Data-driven CBM Tool for risk-informed decision-making in an Electric Arc Furnace

Luca Fumagalli [1], Laura Cattaneo[1,*], Irene Roda[1], Marco Macchi [1] and Maurizio Rondi[2]

[1] Department of Management, Economics and Industrial Engineering, Politecnico di Milano, P.za Leonardo da Vinci 32, 20133 Milano, Italy; {luca1.fumagalli; laura1.cattaneo; irene.roda; marco.macchi}@polimi.it

[2] Tenaris Dalmine S.p.A., Piazza caduti 6 luglio 1944 n.1 - 24044 Dalmine (BG), Italy.

***** Correspondence: laura1.cattaneo@polimi.it; Tel.: +39-02-2399-2728

Abstract: Nowadays maintenance activities and safety management can be supported by a mature state of the art favouring the implementation of Condition Based Maintenance programme, which recommends maintenance decisions based on the information collected through asset life. The main idea, which grounds in the Industry 4.0 paradigm, is to utilize the asset degradation information, extracted and identified through different techniques, to reduce and eliminate costly, unscheduled downtimes and unexpected breakdowns and to avoid risky scenarios. This paper aims at developing and testing a data-driven CBM tool to provide fault diagnostics transforming raw data from the shop-floor into information, finally enabling risk-informed decision-making. The tool relies on a process of knowledge discovery that incorporates both prior knowledge and proper interpretation of data analytics results. Prior knowledge is extracted through a Process Hazard Analysis (PHA), while data analysis deals with Statistical Process Control and Novelty Detection. The model is proposed to integrate some Cyber-Physical System element in the extant plant automation, to exploit its computational capabilities through the continuous monitoring and data analytics. This enables a “watchdog agent” of risky scenario, allowing an on-line risk-assessment of safety-critical components, finally enhancing the intelligence in the industrial process.

Keywords: Condition Based Maintenance; Process Hazard Analysis; Statistical Process Control; Novelty Detection; decision support system; risk assessment.

1. Introduction

During the last years, a large amount of manufacturing data was collected in database management systems and data warehouses, due to the wide use of distributed and automated control systems and the usage of embedded sensors in the physical assets composing the manufacturing plants (in the reminder, also referred to as manufacturing assets). This large amount of data was not fully exploited, since it was more used for daily technical checks and process log fulfilment. Besides, smart devices were not used to construct a continuous and seamless flow of information throughout the entire processes. This was observed because no infrastructure existed for delivering, managing and analyzing the data over a network, even if the devices were networked [1].

More recently, many countries have announced a new wave of development plans in manufacturing. Among these, Germany has proposed the concept of Industry 4.0 [2, 3] whose main goal is to develop smart factories for higher competitiveness and flexibility. In this context, “smart” is strictly connected with information, since the ability to extract information from data is needed in order to make the manufacturing process more intelligent.

The technological basis of Industry 4.0 has been identified as the Internet of Things (IoT) and the Cyber Physical Systems (CPS). IoT proposes to embed electronics, software, sensors, and network connectivity into devices (i.e. things), to allow the collection and exchange of data through the Internet [4, 5]. CPS are smart embedded and networked systems within production systems, operating both in the virtual (cyber) and physical space, interacting with and controlling manufacturing assets, sensing and acting on the real world [6–8]. With the use of these technologies, Industry 4.0 opens the way to real-time monitoring and synchronization of the real world activities to the virtual space thanks to the physical-virtual connection and the networking of CPS elements [8].

The development of IoT and CPS technologies has been contemporary to the development of data storages, computational power and analysis algorithms, which have experienced a fast improvement. Therefore, research in data analysis and artificial intelligence for smart factories has become an unquestionable trend: they aim to analyze raw data in order to discover hidden patterns and relationships among different variables [9, 10]. In other words, the main goal of data analytics – including data analysis based on statistical models and machine learning majorly due to the artificial intelligence techniques [11] – is to extract useful information from raw data and transfer it to effective knowledge to improve the understanding of processes (running in the physical space) and to support decisions (within the cyber-physical space) [12]. In this direction, a manufacturing asset usage could be reported to selected stakeholders (e.g., designers, plant managers, asset/maintenance managers) through proper monitoring and data analytics, with the purpose of improving the design and the management of manufacturing plants. Condition Based Maintenance (CBM) holds great research value in this smart manufacturing context, relying on the prominent role of data-driven approaches for CBM program development. Research in data-driven CBM requires an interdisciplinary background of signal processing, statistics, computer science, and necessary domain knowledge [1].

The model herein proposed relies on such a data-driven CBM to enhance transparency to manufacturing assets' capabilities and their current conditions in order to discover hidden patterns that anticipate risky scenarios. The model builds on a knowledge discovery that incorporates both prior knowledge and proper interpretation of data analytics results. Prior knowledge is extracted through a Process Hazard Analysis (PHA), aimed at identifying the asset critical scenarios according to risk management principles, and at mapping the manufacturing asset behavior in terms of healthy, abnormal, faulty states, and related failure mechanisms determining – as physical and/or chemical transformation processes – the asset degradation. Proper interpretation of data analytics aims at exploiting the huge amount of data in manufacturing databases with the purpose to discover useful information through a set of monitored features recommended to anticipate risky scenarios. This is particularly relevant as risky scenarios, especially those causing catastrophic effects, are typically experiencing very low (or null) availability of Run-To-Failure (RTF) data; this finally leads to reflect on the possibility (and even on the necessity) to characterize risky scenarios by relying not directly on the occurrence of failures, but on indirect measures – i.e. features extracted from the monitored asset conditions – as proxy for the impending failures.

The following hypotheses stand as further requirements to meet the challenges from extant industrial contexts.

1. Data acquisition should not require additional sensors in the physical assets. Indeed, in some industrial contexts, sensors can be hardly installed due to technical constraints (harsh environmental conditions, physical limits in available volumes, …); in some other contexts, also for economic and managerial reasons (such as the need to avoid long shutdowns), the extant plant automation is solely retrofitted by integrating components to manage the network connectivity of the physical assets (by adding hardware, as a gateway, and related software in order to enable data streaming from physical space, i.e. the industrial machinery, to cyber space, where other software are used for monitoring and data analytics).
2. Failure mechanisms determining the asset degradation can be typically identified based on process experts, but can be hardly specified based on mathematical equations expressing the physics of the asset degradation. Even with this limitation, prior knowledge could be an aid to drive the knowledge discovery by means of an initial identification of potential candidates of key measurable parameters (i.e. key variables or key features) to monitor and map the asset behavior. If this can be done, the subsequent requirement is to adopt data analytics to help completing the initial model, complementing the prior knowledge with a data-driven approach to fully understand the asset degradation process, which means finally connecting the physical with the cyber space through proper models.
3. Proper interpretation of data analytics results should be accepted by process experts, knowledgeable of the real-world physics within the industrial process under study. This requirement extends the validation not only on computational capabilities, but also to time and accuracy of data conversion to information. This is particularly relevant in the problem setting of this paper, where the objective is to address a decision making that is informed, based on proper data analytics results, of risky scenarios that are going to happen.

The model is proposed to integrate extant plant automation. In particular, its computational capabilities are assumed to be exploited through a continuous monitoring that extends the extant automation and to enable a "watchdog agent" of risky scenarios. To this end, the hypothesis is to start from a "traditional" automation featuring, in a hierarchy, the supervision and control of the industrial process and asset under study. Relying on it as a base-ground, some CPS elements should be introduced to enable real-time monitoring and data analytics running beside the "traditional" control. This last assumption is driven by needs of industrial deployment: CPS are in fact assumed to be introduced in the "brown field", thus adding data management and computational capabilities to enhance intelligence in the industrial process. The term "watchdog agent" is borrowed from the concept originally introduced by [13]: herein, the term is adopted considering an extended sense of condition monitoring, linking to an on-line risk-assessment of safety-critical components.

Given all the previous assumptions, in a nutshell the research contribution of this paper is to formalize and to test a data-driven CBM tool enabling an on-line control to support a risk-informed decision-making. The test-bed implementation of this data-driven approach is put in the context of the steel making industry, and finally leads to the development of a smart maintenance tool acting as watchdog agent of risky scenarios due to a critical component of an Electric Arc Furnace (EAF). The paper grounds on some first information provided by [14, 15] and on an extension of [16].

The rest of the paper is organized as follows. Section 2 presents the research methodology. Section 3 introduces the state of the art focusing on data analytics for CBM. Section 4 describes the industrial case, introducing the EAF of a steel making plant and its key elements in order to set the problem, while also describing the expected benefits due to the introduction of such CBM tool. Section 5 presents the models developed, highlighting the data analytics methods used to perform CBM. Section 6 describes the deployment of such models in the extant ICT architecture of the plant, presenting the results in practice, and illustrating the output information as well as its usefulness for risk-informed decision making. Conclusions and future developments are reported in Section 7.

2. Research Methodology

The research methodology defined and followed in this paper is composed by three main steps (see Fig. 1), addressed to establish a CPS framework where physical space is logically connected with the virtual (cyber) one, in order to allow interaction and information flow with the purpose to build the data-driven CBM tool and, subsequently, to enhance the risk-informed decision-making.

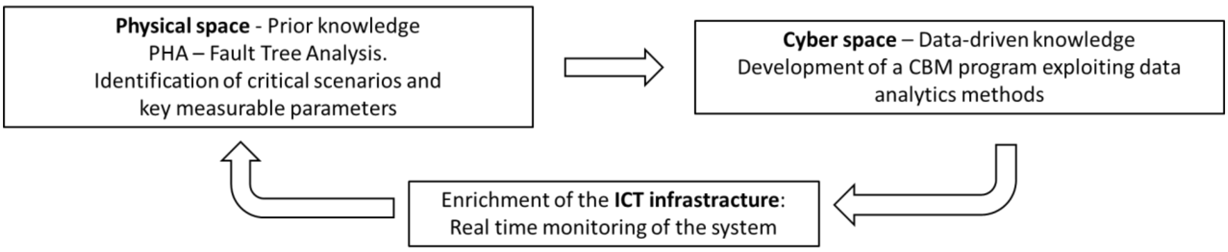


Figure 1. CPS research methodology framework.

i) Physical process description is performed based on prior knowledge. Since we assume that physical models and mathematical equations cannot be easily defined due to the complexity of the manufacturing asset under study, prior knowledge is extracted through a Process Hazard Analysis (PHA). Relying on experts' knowledge, PHA initially aims at identifying the asset critical scenarios according to risk management principles, and then at mapping the manufacturing asset behavior in terms of healthy, abnormal, faulty states, and related failure mechanisms determining – as physical and/or chemical transformation processes – the asset degradation [15]. Among the different PHA techniques to study asset failures (FMEA, HAZOP, Fault Trees, Event Trees, Reliability Diagrams, and so on), the work described in this paper relies on Fault Tree Analysis (FTA), judged as convenient technique for the specific risk-based analysis of the faults of the safety-critical component under study. FTA provides evidence of critical scenarios, describing how the asset under study can fail; as such, it also helps identifying the key measurable parameters (i.e. key variables) to monitor and map the asset behaviour.

ii) Critical scenarios and failure mechanisms, determining the asset degradation, are monitored through the engineering and deployment of a CBM tool: models, exploiting data analytics methods, are then developed to perform the CBM program. In this way, prior knowledge is completed and enhanced with a data-driven approach to fully understand the asset degradation process, moving – through the data analytics methods – from the physical space to the virtual (cyber) one. These techniques are chosen considering that the asset cannot experience catastrophic effects and, as such, historical Run-To-Failure (RTF) data are not available.

Data analytics is then intended to process the features extracted from monitored asset conditions as proxy for the asset risk of failure.

iii) The effective connection between the physical and the cyber space is achieved through the implementation and integration of the developed models in the extant ICT architecture of the plant, thus enabling real-time monitoring and data analytics running beside the “traditional” hierarchical control. In this way, data management and computational capabilities are enhancing intelligence in the industrial process, leading to a new functionality in the cyber space that enables the risk-assessment of safety-critical components while the asset is running (i.e. on-line risk-assessment). From process supervision, it means enhanced capabilities to run the production process with inherent safety.

3. State of the Art of Data Analytics for Condition Based Maintenance

Cost-effective and accurate maintenance shows an increasing importance in improving asset availability, quality products and operational safety, having a direct impact on the competitiveness of organisations. To comply with the cost-effectiveness and accuracy requirements, CBM has become established throughout industry [17, 18], in particular in high-risk sectors [19, 20].

CBM is a program that recommends maintenance decisions based on the information collected through condition monitoring [21]. The main idea is to utilize the asset degradation information, extracted and identified from on-line sensing techniques, to reduce and eliminate costly, unscheduled downtimes and unexpected faults, ultimately to optimize asset utilization in the facility and to achieve near-zero unplanned downtimes [13, 22–24]. Diagnostics and prognostics are two important activities in a CBM program. Diagnostics deals with fault isolation and identification, before or after the failure occurs; prognostics is focused on the failure mode evolution and deals with fault prediction, before it occurs [21, 25]. A CBM program can be used both to diagnose and, when relevant, to prognosticate the asset behavior.

No matter what the final objective of a CBM program is, the process is typically composed by different steps: i) Data Acquisition, ii) Data Manipulation (DM), iii) State Detection (SD), iv) Health Assessment (HA), v) Prognostic Assessment (PA), vi) Advisory Generation (AG) [26]. All these steps are explained as functional blocks by an international standard, the MIMOSA OSA-CBM (Machinery Information Management Open Standards Alliance), corresponding to the ISO 13374 [27]. The last three functions of MIMOSA OSA-CBM process, i.e. from HA to AG, are inherently connected with failure risk management: they provide functionalities relevant for a decision-making support where the risk that an asset will fail can emerge through proper information (e.g. an asset health index representing the asset health status as a proxy for its risk of failure) and, then, determine correspondent mitigation actions [17, 18].

Data analytics represents an important lever in the development of CBM, both in diagnostics and prognostics phase. A variety of models, algorithms and tools are available in the literature to analyze data, and the choice among them mainly depend of the types of data collected and the knowledge available about the asset/system. According to [21], the different data analytics methods can be classified as statistical, model-based and Artificial Intelligence (AI) approaches; other classifications exist in literature [28, 29], where often AI approaches are classified as Machine Learning or Data-Driven approaches.

Statistical approaches deal with the objective to detect if a specific failure is present or not, based on some available condition monitoring data, without intrusive inspection of the asset. Among the different approaches, statistical inference and, in particular, hypothesis testing has been used for fault diagnosis [21]. Another conventional approach is Statistical Process Control (SPC): SPC has been developed in the field of quality control theory, but it has been used also in other contexts of the manufacturing process control, such as diagnostics and fault detection [21, 30]. The main idea of SPC is to control if a signal (or a feature), that should represent normal condition, is within control limits or not. Cluster analysis is another method, consisting of a multivariate statistical analysis that groups signals (or features) into a certain number of heterogeneous groups that possess homogenous contents in terms of signal characteristics. In this way, differences between the groups are identified, while the signals within a single group are similar. The Cluster analysis algorithm computes distance (or similarity) between two signals to assess if the two signals belong to the same group; commonly used distance measures are Euclidean distance, Mahalanobis distance, Kullback–Leibler distance and Bayesian distance. Application of Cluster analysis in machinery fault diagnosis has been discussed in [31–34]. Other statistical algorithms such as Nearest Neighbors and Linear discriminant functions, have been used for fault classification. In this case, two closest groups are fused into a new group where the distance between the two groups is the distance of the two nearest neighbors in the two separate groups. Linear discriminant function is often used as boundary for two adjacent groups [35, 36].

A sharp limit between statistical method and AI/ML algorithms actually does not exist, as also illustrated in [11]. Besides this awareness, we can assess that AI/ML algorithms can be classified in two main categories, depending on the available data to be analyzed, i.e. supervised learning and unsupervised learning algorithms [37].

Supervised Learning is performed using labelled data, i.e. data that are directly related to a specific condition of the asset (normal or healthy/abnormal or degraded/faulty), or for which it is possible to measure and register both some inputs and a related output for a specific number of observations (for example, key features as inputs and the asset health status as related output). Therefore, the supervised learning algorithm receives a set of inputs along with the corresponding correct outputs/labels, and it learns the mapping function that connects them. The goal is to approximate the mapping function so well that the algorithm is able to receive in input new data with no output/label and predict the correspondent output (i.e. assigning the correct output/label). Applying supervised learning in CBM directly means having available historical data of both normal (healthy) condition and degraded (abnormal) condition, needed to train the model [21]. Examples of algorithms that belong to this category and that have been used to make diagnosis in CBM program include K-Nearest Neighbors, Support Vector Machine, Discriminant Analysis, and Neural Networks [37–39].

Unsupervised Learning is performed when data used for training have no label attached. The algorithm receives a set of inputs and, without having corresponding outputs, it explores the dataset with the purpose of detecting a structure or a hidden distribution within the data, and to learn more about them. Regarding CBM, unsupervised learning is suitable when there are no historical data on degradations/failures. Examples of algorithms that belong to this category include K-Means Clustering, Self-Organizing Maps, and Hierarchical Methods [37, 40].

Between the supervised and unsupervised categories, Novelty Detection (ND) exists that “can be defined as the task of recognizing that test data differ in some respect from the data that are available during training” [41]. More precisely, ND algorithms deal with a supervised problem, called one-class classification problem, in which a training dataset is available, with the characteristic of presenting only (or mainly) “normal” behavior outcomes and insufficient data describing the “abnormal” ones. This situation is frequently present in the real industrial world: a machinery monitoring system is likely to collect only data regarding the correct machine behavior (that is, the normal or healthy state), since measurements of failures are not present (i.e. the machine has never broken since it is new or, when the machine is turned-off due to an impending failure, data collection was also turned-off). Therefore, ND represents a powerful set of algorithms and methods that can be used to train datasets where no examples of the novel class are contained [41].

Lastly, physical (or model-based) approaches rely on explicit mathematical models of the monitored asset and of the physics of its degradation. These models can be more reliable and effective than other model-free or data-driven approaches, if the model is correctly built and validated. However, mathematical modelling may not be feasible for complex assets/systems, since it would be very difficult or even impossible physically describing these assets/systems. As reported in [21], different model-based approaches have been applied to fault diagnosis of a variety of mechanical systems, such as gearboxes, bearings, rotor and cutting tools [42].

4. The industrial case

4.1 The industrial setting, the Electric Arc Furnace and the critical component under study

Tenaris is the leading global manufacturer and supplier of tubular products and services used in the drilling, completion and production of oil and gas, in process and power plants and in specialized industrial and automotive applications. An important part of Tenaris’s manufacturing capacity is located in Italy at the state-of-the-art seamless pipe mill in Dalmine, Bergamo. Tenaris Dalmine is the headquarter of Tenaris Italia, which has about 3000 employees, 5 premises and a yearly production capacity of 950,000 tons of finite goods. Tenaris’s process for producing seamless steel pipe is based on the EAF, Ladle Furnace, Vacuum Degassing and Continuous Casting process.

In Tenaris Dalmine there are two EAFs, with the following characteristics: capacity 105 t, diameter 6.1 m, weight of the structure 230 t. Each furnace allows to obtain liquid steel, being charged by selected scrap and varying percentages of pig iron [15], and processing with an average power request of 67 MW per cast, and a maximum power peak of 89 MW. The casting cycle is around 38.5 min and this corresponds to a productivity above 150 t/h. The main contribution of energy for melting the scrap is supplied from the electricity for about 2/3 and for 1/3 by chemical energy supplied by the lances and the burners on board.

The EAF’s structure, represented in Figure 2, is constituted by a lower sheet metal keel coated with refractory bricks, which serves to contain the liquid steel, and is cooled by a cage, with a structural function that supports panels cooled by a water circuit [43]. The panels are positioned around the furnace circumference on two different rows: they are defined by Tenaris Dalmine experts as “down panels”, when they are in the first row close to the bottom part of the furnace, and “up panels” when they are in the second row, close to the furnace roof.

The panels are the critical component of the EAF under study: during the heating process, they could be stressed out and thus fail with a leak of water in the furnace, which is finally leading to risky scenario (risk of furnace explosion). It can be easily understood that the high amount of energy provided to melt the steel, also impacts on the stress that the panels can suffer, leading to the panels' degradation.

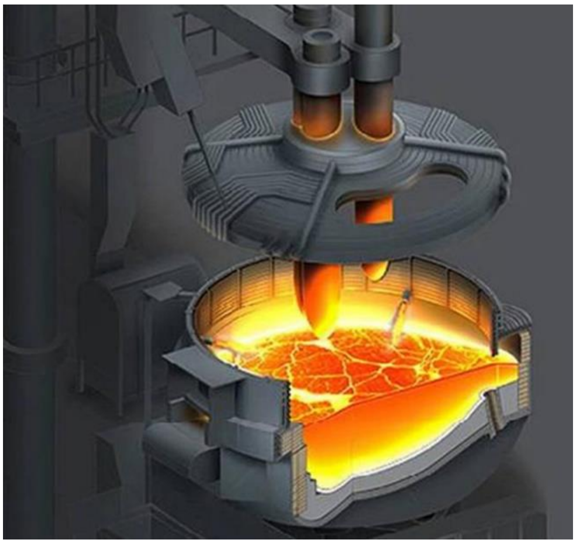


Figure 2. Rendering of the EAF with panels.

4.2 The CBM program development practice and the expected benefits

The maintenance practice to maintain the panels is based on regular inspections, with subsequent revision and repair when a serious degradation (visible on the surface of the panel), or even a hole (inducing a high risk of catastrophic effects, due to the leak of water), has occurred.

To inspect the panels, the furnace must be stopped. As two EAFs are available in Tenaris Dalmine, one is used when the other is under revision and repair. Nevertheless, the process of changing the furnaces, cooling down the one just used, and making it available for maintenance, can take days. This requirement prevents to schedule this activity frequently and may potentially lead to a period in which the plant operation is running in a risky scenario. Nonetheless, risky scenario does not necessarily mean that there is an actual risk of serious dangerous event, in fact: i) the panels experience a (very) long lasting degradation process, and their frequency of inspection has been carefully adjusted after years of plant operation; ii) moreover, a visual inspection is possible, during operations, thanks to the use of some cameras; due to this visual inspection the furnace risks to be stopped in the middle of the period of activity in order to solve problems on the panels.

It is evident that the proposed CBM tool allows, through continuous monitoring, to enhance capabilities to run the production process with inherent safety: it enables the risk-assessment of the panels as safety-critical components while the EAF is running (i.e. on-line risk-assessment). This leads to better control the level of risk during the operations, relying on the data-driven CBM as improved methodology compared with the usual inspection practice. Indeed, the improved CBM program supports two main functions: i) the time when the EAF must be stopped can be dynamically scheduled, also exploiting the on-line risk assessment of the panels as safety-critical components; ii) the panels that are worth of attention – i.e., those that are marked with high risk – can be identified before the casting is stopped, this allows to save time during the inspection and maintenance, because the maintenance operators can focus only on these high risky panels.

In a nutshell, the expected benefits provided by the CBM tool are: i) the decrease of risk during the asset operations and ii) the resource savings during inspection and revision – i.e. efficiency in the resource usage and corresponding cost savings. Furthermore, a better understanding of the stress of the panels based on different operating practices could be achieved. In this way, knowledge is capitalized, directly serving to a better maintenance scheduling: even if the inspection frequency might be adequately planned, the CBM tool is used to dynamically adjust, based on the asset risk, the schedule of inspection and revision activity.

5. Data-driven CBM tool for risk-informed decision-making in an Electric Arc Furnace

Following the research methodology presented in Section 2, a CBM tool has been constructed for the on-line control and the support of risk-informed decision-making, in the EAF in Tenaris Dalmine. In particular, the CBM tool is built to provide fault diagnostics transforming raw data from the shop-floor into information, finally enabling both maintenance decision-making and risk assessment. In particular, we focus on diagnostics before the failure occurs, according to the objective to monitor and to control the EAF health status in order to avoid risky scenarios when the EAF is running.

The first step of the research methodology consists in physically describing the system. This description is built exploiting the expert's knowledge through FTA as PHA technique. FTA develops a tree, called Fault Tree, which allows a graphical representation of the logical path that connects an unwanted event – namely the "root" of the tree, also said "top event" – with the events that are its root causes. The technique follows a hierarchical methodology which first connects the top event with the events that are directly intended as its root causes; such events (intermediate events) are in turn related to other events that are the underlying causes. The hierarchical tree construction is completed when it has reached the so-called events/basic causes – that is to say, the "leaves" of the tree. For the proposed case, FTA was built exploiting the knowledge of company's experts and integrated thermo-mechanical notions to properly describe possible failure mechanisms. The unwanted event, i.e. the "top event" – defined by Tenaris Dalmine experts as the one to be taken under control – is the "holing" of the EAF panels. Developing the FTA for this kind of panels, especially those called "up panels", results in the identification of three different basic causes, related to three different failure mechanisms, such as: i) progressive degradation (this is due to the panel exposition to thermal irradiation; it happens when the "up panel" is not covered by the steel slag); ii) suddenly breakdown by means of a splash (this is a mechanical broken event, neglected in this work due to its high rarity according to experts); iii) corrosion (this is a chemical process that is not actually considered in this work as the panel is clad by a cathode protection). Once identified the "holing" of the panel as the top event and its basic causes, it is needed to identify the key measurable parameters (i.e. the key variables) able to measure the related failure mechanisms. Exploiting again the prior knowledge of Tenaris Dalmine experts, it has been decided to monitor the temperature of the water that flows within the cooling circuit of the panels. This temperature is an indicator of the thermal load felt by the panel during the casting. The thermal load is directly connected with the energy absorbed by the panel and could therefore describe the thermal stress at which the panel is subject, that causes the progressive degradation of the panel condition (i.e. the failure mechanisms i) under study).

The key feature (KF) describing the thermal loading is subsequently defined as:

$$\text{Key Feature (KF)} = \int_{Tstart}^{Tend} (T_{H_20}^{out}\,(t) - T_{H_20}^{in}(t))dt, \quad (1)$$

where T_{H_20} indicates the water temperature within the panel circuit, $T_{H_20}^{out}(t)$ is a function that describes the temperature evolution at the final point of the circuit and $T_{H_20}^{in}(t)$ is another function that describe the same quantity at the initial point of the circuit. The integral is computed during the casting time, defined as T_{start}-T_{end}. This key feature works in background: it will be further manipulated to develop the algorithms providing, in the CBM tool, the data analytics to achieve the proxy for the asset risk of failure.

In particular, progressive degradation, that may lead to the selected top event, is monitored thanks to the deployment of a State Detection (SD) algorithm. The panel thermal state is taken under control by checking if the computed KF are compliant with some target/normal values. The control is performed applying statistical analysis, namely SPC. Statistical analysis is chosen in this case to detect if a different condition is present or not, based on some available condition monitoring data, since we have no availability of historical data of the asset, describing the degradation pattern. The idea is to measure the deviation of the current feature from a reference value representing the normal condition, to see whether the current KF is within the control limit or not. The goal of the SD model is the identification of critical scenarios, i.e. castings with value of KF out of the upper control limit. Indeed, a high value of KF means high thermal load and, in turn, a high risky scenario, since the thermal load is stressing out the panels.

The first step for the SD implementation is to build a frequency histogram, to visualize the computed KF for all the different castings, during a time-frame equal to 2 months. This temporal window allows to ensure a statistically significant amount of data and it is taken as reference/training period, since the EAF is considered working in normal conditions (stress-free). Then, to set the control limits, these reference data are used to compute the mean μ_{ref} and the standard deviation σ_{ref} of the collected KF sample. A data pre-processing phase is necessary to properly clean the data, in particular the data set is depurated by outliers. The clean dataset is then used to update the mean μ'_{ref} and the standard deviation σ'_{ref} and to define the upper threshold of the control limits, fixed as $UP_{ref} = \mu'_{ref} + 3\sigma'_{ref}$. Once set the SPC upper limit, the SD algorithm can run in real-time. The end-user, i.e. the Tenaris Dalmine technician, chooses the period of analysis (e.g. one week) and the SD tool checks all the KF values in this period. If the single value exceeds the upper threshold UP_{ref}, a Critical Events Counter (CEC) will be incremented by a unit. The CEC is representative of the number of events considered critical for the panel (the event/the casting is stressing out the panel, i.e. highly critical). An alarm is then displayed on a dashboard when the counter overcomes a defined threshold TH, being a high CEC representative of a highly stressing condition for the panel, that might cause a progressive degradation mechanism. The TH threshold is fixed not only justifying it with reasons due to thermo-mechanical notions, but moreover exploiting the company risk aversion attitude.

All the main steps of the SD algorithms, both for the SPC setting and the real-time execution, are reported in Figure 3.

SD algorithm- set up of SPC

1. Choose the reference period for data collection: 2months.
2. Compute the value of KF (equation (1)) for each casting during the reference period.
3. Compute the mean μ_{ref} and the standard deviation σ_{ref} for KF in the reference period.
4. Clean the dataset excluding from the sample values of:
   - KF > (μ_{ref} + 3 σ_{ref}) (Outlier)
   - KF < (μ_{ref} - 3 σ_{ref}) (Outlier)
   - KF < 0 (inconsistent value)
5. Compute again the mean and standard deviation of KF in the reference period without outliers, obtaining updated μ'_{ref} and σ'_{ref}
6. Fix the upper reference threshold as $UP_{ref} = \mu'_{ref} + 3\sigma'_{ref}$

SD algorithm- Real-time execution phase

1. Choose the real-time execution phase: 1 week
2. Compute the value of KF for each casting during the execution period.
3. If KF > $\mu'_{ref} + 3\sigma'_{ref}$ increment the critical event counter CEC=CEC+1.
4. If CEC > TH, an alarm is displayed on the dashboard.

Figure 3. SD algorithm – set-up and real-time execution

The SD algorithm is complemented by a HA algorithm. In fact, the SD algorithm aims at identifying a proxy of the failure risk by tracing an accumulation of critical scenarios (represented by the CEC indicator), as specific events – particular castings – causing highly critical thermal load, stressing out the panels. Still, the progressive degradation is not fully characterized: the energy is continuously absorbed by the panels. This is actually considered by the HA algorithm, that complements the information provided by the SD with further information to fully describe and analyze the progressive degradation over time.

The construction of the HA algorithm is based on ND techniques. According to [41], ND can be defined as the task of recognizing that data differ in some respect from data that are available during training normal conditions. In the analyzed industrial case, we deal with examples and collection of data coming only from “normal” condition and we do not have available data that describe “abnormalities” or “degraded” conditions, or RTF data. This is due to the fact that data related to “abnormal” conditions were not registered in the past and, moreover, because in some risky scenarios the plant was turned off, in order to avoid catastrophic effects. A solution to this problem is offered by ND, in which description of normality is built defining a model starting from a sample of normal system behavior. Previously unseen patterns are then recognized, comparing them with the model of normality and assigning a novelty score. This score is then compared with a fixed threshold: if the threshold has been overcome, the new situation is labelled as abnormal. As reported in [41], different kinds of categories of ND exist; in this work we refer to probabilistic ND techniques [44, 45]. These approaches are based on the estimation of a probability density function that fits the normal collected data. The resultant distribution is then used to define a boundary of normality and a proper score. A threshold has to be fixed to test whether a sample belongs to the same distribution or not. In our work, starting from the same KF of the SD model and following a parametric approach [46], a Gaussian distribution is fitted on the reference/training normal data. The Gaussian distribution has been selected as the most proper distribution looking at both the shape of the collected data (suggested by the histogram plot) and from literature reference [41]. Once modelled the normal condition, a new sample of real time data is collected and used to fit an actual Gaussian distribution. Once obtained this new actual distributions, it is possible to sketch on a graph two curves: the first one as reference distribution, coming from data of normal operating conditions of the panel; the second one as distribution representative of the actual asset health status due to the operating conditions. The score of the ND technique is then determined through the measure of the overlapping area between the two Gaussian distributions. This overlapping is measuring the deviation from the normal condition. The novelty threshold is fixed equal to the 70% of overlapping. It is worthy to notice that this threshold on the level of the overlapping is fixed not only justifying it with reasons due to thermo-mechanical notions, but moreover exploiting the company risk aversion. In addition, the HA algorithm also compares the value of the KFs' average in the actual period of analysis with its reference averages (computed during the reference/training period), being the maximum difference an empirical-plant value, derived from previous studies performed by the company and fixed equal to 4000 units (the unit of measure is confidential due to company restrictions). This further score of ND is actually measuring the distance from the "normal" central position. All the steps that defines the HA algorithm, both for setting the normality description and for the real-time execution are reported in Figure 4.

HA algorithm- setup of normality

1. Choose the reference period for data collection: 2 months.
2. Implement all the steps of SD algorithm.
3. Once clean the dataset, plot an histogram of collected KF (equation (1)) during the reference period. Choose a fixed number n of histogram intervals (along x axis) setting:
   - n -1 intervals of width 2000 (0:2000:100000);
   - The last n-th interval from 100000 to KF_{max} value (100000: KF_{max}), where KF_{max} is the maximum KF value. If KF_{max} < 100000 last interval is from 100000 to 102000.
4. Obtain for each n-interval the value of the probability density function pdf_{ref}^{n} in the central position of the interval as:

$$pdf_{ref}^{n} = \left(\frac{1}{\sigma'_{ref}\sqrt{2\pi}}\right) e^{\frac{-(x-\mu'_{ref})^2}{2\sigma'^2_{ref}}}$$

   where x is the central position of the n-th interval in which you are evaluating the pdf.

5. Compute for each n-interval the percentage of Reference Gaussian Area as:

   $\%REF_n = \frac{pdf_{ref}^{n}}{cumRef}$, where $cumRef = \sum_1^n pdf_{ref}^{n}$

HA algorithm- real-time execution

1. Set a real-time execution period of analysis (e.g. one week).
2. Compute the value of KF (equation (1)) for each casting and for each monitored panel.
3. Compute for each panel the actual mean μ_{actual} and the actual standard deviation σ_{actual} of KF for the execution period of analysis, in order to estimate the actual Gaussian distribution.
4. Obtain for each n-th interval (set as the same of the setup phase) the value of the probability density function pdf_{act}^{n} in the central position of the interval, computed as

$$pdf_{act}^{n} = \left(\frac{1}{\sigma_{act}\sqrt{2\pi}}\right) e^{\frac{-(x-\mu_{act})^2}{2\sigma^2_{act}}},$$

   where x is the central position of the n-th interval in which you are calculating the pdf.

5. Compute for each n-th interval the percentage of Actual Gaussian Area as:

   $\%ACT_n = \frac{pdf_{act}^{n}}{cumAct}$, where $cumAct = \sum_1^n pdf_{act}^{n}$
6. Compute the score of similarity ($\%SIM_n$), for each n-th interval. This value is the minimum between $\%REF_n$ and $\%ACT_n$:

   $\%SIM_n = \min(\%REF_n; \%ACT_n)$;
7. Compute the $Overlap_{area} = \sum_1^n \%SIM_n$.
   - If $Overlap_{area}$ <70 % and $\mu_{actual} - \mu'_{ref} \leq 4000$, then Level 2
   - If $Overlap_{area}$ <70 % and $\mu_{actual} - \mu'_{ref} > 4000$, then Level 4
   - If $Overlap_{area} \geq 70$ % and $\mu_{actual} - \mu'_{ref} \leq 4000$, then Level 1
   - If $Overlap_{area} \geq 70$ % and $\mu_{actual} - \mu'_{ref} > 4000$, then Level 3

Figure 4. HA algorithm – set-up and real-time execution.

The combination of the two novelty scores and their thresholds composes a matrix built with four quadrants, each of them corresponding to a precise critical scenario/alert level (Fig. 5).

	$Overlap_{area} \geq 70$ %	**$Overlap_{area} < 70$**
$\mu_{actual} - \mu'_{ref} \leq 4000$	Level 1	Level 2
$\mu_{actual} - \mu'_{ref} > 4000$	Level 3	Level 4

Figure 5. Matrix of Alert Levels.

- *Level 1* : the overlapping area is more than 70% and the difference between the averages of KF is lower than 4000. This means that the actual distribution is really similar to the reference one. The panel is stressed as usual (normal condition), so the risky level is low and the actual maintenance policy is appropriated.

- *Level 2*: the overlapping area is less than 70% and the difference between the averages of KF is lower than 4000. In this scenario and, more precisely, when the difference between the averages of KF is assuming a negative value, the actual process is different from the reference process but since the actual feature is moving to the left (i.e. difference between averages could be negative, so KF values are in average smaller than the usual ones), the process seems to stress the panel *less* than the normal. The risky level is very low and the actual maintenance policy is appropriated. Better, the planner cannot exclude that the frequency of inspections can be reduced.

- *Level 3:* the overlapping area is more than 70% and the difference between the averages of KF is more than 4000. In this scenario the actual process is closer to the reference process but, since the actual feature is moving to the right (i.e. difference between averages is more than 4000, so actual KF values are in average bigger than the reference ones), the process seems to stress the panel *more* than the normal situation. The risky level is increasing and the actual maintenance policy could be no longer appropriated: an alert is generated, that recommends to increase the frequency of inspections to control any possible degradation of the asset.

- *Level 4*: the overlapping area is less than 70% and the difference between the averages of KF is more than 4000. This is the worst scenario, in which the actual process is really different from the reference process and the actual feature is dramatically moving to the right (i.e. difference between averages is more than 4000, so KF values are in average bigger than the usual ones). The process is stressing the panel *much more* than the normal case. The risky level is high and the actual maintenance policy is no longer appropriated. An alarm is generated, that recommends as soon as possible to inspect the asset in order to avoid any possible damage.

The asset health status is in this way directly connected with the risk assessment and the alert generation, that represents the final outcome of both the two algorithms, as reported in Figure 6. The alert generated by the SD algorithm is a proxy of the failure risk: once this alarm is triggered by the CEC, the progressive degradation is fully characterized by the HA algorithm, which provides complete information about the kind of degradation over time and the maintenance policy that should be recommended. In this sense, it is straightforward to conclude which are the scores and the indicators that should be visualized on a proper CBM tool dashboard, for the risk assessment/visualization/mitigation and for the maintenance action suggestion. The dashboard has to show the number of critical events (that is, the CEC) in the period of analysis and the trend of the CEC in a time-based chart, compared with the fixed alert threshold TH. Moreover, the dashboard has to show the overlapping area scores in the period of analysis and the difference between KF's averages (reference and actual). The combination of these two scores has to provide real time information about the Alert Level (i.e. Level 1, Level 2, Level 3, Level 4), following the rules previously described and displayed in Fig. 6.

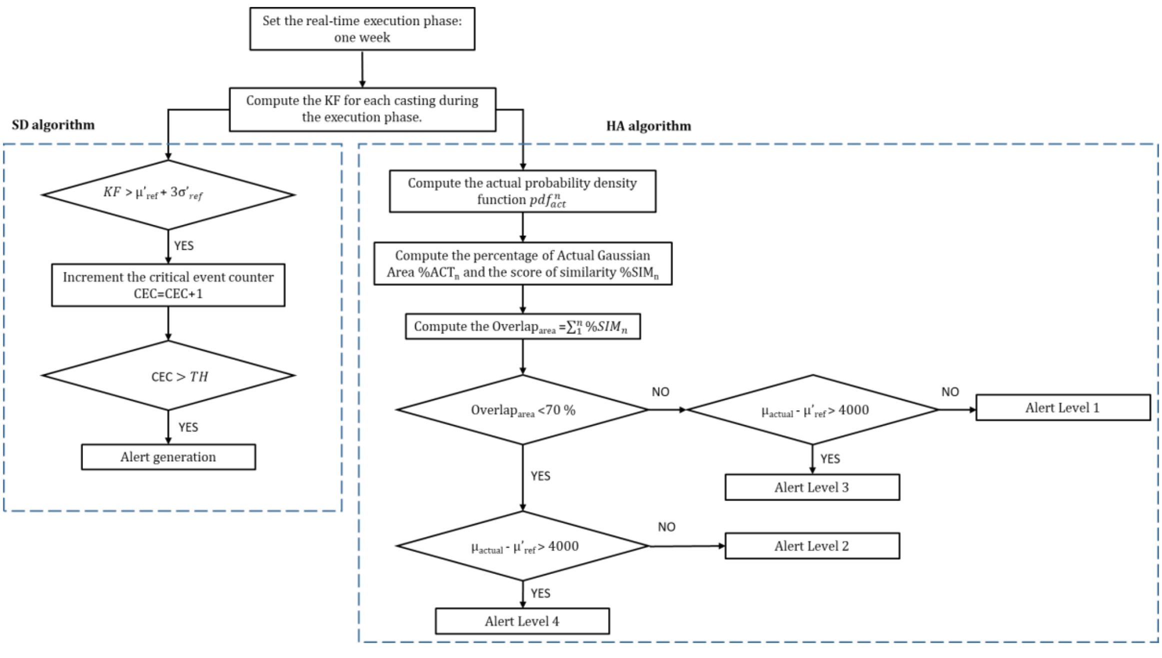


Figure 6. Flowchart of SD and HA algorithms – real time execution.

5. **Deployment of the SD and HA algorithms in the extant ICT architecture of the plant**

The SD and HA algorithms engineered in this work have been coded into a web application designed to be integrated in the information system of Tenaris Dalmine. The application is available to the process operator/maintenance engineer in the control room. In particular, the CBM tool is developed in such a way that is accessible as Level 2 tool in the ICT architecture, as reported in Figure 7.

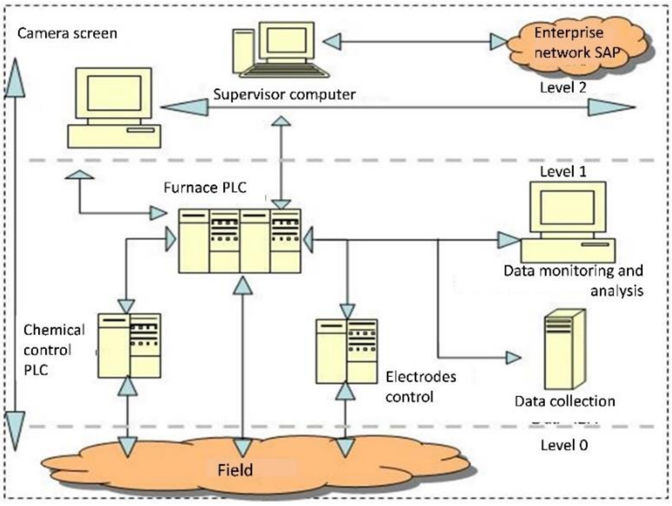
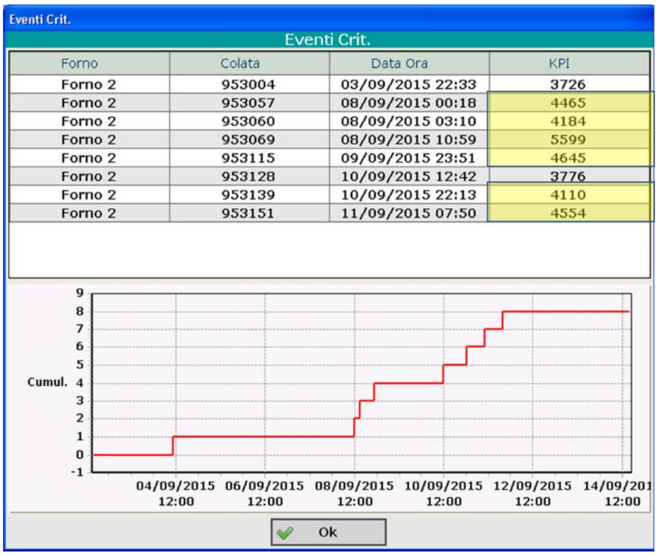


Figure 7. Architecture of Tenaris Dalmine extant ICT system

As web-based tool integrated with the suite used by plant operations in Tenaris Dalmine, it realizes the paradigm postulated since many years about integration of CBM solutions within the information system for plant operations and maintenance management [47]; indeed, a continuous monitoring, extending the extant automation with the presence of a "watchdog agent" processing field data in real time, enables the decision-maker to take risk-informed decisions while the furnace is running. Referring to Figure 7, integration is achieved combining the "traditional" automation featuring the supervision and control of the industrial process/asset, with some CPS elements developed through the web-based CBM tool: data management and computational capabilities are then provided at Level 2 in the ICT architecture to enhance intelligence in the industrial process/asset. The tool has been developed as highly parametrized one: it is possible to apply it also to other similar furnaces, even with differences in configurations (number of panels, setting parameters, etc.). It would be relevant for further developments to other EAFs, beyond Tenaris Dalmine.

In the following Figure 8 and 9, the Human Machine Interface (HMI) of the CBM tool is represented. It allows to check the number of critical events in the period of analysis, thanks to the CEC; it provides the casting number, the timestamp and the KF value for each casting. In the HMI, the KF is indicated as KPI, according to the name used in the plant. The trend of the CEC in the period of analysis is shown in a time-based chart, this is subject to the control of SD algorithm (see Figure 8). The HMI shows also the overlapping area indicator in the period of analysis – subject to the control of HA algorithm – and provides the information about the Alert Level (i.e. Level 1, Level 2, Level 3, Level 4, as explained in the previous section) (see Figure 9). The diagram in the bottom-right of Figure 9 also shows the average and standard deviation of KF during the heating cycles: the trend of this variable allows to know the operational mode of the furnace along time.



Eventi Crit.

Forno	Colata	Data Ora	KPI
Forno 2	953004	03/09/2015 22:33	3726
Forno 2	953057	08/09/2015 00:18	4465
Forno 2	953060	08/09/2015 03:10	4184
Forno 2	953069	08/09/2015 10:59	5599
Forno 2	953115	09/09/2015 23:51	4645
Forno 2	953128	10/09/2015 12:42	3776
Forno 2	953139	10/09/2015 22:13	4110
Forno 2	953151	11/09/2015 07:50	4554

Figure 8. HMI of the CBM tool, SD algorithm (in Italian, as the tool running in the company).

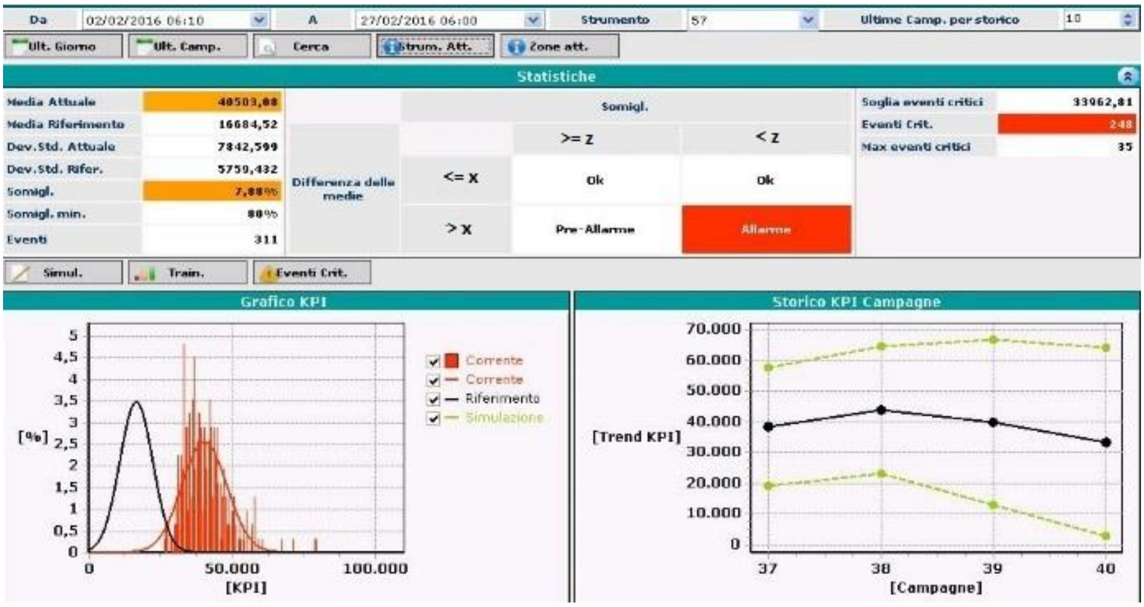


Figure 9. HMI of the CBM tool, HA algorithm (in Italian as the tool running in the company).

For a further description of the tool and its output information, two real cases are now presented, in which the CBM tool has been able to diagnose occurring risky scenarios.

In the first case, during a planned inspection, two holes of small dimensions were discovered in a panel. To properly evaluate this case, it is interesting to check the behavior of the SD algorithm. Figure 8 shows the register of the CEC with KF bigger than the upper threshold UP_{ref} = 4000 (see top of Fig. 8 in the column labeled KPI and highlighted in yellow color) in the days before the inspection, i.e. castings which have stressed out the panel. It is therefore evident that the panels appeared to be subject to stress in these days, and the evidence observed in the planned inspection was reasonably explained by such overstress.

In the second case, the panel under control seems to change its actual feature, showing a behavior different from normal operating conditions. Figure 9 shows the HA algorithm in which the actual Gaussian distribution of KF (red line) and the reference Gaussian of KF (black line) are compared. As the rate of overlapping area is lower and the difference between the average of the KF bigger than the respective thresholds, there is an alert level 4 (i.e. the 4th quadrant of the Matrix of Alert Levels, as reported in the red block of the figure). This is the worst scenario in which the actual process is different from the reference process, and the actual KF distribution is moving to the right: in this case, the process stresses and loads the panel more than the normal operating conditions. This arises attention and a clear message is sent to the operator/maintenance engineer in the control room.

6. Conclusion

Production process in steel making industry involves many variables and operators/engineers cope with the tasks of monitoring, controlling and diagnosing the health status of the production assets. This often results in high difficulties to effectively analyze the current assets' status, to detect and diagnose process anomalies and to anticipate risky scenarios. This eventually impacts on the capability to quickly take appropriate maintenance decisions and control actions. Moreover, in this industrial context, sensors cannot be installed due to technical constraints (harsh environmental conditions of the EAF).

Given all these assumptions, the solution proposed by this research represents a way to interact with and control the manufacturing asset to prevent any risky scenario. The model herein proposed formalizes and tests a data-driven CBM tool enabling an on-line control to support risk-informed decision-making. This tool relies on a process of knowledge discovery that incorporates both prior knowledge on the physical asset and proper interpretation of data analytics results. In a nutshell, the data-driven CBM tool builds fault diagnostics transforming raw data from the shop-floor into information, to enable both maintenance decision-making and risk assessment. Prior knowledge is fundamental to provide evidence of critical scenarios, to describe how the asset under study can fail and, last but not least, to help identifying the key measurable parameters to monitor and map the asset behavior.

Combining prior knowledge and data analytics, a Cyber-Physical integration has been conducted in the brown field, relying on the extant plant automation, and running beside the "traditional" control. Indeed, computational capabilities have been enhanced enabling a “watchdog agent” of risky scenarios, able of performing real-time monitoring and data analytics and allowing the on-line risk-assessment of safety-critical components.

Today, the data-driven CBM tool is running in the plant and the future aim is to correlate the main tool indicators (i.e. the CEC – critical events counter - and the overlapping areas) to the remaining useful life (RUL) of the panel, developing in this way also a prognostic phase. To this end, it is necessary to continue with the data collection to characterize the life of the panels up to their replacement/repair. The tracking of the panels life will then serve to map the indicators information (again the CEC and the overlapping areas) with the actual degradation occurred/observed in the panel. The final aim will be the implementation of an acceptable degradation function of the panels. This would be the basis for the definition of a reliability model of the panels themselves – to express the probability not to degrade before a certain time *t* –, thus for the computation of the subsequent risks. On the whole, it could be eventually possible to obtain (by a computational model) also a prognostic-risk prediction.

A further extension of the data-driven CBM tool could be built by reflecting on the vision of the social network of machines, suggested by [8], to provide EAFs with self-comparison ability. This will be useful for performance rating among the fleet owned by Tenaris Dalmine, and for a similarity identification between the performance and assets operating condition, eventually useful to develop a powerful predictive analytics of the RUL based on the different operating conditions. On the scientific level, it would enable continuing the implementation path towards building CPS in production.

Last but not least, dynamic scheduling could be evaluated and optimized, building on the capabilities due to on-line risk-assessment and a further optimization of maintenance interventions, with the purpose to find a proper balance between the frequency of inspections – done when the furnace is stopped, according to the current practice – and the continuous monitoring while the furnace is running. This may help to optimize the maintenance efficiency while respecting risk constraints.

References

1. Lee J, Jin C, Liu, Zongchang and Davari Ardakani H (2017) Introduction to Data-Driven Methodologies for Prognostics and Health Management. In: Ekwaro-Osire S., Gonçalves A. AF (ed) Probabilistic Prognostics and Health Management of Energy Systems. Springer, Cham, pp 9–32

2. Shrouf F, Ordieres J, Miragliotta G (2014) Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm. IEEE Int Conf Ind Eng Eng Manag 2015-Janua:697–701 . doi: 10.1109/IEEM.2014.7058728

3. Negri E, Fumagalli L, Macchi M (2017) A Review of the Roles of Digital Twin in CPS-based Production Systems. Procedia Manuf 11:939–948 . doi: 10.1016/j.promfg.2017.07.198

4. Sarma S, Brock D, Ashton K (2000) The networked physical world. Auto-ID Cent White Pap MIT- 1–16

5. Ashton K (2009) That “Internet of Things” Thing. RFiD J 4986 . doi: 10.1145/2967977

6. Baheti R, Gill H (2011) Cyber-physical Systems. Impact Control Technol 161--166 . doi: 10.1145/1795194.1795205

7. Jazdi N (2014) Cyber physical systems in the context of Industry 4.0. 2014 IEEE Autom Qual Testing, Robot 2–4 . doi: 10.1109/AQTR.2014.6857843

8. Lee J, Bagheri B, Kao HA (2015) A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. Manuf Lett 3:18–23 . doi: 10.1016/j.mfglet.2014.12.001

9. Kusiak A (2006) Data mining: Manufacturing and service applications. Int J Prod Res 44:4175–4191 . doi: 10.1080/00207540600632216

10. Kuo Y-H, Kusiak A (2018) From data to big data in production research: the past and future trends. Int J Prod Res 7543:1–26 . doi: 10.1080/00207543.2018.1443230

11. Cattaneo L, Fumagalli L, Macchi M, Negri E (2018) Clarifying Data Analytics Concepts for Industrial Engineering. IFAC-PapersOnLine 51:820–825 . doi: 10.1016/j.ifacol.2018.08.440

12. Ge Z, Song Z, Ding SX, Huang B (2017) Data Mining and Analytics in the Process Industry: The Role of Machine Learning. IEEE Access 5:20590–20616 . doi: 10.1109/ACCESS.2017.2756872

13. Djurdjanovic D, Lee J, Ni J (2003) Watchdog agent - An infotronics-based prognostics approach for product performance degradation assessment and prediction. Adv Eng Informatics 17:109–125 . doi: 10.1016/j.aei.2004.07.005

14. Colace C, Fumagalli L, Pala S, Macchi M, Matarazzo NR, Rondi M (2013) An Intelligent Maintenance System to Improve Safety of Operations of an Electric Furnace in the Steel Making Industry. 2013 Progn Heal Manag Conf 33:397–402 . doi: 10.3303/cet1333067

15. Colace C, Fumagalli L, Pala S, Macchi M, Matarazzo NR, Rondi M (2015) Implementation of a condition monitoring system on an electric arc furnace through a risk-based methodology. Proc Inst Mech Eng Part O J Risk Reliab 229:327–342 . doi: 10.1177/1748006X15576441

16. Fumagalli L, Macchi M, Colace C, Rondi M, Alfieri A (2016) A Smart Maintenance tool for a safe Electric Arc Furnace. IFAC-PapersOnLine 49:19–24 . doi: 10.1016/j.ifacol.2016.12.155

17. Heng A, Tan ACC, Mathew J, Montgomery N, Banjevic D, Jardine AKS (2009) Intelligent condition-based prediction of machinery reliability. Mech Syst Signal Process 23:1600–1614 . doi: 10.1016/J.YMSSP.2008.12.006

18. Niu G, Yang B-S, Pecht M (2010) Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance. Reliab Eng Syst Saf 95:786–796 . doi: 10.1016/J.RESS.2010.02.016

19. Arunraj NS, Maiti J (2010) Risk-based maintenance policy selection using AHP and goal programming. Saf Sci 48:238–247 . doi: 10.1016/j.ssci.2009.09.005

20. Marseguerra M, Zio E, Podofillini L (2002) Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation. Reliab Eng Syst Saf 77:151–165 . doi: 10.1016/S0951-8320(02)00043-1

21. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20:1483–1510 . doi: 10.1016/j.ymssp.2005.09.012

22. Lee J, Ni J, Djurdjanovic D, Qiu H, Liao H (2006) Intelligent prognostics tools and e-maintenance. Comput Ind 57:476–489 . doi: 10.1016/j.compind.2006.02.014

23. Deloux E, Castanier B, Bérenguer C (2008) Maintenance policy for a deteriorating system evolving in a stressful environment. Proc Inst Mech Eng Part O J Risk Reliab 222:613–622 . doi: 10.1243/1748006XJRR148

24. Huynh KT, Barros A, Bérenguer C (2012) Adaptive condition-based maintenance decision framework for deteriorating systems operating under variable environment and uncertain condition monitoring. Proc Inst Mech Eng Part O J Risk Reliab 226:602–623 . doi: 10.1177/1748006X12465718

25. Guillén AJ, Crespo A, Macchi M, Gómez J (2016) On the role of Prognostics and Health Management in advanced maintenance systems. Prod Plan Control 27:991–1004 . doi: 10.1080/09537287.2016.1171920

26. Bengtsson M (2004) Standardization issues in condition based maintenance. Proc 16th Int Congr 53

27. ISO/IEC (2003) ISO 13374 - 1 - Condition monitoring and diagnostics of machines -- Data processing, communication and presentation -- Part 1: General guidelines

28. Lei Y, Li N, Guo L, Li N, Yan T, Lin J (2018) Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech Syst Signal Process 104:799–834 . doi: 10.1016/j.ymssp.2017.11.016

29. Dragomir OE, Gouriveau R, Dragomir F, Minca E, Zerhouni N (2018) Review of prognostic problem in condition-based maintenance. 2009 Eur Control Conf 1587–1592 . doi: 10.23919/ecc.2009.7074633

30. Fugate ML, Sohn H, Farrar CR (2001) Vibration-based damage detection using statistical process control. Mech Syst Signal Process 15:707–721 . doi: 10.1006/mssp.2000.1323

31. Skormin VA, Popyack LJ, Gorodetski VI, Araiza ML, Michel JD (2008) Applications of cluster analysis in diagnostics-related problems. 161–168 vol.3 . doi: 10.1109/aero.1999.789775

32. Staszewski WJ, Worden K, Tomlinson GR (1997) TIME–FREQUENCY ANALYSIS IN GEARBOX FAULT DETECTION USING THE WIGNER–VILLE DISTRIBUTION AND PATTERN RECOGNITION. Mech Syst Signal Process 11:673–692 . doi: 10.1006/MSSP.1997.0102

33. Goumas SK, Zervakis ME, Stavrakakis GS (2002) Classification of washing machines vibration signals using

discrete wavelet analysis for feature extraction. IEEE Trans Instrum Meas 51:497–508 . doi: 10.1109/TIM.2002.1017721

34. A.K. Jain, R.P.W. Duin, J. Mao (2000) Statistical Pattern Recognition: A Review. IEEE Trans Pattern Anal Mach Intell 22:4–37

35. Sun Q, Chen P, Zhang D, Xi F (2004) Pattern Recognition for Automatic Machinery Fault Diagnosis. J Vib Acoust 126:307 . doi: 10.1115/1.1687391

36. Mechefske CK, Mathew J (1992) Fault detection and diagnosis in low speed rolling element bearings Part II: The use of nearest neighbour classification. Mech Syst Signal Process 6:309–316 . doi: 10.1016/0888-3270(92)90033-F

37. Hastie T, Tibshirani R (2009) The Elements of Statistical Learning - Data Mining, Inference and Prediction

38. Accorsi R, Manzini R, Pascarella P, Patella M, Sassi S (2017) Data Mining and Machine Learning for Condition-based Maintenance. Procedia Manuf 11:1153–1161 . doi: 10.1016/J.PROMFG.2017.07.239

39. Poyhonen S, Jover P, Hyotyniemi H (2004) Signal processing of vibrations for condition monitoring of an induction motor. 499–502 . doi: 10.1109/isccsp.2004.1296338

40. Qingfeng W, Wenbin L, Xin Z, Jianfeng Y, Qingbin Y (2011) Development and application of equipment maintenance and safety integrity management system. J Loss Prev Process Ind 24:321–332 . doi: 10.1016/J.JLP.2011.01.008

41. Pimentel MAF, Clifton DA, Clifton L, Tarassenko L (2014) A review of novelty detection. Signal Processing 99:215–249 . doi: 10.1016/j.sigpro.2013.12.026

42. Cubillo A, Perinpanayagam S, Esperon-Miguez M (2016) A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. Adv Mech Eng 8:1–21 . doi: 10.1177/1687814016664660

43. Ferro L, Giugliano P, Galbiati P, Memoli, Francesco Giavani C, Maiolo J (2007) The electric arc furnace of Tenaris Dalmine: from the application of the new technologies of digital electrode regulation and multipoint injection to the dynamic control of the process. In: METEC, METEC International Metallurgy Trade Fair with Congress, 2007, InSteelCon, International Steel Conference on new Developments in Metallurgical Process Technologies, 3. Stahlinstitut VDEh, pp 504–512

44. Paalanen P, Kamarainen J-K, Ilonen J, Kälviäinen H (2006) Feature representation and discrimination based on Gaussian mixture model probability densities—Practices and algorithms. Pattern Recognit 39:1346–1358 . doi: 10.1016/J.PATCOG.2006.01.005

45. Filev DP, Tseng F (2006) Real time novelty detection modeling for machine health prognostics. Annu Conf North Am Fuzzy Inf Process Soc - NAFIPS 529–534 . doi: 10.1109/NAFIPS.2006.365465

46. Markou M, Singh S (2003) Novelty detection: a review—part 1: statistical approaches. Signal Processing 83:2481–2497 . doi: 10.1016/J.SIGPRO.2003.07.018

47. Fumagalli L, Macchi M, Rapaccini M (2009) Computerized maintenance management systems in SMEs: A survey in Italy and some remarks for the implementation of Condition Based Maintenance. IFAC