Enhancing the diagnostic performance of Condition Based Maintenance through the fusion of sensor with maintenance data

MSc. Henrik Simon[1] and Sascha Schoenhof[2]

[1,2]*Institute of Flight Systems and Automatic Control, Technical University Darmstadt,*
Otto-Berndt-Str. 2, 64287 Darmstadt, Germany
simon@fsr.tu-darmstadt.de

ABSTRACT

Although Maintenance data is crucial for authoritative reporting reasons and is generally used to optimize maintenance planning in terms of budget, scheduling and logistics, the potentials of the implicit given information for Prognostics and Health Management (PHM) frameworks are not yet completely leveraged. Traditional PHM frameworks typically rely exclusively on sensor data to derive a system's health status, while maintenance, repair and overhaul (MRO) data is not investigated. However, maintenance data contains valuable information on which part of a system is checked, serviced or replaced. In the presented work, a novel approach to fusion maintenance data into a traditional (sensor-based) PHM/condition monitoring framework is introduced. This fusion enables a model update of the condition monitoring framework and hence improves its diagnostics performance in terms of classification accuracy. The presented work uses data from a simulation framework to develop and evaluate the method. A sensitivity analysis shows influences of various sources of uncertainty and constraints of the approach. First results do not show significant improvements compared to a benchmark approach, but the variety of setting parameters in the simulation environment and their influence on uncertainty are subject of further research.

1. INTRODUCTION AND PROBLEM DESCRIPTION

1.1. Introduction

Modern systems, and especially aircraft systems, are becoming more and more complex. With increasing number of sensors for supervising systems and increasing system complexity, the amount of available information for diagnostic purposes is increasing and the diagnostic process itself is getting more complex as well.

One important source of information is maintenance (MX) data. While this data is collected and archived since decades and typically used for reliability engineering and maintenance effectiveness analysis, its potentials for the enhancement of condition monitoring frameworks are not yet completely leveraged. One of the reasons, why this data is stored initially in the aviation industry, is of authoritative nature: to comply with EASA Part-M / Part 145 regulation (European Union, 2014). Furthermore, the data is only collected, when damages are already present while the initial activation point in time is unknown and hence it is not trivial to build prognostic methods based on it. On the other side, especially in case of MRO companies, these data contain the core know-how of the organization, since mechanics and engineers write down their experiences, diagnoses and actions.

In the presented work, a new approach to include the valuable information of MX data into condition monitoring frameworks is proposed. The main hypothesis is that this information gain will reduce fault classification uncertainty and will enable a model update through maintenance evidences.

The remainder of this paper is structured as follows: Section 1 gives an introduction into the topic, the research problem and the state of the art. Section 2 describes the methodology to fusion MX data into a traditional monitoring framework and the corresponding preprocessing steps. In Section 3, a simulative framework and use case is introduced. The paper closes with the discussion of the results in section 4 and conclusions and an outlook on further research in section 5.

1.2. Problem description

The core idea of PHM and condition monitoring frameworks is to enhance the understanding of the systems' health status through constant monitoring and to enable condition based maintenance (CBM). Process models like the Open System Architecture for Condition Based Maintenance (OSA-CBM), or the Cross Industry Standard for Data Mining (CRISP-DM) propose structures and guidelines for the development of

PHM methods. The basic assumption is, that the steps for developing such methods are sequential, starting with a data acquisition and data/problem understanding phase, followed by the diagnostic and prognostic step and finishing with the advisory generation and health management step.

The presented work focusses on the diagnosis part, proposing a new methodology for improving the performance (accuracy) and reducing the uncertainty of damage classification problems. The prognostic step and Remaining Useful Lifetime (RUL) estimation is explicitly not part of this work. The research in the field of PHM is manifold. In this paper we will focus on two specific aspects, before presenting the new approach: On the on hand classical PHM and monitoring frameworks based on time-series (sensor) data and on the other hand the usage of event and MX data, mostly used for planning optimization with a multitude of different goals.

The constant monitoring of the systems health status (and thus the basis for any prognostic efforts) requires up-to-date information from the system itself. In most cases, this is realized through sensors, which deliver data, which is then processed in models and gives indication for the *systems health* status. The research in this context is mainly focused on data based methods with strong focus on RUL prognosis in the recent years and decades. The availability of increasing amounts of data and computing resources enables the usage of artificial intelligence and machine learning.

In (Nguyen and Medjaher, 2019) a Long Short-Term Memory (LSTM) network is presented to build a new dynamic predictive maintenance framework based on sensor measurements. Susto (Susto et al., 2015) proposes a multiple-algorithm classification approach with different prognostic horizons as decision support. With higher regard to environmental and operational parameters influencing maintenance processes, (Reder et al., 2018) are looking for approaches to model weather influences on wind turbine failures and (Verhagen and Boer, 2018) proposes time-dependent proportional hazard models to identify operational factors influencing maintenance event occurrences. Also in the context of wind turbines, (Asgarpour and Sørensen, 2018) present a Bayesian approach for a prognostic model and degradation monitoring. The usage of Decision Trees and Neural Networks and a general review of various PHM algorithms are discussed by (Carvalho et al., 2019) and (Accorsi et al., 2017). The widespread problem of high class imbalance through un underrepresented damage class is analyzed and discussed by (López et al., 2013).

Looking deeper into probabilistic approaches, Bayesian statistics and networks are often used to build diagnostic models. In (Ashasi-Sorkhabi et al., 2017), the authors propose a CBM implementation based on Bayesian statistics for a train gearbox, (Zhang et al., 2014) propose a Bayesian Belief Network for the lifetime prediction of bearings, (D.

Huang et al., 2020) use Bayesian Neural Networks for RUL Prediction in Aircraft Engines.

While most publications focus on the development and application of methods for individual components, systems with interacting degradation effects are often not considered. In (Bian and Gebraeel, 2014) a Bayesian based model is created to estimate remaining useful lifetimes of multi-component systems with interaction of degradation effects. A dynamic reliability assessment for systems of systems through a state model is proposed by (Heier et al., 2018).

The second source of data of particular interest for this work, is MX data. This data is categorized as time-discrete temporal data or event based data. The individual events do happen independently of each other and do not follow a constant time interval. MX data analysis is typically used for optimization of MX intervals, budget, logistics or personnel. In this context, (Dinis et al., 2019) built a maintenance capacity planning algorithm based on Bayesian Networks for the use in CBM, (Eickemeyer et al., 2013) describe the use of Bayesian Networks for the estimation of workload in maintenance processes and (Jones et al., 2010) analyze the influence of various parameters onto the failure rates of systems. The necessary infrastructures and architectures for collecting and processing of MX data are discussed in (Albano et al., 2018) and (Mourtzis and Vlachou, 2018), where modern Cyber Physical Systems and Cloud architectures for Predictive Maintenance applications are discussed.

The collection of maintenance data, especially in the aviation industry, comes with its own complications: One of the core challenges in the large scale use of this data lies in the presence of natural language, which is used to describe a given problem, identified findings and maintenance actions. Maintenance, Repair and Overhaul (MRO) companies, like Lufthansa Technik AG, collect large amount of MX data, which are hardly analyzed automatically. Current research topics in the industry are therefore investigating the potentials of Natural Language Processing (NLP) on MX data. In (Dixit et al., 2021) and (N. B. Niraula et al., 2020) NLP methods are applied in order to extract information like specific parts and conditions from the MX entries, enabling automatic diagnostics. In (Welz et al., 2017; Welz, 2017), the author proposes the use of maintenance action dependent models to incorporate imperfect maintenance actions. In (Martorell et al., 1999) an age-dependent reliability model is introduced incorporating both working conditions and maintenance effectiveness, quantifying the effects on the equipment age of a nuclear power plant. Similarly (Chuang et al., 2020) proposes a new condition based maintenance approach under imperfect maintenance for slowly degrading, continuously monitored systems, including the effect that systems do not restore to a as-good-as-new state after maintenance.

The preprocessing of event based data, like MX data, is one of the core challenges in the presented work. Several

approaches to align event data with time series data are described by (Korvesis et al., 2018), looking explicitly into the preprocessing of Post Flight Reports for diagnostic and prognostic purposes. In (Same and Govaert, 2012) a sequential time series segmentation approach is introduced and (Yu et al., 2011) compare Period-based and Event-driven through a Bayesian prediction model.

2. METHODOLOGY OF THE DIAGNOSTIC MODEL

In the context of monitoring and prognostics, maintenance and inspection information are often used for the data labeling, post-event diagnostics or the selection of use cases for PHM methods. Monitoring and performance assessment methods only rely on sensor data. Labels, which are given through maintenance data, are used to train and validate models and to evaluate the results. In the proposed methodology however, maintenance data will be used in the forward modelling phase of a condition monitoring framework. Hence, the information, which is inherently given in the maintenance data, will be fed to the diagnostic model. The key to achieving this, lies in the correct preprocessing of the maintenance data. To the author's best knowledge, this is the first time that this fusion approach is investigated and compared to a benchmark.

The general idea proposed in the paper and the evaluation process is shown in Figure 1. The traditional processing path is shown on the left, where the sensor data is preprocessed and then fed to a learning model leading to the diagnosis, which is understood as a classification problem: A damage will be classified as active or inactive based on sensor data. The new methodology, shown on the right, includes MX data into this approach, which needs an advanced preprocessing in order to align the temporal data structure with the time series. This leads to a different (ideally more accurate) classification result. The delta between the benchmark and the MX-fusion approach, quantified through classification metrics, can be directly linked to the additional input data. The prosed approach is throughout this paper referred to as MX-fusion approach.
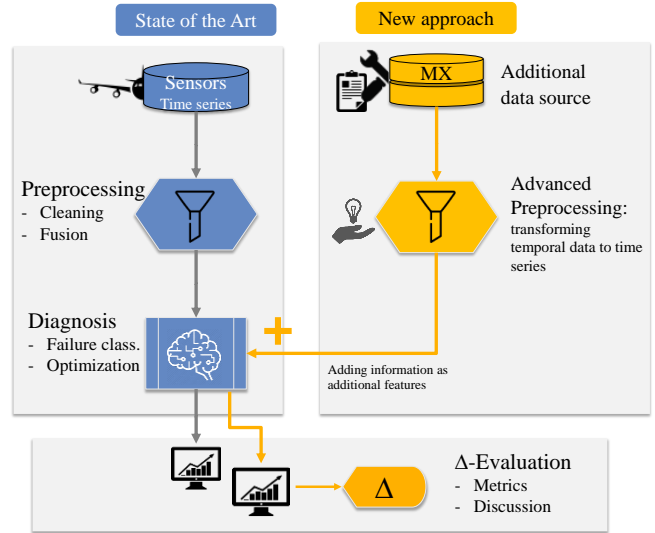

Figure 1: Concept of fusion maintenance data with sensor data

The input data is two-fold: The benchmark algorithm is only considering sensor data from the observed systems directly. The secondary data source is MX data. An exemplary plot of the sensor data, derived in the case study in section 3, is shown in Figure 2. The data shows the typical degrading pattern over time, with sudden jumps, which mark the exchange or repair of damaged components.
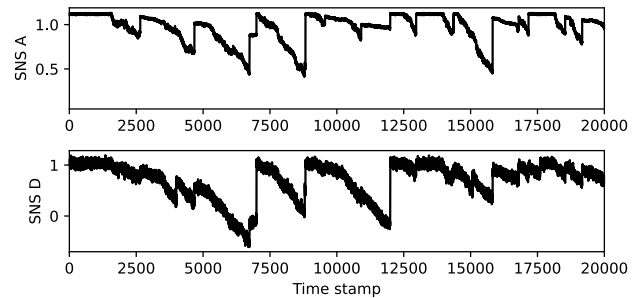

Figure 2: Sensor data over time

The sensor data is preprocessed in order to cope for the evolution of the signal over time. Therefore *time lag* features are taken, meaning that new input features are generated from sensors signals of previous time steps. The interval is varied from 5 to 50 datapoints per sensor signal. The rest of the standard preprocessing steps, like cleaning and smoothing the data, is done through the simulation framework (see section 3).

One of the key challenges to be able to include maintenance data in a monitoring framework is the correct preprocessing: The time-discrete, temporal data have to be aligned with the sensor data and thus have to be translated to time-series data. In a first setup, which is presented in the scope of this work, the event data is preprocessed in such a way, that it can be added to a classification algorithm, which is based on time

series data. In essence, new feature columns are generated and forwarded to the algorithm. The method used for this will be explained in the following.

Maintenance data logs typically consist of a timestamp, categorical mapping parameters (Air Transport Association (ATA) chapters[1] in case of aviation industry), event-ID, an optional reporting, action and finding. An example for a (simplified) maintenance log is shown in Table 1. The timestamp marks the exact time of the maintenance action (and can furthermore include the exact time of the damage occurrence), the categorical mapping parameters allows a rough information on where the damage is located in the system, in aviation ATA codes are used denoting broad categories such as landing gear (ATA 32) or engine (ATA 72). The reporting column contains a problem description and can be filled automatically (fault log), or through personnel with natural language (by pilot, cabin crew or maintenance personnel). The action and finding columns contain information of the maintenance event itself, also mostly reported in natural language. The analysis of the natural language leads to uncertainties in the labelling process, but will not be part of this work. For the purpose of this paper, only timestamp and findings are of interest, assuming that the exact finding can be taken from the technical logs.

Table 1: Maintenance log example

Time stamp	…	ATA	Reporting	Finding/ Action
2021-04-01 15:56	Sys ID	32-50	Landing gear steering issue	bolt broken, replaced
2021-04-05 02:08	Event ID	79-11	/	EOS performed
2021-04-12 21:35	Comp ID	25-35	Galley latch loose	Latch adjusted
2021-04-23 01:04	…	29-11	Hyp pump failure	Perform tests
…	…	…	…	…

In the following, the methodology of preprocessing of the MX data will be explained. It is based on (Korvesis et al., 2018) and adapted slightly, to incorporate for the special characteristics of MX data. Figure 3 shows a visual interpretation of the segmentation process. DMG_x, DMG_y, DMG_z represent three possible damages, their active signal is shown in blue, values of one indicating the damages is active, zero it is inactive. Below each Damage-Signal plot, "x" marks, when a damage is detected and repaired. Our goal is to find the damage $d_? \in DMG$, with $DMG = \{d_1, d_2, ..., d_k\}$ being the set of possible damages. We start by dividing the dataset into individual episodes ep_i^j which correspond to the time between two consecutive (target) damages. Each episode will later be used for training process separation, meaning that each episode represents one

training data set. The figure shows the episodes corresponding to DGM_x. In the next step, the damages within each episode are aggregated over the length of constant time segments seg_t, marked by vertical dashed lines in the figure. The length of the segments is not fixed to the damages and adapted to the problem, in our case a length between 500 and 2000 time steps is considered.
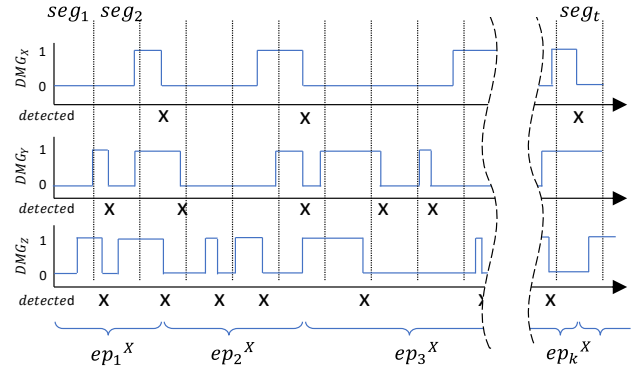

Figure 3: Preprocessing step of maintenance data

Segments are represented by vectors $x \in \mathbb{N}^{|D| \times 1}$ containing the sum of all damages which occurred in the equivalent segment. The segments are identified by the application of the sliding window method in order to oversample the low amount of damages. The main difference compared to (Korvesis et al., 2018) would be, that we take damage occurrences (maintenance actions) instead of fault logs, which will lead to uncertainties (further explained in section 2.3). Additionally to this segmentation approach, a more simple feature is introduced: The time since the last repair, which is just cumulating the time steps since the last maintenance action, comparable with mean time between failure (MTBF). Third, the cumulative sum of undertaken checks after a damage detected/reset event is taken as feature. This should further incorporate the idea of a rising damage probability over time.

Classification

The problem is developed as a multi class classification problem. There are many off the shelf algorithms, which can be used for this. In the first setup, which is presented in this work, random forest (Breiman, 2001), Multi-Layer Perceptron Neural Network and Support Vector Machine classification algorithms are established and compared with each other. Since the development of the algorithm is not in the focus of this work, but the preprocessing methodology and general idea behind the data fusion, no distinct optimization of the algorithms through hyper parameter tuning is done.

Since multiple damages can be active at the same time, the problem is furthermore defined as a *multilabel* classification

---

[1] https://en.wikipedia.org/wiki/ATA_100

problem. This leads to the circumstances, that in praxis, one classifier is fitted for every damage (class).

2.1. Primary contribution

The primary goal of the methodology is to optimize the diagnostic performance of the monitoring framework. Optimization is in this case understood as increasing the overall accuracy and reducing the uncertainties. The resulting research hypothesis is, that an additional information source will be beneficial for the classification algorithm.

2.2. Secondary contribution

The secondary goal is the fundamental analysis and critical discussion of all influencing parameters on the whole maintenance complex. Since the data to build and assess the methodology comes from a generic simulation framework, influences of uncertainties and noise as well as confidences on the results of the fusion methodology can be assessed.

Hence, one goal of this research is to fully understand, quantify and describe the influence of the various sources of uncertainties, concerning the sensor data:

- The number and placement of available sensors.
- The resolution (time and amplitude) of the sensor signal.
- The influence of noise/uncertainty in the sensor signal.

And concerning the maintenance data:

- Faulty labeling during the logging process (not investigated yet).
- Uncertainty due to faulty labeling during NLP (assuming automatic processing of MX logs).
- Influence of inspection interval variations.
- Interaction of multiple damages.
- Difference of routine and non-routine events in the processing setup.
- Ambiguity due to corrective v/s preventive actions taken in the maintenance.
- Ambiguity due to action taken in maintenance was to address the symptom but not the root cause

From these investigations, general guidelines to the data bases and data quality can be derived: For example for a MX operator who is establishing NLP for MX records, it would be very beneficial to quantify a minimum label confidence, which is necessary for useful postprocessing of automatically analyzed and classified maintenance data. This is one of the secondary goals, which are part of ongoing research and planned to be presented in 2022.

2.3. Assumptions and limitations

As for all simulation frameworks, several simplifications were made. Thus the following assumptions and limitations have to be defined to the investigated use case:

- A slowly degrading system is assumed (no random failures).
- The system is component-based (line replaceable units).
- Perfect maintenance is assumed, components are replaced with brand new material
- A minimum amount of system complexity and damage interaction is required, as the methodology is not expected to improve results, when only monitoring single components.
- A maintenance history is generally available: Rather short inspection intervals with reliable findings.

Requirements to the input data:

- Sensor signals are available.
- Component changes with findings (i.e. labels, which damages occurred) are available.
- Mapping of damages to components is generally possible.

Requirements to the methodology:

- Only diagnosis (no RUL prognosis) phase.
- The problem is understood as classification setup, multiclass, meaning there are more then two classes (more than one damage) and multilabel, meaning that one time step can include one or more damages being active.
- A probabilistic approach with confidence interval is desired.
- For the first step we take the MX findings as ideal and trustworthy. In practice, however, one would have to extract the findings from the maintenance logs through NLP, which comes with uncertainties.

3. CASE STUDY / NUMERICAL ANALYSIS

In the following, the underlying data basis will be presented and the case study will be formulated. The data for the development and evaluation of the methodology comes from a MX simulation framework, which will be presented at first.

3.1. Maintenance simulation framework

For the generation of data, a generic simulation framework was developed. The benefits of the development inside a simulation framework are, that the complexity, general noise of the data and the uncertainties can be adapted, allowing to define exact limits of the methodology and to derive requirements concerning the real world data.

The simulation framework consists of four main objects and a system class to combine all objects: sensors, components, damages and inspection/maintenance events. Every object has dependencies to the others and multiple parameters for the tuning of noise and uncertainty. A simplified UML-Diagram is shown in Figure 4.
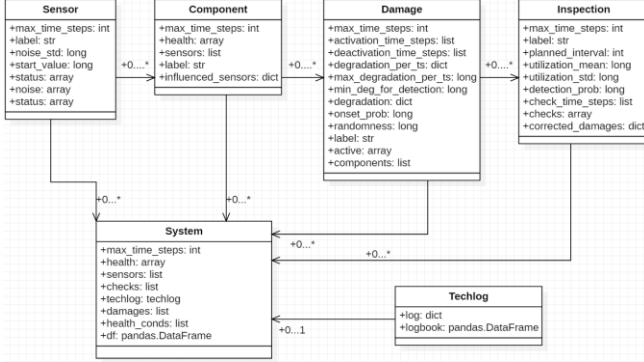

Figure 4: Simplified UML-diagram of the mx-simulation framework

At the initialization, the duration N of the simulation as well as the four main objects sensors, components, damages and inspections are getting defined by the user and fed to a system class, which defines their integration and runs the simulation. Each timestep t of the simulation is defined by an integer. The initial value of a sensor s is drawn from a normal distribution around a reference value and is constant. Subsequently, the signal has assigned a noise value $r_s(t)$ at each timestep. Sensors can be simultaneously affected either positively or negatively by multiple components. The influence of a component c on the sensor s is denoted as proportional factor w_c^s and thus can be set to zero if the component does not influence the sensor at all. One component can also influence several sensors differently. After again adding uncertainty $\tilde{r}_s$ through noise, the value of the sensor s at timestep t is given by eq. (1):

$$s(t) = s(0) + r_s(t) + \sum_c \pm c(t) \cdot w_c^s + \tilde{r}_s(t). \qquad (1)$$

A component c has a starting value of one and can only decrease to a minimum of zero hence reflecting the condition of the component. Each component c degrades due the influence of degradation v_d^c caused by damage d which is modelled by a random walk behavior. This is a further simplification, as the sensor signal (degradation) is directly related to the component health. However, especially vibration sensor signals are underlying a more complex relation to components in real world. More sophisticated models are part of ongoing research. An uncertainty $|r_c(t)|$ is additionally added. Thus, resulting in eq. (2) for the value of component c and timestep t:

$$c(t) = 1 - \sum_d \sum_0^t d(t) \cdot v_d^c - |r_c(t)|. \qquad (2)$$

A damage d can only have a value of zero, indicating a disabled damage, or one, indicating an active damage. A damage d is randomly activated based on a user-specified probability p_d and remains active for upcoming timesteps until it is deactivated by an inspection M. An inspection can either investigate entire components, observing the amount of degradation caused by each damage specifically influencing this component, or damages, observing their influence on the degradation of each system component. An inspection takes place at intervals specified by the user although the exact timing is also subject of uncertainty. A damage can only be detected by an inspection if it is active and has already led to a minimum degradation since the last time it was deactivated. Even if these conditions are met, there is still a probability p_M that the damage d will not be detected by an inspection M. Both, the minimum degradation, as well as the probability p_M can be predefined by the user. If a damage is detected, it is deactivated. The degradation of components caused by the deactivated damage is set to zero indicating a repair of the component. The incident is documented in a Techlog object.

3.2. Simulation setup

The framework is completely generic and thus almost every technical system can be reproduced. For first results, a fictious system is designed. The different objects and causal relations are shown in Figure 5. The setup consists of three sensors, two components, two damages and two inspections. Several levels of interaction, especially of components on sensors, are implemented. The thickness of the connecting lines mark the qualitative size of influence.
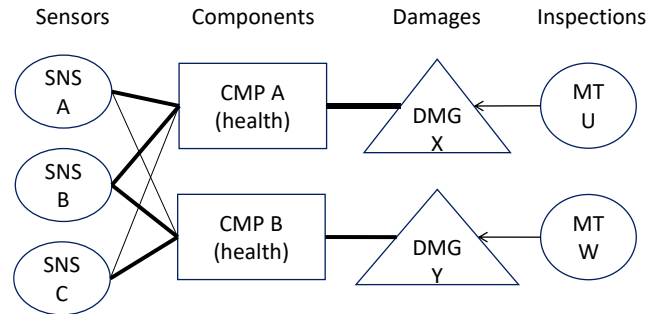

Figure 5: Fictious system simulation design

The setup, which is shown above, is run in a Monte-Carlo like setting. The resulting dataset consists of 100 simulation runs, each with small variations in the input parameters. An example plot with two sensor signals and two damages with corresponding inspections/repairs is shown in Figure 6. The top plot shows the sensor signal of *SNS A* over time, the plot below (second plot) shows the equivalent damage *DMG X*. Values of one corresponding to an active damage, values of zero to an inactive damage. The same is shown in the two lower subplots for sensor *SNS B* and damage *DMG Y*.
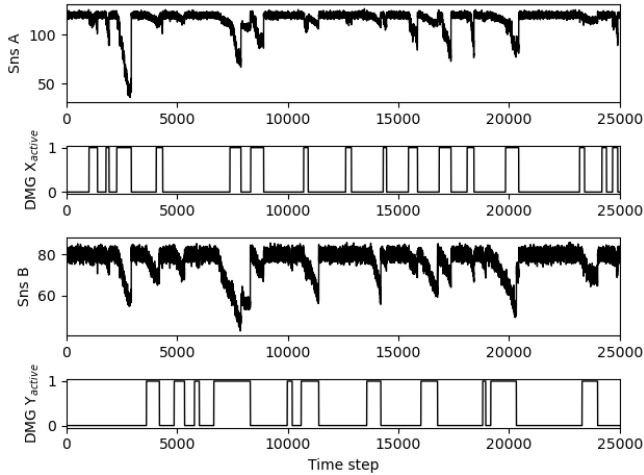
Figure 6: Plot of sensor signals and damage activation for the case study

The data is split into training and validation data. Since multiple simulation sets are available, 80% will be used for training, and 20% are used for validation. The data will be trained only on time steps, were inspections happened. This is important, since this ensures a oversampling of the damage classes, but it reduces the dataset tremendously at the same time. Inspections do not necessarily mean, that an active damage is present. The share of active damages is about 10-20% in the training set, depending on the setting. Due to false negative inspections (a damage is not detected, even though it is active), errors in the training data may be present. Alternatively the complete dataset could be used, which would require methods to address the class imbalance (the share of detected damage entries goes below 0.1%). For the validation set, the whole dataset with no over- or undersampling is taken. The resulting share of active cases varies between 5% and 40%, depending on the damage and the setup.

3.3. Classification results

For the classification a Support Vector Machine (SVM) is used. The training set consists of 1600 datapoints. In case of the benchmark approach, only sensor raw sensor values of the last 10 time steps are used, results in 30 input features. For the MX-Fusion approach, the same sensor values are used plus additional features as described in chapter 2, resulting in 36 input features. The SVM is not further tuned or optimized.

The classification results of the basic setup are shown over time in Figure 7. The two subplots show the results for the individual damages. The true damage activation profile is shown as the solid black line, indicating whether the damage is active or not. The classification results are indicated as vertical bars: blue bars marking classification results of the benchmark approach, orange bars marking results of the MX-fusion approach. True positives are when the damage activation profile is *active* and bar plot of the benchmark or

MX-fusion are also positive. False positives are given when the true signal is healthy, but bar plots are positive, false negatives are given when the true signal is active, but classification results are healthy.
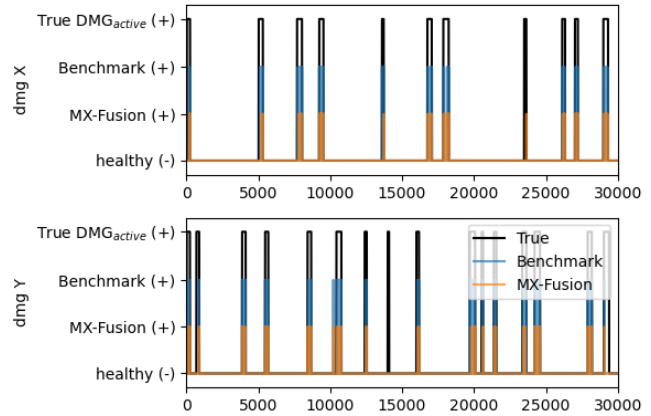

Figure 7: Classification results

The plots over time do not allow an interpretation of the results. The corresponding confusion matrices are shown in Figure 8.
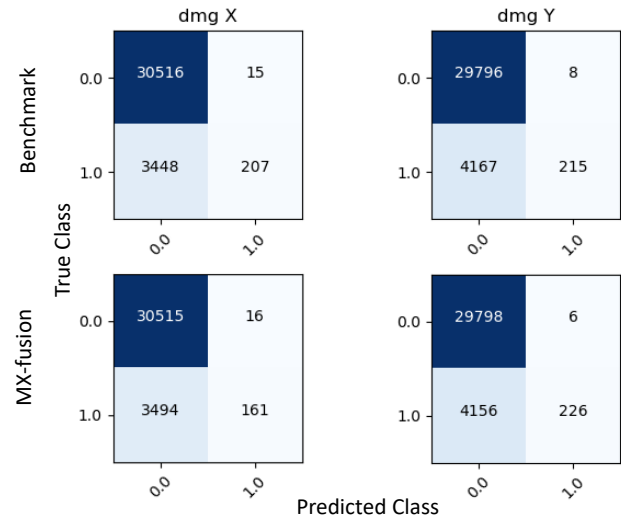

Figure 8: Confusion matrices for the classification results

In the next section the results will be evaluated and discussed based on evaluation metrics.

4. DISCUSSION

The discussion of the results is based on error metrics. For the evaluation of the methodology, we use standard metrics for classification (Saxena et al., 2008):

- Accuracy: $\frac{TP + TN}{TP + TN + FP + FN}$

- Precision: $\frac{TP}{TP + FP}$

- Recall: $\frac{TP}{TP + FN}$

- F1-score: $2 \cdot \frac{precision \cdot recall}{precision + recall}$

The results are evaluated for each damage class, resulting in three metrics per classification run. Even though multiple different setups and variations where investigated, we will only present one result here, being representative for the overall results. The classification results of the basic setup, shown in Figure 7, are listed in Table 2. The metrics show no significant change between the two classification setups. As can be seen from the evaluation metrics, there is only a minor increase in Precision and F1-Score for damage X, otherwise, the accuracy and metrics are poorer for the MX-fusion approach.

Table 2: Diagnostic results in comparison

Method	Accuracy	Precision	Recall	F1-score
DMG_X benchmark	0.89	0.93	0.06	0.11
DMG_X MX-fusion	0.89	0.91	0.04	0.08
DMG_Y benchmark	0.87	0.96	0.05	0.09
DMG_Y MX-fusion	0.88	0.97	0.05	0.10

In general, the overall accuracy of the classification results is mainly driven by the sensor data. Additional MX event data does not significantly influence the results. If there is an influence, most of the time, it is deteriorating compared to the benchmark approach. In a few cases, the results with MX-Fusion are getting better: small increases in F1-Score and Precision can be seen. However, these are cases where the benchmark performance is generally very poor and the algorithm is severely underfitted. The event data can increase robustness in those cases, enabling a minimal amount of classification capability. Especially when taking all available training data points for the algorithm fitting process into account (no undersampling of the healthy class through inspection selection), all datapoints are classified as healthy (false negative) in the validation phase.

A sensitivity analysis of the most obvious influencing parameters of the damage onset probability, the sensor noise and the inspection interval is shown in figure XY. In total, six plots are showing the Accuracy and F1-score of three different parameter variations. In these graphics we focus only on damage Y, blue stars marks the benchmark approach, orange squares the MX-fusion approach. For the reason of clarity, no Precision or Recall metrics are shown in these plots. The default setup (while the other parameter is varied) is the damage onset probability 0.0009, sensor noise set to 1 and the maintenance interval set to 400. We can see, that the overall accuracy decreases with increasing damage onset probability and rising sensor noise and has a maximum at maintenance intervals between 300 and 800 time steps, while decreasing at shorter and longer intervals. Concerning the F1-score we see that there is a slightly increasing trend with increasing damage onset probability (mainly due to increasing Recall values), a clear decreasing trend with

increasing sensor noise and a inversed trend for the maintenance interval compared to the accuracy.
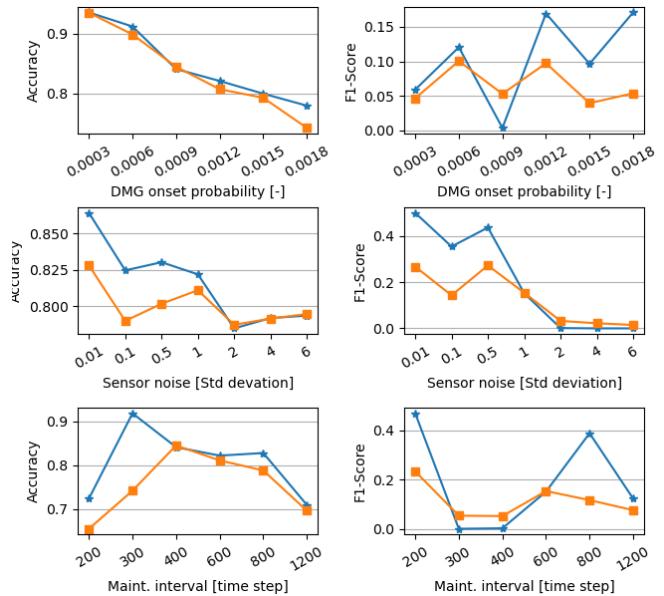

Figure 9: Sensitivity analysis of damage onset probability, sensor noise and maintenance interval

The reasons for the results are manifold: The first source of error and uncertainty lies in the preprocessing step. The preprocessing of the event data is rather rudimentary and more sophisticated methods are being developed. Furthermore, the influence of the window length on the methodology needs to be further analyzed. Additionally, there is a variety of sources of uncertainties in the simulation framework: The noise of the sensor data, the complexity of the system setup, the interaction of several damages, the onset probability of damages and the inspection intervals. Only a few parameters were varied in small scale already: the noise of the sensor signals, the amount of damages and the interactions between components and multiple sensors.

5. CONCLUSIONS AND OUTLOOK

This work presented the first approach to an overall view on the interactions of maintenance data and sensor information for diagnostic frameworks. A first attempt of improving the condition monitoring diagnostic accuracy and a damage classification framework through the fusion of maintenance data was presented. The results do not show a significant increase in accuracy or confidence compared to a benchmark approach based only on sensor data.

The classification results comply with the general problem of applying classification algorithms onto real world problems: Due to the high class imbalance, which is in the nature of a failure monitoring problem, the general accuracy and quality of the results can be poor. One point for further research is therefore the formulation as a regression problem.

Two major topics are part of ongoing research and results will be presented in 2022: On one side, more sophisticated preprocessing approaches for the MX data are in development, and on the other side, the influence of maintenance data quality (also including credibility and label confidence) will be further analyzed. Furthermore, probabilistic approaches, such as (Dynamic) Bayesian Networks and Hidden Markov Models for the classification problem will be investigated.

Further research is also needed to understand the influences of the numerous variation parameters of the simulation framework: the noise and uncertainty of sensor data in combination with corresponding maintenance events, the complexity of the system setup, the interaction of several damages, the onset probability of damages and many more. The field of analysis and further research in this context is wide.

Furthermore, a Bayesian approach seems interesting in this context, allowing a quantification of the class probability and general confidence of the classification. Furthermore, additional features could be added rather easily in a Bayesian Network analyzing setup. This kind of classification setup is in development.

Finally, the assessment through the simulation framework allows to derive requirements to the data quality for MRO internal diagnostic algorithms: For example, the necessary label confidence of automatic maintenance report classification. To investigate this topic, the setup needs to be extended by the uncertainty of technical logbooks and routine and non-routine maintenance problems.

REFERENCES

Accorsi, R., Manzini, R., Pascarella, P., Patella, M. and Sassi, S. (2017) 'Data Mining and Machine Learning for Condition-based Maintenance', *Procedia Manufacturing*, Vol. 11, pp.1153–1161.

Albano, M., Jantunen, E., Papa, G. and Zurutuza, U. (2018) 'The MANTIS Book: Cyber Physical System Based Proactive Collaborative Maintenance', River Publishers, Aalborg.

Asgarpour, M. and Sørensen, J.D. (2018) 'Bayesian based Prognostic Model for Predictive Maintenance of Offshore Wind Farms', *The International Journal of Prognostics and Health Management (IJPHM)*, Vol. 9, No. 1, p.10.

Ashasi-Sorkhabi, A., Fong, S., Prakash, G. and Narasimhan, S. (2017) 'Condition Based Maintenance Implementation for an Automated People Mover Gearbox', *The International Journal of Prognostics and Health Management (IJPHM)*, Vol. 8, No. 3.

Bian, L. and Gebraeel, N. (2014) 'Stochastic modeling and real-time prognostics for multi-component systems with degradation rate interactions', *IIE Transactions*, Vol. 46, No. 5, pp.470–482.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, Vol. 45, No. 1, pp.5–32.

Carvalho, T.P., Soares, Fabrízzio A. A. M. N., Vita, R., Da Francisco, R.P., Basto, J.P. and Alcalá, S.G.S. (2019) 'A systematic literature review of machine learning methods applied to predictive maintenance', *Computers & Industrial Engineering*, Vol. 137 [online] https://www.sciencedirect.com/science/article/pii/S0360835219304838.

Chuang, C., Ningyun, L., Bin, J. and Yin, X. (2020) 'Condition-based maintenance optimization for continuously monitored degrading systems under imperfect maintenance actions', *Journal of Systems Engineering and Electronics*, Vol. 31, No. 4, pp.841–851.

D. Huang, R. Bai, S. Zhao, P. Wen, S. Wang and S. Chen (2020) 'Bayesian Neural Network Based Method of Remaining Useful Life Prediction and Uncertainty Quantification for Aircraft Engine' in *2020 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, pp.1–8.

Dinis, D., Barbosa-Póvoa, A. and Teixeira, Â.P. (2019) 'Valuing data in aircraft maintenance through big data analytics: A probabilistic approach for capacity planning using Bayesian networks', *Computers & Industrial Engineering*, Vol. 128 [online] http://www.sciencedirect.com/science/article/pii/S0360835218304856.

Dixit, S., Mulwad, V. and Saxena, A. (2021) *Extracting Semantics from Maintenance Records*. https://arxiv.org/pdf/2108.05454.

Eickemeyer, S.C., Borcherding, T., Schäfer, S. and Nyhuis, P. (2013) 'Validation of data fusion as a method for forecasting the regeneration workload for complex capital goods', *Production Engineering*, Vol. 7, 2-3, pp.131–139.

European Union (2014) *Commision Regulation (EU) No 1321/2014. Easy Access Rules for Continuing Airworthiness,* Brussels.

Heier, H., Mehringskotter, S. and Preusche, C. (2018) 'The use of PHM for a dynamic reliability assessment' in *IEEE Aerospace Conference*, IEEE, IEEE, pp.1–10.

Jones, B., Jenkinson, I., Yang, Z. and Wang, J. (2010) 'The use of Bayesian network modelling for maintenance planning in a manufacturing industry', *Reliability Engineering & System Safety*, Vol. 95, No. 3 [online]

http://www.sciencedirect.com/science/article/pii/S0951832009002518.

Korvesis, P., Besseau, S. and Vazirgiannis, M. (2018) 'Predictive Maintenance in Aviation: Failure Prediction from Post-Flight Reports' in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, IEEE, IEEE, Piscataway, NJ, pp.1414–1422.

López, V., Fernández, A., García, S., Palade, V. and Herrera, F. (2013) 'An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics', *Information Sciences*, Vol. 250, pp.113–141.

Martorell, S., Sanchez, A. and Serradell, V. (1999) 'Age-dependent reliability model considering effects of maintenance and working conditions', *Reliability Engineering & System Safety*, Vol. 64, No. 1 [online] https://www.sciencedirect.com/science/article/pii/S0951832098000507.

Mourtzis, D. and Vlachou, E. (2018) 'A cloud-based cyber-physical system for adaptive shop-floor scheduling and condition-based maintenance', *Journal of Manufacturing Systems*, Vol. 47 [online] https://www.sciencedirect.com/science/article/pii/S0278612518300700.

N. B. Niraula, A. Kao and D. Whyatt (2020) 'HPart and Condition Extraction from Aircraft Maintenance Records' in *2020 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp.1–7.

Nguyen, K.T. and Medjaher, K. (2019) 'A new dynamic predictive maintenance framework using deep learning for failure prognostics', *Reliability Engineering & System Safety*, Vol. 188, pp.251–262.

Reder, M., Yürüşen, N.Y. and Melero, J.J. (2018) 'Data-driven learning framework for associating weather conditions and wind turbine failures', *Reliability Engineering & System Safety*, Vol. 169, pp.554–569.

Same, A. and Govaert, G. (2012) 'Online Time Series Segmentation Using Temporal Mixture Models and Bayesian Model Selection' in *2012 Eleventh International Conference on Machine Learning and Applications (ICMLA)*, Institute of Electrical and Electronics Engineers; IEEE Computer Society; International Conference on Machine Learning and Applications; ICMLA, IEEE, Piscataway, NJ, pp.602–605.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S. and Schwabacher, M. (2008) 'Metrics for evaluating performance of prognostic techniques' in *2008 International Conference on Prognostics and Health Management (PHM)*, IEEE, IEEE, Piscataway, NJ, pp.1–17.

Susto, G.A., Schirru, A., Pampuri, S., McLoone, S. and Beghi, A. (2015) 'Machine Learning for Predictive Maintenance: A Multiple Classifier Approach', *IEEE Transactions on Industrial Informatics*, Vol. 11, No. 3, pp.812–820.

Verhagen, W.J. and Boer, L.W. de (2018) 'Predictive maintenance for aircraft components using proportional hazard models', *Journal of Industrial Information Integration*, Vol. 12, pp.23–30.

Welz, Z. (2017) 'Integrating Disparate Nuclear Data Sources for Improved Predictive Maintenance Modeling: Maintenance-Based Prognostics for Long-Term Equipment Operation', *Doctoral Dissertations* [online] https://trace.tennessee.edu/utk_graddiss/4667.

Welz, Z., Coble, J., Upadhyaya, B. and Hines, W. (2017) 'Maintenance-based prognostics of nuclear plant equipment for long-term operation', *Nuclear Engineering and Technology*, Vol. 49, No. 5 [online] https://www.sciencedirect.com/science/article/pii/S1738573317303236.

Yu, L., Zheng, Z., Lan, Z. and Coghlan, S. (2011) 'Practical online failure prediction for Blue Gene/P: Period-based vs event-driven' in *2011 IEEE/IFIP 41st International Conference on Dependable Systems and Networks Workshops*, IEEE, IEEE, pp.259–264.

Zhang, X., Kang, J. and Jin, T. (2014) 'Degradation Modeling and Maintenance Decisions Based on Bayesian Belief Networks', *IEEE Transactions on Reliability*, Vol. 63, No. 2, pp.620–633.

BIOGRAPHY

Henrik Simon received his M.Sc. in mechanical and process engineering from Technische Universität Darmstadt, Germany, in 2017. Since 2018 he is working at the Institute of Flight Systems and Automatic Control (FSR). As a Research Associate, he has been involved in several projects with a strong focus on Prognostics and Health Management. His current studies are aligned between PHM, usage and fusion of maintenance data and lectures of Machine Learning Applications for mechanical engineering students.