Journal of Computational and Applied Mathematics 234 (2010) 2202–2212

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Comparison of adaptive filters for gas turbine performance monitoring

S. Borguet *, O. Léonard

Department of Aerospace and Mechanics, Turbomachinery Group, University of Liège, Chemin des Chevreuils 1, B-4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 8 August 2008
Received in revised form 16 December 2008

Keywords:
Gas path analysis
Adaptive estimation
Kalman filter

ABSTRACT

Kalman filters are widely used in the turbine engine community for health monitoring purpose. This algorithm has proven its capability to track gradual deterioration with a good accuracy. On the other hand, its response to rapid deterioration is either a long delay in recognising the fault, and/or a spread of the estimated fault in several components. The main reason of this deficiency lies in the transition model of the parameters that assumes a smooth evolution of the engine's condition. The aim of this contribution is to compare two adaptive diagnosis tools that combine a Kalman filter and a secondary system that monitors the residuals. This auxiliary component implements on one hand a covariance matching scheme and on the other hand a generalised likelihood ratio test to improve the behaviour of the diagnosis tool with respect to abrupt faults.

1. Introduction

In the last years, predictive maintenance has been widely promoted in the jet engine community. A maintenance schedule adapted to the level of deterioration of the engine leads to many advantages such as improved operability and safety or reduced life cycle costs. In this framework, generating a reliable information about the health condition of the engine is a requisite.

In this paper, **Module Performance Analysis** is considered. Its purpose is to detect, isolate and quantify the changes in engine module performance, described by so-called health parameters, on the basis of measurements collected along the gas path of the engine [1]. Typically, the health parameters are correcting factors on the efficiency and the flow capacity of the modules (fan, lpc, hpc, hpt, lpt, nozzle) while the measurements are inter-component temperatures, pressures and shaft speeds. As illustrated in Fig. 1, the diagnosis problem (or health parameter estimation problem) can be regarded as the inverse problem of performance simulation.

Fig. 2 sketches a typical Exhaust Gas Temperature profile versus engine usage time. As far as time scale is considered, engine health variations can be divided into two groups. On one hand, **gradual deterioration** (due to erosion or fouling for instance) occurs during normal operation of the engine and affects all major components at the same time. On the other hand, **accidental events**, caused for instance by hot restarts or foreign/domestic object damage,[1] impact one (at most two) component(s) at a time and occur infrequently. The time of occurrence of the event, as well as the impacted component and the magnitude of the fault are typically unknown.

Among the numerous techniques that have been investigated to solve the diagnosis problem, see [2] for a detailed review, the popular Kalman filter [3] has received special attention. This recursive, minimum-variance algorithm has proven its capability to track gradual deterioration such as wear with good accuracy. Indeed, the Kalman filter embeds a transition model that describes a "relatively slow" evolution of the health parameters. On the other hand, the response of the Kalman

* Corresponding author.
E-mail addresses: s.borguet@ulg.ac.be (S. Borguet), o.leonard@ulg.ac.be (O. Léonard).

[1] A Domestic Object Damage is caused by an element of the engine (*e.g.* part of a blade) that breaks off and strikes a downstream flow path component.

0377-0427/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2009.08.075

Nomenclature

$\widehat{\cdot}$	Estimated value
A8IMP	Nozzle exit area (nominal value : 1.4147 m^2)
EGT	Exhaust Gas Temperature
EKF	Extended Kalman Filter
GLR	Generalised Likelihood Ratio
hpc	High pressure compressor
hpt	High pressure turbine
k	Discrete time index
lpc	Low pressure compressor
lpt	Low pressure turbine
M	Size of the buffer
N	Rotational speed
n_w	Number of health parameters
n_y	Number of measurements
p_i^0	Total pressure at station i
SEi	Efficiency factor of the component whose entry is located at section i (nominal value : 1.0)
SWiR	Flow capacity factor of the component whose entry is located at section i (nominal value : 1.0)
T_i^0	Total temperature at station i
$\mathbf{u}_k$	Actual command parameters
$\mathbf{w}_k$	Actual but unknown health parameters
$\mathbf{y}_k$	Observed measurements
$\boldsymbol{\Delta}_{\mathbf{w}}$	Unknown abrupt fault
$\boldsymbol{\epsilon}_k$	Measurement noise vector
$\boldsymbol{\nu}_k$	Process noise vector
$\boldsymbol{\tau}$	Unknown time of occurrence of the abrupt fault
$\mathcal{N}(\mathbf{m}, \mathbf{R})$	A Gaussian probability density function with mean $\mathbf{m}$ and covariance matrix $\mathbf{R}$

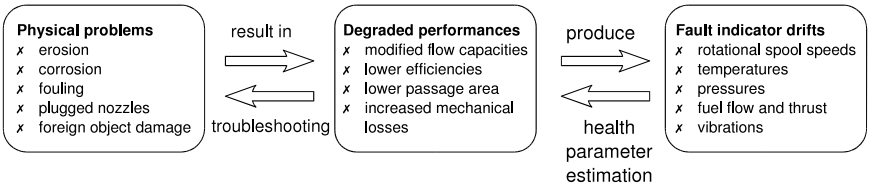


Fig. 1. The Gas Path Analysis approach to jet engine diagnostics.

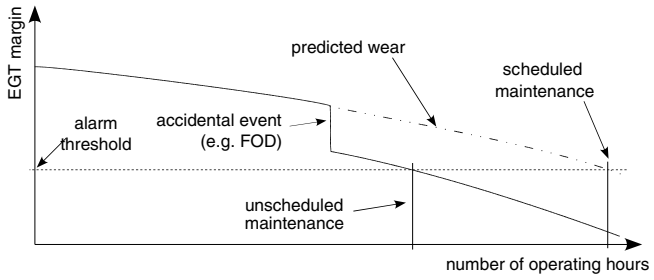


Fig. 2. Typical EGT margin profile showing gradual and abrupt health variations.

filter to short-time-scale variations in the engine condition is either a long delay in recognising the fault, or/and a spread of the estimated fault over several components which is termed the "smearing" effect, see [4].

One way to tackle this problem is to reconsider it in the realm of **adaptive estimation** [5]. The basic idea consists of increasing the mobility of the health parameters momentarily in order to recognise a rapid degradation. Two different approaches have been investigated by the authors in previous researches. In [6], a **Covariance Matching** scheme has been implemented in the Kalman filter. The constraints on the rate of variation of the health parameters are tuned on-line by ensuring consistency between the observed residuals and their statistics. In [7], the transition model for the health

parameters is modified so as to account for possible jumps in the parameters. A **Generalised Likelihood Ratio Test** that detects and estimates abrupt faults is added to the Kalman filter for this improved transition model.

The present contribution aims at comparing both adaptive algorithms at the theoretical, implementation and performance levels. As far as this last issue is concerned, the adaptive diagnosis tools are submitted to a series of simulated fault cases that may be encountered on a commercial aircraft engine.

2. Kalman-filter-based diagnostics

The scope of this section is to describe the diagnosis tool which relies on the celebrated Kalman filter [3]. One of the master pieces of this algorithm is a model of the jet engine. Considering a steady-state operation of the gas turbine, these simulation tools are generally nonlinear aero-thermodynamic models based on mass, energy and momentum conservation laws applied to the engine flow path. Eq. (1) represents such an engine model where k is a discrete time index, $\mathbf{u}_k$ are the parameters defining the operating point of the engine (e.g. fuel flow, altitude, Mach number), $\mathbf{w}_k$ are the health parameters and $\mathbf{y}_k$ are the gas path measurements. A random variable $\boldsymbol{\epsilon}_k \in \mathcal{N}(\mathbf{0}, \mathbf{R}_\mathbf{r})$ which accounts for sensor inaccuracies and modelling errors is added to the deterministic part $\mathcal{G}(\cdot)$ of the model in order to reconcile the observed measurements and the model predictions. Eq. (1) is therefore termed the statistical model.

$$\mathbf{y}_k = \mathcal{G}(\mathbf{u}_k, \mathbf{w}_k) + \boldsymbol{\epsilon}_k. \tag{1}$$

In the frame of turbine engine diagnosis, the quantity of interest is the difference between the actual engine health condition and a reference one. In the *recursive* approach that is proposed here, this reference value is represented by a so-called prior value which designates a value of the health parameters $\widehat{\mathbf{w}}_k^-$ that is available before the measurements $\mathbf{y}_k$ are observed. Assuming a linear relationship between the measurements and the health parameters around the prior values, as well as given operating conditions, the statistical model is reformulated according to Eq. (2).

$$\boldsymbol{\epsilon}_k = \mathbf{r}_k - \mathbf{G}_k(\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) \tag{2}$$

where

$$\mathbf{r}_k \triangleq \mathbf{y}_k - \mathcal{G}(\mathbf{u}_k, \widehat{\mathbf{w}}_k^-) \quad \text{and} \quad \mathbf{G}_k \triangleq \left.\frac{\partial \mathcal{G}(\mathbf{u}_k, \mathbf{w}_k)}{\partial \mathbf{w}_k}\right|_{\mathbf{w}_k = \widehat{\mathbf{w}}_k^-} \tag{3}$$

are respectively the a priori residuals and the Jacobian matrix of the engine model at the prior value $\widehat{\mathbf{w}}_k^-$.

A common interpretation of the Kalman filter is that of a recursive, bayesian algorithm for parameter identification. The health parameters and the measurement noise are considered here as Gaussian random variables.[2] Within this framework, the estimated health parameters are obtained by minimising the following objective function

$$\mathcal{J}(\mathbf{w}_k) = (\mathbf{w}_k - \widehat{\mathbf{w}}_k^-)^\mathrm{T} \left(\mathbf{P}_{\mathbf{w},k}^-\right)^{-1} (\mathbf{w}_k - \widehat{\mathbf{w}}_k^-) + \boldsymbol{\epsilon}_k^\mathrm{T} \mathbf{R}_\mathbf{r}^{-1} \boldsymbol{\epsilon}_k. \tag{4}$$

The first term on the right-hand side of Eq. (4) forces the identified parameters to remain in the neighbourhood of the prior values $\widehat{\mathbf{w}}_k^-$, the prior covariance matrix $\mathbf{P}_{\mathbf{w},k}^-$ specifying the shape of this region. The second term reflects a weighted-least-squares criterion.

To generate the a priori values of the health parameter distribution (*i.e.* mean $\widehat{\mathbf{w}}_k^-$ and covariance $\mathbf{P}_{\mathbf{w},k}^-$), a model describing the temporal evolution of the parameters must be supplied as well. Generally, little information is available about the way the engine degrades which motivates the choice of a random walk model

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \boldsymbol{\nu}_k. \tag{5}$$

The random variable $\boldsymbol{\nu}_k \in \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ is the so-called process noise that provides some adaptability to track a time-evolving fault. In the present application it is assumed that the health parameters vary independently such that the covariance matrix $\mathbf{Q}_k$ is strictly diagonal. Even if the transition model (5) appears quite simple, the covariance matrix $\mathbf{Q}_k$ enables the control of the stochastic character of the time series formed by the health parameters $\mathbf{w}_k$: low values mean slow variations while high values suppose fast variations.

Algorithm 1 summarises in a pseudo-code style the basic processing step of the extended Kalman filter. This algorithm has a predictor–corrector structure and involves only basic linear algebra operations. On line 1, prediction of the prior values of the health parameter distribution are made through the transition model (5). Then the data are acquired and used for building the a priori residuals (lines 2 and 3). The Jacobian matrix is assessed on line 4 and subsequently used in the computation of the covariance matrix of the residuals $\mathbf{P}_{\mathbf{y},k}$ (line 5) and of the Kalman gain $\mathbf{K}_k$ (line 6). Loosely speaking, it weights the uncertainty on the parameters versus the one on the measurements. Finally, the a posteriori distribution is assessed at the corrector step (line 7).

To complete the picture, the block diagram in Fig. 3 shows this closed-loop, predictor–corrector structure. The interested reader may consult Reference [8] for an extensive derivation and additional details.

[2] Statistically, they are hence thoroughly defined by their mean and covariance matrix.

Algorithm 1 : Basic step of the extended Kalman filter

1: $\widehat{\mathbf{w}}_k^- = \mathbf{w}_{k-1}$ and $\mathbf{P}_{\mathbf{w},k}^- = \mathbf{P}_{\mathbf{w},k-1} + \mathbf{Q}_k$
2: acquire $\mathbf{u}_k$ and $\mathbf{y}_k$
3: $\mathbf{r}_k = \mathbf{y}_k - \mathcal{G}(\mathbf{u}_k, \widehat{\mathbf{w}}_k^-)$
4: compute Jacobian matrix as per eqn. 3
5: $\mathbf{P}_{\mathbf{y},k} = \left(\mathbf{G}_k\, \mathbf{P}_{\mathbf{w},k}^-\, \mathbf{G}_k^{\mathrm{T}} + \mathbf{R}_\mathbf{r}\right)$
6: $\mathbf{K}_k = \mathbf{P}_{\mathbf{w},k}^-\, \mathbf{G}_k^{\mathrm{T}}\, \mathbf{P}_{\mathbf{y},k}^{-1}$
7: $\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^- + \mathbf{K}_k\, \mathbf{r}_k$ and $\mathbf{P}_{\mathbf{w},k} = (\mathbf{I} - \mathbf{K}_k\, \mathbf{G}_k)\, \mathbf{P}_{\mathbf{w},k}^-$

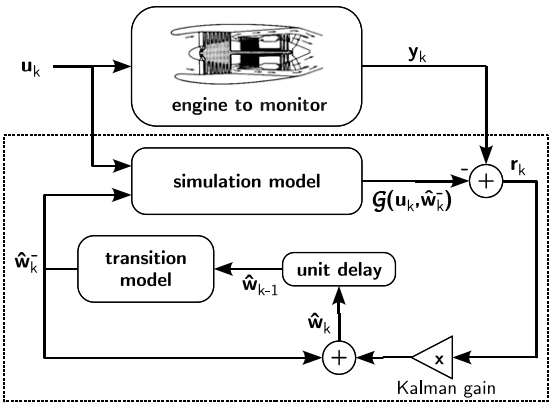


Fig. 3. Performance monitoring tool based on an Extended Kalman Filter.

3. Adaptive algorithms

To improve the tracking abilities of abrupt faults without sacrificing the reliability of the estimation of long-time-scale deterioration, adaptive estimation is considered. The approach is based on the assumption that abrupt faults may occur, but that they occur infrequently. Hence, the core of the adaptive algorithm consists of a Kalman filter, which relies on the assumption of a smooth variation of the engine condition. An auxiliary component complements the design. Basically, this secondary system monitors the residuals of the filter to determine whether an abrupt event has occurred and adjusts the response of the filter accordingly. In the following, two techniques are presented: the first one implements a covariance matching scheme and the second one uses a generalised likelihood ratio test.

3.1. Covariance matching

The adaptive algorithm based on the covariance matching technique is inspired by the work in [9]. It is intended to enforce consistency between the predicted residuals $\mathbf{r}_k$ and their statistics. In short, the adaptive algorithm provides an on-line feedback from the residuals in terms of process noise levels. A thorough description of the methodology being provided in Reference [6], only the major elements are recalled in the following.

The implementation of the adaptive feature relies on a buffer containing the $M+1$ latest residuals. The estimation is hence delayed by M time steps ; it means that at time step k, the most recent estimate is $\widehat{\mathbf{w}}_{k-M-1}$ and that the new estimate $\widehat{\mathbf{w}}_{k-M}$ will be based on the residuals in the buffer. The covariance matching scheme is applied to the averaged residual over the $M+1$ samples of the buffer

$$\bar{\mathbf{r}}_k = \frac{1}{M+1}\sum_{l=0}^{M} \mathbf{r}_{k-l}. \tag{6}$$

This averaging makes the mean $\bar{\mathbf{r}}_k$ less sensitive to the measurement noise. Indeed, it can be shown that the mean residual $\bar{\mathbf{r}}_k$ is a white and Gaussian random variable with zero-mean, $\mathcal{E}(\bar{\mathbf{r}}_k|\mathbf{y}_{k-M-1}) = 0$ and covariance matrix given by

$$\mathcal{E}(\bar{\mathbf{r}}_k\, \bar{\mathbf{r}}_k^{\mathrm{T}}|\mathbf{y}_{k-M-1}) = \overline{\mathbf{G}}_{k,M}\, \mathbf{P}_{\mathbf{w},k-M}\, \overline{\mathbf{G}}_{k,M}^{\mathrm{T}} + \overline{\mathbf{R}}_r + \sum_{l=0}^{M} \overline{\mathbf{G}}_{k,l}\, \mathbf{Q}_k\, \overline{\mathbf{G}}_{k,l}^{\mathrm{T}} \tag{7}$$

where the $n_y \times n_w$ matrix $\overline{\mathbf{G}}_{k,l}$ and the $n_y \times n_y$ matrix $\overline{\mathbf{R}}_r$ are defined as

$$\overline{\mathbf{G}}_{k,l} = \frac{1}{M+1}\sum_{i=0}^{l} \mathbf{G}_{k-i} \quad \text{and} \quad \overline{\mathbf{R}}_r = \frac{1}{M+1}\mathbf{R}_\mathbf{r}. \tag{8}$$

The covariance matching scheme ensures consistency of the residuals with their statistics by determining $\mathbf{Q}_k$ such that

$$\operatorname{diag}\left(\bar{\mathbf{r}}_k\ \bar{\mathbf{r}}_k^{\mathrm{T}}\right) = \operatorname{diag}\left(\mathcal{E}(\bar{\mathbf{r}}_k\ \bar{\mathbf{r}}_k^{\mathrm{T}}|\mathbf{y}_{k-M-1})\right) \tag{9}$$

where the operator diag $(\cdot)$ designates the vector made of the diagonal values of a square matrix. The matching criterion (9) is restricted to the *diagonal* terms of the matrix $(\bar{\mathbf{r}}_k\ \bar{\mathbf{r}}_k^{\mathrm{T}})$ as the off-diagonal terms are sensitive to the measurement noise, even after the averaging performed by Eq. (6). The left-hand side of Eq. (9) is computed from the buffer of residuals, while the right-hand side is the expected theoretical value from Eq. (7) that *does not depend* on the data.

For sake of simplicity, the matrix $\mathbf{Q}_k$ is reduced to a vector containing its diagonal terms which leads to the following equality to be verified

$$\begin{aligned}&\bar{\mathbf{r}}_k^2 - \operatorname{diag}\left(\overline{\mathbf{G}}_{k,M}\ \mathbf{P}_{\mathbf{w},k-M}\ \overline{\mathbf{G}}_{k,M}^{\mathrm{T}} + \overline{\mathbf{R}}_r\right) = \operatorname{diag}\left(\sum_{l=0}^{M} \overline{\mathbf{G}}_{k,l}\ \mathbf{Q}_k\ \overline{\mathbf{G}}_{k,l}^{\mathrm{T}}\right)\\ &\Leftrightarrow \mathbf{d}_k = \mathbf{B}_k\ \mathbf{f}_k\end{aligned} \tag{10}$$

where

$$\mathbf{f}_k = \operatorname{diag}(\mathbf{Q}_k) \quad \text{and} \quad \mathbf{B}_k = \sum_{l=0}^{M} \overline{\mathbf{G}}_{k,l}^2 \tag{11}$$

note that in the computation of $\mathbf{B}_k$, the square operator is applied element-wise.

In turbine engine diagnosis the number of health parameters n_w generally exceeds the number of sensors n_y. As a result, the matrix $\mathbf{B}_k$ is not invertible and Eq. (10) has no unique solution. The diagonal terms of $\mathbf{Q}_k$ are obtained from a maximum a posteriori approach. Obviously, the maximum a posteriori solution is limited to positive values as $\widehat{\mathbf{f}}_k$ is a variance. The prior distribution is specified through its mean value $\mathbf{f}^{min}$ and its covariance matrix $\mathbf{P}_f$. $\mathbf{f}^{min}$ is set to the value of the variance of the process noise $\mathbf{Q}_k^{min}$ that is used to track gradual deterioration. The diagonal terms of $\mathbf{P}_f$ reflect the maximum expected magnitude of an abrupt event.

The residuals $\mathbf{d}_k$ in relation (10) are not Gaussian as they result from a non-linear operation on the residuals $\bar{\mathbf{r}}_k$. Consequently, the maximum a posteriori solution may be quite noisy for small-sized buffers. This generates spurious covariance rises that decrease the quality of the filtering. It can be shown that in the case of progressive degradation, the Mahalanobis distance q_k, defined in Eq. (12), follows a Chi-squared distribution with n_y degrees of freedom.

$$q_k = \bar{\mathbf{r}}_k^{\mathrm{T}}\ \left(\mathbf{P}_{\bar{\mathbf{r}},k}\right)^{-1}\ \bar{\mathbf{r}}_k \tag{12}$$

where covariance matrix $\mathbf{P}_{\bar{\mathbf{r}},k}$ is assessed through Eq. (7) with $\mathbf{Q}_k = \mathbf{Q}_k^{min}$.

The scalar q_k is compared to a threshold η defined by specifying an acceptable misclassification probability P_F that is the probability of obtaining $q_k > \eta$ in the case of gradual deterioration. Typically, P_F is set to a low, but nonzero, value $(e.g.\ 10^{-6})$.

$$P_F = \int_{\eta}^{\infty} p(q_k)\ \mathrm{d}q_k. \tag{13}$$

If an abrupt fault is deemed to have occurred, the maximum a posteriori solution to Eq. (10) is used, otherwise the process noise $\mathbf{Q}_k^{min}$ related to gradual deterioration is selected. This logic is translated mathematically in Eq. (14).

$$\widehat{\mathbf{f}}_k = \begin{cases} \mathbf{f}^{min} + \max\left[0\ ,\ \left(\mathbf{P}_f^{-1} + \mathbf{B}_k^{\mathrm{T}}\ \mathbf{B}_k\right)^{-1}\ \mathbf{B}_k^{\mathrm{T}}\mathbf{d}_k\right] & \text{if } q_k > \eta \\ \mathbf{f}^{min} & \text{otherwise.} \end{cases} \tag{14}$$

3.2. Generalised likelihood ratio

The adaptive algorithm based on the generalised likelihood (GLR) technique is inspired by the work in [10]. It implements a GLR test in order to *detect and estimate* abrupt faults. The milestones of the technique are reported hereafter, the interested reader can find a detailed description in Reference [7].

The adaptive algorithm uses a modified transition model of the health parameters that accounts for possible abrupt faults

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \boldsymbol{\nu}_k + \boldsymbol{\Delta}_{\mathbf{w}}\ \delta_{\tau,k}. \tag{15}$$

The last term in (15) accounts for a possible "jump" in the health parameters:

- $\boldsymbol{\Delta}_{\mathbf{w}}$ is a vector modelling the jump,
- τ is a positive integer that represents its time of occurrence,
- $\delta_{i,j}$ is the Kronecker delta operator.

Note that $\boldsymbol{\Delta}_{\mathbf{w}}$ and τ are regarded here as *unknown* parameters and not as random variables, which means that no prior distribution is attached to them.

The strategy of adaptive estimation comes from viewing the new state-space model according to two different hypotheses:

- H_0: no jump up to now ($\tau > k$)
- H_1: a jump has already occurred ($\tau \leq k$).

Under assumption H_0, the Kalman filter provides an optimal estimation of the health parameters in the least-squares sense. Under assumption H_1, the residuals $\mathbf{r}_k$ are a function of the jump characteristics τ and $\boldsymbol{\Delta}_{\mathbf{w}}$. Given the linearisation of the measurement equation, the residuals $\mathbf{r}_k$ can be expressed as the sum of two terms

$$\mathbf{r}_k = \mathbf{r}_{k,H_0} + \mathbf{H}_{k,\tau}\ \boldsymbol{\Delta}_{\mathbf{w}} \tag{16}$$

where $\mathbf{r}_{k,H_0}$ are the residuals in the no-jump case, distributed as $\mathcal{N}(\mathbf{0}, \mathbf{P}_{\mathbf{y},k})$ and the second term represents the influence of a jump $\boldsymbol{\Delta}_{\mathbf{w}}$ that has occurred at time τ on the residuals at time k. The matrix $\mathbf{H}_{k,\tau}$ can be computed from the state-space model and the Kalman filter equations, see Reference [10] for further details.

In order to determine which hypothesis between H_0 and H_1 is true, a GLR test (see [11]) is applied. In short, it is a statistical test in which a ratio is computed between the maximum probability of a result under two different hypotheses, so that a decision can be made between them based on the value of this ratio. Unlike the classical likelihood ratio test, the generalised one does not require prior distributions for τ and $\boldsymbol{\Delta}_{\mathbf{w}}$ to be specified, but it provides estimates of these quantities. This is definitely an advantage in the present application.

Similar to the approach taken in the covariance matching scheme, the detection of the occurrence of an abrupt event is limited to a sliding window covering the previous M time steps. Essentially, the procedure consists first in computing the maximum likelihood estimates $\widehat{\tau}$ and $\widehat{\boldsymbol{\Delta}}_{\mathbf{w}}$ from the residuals $\mathbf{r}_{k-M}, \ldots, \mathbf{r}_k$ assuming H_1 is true. These values are then substituted into the usual LRT for H_1 versus H_0. Given that all probability densities are Gaussian, the log-likelihood ratio takes the form

$$l_{k,\tau} = \mathbf{d}_{k,\tau}^{\mathrm{T}}\ \mathbf{C}_{k,\tau}^{-1}\ \mathbf{d}_{k,\tau} \quad k - M < \tau \leq k \tag{17}$$

where matrix $\mathbf{C}_{k,\tau}$ is deterministic and *does not depend* on the data while vector $\mathbf{d}_{k,\tau}$ is a linear combination of the residuals

$$\mathbf{C}_{k,\tau} = \sum_{j=\tau}^{k} \mathbf{H}_{j,\tau}^{\mathrm{T}}\ \mathbf{P}_{\mathbf{y},j}^{-1}\ \mathbf{H}_{j,\tau} \quad \text{and} \quad \mathbf{d}_{k,\tau} = \sum_{j=\tau}^{k} \mathbf{H}_{j,\tau}^{\mathrm{T}}\ \mathbf{P}_{\mathbf{y},j}^{-1}\ \mathbf{r}_j. \tag{18}$$

These two equations show that the likelihood ratio (17) actually implements a matched filter *i.e.* a correlation test between the variations in the residuals and the signature of a jump, represented by $\mathbf{H}_{k,\tau}$.

The value $\widehat{\tau}$ that maximises $l_{k,\tau}$ represents the most likely time at which a jump occurred during the last M time steps. The decision rule to choose between H_0 and H_1 is

$$l_{k,\widehat{\tau}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{19}$$

The threshold η is directly related to the probability of false alarms P_F in jump detection through

$$P_F = \int_{\eta}^{\infty} p(l = L|H_0)\ \mathrm{d}L \tag{20}$$

where $p(l = L|H_0)$ is the probability density of $l_{k,\tau}$ conditioned on H_0 which is a central Chi-squared density with n_w degrees of freedom, see [10] for a proof. Specifying an allowable false alarm rate (small, but nonzero) gives the threshold value.

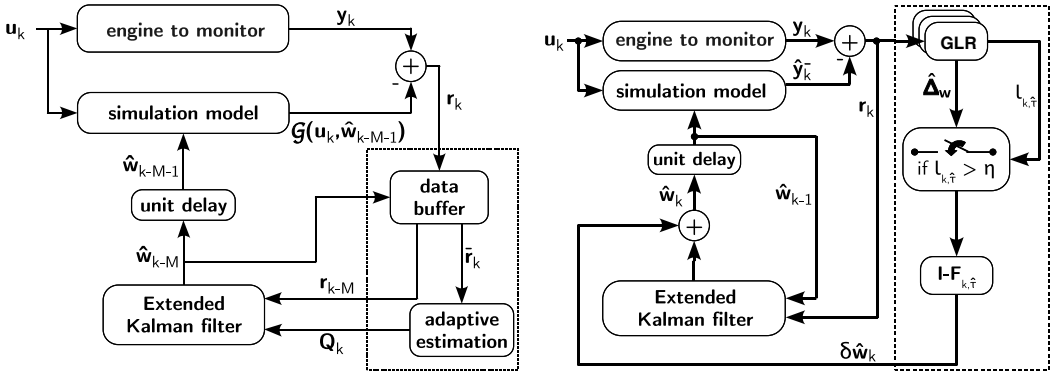
If hypothesis H_1 is verified at time step k, a jump has occurred at estimated time $\widehat{\tau}$ and its maximum likelihood estimate is given by

$$\widehat{\boldsymbol{\Delta}}_{\mathbf{w},k} = \mathbf{C}_{k,\widehat{\tau}}^{-1}\ \mathbf{d}_{k,\widehat{\tau}} \tag{21}$$

the latter relation provides a least-squares estimate of the jump $\boldsymbol{\Delta}_{\mathbf{w}}$ assuming that τ is known and that no prior information is available about the value of the jump. In that case, $\mathbf{C}_{k,\widehat{\tau}}^{-1}$ is the error covariance of the estimate $\widehat{\boldsymbol{\Delta}}_{\mathbf{w}}$.

Once a jump is detected by the GLR test, the maximum likelihood estimates $\widehat{\tau}$ and $\widehat{\boldsymbol{\Delta}}_{\mathbf{w},k}$ can be used directly to increment the parameters estimated by the Kalman filter

$$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_{k,KF} + \underbrace{\left(\mathbf{I} - \mathbf{F}_{k,\widehat{\tau}}\right)\ \widehat{\boldsymbol{\Delta}}_{\mathbf{w},k}}_{\delta\widehat{\mathbf{w}}_k} \tag{22}$$



(a) Covariance matching. (b) Generalised likelihood ratio.

Fig. 4. Block diagram of the adaptive algorithms.

where $\widehat{\mathbf{w}}_{k,KF}$ are the Kalman filter estimates, $\mathbf{I}\ \widehat{\boldsymbol{\Delta}}_{\mathbf{w},k}$ is the contribution to the parameters of a jump that occurs at $\widehat{\tau}$ and $\mathbf{F}_{k,\widehat{\tau}}\ \widehat{\boldsymbol{\Delta}}_{\mathbf{w},k}$ represents the response of the Kalman filter to the jump prior to its detection.

In some cases, the estimate of the jump may be inaccurate. To reflect this degradation in the quality of the estimate, caused by the jump, it is advised in [10] to increase the covariance matrix of the health parameters accordingly. This rise in parameter covariance results in an increased Kalman gain *i.e.* an increased bandwidth. The filter can improve its response to the jump and hence compensate for inaccuracies in $\widehat{\tau}$ and $\widehat{\boldsymbol{\Delta}}_{\mathbf{w},k}$. The covariance modification is done through

$$\mathbf{P}_{\mathbf{w},k} = \mathbf{P}_{\mathbf{w},k,KF} + \underbrace{\left(\mathbf{I} - \mathbf{F}_{k,\widehat{\tau}}\right)^{\mathrm{T}}\ \mathbf{C}_{k,\widehat{\tau}}^{-1}\left(\mathbf{I} - \mathbf{F}_{k,\widehat{\tau}}\right)}_{\delta\mathbf{P}_{\mathbf{w},k}}. \tag{23}$$

As mentioned in the description of the covariance matching scheme, the number of health parameters generally outweighs the number of measurements. As a consequence, the system is only partially observable and matrix $\mathbf{C}_{k,\tau}$ is singular. In that case, the pseudo-inverse usefully replaces the common inverse. The possible jump directions are then restricted to the observable subspace of the parameter space, see [12] for further details. Another possibility is to estimate the jump with a bayesian approach. The introduction of prior knowledge about the jump focuses the search for a solution in a neighbourhood of the a priori values. Note that this solution could also be used for fully observable systems in the case where some a priori information about the possible range of values for $\boldsymbol{\Delta}_{\mathbf{w}}$ is available.

3.3. Practical implementation of the adaptive algorithms

To complete the presentation of the adaptive filters, Fig. 4 sketches the integration of each adaptive component, which comprises all the elements in the dashed box, with the Kalman filter. For sake of clarity, only the most relevant data streams are sketched in the diagram. The left part of the figure is related to the implementation of the covariance matching scheme. The "data buffer" box stores the residuals and compute their averaged values $\bar{\mathbf{r}}_k$. The "adaptive estimation" box implements relation (14).

The right part of the figure depicts the implementation of the generalised likelihood technique. The "GLR" box updates the quantities $\mathbf{d}_{k,\tau}$ and $\mathbf{C}_{k,\tau}$ in the M-sized buffer. Note that recursive relations can be derived, see [10]. The likelihood ratio is assessed through Eq. (17) and its maximum value is searched for. The outputs of the GLR box are the estimated time of occurrence of the jump $\widehat{\tau}$ (not represented), the estimated jump $\widehat{\boldsymbol{\Delta}}_{\mathbf{w},k}$ and the maximum value of the likelihood ratio in the buffer $l_{k,\widehat{\tau}}$.

4. Application

4.1. Engine layout

The application used as a test case is a high bypass ratio, mixed-flow turbofan. The engine performance model has been developed in the frame of the OBIDICOTE[3] project and is detailed in [13]. A schematic of the engine is sketched in Fig. 5 where the location of the eleven health parameters and the station numbering are also indicated. Only one command variable, which is the fuel flow rate fed in the combustor, is considered here.

The sensor suite selected for diagnosing the engine condition is representative of the instrumentation available on-board contemporary turbofan engines and is detailed in Table 1 where the nominal accuracy of each sensor is also reported.

[3] A Brite/Euram project for On-Board Identification, Diagnosis and Control of Turbofan Engine.

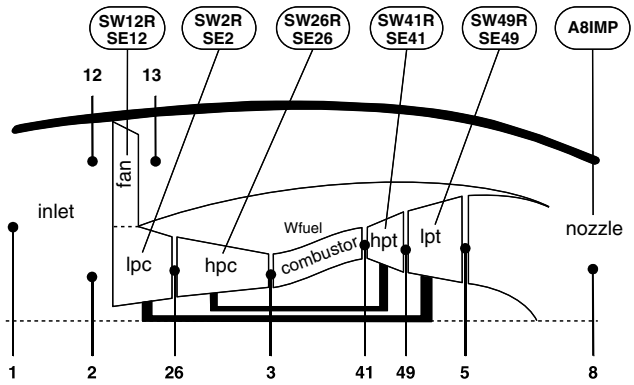


Fig. 5. Turbofan layout with station numbering and health parameters location.

Table 1
Selected sensor suite (uncertainty is three times the standard deviation).

Label	Uncertainty	Label	Uncertainty
p_{13}^0	±100 Pa	T_{13}^0	±2 K
p_3^0	±5000 Pa	T_3^0	±2 K
N_{lp}	±6 RPM	N_{hp}	±12 RPM
T_5^0	±2 K		

Table 2
List of abrupt component faults.

Label	Definition of the fault		Faulty component
a	−1% on SW12R	−0.5% on SE12	FAN
	−0.7% on SW2R	−0.4% on SE2	LPC
b	−1% on SE12		
c	−1% on SW26R	−0.7 % on SE26	HPC
d	−1% on SE26		
e	−1% on SW26R		
f	+1% on SW41R		HPT
g	−1% on SW41R	−1 % on SE41	
h	−1% on SE41		
i	−1% on SE49		LPT
j	−1% on SW49R	−0.4 % on SE49	
k	−1% on SW49R		
l	+1% on SW49R	−0.6 % on SE49	
m	+1% on A8IMP		Nozzle
n	−1% on A8IMP		

4.2. Definition of the test-cases

Simulated data have been generated to examine how both adaptive algorithms perform and to assess the improvements they can achieve with respect to the generic EKF. Cruise conditions (Alt = 10668m, Mach = 0.8, $\Delta T_{ISA} = 0$ K) are assumed. The flight sequence is 5000 s long and the sampling rate is set to 2 Hz. Gaussian noise, whose magnitude is specified in Table 1 is added to the clean simulated measurements to make them representative of real data. No sensor malfunction such as bias or drift is considered in the present study.

Engine wear due to normal operation is simulated by linearly drifting values of nearly all health parameters, starting from a healthy engine (all parameters at their nominal values) at $t = 0$ s and with the following degradation at the end of the sequence ($t = 5000$ s): −0.5 % on SW12R, −0.5% on SE12, −0.4% on SW2R, −0.5% on SE2, −1.0% on SW26R, −0.7% on SE26, +0.4% on SW41R, −0.8% on SE41, −0.5% on SE49. To demonstrate the improvements achieved with the adaptive algorithms, a number of fault cases, summarised in Table 2, are superimposed one at a time to the global performance deterioration. While far from being exhaustive, these cases are representative of typical accidental component faults that can be expected on turbofans and are added as a step change to engine wear at t = 2500 s. This library of degradations (both distributed and localised) has been devised in the frame of the OBIDICOTE project too, see [14] and has been used in a number of studies.

Table 3
Comparison of the figure of merit obtained with the generic and adaptive algorithms.

Case	EKF		EKF + CM		EKF + GLR	
wear (**w**)	0.09 %	✓	0.10 %	✓	0.10 %	✓
w + **a**	0.50 %	-	0.14 %	✓	0.12%	✓
w + **b**	0.27 %	-	0.08 %	✓	0.09 %	✓
w + **c**	0.20 %	✓	0.13 %	✓	0.11 %	✓
w + **d**	0.18 %	✓	0.11 %	✓	0.13 %	✓
w + **e**	0.32 %	-	0.15 %	✓	0.12 %	✓
w + **f**	0.21 %	✓	0.10 %	✓	0.09 %	✓
w + **g**	0.43 %	-	0.12 %	✓	0.11 %	✓
w + **h**	0.23 %	✓	0.12 %	✓	0.10 %	✓
w + **i**	0.20 %	✓	0.10 %	✓	0.09 %	✓
w + **j**	0.53 %	-	0.52 %	-	0.57 %	-
w + **k**	0.45 %	-	0.13 %	✓	0.09 %	✓
w + **l**	0.40 %	-	0.08 %	✓	0.09 %	✓
w + **m**	0.16 %	✓	0.13 %	✓	0.11 %	✓
w + **n**	0.34 %	-	0.14 %	✓	0.15 %	✓

4.3. Definition of a figure of merit

The quality of the estimation performed by the generic and adaptive diagnosis tools is assessed in terms of the maximum root mean square error (RMSE) over the last 50 s of the sequence (*i.e.* over 100 samples):

$$RMSE = \max\left(\sqrt{\frac{1}{100}\sum_{t=4950s}^{5000s}\left(\frac{\mathbf{w}_t - \widehat{\mathbf{w}}_t}{\mathbf{w}^{hl}}\right)^2}\right) \tag{24}$$

where $\mathbf{w}^{hl}$ are the nominal values of the health parameters.

Given the stochastic character of the measurement noise, each test-case has been run twenty times. The RMSEs reported in Table 3 are the average values over the twenty runs in order to guarantee that they are statistically representative. A test-case characterised by an averaged maximum RMSE below 0.25% is declared as successful which is indicated by a check mark. This threshold corresponds to roughly three times the standard deviation of the identified health parameters (*i.e.* the square root of the diagonal terms of the covariance matrix $\mathbf{P}_{\mathbf{w},k}$).

4.4. Results

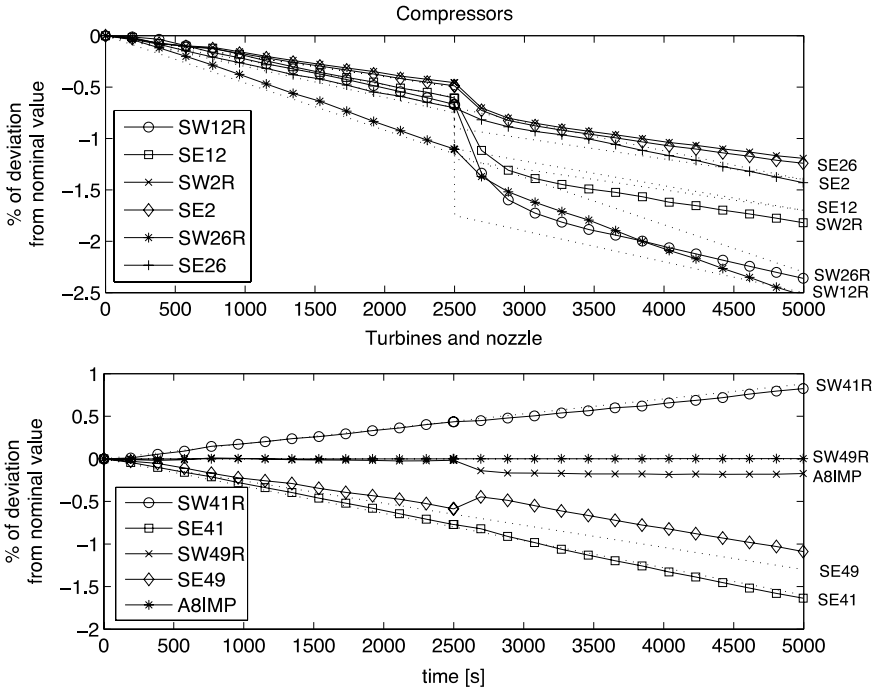
The generic and adaptive diagnosis tools have been run on the aforementioned test-cases. The tuning parameters for the covariance matching scheme are a buffer size $M = 50$ samples and a probability $P_F = 10^{-4}$%. For the GLR detector, the sliding window has a width of $M = 10$ time steps and a probability $P_F = 10^{-4}$% too. These settings were found satisfactory for the level of noise and the magnitude of the abrupt faults.

Table 3 reports the figure of merit defined by (24) for the different fault-cases that have been considered in the present study. The first line is related to the case of engine wear (abbreviated **w** in the subsequent lines). For this case of long-time-scale deterioration, the Kalman filter performs an accurate tracking of the engine condition, which is confirmed by a RMSE of 0.09%. It can also be seen that the adaptive tools have essentially the same performance as the Kalman filter for this test-case. Indeed, as long as the adaptive component does not issue any detection flag, the adaptive algorithm confounds with the generic one. This can clearly be seen from the schematics in Fig. 4.

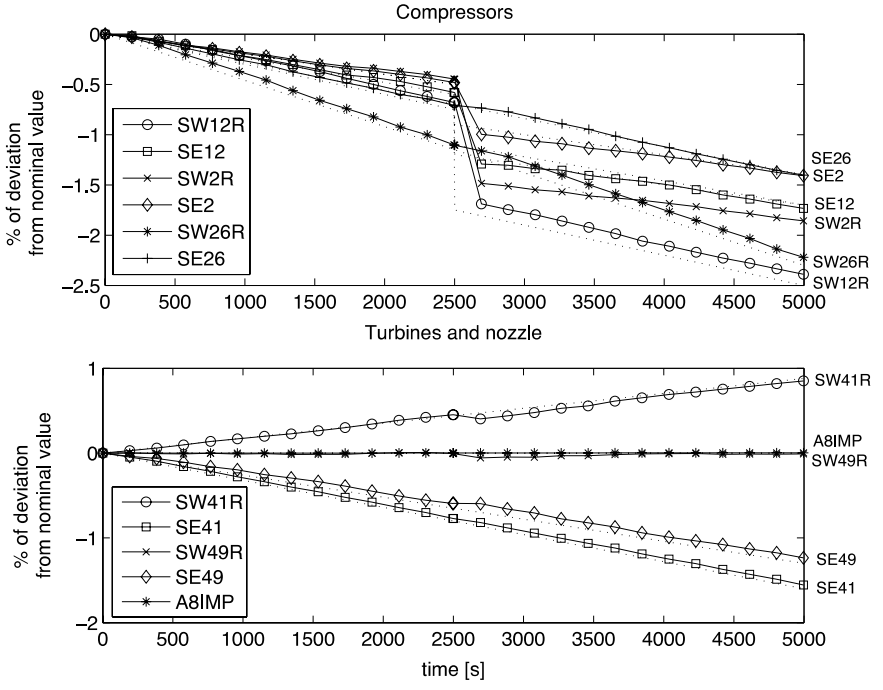
Considering the cases of mixing gradual deterioration and an abrupt component fault, it can be seen that the EKF achieves a reasonable identification for a mere 6 cases out of 14. Furthermore, the figure of merit is a measure of the estimation accuracy at the end of the sequence, but does not reflect the delay in fault recognition. On the other hand, both adaptive tools succeed in solving all test-cases but case **w** + **j**. The meaning of this statement is twofold: the gradual deterioration is effectively tracked and each abrupt fault (but the type **j** ones) is correctly detected and isolated. The figure of merit is about the same as for the case of pure engine wear, which hints at the capability of the adaptive tools to efficiently handle accidental events. For sake of completeness, the misdiagnosis of case **w** + **j** is due to a lack of observability of the health parameters of the turbines with the sensor configuration used here.

Fig. 6 depicts the identification of fault case **w** + **a**. The left graphs are obtained with the generic EKF, while the right ones result from the combination of the EKF and the GLR detector. It can readily be seen that the generic Kalman filter is unable to follow the abrupt change in the performance of the fan and the lpc. This fast variation is indeed not accounted for in the transition model of the health parameters that is blended in the EKF. As a result, the fault is spread over multiple components such as the hpc (SW26R) and the lpt (SE49). Moreover, several hundreds of seconds are needed to (erroneously) recognise the abrupt fault.

The processing of the same fault case by the EKF + GLR algorithm leads to obvious improvements. First, the abrupt fault is localised on the fan and the lpc and its magnitude is quite fairly assessed, the drop on SW2R being slightly overestimated. Second, the responsiveness for recognising the fault has dramatically improved. These two points greatly enhance the



(a) Kalman filter (EKF).



(b) Adaptive filter (EKF + GLR).

Fig. 6. Tracking of engine wear + fault 'a'.

relevance of the results and provide a detailed insight of the temporal evolution of the engine condition. Similar conclusions can be drawn for the adaptive algorithm featuring the covariance matching scheme.

5. Conclusion

In this contribution, two adaptive algorithms for engine health monitoring have been presented and compared. Both combine a Kalman filter, which provides accurate estimation of the health condition for long-time-scale deterioration

(such as engine wear), and an adaptive component which monitors the residuals and looks for abrupt changes in the health condition. On one hand, a covariance matching scheme performs an on-line tuning of the process noise variances. On the other hand, a generalised likelihood ratio test detects and estimates rapid changes in the engine condition. Interestingly, the present approach does not require the set-up of a pre-defined bank of accidental faults. The methodology could also be extended to handle system faults such as stuck bleed valves or mistuned variable stator vanes.

The improvements brought by the adaptive algorithms with respect to a generic Kalman filter have been illustrated on a turbofan application. The accurate estimation of abrupt faults achieved by the adaptive algorithms allows efficient performance monitoring and better component fault isolation. Moreover, the good tracking properties of the Kalman filter are maintained for slow evolutions of the engine's condition.

The implementation of the adaptive algorithms is quite straightforward and involves only basic matrix operations. The computational burden of the generalised likelihood ratio technique is slightly higher than that of the covariance matching scheme. Yet, the former provides a more detailed description about the abrupt fault ; indeed the estimated time of occurrence, the impacted component and the magnitude of the fault are reported while the latter only recognises the occurrence of an accidental event within a memory buffer and increases the covariance of the impacted component. With information fusion in mind, the generalised likelihood ratio approach seems therefore to be more promising.

References

[1] A.J. Volponi, Foundation of gas path analysis (part i and ii), in: von Karman Institute Lecture Series, in: Gas Turbine Condition Monitoring and Fault Diagnosis, vol. 01, 2003.
[2] Y.G. Li, Performance-analysis-based gas turbine diagnostics: A review, Proc. Inst. Mech. Eng. A 216 (5) (2002) 363–377.
[3] R.E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME D 82 (1960) 35–44.
[4] M.J. Provost, The use of optimal estimation techniques in the analysis of gas turbines, Ph.D. Thesis, Cranfield University, 1994.
[5] R.K. Mehra, Approaches to adaptive filtering, IEEE Trans. Automat. Control 17 (5) (1972) 693–698.
[6] O. Léonard, S. Borguet, P. Dewallef, An adaptive estimation algorithm for aircraft engine performance monitoring, AIAA J. Propulsion and Power 28 (4) (2008) 763–769.
[7] S. Borguet, O. Léonard, A generalised likelihood ratio test for adaptive gas engine turbine performance monitoring, ASME J. Eng. Gas Turbines Power 131 (1) (2009) paper ID 011601.
[8] P. Dewallef, Application of the Kalman filter to health monitoring of gas turbine engines: A sequential approach to robust diagnosis, Ph.D. Thesis, University of Liège, 2005.
[9] A.H. Jazwinski, Adaptive filtering, Automatica 5 (1970) 475–485.
[10] A. Willsky, H. Jones, A generalized likelihood ratio approach to state estimation in linear systems subject to abrupt changes, in: 1974 IEEE Conf. Decision and Control, 1974, pp. 846–853.
[11] H.L. van Trees, Detection, Estimation and Modulation Theory, John Wiley & Sons, New-York, 1968.
[12] S. Borguet, O. Léonard, The Fisher information matrix as a relevant tool for sensor selection in engine health monitoring, Internat. J. Rotating Machinery 2008 (2008) paper ID 784749.
[13] A. Stamatis, K. Mathioudakis, J. Ruiz, B. Curnock, Real-time engine model implementation for adaptive control and performance monitoring of large civil turbofans, in: ASME Turbo Expo, Number 2001-GT-0362, 2001.
[14] B. Curnock, Obidicote project — work package 4: Steady-state test cases. Technical Report DNS62433, Rolls-Royce PLc, 2000.