	Approximation of the prior probability density function at the $k$ th cycle
$p(x_k x_{k-1})$	Transition probability distribution defined by the state model
$p(x_k z_{1:k})$	Posterior probability density function at the $k$ th step
$p(z_k x_k)$	Likelihood function of the measurement model
$p(z_k x_k^i)$	Likelihood function of the $i$ th particle at step $k$
$N_{eff}$	Effective sample
$N_{thres}$	Threshold of effective sample
$K(\cdot)$	Kernel density
$K_h$	Rescaled version of kernel density $K(\cdot)$
$h$	Kernel bandwidth
$n_x$	Dimension of the state vector $x$
$K_{opt}$	Optimal Kernel
$c_{n_x}$	Volume of the unit hypersphere in $R^{n_x}$
$h_{opt}$	Optimal bandwidth
$\sigma$	Standard deviation of $n_k$
$T_{EoL\_prognostic}$	Predicted EoL time
$T_{EoL\_true}$	Actual EoL time
$RUL\_prognostic$	Predicted PDF distribution of RUL
$RUL\_true$	Time distance between the true EoL time and the prognostic start time

way, the lifetime of PEMFC can be extended with more adequate usage mechanism. In the full process of PHM, prognostics is the core technology which plays an important role for the following health management [6].

Prognostic techniques can be conditionally classified into long-term prognostic techniques and short-term prognostic techniques. Ref. [7] presents a good overview of short-term prognostic techniques, including neural networks, autoregressive integrated moving average, genetic algorithm, etc. The authors of Ref. [7] also presented an algebraic prognostic approach with mixed smoothing (APMS) for short-term time series. However, in the field of PEMFC, a long-term prognostic is more expected since it can provide more important degradation information to PEMFC users from a long-term

perspective, so that adequate actions can be taken timely to prolong the lifetime of PEMFC. Moreover, APMS has a good performance for non-linear time series, but does not work well with non-stationary time series. Considering the variable working conditions and unknown external disturbance of PEMFC in real applications, the monitoring data of PEMFC are usually non-stationary. Thus, a prognostic approach which can successfully dealing with such uncertainties should be proposed. Most existing prognostic methods for PEMFC can be classified into two main categories: model-based methods, including extended Kalman filter [8], particle filter [6,9], and physical phenomena-based model which is based on the degradation law and electrical equivalent circuit [10]; and data-driven methods, including neural network [11], relevance vector machine [12], neuro-fuzzy inference systems [13] and some machine learning methods [14]. Model-based methods aim at establishing empirical or mechanism models to simulate the degradation process of PEMFC according to the complex degradation mechanism [2]. The advantage of this kind of method is that it doesn't require a large amount of data. It can also provide an accurate prognostic result given an accurate degradation model. However, the construction of accurate degradation model of PEMFC is usually difficult in real applications, since the complex degradation mechanism of PEMFC is not fully understood yet [2]. Data-driven based methods, on the other hand, don't need priori knowledge to establish an accurate degradation model, since they aim at mining the degradation law of PEMFC by learning the available degradation data using some intelligent computation methods. This kind of method doesn't need to fully understand the degradation mechanism of PEMFC and have a good capability to catch the nonlinearities contained in the monitoring signal [6]. However, the main drawback is that the performance of data-driven based method strongly relies on the amount and quality of data in the training process. Considering both advantage and disadvantage of model-based methods and data-driven based methods, this study presents a hybrid prognostic method for PEMFC in order to combine both advantages of model-based methods and data-driven based methods, thus improving the prognostic accuracy of PEMFC.

Particle filter, one of the typical model-based prognostic method, is based on Bayesian technique, which employs a set of weighted particles to form a posterior distribution of the system. Compared with Kalman filter, particle filter shows excellent performance in dealing with nonlinear systems with non-Gaussian noise [15]. Thus, particle filter-based prognostics have become a hot issue in recent years. However, in the particle filter method, particle degeneracy phenomenon and loss of diversity among the particles affects the prognostic accuracy heavily, which restricts its real application. In this study, a modified particle filter method, regularized particle filter, is proposed to solve the above mentioned problem. On the other hand, least square support vector machine (LSSVM) is a data-driven method, which is much easier and computationally simpler than standard support vector machine. Compared with neural networks, LSSVM has a better generalization performance [16]. Nowadays, LSSVM has been widely applied in the field of wind power prediction [16], electronic equipment [17], bearing degradation prediction [18],

and so on. However, studies on its application in the field of PEMFC is very few. Additionally, even though LSSVM can reach a good prognostic result for PEMFC, it cannot quantify the uncertainty of the prognostic result. However, a prognostic result with uncertainty characterization is of great importance in decision-making process. Aiming at obtaining an accurate prognostic result with uncertain characterization, this study combines both advantages of LSSVM and RPF. First, LSSVM is employed to realize a preliminary prognostic for PEMFC. Then, the predicted results of LSSVM are introduced into the prognostic framework of RPF as new observation values to get the final RUL probability distribution of PEMFC.

The remainder of this paper is organized as follows. In Section [Related theories](#), related theories including basic principle and algorithms of LSSVM and RPF are introduced. In Section [Methodology](#), the whole prognostic framework based on LSSVM-RPF for PEMFC are detailed illustrated. The experiment results based on PEMFC dataset provided by FCLAB Research Federation are presented and discussed in Section [Results and Discussion](#). Finally, a conclusion is given in Section [Conclusion](#).

## Related theories

### Least square support vector machine

Support vector machine proposed by Cortes and Vapnik [19] is a powerful tool for both classification and regression. It is based on structural risk minimization principles rather than empirical risk minimization principle, which bring SVM a better generalization property than neural networks [20]. SVM can show high performance when dealing with limited training samples. However, when the amount of training data becomes larger, the complexity of optimization increases proportionally with the number of training points, which makes it time consuming and not suitable for real-time application [18]. In order to overcome the shortage of traditional SVM, Suykens et al. [21] proposed the LSSVM, which converts the optimization problem of SVM to solving linear equations, thus reducing the time consumption and improving the convergence accuracy [17]. The principle of LSSVM can be introduced as follows [20,22].

Consider a given training dataset  $\{\mathbf{x}_i, y_i\}, i = 1, 2, \dots, N$ , where  $\mathbf{x}_i \in \mathbb{R}^d$  is the  $i$ th input data,  $y_i \in \mathbb{R}$  is the corresponding  $i$ th output data, and  $d$  is the dimension of  $\mathbf{x}_i$ . Based on a nonlinear function  $\varphi(\cdot)$ , the regression model which maps the input data to a higher dimensional feature space can be constructed as follows:

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b, \mathbf{w} \in \mathbb{R}^d, b \in \mathbb{R}, \varphi \in \mathbb{R}^d \rightarrow \mathbb{R}^M, M \rightarrow \infty, \quad (1)$$

According to the structural risk minimization principle, the optimization problem in the primal space can be described as follows:

$$\min J(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} C \xi^T \xi \quad (2)$$

which is subjected to

$$y_i = \mathbf{w}^T \varphi(\mathbf{x}_i) + b + \xi_i, i = 1, 2, \dots, N \quad (3)$$

Aiming at solving the above optimization problem, the Lagrange function can be obtained by introducing the Lagrange multipliers  $\alpha_i$ :

$$L(\mathbf{w}, b, \xi, \alpha) = \frac{1}{2} (\mathbf{w}^T \mathbf{w}) + \frac{C}{2} \|\xi\|^2 - \sum_{i=1}^N \alpha_i (\mathbf{w}^T \varphi(\mathbf{x}_i) + b + \xi_i - y_i) \quad (4)$$

The optimal solution of equation (4) can be obtained by taking the partial derivatives of equation (4) with respect to  $\mathbf{w}$ ,  $b$ ,  $\xi$ , and  $\alpha$ , and equal them to zero.

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i \varphi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = C \xi_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \mathbf{w}^T \varphi(\mathbf{x}_i) + b + \xi_i - y_i = 0 \end{cases} \quad (5)$$

Then, equation (5) can be expressed using the following set of linear equations after eliminating  $\mathbf{w}$  and  $\xi_i$ :

$$\begin{bmatrix} \mathbf{K} + \frac{1}{C} \mathbf{I} & \mathbf{e} \\ \mathbf{e}^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{Y} \\ 0 \end{bmatrix} \quad (6)$$

where  $\mathbf{Y} = [y_1, y_2, \dots, y_M]^T$ ,  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]^T$ ,  $\mathbf{e} = [1, 1, \dots, 1]^T$ , and  $\mathbf{K}$  represents the kernel matrix which can be written as:

$$K_{ij} = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \quad (7)$$

Then, the function estimation of LSSVM can be given by

$$y(\mathbf{x}) = \sum_{i=1}^N \alpha_i \phi(\mathbf{x}_i, \mathbf{x}) + b \quad (8)$$

where  $\alpha_i$  and  $b$  can be calculated by equation (6).

The Radial basis function (RBF) kernel can be described as follows

$$K_{ij} = \exp(-\eta \|\mathbf{x}_i - \mathbf{x}_j\|_2^2) \quad (9)$$

For the recursive regression of LSSVM, given the input matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{k-m+1} \\ \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_m \\ \mathbf{x}_2 & \mathbf{x}_3 & \dots & \mathbf{x}_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{k-m+1} & \mathbf{x}_{k-m+2} & \dots & \mathbf{x}_k \end{bmatrix} \quad (10)$$

and output vector

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_{k-m+1} \end{bmatrix} = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_{k+1} \end{bmatrix} \quad (11)$$

then the regressive function is as follows

$$\hat{x}_s = \sum_{i=1}^N \alpha_i \phi(\mathbf{x}_i, \mathbf{x}_s) + b \quad (12)$$

The prediction result at  $t_{k+1}$  can be calculated by the following equation

$$\hat{\mathbf{x}}_{k+1} = \sum_{i=1} \alpha_i \phi(\mathbf{x}_i, \mathbf{x}_{k-m+1}) + \mathbf{b} \quad (13)$$

Given  $\mathbf{x}_{k-m+2} = [\mathbf{x}_{k-m+2} \ \mathbf{x}_{k-m+3} \ \cdots \ \mathbf{x}_k \ \hat{\mathbf{x}}_{k+1}]$ , then the prediction result at  $t_{k+2}$  can be obtained as follows

$$\hat{\mathbf{x}}_{k+2} = \sum_{i=1} \alpha_i \phi(\mathbf{x}_i, \mathbf{x}_{k-m+2}) + \mathbf{b} \quad (14)$$

Recursively, the prediction result at  $t_{k+j}$  is

$$\hat{\mathbf{x}}_{k+j} = \sum_{i=1} \alpha_i \phi(\mathbf{x}_i, \mathbf{x}_{k-m+j}) + \mathbf{b} \quad (15)$$

where  $\mathbf{x}_{k-m+j} = [\mathbf{x}_{k-m+j} \ \cdots \ \mathbf{x}_k \ \hat{\mathbf{x}}_{k+1} \ \cdots \ \hat{\mathbf{x}}_{k+j-1}]$ .

### Particle filter and regularized particle filter

Particle filter is a Monte Carlo-based computation method which is designed for recursively estimating the evolving posterior distribution of a system using a set of weighted particles. Particularly, PF is a promising approach to handle model nonlinearities with non-Gaussian noise. However, the main drawback of PF is the degeneracy phenomenon and loss of diversity among the particles. This section will first introduce the principle of PF method. Then, a modified particle filter method, regularized particle filter will be presented to improve the performance of standard PF.

#### Particle filter

The particle filter method is based on the following system state space model.

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \quad (16)$$

$$\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{n}_k) \quad (17)$$

Both  $\mathbf{v}_{k-1}$  and  $\mathbf{n}_k$  can be either Gaussian or non-Gaussian.

Fig. 1 shows the general process of particle filter with one parameter estimation, which mainly includes two procedures, i.e., Sequential Importance Sampling (SIS) and Resampling [15].

#### (1) Sequential Importance Sampling

Before receiving knowledge of the measurement  $\mathbf{z}_k$ , suppose the posterior probability density function at the  $k$ -1th step is  $p(\mathbf{x}_{k-1} | \mathbf{z}_{1:k-1})$  and  $N$  particles have been generated from the posterior distribution. Then, the approximation of the prior probability density function at the  $k$ th cycle,  $p(\mathbf{x}_k | \mathbf{z}_{1:k-1})$ , can be calculated based on the state model according to the Chapman–Kolmogorov equation:

$$p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1} \quad (18)$$

When the measurement  $\mathbf{z}_k$  is available, the posterior probability density function at the  $k$ th step,  $p(\mathbf{x}_k | \mathbf{z}_{1:k})$ , can be calculated based on the Bayesian algorithm and the Markov assumption:

$$p(\mathbf{x}_k | \mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k | \mathbf{x}_k, \mathbf{z}_{1:k-1}) p(\mathbf{x}_k | \mathbf{z}_{1:k-1})}{p(\mathbf{z}_k | \mathbf{z}_{1:k-1})} = \frac{p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{z}_{1:k-1})}{\int p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) d\mathbf{x}_k} \quad (19)$$

Suppose the measurement noise  $\mathbf{n}_k$  follows a Gaussian distribution,  $\mathbf{n}_k \sim N(0, \sigma)$ , then likelihood function of the  $i$ th particle at step  $k$ ,  $p(\mathbf{z}_k | \mathbf{x}_k^i)$ , can be calculated using the following equation.

$$p(\mathbf{z}_k | \mathbf{x}_k^i) = \frac{1}{\sqrt{2\pi}\sigma_k^i} \exp\left[-\frac{1}{2}\left(\frac{\mathbf{z}_k - \mathbf{x}_k^i}{\sigma_k^i}\right)^2\right] \quad (20)$$

The weight of the  $i$ th particle at step  $k$  can be obtained based on the likelihood function of the measurement  $\mathbf{z}_k$ , as formulated in equation (21).

$$w_k^i = \frac{p(\mathbf{z}_k | \mathbf{x}_k^i)}{\sum_j^N p(\mathbf{z}_k | \mathbf{x}_k^j)} \quad (21)$$

#### (2) Resampling

A problem existing in the SIS framework is the degeneracy phenomenon of the particles. It means that after a few iterations, only few particles will have negligible weight, while the weights of other particles tend to toward zero. The particle degeneracy can be measured using effective sample which can be calculated as follows:

$$N_{eff} = \frac{N}{1 + \text{var}(w_k^i)} \approx \frac{1}{\sum_{i=1}^N (w_k^i)^2} \quad (22)$$

To avoid the degeneracy phenomenon of particles, a resampling step is performed to remove particles with small weights and duplicate particles with large weights when  $N_{eff} < N_{thres}$ . After that, the weights of all particles are set to  $1/N$ .

#### Regularized particle filter

Even though resampling can reduce the degeneracy phenomenon of the particles, it should be noticed that another problem is introduced, that is, the loss of diversity of the particles. This new problem is caused due to the fact that new samples are drawn from a discrete distribution rather than a continuous one. This problem may lead to “particle collapse”, which means that all particles share the same weight, thus giving a poor representation of the posterior distribution [23]. Here, a modified version of particle filter, regularized particle filter is presented to solve the above problem. RPF method realizes resampling from a continuous distribution of the posterior density [23]:

$$p(\mathbf{x}_k | \mathbf{z}_{1:k}) \approx \sum_{i=1}^N w_k^i K_h(\mathbf{x}_k - \mathbf{x}_k^i) \quad (23)$$

$$K_h(x) = \frac{1}{h^{n_x}} K\left(\frac{x}{h}\right) \quad (24)$$

where the kernel density  $K(\cdot)$  is a symmetric probability density function which satisfies:

$$\int xK(x)dx = 0, \int \|x\|^2 K(x)dx < \infty \quad (25)$$

The Kernel and bandwidth are chosen to minimize the mean integrated square error (MISE) between the true posterior density and the corresponding regularized empirical representation. In a special case when all the particles have

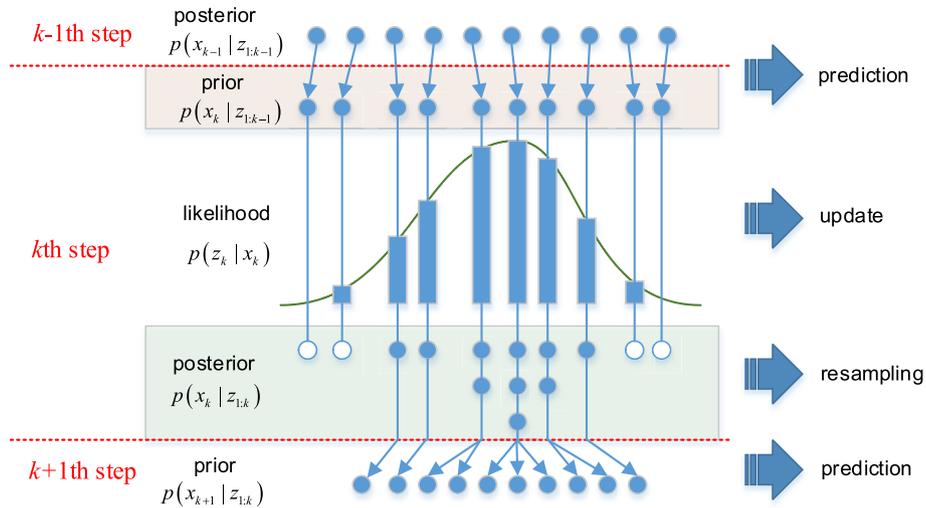


Fig. 1 – General process of particle filter.

the same weight, the optimal Kernel is the Epanechnikov kernel:

$$K_{opt} = \begin{cases} \frac{n_x + 2}{2c_{n_x}} (1 - \|x\|^2) & \text{if } \|x\| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

The optimal bandwidth can be calculated by equation (27) when the underlying density is Gaussian with a unit covariance matrix.

$$h_{opt} = A \cdot N^{1/(n_x+4)} \quad (27)$$

where

$$A = \left[ 8c_{n_x}^{-1} (n_x + 4) (2\sqrt{\pi})^{n_x} \right]^{1/(n_x+4)} \quad (28)$$

### Methodology

In this section, the framework of the proposed prognostic method for PEMFC based on LSSVM-RPF is illustrated. The details of each step are elaborated to give the readers a better understanding.

#### Prognostic framework based on LSSVM-RPF

In order to combine both advantages of data-driven method and model-based method, a prognostic framework based on LSSVM-RPF for PEMFC is proposed in this study, as illustrated in Fig. 2. In the proposed framework, voltage drop is considered as an aging indicator. The original voltage measurements of PEMFC are first preprocessed. Then, the preprocessed

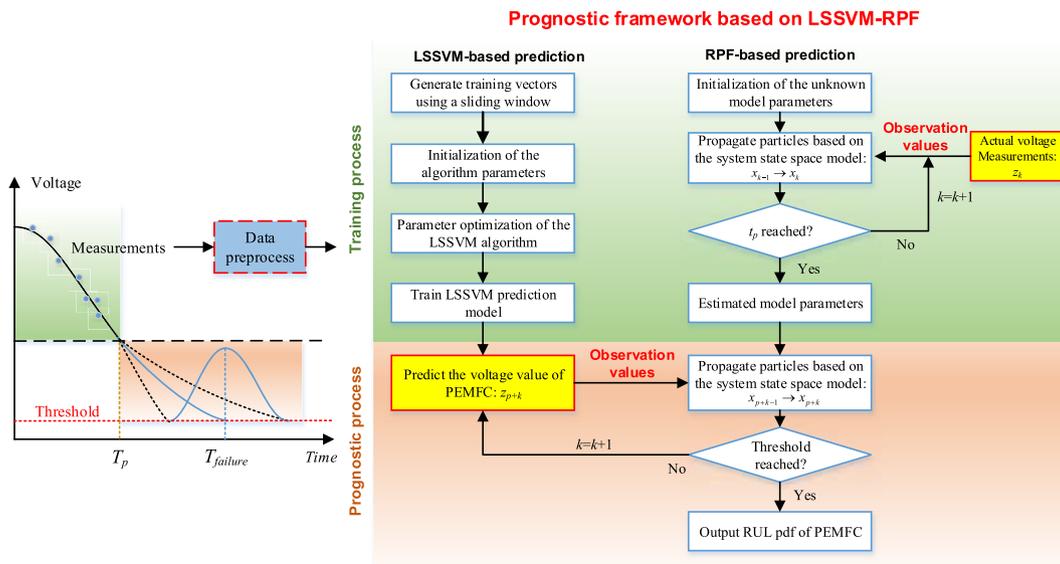


Fig. 2 – Prognostic framework for PEMFC based on LSSVM-RPF.

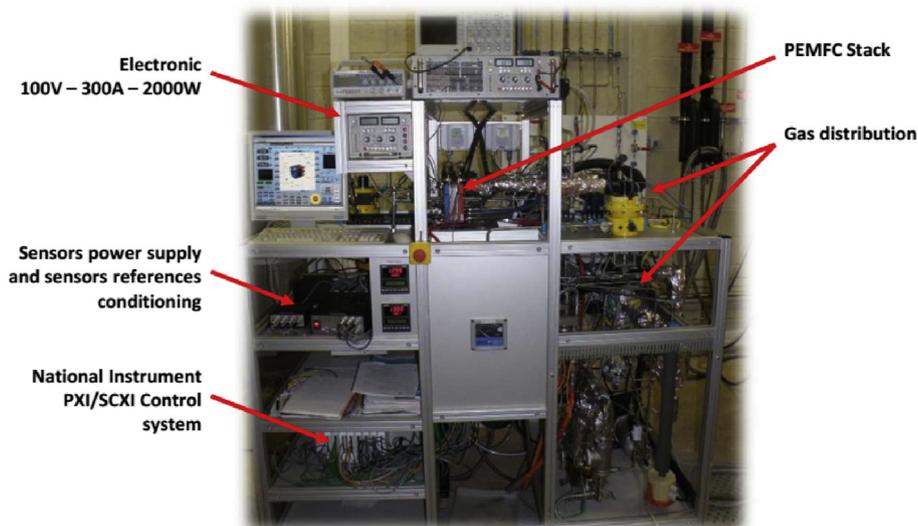
voltage measurements are introduced into the prognostic framework as training dataset.

The prognostic framework can be divided into training process and prognostic process. In the training process, the LSSVM prediction model is trained based on the pre-processed voltage values, while RPF method is performed to track the system state with the measured voltage values input as system observations. In the initialization of RPF method, a set of random values are assigned to the unknown system model parameters. Then, these parameters will update each step in the RPF method until they tend to be stable values. Suppose the prognostic start time is  $T_p$ . The RPF method will stop tracking the system state when the time reaches  $T_p$ , and the model parameters estimated at time  $T_p$  will be used in the following prognostic process. In the prognostic process, each subsequent voltage value of PEMFC is first predicted based on the trained LSSVM prediction model. Then, the predicted voltage values of LSSVM, as new

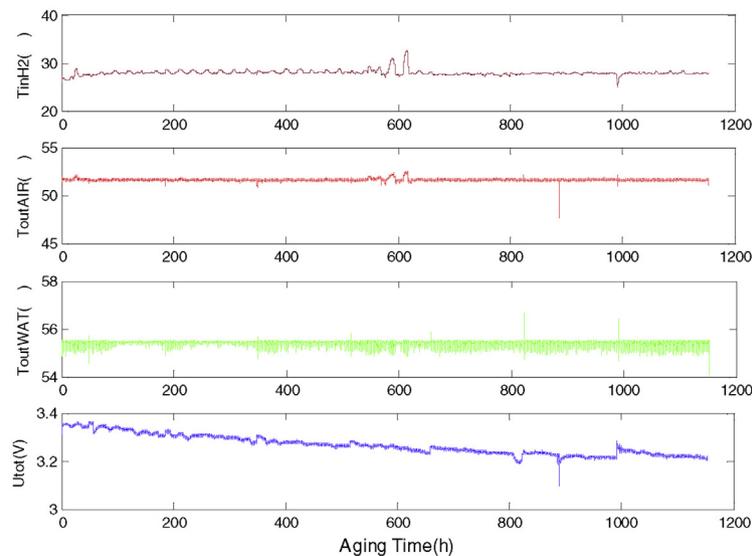
system observation, will be introduced into the prognostic framework of RPF, which receives new observations from LSSVM and predicts the subsequent system state until the given failure threshold is attained. Finally, an uncertainty characterization of the prognostic result in the form of a RUL probability distribution of the PEMFC is provided by RPF method.

### Experimental data and preprocessing

The PEMFC datasets provided by FCLAB Research Federation in IEEE PHM 2014 Data Challenge is utilized for verification of the proposed method. The experimental data are collected from five monolithic fuel cells, each of which has an active area of  $100 \text{ cm}^2$ . The nominal current density of the cells is  $0.70 \text{ A/cm}^2$  and their maximal current density is  $1 \text{ A/cm}^2$ . The test bench is adapted for PEMFC with a power up to  $1 \text{ kW}$ , as shown in Fig. 3(a).



(a)



(b)

Fig. 3 – (a) Test bench of PEMFC provided by FCLAB; (b) Part of the monitored aging parameters of PEMFC.

Two long-term aging tests are performed in the PHM data challenge. In this study, the experimental data collected from the first FC stack, which operates in a stationary condition with a current of 70 A, are utilized for verification. A variety of condition monitoring data (e.g. voltage, temperatures, air flow, gas pressure, etc.) was collected. Fig. 3(b) shows several monitored aging parameters of PEMFC.

As illustrated in Fig. 3(b), an obvious degradation trend of the stack voltage is shown as it continuously drops or fluctuates over time; however, the other parameters exhibit no obvious variation trend as time goes on. Thus, this study adopts stack voltage as the health indicator of PEMFC. Notably, the raw voltage data contain a lot of noise and sharp peaks, and a total of 143,862 data points may also result in time consuming in terms of computation; thus, a data preprocessing is performed on the raw voltage data in this study. The data preprocessing is based on a kernel based smoother which is similar to moving average but the average is weighted [8]. Suppose  $x(t)$  represents the voltage data with  $n$  data points and  $x(t_j)$  is a sample data point at time  $t_j$ , where  $j = 1, 2, 3, \dots, n$ . The filtered voltage data,  $f(t_j)$  can be calculated as follows:

$$f(t_j) = \frac{\sum_{i=1}^n s_i \cdot x(t_j)}{\sum_{i=1}^n s_i} \quad (29)$$

where

$$s = K\left(\frac{t_j - t}{h}\right) \quad (30)$$

$K(\cdot)$  is a Gaussian kernel function with a bandwidth  $h$ , which has the following form:

$$K(t) = \frac{e^{-\frac{t^2}{h^2}}}{\sqrt{2\pi}} \quad (31)$$

After filtering, the cubic spline method is applied to interpolate the filtered data, so that a new voltage dataset with a total of 1155 data points from time 0 h to 1154 h is obtained with a uniform time interval of 1 h. Fig. 4 shows the raw and filtered voltage values of the PEMFC stack.

As it can be seen from Fig. 4, the voltage data is clearly smoothed after data preprocessing. Noise and large peaks are

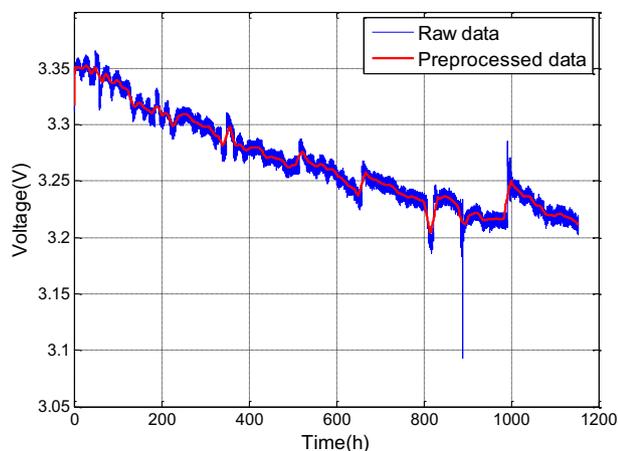


Fig. 4 – Raw and filtered voltage values of the PEMFC stack.

removed while data points are reduced and regularized with a uniform time interval. The preprocessed voltage data are more suitable for the following prognostic, since the uncertainty in the prognostic which may bring disturbance to the prognostic result can be avoided. Failure threshold of the PEMFC stack is set based on the preprocessed voltage data, which means that the RUL estimation also relies on how the data is preprocessed. Additionally, fewer data points in the preprocessed voltage dataset can make the computation more efficiency in the validation experiment.

### Prognostic based on LSSVM

The preprocessed voltage data are first employed as training data of LSSVM to realize the preliminary prognostic, which can be divided into training process and prognostic process. In the training process, a sliding window is used to divide the available voltage data into multiple groups of dataset to form training vectors of LSSVM. Suppose the width of the sliding window is  $n$  and the step length of the sliding window is 1. The voltage data collected before the prognostic start time  $T_p$  are divided into input vectors with  $n$  voltage values and the corresponding output vectors are formed by one subsequent voltage value, as shown in Fig. 5. In this study, a polynomial kernel is chosen for the training process of LSSVM with  $n = 300$ . After the LSSVM model is well trained, the sliding window continues move on, to form the input vector of the subsequent time  $T_p + 1$ . By introducing the new input vector into the trained LSSVM model, the voltage value of time  $T_p + 1$  is predicted. Then, the predicted voltage value is used to form the input vector of time  $T_p + 2$  for further prediction. This procedure continues step by step until the predicted voltage value reaches the given threshold. The prognostic framework of LSSVM method is illustrated in Fig. 5.

### Prognostic based on RPF

Even though LSSVM can reach a good prognostic result of PEMFC, it cannot quantify the uncertainty of the prognostic result. However, uncertainty characterization of the prognostic result is a critical aspect in decision-making process. In this study, RPF method is utilized to realize uncertainty characterization of the prognostic result. Compared with standard PF, RPF method provides a good solution to the degeneracy phenomenon and loss of diversity among the particles, thus improving the accuracy of the prognostic result.

### Voltage drop modelling

State space model is the base of RPF method. In this section, we try to find a proper state model that represents the voltage degradation through time. Several state space models for voltage degradation have been proposed according to relative literature [6,8]:

● State model:

1) Exponential model:

$$x_k = \exp(-\beta \cdot (t_k - t_{k-1})) \cdot x_{k-1} \quad (32)$$

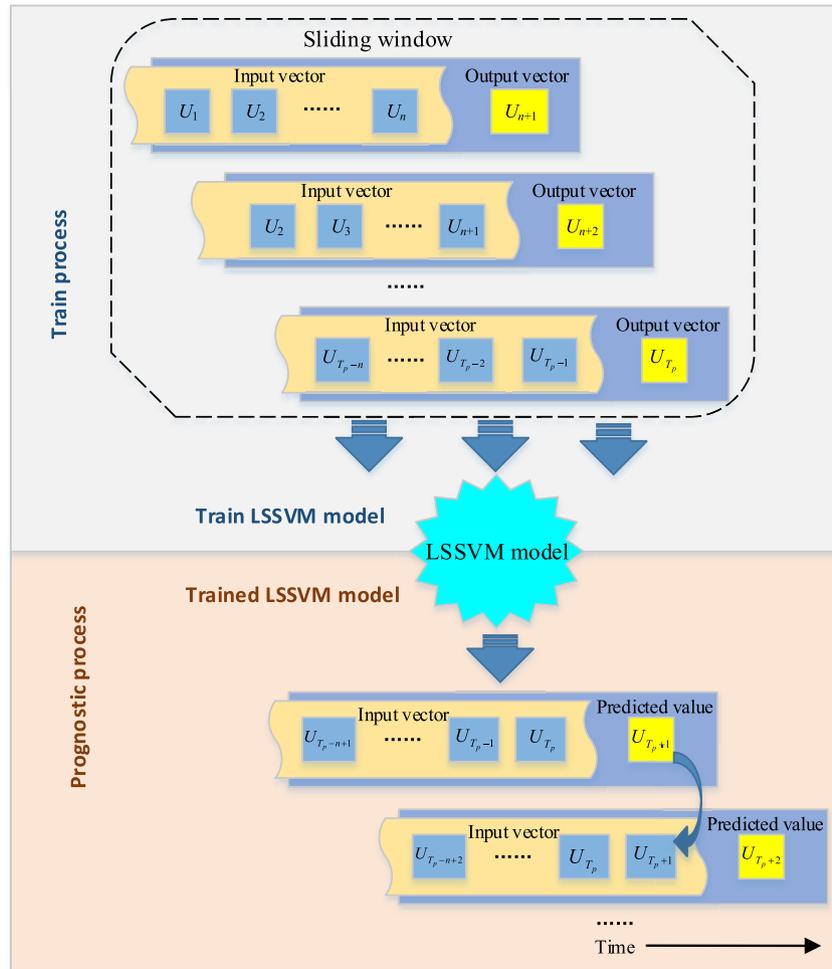


Fig. 5 – Prognostic framework based on LSSVM.

2) Log-linear model:

$$x_k = -\alpha \cdot \ln(t_k/t_{k-1}) - \beta \cdot (t_k - t_{k-1}) + x_{k-1} \quad (33)$$

3) Linear model:

$$x_k = -\beta \cdot (t_k - t_{k-1}) + x_{k-1} \quad (34)$$

The authors of Ref. [6] have tested the above mentioned models using the same dataset of PEMFC. According to the experiment results, the log-linear model shows the best performance. Thus, in this study, the log-linear model is adopted as the state model of voltage degradation. In the log-linear model, the linear part represents the voltage drop under constant current and constant operating condition during aging time, while the logarithm part is used to fit the voltage evolution at the very beginning and end of life.  $\alpha$  and  $\beta$  are system model parameters which need to be estimated using the training dataset. The system noise  $v_k$  is ignored since it can be handled through the uncertainty in the unknown model parameters [6].

● Observation model:

$$z_k = x_k + n_k \quad (35)$$

where  $n_k$  is a Gaussian noise that satisfies  $n_k \sim N(0, \sigma)$ .  $\sigma$  is the unknown standard deviation of the Gaussian noise which also need to be estimated.

#### Parameters initialization and estimation

Initial parameters play an important role to achieve satisfactory prognostic results for both PF and RPF. In this application, the initial parameters include model parameters, i.e.,  $x_0$ ,  $\alpha_0$ ,  $\beta_0$ ,  $\sigma_0$ , as well as algorithm parameters, i.e., the number of particles  $N$  and effective samples  $N_{eff}$ .

The number of particles  $N$  cannot be set too large or too small. A too large number of particles will bring huge calculation, while a too small one will lead to low accuracy. Considering both calculation and accuracy, the number of particles is set  $N = 300$  and the number of effective sample is set  $N_{eff} = 200$ . Suppose the initial system state  $x_0$  follows a uniform distribution centered by the initial voltage with a range of  $\pm 0.05$  V. The unknown parameters  $\alpha_0$ ,  $\beta_0$ , and  $\sigma_0$  are also initialized with a uniform distribution. The ranges of  $\alpha$  and  $\beta$  are obtained by fitting the log-linear model to the initial several voltage measurements. According to the experimental

results, the initial value of  $\alpha$  is set as a uniform distribution with a range of  $[-0.008, -0.007]$  and the initial value of  $\beta$  is set as a uniform distribution with a range of  $[0.0005, 0.0006]$ . The standard deviation  $\sigma$  of the measurement noise is set as a uniform distribution with a range of  $[0.001, 0.002]$ .

For the training process of RPF, the parameters are updated at each step based on the previous voltage data before the starting time  $T_p$ , until they tend to stable values. Finally, the estimated parameters at time  $T_p$  will be used in the following prognostic process.

### RUL estimation based on LSSVM-RPF

In the prognostic process, the LSSVM is first employed to realize the preliminary prediction. Then, the predicted voltage value by LSSVM is introduced into the RPF prognostic framework, acting as system observation. The RPF algorithm receives the system observation value from LSSVM and propagate particles of the next step by extrapolating the system state model based on the estimated model parameters, thus realizing the voltage prediction. If the predicted voltage value of RPF doesn't reach the given end of life (EoL) threshold, the prognostic procedure based on LSSVM-RPF continues until the EoL threshold is reached, as shown in Fig. 6. The two dashed curves in Fig. 6 represent the confidence interval of the degradation state and the probability distribution function (PDF) curve which is filled in yellow color represents the distribution of EoL time when the predicted voltage value reaches the given threshold. The distribution of RUL can be obtained by subtracting the prognostic start time  $T_p$  from the PDF distribution of EoL time.

## Results and discussion

In this section, the performance of the proposed prognostic method based on LSSVM-RPF for PEMFC is evaluated using the datasets provided by FCLAB Research Federation. A comparison is also conducted between the proposed LSSVM-RPF method with RPF, and PF.

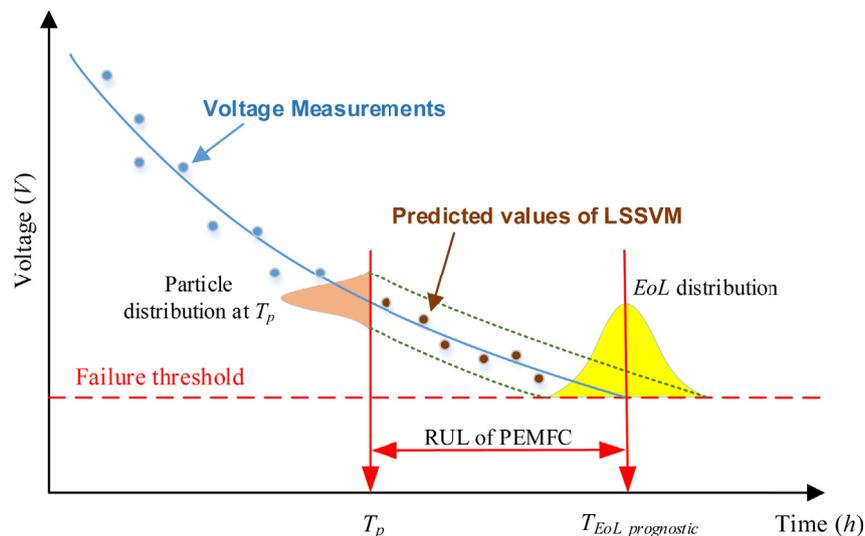


Fig. 6 – Illustration of RUL estimation.

### Comparison of LSSVM-RPF, RPF, and PF at different prognostic start time

In order to investigate the impact of different prediction start time on the prognostic results, this study selects three different times as the prediction start time, that is  $T_p = 400$  h,  $T_p = 500$  h, and  $T_p = 600$  h. Suppose the expected lifetime of PEMFC is 800 h, that is, the actual EoL time is  $T_{EoL\_true} = 800$  h. Then the voltage value at  $T = 800$  h, 3.232 V, is selected as the EoL threshold. The filtered voltage data before the prognostic start time  $T_p$  are employed to realize system state tracking and unknown parameter estimation, while the prognostics are performed from  $T_p + 1$  to 1154 h. State tracking and degradation prognostic results of LSSVM-RPF, RPF, and PF at different prognostic start time are shown in Figs. 7–9. To make a comparison of the prognostic performance of the above mentioned three methods, the relative errors (RE) between the predicted EoL time  $T_{EoL\_prognostic}$  and actual EoL time  $T_{EoL\_true}$  are calculated by equation (36). Root mean square errors (RMSE) between the predicted voltage value  $\hat{z}_k$  and actual experimental data  $z_k$  in the whole lifetime of PEMFC from 0 h to 1154 h are also calculated according to equation (37) as a comprehensive performance indicator which can assess not only the prognostic performance but also system state tracking performance of the above mentioned three methods. The details of the results can be found in Table 1 and Fig. 10.

$$RE = (T_{EoL\_prognostic} - T_{EoL\_true}) / T_{EoL\_true} \quad (36)$$

$$RMSE = \sqrt{\sum_{k=1}^n (\hat{z}_k - z_k)^2 / n} \quad (37)$$

It can be seen from Figs. 7–9 that the proposed prognostic method based on LSSVM-RPF shows a good capability to capture the nonlinear component contained in the voltage measurements and track the system state well based on the available voltage measurements. By learning the degradation evolution law from the training dataset, excellent prognostic results with a similar variation trend to the actual experimental data can be obtained. However, regarding RPF and PF,

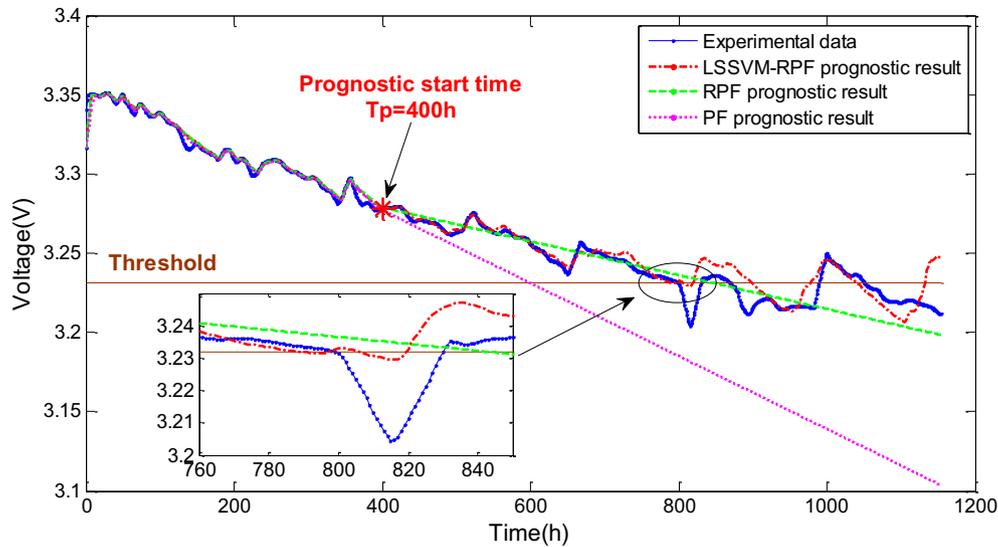


Fig. 7 – Prognostic results of LSSVM-RPF, RPF, and PF at  $T_p = 400$  h.

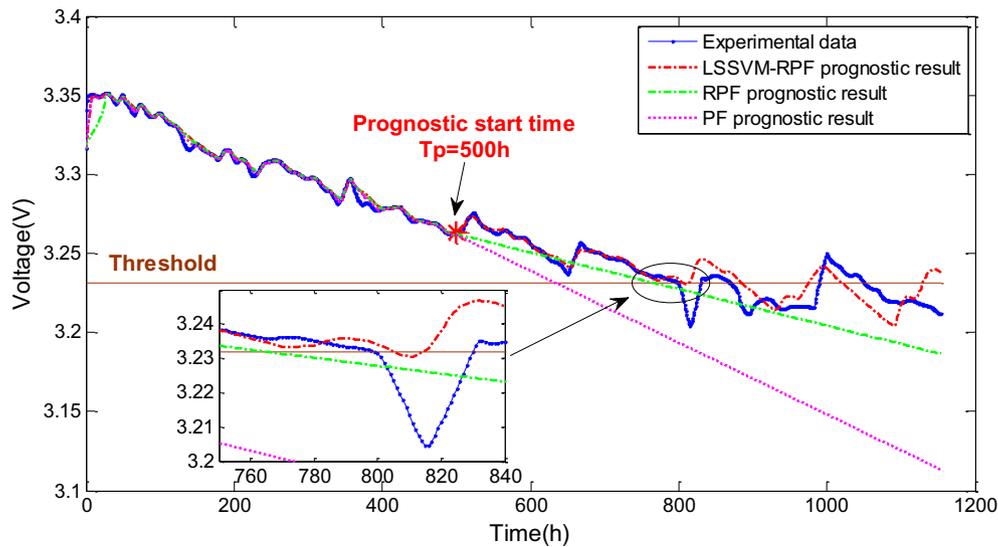


Fig. 8 – Prognostic results of LSSVM-RPF, RPF, and PF at  $T_p = 500$  h.

both methods show a large reliance on the system state model. Even though RPF and PF also exhibit a good capability to track the system states before the prognostic start time, the results of the prognostic process are nearly to a straight line without any fluctuations. This phenomenon is caused due to the linear part of the log-linear system model. Whereas, as an intelligent data-driven based method, LSSVM has a good robust to the nonlinearity of the voltage degradation. Thus, the prognostic accuracy of the proposed LSSVM-RPF method is greatly improved and the reliance on the log-linear system model is reduced by receiving the LSSVM predicted values as system observations.

Additionally, as illustrated in Figs. 7–9, RPF shows a better performance than PF at each prognostic start time, which indicates that the regularization process in RPF has effectively

reduced the degeneracy phenomenon and loss of diversity of the particles. It can also be noticed from Table 1 and Fig. 10 that the later the prognostic start time is, the smaller the REs are, for all of the above mentioned three methods. However, the RMSE values don't show an obvious trend of decrease with the time goes on. That is because characterizations were performed on PEMFC stack at 823 h and 991 h, which cause large fluctuations of the voltage measurements, thus influencing the calculation of RMSE. Nevertheless, both RE and RMSE values of LSSVM-RPF are the lowest. The smallest RE of LSSVM-RPF is 0.0025, obtained at  $T_p = 600$  h, which is nearly 11 times smaller than that of RPF (0.02875), and 34 times smaller than that of PF (0.08625). The smallest RMSE of LSSVM-RPF is 0.0072, obtained at  $T_p = 400$  h, which is nearly 6 times smaller than that of RPF (0.0460), and 64 times smaller than that of PF (0.4601).

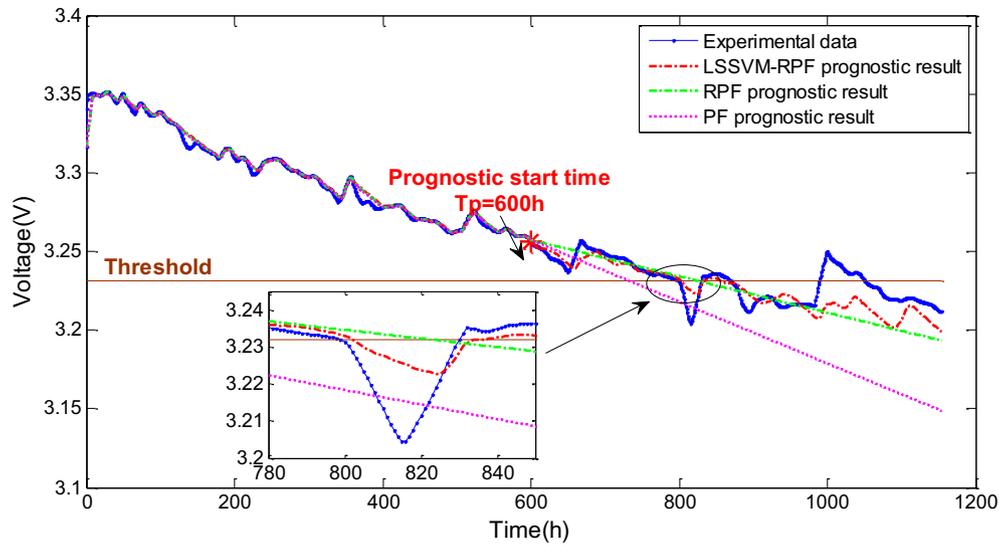


Fig. 9 – Prognostic results of LSSVM-RPF, RPF, and PF at  $T_p = 600$  h.

Uncertainty characterization of RUL

An uncertainty characterization of the prognostic result is a critical aspect in decision-making process, which can provide a better guidance than a single estimated value. In the prognostic process, the RPF method receives new system observations from LSSVM and predicts the subsequent system state by propagating particles for the next step. The distribution of the particles is an approximation of the degradation state of PEMFC at the current time. The particles

are propagated and the voltage values are predicted step by step. Once the predicted voltage value reaches the pre-set EoL threshold, the PDF distribution of predicted EoL time can be formed based on the final distribution of the particles at the last step. Then, the RUL distribution can be obtained by subtracting the prognostic start time  $T_p$  from the PDF distribution of the predicted EoL time. In this study, the RUL probability distribution of PEMFC at different prognostic start time,  $T_p = 400$  h,  $T_p = 500$  h, and  $T_p = 600$  h, are calculated as shown in Fig. 11.

Additionally, confidence interval also plays an important role in uncertainty characterization, which is of great significance for PEMFC users to make good decisions and take adequate actions. In this experiment, the confidence intervals with a significance level of 90% are calculated in both parameter estimation process and prediction process. As illustrated in Fig. 11, the blue line represents the actual experimental data, while the red line represents the predicted voltage values based on LSSVM-RPF. The upper and lower boundaries of the 90% confidence intervals are illustrated using dashed green lines. Table 2 lists the confidence intervals when the predicted voltage value reaches the pre-set EoL threshold at different prognostic start time. As shown in Table 2, the lengths of the confidence intervals become

Table 1 – Prognostic accuracy of LSSVM-RPF, RPF, and PF.

Method	Prognostic result	$T_p = 400$ h	$T_p = 500$ h	$T_p = 600$ h
LSSVM-RPF	$T_{EoL\_Prognostic}$	788	805	802
	RE	0.015	0.00625	0.0025
	RMSE	0.0072	0.0509	0.0644
RPF	$T_{EoL\_Prognostic}$	843	764	823
	RE	0.05375	0.045	0.02875
	RMSE	0.0460	0.2369	0.0892
PF	$T_{EoL\_Prognostic}$	599	633	731
	RE	0.25125	0.20875	0.08625
	RMSE	1.1285	0.9479	0.4601

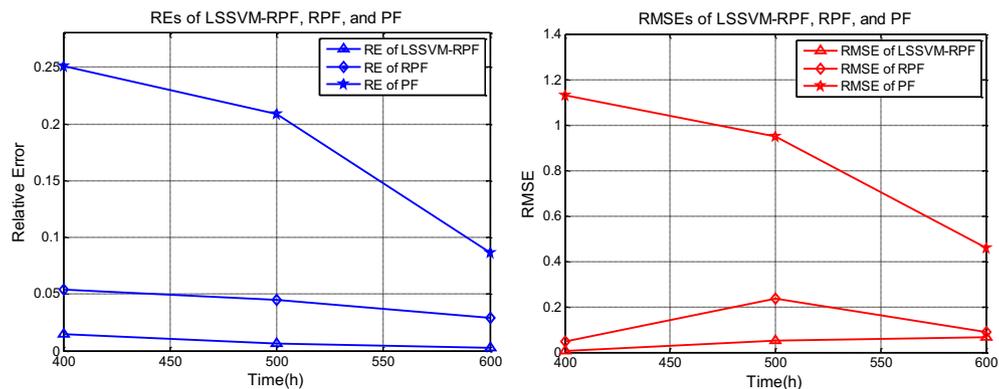
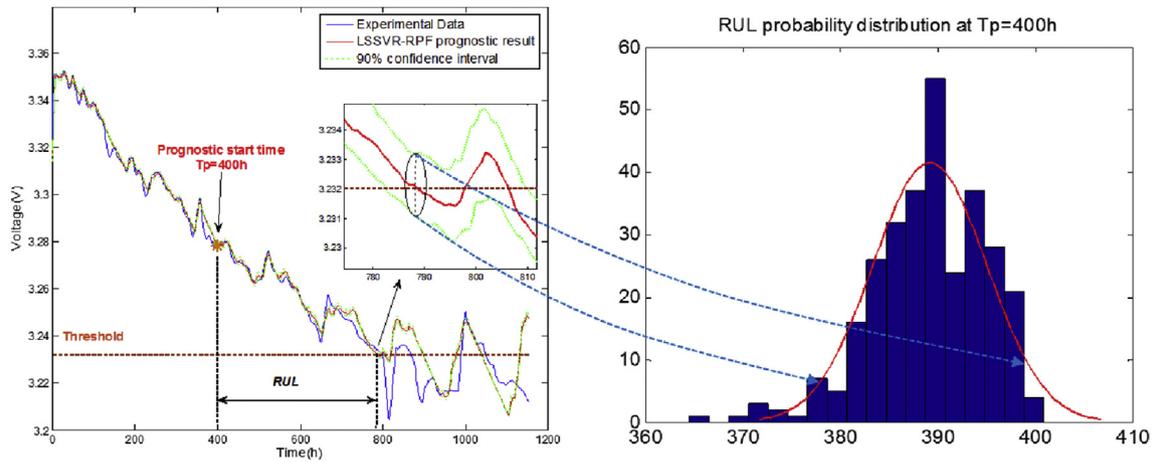
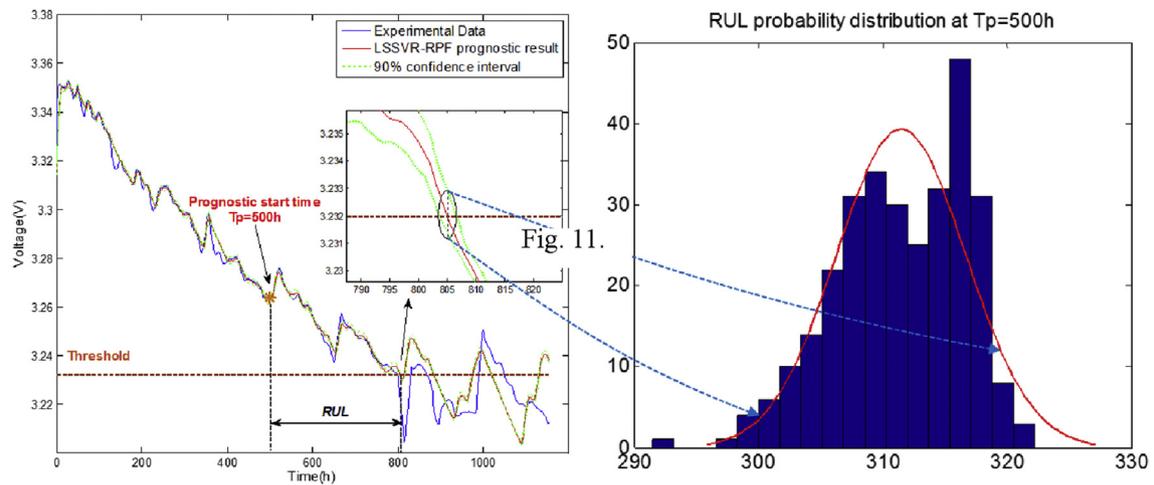


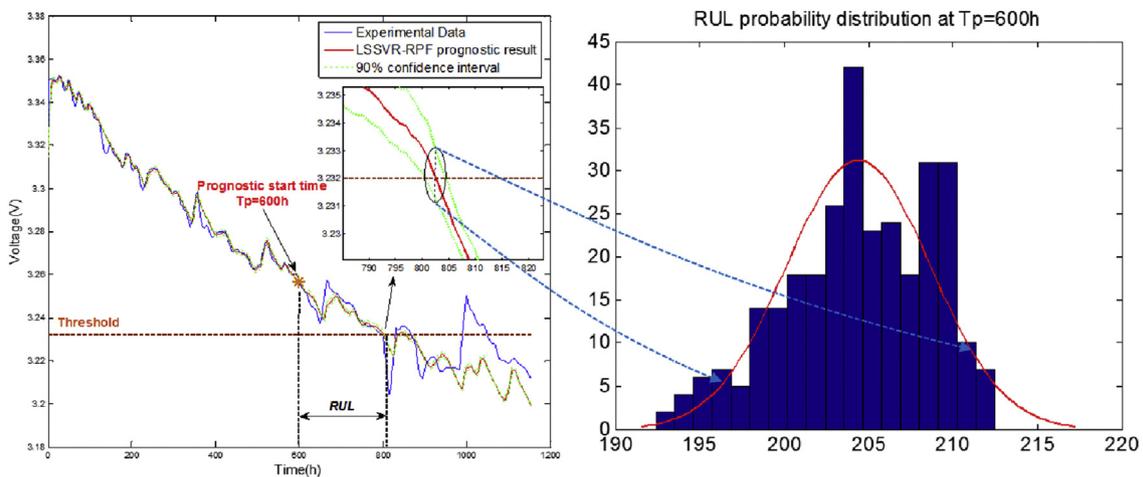
Fig. 10 – REs and RMSEs of LSSVM-RPF, RPF, and PF.



(a)



(b)

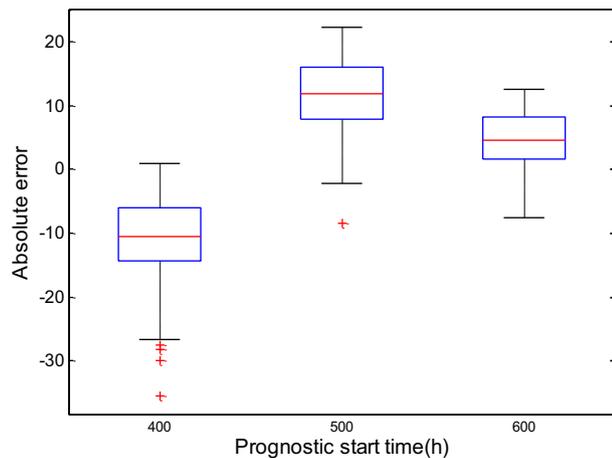


(c)

**Fig. 11** – RUL probability distribution and confidence interval of LSSVM-RPF at (a)  $T_p = 400$  h, (b)  $T_p = 500$  h, and (c)  $T_p = 600$  h.

**Table 2 – Prognostic accuracy of LSSVM-RPF, RPF, and PF.**

		$T_p = 400$ h	$T_p = 500$ h	$T_p = 600$ h
Confidence interval	Upper boundaries	398	318	211
	Lower boundaries	378	302	197
Length of confidence interval		20	16	14

**Fig. 12 – Absolute error of the predicted RUL.**

smaller with the prognostic start time move backward, which indicates that a better prognostic result with a more centered confidence interval can be obtained when more voltage data participate in the parameter estimation and prognostic model construction.

To make a clearer comparison of the distribution of RUL at different prognostic start time, the absolute errors of RUL at each prognostic start time are calculated as follows:

$$AE = RUL_{\text{prognostic}} - RUL_{\text{true}} \quad (38)$$

$$RUL_{\text{true}} = T_{\text{EoL}_{\text{true}}} - T_p \quad (39)$$

The AEs at three different prognostic start times are drawn using a boxplot as illustrated in Fig. 12. In each box, the central red line represents the median value, and the upper and lower edges of each box represent the 25th and 75th percentiles. The individual red crosses are outliers. As illustrated in Fig. 12, with the prognostic start time becomes later, the absolute error of RUL tends to be more centered with a smaller confidence interval and the median value of AE becomes closer to zero, which indicates that a higher accuracy of the prognostic results is achieved.

## Conclusion

This paper presents a hybrid prognostic framework for PEMFC based on LSSVM-RPF. In the proposed prognostic framework, LSSVM is first employed to realize a preliminary prognostic of the PEMFC. Then, RPF method receives the predicted voltage value of LSSVM as new system

observations and output an uncertainty characterization of the prognostic result in the form of a RUL probability distribution. A comparison between the proposed LSSVM-RPF method with RPF and standard PF is conducted based on the PEMFC dataset provided by FCLAB Research Federation. The results show that the proposed prognostic method based on LSSVM-RPF has the best performance.

The main contribution of the proposed method can be summarized as follows:

- 1) The proposed hybrid LSSVM-RPF method combines both the advantages of data-driven method and model-based method. It can provide not only an estimated value of RUL but also an uncertain characterization of RUL with a probability distribution.
- 2) A better prognostic result can be achieved with a lower reliance on system model by receiving the LSSVM predicted values as system observations.
- 3) The proposed hybrid LSSVM-RPF method has a better capability to capture the nonlinearity contained in the voltage degradation data than model-based methods.
- 4) The RPF method has a better performance than the standard particle filter, since RPF method can effectively solve the degeneracy phenomenon and loss of diversity among the particles.

However, limitations also exist in this study, that is, all of the above mentioned three prognostic methods don't consider the variable loading conditions. The verification of the proposed method is also based on PEMFC dataset which is obtained under a stationary condition with a constant current. Whereas, a stationary working condition rarely happens in real applications since PEMFC usually works with a variable load. A typical example is electric vehicles in which PEMFC is the power supply. Thus, the following step of this work will be improving the prognostic framework to realize RUL prognostic and method verification for PEMFC under variable loading conditions, which is a great challenge but of high significance.

## Acknowledgements

This study was supported by the Fundamental Research Funds for the Central Universities (Grant No. YWF-16-BJ-J-18) and the National Natural Science Foundation of China (Grant Nos. 51575021 and 61603016), as well as the China Post-doctoral Science Foundation (Grant Nos. 2017M610033 and 2017T100026).

## REFERENCES

- [1] Jouin M, Gouriveau R, Hissel D, Péra MC, Zerhouni N. Prognostics and health management of PEMFC—state of the art and remaining challenges. *Int J Hydrogen Energy* 2013;38(35):15307–17.
- [2] Liu H, Chen J, Hou M, Shao Z, Su H. Data-based short-term prognostics for proton exchange membrane fuel cells. *Int J Hydrogen Energy* 2017;42(32):20791–808.

- [3] Nigmatullin RR, Martemianov S, Evdokimov YK, Denisov E, Thomas A, Adiutantov N. New approach for PEMFC diagnostics based on quantitative description of quasi-periodic oscillations. *Int J Hydrogen Energy* 2016;41(29):12582–90.
- [4] Jouin M, Gouriveau R, Hissel D, Péra MC, Zerhouni N. Degradations analysis and aging modeling for health assessment and prognostics of PEMFC. *Reliab Eng Syst Saf* 2016;148:78–95.
- [5] Petrone R, Hissel D, Péra MC, Chamagne D, Gouriveau R. Accelerated stress test procedures for PEM fuel cells under actual load constraints: state-of-art and proposals. *Int J Hydrogen Energy* 2015;40(36):12489–505.
- [6] Jouin M, Gouriveau R, Hissel D, Péra MC, Zerhouni N. Prognostics of PEM fuel cell in a particle filtering framework. *Int J Hydrogen Energy* 2014;39(1):481–94.
- [7] Palivonaite R, Lukoseviciute K, Ragulskis M. Short-term time series algebraic forecasting with mixed smoothing. *Neurocomputing* 2016;171:854–65.
- [8] Bressel M, Hilairret M, Hissel D, Bouamama BO. Extended Kalman filter for prognostic of proton exchange membrane fuel cell. *Appl Energy* 2016;164:220–7.
- [9] Kimotho JK, Meyer T, Sextro W. PEM fuel cell prognostics using particle filter with model parameter adaptation// *Prognostics and Health Management (PHM), 2014 IEEE Conference on. IEEE; 2014. p. 1–6.*
- [10] Lechartier E, Laffly E, Péra MC, Gouriveau R, Hissel D, Zerhouni N. Proton exchange membrane fuel cell behavioral model suitable for prognostics. *Int J Hydrogen Energy* 2015;40(26):8384–97.
- [11] Morando S, Jemei S, Hissel D, Gouriveau R, Zerhouni N. ANOVA method applied to proton exchange membrane fuel cell ageing forecasting using an echo state network. *Math Comput Simulat* 2017;131:283–94.
- [12] Wu Y, Breaz E, Gao F, Miraoui A. A modified relevance vector machine for PEM fuel-cell stack aging prediction. *IEEE Trans Ind Appl* 2016;52(3):2573–81.
- [13] Silva RE, Gouriveau R, Jemei S, Hissel D, Boulon L, Agbossou K, et al. Proton exchange membrane fuel cell degradation prediction based on adaptive neuro-fuzzy inference systems. *Int J Hydrogen Energy* 2014;39(21):11128–44.
- [14] Javed K, Gouriveau R, Zerhouni N, Hissel D. Prognostics of Proton Exchange Membrane Fuel Cells stack using an ensemble of constraints based connectionist networks. *J Power Sources* 2016;324:745–57.
- [15] An D, Choi JH, Kim NH. Prognostics 101: a tutorial for particle filter-based prognostics algorithm using Matlab. *Reliab Eng Syst Saf* 2013;115:161–9.
- [16] De Giorgi MG, Campilongo S, Ficarella A, Congedo PM. Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN). *Energies* 2014;7(8):5251–72.
- [17] Yin M, Ye X, Tian Y. Research of fault prognostic based on LSSVM for electronic equipment// *Prognostics and System Health Management Conference (PHM), 2015. IEEE; 2015. p. 1–5.*
- [18] Dong S, Luo T. Bearing degradation process prediction based on the PCA and optimized LS-SVM model. *Measurement* 2013;46(9):3143–52.
- [19] Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20(3):273–97.
- [20] Yuan X, Chen C, Yuan Y, Huang Y, Tan Q. Short-term wind power prediction based on LSSVM–GSA model. *Energy Convers Manag* 2015;101:393–401.
- [21] Suykens JAK, Vandewalle J, De Moor B. Optimal control by least squares support vector machines. *Neural Network* 2001;14(1):23–35.
- [22] Gencoglu MT, Uyar M. Prediction of flashover voltage of insulators using least squares support vector machines. *Expert Syst Appl* 2009;36(7):10789–98.
- [23] Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* 2002;50(2):174–88.