Diagnostic of fuel cell air supply subsystems based on pressure signal records and statistical pattern recognition approach

Djedjiga Benouioua, Fabien Harel, Denis Candusso

Diagnostic of fuel cell air supply subsystems based on pressure signal records and statistical pattern recognition approach

D. Benouioua[1,2,3,4,5], F. Harel[1,3], D. Candusso[1,4,*]

[1] FCLAB, Rue Ernest Thierry Mieg, F-90010 Belfort Cedex, France.

[2] FEMTO-ST, Université de Bourgogne Franche-Comté, Rue Ernest Thierry Mieg, F-90010 Belfort Cedex, France.

[3] AME-Eco7, Université Gustave Eiffel, Université de Lyon, 25 Avenue François Mitterrand, Case24, F-69675 Bron Cedex, France.

[4] SATIE, COSYS, Université Gustave Eiffel, 25 Allée des marronniers, F-78000 Versailles Satory, France.

[5] CSEM SA, Jaquet-Droz 1, 2002 Neuchâtel, Switzerland.

[*] Corresponding author: denis.candusso@univ-eiffel.fr

Abstract:

A data-driven and application-oriented diagnosis tool is developed for Fuel Cell (FC) air supply subsystems. A bench emulating a FC air line is built to study normal and abnormal operations (clogged inlet, air leakage, error in compressor speed control) and data are collected using the air pressure transducer, which is usually implemented in FC generators. A pattern recognition approach is then applied to statistical features extracted from the pressure signal. The performance of the diagnosis strategy is evaluated from confusion matrices, associated to graphs and performance indicators. Two examples of compressors, air subsystem managements, and data records are considered to examine the method portability. Best classification rates (> 95%) are obtained on test profiles, when the pressure regulation is disabled; fault stamps can thus be found in the pressure signal morphology. Regarding the frequency of data logging, both 1 kHz and 100 Hz values are found effective for fault isolations.

Keywords: Fuel cell; Air supply subsystem; Compressor; Diagnosis; Supervised machine-learning.

Highlights:

- A data-driven diagnostic tool is developed for the air supply subsystem of fuel cells.
- The input to the algorithm is the signal from the pressure sensor (at 1000 or 100 Hz).
- Fault scenarios are: air leak, compressor speed control fault, clogged air inlet.
- The portability of the method is shown with 2 compressors and 2 air control modes.
- Highest classification rate (97%) is reached with the pressure regulation disabled.

1. Introduction

Fuel Cell (FC) generator components and subsystems are inevitably subject to slow or abrupt degradations during their life cycle. FCs face mechanical, chemical, and thermal factors contributing to ageing and / or early failure [1, 2]. Among the balance-of-plant components and sections used in PEMFC (Proton Exchange Membrane Fuel Cell) generators, the air supply and conditioning subsystem plays a key-role for the whole FC generator. Its major tasks can be summarized as follows [3-5]:

- Air supply: the air supply system must provide sufficient reactant flows (i.e. to keep the desired oxygen excess ratios) over the full FC power range. A compressor machine is usually used to this end (blowers are adapted for low pressure FCs) and various technologies of compressors are available as centrifugal, screw, claw, lobe, scroll, rotary vane… By the way, it is important to highlight that the air compressor is identified as the most energy-consuming and most expensive ancillary in a FC system [5]. Pressure and flow rate have to be controlled properly in the cathode section. The control of the air supply system requires therefore well-suited sensors. Firstly, pressure and temperature transducers, mass air flow sensors, and a shaft speed sensor shall be used to build compressor maps on dedicated testbenches [6-8]. Then, the air mass flow sensor is usually removed from commercialized FC generators since the air mass flow rate can be deduced from pressure, rotational speed of electric motor, and compressor maps.
- Air cleaning: the air at the inlet of the FC air section must be filtered. Particles, dust, oil content, or chemical substance can be harmful for the FC stack and they need to be removed from the air feeding the generator.
- Cathode pressurization: in any case, the air is supplied to the stack under pressure, ranging usually from low levels (i.e. values slightly above atmospheric pressure) to 1.5 - 2 bar rel. depending on the stack design. The pressure inherently depends on the air flow rate but it can also be adjusted using a back-pressure valve / throttle valve (and possibly with an expander / turbine), downstream of the FC stack.
- Air humidification: the Nafion-based polymer membrane of the Low Temperature - PEMFC has to be maintained in a fully hydrated state. Therefore, the air supply system has to balance the water needed by the Membrane Electrode Assemblies (MEAs), the water produced by the chemical reaction in the cell, and possibly the amount of water injected upstream of the stack. Water can possibly be recovered from the cathode outlet by a drain or a condenser, and a

humidifier can be used between the compressor outlet and the stack air inlet. The humidification depends strongly on the FC operating pressure level, on the air mass flow rate but also on the temperatures of FC air inlet and stack.

As we can see, complex dependencies between these tasks have to be taken into account to design and control the air supply system properly. Control plays therefore a key-role in the air supply management of FC generators [9-13]. Any pressure ripples or abrupt variations of the air flow may induce some loss of performances in FC stacks and generators, and possible mechanical degradations in PEMFC stacks (resulting in early failures due to cracks, pinholes, perforations, and tears in the membranes of the electrochemical assembly). Unequal pressures on both sides of the MEA would also damage the stack. Common faults occurring in the air line sections of FC systems can be related with any dysfunctions in the controls of the compressor machines (inducing wrong rotation speeds and thus air flow values that would be non-adapted to the FC operation), any possible air leakages in the air ducts, or any problems at the inlet of the air section (e.g. clogged filter at air line inlet). The causes of these faults have different possible origins; they may be related with ageing factors or with non-ordinary FC operating conditions. Overall, the effects of these faults on the FC stack operation are mainly related with oxygen starvation or air oversupply phenomena.

The rapid effect of air starvation and air oversupply on the FC behavior can be detected on the evolution of the electrical stack performances [14]. To ensure steady conditions at the air inlet of a FC, it is therefore critical to monitor the State-of-Health (SoH) of the FC air supply and conditioning subsystem. There is currently a strong need for monitoring tools and solutions to keep the air line components healthy, to improve the stack durability and generator efficiency with optimized FC system operations. The SoH of the FC generator air section can be deduced from the monitoring of the FC voltage response. In this case, the FC itself acts as a sensor that can be used for diagnostic purposes. Air starvation or air over-supply conditions can be detected in such a way.

PEMFC fault diagnosis has received much more attention over the past decade [15]. Due to the difficulty of developing accurate, knowledge-based PEMFC models incorporating various failure mode effects, data-driven approaches are widely used for diagnosis purposes. In such methods, features are extracted from the PEMFC measurements using signal processing techniques, and the PEMFC state can be determined by implementing pattern recognition methods to these features. Various pattern recognition algorithms have been applied to the extracted features, such as Gaussian mixture model, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), self-organizing maps, etc. Numerous articles deal with such data-

driven FC diagnostic tools based upon stack voltage measurements [16-19]. Signal-based fault detections can be done for instance from EIS records [20] or by considering the natural oscillations of the FC voltage [21].

Many FC diagnostic tool and diagnosis strategies are based on voltage measurements at stack or cell terminals. But, as an air pressure transmitter is usually implemented in the cathode section of FC generators, it is also possible to use this sensor and its signal in order to assess the SoH of the FC air supply and conditioning subsystem. Moreover, the data provided by the pressure transmitter can possibly be used in redundant mode with the FC voltage data and/or as an additional input to propose a full diagnosis strategy. One may foresee that the pressure data can be used for the detection of specific air line faults and in order to decorrelate such faults from FC failures that may be more or less independent from the air line.

Numerous recent references dealing with FC diagnosis based on pressure drop measurement can also be found in the literature. Indeed, an increase in the pressure drop, particularly at the cathode side of a PEMFC, is a reliable indicator of cell flooding. This parameter can be used to make decisions and propose corrective actions [22]. As an example, in order to propose a non-intrusive flooding diagnosis procedure, Steiner et al. [23] used the comparison between the information given by a differential pressure sensor (measuring the inlet / outlet cathode pressure drop) and the output of a black-box model (estimating the pressure drop using an artificial Neural Network trained with flooding-free data). Of course, various works recommend FC diagnostic tools based on the measurement of the gas pressure and especially on the pressure-induced voltage oscillations [24, 25]. Different FC diagnosis strategies are also based on the measurements of various physical operating parameters, including the air pressure at cathode as input (model-based as in [26] and non-model-based strategies as well).

Although many studies have been conducted on stack failures and performance degradation, relatively few studies have examined Balance-of-Plant components and subsystems [19, 27]. In a recent article, Jinyeon Won et al. propose a hybrid diagnostic method to diagnose faults in the air supply section of a PEMFC system [28]. In this work, an artificial neural network classifier, used as a data-driven diagnostic tool, is combined with a model-based diagnosis method. The goal is to provide faster and more accurate diagnostics compared to the previously developed and solely model-based tool. The results obtained should help to develop an effective FC SoH management system.

In the academic literature dealing with FCs, only few information is available on diagnostic tools, diagnosis strategies focusing on the FC air supply subsystem (i.e. considering its own fault modes), and based on a coupling between cathode pressure measurements and pattern-

recognition algorithms. So, the monitoring tool that we propose for the FC air supply and conditioning subsystem relies on the measurement and record of the air pressure signal at the FC cathode side. It uses signal processing techniques to generate relevant statistical features related with the various explored operating conditions. It relies on a data-driven, pattern-recognition based approach implementing the KNN algorithm (selected as an example of classifier).

In order to develop and assess such a diagnosis method, an experimental testbench emulating a FC air supply and conditioning subsystem was designed and some tests were performed in different operating conditions to build an experimental dataset. More details on this step are given in Section 2. The diagnosis method that we propose to identify different operating modes and faults of the FC air subsystem is described in Section 3. Two examples of possible application are presented in Sections 4 and 5, related with two different technologies of compressor, two air line control modes, and two frequencies of data pressure records. The objective is to examine and discuss the portability of the method. Finally, major conclusions and perspectives are reported in Section 5.

2. Experimental

2.1 Testbench

In order to propose a diagnostic tool based on the measurement of the air pressure signal, a testbench has been developed to emulate a FC air supply and conditioning subsystem. With this experimental mean, different technologies and designs of components to be used in FC generator cathode section (e.g. compressor, back-pressure valve, flow and pressure sensors, pipes) can be implemented and evaluated according to specific FC systems requirements, expressed primarily in terms of air flow and pressure at stack inlet. Moreover, various air flow and pressure control modes can be investigated. Further components can possibly be included into the air line, as external humidifiers if a self-humidifying FC stack is not considered in the design of the studied FC generator. Also, the testbench allows studying any possible fault scenarios related with FC air supply subsystems. A picture and a scheme of the developed testbench are shown in Fig. 1. he testbench includes mainly (in air flow direction):

- An air filter at inlet.
- A compressor (claw compressor or regenerative blower in this study).

- A pressure sensor (piezoresistive pressure transmitter Keller PR 23S). The pressure range of the sensor is 0 - 3 bar abs (with 0 - 10 V full scale measurement). Its output resolution (including analog / digital converter) is about 0.2 mV (16 bits), which leads to a pressure measurement resolution of 0.06 mbar. The limiting frequency is 1 kHz. Such a pressure measurement is typical of FC system applications.
- A temperature sensor (thermocouple type K), which will not be used in this study.
- A pneumatic back-pressure valve (Kämmer series 030).
- An air mass flowmeter (Brooks 5863S).

The testbench is dedicated to investigate compressors with air pressure levels up to 3 bar abs and air flows up to 300 Nl/min. This last value can be related with a 6 kW class PEMFC stack (e.g. 42 cell stack having individual cells of 220 cm² electrochemical surface area, showing voltages of 0.7 V @ 1 $A.cm^{-2}$ current density, and fed with an air stoichiometry rate close to two).

The monitoring and the control of the testbench parameters are done through a National Instruments embedded controller with analog or numeric exchanges and a dedicated control software developed with Labview. The different data measured by the sensors are acquired at a 1 kHz sampling rate. The control-command has an action on the speed of the compressor and on the position of the back-pressure valve. The action on the compressor speed is linked to a control loop of the air flow. The action on the valve position is related to a regulation loop of the pressure in the air line. A Proportional-Integral (PI) regulation process is used in both cases. The power demand from the compressor is carried out from current and voltage measurements on the electrical section of the ancillary. Generally, the knowledge of the compressor power magnitude in its various operating cases (flow / pressure) helps greatly to compare different compressor technologies and to find the best suited ones for the applications to be developed. The test bench functionalities can contribute to this objective. Moreover, even if this is not the topic of this study, electrical measurements on the motor compressor can also be used to identify dysfunctions on the air line.

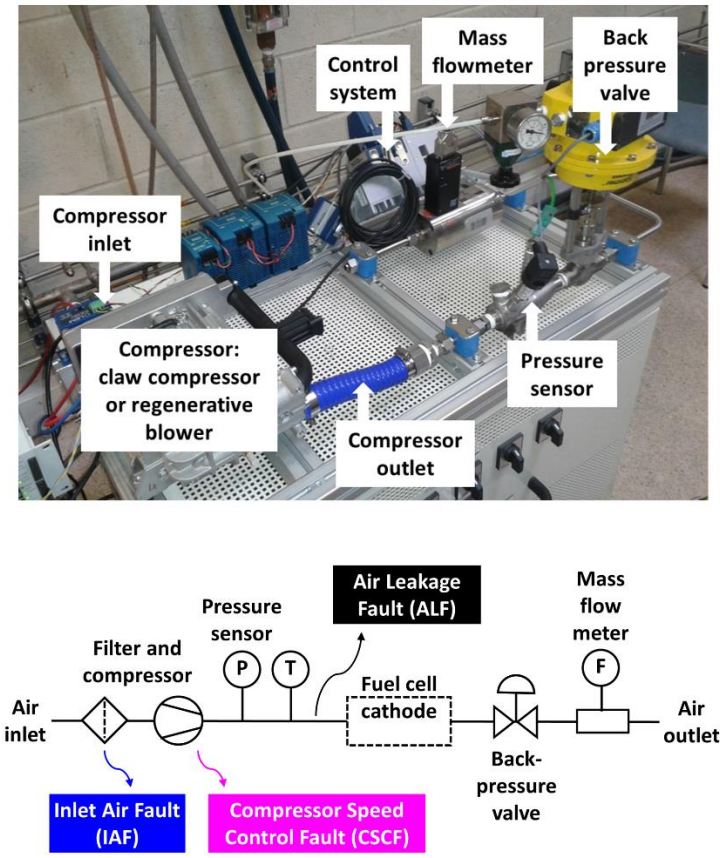

Fig. 1. Picture and scheme of the testbench dedicated to the study of a FC air supply and conditioning subsystem, including possible fault scenarios.

2.2. Operating modes and faults

In this study, four operating scenarios of air supply subsystem are investigated using the testbench shown in Fig. 1. The introduced fault scenarios are related with common faults occurring in the air line section of a FC system. These malfunctions have been identified during various tests and projects that have been carried out on the FC platform of Belfort in recent years; FC generator manufacturers also face these problems. Their occurrence and relative severity have been reported in the literature [28].

In the case of the Normal Operation (NO) scenario, the air line ancillaries operate without any failure (no fault is introduced). In the case of the three other operating scenarios, faults are introduced by intention (Fig. 1.b)). They are related to the following dysfunctions of the FC air supply and conditioning subsystem:

- Compressor Speed Control Fault (CSCF): a drop in the set point of the compressor speed is voluntary caused.

- Air Leakage Fault (ALF): an air leakage is deliberately caused by opening partially a dedicated valve in the air duct.
- Inlet Air Fault (IAF): in this test, the air inlet of the compressor is partially clogged.

The causes of the studied faults may come from different sources and the effect on the FC behavior is mainly related with oxygen starvation. The causes and effects of the studied faults are listed in Table 1.

Table 1. The causes and effects of the studied faults.

Operating faults	Causes	Effects on PEMFC
Compressor Speed Control Fault (CSCF)	Mechanical damage in the air compressor. Control system dysfunction.	Oxygen starvation. Pressure variations at cathode and pressure differences between anode and cathode. Mechanical degradation of the MEA. Air temperature effect on the FC humidity.
Air Leakage Fault (ALF)	Degradation of a sealing gasket in the air line. Leaks in the compressor due to sealing defect in the compression head. Mechanical fatigue of the compression head.	Oxygen starvation. Pressure differences. Air temperature effect on the FC humidity.
Inlet Air Fault (IAF)	Clogged air filter at the compressor inlet.	Oxygen starvation. Pressure differences. Air temperature effect on the FC humidity.

In addition, the Compressor Speed Control Fault and Inlet Air Fault operating scenarios may cause any oscillations due to the rotor imbalance, and resonances in worst-case. This will cause increased wear of bearings and fatigue in structural components of the compressor as well as higher energy consumption. The pressure differences due to the three studied operating scenarios can accelerate the stress of FC system components and therefore induce ageing phenomena. The humidifier section of the FC generator can also be impacted: pressure

differences will put stress on its membranes. Over time, this could cause internal leakage as well as a decline of performance.
As it will be presented in Section 2.3., each operating mode will be associated to a class label for the diagnosis task.

2.3 Experimental database constitution

The database used in this work is built from experimentations conducted with two different air compressor technologies, namely a claw compressor and a regenerative blower. As already mentioned, the air flow and pressure are either regulated (closed loops) or not controlled (open loops). Besides, the pressure can be controlled while the air flow varies in open loop. The data monitored and used for the diagnosis are collected using a pressure transducer, the kind of sensor which is usually implemented in real FC generators (unlike mass flow controllers for example as they are dedicated to in-lab FC testbenches). The acquisition of the air pressure signal is done at two frequencies: 1 kHz and 100 Hz. This frequency range is selected as a trade-off between several constraints. On the one hand, the frequency level must be high enough to be able to capitalize on the possible fault stamps present in the morphology of the pressure transmitter signals. The compressor velocity, the pressure level, any parameters linked with its mechanical design (e.g. number of vanes) impact the frequential signature of the pressure signal (fundamental and harmonics). In this regard, preliminary experiments have shown that a hundred Hertz can be considered as a minimal frequency value for the expected air flows and pressures. On the other hand, a sampling frequency of 1 kHz (and even 100 Hz) can be considered as quite high for FC generators dedicated to on-board applications. Indeed, such FC systems may be equipped with less performing sensors (usually for economic reasons) and high frequency records also mean high computational effort required by the diagnosis algorithm, which may hinder its on-board applications as well.
Two examples of data collected on the FC air line testbench are shown in Fig. 2 and 3. Both figures are related with air flow and pressure data recorded under normal and abnormal operating conditions intentionally introduced during the experimentations. However, Fig. 2 and 3 correspond to distinct types of data. The signals of Fig. 2 come from measurements recorded at 1 kHz, with a claw compressor operated in air pressure and flow open loops whereas Fig. 3 is based on a regenerative blower technology with regulated pressure and flow, and a recording frequency of 100 Hz. In both cases, pressure and air flow reference values are applied, respectively to 200 mbar rel. and 150 Nl/min. In Fig. 2, the faults lead to obvious deviations

from the pressure and flow references while in Fig. 3, the same faults do not lead to strong shifts from the setpoints. Due to the pressure and flow controls, the pressure and the flow signals are much steadier in Fig. 3. (except for the CSCF fault which dynamical effects cannot be mitigated by the control loops; the disturbances remain visible on the signals).

Besides, in Fig. 2 and 3, the time and pressure intervals linked with the different operating conditions are identically reported in colors on the pressure signal plots. These intervals will be used to provide some initial / preliminary labels to the data that will be used in our diagnosis tool, based on the supervised machine learning approach described in the next section. According to the experimenter knowledge, these data labels cannot be associated to every point with the same accuracy. Some uncertainty on the labeling remains during the transient's phases between two different operating conditions.

A last remark on Fig. 2 and 3: the difference between the total durations of the two experiments is due to the fact that, for the two record frequencies considered, we aimed at having approximately the same sizes of data for the diagnosis procedure.

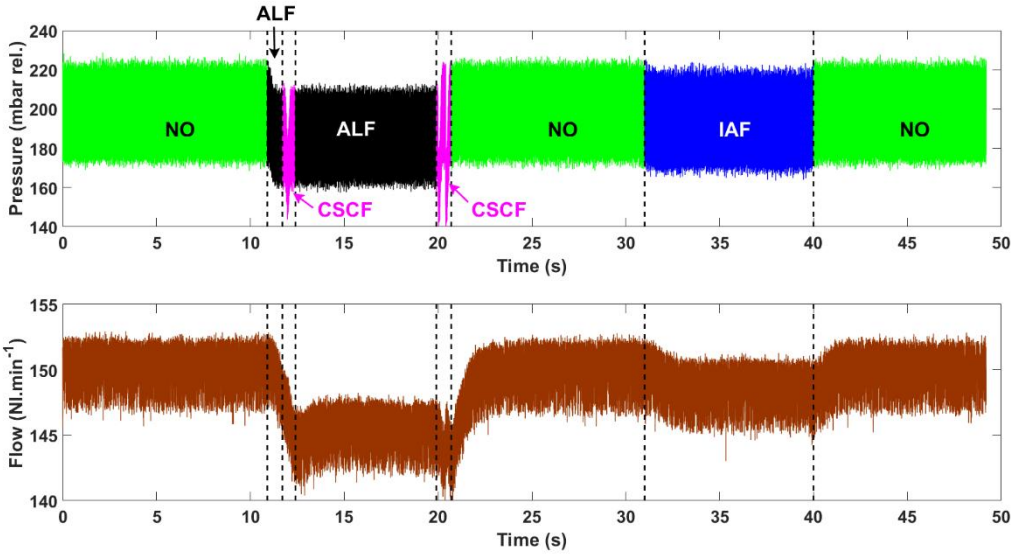

Fig. 2. Example of air pressure and flow signals recorded at 1 kHz with a claw compressor in the air line, operated in open loops on pressure and flow, under normal and abnormal operating conditions. The top figure shows the intervals related with the four operating conditions and the initial / preliminary labels.

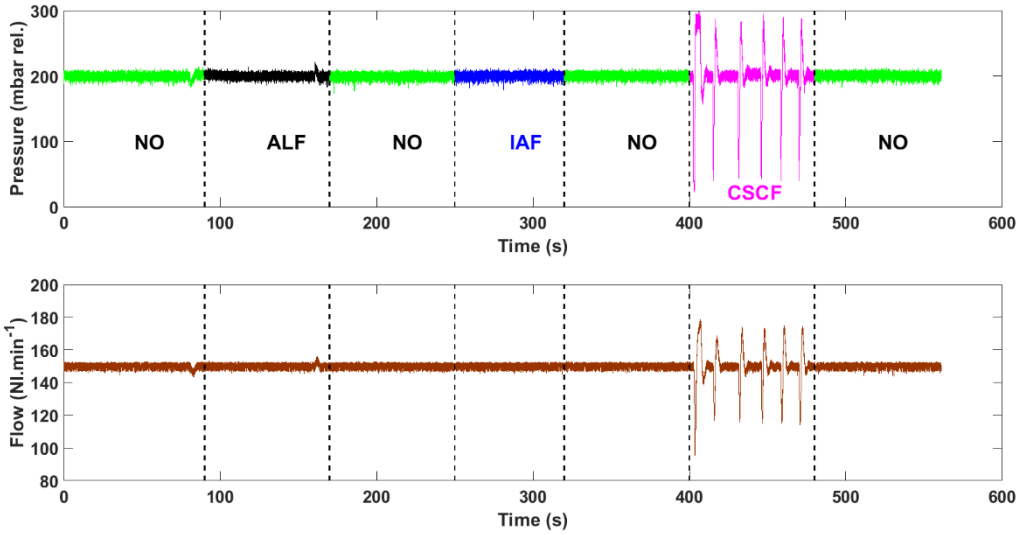

Fig. 3. Second example of air pressure and flow signals. These records are done at 100 Hz and obtained with a regenerative blower in the air line, operated in closed control loops on pressure and flow. The intervals and the initial / preliminary labels linked with the four operating conditions are reported on the top figure.

3. Diagnosis method

3.1 General principle

The diagnosis method that we propose to identify different operating modes and faults in a FC air supply and conditioning subsystem corresponds to a data-driven, pattern-recognition based approach [29]. It includes machine learning [30], instance-based learning algorithms that, instead of performing explicit generalization, compares new problem instances with instances seen in training, which have been stored in memory. In general, the first step of such an approach consists in extracting some features from available measurements done on the investigated system to build a training set of data (in our study: the air pressure signal). At this stage, a feature selection step may sometimes be useful to remove non-relevant or redundant feature(s), in order to increase the quality of the results. In a second step, a learning / training process is applied. Different features extracted from the training data set are associated to a given number of classes (four in our case: NO, ALF, IAF, CSCF classes). A testing step is used to verify that the classification algorithm leads to good results (i.e. high classification rates). Possibly, these tests can be done with features extracted from available measurements that were

not already used in the training phase. In the last phase of the pattern recognition approach, the algorithm is applied to new measurements data and their related features, which are then classified. The algorithm indicates if the new measurements (here, the new pressure values) can be associated to one of the (four) operating conditions.

Our diagnosis method combines a statistical analysis of the gas pressure signal measured in the air line with a learning and classification algorithm. In our study, the K-Nearest Neighbors (KNN) rule is adopted for the learning and classification task. KNN is one of the topmost instance-based (or memory-based machine) learning algorithms. It is a conceptually simple yet very powerful algorithm. KNN is highly accurate and simple to use. It is easy to interpret, understand, and implement. KNN is a non-parametric method used for regression and classification as well, unlike some other supervised learning algorithms [31]. KNN does not make any assumptions about the data, meaning that it can be used for a wide variety of problems and applications such as finance, healthcare, image and video recognition. For those reasons, KNN is one of the most popular machine learning algorithms. Main KNN cons can be presented as follows [32]: since KNN stores most or all of the data, a lot of memory might be required, which makes KNN computationally expensive. Large datasets can also induce long computation durations for the predictions.

3.2 KNN basics

In KNN classification, the output is a class membership. The KNN classifier is a distance-based supervised classifier which only considers the neighborhood around the data point (in the feature space) that we want to classify. In KNN, K is the number of nearest neighbors and the core deciding factor in the classification process. K is a positive integer, typically small. The best choice of K depends upon the data and a good K can be selected by various heuristic techniques. Generally, larger values of K reduce the effect of the noise on the classification but make boundaries between classes less distinct [33].

K=1 is the simplest case. Let suppose that P1 is the point, for which a label needs to predict. First, the closest point to P1 has to be found, with regard to a selected distance measure (e.g. Euclidean distance, Hamming distance, Manhattan distance …), and then the label of the nearest point is assigned to P1. In other words, P1 is simply assigned to the class of its single nearest neighbor.

When K>1, the K closest points to P1 have to be found. Then, the point is classified by a majority vote of its K neighbors. Each object votes for its class and the class with the most votes is selected as the prediction (i.e. the label of the class with the most neighbors is associated to P1). Otherwise stated, P1 is assigned to the class most common among its K nearest neighbors. To find closest similar points, the distance between points is determined using a chosen metric.

A confusion matrix, or error matrix, is a common tool applied in machine learning to gauge the quality of a classifier. With its associated statistics, this matrix can be used to validate the accuracy of the KNN classification. It is represented under the form of a table where each row represents the instances in a predicted class while each column represents the instances in an actual (or true) class (or vice versa) [34]. True and predicted values are then cross–tabulated in the table. The matrix name stems from the fact that the layout allows to observe if the system is confusing two classes (i.e. commonly mislabeling one as another). All correct predictions are located in the diagonal of the confusion matrix. So, it is easy to visually inspect the table to identify the prediction errors as they will be represented by values outside the diagonal.

The simplest confusion matrix is a type 2×2 matrix for a binary classification task (Table 2). In this case, there are two decision classes, Positives and Negatives, with actual (i.e. true) values or classes (P and N) and predicted values ($\widehat{P}$ and $\widehat{N}$). The entries in the table are called True Positives (TP, i.e. correct positive predictions), False Positives (FP, i.e. instances that are actually negative but tagged as positive by the classifier. Said otherwise: irrelevant items that are incorrectly identified as relevant), True Negatives (TN, i.e. genuinely false instances. In other words: irrelevant items that are correctly identified as irrelevant), and False Negatives (FN, i.e. instances tagged as negative by the classifier, but genuinely positive. Said differently: relevant items that are incorrectly identified as irrelevant) [35, 36].

The reason that the confusion matrix is particularly useful is that, unlike other types of classification metrics such as simple accuracy, the confusion matrix generates a more complete picture of how a classifier performs. As a matter of fact, many statistics or metrics are associated with a confusion matrix, such as [37]:

- Accuracy: it is the proportion of correct classifications ((TP+TN) / total population). This metric is not sufficient to express the real performance of a classifier, in particular when the numbers of observations in different classes vary greatly.
- Sensitivity or recall: it is the number of genuinely positive instances divided by the number of false-negative instances and total positive instances (TP / (FN+TP)). In other words, the

sensitivity is representative of the proportion of true positive instances that a machine learning model has classified.

- Specificity: it quantifies the true negative rate or the number of instances that the model defined as negative and that were truly negative. This is calculated by taking the number of instances classified as negative and dividing it by the number of false-positive instances combined with the true negative examples (TN / (FP+TN)).

- Precision or Positive Predictive Value (PPV): it is concerned with how many of the instances that the model labelled positive were truly positive. In order to calculate this metrics, the number of true positive instances are divided by the number of false-positive instances plus true positives ((TP) / (FP+TP)). To make the distinction between sensitivity and precision clearer, precision aims to figure out the percentage of all instances labelled positive that were truly positive, while sensitivity tracks the percent of all true positive examples that the classifier could recognize.

- Negative Predictive Value (NPV): it is defined as the proportion of predicted negatives which are real negatives (TN / (FN+TN)).

Note that various other metrics can also be used (e.g. False Negative Rate, False Positive Rate, False Omission Rate, prevalence, F1 score, etc) to gauge the quality of a classifier.

Table 2. Form of the simplest confusion matrix for a binary classification task, with examples of metrics.

		Actual / true classes		Examples of metrics
	Total population = TP+FN+FP+TN	Positives (P) = TP + FN	Negatives (N) = FP + TN	
Predicted classes	Positives ($\hat{P}$) = TP+FP	True Positives (TP)	False Positives (FP)	Precision or PPV = TP / (TP+FP)
	Negatives ($\hat{N}$) = FN+TN	False Negatives (FN)	True Negatives (TN)	NPV = TN / (FN+TN)
Examples of metrics		Sensitivity = TP / (TP+FN)	Specificity = TN / (FP+TN)	Accuracy = (TP+TN) / Total population

An example of confusion matrix with 4 classes is given in Eq. 1; such a matrix will be used in our study. Regardless of the size of the confusion matrix, the method for interpreting the table is the same. Some programming codes are designed for the formation of a confusion matrix

with two or more classes and for the calculating of the associated performance indicators. One example of such a code developed in the Matlab™ environment is proposed by Abbas Manthiri S [38].

$$Confusion\ matrix\ for\ 4\ classes = \begin{array}{c|cccc} Predicted\ class & & Actual\ class & & \\ & Class1 & Class2 & Class3 & Class4 \\ Class1 & \boldsymbol{M_{11}} & \boldsymbol{M_{12}} & \boldsymbol{M_{13}} & \boldsymbol{M_{14}} \\ Class2 & \boldsymbol{M_{21}} & \boldsymbol{M_{22}} & \boldsymbol{M_{23}} & \boldsymbol{M_{24}} \\ Class3 & \boldsymbol{M_{31}} & \boldsymbol{M_{32}} & \boldsymbol{M_{33}} & \boldsymbol{M_{34}} \\ Class4 & \boldsymbol{M_{41}} & \boldsymbol{M_{42}} & \boldsymbol{M_{43}} & \boldsymbol{M_{44}} \end{array} \quad (1)$$

Therefore, to summarize, KNN has the following basic steps: distance computing, closest neighbors search, vote for labels. The quality of the KNN classifier can be assessed thanks to a confusion matrix, which allows displaying and comparing the actual values with the predicted ones, possibly by the computing of suitable metrics.

3.3 Details on the steps of the diagnosis method

Our diagnosis method involves a series of different steps, which are synthetized by the block diagram of Fig. 4 and are described below. Two application examples of diagnosis will be presented in Sections 4 and 5.

The initial step deals with experimentation and data collection. A compressor technology needs to be selected, as well as a control mode for the air line (open loop(s) / closed loop(s)) for the air pressure and flow, and a frequency for the data record (two values are investigated in our study: 1 kHz and 100 Hz). These selections being made, the air pressure signal is acquired in various operating conditions (NO, ALF, CSCF, IAF) and the labels are affected to the corresponding data.

The feature extraction step is the process of defining a set of features, which will most efficiently or meaningfully represent the information that is important for the analysis and classification. In our case, the features are statistical values: the pressure signal is analyzed and described by first order statistics or moments computed on a moving scanning window, which has to be defined with a proper size (the effect of the scanning window size on the classification rate will be discussed in Section 4.2). In our study, 6 statistical values are computed for each point of the pressure signal on a scanning window defined in the Matlab™ environment using the “moving” instruction of Aslak Grinsted [39]. The choice of the statistical descriptors can be

examined and possibly, the dataset dimension might be slightly reduced; this will be discussed in Section 4.3.

The 6 statistical features are the following ones: maximum and minimum values of the pressure amplitude, mean value of the pressure, variance of pressure samples, skewness of the signal (skewness is a measure of the asymmetry of the pressure data around the mean sample), and kurtosis of the signal (kurtosis is used for describing or estimating a distribution's peakedness and frequency of extreme values). Some examples of plots of the statistical features estimated on the pressure signal by applying a scanning window will be shown in Sections 4 and 5.

The next step consists in the creation of the training and testing bases. The 6 statistical descriptors (mean, max, min, var, skewness, kurtosis) are considered on 4 pressure intervals representative of the 4 operating conditions explored. Each one the 4 intervals can be divided into two parts: with n_training observations for the training phase and n_testing observations for the testing phase. In this case, for each operating condition (or class), the sizes of the training and testing bases are equal to 6×n_training and 6×n_testing respectively. The global training over the complete basis is done using the KNN algorithm applied to the 4×6×n_training elements. The global testing is done using the KNN algorithm applied to the 4×6×n_testing elements. The aim of the testing is to verify that the labels of the testing data, resulting from the KNN classification, are identical to the preliminary labels assigned. In other words, the testing phase allows a first assessment of the KNN algorithm performance. Classification rates of 100% should normally be obtained for a "sufficient" minimal size of the scanning window used to compute the statistical features; this issue will be investigated in the examples of Sections 4 and 5.

Once the quality of the KNN classifier is examined and check, new data (i.e. data out of the training and testing intervals) can be classified. This is the validation step, which is also the last step of our diagnosis method. The new data can be selected using a moving window linked with a set of n_new consecutive observations in the pressure signal. In our study, we have decided to apply a moving window with no superimposition between two successive windows of n_new observations. In any case, the 6 statistical features are computed on the new moving window and the new data set is classified using the KNN algorithm (as it was already done with the testing data). The operating condition on the new interval can be predicted and a classification rate can also be calculated, if the preliminary labels of the new points are known.

In the next two sections, the method is applied to two different concrete cases of the experimental database presented in Section 2.3.

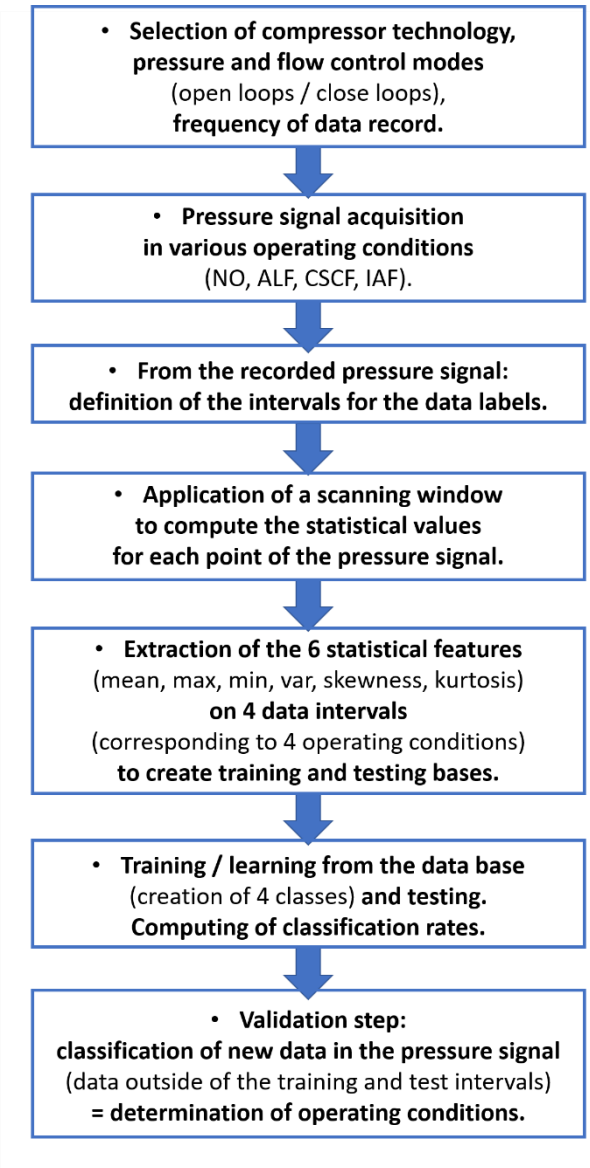

Fig. 4. Block diagram of the diagnosis method.

4. First example of diagnosis (claw compressor, open loops, 1 kHz)

In a first example, the diagnosis method is applied to an air pressure signal recorded at 1 kHz with a claw compressor implemented in the air line. This one is operated in open loops on pressure and flow, under normal and abnormal operating conditions.

4.1 Feature extraction and classification on the full pressure signal

The related pressure signal, which was already presented in Section 2.3. is reported at the top of Fig. 5., a set of subplots that are related with the main steps of the diagnosis method. The

initial labels and time intervals linked with the different applied conditions can be observed in Fig. 5.a). Therefore, in this subplot, each pressure observation has a preliminary label: NO, ALF, IAF, or CSCF. As there is no control through the back-pressure valve in this experiment, the pressure changes due to the various operating conditions can easily be observed on the graph.

The same pressure signal is reported in the second subplot (Fig. 5.b)), which indicates the four intervals of observations selected both for the training and testing. In this case, short data intervals are selected with n_training = 500 observations and n_testing = 200 observations, leading to a global time duration of 0.7s for each one of the 4 intervals. These short durations are mainly chosen as an example but also with respect to the CSCF operating condition. Indeed, the selected CSCF interval encompasses the first occurrence of the CSCF fault. The other and later occurrences of the CSCF faults are not considered for the training and testing. In this example, the NO, ALF and IAF intervals are selected roughly in the middle of the respective time spans of operating conditions so that they can reasonably be considered as representative of these settings (other intervals of similar sizes and chosen close to those presented here lead to similar results). On each interval of 700 observations linked with one operating condition, the training and testing data can be divided either into two time-consecutive sets (first, training data and then, testing data), or into two groups created with a random splitting (i.e. with a random mix of training and testing data).

In the third subplot (Fig. 5.c)), the four training - testing intervals are introduced into the graphs of the 6 statistical descriptors (mean, max, min, var, skewness, kurtosis,) computed using a scanning window with a size equal to 800 observations, i.e. with a duration of 0.8 s. This length is chosen to ensure 100 % classification rates on the testing intervals (additional information will be given on this issue in Section 4.2). All the selected features form a whole data space made of 700×4×6 elements. A first direct, visual examination of the 6 curves in Fig. 5.c) reveals that the mean, maximal, and minimal pressure values show noticeable variations according to the various operating conditions. At first glance, the morphologies of the variance, skewness and kurtosis plots seem to be less impacted by the faults, except on the CSCF intervals.

The classification results obtained with the diagnosis procedure, including the KNN algorithm used with K = 9, applied to the complete pressure signal are displayed in Fig. 6. In this study, the initial pressure signal is divided into consecutive windows of n_new = 100 observations. The 6 statistical features computed in these new windows form some new data sets (matrixes of 6×100 elements) that are classified using the KNN algorithm. One classification rate (in %) is calculated for each one of the 100-observation set. The first point that can be checked from

Fig. 6 is the following one: classification rates of 100% are obtained on the (short) intervals dedicated to the training and testing stages. This means that top results are achieved in this first assessment of the diagnosis method. Mainly 100% classification rates are also obtained outside of the short training and testing intervals. In fact, an average classification rate close to 97% is obtained on the overall pressure signal (Fig. 6). The lower rates are computed during the transients between two operating conditions. The observations at the boundary of two operating modes are the most difficult to classify since the related statistical features have intermediate values with respect to two different conditions. Besides, as it was mentioned in Section 2.3, there is some uncertainty on the preliminary labelling, especially during the transitional phases, and the possibility of mislabeling the data was considered as non-negligible in these cases. Details on the classification are also shown in Fig. 6. For every window of 100 observations, the numbers (and thus the percentage) of elements that are classified as NO, ALF, IAF, and CSCF are given in the four subplots at the bottom of Fig. 6. Moreover, the four graphs indicate with the "True" or "False" mentions whether the labels determined by the KNN algorithm are identical or not to the preliminary labels. Actually, the four graphs are plotted using the elements contained in each confusion matrix computed by the KNN for a window of 100 observations. Let us consider, as an example, the time value t=10.7s in Fig. 6. In this case, all the 100 observations of the related windows used for the KNN computing have NO as preliminary labels (Fig. 5.a)) and the classification rate is equal to 40% (the value is displayed in the first subplot of Fig. 6). The KNN allows a good prediction for 40 NO states: a "True" green square marker is plotted at (t=10,7, y=40) on the subplot which has the y-axis labeled as "Classified as NO (%)". However, 60 observations are "Classified as IAF": a "False" red square marker is plotted at (t=10,7, y=60) on the related subplot. The concordance between the subplot of Fig. 6 and the confusion matrix computed for t=10.7s can be observed in Eq. 2. All correct predictions ("True Positives" and "True Negatives" associated to the green / "True" markers of Fig. 6) are located in the diagonal of the table. The prediction errors (associated to the red / "False" markers of Fig. 6) are represented by values outside the diagonal. The red markers correspond to the "False Positives" for the operating condition X examined on the related subplot (i.e. on the subplot with an ordinate which label name is "Classified as X").

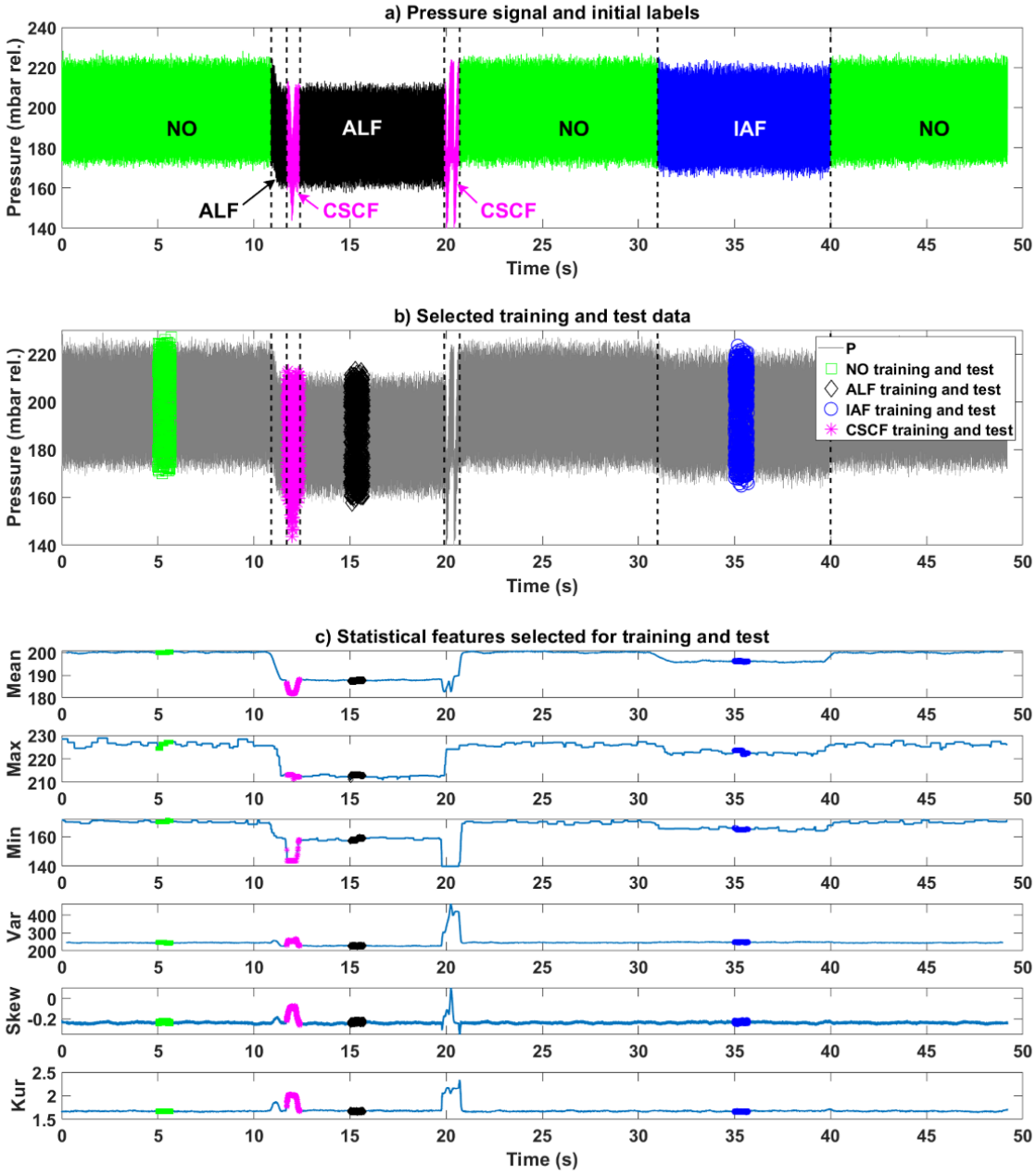

Fig. 5. Example of data features extraction and classification results (Example 1: claw compressor, open loops, 1 kHz): a) Pressure signal and initial labels selected for the operating conditions; b) Pressure signal and selected training and test data; c) Evolution of the 6 statistical features computed for a scanning window with a size of 800 observations.

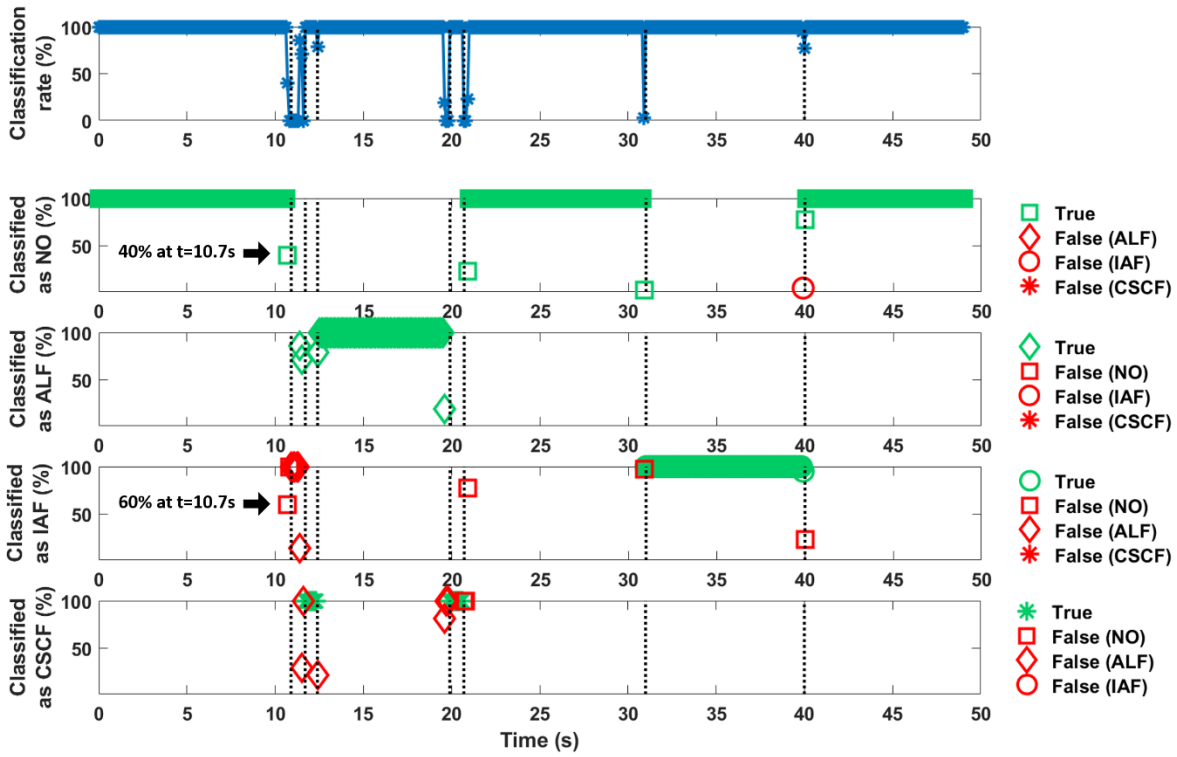

Fig. 6. Example of classification results (Example 1: claw compressor, open loops, 1 kHz).

$$Confusion\ matrix\ computed\ for\ the\ 100\ observations\ at\ t = 10.7\ s\ in\ Fig.6 = \begin{array}{c|cccc} & \multicolumn{4}{c}{Actual\ class} \\ Predicted\ class & NO & ALF & IAF & CSCF \\ \hline NO & \mathbf{40} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ ALF & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ IAF & \mathbf{60} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ CSCF & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \quad (2)$$

A global confusion matrix can be established to summarize all the data samples (i.e. instances) of the full pressure signal profile on the basis of the classification results displayed in Fig. 6. True (or actual) values and values predicted for the 4 classes are cross-tabulated in this matrix (Eq. 3), which has the form of Eq. 1 presented in Section 3.2. All correct predictions for NO, ALF, IAF, and CSCF operating conditions are located in the diagonal of the confusion matrix, while the values outside the diagonal represent the prediction errors (the link between the matrix content and the four last subplots of Fig. 6 can be established quite easily; the "False Positives" are identified with red markers). To express the classifier performance more clearly, the values in the global confusion matrix can also be given in Eq. 4 in percentage of the actual values: $P = [30300 \quad 8300 \quad 9000 \quad 1500]$.

$$\text{Global confusion matrix for Example 1 (Nr. of data samples)} = \text{Predicted class}$$

	Actual class NO	ALF	IAF	CSCF
NO	**29743**	**0**	**5**	**0**
ALF	**0**	**7355**	**0**	**0**
IAF	**357**	**514**	**8995**	**0**
CSCF	**200**	**431**	**0**	**1500**

(3)

$$\text{Global confusion matrix for Example 1 (in \% of actual values)} = \text{Predicted class}$$

	Actual class NO	ALF	IAF	CSCF
NO	**98.2**	**0**	**0.1**	**0**
ALF	**0**	**88.6**	**0**	**0**
IAF	**1.2**	**6.2**	**99.9**	**0**
CSCF	**0.7**	**5.2**	**0**	**100**

(4)

As mentioned in Section 3.2, such a confusion matrix generates a good overall picture of how the KNN algorithm performs on the full pressure profile and it can be used to gauge the quality of the classifier. In Eq. 4, high rates of correct predictions can be observed in the diagonal of the matrix (98.2%, 88.6%, 99.9%, and 100% rates are obtained for NO, ALF, IAF, and CSCF respectively) while the prediction errors (i.e. the values outside the diagonal) are low (the values are ranging from 0 to 6.2%; besides, 7 zero values can be observed outside of the diagonal).

In order to further interpret the KNN results, some statistics or metrics associated to the matrix shall be computed following the notations and mathematical relations presented in Table 2. As the calculations for this 4 classes matrix are a little more complex than in the simple case of a 2×2 binary table, the programming code proposed in [38] can help to compute the associated performance indicators. The calculations are detailed in Table 3. An excellent accuracy rate is obtained (0.97). The sensitivity, specificity, and NPV metrics are also excellent with values equal to 0.97, 0.99, and 0.99 respectively. The precision (equal to $0.90 = mean([1 \quad 1 \quad 0.91 \quad 0.70])$) can be considered as very good (in detail: excellent for NO and ALF with 1, very good for IAF with 0.91, and correct / good for CSCF with 0.70). Note that the very good precision ($(TP./(TP + FP))$) could already be observed on each one of the four last subplots of Fig. 6. Indeed, on each subplot, limited numbers of false predictions ("False Positives") were displayed with red markers while numerous instances were characterized as "True Positives" with green markers.

Table 3. Calculations of performance indicators for the global confusion matrix related with the first example of diagnosis.

Appellations	**Calculations**
Total population	$Total\ Population = $Sum of the values in the matrix $= 49100$
Positives	$P = [30300 \quad 8300 \quad 9000 \quad 1500]$
Negatives	$N = Total\ Population - P = 49100 - P = [18800 \quad 40800 \quad 40100 \quad 47600]$
True Positives	$TP = [29743 \quad 7355 \quad 8995 \quad 1500]$
False Positives	$FP = [(0+5+0) \quad 0 \quad (357+514+0) \quad (200+431+0)] = [5 \quad 0 \quad 871 \quad 631]$
False Negatives	$FN = [(0+357+200) \quad (0+514+431) \quad (5+0+0) \quad (0+0+0)]$ $= [557 \quad 945 \quad 5 \quad 0]$
True Negatives	$TN = N - FP = [18795 \quad 40800 \quad 39229 \quad 46969]$
Predicted Positives	$\hat{P} = [29748 \quad 7355 \quad 9866 \quad 2131]$
Predicted Negatives	$\hat{N} = Total\ Population - \hat{P} = [19352 \quad 41745 \quad 39234 \quad 46969]$
Metrics	**Calculations and results**
Accuracy	$Accuracy = mean\ (TP + TN)\ /\ Total\ Population = 0.97$
Sensitivity	$Sensitivity = mean\ (TP./P) = mean([0.98 \quad 0.88 \quad 0.99 \quad 1]) = 0.97$
Precision	$Precision = mean\ (TP./(TP + FP)) = mean([1 \quad 1 \quad 0.91 \quad 0.70]) = 0.90$
Specificity	$Specificity = mean\ (TN./N) = mean([1 \quad 1 \quad 0.98 \quad 0.99]) = 0.99$
Negative Predictive Value	$NPV = mean\ (TN./(FN + TN)) = mean([0.97 \quad 0.98 \quad 1 \quad 1]) = 0.99$

In this first example of diagnosis, the classification rates obtained on the dynamical pressure profile and displayed in Fig. 6 can be assessed as very good in general (the transient phases generate the most difficult data to process, due to their uncertainties on the preliminary labelling leading to false alarms). The performance indicators calculated from the global confusion matrix can also be considered as very good. These results can be achieved especially because well-suited choices are made in the application of the KNN method, for instance in the selection of the training and testing data intervals. Another parameter of paramount importance in the proposed diagnosis strategy is the length of the scanning window used to compute the 6 statistical descriptors. In this first example, as previously mentioned, a scanning window

including 800 observations was selected to ensure 100 % classification rates on the testing intervals. Further explanations are given on this issue in the following sub-section.

4.2 Effect of the scanning window size on the classification rate

Moving window methods are applied to process data in smaller batches at a time, typically in order to statistically represent a neighborhood of points in the data. The moving average is a common data smoothing technique that slides a window along the data, computing the mean of the points inside of each window. This can be helpful to eliminate insignificant variations from one data point to the next one [40].

In our case, the length of the window used to compute the statistical descriptors should be neither too short (it could not statistically represent a neighborhood of points in the pressure data), nor too long (the duration of the diagnosis process would be increased, which would obviously constitute a disadvantage).

The impact of the scanning window size is studied on the basis of the 4 pressure data intervals selected in the current example for the training and testing. Various window sizes, ranging from 50 to 1000 with steps of 50, are considered and for each size, the whole diagnosis method (i.e. computing of statistical descriptors, plus KNN method) is applied to the data intervals. A KNN classification rate is finally computed. As the 4 data intervals need to be divided into training and testing observations, we decide to evaluate the performance of the diagnosis strategy on 10 different sets from random data splits. Then, for each window size, a mean classification rate can be computed from the 10 KNN values. The effect of the scanning window size on the classification rate can be observed in Fig. 7. The lowest mean KNN rate (about 83%) is computed for the shortest window (50 observations made in 0.05 s). A 100% rate is reached for a window size close to 0.5 s and a steady rate of 100% is reached for a 0.7 s length (see Fig. 7 and the mean values computed from the 10 data splits). For memory, in Fig. 5-6, a value of 0.8 s (800 observations) was selected to ensure 100 % classification rates on the testing intervals.

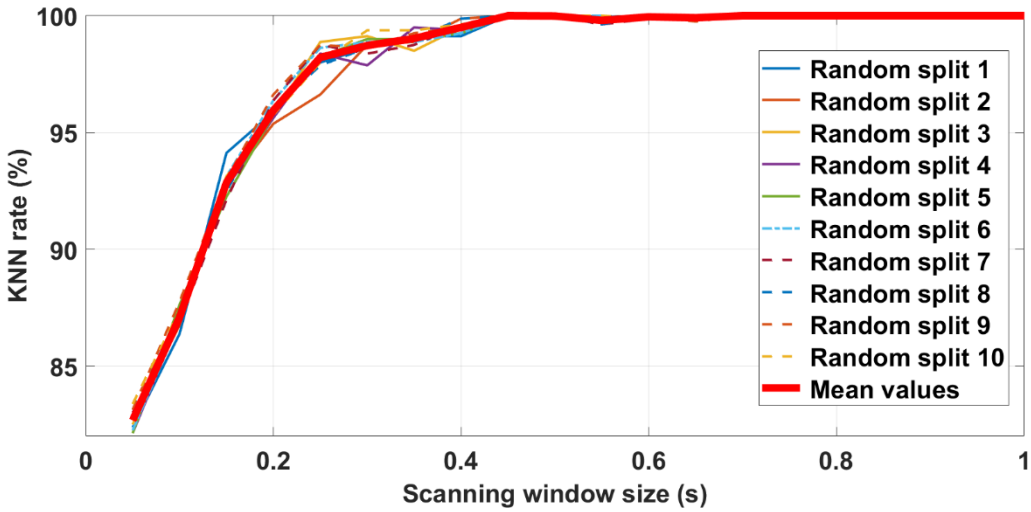

Fig. 7. Impact of the scanning window size for 10 data splits made between training and test data (recorded at 1 kHz).

5. Second example of diagnosis (regenerative blower, closed loops, 100 Hz)

In the second example, the diagnosis method is applied to a different compressor technology, a regenerative blower, tested in the air line that is operated now in closed loops on pressure and flow (on the contrary of the first example), still under normal and abnormal operating conditions. The air pressure signal is now recorded at 100 Hz (i.e. not at 1 kHz anymore): the aim is to evaluate how the proposed algorithm is affected by the sampling frequency reduction (which can be seen as a benefit in terms of sensor and computing costs). An approach similar to that used in the first example is adopted.

Two different cases are considered, depending on the size of the data sets used for the learning - testing steps of the KNN algorithm. In Section 5.1, a first attempt is made with short data sets (i.e. with sizes close to those of the first example). Then, larger data intervals are selected (Section 5.2).

5.1 Preliminary tests with short data intervals for the KNN learning - testing steps

The diagnosis approach is illustrated through Fig. 8. The pressure signal, already presented in Section 2.3. is reported at the top of Fig. 8., the set of subplots related with the main steps of the diagnosis method. The initial labels (NO, ALF, IAF, CSCF) and time intervals linked with the various applied conditions can be observed in Fig. 8.a). Unlike the first example, a pressure

control at 200 mbar rel. is performed using the back-pressure valve. Therefore, it is much more difficult to observe the pressure variations due to the changes of the operating conditions, except for CSCF (in this case, the quick dynamical disturbances linked with the fault cannot effectively be attenuated by the pressure regulation).

The pressure signal is reported in the subplot of Fig. 8.b), with the four training and testing intervals. In this case, slightly longer intervals (14 s) are selected with n_training = 1100 observations and n_testing = 300 observations, for each one of the 4 intervals. The choice is done as a function of the duration of the first CSCF occurrence. The NO, ALF, and IAF training and testing intervals are selected roughly in the middle of the time span of each operating condition in order to be considered as representative of these settings. As in the first example of diagnosis, on each interval of 1400 observations linked with one operating condition, the training and testing data can be divided either into two time-consecutive sets, or into two groups created with a random splitting. The plots of the 6 statistical descriptors (mean, max, min, var, skewness, kurtosis) will not be shown in Fig. 8. but the descriptors are computed using a scanning window with a size equal to 800 observations (equivalent to a duration of 8 s). Like in the first diagnosis example, this value is chosen to ensure 100 % classification rates on the testing intervals. The impact of the scanning window size can possibly be studied following the approach described in Section 4.2. All the selected features lead therefore to a whole data space made of 1400×4×6 elements. The classification results obtained with the diagnosis procedure, including the KNN algorithm (K = 9), applied to the complete pressure signal are displayed in the last subplot (Fig. 8.c)). The initial pressure signal is divided into consecutive windows of n_new = 100 observations. The 6 statistical features computed in these new windows lead to new data sets (6×100 matrixes) that are classified using KNN. Here, one classification rate (in %) corresponds to one 100-observation set. It is then possible to verify that classification rates of 100% are obtained on the short intervals used for the training and testing processes. Outside of the short training and testing intervals, the KNN classifier yields mixed results. A mean classification rate close to 66% is obtained on the overall pressure signal.

The classification results can be further analyzed using the global confusion matrix of Eq. 5 that offers an overall “mathematical picture” of the diagnosis made on the whole pressure profile. The global confusion matrix is also expressed in percentage (Eq. 6) of the P actual values, with $P = [33000 \quad 8000 \quad 7000 \quad 8000]$. In order to gauge the quality of the diagnostic tool, statistics or metrics associated with this global confusion matrix will be calculated and used later in Section 5.2. Comments can be done on the contents of the global confusion matrix, as follows. Elements with NO as initial label are mostly classified as NO.

However, confusions are done with ALF and mainly with IAF. This can be explained as follows: in these two faulty modes, the consequences of the air leakage and clogged air line inlet on the pressure can be well mitigated by the back-pressure and compressor with actions that compensate for the change of pressure (this could already be seen on the pressure signal displayed at the top of Fig. 8.). Elements with ALF initial labels are mostly classified as ALF (confusions can be seen with NO and IAF). Elements with IAF as preliminary label are mostly well classified (there are confusions with NO). Elements with CSCF as initial label are mostly well classified (main confusions are done with IAF). It is also possible to remark that the elements with NO, ALF, IAF as initial labels are barely not classified as CSCF.

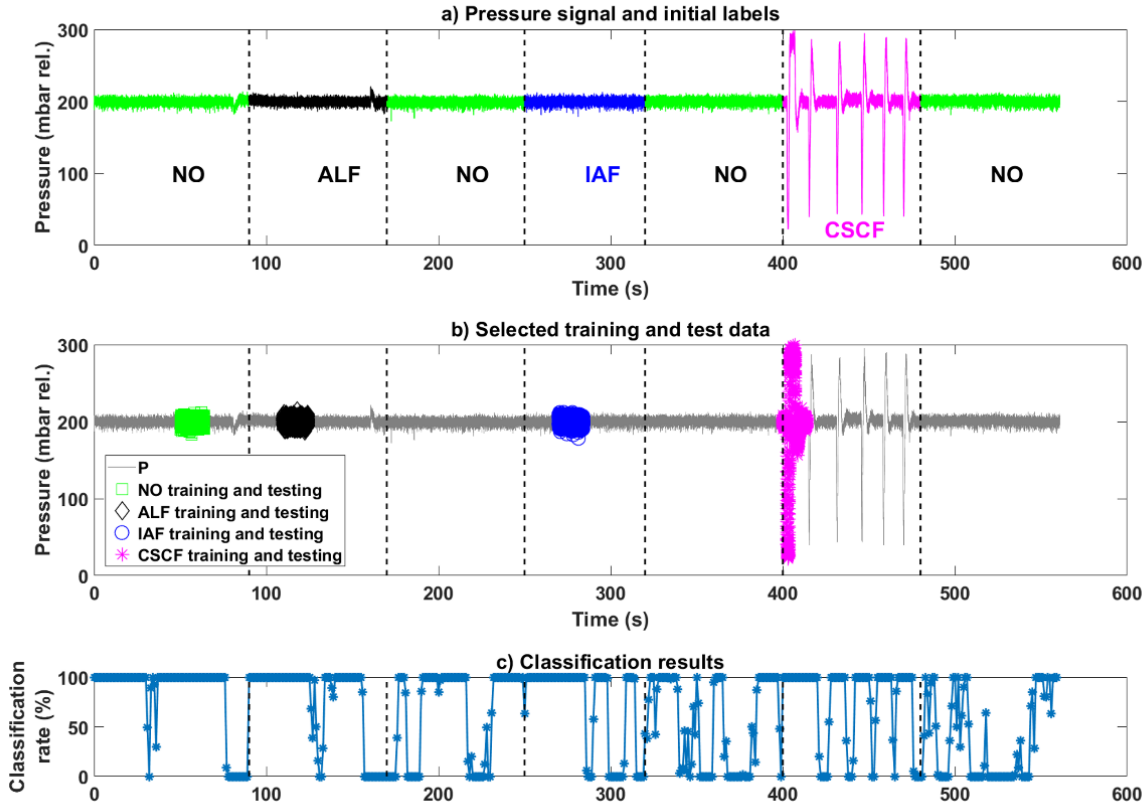

Fig. 8. Example of data features extraction and classification results (regenerative blower, closed loops, 100 Hz): a) Pressure signal and initial labels selected for the operating conditions; b) Pressure signal and selected training and test data (n_training = 1100 and n_testing = 300); c) Classification results using the 6 statistical features computed for a scanning window with a size of 800 observations.

$$\text{Global confusion matrix for Example 2 and Fig. 9 (Nr. of data samples)} = \begin{array}{c|cccc} & & \text{Actual class} & & \\ & NO & ALF & IAF & CSCF \\ \text{Predicted class} & & & & \\ NO & \mathbf{19957} & \mathbf{738} & \mathbf{1964} & \mathbf{69} \\ ALF & \mathbf{839} & \mathbf{6152} & \mathbf{0} & \mathbf{65} \\ IAF & \mathbf{12069} & \mathbf{1110} & \mathbf{5036} & \mathbf{1779} \\ CSCF & \mathbf{135} & \mathbf{0} & \mathbf{0} & \mathbf{6087} \end{array} \quad (5)$$

$$\text{Global confusion matrix for Example 2 and Fig. 9 (in \% of actual values)} = \begin{array}{c|cccc} & & \text{Actual class} & & \\ & NO & ALF & IAF & CSCF \\ \text{Predicted class} & & & & \\ NO & \mathbf{60.5} & \mathbf{9.2} & \mathbf{28.1} & \mathbf{0.9} \\ ALF & \mathbf{2.5} & \mathbf{76.9} & \mathbf{0} & \mathbf{0.8} \\ IAF & \mathbf{36.6} & \mathbf{13.9} & \mathbf{71.9} & \mathbf{22.2} \\ CSCF & \mathbf{0.4} & \mathbf{0} & \mathbf{0} & \mathbf{76.1} \end{array} \quad (6)$$

To conclude the comments on this second example with short data intervals selected for the KNN learning - testing steps, it can be stated that the classification results obtained in this second example illustrating the diagnostic method are not as good as in the first example. The gap of performance is not due to the different compressor technologies nor to the lower data record frequency. Obviously, the lower data record frequency leads to the need of longer durations for the scanning time window used to compute the statistical features. But the main difficulty encountered in successfully applying the diagnosis method comes here from the closed loops applied in the control of air pressure and flow. These regulations tend to erase the impact of operating faults on the morphology of the pressure signal and this one no longer carries as much information as before for the diagnosis.

5.2 Larger data intervals for the KNN learning - testing steps

Assuming that the closed loop controls need to be retained for the air pressure and flow, one possible solution to enhance the diagnosis results is to consider larger data intervals for the learning (and testing) steps in the KNN algorithm. This is illustrated by Fig. 9 where this solution is adopted. Classification rates of 100% are obtained on the larger intervals dedicated to the training and testing stages. Nevertheless, there are still confusions between NO and IAF states: many NO elements outside of the NO training and testing intervals are classified as IAF. Besides, only a few NO and CSCF elements are classified erroneously as ALF. And only a few NO instances are classified erroneously as CSCF. These “errors” of classification are made in the transients between different operating states and they can also be related with any possible initial mislabeling. The results of Fig. 9 indicate the potential of the diagnosis method when extended data bases are considered in the training step.

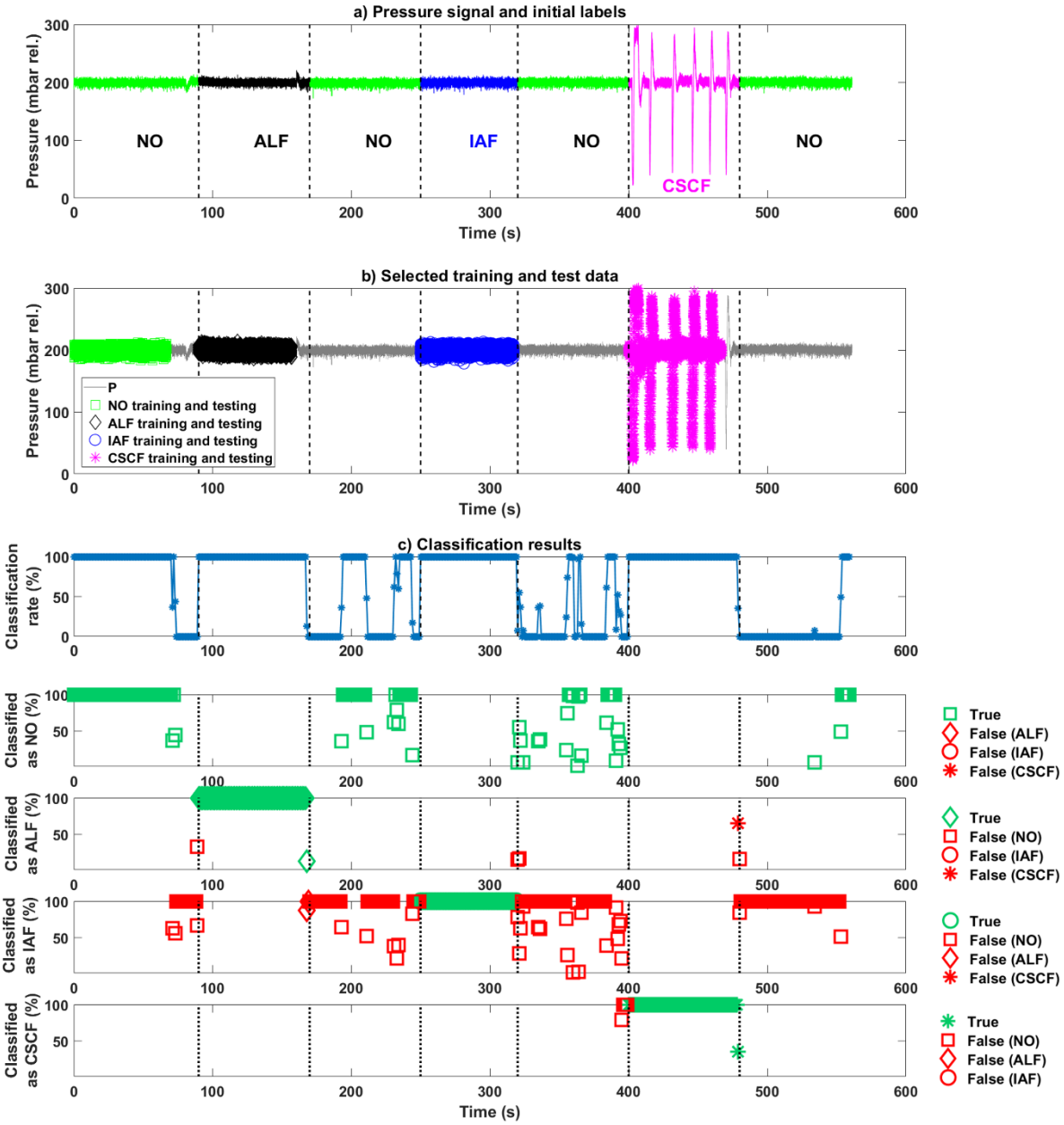

Fig. 9. Example of data features extraction and classification results (regenerative blower, closed loops, 100 Hz): a) Pressure signal and initial labels selected for the operating conditions; b) Pressure signal and selected training and test data (n_training = 6500 and n_testing = 200); c) Classification results using 6 the statistical features computed for a scanning window with a size of 1500 observations.

The global confusion matrix of Eq. 7 summarizes all the data samples resulting from the classification of Fig. 9.c). The global confusion matrix is also expressed in percentage of the actual values (Eq. 8).

$$\text{Global confusion matrix for Example 2 and Fig. 10 (Nr. of data samples)} = \text{Predicted class}$$

	Actual class NO	ALF	IAF	CSCF
NO	**12611**	**0**	**0**	**0**
ALF	**81**	**7813**	**0**	**65**
IAF	**19829**	**187**	**7000**	**0**
CSCF	**479**	**0**	**0**	**7935**

(7)

$$\text{Global confusion matrix for Example 2 and Fig. 10 (in \% of actual values)} = \text{Predicted class}$$

	Actual class NO	ALF	IAF	CSCF
NO	**38.2**	**0**	**0**	**0**
ALF	**0.2**	**97.7**	**0**	**0.8**
IAF	**60.1**	**2.3**	**100**	**0**
CSCF	**1.5**	**0**	**0**	**99.2**

(8)

Some statistics or metrics associated to this global confusion matrix are presented in Table 4; with calculations similar to those already shown in Table 3. To compare and observe the effects of the different data interval lengths selected for the learning - testing steps used in the KNN algorithm, the same statistics are reported in Table 4 for the global confusion matrix of Eq. 5 (that was related with Fig. 8.c)). The metrics of Table 3 computed for the first example (claw compressor, open loops, 1 kHz) are also reported in the last column of Table 4, as a reminder. In the two cases related with the second example of diagnosis, whatever the learning - testing interval sizes, the accuracy rate and the other metrics values are clearly lower than those of the first example where the metrics could be considered as excellent (sensitivity, specificity, and NPV) or very good (precision). This quick comparison between the two examples highlights the net (negative) effect of the pressure control mode on the diagnosis results.

In the second example, the change of the learning - testing interval lengths does not really impact the accuracy rate, nor the (mean) specificity and Negative Predictive Value (NPV) metric (Table 4). In detail, small variations can be observed for the specificity and NPV values computed for each operating case (this cannot be observed on the mean statistical values). For instance, the specificity linked with NO changes positively from 0.88 to 1 with larger learning - testing intervals. On the contrary, the NPV value linked with NO varies negatively from 0.61 to 0.53. Larger learning - testing intervals may therefore have different effects (i.e. either positive or negative) on the statistical measures linked with different operating conditions. The main interest of increasing the learning - testing interval sizes is the improvements of the mean sensitivity value (from 0.71 to 0.84) and mean precision rate (from 0.74 to 0.8). In detail (that is considering the values computed for each operating condition), the increase of the interval lengths is not beneficial for the sensitivity linked with the NO case (from 0.61 to 0.38; there are

more “False Negatives” for NO in Eq. 7 and 8 than in Eq. 5 and 6). However, the increase of the interval lengths has clearly positive impacts for the other operating cases, namely ALF, IAF, and CSCF. In Table 4, one can also note that the precisions for NO and ALF are better with larger learning - testing intervals (increase from 0.88 to 1, and from 0.87 to 0.98 respectively); lower numbers of false predictions (“False Positives”) are thus displayed with red markers in Fig. 9.c) for NO and ALF while numerous instances are characterized with green markers as “True Positives”.

Table 4. Performance indicators for the second example of diagnosis and comparisons between the two cases: short and larger learning / testing intervals. Last column: reminder of the metrics computed for the first example (claw compressor, open loops, 1 kHz).

Metrics	**Results for the confusion matrix of Eq. 5 related with Fig. 8.c) (short learning - testing intervals)**	**Results for the confusion matrix of Eq. 7 related with Fig. 9.c) (larger learning - testing intervals)**	**Results obtained for the first example**
Accuracy	0.66	0.63	0.97
Sensitivity	0.71 = mean ([0.61, 0.77, 0.72, 0.76])	0.84 = mean ([0.38, 0.98, 1, 0.99])	0.97
Precision	0.74 = mean ([0.88, 0.87, 0.25, 0.98])	0.8 = mean ([1, 0.98, 0.26, 0.94])	0.90
Specificity	0.89 = mean ([0.88, 0.98, 0.69, 0.99])	0.89 = mean ([1, 0.99, 0.59, 0.99])	0.99
NPV	0.87 = mean ([0.61, 0.96, 0.95, 0.96])	0.88 = mean ([0.53, 1, 1, 1])	0.99

6. Conclusions

The application-oriented tool presented in this article is proposed to detect and identify typical faults that may occur in FC air supply and conditioning subsystems. The fault scenarios are reproduced on a specific experimental testbench specifically designed to emulate a FC air line and to examine the portability of the diagnostic tool with two various compressor technologies, two pressure / flow mode controls, and two data record frequencies. A data-driven diagnosis approach based on the collected pressure datasets is selected with the aim to utilize the fault stamps present in the morphology of the pressure transmitter signals.

A clear discrimination between the operating modes is obtained when the pressure is not regulated. In this case, excellent classification rates can be obtained if well-suited choices are made in the application of the KNN method, for instance in the selection of the training and

testing data intervals. Mainly the observations at the boundary of two steady-state operating conditions are difficult to classify.

The discrimination between the operating modes is much more difficult when the pressure control mode is applied. In that case, the fault stamps in the morphology of the pressure signal tend to be removed and one possible solution to increase the performance of the diagnostic tool consists in lengthening the data intervals related with the learning step. However, in the view of a real application, the diagnosis procedure should preferably be applied to the FC air supply subsystems operated in open loop, for example before the start-up, after the shut-down or during idle mode phases of the FC generator.

Regarding the frequency of the data record, both 1000 and 100 Hz values are found effective for the isolation of the air subsystem. The highest frequency level allows shorter detection times while the lowest frequency should be more appropriate to reduce the costs of the data acquisition and process system.

Acknowledgments

The work performed was done within the European Giantleap project, entitled “Giantleap Improves Automation of Non-polluting Transportation with Lifetime Extension of Automotive PEM fuel cells”. This project has received funding from the Fuel Cells and Hydrogen 2 Joint Undertaking under grant agreement No 700101. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and Hydrogen Europe and N.ERGHY. The authors would like to show their sincere appreciation to all the partners of this project.

References

[1] M. Mench, E. Caglan, Kumbur T. Nejat Veziroglu. Polymer Electrolyte Fuel Cell Degradation. 1st Edition. USA. Academic Press. 2011. 472 pages.

[2] J. Wang. System integration, durability and reliability of fuel cells: Challenges and solutions. Applied Energy 2017;189:460-479.

[3] B. Blunier, A. Miraoui. Air Management in PEM Fuel Cells: State-of-the-Art and Prospectives. International Aegean Conference on Electrical Machines and Power Electronics. 10-12 Sept. 2007. Bodrum, Turkey. DOI: 10.1109/ACEMP.2007.4510510.

[4] J. Hou, M. Yang, C. Ke, J. Zhang. Control logics and strategies for air supply in PEM fuel cell engines. Applied Energy 2020;269:115059.

[5] J. Larminie, A. Dicks. Fuel Cell Systems Explained, Second Edition. 2003. John Wiley & Sons, Ltd. ISBN: 0-470-84857-X.

[6] J. Schröter, T. Graf, D. Frank, C. Bauer, J. Kallo, C. Willich. Influence of pressure losses on compressor performance in a pressurized fuel cell air supply system for airplane applications. International Journal of Hydrogen Energy 2021;46(40):21151-21159.

[7] Y. Zhang, S. Xu, Y. Wan. Performance improvement of centrifugal compressors for fuel cell vehicles using the aerodynamic optimization and data mining methods. International Journal of Hydrogen Energy 2020;45(19):11276-11286.

[8] Z. Deng, Q. Chen, L. Zhang, Z. Fu. Data driven NARMAX modeling for PEMFC air compressor. International Journal of Hydrogen Energy 2020;45(39):20321-20328.

[9] J. K. Gruber, M. Doll, C. Bordons. Design and experimental validation of a constrained MPC for the air feed of a fuel cell. Control Engineering Practice 2009;17(8):874-885.

[10] I. Matraji, S. Laghrouche, S. Jemeï, M. Wack. Robust control of the PEM fuel cell air-feed system via sub-optimal second order sliding mode. Applied Energy 2013;104:945-57.

[11] S. Laghrouche, I. Matraji, F.S. Ahmed, S. Jemeï, M. Wack. Load governor based on constrained extremum seeking for PEM fuel cell oxygen starvation and compressor surge protection. Int J Hydrogen Energy 2013;38(33):14314-14322.

[12] D. K. Kim, H. E. Min, I. M. Kong, M. K. Lee, C. H. Lee, M. S. Kim, H. H. Song. Parametric study on interaction of blower and back pressure control valve for a 80-kW class PEM fuel cell vehicle. International Journal of Hydrogen Energy 2016; 41(39):17595-17615.

[13] D. Yang, Y. Wang, Z. Chen. Robust fault diagnosis and fault tolerant control for PEMFC system based on an augmented LPV observer. International Journal of Hydrogen Energy 2020;45(24):13508-13522.

[14] H. Chen, X. Zhao, T. Zhang, P. Pei. The reactant starvation of the proton exchange membrane fuel cells for vehicular applications: A review. Energy Conversion and Management 2019;182:282-298.

[15] A. Benmouna, M. Becherif, D. Depernet, F. Gustin, H.S. Ramadan, S. Fukuhara. Fault diagnosis methods for Proton Exchange Membrane Fuel Cell system. International Journal of Hydrogen Energy 2017;42(2):1534-1543.

[16] R.-H. Lin, X.-N. Xi, P.-N. Wang, B.-D. Wu, S.-M. Tian. Review on hydrogen fuel cell condition monitoring and prediction methods. International Journal of Hydrogen Energy 2019;44(11):5488-5498.

[17] T. Sutharssan, D. Montalvao, Y. K. Chen, W.-C. Wang, C. Pisac, H. Elemara. A review on prognostics and health monitoring of proton exchange membrane fuel cell. Renewable and Sustainable Energy Reviews 2017;75:440-450.

[18] Z. Li, R. Outbib, D. Hissel, S. Giurgea. Data-driven diagnosis of PEM fuel cell: A comparative study. Control Engineering Practice 2014;28:1-12.

[19] Z. Zheng, R. Petrone, M.-C. Péra, D. Hissel, M. Becherif, C. Pianese, N. Yousfi Steiner, M. Sorrentino. A review on non-model based diagnosis methodologies for PEM fuel cell stacks and systems. International Journal of Hydrogen Energy 2013;38(21):8914-8926.

[20] D. Hissel, D. Candusso, F. Harel. Fuzzy-Clustering Durability Diagnosis of Polymer Electrolyte Fuel Cells Dedicated to Transportation Applications. IEEE Transactions on Vehicular Technologies 2007;4(5):1211-2420.

[21] D. Benouioua, D. Candusso, F. Harel, P. Picard, X. François. On the issue of the PEMFC operating fault identification: Generic analysis tool based on voltage pointwise singularity strengths. International Journal of Hydrogen Energy 2018;43(25):11606-11613.

[22] F. Barbir, H. Gorgun, X. Wang. Relationship between pressure drop and cell resistance as a diagnostic tool for PEM fuel cells. Journal of Power Sources 2005;141(1):96-101.

[23] N. Yousfi Steiner, D. Candusso, D. Hissel, P. Moçoteguy. Model-based diagnosis for proton exchange membrane fuel cells. Mathematics and Computers in Simulation 2010;81(2):158-170.

[24] A. M. Niroumand, W. Mérida, M. Eikerling, M. Saif. Pressure-voltage oscillations as a diagnostic tool for PEFC cathodes. Electrochemistry Communications 2010;12(1):122-124.

[25] G. Tian, S. Wasterlain, I. Endichi, D. Candusso, F. Harel, X. François, M.-C. Péra, D. Hissel, J.M. Kauffmann. Diagnosis methods dedicated to the localisation of failed cells within FEMFC stacks. Journal of Power Sources 2008;182(2):449-461.

[26] M.-H. Wang, M.-J. Chen. Design of a Fault Diagnosis System for PEM Fuel Cells. International Symposium on Computer, Consumer and Control. 4-6 June 2012. Taichung, Taiwan. DOI: 10.1109/IS3C.2012.122

[27] W.-Y. Lee, H. Oh, M. Kim, Y.-Y. Choi, Y.-J. Sohn, S.-G. Kim. Hierarchical fault diagnostic method for a polymer electrolyte fuel cell system. International Journal of Hydrogen Energy 2020;45(47):25733-25746.

[28] J. Won, H. Oh, J. Hong, M. Kim, W.-Y. Lee, Y.-Y. Choi, S.-B. Han. Hybrid diagnosis method for initial faults of air supply systems in proton exchange membrane fuel cells. Renewable Energy 2021;180:343-352.

[29] K. Fukunaga. Introduction to statistical pattern recognition (2nd ed.), Academic Press Inc., New York (1992).

[30] C. Bishop. Pattern recognition and machine learning. Springer-Verlag New York (2006).

[31] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician 1992;46(3):175-185.

[32] D. Nelson. What is K-Nearest Neighbors? Published on February 23, 2020. https://www.unite.ai/what-is-k-nearest-neighbors/. Last access: 07/06/2021.

[33] B. S. Everitt, S. Landau, M. Leese, D. Stahl. Miscellaneous Clustering Methods, in Cluster Analysis, 5th Edition (2011), John Wiley & Sons, Ltd, Chichester, UK.

[34] David M. W. Powers. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies 2011;2(1):37-63.

[35] I. Düntsch, G. Gediga. Indices for rough set approximation and the application to confusion matrices. International Journal of Approximate Reasoning 2020; 118:155-172.

[36] M. Santini. Machine Learning for Language technology 2015. Lecture 3: Basic concepts of machine learning - induction & evaluation. Uppsala University, Sweden. https://www.slideshare.net/marinasantini1/lecture-3-basic-concepts-of-machine-learning-induction-evaluation. Last access: 07/06/2021.

[37] D. Nelson. What is a Confusion Matrix? Published on December 13, 2019. https://www.unite.ai/what-is-a-confusion-matrix/. Last access: 07/06/2021.

[38] A. Manthiri S (2020). Multi Class Confusion Matrix. MATLAB Central File Exchange. Retrieved February 25, 2020. https://www.mathworks.com/matlabcentral/fileexchange/60900-multi-class-confusion-matrix. Last access: 07/06/2021.

[39] A. Grinsted. Moving averages / Moving median etc. Version 1.1.0.0 (1.66 KB). Calculates moving averages (or median/fun) of a timeseries. https://fr.mathworks.com/matlabcentral/fileexchange/8251-moving-averages-moving-median-etc?focused=5193309&tab=function. Last access: 07/06/2021.

[40] MathWorks - Matlab documentation. https://fr.mathworks.com/help/matlab/data_analysis/data-smoothing-and-outlier-detection.html. Last access: 01/09/2020.