Towards Predictive Maintenance: an Edge-based Vibration Monitoring System in Practice

1st Victor Lorhan Loiola Costa
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—Due to the high acquisition and operation cost of industrial machinery, the cost-effectiveness is highly influenced by the quality and continuity of their production. In this context, Predictive Maintenance emerges as a maintenance strategy aiming to maximize uptime by constantly monitoring a quantity related to the machine's health, such as vibration patterns, in order to perform maintenance stops only when strictly required. The implementation of this strategy, however, faces multiple challenges. One of them is related to the design, installation and operation of the sensoring systems required, which are subjected to budget constraints and also technical constraints such as sensor battery lifetime. Another challenge is given by the intelligence required for analysing real-world data generated within uncontrolled industrial environments and producing a machinery health indicator from it. This work illustrates a vibration monitoring system currently operating in a textile manufacturing machine; proposes a versatile Anomaly-Detection-based procedure for the analysis the real-world data produced by such systems and extraction of a machine health indicator from it; and also proposes an extension to the system in order to adapt it to the connectivity requirements set by the current context of Industry 4.0.**

***Index Terms*—Predictive Maintenance, Anomaly Detection, Vibration Monitoring**

I. Introduction

A. The Way to Predictive Maintenance

Today's industrial machinery are usually capital-intensive. Hence, keeping such equipment in optimal operating condition with the highest availability, performance, and quality is viewed as a crucial part of ensuring the return on investment in the acquisition and operation. In this context, machine maintenance, a vital aspect of the Overall Equipment Effectiveness (OEE), ensures that a facility satisfies production schedules, minimizes machinery downtime, and prevents potential accidents in workplaces.

Different maintenance strategies are proposed as the complexity of machinery and their working environment increase over time. In general, they can be divided into:

1) reactive maintenance: performing maintenance when an issue is presented
2) preventive maintenance: performing maintenance due to failure experience accumulated by machine producers and operators
3) predictive maintenance: performing maintenance based on the evaluation of data collected from equipped sensors and other sources

In comparison with reactive and preventive maintenance, predictive maintenance detours unnecessary maintenance stops and unscheduled machine downtime since the machine is constantly monitored, and the maintenance will solely be performed when failure is imminent.

B. Vibration analysis - a glimpse into the condition of rolling machinery

Here we talk concretely about vibration. DO WE JUST ELIMINATE THIS SUBSECTION?

C. Associated challenges

Developing predictive maintenance concepts is encountered a wide range of challenges, such as cost, operational variability, and data privacy. First, as predictive maintenance is based on machine monitoring data collected from different sensors, the cost of the acquisition and installation of sensors is involved and can vary according to the complexity of the data model applied. Besides, individual machines produced as the same type might operate differently under different conditions. Hence, ensuring the general availability of the data-based model developed for a certain machine type evolves into another problem. Furthermore, accompanied by the increased complexity of methods and the quality of data flow, the requirements for hardware to process are restricted. Finally, data always implicates some sensitive information about production and business. Preventing data tampering and leakage is considered as well.

D. Industry 4.0 - Enabling Predictive Maintenance

In the current context of Industry 4.0, large amounts of devices are internetworked. The term Internet of Things (IoT) emerges as the amalgam of software techniques, communication technologies, and individual devices that constitute these networks which, despite the name, may or may not be connected to the internet.

The emerging chip technology facilitates the development of edge devices which, compared with cloud devices, provide the following advantages: 1) reduced latency, 2) improved security, 3) reduced infrastructure cost, 4) improved reliability, and 5) large autonomy for battery-powered devices. The combination of edge devices and predictive maintenance covers the challenges mentioned above.

E. Our contribution

This paper illustrates an edge-based vibration monitoring system currently operating in a textile manufacturing machine and proposes a versatile anomaly-detection-based procedure for the analysis of real-world data and extraction of machine health status, considering the battery life of sensors.

The remaining part of the paper is structured as follows: Section II provides an overview of the related work regarding predictive maintenance concepts for vibration analysis. In Section III, we propose our concept for data analysis techniques. The implementation details are presented in Section IV. We highlight the results of the data analysis and indicate our discussion in Section V. Consequently, the paper is concluded in Section VI.

II. Related work

Here I will present some articles and other sources about vibration analysis and talk about their shortcomings which we intend to overcome. I will also write about some cases where PM wa0s sucessfully implemented.

III. Concept - Data analysis techiques

The vibration monitoring solution proposed in this work employs a range of techniques within the realm of data and signal analysis. This Section is dedicated to summarizing them, see Figure 1.

First, in Section III-A, we present a discussion about the field of machine vibration analysis, which is required in order to gain insights into how to interpret the vibration signals and learn which of its features can be correlated to the machine's condition. In Section III-B we present an overview of the implemented signal processing techniques to extract the desired features from the vibration signals. In Section III-C we discuss the data clustering techniques used to identify and segregate the potentialy multiple operational modes presented by the machine under study. Finaly, in Section III-D we explore the Anomaly Detection concept and discuss about how to use the multi-variate Gaussian distribution to implement Anomaly Detection in the scenario at hand.

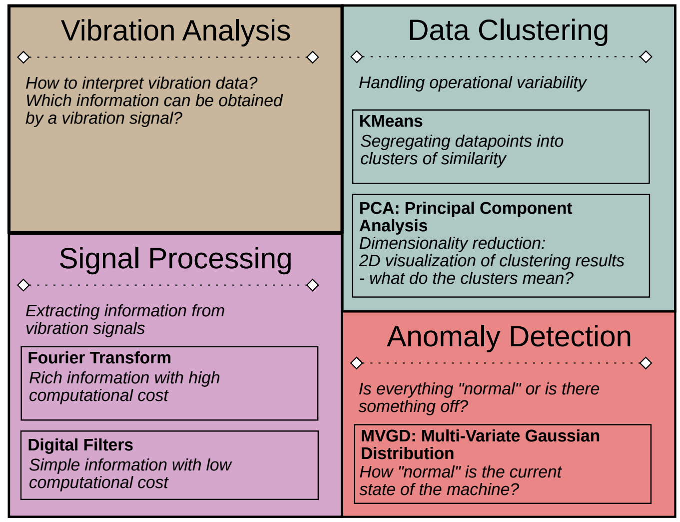

Fig. 1. Overview of the techniques employed in the data analysis.

A. Vibration Analysis

In the context of PM of rotating machinery, vibration analysis is widely considered the single most important technique available [1]. It provides a robust method for monitoring the general health of an equipment by indicating the onset and/or presence of many fault mechanisms, including [1]:

- Shaft misalignment
- Rotor unbalance
- Mechanical looseness
- Bear and gearing damage/degradation
- Inadequate reassembly after maintenance

In order to carry out this kind of analysis, the process starts with the aquisition of vibration signals of the machine under test by means of a vibration transducer. Then, the vibration waveforms can be analysed by multiple techniques.

In the simplest approach, we consider the vibration of rotating machinery is generally to be avoided. Although it is physically impossible to have a rotating machine that does not produce vibration, the intensity of this vibration should be kept within acceptable levels dictated by the mechanical structure of the machine and the environment where it is located.

This approach gives little indication of what the root cause of an underlying problem might be. A more sophisticated approach is the frequency analysis of the vibration waveform. The decomposition of the vibration signal into its frequency components is explained in more detail in Section III-B. Once the frequency components are obtained, one can associate the amplitudes over the frequency spectra to the frequencies where mechanical problems such as the ones listed in this section are expected to appear. Also, while the frequency indicates the source of the vibration, the amplitude indicates the severity of the problem.

The process of associating specific mechanical problems to a machine's vibration frequency spectra involves knowledge of a plethora of characteristics such as bearing and gear geometry and materials, type of grease, maintenance history and so on [1].

A basic characteristic is the rotating speed of the motors. A very common type of motor used in industrial rotating machinery is the four-pole three-phase electric induction motor [2]. This type of machine presents a mechanical rotation slightly less then half that of its alternate current electric supply. Thus, for a 50Hz network, we expect a peak around 25Hz for rotating machinery with such motors. Thus, for the case of machine speed control by means of the motor, the position where the peak around 25Hz appears is directly related to the speed the machine is operated at. This is particularly meaningful, as the machine's speed can exert a high influence on its vibration pattern. Hence, the vibration analysis should take the machine's speed into account [3].

B. Signal Processing

As exposed in Section III-A, the following characteristics of the vibration signal are of particular interest for an estimation of the machine's condition:

- Machine's speed
- Energy content of the vibration signal
- Distribution of the vibration's energy over its frequency spectrum

In order to extract this information from the vibration signals, a set of techniques from the realm of signal processing must be employed. In this section, we describe two techniques: the *Fourier Transform* and the *Digital Filters*.

1) Fourier Transform: Given our interest in the machine's speed, which can be directly related to the frequency peak in the 25 Hz region (see Section III-A), we then set ourselves to find the frequency where this peak happens. For this purpose, we can make use of the Fourier Transform.

The Fourier Transform is a mathematical transform widely used to decompose a time-function into multiple sinusoidal functions distributed over a frequency spectrum. The result of this transform is then not based on time, but on frequency [4]

In the discrete-time case, the FT is called Discrete Fourie Transform, which is expressed as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j\omega kn/N} \quad (1$$

Due to the discrete-time nature of the sampled signals use in signal processing, there is a time-difference T_s between an two consecutive samples. Such T_s value is called *samplin time*, and its inverse $f_s = 1/T_s$ is called *sampling frequency* As stated by the *Nyquist-Shannon sampling theorem*, th highest-frequency component in the image of $G(\omega)$ will have frequency $f_s/2$ [5]. Thus, the sampling frequency f_s of a signal defines the highest frequency component that can be obtained by the Fourier Transform of this signal.

Just like the input signal is of discrete-time, the output signal is of discrete-frequency. The time-resolution of the input signal is equivalent to the sampling time T_s and is thus determined solely by the sampling rate f_s. The frequency-resolution Δf of the output, however, is influenced also by the number N of the samples in the input signal and is expressed Equation 2:

$$\Delta f = f_s/N \quad (2)$$

I STILL NEED TO EXPAND THESE TWO TOPICS:

- Rich but expensive information.
- Explain how to extract a specific range.

2) Digital Filters: Digital filters are algorithms applied to sampled signals to reduce or enhance specific frequency ranges contained in them.

Similarly to the FT, the input to a digital filter is a time-domain signal. The output, however, is another time-domain signal, not a frequency-domain decomposition. This output signal contains, idealy, only the frequency components within specific ranges called cuttoff frequencies, which are embedded in the filter's mathematical description. Briefly explaining, ideal filters work by applying a zero-gain to the frequencies to be removed from the signal, which are called the stop-band, and unity-gain to the frequencies to be preserved, which are called the pass-band.

In contrast to ideal filters, real-world filters presents undesired characteristics such as a transition band, which is a gradual reduction of the gain between the frequencies to be preserved and the ones to be rejected. Other undesired characteristics are ripple and phase distortion. Figure 2 illustrates the transition band and ripple characteristics in a low-pass filter, which is characterised by a single cuttoff frequency, with the stop-band being all the frequencies above the cuttoff frequency, and the pass-band the ones below it.

The non-idealities of a filter implementation can be reduced by increasing the filter's complexity, at the cost, in the digital case, of more computations required. Other trade-offs are presented by the filter's topology, as this choice can result, for example, in a narrower transition band at the cost of a higher phase distortion.

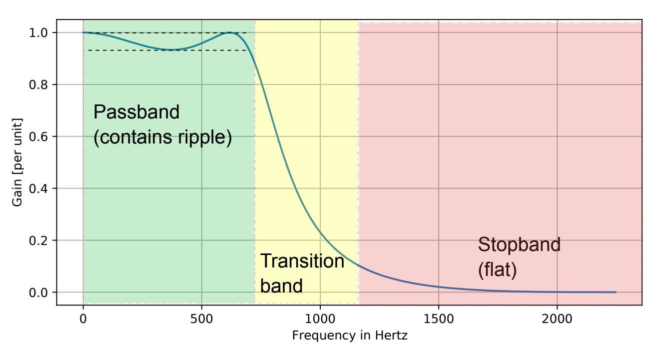

Fig. 2. Example of a frequency response of a non-ideal low-pass filter

While obtaining the filter's output for a given frequency range, the Root Mean Square (RMS) value of this output can be simultaneously computed. This RMS value of the output signal, which is obtained by summing the squares of each sample and then taking the square root of the sum, can be interpreted as the energy content of a signal. Hence, the RMS value can be used to represent the intensity of the vibration in the corresponding frequency range [3].

C. Clustering

As already explained in Section 1.1, industrial machinery usually presents different operational modes. These different types of operations produce differences in vibration signals which must be taken into account in the analysis of the acquired signals. This work then makes use of clustering techniques in order to identify aggregation patterns in the vibration signals which can be associated with different operational modes.

In statistical data analysis, the term ”clustering” refers to the task of dividing a group of data points into subgroups, also called clusters, of data points which are considered similar according to some specification.

In this context, the concept of ”similarity” can be defined in different forms. A simple and intuitive method is the Euclidean distance between the data points, meaning that points located close to each other are considered similar.

This section then describes two specific techniques employed in the clustering task in this work: the K-Means algorithm and the Principal Component Analysis.

1) K-Means: K-Means is a simple and popular technique within the realm of data clustering. This technique works, in essence, by identifying a certain number of positions around which the data points tend to aggregate themselves and then labeling the data points according to the closest of the positions of aggregation identified.

Figure provides an intuition for the working principle behind this technique. It illustrates a toy example with multiple datapoints which aggregate themeselves in two clusters. The objective of K-Means is to find points C_A and C_B, namely the centroids, which minimizes the WCSS, which is determined as per expressed in Equation 3:

$$WCSS = \sum_{i=1}^{m} l_i^2 \tag{3}$$

where m is the number of points in the dataset and l is the Euclidean distance between each point p_i and the nearest centroid to it, as per Equation 4:

$$l_i = min(dist(C_A, p_i), dist(C_B, p_i)) \tag{4}$$

The K-Means technique does not specify the number of clusters in the data. This is an information that must be given as an input to the problem. However, this information is usually not known beforehand, and figuring out how many clusters the underlying patterns in the data present is actually part of what is expected of a data clustering procedure. The Principal Component Analysis, presented in the next session, comes into play in this task by facilitating that humans apply their expertise to identify patterns visually.

2) Principal Component Analysis: Principal Component Analysis is a data dimensionality reduction technique widely used to represent data with more than three dimensions in a graphical way in two or three dimensions. This enables humans to use their sense of vision to try to find insights in the data.

The intuition behind this technique can be explained as applying a change of basis in the data, where this new base is composed of vectors that point in the directions of maximum variance in the data. These vectors, also called Principal Components, are hierarchically sorted based on how much of the total data variance they explain.

Such principal component vectors are equivalent to the eigenvectors of the data's covariance matrix, and the variance explained by them is directly related to their associated eigenvalues [6]. Thus, the math behind PCA can be summarized by building a linear transformation matrix with the eigenvectors of the covariance matrix and then applying this linear transformation to the data in question.

The data's covariance matrix C, of dimensions $m \times m$, where m is given as per Equation 5:

$$C = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \mu)(x^{(i)} - \mu)^T \tag{5}$$

D. Anomaly Detection

Anomaly Detection (AD) is a term used to refer to a number of mathematical techniques which aim to detect outliers in datasets.

These techniques can be summarized in two basic pillars:

1) Providing a mathematical model that captures the underlying patterns in a dataset
2) Using this model to declare if a given datapoint conforms to the model-associated pattern or not, that is, if it is a normal or an anomalous point

1) The Multi-Variate Gaussian Distribution: The Multi-Variate Gaussian Distribution is an extension of the Gaussian Distribution in which the data-points have more than one dimension, in addition to taking to account the correlation of the variables in the dataset for the determination of the probability that the point belongs to the distribution.

This calculation is based on the covariance matrix C, already presented in Equation 5. The data-point's probability is given by Equation 6:

$$p(x|\mu, C) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} \exp\left(\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right) \tag{6}$$

IV. Hardware implementation - Data aquisition setup

The vibration monitoring hardware and software infrastructure used in this work was developed by the company DELTA Systems and is installed in a textile manufacturing plant in the city of Malatya in Turkey.

The hardware setup consists in a set of 85 wireless vibration sensors, which are all installed in a same machine and communicate with 2 gateways over radio. Figure 3 depicts one of the sensors attached to the referred machine.

Fig. 3. Textile machine with electronic sensor attached (Source: DELTA Systems GbR).

The wireless electronic sensors are embedded systems controlled by a microcontroller. The microcontroller program is written in C language, and the compiled machine code is stored in the microcontroller itself. The system also comprises an external flash memory used to store vibration measurements for later analysis and/or transmission.

An important challenge related to the operation of the sensors is their energy autonomy. Due to the small form factor of their sensors, the space available for their battery supply is limited, so that the sensor operation, defined by their firmware application code, must ensure that energy-intensive operations, such as the radio communication, are used as seldom as possible.

The gateway modules are embedded computers with a linux operational system. They run an application written in Python, which is responsible for interacting with a SQLite database, communicating with the field sensors over radio, and also providing a graphical user interface which can be accessed via TCP/IP based connection from a desktop computer connected to the same network.

An overview of this hardware, firmware and software infrastructure available to the vibration monitoring setup is depicted in Figure 4.

V. Results and Discussion

A. Data Exploration - Overview of the vibration measurements

The electronic sensors provide vibration measurements along the three cartesian axes x, y and z, with origin and orientation assossiated to the package of the digital vibration transducer integrated circuit. By letting the fictional axis a be defined as $a = \sqrt{x^2 + y^2 + z^2}$. This fictional axis a then contains all the frequency components found in x, y and z.

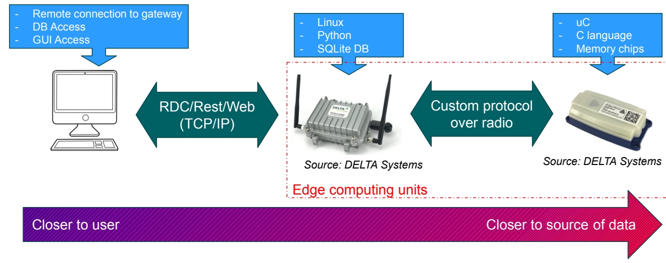

Fig. 4. Overview of the system's architecture.

On the firmware side, the vibration measurements are configured to a sampling frequency of 4500Hz and duration of 2s. Hence, the number of samples for each axis in a measurement is $N = 4500Hz \cdot 2s = 9000$.

In Figure 5 we can see an example of a vibration signal, on both time and frequency domain, for axes x, y, z and a.

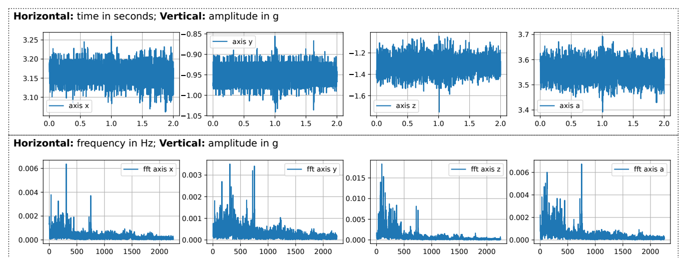

Fig. 5. Plot of an example of a vibration signal over axes x, y, z and a, on both time (top) and frequency (bottom) domain.

B. Machine speed

In order to observe the behavior of f_{motor}, we look more closely into the frequency range from 25 to 35 Hz. A plot for this specific region for different a-axis measurement signals is presented in Figure 6. An analysis of this plot indicates that, in this range, there is sometimes indeed a peak near 30 Hz, and sometimes there is no discernible peak at all. An important takeaway from this observation is the confirmation that f_{motor} can indeed be obtained in the 25 to 35 Hz range, which allows us to restrict the costly FT calculations to this range. As for the measurements with no discernible peak in the observed region, they presumably represent measurements acquired when the machine was not operating, as will be explained in more detail in Section V-D.

C. Digital Filters

In the introduction to Section III-B we presented three vibration signal characteristics of interest for our analysis.

The first of them, the machine's speed, can be represented by f_{motor}, obtained as explained in Section V-B.

The second one, the energy content of the vibration signal, can be easily obtained by the RMS value of the a-axis of the measurement.

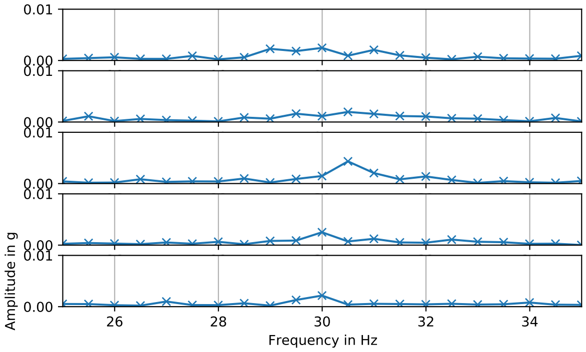

Fig. 6. Observations of the a-axis frequency spectra in the expected range for f_{motor}.

In this section, we detail how to obtain the third and last one: the distribution of the vibration's energy over its frequency spectrum. As already explained in Section III-B2, this can be acomplished with reasonable computational cost by computing the RMS values of the output of digitail filters.

For this purpose, the available frequency range of up to 2250 Hz is divided into three ranges:

- a lower range up to 500 Hz
- a middle range betwen 500 and 1250 Hz
- a higher range from 1250 Hz on

The higher the number of ranges, the more detailed the vibration energy distribution over frequency, at the cost of more computations required.

For each of these ranges, a digital filter was developed. These filters are then named F_{low}, F_{mid} and F_{high}, acording to their associated frequency range.

By taking to account the discussion in Section III-B2, the main design decisions are summarized as follow:

- Topology: The topology chosen was the inverse Chebyshev. The first reason for this choice is the narrow transition band, which is always desirable. The tradeoff for the narrower transition band is a larger phase distortion, which is irrelevant for our case, as we do not take phase into account in our modelling. Finally, we desired a flat passband in order to preserve as much as possible the information of the passband for our modelling.
- Order: As a general rule, the overall performance of the filter is improved with higher order, at the cost of more computations required. By experimentation, the order value was set to 3.
- Overlap: Some overlap was included in the cutoff frequencies of the three filters. This was a decision taken in orther to achieve an optimum configuration within the numerous tradeoffs in filter design. Hence, F_{low} was defined as a low-pass with cuttoff at 700 Hz, F_{mid} was defined as a band-pass with cutoffs at 400 and 1350 Hz, and F_{high} as a high-pass at 1150 Hz.

The frequency response of the three filters is presented in 7. Figure 8 depicts the output of the filters, in time and frequency domain, for an example a-axis measurement input signal.

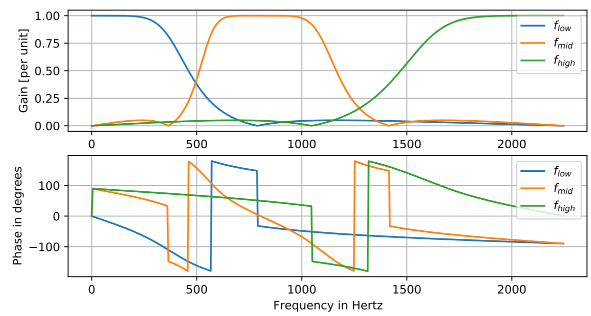

Fig. 7. Frequency response of the three filters.

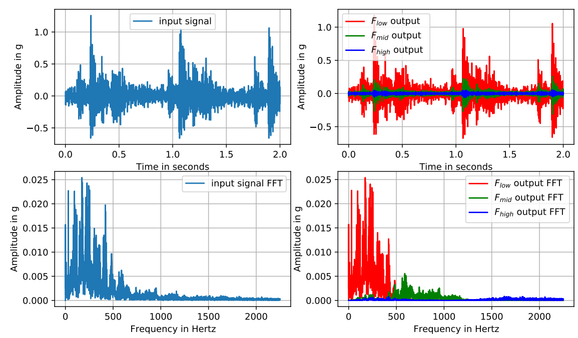

Fig. 8. Filters, in time and frequency domain, for an example a-axis measurement input signal.

D. Clustering: Segregation of Operational Modes

In order to understand the underlying patterns in the data, the first step was to carry out a Principal Component Analysis on the dataset. The input to this analysis contained five features:

- f_{motor}: peak frequency in the 25Hz to 35 Hz range
- RMS_{total}: RMS value of the a axis
- RMS_{low}: RMS value of the output of the f_{low} filter
- RMS_{mid}: RMS value of the output of the f_{mid} filter
- RMS_{high}: RMS value of the output of the f_{high} filter

The resulting PCA plot is presented in 9. This same plot also shows a seggregation of the points based on the results of a Kmeans clustering procedure.

While analysing the measurements assigned to each cluster, it was noted that one of the clusters was characterised by lower vibration levels in general than the other one. This lead to the conclusion that the former represents the "operational" state for when the machine is off, and the later for when the machine is on. This conclusion is supported by the region with large variation, encircled in a green dashed shape. When the machine is off, the motors are not running, so they do not produce the 30 Hz peak, thus the peak in the 25 Hz to 35 Hz range ends up being mostly randomly defined by noise.

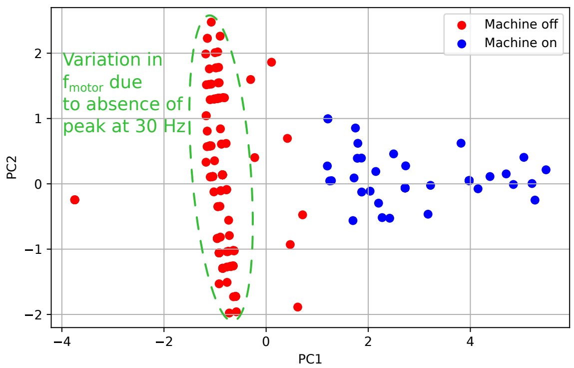

Fig. 9. PCA plot for the described dataset

When the machine is off, its vibration signature is of low relevance for PM. Thus, the analysis from this point on focuses on the cluster that represents the machine in its ”on” state. The segregation can be easily carried out by thresholding the RMS_{total} value. This threshold value was set to $0.015g$. It is worth at this point to recap that the a axis signal has its average value removed, so the $1g$ gravity vibration is not present in the signal.

E. Approach to Anomaly Detection

Feature engineering. Models. Plot 3D. Plot over time.

F. The Way to the Internet of Things

TODO.

VI. Conclusion

Acknowledgment

The authors thank Mr. Georg Sommer, software engineer at DELTA Systems GbR, for his work in writing the gateway’s software infrastructure which enabled the aquisition, storage and retrieval of the data produced by the electronic sensors.

References

[1] R. B. Randall, “Vibration-based condition monitoring: industrial, automotive and aerospace applications,” John Wiley & Sons, 2021.

[2] M. Tsypkin, “The Origin of the Electromagnetic Vibration of Induction Motors Operating in Modern Industry: Practical Experience—Analysis and Diagnostics,” IEEE Transactions on Industry Applications, vol. 53, pp. 1669–1676,2017.

[3] H. P. Bloch and F. K. Geitner, “Machinery Failure Analysis and Troubleshooting: Practical Machinery Management for Process Plants,” Elsevier, 2012.

[4] L. Tan and J. Jiang, “Digital signal processing: fundamentals and applications,” Academic Press, 2018.

[5] A. J. Jerri, “The Shannon sampling theorem—Its various extensions and applications: A tutorial review,” Proceedings of the IEEE, vol. 65, pp. 1565–1596, 1877.

[6] M. E. Wall and A. Rechsteiner and L. M. Rocha, “Singular value decomposition and principal component analysis,” in “A practical approach to microarray data analysis,” p91–109, Springer, 2003.