A Study on Redundancy and Intrinsic Dimension for Data-Driven Fault Diagnosis

Daniel Jung
Linköping University, Sweden

David Axelsson
Linköping University, Sweden

Abstract

Data-driven fault diagnosis of technical systems use training data from nominal and faulty operation to train machine learning models to detect and classify faults. However, data-driven fault diagnosis is complicated by the fact that training data from faults is scarce. The fault diagnosis task is often treated as a standard classification problem. There is a need for methods to design fault detectors using only nominal data. In model-based diagnosis, the ability construct fault detectors depends on analytical redundancy properties. While analytical redundancy is a model property, it describes the diagnosability properties of the system. In this work, the connection between analytical redundancy and the distribution of observations from the system on low-dimensional manifolds in the observation space is studied. It is shown that the intrinsic dimension can be used to identify signal combinations that can be used for constructing residual generators. A data-driven design methodology is proposed where data-driven residual generators candidates are identified using the intrinsic dimension. The method is evaluated using two case studies: a simulated model of a two-tank system and data collected from a fuel injection system. The results demonstrate the ability to diagnose abnormal system behavior and reason about its cause based on selected signal combinations.

2012 ACM Subject Classification Computing methodologies

Keywords and phrases Data-driven diagnosis, intrinsic dimension, model-based diagnosis, structural methods

Digital Object Identifier 10.4230/OASIcs.DX.2024.4

Funding *Daniel Jung*: D. Jung is partially funded by the Swedish excellence center ELLIIT.

1 Introduction

The main objective of fault diagnosis of technical systems is detecting abnormal behavior and identifying its cause. Different scientific communities have approached this problem, e.g. model-based diagnosis and data-driven diagnosis. The general principle is to compare observations with model predictions to detect inconsistencies due to faults. When the abnormal behavior is significant compared to prediction inaccuracies, a fault is detected. After a fault has been detected, fault isolation is performed to compute diagnoses that are consistent with the observations. Model-based diagnosis relies on mathematical models derived from physical insights about the system to compute residuals. Data-driven diagnosis uses training data from relevant fault scenarios to learn the relationship between observations and diagnoses, e.g. using a classifier.

Having access to informative measurements from the system, e.g. sensor and actuator signals, is necessary to draw conclusions about the system's health. Model inaccuracies result in prediction errors or misclassifications which complicate fault detection and isolation. Methods are needed to systematically construct fault detectors based on the measurements. In model-based diagnosis, the concept of redundancy is central for design and analysis of diagnosis systems, see e.g. [17]. Analytical redundancy is necessary to construct residual generators, i.e., to formulate a mathematical model describing the relationship between a set

of known signals. Fault detectability and isolability properties of the model depends on what type of residual generators can be constructed [27]. If faults are not detectable or isolable, more sensors can be added to the system to improve the redundancy properties [18].

In data-driven diagnosis, the fault diagnosis problem is often treated as a general classification problem and the objective is to extract a set of features from the observations and train a model that can distinguish between different fault classes [24]. However, redundancy properties are, in general, not considered in the data-driven diagnosis system design process. If training data is representative of all fault classes, and the set of features is sufficiently rich, a data-driven classifier is expected to find a good separation between the different classes. However, if training data is not representative of the relevant fault scenarios, there is a significant risk that out-of-distribution samples, i.e. samples that deviate from operating conditions covered in training data, will result in false alarms and misclassifications [31].

Data-driven diagnosis have been used in various applications and different survey papers try to summarize the current state-of-the-art, see e.g. [1, 30, 29]. The survey paper [26] focused on machine learning for predictive maintenance. The importance of access to training data for data-driven fault diagnosis and how this affects the selection of data-driven methods is discussed in [7]. In [22], different open set recognition methods for fault diagnosis are discussed.

A common design principle in machine learning is trial and error, i.e. the engineer experimentally try to identify a good combination of features and data-driven models such that satisfactory performance is achieved. There is a need for data-driven methods using the ideas from model-based diagnosis to automate the design process. In e.g. [16], it is shown that redundancy is related to the manifold assumption, i.e. that observations from nominal system behavior are distributed on a low-dimensional manifold in observation space. The strong correlation between observations means that it is possible to predict one signal from the others which is needed to construct residual generators.

Measurement signals are often corrupted by some degree of noise which means that a consistency relation is almost satisfied in the nominal case or that observations are approximately distributed on a manifold. This means that some method is needed to evaluate if observations are likely to be distributed on a manifold. Several methods use Intrinsic Dimension (ID) [5] to estimate how many dimensions are sufficient to model the distribution of data, see e.g. [4]. In this work, intrinsic dimension is proposed to analyze redundancy properties from data and for data-driven design of diagnosis systems.

The outline of this paper is as follows. First, the problem formulation is presented in Section 2. The relationship between analytical redundancy and manifolds is discussed in Section 3. In Section 4, a data-driven diagnosis system design principle using intrinsic dimension is described. The proposed methods are evaluated in Section 5 using both a two-tank simulation model and data from a fuel injection system. The use of faulty data to design residuals that are insensitive to faults is illustrated in Section 6. Section 7 presents the conclusions from this work and proposed future works.

2 Problem formulation

The objective of this work is to investigate how redundancy can be evaluated directly from observations and how this information can be used in a data-driven diagnosis system design process. Specifically, the focus is on the relationship between redundancy and the intrinsic dimension of data, i.e., when the distribution of data can be approximated by a low-dimensional manifold in observation space. Then, a data-driven diagnosis system

design process, based on the principles of residual-based model-based design, is discussed and evaluated. Two case studies are considered: simulated data from a two-tank water tank system and real data from a fuel injection system in a heavy-duty diesel engine.

3 Estimating redundancy using intrinsic dimension

To formulate a redundancy property for data-driven diagnosis, the connection between analytical redundancy and manifolds is discussed. Then, it is described how the Intrinsic Dimension (ID) can be used to analyze the distribution of data including a brief summary of some methods used to estimate ID.

3.1 Analytical redundancy and manifolds

In residual-based diagnosis, residual generators are used to model the nominal relationship between known signals, e.g. actuators u and sensor signals y. This is done by first deriving consistency relations $0 = g(z)$, where $z = (u, y)$ from a model of the system. A consistency relation is, ideally, zero in the nominal case and deviates from zero when a fault occurs in the part of the system that is modeled by the relation [15]. The consistency relation is then used to construct a residual generator $r = g(z)$ to detect faults. A fault f_i is detectable if it is possible to construct a residual generator that is sensitive to the fault. A fault f_i is isolable from another fault f_j if it is possible to construct a residual generator that is sensitive to f_i but insensitive to f_j [15].

Different methods are used to derive consistency relations from a model, see e.g. [17]. For complex non-linear models, structural methods have been shown useful to analyze redundancy and find consistency relations. When constructing residual generators, several methods have focused on those using a minimal subset of observations, e.g. Minimally Structurally Over-determined (MSO) sets [17]. If one measurement is removed from the set, the remaining observations are no longer distributed on a low-dimensional manifold in the remaining observation subspace. Residuals constructed from MSO sets are good for fault isolation since they use a minimal set of observations. This might come at the cost of higher model inaccuracies [8].

The set of observations z that fulfill a set of consistency-relations is also called a solution manifold [23]. Thus, the set of possible observations z that can be explained by a fault-free system are modeled by the consistency relations. If the set of observations z is insensitive to a fault f_j then z will still belong to the solution manifold when the fault is present which is no longer true if the residual is sensitive to the fault [21].

A higher degree of redundancy means that more consistency relations can be constructed from the model. Assume that $g(z)$ is differentiable γ times. If $g(z) : \mathbb{R}^n \to \mathbb{R}^m$, and $\frac{\partial g}{\partial z}$ has full rank m, then the set of z that fulfill $0 = g(z)$ are distributed on a p-dimensional C^γ-manifold [23] where $p = n - m > 1$. Note that if derivatives of signals are needed to construct the consistency relations, then the derivatives are included as separate signals when defining z, see e.g. [16].

▶ **Example 1.** Consider a small system consisting of three sensors, y_1, y_2, and y_3, measuring the state x. The degree of redundancy is equal to 2 and an example of two consistency relations spanning all linear residual generators is $y_1 - y_2$ and $y_1 - y_3$. The solution manifold in the $n = 3$-dimensional observation space is given by $y_1 = y_2 = y_3$ which is a line, i.e. a 1-dimensional manifold. The number of known signals $n = 3$, consistency relations $m = 2$, and the dimension of the manifold $p = 1$ which is consistent with $p = n - m$.

For illustration, consider a linear model. The degree of redundancy of a linear model corresponds to how many linearly independent consistency relations that can be constructed, see e.g. [11]. Let

$$\dot{x} = Ax + Bu \tag{1}$$

where x is a state vector and A has full rank. Let A, and B be matrices of appropriate dimensions. For each value of u there exists values of x such that the equations are satisfied. Thus, the solution set $z = u$ is distributed in the whole observation space, which is spanned by u, i.e. the degree of redundancy is zero. Assume that sensors are added to the model as

$$y = x. \tag{2}$$

By adding sensors to the system, the dimension of the observation space n increases by one for each new signal. Also, the degree of redundancy increases by one for each new sensor, i.e. one more independent consistency relation can be constructed. The dimension of the solution manifold decreases by one with respect to n for each new consistency relation.

If a mathematical model of the system is not available, it is not possible to evaluate analytical redundancy. Instead, the idea here is to analyze when data is distributed on a lower-dimensional manifold to draw conclusions about the redundancy properties of the system. From a set of observations z, it is possible to select a subset of $\tilde{n}$ signals $\tilde{z} \subseteq z$ to formulate an observation space of dimension $\tilde{n}$. If the observations are distributed on a p-dimensional manifold then the degree of redundancy m can be computed as $m = \tilde{n} - p$. However, since signals are corrupted by noise, data dimension analysis methods are needed to estimate the dimension of the manifold. Here, the intrinsic dimension of observations will be used to approximate the dimension of the solution manifold.

3.2 Data dimension analysis

The ID essentially refers to the minimum number of variables needed to represent the underlying structure of a dataset without significant loss of information. Estimating the ID is crucial in various applications such as dimensionality reduction, data visualization, and understanding the complexity of datasets. Below is an outline of the ID estimation methods that are used in this work. The methods are implemented in the *scikit-dimension* python package, and were selected because of their favorable characteristics in the paper [4]. Moreover, After some initial testing, they were found to be the three most computationally efficient methods in this application, while still resulting in a good ID estimate.

Principal component analysis. Principal component analysis (PCA) is a commonly used linear projection method, that is searching for the best subspace projection that minimize the projection error. The main idea is to compute all eigenvalues of the data covariance matrix and based on the dominating eigenvalues, decide the ID. There exist multiple ID estimation methods that are based on PCA, e.g. methods that only return eigenvalues larger than a constant times the largest eigenvalue [12]. The authors of [6] have shown that many PCA estimation methods often result in poor estimations, especially for nonlinear and noisy data. However, there are more involved PCA based ID estimation methods that aims to improve the estimation for more complex data, such as [9].

Maximum likelihood estimation. The maximum likelihood estimation (MLE) is a statistical method used to estimate the ID of a dataset by maximizing the likelihood of the observed data under a given model. This approach focuses on the distribution of distances between neighboring points in the data to determine the ID.

One MLE-based algorithm for ID estimation was developed by Levina and Bickel [19]. This method starts by considering the k-nearest neighbors of each point in the dataset and for a given point, the distances to its nearest neighbors are used to estimate the local dimension around that point. The idea is that, in a space with a certain ID d, the distances to the nearest neighbors should follow a certain distribution related to d. An extension of this algorithm, proposed by Haro et al. [13], enhances the robustness of the original method, particularly in noisy environments.

Method of moments. Method of moments (MoM) is an estimation technique that derives the ID by using sample moments, such as mean and variance, with their theoretical counterparts. In the context of ID estimation, MoM involves using moments of distance distribution between data points. By matching the empirical moments calculated from the data with the expected moments under the assumption of a particular ID, it is possible to estimate the ID [3].

4 Design of data-driven residuals using intrinsic dimension

This section outlines a method for designing residuals based on the ID-based search of signal combinations. The process consists of four key steps: *data preprocessing*, *searching for low-dimensional signal combinations*, *selection of signal combinations*, and *design of residuals using the selected signal combinations*.

4.1 Data preprocessing

The first step is to determine which observations will be used for fault diagnosis and how it will be preprocessed. This preprocessing typically involves filtering to reduce noise, removing outliers, normalizing data, and potentially discarding corrupt or impractical data.

To model dynamic systems, derivatives are needed to form consistency relations. In many applications, the derivatives are not directly measured and must be numerically approximated. If the data is noisy or if higher order derivatives are needed, numerical approximations often results in poor-quality derivatives, which may complicate the evaluation of the ID. In this work, up to second-order derivatives of the observations have been estimated using a Savitzky-Golay filter [25]. The Savitzky-Golay filter is a non-causal smoothing filter that approximates the noise-free signal using low-degree polynomials. The set of observations z is constructed by combining actuator and sensor signals, and their higher-order derivatives.

4.2 Searching for low-dimensional signal combinations

Let $\tilde{z} \subseteq z$ denote a subset of signals such that $\tilde{z} \in \mathbb{R}^{\tilde{n}}$ where $\tilde{n} \leq n$. Based on the set of observations computed in the previous section, the next step involves identifying signal combinations $\tilde{z}$ where the ID is lower than the number of signals $\tilde{n}$ since that means that $\tilde{z}$ is approximately distributed on a low-dimensional manifold.

Since methods for evaluating ID are estimations, they do not provide a binary answer regarding whether the data distribution from a sensor combination is low-dimensional. The ID estimation methods described in Section 3.2 are local methods, i.e. the ID is evaluated by calculating the local ID for each sample using N nearest neighbors and then aggregating the information. The local ID distribution can also be used to calculate a global ID estimate, e.g. by taking the average of local ID distribution. However, taking the average value can be sensitive to outliers. Therefore, the distribution of the estimated local IDs is used to evaluate

the probability that the $\mathrm{ID}_{\tilde{z}}$, which denote the estimated ID for sensor combination $\tilde{z}$, is lower than $\tilde{n}$ minus an offset parameter α. If this probability exceeds a certain confidence threshold

$$\mathbb{P}(\mathrm{ID}_{\tilde{z}} < \tilde{n} - \alpha) \geq \theta, \tag{3}$$

the combination is considered low-dimensional, where θ is the required confidence level. Moreover, the number of nearest neighbors N must be determined. Together, these three parameters, α, θ, and N, affect the search results. The probability in (3) is estimated using a kernel density approximation (KDE) to approximate the distribution [14].

The objective is then to find all signal combinations $\tilde{z}$ such that $\mathrm{ID}_{\tilde{z}} < \tilde{n} - 1$, i.e. if (3) is satisfied. Note that if nominal data of $\tilde{n}$ is distributed on a low-dimensional manifold then all supersets of $\bar{z}$ are also distributed on low-dimensional manifolds. Thus, it is sufficient to find minimal signal combinations such that (3) is satisfied, i.e. when $\mathrm{ID}_{\tilde{z}} = \tilde{n} - 1$, instead of evaluating all signal combinations. Finding the minimal combinations is a combinatorial problem and the search space grows exponentially with n. Another important factor is the curse of dimensionality, i.e., when analyzing data in higher dimensions, the number of samples needed in training data to estimate the data distribution grows exponentially with $\tilde{n}$, see e.g. [28]. Therefore, the combinations $\tilde{z} \subseteq z$ are evaluated using a breadth-fist search principle, starting with the smallest combinations. This ensures that the minimal combinations are found first, and the search is terminated when no more minimal signal combinations can be identified.

4.3 Selection of signal combinations

After all minimal observation combinations have been found, the next step is to select a subset of these combinations for residual design. The sensor combination $\tilde{z}$ can be evaluated using different criteria. Here, each $\tilde{z}$ is used to fit a regression model using one of the signal as a target signal and evaluate the prediction error. For each $\tilde{z}$, all signals are evaluated as target signals predicted using a regression model based on the other signals in the set and the best prediction accuracy, quantified using the normalized root mean squared error (NRMSE), is selected as a metric. Here, linear regression models are used to reduce computation time when evaluating many combinations. The combinations which has a low NRMSE value and $\tilde{n} - \mathrm{ID}_{\tilde{z}} \approx 1$ are then selected for residual generation.

4.4 Residual design

When a number of minimal sensor combinations has been identified, the last step is to design residual generators based on the selected combinations. The primary idea is to use a data-driven approach to predict one the signals using the other within each selected combination, and by doing this, constructing a residual based on the prediction error. This can be achieved through various methods, such as linear regression models, neural networks, or other techniques, depending on the system's complexity.

The objective of this work is illustrating the potential of using ID. Therefore, standard linear regression modeling techniques will be used in the case studies, see e.g. [20].

5 Evaluation of method for data-driven residual design

This section will present the result of using the residual design method described in Section 4. The proposed design method for data-driven residuals is evaluated first using a simulation model of a non-linear two-tank system to compare the results with a structural analysis

of the simulation model. Then, data from a fuel injection system in a heavy-duty truck is used to validate the method. The maximum likelihood estimation method was used when evaluating the ID for both systems.

5.1 Water tank system

The water tank system consists of two connected tanks, as illustrated in Figure 1 where the upper tank is filled by a controlled pump, and the water flows from the upper to the lower tank. The model for this system, derived from first principles, is described by the set of equations:

$$\begin{aligned} \dot{h}_1 &= d_1 u - d_2\sqrt{h_1} + f_a, & y_1 &= h_1 + e_1, \\ \dot{h}_2 &= (1-f_l)d_3\sqrt{h_1} - (1-f_c)d_4\sqrt{h_2}, & y_2 &= h_2 + e_2, \\ & & y_3 &= d_5\sqrt{h_1} + e_3, \\ & & y_4 &= (1-f_c)d_6\sqrt{h_2} + e_4 \end{aligned} \tag{4}$$

where h_1, h_2 are the water levels in the upper and lower tanks, respectively, u is the control signal to the pump, y_i are the measurement signals, d_i are model parameters, e_i are measurement noise, here modeled as Gaussian distributed, f_a is a fault in the pump actuator, f_l is a leakage between the upper tank and sensor y_3, and f_c is a clogging in the output of the lower tank.

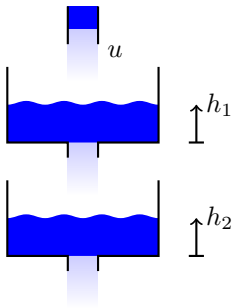

Figure 1 An illustration of the two tank system.

A structural model is derived based on the model equations (4) and implemented in Fault Diagnosis Toolbox [10]. The set of MSO candidates based on the model are identified for comparison with the identified observation combinations identified from data using ID.

5.1.1 Data description

Seven datasets are generated from the simulation model of the water tank system, including fault-free and faulty scenarios. All dataset include the following set of signals:

$$z = \{y_1, y_2, y_3, y_4, \dot{y}_1, \dot{y}_2, \dot{y}_3, \dot{y}_4, \ddot{y}_1\, \ddot{y}_2\}.$$

For simplicity, the set of signals were selected based on the model structure. Since the model had two dynamic equations, the first and second derivatives of the signals were included. Note that higher derivatives of the signals could be included, but the number of signals was intentionally kept low to reduce the number of combinations. Since only the values y_i are directly obtained, the first- and second-order time derivatives are numerically approximated. In this case, they are approximated using a Savitzky–Golay filter [25].

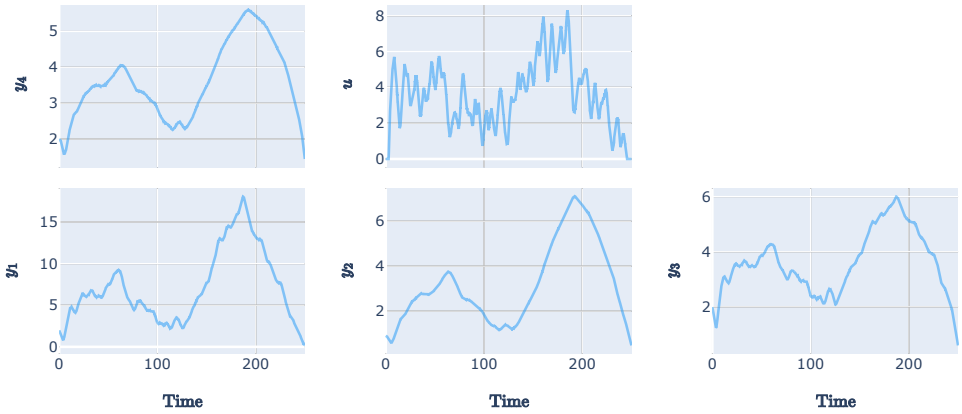

Figure 2 Example of measurements from the simulated water tank system.

The first four datasets represent nominal operation data: two are used for evaluating the ID and selecting combinations for residual design. The third and fourth datasets are used for testing the residuals, with one dataset for training the model and the other for evaluation. The remaining three datasets correspond to three different fault scenarios: a fault in the actuator (f_a), leakage between the upper tank and sensor 3 (f_l), and clogging in output of lower tank (f_c). These datasets are used to evaluate how the designed residuals react to the different faults. Figure 2 shows the signals from one of the nominal datasets, excluding the derivatives.

5.1.2 Intrinsic dimension evaluation

In addition to evaluating the ID for all signal combinations, the Normalized Root Mean Squared Error (NRMSE) of the prediction error is calculated after fitting a linear regression model to the data. Since the water tank system is non-linear, the linear regression model was extended to include the squares and square roots of all signals based on knowledge of the system behavior.

Figure 3 shows the relationship between evaluated ID and NRMSE for all 4083 signal combinations. The NRMSE was normalized by the standard deviation of the target signal to allow for a fair comparison between the different regression models predicting different target signals. The x-axis shows the difference between the number of signals in the set (the color of each point) and the estimated ID from the data. Observation sets where the dimensional difference is large result in regression models utilizing many signals for prediction which gives a better overall prediction accuracy. For dimensional difference one and higher, there are sensor combinations that achieve close to perfect prediction accuracy. For a dimensional difference smaller than one, the best NRMSE increases which indicates that there is insufficient information in the regression models to predict the target signals. This observation agrees with the expected behavior when observations are not distributed on a low-dimensional manifold.

The MSO sets derived from the water tank model (4), represent the minimal equation sets to construct consistency relations. Based on the MSO sets it is also possible to derive the minimal sets of observations needed to construct consistency relations based on the model. The sensor combinations that correspond to the MSO sets derived from the model are highlighted as stars in Figure 3. Ideally, the dimensional difference of the MSO sets should be close to one, and the NRMSE should be small. As shown in the figure, while not perfect, the MSO sets generally follow this expectation.

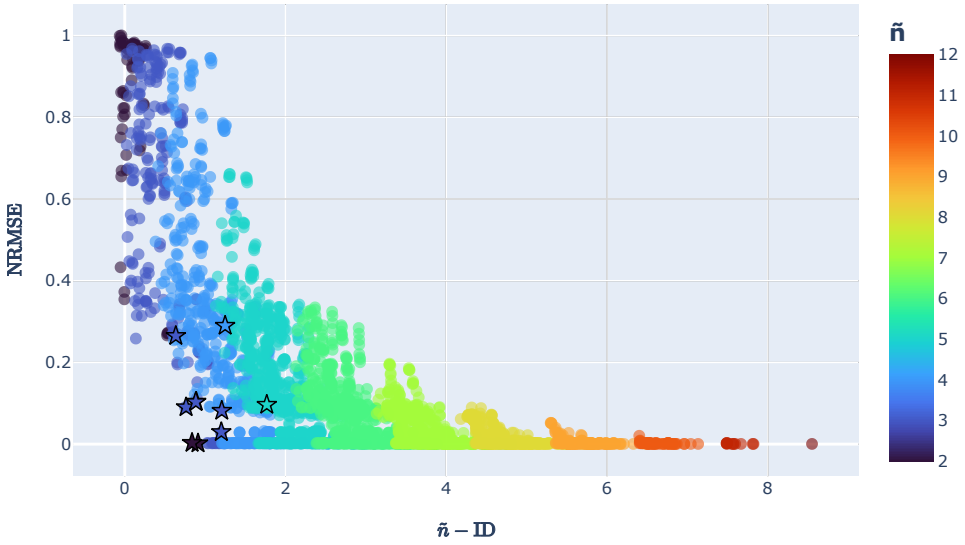

Figure 3 Normalized root mean squared error versus dimensional difference for all signal combinations for the water tank system. The stars represent the sensor combinations corresponding to MSO sets derived from the model (4). Marker color represent combination dimension, i.e. the number of signals in the set.

After evaluating all minimal signal combinations, as described in Section 4.2, 209 combinations were identified as minimal. The top six ranked combinations, those with the highest probability and a NRMSE of less than 0.2, were selected as candidates for residual design.

The parameters used for evaluating all combinations and finding minimal combinations were selected as $N = 70$, $\alpha = 0.5$, and $\theta = 0.70$. These values were identified experimentally based on their impact on the resulting minimal combinations. Both parameters α and θ influence the strictness of the search process similarly: increasing their values makes the search criteria more strict, reducing the likelihood of identifying weaker signal combinations. There is a trade-off between the risk of overlooking important combinations with higher parameter values and the risk of detecting false positives with lower values. As mentioned in [13] increasing the number of neighbors reduces the ID estimation bias when using the MLE method. However, there is a compromise, since increasing N might affect underlying assumptions of constant density in the neighborhood, which might affect the results negatively. This means that tuning of N is required that depends on the properties of the data.

The process for designing residuals using these combinations will be described next.

5.1.3 Residual design and evaluation

The residual generators are constructed based on the six selected sensor combinations using the same linear regression models as to compute the NRMSE.

Figures 4-6 show the performance of the residuals when subjected to the three modeled faults: actuator fault (f_a), clogging (f_c), and leakage (f_l), respectively. In these figures, the blue line is the residual under nominal conditions, while the orange line is the residual affected by the fault. Any overlap between the two lines is indicated by a green color. All three faults are introduced in the system using similar fault scenarios. Each fault signal behaves as a ramp starting at $t = 10$ second that increases to an amplitude of 0.4 over 2

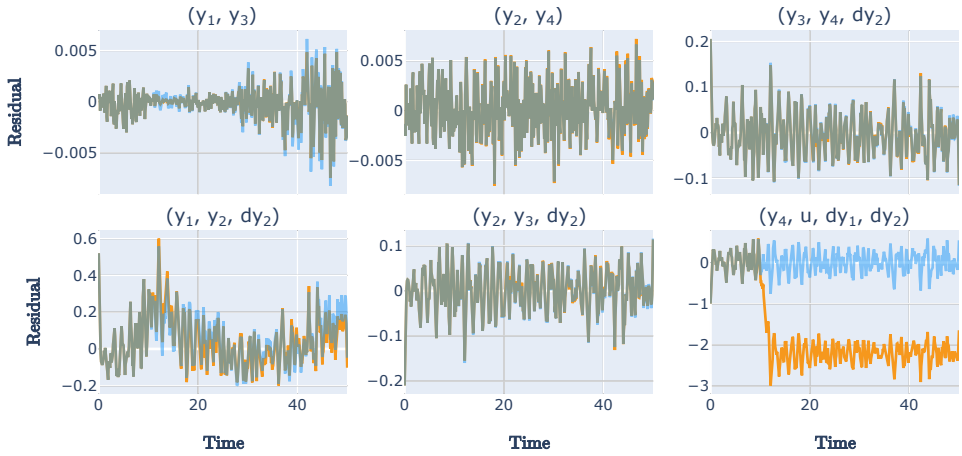

Figure 4 Comparison of residuals between fault-free data and data under an actuator fault, which occurs at $t = 10$ s.

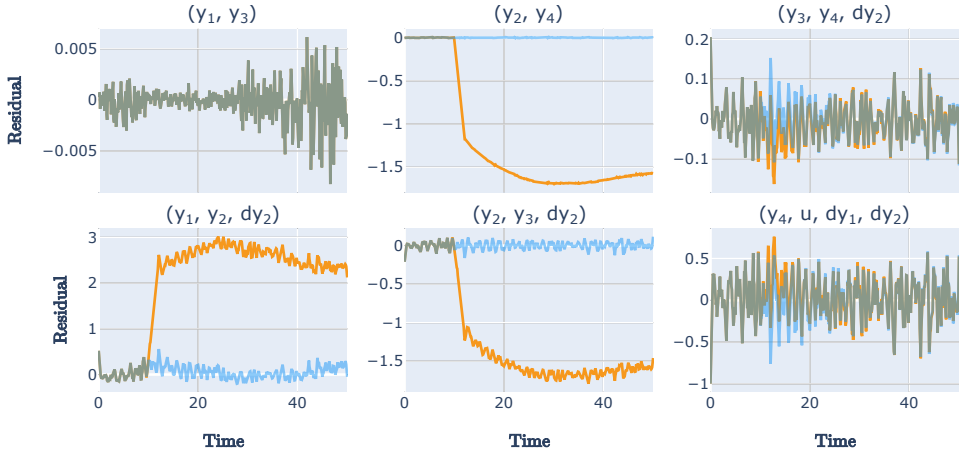

Figure 5 Comparison of residuals between fault-free data and data under a clogging fault, which occurs at $t = 10$ s.

seconds. The results demonstrate that the residuals maintain a low value under nominal conditions, while different residuals reacts to the different faults, indicating their potential for fault isolation.

It is also possible to analyze which part of the system that is modeled in each residual generator based on the selected the sensor combinations. For example the residual generator based on the signals y_1, y_2, and $\dot{y}_2$ is modeling the relationship between signals related to the dynamics of tank two based on the water level of tank one, see (4). This is also visible in the residual plots since the residual reacts to a clogging in tank one in Figure 5, which affects how much water is flowing into tank two, and a leakage in tank two in Figure 6, which affects the behavior the tank. However, the actuator fault in Figure 4 is only affecting how much water that flows into tank one but not the relationship between y_1, y_2, and $\dot{y}_2$. This shows that it is possible reason about the cause of triggered data-driven residuals based on physical insights about the system.

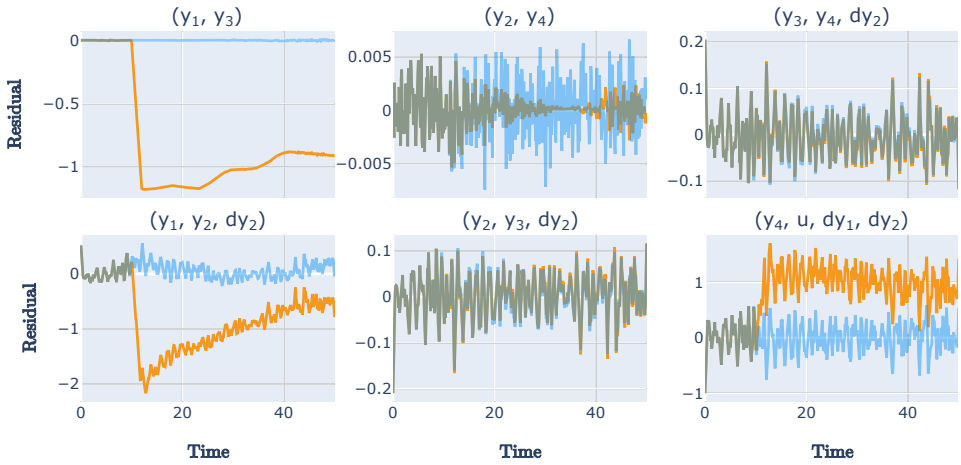

Figure 6 Comparison of residuals between fault-free data and data under a leak fault, which occurs at $t = 10$ s.

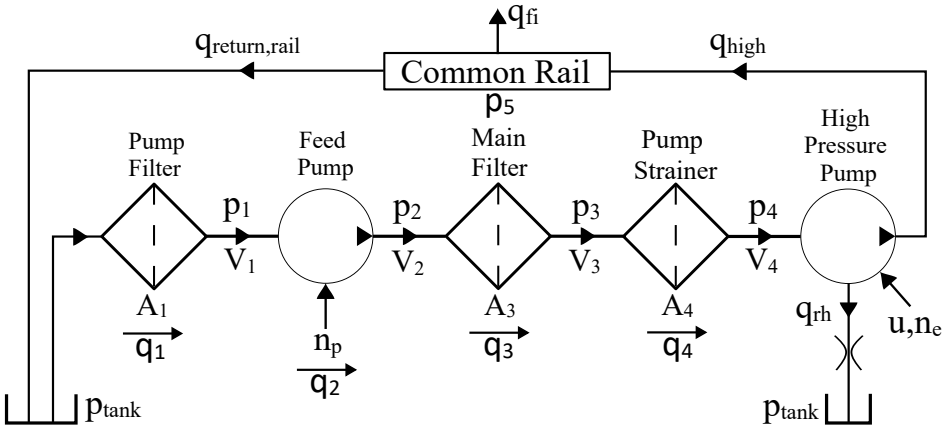

Figure 7 A schematic of the fuel injection system. The figure is used with permission from [2].

5.2 Fuel injection in heavy-duty diesel engine

The second case study is a fuel injection system in a heavy-duty diesel engine. The same case study has previously been analyzed in [2]. The system consists of two pumps used to provide fuel that is injected into the cylinders in a diesel engine. A schematic of the system is shown in Figure 7 where fuel from the tank is pumped through a set of filters and finally to the common rail where it is injected to the cylinders.

The amount of fuel injected into the cylinders is not measured but is controlled by adjusting how long a set of solenoid valves are open. Thus, the system acts to maintain a fixed pressure in the common rail to inject the correct amount of fuel. The available signals from the fuel injection system are:

- Pump speeds y_{np} and y_{ne},
- LP and the HP circuit pressures y_{p2} and y_{p5},
- HP pump solenoids duty cycle u,
- Estimated injected fuel to the cylinders $y_{q,fi}$.

5.3 Data description

Based on the results in [2], it was observed that the dynamics could be approximated by a second order system. Therefore, the extracted set of observations includes the following signals:

$$z = \{y_{np}, y_{ne}, y_{p2}, y_{p5}, y_{q,fi}, u, \dot{y}_{ne}, \dot{y}_{p5}, \dot{y}_{q,fi}, \dot{u}, \ddot{y}_{p5}, \ddot{y}_{q,fi}\},$$

where the derivatives of the signals are numerically approximated, similar as for the water tank system.

The fuel injection system data consists of four datasets, all collected during real-world operation of a heavy-duty truck:

- Fault-free data from 40-tonne load (30 min)
- Fault-free data from no load (20 min)
- Clogging in the main filter (28 min)
- Degradation of the HP pump (4 min)

The first two datasets represent nominal operation: one with the truck under a 40-ton load and the other with no load. The dataset recorded under load conditions is used for performing the ID evaluation and designing the residuals, while the no-load dataset is used to evaluate these residuals. The remaining two datasets were recorded under faulty conditions: one with a clogged main filter in the fuel injection system, and the other with a pump-related failure. These datasets are used to assess how the faults impact the designed residuals. Figure 8 shows the data from the 40-tonne load nominal dataset.

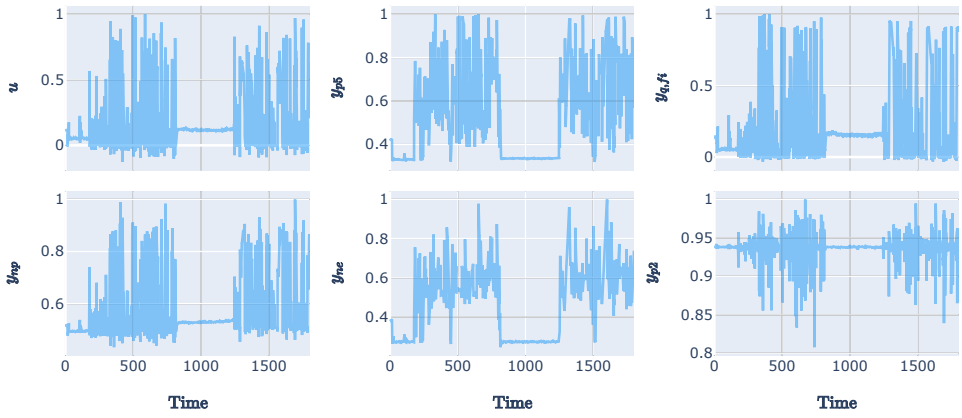

Figure 8 Example of measurements from the fuel injection system in a heavy-duty truck.

The evaluation procedure follows the same methodology used for the water tank system described in Section 5.1. However, there is no comparison of the solution to any MSO sets.

5.3.1 Intrinsic dimension evaluation

The 12 selected signals in z result in 4083 sensor combinations in total. Out of all of these combinations, 462 were classified as low-dimensional with the parameters: $\alpha = 0.5$, $\theta = 0.65$, and $N = 160$. These parameters were based on those used in the water tank system, but were adjusted to account for the larger and noisier dataset. Specifically, the number of neighbors was increased to improve the results, and the search strictness was slightly reduced by lowering θ to avoid overlooking potentially relevant combinations.

Here, data is used to fit a linear regression model to compute the prediction error. Based on the results in [2], it was concluded that a linear model was sufficient to model the system behavior. Figure 9 shows the relationship between NRMSE and dimensional difference for all combinations. The results show the same characteristic behavior as for the water tank system where a higher dimensional difference result in a lower prediction error. Still, the objective is to design residual generators using minimal sensor sets. Analogously to Section 5.1, six candidates were selected to be used for residual design. Because of the higher RMSE in this case, the candidates were chosen to be below a NRMSE of 0.4 instead of 0.2.

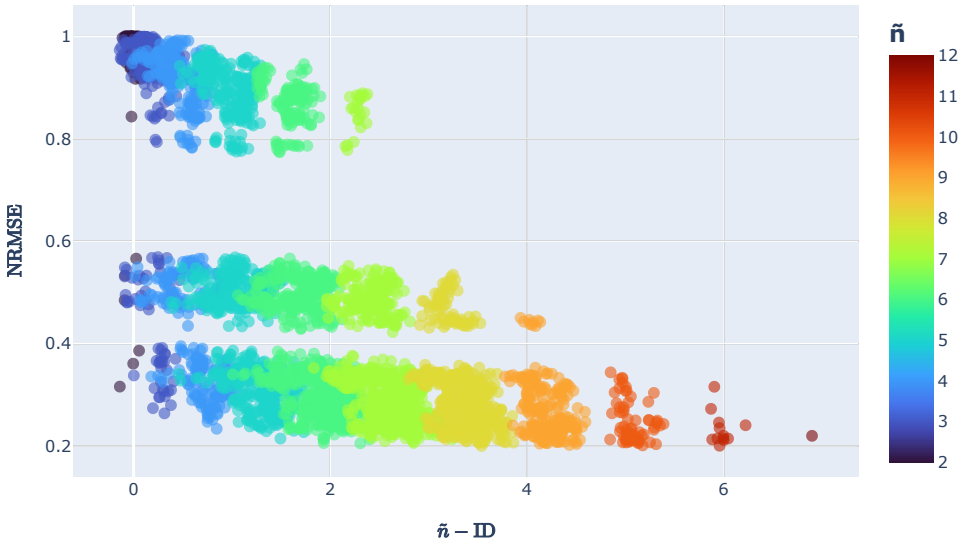

Figure 9 Normalized root mean squared error versus dimensional difference for all signal combinations for the fuel injection system. Marker color represent the dimension of the observation space, i.e. the number of signals $\tilde{n}$.

5.3.2 Residual design and evaluation

Figures 10 and 11 show how the residuals react to the two different faults, both time series data and the distributions of the residual during nominal (blue) and faulty conditions (orange). Even tough the residuals are affected by noise, the faults are detectable. Since each fault is resulting in a bias in the residual output, the noise can be suppressed using a low-pass filter without significantly reducing detection performance. The results from the analysis of the fuel injection system show that the proposed design method can be used for designing data-driven residual generators for fault diagnosis in industrial applications.

6 Identifying sensor combinations for fault isolation using augmented training data

The residual design process for the two case studies in the previous section identified sensor combinations based on the prediction error. This means that there is no information about the fault sensitivity of the constructed residual generators. In [21], it is shown that data from decoupled faults are distributed along the same manifold as fault-free data. Thus, augmenting the training data used in the ID evaluation with data from faults that should be decoupled could be used to find signal combinations that are insensitive to the fault.

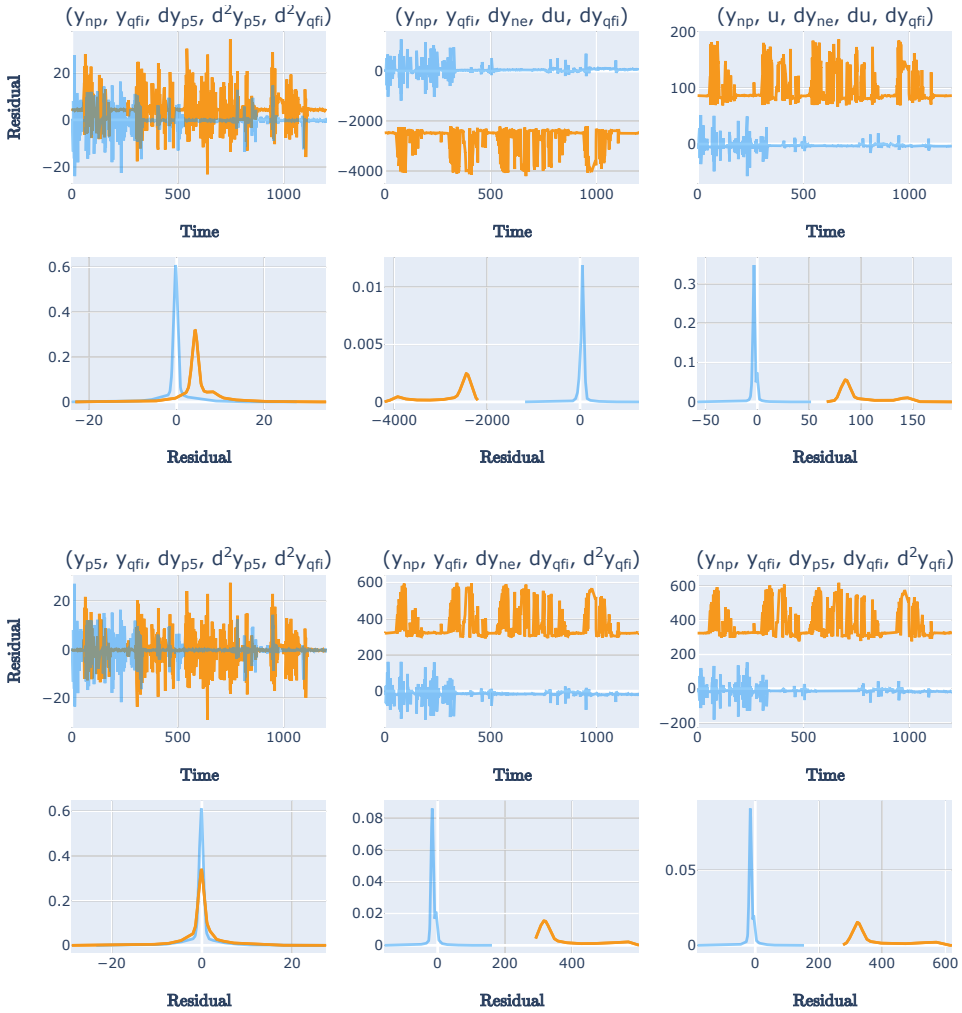

Figure 10 Residual comparison between no fault and clogging in the main filter. Each residual have two separate plots, the upper plot showing the time-series data of the residuals, and the lower plot showing the residual distribution.

The design process in the previous section for the water tank system using training data augmented with data from the clogging fault. A new set of sensor combinations is identified and used to construct residual generators. Figure 12 shows the new selected residuals based on the six top ranked signal combinations, when performing exactly the same procedure as previously. All residuals are insensitive to the clogging fault, demonstrating the potential of finding signal combinations that can be used to form residuals unaffected by a chosen fault. This is promising, since have residuals that are affected by different faults is essential for consistency-based fault isolation.

Experiments also showed that for the ID algorithm to identify sensor combinations that were insensitive to the fault in the augmented training data required sufficient excitation of the fault. For example, a constant fault that is only introducing a shift in the measurements will not affect the estimated ID since the local ID around those measurements will be the same.

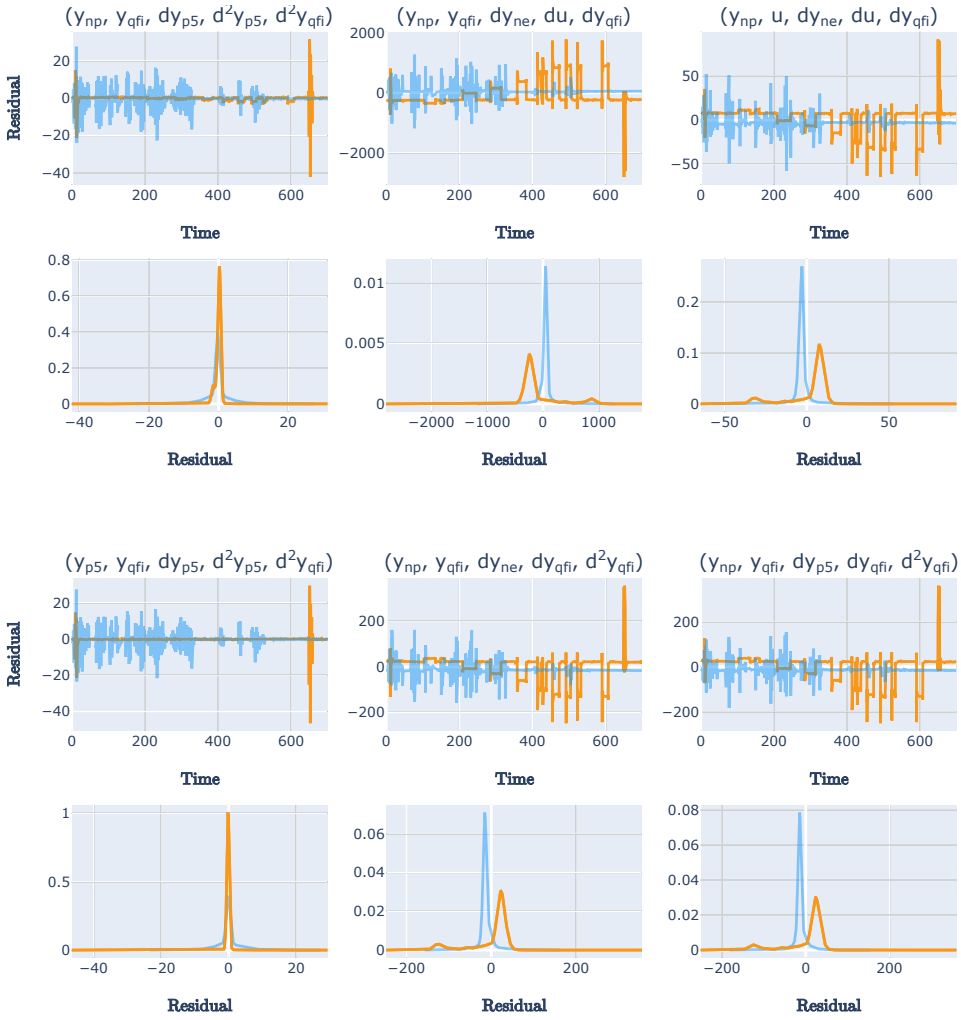

Figure 11 Residual comparison between no fault and a pump related fault.

7 Conclusions and future works

Even though data-driven diagnosis is often treated as a standard classification problem, it is shown that theory from model-based diagnosis, e.g. redundancy, can be used in a data-driven context. When representative training data is limited, there is a risk of misclassifications due to generalization problems. Thus, theory and methods for data-driven fault diagnosis are needed. The connection between analytical redundancy and ID is evaluated using data from a simulated two-tank system. The results show the connection between model-based methods to find consistency relations and the distribution of data. The results when evaluating the prediction performance of different sensor combinations validate the fact that prediction accuracy is improved by utilizing more observations from the system. However, from a fault isolation perspective it is more relevant to identify minimal subsets of observations for residual generation, similar to MSO sets in model-based diagnosis. A data-driven diagnosis system design process is also proposed using the ID to find candidate observation sets for residual generation. The proposed method is validated using both the simulation model of the two tank system and data from a fuel injection system.

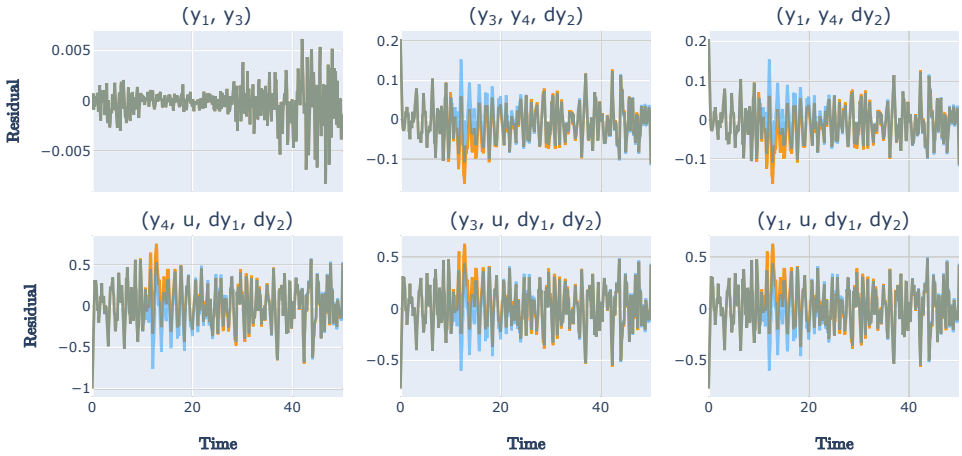

Figure 12 Comparison of residuals between normal data and data under a clogging fault, which occurs at $t = 10$ s.

As future work, more investigations are needed to efficiently estimate the ID using noisy data. Here, derivative causality was considered, i.e. signals and their derivatives, which has limitations for complex dynamic systems. Thus, it should be investigated how to apply ID for integral causality, e.g. how to directly analyze time-series data instead of estimating the derivatives.

References

1 A. Abid, M. Khan, and J. Iqbal. A review on fault detection and diagnosis techniques: basics and beyond. *Artificial Intelligence Review*, 54:3639–3664, 2021. doi:10.1007/S10462-020-09934-2.

2 N. Allansson, A. Mohammadi, D. Jung, and M. Krysander. Fuel injection fault diagnosis using structural analysis and data-driven residuals. *IFAC-PapersOnLine*, 58(4):360–365, 2024.

3 L. Amsaleg, O. Chelly, T. Furon, S. Girard, M. Houle, K. Kawarabayashi, and M. Nett. Extreme-value-theoretic estimation of local intrinsic dimensionality. *Data Mining and Knowledge Discovery*, 32(6):1768–1805, 2018. doi:10.1007/S10618-018-0578-6.

4 J. Bac, E. Mirkes, A. Gorban, I. Tyukin, and A. Zinovyev. Scikit-dimension: a python package for intrinsic dimension estimation. *Entropy*, 23(10):1368, 2021. doi:10.3390/E23101368.

5 R. Bennett. The intrinsic dimensionality of signal collections. *IEEE Transactions on Information Theory*, 15(5):517–525, 1969. doi:10.1109/TIT.1969.1054365.

6 F. Camastra and A. Staiano. Intrinsic dimension estimation: Advances and open problems. *Information Sciences*, 328:26–41, 2016. doi:10.1016/J.INS.2015.08.029.

7 X. Dai and Z. Gao. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics*, 9(4):2226–2238, 2013. doi:10.1109/TII.2013.2243743.

8 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. doi:10.1016/J.AUTOMATICA.2013.02.045.

9 M. Fan, N. Gu, H. Qiao, and B. Zhang. Intrinsic dimension estimation of data by principal component analysis. arxiv. *Preprint]. doi*, 10, 2010.

10 E. Frisk, M. Krysander, and D. Jung. A toolbox for analysis and design of model based diagnosis systems for large scale models. *IFAC-PapersOnLine*, 50(1):3287–3293, 2017.

11 E. Frisk and M. Nyberg. A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems. *Automatica*, 37(9):1417–1424, 2001. doi:10.1016/S0005-1098(01)00078-4.

12 K. Fukunaga and D. Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on computers*, 100(2):176–183, 1971. doi:10.1109/T-C.1971.223208.

13 G. Haro, G. Randall, and G. Sapiro. Translated poisson mixture model for stratification learning. *International Journal of Computer Vision*, 80:358–374, 2008. doi:10.1007/S11263-008-0144-6.

14 T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning: data mining, inference, and prediction, 2017.

15 D. Jung. Isolation and localization of unknown faults using neural network-based residuals. In *Annual Conference of the PHM Society*, volume 11, 2019.

16 D. Jung, M. Krysander, and A. Mohammadi. Fault diagnosis using data-driven residuals for anomaly classification with incomplete training data. *IFAC-PapersOnLine*, 56(2):2903–2908, 2023.

17 M. Krysander, J. Åslund, and M. Nyberg. An efficient algorithm for finding minimal overconstrained subsystems for model-based diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1):197–206, 2007. doi:10.1109/TSMCA.2007.909555.

18 M. Krysander and E. Frisk. Sensor placement for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(6):1398–1410, 2008. doi:10.1109/TSMCA.2008.2003968.

19 E. Levina and P. Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17, 2004.

20 L. Ljung. *System identification (2nd ed.): theory for the user*. Prentice Hall PTR, USA, 1999.

21 A. Mohammadi, M. Krysander, and D. Jung. Analysis of grey-box neural network-based residuals for consistency-based fault diagnosis. *IFAC-PapersOnLine*, 55(6):1–6, 2022.

22 A. Rehman, W. Jiao, J. Sun, H. Pan, and T. Yan. Open set recognition methods for fault diagnosis: A review. In *2023 15th International Conference on Advanced Computational Intelligence (ICACI)*, pages 1–8. IEEE, 2023.

23 W. Rheinboldt. On the computation of multi-dimensional solution manifolds of parametrized equations. *Numerische Mathematik*, 53(1):165–181, 1988.

24 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. doi:10.1109/ACCESS.2015.2422833.

25 A. Savitzky and M. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

26 A. Theissler, J. Pérez-Velázquez, M. Kettelgerdes, and G. Elger. Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. *Reliability engineering & system safety*, 215:107864, 2021. doi:10.1016/J.RESS.2021.107864.

27 L. Travé-Massuyès. Bridging control and artificial intelligence theories for diagnosis: A survey. *Engineering Applications of Artificial Intelligence*, 27:1–16, 2014. doi:10.1016/J.ENGAPPAI.2013.09.018.

28 M. Verleysen and D. François. The curse of dimensionality in data mining and time series prediction. In *International work-conference on artificial neural networks*, pages 758–770. Springer, 2005.

29 Y. Xu, S. Kohtz, J. Boakye, P. Gardoni, and P. Wang. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. *Reliability Engineering & System Safety*, 230:108900, 2023. doi:10.1016/J.RESS.2022.108900.

30 Z. Xu and J. Saleh. Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. *Reliability Engineering & System Safety*, 211:107530, 2021. doi:10.1016/J.RESS.2021.107530.

31 T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, and E. Xu. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA transactions*, 119:152–171, 2022.