Int.J. Thermodynamics,
Vol.7, (No.2), pp.95-106, June-2004

ISSN 1301-9724

Hybrid Semi-Quantitative Monitoring and Diagnostics of Energy Conversion Processes

Enrico Sciubba
Mechanical Engineering Dept., University of Roma 1
Via Eudossiana18, 00184 Roma - Italy
E-mail: enrico.sciubba@uniroma1.it

Abstract

This paper presents a novel development in the field of automatic, "intelligent" process monitoring. It is possible to show that some of the limitations of a totally qualitative approach may be overcome by judicious use of reconciled experimental data and/or of the results of numerical simulations. The major problem in basing the response of a monitoring & diagnostic system on a process simulator is that, to run efficiently in real time, the simulator must introduce some simplification in the process model, and therefore its reliability as a source of "process data" is negatively affected. The approach proposed in this paper consists of adopting a mix (thence the attribute "hybrid" in the title) between reconciled data and physical modeling, to extract a limited number of numerical coefficients that introduce a sufficient degree of "quantification" in a qualitative monitoring system. The result is a fast and reliable intelligent procedure that assists human operators by presenting them a preliminary fault analysis based on a limited set of relevant reconciled process variables. An application to a regenerated gas turbine expansion process is discussed in detail.

Keywords: Diagnostics, prognostics, process health monitoring

1. Introduction

Diagnosis is a field of the research on energy systems devoted to the study of operation anomalies. These anomalies cause the actual performance of an energy system to differ from the expected (e.g. design) one, resulting in an increase of the amount of resources needed to obtain the same product, or, in more general terms, in a decrease of the overall efficiency.

Process monitoring is a logically and physically complex task. It consists of several subtasks, all of equal importance, and all connected with each other in such a way that if one fails, the whole monitoring task fails as well. It is not possible to describe a "unique procedure" on which an intelligent process monitor can be based, because in most cases, at least some of the process features are application dependent. In general, though, one can identify six main phases:

a) Data acquisition;

b) Data filtering and reconciliation;

c) Calculation of a set of proper "process health indicators";

d) Comparison of the calculated set with a reference set, and identification of the discrepancies;

e) Identification of all of the possible fault chains and alarm;

f) Implementation of a series of remedial actions.

(This last step has seen very few practical implementations, because it requires the fielding of an Expert Process Controller).

The first three steps, (a) through (c), do not pose particular problems: there are several accepted techniques for data reconciliation, some based on real-time acquisition of a limited set of "control data" (for instance, T and p in only m<N of the total *N* measuring stations), some on a comparison with a continuously running, on-line numerical process simulator. The set of indicators is obviously very knowledge-intensive: in addition to the knowledge of the Process Engineer, it must also contain sizeable portions of that of the Design Engineer and of the Plant Operator. But, once a certain set has been selected as "relevant" for a given process, it is uniformly used without modifications over the entire operational life of the plant. Points (d) and

(e) are problematic: first of all, because the discrepancies must be not only identified, but also interpreted (values below a certain threshold ought to be neglected, historical trends ought to be recorded, etc.), and second, because the induction from a series of "faulty indicators" of the actual (or possible) physical process faults is a difficult inverse problem.

Let us reconsider the usual diagnostic procedure, at its highest (most aggregated) logical level: we observe and control a set of physical quantities x_j (like pressure, temperature, mass flow rate, concentration, mechanical stress, etc.), but assess the performance of the process by means of a set of indicators I_k (efficiency, emission level, production cost, etc.): the relation between these indicators and the quantities subjected to our physical measures can be computed in many ways: the "best possible" solution would of course be that of having a process simulator running on-line and providing -in real time- numerical results that can be used both for the reconciliation of the measured data and for the fault-identification routine. This solution is though clearly unfeasible, except in very simple applications for which it would be in fact useless (because the fault identification routine is so simple that it can be performed without all the unnecessary complication). Thus, several degrees of simplification have been proposed. The variability of the methods adopted (or simply proposed) to solve $\mathbf{I}=f(\mathbf{X})$ is extremely large, because they are usually very much application-dependent: again, we can identify some major trends. We shall briefly describe (in Section 5 here below) three methods, respectively called the analytical, the quantitative-empirical, and the qualitative approach, to identify their strengths and limitations, before introducing the hybrid, semi-quantitative method object of this paper. To present the matter in a proper perspective, it is though convenient to first develop a formal representation of the "monitoring and diagnostic" activity.

2. Fault Identification

2.1 Definition of terms

We call a *measurable* any property pertaining to either a component of the process or to the (material or immaterial) flows that connect components to each other. Measurables are further divided in *primary* (or thermodynamical) if they express a direct material property (pressure, temperature, mass flow rate, concentration, stress level etc.), and *secondary* (or functional) if they express the value of an operating parameter of the component or flow (power, heat flux, etc.). A *signature* is an ordered set of measurables, each one of which has been attributed a numerical value. An *indicator* is a functional relationship between two or more measurables, usually expressed in the form of a ratio between "actual" and "design" values, which is defined by the Process or Design Engineer as a convenient performance quantifier. We speak of a *fault* when the value of an indicator is outside of an a priori defined "acceptable range": usually, this is due to one or more signatures being "deranged" with respect to their design values. A *fault chain* is the series of events (identified by the values of one or more measurables) that is known to lead to an unacceptable state of the process. A *failure point* is any operating point whose signature, for the given boundary conditions, lies outside of an a *priori* accepted operational range: it can be seen also as an attractor point in the process failure-state space (Angello, 2003).

2.2 The higher-order rules of failure detection

It is known from AI theory (Sriram, 1997, Sciubba and Melli, 1998), that it is convenient to re-organize, wherever possible, the knowledge bits acquired during the Knowledge Acquisition Phase, because such systematization goes in favour of the transparency and the accessibility of the "built-in-logic" of the Expert System. Such a reorganization is called Knowledge Clustering, and it has proven very effective in the implementation of an Intelligent Diagnostic & Prognostic Systems in the past (Sciubba and Melli, 1998, Biagetti and Sciubba 2002, Biagetti and Sciubba 2004). This means that the inference rules are hierarchically ordered in more than one "cluster". We adopt here the following clustering rules of failure detection:

1) There exist a finite number of possible failures, and for each one of them there exists at least one specific signature;

2) There are no sudden failures: every possible failure is "forewarned" by a drifting of the operational state of the plant along a path that leads to a failure point;

3) Each one of these "drifting" processes has a characteristic time scale that depends both on the component and on the type of failure;

4) A convenient way to represent such a drifting is that of employing a proper set of dimensionless indicators;

5) There exists at least one fault chain for each process of "failure formation". Each chain has at least two fuzzy aspects: first, the "causes" it contains are necessary, but not sufficient (for example, for a tube in the superheating section of a steam boiler to burst, it is necessary that the gas temperature at the location of the tube be higher than a certain design limit, but once the temperature exceeds this limit, not all tubes burst). Second, even this necessity is affected by

some degree of uncertainty (for example, a tube may burst even if the gas temperature is below the design limit);

6) Some of the fault chains may be concurrent. That is, the same failure may be originated by one or the other or by a combination of two (or more) fault chains;

7) Many of the fault signatures are non-local: the values of measurables detected at locations physically remote from the point where the failure actually takes place may be affected by the drifting process mentioned in point 2). In this case, we say that these measurables (and the indicators constructed on them) are correlated with the ones immediately affected by the failure.

2.3 Fault indicators

There are no general rules as to how to select the set of process indicators: often, those suggested by the International Industrial Standards do not suffice. Therefore, it is advisable to leave the selection of the indicators to the Process- and Design Engineers of the specific plant, together with the Plant Operator. One important feature is that the set must be complete, i.e., it must uniquely identify each relevant station in the process.

3. The General Diagnostic Procedure

From a critical analysis of several fault identification procedures adopted in present industrial practice, supported by a careful perusal of the most recent literature on this topic (Forsyth and Delaney, 2000, Gülen et al., 2000, Ozgur et al., 2000, Roemer and Kacprzynski, 2000, Tsalavoutas, 2000, Angello, 2003) a "general" task list may be derived on which an Automatic Diagnostic System ought to be based:

1) Identify in real time (in practice, at sufficiently small time intervals) the operational state of the Process. A state of the process is represented by a vector identified by an ordered set of N measurables;

2) Compare, at each time step, the detected operational state with the expected one. To do this, either a pre-determined operational schedule of the process, or a reliable Process Simulator, or both, must provide such an instantaneous "design datum";

3) If the value of the k-th measurable differs from the corresponding "design datum" by more than a preset tolerance, label this occurrence as a failure F, and activate a monitoring-and-control procedure on the component to which this measurable pertains;

4) Verify the presence of the "failure" F just detected in one of the "fault chains" contained in the Knowledge Base. If F belongs to a known fault chain, proceed to step (5) here below. If it does not, simply activate a sub-procedure to monitor k for a prescribed period of time (to see if it persists), and notify the (human) Plant Operator of this action[1];

5) If the event F="k-th measurable out of range" belongs to one or more known fault chains, launch a monitoring-and-control procedure on all measurables i,j,…z that appear together with k in the detected fault chains;

6) If a fault chain is indeed identified as "active": a- notify the Plant Operator; b- consult the Process Knowledge Base to search for remedial actions (e.g., adjustment of other process parameters to compensate for the derangement in k); c- decide whether it is possible to wait for the next scheduled maintenance intervention or if a repair/substitution is immediately necessary.

4. The Mathematical Problem Position

Define the "performance function" $\mathbf{\Pi}_P$ of an energy conversion process **P** as the deterministic mathematical relation between the instantaneous process output(s) and a set of N process parameters (the measurables): $\mathbf{\Pi}_P$ can be thought of as a (generally non-linear) operator that, applied to the vector $\mathbf{X}_N$ of measurables, generates the output vector $\mathbf{Y}_M$ [2] (the mapping is performed by $\mathbf{\Pi}_P$ on the quantity enclosed within the brackets $\langle . \rangle$).

$$\begin{aligned} \mathbf{X} &= (x_i) = (p_j, T_j, m_j, b_j, ..., x_N) \\ \mathbf{Y} &= (y_k) = (y_1, y_2, y_3, ..., y_M) \end{aligned} \tag{1}$$

$$\mathbf{Y} = \mathbf{\Pi}_P \langle \mathbf{X} \rangle \tag{2}$$

(Notice that $\mathbf{\Pi}_P$ must be an MxN matrix). The indicators (assume their total number is S) are generally simple analytical functions of some of the x_k and y_k and of their respective "design" values:

$$I_j = f_j(x_\ell, x_m, \ldots x_n, y_R, y_S, \ldots, y_T;\ x_{\ell o}, x_{mo}, \ldots\ x_{no;}\ y_{R,o}, y_{S,o}, \ldots, y_{T,o}) \tag{3}$$

So that we can symbolically write:

$$\mathbf{I} = \mathbf{A} \begin{vmatrix} \mathbf{X} \\ \mathbf{Y} \end{vmatrix} = \mathbf{AV} \tag{4}$$

where **A** is a (generally very sparse) Sx(N+M) matrix and **V** the augmented (**X**,**Y**) vector.

---

[1] This "fact" (i.e., "*k* must be monitored whenever $\mathcal{F}$ happens") can of course be added to the Knowledge Base.

[2] In energy conversion processes, the elements (y_k) are usually mechanical, electrical or thermal energy fluxes (measured in kW), or specific production costs (in €/kg, €/unit or €/kJ).

Now, denote as **V'** a deranged operational state, in which some of the measurables have taken values slightly (but detectably) different from their "design datum" ($x_{io}\ldots y_{jo}$ in equation 3). We can formally compute the new outputs by:

i. Approximating the new value assumed by the indicators:

$$\mathbf{I'} = \mathbf{I} + \mathbf{J}^1_{I,V}*(\mathbf{V}-\mathbf{V'}) + O(\mathbf{V}-\mathbf{V'})^2 \qquad (5)$$

where $\mathbf{J}^1_{I,V} = \left|\partial I/\partial V\right| = \left|\partial I_i/\partial v_j\right|$ is the -*S*x(*N*+*M*)-first-order Jacobian of the transformation $\mathbf{V}\Rightarrow\mathbf{I}$, and "O" means "of the order of".

ii. Applying the definition of **I'** provided by equation (5) and regrouping:

$$\Delta\mathbf{I}=\mathbf{I}-\mathbf{I'}\approx\mathbf{J}^1_{I,V}*(\mathbf{V}-\mathbf{V'}) = \mathbf{J}^1_{I,V}*\Delta\mathbf{V}\leq\Delta\mathbf{I}_{tol} \quad (6)$$

where $\Delta\mathbf{V}$ is measured[3], $\Delta\mathbf{I}$ is computed and $\Delta\mathbf{I}_{tol}$ is a preassigned maximum derangement vector: equation 6 must be solved iteratively, because of the non-linearity of **J**, and it provides us with the formal solution of the problem.

Such an exact mathematical approach is, however, not applicable in practice, because we do not know the exact form of **J** except in very few, *ideal*, cases, of no practical interest (the so-called "textbook cases"). Thus, we have to recur to iterative procedures (Sciubba and Melli, 1998, Biagetti and Sciubba 2002, Ozgur et al., 2000, Roemer and Kacprzynski, 2000, Tsalavoutas, 2000), that allow us, for a given set of tolerance intervals on each indicator I_j, to check, by using equations (5) and (6), whether the detected derangement in the augmented measurables v_k is acceptable or not, i.e., if it must be considered a failure or not. Notice that it would be more convenient, from the Operator's point of view, to solve the inverse problem instead: find the "acceptable" derangements $\|v'_k - v_k\|$ such that $\|I'_j - I_j\| \leq \Delta I_{j,tol}$: using the same notation as equation (5) above, the position now is:

$$(\mathbf{V}' - \mathbf{V}) = \mathbf{B}*(\mathbf{I}' - \mathbf{I}) \qquad (7)$$

Such a formulation is of course more direct, in that the Diagnostic System is only concerned with the augmented measurables, and the indicators are used in their proper sense, i.e., as "failure flags". Unfortunately, this inverse problem presents even more serious mathematical problems than the direct one: neither **A** in equation (4) nor **J** in equation (6) are square, and to solve them -to obtain **B** in equation (7)- by "best fit" or Singular Matrix methods requires caution because of their high non linearity. Thus, there is real incentive to recur to novel approaches to solve equation (7) "indirectly": in the next Section we shall examine four possible alternative methods. First, though, let us remark that **J** can be made square by adding (*N*+*M*-*S*) "dummy indicators" of the simple form $I_k=x_j$: these indicators shall not be used in the performance evaluation process, and are only needed to make **A** and **J** formally invertible[4]:

$$\mathbf{J}=\begin{vmatrix} \frac{\partial I_1}{\partial v_1} & \frac{\partial I_1}{\partial v_2} & \cdots & \cdots & \cdots & \frac{\partial I_1}{\partial v_S} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial I_S}{\partial v_1} & \frac{\partial I_S}{\partial v_2} & \cdots & \cdots & \cdots & \frac{\partial I_S}{\partial v_{N+M}} \\ \text{---} & \text{---} & \text{---} & \text{---} & \text{---} & \text{---} \\ \frac{\partial I_{S+1}}{\partial v_1}=0 & \frac{\partial I_{S+1}}{\partial v_2}=0 & \cdots & \frac{\partial I_{S+1}}{\partial v_{S+1}}=1 & \cdots & \frac{\partial I_{S+1}}{\partial v_{N+M}}=0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial I_{N+M}}{\partial v_1}=0 & \frac{\partial I_{N+M}}{\partial v_2}=0 & \cdots & \cdots & \cdots & \frac{\partial I_{N+M}}{\partial v_{N+M}}=1 \end{vmatrix}$$

In the following of this paper, therefore, we shall work with these "augmented forms" of both **A** and **J,** and speak of a formal inverse $\mathbf{B}=\mathbf{J}^{-1}$, solely for notational convenience: **J** will never be explicitly inverted.

5. Four Possible Operational Procedures to Solve the Inverse Problem Indirectly

Diagnostics is of course employed in all major industrial processes today, and even in the absence of a complete formal position of the problem, approximate methods are adopted in practice to detect possible failures. All of the current methods are direct, i.e., they proceed from the measurables to the indicators -as in equation (5)- and then use the normalised magnitude of the derangement of the indicators (taken individually or combined into "fault chains") to identify non-acceptable states. This first class of Diagnosers adopts the procedure defined by equations (4), (5) and (6) above, and their practical "intelligent" deployment is discussed for example in (Gülen et al., 2000, Roemer and Kacprzynski, 2000, Tsalavoutas, 2000, Biagetti and Sciubba 2002, Angello, 2003). It is interesting to remark that, from a logical point of view, all of the procedures presently adopted solve (fuzzily) in reality the inverse problem posed by equation (7):

[3] Actually, as it is clear from the previous discussion, it is the x_i that are measured, while the y_j are computed via equation (2).

[4] Note that this is done here simply to simplify the mathematical notation: it is clear that, except in very special cases of no practical usefulness, an "operator matrix" like **A** or **J** is not directly invertible (it may, of course, be inverted by means of iterative numerical procedures, but this is not of interest here).

▪ *Method I* – measure **X**, derive **Y** by using equation (2), construct the set of indicators **I** and bound $\left\| I'_j - I_j \right\|$ on the basis of proper phenomenological considerations. In operation, leave the fault identification to an Expert Operator who heuristically infers "dangerous derangements" of the indicators from the detected derangements of the measurables.

There are though at least three other approaches to the problem of identifying a fault: for broader generality, we shall identify them by their methodological approach, and classify them as "*analytical*", "*qualitative*" and "*quantitative-empirical*" methods respectively.

▪ *Method II* (analytical) – Measure **X**, derive **Y**, and use a simplified thermodynamic process analysis to derive formal expressions for **Π, A** and **J**. Obtain a small number of closed formulas that express some of the relations 7. Rely on Operator's expert judgment for the remaining indicators.

▪ *Method III* (qualitative) – Develop, from phenomenological reasoning, a qualitative Data Influence Matrix **DIM** (Biagetti and Sciubba, 2004) of the following form:

DIM =

Augmented measurables ⇒ ---------- Indicators ⇓	$\mathbf{v_1}$	$\mathbf{v_2}$	…..	$\mathbf{v_{N+M}}$
$\mathbf{I_1}$	↑	U	…	↓
$\mathbf{I_2}$	0	↓	…	↓
….	…	…	…	…
$\mathbf{I_S}$	U	↑	…	↓

Where the symbols indicate respectively whether I_j grows with (↑), decreases with (↓) or is indifferent (0) to a positive variation of v_k. The appearance of the symbol "U" in position j,k indicates that the variation of I_j with v_k is unknown a *priori*. From this "extended truth table" infer (on the basis of formal logical procedures, Sriram, 1997) a qualitative equivalent of equation (7). Notice that such a procedure is highly facilitated by the use of a proper "filter" to identify the respective interactions between measurables: also in qualitative terms, such a filter may be represented by a Measurable Correlation Matrix **MCM**:

MCM =

	$\mathbf{v_1}$	$\mathbf{v_2}$	…..	$\mathbf{v_{N+M}}$
$\mathbf{v_1}$	1	0	…	↓
$\mathbf{v_2}$	0	1	…	↑
….	…	…	1	…
$\mathbf{v_{N+M}}$	↓	↑	…	1

MCM (which is diagonally symmetrical) can be used to reinforce or weaken the "truth value" assigned by **DIM**. For instance:

a) The above **DIM** shows that (we believe that) I_2 decreases if either v_2 or v_{N+M} increase; but **MCM** tells us that v_2 grows with v_{N+M}, and therefore a state identified by a decreasing v_2 and an increasing v_{N+M} may be affected by measurement inconsistencies, and the "truth value" of the corresponding line of **DIM** is negatively affected;

b) The above **DIM** shows that I_1 grows with v_1 and decreases with v_{N+M}, and **MCM** tells us that v_1 decreases with v_{N+M}: therefore, a state identified by an increasing v_1 and a decreasing v_{N+M} is "self-consistent", and the "truth" of the corresponding line of **DIM** is reinforced.

▪ *Method IV* (quantitative-heuristic) - Measure **X**, derive **Y**, and run a process simulator to compute numerically the entries of **Π, A** and **J**. As in Method II, use the expert judgment of the Operator to infer possible unacceptable derangements in the v_k.

Each one of these methods has its advantages and shortcomings:

▪ Method I relies heavily on the alertness and experience of the human Operator, and it may be negatively affected by Operator cognitive overload (Sriram, 1997, Sciubba and Melli, 1998), bias or fixation ("if an event has always led to a certain consequence, it will do so also this time");

▪ Method II reduces this danger, by decreasing the extent of the decision space of the Operator. Often, though, the analytical relations are based on simplifications that are not verified in practical operation;

▪ Method III is of course useful only in qualitative sense, and needs a specially trained Operator to cope with its very complicated truth tables;

▪ Method IV, obviously more deterministic and precise from the point of view of the process calculation, is expensive, and often unfeasible in real-time.

A common trait of all four methods is that they leave the solution of the inverse problem to the human Operator. This is really the motivation of the present paper: to devise a diagnostic procedure that substantially reduces the need to recur to a direct human intervention in the "reasoning" that leads to fault detection. The idea of the Hybrid Semi-Quantitative method ("HSQM") proposed here is that of combining some of the features of Methods II and III above, leaving the "decisional" phase to an Expert System trained separately (and thus, not

necessarily in real-time) on a proper Process Simulator. The steps of the HSQM are as follows:

1) Consider that equation 7 can be rewritten in the form:

$$\Delta v_j = \sum_{k=1}^{S} b_{jk} \Delta I_k \quad \text{(for j= 1, 2 …N+M)} \quad (8)$$

2) For some of the influence coefficients b_{jk}, analytical expressions based on process thermodynamics may be found;

3) The remaining b_{jk} can be computed by a small number of extensive process simulations not necessarily conducted in real time: "guessed" operational curves can be fed to the Process Simulator in order to derive numerical values of the influence coefficients (some of these may be "best fit" values);

4) The b_{jk} may now be introduced *as facts* in the Knowledge Base of an Expert Diagnostic Assistant: the task of the ES is now only that of achieving the goal: $\|\Delta I_k\| < \varepsilon_k$, flagging all the occurrences Δv_j for which the goal is not attained as "derangements leading to possible failures". The Operator's task is thus limited to a higher level diagnostic (identifying all the possible fault chains and remedial actions), which could, in principle, be performed also by another Expert System (Sciubba and Melli, 1998, Roemer and Kacprzynski, 2000, Biagetti and Sciubba, 2004).

6. A Practical Application of HSQM

6.1 A simple expansion process in a turbine

Consider the expansion in a real gas turbine depicted in *Figure 1*. A hot pressurised gas at (p_1,T_1), flowing at a rate of m_g kg/s, undergoes an irreversible expansion, exiting the turbine at (p_2,T_2). Both the specific heat c_p and the mechanical efficiency η_m are assumed to remain constant throughout the expansion, and the process is exactly known at its "design point" (i.e., when p_1, p_2, T_1, T_2 and the turbine polytropic efficiency assume their respective design values). We want to apply the above-defined HSQM to define a possible monitoring procedure for this simple system[5].

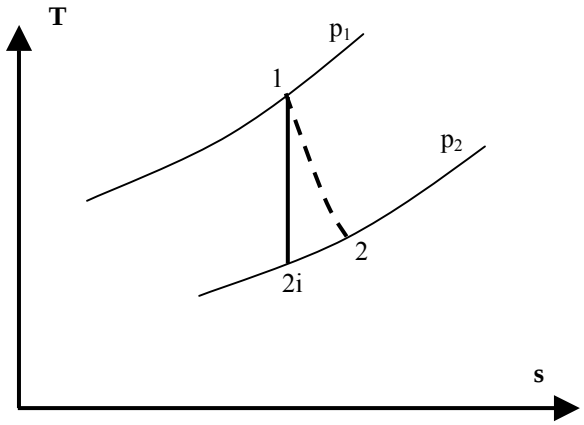

Figure 1. The conventional representation of the expansion of a real gas in a turbine.

6.2 The measurables

In current industrial practice, we are able to measure in real time the following process-related quantities:

- $\mathbf{X} = (p_1, p_2, T_2, m_g)$[6]
- $\mathbf{Y} = (P, T_1, \eta_t)$, where $P = m_g c_p (T_1 - T_2)$ and $\eta_t = (T_1 - T_2)/(T_1 - T_{2i})$[7]. The isentropic final temperature is defined as $T_{2i} = T_1 (p_2/p_1)^{\gamma}$, with $\gamma=(\kappa-1)/\kappa$, $\kappa=c_p/c_v$, and the turbine entry temperature is found from: $T_1 = T_2 (p_1/p_2)^{\alpha}$, with $\alpha=\gamma\eta_{pt}$. In addition, there is a non linear correlation between η_{pt} and $\eta_{t,real}$, as shown in *Figure 2*.

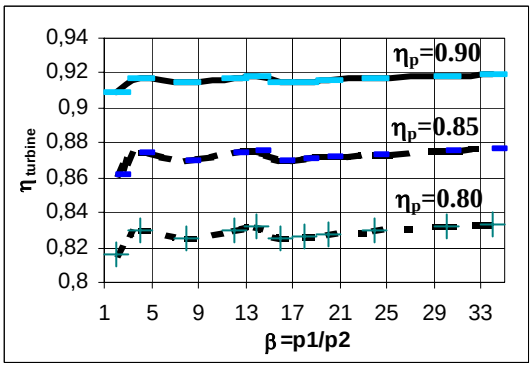

Figure 2. The functional relationship between the total, multi-stage ($\eta_{t,real}$) and the polytropic ($\eta_{p,t}$) efficiency for a gas turbine

$$\eta_{t,real} = \left(1 - \left[1 - \eta_{pt}\left(1 - \beta^{(1-\kappa)/n_{stages}\kappa}\right)\right]^{n_{stages}}\right) \Big/ \left(1 - \beta^{(1-\kappa)/\kappa}\right)$$

where $\eta_{pt} = \left(1 - \beta^{\eta_p (1-\kappa)/\kappa}\right) \Big/ \left(1 - \beta^{(1-\kappa)/\kappa}\right)$

Thus, with the notation adopted in Section 4 above, N=4, M=3, and Π is given by:

[5] Actually, the case of the expander, as presented here, belongs to the class we previously labelled as "textbook cases", in which analytical functions are known for all of the y and I. Such an application has been chosen here because its discussion helps putting the HSQM method into a clearer perspective.

[6] It is assumed here that T_1 cannot be directly measured, but is inferred once T_2 is known.

[7] This is the efficiency of a hypothetical single-stage turbine that performs the entire expansion: it is mentioned here because this is the quantity usually introduced in textbooks. If -as in this example- the expansion ratio is so high that $n_{stages}>1$ must be used, the correct turbine efficiency formula becomes the $\eta_{t,real}$ given in the legend of Figure 2.

$$\Pi = \begin{vmatrix} c_p m_g T_2 \left(\frac{\langle * \rangle}{p_2}\right)^{\alpha} & 0 & 0 & -c_p T_2 \langle * \rangle \\ 0 & T_2 \left(\frac{p_1}{\langle * \rangle}\right)^{\alpha} & 0 & 0 \\ \frac{1-\left(\frac{p_2}{\langle * \rangle}\right)^{\alpha}}{1-\left(\frac{p_2}{\langle * \rangle}\right)^{\gamma}} & 0 & 0 & 0 \end{vmatrix} \quad (9)$$

where the operational brackets indicate the quantity to which the operator defined by the corresponding matrix entry applies. There may be several equivalent forms for Π_P, and we shall adopt the one given above without further justification.

6.3 The Proposed Indicators and the Failure Detection Criteria

There are several possible choices for the set of performance indicators. Let us assume here the following four: the exergetic cost $c_w = \left(c_1 m_g (e_1 - e_2) + Z\right)/P$ of the generated power, the exergy $E_2 = m_g * e_2$ of the "cold" gases, and the dissipation temperatures θ_1 and θ_2 in point 1 and 2 respectively[8]. That is: I = (c_w, E_2, θ_1, θ_2). Thus, with the notation of Section 4, S=4. A convenient choice for the augmented form of A is thus:

$$\mathbf{A} = \begin{vmatrix} A_{11} & A_{12} & A_{13} & A_{14} & 0 & A_{16} & 0 \\ 0 & A_{22} & A_{23} & A_{24} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A_{36} & 0 \\ 0 & 0 & A_{43} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

With:

$$A_{11} = \frac{c_1 \gamma T_o}{T_1 - T_2} \ln\left(\frac{\langle \cdot \rangle}{p_2}\right) \quad (10a)$$

$$A_{12} = \frac{c_1 T_2}{T_1 - T_2} \left(\frac{p_1}{\langle \cdot \rangle}\right) \quad (10b)$$

$$A_{13} = -\frac{c_1 \langle \cdot \rangle}{T_1 - \langle \cdot \rangle} \quad (10c)$$

$$A_{14} = \frac{c_1 Z}{c_p \langle \cdot \rangle (T_1 - T_2)} \quad (10d)$$

$$A_{22} = c_p m_g T_o \gamma \ln\left(\frac{\langle \cdot \rangle}{p_o}\right) \quad (10e)$$

$$A_{23} = c_p m_g \left(\langle \cdot \rangle - T_o \ln\left(\frac{\langle \cdot \rangle}{T_o}\right)\right) \quad (10f)$$

$$A_{24} = -c_p T_o \quad (10g)$$

$$A_{36} = \frac{\langle \cdot \rangle - T_o}{\ln\left(\frac{\langle \cdot \rangle}{T_o}\right) - \gamma \ln\left(\frac{p_1}{p_o}\right)} \quad (10h)$$

$$A_{43} = \frac{\langle \cdot \rangle - T_o}{\ln\left(\frac{\langle \cdot \rangle}{T_o}\right) - \gamma \ln\left(\frac{p_2}{p_o}\right)} \quad (10i)$$

Let us further assume that industrial experience has determined that the following failures can happen:

1) abnormal increase or decrease of p_1 due to burner blockage;
2) abnormal increase or decrease of T_1 due to burner malfunctions;
3) abnormal increase of p_2 due to turbine malfunction;
4) abnormal increase of T_2 due to either turbine fouling or/and corresponding increase of T_1;
5) deterioration of η_{pt} due to turbine fouling, that leads to higher T_2 even for nominal values of p_1 and T_1.

The corresponding "deranged" operational points are shown in *Figure 3*. We shall not be concerned here with the procedure for inducing the actual fault from the detected derangements in the measurables: interested readers are referred to (Biagetti and Sciubba, 2002). Our goal is to identify the correct correlations between these derangements and the values attained by the four selected indicators.

[8] The dissipation temperature in a point R (h_R, s_R) is defined (Royo et al., 1997) as $\theta_R = h_R/s_R$.

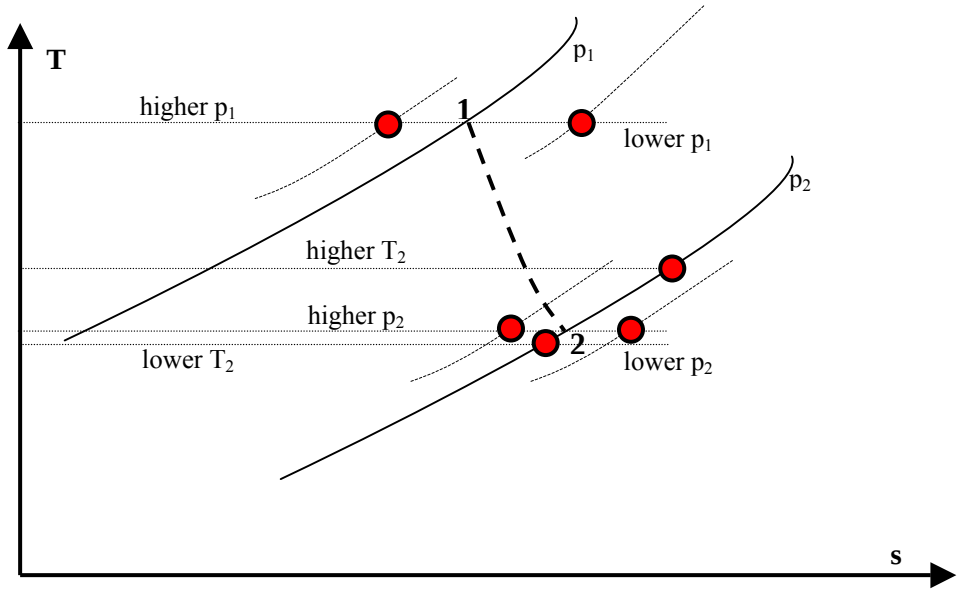

Figure 3. The possible derangements in the measurables considered in this study (extent of the variations exaggerated for clarity)

The augmented jacobian **J** can be written as follows:

$$\mathbf{J}=\begin{vmatrix} J_{11} & J_{12} & J_{13} & J_{14} & 0 & J_{16} & 0 \\ 0 & J_{22} & J_{23} & J_{24} & 0 & 0 & 0 \\ J_{31} & 0 & 0 & 0 & 0 & J_{36} & 0 \\ 0 & J_{42} & J_{43} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & J_{55} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & J_{66} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & J_{77} \end{vmatrix}$$

With[9]:

$$J_{11}=\frac{\partial I_1}{\partial p_1}=-\frac{c_1\gamma T_o}{T_1-T_2}\frac{p_2}{p_1} \qquad (11a)$$

$$J_{12}=\frac{\partial I_1}{\partial p_2}=\frac{c_1\gamma T_o}{(T_1-T_2)\,p_1p_2} \qquad (11b)$$

$$J_{13}=\frac{\partial I_1}{\partial T_2}= \left[\left(-2c_1(T_1-T_2)-c_1T_o\frac{T_1-T_2}{T_1T_2}-c_1T_o\ln\left(\frac{T_1}{T_2}\right)+c_1\gamma\,T_o\,\ln\left(\frac{p_1}{p_2}\right)-\frac{Z}{c_pm_g}\right)\Big/(T_1-T_2)^2\right] \qquad (11c)$$

$$J_{14}=\frac{\partial I_1}{\partial m_g}=-\frac{Z}{c_pm_g^2(T_1-T_2)} \qquad (11d)$$

$$J_{16}=\frac{\partial I_1}{\partial T_1}= \left[\left(-c_1T_oT_2\left(1-\frac{T_2}{T_1}\right)+c_1T_o\ln\left(\frac{T_1}{T_2}\right)-c_1\gamma T_o\ln\left(\frac{p_1}{p_2}\right)+\frac{Z}{c_pm_g}\right)\Big/(T_1-T_2)^2\right] \qquad (11e)$$

$$J_{22}=\frac{\partial I_2}{\partial p_2}=\frac{c_pm_gT_o\gamma}{p_2} \qquad (11f)$$

$$J_{23}=\frac{\partial I_2}{\partial T_2}=c_pm_g\left(1-\frac{T_o^2}{T_2}\right) \qquad (11g)$$

$$J_{24}=\frac{\partial I_2}{\partial m_g}= c_p\left(T_2-T_o-T_o\left(\ln\left(\frac{T_2}{T_o}\right)+\gamma\ln\left(\frac{p_2}{p_o}\right)\right)\right) \qquad (11h)$$

$$J_{31}=\frac{\partial I_3}{\partial p_1}=\frac{p_o(T_1-T_o)}{p_1\left(\ln\left(\frac{T_1}{T_o}\right)-\gamma\ln\left(\frac{p_1}{p_o}\right)\right)^2} \qquad (11i)$$

$$J_{36}=\frac{\partial I_3}{\partial T_1}=\frac{\ln\left(\frac{T_1}{T_o}\right)-(T_1-T_0)\frac{T_o}{T_1}}{\left(\ln\left(\frac{T_1}{T_o}\right)-\gamma\ln\left(\frac{p_1}{p_o}\right)\right)^2} \qquad (11j)$$

[9] Notice that the expressions provided here for J_{ij} do not correspond to the ones that one would obtain by deriving A_{ij} as given by the set of equations (10a-10i). The reason is that there is no unique way of rearranging the expressions for the indicators into a matricial form: here, simply, it was found that the form assumed for **A** was not convenient for **J**. Both forms are though "correct", in the sense that they both reproduce the mathematical formulae that define the indicators.

$$J_{42}=\frac{\partial I_4}{\partial p_2}=\frac{p_o(T_2-T_0)}{p_2\left(\ln\left(\frac{T_2}{T_o}\right)-\gamma\ln\left(\frac{p_2}{p_o}\right)\right)^2} \quad (11k)$$

$$J_{43}=\frac{\partial I_4}{\partial T_2}=\frac{\ln\left(\frac{T_2}{T_o}\right)-(T_2-T_0)\frac{T_o}{T_2}}{\left(\ln\left(\frac{T_2}{T_o}\right)-\gamma\ln\left(\frac{p_2}{p_o}\right)\right)^2} \quad (11l)$$

$$J_{55}=\frac{1}{P_d} \quad (11m)$$

$$J_{66}=\frac{1}{T_{1d}} \quad (11n)$$

$$J_{77}=\frac{1}{\eta_{td}} \quad (11o)$$

where c_1 (in €/kJ) is the (known) cost of the unit exergy of the incoming gas (at p_1,T_1).

6.4 A comparison of the HSQM and of the four "classical" methods described in Section 5

Assume that the following ΔI's have been identified as "dangerous" (the double bars indicate a "norm", like for example the rms, to which a sign can be attached):

$$||I_1'-I_1|| \geq 5\% \; ; \; ||I_2'-I_2|| \geq -5\% \; ;$$
$$||I_3'-I_3|| \geq 50\% \; ; \; ||I_4'-I_4|| \geq -50\%$$

The problem now is to identify the derangements in the augmented measurable that correspond to these maximum acceptable variations in the values of each indicator.

- *Method I* – As remarked above, the fault identification is left to the Expert Operator, who "knows" for example that "higher T_2 ∪ constant p_2 usually means higher T_1", and that "higher T_2 ∪ higher p_2 means turbine fouling": although in this simple case such an approach would suffice, it is apparent that more complex control tasks would pose enormous cognitive loads on Operators, increasing their "stress level" and the probability of errors in their judgements.

- *Method II* (Analytical) – This is what we have already done in Sections 6.2 and 6.3 above, when calculating the individual entries in **Π** and **J**. By looking at Π_{11} and Π_{22}, for instance, the Operator can infer that, if p_2 remains constant, a decrease in the net output power P is determined by a lower-than-normal value of p_1, which indicates burner malfunction..

- *Method III* (qualitative) – The qualitative **DIM** of the process under study is the following:

DIM =

Augmented measurable ⇒ ---------- Indicators ⇓	p_1	p_2	T_2	m_g	P	T_1	η_t
I_1 (c_w)	↑	↓	U	0	↑	↓	↓
I_2 (E_2)	0	↑	↓	↑	0	0	0
I_3 (θ_1)	↑	0	0	0	U	↑	0
I_4 (θ_2)	0	↑	↑	0	U	0	0

Recall that DIM_{jk} shows whether I_j grows with (↑), decreases with (↓), or is indifferent to (0) a positive variation of v_k. DIM_{jk}= U indicates that the variation of I_j with v_k is unknown a priori. From this "extended truth table" the Operator can infer a qualitative equivalent of equation (6). The qualitative filter defined in Section 5 by the influence matrix **MCM** takes the form:

MCM =

	p_1	p_2	T_2	m_g	P	T_1	η_t
p_1	1	U	U	↑	U	U	U
p_2	↑	1	U	↓	↓	U	↓
T_2	↓	↑	1	0	↓	↑	↓
m_g	↑	↓	U	1	↑	↑	↑
P	↑	↓	↓	↑	1	↑	↑
T_1	0	0	↑	0	↑	1	↑
η_t	↑	↓	↓	↑	↑	↑	1

In the present example, it is clear that, for instance, **MCM** reinforces the truth-values of DIM_{51} and weakens those of DIM_{42}.

▪ *Method IV* (quantitative-heuristic) – As stated above, this method actually provides the Operator with the (instantaneous) exact correlation between **I** and **V** (the exactness of the correlation depends of course on the accuracy of the Process Simulator). Again, while for a simple process as the one considered here, running a numerical simulator in real time would not pose any problem, for more complex processes this may be impossible to achieve in practice.

▪ *HSQM* (hybrid semi-quantitative) – In this approach, the entries of both $\mathbf{\Pi}$ and **J** are calculated exactly only for a limited number of "operational points" selected by the Process Engineer with the assistance of the Operator. These values provide an approximate (but very close to reality) estimate of the numerical correlation between the v_k and the I_j. To demonstrate the procedure, let us assume the following operational point: p_1=18 bar; p_2 = 1.1 bar; T_2 = 760 K; m_g = 250 kg/s; c_p = 1.2 kJ/(kg K); κ = 1.4; η_{pt} = 0.85; η_m = 0.9; Z=0.47 €/s; c_1 = 0.05 €/kJ. The entries of **J** take the following values[10]:

$$\mathbf{J}=\begin{vmatrix} -3.54\,10^{-4} & 2.93\,10^{-4} & -1.78\,10^{-8} & -8.52\,10^{-9} & 0 & -1.03\,10^{-2} & 0 \\ 0 & 2.34\,10^{4} & 1.82\,10^{2} & 2.28\,10^{2} & 0 & 0 & 0 \\ 1.09\,10^{2} & 0 & 0 & 0 & 0 & -3.9\,10^{2} & 0 \\ 0 & 5.13\,10^{2} & -3.66\,10^{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5.03\,10^{-6} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6.67\,10^{-4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.12 \end{vmatrix}$$

From equation (9) the reference values of the **Y** can be directly calculated:

$$y_1 = P_o = 198354 \text{ kW}$$
$$y_2 = T_{1o} = 1500 \text{ K}$$
$$y_3 = \eta_{to} = 0.898$$

For comparison, the exergy E_2 of the exhaust gases amounts to 56790 kW. By inspection, we see that (in our linearised assumption) a 5% variation in the exergetic cost of the generated power is "made up" of four terms related to the measurables:

$$(c_w'-c_{wd})_I = -3.54\,10^{-4}*(p_1'-p_1) \quad 10a)$$
$$(c_w'-c_{wd})_{II} = 2.93\,10^{-4}*(T_2'-T_2) \quad 10b)$$
$$(c_w'-c_{wd})_{III} = -1.78\,10^{-8}*(m_g'-m_g) \quad 10c)$$
$$(c_w'-c_{wd})_{IV} = -8.52\,10^{-9}*(T_1'-T_1) \quad 10d)$$

So that it is far more important to closely monitor the derangements in p_1 than those of p_2, T_2 and m_g: in fact, for all other measurables remaining constant, a Δp_1=+1.8 bar (a 10% variation) would cause an increase in c_w of 0.0006 €/kWh (1%), while, for symmetrical conditions, a ΔT_2=+76 K (an unlikely strong variation!) decreases the power exergetic cost c_w of about 2.2%. Similar considerations apply to the remaining indicators. *Figures 4a-4d* show the linearised behaviour of the system in the vicinity of the design point for what I_1 is concerned.

5. Conclusions

The new method proposed here for Process Monitoring is an approximate and semi-quantitative inverse method that consists in providing the plant Operator with instantaneous estimates of the relative order of magnitude of the correlation between the variations of the chosen performance indicators and the variations of the measurables. Formally, this is equivalent to computing the derivatives $\partial I_j/\partial v_k$ (or their finite equivalents, $\Delta I_j/\Delta v_k$) not at all instants in time, i.e., for each single point of the operating curve of the process, but at a series of properly selected representative points. The advantage in terms of computational intensity is substantial, and becomes a discriminating factor for complex processes, in which the implementation of a real-time simulator becomes unfeasible. The method has been presented here as an attempt to find a "direct" solution for the inverse monitoring problem, namely, that of deriving proper bounds not for the selected indicators, but directly for the measurable under monitoring. The solution presented here is not fully "automated": it is rather a convenient way of presenting the relevant Knowledge Base to the Plant Operator in an immediately understandable and quantitative form. Further work is underway to devise an Expert Process Operator that can automatically perform the inference that is here left for the human Operator. Such an Expert Process Operator (an example of which is presented in (Biagetti and Sciubba, 2004), but without the

[10] These are the linearised values in the immediate vicinity of the nominal operating point (i.e., the J_{ij} are calculated using the nominal values of the v_k, but retaining the non-linear expressions): this is obviously an approximation, as shown by equation (5)

HSQM feature discussed here) should in any case be implemented together with an "Expert Fault Identification System" capable of higher level reasoning, whose task would be that of linking the detected derangements to possible fault chains and to diagnose and possibly prognose failures.

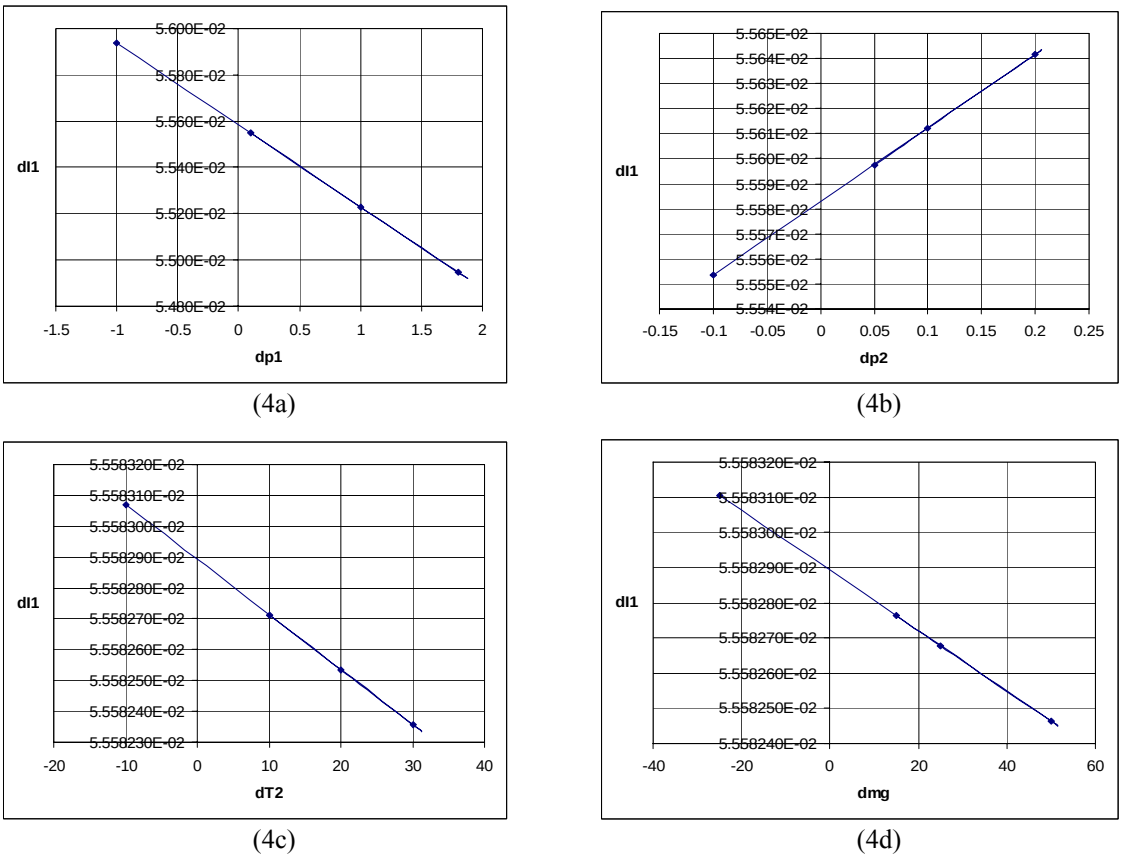

Figure 4. Linearised influence of p_1, p_2, T_2, m_g on c_w. The slope of each line represents $J_{jk} = \partial I_j / \partial v_k$. The variation of I_1 is measured in €¢/kWh; those of p_1, p_2, T_2 and m_g in bar, K and kg/s respectively.

List of Symbols

c_p	Specific Heat, kJ/(kgK)
c_w	Thermo-economic cost, €/kJ
E	Exergy flow, kW
m	Mass flow rate, kg/s
n_{stages}	Number of turbine stages
p	Pressure, Pa
T	Temperature, K
Z	Capital cost rate, €/s
β	Compression ratio
η_{pt}	Polytropic efficiency
η_t	Thermodynamic expansion efficiency
$\eta_{t,real}$	Real multi-stage turbine efficiency
γ	Isentropic exponent
κ	Specific heat ratio
θ	Dissipation temperature, K
$(\ )_d$	Design Conditions

References

Angello L., *Intelligent tools for combustion turbine health monitoring*, Proc. CAME-GT Conf., July 2003

Berleant D., Kujpers B., *Qualitative and quantitative simulation: bridging the gap*, AI Journ., n.95 (2), 1998

Biagetti T., Sciubba E., *PROMISE: a tool for the on-line intelligent performance prediction of cogeneration plants, Energy,* to appear, v.29, 2004

Biagetti T., Sciubba E., *A first step towards unmanned Intelligent Process Management: a procedure for the Diagnostics and Prognostics of Energy Conversion Plants,* IJAT, April 2002

Can S., Gülen et al., *Real time on-line performance diagnostics of heavy-duty industrial gas turbines*, ASME paper 2000-GT-312, 2000

Forsyth G., Delaney J., *Designing diagnostics Expert Systems for long-term supportability*, ASME paper 2000-GT-31, 2000

Kujpers B., *Qualitative Reasoning: Modelling And Simulation with incomplete knowledge*: MIT Press, Cambridge (MA), 1994

Kujpers B., *Qualitative Simulation*, in Meyers R.A., ed.: "Enc. of Physic. Sci. & Techn.", 3rd ed., Academic Press, NY 2001

Ozgur D. et al., *Remote monitoring and diagnostics system for GE heavy-duty gas turbines*, ASME paper 2000-GT-314, 2000

Roemer M.A., Kacprzynski G.J., *Advanced diagnostics and prognostics for gas turbine engine risk Assessment*, ASME paper 2000-GT-30, 2000

Royo J., Valero A., Zaleta A., *The dissipation temperature: a tool for the analysis of malfunctions in thermomnechanical systems*, En.Conv.Mgmt., v.38, n.15-17, 1997

Sciubba E., Melli R., *Artificial Intelligence in Thermal Systems Design: Concepts And Applications*, Nova Science, NY, 1998

Sriram R.D., *Intelligent Systems for Engineering*, Springer NY, 1997

Tsalavoutas A. et al.: *Combining advanced data analysis methods for the constitution of an integrated gas turbine condition monitoring and diagnostic system*, ASME paper 2000-GT-34, 2000