University of Genoa

Polytechnic School

Thesis for final dissertation
Ph.D. Course: Machine and Systems Engineering for Energy, Environment and Transport

Curriculum: Transport and Logistics Systems Engineering

XXX cycle

Risk-Based Optimal Scheduling for the Predictive Maintenance of Railway Infrastructure

Advisor: Professor Angela Di Febbraro

Candidate: Alice Consilvio

May 15th, 2018

PREFACE

The research work presented in this thesis has been carried out at DIME - Department of Mechanical, Energy, Management and Transportation Engineering of the University of Genoa, Polytechnic School, and has been sponsored by the Italian Company Ansaldo STS S.p.A., which is a leader in the design and development of equipment and systems for the signaling and automation of metro networks and railway lines.

I would like to express my gratitude to all the persons that helped me in the development of this thesis.

First, I would like to express my deepest gratitude to my advisor, Professor Angela Di Febbraro and to the Professor Nicola Sacco for the great support and for encouraging me to follow my ambitions starting this challenging adventure in the research sector.

I would like to express enormous thanks also to the 'Innovative Technologies & New Services' Unit of Ansaldo STS. In particular, I am very grateful to Nadia Mazzino and Federico Papa for the great opportunity of collaboration with their Company and for the possibility of participation to important European Projects, in which Ansaldo STS is involved as coordinator or partner.

Other special thanks are for the policy officer Cristina Marolda and her colleagues of the Innovation and Research Unit, Directorate General for Mobility and Transport (DGMOVE) of the European Commission for the opportunity to spend a period in Brussels working within this important European Institution.

Thanks to all my colleagues that shared with me the last years.

Finally, the most important thanks to my family that has always been close to me throughout this experience.

ABSTRACT

In this thesis a risk-based decision support system to schedule the predictive maintenance activities, is proposed. The model deals with the maintenance planning of a railway infrastructure in which the due-dates are defined via failure risk analysis. The novelty of the approach consists of the risk concept introduction in railway maintenance scheduling, according to ISO 55000 guidelines, thus implying that the maintenance priorities are based on asset criticality, determined taking into account the relevant failure probability, related to asset degradation conditions, and the consequent damages. This approach belongs to the framework of "predictive maintenance" which aims at intervening when an asset has reached a certain degradation state, being the future conditions forecasted by appropriate models. Therefore, the objective is to intervene when the asset has reached a certain degradation state and thus preventing faults and possible failures, focusing in particular on the maintenance of rail track.

Two different dimensions of the model are introduced:

- *the scheduling of predictive maintenance related to a railway line.*
- *the scheduling of predictive maintenance related to a railway network.*

The problem is firstly formulated as a Mixed Integer Linear Programming (MILP) optimization problem based on risk minimization. Then, considering the extreme computational complexity of this kind of scheduling problems, a matheuristic methodology, based on the mathematical model, is proposed. The algorithm, compared, where possible, with the general-purpose solver IBM-Ilog Cplex®, is characterized by relevant performances.

Moreover, with the aim of taking into account the stochastic nature of real environments, two different levels of the model are considered: the first level consists of an off-line decision support system to schedule railway predictive maintenance activities, while the second level consists of an on-line recovery decision support system, aimed at the schedule adaption when a failure occurs, or when the off-line plan cannot be met.

In particular, to cope with the stochastic nature of real environments, the parameters representing the degradation state of railway assets are modelled as Non-stationary Stochastic Processes (NSP) and a Rolling Horizon framework is introduced to manage unpredicted faults or delays of the maintenance activities.

In doing so, it has been noted that, the adaptive rescheduling models only partially solve the issue of uncertainty, since they consider deterministic sub-problems of the overall problem and they cannot vary continuously the stochastic input variables. For this reason, the risk-based maintenance planning problem is then formulated in term of stochastic programming. The stochastic formulation of the model, for scheduling predictive and risk-based maintenance activities in the rail sector, introduces stochastic deadlines to consider explicitly the stochastic nature of risk and of real-world maintenance operations.

After providing the formal mathematical description of the model, some experimental results are reported and some applications to a real rail network are described, discussing the results. In the end, some indications about future developments of the study are given.

OUTLINE

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Nowadays, rail transport may contribute to solve mobility problems in Europe and all over the industrialized world, thanks to its positive prerogatives. In fact, it is well known that such a transportation mode can play an important role in reducing congestion and environmental impact [1], by means of local and metropolitan trains, in urban areas, or via high-speed trains, competing effectively with air transportation, wherever possible.

Therefore, the development of an integrated, reliable and performant railway transportation system is a target for Europe and many Countries, all over the world. For this reason, in face of a still strong growing demand of transport, the European Union has identified in railway system strengthening a priority for the European transport development and has finalized research programmes, in particular the Shift2Rail Research Programme[1], with ambitious objectives such as interoperability and punctuality improvement, infrastructure reliability enhancement and life cycle cost reduction.

The achievement of such important strategic role and the realization of safety, security, efficiency and economic goals may be supported through the optimization of rail maintenance management. Maintenance has great impact on rail transport performances, costs and quality. It is indispensable in order to ensure service availability and safety for people and goods. Furthermore, a correct and efficient maintenance management may be a way to reduce costs and improve the quality of services.

Nevertheless, railway infrastructure has the characteristic of being scarcely redundant (with none or very few path alternatives), which implies that any fault may result into dramatic drops of the system performance and capacity. Examples of recent severe events are the cases of Pioltello, Italy, (January 2018), and Rastatt, Germany (August 2017). Moreover, the space-distributed aspect of railway infrastructure should be considered and such a characteristic generates significant difficulties in the organization of maintenance activities and in the management of the relevant supporting resources. Another aspect that makes railway maintenance critical is the time

---

[1] https://shift2rail.org/

constraint. In fact, the available time for maintenance activities is strictly limited due to various factors, such as rail traffic, climate, fulfilment of fixed operation sequences, etc.. Some of these requirements result to be soft constraints, that is, violations can be tolerated if no better choices exist, and some are hard constraints that can never be violated.

Therefore, any asset of a railway system need very carefully planned maintenance activities, aiming at guaranteeing the best performance as possible in any time.

To cope with this problem, many maintenance approaches have been developed in the relevant literature, such as corrective maintenance, performed when a fault occurs, or preventive maintenance which can be subdivided into:

- planned maintenance, performed on a regular fixed time schedule. It can lead to a significant reduction of the useful life of components, due to early replacement and unnecessary, a-priori scheduled maintenance activities;
- condition-based maintenance, performed only when necessary, on the basis of the continuously monitored asset conditions. This approach allows a better usage of infrastructure components, but requires a regular and frequent monitoring of the degradation state of railway assets;
- predictive maintenance, performed only when necessary, on the base of suitable model estimations.

The last approach, which is considered in the present work, guarantees the best reduction of maintenance costs, because maintenance is performed only when necessary and a limited number of monitoring measures is required. In particular, the end of predictive railway maintenance is to minimize the probability of the occurrence, of the so-called mission-critical faults during train service, i.e. those that prevents trains for circulating or that can lead to accidents, while keeping maintenance costs as low as possible.

This thesis is developed in collaboration with Ansaldo STS S.p.A.[2], an Italian Company whose core business is the development of solutions for rail transport signalling and automation. Ansaldo STS is often in charge of the maintenance of its plants and sometimes stipulates also full-maintenance

---

[2] http://www.ansaldo-sts.com

contracts with the infrastructure owner/manager of the railway line in which Ansaldo STS systems are installed.

Ansaldo STS is currently involved in two European Union funded projects related to railway maintenance issue: the Horizon 2020 research project “In2Rail” and the Shift2Rail project “In2Smart”.

IN2RAIL “Innovative Intelligent Rail” Project[3] is funded by European Union’s Horizon 2020 research and innovation programme within the call Mobility for Growth-2.1-2014 - I^{2}I – Intelligent Infrastructure. It is one of the lighthouse projects of SHIFT2RAIL, the first European rail joint technology initiative focused on seeking research and innovation (R&I) and market-driven solutions for the railway sector. The aim of In2Rail project is to lay the foundations for a resilient, consistent, cost-efficient, high capacity European network. Specifically, In2Rail explores innovative technologies and the resulting concepts embedded in a system framework where infrastructure, information management, maintenance techniques, energy, and engineering are integrated, optimised, shared and exploited.

Relatively to the maintenance framework, the project aims to gain:

- Reduction in inspection and maintenance costs of more than 25%.
- Reduction in inspection cost of tunnels and bridges by 20%.
- Improved reliability: degradation is continually monitored, and faults predicted; service disrupting faults reduced by 25 - 40%.
- Reduction of maintenance activity cost by 30% through an adaptive and risk-based maintenance approach, in particular for rail track.

In2Smart “Intelligent Innovative Smart Maintenance of Assets by integRated Technologies” project[4] addresses the call “S2R-CFM-IP3-02-2016, Intelligent Maintenance Systems and Strategies” launched by the Shift2Rail Joint Undertaking (JU) and forms part of the framework of research and innovation projects that will deliver the vision and strategy of Shift2Rail IP3 for cost efficient, sustainable, and reliable high capacity infrastructure.

---

[3] http://www.in2rail.eu/

[4] https://shift2rail.org/projects/in2smart/

The aim of In2Smart project is to deliver an innovative asset management process, meeting the best practice set out in ISO55000 in the railway sector. The project should be an accelerator creating new and optimised strategies, frameworks, processes and methodologies, tools, products and systems for the implementation of a step change in risk based, prescriptive and holistic asset management in the rail sector. The project will develop innovation throughout the asset management cycle of data, information, intelligence, decision and execution delivered by intelligent and autonomous systems harnessing the latest technological developments from more advanced sectors and pushing the boundaries of asset management excellence.

In this context, considering the recognized relevance of the problem at the European level, the aim of this thesis is to study the planning of predictive maintenance activities in railway sector, paying attention to the concept of risk-based and evidence-based optimization and focusing mainly on rail track.

In this thesis, a model for the risk-based optimal scheduling of railway maintenance activities is proposed. Such a model consists of a decision support system to schedule railway predictive maintenance activities, in order to intervene when an asset has reached a certain degradation state, and thus preventing faults and possible failures. Taking into account the stochastic nature of asset degradation, the parameters representing the degradation states of railway assets are represented as Non-stationary Stochastic Processes (NSP). The whole approach is based on the risk minimization framework, according to ISO 55000 guidelines [2], and introduces the concept of risk in railway maintenance activities scheduling. It implies that maintenance activities priorities are based on the certain criticality indexes that take into account both the fault probability and the relevant losses.

Two different dimensions of the model are introduced:

- the scheduling of predictive maintenance related to a railway line.
- the scheduling of predictive maintenance related to a railway network.

Moreover, in order to deal with the occurrence of unexpected events, a Rolling Horizon framework methodology is introduced to manage unpredicted faults occurrences or delays in the execution of the previously determined schedule. However, since the adaptive rescheduling model, considering deterministic sub-problems of the overall problem, only partially solves the issue of uncertainty; the railway maintenance scheduling is formulated introducing stochastic variables.

In particular, this work explicitly considers the stochastic nature of risk and of the real-world degradation process, introducing a stochastic formulation of the problem that considers stochastic deadlines.

Summing up, the overall aim of this thesis is to develop a decision support system for railway predictive risk-based maintenance planning and to apply the model to real case studies.

The detailed objectives, which lead to the overall result of the presented thesis, are:

- To review existing maintenance policies and planning principles, which are used in the maintenance planning concept background, and to introduce the notion of planning levels that are consistently approached by the decision support framework (see Chapter 1).
- To describe the existing maintenance procedures, which are used in rail sector and the main considered assets (see Chapter 2).
- To define the (mathematical) features of the optimisation scheduling models subsequently used to formalise the decision-making (see Chapter 3).
- To describe the theoretical background of the stochastic planning concept, which is the handling of uncertainties through the use of probabilistic information. Mathematical concepts related to optimisation under uncertainty that are applied in maintenance planning are introduced (see Chapter 3).
- To review the existing literature and to analyse the most recent European Projects on maintenance planning (see Chapter 4)
- To analyse existing degradation models and to evaluate a suitable approach to predict future asset condition (see Chapter 5).
- To summarise the functionalities, inputs and outputs of important ambient building blocks of the decision support framework (Chapter 6).
- To mathematically formalize the planning model (see Chapter 7).
- To describe the algorithmic solution approach, based on the mathematical concepts for optimisation under uncertainty (see Chapter 8).

- To validate the concept by means of real-world examples: application of the model to the maintenance planning of a real rail infrastructure (see Chapter 9)

The following chapter describes the state of art of maintenance planning techniques.

1. BACKGROUND ON MAINTENANCE PLANNING

This Chapter describes the background related to the concept of risk-based predictive maintenance planning, which is the maintenance policy and planning principle considered in this work. In defining the planning concept, references to this background will be made and basic assumptions and ideas will be used and refined.

First of all, the definition of maintenance needs to be described.

Maintenance includes all actions necessary for retaining a system or an item in, or restoring it to, a state in which it can perform its required function [3].

Other definitions of the concept of Maintenance are given [4] as:

1. “Any activity – such as tests, measurements, replacements, adjustments and repairs — intended to retain or restore a functional unit in or to a specified state in which the unit can perform its required functions.”
2. “For material — all action taken to retain material in a serviceable condition or to restore it to serviceability. It includes inspection, testing, servicing, classification as to serviceability, repair, rebuilding, and reclamation.”
3. “For material — all supply and repair action taken to keep a force in condition to carry out its mission.”
4. For material — the routine recurring work required to keep a facility (plant, building, structure, ground facility, utility system, or other real property) in such condition that it may be continuously used, at its original or designed capacity and efficiency for its intended purpose.”

The concept of maintenance actions has been subjected to substantive changes during the last decades [5], trespassing the original focus on repairing-replacing actions to preventing activities.

The pressure for achieving technical systems able to work under the highest quality standards produces an increase of maintenance actions with a consequent increase of maintenance costs.

Maintenance policies define the general approach to manage maintenance of single components in order to control their behaviour in a long-term perspective, in accordance with the specifics of the respective assets, their surrounding environment and the organisation's strategic objectives. A policy only defines the overall measures to be undertaken, how activities are triggered and possibly combined, but does not define the details of actual application and realisation. Policies typically used are *reactive/corrective, condition-based, preventive/planned* and *predictive maintenance*.

In contrast, *planning principles* are necessary to manage the maintenance of the whole infrastructure system and are used for the selection, adaption and application of models and methods available for decision support. Principles describe the way in which maintenance is planned and organised, by applying certain maintenance policies, selecting and allocating maintenance activities and interventions and deciding on resource usage etc. Three main planning principles are presented: *risk-based*, *reliability-centred* and *evidence-based*.

Maintenance policies often are also called maintenance strategies, concepts, procedures or simply methods.

It is a common practice to distinguish several planning levels that decompose the overall maintenance planning process into single steps with dedicated tasks and decisions to be made: *strategic, tactical* and *operational (or dynamic)*. At the end of this Chapter, a brief definition of these planning levels and of their scopes and of the boundaries between them are provided.

1.1 MAINTENANCE POLICIES

Maintenance policies can be classified into two main categories [6] [7]:

- Corrective or reactive maintenance,
- Preventive maintenance.

Preventive maintenance, which is carried out before break down occurs, can be divided into three categories:

- planned maintenance,

- condition-based maintenance and
- predictive maintenance.

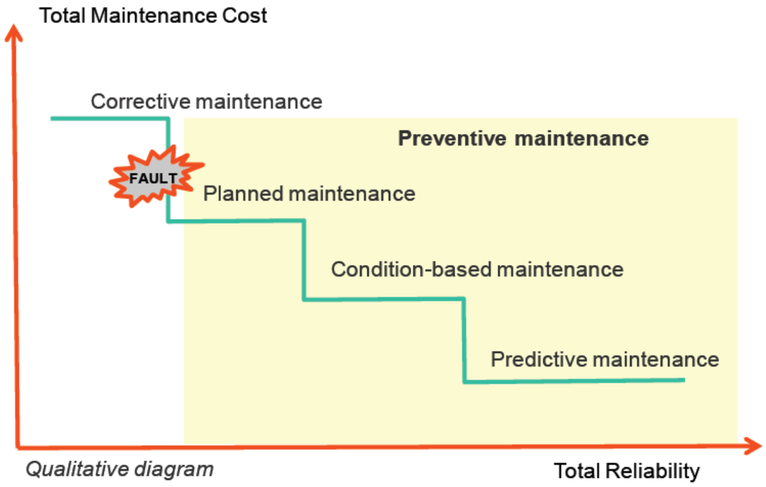

FIGURE 1 MAINTENANCE POLICIES (SOURCE [8])

Corrective maintenance or Reactive Maintenance: The maintenance tasks are carried out after break down. It can be defined as the tasks required when an item has failed or worn out, to bring it back to working condition. This maintenance is triggered by an unscheduled event, such as the failure of an item. With this kind of maintenance policy, the maintenance related costs are usually high due several reasons. First of all, restoring the item or system mostly has to be done urgently, thus the planning of manpower and spare parts is extremely difficult. Secondly, the failure of an item might cause a large amount of consequential damage to other items in the system too. Finally, there are high safety/health dangers caused by the failure and the costs of downtime and penalty associated with the lost production is mostly huge.

To apply reactive maintenance, only few information about the assets are necessary, but the failure will occur unexpected.

Preventive maintenance. This maintenance is carried out before break down occurs. Its aim is to reduce the probability of occurrence of failure. It is used to minimize the disadvantages of the corrective maintenance by reducing the probability of occurrence of failure, preventing sudden failures and discovering hidden failure. The tasks corresponding to this type of maintenance take place under a schedule, in contrary to the unpredicted case of the corrective maintenance, which follows a random failure patterns. Preventive maintenance is carried out to prevent breakdown during operational time, by maintaining the system during the downtime. Therefore, it can be planned ahead and performed when it is convenient. This is very important when work preparation is necessary, so for example new components can be ordered in time and also enough maintenance crew can be available at the planned maintenance execution times.

This type of maintenance can be subclassified into [9]:

- Planned Preventive maintenance or only preventive maintenance.
- condition-based maintenance and
- predictive maintenance.

Planned Preventive maintenance or preventive maintenance: it is characterised by regular maintenance activities that are carried out in predefined time intervals. The aim of preventive maintenance is to extend life-time, to increase asset performance and to avoid unexpected break-down. It includes operations planned in advance. This type of maintenance has many different variations and includes preplanned actions as adjustments, replacements, renewals, inspections. Operations prescribed by the technical specifications of equipments, apparatus and systems are included in this set; it also includes all major operations to be carried on the main assets of railway infrastructure (i.e.: ballast, tamping, geometry). Some of these activities will result in system downtime, whereas others can be done while the system is in operation. The main purpose of this type of maintenance is to avoid failures. It follows a scheme of actions and inspection intervals. There is not guarantee that the equipment will continue to work even if it is maintained according to the maintenance plan, though the probability of failure decreases. To apply preventive maintenance, time intervals values have to be defined, which can be done based on expertise, historical data or scientific results. In Figure 2, the progress of preventive maintenance is shown: in a predefined interval, maintenance activities are executed and the track condition is improved in order to do not fall down a certain condition limit (dashed line).

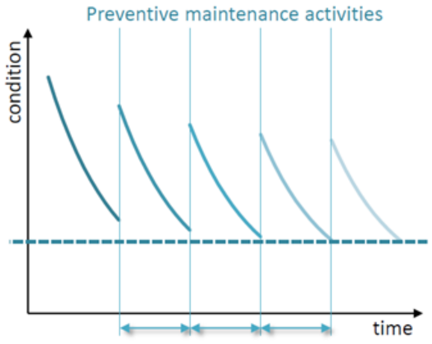

FIGURE 2 PLANNED MAINTENANCE

Condition-based Maintenance (CBM): In condition-based maintenance, failures or break-down will be avoided by maintaining the assets when they show signs of decreasing performance or upcoming failure. Therefore, the assets have to be monitored closely to see condition changes in time, as shown in Figure 3. As soon as monitoring shows that the condition is below the trigger values (dashed line), maintenance is requested and executed. With it, the condition is improved. It is also necessary to define trigger values for the measures that trigger the need of maintenance intervention. Thereby, the time between two condition measurements, as well as the time between maintenance request and execution, has to be considered to ensure punctual maintenance. Indeed, the maintenance activity should be defined in advance. Decision support in resource allocation can help to reduce the time between maintenance request and execution. Then, trigger values can be higher and the maintenance effort can be reduced. Nowadays, the interest for using condition-based maintenance tasks has increased because of the safety requirements and the need to reduce the maintenance costs. Waiting until a component fails may maximise the life of that component, but its failure may cause significant damages to other parts of the system. Moreover, it will cause a disruption of the whole operation. However, having a complicated system (e.g. railway system) with many components, means that it is very difficult (or even impossible) to monitor each of them and to keep all the information in a database. Therefore, conditional maintenance tasks may not be easily usable for complex systems.

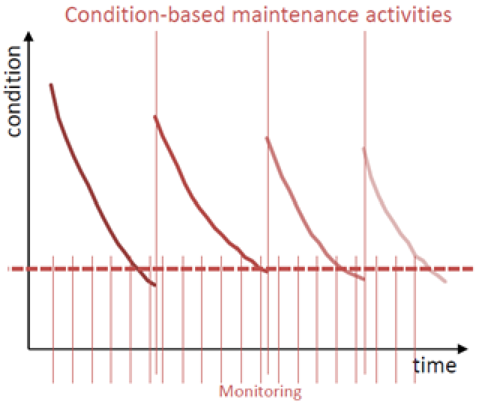

FIGURE 3 CONDITION-BASED MAINTENANCE

Predictive maintenance: Predictive maintenance tasks consists of direct monitoring methods used to determine the exact status of the items, for predicting possible degradations and for discovering those areas where maintenance is needed. The objective is to predict the time at which failures will occur and to take actions based on the predictions. Techniques that help to determine the conditions of in-service equipment and for predicting possible degradations are applied in order to predict when maintenance is needed and should be performed.

Predictive maintenance tries to predict when failure will occur and to plan maintenance activities accordingly. The used techniques are based on statistical analysis and control, to determine at which asset status maintenance will be needed.

Predictive maintenance is a relatively new approach because it requires closely monitoring and a fine understanding of the deterioration process. The main idea is to predict asset condition in order to plan maintenance in advance. In Figure 4, the progress of predictive maintenance is shown. The blue line represents the expected track condition, the red line the real condition. If the prediction reaches the condition limit (dashed line), maintenance will be executed to improve the condition. Thereby, the real condition can be worse or better than predicted, but, with a model close to the real deterioration, the perfect time for maintenance can be approximated. As mentioned, to apply predictive maintenance, a deep understanding of the deterioration process is essential. To develop an efficient degradation model, historical data are necessary. If a good and reliable degradation model exists, asset condition can be predicted with a small variation and maintenance can be planned in advance. The big advantage of the procedure is the high planning ability. The upcoming maintenance activities are known in advance and can be scheduled. The resulting maintenance plans should be robust against uncertainties, like unexpected deterioration or unforeseen events which requires extraordinary maintenance.

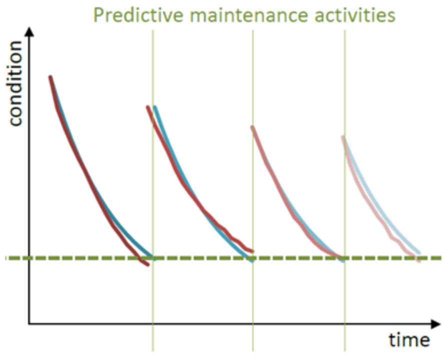

FIGURE 4 PREDICTIVE MAINTENANCE

It is also possible (and usual) to combine strategies, e.g. to request for maintenance if quality falls below the defined condition trigger, but not later than after a predefined time interval without maintenance.

1.2 MAINTENANCE POLICIES COMPARISON

In this Section, the behaviour of the four presented maintenance policies will be analysed. Then, the policies will be compared to each other and some advantages and disadvantages will be shown. In Figure 5, the maintenance policies are evaluated with respect to the planning period. As suggested in the description of the policies, reactive maintenance has to be executed promptly or in short-term. In condition-based maintenance policies, the planning period depends on the selected trigger and the inspection interval. In the most cases, condition-based maintenance has to be planned in short-term, but if the trigger has a large buffer, the planning period can be longer. Preventive and predictive maintenance have longer planning periods. In preventive maintenance, the planning period depends on the length of the time trigger. In predictive maintenance, the planning periods depends on the reliability of the prediction models. It is assumed that predictive maintenance is implemented if and only if the prediction is good enough. Then, the planning period is medium- to long-term.

FIGURE 5 EVALUATION WITH RESPECT TO THE PLANNING PERIOD

In Figure 6, the maintenance policies are evaluated with respect to risks. Reactive maintenance is classified as risky, because in most applications it is too risky to wait until failures occurs. This doesn't mean, that reactive maintenance is never suitable. There are some applications, where reactive maintenance is a good option, for example, when some components are redundant. Condition-based maintenance is safe, if the inspection interval is not too long or the condition trigger is not too low for the degradation rate. To evaluate the inspection interval and the condition trigger value, the degradation rate has to be approximated. If deterioration is underestimated or the chosen parameters are not suitable, maintenance can be requested too late and the risk is higher. Preventive can be safer if maintenance intervals and the usage triggers are chosen in a pessimistic and precautionary way in order to reduce the risk.

Predictive maintenance is safe, because of the longer planning period. With it, possibly misjudgements in the parameter evaluation can be seen in time and the parameters can be adjusted.

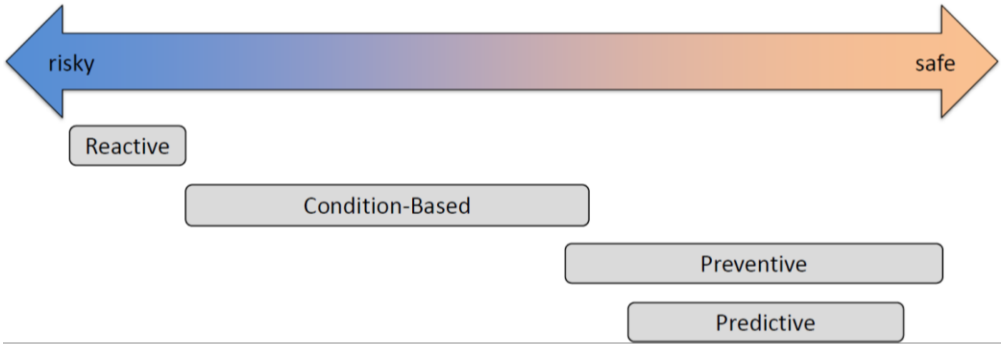

FIGURE 6 EVALUATION WITH RESPECT TO RISK

In Figure 7, the maintenance policies are evaluated with respect to the cost effectiveness. Reactive maintenance is in most cases expensive, because failures are not avoided by improving the asset status through additional maintenance. Then, deterioration can be fast and the assets have to be replaced frequently. Preventive maintenance is in most cases expensive because, to reduce the risk, maintenance intervals and usage triggers should be chosen in a pessimistic way. Doing so, more maintenance is performed than necessary, which increases the costs. Condition-based maintenance has a strong spread of costs. It can be effective, but it can also be expensive, depending on the selected parameters, costs for inspection/monitoring and costs for maintenance. Predictive maintenance can be cost effective, if monitoring is not too expensive. In fact maintenance is planned and executed when it is really needed, since the condition trigger can be chosen lower, and more close to the asset real behaviour, than in condition-based maintenance.

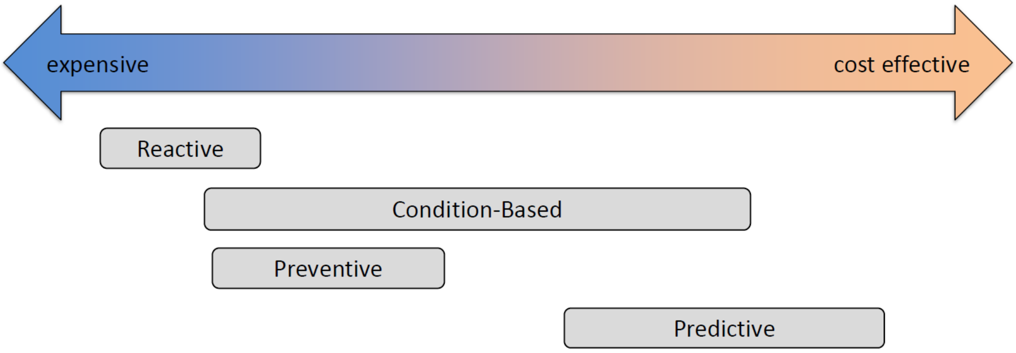

FIGURE 7 EVALUATION WITH RESPECT TO COST EFFECTIVENESS

In Figure 8, the maintenance policies are evaluated with respect to the amount of information that the policy provides regarding the behaviour of the infrastructure system. Reactive maintenance provides no additional information, since only breakdowns are observed. Also, preventive maintenance provides only few information because inspection and monitoring have a secondary role. Condition-based and predictive maintenance give a lot of information about the infrastructure condition. In order to observe the deterioration process, monitoring systems are implemented or inspections are performed. Since, in predictive maintenance, also information about the future condition is provided, this strategy leads to the best informed situation.

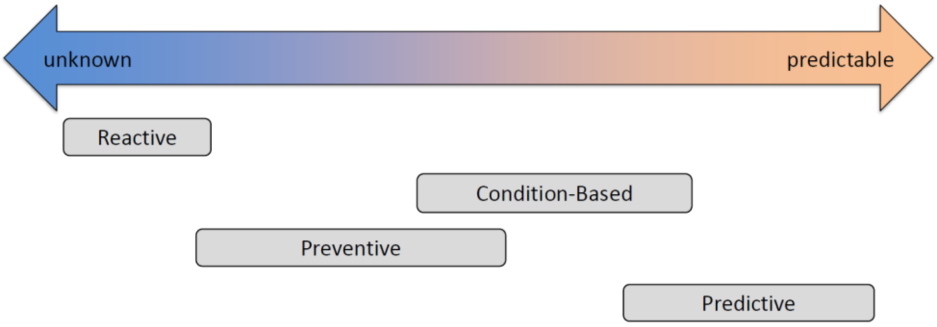

FIGURE 8 EVALUATION WITH RESPECT TO THE AMOUNT OF INFORMATION

In Figure 9, the maintenance policies are evaluated with respect to the implementation effort. Reactive maintenance can be intuitively implemented, without defining trigger values and installing monitoring systems. Only inspections to detect failures are necessary. Also preventive maintenance has a low implementation effort, since time triggers can be defined based on experts knowledge. However, if the maintenance operator increases the effort for implementation, e.g. by analysing the deterioration process to derive better time triggers, the strategy can be improved in terms of costs or risks. The implementation effort for condition-based maintenance is higher, because monitoring systems have to be installed or the assets have to be inspected closely. Predictive maintenance has the highest implementation effort: it requires monitoring, data evaluation and expertise to derive suitable deterioration models in order to predict future condition with a high reliability.

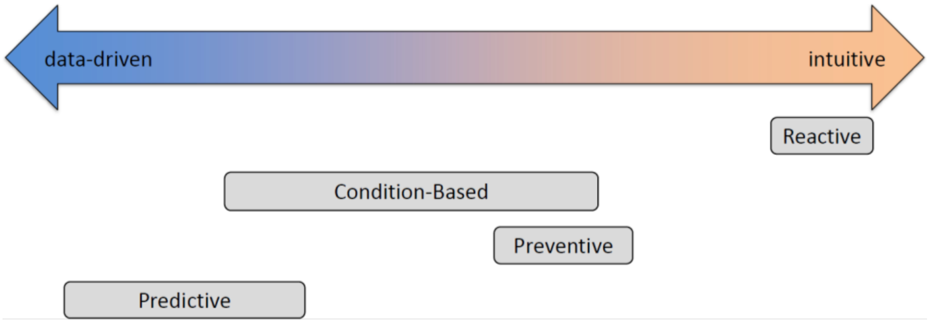

FIGURE 9 EVALUATION WITH RESPECT TO THE IMPLEMENTATION EFFORT

In Figure 10, the different evaluation criteria of the maintenance policies are aggregated into one diagram. The evaluation regarding

- the length of the planning period from short-term to long-term,
- the risk awareness from risky to safe,
- cost effectiveness from expensive to cost effective,
- the amount of infrastructure knowledge from unknown to predictable and
- the implementation effort from data-driven to intuitive

is summarised as a radar chart.

Summing up, reactive maintenance is easy to implement, but the planning period is short, risks are not avoided and no information about condition is given. Thus, reactive maintenance can be used for components with less risk in case of breakdown and less replacement effort.

Preventive maintenance also has a low implementation effort, only the maintenance activities and the time or usage trigger need to be defined. Therewith, maintenance activities can be planned in long-term. Nevertheless, often, this approach is expensive because preventive maintenance is usual done before needed in order to avoid risk. This approach also provides only less information about infrastructure quality. Condition-based maintenance helps to reduce risks and costs, because maintenance is done when necessary. This requires closely monitoring to known the current infrastructure condition. After detecting signs of deterioration, maintenance should be executed promptly, so, the planning period is more short-term up to medium-term.

Predictive maintenance combines long-term planning with condition-based maintenance. Thanks to failure and deterioration models, future infrastructure condition is predicted and based on it maintenance can be planned in advance. To use this approach, a depth understanding of the underlying deterioration processes and the failure models is necessary.

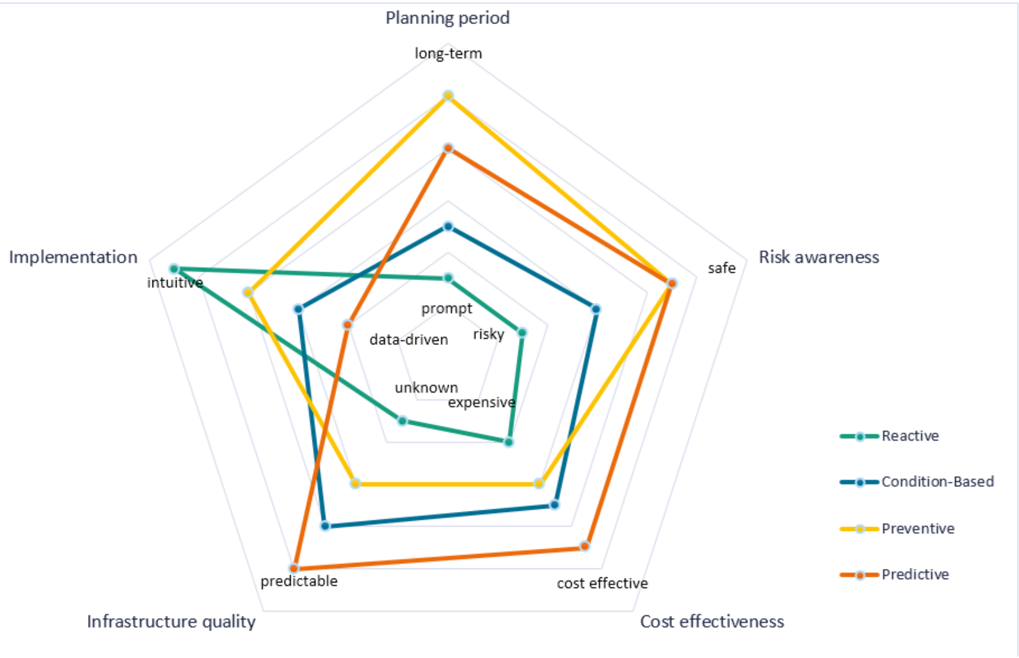

FIGURE 10 AGGREGATED EVALUATION OF MAINTENANCE POLICIES

Table 1 summarises the comparison by listing pros and cons together with possible applications of single policies.

TABLE 1 COMPARISON OF MAINTENANCE POLICIES

	Pro	Cons	Possible applications
Reactive Maintenance	No monitor systems are needed, minor implementation effort	Unexpected Failure and high cost	Assets whose breakdown has a minor influence on the network performance

Condition-Based	Operator knows a lot about network condition	Inspection or monitoring systems are necessary	Assets whose breakdown has an higher influence on the network, but whose deterioration is difficult to predict
Preventive Maintenance	No monitor systems are needed; Activities are planned on long-term	Less condition information; Good choice of trigger value is important	Assets with an estimable deterioration, for which monitoring is too expensive
Predictive Maintenance	Activities are planned on long-term; The best maintenance time with respect to costs and risks can be chosen	Monitoring is necessary; Deterioration has to be predictable; Deterioration should be largely independent from external influences	Assets with monitoring systems and an analysed and recognised deterioration process

1.3 MAINTENANCE PLANNING PRINCIPLES

Depending on the overall asset management framework, the maintenance policies and the necessary parameters are usually defined. Thus, for each kind of component it is decided whether it is maintained in reactive, predictive, condition-based or preventive way, and the related time intervals, triggers, inspection times etc. are evaluated. Furthermore, it is determined which maintenance activities are associated to the different failure modes.

The way in which this detailed planning and organisation is done, depends on the different applied planning principles:

When maintenance decisions are made with focus on risk minimisation, the principle can be called **Risk-Based Maintenance (RBM)**. Risk is calculated based on the probability of failure and the related consequence. Components whose failures have a small probability and less impact are "safe" and could be put aside from maintenance focus. In contrary, components with a high failure probability and/or with drastic failure consequences are "risky" or "critical". In risk-based asset management, these components will be closely inspected and prioritised maintained in order to mitigate risks.

In Figure 11(from [10]), an example of RBM framework is shown. As every maintenance framework, it starts with data collection. By analysing the data, the failure modes are defined and, for each failure, the risk is evaluated. Then, the risks are ranked and a plan for inspection as well as a proposal to mitigate risks are defined, e.g. a set of maintenance strategies and inspection intervals or a proposal to install monitoring systems. At the end of the process, it is checked whether the proposed measures can be realised. If not, a new plan has to be defined.

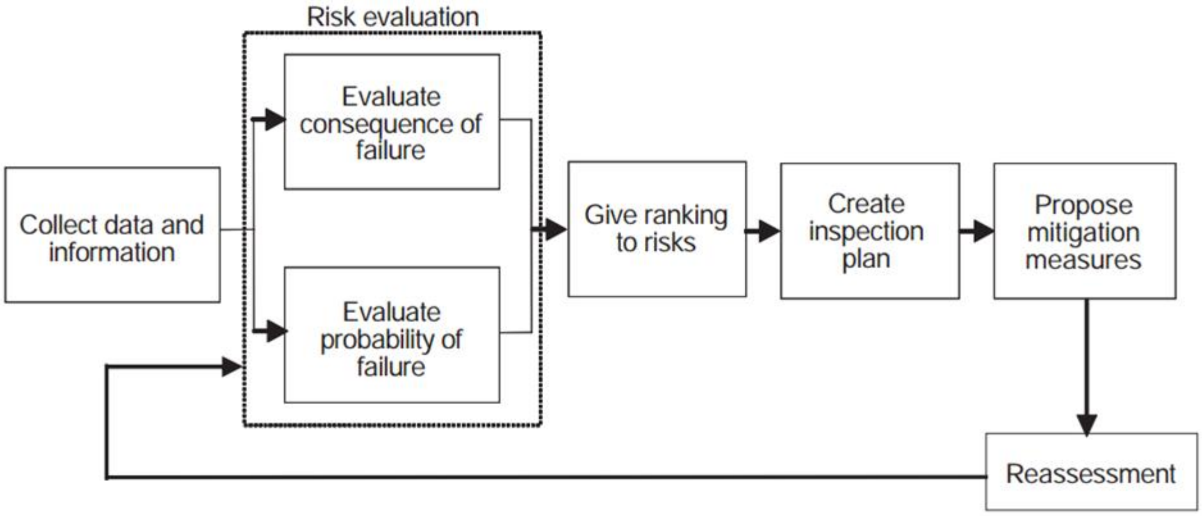

FIGURE 11 RISK-BASED MAINTENANCE

On the other hand, ***Reliability-Centred Maintenance (RCM)*** focuses on system reliability: the aim of maintenance is to ensure the functionality of the infrastructure. Therewith, the focus is not on the components with the highest risk, but on the components with the most important functions to enhance safety and reliability. Maintenance is not necessarily executed to avoid failures; also activities that mitigate the consequences of failures are possible, e.g. speed restrictions or axle load limitations. To reduce costs, unnecessary maintenance is avoided. In order to preserve the system functionality, the failure modes that affect the system function are identified and prioritised. Then, applicable and effective tasks to control the failure modes are selected. At the end, the operator has a ranking of maintenance tasks which is the base for further planning. The seven main questions are:

1. What are the functions and desired performance standards of each asset?
2. How can each asset fail to fulfil its functions?
3. What are the failure modes for each functional failure?
4. What causes each of the failure modes?
5. What are the consequences of each failure?
6. What can and/or should be done to predict or prevent each failure?
7. What should be done if a suitable proactive task cannot be determined?

In the most applications, a mix of risk-based and reliability-centred maintenance will be used because the avoidance of risks and the assurance of system reliability are important for the operator. Both, risk-based and reliability-centred maintenance are rather rule- and experience-based. So, the overall performance of these systems depends heavily on the operator and can be hardly controlled.

Finally, ***Evidence-Based asset management (EBM)*** focuses on data-driven decisions, optimizing clearly defined performance values. Newest scientific findings and mathematical models are applied. This implies the shift of the operator function from maintenance plans definition to maintenance plans control. In order to apply evidence-based maintenance, it is important to develop decision support tools on all levels of maintenance planning. New tools to include new scientific results need to be developed and interfaces for the input and output of the decision support system have to be defined.

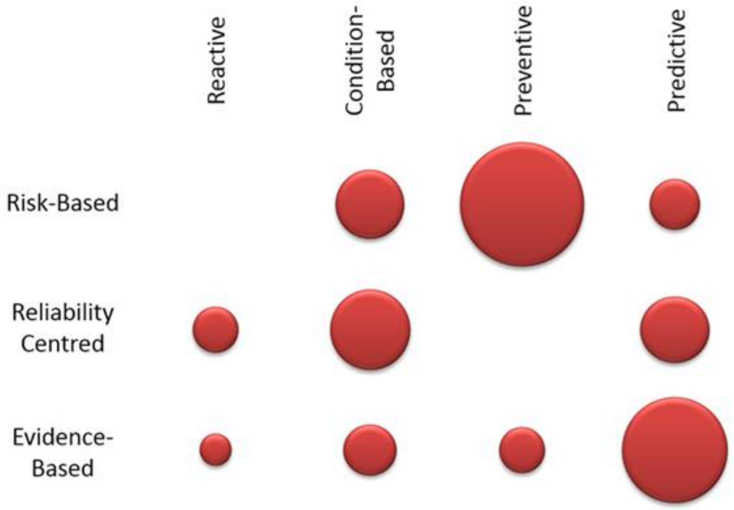

FIGURE 12 ASSESSMENT OF THE USAGE OF PRINCIPLES IN THE DIFFERENT POLICIES

In Figure 12 an assessment for the usage of planning principles in the different maintenance policies is shown. It is expected that risk-based maintenance will focus on preventive and condition-based maintenance because frequent maintenance and inspection leads to a good overall condition that decreases risks. In contrast, reliability centred maintenance is more condition-oriented, so mainly condition-based and predictive maintenance are selected. With it, maintenance can be done when necessary by ensuring a high reliability. Finally, it is assumed that new evidence-based maintenance will be mainly based on predictive maintenance because of the advanced understanding of deterioration processes, while random or unpredictable failures will be maintained in a reactive, condition-based or preventive manner.

1.4 MAINTENANCE PLANNING LEVELS

The scope of maintenance planning and scheduling is to perform all maintenance activities in such a way that the generalized cost is minimized. Planning involves ordering of tasks and resources; scheduling takes care of sequencing the tasks regarding time.

The overall planning process in maintenance management is a complex decision-making process, which cannot be described in a single model and solved in a comprehensive manner. Rather, it is decomposed into several planning steps, each of them dedicated to solve a specific task, thereby focussing on a well-defined set of decisions to be made, constraints and restrictions to be considered and objectives to be achieved.

It is common practice to separate the planning steps into three so-called planning levels: strategic, tactical and operational planning. This distinction to some extend reflects the time horizon of planning decisions to be made within the single step. However, since the actual planning horizon to be considered varies amongst different stakeholders and applications, other criteria are used to distinguish the planning levels, such as the "objectives" and "variables" that are considered in the decision making and the output and consequences of the decision process. An important factor is the level of detail, which increases from strategic to operational.

The most general definition is given as follow:

- ***Strategic planning*** is concerned with decisions that influence the maintenance management in the long-term. No single asset or current failures in the infrastructure network are considered, but the behaviour of asset groups and failure modes under the application of certain maintenance strategies (or policies, methods) is evaluated. Selection of strategies for maintenance and for possession booking, budget allocation, capacity improvement and similar long-term control instruments is done.
- ***Tactical planning*** considers the real network with current and predicted conditions, and decisions are made that are directly related to concrete maintenance activities, according to selected maintenance strategies. Selection, combination and allocation in a mid-term horizon are the typical decisions defined by a tactical plan. Alignment between maintenance and traffic operation is an issue at this level, e.g. possession windows have to be selected and shifted.

- ***Operational planning*** is the most detailed level and considers the actual implementation of single maintenance activities, the scheduling of resources like machinery, staff, material, spare parts. Due to its short-term horizon the possibilities to influence traffic are limited usually given train schedules and possession windows have to be met.

Maintenance plan activities can also be grouped into two classes:

- ***Stationary plans***, which are planned on the long term. The maintenance scheduling can be modelled in an infinite planning horizon, defining static rules which do not change with time. This type of activities allows to group them into related-activity sets and the focus is planning them under an optimum scheme.
- ***Dynamic plan***, carried out on the short term. The plan considers unexpected tasks to carry out corrective operations and other non-corrective planned operations. The maintenance scheduling is modelled in a finite planning horizon, which changes continuously. Real-time information and adaption of existing maintenance schedules is crucial. The activities are also able to be grouped into sets under dynamic optimization overhead management. Dynamic maintenance management models pursuit the optimum scheduling of maintenance operations under a cost objective function subjected to constraints (technological, service, time-window, among others). These models group preplanned operations with unexpected ones in order to rationalize all maintenance activities. When unplanned operations are needed unexpectedly, due to the deterioration of components for instance, the maintenance belongs to the corrective type.

Another classification of maintenance planning is the distinction between ***deterministic*** and ***stochastic*** models.

Deterministic problems are those in which the timing and outcome of the maintenance and replacement actions are assumed to be known with certainty.

Therefore, in the case of deterministic maintenance planning, operation is a-priori known and a deterministic maintenance scheduling can be settled.

On the other hand, the ***stochastic*** **problems** are those where the timing and outcome of the maintenance and replacement actions depend on chance.

Indeed, maintenance operations are envisaged from predictive models under a probabilistic environment, this means that the predictive maintenance tasks will be needed with a certain probability, and a stochastic maintenance scheduling should be considered.

The objective of the stochastic Predictive Maintenance is to give an answer of the most probable required maintenance operations.

1.5 MAINTENANCE TRIGGERS

As mentioned above, the presented maintenance policies require different maintenance triggers:

1. **Condition trigger**: Condition-based and reactive maintenance are based on condition triggers. These triggers can be the achievement of a certain measure value, a dropping below a needed quality level or the occurrence of a failure. Predictive maintenance is also based on condition triggers. Differently to condition-based maintenance, planning starts when the trigger is predicted and not just when it is reached.
2. **Break-down trigger**: A special kind of condition trigger, because the trigger value is the “break-down”. Only usable for uncritical assets in terms of safety and reliability.
3. **Time trigger**: Preventive maintenance is repeated in predefined intervals. These intervals can be defined by time (e.g. once a year).
4. **Usage Trigger**: Preventive maintenance can also be triggered by usage (e.g. every 100,000 switching operations). Triggering by usage has the advantage that for assets with changing workload the definition of a time trigger can be hard. Nevertheless, a usage trigger requires a usage measurement, which leads to a higher implementation effort. Maintenance is performed when the usage limit is reached or when its exceeding is expected.
5. **Event trigger**: Some maintenance activities are triggered by external events, e.g. winter maintenance is triggered by sub-zero temperatures.

If a time period without maintenance is elapsed and the trigger is reached, such as the measured condition falls below a critical value, the asset is broken, or a certain number of trains passed the section, thus a predefined maintenance activity is requested. The selection of the maintenance policy, the definition of trigger values and the selection of resulting maintenance is part of strategic planning.

2. BACKGROUND ON RAILWAY MAINTENANCE

2.1 MAINTENANCE OF RAILWAY INFRASTRUCTURE

A definition of railway infrastructure is given by European Regulation 2598/1970 and comprises routes, tracks and field installations necessary for the safe circulation of trains. Therefore, the term infrastructure covers all the assets that are used for train operation [11]. These assets are:

- Tracks including rails, sleepers, fastenings, switches and crossings, ballast and platform. It can be structured into superstructure and subgrade.
- Switches and crossings,
- Bridges and viaducts: including pillars, decks, foundations
- Tunnels
- Energy supply installations: including catenary and support third rail, substations and control equipment,
- Safety, signalling and telecommunication equipment: including fixed signals, track circuits, train control equipment, signal cables or wires, signal boxes and, for high speed lines, cab signalling systems.

It is a complex infrastructure that requires a high degree of safety and reliability.

The primary objectives of rail infrastructure are:

- first of all, to ensure safe operations of rolling stocks at the scheduled speed,
- to afford conditions for the highest quality and reliability of transport,
- to contribute to a sustainable development.

Therefore, the management of infrastructure should respond to the following objectives:

- to maintain and increase high level of safety,
- to reduce costs, without however decreasing safety standards,
- to improve organization, materials, equipment and staff's qualification in order to respond more efficiently to requirements of operation.

Appropriate maintenance of infrastructure is vital to achieve the aforementioned goals.

The maintenance of this system is a complicated and expensive task which represents an important share of total railway infrastructure costs.

Maintenance of infrastructure can refer to the following components:

- maintenance of track,
- maintenance of electrification equipment,
- maintenance of signalling equipment,
- maintenance of switch and crossings,
- maintenance of bridges and tunnels.

The maintenance of all these subsystems is a complex issue, which makes difficult to plan and execute the maintenance tasks. Factors such as geographical and geological features, topography or climatic conditions need to be considered when planning for maintenance. Furthermore, the availability of the track for maintenance (on possession for maintenance) without disrupting train services is also an important issue to be considered when planning the maintenance tasks to be executed. Maintenance is critical for ensuring safety, train punctuality, overall capacity utilization and lower costs for modern railways.

Regarding railway infrastructure maintenance, a survey conducted among maintenance agencies [12] revealed that most maintenance activities are related to the track system and are concentrated in:

- rail maintenance,
- track geometry maintenance,
- tie and fastener maintenance,
- ballast maintenance,
- track inspection.

Regarding preventive maintenance, the main operations in rail sector can be divided into:

- **Routine (spot) maintenance works,**
- **Project (systematic maintenance).**

Routine (spot) maintenance works consist of inspections and small repairs of the local irregularities carried out manually or using small machines. These are jobs that do not take much time to be performed and are done frequently from once per month to once in a year. For example: switch inspections, switch lubrication, maintenance at level crossings, rectifying track gauge, tamping using vibrating compactors or tamping tines, etc. [13]. Inspections of the track are generally performed at regular intervals.

The purpose of inspections is to determine whether the condition of the track components is satisfactory or unsatisfactory and hence whether further action is required. There are two types of inspections, namely visual inspection and condition measurements.

The frequency of the visual inspection varies depending on speed limit and daily train tonnage from a few weeks on the most important lines to once a month on the least important lines. In general, the visual inspections do not need track possession. Extra inspections are necessary in special cases, such as very hot or very cold weather. The condition measurements, namely ultrasonic rail inspections (with hand equipment or ultrasonic train) are performed in order to check rails for internal defects, rail profile, rail track geometry, rail surface, to carry out measurements on switches. A very important advantage of ultrasonic inspection is that defects can be detected at an early stage, so the repairs can be scheduled on time.

Project (systematic maintenance) include large amount of work that necessitate separate planning. These activities are carried out with heavy track maintenance machines (e.g. tamping machines, ballast regulators, rail-grinding machines, ballast cleaners) approximately once per a couple of years. For example, tamping is done on average every 4 years, grinding every 3 years.

There are also other types of works on the railway infrastructure. These are: renewals and new constructions.

Renewal is done for safety reasons or when the maintenance of different track components is becoming too expensive. The main track renewal activities are: sleepers, rail, ballast renewal and complete or partial renewal of the switches.

New constructions include all activities that are intended to construct completely new tracks, tunnels, bridges, stations, etc. throughout the Country.

Considering the main actors involved in rail maintenance, it is worth mentioning that the liberalization of the railway sector required the distinction between the infrastructure owners and managers, who run the network, and the railways companies, that use it for transporting passenger and goods.

European Union Directive 1991/440/EU was the first step towards the definition of infrastructure manager and service operator as different agents. Therefore, nowadays, the infrastructure manager has to control the rail network and guarantee the required infrastructure performance.

The railway infrastructure manager can perform inspection and maintenance tasks:

- Using its own resources,
- Subcontracting all or part of it (inspection, track, catenary or communications) to different subcontractors,
- Subcontracting everything to only one subcontractor (as it is the case in France).

Usually large railway administrators prefer to subcontract maintenance tasks, that require a lot of Employees, and to perform inspection ones themselves. On the contrary, small railway administrators cannot afford to have (expensive) inspection trains but can perform maintenance activities with their own resources.

In general, the administrators are not technology developers and must buy different instruments and software from different suppliers, which imply different systems and data format. In fact, the big tamping and ballast cleaning machines are designed by a few companies, which usually operate also the maintenance interventions. On the contrary, the measurements instruments involve a lot of companies which must be coordinated.

Summarising, railway maintenance involves many different companies:

- The infrastructure manager, who, in some cases (but not usually), may perform all tasks itself, but in most occasions subcontracts at least parts of the work.

- Inspection companies that owns inspection vehicles or install measurement instrumentation into other company vehicles.
- Specialised track maintenance companies that own tramping and ballast cleaning machines.
- Communications and electric companies that perform maintenance activities on the electrical and communications systems.
- Building companies that take care of subgrade maintenance.

2.2 INFRASTRUCTURE POSSESSION TIME AND TRAIN SERVICES

An important aspect to be considered is that railway infrastructure maintenance works need the **possession of the infrastructure**. The term possession of the infrastructure indicates the use of the infrastructure by some activities. When the infrastructure is used by trains the possession for service takes place, possession for maintenance indicates the use of the infrastructure by maintenance operations. The possession for maintenance can be either partial, when maintenance and trains share the infrastructure, or privative when the maintenance takes full possession of it. The first category implies safety risk conditions to be properly assessed and may be precluded by some/many Administrations. In that follows the term possession for maintenance stands for full possession of the infrastructure by maintenance operations. As train services is the most usual activity supported by the railway infrastructure, single term possession is customary identified with maintenance operations, and is a synonymous of possession for maintenance.

Possession can be divided into several categories [14], ordered according to the severity of the inconveniences carried out by the disruption of the train services:

- **Overnight possession** takes place in the free-of-service periods (time-window). The possession for maintenance depends on the extension of the available time-windows defined by the service of last train on a day and the first train of next day.
- **Weekend possession** makes use of the fact that train services are reduced (may be reduced, re-scheduled or re-routed) respect to labour-day services, therefore larger and more frequent time-windows might be available.

- **Daytime possession**, the shortage of available time-windows makes this possession to be focused on operations that cannot be postponed for latter, such as corrective works.

Currently, travellers are informed days/weeks beforehand if maintenance actions are planned for weekends or weekdays through announcements/leaflets on the stations, media and internet. They can get acquainted with traveling suggestions (e.g. travel via another route, use slow trains instead of intercity trains), alternative traveling possibilities (e.g. travel by special buses) and extra traveling time (e.g. extra 30 minutes).

During track possessions the train operation is likely to be affected. For these days the timetable, the rolling stock and the crew schedules are adjusted and an operational timetable, an operational rolling stock planning and an operational crew planning is made. For a more detailed description of the planning process, refer to Huisman et al. [15]. Thus, in the original (tactical) timetable, trains are inserted, deleted or their departure and/or arrival times are changed. Furthermore, the type, number and order of the train units to be used for each train and the duties of the crew might be different during these maintenance periods than in the original (tactical) planning. Once the maintenance work is finished the train operation should be retaken as quick and smooth as possible. This means that the crew should retake their original duties, the trains should operate again according to their tactical timetable and with a composition prescribed in the tactical rolling stock planning. However, after a possession the rolling stock units may not finish their duties at the location where they were planned to or at the location where they have to start their duties next day. This is not a problem if two units of the same type get switched. In many cases, however, the number of units ending up after a possession at a certain station differs from the number of units that has to start their next day's duty there. To prevent expensive dead-heading trips, i.e. trips where empty units are moved from one location to the other one, it is attractive to consider this balancing issue already at the moment when the operational rolling stock schedule is made.

The main lines are intensively used during the day; therefore, taking out working zones of these lines in the daytime would cause severe disruptions for the railway traffic. Thus, the main lines are maintained at night (the passenger traffic is almost absent in the night and the cargo trains can be grouped: one set of trains per hour in each direction). Nevertheless, some minor rail operators make maintenance on the line during the spare periods of the timetable along the day, such as the the case of the Trento-Malè- Marilleva line.

Therefore, maintenance is critical for ensuring safety, train punctuality, overall capacity utilization and lower costs for modern railways.

The need for more maintenance and the increase of infrastructure possession time to carry it out, is in conflict with the increment of the infrastructure use by train services to satisfy the demand. Longer operating time, higher number of services and trains increase the annual traffic load and accelerate the infrastructure deterioration; this results into an increase of the number, severity and frequency of the needed maintenance operations and, on the other hand, it has decreased the available time for maintenance. In addition to the above issues, the European directive regarding infrastructure charges and capacity allocation [16] defines an organizational and regulatory framework tending to an optimization of the railway infrastructure. More recently, the Fourth Railway Package Technical Pillar, comprising the ERA Regulation 2016/796, the Interoperability Directive 2016/797 [17] and the Rail Safety Directive 2016/798 [18], published in the Official Journal of the EU on May 26 2016 and entered into force on 15 June 2016, deals much with maintenance, though with a relevant attention on rolling stock, but with ample references also to the infrastructure. In particular, the Interoperability Directive 2016/797 and the Rail Safety Directive 2016/798 address maintenance of the rail network, mainly in relation with the operation of rolling stock, highlighting their interrelation.

Maintenance productivity is directly related to the available time-windows of train services. Since time-windows are limited, while the rail transport demand is growing, advance managerial techniques and procedures get into action as a tool to combine the objective pursuit by both need, train service and the maintenance.

Therefore, traffic management as well as maintenance management is related to make use of the existing infrastructure, and thus depend on information of their elements and respective conditions, but both do so in a different way and from different perspectives:

- Traffic management uses infrastructure information to control and assure the operation of planned train services.

- Maintenance management in the long run also aims at increasing infrastructure availability, but in the short term its immediate effect is to restrict train services by occupying time windows for interventions – either booked possessions or during train-free periods.

In this sense, traffic management and maintenance planning can be seen as concurring topics on the infrastructure. Therefore, ideally both planning and decision-making processes should be done together, guaranteeing to be perfectly aligned to each other. It is not possible to globally assign to one process – for example to the traffic management – the priority against the other, but rather a healthy balance has to be found – an issue of so-called concurring planning.

In the practice such an idealised coordinated approach is rarely implemented, because of several reasons:

- from an organisational aspect, both topics involve different stakeholders with different tasks, requirements to consider and objectives to achieve, depending on contractual regulations amongst them which exist in a wide variety;
- from an historical aspect, methodologies and tools to support traffic management resp. maintenance planning have been developed independently from each other, used in different levels of maturity in respective organisations; and
- for practical reasons it is clear that both tasks for its own already are very diversified and constitute complex decision-making processes, so that a computational handling of a fully integrated planning approach is more than ambitious.

Nevertheless, in the scientific literature it is possible to find recent attempts to formulate and solve this problem such as in [19] and in [20].

To enable a broader spectrum of coordinative activities, traffic management has to be considered in a slightly broader context than the usual meaning as real-time traffic control and operation. Traffic management should also include the issues of short-term capacity planning which arise when an adjustment or revision of timetables becomes necessary, often triggered by asset and infrastructure condition and the requirement for maintenance activities. Also, the coordination possibilities to be achieved cover the planning of minor maintenance activities (during train-free periods or asking for extra possessions in the short-term) as well as the operating planning and control of traffic affected by maintenance. Such a broader view allows to identify more coordinative (concurring) planning activities than in the real-time controlling process only, where only a few restricted options are available to react on issues, but actual "proactive" planning capabilities are not given.

In principle, there are two ways for coordinating planning [21]:

- The requirements from traffic management influence and "govern" the maintenance planning model, i.e. the maintenance planning model makes use of information defined by the traffic management, this is the case of the model presented in this thesis.
- Traffic management models include information coming from maintenance, and consider them for decision-making.

The basic approach to concurring planning in mathematical modelling is to integrate elements of one aspect into the model of the other aspect, either in the definition of the decisions or solutions space, or as a component of the objective functions, or as additional constraints to be met.

The following examples underpin the ideas behind this approach:

- Decisions / solution space: To integrate decisions for traffic management into maintenance planning or vice versa means to define decision variables in the model that will be considered and determined in parallel, e.g. a model with variables for the allocation of minor maintenance activities to train-free periods, where the start and end times of the train-free periods also are variables that have to be fixed (within a smaller range). Naturally, this drastically increases the solution space of the underlying mathematical model, and thus leads to decision problems with a high computational complexity, practically hard to solve or even untraceable. The alternative is to separate decision variables in different models and to balance between models.
- Constraint: Into a model for the planning of mid-term maintenance activities, a restriction on the possible time windows – either coming from booked possessions or from train-free periods – can be integrated in order to be aligned with traffic management, this is the case of the model presented in this thesis.

- Constraint: When adjusting timetables and train paths, it is possible to consider asset conditions and restrictions on traffic resulting from them, which means to foresee maintenance activities required or already allocated. This means to add constraints e.g. in the form of non-availability of sections or speed reductions.
- Constraint: In resource scheduling for maintenance activities the model could include constraints on the travel paths for machinery or constraints on the working area for maintenance staff, both induced by considerations coming from traffic management.
- Constraint: Real-time traffic control as a planning problem could consider the actual state of work activities and derive from this constraints on available times, sections, speeds.
- Objective: The short- or mid-term maintenance planning model can define an objective function to minimise costs for extra possession windows when allocating activities.
- Objective: The planning model can define an objective function to minimise train delay due to planned activities, both when planning maintenance activities or when adjusting traffic control, especially in short-term planning levels.
- Objective: The long- or mid-term maintenance planning model can include a bonus for releasing unused possession times in the form of "negative costs".
- Objective: Mid-term planning models to adjust train paths and timetables can aim at minimising the number of affected trains (or more general: expected capacity reductions).

2.3 RAILWAY MAINTENANCE COSTS

Maintenance costs are affected by factors such as:

- increases of maintenance actions due to higher quality standards,
- increases of manpower costs of maintenance personnel,
- increases of management costs.

Preventing maintenance is an area of increasing importance due to the economic interest to reduce maintenance costs. Corrective maintenance tasks will be never avoided because of unexpected failures. These failures provoke disruption of the production/service and cause not only additional costs for production losses but additional malfunctions/damages to other related components and equipment.

Preventive maintenance is a step forward intended to minimize corrective maintenance. Nevertheless, it is most founded in models that predict the risk of failure and residual life of components in order to finally integrate and merge them into a system predictive model. The difficulties encountered in this bottom-up modelling process may end up with conservative scheduling actions that might increase the maintenance actions over those strictly needed. This conservative behaviour gives lieu to an increment in the maintenance costs, due to earlier part replacement with still substantial residual life.

The four maintenance types can be coordinated under a scheduling scheme, by combining activities, in order to minimize the period of time the equipment is idle. This approach helps saving costs. This way of proceeding has also consequences. The replanning of the maintenance scheduling should be carefully optimized in order to avoid non-attention to previous maintenance scheduled tasks, which might additional increase costs.

Orders of magnitude of maintenance costs in some sectors are presented in Cross [22] as a percentage of the total operating costs.

With respect to railway maintenance costs, Zoeteman [23], [24] reports that tracks and switches consume more than 50% of total maintenance costs and 75% of renewal costs due to their high usage, relatively rapid deterioration pattern and high installation cost. Tracks and switches are, in fact, the most important equipment of the railway infrastructure, since they are often sources of failure and traffic disruption.

Track maintenance costs and planning are heavily influenced by track possession and site access windows. An average work window of 4 hours is most common, of which 50 % is expended for displacements to the work site. This means that a planning regarding access, to coordinate maintenance crews and resources, is required in order to keep costs bounded.

The main maintenance costs derive from wear and fatigue rather than secondary sources such as corrosion, derailment, human error and vandalism.

There is a second additional factor that cannot be neglected and influences the overall maintenance budget, it comes from the need of replacement of components due to their inappropriate performance or design during the project deployment.

Finally, the third most influence factor is postponed track maintenance. When this case occurs, corrective maintenance has to be carried out as a matter of urgency, and costs rise up.

These findings are common for almost all type of railway infrastructures.

Summing up, the state of the art reports that maintenance costs are influenced by the following factors:

- **Infrastructure maintenance activities**. The costs associated can vary with system configuration and the technology associated to the infrastructure. Maintenance activities are characterised by the "Maintenance Demand" defined as the level of resources (effort, materials, equipment, organizational and administrative) to provide and acceptable asset condition level.
- **Resources** (labour, material, equipments, organizational and administrative) costs. Work-window costs. It stands for the cost associated to the time-window created to conduct maintenance activities when the infrastructure possession is detracted from railway service.
- **Work-window costs**. It stands for the cost associated to the time-window created to conduct maintenance activities when the infrastructure possession is detracted from railway service.
- **Inspection and maintenance technologies costs**. They are affected by the inspection technology used and the maintenance techniques utilised to carry out the operations.
- **Inspection and maintenance policies costs**. They are affected by the level of maintenance enforced, the frequency of inspections.
- **Operating costs**. They stand for the costs associated to operational optimisation of maintenance operations.
- **Indirect cost.** They represent the costs not directly involved in the maintenance tasks such as preparing crews and materials for a task, mid-level supervision, stocktaking of inventory,

purchasing activities, equipment procurement and maintenance, training and organisation overheads.

Due to very high maintenance costs, it is important that the track maintenance works are scheduled in an effective and efficient manner. This includes short term planning, such as daily scheduling of the activities, as well as the medium to long term planning.

2.4 RAILWAY MAINTENANCE PLANNING

As mentioned, since rail is an important transportation mode, proper maintenance of the existing lines, repairs and replacements carried out in time are all important to ensure efficient operation.

Moreover, since some failures might have a strong impact on the safety of the passengers, it is important to prevent these failures by carrying out in time and according to some predefined schedules preventive maintenance works.

Therefore, any asset of a railway systems needs very carefully planned maintenance activities, aiming at guaranteeing the best performances as possible in any time. In particular, predictive railway maintenance is aimed at minimizing the probability of the occurrence, of the so-called mission-critical faults during train service, i.e. those that prevents trains for circulating or can lead to accidents, while keeping maintenance costs as low as possible.

Nevertheless, the scheduling of the preventive maintenance works on the railway infrastructure is very difficult, since there are many constraints to be considered. Railway infrastructures have the common characteristic of being scarcely redundant (no or very few path alternatives), and this implies that, when a fault occurs, the system performances have a dramatic drop. Moreover, railway maintenance must take into account the space-distributed aspect of railway infrastructure. Railway assets are often not spatially delimited to a point, and this implies difficulties in the organization of maintenance activities and resources. Another aspect that makes railway maintenance critical is the time constraints due to rail traffic operation.

As a matter of fact, carrying out maintenance on the rail infrastructure usually involves many disturbances for the travelers (e.g. delays, canceled trains), and vice versa, the train operation restricts the length and the frequency of the infrastructure possession. Moreover, in the last years due to a couple of severe accidents, the safety regulations for the track workers became very strict. Thus, in some countries no train operation is allowed during maintenance work. Furthermore, the railway infrastructure maintenance costs have increased substantially in the past years.

Since the railway infrastructure maintenance is very difficult to plan and involves high costs, there is a need for developing tools, which help the maintenance planners to come up with optimal maintenance plans. This issue is proved to be an interesting research subject since the results might help to improve the quality of the rail infrastructure. As a consequence, this thesis is focused on railway infrastructure maintenance, however the mathematical models and techniques developed for scheduling railway maintenance can also be applied for maintenance scheduling in other sectors as well.

2.5 MAINTENANCE OF RAILWAY TRACK

The maintenance of railway track is a complicated and expensive task, which represents an important share of total railway infrastructure costs. As a matter of fact, the maintenance of the track represents around 40% of the total maintenance cost of the railway system.

The state of track depends on many factors such as the characteristics and age of the elements, the track geometry, topography and geology, weather conditions and supporting loads. Track maintenance is still a very little automated process, relied on the skills of specialised human operators and based on rules established a long time ago (preventive maintenance) complemented by the execution of on-call corrective tasks whenever there are faults in the system.

Furthermore, the saturation of the capacity of the track sections as a result of increased load of rail services requires intensified maintenance and the planning and coordination of the rail activities, in order to accommodate maintenance tasks to the availability of time windows needed to guarantee technical regulatory levels. Thus, as railway uses increases so does the need for maintenance while the availability of the track for maintenance decreases. As a consequence, the work is mostly carried out outside of daylight conditions and under pressure, increasing the risk of staff accidents. Finally, the maintenance management, based on cyclical preventive works and on corrective maintenance, entails high costs, in both resources reliability and availability of infrastructure.

This situation requires the streamlining of the maintenance management based on monitoring the track condition, automating the planning management and especially monitoring the evolution of the parameters that determine the track condition for predictive maintenance and risk analysis. This scheme would allow evolving the maintenance management model based on corrective/preventive maintenance into a model based on conditions/predictions, helping those responsible for making decisions to achieve optimal maintenance plans that minimise the maintenance costs, ensure a satisfactory safety margin and prevent quick degradation of track quality.

In this section, the structure of railway infrastructure is shown and the main defects and the related maintenance interventions are described.

As regards railway structure, it is possible to distinguish the **superstructure** and the **subgrade**.

The **superstructure**, which supports and distributes train loads and is subject to periodical and maintenance and replacement, consists of:

- The track,
- The track bed.

The **track** consists of:

  - The **rails**, which support and guide the train wheels.
  - The **sleepers** (also called ties, mainly in North America), which distribute the loads applied to the rails and keep them at a constant spacing.
  - The **fastenings** which ensure the rail-sleeper connection.
  - The **switches** and **crossing.**

The **track bed** consists of:

    - The **ballast**, usually consisting of crushed stone and only in exceptional cases of gravel. The ballast should ensure the damping of most of the train vibrations, adequate load distribution and fast drainage of rainwater.

    - The **sub-ballast**, consisting of gravel and sand. It protects the upper layer of the subgrade from the penetration of ballast stones, while at the same time contributes to further distributing external loads and ensuring the quick drainage of rainwater.

The **subgrade**, on which the train loads, after adequate distribution in the superstructure, are transferred and which in principle should not be subjected to interventions during periodical maintenance of the railway track, consists of:

- The **base**, which in the case of the track laid along a cut consists of onsite soil, while in the case of an embankment is composed of soil transported to the site.
- The **formation layer**, used whenever the base soil material is not of appropriate quality.

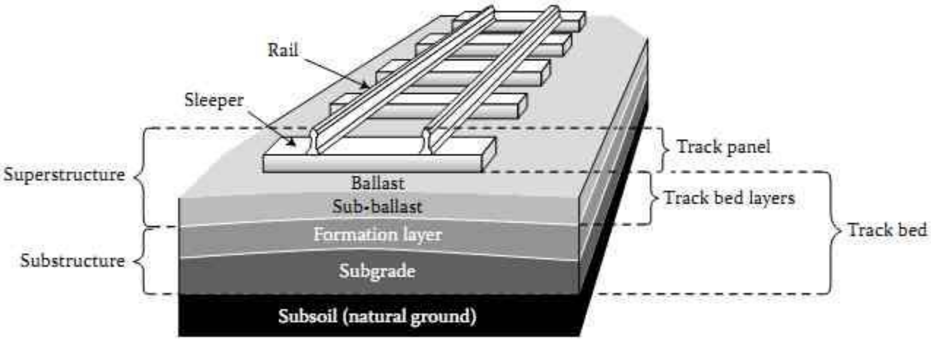

FIGURE 13 SUPERSTRUCTURE AND SUBGRADE

The track usually lies on ballast which provides a flexible support. It is referred as ballasted track. However, it is possible, that the track is supported by a concrete slab, instead of ballast. In this case, the support is inflexible and it is called slab track. Although a slab track is used in certain railways (e.g. the Japanese and the German, among others), it is most effective when used in tunnels, because it allows a smaller cross-section and facilitates maintenance. In most of the tracks worldwide, a ballasted track is still the case, as it ensures flexibility (an important factor in the event of differential settlements) and much lower construction cost, while at the same time offering a very satisfactory transverse resistance, even at high speeds. The problem of noise, which is much greater with the track on concrete slab than with the track on ballast, should not be disregarded. When a slab track is applied (e.g. in the case of a tunnel), the sudden variation in track stiffness (felt by passengers as a jolt) is lessened by placing rubber pads of a suitable thickness along the tunnel entrance and exit.

The choice between ballasted and non-ballasted track should consider construction cost (much greater for non-ballasted track), maintenance cost (much greater for ballasted track), together with technical requirements. Both solutions have pros and cons.

The main defects of railway infrastructure can be divided in:

- Track defect.
- Rail defects.
- Sleepers defects.
- Fastening defects.
- Ballast defects.

Regarding **track defects**, they are defined as the deviations of the actual form from the theoretical values of the track's geometrical characteristics. Track defects are consequence of train traffic, they are of a macroscopic and geometric nature and usually they are rectified by track maintenance.

In particular, geometry **track defects** include [25]:

- Longitudinal defect
- Transverse defect
- Horizontal defect
- Gauge deviations
- Track twist

The **longitudinal defect** (LD), is defined as the difference between the theoretical Z_{th} and the real value Z of track elevation and is given by the equation:

$$LD = Z_{th}(T, x) - Z(T, x)$$

The longitudinal defect is the most reliable in illustrating the effect of the vertical loads on track quality and it is the principal factor (together with the transverse defect, see below, which accompanies longitudinal defects) in determining the magnitude of the track maintenance expenses.

The **transverse defect** (TD), is defined as the difference between the theoretical and the real value of cant. The cant of a railway track is the change in elevation (height) between the internal $Z_{int}$and external rails Z_{ext}.

$$TD = (Z_{int} - Z_{ext})_{th} - (Z_{int} - Z_{ext})$$

This is normally greater where the railway is curved; for rectilinear parts of track layouts, where curvature is zero, the transverse defect is the difference of elevation between internal and external rail: $Z_{int} - Z_{ext}$.

The **horizontal defect** (HD), is defined as the horizontal deviation of real position of the track from its theoretical position. The horizontal defect depends on the transverse track effects (more than the two previous types of defects) and on the characteristics and particularities of the rolling stock.

Certain track **gauge deviations**, affected by the mechanical properties of track materials and the particularities of the rolling stock, are permissible. Gauge values acceptable for standard gauge tracks are defined for each track line.

Since along straight and circular sections (where cant is constant), four point of the track lying on two transverse sections must lie in the same plane, **track twist** is defined as the deviation of one point from the plane defined by the other three.

If i and $i+1$ are two successive transverse sections of the track, spacers Δl apart (e.g. at the positions of two sleepers), track twist is defined as the variations of the transverse defect TD per unit length Δl.

$$Tracktwist = \frac{TD_{i+1} - TD_i}{\Delta l}$$

The risk of derailment is prevented when the real value of twist is smaller than its critical value causing derailment, which depends mainly on speed and to a lesser degree on the type of the track equipment and of rolling stock.

Considering **rail defects**, it is worth mentioning that the rail suffers from stresses that can cause defects and may bring it to failure. The total stresses developed in the rail are the sum of:

1. Stresses at the wheel-rail contact,
2. Stresses resulting from rail bending on the ballast,

3. Stresses resulting from bending of the rail head on the web,
4. Stresses resulting from thermal effect,
5. Plastic stresses, remaining in the rail after the removal of external loads.

Those stresses gradually decrease the mechanical strength of the rail due to repeated loading. That is called fatigue. Once the fatigue limit is reach, the rail is bring to failure.

The effects of these stresses could be:

- **Plastic deformation**. The rail support high stresses during the train circulation. If those stresses are greater than the elastic limit, a flange can appear in the rail head, because below this flange, the stress limit may be exceeded.
- **Rail wear**. The traffic load produces the rail wear that affect the rail profile. There are two type of rail wear:
    - **Vertical wear**. It reduces the rail section and consequently the rail resistance. The maximum permissible vertical wear of the rail is a function of the maximum train speed and of traffic load. There are regulations, in different countries, to limit the maximum permissible vertical wear of the rail.
    - **Lateral wear**. It reduces the rail section and consequently the rail resistance and affect to the gauge of the track too. As in case of vertical wear, the maximum permissible lateral wear of the rail is a function of the maximum train speed and of traffic load. There are regulations, in different countries, to limit the maximum permissible lateral wear of the rail.
- **Surface defects** can be distinguished in:
    - **Short-pitch corrugations.** Their cause is train traffic and they consist of corrugations with a wavelength $\lambda = 3 - 8$ cm. They can provoke many adverse effects: high frequency oscillation of the track, including resonance, and leading to higher rail stresses, concrete sleeper fatigue with cracking in the rail seat area, loosening of fastenings, accelerated wear of pads and clips, premature failure of ballast and the subgrade, and increase by $5 + 15$ dB in noise level. This defect is

  detected either visually or by appropriate recording equipment. It is repaired by passage of special equipment, which grinds and smooths the rail.
    - **Long-pitch corrugations.** They have wavelengths $\lambda = 8 - 30$ cm and occur mainly on the inside rails of curves having a radius of 600 m and smaller. This kind of defect is the most common on suburban and underground railways. Detection and repair processes are similar than those for short-pitch corrugations.

Regarding the **sleeper defects**, it is worth noting that their failures depend on the type of sleepers (steel, or concrete).

Defects in steel sleepers derive for sensitive to chemical attack particularly in industrial and coastal areas. The steel sleepers in a non-aggressive environment have a lifetime of about 50 years. In chemical aggressive environment this lifetime could be only a few years, by the accelerate corrosion.

Defect in concrete sleepers are rare, because of its strict quality control and testing in manufacturing. Generally, defect in concrete sleepers are consequences of defect in the accessories, like loss of fastening, pad or other defects that force concrete sleepers to work out of design conditions. The estimated lifetime for the concrete sleepers are 50 years, which is the time form track total renovation.

The fastenings are the elements used to ensure the anchorage of the rail to the sleeper, to maintain the correct longitudinal and transversal position of sleepers and, if necessary, to ensure the electrical insulation.

One of the most common **fastening defects** is a reduction in the connection between the rail and sleepers, with the risk that it is no longer guaranteed the correct track gauge. In addition, fastenings may be subject to another type of serious defect, which is that of fatigue failure.

Defects in fastening depend on the type:

- Rigid fastenings present two type of defects:
  - broken bolts or nails and

    - gap between nails and rail. Both defects are severe and may cause a derailment.
- Elastic fastening. Defects in this type of fastening may cause either break of a fastening components or loosing of a fastening element.

The most typical and frequent **defect of switches**, is related to the poor lubrication of the mechanical elements that constitute them. This type of defect involves a malfunction of the switch, with subsequent slowdown in the operation of deviation.

Another defect consists of the wear and aging of mechanical elements.

Finally, **ballast defects** are due to the traffic loads, which, along the time, cause the rearrangement of the stone modifying the characteristic of the ballast. With this rearrangement, the ballast loose his original characteristics, increase its flakiness index, decrease its granulometric composition and loose partially the hard corners of the stones. These changes of characteristics worsen its behaviour with the traffic load with respect to stress distribution, vibration attenuation and rainwater retention.

Once the main track defects have been identified, it is possible to describe the most representative maintenance and renewal tasks to be performed on the track related to the above-mentioned defects.

Regarding **rail maintenance operations**, there are four methods to eliminated rail defect:

- Rail weld recharge.
- Rail grinding.
- Rail replacement.
- Rail tamping.

Rail weld recharge is a technique for the cost effective repair of discrete defects on the running surface of rail. The results archived are very much depending on the expertise of the operator.

Repair welding is the most cost-effective method of repair of defects of the tread and guide surfaces of railway tracks and switches.

Two methods for welding repair could be used:

- Manual welding. Manual or automatic welding process with consumable electrode is used to repair railway tracks. The completed works are continuously checked by the use of non-destructive test methods because the quality of work is function of the welding staff capacity. These processes are used in contact surface to repair defect at the initial stage and it is always necessary to make a grinding in the weld zone after the weld process.
- Union welding. This process is similar to joining a continuous welded rail. As in manual weld, it is always necessary to make a grinding in the weld zone after the weld process and before the end of the repair works.

Rail grinding is considered the single most effective maintenance practice to control the effects of rolling contact fatigue, restore profile, and maximize value from the rail asset.

The substantial return on investment from rail grinding is well documented and includes:

- Extended rail life.
- Fuel savings.
- Reduced surfacing cycles.
- Extended track component life.
- Reduced wear on rolling stock.
- Increased axle loads.
- Increased train speeds.
- Improved ride quality and passenger comfort.

Railroads everywhere are facing continued challenges of maintaining track in shortened work blocks with limited resources. A proper rail grinding program is a key component to a maintenance plan. Rail-grinding equipment may be mounted on a single self-propelled vehicle or on a dedicated rail grinding train which, when used on an extensive network, may include crew quarters. The grinding wheels, of which there may be more than one hundred, are set at controlled angles to restore the track to its correct profile. The machines have been in use in Europe since the early

1990s. The rail grinding could be made in large extension of rail line or only in a short location of rail line. There are different types of machine to adjust the rail profile to theory profile according to the extension of the grinding work.

When the rail defects are very severe and the rail profile cannot be recovery by other process, it is necessary the **rail replacement** in the damage section. In this case, it is always necessary to use welding to join the new profile of rail with the existing one. After that, some local grinding on the join is required to achieve appropriate rail profile.

Rail tamping can restore the ballast, the initial sleeper's position, the rail geometry and the vertical rail deviation. The tampering activity, obtained by introducing vibrating blades in the ballast. The tamping is performed using a train called tamper train. In addition, the restoration of the ballast may be executed by refilling stones. The speed of the tamping and ballast cleaning machines is around 0.4-1 km/h.

Regarding **sleepers and fastening maintenance**, it is worth mentioning that the sleepers usually have no more maintenance than replace them when they are exhausted, while the defects of fastenings (rigid or elastic) are usually isolated and must be repaired manually. The operations for fastenings consist in the change of the fastening and, in case of timber sleepers, filling the screw hole with synthetic material. On the other hand, elastic fastenings can be modified, in order to correct horizontal rail deviations, when they are detected by the rail geometry inspection.

Instead, the most typical and frequent activity related to **switches maintenance**, is the lubrication of the mechanical elements that constitute them. Another maintenance activity is the replacement of mechanical elements subjected to wear and aging.

Finally, the **ballast maintenance operations** can be resumed as follows:

- Restoration of its initial geometry or curb profile. The restoration of the ballast geometry is performed by the **ballast profiling machine** which is a special train which can pull, move or replace the ballast by its mechanical blades, and give the ballast the adequate cross section profile.
- Restoration of ballast under the sleepers. The vibrations of trains slowly expel the ballast under the sleepers; the **ballast tamper machine** must return it back by its blades. In addition, the tamper machine can restore the initial sleeper’s position, the rail geometry and the vertical rail deviation.
- Clean the ballast from broken stones using the ballast cleaning machine.

In fact, as mentioned, ballast degradation is a complex mechanism that includes different phenomena:

- Broken edges. The train loads, especially at the beginning of the exploitation, break the stones edges, and decrease the material stiffness.
- Mixing with fine soils and vegetation, which affect the drainage, decrease the stiffness and accelerate the stone movements. The fine soils can be produced by stone cracks and subgrade contamination.
- Stones displacement. The concentrated charges under the sleepers can displace the stones, making an empty space just in the most solicited point.

When the ballast does not work perfectly, the track loses stiffness and vertical regularity. The vertical defects of wavelength greater than 3 meters are usually due to ballast degradation. It must be corrected by tamping machines. As mentioned, the tamping activity is obtained by introducing vibrating blades in the ballast.

In addition, the restoration of the ballast may be executed by refilling stones. The ballast cleaning machines are able to reach around 25 cm below the sleepers and grab stones smaller than 25 mm in order to replace them. Different ballast machines can work together, in a “tamping train”, which increases the performance of all of them. The speed of the tamping and ballast cleaning machines is around 0.4-1 km/h.

3. BASICS ON MATHEMATICAL PROGRAMMING

The objective of this section is to describe the mathematical concepts that are used for maintenance planning. In particular, this is done via the introduction of:

- common features of the optimisation models which are subsequently used to formalise the decision-making tasks,
- algorithmic solution approaches to be applied in decision support systems and tools.

As mentioned before, the objective of this thesis is the study of a maintenance activity scheduling problem, in which not only costs, but also risk issues are considered. The problem is analyzed as a Mixed Integer Programming problem (MIP). Moreover, it has been proven to be a non-polynomial time hard problem (NP-hard). This makes the problem solvable in a reasonable time only for very tiny instance dimensions. Then, it is necessary the implementation of "heuristic" approaches so as to find good solutions in short times, even if sub-optimal. Before showing the core of the problem formulation, it seems to be interesting giving a brief exposition of the main concepts of MIPs, Machine Scheduling, heuristic and matheuristic approach.

Therefore, first, a brief overview on the theoretical background and a brief dissertation on Mathematical Programming is proposed. In particular, the focus is on Mixed Integer Linear Programming and its application in scheduling problems.

3.1 MIXED INTEGER LINEAR PROGRAMMING

The "classical" theory of mathematical optimisation deals with the following abstract and generalised problem: given a set of decisions, find amongst the "feasible" choices one that is "best". Formally, this is expressed as follows: find an ***optimum solution*** $x^* \in M$ such that

$$f(x^*) \leq f(x) \qquad \forall x \in M$$

This compact representation includes all three ingredients of a mathematical optimisation problem: The ***degrees of freedom*** x for which a decision has to be made, the objective function $f(x)$ to be achieved, by which the different choices can be compared to each other, and the restrictions to be met, expressed by the set of feasible solutions (or choices) M. The set M normally is constructed by a number of constraints in the form

$$h_i(x) \leq 0, \qquad i = 1, \dots, m$$

or $h(x) \leq 0$ where $h(x)$ is considered as the vector

$$h(x) = (h_1(x), \dots, h_m(x)).$$

The Mixed Integer Mathematical Programming is the part of mathematical optimization that make possible to formulate many complex optimization problems in which some variables are restricted to be integer.

A MIP problem is defined by a set of variables x and an objective function ($min\ c^T x$) subject to a set of linear constraints ($Ax = b$) and a set of integrity constraints on part of the variables.

A typical formulation is:

$$Z = min\ c^T x \qquad (3.1.1)$$

subject to:

$$Ax + Gy = b \qquad (3.1.2)$$

$$x \geq 0; \qquad x \in Z \qquad (3.1.3)$$

$$y \geq 0; \qquad y \in Z \qquad (3.1.4)$$

where $A \in \mathbb{R}^{mn}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$and $I \subseteq \{1, .., n\}$.

The set $X = \{x \in \mathbb{R}^n Ax \leq b, x \geq 0, x_i \in \mathbb{Z}, \forall i \in I\}$ represents the feasibility region.

This kind of problems are also called Combinatorial Problems. Mixed integer optimization models are suitable to describe situations in which the objective is to optimize the use of nondivisible resources or the choice between discrete alternatives. Frequently, the formulations present Boolean variables, so as the integer variable assume only $0 - 1$ values. These cases, and in particular the one proposed in this thesis, are called $0 - 1$ Programming. In general, binary values are used to describe the occurrence (or not) of the considered event:

$$x = \begin{cases} 1 & if\ the\ event\ occurs \\ 0\ otherwise \end{cases} \qquad (3.1.5)$$

3.2 STOCHASTIC PROGRAMMING

Classical optimisation theory covers a number of specific problem cases, e.g. linear, quadratic, non-linear, combinatorial optimisation, and provides methods and techniques to find optimum or near-optimum solutions in a deterministic and static way: ***Deterministic*** means that it is assumed that all values and parameters occurring in the problem description are known and fixed, ***static*** refers to the fact that it is looked for a single solution for the problem at hand only, the application of the solution method is a one-time procedure.

The optimisation problems that are typically involved in maintenance planning differ from this classical view of the optimisation theory due to the presence of what is called uncertainty: Decisions must be made in face of some unknown elements, without the full knowledge of their consequences. In particular, decisions made now will have consequences that become known in the future only, but there might be opportunities to (partly) correct decisions or adapt them to new situations when more information becomes available. Basically, there are two ways how uncertainties can emerge:

- ***Stochastic***: Some of the variables or parameters in the problem description are random variables, thus the objective function or the restrictions are of stochastic nature.
- ***Dynamic***: The solution of the problem consists in a sequence of decisions and their updating or adaption is performed, once new information becomes available. Thus, the solution procedure is a dynamic process.

In complex decision-making scenarios like maintenance planning mostly a combination of stochastic and dynamic uncertainties has to be modelled.

There are several mathematical concepts dealing with this kind of optimisation under uncertainty, which are the background of the modelling and solutions techniques applied in the risk-based maintenance planning concept.

The concept of **multi-stage programming** [26] consider the case when the decision-maker is able to react on changed situations and new information by adapting the solutions found, dealing with dynamic uncertainty in a very explicit manner: a stage-based, recurrent process of decisions followed by observations is applied, using additional decision variables in each stage.

In the simplest case, the two-stage approach, the decision variables are split into two sets - first stage decisions $x = (x_1, \dots, x_n) \in X$ and second stage decisions $y = (y_1, \dots, y_n) \in Y$ - with the following meaning:

The first stage variables x have to be decided before the realisation of some uncertain parameters can be observed by the decision-maker, thus only with knowledge available in the current situation. Then, after observations leading to new information, a so-called recourse action to react on the changed situation and adapt the overall solution can be made, by deciding on the second stage variables y. The decisions made in the first stage contribute to an objective value $f(x)$, but the recourse action implies an additional value - which can be seen as an additional "cost" of recourse, depending on the realisation of the uncertain parameters as well as on the first and second stage variables: $g(x, y, \xi)$.

The second-stage variables are considered to be determined in a separate optimisation problem, the second-stage problem, giving the optimal value

$$G(x, \xi) = \min_{y} g(x, y, \xi)$$

which is formally used to express the first stage problem:

$$\min_{x} \{ f(x) + \mathbb{E}_{\xi} \left[G(x, \xi) \right] \}$$

i.e. find the solution x^* of first stage variables that minimse the first stage objective plus the expected minimum recourse cost.

Of course, this approach can easily be extended to a general multi-stage process. Here, the solution process consist of the sequence: stage 1 decisions → observation → stage 2 decisions → ... → stage m decisions, and can be expressed mathematically by using variables $x^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})$ for each stage $i = 1, \dots, m$ and recourse functions $f^{(i)}(x^{(1)}, \dots, x^{(i)}, \xi)$ for stages $i = 2, \dots, m$ depending on variables up to stage i (and the uncertainty ξ). Again, stages are solved consecutively using the expected minimum solution from the previous (higher) stage.

One important remark is that still the multi-stage stochastic programming concept defines so-called a-priori-solutions, i.e. solutions explicitly consider dynamic adaptions (in the form of recourse actions), but that are selected in advance of all realisations of uncertain parameters. Thus, they provide means to dynamically planning, but they heavily rely on the validity of all probabilistic information used in the problem formulation even during the execution of solutions.

3.3 SCHEDULING PROBLEM

MILP models are used in a great set of applications, but the scope of this dissertation is focused on a peculiar kind of formulation: sequencing/scheduling problems.

The basic concepts and notation of scheduling problem are introduced (Graham et al. [27] and Blazewicz [28]). In scheduling problem, a set of n jobs/tasks $j = 1, \ldots, n$ and M machines/processors $m = 1, \ldots, M$ are considered. In the scheduling literature, the terms "task" and "job" are often used interchangeably, although in some cases, tasks are decomposed into separate parts called jobs. Each machine may work on a single job at a time, and each job may be processed by a single machine at a time. The schedule is a list of jobs with the times when the jobs are processed by machines, and a feasible schedule satisfies the timing requirements as well as the fundamental assumptions described above.

In one-machine environment there is only one machine that can process one job at time.

If each job must be processed in an uninterrupted time period, the schedule environment is nonpreemptive, whereas, if a job may be processed for a period of time, interrupted and continued in a later point in time, the scheduling is preemptive.

Each task j has the following properties (Conway et al. [29]):

- the task has a vector of processing times with each element of the vector corresponding to the processing on a particular processor, $[p_{j1}, p_{j2}, \ldots, p_{jM}]$,
- the task has an arrival time or ready time, r_j,
- the task has a due date or deadline, d_j
- the task has a weight or priority, w_j,
- the task may be preemptive or non-preemptive, depending on whether preemption is allowed in the schedules (preemption is also referred to as "task splitting"),

- the task may be dependent or independent. Dependence between tasks is specified by means of a precedence tree or a more general precedence graph.

Therefore, a schedule is an assignment of processors to tasks. At each moment, at most one task is assigned to each processor, and at most one processor is assigned to each task. Each task is processed after its arrival time, and all tasks are completed.

The peculiarity of a scheduling problem consists in the way the variables are used to describe precedence/succession constraints, which are defined "disjunctive constraints".

This definition is related to the fact that, while usually MILP formulations require the satisfaction of the whole set of constraints, in this case only a subset is needed to be respected. This is important when there are sequencing activities that cannot be processed at the same time. In fact, supposed to have n operations to be sequenced on a unitary capacity machine. Let p_i be the processing time of the i −th task on the m −th machine, t_i the starting time of the i −th task. So, if the i −th task precedes the j −th, then: $t_j \geq t_i + p_i$.

Vice versa, if j precedes the i then $t_i \geq t_j + p_j$. The representation of these constraints is possible thanks to the following binary variables:

$$x_{ij} = \begin{cases} 1 \ \ if\ i\ is\ processed\ before\ j \\ 0\ otherwise \end{cases} \qquad (3.3.6)$$

Then, the constraints are formulated in this way:

$$Bx_{ij} + t_i - t_j \geq p_j \qquad 1 \leq i < j \leq n \qquad (3.3.7)$$

$$B(1 - x_{ij}) + t_i - t_j \geq p_i \qquad 1 \leq i < j \leq n \qquad (3.3.8)$$

where B is a big real number representing the infinite.

It is clear that, if $x_{ij} = 1$ then the first constraint is always satisfied because $B \gg p_j + t_j - t_i$, while the latter expresses the starting time for the j −th activity. Vice versa when $x_{ij} = 0$. Starting from this formulation for the activity sequencing, it is possible to construct a scheduling problem through the introduction of temporal constraints.

The performance characteristics and performance measures of individual tasks and of schedules can be defined. Each task in a schedule can have:

- a completion time which we denote as C_j,
- a flow time, denoted $F_j = C_j - r_j$,
- a lateness, denoted $L_j = C_j - d_j$
- a tardiness, denoted $T_j = \max(L_j, 0)$
- a unit penalty $U_j = 0 \; if \; C_j \leq d_j$, else 1.

These properties of schedules not only provide measures for evaluating schedules, but also provide criteria for optimization in algorithms that produce schedules.

In particular, schedules are evaluated using

- schedule length or makespan, $C_{max} = max(C_j)$;
- mean flow time, $\bar{F} = \frac{1}{n}\sum_{j=1}^{n} F_j$
- mean weighted flow time, $\bar{F}_w = \frac{\sum_{j=1}^{n} w_j F_j}{\sum_{j=1}^{n} w_j}$
- maximum lateness, $L_{max} = max(L_j)$;
- mean tardiness, $\bar{\mathrm{T}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{T_j}$
- mean weighted tardiness, $\bar{\mathrm{T}}_{\mathrm{w}} = \frac{\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{w_j T_j}}{\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{w_j}}$
- number of tardy tasks, $\bar{\mathrm{U}} = \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{U_j}$

Graham et al. [27] defined a notation for identified scheduling problems.

There are three fields in the problem specification separated by vertical lines or bars:

$$\alpha/\beta/\gamma$$

here α is a string describing the processor environment (uniprocessor, multiprocessor, job shop, etc.), β gives other details about the scheduling environment (preemption, precedence constraints, resource constraints, etc.), and γ specifies the criterion for optimization.

Each of these three fields has subfields, which may or may not include a symbol. The symbol φ is used to indicate the case where no symbol appears in a subfield; this is usually a “default” case.

Hereafter, the different possible values of the three fields are described in details.

$$\alpha = \alpha_1\alpha_2$$

$$\alpha_1 \epsilon \{\phi, P, Q, R, F, J, Q\}$$

For $\alpha_1 \epsilon \{\phi, P, Q, R\}$ each task J_i has a single operation to be executed on processor M_i, and the processing time is p_{ij}

$\alpha_1 = \phi$	single processor $p_{1j} = p_j$, the index for the processor is dropped since there is only one.
$\alpha_1 = P$	identical parallel processors, $p_{ij} = p_j(\ i = 1, .., m)$, the index is dropped since the processing time does not depend on the actual processor used.
$\alpha_1 = Q$	uniform parallel processors, $p_{ij} = q_i p_j(\ i = 1, .., m),)$, where q_i is the speed factor of processor M_i which indicates its (constant) relative processing speed.
$\alpha_1 = R$	unrelated parallel processors.
$\alpha_1 = F$	flow shop, each J_i is composed of a chain of operations $\{\ O_{1j\ldots mj}\}$ where O_{ij} is to be processed on M_i for the computation time p_{ij}, the order in which operations are serviced is fixed by the ordering of the chain.
$\alpha_1 = J$	job shop, each J_i is composed of a chain of operations $\{\ O_{1j\ldots mj}\}$ where O_{ij} is to be processed on processor μ_{ij} for the computation time p_{ij} and $\mu_{i\ldots.1j} \neq \mu_{ij}$ $(\ i = 2, .., m_j)$, the order in which operations are serviced is fixed by the ordering of the chain, but the order may be different for different tasks.
$\alpha_1 = O$	open shop, each J_i is composed of a set of operations $\{\ O_{1j\ldots mj}\}$ where O_{ij} is to be processed on M_i for the computation time p_{ij}, the order in which operations are serviced is arbitrary.

$\alpha_2 \epsilon \{\phi\} \cup \mathcal{N}$ where $\mathcal{N}$ is the set of natural numbers (positive integers).

$\alpha_2 \epsilon \mathcal{N}$	m is constant and equal to α_2.
$\alpha_2 = \phi$	m is variable.

Note that when $\alpha_1 = \phi,\ \alpha_2 = 1$

$$\beta = \beta_1\beta_2\beta_3\beta_4\beta_5\beta_6$$

$$\beta_1 \epsilon \{pmtn, \phi\}$$

$\beta_1 = pmtn$	preemption is allowed in the servicing of the tasks.
$\beta_1 = \phi$	preemption is not allowed.

$$\beta_2 \epsilon \{res, res1, \phi\}$$

$\beta_2 = res$	the system includes s resources, $R_h(\ h = 1,..,s))$ in addition to the processor; Each task J_i needs r_{hj} units of resource R_h for the duration of its service.
$\beta_2 = res1$	the system includes a single additional resource ($s = 1$).
$\beta_2 = \phi$	no additional resources are required by the tasks.

$$\beta_3 \epsilon \{prec, tree, \phi\}$$

$\beta_3 = prec$	a (general) precedence relation exists between the tasks.
$\beta_3 = tree$	a precedence tree describes the precedence relation between tasks.
$\beta_3 = \phi$	there is no precedence relation for the tasks; the tasks are independent.

$$\beta_4 \epsilon \{r_j, \phi\}$$

$\beta_4 = r_j$	arrival times are specified for each task.
$\beta_4 = \phi$	$r_j = 0\ (\ j = 1,..,n)$ all tasks are released at the same time.

$$\beta_5 \epsilon \{m_j \leq \overline{m}, \phi\}$$

$\beta_5 = m_j \leq \overline{m}$	$\overline{m}$ is a constant upper bound on m_j, the number of processors needed for the operations of a task (only in the job shop, $\alpha_1 = J$)
$\beta_5 = \phi$	there is no bound on m_j.

$$\beta_6 \epsilon \left\{\ p_{ij} = 1,\ \underline{p} \leq p_{ij} \leq \overline{p}, \phi\right\}$$

$\beta_6 =\ p_{ij} = 1$	unit processing time

$\beta_6 = \underline{p} \leq p_{ij} \leq \overline{p}$	processing time is bounded by constants $\underline{p}$ and $\overline{p}$
$\beta_6 = \phi$	no bounds on processing time

$$\gamma \epsilon \left\{ C_{max}, L_{max}, \sum C_j, \sum w_j C_j, \sum T_j, \sum w_j T_j \sum U_j, \sum w_j U_j \right\}$$

Some examples of scheduling problems are specified using the notation described above. The references given for the following problems were cited by Graham et al. [27] as sources for the original solutions or complexity proofs.

$1\|\|L_{max}$	On a single processor we wish to minimize the maximum lateness with no preemptions, no precedence constraints (independent tasks), and release dates at $t = 0$. Jackson's rule solves this problem: schedule the tasks using the earliest deadline first selection policy [30].
$1\|prec\|L_{max}$	On a single processor we wish to minimize the maximum lateness L_{max} subject to general precedence constraints on the tasks. Note that the implicit details of this specification include: there is no preemption, tasks are all released at time $t = 0$, and there are no additional resources. There is a polynomial time algorithm for this problem [31].
$1\|\|w_jC_j$	On a single processor, minimize the sum of weighted completion times with no preemptions, no precedence constraints, and equal arrival times. The solution is to use Smith's rule [32] to schedule tasks in order of non-increasing ratios $w_j = p_j$.

$1\vert\vert U_j$	On a single processor, minimize the number of tasks that miss their deadlines (no preemptions, no precedence constraints, and equal release times). Moore [33] provided the solution to this problem: schedule tasks in earliest-deadline-first order, removing the scheduled task with the largest processing time when the most recently added task fails to meet its deadline.
$P2\vert\vert C_{max}$	On 2 identical parallel processors, minimize maximum completion time (with no preemptions, equal release times, and no precedence constraints). This problem is NP-hard [34][35].
$P\vert pmtn\vert C_{max}$	On identical parallel processors, minimize maximum completion time with preemptions allowed (with equal release times and no precedence constraints). McNaughton gave a simple $O(n)$ algorithm [36].
J3$\vert$ p_{ij} = 1$\vert C_{max}$	In a 3-machine job shop, the objective is minimize maximum completion time where tasks have unit processing time. Preemptions are not allowed and tasks are all released at time $t = 0$. This problem is NP-hard [37].

Although for some simple or restricted problems an optimal solution exists, most scheduling problems are very difficult and have been shown to be NP-hard [38].

To try to find approximate solutions to practical problems, various techniques are used which typically relax some of the assumptions that make the problem difficult. For example, it is easier to produce a schedule if preemption is allowed than if preemption is not allowed. So one technique for approximation is to allow preemption in a problem specified as non-preemptive. In this way, the basic shape of the preemptive schedule can be used to construct a schedule with no preemption.

Approximation algorithms are also available which help to organize the search for a solution to a problem, avoiding the complete enumeration, whereas, the technique is to just use an enumerative algorithm if the problem is sufficiently small.

3.4 PARALLEL MACHINE SCHEDULING

In parallel scheduling problems m machines are given. The processors are in one of two configurations: parallel or dedicated (specialized).

In the case of parallel processors, it is necessary to distinguish between identical processors, uniform processors, and unrelated processors. Identical processors have speeds which are constant and which do not depend on the task in service. Uniform processors have constant speeds but the speeds of individual processors may be different; the processor speed does not depend on the job. With unrelated processors, the speed depends on the job.

In the case of dedicated processors, it is necessary to distinguish between flow shops, open shops, and job shops. This nomenclature is taken from the industrial management literature [39].

In the flow shop, each task is processed by all processors in the same order.

In the open shop, each task is processed by all processors, but the order of processing is arbitrary.

In the job shop, each task is processed by an arbitrary subset of the processors, and the order is arbitrary. The specification of the subset and the order is fixed *a priori*.

In this study, maintenance planning is modelled as the scheduling of a set of independent jobs (or maintenance activities) on a set of unrelated parallel machines (or maintenance teams) with sequence dependent setup times. The issue is how to temporally allocate the given jobs to the available machines in order to minimize the total weighted completion time. Actually, the model developed in this thesis, and described in the following Section 6, is complex, but it is interesting to give a background on this kind of problems because they represent the "core" of the presented formulation.

Therefore, the “core” model of the maintenance planning problem is an unrelated parallel machines scheduling based on the following basic assumptions:

- The jobs are simultaneously available at the beginning of the scheduling horizon.
- Each job can be processed on any machines but needs to be processed by one machine only.
- Job preemption is not allowed.
- No processing precedence among jobs on any machine exist (open shop scheduling)
- Each machine can process at most one job at a time.
- The setup times are dependent only on jobs sequences and are machine-independent.
- The scheduling objective is to minimize total weighted mean completion time.

This problem is at least NP-hard in the ordinary sense, since the special case when there are only two identical machines with no setups is NP-hard in the ordinary sense [35].

3.5 RESOLUTION APPROACHES

In this paragraph, the main scheduling resolution approaches are presented. The resolution approaches are in general classified into the following categories:

- Manual/Simulation dispatch;
- Exact Mathematical approaches;
- Heuristic approaches;
- Hybrid approaches.

Manual-dispatch scheduling arises when precise matching of a small set of resources and activities is needed.

Simulation-dispatch essentially computerizes a very simple version of manual-dispatch.

They are both easy and workmen-friendly instruments, but poorly efficient.

Exact mathematical approaches choose an objective to be optimized (makespan, tardiness, utilization etc.,), formalize the resources and constraints, and solve the problem by mathematical programming techniques. This category includes: Integer Programming approach, Dynamic Programming approach, Lagrangian Relaxation approach and Branch and Bound Technique's.

Heuristic approaches try to solve approximately the mathematical problem thus formulated. They include searching techniques such as: Tabu search, Simulated annealings, Genetic algorithms and Approximation algorithms.

Hybrid approaches are methodologies in which two or more techniques are implemented in synergy. The interest is here especially focused on that kind of hybridization that are based on combining Mixed Integer Programming and heuristics.

Regarding the **exact approaches**, an exact resolution technique consists in the explicit enumeration of all solutions for the considered problem, i.e. calculating the value of the objective function in each point of the feasibility region and then choosing the best solution between them. Therefore, this method is applicable only for very small instances.

A consolidated approach is given by the Branch & Bound algorithm. It consists in the partial exploration of the feasible solution set. In particular, the objective function is solved for a subset of the feasible solutions which contains at least an optimal solution. This is possible thanks to the use of bounds for the function to be optimized combined with the value of the current best solution so that it is possible to search parts of the solution space only implicitly. Suppose 10 jobs are to be sequenced on one machine, and that there is a function to evaluate how good a given schedule is. Conceptually, explicit enumeration creates a decision tree with 10 branches for the possible choices for second job, leaving 10 x 9 = 90 second level branches. Then the tree branches again and again. Therefore, finally there are 10! branches at the 10-th level, which is a huge number again. Branch & Bound, instead, is based on the idea that, instead of looking at all possible solutions and evaluating them, it is possible to prove that certain parts of the tree can simply be chopped off or pruned, since they can be shown to have only non-optimal solutions. At any point during the solution process, the status of the solution with respect to the search of the solution space is described by a pool of yet unexplored subset of this and the best solution found so far. Initially only one subset exists, namely the complete solution space. The unexplored subspaces are represented as nodes in a dynamically generated search tree, which initially only contains the root, and each iteration of a classical B&B algorithm processes one such node. The iteration has three main components: selection of the node to process, bound calculation, and branching. The sequence of these may vary according to the strategy chosen for selecting the next node to process. If the selection of next sub problem is based on the bound value of the subproblems, then the first operation of an iteration after choosing the node is branching, i.e. subdivision of the solution space of the node into two or more subspaces to be investigated in a subsequent iteration. For each of these, it is checked whether the subspace consists of a single solution, in which case it is compared to the current best solution keeping the best of these. Otherwise, the bounding function for the subspace is calculated and compared to the current best solution. If it can be established that the subspace cannot contain the optimal solution, the whole subspace is discarded, else it is stored in the pool of live nodes together with its bound. The search terminates when there are no unexplored parts of the solution space left, and the optimal solution is then the one recorded as “current best”. A B&B algorithm for a minimization problem hence consists of three main components:

- a bounding function providing for a given subspace of the solution space a lower bound for the best solution value obtainable in the subspace,
- a strategy for selecting the live solution subspace to be investigated in the current iteration,
- a branching rule to be applied if a subspace after investigation cannot be discarded, hereby subdividing the subspace considered into two or more subspaces to be investigated in subsequent iterations.

The bounding function is the key component of any B&B algorithm. Ideally the value of a bounding function for a given subproblem should equal the value of the best feasible solution to the problem, but since obtaining this value is usually in itself NP-hard. Unfortunately, the problem considered in this dissertation is characterize by very weak bounding function, so that the exact algorithm is not efficient when instance dimensions grow. For example, no computer is available that could check out all permutations of 50 jobs on one machine and, even if such an extraordinary machine were developed, to solve a problem with 55 jobs would require a computer about 300,000,000 times faster. Of course, B&B and modern solvers are cleverer than this, but running times do still go up exponentially with the problem size in the same way.

Therefore, the application of such method is not suitable. For this reason, many scheduling heuristic techniques have been used for managing preventive maintenance activities.

Heuristic problem solving techniques range from simple constructive techniques such as ad-hoc greedy algorithms over local search methods to various metaheuristics. Especially the latter category is well-developed and has proven to be highly useful in practice. As their name suggests, metaheuristics are defined on a higher and basically problem independent level. A meaningful criterion is the division into single-solution based methods (i.e. following a single search trajectory), which are often sophisticated variants of local search either using a single neighborhood or several ones, and population based methods (i.e. multiple search trajectories, usually running in an intertwined way).

Prominent examples of the former class are variable neighborhood search (VNS), tabu search (TS) and simulated annealing (SA), while the broad class of evolutionary algorithms (EAs), swarm intelligence methods such as ant colony optimization (ACO) algorithms and particle swarm optimization (PSO), as well as scatter search (SS) belong to the latter class. Another criterion is whether the solutions are primarily constructed (e.g. greedy randomized adaptive search procedures or ACO) or improved (e.g. VNS, TS, SA, EAs, PSO, SS) [40].

The explanation of this kind of modelling methods are beyond the scope of this thesis. The interested reader shall refer to specialized texts.

Another particular approach is based on the hybridization of mathematical programming and heuristic methodologies. That is the approach used in the present research.

Exact techniques can generate provably optimal solutions for optimization problems, while the field of heuristics is introduced to generate “solutions” for optimization problems that should be “good” but not necessarily provably optimal. The objective of this paragraph is to explore the relationship between these two approaches to problem solving and, in particular, to introduce the idea of hybrid approach that is the basis of the work made in this thesis. Puchinger and Raidl [41] give a useful classification scheme and many examples for such hybrids, distinguishing between the following two main categories (defining the control strategy):

- collaborative (cooperative) combination, where the two methods exchange information, but are not part of each other, thereby running in sequential order or being executed in a parallel or intertwined way;
- and integrative (coercive) combination, with a distinguished master and at least one integrated slave algorithm, where exact algorithms are incorporated in (i.e. are subordinates of) metaheuristics and vice versa.

In a subsequent work [42] the same authors especially focus on hybrids between metaheuristics and (integer) linear programming techniques. More recently such hybrids, exploiting in a suitable way the mathematical model of the problem, are often called matheuristics [43], or, due to their nature, model-based metaheuristics. An interesting survey on matheuristic approches is proposed by Ball [44]: the author identify four classes of methodologies explored in literature.

- The **first class** of methods break down a problem into a sequence of subproblems where each subproblem is modelled as a mathematical program and solved optimally.
- The **second class** of methods are improvement algorithms that solve a mathematical problem to generate an improved solution from a known feasible solution; this class of techniques is also referred to as large-scale neighbourhood search.
- The **third class** of methods employ a mathematical programming algorithm, most notably branch-and-bound, to generate an approximate solution to the problem of interest.
- Finally, the **last class** of methods solve a relaxation to the original problem of interest as a first step in generating a good feasible solution.

The algorithm proposed in this thesis may be inserted in the second class abovementioned. In particular, it can be classified as a Row Partitioning algorithm. The idea is to finding the best solution over a restricted feasible region. More details are given in the following chapters.

Another detailed survey is proposed by El-Ghazali [45], who presents a detailed taxonomy of the heuristic and hybrid approaches. In his work, the author proposes an interesting matheuristics classification:

- **Low-level relay hybrids** (LRH): This class of algorithms represents hybrid schemes in which a metaheuristic approach (resp. exact approach) is embedded into an exact approach (resp. Single Solution-Based metaheuristic approach) to improve the search strategy.
- **Low-level teamwork hybrids** (LTH): In this class of hybrid algorithms, a search component of a Population-Based metaheuristic is replaced by another optimization algorithm. Concerning the combination of P-metaheuristics and MP algorithms, two main hybrid approaches may be considered: exact search hybrid algorithms in which a P-metaheuristic is embedded into an exact algorithm, and heuristic search algorithms in which an exact algorithm is embedded into a P-metaheuristic.
- **High-level relay hybrids** (HRH): In this class, self-contained algorithms are used in sequence, i.e. some information is provided between the two families of algorithms (metaheuristics and MP algorithms) in a sequential way. In the case where the information is provided by the metaheuristics, the most natural and trivial hybrid approach is to start with a metaheuristic to find a "good" bound which will be used by a MP algorithm in the bounding phase; therefore, metaheuristics can be used to reduce the size of the original problem. Then, the exact method can be applied to solve the reduced problem. This allows the reduction of the problem into much smaller problems which can be solved exactly by state-of-the-art mathematical programming algorithms. An example is the partitioning of decision variables: In this strategy, the decision variables are partitioned into two sets X and Y. The metaheuristic will fix the variables of the set X and the exact method will optimize the problem over the set Y. Hence, the generated subproblems are subject to free variables in the set Y and fixed variables in the set X. Those subproblems are solved exactly. A set of high quality solutions may be obtained by a P-metaheuristic or an iterated S-metaheuristic

- **High-level teamwork hybrids** (HTH): this class of hybrids which combines metaheuristics and MP algorithms in a parallel cooperative way. It is a promising class of hybrids, but still mostly unexplored.

The approach proposed in this thesis may be included in the third case and it is based on the idea of Della Croce *et al.* [46], as described in Chapter 8.

3.6 LOWER BOUND FOR PARALLEL MACHINES SCHEDULING PROBLEMS

Consider the situation where a set of n jobs $N = \{1, \dots, n\}$ has to be processed, m parallel machines are available and the objective is to minimize a sum objective function. To each job i, a release date r_i, a processing time p_i, a due date d_i and a weight w_i are associated. Preemption is not allowed and no more than m jobs can be scheduled simultaneously. The tardiness of job i is defined as $T_i = \max(C_i - d_i, 0)$, where C_i is the completion time of job i. As mentioned, several criteria can be considered: minimizing the maximum completion time (makespan), the total (weighted) tardiness and total (weighted) completion time. The goal of the makespan problem is to find a feasible schedule of the n jobs that minimizes C_{max}. A natural extension of the makespan problem is the weighted completion time scheduling problem, where in addition each job i has a positive weight w_i that expresses the importance of that job. These scheduling problems on parallel processors are often intractable and combinatorial lower bounds are needed to guide solvers and evaluate heuristics [47]. Indeed, as scheduling problems are strongly NP-Hard [37] it is essential to have good and fast lower bounds. In this study, a brief survey of existing bounds is proposed.

Some possible strategies to obtain a good lower bound are:

- **Relaxing integrity constraint:** the integrity constraint is relaxed to compute a lower bound.
- **Lagrangian relaxation on resource constraint:** the Lagrangian relaxation on the resource constraint is used to get a lower bound. In particular, constraint of non temporal overlapping is relaxed to be put in the objective function and the Lagrangian multipliers are introduced.
- **Lagrangian relaxation on the number of occurrences constraint:** the constraint stating that a job has to be executed exactly once is relaxed. In this new problem, the optimal schedule

on m parallel machines, where jobs are allowed to be “unprocessed” or to be processed several times, is found. To improve this lower bound, a constraint, stating that a job cannot be processed twice consecutively, can be added.

- **Relaxing all release dates to the earliest release date:** All release dates are relaxed to the minimal release date among all release dates. The relaxed problem reduces to $P_m || \sum C_j$ which is polynomially solvable [48]. All jobs are scheduled according to the Shortest Processing Time (SPT) Rule on the earliest available machine. That is to say that the job with the shortest processing time is scheduled on the machine which is available the earliest. This lower bound can be then computed in $O(n\ log\ n)$.
- **Splitting problem into relaxed sub-problems:** A lower bound is obtained in which several subsets of jobs are build according to the value of their release dates. The main idea is to relax the release dates of jobs according to the minimal release date among the jobs of the same subset, building the optimal solutions of these subsets independently using the SPT rule. The sum of the costs of each of these schedules is then a lower bound of the original problem. This bound can be computed in $O(n2)$ since at most n lower bounds in $O(n)$ are computed, after a single step of sorting jobs in non-decreasing order of processing times.
- **Job splitting:** This lower bound consists in allowing preemption and “Splitting” i.e. simultaneous execution of parts of a same job on several machines [49]. The optimal schedule of this relaxed problem is built by sequencing jobs according to the Extended Shortest Remaining Processing Time rule: at time t, the job k with the shortest remaining processing time is chosen. Then, unit parts of job k are scheduled as soon as possible, possibly on different machines during the same time unit, until the job is completed or a new release date has been reached. The $i-th$ lowest completion time of a job in this optimal schedule is a lower bound.

In [50] the lower bound for a resource-scheduling problem with limited working time are considered:

- The **surface bound** takes the duration of the tasks and the length of the work period into account, but completely ignores the start times and the resource assignment. A very simple lower bound is obtained by calculating the overall amount of work to be scheduled and to divide this by the maximum length of the work period. In principle the sum of all the durations of the tasks can be considered and this number can be divided by the length of the work period. This type of lower bound will be quite good if there are few constraints on the sequence of tasks and the task duration is small compared to the length of the work period. The lower bound calculated is just the number of resources needed to perform all tasks, if no resource is kept waiting and all resources work during the full span of the work time. If the start times of the tasks are constrained a lot or the tasks cannot be combined into groups with the duration of the work period, this lower bound will be too optimistic.

- The **peak bound** is based on the start dates and the duration of the tasks, but ignores the work period length and the resource assignment. Instead of assigning tasks to multiple disjunctive resources, one cumulative resource is considered to which we assign tasks with a resource usage of 1. The limit is obtained by the peak use of the cumulative resource at any one time point.
- The **bin packing** bound uses the duration of the tasks, the work period limit and an estimation of the resource assignment, but ignores the start date of the tasks. Each resource is a bin with fixed size W. Each task is an item that should be put in one bin and which uses Duration space in the bin. A lower bound of the number of bins required for a given set of items is found.

Before entering the core of the dissertation, a literature review is provided.

4. LITERATURE REVIEW

In this chapter, a brief overview of the state of art is proposed. The chapter is divided into three sections: the first is related to the parallel-machine scheduling problems in general, the second focuses on their application in railway maintenance planning, from a deterministic and stochastic point of view, and the third presents the most recent European projects on railway maintenance.

4.1 LITERATURE REVIEW ON PARALLEL MACHINE SCHEDULING

The first comprehensive survey paper on scheduling problems with separate setup times or costs was conducted by Allahverdi *et al*. [51] who reviewed the literature since the mid- 1960s.

Since the appearance of that survey paper, there has been an increasing interest in scheduling problems with setup times (costs) with an average of more than 40 papers per year being added to the literature. More recently, Ali Allahverdi *et al*. [52] realized a detailed survey of scheduling problems in general. The paper classifies scheduling problems into those with batching and non-batching considerations, and with sequence-independent and sequence-dependent setup times. It further categorizes the literature according to shop environments, including single machine, parallel machines, flow shop, no-wait flow shop, flexible flow shop, job shop, open shop, and others. It is worth noting that the open shop class, which is the one the object of this thesis addresses, seems to be poorly investigated. Open shop scheduling problem are the focus of the review in Anand *et al*. [53]. First, the problem is classified as per different measures of performance, i.e., minimization of makespan, minimization of sum of completion times of jobs, minimization of sum of weighted completion times of all jobs, minimization of total tardiness of all jobs, minimization of sum of weighted tardiness of all jobs, minimization of weighted sum of tardy jobs, and miscellaneous measures of the open shop scheduling problem. In each category, the literature is further classified based on approaches used and then the contributions of researchers in the respective categories are presented. The study summarizes the approaches used by each considered author: the interesting fact is that only 4 researches of the 100 considered propose a hybrid approach, and not even one suggest a model-based heuristic.

4.2 LITERATURE REVIEW ON RAILWAY MAINTENANCE PLANNING

Preventive maintenance works are performed in order to reduce the probability of the occurrence of a fault and to improve the overall reliability and availability of a system. Poorly designed preventive maintenance schedules may incur high maintenance and operation costs and loss of safety. Therefore, literature gives great attention to railway maintenance scheduling problem. Authors agree that is essential to move from a planned preventive maintenance to a predictive preventive maintenance [54]. In rail world, as described in the following Chapter 5, automatic monitoring and diagnostic systems, mounted on trains and along the lines, have become over years more and more significant and sophisticated. New suitable models are necessary to use efficiently the new available data, for estimating the time when a fault is likely to occur and adapt maintenance interventions accordingly [55].

Therefore, new predictive maintenance schedule model must be able to reduce both the overall maintenance budget and the safety risk, evaluating the priority of the maintenance tasks, taking into account not only assets degradation conditions but also assets criticalities. A fault can be better tolerated by the system than another, according to the related consequences.

In past literature, the most common optimization criteria for railway maintenance are based on cost minimization, in order to reduce the overall maintenance budget. Nowadays, the aim of predictive maintenance scheduling is to minimize time duration and overall maintenance costs, planning the preventive maintenance activities and assigning them to the different working teams, but also to maximize the system reliability and availability, at the same time.

Moreover, as mentioned, railway maintenance must take into account the space-distributed aspect of railway infrastructure. Railway assets are often not spatially delimited to a point, and this implies difficulties in the organization of maintenance activities and resources. Regarding this topic, it is worth mentioning the study of Camci [56] that presents a model for the maintenance scheduling of geographically distributed assets incorporating the travel time between assets in the scheduling problem when one team is responsible for the maintenance of multiple assets.

As a general consideration coming from the relevant research, it is worth saying that preventive maintenance problems in a large-scale railroad network (i.e., typically including hundreds of segments) involves hundreds of activities and very complex relationships among them, which generates a larger number of side constraints. In addition, such optimization problems have been proven to be Non deterministic Polynomial-time Hard (NP-Hard) problems, as described in the previous Chapter, that is, a class of problems to solve which, a polynomial-time algorithm has never been found. Thus, they require too a long time to be solved and, for this reason, many scheduling heuristic techniques have been used for managing preventive maintenance activities [57], mainly based on:

- genetic algorithms;
- ontology-based modelling;
- strategic gang scheduling;
- other specific heuristic approaches.

However, when the dimension of the problem admits it, a standard solver of optimization problems can be sufficient.

More in detail, Higgins [58], Budai *et al.* [59], Peng *et al*. [60] and Borraz-Sanchez *et al*. [61] develop mathematical models aiming at determining the assignment and the schedules of maintenance teams to minimize the disruption of train operations and the infrastructure possession time. Such models take into account budget constraints, train schedules, working teams travel times, and various interrelations among maintenance activities. The solution approaches consist of heuristic approaches such as tabu search approach and simple greedy heuristic. Nevertheless, they do not take into account data about the infrastructure conditions provided by monitoring systems.

Some steps forward in this direction are done by Dell'Orco *et al*. [62], Umiliacchi *et al*. [63][64] and Jiménez-Redondo *et al*. [65]. In fact, their models aim at minimizing the probability of a mission-critical fault occurring during train service, keeping the railway track at good safety and comfort levels. They use data from special diagnostic trains and new inspection technologies to evaluate track condition.

Instead, Simson *et al*. [66], Burrow *et al*. [67] and Zhang *et al*. [68] develop tools based on predictive degradation models, providing important steps from condition-based maintenance. In these approaches, track condition data allows to predict faults and determine what maintenance work is required for safe train operation and whether safety-related speed restrictions are necessary.

These models enable long-term forecasts of rail condition and strategic maintenance decisions, also considering the uncertainties of the deterioration process but they do not consider assets criticalities and risk evaluation.

On the other hand, Carretero *et al*. [69] and Bharadwaj [70] apply reliability concepts to plan periodic maintenance, taking into account the assets criticalities by means of a FMECA (Failure Mode Effect and Criticality Analysis) procedure.

They develop risk-based methodologies to estimate the optimal time of replacement or repair of a railway structure, but they do not take into account predictive models of assets condition.

Therefore, in literature many different approaches have been proposed to schedule maintenance activities, nevertheless new steps forward are needed to adapt railway maintenance schedule both to new predictive model and to new standards of risk-based assets managements and safety. Therefore, the presented study is aimed at giving a systemic view of how scheduling models can use the significant outputs of predictive tools and degradation models, based on data from railway field, to achieve a risk-based maintenance plan.

In real environments, predictive maintenance scheduling has to deal with the occurrence of unexpected events, due to the intrinsic stochastic nature of the system.

The problem of uncertainty in scheduling process and the discrepancy between theoretical schedule and real system behavior are faced in literature evaluating the risk associated to the occurrence of unfavorable events, such as machine unavailability or delays [71][72] .

Other researches consider a dynamic rescheduling to achieve a fault recovery in real-time systems [73][74].

For what concern maintenance scheduling, Ma *et al*. deal with the fact that maintenance schedule is enacted, based on a statistical average. This still retains the unavoidable risk that the system might fail before criteria are exceeded: a fault might occur unexpectedly [75].

In railway field, rescheduling researches are applied mostly to trains disruptions [76][77] for the management of rail traffic in case of delays and unexpected events, whereas in railway maintenance sector only few studies are available about this topic. Some steps forward are moved by Baldi et al. [78] that introduce a stochastic scheduling problem for railway maintenance tactical planning, and by Andrews *et al*. [79] that develop a stochastic model for railway track asset management.

Nevertheless, much work still has to be done in order to deal with the real environment uncertainties related to risk-based predictive maintenance planning.

Therefore, the present study aims at filling this gap, applying rescheduling techniques to predictive maintenance scheduling disturbances taking into account risk. Indeed, this thesis presents a rolling horizon scheduling model [80] to deal with the stochastic nature of railway maintenance.

Nevertheless, it is worth noting that the adaptive rescheduling models only partially solve the uncertainty issue, since they consider deterministic sub-problems of the overall problem and they cannot vary continuously the stochastic input variables.

Therefore, to cope with this problem, in this work, the risk-based maintenance planning problem is also formulated in term of stochastic programming.

In particular, the so-called two-stage stochastic linear programs with recourse is applied to rail predictive maintenance. This mathematical technique has been already applied to scheduling problem in production sector, for instance by Al-Khamis *et al.* [81].

4.3 EUROPEAN PROJECTS ON MAINTENANCE PLANNING

The European Commission recognized the criticality and the relevance of the problem, funding different European Projects on maintenance planning through its Research Framework Programmes.

TABLE 2 EUROPEAN PROJECTS ON MAINTENANCE PLANNING

Project Acronym	Project Title	European Research Programme	Duration
PM'n'IDEA project	Predictive maintenance employing non-intrusive inspection & data analysis	FP7	2009-2012
ACEM-Rail project	Automated and cost effective maintenance for railway	FP7	2010-2013
AUTOMAIN project	Augmented Usage of Track by Optimisation of Maintenance, Allocation and Inspection of railway Networks	FP7	2011-2014
MAINLINE project	Maintenance, renewal and improvement of rail transport infrastructure to reduce economic and environmental impacts	FP7	2011-2014

INFRALERT project	Linear infrastructure efficiency improvement by automated learning and optimised predictive maintenance techniques	HORIZON 2020	2015-2018
IN2RAIL Project	Intelligent Innovative Rail	HORIZON 2020	2015-2018
AM4INFRA project	Common Framework for a European Life Cycle based Asset Management Approach for transport infrastructure networks	HORIZON 2020	2016-2018
RAGTIME project	Risk based approaches for Asset inteGrity multimodal Transport Infrastructure ManagEment	HORIZON 2020	2016-2019
IN2SMART project	Intelligent Innovative Smart Maintenance of Assets by integrated Technologies	SHIFT2RAIL	2016-2019

PM'n'IDEA: Predictive maintenance employing non-intrusive inspection & data analysis (http://www.pmnidea.eu/.)

The project is launched officially in June 2009. The project aims to reduce the burden on inspection and maintenance of Europe’s increasingly congested rail and tramways through the development of novel sensor technologies and procedures. The project will develop novel inspection and sensor technologies for rail track infrastructure; it is urban oriented but many of the products developed will be applicable to mainline.

Moreover, this project deliver new component designs and maintenance processes that are aimed at improving the integrity of urban rail transport networks through the deployment of intelligent design and sensor technologies into cost effective products and targeted non-intrusive monitoring processes. In particular, the monitoring systems will combine objective automatic visual inspection with examination of internal integrity and the assessment of system and component degradation from a defined datum.

ACEM RAIL: Automated and cost effective maintenance for railway (www.acem-rail.eu)

ACEM-Rail project deals with automation and optimisation of railway infrastructure maintenance. It focuses on the track. The final goal is to reduce costs, time and resources required for maintenance activities and increase the availability of the infrastructure. ACEM-Rail project means an important step forward in railway infrastructure maintenance techniques. Indeed, several technologies for automated and cost effective inspection of the track (subgrade and superstructure) condition and predictive algorithms to estimate the rail defects evolution are developed. Algorithms for an optimal planning of railway infrastructure maintenance tasks are studied in order to obtain appropriate optimisation models for the integrated scheduling of preventive and corrective operations. Modes and tools are introduced in order to monitor the proper execution of corrective and preventive maintenance tasks.

AUTOMAIN: Augmented Usage of Track by Optimisation of Maintenance, Allocation and Inspection of railway Networks (http://automain.uic.org)

The project launched officially in February 2011 has the major goal to optimize and automate maintenance and inspection where possible, also to introduce new planning and scheduling tools and methodology. The project aims to reduce the possession time around 40%. To achieve this, five objectives are set:

1. adopting best practice from other industries in maintenance optimization (e.g. highways, aerospace).
2. developing novel track inspection approaches for freight routes with a scope on in-train measuring and self inspecting switch.
3. researching and assessing innovations that can improve the effectiveness and efficiency of large scale inspection & maintenance processes with a scope on track and switch maintenance, track inspection;
4. further developing of key technologies that will drive the development of modular infrastructure design.
5. developing a new maintenance planning and scheduling tool that is able to optimise the maintenance activities.

MAINLINE: Maintenance, renewal and improvement of rail transport infrastructure to reduce economic and environmental impacts (http://www.mainline-project.eu/)

The main idea of the MAINLINE project is to develop methods and tools that will contribute to a more cost-efficient and effective improvement of European railway infrastructure based on whole life considerations. In view of the scale of renewal dictated by conventional methods over the next decades, it is essential for the project to:

1. facilitate the utilisation of improved assessment and life extension without increasing risk;
2. improve existing knowledge on damage and deterioration mechanisms in order to reduce significantly their effect on asset performance;
3. identify and implement new cost effective replacement/renewal construction methods and logistics, bearing in mind the logistics and operational constraints across an expanding railway network, and the associated political aspirations towards a sustainable low carbon society;
4. identify and compare new surveying and monitoring technologies in order to complement or replace existing techniques;
5. develop methods to determine the whole life environmental and economic impact from track and infrastructure maintenance and renewal through the use of various scenarios and management policies and to create a tool that can compare different maintenance and replacement strategies for track and infrastructure based on traffic situation and whole life evaluation.

INFRALERT- Linear Infrastructure Efficiency Improvement By Automated Learning And Optimised Predictive Maintenance Techniques (http://infralert.eu/)

INFRALERT aims at developing an expert-based information system to support and automate linear asset infrastructure management from measurement to maintenance. This enfolds the collection, storage and analysis of inspection data, the deduction of interventions to keep the performance of the network in optimal condition, and the optimal planning of maintenance interventions.

INFRALERT develops and deploys solutions that enhance the infrastructure performance and adapt its capacity to growing needs by: (i) ensuring the operability under traffic disruptions; (ii) keeping and increasing the availability by optimising operational maintenance interventions and assessing strategic long-term decisions on new construction; and (iii) ensuring service reliability and safety by minimising incidences and failures. INFRALERT is directly applicable by Rail and Road Infrastructure Managers in the field of Intelligent Maintenance and long term strategic planning. Two real pilots (for roads and rail transport systems) are used to validate and demonstrate the results of the research activities.

IN2RAIL Innovative Intelligent Rail (http://www.in2rail.eu/)

IN2RAIL makes advances towards SHIFT2RAIL objectives: enhancing the existing capacity fulfilling user demand; increasing the reliability delivering better and consistent quality of service; reducing the LCC increasing competitiveness of the EU rail system.

To achieve the above, a holistic approach covering Smart Infrastructures, Intelligent Mobility Management (I2M) and Rail Power Supply and Energy Management is applied. Smart Infrastructure addresses the fundamental design of critical assets - switches and crossings and tracks. It researches components capable of meeting future railway demands and utilises modern technologies in the process. Risk and condition-based LEAN approaches to optimise RAMS and LCC in asset maintenance activities is created to tackle the root causes of degradation. It delivers a consistent and holistic approach to asset maintenance improving the reliability of the railway system reducing recurring maintenance costs. The research focuses on an asset maintenance framework, a dynamic model for track system maintenance, and condition and risk based maintenance planning.

AM4INFRA Common Framework for a European Life Cycle based Asset Management Approach for transport infrastructure networks (http://www.am4infra.eu/)

The overall objective of this CSA is to launch a life cycle and risk based Asset Management framework approach enabling effective governance of transport infrastructure networks across

Europe. The framework approach is supported by data management tools which ensure transparency and compatibility of optimization and collaboration actions of the infrastructure authorities within the modes, across the modes and across Europe. In particular, the project deals with life cycle management, risk based approach and asset performance. It assesses the practices adopted in a range of European countries with different types of national networks, identify current tools used and specific risk based and whole life cost tools applied for life cycle planning. The aim is to develop a portfolio of good practices, promoting a framework that allows the application of whole life cost and risk based principles for developing network programmes. Moreover, it deals with data and information management. The challenges are to classify the assets' key information, create an asset data dictionary and to design a common core system model for asset information management. This needs to be faced at European level to guarantee a common approach for asset information management and ensure uniform standards on trans-European networks.

RAGTIME Risk based approaches for Asset inteGrity multimodal Transport Infrastructure ManagEment (**http://ragtime-asset.eu/)**

The main objective of RAGTIME is to develop, demonstrate and validate an innovative management approach and to lay out a whole system planning software platform, based on standard multiscale data models, able to facilitate a holistic management throughout the entire lifecycle of the infrastructure, providing an integrated view of risk based approach, implementing risk based models, resilient concepts and mitigation actions, with specific reference to climate change related threats perspective, and monitored with smart systems, in order to optimize Return Of Investment, management, guarantee Level Of Service and improve resilience through maintaining the service.

IN2SMART Intelligent Innovative Smart Maintenance of Assets by integrated Technologies (**https://shift2rail.org/projects/in2smart/)**

IN2SMART represents the 1st proposal of the Shift2Rail members referred, according to MAAP, to the following Technology Demonstrators (TDs): TD3.7 Railway Information Measuring and Monitoring System (RIMMS), TD3.6 Dynamic Railway Information Management System (DRIMS) and TD3.8 Intelligent Asset Management Strategies (IAMS). These TDs deploy an overall concept for Intelligent Asset Management based on the following three main interlinked layers:

- Measuring and Monitoring systems to collect data from the field related to the railway assets status: IN2SMART develops unmanned systems for “remote” monitoring; track geometry, switches & crossings and signalling monitoring systems; innovative measurement of train parameters and wheel defects combined with rolling stock identifications systems.
- Data management, data mining and data analytics procedures to process data from the field and from other sources: IN2SMART develops standard open interfaces to access heterogeneous maintenance-related data; analytic tools to automatic detect anomalies, discover and describe maintenance workflow processes and predict railway assets decay towards prescriptive maintenance.
- Degradation models and decision-making tools to support maintenance strategies and execution: IN2SMART lays the foundation of a generic framework for asset management and decision support process. This framework specifies the scope, objectives, workflow and outcomes of the decision-making process for maintenance interventions planning, and is the enabler for the development of future decision support tools and systems. IN2SMART also develops an optimised tamping tool and a robot platform for maintenance works.

IN2SMART complements the work of the IN2RAIL lighthouse project to reach a homogeneous TRL4/5 demonstrator. The following Grant will start from IN2SMART to reach the final Integrated Technology Demonstrators that will deploy the overall concept of Intelligent Asset Management.

5. DATA COLLECTION AND DEGRADATION MODELS

This Chapter described the available data from the field that can be used to model the degradation of rail track. Moreover, the existing approaches to predict the future status of rail are analysed, since the asset condition is a fundamental input for the proposed risk-based scheduling model.

5.1 RAILWAY TRACK MONITORING

This section focuses on the current measurement and monitoring techniques to evaluate track condition [40] . In the last decades, inspections were mainly visual, performed by operators walking along the track, sometimes after user's alerts, but nowadays new technologies and practises for the inspection and monitoring of the track have been introduced. This section describes several measurements techniques used to evaluate track defects. These techniques are mainly based on ultrasonic, laser and cameras.

The geometric state of the track is evaluated by the control of some geometrical parameters, for which the railway network regulations establish the permitted values. Analysing the measurements for these parameters and comparing them to the threshold values, it is possible to assess the geometric quality of the track, according to which corrective actions and driving restrictions are programmed. Therefore, if measures are within the threshold limits, then train operation is safe and complies with an adequate level of comfort. Otherwise, maintenance tasks should be carried out.

Track geometry defects are the source for high dynamic forces exchanged between the train and the rails. In most cases such defects are not considered enough to trigger a maintenance activity but only some safety or comfort related interventions, such as the reduction of the train speed, are requested to automatically cover the issue. Actually, this is only partially true. This is one of the reasons for a transition to a maintenance, based on an accurate monitoring of track geometry defects.

In order to evaluate track geometry, some measurement instruments are normally mounted on two types of vehicles:

- Special (and expensive) measurement trains.
- Low speed, inexpensive vehicles.

The low speed, inexpensive vehicles are **small vehicles** (often two axles ones) loaded with track geometry evaluation instruments, while the **measurements trains** are more sophisticated, complex and expensive track inspection vehicles owned by railway administrators to cover the rail network.

These vehicles perform comprehensive inspection of the track. They run on more important lines and its schedule is planed with enough anticipation.

Measurement trains are normally managed centrally, and distribute the results to a central data base and to the regional maintenance offices. On the other hand, low speed, inexpensive vehicles are normally assigned to regional maintenance office to allow prompt measurement in critical area, checking the works done by contractors, etc. It is important to recall that the measurement train could be far away, and it could have its own work plan. Changing the work plan and moving the train would be expensive and inefficient.

Anyway, most of the infrastructure inspection is currently carried out by the measurement trains.

A measurement train carries a wide variety of instruments. Actually, different railways have adopted very different configurations. Some maintenance actors prefer having quite “specialized” trains, e.g. one for the track, one for the catenary, one for the signalling and telecommunications.

Others prefer integrating everything on a single train (to take better advantage of every measurement run). In any case, these trains are running on the entire network of an administration, on a planned routine basis, delivering the data every N days (where N normally depends on the line class).

The new idea is to use **trains in commercial service** to perform inspections on the track. This way to carry out maintenance would save cost based on the following facts:

- expensive measurement train wouldn’t be required;
- cost of performing the operation (crew, traction, etc.) would be saved;
- slots for maintenance operations, difficult to find on busy lines, wouldn’t be required.

Therefore, the track availability for service will increase.

Moreover, since the train would be on service, normally going up and down the same line every day, the frequency of the measurements would be high.

To perform inspections of the track using on-service trains, many of them must be equipped with the instruments to cover the network, normally one equipped train for each line. Moreover, the localization equipment becomes very critical: special techniques must be used, in particular, the

GPS, even where available, could not provide the necessary resolution. The cost of the instruments is more than compensated by the savings in the train and relevant operational expenses.

Obviously such an automatic system is complicated to design, build and manage. The reliability of every component must be much higher than the reliability normally accepted for a measurement train. A large scale experiment on this subject started years ago in the UK (about 30 trains to be equipped), eventually it failed, probably due to unreliable components and wrong software architecture. Other experiments are known, but none is in commercial operation at this time.

Anyway, evaluating track condition using on-service trains can deliver very useful data for an improved maintenance strategy:

- Better trends (more accurate).
- Immediate verification of the works.
- Early detection of unpredictable faults.

Data collected can be reported in electronic format. Severe faults are reported at once and immediate actions are taken. Other non-urgent defects are delivered to the central/regional office where the defects are evaluated by rather simple algorithms and ranked by severity indexes.

Considering the main available **inspection and monitoring technologies,** they can be divided in [82]:

1. Fibre optic sensors laid along the track and other infrastructure elements (such as bridges) [83].
2. Hollow-shaft integrated acoustic sensor system [84].
3. Rail monitoring sensor combining eddy current distance measurement with acceleration data.
4. Laser profiler and inertial pack to monitor the track geometry.
5. Ultrasonic non-destructive fuzzy inspection techniques.
6. Non-contact thermography system for rail surface monitoring.
7. Visual camera.

The first four techniques above are the most promising because they allow the Railway Infrastructure Manager keeping updated information on a daily basis on the track state in a very cost-effective way. Technologies 2, 3 and 4 above could be embarked on commercial trains obtaining an automated and unattended measurement system. This is a step forward in track inspection technologies which will reduce the need for expensive instrumentation trains inspecting the infrastructure during the night when there are not rail services and will increase the capacity of the rail transport. Moreover, the availability of frequent and quality data on the track condition makes possible the condition-based maintenance on the basis of daily updated measurements.

At current state of development, technologies 5 and 6 above can only be run on trains at low speed which make them unsuitable for commercial trains.

Here below, these technologies are described more in detail.

1. Fibre optic sensors

In the railways field, distributed optical fibre sensors can be employed for spatially continuous monitoring of the tracks temperature and deformation, as well as the monitoring of the structural integrity of infrastructures as tunnels, bridges and embankments. To this purpose, a single-mode optical fibre cable has to be attached to the track and/or the structure under test in order to detect both tensile and compressive strains.

2. Acoustic Inspection Techniques

Acoustic methods “listen” to natural sound sources like the rolling noise. If hollow shafts are available (e.g. in some high-speed trains) hollow shaft integrated acoustic sensor systems can be used to detect defects in wheel sets of the rolling stock. The system could include acceleration sensors as well as structure-born sound sensors and uses wireless real-time data transmission. The acoustic part of the system detects and evaluates acoustic signals generated by the rail-wheel contact.

3. Inspection using pulsed eddy currents

Eddy current measurements have been a standard technique for a long time for finding cracks in metals either on the surface or within the material [85]. Recently, eddy current sensors have also become a common method for rail inspection [86]. However, the probe has to be either in contact or very close to the surface (<1mm). Therefore, usage in rail inspection has for a long time been limited to hand-held system or system mounted on manually driven trolleys. To adapt the method of eddy current testing to a train borne platform mounted on a commercial train the distance between sensor and rail has to increase significantly. Anyway, for regular rail inspection reduced performance of the system could be tolerated while at the same time increasing the frequency of inspection which would be possible using commercial trains. Initial experiments carried out by Siemens have shown that eddy currents of lower frequency are quite able to detect larger cracks even when probe-surface distance exceeds 10mm.

4. Laser profilometer

Profilometer is a measuring instrument used to measure a surface's profile, in order to quantify its roughness. Vertical resolution is usually in the nanometre level, though lateral resolution is usually poorer. An optical profilometer is a non-contact method for providing much of the same information as a stylus based profilometer. There are many different techniques which are currently being employed, such as laser triangulation, and confocal microscopy.

5. Ultrasonic inspection

Ultrasonic techniques belong to the most commonly used non-destructive methods with a wide variety of application fields. In most cases a broadband pulse is excited by a piezoelectric transducer and is send into the structure using an appropriate coupling agent like water, oil or viscous paste. The waves interact with interior defects and are reflected back so that they can be detected by a sensor. The latter can be either the same transducer that was used for excitation or an additional sensor. Ultrasonic techniques are well known for the inspection of the rolling stock of high-speed lines in Europe and abroad [87],[88].

Test trains typically include ultrasonic and eddy current systems to automatically scan the rail during run of the train (<100 km/h). The procedure is usually organised in a three tier inspection process. The first tier, fast mapping of the rail is performed by the inspection car traveling at high speed on the track. Once data is recorded and stored it is analyzed off line. The analysis or processing can identify and categorize flaws in a scan. The processing step issues a report which contains a list of all suspect flaws, their location in the scan and their distance from the nearest reference points. This scan report serves the repair team who returns to the relevant section of rail, and has to locate the detected flaw and verify it prior to maintenance operations on the rail. Such approach reduces the amount of time the track is blocked by ultrasonic inspection process, but on the other side it relates very much on experience of analysis team, and also increases the time until final results with defect classification will be available for maintenance planning.

Despite the problems mentioned above automatic test trains or test vehicles still provide useful information about the track and rail condition.

6. Thermographic inspection

Infrared thermography also belongs to the well-known non-contact non-destructive techniques. The material under test is first heated by a flash lamp or an inductive technique. After that the spatial-temporal evolution of the thermal field is monitored by an infrared camera. If defects are present, the thermal conductivity is locally decreased so that “hot spots” of higher temperature can be detected. From the temporal change of the thermal field additional information about depth and size of the defect can be determined in principle [89],[90].

In earlier investigations this technique has already shown its high potential for the characterization of typical flaws in rails. It could be shown that this technique principally allows the characterisation of the rails with a high sensitivity and a high testing speed. Just like ultrasonic and electromagnetic techniques a thermography system can also be integrated in a testing train. With the current hardware an automated testing of the rails and automated defect recognition at speeds up to 20 m/s (about 70 km/h) seems to be possible.

7. Inspection using visual cameras

The main goal for using images of the track is to eliminate, or reduce as much as possible, the visual inspection done by workers walking along the track to detect any fault, missing components, etc. The state of the art of these instruments does not yet allow a complete and safe elimination of the inspection done by humans, but helps a lot and, also, allows detecting a number of risky situation difficult to detect by the human eye.

A number of linear cameras are mounted under the vehicle.

The linear cameras are space triggered (e.g. every 1 mm). Every image is equivalent to a single line of a normal camera. Assembling all this lines in an endless sequence gives a stream image of the track having:

- on the transverse (Y) axis, as many pixels as the camera,

- on the longitudinal (X) axis, as many pixels as the number of mm travelled by the train.

A colour image is normally used to allow a human inspector viewing the track as if he was walking, but with obvious advantage for safety and line capacity. Sometimes it used for the faulty fasteners automatic detection, by machine vision techniques.

A black and white image is normally used for detecting rail surface defects and for every automatic analysis by machine vision techniques (rail surface, fasteners, sleepers, joints). The automatic analysis is useful to focus the attention of the workstation operator, who then goes to examine the relevant colour image and decides the relevant actions.

5.2 TRACK DEGRADATION MODELS

The railway track and infrastructure degrade with age and usage, and this implies that they can become unreliable due to failure. When a failure occurs, the consequences can be significant, including a high cost of railway operation, economic loss, damage to the railway asset and environment and possible loss of human lives. Unreliability may also lead to annoyance, inconvenience and a lasting customer dissatisfaction that can create serious problems for the company's position in the marketplace. An applicable and effective maintenance strategy can guarantee the achievement of reliability goals and compensate for unreliability.

Maintenance actions are used to control the degradation of the track, reduce or eliminate the likelihood of failures, and restore a failed part to an operational state. It is necessary to model track degradation behaviour to select an applicable and effective maintenance policy, but modelling and predicting the track geometry degradation is a complex task, requiring the following information: 1. the interaction of different track components, 2. the effect of maintenance actions on track quality, 3. the heterogeneity factors e.g. environmental factors, soil type and condition.

In addition, higher demand for railway transportation makes higher speed and axle load an essential requirement that, at the other hand, accelerates the track aging process and negatively affects its reliability. Therefore, the increased demand and complexity dictates the need for comprehensive track degradation models.

In the two last decades, a great deal of research has been done in the field of track geometry degradation modelling [91]. Determining an indicator to represent track quality is an essential prerequisite for modelling track degradation. Different indicators are used based on the aim of the research. The indices for representing track quality condition are demonstrated in Fig. 14.

Sadeghi *et al*. [91] proposed a track geometry index that uses the following track geometry parameters: alignment, profile, twist, gauge, and rail cant. Using justified coefficient, they combined the parameters to design the track geometry index.

In order to consider structural defects, Sadeghi *et al*. [93] proposed a quantitative track structural quality index. This index is defined for each track component group, i.e. rail, sleeper, fastener, ballast.

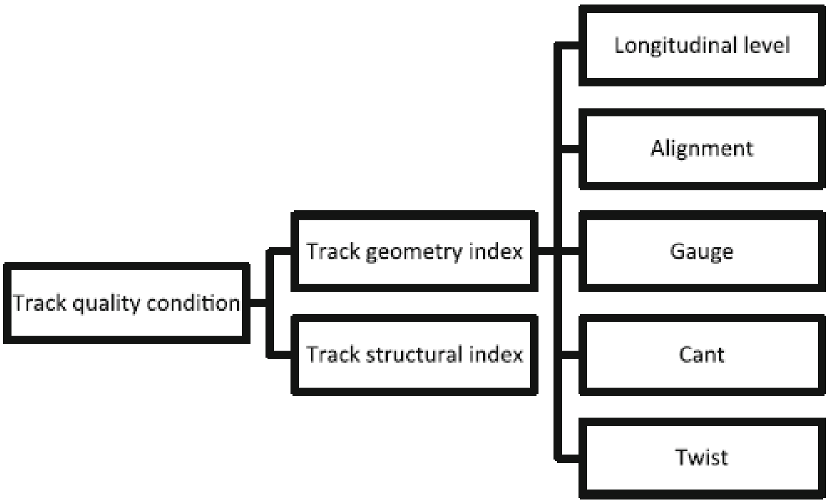

FIGURE 14 TRACK CONDITION MEASURES

Faiz *et al*. [94] studied the geometry parameters used in the UK track maintenance process and applied linear regression analysis to explain their correlations. A Generalized Energy Index (GEI) instead of a Track Quality Index (TQI) for track quality evaluation is proposed by Li *et al*. [95]. The GEI can consider different track irregularity wave-length and speed. Haifeng *et al*. [96] proposed an integral maintenance index (IMI) that considers the distribution of track geometry parameters to evaluate track condition. El-Sibaie *et al*. [97] developed a number of track quality indices to evaluate track quality condition in relation to different track classes.

By looking to the literature, it can be observe that most of the researchers considered short wavelength longitudinal level as the crucial factor in degradation modelling. This issue can be seen in Fig. 15.

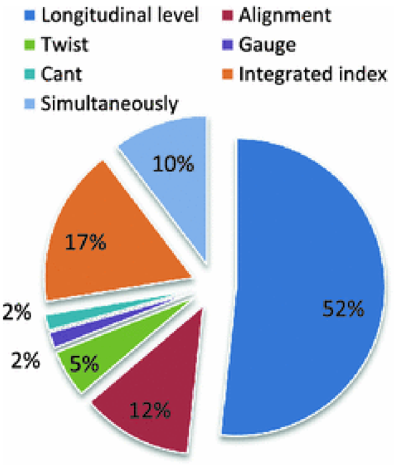

FIGURE 15 DISTRIBUTION OF APPLIED TRACK GEOMETRY MEASURES

After finding the proper track quality measure, a degradation model must be constructed and the effect of different maintenance strategies on track degradation evaluated. There are two major approaches for track geometry degradation modelling, i.e. mechanistic and statistical approaches. Concerning mechanistic approach, a number of researchers tried to find the interactions among track components and their influences on track geometry degradation.

The most important models are those proposed by Shenton [98], Sato [99][99], Chrismer *et al*. [101], Öberg *et al*. [102], and Zhang *et al*. [103]. Dahlberg [104] also provide an extensive review on mechanistic models applied for track geometry degradation.

Concerning statistical approaches, which are the main focus in this thesis, the most commonly applied methods are summarised in Fig. 16.

Andrade *et al*. [105] assessed track geometry degradation and the uncertainty of degradation model parameters. They considered a linear model for track longitudinal level degradation. They performed statistical correlation analysis for each group section and fitted the log-normal distribution to the track's longitudinal level degradation. A multi-stage linear model is applied by

Gou *et al*. [106] to cope with different phases of degradation between two consecutive maintenance interventions and the exponential growth of track irregularity.

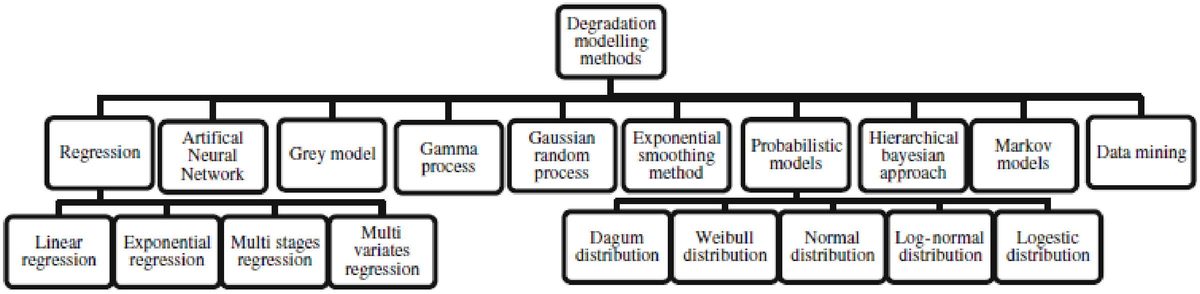

FIGURE 16 TRACK DEGRADATION APPROACHES

Famurewa *et al*. [107] compared the accuracy of linear, exponential, and grey models in the estimation and prediction of track geometry degradation. The comparison demonstrated the grey model has lower mean average percentage error than the linear model and an approximately equal error value with the exponential model.

In [108] the deterioration of track quality at one specific track position is shown over a period from 2001 to 2007. The theoretical exponential function is in good accordance with the real track behaviour since the measured track indexes are well fitted.

Lyngby [109] suggested a methodology for evaluating track degradation in terms of track geometry irregularities and proposed a multivariate regression model to demonstrate the relationship between the track degradation measure variable and influencing variables on track degradation. Since different sections of track are not identical, the track was split into homogenous sections with similar variables. He concluded: (1) axle load has a nonlinear relation with degradation; (2) degradation after tamping is dependent on the number of previous tampings; (3) soil consisting of clay material will settle sooner than other types of soil; (4) light rail tracks degrade faster than heavy rail tracks; (5) harsh rainfall increases degradation rate.

Using waveform data, Liu *et al*. [110] proposed a short range prediction model to estimate any track irregularity index over a short track section length (25 m) and on a day-by-day basis. They concluded the total process of track surface change over track sections is nonlinear and different track sections have different nonlinear process.

Xu *et al*. [111] proposed an approach based on historical changes in track irregularity to predict the short-term track degradation. They estimated the non-linear behaviour of track irregularity during a cycle using a number of short range linear regression models.

Two degradation models to predict track alignment irregularities are proposed in the work by Kawaguchi *et al*. [112]. First, they developed a degradation model based on analysis of lateral track deformation to estimate mean time to maintenance of track alignment irregularities. Second, they designed another degradation model based on the exponential smoothing method to accurately predict the track alignment irregularities a maximum of 1 year in advance.

The comparison of the efficiency of the double exponential smoothing method, a generic degradation model, and an autoregressive model for track degradation prediction is addressed in the work by Quiroga *et al*. [113]. The three models lose their efficiency in track degradation prediction after performing a number of tamping procedures. After considering these issues, they developed a hybrid discrete-continuous framework based on a grey box model. After comparing these four models, they concluded the proposed hybrid model is more efficient in terms of track degradation behaviour prediction.

A stochastic approach based on Dagum distribution is developed by Vale *et al*. [114] to model track longitudinal level degradation over time. The researchers classified the track longitudinal level changes into three speed classes and different inspection intervals.

The Gaussian random process is used by Zhu *et al*. [115] to model track irregularities in vertical profile and alignment. They discussed power spectral density analysis and cross-level statistics about track irregularities to improve track degradation modelling.

A stochastic Markov model is used by Bai *et al*. [116] to evaluate track degradation.

They considered various heterogeneous factors and argued that the existence of these factors caused two maintenance units with the same mileage to show different degradation behaviour. A Markov model is deployed by Yousefikia *et al*. [117] to model tram track degradation and to obtain the optimal maintenance strategy. A model by integrating the grey model and Markov chain is developed by Liu *et al*. [118] to predict track quality condition.

Andrade *et al*. [119] used a Bayesian approach to evaluate a track geometry degradation model and to deal with the uncertainty of its parameters. They considered the track longitudinal level deviation to have a linear relationship with passing tonnage and assumed the initial longitudinal level and degradation rate would take a bivariate log-normal prior distribution. They argued that the parameter uncertainties are significant in the design stage.

Guler [120] used artificial neural networks to model the degradation of different track geometry parameters. The model considered traffic load, velocity, curvature, gradient, cross-level, sleeper type, rail type, rail length, falling rock, land slide, snow, and flood as influencing factors. A modified grey model is developed by Chaolong *et al*. [121] to analyse track irregularity time series data and obtain a medium-long term prediction of track cross levelling. They compared the stochastic linear autoregressive model, Kalman filtering model, and artificial neural network with respect to the short term track cross levelling prediction. They observed the accuracy of the ANN model was higher than the two other models.

A machine learning model based on the characteristics and inspection data of the track using a multi-stage framework is developed by Xu *et al*. [122] to predict changes in track irregularity over time. They defined different stages of track changes based on maintenance thresholds and linear regression is used to predict track degradation in each stage.

Xu *et al*. [123] proposed a track measures data mining model to predict railway track degradation for a short time period. Data mining and time series theories are applied by Chaolong *et al*. [124] to predict track irregularity standard deviation time series data. In order to predict the changing trends of track irregularity, they used the linear recursive model and the linear autoregressive moving average model.

According to data mining techniques, the prediction of the asset condition can be categorised in two ways: nowcasting and forecasting [125]. Nowcasting methods are used to identify faults that will lead to failure within a few hours; this is done for safety reasons and also to extend remaining useful life (RUL). Forecasting can be useful to assess the condition of an asset for the remaining useful life in the long run. There are three types of methods to quantify remaining useful life: data driven, symbolic and physical models. Data driven methods are purely based on the data acquired by sensors; they carry out classification and clustering techniques to identify anomalies. Symbolic methods make use of work orders and other empirical records of maintenance. Finally, physical methods exploit the physical structure of the component to analyse degradation. The combination of symbolic, data driven and physical models into hybrid models is demonstrated to be a good solution for nowcasting and forecasting of asset condition.

Prognostic models, predicting information about the asset condition, providing estimations of current or future values of the relevant parameters, shall also be used in the alert management system to derive failure probability distributions. The future (unknown and random) degradation and defect evolution in this thesis is described as stochastic processes, and functions for the failure probabilities are determined using the output of prognosis models.

6. PREDICTIVE RISK-BASED MAINTENANCE SCHEDULING OF RAILWAY ASSETS

The aim of the proposed model is to plan railway predictive maintenance activities, in order to intervene when an asset has reached a certain degradation state and thus preventing faults and possible failures.

In analysing the proposed model formulation, it is worth remembering that the current practices consider off-line scheduling models that cover long-term horizon, neglecting operative disturbances. In particular, the weaknesses of currents methods are due to their inability of dealing with the following characteristics of railway lines:

- All the different railway infrastructures have the common characteristic of being scarcely redundant (no or very few path alternatives), and this implies that, when a fault occurs, the system performances have a dramatic capacity drop and a consequent loss for railway operators. Therefore, the maintenance plan should be able to minimize the probability of the occurrence, of the so-called mission-critical faults during train service, i.e. those that prevents trains for circulating or can lead to accidents;

- in addition, railway maintenance plan should be able to optimize the resources utilization taking into account the space-distributed aspect of railway infrastructure. In fact, railway assets are often not spatially delimited to a point, and this implies difficulties in the organization of maintenance activities and resources;
- another aspect that makes railway maintenance critical is the time constraint. In fact, the available time for maintenance is very limited due to various factors such as railroad traffic, climate, and interrelations among different maintenance projects. The maintenance planning model should be able to consider these constraints. In particular, some of these requirements result to be soft constraints, that is, violations that can be tolerated if no better choices exist, and some are hard constraints that can never be violated;
- since in the rail sector automatic monitoring and diagnostic systems, mounted on trains and along the lines, have become over years more and more performant and sophisticated, and new algorithms can be used to assess the asset status, the planning model should be able to use the inputs from degradation model able to predict the time when a fault is likely to occur, thanks to the new available data;
- finally, the planning model should be able to adapt maintenance interventions accordingly to updated information received in real time from geographical distributed sites.

The proposed approach has the capability to deal with the above mentioned issues. The resources utilization is optimized, taking into account the different characteristics of the maintenance teams, such as working speed and travel speed, and defining the optimal assignment of the maintenance activities to the available maintenance teams. Moreover, the planning model takes into account the spatially-distributed aspect of railway infrastructure, considering the necessary time for moving from different points of the railway line, which can vary according to the maintenance teams performance. The availability of the maintenance teams is given by the resource manager and can vary during the considered time horizon.

The time constraints are introduced, considering that the available time for maintenance is limited by various factors such as railroad traffic. Therefore, the time horizon is divided into train-free sub-intervals during which the train circulation is forbidden and maintenance activities can be performed. The duration, the starting and final instant of each sub-intervals are assumed to be given by the traffic manager.

The model is based on the risk minimization, according to ISO 55000 guidelines taking into account risk thresholds, evaluated using the inputs from degradation model able to predict the time when a fault is likely to occur. The risk thresholds become constraints of the optimization problem, as described in the Chapter 7.

Finally, the planning model is able to adapt maintenance interventions accordingly to updated information received in real time. In particular, two different scenarios of unexpected events can be addressed:

- work delay scenario. A delay of one activity makes the next activity impossible to be finished in time in the considered working time interval. Therefore, such a maintenance has to be reconsidered in the problem stated at the following decision time for the next working time-period;
- new maintenance activity scenario need to be considered in the planning. This scenario considers an updated input from the predictive model related to new maintenance activities to be performed.

6.1 DECISION SUPPORT FRAMEWORK

In this section, the proposed general architecture of the decision support system is described.

In particular, the architecture consists of two different levels: the first level is an off-line decision support system to schedule railway predictive maintenance activities, while the second level consists of an on-line recovery decision support system, aimed at the adaption of the schedule when a failure occurs, or when the off-line plan cannot be met.

More in details, consider the modular scheme depicted in Fig. 17. In such a scheme, the Infrastructure *Degradation Condition Analysis module* takes as inputs the outcomes of the *Dynamic Railway Information Management module* that forecasts the future degradation scenarios of railway infrastructure via suitable *Track Degradation Models*. In other words, these modules elaborate real-time data about track conditions and the conditions of each asset, providing their current and future states on a-priori determined degradation curves. Finally, based on the particular reached degradation condition of each asset, and on its future trend, the *Infrastructure Degradation Condition Analysis* module provides the failure probability $\mathbb{P}(t_f \leq T)$.

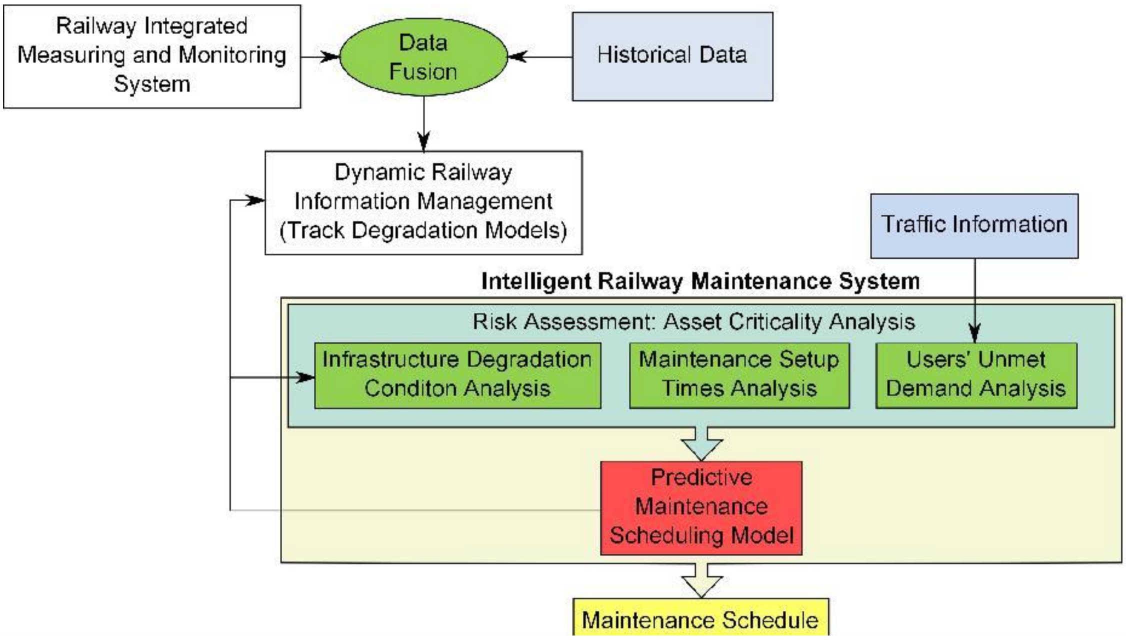

FIGURE 17 FIRST LEVEL MODULAR ARCHITECTURE OF THE DECISION SUPPORT SYSTEM

At the same time, the *Users Unmet Demand Analysis* module takes the railway traffic information as an input to determine which *train-free periods* are available for maintenance interventions. Moreover, the trains that will suffer delays or cancellations in case of fault occurrences are determined, and the consequent users (passengers and freight) unmet demand, and the relevant loss D, necessary for computing the criticality weights ω_i are evaluated.

Finally, the *Maintenance Setup Times Analysis* module determines the maintenance setup times for all the activities, which can be computed as the sum of:

- the removal time of the repair yard of the completed maintenance activity i by the team m, hereafter indicated as C_{im}^{-};
- the preparation time of the repair yard of the next scheduled maintenance activity i by the team m, hereafter indicated as C_{im}^{+};
- the team m displacement time from the asset i to the asset j, hereafter indicated as C_{ijm}. Note that this term depend on the relative positions of the assets i and j. Therefore, since if $h \neq j$, then $C_{ijm} \neq C_{ihm}$, such costs have to be computed for all the couples of activities. Anyway, it is assumed that $C_{ijm} = C_{jim}$.

Note that, all these costs depend also on the characteristics of team m. Therefore, they have to be computed for the different available maintenance teams.

The planning is performed by the *Predictive Maintenance Scheduling Model module*, which determines the optimal *Maintenance Schedule* by taking into account the above determined maintenance priorities, costs, and constraints.

The second level consists of a recovery decision support system for the scheduling adaption in case of failure occurrence or when the off-line schedule cannot be met. This is an online rescheduling process made up by the reactive activities that have to cope with uncertainties and changes of operating conditions, targets and outcomes. Online rescheduling is a real-time decision making function that has to take into account new information, at a regular or irregular frequency, and to implement next decisions, based upon the available information on the new state of the system and new requirements. Therefore, this level implies feedback and feed-forward information structures.

The architecture of the second level is shown in Fig. 18. As the implementation of the off-line schedule can be perturbed by unpredictable occurrences and "failures", such as delays in the beginning and/or ending of some operation executions, the recovery model is made up by a *Real-*

time Control System able to detect, to diagnose, and to correct the consequences of the undesired events.

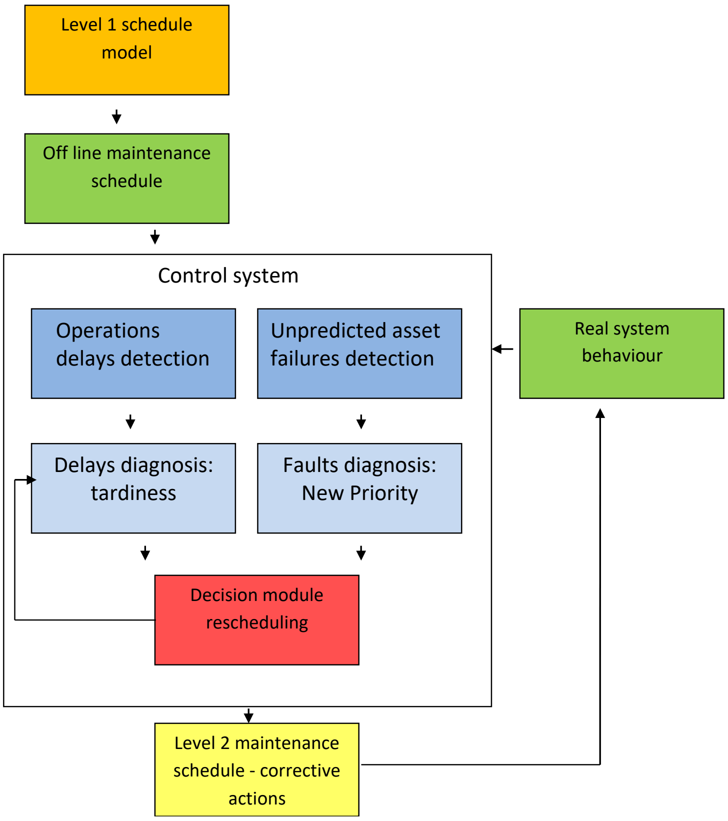

FIGURE 18 SECOND LEVEL MODULAR ARCHITECTURE OF THE DECISION SUPPORT SYSTEM

The inputs of this *Control system* are the outcomes of the first level model and the information from the *Real system behavior monitoring*.

Since the disturbances to the off-line schedule can occur at any time and place, during the operations, it is very important to estimate type, time, and location of any undesired event, and especially its propagation capability. This makes the *fault monitoring and detection activities* of paramount importance. Therefore, the *Control system* is composed of two modules performing fault detection, and two modules performing fault diagnosis functions. The two modules performing fault detection are the *Delays detection module* and the *Unpredicted assets failures* *detection* module, while the modules performing faults diagnosis are the *Delays diagnosis module* and the *Faults diagnosis module*.

The *Operations delays detection* module detects the state of the process, such as delays in the maintenance operations starting time, due to work team delays. Instead, the *Unpredicted assets failures detection* module detects unexpected faults, that degradation models haven't been able to predict or updated information on asset status. Disruptions or delays in work team maintenance activities and travels may produce tardiness in the operations and the consequently risk of exceeding soft or hard degradation thresholds. The occurrence of faults that degradation models are not able to predict, determines unexpected high priority of interventions. Updated information on asset status could also determine changes in intervention priority and new asset to be considered in the planning.

The *Operations delays* detection module is equipped with a clock and an information system consisting of the list of scheduled activities, their timing, and the work team allocated to each operation. Whenever a delay is detected, the detection module sends a warning to the *Delays diagnosis module*. In turn, the *Delays diagnosis* module identifies the resource (work teams) in late and the tardiness value and sends the results to the *Decision module*.

The *Unpredicted assets failures detection module* detects when a track segment reaches unexpectedly unsafety conditions and sends a warning to the *Faults diagnosis module* that defines the absolute priority of this maintenance activity and update the other activities priorities accordingly. Therefore, the feedback information is the update of the maintenance schedule calculated by the level 1 and the update of the priority parameters and tardiness parameters that are assumed in the computation of future reschedules.

Consequently, the information obtained from the running feedback structure is used in the rescheduling process by the *Decision module* to define the corrective actions. The corrective actions consist of the definition of new maintenance activities timetable and new activities allocations to the work teams. The rescheduling procedure can be distinguished in:

- reschedule based on changes in the work team conditions and operations;
- reschedule based on changes of assets conditions, failures, incidents.

Whenever a deviation from the offline schedule is detected, an optimization problem is solved by the *Decision module* to find the schedule that minimizes the deviations and the new expected completion times.

Therefore, the output of the *Control system* is a recovery strategy that can be applied to minimize the delay propagation.

Summing up, the continuous feedback control allows to handle the uncertainties and disturbances due to the existence of external influences that can produce a discrepancy between the assumed off-line schedule models and the actual behavior.

Therefore, the decision support system allows maintenance managers to shift from an off-line long-term planning strategy to a real-time dynamic day-to-day planning strategy, responding dynamically to the increasing understanding of the real world process. The adaption of maintenance schedules to real-time information allows also taking into account the unexpected events related to the stochastic nature of real world operations. Therefore, the proposed planning model is a useful instrument to make decisions at operational level.

When new information arrives or an unexpected event occurs, a new statement of the model is solved and the solution is used to support the real-time decision making.

Whereas the current practices consider off-line scheduling models that cover long-term horizon, neglecting operative disturbances, the presented decision support system is a dynamic rescheduling model that considers the short-term planning with the aim of dealing with real-time data and uncertainties.

6.2 RISK-BASED APPROACH

As mentioned, the model is based on risk minimization, according to ISO 55000 guidelines taking into account risk thresholds, which are evaluated using the inputs from degradation model able to predict the time when a fault is likely to occur.

In this section, the generic risk-based formulations of the optimal scheduling problem for both non-technological and technological assets are described.

RISK-BASED MODEL FOR NON-TECHNOLOGICAL ASSETS

Regarding the risk-based formulation of the optimal scheduling problem for non-technological asset, some assumptions are introduced at different steps, only when necessary, with attention at keeping the model more general as possible.

Then, in a generic time instant t_k the general risk definition for a generic asset i and for a given time $t \geq t_k$, is expressed as the product of the probability that the fault occurs in t, $\mathbb{P}_i\{t|t \geq t_k\}$, for the consequent losses D_i, that is,

$$R_i(t, t_k) = \mathbb{P}_i\{t|t \geq t_k\}D_i \tag{6.2.1}$$

As a consequence, given a maximum tolerable risk $\bar{R}_i$ for the asset i, and the losses D_i generated by its fault, it is easy to compute the maximum tolerable fault probability $\bar{p}_i = \bar{R}_i/D_i$.

Assumption 1: It is possible to estimate the instant at which the fault probability $\mathbb{P}_i\{t|t \geq t_k\}$ reaches the value $\bar{p}_i$ by means of suitable stochastic models which estimate the present and future states of the asset i and provide the so-called assets degradation curve of a vital parameter, hereafter generally indicated as $\delta_i(t, t_k)$, $\forall t \geq t_k$.

The determination of $\delta_i(t, t_k)$ is beyond the scope of this paper. The interested reader can refer to Wang *et al*. [126]and Ke *et al*. [127] for the degradation curve evaluation of a generic asset, and to Famurewa *et al*. [128][129], for the degradation curve estimation of a rail track.

Assumption 2: The fault probability can be expressed as the probability that the parameter $\delta_i(t, t_k)$ reaches a given lower or upper threshold, hereafter indicated as δ_i^L and δ_i^U respectively. Formally,

$$\mathbb{P}_i\{t|t \geq t_k\} \simeq \mathbb{P}\{\delta_i(\bar{t}, t_k) \leq \delta_i^U \vee \delta_i(\bar{t}) \geq \delta_i^U\} \tag{6.2.2}$$

Assumption 3: The parameter $\delta_i(t, t_k)$ can be modeled as Non-stationary Stochastic Process (NSP) with known expectation $\bar{\delta}_i(t, t_k)$ and variance $\sigma_i^2(t, t_k)$.

Such an assumption allows to consider the realistic case in which the degradation model cannot predict the asset state exactly but only via an error-affected estimate of $\bar{\delta}_i(t, t_k)$ computed in t_k for all the instants $t \geq t_k$.

Assumption 4: The parameter $\delta_i(t, t_k)$ can be modeled as Non-stationary Gaussian Stochastic Process (NGSP), that is $\delta_i(t, t_k) \in \mathcal{N}\left(\bar{\delta}_i(t, t_k), \sigma_i^2(t, t_k)\right)$ and interpreted as the residual time-varying stochastic error in the estimate of $\bar{\delta}_i(t, t_k)$.

Therefore, by means of the assumptions 1, 2, 3 and 4, for any value of the variable t, the NGSP $\delta_i(t, t_k)$ turns out to be a simple Gaussian stochastic variable with given expectation and variance. Therefore, it is possible to compute the probability $\mathbb{P}\{t|t \geq t_k\}$ as

$$\begin{aligned}&\mathbb{P}\{\delta_i(t, t_k) \leq \delta_i^U \vee \delta_i(t, t_k) \geq \delta_i^U\} \\ &= 1 - \frac{1}{\sqrt{2\pi\sigma_i^2(t, t_k)}} \int_{\delta_i^L}^{\delta_i^U} \exp\left(\frac{\left(x - \bar{\delta}_i(t, t_k)\right)^2}{2\sigma_i(t, t_k)}\right) dx, \forall\, t \geq t_k\end{aligned} \tag{6.2.3}$$

Then, by varying t, it is possible to derive a curve providing the failure probability $\mathbb{P}_i\{t|t \geq t_k\}$ for all t.

As an example, consider the plot in Fig.19, where:

- it is assumed $t_k = 0$, that is, the asset i is at the beginning of its life cycle;

- for the sake of clearness only some representative instants t are depicted;
- the dot-dashed line represents the shape of $\bar{\delta}_i(t,0)$;
- the dashed and dotted lines represent the bounds δ_i^L and δ_i^U;
- the continuous lines represent, for any t, the density functions of the stochastic variables $\delta_i(t,t_k)$;
- the filled areas represent, for any t, the probabilities $\mathbb{P}\{\delta_i(t,0) \leq \delta_i^U \vee \delta_i(t,0) \geq \delta_i^U\}$.

In addition, in Figure 20, the relevant complete fault probability curve $\mathbb{P}_i\{t|t \geq 0\}$ is depicted.

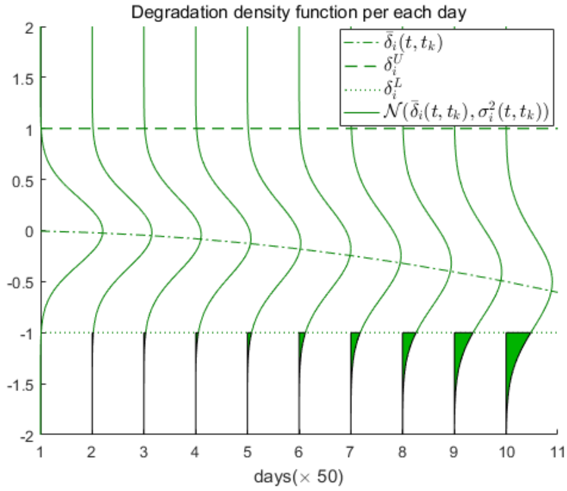

FIGURE 19 ASSET DEGRADATION STOCHASTIC PROCESS

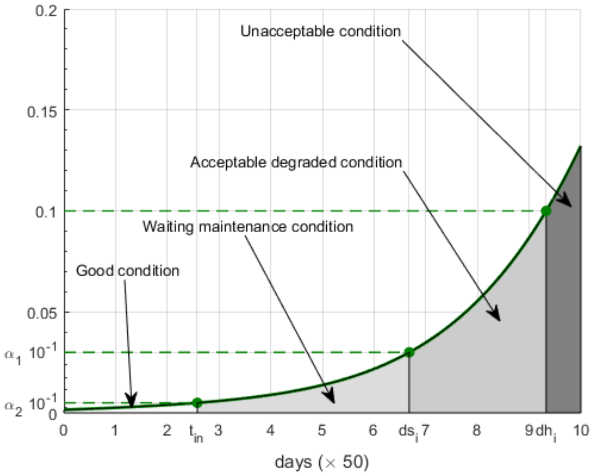

FIGURE 20 ASSET FAULT PROBABILITY CURVE

Coming back to the model, once the fault probability curve is determined for a given asset, it is immediate to determine the instant at which the fault probability reaches the threshold $\bar{p}_i$ determining the hard deadline as the instant $t = dh_i$ such that $\mathbb{P}_i\{t = dh_i | t \geq t_k\} = \bar{p}_i$. Analogously, a soft deadline can be determined as the instant $t = ds_i$ such that $\mathbb{P}_i\{t = ds_i | t \geq t_k\} = \alpha_1 \bar{p}_i$, being $\alpha_1 \in (0,1)$ a chosen parameter indicating a tolerable percentage of the fault probability $\bar{p}_i$. Finally, it is possible to define maintenance requirement starting time as the instant $t = t_{in}$ such that $\mathbb{P}_i\{t = t_{in} | t \geq t_k\} = \alpha_2 \bar{p}_i$, being $\alpha_2 \in (0, \alpha_1)$, a parameter analogous to α_1. Given the above model, each asset can be in one of the following states:

- the *Good condition*, that is, the state characterized by a negligible degradation;
- the *Waiting maintenance condition*, expected to be reached in t_{in}, representing the state in which the considered asset is far from fault, although its degradation process has begun;
- the *Acceptable degraded condition*, expected to be reached in ds_i, representing the state in which the considered asset is still sufficiently far from a probable fault but begins to need a maintenance activity;
- finally, the *Unacceptable condition* that should never be reached, expected to be reached within dh_i.

Once the state of each asset is identified and the thresholds determined, the model proposed in the following Chapter 7 allows to plan the maintenance activities via a Mixed Integer Linear Programming (MILP) problem, in which the risk thresholds become constraints of the optimization problem. In doing so, only the assets that have reached the Waiting maintenance condition are considered.

RISK-BASED MODEL FOR TECHNOLOGICAL ASSETS

Considering technological assets, it is worth noting that deterioration could be difficult to be measured for rail assets, such as signalling equipment or electrical equipment, since their deterioration does not progress continuously and failures seem to occur suddenly. For this reason, predictive maintenance has been usually applied to rail track because track deterioration can be directly measured and continuously perceived. However, recent advances in ICT, especially progress in data analysis technologies, allow to perceive signs and causes of failures even in types of machinery that apparently breaks down suddenly, by continuously measuring certain physical quantities (current, resistance, etc.) [130].

Hereafter, the generic risk-based formulation of the optimal scheduling problem for technological assets is described to show the applicability of the proposed risk-based approach to different stochastic process. The notation used in this section is reported in Tab. 3

TABLE 3 NOTATION OF THE RISK-BASED MODEL

SETS AND DETERMINISTIC PARAMETERS	
$\mathcal{A}$	set of the considered assets and $\lvert\mathcal{A}\rvert$ the relevant cardinality
$i = 0,1,\dots,\lvert\mathcal{A}\rvert$	index of assets
τ	time at which the risk assessment is performed
Δ	generic time interval
$\mathcal{D}_i$	loss caused by the fault of the asset i
$\{t_k^i\}_{k=0,1,\dots}$	sequence of time instants at which the maintenance of the asset i are performed
t_0^i	time of the first installation of the asset i
$R_i(\tau)$	failure risk at τ of asset i
$\bar{R}_i$	maximum tolerable risk of asset i
STOCHASTIC VARIABLES PARAMETERS	
$F_i\left(\tau,\theta(t_k^i)\right)$	Cumulative Distribution Function (CDF) of the random fault process of the asset i, that is, the probability that a fault occurs within the time τ
$\theta(t_k^i)$	vector of CDF parameters depending on previous maintenance activities performed in t_k^i
$\varphi_i(\tau,\Delta,t_k^i)$	conditioned fault probability
OUTPUT PARAMETERS	
$\delta h_i(t_k^i)$	hard deadline of the maintenance activity on asset i after the maintenance occurred in t_k^i
$\delta s_i(t_k^i)$	soft deadline of the maintenance activity on asset i after the maintenance occurred in t_k^i
$\delta g_i(t_k^i)$	instant when it is necessary to start planning the maintenance of asset i, after the maintenance occurred in t_k^i. Such a parameter can be interpreted as the release time of asset i in the scheduling problem

The risk related to a fault occurring in the interval $[\tau,\tau+\Delta]$, given that it has not occurred in the interval $[t_k^i,\tau]$, that is, since the last performed maintenance activity, can be written as

$$R_i(\tau) = \varphi_i\left(\tau, \Delta, t_k^i\right) \cdot \mathcal{D}_i \tag{6.2.4}$$

being $\varphi_i\left(\tau, \Delta, t_k^i\right)$ the probability that a fault will occur within $\tau + \Delta$, conditioned by the fact that it has not occurred yet. Since this process has memory, it turns out that $\varphi_i\left(\tau, \Delta, t_k^i\right)$ is a non-decreasing function.

With this hypothesis, and given the Cumulative Distribution Function (CDF) $F\left(\tau, \theta\left(t_k^i\right)\right)$ of the stochastic variable representing the fault occurrence time, the function $\varphi_i\left(\tau, \Delta, t_k^i\right)$ can be written as the probability that the fault occurs within a certain interval Δ starting in τ, given that the fault has not occurred in τ, that is

$$\varphi_i\left(\tau, \Delta, t_k^i\right) = \mathbb{P}_i\left\{t \leq \tau + \Delta | t \geq \tau, \theta\left(t_k^i\right)\right\} = \frac{F\left(\tau + \Delta, \theta(t_k^i)\right) - F(\tau, \theta(t_k^i))}{1 - F(\tau, \theta(t_k^i))} \tag{6.2.5}$$

It is worth noting that the fault occurrence time results to be a stochastic process since the CDF parameters $\theta\left(t_k^i\right)$ depend on time and, in particular, on the times at which maintenance activities have been performed. Note that the model in (6.2.5) does not introduce assumptions about the shape of the fault occurrence time CDF but only requires, in a very general form, that a time dependence of the considered stochastic variables is able to represent the degradation process.

Coming back to the model in Eq. (6.2.4), given a maximum tolerable risk $\bar{R}_i$ for the asset i, and the losses $\mathcal{D}_i$ generated by its fault, it is easy to compute the maximum tolerable value of the conditioned fault probability in Eq. (6.2.5), as $\bar{\varphi}_i = \bar{R}_i / \mathcal{D}_i$. In any case, is assumed that, given τ, it is possible to determine, analytically or numerically, the interval duration Δ during which $\varphi_i\left(\tau, \Delta, t_k^i\right)$ is below the value $\bar{\varphi}_i$, that is

$$\Delta\left(t_k^i\right) = \varphi_i^{-1}\left(\bar{\varphi}_i, t_k^i\right) - \tau \tag{6.2.6}$$

where the notation $\Delta\left(t_k^i\right)$ points out that this interval depend on the last performed maintenance activity. Nevertheless, aiming at keeping the notation simple, this dependence is indicated only when necessary.

Focusing on technological assets, the fault occurrence time can be described by a Weibull stochastic process:

$$F\left(t, \lambda(t_k^i), \eta(t_k^i)\right) = 1 - \exp\left(-\left(\frac{t}{\lambda(t_k^i)}\right)^{\eta(t_k^i)}\right) \tag{6.2.7}$$

being $\lambda(t_k^i)$ and $\eta(t_k^i)$ the scale and shape parameters determined by the maintenance activity performed in t_k^i.

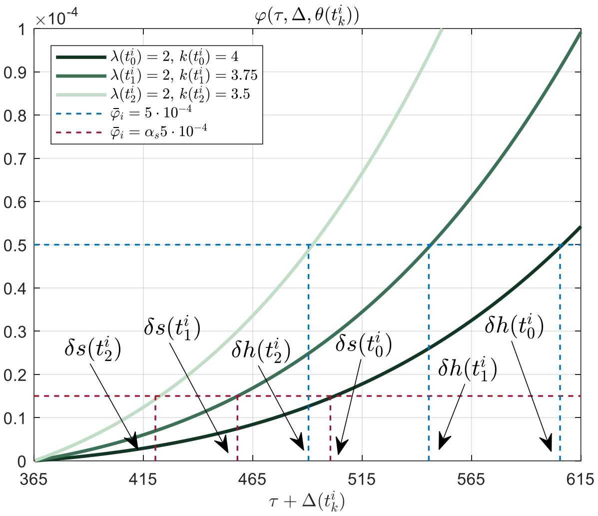

FIGURE 21 COMPARISON BETWEEN THE CONDITIONED PROBABILITIES $\varphi_i(\tau, \Delta(t_0^i), t_0^i)$, $\varphi_i(\tau, \Delta(t_1^i), t_1^i)$, AND $\varphi_i(\tau, \Delta(t_2^i), t_2^i)$ EVALUATED IN $\tau = t_k^i + 365$ DAYS, THAT IS, A YEAR AFTER THE MAINTENANCE

It is well known that the Weibull distribution can be suitably used to represent the time-dependent probability that a fault occurs within a certain interval, given that it is not occurred since the last maintenance activity. In addition, the time dependence of the parameters $\lambda(t_k^i)$ and $\eta(t_k^i)$ allows to model the fact that the maintenance activity performed in t_k^i is not able to restore the initial state of an asset. In other words, the decrease with the time sequence t_k^i, $k = 0,1, \dots$, of the scale and shape parameters $\lambda(t_k^i)$ and $\eta(t_k^i)$ values points out that it is impossible to restore the initial conditions characterizing the new asset. In fact, $\lambda(t_k^i) \leq \lambda(t_{k-1}^i)$ and $\eta(t_k^i) \leq \eta(t_{k-1}^i)$ indicate that, given the same interval Δ, a fault in $[t_k^i, t_k^i + \Delta]$ is more probable than a fault in $[t_{k-1}^i, t_{k-1}^i + \Delta]$.

Once the function $\varphi_i(\tau, \Delta_h, t_k^i)$ is determined and thanks to the assumption that it is possible to compute the time $\tau + \Delta_h$ at which it reaches the threshold $\bar{\varphi}_i$, the so-called *hard deadline* is determined as the instant

$$\delta h_i(t_k^i) = \tau + \Delta_h(t_k^i) \tag{6.8}$$

where the dependency of δh on t_k^i points out that they can vary after each maintenance activity.
Analogously, a soft deadline can be determined as the instant

$$\delta s_i(t_k^i) = \tau + \Delta_s(t_k^i) \tag{6.2.8}$$

such that $\varphi_i(\tau, \Delta_s, t_k^i) = \alpha_s \bar{\varphi}_i$, being $\alpha_s \in (0,1)$ a parameter indicating a tolerable percentage of the fault probability $\bar{\varphi}_i$. Analogously, it is possible to define maintenance requirement starting time as the instant

$$\delta g_i(t_k^i) = \tau + \Delta_g(t_k^i) \tag{6.2.9}$$

such that $\varphi_i(\tau, \Delta_g, t_k^i) = \alpha_g \bar{\varphi}_i$, being $\alpha_g \in (0, a_s)$ a parameter analogous to α_s.

The instant δg expresses the idea that it is possible to avoid considering, in the mathematical programming model, all the assets with release time greater than the planning horizon, as their degradation state will not reach a non-negligible degraded condition before the last maintenance activity of the considered assets has to be concluded. Aiming at simplifying the notation, in the following the explicit dependence of the deadlines and of the release time on the time t_k^i will be dropped whenever not necessary.

Some examples of the soft and hard deadlines for a balise maintenance (characterized by a Weibull fault probability) are reported in Fig. 21, where it is possible to note that the slopes of the functions $\varphi_i(\tau, \Delta_s, t_k^i)$ increases with t_k^i and, consequently, the deadlines decrease.

It is worth mentioning that, considering the above described decrease of the deadlines after the execution of the maintenance activities, the cost reduction can be evaluated comparing the schedule defined by a planned routine maintenance and the proposed approach. In this way, the number of unnecessary maintenance activities that would be executed and the number of failure, that would not be prevented using the cyclical approach, can be evaluated. In particular, Fig.21 shows that considering the dark green curve with parameters $\lambda(t_0^i)$ and $\eta(t_0^i)$ and performing the maintenance activity at $\tau + \Delta\left(t_0^i\right) = 500$ days, the soft deadline and the hard deadline are not exceeded. After the first maintenance intervention, considering the medium green curve (characterized by the parameters $\lambda\left(t_1^i\right) \leq \lambda\left(t_0^i\right)$ and $\eta\left(t_1^i\right) \leq \eta\left(t_0^i\right)$), the execution of the intervention at $\tau + \Delta\left(t_1^i\right) = 500$ days implies that the soft deadline is exceeded while the hard deadline is not reached. Finally, when the second maintenance intervention is executed, if a further maintenance is performed at $\tau + \Delta\left(t_2^i\right) = 500$, both the soft and the hard deadlines are exceeded. To cope with this problem, the inter-time between two consecutive maintenance activities could be reduced. Nevertheless, fix reductions may lead to an avoidable reduction of the asset useful life and to an increase of maintenance costs.

Coming back to the proposed approach, analogously to the previous risk-based approach, being $\tau \geq t_k^i$ the instant at which the last risk assessment has been performed, a generic asset i is in the following states:

- a *good condition* $\forall t \in (\tau, \delta g_i)$, that is, the asset is characterized by a negligible degradation;
- a *maintenance awaiting condition* $\forall t \in (\delta g_i, \delta s_i)$, that is, the asset is characterized by a small degradation, although its degradation process has already begun;
- an *acceptable degraded condition* $\forall t \in (\delta s_i, \delta h_i)$, that is, the asset is characterized by a non-negligible degradation, although it is still sufficiently far from a an high fault risk. In this state, the asset needs a maintenance activity;
- an *unacceptably degraded condition* $\forall t \geq \delta h_i$ that is, the state of the asset is characterized by a significant degradation that should never be reached.

In this way, the soft and hard deadlines are determined according to the asset state, and in particular according to the distance from the thresholds $\bar{\varphi}_i$, $\alpha_s\bar{\varphi}_i$, and $\alpha_g\bar{\varphi}_i$. This means that an asset that is close to exceed the threshold will have a close deadline, implying the possibility of considering different severity levels between assets in the same state.

Once the state of each asset is identified and the thresholds determined, the model proposed in the following Chapter 7 allows to plan the maintenance activities via a Mixed Integer Linear

Programming (MILP) problem. As already mentioned, in doing so, only the assets that have reached the Waiting maintenance condition are considered.

To conclude, it is worth remarking that the model, proposed in this section, can be also applied considering different stochastic processes, provided that they are able to capture the degradation dynamics of assets.

6.3 ROLLING HORIZON APPROACH

As mentioned, the model is used within a rolling horizon framework, taking into account the available real-time information and the updated inputs from the predictive model.

In this section, the Rolling Horizon (RH) framework [80] for the maintenance scheduling optimization is described. Such a generalization of the optimization procedure is necessary in practice since, for instance, some maintenance activities cannot sometimes be performed as planned due to working delays, and/or to sudden, unpredicted, asset faults.

With this approach, ΔT time units before each decision time, that is in $t_k - \Delta T$, $\forall k$, the problem is stated and solved considering the degradation curves and the consequent parameters deriving from the Stochastic Process $\delta_i(t, t_k - \Delta T)$, $\forall i$. It is worth saying that such a solution is determined for the RH window $(t_k, t_k + \Theta)$ but it is applied only in the interval (t_k, t_{k+1}). In fact, as the time passes and the instant $t_{k+1} - \Delta T$ approaches, the parameters of the Stochastic Process are updated determining $\delta_i(t, t_{k+1} - \Delta T)$ and a new instance of the problem is stated and solved for the RH window $(t_{k+1}, t_{k+1} + \Theta)$. In doing so, the assets that have to be maintained in the interval $(t_k + \Theta, t_{k+1} + \Theta)$ are inserted. As regards the initial solution of the problem stated in $t_{k+1} - \Delta T$, it consists of the optimal sequence determined for the previous instance without the assets maintained in (t_k, t_{k+1}) plus the new assets added in the end of the sequence.

Therefore, in t_{k+1} the new solution computed in $(t_{k+1} - \Delta \mathrm{T}, t_{k+1})$ is applied. Finally, by setting $k = k + 1$, the RH framework can manage the maintenance scheduling for an indefinite time. A sketch of the above described framework is reported in Fig.22.

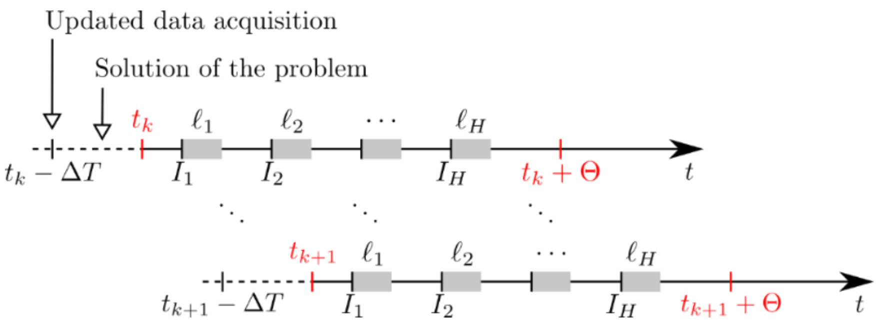

FIGURE 22 ROLLING-HORIZON APPROACH

Figure 23 shows the logical framework of the Rolling-horizon approach applied to the proposed risk-based model. Summing up, given, for example, the Gaussian degradation model, described in Section 6.2, and the functions of the expectation $\bar{\delta}_i(t)$ and the standard deviation $\sigma_i(t)$, it is possible to evaluate the probability of failure, as the probability that the degradation measure is greater than the critical threshold value.

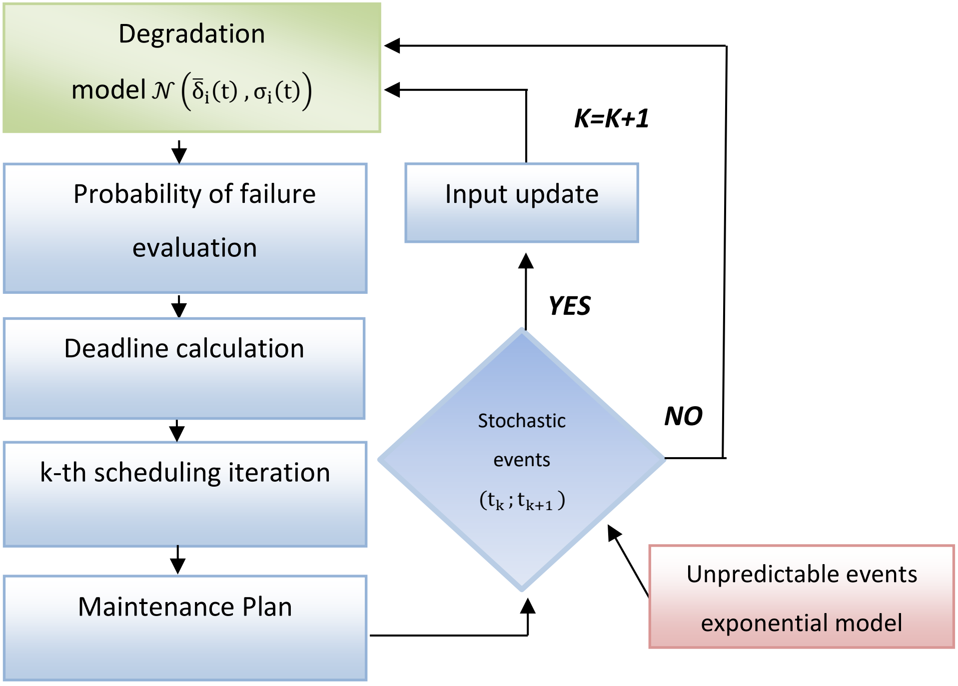

FIGURE 23 MODEL LOGICAL FRAMEWORK.

Therefore, given the above model, the status of each asset can be evaluated.

Once the state of each asset is identified, the failure probabilities can be translated into suitable problem constraints and it is possible to define the soft and hard deadlines of the scheduling problem.

Given these thresholds, the scheduling model allows to plan the maintenance activities by evaluating the completion time of the activity on the asset only when it has reached the *Waiting maintenance condition*, and imposing to finish before the forecast failure, the hard deadline, thus obtaining a Mixed Integer Linear Programming (MILP) problem. In doing so, once fixed a reference time horizon, the risk can be used as a suitable weight, thus allowing that the maintenance activities are properly prioritized within the time horizon.

At the step k, the problem is solved and the maintenance plan is calculated. At the next step, $k + 1$, the possible changes in the input data, due to unpredicted events such as delays or stochastic faults, are checked and, in case, inputs are updated according to new available information. In both cases, at time $\mathrm{t_{k+1}}$ the new expectation and standard deviation values of the degradation parameter are assumed for the evaluation of the new scheduling deadlines.

7. MODEL DESCRIPTION

In this chapter, three different extensions of the optimization model are described considering the following scenarios:

- Deterministic scheduling of maintenance activities along a railway line.
- Deterministic scheduling of maintenance activities along a railway network.
- Stochastic scheduling of maintenance activities along a railway network.

7.1 DETERMINISTIC SCHEDULING PROBLEM FOR A RAILWAY LINE

In this section, the MILP formulation of the deterministic scheduling of predictive maintenance activities along a railway line is described [131]. In doing so, without losing generality, it is assumed $t_k = 0$. The assumptions considered in the MILP problem formulation are the following:

- only the assets that have hard deadline in the interval $(t_k, t_k + \mathrm{T})$ and initial time $t_i \leq t_k + \mathrm{T}$ are considered, being T a suitably chosen horizon;
- the interval $(t_k, t_k + \mathrm{T})$ is subdivided into non-continuous train-free sub-intervals during which the train circulation is forbidden and maintenance activities can be performed;
- all maintenance teams are available in t_k;
- the setup time of each maintenance activity processed by each maintenance team is sequence-dependent;
- the processing times of the maintenance activities by different maintenance teams on the same asset can be different;
- each maintenance team can perform only a maintenance activity at a time;
- each asset is characterized by distinct soft and hard deadlines;
- no preemption of maintenance activities is allowed;
- the maintenance activities can be tardy only with respect to the soft deadlines;
- maintenance teams are available throughout all the scheduling time horizon;
- the maintenance activity on all the assets can be processed by any free work team;

The relevant problem notation is reported in Tab.4.

TABLE 4 NOTATION OF THE SCHEDULING MODEL FOR A RAILWAY LINE

INDEXES	
$i = 0, \dots, \lvert\mathcal{A}\rvert$ $j = 1, \dots, \lvert\mathcal{A}\rvert + 1$	indexes of assets
$i = 0$	dummy activity artificially introduced to correctly identifying the first real one
$j = \lvert\mathcal{A}\rvert + 1$	dummy activity artificially introduced to correctly identifying the last real one
SETS	
$\mathcal{T}$	set of maintenance teams, and $\lvert\mathcal{T}\rvert$ its cardinality
$\mathcal{H}_{t_k,\mathrm{T}}$	set of the train-free sub-intervals of the interval $(t_k, t_k + \mathrm{T})$

CONSTANTS	
π_i	processing time of maintenance activity on asset i
T	maintenance planning horizon
ℓ_h	length of the h^{th} sub-interval in $\mathcal{H}_{t_k,\mathrm{T}}$
I_h	initial time of the h^{th} sub-interval in $\mathcal{H}_{t_k,\mathrm{T}}$
$s_{i,j,m}$	set-up time for assigning activity on asset j after activity on asset i to maintenance team m. Such a set-up time can be interpreted as the sum $s_{i,j,m} = c_{i,j,m} + c_{i,m}^{-} + c_{j,m}^{+}$, where $c_{i,j,m}$ is the team m travel time from the asset i to the asset j, $c_{i,m}^{-}$ is the removal time of the repair yard of the completed maintenance activity on asset i by the team m, and $c_{j,m}^{+}$ is the preparation time of the repair yard of the next scheduled maintenance activity on asset j by team m
ω_i	maintenance activity priority of the asset i
B	integer suitably chosen to approximate +∞ in the constraints
VARIABLES	
c_i	completion time of maintenance activity on the asset i
t_i	starting time of maintenance activity on the asset i
q_i	tardiness of maintenance activity on asset i
$x_{i,j,m}$	binary *sequence variable* equal to 1 if the activity on asset j is performed immediately after the activity on asset i by maintenance team m, and 0 otherwise
$w_{i,h}$	binary variable equal to 1 if the activity on asset i is performed in the time interval h and 0 otherwise

Moreover, as regards the notation, it is assumed that:

- dh_i is the hard deadline of the maintenance activity on asset i;
- ds_i is the soft deadline of the maintenance activity on asset i.

Given the above notation, the optimization problem can be formalized as

$$\min \sum_{i=1}^{|\mathcal{A}|} (\omega_i c_i + \omega_i q_i) \tag{7.1.1}$$

subject to:

$$c_i = t_i + \sum_{m=1}^{|\mathcal{T}|} \sum_{j=1, j\neq i}^{|\mathcal{A}|+1} \pi_i x_{i,j,m} \qquad \forall\, i \tag{7.1.2}$$

$$c_i \leq dh_i \qquad \forall\, i \tag{7.1.3}$$

$$q_i = \max(0, c_i - ds_i) \qquad \forall\, i \tag{7.1.4}$$

$$t_j \geq I_h w_{j,h} + s_{0,j,m} - B(1 - x_{0,j,m}) \qquad \forall\, j, \forall\, m, \forall\, h \tag{7.1.5}$$

$$t_j \geq \max\big(I_h w_{j,h}, t_i + \pi_{i,m}\big) + s_{i,j,m} - B\big(1 - x_{i,j,m}\big) \qquad \forall\, m, \forall\, h, \forall j, \forall\, i, i \neq j \tag{7.1.6}$$

$$\sum_{m=1}^{|\mathcal{T}|} \sum_{i=0, i\neq j}^{|\mathcal{A}|} x_{i,j,m} = 1 \qquad \forall\, j \tag{7.1.7}$$

$$\sum_{m=1}^{|\mathcal{T}|} \sum_{j=1, j\neq i}^{|\mathcal{A}|+1} x_{i,j,m} = 1 \qquad \forall\, i \tag{7.1.8}$$

$$\sum_{j=1}^{|\mathcal{A}|+1} x_{0,j,m} \leq 1 \qquad \forall\, m \tag{7.1.9}$$

$$\sum_{k=1, k\neq j}^{|\mathcal{A}|+1} x_{j,k,m} - \sum_{i=0, i\neq j}^{|\mathcal{A}|} x_{i,j,m} = 0 \qquad \forall\, j, \forall\, m \tag{7.1.10}$$

$$c_i \leq I_h + \ell_h + B\big(\, 1 - w_{i,h}\big) \qquad \forall\, i, \forall\, h \tag{7.1.11}$$

$$\sum_{h=1}^{\mathcal{H}_{t_k,\mathrm{T}}} w_{i,h} = 1 \qquad \forall\, i \tag{7.1.12}$$

$$t_i, c_i, q_i \in \mathbb{R}_{+,0} \qquad \forall\, i \tag{7.1.13}$$

$$x_{i,j,m}, w_{i,h} \in \{0,1\} \qquad \forall\, m, \forall\, h, \forall j, \forall\, i, i \neq j \tag{7.1.14}$$

where:

- the constraints in Eq. (7. 1.2) define the completion times of the maintenance activities;

- the constraints in Eq. (7. 1.3) guarantee that the maintenance activity on each asset is completed before the relevant hard deadline;
- the constraints in Eq. (7. 1.4) define the tardiness of the maintenance activities
- the constraints in Eq. (7. 1.5) define the initial setup time;
- the constraints in Eq. (7. 1.6) define the precedence relation between the activities on the assets i and j and guarantee that if an activity is performed in the train-free interval h, it starts after the beginning of such an interval;
- the constraints in Eq. (7. 1.7) and Eq. (7. 1.8) guarantee that every maintenance activity has exactly one predecessor and one successor, respectively. In addition, these constraints state that each maintenance team can execute at most one activity at a time;
- the constraints in Eq. (7. 1.9) guarantee that only one activity is scheduled as first for each maintenance team;
- the constraints in Eq. (7. 1.10) state that any pair of predecessor/successor activities has to be assigned to the same maintenance team;
- the constraints in Eq. (7. 1.11) guarantee that all maintenance activities finish within the train-free interval;
- the constraints in Eq. (7. 1.12) guarantee that each activity is assigned only to a single time interval;
- the constraints in Eq. (7. 1.13) and Eq. (7. 1.14) define the problem variables.

7.2 DETERMINISTIC SCHEDULING PROBLEM FOR A RAILWAY NETWORK

In this section, the MILP formulation of the proposed risk-based scheduling is described for the maintenance of a railway network [132].

In particular, the rail network is represented as a graph $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$; $\mathcal{N}$ is the set of nodes; $\mathcal{L}$ is the set of links connecting the nodes. The rail stretches are modelled as links and the rail stations, at the beginning and at the end of each rail stretch, are represented as nodes.

It is assumed that the origin and the destination of each maintenance team trip are not a priori known but are optimally chosen by the model.

Let i, $i = 0, \dots, |\mathcal{A}|$ be a generic rail asset to be maintained, and let t_i be the starting time of the related maintenance activity.

The detailed notation is reported in Tab.5 that integrates the definitions already provided in Tab.3.

TABLE 5 NOTATION OF THE SCHEDULING MODEL FOR A RAILWAY NETWORK

INDEXES	
$i = 0, \dots, \vert\mathcal{A}\vert$ $j = 1, \dots, \vert\mathcal{A}\vert + 1$	indexes of assets
$i = 0$	dummy activity artificially introduced to correctly identifying the first real one
$j = \vert\mathcal{A}\vert + 1$	dummy activity artificially introduced to correctly identifying the last real one
SETS	
$\mathcal{T}$	set of maintenance teams, and $\vert\mathcal{T}\vert$ its cardinality
$\mathcal{N}$	set of railway nodes
$\mathcal{L}$	set of railway links
$\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$	graph representing the considered railway network
$\mathcal{H}_{t_k,\mathrm{T}}$	set of the train-free sub-intervals of the interval $(t_k, t_k + \mathrm{T})$
MATRICES	
A	adjacency matrix of the graph: element $A_{n,l}$ is equal to 1 if node $n \in \mathcal{N}$ and node $l \in \mathcal{N}$ are connected by link $(n, l) \in \mathcal{L}$, and 0 otherwise
M	inclusion matrix: element $M_{i,(n,l)}$ is equal to 1 if asset i is located on link $(n, l) \in \mathcal{L}$, and to 0 otherwise
D	matrix of the distances between the rail nodes
CONSTANTS	
$\pi_{i,g}$	processing time of maintenance activity on asset i by maintenance team g

T	maintenance planning horizon
ℓ_h	length of the h^{th} sub-interval in $\mathcal{H}_{t_k,\mathrm{T}}$
I_h	initial time of the h^{th} sub-interval in $\mathcal{H}_{t_k,\mathrm{T}}$
$s_{i,j,m}$	set-up time for assigning activity on asset j after activity on asset i to maintenance team m. Such a set-up time can be interpreted as the sum $s_{i,j,m} = c_{i,j,m} + c^{-}_{i,m} + c^{+}_{j,m}$, where $c_{i,j,m}$ is the team m travel time from the asset i to the asset j, $c^{-}_{i,m}$ is the removal time of the repair yard of the completed maintenance activity on asset i by the team m, and $c^{+}_{j,m}$ is the preparation time of the repair yard of the next scheduled maintenance activity on asset j by t team m
ω_i	maintenance activity priority of the asset i
B	integer suitably chosen to approximate +∞ in the constraints
α_j	weight of the objective function j^{th} term, chosen according to planner intentions
VARIABLES	
$c_{i,m}$	completion time of maintenance activity on the asset i performed by maintenance team m
$t_{i,m}$	starting time of maintenance activity on the asset i performed by maintenance team m
q_i	tardiness of maintenance activity on asset i
$x_{i,j,m}$	binary *sequence variable* equal to 1 if the activity on asset j is performed immediately after the activity on asset i by maintenance team m, and 0 otherwise
$y_{i,m}$	binary assignment variable equal to 1 if the activity on asset i is performed by maintenance team m, and 0 otherwise
$u_{n,m}$	binary variable equal to 1 if node $n \in \mathcal{N}$ is the origin of maintenance team m

$v_{n,m}$	binary variable equal to 1 if node $n \in \mathcal{N}$ is the destination of maintenance team m
$z_{(n,l),m}$	binary variable equal to 1 if link $(n,l) \in \mathcal{L}$ belongs to the optimal path of maintenance team $m \in \mathcal{T}$, and to 0 otherwise
$w_{i,m,h}$	binary variable equal to 1 if the activity on asset i is performed by the maintenance team m in the time interval h and 0 otherwise
P_m	total path among the sites of the assets travelled by the working team $m \in \mathcal{T}$
$\Delta P_{m,g}$	difference between the paths travelled by the working team m and g
$\Delta E_{m,g}$	difference between the number of jobs assigned to working team m and g

The assumptions considered in the MILP problem formulation are the following:

- only the assets that have hard deadline in the interval $(\tau, \tau + \mathrm{T})$ and initial time $\delta g \leq \tau + \mathrm{T}$ are considered, being T a suitably chosen *horizon;*
- the interval $(\tau, \tau + \mathrm{T})$ is subdivided into non-continuous *train-free sub-intervals* during which the train circulation is forbidden and maintenance activities can be performed;
- all maintenance teams are available in τ;
- the set-up time of each maintenance activity processed by each maintenance team is sequence-dependent;
- the processing times of the maintenance activities of different maintenance teams on the same asset can be different;
- each maintenance team can perform only a maintenance activity at a time;
- each asset is characterized by distinct soft and hard deadlines;
- no pre-emption of maintenance activities is allowed;
- the maintenance activities can be tardy only with respect to the soft deadlines;
- the maintenance teams are available throughout all the scheduling time horizon;
- the maintenance activity on all the assets can be processed by any free maintenance team.

The optimization problem is described assuming that the risk analysis is performed in $\tau > t_k^i, \forall i \in \mathcal{A}$, and consequently, the deadlines depend on the last maintenance activities performed, for any asset, in t_k^i. Therefore, given the above notation, the optimization problem can be formalized as

$$\min \alpha_1 \left(\sum_{m=1}^{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{A}|} \omega_i c_{i,m} + \sum_{i=1}^{|\mathcal{A}|} \omega_i q_i \right) + \alpha_2 \sum_{m=1}^{|\mathcal{T}|} P_m + \alpha_3 \sum_{m=1}^{|\mathcal{T}|} \sum_{\substack{g=1, \\ g \neq m}}^{|\mathcal{T}|} \left(\Delta P_{m,g} + \Delta \mathrm{E}_{m,g} \right) \tag{7.2.1}$$

subject to:

$$c_{i,m} = \ max(0, t_{i,m} + \pi_{i,m} - B(1 - y_{i,m}) \qquad \forall i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{7.2.2}$$

$$t_{i,m} \geq \delta g(t_k^i) \qquad \forall i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{7.2.3}$$

$$c_{i,m} \leq \delta h(t_k^i) \qquad \forall i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{7.2.4}$$

$$q_i \ = \max(0, c_{i,m} - \delta s(t_k^i)) \qquad \forall i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{7.2.5}$$

$$t_{j,m} \geq c_{i,m} + \ s_{i,j,m} - B(1 - x_{i,j,m}) \qquad \forall m \in \mathcal{T}, \forall j, i \in \mathcal{A} \tag{7.2.6}$$

$$t_{j,m} \geq \ s_{0,j,m} - B(1 - x_{0,j,m}) \qquad \forall m \in \mathcal{T}, \forall j \in \mathcal{A} \tag{7.2.7}$$

$$x_{i,j,m} \leq min\{y_{i,m}, y_{j,m}\} \qquad \forall m \in \mathcal{T}, \forall i, j \in \mathcal{A}: i \neq j \tag{7.2.8}$$

$$\sum_{j=1}^{|\mathcal{A}|+1} x_{0,j,m} \leq 1 \qquad \forall m \in \mathcal{T} \tag{7.2.9}$$

$$\sum_{m=1}^{|\mathcal{T}|} \sum_{i=0, i \neq j}^{|\mathcal{A}|} x_{i,j,m} = 1 \qquad \forall j \in \mathcal{A} \tag{7.2.10}$$

$$\sum_{m=1}^{|\mathcal{T}|} \sum_{j=1, j \neq i}^{|\mathcal{A}|+1} x_{i,j,m} = 1 \qquad \forall i \in \mathcal{A} \tag{7.2.11}$$

$$\sum_{e=1, e \neq j}^{|\mathcal{A}|+1} x_{j,e,m} - \sum_{i=0, i \neq j}^{|\mathcal{A}|} x_{i,j,m} = 0 \qquad \forall j \in \mathcal{A}, \forall m \in \mathcal{T} \tag{7.2.12}$$

$$t_{j,m} \geq \ I_h w_{j,h} + \ s_{i,j,m} x_{i,j,m} \qquad \forall i, j \in \mathcal{A}, \forall h \in \mathcal{H}_{t_k, \mathrm{T}}, \forall m \in \mathcal{T} \tag{7.2.13}$$

$$c_{i,m} \leq I_h + \ell_h + B(1 - w_{i,h}) \qquad \forall i, \forall h \in \mathcal{H}_{t_k, \mathrm{T}}, \forall m \in \mathcal{T} \tag{7.2.14}$$

$$\sum_{m=1}^{|\mathcal{T}|} \sum_{h=1}^{\left|\mathcal{H}_{t_k,\mathrm{T}}\right|} w_{i,m,h} = 1 \qquad \forall i \in \mathcal{A} \qquad (7.2.15)$$

$$w_{i,m,h} \leq y_{i,m} \qquad \forall h \in \mathcal{H}_{t_k,\mathrm{T}}, \forall m \in \mathcal{T}, \forall i \in \mathcal{A} \qquad (7.2.16)$$

$$\sum_{m=1}^{|\mathcal{T}|} y_{i,m} = 1 \qquad \forall i \in \mathcal{A} \qquad (7.2.17)$$

$$\Delta P_{m,g} = \left|P_m - P_g\right| \qquad \forall g, m \in \mathcal{T} : g \neq m \qquad (7.2.18)$$

$$\Delta \mathrm{E}_{m,g} = \left|\sum_{j=1}^{|\mathcal{A}|} y_{j,m} - \sum_{j=1}^{|\mathcal{A}|} y_{j,g}\right| \qquad \forall g, m \in \mathcal{T} : g \neq m \qquad (7.2.19)$$

$$\sum_{m=1}^{|\mathcal{T}|} \left(z_{(n,l),m} + z_{(l,n),m}\right) \geq 1 \qquad \forall (n,l) \in \mathcal{L} : A_{n,l} = 1 \qquad (7.2.20)$$

$$\sum_{m=1}^{|\mathcal{T}|} \left(z_{(n,l),m} + z_{(l,n),m}\right) \leq |\mathcal{T}| \qquad \forall (n,l) \in \mathcal{L} : A_{n,l} = 1 \qquad (7.2.21)$$

$$z_{(n,l),m} + z_{(l,n),m} \leq 1 \qquad \forall (n,l) \in \mathcal{L} : A_{n,l} = 1 \qquad (7.2.22)$$

$$P_m = \sum_{(n,l) \in \mathcal{L}} z_{(n,l),m} D_{n,l} \qquad \forall m \in \mathcal{T} \qquad (7.2.23)$$

$$\sum_{n=1}^{|\mathcal{N}|} u_{n,m} = 1 \qquad \forall m \in \mathcal{T} \qquad (7.2.24)$$

$$\sum_{n=1}^{|\mathcal{N}|} v_{n,m} = 1 \qquad \forall m \in \mathcal{T} \qquad (7.2.25)$$

$$y_{i,m} - z_{(n,l),m} - z_{(l,n),m} \leq 0 \qquad \begin{array}{l} \forall m \in \mathcal{T}, \forall i \in \mathcal{A}, \\ \forall (n,l) \in \mathcal{L}: \\ A_{n,l} = 1,\ M_{i,(n,l)} = 1 \end{array} \qquad (7.2.26)$$

$$\sum_{\substack{l=1: \\ a_{nl}=1}}^{|\mathcal{N}|} z_{(n,l),m} - \sum_{\substack{l=1: \\ a_{nl}=1}}^{|\mathcal{N}|} z_{(l,n),m} = u_{n,m} - v_{n,m} \qquad \forall n \in \mathcal{N}, \forall m \in \mathcal{T} \qquad (7.2.27)$$

$$u_{n,m} + v_{n,m} \leq 1 \qquad \forall n \in \mathcal{N}, \forall m \in \mathcal{T} \qquad (7.2.28)$$

$$t_{i,m}, c_{i,m}, q_i \in \mathbb{R}_{+,0} \qquad \forall\, i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{7.2.29}$$

$$P_m \in \mathbb{R}_{+,0}, \Delta P_{m,g} \in \mathbb{R}, \Delta \mathrm{E}_{m,g} \in \mathbb{Z} \qquad \forall g, m \in \mathcal{T} \tag{7.2.30}$$

$$x_{i,j,m} \in \{0,1\} \qquad \forall m \in \mathcal{T}, \forall j, i \in \mathcal{A} : i \neq j \tag{7.2.31}$$

$$y_{i,m} \in \{0,1\} \qquad \forall m \in \mathcal{T}, \forall i \in \mathcal{A} \tag{7.2.32}$$

$$w_{i,h} \in \{0,1\} \qquad \forall i \in \mathcal{A}, \forall h \in \mathcal{H}_{t_k,\mathrm{T}} \tag{7.2.33}$$

$$z_{(n,l),m} \in \{0,1\} \qquad \forall m \in \mathcal{T}, \forall (l,n) \in \mathcal{L} \tag{7.2.34}$$

$$u_{n,m}, v_{n,m} \in \{0,1\} \qquad \forall m \in \mathcal{T}, \forall n \in \mathcal{N} \tag{7.2.35}$$

where:

- constraints (7.2.2) define the completion times of the maintenance activities;
- constraints (7.2.3) define the release times of the maintenance activities;
- constraints (7.2.4) guarantee that the maintenance activity on each asset is completed before the relevant hard deadline;
- constraints (7.2.5) define the tardiness of the maintenance activities;
- constraints (7.2.6) define the precedence relation between the activities on the assets i and j;
- constraints (7.2.7) define the initial set-up times;
- constraints (7.2.8) impose that if the activity on asset j is performed soon after the activity on asset i by maintenance team m, both the asset i and the asset j are assigned to the same maintenance team m;
- constraints (7.2.9) guarantee that at most one activity is scheduled as the first work of each maintenance team;
- constraints (7.2.10) and (7.2.11) guarantee that every maintenance activity has exactly one predecessor and one successor, respectively;
- constraints (7.2.12) state that a predecessor/successor pair of activities has to be assigned to the same maintenance team m;
- constraints (7.2.13) guarantee that if an activity is performed in the train-free interval h, it starts after the beginning of such an interval;

- constraints (7.2.14) guarantee that all maintenance activities finish within the train-free interval;
- constraints (7.2.15) guarantee that each activity is only planned in a single time interval;
- constraints (7.2.16) guarantee that only the maintenance activities assigned to the maintenance team m can be planned in a time interval on that machine.
- constraints (7.2.17) guarantee that each activity is only assigned to a single maintenance team;
- constraints (7.2.18) define the path differences between one maintenance team and each other;
- constraints (7.2.19) define the differences in the number of jobs assigned to one maintenance team and the others;
- constraints (7.2.20) and (7.2.21) respectively state that, if $(n, l) \in \mathcal{L}$, (n, l) has to be assigned at least to one and at most to two maintenance teams;
- constraints (7.2.22) guarantee that each maintenance team can cross link (n, l) in only one direction;
- constraints (7.2.23) define the length of the path for each maintenance team;
- constraints (7.2.24) guarantee that each maintenance team m has only one origin;
- constraints (7.2.25) guarantee that each maintenance team m has only one destination;
- constraints (7.2.26) state that if asset i is located on link (n, l) and maintenance team m crosses that link, i may be assigned to maintenance team m or not;
- constraints (7.2.27) define all the possible paths for each maintenance team m from its origin $u_{n,m}$ to its destination $v_{n,m}$. The peculiarity of such constraints stands in the fact that the origin and the destination of all the maintenance teams are optimally chosen by the model. In fact, these constraints depend on the values assumed by the right-hand-side variables, whose values are also constrained by (7.2.28), which ensures that the same node n cannot be origin and destination of the same maintenance team m;
- constraints (7.2.29)-(7.2.35) define the problem variables.

It is worth noting that the weights α_j in the objective function are chosen by the maintenance service provider, according to the strategic goals of its organization. Anyway, this choice cannot affect the safety level of the system performance, since the constraints in (7.2.21) guarantee that the hard deadlines are always fulfilled. If the maintenance service provider gives a low value to the weight α_1, making the soft deadlines fulfillment negligible in the cost function, the safe condition is still guaranteed. In other words, even if the maintenance can be performed very late with respect to the soft deadlines, the assets still remain in an acceptable degraded condition.

Another aspect to be pointed out is that, even if the monetary cost of the operations is not explicitly indicated in the objective function, it is implicitly considered in the model, since planning the maintenance activities before the hard deadline allows to avoid failure and costly service interruptions. Moreover, the cost reduction is modelled in terms of reduction of the time needed to complete the maintenance activities and optimization of resources utilization, through the definition of the path of each maintenance team.

7.3 STOCHASTIC SCHEDULING PROBLEM FOR A RAILWAY NETWORK

In this section, the formulation of the scheduling problem as a stochastic linear programming model is proposed [133]. In doing so, the fundamental formulation and properties of the so-called two-stage stochastic linear programs with recourse are recalled [26].

In such a kind of problems, it is assumed that:

- the optimal values of the so-called first-stage variables x, have to be chosen in advance, with only partial information about some inputs of the problem that are modelled as stochastic variables H;
- the remaining variables y, indicated as second-stage variables, can be chosen later, when more information is available, that is when the realizations η of H become known.

The general problem formulation can be written as

$$\min z(x,\eta)$$

$s.t.$

$$z(x,\eta) = c^T x + E_H[\min q(\eta)^T\, y(\eta)] \tag{7.3.1}$$

$$Ax = b$$

$$Wy(\eta) = g(\eta) - T(\eta)x$$

$$x \geq 0, y(\eta) \geq 0$$

where A, b, and c are the usual constant coefficient of deterministic linear programming problems, whereas $T(\eta)$, W, and $g(\eta)$ are the realization-dependent coefficients that relate the first-stage and second-stage variables. It is worth noting in (7.3.1) that the second stage variables $y(\eta)$ depend on the first stage variables x, via the random matrix $T(\eta)$. Finally, $q(\eta)$ is the vector of the weights of the second-stage variables in the cost function, also depending on the stochastic variables H. The solution approach to the problem in (7.3.1) strongly depends on the kind of stochastic variables H, that is, if they are discrete or continuous. Nevertheless, with the aim of simplification, a discrete approximation of continuous variables could be determined by defining n sets I_i, and computing the probability $p_i = p_i(\eta_i) = \mathbb{P}(\eta \in I_i)$ $i = 1 \dots n$, as depicted in Fig. 24.

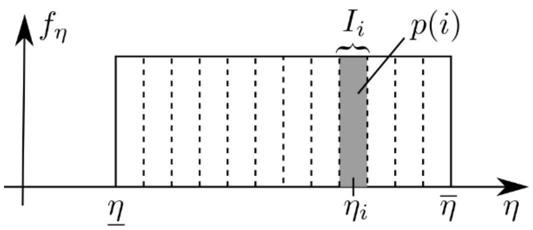

FIGURE 24 DISCRETIZATION OF CONTINUOUS STOCHASTIC VARIABLES.

With this assumption, the expectation in (7.3.1) can be rewritten as for discrete variables, and the general problem takes the form

$$\min z(x,\eta)$$

$s.t.$

$$z(x,\eta) = c^T x + \frac{1}{n}\sum_{i=1}^{n} p_i \,[\min q(\eta_i)^T\, y(\eta_i)] \tag{7.3.2}$$

$$Ax = b$$

$$Wy(\eta_i) = g(\eta_i) - T(\eta_i)x \qquad i = 1 \dots n$$

$$x \geq 0, y(\eta_i) \geq 0 \qquad i = 1 \dots n$$

The stochastic linear problem of (7.3.1) is hence reformulated as a deterministic linear problem, although with the introduction of new vectors of variables $y_i, i = 1 \dots n$. Note that the number of such variables increases with the goodness of the discrete approximation of the continuous stochastic variables.

Therefore, predictive maintenance scheduling is a significant example of such a kind of problems, where:

- the assignments of the maintenance teams are the first-stage variables;
- the maintenance activities sequences are the second stage variables;
- the deadline of interventions are the stochastic variables.

Hereafter the process for evaluating the stochastic deadlines is described.

As mentioned before, the presented approach introduces the concept of risk in railway maintenance scheduling aiming at finding the maintenance schedule that minimizes the risk of failure. In particular, as described in Chapter 6, the general risk definition for a generic asset i and for a given time t, is expressed as the product of the failure probability in any instant $t_f \leq t$, for the consequent lossesd D_i, that is,

$$R_i(t) = \mathbb{P}_i(t_f \leq t) D_i$$

As a consequence, given a maximum tolerable risk $\bar{R}_i$ for the asset i, and the losses D_i generated by its failure, it is easy to compute the maximum tolerable fault probability $\bar{p}_i = \bar{R}_i / D_i$. Therefore, in a *predictive maintenance framework* it is possible to determine the instant at which the failure probability $\mathbb{P}_i(t_f \leq t)$ reaches $\bar{p}_i$ by means of suitable probabilistic models that assess the present and future asset states by means of historical data and real time monitoring data, and providing the so-called *assets degradation curves*.

As a general example such curves, consider the asset degradation curve (Fig.25), where the condition of a generic asset i can be considered unacceptable whenever a certain parameter $\delta_i(t)$ turns out to be greater than a given critical threshold δ^{cr}, that is, $\delta_i(t) \geq \delta^{cr}$.

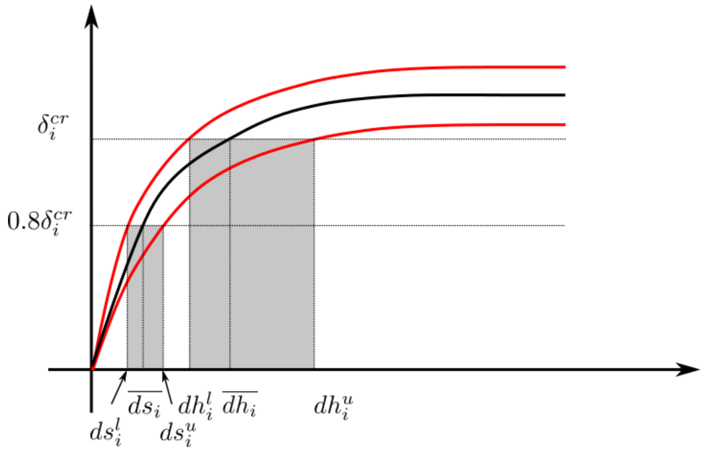

FIGURE 25 ASSET DEGRADATION CURVE

Such a condition must be avoided, since it is associated to a failure probability $\mathbb{P}_i\left(t_f \leq t\right) \geq \bar{p}_i$, and then $R_i(t) \geq \bar{R}_i$.

On the other hand, the asset i can be considered in a tolerable degraded condition whenever the time-varying parameter $\delta_i(t)$ turns out to be greater than a lower threshold $\alpha\delta_i^{cr}$, $0 < \alpha < 1$, that is, $\delta_i(t) \geq \alpha\delta_i^{cr}$. Such a condition can be sometimes reached since it represents the condition $\mathbb{P}_i\left(t_f \leq t\right) < p_i' < \bar{p}_i$ corresponding to a tolerable risk $R_i(t) < R_i' < \bar{R}_i$.

Given these definitions, a *maintenance hard-deadline* dh_{i} can be defined as the instant at which $\delta_i(dh_i) = \delta_i^{cr}$ corresponding to the time limit at which the failure probability $\mathbb{P}_i\left(t_f \leq dh_i\right)$ becomes $\bar{p}_i$. The hard-deadline dh_{i} can never be overcome by the completion time of the maintenance activities of asset i.

Analogously, a *soft-deadline* ds_i can be determined as the instant at which $\delta_i(ds_i) > \alpha\delta_i^{cr}$, corresponding the instant at which $\mathbb{P}_i\left(t_f \leq ds_i\right) = p_i' < p_i$.

As depicted in Fig.25, taking into account the uncertainties in the degradation *trend curves* evaluation represented by the red lines, it is possible to determine the uncertainty intervals $\left[dh_i^l, dh_i^u\right]$ and $\left[ds_i^l, ds_i^u\right]$ gathering the real value of the hard and soft deadlines dh_i and ds_i, which turns out to be stochastic variables. Hence, given the distribution Δ_i of the parameter $\delta_i(t)$

for each t, it is possible to determine the distributions of dh_i and ds_i, via the inverse distribution Δ_i^{-1}.

Hence, being the deadlines stochastic variables, the predictive risk-based maintenance-planning problem is formulated as a stochastic linear programming problem, whose general formulation and notation are described hereafter.

Analogously to the scheduling model described in section 7.2, the rail network is represented as a graph $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$; $\mathcal{N}$ is the set of nodes; $\mathcal{L}$ is the set of links connecting the nodes. The rail stretches are modelled as links and the rail stations, at the beginning and at the end of each rail stretch, are represented as nodes.

It is assumed that the origin and the destination of each maintenance team trip are not a priori known but are optimally chosen by the model.

Let i, $i = 0, \dots, |\mathcal{A}|$ be a generic rail asset to be maintained, and let $t_{k,i}$ be the starting time of the related maintenance activity. Since the stochastic nature of the degradation process, the present and future degradation states of all the considered assets are not perfectly forecast and deterministically known. Hence, as described, the soft and hard deadlines of each maintenance activity are continuous stochastic variables. Nevertheless, as mentioned, with the aim of simplification, a discrete approximation of such variables is defined by considering k samples, $k = 1,2, .. , |\mathcal{K}|$ of the soft and hard deadlines.

Moreover, other assumptions considered in the proposed model are the following:

- all maintenance activities are scheduled within the time horizon T;
- all maintenance teams are available at the initial time;
- the setup time of each maintenance activity processed by each maintenance team is sequence-dependent;
- the process time of each maintenance activity by each maintenance team differs from each other;
- each maintenance team can perform only one maintenance activity at a time;
- each maintenance activity has distinct stochastic soft and hard deadlines;
- no preemption of maintenance activities is allowed;

- tardiness is allowed only with respect to the soft deadlines;
- all the maintenance teams are unrelated and each maintenance activity can be processed by a free maintenance team;
- each maintenance work team is able to perform each maintenance activity. Nevertheless, this assumption can be easily removed by setting some assignation variables to zero.

In Tab.6, the notation considered in the stochastic scheduling problem is introduced.

TABLE 6 NOTATION OF THE STOCHASTIC SCHEDULING MODEL

INDEXES	
$i = 0, \dots, \lvert\mathcal{A}\rvert$ $j = 1, \dots, \lvert\mathcal{A}\rvert + 1$	indexes of assets
$i = 0$	dummy activity artificially introduced to correctly identifying the first real one
$j = \lvert\mathcal{A}\rvert + 1$	dummy activity artificially introduced to correctly identifying the last real one
$k = 1,2,..,\lvert\mathcal{K}\rvert$	Indexes of samples
SETS	
$\mathcal{T}$	set of maintenance teams, and $\lvert\mathcal{T}\rvert$ its cardinality
$\mathcal{N}$	set of railway nodes
$\mathcal{L}$	set of railway links
$\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$	graph representing the considered railway network
$\mathcal{K}$	Set of samples
MATRICES	
A	adjacency matrix of the graph: element $A_{n,l}$ is equal to 1 if node $n \in \mathcal{N}$ and node $l \in \mathcal{N}$ are connected by link $(n, l) \in \mathcal{L}$, and 0 otherwise
M	inclusion matrix: element $M_{i,(n,l)}$ is equal to 1 if asset i is located on link $(n, l) \in \mathcal{L}$, and to 0 otherwise
CONSTANTS	
$\pi_{i,g}$	processing time of maintenance activity on asset i by maintenance team g

T	maintenance planning horizon
$s_{i,j,m}$	set-up time for assigning activity on asset j after activity on asset i to maintenance team m. Such a set-up time can be interpreted as the sum $s_{i,j,m} = c_{i,j,m} + c^{-}_{i,m} + c^{+}_{j,m}$, where $c_{i,j,m}$ is the team m travel time from the asset i to the asset j, $c^{-}_{i,m}$ is the removal time of the repair yard of the completed maintenance activity on asset i by the team m, and $c^{+}_{j,m}$ is the preparation time of the repair yard of the next scheduled maintenance activity on asset j by t team m
ω_i	maintenance activity priority of the asset i
B	integer suitably chosen to approximate +∞ in the constraints
VARIABLES	
$c_{k,i}$	completion time of maintenance activity on the asset i for the k –th sample
$t_{k,i}$	starting time of maintenance activity on the asset i for the k –th sample
$q_{k,i}$	tardiness of maintenance activity on asset i for the k –th sample
$x_{k,i,j,m}$	binary *sequence variable* equal to 1 if the activity on asset j is performed immediately after the activity on asset i by maintenance team m, for the k –th sample, and 0 otherwise
$y_{i,m}$	binary assignment variable equal to 1 if the activity on asset i is performed by maintenance team m, and 0 otherwise
$u_{n,m}$	binary variable equal to 1 if node $n \in \mathcal{N}$ is the origin of maintenance team m
$v_{n,m}$	binary variable equal to 1 if node $n \in \mathcal{N}$ is the destination of maintenance team m
$z_{(n,l),m}$	binary variable equal to 1 if link $(n,l) \in \mathcal{L}$ belongs to the optimal path of maintenance team $m \in \mathcal{T}$, and to 0 otherwise

Moreover, as regards the notation, it is assumed that:

- $dh_{k,i}$ be k –th sample of the hard deadline of the maintenance activity on asset i;

- $ds_{k,i}$ be k –th sample the soft deadline of maintenance activity on asset i.

Given the above notation, the optimization problem can be formalized as

$$\min \sum_{k=1}^{|\mathcal{K}|} \sum_{i=1}^{|\mathcal{A}|} (c_{k,i} + q_{k,i}) \tag{7.3.3}$$

subject to:

$$\sum_{n=1}^{|\mathcal{N}|} u_{n,m} = 1 \qquad \forall\, m \in \mathcal{T} \tag{7.3.4}$$

$$\sum_{n=1}^{|\mathcal{N}|} v_{n,m} = 1 \qquad \forall\, m \in \mathcal{T} \tag{7.3.5}$$

$$\sum_{m=1}^{|\mathcal{T}|} y_{i,m} = 1 \qquad \forall\, i \in \mathcal{A} \tag{7.3.6}$$

$$\sum_{m=1}^{|\mathcal{T}|} z_{(n,l),m} + z_{(l,n),m} = 1 \qquad \begin{array}{c} \forall\, n, l \in \mathcal{L}, \\ \text{s.t } l > n, \\ A_{n,l} = 1 \end{array} \tag{7.3.7}$$

$$y_{i,m} - z_{(n,l),m} - z_{(l,n),m} = 0 \qquad \begin{array}{c} \forall\, i \in \mathcal{A}, \\ \forall\, m \in \mathcal{T}, \\ \forall\, n, l \in \mathcal{L}, \\ \text{s.t. } l > n, \\ A_{n,l} = 1, M_{i,(n,l)} = 1 \end{array} \tag{7.3.8}$$

$$\sum_{\substack{l=1, \\ AD_{nl}=1}}^{|\mathcal{N}|} z_{(n,l),m} - \sum_{\substack{l=1, \\ AD_{nl}=1}}^{|\mathcal{N}|} z_{(l,n),m} = u_{n,m} - v_{n,m} \qquad \begin{array}{c} \forall\, n \in \mathcal{N}, \\ \forall\, m \in \mathcal{T} \end{array} \tag{7.3.9}$$

$$u_{n,m} + v_{n,m} \leq 1 \qquad \begin{array}{c} \forall\, n \in \mathcal{N}, \\ \forall\, m \in \mathcal{T} \end{array} \tag{7.3.10}$$

$$\mathrm{x}_{k,i,j,m} \leq min\{\mathrm{y}_{i,m}, \mathrm{y}_{j,m}\} \qquad \begin{array}{c} \forall\, k \in \mathcal{K}, \\ \forall\, i, j \in \mathcal{A}, \\ \forall\, m \in \mathcal{T} \end{array} \tag{7.3.11}$$

$$c_{k,i} = t_{k,i} + \sum_{m=1}^{M} \sum_{j=1, j \neq i}^{|\mathcal{A}|+1} \pi_{i,m} \, x_{k,i,j,m} \qquad \forall\, k \in \mathcal{K}, \ \forall\, i \in \mathcal{A} \tag{7.3.12}$$

$$c_{k,i} \leq dh_{k,i} \qquad k \in \mathcal{K}, \ \forall\, i \in \mathcal{A} \tag{7.3.13}$$

$$q_{k,i} = \max(0, t_{k,i} - ds_{k,i}) \qquad k \in \mathcal{K}, \ \forall\, i \in \mathcal{A} \tag{7.3.14}$$

$$t_{k,j} \geq s_{0,j,m} - B(1 - x_{k,0,j,m}) \qquad \forall\, k \in K, \ \forall\, j \in \mathcal{A}, \ \forall\, m \in \mathcal{T} \tag{7.3.15}$$

$$t_{k,j} \geq t_{k,i} + \pi_{im} + s_{i,j,m} - B\left(1 - x_{k,i,j,m}\right) \qquad \forall\, k \in \mathcal{K}, \ \forall\, m \in \mathcal{T}, \ \forall i, j \in \mathcal{A}, \ \text{s.t. } i \neq j \tag{7.3.16}$$

$$\sum_{m=1}^{|\mathcal{T}|} \sum_{i=0, i \neq j}^{|\mathcal{A}|} x_{kijm} = 1 \qquad \forall\, k \in \mathcal{K}, \ \forall\, j \in \mathcal{A} \tag{7.3.17}$$

$$\sum_{m=1}^{|\mathcal{T}|} \sum_{j=1, j \neq i}^{|\mathcal{A}|+1} x_{kijm} = 1 \qquad \forall\, k \in \mathcal{K}, \ \forall\, i \in \mathcal{A} \tag{7.3.18}$$

$$\sum_{j=1}^{|\mathcal{A}|+1} x_{k0jm} \leq 1 \qquad \forall\, k \in \mathcal{K}, \ \forall\, m \in \mathcal{T} \tag{7.3.19}$$

$$\sum_{h=1, h \neq j}^{|\mathcal{A}|+1} x_{kjhm} - \sum_{i=0, i \neq j}^{|\mathcal{A}|} x_{kijm} = 0 \qquad \forall\, k \in \mathcal{K}, \ \forall\, j \in \mathcal{A}, \ \forall\, m \in \mathcal{T} \tag{7.3.20}$$

$$t_{ki}, c_{ki}, q_{ki} \in \mathbb{R}_{+,0} \qquad \forall\, k \in \mathcal{K}, \ \forall\, i \in \mathcal{A} \tag{7.3.21}$$

$$x_{kijm}, y_{im} \in \{0,1\} \qquad \forall\, k \in \mathcal{K}, \ \forall\, m \in \mathcal{T}, \ \forall i, j \in \mathcal{A}, \tag{7.3.22}$$

$$z_{nlm}, u_{n,m}, v_{n,m} \in \{0,1\} \qquad \text{s.t. } i \neq j \quad \forall\, m \in \mathcal{T}, \quad \forall n, l \in \mathcal{N} \tag{7.3.23}$$

where:

- the constraints in (7.3.4) guarantee that each maintenance team m has only one origin;
- the constraints in (7.3.5) guarantee that each maintenance team m has only one destination;
- the constraints in (7.3.6) guarantee that each asset i is allocated to only one maintenance team m;
- the constraints in (7.3.7) guarantee that if the arc (n, l) is associated to maintenance team m, the same arc in the opposite direction (l, n) must be assigned to the same maintenance team m;
- the constraints in (7.3.8) guarantee that if the arc (n, l) is associated to maintenance team m and the asset i is situated on that arc, than the asset i must be assigned to the same maintenance team m;
- the constraints in (7.3.9) define all the possible paths for each maintenance team m from his origin $u_{n,m}$ to his destination $v_{n,m}$. The peculiarity of such constraints stands in the fact that the origin and the destination of all the maintenance team are optimally chosen by the model. In fact, these constraints depend on the values assumed by the right-hand-side variables, whose values are also constrained by (7.3.10), which ensures that the same node n cannot be origin and destination of the same maintenance team m;
- the constraints in (7.3.11) define that if the activity on asset j is performed soon after the activity on asset i by work team m for the realization k, both the asset i and the asset j must be assigned to the same work team m;
- the constraints in (7.3.12) define the completion times of the maintenance activities on asset i for the realization k;
- the constraints in (7.3.13) guarantee that each maintenance activity i is completed before the relevant hard deadline of the k –th realization;

- the constraints in (7.3.14) define the tardiness of each activity related to the soft deadline of the k –th realization;
- the constraints in (7.3.15) define the initial setup time;
- the constraints in (7.3.16) define the precedence relation between the activity i and j for the realization k;
- the constraints in (7.3.17) and (7.3.18) guarantee that every maintenance activity has exactly one predecessor and one successor, respectively, for each sample k;
- the constraints in (7.3.19) guarantee that, for each sample k, at most one activity is scheduled as the first work of each work team;
- the constraints in (7.3.20) state that a predecessor/successor couple of activities has to be assigned to the same work team m for each sample k;
- the constraints in (7.3.21), (7.3.22) and (7.3.23) define the problem variables.

7.4 MODEL APPLICABILITY AND DATA REQUIREMENTS

In this section, the real world applicability of the mathematical programming problem and of the relevant solution is briefly discussed, focusing in particular on the input data requirements.

Then, with reference to the sets, the matrixes, and constants in Tab.3-6, the required data results to be:

- the failures occurred in the past;
- the maintenance activities executed in the past;
- the losses caused by the failures;
- the processing time of the maintenance tasks;
- the position of the asset;
- the speed of maintenance teams for moving along the network;
- the availability of maintenance teams;
- the available train-free sub-intervals for the maintenance interventions.

As regards the availability of this information, it is worth mentioning that:

- the failures occurred in the past are usually traced by the infrastructure manager and the maintenance service providers. This means that the failure rates of the assets are usually known;
- the executed maintenance activities are usually known as the infrastructure manager and the maintenance service provider have to take trace of the interventions performed by their teams;
- each maintenance team has usually to communicate the starting and finishing times of the intervention, making possible to know the duration of each maintenance task and, for different kind of interventions, the relevant averages;
- the positions of the assets are known by means the railway line scheme;
- the speed of the maintenance teams moving from an asset site to another depends on the transportation mode they use to travel (working machines, diagnostic trains, etc.);
- the number of available maintenance teams is a-priori established by the maintenance service provider;
- the starting and final times of each train-free sub-interval are provided by the traffic manager.

The losses caused by the failure represent the most difficult input parameter to be defined. The infrastructure manager usually knows the criticality of the different assets and the expected loss caused by their failures, although this information is sensitive and in general not shared. To cope with this problem, the maintenance service provider can estimate these data from field interviewing railway system experts.

To conclude, it is worth noting that some of the above input data, such as the processing time, are not deterministic.

8. RESOLUTION APPROACH

In this Chapter, the resolution approach used to solve the scheduling problems introduced in the previous Chapter 7 is described. As mentioned in Section 3.5, a hybrid approach is considered based on combining Mixed Integer Programming and heuristics.

8.1 ALGORITHMS

8.1.1 MATHEURISTIC FOR RAILWAY LINE MAINTENANCE SCHEDULING

As mentioned, the considered scheduling problem has been proven to be NP-hard and, even for relatively small instances, requires a very long time to be solved. Therefore, in this work a matheuristic solution approach [46] is proposed to solve the instances of the problem. As already mentioned, the matheuristic approach is based on the idea of combining the strength of both approximated metaheuristic and exact methods, leading to an hybrid approach. In particular, the considered matheuristic is described in the algorithm in Fig. 26 [134]. The algorithm starts from an initial admissible top-to-end solution, consisting of executing the activities on the assets as sorted along the line, while assigning the first $N/2$ activities to the first team and the remaining ones to the second team. Then, it iterates until the maximum time limit ΔT is reached or when the impossibility of new improvements is detected. At each iteration, a subset of the sequencing variable is fixed and a subproblem is optimally solved.

It is worth saying that the choice of the top-to-end initial solution is due not only to its simplicity but also because such a strategy is often applied in the real-world maintenance of large geographical distributed assets.

```
Algorithm: Matheuristic algorithm
set h = 0;
set $X_0$ = initial admissible top-to-end solution
repeat
    set $improved = false$;
    repeat
        Randomly chose a subset $X \subseteq X_h$
        Let the variables in $X$ vary and fix the variables in $\bar{X} = X_h \setminus X$
        minimize (7.2.1) with respect to $X$ and subject to (7.1.2) – (7.1.15)
        Let $X^*$ be the optimal solution
        if $f(X^* \cup \bar{X}) < f(X_h)$ then
            set $X_{h+1} = X^* \cup \bar{X}$
            set $improved = true$
        end if
    until $improved$ or all the subsets $X$ have been considered
until not $improved$ or time limit expired
```

FIGURE 26 MATHEURISTIC ALGORITHM.

8.1.2 MATHEURISTIC FOR RAILWAY NETWORK MAINTENANCE SCHEDULING

The optimization scheduling problem on a railway network has also been proven to be NP-hard and, even for relatively small instances, requires a very long time to be solved. Also in this case, a matheuristic solution approach [46], that combines the strength of both approximated metaheuristic and exact methods, leading to an hybrid approach, is proposed to solve the instances of the problem in Eq. (7.2.1-7.2.35). The considered matheuristic algorithm is described in Fig. 27 and Fig. 28.

In particular, the algorithm is based on the decomposition of the main problem into two algorithms, each divided into two main steps:

Algorithm 1 - Find the initial solution:

1. In the first step (Fig. 27), only the paths assignment problem is solved: the assets located on each link are allocated to one maintenance team.
2. In the second step (Fig. 27), the sequencing problem is solved for each team independently. In doing so, if the sequence determined by simply travelling on the directed path from the

origin to the destination of the team m is feasible, such a top to end solution is considered as the initial one for m: in this case, no MILP problems are solved.

3. the third step (Fig. 27), allows to face the case generated by the possible unfeasibility of the step 2.b. If a solution for such a problem does not exist, it is necessary to find different paths and/or different assignments. The tabù search approach makes the solution of the problem in step 1.a, that generates unfeasibility in one of the problems in step 2.b, also unfeasibly for the problem in step 1.a. Note that, in principle, this step could reduce the set of feasible solutions of the problem in step 1.a to the empty set.

```
Algorithm 1: Initial Solution Generation

Initialization  Let $t_{ass}$ be the maximum time for the solution of the task-team
                assignment problem
                Let $t_{ord}$ be the maximum time for the activity sorting problem
Step 1:         Task Assignment problem
1.a             While ($t < t_{ass}$ or the optimum is found)
                    Minimize (8.1.2.1) only with respect to the constraints
                    (7.2.17-7.2.28),(7.2.30),(7.2.32),(7.2.34-7.2.35)
                End while
                Let $[y^{(0),*}, z^{(0),*}]$ be the found solution, where $y^{(0),*}$, be the best
                task-team allocation variable and $y_m^*$ is the subvector relevant
                to the team $m$
                Let $x^{(0),*} = \emptyset$ be the initial sequencing variable set
Step 2:         Independent Sequences Assignment
                Set $m = 1$
2.a             If the "TOP TO END" strategy is feasible, add the relevant
                solution $x_m^*$ to $x^{(0),*}$ go to Step 2.d
                While ($t < t_{ord}$ or the optimum is found)
2.b                 Given $y_m^*$, minimize (8.1.2.2) with respect to the
                    constraints (7.2.2-17),(7.2.29), (7.2.31-7.2.33)
                End while
2.c             If a solution $x_m^*$ does not exist, go on Step 3.a, otherwise add
                $x_m^*$ to $x^{(0),*}$ and go to Step 2.d
2.d             If $m = |\mathcal{T}|$, then $x^{(0),*}$ is a feasible initial solution, otherwise
                set $m = m + 1$ and go to Step 2.a
                End for
Step 3: Tabù strategy
3.a             Define a new constraint making the solution $[y^{(0),*}, z^{(0),*}]$
                unfeasible and go to Step 1.a
```

FIGURE 27 INITIALIZATION ALGORITHM

If Algorithm 1 provides a feasible solution, the scheduling plan improvement is obtained by means of two further steps:

Algorithm 2 - Refine the solution:

1. The first step (Fig. 28) considers a single team m independently from the others and looks for a better activity sequence for it. Such a step is similar to the second step of Algorithm 1 but it looks for the best solution instead of a simply feasible one;
2. The second step (Fig. 28) consists in the solution of the main problem starting from the tasks sequence determined in step 1 by varying a subset of sequence variables associated with different teams. This step applies a general matheuristic approach in which the variables to optimize and the fixed variables are randomly chosen.

As regards the specific optimization problems:

- In the first step of Algorithm 1, only the path travelled by the teams and the balancing of the number of activities assigned to them are considered. The relevant cost function is

$$\min\left(\alpha_2\sum_{m=1}^{|\mathcal{T}|}P_m+\alpha_3\sum_{m=1}^{|\mathcal{T}|}\sum_{\substack{g=1,\\ g\neq m}}^{|\mathcal{T}|}\left(\Delta P_{m,g}+\Delta E_{m,g}\right)\right) \tag{8.1.2.1}$$

- In the second step of Algorithm 1 and in the first step of Algorithm 2 only the completion times are optimized by means of the cost function

$$\min\alpha_1\left(\sum_{m=1}^{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{A}|}\omega_i c_{i,m}+\sum_{i=1}^{|\mathcal{A}|}\omega_i q_i\right) \tag{8.1.2.2}$$

- Finally, the second step of Algorithm 2 considers the complete problem in Eq. (7.2.1-7.2.35).

```
Algorithm 2: Solution refinement
Initialization  Let $K$ be the maximum number of refinement iterations
                Let $t_{max}^{Step\,1}$ be the maximum time for the whole solution refinement
                Let $t_{max}^{Step\,2}$ be the maximum time for the matheuristic solution
                refinement

                While $k \leq K$ and $t < t_{max}^{Step\,1}$
Step 1.             For $m = 1, \ldots, |\mathcal{T}|$
                        Let the variables $x_{i,jm}^{(k)}$ free to vary and fix all the
                        others. Then minimize (8.1.2.2) with respect to the
                        constraints (7.2.2-17),(7.2.29),(7.2.31-7.2.33)
                    End for
Step 2.             While $t < t_{max}^{Step\,2}$
                        Randomly choose a subset $\hat{x} \subseteq x^{(k)}$
                        Fix the variables in $\bar{x} = x^{(k)} \backslash \hat{x}$
                        Minimize (7.2.1) with respect to $\hat{x}$, $y$ and $z$
                        This problem is subject to $\bar{x}$ and to the constraints
                        (7.2.2-7.2.35))
                        Let $[x^*, y^*, z^*]$ be the best found integer solution
                        If the solution $[x^* \cup \bar{x}, y^*, z^*]$ is better than the solution
                        $[x^{(k)}, y^{(k)}, z^{(k)}]$
                            Set $x^{(k+1)} = x^* \cup \bar{x}$
                        End if
                        Set $k = k + 1$
                    End While
                End While
```

FIGURE 28 MATHEURISTIC ALGORITHM

8.2 ALGORITHMS PERFORMANCE

8.2.1 MATHEURISTIC FOR RAILWAY LINE MAINTENANCE SCHEDULING

As regards the performance of the matheuristic to solve the railway line maintenance scheduling problem, the applications of the proposed approach have shown reductions of about the 50% of the cost function with respect to the reference initial top-to-end solution.

Regarding the comparison of the matheuristic performance with respect to those provided by generic branch and bound approach implemented by IBM-Ilog Cplex® solver, consider the outputs reported in Fig. 29 for different instance dimensions of the problem.

In particular, in such a figure, the relative values of the cost functions per each solution are depicted being the relevant labels the time required for finding such solutions. It is possible to note that the generic branch and bound approach implemented in Cplex does not find the optimal solution for instances with $N \geq 20$ and even a feasible solution for instances with $N \geq 35$, considering a maximum running time of 48 hours.

Moreover, it is interesting to note that even the lower bound (that is the solution of a relaxed problem) is very hard to find: in fact, for instances with $N \geq 20$ the solver stops before the 48th hour having reached the maximum memory capacity and, with $N \geq 30$ even much before the first hour of computation.

To conclude, it is worth saying that such a computational analysis has been obtained by implementing the matheuristic algorithm in Matlab® and IBM-Ilog Cplex® on a 3.10 GHz PC with 16 GB RAM.

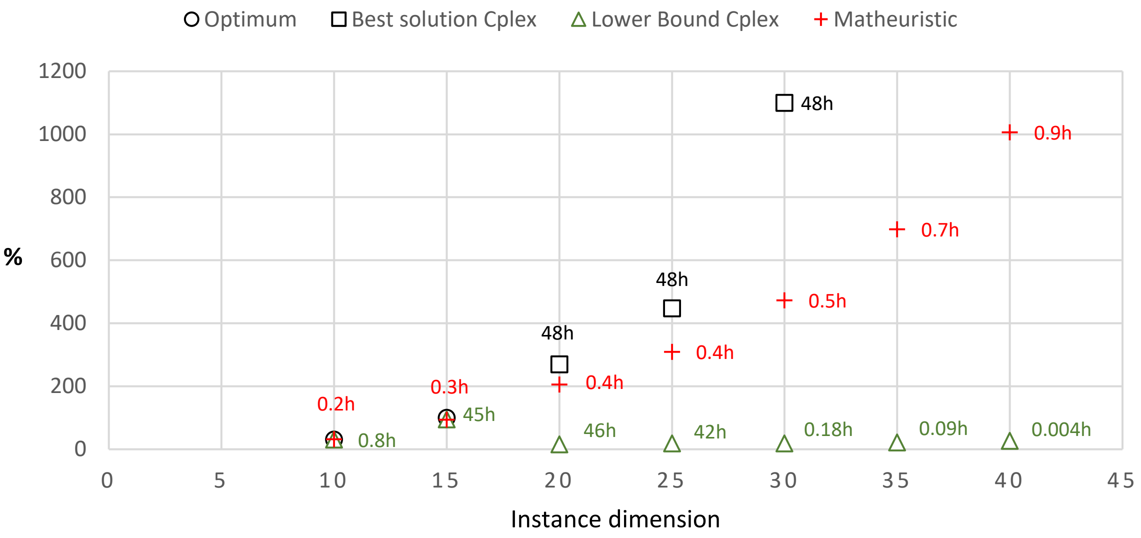

FIGURE 29 CPLEX AND MATHEURISTIC RESULTS COMPARISON WHERE THE REFERENCE POINT (100%) CORRESPONDS TO THE OPTIMUM OF THE INSTANCE WITH DIMENSION OF 15 MAINTENANCE ACTIVITIES.

8.2.2 MATHEURISTIC FOR RAILWAY NETWORK MAINTENANCE SCHEDULING

In this section, the performance of the matheuristic for the scheduling of railway network maintenance are described. As concerns the comparison of the matheuristic performance with those provided by a branch-and-bound approach implemented by the general purpose solver IBM-Ilog Cplex®, consider the outputs reported in Fig. 30 where the cost function values for different instance sizes are reported in a logarithmic scale. Such results have been obtained by implementing the matheuristic algorithm in Matlab® and IBM-Ilog Cplex® on a 3.10 GHz PC with 16 GB RAM.

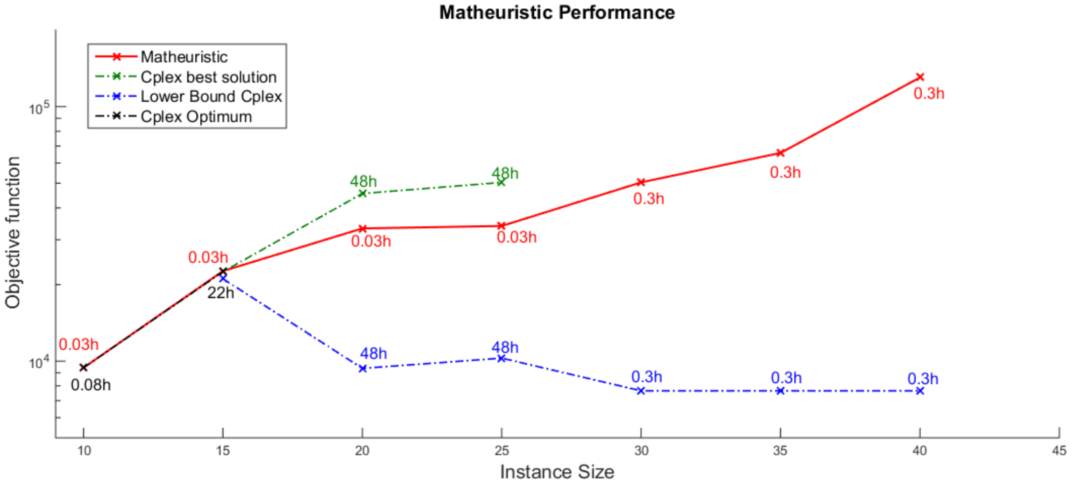

Figure 30 Cplex and matheuristic cost functions comparison. The labels represent the time required for computing the values.

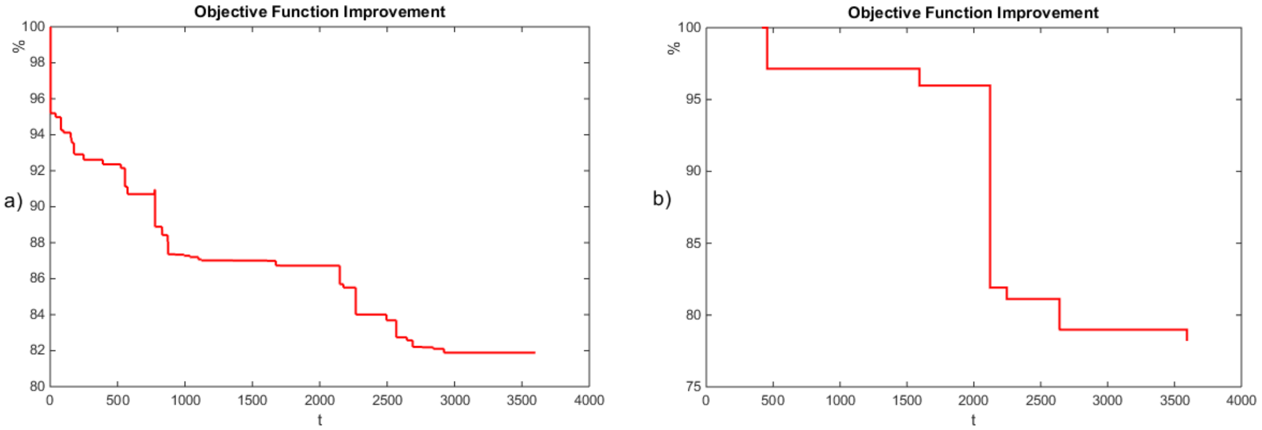

Figure 31 Cost function improvement respect to the initial solution considering 42 maintenance activities a) and 84 maintenance activities b)

In addition, in such a figure, the computational times required by Cplex and by the proposed algorithms to find the solutions or bounds are also indicated. Note that the increasing values of the objective function are due to the increase of the problem size and it does not indicate a worsening of performance. Indeed, the values of the objective function are represented in order to show the gap from the lower bound.

As regards the results, Cplex is able to determine the optimal solution (defined as ‘Cplex optimum’ in the figure), in a short time, for instances that include up to 15 activities, while it is only able to compute feasible solutions (defined as ‘Cplex best solutions’) for instances up to 25 activities. The reported cost function values are the best found after 48 hours, together with the relevant lower bounds. Considering the gap, it is clear that a good lower bound could not be found for big instances of the considered problem. In addition, for instances with more than 25 activities, for which even a feasible solution could not be found by Cplex, the reported values of the lower bound are the best obtained after 0.3 hours, that is, the time needed by the matheuristic to find its best solution.

For graphical reasons, in Fig.30 only the instances until 40 activities are considered even if the matheuristic is able to solve bigger size of the problem, as shown in the following.

For what concerns the matheuristic optimization performance, the applications of the proposed approach have shown, after a computational time of 1 hour, an improvement of about the 20% of the cost function with respect to the reference initial solution provided by Algorithm 1. In Fig.31, the comparison between the initial solution and the final solution found by the matheuristic algorithm is depicted. In looking at this result, it is worth keeping in mind that the values in the graphs consider all the terms in the cost function defined in (7.2.1). A discussion about the risk reduction is specifically reported in Section 9.3.

8.2.3 LOWER BOUND FOR THE CONSIDERED PROBLEM

In this study, maintenance planning is modelled as the scheduling of a set of independent jobs (or maintenance activities) on a set of unrelated parallel machines (or maintenance teams) with sequence dependent setup times.

Although for some simple or restricted instances an optimal solution exists, the problem presented in this thesis is very difficult to solve and have been shown to be NP-hard [38].

Trying to find approximate solutions to big instances of the problem, the above described matheuristic approaches have been introduced.

Therefore, in order to evaluate the efficiency of the proposed approximate matheuristic methods and the goodness of the found solutions, it is essential to have a good lower bound as described in Section 3.6.

In this study, a good lower bound is identified by relaxing the integrity constraints on the sequence variables ($x_{i,j,m} \in \mathbb{R}_{+,0}$) and introducing the following constraints:

$$\sum_{i=1}^{|\mathcal{A}|} w_{i,m,h} \leq N_{max} \qquad \forall h \in \mathcal{H}_{t_k,\mathrm{T}}, \forall m \in \mathcal{T} \qquad (8.2.3.1)$$

Where:

$w_{i,m,h}$	Binary variable equal to 1 if the activity on asset i is performed by the maintenance team m in the time interval h and 0 otherwise
N_{max}	Maximum number of maintenance activities that can be performed in the h^{th} train-free sub-interval in $\mathcal{H}_{t_k,\mathrm{T}}$

The constraints in Eq. 8.2.3.1 state that the number of maintenance activities that can be assigned to the train-free sub-interval h for each maintenance team m is at most equal to N_{max}. The number N_{max} can be easily evaluated considering a bin-packing problem in which each train-free sub-interval h is a bin with fixed size. Each maintenance activity is an item that should be put in one bin and which uses the duration space of the bin.

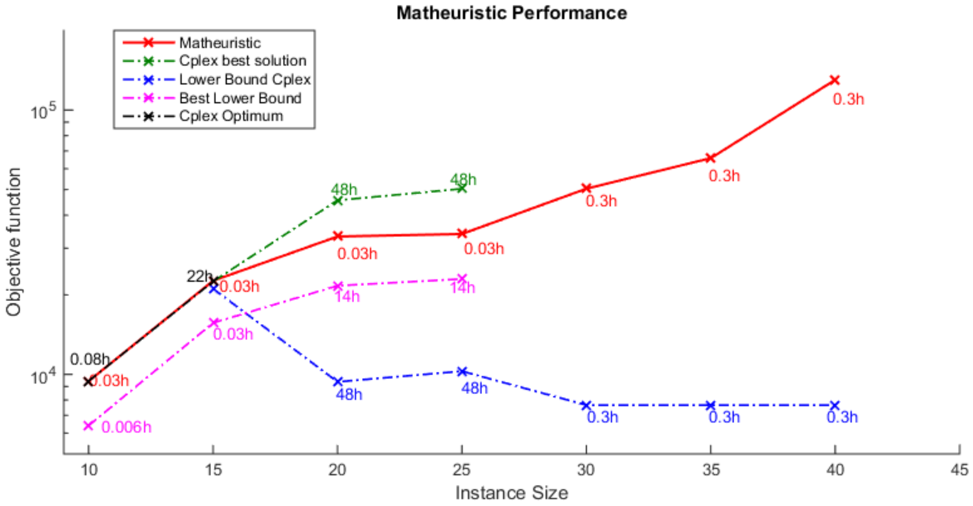

FIGURE 32 CPLEX AND MATHEURISTIC COST FUNCTIONS COMPARISON WITH THE BEST LOWER BOUND

The matheuristic performances for different instances of the scheduling problem for the railway network maintenance are compared with the best solution and the lower bound found by the IBM-Ilog Cplex® solver. Since the lower bound obtained by the common solver relaxing the integer constraints is not efficient, the above described lower bound is used to evaluate the goodness of the approximated solution evaluated through the proposed matheuristic algorithm.

TABLE 7 MATHEURISTIC PERFORMANCE

Instance dimension		**Cplex**		**LB***		**Matheuristic**
Teams	Activities	Time (h)	GAP %	Time	GAP %	Time (h)
2	10	0.08	0%			0.03
2	15	22	0%			0.03
3	20	48	80%	14	55%	0.03
3	25	48	90%	14	60%	0.03
4	30	0.3	97%	-	-	0.3
4	35	0.3	97%	-	-	0.3
4	40	0.3	97%	-	-	0.3
4	60	0.3	97%	-	-	1
6	80	0.3	97%	-	-	1
10	100	0.3	97%	-	-	2

In Tab. 7 the matheuristic performances for different instances of the problem and the gap with respect to the lower bound are reported. Therefore, the proposed approach can be applied also to big instance of the problem. The increase of the maintenance activities, that need to be planned, requires the utilization of a greater number of maintenance teams.

9. CASE STUDIES AND RESULTS

In this chapter, some experimental results are reported, the application of the optimization models to a real rail network is described and the results are discussed to demonstrate the effectiveness of the proposed approach.

9.1 DETERMINISTIC OFF-LINE PLANNING OF RAILWAY LINE MAINTENANCE

The considered case study consists of the tamping maintenance of a railway stretch of 60 km length, subdivided into 30 equal segments of 2 km (Fig. 33). In such a realistic case, it is assumed that 10 segments, randomly generated, require maintenance activities.

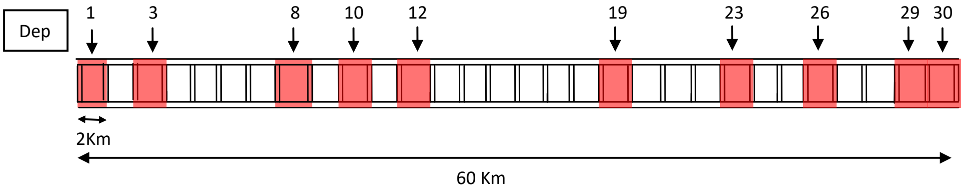

FIGURE 33 FIRST AND SECOND SCENARIO LINE SCHEME.

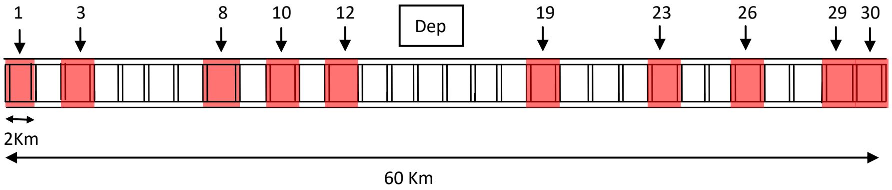

FIGURE 34 THIRD AND FOURTH SCENARIO LINE SCHEME.

In the considered scheduling instance, the random generated segments requiring maintenance are: 1, 3, 8, 10, 12, 19, 23, 26, 29, 30.

These segments are hence assumed to be characterized by non-negligible degradation conditions, and have to be considered in the next planning horizon T of one week. The train-free time intervals start at 0 a.m. and end at 5 a.m.. Moreover, two maintenance teams are available for maintenance activities, each starting its first travel from a depot located along the railway line.

Given these assumptions, four different scenarios are analyzed:

- First scenario: the two maintenance teams have the same working speed of 500 m/h (of tampered railway) and the same travelling speed of 40 km/h (between two rail stretches that have to be maintained), since they are equipped with the same single-sleeper tamper machine. Finally, both the teams start their first travel from a depot at the beginning of the considered railway line (see Fig. 33);

- Second scenario: the two maintenance teams have different working speeds (500 m/h and 1000 m/h, respectively), and different travelling speeds (40 km/h and 60 km/h, respectively). The first team is equipped with a single-sleeper tamper machine, while the second team is equipped with a two-sleeper tamper machine. Finally, both the teams start their first travel from a depot at the beginning of the considered railway line (see Fig. 33);
- Third scenario: this scenario is similar to the first one, but both the working teams start their travel from a depot located in the middle of the considered railway line (see Fig. 34);
- Fourth scenario: this scenario is similar to the second one, but both the working teams start their travel from a depot located in the middle of the considered railway line (see Fig. 34).

As regards the results, the optimal solution of the first scenario is shown in Fig. 35: the two maintenance teams work four nights, each performing five maintenance activities. The completion time is 75.7 hours (in the fourth train-free interval, that is, in four days).

The optimal schedule of the second scenario is shown in Fig. 36.

In this case, the second team, characterized by a higher working speed, performs seven activities in three train-free intervals (equivalent to three days). At the same time, the first team performs its three activities in two train-free intervals. The completion time is 49.7 hours, corresponding to a time reduction of about 34%, with a consequent reduction of the objective function value around 45%.

Similar solutions have been found for the third and fourth scenarios, as shown in Fig. 37 and Fig. 38, respectively. In details, the optimal schedule of the third scenario shows that both teams execute five activities, but the first maintenance team needs four train-free intervals, whereas the second team only needs three train-free intervals, although both the teams have the same working/travelling speeds. This result is due to the central position of the depot, which reduces the initial setup time and let the second team complete two activities within the first train-free interval. The whole completion time is 74.6 hours.


Maintenance activities assigned to Work Team 1

Maintenance activities assigned to Work Team 2

FIGURE 35 FIRST SCENARIO RESULTS.


Maintenance activities assigned to Work Team 1

Maintenance activities assigned to Work Team 2

FIGURE 36 SECOND SCENARIO RESULTS.


Maintenance activities assigned to Work Team 1

Maintenance activities assigned to Work Team 2

FIGURE 37 THIRD SCENARIO RESULTS.

FIGURE 38 FOURTH SCENARIO RESULTS.

Finally, the optimal schedule of the fourth scenario (Fig. 38) shows that the second team, characterized by a higher working speed, performs seven activities in three train-free intervals, while the first team performs three jobs in two train-free intervals, similarly to the first case. The completion time is 49.5 hours.

The obtained results show that the substitution of a maintenance team with a faster one produces a significant reduction of the total completion time, whereas the depot position has a negligible influence on the scheduling in this particular case.

Such results show the capability of the model to perform optimization for different scenarios, thus allowing a complete cost/benefit analysis of the maintenance scheduling problem. This allows, for instance, to prove that it is possible to place the depot in different positions without affecting the maintenance performance, thus allowing to choose the cheapest one (for instance from the building point of view).

9.2 DETERMINISTIC ON-LINE PLANNING OF RAILWAY LINE MAINTENANCE

This case study consists of the *tamping maintenance* of a railway stretch subdivided into equal segments of 2 km each, as shown in Fig.39. Moreover, it is assumed that two maintenance work teams are able to perform activities on all the rail stretches without limitations.

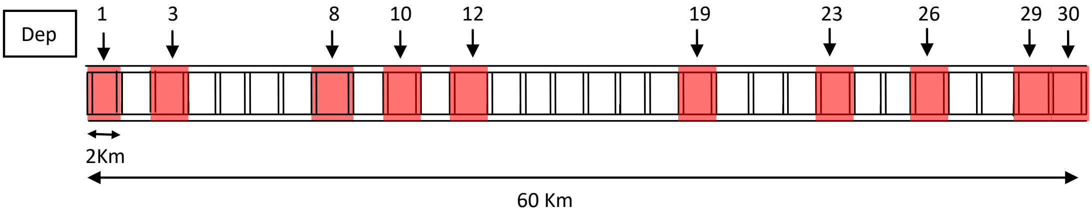

FIGURE 39 RAILWAY LINE SCHEME (60 KM DIVIDED INTO 30 STRETCHES).

The aim of the considered case study is to stress about the capability of the model of recovering when something does not go as planned, or when additional maintenance activities are considered, respectively, by applying the Rolling Horizon (RH) approach described in Section 6.3.

In particular, three scenarios are analyzed:

- the maintenance scheduling of the subset $R = \{1, 3, 8, 10, 12, 19, 23, 26, 29, 30\}$ of rail stretches, highlighted in Fig. 39. For this case, two instances of the optimization problem are considered, being the first due to a work delay *(scenario A)* and the second due to an unexpected rail stretch breakdown *(scenario B)*;
- the third scenario *(scenario C)* consists of an extension of the railway line depicted in Fig. 39 characterized by a set of 100 assets requiring maintenance activity. In this case, 16 instances of the RH approach are considered.

According to the approach described in Chapter 6.2.1, the soft and hard deadlines are evaluated considering the data presented by Famurewa *et al.* in [128]. In particular, it is assumed that the parameter $\delta_i(t, t_k)$ consists of a NGSP modelling the deformation of rail vertical geometry, with constant expectation $\bar{\delta}_i(t, t_k) = \bar{\delta}_i$ and time-varying variance $\sigma_i^2(t, t_k)$. As regards the critical value, it is evaluated considering $\left|\bar{\delta}_i - \delta^{cr}\right| = 10$ mm.

In both the cases, the optimal solution of the optimization problem is obtained via the matheuristic approach described in Section 8.2.1. As regards the model parameters, the horizon Θ is set to 96 hours whereas each train-free interval starts at midnight and lasts 5 hours. As regards the maintenance performances, it is assumed that the teams working speed is 1 km of tampered railway per hour (2 hours per rail stretch), whereas the travelling speed between two rail stretches that have to be maintained is 40 km/h.

Finally, each run of the matheuristic approach lasts 1 hour, that is $\Delta T = 1$ h.

SCENARIO A – WORKING DELAY

As said, the first scenario consists of the scheduling and re-scheduling of the set R. With only 10 maintenance activities to consider, the problem in Chapter 7.1 is able to find the best solution that assigns the activities $M_1 = \{3,10; 19; 12; 8\}$ to the first team and $M_2 = \{30; 23,26; 29; 1\}$ to the second team. In such sets, the maintenance activities assigned to different train-free intervals are separated by semicolon, and the relevant representation is depicted in Fig. 40, where the gray boxes represent the interval in which trains circulate, and it is not possible to perform the maintenance activities.

With reference to the RH horizon framework, the dynamic evolves as follows:

1. At t_k the maintenance work teams start their maintenance activities and, at the end of the first train-free interval, the maintenance activity 30 is finished according to the schedule, while a delay of the activity 3 makes the activity 10 impossible to finish in time. Therefore, such a maintenance has to be reconsidered in the problem stated in $t_{k+1} - \Delta T$ for the interval $(t_{k+1}, t_{k+1} + \Theta)$ beginning with the next day;
2. The new schedule to be applied in t_{k+1} results to be $M_1 = \{\mathbf{3};\ 10, 12; 26; 8\}$ for the first team and $M_2 = \{\mathbf{30}; 19; 23,29; 1\}$ for the second one, where the bold entries indicate the already executed maintenance activities.

Looking at the solutions depicted in Fig. 40, it is possible to note that:

- in the second train-free interval some activities previously assigned to a team are then assigned to the other one (19 is assigned first to team M_1 and then to M_2, whereas 26 is assigned first to team M_2 and then to M_1);
- some activities are anticipated, other delayed (12 is anticipated from the third train-free interval to the second one, whereas 23 and 26 are delayed from the second train-free interval to the third one);
- despite some modifications, the activities scheduler is able to keep all the maintenance activities within the first four train-free intervals.

SCENARIO B – SUDDEN FAULT

At $t_{k+2} - \Delta T$, when the optimization problem instance for the interval (t_{k+2}, t_{k+2} +Θ) is stated, an unexpected breakdown of the previously unscheduled rail stretch 25 is detected. Therefore, it has to be considered in the problem stated in $t_{k+2} - \Delta T$, which provides the solution depicted in Fig. 41. With reference to the RH horizon framework, the dynamic evolves as follows:

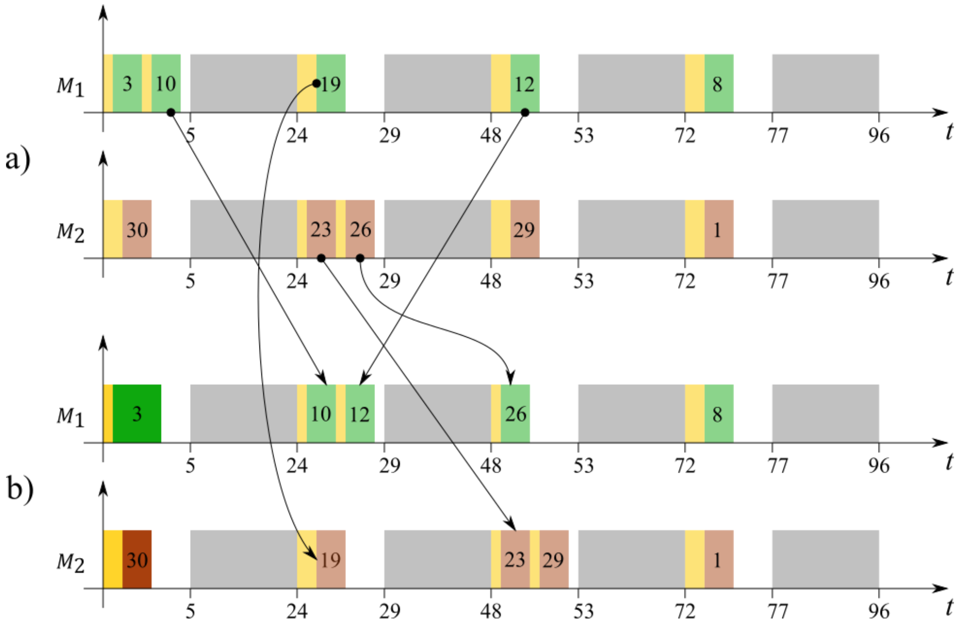

FIGURE 40 ACTIVITY SCHEDULING FOR t_k (A) AND t_{k+1} (B).

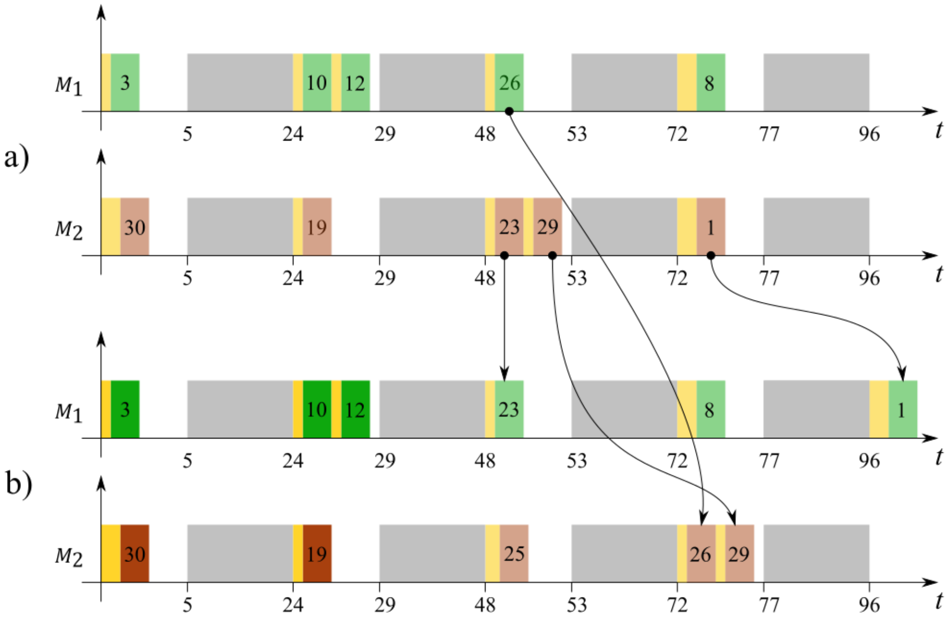

FIGURE 41 ACTIVITY SCHEDULING FOR t_k AND t_{k+1} (A) AND t_{k+2} (B).

1. At $t_{k+2} - \Delta T$ a fault of the rail stretch 25 is detected;
2. Such a maintenance activity is scheduled as the first activity in the next train-free interval due to its high priority;
3. The new schedule results to be $M_1 = \{\mathbf{3}; \mathbf{10}, \mathbf{12}; 23; 8; 1\}$ for the first team and $M_2 = \{\mathbf{30}; \mathbf{19}; 25; 26; 29\}$ for the second one, where the bold entries indicate the already executed maintenance activities.

Looking at the solution depicted in Fig. 41 it is possible to note that:

- the new unpredicted activity (asset 25) is scheduled as first activity of the team M_2 in the third train-free period;
- the maintenance of the stretch 26 is assigned first to team M_1 and then to M_2, whereas the activities on the stretches 23 and 1, previously assigned to team M_2, are then assigned to M_1.
- the maintenance of the stretches 26, 29, and 1 are delayed from the third train-free interval to the fourth one.

SCENARIO C– BIG INSTANCE

In this scenario, the RH approach has been applied to a bigger instance of the problem, considering a set of 100 assets, although in each RH window only 30 maintenance activities are considered with the aim of keeping the computational effort limited. As regards the results, consider the cost function values computed for each iteration of the matheuristic algorithm and for all the RH windows depicted in Fig. 42. In such a figure, each peak represents the transition between two consecutive RH windows $(t_k, t_k + \Theta)$ and $(t_{k+1}, t_{k+1} + \Theta)$. Therefore, it is possible to note that, due to the new introduced rail stretches to be scheduled and to the updated problem parameters, the first value of the cost function in each window results to be much greater than the optimal value computed in the previous window.

As regards the solution, it is interesting to look at the evolution of the maintenance activities sequence reported, for each RH window, in Table 8, where the grey cells indicate, for each RH window, the activities already executed. In such a table, it is possible to note that the introduction of new activities makes the sequence of planned but not executed activities change.

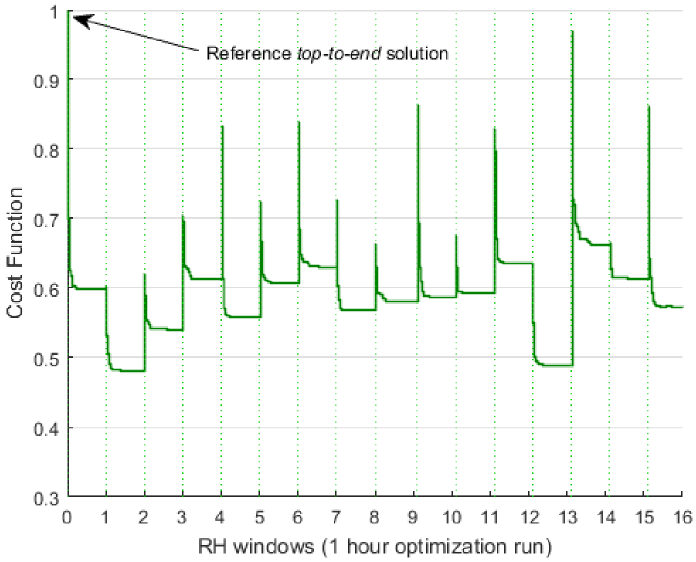

FIGURE 42 COST FUNCTION OF EQ. (4) SHAPE IN THE RH FRAMEWORK

TABLE 8 SCHEDULED MAINTENANCE ACTIVITIES PER EACH MACHINE AND TRAIN FREE INTERVAL; IN LIGHT GREY THE ACTIVITIES ACTUALLY EXECUTED AT THE RH STEP $k-1$, AND IN DARK GREY THE INITIAL TOP-TO-END SOLUTION

k	Team	Train free interval																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	M_1	1 2	3 4	5 6	7 8	9 10	11 12	13 14	15															
	M_2	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30															
1	M_1	11 12	10 16	6 14	4 24	8 2	9 5	22 3	1															
	M_2	18 17	21 19	7 28	27 30	26 13	23 15	20 29	25															
2	M_1	11 12	10 16	28 6	14 4	8 2	9 13	31 20	5 3	1														
	M_2	18 17	32 21	19 7	27 24	30 26	23 33	15 22	34 29	25														
3	M_1	11 12	10 16	13 35	14 23	6 31	4 8	2 15	9 26	33 3	29													
	M_2	18 17	32 21	38 22	25 20	7 28	34 30	27 19	37 36	5 24	1													
4	M_1	11 12	10 16	13 35	14 25	15 28	36 6	34 8	2 19	41 39	9 3													
	M_2	18 17	32 21	38 22	24 26	42 23	7 40	20 31	4 30	5 29	27 37	1 33												
5	M_1	11 12	10 16	13 35	14 25	15 45	19 43	40 6	39 46	8 2	33 9	41 3	36											
	M_2	18 17	32 21	38 22	24 26	27 30	28 7	23 44	4 34	42 37	20 31	5 29	1											
6	M_1	11 12	10 16	13 35	14 25	15 45	19 47	44 6	50 43	40 8	42 2	9 28	49 3											
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	20 33	7 29	4 46	39 23	5 37	34 48	1 41										
7	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	48 6	29 51	43 8	46 2	54 9	49 3	1									
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	23 37	7 33	4 50	42 28	5 41	52 39	44									
8	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	52 6	33 51	48 58	50 8	2 9	57 3	1								
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	28 41	37 56	7 54	4 29	46 44	5 42	49								
9	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	48 6	56 59	37 8	2 33	9 46	57 62	1							

	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	44 41	7 58	4 51	54 60	50 49	5 3	52							
10	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	33 51	63 65	44 6	41 59	8 2	9 37	50 3	1						
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	48 46	64 49	60 7	4 56	62 58	54 66	52 5	57						
11	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	33 51	37 56	41 69	49 6	63 4	2 44	9 67	57 3	1					
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	48 46	50 52	70 68	64 7	66 59	62 8	58 54	5 65	60					
12	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	33 51	37 56	60 73	44 71	54 6	67 4	2 9	63 49	62 3	1				
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	48 46	50 52	41 57	58 68	7 72	70	64 66	8 74	59 5	69 65				
13	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	33 51	37 56	60 73	75 65	49 6	7 4	70 8	2 54	58 64	77 3	69			
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	48 46	50 52	41 57	44 62	63 59	72 71	74 76	68 9	67 66	78 5	1			
14	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	33 51	37 56	60 73	75 65	68 80	54 66	7 6	59 4	74 8	2 67	77 3	72		
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	48 46	50 52	41 57	44 62	49 64	81 63	82 76	78 71	70 58	9 69	79 5	1		
15	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	33 51	37 56	60 73	75 65	68 80	67 84	58 69	7 6	63 4	78 8	2 70	81 3	77	
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	48 46	50 52	41 57	44 62	49 64	54 85	71 83	66 79	82 76	74 59	9 86	72 5	1	
16	M_1	11 12	10 16	13 35	14 25	15 45	19 47	20 53	23 55	28 61	33 51	37 56	60 73	75 65	68 80	67 84	71 70	59 72	88 6	7 66	4 8	9 77	89 3	81
	M_2	18 17	32 21	38 22	24 26	27 30	31 36	34 40	39 43	42 29	48 46	50 52	41 57	44 62	49 64	54 85	87 58	76 90	69 83	86 79	82 2	78 63	5 74	1

Finally, it is worth noting that, the performance of the matheuristic approach is compared with the basic top-to-end solution and results into an about 50% improvement of the relevant cost function values. It is worth recalling that such a basic top-to-end solution is considered only in the first RH window, as also indicated in Fig. 42, while the initial solution in the each other windows is the best of the previous one.

9.3 DETERMINISTIC PLANNING OF RAILWAY NETWORK MAINTENANCE

In this section, the problem described in Chapter 7.2 is applied to two real world case studies and the relevant results are discussed. To this aim, the North-Western Italian rail network is considered. It can be represented by means of a graph with 14 nodes and 21 links, as depicted in Fig. 43.

As regards the problem size in the two cases, the maintenance activities of 42 and 84 technological components, respectively, are considered to be scheduled. The relevant risk model has been obtained by identifying suitable Weibull distributions for all the asset kinds, whose parameters are reported in Tab.9.

TABLE 9 FAULT RATE OF RAILWAY TECHNOLOGICAL ASSETS: WEIBULL DISTRIBUTIONS AFTER THE FIRST INSTALLATION

Asset	**Number of considered assets Case Study 1**	**Number of considered assets Case Study 2**	$\boldsymbol{\lambda(t_0^i)}$ **[days]**	$\boldsymbol{\eta(t_0^i)}$
Radio Block Centre	3	6	1690	3.2
Interlocking	3	6	1120	4.5
Balise	15	30	1378	5.6
Track circuits	10	20	1115	3.6
Axle bearings temperature detection systems	5	10	1251	3.8
Switch & Crossings (IT subsystem)	6	12	1459	4.1

By applying the model described in Chapter 6, the soft and hard deadlines have been calculated for each asset at time $\tau = t_0^i + 365$ days, that is one year after their installation, taking into account a maximum tolerable value of risk of $0.5\ 10^{-4}$. Since the values of the expected losses caused by the failures are sensitive data, only a fictitious risk threshold is reported here.

As regards the parameters for the model in Chapter 7.2, it is assumed that four and six maintenance teams m_i are available in the first and second case study, respectively. Moreover, train-free sub-intervals that starts at each midnight and lasts 5 hours are taking into account. In addition, since a weekly schedule of their activities is required, the horizon T is set to 168 hours so as to determine a short-term operational planning. Finally, the travelling speed of the maintenance teams and the processing time of the maintenance activities are given by the available data.

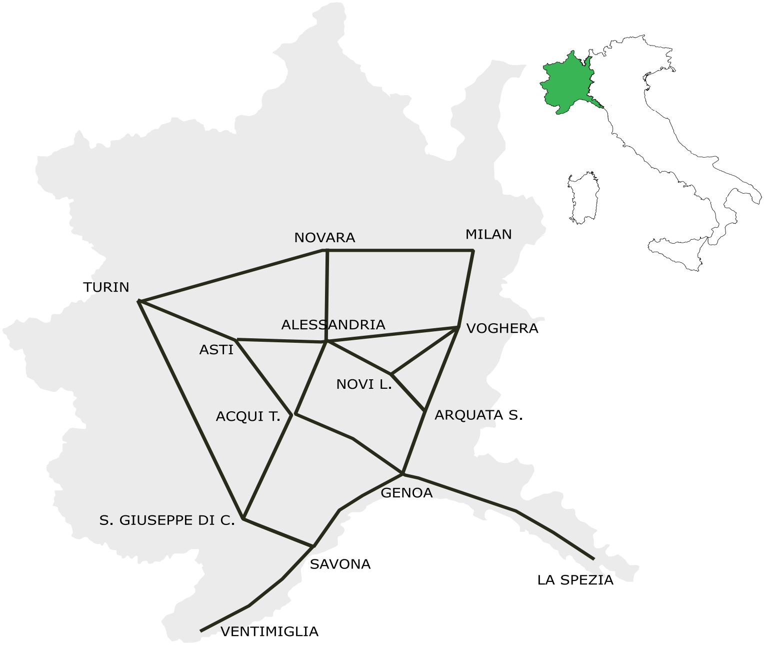

Figure 43 Railway Network

The optimal solution of the optimization problem is obtained via the *matheuristic* approach, described in Section 8.1.2, by considering the following parameter values: $t_{ass} = 0.05$ h, $t_{ord} = 0.03$ h, $K = 50$, $t_{max}^{Step\ 1} = 1$ h, and $t_{max}^{Step\ 2}$= 0.2 h. Note that K is the number of iterations of the total cycle, while the number of iterations of the different steps are into the order of hundreds. In any case, for all the instances, the algorithms stopped before reaching the threshold parameter values since no further improvement was found.

The results depicted in Fig. 44a and Fig. 44b show the links assigned to maintenance teams in the initial solution and final solution. In such figures, it is worth noting that some teams cross the same link. In particular, in Fig. 44a team 2 and team 4 cross the same links S. Giuseppe di Cairo – Savona and Arquata S. – Novi L., whereas in Fig. 44b:

- team 1 and team 4 cross the links Turin – Novara, Novara – Milan, Milan – Voghera;
- team 1 and team 2 cross the link Turin – Asti;
- team 2 and team 3 cross the link Alessandria – Novara;
- team 2 and 4 cross the link S. Giuseppe di Cairo – Savona.

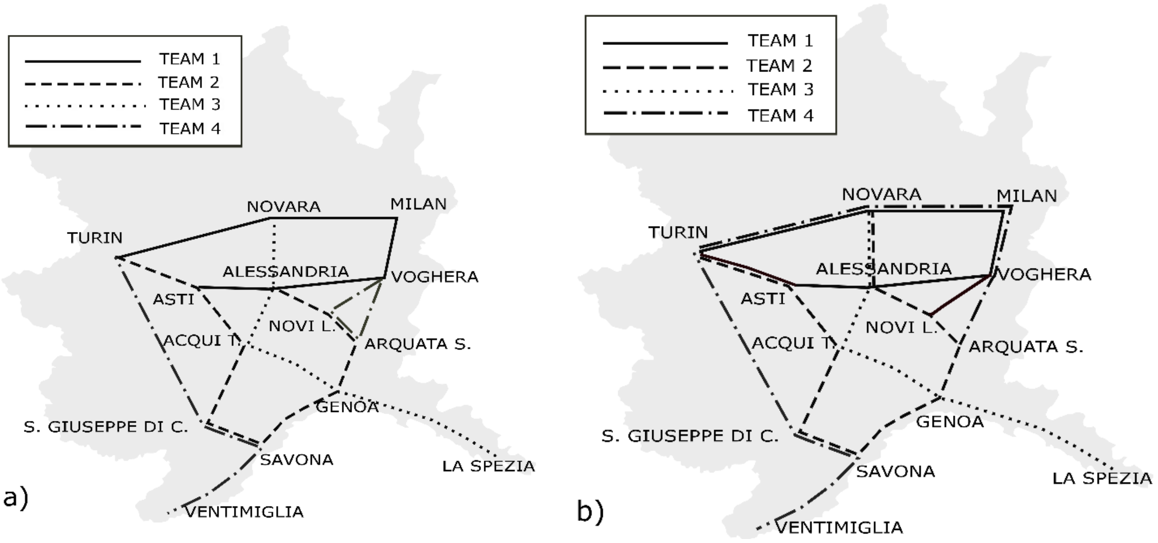

FIGURE 44 LINKS ASSIGNMENT ACCORDING TO THE INITIAL (A) AND FINAL (B) SOLUTIONS

This is allowed by the constraints (7.2.21) where the parameter $|\mathcal{T}|$ is set to 4. In this way, some assets of the same link could be assigned to different maintenance teams in order to satisfy the deadlines and reduce the total completion time. It is interesting to note that the total path in the final solution is greater than the initial solution one (with an increase in length of 16%) since, in order to execute the maintenance activities in a shorter total time, the maintenance teams sometimes need to cross the same links and travel for longer distances.

The results depicted in Fig. 45a and Fig. 45b show the activities assignment to maintenance teams according to the initial and final solution. It is to be noted that in Fig. 45a and in Fig. 45b:

- each numbered rectangle represents the duration of a maintenance activity;
- a different colour is assigned to each team;
- the orange rectangles between each pair of numbered rectangles represent the set-up times;
- the grey rectangles represent the time period in which maintenance is forbidden due to the trains circulation.

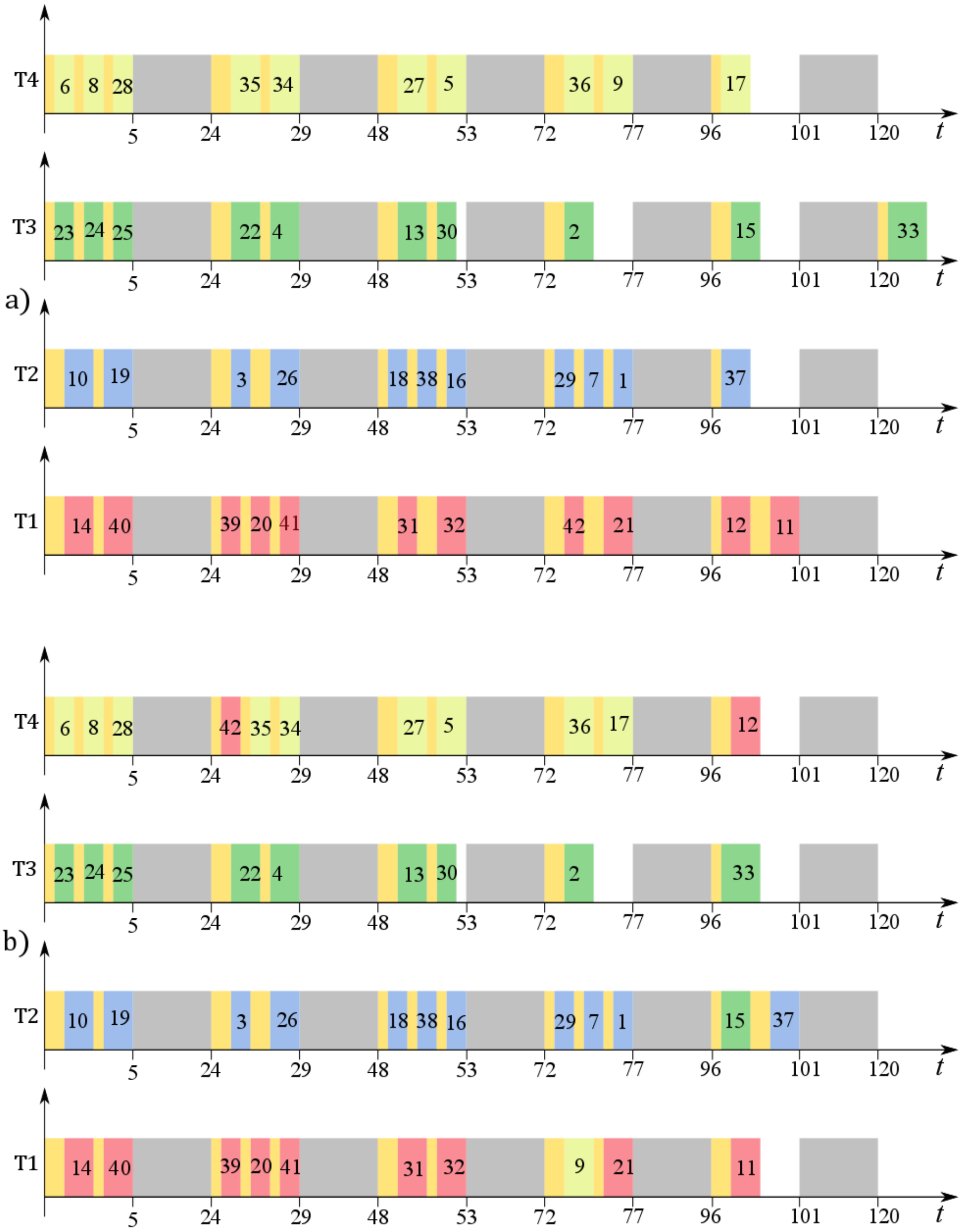

FIGURE 45 MAINTENANCE ACTIVITIES ASSIGNMENT ACCORDING TO THE INITIAL (A) AND FINAL (B) SOLUTIONS

The comparison between the Fig.45a and Fig.45b shows a numerical reduction of the total completion time of 20%. Graphically, such a result is represented by the fact that in the final solution all the jobs are executed within the fifth train-free sub-interval.

More in detail, in the final solution:

- maintenance activities 12 and 42 are assigned to team 4 instead of team 1;
- maintenance activity 9 is assigned to team 1 instead of team 4;
- maintenance activity 15 is assigned to team 2 instead of team 3;
- team 3 finishes its activities in the train free interval 5, instead of interval 6 with a correspondent completion time reduction of 19%;
- team 1 finishes in the same time interval but early, with a completion time reduction of 2%;
- teams 2 and 4 finish in the same time interval with a not relevant increase of their completion times (0.9%);

Regarding the difference of maintenance teams workloads, it is possible to note that:

- the maximum difference between the completion time of the maintenance teams is 25% in the initial solution and 2% in the final one;
- the sum of the differences of the number of maintenance activities assigned to the maintenance teams and the differences of their path lengths provided by the term $\sum_{m=1}^{|\mathcal{T}|}\sum_{\substack{g=1,\\ g\neq m}}^{|\mathcal{T}|}\left(\Delta P_{m,g}+\Delta \mathrm{E}_{m,g}\right)$, shows a reduction of 80%.

Focusing only on the risk reduction achievable with the proposed approach, the following scenarios are considered:

- Scenario 1: all the maintenance activities are completed in correspondence of the hard deadlines.
- Scenario 2: all the maintenance activities are completed in correspondence of the soft deadlines.
- Scenario 3: the maintenance activities are executed according to the proposed initial schedule depicted in Fig.45a.

- Scenario 4: the maintenance activities are executed according to the proposed final schedule depicted in Fig.45b.

In the following discussion, the losses generated are neglected since they are constant. Therefore, keeping in mind the risk approach defined in Chapter 6.2:

- in Scenario 1 each asset has a risk proportional to the probability thresholds $\bar{\varphi}_i$;
- in Scenario 2 the risk for each asset would be proportional to $\alpha_s\bar{\varphi}_i$;
- finally, in Scenario 3 and 4 the risk depends on the specific completion times in the initial and best scheduling solution, respectively.

The risk in Scenario 4 turns out to be -90% comparing to Scenario 1, -80% comparing to Scenario 2, and -25% comparing to Scenario 3. Nevertheless, it is worth underlying that Scenario 1 and Scenario 2 schedules may be non-feasible solutions.

Regarding the weights α_j in the objective function, it is worth giving some indications about how the solution varies, considering, predominantly, only some of the objective function terms.

In Tab. 10, the influence of weights choice is shown, considering, as reference solution, the case in which all the weights have a high value.

In particular, a high value of α_1 and low values of α_2, α_3 determines long distance (+49% respect to the reference solution) covered by the maintenance teams and a not equally distributed work effort between the maintenance teams (+564% of difference in the work load) with a correspondent negligible reduction of the completion time respect to the reference solution. Fig.46 shows the longer paths covered in this scenario by the maintenance teams and the assignation of the same link to a higher number of maintenance teams (some links are assigned to all the available teams). Fig.47, compared to Fig. 45, shows that maintenance team $T4$ finishes its activities in trains-free interval 4, instead of interval 5, due to the not equally distributed workload.

In contrary, low value of α_1 and high value of α_2, α_3 determine higher completion times and a higher number of activities that exceed the soft deadlines (14% of the total number of the considered activities), but a more efficient utilization of maintenance teams. Indeed, the reduction of the covered distances is about -40% and a more equally distributed workload between the maintenance teams is shown (-29% of work load difference).

TABLE 10 INFLUENCE OF WEIGHTS CHOICE ON OBJECTIVE FUNCTION VALUE (DIFFERENCE RESPECT TO THE REFERENCE SOLUTION)

α_1	α_2	α_3	$\sum_{m=1}^{\lvert\mathcal{T}\rvert}\sum_{i=1}^{\lvert\mathcal{A}\rvert}\omega_i c_{i,m}+\sum_{i=1}^{\lvert\mathcal{A}\rvert}\omega_i q_i$	$\sum_{m=1}^{\lvert\mathcal{T}\rvert}P_m$	$\sum_{m=1}^{\lvert\mathcal{T}\rvert}\sum_{g=1,\, g\neq m}^{\lvert\mathcal{T}\rvert}(\Delta P_{m,g}+\Delta E_{m,g})$	Tardy activities respect to soft deadlines
High	High	High	0%	0%	0%	0%
High	Low	Low	-2%	+49%	+564%	0%
High	High	Low	+2%	-28%	+1079%	0%
High	Low	High	+1%	+55%	-7%	0%
Low	High	Low	+2332%	-42%	+1711%	38%
Low	Low	High	+2075%	+60%	-86%	33%
Low	High	High	+2171%	-40%	-29%	14%

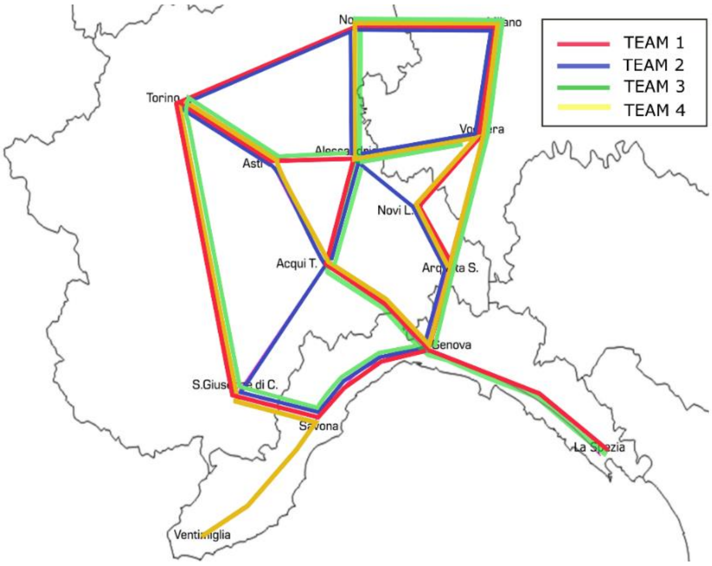

FIGURE 46 LINKS ASSIGNMENT IN CASE OF HIGH VALUE OF α_1 AND LOW VALUES OF α_2 AND α_3

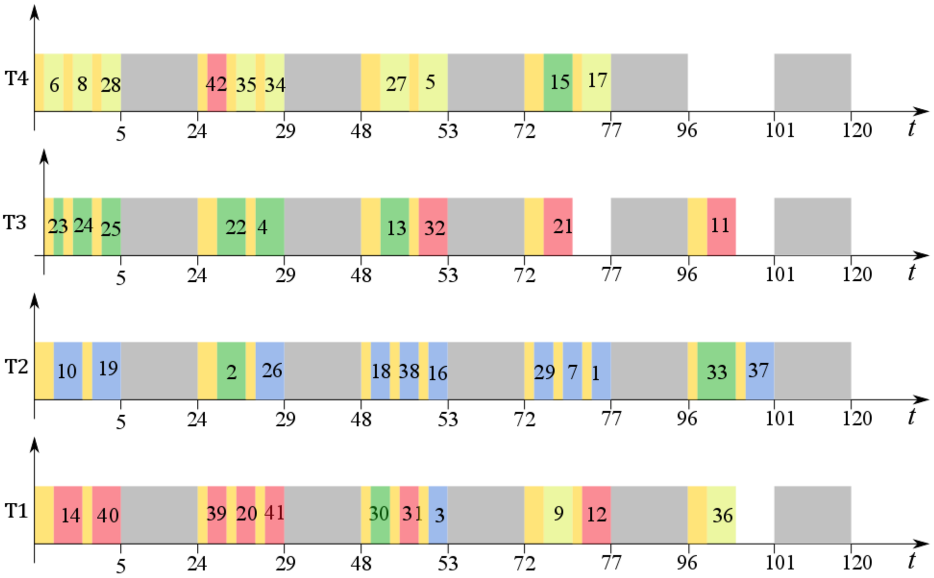

FIGURE 47 MAINTENANCE ACTIVITIES ASSIGNMENT AND SEQUENCE FOR HIGH VALUE OF α_1 AND LOW VALUES OF α_2 AND α_3

Therefore, the weights can be chosen in order to make the terms in objective function comparable or in order to give some priorities according to the organization’s goals.

Note that, as already mentioned, the choice of the weights cannot affect the safety of the system, since the hard deadline will never be exceeded, that is, a tolerable risk level is always guaranteed.

The second case study consists of the scheduling of 84 maintenance activities whose Weibull parameters are depicted in Tab.7. In analysing this case study, the initial and final schedules are shown in Fig.48, while the links assignment is reported in Fig.49.

Comparing to the previous case study, it is worth mentioning that the number of maintenance teams should be increased from four to six in order to find feasible solutions, whereas the algorithms parameters are the same. The results show an average reduction of the completion time of the maintenance teams of around 11% and a reduction of risk of 28% respect to the initial solution. In this case, the comparison with the fictitious schedules of scenario 1 and scenario 2 shows a risk reduction of 83% and 66%, respectively.

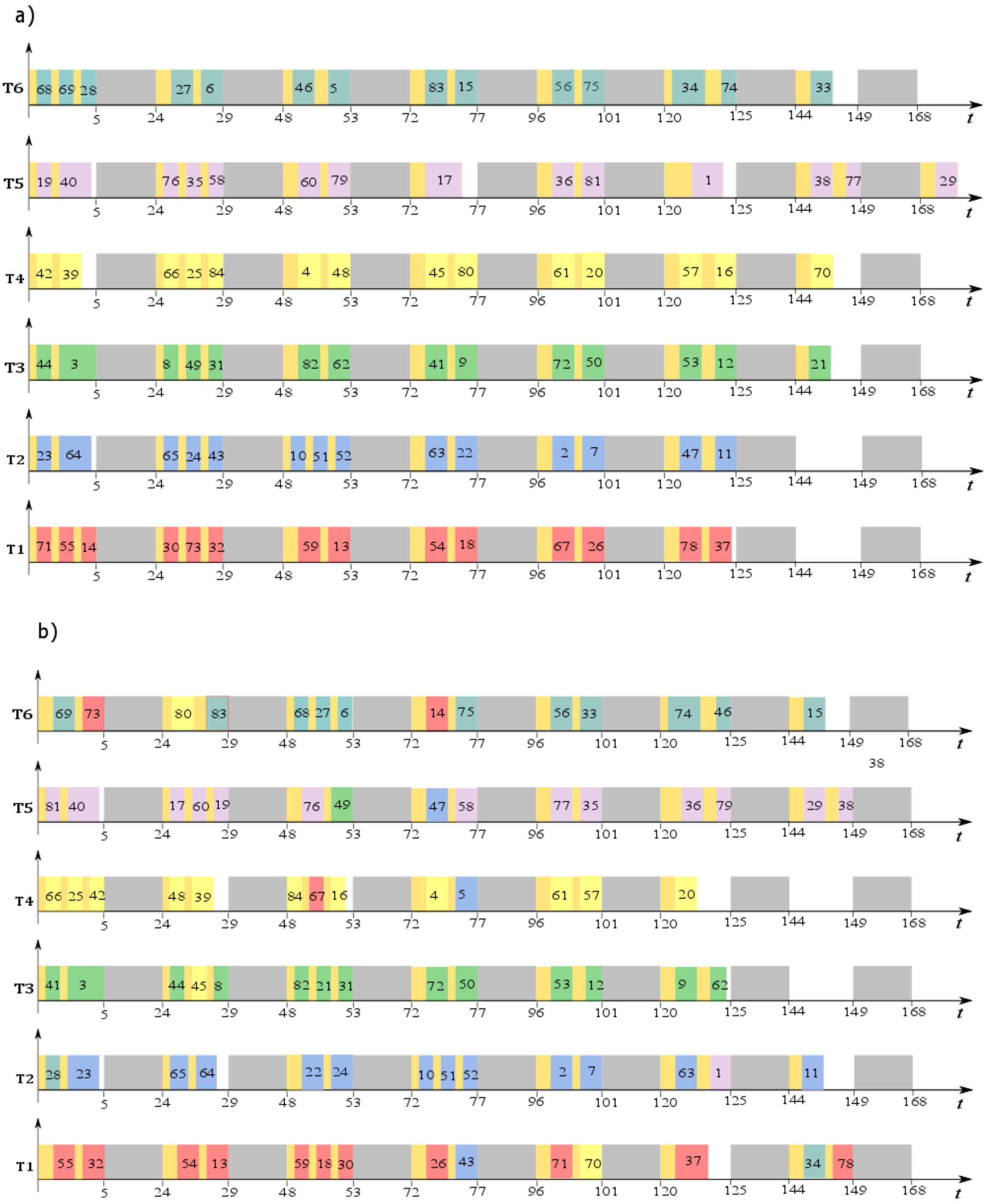

FIGURE 48 MAINTENANCE ACTIVITIES ASSIGNMENT AND SEQUENCE FOR THE INSTANCE 84 ACTIVITIES AND 6 TEAMS ACCORDING TO THE INITIAL (A) AND FINAL (B) SOLUTIONS.

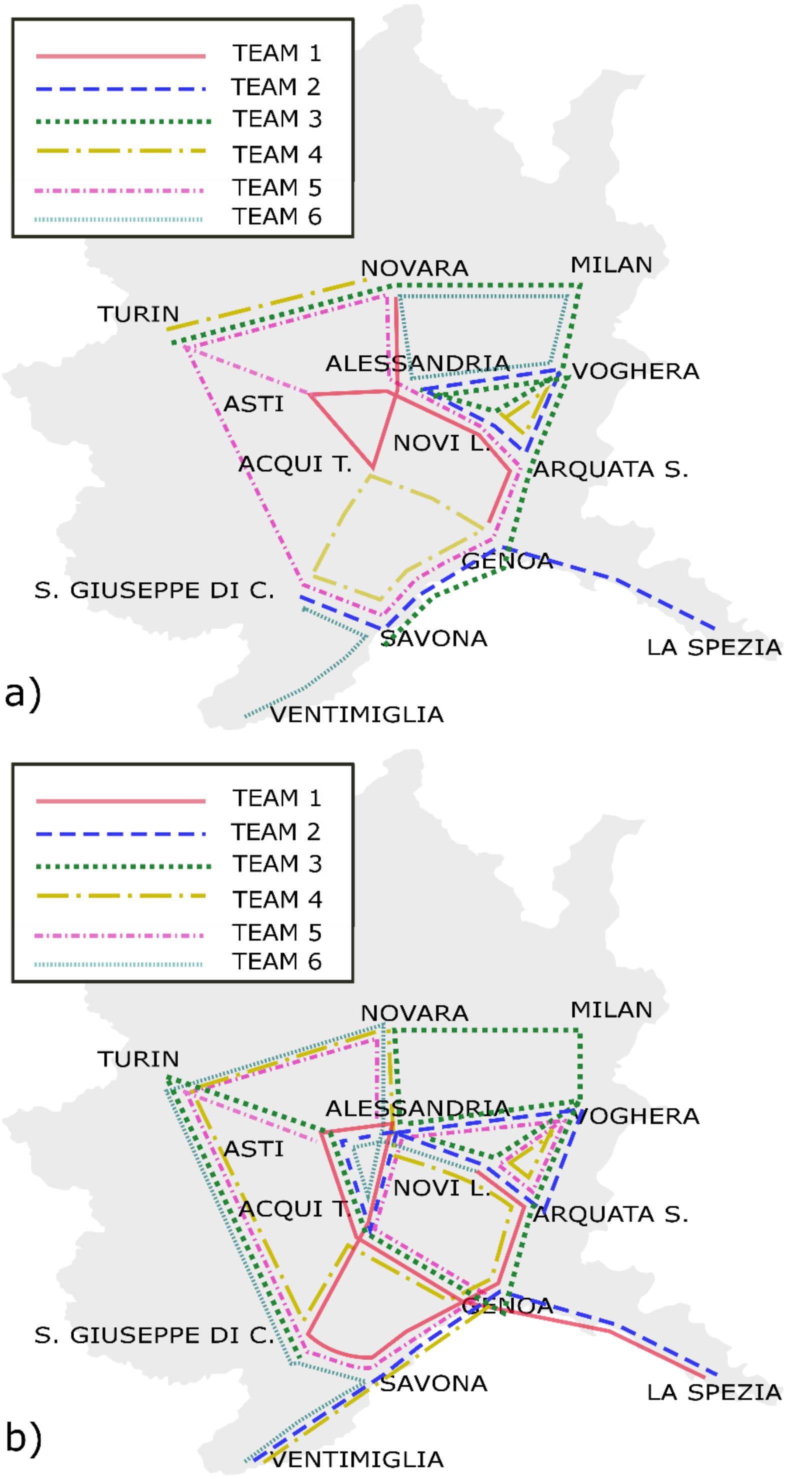

FIGURE 49 LINKS ASSIGNMENT ACCORDING TO THE INITIAL (A) AND FINAL (B) SOLUTIONS

Summarizing, the presented case studies have shown that the proposed modelling and optimization approaches:

- improve the reliability and availability of railway transport in term of reduction of failure risk and of the relevant cost;
- improve railway asset management efficiency through the optimal utilization of the available time for maintenance, with a reduction of the difference between maintenance teams completion times.
- optimize the resources utilization avoiding unnecessary trips of the maintenance teams, balancing the workloads, etc. In particular, the sum of the differences between the number of activities assigned to the maintenance teams and the differences between their path lengths is minimized.

9.4 STOCHASTIC PLANNING OF RAILWAY NETWORK MAINTENANCE

In this section, the stochastic problem described in section 8.3 is applied to a realistic case study and the relevant results are discussed.

To this aim, a rail network that can be represented by means of a graph with 4 nodes and 5 links, depicted in Fig. 50, is considered.

It is supposed that the predictive maintenance of 10 critical assets of this rail network need to be planned.

It is also assumed that two maintenance teams M1, M2 are available for maintenance, each starting its first travel from the depot situated in node 1. The travelling speed between two rail assets that have to be maintained is 40 km/h.

The hard deadline dh_i is represented by means of a uniform variable with expected value $\overline{dh_i}$.

Analogously, the soft deadline ds_i is expressed by means of a uniform variable with expected value $\overline{ds_i}$.

Therefore, for each asset a sample of k hard and k soft deadlines is generated.

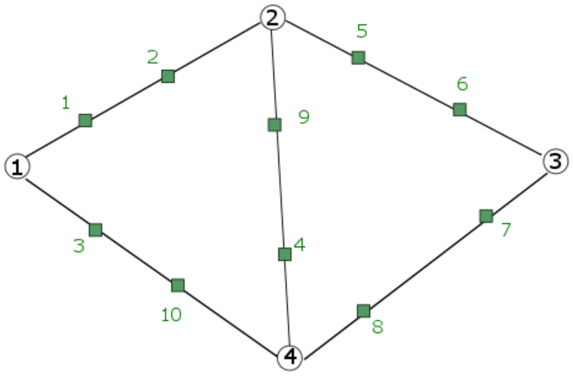

FIGURE 50 GRAPH OF RAIL NETWORK

Two scenarios are analyzed:

1. In the first scenario, the dimension of the considered sample $k = 30$ is fixed, whereas three different standard deviation values (1%, 5%, 10%) are considered.
2. In the second scenario the standard deviation is fixed (5%), whereas the dimension of the sample varies.

The stochastic problem has been implemented on a 2.5 GHz PC with 8 GB RAM, using IBM-Ilog Cplex®. In this section, the results obtained after a running time of ten minutes are described.

In particular, the results depicted in Fig. 51, 52, 53 show the links and activities assignation to maintenance teams M1 and M2 and the path of each maintenance team in scenario 1.

The results depicted in Fig. 52, 54, 55, show the links and activities assignation to maintenance teams M1 and M2 and the path of each maintenance team in scenario 2.

Instead, for the deterministic problem, it has been possible to find the optimal solution that is depicted in Fig.56.

Therefore, the results show that the sample dimension and the sample standard deviation influence the problem solution, demonstrating the convenience of applying stochastic planning approach. Works are in progress to apply the stochastic approach to a bigger instance of the problem and to consider the stochastic nature of other input data in the model formulation.

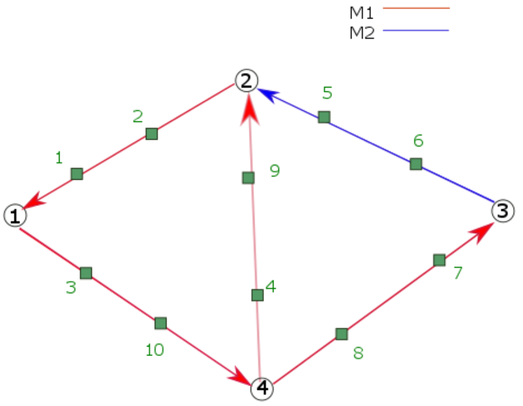

FIGURE 51 RESULT SCENARIO 1: STANDARD DEVIATION 1%

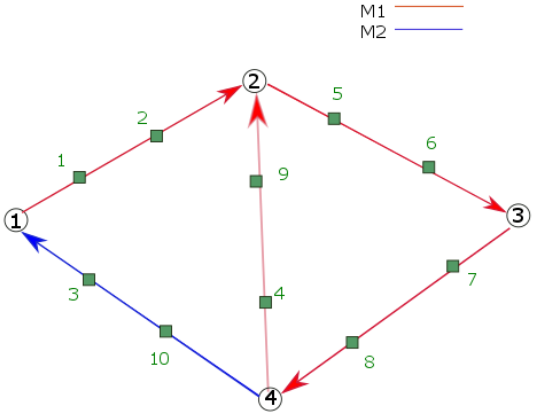

FIGURE 52 RESULT SCENARIO 1: STANDARD DEVIATION 5%

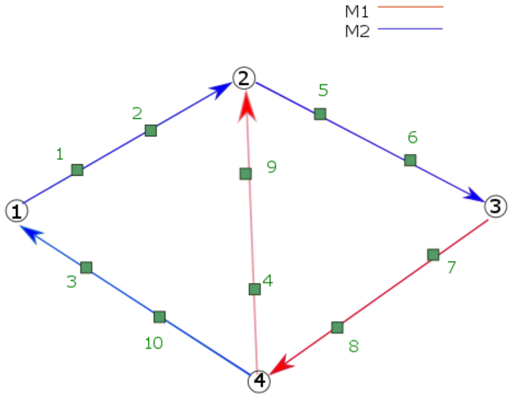

FIGURE 53 RESULT SCENARIO 1: STANDARD DEVIATION 10%

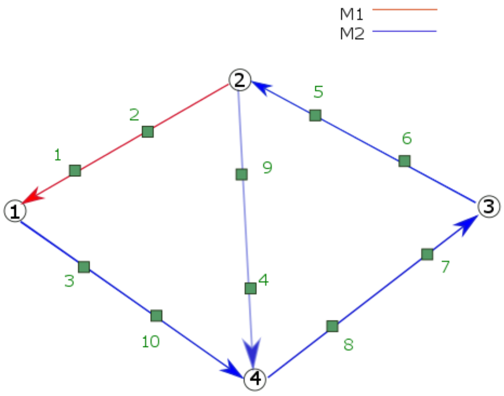

Figure 54 Result scenario 2: 10 samples

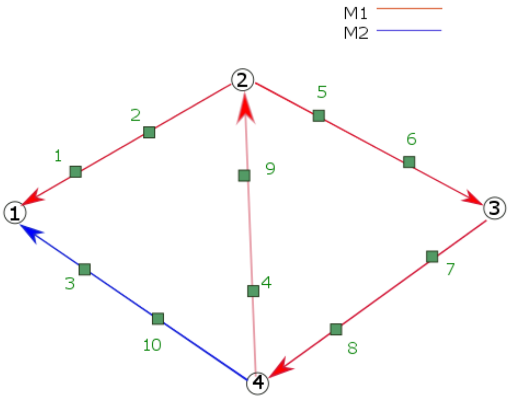

Figure 55 Result scenario 2: 20 samples

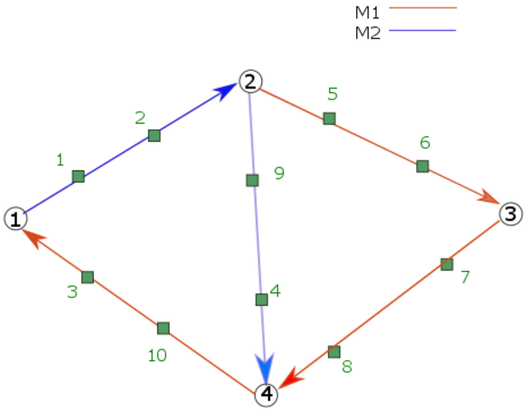

Figure 56 Deterministic problem result

CONCLUSIONS

The study presents a risk-based decision support system to schedule predictive maintenance activities, in order to intervene when a rail asset has reached a certain fault probability, avoiding failures.

The work shows how data, collected from the field, could be used to predict future asset condition and future failure rate trend in order to make decisions on asset management and maintenance execution. More in details, the presented formulation of the maintenance planning problem is able to consider degradation phenomena and failure risk in the evaluation of the optimal maintenance plan.

Moreover, the model allows to consider the space-distributed aspect of railway infrastructure, defining the optimal path along the railway network and the maintenance activities assignment for each maintenance team.

The available resources are optimized, minimizing the time for the maintenance execution and optimizing the utilization of the infrastructure possession time.

In particular, two different extensions of the model are introduced:

- the scheduling of predictive maintenance of a railway line.
- the scheduling of predictive maintenance of a railway network.

Moreover, to deal with the stochastic aspect of the problem a Rolling Horizon framework is introduced to manage unpredicted faults or delays in maintenance activities execution.

In doing so, it has been noted that, the adaptive rescheduling models only partially solve the issue of uncertainty, since they consider deterministic sub-problems of the overall problem and the stochastic input variables cannot vary continuously. For this reason, the risk-based maintenance planning problem is then formulated in term of stochastic programming. The stochastic formulation of the model, for scheduling predictive and risk-based maintenance activities in rail sector, introduces stochastic deadlines to consider explicitly the stochastic nature of risk and of real-world maintenance operations.

The foreseen impacts of the study are:

- Improvement of reliability and availability of railway transport with a consequent reduction of cost for unexpected railway repairs and trains circulation disruptions.
- Enhanced safety, planning the interventions when the failure probability level achieves a certain threshold.
- Improvement of railway asset management efficiency through the optimization of resources utilization, avoiding unnecessary trips of the maintenance teams.
- Improvement of railway capacity through the optimization of the infrastructure possession time.
- Enhanced decision-making process for the rail infrastructure manager.
- The shift from expensive and inefficient corrective and planned maintenance strategies to an effective and efficient predictive maintenance policy.

REFERENCES

[1] B. Dalla Chiara, D. De Franco, N. Coviello, D. Pastrone, (2017). Comparative specific energy consumption between air transport and high-speed rail transport: A practical assessment, Transportation Research Part D: Transport and Environment, vol. 52, Part A, pp. 227-243.

[2] ISO International Standardization Organisation, (2014). ISO 55000:2014 Preview Asset management -- Overview, principles and terminology.

[3] UIC International Union of Railways, Asset management working group, (2016). Practical implementation of Asset Management through ISO 55001.

[4] IN2RAIL Deliverable (2017), D6.5 – Condition- & Risk-based Maintenance Planning (CRIMP).

[5] SHIFT2RAIL Joint Undertaking, (2015). SHIFT2RAIL Multi Annual Action Plan.

[6] K. Kobbacy, D. Murthy (2007). Complex System Maintenance Handbook: Blending Theory with Practice, chap. 1, Springer Verlag, Berlin, pp. 3-18.

[7] M. Corfiati, B. Dalla Chiara, M. Galfrè, (2011). Evolution lines in the maintenance of rolling stock / Linee evolutive nella manutenzione dei rotabili ferroviari, Ingegneria Ferroviaria, n. 9, vol. LXVI, pp.751-772.

[8] P. Umiliacchi, D. Lane F. Romano, (2011). Predictive maintenance of railway subsystems using an ontology based modelling approach. 9th World Conference on Railway Research, May 22-26, Lille, France.

[9] G. Budai-Balke, (2009). Operations Research Models for Scheduling Railway Infrastructure Maintenance. PhD Thesis. Erasmus University Rotterdam.

[10] S. Sakai, (2010). Risk-based maintenance, JR East Technical Review 17.

[11] Improverail (2002). Final Report.

[12] L.E. Daniels, (2008). Track maintenance costs on rail transit properties. TCRP Document 43. TRB.

[13] C. Esveld, (2001). Modern Railway Track, MRT-Productions, Zaltbommel, The Netherlands.

[14] Office of the Rail Regulator (2001). The Possessions Review: Provisional Conclusions on the Schedule Incentive Structure, http://orr.gov.uk/__data/assets/pdf_file/0017/476/sch-4-8-consultation-2012.pdf.

[15] D. Huisman, L. Kroon, R. Lentink, and M. Vromans (2005), Operations Research in Passenger Railway Transportation, Statistica Neerlandica, vol. 59, pp. 467-497.

[16] EC OJ (2001). Official Journal L. 075. Directive 2001/14/EC of the European Parliament and of the Council of 26 February 2001 on the allocation of railway infrastructure capacity and the levying of charges for the use of railway infrastructure and safety certification. 15/03/2001, pp. 0029-0046.

[17] EC Official Journal (2016). Directive (EU) 2016/797 on the interoperability of the rail system within the European Union (Recast of Directive 2008/57/EC).

[18] EC Official Journal (2016). Directive (EU) 2016/798 on railway safety (Recast of Directive 2004/49/EC).

[19] S. Van Akena, N. Bešinović, R. M. P. Goverde, (2017). Designing alternative railway timetables under infrastructure maintenance possessions, Transportation Research Part B: Methodological vol. 98, pp. 224-238.

[20] T. Lidén, M. Joborn, (2017). An optimization model for integrated planning of railway traffic and network maintenance, Transportation Research Part C: Emerging Technologies vol. 74, pp. 327-347.

[21] IN2RAIL Deliverable D6.6 – Dynamic and Real-time Maintenance Planning Concept, 2017.

[22] M. Cross, (1988). Raising the Value of Maintenance in the Corporate Environment, Management Research News, vol.11, pp. 8-11.

[23] A. Zoeteman, (2007). ProRail's Management of Tracks and Turnouts, European Railway Review, http://www.europeanrailwayreview.com/1516/past-issues/prorail%E2%80%99s-management-of-tracks-and-turnouts/.

[24] A. Zoeteman, R. van Zelm (2001). Developing a Cost-effective Renewal Strategy for Railway Track on Conventional Networks - A Practical Example of a Policy Revision at Netherlands Railways, in Proceedings of the 5th World Conference on Railway Research, Cologne.

[25] N. Jimenez, A. Barragan, P. Cembrero, F. G. Benitez, N. Caceres, F. Schubert, A. Simroth, C. Santanera, (2010). ACEM-RAIL W.P.1 Task1.1 D1.1 Report on the state of practice of railway infrastructure maintenance. Seventh

Framework Programme theme sst.2010.5.2.1.

[26] J. R. Birge, F. Louveaux, (1997). Introduction to Stochastic Programming. Springer Series in Operations Research, Springer.

[27] R. L. Graham, E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, (1979). Optimization and Approximation in Deterministic Sequencing and Scheduling: A Survey. Annals of Discrete Mathematics, vol.5, pp. 287–326.

[28] J. Blazewicz. (1987). Selected Topics in Scheduling Theory. Annals of Discrete Mathematics, vol.31, pp.1–60.

[29] R. W. Conway, W. L. Maxwell, L. W. Miller, (1967). Theory of Scheduling. Addison-Wesley.

[30] J. R. Jackson, (1955). Scheduling a Production Line to Minimize Maximum Tardiness. Technical Report Research Report 43, Management Science Research Project, UCLA.

[31] E. L. Lawler (1973). Optimal Sequencing of a Single Machine Subject to Precedence Constraints. Management Science, vol.19, pp.544–546.

[32] W. E. Smith (1956). Various Optimizers for Single-Stage Production. Naval Research Logistics Quarterly, vol. 3, pp.59–66.

[33] J. M. Moore, (1968). An n Job, One Machine Sequencing Algorithm for Minimizing the Number of Late Jobs. Management Science, vol. 15, n.1, pp. 102–109.

[34] J. Bruno, E. G. Coffman, Jr. (1974), R. Sethi. Scheduling Independent Tasks to Reduce Mean Finishing Time. CACM, vol. 17, pp.382–387.

[35] J. K. Lenstra, A. H. G. Rinnooy Kan, P. Brucker (1977). Complexity of Machine Scheduling Problems. Annals of Discrete Mathematics, vol. 4, pp. 281–300.

[36] R. McNaughton. (1959). Scheduling with Deadlines and Loss Functions. Management Science, vol. 6, n.1 pp. 1–12.

[37] J. K. Lenstra, A. H. G. Rinnooy Kan (1979). Computational Complexity of Discrete Optimization Problems. Annals of Discrete Mathematics, vol. 4, pp. 121–140.

[38] C. W. Mercer (1992) An Introduction to Real-Time Operating Systems: Scheduling Theory.

[39] P. Brucker (2006), Scheduling Algorithms, Springer-Verlag Berlin Heidelberg.

[40] S. Pirkwieser, (2012), Hybrid Metaheuristics and Matheuristics for Problems in Bioinformatics and Transportation, Bakk.techn.

[41] J. Puchinger, G. R. Raidl, (2005). Combining metaheuristics and exact algorithms in combinatorial optimization: A survey and classification. In J. Mira and J. Álvarez, editors, Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach, vol. 3562 of LNCS, pp. 113-124. Springer Berlin Heidelberg.

[42] G. R. Raidl, J. Puchinger, (2008). Combining (integer) linear programming techniques and metaheuristics for combinatorial optimization. Studies in Computational Intelligence pp. 31-62. Springer-Verlag Berlin Heidelberg.

[43] V. Maniezzo, T. Stützle, (2010), Matheuristics: Hybridizing Metaheuristics and Mathematical Programming, vol. 10 of Annals of Information Systems, Springer.

[44] M. O. Ball, (2011). Heuristics based on mathematical programming, Elsevier, Surveys in Operations Research and Management Science vol. 16 pp. 21-38.

[45] E. Talbi, (2015). Combining metaheuristics with mathematical programming, constraint programming and machine learning, Springer Science Business Media New York.

[46] F. Della Croce, A. Grosso, F. Salassa, (2014) A matheuristic approach for the two-machine total completion time flow shop problem. Annals of Operation Research, vol. 213, pp. 67–78.

[47] P. Baptiste, A. Jouglet, D. Savourey, (2011) Lower Bounds for Parallel Machine Scheduling Problems.

[48] W.A. Horn (1973). Minimizing average flow time with parallel machines. Operations Research, vol. 21, pp. 846–847.

[49] F. Yalaoui, C. Chu (2005). A new exact method to solve the $P_m | r_i | \sum C_i$ problem. International Journal of Production Economics, vol. 100, n.1, pp. 168-179.

[50] H. Simonis, (1995). Calculating Lower Bounds on a Resource Scheduling Problem.

[51] A. Allahverdi, J.N.D. Gupta, T. Aldowaisan, (1999). A review of scheduling research involving setup considerations, OMEGA The International Journal of Management Sciences vol. 27, pp. 219–239.

[52] A. Allahverdi, C.T. Ng, T.C.E. Cheng, M. Y. Kovalyov, (2008). A survey of scheduling problems with setup times or costs, European Journal of Operational Research vol. 187, pp. 985–1032.

[53] E. Anand, R. Panneerselvam, (2015). Literature Review of Open Shop Scheduling Problems, Intelligent Information Management, Scientific Research Publishing vol. 7, pp. 33-52.

[54] A. Zoeteman, C. Esveld, (2004). State of the art in railway maintenance management: planning systems and their application in Europe. IEEE International Conference on Systems, Man and Cybernetics, vol.5, pp. 4165-4170.

[55] N. Jimenez, A. Barragan, P. Cembrero, F. G. Benitez, N. Caceres, F. Schubert, A. Simroth, C. Santanera, (2010). ACEM-RAIL W.P.1 Task1.1 D1.1 Report on the state of practice of railway infrastructure maintenance. Seventh Framework Programme theme sst.2010.5.2.1.

[56] F. Camci, (2015). Maintenance scheduling of geographically distributed assets with prognostics information. European Journal of Operational Research, vol. 245, pp. 506–516.

[57] S. S. Soh, N.H.M. Radzi, H. Haron, (2012). Review on Scheduling Techniques of Preventive Maintenance Activities of Railway. Proc. 4th International Conference on Computational Intelligence, Modelling and Simulation (CIMSiM), pp.310-315.

[58] A. Higgins, (1998). Scheduling of Railway Track Maintenance Activities and Crews. Journal of Operational Research Society, vol. 49, pp. 1026-33.

[59] G. Budai, D. Huisman, R. (2004). Proc. IEEE International Conference on System, Man and Cybernetics, vol. 5, pp.4171- 4176.

[60] F. Peng, S. Kang, X. Li, Y. Ouyang, (2011). A heuristic approach to the railroad track maintenance scheduling problem. Computer-Aided Civil and Infrastructure Engineering vol. 26, n.12, pp. 129–145.

[61] C. Borraz-Sanchez, D. Yang, D. Klabjan, (2011). Strategic Gang Scheduling for Railroad Maintenance. Operations Research and Management Science, Railway Application Section.

[62] M. Dell'Orco, M. Ottomanelli, L. Caggiani, D. Sassanelli (2008). New Decision Support System for Optimization of Rail Track Maintenance Planning Based on Adaptive Neurofuzzy Inference System. Transportation Research Record: Journal of the Transportation Research Board, n. 2043, Transportation Research Board of the National Academies, Washington, D.C., pp. 49–54.

[63] P. Umiliacchi, D. Lane, F. Romano, (2011). Predictive maintenance of railway subsystems using an ontology based modelling approach. 9th World Conference on Railway Research.

[64] R. Shingler, G. Fadin, P. Umiliacchi, (2008). From RCM to predictive maintenance: The InteGRail approach. 4th IET International Conference on Railway Condition Monitoring, pp. 1-5.

[65] N. Jiménez-Redondo, S. Escriba, F.G. Benítez, F. Cores, N. Cáceres (2014). Towards automated and cost-efficient track maintenance. Final developments of the ACEM-Rail project. Transport Research Arena 2014.

[66] S. A. Simson, L. Ferreira, M. H. Murray (2000). Rail Track Maintenance Planning. An Assessment Model. Transportation Research Record: Journal of the Transportation Research Board, n. 1713, Transportation Research Board of the National Academies, Washington, D.C., pp. 29–35.

[67] M. P. N. Burrow, S. Naito, H. T. Evdorides (2009). Network-Level Railway Track Maintenance Management Model. Transportation Research Record: Journal of the Transportation Research Board, n. 2117, Transportation Research Board of the National Academies, Washington, D.C., pp. 66–76.

[68] T. Zhang, J. Andrews, R. Wang, (2013). Optimal scheduling of track maintenance on a railway network. Quality and Reliability Engineering International, vol. 29, pp. 285-297.

[69] J. Carretero, J. M. Pérez, F. García-Carballeira, A. Calderón, J. Fernández, J.D. García, A. Lozano, L. Cardona, N. Cotaina, P. Prete, (2003). Applying RCM in large scale systems: a case study with railway networks. Reliability Engineering & System Safety, vol. 82, n. 3, pp. 257-273.

[70] U. Bharadwaj, (2007) A risk based approach to maintenance optimisation of business critical railway structures/ equipment. Seminar on the Application of New Technologies to Railways, 2007 Institution of Engineering and Technology, pp. 53-60.

[71] M. Wang, K. Ramamohanarao, J. Chen, (2012). Dependency-based risk evaluation for robust workflow scheduling. Proc. of IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum.

[72] Y. F. Wang, Y. F. Zhang, J.Y. H. Fuh, Z. D. Zhou, P. Lou, L. G. Xue, (2008) An integrated approach to reactive scheduling subject to machine breakdown. IEEE International Conference on Automation and Logistics, ICAL, pp.542-547.

[73] Y. Gao, D. Yu-Si, Z. Hong-Yu, (2009). Job-shop scheduling considering rescheduling in uncertain dynamic environment, Proc. of International Conference on Management Science and Engineering, ICMSE, pp.380-384.

[74] R. M. Pathan, (2007). Recovery of fault-tolerant real-time scheduling algorithm for tolerating multiple transient faults. Proc. of 10th International Conference on Computer and Information Technology, ICCIT, pp. 1-6.

[75] L. Ma, J. Kang, C. Zhao and S. Y. Liu, (2007). Modelling the impact of prognostic errors on CBM effectiveness using discrete-event simulation. Proc. of International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, ICQR2MSE, pp. 520-525.

[76] S. Narayanaswami, N. Rangaraj, (2015). A MAS architecture for dynamic, realtime rescheduling and learning applied to railway transportation, Expert Systems with Applications, vol. 42, n.5, pp.2638-2656.

[77] L. Meng, X. Zhou, (2014). Simultaneous train rerouting and rescheduling on an N-track network: A model reformulation with network-based cumulative flow variables. Transportation Research Part B: Methodological, vol. 67, pp. 208-234.

[78] M. Baldi, F. Heinicke, A. Simroth, R. Tadei. (2016) New heuristics for the Stochastic Tactical Railway Maintenance Problem. Omega, vol. 63, pp. 94-102.

[79] J. Andrews, D. Prescott, F. De Rozières, (2014). A stochastic model for railway track asset management. Reliability Engineering and System Safety, vol.130, pp. 76–84.

[80] C. Cassandras, S. Lafortune, (2008) Introduction to Discrete Event Systems. Springer Science, New York. 800. 10.1007/978-0-387-68612-7.

[81] T. Al-Khamis, R. M'Hallah, (2011). A two-stage stochastic programming model for the parallel machine scheduling problem with machine capacity. Computers & Operations Research, vol. 38, pp. 1747–1759.

[82] N. Jiménez-Redondo, Nicola Bosso, Luigi Zeni, Aldo Minardo, Frank Schubert, Franziska Heinicke, Axel Simroth, (2012). Automated and Cost Effective Maintenance of Railways. Transport Research Arena 2012, Procedia Social and Behavioral Sciences, vol. 48, pp. 1058-1066.

[83] A. Minardo, G. Porcaro, D. Giannetta, R. Bernini, L. Zeni, (2013). Real-time monitoring of railway traffic using slope-assisted Brillouin distributed sensors. Applied Optics, vol.52, n.16, pp. 3770-3776.

[84] F. Schubert, M. Barth, U. Hoenig, H. Neunuebel, M. Stephan, A. Gommlich, L. Schubert, B. Frankestein, (2013). Characterisation of Rail Tracks by Means of Hollow Shaft Integrated Acoustic Sensor System. European Transport Conference 2013, Frankfurt.

[85] B. Rockstroh, W. Kappes, F. Walte, (2008). Ultrasonic and eddy current inspection of rail wheels and wheel set axles, WCNDT 2008.

[86] T. Heckel, H. Thomas, M. Kreutzbruck, S. Rühe, (2009). High Speed Non-Destructive Rail Testing with Advanced Ultrasound And Eddy-Current Testing Techniques, Indian National Seminar & Exhibition on Non-Destructive Evaluation NDE 2009, December 10-12.

[87] H. Salzburger, L. Wang, X. Gao, (2008). In-motion ultrasonic testing of the tread of high-speed railway wheels using the inspection system AUROPA III, Proceedings of WCNDT 2008.

[88] C. Peng, L. Wang, X. Gao, Z. Wang, Q. Zhao, Y. Zhang, J. Peng, K. Yang, (2010). Dynamic Wheel Set Defect Inspection by Applying Ultrasonic Technology, ECNDT 2010.

[89] S. Štarman, V. Matz, (2008) Automatic System for Crack Detection Using Infrared Thermographic Testing, WCNDT 2008.

[90] A. Elballouti, S. Belattar, N. Laaidi, (2010). Thermal Non-Destructive Characterization of the Spherical Defects in the Roadways, ECNDT 2010.

[91] I. Soleimanmeigouni, A. Ahmadi, (2016) A Survey on Track Geometry Degradation Modelling. Current Trends in Reliability, Availability, Maintainability and Safety. Lecture Notes in Mechanical Engineering. Springer, Cham, pp. 3-12.

[92] J. Sadeghi, M. Fathali, N. Boloukian, (2009). Development of a new track geometry assessment technique incorporating rail cant factor. Proc Inst Mech Eng Part F J Rail Rapid Transit vol.223, n.3, pp.255–263.

[93] J. Sadeghi, H. Askarinejad (2011). Development of track condition assessment model based on visual inspection. Struct Infrastruct Eng vol.7, n.12, pp. 895–905.

[94] R. Faiz, S. Singh (2009). Time based predictive maintenance management of UK rail track. In: ICC'09. International Conference on computing, engineering and information, pp. 376–383.

[95] H. Li, T. Xiao, (2014). Improved generalized energy index method for comprehensive evaluation and prediction of track irregularity. J Stat Comput Simul vol.84, n.6, pp.1213–1231.

[96] H. Li, Y. Xu, (2009). Railway Track Integral Maintenance Index and Its Application. Int Conf Transp Eng pp. 2514–2519.

[97] M. El-Sibaie, Y. Zhang, (2004). Objective track quality indices. Transp Res Record J Transp Res Board vol. 1863, pp. 81–87.

[98] M. Shenton, (1985). Ballast deformation and track deterioration. Track Technol pp. 253–265.

[99] Y. Sato, (1997). Optimum track structure considering deterioration in ballasted track. In: Proceedings of 6th international heavy haul conference, Cape Town, SouthAfrica, p 576.

[100] Y. Sato, (1995).Japanese studies on deterioration of ballasted track. Veh Syst Dyn vol. 24, n.1, pp.197–208.

[101] S. Chrismer, E. Selig, (1993). Computer model for ballast maintenance planning. In: Proceedings of 5th International Heavy Haul Railway Conference, Beijing, China, pp. 223–227.

[102] J. Öberg, E .Andersson, (2009). Determining the deterioration cost for railway tracks. Proc Inst Mech Eng Part F, J Rail Rapid Transit vol.223, n.2, pp.121–129.

[103] Y. Zhang, M.H. Murray, L. Ferreira (2000) Modelling rail track performance: an integrated approach. Transp J pp.187–194.

[104] T. Dahlberg, (2001). Some railroad settlement models—a critical review. Proc Inst Mech Eng Part F J Rail Rapid Transit vol.215, n.4, pp.289–300.

[105] A.R. Andrade, P.F. Teixeira, (2011).Uncertainty in rail-track geometry degradation: Lisbonoporto line case study. J Transp Eng. Vol.137, pp.193-200.

[106] R. Guo, B.M. Han, (2013). Multi-stage linear prediction model for railway track irregularity. Appl Mech Mater vol.361, pp.1811–1816.

[107] S.M. Famurewa, T. Xin, M. Rantatalo, U. Kumar, (2013). Comparative study of track geometry quality prediction models. International Conference on Condition Monitoring and Machinery Failure Prevention Technologies.

[108] MAINLINE MAINtenance, renewal and Improvement of rail transport iNfrastructure to reduce Economic and environmental impacts Deliverable 2.2: Degradation and intervention modelling techniques. Seventh Framework Programme Theme SST.2011.5.2-6.: Cost-effective improvement of rail transport infrastructure.

[109] N. Lyngby, (2009). Railway track degradation: Shape and influencing factors. Int J Perform Eng vol.5, n.2, pp.177.

[110] R. Liu, P. Xu, F. Wang, (2010). Research on a short-range prediction model for track irregularity over small track lengths. J Transp Eng vol.136, n.12, pp. 1085–1091.

[111] P. Xu, Q. Sun, R. Liu, F. Wang, (2011). A short-range prediction model for track quality index. Proc Inst Mech Eng Part F J Rail Rapid Transit vol. 225, n.3, pp.277–285.

[112] A. Kawaguchi, M. Miwa, K. Terada, (2005). Actual data analysis of alignment irregularity growth and its prediction model. Q Rep RTRI vol.46, n. 4, pp.262–268.

[113] L. Quiroga, E. Schnieder, (2010). Modelling high speed railroad geometry ageing as a discrete-continuous process. In: Proceedings of the stochastic modeling techniques and data analysis international conference, SMTDA.

[114] C. Vale, S.M. Lurdes, (2013). Stochastic model for the geometrical rail track degradation process in the portuguese railway northern line. Reliab Eng Syst Saf vol.116, pp. 91–98.

[115] M. Zhu, X. Cheng, L. Miao, X. Sun, S. Wang, (2013). Advanced stochastic modeling of railway track irregularities. Adv Mech Eng vol.5, pp.401-637.

[116] L. Bai, R. Liu, Q. Sun, F. Wang, P. Xu, (2015). Markov-based model for the prediction of railway track irregularities. Proc Inst Mech Eng Part F J Rail Rapid Transit vol. 229, n.2, pp.150–159.

[117] M. Yousefikia, S. Moridpour, S. Setunge, E. Mazloumi, (2014). Modeling degradation of tracks for maintenance planning on a tram line. J Traffic Logist Eng vol.2, n.2.

[118] F. Liu, X. Wei, Y. Liu, X. Yin, (2014). Track quality prediction based on center approach Markov-Grey GM (1, 1) model. In: 2014 IEEE international conference on information and automation (ICIA), pp 81–86.

[119] A.R. Andrade, P.F. Teixeira, (2012). A bayesian model to assess rail track geometry degradation through its life-cycle. Res Transp Econ vol.36 n.1, pp.1–8.

[120] H. Guler, (2014). Prediction of railway track geometry deterioration using artificial neural networks: A case study for turkish state railways. Struct Infrastruct Eng vol.10, n.5, pp.614–626.

[121] J. Chaolong, X. Weixiang, W. Lili, W. Hanning, (2013). Study of railway track irregularity standard deviation time series based on data mining and linear model. Mathematical Problems in Engineering, vol. 2013, pp.12.

[122] P. Xu, R. Liu, F. Wang, Q. Sun, H. Teng, (2011). A novel description method for track irregularity evolution. Int J Comput Intell Syst vol.4, n.6, pp.1358–1366.

[123] P. Xu, R. Liu, F. Wang, F. Wang, Q. Sun, (2013). Railroad track deterioration characteristics based track measurement data mining. Mathematical Problems in Engineering, vol. 2013, pp.7.

[124] J. Chaolong, X. Weixiang, W. Futian, W. Hanning, (2012). Track irregularity time series analysis and trend forecasting. Discret Dyn Nat Soc 2012.

[125] E. Fumeo, D. Anguita, D. Galar, N. Mazzino, C. Milani, C. Dambra, (2016) Asset Status Nowcasting and Forecasting in Integrated Railway Systems 11th WCRR 2016, Milan, Italy.

[126] P. Wang, D. W. Coit, (2007). Reliability and Degradation Modeling with Random or Uncertain Failure Threshold. 2007 Annual Reliability and Maintainability Symposium, Orlando, pp. 392-397.

[127] W. Ke, C. Ren, K. Jin, H. Lu, (2007). System performance, degradation, and reliability assessment. Proc. of 2007 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, pp. 1216-1220.

[128] S. M. Famurewa, U. Juntti, A. Nissen, U. Kumar, (2016). Augmented utilization of possession time: analysis for track geometry maintenance, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 230, n.4.

[129] S. M. Famurewa, T. Xin, M. Rantatalo and U. Kumar, "Optimisation of maintenance track possession time: A tamping case study," Proc.Inst.Mech.Eng.Pt.F: J.Rail Rapid Transit, 2013.

[130] A. Yokoyama, (2015). Innovative changes for maintenance of railway by using ICT - To Achieve “Smart Maintenance” - The Fourth International Conference on Through-life Engineering Services, Procedia CIRP, vol. 38, pp. 24 – 29.

[131] A. Consilvio, A. Di Febbraro, N. Sacco, (2015). A modular model to schedule predictive railway maintenance operations. 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp.426-433.

[132] A. Consilvio, A. Di Febbraro, R. Meo, N. Sacco, (2018). Risk-Based Optimal Scheduling of Maintenance Activities In a Railway Network, EURO Journal on Transportation and Logistics, pp.1-31, 2018, DOI 10.1007/s13676-018-0117-z.

[133] A. Consilvio, A. Di Febbraro and N. Sacco, (2016). Stochastic scheduling approach for predictive risk-based railway maintenance, 2016 IEEE International Conference on Intelligent Rail Transportation (ICIRT), Birmingham, pp. 197-203.

[134] A. Consilvio, A. Di Febbraro, N. Sacco, (2016). Two-Level Rolling Horizon Procedure for Railway Predictive Maintenance Scheduling. Transportation Research Board 2016 Annual meeting Compendium of papers 16-2643, Session 596. Washington D.C., January 2016.

ANNEX MATLAB CODE

MATHEURISTIC CODE FOR THE ROLLING HORIZON APPROACH

```
clc
clear
close all

global consolidated
global firstToAdd

n = 30;
consolidated = zeros(1,n); % matrix that has element 1 if the job is already
% executed, 0 otherwise
firstToAdd = n+1;

%initial solution top to end
[fval,X,Y,Ts,C] = resAnalyzer_iniziale30();

%writing of the matrix Sol1 and Sol2 of the jobs assigned to the first and
%second machine according to the initial solution
next = 1;
l = 1;
for j = 1:n
    job = find(X(next,:)==1);
    if job == n+1 % last job
        break
    end
    if isempty(job)
        disp('MAI')
        break
    end
    Sol1(1,l) = job; % jobs assigned to machine 1.
    next = job+1; % next job
    l = l+1;
end
next = n+2;
l = 1;
for j = 1:n
    job = find(X(next,:)==1);
    if job == n+1 % last job
        break
    end
    if isempty(job)
        disp('MAI')
        break
    end
```

```
    Sol2(1,l) = job; % jobs on machine 2
    next = (n+1)+job+1;  % next job
    l = l+1;
end
```

FIGURE 57 INITIAL SOLUTION

```
f = [];
i = 1;
t_tot=0;
tf = 0;
for k = 0:15 % number of rescheduling iteration
    [m,t,d,ds,w,lt,It,p1,p2,S1,S2] = generateData(n,X,Y,Ts,C);
% function that updates the data
    t_sol = 0;
    ni = 0;
    while t_sol < 3600
        writeF(X,n,m,t,d,ds,w,lt,It,p1,p2,S1,S2);
 % function that fixes some variables of the sequences matrix
        tic
        !"C:\Program
Files\IBM\ILOG\CPLEX_Studio125\opl\bin\x64_win64\oplrun.exe"
"C:\Users\Alice.Consilvio\Opl\matheuristic6\matheuristic6.mod"
"C:\Users\Alice.Consilvio\Opl\matheuristic6\matheuristicMatlab.dat"
        time = toc;
        t_sol = t_sol + time
        [fval,X,Y,Ts,C] = resAnalyzer();
        f = [f; fval];
        tf = [tf; tf(end)+time];
        ni = ni +1;
    end
    t_tot = [t_tot t_tot(end)+t_sol];
    i = [i i(end)+ni];
%writing of the matrix Sol1 and Sol2 of the jobs assigned to the first and
%second machine according to the final solution
    next = 1;
    l = 1;
    for j = 1:n
        job = find(X(next,:)==1);
        if job == n+1 % last job
            break
        end
        if isempty(job)
            disp('MAI')
            break
        end
        Sol1(k+2,l) = job;
        next = job+1;  % next job
        l = l+1;
    end
    next = n+2;
    l = 1;
    for j = 1:n
        job = find(X(next,:)==1);
```

```
        if job == n+1 % last job
            break
        end
        if isempty(job)
            disp('MAI')
            break
        end
        Sol2(k+2,l) = job;
        next = (n+1)+job+1;  % last job
        l = l+1;
    end
end
i(end) = i(end)-1;

for kk = 1:n
    if Y(kk,1) == 1
        consolidated(end,kk) = 1;% writing of the already executed jobs
    end
end
%plot of the objective function
stairs(tf(1:end-1),f/max(f),'LineWidth',1)
for k = 1:length(t_tot)
    line([t_tot(k) t_tot(k)],[0 1],'LineStyle','--')
end

save Result30.mat Sol1 Sol2 consolidated f t_tot tf
```

FIGURE 58 SOLUTION IMPROVEMENT

```
function [m,t,d,ds,w,lt,It,p1,p2,S1,S2] = generateData(n,X,Y,Ts,C);

% this function allows to consider the updating input at each decision time.
%In particular: the updated deadlines and priorities of the maintenance
%activities.
global consolidated
global firstToAdd

%% data definition

m = 4;
t= 20;

% the jobs executed during the previous days are deleted and new jobs are
%considered

steps = size(consolidated,1);

if steps == 1
    d = d_static(1:n);
    ds = ds_static(1:n);
    w = w_static(1:n);
    p1 = p1_static(1:n);
    p2 = p2_static(1:n);
    save data_last.mat d ds w p1 p2
    consolidated(end+1,1) = 0;  % new row
```

```
    return
end

for i = 1:n
    if Y(i,1) == 1
        consolidated(steps,i) = 1;
    end
end
consolidated(end+1,1) = 0;  % new row

load data_Last

i = randperm(100);   % permutation on the priorities
wPerm = w_static(i);

for j = n:-1:1
    d(j) = d(j) - 1440; % after one day
    ds(j) = ds(j) - 1440;
    if consolidated(steps,j) == 1
        d(j) = [];
        d(end+1) = d_static(firstToAdd);
        ds(j) = [];
        ds(end+1) = ds_static(firstToAdd);
        w(j) = [];
        w(end+1) = wPerm(firstToAdd);
        p1(j) = [];
        p1(end+1) = p1_static(firstToAdd);
        p2(j) = [];
        p2(end+1) = p2_static(firstToAdd);
        firstToAdd = firstToAdd + 1;
    end
end

return
```

FIGURE 59 UPDATING INPUT AT EACH DECISION TIME

```
function y = writeF(X,n,m,t,d,ds,w,lt,It,p1,p2,S1,S2)
%% data writing
fID =
fopen('C:\Users\Alice.Consilvio\opl\matheuristic6\matheuristicMatlab.dat','wt')
;
fprintf(fID,['n=',num2str(n),';\n']);
fprintf(fID,['m=',num2str(m),';\n']);
fprintf(fID,['t=',num2str(t),';\n']);

fprintf(fID,'p = [[');
for k = 1:n
    if n ~= length(p1)
        keyboard
    end
    fprintf(fID, [num2str(p1(k)), ' ']);
end
fprintf(fID,'],[');
for k = 1:n
```

```
    fprintf(fID, [num2str(p2(k)), ' ']);
end
fprintf(fID,']];\n');

fprintf(fID,'d = [');
for k = 1:n
    if n ~= length(d)
        keyboard
    end
    fprintf(fID, [num2str(d(k)), ' ']);
end
fprintf(fID,'];\n');

fprintf(fID,'ds = [');
for k = 1:n
    fprintf(fID, [num2str(ds(k)), ' ']);
end
fprintf(fID,'];\n');

fprintf(fID,'w = [');
for k = 1:n
    fprintf(fID, [num2str(w(k)), ' ']);
end
fprintf(fID,'];\n');

fprintf(fID,'lt = [');
for k = 1:10
    fprintf(fID, [num2str(lt(k)), ' ']);
end
fprintf(fID,'];\n');

fprintf(fID,'It = [');
for k = 1:1:10
    fprintf(fID, [num2str(It(k)), ' ']);
end
fprintf(fID,'];\n\n');

fprintf(fID,'S=[[');
for i = 1:n+1
    fprintf(fID,'[');
    for j = 1:n
        fprintf(fID, [num2str(S1(i,j)), ' ']);
    end
    fprintf(fID,']\n');
end
fprintf(fID,'],[');
for i = 1:n+1
    fprintf(fID,'[');
    for j = 1:n
        fprintf(fID, [num2str(S2(i,j)), ' ']);
    end
    fprintf(fID,']\n');
end
fprintf(fID,']];');

fprintf(fID,'//matrix for fixing variables\n');
```

```
fprintf(fID,'\n');
fprintf(fID,'///////////////////////////////////////////////////////////////////
/////////////////////////  );\n');

pr = randperm(2*n+2,50);
p = zeros(2*n+2,1);
p(pr) = 1;

fprintf(fID,'F =[[');

for i = 1:n+1
    fprintf(fID,'[');
    if p(i) == 1
        for j = 1:n+1
            if X(i,j) == 1
                fprintf(fID,'1 ');
            else
                fprintf(fID,'2 ');
            end
        end
    else
        for j = 1:n+1
            fprintf(fID,'0 ');
        end
    end
    fprintf(fID,']\n');
end
 fprintf(fID,'],\n[');
for i = n+2:2*n+2
    fprintf(fID,'[');
    if p(i) == 1
        for j = 1:n+1
            if X(i,j) == 1
                fprintf(fID,'1 ');
            else
                fprintf(fID,'2 ');
            end
        end
    else
        for j = 1:n+1
            fprintf(fID,'0 ');
        end
    end
    fprintf(fID,']\n');
end
fprintf(fID,']];\n');

fclose(fID);

return
```

FIGURE 60 FUNCTION FOR FIXING SOME SEQUENCE VARIABLES

MATHEURISTIC CODE FOR THE SCHEDULING ON A RAIL NETWORK

```
clc
    clear all
    close all
%global variables declaration
    global machine
    global path1
    global path2
    global path6
    global path7
    nodes= 14;
    jobs= 42;
    teams= 4;
    timewindows= 10;
    ciclitot=50;

 %  Data files
    Data = 'Genova_4T_42J.dat';

 % Root folder directory
    root =
'C:\Users\Alice.Consilvio\Desktop\Cplex\Assegnazione_multiarco\';
%Data for matheuristics
    path1 = strcat( root, 'ProblemData\Data_', Data);
    path2 = strcat( root,'Matheuristics.dat');
    path3 = strcat( root,'ResAnalyzer.m');
    path4 = strcat( root,'BestSolution.m');
    path5 = strcat( root, 'ResAnalyzer_start.m');
%Data independent sequences assignment
    path6 = strcat( root, '\OrderingData\Ordering_', Data );

    path7 = strcat( root, 'Partial_Heuristics.dat');
%Data for task assignment
    path8 = strcat( root, '\AssignmentData\Ass_', Data );
    path9 = strcat( root, 'Assignment_Data.dat' );
%Cleaning of ResAnalyzer.m and ResAnalyzer_Start.m
    fID3 = fopen(path3 ,'wt');
    fprintf(fID3,'');
    fclose(fID3);
    fID5 = fopen(path5 ,'wt');
    fprintf(fID5,'');
    fclose(fID5);
    fID8 = fopen( path9 ,'wt');
    fprintf(fID8,'');
    fclose(fID8);
    copyfile(path8 , path9);
```

FIGURE 61 INIZIALIZATION

```
    f = [];
    t_tot=0;
    t_sol=0;
    tryals = 0;
    tic
    % Cplex solves the task assignment problem

!"C:\ProgramFiles\IBM\ILOG\CPLEX_Studio125\opl\bin\x64_win64\oplrun.exe
" -p "C:\Users\Alice.Consilvio\Desktop\Cplex\Assegnazione_multiarco"
"Path_Assignment"

    [Y_0, Z_0, A] = Starting_Paths();
    disp('Task Assignment Completed');
```

Figure 62 Task Assignment problem

```
X_par= zeros(jobs+1,jobs+1,teams);
    X_parziale=[];

    % independent sequence matrix

   for machine = 1: teams
        WriteFtoptoend(Y_0,jobs,teams,machine,X_par);
        tic
        !"C:\Program
Files\IBM\ILOG\CPLEX_Studio125\opl\bin\x64_win64\oplrun.exe" -p
"C:\Users\Alice.Consilvio\Desktop\Cplex\Assegnazione_multiarco"
"First_Partial_Ordering"
        time=toc;

        [X_par] = Parziale();

   end
    X_1 = X_par;

    t_seq_iniziale = toc;
    t_iniz= t_seq_iniziale;
    disp('Unrelated machine ordering completed');
    resanalyzer= dir(path5);
```

Figure 63 Independent sequence assignment

```
    % first feasible solution
    while resanalyzer.bytes == 0

        SetUp(X_1,jobs,teams,tryals);
    % at first iteration X_1 is fixed, then it changes randomly
        tic
        !"C:\Program
Files\IBM\ILOG\CPLEX_Studio125\opl\bin\x64_win64\oplrun.exe" -p
"C:\Users\Alice.Consilvio\Desktop\Cplex\Assegnazione_multiarco" "SetUp"
        time_iniz = toc;
```

```
        t_iniz = [t_iniz; t_iniz(end)+time_iniz];
        resanalyzer= dir(path5);
        tryals = tryals +1;

    end

    disp('Feasible solution found');
    [fval,Y_2,X_2,Z_2,TW_2,Ts_2,C_2,proc] = ResAnalyzer_start();
    f = [f; fval];
    tf = [tf; t_iniz(end)];
    Y_2in = Y_2;
    X_3=[];
    iterations=0;
    ciclo_esterno=0;
    tf = t_iniz(end);
    X_3=X_2;
```

FIGURE 64 INITIAL SOLUTION

```
    % iterative process for solution improvement
    while ciclo_esterno <= ciclitot

        while t_sol < 3600

            WriteF(X_3,jobs,teams,iterations);
    % function that fixes some rows of X_3
            tic
            !"C:\Program
Files\IBM\ILOG\CPLEX_Studio125\opl\bin\x64_win64\oplrun.exe" -p
"C:\Users\Alice.Consilvio\Desktop\Cplex\Assegnazione_multiarco"
"Matheuristics"
            time = toc;

            t_sol = t_sol+time;
            [fval,Y_4,X_4,Z_4,TW_4,Ts_4,C_4] = ResAnalyzer();
   % Reads Cplex solution as a function
            f = [f; fval];
            tf = [tf; tf(end)+time];
            iterations = iterations+1;
            X_3= X_4;

            if  fval <= min(f);
 % Check the improvement of the solution and save it in the file
% BestSolution.m
                copyfile(path3,path4);
                disp('Better Solution Found');
                fin = fopen(path3);
                fout = fopen(path4, 'wt');
                while ~feof(fin)
```

```
                    s = fgetl(fin);
                    s = strrep(s, 'ResAnalyzer()', 'BestSolution()');
                    fprintf(fout,'%s\n',s);
                end
                fclose(fin);
                fclose(fout);

            end
        end
        disp('Current Solution');

        ciclo_esterno=ciclo_esterno+1;
        Y_5=Y_4;
if ciclo_esterno == ciclitot
            break
        else
            % independent sequence assignment
            t_ord=0;
            X_5= X_4;
            machine=1;

            for machine = 1: teams
                Partial_Ord2(Y_5,jobs,teams,machine,X_5);
                tic
                !"C:\Program
Files\IBM\ILOG\CPLEX_Studio125\opl\bin\x64_win64\oplrun.exe" -p
"C:\Users\Alice.Consilvio\Desktop\Cplex\Assegnazione_multiarco"
"Partial_Ordering"
                time=toc;

                [X_5] = Parziale();

                t_ord = t_ord+time;

            end
             X_3 = X_5;

            tf(end) = tf(end)+t_ord;
        end
    end
    t_tot = tf(end);
```

FIGURE 65 SOLUTION IMPROVEMENT

```
function F = WriteFtoptoend(Y_0,jobs,teams,machine,X_par)
%% data writing
global path6
global path7

  fID1 = fopen( path7 ,'wt');
  fprintf(fID1,'');
  fclose(fID1);
  copyfile( path6 , path7);
```

```
  fID6 = fopen( path7 ,'at'); % writing of the top to end matrix

  fprintf(fID6,'macchina = %d;',machine); %team not constrained
  fprintf(fID6,'\n');

  fprintf(fID6,'//matrice per fissare le variabili');
  fprintf(fID6,'\n');

fprintf(fID6,'//////////////////////////////////////////////////////////////////
//////////////////////////;\n');
  fprintf(fID6,'Y0 =[[');
   for h= 1:teams
       fprintf(fID6,'');
   for i= 1 : jobs

                 if Y_0(h,i)==1

                      fprintf(fID6,'1 ');

                 else
                    fprintf(fID6,'0 ');

                 end
   end
    fprintf(fID6,']');
         if h<teams
         fprintf(fID6,',\n[');
         else
         fprintf(fID6,'];\n');
         end
   end
  F=zeros(jobs+1,jobs+1,teams);
  fprintf(fID6,'F =[');
  if machine==1
%at the first iteration it writes the top to end sequence matrix according to
%the Y0

      for h= 1:teams

          fprintf(fID6,'[\t');
          %last job counter
          z=1;
         for j = 1:jobs

                   if  Y_0(h,j)==1
                   F(z,j,h)=1;
                   z=j+1;
                   else
                   F(z,j,h)=0;
                   end
          end
             F(z,jobs+1,h)=1;

             k=1;
   for i=1:(jobs+1)
```

```
        for j=1:(jobs+1)
             if j==1
                 fprintf(fID6,'[%d\t',F(i,j,h));
            else
                 fprintf(fID6,'%d\t',F(i,j,h));
             end
       end
       if k==(jobs+1)
           if h==teams
           fprintf(fID6,']]];\n');
           else
           fprintf(fID6,']],\n');
           end
       else
       k=k+1;
       fprintf(fID6,']\n');
       end
   end
      end

  else
%for the next iteration the sequence matrix is equal to the previous determined
%one.
      fprintf(fID6,'[');
      for h= 1:teams
         for i = 1:jobs+1
            fprintf(fID6,'[');

               for j = 1:jobs+1
                   if  X_par(i+(h-1)*(jobs+1),j)==1
                   fprintf(fID6,'1 ');
                   else
                   fprintf(fID6,'0 ');
                   end
               end

             fprintf(fID6,']\n');
         end

         if h<teams
         fprintf(fID6,'],\n[');
         else
         fprintf(fID6,']];\n');
         end
      end
  end

  fclose(fID6);

return
```

FIGURE 66 FUNCTION FOR WRITING THE TOP TO END SEQUENCES MATRIX

```
function y = WriteF(X_3,jobs,teams,iterations)

%% data writing
global path1
global path2

  fID1 = fopen( path2 ,'wt');
   fprintf(fID1,'');
   fclose(fID1);
   copyfile( path1 , path2);

   fID6 = fopen( path2 ,'at');

   fprintf(fID6,'//matrix for fixing some variables \n');
   fprintf(fID6,'\n');

fprintf(fID6,'/////////////////////////////////////////////////////////////////
//////////////////////////  );\n');

if iterations==0;
     fprintf(fID6,'F =[[');
      for h= 1:teams
         for i = 1:jobs+1
            fprintf(fID6,'[');

               for j = 1:jobs+1
                   if  X_3(i+(jobs+1)*(h-1),j)==1
                   fprintf(fID6,'1 ');
                   else
                   fprintf(fID6,'0 ');
                   end
               end

             fprintf(fID6,']\n');
         end

         if h<teams
         fprintf(fID6,'],\n[');
         else
         fprintf(fID6,']];\n');
         end
      end
      else

   pr = randperm(jobs+1, round((jobs+1)*4/5));
   p = zeros(jobs+1,1);
   p(pr) = 1;

   fprintf(fID6,'F =[[');
      for h= 1:teams
         for i = 1:jobs+1
            fprintf(fID6,'[');
             if p(i) == 1
               for j = 1:jobs+1
                   if  X_3(i+(jobs+1)*(h-1),j)==1
                   fprintf(fID6,'1 ');
```

```
                   else
                   fprintf(fID6,'0 ');
                   end
               end
             else
                 for j = 1:jobs+1
                 fprintf(fID6,'2 ');
                 end
             end
             fprintf(fID6,']\n');
         end

         if h<teams
         fprintf(fID6,'],\n[');
         else
         fprintf(fID6,']];\n');
         end
      end
end

  fclose(fID6);

return
```

FIGURE 67 FUNCTION FOR FIXING SOME SEQUENCE VARIABLES