CHAPTER 1
INTRODUCTION

There is an increasing interest in theory and applications of model-based fault detection and fault diagnosis methods, because of economical and safety related matters. In particular, well–established theoretical developments can be seen in many contributions published in the IFAC (International Federation of Automatic Control) Congresses and IFAC Symposium SAFEPROCESS (Fault Detection, Supervision and Safety of Technical Processes) [Isermann and Ballé, 1997, Isermann, 1997, Patton, 1999, Frank *et al.*, 2000].

The developments began at various places in the early 1970s. Beard [Beard, 1971] and Jones [Jones, 1973] reported, for example, the well–known "failure detection filter" approach for linear systems.

A summary of this early development is given by Willsky [Willsky, 1976]. Then Rault and his staff [Rault *et al.*, 1971] have considered the application of identification methods to the fault detection of jet engines. Correlation methods were applied to leak detection [Siebert and Isermann, 1976].

The first book on model-based methods for fault detection and diagnosis with specific application to chemical processes was published by Himmelblau [Himmelblau, 1978]. Sensor failure detection based on the inherent analytical redundancy of multiple observers was shown by Clark [Clark, 1978].

The use of parameter estimation techniques for fault detection of technical systems was demonstrated by Hohmann [Hohmann, 1977], Bakiotis [Bakiotis *et al.*, 1979], Geiger [Geiger, 1982], Filbert and Metzger [Filbert and Metzger, 1982].

The development of process fault detection methods based on modelling, parameter and state estimation was then summarised by Isermann [Isermann, 1984] and [Isermann, 1997]

Parity equation-based methods were treated early [Chow and Willsky, 1984], and then further developed by Patton and Chen [Patton and Chen, 1994b], Gertler [Gertler, 1991], Höfling and Pfeufer [Höfling and Pfeufer, 1994].

Frequency domain methods are typically applied when the effects of faults as well as disturbances have frequency characteristics which differ from each other and thus the frequency spectra serve as criterion to distinguish the faults [Massoumnia *et al.*, 1989, Frank *et al.*, 2000, Ding *et al.*, 2000].

The developments of fault detection and isolation methods to the present time is summarised in the books of Pau [Pau, 1981], then Patton *et al.* [Patton *et al.*, 2000], Basseville and Nikiforov [Basseville and Nikiforov, 1993], Chen and Patton [Chen and Patton, 1999], Gertler [Gertler, 1998], Isermann [Isermann, 1994b] and in survey papers by Gertler [Gertler, 1988], Frank [Frank, 1990] and Isermann [Isermann, 1994a].

Within IFAC, the increasing interest in this field was taken into account by creating first in 1991 a SAFEPROCESS (Fault Detection Supervision and Safety for Technical Processes) Steering Committee which then became a Technical Committee in 1993.

The first IFAC SAFEPROCESS Symposium was held in Baden–Baden, Germany in 1991 [Isermann and Freyermuth, 1992], and the second in Espo, Finland in 1994. The third symposium was scheduled at Hull, UK in 1997 and the fourth one was held in Budapest, Hungary in June 2000. The fifth is expected at Washington DC in July 2003.

Another tri–ennial series of IFAC Workshop exist for "Fault detection and supervision in the chemical process industries". Workshops were held in Newark, Delaware, Newcastle UK, Lyon and Korea between 1992 and 2001.

Most contributions in fault diagnosis rely on the analytical redundancy principle. The basic idea consists of using an accurate model of the system to mimic the real process behaviour. If a fault occurs, the residual signal (*i.e.* the difference between real system and model behaviour) can be used to diagnose and isolate the malfunction.

Model-based method reliability, which also includes false alarm rejection, is strictly related to the "quality" of the model and measurements exploited for fault diagnosis, as model uncertainty and noisy data can prevent an effective application of analytical redundancy methods.

This is not a simple problem, because model-based fault diagnosis methods are designed to detect any discrepancy between real system and model behaviours. It is assumed that this discrepancy signal is related to (has a response from) a fault. However, the same difference signal can respond to model mismatch or noise in real measurements, which are erroneously detected as a fault. These considerations have led to research in the field of "robust" methods, in which particular attention is paid to the discrimination between actual faults and errors due to model mismatch.

On the other hand, the availability of a "good" model of the monitored system can significantly improve the performance of diagnostic tools, minimising the probability of false alarms.

This monograph is devoted to the explanation of what is a "good" model suitable for robust diagnosis of system performance and operation. The book also explains how "robust models" can be obtained from real data. A large amount of attention is paid to the "real system modelling problem", with reference to either linear and non–linear model structures. Special treatment is given to the case in which noise affects the acquired data. The mathematical description of the monitored system is obtained by means of a system identification scheme based on equation error and errors–in–variables models. This is an identification approach which leads to a reliable model of the plant under investigation, as well as the estimation of the variances of the input–output noises affecting the data.

The purpose of the monograph is to provide guidelines for the modelling and identification of real processes for fault diagnosis. Hence, significant attention is paid to practical application of the methods described to real system studies, as reported in the last chapters.

In particular, this first chapter of the book outlines a new a common terminology in the fault diagnosis framework and gives some discussion and summary of developments in the field of fault detection and diagnosis based on papers selected during 1991–2001.

1.1 Nomenclature

By going through the literature, one recognises immediately that the terminology in this field is not consistent. This makes it difficult to understand the goals of the contributions and to compare the different approaches.

The SAFEPROCESS Technical Committee therefore discussed this matter and tried to find commonly accepted definitions. Some basic definitions can be found, for example, in the RAM (Reliability, Availability and Maintainability) dictionary [RAM, 1988], in contributions to IFIP (International Federation for Information Processing) [IFI, 1983].

Some of the terminology used in this book is given below. These are based on information obtained from the SAFEPROCESS Technical Committee and are considered "on–going" in the sense that new definitions and updates are being made.

1. ***States and Signals***

   Fault

   An unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable, usual or standard condition.

   Failure

   A permanent interruption of a system's ability to perform a required function under specified operating conditions.

   Malfunction

   An intermittent irregularity in the fulfilment of a system's desired function.

   Error

   A deviation between a measured or computed value of an output variable and its true or theoretically correct one.

Disturbance
An unknown and uncontrolled input acting on a system.

Residual
A fault indicator, based on a deviation between measurements and model-equation-based computations.

Symptom
A change of an observable quantity from normal behaviour.

2. ***Functions***

Fault detection
Determination of faults present in a system and the time of detection.

Fault isolation
Determination of the kind, location and time of detection of a fault. Follows fault detection.

Fault identification
Determination of the size and time-variant behaviour of a fault. Follows fault isolation.

Fault diagnosis
Determination of the kind, size, location and time of detection of a fault. Follows fault detection. Includes fault detection and identification.

Monitoring
A continuous real-time task of determining the conditions of a physical system, by recording information, recognising and indication anomalies in the behaviour.

Supervision
Monitoring a physical and taking appropriate actions to maintain the operation in the case of fault.

3. ***Models***

Quantitative model
Use of static and dynamic relations among system variables and parameters in order to describe a system's behaviour in quantitative mathematical terms.

Qualitative model
Use of static and dynamic relations among system variables in order to describe a system's behaviour in qualitative terms such as causalities and IF–THEN rules.

Diagnostic model
A set of static or dynamic relations which link specific input variables, *the symptoms*, to specific output variables, the faults.

Analytical redundancy
Use of more (not necessarily identical) ways to determine a variable, where one way uses a mathematical process model in analytical form.

4. ***System properties***

   Reliability

   Ability of a system to perform a required function under stated conditions, within a given scope, during a given period of time.

   Safety

   Ability of a system not to cause danger to persons or equipment or the environment.

   Availability

   Probability that a system or equipment will operate satisfactorily and effectively at any point of time.

5. ***Time dependency of faults***

   Abrupt fault

   Fault modelled as stepwise function. It represents bias in the monitored signal.

   Incipient fault

   Fault modelled by using ramp signals. It represents drift of the monitored signal.

   Intermittent fault

   Combination of impulses with different amplitudes.

6. ***Fault terminology***

   Additive fault

   Influences a variable by an addition of the fault itself. They may represent, *e.g.*, offsets of sensors.

   Multiplicative fault

   Are represented by the product of a variable with the fault itself. They can appear as parameter changes within a process.

1.2 Fault Detection and Identification Methods based on Analytical Redundancy

A traditional approach to fault diagnosis in the wider application context is based on *hardware or physical redundancy* methods which use multiple sensors, actuators, components to measure and control a particular variable. Typically, a voting technique is applied to the hardware redundant system to decide if a fault has occurred and its location among all the redundant system components. The major problems encountered with hardware redundancy are the extra equipment and maintenance cost, as well as the additional space required to accommodate the equipment [Isermann and Ballé, 1997, Isermann, 1997].

In view of the conflict between reliability and the cost of adding more hardware, it is possible to use the dissimilar measured values together to cross-compare each other, rather than replicating each hardware individually. This is the meaning of *analytical or functional redundancy*. It exploits redundant analytical relationships among various measured variables of the monitored process [Patton *et al.*, 1989, Chen and Patton, 1999].

In the analytical redundancy scheme, the resulting difference generated from the comparison of different variables is called a *residual or symptom signal*. The residual should be zero when the system is in normal operation and should be different from zero when a fault has occurred. This property of the residual is used to determine whether or not faults have occurred [Patton *et al.*, 1989, Chen and Patton, 1999].

Consistency checking in analytical redundancy is normally achieved through a comparison between a measured signal with estimated values. The estimation is generated by a mathematical model of the considered plant. The comparison is done using the residual quantities which are computed as differences between the measured signals and the corresponding signals generated by the mathematical model [Patton *et al.*, 1989, Chen and Patton, 1999].

Figure 1.1 illustrates the concepts of hardware and analytical redundancy.

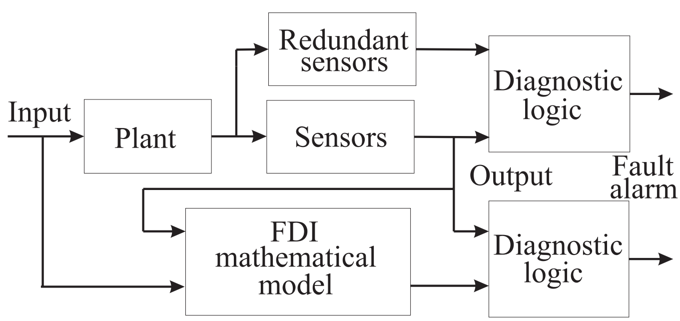

Fig. 1.1. Comparison between hardware and analytical redundancy schemes.

In practice, the most frequently used diagnosis method is to monitor the level (or trend) of the residual and take action when the signal reaches a given threshold. This method of *geometrical analysis*, whilst simple to implement, has a few drawbacks. The most serious is that, in the presence of noise, input variations and change of operating point of the monitored process, false alarms are possible.

The major advantage of the model-based approach is that no additional hardware components are required in order to realize a Fault Detection and Isolation (FDI) algorithm. A model–based FDI algorithm can be implemented via software on a process control computer. In many cases, the measurements necessary to control the process are also sufficient for the FDI algorithm so that no additional sensors have to be installed [Patton *et al.*, 1989, Chen and Patton, 1999, Basseville and Nikiforov, 1993].

Analytical redundancy makes use of a mathematical model of the system under investigation and it is therefore often referred to as the *model–based approach* to fault diagnosis.

1.3 Model-based Fault Detection Methods

The task consists of the detection of faults on the technical process including actuators, components and sensors by measuring the available input and output variables $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$. The principle of the model–based fault detection is depicted in Figure 1.2.

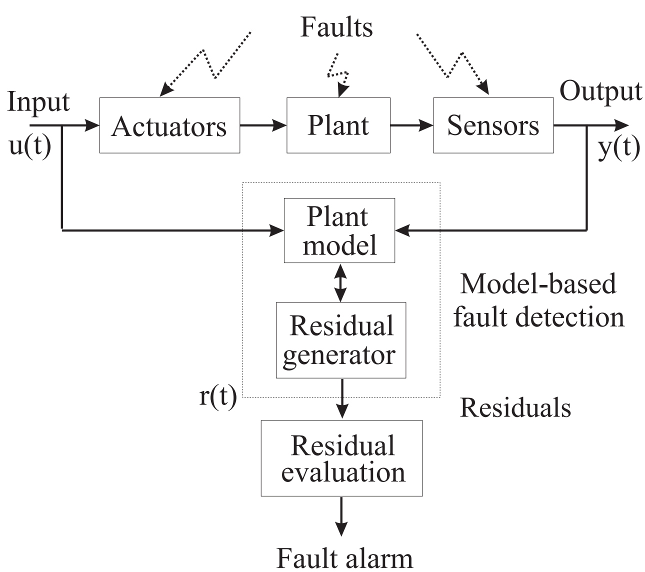

Fig. 1.2. Scheme for the model–based fault detection.

Basic process model–based FDI methods have been described by Patton *et al.* [Patton *et al.*, 1989], Basseville and Nikiforov [Basseville and Nikiforov, 1993], Gertler [Gertler, 1998] and Patton *et al.* [Chen and Patton, 1999, Patton *et al.*, 2000]:

1. Output observers (OO, estimators, filters);
2. Parity equations;
3. Identification and parameter estimation.

They generate residuals for output variables with fixed parametric models under method 1, fixed parametric or nonparametric models under method 2 and adaptive nonparametric or parametric models under method 3.

An important aspect of these methods is the kind of fault to be detected. As noted above, one can distinguish between *additive faults* which influence the variables of the process by a summation and *multiplicative faults* which are products of the process variables. The basic methods show different results, depending on these types of faults.

If only output signals $\boldsymbol{y}(t)$ can be measured, *signal model–based methods* can be applied, *e.g.* vibrations can be detected, which are related to rotating machinery or electrical circuits. Typical signal model–based methods of fault detection are:

1. Bandpass filters;
2. Spectral analysis (FFT);
3. Maximum-entropy estimation.

The characteristic quantities or features from fault detection methods show stochastic behaviour with mean values and variances. Deviations from the normal behaviour must then be detected by methods of *change detection* (residual analysis, Figure 1.2) like:

1. Mean and variance estimation;
2. Likelihood-ratio test, Bayes decision;
3. Run-sum test.

1.4 Model Uncertainty and Fault Detection

Model-based FDI makes use of mathematical models of the system. However, a perfectly accurate mathematical model of a physical system is never available. Usually, the parameters of the system may vary with time and the characteristics of the disturbances and noises are unknown so that they cannot be modelled accurately. Hence, there is always a mismatch between the actual process and its mathematical model even under no fault conditions. Such discrepancies cause difficulties in FDI applications, in particular, since they act as sources of false alarms and missed alarms. The effect of modelling uncertainties, disturbances and noise is therefore the most crucial point in the model–based FDI concept and the solution to this problem is the key for its practical applicability [Chen and Patton, 1999].

To overcome these problems, a model–based FDI scheme has to be insensitive to modelling uncertainty. Sometimes, a reduction of the sensitivity to modelling uncertainty does not solve the problem since the sensitivity reduction may be associated with a reduction of the sensitivity to faults [Chen and Patton, 1999, Gertler, 1998]. A more meaningful formulation of the FDI problem is to increase insensitivity to modelling uncertainty in order to provide increasing fault sensitivity.

The difficulties introduced by model uncertainties, disturbances and noises in model–based FDI have been widely considered during the last 10 years by both academia and industry [Gertler, 1998]. A number of methods have been proposed to tackle this problem, for example the Unknown Input Observer (UIO), eigenstructure assignment and parity relation methods.

An important task of the model–based FDI scheme is to be able to diagnose *incipient faults* in a system. With respect to *abrupt faults*, incipient faults may have a small effect on residuals and they can be hidden by disturbances. On the other hand, hard faults can be detected more easily because their effects are usually larger than modelling uncertainties and a simple fixed threshold is usually enough to diagnose their occurrence by residual analysis.

The presence of incipient faults may not necessarily degrade the performance of the plant, however, they may indicate that the component should be replaced before the probability of more serious malfunctions increases. The successful detection and diagnosis of incipient faults can therefore be considered a challenge for the design and evaluation of FDI algorithms.

1.5 The Robustness Problem in Fault Detection

In this monograph, observer–based approaches to robust FDI in industrial dynamic systems are summarised and applied to simulated and real plants. In the context of automatic control, the term robustness is used to describe the insensitivity or invariance of the performance of control systems with respect to disturbances, model–plant mismatches or parameter variations. Fault diagnosis schemes, on the other hand, must of course also be robust to the mentioned disturbances, but, in contrast to automatic control systems, they must not be robust to actual faults. On the contrary, while generating robustness to disturbances, the designer must maintain or even enhance the sensitivity of fault diagnosis schemes to faults. Furthermore, the robustness as well as the sensitivity properties must be independent of the particular fault and disturbance mode. Generally, the problem of robust FDI can be divided into the tasks of *robust residual generation* followed by *robust residual evaluation*.

In many cases, the disturbances and model–plant mismatches to which robustness must be generated, are due to the use of linear models for describing dynamic behaviour of non–linear processes. In this contribution, modelling errors are avoided from the very beginning by focusing on robust residual generation methods using linear and non–linear process models. This in turn simplifies the problem of residual evaluation without reducing the sensitivity to actual faults.

Effective tools for robust residual generation and even complete decoupling from external disturbances and unknown system parameters can be provided, *e.g.*, by unknown input observers which are introduced and applied to industrial processes. It is shown that the proposed solution to the disturbance de–coupling problem provides, in addition, the solution to both the fault detection and fault isolation problems.

On the other hand, many dynamic processes can only be described effectively using non–linear mathematical models. Most of the existing observer–based FDI techniques, however, are limited to the use of linear process models. The methods that can be found in the literature are based on the assumption that the system under supervision stays, during normal operation, in a neighbourhood of a certain known operating point [Chen and Patton, 1999, Patton *et al.*, 2000]

It is clear that, as almost every process system is non–linear, the modelling errors almost always reduce the accuracy of the linear model and therefore the performance of the FDI algorithm is compromised. Various methods for generating robustness to linearisation have been proposed in the literature and the reader is referred to [Patton *et al.*, 2000, Chap. 7] for a comprehensive treatment of this subject.

This monograph also surveys the state of the art of robustness methods and it presents some important ideas concerning the development of the use of non–linear models and predictors for FDI. In Chapter 4 observer–based approaches to robust FDI for dynamic systems are considered in more detail. In this contribution, the available model–based approaches are generalised, and thus extended to a wider class of dynamic systems.

In order to accommodate the application of robust FDI concepts, disturbances and parameter uncertainties of the monitored plants as well as faults are modelled in the form of unknown input signals. It is shown that, provided certain conditions can be met, complete decoupling of the residual from disturbances as well as from the parameter uncertainties of the process model can be achieved, whilst the sensitivity of the residual to faults is maintained. As the faults are also modelled in the form of external signals, this method additionally provides tools for the purpose of fault isolation. Fault isolation requires the de-coupling of the effects of different faults on the residual [Chen and Patton, 1999] and this, in turn, allows for decisions on which fault or faults out of a given set of possible faults has actually occurred.

These residual properties must be completely independent of the magnitude or frequency of the unknown inputs and the faults. This is crucial, in cases where no *a priori* knowledge about these properties is available. For systems, where the complete decoupling of the remaining unknown inputs or faults from the residual proves impossible, a threshold selection method, employing functional analytic methods and appropriate vector and operator norms can be exploited. This technique provides a tool for the robust evaluation of the residuals which have been generated by unknown input observers. Using the same functional analysis methods as employed for threshold selection, a performance index can be defined which allows for performance evaluation and, to a certain degree, also allows for optimal residual generator design [Patton *et al.*, 2000].

1.6 System Identification for Robust FDI

In earlier sections of this monograph, we have seen that model–based FDI methods formally require a high accuracy mathematic model of the monitored system. The better the model is as a representation of the dynamic behaviour of the system, the better will be the FDI performance. It is difficult to develop a highly accurate model of a complex system and hence the interesting question is: "what is a reasonable model to enable good performance in FDI to be guaranteed?".

It would be attractive to develop a robust FDI technique which is insensitive to modelling uncertainty, *i.e.*, so that a highly accurate mathematical model is no longer required. However, in order to design a robust FDI scheme, we should have a description (*i.e.*, some information) about the uncertainty, *e.g.*, its *distribution matrix* and spectral bandwidth, etc. Furthermore, this description should provide assistance for robust FDI design, *i.e.*, it can be handled in a systematic manner. Chapters 2 and 4 show how a typical uncertainty description makes use of the concept of "unknown inputs" acting upon a nominal linear model of the system. These unknown disturbances describe the uncertainties acting upon the system but disturbance distribution matrices are assumed known since they can be estimated by identification schemes.

It is clear that disturbances and faults act on the system in the same way, and thus we cannot easily discriminate between these excitations unless we know the structure of the disturbance distribution matrix. Once the disturbance distribution matrix is known, we can generate the residual with the disturbance de-coupling (robust) property, *i.e.*, the residual is de-coupled from the disturbance (uncertainty). The robust residual can then be used to achieve reliable FDI.

The theories underlying robust FDI approaches have been very well developed, but for real applications the following problems remain unsolved:

- estimation of reliable model for the monitored process;
- modelling accuracy of the real uncertainty by means of identified disturbance terms when no knowledge of the uncertainty is available;
- estimation of the disturbance terms and the structure of distribution matrices.

This book seeks to answer the above questions. Some simulation and real examples are given to test some of the theoretical results. These problems have to be addressed, otherwise the application domain of the disturbance de–coupling approach for robust FDI is very limited. In fact, few researchers and contributions have presented the application results of robust fault diagnosis to real processes.

As mentioned above, a primary requirement for model–based and disturbance de-coupling approaches to robust FDI is that both the system model and disturbance distribution matrices must be known. It is interesting that, within the framework of international research on this subject, there have been few attempts to address the problem by means of the *identification approach.* This lack of information has obstructed the application of robust FDI in real engineering systems. Chapters 3 and 4 present the research developments surrounding the joint estimation of system and disturbance matrices in order to solve the robust fault diagnosis problem.

Concerning the identification schemes developed and exploited in Chapters 3, 4 and 5, when all observed variables of a dynamic process are affected by uncertainties, the parameter estimation task can be performed by the so–called *errors–in–variables* methods. On the other hand, *equation error* methods can be developed in the case of exactly known plant variables [Simani *et al.*, 2000a]. It is worthwhile noting that less attention has been paid to errors–in–variables schemes.

Under these considerations, Chapters 3, 4 and 5 present the robust FDI results concerning the description of monitored plants by means of equation error and error–in–variables identified models in the presence variable uncertainties. Moreover, for the examples presented, estimates obtained by the errors–in–variables approach and equation error estimates are computed and compared in Chapter 5.

1.7 Fault Identification Methods

If several symptoms change differently for certain faults, a first way of determining them is to use classification methods which indicate changes of symptom vectors.

Some classification methods are [Patton *et al.*, 1989, Basseville and Nikiforov, 1993, Gertler, 1998, Babuška, 1998, Chen and Patton, 1999]:

1. Geometrical distance and probabilistic methods;
2. Artificial neural networks;
3. Fuzzy clustering.

When more information about the relations between symptoms and faults is available in the form of diagnostic models, methods of reasoning can be applied. Diagnostic models then exist in the form of symptom–fault causalities, *e.g.* in the form of symptom-fault tree. The causalities can be expressed as IF–THEN rules. Then analytical as well as heuristic symptoms (from operators) can be processed. By considering these symptoms as vague facts, probabilistic or fuzzy set descriptions lead to a unified symptom representation. By using forward and backward reasoning, probabilities or possibilities of faults are obtained as a result of diagnosis. Typical approximate reasoning methods are [Basseville and Nikiforov, 1993, Chen and Patton, 1999]:

1. Probabilistic reasoning;
2. Possibilistic reasoning with fuzzy logic;
3. Reasoning with artificial neural networks.

This very short consideration shows that many different methods have been developed during the last 20 years. It is also clear that many combinations of them are possible.

Based on more than 100 publications during the last 5 years, it can be stated that parameter estimation and observer-based methods are the most frequently applied techniques for fault detection, especially for the detection of sensor and process faults. Nevertheless, the importance of neural network-based and combined methods for fault detection is steadily growing. In most applications, fault detection is supported by simple threshold logic or hypothesis testing. Fault isolation is often carried out using classification methods. For this task, neural networks are being more and more widely used.

The number of applications using non–linear models is growing, while the trend of using linearised models is diminishing. It seems that analytical redundancy-based methods have their best application areas in mechanical systems where the models of the processes are relatively precise. Most non–linear processes under investigation belong to the group of thermal and fluid dynamic processes. The field of applications to chemical processes has few developments, but the number of applications is growing. The favourite linear process under investigation is the DC motor. In general, the trend is changing from applications to safety-related processes with many measurements, as in nuclear reactors or aerospace systems, to applications in common technical processes with only a few sensors. For diagnosis, classification and rule-based reasoning methods are the most important and the use of neural network classification as well as fuzzy logic-based reasoning is growing.

1.8 Report on FDI Applications

Because of the many publications and increasing number of applications (IFAC Congress and IFAC Symposia SAFEPROCESS) between 1991–2000, it is of interest to show some trends [Patton *et al.*, 1989, Basseville and Nikiforov, 1993, Gertler, 1998, Chen and Patton, 1999, Frank *et al.*, 2000]. Therefore, a literature study of IFAC FDI–related Conferences is briefly presented in the following. Contributions taking into account the applications reported in Table 1.1 were considered. The type of faults considered are distinguished according to Table 1.2. Among all contributions, the fault detection methods were classified as in Table 1.3. The change detection and fault classification methods are indicated by Table 1.4. The reasoning strategies for fault diagnosis are reported in Table 1.5. The contributions considered are summarised in Table

1.6. The evaluation has been limited to the Fault Detection and Diagnosis (FDD) of laboratory, pilot and industrial processes.

Table 1.1. FDI applications and number of contributions.

Application	Number of contributions
Simulation of real processes	55
Large-scale pilot processes	44
Small-scale laboratory processes	18
Full-scale industrial processes	48

Table 1.2. Fault type and number of contributions.

Fault type	Number of contributions
Sensor faults	69
Actuator faults	51
Process faults	83
Control loop or controller faults	8

Table 1.3. FDI methods and number of contributions.

Method type	Number of contributions
Observer	53
Parity space	14
Parameter estimation	51
Frequency spectral analysis	7
Neural networks	9

Table 1.4. Residual evaluation methods and number of contributions.

Evaluation method	Number of contributions
Neural networks	19
Fuzzy logic	5
Bayes classification	4
Hypothesis testing	8

Table 1.6 shows that among mechanical and electrical processes, DC motor applications are mostly investigated. Parameter estimation and observer-based methods are used in the majority of applications on these kind of

Table 1.5. Reasoning strategies and number of contributions.

Reasoning strategy	Number of contributions
Rule based	10
Sign directed graph	3
Fault symptom tree	2
Fuzzy logic	6

Table 1.6. Applications of model–based fault detection.

FDD	Number of contributions
Milling and grinding processes	41
Power plants and thermal processes	46
Fluid dynamic processes	17
Combustion engine and turbines	36
Automotive	8
Inverted pendulum	33
Miscellaneous	42
DC motors	61
Stirred tank reactor	27
Navigation system	25
Nuclear process	10

processes, followed by parity space and combined methods. Thermal and chemical processes are investigated less frequently.

Table 1.3 shows that parameter estimation and observer-based methods are used in nearly 70% of all application considered. Neural networks, parity space and combined methods are significantly less often applied.

More than 50% of sensor faults are detected using observer-based methods, while parameter estimation and parity space and combined methods play a less important role. For the detection of actuator faults, observer-based methods are mostly used, followed by parameter estimation and neural networks methods.

Parity space and combined methods are rarely applied. In general, there are fewer applications for actuator faults than for sensor or process faults. The detection of process faults is mostly carried out with parameter estimation methods. Nearly 50% of all the applications considered use parameter estimation-based methods for detection of process faults. Observer-based, parity space and neural networks-based methods are used less often for this class of faults.

Among all the described processes, linear models have been used much more than non–linear ones. On processes with non–linear models, observer-based methods are mostly applied, but parity equations and neural networks also play an important role. On processes with linear or linearised models, parameter estimation and observer-based methods are mostly used. Parity space and combined methods are also used in several applications, but not to the same extent as observer-based and parameter estimation methods.

Taking into account the system considered, the number of non–linear process applications using non–linear models are decreasing. For linear processes, no significant change can be stated.

The use of neural networks and combinations seems to be increasing.

Concerning the fault diagnosis methods, in recent years, the field of classification approaches, especially with neural networks and fuzzy logic has steadily been growing. Also, rule–based reasoning methods are increasingly being based on fault diagnosis. A growing application of fuzzy rule-based reasoning can be stated. Applications using neural networks for classification are increasing and the trends are analogous to the increasing number of non–linear process investigations. Nevertheless, the classification of generated residuals seems to remain the most important application area for neural networks.

1.9 Outline of the Book

To detect and isolate faults in a dynamic system, based on the use of an analytical model, a residual signal has to be used. It is derived from a comparison between real measurements and the relative estimates (generated by the model). The modelling uncertainty problem can be tackled by designing a FDI scheme, whose residuals are insensitive to uncertainties whilst sensitive to faults. On the other hand, a model with satisfactory accuracy can be estimated using identification procedures [Norton, 1986, Söderström and Stoica, 1987, Ljung, 1999].

The aim of the design of a FDI scheme is to reduce the effects of uncertainties on the residuals and to enhance the effects of faults acting on the residuals. The *main aim of this monograph* is to develop a residual generator for model–based fault diagnosis of a process by means of input and output signals. An accurate model of the process under investigation will be estimated using identification procedures from data affected by noises and acquired from simulated and/or actual plants. The monograph consists of 6 chapters and the main contributions are presented in Chapters 3, 4 and 5. Chapters are devoted to the particular problem in residual generation and the are organised as follows.

Chapter 2 reviews the state of the art of the model–based FDI. The FDI problem is formalised in an uniform framework by presenting the mathematical description and definitions. The fundamental issue of model–based methods is the generation of residuals using the mathematical model of the monitored system. By analysing residuals, fault diagnosis can be performed. Some structures of the residual generator are presented in this Chapter in order to give ideas how to implement the residual generation. A residual generator can be designed for achieving the required diagnosis performances, *e.g.* fault isolation and disturbance decoupling.

In order to design the residual generator, some assumptions about the modelling uncertainties need to be made. The most frequently used hypothesis is that the modelling uncertainty is expressed as a disturbance term in the system dynamic equation. The disturbance vector is unknown whilst its distribution matrix can be estimated by using identification procedures. Based on this assumption, the disturbance decoupling residual generator can be design by using unknown input observer methods [Chen and Patton, 1999, Liu and Patton, 1998].

Chapter 3 demonstrates how to apply dynamic system identification methods in order to estimate an accurate model of the monitored system.

The FDI methods presented require, in fact, a linear mathematical model of the process under investigation, either in state space or input-output form.

In particular, since state space descriptions provide general and mathematically rigorous tools for system modelling, they may be used in the residual generator design, both for the deterministic case (UIO and OO) [Chen and Patton, 1999, Frank, 1990, Luenberger, 1979, Watanabe and Himmelblau, 1982] and the stochastic case (Kalman filters (KF) and unknown input Kalman filters (UIKF)) [Jazwinski, 1970, Xie *et al.*, 1994, Xie and Soh, 1994].

In such a manner, the suggested FDI tool does not require any physical knowledge of the process under observation since the linear models are obtained by means of an identification scheme which exploits equation error (EE) and errors–in–variables (EIV) models. In this situation, the identification technique is based on the rules of the Frisch scheme [Frisch, 1934], traditionally exploited to analyse economic systems. This approach, modified to be applied to dynamic system identification [Kalman, 1982b, Kalman, 1990, Beghelli *et al.*, 1990], gives a reliable model of the plant under investigation, as well as the variances of the input–output noises affecting the data.

For the non–linear case, piecewise affine and fuzzy models will be used as prototypes for the identification. In particular, the multiple-model approach, using several local affine submodels each describing a different operating condition of the process, is exploited.

Chapter 4 aims to define a comprehensive methodology for actuator, process component and sensor fault detection. It is based on an output estimation approach, in conjunction with residual processing schemes, which include a simple threshold detection, in deterministic case, as well as statistical analysis when data are affected by noise. The final result consists of a strategy based on fault diagnosis methods well–known in the literature for generating redundant residuals.

In particular, this Chapter studies the approach to residual generation with the aid of OO, UIO, KF and UIKF. The residual is defined as the *output estimation error*, obtained by difference between the measurement of one output and the relative estimate. This Chapter also presents the design of such estimators both in the deterministic and stochastic environment.

The diagnosis procedure may be further specialised for actuators, input or output sensors and process components. In fact, the fault diagnosis of input sensors and actuators uses a bank of UIO in high signal to noise ratio conditions or a bank of UIKF, otherwise. The i–th UIO or UIKF is designed to be insensitive to the i–th input of the system. On the other hand, output sensor and process component faults affecting a single residual can be detected by means of a OO or a classical KF, driven by a single output and all the inputs of the system.

Chapter 5 shows how the proposed algorithms can be applied to the FDI of actuators, process components and input-output sensors of industrial plants.

In particular, the FDI techniques presented in this book have been tested on time series of data acquired from different simulated and real industrial gas turbine working in parallel with electrical mains, whose linear mathematical description is obtained by using identification procedures.

Results from simulation show that minimum detectable faults are perfectly compatible with the industrial target of this application.

Chapter 6 summarises the contributions and achievements of the monograph providing some suggestions for possible further research topics as an extension of this work.

1.10 Summary

Chapter 1 has provided a common terminology in the fault diagnosis framework in order to comment on some developments in the field of fault detection and diagnosis based on papers selected during the last 10 years.

The structure of the six chapters of this monograph and the main contributions presented have also been outlined briefly.