



HAL
open science

Diagnosis of spindle failure by unsupervised machine learning from in-process monitoring data in machining

Victor Godreau, Mathieu Ritou, Cosme de Castelbajac, Benoit Furet

► To cite this version:

Victor Godreau, Mathieu Ritou, Cosme de Castelbajac, Benoit Furet. Diagnosis of spindle failure by unsupervised machine learning from in-process monitoring data in machining. *International Journal of Advanced Manufacturing Technology*, 2024, 131, pp.749-759. <10.1007/s00170-023-11834-y>. <hal-04200150>

HAL Id: hal-04200150

<https://hal.science/hal-04200150v1>

Submitted on 8 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Diagnosis of spindle failure by unsupervised machine learning from in-process monitoring data in machining

Victor Godreau^{1,2}, Mathieu Ritou^{1*}, Côme de Castelbajac^{1,3}, Benoit Furet¹

¹ LS2N (Laboratory of Digital Sciences of Nantes, UMR CNRS 6004), Nantes Université, 1 Quai de Tourville, 44000 Nantes, France

² Europe Technologies, rue de la fonderie, 44470 Carquefou, France

³ Mitis, 12 rue Johannes Gutenberg, 44340 Bouguenais, France

*Corresponding author: mathieu.ritou@univ-nantes.fr orcid.org/0000-0002-0568-9383

Abstract:

In High Speed Machining (HSM), process performance is closely linked to the optimization of cutting conditions and spindle exploitation. Keeping high levels of productivity and machine availability with limited costs is important. However, machining incidents, such as abnormal vibration or tool failure, can cause spindle failure and machine downtime. Consequently, identifying which kind and which severity of machining incident can damage an HSM spindle is critical (as well as which evolution of spindle vibration signature reveals a damage). For that purpose, in-process monitoring data and spindle condition monitoring data are analyzed by Knowledge Discovery in Database (KDD), with a dedicated method to the machining process. Since daily spindle vibration signatures are measured, the in-process monitoring data needs to be daily aggregated. An original unsupervised co-training by Genetic Algorithm is then proposed for the diagnosis of HSM spindle, in order to determine which machining events are critical for the spindle condition. Afterwards, preventive actions can be taken. The approach was applied to three spindle lifetimes, during which the monitoring data were collected for two years of machining of aeronautic structural components. Two major causes of spindle failure were then identified.

Keywords: Knowledge Discovery in Database, unsupervised Machine Learning, Machining, Monitoring, Maintenance.

1. Introduction

Since the 90's, informatics has progressed in two important ways for the exploitation of in-process manufacturing data. In the one hand, the quantity of data that a monitoring system can treat and store was increased; on the other hand, new methods were proposed for Big Data analytics [1]. Many fields are now improving their expertise using Knowledge Discovery in Database (KDD) methods [2]. For example, in educational sciences, students' performance were analyzed with this kind of strategy using decision tree algorithms [3]. Marketing departments also takes decisions supported by customers habits [4]. Nevertheless, in many fields collected data is still under-exploited.

In particular, in the manufacturing field, there are many applications of KDD methods [5]. The studied issues in the literature can be in varied domain. For example, quality improvement [6] where lots of data are collected, have made significant progress analyzing them. In the shop floor control, association rule induction [7] have optimized their efficiency. Another field is the preventive maintenance [8] where data can be treated in order to anticipate future needs. Concerning machining, recent works start developing such approach for process improvement [9,10]. An abundant literature deals with the in-process monitoring of tool condition monitoring, chatter detection or prediction of surface roughness [11]. Both physics-based approach or Artificial Intelligence (AI) algorithms have been used [12]. For example, an approach based on mechanical model were proposed to detect tool failure by the analysis of eddy current signals in the angular domain in [13]. Also, an effective chatter criterion was obtained from the physical modeling of the process [14]. Other technics like cyclo-stationarity has also been used to detect tool failure or chatter in milling process [15]. On the AI side, Bayesian inferences used in neural network [16] is applied on experimental data to estimate tool wear.

However, the spindle is a key component of a machine-tool, providing its performance, despite a complex dynamic behavior [17]. The monitoring of the spindle condition is also an important issue due to high maintenance costs in the High Speed Machining (HSM). This issue belongs to the larger scientific domain of the condition-based maintenance of rotating machines. The most common technique consists in detecting a bearing failure for its vibration signature, measured with an accelerometer [18]. Other techniques, like lubricant analysis (e.g. applied for gear box) are not suitable for spindle, since their bearing lubrication is based on grease or air-oil mist. Temperature measurement is performed on spindle, but it only enables late detection of severe bearing damage. Therefore, most of vibration-based approaches focus on the early detection of bearing damage [19], with different signal processing methods, generally performed in the frequency domain. Classically, the objective is the early detection of a localized defect on the bearing ring, by monitoring the bearing fault-induced vibration frequencies, such as BPF0 (Ball Pass Frequency Outrace) for the outer ring [20]. Distributed defects can also be observed on HSM spindle and monitored [21]. However, literature approaches focus on the monitoring of the spindle condition, i.e. the detection of an initial damage and its periodic follow-up, and do not address the identification of the cause of the spindle failure. Thus, unsolved problems remain, concerning the causes of the spindle damages, which is an important industrial issue and also difficult to investigate due to their long lifetimes. Besides, the monitoring tasks focus on only one aspect, the process or the machine condition; but both are not studied together, which is the originality of the paper. It could be tackled by KDD and Machine Learning from both the in-process monitoring and the machine-tool condition monitoring data. One difficulty is that, during intensive industrial production, the ground truth cannot generally be known. Indeed, no-one can manually label several months of real-time monitoring data, i.e. indicating the exact points-in-time when machining events occurred. Consequently, unsupervised learning should be performed.

The aim of this article is to propose an original approach for the diagnosis of HSM spindle, from in-process monitoring data and unsupervised co-training by Genetic Algorithm, in order to determine which machining events are critical for the spindle condition. Physics-based monitoring criteria were selected to detect machining incidents that could potentially damage the HSM spindle; such as chatter, tool breakage, collision or excessive vibrations. A recent device was developed to collect the machining in-process monitoring data record daily signatures of spindle vibration, to evaluate the spindle condition. During the developed KDD process, an aggregation operator transforms the in-process monitoring data into meaningful daily indicators, in order to reveal machining incidents that could damage spindles. The latter should be sufficiently robust for dataset of industrial production. Lastly, since the ground truth is unknown, an unsupervised co-training is introduced, by Genetic Algorithm, to identify the critical machining incidents that can damage the spindles. Statistical implication analysis was adopted to optimize the correlations between machining events and spindle condition evolution. The proposed approach was then applied and validated on a two-year database of machining of aeronautic structural components, with three spindle lifetimes.

2. Diagnosis of spindle failure by Knowledge Discovery in Database

The objective is to determine which kind of machining incident and which severity can damage an HSM spindle, on the one side; as well as which evolution of spindle condition signature reveals a damage, on the other side. However, both were unknown at the beginning of the investigation. Indeed, the lifetime of HSM spindle can be about several hundreds or thousands of hours. The evolution of spindle condition is thus difficult to perceive, for a human being, due to these very long periods. Besides, it is difficult to investigate it scientifically, based on realistic spindle failure caused by industrial productions, since it requires the collect of in-process monitoring data and spindle condition monitoring data during long periods in industry.

Such a collect of in-process monitoring data has been performed in this work, during two years of production of a machine-tool in an aeronautic company. The machine-tool was instrumented with additional accelerometers in the spindle. In-process monitoring criteria were computed by real-time signal processing and continuously collected, as well as contextual information, every tenth of a second. The condition of the spindle was also evaluated, through a daily measurement of the spindle vibration signature.

A process of Knowledge Discovery in Database has then been developed to learn from the data which critical machining incidents can damage a spindle, cf. Figure 1. It is assumed that the damages of spindle are caused by critical incidents occurring during machining. Different physics-based monitoring criteria were selected for the study, to detect and recognize various kinds of machining incident that could potentially damage the spindle. Similarly, different physics-based criteria were selected for the monitoring of the spindle condition, to detect different potential failure modes of the spindle, and particularly of its rolling bearings.

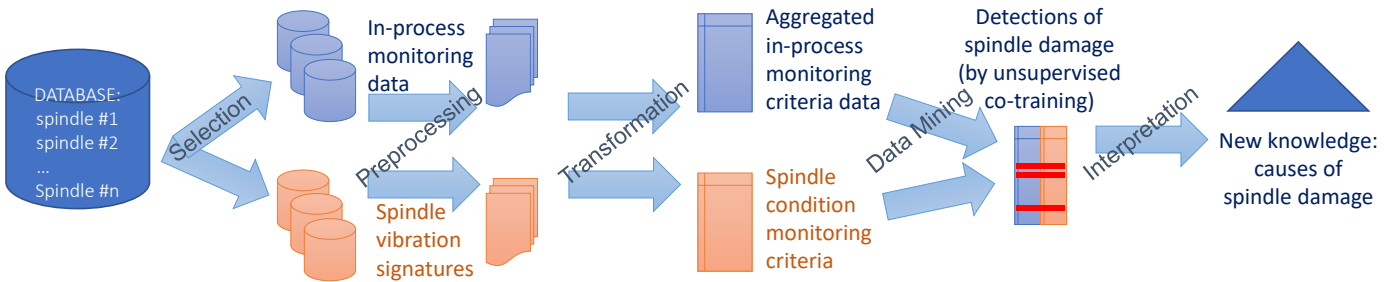


Figure 1. Diagnosis of spindle failure by Knowledge Discovery in Database and unsupervised Machine Learning.

Since the spindle condition is assessed once a day by the vibration signature, the in-process monitoring data thus needs to be aggregated per day, in order to enable daily comparisons. However, machining data of industrial production is complex. A great attention should be paid to data preparation and cleaning. The extraction of classical engineering feature from time series is insufficient. A specific processing is needed to assess the severity of a machining incident, notably through critical machining vibrations, that would exceed a given critical threshold.

As the ground truth is unknown, the dataset is unlabeled and an unsupervised Machine Learning is necessary to determine which day a critical machining incident has damaged the spindle. To do so, a Genetic Algorithm was chosen to perform this task of unsupervised co-training; by determining the critical thresholds beyond which the in-process incidents can damage the spindle, on the one side, and beyond which the evolution of the spindle condition signature reveals a damage of the spindle. Finally, the interpretation of the results of spindle damage detections should conduct to the knowledge of the causes of the spindle failures.

3. Data collect on industrial machine-tool

In the case study, three Fischer MFW-2310 spindles (24kRPM, 70kW) have been monitored during their whole lifetime, during their production in an aeronautic factory. In this way, 0.5 Gb of industrial in-process monitoring data were daily collected for two years. This database is separated in two different sets of data. The first one is the in-process monitoring dataset. The second is the spindle vibration signatures that estimate the spindle condition. Both are collected by EmmaTools device, cf. Figure 2 [10].

The in-process monitoring dataset is composed of two sources of data. Firstly, the HSM spindle is instrumented with four additional accelerometers (two at the front bearing of the shaft and two rear bearings, in both radial directions – in blue Figure 2), measuring the vibrations at a sampling rate of 51.2 kHz, close from the cutting zone. The second source is the sensors already installed in the machine-tool and its control information, which are collected from the Computerized Numerical Command (CNC) by field bus every tenth of a second. In this way, information related to the in-process monitoring (vibration criteria, spindle power or speed, feedrate...) and to the machining context (tool, program, date...) is continuously recorded, 24 hours a day and 365 days a year.

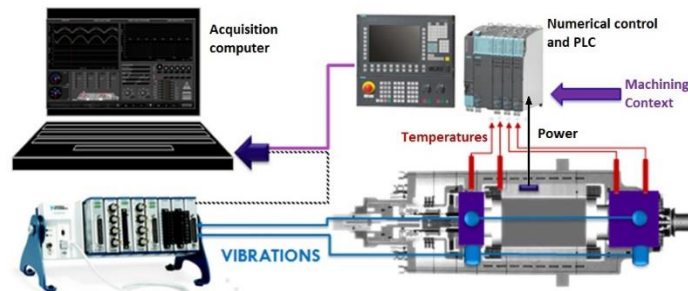


Figure 2. Device for the collect of in-process monitoring data, as well as vibration signatures of the spindle condition.

Besides, the spindle condition is estimated by the spindle vibration signatures, which is classical in condition-based maintenance, using the four accelerometers. To do that, a specific protocol was set up. A dedicated program is run by the CNC once a day. The spindle operates idle rotation at 17500 RPM. In order to obtain reproducible data, the vibration signature is measured every day at the same position, with the same tool and at the same steady state operating temperature, after a warm-up cycle.

4. Monitoring criteria

4.1 In-process monitoring criteria

Several in-process monitoring criteria, presented below, were selected to detect different kinds of incidents during machining, which could potentially damage the spindle.

Model-based in-process monitoring criteria were proposed by Godreau et al. [10] for the detection of chatter (*Nh*) and tool breakage (*Ub*). Since the spindle speed is measured, order tracking can be performed in the frequency spectrum. In this way, the monitoring criteria evaluate respectively the asynchronous vibration level (to detect instability of the cut) and the tool unbalance (as a consequence of a cutter breakage).

The detect of collision is also proposed in the frequency spectrum. The duration of the impact of a collision is assumed to last few tenths of a second, which corresponds to an excitation of a low frequency band, cf. Figure 3. Thus, the proposed criterion estimates the magnitude of the low frequency contributions, in the frequency spectrum. Collisions are detected by the proposed criterion, noted *C*, defined as the sum of the amplitude of the frequency band under the spindle rotation frequency.

The classical criteria for the assessment of a vibration level V_{rms} and A_{rms} (Root Mean Square of vibration Velocity and Acceleration respectively) were also considered in this study, as possible cause of spindle damage.

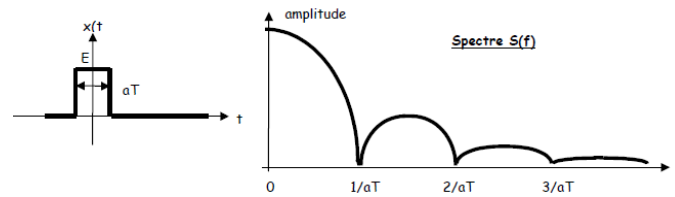


Figure 3. Collision model in time and frequency domains.

The recording of raw accelerometers signals at high sampling rate would lead to excessively big database. Therefore, real-time signal processing is computed on each accelerometer signal, to evaluate the in-process vibration monitoring criteria (every tenth of a second). The criterion value from each accelerometer is noted for example $V_{rms,x,fr}$, for the vibration velocity, of measuring axis x or y and position *front* or *rear* bearing. In this way, all the criteria are calculated and recorded every tenth of a second, for each accelerometer.

4.2 Spindle condition monitoring criteria

In order to evaluate the spindle condition, three criteria are computed on the vibration signatures that are measured every day. The vibration spectrum of rotating machines is composed of several frequencies. Those frequencies are due to the rotations of the parts composing the ball bearings [20]. For example, one of those frequencies corresponds to the passage of each ball on the outer race of the bearing. If the outer race got a defect, a shock will occur each time a ball falls into these defects. This phenomenon is detectable in the vibration spectrum. BPFO, BPF1 and FTF are frequencies that correspond to respectively the defects on the outer race, inner race and cage of the bearings. Another frequency is followed at the shaft speed (f_s), noted 1X, that represents the unbalance of the spindle shaft. A thin frequency bandwidth is monitored around these fault-induced characteristic frequencies and shaft one, in the frequency spectrum of the spindle vibration signatures. As an example, BPFO criterion will be explicitly explained.

To calculate the criterion, the signals acquired during the spindle vibration signatures are filtered around the characteristic frequencies (Table 1). The cutting frequencies of the bandpass filter are calculated to consider variations of temperature that can influence the theoretical frequencies. After the bandpass filtering, the amplitudes of the BPFO peaks in the daily vibration signature signal are averaged. In Figure 4, BPFO is calculated over the lifetime of the third spindle. In condition-based maintenance, it is known that a variation (increase or decrease) of the criterion level can reveal a damage of rotative machine. This is the reason why the variations of BPFO between two consecutive days is computed, noted Δ_{BPFO} .

Bearing or shaft damage	Theoretical frequency		Bandpass filter	
	Front bearing	Rear bearing	Low frequency	High frequency
BPFI	$13.7 f_s$	$13.66 f_s$	$13.25 f_s$	$13.84 f_s$
BPFO	$11.3 f_s$	$11.34 f_s$	$11.19 f_s$	$11.68 f_s$
FTF	$0.45 f_s$	$0.45 f_s$	$0.3 f_s$	$0.6 f_s$
1X	$1 f_s$	$1 f_s$	$0.95 f_s$	$1.05 f_s$

Table 1. Bearing fault-induced frequencies of the spindle (left) and the characteristics of the bandpass filter (right).

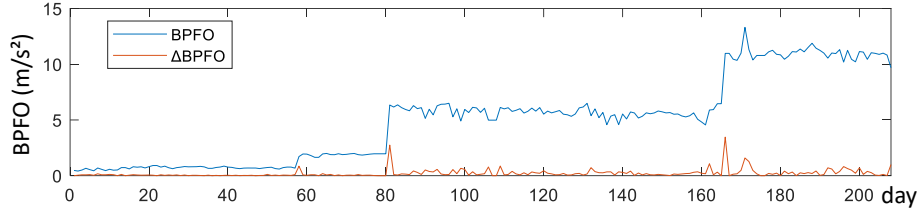


Figure 4. Spindle condition monitoring criterion BPFO during the complete lifetime of the third spindle, and variation Δ_{BPFO} between consecutive days.

5. Aggregation methods and operators

The criteria explained above are computed in real-time and collected. The dataset is thus composed of, on the machining monitoring side, 5 criteria (Vrms, Arms, Nh, Ub, C) calculated on the 4 accelerometers every 0.1s; and on the spindle condition side, 4 criteria (BPFO, BPFI, FTF, 1X) calculated on each accelerometer once a day from the spindle vibration signatures.

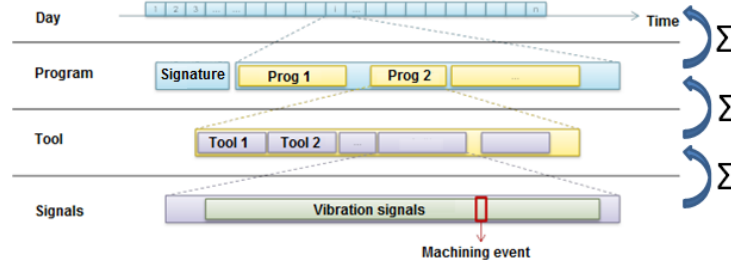


Figure 5. Structure of the in-process monitoring data, which needs to be daily aggregated.

5.1 Aggregation of the in-process monitoring criteria

In order to compare the two datasets, they have to be at the same time scale (i.e. same sampling frequency). To do that, an aggregation operator has been introduced to aggregate the machining vibration monitoring data, once a day. It is called *CO*.

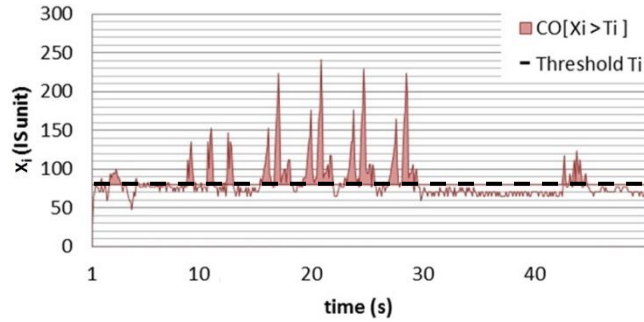


Figure 6. Criticality operator.

The Criticality Operator (*CO*) (Figure 6) considers the duration over a given critical threshold and the quantity over that threshold. Let $X_i = \{x_i(k), k = 0..n\}$ the i^{th} time series representing a monitoring criterion. So, the operator is defined as:

$$CO[X_i > T_{i,j}] = \sum_{k=ti}^{tf} \max(x_i(k) - T_{i,j}, 0) . dt \quad (\text{Eq. 1})$$

with $T_{i,j}$ a critical threshold, ti and tf the initial and final time of the aggregated period. The criticality of a given time series is computed in relation to a critical threshold. However, the value of the critical threshold will only be known at the end of the data mining process. Thus, several thresholds are tested. This is the reason why a set of critical operators are computed for each variable X_i taking into account different value j of the critical threshold noted $T_{i,j}$.

In this paper, *CO* is always calculated on a whole day. Computed this way, *CO* is consequently sensible to a large threshold crossing or a long threshold crossing. When the threshold varies, we can see the evolution of each value of *CO* day by day. With this we can conclude if a peak is a long cross or a large cross of the threshold during a day. For example (Figure 7), some peak well represented compared to other with the lower threshold can disappear when the threshold grows (day 101). This means that there was a long time passed over the threshold but the crossing was not very high. To the contrary, if a peak gain importance with the threshold getting higher (day 151), the event was short in time but very high in value. And if a peak is represented regardless of the threshold (day 230), there was both a long and high overtaking of $T_{i,j}$. In this way, the learning of the critical threshold values should reveal which severity of machining incident can damage a spindle.

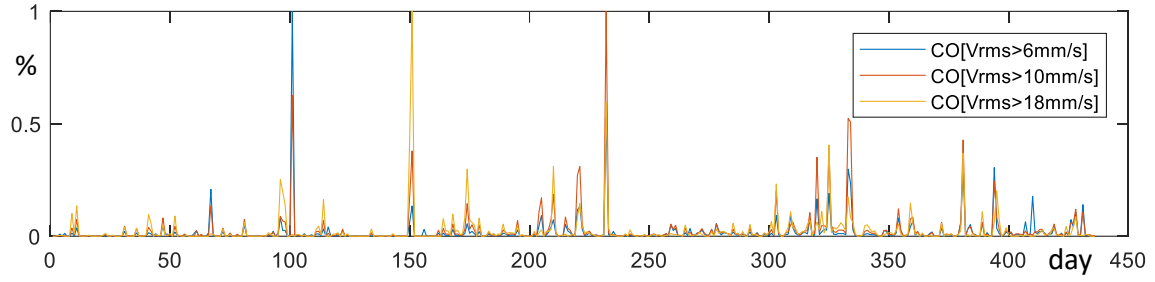


Figure 7. Results of aggregation of vibration level Vrms during machining by CO, for three different values of the critical threshold T_{Vrms} (expressed as a percentage of the maximum for each threshold).

When CO is computed, a criticality of each day concerning the studied criterion is obtained and a low level noise appears regardless the value of $T_{i,j}$. In order to increase the relevance of the information and to correlate the day criticality with the spindle condition, another threshold is needed (noted U_i). It is considered that a machining incident is detected a given day if the value of aggregated data overpasses this second threshold U_i :

$$\text{If } CO[X_i > T_{i,j}] > U_i \text{ then 'incident' = 1} \quad (\text{Eq. 2})$$

Then, this expression has to be correlated with the spindle condition by tuning the critical thresholds $T_{i,j}$ and U_i , during the unsupervised co-training.

Note that, before this aggregation, the time series of in-process monitoring criterion are available for each of the four accelerometers. They are fused, conducting to a unique time series per monitoring criterion. For that purpose, the resultant of the accelerometer values is computed per bearing of the spindle (the front and rear ones) and then the mean value of the two bearings is considered:

$$Z = \left(\sqrt{Z_{x,fr}^2 + Z_{y,fr}^2} + \sqrt{Z_{x,re}^2 + Z_{y,re}^2} \right) / 2 \quad (\text{Eq. 3})$$

where $Z_{x,y,fr/re}$ are the values of monitoring criterion from each accelerometer signal.

5.2 Evolution of spindle condition monitoring criteria

Considering the spindle condition, criteria are evaluated once a day, for each accelerometer. Values per accelerometer are then aggregated in accordance with Eq. 3. Besides, the criteria have to be transformed to reflect daily deteriorations of the spindle condition, instead of an absolute value. The difference between the value for a given day and the previous one is simply computed. In order to learn which evolution of spindle condition monitoring criterion reveals a damage of the spindle, another critical threshold V_k is define as:

$$\text{If } (Y_{i,l} - Y_{i,l-1}) > V_i \text{ then 'damage' = 1} \quad (\text{Eq. 4})$$

Where Y_i is a monitoring criterion of the spindle condition ($BPFO$, $BPFI$, FTF or $1X$) and l the day number.

6. Unsupervised co-training of critical thresholds

Once the aggregation step performed, the aggregated data is composed of one value per criterion, per critical threshold ($T_{i,j}$) and per day; both for the in-process monitoring side and for the spindle condition one. On the in-process monitoring criteria, the operator $CO[X_i > T_{i,j}]$ has been applied on the three database corresponding to the whole lifetimes of the three studied spindles. In order to reduce the computation time during the co-training, aggregations are pre-calculated for discretized values of critical thresholds ($T_{i,j}$), as listed in Table 2, for each monitoring criterion.

Criterion X_i	Vrms (mm/s)	Arms (m/s ²)	Nh (m/s ²)	Ub (m/s ²)	C (m/s ²)
Min	2	20	20	2	5
Max	20	200	200	20	50
Step	2	20	20	2	5

Table 2. Minimum, maximum and step of the discrete values of thresholds $T_{i,j}$ for the co-training.

At this step, the difficulty is to learn all the critical thresholds (T , U and V) in an unsupervised way, in order to optimize the correlations between two following facts:

- A: occurrence of a new degradation of the spindle condition ('damage' = 1).
- B: detection of a machining incident the day before ('incident' = 1).

Since these two conditions should be learnt simultaneously, this task is called "co-training" in Data Sciences [22]. There are too many possible combinations to try to perform the co-training manually. Moreover, the discontinuity of the cost function (due to the change of pre-calculated discretized thresholds) makes difficult the usage of a classic minimization function. Therefore, a Genetic Algorithm has been chosen for the co-training of the critical thresholds.

6.1 Genetic Algorithm

A Genetic Algorithm has been chosen to tune all thresholds (T_{ij} , U_i and V_j) on the criteria in order to maximize the correlations between the 5 machining incident criteria and the 4 spindle degradation criteria. This type of algorithm is needed when the number of solutions to be tested becomes too large. A genetic algorithm is a heuristic method composed of a few steps which explores the possible ensemble and converges towards a good local minimum after the exploration of a lot of possibilities. A total of 14 threshold values need to be learned simultaneously (cf. Table 3). Thus, each individual is depicted by 14 variables. To reduce the computation cost, each threshold (a continuous variable) has been discretized in 10 discrete values, presented in Table 2.

Steps of the algorithm are:

- random generation of the first test population,
- computation of the cost function on each member of the population (Eq.10),
- crossing of members who have the highest score (50%),
- creation of new population adding new children and (5%) highest ranked members of the past population,
- mutation of a part of the new population (5%),
- restarts with the new population.

6.2 Cost function

In every optimization problem, the cost function that evaluates all the members of a population is of major importance. This cost function translates the asked question into a mathematical calculus in order to be executed automatically. In this paper, the question is:

How severe a machining event has to be to damage an HSM spindle?

To answer this question, a statistical approach has been used: the statistical implicative analysis [23]. It quantifies the implication of an event on another one. So, "A implicate B" and "B implicate A" can be implemented and weighted in order to find a good balance in the final cost function. The implication is calculated based on the number of counterexamples ($n_{a\bar{b}}$ number of cases where A is true and B is false) regarding, the size of the explored space (n – the days number), the number of case where A is true (n_a – day with machining incidents) and the number of case where B is false ($n_{\bar{b}}$ days without spindle damage). In this method, it is the unlikelihood to encounter a counter example that is modeled. A standardized random variable representing the number of counter examples can be expressed this way with a Poisson distribution of parameter $\frac{n_a n_{\bar{b}}}{n}$:

$$Q(a, \bar{b}) = \frac{\text{Card}(X \cap \bar{Y}) - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} \quad (\text{Eq. 5})$$

with X and Y two randomly chosen events respecting the number of occurrences of respectively A and B ($\text{Card}(X) = \text{Card}(A)$ & $\text{Card}(Y) = \text{Card}(B)$). This function can be calculated within the particular case where X and Y are respectively A and B. This leads to the:

$$\text{Implication index } A \rightarrow B = q(a, \bar{b}) = \frac{n_{a\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} \quad (\text{Eq. 6})$$

The implication intensity is the probability of having less counter examples in the studied implication than in a random case of same size:

$$\text{Implication intensity of } A \rightarrow B = \varphi(a, b) = 1 - \Pr[Q(a, \bar{b}) \leq q(a, \bar{b})] = \frac{1}{2\pi} \int_{q(a, \bar{b})}^{+\infty} e^{-x^2/2} dx \quad (\text{Eq. 7})$$

with α the confidence level of the implication, this statistical implication analysis shows that the implication of A on B is admissible with a level of confidence of $1-\alpha$ if and only if:

$$\varphi(a, b) \geq 1 - \alpha \quad (\text{Eq. 8})$$

In this study, if only one of those implications is computed as cost function for the algorithm the bijection of the link between A and B cannot be proven. So, it is needed to temper this phenomenon by the calculus of the reciprocal implication:

$$\text{Implication index } B \rightarrow A = q(b, \bar{a}) = \frac{n_{b\bar{a}} - \frac{n_b n_{\bar{a}}}{n}}{\sqrt{\frac{n_b n_{\bar{a}}}{n}}} \quad (\text{Eq. 9})$$

The implication of A on B is favored because the study aims at determining the causes of HSM spindle degradation. Moreover, close to the end of life of a spindle, bearings are already damaged. So, on that period, new machining events may not generate additional damage of the spindle condition. This is reason why a parameter $\beta=0.75$ was introduced in the cost function F (Eq. 10). Finally, a convergence problem can appear, because nothing goes against the increase of A and B, decreasing the critical threshold values to 0 during the training, which conducts to the trivial solution where faulty events would be detected every day. Consequently, a parameter was added to the final cost function to penalize cases where a lot of events would be found and correlated. To find such a compromise, the proposed cost function is:

$$F = -\frac{q(a, \bar{b})}{\beta} - \beta \cdot q(b, \bar{a}) - \sqrt{\frac{n_b n_{\bar{a}}}{n}} \quad (\text{Eq. 10})$$

7. Results

After the co-training by Genetic Algorithm, results of Table 3 are obtained. The critical thresholds have been simultaneously learnt by co-training: T_{ij} and U_i thresholds are for the machining incidents, respectively V_{rms} , Arms, Nh, Ub and C; and V_i ones for the spindle condition monitoring.

In-process monitoring criteria										Spindle condition monitoring criteria			
Vrms		Arms		Nh		Ub		C		Δ_{IX}	Δ_{BPFI}	Δ_{BPFO}	Δ_{FTF}
T_{Vrms}	U_{Vrms}	T_{Arms}	U_{Arms}	T_{Nh}	U_{Nh}	T_{Ub}	U_{Ub}	T_C	U_C	V_{IX}	V_{BPFI}	V_{BPFO}	V_{IX}
mm/s	%	m/s ²	%	m/s ²	%	m/s ²	%	m/s ²	%	m/s ²	m/s ²	m/s ²	m/s ²
4	47	120	35	100	91	0.52	26	/	/	0.60	0.52	1.79	0.5

Table 3. Results of the co-training of critical thresholds, for the detection of machining events and of spindle condition degradation.

Figure 8 presents the results of the co-training, with the critical machining incidents (red circles) detected by each aggregated in-process monitoring criterion, beyond their learnt critical threshold T_i for CO (and a second filtering by U_i , the orange horizontal line), during the three spindle lifetimes (delimited by the green vertical dashed lines). Similarly, Figure 9 presents the results of spindle damages that were simultaneously detected by the co-training, for each spindle condition monitoring criteria.

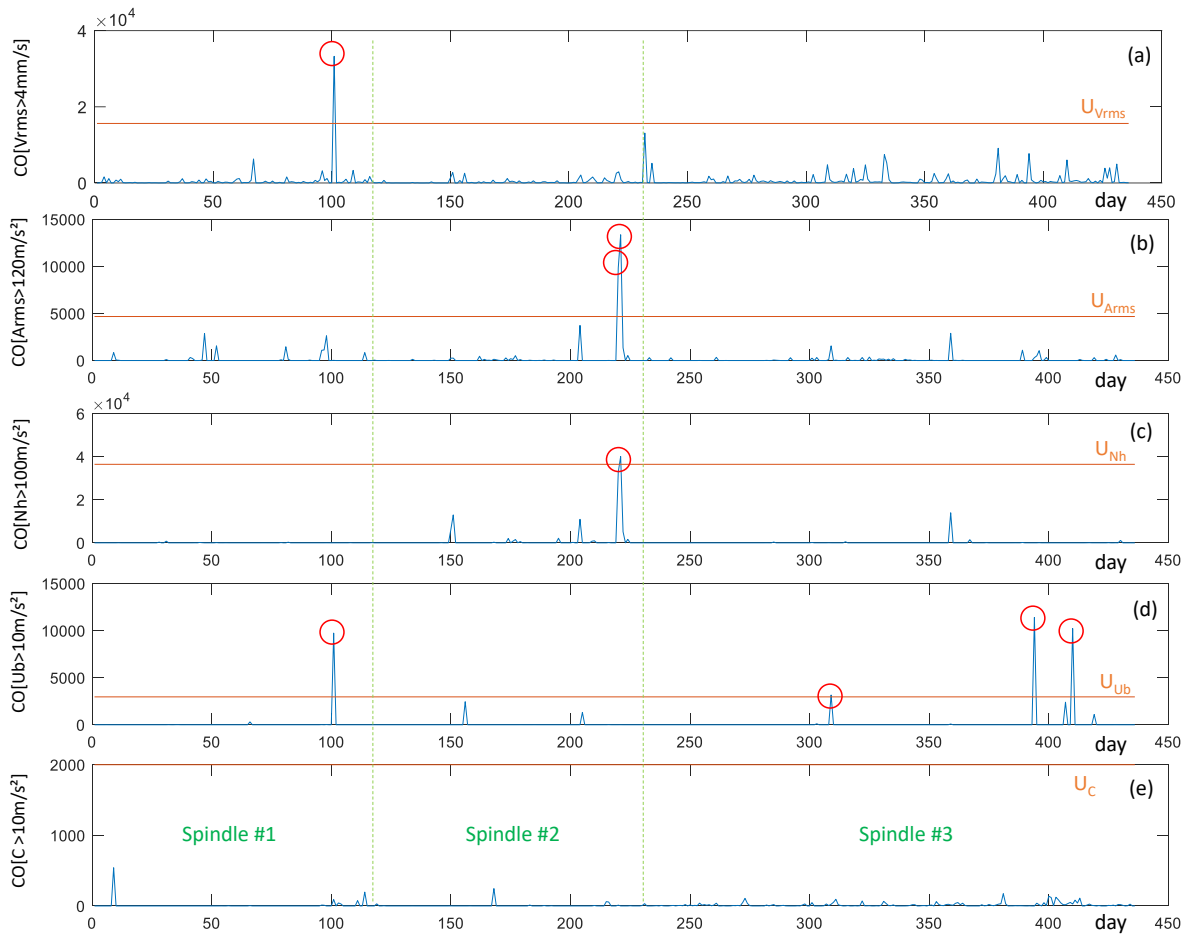


Figure 8. Detection of machining incidents that damage the spindle, from daily aggregations of in-process monitoring criteria and unsupervised co-training by G.A. with the spindle condition, over three lifetimes of HSM spindle in industry.

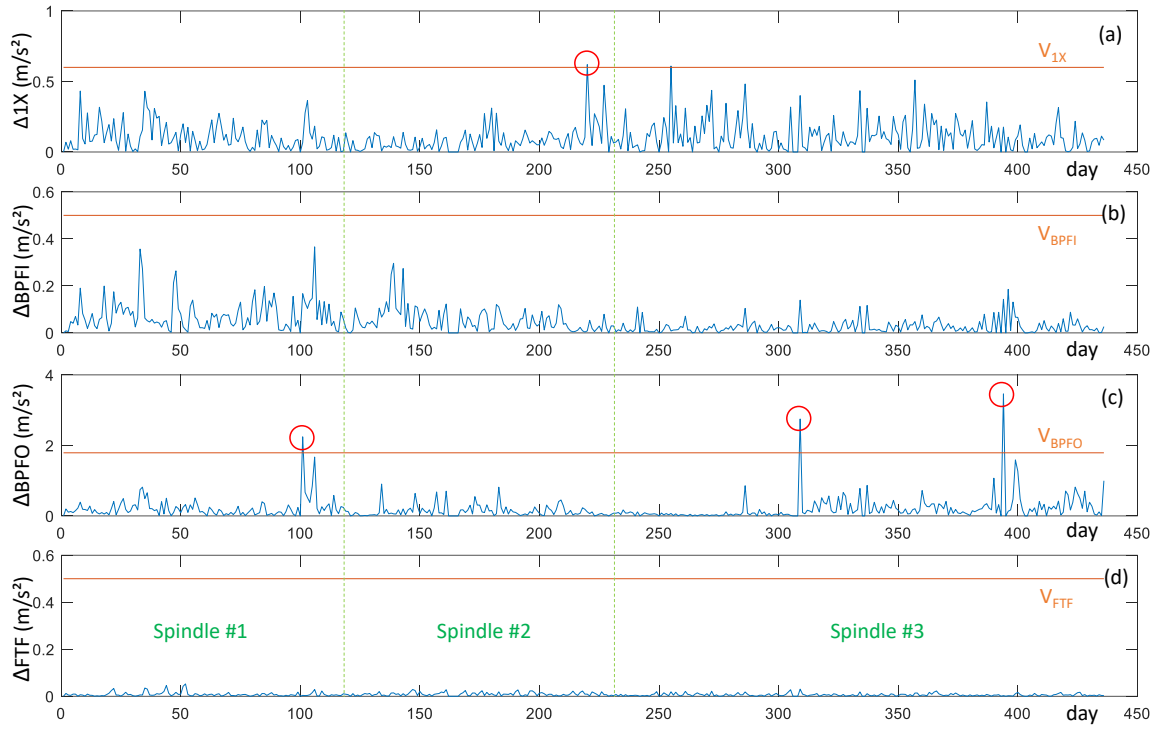


Figure 9. Detection of deterioration of the spindle condition, from the spindle condition monitoring criteria and unsupervised co-training by G.A. with the in-process monitoring, over three lifetimes of HSM spindle in industry.

A total of four spindle degradations ($B, n_b = 4$) were found by the co-training. All of them were explained by the six machining incidents ($A, n_a = 6$) during the $n = 436$ days of the study. It can be noticed that the dataset is strongly unbalanced, with very few events and most of days without problem, during the three spindle lifetimes. In terms of statistic implication, the implication:

- $A \rightarrow B$ is reliable at 99%
- $B \rightarrow A$ is 95%

The high implication of A (machining incidents) on B (spindle damage) means that every spindle damage was caused by a machining incident. The implication of B on A is less reliable. It means that machining incidents were often detected simultaneously with a spindle damage, but not always. The authors explain it by the fact that, if the spindle bearings are already damaged, a new machining incident may not necessarily increase the level of damage.

The critical thresholds T_i have been learnt by the co-training with G.A.. On the side of the in-process monitoring, incidents that damage the spindle are:

- Excessive vibration levels $V_{rms} > 4$ mm/s or $A_{rms} > 120$ m/s²
- Chatter $Nh > 100$ m/s²
- Tool breakage $Ub > 10$ m/s²
- Collision C: no event correlated

For the maintenance point of view, a significant degradation of the spindle condition can be monitored if the thresholds V_i are overpassed:

- $\Delta_{BPFO} > 1.79$ m/s²
- Δ_{BPFI} and Δ_{FTF} : no detection correlated
- $\Delta_{IX} > 0.6$ m/s²

Further analyses can be done. Firstly, three of the four incidents that are correlated with a spindle damage, are due to tool breakages (Ub), cf. the days 101, 309 and 394 in Figure 8. Those events deteriorate the outer ring of the ball bearings, as detected by Δ_{BPFO} criterion. The last event correlated (day 221) was a chatter (Nh) of high amplitude correlated with Δ_{IX} an unbalance or bending of the spindle shaft. It was also detected by A_{rms} . It probably means that this severe chatter has generated very high forces (due to the instability in the cut, conducting to large variations of chip thickness) that have plastically deformed the shalt or the HSK interface. The V_{RMS} criterion has a threshold at 4mm/s, which is reasonable. One event is correlated (day 101). This event was high in amplitude and long, and was also detected Ub as a tool breakage. No collision C was found in the dataset. Likewise, two spindle condition criteria (Δ_{BPFI} and Δ_{FTF}) did not enabled the detection of implication. It can be concluded that the main cause of failure of HSM spindle is some rare tool breakages. In addition, severe chatter can also damage the spindle.

8. Conclusion

In this paper, the cause of the failures of HSM spindle has been investigated, by in-process monitoring of industrial machining productions and data mining. Physics-based in-process monitoring criteria were collected for the detection of machining incidents, as well as for the spindle condition monitoring. A Knowledge Discovery in Database (KDD) approach was developed, in order to identify what king and severity of

machining incidents damage HSM spindles. An unsupervised co-training by (of critical thresholds) by implicative statistics and Genetic Algorithm was proposed, in order to detect correlation between machining incidents and spindle damage. To do so, a dedicated aggregation operator (CO) has been developed to daily aggregate the machining vibrations. The automatic co-training of critical thresholds for the detection of event machining and critical thresholds on the spindle condition criteria has been performed. The principal output of this study is the knowledge of the kind and severity of machining incidents that can deteriorates HSM spindles. Then, based on those results, preventive actions have been taken in the company, and the spindle lifetime has been multiplied by three. Another important aspect is that all the data mining process is automatic. Thus, future databases can be automatically analyzed, to improve the precision of the results obtained in this paper.

Declarations

The authors acknowledge the financial support of the French government for the FUI QuaUsi project. They thank the industrial partners of the project for their contributions. Code or data availability is not applicable. The authors have no conflict of interest or competing interests to declare that would be relevant to the content of this article.

References

- [1] R. X. Gao, L. Wang, M. Helu, and R. Teti (2020) Big data analytics for smart factories of the future, *CIRP Ann.*, 69(2):668–692. doi: 10.1016/j.cirp.2020.05.002.
- [2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth (1996) From data mining to knowledge discovery in databases, *AI Magazine*, 17(3):37–53. <https://doi.org/10.1609/aimag.v17i3.1230>
- [3] B. K. Baradwaj (2011) Mining Educational Data to Analyze Students Performance, *Int. J. Adv. Comput. Sci. Appl.*, 2(6):63–69. <https://doi.org/10.14569/IJACSA.2011.020609>
- [4] M. J. Shaw, C. Subramaniam, G. Woo, and M. E. Welge (2001) Knowledge management and data mining for marketing, *Decision Support Systems*, 31(1):127–137. [https://doi.org/10.1016/S0167-9236\(00\)00123-8](https://doi.org/10.1016/S0167-9236(00)00123-8)
- [5] J. Harding (2006) Data mining in manufacturing: a review, *J. Manuf. Sci. Eng.*, 128:969–976. doi: 10.1115/1.2194554.
- [6] G. Köksal, İ. Batmaz, and M. C. Testik (2011) A review of data mining applications for quality improvement in manufacturing industry, *Expert Syst. Appl.*, 38(10):13448–13467. doi: 10.1016/j.eswa.2011.04.063.
- [7] M.-C. Chen (2003) Configuration of cellular manufacturing systems using association rule induction, *Int. J. Prod. Res.*, 41(2):381–395. doi: 10.1080/0020754021000024184.
- [8] D. Batanov, N. Nagarur, and P. Nitikhunkasem (1994) EXPERT-MM : A knowledge-based system for maintenance management, *Artif. Intell. Eng.*, 8:283–291.
- [9] J. Lenz, T. Wuest, and E. Westkämper (2018) Holistic approach to machine tool data analytics, *J. Manuf. Syst.*, 48:180–191. doi: 10.1016/j.jmsy.2018.03.003.
- [10] V. Godreau, M. Ritou, E. Chové, B. Furet, and D. Dumur (2019) Continuous improvement of HSM process by data mining, *J. Intell. Manuf.*, 30(7): 2781–2788. doi: 10.1007/s10845-018-1426-7.
- [11] R. Teti, K. Jemielniak, G. O’Donnell, and D. Dornfeld (2010) Advanced monitoring of machining operations, *CIRP Ann. - Manuf. Technol.*, 59(2):717–739. doi: 10.1016/j.cirp.2010.05.010.
- [12] D. H. Kim, T.J.Y. Kim, X. Wang et al. (2018) Smart Machining Process Using Machine Learning: A Review and Perspective on Machining Industry, *International Journal of Precision Engineering and Manufacturing-Green Technology*, 5(4):555–568. doi: 10.1007/s40684-018-0057-y.
- [13] M. Ritou, S. Garnier, B. Furet, and J. Y. Hascoet (2014) Angular approach combined to mechanical model for tool breakage detection by eddy current sensors, *Mech. Syst. Signal Process.*, 44(1):211–220 doi: 10.1016/j.ymssp.2013.02.004.
- [14] R. P. H. Faassen, N. van de Wouw, J. a. J. Oosterling, and H. Nijmeijer (2003) Prediction of regenerative chatter by modelling and analysis of high-speed milling, *Int. J. Mach. Tools Manuf.*, 43(14):1437–1446. doi: 10.1016/S0890-6955(03)00171-8.
- [15] M. Lamraoui, M. Thomas, M. El Badaoui, and F. Girardin (2014) Indicators for monitoring chatter in milling based on instantaneous angular speeds, *Mech. Syst. Signal Process.*, 44(1):72–85. doi: 10.1016/j.ymssp.2013.05.002.
- [16] S. Seguy, G. Desein, and L. Arnaud (2008) Surface roughness variation of thin wall milling, related to modal interactions, *Int. J. Mach. Tools Manuf.*, 48(3):261–274. doi: 10.1016/j.ijmactools.2007.09.005.
- [17] Rabréau, C., Noël, D., Loch, S.L., Ritou, M., Furet, B. (2017) Phenomenological model of preloaded spindle behavior at high speed, *International Journal of Advanced Manufacturing Technology*, 90(9):3643–3654. doi: 10.1007/s00170-016-9702-1
- [18] N. Lynagh, H. Rahnejat, M. Ebrahimi, and R. Aini (2000) Bearing induced vibration in precision high speed routing spindles, *Int. J. Mach. Tools Manuf.*, 40(4):561–577. doi: 10.1016/S0890-6955(99)00076-0.
- [19] R. B. Randall and J. Antoni (2011) Rolling element bearing diagnostics-A tutorial, *Mechanical Systems and Signal Processing*, 25(2):485–520. doi: 10.1016/j.ymssp.2010.07.017.
- [20] T. Hoshi (2006) Damage Monitoring of Ball Bearing, *CIRP Annals – Manuf. Technol.*, 55(1):427-430. [https://doi.org/10.1016/S0007-8506\(07\)60451-X](https://doi.org/10.1016/S0007-8506(07)60451-X)
- [21] C. De Castelbajac, M. Ritou, S. Laporte, and B. Furet (2014) Monitoring of distributed defects on HSM spindle bearings, *Appl. Acoust.*, 77:159-168. doi: 10.1016/j.apacoust.2013.07.008.
- [22] S. Yu, B. Krishnapuram, R. Rosales, and R. B. Rao (2011) Bayesian co-training, *Journal of Machine Learning Research*, 12(80):2649-2680. <https://doi.org/10.5555/1953048.2078190>
- [23] R. Gras, E. Suzuki, F. Guillet, and F. Spagnolo (2008) *Statistical Implicative Analysis, Theory and Applications*, Series Studies in Computational Intelligence ed., 127, Springer-Verlag. ISBN 978-3-540-78982-6