
**Condition monitoring and diagnostics
of machines — Data interpretation
and diagnostics techniques —**

**Part 2:
Data-driven applications**

*Surveillance et diagnostic d'état des machines — Interprétation des
données et techniques de diagnostic —*

Partie 2: Systèmes guidés par les données



STANDARDSISO.COM : Click to view the full PDF of ISO 13379-2:2015



COPYRIGHT PROTECTED DOCUMENT

© ISO 2015

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

	Page
Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
4 Procedure to implement data-driven monitoring	2
4.1 Principle of data-driven monitoring methods.....	2
4.2 Asset critical failures and process parameters selection.....	3
4.3 Data cleaning and resampling.....	3
4.3.1 General.....	3
4.3.2 Interpolation errors.....	3
4.3.3 Data quality issues.....	3
4.3.4 Data resampling.....	4
4.4 Model development.....	4
4.4.1 General.....	4
4.4.2 Definition of models and selection of relevant inputs.....	4
4.4.3 Selection of relevant operating conditions and data.....	4
4.4.4 Preparation of the model tests.....	5
4.5 Model performance evaluation.....	5
4.6 Alarm setting.....	5
5 Procedure to implement data-driven diagnosis	6
5.1 General.....	6
5.2 Automated pattern classification approach.....	6
5.3 Simplified automated signature classification approach.....	7
6 General recommendations to implement data-driven monitoring methods	8
Annex A (informative) Example of data-driven monitoring application	9
Annex B (informative) Example of data-driven diagnostic application	11
Bibliography	12

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT), see the following URL: [Foreword - Supplementary information](#).

The committee responsible for this document is ISO/TC 108, *Mechanical vibration, shock and condition monitoring*, Subcommittee SC 5, *Condition monitoring and diagnostics of machine systems*.

ISO 13379 consists of the following parts, under the general title *Condition monitoring and diagnostics of machines — Data interpretation and diagnostics techniques*:

- *Part 1: General guidelines*
- *Part 2: Data-driven applications*
- *Part 3: Knowledge-based applications*

Introduction

This part of ISO 13379 contains general procedures that can be used to determine the condition of a machine relative to a set of baseline parameters. Changes from the baseline values and comparison to alarm criteria are used to indicate anomalous behaviour and to generate alarms: this is usually designated as condition monitoring. Additionally, procedures that identify the cause(s) of the anomalous behaviour are given in order to assist in the determination of the proper corrective action: this is usually designated as diagnostics.

STANDARDSISO.COM : Click to view the full PDF of ISO 13379-2:2015

[STANDARDSISO.COM](https://standardsiso.com) : Click to view the full PDF of ISO 13379-2:2015

Condition monitoring and diagnostics of machines — Data interpretation and diagnostics techniques —

Part 2: Data-driven applications

1 Scope

This part of ISO 13379 gives procedures to implement data-driven monitoring and diagnostic methods to facilitate the work of analysis carried out by specialist staff typically located in a monitoring centre.

Although some of the steps are embedded in existing tools, it is essential to be aware of the following steps for optimum use:

- selection of the asset, the critical failures and the available process parameters;
- data cleaning and resampling;
- model development;
- model initialization and tuning;
- model performance evaluation;
- diagnostics process.

The implementation of these steps does not require a thorough knowledge of the statistical methods. It does require the competence first to build the training models and then to carry out monitoring and diagnostics processes.

The training in data-driven monitoring is carried out on equipment that is exhibiting normal behaviour. In that case, the principle of fault detection is to compare observed data to estimated data. A difference (called residuals) between an observed and expected values of the parameters reveals the presence of an anomaly, which can be related either to equipment or instrument.

The training in data-driven diagnosis is carried out both on equipment that is exhibiting normal behaviour and failures. The principle of the method is not to detect the deviation of a parameter but to identify a fault by comparison of the observed situation to the faults learnt during the training phase. The technique usually applied is pattern recognition followed by pattern classification.

Data can be available from the data historian of the distributed control system (DCS) or from specialized monitoring systems.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 13372, *Condition monitoring and diagnostics of machines — Vocabulary*

ISO 13379-1, *Condition monitoring and diagnostics of machines — Data interpretation and diagnostics techniques — Part 1: General guidelines*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 13372 and ISO 13379-1 apply.

4 Procedure to implement data-driven monitoring

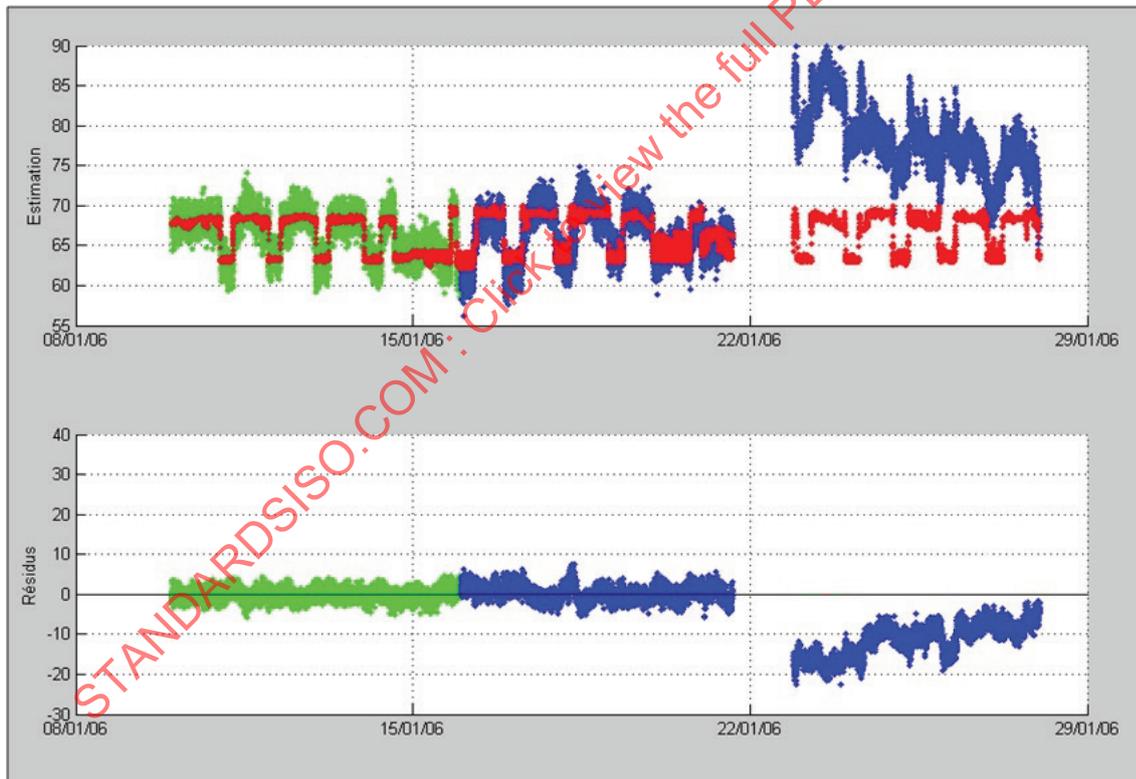
4.1 Principle of data-driven monitoring methods

Advanced statistical methods that simultaneously consider multiple plant signals and model the underlying relationship between them are beginning to replace the classical methods for condition monitoring which are based on the observation of trends of individual signals.

These monitoring methods rely on the same principle to detect a fault, which is to compare observed data to estimated data.

Prior to the monitoring phase, it is required to build the model of the normal equipment behaviour, during a training phase. Faults can thus often be detected as deviations between an observed and an expected value of the parameters of the system.

[Figure 1](#) shows an example of an application on a gas turbine. The objective is to detect abnormal shaft displacements after a shut down. Several inputs are considered in the model: active and reactive power and bearing displacements.



Key
 green training
 blue monitoring
 red prediction

Figure 1 — Gas turbine displacement magnitude and residual

Data-driven monitoring methods generally applied are Auto associative kernel regression (AAKR), cluster and partial least square (PLS), support vector machine (SVM), and/or Mahalanobis-Taguchi (MT) methods.

4.2 Asset critical failures and process parameters selection

The procedure for the implementation of data-driven monitoring is precisely described in ISO 17359. It includes two main audits:

- equipment audit: identify equipment and its function;
- reliability and criticality audit: produce a reliability block diagram, establish equipment criticality and perform failure modes, effects and criticality analysis.

Once this preliminary study is carried out and the list of the critical faults is identified, it is necessary to list the process data available in the data historian or in specialized monitoring systems. An example would be a vibration monitoring system.

It might be necessary to consider the installation of additional sensors or location of existing sensors if the detection scope of the critical faults is not completely covered.

4.3 Data cleaning and resampling

4.3.1 General

In order to build a robust model, one shall first collect data covering all the operating conditions in which the system is expected to run and for which signal validation is desired. These data are historical data that have been collected and stored. In fact, they might not always represent the real plant state due to several anomalies that commonly occur, including interpolation errors, random data errors, missing data, loss of significant figures, stuck data, and others. Data should always be checked and corrected.

WARNING — Caution shall be taken before deleting data.

4.3.2 Interpolation errors

The first problem usually encountered when using historical data for model training is that available conditioned data do not correspond to actual data, but instead, data resulting from compression routines normally implemented in data archival programs. Generally, the data historian creates a data archive that is a time series database. However, all of the data are not stored at each collection time. Only data values that have changed by more than a specified tolerance are stored along with their time stamp. This method requires much less storage but results in a loss of data fidelity. When data are extracted from the historian, data values between logged data points are calculated through either a simple linear interpolation or a step at the time of the second data point. The resulting data appear to be a saw-tooth time series and the correlations between sensors might be severely changed.

As a conclusion, data collected for model training should be actual data and tolerances should be set as small as possible or not used.

4.3.3 Data quality issues

Several of the most common data quality issues are:

- missing data;
- noisy or random data;
- defective sensors for which the data value is not updated or is out of calibration;
- unreasonable data values (out of range).

Most of these data problems can be visually identified or can be detected by a data clean-up utility. These utilities remove bad data or replace it with the most probable data value using an algorithm. It is most common to delete all bad data observations from the training data set. Most software systems include automated tools for data clean-up; these tools easily identify extreme outlying data but are typically insensitive to data errors that occur within the expected region of operation. The addition of bad data points in a training set can invalidate a model.

4.3.4 Data resampling

Once the data have been cleaned, it might be necessary to resample the data at a lower rate determined by the selected operation modes. Thus, it is advised to keep all the time stamps to characterize the transients of the significant operating parameters (e.g. run down of a machine) whereas under steady-state operation, a sample every 10 min (obtained by average or not) might be sufficient.

4.4 Model development

4.4.1 General

Model development is not trivial. There are several steps that need to be performed including:

- selecting relevant features;
- selecting relevant operating regions and training data;
- preparing the model tests.

Construction of a data-driven model requires:

- a set of parameters (sensors) which focus on a specific type of fault (mechanical, electrical, thermal, etc.);
- data samples for a period during which the machine is known to be in good health.

4.4.2 Definition of models and selection of relevant inputs

Once the quality of the data has been validated, model features shall be defined. Features may be the raw sensor values themselves or derived from the sensor values (exponentially weighted moving averages, means, kurtosis, etc.). A large process plant may possess hundreds of parameters that require monitoring for the assessment of critical equipment. Hence, they shall be divided into smaller correlated groups to focus on a specific function of the equipment (thermal, mechanical, cooling, etc.).

Model performance can be strongly affected by the features included. Unnecessary features tend to reduce performance by inducing false alarms or masking real events. Unnecessary features may include non-varying or random features. Missing important features can make some faults impossible to detect.

To obtain accuracy and robustness when building a model, features should be selected bearing in mind the functional aspect (parameters useful for detection of a specific group of faults) as well as the numerical aspect. It is recommended to employ correlated features in the model and to consider the fact that normal change of the machine condition can be explained by an independent parameter of the equipment (such as external process parameters).

4.4.3 Selection of relevant operating conditions and data

The model shall be trained with data covering all operating conditions in which it is expected to be applied. These operating regions can vary significantly between plants since they are defined by system structure, sensor values, and operating procedures.

One example of an operating condition change is the periodic usage of standby pumps or the cycled usage of redundant pumps. A model shall be trained for each operating condition of the system to work

properly, but excessive training on unusual conditions might degrade the performance on the most usual operating conditions. Therefore, some plant line-ups might never be included in the training set.

Operating conditions can also change as a result of equipment repair. In this case, the model shall be retrained to account for the new condition.

Finally, operating conditions also change due to cyclic phenomena, such as seasonal variations. If a model is trained on data collected in the summer, it might not perform well in the winter when the temperature of the air, cooling water, etc. are significantly different. Additionally, unusual variation in ambient conditions might affect model performance, e.g. if a model is trained using typical summer data, and then applied during an unusually hot summer with higher cooling water temperatures, the model might not perform correctly. In this case, data from the new operating conditions shall be added to the training data.

Some particular operating conditions to keep in mind are the following.

- Times when the machine has been serviced should be noted because maintenance might cause significant changes of performance.
- It is necessary to account for machines which have very long intervals between maintenance and that show significant normal deterioration in performance.
- When building a model, be aware of the impact of preceding transients on the current behaviour (for example preceding power or speed changes).

4.4.4 Preparation of the model tests

Once all typical operating regions are identified, the next step is to select the input data and to divide the data between training, validation, and test observations. Generally, the user does not have the choice of the underlying classification method since it depends on the monitoring tool.

The training and validation data are used to develop and test the model of the process equipment. The test set which consists of the validation model with additional deviations is used to evaluate the model detection performance.

4.5 Model performance evaluation

The next step in the monitoring process is to select and optimize the model parameters, which can differ from one method to another (e.g. kernel bandwidth, maximum cluster diameter, number of Eigen values, number of hidden layers).

Models are trained using the training data and their performance is assessed using the validation data. The model parameters are selected such that performance on the validation data is best. Once the final model parameters have been selected using the training and validation data, the test set can be used to assess how well the model will perform in general. It is important that the test data not be used to tune the model parameters; doing so yields a biased assessment of generalized performance. Additionally, the performance of a model may be assessed when there is sensor drift by applying increasing bias to each sensor reading in turn, and using the model output to correct the sensor faults. The predictions using faulty input data are then used to determine the model's robustness and spillover capabilities. The robustness may be interpreted as a measure of a data-driven model's ability to make correct sensor predictions when the respective sensor value is incorrect due to some sort of fault. Spillover measures the effect a faulty sensor input has on the other sensor predictions.

4.6 Alarm setting

Setting the monitoring model alerts and alarms is a critical task in implementing a successful monitoring application. Setting model alerts is an act of balancing model sensitivity with the probability of missed alarms and the probability of false alarms.

The most common approach to setting alerts is to choose a number (threshold) that seems reasonable from an engineering perspective. This setting may be adjusted on individual parameter residuals or on an overall distance between the actual sample and the training model, or both.

Moreover, it is recommended that the performance of a developed data-driven model be quantified and that alert settings be adjusted according to the accuracy of the model's estimations.

The most common approach to alarm definition is through the use of a logic filter specifying the number of alerts that shall occur within a given window. The use of this type of logic window effectively reduces the volume of alerts, filtering out spurious points where large residual values occur.

A final element in setting alerts and alarms is choosing which signals to monitor to present alarms. In general, it is not necessary to present alarms for all of the signals in a given model. Overall, the decision is based on engineering judgment of how indicative a given signal's response is to the health of the equipment being monitored.

5 Procedure to implement data-driven diagnosis

5.1 General

A data-driven model can produce a series of residual data streams with corresponding alerts and alarms according to the model settings. Therefore, the final step is to use the residuals in order to identify the fault that is at the origin of the observed alerts and alarms.

Various causes can have provoked deviations on the parameter residuals. Firstly, one shall be sure that the observed deviations are not due to model tuning. If not, the other root causes can be an instrument or equipment fault.

Current data-driven monitoring experience generally relies on a manual interpretation process.

Many equipment degradation pathways are known and are repeatable events. If failure signatures can be defined for these pathways, the process of anomaly interpretation can be automated.

Two main approaches can be used for diagnosing the condition of the machine and are described in ISO 13379-1:

- data-driven approaches;
- knowledge-based approaches.

Numerous data-driven approaches are fully explained in ISO 13379-1. The objective of the following paragraphs is to present two applications of simple data-driven approaches:

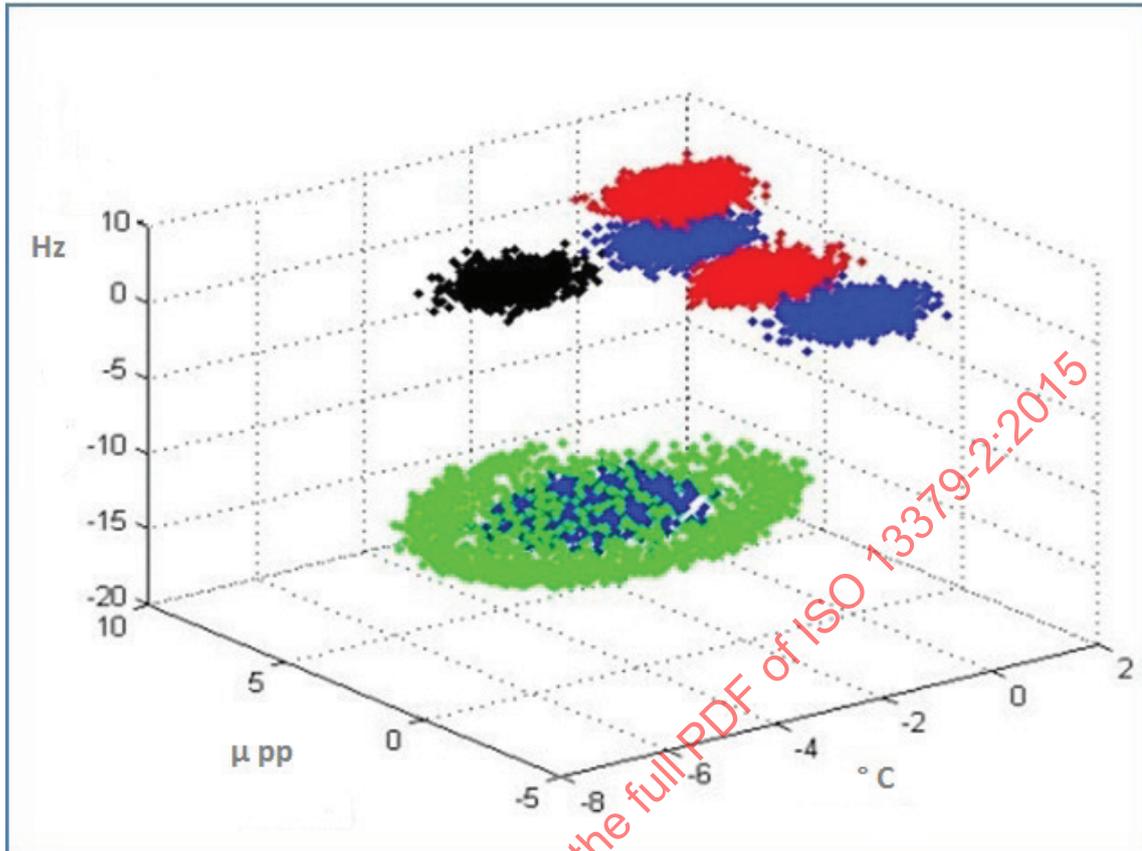
- automated pattern classification;
- simplified automated signature classification.

Another simple way to automate the diagnostics process is to program rules based on all the available stored information (residuals, absolute values, alert and alarm thresholds). Rule-based diagnosis is considered part of knowledge-based diagnosis.

5.2 Automated pattern classification approach

Pattern classification aims to classify data patterns either based on a priori knowledge or on statistical information extracted from the patterns. The patterns to be classified are usually groups of measurements, defining points in an appropriate multidimensional space.

The classification or description scheme is usually based on the availability of a set of patterns that have already been classified or described. This set of patterns is termed the training set (see [Figure 2](#)), and the resulting learning strategy is characterized as supervised learning.



Key

blue	healthy state
red	failure state 1
black	failure state 2
green	failure state 3

Figure 2 — Example of a pattern classification

The patterns can be based on raw values of parameters but also on residual data processed during the data-driven monitoring phase. The acceptable zone boundaries shall be clearly defined to avoid false diagnosis and to allow automatic diagnosis. The use of residual data enables the model to be free from the initial state of the machine and to take into account only the evolution of the behaviour due to the detected fault.

Pattern classification can be easily implemented to faults that can be well characterized by a stable state. The classification of faults with a signature that evolves over time (wear, rubbing, crack, etc.) is possible, but more difficult.

5.3 Simplified automated signature classification approach

One simple and automated classification approach may be carried out using residual data with corresponding alerts and alarms.

The list of alarms corresponds to a set of status indicators for the component process measurements (for example: parameter 1 high (or positive alarm); parameter 2 and 3 no alarm; parameter 4 low (or negative alarm), etc.

One way to use this information is to convert the alert and alarm list into a numerical pattern, which provides [1 0 0 -1]. This representation of the decision implies a logic-based approach to determining the cause of degradation for the component (see [Table 1](#)).

Table 1 — Sample

Failure signature with six sensors						Condition assessment
1	0	0	-1	1	0	Component failure n°1
-1	-1	0	-1	0	0	Component failure n°2
0	0	0	1	1	1	Component failure n°3
1	0	0	0	0	0	Sensor failure X1
0	1	0	0	0	0	Sensor failure X2
0	0	1	0	0	0	Sensor failure X3
0	0	0	1	0	0	Sensor failure X4
0	0	0	0	1	0	Sensor failure X5
0	0	0	0	0	1	Sensor failure X6
-1	0	0	0	0	0	Sensor failure X1
0	-1	0	0	0	0	Sensor failure X2
0	0	-1	0	0	0	Sensor failure X3
0	0	0	-1	0	0	Sensor failure X4
0	0	0	0	-1	0	Sensor failure X5
0	0	0	0	0	-1	Sensor failure X6

6 General recommendations to implement data-driven monitoring methods

For optimum implementation of data-driven methods for monitoring and diagnostics purposes, one should:

- list the critical faults to be identified and the associated process data in the data historian or in specialized monitoring systems;
- build models with less than 20 descriptors for both monitoring and diagnostics purposes;
- integrate correlated descriptors for monitoring data-driven methods in the model;
- clean up historical data corresponding to normal behaviour for detection purposes;
- clean up historical data corresponding to normal and faulty behaviours for diagnostics purposes;
- keep track of the training data and of the monitoring configuration for data interpretation and retraining;
- verify that there is clear separation between patterns corresponding to different failures;
- pay attention to the robustness of the model when integrating trend values of the fault descriptors in a diagnostics model;
- set up reasonable alarm thresholds on residues for detection purposes (for example a three degree threshold could generate numerous alarms).

Annex A (informative)

Example of data-driven monitoring application

A.1 General

This Annex gives an example of the application of data-driven monitoring.

A.2 Vibration behaviour of a combined cycle steam turbine (ST)

During operation, absolute vibrations of this ST bearing reached abnormal levels generating operating alarms.

The purpose of this case study was to identify with empirical models whether the trend was due to operating parameter changes or to a mechanical problem of the intermediate steam turbine bearings.

Several inputs were considered for this case study: active power, input steam flows and backpressure. Output parameters were the vibrations of the two intermediate vibration turbine bearings expressed in mm/s rms.

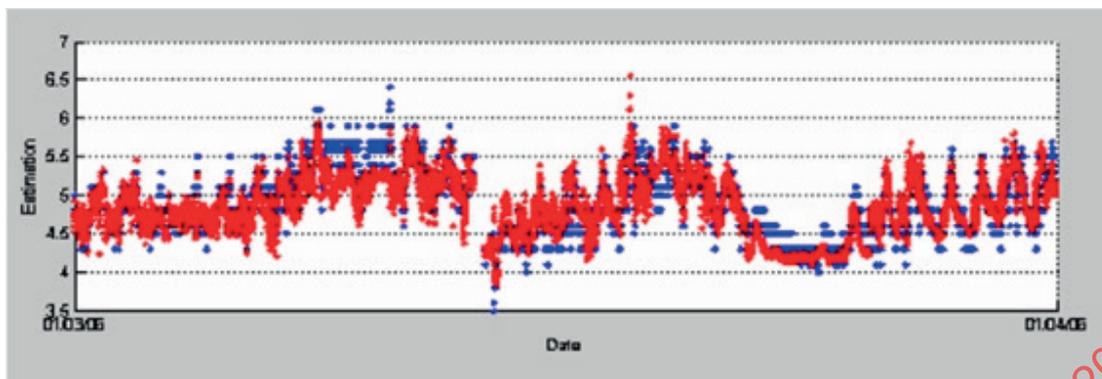
[Figure A.1](#) presents estimated vibration levels in red, during three phases (training period, normal and abnormal behaviour test periods).

Two reasons were raised to explain this behaviour change:

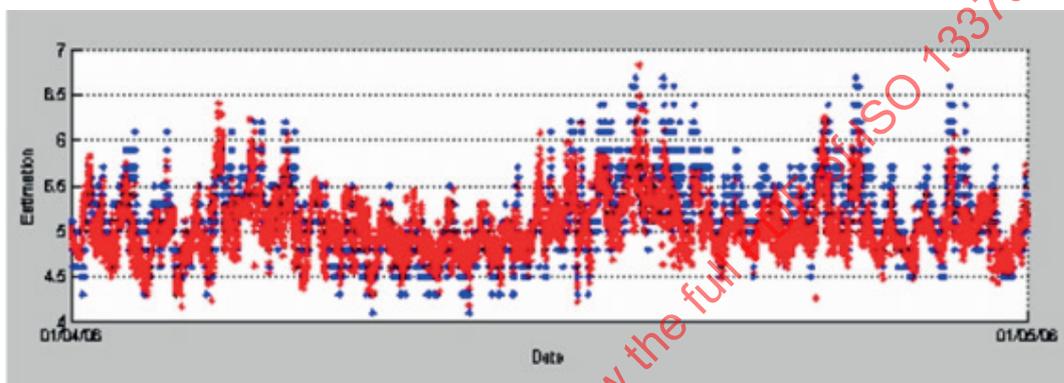
- a) the stiffness of the overall structure has drifted or was modified during the last shutdown, or both;
- b) such a high backpressure has never been reached before.

What was deduced from the analysis of [Figure A.1](#) is that:

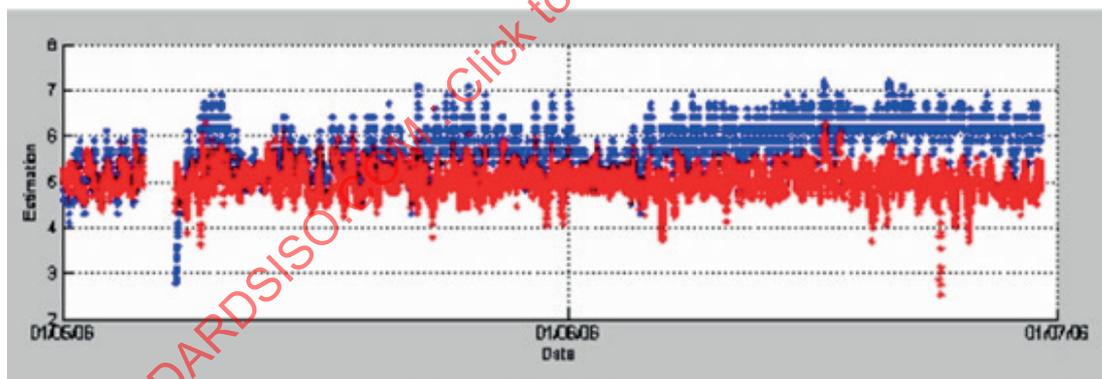
- a slight trend of the residues (difference between red and blue graphs) could have been detected for several months;
- the trend was not related to a long-term change of operating parameters.



a) Training phase



b) Normal behaviour period



c) Abnormal behaviour period

Key
 blue monitoring
 red prediction

Figure A.1 — Estimated vibration levels for three ST states