



HAL
open science

Coupling data-driven and model-based methods to improve fault diagnosis

Mohamed Amine Atoui, Achraf Cohen

► **To cite this version:**

Mohamed Amine Atoui, Achraf Cohen. Coupling data-driven and model-based methods to improve fault diagnosis. *Computers in Industry*, 2021. hal-03175813

HAL Id: hal-03175813

<https://inria.hal.science/hal-03175813>

Submitted on 10 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Coupling data-driven and model-based methods to improve fault diagnosis

M. Amine Atoui^a, Achraf Cohen^b

^a*CNRS Lab-STICC, University of Bretagne-sud*

^b*Mathematics and Statistics Department, University of West Florida, Pensacola, FL 32514, USA*

Abstract

Monitoring a system is often not an easy task and the best approach to address it would be to develop a monitoring system that uses data, expert knowledge, and mathematical models. Combining these three sources of information on the system is often unpractical of various reasons such as in complex systems. In this paper, a hybrid method for diagnosing single and multiple simultaneous faults, while considering unknown operating conditions, is proposed. The proposed method is integrated in a Bayesian network classifier with innovative decision rules combining statistical decisions and fault signature matrix. A heating water process is utilized for illustrative and evaluation purposes, in which several scenarios of operating conditions are simulated. The results, in terms of classification rates, show superior performance for the proposed method over a purely data-driven method. It is worth mentioning that the proposed method shows higher capabilities to isolating multiple simultaneous faults that the data-driven method fails to accomplish.

Keywords: fault detection, fault isolation, Bayesian networks, classifiers, signature matrix, unknown faults

1. Introduction

Industries and businesses continuously increase their interests in using Statistical Process Monitoring (SPM) for purposes of improving process pro-

Email addresses: amine.atoui@univ-ubs.fr (M. Amine Atoui), acohen@uwf.edu (Achraf Cohen)

Preprint submitted to Elsevier

December 19, 2020

ductivity and quality. The tasks involved in SPM include: (1) fault detection, attempts to determine whether the process is in normal or faulty operating conditions; (2) fault diagnosis, identifies the faults that have occurred (3) fault estimation, evaluates the fault magnitude; and (4) fault reconstruction, which estimates the fault-free data to continue monitoring even if some faults have occurred. Two main approaches can be utilized to detect and diagnose faults: model-based and data-driven.

Model-based techniques are based on an analytical model of the system while data-driven methods are made of statistical models using the available process data. Model-based methods compare the observed values to the model output. The resulting inconsistencies are called residuals. These residuals are sensitive to noise, model uncertainties, and faults. Statistical tests are commonly used to monitor these inconsistencies. Model-based methods unlike data-driven methods only require normal operating conditions training set to decide about the process 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, the 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. More details on model-based methods can be found in (Isermann, 2005).

Data-driven methods manage the fault diagnosis problem as a discrimination/classification problem. Models are built from data representing the process in different operating conditions. These models are called classifiers. These classifiers are trained to learn the process behaviour and help to decide about the state (faulty or not) of a new observation. The effectiveness of data-driven methods requires having good quality and reliable data, which can be challenging to obtain in a complex system. Besides, it is difficult to have data that identify all possible faulty operations. Some of the techniques that have been widely studied for process monitoring include principal component analysis (PCA) (Yin et al., 2014; Kano et al., 2002), partial least squares (PLS) (Peng et al., 2015), wavelets analysis (Cohen and Atoui, 2020). Several machine learning methods (Bishop, 2006) have been applied such as support vector machines, neural networks (Heo and Lee, 2018), and decision trees (Liu et al., 2018). These methods are powerful but require a considerable amount of reliable data.

The two methods have advantages and disadvantages. The creation of a framework combining the two categories would allow an improvement of Fault

Detection and Diagnosis (FDD) approaches (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 within a Bayesian network.

Bayesian Network classifiers (BNCs) are classification models (Nielsen and Jensen, 2009) that have been widely applied for FDD. BNCs use probabilistic reasoning and can be designed using physical knowledge or system data or both. Several published research papers have investigated the benefits of fault diagnosis using Bayesian models (Qin et al., 2018; Jun and Kim, 2017; Don and Khan, 2019; Wang et al., 2018; Herrera-Vega et al., 2018). BNs are powerful probabilistic tools to deal with uncertainties (Barber, 2012). For instance, Wang et al. (2019a) proposed a simplified BN for fault isolation based on the correlations among faults and applied it to an engine fuel injection system. Cai et al. (2017) presented a review on use of BNs in fault diagnosis in the last decades. Researchers presented different approaches with a rich bibliography. Bond Graph (BG) was integrated with a BN for fault diagnosis (Bouallegue et al., 2017; Lo et al., 2011). Yu and Zhao (2019) proposed to use an ensemble method into a BN to improve the diagnosis capabilities for complex systems. In Liu et al. (2020) a methodology BN-based for subsea christmas tree is presented. Wang et al. (2019b) presented a discrete Bayesian network for chiller fault diagnosis. An acoustical damage detection Bayesian network of the wind turbine yaw system is proposed in Chen et al. (2020).

An open circuit two layers fault diagnosis Bayesian network is proposed in (Han et al., 2019), where it was applied to inverter insulated gate bipolar transistors. In (Liu et al., 2015), the two layer model is augmented by another layer based on operation procedures. Another two layers BNs are used by Cai et al. (2016) for fault diagnosis of complex electronic systems in the presence of transient and intermittent faults. Nguyen et al. (2020) generated residuals based on the analytical redundancy relations (Ding, 2008) and evaluated them based on the two-layer BN concept to isolate faults. A Bayesian classifier is used to discriminate between possible leak scenarios based on pressure residuals (Soldevila et al., 2017). Similar networks sharing this same topological structure of two-layer BN (Schwall and Gerdes, 2002) has been proposed (Schwall and Gerdes, 2005; Pernestål et al., 2008; Gonzalez et al., 2012; Atoui et al., 2016; Soldevila et al., 2017; Nguyen et al., 2020; Song et al., 2020; Chen et al., 2020; Song et al., 2020).

The symptoms or residual-based Bayesian networks describe the relations

between faults and root-cause variables. They are the marriage between the fault diagnosis task and the probabilistic processing. A few papers discussed hybrid diagnosis systems where data-driven methods and model-based methods are combined (Jung et al., 2018; Tidriri et al., 2018; Atoui et al., 2016).

Most of the mentioned BNs methods rely on a discrete form with a complex structure deduced from the links residuals-faults provided by the signature matrix. Also, the continuous residuals are often discretized. During this process of discretization, some of the information is lost. Furthermore, most BNs mentioned above provide statistical decisions with a false alarm rate higher than expected because of inflation with multiple testing.

In this work, the proposed methodology consists of coupling data-driven and model-based approaches into a BNC structure to detect and diagnose faults. BNC's maximum a posteriori rule is modified to deal with multiple simultaneous testing, which should avoid a higher false alarm rate. A simple BN structure is proposed to incorporate the signature matrix. The proposed algorithm enables single, multiple simultaneous, and unknown faults to be detected and isolated. A thorough search of the relevant literature as demonstrated above yielded none of BN methods were proposed to combine classifier (data driven) and residuals (model-based) decisions for fault diagnosis.

The paper is structured as follows: Section 2 introduces background and definitions. In the third Section, the method and algorithm are presented. Section 4 presents the results of evaluating the performance of the method using a simulated process water heater. Conclusions and perspectives are given in the final Section.

2. Background

2.1. Probabilistic directed acyclic graphical models

Such models are also called Bayesian networks. They are powerful probabilistic tools. They have utilized to model complex systems for faults detection and diagnosis purposes. BNs can be built using the expert(s) knowledge, physical models, or using data about the system. These information are complementary and often improve the modeling to understand the relationships between the process parameters and variables.

Definition 1 *A directed acyclic graph \mathbf{G} is composed of a set of nodes V and a set of arcs E , such as it represents a finite probabilistic space, It consists*

of a set of random variables $\mathbf{r}_1, \dots, \mathbf{r}_m$ representing the nodes of G such that

$$p(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m) = \prod_{i=1}^m p(\mathbf{r}_i | Pa(\mathbf{r}_i)), \quad (1)$$

where $Pa(\mathbf{r}_i)$ indicates the variables parents of \mathbf{r}_i in \mathbf{G} .

Each variable in regards of its parent variables in G is defined by a conditional probability table (CPT). These tables are then updated, through inference algorithms, in respect of the observations of one or several nodes in V .

2.2. Fault signature matrix

A priori knowledge about a process can be characterized 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 is common to consider deviations from the normal operating conditions, *NoC*. The evaluation of these residuals determines the operating state of the process.

Residuals $\mathbf{r}_1, \dots, \mathbf{r}_k$ 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 presence of faults one can isolate them based on structured residuals. These residuals are built in a way that they are sensitive to specific faults and not to all faults. The isolation in this case consists in comparing the symptoms and the characteristics of each fault (Ding, 2008). These characteristics can be assembled into a fault signature matrix reflecting the sensitivity or robustness of the residuals.

| | F_1 | F_2 | F_3 |
|-------|-------|-------|-------|
| r_1 | 0 | 1 | 0 |
| r_2 | 1 | 0 | 0 |
| r_3 | 1 | 0 | 1 |

Table 1: Example of a fault signature matrix

In the signature matrix, we define Γ_{F_k} corresponds to the signature of each operating condition. Its elements are 0 or 1 indicating whether or not a residual is sensitive to an operating condition. Table 1 presents an example of fault signature matrix where $\Gamma_{F_1} = \{1, 1, 0\}$ (column) indicating that residuals 1 and 2 are sensitive to fault 1.

3. Methodology and algorithm

The task of fault diagnosis can be understood as a classification problem where classifiers assign a new observation to a class among the known classes. 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. On the other side, faults can be also isolated based on the process analytical model. This model describes the analytical relations between process variables. Based on these relationships different approaches are possible to generate residuals to be evaluated. A [straightforward](#) and efficient approach is the structured residuals (Ding, 2008). The signature matrix is built such that 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 particular subset of faults, and each fault is just characterised by a certain subset of residuals. Therefore, a new observation belongs to the fault whose residuals are affected. Structured residuals are designed to well isolate the process faults, both single and multiple. But, it is often difficult to identify and decouple all of them when the process 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 simultaneous 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.

We shall present a hybrid Bayesian network-based scheme for fault detection and diagnosis. The proposed network is shown in Figure 1.

Let r be an observed vector of $\mathbf{r}, \mathbf{r} \in \mathbb{R}^m$ and suppose there are $K + 1$ different classes $C_j, C_j \in \{NoC, F_k(Faults)\}, k \in \{1, \dots, K\}$. The network shown in Figure 1 should assign r to class C_j that has the highest a posterior probability $p(\mathbf{D} = C_j | r)$, \mathbf{D} is a discrete variable.

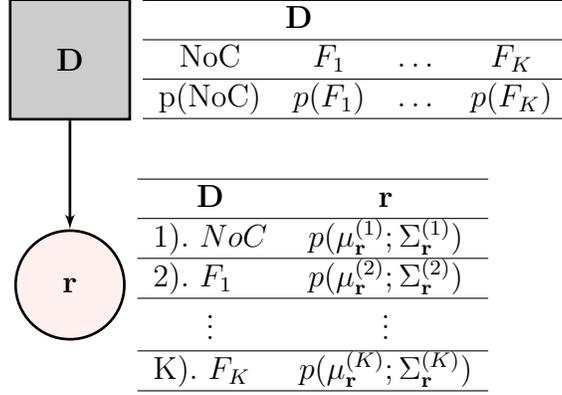


Figure 1: An example of a condensed Bayesian network classifier

Definition 2 *The Maximum A Posterior (MAP) rule (Duda et al., 2012), δ , can be then written as $\delta: r \in C_{j^*}$, with*

$$j^* = \arg \max_{j=1, \dots, K+1} p(C_j)p(C_j|r), \quad (2)$$

where $p(C_j)$ represents the prior probability of the class C_j , $p(r)$ is the normalization factor, and $p(r|C_j)$ is the density function of r given C_j , defined as $p(r|C_j) = 2\pi^{-\frac{m}{2}} |\Sigma_j|^{-\frac{1}{2}} e^{-\frac{1}{2}(r-\mu_j)^T \Sigma_j^{-1} (r-\mu_j)}$ with μ_j and Σ_j are its parameters, respectively.

According to the MAP rule the network only discriminates between known and single faults. Therefore, one could suggest to enhance this rule by including an analytical insight.

3.1. Probabilistic limits

The MAP rule does not treat multiple and unknown faults. **It also does not take into account a fixed false alarm rate.** One way to fix this would be to isolate each class statistically and consolidate the boundaries obtained by a data-driven classifier.

The proposal is based on the Hotelling T-squared statistic associated to each class C_j . This distance is widely used for fault detection. This statistic has a known distribution for multivariate normal data. Then, the statistic can be compared to its upper control limit CL in respect to a given false

alarm rate α . We refer to it by multivariate T^2 . It can be deduced from the proposed network. It is written, in respect to C_j , C_j , $j = \{1, \dots, K + 1\}$, as follows:

$$T_{C_j}^2(r) = (r - \mu_j)^T \Sigma_j^{-1} (r - \mu_j) \quad (3)$$

and follows approximately the distribution given by (when Σ is estimated)

$$\frac{m(N_j^2 - 1)}{N_j(N_j - m)} F_{\alpha_j}(m, N_j - m) \quad (4)$$

where F_{α_j} is the $(1 - \alpha_j)$ th quantile of a Fisher distribution with degrees of freedom m and $N_j - m$, which is the control limit, denoted by $CL_{T_j^2}$.

Definition 3 *The results of the statistical test in (3) and the threshold in (4) are matched by the outputs of the BN given in Figure 1 compared to the probabilistic limit PL_j . This limit is given, with respect of the class C_j , by*

$$PL^{C_k} = \frac{e^{-(CL_{T_{C_k}^2} - T_{C_k}^2(r))}}{1 + \sum_{v \neq k} \omega_v \frac{|\Sigma_k|^{\frac{1}{2}}}{|\Sigma_{C_v}|^{\frac{1}{2}}} e^{-(T_{C_v}^2(r) - T_{C_k}^2(r))}} \quad (5)$$

such as

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

with ω_v a weight related to class C_v .

Furthermore, in this paper, under a multivariate Bayesian network each residual is monitored 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 deviates from its 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. Residuals are built in a way that they are sensitive to specific faults and not to all faults. A residual i is sensitive to a fault j if it deviates from its normal behavior in its presence. Fault j is therefore considered as present if the residuals associated to its characteristics are all tested positive to a change in their normal behavior. 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.

Proposition *A residual i is triggered, declared faulty, if the posterior probability of the residuals regarding the NoC class is lower than the probabilistic limit $PL_{r_i}^{NoC}$*

Proof *Consider the univariate statistic δ , for a given observation r of the residual $r_i, i = 1, \dots, m$ and in respect to class NoC, given by:*

$$\delta_{NoC}^{r_i} = (r_i - m_{NoC}^i)^2 / \sigma_{NoC}^2 \quad (7)$$

where m_{NoC} and σ_{NoC}^2 are the distribution parameters of the class NoC.

Let develop the inequality equation given below

$r_j \in NoC$ implies that

$$\begin{aligned} \delta_{NoC}^{r_i} &\leq CL_{\delta}^{NoC} \\ \frac{1}{\sqrt{2\pi}\sigma_{NoC}} e^{-\frac{1}{2} \frac{(r - m_{NoC}^j)^2}{\sigma_{NoC}^2}} &\geq \frac{1}{\sqrt{2\pi}\sigma_{NoC}} e^{-\frac{1}{2} CL_{\delta}^{NoC}} \\ p(r_i | NoC) &\geq p(r_i^* | NoC) \end{aligned}$$

In respect to the nature of the structured residuals, where $p(r_{/r_i} | NoC) = \prod_{r_s \in r_{/r_i}} p(r_s | NoC)$, the development started above is conducted, in the context of normal operating conditions, as below ($r_{/r_i}$ is the vector r except r_i)

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

where $r = [r_1, \dots, r_m]^T$, r_i is an observation of r_i with $r_i^* \in NOC$. Thus, the following rule is deduced

$$r_i \in NOC, \text{ if } p(NoC | r) \geq PL_{r_i}^{NoC} \quad (8)$$

where

$$PL_{r_i}^{NoC} = \frac{e^{-(CL_{\delta}^{NoC} + T_{NoC}^2(r_{/r_i}) - T_{NoC}^2(r))}}{1 + \sum_{v \neq NoC} \omega_v \frac{|\Sigma_{NoC}|^{\frac{1}{2}}}{|\Sigma_{C_v}|^{\frac{1}{2}}} e^{-(T_v^2(r) - T_{NoC}^2(r))}} \quad (9)$$

3.2. The proposed algorithm for fault diagnosis

In application, it is often challenging to efficiently characterize a process operating conditions, because it could be sometimes hard to obtain data on faulty conditions that are seldom or that come with high risk. Therefore, some of the observations could belong to none of the predefined classes and thus it is relevant to design a monitoring system that consider the possibility that some new observations belong to a new class or in other words does not belong to any of the predefined faults classes.

Atoui et al. (2019) proposed to use $K + 2$ classes, where one of the classes represented the UFC (Unknown Fault Class) unknown or undefined states operating conditions. The decision rules were based on comparing the value of a new observation to the statistical control limit corresponding to the fault that has the highest a posteriori probability. Therefore, if the new observation does not happen to belong to the faulty class with the highest posterior probability, then it is labelled as UFC class.

The proposed algorithm is an extension of the mentioned work to account for multiple simultaneous faults and enhance the decision rules, with a combination of data-driven approach and the structured residual model-based approach. Here, when a fault is detected, a different approach is considered to decide about the belonging of the new observation. First of all, the faulty classes are compared and sorted in E in respect to their posterior probabilities for a given observation r . Next, the posterior probability of the NoC class is compared, in respect to each residuals i , to the limit given in Eq. (8). So, residuals' evaluation is contrasted to each fault's signature Γ_{F_k} . Then, the triggered faults would constitute a ranked subset of E . This subset represents a set of fault that are most likely responsible of the faulty conditions. When this subset is empty we consider that the observation belongs to Unknown Operating Conditions (UoC). We change the Unknown Fault Class (UFC) to Unknown Operating Conditions (UoC) because an observation that does not belong to any defined classes could be faulty or just another normal condition. The proposed algorithm is presented in Figure 2.

This work enhances decision making process under BN. The proposed framework is able to detect and diagnosis single, multiple simultaneous faults and unknown operating conditions. In the next section, we illustrate and assess the performance of the new approach on multiple scenarios of the water heater process (Weber et al., 2008).

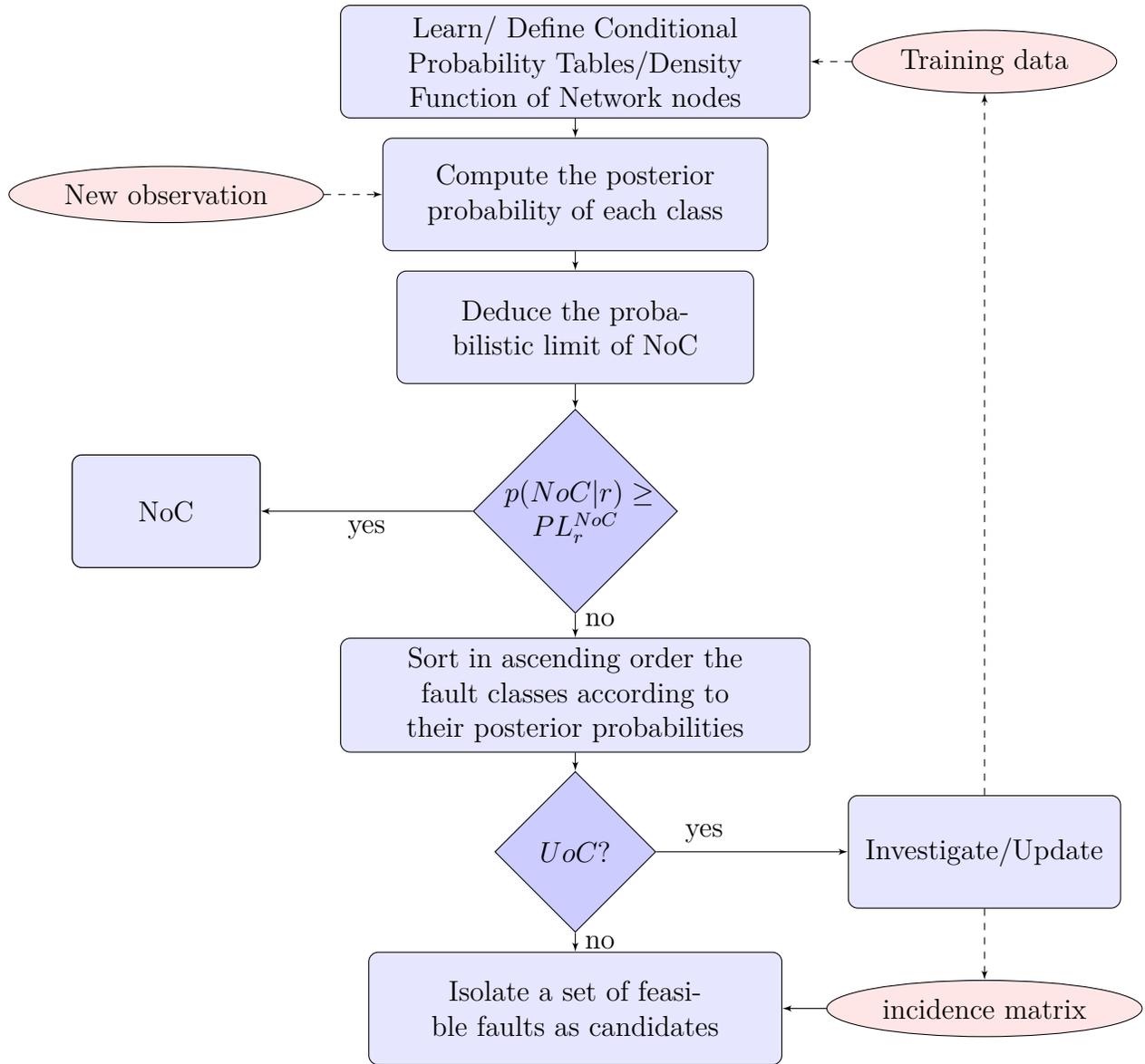


Figure 2: Diagram of the proposed BN strategy to detect and diagnose faults

Algorithm 1 Diagnosis of single, multiple simultaneous and unknown faults

Input: a new observation (residuals) $r : [r_1, \dots, r_m]$

Outputs: one of these sets NoC, Θ_r and UoC to which r belongs, with $\Theta_r \subset \{F_1, \dots, F_K\}$

Calculate $p(\mathbf{D} = C_j|r)$, $C_j \in \{NoC, F_1, \dots, F_K\}$;

if $p(\mathbf{D} = NoC|r) \geq PL_r^{NoC} \parallel \Gamma_{NoC}$ **then**

$r \in NoC$

▷ Detection

else

Sort $E = \{F_1, \dots, F_K\}$ in ascending order according to posterior probabilities

while $i < K - 1$ and UoC **do**

UoC = *False*

if $\Gamma_{E(i)}$ (Residuals triggered) **then**

UoC = *True*

$r \in E(i)$

add $E(i)$ to Θ_r

▷ Diagnosis

if $\Theta_r = \emptyset$ **then**

$r \in UoC$

▷ Rejection

One can collect similar observations and identify new operating conditions.

4. Application and performance

4.1. Process description

To illustrate the proposed approach, a water heater process is utilized (Atoui et al., 2016). It is made up of a tank, see Figure 3. Its inputs are the the heating electric power \mathbf{P} , water flow part \mathbf{Q}_i and water temperature T_i . The temperature \mathbf{T} regulated around an operating point and the water flow part \mathbf{Q}_0 are its outputs.

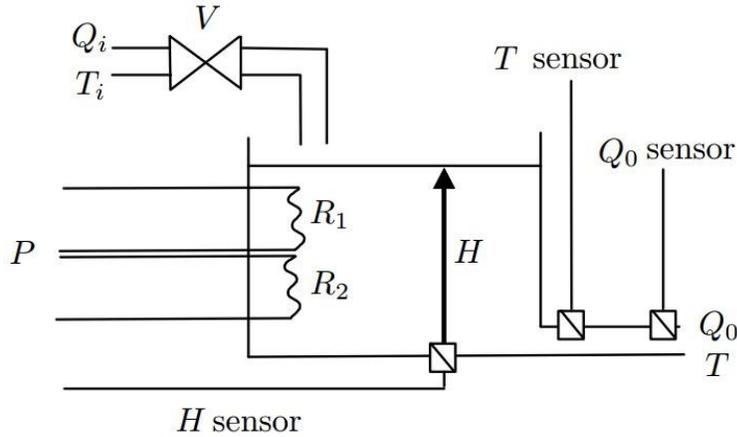


Figure 3: Heating water process

The thermal process goal is to monitor the water flow rate. The hydraulic and thermal equations describing the process are:

$$\begin{cases} S \frac{dH(t)}{dt} = Q_i(t) - Q_0(t) \\ \frac{dT(t)}{dt} = \frac{P(t)}{\rho C S H(t)} - \frac{(T(t) - T_i) \times Q_i(t)}{S H(t)} \end{cases} \quad (10)$$

where S is the section, ρC is a constant, and T_i is equal to $20^\circ C$.

In this investigation, the following sensors are considered: \mathbf{H} water level sensor, \mathbf{T} output temperature sensor, and \mathbf{Q}_0 output flow part sensor.

An observer is introduced, such that for each instant p , a residual vector $[r_1(p) \ r_2(p)] = [T(p) - \hat{T}(p) \ H(p) - \hat{H}(p)]^T$ is generated. These two residuals would be useful to monitor the outcomes of the sensors \mathbf{H} or \mathbf{T} . A third residual $r_3(p) = [Q_0(p) - \hat{Q}_0(p)]$ is produced in respect to the physical equation between output flow part \mathbf{Q}_0 and liquid level \mathbf{H} (determined by using the Torricelli-rule: $Q_0(p) = \eta \sqrt{H(p)}$).

The generator $\mathbf{r}(p) = [\mathbf{r}_1(p) \ \mathbf{r}_2(p) \ \mathbf{r}_3(p)]^T$ outcomes are structured. This means that the residuals are created such as they are sensitive to certain faults and robust to others; each fault signature is unique. The faults signatures are gathered in the incidence matrix given in Table 2.

| | F_T | F_H | F_{Q_0} |
|----------------|-------|-------|-----------|
| \mathbf{r}_1 | 1 | 0 | 0 |
| \mathbf{r}_2 | 0 | 1 | 0 |
| \mathbf{r}_3 | 0 | 1 | 1 |

Table 2: Water heater process fault signature matrix

4.2. Illustrative cases and performance

To evaluate the proposed method, five cases are simulated and presented in Table 3. [case 1] consists of 50 free-fault observations (*NoC*). [case 2] introduces 50 observations from F_T . [case 3] is a multiple fault case where F_T and F_{Q_0} are present. In [case 4] only fault F_{Q_0} is present, and the last case, [case 5], is simulated to model the unknown operating conditions (*UoC*), where F_H is introduced. It is worth to recall that F_H is not part of the training data and the model should not classify observations from F_H as known faults.

| Cases | Samples | Op. cdt. | Fault(s) | Label |
|-------|-----------|-------------------|---------------|-----------|
| 1 | 1 : 50 | NoC | none | NoC |
| 2 | 51 : 100 | F_T | single (=1) | F_T |
| 3 | 101 : 150 | F_T & F_{Q_0} | multiple (>1) | UoC |
| 4 | 151 : 200 | F_{Q_0} | single (=1) | F_{Q_0} |
| 5 | 201 : 250 | UoC | unknown | UoC |

Table 3: The different simulated scenarios ran 100 times

Next, we generate a training data set of 100 residual observations for each of the following operating conditions: NoC, F_T , and F_{Q_0} . This data set is used to learn the model. It is worth to mention that Fault F_H is not part of the training data set because it is considered as unknown operating conditions for testing purposes. To evaluate the model, a testing data set of 250 residual observations is generated. 100 simulations were run to account for the random generation process. We provide the results in terms of the average of **classification rates** and their standard errors.

| Cases | data-driven model labels | | | |
|-------------------|--------------------------|--------------------|--------------------|---------------------|
| | NoC | F_T | F_{Q_0} | UoC |
| NoC | 98.16 (1.97) | 0 (0) | 0.38 (0.92) | 1.46 (1.81) |
| F_T | 1.94 (0.34) | 95.8 (0.47) | 0 (0) | 2.18 (0.63) |
| F_T & F_{Q_0} | 0 (0) | 0.88 (1.45) | 0 (0) | 99.12 (1.45) |
| F_{Q_0} | 2.16 (0.54) | 0 (0) | 85.7 (2.52) | 12.14 (2.49) |
| UoC | 2 (0) | 0 (0) | 0 (0) | 98 (0) |

Table 4: The average classification rates (standard error)% using only the data-driven method

4.3. Results and discussion

We compare the proposed method with a data-driven method proposed in (Atoui et al., 2019). This comparison should provide evidence on whether or not the hybrid approach is efficient compared to a data-driven method in this context. Even a small improvement should count given the fact that the system is not complex. The obtained results are given in Tables 4 and 5. Recall that there are 5 cases, see Table 3.

| Cases | Proposed model labels | | | |
|-------------------|-----------------------|----------------|---------------------|---------------------|
| | NoC | F_T | F_{Q_0} | UoC |
| NoC | 99.4 (1.08) | 0 (0) | 0.26 (0.88) | 0.34 (0.69) |
| F_T | 1.96 (0.28) | 98 (0) | 0.06 (0.79) | 0.04 (0) |
| F_T & F_{Q_0} | 0 (0) | 100 (0) | 97.82 (0.75) | 0 (0) |
| F_{Q_0} | 2.3 (0.82) | 8.96 (1.30) | 97.14 (0.88) | 0.4 (0) |
| UoC | 2 (0) | 0.02 (0.51) | 0.02 (0.51) | 97.98 (0.51) |

Table 5: The average classification rates (standard error)% combining the data-driven and model-based methods

The hybrid method proposed in this paper shows promising and effective diagnosis qualities. It is able to classify correctly single, multiple, and unknown faults with a higher correct classification rate than the only data-driven method.

For instance, [case 1] the normal operating condition scenario: both methods provide similar results, with a slightly improvement in favor of the proposed method 99.4%. In [case 2] Fault F_T is introduced and the proposed method is superior.

In [case 3], multiple simultaneous faults are considered. Faults F_T and F_{Q_0} are simultaneously introduced. One can notice from Tables 4 and 5, the proposed approach is able to recognize both faults (F_T : 100%, F_{Q_0} : 97.82%), however the data-driven labels the case as unknown operating conditions (99.12%). This is indeed one of the advantages of the proposed method.

In fact, data-driven classifiers tend to decide about the belonging of a new observation by discriminating between the defined classes. Only one of them should be picked as the responsible class. Basically, it is challenging to diagnose multiple faults under a data-driven classifier as it would lead us to add classes representing the simultaneous appearance of several faults. In this case, the number of fault classes can tremendously increase, which leads to a larger number of parameters and complex models. On the other side, the proposed method attempts to solve this problem by taking advantage of signature of faults (model-based) as illustrated in [case 3].

In addition, the proposed algorithm does not only well isolate known and unknown faults but also provide further relevant information by ranking the single faults in the multiple faults scenario. This can be useful when the signature matrix would suggest a set of feasible faults as candidates for the diagnosis. The ranking here would help the operator to prioritize the faults with higher occurring probabilities.

[case 4] deals with a single fault F_{Q_0} . Consistently, the powerful capability of the proposed method to discriminate F_{Q_0} is demonstrated. The results in this case are significantly better than the data-driven approach (97.14% versus 85.7%). Unknown faulty operating conditions are considered in [case 5] and both methods perform very well. Unknown faults are often ignored in the literature although they are an important piece in the fault diagnosis puzzle and that is probably because it is impossible to gather data of faulty operating conditions that are not known already. Moreover, structured residuals do not deal with unknown operating conditions. This can be explained by the signature of F_H . F_H affects residual 2 but also residual 3. The latter is also sensitive to fault F_{Q_0} . Therefore, a deviation of the mean of residual 3 from its operating conditions leads to an alarm but a false one in terms of identification.

Therefore, the proposed method allows the diagnosis of single and multiple faults as well as recognizes unknown operating conditions.

5. Conclusion

In this study a Bayesian network classifier for fault detection and diagnosis is proposed. The proposed method combines data-driven and model-based approaches to introduce an hybrid diagnosis tool that is able to effectively diagnose single faults, multiple simultaneous faults, and unknown operation conditions. To demonstrate the credibility of the proposed approach, five scenarios were simulated on a water heater process.

For future work, it would be interesting to extend the proposed algorithm to provide the most minimal set of fault candidates with the most likely explanations in the case of multiple faults. At this stage, the algorithm only provides a ranked set of faults in respect to all the residuals sensitivities. Another relevant path to investigate is to update the classifier in order to handle efficiently the observations that were labelled unknown operating conditions. It would be useful to automatically and gradually add a new class of fault or a new normal operation condition class.

References

- Isermann, R.. Model-based fault-detection and diagnosis—status and applications. *Annual Reviews in control* 2005;29(1):71–85.
- Yin, S., Ding, S.X., Xie, X., Luo, H.. A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics* 2014;61(11):6418–6428.
- Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss, R., et al. Comparison of multivariate statistical process monitoring methods with applications to the eastman challenge problem. *Computers and Chemical Engineering* 2002;26(2):161–174.
- Peng, K., Zhang, K., You, B., Dong, J.. Quality-related prediction and monitoring of multi-mode processes using multiple pls with application to an industrial hot strip mill. *Neurocomputing* 2015;168:1094–1103.
- Cohen, A., Atoui, M.A.. On wavelet-based statistical process monitoring. *Transactions of the Institute of Measurement and Control* 2020;doi:0142331220935708.
- Bishop, C.M.. *Pattern recognition and machine learning*. springer; 2006.

- Heo, S., Lee, J.H.. Fault detection and classification using artificial neural networks. *IFAC-PapersOnLine* 2018;51(18):470–475.
- Liu, S., Xu, L., Li, Q., Zhao, X., Li, D.. Fault diagnosis of water quality monitoring devices based on multiclass support vector machines and rule-based decision trees. *IEEE Access* 2018;6:22184–22195.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., Yin, K.. A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers & chemical engineering* 2003;27(3):327–346.
- Nielsen, T.D., Jensen, F.V.. *Bayesian networks and decision graphs*. Springer Science & Business Media; 2009.
- Qin, A., Hu, Q., Lv, Y., Zhang, Q.. Concurrent fault diagnosis based on bayesian discriminating analysis and time series analysis with dimensionless parameters. *IEEE Sensors Journal* 2018;19(6):2254–2265.
- Jun, H.B., Kim, D.. A bayesian network-based approach for fault analysis. *Expert Systems with Applications* 2017;81:332–348.
- Don, M.G., Khan, F.. Dynamic process fault detection and diagnosis based on a combined approach of hidden markov and bayesian network model. *Chemical Engineering Science* 2019;201:82–96.
- Wang, Z., Wang, L., Liang, K., Tan, Y.. Enhanced chiller fault detection using bayesian network and principal component analysis. *Applied Thermal Engineering* 2018;141:898–905.
- Herrera-Vega, J., Orihuela-Espina, F., Ibarzüengoytia, P.H., García, U.A., Morales, E.F., Sucar, L.E., et al. A local multiscale probabilistic graphical model for data validation and reconstruction, and its application in industry. *Engineering Applications of Artificial Intelligence* 2018;70:1–15.
- Barber, D.. *Bayesian reasoning and machine learning*. Cambridge University Press; 2012.
- Wang, J., Wang, Z., Stetsyuk, V., Ma, X., Gu, F., Li, W.. Exploiting bayesian networks for fault isolation: A diagnostic case study of diesel fuel injection system. *ISA transactions* 2019a;86:276–286.

- Cai, B., Huang, L., Xie, M.. Bayesian networks in fault diagnosis. *IEEE Transactions on Industrial Informatics* 2017;13(5):2227–2240.
- Bouallegue, W., Bouslama, S., Tagina, M.. Robust fault detection and isolation in bond graph modelled processes with bayesian networks. *International Journal of Computer Applications in Technology* 2017;55(1):46–54.
- Lo, C., Wong, Y.K., Rad, A.B.. Bond graph based bayesian network for fault diagnosis. *Applied soft computing* 2011;11(1):1208–1212.
- Yu, W., Zhao, C.. Online fault diagnosis for industrial processes with bayesian network-based probabilistic ensemble learning strategy. *IEEE Transactions on Automation Science and Engineering* 2019;16(4):1922–1932.
- Liu, P., Liu, Y., Cai, B., Wu, X., Wang, K., Wei, X., et al. A dynamic bayesian network based methodology for fault diagnosis of subsea christmas tree. *Applied Ocean Research* 2020;94:101990.
- Wang, Y., Wang, Z., He, S., Wang, Z.. A practical chiller fault diagnosis method based on discrete bayesian network. *International Journal of Refrigeration* 2019b;102:159–167.
- Chen, B., Xie, L., Li, Y., Gao, B.. Acoustical damage detection of wind turbine yaw system using bayesian network. *Renewable Energy* 2020;.
- Han, S., He, Y., Zheng, S.. Open-circuit fault diagnosis of an inverter based on bayesian network. In: *Chinese Intelligent Systems Conference*. Springer; 2019, p. 62–70.
- Liu, Z., Liu, Y., Cai, B., Zheng, C.. An approach for developing diagnostic bayesian network based on operation procedures. *Expert Systems with Applications* 2015;42(4):1917–1926.
- Cai, B., Liu, Y., Xie, M.. A dynamic-bayesian-network-based fault diagnosis methodology considering transient and intermittent faults. *IEEE Transactions on Automation Science and Engineering* 2016;14(1):276–285.
- Nguyen, T.N., Downar, T., Vilim, R.. A probabilistic model-based diagnostic framework for nuclear engineering systems. *Annals of Nuclear Energy* 2020;149:107767.

- Ding, S.X.. Model-based fault diagnosis techniques: design schemes, algorithms, and tools. Springer Science & Business Media; 2008.
- Soldevila, A., Fernandez-Canti, R.M., Blesa, J., Tornil-Sin, S., Puig, V.. Leak localization in water distribution networks using bayesian classifiers. *Journal of Process Control* 2017;55:1–9.
- Schwall, M.L., Gerdes, J.C.. A probabilistic approach to residual processing for vehicle fault detection. In: *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*; vol. 3. IEEE; 2002, p. 2552–2557.
- Schwall, M.L., Gerdes, J.C.. Residual autocorrelation in probabilistic model-based diagnostics. In: *ASME International Mechanical Engineering Congress and Exposition*; vol. 42169. 2005, p. 1255–1264.
- Pernestål, A., Wettig, H., Silander, T., Nyberg, M., Myllymäki, P.. A bayesian approach to learning in fault isolation. In: *19th International Workshop on Principles of Diagnosis*. Citeseer; 2008, p. 143.
- Gonzalez, R., Huang, B., Xu, F., Espejo, A.. Dynamic bayesian approach to gross error detection and compensation with application toward an oil sands process. *Chemical engineering science* 2012;67(1):44–56.
- Atoui, M.A., Verron, S., Kobi, A.. A bayesian network dealing with measurements and residuals for system monitoring. *Transactions of the Institute of Measurement and Control* 2016;38(4):373–384.
- Song, G., Rossi, P.S., Khan, F., Paltrinieri, N., BahooToroody, A.. Model-based information fusion investigation on fault isolation of subsea systems considering the interaction among subsystems and sensors. *Journal of Loss Prevention in the Process Industries* 2020;:104267.
- Jung, D., Ng, K.Y., Frisk, E., Krysander, M.. Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation. *Control Engineering Practice* 2018;80:146–156.
- Tidriri, K., Tiplica, T., Chatti, N., Verron, S.. A generic framework for decision fusion in fault detection and diagnosis. *Engineering Applications of Artificial Intelligence* 2018;71:73–86.

- Duda, R.O., Hart, P.E., Stork, D.G.. Pattern classification. John Wiley & Sons; 2012.
- Atoui, M.A., Cohen, A., Verron, S., Kobi, A.. A single bayesian network classifier for monitoring with unknown classes. Engineering Applications of Artificial Intelligence 2019;85:681–690.
- Weber, P., Theilliol, D., Aubrun, C.. 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 2008;222(2):161–172.