

Fault diagnosis for PEMFC systems in consideration of dynamic behaviors and spatial inhomogeneity

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

Abstract—In this paper, a data-driven fault diagnostic strategy is designed for Proton Exchange Membrane Fuel Cell (PEMFC) systems. In order to take the system dynamics and spatial inhomogeneity into account, the individual cell voltages measured in a sliding diagnosis window are considered integrally as a diagnostic observation. In the proposed diagnostic approach, a time-series analysis tool, named shapelet transform, is used to extract the discriminative features from the diagnostic observations. The classification tool, named Sphere Shaped Multi-class Support Vector Machine (SSM-SVM), is then carried out in the feature space in order to realize both fault detection and fault isolation. The approach is validated on the experimental testbench. The performance of the approach is evaluated and the necessity of considering both dynamic behaviors and the spatial inhomogeneity is highlighted.

Index Terms—PEMFC systems, data-driven diagnosis, dynamics, cell voltages, shapelet transform, SVM classification.

I. INTRODUCTION

Reliability and durability have been considered as two main obstacles for the widespread commercialization of fuel cell technology [1], [2]. Fault diagnosis, especially online fault diagnosis, has been considered as an efficient solution to improve the performance of fuel cells in those two aspects [3], [4].

As multiple physical and time-scale phenomena are involved in a running fuel cell stack, fuel cell systems have been considered as nonlinear time-varying systems [5]. Designing an efficient diagnosis for fuel cell systems is not a trivial task. During the last two decades, a lot of research has been carried out to propose an effective fault diagnosis strategy. Most research works concern the Proton Exchange Membrane Fuel Cell (PEMFC) systems, since this type of fuel cell has received the most attention in both stationary and mobile applications.

One of the most substantial works is dedicated to building model-based fault diagnosis [6], [7]. These diagnosis approaches rely on the mathematical model of the normal system. When such a model is available, fault diagnosis can be achieved by comparing actual observations with the model prediction. However, the physical model is usually built based on a specific fuel cell system and stack. Considering the differences between the different fuel cell stacks, auxiliary components, and control laws, it is not evident to use one model validated in a system to another one. To apply an existing model for a specific fuel cell systems, usually a lot of work is needed for model or parameter identification and adjustment. To avoid building a physical and complex mathematical model, some attempts have been taken to design

the fault diagnosis by building a “black-box” model. One such well known model is realized by using Neuron Networks (NNS) [8]. The approach is efficient on fault detection for a specific fault. However, it fails on multiple faults isolation in the current state. In addition, some phenomena and processes, such as the spatial inhomogeneity, are hard to describe by using a quantitative model.

Other research works were mainly focused on data-driven fault diagnosis. The complex modeling process can be avoided in data-driven diagnosis, although abundant data are usually needed. In the framework of the data-driven diagnosis, one of the key step is to find the features by which the faults can be detected and isolated.

In [9], multivariate analysis was carried out to extract the features for the PEMFC system which was applied in the bus. In the approach, the required data need to be acquired from more than 20 different sensors. The approach is therefore limited to the demonstration items or the cases where a fair amount of sensors are installed into the systems.

Some experimental instruments and methodologies are employed to extract the diagnosis oriented features. In [2], [10], the Electrochemical Impedance Spectroscopy (EIS), known as a powerful electrochemical characterization tool, is used for fault diagnosis. Some fault discriminative features can be extracted thanks to the abundant information contained in the EIS curves. In [11], the EIS test function is integrated into the DC/DC converter and thus the EIS can be acquired without additional instrument. A classical EIS test usually takes at least 10 min, which makes the EIS based approaches unsuitable for online diagnosis of the fast faults. Recently, a broadband method is proposed to reduce the EIS execution time to a couple of seconds [12]. However, the EIS test requires the fuel cell system to be operated with a static output current. Hence, the EIS tool is still hard to be applied in the fuel cell systems running in dynamic processes.

To avoid monitoring a whole EIS curve, some diagnostic approaches are proposed through time-frequency analysis of the output stack voltage signal. For instance, the wavelet transform, which had been widely used for bearing tool fault diagnosis, was extended and adapted to fuel cell diagnosis [13]. Other time-frequency analysis tools, such as the singularity analysis and empirical mode decomposition, were also studied for diagnosis purposes [14], [15]. These techniques were used to analyze the stack output voltage signal to find the frequency components which are discriminative of the faults concerned. However, a significant change in stack voltage may be driven by input variables or by faults. These frequency analysis based methods are risky, as they may lead to loss in

performance in dynamic situations.

Actually, as research is going deeper and deeper, the fuel cell fault diagnosis requirements are becoming more and more critical. An efficient diagnostic approach needs to satisfy the following criteria to the greatest extent: Multiple faults diagnosis capability; High diagnostic accuracy; Unseen fault recognition; Online capability; Good generalization; Self-adaptation. However, most research approaches still can not meet all these criteria.

Moreover, all the above mentioned studies assume, either deliberately or not, that the fuel cells in a stack are identical. The differences among the different fuel cells are not taken into consideration. However, because of the spatial inhomogeneity on fluid, thermal, and humidity distributions, it has been observed that the individual fuel cell voltages are not strictly identical. Moreover, the distributions are highly influenced by the faults [16]. Hence, it is more reasonable to consider the spatial inhomogeneity in the design of the fault diagnosis approach.

In a recent work [17], fault diagnosis algorithms were developed with considering the individual cell voltages sampled at one time point as an observation for fault diagnosis, such that the spatial inhomogeneity is taken into account. The amplitudes and spatial distributions of cell voltages play the role to provide the fault discriminative information. The pattern classification tools are then employed for fault detection and isolation. The proposed approach was verified to satisfy most of the aforementioned criteria. However, in this consideration, the observations were assumed to be temporally independent. Actually, the neighbor samples are correlated temporally when the fuel cell system is operated in dynamic conditions. In the case where the data are non-stationary, it is more reasonable to investigate the features from the time sequences but a single sample.

In this study, a diagnosis strategy is developed taking the system dynamics and spatial inhomogeneity of the fuel cell stacks. To realize this, the individual cell voltages sampled in a sliding window are considered as an observation for diagnosis. The techniques named shapelet transform and Sphere Shaped Multi-class Support Vector Machine (SSM-SVM) are successively used to extract features from the multiple variate time series and classify the features into the classes representing the health states.

The paper is organized as follows: Section II is dedicated to the presentation of the diagnostic approach. The implementation process is dealt with in the section. In Section III, the experiments and the data used to validate the proposed approach are introduced. Following that, the diagnostic results are provided in Section IV. Finally, the paper is concluded in Section V.

II. DIAGNOSTIC APPROACH

The diagnosis problem is regarded as multi-variate time series classification problem. Considering the spatial inhomogeneity and the time correlations between samples, the individual fuel cell voltages sampled in a time interval are considered

as the variables for diagnosis. A diagnostic observation or a diagnostic window is denoted as

$$T = [T(1), \dots, T(l_w)] = [\mathbf{v}_1, \dots, \mathbf{v}_{l_w}] \quad (1)$$

where \mathbf{v}_k is the vector sampled at time index k and composed by cell voltages of number n_{cell} , as $\mathbf{v}_k = [v_k(1), v_k(2), \dots, v_k(n_{cell})]^T$, l_w is the window width. It can be seen that the observation T is a multi-variate time series.

The training dataset consists of N such observations which are distributed in a variety of classes denoted as $\Omega_0, \Omega_1, \Omega_2, \dots, \Omega_C$. The class label 0 corresponds to the fault free or normal state, while 1, 2, \dots , C correspond to the faults of various types. The numbers of diagnostic observations in these classes are respectively N_0, N_1, \dots, N_C , satisfying $N_0 + \dots + N_C = N$. For an observation T_i ($i = 1, \dots, N$), the class label $g_i \in \{0, 1, \dots, C\}$ is assumed to be known in prior.

A. Features extraction based on shapelet transform

In this study, the diagnosis in dynamic processes is concerned. It is therefore necessary to analyze a time series other than a single sample. One of the most promising recent time series analysis methods consists in finding the shapelets within a dataset. Unlike the time-frequency transformation tools, such as wavelet transform, the discriminative local information can be extracted directly by using this tool without any prior configurations or parameterization. In our case, a shapelet is a subsequence of a diagnosis observation, and is identified as being representative of class membership [18], [19]. Fig. 1 shows an example of a shapelet generated from a diagnosis observation which is a 5-variate time-series. In our approach, the most discriminative shapelets are firstly generated from the training data. Based on the shapelets, the shapelet transform is performed to extract features from an arbitrary diagnostic observation.

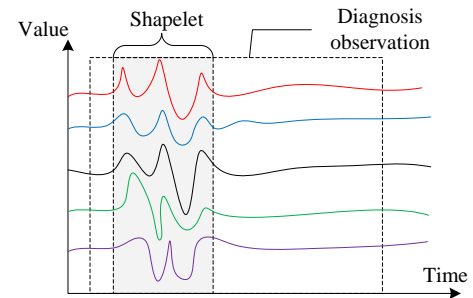


Fig. 1. Schema of shapelet

1) *Generating shapelets*: Finding a shapelet requires generating a set of candidates. Assuming the pursued shapelet width is l_s , the shapelet candidates are defined as the subsequences of the observations in the training dataset. For instance, $l_w - l_s + 1$ shapelet candidates of length l_s for an observation T_i ($i = 1, \dots, N$) can be formulated as

$$SC_{i,j} = \{T_i(j), T_i(j+1), \dots, T_i(j+l_s-1)\}, \quad j = 1, \dots, l_w - l_s + 1 \quad (2)$$

where $SC_{i,j}$ denotes the j th shapelet candidate for T_i . As $SC_{i,j}$ is a subsequence of T_i , its class label is also g_i . Thus,

the total number of the shapelet candidates generated from the training dataset is $(l_w - l_s + 1)N$. Notice that, in [19], the shapelet length is varied between 3 and l_w . In this study, the shapelet length is defined in prior as a fixed value to facilitate calculation without losing performance.

The distance between a shapelet candidate $SC_{i,j}$ ($i = 1, \dots, N$ and $j = 1, \dots, l_w - l_s + 1$) and an observation T_k is the minimum distance between $SC_{i,j}$ and all the shapelet candidates generated from T_k ($k = 1, \dots, N$), as

$$dis(SC_{i,j}, T_k) = \min_{l \in \{1, \dots, l_w - l_s + 1\}} ||SC_{i,j} - SC_{k,l}|| \quad (3)$$

where $||\cdot||$ denotes the Euclidean distance. For each candidate, the distance from it to every observation in the training dataset is calculated. The mean distance from $SC_{i,j}$ to the observations in the same class is calculated as

$$\bar{dis}(SC_{i,j}) = \frac{\sum_{g_k=g_i} dis(SC_{i,j}, T_k)}{N_i} \quad (4)$$

The quality of a shapelet candidate is defined as a measure of the discriminatory power of a shapelet candidate

$$Qual(SC_{i,j}) = \frac{\sum_{g_k \neq g_i} dis(SC_{i,j}, T_k)^2}{\sum_{g_k=g_i} (dis(SC_{i,j}, T_k) - \bar{dis}(SC_{i,j}))^2} \quad (5)$$

It can easily be seen that the quality of a shapelet candidate becomes bigger as it gets closer to the observations in the identical class and far from those in different classes. In other words, the shapelet candidates with higher quality values own more discriminatory power and thus can be selected as the shapelets.

Once all candidates in a class have been assessed, they are sorted in descending order of the quality values. Two shapelet candidates from the same observation may have overlapped indices. In such case, the one with lower quality is removed. From the remaining candidates, some of those with the highest quality values are finally selected as the shapelets.

The whole procedure of shapelet generation is abstracted as Algorithm 1. By applying the algorithm, n_c shapelets are generated for the class labeled c , $c \in \{0, 1, \dots, C\}$. Thus, the total shapelet number is $N_{shapelet} = \sum_{c=0}^C n_c$.

Algorithm 1 Shapelet generation

```

1: Load  $\{T_i | i = 1, \dots, N\}$ ;
2: Initialize  $l_s, n_c$  ( $c = 0, 1, \dots, C$ );
3: for  $c = 0$  to  $C$  do
4:   for  $i = 1$  to  $N$  do
5:     for  $j = 1$  to  $l_w - l_s + 1$  do
6:       Generate  $SC_{i,j}$ ; /* use (2) */
7:       Compute  $Qual(SC_{i,j})$ ; /* use (3)-(5) */
8:     end for
9:   end for
10:  Select  $\arg \max Qual(SC_{i,j})$  as shapelets, denoted as  $\{S_{c,1}, \dots, S_{c,n_c}\}$ ;
11: end for

```

2) *Shapelet transform*: For an arbitrary T , the shapelet features are obtained by calculating the distance from it to all the shapelets generated from the training dataset. The shapelet transform is simply summarized as Algorithm 2. After shapelet transform, the features of the dimensional number $N_{shapelet}$ can be extracted.

Algorithm 2 Shapelet transform

```

1: Get  $T$ ;
2: for  $c = 0$  to  $C$  do
3:   for  $i = 1$  to  $n_c$  do
4:     Compute  $dis(S_{c,i}, T)$  as shapelet feature;
5:   end for
6: end for

```

Notice that several modifications have been made on shapelet transform method to satisfy the requirement of online diagnosis. First, the quality of a shapelet is defined specifically to qualify the discriminatory power. Second, the size of shapelet candidate is fixed to decrease the performing time.

B. SSM-SVM

Once shapelet features are acquired, the classification is carried out in the feature space. Support Vector Machine (SVM) has been considered as a state-of-art classifier thanks to its superior characteristics [20]. In a recent study, one of the SVM variants named SSM-SVM was applied for fault diagnosis [17]. As shown in Fig. 2, the approach achieves classification goal by projecting the data into high-dimensional space by a nonlinear function and seeking class-specific spheres in the projected space. The samples from one specific class are enclosed by the corresponding sphere, while those from other classes are excluded outside. Compared with the traditional SVM classification, SSM-SVM benefits the potential of recognizing the unseen classes (faults), i.e. a new fault type that does not exist in the training dataset [17]. Similar to the traditional SVM method, training SSM-SVM can also be translated into a quadratic problem by introducing a kernel function and playing “kernel trick” [21].

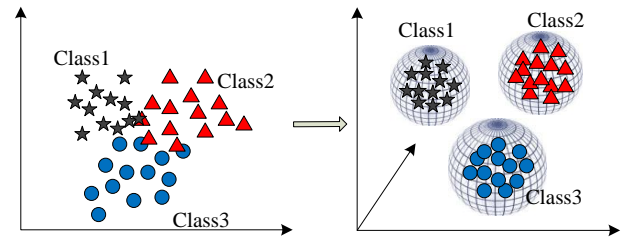


Fig. 2. Schematic of SSM-SVM

In our previous study [17], the diagnostic rule was proposed based on a trained SSM-SVM, which enables not only classifying an example to a known class, but also recognizing an unseen fault class. The principle of the diagnostic rule can be described as follows: For a test example, the distance to each sphere center is calculated in the high-dimensional space. If all the distances exceed a threshold, it is considered that the test

example belongs to an unseen new class. If all the distances are within the threshold, the example belongs to the class with the shortest distance. Mathematically, the class label g of an arbitrary realtime sample T can be expressed as

$$g = \begin{cases} \arg \max_i G_i(d_i(s)) & \text{if } \max_i G_i(d_i(s)) \geq \delta_i \\ \text{new} & \text{if } \max_i G_i(d_i(s)) < \delta_i \end{cases} \quad (6)$$

where s is the vector composed by shapelet features extracted from T , d_i is the distance to the i th sphere center, G_i a smooth decreasing function, δ_i is the predefined threshold. d_i , G_i and δ_i are all determined in the training phase. In the same study, an online learning method was also proposed to achieve the training and online adaptation goals in the whole life span of the concerned stack. The mathematical details about the formulation of SSM-SVM and the proposed diagnostic rule can be found in [17] and the references therein.

C. Implementation of the diagnostic strategy

The diagram representing the implementation of the proposed diagnostic strategy is given in Fig. 3. As conventional data-driven fault diagnosis, the diagnostic strategy proceeds successively from an offline learning stage to an online performing stage.

In the offline stage, the historical data sampled in normal state and in concerned faulty states are collected as the training dataset. Then, the shapelets are generated from the training data. With the shapelets and through the shapelet transform, the shaplet features are extracted from the training data. Based on the shapelet features, SSM-SVM classifier is further trained. The generated shapelets and the trained SSM-SVM classifier are then exported for online stage.

In the online stage, the diagnostic observation is acquired by sliding a diagnostic window of width l_w on the real-time data. For instance, at time $t \geq l_w$, the diagnostic observation T_t is formulated as

$$T_t = [v_{t-l_w+1}, v_{t-l_w+2}, \dots, v_t] \quad (7)$$

where v_t denotes the vector composed by the cell voltages sampled at time t . Although the sliding window size is l_w , the diagnostic window is updated at the sampling frequency. This means that the diagnostic period is 1 s which is sufficiently short to provide an in time diagnostic result for the studied faults.

The shapelet transform is actually achieved by calculating the distance between the shapelet and the diagnostic observation in the sliding window. According to (3), the distance from T_t to $S_{c,i}$ is calculated as

$$\text{dis}(S_{c,i}, T_t) = \min_{j \in \{t-l_w+1, \dots, t-l_s+1\}} \|S_{c,i} - [v_j, \dots, v_{j+l_s-1}]\| \quad (8)$$

It can be seen that $l_w - l_s + 1$ Euclidean distances have to be calculated. In this case, the computation complexity of shapelet transform is $O((l_w - l_s + 1)l_s n_{\text{cell}} N_{\text{shaplet}})$. When the diagnostic window is slid a step ahead to T_{t+1} , the distance from the shapelet $S_{c,i}$ to T_{t+1} is calculated as

$$\text{dis}(S_{c,i}, T_{t+1}) = \min_{j \in \{t-l_w+2, \dots, t-l_s+2\}} \|S_{c,i} - [v_j, \dots, v_{j+l_s-1}]\| \quad (9)$$

Comparing (9) with (8), the Euclidean distances with indices $j \in \{t-l_w+2, \dots, t-l_s+1\}$ have already been computed in the previous diagnostic cycle (8). Only one more distance with $j = t-l_s+2$ needs to be calculated. In this case, the computation complexity turns to $O(l_s n_{\text{cell}} N_{\text{shaplet}})$. Hence, the sliding window procedure can largely reduce the computation complexity.

III. EXPERIMENTS AND DATABASE PRESENTATION

A. Stack and experiment setup

The fuel cell stack studied is a 64-cell PEMFC stack fabricated by the French research organization CEA and dedicated to automotive applications. The stack is operated in dead-end anode, closed cathode mode. The hydrogen pressure is regulated by an inlet side valve. A purge valve functions to expel the vapor and impurities. In the cathode side, the air inlet side compressor and the outlet side valve are regulated to control the air flow rate and pressure. The nominal operating parameters of this stack are summarized in Table I.

TABLE I
NOMINAL CONDITIONS OF THE STACKS

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)	65-70 °C
Anode relative humidity	50%
Cathode relative humidity	50%
Current	90 A
Voltage per cell	0.7 V
Electrical power	4032 W
Active area	180 cm^2

As Fig. 4 shows, the stack is installed into an integration PEMFC system in which various operating parameters, such as pressures, relative humidities, flow rates, and temperatures of the reactants, can be regulated in a sufficiently wide range. A programmable DC load is connected to the output of the stack and emulates various user profiles. The measuring and computing board is equipped with multi-channel Giant Magnetoresistance (GMR) voltage sensors and an FPGA based computing unit. Actually, these two parts are integrated in the form of a 3D integration circuit (IC) in order to meet the critical space requirement for an automotive application. The computing board is devoted to measuring the individual cell voltages and performing the diagnosis approach online. In addition, the measured data can also be transferred and saved in the PC with the help of a Labview based interface. The diagnosis results can be visualized on the screen with the same interface. More details about the testbench can be found in [16].

B. Experimental Data

Besides operating the PEMFC in normal condition, a series of experiments were carried out to emulate the different faults that may occur in the PEMFC system in practice. The different

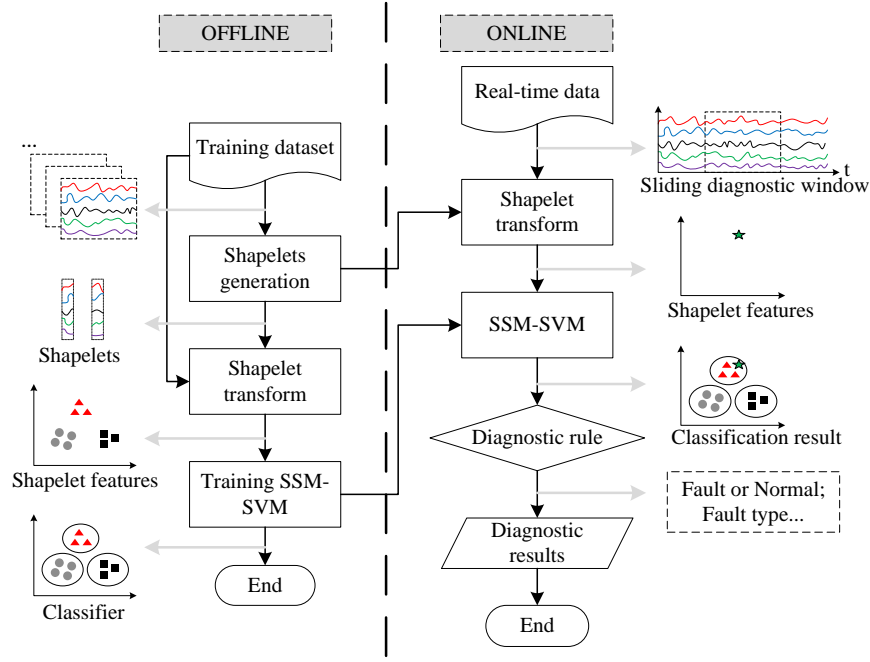


Fig. 3. Implementation diagram for the proposed diagnostic strategy

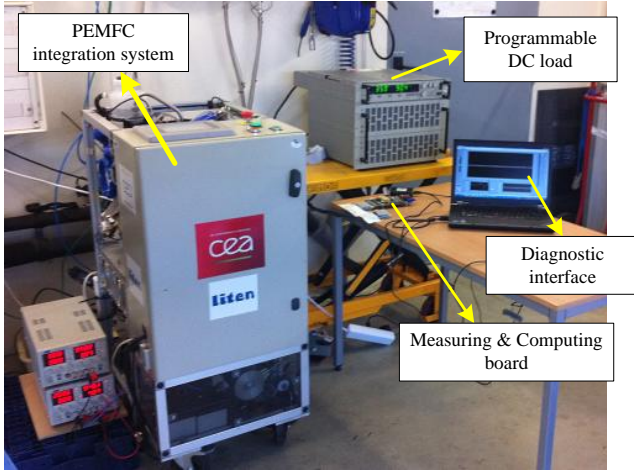


Fig. 4. Overview of the experimental testbench

health states are summarized in Table II. The concerned faults are those related to faulty operations which are recoverable. Actually, those faults are created by varying the input variables, such as pressures, gas flow rates, and temperature, beyond the normal intervals. Each fault was produced by modifying one input variable, while the other ones were kept at the nominal values. For instance, the relative humidity and pressure of air, the stack temperature evolutions are shown in Fig. 5. An efficient online diagnosis can be combined with fault tolerant control or action capable of eliminating the faults before a permanent fault is generated [16].

Through the measuring and computing board, the individual cell voltages were measured in real time and the data were stored into the PC for the usages of training and posterior analysis. The sample frequency was fixed at 1 s to facilitate an

TABLE II
EXPERIMENTS CARRIED OUT IN DIFFERENT HEALTH STATES

Operation	Health state
Nominal condition	Normal state (N1)
Pressure of 1.3 bar on each side	Low pressure fault (F1)
Pressure of 1.7 bar on each side	High pressure fault (F2)
Lower relative humidity	Drying fault (F3)
St. Air 1.5	Low air stoichiometry fault (F4)

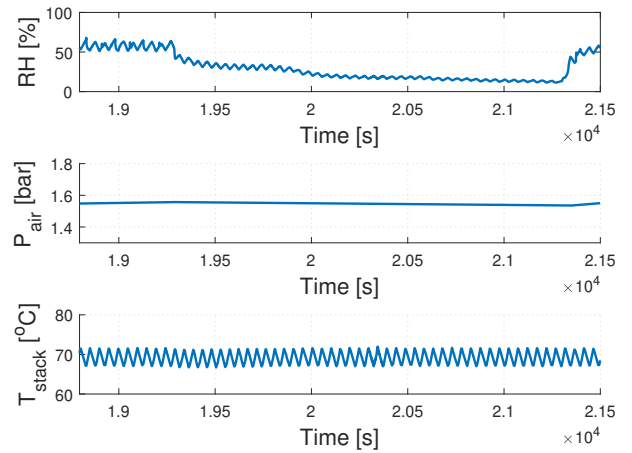


Fig. 5. Relative humidity of air, pressure of air and temperature in F3 fault process

in time diagnosis procedure for most faults and not to cause a too heavy computational burden. An overview of the cell voltages monitored in different health states is shown in Fig. 6(a). In this 1st-generation board, the equipped voltage sensor is capable of monitoring 14 voltage signals. The 14 channels were connected to the cell 1, 3, 5, 6, 7, 9, 10, 11, 12, 18, 19, 20, 21, 22 counting from the cathode side, because it was found experimentally that the cell voltages near to cathode side (gas supply side) are more sensible to the concerned faults. The details of a length of 200 samples are shown in Fig. 6(b). It can be observed that the voltage values differs from cell to cell. Moreover, the voltage signals are not stationary, but present a cyclical dynamic behavior. This can be linked to the purge action which is activated each 90 s. The dynamic evolutions of the cell voltage signals are also different.

In order to obtain the training data and validate the proposed approach, the same experiments were repeated several times. The data in the first run are used to train the diagnostic algorithms, while those in the other runs are used to verify the approach. In this way, the robustness of the approach can be verified, since the data acquired in different runs show some variations because of the different external disturbances, such as environment humidity, temperature, fuel cell ageing effect.

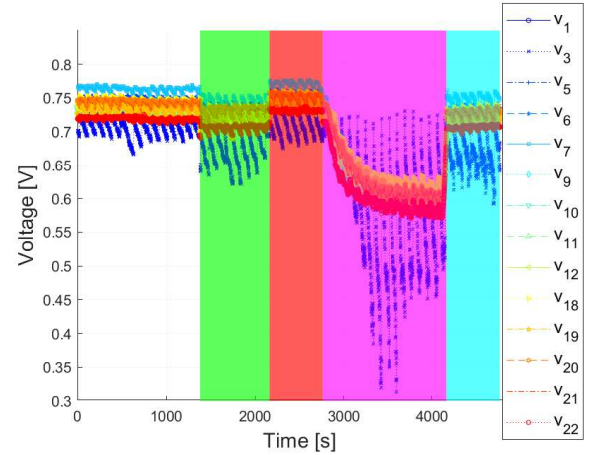
IV. RESULTS AND DISCUSSION

In order to show the superiority of the approach, several tests were carried out and the results were compared. First, the cell voltages measured at one time point are considered as the variables for diagnosis. The dynamics between neighbor samples are thus omitted. In the second test, a time series of the stack voltage measured in a diagnostic window is considered as the diagnosis variable. The spatial inhomogeneity is out of consideration in this case. The last test is dedicated to the proposed diagnostic approach, in which the individual cell voltages measured within a diagnostic window are considered as the diagnosis variables.

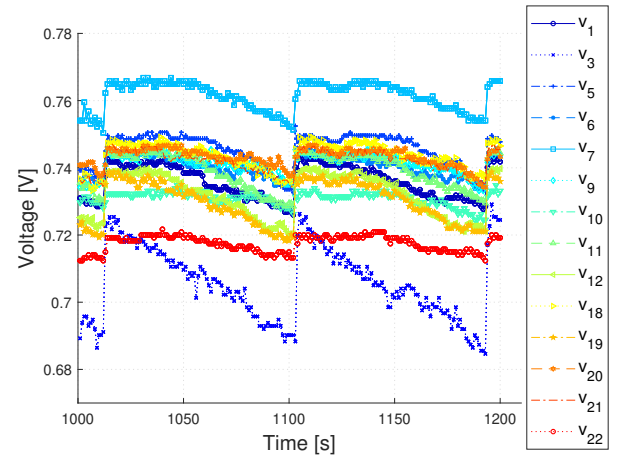
A. Diagnosis without considering dynamic behaviors

In this test, cell voltages measured at a time point are formulated as a vector and considered as the variable for diagnosis. The shapelet transform is not employed in this case, while the SSM-SVM is used directly to process the original data. The dataset used for training consists of 1380 samples in NI class, and 780, 600, 1400, 600 in F1, F2, F3, F4 classes respectively. The parameters of SSM-SVM are initialized using a gradient descent algorithm [22].

After the training process, the overall classification accuracy rate reaches to 99.89%. The classification results on the training data are shown visually in Fig. 7. Only a little part of data in F1 and F4 are misclassified into each other. This can be explained by the fact that similar effects on fuel cell voltages are caused by the low air pressure (F1) and low air stoichiometry (F4). It can also be inferred that the high air pressure fault may also be partly confused by high air flow rate fault. In fact, air pressure and air flow rate are two coupled



(a) Cell voltages at different health states



(b) Details of Cell voltages

Fig. 6. Single cell voltage signals

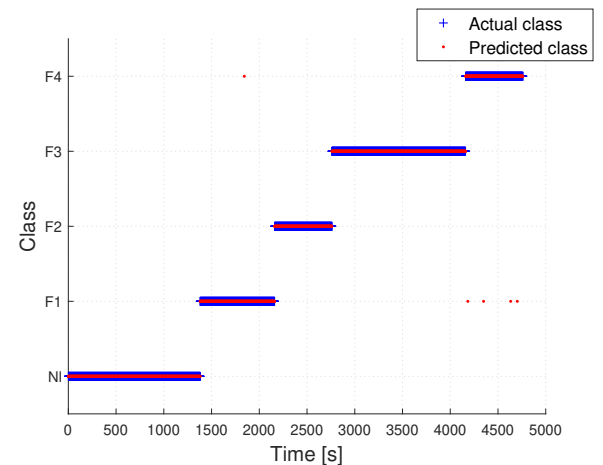


Fig. 7. Training results using cell voltages

variables in the compressor control. The faults concerning on these two variables could be difficult to be separated.

The trained SSM-SVM classifier is then tested on the data obtained from another experiment run. The dataset for test consists of 901 samples in NL, and 794, 306, 2043, 764 in F1, F2, F3, F4. The overall diagnostic accuracy rate being 86.26%. The diagnostic results on the test data are shown in Fig. 8. The confusion matrix on the test results is presented in Table III. The rows in the table show the distribution of the data in an actual class. It can be observed that the false diagnosis happens on the data in NL, F1, F3, and F4, especially the misclassification between F1 and F4 becomes more serious. Obviously, without considering the dynamic behaviors of the cell voltages, the generalization performance is not satisfying.

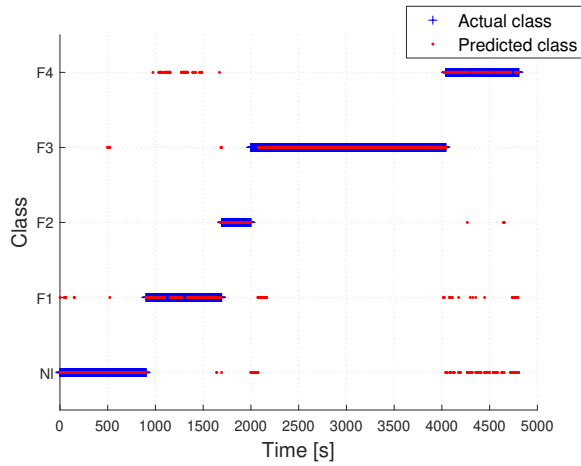


Fig. 8. Diagnostic results without considering dynamics

TABLE III

CONFUSION MATRIX: DIAGNOSTIC RESULTS WITHOUT CONSIDERING DYNAMICS

		Diagnosed classes				
		NL	F1	F2	F3	F4
Actual classes	NL	0.9689	0.0144	0	0.0166	0
	F1	0.0038	0.7778	0.0038	0.0051	0.2096
	F2	0	0	1.0000	0	0
	F3	0.0362	0.0343	0	0.9241	0.0054
	F4	0.3534	0.0340	0.0065	0	0.6060

B. Diagnosis without considering spatial inhomogeneity

In this test, the spatial inhomogeneity of the fuel cell stack is neglected. In other words, the individual cell voltages are assumed to be identical. A series of the stack voltage sampled in a diagnostic window is considered as the diagnosis variable. The implementation of this test follows the flow chart given in Fig. 3 except that the diagnostic observation is an univariate time-series rather than multi-variate time-series.

Considering that the purge cycle is 90 s, the diagnostic window size is set as 100 in order to cover a completed purge cycle. For shapelet transform, the shapelet width is 10. 8 shapelets are generated for each class, and a total number

of 40 shapelets are extracted from the training dataset. With the shapelets, a 40-dimensional feature can be extracted from each diagnostic window. The SSM-SVM classifier is trained and implemented in the shapelet feature space.

According to the results, the accuracy rate on the training data is 93.48%. While the overall diagnostic accuracy rate on the test data is 91.90%. The diagnostic results are shown visually in Fig. 9 and quantitatively in Table IV in the confusion matrix form. Compared with the results of the last test, the diagnostic performance is improved regarding the diagnosis accuracy rates for each class. However, a significant part of data in F4 is misclassified, which results in a low diagnostic rate for this fault type.

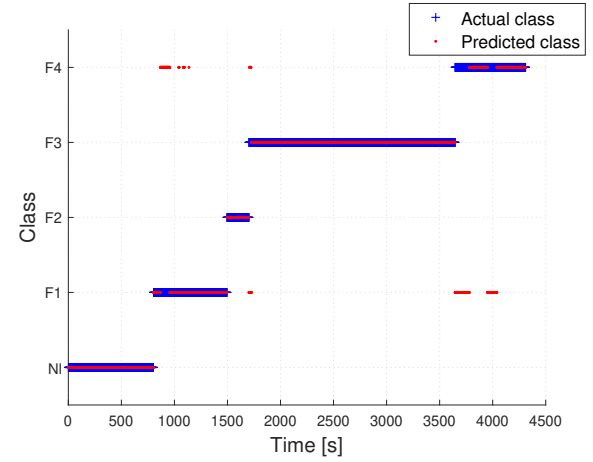


Fig. 9. Diagnostic results without considering spatial inhomogeneity

TABLE IV

CONFUSION MATRIX: DIAGNOSTIC RESULTS WITHOUT CONSIDERING SPATIAL INHOMOGENEITY

		Predicted classes				
		NL	F1	F2	F3	F4
Actual classes	NL	1.0000	0	0	0	0
	F1	0	0.8514	0	0	0.1486
	F2	0	0	1.0000	0	0
	F3	0	0.0057	0	0.9876	0.0067
	F4	0	0.3338	0	0	0.6662

C. Diagnostic considering both dynamics and spatial inhomogeneity

The third test concerns the approach proposed in this study. The diagnosis is implemented based on the individual cell voltages sampled in a diagnostic window, such that both the spatial inhomogeneity and the dynamic behaviors of the cell voltages are taken into account. The implementation process follows exactly the flow chart given in Fig. 3. As in the previous test, the diagnostic window is set at 100. The width of shapelet is 10. 8 shapelets are generated from the training data in each class. With the shapelets, the shapelet features extracted from the training data are shown visually in Fig. 10. It can be seen that the features exhibit class discriminative characteristics. The SSM-SVM is trained in the shapelet feature space and

a classification accuracy rate of 99.98% is obtained on the training dataset.

After implementing the trained approach on the test data, the overall diagnosis accuracy rate of 96.13% is obtained. While the detailed results are given visually in Fig. 11 and quantitatively in Table V. Compared with the previous two tests, the performance is improved with respect to both the overall diagnosis accuracy rate and those for each class. In particular, the diagnosis accuracy rate for the data in F4 reaches almost 90%, which is considered as an acceptable level.

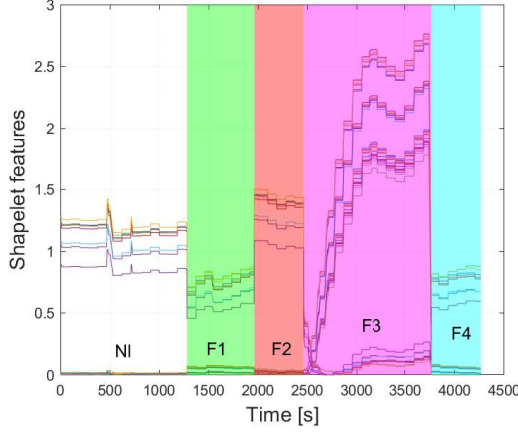


Fig. 10. Shapelet features extracted from the training dataset

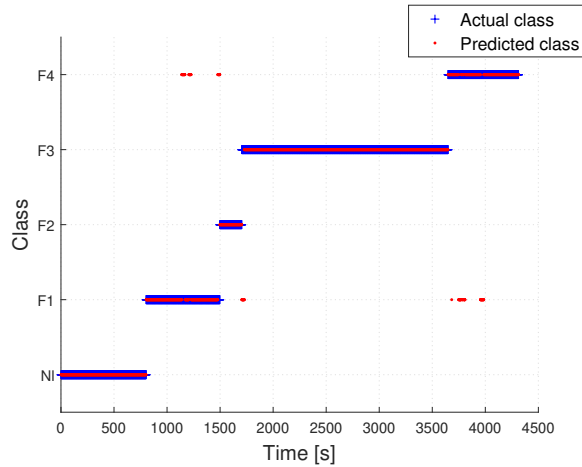


Fig. 11. Diagnostic results using shapelet features from cell voltages

TABLE V
CONFUSION MATRIX: DIAGNOSTIC RESULTS USING SHAPELET FEATURES FROM CELL VOLTAGES

		Predicted classes				
		NI	F1	F2	F3	F4
Actual classes	NI	1.0000	0	0	0	0
	F1	0	0.8993	0	0	0.1007
	F2	0	0	1.0000	0	0
	F3	0	0.0123	0	0.9877	0
	F4	0	0.1113	0	0	0.8887

D. Discussion

In this study, the faults caused by variation of one operating parameter are put much importance, although the other faults that caused by several operating parameters can also occur. In fact, the probability of the latter faults is much lower than that of the former ones. In addition, it is always the case that the faults starts with a fault caused by single abnormal operating parameter.

More faults can be encountered in practice. The proposed approach can be utilized for the diagnosis of more faults only if the data corresponding to the concerned faults are collected in prior [16].

The stack concerned in this study is required to be operated near to the nominal current value. In the case where dynamic load is concerned, the training data should contain sufficient dynamic conditions in order to obtain a reliable diagnosis result. Although the implementation is the same, the acquirement of training data should be put more importance and attention.

Although training the proposed approach usually takes a relatively long time (several minutes), performing the trained diagnostic model is sufficiently fast for online use. According to our experience, the performing time for one diagnosis cycle is within 1 ms by using a PC and Matlab environment. Meanwhile, using the designed IC, the executing time is limited to 10 ms. In this study, the computation time of the proposed diagnosis approach satisfies the requirement of online use and it is not the main concern.

It should be noted that even the voltages of a part of cells over the stack were measured and handled, a satisfying result can still be obtained. As a big cell number is concerned, hundreds cells for instance, some dimension reduction technique can be optionally considered to reduce the complexity of the problem [23].

In the proposed approach, the diagnosis is regarded as a multi-variate time series classification problem. Shapelet transform was selected as the tool to extract the features because it has been regarded as one of the most promising recent methods on time series analysis. It is probable that other time series analysis methods could also be used to extract the features and realize this classification task. From a theoretical point of view, an increasing attention has been paid to the research on multi-variate time series classification.

Although fuel cell diagnosis is talked about in this paper, the proposed approach should not be limited to the application of fuel cell diagnosis. In fact, the diagnosis for many practical systems can be regarded as a multi-variate time series classification problem so long as the necessary experimental data are available.

V. CONCLUSION

The paper has presented and experimentally demonstrated a novel data-driven fault diagnostic strategy for PEMFC systems. Taking both the spatial inhomogeneity and the dynamics into account, the individual cell voltages monitored in a sliding window were considered as the diagnostic observation. The shapelet transform was employed to extract the class discriminative features, which were fed to the SSM-SVM for diagnostic oriented classification procedure.

The experimental test results show that the high diagnosis accuracy rate and robust results can be acquired using the proposed approach. The performing time satisfies the requirement for online implementation. Through comparison, it is also shown that the diagnostic performance is lowered without either considering the spatial inhomogeneity or dynamic behaviors. Combines with the specially designed IC, the strategy is promises to be used in the practical PEMFC systems. Moreover, the proposed approach has the potential to be extended to other diagnosis cases where the dynamics of multiple variables is involved.

REFERENCES

- [1] Y. Wu, E. Breaz, F. Gao, D. Paire, and A. Miraoui, "Nonlinear performance degradation prediction of proton exchange membrane fuel cells using relevance vector machine," *IEEE Transactions on Energy Conversion*, vol. 31, no. 4, pp. 1570–1582, Dec 2016.
- [2] C. de Beer, P. Barendse, and P. Pillay, "Fuel cell condition monitoring using optimized broadband impedance spectroscopy," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5306–5316, Aug 2015.
- [3] J. Wu, X. Z. Yuan, J. J. Martin, H. Wang, J. Zhang, J. Shen, S. Wu, and W. Merida, "A review of PEM fuel cell durability: Degradation mechanisms and mitigation strategies," *Journal of Power Sources*, vol. 184, no. 1, pp. 104 – 119, 2008.
- [4] C. de Beer, P. Barendse, P. Pillay, B. Bullecks, and R. Rengaswamy, "Classification of high temperature pem fuel cell degradation mechanisms using equivalent circuits," *IEEE Trans. Ind. Electron.*, vol. PP, no. 99, pp. 1–1, 2015.
- [5] F. Bianchi, C. Kunusch, C. Ocampo-Martinez, and R. Sanchez-Pena, "A gain-scheduled lpv control for oxygen stoichiometry regulation in pem fuel cell systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 5, pp. 1837–1844, Sept 2014.
- [6] S. de Lira, V. Puig, J. Quevedo, and A. Husar, "LPV observer design for PEM fuel cell system: Application to fault detection," *Journal of Power Sources*, vol. 196, no. 9, pp. 4298 – 4305, 2011.
- [7] D. Rotondo, R. M. Fernandez-Canti, S. Tornil-Sin, J. Blesa, and V. Puig, "Robust fault diagnosis of proton exchange membrane fuel cells using a takagi-sugeno interval observer approach," *International Journal of Hydrogen Energy*, vol. 41, no. 4, pp. 2875 – 2886, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360319915301166>
- [8] N. Yousfi Steiner, D. Hissel, P. Moçotéguy, and D. Candusso, "Diagnosis of polymer electrolyte fuel cells failure modes (flooding & drying out) by neural networks modeling," *International Journal of Hydrogen Energy*, vol. 36, no. 4, pp. 3067–3075, Feb. 2011.
- [9] J. Hua, J. Li, M. Ouyang, L. Lu, and L. Xu, "Proton exchange membrane fuel cell system diagnosis based on the multivariate statistical method," *International Journal of Hydrogen Energy*, vol. 36, no. 16, pp. 9896 – 9905, 2011.
- [10] Z. Zheng, R. Petrone, M. Pera, D. Hissel, M. Becherif, and C. Pianese, "Diagnosis of a commercial pem fuel cell stack via incomplete spectra and fuzzy clustering," in *Industrial Electronics Society, IECON 2013 - 39th Annual Conference of the IEEE*, Nov 2013, pp. 1595–1600.
- [11] Z. Zheng, M.-C. Péra, D. Hissel, M. Becherif, K.-S. Agbli, and Y. Li, "A double-fuzzy diagnostic methodology dedicated to online fault diagnosis of proton exchange membrane fuel cell stacks," *Journal of Power Sources*, vol. 271, pp. 570 – 581, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378775314012117>
- [12] R. A. Nazer, V. Cattin, P. Granjon, M. Montaru, and M. Ranieri, "Lassical eis vs broadband identification comparison based on a well-known reference impedance," in *The International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium*, 2013.
- [13] E. Pahon, N. Y. Steiner, S. Jemei, D. Hissel, and P. Moçotéguy, "A signal-based method for fast PEMFC diagnosis," *Applied Energy*, vol. 165, pp. 748 – 758, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261915016578>
- [14] D. Benouioua, D. Candusso, F. Harel, and L. Oukhellou, "Fuel cell diagnosis method based on multifractal analysis of stack voltage signal," *International Journal of Hydrogen Energy*, vol. 39, no. 5, pp. 2236 – 2245, 2014.
- [15] C. Damour, M. Benne, B. Grondin-Perez, M. Bessafi, D. Hissel, and J.-P. Chabriot, "Polymer electrolyte membrane fuel cell fault diagnosis based on empirical mode decomposition," *Journal of Power Sources*, vol. 299, pp. 596 – 603, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378775315303013>
- [16] Z. Li, R. Outbib, S. Giurgea, D. Hissel, and Y. Li, "Fault detection and isolation for Polymer Electrolyte Membrane Fuel Cell systems by analyzing cell voltage generated space," *Applied Energy*, vol. 148, pp. 260 – 272, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306261915003682>
- [17] Z. Li, R. Outbib, S. Giurgea, and D. Hissel, "Diagnosis for pemfc systems: A data-driven approach with the capabilities of online adaptation and novel fault detection," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5164–5174, Aug 2015.
- [18] L. Ye and E. Keogh, "Time series shapelets: a new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 947–956.
- [19] J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 289–297.
- [20] H. Xue, S. Chen, and Q. Yang, "Structural regularized support vector machine: A framework for structural large margin classifier," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 573–587, April 2011.
- [21] P. Y. Hao and Y. H. Lin, "A new multi-class support vector machine with multi-sphere in the feature space," in *Proceedings of the 20th international conference on Industrial, engineering, and other applications of applied intelligent systems*, ser. IEA/AIE'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 756–765.
- [22] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Mach. Learn.*, vol. 46, no. 1-3, pp. 131–159, Mar. 2002. [Online]. Available: <http://dx.doi.org/10.1023/A:1012450327387>
- [23] Z. Li, R. Outbib, D. Hissel, and S. Giurgea, "Data-driven diagnosis of PEM fuel cell: A comparative study," *Control Engineering Practice*, vol. 28, no. 0, pp. 1 – 12, 2014.