Preliminary report on prediction of failures for the HMC106 centrifugal compressor

Ian Palacín, Jordi Planes, Simone Arena

1 Introduction

The aim of this report is to briefly explain a preliminary study of the data provided by the HMC106 centrifugal compressor and the potential it has when it comes to detecting filures before they happen. The benefits of this application are numerous.
What is needed to understand the content below will be explained throughout the document, so that no previous knowledge is required to understand the general idea.
Each section of this document except for the first and the last one will show the results of the study upon a real different fail of the machine. There will be a total of 4 fails shown:

- **Fail 1** will serve as a point of view of the early stages of the study.
- **Fail 2, 3 and 4** it will serve as a point of view of the late stages of the study.

The first section will explain the most important tools that have been used to generate the results, which are represented visually by two different graphs. These graphs will appear in each of the fails that will be seen in this document, so the first section will explain how to interpret them.
In the appendix there is more information about each fail and the data used to study it.

2 Statistical tools and graphs

A series of tools and techniques have been used to obtain the data that will be represented in the graphs, the most important two are the following:

- **PCA**: Principal Component Analysis, brings together different variables into one in such a way that it maintains the properties and relations of the original variables. Working with one variable instead of many is easier and makes its visual representation feasible. **If you want to know more**... it is a branch of the multivariate statistical analysis based on dimensionality reduction. It makes a linear mapping of the data from the eigenvectors of the covariance matrix that correspond to the largest eigenvalues (main components).
- **Mahalanobis distance**: measures the distance of a specific value from the other values in the dataset. In other words, it measures how anomalous or ”strange” is a specific value compared to the rest. **If you want to know more**... doing the standard deviation we could find how anomalous is a value with respect to the center of mass, but we would be assuming that the sample points are distributed in a spherical manner. When the distribution is non-spherical, for instance ellipsoidal, the probability of the test point depends not only on the center of mass, but also on the direction. The Mahalanobis distance consider these details.

In this report, the result of the prediction of a failure is represented by two graphs. These graphs will be seen in each of the following sections and were obtained using the techniques mentioned above, among others.

2.1 Anomaly metric

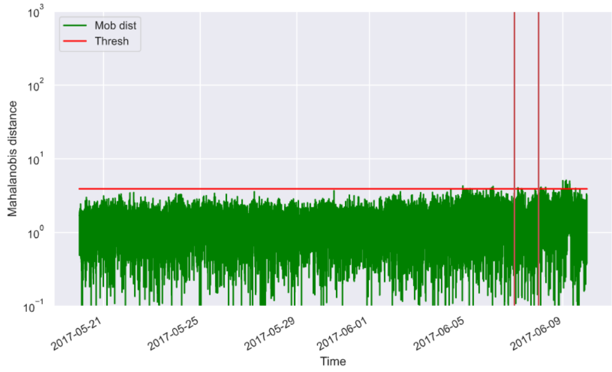

Figure 1: Anomaly metric

The first graph is the Anomaly Metric, it is based on the data of a set of variables, for instance, the temperature and the vibration of the sensor X over a period of time.

The **green line** is the Mahalanobis distance, measuring the "strangeness" of each data point of the original dataset. The **two vertical red lines** represent the beginning and the end of the fail respectively, in this case 24.10.2016 to 27.10.2017. The **horizontal red line** is an example of a threshold, where a notification could be sent.

It can be interpreted as ***"how anomalous is each data point over a period of time"***.

2.2 Anomaly metric tendency

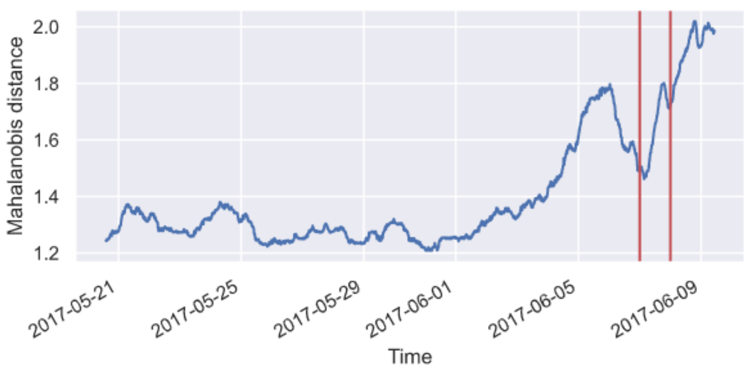

Figure 2: Anomaly tendency

The second graph is the Anomaly Tendency, it shows the tendency of the Anomaly metric graph. It is a more visual and intuitive way to see the general path of the previous graph. The beginning and end of the fail is also represented as the two red vertical lines.
It can be interpreted as ***"how anomalous is the data overall"***.

3 Fail 1

This is the first fail shown in the document, it will serve as a point of view of the first steps of the study. The machine failed from the 24.10.2016 to the 27.10.2017, the data used to analyze it was the pressure and temperature of the compressor[1], and revealed the following graphs:

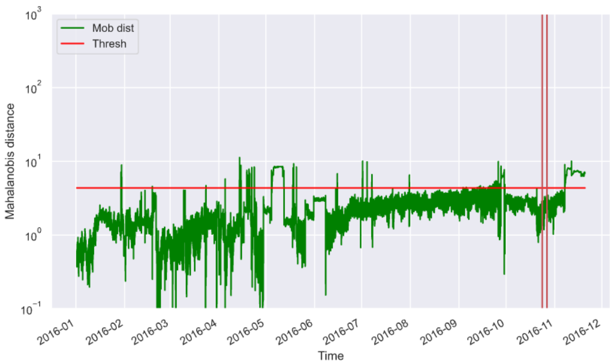

Figure 3: Anomaly metric for fail 1

At the beginning of the study there was a lot of data to filter and to treat, the variables and the general behavior of the data was not clear and all the impediments produced poor results. The poor results can be seen in the graph, where nothing is clear and no feature can be taken out of it.

The same errors are also extrapolated to the second graph, where no failure is appreciated nor detected.

[1]HPI602, HPI605, HTI605

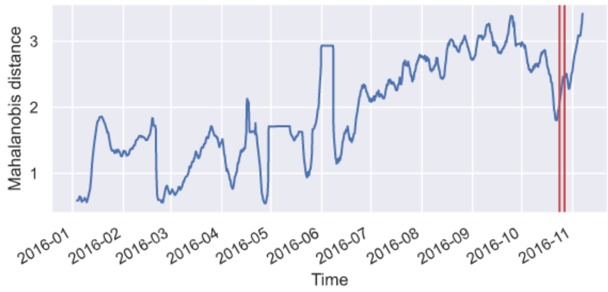

Figure 4: Anomaly tendency for fail 1

4 Fail 2

The fails of the machine that will be shown from now on (3, 3 and 4) will serve as a point of view of the late stages of this study, where the greater knowledge of the important variables, the general behavior of the data and the proper treatment of it led to noticeable results.
The fail that will be shown in this section started the 22.05.2017 and ended three days later, the 25.05.2017. The data used to analyze the failure is from two variables that represent the axial displacement of the compressor[2].

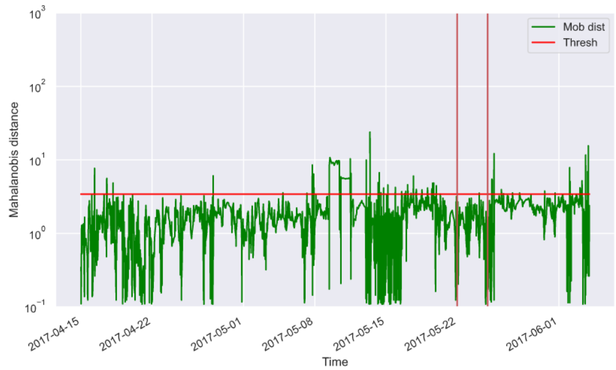

Figure 5: Anomaly metric for fail 2

As the graph below shows how abnormal each data point is separately, the information represented is difficult to see. That is why we have the second graph, where the general tendency is clear, and the results appear more obvious.

[2]HZI877B, HZI877C

Figure 6: Anomaly tendency for fail 2

In the graph above it is clearly seen the exact moment where the machine failed, the peak of the curve. An ascendant curve that could be a visual representation of the deterioration of the machine (or a part of it), culminating in its failure marked with the first red line. The fail could have been detected over 10 days in advance.

5 Fail 3

The next fail started on the 12.11.2018 and ended on the 29.11.2018. This time the data is based on the temperature and vibration of the electric motor[3].

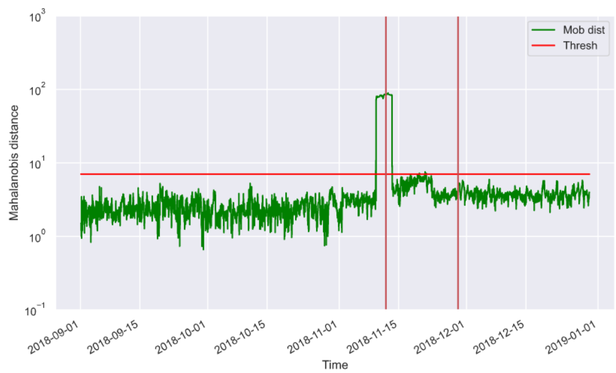

Figure 7: Anomaly metric for fail 3

[3]HTI880A, HTI879B, HTI883A, HVI877X, HVI877Y, HVI878Y

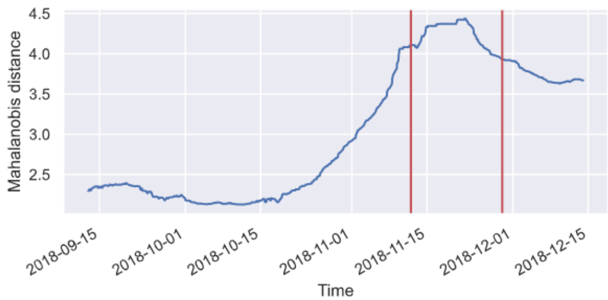

Figure 8: Anomaly tendency for fail 3

The results are similar to the ones in the previous section, a clear deterioration leading to the fail of the machine, foreseeable 10 days before.

6 Fail 4

The last failure seen in this document started on the 12.11.2018 and ended on the 29.11.2018. The data belongs to the electric motor[4], also foreseeable.

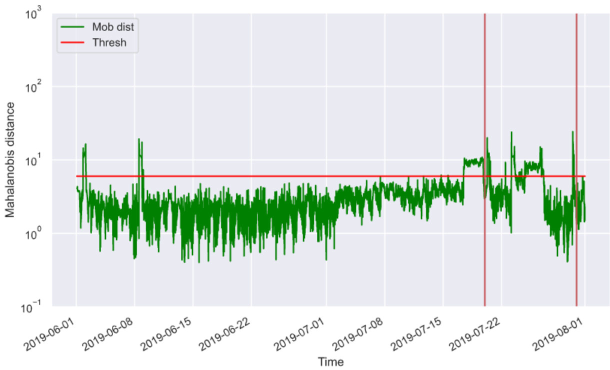

Figure 9: Anomaly metric for fail 4

[4]HTI879A, HTI879B, HTI883A, HVI877X, HVI877Y, HVI878Y

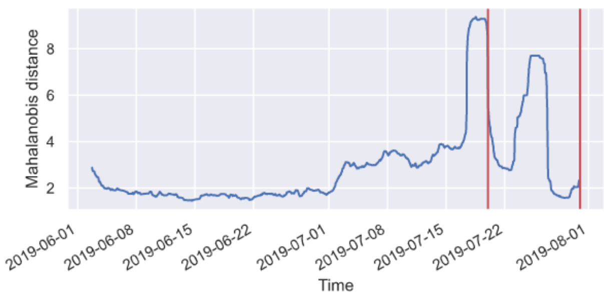

Figure 10: Anomaly tendency for fail 4

7 Other insights

The previous sections showed the capability that the data has at the time to predict when the machine is going to fail. However, the potential is much greater, and the same data can reveal more important information and insights about the operation of the machine. This information can lead directly to machine optimizations and better performance. See the following example.

We saw that the data can predict whether the machine is going to fail in a near future, but with little effort it can also reduce the number of possible culprits of the failure, or even point directly to the specific component that will cause it. Look at the following graphs:

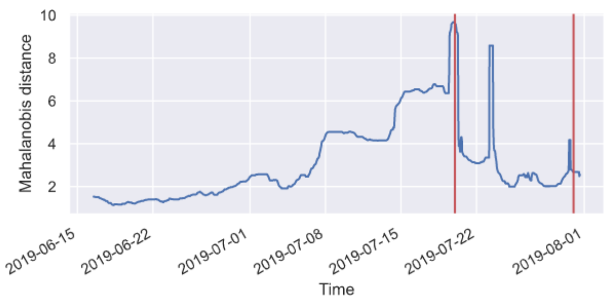

Figure 11: Anomaly tendency for bearing 892B

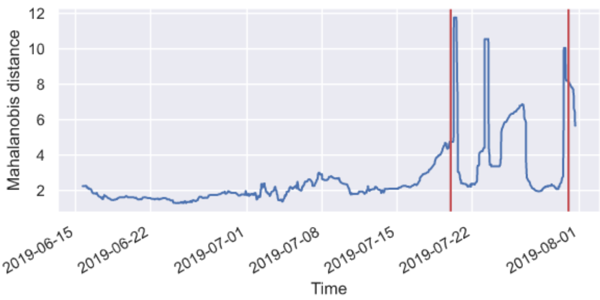

Figure 12: Anomaly tendency for bearing 891B

Both graphs are from the same machine fail, that of 20.07.2019 to 31.07.2019. They are even from the same part of the machine, the gearbox. The difference is that the first is from the bearing **892B**, and the second from the bearing **891B**. In the first graph the deterioration is evident, and the peak occurs **before** the fail. Contrary to the second graph, in which the deterioration is not that obvious, and the peak occurs **after** the fail. This type of precedence gives a lot of information about the culprits of the fail.

8 Conclusions

This was only a preliminary study on the data of the machine aimed to show the potential and maneuverability that it offers, and the amount of valuable information that could provide the continuation of the analysis, with a deeper study of the data. In this document were just presented a few examples.

9 Appendix A

9.1 Details of fail 1

Variables used: HPI602, HPI605, HTI605
Time of the fail: 24.10.2016 - 27.10.2016
Training data graphic:

Figure 13: Training dataset for fail 1

9.2 Details of fail 2

Variables used: HZI877B, HZI877C
Time of the fail: 22.05.2017 - 25.05.2017
Training data graphic:

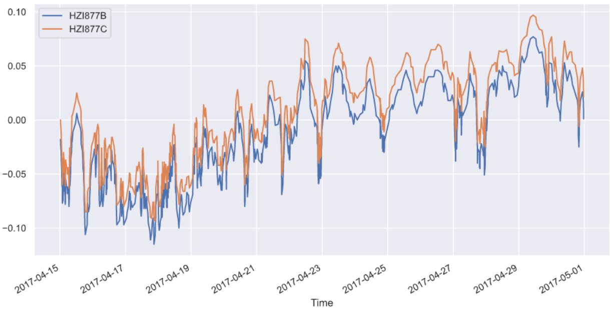

Figure 14: Training dataset for fail 2

9.3 Details of fail 3

Variables used: HTI880A, HTI879B, HTI883A, HVI877X, HVI877Y, HVI878Y
Time of the fail: 12.11.2018 - 29.11.2018
Training data graphic:

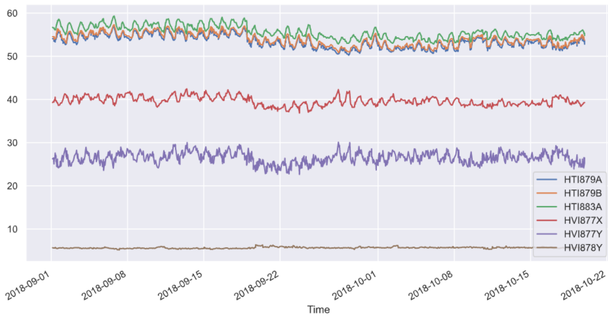

Figure 15: Training dataset for fail 3

9.4 Details of fail 4

Variables used: HTI879A, HTI879B, HTI883A, HVI877X, HVI877Y, HVI878Y
Time of the fail: 20.07.2019 - 31.07.2019
Training data graphic:

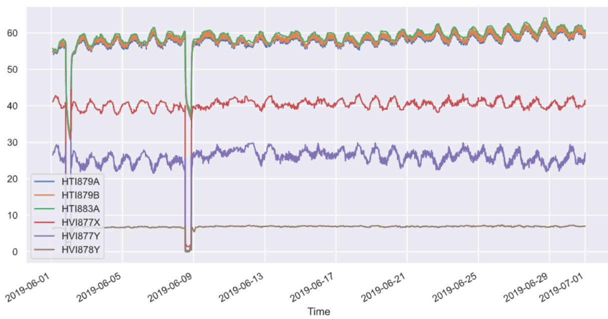

Figure 16: Training dataset for fail 4