Machine Learning for Predictive Maintenance: a cost-oriented model for implementation.

Eleonora Florian[ab], Fabio Sgarbossa[c], Ilenia Zennaro[a],
[a]*Department of Management and Engineering,*
University of Padua, Stradella San Nicola 3, 36100, Vicenza, Italy,
eleonora.florian.1@phd.unipd.it, ilenia.zennaro@unipd.it
[b]*Presenting author*
[c]*Department of Mechanical and Industrial Engineering,*
NTNU, S.P. Andersens vei 3, 7031, Trondheim, Norway,fabio.sgarbossa@ntnu.no

Abstract

Predictive Maintenance (PdM) is a condition-based maintenance policy that carries out maintenance action when needed, avoiding unnecessary preventive action or failure. Machine learning (ML), in the form of advanced monitoring and diagnosis technologies, has become increasingly attractive. Implementing PdM through ML is a difficult and expensive process, especially for those companies which often lack the necessary skills and the financial and labour resources.

Thus, a cost-oriented analysis is required to define the most suitable maintenance policy and quantify the achievable saving. Implementing PdM with ML techniques involves the costs of investment in IT technologies, in addition to those incurred in traditional maintenance, however, no previous works consider ML implementing costs in economic evaluation of PdM. ML performance is evaluated in terms of the faults intercepted, amount of unexploited lifetime and frequency of unexpected breaks.

This paper aims to provide a useful study on data availability, system reliability and costs for companies which are interested in implementing PdM policy through ML. This paper compares preventive maintenance and failure costs with the PdM costs related to the ML.

The impact of investment costs on the ML model performance is investigated and it is integrated in the cost evaluation of PdM policy. Considering PdM costs and ML performance, useful graph about the convenient zone of applying this policy are presented, evaluating different reliability parameters. Finally, the cost model allows to evaluate the desired level of performance of the ML and consequently to evaluate PdM investment costs.

Keywords: Predictive maintenance, Condition-based maintenance, Machine learning, Maintenance costs, Decision-making systems.

1. Introduction

Production systems are affected by degradations and failures caused by operational and environmental conditions. Maintenance policies aims to guarantee the availability of production systems and to guarantee the targeted throughput, with consequently profitability. However, maintenance represents a considerable cost in terms of resources, time, tools and spare parts. In many cases maintenance costs can vary from 15% to 70% of the total production cost, or even exceed the annual net profit (Madu, 2000). Preventive Maintenance (PvM) is the most common policy applied to guarantee the availability and efficiency of manufacturing systems; this policy is based on maintenance activities that are carried out based on the estimation of mean time between failure and failure rate (Cocconcelli *et al.*, 2018). PvM costs and failure costs need to be balanced to find the optimal replacement interval, as the one that minimises the combination of the two costs. This policy is defined as a time-based maintenance as maintenance actions are driven by the replacement interval (Waeyenbergh and Pintelon, 2002). It is a very effective and diffused policy, but it doesn't maximise the availability of production systems, as failures might happened and there could be extra costs due to unnecessary maintenance actions (Susto *et al.*, 2013). Moreover, when data are not available or complete, there could be extra costs due to wrong parameters' estimations (Sgarbossa *et al.*, 2019). On the other hand, in the ideal condition of complete and available data, maintenance parameters are well estimated, but there still be costs due to failures.

Nowadays global market requires an increasing availability of production systems, that are limited by fixed financial budgets; in this context an optimisation of maintenance policy is needed (De Carlo and Arleo, 2013). Aiming to minimise maintenance costs more and more and to increase manufacturing systems' availability, Predictive Maintenance policy (PdM) might be the suitable. It allows to maintain availability and reliability of components as long as they are able to work through condition monitoring, and to intervene if it is necessary, i.e. when the breakdown is approaching. In this way, PdM consents to obtain a saving in costs related to spare parts consumption, production time optimisation and maintenance activities scheduling (Florian *et al*, 2019). To implement PdM as first it is necessary to select the machine or facility of interest and to define the maintenance problem; then it is necessary to define the operational model and to analyse the failure mode for establishing the goal of predictive model (Jardine, Lin and Banjevic, 2006; Accorsi *et al.*, 2017). Afterwards there it is necessary to manipulate the data, i.e. data selection, analysis, processing, modelling and evaluation for testing the feasibility of the model. Finally, there is the implementation of a decision support maintenance system that derives from it (Bousdekis *et al.*, 2018).

In last years, the progress in information technology (IT) has encouraged the spread of real-time control systems in manufacturing companies, increasing the amount of available data about the state of monitored machines and components. Moreover, technological development introduced tools able to collect and analyse big data, and to supply decision support capabilities for large data sets of time series data (Prajapati, Bechtel and Ganesan, 2012). Pozzi and Strozzi 2018 presents the recent revolution of manufacturing systems due to digitalisation of processes and machines. In this context, traditional and smart systems co-habitat in the same industrial reality. The issue is to establish if it is convenient to install sensors and monitoring systems on the first ones, considering the trade-off between costs and benefits. In fact, there are several factors that need to be considered for PdM implementation, first of all the need of high investments in IT architecture and in intelligent systems that should support operators in maintenance decisions (Barraza-Barraza, Limón-Robles and Beruvides, 2014).

Machine Learning (ML) techniques could favour the deployment of PdM. In fact recently the interest of research and industrial community around ML has deeply increased; it refers to a set of algorithms for analysing and process data for clusterisation, classification or prediction purpose (Hofmann, Schölkopf and Smola, 2008). Sala et al. 2018 proposed a framework to select the most suitable ML algorithm based on input data, first and second layers and response type.

Supposing the ML algorithm is selected and developed to monitor a specific production system, this paper aims to investigate when PdM implementation with a data-driven strategy is convenient in terms of maintenance costs reduction and production systems availability increasing. It presents a new analysis that compares costs and benefits of PdM compared to PvM, considering not only maintenance direct costs but also investments costs due to strategies' implementation, supposing that reliability parameters are known. Moreover, it is investigated where it is convenient to install sensors and sophisticated monitoring systems, i.e. it is presented an analysis that carries out maintenance benefits in relation to the investment costs of these systems and the accuracy of monitoring data. Given reliability parameters, PvM and PdM investment costs, preventive and failure costs, the two policies are compared to carried out a parametrical analysis that highlight the convenience of PvM or PdM in order to minimise total maintenance costs. Finally, the cost analysis allows to evaluate the desired level of performance of the ML and consequently to evaluate PdM costs in relation to it. The paper is divided in five sections: after this introduction, the second section presents a literature review of existing models of PvM and PdM costs and about ML algorithm implementation for PdM. Subsequently section 3 presents the model and costs definition, while section 4 presents the parametrical analysis and some graphical representation of the results. Finally, section 5, is about conclusions and furthers research.

2. Literature review

PvM approach is the most common policy in maintenance management to avoid failures, as it provides maintenance actions based on a schedule. Anyway, this approach is not optimal in terms of costs as it does not consider the dynamic state of the production equipment; often it leads to over or less maintenance that causes unnecessary replacements or unproductivity in the manufacturing process (Ding and Kamaruddin, 2014). PdM concept was introduced in 1940s by the Rio Grande Railway Company (Prajapati, Bechtel and Ganesan, 2012). Its goal is to intercept in advance the symptoms of anomaly behaviours of a physical production system for developing a just-in-time maintenance action, in which availability, quality and safety of the equipment are preserved, failure risks decrease, and costs of unnecessary time-based maintenance activities are reduced. (Krishnamurthy *et al.*, 2005). Equipment monitoring requires tools based on historical data and statistical inference methods to define the health status of the system and to detect in advance pending failures with consequent timely pre-failure actions. (Susto, Beghi and De Luca, 2012). The PdM process develops according to 3 key steps (Martin, 1994; Jardine, Lin and Banjevic, 2006). The first one provides the *data acquisition*: data can be provided from different source and can have different nature as monitoring data, maintenance event data, process data. They are equally important and necessary for the implementation of the PdM. Usually, the first one is structured, and their collection is completely automated, the second one is unstructured, and their collection can be partially entrusted to the operator, therefore they need accurate validation. Then there is the *data processing*, as it is necessary to clean the dataset, then to analyse the data and to verify its consistency with the physical phenomenon and, finally, it is carried out the procedure for extracting features and information for the PdM objective. Finally, the *maintenance decision-making* phase is carried out. It is divided into two main categories, according to which different techniques are implemented: diagnostics (fault detection), in which the fault is identified real-time (Stetco *et al.*, 2019) and prognostic (fault prediction), whose goal is to estimate the Remaining Useful Life (RUL) of the monitored component.

Many studies have been carried out about predictive maintenance models. Zhou *et al.* (2007) presented a reliability-oriented model based on a continuous monitoring system for PvM; the model assumes that the system is subject to continuous degradation that can be monitored. Deloux *et al*., (2009) proposed a PdM application combining statistical process control with condition-based maintenance; this work focuses on the monitoring of stress value through a statistical process. The work of Curcuru *et al.* (2010), instead, proposed to model the degradation mode with a stochastic model combined with a Bayesian approach; they carried out a model that minimised total maintenance costs based on this assumption. Yang *et al.*, (2011) assumed that the degradation state is linked to the throughput amount, and so they proposed a joint transformation for lifetime and production volumes over maintenance levels. Sheu *et al.*, (2015) proposed a long-term average cost function for major, minor and imperfect repairs, assuming that the degradation of the equipment follows a non-homogeneous continuous-time Markov process.

The rapid development of Information Technology (IT) and the cyber manufacturing provide a new possibility for innovative PdM methods (Lee *et al.*, 2015). With cyber manufacturing systems large amount of data are available; it is possible to use these data to predict current and future health states of equipment and to carry out effective PvM maintenance strategies (Vogl, Weiss and Helu, 2019). Moreover, industrial strategies, such as Industry 4.0, encouraged investments in smart machines and tools that work on online networks, making possible to share multiple kind of information in real time: operational, environmental and process data. In this context ML is a useful tool to implement PdM models and decrease failures’ costs. Many works have been carried out using different approaches as classification methods, filtering and prediction approaches and regression methods (Jardine, Lin and Banjevic, 2006). Berka and Macek (2011) proposed a model for fault detection and diagnostics of a dynamic stochastic system to carry out effective maintenance actions; the diagnostic is carried out using a Bayesian approach to uncertainty and maintenance strategy is determined by a dynamic programming algorithm. Susto *et al.*, (2013) presented a PdM system based on the availability of the current values of the physical factors acting on the production process and on Support Vector Machines (SVMs), that is a powerful tool for dealing with classification problems; in particular it separate different classes of data deciding the optimal separation boundary. The model used SVMs to separate faulty from non-faulty state of the machines and it gave back the distance from the failure, as the equipment Remaining Useful Life (RUL). He et al., (2018), instead, proposed a PdM decision-making method based on mission reliability state for cyber manufacturing systems; they assumed that the cyber manufacturing system allows big data collection from the process and the transformation of these data in useful and meaningful information through cyber-physical systems development is the basis for PdM model.

PdM success depends on the quality and robustness of the condition monitoring system. In fact, at low level, a single subcomponent can be monitored, or, at uppermost level, it can be controlled the whole asset. Depending on the signal acquisition technique, the monitoring system is classified as intrusive (vibration analysis, oil debris) or non-intrusive (power signal) (Stetco *et al.*, 2019). In general four elements need to be in place: sensor technology for data collection, communication technology for data transfer, computation technology for data processing and management tools and practise to integrate the previous three elements (Pereira and Carro, 2007; Muller, Crespo Marquez and Iung, 2008; Campos, 2009; Widodo and Yang, 2011). These elements and related activities might be expensive and often compromises are necessary (Sirvio, 2015). Compromises are about extent, the frequency and the precision of data collection; data collection is measured in costs, speed and volume: the bigger is the speed and the volume of it the higher are costs related to ML investments.

Many studies have been carried out about PvM costs vs PdM costs, but no one considers investments costs of both strategies. This study fills this research gap and it considers these additional cost items in order to establish the best policy in terms of value and feasibility for the implementation of the PdM with the ML.

The aim of this paper is to compare the Unit-Expected Cost (UEC) of PvM and PdM obtained by reliability and cost data. It is assumed that this data is provided by the company, while the analysis of censored data will be carried out in further studies.

In this context, this paper presents a costs analysis comparing PvM and PdM costs, considering the influences of investment costs and consequently the accuracy of data collection.

3. Maintenance cost estimation

The implementation of the PdM with ML involves additional costs and development times compared to other maintenance strategies. As shown in the previous section, having online models involves the installation and development of an IT infrastructure, with which to run the model, and the use of resources (data scientists, process engineers and maintenance engineers) with various competences.

The proposed approach is based on the estimate of the main reliability functions. Weibull distribution has been chosen to determine the reliability of the components. It is completely defined by two parameters: θ (scale parameter) and β (shape parameter), which can be estimated from the time-to-failure (TTF) of the components (Manzini *et al.*, 2009). Hence, this distribution is very flexible and, for this reason, suitable for this study (Faccio *et al.*, 2014).

The reliability function is defined as follows:

$$R(t) = \exp\left[-\left(\frac{t}{\theta}\right)^{\beta}\right]. \tag{1}$$

3.1. Preventive maintenance Unit-Expected Cost (UEC_{PvM})

In order to evaluate the costs linked to the PvM policy, we have used the UEC model proposed in Faccio *et al.*, 2014, an adaptation of the best-known cost formulation proposed by Barlow and Hunter, 1960 for use-based maintenance in the event of age-based replacement. The following proposed model assumes that the application of PvM requires additional management costs dictated by the need to define maintenance plans and plan operations.

In accordance with Faccio *et al.*, 2014, the UEC for the age-replacement policy is as follows:

$$UEC_{PvM}(t = t_{PvM}) = \frac{C_p R(t_{PvM}) + C_f\left[1 - R(t_{PvM})\right]}{\int_0^{t_{PvM}} R(t)dt} + \frac{C_{PvM}}{T}, \tag{2}$$

where, β and θ are, respectively, shape and scale parameter of Weibull distribution, MTTR is Mean-Time-To-Repair, calculated as the average of the TTR when data is available, $R(t)$ is reliability function of the analysed component, $R(t_{PvM})$ is the reliability of the component at replacement interval time, which minimises UEC and C_{PvM} is fixed cost paid by maintenance management every time a replacement period is defined. C_f is average repair cost at failure defined as sum of C_{sp}, spare parts cost (direct cost), and C_{ma-FBM}, maintenance-action cost in case of failure-based maintenance, while C_p, average cost of preventive action, is the sum of C_{sp} and C_{ma-UBM}, maintenance-action cost in the event of use-based maintenance with age replacement policy.

3.2. Predictive maintenance UEC (UEC_{PdM})

The costs for implementing PdM with ML include the investment costs for the technology and the operating costs closely-linked to the performance of the ML model. These costs have a different importance according to the reliability parameters of the component. It is assumed that the system is non-inspectable; hence, when the model signals an anomaly, the component is replaced.

The replacement time is dictated by anomaly detection carried out by the ML model deployed before the failure occurs. This involves $R(t)$ as equal to 0 and replacement time, t, as equal to $+\infty$. The UEC_{PdM} can be estimated as follows:

$$UEC_{PdM}(t = +\infty) = C_P \times F \times f_{scoring} + \frac{C_f \times (1 - H) + C_P \times H}{\int_0^{t=+\infty} R(t)dt} + \frac{C_{PdM}}{T}. \tag{3}$$

A ML model for PdM can be treated as a binary classification model (classifier) in normal and anomalous behaviour. It classifies a temporal instant as anomalous in the event of incipient failure, thus it is possible to replace the component only if required by its condition and to avoid opportunity costs due to the unexploited useful life or maintenance action costs caused by sudden breakdowns (Susto *et al.*, 2015). It is assumed that the output of the ML model is the prediction of the anomaly in advance of a time range sufficient to allow intervention planning and the chance to apply opportunistic maintenance actions. Therefore, the cost of intervention is C_p, as defined in the previous sub-section.

Typically, the measure of a classifier's performance follows the theory summarised by the ROC curve (Marzban, 2004), which depends on the Hit Rate, H (Sensitivity or Recall), and on the False Alarm Rate (F) definition, which is also defined as the complement to 1 of the Specificity or Precision. If the model classifies samples correctly, the true negatives are the normal samples, while the true positives are the anomalous ones. If the model does not classify correctly, it can make a type II error and provide false negatives (predicting a normality erroneously) or a type I error and provide false positives (predicting an anomaly erroneously). The scenarios described are summarised in the confusion matrix shown in Table 1.

	PREDICTED NORMAL	PREDICTED ANOMALY
ACTUAL NORMAL	$true\ negative$ (tn)	$false\ positive$ (fp)
ACTUAL ANOMALY	$false\ negative$ (fn)	$true\ positive$ (tp)

Table 1: Confusion matrix for anomaly detection.

F and H measure the incidence of type I and type II errors in the predictions provided by the ML model, respectively. A type I error implies unnecessary intervention while a type II error leads to an intercepted fault; therefore, the former leads to an intervention cost equal to $C_p$while the latter to intervention with a cost equal to C_f. Mathematically, F and H are defined as below:

$$H = \frac{tp}{tp + fn}; \tag{4}$$

$$F = \frac{fp}{tn + fp}. \tag{5}$$

Real data, often, provides two classes that are not completely separable; there is an area where the two classes overlap (fig.2). The task of the data scientist is to define the anomaly threshold from which the F and H values derive. Establishing the F and H values for which the PdM model leads to savings, compared to other maintenance policies, can be preparatory to the definition of the anomaly threshold and to the establishment of the feasibility of expected performance.

The ROC curve defines the relationship between F and H: an increase in F involves an increase in H, according to a slope that depends on the degree of the two sets' separability. Separability is measured with AUC, which is equal to 1 if the sets are completely separable and equal to 0 if they are completely overlapping.

The UEC_{PdM}, for the above reasons, is evaluated only for $H \geq F$

H depends on the number of faults, while F depends on the number of the runs of the ML in the period. H is closely-linked to component reliability, while in order to evaluate the impact of F, $f_{scoring}$ has been introduced:

$$f_{scoring} = \frac{run}{T} \tag{6}$$

A component with a slow dynamic failure process involves a low number of runs in the period; the opposite is true for a system with a fast-dynamic failure process. For this study, we will assume $f_{scoring}$ is constant; its impact on final costs will be analysed in further studies.

The fixed cost for PdM with ML includes both the maintenance management costs introduced previously and the investment costs for the ML. The latter include installation costs, costs for model development and system maintenance costs for the reference period. Therefore, these costs are much greater than the maintenance management costs and C_{PvM} becomes negligible, as shown in Eq.7:

$$\begin{gathered}\frac{C_{PdM}}{T} = \frac{C_{PdM}}{T} + \frac{C_{ml}}{T}; \\ \frac{C_{ml}}{T} \gg \frac{C_{PdM}}{T} \Rightarrow \frac{C_{PdM}}{T} = \frac{C_{ml}}{T}\end{gathered} \tag{7}$$

The proposed cost model is representative of extreme cases in which ML does not work or is not present and those in which ML works perfectly.

With a ML model with faulty performance, $H = F = 0$, so the UEC_{PdM} becomes:

$$UEC_{PdM}(t = +\infty) = \frac{C_f}{\int_0^\infty R(t)dt} + \frac{C_{PdM}}{T}. \tag{8}$$

The UEC is equal to failure-based maintenance UEC with the addition of the unit investment costs due to the ML investment.

If a ML model has optimal performance, $H = 1, F = 0$, the UEC_{PdM} becomes:

$$UEC_{PdM}(t = +\infty) = \frac{C_P}{\int_0^\infty R(t)dt} + \frac{C_{PdM}}{T}. \tag{9}$$

In this case, all replacement interventions are planned and there are no unnecessary preventive interventions.

4. Analysis

The main aim of this paper is to evaluate the impact of the PdM parameters on the UEC_{PdM}. This work compares two maintenance policies, PdM and ARP-I, and calculates the savings for the component/system obtained with the application of PdM.

The analysis has been carried out by comparing $UEC_{PdM} \times \theta / C_f$ and $UEC_{PdM} \times \theta / C_f$ for the values set in Table 2:

In accordance with Faccio *et al.*, 2014, the UEC_{PvM} is evaluated for the intervention interval (t_{PvM}) which minimises it and depends on β and C_f / C_p, while the UEC_{PdM} depends on the combination of F and H and of the $C_f/C_{PdM} \times T$ ratio.

β	1, 2, 3
θ	2000h
C_f/C_p	2, 5, 10
C_f/C_{PdM}	2, 4
T	1760h, 3520h
H	$\geq F$

Table 2: UEC_{PdM} parameters for the savings analysis.

4.1 Saving analysis

In the heat-maps shown in figure 3, the iso-cost lines are drawn for the different values of β and C_f/C_p. Each row represents the estimated percentage savings due to the application of ML.

The graphs show how the application of ML is more justified for the low C_p costs. In Sgarbossa *et al.*, 2019, it has been shown that if C_f is much greater than C_p, the application of a preventive policy rather than a corrective one leads to savings of over 50%; with component monitoring, these savings increase. Where C_p is comparable to C_f, there is an advantage in the application of preventive policies only if the probability of identifying the anomaly is very high, otherwise the impact of costs for unnecessary interventions dominates and the investment costs in ML are not justified. Indeed, with lower ML costs (where T = 3520h), ML leads to savings even with low C_f/C_p.

With the same $C_f/C_{PdM} \times T$, the results do not vary. For example, if the work is carried out in 2 shifts, the C_f/C_{PdM} ratio can be halved, since, with the same θ, the avoided failures double.

As C_f/C_p increases with fixed C_f, the slope of the iso-cost lines decreases and becomes almost constant with the variation of F. If the cost of preventive interventions is low, the incidence of false positives decreases; considering the trend of the ROC curve, this factor allows H to be maximise during the anomaly threshold selection. Vice versa, when the C_p is high, the costs are sensitive even to the slight variations of F; the anomaly threshold selection thus minimises F.

As β increases, the slopes of the iso-cost lines are almost constant; however, they move significantly when C_f/C_p is medium-high, while when it is low the lines do not change.

In Figure 3, in the event of $C_f/C_p = 2$, the intercept q with H axes decreases with $\beta = 2$ and increases with $\beta = 3$, while with other C_f/C_p values q increases with increment of β. In accordance with Faccio *et al.*, 2014, with $1 \leq C_f/C_p \leq 4$, UEC_{PvM}, in the event where $\beta = 1$ has a lower value than in the event where $\beta = 2$. Whether $C_f/C_{PdM} \times T$ increases this effect is dampened.

In the case where $\beta = 1$, $t_{PvM} = +\infty$ and the UEC is corrective maintenance UEC.

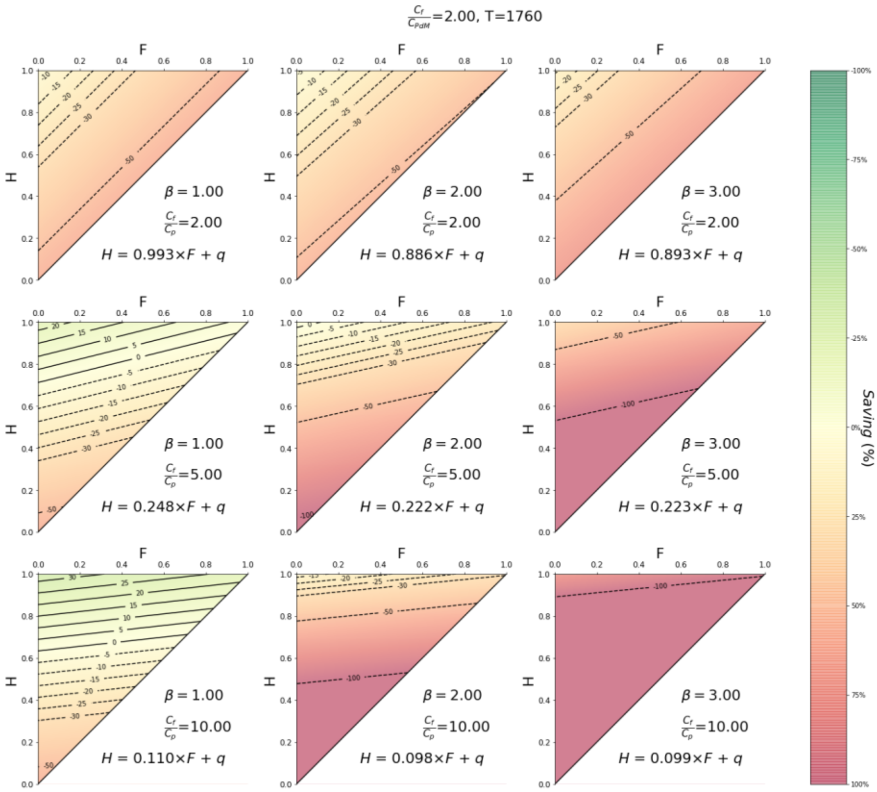

Figure 3: Iso-cost lines for $C_f/C_{PdM} = 2$ and $T = 1760h$

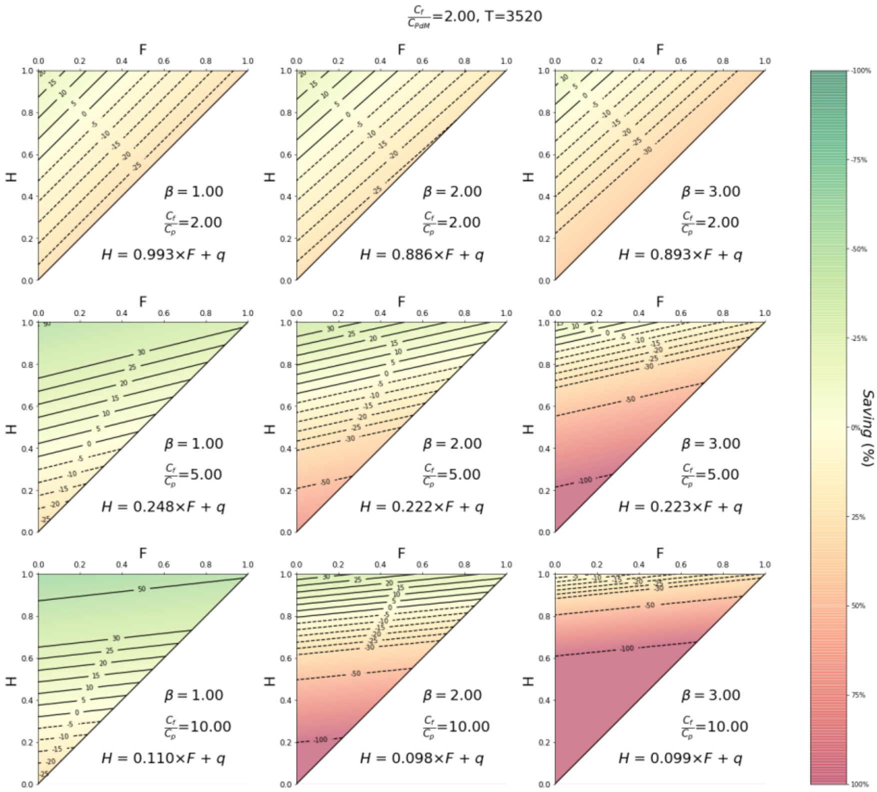

Figure 4: Iso-cost lines for $C_f/C_{PdM} = 2$ and $T = 3520h$

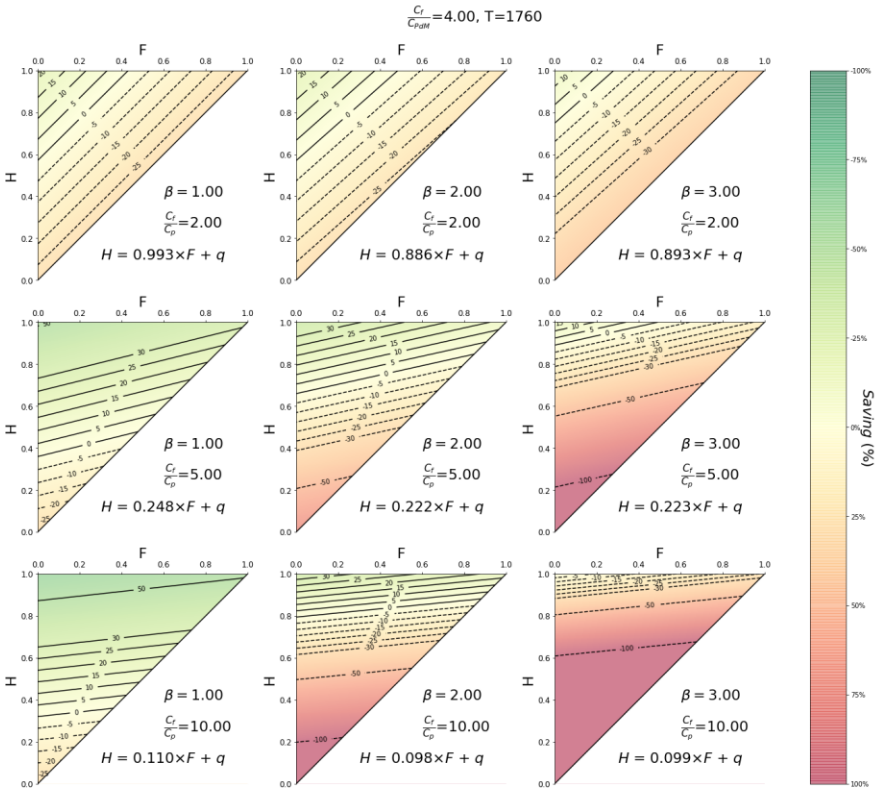

Figure 5: Iso-cost lines for $C_f/C_{PdM} = 4$ and $T = 1760h$

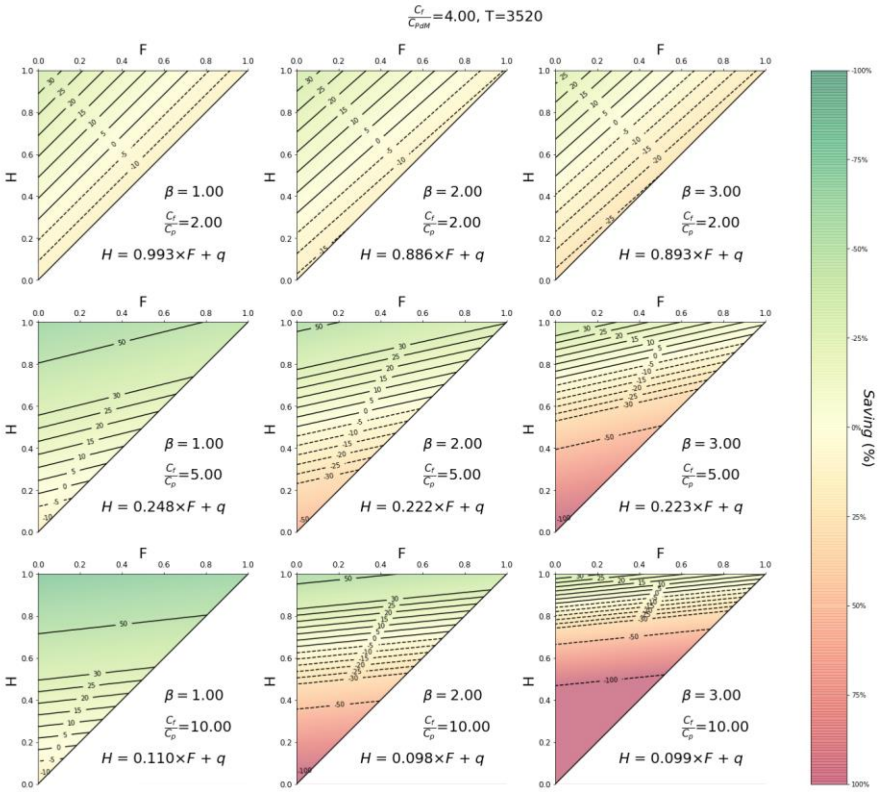

Figure 6: Iso-cost lines for $C_f/C_{PdM} = 4$ and $T = 3520h$

4.2 C_f/C_{PdM} threshold analysis

In a business context, in order to choose whether to apply the PdM or the PvM policies, it is important to know the size of the investment to be faced and the ML model performance necessary to obtain an economic advantage.

For this reason, the C_{PdM}/C_f ratio has been analysed for the different values of F and H. The area for which $UEC_{PdM} \geq UEC_{PvM}$ is represented. Figures 7 and 8 show the trend of C_{PdM}/C_f for which $UEC_{PdM} = UEC_{PvM}$. Having selected F and H, the application of ML is better value for C_{PdM}/C_f values lower than those identified on the iso-threshold upper-bound line.

For the evaluation of F and H, the considerations made in the previous subsection are valid. The iso-threshold lines are equidistant, while as T increases, their distance decreases. The costs are more sensitive to F and H changes: slight deviations lead to the passage between different iso-thresholds. It is, therefore, advisable to be more cautious when evaluating the feasibility of ML performance.

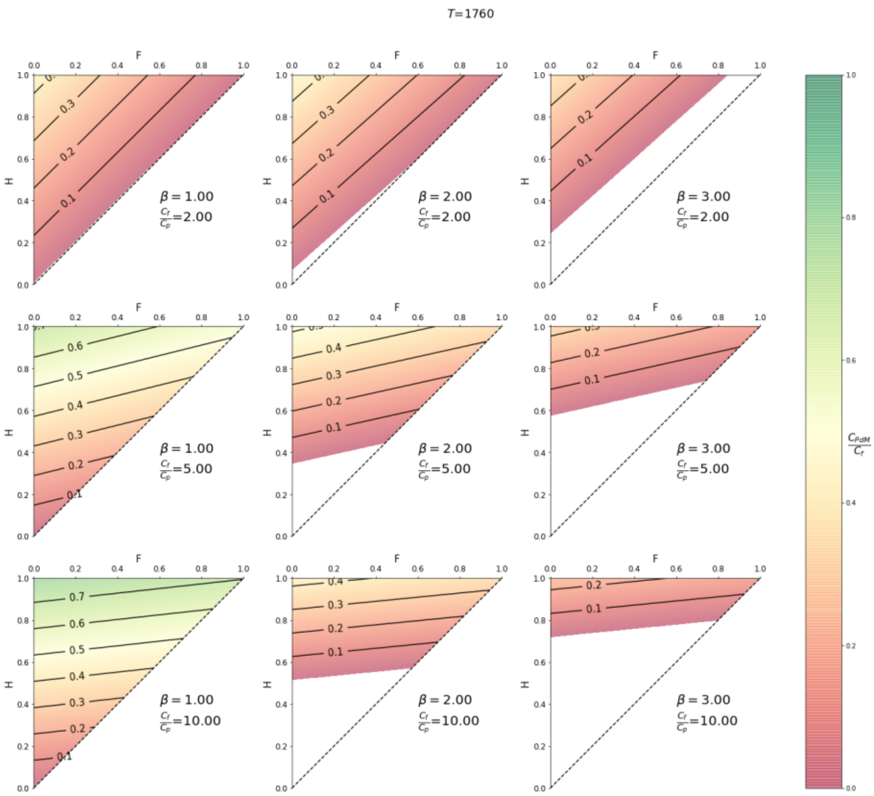

Figure 7: Iso-threshold C_{PdM}/C_f with $T = 1760h$ for different F and H.

5. Conclusions and further research

Nowadays, the global market requires the increasing availability of production systems. Aiming to minimise maintenance costs and to increase the availability of manufacturing systems, PdM might be the most suitable policy. In the last few years, the amount of available data for the state of machines and components and the interest of researchers and industry around ML have greatly increased: the development of PdM could be favoured by this. The issue is to establish whether, once the trade-off between costs and benefits has been considered, the installation of sensors and monitoring systems has an economic value.

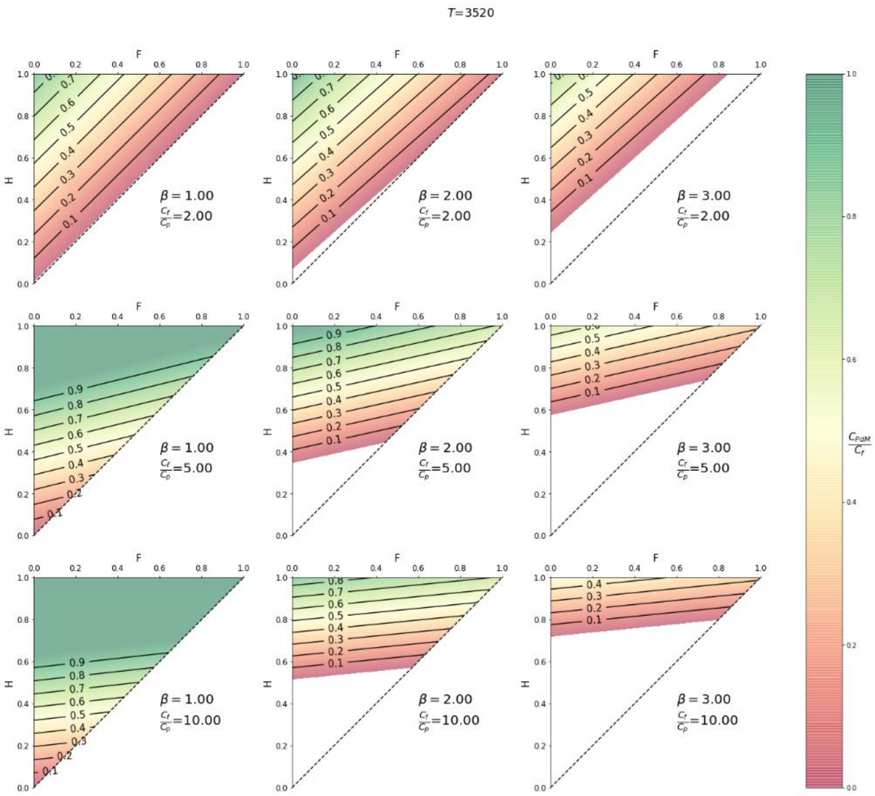

Figure 8: Iso-threshold C_{PdM}/C_f with $T = 3520h$ for different F and H

This paper introduces a cost-oriented model and investigates the impact of ML performance parameters and ML investment costs on total UEC_{PdM}.

It aims, as first, to estimate the costs and benefits of PdM policy studying the impact of intervention cost and reliability parameters, supposing the latter are known, on the positive saving area. From the savings analysis, it is clear that:

- As $C_f/C_{PdM} \times T$ increment increases the area of the positive savings, and the UEC_{PdM} is most suitable maintenance policy also with lower ML performance. These evaluations are valid both in the degradation period ($\beta > 1$) and in the random failure period ($\beta = 1$).
- The C_f/C_p and β increment requires higher H values.
- The C_f/C_p increase reduces the impact of F on the final $UEC_{PdM,}$ providing the opportunity to maximise H.

As second, it aims to define the C_{PdM}/C_f upper-bound below which PdM leads to savings. An easy-to-use tool has been introduced in order to evaluate, fixed C_{PdM}/C_f, the ML performance requires and, fixed F e H, the maximum value that C_{PdM} can have compared to $C_f$in order to obtain a return on investment.

Further researches will analyse the model in case of no data and the relationship between ROC-AUC curve and UEC_{PdM}.

Acknowledgements

The authors heartfelt thanks Icare S.r.l for having make the research possible. The authors are also grateful to the Maintenance Engineering team and all Data Science staff for thei close collaboration and great support.

References

Accorsi, R. *et al.* (2017) 'Data Mining and Machine Learning for Condition-based Maintenance', *Procedia Manufacturing*, 11, pp. 1153–1161.

Barlow, R. and Hunter, L. (1960) 'Optimum Preventive Maintenance Policies', *Operations Research*.

Barraza-Barraza, D., Limón-Robles, J. and Beruvides, M. G. (2014) 'Opportunities and challenges in Condition-Based Maintenance research', in *IIE Annual Conference and Expo 2014*, pp. 3035–3043.

Berka, J. and Macek, K. (2011) 'Effective maintenance of stochastic systems via dynamic programming', in *Proceedings of 19th Technical Computing Prague Conference*.

Bousdekis, A. *et al.* (2018) 'Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance', *Journal of Intelligent Manufacturing*, 29(6), pp. 1303–1316.

Campos, J. (2009) 'Development in the application of ICT in condition monitoring and maintenance', *Computers in Industry*.

De Carlo, F. and Arleo, M. A. (2013) 'Maintenance cost optimization in condition based maintenance: A case study for critical facilities', *International Journal of Engineering and Technology*.

Cocconcelli, M. *et al.* (2018) 'Development of a methodology for condition-based maintenance in a large-scale application field', *Machines*, 6(2).

Curcuru, G., Galante, G. and Lombardo, A. (2010) 'A predictive maintenance policy with imperfect monitoring', *Reliability Engineering and System Safety*.

Deloux, E., Castanier, B. and Bérenguer, C. (2009) 'Predictive maintenance policy for a gradually deteriorating system subject to stress', *Reliability Engineering and System Safety*.

Ding, S. H. and Kamaruddin, S. (2014) 'Maintenance policy optimization—literature review and directions', *International Journal of Advanced Manufacturing Technology*.

E. Florian, F. Sgarbossa, I. Z. (2019) 'Machine learning for predictive maintenance: a methodological framework', in.

Faccio, M. *et al.* (2014) 'Industrial maintenance policy development: A quantitative framework', *International Journal of Production Economics*, 147, pp. 85–93.

He, Y. *et al.* (2018) 'Cost-oriented predictive maintenance based on mission reliability state for cyber manufacturing systems', *Advances in Mechanical Engineering*.

Hofmann, T., Schölkopf, B. and Smola, A. J. (2008) 'Kernel methods in machine learning',

Annals of Statistics.
Jardine, A. K. S., Lin, D. and Banjevic, D. (2006) ‘A review on machinery diagnostics and prognostics implementing condition-based maintenance’, *Mechanical Systems and Signal Processing*, 20(7), pp. 1483–1510.
Krishnamurthy, L. *et al.* (2005) ‘Design and deployment of industrial sensor networks: Experiences from a semiconductor plant and the North Sea’, in *SenSys 2005 - Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems*.
Lee, J. *et al.* (2015) ‘Industrial Big Data Analytics and Cyber-physical Systems for Future Maintenance & Service Innovation’, in *Procedia CIRP*.
Madu, C. N. (2000) ‘Competing through maintenance strategies’, *International Journal of Quality & Reliability Management*.
Manzini, R., Regattieri, A., Pham, H., & Ferrari, E. (2009) *Maintenance for industrial systems.* Edited by S. S. & B. Media.
Martin, K. F. (1994) ‘A review by discussion of condition monitoring and fault diagnosis in machine tools’, *International Journal of Machine Tools and Manufacture*, 34(4), pp. 527–551.
Marzban, C. (2004) ‘The ROC curve and the area under it as performance measures’, *Weather and Forecasting*.
Muller, A., Crespo Marquez, A. and Iung, B. (2008) ‘On the concept of e-maintenance: Review and current research’, *Reliability Engineering and System Safety*.
Pereira, C. E. and Carro, L. (2007) ‘Distributed real-time embedded systems: Recent advances, future trends and their impact on manufacturing plant control’, *Annual Reviews in Control*.
Pozzi, R. and Strozzi, F. (2018) ‘How assembly systems are adopting the technologies of I40: A preliminary landscape’, in *Proceedings of the Summer School Francesco Turco*, pp. 369–375.
Prajapati, A., Bechtel, J. and Ganesan, S. (2012) ‘Condition based maintenance: A survey’, *Journal of Quality in Maintenance Engineering*, 18(4), pp. 384–400.
Sala, R. *et al.* (2018) ‘How to select a suitable machine learning algorithm: A feature-based, scope-oriented selection framework’, in *Proceedings of the Summer School Francesco Turco*, pp. 87–93.
Sgarbossa, F. *et al.* (2019) ‘Age replacement policy in the case of no data: the effect of Weibull parameter estimation’, *International Journal of Production Research*.
Sheu, S. H. *et al.* (2015) ‘Optimal preventive maintenance and repair policies for multi-state systems’, *Reliability Engineering and System Safety*.
Sirvio, K. M. (2015) *Intelligent systems in maintenance planning and management*, *Intelligent Systems Reference Library*.
Stetco, A. *et al.* (2019) ‘Machine learning methods for wind turbine condition monitoring: A review’, *Renewable Energy*, 133, pp. 620–635.
Susto, G. A. *et al.* (2013) ‘Prediction of integral type failures in semiconductor manufacturing through classification methods’, in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*.

Susto, G. A. *et al.* (2015) ‘Machine learning for predictive maintenance: A multiple classifier approach’, *IEEE Transactions on Industrial Informatics*, 11(3), pp. 812–820.

Susto, G. A., Beghi, A. and De Luca, C. (2012) ‘A predictive maintenance system for epitaxy processes based on filtering and prediction techniques’, *IEEE Transactions on Semiconductor Manufacturing*.

Vogl, G. W., Weiss, B. A. and Helu, M. (2019) ‘A review of diagnostic and prognostic capabilities and best practices for manufacturing’, *Journal of Intelligent Manufacturing*, 30(1), pp. 79–95.

Waeyenbergh, G. and Pintelon, L. (2002) ‘A framework for maintenance concept development’, *International Journal of Production Economics*, 77(3), pp. 299–313.

Widodo, A. and Yang, B. S. (2011) ‘Machine health prognostics using survival probability and support vector machine’, *Expert Systems with Applications*.

Yang, Y. *et al.* (2011) ‘A hybrid feature selection scheme for unsupervised learning and its application in bearing fault diagnosis’, *Expert Systems with Applications*, 38(9), pp. 11311–11320.

Zhou, X., Xi, L. and Lee, J. (2007) ‘Reliability-centered predictive maintenance scheduling for a continuously monitored system subject to degradation’, *Reliability Engineering and System Safety*.

Eleonora Florian graduated at the University of Padua in Management Engineering. Currently is a Ph.D. student in University of Padua. The topic of his research concerns the application of machine learning and deep learning solution on maintenance management.

Fabio Sgarbossa is Full Professor of Industrial Logistics at the Department of Mechanical and Industrial Engineering at NTNU (Norway). He is working within the Production Management Group at MTP and he is responsible of the Logistics 4.0 Lab. He has been and he is involved in several European and National Projects. He is author and co-author of about 90 publications in relevant international journals. He is member of Organizing and Scientific Committees of several International Conferences, and he served as guest editor of several Special Issues in relevant International Journals.

Ilenia Zennaro received the Ph.D. degree in Mechatronics and Product Innovation Engineering, with the curricula in Industrial Plants and Logistics, from the University of Padua, Italy, in 2017. From 2018 is Assistant Professor and Researcher with the University of Padua, in the department of Management and Engineering. Her research interest includes industrial and logistics systems design, management and modelling. Her research has been published in various international scientific journals and conference proceedings.