Meas. Sci. Technol. **17** (2006) 1809–1818 doi:10.1088/0957-0233/17/7/020

An approach to model-based fault detection in industrial measurement systems with application to engine test benches

P Angelov[1], V Giglio[2], C Guardiola[3], E Lughofer[4] and J M Luján[3]

[1] Department of Communication Systems, InfoLab21, Lancaster University, South Drive, Lancaster, LA1 4WA, UK
[2] ISTITUTO MOTORI CNR, Spark Ignition Engines and Fuels Department, Viale Marconi 8, 80125 Napoli, Italy
[3] CMT Motores Térmicos, Universidad Politécnica de Valencia, PO Box 22012, E-46071 Valencia, Spain
[4] Department Knowledge-Based Mathematical Systems, Johannes Kepler University of Linz, Altenbergerstrasse 69, A-4040 Linz, Austria

E-mail: edwin.lughofer@jku.at

Received 30 November 2005, in final form 28 February 2006
Published 7 June 2006
Online at stacks.iop.org/MST/17/1809

Abstract
An approach to fault detection (FD) in industrial measurement systems is proposed in this paper which includes an identification strategy for early detection of the appearance of a fault. This approach is model based, i.e. nominal models are used which represent the fault-free state of the on-line measured process. This approach is also suitable for off-line FD. The framework that combines FD with isolation and correction (FDIC) is outlined in this paper. The proposed approach is characterized by automatic threshold determination, ability to analyse local properties of the models, and aggregation of different fault detection statements. The nominal models are built using data-driven and hybrid approaches, combining first principle models with on-line data-driven techniques. At the same time the models are transparent and interpretable. This novel approach is then verified on a number of real and simulated data sets of car engine test benches (both gasoline—Alfa Romeo JTS, and diesel—Caterpillar). It is demonstrated that the approach can work effectively in real industrial measurement systems with data of large dimensions in both on-line and off-line modes.

Keywords: measurement systems, model-based failure detection, data-driven and hybrid modelling, data quality, combustion engines, engine test benches

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Recently, significant growth has been observed in the size and complexity of the technological installations in the automotive, power, chemical and food industries [9]. A side effect of this growth is an increase in the concentration of measuring, processing and control devices. The likelihood of appearance of a fault that may lead to the breakdown of a component, or the whole system, increases with the complexity of the system [8]. To tackle this problem, as well as to address the increasingly restrictive safety and environmental regulations, a significant rise in the demands on automatic fault detection, isolation and correction (FDIC) algorithms has been observed [18, 9]. These systems are required to cope with large dimensionality

0957-0233/06/071809+10$30.00

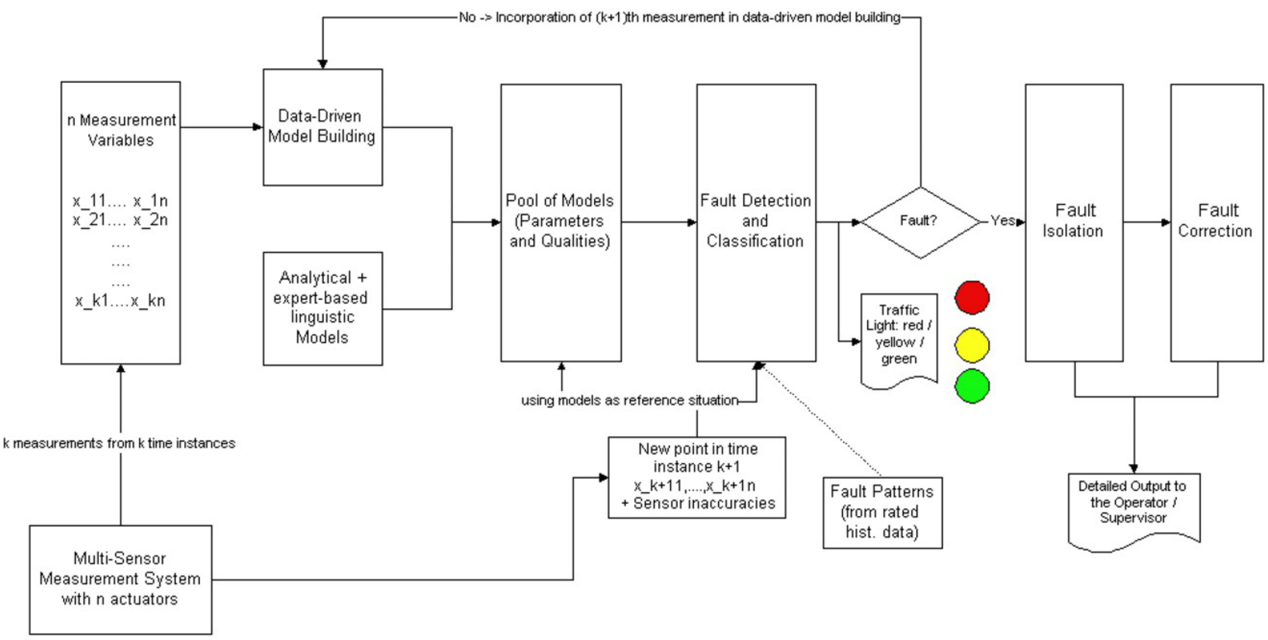

Figure 1. FDIC scheme in an on-line measurement and plausibility check system.

of the measured process variables, high sampling rates, non-stationary patterns, false alarms etc. The automation of diagnostic operations makes it possible to significantly shorten the time of identification and isolation of potential breakdowns, which has economic implications.

It is useful to formally define the fault of a system. The International Federation for Automatic Control, IFAC, has issued by one of its technical committees the following definition, which will be adopted in our paper [1]:

> Fault is an unpermitted deviation of at least one characteristic property or variable of the system from its acceptable/usual/standard behaviour.

In fact, a fault is caused by wrong behaviour of the system, which can even be a serious danger for human beings, e.g. a broken pipe where gas effuses. In this sense, an early detection of faults is indispensable. In this paper, the key assumption is that most of the faults, and especially all significant faults, are reflected in the (on-line recorded) measurement data as untypical patterns, jumps, drifts etc. Hereby, the following three phases are considered:

- Fault detection (FD): concerns the detection and warning of fault presence.
- Fault isolation (FI): concerns the identification of the place of fault appearance.
- Fault correction (FC): concerns the automatic correction of faults (in the measurement data).

The main focus is mainly on FD as this phase is the most important one which is a prerequisite for the other two. The proposed new concepts of automatic fault isolation and automatic fault correction (i.e. delivering correct measurement values) are also introduced.

There is a considerable body of literature concerning fault detection, isolation and correction (FDIC) in industrial measurement systems [8, 9, 12, 14, 18]. The approaches can be loosely divided into the following:

(i) Classification based.
(ii) Hypothesis testing.
(iii) Model based.
(iv) Signal processing based.

The classification-based approaches use techniques such as principal component analysis (PCA), Fisher discriminant analysis (FDA), and work normally off-line (based on a batch set of data collected from the industrial system) [9]. The second group relies on statistical tests of hypotheses about the structure of the model. In [26] such an approach is combined with propositional logic. It has close links with the group of model-based approaches. Model-based approaches are perhaps the most popular type. They are based on the idea of a model that represents the ideal, fault-free process and the comparison of the real measurement with the reference model. The resulting deviations (residuals) can then be observed (in off-line or on-line mode) and used to trigger a fault detection mechanism [8, 29]. Different types of models (regression and correlation models [28], fuzzy models [7, 19], causal-based models [15] or first principle models [18]) are used in different publications. However, thresholds are usually selected manually and residual calculations are very specific [19, 28]. Signal processing and filtering techniques are used, for example, in so-called intelligent sensors [2], to detect sensor faults such as peaks, (mean) drifts or other anomalies [22] in both time and frequency domains.

In this paper, a generic framework for combined FDIC is proposed that is suitable for both on-line and off-line applications, including real-time. This is based on the results presented in [14, 24]. The proposed approach is model based and classification based. The basic structure of the proposed methodology is graphically presented in figure 1.

The approach combines on-line adaptation and evolution of the reference model, which can be built based on the experimental/operational data. The on-line nature and the ability to detect faults as early as possible are the key features of the proposed new approach. As an additional option fault patterns can be used to train the data-driven model to help the process of fault correction (dotted line in figure 1).

The remainder of the paper is structured as follows. In section 2 the goals of the FDIC problem are formally stated. The role of reference models is discussed in section 3.1. The proposed approach for FDIC is presented in section 4. The evaluation methodology and results are presented in section 5. Section 6 concludes the paper and outlines directions for further work.

2. Problem statement

Let $x_1, x_2, \ldots, x_n$ be n-independent measurement variables in an arbitrary industrial process recorded dynamically with certain frequency (time steps). Then, a general static model for a specific variable x_i at time instant k can be defined as

$$\hat{x}_i = f_k(x_{j_1}, \ldots, x_{j_n}), \qquad j_m \in \{1, \ldots, n\}\backslash\{i\}. \tag{1}$$

If we take the time component into account, equation (1) can be detailed as

$$\begin{aligned}\hat{x}_i = {} & f_k(x_{j_1}(k), x_{j_1}(k-1), \ldots, x_{j_1}(k-l), \ldots, x_{j_n}(k), \\ & \times x_{j_n}(k-1), \ldots, x_{j_n}(k-l), \\ & \times x_i(k-1), x_i(k-2), \ldots, x_i(k-l)) \\ & j_m \in \{1, \ldots, n\}\backslash\{i\}.\end{aligned} \tag{2}$$

The constant l defines the order of the model. In both cases a dependence between x_i and a subset of the other measurement variables in the system is described by a model f_k at time instant k. When the model f_k is based on first principles, it represents an analytical formula where all the parameters are known and set *a priori* and $f_k = f_{k+m}$ for all $m \in \mathbb{N}$. That means the parameters of the model, and hence the model itself, do not change with time.

In the alternative *data-driven* approach the parameters of the model f_k are estimated based on data alone [3]. The data may be collected historically, by simulation of the considered process, or in *on-line mode*. When new recorded data $\vec{x}_{k+m}$ are collected the model parameters are adjusted, and, moreover, the model structure may also be adapted/evolved [3], such that $f_k \neq f_{k+m}$ for some $m \in \mathbb{N}$. With this notation, the goal of a fault detection (FD) strategy can be formulated in the following way:

Goal 1. *Let $f_{1,k}, \ldots, f_{m,k}$ be m various models as defined by equations (1) or (2) describing some relationships between different variables of an industrial process at time instant k. Then n newly recorded measurements $\vec{x}_{k+1,\ldots,k+n}$ ($\vec{x}_k$ denotes the row vector containing the kth measurement values for all variables) should be classified using these models as reference, such that the number of correct classifications should be as high as possible.*

If only two classes are considered (one class representing the fault-free case and the other representing all possible faulty cases) the correct classifications can be split into correct detections of faults and into correct detection of no faults, both influencing the *detection rate* and *false detection rate*. This distinction is made, because the detection and false detection rates play different roles with different priorities. The ideal case would be 100% detection rate and 0% false detection rate, which can usually only be achieved with noise-free data and perfect models. Usually, the false detection rate is more important than the detection rate, as in the case of a high false detection rate the operator gets easily irritated and confidence in the system falls.

The goal of the fault isolation strategy can be defined in the following way [12]:

Goal 2. *Let $\vec{x}_{k+i}, i \in \{1, \ldots, l\}, l < m$, represent l faulty measurements out of n measurements in total; then the goal of the fault isolation strategy is to keep the number of correct isolations as high as possible.*

Similarly to goal 1, the number of correct isolations can be divided into (correct) isolations of faulty channels and into (correct) non-isolations of non-faulty channels, both influencing the false isolation rate [12]. As fault isolation is triggered by FD, it is obvious that the performance of the fault isolation with respect to correctly found faulty channels can be only as good as the FD method itself. In goal 2 the focus is kept only on one part of fault isolation, namely the detection of the measured and calculated channels affected by the fault. It should be noted that the time instant when the fault appears is usually known and registered (both in off-line and on-line modes of data collection).

The goal of the fault correction can be defined as follows:

Goal 3. *Let $\vec{x}_{k+i}, i \in \{1, \ldots, l\}, l < m$ be the l faulty measurements and let $x_{j,k+i}, j \in \{1, \ldots, o\}$ be the faulty value in channel j of measurement $k+i$; then the goal of fault correction is to correct the faulty values such that*

$$E = \sum_{i=1}^{l} \sum_{j=1}^{p} |x_{j,k+i}(\text{corr}) - x_{j,k+i}(\text{fault-free})| \to \min \tag{3}$$

is as small as possible, i.e. the deviation of the corrected values from the real fault-free values is minimal.

Here the performance of fault correction depends strongly on both the performance of FD and the performance of fault isolation. The question arises whether it is possible to obtain the real fault-free values. Quite often during the testing and validation phase artificial errors are included although the real values are known. Ideally, the FD approach should detect all failures affecting any of the measurement channels included in the $f_{1,k}, \ldots, f_{m,k}$ set of models. This covers all kinds of faults in the measurement system, and those system faults that affect the measured variables.

3. Models as reference situation

Equations (1) and (2) correspond to general models, where an estimate $\hat{x}_i$ is provided for measurement channel x_i on the basis of the measured values of other channels at the same instant or at previous instants. These models can be modified to include inequalities ($\langle, \rangle$). In this case, instead of a direct estimation of the limits, measured values are provided. According to the nature of the model, they can be classified into several groups: *analytical models*, *knowledge-based*, *data-driven models* or *hybrid models*.

3.1. Analytical models

Analytical (also known as *first principle*) models are functional relationships resulting from a theoretical analysis of the considered phenomenon based on the laws of physics, chemistry, biology etc. For example, mass balance, energy balance and conservation of momentum are often used to build up analytical models for both open and closed systems. Many analytical models refer to transient processes and take the form of integral or differential equations; their use for FD purposes is complex and they are not suitable for the FDIC approach described in this paper. Conversely, analytical models referring to steady-state processes, expressed in the form of formulae/equations, are suitable for FDIC purposes. They have the advantage that their parameters have clear meaning and can remain static. For this type of model there is no need for experimental data to build or tune the model. Their main drawback is the requirement of extensive physical knowledge of the system, since their reliability depends on the correct identification of the variables that influence the phenomenon.

An example of a steady-state analytical model in an engine test bench system is the formula for the calculation of air inlet volumetric flow of the engine using a measuring sharp-edged orifice. In this case, the air flow rate (as the target channel) is expressed as a function of the input channel's air pressure, P_{air}, and temperature, T_{air}, and the pressure drop, ΔP, through the orifice:

$$Q = f(P_{\text{air}}, T_{\text{air}}, \Delta P). \tag{4}$$

3.2. Knowledge-based models

Knowledge-based models are created from linguistic expert knowledge: knowledge expressed in the form of linguistic rules. This collection of rules is coded into binary or fuzzy rule bases triggering decision trees [6, 11] or complete fuzzy systems [25, 27]. From this point of view, they are built once and remain static for the whole application process in a similar way as for analytical models. As opposed to analytical models, they benefit from being applicable to very complex relationships within the system which cannot be described by an analytical model. They allow a good insight into some system behaviour for non-experts since linguistic rules are easily readable by humans. However, the drawback of a high development effort still remains, as expert knowledge is usually collected through extensive meetings and discussions. Additional drawbacks are that sometimes the expert knowledge is not enough to explain extraordinary system states or when the experts have contradictory opinions. In the literature knowledge-based models are also called *weak white box* models. As they are not included in the current FDIC framework, they will not be considered in this paper.

3.3. Data-driven models

Data-driven models are generated from data alone without any prior knowledge or assumptions about the physical attributes and meanings of the measurement channels [4]. This generation process is also called training or learning. Data can be available in the form of batch (off-line collected and pre-recorded) data sets, most commonly stored in data matrices; or in the form of on-line measurements, which are recorded during the process operation. If the latter is the case the models should be kept up to date (usually with incremental learning techniques), especially when tracking highly time-variant system behaviour (figure 1). Data-driven models can be built up generically in the sense that no underlying physical, chemical or other laws about the measurement variables must be known. This produces the generic usability of such methods for any industrial system under the assumption that there exist observed or recorded measurements. However, faults in the data set can erode the trained models such that they lead to incorrect approximation results. This drawback can be avoided using feedback rejection as shown in figure 1. In this way, only those data points which are classified as fault-free [22], are incorporated into the model building process. In the proposed FDIC framework two kinds of data-driven models were applied:

- Linear correlation and regression models.
- Fuzzy models extracted from data.

Both types can be developed by batch learning or in an incremental (on-line) manner. While correlation and regression models [16] possess only linear parameters and therefore have a reduced flexibility, fuzzy models, especially Takagi–Sugeno fuzzy models, can approximate an arbitrary function to any degree of accuracy [30]. Such models are thus feasible choices for modelling nonlinear dependences between variables. Fuzzy models are preferable compared to neural networks because of their transparency and the interpretability of the underlying relationships [3]. Various methods exist in the literature for training fuzzy models in the batch learning mode [3, 10, 17]. For on-line (incremental) learning of fuzzy models refer to [5, 21, 23]. In [21] special attention is paid to the bias errors for specific types of fuzzy models, which will be an essential point in the FD approach presented in section 4. This type of method will be called *fuzzy* for the remainder of this paper.

3.4. Hybrid models

Often the considered steady process involves many variables which are not necessarily independent. Very often the average value of steady variables is influenced by transient phenomena. In these cases, theoretical analysis or empirical knowledge is useful in the identification of which variables or groups of variables are important. The functional structure can also be determined, but not the parameters. In this case, data-driven parameter identification using a reference database is a helpful addition to the analytical or empirical prior knowledge. The models obtained by this procedure are classified as *hybrid models*. In section 5 an example of this type of hybrid model will be demonstrated together with the applied identification strategy.

4. Fault detection isolation and correction approach

4.1. FD approach for models of equalities type

One obvious FD approach, when using models based on equalities with a unique target measurement variable as the fault-free reference situation, is to generate m residuals from m

models for the current measurement (say the kth) by comparing the measured value with the values of the target variables estimated by these models

$$\mathrm{res}_{k,m} = \|\hat{x}_{k,m} - x_{k,m}\| \tag{5}$$

where $\hat{x}_{k,m}$ is the estimated value of the mth model using the kth measurement. This residual can be compared to a percentage threshold on the relative deviation of the measured value from the expected value of the model,

$$\frac{\|\hat{x}_{k,m} - x_{k,m}\|}{\hat{x}_{k,m}} > \mathrm{perc\_thresh} \tag{6}$$

where perc_thresh $\geqslant 0$ can be tuned according to the precision requirements of the FD framework. The threshold is easily interpretable for an operator. As the threshold approaches 0 the risk of false detections becomes higher, whereas faults with lower intensities can be detected. From this point of view, there is always a trade-off between achieving high detection rates and low false detection rates.

The previous formula denotes the condition for a fault. It is only valid in the case of perfect models and noise-free data, otherwise bias and variance errors have to be integrated into the fault condition in order to obtain correct and stable results. Moreover, in the case when $\hat{x}_{k,m} < \epsilon$ with ϵ near zero, the above formula becomes unstable. The bias error is part of the whole model error (also called expected prediction error) which stems from the inflexibility of a model when reproducing a nonlinear process, e.g. in the case that the model has too low parameters or does not have an appropriate structure or appropriate inputs. The variance error is the part of the expected prediction error which is due to noise in the data and therefore yields uncertainty in the parameters of the generated model.

Different levels of sensor inaccuracies can appear for different variables and also in different measurement systems. Thus, in order to guarantee automatic threshold determination and improve the correctness of FD statements, both the bias and the variance errors are incorporated into the estimated values of models:

$$\begin{aligned}\hat{x}_{k,m} &= f_{k,m} \pm \mathrm{model\_error}_m \\ &= f_{k,m} \pm \sqrt{\mathrm{bias\_error}_m^2 + \mathrm{var\_error}_m}.\end{aligned} \tag{7}$$

Obviously, in the case of a high *bias error* the *variance error* can be neglected, while in the case of a low *bias error* the *variance error* produces a significant contribution to the expected prediction error which is important to avoid the well-known overfitting effect for data-driven models [16]. For analytical models, the *variance error* can be estimated by applying an error back-propagation law [1] (neglecting the time component)

$$\varepsilon_{\hat{x}_m} = \left|\frac{\partial f_m}{\partial x_1}\right| \varepsilon_{x_1} + \left|\frac{\partial f_m}{\partial x_2}\right| \varepsilon_{x_2} + \cdots + \left|\frac{\partial f_m}{\partial x_n}\right| \varepsilon_{x_n} \tag{8}$$

which incorporates sensor inaccuracies in all input variables $x_1, \ldots, x_n$ contained in the model f_m. The drawback of this approach lies with the indispensable *a priori* knowledge about sensor inaccuracies for all variables in a system, which is not always available. However, whenever measurement data of the underlying process are available, the expected prediction error can be estimated by the addition of a term to the squared bias

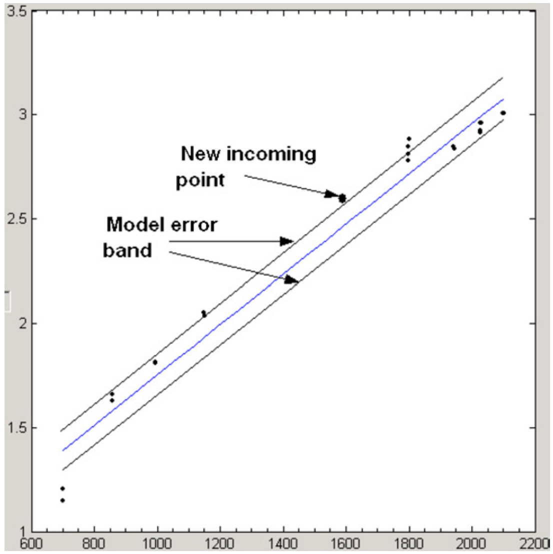

Figure 2. The importance of the error band. The new incoming point (marked as a big dot) is incorrectly classified as faulty because the threshold is set based on a tight model error band.

error which incorporates the covariance between the measured and estimated values from the model [16].

Using this notation and $\tilde{x}_{k,m} = x_{k,m} \pm \varepsilon_{x_m}$, where ε_{x_m} denotes the inaccuracy level of those sensors which sample the model's output variable x_m,

$$\begin{aligned}&\hat{x}_{k,m} - \tilde{x}_{k,m} \\ &\quad = f_{k,m} \pm \sqrt{\mathrm{bias\_error}_m^2 + \mathrm{var\_error}_m} - x_{k,m} \mp \varepsilon_{x_m}\end{aligned} \tag{9}$$

leading to the fault condition

$$\begin{aligned}\exists m\colon\quad & f_{k,m} - x_{k,m} \mp \varepsilon_{x_m} - \sqrt{\mathrm{bias\_error}_m^2 + \mathrm{var\_error}_m} > t \\ \vee\quad & f_{k,m} - x_{k,m} \mp \varepsilon_{x_m} + \sqrt{\mathrm{bias\_error}_m^2 + \mathrm{var\_error}_m} < -t.\end{aligned} \tag{10}$$

The existence operator is used over all models because a significant violation of one model will classify the measurement as faulty. Let us choose (without loss of generality) t to be a positive integer factor of $\sqrt{\mathrm{bias\_error}_m^2 + \mathrm{var\_error}_m}$, say thr; note that ε_y is always positive. This finally leads to

$$\begin{aligned}\exists m:\ & \frac{f_{k,m} - x_{k,m} - \varepsilon_{x_m}}{\sqrt{\mathrm{bias\_error}_m^2 + \mathrm{var\_error}_m}} > thr \\ \vee\ & \frac{f_{k,m} - x_{k,m} + \varepsilon_{x_m}}{\sqrt{\mathrm{bias\_error}_m^2 + \mathrm{var\_error}_m}} < -thr.\end{aligned} \tag{11}$$

If $t = 0$, i.e. one-σ area, is used in order to allow only values inside the 'model error band', this will lead to a threshold which may be too small and result in too many false detections as demonstrated in figure 2.

4.2. FD approach for inequalities-based models

In previous subsections, only models in the form of a set of equations (1) for a static model or (2) for a dynamic model have been considered. However, very often the expert

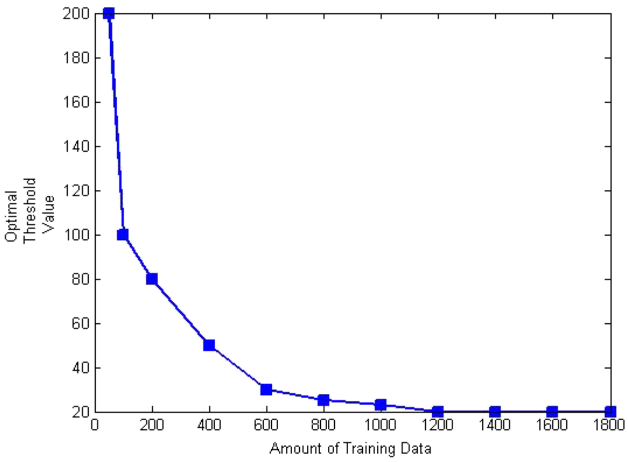

Figure 3. Best threshold value with respect to number of training data points for building high-dimensional reference fuzzy models for specific data sets.

knowledge and analytical models can only be given in the form of inequalities. This happens frequently when the system is only known in a vague way and the available models are inaccurate. The proposed FD scheme is generally able to deal with this class of relationships and only small modifications are required to take this into account. Let us assume that equation (1) becomes

$$x_m < f_k(x_{j_1}, \ldots, x_{j_n}), \qquad j_m \in \{1, \ldots, n\} \backslash \{i\} \tag{12}$$

where $<$ is used without loss of generality. In this case, the residual is considered to be equal to 0 if the inequality is satisfied and only the one-sided distribution is used in the definition of the fault condition. Thus substituting (11)

$$\exists m \frac{f_{m,k-1} - x_{m,k} + \varepsilon_y}{\sqrt{\text{bias\_error}_m^2 + \text{var\_error}_m}} < -thr. \tag{13}$$

4.3. Adaptive thresholding in data-driven models

The estimation of the variance error as described in section 4.1 is not always trustworthy, and sometimes even impossible to calculate. An alternative is to have an *adaptive* threshold *thr* on the right-hand side of the fault conditions (11) and (13). The threshold value depends on the number of training data points in a monotonically decreasing way because of overfitting and hence an increase in the variance error is more likely with a small number of data. This means the larger the number of data points fed into the training algorithm, the smaller the adaptive threshold gets. Ideally, it should ensure rates of detection and false detection similar to when trained with the full number of data points in a batch mode. Based on empirical tests with different data sets a functional dependence between the number of training data points and the best value of the threshold was extracted and was used for the car engine data sets shown later.

4.4. Local model errors

An interesting aspect of the analysis of the data space for potential faults concerns local model errors. For a specific local region in the input/output data space a separate model error can be introduced in the denominator of the fault condition (11) influencing the FD logic. This way a new incoming data sample is checked against the bias error of that part of the actual model which is nearest to this new point (with respect to some distance measure). It can be expected that the detection rate increases, as the model error is usually different for different regions (in some regions the model is more accurate than in others). This was verified using FD based on a fuzzy model where each separate rule denotes a specific local region in the input/output data space. It should be noted, however, that for on-line FD this will inevitably lead to additional computational burden, which may be prohibitive in real-time applications.

4.5. Consolidation of multiple FD outputs

The residuals calculated on the left-hand side of the inequality can be normalized into *error hint fuzzy* values based on different values of the monotonic transfer function as follows:

- transfunc(0) = 0.
- transfunc(thresh) = 0.5; in this case a value greater than 0.5 is classified as fault.
- transfunc(max_res) = 1, where max_res denotes the maximal value of a residual.
- transfunc is monotonic.

This normalization is indispensable considering that different components in a system can interact and contribute to the overall plausibility and fault occurrence in a complex industrial system. The condition defined above can be applied in a generic way in all model-based FD modules. Special attention should be paid to the threshold.

From the aggregated *error hint fuzzy* value the unique *error hint crisp* value is produced which yields a crisp FD statement concerning the current measurement. This value is zero in the case of no fault in the actual measurement or 1 in the case of fault in the actual measurement. Note that an *error hint fuzzy* value greater than 0.5 means fault. In addition to these two values a third value is generated and used within the consolidation process, the *internal quality*, which gives rise to the trustworthiness of the models which were used as fault-free references. One possible choice to calculate the *internal quality* is to evaluate the aggregated model quality and use this as a measure of the *internal quality*. The quality of an analytical or knowledge-based model can be determined using expert knowledge. Alternatively, it can be evaluated based on measurement data through the *r*-squared-adjusted formula. This measure incorporates the degrees of freedom of a model and penalizes more complex models relative to easier ones. It delivers values in the interval [0, 1], where a value near 1 means that a very reliable model is available and a value far below 1 suggests that the model is useless. This can be of essential help when generating data-driven models from a high-dimensional set of variables, where the input and output structures of the models are not given and have to be estimated automatically from data. Hence, models with a high quality can be used in the FD framework while others are skipped.

The values produced by different plausibility check algorithms, modelling and FD are combined directly together in a closed formula. In this sense, the direct consolidation method evaluates an overall plausibility statement for the current measurements. The consolidation of partial FD statements can be based on different aggregation formulae combining the *error hint crisp* values with the Boolean operators OR, AND, a democratic or soft decision combined with the confidence values (=internal qualities) of all n FD modules,

$$\begin{aligned} &\text{If } \sum_{i\in A} \text{int\_qual}_i \geqslant \sum_{i\in B} \text{int\_qual}_i \\ &\quad \text{overall\_error\_crisp} = 1 \\ &\text{Else} \\ &\quad\quad \text{overall\_error\_crisp} = 0 \end{aligned} \tag{14}$$

where $A = \{\text{error\_crisp}_i = 1 | i = 1, 2, \ldots, n\}$ and $B = \{\text{error\_crisp}_i = 0 | i = 1, 2, \ldots, n\}$. In this case the confidences of FD modules classifying a measurement as faulty are compared with those of FD modules classifying a measurement as fault-free. An *error hint fuzzy* value and an *internal quality* value are calculated that serve as overall fault likelihood measures. These two values are shown in the operator's GUI.

4.6. Fault isolation

Fault isolation is triggered by the FD. Indeed, only in the case of a detected fault is it undertaken. The proposed fault isolation approach incorporates the output of the consolidation process in FD, namely the *error hint fuzzy* value and the *internal quality* value produced by each model for each separate measurement. Additionally, the gradient information serves as an indicator of how strongly a certain variable influences the model. Hence, in order to get comparable and range-independent influence information, normalized gradients are computed for each variable in the actual point. The sensitivity vector introduced in [13] is upgraded with the gradient information and an estimate of the size of perturbation for a specific channel is derived using *error hint fuzzy* and *internal quality* in order to classify this channel as faulty (= to isolate the fault). From this estimation a fault likelihood for each channel is calculated. It is in the interval $[-1, 1]$ (+1 for very likely, -1 for not likely; usually a threshold value of zero was used in order to get a crisp decision for each measurement channel). For further details see [12].

4.7. Fault correction strategy

The fault correction strategy for delivering an expected value can be summarized based on the information provided by the reference model and using the proposed FD approach, consolidating the results from each module and causing fault isolation statements as described in section 4.6, as follows:

- If the faulty channel isolated using the fault isolation module is not a target channel (and hence only occurs as an input channel in at least one model and not on the left-hand side as y of any model), then no corrected value can be delivered for this channel. This may happen when no useful functional relationship to at least one other measurement channel in the system exists for the isolated channel which appears as target.
- If the faulty channel isolated using the fault isolation module is a target channel in only one model, take the estimated value from this model (= the target value) and deliver this as the corrected value.
- If the faulty channel isolated by the fault isolation module is a target channel in more than one model, take the model which is the most trustworthy, i.e. with the highest model quality—the other models for that channel can obviously be considered obsolete for fault correction as they possess a lower quality due to the less precise representation or lower flexibility (e.g., if input channels are missing etc).

Combining the FD, the consolidated estimate, fault isolation and correction, the overall FDIC framework is summarized as graphically presented in figure 1.

5. Evaluation and results

5.1. Measure values for evaluation

In order to be able to verify, validate, and compare the performance of the model-based FDIC approach an objective benchmark is needed. Using the labelled data (data sets marked as faulty or error-free) one can define the *detection rate* using the relative frequency of detections as follows,

$$A_{\text{det}} = \frac{N_{\text{FD}}}{N_{\text{FM}}}, \tag{15}$$

where N_{FD} denotes the number of measurements with a correct detection and N_{FM} denotes the number of faulty measurements in the test data set. The *false detection rate* can also be defined using the labelled data and the relative frequency of false detections, leading to

$$A_{\text{false}} = \frac{N_{\text{FOD}}}{N_{\text{FFM}}} \tag{16}$$

where N_{FOD} denotes the number of measurements with false detection and N_{FFM} denotes the number of fault-free measurements in the test data set. A FD method that has both high *detection rate*, A_{det}, and low *false detection rate*, A_{false}, is preferable. For the cases when both *detection rate*, A_{det}, and *false detection rate*, A_{false}, are high it is not obvious which method is preferable. For such cases, one can use the following additional measure called *external FD-method quality*, or simply *external quality*, which combines *detection rate* and *false detection rate* in one value and incorporates the relative frequency of faulty data in the test data set overall,

$$\text{ext}_{\text{qual}} = \frac{w_{\text{false}}(1 - A_{\text{false}}) - \frac{m}{n}(w_{\text{false}}(1 - A_{\text{false}}) - w_{\text{det}} A_{\text{det}})}{1.5 - \frac{m}{n}}, \tag{17}$$

where n is the number of the test data records, m is the number of faulty test data records, $n - m$ is the number of fault-free test data records. The weights w_{false} and w_{det}, both $\in [0, 2]$ with $w_{\text{false}} + w_{\text{det}} = 2$, reflect the impact of the correct detections and false detections on the *external quality* measure. The weights can be adjusted in such a way that the importance of correct detections dominates the importance of false detections or vice versa.

5.2. *Application example*

A real-world application example is used to demonstrate the viability of the proposed FDIC approach. The example is from the area of automotive engine test benches. This is an interesting industrial application which is characterized by high complexity, level of automation, and sophistication. The verification and validation of correct detections and false detections were carried out on data sets obtained through both simulation and real-life tests of three different car engines:

- Simulated engine data including noise-free data and faults with an intensity of 5% and 10%.
- Data set from a direct injection gasoline engine (Alfa Romeo JTS) including 19 channels and faults with intensities of 3%, 5%, 10%, 20% and 50%.
- Real measured data for a heavy diesel engine (Caterpillar) for real on-line check.

For the off-line tests, all data sets were divided into a training data set for generating and adapting data-driven (regression and fuzzy rule-based) models and into testing data set. These test data sets include artificially built-in faulty measurements. In this sense, some good measurement values in the data sets were disturbed by a percentage of their absolute value. For example, a value of 100 was reset artificially to a value 90, representing a (rather small) fault causing 10% disturbance in the data. Hence, from now on we simply speak about a fault level of 10% rather than a fault causing 10% disturbance in the data. According to the experience reported by several engine test specialists (both engineers and technical staff interviewed in several engine manufacturing companies and research institutes), the usual critical failures affect the signal value in one or several measurement channels by more than 10%. Hence, fault levels of 20% and 50% were selected for performance evaluation of our fault detection approach. Moreover, fault levels of 3%, 5% and 10% were also selected in order to evaluate the performance on less significant faults (e.g., resulting from sensor aging or slightly incorrect positioning/connection), which are harder to detect. In this sense, the performance at these fault levels defines the achievable limits for the fault detection approach.

For the Caterpillar data no training data set was used for off-line tests. Instead the data were sent sample by sample in the on-line mode. The model was built from the first 30 or 50 data samples and tested based on the remaining data. In addition, the fault-free data after the 30th or 50th data sample were used to adapt the model in the on-line mode. For all test results, the direct consolidation strategy is used with the ‘OR’ type of aggregation operator. This has been confirmed by a large number of empirical experiments. This is logical, when taking into account that the main focus is laid on keeping the false detection rate close to 0 for each FD module, since otherwise the confidence in the proposed approach amongst the human operators of the system would quickly diminish.

5.3. *Off-line results*

5.3.1. Results on simulated data. The first tests were carried out on simulated (thus noise-free) data for a special diesel engine together with two rated check data sets; one containing fault levels of 10%, the other containing fault levels of 5% in some channels. Hence this data set is a good benchmark for the sensitivity of the methods with respect to less significant faults. The training data set contained around 1000 points, while the test data sets contained 2052 points, where approximately half of them were faulty. Practically, all the methods produced between 0% and 2% false detections using only an optimal value for the threshold. All the detection results are listed in table 1.

Table 1. Comparison results of model-based FD approach among several methods (= components) for simulated car engine data.

Method	Fault level	A_{det}	A_{false}	ext_{qual}
Global correlation	10%	20.47%	0.00%	0.62
	5%	20.04%	0.00%	0.62
Local correlation	10%	49.08%	0.00%	0.76
	5%	37.12%	0.00%	0.7
Local regression	10%	20.67%	0.00%	0.62
	5%	12.88%	0.00%	0.58
Fuzzy model	10%	87.27%	0.00%	0.95
	5%	67.08%	0.00%	0.82
Analytical model	10%	35.13%	0.00%	0.64
	5%	23.92%	0.00%	0.69
Overall	10%	87.27%	0.00%	0.95
	5%	67.08%	0.00%	0.82

From this table it can be concluded that the FD system based on the adaptive fuzzy inference system (ANFIS [17]) and trained off-line produced by far the best results of all methods. This model and the training procedure described in [17] are, however, not suitable for use in the on-line mode because of the iterative nature of the gradient-based search algorithm used. When the data are fed into the system in the on-line mode (sample by sample) the results deteriorate significantly (a drop of detection rate from 87.27% to 46.95% in the case of fault levels of 10% and from 67.08% to 22.60% in the case of fault levels of 5% can be observed). A further analysis of these results is needed for the on-line case, because recent parallel studies [4, 23]) indicate that the incremental on-line learning methods for fuzzy models outperform ANFIS. Surprisingly, the analytical models did not produced stronger results than all the data-driven modelling approaches for all tested fault levels. Furthermore, in the case of fault levels of 10% analytical models were even weaker than the incremental learning method for fuzzy systems (35.13% versus 46.95%).

5.3.2. Results on real measured engine test bench data. The second tests concerned the plausibility check of hybrid (simulation) models. These models possess an analytical structure in the form of regressor terms and unknown linear parameters which should be identified using the training data, which was developed for a spark ignition engine with gasoline direct injection. Six different *hybrid models* were developed for engine-out emission using carbon monoxide, total unburned hydrocarbon and nitric oxide, in-cylinder air trapped mass, exhaust pressure and temperature, all measured as average values in steady operation. Actually, these variables are linked to the transient processes occurring in the cylinder during one or two consecutive crankshaft revolutions, namely the air flow through inlet valves during the down stroke of

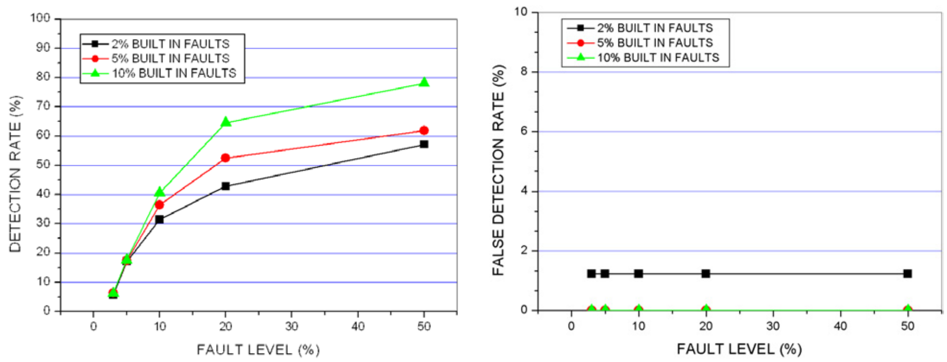

Figure 4. Detection rate (left) and false detection rate (right) depending on the fault level (along the x-axis) and fault quantity (different line styles).

the piston, followed by compression, combustion, expansion, blow down and expulsion of exhaust gases. The common approach in taking into account these transient phenomena is a theoretical analysis leading to a differential equation representing the process. The problem is to find models usable by the FDIC algorithm, i.e. relationships with one target channel as a function of other input channels. Often, an integration of the differential equation does not lead to this result, but allows highlighting of the functional links between the various variables. The alternative is to resort to pre-existent knowledge or experimental analysis on the effects of the variation of each input on the target variable. For inlet air mass flow the first method was followed, namely a preliminary analysis of the energy balance in differential terms, leading to a differential equation. By integration and the proper rearrangement, the following formula can be obtained,

$$\frac{m}{\rho_0 C} = a_1 \left(\frac{p_m}{RT_m}\right)^{0.8} \frac{1}{N^{0.2}} + a_2 \frac{p_{\text{exh}}}{287T_m} + a_3 \frac{p_m}{RT_m} + a_4 \frac{p_{\text{IVC}}}{287T_m}$$

where m denotes the air mass trapped inside the cylinder, ρ_0 is the reference air density, C is the cylinder displacement, p_m is the inlet manifold mean pressure, T_m is the inlet manifold mean temperature, R is the gas constant, N is the engine rotation speed, p_{exh} is the exhaust manifold mean pressure, and p_{IVC} is the pressure in the cylinder at the inlet valve closing. Regarding the exhaust gas temperature at the exhaust port, it is known that this mainly depends on spark advance, relative air/fuel ratio, and load related to inlet manifold pressure. Dependence on heat exchange was also considered including engine rpm. Thus the model

$$T_{\text{exh}} = a_1 p_m + a_2 p_{\text{exh}} + a_3 \Phi_{\text{adv}} + a_4 \lambda + a_5 N + a_6 N^2$$

was proposed, where T_{exh} denotes the exhaust manifold mean temperature, p_{exh} is the exhaust manifold mean pressure, p_m is the inlet manifold mean pressure, Φ_{adv} is the spark advance, λ is the relative air/full ratio and N is the engine rotation speed.

The other four *hybrid models* that were used are not stated here in detail. The linear parameters within these models (see above) could be identified based on an experimental training fault-free data set in a generic manner and in terms of using the same estimation algorithm for all six models. 2204 values (116 data rows times 19 channels) of fault-free data samples were used. It is interesting to note that the resulting linear parameters and hence complete models were almost the same using both off-line training based on least-squares estimation [20] as well as on-line incremental learning [23]. It illustrates that convergence of the on-line incremental learning is quite satisfactory for real practical cases. The test data sets were obtained by introducing fault levels of 3%, 5%, 10%, 20% and 50% (with random + or − sign) in order to achieve 2% of faulty single values with respect to the total number of values 2204. Further random errors were added in the same way to achieve 5% and 10% of faulty single values respectively. This procedure produced 15 test data sets, from which detections were extracted as shown in figure 4.

It can be clearly seen that the false detection rate is negligible, as it is 0 for all cases except in the case of a low rate of faulty data (2%). The detection rate is quite high, especially for fault levels (intensities) of 20% and more (which usually represent the really significant faults on an engine test bench). The relatively lower detection rates for fault levels of 3% and 5% stem from the fact that the model errors consisting of bias and variance error trigger confidence intervals which are larger than 3%, respectively 5%, deviation from the model in certain input regions. However, with the threshold *thr* as defined on the right-hand side of (11), the width of the confidence band can be enlarged. It was set to a value of 5 for all six hybrid simulation models (which was slightly below the default value of 8). It should also be recognized that the fault isolation rates were between 15% and 20% lower than the fault detection rates, but still inside an acceptable range for the operators.

5.4. On-line results

A fuzzy model and a correlation-based model were built based on the first 30 or 50 fault-free data samples. These models were later used for dynamic adaptation and to check the plausibility of the on-line recorded data from a diesel engine. For consolidation of different fault conclusions the 'OR' type of aggregation has been used. Faulty data points were ignored and not used for on-line model adaptation. The following results were obtained for real data coming in on-line from a large diesel engine:

overall detection rate: 64.7%;
overall false detection rate: 0.0%.

This result is very promising bearing in mind that many small faults (with a deviation of 5% or less) were present in this real data set and no fuzzy models and correlation models were pre-built from fault-free and high-quality training data. Moreover, all major faults (with a deviation of 20% or more), which are significant and critical and hence have to be detected, have actually been detected. If the feedback is omitted (see the line on top of figure 1) faulty data points are also incorporated into the model building steps for both the fuzzy model and the correlation models, and the detection rate decreases by about 15–20%. This illustrates the essential role of the feedback for stopping the influence of the faulty data on the update of the FD model.

6. Conclusion and future directions

In this paper, the analytical basis of the FD logic integrated into a FDIC framework was demonstrated on real data from car engine test benches. The proposed FDIC framework has a generic nature and is applicable to any complex contemporary industrial measurement system. The proposed reference model-based framework has been applied to both off-line simulated data and real data in both off-line and on-line modes using vehicle engine test benches for Alfa Romeo and Caterpillar. The results demonstrate the viability of the proposed methodology and its superiority over the conventional correlation- and regression-based models and over analytical (first principles-based) models. The rate of false detections can be expected to be very small. The proposed approach is potentially very useful for early FDIC in real time in various industrial systems. Promising future extension includes development of the local error bars for data-driven models in order to cope with extrapolation and different data densities in different local areas. The application of the proposed approach to practical industrial measurement systems will be addressed in the near future.

Acknowledgment

This work was supported by the European Commission (project GRD1-2001-40034).

References

[1] *Guide to the Expression of Uncertainty in Measurement* 1995 (Geneva: International Organization for Standardization)
[2] Alag S, Agogino A M and Morjaria M 2001 A methodology for intelligent sensor measurement, validation, fusion, and fault detection for equipment monitoring and diagnostics *Artif. Intell. Eng. Design. Anal. Manuf.* **15** 307–20
[3] Angelov P P 2002 *Evolving Rule-Based Models* (Berlin: Springer)
[4] Angelov P P and Buswell R A 2002 On-line identification of evolving fuzzy rule-based models *IEEE Trans. Fuzzy Syst.* **10** 667–77
[5] Angelov P P and Filev D 2004 An approach to online identification of Takagi–Sugeno fuzzy models *IEEE Trans. Fuzzy Syst.* **34** 484–98
[6] Breiman L 1993 *Classification and Regression Trees* (Boca Raton, FL: Chapman and Hall)
[7] Catelani M, Fort A and Alippi C 2002 A fuzzy approach for soft fault detection in analog circuits *Measurement* **32** 73–83
[8] Chen J and Patton R J 1999 *Robust Model-Based Fault Diagnosis for Dynamic Systems* (Norwell, MA: Kluwer)
[9] Chiang L H, Russell E L and Braatz R D 2001 *Fault Detection and Diagnosis in Industrial Systems* (London: Springer)
[10] Chiu S 1994 Fuzzy model identification based on cluster estimation *J. Intell. Fuzzy Syst.* **2**
[11] Drobics M 2000 Fuzzy decision trees—induction and deduction *Technical Report* SCCH-TR-0206, Software Competence Center Hagenberg, A-4232 Hagenberg, Austria
[12] Efendic H, Schrempf A and Del Re L 2003 Data based fault isolation in complex measurement systems using models on demand *Safeprocess 2003 IFAC: Proc. 5th IFAC Symp. Fault Detection Supervision and Safety of Technical Processes* pp 1149–54
[13] Fang C, Ge W and Xiao D 2000 Fault detection and isolation for linear systems using detection observers *Issues of Fault Diagnosis for Dynamic Systems* ed R J Patton, P M Frank and R N Clark (Berlin: Springer) pp 87–113
[14] Galindo J, Lujan J M, Guardiola C, Lughofer E and Klement E P 2004 Fault detection in engine measurement systems by a model-based approach *Proc. SAE 2004*
[15] Gertler J 1998 *Fault Detection and Diagnosis in Engineering Systems* (New York: Dekker)
[16] Hastie T, Tibshirani R and Friedman J 2001 *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (New York: Springer)
[17] Jang J-S R 1993 ANFIS: adaptive-network-based fuzzy inference systems *IEEE Trans. Syst. Man Cybern.* **23** 665–85
[18] Korbicz J, Koscielny J M, Kowalczuk Z and Cholewa W 2004 *Fault Diagnosis—Models, Artificial Intelligence and Applications* (Berlin: Springer)
[19] Laukonen E G, Passino K M, Krishnaswami V, Lub G-C and Rizzoni G 1995 Fault detection and isolation for an experimental internal combustion engine via fuzzy identification *IEEE Trans. Control Syst. Technol.* **3** 347–55
[20] Ljung L 1999 *System Identification: Theory for the User* (Upper Saddle River, NJ: Prentice-Hall)
[21] Lughofer E 2005 Data-driven incremental learning of Takagi–Sugeno fuzzy models *PhD Thesis* Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz
[22] Lughofer E, Efendic H, Del Re L and Klement E P 2003 Filtering of dynamic measurements in intelligent sensors for fault detection based on data-driven models *Proc. IEEE CDC—IEEE CDC Conf. (Maui, HI)* pp 463–8
[23] Lughofer E and Klement E P 2005 FLEXFIS: a variant for incremental learning of Takagi–Sugeno fuzzy systems *Proc. FUZZ-IEEE 2005 (Reno, NV, USA)* pp 463–8
[24] Lughofer E, Klement E P, Lujan J M and Guardiola C 2004 Model-based fault detection in multi-sensor measurement systems *Proc. IEEE IS (Varna, Bulgaria)* pp 184–9
[25] Michels K, Klawonn F, Kruse R and Nürnberger A 2002 *Fuzzy-Regelung: Grundlagen, Entwurf, Analyse* (Berlin: Springer)
[26] Nyberg M 1999 Model based fault diagnosis, methods, theory, and automotive engine application *PhD Thesis* Department of Electrical Engineering Linköping University, SE-581 83 Linköping, Sweden
[27] Piegat A 2001 *Fuzzy Modeling and Control* (Heidelberg: Springer)
[28] Schrempf A, Del Re L, Groißböck W, Lughofer E, Klement E P and Frizberg G 2001 Automatic engine modelling for failure detection *Proc. 2001 ASME Int. Mechanical Engineering Congress and Exposition*
[29] Simani S, Fantuzzi C and Patton R J 2002 *Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques* (Berlin: Springer)
[30] Wang L X 1992 Fuzzy systems are universal approximators *Proc. IEEE Int. Conf. on Fuzzy Systems* pp 1163–9