TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 18
Number 3
September 2024
DOI: 10.12716/1001.18.03.14

Degradation Data Self-Analysis Layer for Integrated Maintenance Activities

J. Szpytko & Y. Salgado Duarte
AGH University of Krakow, Kraków, Poland

ABSTRACT: Reliability-oriented approach based on Monte Carlo simulations is a well-established methodology for coordinating maintenance activities of any technical system. Usually, coordination is conducted using holistic performance indicators, which are obtained from the convolution between the stochastic system availability and the system service required in a time horizon of *t*. Specifically, the system stochastic availability modeling is composed of the degradation process due to the system operation and the planning of the maintenance activities needed to keep the system operating at the desired standards. In the case of the degradation modeling process, given its random nature, it is addressed with predictions, which in practice, consist of generating random samples of the stochastic degradation processes from probability distributions, and the parameterization is usually estimated by fitting the distributions to historical degradation data for each technical component considered. Crucial to forecasting accurate performance indicators is the use of up-to-date information, i.e., the self-update of historical degradation data. In this paper, to address accurate performance indicators, we propose using the machine learning approach to update the adaptable model layers affected by changes in the degradation data. The paper's case study is an overhead crane system of a hot rolling mill process in a steel plant, which operates under hazardous conditions and continuously. We focus on overhead cranes because they are critical components of production processes. The paper's subject is validating the performance of a self-analysis layer, which processes the degradation data of the analyzed technical devices. The engineering solution ensures well-processed inputs for the problem of coordination of maintenance activities of overhead cranes, which is the object of the study of this research.

1 INTRODUCTION

In recent decades, many fields of knowledge have implemented data collection techniques and data analysis methods. Among many, machine learning is one of the most widely applied methods for data self-analysis. Today, machine learning has concrete examples in many engineering systems and intuitively merges concepts such as cyber-physical systems and cognitive computing into complex and smart platforms. The previous statement is, in some ways, the result of advances in the levels of data storage and data acquisition in engineering systems. In today's industry, it is common to find physical engineering systems with hundreds of sensors that collect data on the performance of their operational functions. Therefore, it is expected to find hundreds of contributions and applications that use the information generated to improve the current state of engineering systems, qualitatively or quantitatively.

As an example of engineering systems, cooperative overhead cranes are critical devices in many continuous production industries. They are usually installed in processes with hazardous conditions and difficult access. Each overhead crane manages a section of the ongoing process, is fixed on the warehouse roof, its movements are limited to a specific working range, and its criticality comes from the sometimes-unexpected unavailability of an overhead crane that can stop an entire production process.

Today, these cranes are changing their design due to the new needs of the current industry. For example, the new generation of overhead cranes is equipped with sensors to collect information. However, the data generated by sensor systems are often not exploited adequately. When talking about cooperative overhead cranes, the two main research fields studied are load operation control, such as the examples [9], [11], and [14], and maintenance, such as the examples [18] and [19]. The reason researchers focus their attention on these fields is simple: incorrect load control and weak maintenance are the most common causes of overhead crane failures in today's industry. Some examples of studies that analyze the root cause of crane failures are [3], [10], and [15]. Even when cranes with different working conditions and functions are analyzed in the cited contributions, as a connection between them, we find that weak maintenance cycles and wrong load control are the causes of crane failures.

Focusing on maintenance strategies, a well-established approach to designing suitable maintenance strategies is reliability analysis, as it allows us to measure the risk of possible failures in overhead crane systems and incorporate and evaluate potential risk scenarios for the system. Examples of contributions in this field are [2] and [8]. However, a weakness of the reliability analysis approach is the reliance on data to predict realistic scenarios. In this paper, we propose to integrate reliability-based maintenance coordination and data analysis methods by offering an engineering solution for a cooperative overhead crane system operating in a steel plant, which is embedded in an integrated digital platform, and somehow manage to address the weakness of reliability analysis.

The research conducted here arises from a local request from the maintenance department of a steel plant with organizational issues. Although it is a local solution, the achievements provide a practical example of how a reliability model can be tailored to address a local solution using data generated in the daily work of the maintenance department. The idea presented in this paper aims to propose, through a practical example, how an existing overhead crane system can be adapted to the digital era and contribute to other data-driven applications. Although the proposed model is its broadest conception, it is an oriented engineering solution for coordinating maintenance activities; the paper's subject is the Degradation Data Self-Analysis (DDSA) layer, which ensures the accurate and robust treatment of degradation data.

Focusing on the performance of this layer has a well-justified reason. For instance, the layer avoids human intervention in online filtering and estimation of the degradation-related parameters of the reliability-oriented optimization model. In addition, the layer ensures a comprehensive mathematical and technical connection between the degradation data due to the system operation and the optimization model parameters, considerably decreasing errors while ensuring that up-to-date information is always used when running the model. In addition, the selected frequency models, outputs of the layer, allow one to prolong in time the degradation due to the operation of the system, which in turn provides for evaluation with scenarios of how the planning process will work in the maintenance department of the steel company when unexpected failures are considered. That said, the criticality of this layer in estimating the performance indicator, a variable that holistically measures the quality of the maintenance coordination process, is evident.

The DDSA layer is the process of filtering, processing, and storing the degradation data of the proposed engineering solution. The source of raw degradation data involved in this process comes from two systems used in the daily work of the steel company: SAP (Systems, Applications & Products in Data Processing) and SCADA (Supervisory Control and Data Acquisition). The proposed paper is an extended and more in-depth version of the work presented by [18]. In this paper, we test and validate the DDSA layer, which fills the decision gaps presented in previous work. Moreover, with the validation, we support the conclusions for a specific scenario and all scenarios analyzed afterward.

Specifically, in this paper, we extend and expose the functional connection between the system components, making clear the relevance of the DDSA layer in the developed model and the algorithm implemented to execute the fitting process. In addition, we isolate, test, and show the process of fitting simple distributions in practice with actual data. For illustrative purposes, the selected example and the described data come from a specific 50-ton overhead crane with 59,501 hours of continuous operation (6.8 years) and 355 hours of the replacement or repair process. Also, during the validation, we include the analysis of the impacts introduced by the changes in the data, which consists of changing the input data in a controlled way and re-evaluating the fitting selection process. Mainly two experiments are conducted; the first consists of contaminating the input data with new values generated from the final selections, and the second one consists of removing the last records of the dataset.

Once the scope of the paper has been declared, the remaining sections of the document are presented as follows: First, the context of degradation data is discussed, which highlights the value of the DDSA layer to ensure accurate and robust treatment of the degradation data. Then, using a 50 tons overhead crane as a case study, the DDSA layer is tested in practice. Finally, the conclusions highlight the main results of this contribution and the connections to future work.

2 MATERIALS AND METHODS

2.1 *Maintenance – Degradation Modelling*

In continuous-process productions, common indicators for monitoring the performance of a technical system, which provides a service or supports a production line process, are based on the system's availability or its components over time. The availability $A(t)$ of a technical system over time can be impacted by two main reasons: planned maintenance or unexpected failures. From the point of view of state diagrams, a technological system can have three possible states: available, unavailable due to unexpected failures, and unavailable due to planned maintenance.

In the case of maintenance activities, imagining a sequence in time, maintenance schedules can be represented by the variables M maintenance-start-time and D maintenance-duration-time, where $M = \{m_1, m_2, \ldots, m_k\}$, $D = \{d_1, d_2, \ldots, d_k\}$ and k number of maintenances. Both variables M and D are planned and can be selected according to maintenance strategies in the analyzed process.

In parallel, degradation can be treated similarly, knowing that degradation is inherent in a technical system. The degradation over time can be represented by the variables F time-to-failure and R time-to-repair, where $F = \{f_1, f_2, \ldots, f_n\}$, $R = \{r_1, r_2, \ldots, r_n\}$ and n is the number of failures. In particular, variables F and R are considered random variables. In the case of F, unexpected failures are related to component degradation due to system operation and are unpredictable in almost all cases. In the case of R, the time depends on the magnitude of the failure, the expertise of the workers who repair the failure, and the logistics behind it. In conclusion, it is also defined as a random phenomenon with certain thresholds. Since the F and R variables are defined as random, the system's availability $A(t)$ over time is defined as a stochastic process.

All the variables M, D, F, and R presented above are times and can be defined as non-negative variables. Maintenance scheduling is usually a planning process, i.e., depending on the planning window (monthly, quarterly, or annual), planners propose the sequence to be executed in the next window. Coordination of the maintenance process is crucial to ensure the life cycle of any system, and its optimization is a daily task in any technical system. It is not surprising that availability $A(t)$ plays an essential role in maintenance coordination and appears in several approaches used to coordinate this process, such as the examples of [6], [12], [7], [4], and [1]. In approaches in which the modeling of the availability $A(t)$ is included, it is also necessary to make decisions related to the random phenomenon, that is, the random variables F and R. The usual approach is to assume the progression of the degradation process in the planned window based on the historical degradation data of the analyzed system. The way to include degradation is to fit some model to the historical degradation data, resulting in one model representing F and another for R, and then, based on the fitted model, $\sim F$ potential failures and $\sim R$ repair times are simulated. The fitting and simulating processes are performed for each component of the analyzed system. The simulated values are convoluted with the proposed sequence of planned maintenance, which allows for modeling the system availability $A(t)$ and assessing the maintenance scheduled impact.

Focusing our attention on variables F and R and the modeling of the fitting process, the problem to be solved in this case is to find the best model to represent the historical data, which is then used to simulate a potential degradation process. The only known information is that we are dealing with non-negative variables, and the stochastic process $A(t)$ is continuous in time. Fitting models can be divided into non-parametric and parametric.

A contribution supporting this classification is [5], which tested the performance and covered convergence issues during the fitting process in both approaches. The research in the above contribution concludes an introductory statement that the convergence of the fitting process depends on the features of the data. They end that the non-parametric approach should be used when the data is dense; otherwise, parametric is the way to go. In our case, degradation data, the subject of study here, are usually sparse; even assumptions are sometimes needed for highly reliable systems. Therefore, given the features of the data, parametric models are the tentative choice for our problem. However, the definition of dense data is not well defined. The historical degradation data, represented in our case by F and R, are non-negative random variables. These recorded values can be assumed as samples of a generating function, which means the degradation process, and in which the order of the failures is irrelevant. Given the constraints and features of the degradation data, the F and R variables are usually modeled with frequency models. Knowing also that F and R are used to model, in our case, a continuous stochastic process, we further restrict the option space to continuous frequency models.

A comprehensive list of parametric frequency models is continuous probability distributions. It should be noted that parametric frequency models were used in all references cited above, which are related to the concept of $A(t)$ availability discussed. On the other hand, non-parametric ones are kernels, splines, neural networks, a simple frequency histogram, and the empirical distribution, which have been applied in [22] and [16]. The selection of the final approach and model to be used is discretionary and is guided by the individual viewpoint of the research conducted. Based on expertise in the field and the references cited, we consider parametric and continuous frequency models to be the models that have the most practical applications, are the most transparent, and consequently achieve the expected results.

Deciding on the set of models to be used to model the variables F and R does not end the discussion. A parametric model (as well as a non-parametric one) is estimated based on the available historical data, and it is well-known that the inference error decreases when the degradation data are more representative of the analyzed phenomenon. Consequently, when more degradation data are available, a good strategy is re-estimating the model's parameterization, always seeking to get closer to the actual phenomenon. Here, machine learning approaches play an essential role in the discussion. In current practice, integrating data monitoring and processing is an important goal. The main idea is to create smart data processing layers to update the models' parameterization based on the newly available data. Several contributions in this field, such as [17], [21], [13], and [20], have proposed practical applications, and the research presented in this paper is another contribution in the same direction.

Here, we propose a smart data-driven algorithm to find the most suitable parametric model for the degradation variables *F* and *R*. The algorithm is encapsulated in the DDSA layer introduced in the previous section, which runs online and is fully connected to the SAP-SCADA systems, which makes it possible to update the data when a scenario is run on the integrated digital platform. The entire DDSA layer tested and validated in this paper is encapsulated in functions implemented in MATrix LABoratory (MATLAB) that work without interaction with the model user, and we insist that the validation of their performance is important in this research.

Finally, before finishing this section, in section 2.2, we present the description of the algorithm implemented in MATLAB, which is used in the DDSA layer, summarizing the standardization efforts of the fitting process. In section 2.3, we contextualize the relevance of the fitting process in the integrated maintenance platform by presenting the functional modeling at the component and system levels, discussing, and describing the connection between them.

2.2 *Fitting Single Distributions Algorithm*

1. For each *i*-th overhead crane, the sequence of F_i and R_i is filtered from the SCADA-SAP systems by means of a unique identifier (ID). By construction, both random number vectors have the same length.
2. Independently, for each streamed sequence, the following steps are applied:
   - Given a random variable sample $\sim X = (x_1, x_2, \ldots, x_n)$, in our case, either the *F* or *R* random sequence filtered and a set of *k*-th predefined and preselected single continuous distributions, we apply for each *k*-th case the following steps:
   - Fit the data (either *F* or *R*) to the *k*-th single distribution by maximizing the log-likelihood function $\ln \mathcal{L}_n(x \mid \theta)$, i.e., the negative logarithm value of the product of the probability of the sample data (*X*), given the parameters θ of the distribution. If the fit does not converge for the given parametric distribution, the process ends; otherwise, it continues as follows:
     - Save the *k*-th index for ID purposes, the estimated negative log-likelihood value when the maximum likelihood estimation (MLE) method was applied (to record the process performance), and the parametrized single distribution structure.
     - II. Apply two goodness-of-fit tests to analyze the results of the fit: first, the one-sample non-parametric Kolmogorov-Smirnov (KS) test, defined as

       $D^* = \max\left(\left|\hat{F}(x) - F(x)\right|\right)$,

       where $\hat{F}(x)$ is the empirical cumulative distribution function of the data and *F(x)* is the cumulative distribution function of the fitted single parametrized distribution; and then the Anderson-Darling (AD) test, defined as

       $n\int_{-\infty}^{\infty}\left(\hat{F}(x) - F(x)\right)^2 w(x)\,dF(x)$,

       where *n* is the number of data points in the sample, $\hat{F}(x)$ and *F(x)* as described above, and *w(x)* is a weight function defined as

       $w(x) = \left[F(x)\left(1 - F(x)\right)\right]^{-1}$.

       In the case of the AD test, the data is the ordered sample.
     - Apply the Akaike information criterion (AIC) defined as

       $-2\log L\left(\hat{\theta}\right) + 2k$,

       where log $L(\hat{\theta})$ denotes the optimal log-likelihood objective function value, and *k* is the number of parameters of the single-fit distribution.
     - Estimate the parametrized fit distribution structure's theoretical mean μ and variance σ^2.
     - Store the frequency of the data (number of records) and the sum of the data (in the case of variable *F*, we are storing the hours of operation).
     - Add a logical check value following the rule: if the *k*-th fitted distribution has finite μ and σ^2, $Flag_k = 1$; otherwise, $Flag_k = 0$.
   - Reject distributions with $Flag_k = 0$ (fits with infinite mean or variance).
   - If the data records are less than ten, the best fit is the exponential distribution. Otherwise, the best fit is the *k*-th distribution with minimum AIC value.
3. End of the fitting process algorithm.

2.3 *Functional Modelling of the Integrated Maintenance Platform*

Looking at contextualizing the integrated digital platform, we can say that the platform is adapted to support online maintenance activities coordination. In the system analyzed, in our case, a set of overhead cranes in a selected manufacturing industry, the maintenance department, which uses the digital platform for planning purposes, has risk circumstances when unanticipated failures occur during scheduled maintenance activities.

The platform was created to reduce the overworking days in the maintenance department by minimizing the convolution between scheduled maintenance activities and unanticipated failures. The impact of risk situations exists because the set of overhead cranes is critical for moving demanding loads on the production line of the manufacturing industry. This undesirable situation stresses workers because they work under pressure during the interaction between scheduled maintenance activities and unanticipated failures.

The integrated digital platform comprises three blocks of self-analysis: data filtering and synthesis, model simulation through scenario evaluation, and optimization layer to coordinate maintenance activities. Each block is supported by independent algorithms implemented in MATLAB. Also, the data sources are mainly two professional systems: SAP (Systems, Applications, and Products in Data Processing) and SCADA (Supervisory Control and Data Acquisition). Integrating all blocks with a dynamic window that visualizes the model performance conforms to the digital platform and connects all blocks through the integrated digital platform.

The data processing (filtering and synthesis) block ensures the existence of all the necessary parameters to evaluate the modeled scenario. Therefore, as we can see, each time a scenario is loaded for evaluation, the model parameterization is calibrated with the updated information.

In addition, the digital platform provides, for a given proposed scenario to be evaluated, the option of using an optimization algorithm (heuristic in our case) to find the best maintenance activities schedule for the scenario. In this case, the optimization algorithm minimizes the interaction between unexpected failures and the maintenance activities scheduled for the set of overhead cranes considered.

Having briefly contextualized the integrated digital platform, we focus on the first block, the data processing block, specifically the degradation data processing.

Previously, we introduced how the measurement of system availability *A(t)* enables coherent coordination of the maintenance process. Now, we intend to describe and deepen the integrated platform's functional modeling point of view. For this purpose, we divide the description into two levels, component, and system, making visually clear the inner workings of the digital platform. However, although the definition has been divided, everything is connected. Furthermore, it is necessary to emphasize that the platform operates without human intervention in data processing, and the platform user only interacts with it through the system-level settings of the parameters of the assessed scenario.

Fig. 1 shows the model composition of each component considered on the digital platform, in our case, overhead cranes.

The stochastic functional capacity of each overhead crane $z_1 = f(t \mid \theta_1)$ is composed of the convolution between the degradation process $C_D = f(t \mid \theta)$ and the planned maintenance process $C_M = f(t \mid \theta)$, where t is the time and θ in both cases is a set of parameters that depend on the function describing the underlying process, either degradation or maintenance.

In the case of the maintenance process, a deterministic concatenation of the maintenance lifecycles $M = f(t \mid \theta)$ and maintenance duration $D = f(t \mid \theta)$ composes the maintenance activity scheduling. Sometimes, the functional variables M and D use predictive models or are fixed standard times provided by the overhead crane manufacturer. In any of them, we deal with functions that depend on time t and certain parameters θ. Independently, but following the same idea, the degradation process is a probabilistic concatenation between the time-to-failures $F = f(t \mid \theta)$ and time-to-repair $R = f(t \mid \theta)$. This time, the variables are modeled with frequency models, i.e., probability distributions. At this point, this is where the connection and relevance of the DDSA layer are present. As we can see, the fitting process ensures a parametric distribution selection close to the shape of the actual data filtered from the SAP-SCADA systems. At this point of the description, the entry points of the integrated digital platform, which are: degradation data and planned maintenance data, are also evident.

Crucial for sensible coordination of maintenance activities is the starting point of the maintenance scheduling of the component, overhead cranes in our case, and the degradation modeling process. The starting point for maintenance scheduling is provided by the optimization model implemented at the system level when coordination is requested. The degradation modeling process is managed by the DDSA layer using the updated information available in the monitoring systems. In summary, we can say that the overhead crane capacity model is a stochastic Markov chain Monte Carlo (MCMC) process.

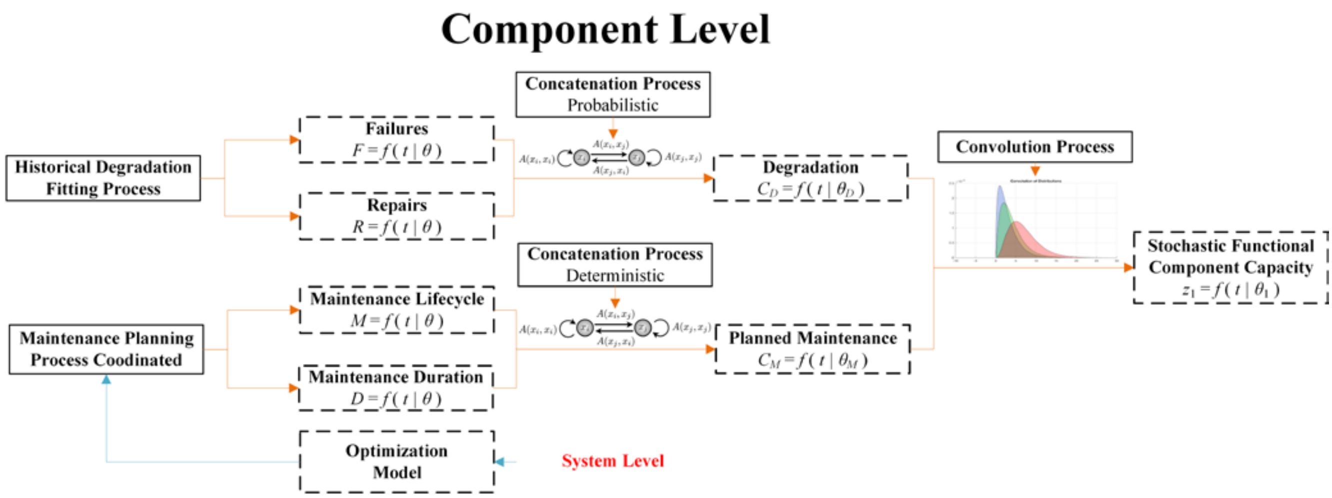

Figure 1. Functional view at the component level.

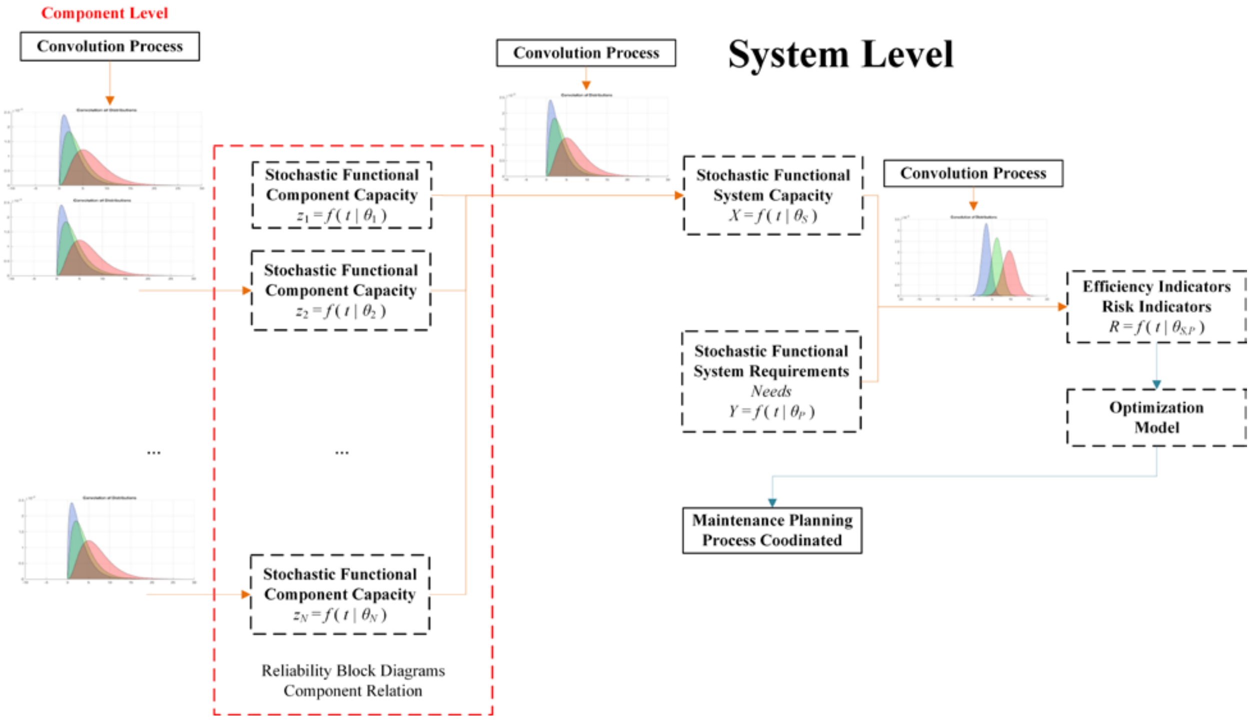

Figure 2. Functional view at the system level.

Once the number of overhead cranes involved in the system is known, the modeling steps described above are applied for each overhead crane independently $z_1 = f(t \mid \theta_1)$, $z_2 = f(t \mid \theta_2)$, …, $z_N = f(t \mid \theta_N)$.

Fig. 2 shows the system-level modeling composition.

Once the component-level modeling step is achieved for each crane, and given the relationship between them, i.e., the series-parallel block diagram, it is possible by convolution to build the stochastic functional system capacity $X = f(t \mid \theta_S)$, where again t is the time and θ_S is a set of parameters that depends on other parameters modeled in previous steps. Once system-level availability $A(t)$ is obtained, which in our case is a stochastic process $X = f(t \mid \theta_S)$ when Monte Carlo simulations are used, and knowing the function that describes the needs of the requested service $Y = f(t \mid \theta_P)$, again using a convolution process, we can estimate performance metrics $R = f(t \mid \theta_{S,P})$ that measure the efficiency and adequacy of the system providing the service, in our case how adequate the actual system is to fulfill the requested service. As we can deduce, if the metric performance measures the system's adequacy, then we can use the metric to find the best operating point for the system.

Knowing that we are modeling two contributors, degradation and maintenance and that one is random, referring to degradation, we can use this approach to find the best maintenance scheduling for the system such that the performance metric provides the lowest/highest possible value. In the end, the optimization model manages the starting point of the maintenance schedule for each overhead crane considered to achieve the goal. Having described the process from the most granular to the highest point, we can state that the DDSA layer plays an essential role in the system performance assessment because it consequently ensures the search for a sensible maintenance scheduling for the system.

3 RESULTS AND DISCUSSION

In this section, we apply the proposed fitting methodology in a case study, i.e., the DDSA layer, as an example of its implementation in practice. The starting point is the actual filtered degradation data. The system under study is made up of 33 different overhead cranes. The applied fitting process is the same, following the flow diagram in Fig. 3 of reference [18] and the algorithm presented in the previous section. For illustrative purposes, the selected example and the data described below are from a specific 50 tons overhead crane. Tab. 1 shows the raw degradation data for the overhead crane analyzed, with 59,501 hours of continuous operation (6.8 years) and 355 hours of the replacement or repair process.

Table 1. Raw Degradation Data

Time-to-Failure (hours)	Time-to-Repair (hours)
22; 487; 1,636; 635; 505; 1,044; 98; 913; 79; 18; 170; 333; 484; 249; 430; 50; 65; 2,382; 1,044; 1,150; 2,663; 228; 1,559; 688; 134; 1,062; 595; 9; 353; 40; 126; 1,264; 1,244; 67; 2,592; 90; 1,919; 2,943; 383; 324; 89; 1,292; 433; 351; 2,109; 1,060; 1,204; 1,054; 1,550; 471; 240; 2,094; 1,405; 3,220; 829; 223; 29; 462; 657; 511; 68; 766; 1,842; 2,476; 280; 486; 50; 206; 279; 917; 104; 24; 237; 660; 125; 127; 763; 205; 332; 204	49; 4; 5; 7; 4; 6; 7; 2; 29; 6; 3; 4; 7; 2; 30; 7; 4; 4; 2; 1; 4; 1; 1; 10; 1; 1; 12; 7; 7; 2; 1; 4; 4; 1; 5; 1; 1; 2; 2; 3; 4; 6; 2; 4; 2; 4; 3; 2; 2; 8; 2; 3; 4; 3; 1; 3; 1; 1; 3; 5; 3; 8; 7; 1; 2; 1; 2; 1; 1; 1; 1; 1; 1; 3; 1; 4; 3; 1; 1; 1

As described in previous sections, F and R variables are analyzed independently and are the inputs of the DDSA layer. Consequently, we apply the same flow diagram for each variable independently. Tab. 1 highlights that the failure frequency is higher than ten, so we fit the degradation data to all the possible single distributions available in the list (the case study was selected to apply the complete diagram).

The complexity and size of a 50 tons overhead crane mean that, for certain failures, more than one group of workers must fix the unexpected failure (multiple or parallel actions). This condition of parallel repair tasks introduces noise into the stored data. For example, one group of workers completes the repair task, and the other is still working, or during the repair time, other potential problems are found, and the repair time is extended in one of the groups. All situations described above are treated with filters in the data just before we calculate the F and R variables. For example, in cases with multi-actions (multiple tasks simultaneously), we set the failure event to the earliest action and the repair event to the latest action.

The fitting process of the single theoretical distributions can be defined as a constrained multivariate non-linear objective function optimization problem. This investigation solved the parameter estimation based on the maximum likelihood estimation (MLE) method for each fitted distribution with a Sequential Quadratic Programming method. Knowing the features underlying the fitting process, Tab. 2 and Tab. 3 list the final fitted distributions after we apply the infinite mean and variance filter to the data presented in Tab. 1.

Table 2. Best distribution fitted for historical data F.

ID	Parameters	Name	AIC Test
3	$\mu = 743.7660$	Exponential	1219.87

Table 3. Best distribution fitted for historical data R.

ID	Parameters	Name	AIC Test
12	$\mu = 4.4402$ $\lambda = 2.7741$	Inverse Gaussian	377.64

In addition to the AIC criterion for selecting the final decision, Tab. 4 and 5 also include two well-established goodness-of-fit tests in the literature, the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test. As expected, not all goodness-of-fit tests are aligned, and, as we know, all have strengths and weaknesses depending on the object of study. While the AD test focuses on how well the tail of the distribution fits the data, the KS test relies on the full support of the distribution. However, the AIC criterion populated stable selections for the historical degradation data tested.

Table 4. Best distribution for F (contamination vs. data)

ID	Parameters	Name	AIC	KS	AD
3 (Data)	$\mu = 743.77$	Exponential	1219.87	0.5782	0.4607
3 (Cont.)	$\mu = 791.07$	Exponential	2457.48	0.7694	0.8847

Table 5. Best distribution for R (contamination vs. data)

ID	Parameters	Name	AIC	KS	AD
12 (Data)	$\mu = 4.44$ $\lambda = 2.77$	Inverse Gaussian	377.64	0.2063	0.3193
12 (Cont.)	$\mu = 4.42$ $\lambda = 2.75$	Inverse Gaussian	755.97	0.6189	0.5224

As a result of the implemented fitting process, Tab. 2 and Tab. 3 show the parameters of the estimated distribution based on the MLE method for the best fit according to the AIC criterion for the historical F and R degradation data.

Additionally, Figs. 3 and 4 show the visualization of the empirical cumulative distribution function (CDF) versus the theoretically fitted CDF (including the confidence interval), demonstrating a coherent approach to selecting the theoretical fit. As a result, we obtain the closest fit to the degradation data, and meanwhile, we introduce additional complexity into the model only when necessary to achieve higher accuracy (parsimony). The results generated in this section are evidence of how the implemented DDSA layer guarantees robust and accurate final selections in the fitting process, which leaves the database structure ready to be used by the optimization model in the final stage of the process (coordination of maintenance strategies).

There are additional validations to evaluate the performance of the self-analysis layer, which consists of changing the input data in a controlled way and then re-evaluating the fitting selection process. Here, we conduct two tests. The first one consists of contaminating the input data with new values generated from the final selections, and the second one consists of removing one by one the last records in the dataset. In both cases, the process is re-assessed after the change. For the first additional validation, we conducted a simple experiment. We generate a new sample from the final selections (for both cases, F and R) with the same size as the original data, then combine both samples (real data and generated data) into one, and then the DDSA layer is used again following the same process. Tab. 4 and Tab. 5 show the results of this experiment. The parameters of the distribution change, but the final selection is the same. Moreover, both goodness-of-fit improved their results as expected.

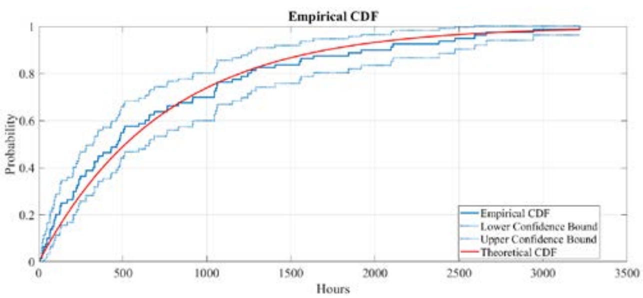

Figure 3. Empirical versus Theoretical CDF (Best fit) for historical data F.

In addition, Figs. 5 and 6 show the new fit plots of the empirical CDF versus theoretical CDF (now contaminated data), evidencing the reduction of the confidence intervals compared to Fig. 3 and Fig. 4.

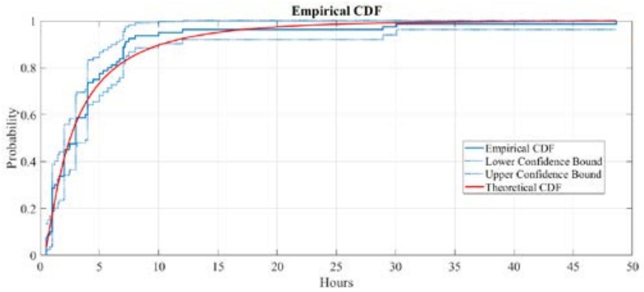

Figure 4. Empirical versus Theoretical CDF (Best fit) for historical data *R*.

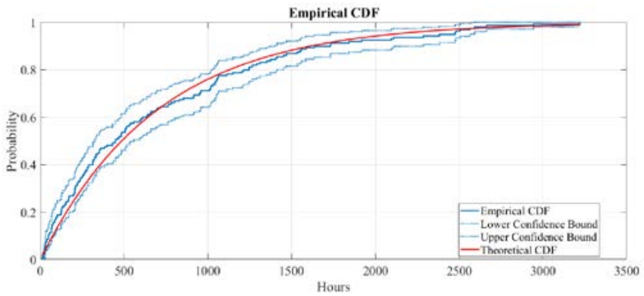

Figure 5. Empirical versus Theoretical CDF (Best fit) for *F*-contaminated data.

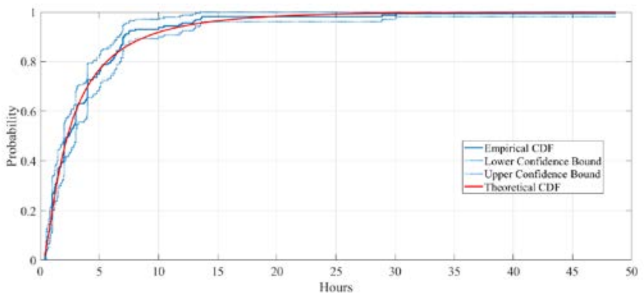

Figure 6. Empirical versus Theoretical CDF (Best fit) for *R*-contaminated data.

In the second additional validation, the experiment consists of removing the last records in the dataset one by one and assessing the changes in the final selected distributions. This experiment is intended to check what happens when we recalibrate the parametric distributions, i.e., equivalent to checking what happens when new failure records appear. In particular, this experiment applies to failure data. The test results are shown in Tab. 6. As we can see, even by removing the last ten records from the dataset, the final selected distribution remains the same.

Table 6. Best distribution fitted for historical data F

Scenario	Name	Parameter	AIC Test
Dataset	Exponential	$\mu = 743.77$	1,219.87
1 point		$\mu = 750.60$	1,206.10
2 points		$\mu = 755.98$	1,191.97
3 points		$\mu = 763.13$	1,178.16
4 points		$\mu = 763.13$	1,162.89
5 points		$\mu = 771.62$	1,149.27
6 points		$\mu = 780.36$	1,135.64
7 points		$\mu = 780.36$	1,135.64
8 points		$\mu = 782.01$	1,120.63
9 points		$\mu = 789.58$	1,106.70
10 points		$\mu = 800.37$	1,093.28

This experiment is an additional result that supports the decision diagram implemented for fitting, which can obtain sensitive results when the data changes. Moreover, the results show how crucial online self-calibration is in capturing the changes in the data. It is clear from Tab. 6 how the parameterization (μ) changes with the data.

At this point of the investigation, we can conclude that the data are correctly, accurately, and robustly filtered, processed, and stored in the database, which ensures clean inputs for the optimization model used in the final stage of the investigation (coordination of maintenance activities). Knowing that the risk model implemented on the integrated digital platform relies on historical data to estimate risk indicators, the approach implemented on the DDSA layer is crucial to ensure robust data processing to guarantee accurate predictions.

4 CONCLUSIONS

The research presented here validates the design and implementation of the DDSA layer that ensures robustness in the filtering and synthesis of the degradation data (time-to-failure and time-to-repair) of the set of overhead cranes considered in this study. The DDSA layer outcomes are probability distributions used to simulate probable failures in the group of overhead cranes, allowing the holistic coordination of maintenance activities by evaluating global system risk indicators. The main research result is validating the robust and accurate filtering and synthesis of the degradation data used to coordinate maintenance activities.

Robustness features come from a formal predefined fitting process (implementation tested in practice in this contribution using the data presented in Table 1), which allows us to obtain coherent distributions to simulate the degradation process due to system operation, given the online historical degradation data. The accuracy comes from a formal technological and modeling connection between degradation due to the system operation, maintenance activities schedule, and management process (selected manufacturing industry needs) through the intermediary database created for maintenance activities coordination.

The proposed solution for data filtering, synthesis, and self-analysis (the DDSA layer) illustrates the practice of self-decision-making focused on operating technical objects. It is an example of the process of adaptation and transition to the digital industry through machine learning approaches. The results achieved so far in this paper through the presented validations (degradation fitting) complement and guarantee the robustness of other processes of the engineering solution (coordination of maintenance activities).

Thanks to the self-contained process designed and shared in this paper, we will be able to analyze in future works the impacts introduced by changes in the data over time, the data window, the criticality of each overhead crane in the system, and the system degradation over time.

ACKNOWLEDGMENT

The Polish Ministry of Science and Higher Education has financially supported the work.

REFERENCES

[1] Canek Jackson, Rodrigo Pascual. (2021). Joint pricing and maintenance strategies in availability-based product-service systems under different overhaul conditions. Reliability Engineering & System Safety, 216, 107817.

[2] Deng X., Yanga X., Zhang Y., Lia Y., and Lua Z. (2019). Risk propagation mechanisms and risk management strategies for a sustainable perishable products supply chain. Computers & Industrial Engineering, 135, 1175–1187.

[3] Domazet, Z., Luka, F., and Bugarin, M. (2014). Failure of two overhead crane shafts. Engineering Failure Analysis, 44, 125-135.

[4] Hanumant P. Jagtap, Anand K. Bewoor, Ravinder Kumar, Mohammad Hossein Ahmadi, Lingen Chen. (2020). Performance analysis and availability optimization to improve maintenance schedule for the turbo-generator subsystem of a thermal power plant using particle swarm optimization. Reliability Engineering & System Safety, 204, 107130.

[5] Hsu, K., Novara, C., Vincent, T., Milanese, M., Poolla, K. (2006). Parametric and non-parametric curve fitting. Automatica, 42, 1869 – 1873.

[6] Jingyuan Shen, Lirong Cui, Yizhong Ma. (2019). Availability and optimal maintenance policy for systems degrading in dynamic environments. European Journal of Operational Research, 276 (1), 133-143.

[7] Kimiya Zakikhani, Fuzhan Nasiri, Tarek Zayed. (2020). Availability-based reliability-centered maintenance planning for gas transmission pipelines. International Journal of Pressure Vessels and Piping, 183, 104105.

[8] Lei X. and MacKenzie C.A. (2019). Assessing risk in different types of supply chains with a dynamic fault tree. Computers & Industrial Engineering, 137, 106061.

[9] Lu, B., Fang, Y., and Sun, N. (2018). Modeling and non-linear coordination control for an underactuated dual overhead crane system. Automatica, 91, 244-255.

[10] Marquez, A., Venturino, P., and Otegui, J. (2014). Common root causes in recent failures of cranes. Engineering Failure Analysis, 39, 55-64.

[11] Mori, Y. and Tagawa, Y. (2018). Vibration controller for overhead cranes considering limited horizontal acceleration. Control Engineering Practice, 81, 256-263.

[12] Naichao Wang, Jiawen Hu, Lin Ma, Boping Xiao, Haitao Liao. (2020). Availability Analysis and Preventive Maintenance Planning for Systems with General Time Distributions. Reliability Engineering & System Safety, 201, 106993.

[13] Putnik, G.D., Manupati V.K., Pabba S.K., Varela L., Ferreira, F. (2021). Semi-Double-loop machine learning based CPS approach for predictive maintenance in manufacturing system based on machine status indications. CIRP Annals - Manufacturing Technology,70, 365-368.

[14] Qian, D. and Yi, J. (2016). Hierarchical Sliding Mode Control for Underactuated Cranes: Design, Analysis and Simulation. Springer, Berlin.

[15] Rusinski, E., Iluk, A., Malcher, K., and Pietrusiak, D. (2013). Failure analysis of an overhead traveling crane lifting system operating in a turbo generator hall. Engineering Failure Analysis, 31, 90-100.

[16] Sidibé I.B., Khatab A., Diallo C., Adjallah K.H. (2016). Kernel estimator of maintenance optimization model for a stochastically degrading system under different operating environments. Reliability Engineering & System Safety, 147, 109-116.

[17] Steurtewagen, B., Van den Poel, D. (2021). Adding interpretability to predictive maintenance by machine learning on sensor data. Computers and Chemical Engineering, 152, 107381.

[18] Szpytko J. and Salgado Duarte Y. (2020a): Integrated maintenance platform for critical cranes under operation: Database for maintenance purposes. IFAC-PapersOnLine, 53(3), 167-172.

[19] Szpytko, J. and Salgado Duarte, Y. (2020b). Exploitation Efficiency System of Crane based on Risk Management. In Proceeding of International Conference on Innovative Intelligent Industrial Production and Logistics (IN4PL 2020), 24-31. ISBN: 978-989-758-476-3.

[20] Theissler, A., Pérez-Velázquez J., Kettelgerdes M., Elger G. (2021). Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry. Reliability Engineering and System Safety, 215, 107864.

[21] Wang H., Yan Q., Zhang S. (2021). Integrated scheduling and flexible maintenance in deteriorating multi-state single machine system using a reinforcement learning approach. Advanced Engineering Informatics, 49, 101339.

[22] Zonta T., Andre da Costa C., Zeiser F.A., de Oliveira Ramos G., Kunst R., da Rosa Righi R. (2022). A predictive maintenance model for optimizing production schedule using deep neural networks. Journal of Manufacturing Systems, 62, 450-462.