Diagnosis for PEMFC Systems: A Data-Driven Approach With the Capabilities of Online Adaptation and Novel Fault Detection

Zhongliang Li, Rachid Outbib, Stefan Giurgea, and Daniel Hissel, *Senior Member, IEEE*

***Abstract*—In this paper, a data-driven strategy is proposed for polymer electrolyte membrane fuel cell system diagnosis. In the strategy, features are first extracted from the individual cell voltages using *Fisher discriminant analy-sis*. Then, a classification method named *spherical-shaped multiple-class support vector machine* is used to classify the extracted features into various classes related to health states. Using the diagnostic decision rules, the potential novel failure mode can be also detected. Moreover, an online adaptation method is proposed for the diagnosis approach to maintain the diagnostic performance. Finally, the experimental data from a 40-cell stack are proposed to verify the approach relevance.**

***Index Terms*—Classification, data-driven diagnosis, feature extraction, novel fault detection, online adaptation, polymer electrolyte membrane fuel cell (PEMFC) systems.**

I. Introduction

INCREASING environment and resource issues have motivated the development and commercialization of fuel cell technologies. Among the various categories of fuel cells, polymer electrolyte membrane fuel cell (PEMFC) is one of the most promising fuel cells, particularly for mobile applications. To meet the requirements of reliability and durability for practical uses, much effort is being put into research on PEMFC material degradation mechanisms, as well as the design and assembly of fuel cells [1]. In addition to this, the topic of fault diagnosis for PEMFC systems is currently receiving considerably increasing attention. It has been recognized that an efficient fault diagnosis strategy can help PEMFCs (or stacks) operating in relatively optimal and efficient conditions, and, thus, mitigate the performance degradation of fuel cells.

In the literature on PEMFCs, several methods have been proposed for fault diagnosis. Among the most substantial, the approach-oriented models have been explored [2]. This approach is interesting; however, due to the complexity of identification of fuel cell inner parameters, a sufficiently accurate and diagnosis-oriented model is usually difficult to obtain [3]. Therefore, data-driven fault diagnosis methodologies have been drawing the attention of researchers. Several data-driven methods have been proposed for PEMFC diagnosis during the last decade [4]–[6]. Although these preliminary results have been obtained, more effort is yet to be made in this direction. For instance, from a practical point of view, the diagnostic method performance, such as diagnosis accuracy and computational cost, should be seriously evaluated. However, these aspects have been usually unclear or omitted in the literature.

Within the scope of data-driven diagnosis, pattern classification techniques have been widely used for fault diagnosis [7]–[10]. The classification-based diagnostic procedure usually proceeds in two steps. First, an empirical classifier is established from prior knowledge and/or history data. This is considered as the training process. Then, by using the classifier obtained, the real-time data are classified into certain classes that correspond to the health states (normal state or various fault states). Thus, fault detection and isolation can be achieved. In addition, it has been also noticed that the classification performance can be improved by combining some signal analysis and/or feature extraction methods (see, for instance, [5], [11], and [12]). In [12], the feature extraction method, i.e., *Fisher discriminant analysis* (FDA), and the classification method, i.e., *support vector machine* (SVM), were adopted to detect the faults, such as flooding and membrane drying. In [13], the combination of FDA plus SVM is compared with several other representative methods and justified to possess the advantages of both high diagnosis accuracy and low computational cost. Thus, this approach is a promising candidate as an online diagnostic tool for PEMFC systems. Nevertheless, some aspects of the approach still need to be improved.

One of the main limits of the conventional classification methods is that the classifiers trained using the preexisting data can only be used to recognize the known classes. An arbitrary example will be classified into a known class even if it belongs

Manuscript received May 23, 2014; revised September 17, 2014, January 20, 2015, and February 24, 2015; accepted March 8, 2015. Date of publication April 2, 2015; date of current version June 26, 2015. This work was supported by The French National Research Agency (ANR) under Project DIAPASON2 (ANR PAN-H 006-04).

Z. Li is with the LSIS Laboratory, UMR CNRS 6168, Aix-Marseille University, 13397 Marseille Cedex 20, France, with FEMTO-ST/ENERGY (UMR CNRS 6174), 25030 Besançon, France, and also with FCLAB (CNRS 3539), 90010 Belfort Cedex, France (e-mail: zhongliang.li@lsis.org).

R. Outbib is with the LSIS Laboratory, UMR CNRS 6168, Aix-Marseille University, 13397 Marseille Cedex 20, France (e-mail: rachid.outbib@lsis.org).

S. Giurgea is with the University of Technology of Belfort-Montbéliard (UTBM), 90400 Sevenans, France, with FEMTO-ST/ENERGY, 25030 Besançon, France, and also with FCLAB, 90010 Belfort Cedex, France (e-mail: stefan.giurgea@utbm.fr).

D. Hissel is with the University of Franche-Comté (UFC), 25000 Besançon, France, with FEMTO-ST/ENERGY, 25030 Besançon, France, and also with FCLAB, 90010 Belfort Cedex, France (e-mail: daniel.hissel@univ-fcomte.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2015.2418324

to a new cluster, which strongly differs from the samples of known classes.

Specifically for PEMFC systems, multiple physical and chemical processes are involved in the fuel cells, and the systems consist of a set of auxiliary components. A variety of faults could be encountered on different parts of the systems [14]. It is usually not possible to get the data in all the failure modes at the training stage. The data from unseen failure modes would always be falsely diagnosed in such cases.

Moreover, specifically for PEMFCs, the data in the normal operating state are nonstationary when the aging effect is taken into account. For instance, the cell voltages decrease to some degree after a period of time operation. The initially trained diagnosis strategy may gradually lose its efficiency.

To solve the aforementioned problems, in this work, a data-driven fault diagnosis strategy for the PEMFC system is proposed. As in [12], the individual cell voltages are employed as the original variables for diagnosis, and FDA is used to extract the features for classification. A classifier, which is named *spherical-shaped multiple-class SVM* (SSM-SVM), is adopted to classify the features into various classes to fulfill the diagnostic tasks, including the detection and isolation of the known faults and the recognition of the potential novel failure modes. Moreover, an online adaptation procedure is proposed to update the diagnosis models in real time.

The contributions of this work are summarized as follows.

- By using the proposed approach, multiple faults involving various parts of the PEMFC systems can be detected and isolated with high diagnostic accuracy.
- The procedure to detect a novel failure mode is proposed and added to the approach. With the procedure, the false diagnosis can be avoided in the case where a novel fault occurs.
- By using the proposed online updating procedure, the efficiency of the strategy can be kept, when the aging effect is taken into account during long-term operation.
- The low computational cost can certify the real-time implementation of the approach in an embedded system.

This paper is organized as follows: In Section II, the diagnosis strategy is presented. Then, the procedure for adapting the diagnostic approach online is introduced in Section III. Section IV is dedicated to the presentation of the investigated PEMFC system and the experimental database. The diagnosis results and corresponding discussion are given in Section V. Finally, we conclude the work in Section VI.

II. Diagnostic Strategy

A. Framework of the Strategy

The framework of the proposed diagnostic strategy is summarized in Fig. 1. The strategy contains an offline initial training stage, an online performing stage, and an online adapting stage. In the offline training stage, the initial models of FDA and SSM-SVM are successively trained based on the basic training database. Historical samples of cell voltages, which are distributed in the normal class and various fault classes, form the basic training database.

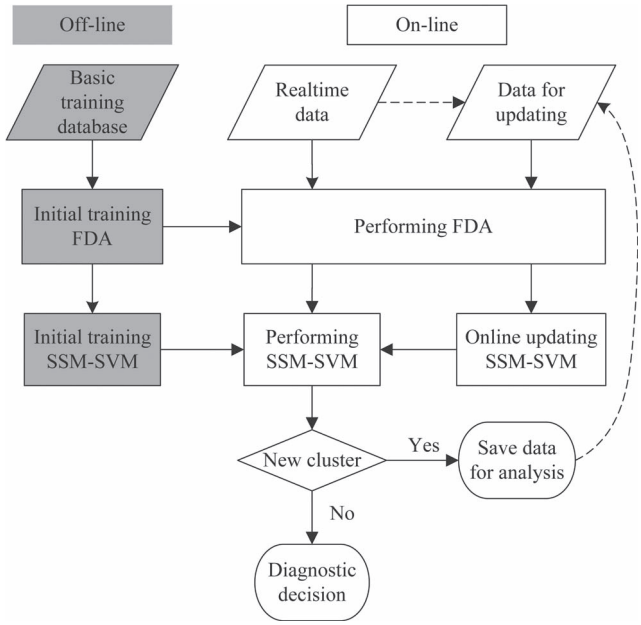


Fig. 1. Flowchart of the proposed diagnostic strategy.

In the online performing stage, the real-time sample (cell voltages) is first processed by the trained FDA model, through which features can be extracted from raw data. Then, with the aid of the trained SSM-SVM model, the features are assigned either to a certain known class to get the diagnostic decision or to a potential novel failure class.

As for the online adapting stage, the labeled data for adapting the models are first collected. The data can either be prepared from real-time samples or from the potential novel failure class. These data are processed using the trained FDA model. Then, the extracted features will be explored to update the SSM-SVM model.

B. Mathematical Description of the Problem

The diagnostic problem can be mathematically abstracted as follows. Let $H \in \mathbb{N}$. Suppose that we have a training data set of N H-dimensional samples $\boldsymbol{x}_n (n \in \mathcal{T} = \{1, \ldots, N\})$, which are distributed in C classes denoted by $\Omega_1, \Omega_2, \ldots, \Omega_C$. In the sequel, for a set Ω, the cardinal (i.e., the number of the elements in Ω) will be denoted as $|\Omega|$. FDA and SSM-SVM models are trained based on the training data set. Through the trained models, a real-time sample $\boldsymbol{x}$ can be classified into a defined class $\Omega_i, i = 1, \ldots, C$ or a novel cluster denoted by Ω_{novel}.

C. FDA

FDA is a technique developed for feature extraction or dimension reduction in the hope of obtaining a more manageable classification problem. Through FDA, the original H-dimensional samples are projected onto L-dimensional space, where $L \in \mathbb{N}$, with $L < H$, such that

$$\boldsymbol{z}_n = \left[\boldsymbol{w}_1^T \boldsymbol{x}_n, \boldsymbol{w}_2^T \boldsymbol{x}_n, \ldots, \boldsymbol{w}_L^T \boldsymbol{x}_n\right]^T \quad n = 1, \ldots, N \tag{1}$$

where $\boldsymbol{z}_n$ is the *projected vector* corresponding to $\boldsymbol{x}_n$, and $\boldsymbol{w}_i$ $(i = 1, \ldots, L)$ are unit *projecting vectors*. The elements of

$\boldsymbol{z}_n$ $(n \in \mathcal{T})$ are named *features*, and the L-dimensional projected space is also called *feature space*.

The objective of training FDA is to find the *projecting vectors*, so that the *projected vectors* of the same class are concentrated, whereas those in different classes are as separated as possible [15].

Avoiding the detailed theoretical deducing process (which can be found in [15]), the seeking of the *projecting vectors* is converted into the following eigenvector problem:

$$\boldsymbol{S}_b \boldsymbol{w}_i = \lambda_i \boldsymbol{S}_w \boldsymbol{w}_i \; i = 1, \ldots, L \text{ and } \lambda_1 \geq \cdots \geq \lambda_L > 0 \quad (2)$$

where $\boldsymbol{S}_w$ and $\boldsymbol{S}_b$ are the *within-class* covariance matrix and the *between-class* covariance matrix, which are given by

$$\boldsymbol{S}_w = \sum_{i=1}^{C} \sum_{\boldsymbol{x}_n \in \Omega_i} (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)^T$$

$$\boldsymbol{S}_b = \sum_{i=1}^{C} |\Omega_i| (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})^T$$

where $\bar{\boldsymbol{x}}_i (i = 1, \ldots, C)$ and $\bar{\boldsymbol{x}}$ are the mean vector of class Ω_i and the mean vector of the total data set, respectively, and are defined by $\bar{\boldsymbol{x}}_i = (1/N_i) \sum_{\boldsymbol{x}_n \in \Omega_i} \boldsymbol{x}_n$ and $\bar{\boldsymbol{x}} = (1/N) \sum_{n=1}^{N} \boldsymbol{x}_n$.

It is found that no more than $C - 1$ of the eigenvalues are nonzero, and the *projecting vectors* correspond to these nonzero eigenvalues [15]. Hence, the dimension of the *feature space* L should satisfy the following constraint:

$$L \leq C - 1. \quad (3)$$

In this paper, the equalization case is preferred. After training FDA, the *projected vectors* $\boldsymbol{z}_n$ $(n \in \mathcal{T})$ corresponding to $\boldsymbol{x}_n$ $(n \in \mathcal{T})$ are exported to the next classification step. *Projecting vectors* $\boldsymbol{w}_i$ $(i = 1, \ldots, L)$ are saved for the use of performing.

D. SSM SVM

SVM is considered as a powerful classifier due to its superior characteristics, such as local minima can be avoided, the solutions can be sparsely represented, and good generalization performance can be achieved [16]. SVM was originally designed for binary classification. By combining the basic binary SVM, multiclass SVM classification can be achieved [17]. Although the classifiers based on the binary SVM can classify the data from the known classes, the capability of detecting an unseen cluster seems to be defective. As Fig. 2(a) shows, the bounders among the trained classes (i.e., Class1, Class2, and Class3) can be affirmed by a multiclass classifier based on binary SVM. According to the decision of the trained classifier, an arbitrary sample will be classified into one of the three classes, even if the data are from a novel cluster, as shown in Fig. 2(b).

Hao and Lin in [18] proposed SSM-SVM. Different from the binary-SVM-based classifier, the approach achieves the classification goal by seeking class-specific spheres. The samples from one specific class are enclosed by the corresponding sphere, whereas those from other classes are excluded outside. As Fig. 2(c) shows, the closed bounders for all of the known classes

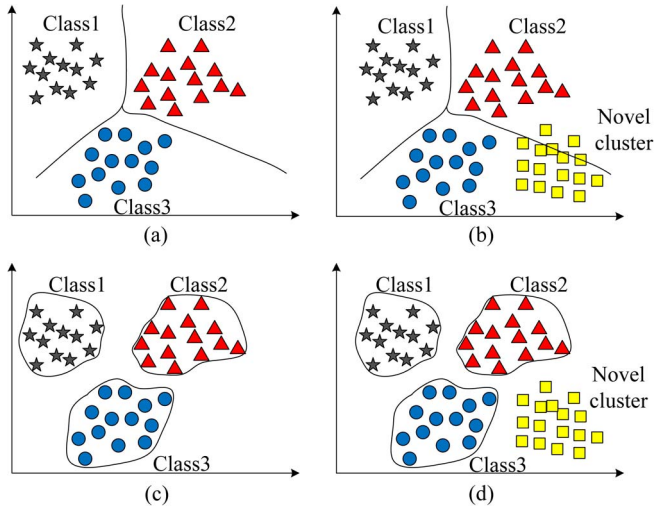


Fig. 2. Schematic diagrams of a conventional binary-SVM-based multiclass classifier and SSM-SVM.

can be found by training SSM-SVM. Thus, the samples from a novel cluster could probably be detected if they are outside all the closed bounders, as shown in Fig. 2(d).

Following the FDA step, the training of SSM-SVM is based on the *projected vectors* $\boldsymbol{z}_n$ $(n \in \mathcal{T})$. More precisely, to solve the nonlinear classification problem, the *projected vectors* are first projected onto high-dimensional space via a nonlinear transform Φ [18]. Taking the ith class for example, the method is realized by seeking the sphere with the minimal radius in the high-dimensional space. The sphere encloses all data points in the ith class and leaves the other data points outside. That is,

$$\begin{cases} \|\Phi(\boldsymbol{z}_n) - \boldsymbol{a}_i\|^2 \leq R_i^2 + \xi_n^i & \text{if } \boldsymbol{z}_n \in \Omega_i \\ \|\Phi(\boldsymbol{z}_n) - \boldsymbol{a}_i\|^2 \geq R_i^2 - \xi_n^i & \text{if } \boldsymbol{z}_n \notin \Omega_i \end{cases} \quad (4)$$

where R_i and $\boldsymbol{a}_i$ are the radius and center of the ith sphere, respectively, and ξ_n^i, which satisfy $\xi_n^i \geq 0$, are the slack variables corresponding to the training data point $\boldsymbol{z}_n$ $(n \in \mathcal{T})$. The slack variables permit the occurrence of errors. For instance, the data in class i could be outside the sphere, whereas the data outside class i could be inside the sphere.

This amounts to solving the following optimization problem (see [18]):

$$\min_{R_i, \boldsymbol{a}_i} \left(R_i^2 + D \sum_{n \in \mathcal{T}} \xi_n^i \right)$$

$$\text{s.t.} \begin{cases} c_n^i \left(\|\Phi(\boldsymbol{z}_n) - \boldsymbol{a}_i\|^2 - R_i^2 \right) - \xi_n^i \leq 0 \\ \xi_n^i \geq 0 \end{cases} \quad \text{for } n \in \mathcal{T} \quad (5)$$

where $c_n^i = 1$ if $\boldsymbol{z}_n \in \Omega_i$ and $c_n^i = -1$ if $\boldsymbol{z}_n \notin \Omega_i$; D is a parameter controling the penalty of errors [18].

By using the Lagrange multipliers, one can deduce that the solutions for problem (5) are given by (see [18])

$$\boldsymbol{a}_i = \sum_{n \in \mathcal{T}} \alpha_n^i c_n^i \Phi(\boldsymbol{z}_n) \quad (6)$$

where $\alpha_n^i (n \in \mathcal{T})$ denote the Lagrange multipliers, and

$$R_i = \|\Phi(\boldsymbol{z}_n) - \boldsymbol{a}_i\| \quad \text{for some } \boldsymbol{z}_n \text{ so that } \alpha_n^i \in (0, D). \quad (7)$$

E. Diagnostic Rules

The goal of this section is to present the diagnostic rules, including that for novel cluster detection.

Let $F_i : \mathbb{R}_+ \to \mathbb{R}_+$ be a smooth function such that F_i is decreasing with $\lim_{\tau\to\infty} F_i(\tau) = 0$. In the classical approach, i.e., without detection of a novel cluster, a general sample $\boldsymbol{z}$ is allotted to a class using the following criterion:

$$\text{Class}(\boldsymbol{z}) = \arg\max_i F_i\left(\|\Phi(\boldsymbol{z}) - \boldsymbol{a}_i\|\right). \quad (8)$$

It should be noted that in this classical approach, the number of classes is fixed, and the sample $\boldsymbol{z}$ is associated to a class even if the distances to the different centers are very large.

In this paper, it is assumed that the classes are not limited to those initially defined. Furthermore, the fact that the distances between a sample $\boldsymbol{z}$ and the different centers are very large, i.e., $\max F_i(\|\Phi(\boldsymbol{z}) - \boldsymbol{a}_i\|)$ is very small and, in a meaning to be defined, can mean the appearance of a novel cluster.

The principle for deciding that a sample belongs to a defined class or to a new cluster can be described as follows. For each class Ω_i, $\delta_i \in \mathbb{R}_+$ is considered to denote the threshold from which a sample $\boldsymbol{z}$ is considered to be definitely outside the class. More precisely, it is assumed that $\boldsymbol{z}$ is outside Ω_i if $F_i(\|\Phi(\boldsymbol{z}) - \boldsymbol{a}_i\|) < \delta_i$. The value of the threshold can be determined based on a calibration data set, and a way to fix its value is to use the *3-sigma law*, i.e.,

$$\delta_i = M_i - 3\sqrt{\frac{1}{|\Omega_i|}\sum_{z_n\in\Omega_i}\left(F_i\left(\|\Phi(\boldsymbol{z}) - \boldsymbol{a}_i\|\right) - M_i\right)^2} \quad (9)$$

with $M_i = (1/|\Omega_i|)\sum_{z_n\in\Omega_i} F_i(\|\Phi(\boldsymbol{z}) - \boldsymbol{a}_i\|)$.

Within this approach, the decision rule is defined by

$$\text{Class}(\boldsymbol{z}) = \begin{cases} \arg\max_i F_i(\boldsymbol{z}) & \text{if } \max F_i(\boldsymbol{z}) \geq \delta_i \\ \text{novel} \quad \text{if } \max F_i & (\|\Phi(\boldsymbol{z}) - \boldsymbol{a}_i\|) < \delta_i. \end{cases} \quad (10)$$

Note that by considering the threshold given by (9), less than 0.15% of the samples in the training data set are misclassified to the novel class.

As for function F_i, several expressions are possible. For the implementation part, that proposed in [18] is used.

Here, the principle for the decision rule is presented. However, the strategy is presented by using the nonlinear transform Φ and without explicitly giving the solutions depending on the available data. Thus, the goal of the training process is to solve the quadratic problem (QP) expressed by (5). In the following section, an online learning method is used to achieve the training procedure.

III. Online Adaptation Method

From (6) and (7), it is observed that the solutions of problem (5) correspond to the multipliers $\alpha_n^i (n \in \mathcal{T})$ and nonlinear transform Φ. By introducing the *kernel function*, the dual problem of (5), which is in terms of $\alpha_n^i (n \in \mathcal{T})$, is deduced (see [18]) as

$$\min \left(\sum_{n,m\in\mathcal{T}} \alpha_n^i Q_{n,m}^i \alpha_m^i - \sum_{n\in\mathcal{T}} \alpha_n^i c_n^i k(\boldsymbol{z}_n, \boldsymbol{z}_n) \right)$$
$$\text{s.t.} \sum_{n\in\mathcal{T}} \alpha_n^i c_n^i = 1 \quad \text{and} \quad 0 \leq \alpha_n^i \leq D \qquad \forall n \quad (11)$$

where $Q_{n,m}^i = c_n^i c_m^i k(\boldsymbol{z}_n, \boldsymbol{z}_m)$; $k(\boldsymbol{z}_n, \boldsymbol{z}_m) = \Phi(\boldsymbol{z}_n)\Phi(\boldsymbol{z}_m)$ is known as the *kernel function*, which implicitly defines the transformation Φ. Throughout this paper, the Gaussian kernel $k(\boldsymbol{z}_n, \boldsymbol{z}_m) = \exp(-\|\boldsymbol{z}_n - \boldsymbol{z}_m\|^2/\sigma)$ will be used.

In [18], the sequential minimal optimization method is proposed to solve the QP problem (11) so as to train SSM-SVM. This method is known as a batch training approach, which is performed in one batch. It implies that if more training data arrive subsequently, the SSM-SVM classifier should be retrained from scratch. This is considered to be computationally inefficient [19]. In [20], an incremental learning method is proposed for training the classic binary SVM. In this method, the solution for $N + 1$ training data could be formulated in terms of that for N data and one new data point. Here, the incremental training method is extended to the training and online adaptation for SSM-SVM without much modification. In addition, a procedure is also proposed to further lower the real-time computational cost.

A. Incremental Learning Method for SSM-SVM

By introducing another Lagrange multiplier b_i, problem (11) can be reexpressed as

$$\min_{0\leq\alpha_n^i,\alpha_m^i\leq D} W_i = \frac{1}{2}\sum_{n,m\in\mathcal{T}} \alpha_n^i Q_{n,m}^i \alpha_m^i + b_i\left(\sum_{n\in\mathcal{T}} c_n^i \alpha_n^i - 1\right). \quad (12)$$

The goal of incremental learning is to keep the Kuhn–Tucker (KT) conditions of (12) satisfied when a new training sample is added to the current training data.

1) Incremental Procedure: To solve the optimization problem (12), we proceed as in [20]. Let $g_n^i (n \in \mathcal{T})$ and h_i be the quantities defined by

$$g_n^i = \frac{\partial W_i}{\partial \alpha_n^i} = \sum_{m\in\mathcal{T}} Q_{n,m}^i \alpha_m + c_n^i b_i \quad (13)$$

$$h_i = \frac{\partial W_i}{\partial b_i} = \sum_{n\in\mathcal{T}} c_n^i \alpha_n^i - 1 = 0. \quad (14)$$

According to the value of g_n^i, set $\mathcal{T}$ is partitioned into three sets, i.e.,

$$\mathcal{T} = \mathcal{T}_m^i \cup \mathcal{T}_E^i \cup \mathcal{T}_R^i$$
$$= \left\{n\in\mathcal{T} : g_n^i = 0\right\} \cup \left\{n\in\mathcal{T} : g_n^i \leq 0\right\} \cup \left\{n\in\mathcal{T} : g_n^i \geq 0\right\}. \quad (15)$$

It should be noted that the solutions for (12) (see, for instance, [20]) are such that

$$\begin{cases} \alpha_n^i = 0 \text{ for } n \in \mathcal{T}_R^i \\ \alpha_n^i \in (0, D) \text{ for } n \in \mathcal{T}_M^i \\ \alpha_n^i = D \text{ for } n \in \mathcal{T}_E^i. \end{cases} \tag{16}$$

We use $(t_j^M)_{j=1,\ldots,|\mathcal{T}_M^i|}$ as the denotations of the elements of $\mathcal{T}_M^i$, and $(\bar{t}_j^M)_{j=1,\ldots,|\mathcal{T}_E^i \cup \mathcal{T}_R^i|}$ as those of $\mathcal{T}_E^i \cup \mathcal{T}_R^i$.

The matrix $\boldsymbol{P}^i \in \mathbb{R}^{(|\mathcal{T}_M^i|+1)\times(|\mathcal{T}_M^i|+1)}$, which will be used later, is defined as

$$\begin{cases} P_{1,1}^i = 0, P_{1,k+1}^i = P_{k+1,1}^i = c_k^i, \text{ and } P_{l+1,k+1}^i = Q_{t_l^M, t_k^M}^i \\ \text{for } k, l \in \left\{1, \ldots, \left|\mathcal{T}_M^i\right|\right\}. \end{cases} \tag{17}$$

In the following, $\mathcal{U}$ will denote the index set of unlearned vectors. Let $\boldsymbol{z}_s$ $(s \in \mathcal{U})$ be a new sample to be added to the learned data. Moreover, let $\alpha_s^i = 0$ be the coefficient assigned to $\boldsymbol{z}_s$ and g_s^i be the quantity associated to α_s^i and determined using (13). If $g_s^i \geq 0$, s will be added to $\mathcal{T}_R^i$, and thus, the KT conditions are satisfied. Otherwise, the KT conditions are maintained by varying the *margin vector* coefficients α_n^i $(n \in \mathcal{T}_M^i)$ and b_i in response to the perturbation imparted by the incremented new coefficient $\Delta\alpha_s^i$, until s enters into $\mathcal{T}_E^i$ or $\mathcal{T}_M^i$.

Taking into account the perturbation caused by the incremental step, the coefficient differences $\Delta\alpha_l^i, \Delta g_l^i$ $(l \in \mathcal{T} \cup \{s\})$, and Δb_i are introduced. Furthermore, for $\tau \in \{\Delta b_i, \Delta\alpha_n^i, \Delta g_n^i; n \in \mathcal{T}\}$, *coefficient sensitivities* $\bar{\tau}$ are defined so that $\tau = \bar{\tau}\Delta\alpha_s^i$.

The KT conditions (13) and (14) can be differentially expressed as

$$\Delta g_n^i = \sum_{m \in \mathcal{T}_M^i \cup \{s\}} Q_{n,m}^i \Delta\alpha_m^i + c_n^i \Delta b_i \quad \forall n \in \mathcal{T} \cup \{s\} \tag{18}$$

$$\Delta h_i = \sum_{m \in \mathcal{T}_M^i \cup \{s\}} c_m^i \Delta\alpha_m^i = 0. \tag{19}$$

For all $n \in \mathcal{T}_M^i$, the condition $g_n^i \equiv 0$ should be maintained. Thus, it can be deduced from (18) and (19) that

$$\boldsymbol{P}^i \begin{pmatrix} \Delta b_i \\ \Delta\alpha_{t_1^M}^i \\ \vdots \\ \Delta\alpha_{t_{|\mathcal{T}_M^i|}^M}^i \end{pmatrix} = -\begin{pmatrix} c_s^i \\ Q_{t_1^M, s}^i \\ \vdots \\ Q_{t_{|\mathcal{T}_M^i|}^M, s}^i \end{pmatrix} \Delta\alpha_s^i \tag{20}$$

or, in terms of coefficient sensitivities, we have

$$\begin{pmatrix} \bar{b}_i \\ \bar{\alpha}_{t_1^M}^i \\ \vdots \\ \bar{\alpha}_{t_{|\mathcal{T}_M^i|}^M}^i \end{pmatrix} = \bar{\boldsymbol{P}}^i \begin{pmatrix} c_s^i \\ Q_{t_1^M, s}^i \\ \vdots \\ Q_{t_{|\mathcal{T}_M^i|}^M, s}^i \end{pmatrix} \tag{21}$$

where $\bar{\boldsymbol{P}}^i = -\{\boldsymbol{P}^i\}^{-1}$. Note that $\bar{\alpha}_n^i \equiv 0$ for all $n \in \mathcal{T}_E^i \cup \mathcal{T}_R^i$.

According to (18) and (21), it can be deduced that

$$\bar{g}_n^i = \sum_{l \in \mathcal{T}_M^i \cup \{s\}} Q_{n,l}^i \bar{\alpha}_l^i + c_n^i \bar{b}_i, \qquad \forall n \in \mathcal{T}_E^i \cup \mathcal{T}_R^i \cup \{s\} \tag{22}$$

and $\bar{g}_n^i \equiv 0$ for all $n \in \mathcal{T}_M^i$.

From the given explanations, it can be seen that $\Delta\alpha_s^i$ can be absorbed by varying α_n^i $(n \in \mathcal{T}_M^i)$ and b_i. Meanwhile, g_n^i $(n \notin T_M^i)$ vary accordingly. Thus, within several incremental steps, s will be added to category $\mathcal{T}_E^i$ when $\alpha_s^i = D$ or to $\mathcal{T}_M^i$ when $g_s^i = 0$ [20].

Remarks: As in [20], some procedures in the incremental learning algorithm can be used here without modifications. For instance, we have the following.

- The composition of sets $\mathcal{T}_M^i$, $\mathcal{T}_E^i$, and $\mathcal{T}_R^i$ can be changed with the change of α_l^i $(l \in \mathcal{T}_M^i \cup \{s\})$ and g_n^i $(n \in \mathcal{T}_E^i \cup \mathcal{T}_R^i \cup \{s\})$. The procedure to determine the maximum $\Delta\alpha_s^i$ such that the index of some sample migrates can be utilized here.
- To account for the new composition of $\mathcal{T}_M^i$, matrix $\bar{\boldsymbol{P}}^i$ must be recomputed. The procedure of recursive computing of $\bar{\boldsymbol{P}}^i$ proposed in [20] can be also adopted in our case.

2) Incremental Learning Algorithm: The overall incremental procedure for learning the new sample $\boldsymbol{z}_s$ can be summarized as Algorithm 1. The samples in $\mathcal{U}$ can be sequentially learned using the algorithm.

Algorithm 1 Incremental SSM-SVM algorithm

1: Assign $\alpha_s^i = 0$;
2: Compute g_s^i;
3: **if** $g_s^i \geq 0$ **then**
4: Add s to $\mathcal{T}_R^i$;
5: **end if**
6: **while** $g_s^i < 0$ & $\alpha_s^i < D$ **do**
7: Compute $M_{\Delta\alpha_s^i} = \max \Delta\alpha_s^i$; /* See [20] */
8: Update $\alpha_s^i + M_{\Delta\alpha_s^i} \to \alpha_s^i$;
9: Compute $\bar{b}_i$;
10: Update $b_i + \bar{b}_i M_{\Delta\alpha_s^i} \to b_i$;
11: **for** $l = 1$ to $|\mathcal{T}_M^i|$ **do**
12: Compute $\bar{\alpha}_{t_l^M}^i$;
13: Update $\alpha_{t_l^M}^i + \bar{\alpha}_{t_l^M}^i M_{\Delta\alpha_s^i} \to \alpha_{t_l^M}^i$;
14: **end for**
15: **for** $l = 1$ to $|\mathcal{T}_E^i \cup \mathcal{T}_R^i|$ **do**
16: Compute $\bar{g}_{\bar{t}_l^M}^i$;
17: Update $g_{\bar{t}_l^M}^i + \bar{g}_{\bar{t}_l^M}^i M_{\Delta\alpha_s^i} \to \alpha_{\bar{t}_l^M}^i$;
18: **end for**
19: Update $\mathcal{T}_M^i$, $\mathcal{T}_E^i$, $\mathcal{T}_R^i$;
20: Update $\boldsymbol{P}^i$; /* Use recursive procedure (see [20]) */
21: **end while**

3) Initialization Procedure: The KT conditions (13) and (14) are assumed to be satisfied on $\mathcal{T}$ before carrying out Algorithm 1. However, the conditions are not satisfied initially by default, when the training of a new SSM-SVM classifier is

launched. An initial procedure is therefore necessary to make the KT conditions fulfilled for a certain number (N^i) of training samples. The initialization procedure proposed in this study is summarized as Algorithm 2.[1]

Algorithm 2 Initialization of incremental SSM-SVM

Assign $N^i = \lfloor 1/D \rfloor + 1$;
Select $\boldsymbol{z}_1^i, \ldots, \boldsymbol{z}_{N^i}^i$ from Ω_i;
Assign $\alpha_1^i = 1 - \lfloor 1/D \rfloor D$;
for $n = 2$ to N^i **do**
  Assign $\alpha_n^i = D$;
end for
for $n = 1$ to N^i **do**
  Compute $g_n^i = \sum_{m=1}^{N_{ini}} Q_{n,m}^i \alpha_m^i$;
end for
Assign $b_i = -\max\{g_n^i\}$;
for $n = 1$ to N^i **do**
  Update $g_n^i + b_i \rightarrow g_n^i$;
end for
if $\arg\max_n g_n^i = 1$ **then**
  Assign $\mathcal{T}_M^i = \{1\}$, Assign $\mathcal{T}_E^i = \{2, \ldots, N^i\}$;
else
  Assign $\mathcal{T}_M^i = \{\arg\max_n g_n^i\}$, $\mathcal{T}_E^i = \{2, \ldots, N^i\} - \{\arg\max_n g_n^i\}$;
  Assign $s = 1$, go to step 6 of Algorithm 1;
end if

B. Improvement of Real-Time Learning Performance

It can be noticed that the ith sphere ($\boldsymbol{a}_i$, R_i) is determined by the samples associated to $\mathcal{T}_M^i$ and $\mathcal{T}_E^i$, whereas Algorithm 1 involves the whole set $\mathcal{T}$. Based on the tacit assumption that the sphere surface before incremental training does not move much after the incremental procedure, the samples that are associated to $\mathcal{T}_R^i$ and are far away from the sphere surface have little impact on the training results. Thus, these samples could be discarded to reduce the memory consumption and the computation time. Nevertheless, the potential candidates for $\mathcal{T}_M^i$ and $\mathcal{T}_E^i$ should not be deleted in the deletion process.

In this paper, a procedure is proposed to keep the training candidates and to discard the useless ones whose indexes are from $\mathcal{T}_R^i$. The samples nearest the sphere surface are kept, for they are the more promising candidates. To find these samples, the square of the distance from $\Phi(\boldsymbol{z}_n)$ to the center $\boldsymbol{a}_i$ is defined as

$$ {d_n^i}^2 = \left\| \Phi(\boldsymbol{z}_n) - \boldsymbol{a}_i \right\|^2 . \tag{23} $$

The difference of ${d_n^i}^2$ and ${R_i}^2$ is deduced from (6), (7), (13), and (23) (see the Appendix), i.e.,

$$ \left| {d_n^i}^2 - R_i^2 \right| = 2 g_n^i \quad n \in \mathcal{T}_R^i. \tag{24} $$

[1]In the algorithm, $\lfloor x \rfloor$ denotes the floor function largest integer not greater than x.

Hence, the samples corresponding to the $|\mathcal{T}_R^i|$ smallest g_n^i are kept, whereas the others are discarded. Here, the maximum $|\mathcal{T}_R^i|$ is set as twice the value $|\mathcal{T}_M^i \cup \mathcal{T}_E^i|$, as

$$ \left|\mathcal{T}_R^i\right| = \begin{cases} \left|\mathcal{T}_R^i\right|, & \text{if } \left|\mathcal{T}_R^i\right| \le 2\left|\mathcal{T}_M^i \cup \mathcal{T}_E^i\right| \\ 2\left|\mathcal{T}_M^i \cup \mathcal{T}_E^i\right|, & \text{if } \left|\mathcal{T}_R^i\right| > 2\left|\mathcal{T}_M^i \cup \mathcal{T}_E^i\right| \end{cases} . \tag{25} $$

Since $|\mathcal{T}_M^i \cup \mathcal{T}_E^i|$ is usually small, this procedure can confirm the slight memory consumption and computation.

IV. Experiments and Database

A. PEMFC Stack and Test Bench

A 5-kW PEMFC test bench was used to carry out the various experiments (see Fig. 3 for an overall view). As shown in Fig. 4, this test bench is composed of a fuel cell stack and of the following subsystems.

- Air supply subsystem: The flow rate, pressure, temperature, and hygrometry level at the air inlet can be regulated to the required conditions.
- Hydrogen supply subsystem: The flow rate and pressure at the hydrogen inlet, temperature, and hygrometry level at the hydrogen inlet can be regulated to the required conditions. Due to the compressor and the flow controller (mass flow controller 2 in Fig. 4), hydrogen recirculation can be achieved.
- Temperature subsystem: This subsystem is dedicated to the control of the stack temperature.
- Electronic load: The load current can be flexibly varied through an electronic load.
- Control/supervision unit: The controls of the test bench and the parameter monitoring are fulfilled using National Instruments Materials and Labview software.

A more detailed description of the test bench can be found in [21].

A 40-cell PEMFC stack was investigated.[2] The active area of the stack/cell is 220 cm^2. The nominal operating conditions of the stack are summarized in Table I.

B. Experimental Database

The experiments in the normal state and various faulty states were carried out on the test bench. The concerned health states are listed in Table II. Note that the normal tests were carried out at four different time points to take the influence of aging into account. The faults were manually induced by varying the operation conditions. For instance, the fault denoted by F_3 in Table II was deduced by setting the air stoichiometry value to 4. The data sampled during the faulty operation periods are collected as the database of the corresponding states.

The experiment in each state was repeated several times. Although various physical variables had been sampled and collected from the test bench, only the cell voltages sampled during

[2]The stack was fabricated by the French research organization Alternative Energies and Atomic Energy Commission (CEA) within the framework of the French ANR DIAPASON project.

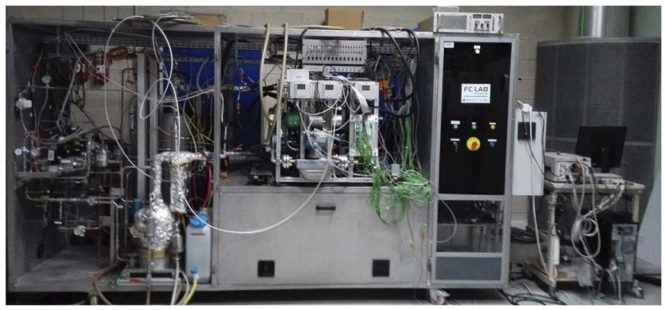


Fig. 3. Photograph of the employed 5-kW PEMFC test bench.

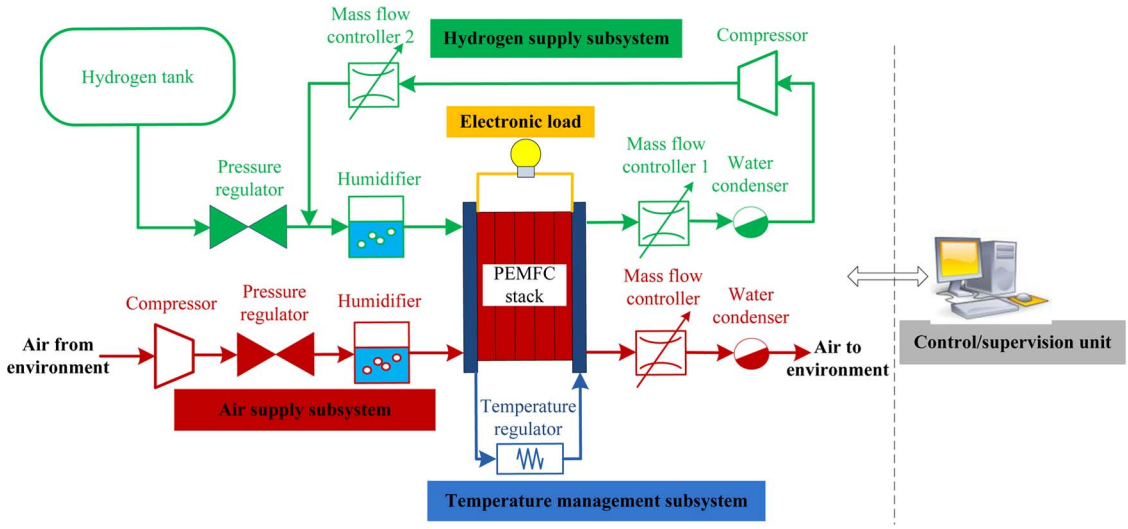


Fig. 4. Schematic of the used PEMFC system.

TABLE I
NOMINAL CONDITIONS OF THE STACK

Parameter	Value
Stoichiometry H_2	1.5
Stoichiometry Air	2
Pressure at H_2 inlet	150 kPa
Pressure at Air inlet	150 kPa
Temperature (exit of cooling circuit)	80 °C
Anode relative humidity	50%
Cathode relative humidity	50%
Current	110 A
Voltage per cell	0.7 V
Electrical power	3080 W

the experiments were drawn to construct the training data set and the test data set. The absolute and relative deviations of the cell voltage measurements are, respectively, 4 mV and 0.57%. For each state, data from one (or several) experiment(s) were used as training data, whereas data from the others were considered as test data.

The evolution of cell voltages in different states is shown in Fig. 5. It should be noted that those for Nl_2, Nl_3, and Nl_4 are not given here for they are visually little varied from that of Nl_1. Among them, F_1 and F_2 occurred in between the experiments, whereas Nl_1, F_3, and F_4 were maintained during the whole corresponding experiments. It can be found that the voltages of different cells have different magnitudes and distributions in different health states. In fact, different spatial distributions of temperature, humidity, and gas fluids led by different health states can result in different behaviors of the cell voltages. Essentially, this character is utilized for fault diagnosis.

V. RESULTS AND DISCUSSION

A. Multifault Detection and Isolation

First, the performance of multifault detection and isolation was investigated. Four faulty states (F_1, F_2, F_3, and F_4) and the normal state at time point 1 (Nl_1) were taken into consideration. After training FDA, the data of five states were projected onto 4-D *feature space*. Fig. 6 shows the first three features of the *projecting vectors*. Globally, the samples from different classes are isolated from the visual point of view.

Following the FDA training process, the SSM-SVM classifier was then trained in *feature space*. The training process could be considered as the initial training, which is offline.

With the trained FDA and SSM-SVM models, the diagnostic accuracy of the test data set was evaluated. The *confusion matrix*, which allows visualization of the classification performance, is shown in Table III. Each row of the table represents the classification distribution of the samples in an actual class. From the table, it can be seen that the diagnostic accuracy is more than 95% for each class, except for class F_4. In fact, F_4 is the lightest fault, which has some overlaps with the normal

TABLE II
CONCERNED STATES (CLASSES)

Health state description	Location	Notation	Sample number (for training)	Sample number (for test)
Nominal operating (time 1)	Whole system	Nl_1	800	2200
Nominal operating (time 2)	Whole system	Nl_2	400	400
Nominal operating (time 3)	Whole system	Nl_3	501	500
Nominal operating (time 4)	Whole system	Nl_4	1002	701
High current pulse or short circuit	Electric circuit	F_1	101	31
Stop cooling water	Temperature subsystem	F_2	21	302
High air stoichiometry (4)	Air supply subsystem	F_3	400	2500
Low air stoichiometry (1.4)	Air supply subsystem	F_4	400	2200

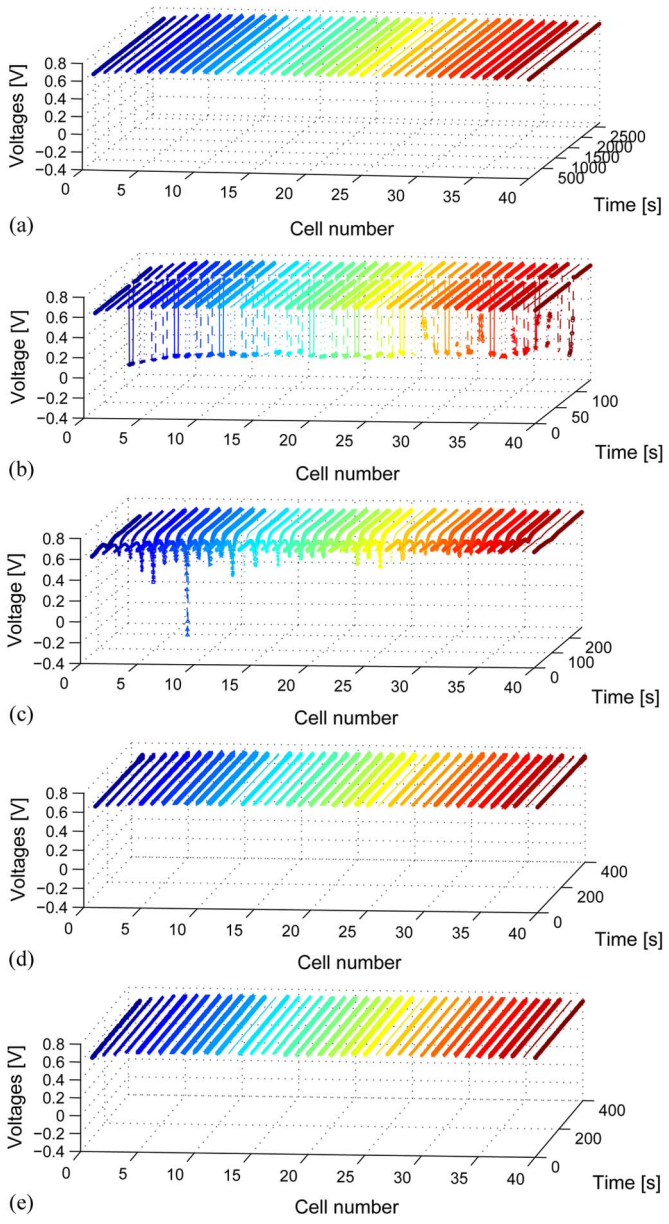


Fig. 5. Evolution of cell voltages in different states. (a) Nl_1. (b) F_1. (c) F_2. (d) F_3. (e) F_4.

state. The overlaps can also be observed in Fig. 6. Moreover, as the decision rule given by (10) was used, a tiny fraction of the samples was misclassified or misdiagnosed to the new fault class. Table IV gives the results without the new fault detection procedure. Comparing the two tables, it can be found that the

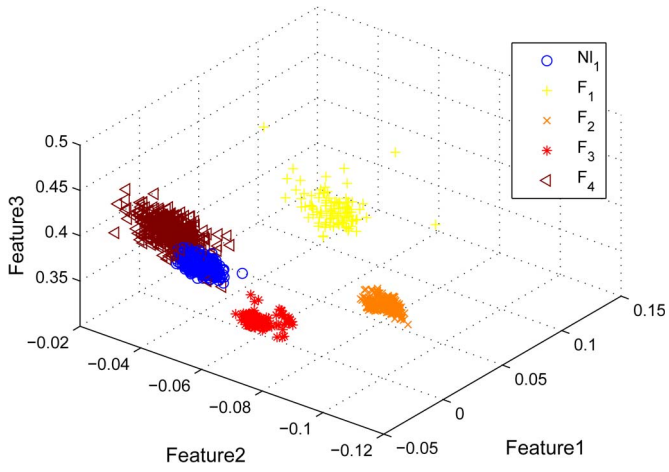


Fig. 6. First three features of the *projecting vectors* from five different health states.

TABLE III
CONFUSION MATRIX (%) WITH NEW FAULT DETECTION

Actual class	Diagnosed class					
	Nl_1	F_1	F_2	F_3	F_4	New fault
Nl_1	97.91	0	0	0.59	1.50	0
F_1	0	96.77	0	0	3.23	0
F_2	0	0	95.24	0	0	4.76
F_3	0	0	0	99.68	0	0.32
F_4	10.00	0	0	0.05	89.32	0.64

TABLE IV
CONFUSION MATRIX (%) WITHOUT NEW FAULT DETECTION

Actual class	Diagnosed class				
	Nl_1	F_1	F_2	F_3	F_4
Nl_1	97.91	0	0	0.59	1.50
F_1	0	96.77	0	0	3.23
F_2	0	0	95.24	0	4.76
F_3	0	0	0	99.68	0.32
F_4	10.00	0	0	0.05	89.95

samples that were misclassified to the new fault class were also mostly misclassified without the new fault detection procedure.

B. Online Adaptation

When the aging effect is taken into account, performance degradation arises over time. Fig. 7 shows the stack voltage sampled at four different time points. The variation of data sampled at each time is caused by system disturbance and

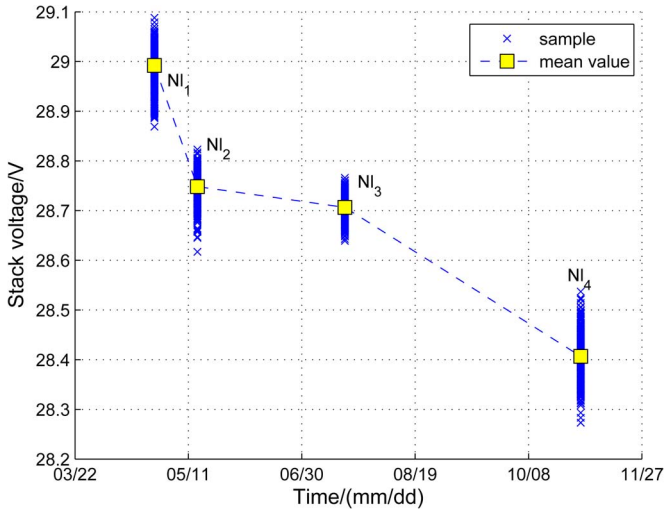


Fig. 7. Evolution of stack voltage over time.

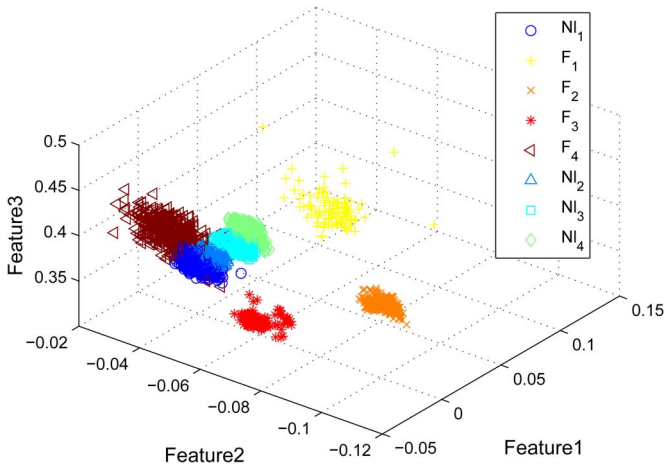


Fig. 8. First three features of the *projecting vectors* from eight different health states.

noise. Globally, a decrease can be observed over time. Since the operating frequency is not homogeneous, the descent speed of stack voltage is varying during the time.

The performance degradation due to the aging effect is usually considered as normal degradation, which is acceptable within certain limits. By using the trained FDA, the data from states Nl_2, Nl_3, and Nl_4 can be projected onto the *feature space*. Fig. 8 shows the first three features of the *projected vectors*. It could be seen that the data from normal classes shift over time. The diagnostic models should be updated to avoid misclassifying the data in normal functioning, but collected after normal aging into the classes representing faults.

Here, to test the efficiency of the updating procedure, the data from classes Nl_1, Nl_2, Nl_3, and Nl_4 were tested by using the diagnostic models that were updated at different times. $Model_1, Model_2, Model_3$, and $Model_4$ were the models trained or updated at time 1, time 2, time 3, and time 4. Each model was updated by using the previous model and the current samples. For instance, $Model_3$ was updated by using $Model_2$ and data from Nl_2.

TABLE V
CLASSIFICATION ACCURACY (%) OF NORMAL DATA SETS USING THE MODELS UPDATED AT DIFFERENT TIMES

Class	$Model_1$	$Model_2$	$Model_3$	$Model_4$
Nl_1	97.91			
Nl_2	65.50	99.75		
Nl_3	0	0.20	1	
Nl_4	0	0	36.03	1

TABLE VI
CONFUSION MATRIX (%) WITH F_1 AS NEW FAULT CLASS

Actual class	Diagnosed class				
	Nl	F_2	F_3	F_4	New fault
Nl	98.91	0	0	1.09	0
F_2	0	95.24	0	0	4.76
F_3	0	0	99.00	0	1
F_4	6.77	0	0	93.23	0
New fault	3.23	0	0	0	96.77

TABLE VII
CONFUSION MATRIX (%) WITH F_2 AS NEW FAULT CLASS

Actual class	Diagnosed class				
	Nl	F_1	F_3	F_4	New fault
Nl	94.95	0	0	4.32	0.73
F_1	3.32	83.87	0	0	12.9
F_3	0	0	99.00	0	1.00
F_4	8.55	0	0	90.36	1.09
New fault	0	0	0	0	1

TABLE VIII
CONFUSION MATRIX (%) WITH F_3 AS NEW FAULT CLASS

Actual class	Diagnosed class				
	Nl	F_1	F_2	F_4	New fault
Nl	94.68	0	0	4.68	0.64
F_1	3.23	90.32	0	0	6.45
F_2	0	0	95.24	0	4.76
F_4	9.73	0	0	89.45	0.82
New fault	0.12	0	0	4.52	95.36

The test results are summarized in Table V. It can be seen that the diagnostic accuracy is low without an updating procedure at each time. On the contrary, with an updating procedure, the diagnosis accuracy could be significantly improved. Hence, the updating procedure is therefore justified to be useful and efficient.

C. Detection of a Novel Failure Mode

Here, classes Nl_R $(R = 1, \ldots, 4)$ are combined as a unique class denoted by Nl. To test the performance of the strategy proposed for detecting a novel failure mode, we proceed as follows. Let $j \in \{1, \ldots, 4\}$. Let us assume that the fault represented by class F_j was initially an unknown fault. Hence, the initial step dedicated to training was realized with the data that were representative of classes Nl and F_i with $i \in \{1, \ldots, 4\} - \{j\}$. After that, the data from various classes including those used in the training phase and the unknown class F_j were treated. Tables VI–IX show the confusion matrices for the different values of j.

TABLE IX
CONFUSION MATRIX (%) WITH F_4 AS NEW FAULT CLASS

Actual class	Diagnosed class				
	Nl	F_1	F_2	F_3	New fault
Nl	99.59	0	0	0	0.41
F_1	3.23	96.77	0	0	0
F_2	0	0	95.24	0	4.76
F_3	0	0	0	99.16	0.84
New fault	46.91	13.23	0	0	39.86

TABLE X
OCCUPIED MEMORY AND COMPUTING TIME

	$Model_1$	$Model_2$	$Model_3$	$Model_4$
Occupied memory	30.4 kb	125 kb	125 kb	125 kb
Performing time	0.49 ms	0.53 ms	0.56 ms	0.53 ms
Updating time		2.93 ms	3.98 ms	5.32 ms

For all the cases, the probabilities that the data located in the known classes were misclassified to the novel classes are generally low. It should be noted that for the cases where F_1, F_2, and F_3 were considered as novel classes, the probabilities of detection of the novel class were more than 95%. However, for the case where F_4 was considered as a novel class, the probability was only equal to 39.86%, which is a low level. It can be deduced that it is relatively difficult to recognize the data in the novel class when they are too close to the known classes.

D. Real-Time Capability

To implement the proposed approach online in an embedded system, the computational cost should be carefully evaluated. In this paper, the computational costs of the two online procedures (performing process and updating process), were evaluated from the perspectives of occupied memory and computing time. The tests were carried out under a 64-bit MATLAB 2010b environment with a 2.7-GHz CPU and 8-G RAM. The results are summarized in Table X ($Model_1$ is the initial model that was trained offline).

It can be found that the updating time is longer than the performing time and is thus the main part of the computing burden. To our knowledge, the diagnostic period of 1 s can satisfy the requirements of diagnosis for most PEMFC systems. Moreover, the memory used by most embedded systems can achieve the storage task easily. Hence, the proposed strategy is suitable for online implementation.

E. Discussion

In practice, the diagnosis of the other faults different from those studied can also be relevant, for instance, the faults related to water management and hydrogen supply. If only the cell voltage magnitudes and/or distributions caused by a fault are different from those of the normal state and other fault states, the fault is detectable and isolable by using the proposed strategy. Hence, it is feasible to extend the approach to the diagnosis of other fault types.

The proposed diagnosis approach is dependent on data. If we want to diagnose a specific fault, the data in the concerned fault mode must be prepared. This could be the weak point of the proposed approach. In practice, the diagnosis of several most common faults could be initially considered, such that the system failure rate could be significantly lowered through successful diagnosis of these faults. Then, the diagnosis of new faults could be added to the existing approach when the corresponding data are collected.

VI. CONCLUSION

This paper has presented a novel data-driven diagnostic strategy for PEMFC systems. The FDA and SSM-SVM methods were used successively to extract the features from individual cell voltages and to classify the extracted features into different classes corresponding to the known health states and the potential novel failure mode. By the incremental learning of SSM-SVM, the online adaptation of the diagnostic approach is realized.

The test results for a 40-cell PEMFC stack show that different faults can be detected and isolated with high accuracy, and the data from the potential novel failure modes can be recognized in most cases. Using an online adaptation procedure, the diagnostic approach can be adapted over the operating time, and the diagnostic performance can be maintained. Moreover, the computational cost is justified to be suitable for online implementation.

APPENDIX
PROOF OF (24)

According to (23) and (7), the left-hand side of (24) can be expressed as

$$\left|{d_n^i}^2 - R_i^2\right| = \left|-2\Phi(\boldsymbol{z}_n)\boldsymbol{a}_i + 2\Phi(\boldsymbol{z}_l)\boldsymbol{a}_i\right| \text{ for } l \in \mathcal{T}_M^i. \quad (26)$$

Substituting (6) and taking into account (13), we obtain

$$\Phi(\boldsymbol{z}_l)\boldsymbol{a}_i = \Phi(\boldsymbol{z}_l)\sum_{m\in\mathcal{T}} c_m^i\alpha_m^i\Phi(\boldsymbol{z}_m) = \frac{\sum_{m\in\mathcal{T}} Q_{m,l}^i}{c_l^i} = \frac{g_l^i - c_l^i b_i}{c_l^i} = \frac{0 - c_l^i b_i}{c_l^i} = -b_i. \quad (27)$$

Substituting (27) and (6) and taking into account (13), the right-hand side of (26) can be expressed by

$$\left|-2\Phi(\boldsymbol{z}_n)\sum_{m\in\mathcal{T}} c_m^i\alpha_m^i\Phi(\boldsymbol{z}_m) - 2b_i\right| = 2\left|\sum_{m\in\mathcal{T}} Q_{m,n} + c_n^i b_i\right| = 2\left|g_n^i\right| = 2g_n^i \quad \text{for } n \in \mathcal{T}_R^i. \quad (28)$$

ACKNOWLEDGMENT

The authors would like to thank the partners for providing experimental data.

REFERENCES

[1] C. de Beer, P. Barendse, P. Pillay, B. Bullecks, and R. Rengaswamy, "Classification of high temperature PEM fuel cell degradation mech-anisms using equivalent circuits," *IEEE Trans. Ind. Electron.*, to be published, DOI: 10.1109/TIE.2015.2393292.

[2] S. de Lira, V. Puig, J. Quevedo, and A. Husar, "LPV observer design for PEM fuel cell system: Application to fault detection," *J. Power Sources*, vol. 196, no. 9, pp. 4298–4305, May 2011.

[3] D. Hissel, D. Candusso, and F. Harel, "Fuzzy-clustering durability diagnosis of polymer electrolyte fuel cells dedicated to transportation applications," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2414–2420, Sep. 2007.

[4] N. Y. Steiner, D. Hissel, P. Mocoteguy, and D. Candusso, "Diagnosis of polymer electrolyte fuel cells failure modes (flooding and drying out) by neural networks modeling," *Int. J. Hydrogen Energy*, vol. 36, no. 4, pp. 3067–3075, Feb. 2011.

[5] D. Benouioua, D. Candusso, F. Harel, and L. Oukhellou, "Fuel cell diag-nosis method based on multifractal analysis of stack voltage signal," *Int. J. Hydrogen Energy*, vol. 39, no. 5, pp. 2236–2245, Feb. 2014.

[6] J. Hua, J. Li, M. Ouyang, L. Lu, and L. Xu, "Proton exchange membrane fuel cell system diagnosis based on the multivariate statistical method," *Int. J. Hydrogen Energy*, vol. 36, no. 16, pp. 9896–9905, Aug. 2011.

[7] C.-M. Vong, P.-K. Wong, and W.-F. Ip, "A new framework of simultaneous-fault diagnosis using pairwise probabilistic multi-label classification for time-dependent patterns," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3372–3385, Aug. 2013.

[8] T. Boukra, A. Lebaroud, and G. Clerc, "Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4034–4042, Sep. 2013.

[9] M. Biet, "Rotor faults diagnosis using feature selection and nearest neigh-bors rule: Application to a turbogenerator," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4063–4073, Sep. 2013.

[10] M. Prieto, G. Cirrincione, A. Espinosa, J. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.

[11] X. Gong and W. Qiao, "Bearing fault diagnosis for direct-drive wind turbines via current-demodulated signals," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3419–3428, Aug. 2013.

[12] Z. Li, S. Giurgea, R. Outbib, and D. Hissel, "Online diagnosis of PEMFC by combining support vector machine and fluidic model," *Fuel Cells*, vol. 14, no. 13, pp. 448–456, Jun. 2014.

[13] Z. Li, R. Outbib, D. Hissel, and S. Giurgea, "Data-driven diagnosis of PEM fuel cell: A comparative study," *Control Eng. Practice*, vol. 28, pp. 1–12, Jul. 2014.

[14] J. Wu *et al.*, "A review of PEM fuel cell durability: Degradation mech-anisms and mitigation strategies," *J. Power Sources*, vol. 184, no. 1, pp. 104–119, 2008.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ, USA: Wiley, 2001.

[16] L. Cao, K. Chua, W. Chong, H. Lee, and Q. Gu, "A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine," *Neurocomputing*, vol. 55, no. 1, pp. 321–336, 2003.

[17] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[18] P. Y. Hao and Y. H. Lin, "A new multi-class support vector machine with multi-sphere in the feature space," in *Proc. 20th IEA/AIE*, 2007, pp. 756–765.

[19] S. Nikitidis, N. Nikolaidis, and I. Pitas, "Multiplicative update rules for incremental training of multiclass support vector machines," *Pattern Recognit.*, vol. 45, no. 5, pp. 1838–1852, May 2012.

[20] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in *Proc. Adv. NIPS*, vol. 13, 2001, pp. 409–415, MIT Press.

[21] D. Candusso *et al.*, "Characterisation and modeling of a 5 kW PEMFC for transportation applications," *Int. J. Hydrogen Energy*, vol. 31, no. 8, pp. 1019–1030, 2006.