Development of data-driven PHM solutions for robot hemming in automotive production lines

Luca Actis Grosso[1], Andrea De Martin[2], Giovanni Jacazio[3] and Massimo Sorli[4]

[1]*Comau, Grugliasco, 10095, Italy*
luca.actisgrosso@comau.com

[2,3,4]*Politecnico di Torino, Torino, 10129, Italy*
andrea.demartin@polito.it
giovanni.jacazio@polito.it
massimo.sorli@polito.it

ABSTRACT

Robotic roller hemming is currently one of the most used solution for joining metal sheets in automotive industry, especially for those production lines which need to favor flexibility with respect to raw productivity and is mostly employed to assemble car doors. Hemming is a fairly delicate process since it does not only suffice a technical requirement – to join two panels together – but also an aesthetical one, since the joint panels are an integral part of the vehicle design and as such an important selling point. An unpredicted rupture or an advanced degradation condition of the system would lead to a significant loss in the quality of the final product or to a sudden stoppage of the production line. The development of a PHM system for hemming devices would hence provide a significant advantage, especially if designed to work for both new and legacy equipment. In this paper, we provide the results of a preliminary analysis of a new PHM framework for robotic roller hemming studied to work without having access to PLC data; the employed data-driven methodology is detailed and applied to the case of increasing wear in the head finger roll. Results from different prognostics routines are hence presented and compared.

1. INTRODUCTION

Maintenance of production systems has a critical impact on the operations and on the performance of manufacturing lines, since its scheduling and effectiveness affect the plants productivity (Gopalakrishnan, Skoohj, and Laroque, 2013), product quality/quality costs (De Ruyter, Cardwell-Hall, and Hodgson, 2010) and the overall production costs (Bevilacqua and Braglia, 2007). To better highlight the importance of maintenance, a recent study by Thomas (2018), estimates that maintenance costs usually represent between the 15% and the 70% of the final cost of goods sold. Traditionally, maintenance on production equipment follows a fixed schedule optimized to minimize the maintenance cost and the occurrence of unpredicted downtimes (Pascual, Meruane, and Rey, 2008). However, almost 50% of the preventive maintenance intervention are deemed as unnecessary (Vogl, Weiss, and Helu, 2016), and unpredicted downtime still represents a significant portion of the total cost of manufacturing and of the production time (estimated in 23.9 % and 13.3% for Sweden by Tabikh, 2014). It is hence clear the appeal that brought to a significant increase of manufacturers interest in developing effective PHM solutions for production systems (Thomas, 2018). The same can hence be observed in literature, with an increased number of contribution in the last few years, covering several fields of applications and different approaches to the problem. In example, Alzahrani, Liu and Kolodziej (2018) recently proposed a data-driven technique to detect and classify tools wear in end mills through the analysis of acoustic emission. A model based approach was instead proposed by Longo, Serpi, Jacazio and Sorli (2018) to enhance the maintenance of automotive industry plants, while data-driven PHM solutions in the semiconductors and electronics business have been recently presented as well (Li and Wu, 2018), (Singh, Selvanathan, Zope, Nistala, & Runkana, 2018). Focusing on the automotive industry, one of the most critical operations performed during cars assembly is the hemming, that is the technology used to join inner and outer closure panels together by bending or folding the flange of the outer panel over the inner one. (Le Maoût, Thuiller and Manach, 2010). Among the available solutions, roller hemming has become one of the most frequently adopted (Esquivel, Carbone, Ceccarelli, Jáuregui, 2016); this process makes use of a rolling element to bend the metal sheets, providing a less traumatic deformation to the panels and hence resulting more suitable for low weight designs and materials (Montalbo, Roth and Kirchain, 2008). To allow for maximum flexibility of the manufacturing lines, roller hemming is nowadays mostly performed by means of a robotic arm (Saboori, Saboori, Carlson and Söderberg, 2009). Such solution is however extremely sensitive to change in the robot performances (Drossel, Pfeifer, Findeisen, Rössinger, Eckert and Barth, 2014) making it a case of great interest for both the PHM community and the manufacturers. This article explores a possible PHM solution for the problem, detailing the followed approach and the first results.

2. MOTIVATION

The World Class Manufacturing Association is a not-for-profit organization of manufacturing companies in May 2006, which objective is to foster development and application of the best production practices, contributing to a more competitive system of production to the benefit of member companies, their plants and, of course, end customers. The WCM Association promotes knowledge sharing among member companies. In addition, it appoints an official Auditor to support the program and assign WCM Awards to the leading plants. FCA plants worldwide adopted the World Class Manufacturing as production methodology, which involves the entire organization, from safety to environment, maintenance, logistics and quality. The primary objective of WCM is continuous improvement in all areas of production in order to guarantee the quality of the final product and meet customer expectations. Projects developed under the WCM, which rely on a high level of employee involvement, target the elimination of all forms of waste and loss with the ultimate objective of achieving zero accidents, zero waste, zero breakdowns and zero inventory. World Class Manufacturing has defined seven steps for improving maintenance, described in figure 1 (FCA, 2016).

- STEP 1 - Preventive elimination of forced deterioration
- STEP 2 - Preventive breakdown analysis
- STEP 3 - Preventive definition of maintenance standard
- STEP 4 - Countermeasures against weak points of the machine and lengthened equipment life
- STEP 5 - Build a periodic maintenance system
- STEP 6 - Build a CBM system
- STEP 7 - Maintenance cost management

The first three steps are usually reached after the ramp-up phase in new plants, while older ones starts from the first. Prognostics (as CBM); seats at the sixth step in the chain, and its development is admitted during production. This means that prognostics solutions might be developed for in-service or legacy equipment, hence posing additional challenges in terms of signals availability or access to the machine.

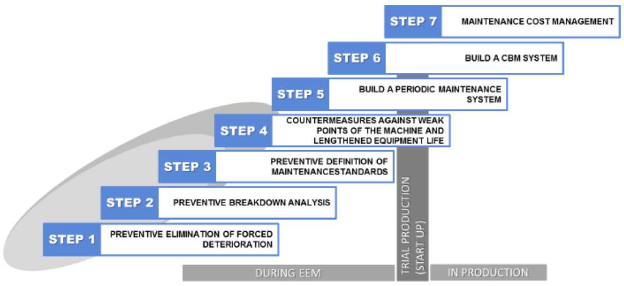

Figure 1. Steps for maintenance improvement according to WCM

The application that will be presented in Section 4 is indeed part of this class of machinery, being already in service in several production lines. Given the issues in having access to the PLC data, the proposed PHM system is based on the use only a handful of sensors (which might be additional or pre-existent depending on the application), placed in easily accessible sections of the machine.

3. METHODOLOGY

Developing a PHM system for in-service machinery presents the advantage of having easy-access to data from the field, hence providing consolidated input to FMECA analysis and a nice assets of information to assess the nominal behavior of the hemming system. Almost more importantly, the availability of these data allows to highlight the occurrence of practical issues connected to the use of these devices which are not usually considered in simulation-only environments, hence boosting considerably the confidence in the developed feature extraction method. At first, a FMECA analysis is used to study the known issues that may occur in the hemming machine and hence to rank them based on severity, frequency and observability criteria (Vachtsevanos, Lewis, Roemer, Hess, and Wu, 2006).

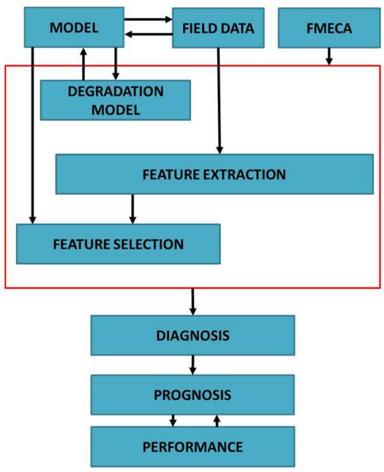

Figure 2. Methodology

A physical model of the system is then built and validated on top of those data, and used to generate increasingly degraded conditions for a few notable failure modes describing the fault evolution according to well-known, robust physical representation available in literature. This approach was deemed necessary due to the lack of meaningful historical data and follows a similar methodology applied with success in the aeronautic field (Autin, Sochelau, Dellacasa, De Martin, Jacazio, and Vachtsevanos , 2018). Since the application will not involve exchange of information with the PLC controlling and monitoring the behavior of the machine, significant effort has been reserved to the study of the feature extraction process, to ensure the capability to correctly recognize beginning and end of each working cycle and to manage data obtained during eventual interruptions of the hemming process. Feature selection has then been performed, paying attention to the behavior of the feature candidates with respect to the occurrence of possible degradations in the considered sensor(s). Hence, fault detection routines have been prepared and tested on simulated data. Finally, prognosis has been achieved by means of two different methods, which merits and issues are presented and compared. A brief overview of the employed approach is depicted in figure 2.

4. COMAU RHEVO APPLICATION

The object of this research is a COMAU RHEvo Roller Hemming head depicted in figure 3. The hemming head is the COMAU product for robotic roller hemming; attached to the last link of a 5 degrees of freedom articulated arm, it allows for a flexible and easily deployable solution. The device can be functionally described as a double spring mechanical structure that permits to use a single tool for Push and Pull applications. A rotating element is attached to this assembly and brought to contact with the metal sheets or panels to be jointed.

Figure 3. The RHEvo

Combined with the position controlled robotic arm, the internal springs allows a full force control over the tool path with a total force range of ± 2200 [N]. The RHEvo is equipped with two load cells used to monitor the exerted force, which is needed to control the system behavior during in-line service and to aid the robot calibration during the installation phase.

4.1. FMECA analysis

FMECA analysis are a necessary first step in the development of a novel PHM application. Vachtsevanos et al. (2006), suggest to study and rank the possible degradations according to four criteria: severity (S), frequency of occurrence (F), detectability (D) and Replaceability of the faulty component (R). For this application, we assigned to each factor a value ranging from 1 (less critical) to 5 (most critical). The results are hence computed as the product of each component and reported in Table 1. It emerges that RHEvo fault modes can be grossly divided into three major categories. Faults causing variations in the system stiffness, faults causing offsets in the exerted force and faults affecting the process quality but not the RHEvo performances. The first category comprehends the spring yielding and the wear of internal components, in the second we can find the wear of the roller while the remaining fault modes fall in the last class. The effects of most of the presented faults on the system performances are not expected to be significantly different from each other, since they mostly revolve around variations in the overall dynamic stiffness. At the same time, replacement of most of the eventually faulty RHEvo components demands the disassembling of the whole hemming head, with the sole exception of the faults affecting only the roller. As such, the interest in exactly classifying the issues possibly affecting the internal parts of the device is low. Priority is hence given to monitoring the overall health status of each sub-system, eventually studying dedicated prognostics routines.

Table 1. FMECA results

Failure mode	Effect	F	S	D	R	OVR
Spring yielding	Alteration of the exerted force	4	4	2	4	128
Internal components wear (bushes/pad/ finger roll)	Alteration of the exerted force	3	4	2	5	120
Bearings wear	Vibrations increase, worsening of hemming results	2	3	1	4	24
Roller wear	Variation of the exerted force	2	4	2	2	32
Roller surface degradation (dirt)	Loss of process quality	4	2	5	1	40

5. PHYSICAL MODEL

The roller hemming head is a mechanically complex device which can be in first approximation modelled as the simple two-springs, two-dampers, one mass system reported in Figure 4; x_0 represents the position of the robot attachment, x the position of the internal mechanical link and y the position of the roller. Bending effects due to the positioning of the roller are in first approximation neglected. As such, the mechanical equilibrium of the system can be simply described as a system of two equations,

$$\begin{cases} F = F_{0x} + k_0(x - x_0) + c_0(\dot{x} - \dot{x}_0) \\ m\ddot{x} + F + F_{0xy}k(x - y) + c(\dot{x} - \dot{y}) + F_{fr} = 0 \end{cases} \quad (1)$$

Where F is the measured force and F_{fr} the friction due to direct contact between mechanical components in relative motion, while F_{0x} and F_{0xy} are the values of preload for the two springs. Springs stiffness and damping coefficients are respectively k_0, k and c_0, c, while m is the translating mass of the system. The model parameters have then been tuned according to a large experimental data set. A small example is reported in figure 5, where the comparison between the simulated and the experimental values for the exerted force have been compared over a portion of a real working cycle. Please notice that the force values have been normalized with respect to a constant which will not be disclosed in this paper and that the time scale has been altered. The force sensor is modelled as a second order transfer function, with an over-imposed Gaussian noise. Although simple, this model allows to easily implement the most significant fault modes described in Section 4, which are hereby recalled and mathematically described.

5.1. Spring yielding/cracking

Yielding and eventually the inception and growth of a crack causes a loss of the spring stiffness directly proportional to the extent of the damage.

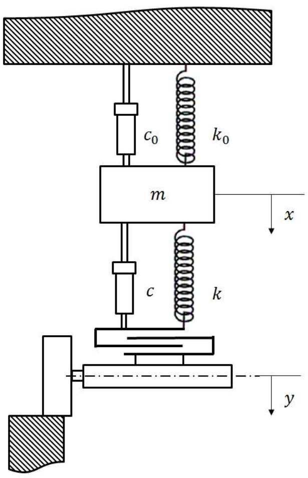

Figure 4. Dynamic model

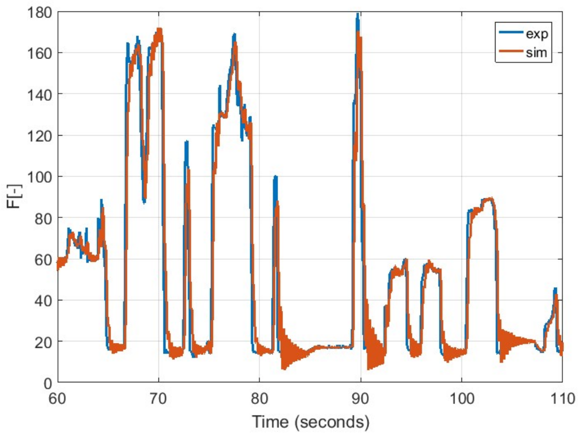

Figure 5. Model validation

Cracks originates due to fatigue associated with the repeated cycles of exerted force. The presence of a crack affects the spring stiffness by reducing the size and changing the shape of the spring section up until the final static rupture. Given the results of the model, crack growth can be estimated by incorporating the Paris' Law inside the model as done in (Autin, Sochelau, Dellacasa, De Martin, Jacazio, and Vachtsevanos , 2018).

5.2. Wear in several components

Wear of the internal components is due to the contact between parts in relative motion and can be modelled as an increasing backlash at the attachments of the two springs, which growth depends on Archard's law. Wear of the roller, on the other hand, is modelled by translating the value of y by an increasing displacement δy, function of the growth coefficient K_w as,

$$\delta y = K_w \int_t vF_c \, dt \; (1 + \alpha \sin(\omega t)) \quad (2)$$

Where v is the linear speed of the roller, F_c the contact force, ω the roller angular speed and α a parameter to simulate a non-uniform distribution of wear along the roller surface. This finds justification when considering that the material of the roller is extremely hard and chosen to survive through prolonged usage without suffering from significant damage for abrasive wear processes. However, localized abrasive or adhesive phenomenon could still occur, leading to small non-uniform deformations.

6. EXPERIMENTAL DATASET

The experimental data set is made of a large number of acquisitions of the exerted force signal captured on healthy, in-service roller-hemming machines. Signals are sampled at 25 Hz for force monitoring, hence filtering the high-frequency components. Although available inside the robot PLC, we don't have a direct access to data such as joint position, motor currents, flags for start and end of each cycle etc. An example of the force signal for one working cycle is reported in figure 6.

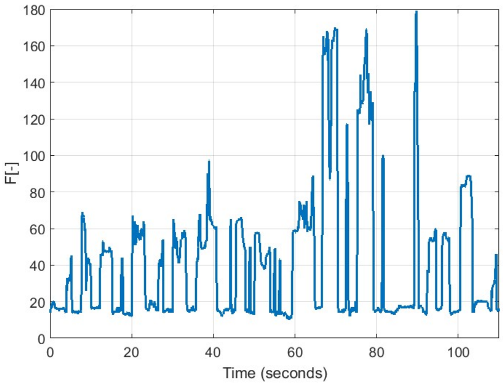

Figure 6. Data set for one working cycle

7. OPERATIONAL SCENARIO

The chosen operational scenario is that of a robotic roller-hemming machine working in an automotive assembly line. In this application, the hemming head is used to perform joints on the car doors. Operations are not completely automatized, and the robot can only work under human supervision. As such, the order through which the doors are worked is unknown a-priori and is subject to possible changes during production, while the number of operations for each car can be assumed constant. In the same way, the duration of each working cycle is not fixed and can significantly vary. One of the most common occurrences is that a working cycle might get paused due to maintenance to the robot or other parts of the assembly line or due to the mandatory breaks that the human supervisor must take during his or her shift; when this happens, the robot is paused and kept still in a neutral position, while sensors signals are still recorded. Data for healthy systems used in this paper naturally incorporates these issues, since they come from real in-service devices. Other possibly occurring disturbances, such as electrical noise, slight temperature variations and off-nominal working conditions are present as well. As stated in Section 2, the PHM system won't have access to PLC data to allow for easier implementation in already operating lines. As such, it will rely only on the signals provided by the roller-hemming head (the exerted force signal), while no information about the robot (motor currents, joints position etc.) is available. The PHM routines should be computationally cheap, so that they can be run directly on-site without resorting to external servers and without the need to connect the robotic hemming system with a central hub for data-mining. The reliance on only one sensor means that additional care has to be put in the processing of raw data and in the feature extraction, and that features must be influenced as little as possible by possible degradations of the force sensor. Given the scenario, the PHM system will rely on the feature(s) extracted from a single force signal to detect a set number of anomalous behaviors. A reasoner will hence be responsible for fault isolation and, whenever possible, fault identification. Prognosis is then performed and the plant maintenance office/system is alerted. A representative flowchart for the PHM framework is reported in figure 7.

8. DATA MINING AND FEATURE SELECTION

Feature selection is the core process of any PHM system and can be divided in two consequent off-line steps. The study of a reliable procedure to extract the feature candidates and the choice of the features from the previously obtained pool. Relying on data coming from only one signal, a critical part of this research project is the study of proper feature extraction techniques. Due to the uncertain

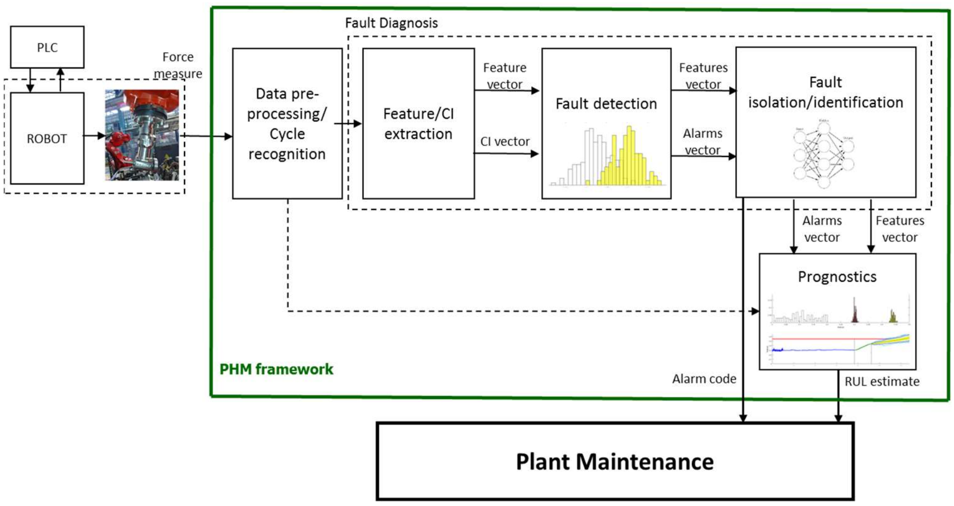

Figure 7. The PHM framework

and wide varying nature of the time scale of the problem, the number of working cycles provides a much more meaningful measure of the system usage than the number of working hours. As such, it is preferable to evaluate any feature(s) candidate over each working cycle; at the same time, it makes more physical sense to correlate the fault growth with each completed cycle rather than with respect to working hours, since the considered mechanical degradations are governed by stress cycles and are not affected by idle time. Working cycles are easily identified by flags in the PLC controlling the robot. However, lacking the access to PLC data, the PHM system must be able to autonomously identify the start and the end of each cycle. Hence, feature selection process according to state-of-the art criteria can be performed (Vachtsevanos et al, 2006). Both experimental data coming from in-service equipment and synthetic data provided by the validated simulation model have been used for these tasks. In this section we look at these problems and describe the adopted approach, detailing the peculiarities of each step and discussing their results.

8.1. Recognition of working cycles

We define as working cycle for this application the whole set of operations that the robotic roller-hemming device must perform to complete the assembly tasks for one car door. From common practice experience we can observe that

- The number of operations performed for each car remains the same (same car type)
- The order of the operations performed may change from one car to the following
- Cycles can be interrupted and reprised following specific occurrences, as shown in figure 8 (maintenance on the production line, operators breaks etc.)
- Time duration can be variable

To recognize and isolate a single working cycle, the algorithm works following an iterative procedure while,

- Searching for the cycle start
- Searching and eliminating pauses in the operations
- Searching for the end point of the cycle
- Eventually adjusting the end point position

The algorithm works having data from at least two consecutive cycles and data from at least one reference cycle. Through the use of rolling variance we can easily identify time spans in which data are characterized by low variance, such as the cycle start, the cycle end and possible breaks in between. In particular, cycle starts are identified as the first data point for which the rolling variance overcomes a fixed threshold, comparable with the measuring noise. In this way, the system searches for the first data point for which the variation of the signal behavior cannot be

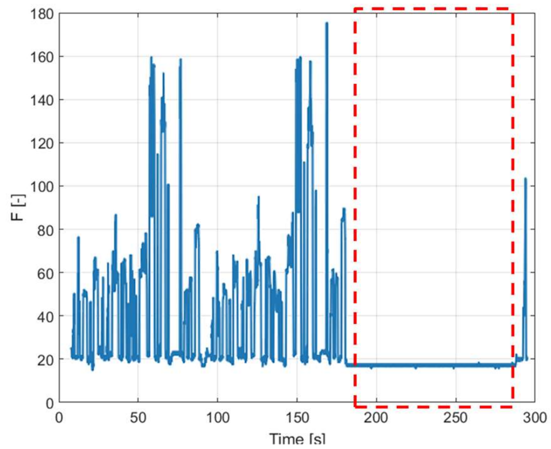

Figure 8. Example of working cycle with short pause

explained as a product of the noise affecting the measurement, hence representing a change in the machine operations. Breaks are detected whenever we have more than a fixed number of consecutive data points with low rolling variance, and those points are hence eliminated from the analysis. As shown in Fig. 8, the presence of several consecutive points with low variance is a good indication of the lack of movements of the hemming head. To detect the periodicity of the signal and hence find the point correspondent to the end of the working cycle we employed a two-steps routine. In the first step we make use of the autocorrelation function, defined as follows,

$$R(t, t+\Delta t) = \frac{E[(X_t - \mu_t)(X_{t+\Delta t} - \mu_{t+\Delta t})]}{\sigma_t \sigma_{t+\Delta t}} \tag{2}$$

Where X, μ and σ are respectively the realization of the process, its mean value and its standard deviation at times t and $t+\Delta t$. Through autocorrelation we can automatically isolate a certain number of candidates Δt corresponding to a potential periodicity by searching for peaks as shown in figure 9.

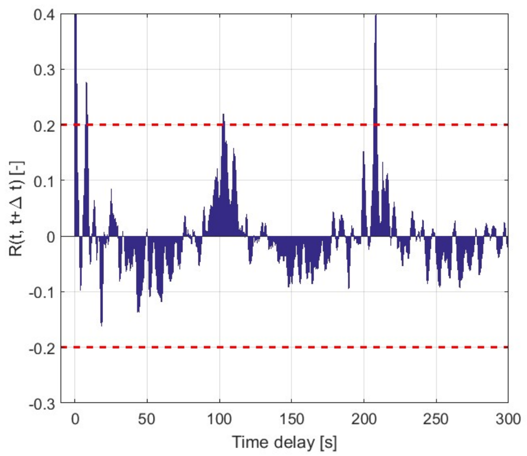

Figure 9. Autocorrelation over a 300 s signal

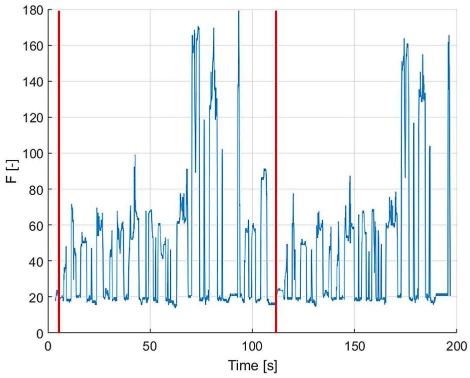

Figure 10. Cycle recognition - results

Hence the search for the right value can be further refined by looking in this pool of time instants for the one associated with the moving average value of the force signal closer to that of the start. It is in fact expected that the force measured in correspondence of the end cycle should be extremely close, if not equal to, the value measured at the start of each cycle; in those conditions the hemming head is not pressed against the metal sheets, and the value of the measured force is theoretically equal to the preload of the springs, minus the effect of friction components. The algorithm has been tested on both healthy experimental data and faulty conditions simulations, providing reliable results in both cases. In some heavily disturbed instances, the algorithm can make mistakes by overestimating the time instant of the cycle end. To avoid this occurrence, a subroutine is called that perform a similarity check with a pool of reference signals. If the similarity is not sufficient, the time instant associated with the end of the cycle is rolled back until the similarity does not increase any more in a significant way. A way to fasten this process is to search for time instant associated with low values of the rolling variance, since the signal at the end of the working cycle should feature low variations. An example of the results of the cycle recognition process is reported in figure 10.

8.2. Feature selection

The feature selection process is probably the most critical step in the definition of a PHM system. During this stage of the development, we search for indexes which must be representative of the health status of the system; aiming at prognosis means that these indexes must present a high correlation with the growth of the faults associated with each of them and possibly a low correlation with every other faults. The first step in the feature selection process is to obtain data in degraded conditions and compare these data with those sampled for healthy (also called baseline) conditions. Healthy conditions data have been obtained directly from the field, while degraded conditions have been simulated by means of the validated model described in Section 4. Relying on only one sensor, we have to take into account the possibility of a degradation of its characteristics in time and eventually its failure; as such, the feature(s) that we might employ in the PHM system should be ideally unaffected by the most common types of faults affecting the sensor and its signal, such as zero-errors, increased noisiness of the output, variations from the sensor's nominal measuring characteristics. The effects of these fault types have been introduced by corrupting both the healthy and the degraded data sets by over imposing a growing noise and a growing negative off-set on the force signal. Example of the considered conditions of the force signal (healthy, faulty, noise increase and off-set presence) are reported in figure 11. The faulty signal is related to wear in one of the RHEvo components. However, the effects of the considered degradations of the RHEvo on the physics of the system are unfortunately quite similar, since they all revolve around a loss of the rolling head stiffness. Given the reliance on only one sensor, difficulties are expected in obtaining a complete disambiguation of the possible faults. Several candidates, such as the mean value of the force signal over each cycle, the variance, peaks amplitudes, have been at first analyzed without success. Hence we adopted a different approach based on the use of a reference signal for healthy condition and the Kullback-Leibler divergence. The Kullback-Leibler divergence is a distribution-wise asymmetric measure of divergence between two probability distributions P and Q (Kullback, & Leibler, 1951); its value can range from 0 (similar behavior of the distributions) to 1 (expectations given distribution approach zero). For discrete probability distribution, it can be expressed as,

$$D_{KL}(P||Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)} \tag{3}$$

In our study, data distributions correspondent to each working cycle are compared with that of a reference signal for healthy conditions. We evaluated the behavior of this feature with respect to the previously introduced four conditions (healthy, failed, noisy sensor, offset in the sensor), and reported an example of the results in figure 12.

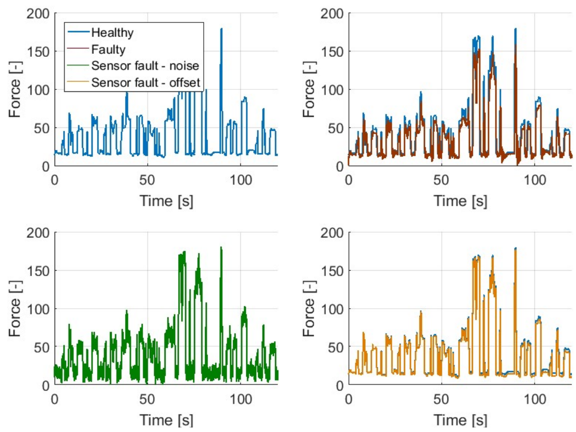

Figure 11. Analyzed signals types

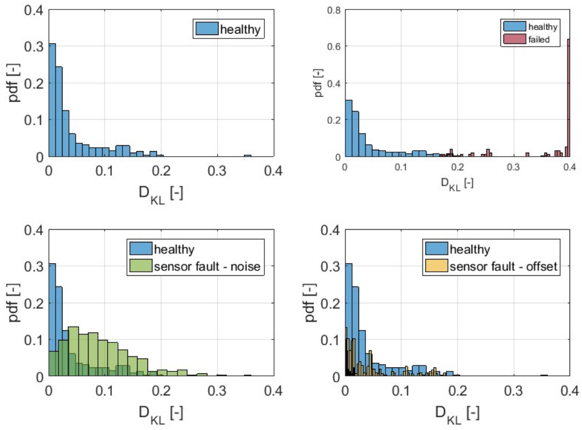

Figure 12. Feature behavior in different conditions

We can observe a good separation for the failed conditions, associated with advanced wear conditions able to severely deteriorate the hemming operations. We can similarly observe that the eventual occurrence of issues in the force sensor is not able to trigger false alarms on the feature, since its values remains inside the healthy baseline. The chosen feature is moreover strongly correlated with the fault growth process, presenting a correlation factor equal to 0.899 and can be as such considered suitable for prognosis. Interestingly, some of the feature candidates discarded for the fault detection of the selected RHEvo degradations can be used to track the evolution in time of possible issues in the force sensor; the average value of the signal over a cycle can be used to detect the occurrence of off-set errors, while rolling variance can observe the increase of the signal noise.

9. FAULT DETECTION

Fault detection is the set of operations performed by the PHM algorithm to detect the anomalous behavior of the selected features. In this case, fault detection is performed by means of a purely data driven approach, where current data distribution obtained over moving windows are compared with a baseline representative of the healthy conditions. In this case, a fault is declared when the 80% of the moving distribution overcomes the 80th percentile of the baseline for healthy conditions, hence accepting 20% probability of false positive and 20% probability of false negative. These percentages are considered acceptable since the selected fault modes (wear and spring yielding) are expected to develop slowly according to extensive industrial experience. We run tests on a few degradation patterns, focusing on wear processes, obtaining an average fault size at detection of 42% of the critical size. Although high, this value does not represent an issue of the system performances, since the fault evolves slowly during operations. An example of fault detection performed over synthetic data obtained by enabling wear growth during simulations is reported in figure 13. The two most significant distributions (baseline and at detection) are highlighted along the behavior of the confidence associated with the fault declaration and the fault detection flag. The fault ratio is thought as the ratio between the wear extension and its critical value, associated with significant performance degradation of the hemming procedure.

We hence performed a few tests running the fault detection algorithm over data suddenly affected by an increased variance in their distribution due to a sudden loss of accuracy of the sensor output. We considered an increase of the signal variance up to five times its nominal value, possibly caused by issues in the signal conditioning module.

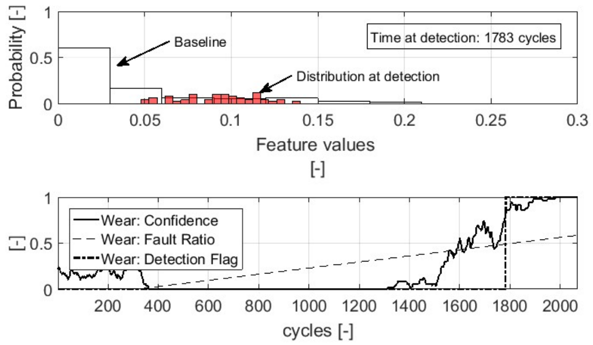

Figure 13. Fault detection

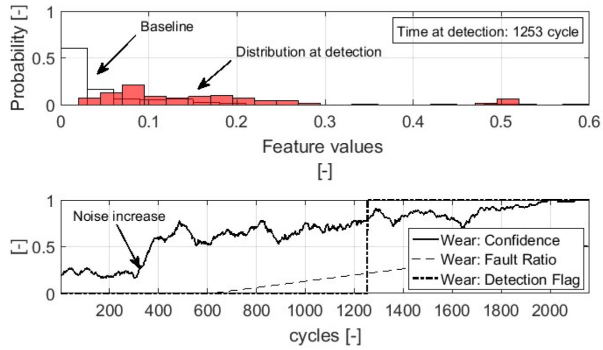

Figure 14. Fault detection with noisy sensor data

Result of one of the trials is reported in figure 14, where we can observe that the sudden occurrence of the increased variance issue is not able to trigger a false alarms, even if it raises the probability associated with the fault declaration, which is hence provided correctly once that the wear begins to develop. The presence of zero-errors of the sensor, covered by the increased off-set case, does not provide instead a significant contribution to the variation of the stochastic behavior of the selected feature; as such, no further tests on this particular issues have been performed.

10. FAILURE PROGNOSIS

Two different techniques have been considered to perform the failure prognosis; an advanced declination of particle filtering and a Long-Short Term Memory (LSTM) artificial neural network. While particle filters are the current state-of-the-art for long term prediction within the PHM community, LSTM tend to exhibit better performances for failure modes affected by abrupt ends, which can be the case for several mechanical degradations, such as crack progression (Huang, Khorasgani, Gupta, Faraht, & Zheng, 2018), hence resulting of particular interest in the context of a purely mechanical device. The theoretical aspects of these methods are hereby briefly recalled and their results compared.

10.1. LSTM for failure prognosis

Long Short Term Memory networks, usually just called LSTMs, are a special kind of Recursive Neural Networks (RNNs), capable of learning long-term dependencies. Firstly introduced by Hochreiter and Schmidhuber (1997), and were further refined and popularized in several research fields (Gers, Schmidhuber, & Cummings, 2000), (Courville, Goodfellow, & Bengio, 2016). LSTMs are explicitly designed to avoid the long-term dependency problem by means of a specific four layer structure.

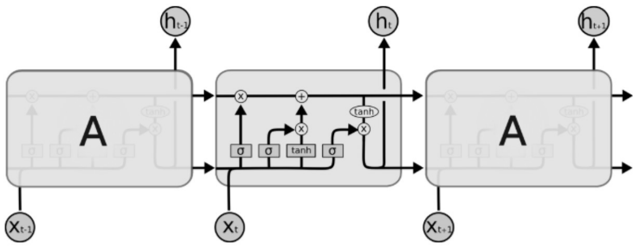

Figure 15. Repeating modules in LSTMs

All RNNs type have the form of a chain of repeating modules of neural network, where each module feeds off the results of the previous one and provides input to the following. In standard RNNs, this repeating module usually have a very simple structure, i.e. a single hyperbolic tangent layer. The four layers of the LSTMs are instead more complex, as shown in figure 15, each of them performing a different function. Each module communicates with the surrounding ones through two ports: a “cell state” port, providing the state of the system as computed in the previous module, and a “previous output” port, providing the output of the previous module. We can observe a “forget” layer, defining which information coming from the previous modules are to be used, an “input” layer defining which states will be updated, a “candidate” layer, which creates a vector of new state candidates and an “output” layer which provides the output of the single module. LSTMs have found applications within the PHM community since they are relatively easy to implement and are able to provide better results than other traditional techniques in system possibly subjected to abrupt failures. In this case we trained the network to output, given the feature value, the distance between the moving distribution of the selected feature and the baseline, normalized with respect to the failure conditions, with the aim of predicting the RUL with respect to a threshold equal to 0.9. The training phase is critical for the long-term prediction performance of the LSTMs, since the network capability to accurately forecast the future behavior of the degradation pattern is heavily influenced by the employed batch size and the used number of epochs. A comparative study of the LSTM behavior for different combinations of epochs (between 50 and 150) and batch size (between 20% and 85% of the training set) was performed over a few degradation patterns. Results shows that the LSTMs need to work with a batch size covering a significant portion of the training set (above 75%) to provide acceptable results, while in the case under analysis performance increases were not appreciable increasing the epochs number above 150.

10.2. Particle filtering for failure prognosis

Particle filters are a class of Bayesian estimators for non-linear, non-Gaussian stochastic problems which can be used to approximate the hidden state(s) of the system (Arumpalam, Maskell, Gordon, & Klapp, 2017) and predict their evolution in time (Orchard, 2007). The filter estimates the current state(s) of the system by performing two sequential steps, prediction and filtering, combining real-time observations of the selected feature y_t with a state model and a process model,

$$\begin{cases} x(t) = f_t\big(x(t-1), \omega(t)\big) \\ y(t) = h_t\big(x(t), \nu(t)\big) \end{cases} \tag{4}$$

Where x is the hidden state, ω and ν are the measure and process noise, while f_t and h_t are non-linear mappings. During prediction, the filter makes use of both the knowledge of the previous state estimate (or its initialization value) $p(x_{0:t-1}|y_{0:t-1})$ and the process model to estimate the next time instant,

$$p(x_{0:t}|y_{1:t-1}) = \int p(x_t|y_{t-1})p(x_{0:t-1}|y_{1:t-1})dx_{0:t-1} \tag{5}$$

Since an analytical solution for this expression does not exist in most of the cases, sequential Monte Carlo techniques, or particle filters, can be used (Roemer, Byington, Kacprszyinski, Vachtsevanos, & Goebel, 2011). Particle filters approximate the state pdf using particles associated with discrete probability masses, also called weights $\widetilde{w}$,

$$p(x_t|y_{1:t}) \approx \widetilde{w}_t\big(x_{0:t}^i\big)\delta\big(x_{0:t} - x_{0:t}^i\big)dx_{0:t-1} \tag{6}$$

Where $x_{0:t}^i$ are the state trajectories. During the second stage of the particle filters, the “filtering” step, the algorithm updates the weights using one of several resampling techniques. Although not the best in performance, Sequential Importance Re-sampling (SIR) provides a good compromise between end results and required computational effort (Vu, Do, Jha, Theilliol, & Peysson, 2018) and has been as such chosen inside this framework.

Long-term predictions are used to estimate the probability of failure in a system given a hazard zone that is defined via a probability density function with lower and upper bounds for the domain of the random variable, denoted as H_{lb} and H_{up}, respectively. Traditionally, the probability of failure at any future time instant is estimated by combining both the weights of predicted trajectories and specifications for the hazard zone through the application of the Law of Total Probabilities as,

$$\hat{p}_{RUL,t} = \sum_{i=1}^{N} p\left(F|X = \hat{x}^{i}_{tRUL}, H_{lb}, H_{up}\right) \tag{7}$$

Acuña and Orchard (2017, 2018) recently underlined that this expression is mathematically incorrect, since it has been misinterpreted as a Cumulative Mass Function. In (Acuña and Orchard, 2017) the expression to compute the Probability Mass Function for the occurrence of a failure $p(F_k)$ at time instant k with prediction started at time instant k_p has been redefined as,

$$\mathcal{P}(F_k) = \mathcal{P}\left(F_k \middle| H_{k_p:k-1}\right)\mathcal{P}\left(H_{k_p:k-1}\right), \quad \forall k > k_p \tag{8}$$

Where $\mathcal{P}\left(F_k \middle| H_{k_p:k-1}\right)$ is the probability of the system to encounter a failure at time instant k under the hypothesis that the system remains healthy up to time instant *k-1*, equivalent to the traditionally adopted expression (7). $\mathcal{P}\left(H_{k_p:k-1}\right)$ is hence the probability that the system has not failed up to time instant *k-1*. The use of this expression allows to easily compute the Acuña's discrete-time risk-of-Failure (Acuña and Orchard, 2018) associated with the RUL estimate for time instant k given the start of the prediction at time instant k_p as,

$$\mathcal{R}(k|y_{1:k}) = \sum_{i=k_p+1}^{k} \mathcal{P}\left(F_i | y_{1:k_p}\right) \tag{9}$$

The degradation model used inside the framework is derived through extensive use of symbolic regression techniques to search for expressions to find the model that best fits the available datasets paying attention to both accuracy and simplicity. We exploit the tool provided by Schmidt and Lipson (2009) to represent the fault size as a function of the number of working cycles by assembling the model by combining basic blocks such as mathematical operators, analytical functions, constants and state variables. The particle filtering framework under analysis makes use of a few additional techniques; the degradation model is continuously and automatically tuned as data streams in, while an outer correction loop is used to fine tune the process and the measure noise by combining the result of short-term predictions obtained for previous time instances with the state estimates (De Martin, Jacazio, & Sorli, 2018). The particle filtering framework is hence applied to the simulated degradation data, making use of a moving window to represent the feature pdf between time instants *t* and *t-M*, where *M* is the moving window size equal to 100 cycles. The algorithm runs in background and long-term prognosis is recalled periodically after the fault detection. Failure is declared in correspondence of a fixed threshold, obtained as the 50th percentile of the feature values for which the degradation of the simulated roller-hemming performance may begin to cause issues to the production process. Since the tuning of the cycle-dependent degradation model is performed automatically within the algorithm, no further action have been required. An example of the particle filter behavior and its application to failure prognosis is reported in Fig. 16.

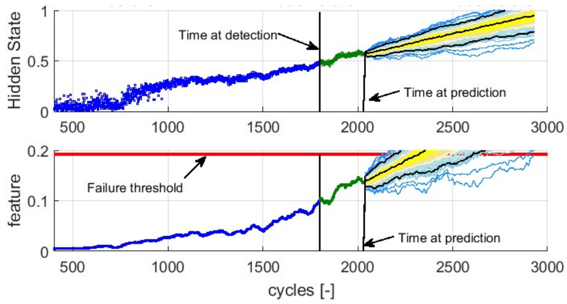

Figure 16. Failure prognosis through particle filtering

10.3. Performance metrics

Traditional performance metrics for PHM systems revolve around the idea of measuring the accuracy of the algorithms in forecasting the time of failure (or End of Life, EoL) of a specific case or set of cases. Most of these metrics can be directly inferred from the α–λ analysis proposed in (Saxena Celaya, Balaban, Goebel, Saha, & Scwabacher, 2008). Between these metrics we can find the Prognostic Horizon (PH), defined as the time span during which the RUL estimate falls within an accuracy band usually defined as $\pm 20\%$ of its real value. The Relative Accuracy (RA), defined in equation (10) and its average value over the total number of predictions.

$$RA = 1 - \frac{\left|\left(\mathrm{EoL} - t_p\right) - \mathrm{RUL}\right|}{\left(\mathrm{EoL} - t_p\right)} \tag{10}$$

The Cumulative Relative Accuracy (CRA) is instead a used to assess the behavior of the prognosis towards the later stages of the system useful life, by using a weighted average of RA where the weights are skewed to favor higher values of λ, often being the values of λ themselves.

$$CRA = \frac{1}{\sum_i \lambda_i}\sum_i \lambda_i \mathrm{RA}_i \tag{11}$$

10.4. Results

We compared the accuracy of the two algorithms in figure 17, where the particle filter is described by the trajectory of the 50th percentile of the particles population. The particle filtering algorithm makes use of 200 particles; the LSTM is trained over a window of 650 cycles using 150 epochs. The LSTM requires a long window of training data before starting to provide accurate results, but provides higher accuracy in predicting the later stages of the fault growth. It is of interest to notice that the traditional advised maintenance indication is in truth associated with a high value of Acuña's risk-of-Failure. The average RA for the traditionally defined advised maintenance through particle filter (50th percentile of the projected particles) is equal to 81.28%, while that of the LSTM is much lower. However, if we restraint the RA calculation to the last 150 cycles, the LSTM is actually able to reach an average value of RA equal to 96.6%. The CRA for the particle filtering algorithm is equal to 85.52%, while the LSTM provides a 94.3% result. Particle filtering techniques benefits from a higher value of the Prognostic Horizon, in this case around 600 cycles (excluding some minor over estimates in few instances), while the LSTM provides a 110 cycles PH. LSTM suffers from higher computational costs, since the training phase, absent from the particle filter, requires an average of 178 seconds to complete, while the prediction phase lasts an average of 1.05 seconds. A single forecast of the particle filter needs an average of 1.27 seconds to complete, while the prediction/filtering cycle used to estimate the system states is almost real-time, requiring less than 0.016 seconds in average to complete. The trial configuration is an Intel Core i7-4770 @ 3.40 Ghz with 8.00 GB of RAM. Finally, an obvious benefit of the Particle Filtering method is the capability to provide an estimate of the RUL pdf and of the risk associated with taking decisions based on its prediction, that could be of critical importance for the objective of zero unexpected failures (and hence unexpected line downtimes) defined by the WCM. Both techniques seems viable for the application in "guessing" the failure time; the particle filter algorithm, being less computationally expensive appears as a preferable choice for a localized PHM solution, where a small computer

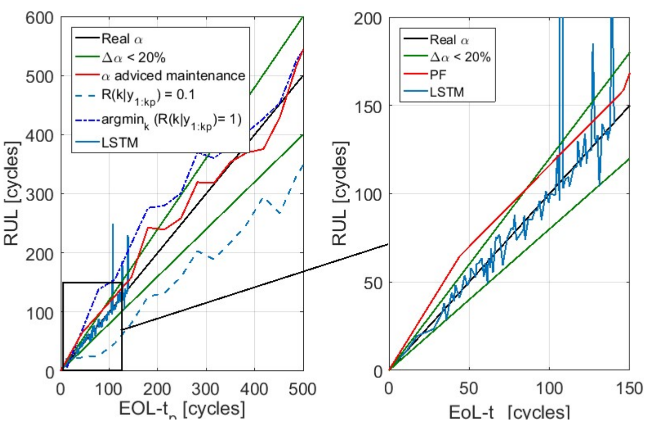

Figure 17. α−λ analysis

attached to the machine or used by maintenance operator cha download data when needed and quickly provide an estimate of the RUL. On the other hand, LSTM provide interesting results when nearing the end-of-life conditions and can support the results of the Particle Filter algorithm; due to the increased computational load, this solution appears to be better-suited for a server-based solution, where data coming from the machine are manually or automatically downloaded and sent to a central hub dedicated to PHM data elaboration for the plant.

11. CONCLUSIONS

The development of PHM techniques for robotic roller hemming applications is an interesting challenge for the PHM community and potentially a significant improvement for production plant management for the industry of automotive assembly. Although the economic benefit is tough to gauge due to the difficulty in accessing companies production data, the introduction of an effective PHM system for roller hemming would significantly affect the plant productivity and the related costs by reducing the occurrence of unexpected downtimes and a more cost-effective planning of the maintenance operations. We describe in this paper the early results of an on-going activity in this field. The proposed PHM framework is purely data driven and designed to work without accessing the PLC data of the robot, by correctly identifying the working cycles and the required features only through post-processing of raw sensor data. The chosen feature is robust to sensor issues and present a good correlation with the fault under examination. A simple data driven technique is used to perform the fault detection, while two possible algorithms, based on the use of particle filtering and of LSTMs are proposed and compared for the RUL estimate; both algorithm are deemed as suitable for the selected case, while their possible implementation inside the plant maintenance system is discussed. Further investigations are of course needed before closing the first phase of the development cycle of the PHM system; for once, it is critical to assess the effects of possible faults inside the robotic arm attached to the roller hemming head on the selected features. In the same way, the experimental data on faulty roller hemming heads needed to validate the model in degraded conditions are still missing, although the relative simplicity of the device design provides a good confidence to the results of the available model. Further studies are also needed to better understand which implementation of the PHM technologies (on-board, localized, server based) is best for the application.

ACKNOWLEDGEMENT

Authors would like to thank Nunzio Magnano from Comau S.p.A. for its support during the development of the study.

REFERENCES

Acuña, D.E., & Orchard, M. (2017) Particle-filtering based failure prognosis via sigma-points: application to Lithium-Ion battery State-of-Charge monitoring. *Mechanical Systems and Signal Processing*, vol. 85, pp.827-848.

Acuña, D.E., & Orchard, M. (2018) A theoretically rigorous approach to failure prognosis. *Annual conference of the PHM society*. September 24-27, Philadelphia, PA, USA.

Alzahrani, A.M., Liu, R., & Kolodziej, J.R. (2018) Acoustic assessment of an End Mill for Analysis of tool wear. *Annual conference of the PHM society*. September 24-27, Philadelphia, PA, USA.

Autin, S., Socheleau, J., Dellacasa, A., De Martin, A., Jacazio, G., & Vachtsevanos (2018) Feasibility study of a PHM system for electro-hydraulic servo-actuators for primary flight controls. *Annual Conference of the PHM society*. September 24-27, Philadelphia, PA, USA.

Arumpalam, S., Maskell, S., Gordon. S., & Clapp, N.J. (2002) A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking. *IEEE Trans. On Signal Processing*, vol. 50, no. 2, pp.174-188.

Bevilacqua, M. and Braglia, M. (2007) The analytic hierarchy process applied to maintenance strategy selection. *Reliability Engineering and System Safety*, vol. 70, no. 1, pp. 71-83.

Courville, A., Goodfellow, I., & Bengio, Y. (2016). Deep Learning. The MIT Press.

De Martin, A., Jacazio, G., & Sorli, M. (2018). Enhanced Particle Filter framework for improved prognosis of electro-mechanical flight controls actuators. *Proceedings of the 4th European conference of the PHM society*. July 3-6, Utrecht, NL.

De Ruyte, A.S., Cardew-Hall, M.J., & Hodgson, P.D. (2002) Estimating quality costs in an automotive stamping plant through the use of simulation. *International Journal of Production Research*, vol. 40, no. 15, pp.3855-3848.

Drossel, W.-G., Pfeifer, M., Findeisen, M., Rössinger, M., Eckert, A., & Barth, D. (2014) The influence of the robot's stiffness on roller hemming processes. *ISR/Robotik 2014; 41st International Symposium on Robotics*, Munich, Germany, 2014, pp. 1-8.

Esquivel, E., Carbone, G., Ceccarelli, M., & Jáuregui (2016) Requirements and constraints for a robotized roll hemming solution. *Advances in Robot Design and Intelligent Control. RAAD 2016. Advances in Intelligent Systems and Computing*, vol. 540. Springer, Cham.

FCA (2016) WCM in FCA. Retrieved from https://www.wcm.fcagroup.com/en-us/wcm_at_fca

Gers, F.A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: continual prediction with LSTM. *IET Conference Proceedings of 9th International Conference on Artificial Neural Networks: ICANN '99*, p. 850-855, DOI: 10.1049/cp:19991218.

Hochreiter, S. & Schmidhuber, Jü. (1997). Long short-term memory. *Neural computation*, vol. 9, pp. 1735-1780.

Huang, W., Khorasgani, H., Gupta, C., Faraht, A., & Zheng, S. (2018) Remaining Useful Life Estimation of Systems with Abrupt Failures. *Annual conference of the PHM society*. September 24-27, Philadelphia, PA, USA.

Kullback, S., & Leibler, R.A (1951) On information and sufficiency. Ann. Math. Stat. vol. 22, pp. 79–86.

Le Maoût, N., Thuillier, S., & Manach, P.Y. (2010) Drawing, flanging and hemming of metallic thin sheets: a multi-step process. *Material & Design*, vol. 31, no. 6, pp. 2725-2736.

Li, Z., Wu, D. (2018) A Data-driven approach to material removal rate prediction in chemical mechanical polishing. *Annual conference of the PHM society*. September 24-27, Philadelphia, PA, USA.

Longo, N., Serpi, V., Jacazio, G., & Sorli, M. (2018) Model-based predictive maintenance tecniques applied to automotive industry. *Proceedings of the 4th European conference of the PHM society*. July 3-6, Utrecht, NL.

Montalbo, T., Roth, R., & Kirchain, R. (2008) Modelling Costs and Fuel Economy Benefits of Lightweighting Vehicle Closure Panels. *SAE Technical Paper 2008-01-0370. 2008 World Congress*. April 14-17, Detroit, MI, USA.

Orchard, M. (2007) *A particle filtering-based framework for on-line fault diagnosis and failure prognosis*, Doctoral dissertation. School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

Pascual, R., Meruane. V., & Rey, P.A. (2008) On the effect of downtime costs and budget constraint on preventive and replacement policies. *Reliability and System Safety*, vol. 93, pp. 144-151.

Roemer, M., Byington, C., Kacprszynksi, G., Vachtsevanos, G., & Goebel., K. (2011) Prognostics, in *System Health Management with Aerospace Applications*, Wiley, pp. 281-295.

Roy, R., Souchoroukov, P., & Shehab, E. (2011) Detailed cost estimating in the automotive industry: Data and information requirements. *International Journal of Production Economics*, vol. 133, pp. 694-707.

Thomas, D. (2018) Advanced maintenance in manufacturing: costs and benefits. *Annual conference of the PHM society*. September 24-27, Philadelphia, PA, USA.

Saboori, B., Saboori, B., Carlson, J.S., & Söderberg, R. (2009) Introducing fast robot roller hemming process in automotive industry. *World Academy of Science, Engineering and Technology*, vol. 34, pp. 503-506.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., & Scwabacher, M. (2008) Metrics for evaluating performance of prognostic techniques. *International*

Conference on Prognostics and Health Management, October 6-8, Denver, CO, USA.

Singh, K., Selvanathan, B, Zope, K., Nistala, & S.H., Runkana, V. (2018). Concurrent estimation of Remaining Useful Life for multiple faults in an Ion Etch Mill: a data-driven approach. *Annual conference of the PHM society*. September 24-27, Philadelphia, PA, USA.

Tabikh, M. (2014) Downtime cost and reduction analysis; survey results. Master Thesis. Mälardalen University, Sweden.

Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering system*. Hoboken, NJ: John Wiley & Sons, Inc

Vogl, G., Weiss, B., Helu, M. (2016) A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*.

Vu, H.C., Do, P., Jha, M.S., Theilliol, D., & Peysson, F. (2018) On the use of particle filters for prognostics in industrial applications. *Proceedings of the 4th European conference of the PHM society*. July 3-6, Utrecht, NL.

BIOGRAPHIES

Luca Actis Grosso graduated in Mathematical Engineering from Politecnico di Torino in 2016 and further refined his studies through the third-level master in Industrial Automation held by the same university. He is currently employed as data scientist in Comau S.p.A., where he work on the development of machine learning techniques for a wide variety of industrial applications, currently focusing on prognostics for robotic devices.

Andrea De Martin obtained his master degree in Mechanical Engineering at Politecnico di Torino, Italy, in 2013 and has since been member of the Mechatronics and Servosystems research group in the same university as a PhD student and hence post-doc fellow. His main research interests are in the area of Prognostics and Health Management for industrial and aerospace applications, focusing on the development of new PHM schemes for flight control systems.

Giovanni Jacazio has been full professor of Applied Mechanics and lecturer of control systems at Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy, until 2015, now retired. He is lecturing courses on fly-by-wire flight control systems and PHM in the Doctoral School of Politecnico di Torino and consulting for UTAS. His main research activities are in the areas of flight control systems for aerospace applications and of prognostics and health management. He has been Research Leader for several European and national research programs, Coordinator of research and consulting activities for several engineering industries. He is member of the SAE A-6 committee on Aerospace Actuation Control and Fluid Power Systems, and fellow of the Prognostics and Health Management Society.

Massimo Sorli born in Sanremo (Italy) in 1956, has been Full Professor of Applied Mechanics in the Politecnico di Torino, Italy, since 2000. He graduated in Mechanical Engineering at the Politecnico di Torino in 1981 and received his PhD in Applied Mechanics in 1987 at the same university. Since October 2007 until December 2011 he has been Director of the Department of Mechanics, from January 2012 until October 2015 he has been Director of the Department of Mechanical and Aerospace Engineering (DIMEAS), both at Politecnico di Torino. His scientific activity is mainly focused in the area of mechatronics, mechanical and fluid (pneumatic, hydraulic and electromechanical) servo-systems for automotive and aeronautical applications, innovative devices in flight control systems, prognostics of servo-actuators, Rendez-Vous & Docking Systems for spacecraft. He has been author or coauthor of more than 230 papers.