

Research Article

Joint Feature and Model Selection for SVM Fault Diagnosis in Solid Oxide Fuel Cell Systems

**Gabriele Moser,¹ Paola Costamagna,² Andrea De Giorgi,¹ Andrea Greco,³
Loredana Magistri,³ Lissy Pellaco,¹ and Andrea Trucco^{1,4}**

¹DITEN, University of Genoa, 16145 Genoa, Italy

²DICCA, University of Genoa, 16145 Genoa, Italy

³DIME, University of Genoa, 16145 Genoa, Italy

⁴PAVIS, Istituto Italiano di Tecnologia, 16163 Genoa, Italy

Correspondence should be addressed to Andrea Trucco; trucco@dibe.unige.it

Received 5 September 2014; Accepted 27 January 2015

Academic Editor: Vincent Cocquempot

Copyright © 2015 Gabriele Moser et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper describes an original technique for the joint feature and model selection in the context of support vector machine (SVM) classification applied as a diagnosis strategy in model-based fault detection and isolation (FDI). We demonstrate that the proposed technique contributes to the solution of an open research problem: to design a robust FDI procedure, correctly functioning with different operating conditions and fault sizes, specifically settled for an electric generation system based on solid oxide fuel cells (SOFCs). By using a quantitative model of the generation system coupled to an optimized SVM classifier, a satisfactory FDI procedure is achieved, which is robust against modeling and measurement errors and is compliant with practical deployment.

1. Introduction

The interest in electric generation plants based on the fuel cell (FC) technology [1, 2] is constantly growing owing to their high energy conversion efficiency and environmental compatibility. However, plants based on FC stacks still suffer from a low reliability and a limited lifetime, and thus the development of specific methods for the automatic online fault diagnosis is of paramount importance for their commercial diffusion. According to [3, 4], among the possible FDI approaches, the model-based scheme [5, 6] is currently the preferred one in the context of the FC technology.

Although systems based on SOFC stacks are universally reputed to be one of the best options for distributed electric generation plants, the literature regarding the FDI procedure in these systems is still scarce [7–10]. Moreover, in many of these papers [7, 8], the proposed diagnosis strategy [6] is limited to inference approaches that use a binary fault signature matrix arranged according to a fault tree analysis (i.e., a deductive top-down tool, typically used in safety and reliability engineering) or an improved version of such

a matrix [10]. To overcome the weaknesses of the binary signature matrix [11], we recently proposed [9] a supervised classification approach, implemented through a SVM, as a possible diagnosis strategy. We demonstrated that the detection and isolation of faults of random size, occurring in an SOFC generation system that works under many different steady-state operating conditions, is possible by using a supervised SVM classifier.

However, only preliminary results (obtained under the assumption that the model provides exact predictions) are reported in [9], leaving many issues unresolved. One of them is the choice of the best physicochemical variables to be measured during the system operation and used for the FDI procedure. Although virtually all the physicochemical variables that characterize the functioning of an SOFC plant can be predicted by the related mathematical model [6], the practical measurement of these variables offers different levels of difficulty, from variables that are easy to measure to variables that are extremely difficult to measure. Therefore, the evaluation of the contribution that each of these variables provides to the FDI procedure is of crucial importance.

An additional open issue regards the model selection for the SVM classifier. Model selection generally consists in tuning the parameters of the SVM classifier, affecting the FDI performance and possibly choosing the kernel function. When the kernel is predefined, the values of the parameters of the SVM need to be optimized prior to the training phase of the classifier. In turn, this optimization depends on the features that are chosen for the classification procedure. Because in model-based FDI the residuals (i.e., the differences between the values of the physicochemical variables measured during SOFC system operation and the values of the same variables predicted by the system model) are used as features [6], it is evident that the selection of the variables to be used and the parameter optimization are strictly interconnected and need a combined investigation.

In this paper, we propose an original technique to carry out feature selection and parameter optimization jointly, in the context of a model-based FDI procedure applied to SOFC systems and performed through SVM classification. Dimensionality reduction has received great attention in the fault diagnosis field [12–15]. It can be performed with projections that create a set of a few synthetic features (i.e., feature transformation methods [12, 13, 15]) or by selecting the most relevant features (i.e., a subset of residuals) for diagnosis [14, 16, 17].

To determine, among the variables that can be measured in the SOFC system, those which play the most critical role in view of the discrimination among the considered fault classes, we focus on feature selection. This approach preserves the physical meaning and interpretability of the features used for classification purposes. Here, a technique based on the minimization of an analytical error bound [18, 19] is proposed. Many previous feature selection methods apply discrete optimization algorithms either to interclass distance measures (e.g., Bhattacharyya or Jeffries-Matusita distances), computed through parametric (usually multivariate Gaussian) models for the feature statistics conditioned to class membership, or to the classification accuracy on a validation set [17]. However, the former method does not fit the FDI procedure addressed here well because no well-defined parametric model is available for the joint statistics of highly heterogeneous physicochemical variables of an SOFC system. The latter method would be pursuable in the present application, but it would require setting aside part of the samples available for the training phase and using them only for validation purposes. Instead, the proposed technique extends the algorithm in [20] by combining it with a nonparametric error bound that can be derived as a by-product of the SVM training. In this way, the feature selection is performed without reserving a portion of the training samples for the validation.

Moreover, while a given kernel function (e.g., a Gaussian radial basis function, RBF) is predefined, the feature selection technique is also integrated with the optimization of the parameters of the SVM classifier (i.e., the model selection) in a single, innovative method. Working in this way, feature selection indicates which variables need to be measured in the SOFC system to attain a given (optimized) level of performance. In addition, by combining this outcome with

the measurement difficulty, it is possible to derive subsets of variables that provide a satisfactory performance with an acceptable measurement difficulty.

The present paper is organized as follows. Section 2 introduces the background concepts, describes the joint feature and model selection, and presents the SOFC system, its quantitative model, and the faults considered. The generation of the dataset and the achieved results with different feature subsets are described in Section 3. Finally, the conclusions are drawn in Section 4.

2. Materials and Methods

2.1. Model-Based FDI. In model-based fault diagnosis [5, 6], a model of the monitored system (encompassing all the system components) is used to predict the values of several physicochemical variables that characterize the behavior of the nonfaulty system under different operating conditions. The predicted values of these variables are then compared to the real values and measured during plant operation, and the residuals (i.e., indicators of deviations between the measurements and model-based predictions) are used in the FDI procedure through an appropriate diagnosis strategy. When the model is run in parallel with the real system (with the same inputs), the residuals can be computed for each variable through parity equations [6], whereby the values predicted by the model are subtracted from the those measured in the real system.

Because the model and the real system receive the same inputs and because the model simulates a nonfaulty system, when the real system is nonfaulty as well, the residuals are zero or less than the model uncertainty and measurement tolerance. Instead, when a fault occurs, the magnitudes of one or more residuals increase, allowing for fault detection and possible isolation. As mentioned above, the diagnosis strategy can be implemented through an inference or classification approach [6]. In the present study, we propose a classification approach; specifically, we use the classify-before-detect paradigm [21, 22]. According to this paradigm, the detection and classification are not a sequence of distinct tasks but are instead performed jointly, defining the nonfaulty state as a class of the classification scheme. The advantage of this approach is that setting (fixed or adaptive) thresholds for the detection task is no longer necessary.

To train and test an FDI procedure, samples (i.e., sets of residuals) obtained from the monitored system working under several different faulty operating conditions are necessary. To circumvent the problems arising from implanting real faults in real systems (i.e., irreparable damage with related economic loss), we follow the approach described in [23]. Indeed, fictitious faults can be implanted into the same model used to predict the physicochemical variables of the healthy system. In this way, as illustrated in Figure 1, two models are run in parallel: the first in place of the real system, operating under healthy or faulty conditions; the second to predict the variables characterizing a healthy system. The replacement of the real system with a model able to simulate also the faulty conditions is effective only if the model is reliable and accurate. This requirement is typically satisfied when

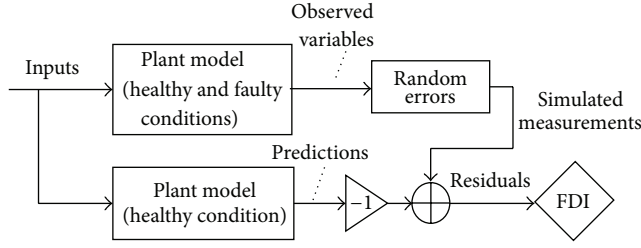


FIGURE 1: Schematic of the simulated deployment of the model-based FDI procedure in industrial plants, according to the approach of parity equations [6].

a quantitative mathematical model (i.e., a model that is based on the physical equations ruling all the processes governing the real system, also called the “first-principle” or “white-box” model [4]) is available and has been validated with experimental trials encompassing several operating conditions. Section 2.5 reports further details about the model of the electric generation SOFC system under consideration in the present paper.

The scheme in Figure 1 can be used to both train and test the FDI procedure [23]. However, the samples used for the SVM classifier training (i.e., the training set, exploited also for the joint feature and model selection) should be different from the samples used to evaluate the FDI performance (i.e., the test set).

Training the classifier requires a substantial amount of data, representative of the possible combinations among operating conditions, fault classes, and fault sizes. Large data collections joined with pattern recognition techniques are typical of data-driven FDI approaches [23–25]. Although model-based and data-driven approaches are traditionally considered distinct, alternative approaches to the FDI problem such as our model-based approach include some aspects of data-driven approaches. Some authors identified this case as the *hybrid* or *integrated approach* [24, 25]. More precisely, in the FDI system we propose, data produced by quantitative experimental validated physically based simulations are used to assemble a statistical knowledge of the relationships between residuals and faults, that is, to train the SVM classifier [23].

2.2. SVM Classification. SVMs are a family of learning machines that are based on methodological contributions from statistical learning, kernel-based processing, functional analysis, and optimization theory [26]. SVM approaches have been proposed for classification [19, 26–28], regression [26, 29], and probability density modeling [30].

Focusing first on binary classification, let ℓ samples, associated with faulty or nonfaulty situations, be generated through the quantitative model. Let d be the number of features (here, the residuals), let \mathbf{x}_i ($i = 1, 2, \dots, \ell$) be the d -dimensional vector collecting the d residuals, and let y_i be a binary variable (named label) that takes on the value $+1$ or value -1 depending on the membership of the i th sample to either one of the two classes. The set $\{(\mathbf{x}_i, y_i)\}_{i=1}^{\ell}$ is the training

set. An SVM classifier assigns an unknown sample $\mathbf{x} \in \mathbb{R}^d$ the class label $\hat{y}(\mathbf{x}) = \text{sgn } f(\mathbf{x})$, where the discriminant function $f(\cdot)$ is the following kernel expansion:

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b, \quad (1)$$

and $K(\cdot, \cdot)$ is a kernel function. The coefficients α_i ($i = 1, 2, \dots, \ell$) are determined by solving the following quadratic programming (QP) problem (multidimensional vectors are considered column vectors, the i th component of $\mathbf{u} \in \mathbb{R}^n$ is denoted as u_i ($i = 1, 2, \dots, n$), and the superscript “ T ” denotes matrix transpose):

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \left(\frac{1}{2} \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} - \mathbf{1}^T \boldsymbol{\alpha} \right), \\ \mathbf{y}^T \boldsymbol{\alpha} = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, \ell, \end{aligned} \quad (2)$$

and the bias b is derived as a by-product of this solution. The matrix Q is the $\ell \times \ell$ matrix whose (i, j) th entry is $Q_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{1}$ is an ℓ -dimensional vector with unitary components, \mathbf{y} is the vector of the labels of the ℓ training samples, and C is a parameter [26]. The expansion (1) is typically sparse; that is, $\alpha_i = 0$ for the majority of training samples; those for which $\alpha_i > 0$ are named support vectors.

A function K of two vectors is a kernel if it is equivalent to the evaluation of an inner-product in some nonlinearly transformed space. Specifically, there are a separable Hilbert space \mathcal{H} and a mapping $\Phi : \mathcal{X} \rightarrow \mathcal{H}$ from a compact subset \mathcal{X} of \mathbb{R}^d to \mathcal{H} , such that $K(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$ for all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ (where $\langle \cdot, \cdot \rangle$ denotes the inner product on \mathcal{H}) [26, 31]. The compactness of \mathcal{X} yields no loss of generality because a compact subset (e.g., a box or a closed ball) that includes all training samples always exists. The so-called Mercer’s conditions are known for a function K to be a kernel. Details can be found in [26]. Here, we only recall that a well-known example is the Gaussian RBF kernel:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (\mathbf{x}, \mathbf{x}' \in \mathcal{X}), \quad (3)$$

where σ is a positive parameter.

It is also possible to prove that (1) is equivalent to a linear discriminant function in the transformed space \mathcal{H} :

$$f(\mathbf{x}) = \langle w, \Phi(\mathbf{x}) \rangle + b, \quad (4)$$

where $w \in \mathcal{H}$ and the bias $b \in \mathbb{R}$ solve the following minimization problem [26]:

$$\min_{w, \xi, b} \left(\frac{1}{2} \langle w, w \rangle + C \cdot \mathbf{1}^T \boldsymbol{\xi} \right), \quad (5)$$

$$y_i (\langle w, \Phi(\mathbf{x}_i) \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, \ell.$$

$\{\xi_i\}_{i=1}^{\ell}$ is a set of slack variables that determine if and how much the discriminant function erroneously classifies the training samples. On one hand, the term $\mathbf{1}^T \boldsymbol{\xi}$ to the objective

function in (5) favors fitting the discriminant function with the available training set. On the other hand, it can be proven that the term $\langle w, w \rangle$ in (5) favors minimizing the expectation (over the probability distribution of the training samples) of the error on an unknown test sample and minimizing overfitting [26]. C tunes the tradeoff between the two terms.

Mapping to the usually higher (possibly infinite) dimensional space \mathcal{H} optimizes the chances that a linear decision boundary can effectively discriminate the classes, while equivalently providing a flexible nonlinear decision boundary in the original space \mathbb{R}^d . At the same time, \mathcal{H} is never computationally involved because all calculations only use the kernel K (see (1) and (2)).

Unlike other popular approaches to nonparametric learning, such as neural networks, the training problem (2) is quadratic and is not plagued with many local minima. Case-specific numerical algorithms have also been proposed to efficiently address it [32].

Generalization to M classes ($M > 2$) is usually achieved by decomposing the multiclass problem into a collection of binary subproblems [19, 26, 28]. Here, the one-against-one (OAO) approach is used, which is usually a good tradeoff between accuracy and computational burden. First, a binary discriminant function $f_{hk}(\cdot)$ is separately determined to discriminate between the h th and k th classes ($h, k = 1, 2, \dots, M; k > h$). Then, to label an unknown sample \mathbf{x} , each function $f_{hk}(\cdot)$ is applied to \mathbf{x} and a vote is cast in favor of either the h th or the k th class depending on the sign of $f_{hk}(\mathbf{x})$. Finally, \mathbf{x} is assigned to the class that received the most votes.

2.3. Feature and Model Selection. A relevant issue in the present FDI scheme is to determine which of the residuals are most informative with respect to the discrimination of the considered classes to minimize both the number of measurements to be taken on the SOFC system and the memory and computational requirements of the FDI procedure. At the same time, the SVM classifier involves parameters (i.e., C in the QP problem (2)) and possible additional parameters in the kernel (e.g., σ in (3)). Their values generally affect the classification performance and need to be set prior to training.

Here, a novel method is developed and embedded into the proposed FDI procedure to select jointly the most informative residuals (i.e., to perform feature selection) and optimize the parameters (i.e., to perform model selection, after fixing the kernel type). The goal is to achieve full automation of the FDI procedure. The key idea is to identify the feature subset and the parameter configuration that minimize an analytical error bound, that is, the so-called span bound. Under mild assumptions, the span bound can be proven to be a tight upper bound on the leave-one-out error rate [18]. It also exhibits a usually high correlation with the error rate on test samples disjointed from the training samples, provided that they are drawn from the same distribution [19, 29]. However, its computation remarkably involves only training samples and no additional validation data. To minimize the span bound, the proposed method combines the approaches introduced in [19, 20] for SVM parameter optimization and feature selection, respectively.

Let \mathcal{R} , \mathcal{S} , and $\boldsymbol{\theta}$ be the set of all d features (i.e., the residuals), a subset of m features ($\mathcal{S} \subset \mathcal{R}$), and a vector collecting the input SVM parameters (e.g., C and σ , if the Gaussian RBF kernel is used), respectively. Considering again the binary case first, let us explicitly stress the dependence on \mathcal{S} and $\boldsymbol{\theta}$ by denoting as $\alpha_i(\mathcal{S}, \boldsymbol{\theta})$ ($i = 1, 2, \dots, \ell$) and $f(\cdot | \mathcal{S}, \boldsymbol{\theta})$ the solution of the QP problem in (2) and the discriminant function in (1), respectively, when the SVM is trained using the m features in \mathcal{S} and the input parameter vector $\boldsymbol{\theta}$. The span bound is defined as the fraction $\mathcal{F}(\mathcal{S}, \boldsymbol{\theta})$ of the training samples such that $\alpha_i(\mathcal{S}, \boldsymbol{\theta}) > 0$ and ($i = 1, 2, \dots, \ell$):

$$\alpha_i(\mathcal{S}, \boldsymbol{\theta}) S_i^2(\mathcal{S}, \boldsymbol{\theta}) \geq y_i f(\mathbf{x}_i | \mathcal{S}, \boldsymbol{\theta}), \quad (6)$$

where the coefficient $S_i^2(\mathcal{S}, \boldsymbol{\theta})$ (named span) can be obtained as a by-product of the training phase by solving a further QP problem [18] or through a fast linear algebra argument [33]. In the multiclass case, $\mathcal{F}(\mathcal{S}, \boldsymbol{\theta})$ is computed through OAO as a weighted average of the span bound values obtained separately for each pair of distinct classes, the weights being proportional to the relative frequencies of the classes in the training set [19]. Further details on this point can be found in [19, 33].

Indeed, $\mathcal{F}(\mathcal{S}, \boldsymbol{\theta})$ is a nondifferentiable function of $\boldsymbol{\theta}$ [18]. According to the definition recalled above for the binary case, $\ell \cdot \mathcal{F}(\mathcal{S}, \boldsymbol{\theta})$ is integer-valued and $\mathcal{F}(\mathcal{S}, \boldsymbol{\theta})$ proves to be piecewise constant on the space of the admissible parameter vectors $\boldsymbol{\theta}$. Similar comments hold in the multiclass case as well. This prevents applying numerical minimization algorithms that make use of derivatives (e.g., the gradient or the Newton-Raphson's methods) [34]. In general, suitable numerical gradients and difference quotients might be used to replace gradients and derivatives, but *ad hoc* convergence theorems would be necessary for their specific application to the span bound. In [34], a regularized differentiable version of the span bound is introduced to allow gradient descent to be applied, but an additional regularization parameter, which has to be manually tuned, is necessary. Here, similar to [19], the Powell's algorithm is used to minimize $\mathcal{F}(\mathcal{S}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Powell's method is an unconstrained minimization technique that emulates the behavior of the conjugate gradient method without using derivatives and converges, under mild assumptions, to a local minimum [35].

To minimize the resulting functional,

$$\mathcal{F}_*(\mathcal{S}) = \min_{\boldsymbol{\theta}} \mathcal{F}(\mathcal{S}, \boldsymbol{\theta}), \quad (7)$$

with respect to \mathcal{S} , the steepest ascent algorithm in [20] is adapted and extended. It is an iterative algorithm, initialized with a preliminary subset of m features, which has been demonstrated to be effective when applied to the maximization of Bayesian interclass distance measures in problems of remote sensing image classification [20, 36]. Here, it is extended to the minimization (steepest descent) of the span bound functional, combined with the Powell's algorithm, and integrated in the proposed FDI procedure.

Specifically, given a subset \mathcal{S} of m features, the proposed feature selection and parameter optimization method evaluates each possible replacement of one of the m features in

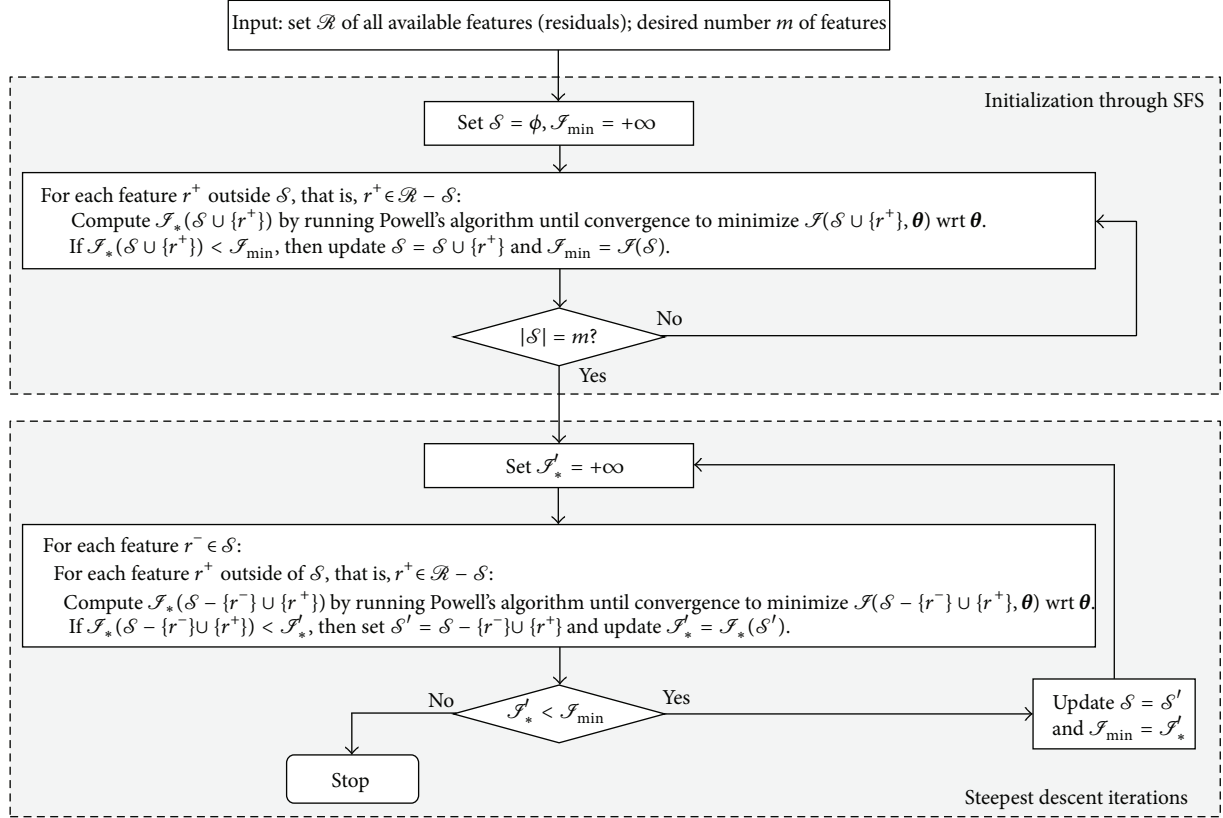


FIGURE 2: Flowchart of the proposed joint feature and model selection algorithm for FDI in SOFC systems. \mathcal{R} , \mathcal{S} , and m denote the set of all available features (residuals), a feature subset, and the desired number of features to be selected, respectively. $|\mathcal{S}|$ denotes the cardinality of \mathcal{S} .

\mathcal{S} by one of the $(d - m)$ features outside \mathcal{S} (i.e., in $\mathcal{R} - \mathcal{S}$) by computing the corresponding value of $\mathcal{J}_*(\cdot)$, that is, by running Powell's algorithm until convergence (see (7)). Let \mathcal{J}'_* be the minimum span bound obtained across all these $m(d - m)$ possible replacements. If $\mathcal{J}'_* < \mathcal{J}_*(\mathcal{S})$, the replacement is performed and \mathcal{S} is correspondingly updated. This procedure is iteratively repeated, while reductions in the span bound are feasible through some replacement of a feature inside by a feature outside the current subset.

The collection of the subsets of m features is finite, so finite-time termination is guaranteed. As discussed in [20], convergence to a local minimum of $\mathcal{J}_*(\cdot)$ is also guaranteed, whereby the notion of the local minimum is interpreted by endowing the discrete space of the subsets of \mathcal{R} with a metric-space topology through the well-known Hamming distance. This property, together with the aforementioned convergence behavior of the Powell's algorithm, suggests that, at the least, local minima of the span bound functional are identified by the proposed method in the searches for both a feature subset and a parameter vector.

Initialization of the method is performed through the sequential forward selection (SFS) algorithm, that is, a well-known suboptimal approach to feature selection [17]. First, SFS starts from an empty subset of features, separately computes the values of $\mathcal{J}_*(\cdot)$ associated with the d subsets composed of one feature each, and selects the feature corresponding to the smallest value of $\mathcal{J}_*(\cdot)$. Then, it evaluates

$\mathcal{J}_*(\cdot)$ for all the $(d - 1)$ subsets of two features, which are obtained by separately pairing the previously selected feature with each other feature. Again, the resulting feature pair with the smallest value of $\mathcal{J}_*(\cdot)$ is selected. Then, the procedure is repeated iteratively, progressively adding one feature at a time until the desired number m of features is reached. Figure 2 displays a flowchart of the proposed feature selection and parameter optimization algorithm. Further details on the SFS and steepest descent (ascent) algorithms can be found in [20, 36].

2.4. SOFC System Description. In this study, the model-based FDI scheme and SVM classification are applied to an electric generation system consisting basically of a reformer, an SOFC stack (formed by a number of planar cells superimposed to each other), and a post burner. The reformer contains a suitable catalyst that promotes the steam reforming reaction within the feeding mixture (3:1 vol. methane and steam), which is partially converted into hydrogen and carbon monoxide. This partially reformed fuel is then fed into the anode compartment of the SOFC stack, while air is fed into the cathode side. The SOFC stack (average operating temperature of 850°C) is composed of a number of rectangular planar cells superimposed onto each other, each of which develops an electrochemical reaction producing steam and carbon dioxide along with electrical power and heat. Nevertheless, the anode exhaust contains a significant

percentage of flammables. Thus, the anode exhaust is mixed with the cathode exhaust and burned in an off-gas burner to reduce the release of pollutants and increase the temperature of the flue gas for further utilization or energy recovery in subsequent components (not considered here). The system scheme is displayed in Figure 3; further details can be found in [37].

2.5. Quantitative Model and Fault Classes. The model of the electric generation SOFC system is a quantitative model, which embeds the physical equations of the phenomena occurring in the process, earning the classification of “first-principle” or “white-box” model. Overall, the model is obtained by coupling the quantitative models of the three components, that is, reformer, fuel cell stack, and burner. The SOFC stack model assumes all cells are identical along the stack. The single cell model is then developed according to a typical scheme applied to chemical reactors [38] and includes the equations of the local chemical and electrochemical reaction kinetics. In turn, the latter includes the evaluation of the Nernst voltage and of all the sources of losses (anodic activation, cathodic activation, and ohmic). The local kinetics is then coupled to local mass, energy, and momentum balances. A partial differential/algebraic system of equations is obtained, integrated using a relaxation method for the energy balance, and combined with a finite difference method to solve the other equations. Reformer and burner are simulated through macroscopic mass and energy balances. In the previous case, the steam reforming reaction is assumed to be at thermodynamic equilibrium. In the latter case, all the flammables are considered to be combusted completely. Further details can be found in [37].

In principle, the model is developed to simulate nonfaulty operating conditions, but it can also be extended to simulate faulty operating conditions by including suitable equations for the simulation of typical faults occurring in SOFC systems [37]. Thus, the following four main fault classes have been simulated [37].

SOFC Stack Degradation. A number of different faults can occur inside the SOFC stack, affecting the cell structure and, in particular, the electrolyte/electrodes coupling in different ways. A comprehensive overview is given in [37], in which it is also demonstrated that, for the purposes of plant simulation under faulty conditions, the effect of all these different faults is correctly simulated with an increase of the overall internal stack losses. Thus, the latter have been increased between 105% and 160% of their nominal value.

Air Leakage. In addition to providing the oxidant necessary for the electrochemical reaction, the cathode air flow allows the temperature of the stack to be controlled. A potential air leak between the air flow meter and the SOFC stack has been simulated by reducing the flow entering the stack to between 50% and 95% of its nominal rate.

Fuel Leakage. Inside the reformer, the methane/steam feed is converted into a mixture with a high percentage of hydrogen. Fuel leakage is likely to occur because of the low dimension

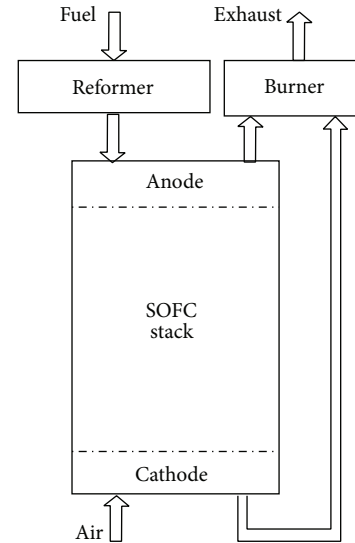


FIGURE 3: Schematic of the SOFC system.

and high diffusivity of the hydrogen molecule. A leak between the exit of the reformer and the entrance of the stack has been simulated by reducing the flow rate to between 75% and 95% of its nominal value.

Reformer Degradation. A reduced conversion of methane can be achieved in the reformer due to a number of faulty effects (e.g., catalyst degradation, carbon deposition, or sulfur poisoning). Several possible degrees of deviation from equilibrium have been simulated by reducing the equilibrium constant of the steam reforming reaction to between 30% and 95% of its thermodynamic value.

As indicated in Figure 3, the SOFC plant under consideration includes an SOFC stack, a methane steam reformer, a burner, a fuel feeding system, and an air feeding system. Given that the burner is a well-tested and mature technology, its faults are not taken into account. On the other hand, the faults of the other four plant components are all considered, one fault for each component. As mentioned above, the different faults that can occur in the SOFC stack end up giving similar effects [37]. Thus, they can be lumped together as one single fault class. Similar considerations hold for the steam reforming reactor. Consequently, the FDI procedure proposed here cannot be used to distinguish among the different types of faults that can occur in the SOFC stack or in the methane steam reformer. If necessary, further investigations (e.g., chemical or electrochemical tests) must be carried out to identify in detail the microscopic cause of the failure, as suggested in [37].

The reliability of the predictions obtained by the SOFC system model is assured by its validation against experimental data. This validation was performed under both steady-state and dynamic working regimes. Moreover, as described in [37], for these working regimes, different operating conditions have been considered. Circumscribed experimental validation under faulty operating conditions has been performed as well [39].

Even if our model of the SOFC system is able to simulate faults occurring during transient system operation, as discussed in [39], the development of an FDI procedure for application in transient conditions requires an analysis of the time-dependent behavior of residuals that is beyond the scope of the present paper. The latter is devoted to an investigation of faults occurring during steady-state operation of the system. Starting with the system working in an unfaulty steady-state operating condition, we simulate the occurrence of a fault, which triggers a transient behavior that is simulated by our model until a new (faulty) steady-state condition is reached. Residuals that we consider are those calculated as the difference between values of the observed physicochemical variables in the faulty and unfaulty steady states.

The experimental SOFC plant that provided the experimental data is identical to the plant described above (see Figure 3). The plant has been manufactured by Staxera GmbH (D) [40] and tested by EBZ GmbH (D) within the European project GENIUS [41].

2.6. Computation of Residuals. As discussed previously and according to [23], in this study, the real plant is replaced by a copy of the quantitative model, modified in view of simulating the effect of different faults of various size occurring in the plant (see Figure 1). Whereas ideal conditions were assumed in [9], here, the model uncertainty and measurement tolerance are considered by adding random errors to the values of the physicochemical variables simulated for the real plant (see Figure 1). This allows us to attain realistic residuals and to investigate the sensitivity of the FDI procedure by testing different error magnitudes.

To introduce the model uncertainty and measurement tolerance, each observed variable calculated by the model used in place of the real plant is multiplied by a random variable γ_ε , uniformly distributed in $[1 - \varepsilon, 1 + \varepsilon]$, where 100ε represents the maximum percentage error. Therefore, the residual δ_V related to a given monitored variable V is computed as follows:

$$\delta_V = V_m \gamma_\varepsilon - V_p, \quad (8)$$

where V_m is the observed variable calculated by the (simulated) plant under faulty and nonfaulty conditions and V_p is the same variable predicted by the model for a nonfaulty plant. The value assumed by the random variable γ_ε is updated each time the residual δ_V is computed.

The residuals of the ten physicochemical variables of the SOFC system, listed in Table 1, are used as features for the classification algorithm. Although these ten variables are easy to access in numerical simulations, their measurement in a real plant raises some complexity concerns. Thus, we have classified them as easy to extremely difficult to measure, as reported in Table 1. The feature selection method proposed here makes it possible to assess the contribution provided by each of these variables toward a correct classification. This information, combined with the measurement difficulty, allows us to select the variables to be monitored in a real application of the FDI procedure.

TABLE 1: Residuals used for the FDI procedure with the related measurement difficulty.

Number	Residual	Measurement difficulty
1	Generated electric power	Easy
2	Air flow rate entering the stack	Easy
3	Fuel flow rate entering the stack	Easy
4	Air pressure loss between the inlet and outlet of the FC stack	Easy
5	Temperature at the burner outlet	Easy
6	Average stack temperature	Difficult
7	Anodic activation losses	Extremely difficult
8	Cathodic activation losses	Extremely difficult
9	Ohmic losses	Extremely difficult
10	Nernst voltage	Extremely difficult

2.7. Control Strategies and Operating Conditions. In this study, both typical control strategies employed in FC generation plants (i.e., constant-current and constant-voltage) have been considered. The generation plant is assumed to be equipped with a power inverter that, in addition to the task of converting the DC electrical power to AC electrical power, controls the power supplied by the plant and the operation mode of the stack. Thus, the inverter can force the stack to generate a constant voltage or constant current according to the V - I characteristic curves of the SOFC stack.

The reference operating condition for the SOFC plant described above includes a constant electrical current of 26.2 A (or a constant SOFC stack potential of 42.5 V), a fuel utilization factor of 0.75, a reformer temperature of 650°C, and an average SOFC stack temperature of 850°C. In this condition, the plant generates an electrical power of approximately 1.1 kW. Starting from this reference operating condition, for each control strategy, different operating conditions are defined by tuning the values imposed for the constant current (or constant voltage) and the fuel utilization factor. During operation, the reformer is kept at fixed temperature (650°C) through an electrical heater, and the SOFC stack temperature is maintained at the desired level by regulating the air inlet flow rate. For the constant-current control strategy, ten operating conditions are considered by modifying the electrical current (ranging from 6 to 30 A) and/or fuel utilization factor (ranging from 0.35 to 0.75) with respect to their reference values. For the constant-voltage control strategy, ten operating conditions are considered as well. The stack potential is modified between 41 and 53 V, and the fuel utilization factor is modified between 0.35 and 0.75.

Further details regarding the control strategy and operating conditions of the SOFC generation plant under consideration are given in [37].

3. Results and Discussion

3.1. Model Uncertainty. As mentioned in Section 2.5, the SOFC system model has been validated experimentally. The maximum difference between the measured and predicted values of the monitored variables is approximately 3% [37]. On the basis of this value, the maximum magnitude for the random errors to be introduced when the measurement of the ten observed variables is simulated (see Figure 1) can be set. Here, the following three values are considered for the maximum percentage error: 2%, 4%, and 6%.

3.2. Dataset Composition. Due to a generalized lack of experimental faulty data in SOFC systems [41], which in turn is strictly related to their high cost, the simulated approach depicted in Figure 1 is used to assemble the dataset. Although the available experimental data are not sufficient to train and test an SVM classifier, they are adequate for validating a physics-based quantitative modeling tool, able to simulate unfaulty [37] and faulty [39] operating conditions, which can then produce the dataset pool necessary for training and validating the SVM classifier.

For each operating condition, ten different sizes of a given fault (inside the range defined in Section 2.5) are considered. Thus, for a given control strategy, 100 combinations of steady-state operating conditions and fault sizes are identified for each fault. After setting the maximum error magnitude, independent random errors are introduced for each monitored variable of each combination. As a result, for each control strategy, a dataset of approximately 500 feature vectors is available (5 classes are indeed considered: 4 faulty classes and the nonfaulty class). The hundred vectors for the nonfaulty class are generated by repeating the realization of the random errors for each operating condition ten times.

Subsequent realizations of the independent random errors enable the generation of an arbitrary number of datasets. Thus, we have produced a pool of datasets for each control strategy (i.e., constant-voltage and constant-current) and for each maximum percentage error (i.e., 2%, 4%, and 6%). Each dataset is composed of approximately 500 feature vectors, and each feature vector is composed of ten features (see Table 1). In each resulting dataset, the sets of training samples of the various classes have approximately the same size. A fuel cell will for the most part operate in a nonfaulty state, so the nonfaulty class is expected to have a larger prior probability than the individual faulty classes, a common situation in a detection problem. This may suggest using a larger training set size for “no-fault” than for the other classes. However, SVM-based classifiers are known to be generally sensitive to significantly unbalanced numbers of training samples per class. Therefore, the aforementioned approach was used to take benefit from the opportunity to actively generate the training set and make sure that balanced sets were constructed for the various classes. Indeed, if a different strategy was used and more training samples were available for the nonfaulty class than for the other classes, sample pre-selection algorithms such as [42] could be applied to make sure that strongly unbalanced classes are avoided.

All the features of a given dataset have been preliminarily normalized to ensure that each has a zero mean and unitary variance. This normalization is necessary because of the significantly different orders of magnitude of the measured residuals. It also helps preventing overflow and favors numerical stability in the solution of the QP problem for SVM training.

3.3. Fault Detection and Isolation. In real systems, the design of a classifier that, working with a given residual subset, provides a satisfactory performance when the error magnitude is not perfectly known and potentially varies over time has great relevance. To this end, a training set has been composed for each control strategy by joining three datasets together: one dataset for each maximum percentage error (i.e., 2%, 4%, and 6%). Here, this training set is used to train SVM classifiers in which the Gaussian RBF kernel in (3) is adopted. The joint feature and model selection method described in Section 2.3 provides the optimal feature subset, when the number of residuals m is increased from 1 to 10.

For each control strategy and optimal feature subset, the SVM classifier trained using the aforementioned mixture of error magnitudes is tested with several datasets (excluding those already used to compose the training set), characterized by given error magnitudes. The performance of the classifier is evaluated through the overall accuracy (OA), that is, the fraction of the correctly classified test samples (an estimate of the probability of correct classification). An average OA (referred to as OA_{avg}) is introduced by averaging the three OA values obtained with regard to the 2%, 4%, and 6% error magnitudes. Finally, to evaluate the classifier performance for each specific class, the average producer accuracy (PA_{avg}) is introduced; that is, the fraction of the test samples belonging to a given class that are correctly classified, averaged over the three error magnitudes (2%, 4%, and 6%).

The OA_{avg} value of the classifiers, trained with the mixture of error magnitudes, with the number of adopted features increasing from 1 to 10, is reported in Table 2 for the constant-voltage case and in Table 3 for the constant-current case. The subset of residuals obtained by the feature selection technique and the value of OA for each error magnitude are also listed. For the sake of brevity, the corresponding parameter configurations obtained by the proposed joint model and feature selection algorithm are not listed. However, as detailed in [19, 29], the parameter vector obtained by minimizing the span bound through Powell’s algorithm typically yield classification accuracies very similar to those obtained by time-expensive grid searches for the minimum cross-validation error rate over a predefined grid in the parameter space. Details on this aspect can be found in [19, 29] with regard to SVM for classification and regression, respectively.

Except for $m = 1$, the classification performance decreases as the error magnitude increases, as expected, and the OA values obtained for the constant-voltage control are higher than the related values obtained for the constant-current control.

The performance behavior as a function of the number of features m is more complex. For both control strategies,

TABLE 2: Residuals and OAs, with the number of features m increasing from 1 to 10, for the constant-voltage case. Training set including a mixture of error magnitudes.

m	Residual subset	Overall accuracy (OA)			
		Test-set max. error			OA_{avg}
		2%	4%	6%	
1	9	0.41	0.42	0.42	0.42
2	1, 2	0.90	0.88	0.82	0.87
3	2, 3, 8	0.95	0.89	0.80	0.88
4	2, 3, 4, 8	0.97	0.92	0.84	0.91
5	1, 2, 3, 4, 8	0.97	0.90	0.83	0.90
6	1, 2, 3, 4, 8, 9	0.98	0.89	0.80	0.89
7	1, 2, 3, 4, 7, 8, 9	0.98	0.88	0.79	0.89
8	1, 2, 3, 4, 5, 7, 8, 9	0.98	0.92	0.82	0.91
9	1, 2, 3, 4, 5, 6, 7, 8, 9	0.98	0.92	0.86	0.92
10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	0.98	0.92	0.86	0.92

TABLE 3: Residuals and OAs, with the number of features m increasing from 1 to 10, for the constant-current case. Training including a mixture of error magnitudes.

m	Residual subset	Overall accuracy (OA)			
		Test-set max. error			OA_{avg}
		2%	4%	6%	
1	2	0.61	0.59	0.58	0.59
2	2, 3	0.76	0.76	0.72	0.75
3	2, 3, 4	0.83	0.83	0.79	0.82
4	2, 3, 4, 7	0.90	0.84	0.75	0.83
5	1, 2, 3, 4, 7	0.90	0.81	0.72	0.81
6	1, 2, 3, 4, 7, 8	0.90	0.83	0.72	0.82
7	1, 2, 3, 4, 5, 7, 8	0.92	0.81	0.73	0.82
8	1, 2, 3, 4, 5, 7, 8, 9	0.92	0.80	0.73	0.82
9	1, 2, 3, 4, 5, 6, 7, 8, 9	0.69	0.68	0.63	0.67
10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	0.76	0.75	0.72	0.74

the value of OA_{avg} monotonically increases moving from $m = 1$ to $m = 4$. In the constant-current case, for $4 \leq m \leq 8$, the OA_{avg} value is nearly constant, exhibiting small oscillations that do not exceed 0.02. For $m > 8$, the OA_{avg} value sharply decreases. In the constant-voltage case, for $m \geq 4$, the OA_{avg} value is nearly constant, exhibiting small oscillations that do not exceed 0.02. Concerning the residuals that are selected when the value of m is increased, we can observe that, in the constant-current case, the residuals selected for a given m are those selected for $m - 1$ plus one new residual. This behavior is not forced by the feature and model selection method presented in Section 2.3. In fact, in the constant-voltage case, this behavior is observed only for $m > 3$, where the performance approaches its maximum.

Because the use of a number of residuals larger than four does not increase or increases negligibly the performance, in the following sections of this paper, we focus on the performance achieved for $m = 4$ and the related residuals.

TABLE 4: OA when two subsets of easy-to-measure residuals are adopted. The training set included a mixture of error magnitudes.

Control str.	Residual subset	Overall accuracy (OA)			
		Test-set max. error			OA_{avg}
		2%	4%	6%	
Constant-voltage	2, 3, 4	0.87	0.78	0.70	0.78
	1, 2, 3, 4	0.97	0.89	0.82	0.89
Constant-current	2, 3, 4	0.83	0.83	0.79	0.82
	1, 2, 3, 4	0.90	0.82	0.74	0.82

In the constant-voltage case, OA_{avg} is equal to 0.91 and OA decreases by 0.13 when increasing the error magnitude from 2% to 6%. The selected residuals are those numbered with 2, 3, 4, and 8 in Table 1. In the constant-current case, OA_{avg} is equal to 0.83, and the OA decreases by 0.15 when increasing the error magnitude from 2% to 6%. The selected residuals are those numbered with 2, 3, 4, and 7. We note that residuals 2, 3, and 4 are selected for both control strategies. These residuals are considered easy to measure in Table 1 (as opposed to residuals 7 and 8, which are extremely difficult to measure). Moreover, when m increases from 4 to 5, the feature that is selected is residual 1, which is easy to measure. We can conclude that among the easy-to-measure residuals, those numbered with 2, 3, and 4 are the most informative for both control strategies. In addition, the residual numbered with 1 is potentially useful.

3.4. Performance of Easy-to-Measure Residuals. The results obtained by the joint feature and model selection can be used to design a classifier that working with a limited number of easy-to-measure residuals provides performance close to the optimal one. In other words, we attempt to exclude the residuals that are not easy to measure, limiting the performance decrease.

On the basis of the considerations drawn in Section 3.3, we analyze the performance of the SVM classifiers that use the two following subsets of residuals, (2, 3, 4) and (1, 2, 3, 4), and are trained by the aforementioned mixture of error magnitudes (see Table 4). While the residual subset (2, 3, 4) is the optimum choice for the constant-current case with $m = 3$ (see Table 3), all the three remaining combinations reported in Table 4 are surely suboptimal. To maximize the performance, for the four possible combinations between control strategy and residual subset, the model parameters of the SVM classifiers have been optimized by minimizing the span bound (computed for a given subset of features) through the Powell's algorithm, according to the description in Section 2.3.

Table 4 displays the performance of the four classifiers when tested with samples with a given error magnitude as well as the average figure OA_{avg} . Comparing Tables 2 and 4, we observe that for the constant-voltage case, the performance obtained with the residual subset (2, 3, 4) is significantly lower than that obtained with the optimum subset (2, 3, 8). Instead, when four features are adopted, the use of the easy-to-measure subset (1, 2, 3, 4) provides an OA_{avg}

TABLE 5: PA_{avg} when two subsets of easy-to-measure residuals are adopted. The training set included a mixture of error magnitudes. Class number meaning: 0: nonfaulty system; 1: SOFC stack degradation; 2: air leakage; 3: fuel leakage; 4: reformer degradation.

Control str.	Residual subset	Average producer accuracy (PA_{avg})				
		Class number				
		0	1	2	3	4
Constant-voltage	2, 3, 4	0.87	0.68	0.88	0.88	0.49
	1, 2, 3, 4	0.89	0.89	0.89	0.92	0.85
Constant-current	2, 3, 4	0.97	0.45	0.83	0.89	0.97
	1, 2, 3, 4	0.85	0.57	0.90	0.86	0.99

that is only marginally lower than that obtained with the optimum subset for $m = 4$ (i.e., 0.89 against 0.91). In addition, in the constant-current case, the use of the easy-to-measure subset (1, 2, 3, 4) provides an OA_{avg} that is only marginally lower than that obtained with the optimum subset for $m = 4$ (i.e., 0.82 against 0.83). However, this performance is equal to that obtained by using the optimum subset (2, 3, 4) composed of only three features. In the constant-current case, the choice between three and four features can be driven by the error magnitude that is expected in real applications. The subset (2, 3, 4) provides a higher OA when the maximum error tends to 6%; the subset (1, 2, 3, 4) is preferable when the maximum error tends to 2%. On the contrary, in the constant-voltage case, the results reported in Table 4 indicate that the subset (1, 2, 3, 4) is always preferable.

While OA_{avg} encompasses the results related to the five considered classes, PA_{avg} provides a performance evaluation for each specific class. Table 5 displays the PA_{avg} values for the four classifiers that work with a limited number of easy-to-measure residuals, that is, the classifiers considered in Table 4. We observe that in the constant-voltage case, the addition of the feature number 1 improves all the PA_{avg} values, especially those related to classes numbers 1 and 4. With the feature subset (1, 2, 3, 4), the nonfaulty status and all the considered faults are identified with an average accuracy higher than 85%. In the constant-current case, the addition of feature number 1 improves the PA_{avg} values for classes 1, 2, and 4 and worsens the PA_{avg} values for classes 0 and 3. Although the OA_{avg} does not change (see Table 4), better performance uniformity is obtained. All the considered faults except for SOFC stack degradation are identified with an average accuracy higher than 85%. PA_{avg} for SOFC stack degradation does not exceed 57% when the constant-current control is adopted, demonstrating that in this specific case the identification of this fault through four easy-to-measure residuals is critical. We also verified that the addition of the last easy-to-measure residual (i.e., feature number 5) to the subset (1, 2, 3, 4) does not improve this performance. A further investigation of this specific misclassification shows that the feature vectors for the SOFC stack degradation fault are sometimes assigned to the fuel leakage fault.

More generally, the addition of residual 5 to the subset (1, 2, 3, 4) does not provide any improvement in the obtained performance. In the constant-voltage case, the addition of

residual 5 slightly decreases the OA values, especially when the maximum error is equal to 6%.

Finally, it is worth noting that the PAs for the nonfaulty class (i.e., the class number 0) were generally high; that is, a few false alarms were generated by the proposed system, although the nonfaulty class is expected to have a larger prior probability than individual faulty classes. This confirms the discrimination capability of the SVM approach and suggests that the generation of balanced sets of training samples per class did not erroneously favor the faulty classes.

4. Conclusions

In the context of SVM classification, we propose an original technique for the joint feature and model selection. This technique is applied to an open research problem: the design of a model-based FDI procedure, integrated with a data-driven diagnosis, for an SOFC electric generation plant. The FDI procedure is demonstrated to perform properly in a wide range of steady-state operating conditions and fault sizes and for various statistics of the random errors affecting the model predictions.

The joint feature and model selection make it possible to evaluate the relative importance of the residuals used as features for the robust FDI procedure, assuring the optimum performance for every cardinality of the feature vector. Moreover, for a given residual subset, if the maximum prediction error is increased by a factor of 3 (from 2% to 6%), the related OA decreases by a factor that does not exceed 1.26. Thus, because the performance reduction is very gradual despite the wide range of operating conditions and fault sizes that are considered, the robustness of the FDI procedure against modeling and measurement errors is proven.

Joining the results achieved by the feature selection to the evaluation of the difficulty in measuring the different residuals, we conclude that the use of the easy-to-measure residuals numbered 1, 2, 3, and 4 altogether are sufficient to achieve an overall performance that is very close to the absolute maximum observed when the use of all the residuals is enabled. This holds for both constant-voltage and constant-current control strategies. The only exception is the identification of the SOFC stack degradation under constant-current control. The accuracy for this fault is significantly lower than those observed for other faults. Further electrochemical investigations will be devoted to individuate new residuals able to provide a better discrimination of the SOFC stack degradation fault.

When a large number of combinations between the operating conditions and fault sizes are considered (for both training and testing), the SVM classifier using the four aforementioned residuals provides an OA, averaged over three values for the maximum prediction error (i.e., 2%, 4%, and 6%) that is higher than 80% for the constant-current case and is close to 90% for the constant-voltage case.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been supported jointly by the Departments DICCA, DIME, and DITEN of the University of Genoa, Italy.

References

- [1] S. Srinivasan, *Fuel Cells: From Fundamentals to Applications*, Springer, New York, NY, USA, 2006.
- [2] N. Behling, *Fuel Cells: Current Technology Challenges and Future Research Needs*, Elsevier, Oxford, UK, 2012.
- [3] D. Marra, M. Sorrentino, C. Pianese, and B. Iwanschitz, "A neural network estimator of Solid Oxide Fuel Cell performance for on-field diagnostics and prognostics applications," *Journal of Power Sources*, vol. 241, pp. 320–329, 2013.
- [4] R. Petrone, Z. Zheng, D. Hissel et al., "A review on model-based diagnosis methodologies for PEMFCs," *International Journal of Hydrogen Energy*, vol. 38, no. 17, pp. 7077–7091, 2013.
- [5] R. Isermann and P. Ballé, "Trends in the application of model-based fault detection and diagnosis of technical processes," *Control Engineering Practice*, vol. 5, no. 5, pp. 709–719, 1997.
- [6] R. Isermann, "Model-based fault-detection and diagnosis—status and applications," *Annual Reviews in Control*, vol. 29, no. 1, pp. 71–85, 2005.
- [7] I. Arsie, A. Di Filippi, D. Marra, C. Pianese, and M. Sorrentino, "Fault tree analysis aimed to design and implement on-field fault detection and isolation schemes for SOFC systems," in *Proceedings of the ASME 8th International Conference on Fuel Cell Science, Engineering and Technology (FUELCELL '10)*, pp. 389–399, New York, NY, USA, June 2010.
- [8] N. Y. Steiner, D. Hissel, P. Moçotéguy et al., "Application of fault tree analysis to fuel cell diagnosis," *Fuel Cells*, vol. 12, no. 2, pp. 302–309, 2012.
- [9] L. Pellaco, P. Costamagna, A. de Giorgi et al., "Fault diagnosis in fuel cell systems using quantitative models and support vector machines," *Electronics Letters*, vol. 50, no. 11, pp. 824–826, 2014.
- [10] A. Sorce, A. Greco, and L. Magistri, "Solid oxide fuel cell systems fault detection and identification through hybrid model and explicit rules," in *Proceedings of the 5th International Conference on Fundamentals & Development of Fuel Cells*, pp. 1–9, Karlsruhe, Germany, 2013.
- [11] T. Escobet, D. Feroldi, S. de Lira et al., "Model-based fault diagnosis in PEM fuel cell systems," *Journal of Power Sources*, vol. 192, no. 1, pp. 216–223, 2009.
- [12] A. Widodo and B.-S. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Systems with Applications*, vol. 33, no. 1, pp. 241–250, 2007.
- [13] J.-D. Son, G. Niu, B.-S. Yang, D.-H. Hwang, and D.-S. Kang, "Development of smart sensors system for machine fault diagnosis," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11981–11991, 2009.
- [14] O. Ondel, E. Boutleux, E. Blanco, and G. Clerc, "Coupling pattern recognition with state estimation using Kalman filter for fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4293–4300, 2012.
- [15] J. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 5, pp. 2363–2376, 2012.
- [16] X. Zhang, W. Li, J. Xi, Z. Zhang, and X. Fan, "Surface defect target identification on copper strip based on adaptive genetic algorithm and feature saliency," *Mathematical Problems in Engineering*, vol. 2013, Article ID 504895, 10 pages, 2013.
- [17] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Transactions on Computers*, vol. 20, pp. 1100–1103, 1971.
- [18] V. Vapnik and O. Chapelle, "Bounds on error expectation for support vector machines," *Neural Computation*, vol. 12, no. 9, pp. 2013–2036, 2000.
- [19] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2734–2752, 2013.
- [20] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1360–1367, 2001.
- [21] F. B. Shin and D. H. Kil, "Full-spectrum signal processing using a classify-before-detect paradigm," *The Journal of the Acoustical Society of America*, vol. 99, no. 4, pp. 2188–2197, 1996.
- [22] A. Trucco, "Detection of objects buried in the seafloor by a pattern-recognition approach," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 769–782, 2001.
- [23] L. H. Chiang, E. L. Russell, and R. D. Braaz, *Fault Detection and Diagnosis in Industrial Systems*, Springer, New York, NY, USA, 2001.
- [24] E. Sobhani-Tehrani and K. Khorasani, *Fault Diagnosis of Non-linear Systems Using a Hybrid Approach*, vol. 383 of *Lecture Notes in Control and Information Sciences*, Springer, New York, NY, USA, 2009.
- [25] J. Luo, M. Namburu, K. R. Pattipati, L. Qiao, and S. Chigusa, "Integrated model-based and data-driven diagnosis of automotive antilock braking systems," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 40, no. 2, pp. 321–336, 2010.
- [26] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, NY, USA, 1998.
- [27] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [28] J. Cui, M. Shan, R. Yan, and Y. Wu, "Aero-engine fault diagnosis using improved local discriminant bases and support vector machine," *Mathematical Problems in Engineering*, vol. 2014, Article ID 283728, 9 pages, 2014.
- [29] G. Moser and S. B. Serpico, "Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 909–921, 2009.
- [30] P. Mantero, G. Moser, and S. B. Serpico, "Partially supervised classification of remote sensing images through SVM-based probability density estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 559–570, 2005.
- [31] L. Debnath and P. Mikusinski, *Introduction to Hilbert Spaces*, Academic Press, New York, NY, USA, 1990.
- [32] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *Journal of Machine Learning Research*, vol. 6, pp. 1889–1918, 2005.

- [33] J. Guo, N. Takahashi, and T. Nishi, "An efficient method for simplifying decision functions of support vector machines," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 10, pp. 2795–2802, 2006.
- [34] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 131–159, 2002.
- [35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++*, Cambridge University Press, New York, NY, USA, 2nd edition, 2002.
- [36] S. B. Serpico, M. D'Inca, F. Melgani, and G. Moser, "Comparison of feature reduction techniques for classification of hyperspectral remote sensing data," in *8th Image and Signal Processing for Remote Sensing*, vol. 4885 of *Proceedings of SPIE*, pp. 347–358, Crete, Greece, September 2002.
- [37] A. Sorce, A. Greco, L. Magistri, and P. Costamagna, "FDI oriented modeling of an experimental SOFC system, model validation and simulation of faulty states," *Applied Energy*, vol. 136, 908, no. 894, 2014.
- [38] H. S. Fogler, *Elements of Chemical Reaction Engineering*, Prentice Hall, New York, NY, USA, 4th edition, 2006.
- [39] A. Greco, A. Sorce, R. Littwin, P. Costamagna, and L. Magistri, "Reformer faults in SOFC systems: experimental and modelling analysis, and simulated fault maps," *International Journal of Hydrogen Energy*, vol. 136, pp. 894–908, 2014.
- [40] *Technical Documentation Integrated Stack Module (ISM)*, Staxera GmbH, Dresden, Germany, 2010, http://www.fuelcellmarkets.com/content/images/articles/00864_AS_Dokumentation_ISM.pdf.
- [41] European Union, "Generic diagnosis instrument for SOFC systems (GENIUS)," Collaborative Project FCH JU 245128, 2010, <http://www.fch-ju.eu/project/generic-diagnosis-instrument-sofc-systems>.
- [42] G. Moser and S. B. Serpico, "Contextual remote-sensing image classification by support vector machines and Markov random fields," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '10)*, pp. 3728–3731, Honolulu, Hawaii, USA, July 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

