

A Combined Data-Driven and Model-Based Residual Selection Algorithm for Fault Detection and Isolation

Daniel Jung and Christofer Sundström

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-149582>

N.B.: When citing this work, cite the original publication.

Jung, D., Sundström, C., (2017), A Combined Data-Driven and Model-Based Residual Selection Algorithm for Fault Detection and Isolation, *IEEE Transactions on Control Systems Technology*, PP(99), 1-15. <https://doi.org/10.1109/TCST.2017.2773514>

Original publication available at:

<https://doi.org/10.1109/TCST.2017.2773514>

Copyright: Institute of Electrical and Electronics Engineers (IEEE)

<http://www.ieee.org/index.html>

©2017 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.



A Combined Data-Driven and Model-Based Residual Selection Algorithm for Fault Detection and Isolation

Daniel Jung and Christofer Sundström

Abstract—Selecting residual generators for detecting and isolating faults in a system is an important step when designing model-based diagnosis systems. However, finding a suitable set of residual generators to fulfill performance requirements is complicated by model uncertainties and measurement noise which have negative impact on fault detection performance. The main contribution is an algorithm for residual selection which combines model-based and data-driven methods to find a set of residual generators that maximizes fault detection and isolation performance. Based on the solution from the residual selection algorithm, a generalized diagnosis system design is proposed where test quantities are designed using multi-variate residual information to improve detection performance. To illustrate the usefulness of the proposed residual selection algorithm, it is applied to find a set of residual generators to monitor the air path through an internal combustion engine.

Index Terms—Fault diagnosis, Fault detection, Machine learning, Change detection algorithms, Automotive applications.

I. INTRODUCTION

FAULT diagnosis is used to detect faults that occur in a system, but also pinpoint what part of the system that is faulty. There are data-driven diagnostic approaches available as well as model-based approaches.

In model-based diagnosis, a model of the system to be monitored is used to compute residuals to detect inconsistencies between the model and system measurements to detect when faults occur [21]. A general model-based design of diagnosis systems is based on a set of residual generators where different residual generators are sensitive to, i.e. they should respond to, different sets of faults that can occur in the system, see for example [28] and [41]. Based on the residuals that have triggered, a fault isolation algorithm [8] computes a set of diagnosis candidates, or fault hypotheses, that can explain the triggered residuals.

There are efficient methods based on the so called structural analysis, described in for example [1], [24], [25], [35], and [40], to find sets of residual generators to ideally achieve a high degree of fault isolability. However, these methods only consider which faults each residual generator is sensitive to, in the ideal case. Therefore, the signal to noise ratio in the residuals is not taken into consideration in the diagnosis system design process. Fig. 1 shows a set of residuals, computed from a set

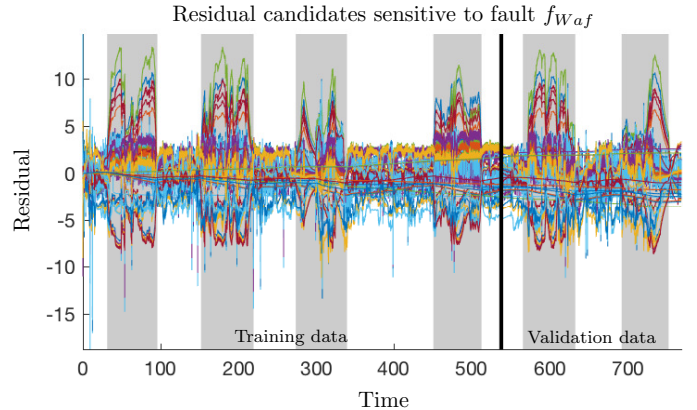


Fig. 1. A comparison of residuals sensitive to the fault f_{Waf} , but with different detection performance, when evaluated using data with fault f_{Waf} . The intervals with intermittent faults are gray-shaded.

of model-based residual generators, where the presence of an intermittent fault is highlighted. All residuals in the figure are ideally sensitive to the fault, but it is clear that only a few of the residuals significantly deviate from their nominal behavior presented in the non-shaded areas. Residual selection considers the problem of finding residual generators that fulfill some given fault detection and isolation performance requirements that should be achieved by a diagnosis system. The figure motivates why it is important to include the quantitative residual performance when formulating the residual selection problem.

To reduce complexity of the diagnostic system, it is preferred to use as small set of residual generators as possible fulfilling the performance requirements. Different search algorithms finding such sets have been proposed, for example Binary Integer Linear Programming (BILP) [31] and different greedy search algorithms [43], [22], [33]. However, in these papers it is assumed that fault detection performance is equal for all residual generator candidates. With respect to previous works, the proposed residual selection algorithm takes quantitative residual performance into consideration.

On the other hand, data driven methods use measurement data to model how the model accuracy and measurement noise affect the diagnostic performance [7], [37]. There are several previous works proposing data-driven classifiers for fault diagnosis, see for example [17] and [47]. One advantage of data-driven fault diagnosis methods is that they can be applied in systems where models are not available [49], [50].

D. Jung was with the Department of Electrical Engineering, Linköping University, Sweden. He is now with the Center for Automotive Research, The Ohio State University, Columbus, OH, USA. (e-mail: daniel.jung@liu.se)

C. Sundström is with the Department of Electrical Engineering, Linköping University, Sweden. (e-mail: christofer.sundstrom@liu.se)

However, the performance of data-driven classifiers depends on training data. It is often difficult and time-consuming to collect a sufficient amount of faulty data to represent each fault mode. Thus, training data is often limited and not always representative of the possible fault realizations that could occur in the system [38]. In this case, a data-driven classifier trained on this data is not expected to achieve reliable fault detection and isolation performance [44].

Given a system model, it is possible to construct residual generators with different fault sensitivity properties meaning that different residual generators will be sensitive to different sets of faults. There are multiple methods to design residual generators, for example Kalman Filters [16] and Particle Filters [48], [51]. By using computer-aided tools, such as [15], it is possible to automatically generate residual generators as executable code in, for example, C or Matlab. Note that the number of residual generators can be significantly larger compared to the set of sensors in the system [24]. Thus, as illustrated in Fig. 1, residual selection is an important step in the diagnosis system design process where the detection performance of the different residual generators must be taken into consideration, for example from faulty data.

Even though training data is not representative of all possible realizations of each fault, it is possible to achieve a diagnosis system that can isolate a fault by utilizing the structural information of the different model-based residual generators. In machine learning, a related problem to residual selection is referred to as *feature selection* [18]. Feature selection is an important topic when designing data-driven models to reduce complexity and the risk of over-fitting. Using data-driven feature selection to select a suitable set of model-based residual generators should take advantage of both model-based and data-driven approaches to maximize fault detection and isolation performance.

A. Main idea and contributions

The main idea in this work is to combine machine learning and model-based methods for residual selection in uncertain systems where model uncertainties and measurement noise cannot be neglected. It is assumed that there exists a model that is used for residual generation, as well as measurement data collected from nominal and faulty system behavior. For example, model information is used to find sets of residual generators that ideally are sensitive to specific faults, but not sensitive to other faults in order to achieve fault isolability properties. A data-driven approach is then used to select residual sets for detecting and isolating different faults, based on the model analysis as well as training data.

The residuals are post-processed to form test quantities. Traditionally, the test quantities are based on selecting the, in some sense, single best residual generators found in the structural analysis. Here, the correlation between different residual generators is considered when designing the diagnostic system to improve detection performance without necessarily adding more residual generators to the diagnostic system.

The proposed diagnostic method is illustrated by designing a diagnosis system to monitor the air path through an

internal combustion engine. Real measurement data have been collected from an engine test bench, including data with injected faults. The residual selection approach is described and demonstrated by applying it to the engine system. The results from the case study shows that the proposed method works well for a system with significant model uncertainties and is able to identify sets of residuals which give good fault detection and isolation performance.

B. Outline

A short summary of basic fault diagnosis definitions is presented in Section II. The combustion engine and the model used is described in Section III. In Section IV residual generator candidates are designed and in Section V the method for selecting the residuals to be used in the diagnosis system is described. Validation of the fault isolability properties using real residual data is presented in Section VI and examples of how the residual generators can be used in the design of the diagnosis system is covered in Section VII, and finally the conclusions are given in Section VIII.

II. BACKGROUND

Before describing the proposed method, some notations and definitions are presented that will be used. Consider a system and a set of n_f faults $\mathcal{F} = \{f_1, f_2, \dots, f_{n_f}\}$ to be monitored by a diagnosis system. A residual generator is defined as follows [43].

Definition 1 (Residual generator): A residual generator $r(z)$ for a given system is a function of sensor and actuator data z where a fault-free system implies that the residual output $r(z) = 0$.

An important property of a residual generator is whether or not it will respond to the presence of a fault in the system.

Definition 2 (Fault sensitivity): A residual generator $r(z)$ is sensitive to a fault f_i if the fault implies that the residual output $r(z) \neq 0$.

It is assumed there exists a set of n_r residual generator candidates $\mathcal{R}_{\text{all}} = \{r_1, r_2, \dots, r_{n_r}\}$ and a description of which faults $\mathcal{F}_k \subseteq \mathcal{F}$ each residual generator $r_k \in \mathcal{R}_{\text{all}}$ is sensitive to. Since the detection performance varies significantly between different residuals, the exoneration assumption is not valid [6], i.e. it is not certain that a fault will trigger all residuals sensitive to the fault.

Since the residual selection problem is considered, the following definitions of fault detectability and isolability are used [43].

Definition 3 (Fault detectability with residual generators \mathcal{R}): A fault f_i is detectable given a residual set \mathcal{R} if there exists a residual generator $r_k \in \mathcal{R}$ that is sensitive to f_i .

Definition 4 (Fault isolability with residual generators \mathcal{R}): A fault f_i is isolable from another fault f_j given a residual set \mathcal{R} if there exists a residual generator $r_k \in \mathcal{R}$ that is sensitive to f_i but not f_j .

In order to isolate faults from each other in model-based diagnosis, residual generators are needed that are sensitive to different sets of faults. If a residual generator r_k is not sensitive to a fault f_j it is said that f_j is *decoupled* in r_k . If the residual

TABLE I
THE INCIDENCE MATRIX OF THE SMALL EXAMPLE MODEL GIVEN IN
EXAMPLE 1.

	x_1	x_2	y	u	f
e_1	X	X			X
e_2		X		X	
e_3	X		X		

generator r_k , where f_j is decoupled, is sensitive to a fault f_i , then, it can be used to isolate the fault f_i from fault f_j .

A. Structural analysis of engine model

To analyze fault diagnosis properties of complex models, several papers propose the use of a structural representation of the system model, see for example, [26] and [28]. A structural model describes which variables that are included in each model equation and can be represented by a logical matrix where an X at position (i, j) means that a variable x_j is included in equation e_i [23], [24]. A small example is used to illustrate the concept.

Example 1: Consider the model

$$\begin{aligned} e_1 : \quad & x_1 = g_1(x_2) + f \\ e_2 : \quad & x_2 = g_2(u) \\ e_3 : \quad & y = x_1 \end{aligned}$$

with 3 equations, 2 unknown variables x_1 and x_2 , known input variable u , measurement variable y , and arbitrary functions g_i . The variable f models a system fault such that, if the fault is present, it will affect equation e_1 . The incidence matrix for this model is given in Table I. \square

The structural analysis is based on this information instead of the model equations and is here performed using the Fault Diagnosis Toolbox in Matlab [15].

III. COMBUSTION ENGINE

A passenger car four cylinder turbo-charged internal combustion engine is used to illustrate the proposed diagnostic approach. This section includes a brief description of the engine as well as the available data.

The engine is mounted in a test bench, see Fig. 2, and can be operated both at steady state and at transients, for example, during different driving cycles. The available measurements from the engine are the following eight sensor signals:

- Pressure before throttle y_{pic}
- Pressure in intake manifold y_{pim}
- Ambient pressure y_{pamb}
- Temperature before throttle y_{Tic}
- Ambient temperature y_{Tamb}
- Air mass flow after air filter y_{Waf}
- Engine speed y_ω
- Throttle position y_{xpos}

and the two actuator signals:

- Wastegate actuator u_{wg}
- Injected fuel mass into the cylinders u_{mf}

These signals represent a standard setup in a production vehicle.

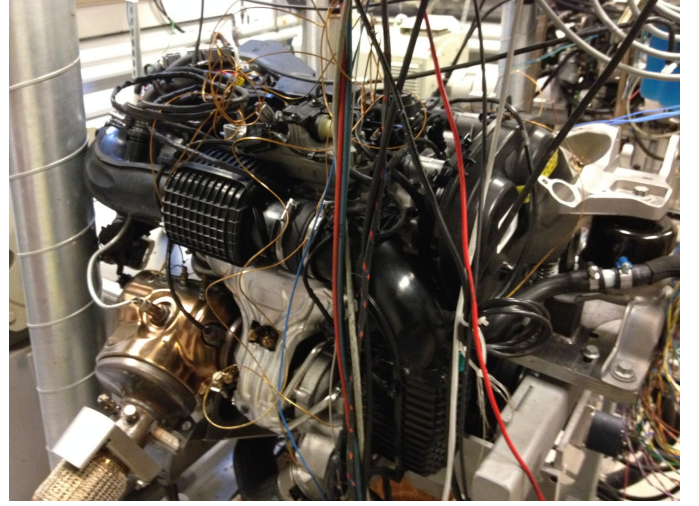


Fig. 2. A picture of the engine test bench.

A. Engine model summary

A mathematical model used here describes the air flow through the engine. The model structure is similar to the model described in [11], and is based on six control volumes and mass and energy flows given by restrictions.

The model has 94 equations, 14 states, and 10 known signals. Non-linear relations, such as if-constraints, and maps, are included in the model. A schematic illustration of the model is shown in Fig. 3, where p_{af} , p_c , p_{ic} , p_{im} , p_{em} , and p_t denote the pressures in the air filter, after the compressor, intercooler, intake manifold, exhaust manifold, and after the turbine, respectively. These pressures indicate where the control volumes are modelled.

In this case study, four sensor faults are considered: A fault in the sensor measuring the air mass flow f_{Waf} , the pressures at the intercooler f_{pic} and the intake manifold f_{pim} , and the temperature at the intercooler f_{Tic} . It is possible to also consider other types of faults, such as, leakages, clogging, and actuator faults. However, the four sensor faults are considered to easier illustrate the concept of the proposed diagnostic approach.

B. Data collection

The engine is controlled to follow a selected driving cycle using a simple driver model and longitudinal vehicle model implemented in Simulink. Measurement data is generated when the FTP75 highway cycle (see for example [19] for the speed profile) is used as a speed reference. Intermittent sensor faults are injected one by one in the engine control unit. The faults f_{Waf} , f_{pic} , and f_{pim} , are injected as multiplicative faults $y_i(t) = (1 + f_i)x_i(t)$ with a 20% change in the measured value and the fault f_{Tic} as a sensor bias $y_{Tic}(t) = x_{Tic}(t) + f_{Tic}$ of 20° .

An example of sensor data from y_{Waf} is shown in Fig. 4 with an intermittent fault f_{Waf} . The signal fluctuates significantly due to the transients in the requested torque from the engine caused by the transients in the driving cycle. These fluctuations in the signal makes it difficult to, for example,

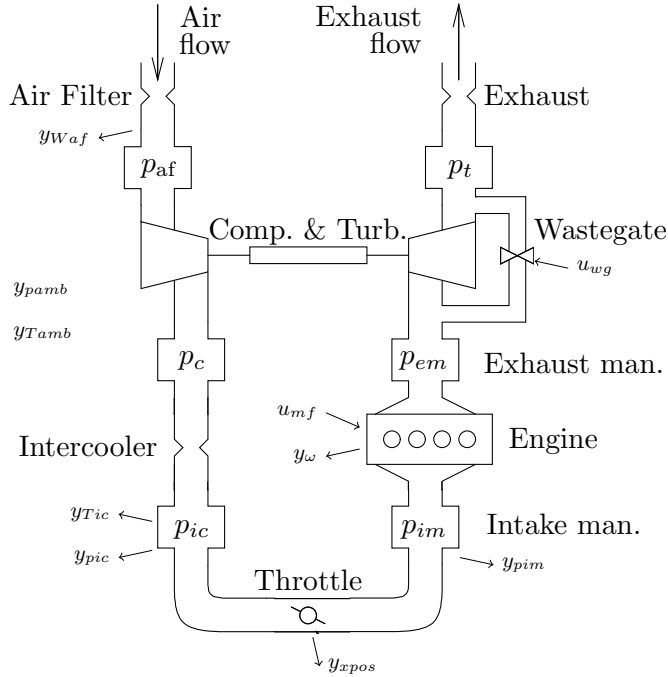


Fig. 3. A schematic of the model of the air flow through the model. This figure is used with permission from [12].

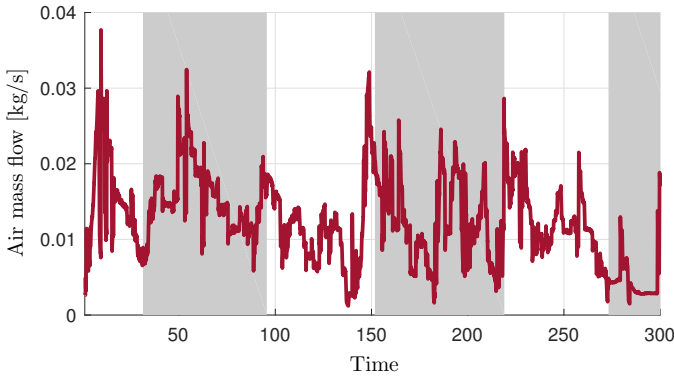


Fig. 4. The air mass sensor data y_{Waf} with an intermittent fault f_{Waf} where the shaded intervals highlight when the fault is present.

threshold the sensor signal to detect whether the sensor is faulty or not. The signal for y_{pim} also varies in a wide span as can be seen in Fig. 5, while the signals for y_{pic} and y_{Tic} are more constant.

The data collected in the study consist of four runs of the highway part of FTP75, one run for each fault. Each data set is 765 seconds and the fault induced in the specific data set is assumed to be intermittent and active approximately half of the time, see for example Fig. 1 where the shaded areas indicate the time slots when the fault f_{Waf} is active.

IV. RESIDUAL GENERATOR CANDIDATES

Generating model-based residual generators require a set of equations with analytical redundancy, i.e., a set of equations where the number of equations is larger than the number of unknown variables. One specific type of such equation sets are

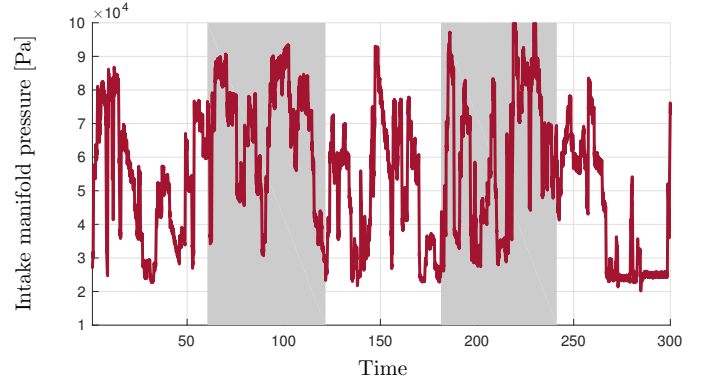


Fig. 5. The air mass sensor data y_{pim} with an intermittent fault f_{pim} .

those that have no redundancy if any additional equation is removed from the set. These sets are referred to as minimally structurally over-determined (MSO) sets of equations and [24] describes an algorithm that finds all MSO sets for a given structural model. Other methods for finding candidate sets of redundancy equations are described in, for example, [25] and [35].

One complicating factor is that the number of redundancy equation sets grows exponentially with the level of redundancy of a system. For the given engine model, the number of MSO sets is 4496, but by adding two additional sensors to the model, the number of MSO sets can increase to approximately 100 000 [22]. Based on an MSO set it is possible to design several residual generators. If the model is non-linear, the different residual generators are likely to have different signal to noise ratios, even though they are generated from the same equation set. Thus, since there commonly are thousands of residual generator candidates, residual selection is a non-trivial problem even for systems with relatively low redundancy.

The residual selection algorithm and test design approach described in this paper is generic and independent on the method used for finding residual generator candidates. However, to illustrate the overall approach, the Fault Diagnosis Toolbox [15] is here used to generate a set of sequential residual generator candidates. A sequential residual generator is a sequential computation form of a residual generator based on an MSO set [42]. One equation is selected as the redundant equation, and the remaining set of equations is ordered such that all unknown variables can be computed sequentially based on known signals. This is illustrated by the following example.

Example 2: Note that the nominal model in Example 1 is an MSO set which can be used for residual generation. A sequential residual generator can be formulated in the following computational form

$$\begin{aligned} e_2 : & \quad x_2 := g_2(u) \\ e_1 : & \quad x_1 := g_1(x_2) \\ e_3 : & \quad r := y - x_1 \end{aligned}$$

where the nominal equations are solved in the following order, e_2, e_1, e_3 . The last equation e_3 is the redundant equation used for computing the residual r . Note that since equation e_1

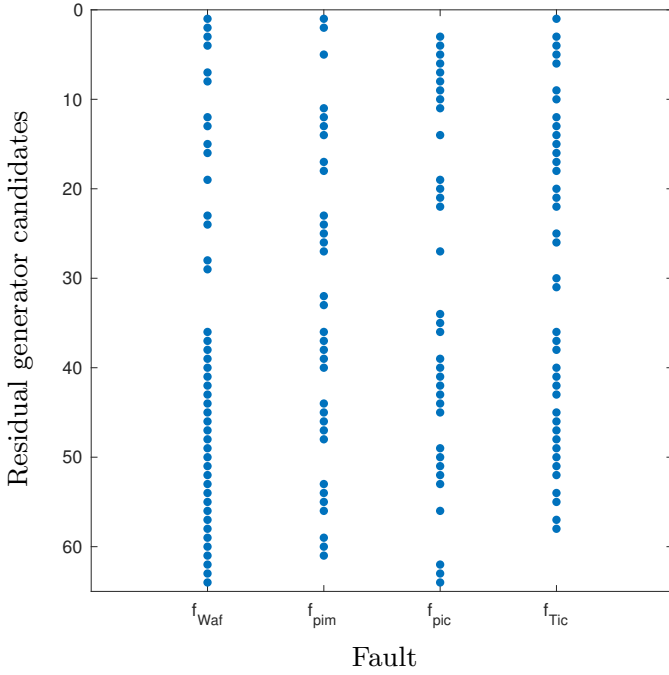


Fig. 6. Fault signature matrix for all residual generator candidates.

is used, where fault f is included, the sequential residual generator is sensitive to the fault f . \square

By selecting different redundant equations, each MSO set can be used to generate different sequential residual generators with different diagnostic performance. Each sequential residual generator is automatically generated from the Fault Diagnosis Toolbox and implemented as c-code. Some of the sequential residual generator candidates are not realizable, for example, if a variable is to be computed from an equation that is not invertible.

In this specific case, a set of 64 residual generator candidates is generated and the corresponding fault signature matrix is shown in Fig. 6. Since some residual generators are based on the same MSO set, they are sensitive to the same set of faults. All dynamic equations in the used residual generators are computed by integration, i.e., no differentiation is carried out in the computation of the residuals [14].

Evaluating one residual using approximately 13 min data (FTP75 highway driving cycle) with sampling rate 1 kHz takes around 0.8 s on a standard laptop. The residuals are evaluated using data from the different fault scenarios described above with intermittent faults. Fig. 1 shows data including intermittent fault f_{Waf} from the subset of residuals that are sensitive to f_{Waf} , and the residuals not sensitive to the fault in Fig. 7. Each data set is divided into training data and validation data, which is illustrated by the vertical black line in the figures. The corresponding residuals with fault f_{pim} are shown in Fig. 8 and Fig. 9, respectively. It is visible that the residual outputs where the fault is decoupled, i.e. residuals not expected to react to the fault, do not change significantly. This is expected, however, note that only a few of the residuals sensitive to a specific fault significantly reacts to the fault. This indicates that measurement data is needed in the design process of selecting

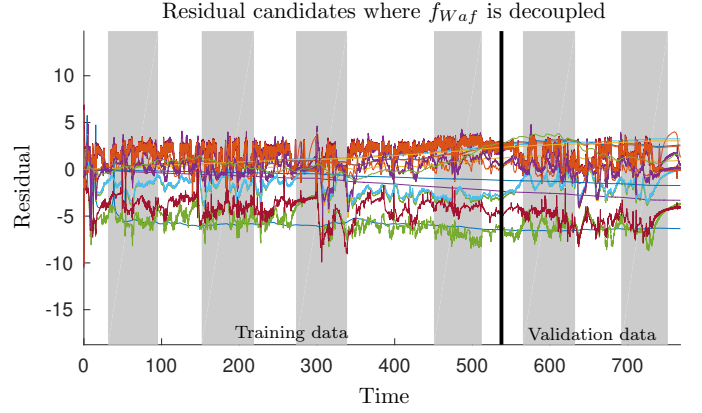


Fig. 7. The residual outputs from the subset of residual generators where fault f_{Waf} is decoupled, when evaluated using data with fault f_{Waf} .

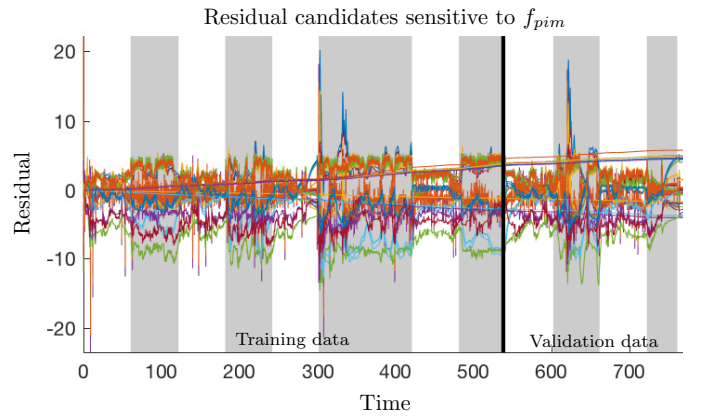


Fig. 8. The residual outputs from the subset of residual generators sensitive to f_{pim} when evaluated using data with fault f_{pim} .

the residual generators to be used in the diagnosis system.

V. RESIDUAL SELECTION

If the performance of each residual is considered independently of each other, the problem of finding a residual set that fulfills a set of performance requirements can be formulated as a minimal hitting set problem [42]. One complicating factor is

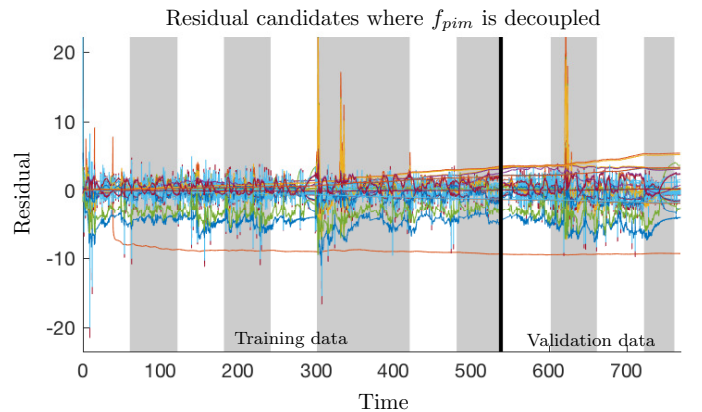


Fig. 9. The residual outputs from the subset of residual generators where f_{pim} is decoupled when evaluated using data with fault f_{pim} .

that solving the minimal hitting set problem is NP-complete, and thereby it is not feasible to find the minimal solution even for relatively small systems. The two common approaches are either to apply a heuristic search strategy to the residual selection process or try to relax the optimization problem to a form that is easier to solve but still gives relevant results. The approach proposed here is to reformulate the residual selection problem as a relaxed convex optimization problem.

To take quantitative detection performance of the candidate residual generators into consideration, a data-driven approach is proposed for the residual selection problem. Since the available data from faulty cases is limited, the fault signature matrix in Fig. 6 is used to find residual generator candidates with specific fault isolability properties. The residual selection problem is formulated as a convex optimization problem which guarantees that any local optimum is also global. Each fault isolability requirement is solved individually and a number of different candidate residual sets are computed with varying trade offs between solution cardinality and detection performance. The best residual set is then selected using cross-validation.

A. Data-driven residual selection

The problem of finding a subset of residuals that achieves satisfactory detection performance is similar to a research problem in machine learning usually referred to as the feature selection problem [18]. There are a couple of reasons why it is relevant to apply a feature selection algorithm instead of using all available residual generator candidates. Two of the most important factors from a residual selection perspective are

- Robustness against overfitting
- Computational cost

Overfitting is a general problem in feature selection when a model becomes dependent of artifacts in the training data and is not able to make reliable predictions on validation data. In this case, it is important to find a set of residual generators that can distinguish faulty behavior from model uncertainties and measurement noise. The second aspect of reducing the number of residuals is that on-line computational cost is reduced if only a small set of residual generators are computed in the diagnosis system.

Since the purpose of the diagnosis system is to detect and isolate a set of different faults, one option is to use data from all faults at once and train one multi-class data-driven classifier, see for example [17]. However, if the amount of training data only contains a limited set of fault realizations from the different faults, it is possible that the classifier incorrectly rejects a fault if it occurs with a different fault realization compared to training data. Thus, it is not desirable that the performance of a test quantity is too dependent on available training data.

Evaluating fault detection performance of a set of residual generators corresponds to evaluating their ability to distinguish faulty behavior from nominal behavior. Here, the residual selection problem is formulated independently for each fault isolation performance requirement, i.e., the goal is to find

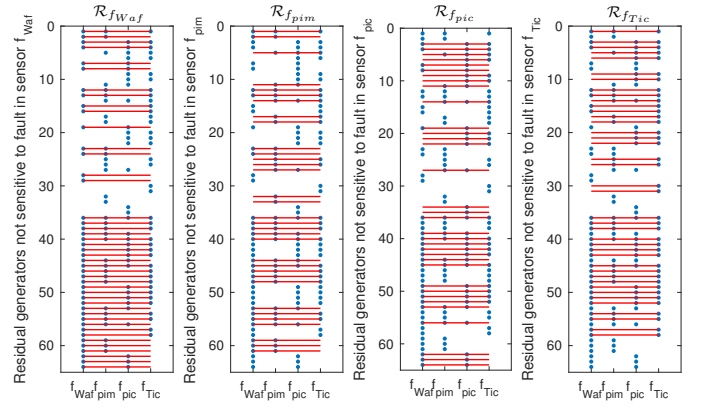


Fig. 10. Sets of remaining residual generator candidates \mathcal{R}_{f_j} when decoupling one fault at the time.

different residual sets for isolating each fault $f_i \in \mathcal{F}$ from another fault $f_j \in \mathcal{F}$ where $f_i \neq f_j$. Let $\mathcal{R}_{f_j} \subseteq \mathcal{R}_{\text{all}}$ denote the subset of residual generators where the fault f_j is decoupled. These subsets of residuals are used in the residual selection algorithm to select residuals with specific isolability properties. The corresponding subsets \mathcal{R}_{f_j} , when the residual generators sensitive to each of the four faults are removed, are shown in Fig. 10.

B. A convex relaxation of the residual selection problem

Let $\bar{r}[t]$ be a column vector representing a sample at time t from the residual generator candidates $\mathcal{R} \subseteq \mathcal{R}_{\text{all}}$ from the model analysis. It is assumed that all residuals in \mathcal{R}_{all} are normalized to have the same noise variance in the nominal case. Let $\beta_0 + \beta^T \bar{r}[t]$ be an affine function of vector \bar{r} such that the sample $\bar{r}[t]$ belongs to Class 0 if $\beta_0 + \beta^T \bar{r}[t] \geq 0$ and Class 1 if $\beta_0 + \beta^T \bar{r}[t] < 0$. The vector β is a column vector with the same number of elements as the number of elements in \mathcal{R} where element β_m in β corresponds to residual generator $r_m \in \mathcal{R}$. The parameter β_0 can be interpreted as a threshold that divides the two classes of data.

There are different methods of quantifying residual detection performance, such as the Kullback-Leibler divergence [2], [10] or power functions [42]. The approach used here is to use the logistic function [20] to evaluate fault detection performance.

The logistic function can be written as

$$\sigma(\bar{r}[t]) = \frac{1}{1 + e^{-(\beta_0 + \beta^T \bar{r}[t])}} \quad (1)$$

which maps any real value \mathbb{R} , to the interval $[0, 1]$. The logistic function can be used to model a probabilistic binary classifier as

$$\begin{aligned} P(\text{Class} = 1 | \bar{r}[t]; \beta_0, \beta) &= \frac{e^{\beta_0 + \beta^T \bar{r}[t]}}{1 + e^{\beta_0 + \beta^T \bar{r}[t]}}, \\ P(\text{Class} = 0 | \bar{r}[t]; \beta_0, \beta) &= \frac{1}{1 + e^{\beta_0 + \beta^T \bar{r}[t]}} \end{aligned} \quad (2)$$

which is called *logistic regression* where $P(\text{Class} = i | \bar{r}; \beta_0, \beta)$ denotes the conditional probability that Class = i given sample \bar{r} and parameters β_0 and β .

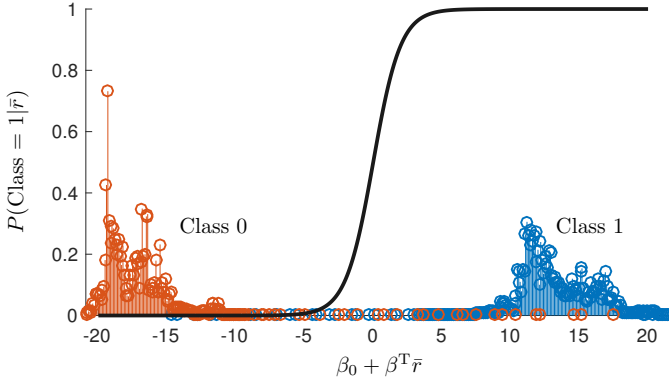


Fig. 11. A logistic regression model.

An example of a logistic regression model is shown in Fig. 11 where the two histograms represent two classes of data and the curve is the probability that data belongs to the right class and one minus the curve that data belongs to the left class. Here, Class 0 represents the fault-free case and Class 1 the faulty case.

Let $\bar{r}[1], \bar{r}[2], \dots, \bar{r}[N]$ denote N samples and $c[t]$ is a class variable which is one if the sample $\bar{r}[t]$ at time t belongs to Class 1 and zero if it belongs to Class 0. The maximum likelihood estimation of the parameters β_0, β can be found by maximizing the log-likelihood

$$\begin{aligned} \max_{\beta_0, \beta} \sum_{t=1}^N \log p_c(\bar{r}[t]; \beta_0, \beta) = \\ \min_{\beta_0, \beta} - \sum_{t=1}^N \left[c[t](\beta_0 + \beta^T \bar{r}[t]) - \log(1 + e^{\beta_0 + \beta^T \bar{r}[t]}) \right] \end{aligned} \quad (3)$$

where $p_c(\bar{r}[t]; \beta_0, \beta) = P(\text{Class} = c[t] | \bar{r}[t]; \beta_0, \beta)$ [20].

An advantage of the logistic regression model is that it is convex [4] which guarantees that any locally optimal parameters β_0 and β are also globally optimal. If training data contains both fault-free data and data from fault f_i , the weights β in (3) can be interpreted as a measure of how important each candidate residual generator in \mathcal{R}_{f_j} is to detect the fault f_i in the training data set.

Since the goal is to find a subset of the best residuals in \mathcal{R}_{f_j} to detect a fault f_i , an L1 penalty to the parameter vector β is added in (3) which gives the L1-regularized logistic regression [20]. Let $\bar{r}_j[t]$ denote a column vector which is a sample of all residuals in \mathcal{R}_{f_j} , i.e., all residuals where f_j is decoupled. Then, the L1-regularized logistic regression can be written as

$$\begin{aligned} \min_{\beta_0, \beta} \left\{ - \sum_{t=1}^N \left[c[t](\beta_0 + \beta^T \bar{r}_j[t]) - \log(1 + e^{\beta_0 + \beta^T \bar{r}_j[t]}) \right] \right. \\ \left. + \lambda \sum_{\forall r_k \in \bar{r}_j} |\beta_k| \right\}. \end{aligned} \quad (4)$$

Thus, by performing residual selection using training data including fault f_i on the subset of residuals \mathcal{R}_{f_j} where f_j is decoupled, the solution will be a residual set which can isolate f_i from f_j . This is an important step in the residual selection

approach that the structural information is used to assure that the solution set to (4) has certain isolability properties. By solving (4) for different faults f_i and f_j , different solution sets can be found with different fault isolability properties. Note that the final residual set in the diagnosis system is the union of all selected solution sets to (4) for different $f_i, f_j \in \mathcal{F}$.

A residual r_k is considered part of the solution set to (4) if the corresponding parameter $\beta_k \neq 0$. The parameter λ is a penalty parameter forcing sparsity to the solution β [20]. A large value of λ will result in a solution with few non-zero elements in β corresponding to most important residuals to detect f_i . By decreasing the value of λ , more elements in β will become non-zero. However, it is non-trivial to select λ that achieves a solution that gives the best trade-off between detection performance and number of residuals.

C. Identify candidate sets using regularization paths

Solving (4) for different values of λ will give different solution sets with different trade-offs between detection performance and number of residual generators. In [9], an algorithm is proposed that efficiently finds the regularization path of the β vector for linear models, i.e., how the solution β depends on λ . In [13] an algorithm for generalized linear models, including logistic regression, is presented and the analysis is performed using the implementation in the GLMnet toolbox in Matlab [36].

The regularization path of each element in the vector β based on data from fault f_{Waf} is shown in Fig. 12 and from fault f_{pim} in Fig. 13. For each value of λ where one of the elements in β goes from zero to non-zero, or vice versa, is marked in the figure. Note that in each interval between the vertical lines, the set of non-zero elements in β is the same. Then, each interval represents a candidate set of residuals where the leftmost interval is the candidate set with highest λ penalty. For each following interval, the corresponding candidate residual sets are ordered based on reduced sparsity penalty, i.e. with reducing λ . In most cases, the cardinality of the candidate set increases when λ decreases. However, depending on the L1 penalty cost of adding residuals to the solution set, there are cases where a parameter β_i becomes zero again (while the absolute value of another non-zero parameter β_j increases) and the cardinality instead decreases.

The corresponding candidate residual generator sets for Fig. 12 and Fig. 13, i.e., the non-zero elements in β in each interval are summarized in Table II and Table III, respectively. Note that the cardinality does not increase for each candidate residual set when λ decreases. Even though the sets can be smaller, it is important to note that since the penalty λ decreases for higher candidate set indices, the risk of overfitting increases. This can be avoided by increasing the amount of training data or to use cross-validation [29].

D. Candidate residual set selection using cross-validation

To decide which residual candidate set to select, the different candidates are evaluated using cross-validation [29]. Validation data is selected as shown in Fig. 1 and Fig. 7 for f_{Waf} , and in Fig. 8 and Fig. 9 for f_{pim} , respectively. Then, a

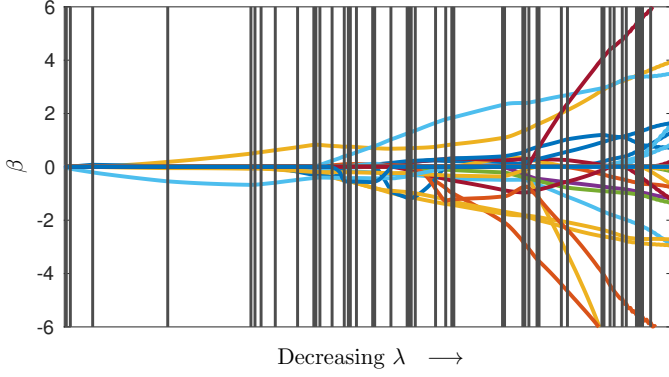


Fig. 12. The regularization path for each element in β in (4) given data from f_{Waf} . Each vertical line marks when one curve becomes either zero or non-zero, meaning that the number of non-zero elements changes in β . Each of the 48 intervals corresponds to one candidate set of residuals.

TABLE II
RESIDUAL GENERATOR CANDIDATE SETS CORRESPONDING TO THE NON-ZERO ELEMENTS OF β IN EACH INTERVAL IN FIG. 12. THE HIGHLIGHTED SET IS THE SELECTED RESIDUAL SET.

Candidate set index	Set of residual generators
1	{62}
2	{29, 62}
3	{24, 29, 45, 62}
4	{24, 45, 62}
5	{24, 38, 45, 62}
6	{24, 38, 45, 59, 62}
7	{6, 24, 38, 45, 59, 62}
8	{6, 24, 38, 45, 50, 59, 62}
9	{6, 24, 30, 38, 45, 50, 59, 62}
10	{6, 9, 24, 30, 38, 45, 50, 59, 62}
11	{6, 9, 24, 30, 38, 45, 50, 55, 59, 62}
12	{6, 9, 24, 38, 45, 50, 55, 59, 62}
13	{6, 9, 24, 38, 45, 50, 55, 59, 62, 64}
...	...
48	{1, 3, 6, 8, 10, 11, 12, 13, 15, 23, 24, 26, 34, 37, 38, 45, 49, 55, 57, 58, 63}

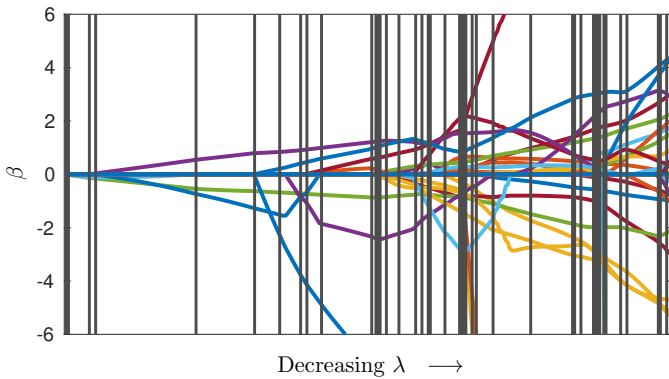


Fig. 13. The regularization path for each element in β in (4) given data from f_{pim} . Each of the 43 intervals corresponds to one candidate set of residuals.

TABLE III
RESIDUAL GENERATOR CANDIDATE SETS CORRESPONDING TO THE NON-ZERO ELEMENTS OF β IN EACH INTERVAL IN FIG. 13. THE HIGHLIGHTED SET IS THE SELECTED RESIDUAL SET.

Candidate set index	Set of residual generators
1	{27}
2	{26, 27}
3	{26, 27, 53}
4	{26, 27, 36, 53}
5	{26, 36, 53}
6	{26, 36, 43, 53, 57}
7	{26, 36, 39, 43, 53, 57}
8	{16, 26, 36, 39, 43, 53, 57}
9	{16, 26, 35, 36, 39, 43, 53, 57}
10	{16, 26, 35, 39, 43, 53, 57}
11	{12, 16, 26, 35, 39, 43, 53, 57}
12	{12, 16, 26, 35, 39, 43, 52, 53, 57}
13	{12, 14, 16, 26, 35, 39, 43, 52, 53, 57}
...	...
43	{1, 3, 7, 9, 10, 12, 13, 14, 15, 16, 19, 26, 28, 30, 32, 34, 35, 36, 37, 38, 41, 42, 43, 45, 52, 53, 55, 57, 60, 63}

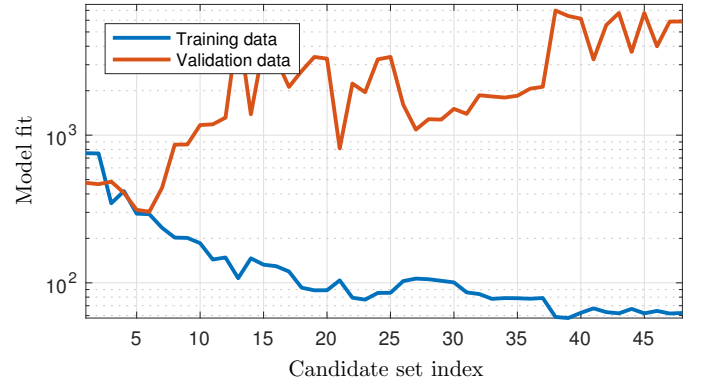


Fig. 14. Cross-validation of residual sets using data with fault f_{Waf} showing model fit with respect to residual set.

new logistic regression model (3), without any regularization term, is trained using only the residual generator candidates \bar{r} in each candidate set. The results from the cross-validation are shown in Fig. 14 for each candidate set in Table II. Evaluation of validation data shows that candidate sets after index six have over-fitting behavior. Note that the candidate sets have been selected using L1-regularization and the cost without the regularization term is not monotonically decreasing when evaluated using training data. However, the cross-validation still gives an indication of the relation between fault detection performance and number of residual generators.

For easier interpretation of the cross-validation, from a fault detection point of view, the mis-classification rate of the logistic regression model for each residual set is computed for both the training set and validation set and the results are shown in Fig. 15 and Fig. 16, for detection of f_{Waf} and f_{pim} respectively. This is computed by counting the number of mis-classified samples in the solution to (3). For each fault in Fig. 15 and Fig. 16 respectively, the mis-classification rate starts to increase for validation data after around candidate set five which indicates that those models suffer from over-fitting.

In this case study, since the difference in performance

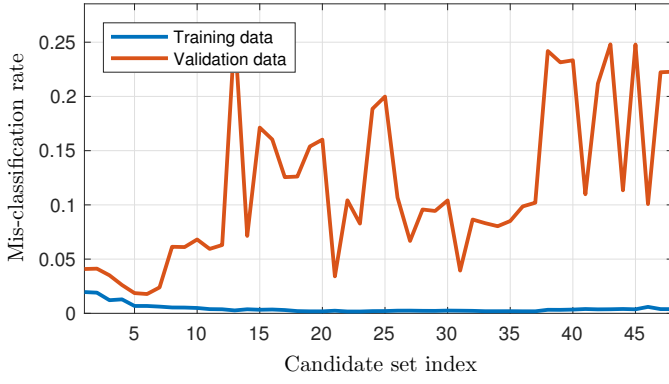


Fig. 15. Cross-validation of residual sets using data with fault f_{Waf} showing mis-classification rate with respect to residual set.

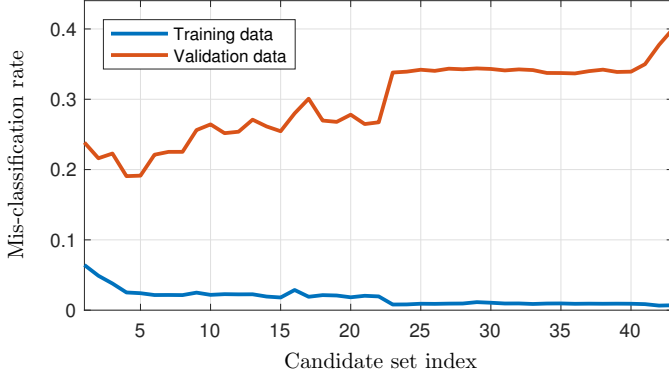


Fig. 16. Cross-validation of residual sets using data with fault f_{pim} showing mis-classification rate with respect to residual set.

is relatively small for the different candidate sets with low cardinality, candidate sets with either one or two residuals are selected for each fault detection and isolation case.

The residual selection algorithm is applied for each fault isolation case, including fault detection and single-fault isolation. The results from the residual selection are summarized in Table IV where the set in position (i, j) is the selected set of residual generators to isolate fault f_i from f_j . In total, seven residuals were selected, $\mathcal{R}^* = \{r_{24}, r_{26}, r_{27}, r_{29}, r_{30}, r_{34}, r_{62}\}$, that together achieve satisfactory detection and isolation performance of all considered faults. The fault signature matrix for \mathcal{R}^* is shown in Table V.

To illustrate fault detection performance of the selected residual sets in Table IV, validation data from two different sets of two residuals are plotted against each other. Fig. 17 and Fig. 18 show the residual sets $\{r_{29}, r_{62}\}$ and $\{r_{26}, r_{27}\}$ which are selected to detect sensor faults f_{Waf} and f_{pim} , respectively, see Table IV. The two figures show the correlation

TABLE IV
SELECTED RESIDUAL SETS FOR EACH CASE OF DETECTING OR ISOLATING EACH FAULT.

	NF	f_{Waf}	f_{pim}	f_{pic}	f_{Tic}
f_{Waf}	$\{r_{29}, r_{62}\}$	-	$\{r_{29}, r_{62}\}$	$\{r_{24}, r_{29}\}$	$\{r_{29}, r_{62}\}$
f_{pim}	$\{r_{26}, r_{27}\}$	$\{r_{26}, r_{27}\}$	-	r_{26}	r_{27}
f_{pic}	r_{34}	r_{34}	r_{34}	-	r_{34}
f_{Tic}	$\{r_{26}, r_{30}\}$	r_{30}	$\{r_{26}, r_{30}\}$	$\{r_{26}, r_{30}\}$	-

TABLE V
FAULT SIGNATURE MATRIX OF RESIDUAL SET \mathcal{R}^* .

Residual	f_{Waf}	f_{pim}	f_{pic}	f_{Tic}
r_{24}	X	X		
r_{26}		X		X
r_{27}		X	X	
r_{29}	X			
r_{30}				X
r_{34}			X	
r_{62}	X		X	

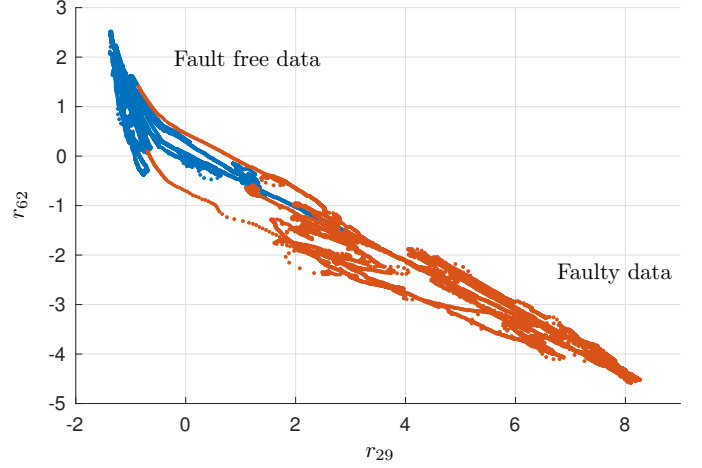


Fig. 17. Residuals r_{29} and r_{62} plotted against each other for a data set with intermittent fault f_{Waf} .

between the residual generators which illustrates that detection of the two faults is improved by analyzing the multi-variate information instead of analyzing each residual individually.

E. Summing up

The data-driven residual selection strategy to find a residual set to isolate a fault f_i from a fault f_j is summarized as follows.

- 1) Evaluate the selected residual generators $\mathcal{R}_{f_j} \subseteq \mathcal{R}_{all}$ where f_j is decoupled using data including both nominal

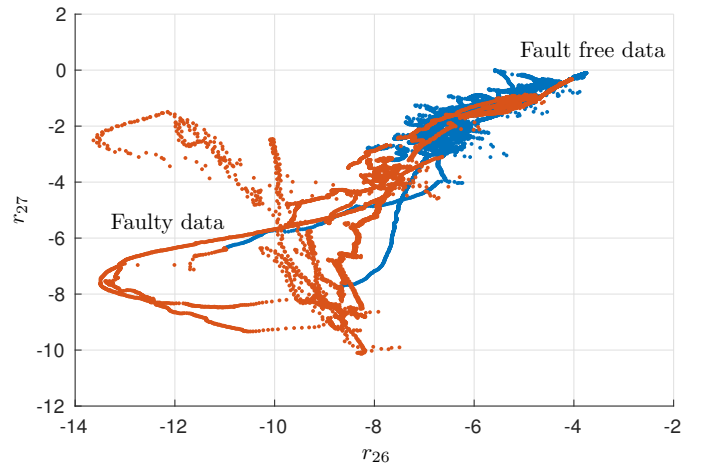


Fig. 18. Residuals r_{26} and r_{27} against each other for a data set with intermittent fault f_{pim} .

data and fault f_i .

- 2) Compute regularization paths of (4), using for example [36], to get candidate residual generator sets.
- 3) Evaluate performance using cross-validation for all candidate sets and select the smallest set with satisfactory performance for validation data.

This procedure is performed for all fault pairs (f_i, f_j) and the final residual set \mathcal{R}^* is the union of the solution sets selected for each fault isolation requirement.

VI. A DATA-DRIVEN VALIDATION OF STRUCTURAL RESIDUAL PERFORMANCE

The set of residuals in Table V is selected to detect and isolate a set of faults with satisfactory detection and isolation performance. This is based on the assumption that they fulfill the structural detectability and isolability performance requirements specified in the fault signature matrix. Here, a data-driven approach is proposed to analyze if the fault isolability properties are fulfilled in the situation where the residual generators are affected by model uncertainties and measurement noise.

Even though faults are ideally decoupled in each residual generator, model uncertainties could be significant. This could result in faults influencing residual generators even though they should not. If this is true, the residuals will cause false alarms and increase the risk of falsely rejecting the true fault hypothesis. If the model uncertainties cannot be neglected, it is relevant to verify that the faults are correctly decoupled in the different residual generators, i.e., the fault signature matrix is correct. The approach here is to evaluate all residuals with data from the different faults and analyze whether the output from the residual set deviates from its nominal behavior or not when each fault is decoupled.

A common approach to visualize fault detection and isolation performance of a set of residuals is to draw each residual in a separate plot to compare the residual distribution in the nominal and the faulty case. This can be evaluated using, for example, the Receiver Operating Characteristic (ROC) curve [3] or power functions [42]. One limitation with applying these methods is that important information regarding correlation between residuals is lost. Thus, it is relevant to plot residual outputs in a way that can visualize this type of multi-variate information.

A. t-Student Stochastic Nearest Embedding

Visualizing multi-dimensional data is a difficult task for dimensions larger than three. An interesting unsupervised non-linear visualization method is the t-Student Stochastic Nearest Embedding (t-SNE) algorithm [46], which can transform multi-dimensional data to low-dimensional space. The algorithm tries to preserve local structures in data using the Kullback-Leibler divergence as a similarity measure. Thus, samples that are similar in the original space are kept close to each other. A fast heuristic implementation of the t-SNE algorithm to handle larger data sets is used here [45].

The results of the t-SNE algorithm when analyzing residual time data, including all four sensor fault scenarios, from all

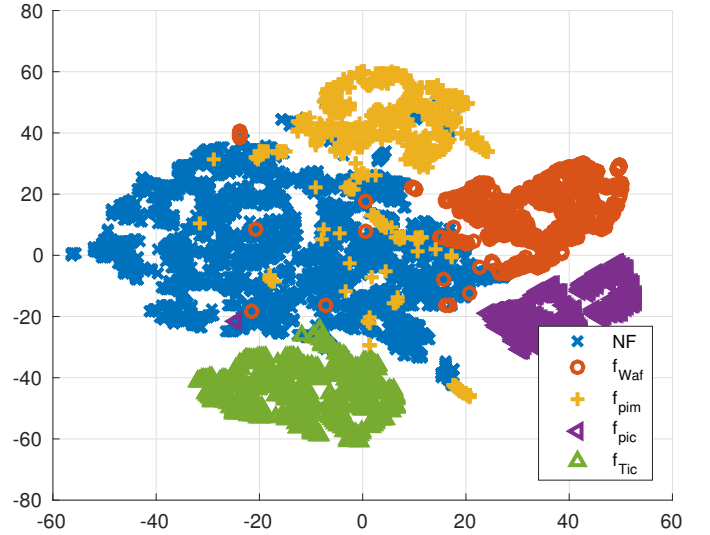


Fig. 19. t-SNE plot analyzing faulty data from the residual set in Table IV.

seven residuals \mathcal{R}^* at the same time are shown in Fig. 19 for the residual set in Table V. To decrease the size of the figure, the information is down-sampled to every 400th sample. Each point is a time sample of the residual set evaluated with real engine data and the different colors represent different fault modes, i.e., the fault that is present in each sample. Note that the dimensionality reduction is performed using non-linear optimization, meaning that the generated plot will look different in each run [46].

The t-SNE plot in Fig. 19 is interpreted as follows. If data from two classes are on top of each other, these samples are overlapping in the original space as well. This means that they are difficult to distinguish from each other, i.e., the residual set outputs are similar. If there are data from two classes that do not lie on top of each other, i.e., they are located at different areas in the t-SNE figure, these samples of data are distinguishable from each other. It is visible that some samples from fault f_{pim} are located among the fault-free data in Fig. 19. This is expected since the results from the residual selection showed that this fault is more difficult to detect. However, most of the faulty data are located in a different location compared to the fault-free case showing that the fault is detectable as concluded in the residual selection.

As discussed in the introduction in Section I, note that since residual data contain a limited amount of fault realizations it is only possible to evaluate if a fault is detectable, for a given residual set, in the t-SNE plot. This means that if there are regions of faulty data that are not overlapping with nominal data, the fault should be detectable. Even though data from the different faults are located in different parts of Fig. 19, it is not possible to state that the faults are isolable from each other, since the faulty data is not representative of all fault realizations. To evaluate fault isolation properties using t-SNE, the structural information in the fault signature matrix is used in the analysis.

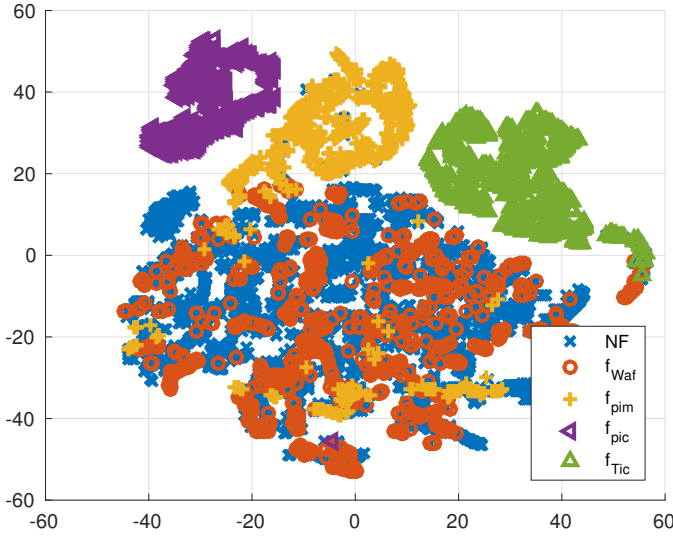


Fig. 20. t-SNE plot visualizing data from the subset of residuals in Table IV where f_{Waf} is decoupled.

B. Analyzing fault isolation properties using t-SNE

The fault signature matrix shows the ideal fault sensitivity of each residual. If a fault is decoupled, then faulty data should not be different from nominal data. By selecting the subset of residuals where a fault f_j is decoupled and analyze the output from the subset using the t-SNE algorithm, then residual data from fault f_j should not differ from nominal residual data if the fault is ideally decoupled. Then, if the other faults are detectable when f_j is decoupled, it is possible to isolate the other faults from f_j . If data from f_j is separated from nominal data, then it is an indication that the fault is not correctly decoupled and the model must be improved.

Using the fault signature matrix makes it possible to evaluate fault isolation properties using t-SNE on the residual set in the same way as for fault detection in Fig. 19. As an example, Fig. 20 shows the t-SNE plot of residuals $\{r_{26}, r_{27}, r_{30}, r_{34}\}$, i.e., the residuals where f_{Waf} is decoupled. It is visible in the plot that f_{Waf} is decoupled since data from the fault lies on top of fault-free data. Since the other faults are not overlapping the nominal data they are still detectable using only the subset $\{r_{26}, r_{27}, r_{30}, r_{34}\}$, i.e., they are isolable from f_{Waf} . Similar conclusions can be drawn when analyzing the residuals $\{r_{29}, r_{30}, r_{34}, r_{64}\}$ where f_{pim} is decoupled since residual data during fault f_{pim} is now on top of fault-free data in Fig. 21.

Note that evaluating the residual set using t-SNE is not really part of the residual selection problem. However, it is an important step to validate fault detectability and isolability properties of the residual generators in cases when model uncertainties cannot be neglected. The procedure to evaluate the properties of isolating fault f_i from another fault f_j for the selected residual set \mathcal{R}^* can be summarized in the following steps.

- 1) Select the subset of residuals in \mathcal{R}^* where f_j is decoupled and evaluate the residuals using data including both nominal data and fault f_j .

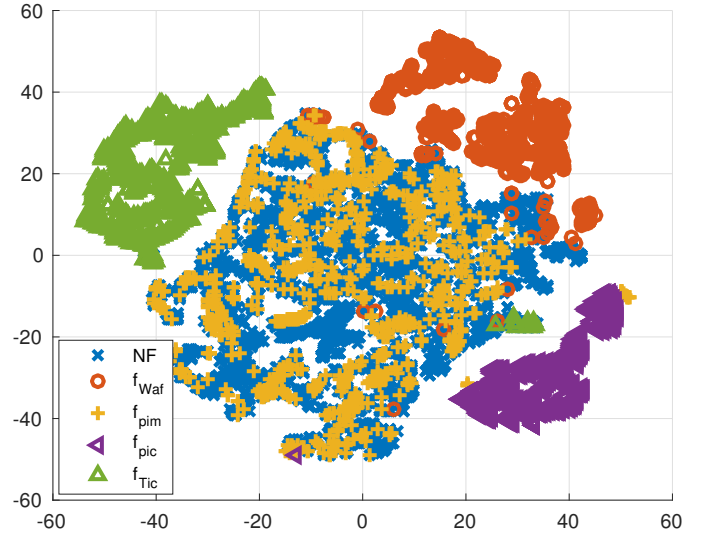


Fig. 21. t-SNE plot visualizing data from the subset of residuals in Table IV where f_{pim} is decoupled.

- 2) Apply the t-SNE algorithm [46] to visualize the residual data. The fault f_i is isolable from f_j if there are sets of data points from fault f_i which are separated from nominal data.

One of the main results from this analysis is that the fault detection and isolation performance of the set of selected residuals is consistent with the fault signature matrix in Table V.

VII. DIAGNOSIS SYSTEM DESIGN

Section V described how the residual selection problem can be relaxed and formulated as a convex optimization problem. Also, for the engine case study, a set of residual generators with satisfactory fault isolation performance was chosen for each fault pair. Note that Fig. 15 and Fig. 16, show that it is possible to improve fault detection and isolation performance by taking multi-variate information from several residuals into consideration instead of only considering the detection performance of the residuals individually. Here, a modified model-based diagnosis system design is proposed to take the multi-variate information of the residual set into consideration. The case study will be used to illustrate the proposed diagnosis system design.

A. Diagnosis system design

A traditional model-based diagnosis system structure is to post-process each residual $r_l \in \mathcal{R}^*$ independently by forming a test quantity $T_l(r_l)$, for example, by using a CUMulative SUM (CUSUM) test quantity [3], [32]. Then, given the test quantities that have triggered, i.e. the test quantities T_l that have exceeded their thresholds J_l , a set of diagnosis candidates is computed, for example, using the fault isolation algorithm described in [8].

The proposed modification of the traditional diagnosis system structure is that the test quantities T_l can be functions of multiple residuals. As an example, from the selected residual

generator sets in Table IV, a test quantity is designed for detecting and isolating each fault. The same residual set is sometimes selected for isolating different faults from each other, for example $\{r_{26}, r_{27}\}$ which has been selected for both detecting fault f_{pim} and isolating it from f_{Waf} . The total number of unique residual generator sets in this case is eight, one more than the number of residual generators in the solution.

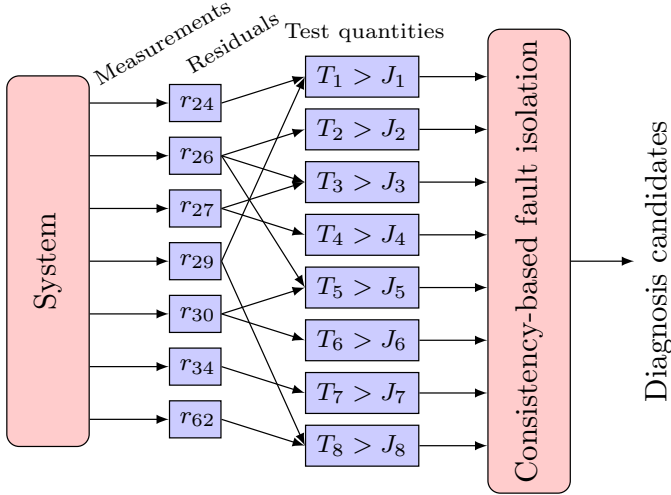


Fig. 22. A schematic of the diagnosis system design where the test quantities are based on residual sets given in Table IV. A test quantity T_i has here reacted when $T_i > J_i$ where J_i is a design parameter

The results from the cross-validation in Section V-D showed that the detection performance can be improved by combining multiple residuals when designing test quantities. The proposed diagnosis system structure is illustrated in Fig. 22. The sensor data from the system are first processed by the selected set of residual generators in Table V and then a set of test quantities are computed based on the different subsets of residuals. The residual subsets are given by the solution sets in Table IV. However, note that the approach is generic and that it is not necessary that the test quantities are designed based on logistic regression.

The fault signature matrix of the designed test quantities in Fig. 22 is shown in Table VI. The fault sensitivity of each test quantity in Table VI is given by the union of the fault sensitivities in Table V for the residuals used in the test quantity. Thus, fault isolation can still be performed based on the multi-variate set of test quantities using existing model-based fault isolation algorithms, for example, consistency-based diagnosis [8]. Furthermore, note that even though a residual is used in several test quantities, it is only necessary to compute it once every time step.

B. Multi-variate test quantity design

As discussed in Section I-A, since the amount of faulty scenarios in the training data is limited, it is not a good approach to train a binary or multi-class classifier using training data without taking the fault sensitivity of the different residual generators into consideration. Since the residual sets in Table IV have good detection performance on training data,

TABLE VI
FAULT SIGNATURE MATRIX OF THE TEST QUANTITIES IN FIG. 22.

Test quantity	f_{Waf}	f_{pim}	f_{pic}	f_{Tic}
T_1	X	X		
T_2		X		X
T_3		X	X	X
T_4		X	X	
T_5		X		X
T_6				X
T_7			X	
T_8	X		X	

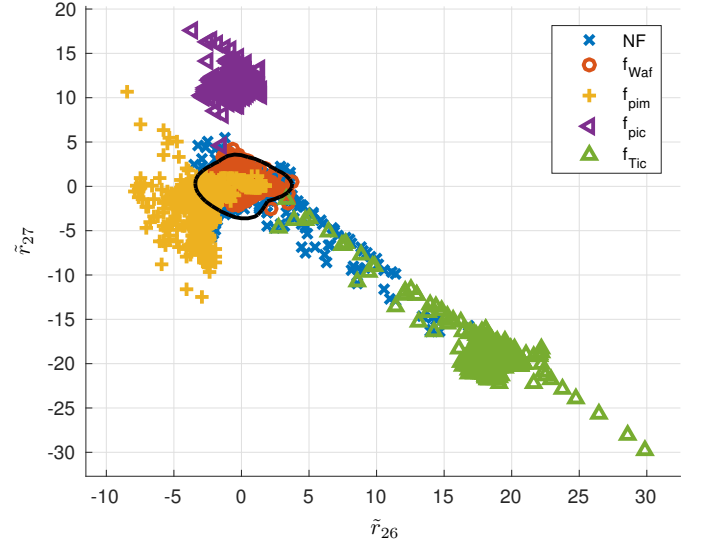


Fig. 23. Residuals r_{26} and r_{27} evaluated with data from different faults are plotted against each other. The two residuals are normalized to have identity covariance matrix and are denoted \tilde{r}_{26} and \tilde{r}_{27} . The black curve represent the boundary of a one-class support vector machine trained on fault-free data to have 1% false alarm rate.

it is assumed that they should have good detection performance, i.e., they will deviate from their nominal behavior, on other realizations of the same faults as well. Therefore, using only nominal data to calibrate the threshold J_i of each test quantity T_i should be satisfactory to assure high diagnostic performance.

As an example, consider the design of test quantity T_3 in Fig. 22 which is a function of residual generators r_{26} and r_{27} . Fig. 23 shows residuals r_{26} and r_{27} plotted against each other with data from different faults. Since the residual generators in this case study are computed without feedback, there is a bias between the different residuals caused by incorrect initial state values. Therefore, the residual outputs for the different fault data sets in the figure are normalized to have the same mean when there is no fault in the data. The two residuals are also normalized to have identity covariance matrix, and the normalized residuals are denoted \tilde{r}_{26} and \tilde{r}_{27} . The fault f_{Waf} is decoupled, which is visible as the data from the fault lies on top of the fault-free data. The three other faults affect the two residuals in different directions.

For linear systems with additive faults, each fault is projected in a specific direction in the linear residual space. Then, the optimal linear residual generator to detect a specific fault is the one corresponding to the vector pointing in the same

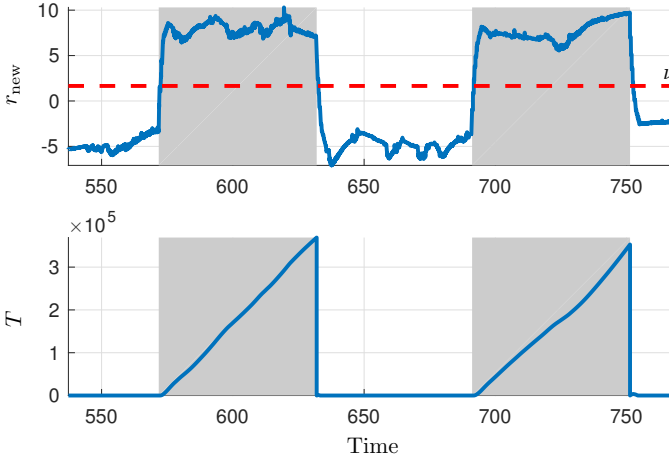


Fig. 24. An example of a CUSUM test applied to a linear combination of r_{26} and r_{27} applied to data including an intermittent fault f_{Tic} . Note that the CUSUM test is manually reset after each intermittent fault disappears.

direction as the fault [10]. Since each fault appears to move the residual outputs in a specific direction, one approach is to generate different test quantities based on a linear combination of the original set of residuals such that it maximizes the detection of each fault, i.e.,

$$r_{\text{new}}[t] = \beta^T r[t]. \quad (5)$$

where β can be determined using, for example, logistic regression (3). Then, for example, a CUSUM test can be applied to the new residual $r_{\text{new}}[t]$

$$T[t] = \max(0, T[t-1] + r_{\text{new}}[t] - \nu), \quad T[0] = 0 \quad (6)$$

where ν is a design parameter. As an example, Fig. 24 shows a CUSUM test applied to a residual r_{new} , optimized to detect f_{Tic} based on r_{26} and r_{27} using logistic regression, as

$$r_{\text{new}}[t] = 0.64r_{26}[t] + 0.38r_{27}[t] \quad (7)$$

and $\nu = 1.65$. To illustrate the concept, the CUSUM test is manually reset when the intermittent fault disappears. The new row in the fault signature matrix for each new test quantity will be the union of the fault sensitivities for the included residual generators.

A data-driven approach to train classifiers when there is mainly data from the nominal behavior and not much from faults is often referred to as anomaly, or novelty, detection [34]. Some examples of anomaly detection algorithms are Principal Component Analysis (PCA) [27], Partial Least Squares (PLS) [30], and k-means [5]. The main principle is to generate a model based on the nominal system behavior and detect when data starts to deviate. An advantage of using the residuals instead of the original measurements is that many of the system non-linearities are captured by the residual generators. This means that a less complex classifier should in many situations be sufficient to detect faults.

One interesting anomaly classifier is the one-class support vector machine [39] which uses support vectors to model one-class training data from a given set of features, in this case the residual set. In general when training one-class support vector

machines, there is no knowledge which features are the most relevant to detect anomalies and all features will be used in the model. However, if the residual selection step in Section V is performed first, an anomaly classifier can be generated based on a selected subset of residuals which is known to be relevant for detecting or isolating a given fault. Thus, the anomaly classifier is trained using only nominal data, but still detects when faults occur. An example is shown in Fig. 23 where the black curve represents the boundary of the one-class support vector machine which is trained to correspond to a 1% false alarm rate for the nominal training data. When evaluated on the validation data shown in the figure the false alarm rate is approximately 2%.

The proposed diagnosis system design in Fig. 22 allows for both classical test quantity design using, for example, CUSUM tests, but also multi-variate methods, such as one-class support vector machines. After a residual set \mathcal{R}^* has been selected, the following diagnosis system design process can be summarized in the following steps.

- 1) Implement all residual generators in \mathcal{R}^* in the diagnosis system.
- 2) For each fault pair (f_i, f_j) :
 - a) Select the subset of residuals in \mathcal{R}^* that was found to isolate fault f_i from f_j as described in Section V-E.
 - b) Design a test quantity, based on a combination of the residual subset found in a), that maximizes detection of fault f_i and still achieves a satisfactory false-alarm rate.

To illustrate the diagnosis system design shown in Fig. 22 a set of eight test quantities is generated based on the seven residual generators as described in the figure. The test quantities are evaluated using data from an intermittent fault f_{Tic} and the results are shown in Fig. 25. The dashed lines represent thresholds tuned using nominal data. The fault signature matrix in Table VI shows that test quantities T_2 , T_3 , T_5 , and T_6 , are sensitive to the fault, which is also visible in the figure. Also, note that the test quantities where f_{Tic} is decoupled are not affected by the fault. This shows that the selected set of residual generators, and the generated test quantities, work as expected and are able to detect and isolate the fault.

VIII. CONCLUSIONS

Finding a suitable set of residual generators to design a diagnosis system is crucial to be able to achieve satisfactory fault detection and isolation performance. A residual selection algorithm is proposed which combines model-based and data-driven methods to find residuals with good fault detection performance, where the set of residuals also fulfill certain isolability properties. A main contribution is that structural information, describing which faults affect which residuals, is combined with training data to identify residual sets for fault detection and isolation, even though training data is limited. The engine case study shows the importance of taking residual detection performance into consideration in the residual selection process. The t-SNE algorithm is shown to be a useful tool to analyze if faults are correctly decoupled. This

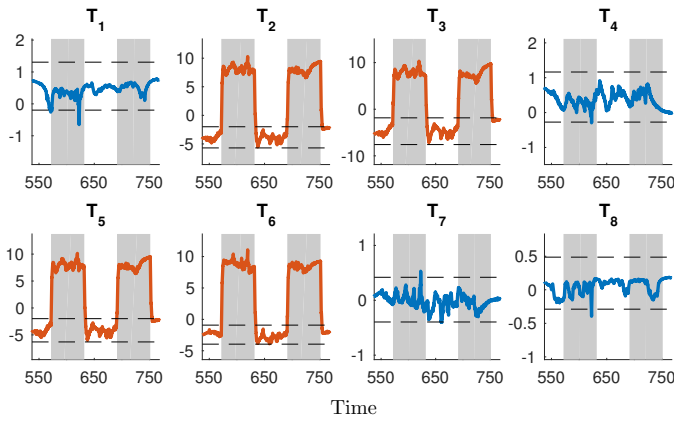


Fig. 25. A set of test quantities, designed as described in Fig. 25, when evaluated with data including an intermittent fault f_{Tic} .

is important, for example, if a simple model structure is used for a complex system where the model uncertainties cannot be neglected. A proposed model-based diagnosis system design uses multi-variate information from several residuals, instead of evaluating each residual independently, to improve fault detection performance.

ACKNOWLEDGMENT

The research has been funded by Volvo Car Corporation in Gothenburg, Sweden.

REFERENCES

- [1] J. Armengol Llobet, A. Bregon, T. Escobet, E. Gelso, M. Krysander, M. Nyberg, X. Olive, B. Pulido, and L. Trave-Massuyes. Minimal structurally overdetermined sets for residual A comparison of alternative approaches. In *Proceedings of IFAC Safeprocess'09*, Barcelona, Spain, 2009.
- [2] M. Basseville. On fault detectability and isolability. *European Journal of Control*, 7(6):625–637, 2001.
- [3] M. Basseville, I. Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.
- [4] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [5] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.
- [6] M. Cordier, P. Dague, F. Lévy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès. 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, 2004.
- [7] X. Dai and Z. Gao. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *Industrial Informatics, IEEE Transactions on*, 9(4):2226–2238, Nov 2013.
- [8] J. De Kleer and B. Williams. Diagnosing multiple faults. *Artificial intelligence*, 32(1):97–130, 1987.
- [9] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.
- [10] D. Eriksson, E. Frisk, and M. Krysander. A method for quantitative fault diagnosability analysis of stochastic linear descriptor models. *Automatica*, 49(6):1591–1600, 2013.
- [11] L. Eriksson. Modeling and control of turbocharged SI and DI engines. *OGST-Revue de l'IFP*, 62(4):523–538, 2007.
- [12] L. Eriksson, S. Frei, C. Onder, and L. Guzzella. Control and optimization of turbo charged spark ignited engines. In *IFAC world congress*, 2002.
- [13] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.
- [14] E. Frisk, A. Bregon, J. Åslund, M. Krysander, B. Pulido, and G. Biswas. Diagnosability analysis considering causal interpretations for differential constraints. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 42(5):1216–1229, 2012.
- [15] E. Frisk, M. Krysander, and D. Jung. A toolbox for analysis and design of model based diagnosis systems for large scale models. In *IFAC World Congress*, Toulouse, France, 2017.
- [16] Z. Gao, C. Cecati, and S. Ding. A survey of fault diagnosis and fault-tolerant techniquespart i: Fault diagnosis with model-based and signal-based approaches. *IEEE Transactions on Industrial Electronics*, 62(6):3757–3767, 2015.
- [17] D. Gorinevsky. Fault isolation in data-driven multivariate process monitoring. *Control Systems Technology, IEEE Transactions on*, 23(5):1840–1852, 2015.
- [18] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.
- [19] L. Guzzella and A. Sciarretta. *Vehicle Populsion System, Introduction to Modelling and Optimization*. Springer Verlag, Berlin, Germany, 3rd edition, 2013.
- [20] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.
- [21] I. Hwang, S. Kim, Y. Kim, and C. Seah. A survey of fault detection, isolation, and reconfiguration methods. *Control Systems Technology, IEEE Transactions on*, 18(3):636–653, 2010.
- [22] D. Jung. A generalized fault isolability matrix for improved fault diagnosability analysis. Conference on Control and Fault-Tolerant Systems (SysTol'16), Barcelona, Spain, 2016.
- [23] M. Krysander. *Design and Analysis of Diagnosis Systems Using Structural Methods*. PhD thesis, Linköpings universitet, June 2006.
- [24] M. Krysander, J. Åslund, and M. Nyberg. An efficient algorithm for finding minimal over-constrained subsystems for model-based diagnosis. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 38(1):197–206, 2008.
- [25] M. Krysander, J. Åslund, and E. Frisk. A structural algorithm for finding testable sub-models and multiple fault isolability analysis. In *21st International Workshop on Principles of Diagnosis (DX-10)*, Portland, Oregon, USA, pages 17–18, 2010.
- [26] E. Larsson, J. Åslund, E. Frisk, and L. Eriksson. Gas turbine modeling for diagnosis and control. *Journal of engineering for gas turbines and power*, 136(7):071601, 2014.
- [27] S. Li and J. Wen. A model-based fault detection and diagnostic methodology based on pca method and wavelet transform. *Energy and Buildings*, 68:63–71, 2014.
- [28] Z. Liu, Q. Ahmed, J. Zhang, G. Rizzoni, and H. He. Structural analysis based sensors fault detection and isolation of cylindrical lithium-ion batteries in automotive applications. *Control Engineering Practice*, 52:46–58, 2016.
- [29] L. Ljung. System identification. In *Signal Analysis and Prediction*, pages 163–173. Springer, 1998.
- [30] Riccardo Muradore and Paolo Fiorini. A pls-based statistical approach for fault detection and isolation of robotic manipulators. *IEEE Transactions on Industrial Electronics*, 59(8):3167–3175, 2012.
- [31] F. Nejjar, R. Sarrate, and A. Rosich. Optimal sensor placement for fuel cell system diagnosis using bilp formulation. In *Control & Automation (MED), 2010 18th Mediterranean Conference on*, pages 1296–1301. IEEE, 2010.
- [32] E.S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.
- [33] L. Perelman, W. Abbas, X. Koutsoukos, and S. Amin. Sensor placement for fault location identification in water networks: A minimum test cover approach. *Automatica*, 72:166–176, 2016.
- [34] M. Pimentel, D. Clifton, L. Clifton, and L. Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.
- [35] B. Pulido and C. González. Possible conflicts: a compilation technique for consistency-based diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(5):2192–2206, 2004.
- [36] J. Qian, T. Hastie, J. Friedman, R. Tibshirani, and N. Simon. Glmnet for matlab, 2013.
- [37] J. Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2):220 – 234, 2012.
- [38] C. Sankavaram, A. Kodali, K. Pattipati, and S. Singh. Incremental classifiers for data-driven fault diagnosis applied to automotive systems. *IEEE Access*, 3:407–419, 2015.

- [39] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, et al. Support vector method for novelty detection. In *NIPS*, volume 12, pages 582–588. Citeseer, 1999.
- [40] M. Staroswiecki and G. Comtet-Varga. Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica*, 37(5):687–699, 2001.
- [41] C. Sundström, E. Frisk, and L. Nielsen. Selecting and utilizing sequential residual generators in FDI applied to hybrid vehicles. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, 44(2):172–185, 2014.
- [42] C. Svärd and M. Nyberg. Residual generators for fault diagnosis using computation sequences with mixed causality applied to automotive systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(6):1310–1328, 2010.
- [43] C. Svärd, M. Nyberg, and E. Frisk. Realizability constrained selection of residual generators for fault diagnosis with an automotive engine application. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(6):1354–1369, 2013.
- [44] K. Tidiriri, N. Chatti, S. Verron, and T. Tiplica. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. *Annual Reviews in Control*, 42:63–81, 2016.
- [45] L. Van Der Maaten. Accelerating t-sne using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1):3221–3245, 2014.
- [46] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.
- [47] S. Yin, S. Ding, X. Xie, and H. Luo. A review on basic data-driven approaches for industrial process monitoring. *Industrial Electronics, IEEE Transactions on*, 61(11):6418–6428, 2014.
- [48] S. Yin and X. Zhu. Intelligent particle filter and its application to fault detection of nonlinear system. *IEEE Transactions on Industrial Electronics*, 62(6):3852–3861, 2015.
- [49] S. Yin, X. Zhu, and O. Kaynak. Improved pls focused on key-performance-indicator-related fault diagnosis. *IEEE Transactions on Industrial Electronics*, 62(3):1651–1658, 2015.
- [50] T. Yuan and S. Qin. Root cause diagnosis of plant-wide oscillations using granger causality. *Journal of Process Control*, 24(2):450–459, 2014.
- [51] B. Zhao, R. Skjetne, M. Blanke, and F. Dukan. Particle filter for fault diagnosis and robust navigation of underwater robot. *IEEE Transactions on Control Systems Technology*, 22(6):2399–2407, 2014.