Fuel Cells

Fault Diagnosis of Practical Polymer Electrolyte Membrane (PEM) Fuel Cell System with Data-driven Approaches

L. Mao[1]*, L. Jackson[1], S. Dunnett[1]

[1] Department of Aeronautical and Automotive Engineering, Loughborough University, Leicestershire, LE11 3TU, UK

Abstract

In this paper, data-driven approaches are applied to identify faults of a practical PEM fuel cell system. Signal processing approaches are selected and employed to multiple sensor measurements, including methodologies reducing the dimension of the original dataset, and techniques extracting features. Both supervised and unsupervised techniques are applied in this study to investigate the robustness of the diagnostic procedure. Moreover, due to the fact that a series of features can be extracted from these sensors, the singular value decomposition (SVD) technique is applied to select features providing better diagnostic performance. Results demonstrate that with features selected from SVD, fuel cell system faults can be detected more effectively, and various fuel cell faults can also be discriminated with good quality. From the findings, conclusions are made and further work suggested.

Keywords: Data-driven Approach, Fault Diagnosis, Feature Selection, PEM Fuel Cell, Signal Processing Technique

1 Introduction

The demand for less polluting energy generation technologies has increased rapidly in order to deal with the issues raised due to environmental change. Among these technologies, the Proton Exchange Membrane Fuel Cell (PEMFC) receives much attention, as it can convert the energy from fuel into electricity with high efficiency and produce only water and heat. In the past few years, PEMFC's have been used in many applications, such as automotive, stationary power, and consumer electronics. However, durability and reliability of the fuel cell are still two main barriers blocking its wider application.

As a solution to guarantee the safety of the fuel cell during its operation and extend the fuel cell lifetime, fault diagnosis of fuel cells has attracted more and more research. A set of diagnostic approaches have been proposed in the last few decades to detect and isolate faults of fuel cells under various operation conditions. Based on the approaches used for fault diagnosis, these studies can be loosely divided into two categories, model-based approaches and data-driven techniques [1–22]. In model-based approaches, a mathematical model should be developed to simulate and predict the system performance, by evaluating residuals between the model and system results, the fault occurrence can be detected. For this purpose, various models can be generated, including the white box models containing a series of space differential equations for simulating the physical behaviors of the system, the black box models deriving relationships between the system inputs and outputs from experimental data, and gray box models combining both advantages of the physical white box models and the empirical black box models. Petrone et al. [1] summarized the state of art of model-based techniques used for fuel cell fault diagnosis; in white box models, parameter identification models, observer-based models, and parity space models are presented, while neural network, fuzzy logic, adaptive neuro-fuzzy inference system (ANFIS) and support vector machine (SVM) are introduced as black box models. Forrai et al. [2], Fouquet et al. [3], Ordonez et al. [4] and Robio et al. [5] employed the equivalent-circuit model to simulate the impedance measurement from fuel cells, by configuring the model parameters using test data, the evolution of model parameters can be used to detect and isolate the fault occurrence. Lee et al. [6] and Zeller et al. [7] used quasi-static circuit models to determine the fuel cell output voltage, and identified model parameters which can be employed to evalu-

[*] Corresponding author, l.mao@lboro.ac.uk

ate the different losses during fuel cell operation. Riascos et al. [8,9] derived the qualitative and quantitative relationship between fuel cell inputs and outputs from a fuel cell model. From the relationship a Bayesian network was developed and used to detect and isolate the fuel cell faults during operation. In other studies the relationship between the fuel cell inputs and outputs was determined using a neural network, the fault diagnosis could then be performed using the trained neural network [10–12].

Due to the complexity of a fuel cell system and its electrochemical behaviors, it is difficult to develop a reliable fuel cell model and derive the relationship between the inputs and outputs. In this case, data-driven approaches can be used as an alternative method for fuel cell fault diagnosis. With this approach, the features indicating fuel cell states are extracted from the measurements and used for fault detection. Zheng et al. [13] provided an overview of non-model based approaches for fuel cell fault diagnosis, in which the current state of data-driven approaches was introduced. Statistical methods like principal component analysis (PCA), fisher discriminant analysis (FDA), and signal processing methods such as fast Fourier transform (FFT), short-time Fourier transform (STFT) and wavelet transform (WT) were considered. Narjiss et al. [14] proposed a method to detect membrane humidification and reactive gas feeding based on the variation of impedance measurements. Furthermore, Onanena et al. [15] employed self organizing maps to separate fuel cell normal operations from faulty conditions (flooding and drying out), the level of faults, such as minor and moderate faults, could also be classified. Zheng et al. [16] also proposed a method to identify the fuel cell flooding from its normal operation based on impedance measurements, several features were extracted from impedance and the fuzzy clustering approach was then applied to determine the fuel cell state. As fuel cell output voltage can be used to indicate the fuel cell performance, several approaches have been proposed to identify fuel cell faults based on voltage measurements. Frappe et al. [17] proposed a method to distinguish fuel cell normal states with faulty conditions (flooding and drying out) based on differentials between voltage measurements at different places. Steiner et al. [18] used wavelet packet transform to extract features from fuel cell voltage, and fuel cell faults (flooding in the paper) could be identified using the extracted features. Li et al. [19] employed FDA to extract features from fuel cell voltage, and a Gaussian mixture model (GMM) was applied to identify different fuel cell states using the extracted features. With this method, normal, transition and faulty fuel cell operations could be classified. Li et al. [20] further compared the performance of different data-driven approaches in fuel cell fault diagnosis using output voltage, in the analysis, PCA, kernel PCA (KPCA), FDA, kernel FDA (KFDA) were used to extract features from the fuel cell voltage, GMM, SVM, k-nearest neighbour (KNN) were employed to make classifications with the extracted features. Results showed that FDA and SVM could give higher performance and less computation cost than the other methods. Several papers also investigate the applica-

tion of data-driven approaches in real fuel cell systems, where a set of sensors are placed to capture the fuel cell performance. Placca et al. [21] and Hua et al. [22] proposed PCA to deal with a large amount of data from fuel cell system for fault diagnosis.

According to previous studies regarding fuel cell fault diagnosis based on data-driven approaches, only limited research has been devoted to performing fault diagnosis using a large amount of sensor measurements, which is usually the case in the practical fuel cell systems due to the fact that a single sensor may not provide complete information about the systems. As multiple sensor measurements can be collected from the fuel cell performance, diagnostic approaches cannot be applied directly due to the high computation cost, since the collected dataset may have high dimensions and contain various information. Thus before performing diagnosis, methodologies reducing the dimension of the original dataset should be performed to reduce computation cost and complexity without losing significant information in the measurements. Based on previous research [12, 19, 20–23], approaches for this purpose include PCA, KPCA, FDA, and KFDA. Moreover, from previous studies, a series of features can be extracted from the sensor measurements using wavelet analysis [29–32], but these features may provide different diagnostic performances. Considering the fact that involvement of all features will make the diagnostic analysis complex and time-consuming, the features with better diagnostic performance should be selected using proper techniques, this will also be studied in this paper.

Based on the previous studies, it can be concluded that although several research has been performed in fuel cell diagnosis, only few sensor measurements are included in the analysis, this may cause the misleading diagnostic results, especially when discriminating various fuel cell failure modes, as limited sensors may not give the complete information about the whole fuel cell system. Given there are many sensors installed in the practical fuel cell system, this sensor information needs to be better utilized and hence is one focus of this paper. Meanwhile, the computational time and complexity should also be considered in the diagnosis of practical fuel cell systems, especially with involvement of many sensor measurements in the analysis, which may require extracting information from original measurements and selecting features containing the most useful information in the analysis, these issues will be addressed in this paper. Moreover, considering the fact that the knowledge gap between various fault diagnostic approaches and practical fuel cell system performance has not been fully studied, the performance of a diagnostic framework including several data-driven approaches is investigated with sensor measurements from practical fuel cell systems in this paper.

The contribution of this paper is to propose a general fault diagnostic framework, which can deal with multiple sensor measurements, and the performance of the fault diagnostic framework is studied using test data from a practical fuel cell system. Compared to the fault diagnostics using EIS, etc., this

Mao et al.: Fault Diagnosis of Practical Polymer Electrolyte Membrane (PEM) Fuel Cell System with Data-driven Approaches

Fuel Cells

fault diagnostic framework can identify the fuel cell faults without interrupting the fuel cell operation, while measurement of EIS requires a relatively specialized technique with expensive and bulky equipment [33–36]. Moreover, the fuel cell fault diagnosis under dynamic load current conditions can be performed with the fault diagnostic framework, this makes this approach more suitable for practical fuel cells system subjected to dynamic load conditions, especially in automotive applications.

The fault diagnostic framework is robust since the methodologies in the framework can be selected based on the practical applications, i.e. without or without prior knowledge about possible fuel cell faults. To achieve the better fault diagnostic performance, the singular value decomposition (SVD) technique is used to select the features from wavelet-related analysis for the diagnostic analysis. The practical fuel cell system and recorded measurements under its steady state operation will be presented in Section 2. Section 3 will describe data-driven approaches used in the analysis, including KPCA and KFDA for dataset dimension reduction, wavelet packet (WP) transform for feature extraction, and singular value decomposition (SVD) for feature selection. In Section 4, results from the proposed framework will be given to demonstrate its performance in identifying faults of practical fuel cell systems, especially the effectiveness of the features selected from SVD in detecting and isolating fuel cell faults. Finally, conclusions will be given based on the diagnostic results.

2 Description of Practical Fuel Cell System

In the study, the evaporatively cooled (EC) fuel cell system is used for the analysis. The EC fuel cell system is designed for high volume, low cost manufacturing applications, where it utilizes the benefit of heat from vaporization instead of circulating coolant through the cells, thus reducing complexity, mass and cost. Figure 1(a) depicts a block diagram of the EC fuel cell system, and Figure 1(b) shows a typical EC fuel cell system.

During EC fuel cell operation, hydrogen and air are injected into the anode and cathode electrodes of the stack, respectively. Water is also injected and partly evaporated within the stack to form a liquid/vapour mix, which can remove heat from the stack. Moreover, the stack exhaust is condensed in the heat exchanger to recover and return sufficient water to the water tank in order to maintain water balance in the system.

During operation of the EC fuel cell system, multiple sensors are placed at the inlets and outlets of the stack can be used to capture the stack performance, including stack voltage, pressure, flow rate, temperature, etc. The reason of using these sensors is that they are usually installed in the practical fuel cells for collecting complete information. Since it is not fully understood which sensors can make the most contribution to the fuel cell fault diagnostics, all the recorded measurements are included in the analysis herein, which are listed in Table 1, where # expresses the stack number, as the EC fuel cell system used in the study contains two stacks, and each stack includes 300 fuel cells, and the locations of these sensors are depicted in Figure 2. It should be noted that the pressure, temperature, gas flow rate, voltage and current can be measured directly using the sensors shown in Figure 2, while the cathode stoichiometry is calculated using the measured gas flows from anode and cathode sides.

From Table 1, it can be seen that 22 sensor measurements can be monitored during the fuel cell system operation. The use of a dataset with such a high dimension in computations will be time consuming and computationally expensive.
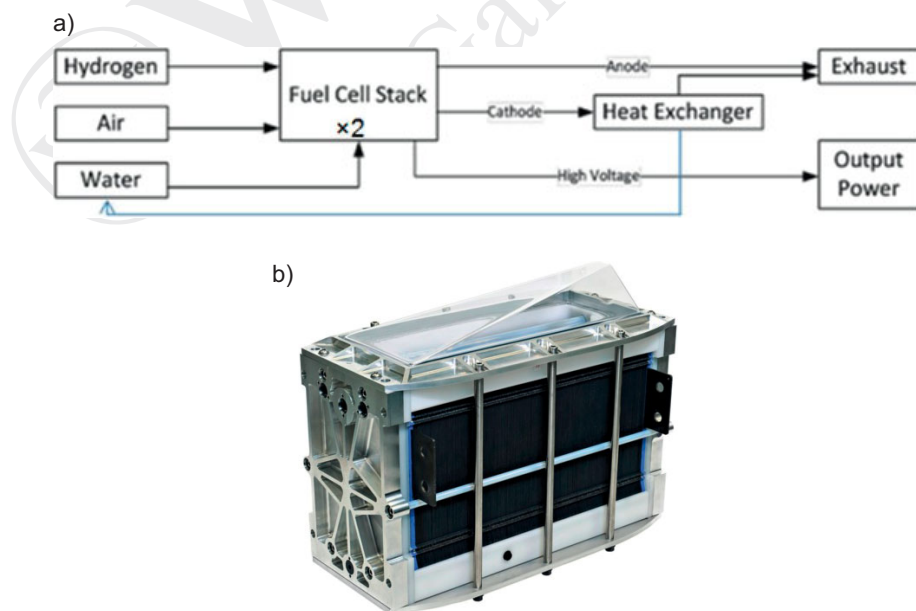

Fig. 1 a) Block diagram of EC fuel cell system; b) EC fuel cell system.

Table 1 Recorded measurements from the EC fuel cell system.

Sensor measurement	unit	Sensor measurement	unit
Stack voltage	V	Cathode inlet pressure #1	mbars
Load current	A	Cathode inlet pressure #2	mbars
Anode reactant flow	SLPM	Cathode outlet pressure #1	mbars
Anode inlet pressure #1	mbars	Cathode outlet pressure #2	mbars
Anode inlet pressure #2	mbars	Cathode stoichiometry	N/A
Anode outlet pressure #1	mbars	Cathode inlet temperature #1	°C
Anode outlet pressure #2	mbars	Cathode inlet temperature #2	°C
Cathode air inlet flow	SLPM	Cathode outlet temperature #1	°C
Water inlet temperature	°C	Cathode outlet temperature #2	°C
Primary water inlet pressure #1	mbars	Primary water inlet flow #1	SLPM
Primary water inlet pressure #2	mbars	Primary water inlet flow #2	SLPM

Therefore, appropriate data-driven approaches should be selected to provide reliable diagnostic results with reasonable computing cost, and these methodologies will be described in the next section.

3 Data-driven Approaches

With recorded measurements from the EC fuel cell system, the data-driven approaches can be applied to extract features for fault identification. However, due to the high dimensions of the measurement dataset (22 sensors in this study), appropriate approaches should be used to reduce its dimension before extracting features. Furthermore, features containing the most useful information will be selected to identify system abnormal performance. The procedure of fault diagnostics is represented with a flowchart depicted in Figure 3. In this section, the data-driven methodologies used in the analysis will be described.

3.1 Reduction of High Dimensional Dataset

The dimension of the original dataset should be reduced before performing fault diagnosis, several techniques, such as PCA, KPCA, FDA, and KFDA can be used for this purpose [20–23]. From previous studies, PCA and KPCA can be used for unsupervised dimension reduction (where the prior knowledge about the state of the measurement datasets cannot be accessed), while FDA and KFDA are more suitable for supervised reduction (where the state of collected measurements is known in advance). Moreover, KPCA and KFDA can provide more reliable diagnostic performance than PCA and FDA due to their ability in dealing with non-linear problems, that is, KPCA and KFDA is extremely useful when the data has a complicated structure and cannot be represented in a linear subspace [20–23].

In this study, as prior knowledge of dataset state can be accessed (normal or faulty), KFDA is used herein for fuel cell fault diagnosis. Moreover, the performance of KPCA is also investigated in the paper to study robustness of the proposed
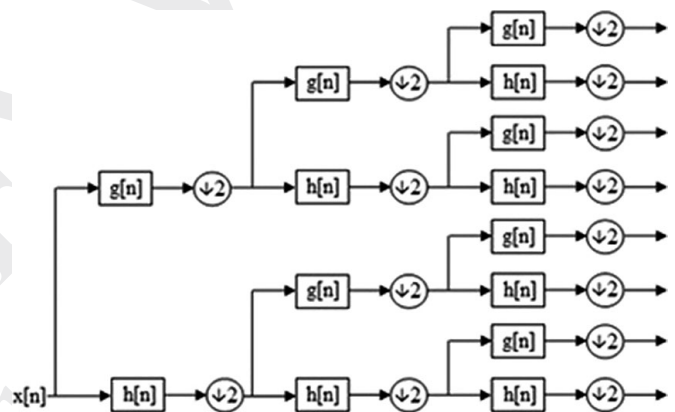

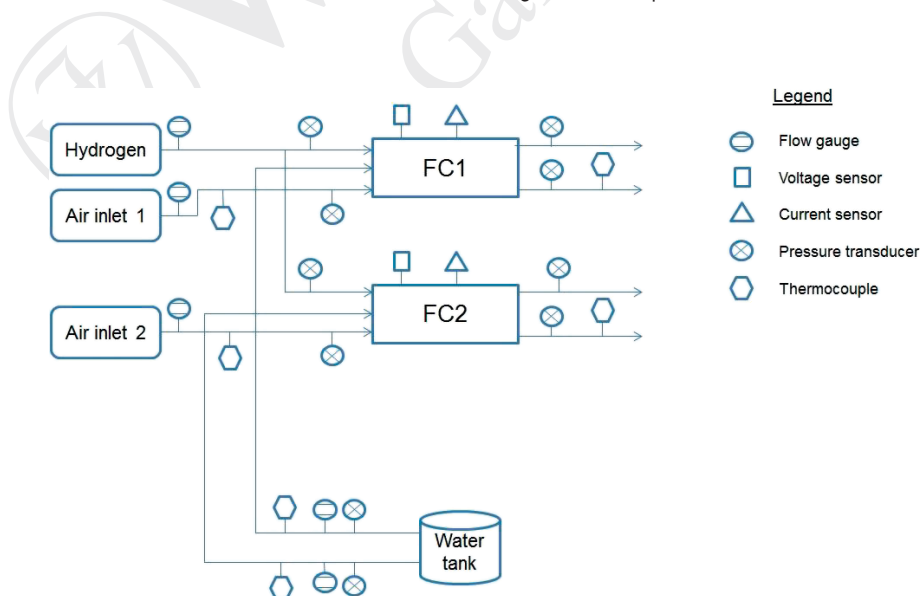Fig. 3 Wavelet packet transform over 3 levels.

Fig. 2 Location of sensors used in EC fuel cell system.

diagnostic procedure, as it may not be possible to access the state of the dataset in advance in practical fuel cell systems.

The general idea of KPCA is a non-linear mapping of the original data to a higher-dimension space, where they vary linearly, then KPCA is applied in the new space, its procedure can be described as:

(i) Construct the kernel matrix K;

$$K_{i,j} = \kappa\left(x_i, x_j\right) \tag{1}$$

where κ is the kernel function, $\kappa\left(x_i, x_j\right) = \phi(x_i)^T\phi(x_j)$, ϕ creates linearly independent variables from the original data x_i and x_j;

(ii) Modify the kernel matrix to generate $\tilde{K}$;

$$\tilde{K} = K - 1_nK - K1_n + 1_nK1_n \tag{2}$$

where 1_n is the n × n matrix where all elements take the value of $1/n$;

(iii) Calculate the highest L eigenvalues and corresponding eigenvectors ($a_1, a_2, \cdots, a_L$) from $\tilde{K}$;

(iv) Project the data to the new direction;

$$z_l = \sum_{i=1}^{n} a_{ln}\kappa\left(x_i, x\right) \tag{3}$$

where z_l is the lth element of the projected vector ($l \in 1, 2, \ldots L$), a_{ln} is the corresponding value in the above calculated eigenvectors.

With the procedure, the high dimension dataset can be reduced significantly without losing useful information, which can be evaluated using Eq. (4),

$$\frac{\sum_{i=1}^{L} \lambda_i}{\sum_{i=1}^{n} \lambda_i} < T \tag{4}$$

where λ_i is the ith principal component, n is the number of total principal component, and L is the selected number of principal components (with the selected principal components, useful information will not be lost), T is the threshold value (0.95 is selected in this case based on previous studies [20, 21]).

Based on Eq. (4), four principal components are selected herein to express the information in the original dataset with size of 22, and in the following analysis, the fault diagnosis will be performed at these four principal directions.

While in KFDA, a direction should be found to maximize the class separation so that the original dataset can be projected into the determined direction, thus when using KFDA, the data should be labeled to represent different classes (like classes C_1 and C_2), their mean value can be written as:

$$m_i = \frac{1}{l_i}\sum_{n=1}^{l_i} x_n^i \tag{5}$$

where l_i is the number of data points of class C_i, x_n^i is the nth data point with C_i.

Since the goal of KFDA is to find the large separation to different classes while keeping the small in-class variance, this can be formulated as:

$$J(w) = \frac{w^T S_B^{\emptyset} W}{w^T S_w^{\emptyset} W} \tag{6}$$

where $S_B^{\emptyset} = (m_2^{\emptyset} - m_1^{\emptyset})(m_2^{\emptyset} - m_1^{\emptyset})^T$, is the between-class covariance matrix, $S_w^{\emptyset} = \sum_{i=1,2} \sum_{n=1}^{l_u} (\emptyset(x_n^i) - m_i^{\emptyset})(\emptyset(x_n^i) - m_i^{\emptyset})^T$, is the within-class covariance matrix, and $m_i^{\emptyset} = \frac{1}{l_i} \sum_{n=1}^{l_i} \emptyset(x_n^i)$.

In this process, the kernel function is used to compute the mapping $\emptyset(x_n^i)$, this can be written as $k(x, y) = \emptyset(x).\emptyset(y)$, with this technique, w in Eq. (6) can be written as:

$$w = \sum_{i=1}^{1} \propto_i\emptyset(x_i) \tag{7}$$

By differentiating $J(w)$ with respect to w, and setting it to zero, the projection of new data can be obtained as:

$$y(x) = (w \cdot \emptyset(x)) = \sum_{i=1}^{1} \propto_i k(x_i, x) \tag{8}$$

In this study, as the state of the dataset is obtained in advance, corresponding class labels (two classes in this case, including normal and faulty states) are assigned to different datasets, and the original dataset is projected to the direction calculated using Eq. (8).

3.2 Feature Extraction and Generation

After reducing the dimensions of the original dataset, features can be extracted to indicate system fault occurrence. In this paper, wavelet packet (WP) transform is used to extract coefficients from the signals, which are extracted from the original dataset using KPCA or KFDA. The principle of WP transform is letting the original data pass through filters to get low-pass results (approximation) and high-pass results (detail). Compared to conventional wavelet transform, WP transform can give more coefficients as both approximation and detail will be filtered to get the next level approximation and detail, respectively [24, 25], this can be expressed in Figure 3, where WP transform is applied to the three levels using Daubechies wavelet ('db1' in the Matlab wpdec function), as it can provide sufficient information for fault diagnosis with reasonable computation time based on the trail-and-error method. For each transform, both low pass coefficient g[n] and high pass coefficient h[n] are obtained.

After obtaining these wavelet coefficients, features can be generated. In this study, normalized energy is generated with the extracted wavelet coefficients, as it can be calculated easily and provides the energy decomposition by wavelet coeffi-
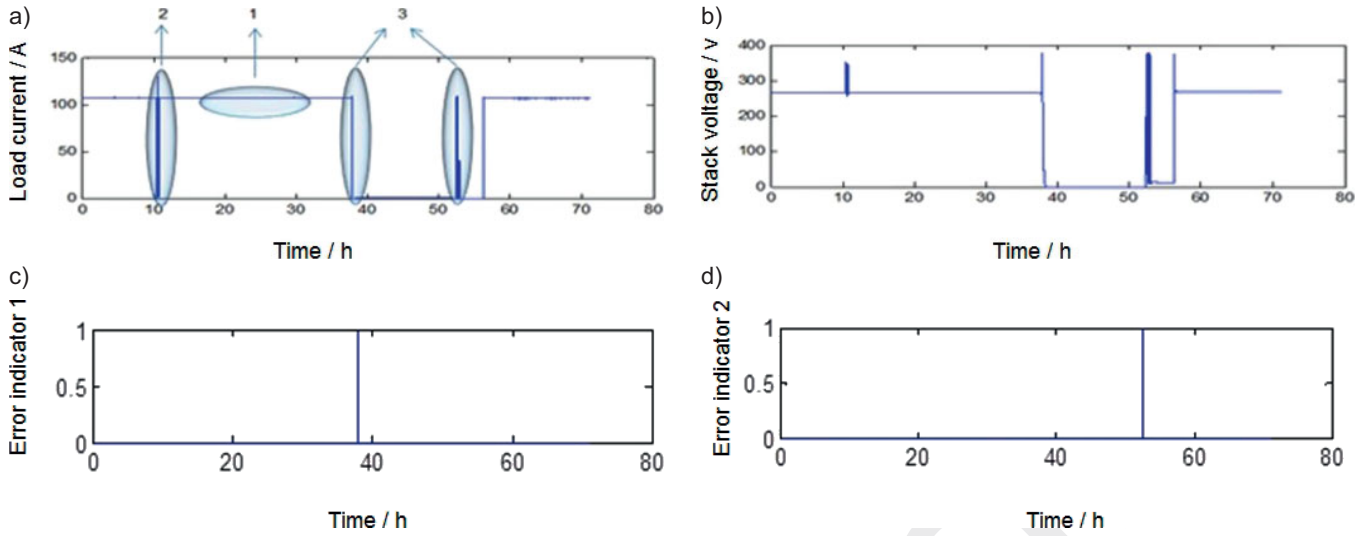
Fig. 4 Wavelet packet transform over 3 levels.

cients [18, 19]. Normalized energy can be computed with the following Eq. (9):

$$E^P = \frac{1}{N_P} \sum_{j \cdot k} \left| C_{j,k}^P \right|^2 \tag{9}$$

where E^P is the normalized energy for specific wavelet packet p, N_P is the number of coefficients in wavelet packet p, and $C_{j,k}^P$ is the coefficient in wavelet packet p.

With Eq. (9), normalized energies can be calculated and used as features to indicate system faults. However, it should be noted that a set of features can be obtained (in this case, 14 normalized energies can be calculated, since WP transform is applied over 3 levels, giving 14 sets of low-pass and high-pass coefficients, which is depicted in Figure 3), as these features might contain misleading information such as measurement noise, process uncertainties, etc., and confusing results may be given when all the features are included in the analysis. Therefore, features containing the most useful information should be selected and used to indicate the system abnormal performance. Singular value decomposition (SVD) is used for this purpose.

3.3 Feature Selection

SVD is a method to find the best approximation of the original data with fewer dimensions. Based on previous studies, SVD can be used effectively to separate noise from measurement signals, and sort useful information [26–28]. In this study, SVD is applied to identify and sort the dimensions containing the most useful information, which can be used as a basis to select features with high significance, its procedure is described as follows:

(i) Generate normalized energy matrix E_t for each dataset, which includes the four principal components after applying KPCA to the original sensor measurements ($E_t = [E_{t,1}^P \; E_{t,2}^P \; \ldots, \; E_{t,M}^P]$, E_t is B × M matrix, B is number of wavelet packets, M is number of signals in class t);

(ii) Apply SVD to E_t to get $E_t = U\Sigma V$ (where Σ is B × M singular value matrix, the rank of normalized matrix M_t is the number of non-zero singular values, U is B × B orthogonal matrix with first M_t columns spanning the range of normalized energy matrix, and V is M × M matrix, which can be used to re-generate the E_t with the first M_t singular values);

(iii) Plot distribution of singular vectors U at the first M_t columns;

(iv) Find energy peaks in above singular vector distribution from each class;

(v) If same energy peaks are included in all classes, they will be selected for the classification; if not, find parent energies containing the most energy peaks for the analysis.

In the next section, the effectiveness of SVD in selecting features for fuel cell fault diagnostics will be studied, the performance of features selected from SVD will be compared to those features randomly selected from the feature dataset.

4 Result and Discussion

4.1 Description of Sensor Measurements

In this section, fault diagnosis of the EC fuel cell system will be performed using the data-driven approaches outlined. In the test, several sensors (listed in Table 1) are installed in a EC fuel cell system to collect the information continuously, generally the constant load current is applied to the system to get the steady-state condition, but during the test the system may experience some quick fluctuations in the load current (transient load), some of these fluctuations will cause the fuel cell fault and lead to the voltage drop of the system. In the test, a new fuel cell stack (contain a series of single fuel cells) is used to replace the existing stack every few months to keep the stable performance of the system. Figure 4 depicts some measurements from the system, including load current (Figure 4(a)), stack voltage (Figure 4(b)), and fault indicators

Mao et al.: Fault Diagnosis of Practical Polymer Electrolyte Membrane (PEM) Fuel Cell System with Data-driven Approaches

Fuel Cells

(Figure 4(c) and Figure 4(d)). In Figure 4(c) and Figure 4(d) it should be noted that 'Error Indicator' is used to indicate the system fault, different faults can be identified with a corresponding error indicators jump to a value of 1. Moreover, only the fuel cell current and voltage measurements are depicted here to illustrate different states of the fuel cell system during operation, while in the analysis, all the 22 sensors are used for diagnosis, which are listed in Table 1.

In this study, the purpose of fault diagnosis is to distinguish fuel cell healthy state and faulty state, and moreover to classify various faults encountered in the tests (in this case two fuel cell faults are studied, including membrane drying out and hydrogen supply problem), which is usually obtained by post-analysis in practical fuel cell systems.

Furthermore, it can be observed from Figure 4(a) that during system operation, transient load current or constant load current can be observed in the fuel cell healthy state (parts 1 and 2 highlighted in Figure 4(a)), which means fuel cell healthy state may provide different fuel cell responses, thus it is necessary to further classify the fuel cell healthy states based on various fuel cell responses. With the varying load current shown in Figure 4(a) and the system fault indicators shown in Figure 4(c) and Figure 4(d), the measurement signals are divided into three groups, these are highlighted in Figure 4(a). It should be mentioned that in this study, each sensor measurement contains 300 sample points, and the sampling frequency in the test is 1Hz, indicating the each sensor measurement contains the fuel cell performance of 300 s (5 min).

(i) Normal part (with constant load current and without fault);
(ii) Fault part (with transient load current and fault);
(iii) Unknown part (with transient load current but without fault), 'unknown' is defined since this data has transient current load similar to the faulty state, but actually it belongs to the healthy state as no fault is observed.

From the sensor measurements, the normal and fault parts can be distinguished clearly since faulty signals have completely different behavior due to the transient current. However, as the unknown part also contains transient load current, it will be difficult to separate this part from the faulty case. Therefore, the performance of data-driven approaches can be studied by separating faulty signals from unknown parts, since unknown parts represent system normal operation without any fault, based on the error indicators.

It should be noted that the datasets used for the analysis are obtained from the EC fuel cell system lifetime test, which includes 25 normal cases, 25 unknown cases, and 11 faulty cases collected at different system operating times, this can be used to better study the robustness of diagnostic procedure shown in Figure 2, as the effects of several external factors, such as room temperature, air humidity, sensor reliability, are already included in the measurements.

4.2 Results and Discussion

From descriptions in Section 3, KPCA and KFDA are used to reduce the dimension of the measurement dataset (22 in this study, which is listed in Table 1). In this analysis, the first four principal components are selected to represent the original dataset with KPCA, thus the original data will be projected to the first four principal directions, while with KFDA, the principal direction for maximizing the class separation is determined to represent the labelled dataset.

As shown in Figure 3, after applying KPCA/KFDA, WP transform is used to the dataset at four principal directions
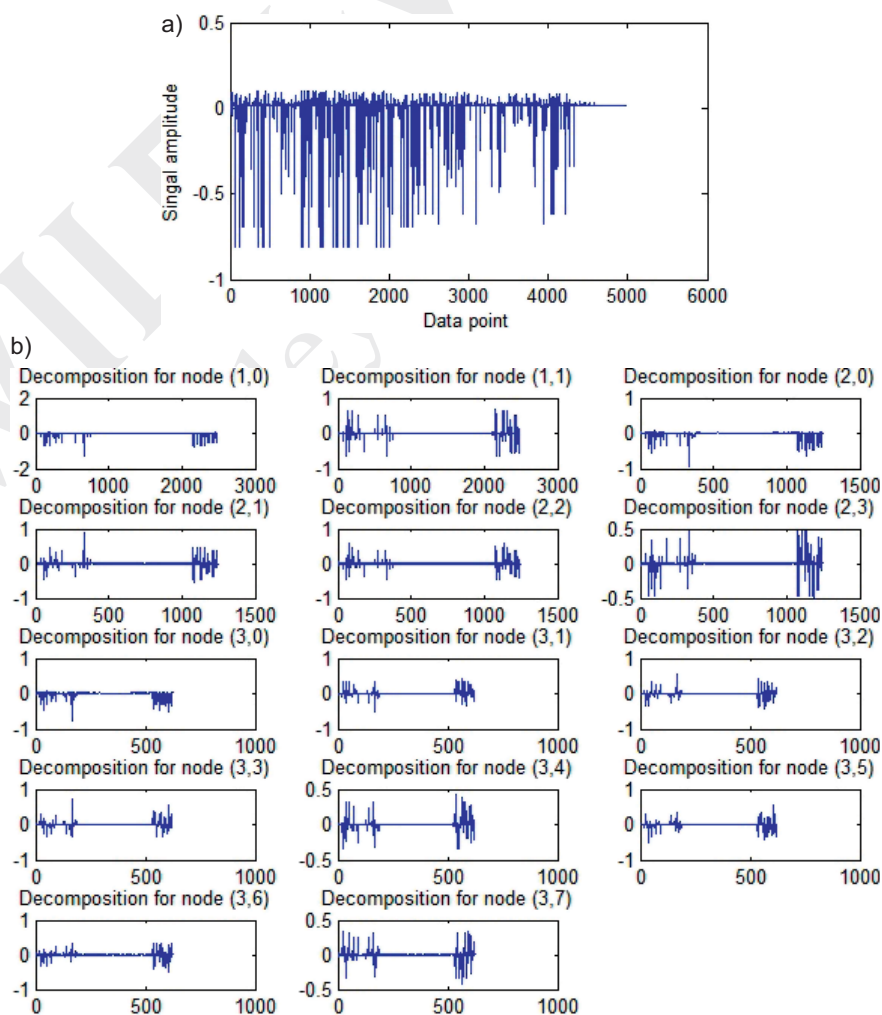

Fig. 5 a) Fuel cell current; b) Fuel cell voltage; c) Error indicator 1 (caused by membrane drying out); d) Error indicator 2 (caused by hydrogen issue).

over three levels, which gives 14 normalized features at each principal direction. For illustration purpose, Figure 5(a) depicts one signal at the 1st principal direction after applying KPCA to the original dataset with normal state, and Figure 5(b) shows the 14 decompositions of the signal after using WP transform.

In the study, two features will be used to discriminate various fuel cell system states shown in Figure 4(a). The reason for using only two features is that these two features can express most information contained in the original signal, this is evaluated using Eq. (10).

$$\frac{\sum_{i=1}^{L} \zeta_i}{\sum_{i=1}^{M_t} \zeta_i} < T \tag{10}$$

where ζ_i is the ith singular value decomposed with SVD, M_t is the number of total non-zero singular values, and L is the selected number of singular values, T is the threshold value (0.95 is selected herein).

In order to study the effectiveness of SVD in fuel cell fault diagnostics, the performance of two features selected using SVD are compared to the features selected randomly. The performance of two random features in distinguishing various fuel cell system states using supervised and unsupervised data dimension reduction techniques is investigated (shown in Figure 6(a1), Figure 6(a2), Figure 6(a3), Figure 6(a4) and Figure 6(b1), Figure 6(b2), Figure 6(b3), Figure 6(b4), Figure 7(a), Figure 7(b)). It should be mentioned that the original dataset (containing 22 sensors) is projected at four principle directions with KPCA, for each principal component, 14 features will be obtained using Eq. (9), and two features are selected randomly at each principal direction and used for classification. While from KFDA, as only 1 principal direction is used, thus two features are selected randomly for the classification, and results using random selected features from KFDA are depicted in Figure 6. Moreover, only two sets of features are shown herein to illustrate the performance of randomly selected features, as there exists a large number of possible combinations for randomly selected features.

From Figure 6, with selected feature sets from unsupervised dimension reduction technique (KPCA), different fuel cell system states are distinguished at four principal directions. It can be seen that normal and unknown states can be roughly distinguished using the randomly selected feature set. However, the faulty state cannot be discriminated from the other two states, and most faulty cases herein are misclassified into normal states. Moreover, it can be seen from Figure 7(a) and Figure 7(b) that even by labelling the different dataset classes when reducing the dimension of the original dataset using KFDA, fuel cell faulty condition cannot be discriminated from the normal conditions using randomly selected features.

As described in Section 4.1, the main purpose of this study is to distinguish the faulty state from the other two states, especially from the unknown state (since the unknown state has transient load current, which makes its response similar to

the faulty state, but it actually represents the fuel cell healthy state, as no fault is observed in the unknown state), but features selected randomly cannot provide reasonable diagnostic results, indicating that faults in the fuel cell system cannot be detected with good quality using randomly selected features.

Performance of features selected using SVD technique (from KFDA) is shown in Figure 8, where it can be seen that with selected features using SVD, different states of fuel cell datasets can be separated with good quality, indicating that by labelling dataset stats in data dimension reduction technique, selected features from SVD can determine the fuel cell state accurately, and three different fuel cell states can be separated clearly.

Moreover, the performance of features selected using SVD technique (from KPCA) is depicted in Figure 9(a), Figure 9(b), Figure 9(c) and Figure 9(d) to illustrate the robustness of the diagnostic procedure without the prior knowledge of dataset state, where classification of various fuel cell system states is performed at four principal directions. It can be seen that consistent diagnostic results can be obtained using features selected from SVD, even without the prior knowledge about the fuel cell faults, most fault signals can be separated from the other two groups, while normal and unknown cases have similar feature values, indicating the healthy state of the system. However, it should be noted that 5 unknown cases are misclassified to the faulty state (this is also observed in Figure 8, where KFDA is used to reduce the dimension of the original dataset), the possible reason maybe that the current approaches are not able to discriminate these states; also these unknown states may actually represent some kind of fuel cell faults which are not indicated in the practical fuel cell system, this requires further investigation.

By comparing the diagnostic performance of KPCA and KFDA, the robustness of the diagnostic procedure can be studied. With prior knowledge about fuel cell faults, the diagnostic procedure using KFDA can give perfect classification results, which is shown in Figure 7, while in cases where prior knowledge about fuel cell faults cannot be achieved, which is always true in practical fuel cell systems, the diagnostic procedure using KPCA can still identify the fuel cell faults with good quality, thus the diagnostic procedure can be used for fault diagnosis at different conditions depending on practical fuel cell systems.

The performance of selected features (from KPCA) is further evaluated using a confusion matrix, which allows the visualization of the performance of various feature sets , Tables 2–4 lists the confusion matrix using two randomly selected features (shown in Figure 6(a1), Figure 6(a2), Figure 6(a3), Figure 6(a4) and Figure 6(b1), Figure 6(b2), Figure 6(b3), Figure 6(b4)) and features selected from SVD (Figure 9(a), Figure 9(b), Figure 9(c), and Figure 9(d)). It should be noted that results from the four principal directions are averaged for each feature scenario. Moreover, as in the actual case, the fuel cell system can only have healthy or faulty state, and the unknown state defined in the analysis belongs to the healthy state, as no fault is observed in this state.

Mao et al.: Fault Diagnosis of Practical Polymer Electrolyte Membrane (PEM) Fuel Cell System with Data-driven Approaches **Fuel Cells**
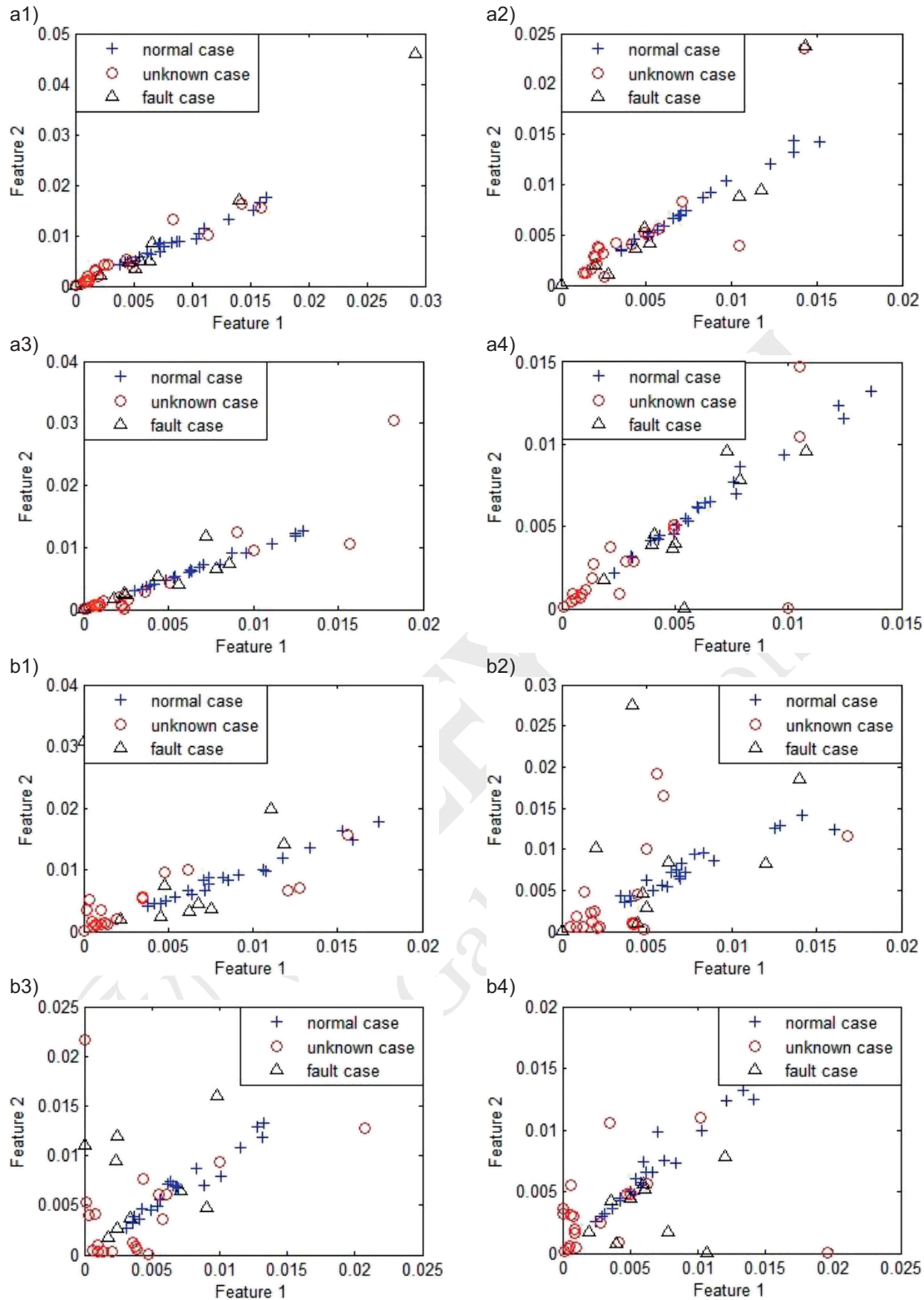
ORIGINAL RESEARCH PAPER

Fig. 6   a) The 1st principal component after applying KPCA; b) Corresponding wavelet packet decomposition.

It can be seen from the results that with selected features from SVD, significant improvement can be achieved in distinguishing unknown states and faulty states, the rate of correct classification of three categories (normal, unknown and faulty)

using two random feature sets (shown in Table 2 and Table 3) are 88%, 20%, 36.3% and 84% 28%, 54.5%, while the rate of correct classification is 100%, 80% and 81.8% for normal, unknown and faulty cases, respectively (shown in Table 4),
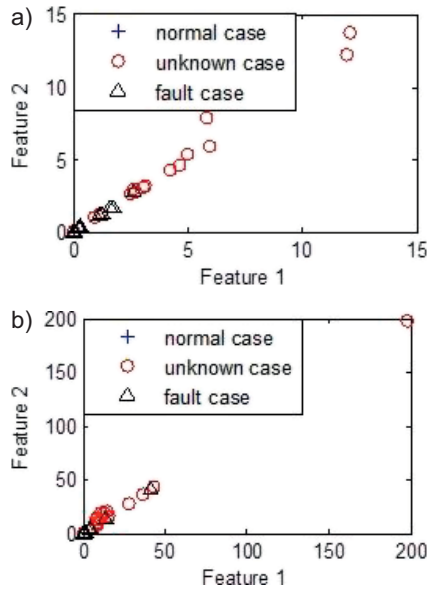
Fig. 7 a1) Classification results (from KPCA) at 1st principal direction; a2) Classification results (from KPCA) at 2nd principal direction; a3) Classification results (from KPCA) at 3rd principal direction; a4) Classification results (from KPCA) at 4th principal direction; b1) Classification results (from KPCA) at 1st principal direction; b2) Classification results (from KPCA) at 2nd principal direction; b3) Classification results (from KPCA) at 3rd principal direction; b4) Classification results (from KPCA) at 4th principal direction.

which further confirms the necessity of selecting features in the fuel cell fault diagnosis. To further improve the performance of diagnostic procedure, fewer sensors can be selected
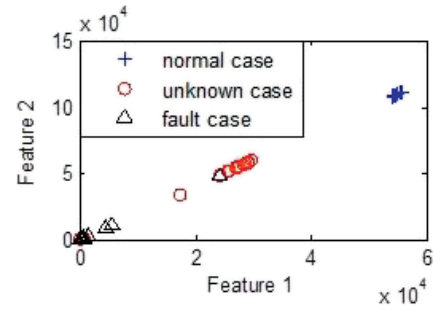

Fig. 8 a) Classification results (from KFDA) with two random feature sets; b) Classification results (from KFDA) with two random feature sets.

Table 2 Confusion matrix with randomly selected features set 1 (shown in Figure 5a).

		Predicted	
		Healthy	Faulty
Actual	Normal	0.88	0.12
	Unknown	0.20	0.80
	Faulty	0.64	0.36

and used in diagnostic analysis, which should be more sensitive to the system performance change due to different fuel cell faults, and removal of less sensitive sensor from analysis can also reduce the noise influence, which can be from the measurement procedure, environment, or sensor itself. This work is currently under investigation.
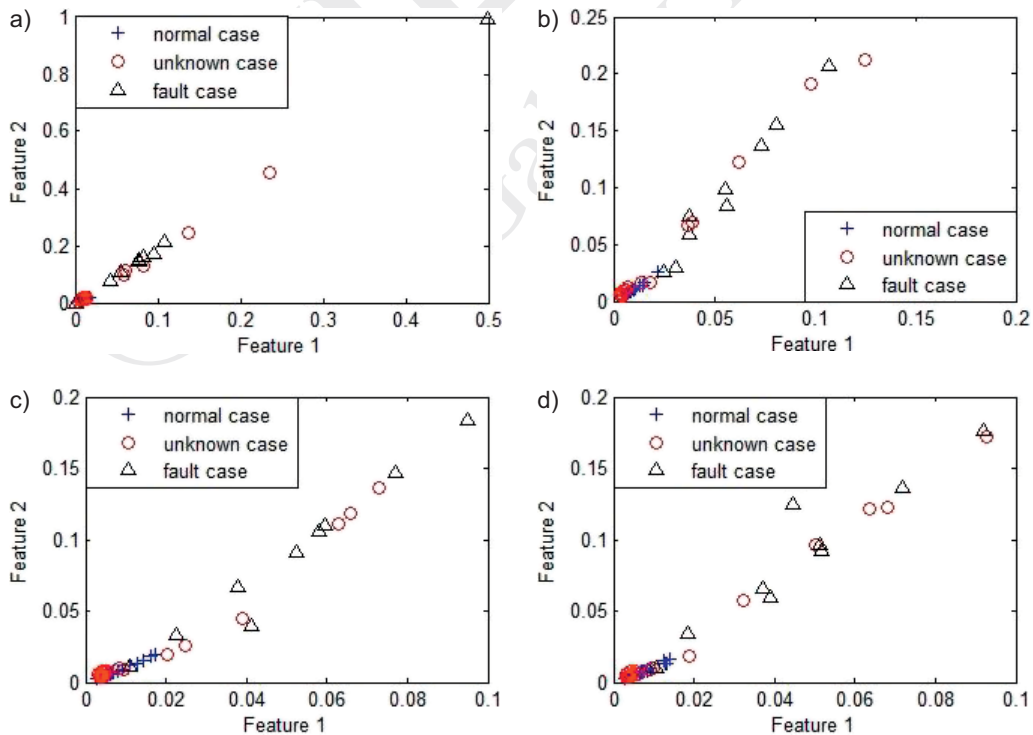

Fig. 9 Classification results (from KFDA) with feature set selected using SVD.

Mao et al.: Fault Diagnosis of Practical Polymer Electrolyte Membrane (PEM) Fuel Cell System with Data-driven Approaches **Fuel Cells**

ORIGINAL RESEARCH PAPER

Table 3 Confusion matrix with randomly selected features set 2 (shown in Figure 5b).

		Predicted	
		Healthy	Faulty
Actual	Normal	0.84	0.16
	Unknown	0.28	0.72
	Faulty	0.55	0.45

Table 4 Confusion matrix with features set selected using SVD (shown in Figure 8).

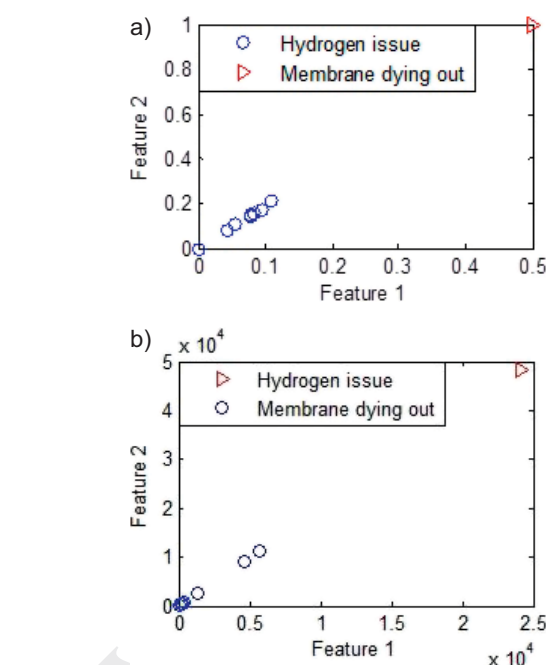|			Predicted		
			Healthy	Faulty	
				Membrane drying out	Hydrogen issue
Actual	Normal		1	0	0
	Unknown		0.80	0	0.20
	Faulty	Membrane drying out	0	1	0
		Hydrogen issue	0.20	0	0.80

Fig. 10 a) Classification results at 1st principal direction; b) Classification results at 2nd principal direction; c) Classification results at 3rd principal direction; d) Classification results at 4th principal direction.

It should be mentioned that some other classification techniques, like KNN, are also used in the analysis to distinguish various fuel cell states using the features selected from SVD and randomly selected features, and similar results can be obtained, i.e. better diagnostic performance can be provided with features selected using SVD. Since no diagnostic performance is not improved using other classification techniques, they will not be described herein.

Moreover, the capability of discriminating different fuel cell faults is investigated with dataset reduced using KFDA and KPCA, and results are depicted in Figure 10(a) and Figure 10(b). It should be mentioned that features used in Figure 10(a) and Figure 10(b) are selected with SVD. From the results it can be seen that with both supervised and unsupervised data dimension techniques, the features selected from SVD can discriminate various fuel cell faults, i.e. hydrogen issue and membrane drying out in this study, indicating the fuel cell faults can be detected and isolated with good quality using the diagnostic framework depicted in Figure 2.

With the described data-driven approaches, the features from the original measurements can be extracted and selected, by classifying these features, the corresponding fuel cell state can be determined. In the practical fuel cell system, by repeating the diagnostic analysis, features representing various fuel cell states can be accumulated, which can distinguish various fuel cell states more clearly. With input of new features, its state can be easily determined by calculating its minimum distance to different fuel cell states.

It should be noted that without prior knowledge about fuel cell faults, the diagnostic procedure will take about 25 min to determine the fuel cell states, based on the previous studies [33, 34], fuel cell flooding can cause the most rapid performance degradation, which is about 0.39 V/h, i.e. the fuel cell voltage will drop 0.15 V at 25 min. With the critical failure threshold of 20% power drop, the fuel cell faults can be identified before reaching the critical level using the diagnostic procedure.

5 Conclusion

In the paper, a fault diagnostic framework based on data-driven approaches is applied in practical fuel cell systems. In this framework, several commonly used data-based diagnostic techniques, including KPCA/KFDA and WP transform, are applied to reduce the dimension of original dataset and extract the features. Moreover, SVD technique is employed in the study to select features with better prognostic results. Test data from a practical EC fuel cell system is used in the analysis.

Results demonstrate that diagnostic performance using different features extracted from WP transform shows clearly variation. Compared to the randomly selected features, features selected using SVD can provide more effective classification results, most faulty cases can be separated from the normal and unknown system states, which indicates that the system faulty state detected with good quality using features selected from SVD technique. Moreover, with labelled datasets (using KFDA), the selected features from SVD can clearly indicate the states of different datasets, while without prior knowledge about dataset state (using KPCA), fuel cell faults can still be discriminated with good quality using the diagnostic pro-

ORIGINAL RESEARCH PAPER

cedure, this further confirm the robustness of the diagnostic procedure in practical fuel cell systems. Furthermore, with features selected from SVD using both KPCA and KFDA, different fuel cell faults can also be determined with good quality.

In the future work, the performance of this diagnostic framework will be studied in real fuel cell systems, and fewer sensors will be used for fuel cell fault diagnosis using the sensor selection algorithm, which will consider both the diagnostic performance and the cost of sensors, so that the fuel cell faults can be identified with less computational time to meet the requirement of on-line health monitoring tasks, and selected sensors can be used to reduce the cost of the monitoring system..

Acknowledgement

References

[1] R. Petrone, Z. Zheng, D. Hissel, M. C. Pera, C. Pianese, M. Sorrentino, M. Becherif, N. Yousfi-Steiner, *International Journal of Hydrogen Energy* **2013**, *38*, 7077.

[2] A. Forrai, H. Funato, Y. Yanagita, Y. Kato, *IEEE Transactions on Energy Conversion* **2005**, *20*, 668.

[3] N. Fouquet, C. Doulet, C. Nouillant, G. D. Tanguy, B. O. Bouamama, *Journal of Power Sources* **2006**, *159*, 905.

[4] M. Ordonez, M. O. Sonnaillon, J. E. Quaicoe, M. T. Iqbal, *IEEE Transactions on Industrial Electronics* **2010**, *57*, 1925.

[5] M. A. Rubio, A. Urquia, S. Dormido, *International Journal of Hydrogen Energy* **2010**, *35*, 2586.

[6] T. W. Lee, A. A. Tseng, K. S. Bae, Y. H. Do, *Energy Fuels* **2010**, *24*, 1882.

[7] A. Zeller, O. Rallieres, J. Regnier, C. Turpin, *Vehicle Power and Propulsion Conference (VPPC)*, Lille, France, **2010**.

[8] L. A. M. Riascos, M. G. Simoes, P. E. Miyagi, *Journal of Power Sources* **2007**, *165*, 267.

[9] L. A. M. Riascos, M. G. Simoes, P. E. Miyagi, *Journal of Power Sources* **2008**, *175*, 419.

[10] N. Y. Steiner, D. Hissel, P. Mocoteguy, D. Candusso, *International Journal of Hydrogen Energy* **2011**, *36*, 3067.

[11] M. M. Kamal, D. Yu, *International Journal of Engineering Science and Technology* **2011**, *3*, 1.

[12] A. Mohammadi, A. Djerdir, D. Bouquain, B. Bouriot, D. Khaburi, *Transportation Electrification Conference and Expo (ITEC)*, Detroit, **2013**.

[13] Z. Zheng, R. Petrone, M. C. Pera, D. Hissel, M. Becherif, C. Pianese, N. Y. Steiner, M. Sorrentino, *International Journal of Hydrogen Energy* **2013**, *38*, 8914.

[14] A. Narjiss, D. Depernet, D. Candusso, F. Gustin, D. Hissel, *13th International Power Electronics and Motion Control Conference*, Poznan, Poland, **2008**.

[15] R. Onanena, L. Oukhellou, E. E. Come, S. Jemei, D. Candusso, D. Hissel, P. Aknin, *Chemical Engineering Transactions* **2013**, *33*, 1021.

[16] Z. Zheng, R. Petrone, M. C. Pera, D. Hissel, M. Becherif, C. Pianese, *39th Annual Conference of the IEEE industrial Electronics Society (IECON)*, Vienna, Austria, **2013**.

[17] E. Frappe, A. D. Bernardinis, O. Bethoux, D. Candusso, F. Harel, C. Marchand, G. Coquery, *The European Physical Journal Applied Physics* **2011**, *54*, 1.

[18] N. Y. Steiner, D. Hissel, P. Mocoteguy, D. Candusso, *International Journal of Hydrogen Energy* **2011**, *36*, 740.

[19] Z. Li, R. Outbib, D. Hissel, S. Giurgea, *European Control Conference*, Zurich, Switzerland, **2013**.

[20] Z. Li, R. Outbib, D. Hissel, S. Giurgea, *Control Engineering Practice* **2014**, *28*, 1.

[21] L. Placca, R. Kouta, D. Candusso, J. F. Blachot, W. Charon, *International Journal of Hydrogen Energy* **2010**, *35*, 4582.

[22] J. Hua, J. Li, M. Ouyang, L. Lu, L. Xu, *International Journal of Hydrogen Energy* **2011**, *36*, 9896.

[23] J. Li, P. Cui, *Expert Systems with Applications* **2009**, *36*, 1423.

[24] R. E. Learned, S. Willsky, *Applied and Computational Harmonic Analysis* **1995**, *2*, 265.

[25] C. Torrence, G. P. Compo, *Bulletin of the American Meteorological Society* **1998**, *79*, 61.

[26] W. X. Yang, P. W. Tse, *DT & E International* **2003**, *36*, 419.

[27] J. Walton, N. Fairley, *Journal of Electron Spectroscopy and Related Phenomena* **2005**, *148*, 29.

[28] M. Bydder, J. Du, *Magnetic Resonance Imaging* **2006**, *24*, 849.

[29] P. Fenglai, W. Nan, Z. Su, *Biotechnology: An Indian Journal* **2014**, *10*, 6236.

[30] J. Kim, Y. Tak, *International Journal of Hydrogen Energy* **2014**, *39*, 10664.

[31] M. Gašperin, P. Boškoski, A. Debenjak, J. Petrovčič, *Fuel Cells* **2014**, *14*, 457.

[32] F. Pei, Z. Li, S. Zhou, *Applied Mechanics and Materials* **2012**, *217–219*, 770.

[33] M. G. Santarelli, M. F. Torchio, P. Cochis, *Journal of Power Sources* **2006**, *159*, 824.

[34] L. Lu, X. Han, J. Li, J. Hua, M. Ouyang, Y. Zheng, *Journal of Power Sources* **2013**, *42*, 272.

[35] G. Mousa, J. DeVaal, F. Golnaraghi, *International Journal of Hydrogen Energy* **2014**, *39*, 20116.

[36] J. M. L. Canut, R. M. Abouatallah, D. A. Harrington, *Journal of the Electrochemical Society* **2006**, *153*, 5.