# DIAGNOSIS OF PEMFC BY USING STATISTICAL ANALYSIS

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

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

zhongliang.li@lsis.org, rachid.outbib@lsis.org, daniel.hissel@univ-fcomte.fr, Stefan.Giurgea@utbm.fr

**ABSTRACT**
Fault diagnosis, especially on-line fault diagnosis is an essential issue for practical application of Polymer Electrolyte Membrane Fuel Cell (PEMFC) system. This paper proposes a diagnosis approach for PEMFC to handle the flooding fault which is considered to be a common fault. In this procedure, both fault detection and fault isolation are considered. For fault detection, the statistical characters of cell voltages distribution of a 20-cell PEMFC stack are analyzed. Parameters for describing voltage distribution characters are extracted. After that, a subset of parameters is selected in the orientation that the definition of fault is as correct as possible. A popular clustering methodology named K-means clustering (KMC) is adopted to make definition of flooding fault zone. For fault isolation, Support vector machine (SVM) classifier is trained to handle the cell voltages constructed vectors. Two different causes of flooding: increasing air humidity and decreasing of stack temperature can be discriminated by the classifier with a high correctness.

Keywords: PEMFC, diagnosis, flooding, voltage distribution, fault isolation

## 1. INTRODUCTION

Environmental issues have increased the demand for less polluting energy generation technologies. Developing fuel cell generator seems to be a possible solution, because they can offer substantially lower emissions, particularly of $CO_2$. Among different kinds of fuel cell, the PEMFC has been drawing more attention because of its high efficiency and its high power density (Wahdame et al. 2008).

However, to be widely adapted to transportation and stationary applications, the reliability and durability have to be improved. Fault diagnosis is an efficient solution to achieve this: fault diagnosis can be combined with control regulation to avoid the serious problem of system (Samy et al. 2011), it is also helpful to understand the physical and chemical mechanism (Wasterlain et al. 2011), so as to speed up the development cycle of new technologies. In addition, fault diagnosis can improve user support and acceptance by reducing down time (Tian et al. 2008).

In PEMFC diagnosis, several points must be satisfied: For transportation application of PEMFC, which aims at minimizing the embedded instrumentation to improve the reliability and decrease the cost, easy-to-monitor parameters are highly desired. For on-line diagnosis, computing time must be restricted so as to achieve diagnosis calculation within a sample time cycle. For control oriented diagnosis, fault isolation, which is to point the type of fault and its location, is an important component other than fault detection (Samy et al. 2011).

Some literatures have provided several fuel cell stack and system diagnostic methods. Physical modeling is an intuitional way to realize the aim of diagnosis (Escobet et al. 2009), however, parameters estimation is a barrier in PEMFC system. Some other works have been done to overcome the shortage of physical modeling. G. Tian et al, see (Tian et al. 2010), developed a methodology based on the analysis of the Open Circuit Voltage (OCV) in order to detect leakage fault, but it is not for on-line diagnosis. In (Yousfi Steiner et al. 2011), N. Yousfi Steiner et al. proposed neural networks model based diagnosis procedure to handle water management issues, but a series of variables, including pressures, are needed to be monitored.

In (Hissel et al. 2004), D. Hissel et al. proposed a fuzzy diagnostic model tuned by genetic algorithm, which is used to diagnose drying of membrane and accumulation of $N_2/H_2O$ in the anode compartment. L. Alberto et al, in paper (Alberto et al. 2008), proposed an approach based on Bayesian networks, which can handle four types of faults in PEMFC system. In (Steiner et al. 2011), N. Yousfi Steiner et al. proposed a wavelet package translating methodology, flooding fault can be detected by analyzing the behavior of fuel cell stack voltage. J. Hua et al., see (Hua et al. 2011), proposed a multivariate statistical method, in which faults can be detected by analyzing principal components. Although above approaches can efficiently achieve the tasks of diagnosis, most of the literatures consider the fuel cell stack as integration. However, fuel cell stack in practice are always composed by series fuel cells, the behaviors of cells are different actually, even in normal operating state. This can be observed from the distribution of individual cell voltages (Hernandez et al.

2006). For this reason, assuming fuel stack as one block constituted by homogeneous cells seems to be arbitrary, additionally, the analysis of the statistical features of individual cells can support us more information for diagnosis.

The goal of this work is to extend the previous studies (Hernandez 2006), in which some statistical qualities of cell voltages distribution and their relations to flooding fault have been found. In this paper, methodologies are used to improve the diagnosis performance and make the previous work more completed. Summarily, the proposed approach is suitable for real-time diagnosis, and is based on data. Specifically, in this paper, four statistical parameters are extracted to analyze the distribution of cell voltages. Fault definition methodology is proposed, in which feature selection and KMC are adopted. Fault isolation is also accomplished with a high correctness by training a SVM classifier, such that the two causes of flooding fault can be discriminated.

This paper is organized as follows: In section 2, the experiments of PEMFC are introduced. In section 3, the diagnosis approach is expounded, including detailed presentations of methodologies and their using in fault detection and isolation. The results of diagnosis are given in the next section. Finally, we conclude the work in section 5.

## 2. DESCRIPTION OF EXPERIMENTS

The data used are collected from a 20-cells PEMFC stack, whose nominal output power is 500 W. Flooding experiments with a certain load (40A) were carried out on the test bench. Some fault controls can result in the flooding inside the fuel. Here, the flooding was induced by two ways: First, increasing the inlet air humidity in order to favor water condensation. Second, the water vapor condensation in the fuel cell is caused by decreasing the temperature of fuel cell stack. The cell voltages were sampled in the flooding evolutions. Figure 1 and figure 2 show cell voltages in flooding evolutions caused by inlet air humidity variation and stack temperature variation. The goal of this paper is to realize fault diagnosis by statistical analysis of cell voltages distribution in these two evolutions.

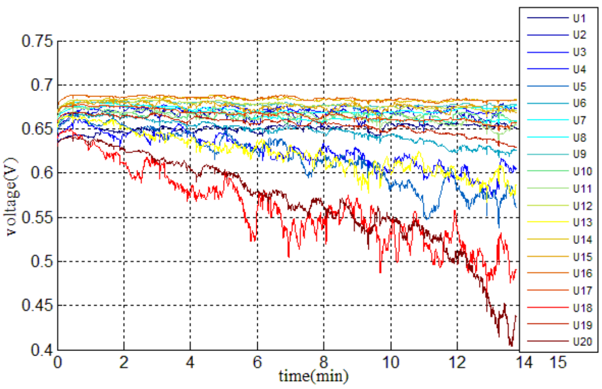


Figure 1: Cell voltages in flooding evolution caused by increasing the inlet air humidity

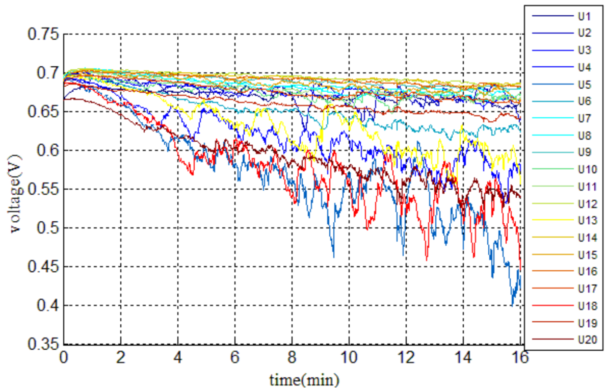


Figure 2: Cell voltages in flooding evolution caused by decreasing the stack temperature

## 3. THE PROPOSED APPROACH

### 3.1. Principle of the approach

The objective of this paper is to realize online diagnosis of PEMFC. The flow chart of diagnosis, showed in figure 3, is as follows: after the real-time samples (just individual cell voltages in this paper) are imported, some diagnosis oriented features are computed. The fault (flooding fault) can be detected if the features of the sample are in the fault zone. After fault is detected, fault isolation can be achieved by a SVM classifier. By using the fault isolation, two fault causes of flooding were been considered: the variation of inlet air humidity (cause1) and the variation of the stack temperature (cause2).

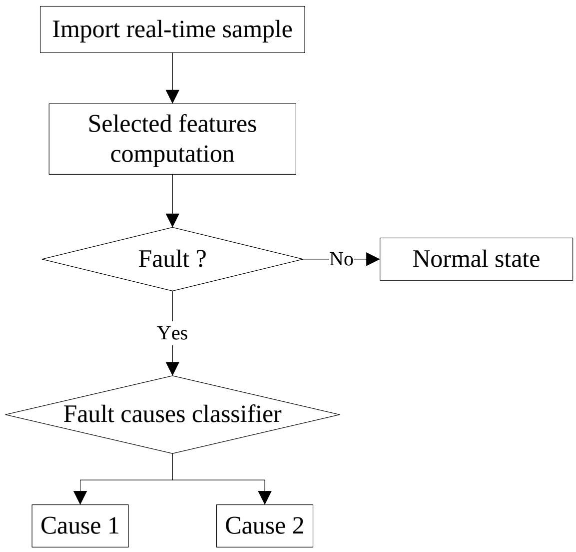


Figure 3: On-line diagnosis flow chart

In the online diagnosis flow chart, some training procedures need to be accomplished off-line. As the figure 4 shows, for the fault detection (the left figure), several statistical parameters are extracted firstly, then, a diagnosis oriented subset of parameters is selected. After that, KMC is adopted for fault definition. For fault isolation (the right figure), vectors composed by cell voltages are constructed firstly, then, SVM classifier is trained based on the vectors.

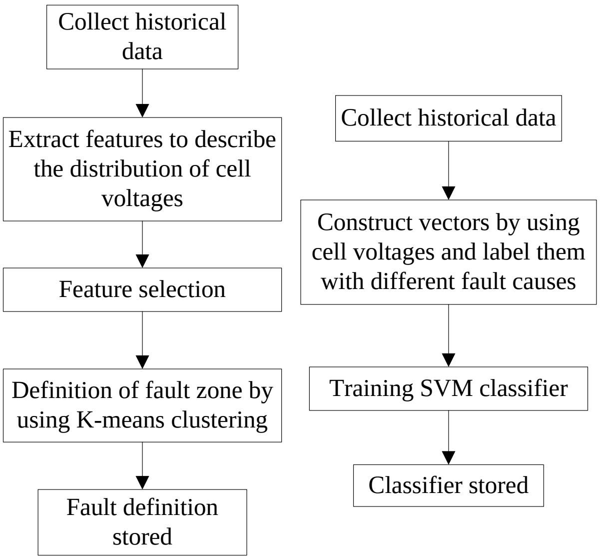


Figure 4: Flow chart of off-line training procedures

### 3.2. Description of cell voltages distribution

In order to describe the statistical characters of cell voltages distribution, some description variables are extracted.

#### 3.2.1. The hypothesis test of Normal distribution

Normal distribution is a most widely used distribution model. A specific normal distribution can be described by two parameters: $\mu$, the mean or the expectation, and $\sigma^2$, the variance. In order to extract the features to describe the distribution of the cell voltages, the hypothesis test of normal distribution is carried out, if the cell voltages follow the normal distribution, $\mu$ and $\sigma^2$ would be extracted; otherwise, some other parameters would be seeked other than these two parameters.

The *Lilliefors test* is used to test the null hypothesis that data come from a normally distributed population, when the null hypothesis does not specify which normal distribution (Lilliefors 1967).

After hypothesis test, the null hypotheses of 84.5% data samples in the flooding evolution are rejected with significance level $\alpha = 0.05$. Hence, it is arbitrary to consider the cell voltages are from normal distribution.

#### 3.2.2. Cell voltages distribution features

The mean $\mu$ and the variance $\sigma^2$ are significant descriptive measures that locate the center and describe the dispersion of probability density function. However, they do not provide a unique characterization of the distribution. To better approximate the probability distribution of a random variable, *skewness* and *kurtosis*, two parameters that are usually used for describing a distribution statistical characters (Ramachandran and Tsokos 2009), are considered other than $\mu$ and $\sigma^2$.

Let $X$ be a random variable, *Skewness*, which is denoted by *Sk*, is defined as

$$Sk = \frac{E[(X-\mu)^3]}{\sigma^3} \tag{1}$$

where $E$ denotes expected value, $\mu$ and $\sigma^2$ are mean value and variance of $X$. *Skewness* is used as a measure of the symmetry of a density function about its mean. Recall that a distribution is symmetric if it looks the same to the left and right of the center point. If $Sk = 0$, then the distribution is symmetric about the mean, if $Sk > 0$, the distribution has a longer right tail, and if $Sk < 0$, the distribution has a longer left tail. The *Skewness* of a normal distribution is zero.

*Kurtosis*, which is denoted by *Ku*, is defined as

$$Ku = \frac{E[(X-\mu)^4]}{\sigma^4} \tag{2}$$

*Kurtosis* is a measure of whether the distribution is peaked or flat relative to a normal distribution. *Kurtosis* is based on the size of a distribution's tails. A high *kurtosis* distribution has longer, fatter tails. A low *kurtosis* distribution has shorter, thinner tails. A distribution which has the same *kurtosis* as a normal distribution is known as *mesokurtic*. It is known that the kurtosis for a standard normal distribution is $Ku = 3$. A distribution with positive excess of 3 is called *leptokurtic*, while a distribution with negative excess of 3 is called *platykurtic* (Ramachandran and Tsokos 2009).

Four parameters: $\mu$, $\sigma^2$, *Sk* and *Ku* of cell voltages in the flooding evolutions caused by increasing the humidity of inlet air and decreasing the temperature of stack are as figure 5 and figure 6.

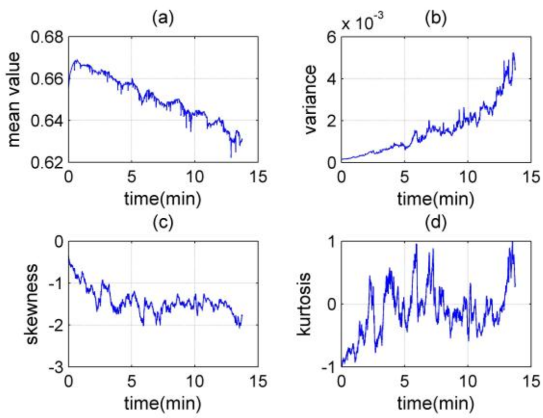


Figure 5: Four statistical variables of cell voltages in flooding evolution caused by humidity variation

From the figures, some intuitional characteristics can be observed:

1. Mean value decreases and variance increases in the flooding evolution.
2. *Skewness* is always negative, which means the probability density distribution function of cell voltages has a longer left tail.
3. *Kurtosis* is always with a negative excess of 3 which means that the probability density function is *leptokurtic*.
4. *Skewness* and *kurtosis* show no monotonous variation in the flooding evolution.

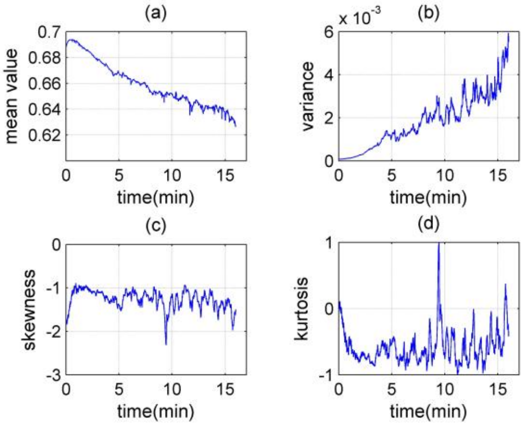

Figure 6: Four distribution parameters in flooding evolution caused by temperature variation

Until now, four typical parameters to describe the properties of cell voltages distribution have been extracted. Fault diagnosis will be preceded based on the four parameters.

### 3.3. Feature selection

Although all these four parameters: $\mu$, $\sigma^2$, *Sk* and *Ku* are useful for cell voltages distribution description, for the fault diagnosis, just a subset of the parameters has a significant role. The goal of feature selection procedure is to find and to select the subset which improves the performance of fault diagnosis. In order to evaluate the performance of feature selection quantificationally, a criterion denoted as *Cr* is defined. The number of non-void subsets of dataset $\{\mu, \sigma^2, Sk, Ku\}$ is $2^4-1 = 15$. The criterions of each subset are calculated and compared, the parameters of the subset which has the most optimal criterion will be chosen as the final features for diagnosis.

The computation of the criterion is connected to the definition of the fault. In the flooding evolutions, it is considered that the vapor condenses as time, which results in the intensive to flooding. So if we allocate the samples in flooding evolution into two clusters (normal cluster and fault cluster), the normal cluster must be the one of which the major part is in the front of the time axis, while the major part of fault cluster is in the tail of the time axis. Consequently, the performance of the clustering can be evaluated by the proportion of overlaps of the two clusters in the time domain. More specifically, the evaluation criterion of fault definition is defined as

$$Cr = N_{error}(\frac{1}{N_1} + \frac{1}{N_2}) \tag{3}$$

where $N_1$ and $N_2$ are the numbers of samples in normal cluster and fault cluster, $N_{error}$ is the number of samples in normal cluster whose indexes exceed $N_1$.

### 3.4. Fault definition

#### 3.4.1. KMC

Clustering is an unsupervised learning algorithm which allocates data points to a certain number of clusters. It is used widely to collect similar data. So clustering is suited for fault definition. KMC is one of the simplest and most popular to solve clustering problem (Macqueen 1967).

The KMC problem aims at allocating a dataset $\{x_1, x_2, ..., x_N\}$ into $C$ clusters, which are denoted as $\omega_1,\ldots,\omega_C$, so as to minimize the within-cluster sum of squares which is defined as

$$J = \sum_{i=1}^{C} \sum_{x_n \in \omega_i} \| x_n - \bar{x}_i \|^2 \tag{4}$$

where $\bar{x}_i$ is the center of cluster $\omega_i$,

$$\bar{x}_i = \frac{1}{N_i} \sum_{x_n \in \omega_i} x_n \tag{5}$$

$N_i$ is number of samples in cluster $\omega_i$.

The implementation of KMC is as follows (Güneş et al. 2010):

**Stage1**: Choose $C$ initial cluster centers $\bar{x}_1, ..., \bar{x}_C$ randomly from the $N$ points $\{x_1, x_2,\ldots, x_N\}$.

**Stage2**: Assign point $x_n$ to the cluster $\omega_i$, $i = 1,\ldots,C$, if

$$\| x_n - \bar{x}_i \| < \| x_n - \bar{x}_j \|, \quad j = 1, ..., C, j \neq i \tag{6}$$

**Stage3**: Compute new cluster centers

$$\bar{x}_i^{new} = \frac{1}{N_i} \sum_{x_n \in \omega_i} x_n \tag{7}$$

**Stage4**: Repeat stage2 and 3 until the centroids no longer move. If

$$\| \bar{x}_i^{new} - \bar{x}_i \| < \varepsilon, \quad i = 1, ..., C \tag{8}$$

the computation process is terminated, otherwise back to stage2.

Although there is no guarantee of achieving a global minimum, the convergence of the algorithm is ensured (Yiakopoulos et al. 2011).

### 3.4.2. Application of KMC

KMC is used for fault definition, the goal of which is to divide "normal state" zone and "fault state" zone in the selected feature space. In this artical, fault definition is accomplished by two stages: Firstly, data in flooding evolution are labeled as "normal state" or "fault state". This is achieved by KMC procedure. After that, different zones are gotten based on the centroids of clusters. The points of the boundary between two zones are equidistant from their centroids.

## 3.5. Fault isolation

The tasks of fault diagnosis contain not only the fault detection, but also fault isolation which is to fix the position and the causes of faults. Fault isolation is beneficial to modify the control in order to slack or eliminate the fault, also it play a positive part in the design and manufacture of products. In this article, the flooding fault is formed by two causes: increasing of the humidity of the inlet air and decreasing of stack temperature. The objective of this part is to isolate these two causes. The fault data can be obtained after fault definition, and data of these two situations have been labeled, hence, the problem is a classification problem.

### 3.5.1. SVM

SVM is a classification method developed by V. Vapnik (Vapnik 1999) and has been widely applied last two decades. As figure 7 shows, the general description of a 2-class SVM is: SVM establishes the optimal separating hyperplane that allocates the majority of points of the same class in the same side, whilst make the smallest distances between the two classes and the hyperplane, which is called margin, to be maximized. The points of the two classes which create the hyperplane are called support vectors.

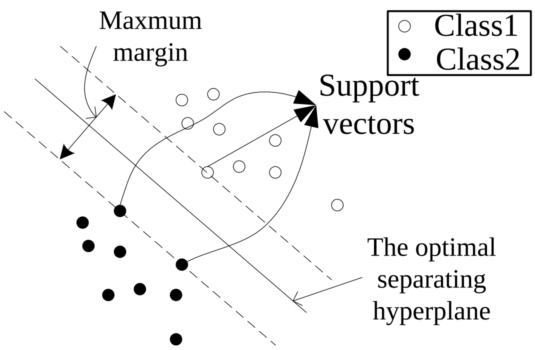


Figure 7: Classification of two classes using SVM

Given the training data $\{\mathbf{x}_n, y_n\}$, $n$ = 1,..., $N$. where the $\mathbf{x}_n \in \mathbf{R}^M$ and $y_n \in \{1;-1\}$. $\mathbf{x}_n$ is an input vector and $y_n$ indicates the class of $\mathbf{x}_n$. It is possible to determine the hyperplane separating data as

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0 \tag{9}$$

where $\mathbf{w} \in \mathbf{R}^M$ is normal to the hyperplane, $b$ is a bias. A distinct separating hyperplane should satisfy the constraints

$$f(\mathbf{x}_n) \geq 1 \quad \text{if } y_n = 1 \tag{10}$$

$$f(\mathbf{x}_n) \leq -1 \quad \text{if } y_n = -1 \tag{11}$$

Support vectors lie on the planes $\mathbf{w}^T\mathbf{x}_n + b = \pm 1$. The margin is $2/\|\mathbf{w}\|$, consequently, the problem can be converted to find the minimum of $\|\mathbf{w}\|^2/2$. Taking into account the misclassification errors with slack variables $\xi_n$ and the error penalty $D$, the optimal hyperplane separating the data can be obtained as a solution to the following optimization problem:

$$\min \quad D\sum_{n=1}^{N}\xi_n + \frac{1}{2}\|\mathbf{w}\|^2 \tag{12}$$

subject to

$$\begin{cases} y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1 - \xi_n \\ \xi_n > 0 \end{cases} \tag{13}$$

After introducing kernel functions to extend the SVM to nonlinear classification domain, the optimization problem (12-13) can be converted to a dual quadratic optimization problem as

$$\max L(\mathbf{a}) = \sum_{n=1}^{N} a_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m y_n y_m k(\mathbf{x}_n, \mathbf{x}_m) \tag{14}$$

subject to

$$\begin{cases} 0 \leq a_n \leq D \\ \sum_{n=1}^{N} a_n y_n = 0 \end{cases} \tag{15}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T$ satisfies $\mathbf{w} = \sum_{n=1}^{N} a_n y_n \mathbf{x}_n$, and $k(\mathbf{x}_n, \mathbf{x}_m)$ denotes kernel function on $\mathbf{x}_n$ and $\mathbf{x}_m$.

The associated class label $y$ of a new vector $\mathbf{x}$ is calculated as follow

$$y = sign\{\sum_{n=1}^{N} a_n y_n k(\mathbf{x}_n, \mathbf{x}) + b\} \tag{16}$$

Three typical kernel functions are presented in table1.

Table 1: Three typical kernel functions

| Kernel | $k(\mathbf{x}_i,\mathbf{x}_j)$ |
|---|---|
| Linear | $\mathbf{x}_i^T\mathbf{x}_j$ |
| Polynomial | $(\mathbf{x}_i^T\mathbf{x}_j + 1)^d$ |
| Gaussian | $\exp(-\|\mathbf{x}_i-\mathbf{x}_j\|^2/c)$ |

Direct solution of the quadratic programming problem (14-15) using traditional techniques is often infeasible due to the demanding computation and memory requirements, a more practical approach sequential minimal optimization (SMO) is used (Platt 1998).

The basic SVM is a binary classifier. There are a couple of approaches to expand this methodology to multi-classification situation such as "one-against-all", "one-against-one", and directed acyclic graph SVM (DAGSVM). In (Hsu and Lin 2002), C. Hsu et al. made a conclusion and a comparison among different approaches.

### 3.5.2. Application of SVM

In order to find the differences between humidity variation caused flooding and temperature variation caused flooding, cell voltage evolutions in the two situations are presented by 3-D figures as figure 8 and figure 9.

The index number of cell is considered as representative of the distance from air entrance to the cell. From the figures, it can be seen that the cells near to the exit are more likely to be fault in the situation of humidity variation; while the cells near to air entrance are more likely to be fault in the temperature variation situation. Hence, the spatial distributions of cell voltages in the two situations are various. If we denote the cell voltages from air entrance to air exit as a vector $\mathbf{v} = [v_1, v_2,\ldots, v_{20}]^T$, then, the spatial distribution information can be kept by the indexes of elements in the vector. Consequently the cell voltages constructed vectors are used as training data for SVM.

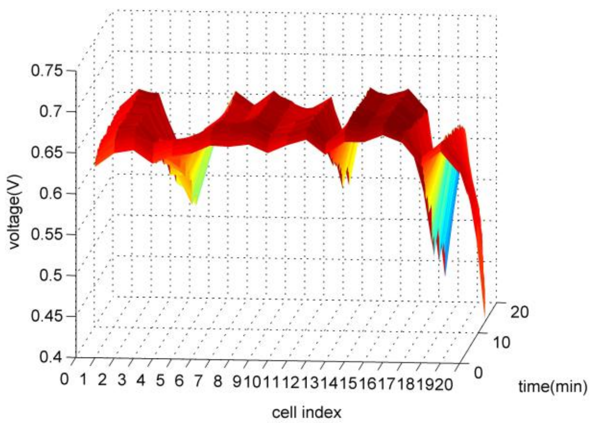


Figure 8: Cell voltages spatial distribution in the flooding evolution caused by humidity variation

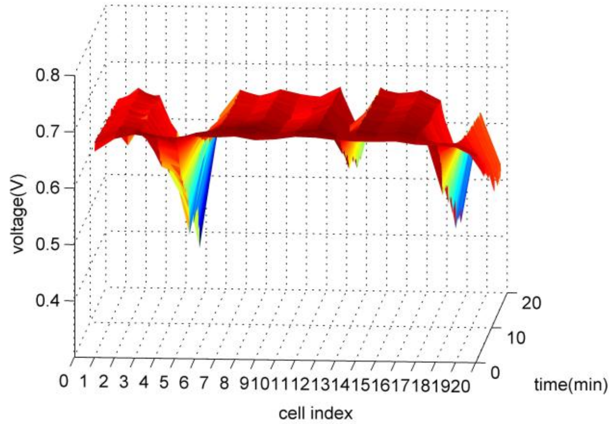


Figure 9: Cell voltages spatial distribution in the flooding evolution caused by temperature variation

### 3.5.3. Evaluation of the classifier

To evaluate the performance and of the SVM classifier, and also to estimate how accurately it will perform in practice, a popular cross-validation methodology which named K-fold cross-validation is used (Kohavi 1995).

In K-fold cross-validation, the total data is randomly divided into $K$ subsets. Of the $K$ subsets, $K-1$ are chosen as training data and the rest one subset is used for test the classifier. The training and test process is then repeated $K$ times, so that each of the $K$ subsets is used once as the test data. The averaged test result is then obtained to evaluate the classifier. The advantage of K-fold cross-validation ensures that all data are used for both training and test, each observation is used for test exactly once.

## 4. RESULTS

### 4.1. Feature selection result

Considering the fault detection, the data in two experiments are analyzed together. The four statistical parameters presented in 3.2 are normalized to $[-1,1]$ firstly. After computing and comparing the criterions of 15 different subsets, the criterion of the subset composed by $\mu$ and $\sigma^2$ is the smallest of all. In other word, the combination of these two parameters is most suitable for fault diagnosis.

### 4.2. Fault definition result

KMC is carried out in the 2-dimension space composed by two features: $\mu$ and $\sigma^2$. The clustering result is as figure 10 shows, the criterion variable $Cr$ of clustering is 0.048, which indicates a good performance of clustering.

After K-means clustering procedure, two centroids are obtained. The “fault zone” and “normal zone” can be divided by drawing a line in the middle of these two centroids as figure 10 and 11 show. For a new sample, it is fault or not can be told by its location is above or below the line.

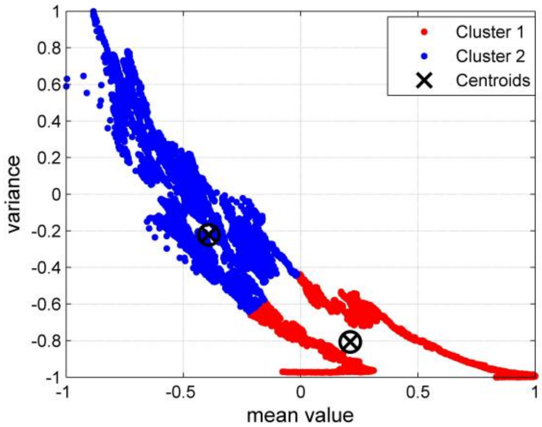

Figure 10: Cluster result on 2-dimenstion space of $\mu$ and $\sigma^2$

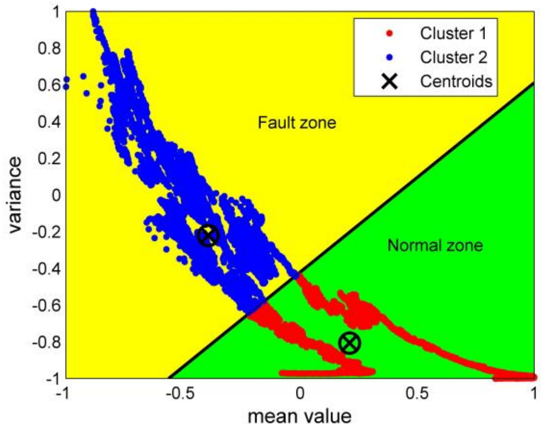

Figure 11: Fault zone and normal zone partition

### 4.3. Fault isolation result

The SVM is carried out on cell voltages composed vectors, Gaussian kernel (as in table 1) is selected with parameter $c$ = 0.5. Error penalty parameter $D$ is chosen to be $10^4$ after several attempts.

In the validation procedure, K-fold cross-validation is used, the number of folds $K$ is set to be 10. On account of that the different folds are chosen randomly, 5 independent tests are preceded. The average error rates of all the tests are kept to be 0. Hence, the performance of SVM for discriminating the causes of flooding is excellent.

After training the total 628 fault vectors, only 9 support vectors which are used for classify new vectors are obtained. Consequently, the computation time will be short enough for online diagnosis.

## 5. CONCLUSION

This paper presents a fault diagnosis approach for PEMFC stack based on statistical analysis. For fault detection, four statistical parameters of cell voltages: $\mu$, $\sigma^2$, $Sk$, $Ku$, are extracted to describe the distribution. Feature selection and KMC are used to define fault zone. For fault isolation, SVM is adopted to classify the cell voltages composed vectors into different fault cause classes. The proposed diagnosis approach can detect flooding fault in fuel cell stack and efficiently classify the causes of flooding fault into two classes: increasing the humidity of inlet air and decreasing the temperature of fuel cell stack. The proposed approach takes account of otherness of cells in the fuel cells stack, and is suited for online diagnosis.

## REFERENCES


Alberto, L. et al., 2008. On-line fault diagnostic system for proton exchange membrane fuel cells. *Journal of Power Sources*, 175, pp.419-429.

Escobet, T. et al., 2009. Model-based fault diagnosis in PEM fuel cell systems. *Journal of Power Sources*, 192(1), pp.216-223.

Güneş, S., Polat, K. & Yosunkaya, Ş., 2010. Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting. *Expert Systems with Applications*, 37(12), pp.7922-7928.

Hernandez, A., 2006. *Diagnostic d'une pile a combustible de type PEFC*. Thesis (PhD). UNIVERSITE DE TECHNOLOGIE DE BELFORT MONTBELIARD.

Hernandez, A. et al., 2006. Fuel cell fault diagnosis : A stochastic approach . In *Industrial Electronics, 2006 IEEE International Symposium on*. pp. 1984-1989.

Hissel, D., Péra, M.C. & Kauffmann, J.M., 2004. Diagnosis of automotive fuel cell power generators. *Journal of Power Sources*, 128, pp.239-246.

Hsu, C.-wei & Lin, C.-jen, 2002. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transaction on Neural Networks*, 13(2), pp.415-425.

Hua, J. et al., 2011. Proton exchange membrane fuel cell system diagnosis based on the multivariate statistical method. *International Journal of Hydrogen Energy*, pp.1-10.

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. pp. 1137-1143.

Lilliefors, H.W., 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, pp.399-402.

Macqueen, J., 1967. SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. In *Proceedings of 5th Berkeley symposium on mathematical statistics and probability*. pp. 281-297.

Platt, J.C., 1998. Sequential Minimal Optimization : A Fast Algorithm for Training Support Vector Machines. *Technical Report MSR-TR-98-14, Microsoft Research*, pp.1-21.

Ramachandran, K.M. & Tsokos, C.P., 2009. *Mathematical Statistics with Applications*, Academic Press.

Samy, I., Postlethwaite, I. & Gu, D.-W., 2011. Survey and application of sensor fault detection and isolation schemes. *Control Engineering Practice*, 19(7), pp.658-674.

Steiner, N.Y. et al., 2011. Non intrusive diagnosis of polymer electrolyte fuel cells by wavelet packet transform. *International Journal of Hydrogen Energy*, 36(1), pp.740-746.

Tian, G. et al., 2008. Diagnosis methods dedicated to the localisation of failed cells within PEMFC stacks. *Journal of Power Sources*, 182, pp.449-461.

Tian, G. et al., 2010. Identification of failed cells inside PEMFC stacks in two cases: Anode/cathode crossover and anode/cooling compartment leak. *International Journal of Hydrogen Energy*, 35(7), pp.2772-2776.

Vapnik, V.N., 1999. An overview of statistical learning theory. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 10(5), pp.988-99.

Wahdame, B. et al., 2008. Analysis of a Fuel Cell Durability Test Based on Design of Experiment Approach. *IEEE Transaction on Energy Conversion*, 23(4), pp.1093-1104.

Wasterlain, S. et al., 2011. Development of new test instruments and protocols for the diagnostic of fuel cell stacks. *Journal of Power Sources*, 196, pp.5325-5333.

Yiakopoulos, C.T., Gryllias, K.C. & Antoniadis, I. a., 2011. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, 38(3), pp.2888-2911.

Yousfi Steiner, N. et al., 2011. Diagnosis of polymer electrolyte fuel cells failure modes (flooding & drying out) by neural networks modeling. *International Journal of Hydrogen Energy*, 36(4), pp.3067-3075.