Int. J. Appl. Math. Comput. Sci., 2025, Vol. 35, No. 2, 235–250
DOI: 10.61822/amcs-2025-0017

INVESTIGATING THE PROBLEM OF MISDIAGNOSIS IN MODEL–BASED FAULT DIAGNOSIS

JAN MACIEJ KOŚCIELNY [a], MICHAŁ BARTYŚ [a,*]

[a] Institute of Automatic Control
Warsaw University of Technology
Boboli 8, 02-525 Warsaw, Poland
e-mail: {jan.koscielny,michal.bartys}@pw.edu.pl

This paper deals with one partly unconscious property of the model-based diagnosis. It discusses occasional contradictions between diagnoses that are logically correct but, in fact, are not consistent with the physical state of the system being diagnosed. This property is studied and discussed based on the analysis of diagnoses generated by four selected approaches using binary and trivalent diagnostic signals. The authors attribute the reasons for this inconsistency to the effect of compensation of fault impacts. The analysis and simulation studies carried out confirmed this assumption. To address this problem, new definitions of diagnoses have been proposed that reflect the different degrees to which diagnoses relate to the actual physical state of the system being diagnosed. In this context, several new metrics for assessing the quality of diagnoses have also been proposed. It is pointed out that, from the utilitarian point of view, only those diagnoses that are logically consistent and have the attribute of physicality are valuable. The problem of misdiagnosis was illustrated on an example of a two-tank system.

Keywords: diagnostic reasoning, misdiagnosis, fault isolation, fault distinguishability.

1. Introduction

The implementation of diagnostic inference, which will at least reduce the number of diagnoses that are inconsistent with the physical state of the system being diagnosed, is of great practical importance. For example, inconsistent diagnoses may lead to erroneous decisions by process operators. In the process industry, a properly designed diagnostic system can prevent critical failures that threaten the health and life of humans, technological installations, and the environment (Song and Jiang, 2022; Xia and Fu, 2024). This implies the need to generate correct diagnoses, as rational safety measures can only be taken if the causes of observed anomalies are correctly identified (Blanke *et al.*, 2015). This also places high demands on the accuracy and reliability of diagnostics generated by advisory diagnostic systems.

In this paper, we will discuss the problem to what extent logically correct diagnoses are really consistent with the physical state of the system being diagnosed. This problem was hinted at some time ago in an article by Struss and Dressier (1989), although to the best of our knowledge it has not been discussed further. This paper points out that even formally correct diagnoses are not necessarily true in terms of their conformity with the physical state of the diagnosed system.

Most of the model-based diagnostic systems in use infer about faults based on binary evaluated residuals. The values of the binary residuals evaluated hereafter will be referred to as binary diagnostic signals. Diagnosis with binary diagnostic signals usually results in the generation of a large number of potential diagnoses. Potential diagnoses indicate alternative causes for the observed diagnostic signals. However, a significant number of potential diagnoses indicate physically impossible states of the system being diagnosed. In such cases, the risk of inappropriate operator reaction increases. Unfortunately, there is a lack of awareness of these risks. The problem will be illustrated in this paper by an example of the diagnosis of a set of serially interconnected tanks introduced by Kościelny and Bartyś (2023).

An additional motivation for this work is the observation that the effect of compensation of fault

*Corresponding author

impacts on residuals is one of the important and still unsolved diagnostic problems. We will show that this effect has a significant impact on the outcome and credibility of the diagnoses.

Hypothesis. *Fault isolation methods based on binary diagnostic signals or binary observations, although logically correct, can generate diagnoses that indicate physically impossible states of the system being diagnosed.*

Hence, the primary motivation for this paper is to provide a new perspective and revise the views on the problem of consistency of correct logical diagnostics with the physical state of the diagnosed system. In particular, in this paper we:

1. demonstrate that generating potential diagnoses of physically impossible states is an immanent feature of model-based fault isolation approaches, where fault isolation is based on binary valued residuals,
2. demonstrate that even the absence of modelling errors, disturbances, and measurement noise does not preclude the possibility of generating formally correct potential diagnoses that nevertheless point to impossible physical states,
3. demonstrate that consistency-based approaches based on Reiter's theory can generate potential diagnoses of physically impossible states by a fault compensation effect.
4. indicate that it is advisable to diagnose based on trinary diagnostic signals.

The structure of this paper is as follows. The motivation, novelty, and contribution to the field of diagnostics are presented in the Section 1. A brief characterization of the state-of-the-art in the fault isolation research area is given in Section 2. Section 3 defines the basic concepts and presents the adopted research methodology. Section 4 describes a phenomenological model of a set of two interconnected tanks; further, this model is used for the simulations reported in Section 5. A detailed description of the diagnostic approaches discussed in this paper is given in Section 6. Section 7 discusses the genesis of diagnoses of physically impossible states, while Section 8 provides a breakdown of the diagnosis results based on the fault distinguishability metrics defined in Appendix A. The comments and perspectives of other works given in Section 9 conclude the paper.

2. Review of diagnostic methods

Model-based approaches to Fault Detection and Diagnosis (FDD) can be divided into two classes, which fundamentally differ in the way in which knowledge of the relationship between faults and diagnostic signals is obtained.

The first class consists of approaches that are based on learning (most recently deep learning) from data. Data used for learning purposes are acquired from normal (nominal) states and states with faults in the diagnosed system, and then used for fault classification (Song and Jiang, 2022; Zheng and Zhao, 2022; Liu *et al*., 2023; Kunpeng *et al*., 2023; Eskandari *et al*., 2024). However, for many industrial installations, especially critical facilities such as nuclear power plants or chemical reactors, the above class of approaches has significant application limitations, as it is very difficult and often impossible to acquire data representing emergency conditions in the installations.

The second class of FDD includes approaches that are mainly knowledge-based. They make use of the relationship between faults and diagnostic signals allowing for diagnostic inference in real-time. The approaches in which this relationship is defined based on expert knowledge are of fundamental practical importance in this case. This class is the subject of this paper.

Here, the models that represent the fault-free state of the diagnosed system are used for fault detection. Fault isolation is most often performed on the basis of the fault-binary-valued diagnostic signal relationship. To this class belong:

- works and approaches derived from control theory, referred to as Fault Detection and Isolation (FDI) (Frank, 1990; Gertler, 1998; Chen and Patton, 1999; Su and Chen, 2019; Jia *et al*., 2023; Tatara and Kowalczuk, 2024) in which a binary relationship between faults and diagnostic signals is defined as the fault signature matrix (FSM) (Cordier *et al*., 2004; Travè-Massuyès, 2014) or structure of residual sets (Gertler, 1998), the Boolean decision table (Chen and Patton, 1999), the coding set (Gertler, 1991), or the binary diagnostic matrix (BDM) (Kościelny, 1995; Korbicz *et al*., 2004) is used to isolate the faults;
- key approaches based on formal logic and artificial intelligence known as consistency based reasoning (CBR) (Struss and Dressier, 1989; Reiter, 1987; de Kleer and Williams, 1987; de Kleer *et al*., 1992; de Kleer and Kurien, 2003);
- approaches derived from a structural analysis (SA) (Blanke *et al*., 2015; Düstegör *et al*., 2006; Krysander *et al*., 2007; Armengol *et al*., 2009; Bregón *et al*., 2013; Bregón *et al*., 2014; Pulido and González, 2004), which are used for solving sensor placement problems as well as used for the design and analysis of structures of models intended for

fault detection and isolation as well as for solving sensor placement problems. The SA analysis can be used in both the FDI and CBR approaches;

- other approaches of diagnosing based on binary diagnostic signals (Kościelny, 1995; Bartyś, 2013).

The aforementioned approaches differ, among others, in the formal description of the diagnosed system, diagnostic reasoning, adopted assumptions, and models. The exhaustive analysis and comparison of the approaches FDI and CBR was carried out in (Cordier *et al*., 2004; Travè-Massuyès, 2014).

In addition to binary, evaluated trivalent residuals are also used for fault isolation. The trivalent residuals (crispy and fuzzy) have been deliberated, among others, in (Kościelny and Bartyś, 2023; Bregón *et al*., 2013; Biswas *et al*., 1997; Kościelny, 1999; Puig *et al*., 2005; Kościelny *et al*., 2006; Daigle *et al*., 2009; Bartyś, 2014).

A generalization of binary and trinary evaluated residuals are multivalued residuals. The relationship between faults and multivalued residuals is used in the fault isolation system (FIS) (Korbicz *et al*., 2004; Kościelny *et al*., 2006). In fact, the FIS is an adaptation of the information system introduced by Pawlak (1991).

Inference based on the analysis of fault signatures is commonly used for FDI, while rather row-based fault inference is typical for CBR approaches. In all the aforementioned approaches, it is assumed that by the absence of false values of diagnostic signals arising from multivalued residuals or conflicts in CBR, the generated diagnoses are formally correct and true in the sense of their relationship to reality. In this paper, we challenge the validity of such an assumption.

3. Methodology

We assume that the diagnosis, regardless of the inference approach used, is made up of possible diagnoses. Each potential diagnosis is an alternative hypothesis consistent with the observations of the diagnosed system. Thus, a potential diagnosis indicates a possible state of a system that is being diagnosed by determining a subset of the faults existing in that state (FDI) or a subset of components (CBR) suspected of being faulty.

The primary goal of fault isolation is to obtain a high-precision diagnosis. Generally, fault distinguishability and precision of diagnosis are in a direct mutual relation, i.e., with an increase of fault distinguishability, the precision of diagnosis increases. The definitions of distinguishability and precision of diagnoses have been discussed, among others, in (Kościelny *et al*., 2019; Kościelny *et al*., 2016).

In this paper, we assume that the analyzed fault isolation approaches are formally correct. We understand the formal correctness of diagnostic inference as complying with the adopted principles of logical calculus (the principles of propositional calculus and the principles of quantifier calculus). However, a fundamental question arises whether a formally correct diagnosis reflects the existing state of the diagnosed system.

For the sake of further analysis, we will introduce a few definitions.

Definition 1. A diagnosis is *correct* if there exists a potential diagnosis indicating the physical state of the system being diagnosed.

Definition 2. A diagnosis is *incorrect* if there is no possible diagnosis indicating the actual physical condition of the system being diagnosed.

Definition 3. A diagnosis is *inclusive* if all faults that actually occurred are in the union of subsets of faults indicated in all potential diagnoses.

Definition 4. A diagnosis is *non-inclusive* if not all real faults that really occurred are in the union of the subsets of faults indicated in all potential diagnoses.

Definitions 1 and 3 show the reversibility of the correctness and inclusiveness of the diagnoses. Every correct diagnosis is inclusive, but not every diagnosis that is inclusive is correct.

Example 1. Let us assume that there is a real state with faults $\{f_4 \wedge f_9\}$. According to Definitions 1–4, the diagnosis

$\Delta_a = \{\mathbf{f_4} \wedge \mathbf{f_9}, f_7 \wedge f_{12}\}$ is correct and inclusive,

$\Delta_b = \{\mathbf{f_4} \wedge f_7, \mathbf{f_9} \wedge f_{12}\}$ is incorrect but inclusive,

$\Delta_c = \{\mathbf{f_4} \wedge f_7, f_8 \wedge f_{12}\}$ is incorrect and non-inclusive. ♦

Definition 5. *The diagnosis of a physically possible state* is a potential diagnosis that indicates the state of the system which is physically possible with the observed values of diagnostic signals.

Definition 6. *The diagnosis of a physically impossible state* is a potential diagnosis that indicates the state of the system which is not physically possible with observed values of diagnostic signals.

The following assumptions have been adopted in this paper in order to analyze the correctness and inclusiveness of diagnoses:

(i) the accurate analytical partial models will be applied for fault detection, and

(ii) the effects of disturbances and measurement noise on the models will be neglected.

Clearly, with these assumptions, the uncertainties of diagnostic signals as well as the uncertainties of decisions regarding conflicts may not be considered. Such idealistic assumptions were made to show that even under such conditions misdiagnoses can occur.

The simulation research of a faulty system will be performed with a nonlinear phenomenological model of a system composed of two interconnected buffer tanks, as described in Section 4. This model reflects the impact of faults on residuals. The model generates continuous residual values as well as their discrete bi- and three-valued representatives (diagnostic signals). The four approaches that will be explored in this article will be based on the following:

- binary signatures of faults (BSR) (Gertler, 1998; Cordier *et al*., 2004),
- consistency-based approach (CBR) (Reiter, 1987),
- trinary signatures of faults (TSR) (Korbicz *et al*., 2004; Kościelny, Bartyś and Grudziak, 2021),
- conflicts and trinary residuals (HIS) (Kościelny and Bartyś, 2023).

The above approaches make use of diagnostic signals and ignore the knowledge of the symptom sequences. However, it should be noted that this knowledge is useful in increasing the distinguishability of faults (Puig *et al*., 2005; Kościelny, Syfert and Wnuk, 2021).

The states of the system being diagnosed are defined by all faults that exist in this state. In addition, we will discuss only the fault-free state and the states with single and double faults. We will also consider all physically possible combinations of the diagnostic signal values in these states. However, for brevity, in this study we will limit our consideration of the impact of the fault compensation effect to double faults only, without losing the generality of conclusions.

The signatures of states with double faults will be derived from the signatures of single faults. Each binary signature will be created as a Boolean alternative of all binary signatures of faults that exist in that state. This approach is commonly adopted in (Gertler, 1998; Kościelny *et al*., 2012; Bartyś, 2013) with full awareness of its unreliability in cases in which a fault compensation effect occurs.

As a result of the trivalent evaluation of a residual, a trivalent diagnostic signal is generated. Thus, the set v of values for each diagnostic signal is a subset of $\{-1, 0, +1\}$. The zero value of the diagnostic signal is interpreted as the insensitivity of the residuum to a fault. The remaining values of diagnostic signals are interpreted as symptoms of faults.

Therefore, the values of diagnostic signals for system states with single faults can take values belonging to the subsets $\{0\}$, $\{-1\}$, $\{+1\}$, or $\{-1, +1\}$.

Table 1. Principles of determining three-valued signatures of double faults.

v_j/v_k	0	-1	$+1$	$-1,+1$
0	0	-1	$+1$	$-1,+1$
-1	-1	-1	$-1,0,+1$	$-1,0,+1$
$+1$	$+1$	$-1,0,+1$	$+1$	$-1,0,+1$
$-1,+1$	$+1,-1$	$-1,0,+1$	$-1,0,+1$	$-1,0,+1$

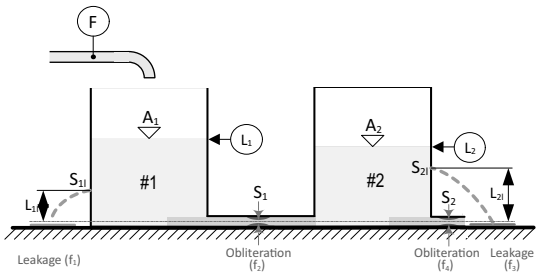

Fig. 1. Diagram of the set of two buffer tanks.

The principles for determining the values of double fault signatures v_{jk} based on the values of diagnostic signals of single faults v_j and v_k are shown in Table 1. These principles boil down to what is called alternative fault signatures. A generalization of the arithmetic of alternative signatures can be found in (Bartyś, 2013). In the case where the diagnostic signals for single faults are mutually contrary, the diagnostic signal for double fault beyond the values of $\{-1\}$ and $\{+1\}$ may take a value of $\{0\}$. This provides an opportunity to consider the effect of fault compensation in the process of diagnostic inference and thus gives a chance to increase fault distinguishability. This aspect will be discussed in more detail in Section 7.

4. Phenomenological model of the set of two buffer tanks

Consider the configuration of two serially interconnected buffer tanks, shown schematically in Fig. 1. Two types of fault are assumed for both tanks: liquid leaks f_1 and f_3 and obliteration of pipelines f_2 and f_4. The fault entries are indicated in Fig. 1. Here, we are dealing with the following:

- holes in tanks with cross-section areas of S_{1l}, S_{2l} located at heights of L_{1l}, L_{2l} relative to the axes of the outlet pipes of both tanks, respectively;
- pipeline obliterations narrowing cross-section pipeline areas S_1, S_2 to S_{1o}, S_{2o}, respectively.

In the following analysis, we assume infallible measurements and ignore leaks in the pipeline connecting

the tanks and leaks in the outlet pipeline. These assumptions allow us to build simple partial models that will be used to make the hypothesis presented in Section 1 plausible.

We assume knowledge of:

- the cross-sectional area A_1 of Tank 1,
- the cross-sectional area A_2 of Tank 2,
- the nominal diameter S_1 of the pipeline connecting both tanks,
- the nominal diameter S_2 of the outlet pipeline,
- the outflow coefficient α_1 from Tank 1,
- the outflow coefficient α_2 from Tank 2,
- the liquid levels L_1 and L_2 in both tanks,
- the flow rate F of liquid entering Tank 1.

4.1. Reference models of the two tank set. Let us define a set of idealized phenomenological reference partial models of the system of two tanks without faults. The fluid mechanics laws of continuous and inviscid flows will be used to describe the flow rate of the liquid. The reference partial models of the volumetric flows for both tanks and for the set of both tanks are

$$A_1\frac{\mathrm{d}L_1}{\mathrm{d}t} - F + \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)} = 0, \quad (1)$$

$$A_2\frac{\mathrm{d}L_2}{\mathrm{d}t} - \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)} + \alpha_2 \cdot S_2\sqrt{2g \cdot L_2} = 0, \quad (2)$$

$$A_1\frac{\mathrm{d}L_1}{\mathrm{d}t} + A_2\frac{\mathrm{d}L_2}{\mathrm{d}t} - F + \alpha_2 \cdot S_2\sqrt{2g \cdot L_2} = 0\,. \quad (3)$$

Equations (1–3) are useful for the qualitative analysis of the impacts of the faults. However, they are not useful for quantitative analysis. Therefore, we will rewrite these equations in the so-called internal form explicitly showing the impact of the faults on residuals.

4.2. Model of two tank system in an internal form. We will begin construction of the model in internal form (Gertler, 1998) with the definition of normalized faults that are easy to interpret and convenient for simulation studies. We will define the faults as follows:

- relative leakage from Tank 1
  $f_1 = \dfrac{S_{1l}}{S_1}$,
- relative obliteration of the inter tank pipeline
  $f_2 = 1 - \dfrac{S_{1o}}{S_1}$,
- relative leakage from Tank 2
  $f_3 = \dfrac{S_{2l}}{S_2}$,
- relative obliteration of the outlet pipeline
  $f_4 = 1 - \dfrac{S_{2o}}{S_2}$.

Now we will perform a transformation of Eqns. (1)–(3) to the set of residual equations

$$\begin{aligned} r_1 = {} & A_1\frac{\mathrm{d}L_1}{\mathrm{d}t} - F + \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)} \\ & + f_1 \cdot \alpha_{1l} \cdot S_1\sqrt{2g(L_1 - L_{1l})} \\ & - f_2 \cdot \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)}, \end{aligned} \quad (4)$$

$$\begin{aligned} r_2 = {} & A_2\frac{\mathrm{d}L_2}{\mathrm{d}t} - \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)} \\ & + \alpha_2 \cdot S_2\sqrt{2g \cdot L_2} \\ & - f_2 \cdot \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)} \\ & + f_3 \cdot \alpha_{2l} \cdot S_2\sqrt{2g(L_2 - L_{2l})} \\ & - f_4 \cdot \alpha_2 \cdot S_2\sqrt{2g \cdot L_2}, \end{aligned} \quad (5)$$

$$\begin{aligned} r_3 = {} & A_1\frac{\mathrm{d}L_1}{\mathrm{d}t} + A_2\frac{\mathrm{d}L_2}{\mathrm{d}t} \\ & - F + \alpha_2 \cdot S_2\sqrt{2g \cdot L_2} \\ & + f_1 \cdot \alpha_{1l} \cdot S_1\sqrt{2g(L_1 - L_{1l})} \\ & + f_3 \cdot \alpha_{2l} \cdot S_2\sqrt{2g(L_2 - L_{2l})} \\ & - f_4 \cdot \alpha_2 \cdot S_2\sqrt{2g \cdot L_2}. \end{aligned} \quad (6)$$

In these equations, it is assumed that the liquid inflow rate F is balanced by the outflow rate, the change in the accumulation of liquid in both tanks and the sum of leaks from both tanks.

4.3. Effect of compensation of fault impact on residuals. The conditions for the compensation of the impact of the fault on the residuals can be obtained directly from (4)–(6) by assigning zero values to the residuals. Each of these equations meets the necessary condition for compensation if the impacts of faults on residuals are opposite. The necessary conditions for the compensation effect for individual residuals are

$$\begin{aligned} r_1 = {} & f_1 \cdot \alpha_{1l} \cdot S_1\sqrt{2g(L_1 - L_{1l})} \\ & - f_2 \cdot \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)} = 0, \end{aligned} \quad (7)$$

$$\begin{aligned} r_2 = {} & -f_2 \cdot \alpha_1 \cdot S_1\sqrt{2g(L_1 - L_2)} \\ & + f_3 \cdot \alpha_{2l} \cdot S_2\sqrt{2g(L_2 - L_{2l})} \\ & - f_4 \cdot \alpha_2 \cdot S_2\sqrt{2g \cdot L_2} = 0, \end{aligned} \quad (8)$$

$$\begin{aligned} r_3 = {} & f_1 \cdot \alpha_{1l} \cdot S_1\sqrt{2g(L_1 - L_{1l})} \\ & + f_3 \cdot \alpha_{2l} \cdot S_2\sqrt{2g(L_2 - L_{2l})} \\ & - f_4 \cdot \alpha_2 \cdot S_2\sqrt{2g \cdot L_2} = 0. \end{aligned} \quad (9)$$

A fault compensation effect is expected when the value of at least one residuum is zero, despite the existence

of fault(s) to which this residuum is sensitive. The fault compensation conditions for double faults apply to

$$\begin{aligned}
&\{f_1 \wedge f_2\} \quad \text{if} \quad r_1 = 0,\\
&\{f_2 \wedge f_3\} \quad \text{if} \quad r_2 = 0 \cap f_4 = 0,\\
&\{f_1 \wedge f_4\} \quad \text{if} \quad r_3 = 0 \cap f_3 = 0,\\
&\{f_3 \wedge f_4\} \quad \text{if} \quad r_2 = 0 \cap f_2 = 0 \cup r_3 = 0 \cap f_1 = 0.
\end{aligned}$$

In addition, Eqns. (7)–(9) show that fault compensation it is not possible for double faults it $\{f_1 \wedge f_3\}$, $\{f_2 \wedge f_3\}$, although is possible for triple faults

$$\begin{aligned}
&\{f_2 \wedge f_3 \wedge f_4\} \quad \text{if} \quad r_2 = 0,\\
&\{f_1 \wedge f_3 \wedge f_4\} \quad \text{if} \quad r_3 = 0,
\end{aligned}$$

and for quadruple fault

$$\{f_1 \wedge f_2 \wedge f_3 \wedge f_4\} \quad \text{if } r_1 = 0 \cap r_2 = 0 \cup r_2 = 0 \cap r_3 = 0.$$

The detailed conditions for the compensation of double faults impacts on residuals for the system of two tanks are depicted in Table 2.

4.4. Simulation model. The simulation model of a set of two tanks was developed in a MATLAB-Simulink environment. The resulting simulation flowchart is shown in Fig. 2. The model reflects the physical structure of the set of buffer tanks shown in Fig. 1. The model consists of two interconnected universal submodels accompanied by a diagnostic system block. For the clarity of the model, the tank parameters and diagnostic inference calculations are hidden under the masks. Each submodel of the tank implements two faults: a leakage from the tank and an obliteration of the outlet pipe. Each submodel provides the possibility of defining the specific behavior of both faults. As a result, the simulation model can be freely extended and used to simulate sets of any number of serially connected and individually parametrised tanks.

The tank model generates an output vector, which includes liquid levels in both tanks, liquid outflow rates, residuals, and diagnostic signals. The change in the level of liquid in the tank is due to the change in the accumulation of liquid. The level of liquid in each tank can therefore be determined by integrating the dynamic accumulation of the liquid, i.e., by integrating the difference in the flow rate of liquid entering and leaving the tank. Let F_{in_i} be a volumetric inflow rate of the liquid, F_{out_i} be the liquid outflow rate, and F_{l_i} be the leakage flow rate from the i-th tank. Therefore, for a continuous and inviscid flow,

$$L_i(t) = \frac{1}{A_i}\int_0^t \left(F_{\text{in}_i} - F_{\text{out}_i} - F_{l_i}\right) \mathrm{d}t. \tag{10}$$

The volumetric flow rate of the liquid F_{out} depends, among others, on its geometry, existing faults, and the level of liquid in the tank in which it flows. In general, the higher the liquid level in the succeeding tank, the smaller the outflow rate F_{out} of the preceding tank. It was therefore necessary to consider this coupling effect when the two submodels of individual tanks were integrated into one model. The liquid levels are determined from the transformed Eqn. (10):

$$\begin{aligned}
L_i(t) = \frac{1}{A_i}\int_0^t \Big(&F_{\text{in}_i} - \alpha_i S_i \sqrt{2g\left(L_i - L_{(i+1)}\right)}\\
&+ f_1 \alpha_{li} S_i \sqrt{2g\left(L_i - L_{li}\right)}\\
&+ f_2 \alpha_i S_i \sqrt{2g\left(L_i - L_{(i+1)}\right)}\Big)\,\mathrm{d}t.
\end{aligned} \tag{11}$$

In addition, the following bonds (couplings) were imposed on the set of two tanks:

$$\begin{cases}
F_{\text{out}_i} = F_{\text{in}_{(i+1)}},\\
L_{\text{in}_i} = L_{\text{in}_{(i+1)}},\\
L_{\text{in}_{(i+2)}} = 0.
\end{cases} \tag{12}$$

A simple residual assessment approach with an arbitrarily selected nonnegative threshold δ was applied for the determination of diagnostic signals. The applied principles for binary and trinary assessments are respectively, given by

$$\begin{cases}
s = 0 \leftrightarrow |r| < \delta,\\
s = 1 \leftrightarrow |r| \geq \delta,
\end{cases} \tag{13}$$

$$\begin{cases}
s = +1 \leftrightarrow r > \delta,\\
s = \ \ 0 \leftrightarrow |r| \leq \delta,\\
s = -1 \leftrightarrow r < -\delta.
\end{cases} \tag{14}$$

Table 2. Conditions for compensation of double fault impacts on residuals.

residual	faults	condition
$r_1 = 0$	$\{f_1 \wedge f_2\}$	$\frac{f_1}{f_2} = \frac{\alpha_1}{\alpha_{1l}}\sqrt{\frac{L_1 - L_2}{L_1 - L_{1l}}}$
$r_2 = 0$	$\{f_2 \wedge f_4\}$	$\frac{f_4}{f_2} = \frac{\alpha_1}{\alpha_2}\frac{S_1}{S_2}\sqrt{\frac{L_1 - L_2}{L_2}}$
$r_2 = 0$	$\{f_3 \wedge f_4\}$	$\frac{f_4}{f_3} = \frac{\alpha_{2l}}{\alpha_2}\sqrt{\frac{L_2 - L_{2l}}{L_2}}$
$r_3 = 0$	$\{f_3 \wedge f_4\}$	$\frac{f_4}{f_3} = \frac{\alpha_{1l}}{\alpha_2}\frac{S_1}{S_2}\sqrt{\frac{L_1 - L_{1l}}{L_2}}$
$r_3 = 0$	$\{f_1 \wedge f_4\}$	$\frac{f_4}{f_1} = \frac{\alpha_{2l}}{\alpha_2}\sqrt{\frac{L_2 - L_{2l}}{L_2}}$

5. Characteristics of fault inference approaches

The diagnostic properties of the set of two tanks presented in Fig. 1 will be discussed based on diagnoses generated by four selected fault inference approaches. This section briefly describes the approaches we are referring to.

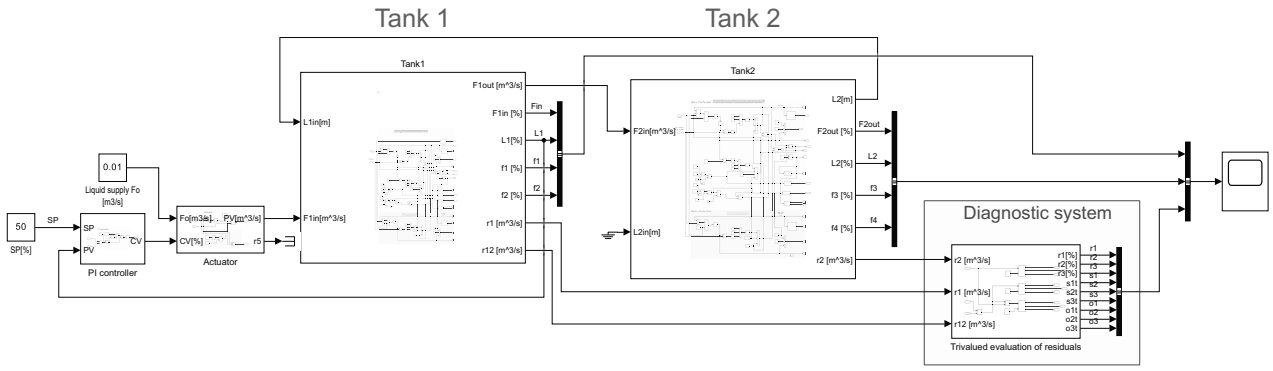

Fig. 2. Simulation diagram of two serially interconnected buffer tanks.

5.1. Diagnosing with binary signatures (BSR). This approach is specific for the FDI community (Gertler, 1998; Cordier *et al.*, 2004). The fault isolation is mainly based on a binary FSM. The columns of the FSM represent the signatures of the system states with faults. Signatures of multiple faults are Boolean alternatives of signatures of single faults that are present in this state. The assumption of the absence of fault compensation effect is adopted for multiple faults.

Reasoning about faults is based on knowledge of fault symptoms (diagnostic signal values equal 1) and the lack of symptoms (zero values of diagnostic signals). Diagnostic signals having zero values reject from diagnosis those states for which the corresponding value of fault-specific diagnostic signal reference values in fault signatures are equal to 1. This complies with the ARR-based exoneration assumption (Cordier *et al.*, 2004).

Table 3 shows the binary signatures of the state of the two-tank system described in Section 4. This table contains signatures of the fault-free state and states with single and double faults.

The reasoning regarding the system state is represented by the following set of rules:

$$\text{if } (s_1 = v_{1i}) \wedge \cdots \wedge (s_J = v_{Ji}) \text{ then } z_i(\Phi_i), \quad (15)$$

where Φ_i is a set of faults in a given state z_i, and $v_{ji} \in \{0, 1\}$ is the binary value of diagnostic signal s_j of the signature of i-th state. The diagnosis is as follows:

$$\Delta_{\text{BSR}} = \left\{ z_i \in Z : \bigvee_{j=1}^{J} (s_{ji} = v_{ji}) \right\}. \quad (16)$$

Table 3. Binary signatures of states for the two tank system.

Φ	$\emptyset$	f_1	f_2	f_3	f_4	f_1 f_2	f_1 f_3	f_1 f_4	f_2 f_3	f_2 f_4	f_3 f_4
Z	z_0	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
s_1	0	1	1	0	0	1	1	1	1	1	0
s_2	0	0	1	1	1	1	1	1	1	1	1
s_3	0	1	0	1	1	1	1	1	1	1	1

5.2. Consistency-based diagnostic approach (CBR). Consistency-based approaches (Reiter, 1987; de Kleer and Williams, 1987; de Kleer and Kurien, 2003; de Kleer *et al.*, 1992) belong to diagnostic reasoning methods based on the theory of diagnosis of the first principles (Reiter, 1987). The essence of these approaches consists of searching for the minimal sets of components or fault types inconsistent with the observations of the diagnosed system.

The CBR introduces the concept of a conflict set. The conflict set contains those system components that explain the inconsistency of the system description and observations.

If we identify or represent components of the system by their faults, then the j-th conflict set C_j can be interpreted as being equivalent to a subset of faults to which the j-th diagnostic signal is sensitive, i.e., $F(s_j = 1)$. We have

$$C_j = F(s_j = 1) = \{ f_k : v_{kj} = 1 \}. \quad (17)$$

Therefore, $\forall j \in \{1, \ldots, J\}$, the rules

$$\text{if } (s_j = 1) \text{ then } F(s_j = 1) \quad (18)$$

define conflict sets, each of which contains at least one fault. Diagnoses are generated as the minimal hitting sets of all minimal conflict sets that have been observed. The hitting set for the conflict sets

$$CS = \{ C_j : s_j = 1 \wedge s_j \in S \}. \quad (19)$$

is $HS_i \subseteq \bigcup C_j \in CS$ such, that the intersection of this set with each conflict set is not empty,

$$\bigwedge_{ij} HS_i \cap C_j \neq \emptyset. \quad (20)$$

The hitting set HS_i is minimal if and only if none of its proper subsets is the hitting set for the set of currently existing conflicts. Thus, the minimal hitting set contains the lowest cardinality subsets of faults that explain all conflicts. This is a potential diagnosis. Further, we will consider only minimal hitting sets. However, all supersets of minimal hitting sets are potential diagnoses as well.

Diagnosis Δ_{CBR} is a set of all potential diagnoses

$$\Delta_{\text{CBR}} = \{HS_i\}. \tag{21}$$

Only conflicts that have appeared are considered in consistency-based approaches. Information that conflicts do not appear is not used in making the diagnosis, unlike in the fault signature-based approaches (BSR and TSR). The conflict sets derived from Table 3 for the two-tank system are as follows:

$$\begin{array}{rl} \{f_1, f_2\} & \text{corresponding to } s_1 = 1, \\ \{f_2, f_3, f_4\} & \text{corresponding to } s_2 = 1, \\ \{f_1, f_3, f_4\} & \text{corresponding to } s_3 = 1. \end{array}$$

5.3. Diagnosing with trivalent signatures (TSR). The approach to diagnosing with trivalent signatures was exhaustively presented in (Korbicz *et al.*, 2004; Kościelny, Bartyś and Grudziak, 2021). In this study, the trivalent signatures of single faults will be derived from (4)–(6) or (7)–(9). They are depicted in the first four columns in Table 4. Based on these signatures and according to the truth table (Table 1), we can specify the signatures of states with double faults.

The diagnostic reasoning is performed according to the following set of rules

$$\text{if } (s_1 = V_{1i}) \wedge \cdots \wedge (s_j = V_{Ji}) \text{ then } z_i(\Phi_i), \tag{22}$$

where V_{ji} is the subset of the diagnostic signal values s_j in state $z_i(\Phi_i)$, and Φ_i is the subset of faults in state z_i.

The diagnosis indicates all states z_i for which the current values of the diagnostic signals match their signatures,

$$\Delta_{\text{TSR}} = \left\{ z_i \in Z : \bigvee_j s_j \in V_{ji} \right\}, \quad s_j \in S. \tag{23}$$

5.4. Hybrid approach (HIS) based on conflicts and trinary valued signatures. The hybrid approach to diagnosing (HIS) has been presented by Kościelny and Bartyś (2023). The strength of this approach is in the synergy gained from integration of the diagnosing based on the fault-symptoms representation in the form of FIS combined with an edge-and node-labelled HS-tree based inference (Reiter, 1987). The inference of faults is carried out based on conflicts and the use of the trinary-valued diagnostic signals. The use of the trinary diagnostic signals yields a twofold increase in the number of conflict sets compared with the CBR. This is because the signs of the trivalent residuals can be taken into account. However, the powers of these sets are not higher than the powers of the conflict sets in the case of binary diagnostic signals. Conflicts are indicated by -1 or $+1$ values of diagnostic signals. Analogously to (18), we can propose the set of the following rules:

$$\begin{cases} \text{if } (s_j = -1) \text{ then } F(s_j = -1), \\ \text{if } (s_j = +1) \text{ then } F(s_j = +1). \end{cases} \tag{24}$$

The diagnosis is generated in the same way as for binary conflicts, i.e., by determining the minimal hitting sets for the CS conflicts (19). The diagnosis takes the form of a set of potential diagnoses (21).

Using the directional impacts of faults on residuals (residual signs), in the scope of consistency-based approaches, is qualified as exploiting fault models (Struss and Dressier, 1989). However, in the case of HIS, these are qualitative models that do not require quantification of the impact of faults on the residuals as in (4)–(6). Usually, it is sufficient to acquire expert knowledge regarding the impact of faults on residual signs or obtain the residual equations.

In the example under consideration, instead of three sets of conflicts generated by the CBR, we obtain as many as six such sets by the HIS approach:

$$\begin{array}{rr} \{f_2\} \text{ if } s_1 = -1, & \{f_1\} \text{ if } s_1 = +1, \\ \{f_4\} \text{ if } s_2 = -1, & \{f_2, f_3\} \text{ if } s_2 = +1, \\ \{f_4\} \text{ if } s_3 = -1, & \{f_1, f_3\} \text{ if } s_3 = +1. \end{array}$$

Table 4. Trivalent signatures of diagnostic states for the two-tank system. Here $v = \{-1, 0, +1\}$.

Φ	$\emptyset$	f_1	f_2	f_3	f_4	f_1 f_2	f_1 f_3	f_1 f_4	f_2 f_3	f_2 f_4	f_3 f_4
Z	z_0	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
s_1	0	$+1$	-1	0	0	v	$+1$	$+1$	-1	-1	0
s_2	0	0	$+1$	$+1$	-1	$+1$	$+1$	-1	$+1$	v	v
s_3	0	$+1$	0	$+1$	-1	$+1$	$+1$	v	$+1$	-1	v

6. Simulations

This section demonstrates the chosen results of simulation tests which confirm the hypothesis formulated in Section 1.

Example 2. (*Single fault and potential diagnoses of physically impossible states generated by the consistency-based approach*) The objective of this example is to demonstrate that the CBR approach may indicate potential diagnoses of physically impossible states due to the fault compensation effect.

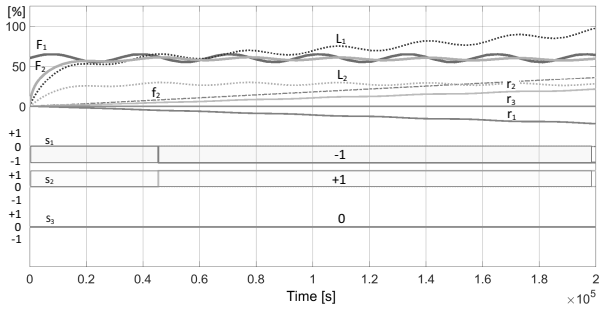

Fig. 3. Example of a simulation of a single fault f_2. Notation: F_1 – liquid inflow rate into tank No. 1; F_2 – outflow rate from tank No. 2; L_1 –liquid level in tank No. 1; L_2 – liquid level in tank No. 2; f_2 – the obliteration of the pipeline connecting both tanks; residuals: r_1, r_2, r_3; diagnostic signals: s_1, s_2, s_3.

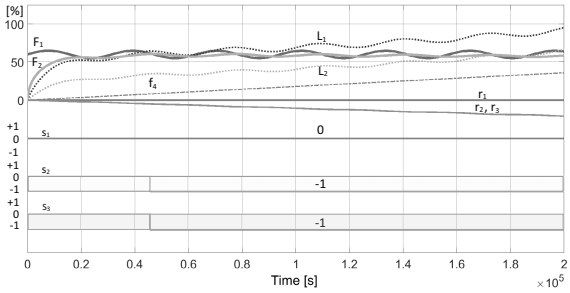

Fig. 4. Example of a simulation of a single fault f_4. The legend as in Fig. 3.

Consider the single fault condition z_2 of the diagnosed two-tank system. Figure 3 provides the result of a simulation of a single idealized incipient fault f_2 representing a relative reduction in the cross-section of the pipeline connecting both tanks. This case corresponds to Diagnosis 3 in Table 5.

In the time interval from 0 to $0.44 \cdot 10^5$ s, all diagnostic signals take zero values. In the time interval from $0.44 \cdot 10^5$ to $2.00 \cdot 10^5$ s, the diagnostic signal takes the value from $(s_1 = -1, s_2 = +1, s_3 = 0)$. Diagnoses based on the approaches BSR, TSR and HIS are accurate and indicate fault f_2.

In turn, the CBR diagnosis indicates potential diagnoses $z_2 = \{f_2\}$, $z_6 = \{f_1 \wedge f_3\}$ and $z_7 = \{f_1 \wedge f_4\}$, of which only the first is physically possible and the last two are physically impossible.

Potential diagnoses $\{f_1 \wedge f_3\}$, according to (9), are physically impossible because both faults exhibit a unidirectional impact on the residual value. Therefore, in this case, it is also not possible to mutually compensate for the impacts of both faults. However, fault compensation is theoretically possible for the state $z_7 = \{f_1 \wedge f_4\}$. The occurrence of this double fault should lead to the generation of diagnostic signal values $(s_1 = +1, s_2 = -1)$, i.e., different from those observed. ♦

Example 3. (*Single fault and potential diagnoses of physically impossible states generated by the BSR and CBR.*) The objective of this example is to demonstrate that in the case of a single fault, the BSR and CBR approaches may produce diagnoses of physically impossible states.

Consider the single fault state $z_4 = \{f_4\}$. This case corresponds to Diagnoses No. 5 in Table 5. Figure 4 provides an example of a simulation of a single incipient fault f_4 representing a reduced cross-section of the outflow pipe of the second tank. For comparison, the slope of f_4 is identical to that of fault f_2 in Example 2.

In the time interval from 0 to $0.44 \cdot 10^5$s, all diagnostic signals take zero values. In the time interval from $0.44 \cdot 10^5$ to $2.00 \cdot 10^5$ s, the values of the diagnostic signal are $(s_1 = 0, s_2 = -1, s_3 = -1)$. The diagnosis generated by the approach BSR indicates states $z_3 = \{f_3\}$, $z_4 = \{f_4\}$, and $z_{10} = \{f_3 \wedge f_4\}$. The potential diagnoses generated by the approach CBR are $z_3 = \{f_3\}$, $z_4 = \{f_4\}$ and $z_5 = \{f_1 \wedge f_2\}$. The TSR indicates potential diagnoses $z_4 = \{f_4\}$ and $z_{10} = \{f_3 \wedge f_4\}$. In turn, the diagnosis formulated by the HIS precisely indicates the real state z_4. All the above diagnoses are correct because all indicate fault f_4.

On the other hand, the state z_3 is physically impossible. This state was incorrectly indicated by BSR and CBR.

In addition, the state z_5 with the double fault in the diagnosis CBR is also not physically possible. According to (7), fault compensation can take place. The occurrence of this state should lead to the generation of diagnostic signal values $(s_2 = +1, s_3 = +1)$, that is, different from those actually observed.

The state z_{10} indicated by the BSR and TSR approaches is physically possible. However, it is not a minimal hitting set and therefore does not occur in the diagnosis CBR. It should be emphasized that all potential diagnoses generated in this example by the TSR and HIS approaches indicate exclusively physically possible states.

This example confirms that the approaches CBR and BSR can generate diagnoses of physically impossible states in the case of single faults. ♦

Example 4. (*Double fault and incorrect diagnosis generated by the BSR and potential diagnoses of physically impossible states generated by the CBR and BSR.*) The objective of this example is to demonstrate that in the case of a double fault, the BSR may fail and the BSR and CBR approaches can generate potential diagnoses of physically impossible states.

Consider now double fault state $z_9 = \{f_2 \wedge f_4\}$ in

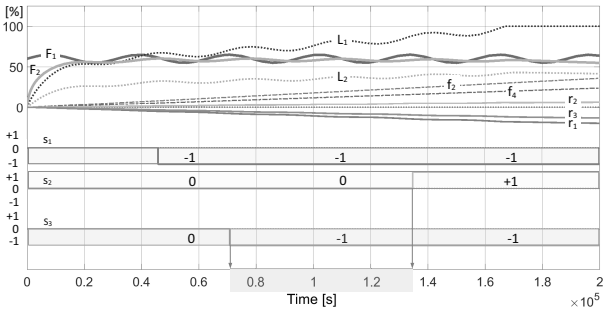

Fig. 5. Example of a simulation of a state with a double fault $z_9 = \{f_2 \wedge f_4\}$. The legend as in Fig. 3. The area marked in grey indicates the time interval for the compensation of impacts of faults f_2 and f_4 on the residual r_2.

a case of compensation of impacts of faults on residual r_2. Figure 5 depicts the result of a simulation of incipient faults f_2 and f_4 representing respectively obliteration of the pipeline connecting the tanks and the outlet pipeline from the second tank. This case corresponds to Diagnoses 14 and 15 in Table 5.

In the time interval from 0 to $0.44 \cdot 10^5$ s, all diagnostic signals take zero values. In the time interval from 0.44 to $10^5 \ldots 0.70 \cdot 10^5$ s, the diagnostic signal values $(s_1 = -1, s_2 = 0, s_3 = 0)$ are pointing to an unknown state. In the time interval from 0.70 to $10^5 \ldots 1.36 \cdot 10^5$ s, the diagnostic signal values are $(s_1 = -1, s_2 = 0, s_3 = -1)$. In this time interval, the effect of compensation of impacts of both simulated faults on residuum r_2 takes place. The diagnosis generated by the BSR exhibits only one potential diagnosis. It indicates state $z_1 = \{f_1\}$, and therefore it is a potential diagnosis of a physically impossible state. Ultimately, the diagnosis is incorrect and non-inclusive.

In turn, the diagnosis generated by the CBR indicates potential diagnosis of a possible physical state $z_9 = \{f_2 \wedge f_4\}$, and two potential diagnoses of the physically impossible states $z_1 = \{f_1\}$ and $z_8 = \{f_2 \wedge f_3\}$. On the other hand, the diagnoses generated by the TSR and HIS are correct.

In the time interval from 1.36 to $10^5 \ldots 2.00 \cdot 10^5$ s, the diagnostic signal values are $(s_1 = -1, s_2 = +1, s_3 = -1)$. The diagnoses obtained by BSR and CBR correctly indicate the state z_9 with double faults $\{f_2 \wedge f_4\}$ and additionally four potential diagnoses of physically impossible states with double faults. It should be noted that the diagnoses generated by the TSR and HIS precisely indicate the physical states of the system.

Example 4 indicates that diagnosing with the BSR and CBR may lead to the generation of potential diagnoses of physically impossible states. ♦

In summarizing this section, it should be emphasized that in all the mentioned examples, the CBR diagnosis is always consistent with the observations. In turn, based on the analysis of the simulation results, the BSR and CBR fault isolation approaches do not guarantee the indication of physically possible states, neither for single nor double faults.

7. Sources of diagnoses of physically impossible states

Principally, there are two main reasons for generating potential diagnoses of physically impossible states in the case of inferring with binary diagnostic signals:

- resulting from a loss of information by the transformation of the continuous residuals into binary diagnostic signals, and
- related to a fault compensation effect.

The binary evaluation of residuals relies on the assigning of a binary number to a diagnostic signal value according to the result of the comparison of the absolute value of residual with a certain nonnegative threshold. Hence, the evaluation result is independent of whether the residual is positive or negative. Therefore, it must be determined whether the loss of information with respect to the sign of the residual affects the result of the diagnostic inference.

In fact, the fault can manifest itself by a permanently unidirectional or bidirectional deviation of the residual value from zero. For example, obliteration of the pipeline causes a unidirectional change in the residual, whereas parametric faults of instruments can cause both an increase and a decrease in the residual value.

Therefore, the binary evaluation of residuals is informatively lossy in the sense that identical binary signatures can be attributed to different faults despite the physical constraints. It is worth noticing that this may not be the case with three-valued signatures.

Thus, a diagnosis based on binary signatures may indicate states that are physically impossible. From these considerations, it we deduce the following:

Conclusion 1. *Binary evaluation of residuals may lead to the generation of potential diagnoses of physically impossible states.*

The fault compensation effect applies exclusively to multiple faults. This effect is usually ignored unjustly in approaches that use binary signatures to isolate multiple faults. Clearly, such approaches are structurally not resistant to fault compensation effects. This problem can be solved by an appropriate adjustment of the fault signature matrices. It is possible, for example, to apply the alternative signature approach proposed in (Bartyś, 2013;

Bartyś, 2014; Bartyś, 2021), which is immune to the fault compensation effect.

FDI approaches ignored the fault compensation effects when developing signatures for states with multiple faults. In turn, the CBR approaches assume the possibility of compensation impacts of faults on the residuals (Cordier *et al*., 2004).

A necessary but insufficient condition for the fault compensation effect is the sensitivity of the residual to at least two faults. In fact, the compensation effect can occur only if the impact of faults on the residual value is opposite as shown in Example 4.

Therefore, the property of complete immunity to the effect of fault compensation is not confirmed in the case of their unidirectional impact on the residuals. It is a case where potential diagnoses of physically impossible states may be generated.

Conclusion 2. *The complete immunity to the effect of fault compensation attributed to the CBR approach is not justified.*

8. Comparative study

The comparative study of the characteristics of diagnostic approaches that are in the area of interest of this paper will be carried out on the example of the diagnosis of a two-buffer tank system set shown in Fig. 1. The four diagnostic approaches, i.e., BSR, CBR, TSR, and HIS, introduced in Section 5 will be considered. The models (1)–(3) will be used for diagnosis.

The study will be carried out according to the methodology presented in Section 3. The fault-free and all physically possible states of the diagnosed system with single and double faults were examined for all combinations of diagnostic signal values that are physically possible in these states. Table 5 provides a breakdown of the diagnoses generated.

Incorrect diagnoses are shaded grey in Tab. 5, while all potential diagnoses of impossible physical system states are highlighted in bold. The zero values of diagnostic signals resulting from the fault compensation effect are in bold and are in bold. According to the note in Tab. 4, it was assumed that a fault diagnosis is generated only if at least one diagnostic signal value deviates from zero. Therefore, Diagnoses 1 and 17 in Table 5 indicate a fault-free state.

The summary of the incorrect diagnoses obtained is shown in Table 6. The number of diagnoses in this table corresponds to the number of diagnoses in Table 5. Table 6 lists the following:

- incorrect diagnoses,
- incorrect but inclusive diagnoses,
- potential diagnoses of the impossible physical states,

Table 5. List of diagnoses of the two-tank system.

Z	s_1, s_2, s_3 trinary	s_1, s_2, s_3 binary	BSR	CBR	TSR	HIS	No
z_0	$0, 0, 0$	$0, 0, 0$	z_0	z_0	z_0	z_0	1
z_1	$+1, 0, +1$	$1, 0, 1$	z_1	z_1 $\mathbf{z_8}$ $\mathbf{z_9}$	z_1	z_1	2
z_2	$-1, +1, 0$	$1, 1, 0$	z_2	z_2 $\mathbf{z_6}$ $\mathbf{z_7}$	z_2	z_2	3
z_3	$0, +1, +1$	$0, 1, 1$	z_3 $\mathbf{z_4}$ z_{10}	z_3 $\mathbf{z_4}$ z_5	z_3 z_5 z_{10}	z_3 z_5	4
z_4	$0, -1, -1$	$0, 1, 1$	z_4 $\mathbf{z_3}$ z_{10}	z_4 $\mathbf{z_3}$ $\mathbf{z_5}$	z_4 z_{10}	z_4	5
z_5	$\mathbf{0}, +1, +1$	$\mathbf{0}, 1, 1$	z_3 $\mathbf{z_4}$ z_{10}	z_5 z_3 $\mathbf{z_4}$	z_5 z_3 z_{10}	z_5 z_3	6
	$+1, +1+1$	$1, 1, 1$	z_5 z_6 $\mathbf{z_7}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_5 z_6 $\mathbf{z_7}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_5 z_6	z_5 z_6	7
	$-1, +1, +1$	$1, 1, 1$	z_5 $\mathbf{z_6}$ $\mathbf{z_7}$ z_8 $\mathbf{z_9}$	z_5 $\mathbf{z_6}$ $\mathbf{z_7}$ z_8 $\mathbf{z_9}$	z_5 z_8	z_5 z_8	8
z_6	$+1, +1, +1$	$1, 1, 1$	z_6 z_5 $\mathbf{z_7}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_6 z_5 $\mathbf{z_7}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_6 z_5	z_6 z_5	9
z_7	$+1, -1, \mathbf{0}$	$1, 1, \mathbf{0}$	$\mathbf{z_2}$	z_7 $\mathbf{z_2}$ $\mathbf{z_6}$	z_7	z_7	10
	$+1, -1, +1$	$1, 1, 1$	z_7 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_7 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_7	z_7	11
	$+1, -1, -1$	$1, 1, 1$	z_7 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_7 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_8}$ $\mathbf{z_9}$	z_7	z_7	12
z_8	$-1, +1, +1$	$1, 1, 1$	z_8 z_5 $\mathbf{z_6}$ $\mathbf{z_7}$ $\mathbf{z_9}$	z_8 z_5 $\mathbf{z_6}$ $\mathbf{z_7}$ $\mathbf{z_9}$	z_8 z_5	z_8 z_5	13
z_9	$-1, \mathbf{0}, -1$	$1, \mathbf{0}, 1$	$\mathbf{z_1}$	z_9 $\mathbf{z_1}$ $\mathbf{z_8}$	z_9	z_9	14
	$-1, +1, -1$	$1, 1, 1$	z_9 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_7}$ $\mathbf{z_8}$	z_9 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_7}$ $\mathbf{z_8}$	z_9	z_9	15
	$-1, -1, -1$	$1, 1, 1$	z_9 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_7}$ $\mathbf{z_8}$	z_9 $\mathbf{z_5}$ $\mathbf{z_6}$ $\mathbf{z_7}$ $\mathbf{z_8}$	z_9	z_9	16
z_{10}	$0, \mathbf{0}, \mathbf{0}$	$0, 0, \mathbf{0}$	z_0	z_0	z_{10} z_0	z_0	17
	$0, +1, +1$	$0, 1, 1$	z_{10} z_3 z_4	z_3 z_4 z_5	z_{10} z_3 z_5	z_3 z_5	18
	$0, -1, -1$	$0, 1, 1$	z_{10} z_3 z_4	z_3 z_4 z_5	z_{10} z_4	z_4	19

for all diagnostic fault isolation approaches being studied.

In Table 5, for the state z_{10}, that is, $\{f_3 \wedge f_4\}$, only three combinations of diagnostic signal values that may actually occur are specified. Other combinations are physically impossible due to the unidirectional impact of the faults f_3 and f_4 on the residuals r_2 and r_3, as can be seen in (5)–(6).

Table 5 allows for the determination of the values of the diagnosis quality indices defined in Appendix A. The calculated values are summarized in Table 7.

As can be seen in Table 7, the index Θ expressing the average share of potential diagnoses of impossible physical states in the total number of diagnoses is different from zero if diagnosing with BSR. This also applies to the CBR approach which is based on binary observations. By contrast, potential diagnoses of impossible physical states are not recorded for the approaches TSR and HIS.

It was also shown that the CBR approach may generate potential logically correct diagnoses of physically impossible states in the case of the fault compensation effect. Table 6 also shows that the application of the TSR and HIS approaches based on trivalent diagnostic signals does not show potential diagnoses of physically impossible states. The index of incorrect diagnoses Ψ expresses the share of diagnoses that do not indicate the real physical state. From Table 5, it follows that the BSR and CBR approaches generate logically correct diagnoses, however of physically impossible states as shown in Table 6. In all these cases, an effect of compensation for the impact of faults on residuals was observed.

The percentage of incorrect diagnoses generated in fault compensation cases (index χ) exhibits the complete lack of robustness in this range demonstrated by the approach BSR. The other approaches proved to be robust in some way in this aspect. In the case study under review, only the TSR approach ensures the full veracity of diagnoses.

The measure of fault distinguishability is the index of theoretical accuracy of diagnosis D. Any diagnosis of a physically impossible state results in a reduction in the value of this index. This is the case of BSR and CBR. This also explains the values of the indicator D in Table 7 for binary approaches compared to their trinary counterparts TSR and HIS. This is in line with the results of previous studies (Bregón *et al.*, 2013; Kościelny *et al.*, 2016; Kościelny *et al.*, 2019).

Surprisingly, however, is the higher value of index D obtained by the HIS in comparison with the TSR approach. It was rather expected that the introduction of the exoneration assumption in the TSR would increase, not decrease, the fault distinguishability.

The explanation is as follows. If the potential diagnoses are minimal hitting sets and if in the real world there is a subset of faults that is a superset of the minimal hitting set, then none of the potential diagnoses will indicate the actual state. This is illustrated in Example 5.

Example 5. Suppose simultaneous faults a and b and two conflict sets $\{a, b, c\}$ and $\{a, b, d\}$. The minimal hitting sets that constitute potential diagnoses are $\{a\}$, $\{b\}$, $\{c, d\}$. The set $\{a, b\}$ is not a minimal hitting set as it is a superset of the minimal hitting sets. ♦

The set of faults $\{f_3, f_4\}$ of the state z_{10} is a superset of the sets $\{f_3\}, \{f_4\}$, and the empty set of the state z_0. Therefore, it is not in the diagnoses of CBR and HIS.

Example 6. *Consequences of a false and non-inclusive diagnosis.* Suppose that Fig. 1 shows a simplified diagram of a configuration of aviation fuel tanks. Assume that and emergency condition occurred with two faults: a leakage from Tank 1 and the clogging of the outlet pipe from Tank 2. Diagnoses are generated by the binary signature approach BS. As a result of the opposite effect of these faults on the value of residual 3 in (6), a false and non-inclusive diagnosis is obtained. This diagnosis indicates a clogging of the channel between tanks (Table 6, Diagnosis 10). The suggested safety precaution is to cut off the fuel supply when the upper alarm limit is exceeded in Tank 1. In fact, there is a leak that threatens to ignite and cause a fire with the consequences similar to the accident in Buncefield, England, in December 2005. ♦

In general, it can be stated that in BSR and TSR all combinations of faults limited only by assumed multiplicity are considered. Only some of these combinations are used in consistency-based methods, despite the unlimited multiplicity of faults they assume.

On the one hand, the difference in the inference method leads to obtaining more precise diagnoses in the case of HIS compared to with TSR (without an indication of the state z_{10}). On the other hand, incorrect diagnoses (17)–(19) are generated for the HIS approach when the state is z_{10}. These reflect increased values of the Ψ and Φ indices for HIS.

Non-zero values of share χ of incorrect diagnoses due to the fault compensation effect indicate that the BSR approach is not robust to this effect. The robustness of the CBR and HIS approaches is definitely better. The TSR approach exhibited total robustness in this case study.

Based on the performed discussion, the following conclusions can be drawn:

(a) diagnoses of physically impossible states may result in incorrect and non-inclusive diagnoses by BSR and CBR approaches;

(b) by CBR and HIS approaches, the incorrect diagnoses may occur only if in reality, there exist faults that are not in the minimal hitting set;

Table 6. Summary of diagnoses obtained for the two-tank system.

Diagnosis	BSR	CBR	TSR	HIS
incorrect	6, 10, 14, 17	17, 18, 19	– –	17, 18, 19
incorrect and inclusive	–	18, 19	–	–
diagnoses of physically impossible states	4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19	–	–

Table 7. Diagnosis quality indices.

Index	BSR	CBR	TSR	HIS
Θ	0.487	0.540	**0.000**	0.000
Ψ	0.211	0.158	**0.000**	0.158
χ	1.00	0.250	**0.000**	0.250
D	0.487	0.347	0.684	**0.816**
Φ	0.211	0.053	**0.000**	0.158

(c) in the studied case, the usage of trivalent diagnostic signals rejects the potential diagnoses of impossible physical states;

(d) the fault distinguishability obtained by inferring based on rows need not be less than that obtained by signature-based diagnosing;

(e) HIS approach with trivalent diagnostic signals does not guarantee exclusively correct diagnoses.

The research carried out does not cover all the aspects that should be taken into account when evaluating diagnostic approaches. For example, they do not capture the impact of false diagnostic signals on diagnoses. However, this is beyond the scope of this paper.

9. Summary

The paper discusses the consistency of the diagnoses obtained using selected model-based diagnostic approaches with the real state of the system being diagnosed.

The contribution of this work is a new look and revision of views on the problem of the generation of diagnoses that are formally correct but not true, i.e., inconsistent with the physically existing state.

The results reported in this paper contribute to the growth of knowledge regarding the identification of the root causes of the generation of misdiagnoses. They may be useful in assessing the robustness of diagnostic approaches against the generation of incorrect diagnoses.

As a result of the study, it was found that approaches based on binary diagnostic signals or observations generate diagnoses of potential physically impossible states, despite the absence of modeling errors, disturbances, and measurement noise. These are the result of binary residual evaluation and/or the effect of compensation of fault impacts on residual values.

In the example studied, it was discovered that diagnosis based on trivalent residuals (TSR and HIS) does not indicate any physically impossible condition and, consequently, does not lead to the generation of misdiagnoses.

The reduction in the number of physically impossible potential diagnoses achieved by applying trivalent residuals affects not only the indices characterizing the share of incorrect diagnoses, but also the increased distinguishability of faults. High distinguishability is, of course, critical for maintaining the safety of the diagnosed processes.

Based on the results obtained in this work, the following working hypotheses can be formulated:

(i) Diagnostic inference based on fault signatures and trivalent residual evaluation rejects incorrect diagnoses.

(ii) Diagnostic approaches based on rows and the HS tree demonstrate higher fault distinguishability compared with diagnosis based on fault signatures and trivalent residual evaluation.

However, both hypotheses require further theoretical and experimental studies.

Acknowledgment

This work was supported by the Institute of Automatic Control and Robotics of the Warsaw University of Technology.

References

Armengol, J., Bregón, A., Escobet, T., Gelso, E., Krysander, M., Nyberg, M., Olive, X., Pulido, B. and Travè-Massuyès, L. (2009). Minimal structurally overdetermined sets for residual generation: A comparison of alternative approaches, *IFAC Proceedings Volumes* **42**(8): 1480–1485.

Bartyś, M. (2013). Generalised reasoning about faults based on diagnostic matrix, *International Journal of Applied Mathematics and Computer Science* **23**(2): 407–417.

Bartyś, M. (2014). *Selected Issues of Fault Isolation*, Polish Scientific Publishers, Warsaw.

Bartyś, M. (2021). Fault compensation effect in fault detection and isolation, *Acta IMEKO* **10**(3): 45–53.

Biswas, G., Kapadia, R. and Yu, X. (1997). Combined qualitative-quantitative steady-state diagnosis of continuous-valued systems, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* **27**(2): 167–185.

Blanke, M., Kinnaert, M., Lunze, J. and Staroswiecki, M. (2015). *Diagnosis and Fault-Tolerant Control*, Springer, New York.

Bregón, A., Alonso-González, C.J. and Pulido, B. (2014). Integration of simulation and state observers for online fault detection of nonlinear continuous systems, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **44**(12): 1553–1568.

Bregón, A., Biswas, G., Pulido, B., Alonso-Gonzalez, C. and Khorasgani, H. (2013). A common framework for compilation techniques applied to diagnosis of linear dynamic systems, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **44**(7): 863–876.

Chen, J. and Patton, R. (1999). *Robust model Based Fault Diagnosis for Dynamic Systems*, Kluwer Akademic Publishers, Boston.

Cordier, M., Dague, P., Lévy, F., Montmain, J., Staroswiecki, M. and Travé-Massuyés, L. (2004). Conflicts versus analytical redundancy relations: A comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives, *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics* **34**(5): 2163–2177.

Daigle, M., Koutsoukos, X. and Biswas, G. (2009). A qualitative event-based approach to continuous systems diagnosis, *IEEE Transactions on Control Systems Technology* **17**(4): 780–793.

de Kleer, J. and Kurien, J. (2003). Fundamentals of model-based diagnosis, *IFAC Proceedings Volumes* **36**(5): 25–36.

de Kleer, J., Mackworth, A.K. and Reiter, R. (1992). Characterizing diagnoses and systems, *Artificial Intelligence* **56**(2): 197–222.

de Kleer, J. and Williams, B. (1987). Diagnosing multiple faults, *Artificial Intelligence* **32**(1): 97–130.

Düstegör, D., Frisk, E., Cocquempot, V., Krysander, M. and Staroswiecki, M. (2006). Structural analysis of fault isolability in the damadics benchmark, *Control Engineering Practice* **14**(6): 597–608.

Eskandari, A., Nedaei, A., Milimonfared, J. and Aghaei, M. (2024). A multilayer integrative approach for diagnosis, classification and severity detection of electrical faults in photovoltaic arrays, *Expert Systems with Applications* **252**(Part A): 124111.

Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy, *Automatica* **26**(3): 459–474.

Gertler, J. (1991). Analitical redunduncy methods in fault detection and isolation, *IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS'91*, Baden-Baden, pp. 9–21.

Gertler, J. (1998). *Fault Detection and Diagnosis in Engineering Systems*, Marcel Dekker, New York.

Jia, F., Cao, F., Lyu, G. and He, X. (2023). A novel framework of cooperative design: Bringing active fault diagnosis into fault-tolerant control, *IEEE Transactions on Cybernetics* **53**(5): 3301–3310.

Korbicz, J., Kościelny, J.M., Kowalczuk, Z. and Cholewa, W. (Eds) (2004). *Fault Diagnosis. Models, Artificial Intelligence, Applications*, Springer, Berlin.

Kościelny, J.M. (1995). Fault isolation in industrial processes by dynamic table of states method, *Automatica* **31**(5): 747–753.

Kościelny, J.M. (1999). Application of fuzzy logic fault isolation in a three-tank system, *IFAC Proceedings Volumens* **32**(2): 7754–7759.

Kościelny, J.M. and Bartyś, M. (2023). A new method of diagnostic row reasoning based on trivalent residuals, *Expert Systems with Applications* **214**: 119116.

Kościelny, J. M., Bartyś, M. and Grudziak, Z. (2021). Tri–valued evaluation of residuals as a method of addressing the problem of fault compensation effect, *in* J. Korbicz, K. Patan and M. Luzar (Eds), *Advances in Diagnostics of Processes and Systems*, Springer, Cham, pp. 31–44.

Kościelny, J.M., Bartyś, M. and Rostek, K. (2019). The comparison of fault distinguishability approaches – Case study, *Bulletin of the Polish Academy of Sciences Technical Sciences* **67**(6): 1059–1068.

Kościelny, J. M., Bartyś, M., Rzepiejewski, P. and da Costa, J. S. (2006). Actuator fault distinguishability study of the damadics benchmark problem, *Control Engineering Practice* **14**(6): 645–652.

Kościelny, J.M., Bartyś, M. and Syfert, M. (2012). Methods of multiple fault isolation in large scale systems, *IEEE Transactions On Control Systems Technology* **20**(5): 1302–1310.

Kościelny, J.M., Syfert, M., Rostek, K. and Sztyber, A. (2016). Fault isolability with different forms of faults-symptoms relation, *International Journal of Applied Mathematics and Computer Science* **26**(4): 815–826.

Kościelny, J.M., Syfert, M. and Wnuk, P. (2021). Diagnostic row reasoning method based on multiple-valued evaluation of residuals and elementary symptoms sequence, *Energies* **14**(2476).

Krysander, M., Aslund, J. and Nyberg, M. (2007). An efficient algorithm for finding minimal overconstrained subsystems for model-based diagnosis, *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* **38**(1): 197–206.

Kunpeng, Z., Bin, J., Fuyang, C. and Hui, Y. (2023). Directed-graph-learning-based diagnosis of multiple faults for high speed train with switched dynamics, *IEEE Transactions on Cybernetics* **53**(3): 1712–1724.

Liu, J., Wang, X., Wu, S., Wan, L. and Xie, F. (2023). Wind turbine fault detection based on deep residual networks, *Expert Systems with Applications* **213**: 119102.

Pawlak, Z. (1991). *Rough Sets. Theoretical Aspects of Reasoning About Data*, Kluwer Academic Publishers, Boston.

Puig, V., Schmid, F., Quevedo, J. and Pulido, B. (2005). A new fault diagnosis algorithm that improves the integration of fault detection and isolation, *44th IEEE Conference on Decision and Control, Seville, Spain*, pp. 3809–3814.

Pulido, B. and González, C. (2004). Possible conflicts: a compilation technique for consistency-based diagnosis, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **34**(5): 2192–2206.

Reiter, R.A. (1987). Theory of diagnosis from first principles, *Artificial Intelligence* **32**(1): 57–95.

Song, Q. and Jiang, P. (2022). A multi-scale convolutional neural network based fault diagnosis model for complex chemical processes, *Process Safety and Environmental Protection* **159**: 575–584.

Struss, P. and Dressier, O. (1992). "Physical negation": Integrating fault models into the general diagnostic system, *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Vol.2, pp. 1318–1323.

Su, J. and Chen, W. (2019). Model-Based Fault Diagnosis System Verification Using Reachability Analysis, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **49**(4): 742–751.

Tatara, M.S. and Kowalczuk, Z. (2024). Approximate and analytic flow models for leak detection and identification, *International Journal of Applied Mathematics and Computer Science* **34**(3): 391–407.

Travè-Massuyès, L. (2014). Bridges between diagnosis theories from control and AI perspectives, *in* J. Korbicz and M. Kowal (Eds), *Intelligent Systems in Technical and Medical Diagnostics,* Berlin/Heidelberg, pp. 3–28.

Xia, D. and Fu, X. (2024). Observer-based sliding-mode fault-tolerant consistent control for hybrid event-triggered multi-agent systems, *International Journal of Applied Mathematics and Computer Science* **34**(3): 361–373.

Zheng, S. and Zhao, J. (2022). High-fidelity positive-unlabeled deep learning for semi-supervised fault detection of chemical processes, *Process Safety and Environmental Protection* **165**: 191–204.

Jan M. Kościelny has been with the Institute of Automatic Control and Robotics at the Warsaw University of Technology since 1973. He is a professor and leader of the Research Working Group on Diagnostics of Industrial Processes. His research activities are focused mainly on the fields of fault detection and isolation, fault-tolerant control, and decentralized systems. He is the author or a co-author of 7 monographs, over 300 papers, and 3 patents. He is a member of the Committee of Automatics and Robotics of the Polish Academy of Sciences and Technical Committee TC 6.4. of IFAC.

Michał Z. Bartyś holds ME, MSc, PhD, and DSc degrees and has been with the Institute of Automatic Control and Robotics, Warsaw University of Technology since 1973. His research activities have focused mainly on process control, fault detection and isolation, functional safety, fault-tolerant systems, fieldbus network systems, fuzzy logic applications, and intelligent final control elements. He has authored and contributed to 17 books, 3 handbooks, 23 textbooks, 205 papers, 7 patents, and has designed 107 mechatronic devices.

Appendix

Metrics of diagnostic quality

In this appendix, several indices for assessing the quality of diagnosis are defined for our comparative studies.

1. The mean share of potential diagnoses of the impossible physical states

$$\Theta = \frac{1}{N}\sum_{i=1}^{N}\frac{n_i^p}{n_i}, \tag{A1}$$

where n_i is the number of potential diagnoses in the i-th diagnosis, n_i^p is the number of potential diagnoses of physically impossible states in the i-th diagnosis, N is the total number of all combinations of diagnostic signal values in all the considered states.

$$N = \sum_{i:z_i \in Z} n_i. \tag{A2}$$

2. Index of the theoretical accuracy of a diagnosis

The theoretical accuracy of a single diagnosis is defined as the reciprocal of the number of d_i states indicated in the diagnosis. The index of theoretical precision of a diagnosis D is defined as the mean value of the diagnostic precision for all diagnoses N,

$$D = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{d_i}. \tag{A3}$$

We define the theoretical accuracy of a diagnosis, because we do not consider here modelling errors, disturbances, measurement noise, uncertainties, etc.

3. The share of incorrect diagnoses

The share of incorrect diagnoses Ψ is the ratio of the number of incorrect diagnoses n_ψ to the number of all possible diagnoses N,

$$\Psi = \frac{n_\psi}{N}. \tag{A4}$$

Received: 28 September 2024
Revised: 9 December 2024
Re-revised: 22 January 2025
Accepted: 22 January 2025