**Preprints of the 21st IFAC World Congress (Virtual)**
**Berlin, Germany, July 12-17, 2020**

# Fault diagnosis integrating physical insights into a data-driven classifier

**M. Amine Atoui** [*] **A. Cohen** [**]

[*] *Lab-STICC-CNRS, UMR 6285, FR (e-mail: amine.atoui@univ-ubs.fr*
[**] *Mathematics and Statistics Department, University of West Florida, Pensacola, FL 32514, USA(e-mail: acohen@uwf.edu)*

**Abstract:**
The main goal of this paper is to present a new method for fault detection and isolation with a Bayesian network (BN). This method combines model-based and data-driven frameworks to detect and diagnose single, multiple and unknown faults. We propose an original BN structure with new decision rules. These rules are constructed to take advantage of the prior model knowledge and the available data. Our network presents new perspective to detect unknown fault and outperforms some recent work proposed in Bayesian networks literature. The performance of the method is illustrated on a heating water process simulating several scenarios of operating conditions.

*Keywords:* Bayesian networks, statistical tests, signature matrix, unknown faults

## 1. INTRODUCTION

Complex processes are very hard to manage and monitor due to the different interactions and interconnections between their components. Therefore, industries and businesses increased their interest on the use of fault detection and diagnosis techniques in order to increase processes' productivity and sustainability. Fault detection attempts to determine whether the process is in normal or faulty operating conditions. Fault diagnosis identifies the fault that has occurred in the process.

Two worlds of thinking exist to detect and diagnose faults: model-based and data-driven. Model-based methods are based on an analytical model of the process while data-driven methods are made of statistical models using the available process data. Model-based methods compare the observed behaviour and values of their variables to the model output. The resulting inconsistencies are called residuals. Theses residuals are sensitive to noise, model uncertainties, and faults. Statistical tests are commonly used to monitor these inconsistencies. Data-driven methods approach the fault diagnosis problem as a discrimination/classification problem. Here models are built from data representing the process in different operating conditions. These models are called classifiers. These classifiers are rigorously trained to learn the process's behaviour and decide about the belonging of a new observation.

Both worlds have advantages and disadvantages. In fact, the effectiveness of data-driven methods consists of having good quality and reliable data, which can be challenging to obtain in a complex system. Besides it is difficult to identify all faulty data. It is clear that these methods are highly dependent on the quality and quantity of data available on the process.

[⋆] This work was not supported by any organization

Model-based methods unlike data-driven methods require normal operating conditions training set to decide about the process's behaviour. Model-based methods are preferred to data-driven methods due to their reliability to describe the dynamic of the process with a physical understanding. However, this physical foundations can be a burden to these methods when the system is complex. An accurate analytical model without uncertainties can be hard to obtain, particularly for processes with a huge number of complex interactions.

Therefore, many papers suggested that the creation of a framework combining the two classes of methods would allow an improvement of fault detection and diagnosis approaches Atoui et al. (2015); Ding et al. (2009); Venkatasubramanian et al. (2003). However, in the literature, the majority of contributions is focused on the development or improvement of one of the two classes of methods. Hybrid diagnosis systems have been proposed and discussed in Jung et al. (2018); Tidriri et al. (2018); Atoui et al. (2015); Atoui (2015). In this paper, we propose a new FDD method based on Bayesian networks. Our study shows that physical insights integrated with a data-driven BN classifier improve significantly decision making. This method also handles explicitly multiple faults.

The paper is structured as follows: section 2 introduces the concepts; In the third section we describe the proposed FD method; Section 4 is an evaluation of the proposed method on a simulated process water heater. Conclusions and perspectives are outlined in the final section.

## 2. METHODS AND TOOLS

### *2.1 Bayesian networks*

A Bayesian Network (BN) (Nielsen and Jensen, 2009) is a probabilistic graphical model. It consists of the following

- a directed acyclic graph $\mathbf{G}$, $\mathbf{G}=(V,E)$, where $V$ is the set of nodes of $\mathbf{G}$ (nodes), and $E$ is the set of arcs of $\mathbf{G}$,
- a finite probabilistic space $(\Omega, \mathbb{Z}, p)$, with $\Omega$ a non-empty space, $\mathbb{Z}$ a collection of the subspaces of $\Omega$ and, $p$ a probability measure on $\mathbb{Z}$ with $p(\Omega)=1$,
- a set of random variables $\mathbf{x} = \mathbf{x}_1, \ldots, \mathbf{x}_m$ associated with the nodes of the graph $\mathbf{G}$ and defined on $(\Omega, \mathbb{Z}, p)$, such that
$$p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = \prod_{i=1}^{m} p(\mathbf{x}_i | Pa(\mathbf{x}_i)) \qquad (1)$$
where $Pa(\mathbf{x}_i)$ is the set of the parent nodes of $\mathbf{x}_i$ in $\mathbf{G}$,
- a conditional probability table (CPT) associate to each node, given its parents, describing probabilistic dependencies between variables,
- calculations named inference, performed in respect of the new evidence about one or several nodes of $\mathbf{G}$, to update the network.

BNs are powerful probabilistic tools. They have shown great abilities to detect and diagnose faults in processes (Atoui et al., 2019; Atoui. et al., 2019; Cai et al., 2017). B

### 2.2 Matrix of signatures

A prior knowledge about a process can be described by a set of mathematical equations explaining its dynamic and the relations between its variables. This set of mathematical equations is the foundation of the model-based methods to generate residuals, residual generator. The residuals are the differences between the observed variables and their estimates when the process is on the $k$th operating conditions. It's common to consider estimates of the normal operating conditions, $NoC$. The evaluation of these residuals determines the operating state of the process.

Residuals $\mathbf{r}_1, \ldots, \mathbf{r}_i$ corresponding to the normal operating conditions are usually considered statistically equal to zero as they might be sensitive to errors, noise, and modeling errors. Statistical tests are then used to compare the residuals to their corresponding thresholds. The process is under normal operating conditions if none of the residuals is out of their limits. Obviously, residuals are also sensitive to the presence of faulty operating conditions. To identify the present of faults one can isolate them based on structured residuals. These residuals are built in a way that they are sensitive to certain faults and not to others. The isolation in this case consists in comparing the symptoms and the characteristics of each fault. These characteristics can be assembled into a Signature Matrix (SM), see Table 1 reflecting the sensitivity or robustness of the residuals. This signature matrix is obtained during a learning stage using heuristics and/or analytical knowledge of the process.

| | NoC | $F_1$ | $\cdots$ | $F_K$ |
|---|---|---|---|---|
| $r_1$ | 0 | $l_{1\cdot 1}$ | $\ldots$ | $l_{1\cdot K}$ |
| $\ldots$ | 0 | $\ldots$ | $\ldots$ | $\ldots$ |
| $r_m$ | 0 | $l_{m\cdot 1}$ | $\ldots$ | $l_{m\cdot K}$ |

Table 1. signature matrix

In the signature matrix, each column corresponds to the characteristics of each fault, $\Gamma_k$, and isolate it to differentiate it from the other faults. It represents the relationship, $l_{i\cdot j}, l_{i\cdot j} \in 0,1$ between each fault and the residuals. In the following $l_{i\cdot j} = 0$ refers to the residual $i$ that is not sensitive to fault $j$.

### 2.3 Relationships between BN and SM

Previous studies have proposed various BN structures for fault diagnosis. A discrete BN discriminating between faults in respect of new symptoms (result of the residuals evaluation) and including components reliability was proposed in (Zaidi et al., 2012; Weber et al., 2008). Similarly, based on the signautre matrix, many authors proposed to directly analyze the residuals within a Bayesian network (Atoui et al., 2016; Zhao et al., 2013; Gonzalez et al., 2012; Schwall and Gerdes, 2002). In (Pernestål et al., 2006), the authors proposed a method to learn the relations between residuals and faults directly form data, where several BN structures have been evaluated and compared using the Diesel engine data. (Kawahara et al., 2005), as well, proposed to use data and physical and expert knowledge to diagnose faults in a spacecraft.

Most of these BNs rely on a discrete Bayesian network with a structure deduced from the links residuals-faults provided by the signature matrix. They also provide decisions about the process behaviour without respecting a false alarm rate. None of these networks consider the possibility of emergence of new observations belonging to unknown operating conditions.

## 3. THE METHODOLOGY

### 3.1 Bayesian network scheme

Fault diagnosis can be seen as a classification problem. Data-driven classifiers assign a new faulty observation to one of the known classes. They try to define boundaries based on training data and to accurately distinguish between the process operating conditions using new data. It is then obvious that the decisions made depend tremendously on the available data. If this data is not representative and reliable then it would lead to inaccurate classification.

Faults can be also isolated based on the process's analytical model. This model describes the analytical relations between process's variables. Based on these relationships, different approaches are possible to generate residuals to be evaluated. A straight forward and efficient approach is the structured residuals (Ding, 2008). Such approach is achieved without the use of fault data operating conditions. Here the signature matrix is built such as residuals are, simultaneously or not, sensitive to different subset of faults. Basically, faults are decoupled in a set of tests such as residuals are sensitive to a specific subset of faults, and to any fault is only characterised by a specific subset of residuals. Fault diagnosis here depends a lot on the decoupling of faults. Indeed, structured residuals are designed to well isolate the process's faults, both single and multiple. But, it's difficult to identify and decouple all of them when the process's analytical model is complex, not accurate or not completely available.

On the other hand, data-driven classifiers are very useful when faulty operating conditions data are available but struggle to identify multiple and unknown faults when there is not such relevant data categorized, simulated or known. Therefore, it can be interesting to boost classifiers by the concept of structured residuals to enhance decision making and handle single, multiple and unknown faults.

In the following, we propose a hybrid Bayesian network-based scheme for fault detection and diagnosis. Among efficient data-driven classifiers is the conditional Gaussian classifier (CGN), which is a special configuration of Bayesian networks. CGNs efficiently models the relationships between process variables and could be naturally used to solve classification problems. The proposed network is the one shown in Figure 1. More advanced structures and network can be also considered.

Let a new observation vector $r$ of $\mathbf{r}, \mathbf{r} \in \mathbb{R}^m$ and $K+1$ different classes $C_j, C_j \in \{NoC,\, F_k\}, k \in \{1,\ldots,K\}$. The network shown in Figure 1 could assigns $r$ to the class $C_j$ with the highest a posterior probability $p(\mathbf{D} = C_j|r)$, $\mathbf{D}$ is a discrete variable. The Maximum A Posterior (MAP) rule, $\delta$, can be written as follows

$$\delta : r \in C_{j^*}, \text{where } j^* = \underset{j=1,\ldots,K}{\operatorname{argmax}} \frac{p(C_j)p(r|C_j)}{p(r)} \qquad (2)$$

where $p(C_j)$ represents the a prior probability of the class $C_j, p(r)$ is the normalization factor, which does not affect the decision, and $p(r|C_j)$ is the multivariate normal probability density function of $r$ given $C_j$, with $\mu_j$ and $\Sigma_j$, respectively, the mean vector and the covariance matrix of the class $C_j$.

Based on the MAP rule, the network is then only deciding and discriminating between known and single faults. Therefore, we suggest to enhance this rule and bring to it an analytical insight.

### *3.2 Probabilistic limits*

The MAP rule do not treat multiple and unknown faults. It also do not respect a false alarm rate. For that, we propose to isolate each class statistically and consolidate the boundaries obtained by the classifier.

Our proposal is based on the Hotelling T-squared statistic associated to each class $C_k$. This distance is widely used for fault detection. It's first calculated and then compared to their predefined threshold (control limit $CL$) in respect to a given false alarm rate $\alpha$. We refer to it by multivariate $T^2$. It can be deduced from our proposed network. It's written, in respect to $C_j$, as below

$$(r-\mu_j)^T \Sigma_j^{-1} (r-\mu_j) \qquad (3)$$

and follows approximately the distribution given by

$$\frac{a(N^2-1)}{N(N-a)} F_\alpha(a, N-a) \qquad (4)$$

where $F_\alpha$ is the $1-\alpha$ quantile of a Fisher distribution, which is the control limit, denoted by $CL_{\mathrm{T}^2}$.

Based on equation (3) and the derivation provided in the appendix A, we developed a probabilistic limit $PL_j$ adapted to our network and matching the control limit in (4). The probabilistic limit $PL_j$ with respect of $C_j$ is given by

$$PL_{C_j} = \frac{\frac{p(C_j)}{|\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}CL_{C_j}}}{\sum_p^{K/\{k\}} \frac{p(p)}{|\Sigma_p|^{\frac{1}{2}}} e^{-\frac{1}{2}T_p^2}} \qquad (5)$$

with $K/\{k\}$ means all the $K+1$ classes except the class $k$.

Furthermore, in this paper, we also propose to monitor under a multivariate Bayesian network each residual in respect to its sensitivity to faults. We came up with original probabilistic limits matching the decisions made by univariate $T^2$-based statistical tests associated to residuals. The proposed limits are compared to the posterior probability of the $NoC$ class to decide whether a residual $i$ deviate from it's normal behaviour. It's worth to mention that structured residuals are by nature assumed to be conditionally independent and that here only data from normal operating conditions are considered. These probabilistic limits can be deduced and obtained from Appendix B. We refer to this probabilistic limit by $PL_{r_j}^{NoC}$. In the following, we note the subset of tests associated to a fault $F_k$, $\Gamma(F_k)$. $\Gamma(F_k)$ is a Boolean variable. Its value depends on whether the residuals corresponding to $F_k$ are triggered or not.

### *3.3 The proposed algorithm for fault diagnosis*

In practice, it's not obvious to identify exactly and describe efficiently a process's operating conditions. Hence, it can be interesting to design a hybrid system and consider that some new observations do not belong to any of the considered operating conditions.

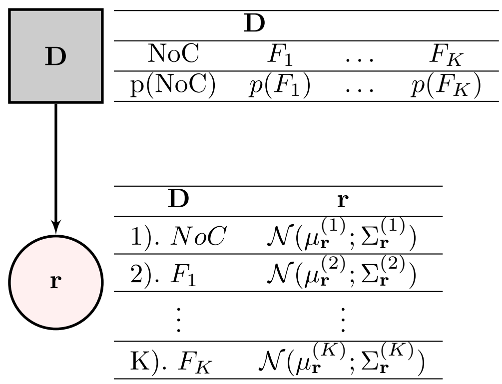


Fig. 1. A CGN classifier

In (Atoui et al., 2019), authors divide the decision space into $K+1$ sub-spaces, where a new sub-space represents the class $UFC$, unknown/ not defined operating conditions. They compare a new observation to the statistical boundary associated to the class of fault with the highest posterior probability. Therefore, if it does not belong to it then it belongs to the class $UFC$.

This rule is interesting as it can handle unknown faults and enhance Bayesian network decision making process. In this paper, we expand this rule to consider as well the multiple fault problem. We combined this data-driven approach to the structured residual model-based concept.

Our methodology differs from the BNlimit in terms of the fault diagnosis methodology. Once a fault is detected, a sorting in ascending order is performed to the outputs of the network except the normal operating conditions. Basically, we collect in $E$ the faults with the highest posterior probability to the one with the lowest posterior probability. Based on the class' weight (posterior probability), we compare in descending order the posterior probability of each class in $E$ to the fault signature corresponding limits (see Appendix B). One can obviously notice that $\alpha$ control the degree of acceptation/ exclusion, a higher value of $\alpha$ would lead to more rejection. These rules are given in Algorithm 3.3, which shows the steps we propose to detect and isolate single, multiple and unknown operating conditions under a BN, for instance, as the one proposed in Fig. 1.

**Algorithm** Fault detection and isolation of known and unknown faults

**Input**: a new observation (residuals) $r : [r_1, \dots, r_m]$
**Outputs**: the class $C_j, C_j \in \{NoC, F_1, \dots, F_K, UoC\}$ to which $r$ belongs

Calculate $p(\mathbf{D} = C_j|r)$;
**if** $p(\mathbf{D} = NoC|r) \geq PL_r^{NoC}$ **then**
    $r \in NoC$ ▷ Detection
**else**
    Sort on E, $\{F_1, \dots, F_K\}$ in ascending order
    **while** $e < K - 1$ **do**
        **if** $\Gamma(F_{E(e)})$ **then**
            UoC $= false$
            $r \in$ E($e$) ▷ Diagnosis
**if** Uoc **then**
    $r \in$ UoC ▷ Rejection

In the next section, we shall demonstrate the applicability of the proposed approach using the heating water process.

## 4. APPLICATION

### *4.1 process description*

To illustrate the interest of our approach, we use a water heater process (Atoui et al., 2016). It consists of a tank (see Figure 2) equipped with two resistors $\mathbf{R}_1$ and $\mathbf{R}_2$. The inputs are the water flow rate $\mathbf{Q}_i$, the electric power for heating $\mathbf{P}$ and the water temperature $T_i$. The outputs are the water flow rate $\mathbf{Q}_0$ and the temperature $\mathbf{T}$ regulated around an operating point. The thermal process objective is to assure a constant water flow rate at a given temperature.

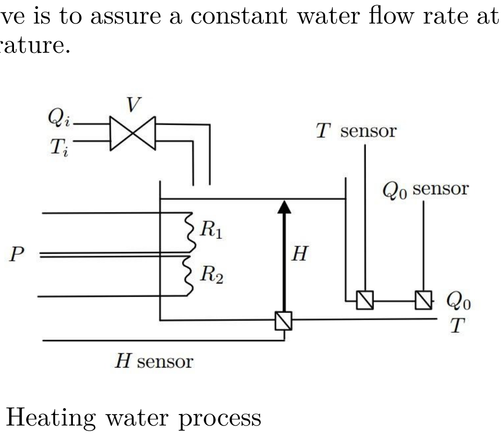


Fig. 2. Heating water process

In this analysis, only sensor and components faults are considered: $\mathbf{H}$ water level sensor, $\mathbf{T}$ output temperature sensor, $\mathbf{Q}_0$ output flow rate sensor. Using Luenberger observer, for each instant $p$, a residuals vector $[r_1(p)\ r_2(p)] = [T(p) - \hat{T}(p)\ H(p) - \hat{H}(p)]^T$ is generated to detect a fault occurrence on $\mathbf{H}$ level sensor or $\mathbf{T}$ temperature sensor. Moreover, according to the physical equation between output flow rate $\mathbf{Q}_0$ and liquid level $\mathbf{H}$ (determined by using the Torriceli-rule: $Q_0(p) = \eta\sqrt{H(p)}$), another residual $r_3(p) = [Q_0(p) - \hat{Q_0}(p)]$ is generated.

The residuals generator $\mathbf{r}(p) = [\mathbf{r}_1(p)\ \mathbf{r}_2(p)\ \mathbf{r}_3(p)]^T$ is structured and evaluated in respect to its corresponding signature matrix. The residuals are sensitive to faults, $l_{i \cdot j} = 1$, as below

$$\mathbf{r}_1 = \{F_T\}, l_{1 \cdot 1} = 1,$$
$$\mathbf{r}_2 = \{F_H\}, l_{2 \cdot 2} = 1,$$
$$\mathbf{r}_3 = \{F_H, F_{Q_0}\}, l_{3 \cdot 2} = 1, l_{3 \cdot 3} = 1$$

### *4.2 Scenarios*

To demonstrate our proposal, we generate samples as training data as follows

- Normal operating Conditions: 80 samples of each residual;
- $\mathrm{F}_T$ operating conditions : 80 samples of each residual;
- $\mathrm{F}_{Q_0}$ operating conditions: 80 samples of each residual.

This data are used to learn our classifier. One can notice the absence of the fault 2, $F_H$. We choose to consider it as an unknown operating condition to test our approach. We recall the signature of $F_H$ involve two residuals, $\mathbf{r}_1$ and $\mathbf{r}_2$.

To compare our approach and present its interest we generated 250 residuals as test data. We have considered the following cases/ scenarios (see table 2) as mentioned in the following

- case 1: normal operating conditions case - we simulated free fault samples.
- case 2: we simulate faulty operating conditions - we consider a single fault: fault 2, $\mathrm{F}_T$.
- case 3: in this case we consider multiple faults - We simultaneously consider fault 1, $\mathrm{F}_T$, and fault 3, $\mathrm{F}_{Q_0}$.
- case 4: we are still in faulty operating conditions. But here we consider only fault $\mathrm{F}_{Q_0}$.
- case 5: we simulated unknown faulty operating conditions. We simulate Fault $\mathrm{F}_H$, and collect residuals.

| cases | samples | Operating cdt. | Fault(s) | Label |
|---|---|---|---|---|
| 1 | 1 : 50 | NoC | none | NoC |
| 2 | 51 : 100 | $\mathrm{F}_T$ | single (=1) | $\mathrm{F}_T$ |
| 3 | 101 : 150 | $\mathrm{F}_T$ & $\mathrm{F}_{Q_0}$ | multiple (>1) | UoC |
| 4 | 151 : 200 | $\mathrm{F}_{Q_0}$ | single (=1) | $\mathrm{F}_{Q_0}$ |
| 5 | 201 : 250 | UoC | unknown | UoC |

Table 2. The different proposed scenarios

### *4.3 Results*

We have compared our new Bayesian network to the BN-based method proposed in (Atoui et al., 2019). The obtained results are given in Figures 3 and 4.

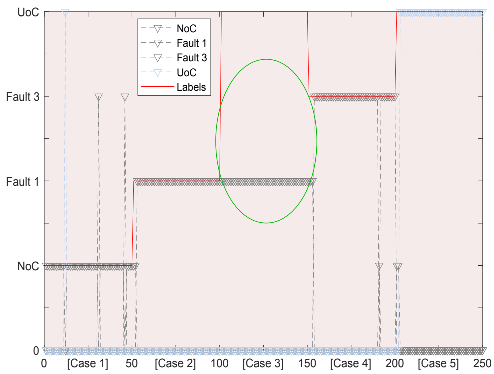


Fig. 3. Results of the approach proposed by (Atoui et al., 2019)

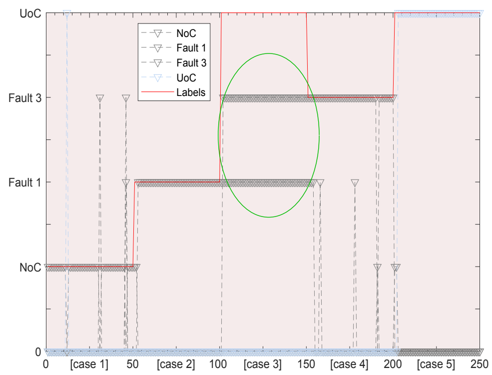


Fig. 4. Results of our proposed approach

Our approach is able to classify correctly single, multiple and unknown faults under a same tool. It combines the faults signature, a model-based concept, with the discrimination ability of data-driven classifiers. The obtained results are very promising.

In the normal operating conditions scenarios, the two methods provide similar results. They are both able to respect a false alarm rate. Therefore they statistically decide about the presence of faults. In case 2 where we have introduced Fault $F_T$, similarly, both approaches were able to recognise it.

We have considered multiple faults in scenario 3. We simultaneously introduced fault $F_T$ and fault $F_{Q_0}$. One can notice in the figures that only our approach is able to recognize both faults while the data-driven approach proposed in (Atoui et al., 2019) is able to only recognize $F_T$. Our proposal gives a new perspective to enhance decisions under BN classifiers. It relies on the signature of faults learned from the knowledge of the process. This way the proposed method drive up a BN classifier to handle multiple faults.

In case 4 we are still in faulty operating conditions. But here we consider only fault 3, $F_{Q_0}$. Once again, we show the capacity of the methods to discriminate between faults. We simulated unknown faulty operating conditions in case 5. We have simulated Fault 2, $F_H$, and collected residuals. These residuals are presented for the first time to both methods. Unknown faults are not well considered in the FDD literature whereas it's an important piece in the fault diagnosis puzzle. We can notice that our approach similarly to BNlimit are able to consider the new observations simulated in case 5 as unknown operating conditions.

## 5. CONCLUSION

In this paper we proposed an interesting Bayesian network scheme for fault detection and diagnosis. Our proposal outperforms the state of the art. The hybrid diagnosis BN-based scheme can be a serious solution to diagnose multiple and unknown faults in a multivariate process. The obtained results are promising and incite for further investigation.

## Appendix A. FAULT ISOLATION

Let us consider the statistic $\Delta_k$ associated to a class $C_j$, $j = \{1, \ldots, K+1\}$, given by:

$$\Delta_j^2 = (\mathbf{r} - m_j)^T S_j^{-1} (\mathbf{r} - m_{C_j}) \tag{A.1}$$

where $m_j$ and $S_j$ are the considered parameters of the distribution of the class $C_j$.

The control limit, $CL_{\Delta_j}$, associated to $\Delta^2$, in respect to $\alpha$, is given by

$$\frac{a(N^2-1)}{N(N-a)} F_\alpha(a, N-a) \tag{A.2}$$

with $F_\alpha$ is the $1-\alpha$ quantile of a Fisher distribution, which is the control limit, denoted by $CL_{\mathrm{T}_j^2}$.

Then, an observation $r$ belongs to the class $C_j$ if $T^2$ is lower or equal to its control limit $CL_{\mathrm{T}_j^2}$. Let's develop the inequality equation as below

$\mathbf{r} \in C_j$, if,

$$\mathrm{e}^{-\frac{1}{2}\Delta_{C_j}} \geq \mathrm{e}^{-\frac{1}{2}CL_\Delta^{C_j}}$$

$$\frac{1}{2\pi^{\frac{m}{2}}|S_{C_j}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}\Delta^{C_j}} \geq \frac{1}{2\pi^{\frac{m}{2}}|S_{C_k}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}CL_\Delta^{C_j}}$$

$$p(r|\mathbf{D} = C_j) \geq p(r^*|\mathbf{D} = C_j) \tag{A.3}$$

where $r^*$ is an observation of $\mathbf{r}$ with $r^* \in C_j$ and $\Delta_{C_j} = CL_\Delta^{C_k}$.

Let's multiply each side of (A.3) by $p(r)$ as below

$$\frac{p(\mathbf{r}|\mathbf{D} = C_j)p(\mathbf{D} = C_j)}{p(\mathbf{r})} \geq \frac{p(\mathbf{r}^*|\mathbf{D} = C_j)p(\mathbf{D} = C_j)}{p(\mathbf{r})} \tag{A.4}$$

where

$$p(\mathbf{r}) = \sum_{K}^{k=1} p(\mathbf{r}|\mathbf{D} = C_j)p(\mathbf{D} = C_j) \tag{A.5}$$

Thus, we deduce the following rule

$$\mathbf{r} \in C_j, \text{ if } p(\mathbf{D} = C_j|\mathbf{r}) \geq PL^{C_k} \tag{A.6}$$

with

$$PL^{C_k} = \frac{p(\mathbf{r}^*|\mathbf{D} = C_j)p(\mathbf{D} = C_j)}{p(\mathbf{r})} \tag{A.7}$$

This rule allows to test statically the membership of a new observation to the class $C_j$. It's worth to mention that $p(\mathbf{D} = C_j|\mathbf{r})$ corresponds to the posterior probability of an observation $r$ of $\mathbf{r}$ given the value $C_j$ of the node $\mathbf{D}$.

## Appendix B. DETECTION OF FAULTY RESIDUALS

In the following we will focus only on the normal operating class $NOC$ where residuals are assumed to be conditionally independent.

Consider a univaraite statistic $\delta$, for a given residual $\mathbf{r}_i, i = 1, \ldots, m$ and in respect to class NoC, given by:

$$\delta_{NoC}^{\mathbf{r}_i} = (\mathbf{r}_i - m_{NoC}^i)^2/\sigma_{NoC}^2 \tag{B.1}$$

where $m_{NoC}$ and $\sigma_{NoC}$ are the considered parameters of the distribution of $\mathbf{r}_i$ under the class $NoC$, assumed to follow a normal distribution.

Let's develop the inequality equation given below

$$\begin{aligned} r_j \in NoC, &\text{if,} \\ \delta_{NoC}^{\mathbf{r}_i} &\leq CL_\delta^{NOC} \\ \mathrm{e}^{-\frac{1}{2}\delta_{NoC}^{\mathbf{r}_i}} &\geq \mathrm{e}^{-\frac{1}{2}CL_\delta^{NoC}} \\ \frac{1}{2\pi^{\frac{m^i}{2}}\sigma_{NoC}}\mathrm{e}^{-\frac{1}{2}\frac{(x-m_{C_i}^j)^2}{\sigma_{NoC}^2}} &\geq \frac{1}{2\pi^{\frac{m^i}{2}}\sigma_{NoC}}\mathrm{e}^{-\frac{1}{2}CL_\delta^{NoC}} \\ p(\mathbf{r}_i|NoC) &\geq p(\mathbf{r}_i^*|NoC) \end{aligned}$$

In respect to the nature of the structured residuals, we continue the development started above, in the context of normal operating conditions, as below

$$\begin{aligned} p(\mathbf{r}_{/\mathbf{r}_j}|\mathbf{r}_j, NoC)(\mathbf{r}_j|NoC) &\geq p(\mathbf{r}_{/\mathbf{r}_j}|\mathbf{r}_j, NoC)p(\mathbf{r}_j^*|NoC) \\ p(NoC|r) &\geq \frac{p(\mathbf{r}_{/\mathbf{r}_j}|NoC)p(\mathbf{r}_j^*|NoC)}{p(\mathbf{r})} \end{aligned}$$

where $\mathbf{r} = [\mathbf{r}_1, \ldots, \mathbf{r}_m]^T$, $r^j$ is an observation of $\mathbf{r}_j$ with $r_j^* \in NOC$ and $\Delta_{NoC} = CL_\Delta^{NoC}$. $\mathbf{r}_{/\mathbf{r}_j}$ is the vector $\mathbf{r}$ except $\mathbf{r}_j$. The distribution of $\mathbf{r}_j^*$ is given by

$$p(\mathbf{r}_{/\mathbf{r}_j}|NoC) = \prod_{\mathbf{r}_s \in \mathbf{r}_{/\mathbf{r}_j}} p(\mathbf{r}_s|NoC) \tag{B.2}$$

Thus, we deduce the following rule

$$r_j \in NOC, \text{ if } p(NOC|\mathbf{r}) \geq PL_{\mathbf{r}_j}^{NOC} \tag{B.3}$$

with

$$PL_{\mathbf{r}_j}^{NOC} = \frac{p(\mathbf{r}_{/\mathbf{r}_j}|NOC)p(\mathbf{r}_j^*|NOC)p(NOC)}{p(\mathbf{r})} \tag{B.4}$$

## REFERENCES

Atoui., M.A., Cohen., A., Rauffet., P., and Berruet., P. (2019). Fault diagnosis by bayesian network classifiers with a distance rejection criterion. In *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO,*, 463–468. INSTICC, SciTePress. doi:10.5220/0008053304630468.

Atoui, M.A., Cohen, A., Verron, S., and Kobi, A. (2019). A single bayesian network classifier for monitoring with unknown classes. *Engineering Applications of Artificial Intelligence*, 85, 681–690.

Atoui, M.A. (2015). *Contributions à la détection et au diagnostic de fautes dans les systèmes par réseaux Bayésiens*. Ph.D. thesis, Université d'Angers.

Atoui, M.A., Verron, S., and Kobi, A. (2015). Fault detection with conditional gaussian network. *Engineering Applications of Artificial Intelligence*, 45, 473–481.

Atoui, M.A., Verron, S., and Kobi, A. (2016). A bayesian network dealing with measurements and residuals for system monitoring. *Transactions of the Institute of Measurement and Control*, 38(4), 373–384.

Cai, B., Huang, L., and Xie, M. (2017). Bayesian networks in fault diagnosis. *IEEE Transactions on Industrial Informatics*, 13(5), 2227–2240.

Ding, S.X. (2008). *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media.

Ding, S., Zhang, P., Naik, A., Ding, E., and Huang, B. (2009). Subspace method aided data-driven design of fault detection and isolation systems. *Journal of process control*, 19(9), 1496–1510.

Gonzalez, R., Huang, B., Xu, F., and Espejo, A. (2012). Dynamic bayesian approach to gross error detection and compensation with application toward an oil sands process. *Chemical engineering science*, 67(1), 44–56.

Jung, D., Ng, K.Y., Frisk, E., and Krysander, M. (2018). Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation. *Control Engineering Practice*, 80, 146–156.

Kawahara, Y., Yairi, T., and Machida, K. (2005). Diagnosis method for spacecraft using dynamic bayesian networks. In *Proc. of 8th International symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*. Citeseer.

Nielsen, T.D. and Jensen, F.V. (2009). *Bayesian networks and decision graphs*. Springer Science & Business Media.

Pernestål, A., Nyberg, M., and Wahlberg, B. (2006). A bayesian approach to fault isolation with application to diesel engine diagnosis. In *17th Workshop on Principles for Diagnosis, Peñaranda de Duero, Burgos, Spain, June 26-28, 2006*, 211–218.

Schwall, M.L. and Gerdes, J.C. (2002). A probabilistic approach to residual processing for vehicle fault detection. In *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*, volume 3, 2552–2557. IEEE.

Tidriri, K., Tiplica, T., Chatti, N., and Verron, S. (2018). A generic framework for decision fusion in fault detection and diagnosis. *Engineering Applications of Artificial Intelligence*, 71, 73–86.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., and Yin, K. (2003). A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers & chemical engineering*, 27(3), 327–346.

Weber, P., Theilliol, D., and Aubrun, C. (2008). Component reliability in fault-diagnosis decision making based on dynamic bayesian networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 222(2), 161–172.

Zaidi, A., Bouamama, B.O., and Tagina, M. (2012). Bayesian reliability models of weibull systems: State of the art. *International Journal of Applied Mathematics and Computer Science*, 22(3), 585–600.

Zhao, Y., Xiao, F., and Wang, S. (2013). An intelligent chiller fault detection and diagnosis methodology using bayesian belief network. *Energy and Buildings*, 57, 278–288.