ANNALS of the ORADEA UNIVERSITY.
Fascicle of Management and Technological Engineering, Volume VII (XVII), 2008

ASPECTS OF MODEL-BASED DIAGNOSTICS IN CONDITION BASED MAINTENANCE

Gheorghe DONCA, Ioan MIHĂILĂ, Marius NICA
University of Oradea

Keywords : condition based maintenance, model-based, sensors, diagnostics, prognostics

Abstract : The objective of maintenance is to reduce the number of unexpected breakdowns due to failures, which may be catastrophic and may incur huge loss. Many industrial companies have shifted their maintenance programs to condition-based maintenance (CBM), which, if correctly and effectively implemented, can significantly reduce the maintenance cost by cutting down the number of unnecessary scheduled preventive maintenance operations. Model-based fault and health diagnostics and prognostics of systems health evolution will help to achieve best reliability for the new complex systems.

1. INTRODUCTION

The most important step between condition monitoring and condition based maintenance is the diagnosis module. Interest in diagnosing and prognosticating faults in engineering systems is as old as engineering systems themselves. Designers and users alike have an interest in preventing the occurrence of *failure* of a mechanism, a machine or any kind of device. Several approaches can be taken, the most obvious of which is to stop the system whenever an *anomaly* is *observed*, i.e., a *fault* is *detected* as a difference in the performance of the system from its *normal behavior*. The alternative approaches consider a variety of situations. For example, what happens when an operator cannot sense or detect the fault? Or, what should be done if it were desired to keep a machine running while the damage is not yet critical? In many situations, making the correct diagnosis is a life or death decision, like when an aircraft in flight undergoes damage.

There are several ways of classifying approaches to the problem of diagnosing an engineering system. Health assessment operations can be classified as either online or offline, depending on whether they are performed while a system is in operation or not, respectively [1].

Another way of classifying diagnostic techniques depends on whether the diagnosis assessment is based on deterministic information (e.g., one obtained from a model) or on stochastic information (e.g., historical, statistical data). The first of these two has been termed a “white box” approach, while the second has been called a “black box” approach.

We can think of existing solutions to the problems of performing diagnostics and prognostics as belonging to one (or perhaps even both) of two types: *data-driven* – also called *model-free* – techniques and *model-based* techniques, although other classifications exist. Data-driven techniques include, for example, signal processing algorithms and knowledge-based methodologies. Model-based techniques more commonly involve the description of a system through mathematical models of the physical laws governing its behavior.

Data-driven techniques rely on comparative assessments of the status of a system under testing with other known occurrences. For as long as the behavior of the system under testing remains similar to that of a previously known, healthy configuration, the former is deemed to be healthy. When the measured behavior deviates from this reference, a fault is detected, and a comparison with the conditions previously observed in analogous faulted systems can take place. Under the appropriate conditions, this new comparison hast the potential to isolate and identify the fault efficiently. Thus, the ability of data-driven techniques to perform the task of diagnosis is given by the *training* of a classification algorithm.

The training algorithms used by data-driven decision processes are highly automated tasks for which extensive literature exists. Intelligent algorithms in support of this duty abound, and implementation is generally a straightforward and proven activity. Even more appealing is the fact that the data-driven effort typically avoids the need to understand the underlying physical mechanisms that describe the behavior of a system; diagnostics are performed regardless of the causes of a fault. Furthermore, data-driven algorithms can continue to ,,learn" as they operate, ideally making their assessments more reliable with each fault detection attempt.

The model-based approach is generally more robust in the sense that it can sort out new or unforeseen situations more easily, since this technique can incorporate and replicate, per its mathematical models, a wider range of behaviors, even if previously unobserved in actual systems. If the state of a system deviates from expected operational ranges, model-based techniques can continue to work by updating physical parameters that describe the new situation. Because of this ability, model-based techniques can also dispense with the use of the extensive training and historical information required by the data-driven approach, and is less prone to the kind of difficulties introduced by under- or over-training.

Yet, setting up an accurate model that describes the physics-of-failure mechanisms (i.e., a physics-based model) requires so much effort and expertise, that many simplifying assumptions have to be made, which degrade the model's reliability or applicability to the real situation. This effort is typically beyond that required by data-driven techniques. All the observed occurrences of a fault in past instances become useless to the modeling effort if the physics behind such behavior are not well understood. A comparison between the applicability of the data-driven and the model-driven approaches is given graphically in Figure 1 [1].

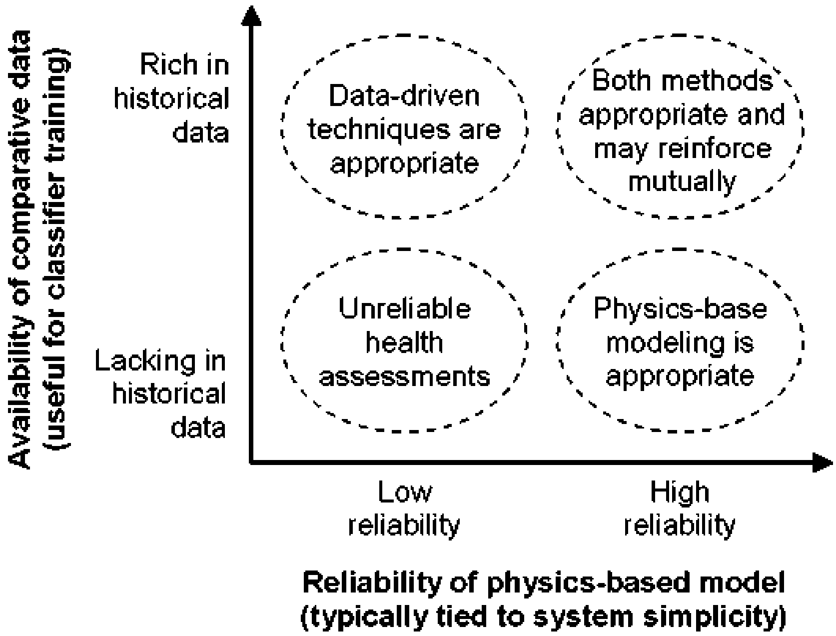

Fig. 1. Comparison of situations where the physics-based approach or the data-driven approach are appropriate [1].

2. MODEL-BASED DIAGNOSTICS

Quantitative engineering diagnosis techniques operate by comparing a particular measurement from a plant with an expected value. When there is a difference, the plant is declared to be at fault. The comparison is typically performed using one of two techniques, *parameter estimation* or *residual evaluation*, as represented in Figure 2 [3].

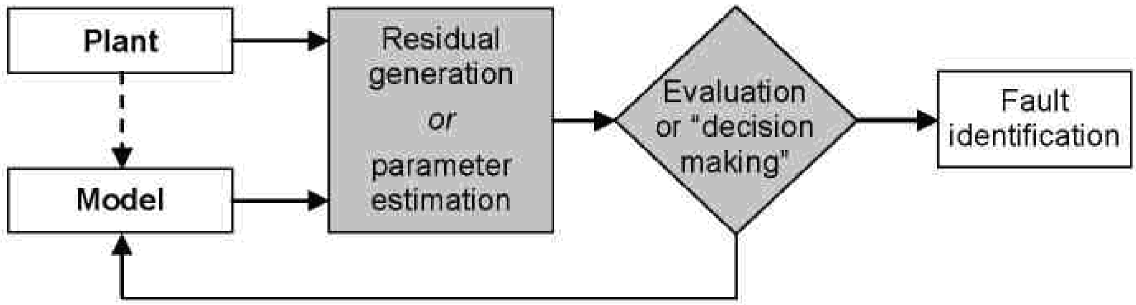

Fig. 2. Overview of the general approach to performing model-based fault detection [3].

Although the "model" block of Figure 2 can refer to a mathematical description of a system, this need not be so. This model can be any realization that replicates certain behaviors (i.e., variables, signals or any other measurable activity) of an engineering system; it may be, for example, a computer simulation tying different processes, even if described without equations; it may be linear or non-linear; it may be deterministic or stochastic, etc. The approach of parameter estimation for performing model-based diagnostics is represented in Figure 3 [5]. A *system identification* model is updated with observations from the plant to try to determine some internal parameters of the plant's process and ensure that they remain within specified bounds, or else a fault may be declared.

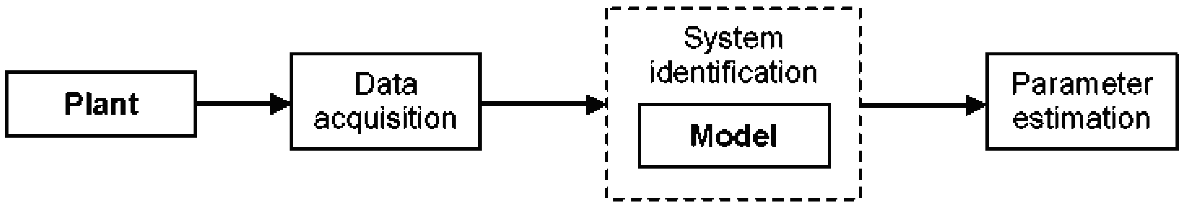

Fig. 3. Parameter estimation procedure [5].

A model to be used for realizing model-based diagnostics must provide the ability to simulate an engineering system under varying types of faults and varying amounts of damage, because, comprehensive fault diagnostics involve not just the aspect of *fault detection*, but *fault isolation* and *fault identification* as well. The design of a diagnostic architecture that can use a so-called physics-based model to simulate the changing behavior of a system with a developing fault is presented in Figure 4, which is *generally applicable to many kinds of systems*.

Any model can be cast in the form of a function, with inputs, outputs, variables and parameters. Physics-based models are functions referencing and relating variables and parameters having physical significance, i.e., that can be associated with physical laws governing the behavior of the system that is modeled, like the laws of motion, mechanics, electromagnetism, etc. To be able to use a physics-based model to replicate the behavior of a system experiencing a fault, we need to apply values that are indicative of the presence of the fault to some of the variables and parameters. This is the reason why it is fitting to classify these according to whether they are sensitive to the fault or not, as presented in Figure 4. This classification will allow the study of the effect of a fault to focus on significant components of the model while simultaneously isolating them to facilitate their consideration.

The most basic underlying assumption is that the model can replicate the behavior of the system under the fault. Models are often available for operational engineering systems, but it is more difficult to ensure that they truly reflect the changing behavior of a system when a fault is present. If the model were not ready to do this, then it must be modified to allow it to simulate the fault conditions. In many occasions, system models based in physical behaviors should have this ability, but in the design stage of an engineering system, the model is reduced or simplified to simulate the normal operating conditions that designers wish to study. Another assumption is that the parameters describing the system have been correctly classified into fault-sensitive or fault-insensitive. This may not be simple, however. It is further assumed that it is possible to study the response of the fault-sensitive parameters to a fault, and that the fault-insensitive parameters either are known or can be determined experimentally.

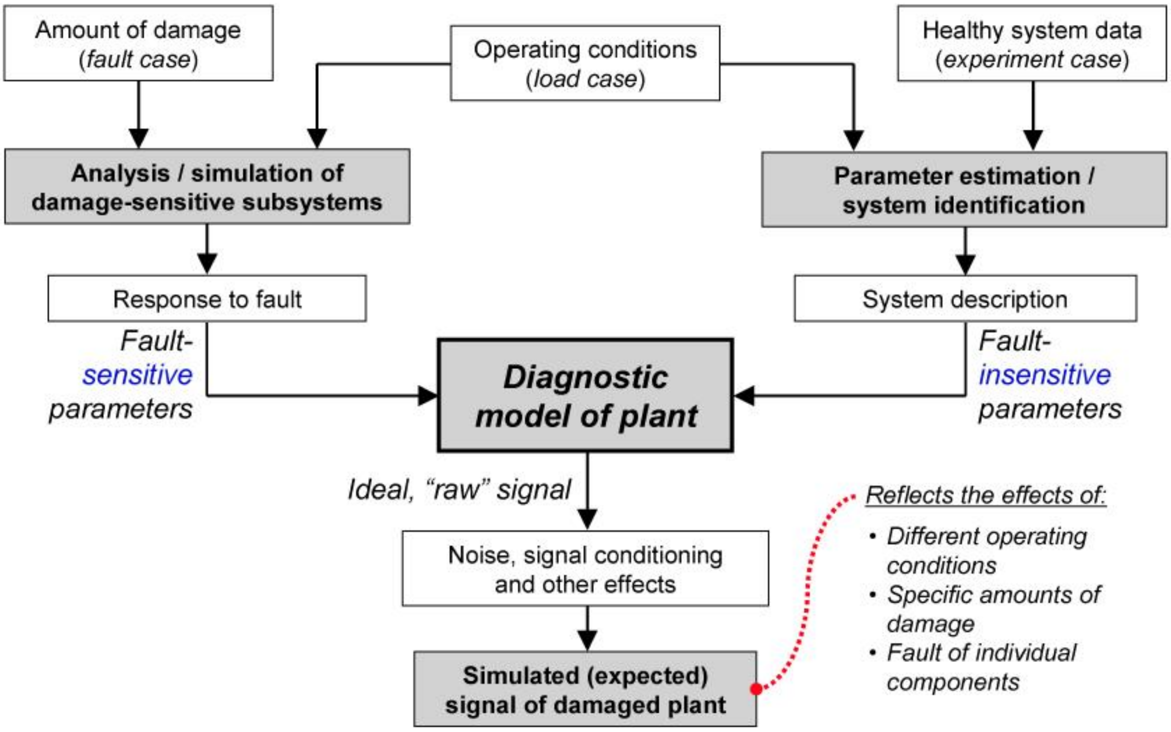

Fig. 4. Suggested architecture for performing model-based simulation of systems [3].

Experimental data must be used to approximate all the *fault-insensitive parameters* describing the system. Of course, experimental data are used only when the values of these parameters are not readily known quantities like invariant material properties or physical constants. This step is included considering that we want to calibrate a model that requires the values of parameters that are specific to a particular system realization or arrangement. The determination of the fault-insensitive parameters is done through any reliable technique for parameter estimation or system identification. These parameters should be determined for particular cases of experimental data that were obtained at known operating conditions of interest. These known operating conditions, given by the application of a particular *loading case*, are represented in the diagram of Figure 4 in the center-top block. Because the parameters to be determined are supposed to be insensitive to the fault, it should not matter whether the plant from which the experimental data were retrieved was at fault or not. However, if some of the parameters are sensitive to the fault to some extent but for reasons of simplicity were assumed insensitive, then it is up to the diagnostic system designer to choose whether it would be better to use experimental data from a system with or without a fault. The decision would be made after considering which of the two instances (with or without a fault) reflects the most adequate values of the parameters chosen as "fault-insensitive". Data from one of these particular *experimental* *cases* enter the methodology on the upper-right corner of the diagram of Figure 4.

To use the model for simulating the behavior of the plant under a fault with a particular degree of development, one must quantify this development as a specific *amount of damage* in the system. The exact amount of damage is used to describe the size of a particular fault, i.e., it is a specific *fault case*, and is entered in the diagnostic algorithm of Figure 4 through the top-left corner. This would be done if the effect of a fault is negligible in such parameters.

The operating conditions (load case) and the amount of damage (fault case) are used to analyze the response of *different components or subsystems* of the plant to the presence of the specific fault size considered. The objective of this analysis is to determine the values of *fault-sensitive parameters* describing the system for the model. Such an analysis is particular to the plant, and may require support from a variety of engineering disciplines.

Once the fault-insensitive and fault-sensitive parameters have been determined, the parameters are input to the plant model at the center of Figure 4. The model is used to simulate the behavior of the plant for a specific fault case in individual components or subsystems and under a particular load case. It is possible to run simulations for particular faults and loads because the parameters were determined also for particular faults and loads. The simulation will produce a signal, or the values of a set of system variables, replicating those from the plant under similar conditions. However, in the real world, the corresponding values of this signal or set of variables are observed (i.e., measured) in the plant after they have been affected by different events, including noise, sensor response characteristics, signal processing, etc. Thus, the signals or variables obtained from the model should be treated or modified in a similar manner to ensure that the simulation truly replicates the plant and allow for fair comparisons between the model and plant. Calibration of the model-based diagnostic architecture is hence due here. Take for example the case of noise. If we know that the signal the plant produces is distorted by extraneous signals that can be characterized with a statistical distribution, the same distribution may be added to the model signal to have it approximate better the behavior of the plant. Omitting the calibration or signal conditioning step described has the potential to invalidate any comparisons between simulation results and plant measurements.

The architecture of Figure 4 is used to simulate the behavior of a plant for a specific fault case in individual components or subsystems and under a particular load case. The simulation produces a signal or values that replicate those of the plant when operating under such conditions. However, the plant may not be operating under such conditions. It is the intent of a diagnostic effort to determine the conditions of the plant, and the simulation does not directly provide information about this. To arrive at a diagnosis of the plant, the signal or values resulting from the model simulations must still be interpreted and compared to those coming from the plant.

It is useful for determining whether a fault exists or not in a plant. This detection approach was depicted earlier in Figure 2. We are here interested, however, in assessing the size of the fault or the severity of damage as well, i.e., we want to perform *fault identification* in addition to fault detection. The idea is to use the model repetitively with different amounts of damage to generate a series of residuals or parameter estimates. The process stops when the evaluation block decides that a residual or parameter estimate has an “adequate” value. At that point, the fault size used in the model to generate the residual or parameter estimate is deemed to approximate the fault size present in the plant. By “adequate value”, different things are meant when using parameter estimation or residual generation. In the case of parameter estimation, an adequate value would be that of a parameter remaining within previously specified bounds. For residual generation, the adequate value approaches zero.

The unknown amount of damage present in the plant is assessed by running repeated trials with different amounts of damage in the model. The trial with an amount of damage in the model that minimizes the absolute value of the residual is taken to be the amount of damage present in the plant.

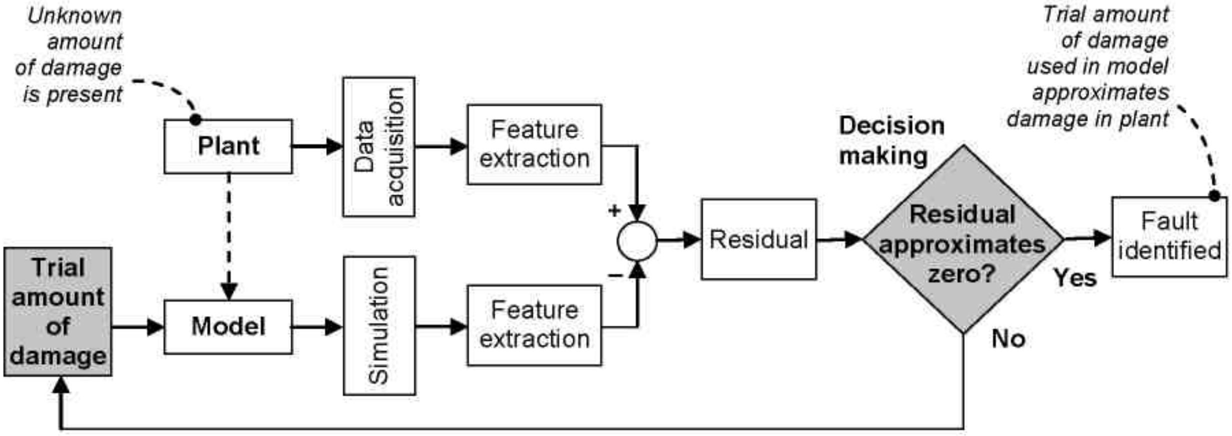

Fig. 5. Reverse engineering to fault identification through residual evaluation [3].

This approach to modeling makes it possible to use feature extraction techniques to diagnose a specific size of fault in the plant by searching for a correspondence of the behaviors observed in the plant among the behaviors of several simulation trials. The simulation trials represent varying degrees of damage, and the one that exhibits the closest correspondence is taken to be the one representing the size of the fault in the plant. Thus, this “searching” technique can be likened to a *reverse engineering process* of fault identification [3]. The model must be able to replicate the behaviors that the plant would exhibit under different amounts of damage and specific operating conditions.

3. CONCLUSIONS

The performance of CBM systems relies on the diagnostic and prognostic techniques used and the sensors selected. First step for an intelligent CBM is the diagnostics module. The most important disadvantages of the model-based approach to diagnostics is that modeling can be a complex and involved task. The models for diagnostics required much consideration and careful adaptations to enable their practical implementation. To implement the diagnostic architecture, it’s need to characterize changes in the sensors signal and establish adequate *feature extraction* routines. An integrated model-based diagnostic and prognostic architecture is the future for all complex and intelligent CBM systems.

REFERENCES

[1] Inman, J. D., et al., *Damage Prognosis*, John Wiley & Sons, Ltd, 2005.
[2] Kothamasu, R., *Intelligent Condition Based Maintenance - A Soft Computing Approach to System Diagnosis and Prognosis*, Georgia Institute of Technology, 2005.
[3] Patrick, R., *A Model Based Framework for Fault Diagnosis and Prognosis of Dynamical Systems with an Application to Helicopter Transmissions*, Georgia Institute of Technology, 2007.
[4] Takata, S., Kimura, F., Houten, F., Westkämper, E., Shpitalni, M., Lee, J., *Maintenance: Changing role in life cycle management*, Manufacturing Technology 53, Nr. 2, 2004.
[5] Vachtsevanos, G., F. L. Lewis, M. J. Roemer, A. Hess, B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, Hoboken, NJ, John Wiley & Sons, 2006.