UNIVERSITY OF GENOA

C.I.E.L.I.
Italian Centre of Excellence on Logistics,
Transport and Infrastructures

Thesis for final dissertation
Ph.D. Course: Logistics and Transport

XXXIV cycle

Discrete Event Simulation and Optimization Approaches for the Predictive Maintenance of Railway Infrastructure

Advisors:
Professor Alice Consilvio

Candidate:
Emanuela Baglietto

2nd December 2022

Abstract

This thesis is carried out within the PhD Course in Logistics and Transport at CIELI - Italian Centre of Excellence on Logistics, Transport and Infrastructures, University of Genoa.

In this work, a discrete event simulation and optimization model is created to schedule the predictive maintenance activities.

Nowadays, after a severe decrease of transport demand during the pandemic period, rail public transport is resuming a central role for both freight and passenger transport.

To cope with this increase in demand, to maintain high safety standards and to avoid unnecessary costs, the idea is to switch to predictive maintenance strategy, intervening before an asset failure and when it has reached a certain state of degradation. The degradation and asset future conditions are predicted according to probabilistic models and maintenance deadlines are defined by applying a risk based approach.

The problem is first formulated as a MILP (Mixed Integer Linear Programming) optimization problem and then transformed into a simulation-based optimization problem using the ExtendSim software. Different simulative models are created to take into account the stochastic nature of some variables in real processes.

After the formal description of the models, some real-world applications are presented. Finally, considerations on the proposed approach are reported highlighting limits and challenges in predictive maintenance planning, such as lack of data and the stochastic and complex environment.

Contents

List of figures

List of table

Introduction

The rail transport could help to solve mobility problems in Europe and all over the industrialized world. Today, it is known that such a transportation mode can play an important role in reducing congestion and environmental impact [1], by use of local and metropolitan trains, in urban areas, or via high-speed trains, competing effectively with air transportation, anywhere possible.

For that reason, the strengthening of an integrated, reliable and performant railway transportation system is a priority for the European transport development: therefore, many research programmes, in particular the Shift2Rail Research Programme, have been putting in place to improve interoperability and punctuality, infrastructure reliability enhancement and life cycle cost reduction.

The optimization of the rail maintenance management can support achieving such important strategic role and the realization of safety, security, efficiency and economic goals: in fact, maintenance has great impact on rail transport performances, costs and quality, as well as on ensuring service availability and safety for people and goods.

However, various factors make railway maintenance critical, for example railway infrastructure has the following characteristic:

- Be scarcely redundant (with none or very few path alternatives), meaning that any fault may result into dramatic drops of the system performance and capacity.
- Heterogeneous space-distribution.
- Time constraint, due to rail traffic, climate, fulfilment of fixed operation sequences, etc..

Some of these characteristic are soft constraints, violations can be tolerated if no better choices exist, and some are hard constraints that can never be violated.

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

In the relevant literature, many maintenance approaches have been developed, such as corrective maintenance, performed when a fault occurs, or preventive maintenance which can be subdivided into:

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

Predictive maintenance, which is considered in this thesis, ensures the best reduction of maintenance costs, because maintenance is performed just when necessary and a limited number of monitoring measures is required. The end of predictive railway maintenance is to minimize the probability of the occurrence, of the so-called mission-critical faults during train service, that is those prevents trains for circulating or that can lead to accidents, while keeping maintenance costs as low as possible.

In this framework, considering the acknowledged importance of the problem at the world level, the aim of this thesis is to study the planning of predictive maintenance activities in railway sector through simulative and optimization approaches.

The thesis describes the achievement of a simulation model for scheduling the maintenance of a railway line and some examples of models are reported.

In detail, the following topics will be discussed:

- A description of the evolution of maintenance planning until today, from the maintenance strategies used and the existing planning principles to maintenance triggers (Chapter 1).
- An overview of railway infrastructure maintenance, how it is managed, realized, costs and planning, furthermore it is reported a list of possible defects of the different components of infrastructure (Chapter 2).
- A background on simulation models, elements, classification and structure of simulation models, discrete event simulation and optimization, pros and cons of optimization, and an analysis of the literature on optimization models. (Chapter 3).
- A framework of the railway infrastructure monitoring methodologies, the study of infrastructure degradation and the models in the literature, the risk-based approach model for the deformation of the rail to determine maintenance deadlines (Chapter 4).
- A description of the maintenance scheduling model used: notations, annotations and problem formulation (Chapter 5).
- An analysis of the maintenance simulation model made with the different alternatives in its construction (Chapter 6).
- An analysis of the realization of the models using the Extendsim software. (Chapter 7).
- Applications and tests of the maintenance simulation model created with the analysis of different models related to the real world. (Chapter 8).

1. Maintenance planning

This Chapter describes the framework, including references, assumptions and ideas, related to the concept of risk-based predictive maintenance planning.

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

Maintenance comprehend 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 [2].

Different definitions of the term of maintenance are given [3] as:

a) “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.”
b) “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.”
c) “All supply and repair action taken to keep a force in condition to carry out its mission.”
d) 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.”

During the last decades, the standard repairing-replacing maintenance actions have become preventing activities, with a consequent increase in expected quality standards and in maintenance costs [4].

The management of maintenance activities aim to control the long-term behaviour of single components, considering the overall measures, eventual triggers and possible combinations, are defined in the maintenance policies, also called strategies, concepts, procedures or methods.

Strategies typically used are: reactive/corrective, preventive/planned, condition-based and predictive maintenance.

A strategies only defines the overall measures to be undertaken, how activities are triggered and possibly combined, but to define the details of actual application and realisation are necessary planning principles.

Planning principles defines the planning and organisation of the activities to be performed 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, as for selecting and allocating maintenance activities and interventions and deciding on resource usage etc. Three main planning principles are shown next: risk-based, reliability-centred and evidence-based.

It is possible to differentiate 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).

In this Chapter, a brief definition of strategies of maintenance, of planning principles and levels, of their scopes and of the boundaries between them are provided.

Every maintenance strategies require different triggers that are explain in the last paragraph of this Chapter.

1.1 Types of maintenance strategies

As mentioned before, maintenance strategies can be classified into two big categories [5] [6]:

- Corrective or reactive maintenance,
- Preventive maintenance.

Preventive maintenance, carried out before fault occurs, can be divided into three classes:

- Planned maintenance,
- Condition-based maintenance,
- Predictive maintenance.

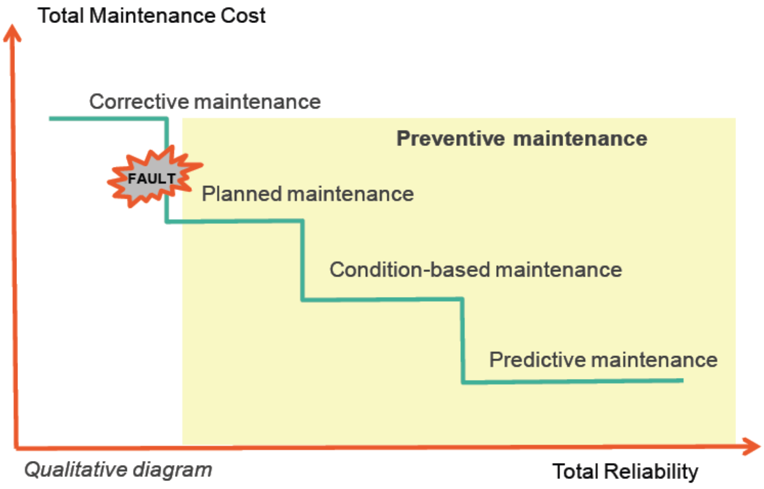


Figure 1: Maintenance strategies [7]

Corrective/Reactive Maintenance: The maintenance tasks are carried out after a fault. It can be defined as the tasks required when a component has failed or worn out, to bring it back to working condition. This maintenance is triggered by an unscheduled event, like as the failure of a component. With this kind of maintenance strategies, the maintenance related costs are usually high due several reasons. Firstly, restoring the component or system mostly has to be done urgently, hence the planning of manpower and spare parts is extremely difficult. Secondly, the failure of a component might cause a large amount of consequential damage to other components 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 vast.

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 fault occurs. Its aim is to reduce the probability of occurrence of failure. It is used to minimize the weaknesses of the corrective maintenance by reducing the probability of occurrence of failure, preventing sudden failures and discovering hidden failure. The tasks corresponding to this strategies of maintenance take place under a schedule, in contrary to the unpredicted case of the corrective maintenance, which follows a random failure models.

Preventive maintenance is carried out to prevent fault during operational time, by maintaining the system during the downtime. Thus, 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 adequate maintenance crew can be available at the planned maintenance execution times.

This type of maintenance can be sub classified into [8]:

- Planned preventive maintenance,
- Condition-based maintenance,
- Predictive maintenance,

which are explained below.

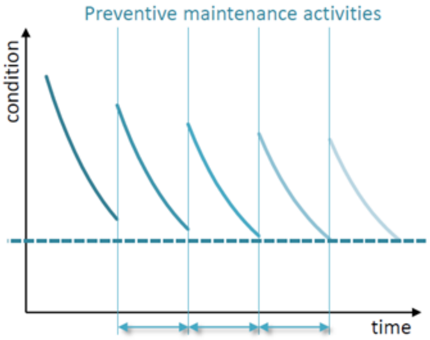


Figure 2: Preventive maintenance

Planned preventive maintenance / Preventive maintenance: It is characterized by regular maintenance activities that are carried out in predefined time intervals. The purpose of preventive maintenance is to extend life-time, to increase asset performance and to avoid unexpected fault. It includes operations planned in anticipation. This type of maintenance has many different alternatives and includes pre-planned actions as replacements, adjustments, inspections, renewals. Some of these activities will result in system downtime, while others can be done while the system is in operation. The main purpose of this type of maintenance is to avoid faults, it follows a scheme of actions and inspection intervals.

There is no assurance that the equipment will continue to work even if it is maintained according to the maintenance plan, although 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. The progress of preventive maintenance is shown in *Figure 2*: in a predefined interval, maintenance activities are executed and the system condition is improved in order to do not fall down a certain condition limit (dashed line).

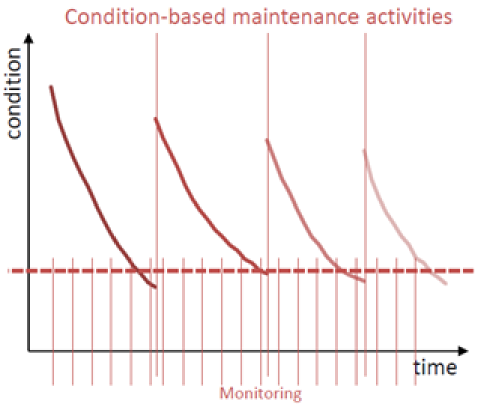


Figure 3: Condition-based maintenance

Condition-based Maintenance (CBM): Failures will be avoided by maintaining the assets when they show signs of decreasing performance or upcoming failure. As shown in *Figure 3*, the assets have to be monitored carefully to see condition changes in time. As soon as monitoring shows that the condition is below the limit values (dashed line), maintenance is requested and executed. With this strategy, the condition is improved. It is also necessary to define limit values for the measures that trigger the need of maintenance intervention. Thus, the time between two condition measures, and the time between maintenance request and execution, has to be considered to ensure punctual maintenance. 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, limit values can be higher and the maintenance effort can be reduced. Today, 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 faults may maximise the life of that component, but its failure may cause significant damages to other parts of the system, it will cause a disruption of the whole process. 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. For that reason, conditional maintenance tasks may not be easily usable for complex systems.

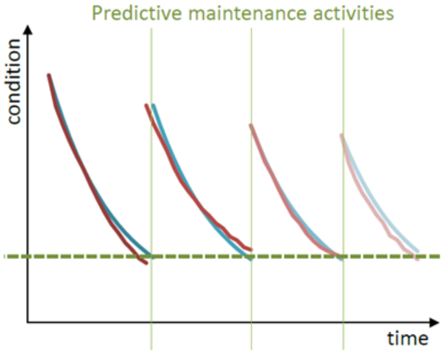


Figure 4: Predictive maintenance

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

Predictive maintenance attempts to predict when fault will occur and to plan maintenance activities consequently. The used techniques are based on statistical analysis and control, to determine at which resource status maintenance will be needed.

This strategies is a relatively new approach because it requires carefully monitoring and an accurate understanding of the deterioration process. The main intention is to predict asset condition in order to plan maintenance in advance. The trend of predictive maintenance is shown in *Figure 4*: The blue line represents the expected system condition, the red line the real condition. If the prediction reaches the condition limit (dashed line), maintenance will be executed to improve the condition.

The real condition can be better or worse than predicted, but with a model close to the real deterioration, the perfect time for maintenance can be approximated. As mentioned before, to apply predictive maintenance, a large understanding of the deterioration process is essential. Historical data are necessary to develop an efficient degradation model. 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 benefit of the procedure is the high planning ability. The upcoming maintenance activities are known in advance and can be scheduled. The resulting maintenance plans could be robust against uncertainties, like unexpected deterioration or unforeseen events which requires extraordinary maintenance.

Combine different strategies is possible and usual, e.g. if quality falls below the defined condition limit, but not later than after a predefined time interval without maintenance is possible request maintenance.

1.2 Maintenance strategies comparison

In this paragraph, the behaviour of the four presented maintenance strategies will be analysed. The strategies will be compared to each other and some advantages and disadvantages will be described.

In *Figure 5*, the maintenance strategies are evaluated with respect to the **planning period**:

- Reactive maintenance must be executed promptly or in short-term.
- Condition-based maintenance must be planned in short-term. Anyway, the planning period depends on the selected trigger and the inspection interval.
- 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 period (medium to long-term) depends on the reliability of the prediction models.

Figure 5: Assessment with respect to the planning period

In *Figure 6*, the maintenance strategies are evaluated with respect to **risks**:

- Reactive maintenance is risky, because in most applications it is too risky to wait until failures occurs. Reactive maintenance is a good options, for example, when some components are redundant.
- Preventive is safe if maintenance intervals and the usage triggers are chosen in a pessimistic and precautionary way.
- 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.
- Predictive maintenance is safe, because of the longer planning period.

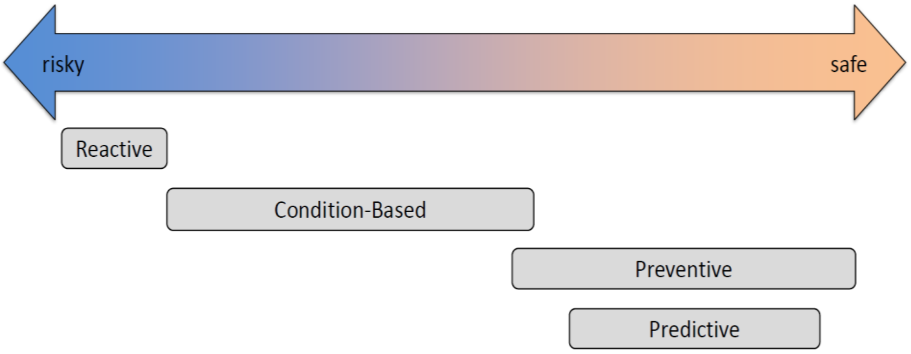


Figure 6: Assessment with respect to risk

In *Figure 7*, the maintenance strategies are evaluated with respect to the **cost effectiveness**:

- Reactive maintenance is in most cases expensive: without adding additional maintenance, deterioration can be fast and the assets have to be replaced frequently.

- Preventive maintenance is in most cases expensive because maintenance intervals and usage triggers should be chosen in a pessimistic way, leading to perform more maintenance and to cost increase.
- Condition-based maintenance has a strong spread of costs. Its effectiveness depends on the selected parameters, costs for inspection/monitoring and costs for maintenance.
- Predictive maintenance can be cost/effective, since maintenance is planned and executed when it is really needed.

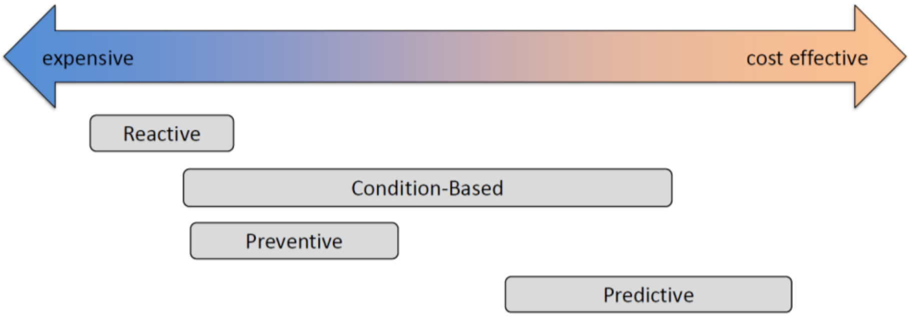


Figure 7: Assessment with respect to cost effectiveness

In *Figure 8*, the maintenance strategies are evaluated with respect to the **amount of information** that the strategy provides regarding the behaviour of the infrastructure system:

- Reactive maintenance provides no additional information, since only breakdowns are observed.
- 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.

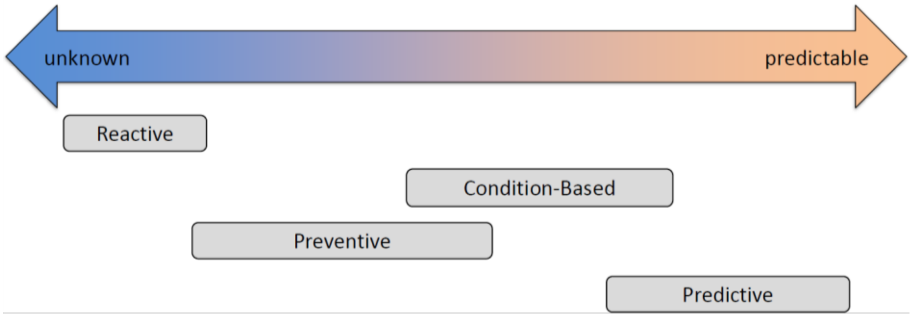


Figure 8: Assessment with respect to the amount of information

In *Figure 9*, the maintenance strategies are evaluated with respect to the implementation effort:

- Reactive maintenance can be intuitively implemented, planning only inspections necessary to detect failures.
- Preventive maintenance has a low implementation effort, since time triggers can be defined based on expert's knowledge.
- 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.

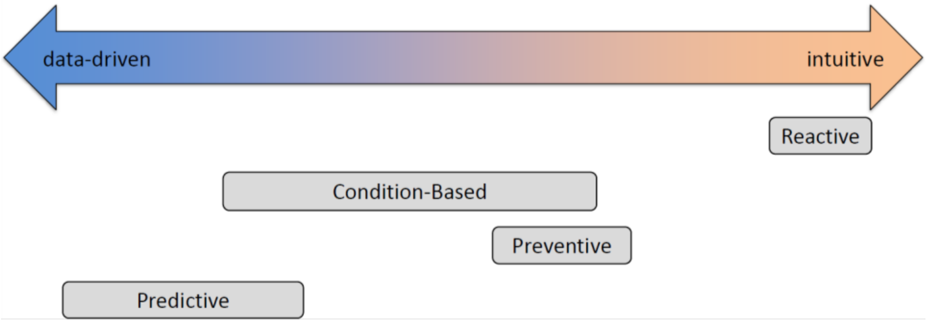


Figure 9: Assessment with respect to the implementation effort

In *Figure 10*, the different assessment criteria of the maintenance strategies are aggregated into one radar chart, taking into consideration:

- 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 system knowledge from unknown to predictable;
- The implementation effort from data-driven to intuitive.

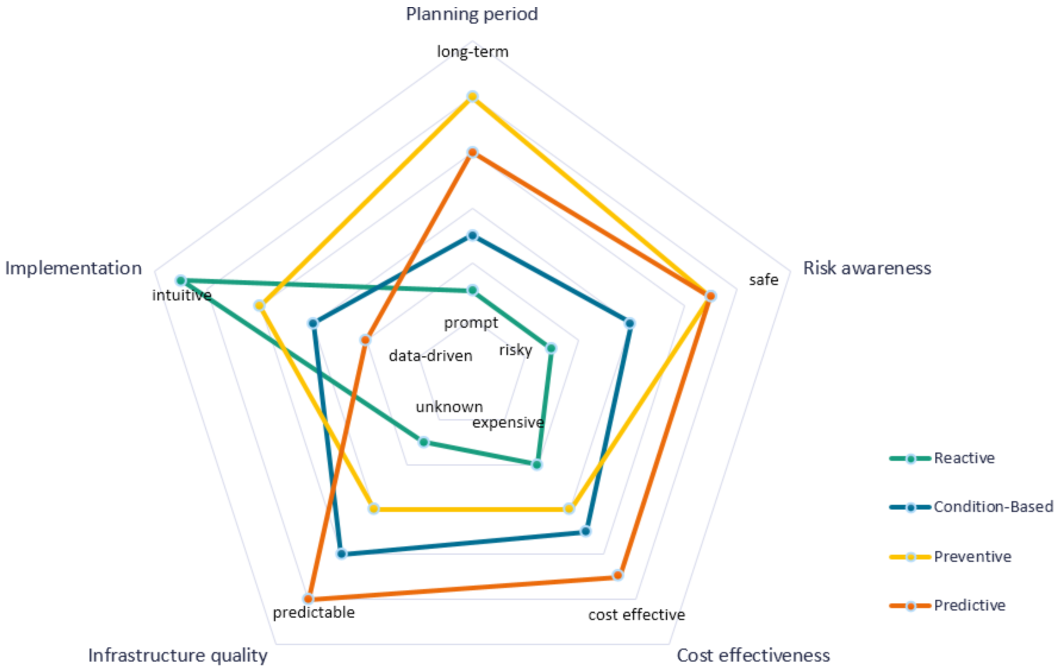


Figure 10: Aggregated assessment of maintenance strategies

Summarising,

- Reactive maintenance is easy to implement, but the planning period is short, risks are not avoided and no information about condition is given. Therefore, reactive maintenance can be used for components with less risk in case of breakdown and less replacement effort.
- Preventive maintenance has a low implementation effort, only the maintenance activities and the time or usage trigger need to be defined. Thereby, maintenance activities can be planned in long-term. But often, this approach is expensive because preventive maintenance is usual done before needed in order to avoid risk. This approach also provides less information about system quality.
- Condition-based maintenance helps to reduce risks and costs, because maintenance is done when necessary. This requires closely monitoring to known the current system 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 fault and deterioration models, future system 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.

In *Table 1* is summarised the confrontation by listing advantages and disadvantages with possible applications of single strategies.

Table 1: Comparison of maintenance strategies

Strategies	**Advantages**	**Disadvantages**	**Possible applications**
Corrective / Reactive Maintenance	Minor implementation labour. No monitor systems are needed.	Unexpected fault and high cost.	Assets whose breakdown has a minor influence on the system performance.
Preventive Maintenance	Activities are planned on long-term. No monitor systems are needed.	Less condition information. Good choice of limit value is indispensable.	Assets with an predictable deterioration, for which monitoring is too expensive.
Condition-Based Maintenance	Extensive knowledge about system condition.	Inspection and monitoring process are necessary.	Assets whose breakdown has an higher influence on the system, but whose deterioration is difficult to predict.
Predictive Maintenance	Activities are planned on long-term. The best maintenance time with respect to risks and costs can be selected.	Monitoring is necessary. Deterioration has to be predictable and 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 strategies and the necessary parameters are usually chosen. For each type of asset it is decided if it is maintained in reactive, predictive, condition-based or preventive way, and the related time intervals, triggers, inspection times etc. are evaluated. Then, it is determined which maintenance activities are associated to the different fault modes.

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

- Risk-Based Maintenance (RBM),
- Reliability-Centred Maintenance (RCM),
- Evidence-Based asset management (EBM).

Risk-Based Maintenance (RBM): Maintenance decisions are made with focus on risk minimisation, 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. Quite the opposite, components with a high fault probability and/or with drastic fault consequences are “risky” or “critical” and will be closely inspected and maintained.

An example of RBM framework is shown in *Figure 11* (from [9]): it starts with data collection, data analysing, definition of fault modes, and risk evaluation for each failure. After, 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. A new plan has to be defined if the proposed measures can't be realized.

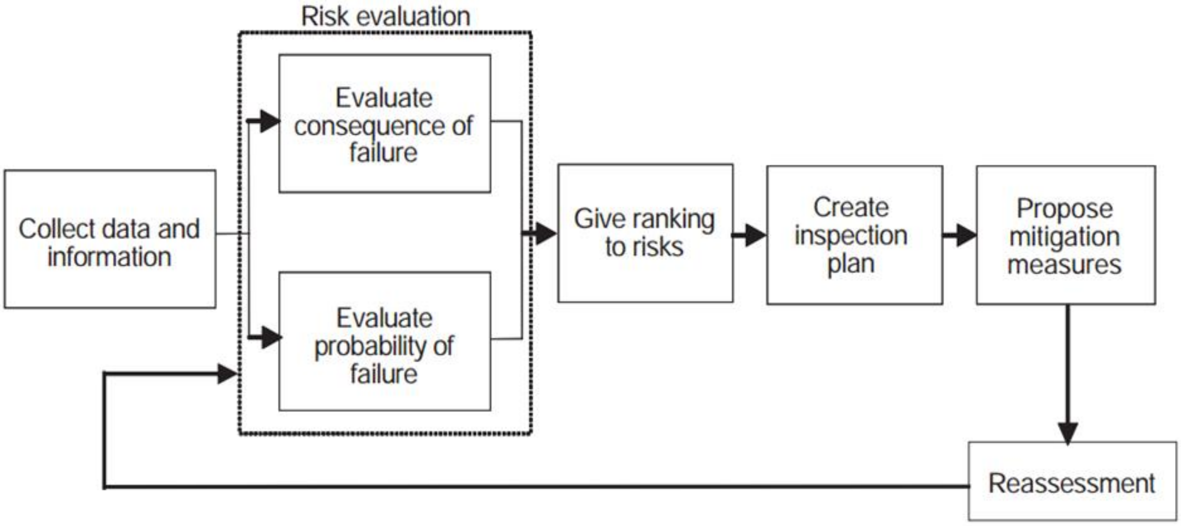


Figure 11: Risk-Based Maintenance

Reliability-Centred Maintenance (RCM): Is focuses on reliability and on ensuring the functionality of system. Thus, 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. In order to preserve the system functionality, the failure modes that affect the system function are identified and prioritised, applicable and effective tasks to control the failure modes are selected, and the operator is able to plan based on a ranking of maintenance tasks.

The seven main questions that guide this planning method are:

a) What are the functions and desired performance standards of each asset?
b) How can each asset fail to fulfil its functions?
c) What are the failure modes for each functional fault?
d) What causes each of the failure modes?
e) What are the consequences of each failure?
f) What can and/or should be done to predict or prevent each failure?
g) What should be done if a suitable proactive task cannot be determined?

Usually, 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. Both, risk-based and reliability-centred maintenance are rather rule and experience-based. The overall performance of these systems depends heavily on the maintainer and can be hardly controlled.

Evidence-Based asset management (EBM): is focuses on data-driven decisions, optimizing clearly defined performance values, and shifting the maintainer function to maintenance plans control. It is important to develop decision support tools on all levels of maintenance planning to apply evidence-based maintenance.

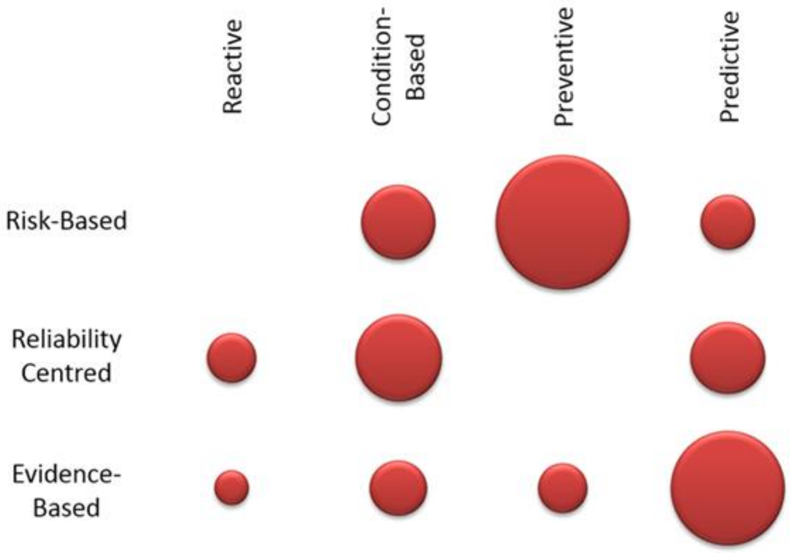


Figure 12: Evaluation of the usage of principles in the different strategies

An evaluation for the usage of planning principles in the different maintenance strategies is shown in *Figure 12*. Since frequent maintenance and inspection leads to a good overall condition that decreases risks, it is expected that risk-based maintenance will focus on preventive and condition-based maintenance. On the other side, condition-based and predictive maintenance are selected in reliability centred maintenance. 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

Maintenance planning (tasks and resources) and scheduling (time) ultimate scope is to minimize the general costs.

The overall planning process in maintenance management is a complex decision-making process, decomposed into three main planning steps, dedicated to solve a specific task: strategic, tactical and operational planning.

- **Strategic planning** evaluates, long-term, the behaviour of asset groups and failure modes under the application of certain maintenance strategies (or methods, or policies).
- **Tactical planning** is directly related to concrete mid-term maintenance activities, according to selected maintenance strategies. Alignment between maintenance and traffic operation is an issue at this level, e.g. possession windows have to be selected and shifted.
- **Operational planning** considers the actual short-term implementation of single maintenance activities, the scheduling of resources like machinery, staff, material, spare parts.

Considering the time horizon, maintenance plan activities can be grouped into two classes:

- **Stationary plans** (long term) defines static rules which do not change with time in an infinite time horizon.
- **Dynamic plans** (short term) considers unexpected tasks to carry out corrective operations and other non-corrective planned operations in a finite planning horizon, which changes continuously. These models group pre-planned operations with unexpected ones in order to rationalize all maintenance activities.

Certainty influences the planning making 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: a deterministic maintenance scheduling can be settled.

- **Stochastic problems** are those where the timing and outcome of the maintenance and replacement actions depend on chance.

Maintenance operations are predicted from predictive models under a probabilistic environment, this indicate 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 described before, the presented maintenance strategies require different maintenance triggers:

a) **Condition trigger:** For condition-based and reactive maintenance. These triggers can be: the achievement of a certain measure value, a dropping below a needed quality level or the occurrence of a fault.

b) **Break-down trigger:** The trigger value is the “break-down”. Only usable for uncritical assets in terms of safety and reliability.

c) **Time trigger:** Preventive maintenance is repeated in predefined time intervals.

d) **Usage Trigger:** Maintenance is performed when the usage limit is reached or when its exceeding is expected.

e) **Event trigger:** Some maintenance activities are triggered by external events, e.g. winter maintenance is triggered by sub-zero temperatures.

The selection of the maintenance strategies, the definition of trigger values and the selection of resulting maintenance is part of Strategic Planning.

2. Railway maintenance

This Chapter describes the framework related to the concept of railway maintenance. First of all, the definition of a railway infrastructure needs to be described. In the second paragraph is explain infrastructure possession time and train services. Then a presentation about costs and planning of railway maintenance is done. The last two paragraphs show the structure of the railway infrastructure and the main defects it may have.

2.1 Railway infrastructure and maintenance

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. Thus, the term infrastructure covers all the assets that are used for train circulation [10]. These assets are:

- Tracks including rails, fastenings, sleepers, switches and crossings, ballast and platform. It can be structured into superstructure and subgrade
- Switches and crossings,
- Tunnels
- Bridges and viaducts: including pillars, foundations, decks
- Energy supply installations: including catenary and support third rail, control equipment and substations
- Safety, signalling and telecommunication equipment: including fixed signals, train control equipment, track circuits, 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 reliability and safety. The primary objectives of rail infrastructure are:

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

Hence, the management of infrastructure should aim to:

- To increase and maintain high level of safety,
- To reduce costs, without decreasing safety standards,
- To improve organization, equipment, materials and staff’s qualification.

Appropriate maintenance of infrastructure is important to achieve the above goals.

The maintenance of railway infrastructure is a complicated and expensive task, it 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 complex, that makes difficult to plan and execute the maintenance tasks.

Factors such as topography or climatic conditions, geographical and geological features need to be considered when planning for maintenance.

Also, the availability of the track for maintenance (on possession for maintenance) without interrupting train services is a fundamental issue to be considered when planning the maintenance tasks to be done.

Regarding railway infrastructure maintenance, a research conducted among maintenance agencies [11] shown that most maintenance activities are related to the track system and are concentrated in: rail track maintenance, ballast maintenance, track geometry maintenance, tie and fastener maintenance, track inspection.

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

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

Routine (spot) maintenance works include inspections and small repairs of the local irregularities carried out manually or using small machines. These are jobs that take a short time to be performed and are done frequently. For example: switch inspections, switch lubrication, rectifying track gauge, maintenance at level crossings, tamping using vibrating compactors or tamping tines, etc. [12].

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, they are performed at regular intervals. There are two types of inspections: visual inspection and condition measurements.

Project (systematic maintenance) consist of a large amount of work that involve separate planning. These activities are carried out with heavy track maintenance machines (e.g. tamping machines, rail-grinding machines, ballast cleaners, ballast regulators) approximately once per a couple of years.

There are also other kinds of works on the railway infrastructure, these are renewal and new constructions.

2.2 Infrastructure possession time and train services

Some restrictions, in term of space and time, for maintenance works is regulated by the possession of the infrastructure, meaning the use of the infrastructure by maintenance operations. The possession can be:

- **Partial**, when maintenance and trains share the infrastructure, implying safety risks to be properly assessed.
- **Privative**, when the maintenance takes full possession of it. Single term "possession" is usually identified with maintenance operations, and is a synonymous of possession for maintenance. The term possession for maintenance stands for full possession of the infrastructure by maintenance operations.

Possession can be classified into three categories [13], 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), it depends on the extension of the available time-windows defined by the ride 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 in the weekend (may be reduced, re-scheduled or re-routed) respect to labour-day services.

- **Daytime possession**, the lack of available time-windows makes this possession to be focused on operations that cannot be postponed for latter.

During track possessions the train operation is likely to be influenced: the timetable, the crew schedules and the rolling stock are adjusted and an operational timetable, an operational crew planning is made and an operational rolling stock planning. For a more exhaustive description of the planning process, refer to Huisman et al. [14].

As a rule, the main lines, or the ones intensively used during the day, are maintained at night. Taking out of these lines in the daytime would cause severe disruptions for the railway traffic. Maintenance is fundamental for ensuring safety, overall capacity utilization, train punctuality and lower costs for railways.

The requirement for more maintenance and the grow of infrastructure possession time to carry it out, is discordant with the increase of the infrastructure use by train services to satisfy the demand. The infrastructure deterioration is accelerated by longer operating time, higher number of services and trains and raise and the annual traffic load; this consequences into an increase of the number, frequency and severity of the needed maintenance operations. Furthermore, the European directive concerning infrastructure charges and capacity allocation [15] defines a regulatory and organizational framework to an optimization of the railway infrastructure.

Lately, the Fourth Railway Package Technical Pillar, comprising the ERA Regulation 2016/796, the Interoperability Directive 2016/797 [16] and the Rail Safety Directive 2016/798 [17], published in the Official Journal of the EU on May 26 2016 and entered into force on 15 June 2016, talks much about maintenance, though with a relevant attention on rolling stock, but with large references also to the infrastructure. Particularly, 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, putting in evidence their interrelation.

Maintenance productivity is directly related to the available time-windows of train services: due to a limited availability of time-windows and to a growing rail transport demand, advance managerial techniques and procedures are required:

- Traffic management: uses infrastructure information to check and guarantee the operation of planned train services.
- Maintenance management: in the long run pitches at increasing infrastructure availability, but in the short term its immediate effect is to restrict train services by occupying time windows for interventions.

Since traffic management and maintenance planning can be seen as concurring topics, ideally both planning and decision-making processes should be done together, avoiding prioritizing one over the other.

In the practice such an idealised coordinated approach is rarely implemented, because of several organisational, historical and practical reasons. In the scientific literature it is possible to find recent attempts to formulate and solve the problem of a computational handling of a fully integrated planning approach such as in [18] and in [19].

Traffic management has to be considered in a lightly broader context than the usual meaning as real-time traffic control and operation: a broader view allows to identify more coordinative planning activities than in the real-time controlling process only, where only a few restricted options are available to react on issues, and actual "proactive" planning capabilities are not given.

There are two ways for coordinating planning [20]:

- The requirements from traffic management affect and "command" 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 realized in this thesis.
- Traffic management models comprehend information coming from maintenance, and consider them for decision-making.

The primary approach to concurring planning in mathematical modelling is to incorporate elements of one aspect into the model of the another aspect, both in the definition of the decisions or solutions space, or as a component of the objective functions, or as additional constraints to be respected.

For example, in this thesis the model presented use the ideas behind this approach with the respect of a constraint: in a model for the planning of mid-term maintenance activities, a constraint on the possible time windows can be integrated in order to be aligned with traffic management.

2.3 Railway maintenance costs and planning

Maintenance costs of railway are influenced by different factors as:

- Raise of manpower costs of maintenance staff,
- Raise of maintenance actions due to greater quality standards,
- Raise of management costs.

Thanks to the interest to reduce maintenance costs, preventing maintenance is an area of increasing importance. Anyway, also in a preventive maintenance context, corrective maintenance tasks will be never avoided because of unexpected failures.

To avoid an overly strictly maintenance model, the four maintenance types can be combined under a scheduling scheme, to minimize the idling time. This way of proceeding can efficiently reduce costs but needs a careful optimization of the replanning of the maintenance schedule.

Maintenance costs in some sectors are presented by Cross [21] as a percentage of the total operating costs.

With regard to railway maintenance costs, Zoeteman [22], [23] reports that tracks and switches use more than 50% of total maintenance costs and 75% of renewal costs due to their elevated utilization, rapid deterioration pattern and high installation cost. Indeed, the most important equipment of the railway infrastructure are tracks and switches, since they are often traffic disruption and sources of failure.

Track maintenance costs and planning are much influenced by track possession and access windows. An average work window of 4 hours is most usual, of which 50% is expended for displacements to the work site. This indicates that a planning regarding access, to coordinate maintenance teams and resources, is necessary in order to keep costs down.

In short, maintenance costs are determined by the following points:

- **Infrastructure maintenance activities**: the costs corresponding can vary with system configuration and the technology associated to the infrastructure. They are correlated at the “Maintenance Demand” defined as the level of resources to give an tolerable asset state level.
- **Resources costs**: material, labour, equipment, administrative and organizational.
- **Work window costs**: it represent the cost associated to the time-window established to make maintenance activities when the infrastructure possession is detracted from railway circulation.
- **Inspection and maintenance technologies costs**: They are influenced by the inspection technology and the maintenance techniques employed to make the operations.
- **Inspection and maintenance policies costs**: They are influenced by the level of maintenance applied and by the number of inspections.
- **Operating costs**: They represent the costs associated to operational optimisation of maintenance operations.
- **Indirect cost**: They stand for the costs not directly involved in the maintenance tasks.

It is important that the track maintenance are scheduled in an effective and efficient way because of very high maintenance costs. This means short term planning, like daily scheduling of the activities, but also the medium to long term planning.

The railway infrastructure maintenance is very difficult to plan and involves high costs, therefore is necessary developing tools which help the maintenance planners to realize optimal maintenance plans. This question is an interesting research subject because the results could help to get better the quality of the rail infrastructure. For the above reason, this study is focused on railway infrastructure maintenance. The mathematical techniques and models developed for planning railway maintenance can also be referred for maintenance planning in other sectors as well.

2.4 Railway infrastructure components

In this paragraph, the structure of railway infrastructure is illustrated. As shown in *Figure 13*, it is possible to recognize the **superstructure** and the **substructure** of the railway infrastructure.

The **superstructure**, which distributes and supports train loads and is subject to periodical replacement and maintenance, it is composed by the track and the track bed.

The **track** include:

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

The **track bed** include:

- The **ballast**, composing of crushed stone or gravel. The ballast ensure adequate load distribution, the damping of most of the train vibrations and fast drainage of rainwater.
- The **sub-ballast**, composing of sand and gravel. It protects the upper layer of the subgrade from the penetration of ballast stones, while at the same time concurs to further distributing external loads and guaranteeing the quick drainage of rainwater.

The **substructure**, where train load is conveyed, after adequate distribution in the superstructure, and which should not be subjected to interventions during periodical maintenance of the railway track, include:

- The **formation layer**, employed whenever the base ground material is not of good quality.
- The **base** or **subgrade**, which can be onsite soil or soil transported to the site in case of embankment.

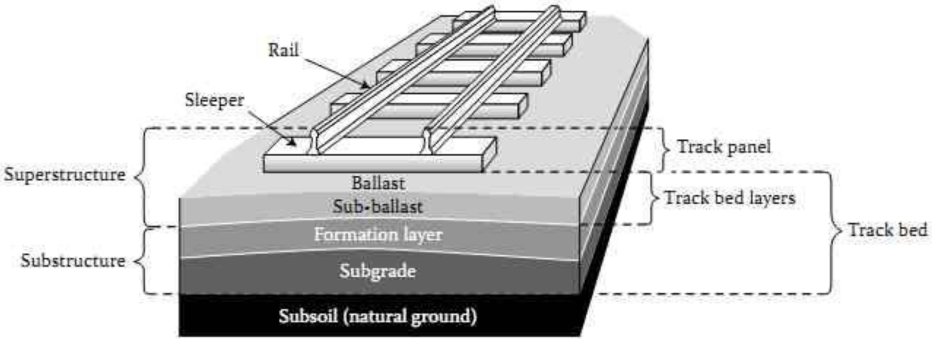


Figure 13: Superstructure and subgrade of railway

The track normally stands on ballast which gives a flexible support. It is called as ballasted track. It is possible that the track is stand by a concrete slab in lieu of ballast. In this case, it is referred as slab track and the support is inflexible. A slab track is used in certain railways e.g. the Japanese and the German, but in most of the tracks worldwide, a ballasted track is still highly use, as it assures flexibility and lower construction cost, and at the same time offering a acceptable transverse resistance, even at high speeds.

The problem of noise should not be forgotten: noise is much greater with the track on concrete slab than with the track on ballast.

The choice between ballasted and non-ballasted track should consider construction cost (much bigger for non-ballasted track), maintenance cost (much bigger for ballasted track), together with technical requirements. Both resolutions have vantages and disadvantages.

2.5 Railway defects

The maintenance of railway track is an important part of total railway infrastructure costs (about 40% of total maintenance cost), and it is an expensive and complex operation. The condition of track is related to many factors such as the age and features of the elements, the track geometry, geology and topography, atmospheric conditions and supporting loads. Track maintenance still relies on the skills of human operators and on rules set long time before (preventive maintenance) in addition to the execution of on-call corrective operation when there are defects in the system.

The increasing saturation of the track capacity requires intensified maintenance to be accommodated during shorter time windows. Consequently, the work is mostly realized in nightlight conditions and under pressure, increasing the risk of staff accidents.

This circumstances requires better maintenance management based on monitoring of the track condition, automating planning management and especially monitoring the evolution of the parameters that determine the track condition for predictive maintenance and risk analysis.

This system would allow develop the maintenance management model founded on corrective/preventive maintenance into a model founded on conditions/predictions, helping to achieve optimal maintenance plans that reduce the maintenance costs, guarantee a good safety margin and prevent fast degradation of track quality.

The major defects of railway infrastructure can be shared in:

- Track defects
- Rail defects
- Sleepers defects
- Fastening defects
- Switches defects
- Ballast defects

2.5.1. Track defects

Track defects are defined as the divergent of the actual shape from the theoretic values of geometrical characteristics of the track. Track defects are consequence of train circulation, they are of a geometric and macroscopic nature and usually they are corrected by track maintenance.

In detailed, geometry track defects comprehend [24]:

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

Longitudinal defect (LD): is defined as the difference between the theoretic Z_{th} and the real value Z of track elevation and is given by the formula:

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

The longitudinal defect is used for illustrating the effect of the vertical loads on track quality and it is the principal coefficient (with the transverse defect which accompanies longitudinal defects) in defining the magnitude of the track maintenance cost.

Transverse defect (TD): is defined as the difference between the theoretic 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 Z_{texth} rails.

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

This is usually bigger where the railroad is curved, for straight parts of track layouts, where curvature is null, the transverse defect is the difference of elevation between internal and external rail: $Z_{int} - Z_{ext}$.

Horizontal defect (HD): is defined as the horizontal deflection of real position of the track from its theoretic position. The horizontal defect is conditional on the transverse track effects (more than the two precedent kind of defects) and on the particularities and characteristics of the rolling stock.

Gauge deviations: some track gauge deviation influenced by the particularities of the rolling stock and the mechanical properties of track materials, are permissible. Gauge values tolerable for standard gauge tracks are established for each track line.

Track twist (TW): Along rectilinear and circular sections (where cant is constant), four point of the track laying on two transverse sections must laid 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 consecutive transverse sections of the track, distant Δl each other, track twist is defined as the variations of the transverse defect TD per unit length Δl.

$$TW = \frac{TD_{I+1} - TD_I}{\Delta l}$$

The risk of derailment is avoided when the real value of twist is smaller than its limit value causing derailment, which depends on speed and on the type of the track equipment and of rolling stock.

2.5.2. Rail defects

The rail endure from effort that can produce defects and may conduce it to fault. The total stresses developed in the rail are the addition of:

- Stresses at the wheel-rail contact,
- Stresses coming from rail flexion on the ballast,
- Stresses coming from flexion of the rail head on the web,
- Stresses coming from thermal effect,
- Plastic stresses, residual in the rail after the removal of external loads.

In this context is possible to define the term "fatigue", it is the progressively decrease of the mechanical strength of the rail due to replicated stressed, once the fatigue limit is reach, the rail is lead to failure.

The consequences of these stresses can be:

- **Plastic deformation:** The rail endure high stresses of the train traffic. If those stresses are higher than the elastic limit, a flange can occur in the rail head, because under this flange, the stress limit is exceeded.
- **Rail wear:** The traffic weight produces the rail wear that influence the rail profile. There are two kind of rail wear: Vertical wear that decreases the rail section and in consequence the rail resistance, Lateral wear that, in addition to the previous two, influence also the gauge of the track. The highest admissible vertical and lateral wear of the rail depends from the maximum train speed and the traffic load.
- **Surface defects**, that can be divided in:
    - Short-pitch corrugations. Their origin is train traffic and they involve of corrugations with a wavelength λ of 3-8 cm. They can produce many negative effects: high frequency oscillation of the track, including resonance, and leading to greater rail stresses, concrete sleeper fatigue with cracking in the rail bottom area, slackening of fastenings,

accelerated wear of clips and pads, premature failure of ballast and the subgrade, and increase of noise level.

- Long-pitch corrugations. They have wavelengths λ of 8-30 cm and happen mainly on the inside rails of curves having a ray of 600 m and smaller. This type of defect is the most ordinary on suburban and underground railways.

  These defects are detected either visually or by appropriate recording equipment. They are fixed by passage of special equipment, which grinds and smooths the rail.

There are four approaches to removed rail defect

1. Rail weld recharge,
2. Rail grinding,
3. Rail replacement,
4. Rail tamping.

Rail weld recharge is a method for cheap repair of discrete defects on the running surface of rail. The results obtained are very much depending on the competence of the maintainer.

Repair welding is the most cost-effective technique for fix defects of the tread and guide surfaces of tracks and switches.

It's possible to used two methods for welding repair: manual welding and union welding.

Rail grinding is the effective maintenance practice to monitor the effects of rolling contact fatigue, renew profile, and maximise value from the rail asset.

The considerable benefit on investment from rail grinding comprises:

- Extended rail life,
- Fuel savings,
- Greater track component life,
- Greater ride quality and passenger comfort,
- Lessen surfacing cycles,
- Lessen wear on rolling stock,
- Raised axle loads,
- Raised train speeds.

An appropriate rail grinding scheduling is fundamental to a maintenance plan. Rail-grinding equipment could be mounted on a single self-propelled vehicle or on a dedicated rail grinding train. These machines have been in use in Europe since the early 1990s, there are different types of machine to regulate the rail profile to theoretical profile according to the extension of the grinding work.

When the rail defects are very serious and the rail profile cannot be fixing by other process, it is required the rail replacement in the damage section. In this occasion, it is always required to employ welding to join the new profile of rail with the existing one. Then, some local grinding on the join is necessary to achieve appropriate rail profile.

Rail tamping can renew the ballast, the initial sleeper’s position, the vertical rail deviation and the rail geometry. The tampering activity, achieve by introducing vibrating blades in the ballast. The tamping is executed using a train named tamper train. Furthermore, the renovation of the ballast could be performed by refilling stones.

2.5.3. Sleepers defects

For the sleeper defects it possible note that their failures depend on the material of sleepers: steel or concrete.

Defects in steel sleepers derive from sensibility to chemical attack especially in industrial and coastal areas. The steel sleepers in a non-rude environment have a lifetime of about 50 years. In chemical rude environment this lifetime could be only a few years, because of the accelerate corrosion.

Defect in concrete sleepers are unusual, for of its severe quality control and testing in manufacturing. Usually, defect in concrete sleepers are consequences of defect in the accessories, like loss of fastening, pad or other defects that obligate concrete sleepers to work out of design conditions. The estimated lifetime for the concrete sleepers are about 50 years, which is the time form track total restructuration.

The sleepers generally have no more maintenance than replace them when they are exhausted.

2.5.4. Fastening defects

The fastenings are the components worn to ensure the anchorage of the rail to the sleeper, to keep the correct longitudinal and transversal position of sleepers and, if required, to guarantee the electrical insulation.

The most common fastening defects is a reduction in the link between the rail and sleepers, with the risk that it is no longer ensured the correct track gauge. Furthermore, fastenings could be subject to another type of serious defect, which is that of fatigue break.

Defects in fastening be determined by the kind:

- **Rigid fastenings** present two type of defects: broken bolts or nails and gap between nails and rail. Both defects are serious and can cause a derailment.
- **Elastic fastenings**: Defects in this type of fastening may cause break or loosing of a fastening components.

The defects of fastenings are normally isolated and must be repaired manually. The interventions for fastenings consist in the change of the fastening and for timber sleepers in filling the screw hole with synthetic material. However, elastic fastenings can be change in order to correct horizontal rail deviations.

2.5.5. Switches defects

A railroad switch is a mechanical installation allowing trains to be moved from one track to another, such as at a railway node or where a spur or siding branches off.

The most typical and frequent defect of switches, is relative to the low lubrication of the mechanical elements that compose them. This kind of defect causes a malfunction of the switch, with result slowdown in the operation of deviation.

A second defect consists of the aging and wear of mechanical components.

2.5.6. Ballast defects

The ballast defects are caused to the traffic loads, which in time origin the readjustment of the stone changing the characteristic of the ballast. With this readjustment, the ballast loses his original characteristics, raise its flakiness index, reduces its granulomere composition and smooths partially the hard corners of the stones. These modifies of characteristics worsen its operation with the traffic load with respect to stress distribution, rainwater retention and vibration attenuation.

The ballast maintenance operations can be listed as below:

- Renovation of its initial geometry or curb profile: The renovation of the ballast geometry is executed 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.
- Renovation of ballast under the sleepers: The vibrations of trains slowly eject the ballast below the sleepers; the ballast tamper machine must restore it back with its blades.
- Clean the ballast from broken stones using the ballast cleaning machine.

Furthermore, ballast degradation is a tricky mechanism that comprehend different events: broken edges, stones displacement and blending with fine soils and vegetation, which influence the drainage, reduce the stiffness and rush the stone movements.

3. Proposed methodology

The methodology proposed in this thesis work involves the use of a simulation model. What simulation models are, how they are made, how they are classified and the pros and cons of simulation models are explained below.

The term simulation means the reproduction of a system's behaviour.

With “simulation” we mean the activity of replicating an existing reality or a reality to be designed by means of suitable models, in order to study, in the first case, the effects of possible interventions or events in some way foreseeable, or, in the second, to evaluate several possible alternative design choices.

To simulate the behaviour of a system it is necessary to build a simulation model. The simulation model must allow to describe the operations of a system and how they must be simulated. The model must be complex enough to represent real behaviour, but it must nevertheless remain as simple as possible. The limits of use of the model must also be clear.

Simulation is a very flexible tool: it can be used to study most existing systems. It is impossible to enumerate all the specific areas in which the simulation can be used. Examples of some important typical categories of applications in which simulation is used are given below:

- Design and definition of the operating procedures of a service system
- Management of inventory systems
- Design and definition of the operating procedures of production systems
- Design and operation of the distribution systems
- Financial risk analysis
- Project management.

The simulation models are analysed by numerical methods rather than analytical methods. Analytical methods use the deductive reasoning of mathematics to solve the model. For example, differential calculus can be employed to compute the minimum-cost policy for some inventory models. Numerical methods employ computational procedures to solve mathematical models. In the case of simulation models, which employ numerical methods, models are “run” rather than solved, that is, an artificial history of the system is generated from the model assumptions, and observations are collected to be analysed and to estimate the true system performance measures. Real-world simulation models are rather large, and the amount of data stored and manipulated is vast, so such runs are usually conducted with by using a computer.

This Chapter describes the framework related to simulation models: what elements make it up, how the models are classified, their structure, pros and cons of simulation model. Furthermore, the discrete event simulation model and simulation of the optimization problem which are the object of study of this thesis are introduced.

3.1 Elements of a simulation model

Some elements that form a simulation model are listed below.

- State variables

A system is described at each instant of time by a set of variables which are called state variables. For example, with reference to a queue system, the number of users present in the system in a certain instant of time is a status variable. There are discrete systems in which the variables change instantaneously at precise instants of time that are finite or belonging to a countable set and continuous systems in which the variables vary continuously over time.

- Events

An event is defined as any instantaneous occurrence that changes the value of at least one of the state variables. The arrival of a user to a queue system is an event, as is the completion of a service. There are events external to the system (exogenous events) and internal events (endogenous events). For example, the start of the service to a user who is queued in a queue system is an endogenous event, because it is internal to the system; the arrival of a user to a queue system is an exogenous event.

- Entities and attributes

Entities are individual elements of the system that need to be defined. An example of an entity is a user at a queue system, or it can be a server. In the first case the entity flows within the system and is called a dynamic entity, in the second case we speak of a static entity. Entities can be characterized by attributes that provide a data value assigned to the entity. For example, in a single-server queue system where the entities are the server and the users, an attribute of a "user" entity could be its time of arrival to the system, while the server is characterized by the "status" attribute which it assumes the value of "free" or "occupied". Some attributes may be of interest in some cases and not in others.

Entities can be grouped into classes that are sets of entities of the same type, that is, entities can be grouped based on attributes. For example, if we consider male and female people as users of a queue system, since the entities are people, they can be grouped into two classes based on the "gender" attribute.

- Resources

Resources are elements of the system that provide a service to entities. An entity can request one or more resource units and if this is not available the entity will have to put itself in a queue waiting for it to become available, or take another action.

If, on the other hand, the resource is available, it is "captured" by the entity, "held" for the necessary time and then "released". An example of a resource could be given by a worker who oversees the operation of a machine that cannot function without the worker; when the use of this machine is required, if the resource "worker" is available then the execution of the work is carried out otherwise it is expected that there is a resource (worker) available.

The worker will be "detained" for the duration of the execution of the work and then "released".

- Activities and delays

An activity is an operation whose duration is known a priori at the beginning of the execution of the activity itself. This duration can be a constant, a random value generated by a probability distribution, or given as input or calculated on the basis of other events that occur in the system. An example is the service time in a queue system.

A delay is a period of time of indefinite duration that is determined by the conditions of the system itself. The time that an entity spends at a queue before a resource it needs is free is a delay.

3.2 Classification of simulation models

Simulation models can be classified according to various criteria. A first classification is between:

- Continuous event models, in which the value of the state variables varies continuously;
- Discrete event models, in which the value of the state variables changes in defined instants of time.

A second classification is between:

- Continuous time models, the state of the system is always defined in each t;
- Discrete time models, the state of the system is defined in discrete time instants: t, t+1, ...

Another classification is between:

- Static models, which represent a system at a particular instant in time;
- Dynamic models, which represent a system evolving over time.

Finally, there are:

- Deterministic models, which do not contain probabilistic components;
- Stochastic models, which present elements subject to randomness.

It should be noted right now that the choice of a continuous or discrete model to be used is not necessarily required by the type of system; for example, you can decide to build a discrete model for a continuous system depending on the study you want to carry out. A typical example is the case in which in representing a railway line, the position of the train can be described by a real variable that provides the distance from the station of origin, or by binary variables that describe the free-busy state of each of the sections of block into which the line is divided.

In this thesis we will consider discrete, dynamic, deterministic simulation models and discrete, dynamic, stochastic simulation models, the latter being commonly called discrete event simulation models.

3.3 Discrete event simulation

In Discrete Event Simulation (DES) the system is represented, in its evolution over time, with variables that change their value instantly in defined instants of time belonging to a countable set. These moments are those in which events happen. It is clear that, being these models of a dynamic nature, it is necessary to record, and keep memory, of the (simulated) time that proceeds.

In particular, it will be necessary to define a mechanism for time advance to make the simulated time proceed from one value to another. The variable that in a simulation model provides the current value of the simulated time is called "simulation clock", and there are two ways to define its progress: advancement of time to the next event or advancement of time in predetermined increments.

The first is the most widespread and is the one we will refer to. In this case, the "simulation clock" is initialized to zero and is advanced to the time of the occurrence of the first of the future events; then the system is updated taking into account the event that happened, the times of future events are updated and the procedure is iterated. on the other hand, advancement in predetermined increments, periods of inactivity are not considered.

For example, in the queue system where events are the arrival of a customer and the conclusion of a service; both are events because they cause the value of some state variable to change. In this case, the time advance mechanism follows the occurrence of these two events in the chronological order in which they occur.

Formally the set of the N admissible events, namely the event space, is denoted as $E = \{e_1, e_2, \dots, e_N\}$, while the set of states, that is the state space, is indicated as $Y = \{y_1, y_2, \dots, y_M\}$.

The system dynamics can be described by the general state equation:

$$y_t = \delta(y_{t-1}, e_t), \qquad t = 0,1,2 \dots \tag{3.3.1}$$

which gives the state $y_t \in Y$ as a function of the occurred event $e_t \in E$ and the previous state y_{t-1}.

In DES the time variable is updated only when an event occurs. DES are usually too complex to be studied analytically, then they are often simulated by means of simulation models as in this case.

3.4 Simulation problem structure

This paragraph shows a list that describes the succession of the various phases that characterize a study based on simulation.

1. Problem analysis

It consists of understanding the problem by trying to understand what the purposes of the study are, identifying what are the essential components and what performance measures are of interest. If a version of the system is already operational, this system must be observed to deduce its fundamental characteristics.

2. Formulation of the simulation model

Since we are dealing with stochastic systems, to formulate a simulation model it is necessary to know the probability distributions of the quantities of interest. Indeed, to generate various representative scenarios of how a system works, it is essential that a simulation generate random observations from these distributions. For example, in queue systems it is required the distribution of inter-arrival times and service times; in inventory management, probability distributions of product demand and the distribution of time between an order and receipt of the goods are necessary; in the management of production systems with machines that can fail, it will be necessary to know the distribution of the time until a machine fails and the distribution of repair times. Generally, it is only possible to estimate these distributions by deriving them, for example, from the observation of already existing similar systems. If from the analysis of the data it is observed that the shape of this distribution approximates a standard type distribution, the standard theoretical distribution can be used by carrying out a statistical test to verify whether the data can be represented well by means of that probability distribution. If there are no similar systems from which to obtain observable data, other sources of information must be used: machine specifications, machine instruction manuals, experimental studies, etc. Building a simulation model is a complex process.

In particular, referring to discrete event simulation, the construction of a model involves the following phases:

i) Definition of state variables.
ii) Identification of the values that can be assumed by the state variables.
iii) Identification of possible events that cause the state of the system to change.
iv) Realization of a simulated time measurement, “simulation clock”, which records the flow of simulated time.
v) Implementation of a method for randomly generating events.
vi) Identification of state transitions generated by events.

3. Analysis of the simulation model

In the model analysis phase, the accuracy of the model created must be verified in different ways. Usually this is done through a conceptual analysis of the model in order to highlight any errors and/or omissions.

4. Choosing the software and building a program

After building the model, it must be translated into a program. Various tools can be used for this:

i) General purpose languages: Languages like C ++, FORTRAN, etc. They were widely used at the birth of simulation but require a lot of programming time and therefore it is generally preferred to use specific languages for simulation.

ii) General simulation languages: They provide many features necessary to create a simulation model thus reducing the construction time; examples are MODSIM, GPSS, SIMSCRIPT, etc. Although less flexible than "general purpose" languages, they are the most natural way to build a simulation model.

iii) Spreadsheets: When you have small problems, you can also use spreadsheets, such as Excel, to get an idea of how a system works.

iv) Simulators: They are application oriented simulation packages. There are numerous interactive software packages for simulation such as ARENA, WITNESS, EXTEND, MICRO SAINT. Some are quite general even if dedicated to specific types of systems such as industrial plants, communication systems, while others are very specific as, for example, in the case of nuclear power plant simulators or cardiovascular physiology simulators. Simulators allow you to build a simulation program using graphical menus without the need for programming. They are fairly easy to learn, and a drawback that many of them have is that they are limited to modeling those systems predicted by their standard features. In any case, some simulators provide the possibility of incorporating routines written in a general purpose language to deal with non-standard elements. Often they also have animation capabilities to show the simulation in action and this allows you to easily illustrate the simulation even to inexperienced people.

In this thesis we will use the ExtendSim simulator, which will be described in the following paragraphs.

5. Validation of the simulation model

In the next phase it is necessary to verify if the model that has been created provides valid results for the system under examination. In particular, it must be verified whether the performance measures of the real system are well approximated by the measures generated by the simulation model. This is very difficult to do, especially at the design stage when the real system does not exist.

6. Simulation design

Before moving on to running the simulation, it is necessary to decide how to conduct the simulation. Often a simulation is a process that evolves during its realization and where the initial results help to lead the simulation towards more complex configurations. There are also statistical problems:

- The determination of the length of the transient of the system before reaching stationarity conditions, the moment from which you start to collect data if you want performance measures of the system in steady state.
- The determination of the length of the simulation (duration) after the system has reached equilibrium. In fact, it must always be borne in mind that the simulation does not produce exact values of the performance

measures of a system as each single simulation can be seen as a "statistical experiment" that generates statistical observations on the performance of the system. These observations are then used to produce estimates of the performance measures and of course increasing the duration of the simulation can increase the accuracy of these estimates.

7. Execution of the simulation and analysis of the results

The simulation output provides statistical estimates of a system's performance measures. A fundamental point is that each measurement is accompanied by the "confidence interval" within which it can vary. These results may immediately highlight a system configuration that is better than the others, but more often more than one configuration will be identified as being the best candidate. In this case, further investigation may be required to compare these configurations.

8. Presentation of the conclusions

In conclusion, it is useful to draw up a report and presentation that summarizes the study carried out, how it was conducted and including the necessary documentation. Including an animation of a simulation in the presentation is usually very effective.

3.5 Pros and cons of simulation

The relevant aspects that make simulation a widely used tool are linked to the fact that it allows to represent real systems, even complex ones, also taking into account the sources of uncertainty and to reproduce the behaviour of a system in reference to situations that cannot be experienced directly.

On the other hand, it must always be kept in mind that the simulation provides information on the behaviour of the system, but not exact "answers". The analysis of the output of a simulation could be complex and it could be difficult to identify which configuration is the best. Furthermore, the implementation of a simulation model could be laborious and high computation times may be required to carry out a meaningful simulation. In summary, some advantages of simulation are:

- Possibility to carry out system tests before investing in structures
- Possibility of zooming in on simulated time
- Possibility of understanding the causes of events
- Explore new management policies
- Problem diagnosis
- Requirements identification
- Understanding the system
- Repeatability

Some disadvantages can be:

- Long and not easy model building
- Long and complex interpretation of results.

3.6 Simulation-based optimization

Simulation projects often aim to answer questions related to the optimization of specific characteristics that represent “what if” scenarios to the proposed system. Optimization is defined as the minimization or maximization or both related to a one or multi-objective function that summarizes, in a mathematical form, the questions made for the system. If so, different combinations of alternatives are considered viable if it satisfies all the restrictions of the problem, or unviable if at least one restriction is not satisfied. The alternative that has the best value for the objective function is considered optimal. If the simulation has sufficient data to represent the analysed system, the best-simulated solution can be inferred as optimal, and have good chances to be implemented in a real system, performing the goal to be an excellent tool to help decision making. To find the optimal solution, a search space made from the combination of the possible values from the variables is evaluated. The size of this search space can be a problem regarding the resources necessary to perform a full search covering all the possible solutions, to find the best one. The resources, in this case, are commonly related to the computational power available to perform all the possible solutions that represent a quantity of time that the decision-making person could not have. Those types of problem are considered NP-hard. Both simulation and optimization are used for performance improvement in management, planning, control and methods for decision-making.

3.7 Literature review on simulation and optimization models

This paragraph presents some discrete event models created with the use of the ExtendSim software. There are also some examples of optimization models built with ExtendSim. No specific examples relating to railway maintenance have been found, therefore some articles are reported that concern other topics but which use the same methodology proposed.

On DES models, S. Alodhaibi et al. [25] focused upon passenger flow issues within airport terminals and included all events occurring between curb-side and boarding. To improve passenger flow and associated planning activities, a simulation framework is developed using Discrete Event Simulation (DES) Busato et al. [26] realized a simulation model for a rice-harvesting chain, using simulation in ExtendSim but not optimizer block, for minimisation of manpower in large scale field operations, such as crop harvesting and transportation.

Bourque et al. [27] implemented the production planning of metallic nuclear fuels (MNF) and demonstrate how SCALE and MCNP nuclear physics model outputs can be integrated into ExtendSim discrete-event simulation (DES) models of the fuel fabrication process to determine the optimal number of staff hired to ensure fuel production goals are met, operations comply with effective dose limit regulations, and overall project costs are reduced.

Liu et al. [28] adopted a simulation method and surrogate technology to transform the black box command and control system into a white box model and try to optimize the seat allocation problem by maximizing the efficiency. Firstly, the dynamic command and control process is modeled based on the simulation tool of ExtendSim, by mining key elements and relations in command and control process. Secondly, with training data generated from the ExtendSim, the artificial neural network (ANN) based surrogate model is adopted to optimized the relation between allocation solution inputs and efficiency outputs. This model realizes through ExtendSim the simulation of a process, while for the optimization it makes use of artificial neural network.

Jiang et al. [29] proposed a campus traffic system simulation model that considering mixed traffic flow and pedestrians' behavioural characteristics. The model is simulated by ExtendSim simulation software. The characteristics and rules of campus traffic layout are established, which provide references for the layout and optimization of facilities, achieving goals of safety and high-performance.

Some examples of models created with the use of the ExtendSim software and its optimization function are reported below.

Hu and Yi [30] realized a simulation model of Kanban system for steel rolling production line is established based on discrete event simulation theory and ExtendSim simulation software. The optimal Kanban quantity is obtained with the genetic algorithm optimizer module of the ExtendSim, which found the WIP stocks and the end product stocks in the production line.

Hu and Linwei [31] are established a simulation model of Noshery (Tang Folkway) service system with Extendsim simulation software, based on discrete event simulation theory and the service flow of a typical Noshery, . By the means of queuing strategy and the optimization module the software possesses, the best configuration of the system resources is obtained.

Hou and Zeng [32], according to the characteristics of the perishable agricultural products and applying ExtendSim system simulation software, built a metamorphic rate (t, R, S) inventory system model of agricultural products. In the initiali inventory demand random, given inventory level condition, they used a day on average total cost in inventory as the objective function. Through the ExtendSim Optimizer optimization simulation module, they obtained the best inventory strategy.

Finally, Li and Li [33] established the inventory control model based on random demand and random lead time. The economic order volume, order point, and safety inventory are obtained based on cost consideration so that the inventory cost can be controlled, and then the simulation verification based on ExtendSim software is established. Then, a reasonable and effective inventory management control strategy for the enterprise rolling workshop spare parts inventory reference is provided.

4. Risk-based approach

In this Chapter, the technologies used to monitoring and to collecting data from the field, are introduced. These data are employed to realize models of the degradation of rail track. In addition, the existing approaches to predict the future status of rail are described, since the asset status is a primary input for the proposed risk-based scheduling model.

4.1 Track monitoring

This paragraph illustrates the existing measurement and monitoring techniques to assess track condition [34].

Although visual inspections executed by operators walking along the track remain the most used and efficient, nowadays new technologies and methods for the control and monitoring of the track are used. These new techniques are mainly focus on ultrasonic, laser and cameras.

The geometric state of the track is assessed by the check of some geometrical parameters determined by the railway network regulations: analysing the values of these parameters and verifying them with threshold values, it is possible to evaluate the geometric quality of the track, under which are programming improvement actions and driving limitations.

It is important do an accurate monitoring of track geometry defects to carry out maintenance when the solutions described above are no longer enough.

There are two kinds of vehicles used to assess track geometry with measurement instruments mounted above, these are :

- Low speed vehicles: they are small and economic vehicles (usually two axles ones) with evaluation instruments of track geometry. They are usually allocated to regional maintenance office to enable instant measurement in critical area.
- Special measurement trains: they are complex, sophisticated and expensive track inspection vehicles possessed by railway administrators. These vehicles execute global inspection of the track, they are usually managed centrally and sent the results to a central data base and to the

regional maintenance offices. Most of the infrastructure inspection is nowadays realized by the measurement trains.

Railways take on very different configurations: various maintenance actors choose having fully “specialized” trains, e.g. one for the track, one for the catenary, one for the telecommunications and signalling.

Others choose combining everything on a single train. Anyway, these trains are circulating on the whole network of an administration, on a schedule routine basis and delivering the data every N days.

The new innovation is to employ trains in commercial service to execute inspections on the track. This way to do control would save cost because of the following facts:

- Expensive special measurement train wouldn’t be necessary;
- Cost of inspections (crew, traction, etc.) would be saved;
- Slots for inspections wouldn’t be necessary difficult to find on busy lines.

Thus, the track availability for service would grow and since the train would be travel the same line every day, the frequency of the check would be high.

To execute inspections of the line employing on-service trains, they must be provided with the equipment to cover the network, usually one train with instruments for each line. Furthermore, the localization becomes very critical: special techniques must be used and the GPS (where available) could not provide the necessary resolution. In this context, the expense of the instruments is covered by the savings of special train and relevant operational expenses.

Such an automatic system (on-service trains used for inspections) is complex to plan, assemble and manage. Every component must have much higher reliability than the usually accepted for component of a special train.

However, assessing track condition with on-service trains can provide very helpful data for a better maintenance strategy:

- Better trends (more precise),
- Immediate check of the works,
- Early finding of unpredictable faults.

Data can be collected in electronic format. Serious faults are signalled at once and corrective actions are made immediately. Other non-serious defects are provided to the central office where the defects are estimated by algorithms and classified by gravity indexes.

The most used inspection and monitoring technologies of track are the following [35]:

1. Fibre optic sensors positioned on the track and other infrastructure parts (like bridges) [36].
2. Hollow-shaft acoustic sensor system [37].
3. Rail monitoring sensor joining eddy current distance measurement and acceleration data.
4. Laser profiler with inertial pack to monitor the track geometry.
5. Ultrasonic non-destructive inspection techniques.
6. Rail surface monitoring with non-contact thermography system.
7. Visual camera.

The most promising techniques are the first four because they let to the Railway Infrastructure Manager obtaining updated information every day basis on the track state in a cheap way. Acoustic methods, pulsed eddy currents and laser profiler inspection could be installed on commercial trains allowing an unattended and automated measurement system. This is a good innovation in track inspection technologies which will decrease the use of expensive instrumentation trains inspecting the infrastructure in the night when rail services is suspended and will raise the volume of the rail transport. Condition-based maintenance become possible thanks to the availability of frequent and quality data on the track condition with updated measurements every day.

Nowadays, ultrasonic and thermographic technologies can only be used with very low speed trains that give them unusable for commercial trains.

The technologies listed above are described more in detail below.

1. Fibre optic sensors

Fibre optic sensors distributed along the track can be used for spatially continuous monitoring of the tracks deformation and temperature, they are used also for monitoring of the structural solidity of infrastructures as bridges, tunnels and embankments. A single-mode optical fibre cable has to be attached to the track and/or the structure considered with the purpose of detect all the strains.

2. Acoustic Inspection Techniques

Acoustic methods “hear” to natural sound sources as the rolling noise. If there are hollow shafts (for example, in some high-speed trains) integrated acoustic sensor systems for hollow shaft can be employed to identify defects in wheelsets of the rolling stock. The system could comprise acceleration sensors and structure sound sensors, it uses wireless real-time data transmission. The acoustic part of the system identify and assesses acoustic signals produced by the rail-wheel contact.

This methods can find wheel defects and it is used for monitor train's wheelset, but it can be useful for track safe, because a wheel defects can damage the rail.

3. Inspection using pulsed eddy currents

Eddy current measurements are a standard technique for finding cracks in metals either on the surface or within the material [38]. Nowadays, eddy current sensors have become a usual method for rail inspection [39]. The sonde has to be in contact or very near to the surface (<1mm), for this reason, use in rail inspection has for a lot of time been restricted to hand-held system or system mounted on manually driven trolleys. The distance between sensor and rail has to raise to adapt the technique of eddy current testing to a train borne platform mounted on a commercial train. However, with frequent rail inspection that would be possible using commercial trains, decreased performance of the system could be tolerated. Some experiments have demonstrated that eddy currents of lower frequency are enough able to identify larger cracks even when sonde-surface distance exceeds 10mm.

4. Laser profilometer

Profilometer is an instrument utilized to measure a surface's profile, with the purpose of quantify its roughness. An optical profilometer is a non-contact method for producing much of the same information as a stylus based profilometer. There are many different techniques which are currently being used, like confocal microscopy and laser triangulation.

5. Ultrasonic inspection

Ultrasonic techniques are the most commonly employed non-destructive methods with a large variety of use fields. Usually a broadband pulse is excited with a piezoelectric transducer and is transmit into the structure using an appropriate connection agent like water, oil or viscous paste. The waves react to interior defects and are sent back so that they can be revealed by a sensor, this can be the same transducer that was employed for excitation or an additional sensor. Ultrasonic techniques are well known for the inspection of high-speed lines in Europe and overseas [40], [41].

Test trains generally comprise ultrasonic and eddy current systems to automatically analyse the rail during run of the train (<100 km/h). The procedure is organised in a three level inspection process.

6. Thermographic inspection

Another non-destructive and non-contact technique used is infrared thermography. First, the material is warmed by a flash lamp or with inductive technique. Second, the spatial-temporal trend of the thermal field is controlled by an infrared camera. Finally, if there are defects, “hot spots” of higher temperature can be detected reducing locally the thermal conductivity. Extra information about size and depth of the defect can be defined from the temporal variation of the thermal field [42], [43].

Thermographic investigation has before shown its great potential for the characterization of typical defects of the rail. It could be demonstrated that this method mainly enables the characterisation of the rails with a elevated sensitivity and a elevated testing speed. This system can be integrated in a testing train like ultrasonic and electromagnetic techniques. Nowadays, an automated testing of the rails and automated defect recognition at speeds up to 20 m/s (about 70 km/h) looks to be possible.

7. Inspection using visual camera

The inspection of the track is providing by visual cameras that are important to reduce or eliminate visual control performed by men walking along the line to check the rail, default or missing components, etc. Nowadays, these instruments does not yet consent total and certain elimination of the inspection done by workers, but can be useful to detecting a lot of dangerous situation difficult to find by the human eye. There are a lot of linear cameras mounted under trains.

A colour image is usually used to enable an inspector viewing the track as if he was walking, but with clear advantage for line capacity and safety.

A black and white image is usually employed for detecting rail surface faults and for every automatic analysis by machine vision techniques (rail surface, sleepers, fasteners, joints). The automatic analysis is helpful to focus the attention of the operator, who then goes to examine the relevant colour image and establishes the adequate actions.

4.2 Degradation models of track and literature review

With age and usage of the railway infrastructure and track, they can become unreliable due to fault cause from degrade. A high cost of railway operation, economic loss, damage to the railway asset and environment, are just someone of possible consequences of a failure. Grave problems for the society’s position in the marketplace can be caused by unreliability railway, because of annoyance, inconvenience and a lasting customer dissatisfaction. The realization of reliability purposes and compensate for unreliability can be guaranteed from an appropriate and successful maintenance strategy.

To restore a defect part to an operational condition, decrease or remove the likelihood of fails and control the degradation of the track are used maintenance actions. It is requited to model track degradation behaviour to choose an appropriate and successful maintenance policy, but foreseeing the degradation of track is a complicated job, that needs the information below:

1. The relations between every track components,
2. The result of maintenance actions on quality of the track,
3. The different factors like environmental, soil type and condition.

The increased of demand and complexity requires the necessity for complete track degradation models, because of higher speed and axle load that accelerates the track aging process and negatively influences its reliability.

In last years, several research has been made in track geometry degradation modelling [44], some of the most significant researches are cited below.

An essential requisite for modelling track degradation is defining an indicator to figure track quality. Indicators can be different and they are chosen based on the scope of the research. *Figure 14* showed some indices for figuring track quality condition.

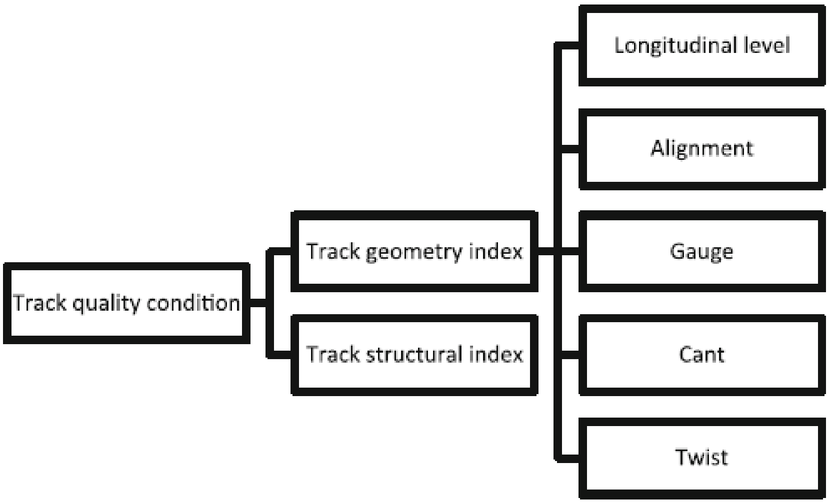


Figure 14: Measures of a railway track

Sadeghi et al. [45] presented some indicators of track geometry that employ the following track geometry parameters: alignment, profile, gauge, twist and rail cant, they combined the parameters to design the track geometry index. Sadeghi et al. [46] submitted a quantitative track structural quality index to consider structural defects. This index is different for each track component group: rail, sleeper, fastener, ballast, etc. Faiz et al. [47] laboured the geometry parameters worn in the UK track maintenance process and used linear regression analysis to demonstrate their correlations. Li et al. [48] proposed a Generalized Energy Index - GEI, that can consider different track irregularity wave-length and speed, in place of a Track Quality Index - TQI for quality evaluation of track. El-Sibaie et al. [49] created a some track quality indices to evaluate quality condition of track in relation to different track classes.

In *Figure 15*, it can be note that most of the authors considered short wavelength longitudinal level as the critical factor in model of degradation.

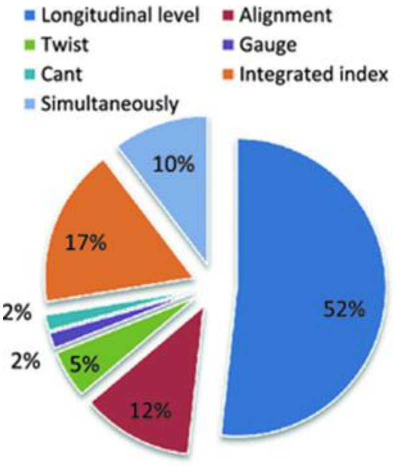


Figure 15: Most used track geometry measures

Once track quality measure have been established, a degradation model have to be realized and the result of different maintenance strategies on track degradation assessed. Mechanistic and statistical approaches are two main approaches for track geometry degradation modelling:

- Mechanistic approaches: some researchers searched to find the interactions between components of track and their effects on track geometry degradation. For example some models are proposed by Shenton [50], Sato [51],[52], Chrismer et al. [53], Öberg et al. [54], Zhang et al. [55] and Dahlberg [56].
- Statistical approaches: the most used methods are schematized in *Figure 16*.

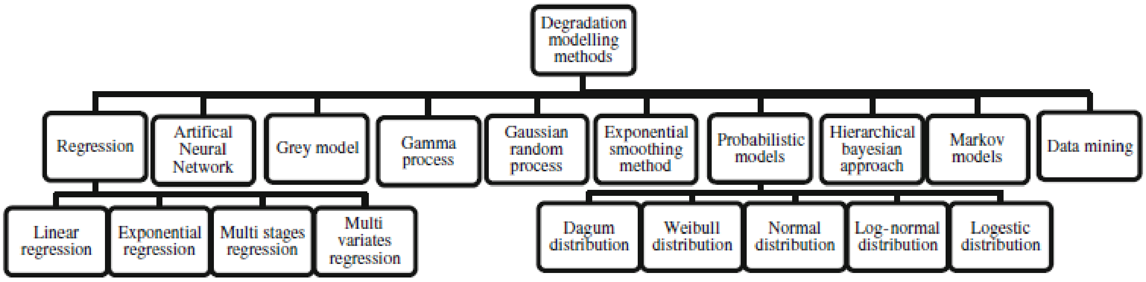


Figure 16: Statistical approaches

Andrade et al. [57] evaluated track geometry degradation and for degradation model parameters their uncertainty. To do that they regarded a linear model for longitudinal level degradation of track, executed statistical correlation analysis for each group section and conformed the log-normal distribution to the longitudinal level degradation of track.

Gou et al. [58] employed a multi-stage linear model to face with various phases of degradation between two consecutive maintenance interventions and the exponential rise of irregularity for track.

Famurewa et al. [59] confronted the precision between linear, exponential and grey models in the estimate and prediction of track geometry degradation. The compare proved that the grey model has a roughly equal error value with the exponential model and lower mean average percentage error than the linear model.

In [60] the track quality deterioration of one specific track position is presented during a period of six years: the theoretical exponential function is a well approximation with the real track behaviour.

A multivariate regression model to prove the connection between the track degradation measure variable and influencing variables on track degradation is proposed from Lyngby [61]. He claimed that every different sections of track are not equal, the track is split into homogenous sections with similar variables. Lyngby proved that: axle load has a non-linear relation with degradation, degradation after tamping is conditioned on the quantity of antecedent tamping, soil composed by clay material will stand sooner than other types of soil, light rail tracks deteriorate faster than heavy rail tracks, severe rainfall increases degradation rate.

Liu et al. [62] created a short range prediction model, with using waveform data, to assess any track irregularity index over a brief track section length (25 m) and on a daily base. They found that the process of track surface variation over track sections is non-linear and every track sections have various non-linear process.

Xu et al. [63] suggested an historical changes approach of track irregularity to forecast the short-term track degradation. They calculated the non-linear behaviour of track irregularity along a cycle employing a number of short range linear regression models.

Kawaguchi et al. [64] realised in their study two degradation models to predict track alignment irregularities. In the first model, they proposed a degradation model based on analysis of lateral track deformation to evaluate mean time to maintenance of track alignment irregularities. In the second model, they developed different degradation model based on the exponential smoothing method to accurately forecast the track alignment irregularities a maximum of a year in advance.

Quiroga et al. [65] used three models: a comparison of the efficiency of the double exponential smoothing method, a general degradation model and an autoregressive model for track degradation prediction. Every models miss their efficiency in track degradation forecast after performing a number of tamping procedures. After considering these problems, they worked on a hybrid discrete-continuous framework based on a grey box model. Comparing these four models, they concluded that their proposed hybrid model is more efficient in terms of track degradation behaviour forecast.

A classification of the track longitudinal level degradation over time, into three different speed and different inspection intervals, with a stochastic approach based on Dagum distribution is developed by Vale et al. [66].

Power spectral density analysis and cross-level statistics about track irregularities to improve track degradation modelling are discussed by Zhu et al. [67] with a Gaussian random process.

A Bayesian approach to assess a track geometry degradation model and to deal with the uncertainty of its parameters is developed by Andrade et al. [68]. They considered the track longitudinal level deviation in a linear relationship with passing tonnage and taked the initial longitudinal level and degradation rate would assumed a bi-variate log-normal prior distribution. They justified that the parameter uncertainties are important in the design phase.

Some stochastic Markov model are used by: Bai et al. [69] to evaluate track degradation, they considered different heterogeneous factors and justify that the existence of these factors caused maintenance units with the same mileage to show different degradation behaviour; Yousefikia et al. [70] to model tram track degradation and to obtain the optimal maintenance strategy; Liu et al. [71] to forecast track quality condition used a model by integrating the grey model and Markov chain.

Artificial neural networks are used by Guler [72] to model the degradation of different track geometry parameters. The model evaluate traffic load, curvature, velocity, gradient, cross-level, rail type, sleeper type, rail length, falling rock, land slide, flood, and snow as influencing factors.

Chaolong et al. [73] developed a modified grey model to study track irregularity time series data and obtain a medium-long term forecasting of track cross levelling. They confronted the stochastic linear autoregressive model with Kalman filtering model and artificial neural network with regard to the short term track cross levelling forecast. They observed the accuracy of the artificial neural network model was greater than the other models.

Xu et al. [74] proposed a machine learning model based on the requirements and inspection data of the track using a multi-stage framework to forecast changes in track irregularity over time. They defined different stages of track changes based on maintenance limits and linear regression is applied to predict track degradation in each stage. Xu et al. [75] developed also a track measures data mining model to forecst railway track degradation for a short time period.

Data mining and time series theories are used by Chaolong et al. [76] to forecast track irregularity standard deviation time series data. They used the linear recursive model and the linear autoregressive moving average model in order to forecast the changing tendency of track irregularity.

Based on data mining techniques, the prediction of the asset condition can be categorised in two mode: nowcasting and forecasting [77]:

- Nowcasting models are applied to identify failures that will carry out to fault within a few hours, this is useful for safety reasons and even to extend remaining useful life (RUL).
- Forecasting models can be helpful to evaluate the condition of an asset for the remaining useful life in the long time.

Three are the methods to quantify remaining useful life:

- Data driven methods are totally based on the data acquired by sensors, they realize classification and clustering techniques to identify anomalies.

- Symbolic methods using of work orders and other empirical records of maintenance.
- Physical methods use the physical structure of the component to analyse its degradation.

The combination of data driven, symbolic and physical models into hybrid models is a good solution for nowcasting and forecasting of asset condition.

Into alert management system should also be used prognostic models, forecasting information about the asset status, providing estimations of current or future values of the relevant parameters, to get breakdown probability distributions. The future degradation and defect evolution are described as stochastic processes.

4.3 Risk-based model

Consilvio et al. [78] used the risk-based approach to determining the set of rail stretches to be maintained starting from a stochastic deformation model. In particular their considered the rail vertical deformation model provided by Famurewa et al. [79], [80]. This model provides the rail deformation over time of the given rail stretches.

However, though the researchers introduced the notion of deformation uncertainty, their determination of the deadlines for maintenance interventions neglected this factor and considered only the average deformation.

From the physical point of view, it is then possible that the average deformation relapses within the thresholds, while at some points the real profile may exceed these limits. This condition is represented in *Figure 17*, rail profile with comparison of:

- The nominal rail profile (dashed line),
- The predicted average deformation (dashed-dotted line),
- The positive/negative thresholds (dotted lines),
- The real rail profile along the 200 m rail stretch (thick continuous line).

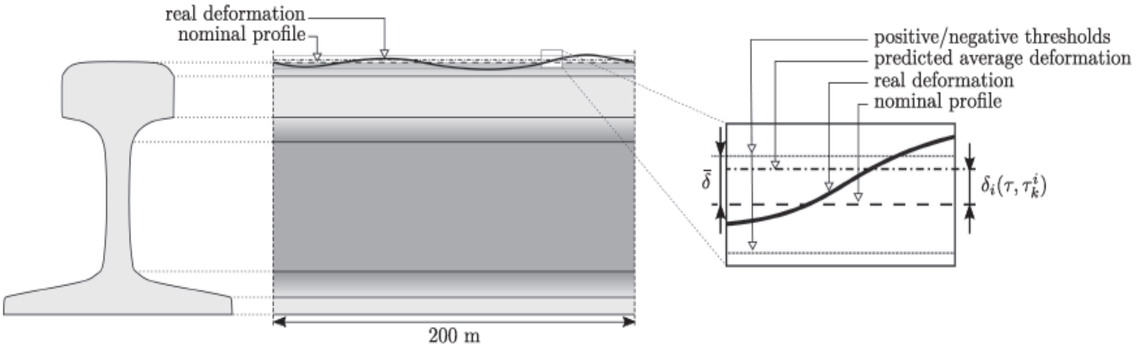


Figure 17: Rail profile degradation

To assess the probability that the deformation overcomes the relevant threshold, the model uncertainty is explicitly considered. The utilized notations is summarized in *Table 2*:

Table 2: Notations of deformation model

Variable	
$\mathcal{A}$	set of the assets needing maintenance and $\lvert\mathcal{A}\rvert$ the relevant cardinality
$\mathcal{R}$	set of railway assets
τ_k^i	time instant of the last maintenance activity on the asset i
$\delta_i(\tau_k^i)$	residual deformation after the maintenance activity in τ_k^i
$\delta_i(\tau, \tau_k^i)$	deformation model
ε	Gaussian r.v. representing the deformation model error
f_i	fault event of the asset i
$\overline{Pr\{f_i\}}$	maximum admissible fault probability (hard deadline)
γ_S	parameters stating the relative admissible fault probability for the soft deadline and the release time
$\tau_H^i(\overline{Pr\{f_i\}})$	hard deadline for maintenance activity of $i \in \mathcal{A}$
$\tau_S^i(\gamma_S\overline{Pr\{f_i\}})$	soft deadline for maintenance activity of $i \in \mathcal{A}$
$\tau_R^i(\gamma_R\overline{Pr\{f_i\}})$	release time for maintenance activity of $i \in \mathcal{A}$

The vertical deformation model explained in [79], [80] for determining the spatial average (computed over the whole length) of the vertical rail deformation of the generic rail stretch $i \in \mathcal{R}$ is a random variable and is expressed as:

$$\delta_i\left(\tau, \tau_k^i\right) = \delta_i\left(\tau_k^i\right) exp(\alpha_i \tau) + \varepsilon \qquad (4.3.1)$$

Where:

- $\delta_i\left(\tau_k^i\right)$ is the deformation along the rail stretch i after the maintenance performed in τ_k^i,
- $\delta_i\left(\tau, \tau_k^i\right)$ average deformation in a generic instant $\tau > \tau_k^i$,
- $\varepsilon \in \mathcal{N}(0, \sigma^2)$ is a Gaussian random variable (r.v.) with null expectation representing the model error modeling the deformation uncertainty due to the model approximations, the measurement errors, and punctual deviation of the real value of the deformation with respect to the average predicted by (4.3.1). It is assumed that the r.v. ε is independent on the particular rail stretch.

The model parameters: the initial deformation $\delta_i\left({\tau^i}_k\right)$, the coefficient α_i, and the variance σ^2, are determined via field data from Famurewa et al. and are assumed to be known.

Agreeing to the model in (4.3.1), the deformation $\delta_i\left(\tau, \tau_k^i\right)$ results in a Gaussian stochastic process with time dependent expectation $E[\delta_i\left(\tau, \tau_k^i\right)] = \delta_i\left(\tau_k^i\right) \cdot exp(\alpha_i \tau) + \varepsilon$ and constant variance σ^2.

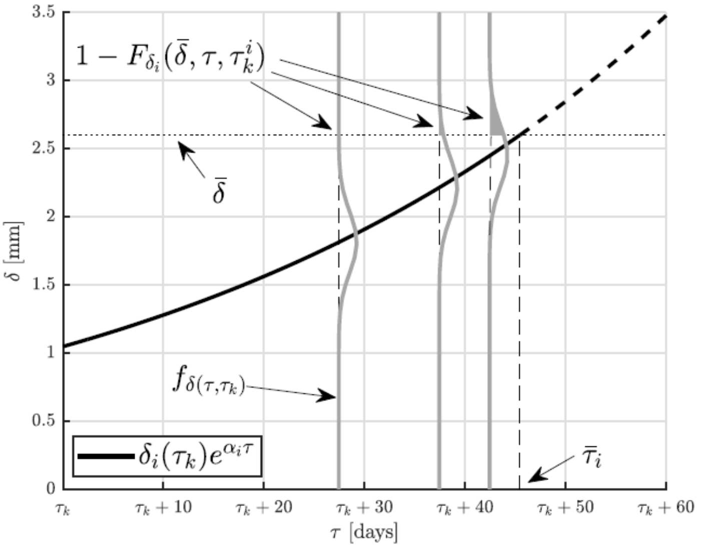


Figure 18: Gaussian stochastic process

As a result, at any time τ, there is a non-zero probability $Pr\{\delta_i(\tau,\tau_k^i) \geq \bar{\delta}\} = 1 - F_{\delta_i}(\bar{\delta},\tau,\tau_k^i)$ that the actual deformation is bigger than the threshold $\bar{\delta}$, as represented by the colored area in *Figure 18*: the thick line represents the expectation of the different Gaussian random variables $\delta_i(\tau,\tau_k^i), \forall\tau$.

The deadline for maintenance intervention, that is the time instant at which the maximum deformation is achieved, can be calculated by considering a maximum threshold of such a probability. Therefore, let $T_{\bar{\delta}}^i$ be a random variable representing the time instant at which the deformation of the rail stretch i reaches the threshold $\bar{\delta}$, is found:

$$\bar{\delta} = \delta_i(\tau_k^i)\ exp(\alpha_i T_{\bar{\delta}}^i) + \varepsilon \tag{4.3.2}$$

With of a simple rotation of (4.3.2) gets:

$$T_{\bar{\delta}}^i = g(\varepsilon) = \frac{1}{\alpha_i}\ln\frac{\bar{\delta}-\varepsilon}{\delta_i(\tau_k^i)} \tag{4.3.3}$$

which is defined if the argument of the logarithm is positive.

Such assumption (4.3.2) is reasonable since, for Chebyshev's inequality $Pr\{\bar{\delta}-\varepsilon \leq 0\} \leq \frac{σ^2}{\delta^2}$, the probability that $\varepsilon \geq \bar{\delta}$ turns out to be neglectable for realistic values of the uncertainty $σ^2$ and of the threshold $\bar{\delta}$, is occured whenever the model error ε is small compared to the admissible deformation.

Because the function $g(\varepsilon)$ in (4.3.3) is continuous and decreases monotonically, the cumulative distribution function (cdf) of $T_{\bar{\delta}}^i$ can be defined as:

$$\begin{aligned} F_{T_{\bar{\delta}}^i}(t) = {} & Pr\{T_{\bar{\delta}}^i \leq \tau\} = Pr\left\{\frac{1}{\alpha_i}\ln\frac{\bar{\delta}-\varepsilon}{\delta_i(\tau_k^i)} \leq t\right\} = \\ & = Pr\{\varepsilon \geq \bar{\delta} - \delta_i(\tau_k^i)\ exp(\alpha_i T_{\bar{\delta}}^i)\} = \\ & = 1 - F_\varepsilon(\bar{\delta} - \delta_i(\tau_k^i)\ exp(\alpha_i T_{\bar{\delta}}^i)) \end{aligned} \tag{4.3.4}$$

where $F_\varepsilon(.)$ is the cdf of the Gaussian random variable.

Therefore, the relevant probability density function (pdf) is:

$$f_{T_{\bar{\delta}}^i}(\tau) = \frac{dF_{T_{\bar{\delta}}^i}}{d\tau} = \frac{\alpha_i\delta_i(\tau_k^i)\ exp(\alpha_i T_{\bar{\delta}}^i)}{\sqrt{2\pi\, σ^2}}\exp(-\frac{(\bar{\delta} - \delta_i(\tau_k^i)\ exp(\alpha_i\tau))^2}{2\, σ^2}) \tag{4.3.5}$$

whose trend is shown in *Figure 19*.

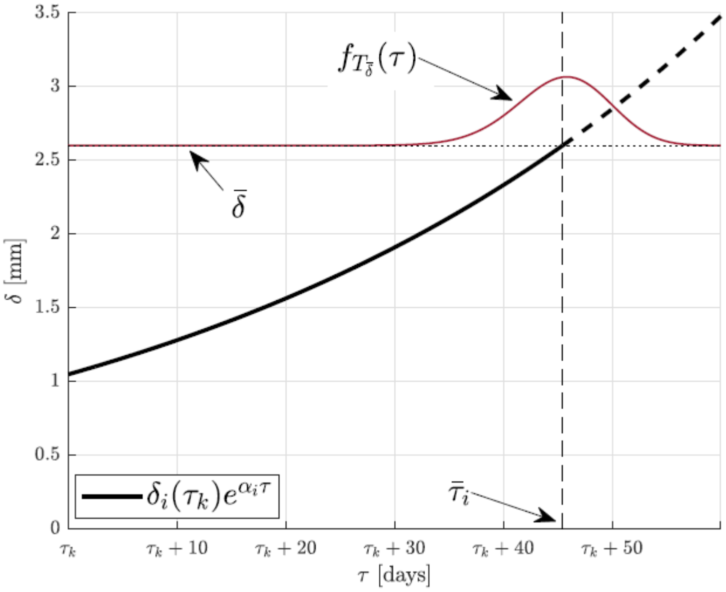


Figure 19: Shape of the pdf $\boldsymbol{f_{T^i_{\bar{\delta}}}(\tau)}$

The pdf in (4.3.5) allows to calculate the median, which, thanks to the symmetry of the Gaussian pdf of ε, always coincides with the instant $\bar{\tau}_\iota = \frac{1}{\alpha_i} \ln \frac{\bar{\delta}-\varepsilon}{\delta_i(\tau_k^i)}$ at which the expectation of $\delta_i(\tau_k^i)$ gets to the threshold $\bar{\delta}$. Note that $\bar{\tau}_\iota$ is always bigger than the expectation $E[T^i_{\bar{\delta}}]$, meaning that neglecting the uncertainty in (4.3.1) leads to underestimation of the fault probability.

Then let f_i be the fault event of the rail i and $\bar{R}_\iota$ be the maximum tolerable value for the fault risk $R_i = D_i(f_i)Pr\{f_i\}$. Assuming that the loss $D_i(f_i)$ is note, the maximum failure probability is:

$$\overline{Pr\{f_\iota\}} = \frac{\bar{R}_\iota}{D_i(f_i)} \qquad (4.3.6)$$

Therefore, the model allows to evaluate the hard deadline $\tau_H^i(\overline{Pr\{f_i\}})$ for the maintenance activity on the rail i consisting of the instant at which $F_{T^i_{\bar{\delta}}}(t) \geq \overline{Pr\{f_\iota\}}$.

Similarly, the soft deadline is defined as the time instant $\tau_s^i\left(\overline{Pr\{f_i\}}\right) = \tau_H^i(\gamma_s \overline{Pr\{f_i\}})$ at which $F_{T^i_{\bar{\delta}}}(t) \geq \gamma_s \overline{Pr\{f_\iota\}}, \gamma_s < 1$.

The soft deadlines should be respected, although not in an obligatory way, to minimize the failure probability even if the threshold $\overline{Pr\{f_\iota\}}$ is always guaranteed.

The same approach can be applied to estimate the release time, which is defined as the time instant τ_R^i at which, given the degradation process $F_{T^i_{\bar{\delta}}}(t) \geq$

$\gamma_r \overline{Pr\{f_i\}}, \gamma_r < \gamma_s$, the failure probability of a generic rail stretch i becomes unneglectable.

By iteratively updating these release times, the subgroup A ⊆ R of rail stretches whose maintenance has to be scheduled can be identified.

5. Maintenance scheduling problem

In this chapter, the Mixed Integer Linear Programming (MILP) problem to define the optimal assignment of the maintenance activities to the available maintenance teams and the relevant schedule is presented. The problem derives from [81], in which a deterministic scheduling of predictive maintenance activities along a railway line is describe but also from [82] in which a risk-based scheduling is described for the maintenance of a railway network is proposed.

The deadlines determined by means of the degradation model are considered as constraints or terms of the cost function.

There are limited periods for maintenance, of this constraint is taking into account by introducing train-free sub-intervals during which the train circulation is forbidden and the maintenance activities can be performed.

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, in the next paragraph are list notations and assumptions of the scheduling model for railway, and in then the optimization problem is describe.

5.1 Problem notations and assumptions

In *Table 3* below are described notations of the MILP model for railway maintenance:

Table 3: Notations of MILP model

Indexes	
$i = 0, \dots, \lvert\mathcal{A}\rvert$	indexes of assets
$j = 1, \dots, \lvert\mathcal{A}\rvert + 1$	
$i = 0$	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	
$p_{i,g}$	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
$\delta h_i(t_k^i)$	hard deadline of the maintenance activity on asset i after the maintenance occurred in $\mathrm{t}_\mathrm{k}^\mathrm{i}$
$\delta s_i(t_k^i)$	soft deadline of the maintenance activity on asset i after the maintenance occurred in $\mathrm{t}_\mathrm{k}^\mathrm{i}$
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
$w_{i,m,h}$	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

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 asset is characterized by distinct soft and hard deadlines.
- Each maintenance team can perform only a maintenance activity at a time.
- No pre-emption of maintenance activities is allowed.
- The maintenance activities can be tardy only with respect to the soft deadlines.
- The maintenance activity on all the assets can be processed by any free maintenance team.
- The maintenance teams are available throughout all the scheduling time horizon.

5.2 Deterministic problem formulation

The optimization problem is described assuming that the risk analysis is performed in $\tau > t_k^i, \forall i \in \mathcal{A}$, and as a consequence, the deadlines depend on the last maintenance activities performed, for any asset, in t_k^i. Therefore, given the notations list in the previous paragraph, the optimization problem is formalized in the following way:

$$\min\left(\sum_{m=1}^{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{A}|}\omega_i c_{i,m} + \sum_{i=1}^{|\mathcal{A}|}\omega_i q_i\right) \tag{5.2.1}$$

subject to:

$$c_{i,m} = \ max(0, t_{i,m} + p_{i,m}x_{i,j,m}) \qquad \forall i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{5.2.2}$$

$$c_{i,m} \leq \delta h\left(t_k^i\right) \qquad \forall i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{5.2.3}$$

$$q_i \ = \max(0, c_{i,m} - \delta s\left(t_k^i\right)) \qquad \forall i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{5.2.4}$$

$$t_{j,m} \geq c_{i,m} + \ s_{i,j,m} - B\left(1 - x_{i,j,m}\right) \qquad \forall m \in \mathcal{T}, \forall j, i \in \mathcal{A} \tag{5.2.5}$$

$$t_{j,m} \geq \ s_{0,j,m} - B(1 - x_{0,j,m}) \qquad \forall m \in \mathcal{T}, \forall j \in \mathcal{A} \tag{5.2.6}$$

$$\sum_{j=1}^{|\mathcal{A}|+1} x_{0,j,m} \leq 1 \qquad \forall m \in \mathcal{T} \tag{5.2.7}$$

$$\sum_{m=1}^{|\mathcal{T}|}\sum_{i=0, i\neq j}^{|\mathcal{A}|} x_{i,j,m} = 1 \qquad \forall j \in \mathcal{A} \tag{5.2.8}$$

$$\sum_{m=1}^{|\mathcal{T}|}\sum_{j=1, j\neq i}^{|\mathcal{A}|+1} x_{i,j,m} = 1 \qquad \forall i \in \mathcal{A} \tag{5.2.9}$$

$$\sum_{e=1, e\neq j}^{|\mathcal{A}|+1} x_{j,e,m} - \sum_{i=0, i\neq j}^{|\mathcal{A}|} x_{i,j,m} = 0 \qquad \forall j \in \mathcal{A}, \forall m \in \mathcal{T} \tag{5.2.10}$$

$$t_{j,m} \geq \ I_h w_{j,h} + \ s_{i,j,m}x_{i,j,m} \qquad \forall i, j \in \mathcal{A}, \forall h \in \mathcal{H}_{t_k,\mathrm{T}}, \forall m \in \mathcal{T} \tag{5.2.11}$$

$$c_{i,m} \leq I_h + \ell_h + B\left(\ 1 - w_{i,h}\right) \qquad \forall i, \forall\ h \in \mathcal{H}_{t_k,\mathrm{T}}, \forall m \in \mathcal{T} \tag{5.2.12}$$

$$\sum_{m=1}^{|\mathcal{T}|}\sum_{h=1}^{\left|\mathcal{H}_{t_k,\mathrm{T}}\right|} w_{i,m,h} = 1 \qquad \forall i \in \mathcal{A} \tag{5.2.13}$$

$$t_{i,m}, c_{i,m}, q_i \in \mathbb{R}_{+,0} \qquad \forall\ i \in \mathcal{A}, \forall m \in \mathcal{T} \tag{5.2.14}$$

$$x_{i,j,m} \in \{0,1\} \qquad \forall m \in \mathcal{T}, \forall j, i \in \mathcal{A}: i \neq j \qquad (5.2.15)$$

$$w_{i,h} \in \{0,1\} \qquad \forall i \in \mathcal{A}, \forall h \in \mathcal{H}_{t_k,\mathrm{T}} \qquad (5.2.16)$$

where:

- (5.2.2) define the completion times of the maintenance activities.
- (5.2.3) guarantee that the maintenance activity on each asset is completed before the relevant hard deadline.
- (5.2.4) define the tardiness of the maintenance activities.
- (5.2.5) define the precedence relation between the activities on the assets i and j.
- (5.2.6) define the initial set-up times.
- (5.2.7) guarantee that at most one activity is scheduled as the first work of each maintenance team.
- (5.2.8) and (5.2.9) guarantee that every maintenance activity has exactly one predecessor and one successor, respectively.
- (5.2.10) state that a predecessor/successor pair of activities has to be assigned to the same maintenance team m.
- (5.2.11) guarantee that if an activity is performed in the train-free interval h, it starts after the beginning of such an interval.
- (5.2.12) guarantee that all maintenance activities finish within the train-free interval.
- (5.2.13) guarantee that each activity is only planned in a single time interval.
- (5.2.14)-(5.2.15)-(5.2.16) define the problem variables.

It is important underline that, even if the monetary cost of maintenance is not expressly indicated in the objective function, it is implicitly regarded in the model, because planning the maintenance activities before the hard deadline enables to avoid failure and expensive service interruptions. Furthermore, the cost reduction is modelled in terms of reduction of the time needed to complete the maintenance activities and optimization of resources utilization.

5.3 Problem formulation with stochastic constraints

As mentioned above, 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. For example, the soft and hard deadlines of each maintenance activity are continuous stochastic variables.

Another stochastic variables is the processing time of maintenance activity on asset i by maintenance team g.

The two examples above are not the only stochastic variables in an optimization problem like this. In reality, almost any data would be considered stochastic. In this thesis, stochastic deadlines and stochastic processing time are used in the model described in Paragraph 8.1.

Paragraph 4.3 describes how the stochastic deadlines are calculated.

Regarding the stochastic processing time $p(i)$, it is calculated with a normal distribution with the average μ of the initial processing time and a variance σ^2 of 0.1 of the initial processing time.

$$p(i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(i-\mu)^2}{2\sigma^2}} \qquad i \in \mathbb{R} \tag{5.3.1}$$

5.4 Data availability in real world scenarios

This paragraph analyzing in detail the input data necessary for the problem resolution. With concern to the chapters above required data are the following:

- Faults that occurred up to an instant of time;
- Maintenance activities executed before;
- Exact position of an asset;
- Processing time of the maintenance activities;
- Disposability of maintenance teams;
- Speed of maintenance teams for a move;
- Disposability train-free sub-intervals for the maintenance activities.
- Losses caused by the faults;

Refers to this data, it is good to remember that:

- Faults occurred before are regularly signed in, this means the faults rates of the assets are often noted;
- Maintenance activities performed are known because the interventions performed by every teams are take trace;
- Positions of every assets are known through the railway line layout;
- Each maintenance team declare the starting and finishing times of its intervention, making possible to know the duration of each maintenance activity and, for different type of interventions, the averages of processing time;
- Number of available maintenance teams is chosen by the maintenance provider;
- Speed of the maintenance teams travelling from an asset site to another depends on the transportation mode they use to move;
- Start and final times of each train-free sub-interval are provided by the traffic manager.
- The losses caused by the fault are the most difficult input parameter to be defined. The infrastructure manager often knows the criticality of the different assets and the expected loss caused by their breakdowns, but this information is delicate and in general not public. Therefore, the maintenance provider can evaluate these data from field interviewing.

It is important to note that some of the listed input data, such for example the processing time, are not deterministic. In this thesis these data are used in the two variables, deterministic and stochastic, highlighting the differences between the two modalities.

6. Models description

To solve the scheduling problems introduced in the previous Chapter 5 a heuristics solution algorithm is usually used.

The optimization scheduling problem on a railway network has been proven to be NP-hard and, even for relatively small instances, requires a very long time to be solved. A matheuristic solution approach, 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 and is described in [83].

In this thesis, a method for solving the optimization problem with a simulation approach is proposed. The idea is to create a discrete event simulation model using the ExtendSim simulation program.

Three different versions of the simulation model are created:

- Optimization Deterministic programming of maintenance activities with the matheuristic algorithm and simulation of the results with the ExtendSim program, (in the next chapter it is called the initial model).
- Assignment of the maintenance activities to teams carried out with algorithm and sequencing of the activities found with the simulation on ExtendSim.
- Team assignment and task sequencing optimization achieved by the ExtendSim simulation model.

For the first point, the versions with the insertion of stochastic deadlines and stochastic processing times are also reported.

6.1 Models comparison

The first model created only depicts the results of the optimization problem, that is, it is a simple simulation of the solution of the matheuristic algorithm.

In the input data we have all the problem data, constants and variables. If the simulation is done *n* times, the result will be identical every *n* times since there are no variables inserted in the model and giving everything as input.

Two different versions of this model are then make:

- In the first version, the deadline within which the maintenance activity must be completed is set as a stochastic variable, in this case the output of the model that will vary is the tardiness (or advance) with respect to the deadline of the end of the maintenance activity (for the influence of the deadline component in the formula). The remaining data of the model will not be modified, therefore it is possible to highlight the case in which a maintenance activity suddenly becomes urgent, if with the fixed schedule the work is finished late with respect to the new maintenance deadline or in any case respects the deadline.
- In the second version, the processing time of the maintenance activity is set as a stochastic variable. In this case, as this input value changes, the completion time, the start time of the maintenance activity and the tardiness with respect to the deadline that varies due to the variation of the processing time component in the formula will vary. With this case it is possible to highlight when a maintenance activity that is taking place at this moment is delayed with respect to the expected time and how much it changes the completion and start times of the subsequent maintenance activities to be carried out.

In neither of the two cases described above will the sequencing of job activities and the assignment of work to the teams vary, which are already assigned as input to the simulation, using the result of the matheuristic algorithm as input data.

Two other models are then created starting from the deterministic model, these models include the possibility of varying and trying to optimize, with the software optimization function, the sequencing of jobs in one, and the sequencing of jobs and assignment to teams in the other one.

The model that optimizes the sequencing of jobs takes as input, in addition to the problem data, the assignment to the teams obtained by solving the problem with the matheuristic algorithm.

Thanks to the optimization carried out with the Extend Sim simulator, the model simulates different possible sequencing of jobs and finds the best solution with respect to the given objective function. Therefore it will choose the sequencing of jobs that optimizes the sum between the completion time of all job activities and the tardiness to finish those jobs.

The complete model for optimizing the sequencing of jobs and assignment to teams, on the other hand, has as input only the data of the problem (such as, for example, number of activities to be carried out, number of teams available, time required for each activity, set up time, deadline, ...).

This model therefore optimizes both the assignment to the teams and the sequencing of the jobs, testing every possible alternative during the simulation and choosing the best one compared to what is required by the objective function.

What has just been described is summarized in the following *Figure 20* where:

- The models made are represented within the red rectangles.
- The input (before the left arrows) and output (after the right arrows) data of each of the models are inside the green ellipses.
- The input data of the problem are inside the blue ellipse.

Each of the models has the following input data, which are common in each model:

- $\mathcal{A}$: number of assets, chosen based on the number of maintenance activities to be performed;
- $\mathcal{T}$: number of maintenance teams, depending on the number of teams available;
- $\mathcal{H}_{t_k,\mathrm{T}}$: number of the train-free sub-intervals of the interval $(t_k, t_k + \mathrm{T})$;
- T: maintenance planning horizon;
- ℓ_h: length of the h^{th} sub-interval in $\mathcal{H}_{t_k,\mathrm{T}}$, times available for maintenance activities;
- I_h: initial time of the h^{th} sub-interval in $\mathcal{H}_{t_k,\mathrm{T}}$;
- $s_{i,j,m}$: set-up time, travel time to reach the maintenance activity plus the preparation time to start the activity;
- ω_i: maintenance activity priority of the asset i;
- B: integer suitably chosen to approximate +∞ in the constraints.

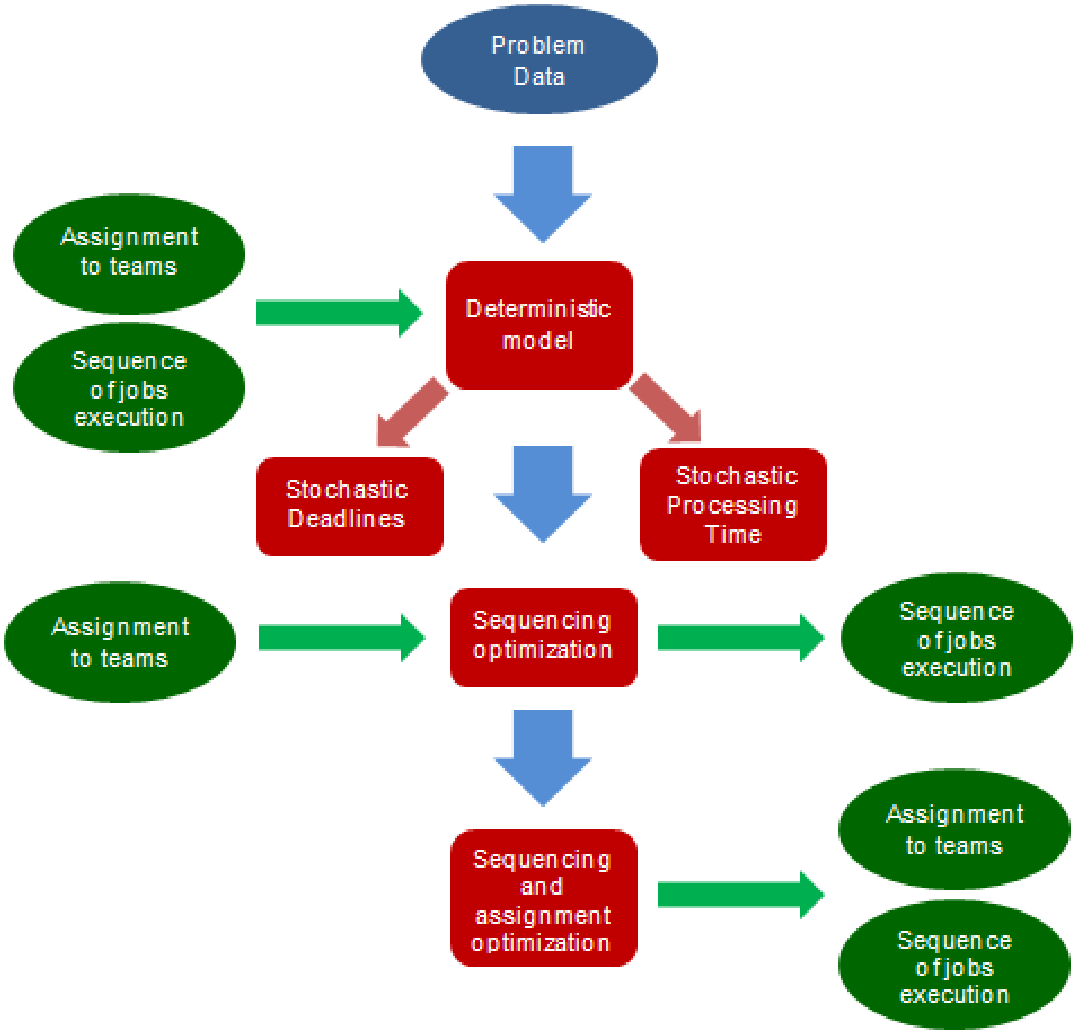


Figure 20: Comparison of models

Furthermore, the following data can be used as variables in certain models or as input in others, the details are shown in the following *Table 4*.

- $p_{i,g}$: processing time of maintenance activity;
- $c_{i,m}$: completion time of maintenance activity;
- $t_{i,m}$: starting time of maintenance activity;
- $\delta h_i(t_k^i)$: hard deadline of the maintenance activity;
- $\delta s_i(t_k^i)$: soft deadline of the maintenance activity;
- q_i: tardiness of maintenance activity;
- $x_{i,j,m}$: binary variable for sequencing maintenance activities;
- $w_{i,m,h}$: binary variable for the assignment of maintenance activities to teams.

Table 4: Data of models

Model	Input Data		Output Data
	Deterministic Data	Stochastic Data	
Deterministic Model	$p_{i,g}$ $c_{i,m}$ $t_{i,m}$ $\delta h_i(t_k^i)$ $\delta s_i(t_k^i)$ q_i $x_{i,j,m}$ $w_{i,m,h}$	–	–
Stochastic Deadlines	$p_{i,g}$ $c_{i,m}$ $t_{i,m}$ $x_{i,j,m}$ $w_{i,m,h}$	$\delta h_i(t_k^i)$ $\delta s_i(t_k^i)$	q_i
Stochastic Processing Time	$\delta h_i(t_k^i)$ $\delta s_i(t_k^i)$ $x_{i,j,m}$ $w_{i,m,h}$	$p_{i,g}$	$c_{i,m}$ $t_{i,m}$ q_i
Sequence of execution of jobs	$p_{i,g}$ $\delta h_i(t_k^i)$ $\delta s_i(t_k^i)$ $w_{i,m,h}$	–	$c_{i,m}$ $t_{i,m}$ q_i $x_{i,j,m}$
Assignment to the teams and sequence of execution of jobs	$p_{i,g}$ $\delta h_i(t_k^i)$ $\delta s_i(t_k^i)$	–	$c_{i,m}$ $t_{i,m}$ q_i $x_{i,j,m}$ $w_{i,m,h}$

6.2 Steps of the simulations

As already anticipated above, the first model is a simple reproduction of the solution of the optimization problem.

The built model simulates the creation of job activities to be carried out, the subdivision into teams that will carry out the maintenance activity, the time to reach the place where the activity is to be carried out, the time to carry out the activity and, finally, the exit from the model, thus the end of the job activity.

For the last two models created, optimization is introduced within the simulation, the model follows the following steps to find the best solution:

- *Initial solution*: the first solution is the one that is given as input in the problem data, it is a feasible solution of the assignment problem but it does not necessarily meet all the requirements of the optimization problem. For example, as a first input, considering that there are 4 teams and 20 work activities available, it is possible to assign the first 5 activities to the first team, the other 5 to the second and so on, going in progressive order.
- *Solution research*: this is the longest phase of the simulation and is the fundamental phase in which the best solution of the problem is sought. In this phase, the model performs various simulations, changing the sequencing of activities and the assignment of teams, and recording the result, if better than the previous solution found.

  The model runs the simulations until the time t that has been set in the simulation is reached and/or until it reaches the set convergence.
- *Best solution*: when the time for the simulation is over, the model chooses the best solution among all those found and the simulation ends.

The *Figure 21* summarizes the steps of the simulation model.

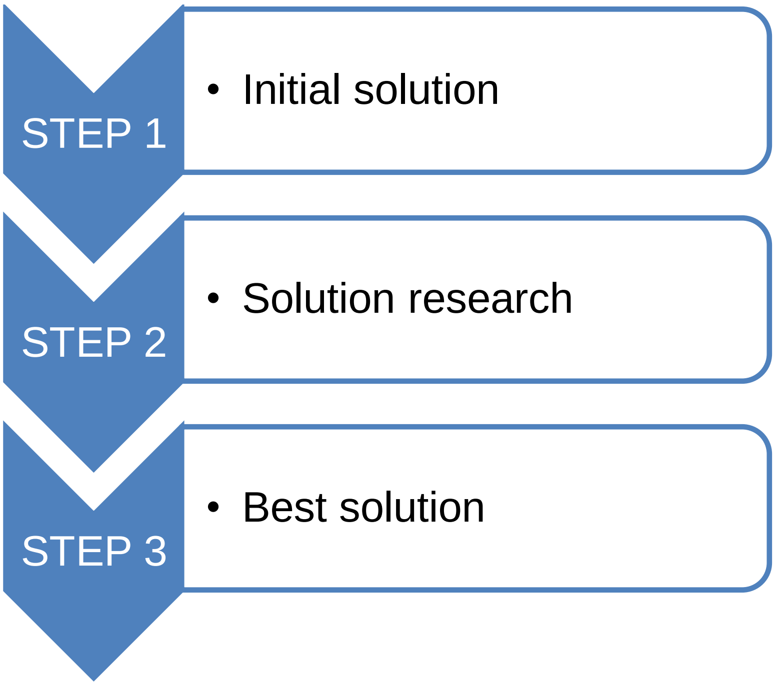


Figure 21: Steps of simulations

6.3 Examples of real applications

Some cases of possible applications to the real world of each of the models made are now described.

Think of the case in which the program of maintenance activities to be carried out in the next few days has already been determined and the agreements with the various companies that have to intervene have already been made.

With the first model it is possible to determine how the completion times of the maintenance activities to be carried out vary, without giving the possibility to redistribute the activities.

For example, in the case in which it is realized that a maintenance activity will take longer than expected, thus increasing its processing time, or in the event that there are delays in the movement to reach the activity, increasing the set up time , by modifying the data and starting the simulation, it is possible to determine how much this will affect the subsequent activities to be carried out and whether it will cause a shift in the next round of activities.

With the model on the sequencing of activities but assigned to teams, for example, it represents the case in which there are several teams available with different qualifications, equipment or skills and to have different jobs to perform. Each team may have the ability to perform a certain type of work but cannot do others. In this way, as input I already have the assignment to the teams but not the sequencing of the activities, so this is a possible application of this model.

The latest model created is the complete model with which it is possible to carry out both the assignment to the teams and the sequencing of activities.

For this model, a real case to apply it to is related to the organization of a certain amount of activities that can be carried out by a certain number of teams, with completely interchangeable skills, qualifications and equipment, and therefore the case in which each team can carry out any work assigned to it.

The implementation of the three models in ExtendSim is shown in the following Chapter 7.

7. Simulation-based optimization models

This chapter explains how the models are created on a small-scale explanatory problem, while in the next chapter the case study used will be reported.

The details of the three models realized with ExtendSim simulator are shown in the following paragraphs.

7.1 Maintenance process simulation model

The first of the models developed uses the matheuristic optimization algorithm solution as input.

Therefore, the following are used as input: the assignment to the teams and the sequence of execution of the jobs, in addition to the problem data.

In the next two paragraph, the model explained in this paragraph is implemented by inserting the optimizer on the simulation model, first by optimizing only the part of the job execution sequence, then also the part of assigning to the teams. *Figure 22* shows the initial model, now is explained in detail how the model is made.

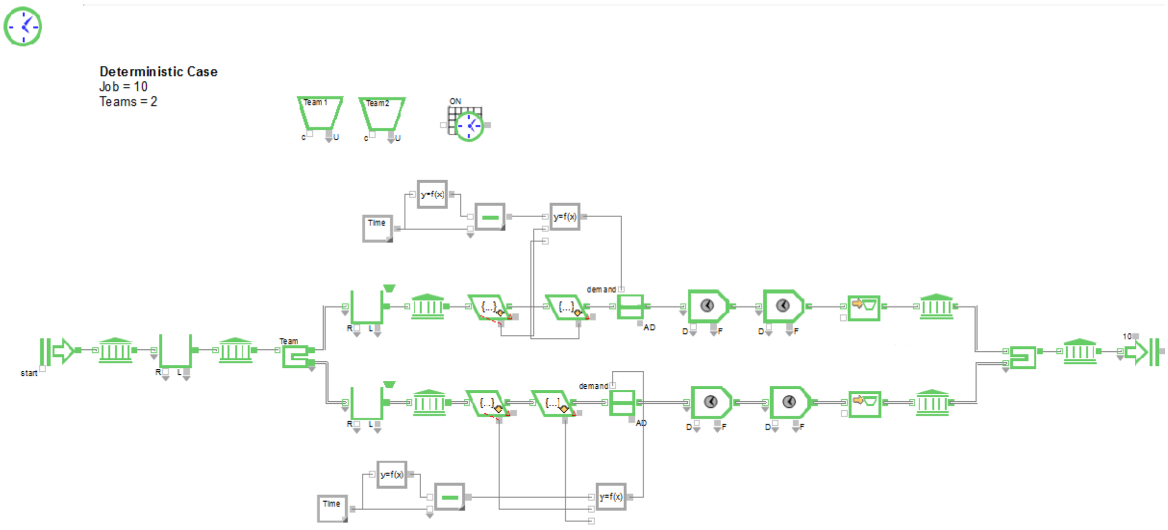


Figure 22: Initial model

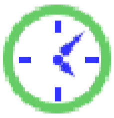
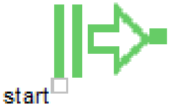
ExtendSim models are made up of blocks and connection. As the model runs, information goes into a block, is processed and/or modified, and is then sent on to the next block via a connection, at the end of simulation it possible to read the results of model.

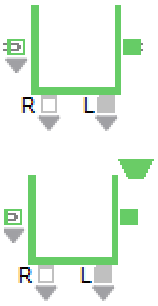
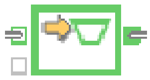
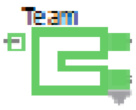
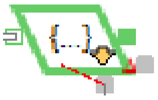
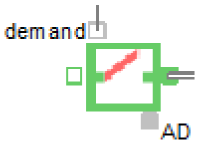
The model simulates the creation of maintenance activities to be done, the process of carrying out each individual activity, by one or the other team, until the end of the activities to be performed. First the job is assigned to a team, secondly it is assigned a processing sequence.

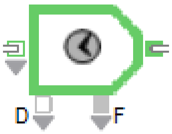
For the first job to be carried out, the team starts from the maintenance depot, while for the following jobs it goes directly from one point to another on the network.

Each block in ExtendSim represents a portion of the problem that is being modelled. Every block have name that indicate the function it perform, in *Table 5* are explained their function.

Table 5: ExtendSim Blocks

	***Executive*** It is used to set the duration of the simulation, in this case the simulation ends when all the items leave the system. It is also used for attribute management, for renaming or deleting attributes or locating where they are used in the model.
	***Create*** Enters items in the model. Items are maintenance job activity that must be execute, in this reduced explanation model the jobs are 10 and they are all introduced into the model at the same instant of time (initial time = 0). Newly created items have many attributes: • Soft Deadline, • Hard Deadline, • Position: that is simply the progressive number of a job to execute, • Processing Time: work time of maintenance activity on an asset by a maintenance team, • Travelling Time: travel and setup time of maintenance activity

	from an asset i to an asset j by a maintenance team, • Team: team at which the job is assigned, • Sequence: order of execution of the job.
	***History*** This block is used to get information of the item that pass through it, it is possible to view its attributes and arrival time in the block.
	***Queue*** A queue stores item until is its turn of processing. Are used two type of queue: • Sorted queue: sorts items based on their attribute choose, • Resource pool queue: allows an item to continue when a resource is available to process it, in this case a team.
	***Resource Pool*** Stores resources for the model, in this case resources are 2 maintenance teams. The team are taken by the Resource pool queue and released by the Resource Pool Release block at some later point in the model, when maintenance activity is end.
	***Resource Pool Release*** Released the Resource at when maintenance activity is end in order to can be used for another job.
	***Select Item Out*** This block sending item to that queue based on the team to which item is assigned.
	***Get*** Takes the value of item attributes (processing time and travel time) to use in next block.
	***Gate*** Controls the flow of items in a portion of the model based on conditions of block "Shift" (conditional gating). Allows a team to

	start set up for a job.
	***Shift*** When it is set "off" it means that are trains circulation so maintenance activities cannot be executed, when it is set to "on" the activities can be executed. The maintenance time horizon is 300 min (5 hours) overnight.
	***Activity*** Holds on one item and passes it out based on: • The travel time for the first activity, • The process time for the second activity.
	***Select Item In*** Merges items from two input to one output.
	***Exit*** Passes items out of the simulation. The total number of items crossing this block is reported on the number over. When all the items/jobs pass through this block, the maintenance activities are finished and the simulation ends.

To determine if at the end of a maintenance activity there is still time in the work shift to start and finish another one (it is not allowed to start the travel or the setup if the activity cannot be concluded), the difference between the time remaining at the end of the shift and the sum of travel time and processing time is calculated, if the time is enough the activity is performed, otherwise it is waited for the next work shift to carry it out.

This condition allows opening and closing of the "Gate" and is inserted as input of this block as shown in *Figure 23*.

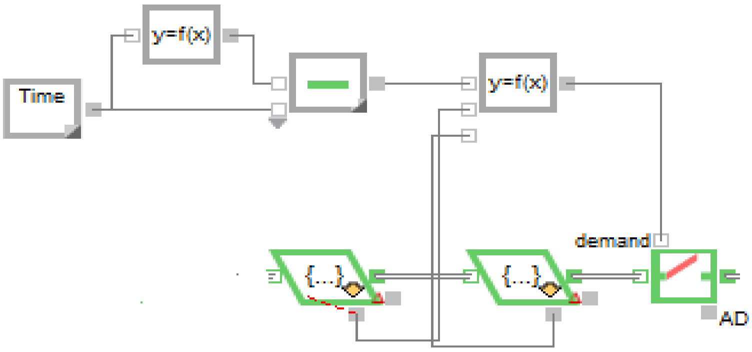


Figure 23: Open and close gate condition

In this model as it is made there are no variables, because all the values are assigned in input, so running the simulation it always get the same result.

As already mentioned in Chapter 5.3, there are some variables that are considered deterministic, which actually have a stochastic nature:

- The deadline of interventions,
- The processing time of maintenance activity on asset i by maintenance team g.

Below are two models that use the uncertainty of these two variables in the initial model.

7.1.1. Stochastic deadlines

The *Figure 24* shows the model with the insertion of the stochastic deadlines according to the Equation (4.3.3) introduced previously.

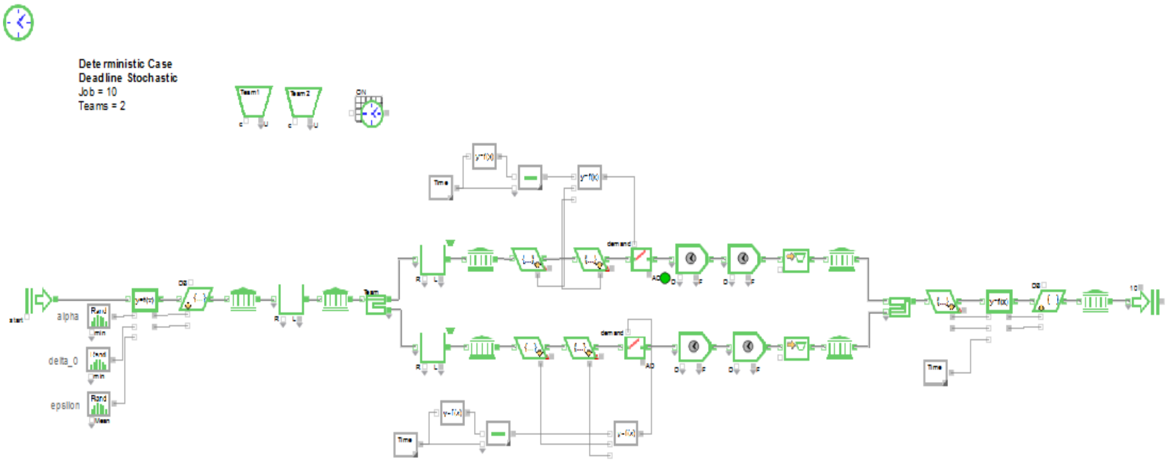


Figure 24: Model with stochastic deadlines

With the help of the blocks in the *Figure 25*, at the beginning of the simulation hard and soft deadlines are assigned in a stochastic way.

The hard deadline is calculated with the formula $\overline{\tau_\iota} = \frac{1}{\alpha_i} \ln \frac{\bar{\delta}-\varepsilon}{\delta_i\left(\tau_k^i\right)}$, in the model the value is assigned to the variables alpha α, epsilon ε, delta δ and delta marked $\bar{\delta}$. The soft deadline is set as 0.3 times the hard deadline.

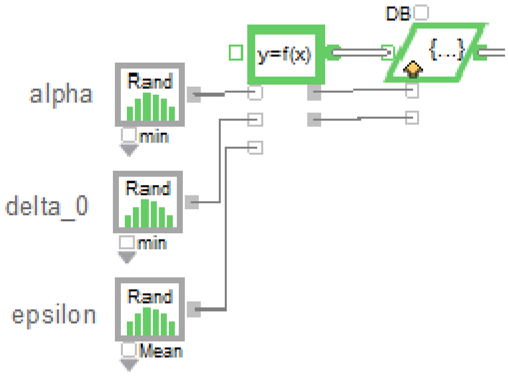


Figure 25: Uncertainty insertion in the deadline

7.1.2. Stochastic processing time

The *Figure 26* shows the model with the insertion of the stochastic processing time.

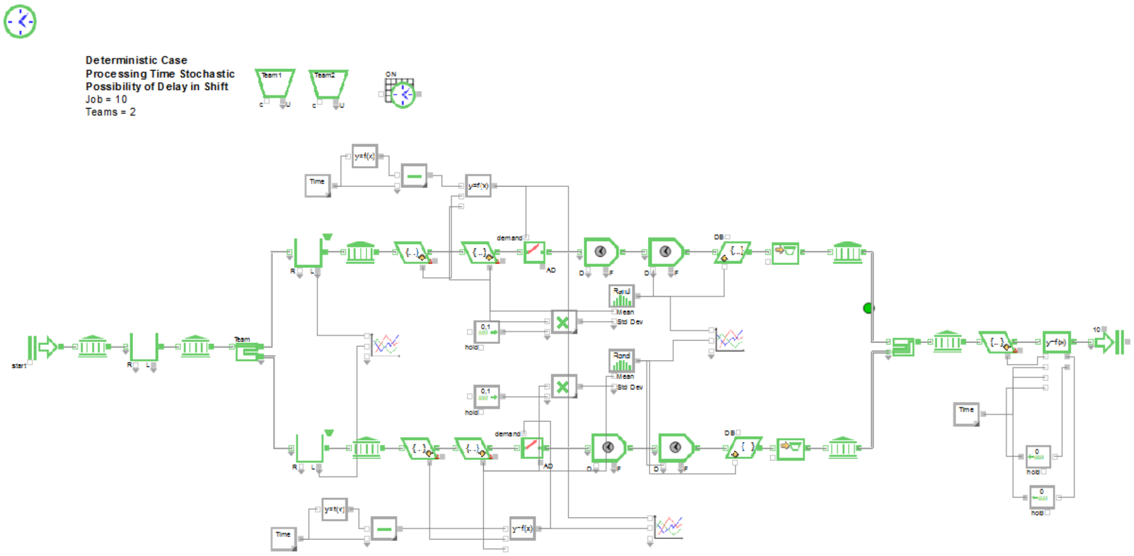


Figure 26: Model with stochastic processing time

The uncertainty in the processing time is added by assigning a normal distribution to the initial processing time, using the Equation (5.3.1).

The mean of the distribution is the initial processing time, the standard deviation is fixed as 0.1 of the initial processing time.

The *Figure 27* shows the blocks that are used for the purpose described above.

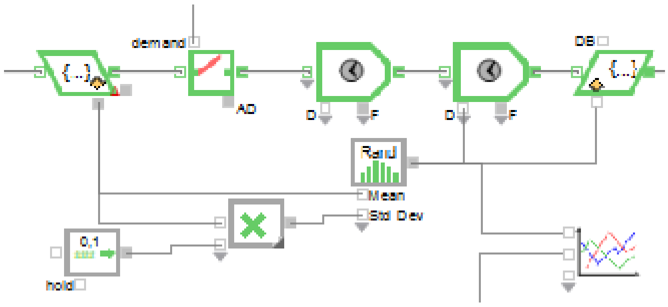


Figure 27: Stochastic insertion in the processing time

7.2 Simulation-based optimization – Sequencing optimization

This paragraph shows how the optimizer is inserted on ExtendSim, and how job sequencing is found through simulation. *Figure 28* shows how to modify the model for the simulation with optimizer. The three modifications to the model are circled in red in the figure, under the detail of each new block inserted.

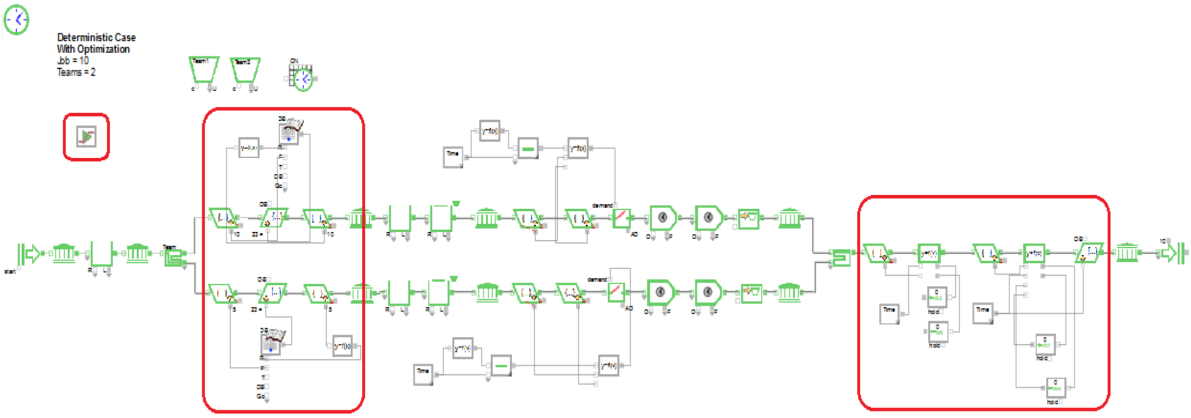


Figure 28: Model with sequencing optimization

To carry out the optimization with ExtendSim the "Optimizer Block" is inserted, it is represented in *Figure 29*.

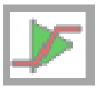


Figure 29: Optimizer Block

Inside the optimizer block the objective function is inserted, taken from (5.2.1), in this model the assignment to the teams is considered as input data.

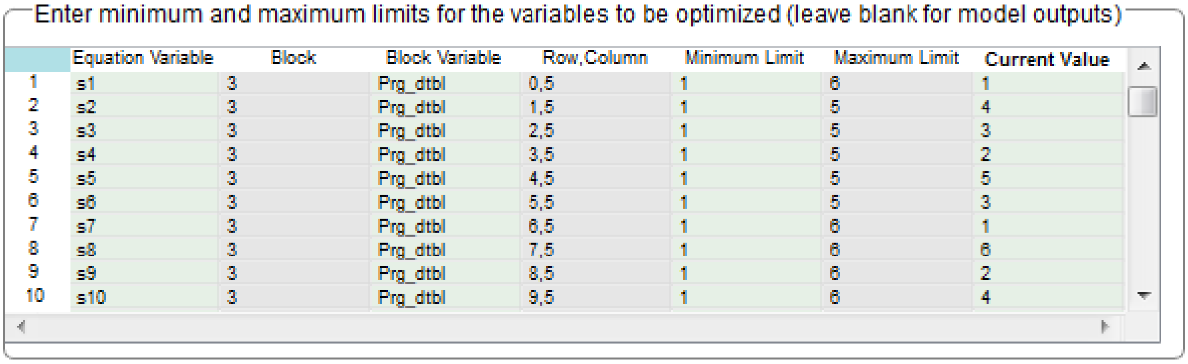
The objective function inserted in the model is therefore:

$$\min\left(\sum_{m=1}^{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{A}|}\omega_i c_{i,m}+\sum_{i=1}^{|\mathcal{A}|}\omega_i q_i\right) \tag{6.2.1}$$

given by the minimization of the completion time of each maintenance activity added to the tardiness compared to the soft deadline of each maintenance activity. The optimizer block also contains all the variables of the simulation and some constraints that are not already intrinsic in the construction of the model.

The completion time, the starting time of the maintenance activity and the tardiness are variables calculated during the simulation, the sequence attributed to each job is also set as a variable. During the simulations the sequence is varied several times, trying to find which is the optimal one with respect to the objective function and the imposed constraints.

The *Figure 30* shows the setting of the "sequence" variable for each job, the maximum and minimum limits of the variables to be optimized and the current value assigned by the simulation.



Enter minimum and maximum limits for the variables to be optimized (leave blank for model outputs)

	Equation Variable	Block	Block Variable	Row,Column	Minimum Limit	Maximum Limit	Current Value
1	s1	3	Prg_dtbl	0,5	1	6	1
2	s2	3	Prg_dtbl	1,5	1	5	4
3	s3	3	Prg_dtbl	2,5	1	5	3
4	s4	3	Prg_dtbl	3,5	1	5	2
5	s5	3	Prg_dtbl	4,5	1	5	5
6	s6	3	Prg_dtbl	5,5	1	5	3
7	s7	3	Prg_dtbl	6,5	1	6	1
8	s8	3	Prg_dtbl	7,5	1	6	6
9	s9	3	Prg_dtbl	8,5	1	6	2
10	s10	3	Prg_dtbl	9,5	1	6	4

Figure 30: Variables to be optimized

The second modification to the basic model is the allocation of travel time directly in the simulation, based on the previous maintenance activity performed and the activity to be performed in that moment.

In the *Figure 31* it is possible to see the blocks for storing the previous job performed and the current one to be performed, the two values are given as input to the database in which the travel time between each job are stored.

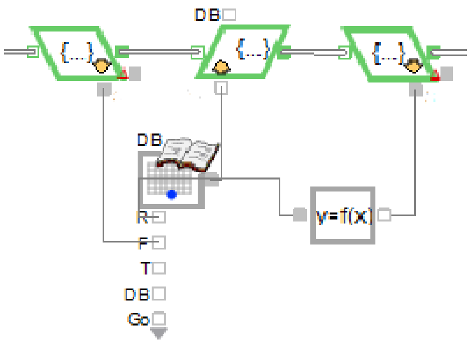
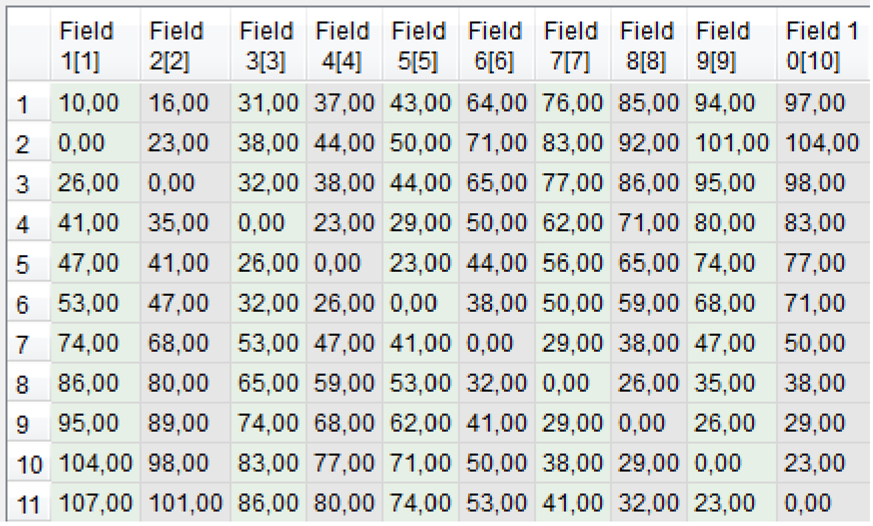


Figure 31: Travel time allocation

With the two inputs, row and column of the matrix, the travel time/setup time to be associated with the activity is found in the database in the *Figure 32*.



	Field 1[1]	Field 2[2]	Field 3[3]	Field 4[4]	Field 5[5]	Field 6[6]	Field 7[7]	Field 8[8]	Field 9[9]	Field 1 0[10]
1	10,00	16,00	31,00	37,00	43,00	64,00	76,00	85,00	94,00	97,00
2	0,00	23,00	38,00	44,00	50,00	71,00	83,00	92,00	101,00	104,00
3	26,00	0,00	32,00	38,00	44,00	65,00	77,00	86,00	95,00	98,00
4	41,00	35,00	0,00	23,00	29,00	50,00	62,00	71,00	80,00	83,00
5	47,00	41,00	26,00	0,00	23,00	44,00	56,00	65,00	74,00	77,00
6	53,00	47,00	32,00	26,00	0,00	38,00	50,00	59,00	68,00	71,00
7	74,00	68,00	53,00	47,00	41,00	0,00	29,00	38,00	47,00	50,00
8	86,00	80,00	65,00	59,00	53,00	32,00	0,00	26,00	35,00	38,00
9	95,00	89,00	74,00	68,00	62,00	41,00	29,00	0,00	26,00	29,00
10	104,00	98,00	83,00	77,00	71,00	50,00	38,00	29,00	0,00	23,00
11	107,00	101,00	86,00	80,00	74,00	53,00	41,00	32,00	23,00	0,00

Figure 32: Travel time database

The third circle contains the blocks that have been inserted to count different values such as: completion time of a job, total completion time of all jobs, tardiness, saved time.

These values are used as a reference by the Optimizer Block to carry out the simulation to find the optimization of the problem.

In the first group of blocks in the *Figure 33* , the overcoming or not of the hard deadline is calculated, which is a very important constraint that must be respected. In the second group of blocks in the figure there is instead the delay, or on the opposite the time saved, with respect to the soft deadline.

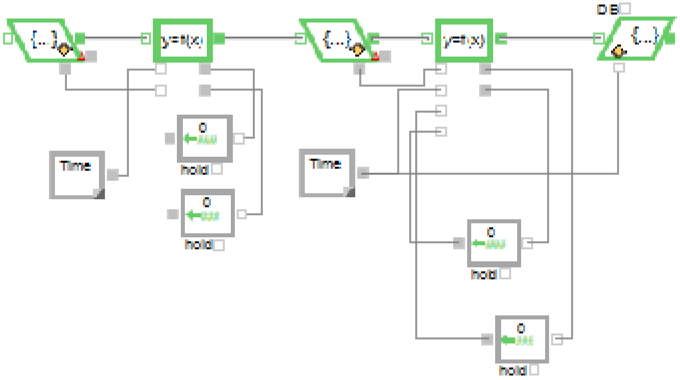


Figure 33: Blocks for counting different elements of the problem

Now let's compare how each constraint of the optimization problem introduced in Chapter 5.2 is inserted into the simulation with ExtendSim.

Some constraints are already based on how the model is built, others are inserted with some blocks, still others are inserted directly into the Optimizer Block, they are explained in detail below.

- (5.2.2) – Completion times of the maintenance activities:

  it is calculated during the simulation when the activity to be carried out crosses the entire model.
- (5.2.3) – Maintenance activity on each asset is completed before the relevant hard deadline:

  this constraint is inserted in the objective function with a very high weight to give importance to this constraint and make it mandatory.
- (5.2.4) – Tardiness of the maintenance activities:

  like the completion time, it is calculated during the simulation when the activity to be carried out crosses the entire model.
- (5.2.5) – Precedence relation among the activities on the assets i and j:

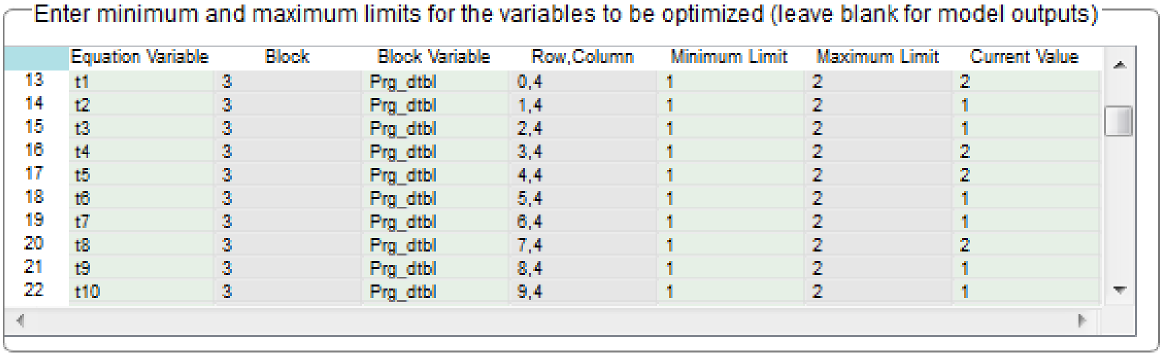
  it derives from the layout of the model, a job is executed once the previous one is finished.
- (5.2.6) – Initial set-up times:

  for the first job that is executed the setup time / travel time is considered from the starting point of the maintenance team, the deposit, this is present in the travel time database in the first row.

- (5.2.7) – At most one activity is scheduled as the first work of each maintenance team:
  due to the way the model is built, only one activity can be performed at a time.
- (5.2.8) and (5.2.9) – every maintenance activity has exactly one predecessor and one successor, respectively:
  it is respected as is the layout of the model.
- (5.2.10) – a predecessor/successor pair of activities has to be assigned to the same maintenance team m:
  as above, it is respected as is the configuration of the model.
- (5.2.11) – if an activity is performed in the train-free interval h, it starts after the beginning of such an interval:
  with Gate and Shift blocks all the free time from train traffic is regulated and maintenance activities can be carried out.
- (5.2.12) – all maintenance activities finish within the train-free interval:
  it is regulated from Gate and Shift blocks.
- (5.2.13) – each activity is only planned in a single time interval
  before starting the move to a new activity it is calculated if there is enough time to start and finish the next job, otherwise the next shift is waited.
- (5.2.14), (5.2.15) and (5.2.16) – define the problem variables:
  they vary in the simulation.

7.3 Simulation-based optimization – Sequencing and assignment optimization

The case of sequencing and assignment optimization of jobs is achieved by including the possibility of varying the assignment to teams within the model.
The layout of the model remains the same as in the case previously described in the *Figure 28*. While within the variables of the model there is also the possibility of varying the assignment to teams as shown in the *Figure 34*.



Enter minimum and maximum limits for the variables to be optimized (leave blank for model outputs)

	Equation Variable	Block	Block Variable	Row,Column	Minimum Limit	Maximum Limit	Current Value
13	t1	3	Prg_dtbl	0,4	1	2	2
14	t2	3	Prg_dtbl	1,4	1	2	1
15	t3	3	Prg_dtbl	2,4	1	2	1
16	t4	3	Prg_dtbl	3,4	1	2	2
17	t5	3	Prg_dtbl	4,4	1	2	2
18	t6	3	Prg_dtbl	5,4	1	2	1
19	t7	3	Prg_dtbl	6,4	1	2	1
20	t8	3	Prg_dtbl	7,4	1	2	2
21	t9	3	Prg_dtbl	8,4	1	2	1
22	t10	3	Prg_dtbl	9,4	1	2	1

Figure 34: Variable assignment to teams

8. Case study and results

In this chapter, some of the results obtained are reported and discussed. In relation to the 3 models explained in Chapter 6 and Chapter 7, 3 different analyses are reported:

- **Comparison of deterministic and stochastic scenarios:** Starting from the model introduced in Paragraph 7.1, a comparison is made with the initial deterministic case and the case with uncertainty in the processing time and in the deadline.
- **Sensitivity analysis:** Starting from the model in Paragraph 7.2, the weight of the two components of the objective function (completion time and tardiness) are varied and the different results are compared.
- **Problem size variation:** Starting from the model in Paragraph 7.3, the number of teams, or the number of jobs or both are varied and the results are compared.

As mentioned before, the case study regards the activity of maintenance of a railway line during the stop of railway traffic in the night. The work shifts are therefore scheduled every night from midnight to five, for a duration of 300 minutes.

As real-world application, the Italian railway network in the Northwest part is considered and the maintenance planning of this part of the line is simulated.

It is assumed the need of 42 maintenance activities of different assets along the entire network as shown in *Figure 35*, there are 4 teams available to do maintenance activities.

The 4 teams that execute the maintenance activities have different equipment and therefore: Team 2 and Team 4 are equipped with more performing machines and their travel time is half the travel time of Team 1 and Team 3. Hence Team 2 and Team 4 reach the workstations in a shorter time.

The depot, from which the teams start the beginning of the first night, is fixed in Genoa.

Each maintenance activity has a different duration depending on the type of work to be performed, this time is considered in the model as Processing Time.

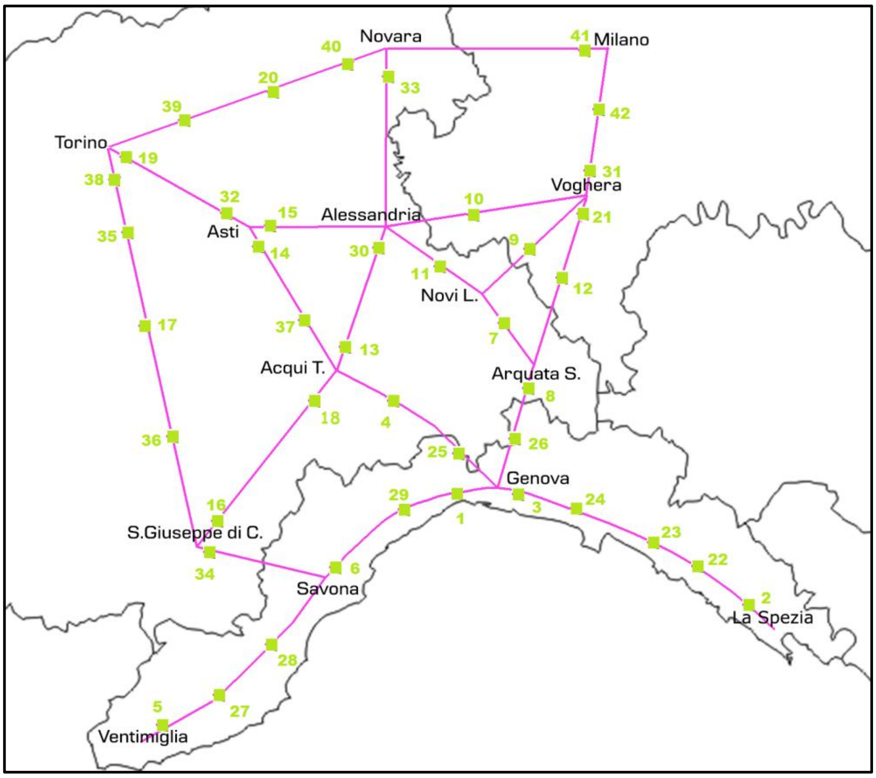


Figure 35: Distribution of jobs to be maintained along the railway network

By applying the Risk-Based Approach described in Paragraph 4.3, the Soft Deadlines and the Hard Deadlines to be used in the simulations are determined.

8.1 Comparison between deterministic and stochastic scenarios

The objective of this model is to compare the result of deterministic optimization with the same result but with the use of stochastic variables.

In detail, two scenarios are analysed:

- **Scenario 1:** From deterministic deadlines to stochastic deadlines,
- **Scenario 2:** From deterministic processing times to stochastic processing times.

For both scenarios the model introduced in Paragraph 7.1 is used.

8.1.1. *Scenario 1 - Deadline*

Stochastic deadlines are setting with the Equation (4.3.3) introduced previously, hard deadline is calculated with the formula $\overline{\tau_\imath} = \frac{1}{\alpha_i} \ln \frac{\bar{\delta}-\varepsilon}{\delta_i\left(\tau_k^i\right)}$ in the model and soft deadline is set as 0.3 times the hard deadline.

The variables alpha α, epsilon ε and delta δ_0 are defined as:

- α is a uniform distribution between 0.05 and 0.12;
- ε is a normal distribution with mean 0 and standard deviation 0.2;
- δ_0 is a uniform distribution between 0.8 and 1.5.

The *Table 6* shows the increase in the sum of the "Hard Delay" and "Soft Delay", with the variation of the alpha, epsilon and delta variables, calculated respectively as:

- Hard Delay is the difference between the Hard Deadline and the job Completion Time, it is the delay compared to the hard deadline. The Total Hard Delay is the sum of the Hard Delay of all jobs.
- Soft Delay is the difference between the Soft Deadline and the job Completion Time, it is the delay with respect to the soft deadline. The Total Soft Delay is the sum of the Soft Delay of all jobs.

Table 6: Some results with variation of $\boldsymbol{\alpha}$ *and* $\boldsymbol{\delta_0}$

	α	δ_0	Total Hard Delay (min.)	Total Soft Delay (min.)
Run 1	0,05	0,8	0	0
Run 2	0,06	0,9	0	2325,44
Run 3	0,07	1,0	0	9648,05
Run 4	0,08	1,1	0	24331,48
Run 5	0,09	1,2	0	38270,57
Run 6	0,10	1,3	0	50690,77
Run 7	0,11	1,4	**1927,64**	69205,69
Run 8	0,12	1,5	**3478,59**	77341,27

From the table it can be seen that as the alpha and delta values increase, the deadlines decrease, thus increasing the delay in the execution of the jobs. The

Soft Deadline is overcome almost immediately, and this is a condition granted by the constraints of the problem. The Hard Deadline is exceeded in the last two examples shown in the table, this condition is not admissible for the problem that requires not to exceed the hard deadline.

Table 7 shows the results of the deterministic problem, which must be compared with the results of the simulations performed.

As already mentioned, the Hard Delay must necessarily be equal to zero because due to the constraints of the problem it is not possible to exceed it, while the Soft Delay can assume a certain positive value.

Table 7: Results of Deterministic Problem

Total Hard Delay (min.)	Total Soft Delay (min.)
0	13552

8.1.2. Scenario 2 - Processing time

Stochastic processing time are setting with the Equation (5.3.1) introduced previously.

By modifying some settings of the Activity block two types of simulations are performed:

- Run from 1 to 5: If there is not enough time to start and finish the maintenance task, the team waits for the next shift to get the job done. This happens in reality, for example, when the team realizes even before starting that time would not be enough because it takes more time than expected or they do not have the equipment ready and it would be necessary to waste more time preparing them.
- Run from 6 to 10: When it is realized late that the maintenance activity takes longer than expected, in this case the team has the opportunity to exceed the interval of passage of the trains. In the real world, it happens if a job isn't done and the team doesn't have enough time to secure the line for the regular passage of trains. This case involves delays in traffic and high costs for loss of regular service.

Results for Run from 1 to 5:

In this simulation it can be seen how some jobs are terminated earlier or later than the deterministic case, due to the variation imposed in the processing time. As an example, *Figure 36* shows the graph with the completion times of Team 1 for each Run performed. Each job is represented with a different color, gray is the time between the end of a job and the start of the next job in the next shift. Comparing the results of the 5 Runs with the results of the deterministic case of *Figure 37* it can be observed that:

- Run 1 and Run 2: Team 1 starts and ends the jobs in the same shifts as the comparison case, the delay of the processing time has no influence on the change of shifts.
- Run 3 and Run 5: Team 1 up to the third shift maintains the same trend as the comparison case, from the fourth night it even anticipates a job and in the end it saves a work shift.
- Run 4: Team 1 is late on the first night of shift and consequently all jobs are delayed for the following days. In the sixth shift the team manages to recover some work, managing to complete the jobs as in the case of comparison.

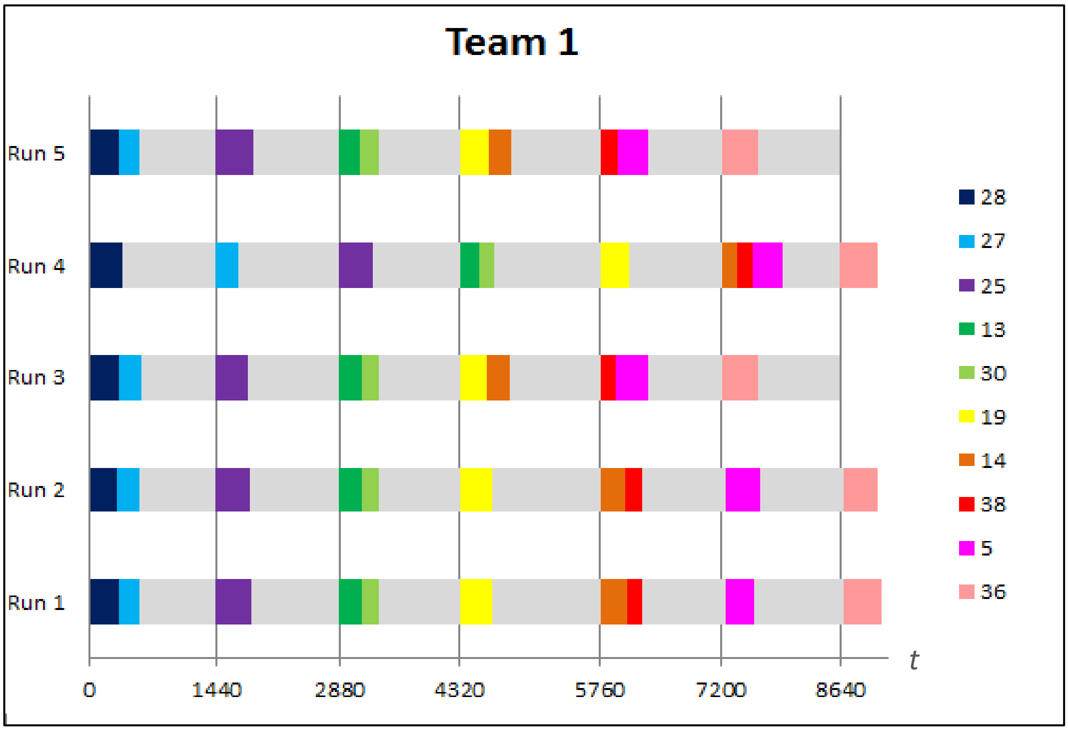


Figure 36: Jobs Sequencing of Team 1- Run from 1 to 5

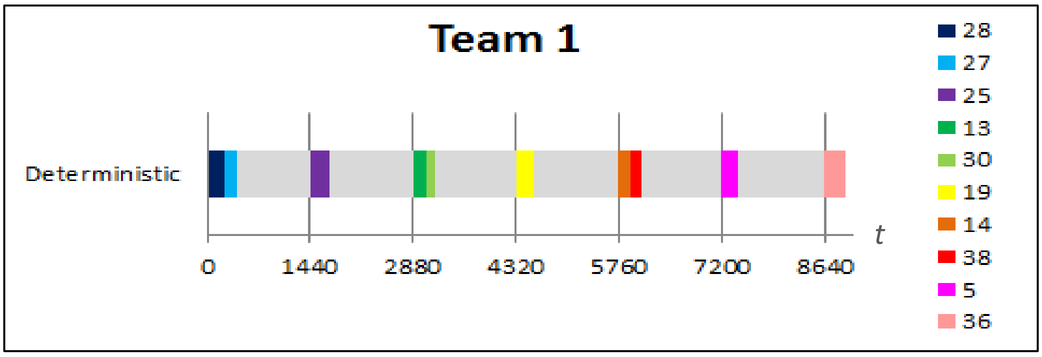


Figure 37: Jobs Sequencing of Team 1 - Deterministic Problem

Regarding the other teams, it is observed that Team 2 always remains aligned with the comparative case, except for Run 3 where from the fourth shift onwards it accumulates a small delay. Team 3 remains aligned to the comparative case except for a small delay in Run 4 for the last shift. Finally, for Team 4 it is observed that in Run 1, Run 2 and Run 4 it is slightly ahead of the deterministic case.

Results for Run from 6 to 10:

In the simulation with Runs from 6 to 10, it is noted that the jobs for which the maintenance activity is delayed, are in any case terminated within the maintenance shift, thus causing a delay in the restart of railway traffic.

In Run 6 and Run 10 there are two jobs that exceed the time of the interruption. In Run 8 and Run 9 only one exceeds the time of the interruption. For Run 7, there are even 3 jobs that terminate the maintenance activity beyond the established time. All these delays in real world lead to an economic loss for the railway company.

Table 8 lists the 3 jobs of Run 7 for which the Teams are late with regard to maintenance activities, they result in delays of 10, 9 and 16 minutes respectively.

Table 8: Job out of shift for Run 7

Job number	Team	Arrival Time	End of shift	Minutes late
27	1	310	300	10
42	4	1749	1740	9
20	3	7516	7500	16

Figure 38 shows the graph with the completion times of Team 1 for each Run performed. Each job is represented with a different color, gray is the time between the end of a job and the start of the next job in the next shift.

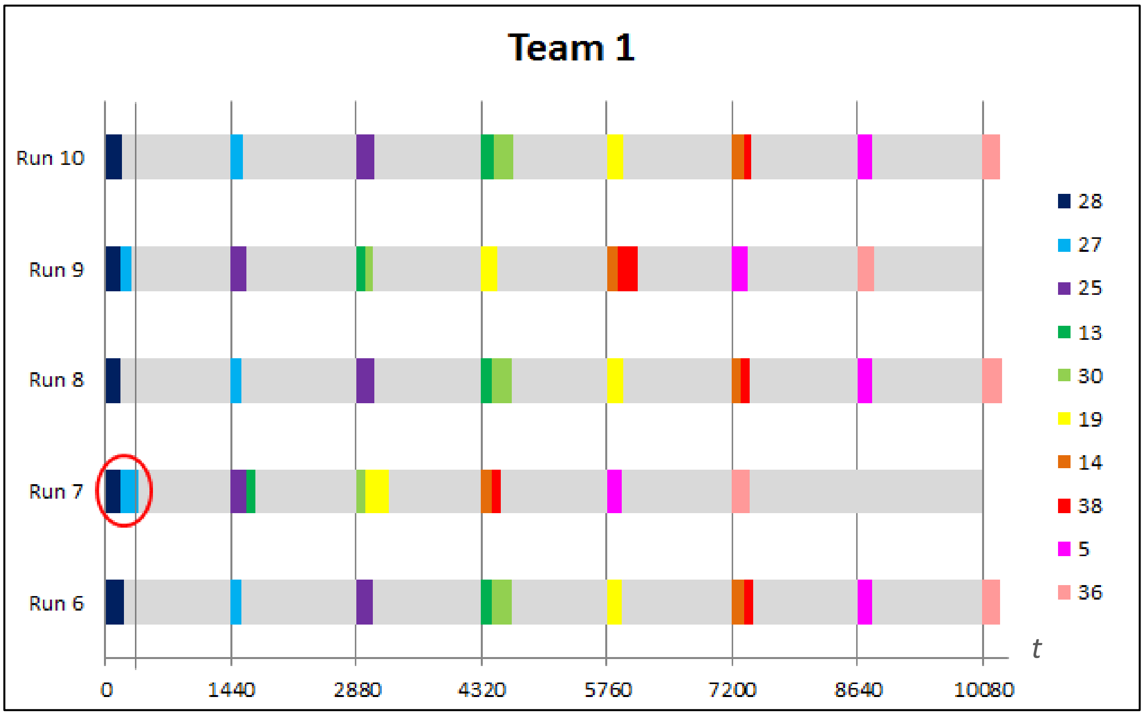


Figure 38: Jobs Sequencing of Team 1- Run from 6 to 10

To better understand what the event of a delay in the completion of an activity, in the red circle in *Figure 39* it is possible to see job 27 (light blue rectangle) which is finished 10 minutes later than the end of the shift (300 min), and therefore during the period in which railway traffic should restart.

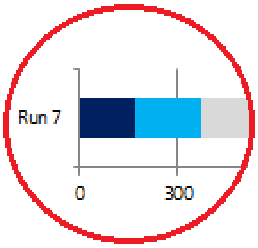


Figure 39: Detail of Run 7 jobs 28 and 27

8.2 Sensitivity analysis – variation of weights in the objective function

The objective of this model is to compare the different result obtained changing the weight of the two components of the Objective Function: Completion Time and Tardiness. The model with the optimization of job sequencing, introduced in Paragraph 7.2, is used.

To the Objective Function, given by Equation (5.2.1), two weights $\alpha_1 \in \mathbb{R}_+$and $\alpha_2 \in \mathbb{R}_+$ are added, the equation becomes:

$$\min\left(\alpha_1 \sum_{m=1}^{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{A}|} \omega_i c_{i,m} + \alpha_2 \sum_{i=1}^{|\mathcal{A}|} \omega_i q_i\right) \qquad (6.2.1)$$

Where:

- $c_{i,m} = Completion\ Time$
- $q_i = Tardiness$

Table 9 indicates the results of different simulations found by changing the two values of the weights.

Table 9: Change of the Objective Function with variation of $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ weights

α_1	α_2	$\sum_{m=1}^{\|\mathcal{T}\|}\sum_{i=1}^{\|\mathcal{A}\|} \omega_i c_{i,m}$	$\sum_{i=1}^{\|\mathcal{A}\|} \omega_i q_i$
Equal weights		0%	0%
Low	High	+11%	-15%
Low	Medium	+2%	+86%
Medium	Low	+5%	+13%
High	Low	-0,08%	+55%

Figure 40 shows the trend of the component of the objective function relating to completion time. *Figure 41* shows the trend of the component of the objective function relating to tardiness.

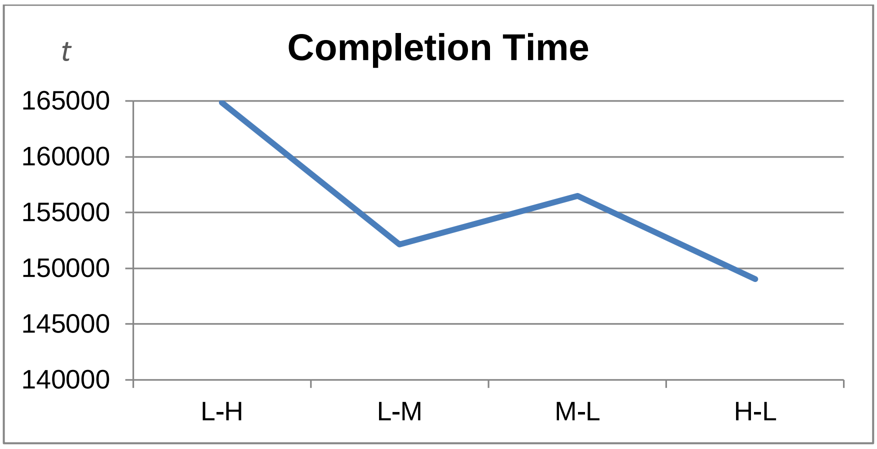


Figure 40: Completion Time trend

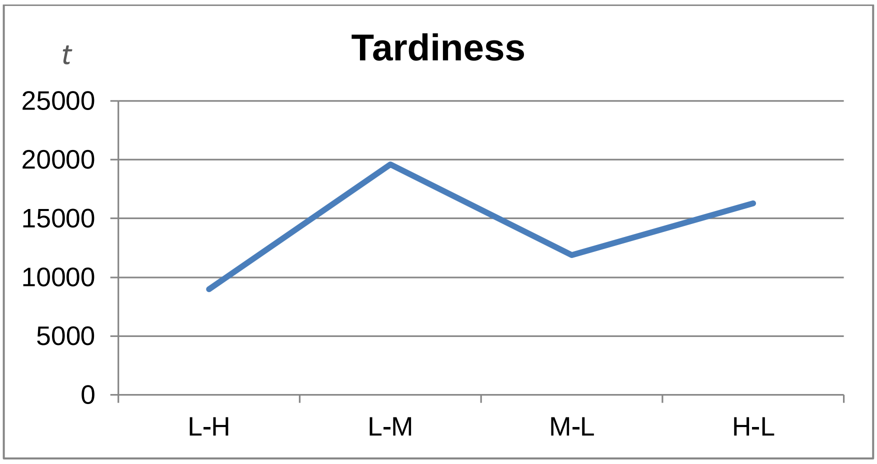


Figure 41: Tardiness trend

It can be observed that by varying the two weights of the components of the objective function, one could expect a totally increasing or decreasing trend of the two components. This does not happen because both components depend on the completion time which therefore affect both components.

In general, comparing the table and the graphs, it can be observed that when α_2 weight is high, the tardiness component decreases, this case is reported in the first row of the *Table 9* and in the first column of the graph of *Figure 41*. In this case the component related to the completion time assumes the maximum value, with a low α_1 weight.

Conversely, when the α_1 weight is high, the completion time component decreases, as reported in the last row of the *Table 9* and in the last column of the graph in *Figure 40*. The tardiness component takes on the maximum value.

Clearly, the change in the weights of the two components of the objective function cause the reordering of the sequencing of jobs during work shifts.

By way of example, the following tables show the change in the sequencing of Team 1 according to the change in the weights in the objective function.

Table 10: Team 1 - Low Weight of Completion Time and High Weight of Tardiness

Job number	Sequence	Processing Time	Travelling Time	Cycle Time	Hard Deadline	Soft Deadline	Hard Delay	Soft Delay
13	1	101	47	148	8640	1440	0	0
30	2	70	23	241	15840	1440	0	0
27	3	110	100	1650	15840	1440	0	210
28	4	126	13	3019	15840	1440	0	1579
25	5	99	70	4489	15840	1440	0	3049
5	6	92	117	5969	15840	8640	0	0
14	7	106	144	7450	23040	8640	0	0
38	8	73	23	8736	37440	15840	0	0
19	9	136	10	8872	37440	15840	0	0
36	10	81	29	10190	37440	15840	0	0
				50764				4838

Table 11: Team 1 - Low Weight of Completion Time and Medium Weight of Tardiness

Job number	Sequence	Processing Time	Travelling Time	Cycle Time	Hard Deadline	Soft Deadline	Hard Delay	Soft Delay
30	1	70	45	115	15840	1440	0	0
36	2	81	92	288	37440	15840	0	0
19	3	136	29	1605	37440	15840	0	0
14	4	106	23	1734	23040	8640	0	0
27	5	110	121	3111	15840	1440	0	1671
5	6	92	23	4435	15840	8640	0	0
28	7	126	47	4608	15840	1440	0	3168
25	8	99	70	5929	15840	1440	0	4489
13	9	101	26	6056	8640	1440	0	4616
38	10	73	65	7338	37440	15840	0	0
				35219				13944

Table 12: Team 1 - Medium Weight of Completion Time and Low Weight of Tardiness

Job number	Sequence	Processing Time	Travelling Time	Cycle Time	Hard Deadline	Soft Deadline	Hard Delay	Soft Delay
13	1	101	47	148	8640	1440	0	0
25	2	99	26	273	15840	1440	0	0
27	3	110	93	1643	15840	1440	0	203
28	4	126	13	3019	15840	1440	0	1579
5	5	92	47	3158	15840	8640	0	0
36	6	81	120	4521	37440	15840	0	0
30	7	70	92	5922	15840	1440	0	4482
38	8	73	36	6031	37440	15840	0	0
14	9	106	23	7329	23040	8640	0	0
19	10	136	23	7488	37440	15840	0	0
				39532				6264

Table 13: Team 1 - High Weight of Completion Time and Low Weight of Tardiness

Job number	Sequence	Processing Time	Travelling Time	Cycle Time	Hard Deadline	Soft Deadline	Hard Delay	Soft Delay
25	1	99	21	120	15840	1440	0	0
13	2	101	26	247	8640	1440	0	0
30	3	70	23	1533	15840	1440	0	93
14	4	106	13	1652	23040	8640	0	0
19	5	136	23	3039	37440	15840	0	0
38	6	73	10	3112	37440	15840	0	0
27	7	110	130	4560	15840	1440	0	3120
28	8	126	13	5899	15840	1440	0	4459
36	9	81	73	6053	37440	15840	0	0
5	10	92	120	7412	15840	8640	0	0
				33627				7672

From *Table 10* it is possible to observe how the jobs that have a shorter deadline are executed in the first few days, while in the following tables it is possible to see how the importance of the Deadline decreases, to reduce the Completion Time of the activities (in the table referred to as "Cycle Time").

Therefore, the weights can be chosen in order to give some priorities according to the maintenance's goals.

As already mentioned, is important to underline that the choice of the weights cannot influence the safety of the railway: the Hard Deadline will never be exceeded, whereby a tolerable risk level is always guaranteed. In the tables above, this is represented by the "Hard Delay" column, to ensure this constraint must always be zero.

8.3 Variation in the problem size – travel time, jobs and teams

Using the model described in Paragraph 7.3 as a starting point, the objective of this model is to compare different scenarios that are created by changing the number of some of the components of the model.

The model used is the one that realizes the complete optimization of the maintenance plan, realizing the assignment to the teams and the sequencing of the jobs.

Specifically, 3 different scenarios are analysed:

- Scenario 1: Variation of the transport speed of the 4 Teams performing maintenance;
- Scenario 2: Change in the number of Teams working on a fixed number of jobs;
- Scenario 3: Variation in the number of Jobs to be carried out by the 4 maintenance teams.

8.3.1. Scenario 1 – Travel time

As mentioned above, the 4 teams that execute the maintenance activities have different equipment and therefore the four teams have different speed, and movement time, in the model defined as "Travel Time". Team 2 and Team 4 reach the workstations in a shorter time than Team 1 and Team 3.

In *Table 14* is shown the case with the Team with two different speed, it should be noted that by solving the optimization, the simulation assigns to teams 2 and 4, which are the fastest moving teams, more jobs. Fast-moving teams have 12 jobs to complete, while slower-moving teams have 9 jobs to complete.

Table 14: Team with different Travel Time

Team 1		Team 2		Team 3		Team 4	
Job number	Sequence	Job number	Sequence	Job number	Sequence	Job number	Sequence
7	1	4	1	3	1	40	1
8	2	22	2	13	2	17	2
11	3	28	3	37	3	34	3
33	4	27	4	39	4	14	4
20	5	15	5	6	5	18	5
41	6	30	6	9	6	2	6
25	7	16	7	5	7	10	7
32	8	26	8	29	8	12	8
35	9	31	9	42	9	21	9
		1	10			38	10
		23	11			19	11
		36	12			24	12

Table 15: Team with same Travel Time

Team 1		Team 2		Team 3		Team 4	
Job number	Sequence	Job number	Sequence	Job number	Sequence	Job number	Sequence
10	1	14	1	35	1	41	1
11	2	39	2	37	2	13	2
3	3	7	3	28	3	25	3
27	4	22	4	30	4	17	4
6	5	4	5	2	5	19	5
9	6	8	6	31	6	1	6
23	7	18	7	32	7	42	7
36	8	20	8	16	8	12	8
33	9	21	9	40	9	34	9
15	10	29	10	24	10	26	10
38	11					5	11

In *Table 15* is reported the case with the teams with same speed, by solving the optimization, the simulation assigns equally the jobs to every team.

However, the total number of jobs 42 is not divisible by 4, the simulation assigns 11 jobs to Team 1 and Team 4 and 10 jobs to Team 2 and Team 3.

Remember that the optimization for the assignment of jobs is also done on the basis of the processing time of the jobs and not only in relation to the travel time.

8.3.2. Scenario 2 - Teams

Scenario 2 shows the result that is found by changing the number of teams that perform maintenance activities: the assignment of teams and jobs is recalculated with the optimizer.

Three simulations were carried out with the same number of jobs, 42 jobs, and with different number of teams: 2, 4, 6. The following figures show:

- The grey color represents the time in which trains are in circulation and it is not possible to carry out maintenance activities.
- The yellow color represents the travel time between the deposit and the first job, and between one job and another.
- Each team is assigned a different color: Red for Team 1, Cyan for Team 2, Green for Team 3, Blue for Team 4, Orange for Team 5 and Purple for Team 6.
- The number on the rectangle represents the number of the executed job.

Figure 42 shows the result of the optimization of 42 jobs with 4 teams. Job maintenance activities are completed in 7 nights.

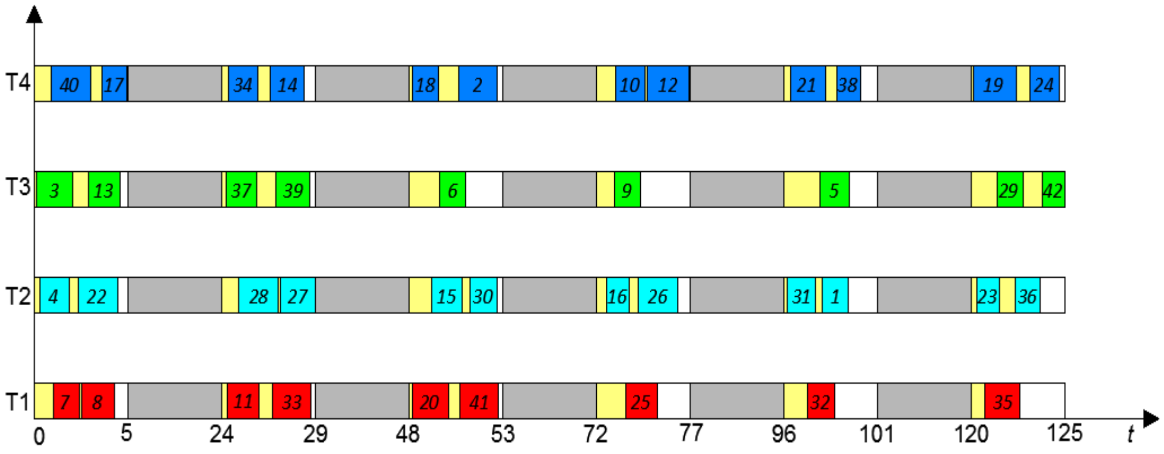


Figure 42: Optimization with 42 jobs and 4 teams

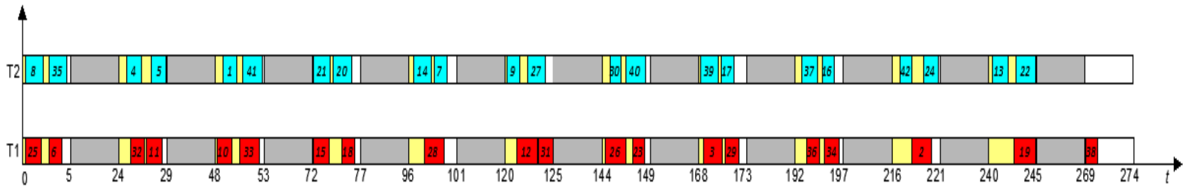


Figure 43: Optimization with 42 jobs and 2 teams

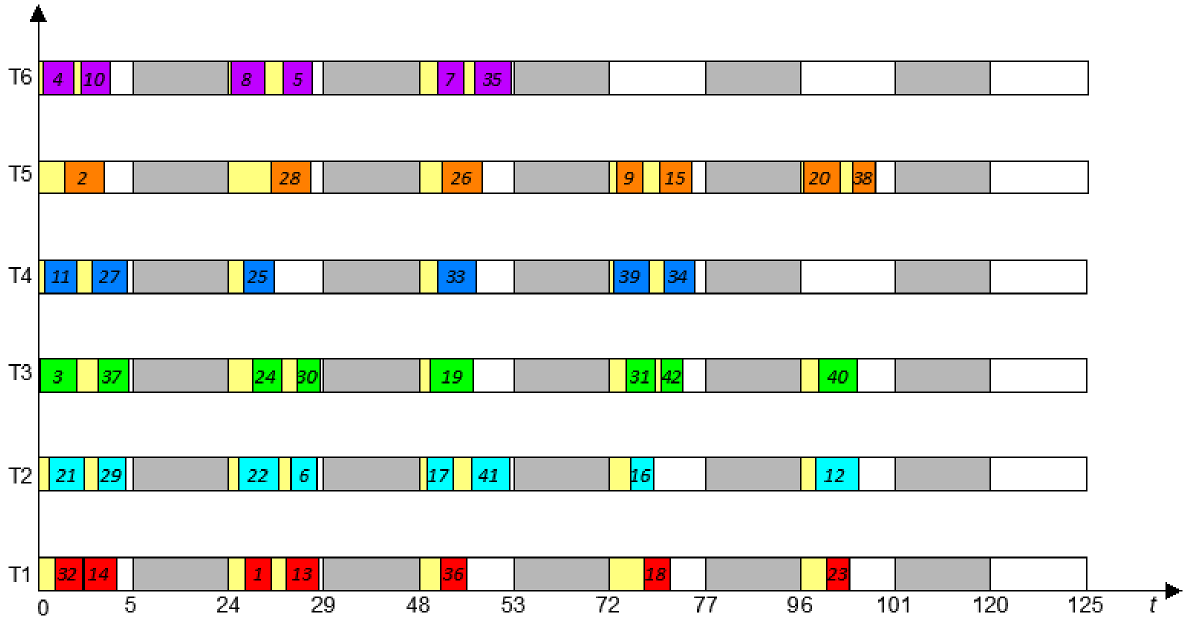


Figure 44: Optimization with 42 jobs and 6 teams

Figure 43 shows the result of the optimization of 42 jobs with 2 teams. Job maintenance activities last for several shifts and are completed in 11 and 12 nights. In the event that there are more maintenance teams available, the maintenance activity of the jobs ends in less time, therefore fewer nights are required, as shown in the *Figure 44*.

It is clear that if more resources are available to carry out maintenance activities, the time required to do them will certainly be less. The optimization is rerun whenever the number of available resources is known, to better assign the work to each team.

8.3.3. Scenario 3 - Jobs

In scenario 3, the number of teams is kept equal to 4, while the number of jobs on which maintenance is required varies. The simulation is made for 21, 32 and 42 jobs.

The reference scenario with 4 teams and 42 jobs is the one represented in *Figure 42*.

The simulation performed for 32 jobs and 4 teams is shown in *Figure 45*.

The graph shows that 5 working nights are required to complete the maintenance activity on the 32 jobs.

Finally, *Figure 46* shows the simulation performed for 21 jobs and 4 teams. Only 3 working nights are necessary to complete the maintenance activity on the 21 jobs.

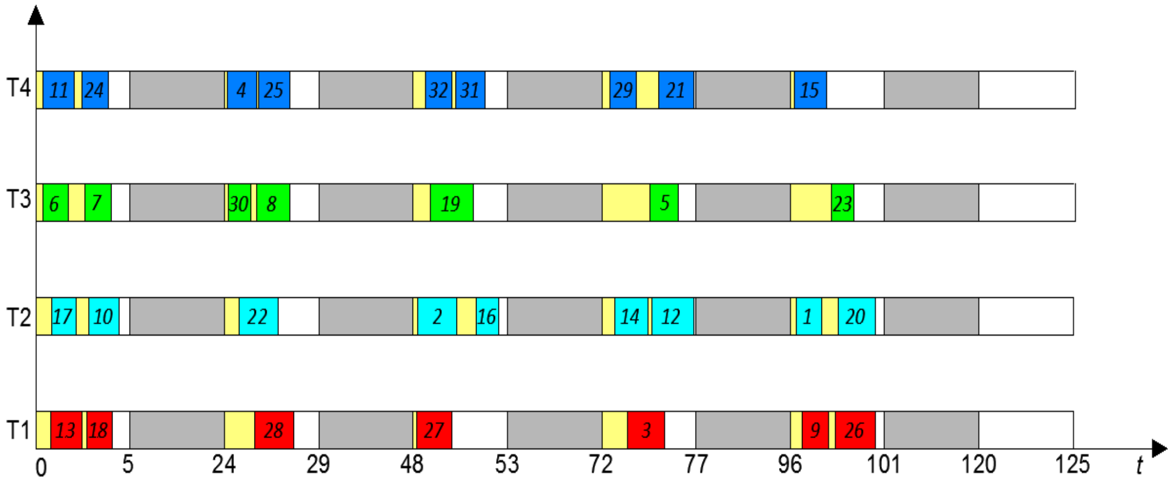


Figure 45: Optimization with 32 jobs and 4 teams

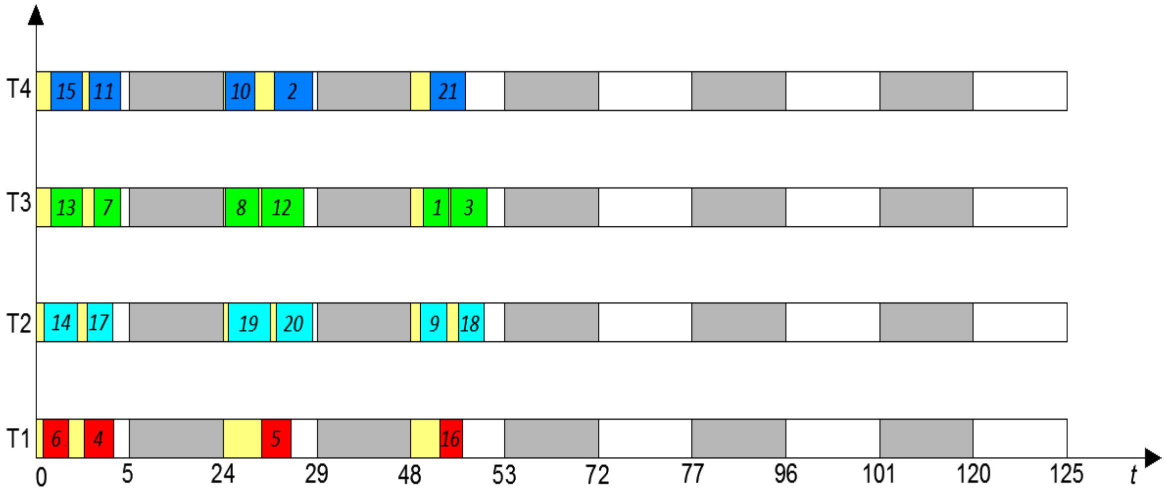


Figure 46: Optimization with 21 jobs and 4 teams

It should be noted that if the maintenance activities to be carried out increase or decrease, the time required to perform them will change and the assignment to the teams of each job will change. The optimization is rerun every time the number of maintenance activities necessary is determined, to assign the work to each team, optimizing as much as possible.

8.4 Performance and computing times

The used railway line maintenance scheduling problem, described in Chapter 5, has an NP-hard complexity and requires a long time to be solved, even for small models.

The matheuristic algorithm that determines the solution of the maintenance optimization problem is implemented with the Matlab and IBM ILOG Cplex programs. The algorithm uses the branch and bound approach, and does not find the optimal solution for a number of jobs greater than 20, but only a feasible solution. The time it takes the algorithm to return a feasible solution is about 2 hours.

Figure 47 shows the solution found using the matheuristic algorithm, also described in [83]. The algorithm performs the solution using a greater number of constraints than the solution found with simulation on ExtendSim.

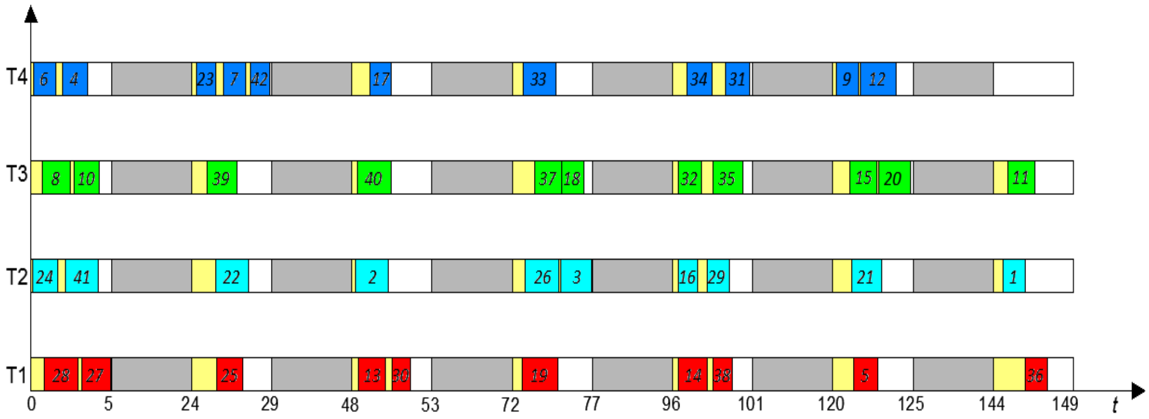


Figure 47: Matheuristic algorithm solution

The ExtendSim optimizer that determines the solution of the maintenance scheduling problem uses the simulation approach, verifying the convergence after a number of attempts set by the user, it still finds a feasible solution to the problem that is not necessarily the optimal solution. The time it takes the simulation to return a solution depends on the percentage of convergence that is entered and how many cases are required to simulate.

For the model created, 1000 simulation cases and a convergence of 100% are required, the simulation lasts between 20/25 minutes.

Figure 48 shows the solution to the job sequencing problem found through ExtendSim, using the assignment to the teams given by the matheuristic algorithm. The complete optimization solution made by ExtendSim is plotted in *Figure 49*.

Comparing the three graphs it is possible to see how the optimization of the job sequencing carried out with ExtendSim finds a better solution for the completion time component, the solution with ExtendSim is also better for the component of the objective function of the deadlines.

However, it must be remembered that in the matheuristic model there are some more restrictive constraints regarding the path of the teams, which have not been included in the ExtendSim program.

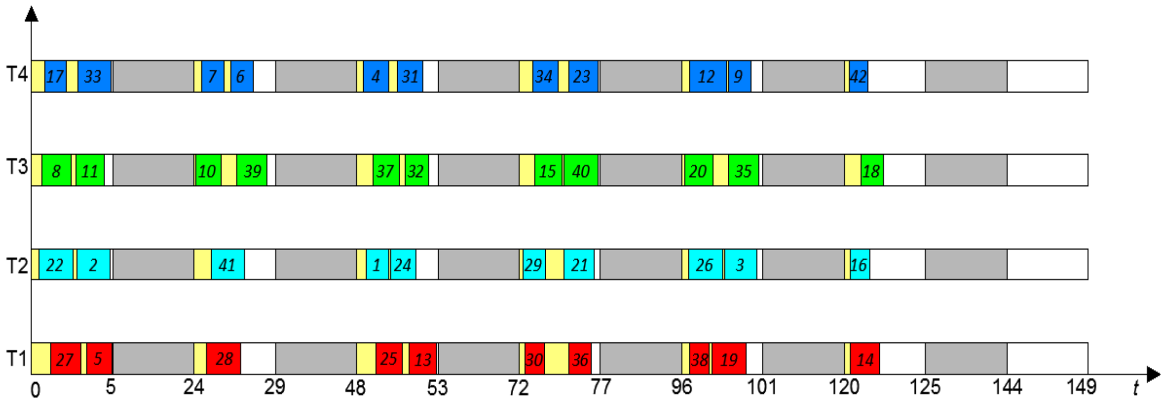


Figure 48: Sequencing optimization

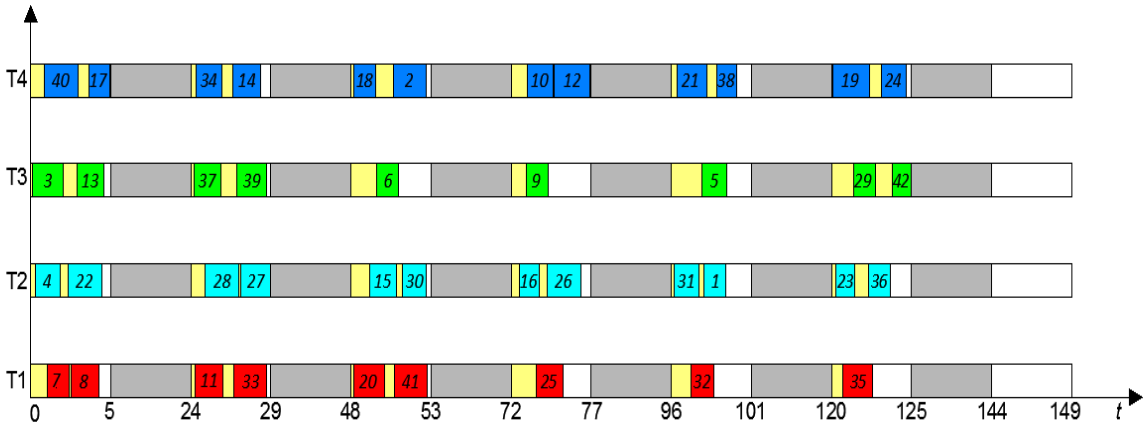


Figure 49: Sequencing and assignment optimization

Figure 50 represents the trend of the value of the objective function in the simulation with the assignment to the teams and the sequencing of the jobs.

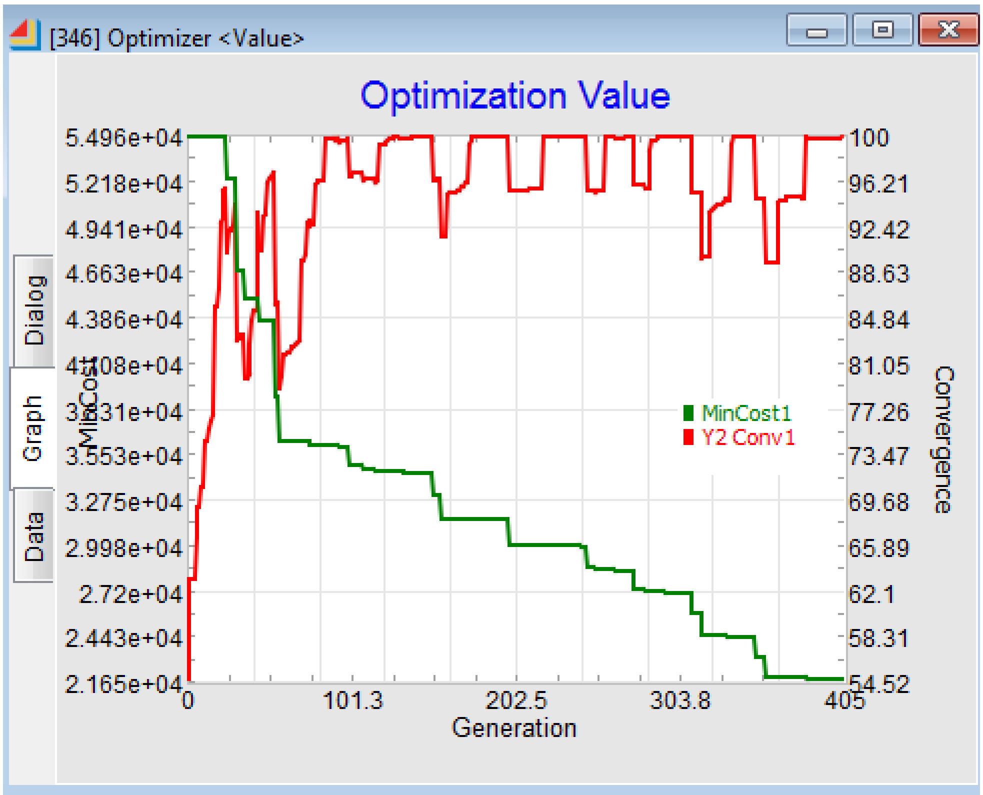


Figure 50: Trend of the objective function and convergence

Finally, the differences between the two optimization models, matheuristic algorithm and simulator, used are summarized:

- The matheuristic algorithm requires a longer running time for greater computational complexity
- The simulation seems to find a better solution, but relaxing some constraints
- However, the simulation has a maximum number of variables equal to 100, as the model is created, the maximum number of jobs that can be scheduled is 48 (with any number of teams).
- With the matheuristic algorithm it is easier to insert a constraint through a formula, while in the simulation the insertion of some constraints is complicated.

The *Table 16* shows the completion times of the 3 models created and of the matheuristic algorithm, to compare the duration of the simulation of each of them. The duration of the simulation depends on the number of assets and teams to realize the planning of the maintenance activity.

Table 16: Computing times of the models

Model	Computing times
Deterministic model	< 2 min
Sequence of execution of jobs	≃ 15 min
Assignment to the teams and sequence of execution of jobs	≃ 20 min
Matheuristic algorithm	≃ 2 hrs

Conclusions

In conclusion, the applied simulation model allows to obtain, in a short time, a predictive railway maintenance scheduling.

The solution is obtained minimizing a cost function that consists in a weighted sum of the completion time and the tardiness with respect to the deadlines of the maintenance activities to be carried out. A sensitivity analysis is performed by varying the cost function weights according to the infrastructure manager targets.

The model makes it possible to scheduling maintenance activities when the asset reaches a certain state of degradation, before its failure, thus avoiding costly railway interruptions. The approach therefore allows to:

- Reduce railway maintenance costs by also improving rail infrastructure availability and reliability.
- Make the best use of the time windows available to carry out the maintenance activity.
- Optimize the use of available resources (teams and machines), balancing their workload, reducing their movements along the line and the time required to carry out maintenance.
- Increase the efficiency of the maintenance process by avoiding unnecessary and costly movements and interventions.
- Decrease the possession time of the infrastructure.

Finally, a real-world case study of railway maintenance is reported.

In the case study, some variables are considered at first as deterministic and then as stochastic to compare the results.

In real world scenarios, the uncertainty is present in almost every input data of the problem. An example is the duration of rail traffic interruption, which is set 300 minutes (5 hours) in the simulation models, the time interval can vary in case of trains delays. This aspect does not depend on the maintenance process.

Other examples can instead be attributed to internal problems of the maintenance process, such as machines or equipment breakdowns, lack of material or human resources, etc. These uncertainties make long-term maintenance activities not easily predictable and programmable.

Therefore, infrastructure managers have to deal quite often with deviations from the predefined maintenance plan.

Moreover, even if a predictive maintenance strategy is applied, unexpected failures may happen due to the uncertainty of asset status estimation, performed by predictive models. The deadline of maintenance interventions should therefore be considered as a stochastic variable.

This present study moves some steps towards the uncertainty management of maintenance process in the rail sector.

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] UIC International Union of Railways, Asset management working group, (2016). Practical implementation of Asset Management through ISO 55001.

[3] IN2RAIL Deliverable (2017), D6.5 – Condition- & Risk-based Maintenance Planning (CRIMP).

[4] SHIFT2RAIL Joint Undertaking, (2015). SHIFT2RAIL Multi Annual Action Plan.

[5] K. Kobbacy, D. Murthy (2007). Complex System Maintenance Handbook: Blending Theory with Practice, chap. 1, Springer Verlag, Berlin, pp. 3-18.

[6] 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.

[7] 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.

[8] G. Budai-Balke, (2009). Operations Research Models for Scheduling Railway Infrastructure Maintenance. PhD Thesis. Erasmus University Rotterdam.

[9] S. Sakai, (2010). Risk-based maintenance, JR East Technical Review 17.

[10] Improverail (2002). Final Report.

[11] L.E. Daniels, (2008). Track maintenance costs on rail transit properties. TCRP Document 43. TRB.

[12] C. Esveld, (2001). Modern Railway Track, MRT-Productions, Zaltbommel, The Netherlands.

[13] Office of the Rail Regulator (2001). The Possessions Review: Provisional Conclusions on the Schedule Incentive Structure.

[14] D. Huisman, L. Kroon, R. Lentink, and M. Vromans (2005), Operations Research in Passenger Railway Transportation, Statistica Neerlandica, vol. 59, pp. 467-497.

[15] 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.

[16] 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).

[17] EC Official Journal (2016). Directive (EU) 2016/798 on railway safety (Recast of Directive 2004/49/EC).

[18] 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.

[19] 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.

[20] IN2RAIL Deliverable D6.6 – Dynamic and Real-time Maintenance Planning Concept, 2017.

[21] M. Cross, (1988). Raising the Value of Maintenance in the Corporate Environment, Management Research News, vol.11, pp. 8-11.

[22] 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/.

[23] 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.

[24] 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.

[25] S. Alodhaibi, R. L. Burdett, P. KDV. Yarlagadda, Framework for Airport Outbound Passenger Flow Modelling, Procedia Engineering, Volume 174, 2017, Pages 1100-1109, ISSN 1877-7058.

[26] P. Busato, R. Berruto, Minimising manpower in rice harvesting and transportation operations, Biosystems Engineering, Volume 151, 2016, Pages 435-445, ISSN 1537-5110.

[27] C. M. Bourque, K. T. Clarno, B. D. Leibowicz, High-level fuel fabrication facility designs from discrete-event simulation, Annals of Nuclear Energy, Volume 168, 2022, 108893, ISSN 0306-4549.

[28] P. Liu, B. Xia, J. Li, D. Zhao, H. Wei and M. Li, "ANN-based Command and Control Seat Allocation Optimization with Simulation Data," 2020 IEEE International Systems Conference (SysCon), 2020, pp. 1-7, doi: 10.1109/SysCon47679.2020.9275844.

[29] L. Jiang, G. Sun and N. Zhang, "Layout research of campus traffic system based on system simulation," *International Technology and Innovation Conference 2009 (ITIC 2009)*, 2009, pp. 1-5, doi: 10.1049/cp.2009.1435.

[30] S. Hu and W. Yi, "Simulation and Optimization of Kanban System for Steel Rolling," 2009 International Workshop on Intelligent Systems and Applications, 2009, pp. 1-4, doi: 10.1109/IWISA.2009.5073111.

[31] S. Hu and X. Linwei, "Simulation and Optimization for Noshery Service System," *2009 International Symposium on Information Engineering and Electronic Commerce*, 2009, pp. 721-723, doi: 10.1109/IEEC.2009.157.

[32] W. Hou and D. Zeng, "ExtendSim Computer Simulation Technology Applied in the Inventory System of Agricultural Products," *2013 International Conference on Computational and Information Sciences*, 2013, pp. 1684-1686, doi: 10.1109/ICCIS.2013.439.

[33] X. Li and P. Li, "Simulation Optimization Under Random Conditions TG Business Model of Spare Parts Inventory," *2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 2019, pp. 1025-10253, doi: 10.1109/ICMCCE48743.2019.00229.

[34] S. Pirkwieser, (2012), Hybrid Metaheuristics and Matheuristics for Problems in Bioinformatics and Transportation, Bakk.techn.

[35] 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.

[36] 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.

[37] 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.

[38] B. Rockstroh, W. Kappes, F. Walte, (2008). Ultrasonic and eddy current inspection of rail wheels and wheel set axles, WCNDT 2008.

[39] 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.

[40] 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.

[41] 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.

[42] S. Štarman, V. Matz, (2008) Automatic System for Crack Detection Using Infrared Thermographic Testing, WCNDT 2008.

[43] A. Elballouti, S. Belattar, N. Laaidi, (2010). Thermal Non-Destructive Characterization of the Spherical Defects in the Roadways, ECNDT 2010.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] M. El-Sibaie, Y. Zhang, (2004). Objective track quality indices. Transp Res Record J Transp Res Board vol. 1863, pp. 81–87.

[50] M. Shenton, (1985). Ballast deformation and track deterioration. Track Technol pp. 253–265.

[51] Y. Sato, (1997). Optimum track structure considering deterioration in ballasted track. In: Proceedings of 6th international heavy haul conference, Cape Town, SouthAfrica, p 576.

[52] Y. Sato, (1995).Japanese studies on deterioration of ballasted track. Veh Syst Dyn vol. 24, n.1, pp.197–208.

[53] 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.

[54] 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.

[55] Y. Zhang, M.H. Murray, L. Ferreira (2000) Modelling rail track performance: an integrated approach. Transp J pp.187–194.

[56] 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.

[57] 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.

[58] R. Guo, B.M. Han, (2013). Multi-stage linear prediction model for railway track irregularity. Appl Mech Mater vol.361, pp.1811–1816.

[59] 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.

[60] 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.

[61] N. Lyngby, (2009). Railway track degradation: Shape and influencing factors. Int J Perform Eng vol.5, n.2, pp.177.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] J. Chaolong, X. Weixiang, W. Futian, W. Hanning, (2012). Track irregularity time series analysis and trend forecasting. Discret Dyn Nat Soc 2012.

[77] 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.

[78] A. Consilvio, A. D. Febbraro and N. Sacco, "A Rolling-Horizon Approach for Predictive Maintenance Planning to Reduce the Risk of Rail Service Disruptions," in IEEE Transactions on Reliability, vol. 70, no. 3, pp. 875-886, Sept. 2021, doi: 10.1109/TR.2020.3007504.

[79] S. M. Famurewa, T. Xin, M. Rantatalo, and U. Kumar, “Optimisation of maintenance track possession time: A tamping case study,” Proc. Institution Mech. Eng., F, J. Rail Rapid Transit, vol. 229, no. 1, pp. 12–22, 2013.

[80] S. M. Famurewa, U. Juntti, A. Nissen, and U. Kumar, “Augmented utilization of possession time: Analysis for track geometry maintenance,”Proc. Institution Mech. Eng., F, J. Rail Rapid Transit, vol. 230, no. 4, pp. 1118–1130, 2015.

[81] 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.

[82] 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.

[83] 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

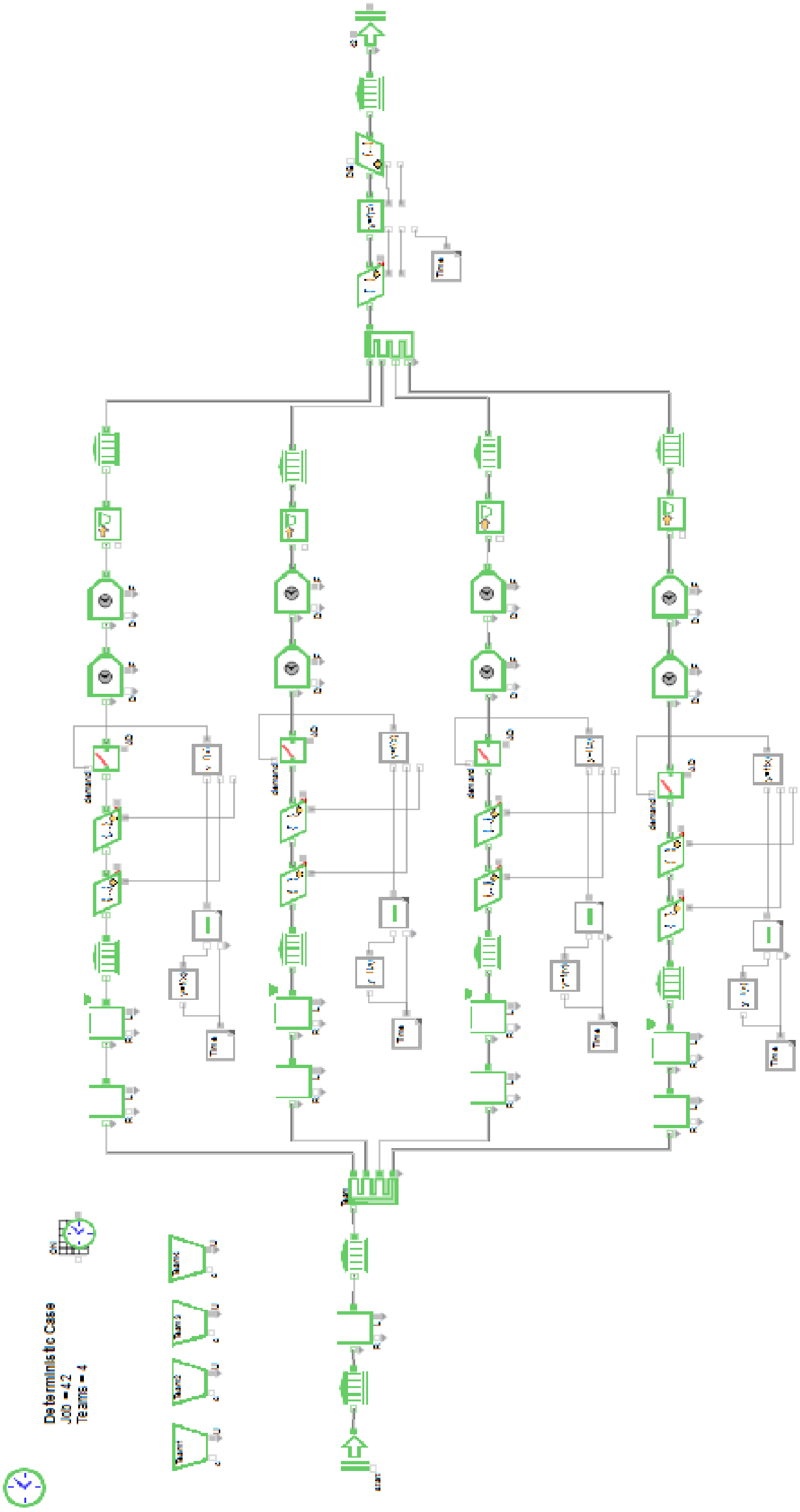


Figure 51: 42 jobs and 4 teams – Deterministic Case

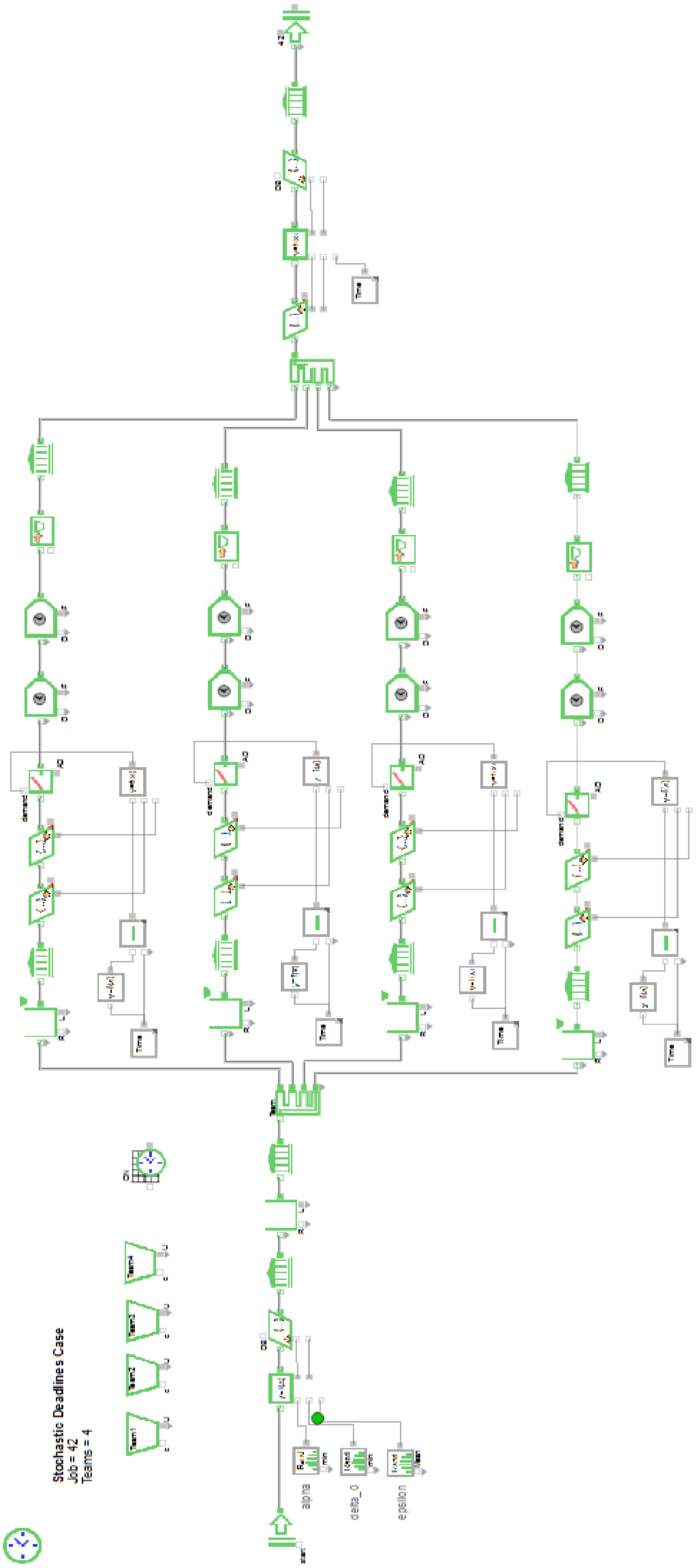


Figure 52: 42 jobs and 4 teams – Stochastic Deadline Case

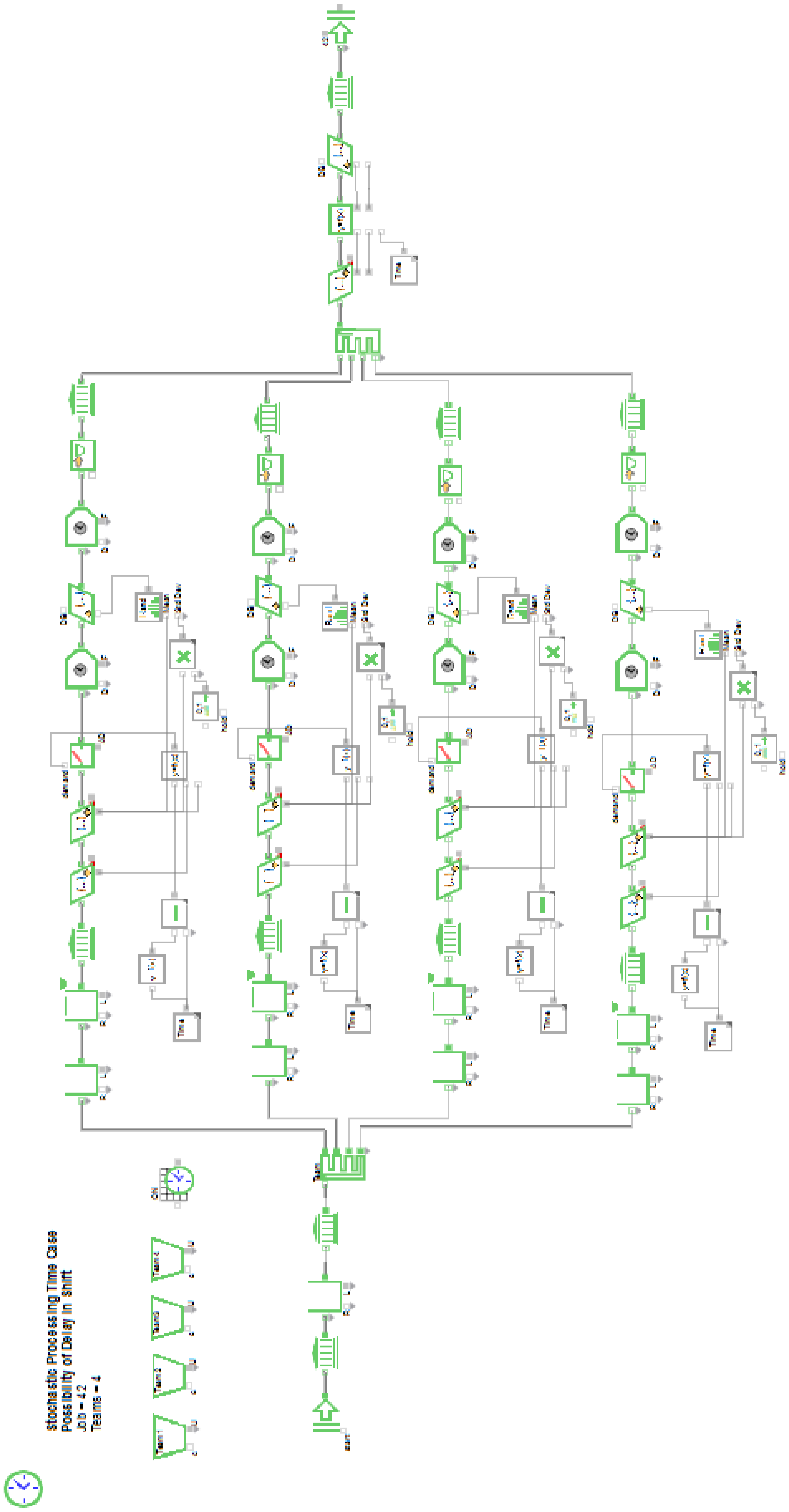


Figure 53: 42 jobs and 4 teams – Stochastic Processing Time Case

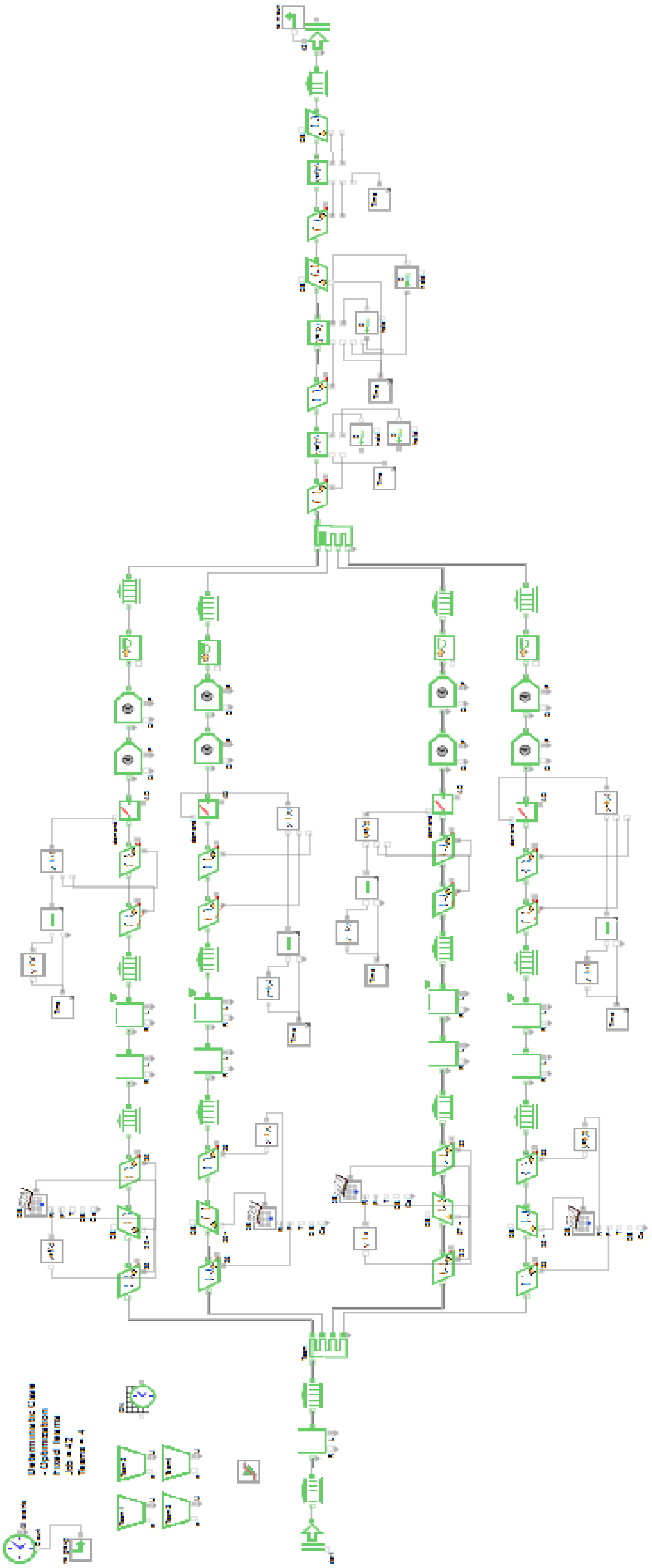


Figure 54: 42 jobs and 4 teams – Deterministic Optimization Case

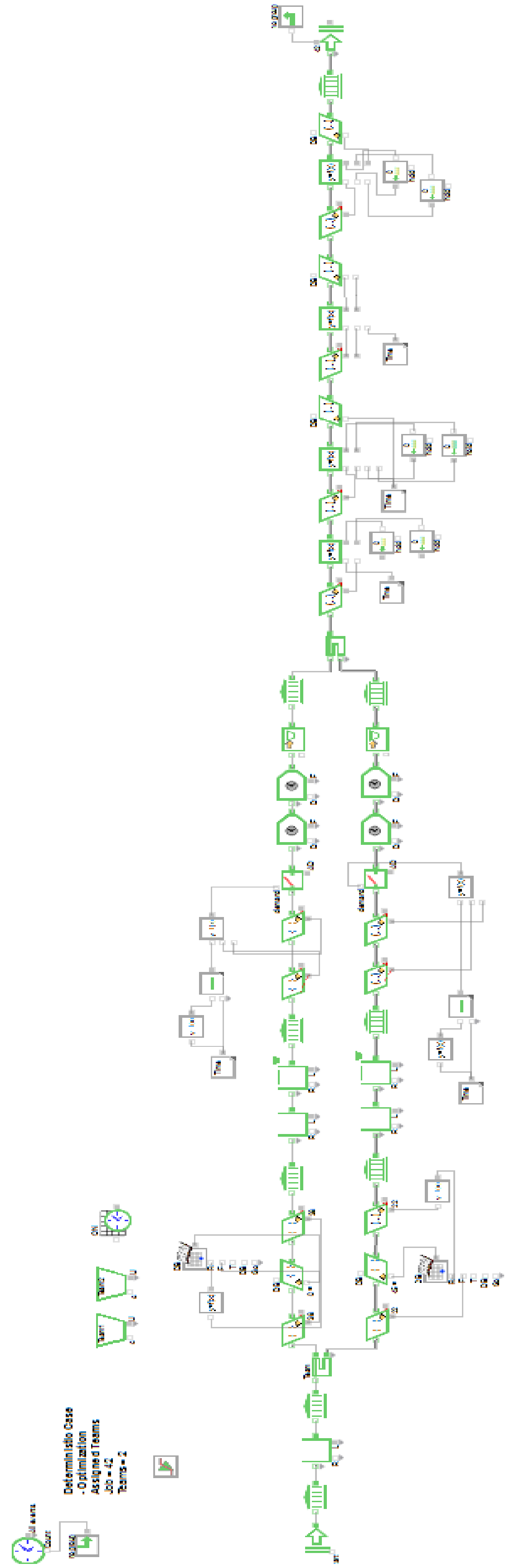


Figure 55: 42 jobs and 2 teams – Deterministic Optimization Case

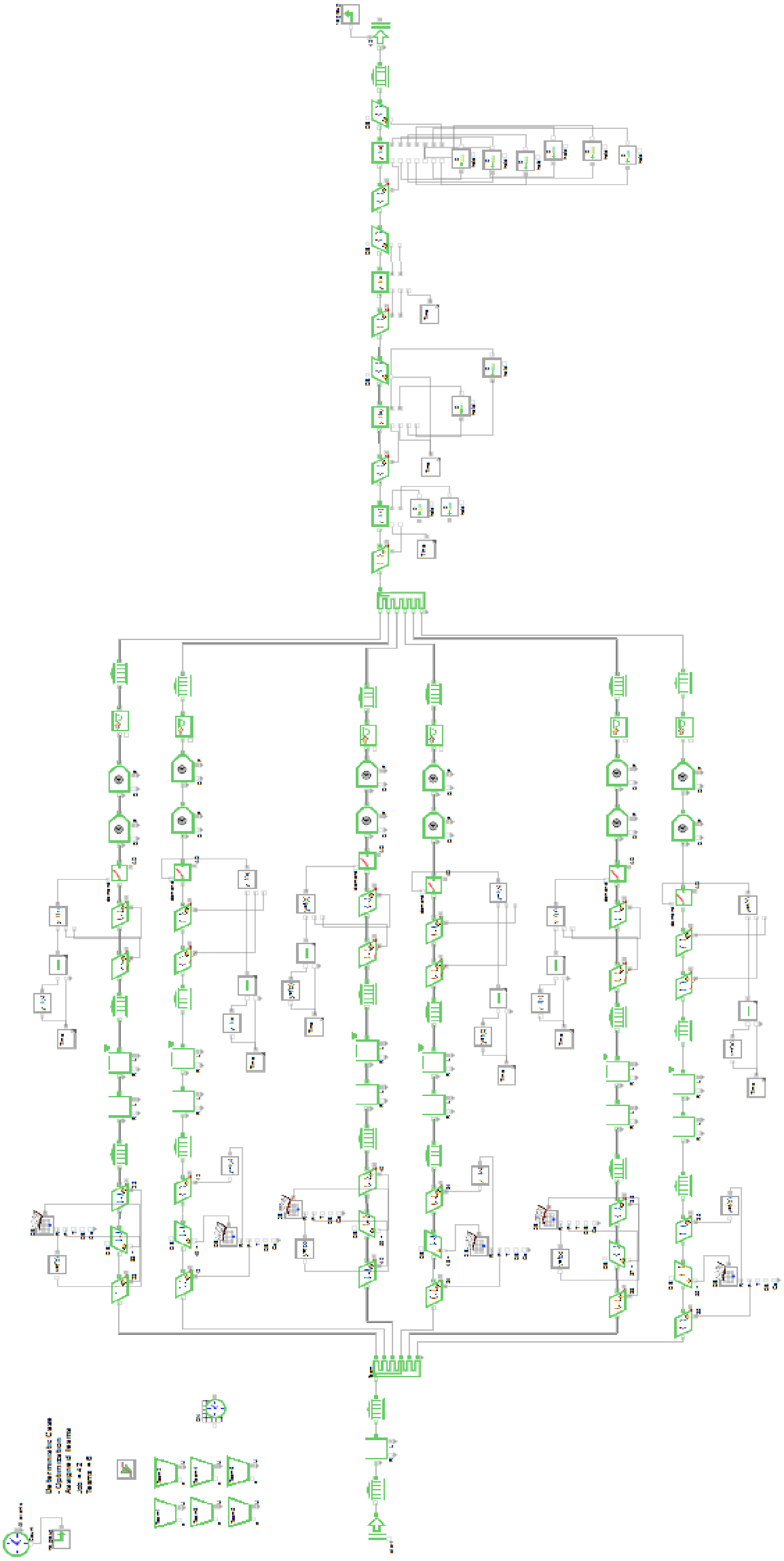


Figure 56: 42 jobs and 6 teams – Deterministic Optimization Case