From Data to Physics: Signal Processing for Measurement Head Degradation Detection

Gabriel Michau[1]

[1] *Zurich University of Applied Sciences, Rosenstr. 3, Winterthur, 8401, Switzerland*
gabriel.michau@zhaw.ch

ABSTRACT

Data-driven Prognostic and Health Management (PHM) is gaining more and more popularity, to the point where it is often seen as a potential solution for any maintenance problem. To the opposite, it is known that, when the physic of a system is known and tractable, it is usually the most efficient way to infer relevant indicators for detecting system behaviour changes. And sometimes, when the physic of a system is not known, data-driven approaches can highlight critical information leading to the right physical assumptions, and then, to the right health indicators.

This paper presents a real-life and successful such "reversed" case study. With focus on the estimation of a measurement system head degradation, a data-driven approach helped to identify the right signal processing tools with which to analyse the data. Using spectral analysis with windowed Fourier transform, the physics of the recalibration events could be hypothesized, and then used for deriving the most relevant degradation indicators. These indicators have been implemented online and so far used successfully on several machines encountering degradation of their measurement head.

1. INTRODUCTION

This paper focuses on Prognostic and Health Management (PHM) with the case study of a measurement head system. High precision measurement heads combine usually two different position measurements to achieve high-accuracy. First, an absolute grading (physical or magnetic) gives at regular interval the precise position of the head. Second, in between each reading, the head interpolates its location. This could be done thanks to a Doppler-effect measurement or thanks to electromagnetic interferences.

In many cases, these systems are used in industrial machines achieving precise task: product measurement, cutting, printing, etc... To the opposite of applications, like product measurement, were only the final position is necessary, for applications where machines perform their tasks continuously (like 3D printing for example), the position of the operating head should be known with high precision at every time. This is paramount to high-quality products and requires a very good match between both the interpolated and the absolute position. When the system is well-calibrated, this is usually the case but events can threaten this calibration. Deteriorations could arise among others, from a bad recalibration after a maintenance task, from displacement after a shock or with time or maybe from dirt masking the measurement head (if operated in a dirty environment, or if manipulated with dirty tools).

When such problems occur, the miss-match between the two combined readings can have unexpected impacts on the system, in particular through the feedback loop between the reading and the control, and eventually, the job quality is decreased. To avoid such degradations, preventive maintenance is often the answer but it has its known drawbacks, like the risk of over-maintaining the system, leading to high maintenance costs, or the lack of reaction when sudden changes occur. Therefore, the detection of these calibration problems is an important challenge for the industry. Recent technological advances have overcome most of the technical requirements for real-time monitoring of fleet of systems, in particular, it is much easier to connect the machines to the internet and to analyse data sent from the customer sites in a single place.

In the rest of this paper, we present how we faced this problem, starting with as little knowledge on the physical system as the one presented above. After a first attempt to solve the problem with machine learning, applying neural networks in Section 2, we identified the system well-defined behaviour in Section 3 with signal processing and spectral analysis. This lead us to propose a first health monitoring indicator. Refining our understanding of the data, we propose additional suitable health monitoring indicators and a implementation framework in Section 4.

2. FIRST APPROACH: EXTREME LEARNING MACHINES

Previous works have shown that neural networks could be used for health monitoring. In particular, Hierarchical Machine Learning (HELM) have proven to be efficient for learning "normal" conditions and then detecting any deviation to these conditions (Michau, Palmé & Fink, 2017; Michau, Yang, Palmé & Fink, 2018). Thus, as a first answer to the monitoring of the measurement head health monitoring, we trained a simple HELM. The reasons why HELM is a good candidate are mainly twofold: First, HELM first layer is an auto-encoder with sparse feature-to-variable connections (ℓ_1-regularisation), making it efficient to handle a vast amount of correlated inputs. It is thus a robust approach that does not require much pre-processing or feature pre-engineering. As the system we are working with, records over 50 variables on 3 measurement heads at the same time, such efficient dimensionality reduction seemed necessary. The system is in fact made of three engines used to move the machine, each engine with one measurement head. For each head, many variables are recorded, among others, position, speed, torque, electric current,... The second reason lies in that the HELM last layer is a one-class classifier, efficient to measure similarities between the training and the testing data. Therefore, if for any reason the testing data are behaving differently from the training, it should be detected.

Figure 1a represents an illustration of the HELM architecture. HELM is a superposition of stacked random neural networks (or Extreme Learning Machines) with as many stacked auto-encoder as one wishes and last, a classifier layer. In this case study, the HELM we used is composed of one auto-encoder and one one-class classifier.

Applying this HELM to laboratory experiments brought very encouraging results. Training the HELM with well-calibrated heads on a controlled movement plan led to very good detection rates as soon as one of the three head was manually deteriorated. Figure 1b illustrates the HELM output for this first case study. The one-class classifier output can be interpreted as a distance to the normal class. A threshold is defined based on a clean validation dataset. When the threshold is exceeded, this is considered as an anomaly detection. For the three deteriorated datasets, the detection rate is very high (in Figure 1b, most of the data-points are above threshold).

However, if it is known that HELM is very good at learning the data it is trained with, it lacks, as many machine learning tools, of the capacity to extrapolate its learning to achieve robustness toward slightly different operating conditions. In Figure 1c, an additional dataset is tested. The head were well calibrated but a different movement plan was used, with different speed settings. This dataset is detected as abnormal.

If there are some already existing trails for achieving better robustness to various operating conditions or for fleet training of the HELM (Michau, Palmé & Fink, 2018), a characteristic of the HELM output stroked. Zooming strongly on the HELM output for deteriorated head datasets, highlighted some periodic oscillations. It is not surprising that problems in reading the position for a moving system would lead to vibratory effects in the machine, it is actually, why such misreading are problematic, but the period of these oscillations made it worth to have a closer look from a spectral analysis perspective.

3. SPECTRAL ANALYSIS: WINDOWED FOURIER TRANSFORM

The vibratory behaviour of the HELM output, led us to have a deeper look at the different available variables from a spectral analysis perspective. In PHM, many machineries include rotary parts and therefore, vibrations are usually monitored: they are usually both symptoms of developing problems, when the spectral composition changes over time, and cause for potential additional problems if their amplitude increase too much. This is particularly true for machines with bearings. Therefore, in traditional PHM, vibrations of rotating parts have played an important role (Dalpiaz & Rivola, 1997; Goyal & Pabla, 2016; Jardine, Lin & Banjevic, 2006; Newland, 1989; Scheffer & Girdhar, 2004). To the opposite, vibrations caused by feedback loops, as it seems to be the case here, have been far less explored for condition monitoring. In most cases, focus is on preventing them from the control (Mirafzal, Khorasani & Ghasemi, 2016).

In the following we revisit the theoretical foundation of the Fourier Transform. If the theoretical tool has already been well studied, as well as vibrations in condition maintenance, it is used here for the understanding of the system behaviour. In addition, highlighting the major strengths and limitations of that tool, we propose a framework on how it can be applied to our case study.

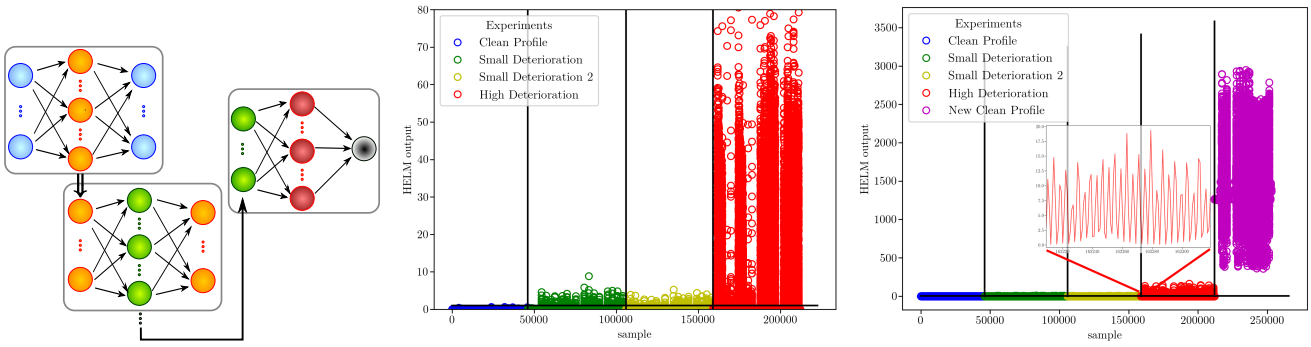
3.1. About the Windowed Fourier Transform

The Fourier Transform is a reversible mathematical operation, which consists in describing a time varying function (signal) from a frequency perspective. If $X(t)$ is a signal, then its Fourier Transform $F(\nu)$ is given as

$$\forall\, \nu \in \mathbb{R}, \quad F(\nu) = \int_{-\infty}^{+\infty} X(t) e^{-2\pi i \nu t} dt \qquad (1)$$

It has an direct physical interpretation and is reversible, giving a one-to-one correspondence between the time and the frequency domain.

If in its primary formulation neither the frequency, nor time



(a) (b) (c)

Figure 1. HELM: (a) Structure of an HELM. HELM consists in stacked auto-encoders for unsupervised learning with a last supervised layer (regression or classification). (b) 1-class HELM output for four datasets with same movement profile. The horizontal blackline represent the threshold above which abnormal movement conditions are detected. (c) HELM output for the 4 four same datasets and an additional dataset, from a clean head, but with a different movement profile and different parameters of the machine.

domains are bounded, it can be adapted to sampled data as:

$$\forall\, \nu \in \mathbb{R},\ F(\nu) = \frac{1}{t_{\text{end}}} \sum_{t=-\infty}^{+\infty} X(t) \cdot \mathbb{I}_{t=0}^{t_{\text{end}}}(t)\, e^{-2\pi i \nu t} \delta_t \quad (2)$$

$$= \frac{1}{t_{\text{end}}} \sum_{t=0}^{t_{\text{end}}} X(t) e^{-2\pi i \nu t} \delta_t \quad (3)$$

$$= \frac{1}{N} \sum_{n=0}^{N} X(n) e^{-2\pi i \nu \frac{n}{F_s}} \quad (4)$$

where

$$\forall t \in \mathbb{R}, \quad \mathbb{I}_{t=0}^{t_{\text{end}}}(t) = \begin{cases} 1 & \text{if } t \text{ in } [0, t_{\text{end}}], \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

the sampling increment is $\delta_t = \frac{1}{F_s}$, with sample N, the last at $t_{\text{end}} = N \cdot \delta_t$.

The function $F(\nu)$ as defined here is periodic with period F_s and, if X, the signal, is real-valued, $F(\nu)$ is in addition symmetrical around $F_s/2$. With N sampling points for frequencies $\nu \in [0, F_s/2]$, this leads to a frequency resolution: $\nu = \frac{k}{N} \cdot F_s$ where $k \in [\![1, \frac{N}{2}]\!]$.

The Discrete Fourier Transform for real signals is therefore:

$$\forall\, k \in [\![1, \tfrac{N}{2}]\!], \quad F(k) = \tfrac{2}{N} \textstyle\sum_{n=0}^{N} X(n) e^{-2\pi i k \frac{n}{N}} \quad (6)$$

Those two last points are actually highlighting the two most important property of the discrete Fourier transform: First, one can only sample frequencies up to $F_s/2$. Above this frequency, there will be spectral aliasing. Second, the spectral resolution depends only from N, the number of points from which the signal is made. To achieve higher temporal resolution of the spectrum, it is customary to use as input to the Fourier transform part of the signal (or a window), yet this is at the expense of spectral resolution. Convolving the signal with a window function and applying the Fourier transform leads to the construction of a spectrogram, whose temporal resolution (resp. spectral) becomes higher (resp. lower) as the windows becomes smaller.

The results of the windowed Fourier Transform depends therefore on these three major points:

1. Events impacting the signal with frequencies below half the sampling frequency only can be observed[1].

2. High temporal precision in the spectrum is achieved with small windows but at the expense of the spectral resolution ($\delta_\nu = \frac{1}{N \cdot F_s}$).

3. The convolution with a function to create a window is not a innocent operation: one can easily show that, $FT\left((f \star g)(t)\right) = FT\left(f(t)\right) \cdot FT\left(g(t)\right)$, where $\star$ is the convolution operation[2]. The spectral composition of the window is multiplied with that of the signal, causing some artefacts in the spectrogram. If by default, the Hamming windows is usually applied for its good spectral localisation, this choice needs to be justified by the case study. The choice of the window relies on a trade-off between better frequency amplitude estimation, artefacts and cut-off frequency (that is a frequency under which all their Fourier coefficients are close to zero). We will see that, in this paper, the rectangle Fourier transform might later be preferred, despite its lack of spectral localisation, as it better estimates the amplitude coefficients and as it has a lower cut-off frequency.

The choice of the window is illustrated in Figure 2 for the

[1] This is actually intuitive: to observe a frequency, it requires at least two points per period (one up, one down)

[2] $(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) g(t - \tau)\, d\tau = \int_{-\infty}^{\infty} f(t - \tau) g(\tau)\, d\tau$

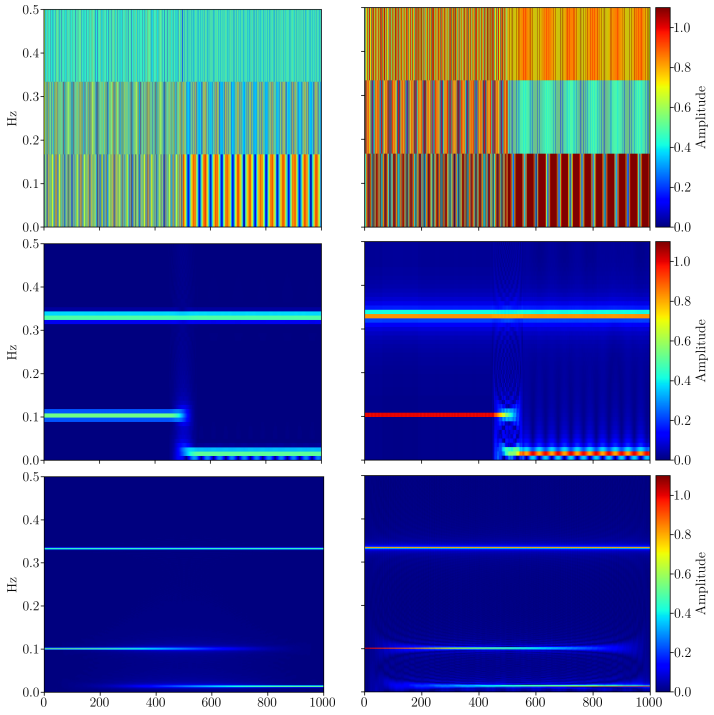


Figure 2. Spectrograms of the signal in Equation (7), with Hamming windows (left) or rectangle windows (right). The windows size are 5, 100 and 500 (top to bottom). Bigger time windows bring better spectral resolution, but worse temporal localisation of the frequency change. The rectangle windows may have artefacts but give a better estimate of the signal component amplitudes.

signal

$$\forall\, t \in [0, 1000],\ X(t) = \begin{cases} sin(2\pi \frac{t}{10}) + sin(2\pi \frac{t}{3}) & \text{if } t < 500, \\ sin(2\pi \frac{t}{80}) + sin(2\pi \frac{t}{3}) & \text{if } t \geq 500. \end{cases} \quad (7)$$

The left column is when using a Hamming windows, the right when using a rectangle windows. The first line is for a window of size 5, the second of size 100, and the third of size 500. As the windows size increases, localising the spectral transition at $t = 500$ becomes harder, but the frequencies are more and more detailed. In the left-column, where the Hamming windows has been used, the spectrogram looks cleaner than with the rectangle window (fewer artefacts) but the amplitude of the original signal components are under-evaluated.

3.2. Applying the Windowed Fourier Transform

By doing the spectrogram of the deteriorated head speed signal as represented in Figure 3, clear frequency patterns appeared. These patterns could be matched with the speed profile with a perfect linear correlation (*cf.* Figure 3, dashed magenta line).

To explain this phenomenon, remember that the measurement system is composed of two scales, one with absolute position

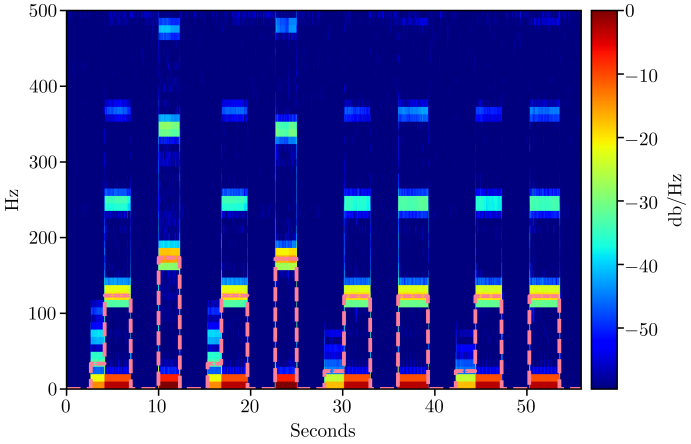


Figure 3. Spectrogram of the Speed signal. In dashed pink, the speed profile superposed and stretched to match the observable patterns (1Hz = 10^{-3}m/s).

reading and one induced by interpolation in-between absolute readings. If we assume that, when the head deteriorates, its interpolation loses accuracy, we could expect the system to recalibrate its position at absolute readings. In fact, when the head is moving at a local speed v, we could expect recalibration events to occur every $\frac{\delta_x}{v}$ where δ_x is the distance between two absolute readings. By computing for small time windows (e.g., , the one used to perform the spectrogram) the average local speed as $\bar{v}(t)$, we can compute the expected recalibration frequency as

$$f_{\text{rec}}(t) = \frac{\bar{v}(t)}{\delta_x} \quad (8)$$

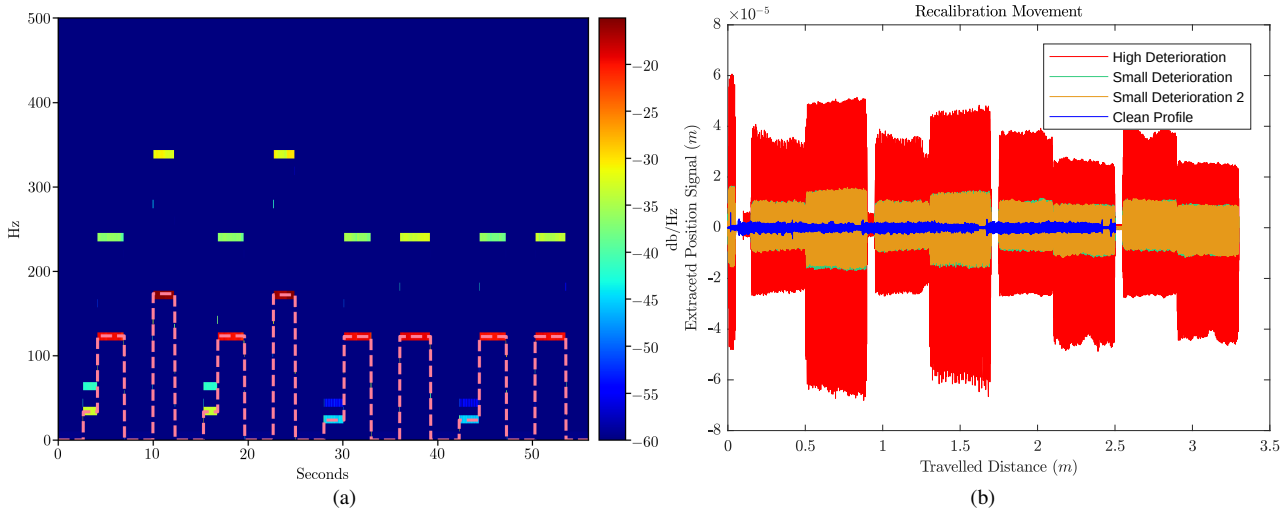
If we set δ_x to 1mm, we have a perfect superposition of the expected recalibration event frequency with frequency patterns in the spectrogram(*cf.* Figure 3). This value for the absolute scale has been confirmed by the system manufacturer.

Note that, if δ_x is 1mm and if F_s is 1000Hz, then the maximum frequency identifiable with the Fourier Transform is 500Hz, which bounds the value of $\bar{v}$ to 0.5m/s.

As we now have identified a particular frequency of interest, it justifies the use of rectangular windows. We are more interested by a correct estimation of the amplitude of this specific frequency than by the exact spectral decomposition of the signal.

3.3. Estimating the recalibration

If our main assumption holds, that is, that deteriorations of the head affect mostly the interpolated position, we could expect a correlation between recalibration amplitude and the measurement head degradation. Therefore, we extracted the part of the spectrogram corresponding to the expected recalibration frequencies and, reconstructed it by applying the Inverse Discrete Fourier Transform. The extracted frequencies and the



(a) (b)

Figure 4. Reconstructed recalibration signals: After extracting frequencies corresponding to expected recalibration frequencies from the spectrogram (a), we apply the Inverse Windowed Fourier Transform (b).

reconstructed signals are represented in Figure 4 for various experiments.

In Figure 4, two observations are important: First, we indeed observe that deteriorated heads tend to increase the amplitude of the recalibration signal. This could be a way to detect such problems. Yet, the second observation is that the amplitude of the recalibration events is varying over time. A simple threshold based detection method might therefore miss some problematic situations.

By comparing those amplitude variations with the speed profile, again, it has been possible to find an additional clear linear correlation between the local average speed and the amplitude of the signal components at frequencies corresponding to the recalibration events. Figure 6a illustrates this correlation.

This led to the first major results of this study: There is a double dependency between the recalibration events and the speed of the measurement head: The frequency of the recalibration events and the amplitude of those frequencies are both linearly correlated with the local average speed. Taking into account these two dependencies, we proposed a first threshold-based detection of problematic measurement: by monitoring the linear coefficient between the Fourier Coefficient at frequencies corresponding to expected recalibration events and the local average speed.

4. SIGNAL CLEANING: THE MEAN ABSOLUTE DEVIATION

4.1. Cleaning the Spectrogram

The spectrogram in Figure 3 contains also frequencies lower than that corresponding to expected recalibration events.

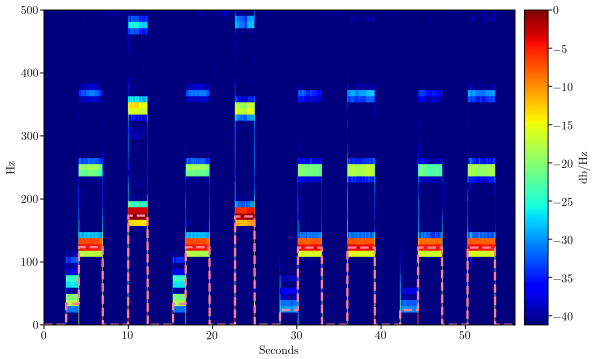


Figure 5. Spectrogram on the difference between control and read speed. Most frequencies but the recalibration event frequencies and harmonics disappeared

These frequencies could come from the windows but also from the head movement profile. Fortunately, among the available variables, the position and speed set by the controller are also accessible. After shifting adequately the time to compensate delays between controls and read speed, we did the same analysis on their difference, using the control speed for computing the local average speed $\bar{v}$. The resulting spectrogram is represented in Figure 5. Most of the frequencies but the ones corresponding to the expected recalibration events and their harmonics disappeared.

Going on with the same process as above led to very similar but cleaner results, as illustrated in Figure 6b, and we could thus suggest an improved indicator based on the Windowed Fourier Transform Coefficient and the local average speed.

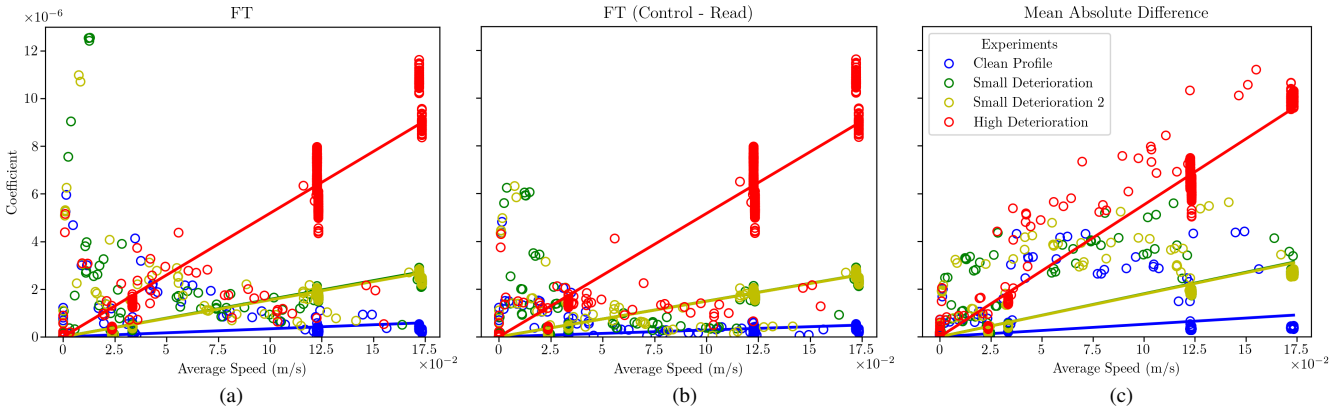


Figure 6. Linear regression between the three proposed indicators and the average local speed. The Fourier Transform Coefficient of the speed (a), the Fourier Transform Coefficient of the difference between control and read speed (b) and the mean absolute difference (c).

4.2. Mean Absolute Deviation

In truth, looking more carefully at the spectrogram in Figure 5, it appears *a posteriori* that the *only* difference between control and read speeds are the expected recalibration events. If this holds, then it seems like the Fourier Transform is not needed at all, one could simply compute the windowed mean absolute difference (MAD). The MAD is defined as:

$$\text{MAD}(n) = \frac{1}{N_w} \sum_{k=n-\frac{N_w}{2}}^{n+\frac{N_w}{2}} |v(k) - \hat{v}(k)|, \tag{9}$$

where N_w is the length of the window, and $\hat{v}$ is the control speed.

Very similarly to the work done so far, a very good linear correlation could be identified between the MAD and the local average speed, and the amplitude of that correlation could also be linked to problematic measurement situations, as illustrated in Figure 6c. These results led us to propose a new indicator for head measurement degradation detection as the linear coefficient between MAD and local average speed.

In Figure 6 comparing the three discussed coefficients, there seems to be, at low speed, a large scattering of the points, especially in Figure 6a. This can be explained to the extent that low speed has been shown to be equivalent to low recalibration frequencies, which could cause a mixture between signals components due to the movement profile with those due to the recalibration. In fact, this scattering decreases with the coefficients computed on the difference between control and read speed. In any case, the density of outsiders is low, as both linear and affine regression converges to the displayed linear models. The remaining scattering can be due to an uneven distribution of the recalibration event causes along the head axis.

4.3. Theoretical Assumptions

The two last indicators, even if stemming from different signal processing tools, look very similar. This encourages us to look for a physical model of the recalibration process, which could explain both results.

Let us focus on a single measurement head and let us assume it is trying to follow the profile $\hat{p}$ set by the controller. If we now assume that deteriorations on the measurement head have for consequences an interpolation error on the position with a coefficient α that only depends on the level of deterioration, then we could expect that at each absolute reading, this error is corrected by the system and brought back to zeros. It could therefore be modelled with periodic function h of period 1 such that the position of the head $p(t)$ follows:

$$\forall\, t \in \mathbb{R}, \quad p(t) = \hat{p}(t) + \alpha \cdot h\left(f_{\text{rec}} t\right) \tag{10}$$

Therefore we have:

$$\forall\, t \in \mathbb{R}, \quad v(t) = \hat{v}(t) + \alpha \cdot f_{\text{rec}} \cdot h'\left(f_{\text{rec}} t\right), \tag{11}$$

The MAD:
Equation (11), can be numerically integrated over N time steps of size $\delta_t = \frac{1}{F_s}$, and with the assumption that $N\delta_t >> \frac{1}{f_{\text{rec}}}$, have:

$$\sum_{n=0}^{N} |v(n) - \hat{v}(n)|\delta_t \simeq \alpha \cdot f_{\text{rec}} \cdot \frac{\beta}{f_{\text{rec}}} \cdot (N\delta_t f_{\text{rec}}) \tag{12}$$

where $\frac{\beta}{f_{\text{rec}}}$ is the integral of h' over one period (of size 1) and $N\delta_t f_{\text{rec}}$ is the number of period of h' in N time steps. Whether h is a sinusoid ($h(x) = sin(2\pi \cdot x)$), a rectangle or a triangle, we have $\beta = 4$.

And the MAD is therefore

$$\begin{aligned} \text{MAD} &= \frac{1}{N}\sum_{n=0}^{N} |v(n) - \widehat{v}(n)|\delta_t \\ &\simeq \alpha \cdot f_{\text{rec}} \cdot \beta \\ &\simeq \alpha \frac{\bar{v}}{1\text{mm}} \cdot \beta \end{aligned} \tag{13}$$

The Fourier Transform:
Under the assumption that the error h is a periodic function, it has a decomposition in Fourier Series:

$$\forall\, x \in \mathbb{R}, \quad h(x) \simeq \sum_{n=0}^{+\infty} C_n e^{-2i\pi n \cdot f_{\text{rec}} \cdot x} \tag{14}$$

Therefore

$$\forall\, x \in \mathbb{R}, \quad h'(x) = \sum_{n=0}^{+\infty} 2\pi n f_{\text{rec}} C_n e^{-2i\pi n \cdot f_{\text{rec}} \cdot x} \tag{15}$$

It is then easy to convince oneself that the Fourier Transform coefficient of h' at frequency f_{rec} (in the series, when $n = 1$) is actually $2\pi f_{\text{rec}} C_1$.

$$\begin{aligned} FT(h')(\nu = f_{\text{rec}}) &= 2\pi f_{\text{rec}} C_1 \\ &\simeq 2\pi \frac{\bar{v}}{1\text{mm}} \alpha \end{aligned} \tag{16}$$

C_1 being the amplitude of the main mode ($n = 1$), we expect $C_1 \simeq \alpha$.

Finally, combining Equations (13) and (16), we found that,

$$\alpha = \begin{cases} \frac{\text{MAD}}{4000 \cdot \bar{v}}, \\ \\ \frac{\text{FT}(|v - \bar{v}|)(\nu = f_{\text{rec}})}{2000\pi \cdot \bar{v}}. \end{cases} \tag{17}$$

As a last remark, when looking at Equation (10), one could ask why the analysis is not directly done with the position signal. From a computational point of view, the difference would be minimal as speed is anyhow needed to compute the expected recalibration frequency $f_{\text{rec}} = \frac{\bar{v}}{10^{-3}}$. The second reason is experimental, working with the speed always gave us much cleaner and clearer results. Both ways, the results were in any case consistent.

5. CASE STUDY SUMMARY

5.1. Results and Theory matches

The works and analysis so far led us to engineer three different indicators for each head. Two are based on the Windowed Fourier Transform and one is on MAD. The indicators themselves are the coefficient stemming from the linear regression on the Fourier Coefficients or the MAD and the average local speed (α above).

In laboratory-controlled experiments, we could find a good match between results and the proposed theory. For example, Figure 7 represents for many experiments ordered by level of recalibration amplitude, the very good correspondence between the three indicators, which also matched observed behaviour. Thresholds could be proposed to separate acceptable recalibration amplitudes (blue datasets), increased but non-problematic recalibration amplitudes (green to yellow datasets) and problematic situations (orange to red).

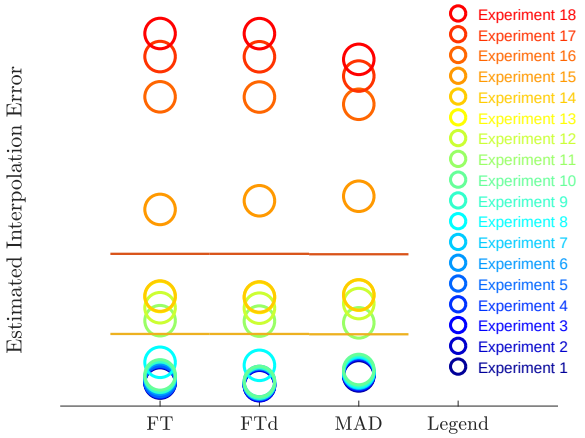


Figure 7. Comparison of 18 different experiments with varied level of deterioration. The three columns correspond to the three way to estimate recalibration event amplitudes: Linear regression on (1) the Fourier Transform Coefficient of the speed (FT), (2) the Fourier Transform Coefficient of the difference between control and read speed (FTd) and (3) the Mean Absolute Difference (MAD).

5.2. Indicator Selection and Degradation Detection Framework

If in laboratory-controlled experiments, we found an excellent match between the three proposed indicators, they are actually not equivalent. The Fourier Transform based indicators have been designed to detect and quantify recalibration events happening at a specific frequency f_{rec}, which is the frequency at which we would expect those events to occur. As such, they would probably not detect other anomalies with different impacts. Yet, the two indicators seem redundant, and as the one based on the difference between control and read speed gave cleaner results, it was the one kept.

To the opposite, the MAD-based indicator relies on the observation that, in our experiments, the only difference observed between control and read speeds, stemmed from recalibration events. Nonetheless, any causes that would impact one of the two speeds would make this indicator rise and we would detect the abnormal situation.

As both indicators have different meaning, we proposed a detection framework that would take advantage of all. This framework is represented in Figure 8. First, for each head, the read speed signal is cleaned thanks to high pass filter. If control speed signal is available, a simple difference between

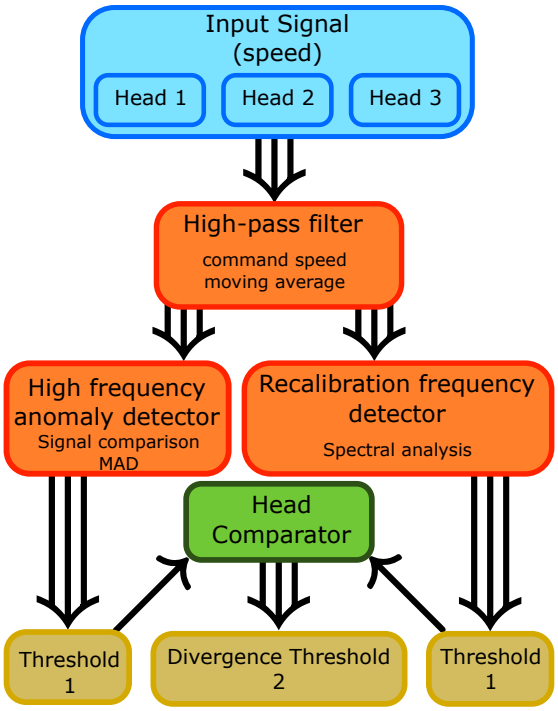


Figure 8. Proposed and implemented framework: The speed is filtered from "expected" mouvement thanks to control speed or a moving average filter. Then both the Fourier Transform based indicator and the MAD based indicator are computed. In addition, these indicators are jointly monitored for the three heads to detect growing divergences.

read an control is a way to remove any normal behaviour of the signal. If the control speed is not available, we could also show that removing from the read signal its moving average would achieve very similar results. Second, the windowed MAD and the coefficients of the Windowed Fourier Transform are computed on this filtered signal, and a linear regression with the local average speeds is performed. This gives two estimates of recalibration events, which are compared to detection thresholds. If both the MAD and the Fourier Transform indicators exceed the threshold, there is a recalibration problem and the measurement head should be cleaned and recalibrated. If only the MAD indicator exceed the threshold, then there is another problem, which is not likely a recalibration problem. The Fourier Transform indicator should normally not be exceeded alone. As a last step, monitoring and comparing the indicators on the three heads of the machine is a way to detect growing divergence and the degradation process of a single head. This information can later be used for predictive maintenance.

This framework has been implemented for a fleet of machines working in real condition and has been so far giving satisfying results.

6. CONCLUSION

The process described in this paper demonstrated how machine learning first, then signal processing, could be used as tools to understand a system and help design an efficient solution for a problem on which there was, at first, very little knowledge. From a functional but not ideal machine learning solution, using Hierarchical Extreme Learning Machines, we could identify behaviours in the data that led us toward more adapted tools and processes. If at first it looked like the difficulty of the problem at hand would be transfer learning to detect abnormal measurements in various operating conditions and for various machines, it appeared eventually that well designed indicators, based on spectral analysis and on data comparison could provide excellent results. This case study is a strong reminder that powerful tools can be a solution, but they also can bring valuable information that would help to design solutions more adapted to the problem at hand.

The solutions implemented here are only limited by the physic of the system, that is the impossibility to detect in a signal, a frequency above half that of the sampling frequency. This corresponds in our case to 500 Hz, or 0.5m/s, a speed easily attainable. The results are also dependent on the right choice for the windows length on which to compute the Fourier Transform and the Mean Absolute Difference. Having already quite a strict limitation on the upper frequency detectable, we also demonstrated that both the frequency of interest and its amplitude are proportional to the speed. As such, low frequencies (in particular under 50Hz) means low speeds but also small amplitudes and higher scarcity of the data (*cf.* Fig. 6). These two boundaries make the choice of a window size very limited, and as such, not really a challenging decision. These strong constraints, both on the window size and on the frequency of interest (the recalibration frequency only) made the use wavelets not deemed relevant, although more powerful for spectral analysis (Peng & Chu, 2004)[3].

Further research would aim at analysing more in depth the root causes for data scattering and study what would be the impact of errors in the read speed, so far considered as reliable, on the linear regression and on the detection reliability. Such considerations could also be included in frameworks comparing the head position distribution such as Matching Pursuit or in probabilistic frameworks such as Bayesian inference.

ACKNOWLEDGEMENT

This research was funded by the Swiss Commission for Technology and Innovation under Grant no. 25728.1 PFES-ES.

[3] instead of using a window of fixed size to convolve with the signal and the pure harmonics (the term $e^{-2i\pi n}$), families of oscillating functions whose limited supports, are used for the transform.

REFERENCES

Dalpiaz, G. & Rivola, A. (1997, January 1). Condition Monitoring and Diagnostics in Automatic Machines: Comparison of Vibration Analysis Techniques. *Mechanical Systems and Signal Processing*, *11*(1), 53–73.

Goyal, D. & Pabla, B. S. (2016, December 1). The Vibration Monitoring Methods and Signal Processing Techniques for Structural Health Monitoring: A Review. *Archives of Computational Methods in Engineering*, *23*(4), 585–594.

Jardine, A. K. S., Lin, D. & Banjevic, D. (2006, October 1). A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510.

Michau, G., Palmé, T. & Fink, O. (2017, October). Deep Feature Learning Network for Fault Detection and Isolation. In *Annual Conference of the Prognostics and Health Management Society 2017*. Annual Conference of the Prognostics and Health Management Society 2017, St. Petersburg, Florida.

Michau, G., Palmé, T. & Fink, O. (2018, July 2–5). Fleet PHM for Critical Systems: Bi-level Deep Learning Approach for Fault Detection. In *In review*. Fourth european conference of the prognostics and health management society 2018, Utrecht, The Netherlands.

Michau, G., Yang, H., Palmé, T. & Fink, O. (2018, February). Feature learning for fault detection in high-dimensional condition-monitoring signals. *submitted*, 1–10.

Mirafzal, S. H., Khorasani, A. M. & Ghasemi, A. H. (2016, November 1). Optimizing Time Delay Feedback for Active Vibration Control of a Cantilever Beam Using a Genetic Algorithm. *Journal of Vibration and Control*, *22*(19), 4047–4061.

Newland, D. E. (1989, August 1). *Mechanical Vibration Analysis and Computation*. Harlow, Essex: Addison-Wesley Longman Ltd.

Peng, Z. K. & Chu, F. L. (2004, March 1). Application of the Wavelet Transform in Machine Condition Monitoring and Fault Diagnostics: A Review with Bibliography. *Mechanical Systems and Signal Processing*, *18*(2), 199–221.

Scheffer, C. & Girdhar, P. (2004, September 23). *Practical Machinery Vibration Analysis and Predictive Maintenance* (1 edition). Amsterdam ; Boston : Burlington, MA: Newnes.