Proceedings of the American Control Conference
Chicago, Illinois • June 2000

A case study of physical diagnosis for aircraft engines

Guillermo Gómez† Guy Campion Michel Gevers Pierre Willems
Centre for Systems Engineering and Applied Mechanics (CESAME)
Catholic University of Louvain, 1348 Louvain-la-Neuve, Belgium
† gomez@csam.ucl.ac.be

Abstract

This paper examines an ongoing research and development project concerned with identification, control and diagnosis in aircraft engines. Within these topics, this paper will focus on physical diagnosis, where one of the main motivations of the project is to introduce innovation and update techniques currently used in industry. Several fault diagnosis methods are considered and compared in terms of their possible application to the project. Further details are provided for the parametric statistical approach which, according to this initial study, is one approach flexible enough and with good foundations in order to achieve the desired objectives.

1. INTRODUCTION

The presentation in this paper is directly related to an ongoing research and development project with joint participation of both industrial and academic institutions. This project deals with matters such as identification, control and diagnosis applied to turbine engines. The focus in this paper will be on the topic of diagnosis.

The field of fault diagnosis has expanded continuously over the last years, with a variety of techniques, new results and innovations being reported. Proof of this is the list of survey and tutorial papers available (see, e.g. , [26], [17], [1], [16], [10], [18]) and also an ever growing list of applications. The motivation is clear, the increasing complexity of today systems and processes and also increasing interest and demand for safety and reliability.

The aerospace industry is no stranger to these developments. In fact, many of the industries involved in the project have their methods for condition monitoring and diagnosis. There is, however, consensual agreement that practical results have not been altogether good, but also that significant potential for improvement still exists.

Current techniques rely on simplified models, often linearized, which already impose limitations due to the physical complexity of a turbine engine and its inherent non-linear characteristics. These limitations may be avoided if more sophisticated models are used. The knowledge needed to describe and characterize these more sophisticated models is in the hands of today's main engine manufacturers and aircraft companies, which have accumulated a great deal of experience over the years. It is thus believed that this experience and knowledge, and derived models, can be a cornerstone in new developments.

The importance of modelling considerations is evident, because the fundamental concept behind many fault diagnosis methodologies is that of Analytical Redundancy. Analytical redundancy allows one to use measured process variables and analyze them in the light of a model that relates them. This analysis then enables one to check for incompatibilities in the data and, eventually, to take a decision about the cause of such an incompatibility.

We thus see that key elements for potential benefits are present. Industry has the knowledge to develop models that predict, with what appears as sufficient accuracy, the operation and behaviour of an engine. They also have the experience to pin-point elements of the model whose monitoring could guide subsequent diagnosis and maintenance. In parallel, a number of methods exist to exploit analytical redundancy for diagnosis purposes. Thus, the current project searches to benefit from the combination of all these elements.

The discussion in this paper will start by describing the general problem. A description of the engine model is also included to the extent needed for the subsequent discussion. Then we will describe what has been our contribution to the project so far. This starts by a brief summary of the basic elements of various approaches to fault diagnosis, describing also how each one of these methodologies could be applied to the diagnosis problem in the project. Among these methods, we will then provide further discussion and some simulation results for the parametric statistical approach, which appears as an interesting candidate thus far. This paper intends also to illustrate how part of the available theory can be considered in the light of an interesting industry application.

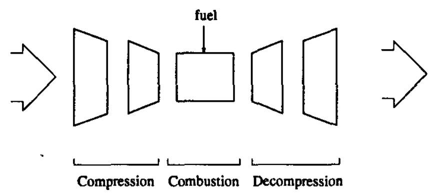

Fig. 1. General schema of an aircraft engine.

2. THE PROBLEM

The system under study is a 'standard' civil aircraft engine, and the general objective is to use a physical model of the engine for diagnosis purposes (i.e. physical diagnosis). The diagnosis is expected to provide information about 'undesired' changes that affect the engine during its life cycle. These changes can be either sudden faults or changes due to engine deterioration.

A clear challenge of the current application comes from the limited number of available measurements currently used in civil aviation (limited in comparison to the large number of variables present in a turbine engine). The benefits from the use of analytical redundancy don't come as an easy reward when the number of faulty situations considered in a diagnosis model grows, but the available measurements remain limited. One might be tempted, for the obvious reasons, to suggest increasing the number of measured variables. However, this would entail not only costs, but also risks since some of these variables may require devices placed in sensitive locations that could also compromise safety in engine operation, thus defeating the purpose of applying fault diagnosis techniques. The partners involved in the project do not envisage significant changes in current measurement sets, so the study has focused on approaches that consider a 'reasonably standard' set of measured variables.

With current methods the diagnosis is usually performed on-ground, so one of the potential innovations is precisely the possibility to change this. Diagnosis should be done on-board by processing the data collected during flights. Another important element is, as already mentioned, to develop the new diagnosis methods in the light of more sophisticated nonlinear models of the engine.

Since modeling is an essential element, we will first describe the generalities of an aircraft engine and the corresponding model.

3. THE ENGINE MODEL

Figure 1 shows the three main stages of a typical aircraft engine. In the model considered, the air is compressed with the aid of compressors. The compressed air then enters a combustion chamber where, in combination with fuel, it undergoes an exothermic reaction (combustion). The subsequent decompression takes place in turbines, where part of the 'thermodynamic' energy of the gas (air plus fuel) is transfered through the engine shafts back to other mechanical parts (like compressors). An important energy balance is that the energy needed for air compression comes from the energy transfered to the shafts in the turbines. If the engine is in steady-state (i.e. there is no shaft acceleration), then the energy needed in the compressors is equal to the energy 'produced' in the turbines.

As a result of its passing through the engine, the gas changes its speed, pressure, temperature, energy, etc. Some of these changes are related to the propulsion purpose of an aircraft engine. The thrust produced is, roughly explained, a result of the change of momentum of the air (air/fuel) entering and leaving the engine.

It is difficult to quantitatively describe the operation of an engine and all the variables involved. Nonetheless, a reasonably good model exists for the steady-state operation of the engine [1]. It is with this steady-state model that our first studies have been conducted.

[1] A dynamic model is also available, but less accurate.

0-7803-5519-9/00 $10.00 © 2000 AACC

The details of the model are not relevant for the current discussion. From a systemic point of view it is sufficient to indicate that this model involves equations of the form

$$X_p = G(X_u, U, \theta), \quad F(X_u, U, \theta) = 0, \tag{1}$$

where the map G describes the operation of the various engine components and stages, while F is an expression of balance equations including energy and mass conservation that constrain the interaction between the various components. The variables X_u represent model 'unknowns' that need to be solved in order to characterize a steady-state operating condition. The variables X_p are other process variables including temperatures, pressures, etc. The parameter vector θ characterizes the various equations and relationships.

Again, we won't give details about the solution of the equations involved. What is relevant for the current discussion is that the solution of such equations can be represented by a non-linear map of the form $X = [X_u^T, X_p^T]^T = \Phi_x(U, \theta)$. We explicitly indicate the dependence on θ because it clearly affects the solution of the model equations. We recall also that, in addition to the input conditions U, only some of the variables in X are measured. If we denote by Y the measured variables, then they can be described by

$$Y = \Phi(U, \theta). \tag{2}$$

Information about the effects of those faults which we are interested in monitoring has been described by the industry partners. The effects expected can be described by changes in certain key parameters of the engine model. Because of this, one may be tempted to directly link this type of diagnosis problem to system identification. Although the hitherto proposed answer is not unrelated to this idea, the first steps undertaken were to survey general fault diagnosis methods. We will first briefly summarize the ideas captured in this survey, discussing also the possible application of the surveyed methods in our diagnosis problem. Then we will focus on that method which, so far, appears as the best suited to our problem.

4. THE CANDIDATE METHODS

By candidate methods we don't mean each and every one of the methods described in the literature, where the list is enormous. Rather, we mean the basic ideas upon which are based a great part of the existing methods.

Naturally, the fundamental concept is that of redundancy. Two related concepts are used in applications, physical redundancy and analytical redundancy. Physical redundancy consists in exploiting the availability of numerous measurements to check, for instance, the existence of contradictions between sensors measuring the same variable. Analytical redundancy, on the other hand, makes use of a model that characterizes the system (or data), checking the compatibility of measured quantities with the system (or data) model.

For practical situations one can consider that a combination of both analytical and physical redundancy is used. The point is that the incorporation of additional measurements can provide more information about a particular system if the relationship (in other words, a model) of these signals with the others is understood. Thus, we see that a key issue is how models are used for diagnosis purposes. This is what differentiates the general methods considered here. Although there are connections, the available signals are processed in different ways to extract information that can indicate the existence of an abnormal situation. This is normally expressed in the form of so called 'residuals', which are signals that are ideally zero when no faults are present.

4.1. Parity equations

The use of parity equations is among the earliest ideas in fault diagnosis. In general terms, a set of parity equations describes an implicit mathematical relationship between the different variables of a system. Therefore, parity equations are an alternative way to model a particular system in the following way:

$$H(Y, U) = 0, \tag{3}$$

where Y and U are, for instance, the inputs and outputs of the system. Of course, such an implicit relation can be either linear or nonlinear, dynamical or static. Early methods were mostly focused on linear relationships for dynamical systems.

A good discussion about the formalism and use of parity equations for residual generation is contained in [6], where connections with methods based on state-estimators are also exposed. Other interesting ideas are included in [15], where some ideas that would also enable fault isolation are also discussed (see also [12], [14]).

Perhaps the most obvious way to generate residuals by means of parity equations is by defining the residual r as

$$r = H(Y, U). \tag{4}$$

Other more sophisticated approaches could include the use of filtered versions of U, Y and/or r, which could be seen as an attempt to enhance the properties of the new residual in terms of noise rejection and/or isolation properties. At least for linear systems, some connections with the use of state-estimators can thus be envisaged, since filtering U and Y 'appropriately' is closely related to the use of observers. Some more insight into these ideas can be found in [11], where a parametrization of all (linearly generated) residuals is presented (see also [12]).

In terms of applying this approach to our engine diagnosis problem, one of the limitations comes from the fact that the number of variables in X_p and X_u is very large, but only few of them are measured. This limits the way in which the model described in (1) could be used for residual generation. Being a static problem, almost the only residual that could be produced following the idea in equation (4) would be $r = Y - \Phi(U, \theta)$. In other words, the residual r would be the prediction error. Such residual would be interesting at least for detection problems. However, this residual is not very interesting for isolation purposes and there are some observations from the statistical approach, which will be considered later, that indicate that such a residual is indeed insufficient for isolation, especially for parametric changes in nonlinear systems [4]. An alternative would then be to consider a *filtered*, or transformed, residual based on r. The problem is that it is not straightforward to see what kind of transformation would be needed. An answer to this can be found in connection to the statistical approach discussed later.

4.2. Observers

Another methodology that supports numerous techniques for fault diagnosis is the use of observers (state-estimators). The early contributions were also focused on linear systems, but its use has motivated the application of similar ideas for nonlinear systems.

Residuals in this approach are generally based on innovation errors, and the motivation for doing so can be explained as follows. Consider the following linear system

$$x(k+1) = Ax(k) + Bu(k), \quad y(k) = Cx(k), \tag{5}$$

where y, u and x are, respectively, the system's output, input and state. The state x is estimated by an observer which, in a noisy environment, is usually implemented as a Kalman filter yielding an state-estimate $\hat{x}$. The innovation $\eta = y - C\hat{x}$ converges, under the no-fault hypothesis, to a zero mean uncorrelated sequence. This property can then be used for purposes of diagnosis.

We won't discuss further details of this approach since its application to our problem, with a static model, is limited. A possibility would still exist by considering parameter estimation as a state estimation problem. This usually involves a quadratic cost function used for optimization. Since this would indeed be an approach based on parameter estimation, we will discuss it with some more detail when we consider the corresponding approach.

4.3. Principal Component Analysis (PCA)

This is a tool that, though widely recognized in chemical engineering, had been somehow overlooked by the 'traditional' control literature on diagnosis problems. More recently, however, it has been taking a positive spin within the control community and related literature (see e.g. [27], [21], [22], [8], [9], [28]).

PCA is mainly a data analysis tool within a statistical framework. For diagnosis purposes, it is usually used first to generate a model for the data (the so called Principal Component Model-PCM) under normal operating conditions. Such a model can be generated with the aid of singular value decomposition (SVD), which reveals that the use of PCA is based on searching linear relationships among the available measurements (outputs and/or inputs).

The PCM can be described in terms of subspaces, the Principal Component Subspace (PCS) and the Residual Subspace (RS). The PCS describes a linear subspace wherein measured variables lie, while the RS is orthogonal to the PCS and is such that the orthogonal projection of data onto the RS is negligible (zero in an ideal situation with a perfect linear underlying system and no noise).

This subspace representation of PCA is useful to appreciate some straightforward connections between PCA and parity equations that, until recently, seemed to have been overlooked [13]. Let the available data (measurements) be represented by column vectors x_k, where each entry in x represents a measured variable (input or output) and k is a time index. Let C be the linear map (matrix) that projects the data onto the PCS and let $\hat{x}_k$ be the resulting projection. Then we can see that, under an ideal situation of perfect linear model and no noise, the data should satisfy $\hat{x}_k \doteq Cx_k = x_k$. Additionally, the corresponding projection to the RS would be $(I - C)x_k = 0$. It is then straightforward to show that the generation of the PCA model is an identification procedure for linear relationships between the data. Since this model can be obtained via SVD, we note that the identification of the PCA model is related to a quadratic cost function. Clearly, the underlying identified model has the form $Mx = 0$, where the rows of M are linearly independent rows of $(I - C)$ (or combinations of them). They are, therefore, linear parity equations.

Trying to use a PCA-based method in our diagnosis problem faces two complications. One is that PCA is a linear technique. This, however, could be surmounted by trying to use new, PCA-like techniques that extend some of the basic ideas of PCA to nonlinear cases (projections, principal curves, principal surfaces, etc [7], [20]). A second, perhaps more fundamental problem is related to the number of available measurements, which is limited. All together, we are left in a very similar position as we were with the use of parity equations and, in addition, it is not clear how faults could then be isolated and/or how the available physical model could be used to assist in this function.

4.4. Parameter estimation

A parameter estimation algorithm is a filter of input and output variables to produce estimates of model parameters for a given description of a system. If the result of the estimation is compared to what is known, or expected, of the system and a significant difference is encountered, then there are good grounds to argue that a change has occurred. This observation reveals why parameter estimation methodologies are of relevance for diagnosis problems, and why they seem particularly suited for faults connected to parametric changes.

The choice of the model structure is one of the key issues for diagnosis purposes, because it affects the definition of the model parameters. If the model is based on physical knowledge and is structured accordingly, then we can expect the corresponding parameters to have physical meaning. Thus, detecting changes in these parameters provides a meaningful tool to detect relevant changes (fault detection) and to indicate where they occurred (fault isolation).

A difficulty arises when the physical parameters are, due to the model structure, difficult to identify. A way to overcome this in system identification is by means of alternative model structures that could approximate the input-output behaviour of the system (polynomials, artificial neural networks, fuzzy systems, etc). The problem with this, however, is that the parameters of the newly structured model may not be meaningful for physical diagnosis and, additionally, their relationship with the physical parameters is usually unknown and hard to use for diagnosis purposes. For more related discussion we refer the reader to, for instance [19], [25], [29], [4].

The use of parameter estimation approaches is very important for our problem, especially as a mean to determine nominal values of model parameters under a normal (non-faulty) condition. For monitoring purposes, however, a slight change in focus has motivated us to consider the Parametric Statistical Approach. The point is that while most methods based on parameter estimation imply repeated estimation of model parameters, the statistical approach relies on testing hypothetical changes. This has the advantage of allowing one to concentrate on model structures which are suitable for physical diagnosis rather than for system identification. Drawing a line that would separate these approaches is indeed a tricky affair, since they are without doubt closely connected.

4.5. Parametric statistical approach - asymptotic local approach

This approach is quite general and can reveal connections between the various approaches discussed so far. The core of the methodology is the reduction of problems to simpler cases, where change detection and isolation can be treated in a unified way. Using the adequate information, together with insight from the *Asymptotic Local Approach*, most problems can be reduced to that of detecting changes in the mean of a Gaussian variable. This variable is used as a residual and is built with the aid of a model whose the parameters reflect the monitoring objectives[2]. The mean value of this residual depends on the monitored parameters and, therefore, faults affecting these parameters can be monitored. The resulting Gaussian distribution of the residual is a consequence of 'laws of large numbers'.

One of the key issues highlighted in this approach is the use of residuals that are relevant from an information point of view. In statistical terms, residuals should be sufficient statistics [4]. In the case of additive faults in linear systems (actuator and sensor faults), the use of the innovations of a Kalman filter makes sense because it provides such a sufficient statistic. However, this is in general not the case when faults correspond to components faults. The statistical approach could also shed some light in the use of non-linear state estimators for FDI purposes, raising the question about the feasibility of using the innovations as residual.

Another important idea, as we have recently indicated, is the slight difference in focus compared to parameter estimation approaches. Here, the focus is on monitoring rather than on repeated identification [4]. Although a system, and a particular model structure, may be difficult to identify, it may be less difficult to monitor. This is important because, for monitoring purposes, one can then concentrate on models with a physically motivated structure.

So far, it is this approach that we have deemed better suited to our problem. We will discuss with more detail how this approach can be used in the project, and will also show some simulation results. More references and general discussion about this methodology can be found in [5], [1], [2], [29], [4]. Some connections with parity equations and with the use of observers can be found in [32], [30].

5. AN INTERESTING CANDIDATE - THE PARAMETRIC STATISTICAL APPROACH

As already indicated, a key feature of this approach is the possibility to deal with various types of diagnosis problems. This is, to some extent, a result of the way in which residuals are generated. The statistical approach distinguishes between two residuals: primary residuals and improved residuals. This distinction is functional in the sense that the key feature of a primary residual is that it contains sufficient information for diagnosis purposes. On the other hand, improved residuals built from primary residuals have statistical properties that facilitate problem reduction and, thus, open the door to the solution of diagnosis problems in a unified way.

We will illustrate this approach by dealing directly with our diagnosis problem and the static model of the engine.

A possibility for a primary residual is to use the gradient of the least-squares prediction error. This residual, which we will denote by K, can be expressed as follows:

$$K(\theta, Y_k) = -\frac{1}{2}\frac{\partial}{\partial\theta}\left[\left(Y_k - \hat{Y}_k(\theta)\right)^T \left(Y_k - \hat{Y}_k(\theta)\right)\right], \qquad (6)$$

where Y_k is the vector of measurements at a time instant k, and $\hat{Y}_k(\theta)$ is the predicted value according to the map Φ in (2) for input values U_k. Note that this residual is given by

$$K(\theta, Y_k) = \left(\frac{\partial}{\partial\theta}\Phi(U_k, \theta)\right)^T (Y_k - \hat{Y}_k(\theta)), \qquad (7)$$

which we can interpret as a filtered (transformed) version of a residual obtained according to the parity equation $Y - \Phi(U, \theta) = 0$. This is the connection we had indicated before when we discussed the possibilities of using parity equations as an approach for our diagnosis problem. The parametric statistical approach thus provides a way to find a 'useful filter'.

The corresponding improved residual built from K is:

$$r_N(\theta) = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} K(\theta, X_k), \qquad (8)$$

where N is the size of the sample. For applications where one needs to constantly monitor for changes, it is possible to use schemes based on moving windows that include the last N samples available for diagnosis. Note that due to the static nature of our problem, there are no major restrictions on the interval between samples.

The improved residual r_N is the key to subsequent steps of change (fault) detection and isolation. The essential feature that

[2] These parameters may correspond to physical parameters or parameters that characterize the effect of additive faults (sensors and/or actuators' faults)

facilitates these steps is the probability distribution function (pdf) of r_N. Under appropriate assumptions and for sample's sizes large enough, r_N has the following pdf:

$$r_N \rightarrow \begin{cases} \mathcal{N}(0, \Sigma) & \text{when there is no parameter change} \\ \mathcal{N}(M\Delta\theta, \Sigma) & \text{when } \theta = \theta_0 + \Delta\theta/\sqrt{N} \end{cases} \quad (9)$$

where θ_0 is the known nominal (or identified) value of the parameter vector θ. The matrix M can be evaluated based on the available data and, in general, it involves a time average of $\frac{\partial}{\partial\theta}\Phi(U_k, \theta)^T \frac{\partial}{\partial\theta}\Phi(U_k, \theta)$. Due to the number of monitored parameters and the limited number of available measurements, the only way for M not to be singular is to use data from changing operating conditions (this is directly related to problems of persistent excitation in parameter estimation). The covariance matrix Σ can also be estimated from the data, though this depends on certain assumptions regarding, for instance, the absence of time correlation in the noise. Details are not really relevant for this presentation. Interesting information on this last issue is contained, for instance, in [31].

Remark 1: A clarification is due in relation to the sense of eq. (9). The parameter θ used in the residual r_N is a parameter of the model of the system and not of the real system. Therefore, the expression 'when $\theta = \theta_0 + \Delta\theta/\sqrt{N}$' must be understood as a reference to the change needed in the model to explain the deviations of the residual r_N, whenever changes have affected the real system.

Then, it should be clear that the residual r_N is evaluated at the nominal (or identified) value $\theta = \theta_0$.

5.1. *Residual evaluation - detection and isolation*

The residual r_N is the key for subsequent detection and isolation of changes. The basic idea behind the methodology is, in both cases, to test hypothesis about the mean of r_N (more insight and discussion can be found in, for instance, [3], [4]).

5.1.1. *Fault detection*

For fault detection, the idea is to check if the mean value of the residual is zero or not, which can be put as testing the (null) hypothesis H_0 ($\mu = M\theta = 0$) against the alternative hypothesis H_1 ($\mu = M\theta \neq 0$). The decision in this case can be taken based on the Generalized Likelihood Ratio (GLR)

$$t = 2 \ln \frac{\max_{\Delta\theta} p_{H_1}(r_N)}{p_{H_0}(r_N)}, \quad (10)$$

where $p_{H_i}(r_N)$ denotes the pdf of the residual r_N under Hypothesis H_i. For the normal pdf's shown in (9) it follows that

$$t = r_N^T \Sigma^{-1} M F^{-1} M^T \Sigma^{-1} r_N, \quad (11)$$

where $F = M^T \Sigma^{-1} M$ is the Fisher Information Matrix. When M is an invertible matrix the test t reduces to $t = r_N^T \Sigma^{-1} r_N$.

Under both hypotheses (H_0 and H_1), t has a $\mathcal{X}^2$ pdf. It is a central distribution for H_0 and a non-central one for H_1, with non-centrality parameter $\Delta\theta^T F \Delta\theta$ (the number of degrees of freedom is the number of monitored parameters). This information is useful to define detection thresholds according to false alarm probabilities.

5.1.2. *Fault isolation*

The problem of isolation is more delicate than that of detection. In the light of basic ideas from decision theory, the greater complexity in this game between 'nature' and the 'statistician' is associated to the increased number of actions (decisions) that are at stake. It is not obvious how to see the problem of isolation as simple hypothesis tests, but some methods such as multiple comparison procedures are certainly related to this [23], [24].

In spite of the difficulty, there are two basic tools that can assist in the task of isolation. These tools are the *Sensitivity* test and the *Min-max* test, which consider a partition of the parameter $\theta^T = [\theta_a^T, \theta_b^T]$.

Sensitivity test The basic idea behind this test is to attribute any changes to a portion of the parameters such as, for example, the part θ_a. In this case, one tests between hypothesis H_{a0} ($\Delta\theta_a = 0$, $\Delta\theta_b = 0$) and H_{a1} ($\Delta\theta_a \neq 0$, $\Delta\theta_b = 0$). This can be solved with the following test statistic:

$$\tilde{t}_a = 2 \ln \frac{\max_{\Delta\theta} p_{H_{a1}}(r_N)}{p_{H_{a0}}(r_N)}, \quad (12)$$

$$= r_N^T \Sigma^{-1} M_a (M_a^T \Sigma^{-1} M_a)^{-1} M_a^T \Sigma^{-1} r_N, \quad (13)$$

where the matrix M is partitioned as $M = [M_a \; M_b]$, according to the parameter partition. The test statistic $\tilde{t}_a$ is subsequently used in tests that compare its value either to a threshold or to similar test statistics (for example, sensitivity test statistics for a different portion of the parameters).

Min-max test The basic idea is similar to that of the Sensitivity test, but the (key) difference is that the remaining portion of the parameter (e.g. θ_b) is considered as an unknown nuisance parameter. A common approach for this is to replace the nuisance parameter for the value that minimizes (for an hypothetical change in θ_a) the power of the (detection) test (i.e. the value that minimizes the probability of correctly detecting a fault). The corresponding test statistic is numerically equivalent to the following GLR [4]

$$t_a^* = 2 \ln \frac{\max_{\Delta\theta_a, \Delta\theta_b} p_{\Delta\theta_a, \Delta\theta_b}(r_N)}{\max_{\Delta\theta_b} p_{0, \Delta\theta_b}(r_N)}. \quad (14)$$

Thus, the min-max test statistic can be evaluated as $t_a^* = r_{a*}^T (F_a^*)^{-1} r_{a*}$, where $r_{a*} = [I, \; -F_{ab} F_{bb}^{-1}] M^T \Sigma^{-1} r_N$, $F_a^* = F_{aa} - F_{ab} F_{bb}^{-1} F_{ba}$.

As for the Sensitivity test, the min-max test statistic t_a^* is subsequently used in comparison tests.

Based on these two different tests, one can conceive various approaches for isolation. For cases when the number of faults (number of changed parameters) is known (let q be this number), the decision about the faulty group of parameters can be made by choosing the group of q parameters with the biggest sensitivity test statistic. This choice is related to a Bayes' approach of choosing the group that yields the maximum *a posteriori* probability of the residual. According to simulations that we have carried out with the engine model, this approach works well (for the considered case of a known number of faults).

The situation is more difficult when the number of changed parameters is, as expected in practice, unknown. A possibility is to use the ideas of multiple comparisons [23], [24] and the min-max test. The decision about changes in each parameter is based on the min-max test statistic of each individual parameter. A problem with this approach is that min-max tests for individual parameters can be very conservative and, therefore, not detect many practical cases. This was apparent in the simulations with our model.

A third alternative is a combined approach that would try to take advantage of the good properties of both tests. The sensitivity approach has good isolation performance when the number of faults is known, while the min-max approach can enhance robustness to changes in other parameters. The *incremental diagnosis* algorithm presented here is along the lines of the one presented in [23] and is based on the following three steps:

Step 1. Assume that there are i changed parameters (the initial value for i is 1) and calculate the corresponding sensitivity test statistics for all possible combinations of i parameters.

Step 2. Choose as faulty parameters the combination with the biggest sensitivity test statistic. Do a partition of the vector parameter as $\theta = [\theta_a, \theta_b]$, where θ_a contains the hitherto isolated parameters and θ_b the remaining ones. Evaluate the rejection test statistic corresponding to θ_b.

Step 3. If the rejection test statistic for θ_b breaches a predefined threshold then it is decided that a change exists in θ_b, i is incremented in one unit and the algorithm returns to Step 1. Otherwise, the algorithm stops and the result of the isolation procedure is θ_a.

To illustrate these ideas, we summarize some results obtained with the proposed incremental algorithm. Several cases were simulated wherein parameter changes were introduced to the model used to generate data (recall that the diagnosis is performed based on the nominal model). The simulated experiment for each case consisted in adding random noise (uniformly distributed with a range of approximately 1% of maximum readings) to the generated data from the faulty model, and then carry on with the incremental diagnosis. For each case, the corresponding experiment was repeated several times. The model considers 9 measurements and 11 monitored parameters θ_1 to θ_{11} representing various parameters involved in thermodynamical equations along the engine.

Case 1

- θ_3 was decreased in 1% from its nominal value.
- Number of repeated experiments : 1000.
- Fault detection (based on global test) : In all experiments the fault is detected.

• Number of exact isolations : 968 (out of 968 cases with precisely 1 isolated parameter).
• Parameter θ_3 in isolated parameters : 1000.
• Individually, the following table summarizes how many times each parameter is part of the isolated parameters.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
0	8	1000	7	5	2	8	1	10	7	1

Case 2.
• Parameters θ_1, θ_2 and θ_{11} were decreased in 1% from their nominal values.
• Number of repeated experiments : 500.
• Fault detection (based on global test) : In all experiments the fault is detected.
• Number of exact isolations : 446 (out of 447 cases with precisely 3 isolated parameters).
• Parameters θ_1, θ_2 and θ_{11} in the isolated parameters : 446.
• Individually, the following table summarizes how many times each parameter is part of the isolated parameters.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
500	446	0	0	0	0	0	0	0	1	500

Case 3.
• Parameters θ_3, θ_5, θ_7 and θ_{10} were decreased in 1% from their nominal values.
• Number of repeated experiments : 928.
• Fault detection (based on global test): In all experiments the fault is detected.
• Number of exact isolations : 511 (out of 512 cases with precisely 4 isolated parameters) .
• Parameters θ_3, θ_5, θ_7 and θ_{10} in the isolated parameters : 852.
• Individually, the following table summarizes how many times each parameter is part of the isolated parameters.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
14	66	884	106	862	65	926	227	232	918	12
1.51%	7.11%	95.26%	11.42%	92.89%	7.00%	99.78%	24.46%	25%	98.92%	1.29%

• The following table is similar to the previous one, but only for those cases where θ_3, θ_5, θ_7 and θ_{10} are in the isolated parameters.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
12	50	852	53	852	59	852	154	169	852	0
1.41%	5.87%	100%	6.22%	100%	6.92%	100%	18.08%	19.84%	100%	0%

Case 4.
• All the first 10 parameters were reduced in 1% from their nominal value.
• Number of repeated experiments : 500.
• Fault detection (based on global tests): In all experiments the fault is detected.
• Number of exact isolations : 0.
• Parameters 1 to 8 (and only these ones) were isolated in 260 experiments.
• Parameters 1 to 8 in the isolated parameters : 261 times (260 times with parameters 1 to 8, plus 1 time parameters 1 to 9).
• Individually, the following table summarizes how many times each parameter is part of the isolated parameters.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}
500	295	433	499	319	500	475	325	238	238	49

It can be observed in Cases 1, 2 and 3 that correct isolation can achieve an extremely successful rate for cases where the number of changed parameters is known. Overall, cases 1 and 2 show very good isolation rates (over 89%), while case 3 is around 55% for exact isolations. Besides, the ratio of cases where the isolated parameters contain the faulty ones is still high. Case 4 is much less successful, but it should be noted that it is a much more extreme case for the isolation algorithm.

Another point we wish to bring out, although not shown in the above simulation results, is that the isolation results are in general better with this approach than with a 'blind' parameter estimation approach, where detection/isolation is based on the simultaneous re-estimation of all parameters.

6. Conclusion

We have described an industry motivated problem of fault diagnosis, which is intended to introduce innovations in the current practice for aircraft engines and, eventually, for similar applications. The necessary modelling information was presented and, subsequently, various approaches to fault diagnosis were briefly summarized and considered in the light of our engine diagnosis problem. The discussion then was focused on a method that has been considered well suited for our particular application, the parametric statistical approach. The results obtained so far are encouraging, but it still remains to study what the possibilities are if similar ideas are applied together with a dynamical model of the engine.

Acknowledgements. This research is being developed with the support of contract BRPR-CT97-0601, BRITE/EURAM Project BE 97-4077, with partial financing from the Commission of the European Communities. We greatly acknowledge the participation of all the partners involved, namely Snecma, Rolls-Royce plc, Motoren-und Turbinen-Union (MTU), Volvo Aero Corporation, Fiat Avio, Techspace Aero S.A., National Technical University of Athens, Lufthansa Technik AG, Technische Universitat Munchen, Universitat Stuttgart and Chalmers University of Technology AB.

References

[1] M. Basseville. Detecting changes in signals and systems - a survey. *Automatica*, 24:309–326, 1988.
[2] M. Basseville. A unified framework for statistical change detection. In *30th CDC*, pp. 2586–2591, Brighton, 1991.
[3] M. Basseville. Information criteria for residual generation and fault detection and isolation. *Automatica*, 33(5):783–803, 1997.
[4] M. Basseville. On-board component fault detection and isolation using the statistical local approach. *Automatica*, 34:1391–1415, 1998.
[5] A. Benveniste, M. Basseville, and G. Moustakides. The asymptotic local approach to change detection and model validation. *IEEE Tr. on Automatic Control*, 32:583–592, 1987.
[6] E. Y. Chow and A. S. Willsky. Analytical redundancy and the design of robust failure detection systems. *IEEE Tr. on Automatic Control*, 29:603–614, 1984.
[7] D. Dong and T. J. McAvoy. Nonlinear principal component analysis - based on principal curves and neural networks. *ACC*, pp. 1284–1288, Baltimore, 1994.
[8] R. Dunia, S. Qin, T. F. Edgar, and T. McAvoy. Identification of faulty sensors using principal components analysis. *AIChE Journal*, 42:2797–2812, 1996.
[9] R. Dunia and S. J. Qin. Multi-dimensional fault diagnosis using a subspace approach. *ACC*, Albuquerque, 1997.
[10] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results. *Automatica*, 26(3):459–474, 1990.
[11] P. M. Frank and X. Ding. Frequency domain approach to optimally robust residual generation and evaluation for model-based fault diagnosis. *Automatica*, 30:789–804, 1994.
[12] J. Gertler. Fault detection and isolation using parity relations. *Control Eng. Practice*, 5(5):653–661, 1997.
[13] J. Gertler, W. Li, Y. Huang, and T. McAvoy. Isolation enhanced principal component analysis. *AIChE Journal*, 45(2):323–334, 1999.
[14] J. Gertler and Q. Luo. Direct identification of nonlinear parity relations - a way arround the robustness problem. *37th CDC*, Tampa, Florida., 1998.
[15] J. Gertler and D. Singer. A new structural framework for parity equation-based failure detection and isolation. *Automatica*, 26:381–388, 1990.
[16] J. J. Gertler. Survey of model-based failure detection and isolation in complex plants. *IEEE Control Systems Magazine*, 1988.
[17] R. Isermann. Proces fault detection based on modeling and estimation methods - a survey. *Automatica*, 20(4):387–404, 1984.
[18] R. Isermann. Fault diagnosis of machines via parameter estimation and knowledge processing - tutorial paper. *Automatica*, 29:815–835, 1993.
[19] R. Isermann and B. Freyermuth. Proces fault diagnosis based on process model knowledge - Part I: principles for fault diagnosis with parameter estimation. *Tr. of the ASME (J. of Dynamic Systems)*, 113:620–626, 1991.
[20] F. Jia, E. B. Martin, and A. J. Morris. Non-linear principal component analysis for process fault detection. *Comp. & Chem. Eng.*, 22:S851–S854, 1998.
[21] K. A. Kosanovich and M. J. Piovoso. Process data analysis using multivariate statistical methods. *ACC*, pp 721–724, Boston, 1991.
[22] J. V. Kresta, J. F. MacGregor, and T. E. Marlin. Multivariate statistical monitoring of process operating performance. *The Can. J. of Chem. Eng.*, 69:35–47, 1991.
[23] Gaël Mathis. *Outils de Détection de Rupture et de Diagnostic: Application à la Surveillance de Turbines à Gaz.* PhD thesis, Université de Rennes I., 1994.
[24] E. Spjϕotvoll. On the optimality of some multiple comparison procedures. *The Annal of Mathematical Statistics*, 43(2):398–411, 1972.
[25] A. Srinivasan and C. Batur. Hopfield/art-1 neural network based fault detection and isolation. *IEEE Tr. on Neural Networks*, 5:890–899, 1994.
[26] A.S. Willsky. A survey of design methods for failure detection systems. *Automatica*, 12(6):601–611, 1976.
[27] B. M. Wise, N. L. Ricker, D. F. Veltkamp, and B. R. Kowalsky. A theoretical basis for the use of principal component models for monitoring multivariate processes. *Proces Control and Quality*, pp. 41–51, 1990.
[28] J. Zhang, E. B. Martin, and A. J. Morris. Fault detection and diagnosis using multivariate statistical techniques. *Tr. of the Institution of Chemical Engineers (Part A)*, 74:89–96, 1996.
[29] Q. Zhang. Fault detection and isolation with non-linear black-box models. In *IFAC Conference on System Identification (SYSID97)*, 1997.
[30] Q. Zhang and M. Bassevile. Monitoring nonlinear dynamical systems: a combined observer-based and local approach. *37th CDC*, Tampa, Florida USA, 1998.
[31] Q. Zhang, M. Basseville, and A. Benveniste. Early warnings of slight changes in systems. *Automatica*, 30(1):95–113, 1994.
[32] Q. Zhang, M. Basseville, and A. Benveniste. Fault detection and isolation in nonlinear dynamic systems: A combined input-output and local approach. *Automatica*, 34(11):1359–1373, 1998.