Preprints, 10th IFAC International Symposium on
Advances in Automotive Control
Columbus, Ohio, USA, August 28-30, 2022

A flexi-pipe model for residual-based engine fault diagnosis to handle incomplete data and class overlapping

Daniel Jung Joakim Säfdal

Dept. of Electrical Engineering, Linköping University, SE-581 83, Linköping, Sweden.
(e-mail: daniel.jung@liu.se, joakim.safdal@gmail.com).

Abstract: Data-driven fault diagnosis of dynamic systems is complicated by incomplete training data, unknown faults, and overlapping classes. Many existing machine learning models and data-driven classifiers are not expected to perform well if training data is not representative of all relevant fault realizations. In this work, a data-driven model, called a *flexi-pipe model*, is proposed to capture the variability of data in residual space from a few realizations of each fault class. A diagnosis system is developed as an open set classification algorithm that can handle both incomplete training data and overlapping fault classes. Data from different fault scenarios in an engine test bench is used to evaluate the performance of the proposed methods. Results show that the proposed fault class models generalize to new fault realizations when training data only contains a few realizations of each fault class.

Keywords: AI/ML application to automotive and transportation systems, Model-based diagnostics, Open set classification, Engine fault diagnosis.

1. INTRODUCTION

Fault diagnosis of dynamic systems considers the problem of detecting abnormal system behavior at an early stage and identify its root cause. A common approach is to use residual generators to detect inconsistencies between sensor data and model predictions. Dynamic systems can have a wide operating range, including different operating conditions and transient behavior. Predictive models can be used to compute residuals that filter out the system dynamics while still being sensitive to faults. An advantage of using residuals as features is that fault-free data are distributed around the origin while data from different fault classes are projected into different manifolds of residual space. If the predictive model is derived from physical insights this is often referred to as model-based diagnosis and if a black-box model is trained using previously collected operational data, it is referred to as data-driven diagnosis.

An advantage of model-based diagnosis is that it is possible to isolate unknown faults by using model analysis (Jung et al., 2018). However, deriving high fidelity models from physical insights for fault diagnosis applications is a time-consuming process. Data-driven models are attractive because they can learn complex information from data. A complicating factor of data-driven fault diagnosis is that faults are rare events. This means that available training data from faults is often imbalanced and not representative of all relevant realizations of each fault class, especially during early system life before any fault has occurred (Jung et al., 2018; Sankavaram et al., 2015). One solution is to train data-driven residual generators from fault-free data as features for fault detection and fault classification. Machine learning and data-driven classification algorithms rely on representative training data to determine decision boundaries to distinguish between different classes. Classification models try to map different sets of feature outputs to class labels that best can explain the outputs. Selecting a suitable model to capture the relevant properties of feature data in a specific application is a non-trivial problem. Many data-driven classifiers do not generalize well which results in unreliable predictions when new feature outputs significantly deviate from training data. One solution is to use open set classification algorithms that model the data support of classes available in training data (Jung et al., 2018; Scheirer et al., 2013). However, using only the data support to model fault classes will result in many fault scenarios are identified as unknown faults if training data only consists of a few realizations of each fault class. Identifying which models are appropriate for a given problem is important to increase model interpretability and avoid misclassifications.

Another complicating factor of fault diagnosis is class overlapping, i.e., that different fault classes can result in similar residual outputs. This happens, e.g., in early stages of system degradation when it is difficult to distinguish small faults from nominal system behavior. An example is shown in Fig. 1 where residual data from an engine test bench is plotted against each other. Each color represents one fault class and data have been collected from different fault magnitudes. Fault-free data (No Fault - NF) and data from different fault classes with small magnitudes are located close to the origin illustrating the overlapping of different classes. Multi-class classification algorithms that only identifies the most likely class that can explain data, e.g., Random Forests, could result in unnecessary

misclassifications. One approach is to model data from each class separately and identify which data classes that can explain the residual outputs (Jung et al., 2018).

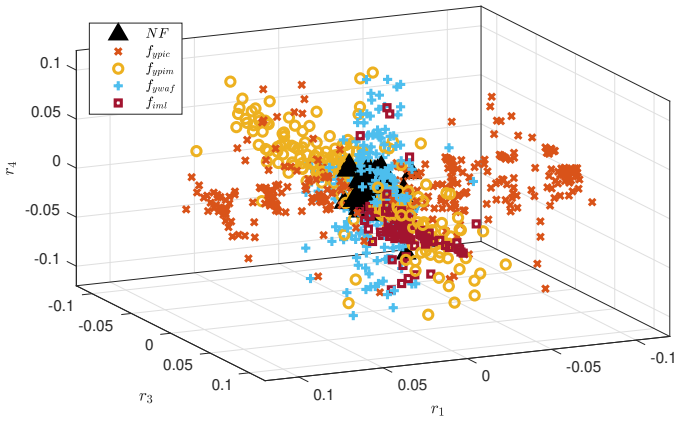

Fig. 1. Residual data collected from the engine test bench (Jung, 2020).

To handle the complicating factors of data-driven fault diagnosis, there are several works investigating the benefits of combining model-based and data-driven methods. A general approach to diagnose dynamic systems is to use sensor data to compute residuals to filter out system dynamics. The residual outputs can then be used as input to a data-driven classifier (Jung et al., 2018). In Slimani et al. (2018), a Bayesian fusion strategy for hybrid fault diagnosis is proposed when combining model-based and data-driven techniques. A hybrid diagnosis system design is developed in Luo et al. (2009) for an ABS system combining support vector machines and model-based residuals. In Khorasgani and Biswas (2018), a hybrid diagnosis system design is proposed for monitoring of smart buildings where model-based residuals are developed based on available system models combined with additional data-driven anomaly classifiers for non-modeled parts of the system. With respect to previous work, this paper focuses on the development of a data-driven model to capture the variability of residual data from different fault classes that can be used for open set classification.

2. PROBLEM STATEMENT

The main contribution in this work is to develop a data-driven model for fault class modeling using residual data. The purpose is to handle the complicating properties of data-driven fault diagnosis of dynamic systems that are summarized in the following bullets:

- Imbalanced data (Sankavaram et al., 2015)
- Unknown fault classes (Scheirer et al., 2013)
- Class overlapping (Lundgren and Jung, 2022)

The purpose of the proposed data-driven model is to improve classification accuracy when training data from faults are limited and only consists of a few realizations from each fault class. An advantage of the proposed model is that it takes into consideration variations in residual data that depend on different realizations of each fault class, referred to as the fault signature. Residual data has been collected from different fault scenarios on an engine test bench (Jung, 2020) to evaluate the proposed method, see Fig. 1.

3. BACKGROUND

Before describing the proposed data-driven fault model, the principles of model-based fault diagnosis using residuals are discussed. Then, a summary of two machine learning methods, Principal Component Analysis and Gaussian Processes, that will be used to formulate the proposed fault class model is given.

3.1 Residual-based fault classification

Residual-based fault detection and isolation is one of the main approaches in model-based diagnosis. Residual outputs r are used to diagnose faults by detecting inconsistencies between sensor data y and model predictions $\hat{y}$. This is illustrated in Fig. 2 where f denotes faults affecting the system and u denotes the control signals.

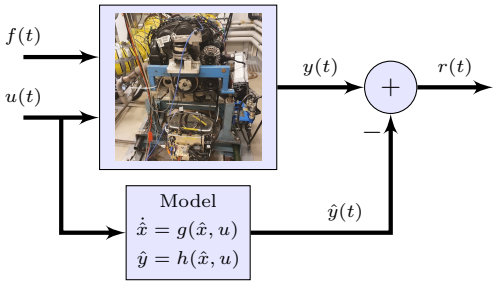

Fig. 2. A residual $r(t)$ compares measurements from the system $y(t)$ with model predictions $\hat{y}(t)$.

When faults occur in different parts of the system, they will have various impact on the overall system behavior. Fault classification is performed by analyzing the residual patterns, for example using consistency-based methods, see, e.g., Pulido and González (2004), to identify a set of fault hypotheses. In practice, it is likely that both the residual mean, and covariance will vary, and not only as a function of fault size but also with varying operating conditions of the system and modeling errors.

3.2 Modeling data variability using PCA

To model the manifold in residual space that describes the distribution of residual data for a given fault class, it is assumed here that the main variability of residual data can be described by a linear subspace, similar to the fault signature for linear systems. Principal Component Analysis (PCA) is the process of finding an ordered set of orthonormal vectors, denoted principal components, along which a given set of data points are linearly uncorrelated. The principal components are ordered such that the first component represents the largest eigenvector of the data's covariance matrix, the second component represents the second largest eigenvector, and so on. Thus, PCA can be used to perform a change of basis and is a popular tool for dimension reduction by only keeping the dominant principal components (Hastie et al., 2009).

3.3 Gaussian Processes

Gaussian Processes (GP) is a non-parametric machine learning method which models data as a stochastic process where any finite collection of random variables from that process has a multivariate normal distribution (Williams and Rasmussen, 2006). A GP model can be used to model a spatially correlated function $y = g(u)$ and its behavior is defined by the mean $\mu(u)$ and covariance function $k(u, u')$. The GP is denoted as

$$\phi(u) \sim \mathcal{GP}\left(\mu(u), k(u, u')\right) \tag{1}$$

Here, the parameters that are used to describe the distribution of residual data from one fault class are modeled using GP to capture the uncertainties in the residual model and sensor noise for different fault realizations.

4. DATA-DRIVEN MODELING OF FAULT CLASSES USING FAULT SIGNATURES

In this section, a data-driven model of fault classes is proposed based on the idea that the variability of residual data from one fault class can be described by a manifold. Here, this manifold represents the non-linear version of the fault signature for the linear cases.

4.1 Modeling fault classes

The distribution of residual data given a specific fault class depends on the fault magnitude and it is here assumed that the main variability of data from one fault class can be represented by a fault signature vector. The distribution of residual output $\bar{r}$ is not only conditionally depending on fault class F but also fault magnitude θ_F. For small fault magnitudes, the residual outputs will have a distribution that is located close to the origin and similar to the nominal (fault-free) class. Thus, fault-free training data represents the asymptotic distribution of residual data when the fault magnitude goes to zero. Therefore, training data from the fault-free case is included when modeling data from each fault class.

To model fault class F, PCA is applied to training data from that class (including fault-free data) to derive a transition matrix A^F based on the principal vectors to perform a change of basis as

$$\bar{r}_F = A^F \bar{r} \tag{2}$$

The first principal vector, represented by the first row in A^F denoted A_1^F, is used to model the fault signature. Then, $\bar{r}_{\|}^F = \bar{r}_1^F = A_1^F \bar{r}$ represents the location of the residual outputs along the main principal vector. Data along the remaining principal components is denoted $\bar{r}_{\perp}^F$.

The conditional distribution $p(\bar{r}_{\perp}^F | \bar{r}_{\|}^F)$ is modeled as a multivariate normal distribution

$$p(\bar{r}_{\perp}^F | \bar{r}_{\|}^F) \sim N(\mu_{\perp}^F(\bar{r}_{\|}^F), \Sigma_{\perp}^F(\bar{r}_{\|}^F)) \tag{3}$$

where the elements of $\mu_{\perp}^F \in \mathbb{R}^{n-1}$ and $\Sigma_{\perp}^F \in \mathbb{R}^{n-1 \times n-1}$ are modeled as functions of $\bar{r}_{\|}^F$. Since there is, in general, no parametric model of how the distribution of residual data varies along the principal vector, GP is used as a non-parametric model of each distribution parameter. To assure that $\Sigma_{\perp}^F$ is invertible and positive definite, a set of GP models are trained to estimate the parameters of the Cholesky decomposition $\Gamma_{\perp}^F(\bar{r}_{\|}^F)$. If $\Gamma_{\perp}^F(\bar{r}_{\|}^F)$ is lower triangular with positive diagonal entries, the multiplication $\Sigma_{\perp}^F(\bar{r}_{\|}^F) = \Gamma_{\perp}^F(\bar{r}_{\|}^F)\Gamma_{\perp}^F(\bar{r}_{\|}^F)^T$ will be positive definite. For each fault class model, a set of $n-1$ GP models is needed to model the mean and $\frac{n(n-1)}{2}$ GP models for the Cholesky factorization of the covariance matrix.

To illustrate the proposed fault class model (3), a simulation study with residual data from three different fault classes are used. Figure 3 shows generated data from three simulated fault classes with varying fault magnitudes. The samples from a subset of fault realizations, represented by 'x' are used as training data while validation data are shown as opaque '·'.

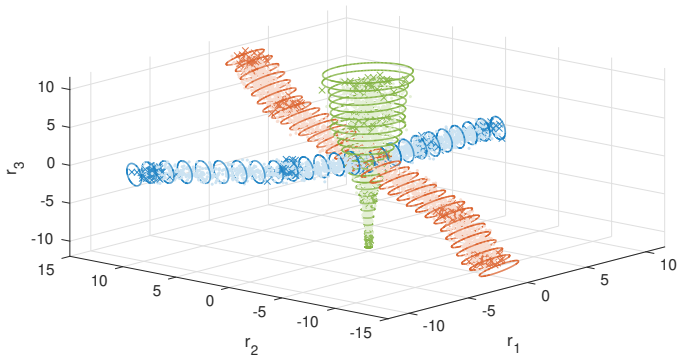

Fig. 3. Simulated residual data from three different fault classes where 'x' denotes training samples. An illustration of the flexi-pipe model trained using residual data from each fault class. The ellipses represent the modeled data distribution along the fault signature estimated for each fault class.

A model (3) for each fault class is trained by first applying PCA to derive a transition matrix A^F and then generate a set of bootstrap samples that are used to estimate distribution mean and covariance of residual data along the main principal vector. These estimates are then fed into a set of GP models to predict the distribution parameters.

The trained fault models are shown in Fig. 3 where the ellipses represent the modeled covariance along the main principal component (fault signature). Figure 3 shows that the proposed model captures the residual behavior, including varying noise levels and the shape of the fault signature. In addition, it gives an approximation of the distribution of fault realizations in between training data. Because of the visual resemblance, the fault model is referred to as a *flexi-pipe* model.

4.2 Training of the flexi-pipe model

Here the process of training the GP models for μ and Γ is described for the flexi-pipe model using training data from fault class F (including data from the fault-free class). First, all training data from that fault class is used to compute the transition matrix A^F which is then used to represent residual data in that new base. Then, a set of bootstrap estimates of μ and Γ are generated for different inputs $\bar{r}_{\|}^F$. Consider the subset of training samples $\bar{r}_{\perp}^F$ such that $\left|\bar{r}_{\|}^F - \bar{r}_{\|,0}^F\right| \leq \varepsilon$ for some value of $\bar{r}_{\|,0}^F$ and $\varepsilon > 0$. If the number of samples in that interval is sufficiently large, a

set of bootstrap estimates are computed for different values of $\bar{r}^F_{\parallel,0}$. The generated set of bootstrap estimates are used to train the set of GP models modeling each element in $\mu(\bar{r}^F_{\parallel})$ and $\Gamma(\bar{r}^F_{\parallel})$, respectively.

5. DATA-DRIVEN FAULT DIAGNOSIS USING INCOMPLETE DATA

Fault classification is performed in two steps. First, a fault detection phase is used to monitor residuals to detect abnormal behavior. When a fault is detected, the fault isolation phase is activated where the residual outputs are used to rank different fault hypotheses, including the hypothesis that an unknown fault has occurred. Here, the known fault classes are modeled using the flexi-pipe model to capture the variability of residual data from a limited set of training data of each fault. The set of fault hypotheses are ranked based on how well each fault class can explain the residual outputs.

5.1 Ranking of fault hypotheses using the flexi-pipe model

To rank the different fault hypotheses, a one-class classifier is implemented based on the flexi-pipe model. A decision function is formulated using the Mahalanobis distance, see for example Murphy (2012), where a sample of the residual outputs $\bar{r}$ is said to be explained by fault class f if

$$(\bar{r}^F_{\perp} - \mu(\bar{r}^F_{\parallel}))^T \Sigma^{-1}(\bar{r}^F_{\parallel})(\bar{r}^F_{\perp} - \mu(\bar{r}^F_{\parallel})) \leq J^2 \qquad (4)$$

where J is a threshold representing how many standard deviations that a sample is allowed to deviate from the mean. Since residual data are tested with respect to all known fault classes, there could be more than one fault class that can explain the residual output if (4) is true for multiple fault classes f. When no known fault class can explain a sample $\bar{r}$, i.e., (4) is not fulfilled for any known fault class, the sample is said to belong to an unknown fault class Jung et al. (2018). This means that each fault hypothesis representing a known fault class is ranked between $0\% - 100\%$ depending on how many samples that can be explained by that class. The unknown fault class is ranked between 0% and one minus the highest rank for any known fault class.

6. CASE STUDY

The proposed flexi-pipe models and diagnosis system are evaluated by using experimental data collected from an engine test bench, see Fig. 4. The engine is a commercial, turbo charged, four-cylinder, internal combustion engine from Volvo Cars. The sensor and actuator setup is the standard commercial configuration for the engine (Jung et al., 2018).

Several data sets, one for each fault class and magnitude, has been collected for different fault scenarios where the engine is operated both stationary and transient modes by following the Worldwide Harmonized Light Vehicles Test Procedure (WLTP) driving cycle. Sensor faults are injected in the engine control unit as multiplicative faults while leakages are implemented using valves of varying orifices. A description of the faults is given in Table 1. Data sets have been collected for realizations of each fault class with magnitudes between -20% and 20% for sensor faults and leakages with an orifice diameter of 4mm and 6mm, respectively.

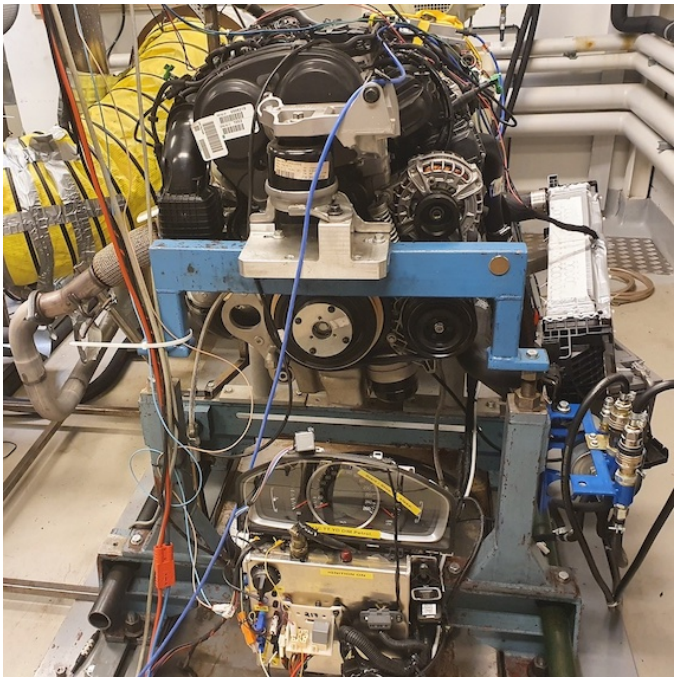

Fig. 4. The engine test bench which was used for data collection. The engine is a commercial four cylinder combustion engine with standard sensor and actuator configuration (Jung, 2020).

Table 1. Fault classes considered in the case study. All sensor faults are induced as multiplicative faults.

Fault Class	Description
NF	Fault-free class
f_{ypim}	Fault in intake manifold pressure sensor
f_{ypic}	Fault in intercooler pressure sensor
f_{ywaf}	Fault in air-mass flow sensor
f_{iml}	Leakage in the intake manifold

6.1 Residual data generation

A set of three residual generators has been implemented by training a set of recurrent neural networks (RNN) for regression using fault-free data as described in Jung (2020). Figure 5 shows an example of the residual r_2 evaluated on fault-free data and data from the sensor fault f_{ypim} with a magnitude of 10%. The upper plots show the output from a sensor measuring the pressure at the intake manifold and the corresponding model prediction. The lower plots show the resulting residual outputs. There is a visible deviation in the residual output between the nominal and the faulty case.

Figure 6 shows an example of when the flexi-pipe model has been fit to residual data from different fault classes. The model captures the variability in residual data for each fault class and the figure also illustrates how each model extrapolates beyond training data.

7. EVALUATION

To evaluate the proposed flexi-pipe model, a set of experiments are conducted using the case study to show the

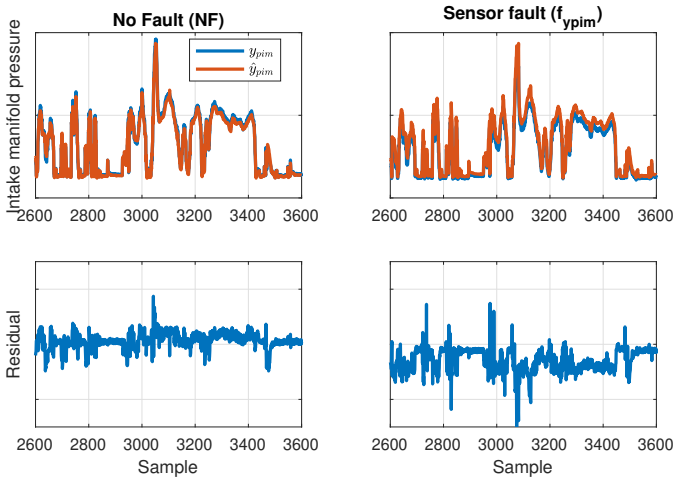

Fig. 5. Sensor data y_{pim} and model prediction $\hat{y}_{pim}$ from an RNN regression model used in r_2 during nominal and faulty operation. The lower plots show the resulting residual outputs.

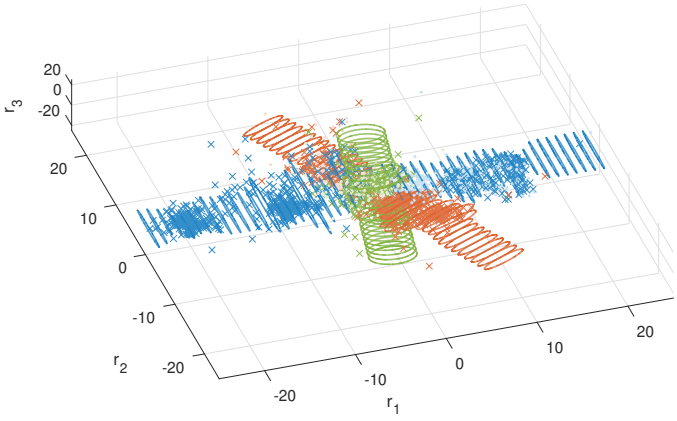

Fig. 6. An example of the flexi-pipe model fit to residual data from different fault classes.

performance with imbalanced training data and unknown faults. The residual outputs have been normalized to have a variance equal to one in the fault-free case. Because the flexi-pipe model is used for modeling fault classes and not the fault-free class, the focus of the case study will be on the fault isolation step and the fault size estimation step.

In the evaluation, the training data set only consists of fault-free data and sensor fault data of magnitude -20%. The remaining data sets are used as test data. The leakage datasets will be used to simulate an unknown fault and are not included in training data.

To compare the results of the flexi-pipe model with other data-driven classification principles, an open set fault classification scenario is evaluated, taking into consideration that there can be unknown fault classes, and a closed set fault classification scenario, where all fault classes are represented in training data. In the closed set scenario, a Random Forest (RF) classifier with 100 trees is evaluated to represent a multi-class classifier. In the open set scenario, the open set classification algorithm proposed in Jung et al. (2018) is evaluated where the fault classes are modeled using a set of one-class support vector machines.

7.1 Closed set fault classification

The first analysis will focus on the closed set case, which is also referred to as the closed world assumption, where it is assumed that there are no unknown fault classes. The two classifiers have been trained on the same dataset. Note that fault-free data have not been included in the training set for the RF classifier. Each fault class is ranked in the different fault scenarios based on how many samples that can be explained by each fault class. The total ranking of all classes for both the RF classifiers sum to 100% since each sample is only classified to one fault class.

The results from the closed set case study are shown in Fig. 7. The curves show the ranking of each fault class for different fault scenarios and fault sizes, where the curve corresponding to the true fault class is highlighted. The MB classifier gives the true fault a high rank in all fault scenarios. However, when analyzing faults f_{ypim} and f_{ywaf} it is visible that there is a significant overlap between the different classifiers, especially for small fault sizes, since all fault classes have a high rank. However, the performance of RF classifier significantly degrades for positive fault sizes and is not able to identify f_{ypic} and f_{ypim}. The RF classifier performs well when ranking fault class f_{ywaf}, even though the ranking goes down for positive faults, which is reasonable since the residual output is not changing significantly for different fault sizes. Note that the RF classifier cannot handle overlapping classes which is visible by that fault class f_{ywaf} receives a high rank for small sizes of the other fault classes.

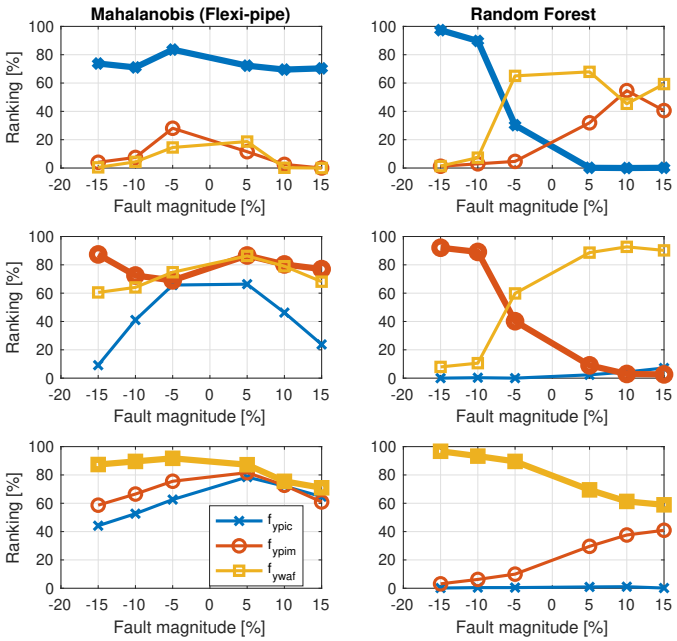

Fig. 7. Evaluation of closed set fault classification problem. The plots show the ranking of different fault classes for the different classifiers as a function of fault size where the true fault class is marked in each subplot.

7.2 Open set fault classification

In the open set case, the unknown fault class is added as a fault hypothesis. Here, the MB is implemented where the

unknown fault hypothesis is included. The RF classifier cannot be used for open set classification. Instead, a set of one-class support vector machines (1SVM) are trained for each fault class. Fault hypotheses are ranked in the same way as for the MB classifier but where the unknown fault class is ranked by counting the samples that cannot be explained by any of the known fault classes. The leakage fault f_{iml} is introduced as an unknown fault f_x.

The results are shown in Fig. 8 where the MB classifier gives the same ranking of the known fault classes as in the closed set scenario. The leakage fault f_{iml} receives a low ranking as an unknown fault. This can be explained by that the residual outputs from the fault f_{iml} are similar to f_{ypim} and f_{ywaf}, see Fig. 1. An interesting observation is that the 1SVM classifier, which models the training data support for each fault class degrades for realizations of fault f_{ypic} that deviate from training data which instead are identified as a likely unknown fault. This means that if training data only contains realizations from a few fault magnitudes, fault ranking using 1SVM does not generalize well and samples from other magnitudes of the same fault class will likely be classified as an unknown fault. The leakage data is ranked higher as an unknown fault compared to the result of the MB classifier.

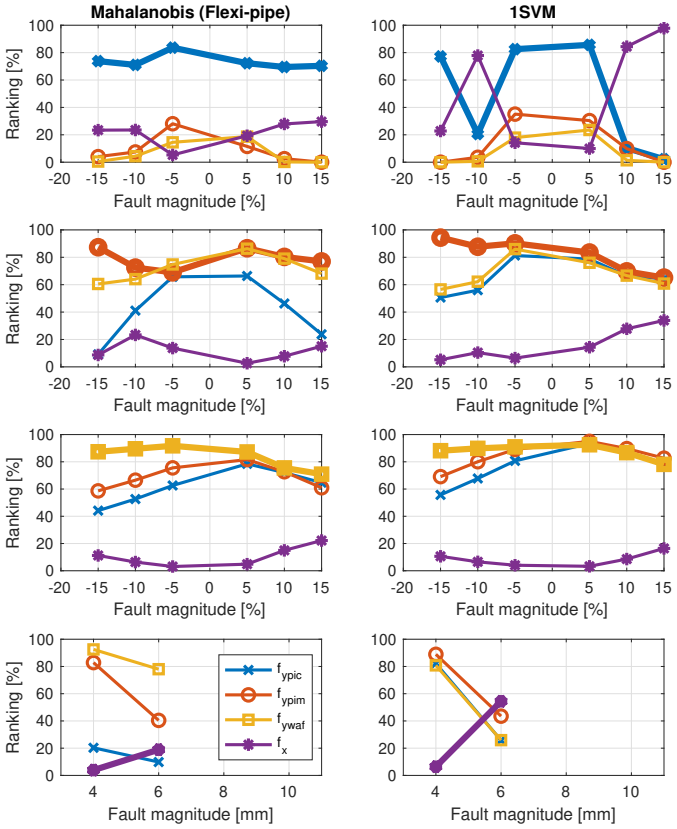

Fig. 8. Evaluation of open set fault classification problem. The different plots show the ranking of different fault classes, including the unknown fault class, for the different classifiers as a function of fault size where the true fault class is marked in each subplot.

8. CONCLUSIONS

The complicating factors of the fault diagnosis problem motivates the use of open set classification principles that can handle imbalanced data and overlapping classes. Experiments using both simulated and engine data show that the proposed flexi-pipe model is suitable for residual-based fault classification since it captures the variability of residual data, even though training data is limited. Results show that the flexi-pipe modeling approach generalizes better than the conventional classifier approaches, both in the closed set and open set fault classification scenarios.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Max Johansson for his insightful inputs during the development of the flexi-pipe model.

REFERENCES

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Jung, D. (2020). Residual generation using physically-based grey-box recurrent neural networks for engine fault diagnosis. *arXiv preprint arXiv:2008.04644*.

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

Khorasgani, H. and Biswas, G. (2018). A methodology for monitoring smart buildings with incomplete models. *Applied Soft Computing*, 71, 396–406.

Lundgren, A. and Jung, D. (2022). Data-driven fault diagnosis analysis and open-set classification of time-series data. *Control Engineering Practice*, 121, 105006.

Luo, J., Namburu, M., Pattipati, K., Qiao, L., and Chigusa, S. (2009). Integrated model-based and data-driven diagnosis of automotive antilock braking systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(2), 321–336.

Murphy, K. (2012). *Machine learning: a probabilistic perspective*. MIT press.

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

Sankavaram, C., Kodali, A., Pattipati, K., and Singh, S. (2015). Incremental classifiers for data-driven fault diagnosis applied to automotive systems. *IEEE access*, 3, 407–419.

Scheirer, W., de Rezende Rocha, A., Sapkota, A., and Boult, T. (2013). Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7), 1757–1772.

Slimani, A., Ribot, P., Chanthery, E., and Rachedi, N. (2018). Fusion of model-based and data-based fault diagnosis approaches. *IFAC-PapersOnLine*, 51(24), 1205–1211.

Williams, C. and Rasmussen, C. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.