BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 62, No. 3, 2014
DOI: 10.2478/bpasts-2014-0062

Application properties of methods for fault detection and isolation in the diagnosis of complex large-scale processes

J.M. KOŚCIELNY* and M. SYFERT

Institute of Automatic Control and Robotics, Warsaw University of Technology, 8 św. Andrzeja Boboli St., 02-525 Warszawa, Poland

Abstract. The survey presents a selection of the methods of the fault detection and isolation suitable to be useful for the diagnostics of the complex, large scale industrial processes. The paper focuses on these methods that have appropriately high level of potential applicability in industrial practice. The novelty of the paper relies on the discussion of the dependency of the level of knowledge about diagnosed process and recommended diagnostic approaches. Appropriate recommendations were given in the convenient form of the table.

Key words: fault diagnosis, fault detection and isolation, complex large-scale processes, process modelling.

1. Introduction

In the last twenty years there has been observed a rapid development of the methods of fault diagnostics, originating from the theory of modeling and identification, as well as artificial intelligence technology approaches. The most detailed description of these methods can be found in the books [1–10]. There had also been written a lot of survey papers on the methods of fault diagnostics [11–18]. As a valuable source of information may serve proceedings from IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes – SafeProcess, which has been organized every 3 years, since 1991. In parallel to the SafeProcess series of conferences, the Polish track of Diagnostics of Processes and Systems has been organized since 1996. Many methods have been developed for fault detection, isolation and identification, but their practical usefulness is usually limited to the certain class of processes. The possibilities of application of particular methods are determined by the specific nature of the process being diagnosed on one hand, and on the other, the form and degree of knowledge about the process required by the given method.

This work concerns the evaluation of usefulness of different methods of detection and isolation of faults for diagnostics of the complex processes (LSS – Large Scale Systems), applied in the chemical, petrochemical, energy industry etc. Such processes include hundreds or even thousands of devices usually operating in difficult and variable conditions and the diversity and number of faults is very high. Diagnostics of such complex technological systems is a very difficult task. There occurs a number of specific problems, conditions and restrictions, which are not important while diagnosing low scale processes, as well as machines or devices. These problems and restrictions influence in a significant way the choice of methods of the faults detection and isolation. Among these problems there are to be considered [7, 19, 20]:

- complexity of systems being diagnosed, which contain thousands of devices operating usually in difficult and variable conditions. Thus, the number of faults may be very high;
- the requirement of realizing diagnostic tasks on-line, with the use of working data only, as one can't disturb the operation of a process by the test impulses;
- changeability of the structure of the process connected with the turning on and off the technological devices, disconnecting measuring instruments etc. Such changeability of structure constitutes very essential impediment during designing a diagnostic system;
- uncertainties of measurements and symptoms of faults,
- lack of data for emergency states. In databases of the systems for automatic control (DCS and SCADA) there are available rich sets of measurement data, but the archived time series concern mainly regular process operation and few referrer to registered abnormal or emergency states. A diagnostic system should detect and recognize serious breakdowns, for which there is no learning data, comprising also potential breakdowns, which had never occurred before;
- the delays of the faults symptoms. The diagnosed process is a dynamic system, so a certain amount of time elapses from appearance of the fault to the moment of arising of its measurable symptoms. The time is dependent on the dynamic properties of the part of the process being tested. The same fault is detected by different diagnostic signals in different time instances. The algorithm of fault isolation can generate false diagnoses if there is no built-in mechanism, which makes the course of inference robust to symptoms' delays.

The algorithms of diagnostic inference intended for diagnosing complex technological systems should acknowledge the above mentioned problems and effectively solve them.

The most significant, while choosing the methods of faults detection and isolation, is the degree of available knowledge

*e-mail: jmk@mchtr.pw.edu.pl

about the process being diagnosed. Simple methods of detection base only on the analysis of thresholds (constraints) or the statistical or spectral analysis of the particular process variables. The advanced methods make use of models for faults detection. These are both – models designed on the basis of the knowledge of equations describing physical phenomena taking place in the process, and the models created on the basis of the process data. One only need to know the relation between the inputs of the particular part of the process and the modeled output for a detection.

In order to isolate the fault it is essential to know the diagnostic relation, i.e. relation between faults and values of the diagnostic signals. This relation can be determined in three ways: as a result of process modeling with the influence of faults; on the basis of learning with the use of measurement data from emergency states; or on the basis of expert knowledge. The way of acquiring such knowledge to a high extent determines the forms of its notation and the choice of method of faults isolation. Thus, these methods determine the obtained quality of diagnosing, with faults' distinguishability among others.

While classifying the knowledge on the diagnosed process one can differentiate four main cases:

- there are not known models of the diagnosed process, only the constraints are known,
- only the qualitative model of the diagnosed process and the constraints are known,
- the quantitative models of the diagnosed process are known, but the influence of the fault is not taken into account,
- the quantitative models of the diagnosed process taking into consideration the influence of faults are known.

Below, one can find the characteristics of the application properties of the methods for the fault detection and isolation for the four above mentioned cases, in respect to diagnostics of the complex industrial processes.

2. Diagnostics of the processes without using models

In the case of not knowing the models of the process mostly are used the control of thresholds of the process variables and control deviations for detecting the faults. Sometimes, the statistical or spectral analysis of the measurement signals are applied. There are used logical functions of alarms or rules for faults isolation, in which premises are the results of constraints checking. Described scheme of diagnosing is shown in Fig. 1.

For recognition of abnormal or emergency states in the systems of industrial processes automation (SCADA, DCS) there is frequently used an alarm signalizing system, which is a simplified version of a diagnostic system. The methods for controlling the limits, which are generally used in the alarm systems for fault detection, have a lot of disadvantages [8], such as:

- occurrence of a very large number of alarms within a short period of time in the states of serious faults. According to EEMUA (The Engineering Equipment and Materials Users' Association) the average daily number of alarms in petrochemical industry is estimated on 1500, and in energy industry on 2000, whereas max. 144 is recommended;
- no possibility of detecting some of the faults due to symptom cancelling effects caused by the control systems (the example of such fault can be the leakage of a toxic substance from the tank of a controlled level),
- long fault detection delays,
- a very large number of possible causes for some alarms.

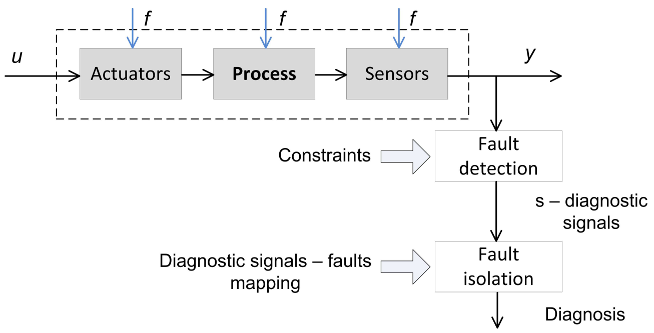

Fig. 1. The scheme of diagnosing based on checking the constraints

Interpreting a large amount of alarms occurring within a very short period of time is a serious problem for the operators, the more because the alarm systems are usually devoid of the mechanisms for faults isolation, i.e. they are unable to point out the real cause of faulty state. The phenomena taking place here is called information overload, what can result in stress [8]. It can lead to additional operator's mistakes, which accumulated with the previously obtained faults may cause serious accidents. The mechanism of such unfavorable (positive) feedback was the cause of many serious breakdowns in nuclear or conventional power stations and chemical plants.

A fault isolation is implemented only in few alarm systems. The diagnostic signals are the results of the control of limits interpreted in a binary or tri-state way {-1, 0, +1}. These signals are the inputs of the algorithm of the fault isolation. The relation between values of the diagnostic signals and the faults takes the form of the logical function (with binary diagnostic signals) or "*IF the values of the diagnostic signals THEN fault(s)*" type of rules with trivalent diagnostic signals.

The diagnostic relation in the form of logical function has been used for a long time, among others in the works [21–23]. Logical function can, in a general case, combine binary diagnostic signals generated with the use of different methods, including the methods of signal analysis. Logical function is individually designed for each fault. This way one can get the procedures for the particular faults isolation without conducting comprehensive diagnostic analysis. However, the logical function might be the same for a variety of faults. When designing these functions for only chosen faults one can be unable to state the lack of isolability between them and the omitted faults. The method does not take into account the uncertainties of symptoms and faults-symptoms relation. It also does not ensure the robustness to the changes of the process's structure and the set of measurements. Applying the rules in respect to fault isolation conducted on the basis of the limits control results and signals analysis has similar disadvantages.

3. Process diagnostics with the use of qualitative models

The qualitative models used in diagnostics of industrial processes are Signed Directed Graphs (SDG) [24, 25]. They represent cause-and-effect relationships between process variables in technological systems. The nodes of the graph are process variables. The nodes corresponding to the measurable variables are observable nodes. The arcs of the graph represent direct relations between the variables in the supervised process. The arrows of the arcs reflect the directions of interactions between the process variables, thus the directions of propagation of the alarms in the process. The arcs of the graph are marked, "+" what stands for the consistent and "$-$" for the opposite direction of changes of the adjoining variables.

SDG graphs are used mainly for the fault isolation in conjunction with the detection conducted on the basis of the control of limits. Additionally, on the basis of the cause-and-effect graphs there can be applied the control of compatibility of directions of changes of the values of process variables. The faults which cause incompatibilities with directions are detected in this way. As an example may serve the control of compatibility of directions of changes in the control signal passed to the control valve and the controlled medium flow. However such test cannot detect slow valve sedimentation process, but is sensitive for all abrupt faults, e.g. blocking of the actuator's piston rod.

The status of the variables in the basic version of the method takes 3 values $\{+, 0, -\}$, where "0" means normal state, "+" means exceeding the high limit of the alarm and "$-$" means exceeding the low limit for the alarm. Faults, which are not clearly represented on the SDG graph, cause the alarms. The fault is manifested by the occurrence of a sequence of alarms. Identification of the initial (causal) alarm enables indicating the faults, which caused this alarm. The diagnosis is then formulated on the basis of the graph analysis. One determines the paths of the graph, which are consistent with the values in the observable nodes. The initial node in such path constitutes causal alarm. There can be many solutions (paths of the graph) consistent with the observed alarms. The diagnosis indicates then a subset of indistinguishable faults. The method has been discussed in the works [23–28]. Shizoka and others [24] suggested the improvement of the method through applying 5 statuses of each variable. It is also possible to implement inference on the basis of the graph with the use of fuzzy logic [26].

In the works [29, 30], the use of GP process graphs, which are the extensions of SDG graphs, was proposed. In the GP graphs are also represented faults, which constitute a separated group of inputs (Fig. 2). This approach, in a clearer way, presents the relations between alarms and faults. GP graph can be used to isolate faults also in relation to detection conducted on the basis of the partial models of the process.

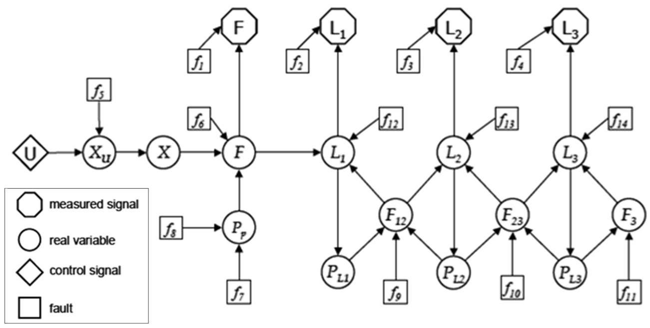

Fig. 2. Exemplary GP graph for the simple process

SDP graphs can be developed on the basis of expert's knowledge: designers, process engineers, control engineers and process operators. That is why these methods are particularly useful in the first stage of diagnostic system design. They can be implemented quickly and at reasonable cost, while applying the methods making use of processes' models require longer time for identification of models and larger financing. That is why the simple methods are recommended to be used in LSS diagnostics.

Qualitative models of the complex processes are much easier to elaborate than the analytical ones. These models describe the relations between process variables and explicitly (GP graphs) or implicitly (SDG graphs) between faults and alarms. Their advantage is a possibility to determine the sequence of alarms for particular faults. Such knowledge can be additionally used in fault inference, what allows for increasing faults isolability in comparison to inference on the basis of the fault-diagnostic signal values (alarms) relation. However, this method can be unreliable. It is difficult to properly define the alarm limits. The values of limits for particular variables are interrelated and defining them in a wrong way may lead to false diagnoses. In addition, the operating control system compensates the influence of some faults, resulting in a lack of certain alarms despite the existence of their causes. Lack of algorithm robustness to the changes of sets of measureable variables (observable nodes) is another disadvantage of inference on the basis of the graphs. It is necessary to modify the graph on a current basis when the above mentioned changes occur.

Apart from the SDG graphs there is also applied abduction logical inference with the use of cause-and-effect graphs (also called AND/OR/NOT graphs). This method was developed in the works of Ligęza and his co-workers [31–35]. Cause-and-effect graph is a model of logical relation between the symptoms (phenomena) describing the behavior of the process. Abduction aims at determining the causes of the discovered breakdown. Thus, this method cannot be used for early fault recognition, e.g. in order to secure the process or re-configure the structure of the control system. The bond-graphs are the other qualitative models used in the diagnosis of processes [36, 37].

4. Diagnostics with the use of quantitative models without taking into consideration the influence of faults

4.1. General scheme. The methods of detection making use of the models of the systems play fundamental role in the diagnostics of processes, whereas these are models built for the normal state of the process [3–16]. Such methods allow for early detection of the small size faults, before they reveal their negative effects. The general scheme of diagnosing with the use of quantitative models not including the faults influence is shown in Fig. 3.

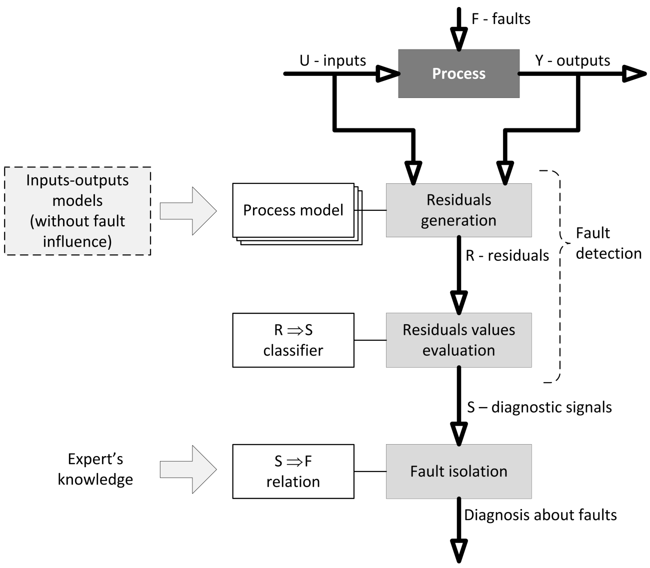

Fig. 3. The general scheme of diagnosing with the use of quantitative models not including the faults influence

4.2. Fault detection. A set of partial models is usually used for fault detection in the complex processes. These models are designed for particular devices and assemblies in their normal state (without any faults). The collection of these models should cover the whole process being diagnosed. The diagnostics based on the partial models has many advantages in relation to the diagnostics based on global models, i.e. of the whole process, such as: shorter detection time, simpler models, lower design costs, greater flexibility of the diagnostic system.

Different kinds of process models can be used for fault detection. The most comprehensive model of the process can be derived directly from the physical equations, e.g. balance equations. Such a model reflects the properties of the process in the whole range of operation. If the equations describing the process have confounding form, then the residuals are calculated as a difference between the left and the right side of the equations (Fig. 4). Designing the models based on the description of the physical phenomena is, for many systems, very difficult or even impossible, because the nature of some of the phenomena occurring in the industrial processes is not known. This limits the application, of this method to processes, which are described by relatively simple relations.

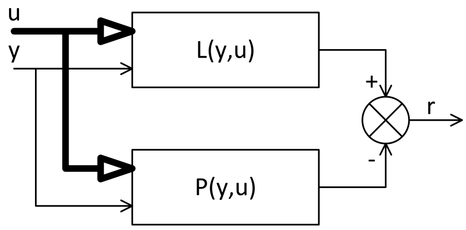

Fig. 4. Generation of residuals on the basis of the equations (in a confounding form) describing physical phenomena

Linear models in the form of state equations, operator transmittances, state observers or Kalman filters [3, 10, 11] have limited applicability in the diagnostics of industrial processes due to non-linearity of the processes and variable operating point. In the case of such models, the residuals are specified as a difference between the measured and modeled values of the output signals (Fig. 5).

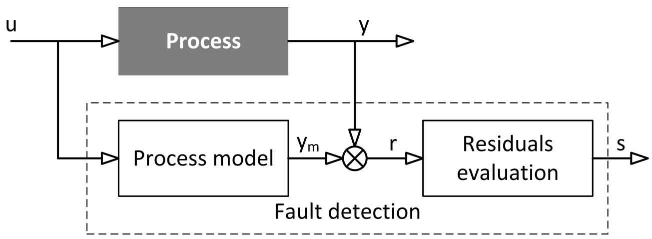

Fig. 5. The diagram of residues' generation and their evaluation

The fault detection based on the estimation of the parameters of the process models also has limited applicability [38]. In this case, the residuals constitute the differences between the nominal and estimated values of the parameters of the process models or physical factors evaluated on their basis. This method is mainly used for the well-defined processes, such as mechanical and electric processes, but rarely for the thermal or chemical ones, because it is difficult to define the appropriate models. Its disadvantages are: detection delay, high computational power expenditures together with the necessity of on-line identification of the parameters of the process models and problems with detecting additive faults.

Approximation models are applied due to the difficulties and restrictions in use of the analytical models. They represent only selected functional features of the real process with proper accuracy. They are created on the basis of the measurement data and expert's knowledge about the model's structure. The range of usefulness of this type of models is limited to the range of input and output signals, on the basis of which the model were created.

Neural and fuzzy models and their combinations [6–8, 11, 14, 18] are of particular practical importance. Such models are tuned based on the measurement data recorded during process exploitation. They give a good image of the system in the scope of variability of the signals, on the basis of which they have been trained. Industrial applications are dominated by unidirectional, multilayer neural networks with delays in the input signals. Another solution are unidirectional, multilayer networks with dynamic neurons [6, 18]. There are also used RBF radial networks and GMDH networks [8, 10, 16, 39–41].

Neural-fuzzy systems [5–8, 11, 14, 42] are a convenient tools for modeling for the purpose of residual generation on the basis of the measurement data and expert's knowledge. They have a lot of well-known advantages and a disadvantage relaying in the fact that the number of rules grows rapidly with the increase in the number of inputs and the number of fuzzy sets defined for particular inputs. This limits their application to the systems with the relatively small number of inputs. However, in diagnostics of industrial processes, partial models are utilized. Due to this fact, restrictions of neural-fuzzy systems are not that important. Moreover, the effective method of limitation of the number of model inputs is the variable aggregation approach [8]. It relies on replacing the subset of input signals with the signal being its properly selected function. In order to make use from this approach one needs to use knowledge concerning physical phenomena taking place in the process.

The additive model is a new developed and highly promising method of modeling of the processes for the purpose of fault detection. They had been discussed in the work [43], whereas Łabęda-Grudziak had applied them for the fault detection [44, 45]. The additive model of the MISO structure for input signals $X_1, X_2, ..., X_p$ and one output signal Y is as follows:

$$Y = \alpha + \sum_{j=1}^{p} \varphi_j(X_j) + \varepsilon, \tag{1}$$

where: error ε is independent from $(X_1, X_2, ..., X_p)$, $E(\varepsilon)+0$, $Var(\varepsilon) = \sigma^2$ and φ_j are one-dimensional functions of X_j variable, not necessarily linear, estimated on the basis of data. The assumption of independence of the input signals is not required. Relations between the output and input signals are estimated by using non-parametric techniques of smoothing, such as natural cubic splines [44, 45]. The model is tuned with the use of the iteration backfitting algorithm [42]. Research conducted by Łabęda-Grudziak [45] shows high practical usefulness of this method for industrial applications.

All mentioned methods of detection, based on the quantitative models are quite troublesome in exploitation. Each maintenance and modernization of the technological installation require repeated tuning of these models. When a nonstationary systems are considered, the passage of time is also such a factor.

A very important issue is ensuring the robustness of detection algorithms for inaccuracy of modeling, disturbances and measurement noises. The known active approaches [3, 13], consisting in generating such signal of residual that will not be sensitive to disturbances and inaccuracy of the model and at the same time will detect the faults, are in practice very difficult to obtain. They require deep knowledge on the process, e.g. observers of unknown input or parity equations in the internal form. Passive methods have much important practical significance. The robustness of the algorithm is achieved in the decision phase and not in the phase of residual generation.

One of the approaches to build robust models is the model identification with limited value of identification error [46, 47]. The approach based on determination of statistical error bounds has the greatest practical importance [7, 48–53]. The identification process is carried out without consideration of its uncertainty, while, in the second step, the model uncertainty is modeled (error model) based on residual signal. The robustness is determined by the adaptive threshold signal applied to residual. The methodology of forming the envelope of uncertainty in the time domain in respect to fuzzy and neural models is intensively developed at the University of Zielona Góra [48–52]. Fault signaling takes place after exceeding by the residual value upper or lower envelope of the area of uncertainty – adaptation limit.

However, in the case of large scale systems, on-line determination of residuals decision thresholds requires high computational power. The simpler, and comparably efficient solution [45] is the application of fuzzy evaluation of the residuals [6–8, 14, 54] in conjunction with experimental determination of the parameters of fuzzy sets. The fuzzy evaluation enables taking into account the uncertainty of modeling errors, disturbances, measurement noise and the problem of the proper specification of the threshold values. In the simplest case there is used fuzzy bi-valued evaluation of the absolute value of the residual. In the case of tri-valued evaluation (Fig. 6) the sign of residual is additionally taken into account, what can increase fault distinguishability indices [8—55].

The result of fuzzy residuals evaluation are fuzzy diagnostic signals. The value of fuzzy signal is therefore determined by the factors of membership of the calculated residual value to particular fuzzy sets. The advantage of this approach is a possibility of experimental assessment of parameters of the fuzzy sets used while making decision about the occurrence of a fault. These parameters can be fixed automatically on the basis of the analysis of statistical parameters of the residual time series in a normal process state, what is presented in Fig. 7.

Making decision not on the basis of current value of residuals, but on the basis of its value in the time window of specific length is another simple method for increasing the robustness of fault detection.

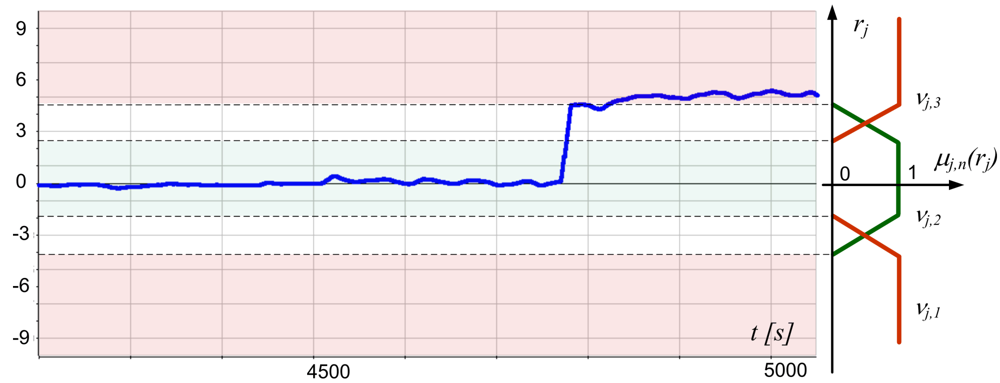

Fig. 6. The schematics of tri-valued evaluation of residuals

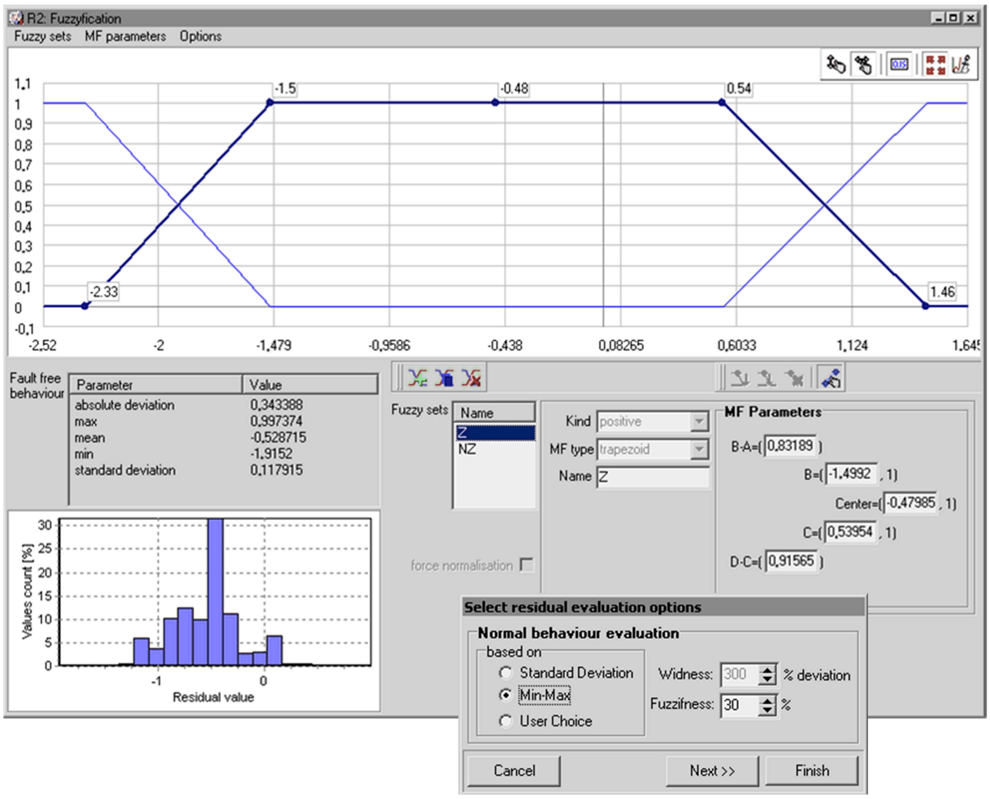

Fig. 7. Graphic interface for the determining the parameters of fuzzyfication of the residuals in AMandD system

4.3. Fault isolation. There are two possible approaches to specifying the faults-diagnostic signals values relation in the case where fault detection is model-based without taking into consideration the influence of faults. The first one rely on learning on the basis of the sequence of measurement data for the normal condition and conditions with faults, and the second on using expert's knowledge.

The method of learning [5, 6, 14, 18] is a very attractive way of getting knowledge on relations between diagnostic signals values and faults. However, in order to learn, one needs to have measurement data characterizing all the states of the process, which should be recognized, so the normal condition of the process, as well as states with faults. Getting such data from exploitation of the complex process is usually impossible. Unacceptable, and often unworkable is introducing faults in the real processes. The condition of applying this method is then the knowledge on the process's model, which allows for the simulation of the faults. This means that we have to dispose of models taking into account the influence of the faults.

When it comes to industrial processes, gathering measurement data for all the states of the process is impossible. The number of the possible faults is very high and particular abnormal or emergency conditions occur very rarely. Moreover, technological installations in chemical, power or food industries are, in major part, individual solutions or are realised in rarely series. It all makes getting the sequence of leaning data, representing particular emergency conditions – impossible. Diagnostic system should therefore detect and recognize serious breakdowns, which had never occurred before. Methods requiring determination of symptoms-faults relation in the stage of learning have then limited application in the diagnostics of industrial processes. However, they are highly useful in diagnostics in serial production of e.g. engines, pumps.

In diagnostics of complex technological systems the most significant are methods utilizing expert's knowledge in the process of designing fault-symptoms relation. Good knowledge about the process allows for determining this relation in a relatively simple way. The designer of a diagnostic system can additionally use the knowledge of process engineers, process operators and maintenance staff.

Expert's knowledge about faults-symptoms relation can be presented in many different forms. When using binary evaluation of residuals, the diagnostic relation can take a form of [5–8]: logical function, diagnostic trees, binary diagnostic matrix or rules of different forms. The most often used are the rules that correspond to columns (2) or rows (3) of the binary diagnostic matrix:

$$\begin{gathered}\text{if} \quad (s_1 = \nu_{1,k})... \wedge (s_j = \nu_{j,k})... \wedge (s_J = \nu_{J,k}) \\ \text{then} \quad (f_k); \nu_j \in 0, 1,\end{gathered} \tag{2}$$

$$\text{if} \quad (s_j = 1) \quad \text{then} \quad (f_a \vee ... \vee f_k \vee f_n). \tag{3}$$

Similar forms of presentation are used with multi-valued evaluation of the residuals. The extension of a binary diagnostic matrix is Fault Isolation System (FIS) [6–8, 55, 56]. The extensions in respect to the binary diagnostic matrix are as follows:

a) with every diagnostic signal can be associated an individual set of its values V_j,
b) set V_j of j-th value of a diagnostic signal can be multi-valued,
c) any FIS element can contain both – single value or subset of values of a diagnostic signals.

FIS rules have the form of:

$$\begin{gathered}\text{if} \quad (s_1 \in V_{1,k})... \wedge (s_j \in V_{j,k})... \wedge (s_J \in V_{J,k}) \\ \text{then} \quad (f_k),\end{gathered} \tag{4}$$

where $V_{j,k}$ is a subset of possible values of s_j in the state of fault f_k, or:

$$\text{if} \quad (s_j = \nu) \quad \text{then} \quad (f_a \vee ... \vee f_k \vee f_n). \tag{5}$$

Rule (5) corresponds to subset of faults $F(s_j = \nu)$, which can generate symptoms $s_j = \nu$.

In the case of fuzzy bi-valued evaluation of the residuals applied in the record of diagnostic relation, there are used binary diagnostic matrix or principles (2) or (3), whereas when fuzzy multi-valed evaluation – FIS system or principles (4), or, eventually (5).

When choosing the form of notation of relation between faults and diagnostic signals values a deciding factor is the robustness of diagnostic system to changes in the structure of the process being diagnosed, including the changes in the set of fault-less measurement paths. Logical functions, diagnostic trees, binary diagnostic matrix, information system, which are specified at the stage of designing of the diagnostic system are rigid and not robust to the changes of structure of the process. Also in the rules of (2) or (4), together with changes of the set of the accessible measurement signals, there is also a change in the set of premises. Moreover, in the case of large scale systems, where the number of realized tests is huge, the rules corresponding to the columns of binary diagnostic matrix or information system are inconvenient due to a very large number of premises.

The method of diagnostic relation notation, robust in the sense of the possible changes of the process's structure, are rules of a (3) and (5) types, where the particular symptoms correspond to subsets of faults that cause these symptoms. This dependency is constant. In the case of changing the structure of the process or as a result of previous diagnoses, this rule can be temporarily eliminated from the set of active rules, but its form is constant. What is more, this rule has a compact form, because the number of possible faults presented in the conclusion is not high, especially in the case of using partial models.

On the basis of rules (3), (5) one can automatically reconstruct the binary diagnostic matrix or FIS, thus the rules corresponding to the fault signatures – (2), (4) respectively. Such rules can be contradictory, i.e. they may have the same premises and different conclusions. They correspond to the unisolable faults.

There is a wide range of methods for faults isolation. In work [42] there are 2 main groups: classification methods and methods of automatic inference. The classification methods are difficult to apply in LSS due to the above mentioned difficulties with specifying pattern data for the states with faults. Automatic inference is conducted on the basis of rules. Due to the variability of the structure of a diagnosed process, the methods of automatic inference are better suitable in the industrial processes diagnoses. Uncertainties of diagnostic signals inclines the application of fuzzy evaluation of residuals and inference with the use of fuzzy logic.

The general diagram of inference based on the application of partial models to fault detection and fuzzy inference on the faults is presented in Fig. 8. Characteristic is the lack of the defuzzyfication block. The diagnosis shows faults and corresponding degrees of activation δ of rules, which are interpreted as factors of conviction on appearance of the certain faults.

Different algorithms of diagnostic inference destined to diagnose complex industrial systems are presented in the works [6–8, 55, 57]. Discussing the specific diagnostic algorithm, taking into consideration all the problems included in Sec. 2 exceeds the frames of this paper.

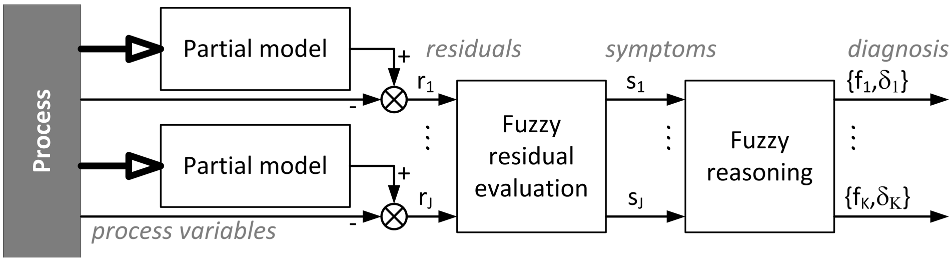

Fig. 8. The diagram of diagnosing with the use of fuzzy logic

5. Diagnostics with the use of quantitative models including the influence of faults

The greatest potential for achieving quick and accurate diagnosis gives the use of analytical models, taking into consideration the influence of faults (6) on the value of the process's outputs:

$$\begin{aligned}\dot{\boldsymbol{x}}(t) &= \phi[\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{f}(t)],\\ \boldsymbol{y}(t) &= \psi[\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{f}(t)]).\end{aligned} \tag{6}$$

From these models one can directly determine the relation between the values of residuals and faults. Therefore, applied scheme of diagnosing (Fig. 9) is similar to this one presented in Fig. 3. The difference lies in the sources of knowledge about the fault-symptoms relation.

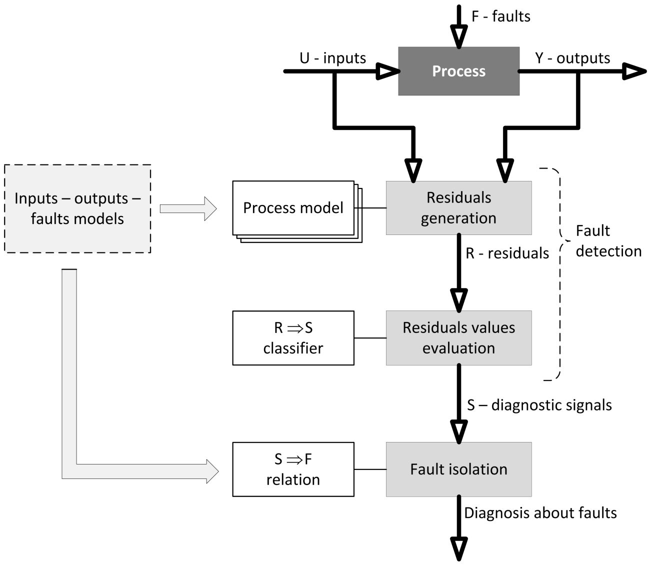

Fig. 9. The general scheme of diagnosing with the use of quantitative models including the faults influence

By carrying out linearization of an equation in the working point of the process and Laplace Transform one achieves transmittance model in the following form:

$$\boldsymbol{y}(s) = \boldsymbol{G}(s)\boldsymbol{u}(s) + \boldsymbol{H}(s)\boldsymbol{f}(s). \tag{7}$$

Residuals resulting from (7) can be presented in two forms [4] – analytical and internal:

$$\boldsymbol{r} = \boldsymbol{y}(s) - \boldsymbol{G}(s)\boldsymbol{u}(s) = \boldsymbol{H}(s)\boldsymbol{f}(s). \tag{8}$$

Analytical form $\boldsymbol{r} = \boldsymbol{y}(s) - \boldsymbol{G}(s)\boldsymbol{u}(s)$ shows the dependency of residual from input and output signals and can be used to faults detection, whereas internal form $\boldsymbol{r} = \boldsymbol{H}(s)\boldsymbol{f}(s)$ defines the dependency of residual from faults and is a basis for fault isolation.

Table 1
Internal form of the residuals

	f_1	...	f_k	...	f_K
r_1	$H_{1,1}$		$H_{1,k}$		$H_{1,K}$
...					
r_j	$H_{j,1}$		$H_{j,k}$		$H_{j,K}$
...					
r_J	$H_{J,1}$		$H_{J,k}$		$H_{J,K}$

In practice, application of linear models for fault detection in industrial processes, usually non-linear, is not recommended. One usually uses models describing physical phenomena without conducting linearization. If they were designed with fault influence taken into consideration, then one assumes that fault vector $f(t) = 0$. The alternative are neural or fuzzy models.

One can claim that in the case of knowing the internal form, the level of knowledge on diagnostic relation is higher than in other forms of notation of this relation. The internal form includes information on dynamics of the fault influence on the residuals. One can, of course, simplify this relation and derive other forms of notation of the faults – diagnostic signals values relation, such as binary diagnostic matrix, FIS system, rules (2), (3) and others.

The knowledge on the internal form allows also for application of, apart from the method of structural residuals, the method of directional [3, 4] or sequence residuals [58, 59]. All these methods enable designing secondary residuals on the basis of primary ones what allows for forming residuals sensitivity for different subsets of faults and leads to increasing the distinguishability of the faults.

The methods of structural and directional residues are commonly known. In the case of structural residuals one uses only information on sensitivity or non-sensitivity of the residuals for particular faults. In this case H matrix is simplified to binary diagnostic matrix. The evaluation of the current values of the residuals is binary. Isolability of faults is achieved when the signatures of the faults (columns of the binary diagnostic matrix) are different for particular faults [4, 6–8].

In the case of directional residuals the isolation is conducted on the basis of the analysis of position of directional vectors in the space of residuals [3, 4, 6, 9]. Directional vectors are determined by residuals gains for particular faults.

It means that instead of $H_{j,k}$ transmittance there is used only $c_{j,k}$ gains of the particular residual on k-th faults. It is assumed that particular $H_{j,k}$ transmittances are of static character, i.e. don't possess integrate part. A given fault, after decaying of the transient states appears always at this direction. In order to get a possibility of fault isolation, the set of residuals is designed in such a way that particular faults result in orientation specific for them in the parity space.

The above mentioned methods did not take into consideration dynamic influence of faults on residuals. The method of sequential residuals [52, 53] is based on the analysis of sequences of occurring symptoms. The method allows for designing the secondary residuals in such a way so as to achieve individual symptom sequences for particular faults. This method, similar to structural and directional residuals requires the knowledge on residuals in the internal form. Due to this fact the costs of design are high. The method can then be applied in the cases of well-recognized processes, operating in working point vicinity, for which necessary mathematical models can be created. Another area of application are critical processes, where the costs of design are negligible compared to the potential losses in the emergency states.

It can be stated that in diagnostics of large scale processes do not apply well established theoretical methods of fault recognition with the use of structural, directional, sequential residuals, observers of the unknown inputs etc. [3, 4, 6, 9, 10, 12, 13, 60]. It results from the difficulty of obtaining the description of the process taking into account the influence of faults on the value of residuals. It has to be therefore highlighted that only in the case of the knowledge on the internal form of residuals it is possible to identify the faults, i.e. defining its size and time of development.

6. Summary

In this work application properties of the methods for fault detection and isolation for the aforementioned four cases, differing in the degree of knowledge about the process with respect to the problem of diagnostics of the complex industrial processes, are described. Table 2 presents these methods and their properties.

Main detection and isolation methods and their exploitation properties corresponding to the four highlighted in Sec. 1 cases of diagnosing, characterized by various degrees of knowledge about the diagnosed process, are summarized in Table 2. The degree of knowledge about the process to a decisive extent determines the achievable quality of diagnosing, defined as the ability of early detection and precise isolation of faults.

Column 1 corresponds to the situation where there are no known models of the diagnosed object and only limits are known. This is the simplest approach, but it has many drawbacks. There is a possibility of using simple fault isolation methods in the case when the qualitative model of the diagnosed process and limits are known (column 2). However, the quality of diagnosing increases slightly, due to the limitation of fault detection for limits control and known weaknesses of this method. The quality of diagnosing is growing strongly in the case of the use of quantitative models for fault detection and isolation algorithms using experts' knowledge (column 3). In this case, early detection of a small-size faults and the diagnosis indicating possible faults are obtained. The precision of diagnosis is dependent on the obtained fault distinguishability, which is decisively influenced by the selection of a set of measurements and the appropriate design of the set of detection algorithms and the diagnostic relation. The highest quality of diagnosis is obtained when the quantitative models of the diagnosed object taking into account the fault influence are known (column 4). In this case, the diagnostic relation is derived from the internal form of the equations describing the process. The most advanced methods of fault isolation and identification may be used. However, this approach is extremely difficult for practical use, because achieving models with fault influence for complex processes is very difficult and expensive.

The simplest approach without models of a diagnosed process applied, but only with the limits of the process variables is commonly used. It seems that the methods based on quantitative models applied for fault detection in relation with the isolation methods making use of expert's knowledge about the relations between faults and diagnostic signals values will dominate in the nearest future. Methods based on the analytical models taking into account the influence of faults will be applied only for critical objects that pose a serious threat to the human safety, environment and the technology installation.

Problems connected with application of diagnostic systems for large scale processes will force new fields of research to appear, such as:

- the development of diagnostic methods for variable structure processes, batch processes and hybrid processes (continuous-discrete),
- ensuring great robustness of diagnostic algorithms to measurement uncertainties, process models and faults-symptoms relation, and also for the symptoms delays, multiple faults, changes of the process' structure etc.,
- the development of methods for diagnosing multiple faults,
- the development of methods of distributed diagnostics,
- the development of specialized sensors for detecting and/or measurement of different kinds of destructive phenomena,
- the development of software tools for designing on-line diagnostics of the process and fault tolerant systems,
- integration of diagnostic systems with the process control systems and software for analysing the process' safety.

Knowledge about the industrial processes diagnostics is now so well established, that one can expect fast replacement of the simple alarm systems by the advanced diagnostic systems.

Table 2
The comparison of the applied methods and properties of on-line diagnostics depending on the level of knowledge about the diagnosed process

Methods/properties	The degree of knowledge about the diagnosed process			
	Small	Medium	Large	Very large
The type of available knowledge	The models are not known, the limitations are known	Qualitative model and the restrictions are known	Quantitative models not including the influence of the faults are known	Quantitative models including the influence of the faults are known
The applied detection methods	• limits control • spectral analysis	• limits control • qualitative models and simple heuristic dependencies	• analytical models • neural models • fuzzy models • statistical models	• analytical models • other models
The methods of acquiring knowledge about the fault-symptoms relation	• expert knowledge • observations – monitoring of theprocess	• expert knowledge • observations – monitoring of the process	• expert knowledge • training on the basis of measurement data	• modelling, including the influence of the faults
The form of notation of faults-symptoms relation	• logical functions • rules	• logical functions • rules • binary diagnostic matrix	• binary diagnostic matrix • FIS information system • rules, logical functions • diagnostic trees • areas in the space of residuals	• internal form of structural residuals • directional residuals • sequential residuals
The applied isolation methods	• inference on the basis of logical functions • inference on the basis of rules	• inference on the basis of rules • analysis of propagation of alarms and directions of interactions between variables – specifying root alarm	• inference on the basis of signatures (classification) • automatic reasoning	• classification • automatic reasoning
The possibility of faults identification	not possible	not possible	limited	exists
The quality of diagnosis	low	medium	high	very high
Remarks	• concealing the symptoms by control loops • detecting medium and large faults • low distinguishability of faults • long time of detection and isolation • detection algorithms do not require tuning after maintenance of the process	• concealing the symptoms by control loops • medium distinguishability of faults • long time of detection and isolation • detection algorithms do not require tuning after maintenance of the process	• a possibility of early fault detection • a possibility of detecting small faults • high distinguishability of faults • models require frequent tuning (periodically, after maintenance)	• a possibility of early fault detection • a possibility of detecting small faults • a possibility of achieving the greatest distinguishability of faults (structuring of residuals) • costs and difficulties with gaining models including the influence of faults

Acknowledgements. This work was supported in part by the National Science Center under the project number DEC-2011/01/B/ST7/06183.

REFERENCES

[1] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes – Theory and Application*, Prentice-Hall, Englewood Cliffs, 1993.

[2] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*, Springer, Berlin, 2004.

[3] J. Chen and R. Patton, *Robust Model Based Fault Diagnosis for Dynamic Systems*, Kluwer Academic Publishers, Boston, 1999.

[4] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*, Marcel Dekker, Inc., New York, 1998.

[5] R. Isermann, *Fault Diagnosis Systems. An Introduction from Fault Detection to Fault Tolerance*, Springer, Berlin, 2006.

[6] J. Korbicz, J.M. Kościelny, Z. Kowalczuk, and W. Cholewa, *Fault Diagnosis: Models, Artificial Intelligence Methods, Applications*, Springer, Berlin, 2004.

[7] J. Korbicz and J.M. Kościelny, *Modeling, Diagnostics and Process Control. Implementation in the DiaSter System*, Springer, Berlin, 2010.

[8] J.M. Kościelny, *Diagnostics of Automated Industrial Processes*, Exit, Warszawa, 2001, (in Polish).

[9] R. Patton, P. Frank, and R. Clark, *Issues of Fault Diagnosis for Dynamic Systems*, Springer, Berlin, 2000.

[10] M. Witczak, *Modelling and Estimation Strategies for Fault Diagnosis of Non-Linear Systems, From analytical to soft computing approaches*, Springer, Berlin, 2007.

[11] J.M.F. Calado, J. Korbicz, K. Patan, R.J. Patton, and J.M.G. Sá da Costa, "Soft computing approaches to fault diagnosis for dynamic systems", *Eur. J. Control* 7 (2–3), 248–286 (2001).

[12] P.M. Frank, "Fault diagnosis in dynamic systems via state estimations methods. A survey", in *System Fault Diagnostics, Reliability and Related Knowledge Based Approaches*, eds. S.G. Tzafestas, vol. 2, D. Reidl Publishing Company, Dordrecht, 1987.

[13] P.M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy", *Automatica* 26, 459–474, (1990).

[14] P.M. Frank and T. Marcu, "Diagnosis strategies and system: principle, fuzzy and neural approaches", in *Intelligent Systems and Interfaces*, eds. H.-N Teodorescu, Kulwer, Boston, 2000.

[15] R. Isermann and P. Ball, "Trends in the application of model-based fault detection and diagnosis of technical process", *Control Eng. Practice* 5 (5), 709–719, (1997).

[16] J. Korbicz, "Robust fault detection using analytical and soft computing methods", *Bull. Pol. Ac.: Tech.* 54 (1), 75–88, (2006).

[17] S. Leonhardt and M. Ayoubi, "Methods of fault diagnosis", *Control Eng. Practice* 5 (5), 683–692 (1997).

[18] R.J. Patton, C.J. Lopez-Toribio, and F.J. Uppal, "Artificial intelligence approaches to fault diagnosis for dynamic systems", *Int. J. Appl. Mathematics and Computer Science* 9 (3), 471–518, (1999).

[19] J.M. Kościelny, M. Bartyś, and M. Syfert, "Practical problems of faults isolation in complex industrial systems", Chapter 10 in: *Intelligent Information Extraction for Diagnostic Purposes*, eds. Z. Kowalczuk and B. Wiszniewski, pp. 167–185, PWNT, Gdańsk, 2007.

[20] J.M. Kościelny and M. Syfert, "Diagnosis of industrial processes from theory and application perspectives", *9th Workshop on Advanced Control and Diagnosis, ACD'2011* 1, CD-ROM (2011).

[21] P.K. Andow and F.P. Lees, "Process computer alarm analysis: outline of a method based on list processing", *Trans. Instn. Chem. Eng.* 53, 195–205 (1975).

[22] K. Plamping and P.K. Andow, "The design of process alarm systems", *Trans. Inst. MC* 5 (3), 161–168 (1983).

[23] M. Iri, K. Aoki, E. O'Shima, and H. Matsuyama, "An algorithm for diagnosis of system failures in the chemical process", *Computers & Chemical Engineering* 3, 489–493 (1979).

[24] J. Shiozaki, H. Matsuyama, K. Tano, and O'Shima, "Fault diagnosis of chemical processes by the use of signed directed graphs: extension to five-range patterns of abnormality", *Int. Chemical Engineering* 37 (4), 651–659 (1985).

[25] B. Shibata, S. Tateno, Y. Tsuge, and H. Matsuyama, "Fault diagnosis of the chemical process utilizing the signed directed graph – improvement and evaluation of the diagnosis accuary", *IFAC Symp. on Fault Detection, Supervision and Safety for Technical Process – SAFEPROCESS' 1991* 2, 381–386 (1991).

[26] J. Montmain and L. Leyval, "Causal graphs for model based diagnosis", *IFAC Symp. Fault Detection, Supervision and Safety for Technical Process – SAFEPROCESS'94* 1, 347–355 (1994).

[27] K. Takeda, B. Shibata, Y. Tsuge, and H. Matsuyama, "The improvement of fault diagnosis algorithm using signed directed graph", *IFAC Symp. Fault Detection, Supervision and Safety for Technical Process – SAFEPROCESS'94* 1, 368–373 (1994).

[28] S. Tateno, B. Shibata, Y. Tsuge, and H. Matsuyama, "Optimal allocation of sensor for fault diagnosis system using signed directed graph", *IFAC Symp. Fault Detection, Supervision and Safety for Technical Process – SAFEPROCESS'94* 2, 736–741 (1994).

[29] J.M. Kościelny and A. Ostasz, "Application of causal graph for description of diagnosed process", 5^{th} *IFAC Symp. Fault Detection, Supervision and Safety for Technical Process – SAFEPROCESS' 2003* 2, 879–884 (2003).

[30] A. Ostasz, "Cause-effect graph of the process and its utilization for determining a set of residuals and diagnostic relations", *PhD Thesis*, Warsaw University of Technology, Warszawa, 2006, (in Polish).

[31] P. Fuster-Para and A. Ligęza, "Diagnostic knowledge representation and reasoning with use of and/or/not causal graphs", *System Science* 22 (3), 53–61 (1996).

[32] A. Ligęza and P. Fuster-Parra, "AND/OR/NOT causal graphs – a model for diagnostic reasoning", *Applied Mathematics and Computer Science* 7, 185–203 (1997).

[33] A. Ligęza and M. Kirchner, "Diagnosis as a process of searching a cause and effect graph. Selected problems and their solutions", *1st National Scientific-Technical Conf. Diagnostics of Industrial Processes* 1, 72–75 (1996), (in Polish).

[34] A. Ligęza, "Functional causal graphs. Yet anpther model for diagnostic reasoning", 3^{th} *IFAC Symp. on Fault Detection, Supervision and Safety for Technical Process – SAFEPROCESS' 2000* 2, 1102–1107 (1997).

[35] A. Ligęza, "Abduction diagnostic inference for the model in the form of function causal-effect graphs", *3rd National Scientific-Technical Conf. Diagnostics of Industrial Processes* 1, 67–72 (1998), (in Polish).

[36] M. Khemliche, B.O. Bouamama, and H. Haffaf, "Sensor placement for component diagnosability using bond-graph", *Sensors and Actuators A: Physical* 132 (2), 547–556 (2006).

[37] P.J. Mosterman, R. Kapadia, and G. Biswas, "Using bond graphs for diagnosis of dynamic physical systems", *Principles of Diagnosis* 1, CD-ROM (1995).

[38] R. Isermann, "Process fault detection based on modeling and estimation. Methods – a survey", *Automatica* 20 (4), 387–404 (1984).

[39] M. Witczak, J. Korbicz, M. Mrugalski, and R. Patton, "A GMDH neural network-based approach to robust fault diagnosis: application to the DAMADICS benchmark problem", *Control Eng. Practice* 14 (6), 671–683 (2006).

[40] C.T. Kowalski and M. Kaminski, "Rotor fault detector of the converter-fed induction motor based on RBF neural network", *Bull. Pol. Ac.: Tech.* 62 (1), 69–76, DOI: 10.2478/bpasts-2014-0008 (2014).

[41] K. Bartecki, "PCA-based approximation of a class of distributed parameter systems: classical vs. neural network approach", *Bull. Pol. Ac.: Tech.* 60 (3), 651–660, DOI: 10.2478/v10175-012-0077-7 (2012).

[42] S. Leonhardt and M. Ayoubi, "Methods of fault diagnosis", *Control Eng. Practice.* 5, 683–692 (1997).

[43] T. Hastie and R. Tibshirani, *Generalized Additive Models*, Chapman and Hall, London, 1990.

[44] Z.M. Łabęda-Grudziak, "Identification of dynamic system additive models by kdd methods", *Measurement Automation and Control* 57 (3), 249–252 (2010).

[45] Z.M. Łabęda-Grudziak, "The use of additive regression model to generate residuals for the purpose of fault detection", *PhD Thesis*, Warsaw University of Technology, Warszawa, 2011, (in Polish).

[46] K. Patan and J. Korbicz, "Fault detection in catalytic cracking converter by means of probability density approximation", *Engineering Applications of Artificial Intelligence* 20 (7), 912–923 (2007).

[47] K. Patan, *Artificial Neural Networks for the Modelling and Fault Diagnosis of Technical Processes*, vol. 377, Lecture Notes in Control and Information Sciences, Springer, Berlin, 2008.

[48] K. Patan, M. Witczak, and J. Korbicz, "Towards robustness in neural network based fault diagnosis", *Int. J. Applied Mathematics and Computer Science* 18 (4), 443–454 (2008).

[49] K. Duzinkiewicz, "Set of parameters and variables in dynamic networks by recursive algorithms with a moving measurement window", *Int. J. Applied Mathematics and Computer Science* 16 (2), 209–217, (2006).

[50] M. Milanese, "Set membership identification of nonlinear systems", *Automatica* 40 (6), 957–975 (2004).

[51] M. Mrugalski, "An unscented Kalman filter in designing dynamic GMDH neural networks for robust fault detection", *Int. J. Applied Mathematics and Computer Science* 23, 157–169 (2013).

[52] M. Mrugalski, "Advanced neural network-based computational schemes for robust fault diagnosis", in *Studies in Computational Intelligence*, ed. J. Kasprzyk, vol. 510, Springer International Publishing, Berlin, 2014.

[53] V. Puig, A. Stancu, T. Escobet, F. Nejjari, J. Quevedo, and R.J. Patton, "Passive robust fault detection using interval observers: Application to the DAMADICS benchmark problem", *Control Engineering Practice* 14 (6), 621–633 (2006).

[54] J.M. Kościelny and M. Syfert, "Fuzzy Diagnostic Reasoning that Takes Into Account the Uncertainty of the Faults-Symptoms Relation", *Int. J. Applied Mathematics and Computer Science* 16 (1), 27–35 (2006).

[55] J.M. Kościelny, D. Sędziak, and K. Zakroczymski, "Fuzzy logic fault isolation in large scale systems", *Int. J. Applied Mathematics and Computer Science* 9, 637–652 (1999).

[56] J.M. Kościelny and M. Syfert, "Fuzzy logic application to diagnostics of industrial processes", *5th IFAC Symp. Fault Detection, Supervision and Safety of Technical Processes SAFEPROCESS'2003* 2, 771–776, (2003).

[57] J.M. Kościelny, "Fault Isolation in Industrial Processes by Dynamic Table of States Method", *Automatica* 31, 747–753 (1995).

[58] J.M. Kościelny, M. Syfert, and Ł. Tabor, "Application of knowledge about residual dynamics for fault isolation and identification", *2nd Conf. Control and Fault-Tolerant Systems (SysTol'2013* 1, 275–280 (2013).

[59] J.M. Kościelny, M. Syfert, and Ł. Tabor, "Sequential residual design method for linear systems", *Conf. Control and Fault-Tolerant Systems SysTol'10, Nice* 1–6, CD-ROM (2010).

[60] D. Krokavec and A. Filasová, "Novel fault detection criteria based on linear quadratic control performances", *Int. J. Applied Mathematics and Computer Science* 22, 929–938 (2012).