INTERNATIONAL STANDARD

ISO 13379-1

First edition
2012-05-01

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

Part 1: **General guidelines**

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

Partie 1: Lignes directrices générales

Reference number
ISO 13379-1:2012(E)

Contents

Page

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 13379-1 was prepared by Technical Committee ISO/TC 108, *Mechanical vibration, shock and condition monitoring*, Subcommittee SC 5, *Condition monitoring and diagnostics of machines*.

This first edition of ISO 13379-1 cancels and replaces ISO 13379:2003, which has been technically revised.

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*

The following parts are planned:

— *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.

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

Part 1:
General guidelines

1 Scope

This part of ISO 13379 gives guidelines for the data interpretation and diagnostics of machines. It is intended to

— allow the users and manufacturers of condition monitoring and diagnostics systems to share common concepts in the fields of machine diagnostics;

— enable users to prepare the necessary technical characteristics that are used for the further diagnosis of the condition of the machine;

— give an appropriate approach to achieve a diagnosis of machine faults.

Since these are general guidelines, a list of the machine types addressed is not included. However, the machine sets covered by this part of ISO 13379 normally include industrial machines such as turbines, compressors, pumps, generators, electrical motors, blowers, gearboxes, and fans.

2 Normative references

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

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

3 Terms and definitions

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

4 Condition monitoring set-up and diagnostics requirements

4.1 Role of diagnostics in operation and maintenance

Diagnostics have an essential role in decision making for operation and maintenance tasks. In order to be effective, diagnostics procedures should be set up according to the faults that can occur in the machine. Therefore, it is strongly recommended that a preliminary study be carried out when preparing the requirements for the condition monitoring and diagnostics system of a machine.

4.2 Establishing diagnostics needs

The principle of this study is shown in Figure 1. The V-shape has been intentionally chosen to represent the high-level concerns (maintenance: machine, risk assessment) and the “low level” ones (measurements: monitoring, periodical tests, data processing).

The left branch corresponds to the preliminary study, which prepares, for a particular machine, the necessary data for condition monitoring and diagnostics. The right branch of the sketch corresponds to the condition monitoring and diagnostic activities that are normally undertaken after the machine has been commissioned. Each layer consists of a preparatory design phase (left) and a usage phase (right).

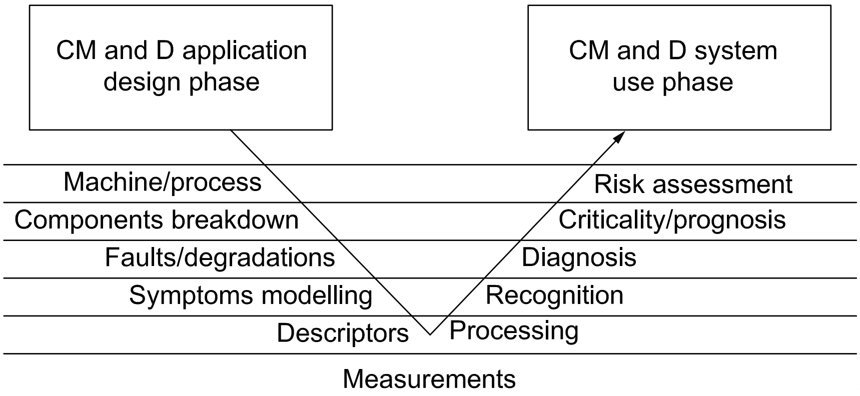


Figure 1 — Condition monitoring and diagnostics (CM and D) cycle: design and use of the application on a machine

The generic steps of the diagnostic study include the following:

a) analyse the machine availability, maintainability, criticality with respect to the whole process;

b) list the major components and their functions;

c) analyse the failure modes and their causes as component faults;

d) express the criticality, taking into account the significance (safety, availability, maintenance costs, production quality) and the occurrence;

e) decide accordingly which faults should be covered by diagnostics (“diagnosable”);

f) analyse under which operating conditions the different faults can be best observed and define reference conditions;

g) express the symptoms that can serve in assessing the condition of the machine and that are used for diagnostics;

h) list the descriptors that are used to evaluate (recognize) the different symptoms;

i) identify the necessary measurements and transducers from which the descriptors are derived or computed.

The steps given in a), b), c) and d) may be followed using maintenance optimization such as FMEA (failure modes and effects analysis) or FMECA (failure modes, their effects and criticality analysis). They also may be accomplished within a more general process of maintenance optimization like RCM (reliability-centred maintenance).

NOTE The FMEA and FMECA procedures are outlined in IEC 60812[6].

The steps given in c), d), e), f), g), h) and i) may be followed using the FMSA (failure mode symptoms analysis) methodology explained in 4.3.

4.3 Failure mode symptoms analysis

4.3.1 Process of failure mode symptoms analysis

The aim of this process is to select monitoring technologies and strategies that maximize the confidence level in the diagnosis and prognosis of any given failure mode.

This methodology is designed to assist with the selection of monitoring techniques that provide the greatest sensitivity to detection and rate of change of a given symptom. Where the confidence in a technique's sensitivity and resulting diagnosis/prognosis accuracy is questionable, the use of additional techniques for further correlation is recommended.

This process is essentially a modification of a FMECA process with a focus on the symptoms produced by each failure mode identified and the subsequent selection of the most appropriate detection and monitoring techniques and strategies.

This tool should be used in conjunction with an existing FMECA analysis that has already identified and ranked possible failure modes.

4.3.2 Guide for usage

This process is best represented by Table A.1. The essential items are as follows:

— listing the components involved;

— listing the possible failure modes for each component;

— listing the effects of each failure mode;

— listing the causes of each failure mode;

— listing the symptoms produced by each failure mode;

— listing the most appropriate monitoring technique;

— listing the estimated frequency of monitoring;

— ranking each failure mode by detection, severity, diagnosis confidence and prognosis confidence resulting in a monitoring priority number (MPN);

— listing the most appropriate correlation techniques;

— listing the frequency of monitoring for the correlation techniques.

The greatest difficulty arises in establishing the correct terms for failure mode, effect, and cause. The failure mode is a definition of how the failure would be observed, i.e. bent, corroded, etc. In the FMECA processes that should have been carried out prior to the FMSA process, there are areas of overlap between the terms used for the failure modes, effects and causes. An item may appear as a *cause of failure* in one line when considering a component and as a *failure mode* in another. A term may also appear as an *effect* in one line when dealing with a component and as a *failure mode* when dealing with an assembly. This also remains true for the FMSA process.

Care shall be taken to avoid duplication of failure mode and cause on the same line. For any one item the failure mode, effect, and cause shall read logically across the page. It can help to use the following form:

— a *failure mode* could result in an *effect* due to a *cause*.

When considering monitoring strategies, the following form can also be used:

— a *failure mode* produces *symptoms*, which are best detectable by a *primary monitoring technique* resulting in a high diagnosis and prognosis confidence when monitored at a given *monitoring frequency*;

— increased diagnosis and prognosis confidence can be gained by using “correlation techniques” when monitored at a given “monitoring frequency”.

4.3.3 Guide for rating

4.3.3.1 General

A rating which estimates the likelihood of detection, prognosis accuracy, and the degree of severity is assigned to each column. Provided that a user applies a consistent rating throughout all analyses, the higher risk categories reflect a higher MPN.

4.3.3.2 Rating detection (DET)

The likelihood of detection is rated from 1 to 5 and is designed to reflect the overall detectability of a failure mode irrespective of the following accuracy of diagnosis or prognosis. This rating is designed to highlight failure modes which:

— produce symptoms that are detectable but unrepeatable;

— produce symptoms that are undetectable;

— produce symptoms that are not measurable in practice; or

— produce symptoms that may be masked by other failure mode symptoms.

This is estimated on a scale of 1 to 5, where:

— 1 means “There is a REMOTE LIKELIHOOD that this failure mode will be detected.”

— 2 means “There is a LOW LIKELIHOOD that this failure mode will be detected.”

— 3 means “There is a MODERATE LIKELIHOOD that this failure mode will be detected.”

— 4 means “There is a HIGH LIKELIHOOD that this failure mode will be detected.”

— 5 means “It is VIRTUALLY CERTAIN that this failure mode will be detected.”

4.3.3.3 Severity of failure (SEV)

This ranking should reflect any previous FMECA analysis and is designed to rank individual failure modes by risk.

This is estimated on a scale of 1 to 4, where

— 1 means “Any event which could cause degradation of system performance function(s) resulting in negligible damage to either system or its environment, and no damage to life or limb.”

— 2 means “Any event which degrades system performance function(s) without appreciable damage to either system or life or limb.”

— 3 means “Any event which could potentially cause the loss of primary system function(s) resulting in significant damage to the said system or its environment and negligible hazard to life or limb.”

— 4 means “Any event which could potentially cause the loss of primary system function(s) resulting in significant damage to the system or its environment, and/or cause the loss of life or limb.”

4.3.3.4 Diagnosis confidence (DGN)

The predicted accuracy of the diagnosis is also rated from 1 to 5. This rating is designed to identify failure modes with:

— detectable, but unrepeatable, symptoms;

— unknown symptoms; or

— symptoms that are not distinguishable from other failure mode symptoms.

This is estimated on a scale of 1 to 5, where:

— 1 means “There is a REMOTE LIKELIHOOD that this failure mode diagnosis will be accurate.”

— 2 means “There is a LOW LIKELIHOOD that this failure mode diagnosis will be accurate.”

— 3 means “There is a MODERATE LIKELIHOOD that this failure mode diagnosis will be accurate.”

— 4 means “There is a HIGH LIKELIHOOD that this failure mode diagnosis will be accurate.”

— 5 means "It is VIRTUALLY CERTAIN that this failure mode diagnosis will be accurate."

4.3.3.5 Prognosis confidence (PGN)

The predicted accuracy of the prognosis is also rated from 1 to 5. This rating is designed to identify failure modes with:

— detectable, but unrepeatable, symptoms;

— symptoms that are not sensitive to changes in degradation;

— unknown failure rates; or

— symptoms that are not distinguishable from other failure mode symptoms.

This is estimated on a scale of 1 to 5, where:

— 1 means "There is a REMOTE LIKELIHOOD that this failure mode prognosis will be accurate."

— 2 means "There is a LOW LIKELIHOOD that this failure mode prognosis will be accurate."

— 3 means "There is a MODERATE LIKELIHOOD that this failure mode prognosis will be accurate."

— 4 means "There is a HIGH LIKELIHOOD that this failure mode prognosis will be accurate."

— 5 means "It is VIRTUALLY CERTAIN that this failure mode prognosis will be accurate."

The frequency of monitoring also contributes to the determination of the accuracy of expected prognosis, i.e. the greater the frequency of monitoring used, the higher the confidence in the expected failure rate and prognosis.

4.3.3.6 Monitoring priority number (MPN)

The ranking is the multiplication of the four preceding rankings and results in an overall rating of each failure mode.

A high MPN value indicates that the nominated technique is the most suitable for the detection, diagnosis, and prognosis of the associated failure mode.

It should be noted that a low MPN value does not imply that monitoring is not necessary but rather that a low confidence level for detection, analysis, and prognosis can be expected with the nominated monitoring technique and frequency.

The least favourable case is a failure mode with high severity, low detectability, low diagnosis confidence and low prognosis confidence.

The most favourable case is a failure mode with low severity, easily detectable, with known failure modes and associated patterns and, therefore, high diagnosis and prognosis confidence levels.

The implementation of a FMSA review and monitoring system design should therefore be carried out taking into consideration:

— safety risk of each failure mode;

— expected rate of deterioration of each failure mode;

— mean time between failures (MTBF) for each failure mode;

— secondary/subsequent failure modes;

— failure mode interrelationships;

— maintenance lead time required;

— availability of spare parts;

— required reliability and availability.

Continuous reassessment should be carried out when experience with a new installation has been gained or when a modification has been carried out.

4.4 Diagnostics requirements report

It is recommended that the synthesis of the preliminary study be stored in a “diagnostics requirements report”. This report should normally:

a) present the adopted breakdown of the machine into components;

b) list the faults associated with these components;

c) give the potentially observable symptoms for each fault;

d) name the condition monitoring descriptors that are to be used; and

e) indicate the method and parameters used for calculation of the descriptors.

It may arise that all the critical faults are not covered by condition monitoring and, as such, are not diagnosable. For this reason, it is strongly recommended that the faults which are addressed and those which are not be emphasized clearly in the report. It may be of use to re-evaluate the value of adding the capability to detect specific faults.

Formally, the diagnostics requirements report can be composed of two parts:

1) a machine description [corresponding to items a) to b) of 4.2]: identification, role in the process, components, criticality analysis;

2) a failure modes/symptoms analysis [corresponding to items c) to i) listed in 4.2]: failure modes, symptoms, descriptors and measurements that are to be used for diagnostics.

Item b) may be realized easily with the FMSA chart given in Annex A.

It is also recommended that the theoretical effectiveness of the diagnostics system be calculated. For this purpose, a proposal for a criterion for the effectiveness of a diagnostics system is given in Annex B.

5 Elements used for diagnostics

5.1 Condition monitoring data

5.1.1 Measurements

All the measurements used for condition monitoring are generally suitable for diagnostics. Descriptors are preferred instead of raw measurements for diagnostics as they offer greater selectivity with respect to faults.

Table 1 gives, as an example, a set of various measurements and parameters used for condition monitoring and diagnostics of a machine.

Table 1 — Example of measurements and parameters used for diagnostics

Performance	Mechanical	Electrical	Oil analysis, product quality and others
Power consumption	Thermal expansion	Current	Oil analysis
Efficiency	Position	Voltage	Ferrography wear debris analysis
Temperature	Fluid level	Resistance	Product dimensions
Pressure	Temperature	Inductance	Product physical properties
IR thermography	Vibration displacement	IR thermography	Product chemical properties
Flow	IR thermography	Capacitance	— colour
	Vibration velocity	Magnetic field	— visual aspect
	Vibration acceleration	Insulation resistance	— smell
	Audible noise	Partial discharge	— other non-destructive testing
	Ultrasonic waves		

5.1.2 Descriptors

Descriptors can be obtained from the condition monitoring system, either directly or after the processing of the measurements. Descriptors are often preferred to measurements for reasons of selectivity. The more selective the descriptors are, the more selective the symptoms and, therefore, the easier the diagnosis. The descriptor selectivity reduces the number of fault hypotheses when inferring from symptoms to fault.

EXAMPLES Amplitude of the first harmonic of the shaft displacement of vibration, crest factor of the acceleration of the vibration, oil total acid number, rotational speed, rolling element bearing damage factor, temperature gradient on an infrared thermography trace.

5.1.3 Symptoms

A symptom can be expressed in the following terms.

a) Time characteristic: the time constant of the evolution of the descriptor.

EXAMPLES 1 h; 10 days; slow.

b) Type of evolution and magnitude change

EXAMPLES Presence; absence; regular increase; decrease; stability; >10; <200; 40 µm cyclic evolution.

c) Descriptor: the descriptor used.

EXAMPLES Temperature; first harmonic of the displacement of the vibration.

d) Location: where the symptom is observable on the machine.

EXAMPLES Shaftline at bearing No. 3 vertical direction; bearing pedestal No. 4; high-pressure body (front left); bearing No. 2.

e) Circumstance: operating conditions in which the symptom is seen.

EXAMPLES During run down; within 1 h after cold start-up; at 100 % power; any circumstance.

When preparing the selection of symptoms for a fault, care should be taken to avoid taking two or several symptoms that may be too dependent (highly correlated), as the evaluation of dependent symptoms does not give more information and, thus, does not allow the diagnosis to progress.

EXAMPLES Slow and regular evolution of first harmonic vector of shaft displacement; bearing temperature is 10 °C above usual value in nominal conditions; a 2 mm/s instantaneous change in pedestal vibration velocity; cyclic evolution of the first harmonic of the displacement of the vibration (>10 µm, after a change in power delivered by the machine); unusual noise; dark colour of the lubricant oil.

5.1.4 Fault

A fault can be expressed in the following terms.

a) Machine: the name or the identifier of the machine.

EXAMPLES Unit No. 1 turbine; boiler feed water pump No. 2; BFW PU2; circulation pump; coal crusher No. 5.

b) Component: name or identifier of the component of the machine on which the fault occurs.

EXAMPLES Bearing No. 3; shaft; piston; low pressure body; seal No. 2.

c) Type of degradation of the component of the machine (compulsory).

EXAMPLES Incorrect clearance, fractures, unbalance, misalignment, opening of contacts.

d) Severity: integer number for example (defined in 4.3.3.3) representative of the magnitude of the degradation or failure mode.

5.1.5 Operational parameters

Operational parameters are often used for diagnostics. They are used both for:

— establishing some descriptors; and

— establishing the operating conditions in which the symptoms appear (circumstance).

Care should be taken when considering operational parameters. When it is a descriptor or enters the computation of a descriptor, the parameter is an output. It is an input when it characterizes an operating condition. This should be considered in order to avoid using an operating condition as a descriptor. For example, the turbine body temperature is a descriptor when monitoring and diagnosing the body. It becomes an operating condition when monitoring the bearing for it has an influence on the work of the bearing, but is not descriptive of bearing faults.

5.2 Machine data

Knowledge of specific data of the machine is often necessary for diagnostics. This is the case, for example:

— for vibrations — data regarding the kinematics of the components of the machine such as rotational speeds, number of teeth on gears, ball bearings characteristic frequencies;

— for oil analysis — data regarding the oil path of the machine, flows, metal composition, filters disposition and fineness;

— for thermography — IR emissivity of a surface.

A distinction should be made between data related to the techniques used for processing descriptors and data related to the configuration of the machine. It is important to record both for the purpose of diagnostics. Data related to the configuration of the machine are normally recorded in the machine file, as it is preferable to record machine data related to condition monitoring techniques within diagnostics requirements, when specifying the descriptors.

5.3 Machine history

Fault occurrence can be linked not only to operation, but also to maintenance of the machine. It may arise that a fault has been introduced during an overhaul or a particular situation. Therefore, it is important to keep a record of the fault history, operational history and maintenance history of the machine in order to take these facts into account for diagnostics.

6 Diagnostic approaches

6.1 Two types of approaches

The diagnosis process is generally triggered by detection of an anomaly during routine monitoring, routine analysis, random analysis or human perception. This detection is carried out by making a comparison between the present descriptors of a machine and reference values (generally called baseline values or data) chosen from experience, from the manufacturer's specifications, from commissioning tests or computed from statistical data (e.g. long-term average).

The possible result of any diagnostic process should include a scenario when no fault exists.

Two main approaches can be used for diagnosing the condition of the machine.

a) **Data-driven** approaches (simple trending, neural network, pattern recognition, statistical, histographic Pareto approach or other numerical approaches). These methods are generally automated, do not require deep knowledge of the mechanism of fault initiation and propagation, but do require training the algorithm using a large set of observed fault data.

b) **Knowledge-based** approaches, which rely on an explicit representation fault behaviour or symptoms through, for example, fault models, correct behaviour models or case description.

It is recognized that these methods may overlap and solutions may be developed which utilize a combination of several approaches.

It should be noted that, in order to analyse the entire process of modelling for the purpose of diagnosis, a definition of an **"observation specification"** is also needed. This specification can provide a procedure that describes and interprets the observations from acquired data. Different levels of detail may be needed: e.g. how the observation model should handle descriptions of parameter evolution, shape, time-related information, levels, and correlations.

The different usable models are described in 6.2 to 6.4.

6.2 General guidelines for the selection of appropriate diagnostic approaches

The choice of appropriate diagnostic approaches (see Figure 2) depends on the:

— application or the equipment;

— end-user of the diagnostic approach;

— monitoring technique;

— complexity of the knowledge to be modelled;

— need for having an explanatory model;

— need to retrain the model;

— availability of existing data with known faults and normal operation.

NOTE It is very important to consider in the initial design how to split online and off-line fault diagnostic processing, accounting for software and hardware capability. Once it is defined, it is important to consider how to combine them.

Some additional guidelines are given in Annexes C and D.

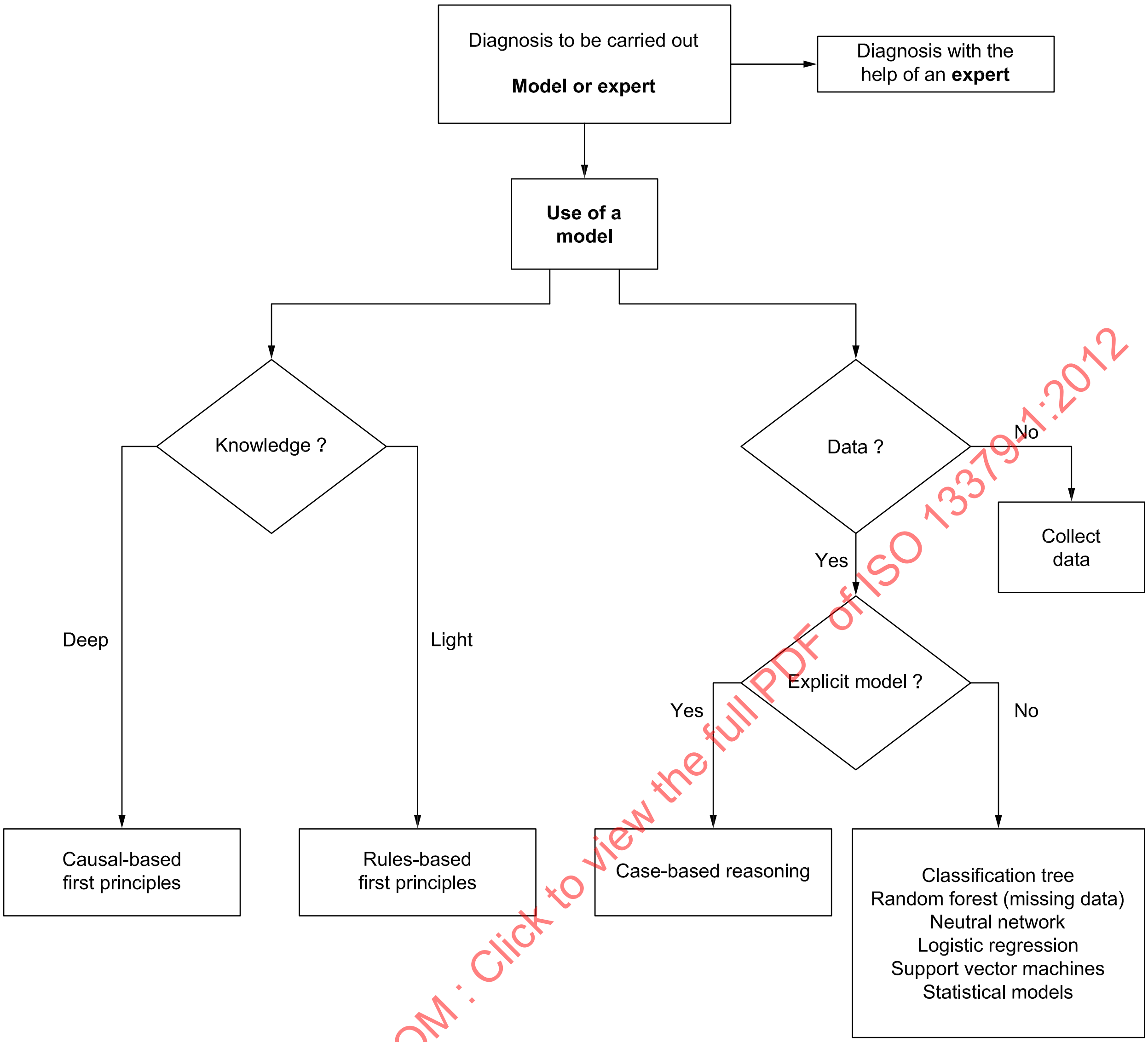


Figure 2 — General approach to the choice of a diagnostic model

6.3 Data-driven approaches

6.3.1 List of methods

This subclause presents several data-driven approaches:

— statistical data analysis and case-based reasoning;

— neural networks;

— classification trees;

— random forests (RF);

— logistic regression (LR); and

— support vector machines (SVM).

Note that there are numerous alternative data-driven approaches; these are merely reference examples of common ones.

The basic principle underlying all data-driven approaches is the use of a model to classify various operating conditions of a machine, normal, fault one, fault two, etc. This is accomplished by first training the model on previously recorded data from each condition, and then feeding the model new data to be classified.

The strong and weak points common to all the data-driven approaches are first described. Then for each approach, the following paragraphs detail the principle, some guidelines to build the model and the specific strong and weak points, when necessary.

6.3.2 Strengths and weaknesses common to all data-driven approaches

Compared to knowledge-based diagnostic approaches, data-driven methods have the advantage of not requiring in-depth knowledge of the system to be diagnosed.

Additionally, they do not have constraints on the format of the independent variables. They can be dichotomous (e.g. on or off), categorical (e.g. machine state, such as “warm up”, “normal operation”, “max power”, “idle”) or continuous (e.g. sensed temperatures, pressures, speeds).

These types of modelling imply some drawbacks for machine diagnostics.

— Learning, calibration and potentially tuning phases are necessary prior to using the model. It may be necessary to repeat these phases, should the equipment or its usage evolve significantly or when new cases are encountered.

— A relatively large number of identified fault and non-fault samples are necessary to build the model.

— The model does not produce an explicative diagnosis.

— The model can be quite computationally intensive to train.

Several approaches (LR, neural network, classification tree, RF and SVM) are best used when there is no mathematical formulation, but rather a base of examples obtained from experience or simulation (tests of quality “pass-no pass”, testing of a chemical reactor, etc.).

6.3.3 Statistical data analysis and case-based reasoning

6.3.3.1 General description

The principle is to exploit the similarity between an observed situation and cases already known and resolved (“the present condition of the equipment resembles other cases”).

A case may be described using basic observed data (mono- or multidimensional, trends, patterns, etc.) or processed data (aggregated into symptoms).

This approach usually identifies a case or a set of cases in the database closely resembling the case one wishes to diagnose. These models require a learning phase based on good feedback, which is to say several well-described cases.

6.3.3.2 Building the model

The first step is to identify the data that will be used to describe a case.

The second is to select the type of similarity or relationship metric (commonly called distance) and classification mechanism that will support the resemblance between cases.

Finally, calibration and potentially iterative tuning of the model occurs during the learning period during which known cases are submitted to the system.

6.3.3.3 Specific strong and weak points

Compared to the other data-driven approaches, this approach requires the cases to be well structured and described.

6.3.4 Neural networks

6.3.4.1 General description

A neural network (also known as artificial neural network) is a non-linear statistical data model, which can be used to model complex relationships. Neural networks are associations of many elementary processors (neurons), whose collective behaviour approximates a given function.

6.3.4.2 Building the model

A neural network shall undergo a training phase where the model is tuned by examining normal and faulty data and changing the internal weights of the model such that the output of the model matches the state of the input data. Once the training process is completed, the network behaves as a “black box”, generating output related to the operational state of the machine (consistent with the training data) as new data are entered.

6.3.5 Classification trees

6.3.5.1 General description

Classification trees (Reference [9]) constitute a non-parametric technique that recursively partitions the data into subsets of greater class purity (normal, fault one, fault two, etc.), examining all variables and all possible splits at each node, and selecting the single best variable and value to split on. Not all variables are necessarily used to partition the data, and some variables may be used multiple times in different nodes.

Trees grown to 100 % node purity will naturally overfit the training data. To avoid this, cross-validation can be used to remove lower branches of the tree so that the classifier generalizes well to new data. This process is known as “pruning”.

6.3.5.2 Building the model

A classification tree shall undergo a training phase, i.e. the algorithm shall build the tree, generally to full purity, and then prune it back to optimal depth using cross-validation. Once the training and pruning steps are completed, new data for monitoring and diagnostic purposes can be run through the tree and classified.

6.3.6 Random forests

6.3.6.1 General description

Random forest (RF, Reference [7]) is a non-parametric classification method that applies bagging (Reference [8]) to a variation of classification trees (Reference [9]). A standard classification tree is constructed by splitting the data on the best of all possible features at each node. For RF, only a randomly selected subset (always chosen from the full set) of features is eligible to split each node. Moreover, each individual tree is constructed on a bootstrap sample of the data. A bootstrap sample is created from a particular data set by selecting cases at random with replacement. Each same-size bootstrap sample will be slightly different from one another. By random selection with replacement from the sample, some cases occur more than once in a bootstrap sample, and some cases do not occur at all.

Finally, in contrast to standard classification trees, the individual RF trees are not pruned; rather, they are typically grown to 100 % node purity. Although hundreds of trees may be developed, RFs are very quick to train (e.g. much faster than typical neural networks for a given data set and processor). Predictions are made by aggregating the predictions of the ensemble (majority vote for classification). Random forest generally exhibits a substantial performance improvement over the single classification tree classifier.

6.3.6.2 Building the model

A bootstrap sample (a sample, drawn with replacement, of the same length as the original data) of cases is chosen from the full set of cases. A small number of features (approximately the square root of the number of features) are randomly selected from the full set of features. The single best split point and feature is used to partition the sample into two more pure nodes. This feature selection and splitting procedure is repeated until each leaf node of the tree is at 100 % purity. Then a new bootstrap sample is selected, and a new tree is built for that sample. To classify a case, it is run through each tree, and the "votes" from each tree are counted. The case is classified as the class with the greatest number of votes.

6.3.6.3 Specific strong and weak points

Compared to classification trees, one substantial advantage of the RF algorithm is that it provides an estimate of variable importance, which can be used for variable selection.

Another substantial advantage is that because of the bagging procedure, RF has the advantage of being able to use all of the data to train on, i.e. it is not necesssary to reserve data for a test set for cross-validation, etc.

Another advantage of RF is that, except in rare cases, unlike many classification algorithms, RF cannot be overtrained. Rather, as training continues, performance oscillates (due to noise) around some minimum error rate.

Finally, in sharp contrast to many classification algorithms, RF has only two parameters to tune (number of trees and number of variables to examine at each node), and the performance of the RF is not very sensitive to these parameters (i.e. there is an optimum, but performance falls off slowly).

A weak point of RF is that, as with many data-driven classifiers, it is a "black box", i.e. given a particular case, it is not easy to explain how a particular classification was arrived at by examining the model.

6.3.7 Logistic regression (LR)

6.3.7.1 General description

Logistic regression (Reference [10]) models the relationship between a set of variables x_i to the expected value $E(y_i)$, with logistic function. The LR model can be expressed as:

$$E(y) = \frac{\exp(\alpha + \beta \boldsymbol{x})}{1 + \exp(\alpha + \beta \boldsymbol{x})} \tag{1}$$

After a transformation, it becomes

$$\ln\left[\frac{E(y)}{1 - E(y)}\right] = \alpha + \beta \boldsymbol{x} \tag{2}$$

Note that the dependent data are expected to be dichotomous (0 or 1), and that the logistic model output [the expected conditional mean, $E(y_i)$] is between 0 and 1. An appropriate threshold needs to be selected to transform the logistic regression output to a binary decision. Because of this, in the cases of multiple fault modes, it is necessary to use a series of models, typically with a supervisory model classifying normal or abnormal behaviour, and an ensemble of models classifying "fault one" or one of all other faults, "fault two" or all other faults, and so on.

6.3.7.2 Building the model

Maximum likelihood method is usually used to estimate the parameters in LR model. In practice, any good statistical package has the ability to generate an LR model.

6.3.7.3 Specific strong and weak points

The coefficients of the LR model provide insight into the importance of making a unit change in an independent variable.

Care shall be taken when selecting features to train the LR model on. Generally, some explicit "feature selection" procedure (e.g. employing cross-validation) is necessary to optimally tune the model.

6.3.8 Support vector machines (SVM)

6.3.8.1 General description

The support vector machine (Reference [11]) algorithm non-linearly maps an input characteristic data set into a higher dimensional set. A linear classifier is then constructed in the higher dimensional set to separate classes. Once the SVM is trained, it is very efficient for the evaluation of new cases.

6.3.8.2 Building the model

Gradient descent methods are generally used to estimate the coefficients in SVM models. In practice, there are many computer-based packages available that build SVM models.

6.3.8.3 Specific strong and weak points

SVMs have only one solution for a given problem (cf. neural networks, which can have multiple local optima). Moreover, in contrast to neural networks, the computational complexity of SVMs does not depend on the dimensionality of the input data.

The biggest drawback to SVM is that it is directly applicable for two-class problems only. Thus, for multi-class problems (e.g. multi-fault isolation problems), the problem needs to be portioned into multiple binary problems.

6.4 Knowledge-based approaches

6.4.1 Fault/symptom diagnostic approach

6.4.1.1 General description

This approach is based on the exploitation of fault/symptom relationships and is known as an associative knowledge model, because the relationships between faults and symptoms are associations. The diagnostic activity results from different tasks, each of which is devoted to a particular aspect. The main tasks are listed and explained below. Figure 3 gives an illustration of the phases of the fault/symptom association approach.

The starting point for the diagnostics is taken to be either:

— the presence of a real anomaly, alarm or abnormal behaviour; or

— a suspicion expressed like an anomaly in order to assess the condition of the machine.

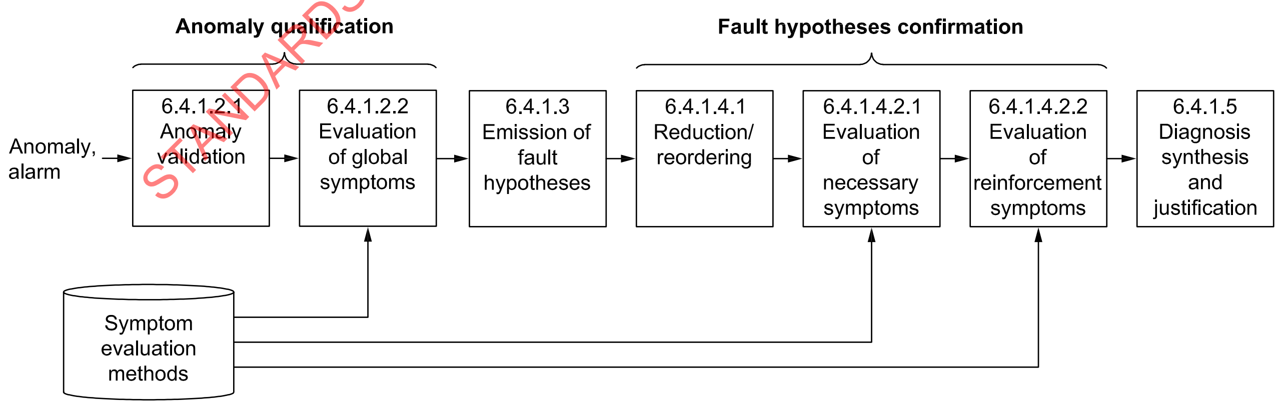


Figure 3 — Fault/symptom diagnostic approach

6.4.1.2 Qualification of the detected anomaly

6.4.1.2.1 Anomaly validation

The anomaly can be:

— derived from descriptors;

— an abnormal change in the data, without reaching alarm levels;

— a human perception of a change on the machine (noise, smell, temperature, moisture, leakage, etc.).

The process generally consists of validating the data from which the anomaly is derived (plausibility of the measurement, correlation with other measurements, alarm criterion check, transducer verification, etc.).

6.4.1.2.2 Evaluation of global symptoms

This step is intended to enable the production of fault hypotheses. A small set of global symptoms is evaluated. These symptoms, called macrosymptoms (grouping of symptoms), are evaluated using specified methods, as for symptoms.

6.4.1.3 Emission of fault hypotheses

Once macrosymptoms have been evaluated, the macrosymptoms/faults association is used to produce a list of fault hypotheses.

6.4.1.4 Confirmation of fault hypotheses

6.4.1.4.1 Reduction/reordering of the fault hypotheses list

This step is optional. It consists of reducing the diagnostic time. From the exhaustive list of fault hypotheses that has been found, a reduction or reordering can be made regarding the following:

— the probability of occurrence of the fault, from the feedback data, on the same type of machine, under the same service and operating conditions;

— the severity of the fault, from the criticality analysis.

When reducing the number of fault hypotheses, great expertise is needed, as the result can be an initial rejection of a fault hypothesis. (This is particularly the case for rare faults, which may, nevertheless, be critical.)

6.4.1.4.2 Evaluation of the fault hypotheses

6.4.1.4.2.1 Evaluation of necessary symptoms

All the necessary (i.e. required or "shall be present") symptoms are examined first. If all the necessary symptoms have been validated, then the fault hypothesis is validated. If one (or more) necessary symptom has been invalidated then the fault hypothesis is rejected.

When several methods exist to evaluate a symptom, the best performance method is preferred.

6.4.1.4.2.2 Evaluation of reinforcement symptoms

Once all the necessary symptoms have been validated, the reinforcement symptoms should be evaluated. These may reinforce the presumption of a particular fault at the final diagnosis step. Unlike the necessary symptoms, if one or more of the reinforcement symptom has not been validated, the fault is not rejected.

6.4.1.5 Diagnosis synthesis and justification

This is the last step in the diagnostics process. The objective is to summarize the realized diagnosis.

The elements that have been evaluated and validated are included in a formal diagnostic report. These elements include:

a) the anomaly that triggered the diagnostic;

b) the global symptoms that have been validated;

c) the rejected faults with the invalidated symptoms; and

d) the validated faults with their respective probabilities.

The report should also state other elements considered during the final stage of the synthesis phase. These elements are used to weight the validated hypotheses according to:

1) machine history;

2) similar cases encountered; and

3) probability and criticality of faults.

A conclusion should be reached. In this conclusion, the faults are given in reverse order of plausibility. A confidence factor (subjective, but based on all the objective previous elements) may be given for each one.

Corrective operation or maintenance actions should be proposed or, if maintenance work is needed but can be delayed, the delay is given and recommendations regarding operation are formulated, if required.

An example of a diagnostic report is given in Annex E.

6.4.2 Causal tree diagnostic approach

6.4.2.1 General description

When an in-depth knowledge of the mechanism of initiation and fault propagation is required, the simple fault-symptom approach is no longer satisfactory. A causal tree diagnostic approach might then be used.

When used in diagnostics, the causal tree analysis method is a process of determining the root cause based on an existing set of failure modes. Causal tree analysis flow-charting is normally used in the retrospective (diagnostic) sense in that the method is used to look at the “caused by” or “influenced by” relationship between failure modes. The data for this process already exist and therefore are not estimated. In the prognosis process, the method differs as the data have to be forecast.

A causal tree models the knowledge as follows:

— in the past, the root cause has “initiated” one or more failure modes;

— the relationship between failure modes can be described by “influence factors” or “initiation criteria”;

— failure mode symptoms can “initiate”, “influence other failure modes” or “have no effect”.

Figure 4 shows an example of a causal tree structure for diagnostics.

The links can be characterized by:

— a delay value representing the time lag between the causes and the effects; and

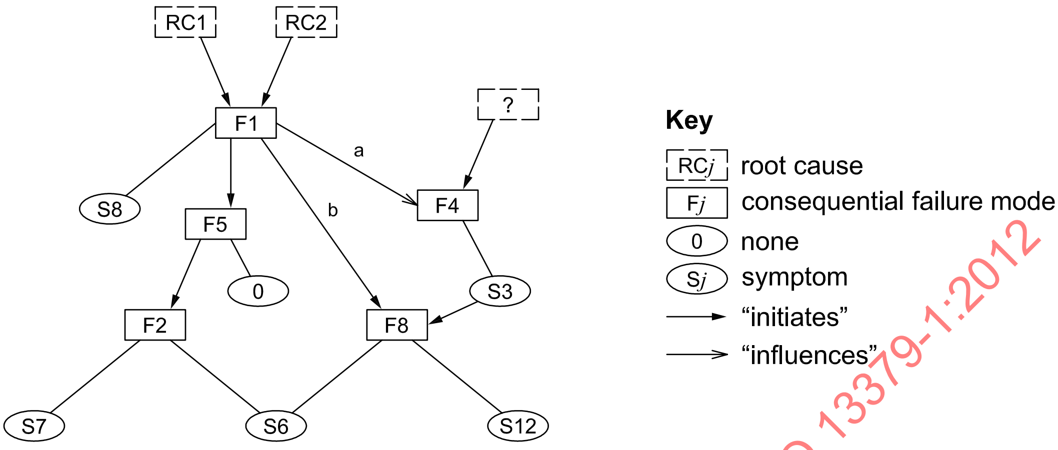
— a probability value representing the probability that this cause has this effect (“initiates” and “induces” only).

A causal tree model is rarely complete since:

— each fault is not systematically linked with a symptom; and

— the root cause of failure modes is not always known.

An example of causal tree modelling is given in Annex F.



a Probability, p = 0,75.

b The delay is 3 days.

Figure 4 — Example of causal tree modelling used for diagnostics

Typically, such models are exploited using logical solving methods: beginning with real observations, the model is run to identify possible causes; conclusions are then drawn as to their expected consequences and their occurrence confirmed or rejected by comparing actual observations with model projections. The process is then repeated until the prime causes (the initial faults) have been established.

6.4.2.2 Building the model

These models are based directly on transcriptions of expert knowledge about faults. The first step is therefore to identify whether such expertise exists and which experts should be consulted, before working with them to gather, formalize and validate the expertise.

NOTE The causal tree model can be transposed into a chart model.

6.4.2.3 Strong and weak points

These models are appropriate under the following circumstances.

— The expertise exists: this applies, therefore, to tested equipment in which significant faults have already been observed and analysed.

— Extensive quantitative feedback is not necessarily available (a typical example: large components such as the turbine or the reactor coolant pump set, for which there is little quantitative feedback on certain faults).

— It is desired to make an "explicative" diagnosis, i.e. one in which the faults identified justify the observations made.

— A visual representation of the fault relationship is desired.

— Detailed knowledge of the failure mechanism, modes, and behaviour is required. The information can be found in FMECA studies, which need to be regularly updated.

In this context, the choice of model depth, i.e. the level of detail sought (symptom/fault associations correspond to level 1 depth) is an important parameter:

— too little depth: the model's explicative capacity will be insufficient;

— too much: the model will be too complex, both for validation and for exploitation.

When the model becomes relatively complex, there may be too many possible combinations of processed information and the models are difficult to exploit.

6.4.3 First principle models

6.4.3.1 General description

The aim is to model the behaviour of equipment operating properly, on the basis of first principle relationships or mathematical equations (e.g. vibratory equation of rotors or civil engineering structures).

A model of system structure and component behaviour is built. It is then used to derive a description of the system expressed in logic. Combined with a set of observations of the system, this description provides the basis for computing a minimal set of possible hypotheses about component malfunctions that can explain observed deviations from the predicted (normal) behaviour.

These models can represent qualitative or quantitative information about expected physical behaviour.

This diagnostics process enables identifying sets of faulty components within the system.

6.4.3.2 Building the model

To build the model, an adequate level of decomposition into submodels has to be carried out in close interaction with designers or physicists.

Depending on the complexity of the various equations representing the system, it is necessary to elaborate adapted resolution algorithms.

6.4.3.3 Strong and weak points

This type of model is appropriate when operation of the equipment can be described as a combination of simple “transfer functions”. It is well suited to operation, which can be expressed in terms of “fluxes” (of data, current, fluid, etc.). They do not require feedback on faults, so such models can be used for equipment still in the design phase. They have been widely used in the domain of electrical, in general, and more precisely digital circuit troubleshooting, as well as in automotive engines. A model of normal behaviour implies the possibility of using the same model not only for diagnosis, but also for FMEA, test generation, diagnosis analysis, etc.

The approach has a drawback for machine diagnostics in that the initial set-up effort and cost is high.

This drawback is offset by the ease of modification once the initial model has been created and the flexibility with which the model can be used (FMECA generation).

6.5 Confidence factor determination

This figure of merit essentially represents the cumulative effect of error sources on the final certainty of confidence in the accuracy of the diagnosis. It can be determined algorithmically or via a weighted assessment system. An example of a weighted assessment is given in Annex G.

The confidence factor should be determined from the following elements:

— maintenance history including experience of same faults on similar machines;

— design and failure modes assessment;

— analysis technique or descriptor used;

— severity limits used;

— measurement interval;

— database set-up;

— data acquisition;

— severity assessment process;

— trend assessment;

— diagnosis process.

Annex A
(informative)

Failure mode and symptoms analysis (FMSA)

Table A.1 — Analysis sheet

Part or assembly name						Suppliers					FMSA number __________	
Part or assembly number											Sheet __________ of __________	
Drawing issue												
											FMSA Committee	
Amendments	1	2	3	4	5	6	7	8	9	10	FMSA Approved	Date
Date											Name	Signature

Item	Part No. name issue	Function or process	Failure mode	Effect of failure	Cause of failure	Failure symp- toms	Primary tech- nique	Fre- quency of moni- toring	Primary MPN[a]					Correla- tion tech- niques	Fre- quency of moni- toring	Correlation MPN[a]				
									DET	SEV	DGN	PGN	MPN			DET	SEV	DGN	PGN	MPN

Table A.1 — Table A.1 (*continued*)

FMSA number __________

Sheet __________ of __________

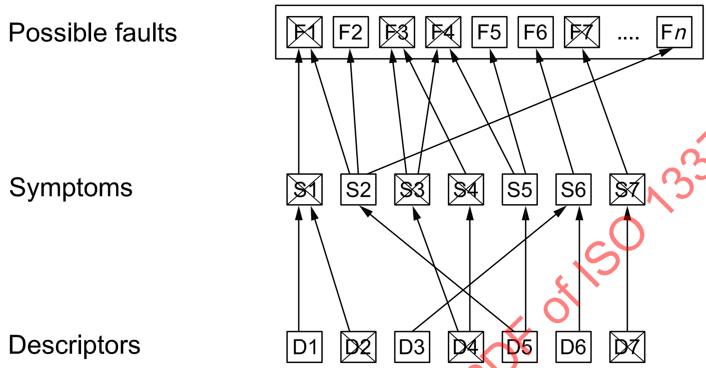
Item	Part No. name issue	Function or process	Failure mode	Effect of failure	Cause of failure	Failure symp-toms	Primary tech-nique	Fre-quency of moni-toring	Primary MPN[a]					Correla-tion tech-niques	Fre-quency of moni-toring	Correlation MPN[a]				
									DET	SEV	DGN	PGN	MPN			DET	SEV	DGN	PGN	MPN

[a] DET ≡ detection; SEV ≡ severity; DGN ≡ diagnosis; PGN ≡ prognosis; MPN ≡ monitoring priority number.

Annex B
(informative)

Effectiveness of the diagnostics system

Each fault can be diagnosed if its symptoms (and thus the descriptors used to evaluate these symptoms) are available. See Figure B.1.



Key

An uncrossed box indicates:

D descriptor

F possible fault

S symptom

A crossed box indicates:

D descriptor not present

F fault not diagnosed

S symptom not verified

Figure B.1 — Example of faults/symptoms/descriptors relationship

Assuming fault F_i has probability of occurrence p_i and severity S_i, it is possible to express a performance criterion for the overall diagnosis process, the diagnostic system effectiveness (DSE), as:

$$\frac{\sum_{D_F} S_i p_i d_i}{\sum_{F} S_i p_i} \tag{B.1}$$

where

F is the possible faults set obtained by the FMEA or FMECA analysis;

D_F is the diagnosable faults set, a subset of F;

d_i (between 0 and 1) is the diagnostic reliability.

The severity, S_i, may be obtained by:

$$S_i = r_\mathrm{f} f_\mathrm{c} f_\mathrm{s} f_\mathrm{sd} \tag{B.2}$$

where

r_f is the failure rate (i.e. number of failures per hour);

f_c is the cost factor, including maintenance and unavailability costs, ranked from 1 to 3 (low, medium, high);

f_s is the safety factor, ranked from 1 to 3 (low, medium, high);

f_{sd} is the secondary damage factor, ranked from 1 to 3 (low, medium, high).

Annex C
(informative)

Comparative analysis of diagnostic models

Diagnostic method	Knowledge used	Strong points	Weak points	Typical applications and references
Rule-based	Human expertise	— Relatively simple to implement	— Incompleteness — Difficulty in explaining multiple faults — Poor explicative capacity — Brittleness to system changes	— Rotating machinery diagnosis — Medical diagnosis
Causal fault	Description of fault mechanism and propagation	— Explicative diagnosis — Handling of multiple independent faults	— Requires good knowledge of possible faults (tested equipment) — Incomplete	— Rotating machinery diagnosis — Medical diagnosis
First principles	Decomposition and transfer function of equipment	— Does not require knowledge of faults (new equipment) — Handles multiple faults well — Gives flexibility to system modification, FMEA, test generation, diagnosis analysis	— Non-explicative diagnosis — Possible aberrant diagnosis — Model complexity in certain domains	— Electronic or fluid circuit diagnosis — Automotive engines and control systems
Statistical **Case-based reasoning**	Samples of significant past diagnosis cases	— Approach well understood — Does not require in-depth knowledge of dysfunctions	— Difficulty in obtaining a sufficient number of significant, well-described cases	— Aeroplane engine diagnosis
Classification trees **Random forests (RFs)** **Logistic regression (LR)** **Neural networks** **Support vector machines (SVMs)**	Samples of significant past diagnosis cases and associated data	— Does not require in-depth knowledge of dysfunctions — RF can accommodate missing data	— Non-explicative diagnosis — Difficulty in obtaining a sufficient number of significant, well-described cases	— Any application

Annex D
(informative)

Most commonly used diagnostic models by monitoring technique

Diagnostic model/ monitoring technique	Knowledge-based			Data-driven						
	Rule-based	Causal fault	First principle	Statistical methods	Case-based reasoning	Neural network	Classification trees	Random forest	Logistic regression	Support vector machines
Vibration	M	D	P	M	D	D	—	D	—	—
Thermography	M	—	—	M	—	D	—	P	—	—
Oil analysis	M	P	—	M	D	D	—	D	D	D
Process parameters	M	—	D	M	M	M	M	M	M	M
Performance	M	—	D	M	M	M	M	M	M	M
Acoustic emission	M	—	—	M	—	D	P	D	—	—
Acoustic monitoring	M	—	—	M	—	D	—	D	—	—
Electrical monitoring	M	—	—	M	—	D	—	—	—	—

M: Mature and commonly applied in industrial applications.

D: Under development and some initial applications.

P: Promising and potential.