Preprints, 9th IFAC Symposium on Fault Detection, Supervision and
Safety of Technical Processes
September 2-4, 2015. Arts et Métiers ParisTech, Paris, France

Analysis of fault isolation assumptions when comparing model-based design approaches of diagnosis systems

Daniel Jung * **Hamed Khorasgani** ** **Erik Frisk** *
Mattias Krysander * **Gautam Biswas** **

* *Dept. of Electrical Engineering, Linköping University, Sweden, e-mail:* {daner,frisk,matkr}@isy.liu.se
** *Inst. of Software-integrated Systems, Vanderbilt Univ., USA, e-mail:* {hamed.g.khorasgani,gautam.biswas}@vanderbilt.edu

Abstract: Most model-based diagnosis approaches reported in the literature adopt a generic architecture and approach. However, the fault hypotheses generated by these methods may differ. This is not only due to the methods, but also on the basic assumptions made by different diagnostic algorithms on fault manifestation and evolution. While comparing different diagnosis approaches, the assumptions made in each case will have a significant effect on fault diagnosability performance and must therefore also be taken into consideration. Thus, to make a fair comparison, the different approaches should be designed based on the same assumptions. This paper studies the relation between a set of commonly made assumptions and fault isolability performance in order to compare different diagnosis approaches. As a case study, five developed diagnosis systems for a wind turbine benchmark problem are evaluated to analyze the type of assumptions that are applied in the different designs.

Keywords: Model-based diagnosis, fault detection and isolation, fault diagnosability analysis.

1. INTRODUCTION

In model-based diagnosis, a mathematical model of the system to be supervised is developed. Then, to detect faults, residuals are designed based on analytical redundancy in the model. Residuals are then commonly designed using, for example, analytical redundancy relations (ARR) (Staroswiecki and Comtet-Varga, 2001; Cordier et al., 2004), possible conflicts (PC) (Pulido and González, 2004; Bregon et al., 2013), or observer-based methods (Frank, 1996; Isermann, 1997). Then a fault isolation algorithm, using some decision logic, computes one or several fault hypothesis based on the residual outputs (or features). A fault hypothesis is a set of faults that can explain the observed residual outputs.

In many applications, full knowledge about the fault type and the nature of fault manifestation, for example, possible faults and fault profiles (i.e., the temporal characteristic of the fault) are not available. Most of the time, only data from the nominal system behavior is available in practice as the model, describing the faults, is constructed mathematically alongside the equations for the system under supervision. Therefore, the conclusions made by the diagnosis system are based on some set of assumptions made about the faults. As a result, it is difficult to make a fair comparison amongst different diagnosis approaches as various design choices can be utilized by each method on the assumptions about the faults.

⋆ The work is partially supported by the Swedish Research Council within the Linnaeus Center CADICS.

Some common assumptions made in different diagnosis approaches are, for example, single-fault assumption and exoneration. The exoneration assumption means that a fault will always trigger *all* residuals sensitive to that fault, i.e. there will never be only a subset of residuals sensitive to the fault that will trigger. The same type of fault assumption can be applied in different diagnosis algorithms to different extends. The type of assumptions that are applied will not only have a significant impact on the fault isolability performance and robustness of the diagnosis system, but also the risk of fault misclassifications. This has previously been highlighted in Cordier et al. (2004) when discussing the bridge between model-based diagnosis approaches from the FDI and DX communities. In contrast to the previous work which focus on assumptions made in specific diagnosis approaches from the two communities, this paper analyzes the relations between assumptions and the measurement as well as the residual output spaces using a more general framework.

Previous works, such as Gertler (1991); de Kleer and Kurien (2003); Cordier et al. (2004, 2006); Llobet et al. (2009); Bregon et al. (2013) have discussed frameworks that bridge the approaches developed by the different research communities. The main focus of these papers have been on studying the similarities and differences between methods developed by the FDI (system diagnosis) and the DX (AI approach to diagnosis) communities. In contrast to previous works, this paper analyzes the relation between some common assumptions which is important when comparing different diagnosis solutions.

One main contribution is the analysis of why different diagnosis approaches can generate different fault hypotheses from the same system. In order to analyze the different approaches, diagnosability properties are considered given the spaces of possible measurements and residual outputs (features). The focus is not to analyze the properties of different residual-design methods but instead the effects of assumptions made. As a case study, five diagnosis systems developed to monitor a wind turbine benchmark model, which participated in a diagnosis competition (Odgaard and Stoustrup, 2012), are analyzed.

Benchmark problems have been used in academic competitions to compare the performance of different implemented approaches (Bartyś et al., 2006; Kurtoglu et al., 2009; Odgaard et al., 2009). Benchmark problems produce results that are useful in practical applications. However, since the comparison is based purely on performance metrics, the results are biased because different solutions are based on different assumptions about the system dynamics and fault manifestations. These papers make no attempt to explain the underlying causes or differences between the different diagnosis methods.

The outline of this paper is as follows. First, the problem formulation is presented in Section 2. Definitions of diagnosability properties are presented in Section 3 and an analysis of different assumptions is made in Section 4. Then, the case study is presented in Section 5 and the results from the analysis are discussed in Section 6. Finally, some conclusions and future work are presented in Section 7.

2. PROBLEM FORMULATION

The goal is to analyze how the assumptions affect the fault isolability performance of a diagnosis system, such as: *which other faults a fault can be isolated from*Here, the assumptions considered are design parameters used during the development of the diagnosis system to simplify the fault isolation problem.

To limit the analysis, different assumptions made in the residual design, such as noise distributions and uncertainties, are not considered here. The effects of applying assumptions that are not valid for a given problem are also not examined. Also, the effects in dynamic systems, such as fault propagation causing delays of different residuals before triggering, are not considered in the analysis.

A short description of the assumptions considered here is as follows. Note that this list is in no way exhaustive but covers a set of the most common assumptions made in different model-based diagnosis approaches.

Closed world assumption The closed world assumption means that the diagnosis system has full knowledge of all possible faults that could occur in the system.

Single-fault assumption In many practical cases multiple faults occur rarely. Therefore, it is commonly assumed that maximally one fault can be present in the system at any given time.

Exoneration The assumption that a fault always triggers all the residuals that are sensitive to the fault.

Limitation of possible fault realisations In many applications, the number of possible fault magnitudes and manifestations are limited, which then limits the set of possible measurement values that can be made for each fault mode. For example, a fault representing an increase in mechanical friction can not be negative. There are also other common assumptions made, such as faults occur either abruptly, are slowly varying, or always have the same magnitudes as a given set of training data.

3. BASIC DEFINITIONS ON FAULT DETECTABILITY AND ISOLABILITY

Here, fault detectability and isolability are defined as properties of the set of possible measurements from the system. Then, these definitions are extended to properties of residual outputs. The goal is to have a set of definitions, which are independent of diagnosis approach to describe the effects of assumptions made. First, the general design of diagnosis systems considered here is described.

3.1 Diagnosis system

The type of model-based diagnosis approaches considered here are consistency-based where the diagnosis system structure can be represented by Fig. 1. The monitored system can be affected by a combination of possible faults $\{f_1, f_2, \ldots, f_{n_f}\}$. A *fault mode* $F_i \subseteq \{f_1, f_2, \ldots, f_{n_f}\}$ represents a specific set of faults that is present in the system, which can be both single faults and multiple faults. The fault mode representing the nominal fault free case is explicitly denoted as NF (No Fault). Note that all faults might not be known by the diagnosis system. The figure shows how a fault that occurs in the system will result in different possible measurements represented by the different ellipses. There can be several fault modes that that can cause the same measurements which are represented by the overlapping ellipses. The measurements $z = (y, u)$ from the system are a combination of available sensors y and known actuators u. The measurements from the fault-free systems are represented by the dark ellipse. Note that fault analysis such as, what specific type of fault realizations that are causing the measurements, for example fault magnitudes and trajectories, is not considered here, only which measurements that can be explained by different types of fault modes.

Then, residuals $\mathcal{T} = \{T_1(z), T_2(z), \ldots, T_p(z)\}$ are used to map the measurements to some residual outputs, or features. Note that the number of residuals can be one, several, or dynamic over time. The features also have different possible values that can be explained by the system being in different fault modes which are also represented by ellipses corresponding to the measurement sets. Different residuals maps different measurements to different feature sets. The features are used in different hypothesis tests to determine if a fault has occurred or not.

Model uncertainties, measurement noise, and process noise complicate distinguishing faults from nominal system behavior. Therefore, one or more hypothesis tests based on the residual outputs are used to determine if a fault has occurred. A common type of hypothesis test is the use of

a threshold, where the threshold can be established using more or less sophisticated methods, such as maximum likelihood estimators or CUSUM tests (Basseville and Nikiforov, 1993). A fault isolation algorithm using some decision logic computes one or several fault hypotheses based on the residual outputs. There are many different methods for fault isolation, see for example De Kleer and Williams (1987); Cordier et al. (2004); Mosterman and Biswas (1999).

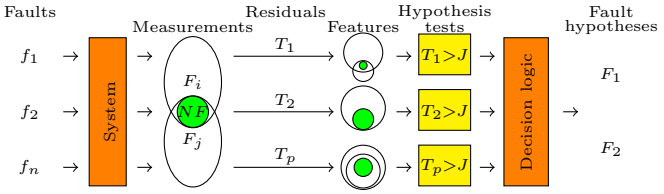

Fig. 1. The diagnosis systems considered here computes diagnosis candidates based on a set of residuals.

3.2 Fault diagnosability given a system

Consider a system $\mathcal{S}$. Let $\Omega_\mathcal{S}$ denote the multi-dimensional observation space of all possible measurements z of $\mathcal{S}$, i.e. $z = (y, u) \in \Omega_\mathcal{S}$. This notation for representing observations consistent with different fault modes is similar to the observation sets used in for example Nyberg and Frisk (2006).

Let $\mathcal{F}_{\text{all}} = \{NF, F_1, F_2, \dots F_q\}$ be the set of all possible fault modes the system $\mathcal{S}$ can be in. However, since all fault modes are not always known, the subset of known fault modes when developing the diagnosis system is denoted $\mathcal{F} \subseteq \mathcal{F}_{\text{all}}$. Given the closed world assumption, $\mathcal{F} = \mathcal{F}_{\text{all}}$.

The fault mode the system is in will affect the measurements made, i.e. different fault modes can generate different measurement values. For each $F_i \in \mathcal{F}$, let $\Phi_\mathcal{S}(F_i) \subseteq \Omega_\mathcal{S}$ represent the subset of measurement values consistent with (the system being in) fault mode F_i.

Example 1. Consider a system of two sensors, y_1 and y_2, measuring the same real-valued quantity, i.e. $y_1, y_2 \in \mathbb{R}$. The fault-free case corresponds to all measurements where the two sensors have the same output, i.e. $y_1 = y_2$. However, a fault in any of the sensors can result in different outputs from the two sensors, i.e., $y_1 \neq y_2$. However, a fault in a sensor is not always visible, for example if the sensor have glitches. Another case is when the two sensors have the same fault, resulting in the same bias in both sensors. In both cases, the two sensors can have the same outputs even though at least one of them is faulty, i.e., $y_1 = y_2$ can also be explained by a faulty sensors. Let F_1 denote the fault mode when y_1 is faulty, F_2 when y_2 is faulty, F_3 when both sensors are faulty, and NF the fault-free case. Then, $\Omega_\mathcal{S} = \{\forall y_1 \in \mathbb{R}, \forall y_2 \in \mathbb{R}\}$, and the measurement sets corresponding to the different fault modes can be defined as

$$\Phi_\mathcal{S}(NF) = \{\forall y_1 \in \mathbb{R}, \forall y_2 \in \mathbb{R} : y_1 = y_2\}, \text{ and}$$
$$\Phi_\mathcal{S}(F_1) = \Phi_\mathcal{S}(F_2) = \Phi_\mathcal{S}(F_3) = \Omega_\mathcal{S}.$$

□

Given the closed world assumption, it is assumed that each $z \in \Omega_\mathcal{S}$ can be explained by at least one fault mode F_i, i.e.

$$\bigcup_{\forall F_i \in \mathcal{F}} \Phi_\mathcal{S}(F_i) = \Omega_\mathcal{S}. \tag{1}$$

Based on the measurement subsets $\Phi_\mathcal{S}(F_i)$ for each fault mode $F_i \in \mathcal{F}$, a fundamental criteria for fault detectability and isolability is defined.

Definition 2. A fault mode F_i is isolable from F_j if

$$\Phi_\mathcal{S}(F_i) \not\subseteq \Phi_\mathcal{S}(F_j). \tag{2}$$

A fault mode F_i is said to be detectable if it is isolable from NF. □

Thus, in order for a fault mode F_i to be isolable from another fault mode F_j there must exist a measurement z that can be explained by F_i but not F_j.

In Trave-Massuyes et al. (2006), a definition related to isolability in the observation space is used, called *discriminability*. Discriminability is defined for a pair of faults and three levels are considered: strongly, weakly, and non-discriminable fault pairs. Weakly and non-discriminable fault pairs are covered by the isolability definition in Definition 2. However, strong discriminability is more restrictive than the definition of isolability and is defined as follows.

Definition 3. Two fault modes, F_i and F_j, are *strongly discriminable* from each other if

$$\Phi_\mathcal{S}(F_i) \cap \Phi_\mathcal{S}(F_j) = \emptyset. \tag{3}$$

□

The definition of strong discriminability says that two strongly discriminable faults will never generate the same measurements from the system. Strong discriminability is symmetric, in contrast to isolability, as shown in Fig. 2, and is closely related to the exoneration assumption that will be discussed later.

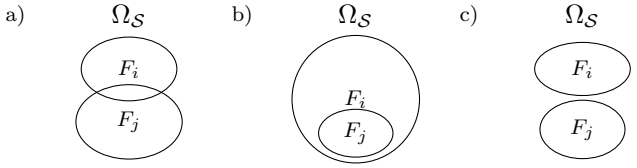

Fig. 2. In case a) F_i and F_j are isolable from each other, in case b) F_i is isolable from F_j but not vice versa, and in case c) F_i and F_j are strongly discriminable from each other.

Often the knowledge about $\Omega_\mathcal{S}$ and all subsets $\Phi_\mathcal{S}(F_i)$ is limited. This includes information about which faults that could occur and how they manifest in the system and evolves over time. The lack of knowledge is often due to lack of available data from different fault modes since faults occur rarely. Also, collecting data that describes all possible measurements is not feasible. It is often difficult to use the measurements without post-processing for detecting faults. Therefore, residuals are designed where the outputs are easier to interpret than the measurements whether there are faults present in the system or not.

3.3 Fault diagnosability properties of diagnosis tests

Each diagnosis test (residual) $T_k : \Omega_\mathcal{S} \to \Omega_{T_k}$ in $\mathcal{T}$ maps the observation set $\Omega_\mathcal{S}$ to the feature set Ω_{T_k} which is the

set of all possible values of T_k. For each fault mode F_i, let $\Phi_{T_k}(F_i)$ denote the projection of the subset $\Phi_{\mathcal{S}}(F_i)$ using T_k, i.e., $T_k : \Phi_{\mathcal{S}}(F_i) \rightarrow \Phi_{T_k}(F_i)$. Fault isolability for a given diagnosis test T_k is then defined as follows.

Definition 4. A fault mode F_i is isolable from F_j with a test T_k if

$$\Phi_{T_k}(F_i) \not\subseteq \Phi_{T_k}(F_j). \tag{4}$$

A fault mode F_i is said to be detectable with a test T_k if it is isolable from NF. □

If a fault mode F_i is detectable with a test T_k, it is said that T_k *is sensitive to* F_i. A graphical interpretation of fault isolability with a test T_k is shown in Fig. 3. The figure represents that if a fault mode F_i is isolable from another fault mode F_j then there are measurements that can be explained by the system being in fault mode F_i but not F_j. Then, if a residual can isolate a fault mode F_i from another fault mode F_j then there are residual outputs that can be explained by the system being in fault mode F_i but not F_j.

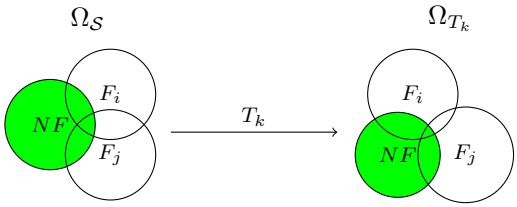

Fig. 3. Two fault modes F_i and F_j are isolable from each other with a test T_k if the observations in $\Omega_{\mathcal{S}}$ that can be explained by each fault mode, are mapped to different subsets of Ω_{T_k}.

In model-based diagnosis, a common approach to perform fault isolation is to design a set of diagnosis tests which are sensitive to different sets of fault modes, such as structured residuals (Gertler and Singer, 1990; Staroswiecki and Comtet-Varga, 2001). If T_k is not sensitive to a fault mode F_j it is said that F_j is decoupled and is here defined as follows.

Definition 5. (Fault mode decoupling). A fault mode F_j is said to be decoupled from T_k if

$$\Phi_{T_k}(F_j) = \Phi_{T_k}(NF). \tag{5}$$

□

A graphical interpretation of decoupling faults is shown in Fig. 4 where the measurements of fault mode F_j, decoupled from T_k, is projected to a subset of the projected features from the fault-free mode. Commonly in residual-based approaches, it is assumed that $\Phi_{T_k}(F_i) = \Phi_{T_k}(F_j)$ if F_i and F_j are non-decoupled fault modes of T_k, and sometimes also $\Phi_{T_k}(NF) \subseteq \Phi_{T_k}(F_i)$ for all fault modes $F_i \in \mathcal{F}$. In this case, fault decoupling is necessary to isolate fault modes from each other and to identify the true fault mode. This case is discussed more in the next section.

4. EFFECTS OF ASSUMPTIONS ON FAULT DIAGNOSABILITY PROPERTIES

Here, the definitions from the previous sections are used to describe how the different measurement sets and residual output sets are related to different common assumptions about faults.

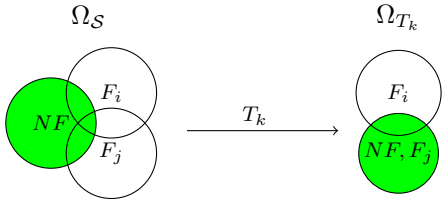

Fig. 4. The fault mode F_j is decoupled from T_k, $\Phi_{T_k}(F_j) = \Phi_{T_k}(NF)$.

4.1 No assumptions

First, consider the case where no assumptions are made about faults (except the closed world assumption). One or several of the possible faults could occur at the same time and there are no limitations assumed about fault magnitudes or manifestations. This means that any residual output can be explained by all the fault modes the residual is sensitive to. Thus, conclusions used for fault isolation are only made when a residual deviates from the fault-free case and decoupled fault modes.

Given the no assumptions case where all fault modes F_i, each residual T_k is sensitive to, are consistent with any value of T_k, i.e. $\Phi_{T_k}(F_i) = \Omega_{T_k}$, then for each fault mode $\bar{F}_i$ there exists a relation such that when $F_i \subseteq \bar{F}_i$, then

$$\Phi_{\mathcal{S}}(F_i) \subseteq \Phi_{\mathcal{S}}(\bar{F}_i). \tag{6}$$

This indicates that should there be no assumptions made about the residual output spaces of non-decoupled fault modes. Also, if T_k is sensitive to F_i, then T_k is sensitive to all fault modes $\bar{F}_i$ representing a superset of faults. For example, if $\{f_1\}$ is a fault hypothesis, so is also $\{f_1, f_2\}$. This means also that any set of faults is a fault hypothesis during the fault-free case since faults might not always be visible in the measurements or several faults might be canceling out each other. The no assumption case is found in the fault isolation algorithms discussed in De Kleer and Williams (1987) where only residuals deviating from the nominal case are used for fault isolation.

4.2 Assumptions about possible fault modes

One common assumption is that no other unknown fault mode can occur beside the defined set of known fault modes, which is called the *closed world assumption*.

Note that if it is assumed that any combination of the possible faults in $\mathcal{F}_{\text{all}}$ can be present in the system at the same time, then the number of fault modes is 2^{n_f}, i.e. the number of fault modes grows exponentially with the number of faults. In order to reduce the complexity of the fault isolation procedure, the number of fault modes can be reduced, for example by only consider a subset of faults $\mathcal{F} \subset \mathcal{F}_{\text{all}}$ and use the closed world assumption. One assumption is that only certain types of faults can occur, such as, sensor faults or actuator faults.

Another common assumption is that only one fault can be present at any given time, i.e. each fault mode represents only one fault, called the *single fault assumption*. Then, the number of fault modes is equal to the number of possible faults in the system (plus the fault-free case), i.e. $\mathcal{F} = \{NF, F_1, F_2, \ldots, F_{n_f}\}$ where $F_i = \{f_i\}, \forall i = 1, 2, \ldots, n_f$. An example where the set of possible fault modes is limited

is when using a bank of Kalman filters to model each possible fault mode, see Isermann (1997). Then, the bank of Kalman filters is used to identify which Kalman filter which do not deviate from the nominal behavior when one of the fault modes occurs.

4.3 Exoneration

In some cases, it is assumed that a fault will *always* cause a diagnosis test to deviate from the fault-free case which is called *exoneration* (Cordier et al., 2004). In dynamic systems the fault propagation can cause delays between when different diagnosis tests will trigger to the fault. Therefore, in these cases, the exoneration assumption can be considered such that the diagnosis tests should trigger within a given time interval. However, for the analysis in this work, only the static case is considered. Exoneration is considered both in the single-fault case and multiple-fault case and is defined as follows.

Definition 6. (Exoneration). Exoneration means that for each non-decoupled fault mode F_i of a test T_k,

$$\Phi_{T_k}(F_i) \bigcap \Phi_{T_k}(NF) = \emptyset. \tag{7}$$

□

Equation (7) shows that if the value of a residual $r \in \Phi_{T_k}(NF)$, then the fault mode F_i can not explain the residual outputs. This means that the exoneration assumption allows conclusions to be drawn from residuals which have outputs that have not deviated from nominal behavior, which was not the case in the no assumption case in Section 4.1.

Note that since exoneration assumes that a fault will always trigger the residual outputs to deviate from $\Phi_{T_k}(NF)$, then the following can be stated about $\Omega_{\mathcal{S}}$.

Theorem 7. Assume that the exoneration assumption is valid for a set of diagnosis tests $\mathcal{T}$. If there exists a diagnosis test $T_k \in \mathcal{T}$, sensitive to F_i but not F_j, then

$$\Phi_{\mathcal{S}}(F_i) \bigcap \Phi_{\mathcal{S}}(F_j) = \emptyset, \tag{8}$$

i.e. the fault modes F_i and F_j are strongly discriminable from each other. □

Proof. Theorem 7 is proved by contradiction. If there exists a diagnosis test T_k that is sensitive to F_i where F_j is decoupled, then

$$\Phi_{T_k}(F_i) \bigcap \Phi_{T_k}(F_j) = \emptyset. \tag{9}$$

Assume that there exists an element z in both $\Phi_{\mathcal{S}}(F_i)$ and $\Phi_{\mathcal{S}}(F_j)$. Since $T_k : \Phi_{\mathcal{S}}(F_i) \rightarrow \Phi_{T_k}(F_i)$ and $T_k : \Phi_{\mathcal{S}}(F_j) \rightarrow \Phi_{T_k}(F_j)$, then the projection of z should lie in both $\Phi_{T_k}(F_i)$ and $\Phi_{T_k}(F_j)$, which is a contradiction. ■

Theorem 7 shows the close relation between the exoneration assumption and strong discriminability of fault modes. A graphical interpretation of exoneration is shown in Fig. 5 where fault mode F_i fulfills the exoneration assumption since no measurement values given fault mode F_i overlaps with measurement values from the other fault modes. That is, there are no measurements or residual outputs when the system is in fault mode F_i that can be explained by any other fault mode.

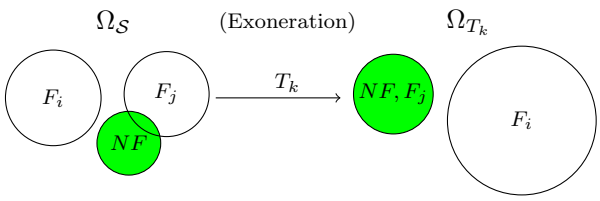

Fig. 5. If F_j is decoupled from T_k, then exoneration means that for fault mode F_i, which is not decoupled, it can be expressed that $\Phi_{T_k}(F_i) \bigcap \Phi_{T_k}(NF) = \emptyset$.

In Cordier et al. (2000), a slightly different definition of exoneration is proposed where only single fault is considered. To assure that two or more faults do not "cancel out" each other, another assumption is also considered together with exoneration, called *no compensation*. This assumption states that if there are two or more faults present in the system, the effects of the faults can not cancel each other out in any of the diagnosis tests. This situation is included in the exoneration assumption in Definition 6 since single fault and multiple faults are just considered as different fault modes.

4.4 Fault magnitudes and manifestation

Another approach to improve isolability performance is to specify possible fault magnitudes and manifestations that a fault could have, e.g., only steps or ramps (Frank, 1990). In many situations, faults can only affect a system in a certain way. For example, leakages often induce a mass flow following a pressure gradient, increased friction in a joint implies an increase of the friction parameter and not a decrease, etc. This type of knowledge or assumption about faults is useful to isolate faults when some faults can not be decoupled. This type of assumptions will limit the measurement sets of different fault modes such that the measurement sets or residual output sets are not overlapping as much as they would otherwise.

Methods designed using training data from *both* fault-free and faulty cases are also considered here since it is often assumed that the training data covers all necessary cases from different fault realizations to perform detection and isolation. Thus, the measurement set or residual output set is determined by the training data. However, it is not considered here that any assumption about fault magnitudes or manifestations is applied if only fault-free data is used in the diagnosis system design. This is because model development is usually made using data from the fault-free system and the assumption would include all model-based diagnosis methods.

5. CASE STUDY: WIND TURBINE BENCHMARK COMPETITION

This analysis is performed based on an analysis of the papers from the five participants in the wind turbine benchmark competition (Odgaard and Stoustrup, 2012). The benchmark describes a wind turbine model and a number of faults that can occur in the system is found in Odgaard et al. (2009). Also, a simulation model to generate data and a set of fault scenarios is provided covering eight single fault scenarios and one double fault scenario. A short

evaluation of the performances of the different solutions is presented in Odgaard and Stoustrup (2012).

The analysis here is made based on the presentations of each solution in each corresponding paper. Each diagnosis system solution is evaluated on, whether they apply any of the assumptions discussed in Section 4 or not, denoted closed world (CW), single fault (SF), exoneration (EX), and fault manifestation (FM) respectively. The analysis focuses on the type of conclusions made when residuals have or have not been triggered. Several solutions apply combinations of different design methods and some assumptions are only partially made for a subset of tests or fault modes. The results presented here are based on the authors analysis of the solutions presented in each paper.

5.1 Analyzed diagnosis systems

The same notation is used here as in Odgaard and Stoustrup (2012) when referring to the different solutions GKSV for Laouti et al. (2011), EB for Zhang et al. (2011), UCB for Ozdemir et al. (2011), COK for Chen et al. (2011), and GFM for Svärd and Nyberg (2011). A short summary of the analysis of the different solutions is presented. All methods are designed based on the CW assumption where all possible faults are given by the benchmark.

GKSV The diagnosis tests are designed using support vector machines (SVM) to classify if a fault is present. For training of each test both fault-free data and faulty data are used as training data using the FM assumption. Most tests are designed to detect only one fault. In some cases where one test is sensitive to two faults, a second test is activated to identify which fault is present. In these cases, the SF assumption is applied since the isolation test is expected to identify only one present fault.

EB The SF assumption is explicitly assumed where a set of observers, modeling each fault mode, are used to isolate the present fault. Different subsets of faults are isolated by comparing the estimated parameters of the different observers. In some cases, the decision logic identifies the fault by identifying which parameter estimation errors that are close to zero and which has significantly deviate from zero. Each fault corresponds to one combination of estimation errors, i.e. one fault signature, used to classify the present fault which implies EX.

UCB A combination of model-based and hardware redundancy-based residuals are used to detect faults. One residual to detect a drivetrain system fault is developed using a data-driven approach. However, the design of the residual is only based on fault-free data and thus the FM assumption is not applied. Fault isolation is performed automatically for some residuals since they are only sensitive to one fault. For some faults, a fault symptom table is used to identify the present fault, i.e. SF is applied. However, it is not clear when reading the paper if the decision logic for the fault signature matrix draws conclusions from non-triggered residuals, i.e. if it applies EX or not.

COK This solution is based on a set of Kalman filters and observers with different fault sensitivities. Fault detection is performed for a moving time window of data and the fault isolation logic uses a column matching approach to identify the present fault given a set of residual outputs. Faults in different subsystems are isolated independently from each other which implies both EX and SF assumptions for each subsystem, but not generally for the whole system.

GFM A set of automatically designed model-based residuals with different fault sensitivities are used. No knowledge about faults is assumed to be known and faults are detected by comparing residual distributions to distributions from fault-free training data. Thus, FM is not applied since no faulty data is used. Only triggered residuals are considered in the the fault isolation algorithm, which computes all fault hypotheses of minimal cardinality and considers multiple-faults if no single fault can explain the triggered residuals, i.e no EX or SF assumptions are applied.

5.2 Comparing assumptions in different solutions

The five diagnosis system solutions are using different combinations of the considered assumptions. Also, the assumptions are applied differently in the different solutions. However, the purpose of applying the assumption to the fault isolation problem in each solution is similar. The closed world assumption is applied in all diagnosis system solutions since the benchmark problem specifies which faults that can occur in the system.

All solutions except GFM are using the SF assumption. In some solutions, the SF assumption is not used in the whole diagnosis system but only when isolating some of the faults. In GKSV, the SF assumption is used in those cases where no diagnosis test can be designed to be sensitive to a single fault to classify which fault that is present. When one of those faults occurs, the first diagnosis test which is sensitive to several faults triggers. Then the test used for isolation is activated where the test is used to classify which of the possible faults that have occurred. The design of the solution in EB is based on the SF assumption since each observer models and estimates one specific fault. The UCB and COK solutions uses the SF assumption when they define and use the fault signature matrix together with column-matching during fault isolation. Thus, SF assumption is used in the GKSV, UCB, and COK solutions in cases where the fault isolability requirements are not otherwise fulfilled.

The EX assumption is applied in at least the EB and COK solutions. EX is applied when using column-matching in the fault signature matrices during fault isolation since all diagnosis tests sensitive to the fault are expected to trigger to accurately isolate the present fault. For example in EB, a fault in one of two sensors or an actuator is isolated depending on how two tests triggers. If one of them triggers it is one of the sensors that is faulty and if both triggers it is the actuator that is faulty.

Since GKSV uses support vector machines to design the diagnosis tests, which requires both fault-free data and faulty data, the solution is based on the FM assumption. The UCB solution uses fault-free data to design one of the residuals and GFM use fault-free data to design the

hypothesis tests. However, since no faulty data is used, the FM assumption is not considered to be used in these cases.

When comparing the different solutions it is clear that there is a difference in how different assumptions are used. The first case is when the method is designed based on a given assumption, for example in CKSV where faulty data are required to train the support vector machines, or in EB where the bank of Kalman filters is selected based on the SF assumption. In the second case, the assumptions can be viewed as tools during the diagnosis system design that are used when necessary, for example, to achieve isolability requirements which are otherwise not fulfilled.

5.3 Results

A summary of the analysis is presented in Table 1 where an X represents that an assumption is made in the diagnosis system solution. In cases where an assumption is only applied partially in the diagnosis system, this is marked using (X). As for the case where it is not clear if an assumption is applied or not, * is used. Note that all training and evaluation scenarios of the benchmark problem mainly consider single faults. The result of the competition is available at com (2014) and the top three positions are presented in Table 1.

Table 1. A summary of which assumptions, described in Section 4, that are utilized in different diagnosis system designs.

Design	CW	SF	EX	FM	Position
GKSV	X	(X)		X	1
EB	X	X	(X)		2
UCB	X	(X)	*		3
COK	X	(X)	(X)		
GFM	X				

This simple analysis indicates that the different solutions are utilizing different assumptions about faults when performing fault isolation. Since the benchmark problem mainly considers single fault scenarios, this is also implemented in almost all solutions. Exoneration is applied at least partially in two of the fault isolation logics in order to isolate the present fault. The SVM design in GKSV is based on training data from the faulty system. The GFM solution applies a few assumptions and is described in Odgaard and Stoustrup (2012) as having relatively slow fault detection. However, it also mentions that it generally performs relatively better when faults occurs at other conditions compared to training data.

6. DISCUSSION

The case study is a good example to show that the solution space of possible diagnosis system designs for a given application is huge. Some assumptions have traditionally been widely used in different research areas, such as the exoneration and single fault assumptions in the FDI community (Cordier et al., 2004). In Section 4, the examples have shown that the assumptions applied in different diagnosis approaches can have a significant impact on both detectability and isolability performance of the developed diagnosis system. Different problems and applications can justify the use of different assumptions in order to improve diagnosability performance without significantly increasing the risk of making faulty conclusions. However, if the applied assumptions are not valid for the given system, the conclusion made by the diagnosis system can not be trusted. Balancing diagnosability performance against robustness with respect to the assumptions made about the system is an important factor in the diagnosis system design. Therefore, it is difficult to make a fair comparison of different solutions without taking the assumptions made in each case into consideration.

When comparing the performance of different solutions using benchmark systems, there are sometimes uncertainties when analyzing which assumptions they have utilized. An analysis, made as in the previous section, could be a complement to other performance metrics as a measure of robustness. Another solution is to have clear specifications of all performance requirements, such as: minimum fault magnitudes, fault time profiles, how many faults that could occur at the same time, etc. Then, it is more clear which assumptions that are valid in each problem. Especially, if the available data from different fault scenarios do not cover all requirements. In this way, it is easier to evaluate which assumptions are more suitable for a given problem, and also which diagnosis approaches that works well for different problem formulations.

7. CONCLUSIONS AND FUTURE WORK

A framework has been proposed to analyze different model-based design strategies of diagnosis systems based on how the observation and feature spaces of residuals are defined. Different diagnosis approaches apply different assumptions about the observation and residual output spaces which have a significant impact on the design as well as the fault hypotheses that are generated. The case study shows that many different solutions can be designed for the same problem where different assumptions are applied. To make a fair comparison of the different solutions it is important to take the different applied assumptions into consideration since they have a significant impact on the performance of each solution.

Future work includes extending the framework to cover other diagnosis approaches which are not covered here, for example, multi-dimensional test quantities or data-driven methods. Also, different benchmark systems proposed in the literature should be analyzed and compared with respect to motivated assumptions.

REFERENCES

(2014). Participation part i (no longer active). URL `http://www.kk-electronic.com/wind-turbine-control/competition-on-fault-detection/participation-part-i-(no-longer-active).aspx`.

Bartyś, M., Patton, R., Syfert, M., de las Heras, S., and Quevedo, J. (2006). Introduction to the damadics actuator fdi benchmark study. *Control Engineering Practice*, 14(6), 577–596.

Basseville, M.E. and Nikiforov, I.V. (1993). Detection of abrupt changes: theory and application.

Bregon, 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. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, PP(99), 1–1.

Chen, W., Ding, S.X., Sari, A., Naik, A., Khan, A.Q., and Yin, S. (2011). Observer-based fdi schemes for wind turbine benchmark. In *Proceedings of IFAC World Congress*, 7073–7078.

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

Cordier, M.O., Travé-Massuyès, L., Pucel, X., et al. (2006). Comparing diagnosability in continuous and discrete-event systems. In *17th Int Workshop on Principles of Diagnosis (DX-06)*, 55–60.

Cordier, M., Dague, P., Dumas, M., Levy, F., Montmain, J., Staroswiecki, M., and Trave-Massuyes, L. (2000). Ai and automatic control approaches of model-based diagnosis: Links and underlying hypotheses. *4th IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes*, 1, 274–279.

de Kleer, J. and Kurien, J. (2003). Fundamentals of model-based diagnosis. *Proceedings of the fifth IFAC symposium on Fault Detection, Supervision, and Safety of technical Processes*, 25–36.

De Kleer, J. and Williams, B.C. (1987). Diagnosing multiple faults. *Artificial intelligence*, 32(1), 97–130.

Frank, P.M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *AUT*, 26(3), 459 – 474.

Frank, P. (1996). Analytical and qualitative model-based fault diagnosis a survey and some new results. *European Journal of Control*, 2(1), 6 – 28.

Gertler, J. (1991). Analytical redundancy methods in fault detection and isolation. In *Proceedings of IFAC/IAMCS symposium on safe process*, volume 1, 9–21.

Gertler, J. and Singer, D. (1990). A new structural framework for parity equation-based failure detection and isolation. *Automatica*, 26(2), 381 – 388.

Isermann, R. (1997). Supervision, fault-detection and fault-diagnosis methods an introduction. *Control Eng Pract*, 5(5), 639 – 652.

Kurtoglu, T., Narasimhan, S., Poll, S., Garcia, D., Kuhn, L., de Kleer, J., van Gemund, A., and Feldman, A. (2009). Towards a framework for evaluating and comparing diagnosis algorithms. *Proc. of the 20th Intern. Workshop on Principles of Diagnosis (DX-09)*, 373–382.

Laouti, N., Sheibat-Othman, N., and Othman, S. (2011). Support vector machines for fault detection in wind turbines. In *Proceedings of IFAC World Congress*, volume 2011, 7067–7072.

Llobet, J.A., Bregon, A., Escobet, T., Gelso, E.R., Krysander, M., Nyberg, M., Olive, X., Pulido, B., and Trave-Massuyes, L. (2009). Minimal structurally overdetermined sets for residual generation: A comparison of alternative approaches. In *Proceedings of IFAC Safeprocess'09*. Barcelona, Spain.

Mosterman, P.J. and Biswas, G. (1999). Diagnosis of continuous valued systems in transient operating regions. *IEEE T-SMCA*, 29(6), 554–565.

Nyberg, M. and Frisk, E. (2006). Residual generation for fault diagnosis of systems described by linear differential-algebraic equations. *IEEE Transactions on Automatic Control*, 51(12), 1995–2000.

Odgaard, P.F. and Stoustrup, J. (2012). Results of a wind turbine fdi competition. In *Proc. of Safeprocess*, volume 2012.

Odgaard, P.F., Stoustrup, J., Kinnaert, M., and de Bruxelles, U.L. (2009). Fault tolerant control of wind turbines–a benchmark model. In *Proc. of Safeprocess*.

Ozdemir, A.A., Seiler, P., and Balas, G.J. (2011). Wind turbine fault detection using counter-based residual thresholding. In *Proceedings of IFAC World Congress*, 8289–8294.

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

Staroswiecki, M. and Comtet-Varga, G. (2001). Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *AUT*, 37(5), 687 – 699.

Svärd, C. and Nyberg, M. (2011). Automated design of an fdi-system for the wind turbine benchmark. In *Proceedings of the 18th World Congress of the International Federation of Automatic Control (IFAC)*, volume 18, 8307–8315. Elsevier.

Trave-Massuyes, L., Escobet, T., and Olive, X. (2006). Diagnosability analysis based on component-supported analytical redundancy relations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 36(6), 1146–1160.

Zhang, X., Zhang, Q., Zhao, S., Ferrari, R.M., Polycarpou, M.M., and Parisini, T. (2011). Fault detection and isolation of the wind turbine benchmark: An estimation-based approach. In *Proceedings of IFAC World Congress*, 8295–8300.