Control Engineering Practice 28 (2014) 1–12

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Data-driven diagnosis of PEM fuel cell: A comparative study

Zhongliang Li [a,b,c,*], Rachid Outbib [a], Daniel Hissel [b,c], Stefan Giurgea [b,c]

[a] Laboratoire des Sciences de l'Information et des Systemes (LSIS), University of Aix-Marseille, France
[b] FEMTO-ST (UMR CNRS 6174), ENERGY Department, University of Franche-Comte, France
[c] FCLAB (Fuel Cell Lab) Research Federation, FR CNRS 3539, rue Thierry Mieg, 90010 Belfort Cedex, France

ARTICLE INFO

Article history:
Received 19 June 2013
Accepted 25 February 2014
Available online 27 March 2014

Keywords:
Fault diagnosis
PEMFC
Water management
Classification
Feature extraction

ABSTRACT

This paper is dedicated to data-driven diagnosis for Polymer Electrolyte Membrane Fuel Cell (PEMFC). More precisely, it deals with water related faults (flooding and membrane drying) by using pattern classification methodologies. Firstly, a method based on physical considerations is defined to label the training data. Secondly, a feature extraction procedure is carried out to pick up the significant features from vectors constructed by individual cell voltages. Finally, a classification is adopted in the feature space to realize the fault diagnosis. Various feature extraction and classification methodologies are employed on a 20-cell PEMFC stack. The performances of these methodologies are compared.

1. Introduction

Polymer Electrolyte Membrane Fuel Cell (PEMFC) is a promising alternative power generator, thanks to its high power density, high efficiency and environmental friendly property. However, reliability and durability are still two barriers which block its wide application. Thus, fault diagnosis is an efficient solution to guarantee the safe operation of the fuel cell. Indeed, more serious faults can be avoided thanks to an inchoate fault diagnosis. Furthermore, the diagnosis results can be supplied to the control unit, thus helping adjust the control commands to make the fuel cell operate efficiently and safely.

The faults that can occur on PEMFC are of two types: first, the irreversible faults that lead to degeneration of the system (for instance, membrane tearing). Second, those that can be corrected and supervised. Among the important faults are those related to water (i.e. flooding and membrane drying). More precisely, these faults have been considered as a major cause of power decay and consequently have drawn considerable attention over the last few decades. Hence, the problem of PEMFC has received intensive studies and several strategies have been proposed.

Among the proposed strategies, a first category is obtained using the analytic models (see for instance Escobet et al., 2009; Hernandez, Hissel, & Outbib, 2010). This approach consists in analyzing the residuals that are obtained by comparing measured inputs and outputs using analytical relationships.

A second category is based on expert knowledge, mainly, fuzzy logic (see Hissel, Péra, & Kauffmann, 2004), Bayesian networks (see Riascos, Simoes, & Miyagi, 2007; Wasterlain, Candusso, Harel, François, & Hissel, 2010), and neural networks (see Yousfi-Steiner, Hissel, Moçotéguy, & Candusso, 2011a and references therein) methodologies.

A third category contains the strategies that are achieved using signal processing. Electrochemical Impedance Spectroscopy (EIS) is considered as a powerful tool for analyzing the behavior of fuel cells and some results were obtained by using the fact that the impedance of fuel cell stacks in certain frequencies is sensitive to faults (Asghari, Mokmeli, & Samavati, 2010; Mérida, Harrington, Canut, & McLean, 2006). Besides, and recently, new results that direct voltage signal analysis thanks to wavelet have been proposed (see Yousfi-Steiner, Hissel, Moçotéguy, & Candusso, 2011b).

Other approaches have also been proposed, for instance those using some novel micro-sensors (Lee & Lee, 2012) or those using multivariate statistical methods (see for instance Hua, Lu, Ouyang, Li, & Xu, 2011).

Generally speaking, and independent of PEMFC, data-driven diagnosis has attracted the interest of many authors. This approach offers the advantage of being relevant for industrial applications. In this framework, various pattern classification techniques have attracted more and more attention in the diagnosis domain (Abbasion, Rafsanjani, Farshidianfar, & Irani, 2007; Li & Xiao, 2012). The common classification procedure usually proceeds in two steps. Firstly, an empirical classifier is established from prior

* Corresponding author at: Laboratoire des Sciences de l'Information et des Systemes (LSIS), University of Aix-Marseille, France. Tel.: +33 3 84 58 36 28; fax: +33 3 84 58 36 36.
E-mail addresses: zhongliang.li@lsis.org (Z. Li), rachid.outbib@lsis.org (R. Outbib), daniel.hissel@univ-fcomte.fr (D. Hissel), stefan.giurgea@utbm.fr (S. Giurgea).

http://dx.doi.org/10.1016/j.conengprac.2014.02.019
0967-0661

knowledge and history data, which is considered as a training process. Secondly, the real-time data are processed by the classifier in order to determine whether and which faults occur. Notice that, in addition to the classification, some feature extraction procedures are usually carried out as a previous step to get relevant features from the raw data (Cao, Chua, Chong, Lee, & Gu, 2003).

This paper is a contribution to the PEMFC diagnosis problem. The proposed approach is data-driven and combines feature extraction and classification. In our strategy, features are firstly extracted from data related to the individual cell voltages. Then, a classifier is used in order to distinguish the features of the possible states (i.e. "normal state", "flooding state" and "membrane drying state"). To do so, we consider for automatic feature extraction four methodologies that are Principal Component Analysis (PCA), Fisher Discrimination Analysis (FDA), Kernel Principal Component Analysis (KPCA), Kernel Fisher Discrimination Analysis (KFDA). In addition, we employ three typical classification methodologies, which are Gaussian Mixture Model (GMM), k-Nearest Neighbor (kNN), Support Vector Machine (SVM). Our aim is to compare the performances of these methodologies, from the point of view of diagnosis precision and computation cost, so as to get a relevant tool for online diagnosis of a PEMFC stack.

The paper is organized as follows. In Section 2, the fuel cell system concerned and the faults are presented. Section 3 is devoted to present methodologies used in this work, including methods for labeling training data, feature extraction, and classification. In Section 4, diagnosis results from the concerned PEMFC system are given. The performances of different methods are also compared and discussed in this section. Finally, Section 5 gives a conclusion.

2. PEMFC operation

2.1. PEMFC system

A running PEMFC is usually fed continuously with hydrogen on the anode side and with air on the cathode side. On the anode side, hydrogen is oxidized:

$$H_2 \rightarrow 2H^+ + 2e^-$$

With the protons transferred through the membrane and the electrons through the external circuit, the oxygen is reduced on the cathode side:

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$$

With the conversion of chemical energy to electrical energy, the by-product water is generated and expelled mostly with the unreacted air from the cathode side.

To produce a useful voltage or power, many cells have to be connected in series, which functions as a fuel cell stack. In addition to the stack, practical fuel cell systems also contain several other ancillary subsystems: hydrogen supply subsystem, air supply subsystem, temperature management subsystem. As shown in Fig. 1, in the hydrogen supply subsystem and air supply subsystem, gas flows and pressures at stack inlets can be regulated respectively by using mass flow controllers and back-pressure valves. The relative humidity of the input air is regulated by a humidifier. The temperature management subsystem can make the fuel cell stack to operate at an appropriate temperature.

Various faults may occur in the fuel cell system. Faults may happen either inside the fuel cell stack or in the ancillary subsystems. The fuel cell stack is the heart of the whole system and the faults in the fuel cell stack can somehow reflect the faults in the ancillary components. Consequently, the diagnosis, especially the online diagnosis of the fuel cell stack is crucial and thus the focus of this paper.

2.2. Concerned faults

Water management has been considered as one of the most important issues in PEMFC. Flooding and membrane drying are the two main degradation mechanisms that occur when water management is not adequate (see for instance Yousfi-Steiner et al., 2008). In this work, the diagnosis of the two typical faults is investigated through data-driven relevant methodologies.

As shown in Fig. 2, a typical PEMFC consists of bipolar plates (BPs), gas diffusion layers (GDLs), catalyst layers (CLs), and a membrane. On both sides of BPs, gas channels (GCs) are grooved for gas flow. In a proper functioning PEMFC, the membrane should keep a certain water content to make the protons transfer through it with low ohmic resistance. Hence, air is usually humidified before being fed into the fuel cell. At the same time, liquid water is generated in the cathode and expelled from the fuel cell by the flow of the unreacted air and hydrogen or purged at regular intervals. Inside the fuel cell, water travels through different layers and moves between anode and cathode. The water content in the

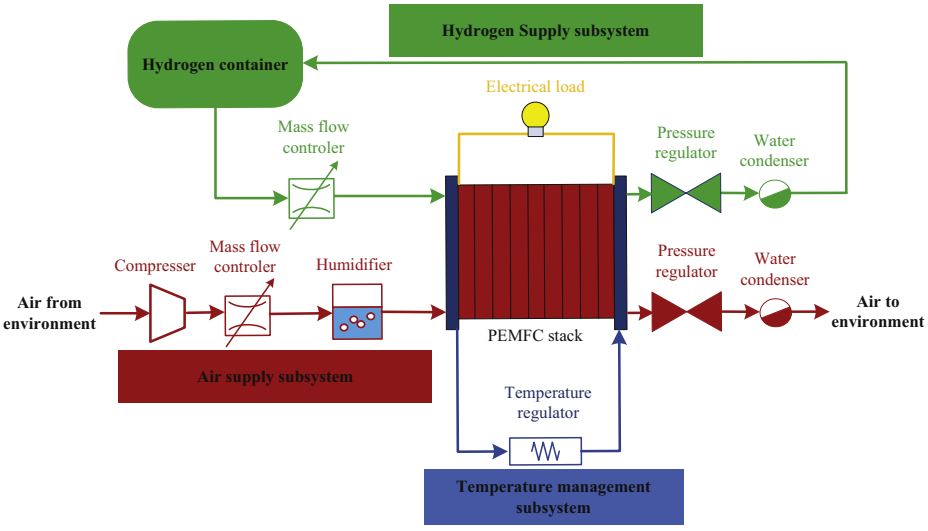

Fig. 1. The schematic of PEMFC system.

channels and different layers can be influenced by a number of factors, such as air humidity and flow rate, stack temperature, and load current. Inadequate water management may cause the degradation of the PEMFC. On the one hand, a dry membrane increases the ohmic losses, thus induces a "membrane drying" fault. On the other hand, the accumulation of liquid water in the GCs and/or gas porosities of GDLs and CLs results in a "flooding" fault. Excessive liquid water will block the reactant pathways, thus making the fuel cell stack degraded. As water is generated and expelled mostly at the cathode side, flooding happens generally at the cathode side (Yousfi-Steiner et al., 2008).

3. Descriptions of diagnosis methodologies

3.1. Approach principle

The approach proposed in this paper is based on experimental data and contains three steps (see Fig. 3): data labeling process, model training process, and diagnosis process. The first two processes are off-line, while the third one is on-line.

The pattern classification methods used belong to supervised ones. In supervised learning, the samples for training should be provided with their category labels before the training procedure. Consequently, it is necessary to define the classes and label the training data. In the labeling stage, a two-phase pressure drop model combined with statistical analysis is used to achieve this (Ito et al., 2008; Wang, Basu, & Wang, 2008). After the data labeling process, the training data can be labeled to three classes: normal state, flooding fault state, and membrane drying state.

In the model training and diagnosis stages, the individual cell voltages are the used variables. Cell voltages are chosen here on account of several points: first, fuel cell voltage signals are highly dependent on the current, the electrochemical characteristics, the temperature, and the aging effect. The voltage variances can be used to determine the magnitude of the parameters for the fuel cell model (Kim, Lee, Tak, & Cho, 2012). In other words, the individual cell voltage can be seen as sensors inside the fuel cell stack. Second, the importance of monitoring individual cell voltages is stressed. Since the cell with the lowest cell voltage in the stack restricts the maximum power output of the stack (Rodatz, Büchi, Onder, & Guzzella, 2004), it is necessary to monitor every single cell voltage (or several neighboring cell voltages together) to some degree. Third, it is observed that the water management faults change the flow distributions of the gases and then make the distribution cell voltages vary (Hernandez, Hissel, & Outbib, 2006). Moreover, the cost of measurement of cell voltages is relatively low. Based on the above points, the vectors constructed by individual cell voltages are considered as the original data for diagnosis.

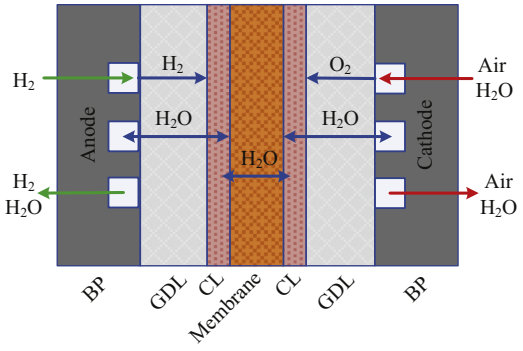

Fig. 2. Schematic picture of water movement inside a PEMFC.

In the model training process, two kinds of model, which are feature extraction and classification, need to be trained. In feature extraction, the goal is to find certain projecting vectors to map the original high-dimensional vectors to feature space, which is of low-dimension. Notice that in the classification, the classifiers will be trained in the feature space. By the trained classifiers, a new point in feature space can be fixed to one of the three predefined classes. In the literature, there are various kinds of methodologies for feature extraction and classification. Hence, this paper presents several relevant and representative methodologies that can be used to compare their performances on the PEMFC diagnosis.

In the diagnosis process, the real-time cell voltages are sampled and represented by vectors. Then, the feature extraction and classification procedures are respectively carried out by using the models obtained in the training process. The real-time data can thus be classified into three classes, and the diagnosis is realized accordingly after these two procedures.

3.2. Data labeling

In order to label the training data to the three classes: "normal", "flooding", and "membrane drying", the normal range of liquid water inside the fuel cell must be evaluated. To do so, an approach that combines pressure drop model with statistical analysis is proposed.

3.2.1. Pressure drop model

The pressure drop between the inlet and outlet channels is significant of the gases removal out of the fuel cell, and it is relevant to the content of liquid water in the flow fields (Yousfi-Steiner et al., 2008). Since only the cathode inlet gas is humidified in the PEMFC system studied, and the water generated is mostly expelled from the cathode side, the cathode side is more relevant to the water management issues. Hence, the pressure drop model at the cathode side is considered.

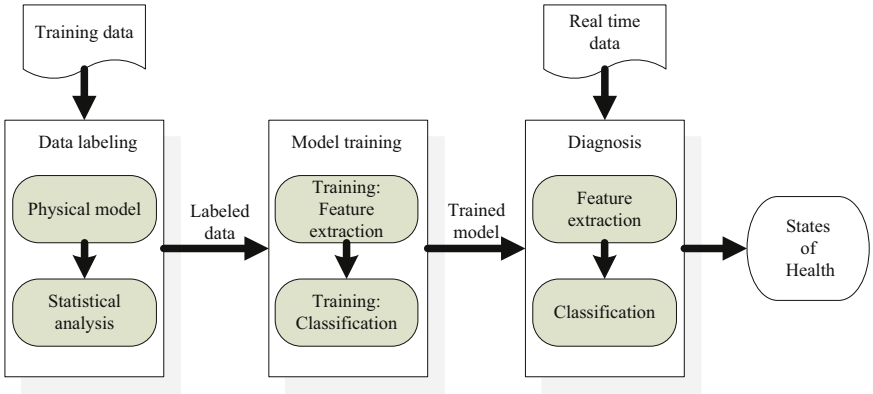

Fig. 3. The framework of the diagnosis approach.

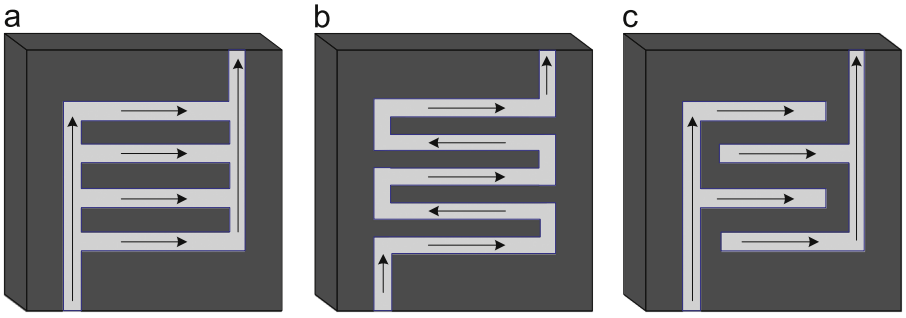

Fig. 4. Three kinds of flow field structures. (a) Parallel flow field. (b) Serpentine flow field. (c) Interdigitated flow field.

GCs are grooved on both sides of BPs for gas flow and different structures can be used. Fig. 4 depicts three classic structures: parallel, serpentine and interdigitated flow fields. Notice that the pressure drop model is dependent on the structures considered. For the parallel and serpentine flow fields, the air passes from the GCs, and the pressure drops throughout the GCs. The major part of pressure loss is associated with the frictional losses along the channel pipe (Tirnovan & Giurgea, 2012). In this case, the pressure drop model based on the Darcy law is given by (Wang et al., 2008):

$$\Delta P = \frac{\mu}{K_{C_0}(1-s)^{n_k}} v_D L_{GC} \tag{1}$$

where v_D is the air flow rate, K_{C_0} is the permeability, which is impacted by the sizes and the structures of the flow fields. μ denotes the viscosity of air and L_{GC} is the length of the channel. $s \in [0,1)$ is defined as the volume fraction of GC occupied by liquid water, which is a key parameter characterizing the water quantity. n_k is a constant ranging between 4.5 and 5.0 (see Wang et al., 2008).

For the interdigitated flow field, the pressure drop mainly occurs in the GDL. In this case, the pressure drop can be denoted as given in (Ito et al., 2008):

$$\Delta P = \frac{150(1-\varepsilon(1-s))^2 \mu}{\varepsilon^3 (1-s)^3 d_0^2} v_D L_{GDL} \tag{2}$$

where d_0 is the representative pore diameter of GDL, L_{GDL} is the rib length of the BP. ε is a constant that reflects the porosity of GDL. $s \in [0,1)$ quantifies the portion of pores (in GDL) occupied by liquid water. Undoubtedly, in (1) and (2), s can be considered as a criterion to quantify the water content inside the fuel cells.

3.2.2. A statistical method

From (1) and (2), the quantity $\Delta P/v_D$ can be considered as a function s:

$$W(s) = \frac{\Delta P}{v_D}$$

where

$$W^{(1)}(s) = \frac{\alpha}{(1-s)^{n_k}} \quad \text{and} \quad W^{(2)}(s) = \beta \frac{(1-\varepsilon(1-s))^2}{(1-s)^3} \tag{3}$$

with

$$\alpha = \frac{\mu L_{GC}}{K_0} \quad \text{and} \quad \beta = 150 \frac{\mu L_{GDL}}{\varepsilon^3 d_0^2}$$

Clearly, $W^{(1)}$ and $W^{(2)}$ defined by (3) are increasing functions for $s \in [0,1)$. Thus, W can replace s to express water quantity. In a normal state, it is considered that the fuel cell can operate in a range of s, so the values of W also distribute in a normal range. If the values of W follow a normal distribution, which is the case in our study (see Section 5), a common statistical method "3-sigma"

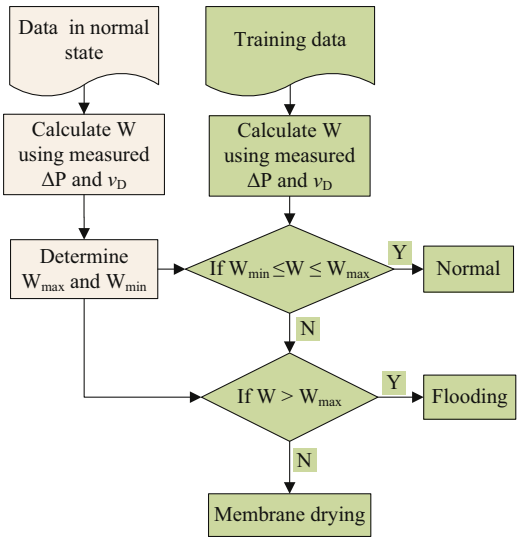

Fig. 5. Training data labeling process.

can be used in order to evaluate the limits of W:

$$W_{max} = \overline{W} + 3\sigma_W, \quad W_{min} = \overline{W} - 3\sigma_W \tag{4}$$

where $\overline{W}$ and σ_W are respectively the average value and the standard deviation of W samples in the normal state.

In the labeling process, W_{max} is the threshold for flooding diagnosis while W_{min} is the one for drying fault diagnosis. The flow chart of the training data labeling procedure can be depicted as in Fig. 5.

3.3. Feature extraction

For high power applications, for instance vehicles, a large number of fuel cells are usually stacked in series to meet the power requirement. Hence, large dimension data has to be handled when individual cell voltages serve as the variables for diagnosis. In order to reduce the complexities of computations, it is necessary to lower the data dimension by some means of feature extraction. At the same time, the feature extraction procedure is motivated to draw useful features for diagnosis. Based on the above considerations, four representative feature extraction methodologies, which can meet these two needs, are presented in this subsection. More precisely, two typical unsupervised and supervised methodologies: PCA, FDA, and their nonlinear forms KPCA and KFDA, are considered.

The feature extraction problem can be described as follows: collect N training samples $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N \in \mathbb{R}^M$, which are distributed in C classes denoted by V_1, V_2,…,V_C. Sample $\boldsymbol{v}_n$ is a vector constructed by M cell voltages (cell number is M):

$$\boldsymbol{v}_n = [v_{n1}, v_{n2}, \ldots, v_{nM}]^T \tag{5}$$

The sample number in V_i is N_i, which satisfies $\sum_{i=1}^{C} N_i = N$. The class index of $\boldsymbol{v}_n$ is denoted by h_n, $h_n \in \{1, 2, \ldots, C\}$. The objective of the training process is to find L ($L < M$) M-dimension unit projecting vectors: $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_L\}$. With these vectors, a real-time sample $\boldsymbol{v}$ can be projected to an L-dimension feature space, the projected vector $\boldsymbol{z}$ is expressed:

$$\boldsymbol{z} = [\boldsymbol{w}_1^T\boldsymbol{v}, \boldsymbol{w}_2^T\boldsymbol{v}, \ldots, \boldsymbol{w}_L^T\boldsymbol{v}]^T \tag{6}$$

3.3.1. PCA

PCA is an unsupervised dimensionality reduction and feature extraction technique that preserves the significant variability information in the original data set. It changes more relevant variables into seldom uncorrelated variables according to the lowest data missing rule (Hua et al., 2011). The PCA procedure used in this study can be summarized by Algorithm 1.

Algorithm 1. PCA

Training:

1: Collect samples $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$.

2: Perform singular value decomposition on covariance matrix: $\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{v}_n - \overline{\boldsymbol{v}})(\boldsymbol{v}_n - \overline{\boldsymbol{v}})^T = \boldsymbol{P}\Lambda\boldsymbol{P}^T$ (7)

where $\overline{\boldsymbol{v}} = \sum_{n=1}^{N}\boldsymbol{v}_n/N$, $\boldsymbol{P} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M]$, $\Lambda = diag(\lambda_1, \ldots, \lambda_M)$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$.

3: Determine the number of principal components L by respecting: $\frac{\sum_{i=1}^{L}\lambda_i}{\sum_{i=1}^{M}\lambda_i} \geq Th, \frac{\sum_{i=1}^{L-1}\lambda_i}{\sum_{i=1}^{M}\lambda_i} < Th$ (8)

where Th is a pre-set threshold, whose value is near but less than 1 (here is set at 0.9).

4: Save vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L$.

Performing:

Calculate the projected vector of a new sample $\boldsymbol{v}$ as (6).

3.3.2. FDA

FDA is a supervised technique developed to reduce the dimensions of the data in the hope of obtaining a more manageable classification problem. The objective of FDA is to find the mapping vectors that make the data in the same class concentrated while the data in different classes separated (Bishop, 2006; Duda, Hart, & Stork, 2001). The FDA procedure in this paper can be briefly formulated by Algorithm 2.

Algorithm 2. FDA

Training:

1: Collect labeled samples: $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$.

2: Calculate within-class-scatter matrix $\boldsymbol{S}_w$ and between-class-scatter matrix

$$\boldsymbol{S}_b.\boldsymbol{S}_w = \sum_{i=1}^{C}\sum_{\boldsymbol{v}_n \in V_i}(\boldsymbol{v}_n - \overline{\boldsymbol{v}}_i)(\boldsymbol{v}_n - \overline{\boldsymbol{v}}_i)^T$$

$$\boldsymbol{S}_b = \sum_{i=1}^{C} N_i(\overline{\boldsymbol{v}}_i - \overline{\boldsymbol{v}})(\overline{\boldsymbol{v}}_i - \overline{\boldsymbol{v}})^T$$

where $\overline{\boldsymbol{v}} = \sum_{n=1}^{N}\boldsymbol{v}_n/N$, and $\overline{\boldsymbol{v}}_i = \sum_{\boldsymbol{v}_n \in V_i}\boldsymbol{v}_n/N_i$.

3: Set $L = C - 1$, and find the L eigenvectors of $\boldsymbol{S}_w^{-1}\boldsymbol{S}_b$ with non-zero eigenvalues: $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_L$.

Performing:

Calculate the projected vector of a new sample as (6).

3.3.3. KPCA

KPCA is an extension of PCA, which aims to solve nonlinear PCA. The key idea of KPCA is intuitive and generic. In general, the nonlinear correlated data can always be mapped to a higher-dimensional space in which they vary linearly via nonlinear mapping (Cover, 1965). After that, the PCA procedure can be carried out in the new space. Actually, this two-step process can be realized by introducing kernel functions and playing "kernel trick" (Campbell, 2002). The KPCA implemented in our study can be summarized by algorithm 3.

Algorithm 3. KPCA

Training:

1: Collect $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$.

2: Get $\boldsymbol{K} \in \mathbb{R}^{N\times N} : \boldsymbol{K}_{ij} = k(\boldsymbol{v_i}, \boldsymbol{v_j})$, where $k(\boldsymbol{v_i}, \boldsymbol{v_j})$ is a kernal function.

3: Modify $\boldsymbol{K}\tilde{\boldsymbol{K}} = \boldsymbol{K} - \boldsymbol{1}_N\boldsymbol{K} - \boldsymbol{K}\boldsymbol{1}_N + \boldsymbol{1}_N\boldsymbol{K}\boldsymbol{1}_N$ (9)

where $\boldsymbol{1}_N \in \mathbb{R}^{N\times N}$ with terms all equal to $1/N$.

4: Find L eigenvectors of $\tilde{\boldsymbol{K}}$ with the largest eigenvalues, which are denoted as $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_L \in \mathbb{R}^N$.

Performing:

Calculate the projected vector of a new sample $\boldsymbol{v}$. The lth ($l \in 1, 2, \ldots, L$) element of the projected vector $\boldsymbol{y}$ can be calculated as $z_l = \sum_{n=1}^{N}\alpha_{ln}k(\boldsymbol{v}_n, \boldsymbol{v})$ (10)

where α_{kn} is the n^{th} element of eigenvector $\boldsymbol{\alpha}_l$.

The kernel functions correspond to nonlinear mappings. In various kernel functions, the Gaussian kernel function is usually the first choice because of its high performance in most cases (see for instance Wu & Wang, 2009). Hence, this popular kernel function is involved in this paper:

$$k(\boldsymbol{v}_i, \boldsymbol{v}_j) = \exp\left(-\frac{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|^2}{\sigma}\right) \tag{11}$$

where σ is a constant that needs to be initialized.

3.3.4. KFDA

Like KPCA, the key idea of KFDA is also to map the data to a new space by nonlinear mapping first, and then carry out the FDA procedure in the new space. Kernel trick helps to realize the KFDA process in the same way as KPCA (Baudat & Anouar, 2000).

Algorithm 4. KFDA

Training:

1: Collect labeled samples: $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$.

2: Get kernel matrix $\boldsymbol{K}$.

3: Modify $\boldsymbol{K}$ as (9)

4: Get matrix $\boldsymbol{U}$:

$\boldsymbol{U} = diag(\boldsymbol{U}_1, \boldsymbol{U}_2 \ldots, \boldsymbol{U}_C)$ (12)

where $\boldsymbol{U}_i \in \mathbb{R}^{N_i\times N_i}$ with terms all equal to $1/N_i$.

5: Find L eigenvectors of $(\tilde{\boldsymbol{K}}\tilde{\boldsymbol{K}})^{-1}\tilde{\boldsymbol{K}}\boldsymbol{U}\tilde{\boldsymbol{K}}$ with the largest eigenvalues, which are denoted by $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_L \in \mathbb{R}^N$

Performing:

Calculate the projected vector of a new sample $\boldsymbol{v}$. The lth ($l \in 1, 2, \ldots, L$) element of the projected vector $\boldsymbol{z}$ can be calculated as (10).

3.3.5. Remarks on feature extraction methods

1. It was verified in Yang, Jin, yu Yang, Zhang, and Frangi (2004) that KFDA is equivalent to KPCA plus FDA. That is, KPCA is

performed first, then FDA is carried out in the feature space obtained by KPCA.
2. PCA and FDA can be seen as the special situations of KPCA and KFDA using the linear kernel function $k(\boldsymbol{v}_i, \boldsymbol{v}_j) = \boldsymbol{v}_i^T \boldsymbol{v}_j$. So FDA can be seen as the procedure of FDA in the PCA mapped space.
3. The performances of both KPCA and KFDA are highly related to the choice of the kernel function and the parameters in the kernel function.
4. Both KPCA and KFDA have two drawbacks that the computation time may increase with the number of training samples, and the data patterns in the feature space are rather hard to interpret in the input data space (Zhu & Song, 2010).

3.4. Classification

The classification proceeds after the feature extraction step. In this step, classifiers are trained in the feature space. The classification procedure can be described as follows: N samples $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N \in \mathbb{R}^L$ distributed in C classes: $Z_1, Z_2, \ldots, Z_C$, are given for training. The sample number in Z_i is N_i, which satisfies $\sum_{i=1}^{C} N_i = N$. The class index of $\boldsymbol{z}_n$ is denoted by h_n, $h_n \in \{1, 2, \ldots, C\}$. The objective of the training process is to get a classifier. With the classifier, the class index h of a real-time sample $\boldsymbol{z}$ can be obtained. In order to make a comparison among different classifiers, three representative classifiers GMM, kNN, and SVM, are under consideration. Without loss of generality, GMM is a preferable parametric classification method, while kNN and SVM are two typical non-parametric ones. kNN is a widely used method, thanks to its simplicity and flexibility. The remarkable characteristics of SVM, such as the good generalization performance, the absence of local minima and the sparse representation of solution, have attracted a lot of attention in recent years (Cao et al., 2003).

3.4.1. GMM

GMM is a parametric classification methodology based on Bayes decision theory (Koski & Noble, 2009). The classification is realized by calculating and comparing the class-conditional probabilities $p(\boldsymbol{z}|Z_i), i = 1, \ldots, C$. In GMM, this density is represented as a weighted sum of R_i component Gaussian densities in the following equation:

$$p(\boldsymbol{z}|Z_i) = \sum_{j=1}^{R_i} p(c_j|Z_i) p(\boldsymbol{z}|c_j, Z_i) \tag{13}$$

where $p(c_j|Z_i), j = 1, \ldots, R_i$ are the mixture weights, which satisfies $\sum_{j=1}^{R_i} p(c_j|Z_i) = 1$, $p(\boldsymbol{z}|c_j, Z_i)$ are the component Gaussian densities. Each component density is an L-variate Gaussian function of the form,

$$p(\boldsymbol{z}|c_j, Z_i) = \frac{1}{(2\pi)^{M/2} |\Sigma_j|^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu}_j)^T \Sigma_i (\boldsymbol{z} - \boldsymbol{\mu}_j) \right\} \tag{14}$$

with mean vector $\boldsymbol{\mu}_j$ and covariance matrix Σ_j. Parameters $\boldsymbol{\mu}_j$, Σ_j and $p(c_j)$ are collectively represented by the notation ζ_i:

$$\zeta_i = \{p(c_j), \boldsymbol{\mu}_j, \Sigma_j\}, \quad j = 1, \ldots, R_i$$

The configuration of R_i is often determined by the complexity of the data distribution. A complex distribution could be described by a choice of large R_i. The Expectation-Maximization (EM) algorithm, whose details can be found in McLachlan and Peel (2004), is adopted to estimate ζ_i. The GMM classification procedure used in this study is described by Algorithm 5.

Algorithm 5. GMM

Training:
1: Collect labeled samples $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N$
2: Initial R_i for $i = 1, \ldots, C$.
3: **for** $i = 1$ to C
Estimate and save ζ_i by using EM algorithm.
end for

Performing:
1: For a new sample $\boldsymbol{z}$, calculate $p(\boldsymbol{z}|Z_i, \zeta_i)$.
2: Class index h is assigned:

$$h = \arg\left\{ \max_{i \in \{1,\ldots,C\}} N_i p(\boldsymbol{z}|Z_i, \zeta_i) \right\} \tag{15}$$

3.4.2. kNN

kNN is a widely-used nonparametric classifier (Bishop, 2006). In the kNN procedure, the classification decision is based on the N training samples. The training step is unnecessary for kNN, and the procedure is given by Algorithm 6.

Algorithm 6. kNN

Training:
1: Collect and save labeled samples: $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N$.

Performing:
1: For a new sample $\boldsymbol{z}$, calculate its Euclidean distances to $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N$.
2: Find the nearest k neighbors of $\boldsymbol{z}$ that are at the minimum Euclidean distances: $\boldsymbol{z}_1^k, \boldsymbol{z}_2^k, \ldots, \boldsymbol{z}_k^k$, whose class indexes are $h_1^k, h_2^k, \ldots, h_k^k$.
3: $\boldsymbol{z}$ is assigned to a class to which most of the neighbors belong:

$$h = \arg\left\{ \max_{j \in \{1,\ldots,C\}} \sum_{i=1}^{k} \delta(h_i^k, j) \right\} \tag{16}$$

3.4.3. SVM

SVM is a classification method developed by Platt (1998) and has been widely applied over the last two decades. The basic theory comes from a binary classification problem. As shown in Fig. 6, data samples are distributed in two classes. Let us suppose we have some hyperplane which separates the points. A SVM looks for an optimal decision hyperplane, which separates the two classes, while at the same time, maximizes the margin between itself and the nearest training examples of each class. These nearest training samples that lie on the margin are called support vectors. For instance, take the samples from the first two classes Z_1 and Z_2, the binary SVM could be described by Algorithm 7.

Algorithm 7. Binary SVM

Training:
1: Collect $(N_1 + N_2)$ labeled sample $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{N_1+N_2}$ from classes Z_1 and Z_2. $g_n \in \{-1, 1\}$, is the class label of sample $\boldsymbol{z}_n$ (-1 for class 1, 1 for class 2). Initial D.
2: Solve the quadratic problem:

$$\min J(\boldsymbol{a}) = \frac{1}{2} \sum_{n=1}^{N_1+N_2} \sum_{m=1}^{N_1+N_2} a_n a_m g_n g_m k(\boldsymbol{z}_n, \boldsymbol{z}_m) - \sum_{n=1}^{N_1+N_2} a_n$$
$$\text{s.t.} \quad \sum_{n=1}^{N} a_n g_n = 0, \quad 0 \le a_n \le D \quad \text{for } n = 1, 2, \ldots, N \tag{17}$$

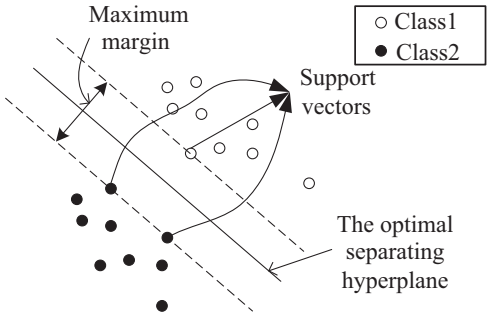

Fig. 6. SVM schematic diagram.

where $\boldsymbol{a}=[a_1,a_2,\ldots,a_{N_1+N_2}]^T$, $k(\boldsymbol{z}_n,\boldsymbol{z}_m)$ is a kernel function.

3: Save support vectors: $\boldsymbol{z}_1^s,\boldsymbol{z}_2^s,\ldots,\boldsymbol{z}_S^s$ and corresponding g_n and a_n, which are denoted by $\{g_n^s\}$ and $\{a_n^s\}$. Support vectors are those samples whose corresponding $a_n>0$.

Performing:

For a new sample $\boldsymbol{z}$, its class label is

$$\text{determined:} g=\text{sign}\left\{\sum_{n=1}^{S} a_n^s g_n^s k(\boldsymbol{z}_n^s,\boldsymbol{z})+b\right\} \tag{18}$$

$$\text{where } b=\frac{1}{S}\sum_{j=1}^{S}\left(g_j^s-\sum_{n=1}^{S} a_n^s g_n^s k(\boldsymbol{z}_n^s,\boldsymbol{z}_j^s)\right)$$

There are several points that need to be emphasized for SVM: first, notice that the real computation process is just correlated to the support vectors. This property is central to the practical applicability of SVM. Second, for solving the quadratic programming problem, a practical approach, Sequential Minimal Optimization (SMO) was used in our study (Platt, 1998). Third, the basic SVM is a binary classifier. To extend the binary classifier to multi-classification situations, a method named "One-Against-One" was adopted in our study. The technique classifies the classes in pairs by using binary SVM. The final classification is obtained by voting all binary classification results. The details can be found in Hsu and Lin (2002).

4. Results and discussion

4.1. The PEMFC under consideration

A 1 kW test bench is used to test a 20-cell PEMFC stack. Table 1 summarizes the parameters of the fuel cell stack investigated. In the test bench, a number of physical parameters impacting or expressing stack performances can be controlled and monitored. Stack temperature (T_{fc}), stoichiometries of hydrogen and air (S_h, S_a), relative humidity of the inlet air (RH), load current (I) can be set. Inlet and outlet pressures of hydrogen and air (P_h, P_a), stack temperatures, current, stack voltage (V_s) and single cell voltages can be monitored. The sample time of the test bench is 150 ms.

The fuel cell temperature is considered to be the value of the temperature at the cooling water (demineralized) outlet. The control of the relative humidity of input air is by means of regulating air dew point temperature. (The input air is set at the same temperature as the fuel cell.) Additional details about the test bench and the test protocol have been previously published in Candusso et al. (2008).

4.2. The results of data labeling

Based on the experimental test bench, experiments in normal conditions were firstly carried out. In the experiments, the stack temperature was set at $T_{fc}=40\ ^\circ$C, the stoichiometries of hydrogen and air were set at the nominal values as in Table 1, relative humidity RH was situated between 75% and 98%, which is considered as a normal range. The output current was configured from 40 A to 2 A.

Table 1
The parameters of the investigated fuel cell stack.

Cell area	100 cm^2
Cell number	20
Flow field structure	Serpentine
Nominal output power	500 W
Nominal operating temperature	40 °C
Operating temperature region	20–65 °C
Maximum operating pressures	1.5 bar
Anode stoichiometry	2
Cathode stoichiometry	4

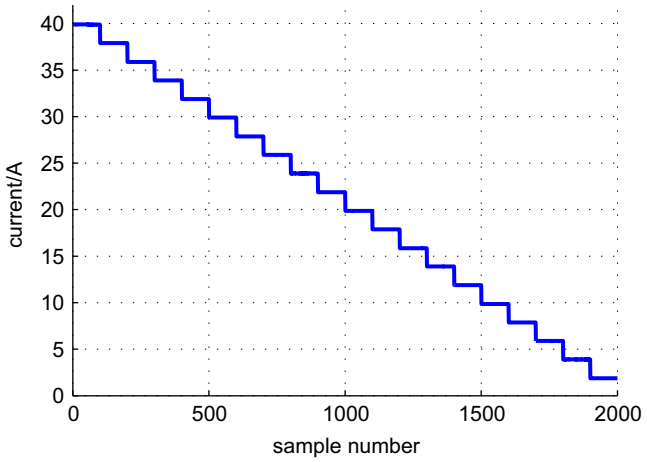

Fig. 7. Current evolution in normal experiment.

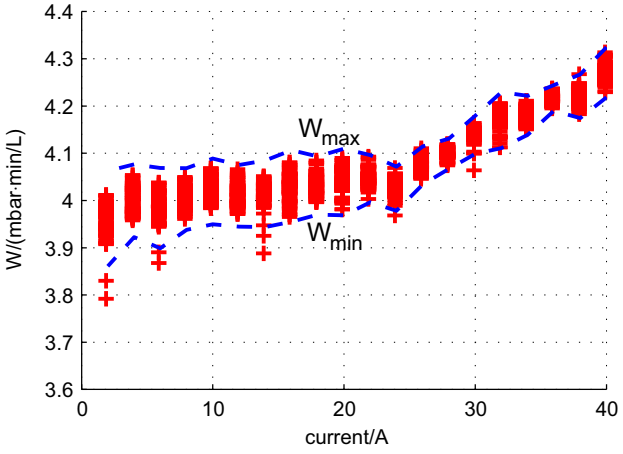

Fig. 8. Values of parameter W in normal condition and their limits.

As shown in Fig. 7, the current was set at 20 discrete points, for every current point, 100 samples were collected into one group.

In the normal state experiments, water indicator W was calculated for every sample. The Lilliefors tests were used to test the null hypothesis that W follows a normally distributed population for each group (Lilliefors, 1967). The null hypotheses were not rejected with significance level 0.05 for all the 20 groups. Hence, it was reasonable to define the up and down limits as (4).

The values W of normal condition in different current points are as in Fig. 8. The up limit W_{max} and down limit W_{min} are also shown in this figure. It can be seen that W_{max} and W_{min} increase globally with increasing current. This could be because water volume generated increases with the rise of the current.

In order to obtain the training data, fault experiments were carried out afterwards. In the fault experiments, the output current was fixed at 40 A; stack temperature was set at $T_{fc}=40\ ^\circ$C; the stoichiometries of the hydrogen and the air were set at the nominal values as in Table 1. The relative humidity RH was situated between 85% and 110%, which would induce flooding by importing some liquid water with the inlet air. The data acquired from a fault experiment was used for training procedure. Several independent fault experiments were done to further test and verify the approach.

The values of W in the fault experiments were compared with the corresponding limits W_{max} and W_{min}. The parameter W in a fault experiment was shown in Fig. 9. From the figure, it can be observed that the data points can be labeled with three class labels and the stack went through three successive states: membrane

drying state, normal state, and flooding state. Actually, it is considered in our study that there is a proper amount of liquid water that exists in the air channels in the normal operating state. At the beginning of the experiment, some time was needed to construct the necessary water environment. Hence, the stack showed a membrane drying state in this period. After this period, with the help of high humidified inlet air, the liquid water accumulated in the air paths. The flooding was therefore induced after a period of normal state.

Although the variable W deduced from the pressure drop model can be used for monitoring the flooding and membrane drying faults, pressure sensors and the instruments for air flow measurement must be settled at both sides of the air stream. These sensors would increase the cost of the fuel cell system. Based on this consideration, an effort was made to realize the fault diagnosis by analyzing only cell voltages.

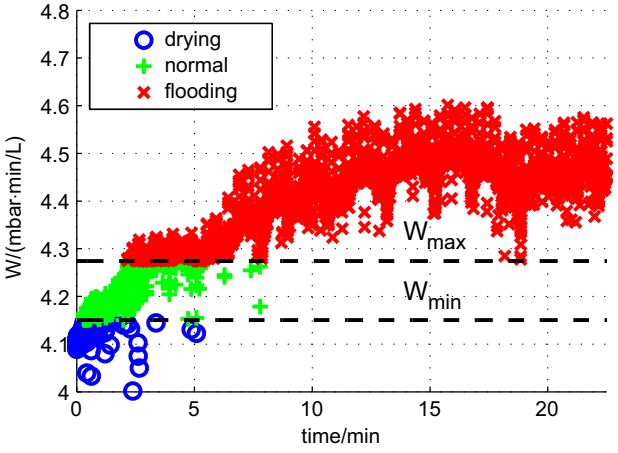

Fig. 9. Parameter W of data in fault process.

4.3. The results of feature extraction and classification

After the samples in fault experiments were labeled, we got to the model training step. As aforementioned, the analytic targets were individual cell voltages in this step. The individual cell voltages in normal experiment and fault experiment are respectively as in Figs. 10 and 11. It can be observed that, from an overall point of view the amplitudes of cell voltages in a normal experiment are more homogeneous than in a fault experiment.

In order to explore the information sufficiently, the data set from one fault experiment is used for the training process. The vectors constructed by cell voltages were used. As mentioned in Section 3, in this study, the dimension of original data is $M=20$, and the number of training samples is $N=9000$. After the labeling process, the training samples were labeled into three classes, so the class number $C=3$. To evaluate the performance of the feature extraction methods and the different classifiers, a criterion error diagnosis rate (EDR) is defined. The error diagnosis points refer to the points which are wrongly diagnosed. EDR means the proportion of the error diagnosis points to total data points.

With the training data set, the feature extraction models and the classification models were trained successively. Some parameters were initialized first. The kernel functions used in KPCA, KFDA and SVM are all Gaussian kernels. In the work of Keerthi and Lin (2003), it has been proven that a Gaussian kernel can approximate most types of kernel functions if the kernel parameter is chosen appropriately. From the similar experience in Yang et al. (2004), the parameters (σ in (11)) of the kernel functions in KPCA and KFDA were set as 0.5 M ($M=20$ is the data dimension of original data space) in this study. The number of Gaussian components (R_i in (13)) in GMM is set at 1 for all the feature extraction methods considering the distribution characteristics of data in feature spaces. In fact, from a visualization point of view, it was found that the within-class data distributions in feature spaces (see Figs. 12–15) are not complex and one Gaussian

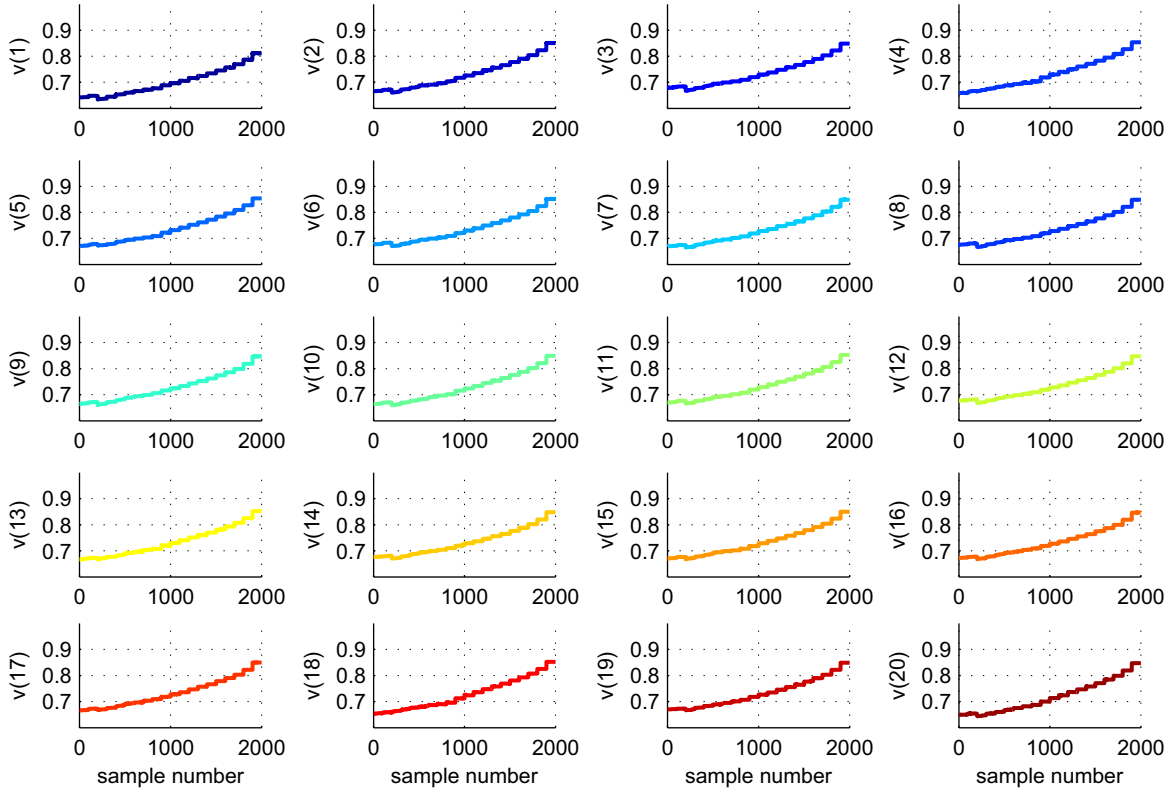

Fig. 10. Cell voltages in a normal experiment. $v(n)$ denotes the voltage behavior of nth cell counting from air entrance to air exit.

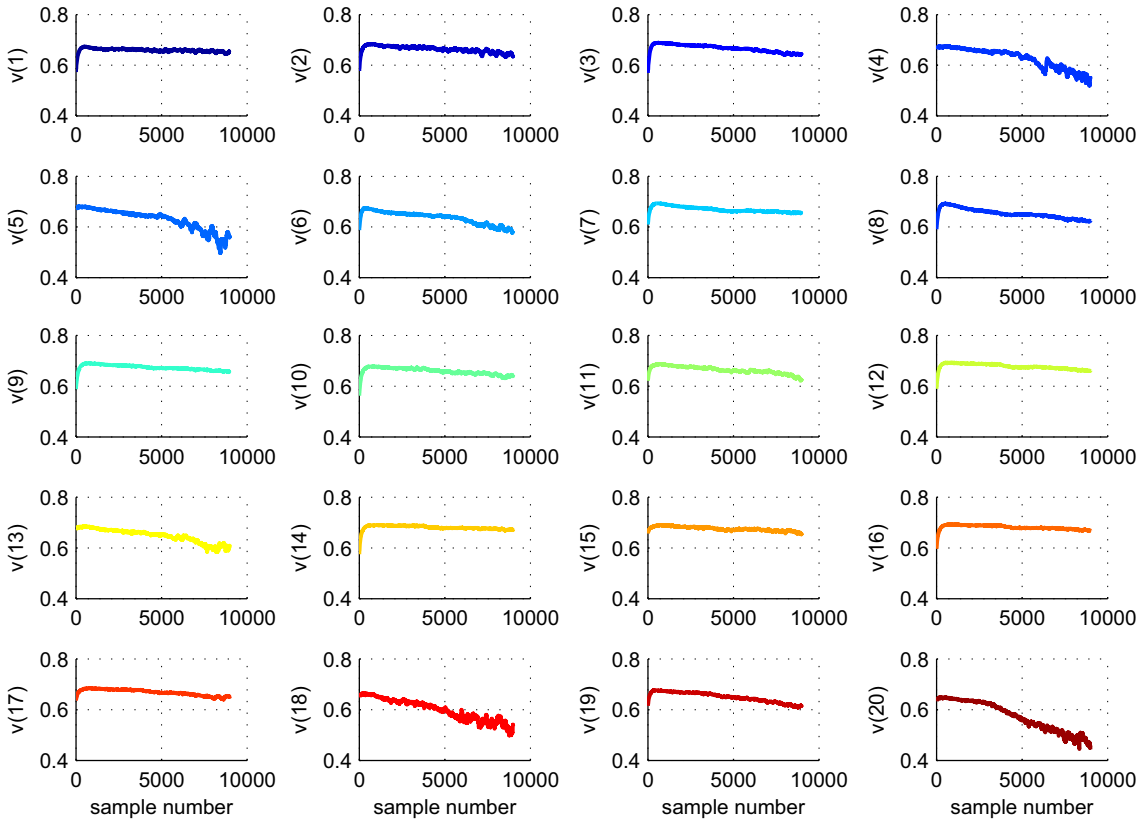

Fig. 11. Cell voltages in a fault experiment.

component could describe the distributions with an acceptable precision. In kNN, the values of k were set at 5, 7, 1, and 15 respectively for the methods PCA, FDA, KPCA, and KFDA, with respecting to the leave-one-out error (Duda et al., 2001). In SVM, the values of kernel parameters (σ) and the parameter D in (17) were determined by trying a group of combinations (see for instance Hsu & Lin, 2002). The pair of (σ, D) that achieved the lowest EDR for test data was chosen. Here the values of σ and D were set at 2 and 10 000 respectively for all the feature extraction methods.

Through feature extraction, the original 20-dimensional data was projected to a 2-dimension space. Figs. 12–15 show respectively the data of a fault experiment in feature spaces generated by PCA, FDA, KPCA, and KFDA.

It can be seen that the data points in Figs. 12 and 14 disperse over the whole scale, whereas the overlap regions between data in the normal state and the other two fault states are large. In contrast, from Figs. 13 and 15, the points in the same class are more concentrated, the amounts of overlapping points are small, which means points in different classes are decentralized.

Classification methods, GMM, kNN, and SVM were carried out in the different feature spaces. For instance, Figs. 16, 17, and 18 show respectively the visualization results of GMM, kNN, SVM classifications in FDA feature space. It can be seen that the feature space is divided into three zones, which denotes different states, and the boundaries determined by different classifiers are different.

Table 2 shows EDRs for different combinations of the feature extraction methods and classification methods. In order to evaluate the robustness of the approach, the data, which is acquired from the other fault experiments than the one for training, were handled as test data. The test data were firstly labeled. Then, the trained feature extraction and classification models were used to process the test data. The EDRs of test data were thus obtained by comparing the diagnostic results with the labeling results.

From Table 2, the performances of the feature extraction methods can be compared. For each classification methodology, the error rates using FDA and KFDA as feature extraction tools are generally lower than those using PCA and KPCA. The reason is that PCA and KPCA are unsupervised methodologies, the training samples are treated equally without considering the label of each point, while FDA and KFDA are supervised methodologies, the labeling information is utilized sufficiently. Hence, we can consider that FDA and KFDA are more suitable for classification problems such as fault diagnosis.

In practice, it is difficult to choose the highest performance classifier. Various problems may have different suitable classification solutions. Concerning our case, the choice of the proper classifier can be achieved firstly by comparing the EDRs of classifiers with FDA and KFDA as foregoing procedures. It can be observed that EDRs obtained using kNN and SVM are always lower than when using GMM.

4.4. Discussion about computation costs

Apart from EDR, computation cost is a crucial factor that needs to be taken into account for real-time implementation. In our approach, the training process is usually out of consideration, since it is completed off-line. Concerning the online diagnosis process, feature extraction methodologies and classification methodologies are considered respectively. The order notation $O()$ is used here to describe the computation cost. $f(x) = O(h(x))$ denotes that there are x_0 and c_0, such that $|f(x)| \leq c_0|h(x)|$ for all $x > x_0$. The order notation is used to give a bound on the limiting behavior of a function.

As in Tables 3 and 4, different methodologies are evaluated from the perspectives of occupied memory and computation time (the notations are described in Section 3). From this table, it can be seen that among different feature extraction methods, needed

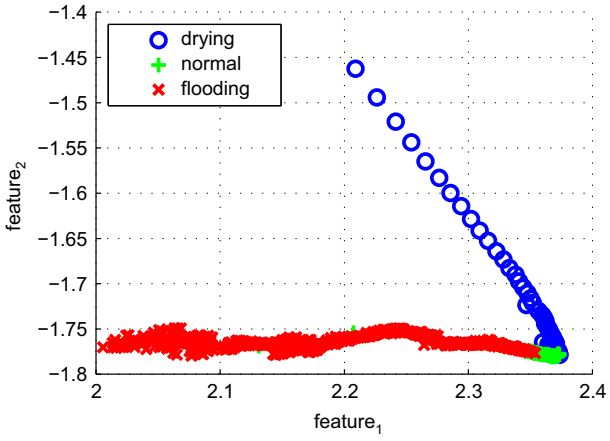

Fig. 12. The features obtained by adopting PCA.

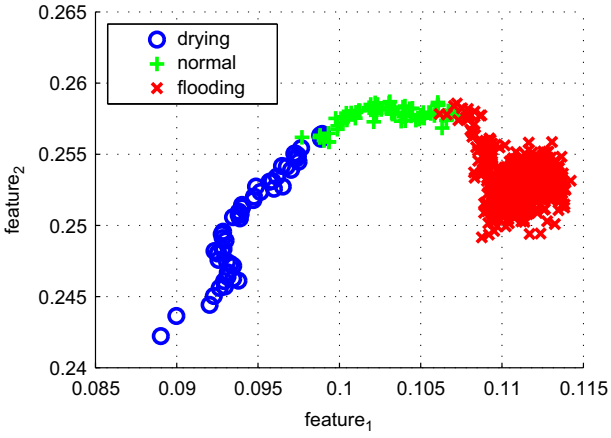

Fig. 13. The features obtained by adopting FDA.

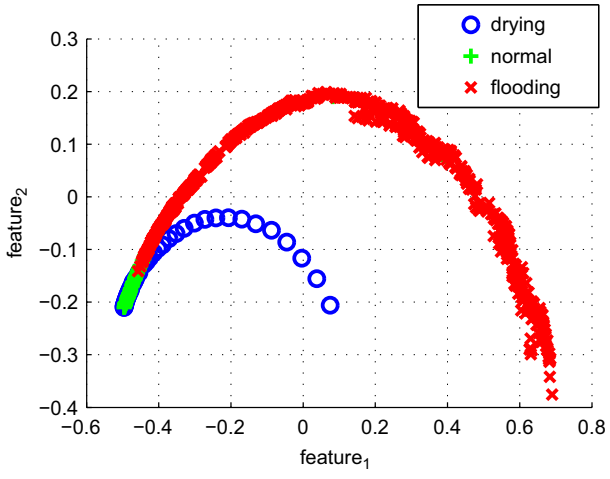

Fig. 14. The features obtained by adopting KPCA.

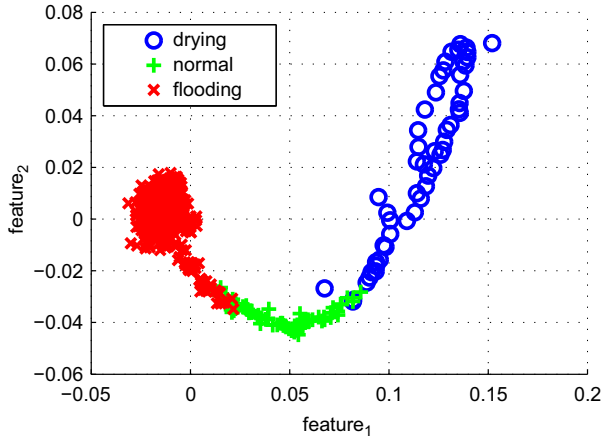

Fig. 15. The features obtained by adopting KFDA.

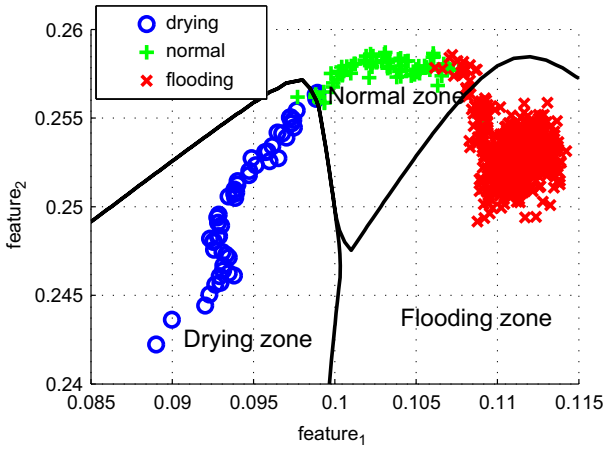

Fig. 16. Classification results in FDA feature space by GMM.

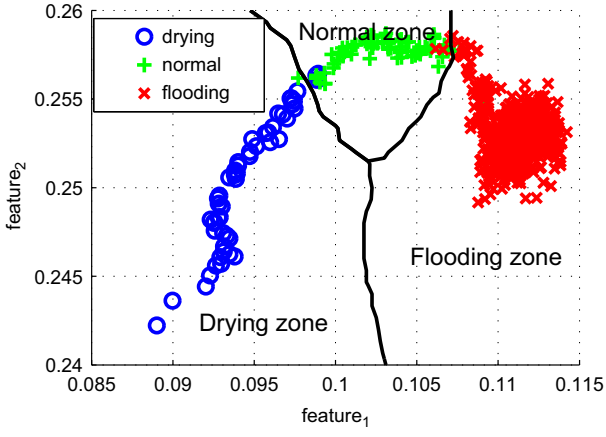

Fig. 17. Classification results in FDA feature space by kNN.

memory and computation time of KPCA and KFDA are in proportion to the number of training samples, and thus are usually large. While for classifiers, the needed memory and computation time of kNN are in proportion to the number of training samples. These methodologies are less suitable than the others for online diagnosis. Hence, considering synthetically the performances of EDR and feasibility of online implementation, FDA combined by SVM can be chosen as final solution in our case.

5. Conclusion

This paper presents an approach for the diagnosis of water management faults in PEMFC stacks. The procedure is realized by classifying the features that are extracted from the vectors constructed by individual cell voltages.

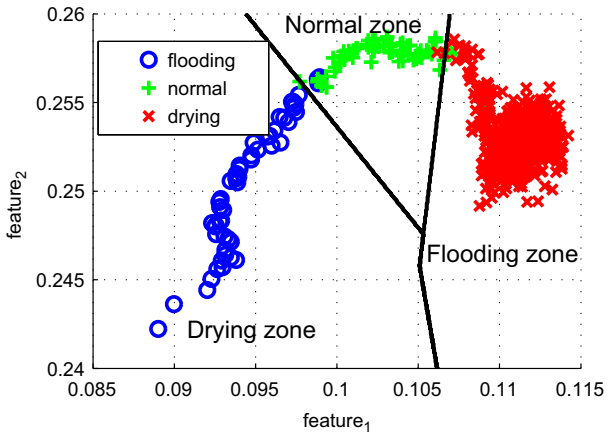

Fig. 18. Classification results in FDA feature space by SVM.

Table 2
Results of varied classifications in different feature spaces.

Feature extraction	Classification	EDR of training data	EDR of test data
PCA	GMM	0.051	0.110
	kNN	0.016	0.110
	SVM	0.015	0.129
FDA	GMM	0.032	0.087
	kNN	0.013	0.070
	SVM	0.014	0.070
KPCA	GMM	0.089	0.058
	kNN	0.052	0.121
	SVM	0.058	0.113
KFDA	GMM	0.034	0.085
	kNN	0.014	0.082
	SVM	0.016	0.075

Table 3
Computation costs of the feature extraction methodologies.

Methodologies	Feature extraction			
	PCA	FDA	KPCA	KFDA
Occupied memory	$O(ML)$	$O(ML)$	$O(MN+NL)$	$O(MN+NL)$
Computation time	$O(ML)$	$O(ML)$	$O(MN+NL)$	$O(MN+NL)$

Table 4
Computation costs of the classification methodologies.

Methodologies	Classification		
	GMM	kNN	SVM
Occupied memory	$O\left(CR\left(\frac{(1+M)(M+2)}{2}\right)\right)$	$O(MN)$	$O(MS+S)$
Computation time	$O\left(CR\left(\frac{(1+M)(M+2)}{2}\right)\right)$	$O(MN)$	$O(MS+S)$

In this approach, the water indicator W is defined to describe the quantity of water inside the fuel cell stack and label the training data. Taking into account the uneven character of cell voltage amplitudes, individual cell voltages are chosen as original variables for diagnosis. Different feature extraction and classification methods are employed and compared. The test results for a 20-cell stack show that FDA and SVM have higher performance and less computation costs comparing with other methods in our case. The EDR of diagnosis by using such an approach is always below 10%. It is therefore a very promising diagnostic proposal to diagnose the faults associated with water management for PEMFC.

Works are in progress in order to extend the proposed approach to other type of faults by increasing the number of classes and the training data related to the corresponding faults.

References

Abbasion, S., Rafsanjani, A., Farshidianfar, A., & Irani, N. (2007). Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing*, *21*(7), 2933–2945 doi:http://dx.doi.org/10.1016/j.ymssp.2007.02.003.

Asghari, S., Mokmeli, A., & Samavati, M. (2010). Study of PEM fuel cell performance by electrochemical impedance spectroscopy. *International Journal of Hydrogen Energy*, *35*(17), 9283–9290 doi:http://dx.doi.org/10.1016/j.ijhydene.2010.03.069.

Baudat, G., & Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. *Neural Computation*, *12*(10), 2385–2404, http://dx.doi.org/10.1162/089976600300014980.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Campbell, C. (2002). Kernel methods: A survey of current techniques. *Neurocomputing*, *48*(1–4), 63–84 doi:http://dx.doi.org/10.1016/S0925-2312(01)00643-9.

Candusso, D., De Bernardinis, A., Péra, M.-C., Harel, F., François, X., Hissel, D., et al. (2008). Fuel cell operation under degraded working modes and study of diode by-pass circuit dedicated to multi-stack association. *Energy Conversion and Management*, *49*(4), 880–895 doi:http://dx.doi.org/10.1016/j.enconman.2007.10.007.

Cao, L., Chua, K., Chong, W., Lee, H., & Gu, Q. (2003). A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, *55*(1), 321–336 doi:http://dx.doi.org/10.1016/S0925-2312(03)00433-8.

Cover, T. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, *EC-14*(3), 326–334, http://dx.doi.org/10.1109/PGEC.1965.264137.

Duda, R., Hart, P., & Stork, D. (2001). *Pattern classification*. Wiley, New York, 2nd edition.

Escobet, T., Feroldi, D., de Lira, S., Puig, V., Quevedo, J., Riera, J., et al. (2009). Model-based fault diagnosis in PEM fuel cell systems. *Journal of Power Sources*, *192*(1), 216–223 doi:http://dx.doi.org/10.1016/j.jpowsour.2008.12.014.

Hernandez, A., Hissel, D., & Outbib, R. (2006). Fuel cell fault diagnosis: A stochastic approach. In *IEEE international symposium on industrial electronics*, *2006* (Vol. 3, pp. 1984–1989). doi:http://dx.doi.org/10.1109/ISIE.2006.295877.

Hernandez, A., Hissel, D., & Outbib, R. (2010). Modeling and fault diagnosis of a polymer electrolyte fuel cell using electrical equivalent analysis. *IEEE Transaction on Energy Conversion*, *25*(1), 148–160, http://dx.doi.org/10.1109/TEC.2009.2016121.

Hissel, D., Péra, M. C., & Kauffmann, J. M. (2004). Diagnosis of automotive fuel cell power generators. *Journal of Power Sources*, *128*(2), 239–246 doi:http://dx.doi.org/10.1016/j.jpowsour.2003.10.001.

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 415–425, http://dx.doi.org/10.1109/72.991427.

Hua, J., Lu, L., Ouyang, M., Li, J., & Xu, L. (2011). Proton exchange membrane fuel cell system diagnosis based on the signed directed graph method. *Journal of Power Sources*, *196*(14), 5881–5888 doi:http://dx.doi.org/10.1016/j.jpowsour.2011.03.008.

Ito, K., Ashikaga, K., Masuda, H., Oshima, T., Kakimoto, Y., & Sasaki, K. (2008). Estimation of flooding in PEMFC gas diffusion layer by differential pressure measurement. *Journal of Power Sources*, *175*(2), 732–738 doi:http://dx.doi.org/10.1016/j.jpowsour.2007.10.019.

Keerthi, S. S., & Lin, C.-J. (2003). Asymptotic behaviors of support vector machines with gaussian kernel. *Neural Computation*, *15*(7), 1667–1689, http://dx.doi.org/10.1162/089976603321891855.

Kim, J., Lee, I., Tak, Y., & Cho, B. (2012). State-of-health diagnosis based on hamming neural network using output voltage pattern recognition for a PEM fuel cell. *International Journal of Hydrogen Energy*, *37*(5), 4280–4289 doi:http://dx.doi.org/10.1016/j.ijhydene.2011.11.092.

Koski, T., & Noble, J. M. (2009). *Bayesian networks: An introduction*. Wiley, Chichester, West Sussex, UK.

Lee, C.-Y., & Lee, Y.-M. (2012). In-situ diagnosis of local fuel cell performance using novel micro sensors. *International Journal of Hydrogen Energy*, *37*(5), 4448–4456 doi:http://dx.doi.org/10.1016/j.ijhydene.2011.11.098.

Li, H., & Xiao, D. (2012). Fault diagnosis using pattern classification based on one-dimensional adaptive rank-order morphological filter. *Journal of Process Control*, *22*(2), 436–449 doi:http://dx.doi.org/10.1016/j.jprocont.2011.12.005.

Lilliefors, H. (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*(318), 399–402, http://dx.doi.org/10.1080/01621459.1967.10482916.

McLachlan, G., & Peel, D. (2004). *Finite mixture models*. New York: Wiley.

Mérida, W., Harrington, D., Canut, J. L., & McLean, G. (2006). Characterisation of proton exchange membrane fuel cell (PEMFC) failures via electrochemical impedance spectroscopy. *Journal of Power Sources*, *161*(1), 264–274 doi:http://dx.doi.org/10.1016/j.jpowsour.2006.03.067.

Platt, J. C. (1998). *Sequential minimal optimization: A fast algorithm for training support vector machines* (pp. 1–21). Technical Report MSR-TR-98-14, Microsoft Research.

Riascos, L. A. M., Simoes, M. G., & Miyagi, P. E. (2007). A Bayesian network fault diagnostic system for proton exchange membrane fuel cells. *Journal of Power Sources*, *165*(1), 267–278 doi:http://dx.doi.org/10.1016/j.jpowsour.2006.12.003.

Rodatz, P., Büchi, F., Onder, C., & Guzzella, L. (2004). Operational aspects of a large PEFC stack under practical conditions. *Journal of Power Sources*, *128*(2), 208–217 doi:http://dx.doi.org/10.1016/j.jpowsour.2003.09.060.

Tirnovan, R., & Giurgea, S. (2012). Efficiency improvement of a PEMFC power source by optimization of the air management. *International Journal of Hydrogen Energy*, *37*(9), 7745–7756 doi:http://dx.doi.org/10.1016/j.ijhydene.2012.02.029.

Wang, Y., Basu, S., & Wang, C.-Y. (2008). Modeling two-phase flow in PEM fuel cell channels. *Journal of Power Sources*, *179*(2), 603–617 doi:http://dx.doi.org/10.1016/j.jpowsour.2008.01.047.

Wasterlain, S., Candusso, D., Harel, F., François, X., & Hissel, D. (2010). Diagnosis of a fuel cell stack using electrochemical impedance spectroscopy and bayesian networks. In *Vehicle power and propulsion conference (VPPC)*, IEEE, 2010 (pp. 1–6). doi:http://dx.doi.org/10.1109/VPPC.2010.5729184.

Wu, K.-P., & Wang, S.-D. (2009). Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. *Pattern Recognition*, *42*(5), 710–717 doi:http://dx.doi.org/10.1016/j.patcog.2008.08.030.

Yang, J., Jin, Z., yu Yang, J., Zhang, D., & Frangi, A. F. (2004). Essence of kernel Fisher discriminant: KPCA plus LDA. *Pattern Recognition*, *37*(10), 2097–2100 doi:http://dx.doi.org/10.1016/j.patcog.2003.10.015.

Yousfi-Steiner, N., Hissel, D., Moçotéguy, P., & Candusso, D. (2011a). Diagnosis of polymer electrolyte fuel cells failure modes (flooding & drying out) by neural networks modeling. *International Journal of Hydrogen Energy*, *36*(4), 3067–3075 doi:http://dx.doi.org/10.1016/j.ijhydene.2010.10.077.

Yousfi-Steiner, N., Hissel, D., Moçotéguy, P., & Candusso, D. (2011b). Non intrusive diagnosis of polymer electrolyte fuel cells by wavelet packet transform. *International Journal of Hydrogen Energy*, *36*(1), 740–746 doi:http://dx.doi.org/10.1016/j.ijhydene.2010.10.033.

Yousfi-Steiner, N., Moçotéguy, P., Candusso, D., Hissel, D., Hernandez, A., & Aslanides, A. (2008). A review on PEM voltage degradation associated with water management: Impacts, influent factors and characterization. *Journal of Power Sources*, *183*(1), 260–274 doi:http://dx.doi.org/10.1016/j.jpowsour.2008.04.037.

Zhu, Z.-B., & Song, Z.-H. (2010). Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis. *Chemical Engineering Research and Design*, *88*(8), 936–951 doi:http://dx.doi.org/10.1016/j.cherd.2010.01.005.