I.I. LOBODA

National Polytechnic Institute, School of Mechanical and Electrical Engineering, Mexico

A MORE REALISTIC PRESENTATION OF MEASUREMENT DEVIATION ERRORS IN GAS TURBINE DIAGNOSTIC ALGORITHMS

Gas path fault localization algorithms based on the pattern recognition theory are an important component of gas turbine monitoring systems. To simulate random measurement errors (noise) in description of fault classes, these algorithms usually involve theoretical random number distributions, like the Gaussian probability density function. A level of the simulated noise is determined on the basis of known information on typical maximum errors of different gas path sensors. However, not measurements themselves but their deviations from an engine baseline are input parameters for diagnostic algorithms. These deviations computed for real data have other error components in addition to simulated measurement inaccuracy. In this way, simulated and real deviation errors differ by an amplitude and distribution. Consequently, with such simulation, the performance of a diagnostic algorithm is poorly estimated, and therefore, the conclusion on algorithm efficiency may be wrong. To understand better noise peculiarities, plots of deviations of real measurements are tracked in the present paper. Additionally, possible deviation errors are surely analyzed analytically. To make noise presentation more realistic, it is proposed to extract random errors from real deviations and to integrate these errors in fault description. Finally, the effect of the new noise representation mode on gas turbine diagnosis reliability is estimated.

Key words: *gas turbine, gas path diagnosis, monitored variable deviation, deviation error.*

Introduction

Application of gas turbine health monitoring systems is a standard worldwide practice. In these systems, diagnostic algorithms based on gas path measured variables (temperature, pressure, rotation speed, fuel consumption, etc.) are considered as principle. Some measured variables are used to set an engine operation point and are called operating conditions. The rest of measured gas path variables are available for diagnostic analysis and are typically called monitored variables.

A total diagnostic process usually includes three principal stages of fault detection, fault identification, and prognostics [1]. They are preceded by an additional stage of measurement data validation and computing deviations. The deviation δY^* is calculated for a monitored variable Y as a relative discrepancy $\left(Y^* - Y_0\right)/Y_0$ between a measured value Y^* and a base-line value Y_0. In contrast to the monitored variables strongly depending on engine operating mode, the deviations, when properly computed, are almost free of the influence of the operating conditions and can be good indicators of engine gradual degradation or abrupt faults. The present paper deals with fault identification algorithms based on the pattern recognition theory. The described deviations are input parameters to these algorithms.

Monitoring systems' effectiveness obviously depends on accuracy of the diagnostic decisions made.

That is why, when a new algorithm is proposed, it is usually tailored and subjected to verification. In the corresponding investigations, operation of the proposed algorithm as well as a whole diagnostic process is simulated. To simulate gas path faults, a gas turbine model computes the monitored variables corresponding to the embedded faults.

The most of researchers also take into account random errors in the monitored variables and operating conditions applying the Gaussian distribution to that end. Such noise simulation has the following limitations. First, the level of simulated noise may differ from the level of random measurement errors that are peculiar to an analyzed gas turbine. Second, not monitored variables themselves but their deviations are input parameters for diagnostic algorithms, and, apart from measurement errors, deviations' errors include other uncertainty components. Third, an error distribution in the deviations based on real measurements is pretty irregular and differs a lot from theoretical distributions.

For a long period of time we have analyzed quality of recorded data and the deviation accuracy problem of a gas turbine power plant for natural gas pumping [2]. Possible error sources were examined and some algorithms were proposed to enhance the deviation quality.

The present paper focuses on more realistic noise representation. The same power plant has been chosen as a test case. Its nonlinear static model and field data recorded at steady states were employed in the investigations. To better understand types and sources of the

deviation errors, the paper looks at deviation graphs plotted for real measurements against power plant operation time. Additionally, the process of computing the deviations is analyzed analytically in order to clearly determine all error components and their nature. As a result of the analysis, it is proposed to draw a noise part from the deviations and integrate it into the description of simulated fault classes. Finally, such a novel mode to describe gas turbine faults is comprehensively discussed.

1. Common approach to a gas path fault recognition problem

For the purposes of diagnosis existing variety of engine faults should be broken down into a limited number of classes. The following hypothesis commonly used in the pattern recognition theory is also accepted in gas turbine diagnostics. It supposes that a system state D can belong only to one of q classes

$$D_1, D_2, ..., D_q \qquad (1)$$

that are set beforehand. As a rule, each fault class corresponds to one engine component.

As mentioned in the introduction, the deviations can potentially be good indicators of engine faults. That is why the deviations computed for m available monitored variables Y_i could form an appropriate space to recognize the faults. An additional operation of normalization $Z_i^* = \delta Y_i^* / a_{Yi}$ can further enhance the space. When a parameter a_{Yi} is a maximal random error of the deviation variable δY_i^*, maximal error amplitudes of all normalized deviation variables Z_i^*, $i = 1, m$ will be equal to one. Such normalization simplifies fault class description and enhances diagnosis reliability. On the basis of the above considerations, a vector $\vec{Z}^*$ that unites elemental variables of the normalized deviations is chosen to form a fault recognition space (diagnostic space). One value of the vector $\vec{Z}^*$ can be considered as a pattern to be recognized.

There are two scenarios to describe the fault classification in the space $\vec{Z}^*$; they can conditionally be called as probabilistic and statistical. The Bayesian approach exemplifies the first scenario [3]. It needs that each fault class D_j be described by its probability density function $f(\vec{Z}^* / D_j)$. The difficulty of this approach is related with the density functions themselves because it is a principal problem of mathematical statistics to assess them. That is why the first scenario can be realized only for a simplified fault classes.

In the second scenario the classes are given by samples of patterns namely vectors $\vec{Z}^*$. In this way, a whole fault classification is a union of pattern samples of all classes. Apart from the simplification of a class formation process, the replacement of the density functions by pattern samples allows creating more complex fault classes only on the basis of real data.

However, gas turbine faults are still often simulated mathematically because of rare appearance of real faults and high costs of physical fault simulation. Among different mathematical models used to simulate the faults, a so-called thermodynamic model can be considered as principal. This static nonlinear one-dimensional component-based model can be structurally presented as

$$\vec{Y} = F\left(\vec{U}_m, \vec{\Theta}\right). \qquad (2)$$

It computes the monitored variables as a function of steady state operating conditions (power set variables and ambient conditions) denoted by a (n×1)-vector $\vec{U}_m$, and engine health parameters $\vec{\Theta} = \vec{\Theta}_0 + \Delta\vec{\Theta}$. Nominal values $\vec{\Theta}_0$ correspond to an engine baseline. Changes $\Delta\vec{\Theta}$ called fault parameters provide some shifting of the performances of engine components (compressors, combustor, turbines, etc.) that results in the corresponding changes of monitored variables. Each fault class D_j is formed by growing values of its own vector $\Delta\vec{\Theta}_j$. Typically, all possible faults of one component are described by two its fault parameters, namely, a flow parameter ΔA and an efficiency parameter $\Delta\eta$.

In this way, the fault parameters embedded into the model allow simulating faults of variable severity for different components.

The normalized deviations induced by the fault parameter vector $\Delta\vec{\Theta}$ can be written as

$$Z_i = \frac{Y_i(\vec{U}_m, \vec{\Theta}_0 + \Delta\vec{\Theta}) - Y_i(\vec{U}_m, \vec{\Theta}_0)}{Y_i(\vec{U}_m, \vec{\Theta}_0) a_{Yi}}, i = 1, m. \qquad (3)$$

To take into consideration random deviation errors, a noise component ε_i should be added, thus resulting in

$$Z_i^* = Z_i + E_{Z_i}, i = 1, m. \qquad (4)$$

As mentioned before, amplitudes of all variables E_{Zi} are equal to one.

The deviations (4) form a vector $\vec{Z}^* = \vec{Z} + \vec{E}_Z$, which is a pattern to be recognized and an element to construct the classification (1). During the generating numerous patterns to represent the fault classes, a variable fault severity is usually determined by the uniform distribution and measurement errors are generated ac-

cording to the Gaussian distribution. A totality **Zl** of classification's patterns is typically called a learning set because it is applied to train or adjust the used recognition technique, for example, a neural network. In addition to the pattern observed and the fault classification accepted beforehand, the recognition technique is an integral part of a whole gas turbine diagnostic process.

Although the technique trained on the learning set data is ready for application, one more set is required to verify and validate it. The necessary set **Zv**, called a validation set, is created in the same way as the set **Zl**. The only difference is that other series of random numbers is generated to simulate fault severity and errors in the deviations. The technique makes a diagnosis for each pattern $\vec{Z}^*$ of the set **Zv**. A nomenclature of possible diagnoses $d_1, d_2, ..., d_q$ corresponds to the accepted classification (1). The diagnosis d_l may differ from a known class D_j due to pattern and classification random errors as well as inherent errors in the technique. Comparing the diagnoses and the classes for a great number of the validation set's patterns, we can compute diagnosis probabilities $Pd_{lj} = P(d_l / D_j)$ and compose a so-called confusion matrix **P**. Its diagonal elements P_{ll} form a vector $\vec{P}$ of true diagnosis probabilities that are indices of classes' distinguishability. A mean number of these elements – scalar $\overline{P}$ – determines total engine diagnosability. The described probabilities not only characterize the chosen recognition techniques, but they also are performances of the engine fault classification and a whole diagnostic process.

When the technique is adequate and well tailored, the diagnostic performance (diagnosis reliability) is mainly determined by the analyzed pattern and the classification. Since the deviation noise is a part of the pattern and classification, accuracy of the performance strongly depends on how realistic is noise simulation.

2. Deviations based on real data

In an effort to better understand peculiarities of the deviation noise, let us look at the deviations computed on the basis of real data. These data were recorded in field conditions for the gas turbine chosen as a test case, namely a power plant for natural gas pumping. It is an aeroderivative two shaft engine with a power turbine.

Figure 1 firstly presented in [4] helps to illustrate behavior of the deviation variable δY^*. This deviation was computed for an exhaust gas temperature (EGT) and is given here against power plant operation time t. With the values Y^* and $\vec{U}_m^*$ measured each hour, the deviations were computed according to an expression

$$\delta Y^* = \frac{Y^* - \hat{Y}_0(\vec{U}_m^*)}{\hat{Y}_0(\vec{U}_m^*)}. \quad (5)$$

In this figure a gray color curve means the deviation itself δY^* while the systematic influence of compressor fouling δY corresponds to a bold line with a maximum change designated as δ_0. In this way, a difference

$$E_{\delta Y} = \delta Y^* - \delta Y \quad (6)$$

can be interpreted as a deviation error.
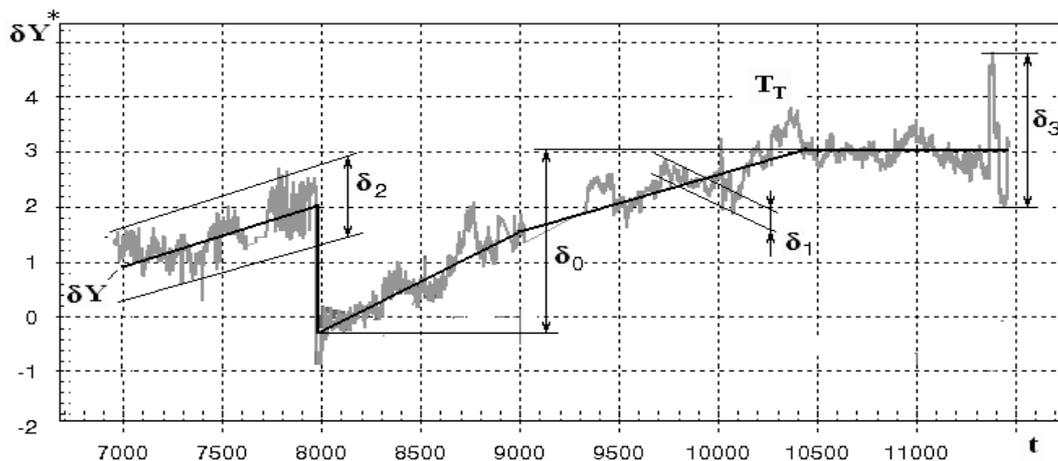
Fig. 1. Deviations plotted in % against the time of operation (hours)

A baseline functions $\hat{Y}_0(\vec{U}_m)$ is of a polynomial type. A vector $\vec{U}_m$ of functions' arguments comprises variables of ambient air pressure p_H^*, engine inlet temperature T_{in}^*, power turbine rotation speed n_{PT} and fuel consumption G_f. Unknown polynomial coefficients

were estimated by the least square method with healthy engine data (reference set).

The baseline functions were determined and the deviations were computed for all 6 gas path variables available for monitoring in the analyzed power plant. Table 1 contains the list of these variables with designa-

tions of the corresponding deviations and normalization parameters a_Y.

Table 1

Monitored variables

No	Variable's name	a_Y	Designations	
			Relative deviations	Normalized deviations
1	Compressor temperature T^*_C	0,00525	dTc	Z1
2	Exhaust gas temperature T^*_{HPT}	0,00453	dTt	Z2
3	Power turbine temperature T^*_{LPT}	0,00502	dTpt	Z3
4	Gas generator rotation speed n_{HP}	0,00347	dNhp	Z4
5	Compressor pressure p^*_C	0,00869	dPc	Z5
6	Exhaust gas pressure p^*_{HPT}	0,00775	dPt	Z6

The EGT deviation plotted in Fig.1 is a result of great efforts to enhance deviation quality. For instance, some cases of sensors' abnormal functioning were detected and the corresponding data were excluded from the analysis. The baseline functions were also optimized by choosing the best function type, arguments, and reference set to determine the function.

As a result of the optimization, the deviations have become good indices of engine deterioration. In Fig.1 we can clearly see two periods of EGT increase that is a result of compressor fouling, which is practically permanent and the most intensive deterioration mechanism of stationary gas turbines [5]. The periods are divided by a compressor washing in the time point t = 7970 hours.

Figure 1 also helps us to quantify quality of the deviations and specify deviation errors. The deviation quality can be expressed by a ratio $\delta_0 / \delta_\varepsilon$ (signal-to-noise ratio) of the maximum systematic change δ_0 to a spread δ_ε of deviation fluctuations.

According to a frequency and scatter, the fluctuations may be conditionally divided into three groups: 1) high frequency noise that is observed in every time point and has a scatter $\delta_{\varepsilon 1} < 0,3\%$; 2) slower fluctuations with the period of 30-300 hours and a scatter $\delta_{\varepsilon 2} < 1,5\%$; 3) single spikes with a scatter $\delta_{\varepsilon 3} > 1,5\%$. Since the spikes have the largest scatter, they can nearly always be detected, identified and excluded from the analyzed data. Generally, they are results of sensor malfunctions. To the contrary, the fluctuations $\delta_{\varepsilon 1}$ resulted from permanent measurement noise can not be removed. Being small, these fluctuations do not however considerably affect diagnosis accuracy. A main obstacle in the way to a correct diagnosis is related with the fluctuations $\delta_{\varepsilon 2}$. On the one hand, their effect is sufficiently great; on the other hand, it is often difficult to identify their origin. That is why these fluctuations can be mistaken for the effects of engine deterioration resulting in a misdiagnosis.

In addition to the graphical analysis conducted above, let us theoretically analyze possible causes and

sources of the deviation errors that can take place in practice. This will help to understand their behavior and to take them into account with higher accuracy.

3. Theoretical analysis of possible errors in real deviations

This analysis takes into consideration our previous studies on deviation accuracy [4,6] and is performed below on the basis of expression (5) used to compute the deviations in real conditions. Although the expression looks to be simple, the analysis will not be so trivial.

3.1. Error types

For a monitored variable Y, expression (5) can be rewritten as

$$\delta Y^* = \frac{Y^*}{\hat{Y}_0(\vec{U}^*_m)} - 1 . \qquad (7)$$

This equation shows that inaccuracy of the deviation is completely determined by errors in a term $Y^* / \hat{Y}_0(\vec{U}^*_m)$. It will be shown below that these errors can be divided into four types. One type is connected with a measured value Y^* and the other three types are related to a function $\hat{Y}_0(\vec{U}^*_m)$.

The measurement Y^* differs from a true value Y by an error E_Y called in this paper as a Type I error. In its turn, the true value depends on a vector $\vec{U}$ of real operating conditions and on engine health conditions given by the vector $\Delta\vec{\Theta}$. As a consequence, the value Y^* can be determined as

$$Y^* = Y(\vec{U}, \Delta\vec{\Theta}) + E_Y(\vec{U}, \Delta\vec{\Theta}) . \qquad (8)$$

The error E_Y is defined here as a function because, in general, measurement errors may depend on the value Y and, consequently, on the variables $\vec{U}$ and $\Delta\vec{\Theta}$.

One more obvious cause of the deviation inaccuracy is related with measurement errors in operating conditions presented in equation (7) by the vector $\vec{U}^*_m$. Given a vector of measurement errors $\vec{E}_{Um}$, which presents Type II errors, the measured operating conditions are written as

$$\vec{U}^*_m = \vec{U}_m + \vec{E}_{Um} . \qquad (9)$$

The next error type (Type III) is also related to engine operating conditions however it is not so evident. The point is that not all real operating conditions denominated in the present paper by a [(n+k)×1] – vector

$\vec{U}$ can be included as arguments of the baseline function. Some variables of real operating conditions are not always measured or recorded, for example, inlet air humidity, air bleeding and bypass valves' positions, and engine box temperature. Let us unite all these additional variables in a (k×1)-vector $\vec{E}_U$. Since such variables exert influence upon a real engine and its measured variable Y^* but are not taken into consideration in the baseline function $\widehat{Y}_0\left(\vec{U}_m^*\right)$, the corresponding deviation errors take place. A similar negative effect can occur if sensor systematic error changes in time.

Given that $\vec{U} = \vec{U}_m \cup \vec{E}_U$, the vector $\vec{U}_m$ can be given by $\vec{U}_m = \vec{U} \setminus \vec{E}_U$ and the equation (9) is converted to a form

$$\vec{U}_m^* = \vec{U} \setminus \vec{E}_U + \vec{E}_{Um} . \qquad (10)$$

Apart from the described errors related to the arguments of the function $\widehat{Y}_0(\vec{U}_m^*)$, the function has a proper error E_{Y_0} (Type IV error). It can result from such factors as a systematic error in measurements of the variable Y, inadequate function type, improper algorithm for estimating function's coefficient, errors in the reference set, limited volume of the set data, and influence of engine deterioration on these data. Given E_{Y_0} and a true function Y_0, the function estimation $\widehat{Y}_0$ can be written as

$$\widehat{Y}_0(\vec{U}_m^*) = Y_0(\vec{U}_m^*) + E_{Y_0}(\vec{U}_m^*) . \qquad (11)$$

3.2. Deviation formula

Let us now substitute equations (8), (10), and (11) into expression (7). As a result, the deviation δY^* is written as

$$\delta Y^* = \frac{Y(\vec{U}, \Delta\vec{\Theta}) + E_Y(\vec{U}_m^* - \vec{E}_{Um} + \vec{E}_U, \Delta\vec{\Theta})}{Y_0(\vec{U} \setminus \vec{E}_U + \vec{E}_{Um}) + E_{Y_0}(\vec{U}_m^*)} - 1 . \quad (12)$$

A dependency $E_Y(\vec{U}_m^* - \vec{E}_{Um} + \vec{E}_U, \Delta\vec{\Theta})$ in this expression can be simplified because of the following reasons:

a) $E_Y \ll Y$,

b) $\left\| \vec{E}_{Um} \right\| \ll \left\| \vec{U}_m^* \right\|$,

c) The influence of $\vec{E}_U$ and $\Delta\vec{\Theta}$ on Y and, consequently, on E_Y is significantly smaller then the influence of $\vec{U}_m^*$.

Taking into account the considerations made, we arrive to a final expression for the deviation

$$\delta Y^* = \frac{Y(\vec{U}, \Delta\vec{\Theta}) + E_Y(\vec{U}_m^*)}{Y_0(\vec{U} \setminus \vec{E}_U + \vec{E}_{Um}) + E_{Y_0}(\vec{U}_m^*)} - 1 . \quad (13)$$

This expression includes four error types introduced above, namely $E_Y, \vec{E}_{Um}, \vec{E}_U$, and E_{Y_0}. Let us now analyze how each error can influence on inaccuracy of the deviation δY^*.

3.3. Influence of different error types

The influence of different error sources on the deviations are analyzed in the sequel under the following assumptions commonly applied in gas turbine diagnostics. First, the same sensors were employed to measure currently analyzed values Y^* and $\vec{U}_m^*$ as well as the reference set data. Second, gross errors (e.g. spikes) have been filtered out. Third, a systematic error and amplitude of random errors in Y^* and $\vec{U}_m^*$ do not depend on engine operating time.

Type I error. Since the sensor performance is invariable, every systematic change of the error E_Y will be accompanied by the same change in E_{Y_0}. As a consequence, accuracy of the deviation δY^* will not be affected by the systematic component of E_Y. As to the random component, it is usually given by the Gaussian distribution. It is also believed that random errors of different variables Y are independent and are described by the multidimensional Gaussian distribution. That is why, the corresponding errors in the deviations δY^* of these variables can also be described by this distribution.

Type II errors. Errors $\vec{E}_{Um}$ can be analyzed in the same way as the Type I errors, separately for systematic and random components. Obviously, the objective of a function determination method (e.g. least square method) is to minimize the distance between the reference set data and function outputs. That is why, a baseline function will correctly describe reference data regardless of the systematic errors in function arguments (systematic component of the error $\vec{E}_{Um}$). Since the systematic error component is the same in the reference set and in a currently measured argument vector $\vec{U}_m^*$, a function output $\widehat{Y}_0(\vec{U}_m^*)$ will be adequate to a measured value Y^*. In this way, the system component of the errors $\vec{E}_{Um}$ cannot influence a lot the deviation δY^*.

As to the random component, it can be described by the multidimensional Gaussian distribution, as in the

case of the monitored variables Y. Because every change of the arguments $\vec{U}_m^*$ has an influence on baseline values of all monitored variables, their baseline values $\hat{Y}_0$ and, consequently, deviations δY^* may have correlation. Thus, independent random errors of measured operating conditions can induce correlated deviation errors that cannot be described by the multidimensional Gaussian distribution.

It is very likely that the noise with a scatter $\delta_{\epsilon 1}$ observed in Fig.1 results from a random component of the errors of Type I and Type II.

Type III error. Presence of such an error has been confirmed after analyzing all other error types. This error occurs because the additional operating conditions $\vec{E}_U$ do not change baseline function but exert influence on a real engine and, accordingly, on all variables Y. For this reason, any change of $\vec{E}_U$ can induce synchronous errors of the deviations δY^* of all monitored variables. It is very likely that most fluctuations with the scatter $\delta_{\epsilon 2}$ (see Fig.1) origin from the Type III errors.

Type IV error. The issue of the baseline function non-adequacy {error $E_{Y0}(\vec{U}_m^*)$} is a particular case of a well studied mathematical problem of the function estimation with empirical data [7]. This error varies in time along with changes in the operating conditions $\vec{U}_m^*$ producing perturbations in the deviation variable δY^*. These perturbation can be both independent and correlated depending on particular causes of the error E_{Y0}. Although the baseline function adequacy is a challenge, the error can be reduced to an acceptable level by applying a proper function type and using a representative reference set.

A deviation plot in Fig.1 is a result of multiple attempts to enhance deviation quality. The achieved deviation accuracy is not inferior to the level known from the literature and is sufficient for reliable monitoring of the power plant under analysis. Thus, we can conclude that Fig. 1 gives an example of deviation errors expected in a real situation. Therefore, to obtain realistic results of gas turbine diagnosis, simulated noise should be as close as possible to such real errors. This is verified below by comparing different schemes to represent deviation errors.

4. Noise representation schemes

4.1. Real error extraction

To extract an error component from the deviations based on real data, a model $Y(\vec{U}_m, \overline{t})$ of an degraded

power plant has been firstly determined as shown in [4, 6]. In addition to the operating conditions $\vec{U}_m$, the monitored variable Y depends in this model on engine operation time after the last washing $\overline{t}$. Model's coefficient were computed by the least square method with the reference set that includes the first 2500 operating points presented in Fig. 1. A baseline model $Y_0(\vec{U}_m)$ was then simply determined by putting $\overline{t}$ equal to zero.

With the described model and equation (6), a relative deviation error $E_{\delta Y}$ is written as

$$E_{\delta Y} = \frac{Y^* - Y_0(\vec{U}_m)}{Y_0(\vec{U}_m)} - \frac{Y(\vec{U}_m, \overline{t}) - Y_0(\vec{U}_m)}{Y_0(\vec{U}_m)} =$$

$$= \frac{Y^* - Y(\vec{U}_m, \overline{t})}{Y_0(\vec{U}_m)}. \qquad (14)$$

The errors $E_{\delta Y}$ of all monitored variables were computed for the 2500 points of the reference set as well as for 1400 subsequent operating points of an additional sample called a testing set. Plots of Fig. 2 illustrate the relative errors $E_{\delta Y}$ of the reference set. With these errors the normalization parameters were estimated for each monitored variable according to an expression $a_Y = 3\sigma_E$, where σ_E denotes a standard deviation of the variable $E_{\delta Y}$. The resulting values a_Y are given in Table 1.
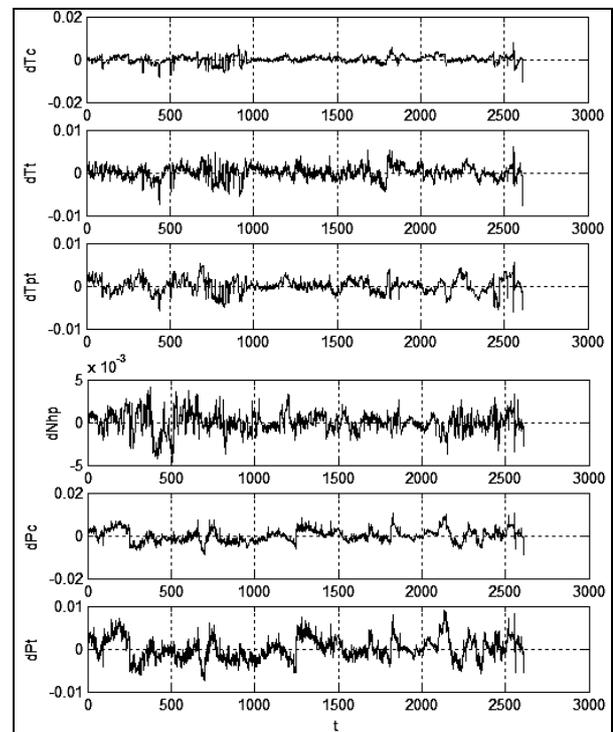
Fig. 2. Relative errors computed
for the reference set

Three error representation schemes are realized and examined below. Since a diagnostic space is formed by the normalized deviations Z^* [see equation (4)], the corresponding normalized errors E_Z are considered in all schemes. These errors, both real and simulated, were computed with the same normalization parameters of Table 1.

4.2. Scheme A: sensor error simulation

This scheme is the most widely applied in gas turbine fault recognition algorithms. The errors of each measured quantity, monitored variable Y or operating condition U, are usually given by the normal distribution. To simulate these errors (errors of Type I and Type II), we used the standard deviations σ of sensor uncertainties given in Table 2. These parameters were chosen in our previous work [8] on the basis of multiple literature sources. The influence of errors of the operating conditions on the monitored variables was estimated with the thermodynamic model described in section 1.

Table 2

Measurements uncertainties (σ,%)

p^*_H	T^*_{in}	n_{PT}	G_f	T^*_C	T^*_{HPT}	T^*_{LPT}	n_{HP}	p^*_C	p^*_{HPT}
0,03	0,2	0,1	0,5	0,2	0,25	0,2	0,05	0,2	0,3

Figures 3 and 4, a illustrate the considered schemes. It is clearly seen in Fig.4a that the presented deviation errors (deviations of exhaust gas temperature and power turbine temperature) have correlation. It is also visible that the error span considerably exceeds the interval (-1,0; 1,0), i.e. the deviation errors induced by the simulated sensor noise are more dispersed than the real errors computed for the reference set data.
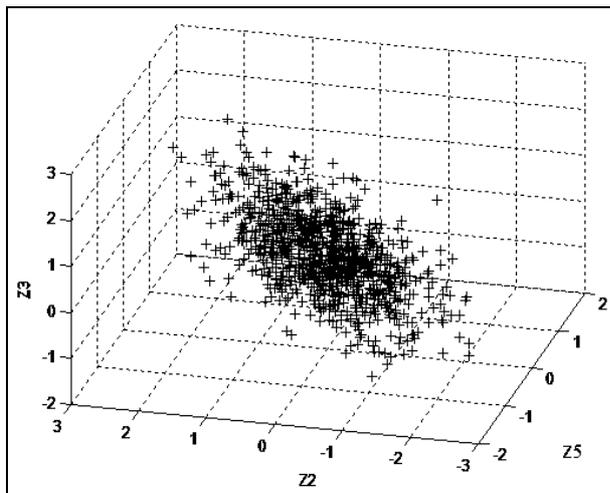
Fig. 3. 3D plot of the normalized deviation errors according to Scheme A (deviation designations Z5, Z2 and Z3 correspond to Table 1)

4.3. Scheme B: direct simulation of the deviation errors

This scheme was applied to simulate fault classes in our previous works (see, for instance, [9]). The deviation errors are given by the multidimensional normal distribution. The same standard deviations that were obtained for real noise are chosen. This allows exact adjustment of simulated errors to real ones.

This scheme is illustrated by Fig.4b. As it was expected, practically all simulated normalized errors are distributed inside the intervals (-1,0; 1,0) and no correlation is observed. The latter can be considered as a disadvantage because the correlation produced by Type II errors and expected in real deviations is absent.

4.4. Scheme C: errors of real deviations

This scheme is proposed and it means the integration of the normalized deviation errors computed with real data in the description of simulated faults. The scheme was realized separately for the cases of the reference and testing sets. The corresponding deviation errors are illustrated by Fig. 4, c and Fig. 4, d. As expected, the errors computed for the reference set (Fig. 4, c) are mostly localized inside the intervals (–1,0; 1,0) while the errors of the training set have significantly wider dispersion. Both figures show visible error correlation between the presented deviations. It also can be seen that the distribution of real errors, especially for the case of the testing set, is les regular than the simulated error distributions presented in Fig. 4, a and Fig. 4, b.

In this way, we can conclude that simulated deviation errors can differ a lot from real errors. Consequently, this can affect the accuracy of estimated indicators of gas turbine diagnosis reliability.

4.5. Influence of different noise representation schemes on diagnosis reliability: first results

With three described above schemes of noise representation, three corresponding fault classifications have been formed for the analyzed power plant. Namely, three variations of the learning and validation sets were created. Each classification includes 9 classes and each class is simulated by the gradual change of the corresponding fault parameter in the thermodynamic model. Four such classes are shown in Fig.5 in the space of three normalized deviations. The deviation errors correspond to scheme A.

Multilayer perceptron, the most widely used network, was chosen to recognize the faults. It was trained consequently with each variation of learning data. The probabilities of true diagnosis $\vec{P}$ and $\overline{P}$ (see section 1) have been computed by applying this net-

work to the corresponding variation of validation data.

Preliminary calculations have shown that the distinguishability of fault classes can change by up to 6% when real errors are replaced by simulated errors. Thus, the diagnostic performance estimated with simulated noise can be inaccurate.
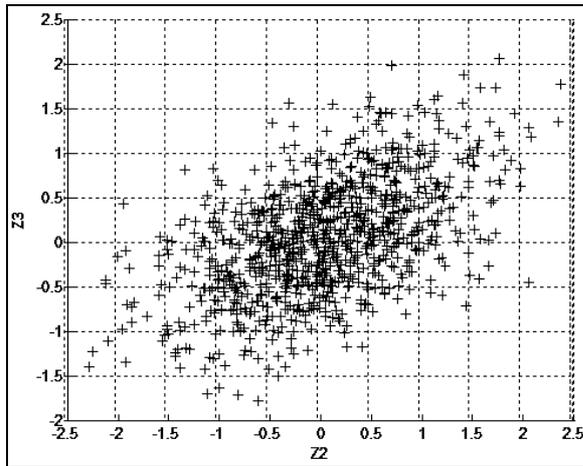
a – Scheme A: sensor error simulation
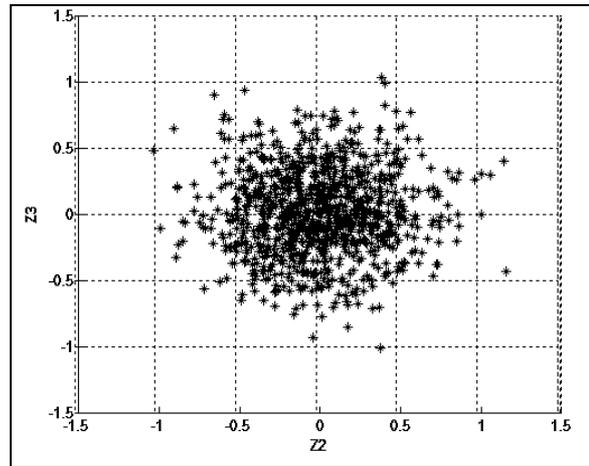
b – Scheme B: deviation error simulation
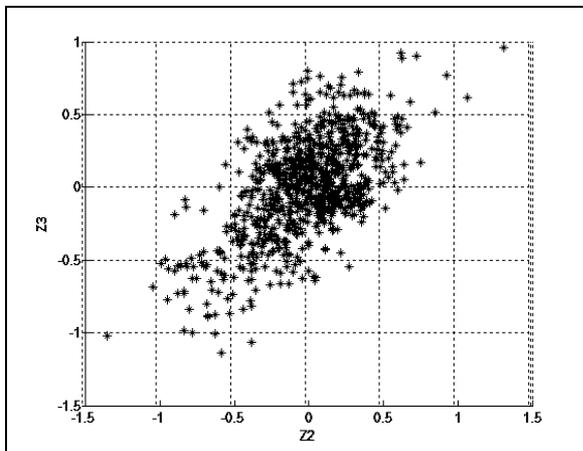
c – Scheme C: real deviation errors (reference set)
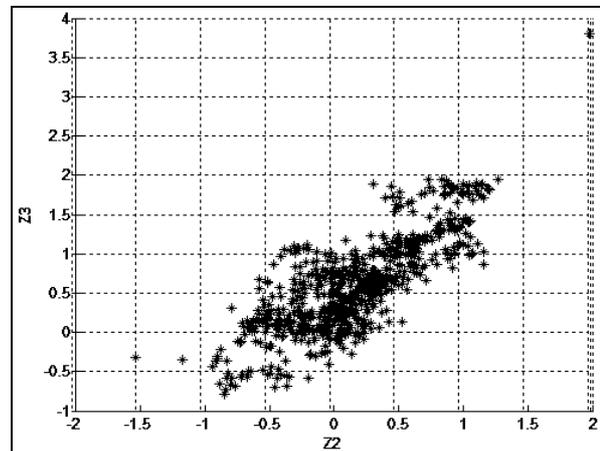
d – Scheme C: real deviation errors (testing set)

Fig. 4. 2D plots for different schemes of deviation error representation (deviation designations Z2 and Z3 correspond to Table 1)
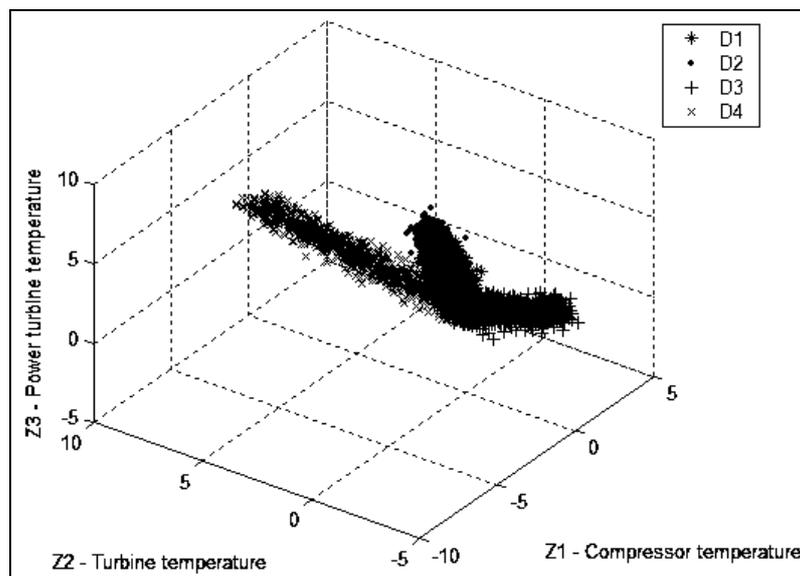
Fig. 5. 3D plot of fault classes in the space of normalized deviations (Scheme A)

The case was also investigated of learning data with reference set errors (see Fig.4c) and validation data with testing set errors (Fig.4d). Since real errors obtained from the testing set are more dispersed, we expected some degradation of the power plant diagnosability. The degradation was found drastic: from $\overline{P}$ = 90% - 94% in the previous cases to $\overline{P}$ = 59%. It happened because the model of degraded engine determined on the reference set has lost its accuracy on the testing set. Such a problem seems to be very probable in real diagnosis and we should be careful to avoid or mitigate it.

Conclusions

Thus, possible errors in deviations of gas turbine monitored variables have been analyzed in this paper. The problem of deviation accuracy is important because no monitored variables themselves but their deviations are input parameters in diagnostic algorithms.

A power plant for natural gas pumping has been chosen as a test case. It was presented in the present study by its nonlinear thermodynamic model and the data recorded under field conditions.

Possible deviation errors have been investigated theoretically and graphically. All error sources were thoroughly examined and classified into four types. We succeeded in finding a single mathematical expression to relate the deviation with its typical errors.

Three alternative schemes, two existing and one new, of deviation error representation in diagnostic algorithms have been realized. They were compared with the use of graphical means and probabilities of correct diagnosis. Preliminary results show that the existing schemes of error simulation do not always ensure the necessary accuracy of estimated engine diagnosability. The new scheme enhances the accuracy by including the noise component obtained from real data into the description of fault classes.

Although the proposed scheme is more realistic, it cannot automatically replace existing noise simulation modes. This scheme is more complex for realization. Additionally, it needs both the thermodynamic model and extensive real data, two things rarely available together. In this way, the proposed scheme of deviation error representation can rather be recommended for a final precise estimation of gas turbine diagnosability.

This paper can only be considered as a preliminary study. The investigations will be continued to better investigate this new scheme and to draw the final conclusion on its applicability in gas turbine diagnostics.

Acknowledgments

References

1. Roemer, M.J. An Overview of Selected Prognostic Technologies with Application to Engine Health Management [Text] / M.J. Roemer, C.S. Byingto, G. Kacprzynski, G. Vachtsevanos. − Proc. ASME Turbo Expo 2006, Barcelona, Spain, 2006. − 9 p.

2. Miro Zarate, L.A. Mejoramiento de la simulacion de fallas de turbinas de gas [Text] / L.A. Miro Zarate, I. Loboda, E. Torres Garcia. − Memorias del 12º Congreso Nacional de Ingenieria Electromecanica y de Sistemas, Mexico, 2010. − P. 576 – 581.

3. Duda, R.O. Pattern Classification [Text] / R.O. Duda, P.E. Hart, D.G. Stork. − Wiley–Interscience, New York, 2001.

4. Loboda, I. Deviation problem in gas turbine health monitoring [Text] / I. Loboda, S. Yepifanov, Y. Feldshteyn. − Proc. IASTED International Conference on Power and Energy Systems, Clearwater Beach, Florida, USA, 2004. − P. 335 – 340.

5. Mejer–Homji, C.B. Gas turbine performance deterioration [Text] / C.B. Mejer–Homji, M.A. Chaker, H. M. Motiwala. − Proc. Thirtieth Turbomachinery Symposium, Texas, USA, 2001. − P. 139 – 175.

6. Loboda, I. Diagnostic analysis of maintenance data of a gas turbine for driving an electric generator [Text] / I. Loboda, S. Yepifanov, Y. Feldshteyn // International Journal of Turbo & Jet Engines. − 2009. − Vol. 26, Is. 4. − P. 235 – 251.

7. Vapnik, V. Estimation of Dependencies Based on Empirical Data [Text] / V. Vapnik. − Springer Verlag, New York, 1982.

8. Maravilla Herrera, C. A comparative analysis of turbine rotor inlet temperature models [Text] / C. Maravilla Herrera, S. Yepifanov, I. Loboda. − Proc. ASME Conference Turbo Expo 2011, Vancouver, Canada, 2011. − 11 p.

9. Loboda, I. A generalized fault classification for gas turbine diagnostics on steady states and transients [Text] / I. Loboda, S. Yepifanov, Y. Feldshteyn // Journal of Engineering for Gas Turbines and Power. − 2007. − Vol. 129, Is. 4. − P. 977 – 985.

БІЛЬШ РЕАЛІСТИЧНЕ ПОДАННЯ ПОМИЛОК ВІДХИЛЕНЬ ВИМІРЮВАНОГО ПАРАМЕТРА В АЛГОРИТМ ДІАГНОСТИКИ ГТД

І.І. Лобода

Алгоритми локалізації дефектів проточної частини, засновані на теорії розпізнавання образів, є важливим компонентом системи контролю ГТД. Ці алгоритми зазвичай залучають теоретичні розподіли випадкових чисел для моделювання випадкових помилок (шуму) в описі класів дефектів. Рівень модельованого шуму визначається на основі відомої статистичної інформації про помилки різних датчиків проточної частини. Однак, не самі виміри, а їх відхилення від нормальних значень є вхідними параметрами для діагностичних алгоритмів. Ці відхилення, розраховані для реальних даних, мають інші складові помилок в додаток до модельованої неточності вимірювань. Таким чином, модельовані і реальні помилки відхилень відрізняються амплітудою та розподілом. Отже, при такому моделюванні характеристики діагностичного алгоритму будуть оцінені не точно, і, тому, висновок про ефективність алгоритму може виявитися помилковим. У даній статті для того, щоб краще побачити особливості шуму, будуються та вивчаються графіки відхилень для реальних даних. Можливі помилки відхилень також ретельно аналізуються аналітично. Для того, щоб зробити подання шуму більш реалістичним, пропонується виділити випадкові помилки з реальних відхилень і інтегрувати їх в опис дефектів. У висновку оцінюється вплив нового способу завдання шуму на достовірність діагностування ГТД.

Ключові слова: ГТД, діагностування проточної частини, відхилення вимірюваних параметрів, помилки відхилень.

БОЛЕЕ РЕАЛИСТИЧНОЕ ПРЕДСТАВЛЕНИЕ ОШИБОК ОТКЛОНЕНИЙ ИЗМЕРЯЕМЫХ ПАРАМЕТРОВ В АЛГОРИТМАХ ДИАГНОСТИКИ ГТД

И.И. Лобода

Алгоритмы локализации дефектов проточной части, основанные на теории распознавания образов, являются важным компонентом системы контроля ГТД. Эти алгоритмы обычно привлекают теоретические распределения случайных чисел для моделирования случайных ошибок (шума) в описании классов дефектов. Уровень моделируемого шума определяется на основе известной статистической информации об ошибках различных датчиков проточной части. Однако, не сами измерения, а их отклонения от нормальных значений являются входными параметрами для диагностических алгоритмов. Эти отклонеия, рассчитанные для реальных данных, имеют другие составляющие ошибок в дополнении к моделируемой неточности измерений. Таким образом, моделируемые и реальные ошибки отклонений отличаюся амплитудой и распределением. Следовательно, при таком моделировании характеристики диагностического алгоритма будут оценены не точно, и, поэтому, заключение об эффективности алгоритма может оказаться ошибочным. В данной статье для того, чтобы лучше увидеть особенности шума, строятся и изучаются графики отклонений для реальных данных. Возможные ошибки отклонений также тщательно анализируются аналитически. Для того, чтобы сделать представление шума более реалистичным, предлагается выделить случайные ошибки из реальных отклонений и интегрировать их в описание дефектов. В заключении оценивается влияние нового способа задания шума на достоверность диагностирования ГТД.

Ключевые слова: ГТД, диагностирование проточной части, отклонения измеряемых параметров, ошибки отклонений.

Лобода Игорь Игоревич – канд. техн. наук, доцент, преподаватель Национального политехнического института, Мехико, Мексика, e-mail: iloboda@ipn.mx.