H2020-MG-8.1a-2014

INFRALERT: Linear Infrastructure Efficiency Improvement by Automated Learning and Optimised Predictive Maintenance Techniques

Grant Agreement number: 636496

Deliverable D6.1

Smart decision support framework

Work Package	WP6 Smart operation and maintenance decision support		
Task	T6.1 Definition of a concept for condition- and risk-based interventions planning T6.2 Design of a framework for smart operation and maintenance decision support		
Revision	6	**Due date**	30/04/2016
Revision date	29/04/2016	**Submission date**	29/04/2016
Deliverable type	Report		

Deliverable Leader	FHG
Contributing partners	CEM, USE, LTU, REG

Dissemination Level		
PU	Public	**X**
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC.	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 636496

Document status

Revision	Date	Description
1	10/03/2016	Version v00
2	11/03/2016	Version v01
3	08/04/2016	Input from USE, REG, LTU integrated
4	13/04/2016	Version v03 for internal review, including input from CEM
5	22/04/2016	Final draft
6	29/04/2016	Final version
Status	Final	

Executive Summary

Results from this Deliverable are related to Tasks 6.1 "Definition of a concept for condition- and risk-based interventions planning" and 6.2 "Design of a framework for smart operation and maintenance decision support".

The condition- and risk-based planning concept forms the "backbone" of the decision support tools. It defines guiding principles and features as well as provides a common structure to be followed when deploying and applying decision support systems to the process of maintenance planning. The concept has been defined in alignment with the specifications made for the Alert management (WP4) and the RAMS&LCC analysis (WP5). To this end, the following subtasks have been carried out:

At first, the background of INFRALERT's concept for maintenance planning has been reviewed, which are the maintenance policies and planning principles pre-existing in theory and practice. When defining the novel planning concept, references to these groundwork are made and basic assumptions and ideas are used and refined. It is common practice to distinguish several planning levels that decompose the overall maintenance planning process into single steps with dedicated tasks and decisions to be made: strategic, tactical and operational (or dynamic). We provide a brief definition of these planning levels and show their scope and the boundaries between them.

The concept for maintenance planning and the decision support framework as developed in this Deliverable are based on several so-called building blocks, each of them providing a dedicated functionality, processing specified input and delivering specified output. The framework basically defines the linking between these blocks to conduct the maintenance planning process. Besides the building blocks that are at the core of the planning concept - dealing with strategic, tactical or operational planning tasks - there are also important ambient blocks feeding the planning tasks with information, doing certain computations in the form of pre- or post-processing within the overall complex process. The most important amongst them - RAMS&LCC analysis, risk assessment, nowcasting and forecasting, alert management - have been summarised and the proper use of the information provided by these ambient building blocks are described.

A mathematical methodology to use probabilistic data for decision-making in optimisation problems is at the core of the concept for condition- and risk-based planning. This methodology allows to deal with uncertainty in maintenance and interventions planning, which is a necessity that arises in practical applications. The document in detail describes the integration of probabilistic information about maintenance alerts, asset conditions, defects and their transition, maintenance interventions and respective costs and resources as well as RAMS parameters and how they will be modelled as stochastic variables for optimisation under uncertainty.

To deal with such a setting, a solution approach has been developed that makes it possible to have "a look into the future" and to estimate or determine the consequences of decisions to be expected in the future. In this way, decisions to be taken now and having an effect in the future can be balanced with regard to robustness of planning.

The developed planning concept follows a nested and adaptive approach: dynamic, tactical and strategic planning levels will get linked together, providing feedback to each other in terms of input and results, planning at each level is done in a rolling time horizon.

Finally, the smart decision support tools to be developed will be a part of the automated data processing chain of the INFRALERT eIMS. Since maintenance and intervention planning is the end point of this chain there is high demand for interactivity with the user of the system. To assure a high

acceptance and usability of the planning tools, a framework has been designed to integrate smart decision support with existing procedures and building blocks. The resulting framework is general enough to be easily adapted and applied to a wide range of maintenance planning scenarios. It provides the basis for the development of specific optimisation models following the condition- and risk-based planning concept.

Table of content

List of tables

List of figures

Abbreviations and acronyms

Abbreviation / Acronym	Description
eIMS	Expert-based Infrastructure Management System
WP	Work Package
RAMS	Reliability, Availability, Maintainability and Safety
LCC	Life Cycle Cost
ESAL	Equivalent Single Axle Load
PSI	Present Serviceability Index
RL	Remaining Life
PDF	Probability Distribution Function
ANN	Artificial Neural Network
MDP	Markov Decision Process
UML	Unified Modeling Language

1 Background

The INFRALERT project aims to develop an expert-based information system to support and automate linear asset infrastructure management from measurement to maintenance. This enfolds the collection, storage and analysis of inspection data, the deduction of interventions to keep the performance of the network in optimal condition, and the optimal planning of maintenance interventions. It will also assess new construction strategic decisions.

For this purpose, in WP6 a concept for condition- and risk-based maintenance and interventions planning is developed, which will be used within a framework for the smart decision support. Results from this Deliverable are related to Tasks 6.1 "Definition of a concept for condition- and risk-based interventions planning" and 6.2 "Design of a framework for smart operation and maintenance decision support".

The ***condition- and risk-based planning concept*** forms the "backbone" of the decision support tools. It defines guiding principles and features as well as provides a common structure to be followed when deploying and applying decision support systems to the process of maintenance planning. The concept will be defined in alignment with the specifications made for the Alert management (WP4) and the RAMS&LCC analysis (WP5). The output from Alert management and RAMS&LCC analysis has to be translated into input for the optimisation models in interventions and maintenance planning. To this end, the following subtasks are carried out:

- Definition of the probabilistic information about asset conditions and parameters, characteristics and effects of maintenance activities as stochastic input parameters for maintenance planning models.
- Analysis and specification of a risk assesment methodology to be applied within the maintenance planning models based on parameter derived from RAMS analysis and simulation.
- Definition of objective functions to handle the trade-off between risks for postponing and costs for preponing interventions.
- Discussion of different uncertainty measures to be applied within the approach in order to evaluate the consequences of unexpected events during infrastructure operation and maintenance execution.

The smart decision support tools to be developed will be a part of the automated data processing chain of the INFRALERT eIMS. Since maintenance and intervention planning is the end point of this chain there is high demand for interactivity with the user of the system. To assure a high acceptance and usability of the planning tools, a ***framework*** will be designed to integrate ***smart decision support*** with existing procedures and building blocks. The resulting framework will be general enough to be easily adapted and applied to a wide range of maintenance planning scenarios. It provides the basis for the development of specific optimisation models following the condition- and risk-based planning concept.

2 Objectives

The overall aim of this deliverable is to specify the concept for condition- and risk-based maintenance planning together with the framework for smart decision support.

The single objectives that lead to this overall result presented in this document are:

- To review existing maintenance policies and planning principles, which are used in he background for INFRALERT's planning concept, and to introduce the notion of planning levels that are consistently approached by the decision support framework (see Section 3).
- To summarise the functionalities, inputs and outputs of important ambient building blocks of the decision support framework: The framework itself defines the linking between so-called building blocks to conduct the maintenance planning process. Ambient blocks, responsible for feeding the planning tasks with information, doing certain pre- or post-processing, are specified in detail in dedicated Deliverables. Here, the relation to decision support framework is given for the most important ones: RAMS and LCC analysis, risk assessment, asset condition and alert generation (Section 4).
- To describe the theoretical background of INFRALERT's planning concept, which is the handling of uncertainties through the use of probabilistic information: Mathematical concepts related to optimisation under uncertainty that will be applied in maintenance planning are introduced (see Section 0).
- To define the (mathematical) features of the optimisation models, subsequently used to formalise the decision-making (see Section 0).
- To describe the algorithmic solution approach and procedures for robust planning, based on the mathematical concepts for optimisation under uncertainty, including a consideration of uncertainty management (see Section 5.3 and 5.4).
- To finally conclude with the framework, being a modular linking of the building blocks and concepts discussed before: The framework is described on two degrees of granularity, to cope with the overall nested planning process as well as to detail on single planning levels (see Section 6).

3 Background on Maintenance Management

This section reviews the background of INFRALERT's concept for maintenance planning, which are the maintenance policies and planning principles pre-existing in theory and practice. When defining the novel planning concept, references to these groundwork will be made and basic assumptions and ideas will be used and refined.

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

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

It is worth noting that wording here is not consistent, neither in practice nor in literature: Maintenance policies often are also referred to as maintenance strategies, concepts, procedures or simply methods, whereas planning principles is a term introduced in this context here.

It is common practice to distinguish several ***planning levels*** that decompose the overall maintenance planning process into single steps with dedicated tasks and decisions to be made: strategic, tactical and operational (or dynamic). At the end of this section we provide a brief definition of these planning levels and show their scope and the boundaries between them. In later sections we will refer to the levels defined in order to separately address the differing features and principles applied. Finally, Section 6 provides a detailed description of each of the planning levels in the context of the INFRALERT framework for smart operation and maintenance decision support.

3.1 Maintenance Policies

Reactive Maintenance (RM)*:* The easiest procedure is reactive maintenance: assets are maintained when they are broken. Basically, there are two kinds of reactive maintenance. On the one hand, the concept of „run to failure“, where assets will be replaced after breaking down. For example components of the signal system are renewed after failure. On the other hand, the routine maintenance activities: Small failures will be repaired when they appear, e.g. potholes are patched every spring to restore pavement smoothness. To apply reactive maintenance, only less information about the assets are necessary, but the failure will occur unexpected.

Condition-based Maintenance (CBM): In condition-based maintenance, failures or break-down will be avoided by maintaining the assets when they show signs of decreasing performance or upcoming failure. Therefore, the assets have to be monitored closely to see condition changes in time, as shown in Figure 1. As soon as monitoring shows that the condition is below the trigger values (dashed line), maintenance is requested and executed. With it, the condition is improved. It is also necessary to define trigger values for the measures that initialise need for maintenance. Thereby, the time between two condition measurements, as well as the time between maintenance request and execution, has to be considered to ensure punctual maintenance. Also the best maintenance activity should be defined in advance. Decision support in resource allocation can help to reduce the time between maintenance request and execution. Then, trigger values can be higher and the maintenance effort can be reduced.

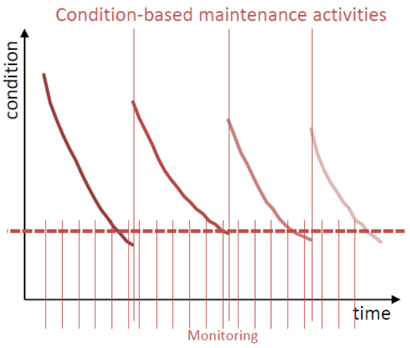

Figure 1. Schematic CBM

Preventive maintenance (PM) is characterised by regular maintenance activities to improve the assets. Often this is done in predefined time intervals, but it is also possible to do preventive maintenance depended on usage triggers, for example after a certain number of trains pass a track. Aim of preventive maintenance is to extend life-time, to increase asset performance and to avoid unexpected break-down. In Figure 2, the progress of preventive maintenance is shown: in a predefined interval, maintenance activities are executed and the track condition is improved in order to do not fall down a certain condition limit (dashed line). To apply preventive maintenance, time intervals or usage trigger values have to be defined, which can be done based on expertise, historical data or scientific results. There are a lot of papers dealing with degradation models for railway tracks and road pavement and the calculation of perfect preventive strategies. The big advantage of the procedure is the high planning ability. The upcoming maintenance activities are known in advance and can be scheduled with renewal and construction in mind.

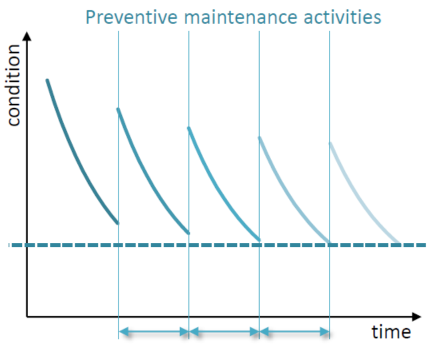

Figure 2. Schematic PM

Predictive maintenance (PdM) is a relatively new approach because it requires closely monitoring and a fine grasp for the deterioration process. The main idea is to predict asset condition in order to plan in advance condition-based maintenance. In Figure 3, the progress of predictive maintenance is shown. The blue line represents the expected track condition, the red line the real condition. If the prediction reaches the condition limit (dashed line), maintenance will be executed to improve the condition. Thereby, the real condition can be worse or better than predicted, but with a model close to the real deterioration the perfect time for maintenance can be approximated. Predictive maintenance has the advantages from condition based and preventive maintenance. To apply predictive maintenance, a deep understanding of the underlying deterioration process is essential. To underpin the results of the degradation models, historical data is necessary. If a good and reliable degradation

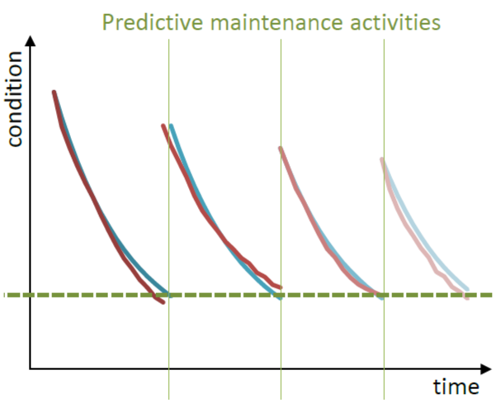

Figure 3. Schematic PdM

model exists, asset condition can be predicted with a small variation and maintenance can be planned in advance. The resulting maintenance plans should be robust against uncertainties like unexpected deterioration or unforeseen events which requires extraordinary maintenance.

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

3.1.1 Maintenance triggers

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

- Condition trigger: Condition-based and reactive maintenance is based on condition triggers. This can be the achievement of a certain measure value, a dropping below a needed quality level or the occurrence of a failure.

  Predictive maintenance also will be based on condition triggers. In difference to condition-based maintenance, planning starts when the trigger is predicted and not just when it is reached.

- Break-down trigger: A special kind of condition trigger, because the trigger value is “break-down”. Only usable for uncritical assets in terms of safety and reliability.

- Time trigger: Preventive maintenance is repeated in predefined intervals. This intervals can be defined by time (e.g. once a year).

- Usage Trigger: Preventive maintenance can also be triggered by usage (e.g. every 100,000 switching operations). Triggering by usage has the advantage that for assets with changing workload the definition of a time trigger can be hard. But a usage trigger requires a usage measurement, which leads to a higher implementation effort. Furthermore, usage trigger can shorten the planning period of preventive maintenance. Planning will start when the usage limit is reached or the time, when it is reached, can be expected.

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

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

3.1.2 Evaluation

In this subsection, the behaviour of the four presented maintenance policies will be analysed. Then, the policies will be compared to each other and some advantages and disadvantages will be shown.

In Figure 4, the maintenance policies are evaluated with respect to the planning period. As suggested in the description of the policies, reactive maintenance has to be executed promptly or short-term. In condition-based maintenance policies, the planning period depends on the selected trigger and the inspection interval. In the most cases, condition-based maintenance has to be planned in short-term, but if the trigger has a large buffer, the planning period can be longer. Preventive and predictive maintenance have longer planning periods. In preventive maintenance, the planning period depends on the length of the time trigger. If a usage trigger is used, the planning period can be shorter because it has to be estimated when the usage trigger will be reached. In predictive maintenance, the planning periods depends on the credibility of the prediction models. It is assumed that predictive maintenance is implemented if and only if the prediction is good enough. Then, the planning period is medium- to long-term.

Figure 4. Evaluation with respect to the Planning Period

In Figure 5, the maintenance policies are evaluated with respect to risks. Reactive maintenance is classified as risky, because in the most applications it is too risky to wait with maintenance until failures occurs. That doesn't mean, that reactive maintenance is never suitable. There are some applications, where reactive maintenance is a good option, e.g. the exchange of traffic lights, the replacement of reflector posts or fastenings. Condition-based maintenance is safe, if the inspection interval is not too long and the condition trigger is not too low for the degradation rate. To evaluate the inspection interval and the condition trigger value, the degradation rate has to be approximated. If deterioration is underestimated or the chosen parameters are not suitable, maintenance can be requested too late and the risk is higher. Preventive and predictive maintenance are safe, because of the longer planning period. With it, possibly misjudgments in the parameter evaluation can be seen in time and the parameters can be adjusted.

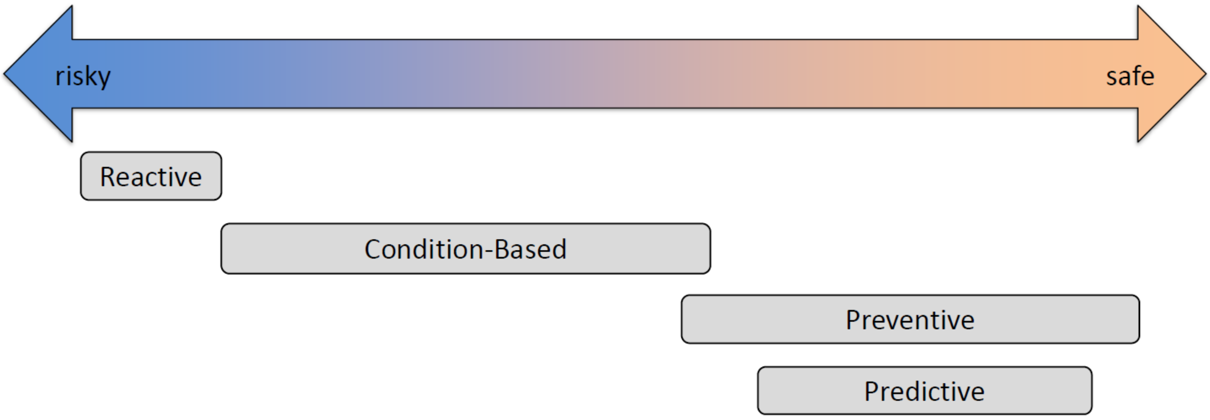

Figure 5. Evaluation with respect to Risk

In Figure 6, the maintenance policies are evaluated with respect to the cost effectiveness. Reactive maintenance is in the most cases expensive, because failures are not avoided by improving the asset through additional maintenance. Then, deterioration is fast and the assets have to be replaced frequently. Condition-based maintenance has a strong spread of costs. It can be effective, but it can also be expensive – depending on the selected parameters, costs for inspection and costs for maintenance. Preventive maintenance is in the most cases expensive. To reduce risks, the maintenance intervals and usage triggers should be chosen pessimistic. With it, more maintenance is done than necessary which increases the costs. Predictive maintenance can be cost effective, if monitoring is not too expensive. This results from the long-term planning of maintenance when needed. With it, the condition trigger can be chosen lower than in condition-based maintenance.

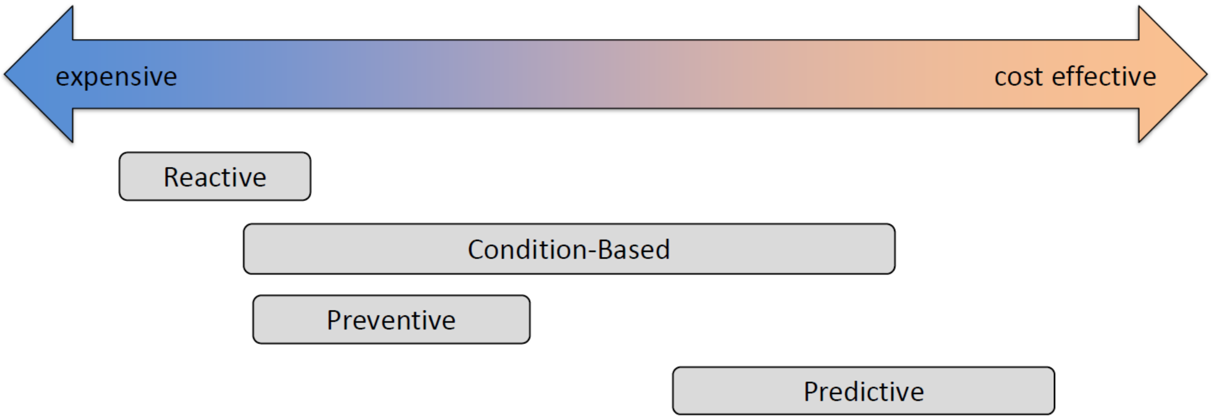

Figure 6. Evaluation with respect to Cost Effectiveness

In Figure 7, the maintenance policies are evaluated with respect to the amount of information the policy provides regarding the behaviour of the infrastructure system. Reactive maintenance provides no additional information since only break-downs are observed. Also, preventive maintenance provides only few information because inspection and monitoring have a secondary role. Condition-based and predictive maintenance gives a lot of information about the infrastructure condition. To observe the deterioration process, monitoring systems are implemented or inspections are performed.

Because in predictive maintenance also information about the future condition is provided, this strategy leads to the best informed situation.

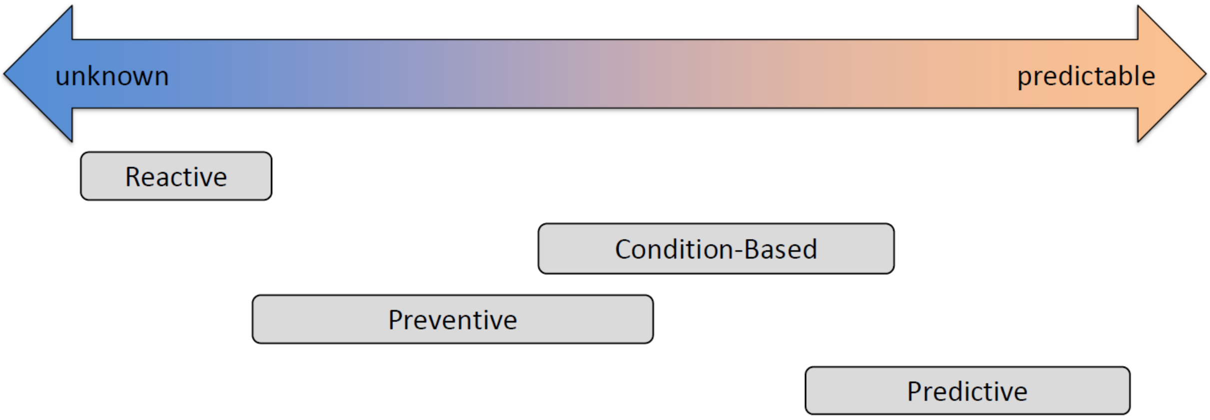

FIGURE 7. EVALUATION WITH RESPECT TO THE AMOUNT OF INFORMATION

In Figure 8, the maintenance policies are evaluated with respect to the implementation effort. Reactive maintenance can be implemented intuitive without defining trigger values and installing monitoring systems. Only inspection to obtain failures is necessary. Also preventive maintenance has a low implementation effort, time triggers can be defined based on expert knowledge. But if the maintenance operator increases the effort for implementation, e.g. by analysing the deterioration process to derive better time triggers, the strategy can be improved in terms of costs or risks. The implementation effort for condition-based maintenance is higher, because monitoring systems have to be installed or the assets have to be inspected closely. The highest implementation effort has predictive maintenance: it requires monitoring, data evaluation and expertise to derive suitable deterioration models in order to predict future condition with a high quality.

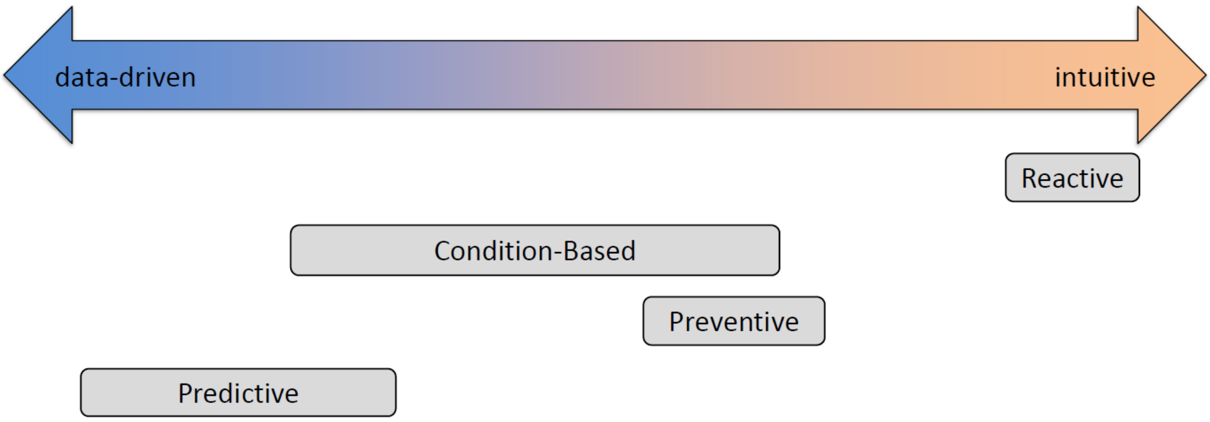

FIGURE 8. EVALUATION WITH RESPECT TO THE IMPLEMENTATION EFFORT

In Figure 9, the different evaluation criteria of the maintenance policies are aggregated into one diagram. The evaluation regarding

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

is summarised as a radar chart.

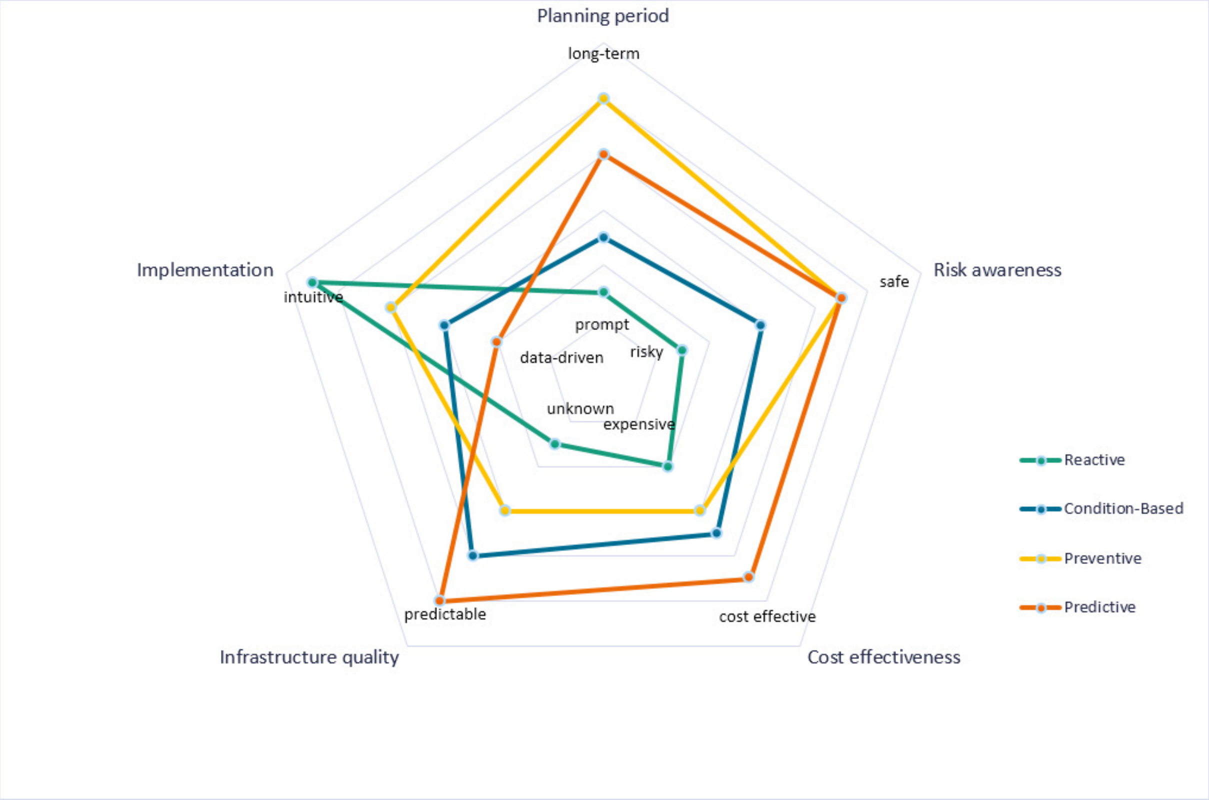

FIGURE 9. AGGREGATED EVALUATION OF MAINTENANCE POLICIES

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

Preventive maintenance also has a low implementation effort, only the maintenance activities and the time or usage trigger have to be defined. Therewith, maintenance activities can be planned in long-term and risks can be avoided. But often, this approach is expensive because preventive maintenance is usual done before needed. This approach also provides only less information about infrastructure quality.

Condition-based maintenance helps to reduce risks and costs, because maintenance is done when necessary. This requires closely monitoring to known the current infrastructure condition. After detecting signs of deterioration, maintenance should be executed promptly. So, the planning period is more short- up to medium-term.

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

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

TABLE 1. COMPARISON OF MAINTENANCE POLICIES

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

3.2 PLANNING PRINCIPLES

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

The way this detailled planning and organisation is done is described by the different planning principles to be applied:

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

In Figure 10 (from [1]), the RBM framework is shown. As every maintenance framework, it starts with data collection. By analysing the data, the failure modes are defined and for each failure the risk is evaluated. Then, the risks are ranked and a plan for inspection as well as a proposal to mitigate risks are defined, e.g. a set of maintenance strategies and inspection intervals or a proposal to install monitoring systems. At the end of the process it is checked whether the maintenance approach can be realised. If not, a new approach has to be defined.

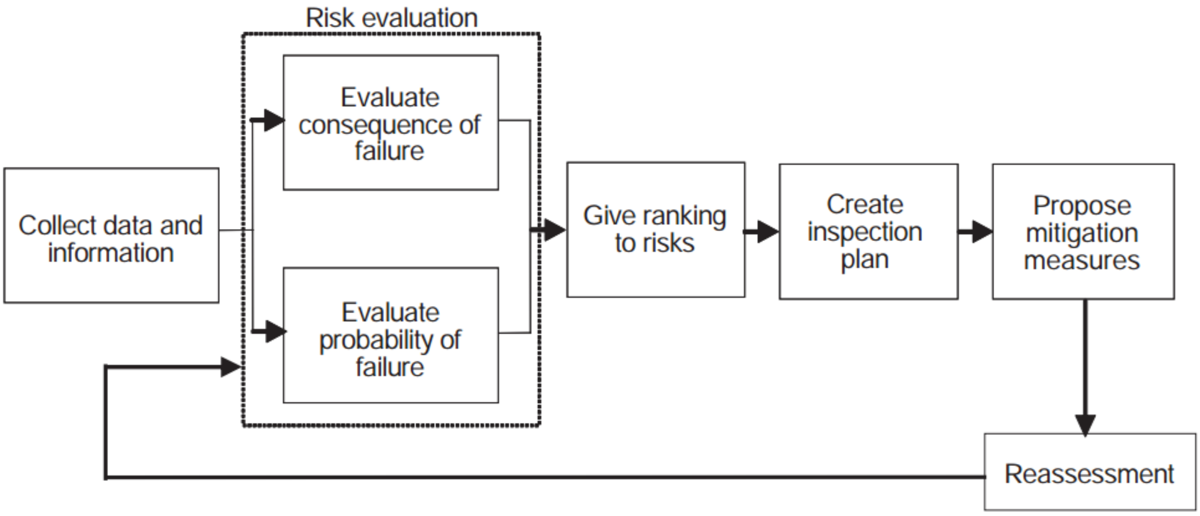

FIGURE 10. RISK-BASED MAINTENANCE

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

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

In the most applications, a mix of risk-based and reliability-centred maintenance will be used because the avoidance of risks and the assurance of system reliability are important for the operator. Both, risk-based and reliability-centred maintenance are rather rule- and experience-based. So, overall performance of these systems depends heavily on the operator and can be hardly controlled. Another problem is that an acceptable working implementation tends to get stuck in itself. With it, innovations and new scientific findings are missing.

In contrast, ***evidence-based asset management (EBM)*** focus on data-driven decisions optimizing clearly defined performance values. Newest scientific findings and mathematical models will be applied. This shifts the function of the operator from preparing of maintenance plans to control them.

To shift to evidence-based maintenance, it is important to develop decision support tools on all levels of maintenance planning. To ensure the adaptability of the different planning tasks to the newest scientific finding, a component-by-component approach is indispensable. Therefore, the task of each component has to be specified and interfaces for the input and output has to be defined. Then, single components can be replaced by new tools to include new scientific results and integrate new construction and maintenance techniques or materials.

In Figure 11 an assessment for the usage of planning principles in the different maintenance policies is made. It is expected that risk-based maintenance will focus on preventive and condition-based maintenance because frequently maintenance and inspection leads to a good overall condition that decreases risks. In contrast, reliability centred maintenance is more condition-oriented, so mainly condition-based and predictive maintenance is selected. With it, maintenance can be done when necessary by ensuring a high reliability. We assume that in future evidence-based maintenance will mainly base on predictive maintenance because of the advanced understanding of deterioration processes. Random or unpredictable failures will be maintained in a reactive, condition-based or preventive manner, depending on the newest scientific findings.

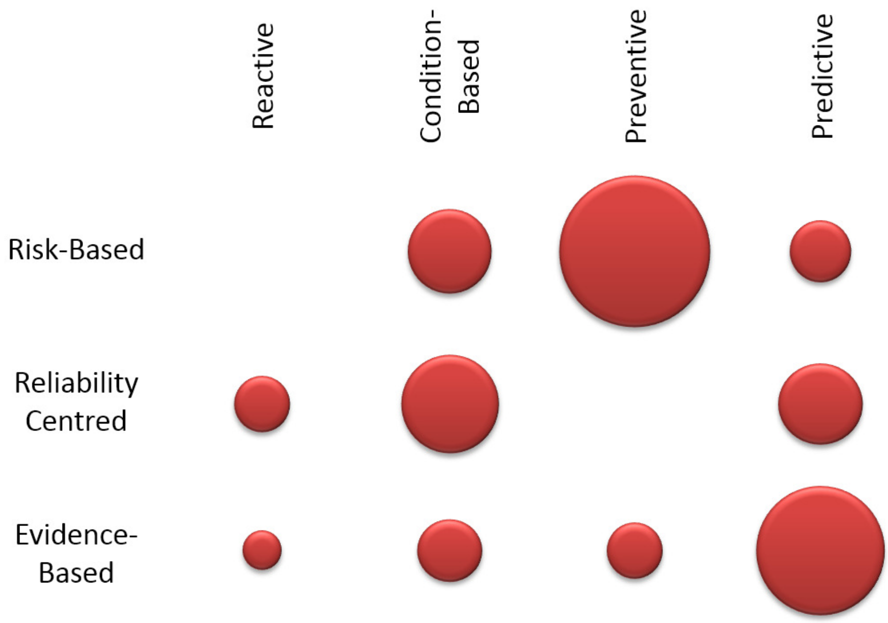

FIGURE 11. ASSESSMENT OF THE USAGE OF PRINCIPLES IN THE DIFFERENT POLICIES

3.3 PLANNING LEVELS

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

It is common practice to separate the planning steps into three so-called planning levels: strategic, tactical and operational (or dynamic) planning. This distinction to some extend reflects the time horizon of planning decisions to be made within the single steps. But since the actual planning horizon to be considered varies amongst different stakeholders and application scenarios - and thus has to be regarded as a variable in the system - other criteria are used to distinguish: the "objects" or "variables" that are subject to planning decisions, the preliminaries under which they are done and the consequences they have in the overall process. An important factor is the level of detail, which increases from strategic to operational.

The most general definition, which is used consistently within the project and all further documentation about it, is given as follow:

- ***Strategic planning*** is concerned with decisions that influence the maintenance management in the long-term. No concrete assets or actual failures in the underlying infrastructure network will be considered, but the behaviour of asset groups and failure modes under the application of certain maintenance strategies (or policies, methods) is evaluated. Planning takes control over the parameters of strategies. Selection of strategies for maintenance and for possession booking, budget allocation, capacity improvement and similar long-term control instruments is done.
- ***Tactical planning*** considers the real network with actual and predicted conditions, and decisions are made that are directly related to concrete maintenance activities, according to selected maintenance strategies. Selection, combination and allocation in a mid-term horizon are the typical decisions defined by a tactical plan. Alignment between maintenance and traffic operation is an issue at this level, whereby e.g. possession windows have to be selected and shifted.
- ***Operational*** (or ***dynamic***) ***planning*** is the most detailed level and considers the actual implementation of single maintenance activities, the scheduling of resources like machinery, staff, material, spare parts. Due to its short-term horizon the possibilities to influence traffic are limited (if any), rather e.g. given train schedules and possession windows have to be met. Real-time information and adaption of existing maintenance schedules is crucial.

4 Ambient Building Blocks

The concept for maintenance planning and the decision support framework as described in Sections 5 and 6 are based on several so-called ***building blocks***, each of them providing a dedicated functionality, processing specified input and delivering specified output. The framework basically defines the linking between these blocks to conduct the maintenance planning process. Besides the building blocks that are at the core of the planning concept - dealing with strategic, tactical or operational planning tasks - there are also important ***ambient blocks*** feeding the planning tasks with information, doing certain computations in the form of pre- or post-processing within the overall complex process.

The functionalities of ambient blocks will not be described in detail here, since they are subject of different Work Packages of INFRALERT's and thus will be handled in different reports. But as they play an important role in the context of the planning concept some of them will be summarised in the following subsections. The proper use of the information provided by these ambient building blocks will be described in Section 5.

4.1 RAMS and LCC

4.1.1 Integrated RAMS and LCC study

Safe and reliable networks with sufficient capacity and availability are a major requirement of today railway and road infrastructures. Due to their complexity and the difficulties to modify the initial design, the performance of the infrastructure depends largely on the maintenance and renewal decisions taken during its life span.

Maintenance decisions are usually leaded by economic criteria in the short-term. Recent changes in the transport sector, e.g. PPP contracts, require to base maintenance strategies in performance criteria. To this end, it is necessary to assess reliability and costs in a holistic way, i.e. considering the whole system and the whole life cycle of the assets. The methodologies used for the assessment of reliability and costs are respectively Reliability, Availability, Maintainability, and Safety (RAMS) and Life Cycle Cost (LCC).

Although RAMS parameters are treated in more detail in WP5, there are several concepts that should be clarified here. According to the standards [2], [3], we define :

- ***Reliability (R)*** as the ability of an item or system to perform a required function, under given environmental and operational conditions, for a stated period of time.
- ***Availability (A)*** as the ability of an item or system to perform its required function, at a stated instant of time or over a stated period of time.
- ***Maintainability (M)*** as the ability of an item or system, under stated conditions of use, to be retained in, or restored to, a state in which it can perform its required functions, when maintenance is performed under stated conditions and using prescribed procedures and resources. It is related with duration and effort required by corrective actions.
- ***Safety (S)*** as the freedom from those conditions that can cause death, injury, occupational illness, or damage to or loss of equipment or property.

RAMS techniques are a central element in maintenance in many different application areas, ranging from manufacturing, electrical engineering, transport, and process industry, to nuclear and space industry [4]. Although widely used in the chemical, electronics or nuclear industry, they are relatively new in the construction sector [5] or infrastructure management [6]. LCC analysis, on the other hand, estimates the most cost-effective option among competing alternatives in maintenance, rehabilitation or construction for a single project and they are common practice in Engineering and Industry.

Traditionally maintenance decisions for railway and road infrastructures have been based on the past experience and expert knowledge. Systematic RAMS&LCC analyses provide a way to integrate this type of analysis in decision support systems, contributing to optimise maintenance strategies and cost-effective decision making.

Recently, RAMS&LCC analyses have attracted much attention in the railway sector [7], with a large number of projects devoted to their development and applications, although their main focus has been on LCC methodologies. On the contrary, few experiences of implementing RAMS&LCC approaches are known in the road sector. There is therefore a need for an integrated study of RAMS&LCC to enhance the cost effectiveness of these infrastructure systems. Moreover, traditional applications of RAMS&LCC in transport infrastructures have followed a deterministic approach in part due to data availability and computer processing capabilities. The predicting deficiencies inherent of such approaches can be overcome using a probabilistic point of view, where RAMS&LCC are described by probability functions expressing the likelihood that a particular RAMS or LCC state will actually occur.

4.1.2 RAMS and LCC Methodologies

As it has been proposed in [8] the idea is that the results of the RAMS and LCC analyses of the system will eventually produce an integrated RAMS&LCC.

RAMS analysis

In an initial phase a combined R and A analysis is carried out to determine how often the system is not achieving the required performance conditions and the possible causes of its poor performance. It is important in this phase to define explicit criteria for failures. Depending on the analysis, failures as consequence of technical equipment, processes, weather or environmental conditions, third parties or a combination of these can be considered. This analysis starts with a determination of the frequency of failures and the non-availability of whole the system. The operational reliability figures are linked to functions executed by the system, and fault trees and other methodologies such as Monte Carlo simulation models can be explored to access relevant R and A parameters of the system.

The M analysis is part of the RAMS and is intended to obtain information of the functional recovery time when failures occur. Several components are included in this analysis such as the following (this will be refined throughout the project): an overview of the maintenance activities, frequency and required maintenance hours depending on the failure mode; risks derived from maintenance activities; resources required by maintenance activities; and a maintenance costs estimation (which will be inputs for LCC).

Finally, an S analysis is carried out to map out system and occupational safety risks for persons and equipment. In this phase, a definition of what is understood by "safe" is important. The analysis provides insights into the possible hazards and the safety levels in terms of personal or social hazards. Safety goals must be established in order to determine the influence of the system safety in LCC and to understand the performance of the system in terms of safety.

LCC analysis

This analysis is intended to determine the total LCC of the system, but it can be determines per functionality or sub-system as well. LCC analysis involves the analysis of the costs of the system or its components over its entire life span. Typical costs may include: acquisition costs (or design and development costs), operating costs (cost of failures, cost of repairs, cost for spares, downtime costs, maintenance costs), and disposal or demolition costs.

Combined RAMS&LCC analysis

In this analysis the cost for non-availability of the system is implemented, which includes as input the previous results for RA, M, S and LCC analyses. The analysis aims at determining what RAM quality and safety (S) level can be achieved for what life cycle costs. The analysis must highlight the cost for the final clients (in our cases the users of the railway and road infrastructure) and attention must be paid to the interface between RAM and Safety, because choices related to Safety may have a direct influence in the Reliability, Availability and Maintainability of the system.

4.1.3 RAMS and LCC Data Flow

The goal of analysing RAMS in a system is to generate input data for the assessment of the sustainability of the system in its life cycle. The outputs of the RAMS allow us to calculate costs and make cost-benefit predictions and analyses. We can define four phases in the RAMS&LCC cycle (see Figure 12), these are: data collection, RAMS simulations, RAMS targets and LCC assessment.

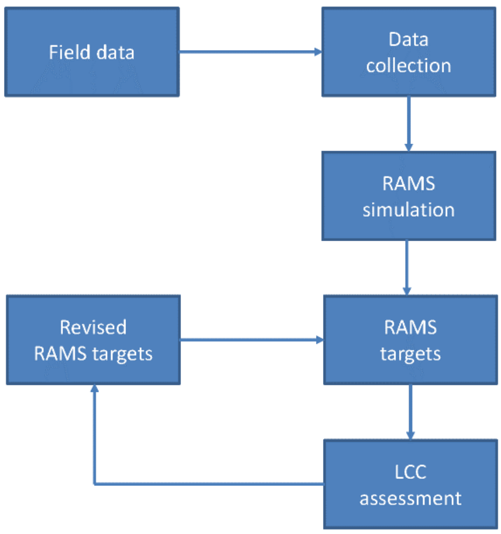

Figure 12. From Component to System Failure

In the first stage ***RAMS data collection*** a study at component level is carried out from data coming from field operation and maintenance feedback. The study of individual component failures provides information about failure rates and other reliability and availability parameters. This information is fed to a simulation with the aim of obtaining RAMS parameter estimation at system level.

In the second stage a ***RAMS simulation*** at system level is carried out. Firstly a model for the whole system is proposed in which interactions between components are established. Different approaches are usually used such as reliability block diagrams, fault trees, Markov chain method, flow networks, Petri nets or Monte Carlo next event simulation.

In the third stage ***RAMS targets***, a comparison with RAMS targets for operation previously established is made. Based on the results from the simulation a revision of these targets may be in order, depending on the performance of the system.

Finally, ***LCC assessment*** is evaluated with the help of the previously calculated RAMS.

4.2 Risk Assessment

4.2.1 The Concept of Risk

Risk can be defined as the effect of uncertainty on the objectives of an organization or project and is generally characterized by certain events and resulting consequences [9]. The result of a risk can be negative or positive and is often measured by the associated likelihood of its occurrence together with the consequences of the occurence. With this strict definition, risk can be seen as a quantity, which can be expressed, measured or calculated using a mathematical formulation. But sometimes also a more qualitative interpretation of risk is used in this sense.

Risk management aims primarily at maximizing control over uncertainty and thereby reducing negative outcomes as well as finding an appropriate treatment when certain events do occur [10]. The fundamental process within the risk management framework is risk assessment, which consists of the identification, analysis and evaluation of possible risks. This provides an understanding of the causes, consequences and probabilities of risks and therefore constitutes the required basis for informed decision-making and the appropriate reaction to certain events [11]. Nevertheless, risk assessment cannot be viewed as an isolated function, but must be realized as an element of the integral risk management process. Risk assessment is usually done in three steps (see Figure 13):

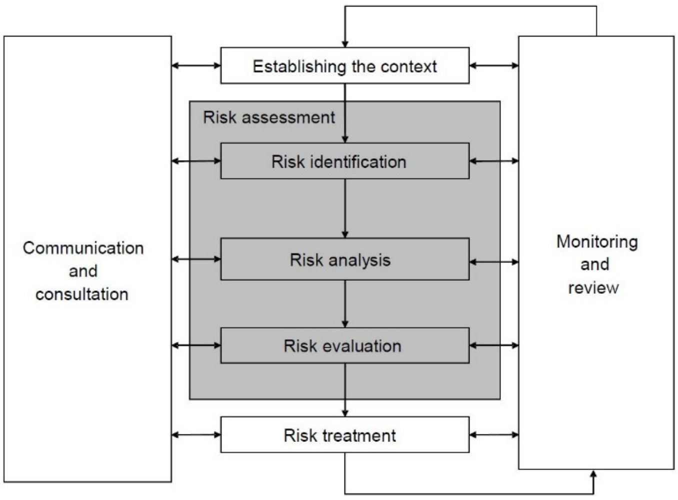

Figure 13. Risk Management Framework

The first step in risk assessment is the ***risk identification***, which aims at generating a comprehensive list of risks based on events that might impact the objectives of the organization or project. In addition to identifying the actual risk, it is also necessary to determine its causes and sources, whether or not these are evident or under the control of the organization, as well as the outcome of the risk [9]. The better the understanding of the sources, the better the outcomes of the risk assessment process and the more meaningful and effective will be the management of risks. Extensive identification of risks is critical, because a risk that is not identified is not included in further analyses. In many application,

risks are linked to a sound concept of ***failures***, which can be seen as triggers of events and thus sources of risks. An analysis of failure modes and their causes supports the risk identification.

The objective of ***risk analysis*** is to develop a quantitative or qualitative understanding of a risk in the context of the organisation or process under consideration. It consists of identifying relevant consequences and their associated probabilities of occurrence so as to determine a level of risk, while taking into account the presence (or absence) and effectiveness of controls [11]. This also includes recognizing the factors that affect consequences and probabilities.

The understanding of risk obtained in risk analysis is used for ***risk evaluation*** to determine the significance of the level and type of risk. To this end, the estimated level of risk is compared to some risk criteria, which are to be defined prior to risk assessment and may be imposed by or derived from legal and regulatory requirements. Risk evaluation establishes the basis for decisions on whether risks require treatment and priorities for treatment [11].

4.2.2 Methodologies / Techniques and Their Application

For the purpose of the planning concept risk assessment can be seen as a building block providing useful information to be used in the decision-making process (as being discussed in Section 5). For this, it is important to understand what are the possible outcomes of risk assessment, and thus what methodologies and techniques are used for risk assessment that could be available for integration into the maintenance planning process.

A highly diverse bunch of methodologies and techniques for risk assessment exists in theory and practice, developed for and applied to diverse domains and industries, amongst others project management, banking and insurance, production and manufacturing systems, maintenance [12]. Appendix 1 gives a brief overview on methods used in the oil and gas industry.

Their suitability for the respective domain and the way they are applied basically depends on the nature of risks and their perception: Risks may be perceived on a technological and operational level, but also on safety, health or political level, and a fundamental distinction can be made whether risks are due to specific malfunctions or "error mechanisms", or whether they are due to unfortunate coincidences [13]. Therefore, the way of assessing risks differ considerably e.g. in civil engineering projects or selecting maintenance strategies. In the domain of interest for INFRALERT - maintenance decision-making - risk mostly is related to technological issues and linked to the operational phase of assets. Thus, used techniques focus on failure modes, effects and root causes.

In the following we give a brief overview on a broad range of risk assessment techniques, which are commonly classified into three main categories: (a) qualitative, (b) quantitative, (c) hybrid or semi-quantitative. The classification is according to the output of the methods: Whilst quantitative methods assign numerical values to risks and apply mathematical models to determine these values, qualitative methods provide rankings or other subjective assessments.

(a) Qualitative techniques

What-if-analysis generates qualitative descriptions of potential problems in the form of questions and responses, and lists recommendations for preventative measures. Brings together a team of experts to analyse a system or process in a comprehensive manner from different perspectives. Mostly used in combination with other, more structured (quantitative) techniques.

Safety audits aim to identify operational procedures or equipment conditions that may lead to hazards or other safety-relevant problems. Resulting reports give an overview on the level of performance for various safety aspects of operations.

Task analysis describes how operators interact both with th system itself and other personnel. The resulting "Task model" provides a formal representation of activities and communications undertaken by operators, mostly in a comprehensive format like a graph.

Sequential Timed Event Plotting (STEP) gives an overview of the sequencing and timing of events that lead to an accident or hazard. The main idea behind the STEP concept is to identify the initiation of the accident through disrupting events or changes. A STEP worksheet contains the time and duration of each event together with the agent causing the event.

The idea behind conducting a ***Hazard and Operability (HAZOP) study*** is that hazards in systems always arise due to some deviation from normal behaviour or operation. To identify possible deviations thus is the main objective of this analytical thinking. The second step is then to ensure appropriate means of prevention for any abnormal case identified. This study is done section-by-section until the whole system or process has been considered.

(b) Quantitative techniques

The ***Proportional Risk-Assessment Technique (PRAT)*** uses a simple formula to quantify risks *R* as

$$R = S \cdot F \cdot P$$

where *S* ist the severity of the (potential) harm caused by the risk, *F* is the frequency factor of the risk and *P* is a factor indicating the probability that the harm actually occurs. All three factors *S*, *F* and *P* are valued in the range 1-10 and are derived from subjective estimations, based on (historical) knowledge and expertise. Therefore, the validity of the risk value *R* strongly depends on the validity of the estimates for these factors. Also, due to the ranking of the factors there are some mathematical problems with the interpretation and comparison of the resulting values *R*.

The ***Decision Matrix Risk-Assessment (DMRA)*** technique assigns a severity rating *S* (as the consequence of failure) and a probability rating *P* (as the probability of failure) to quantify risk R as

$$R = S \cdot P$$

From a risk matrix similar to that of Table 2 the level of risk can be read off.

TABLE 2. RISK MATRIX

	Severity *S*			
Probability *P*	Negligible (1)	Marginal (2)	Critical (3)	Catastrophic (4)
Certain (5)	5	10	15	20
Likely (4)	4	8	12	16
Possible (3)	3	6	9	12
Unlikely (2)	2	4	6	8
Rare (1)	1	2	3	4

(red)	Unacceptable	15-20
(orange)	Undesirable	8-14
(yellow)	Acceptable with controls	4-7
(green)	Acceptable	1-3

Quantitative Assessment of Domino Scenarios (QADS) is used to assess severe hazards or failures that occure as a consequence of the "domino effect", where minor primary events lead to a subsequent cascading sequence of secondary events with escalating severities. The propagation mechanism of this domino effect has to be understood in order to consider the hazard by means of QADS. The method uses the combinatorial properties of the underlying structure of domino scenarios to calculate the overall expected frequencies of secondary events [14].

The ***Clinical Risk and Error Analysis (CREA)*** is a modified version of DRMA specifically designed for the application in the medical domain, using a dedicated method to derive probabilities and severities of so-called error modes occurring during the therapy process [15].

The ***Predictive, Epimistic Approach (PEA)*** provides a methodological framework for combining hard, objective data about rare "abnormal accidental actions" with subjective information expressed as expert knowledge. Probability of failure is seen to be afflicted with an epimistic uncertainty or the "degree of belief" from the engineer's perspective, which can be modelled and simulated. The resulting failure probabilities for rare events will get more reliable using this approach [16].

The ***Weighted Risk Analysis (WRA)*** is based on the idea to compare different types of risk (e.g. societal, environmental, economic, safety) in a certain, well-defined dimension, e.g. on a monetary basis. To find proper weighting factors to balance different risk categories is an open problem, and very much depends on subjective perceptions of decision-makers.

(c) Hybrid techniques

The aim of the ***Fault Tree Analysis (FTA)*** is to reveal the main causes of a failure by deductively identifying the factors leading to an undesired "top event". A fault tree is constructed from events and gates: Events represent basic technical failures, conditions or operator errors, categorised in several classes according to their meaning (e.g. primary, secondary, triggering, commanded), while gates are used to combine events in the logical structure of the underlying technical system or process. Most common combinations are OR and AND gates. Then, the resulting fault tree will be analysed in order to identify events that are causes for failures - occurring singular or in combination. Dependencies between events are recognised by determining minimum cuts in the failure tree. The tree normally will be reduced or re-arranged during this iterative process. Finally, a quantitative analysis of the performance of the process or system in terms of failure probabilities is made on the basis of statistical properties of events. For highly complex systems this calculation is made via simulation.

The ***Event Tree Analysis (ETA)*** uses an inductive approach to determine the consequences of an undesired event and can also be applied qualitatively, quantitatively as well as both combined. ETA is related to FTA in the sense that both methods are complimentary to each other, and ETA focuses on the opposite side of an undesired "top event": The decision tree derived from ETA contains the multiple consequences that emerge after the event occurring. It is used to identify resulting accident sequences and to derive mitigating activities, addressing the relevant, critical failures. Similar calculations as in FTA using probabilities of events will be made to quantify frequencies of possible overall outcomes in the system.

When applying risk assessment and - prior to this - selecting appropriate methods it is crucial to consider issues related to quality and availability of data required for the underlying models. Vague and ambiguous input data lead to poor output, resulting in unacceptable risk indicators [17]. As an example, measuring or even estimating the probabilities of rare events (like failure occurrences, accidents) requires a huge amount of data, which is often not available due to limited time horizon or

simply limited number of occurrences. In many methods the probabilities of risks are rather estimated than reliable, calculated numbers. This means that the significance of statistical values derived from existing data might be very restricted, which has to be considered in the following steps processing these information. The same issue happens with the use of physical or functional models to explain and predict certain events, failures, behaviour etc. that lead to risks. On the one hand, more abstract models could lead to inaccurate results of risk analysis. But on the other hand, there is also a danger that very detailed, complex models could be based on wrong assumptions or make use of data which is available in low-quality only. Then again, resulting risk indicators have to be handled with care. In overall, it has to be studied and evaluated whether the models and data available are suited for the level of detail the resulting risk indicators provide. Or the other way around: The availability and quality of models and data should guide the selection of the scope of risk assessment in further decision-making process, and could even define a limitation in its applicability and usability.

Another point to mention with respect to the application of risk assessment in the planning context is related to the updating of results: Risk assessment often is seen as a one-off, non-repetitive exercise with static results. But in operational environments like maintenance with emergence of new sources of risks and permanently incoming information on events, failures, operations, it should be seen as a dynamic process. Therefore, by integrating risk assessment techniques and their results into the overall maintenance planning concept it has to be assured that there is a regular updating mechanism for risk indicators. Most methods and techniques in use are suited to be applied in such a way, although it has to be checked whether the effort for updating justifies the benefit of "better" results.

4.3 Condition Information

This section presents a summary of models used to assess the condition of the assets (nowcasting) and the prediction for future decisions (prognosis or forecasting). Examples for road and rail assets are presented where the input data and the output are described.

4.3.1 Nowcasting

Nowcasting focuses on the known and knowable and is the basis of a robust decision-making process. From infrastructure management viewpoint, there are several inspection methods and systems that can be used for nowcasting, i.e. assessing the current condition of the assets. The diagnosis carried out to obtain the asset condition is a complex job that involves in-depth analysis of the system and often involves experts performing inspections and analyzing features with regard to established thresholds. For linear assets such as road and railway, nowcasting is still a highly manual procedure. Examples of inspections used for nowcasting is presented in Table 3. For a full description of condition indicators see Deliverable D3.1. The data of geometry monitoring of the road and rail infrastructure carried out using standardized measurement vehicles are used for forecasting as discussed in section 4.3.2.

TABLE 3. NOWCASTING ON LINEAR ASSETS USING EXPERT ASSESSMENT DURING INSPECTIONS

Infrastructure asset	Feature	Method/Inspection	Output
Road pavement	Surface defects and structural conditions, e.g. cracking	Visual inspections (or automatic assessment based on image processing)	Proportion of defects and cracks within a certain area. Estimated cracking rate (%) and surface defects rate (%) Severity level of defect
Railway track	Rail profile, gauge, surface defects (RCF, head checks, squats etc.) and corrugations	Ocular inspections, contact based measurement systems and optical systems	Severity level of defects [no. or no./km]
Railway switches and crossings	Switch blade position	Continuous monitoring using sensors	Malfunctions per switch type over a short time horizon [no. or %]

In addition to the existing expert-based assessment for nowcasting, there are other unsupervised and semi-supervised methods that can be used when measurement vehicles are employed for collecting large volume of data. These methods can be based on residual analysis where the deviation of a measured signal from a reference signal representing a healthy condition is analyzed, e.g. utilizing physics-based models. Other techniques are symbolic models (fuzzy logic), multivariate statistical analysis, machine learning, and artificial neural networks. These techniques depend on high amount of relevant data for both healthy and unhealthy conditions. Features derived from inspections and condition monitoring data are combined with asset information (e.g. type and age of various asset types), operational data (e.g. number of axles, and line speed), maintenance data (e.g. number of repairs, repair level), and contextual data, (e.g. weather, location of asset). Data pre-processing is an import step for the implementation of these techniques.

4.3.2 FORECASTING

Forecasting can be done by using four main approaches: physics-based, data-driven, symbolic, and hybrid. Condition indicators commonly used for forecasting of road and railway assets are pavement and track geometry, respectively. The main indicators for roads are the International roughness index (IRI; longitudinal unevenness) and rut depth (transverse unevenness). Condition indicators for track geometry are: Gauge, Cross level, Cant, Longitudinal level, Twist, and Alignment. The deviations of the above geometrical condition indicators from the mean or designed geometrical characteristics are often used for the purpose of maintenance prediction and planning. Furthermore, these geometrical parameters can be aggregated to a quality index for benchmarking, monitoring maintenance and renewal and strategic decision making and for both road and railway infrastructure.

Physics-based approaches focus on the load of traffic since a large part of infrastructure degradation depends on the load of every axle passage. When the material is subjected to heavy loads it deforms and the deformation depends on the magnitude of the stresses and/or strains the material is exposed to. Unfortunately, it is difficult to determine the resulting deformation even if the load can be determined. Therefore, empirical data are necessary for physics-based approaches. The stresses and

strains can be calculated analytically, but the relationship between the stress/strain and the degradation needs to be estimated from experimental data.

- For road, AASHTO Road design guide is a commonly used model. The properties of the base and subbase are described by a structural number, which is a function of the thickness and modulus of each layer and the drainage conditions. From the model, number of equivalent single axle load (ESAL) that the road tolerates can be extracted, i.e. total ESAL to pavement failure. The failure is based on the Present Serviceability Index (PSI). The PSI is related to the road geometry features IRI and rut depth (RD) as exemplified in the following equation

$$PSI = a_1 \cdot e^{b_1 \cdot IRI} + a_2 RD^2 + a_3\sqrt{C+S+P} \tag{1}$$

  where C is the total cracked pavement area, S is the total pavement disintegrated area (pot-holes and raveling), P is the pavement patching area, and a and b are experimental constants. The remaining life (RL) can then be calculated based on the total ESAL to date (N_p) and total ESAL to failure ($N_{1.5}$), where 1.5 denotes the PSI threshold, as in

$$RL(\%) = 100\left(1 - \frac{N_p}{N_{1.5}}\right) \tag{2}$$

- For railway track deterioration, there are several physics-based models that have been proposed in the past and are still in use. The track deterioration model suggested by ORE (Office de Recherches et d'Essais de l'Union Internationale des Chemins de Fèr) is one of the commonly applied models. The deterioration model as given in equation (3) below accounts for two phenomena: the deterioration directly after maintenance (tamping) e_o and the traffic induced deterioration. The second part depends on traffic volume T, dynamic axle load Q and speed v.

$$e = e_0 + hT^{\alpha}(2Q)^{\beta}v^{\gamma} \tag{3}$$

  where h is a constant and the parameters α, β, and γ have to be estimated from experimental data. This model can be adapted to give the kind of outputs that is useful for the requirements of INFRALERT. Given a specific operational history of a railway section, the remaining axle load or traffic volume that should be anticipated before reaching a predefined intervention or alert threshold can be estimated. The remaining traffic volume to reach the threshold is referred to as remaining useful life and can be used for alert management and maintenance planning.

The deterministic models can be enriched with stochastic information by defining parameters or variables as stochastic variable, e.g. the parameters *a* and *b* for the calculation of PSI or the traffic volume per time to define a time-dependent stochastic degradation with equation (3). The resulting stochastic functions can be analysed by means of simulation to get the requested stochastic information.

An overall view of data-driven models for road and rail track degradation and corresponding outputs are presented in Table 4.

- The input to the models is time series of geometry data. In order to establish these time series, the infrastructure network is segmented into homogenous segments that could be of several hundred meters.

- The output is commonly a yearly increment of the geometry features from which an estimation of remaining life (RL) can be deduced. To be used for decision support, the output of a forecasting model should preferably include a confidence interval associated to the estimated RL. The confidence or prediction intervals describe the inherent uncertainty of the

deterioration process and future operation as well as the errors associated with forecasting techniques being used. The decisions based on forecasting are then based on the bounds of these confidence intervals rather than a specific deterministic value.

TABLE 4. DATA-DRIVEN APPROACHES FOR ROAD AND RAIL TRACK DEGRADATION

Input	Method	Description	Output
Road and track geometry (time series data)	Trend analysis	The simplest form of forecasting is trend analysis of a condition indicator correlated to the remaining life. The trend is calculated based on time series of data using regression analysis and then extrapolated.	Yearly increment
	Markov models	Markov models assume that a system can be - at any instant - in only one of a finite number of states. By defining probabilities associated with each state, as well as probabilities associated with transitioning from one to another, probabilities of future failure can be estimated.	Average yearly increment
	Bayesian estimation	Two common approaches for implementing Bayesian network models are Kalman and Particle filtering. The Kalman filters estimate the underlying state of a dynamic system (a-posteriori PDF) by extrapolating from prior state and noisy measurements. Kalman filtering assumes that the dynamic system is linear and that the noise is Gaussian; therefore a number of modifications and alternatives have been explored, one being the Monte Carlo-based particle filter.	An estimated time to failure with associated confidence interval
	Artificial Neural Networks	Artificial Neural Networks (ANN) estimate the remaining life from a mathematical representation of a system derived from observation data. ANN relies on sufficient and adequate training data to adjust network node weights.	Yearly increment

4.4 ALERT AND INTERVENTIONS MANAGEMENT

One of the pillars of the project is the development of an alert management system (WP4). This system will analyse asset condition and operational information to provide alerts whenever the infrastructure reaches, or is close to reaching, a critical level in the present time or in the near future (or for a given scenario). This threaten is identified when the evolving behaviour of one or several features overcome the limits of standardised thresholds. The severity of an alert can be theoretical assessed based on the proximity to those defined limits.

In methodological predictive maintenance, the appearance of alerts is just based on forecasting the evolving behaviour of some features, and the know-how of maintenance staff from prior situations, which in some cases might not be reliable enough. This situation can be improved by the use, in a methodological way, of recorded historical maintenance information adequately rated and assessed. Therefore, the alert management system will be able to combine the current and predicted asset condition with operational and historical maintenance data, to get information about the maintenance tasks that are necessary to avoid later severe degradation or mismatching of safety and/or comfort conditions. By means of data mining methodologies, a prioritised listing (ranked on severity level) corresponding to the alerts generated by all assets of a linear transport infrastructure will be generated.

This section explains the modules for alert generation, which are:

- ***Alerts***: Levels & Reliability & Severity
- ***Prediction***
- ***Prioritising alerts***

The flow diagram of the first module ***Alerts*** is shown in Figure 14. For the asset a_i, the system makes a request to the data base, which includes the functional tree, for all its hierarchical related assets. For each element of the resulting set, all the historical features and historical interventions are collected. Furthermore, the condition features (nowcasting and forecasting) for asset a_i is requested to WP3. This input information related to the given asset will be the basis for the segmentation procedure. This procedure is responsible for breaking down the road/rail network into a series of segments. These segments are classified into groups with homogeneous characteristics in order to be used in prediction models.

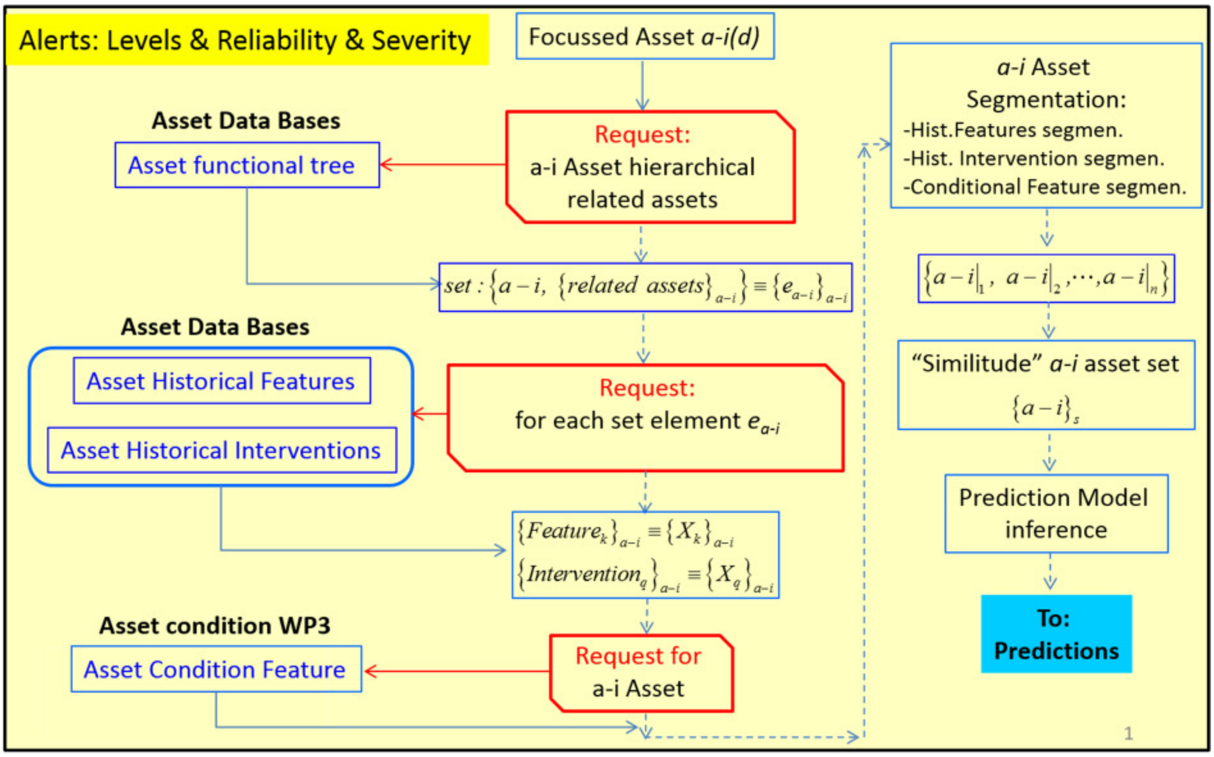

FIGURE 14. MODULE ALERTS: LEVELS & RELIABILITY & SEVERITY

To execute the prediction model it is necessary to specify the time scenario t_{m+k} (e.g. short, medium or long term). The module ***Prediction*** (see Figure 15) automatically determines the foreseeable time to surpass the limit reference (e.g. exceptional limit).

In a first level, the severity is calculated as a function of the difference between the standard limit and the most probable estimated value of the feature (***technical severity***). Moreover, an external criterion also applies for determining the severity level of the alert (***additional severity***). This field is optional and can be used by other modules of the software. For example, this criterion may take into account the cost of do-nothing, the cost of the intervention, etc.

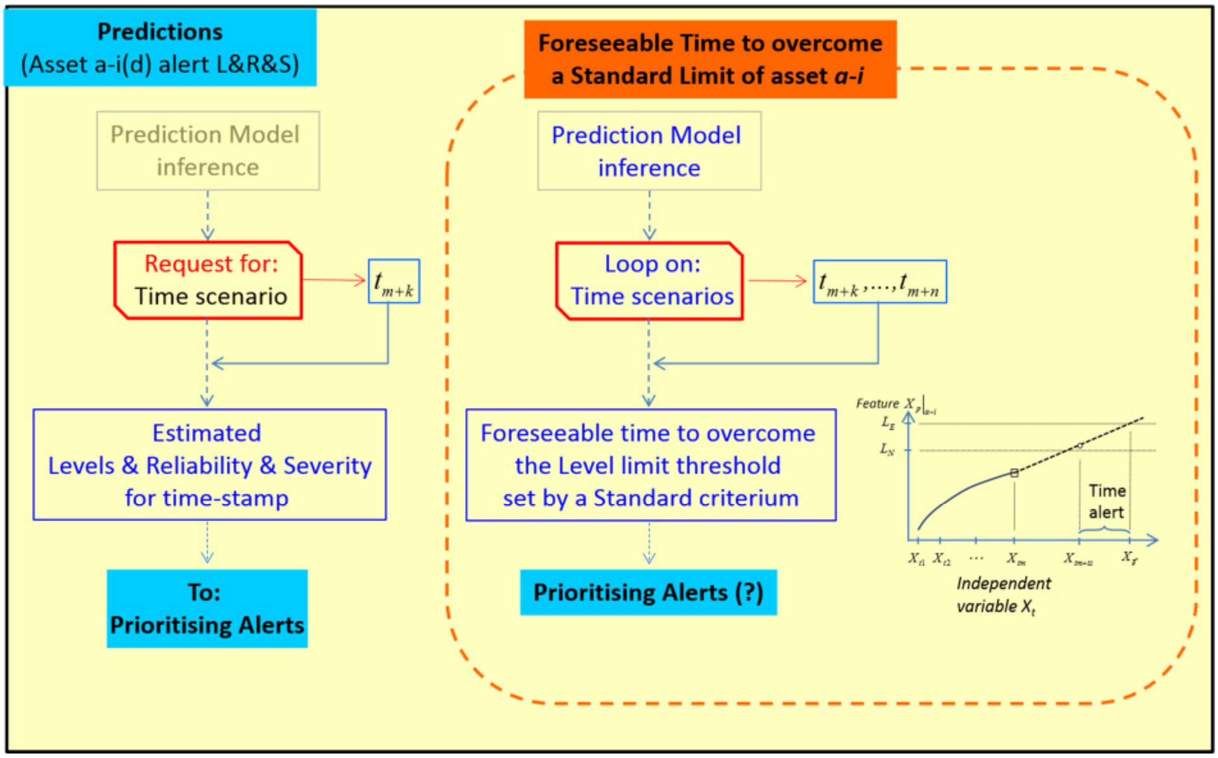

Figure 15. Module Prediction

This information and the previous one generated in the module ***Alerts*** are used as input data of the module ***Prioritising alerts*** (see Figure 16). This module infers the two most likely interventions (based on historical data) associated to the alert. Furthermore, looking into the cost databases, the module determines three different costs for each most likely intervention. Doing this for all assets, the prioritised listing is generated.

Due to the fact that asset historical features and the asset historical maintenance interventions are known, it is possible to associate a combination of levels of features with a failure. In this way, other field of the prioritised listing will be the two (for instance) most probable failures associated to the alert, as well as the maintenance interventions.

Therefore, the output of WP4 is one table that stores the prioritized list of alerts with a level of severity associated to each alert, the most probable failure, the foreseeable time until this failure and the most probable intervention needed. A more detailed explanation of WP4 output can be found in D4.1.

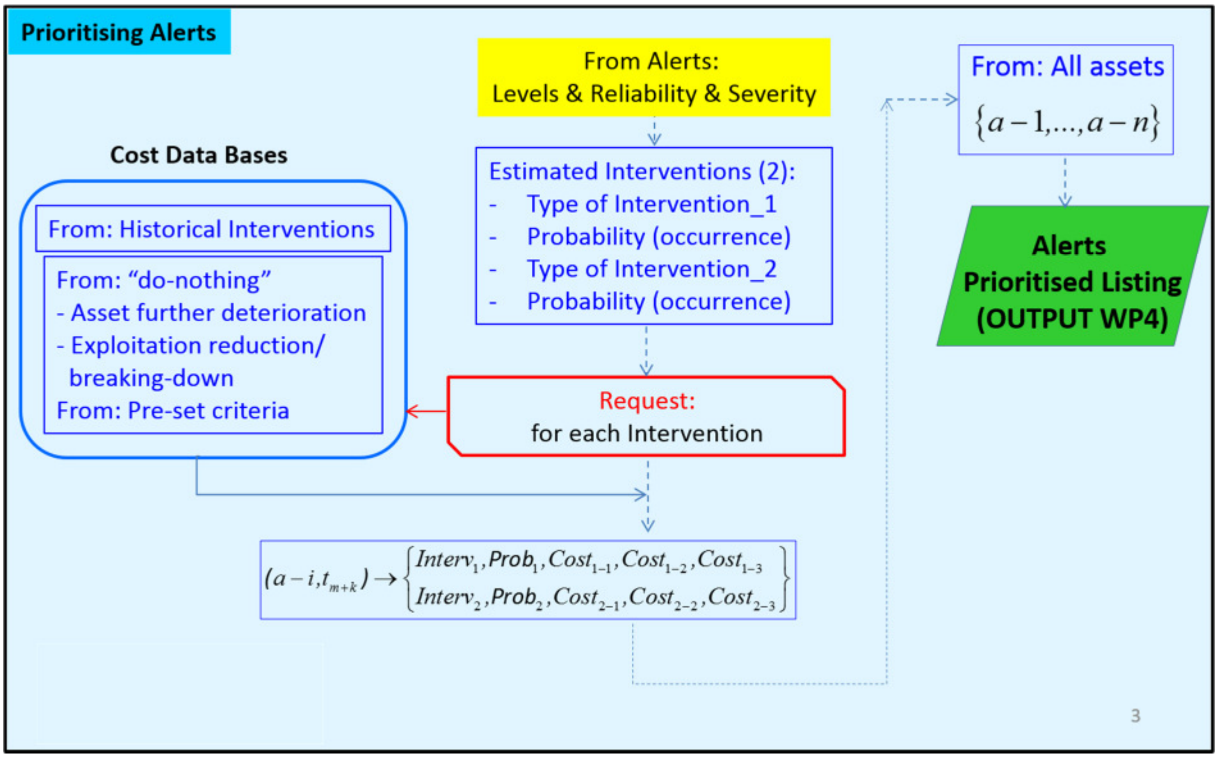

FIGURE 16. MODULE PRIORITISING ALERTS

5 Concept for Condition- and Risk-based Maintenance Planning

The core of INFRALERT's concept for condition- and risk-based planning is a mathematical methodology to use probabilistic data for decision-making in optimisation problems. This methodology allows to deal with uncertainty in maintenance and interventions planning, which is a necessity that arises in practical applications.

"Use of probabilistic data" is a recurrent theme in INFRALERT developments, appearing all over from asset condition evaluation to decision support as seen in the previous Section 4:

Models for nowcasting and forecasting developed in the asset condition component provide not only single estimations of current or future values for relevant parameters, but they will also be used in the alert management system to derive probability distributions (or their characteristic parameters). The future (unknown and random) degradation and defect evolution will be described as stochastic processes, and functions for transition probabilities can be determined using the output of prognosis models. In the same way, RAMS/LCC models will deliver probabilistic information, since it describes output parameters by means of stochastics.

Finally, in the optimisation models for decision support all this probabilistic information describing maintenance alerts including conditions, defects and their transition, variables for maintenance interventions (like costs, resources) and RAMS parameters will be modelled as stochastic variables. Rather than traditional optimisation with deterministic, static variables and objectives an approach based on stochastic modelling for optimisation under uncertainties is applied. In such a setting, a solution methodology has to be developed that makes it possible to have "a look into the future" and to estimate or determine the consequences of decisions to be expected in the future. In INFRALERT, this will be done e.g. by applying the Monte Carlo Rollout method, a well-established technique for optimisation under uncertainties that combines classical heuristic planning with scenario simulation. In this way, decisions to be taken now and having an effect in the future can be balanced with regard to robustness of planning.

The objective of this section is to describe the ideas of the concept for maintenance planning used in INFRALERT. In particular, this is done via a translation into

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

First, we give a brief overview on the theoretical background that is used for the mathematical concept.

5.1 Theoretical Background: Optimisation under Uncertainty

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

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

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

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

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

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

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

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

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

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

There are several mathematical concepts dealing with this kind of optimisation under uncertainty, which are the background of the modelling and solutions techniques applied in INFRALERT's planning concept. The most important ones are considered in the following.

5.1.1 Robust Optimisation

The idea behind this concept is that uncertainty in input data leads to uncertainty in the output, i.e. in the objective values achieved by solutions. To tackle this, one has to look for solutions that are robust or insensitive with respect to input uncertainty, i.e. that have a small deviation in the possible output.

Theory discusses two main cases, distinguishing whether or not input uncertainty can be modelled stochastically by random variables with known probability distribution functions:

If not, uncertain parameters are threated as belonging to a certain ***range of values*** instead of using inappropriate point estimates as exact values. Consider the general optimisation problem from above, now the objective function $f(x)$ is replaced by its uncertain version $f(x,u)$ depending on decision variables $x \in M$ and uncertain parameters $u \in U$. The standard formulation of a robust optimisation problem is to find a ***robust solution*** $x^* \in M$ such that

$$\max_{u\in U} f(x^*,u) \leq \max_{u\in U} f(x,u) \quad \forall x \in M$$

This so-called ***Minimax principle*** is an extremely conservative approach, as it looks for such solutions that protect against worst-case scenarios related to possible realisations of $u \in U$. This can lead to solutions x_1 that provide a rather poor objective $f(x_1,u)$ for all of the potential values for $u \in U$, but there could be x_2 with much better objective values $f(x_2,u) \ll f(x_1,u)$ for "most" of $u \in U$ and only slightly worse $f(x_2,u) \geq f(x_1,u)$ for "some" u.

For this reason, there are several techniques to relax this rather demanding formulation occurring in the minimax principle. One is based on the observation that for some applications the uncertain parameters $u = (u_1, \dots, u_k)$ have a much smaller aggregated deviation than suggested by the potential range for single parameters in the worst-case. As an example, if the single u_i are real-valued parameters from a potential range $u_i \in [a_i, b_i]$ then it is unlikely that all parameters u_i are at the border of its range. In such a case, an artificial constraint on the aggregated deviation - for example on the value

$$\sum_{i=1}^{k} |u_i - \frac{b_i - a_i}{2}|$$

- is constructed, leading to a smaller subset of potential parameter values $\overline{U} \subset U$. This set simply replaces U in the robust problem formulation:

$$\max_{u\in \overline{U}} f(x^*,u) \leq \max_{u\in \overline{U}} f(x,u) \quad \forall x \in M$$

A different technique uses the fact that there is - besides the "potential" range U - sometimes also a natural "nominal" range $N \subset U$ from which uncertain parameters "normally" take their values. This much higher relevance of N compared to U shall be taken into account somehow by the optimisation modelling, in the sense that if a solution x_1 performs very good for all "normal" $u \in N$, but only slightly worse for $u \in U \backslash N$ than the robust minimax solution, this could nevertheless be preferred according to the height of the deviation in objective values. This consideration leads to the formulation of an adapted objective function $\tilde{f}(x,u)$, e.g. by modifying the original $f(x,u)$ in the form

$$\tilde{f}(x,u) \coloneqq f\big(x, \pi(u)\big) + g(x,u)$$

where $\pi: U \to N$ is a projection from $u \in U$ to "its" nominal value $\pi(u) \in N$ (with $\pi(u) = u$ for $u \in N$), and $g(x,u)$ is a penalty term, balancing between the actual objective value $f(x,u)$ and the projected one $f(x,\pi(u))$ (with $g(x,u)) = 0$ for $u \in N$).

The second main case evolves if it is possible (or appropriate) to apply a ***stochastic modelling*** to uncertain parameters, using known or presumed probability distributions. The common notation differs from the one used above, to reflect this stochasticity and refer to the usual conventions: Uncertain parameters are expressed as random variables ξ on a probability space $(\Omega, \mathcal{F}, P)$, for short we write $\xi \in \Omega$ with probability measure $P_\xi = P$.

The main approach to handle stochastic uncertainties in a robust manner here is to introduce a functional Φ which is applied to the stochastic objective function $f(x,\xi)$ in order to obtain a deterministic version of it, which leads to a "classical" deterministic optimisation problem to find a solution $x^* \in M$ such that

$$\Phi(f(x^*,\xi)) \leq \Phi(f(x,\xi)) \quad \forall x \in M$$

(Note that the functional $\Phi(\cdot) \coloneqq \max_{\Omega}(\cdot)$ is the special case leading to the minimax principle discussed above.)

Let

$$\mu(x) \coloneqq \int_\Omega f(x,\xi) dP_\xi$$

be the mean value of $f(x,\xi)$ and

$$m^k(x) \coloneqq \int_\Omega \big(f(x,\xi) - \mu(x)\big)^k \, dP_\xi$$

be the k-th order moments of the random function $f(x,\xi)$, in particular $\sigma^2(x) = m^2(x)$ is the variance of $f(x,\xi)$. A typical and obvious choice for the functional Φ is

$$\Phi(f(x,\xi)) \coloneqq \mu(x)$$

The resulting optimisation problem is referred to as ***mean value optimisation***. This version uses only very limited, almost minimalistic information about the probabilistic features of $f(x,\xi)$, since the mean value is by far not a sufficient representation of the probability distribution function behind. The main advantage, on the other hand, is that the mean value has a very obvious interpretation and thus is easily accepted by decision-makers. From a mathematical point of view it has the nice property that in case an approximation approach is used - which is the normal way of solving complex problems - the mean value is a fast converging statistical parameter. Thus, an accurate prediction quality can be achieved with low computational effort.

By including higher order moments the generalisation to the ***mean value penalty optimisation*** is achieved, defined by the functional

$$\Phi\big(f(x,\xi)\big) \coloneqq w_1 \mu(x) + \sqrt{\sum_{k=2}^{n} w_i m^k(x)}$$

where $w_1, \dots, w_n$ are weights to be selected accordingly. By increasing n the amount of information about the probability distribution can be increased, but of course to the cost of computational complexity. An important special case of this is obtained by setting $n = 2$ and $w_1 = w_2 = 1$:

$$\Phi(x) = \mu(x) + \sigma(x)$$

which again has an intuitive meaning: In this case, robust solutions are the ones that minimise the mean value plus the standard deviation, which adds some "safety" to the objective values that are to be expected. As always, whether this additional probabilistic information will be seen as sufficient or not strongly depends on the subjective evaluation of the decision-maker, and should correlate to his knowledge about the underlying stochastics of the uncertainty, i.e. how strong he "beliefs" in the relevance of this statistical concept and e.g. the sigma rules.

Similar to the first case of range modelling, these stochastic models can be combined with constrained and multi-objective optimisation:

For instance, a typical constrained optimisation problem evolves using the following formulation: Find $x^* \in M$ such that

$$\begin{aligned} \mu(x^*) &\le \mu(x) \quad \forall x \in M \\ m^k(x^*) &\le c_k \qquad k = 2, \dots, n \end{aligned}$$

Here, the constants c_k play the role of bounding the k-th order moments of the objective values $f(x, \xi)$ obtained. Again, additional probabilistic information similar as in the mean value penalty optimisation formulation is integrated, but it is challenging to find proper values c_k to bound the moments, since it is not clear whether a feasible solutions can be found at all and if so, whether the bounds are too restrictive in order to yield to a "good" objective value $\mu(x)$.

Therefore, multi-objective approaches can be helpful to generate the Pareto set of optimum solutions with regard to the different moments, i.e.

$$\begin{cases} \min_{x \in M} \mu(x) \\ \min_{x \in M} m^k(x) \quad k = 2, \dots, n \end{cases}$$

Of course, this increases the complexity of the optimisation problem and introduces an additional decision-making criteria: how to select candidates from the Pareto set.

5.1.2 Probabilistic Constraint Programming

The above optimisation models considered the consequences of uncertainty as uncertain output, thus understood the objective function as a random function $f(x, \xi)$ depending on a random input variable ξ. But in most applications (in particular related to maintenance planning) the uncertain input also affects the feasibility of the resulting solutions. In mathematical terms, this means an uncertainty in the underlying constraints of the optimisation problem, so the feasible set of solutions $M = M(\xi)$ now depends on uncertain input parameters to be modelled as random variables. Then, the membership of a solution $x \in M(\xi)$ itself is a stochastic property which is met with a certain probability only.

This leads to a mathematical formulation using ***probabilistic constraints***: Find x^* such that

$$\begin{aligned} f(x^*) &\le f(x) \\ P\big(x^* \in M(\xi)\big) &\ge p \end{aligned}$$

for some probability bound $0 < p < 1$ or - using the explicit formulation of the feasible set via constraints:

$$\begin{aligned} f(x^*) &\le f(x) \\ P(h_i(x^*, \xi) \le 0) &\ge p_i \quad \forall i = 1, \dots, m \end{aligned}$$

with m probabilistic constraints.

Of course, any combination with random objective function modelling from the above approaches is possible. But it is also possible to handle uncertain output as a probabilistic constraint in the following form: Define a maximum objective f^* value which should be guaranteed as an upper bound. Then simply add to the "normal" probabilistic constraint or robust optimisation problem the additional probabilistic constraint

$$P(f(x,\xi) \geq f^*) \leq p_f$$

which bounds the probability of exceeding f^* by at most p_f. Often, such an approach is embedded in an iterative optimisation process where a proper value for f^* is determined stepwise in order to find a minimum acceptable upper bound that allows for feasible overall solutions x^*.

5.1.3 Multi-Stage Stochastic Programming

All the approaches described so far do not consider the case when the decision-maker is able to react on changed situations and new information by adapting the solutions found, be it because they become infeasible or because they lead to bad solutions that could be revised and improved. This case represents the ***dynamic aspect*** of uncertainty, where the solution of the optimisation problem is turned into a solution process.

The concept of multi-stage programming deals with this dynamic uncertainty in a very explicit manner: A stagewise, recurrent process of decisions followed by observations is applied, using additional decision variables in each stage.

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

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

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

$$G(x,\xi) = \min_y g(x,y,\xi)$$

which is formally used to express the first stage problem:

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

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

Of course, this approach can easily be extended to a general multi-stage process. Here, the solution process consist of the sequence

stage 1 decisions → observation → stage 2 decisions → ... → stage m decisions

and can be expressed mathematically by using variables $x^{(i)} = (x_1^{(i)}, \dots, x_n^{(i)})$ for each stage $i = 1, \dots, m$ and recourse functions $f^{(i)}(x^{(1)}, \dots, x^{(i)}, \xi)$ for stages $i = 2, \dots, m$ depending on variables up to stage i (and the uncertainty ξ). Again, stages are solved consecutively using the expected minimum solution from the previous (higher) stage.

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

5.1.4 Markov Decision Process

Markov Decision Processes (MDP) are another basic concept to deal both with stochastic and dynamic uncertainty. It provides a very general instrument to model a dynamic and stochastic planning process with subsequent actions taken by a decision-maker and observations of the surrounding "nature", again leading to actions as reactions to the consequences.

More formally, an MDP can be defined by a 5-tupel (S, A, P, R, γ) where S is a set of states, A is a set of actions, P is a probabilistic state transition function, R is a reward function and γ is a discont factor. An MDP represents a time-discrete control process. At each point in time t the process is in a certain state $s_t \in S$, and the decision-maker has to select an action $a_t \in A$. Then, a new state $s_{t+1} \in S$ will be reached at time $t + 1$ by randomly moving according to the state transition function, i.e. each state s_{t+1} will be reached with probability $P(s_{t+1}| \, s_t, a_t)$. Taking the action gives the decision-maker a reward $R(s_t, a_t, s_{t+1})$ in each step.

The process satisfies the stochastic Markov property in the sense that the transition to the next state s_{t+1} only depends on the current state s_t and the action a_t chosen, but not on the complete past of the process.

A simple example of an MDP with three state $S = \{s_0, s_1, s_2\}$ and two actions $A = \{a_0, a_1\}$ is shown in Figure 17. Here, for instance a transition from state s_0 to s_2 by taking action a_0 is possible with probabilty $P(s_2|s_0, a_0) = 0.5$. Rewards are achieved only by transitions $s_1 \overset{a_0}{\to} s_0$ with $R(s_1, a_0, s_0) = 5$ and by $s_1 \overset{a_0}{\to} s_0$ with $R(s_2, a_1, s_0) = -1$.

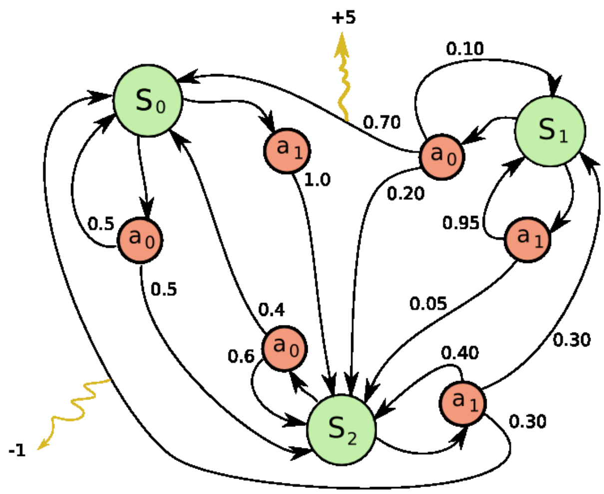

FIGURE 17. EXAMPLE FOR A MARKOV DECISION PROCESS

The overall goal in an MDP is to find a so-called ***policy***, which is a function

$$\pi: S \to A$$

that specifies the action $a = \pi(s)$ to take at each state $s \in S$. Considering a (possibly infinite) time horizon, the overall disconted reward of policy π is

$$\sum_{t=0}^{T} \gamma^t R(s_t, \pi(s_t), s_{t+1})$$

which has to be maximised. Note that this value is a random variable, thus the task is to maximise its expectation. Theoretically, it is possible to find optimal policies π^* by means of dynamic programming. But in practice this is only possible for MDPs with very small state and action spaces and finite time horizons, which is mostly not the case. Therefore, approximation methods like value iteration, policy iteration, prioritised sweeping are used.

There are several generalisations to the basic MDP model, which handle e.g. the case that the resulting states $\pi(s)$ are only observable with a given probability (Partial Observable MDP) or that the transition probabilities T or rewards R are unknown and will be learned during execution (Reinforcement Learning).

5.1.5 Scenario Generation and Sampling Techniques

A fundamental technique used for the solution of optimisation problems under uncertainty as discussed in the above subsection is that of ***sampling***: Sampling refers to the generation of concrete realisations for random variables, either as single numerical values or in a combined manner by generating complex ***scenarios***. In this meaning, a scenario is a representation (or realisation) of the world modelled in an optimisation problem for a particular possible future. When generating scenarios via sampling, often several single probability distributions are combined to a more or less complex probabilistic simulation model, and the sampling process itself is referred to as ***simulation***.

When it comes to the use of random scenarios to derive optimal (or near-optimal) decisions for maintenance planning, the challenge arises how to do the sampling in a way that the underlying "real world" is reflected best. In this specific area and the class of problems derived from there we have to keep in mind the following circumstances:

- Planning models are very complex and hard to solve, even with the use of approximation and simulation techniques. Therefore, only a limited number of scenarios can be used, not covering the full spectrum of possible future developments.
- On the other hand, events relevant in maintenance planning such as failures, defects, break-downs, accidents, etc. are more or less rare events, that have a relatively low probability. In random scenarios they appear rather seldom, but have a huge influence or cause important consequences. In mathematical terms, they are related to so-called heavy-tailed (or long-tailed, fat-tailed) probability distributions, meaning that outliers still have sufficient influence on the overall expected result, despite having low probabilities. In other applications of optimisation under uncertainty and risk this phenomenon is referred to as "black swans": a surprising event with major effects [18].
- Often the uncertainty in a planning problem is induced by a complicated system or process (e.g. due to complex physics or mechanics) or comes from a context that can not be well-described by mathematical formulations (e.g. human behaviour, external influences). In such

cases, the underlying probability distributions of uncertain parameters are not known explicitly, and cannot be derived formally. There are just historical observations and assumptions available.

From this it follows that the design of sampling techniques for scenario generation has to be carefully reviewed, since otherwise wrong conclusions will be drawn from the samples used:

As a first consequence, the overly well-accepted assumption and usage of normal (Gaussian) distributed randomness (or otherwise "smooth" probability distributions like beta, exponential etc.) is not appropriate. In particular, normal distributions will often be used carelessly - justified by the central limit theorem (CLT) - but without providing real evidence for its applicability. In fact, even the CLT, making a statement about the infinite, does not have to hold in a particular situation, where only a limited number of random influences leads to a combined uncertainty, not following the limit distribution. Also, the underlying assumption of stochastic independence seldom is satisfied strictly. (Even if this condition can be weakened mathematically, it's not clear in practical applications whether it then holds or not.) Furthermore, normal distributions and similar well-studied functions do not account for phenomena like the black swan, since they are the ones that are not heavy-tailed, thus are not suited to predict rare but major events.

Instead of relying on the validity of "conventional" and "established" probability distributions the integration of historical data and expert knowledge to enrich the used distributions should be employed. In this way, more specific information can be used to design e.g. stochastic simulation models which account for the characteristics of the problem to be modelled.

A second consequence of the above considerations is the fact that plain Monte Carlo simulations could fail to provide meaningful scenarios. Instead of sampling from the given probability distribution itself and trying to spread the scenarios evenly amongst the probability space (as usually done in Monte Carlo simulations) more specific sampling techniques are available that are known under the term variance reduction. Examples amongst them are importance sampling, stratification, antithetic sampling or conditioning. For instance, importance sampling modifies the underlying probability distribution in order to get samples from an "interesting" region of the random variable (this could be a region with rare events). In parallel to sampling itself, the method allows to derive insights from the distribution itself, and thus can be used for uncertainty quantification and sensitivity analysis (see Section 5.4.1) [19].

Black swans are especially difficult to predict using plain Monte Carlo simulation for the reasons mentioned above. Techniques like importance sampling may be of help, once the regions of interest for sampling are known. But since black swans are more or less singular events, making these regions rather points, it could be impossible to identify them in advance and modifying probability distributions accordingly. A more rigorous approach to handle black swans could be to explicitly create (with the help of expert knowledge) "bad" scenarios in order to guide the sampling process to the interesting events. Samples generated this way help to show the optimisation process what are the scenarios to build robustness against - which is one of the basic ideas of optimisation under uncertainty (so-called hedging).

5.2 Features of the Optimisation Models

5.2.1 Stochastic RAMS/LCC

RAMS and LCC analyses will follow a stochastic approach. In a probabilistic approach to system reliability analysis, the primary objective is to obtain a failure distribution of the entire system based on the failure distributions of its components, which is the starting point (see Figure 18).

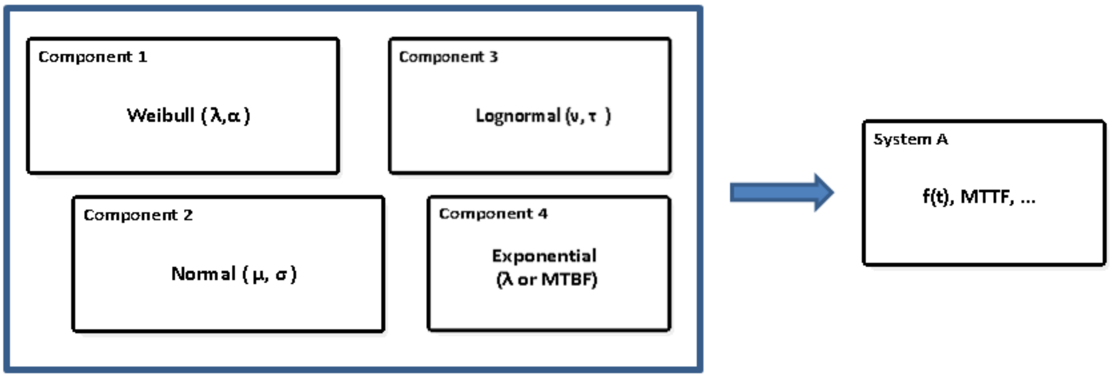

Figure 18. From Component to System Failure

Deliverable D5.1 "Report on RAMS data collection and failure rate analysis at component level" gives an overview of the RAMS parameters estimation at component level. The final goal will be to derive similar parameters but at system level.

For the sake of argument, in what follows we assume that the information at component level (the starting point) is already known. The availability of the (whole) system can be analysed by different approaches (reliability block diagrams, fault tree, Markov chain method, flow networks, Petri nets or Monte Carlo next event simulation, among others). They all depend on factors such as the complexity of repair strategies, adequacy of the "as good as new" approximation for repaired items, adequacy of the constant failure- and repair-rate approximation, etc. Among the most flexible ones, which can be used to analyse almost any type of system, is the ***Monte Carlo next event simulation***. Thanks to this flexibility, this method will be a good candidate to be used in our analyses of System Reliability.

Monte Carlo next event simulation is carried out by simulating typical lifetime scenarios for the system under study. The starting point is a model for the system in the form of a flow diagram or reliability block diagram. This diagram will basically contain the topology or structure of the system under study. Then, random events (i.e., events associated to item failures) are generated which together with scheduled events (e.g., preventive maintenance actions) and conditional events (i.e., initiated by the occurrence of other events) are included to create a scenario as close as possible to reality. Several types of input data have to be available:

- A model for the whole system in the form of flow or block diagram.
- Knowledge of: failure modes, failure effects, and consequences.
- Component failure and repair data (failure modes, downtimes distributions and estimates of the required parameters).
- Interventions and repair strategies as well as the durations for the various failure modes.

- Frequency and duration of inspections and planned maintenance actions.
- Resource data (e.g., availability of spare parts and maintenance resources).
- Throughput data and system/component capacities.

With this information we can simulate a "real experiment" from which performance measures can be derived. We can for instance calculate:

- The observed availability of the system in the simulated time period (e.g. observed uptime over the simulated time period).
- The number of system failures.
- The number of failures for each component.
- The contribution to system unavailability for each of the components.
- The use of maintenance resources.
- The system performance or throughput as a function of time.

The simulation can be repeated a number of times, therefore generating a big number of "independent" experiments from which we can infer rich probabilistic information of our system. These are for instance, estimates of performance measures of interest as mean values (MTBF, MTTF, MRL), confidence limits, higher order moments, p-values, statistical tests (ANOVA), underlying probability distribution functions (distributions of system failures).

Monte Carlo next event simulation method is appropriate for complex systems as it allows including many different factors and situations that can be considered and implemented in the simulations. The outputs of the simulation are probabilistic estimates for RAMS parameters at system level and as such are subject to uncertainty. These can be used the stochastic optimization (optimization under uncertainties) that will be carried out in the Smart decision support framework. Moreover, thanks to the simulation, we will also be able to learn what failure modes are more prone to happen. This information will be of great importance because it can be used to predict faults and condition evolving in specific asset components, as well as to improve the predicting capability of asset's alerts and the efficiency of the asset management system.

The main advantages of the simulation approach are among others:

- The possibility to easily derive information about the underlying probability distributions of RAMS parameters beyond mean values as outlined above, and to easily estimate the error in our models.
- They can accommodate highly complex scenarios involving a multitude of probabilistic events, such as corrective maintenance, preventive maintenance, inspections, imperfect repairs, crew response times, spare part availability, etc. When events such as these are considered, analytical (deterministic) solutions become impossible when dealing with real systems of sufficient complexity.
- The discrete event simulation also has the capabilities for: examining resource usage, efficiency and costs, optimizing procedures and resource allocation, analysing relationships between systems and components, maximizing throughput and minimizing work downtimes.

The stochastic Monte Carlo method outlined above feeds on two key elements: random number generation and event simulation. Although these are rather common practice methods in a large variety of fields, we will quickly review them for the sake of completeness.

Generation of random variables with arbitrary probability distribution

The inverse transform sampling algorithm: This method is based in the following simple observation: If T is a continuous random variable with (cumulative) distribution function F_T, strictly increasing for all t, then the random variable $Y = F_T(t)$ is uniquely determined and has uniform distribution $U(0,1)$. So by inversion, if Y is $U(0,1)$ and T has distribution function F_T, then the random variable $F_T^{-1}(Y)$ has the same distribution as T (see Figure 19).

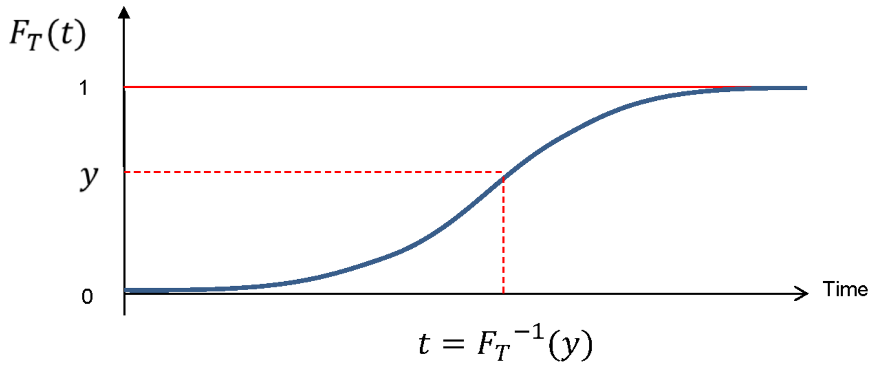

FIGURE 19. INVERSE SAMPLING

This result can be used to generate random deviates for an arbitrary distribution, starting from random variables uniformly distributed, and provided that the inverse function does exist. The algorithm that implements these steps is the following:

While *number of random variables has not been reached* **do:**
1. Generate a uniformly random deviate *u* ~ *U(0,1).*
2. Compute the value *t* such that $F_T(t)=u$ (by inversion).
3. Take *t* to be a random number drawn from the distribution described by F_T.
4. number of random variables <- number of random variables + 1

EndWhile

In some cases this method is not straightforward. In fact, there are many situations in which the inversion method is complicated to program or excessively inefficient to run. The inversion method is only really useful if the inverse distribution function is easy to compute (analytically or numerically). In the case where the inverse function can be calculated, the algorithm is very efficient because we only need a uniform to generate a random deviate of our desired random distribution.

Metropolis-Hasting algorithm: The Metropolis-Hasting algorithm is a powerful Markov chain Monte Carlo-based method that although less efficient than the previous algorithm, it is applicable in most cases and, in particular, very useful when the sampling distribution is very complicated. The algorithm can draw samples from any probability distribution provided we know it up to proportionality constant. The algorithm can be outlined as follows:

Suppose that our target distribution has density *f(x)*. Then given X_n, a proposed value Y_{n+1} is generated from some pre-specified "proposal" density $q(X_n/y)$, and then is accepted with probability

$$\alpha(x,y) = \begin{cases} \min\left\{\frac{\pi(y)q(y|x)}{\pi(x)q(x|y)}, 1\right\}, & \pi(x)q(x|y) > 0 \\ 1, & \pi(x)q(x|y) = 0. \end{cases}$$

If the proposed value is accepted, we set, $X_{n+1}= Y_{n+1}$; otherwise, we set $X_{n+1}= X_n$. The function $\alpha(x,y)$ is chosen to ensure that the Markov chain $X_0, X_1, \ldots$ is reversible with respect to the target density $f(y)$, so that the target density is stationary for the chain. When the proposal density is chosen to be symmetric, i.e., $q(x|y) = q(y|x)$ the algorithm is called simply **Metropolis algorithm**.

The right choice for the proposal density function $q(x|y)$ is crucial to achieve rapid convergence. Usually, what is chosen is a *symmetric random-walk* for the Metropolis algorithm, in which the proposal value is given by $Y_{n+1} = X_n + Z_{n+1}$, where the steps $\{Z_{n+1}\}$ are drawn from some symmetry distribution (e.g., $U(0,d), N(0,\sigma^2)$, ...).

Most of these algorithms are implemented in most modern statistics and computing libraries in diverse languages (R, Python, FORTRAN, C++, Java).

Simulation of new events

New events can be simulated using the Monte Carlo techniques outlined above. For instance, imagine we want to simulate a single repairable item with only one failure mode. A lifetime scenario for such item can be simulated as follows:

We start the simulation at $t = 0$ and the item is assumed to be functioning at $t = 0$.

The time t_1 to the first failure (TTF) is generated from the life distribution. This life distribution has to be specified by us based on previous knowledge and analyses in validation data sets on failure events at component level. Different distributions can be studied but usually a Weibull distribution is used for time to failure events.

The repair or restoration time d_1 is generated from a specified repair time distribution (usually a lognormal distribution), where the repair time distribution is also specified by us, and may for instance depend on the season (the date) or the time of the day of the failure among other factors. For example, the repair time may be longer for a failure that occurs during the night than the same failure occurring at ordinary working hours. The simulator clock is then set to $t_1 + d_1$.

The time t_2 to the second failure is generated from the life distribution. The item may not be "as good as new" after the repair action and the life distribution may therefore be different from the one used before. The simulator clock is set to $t_1 + d_1 + t_2$.

The repair or restoration time d_2 is generated from the repair time distribution.

The simulation continues until the simulator clock reaches a predetermined time (10 years for example) and we can create a log file to track all the events (failures, repairs) at each simulator clock time. From this file we are able to calculate the number of failures, the accumulated use of repair resources and utilities, the observed availability, and so on for a specific scenario. For instance, the observed availability A_1 is calculated as the accumulated time the item has been functioning divided by the length of the simulated period. The mean availability for an item can be estimated by averaging over samples once we have run the simulation a number *n* of times. We can also calculate the availability within periods of time by splitting the simulation into a number of time intervals.

The simulation can theoretically take into account any contingency variable, such as:

- Seasonal and daily variations.
- Variation on loadings.
- Periodic testing and interventions into the items.
- Planned shutdown periods.
- Interactions between components in the system.
- Dependencies between functioning times and downtimes, …

The number of input data needed for the simulation of life scenarios depends on the complexity of our system. Some decision rules must be established to account for the various events and combinations of events. These rules must state which actions should be a consequence of each event, as for example:

- Setting priorities between repair actions of simultaneous failures when we have limited repair resources.
- Switching policies between standby items.
- Deciding to replace or refurbish some additional component of the same subsystem when a component fails.
- Declining to shut down a whole subsystem after a failure of a component.

5.2.2 Stochastic Asset Condition

For predictive maintenance, or condition-based interventions planning in advance, the future asset condition will be estimated. As estimation implies, the condition is afflicted with uncertainties and underlies random impacts. So, the asset condition used in long-term and medium-term maintenance planning should be of stochastic nature.

Some literature describing stochastic models for asset condition development exists. Often, these models are used in strategic planning to deduce optimal maintenance strategies (in terms of LCC and RAMS). Sometimes, the models are derived from deterministic models (which are also common). Examples for stochastic models are as follows:

Stochastic functions: Deterministic functions are afflicted with some stochastic parameters, which are varied during a simulation process. E.g. in [20] a degradation model for track geometry is shown. The condition after the n-th tamping action is normal distributed with $NL_{init,n} \sim N(\mu_{NL_{init}}(n), \sigma_{NL_{init}}(n))$ and the deterioration after the n-th tamping follows $NL(t) = NL_{init,n} \cdot e^{b_n(t-t_n)}$ with $b_n \sim N(\mu_b(n), \sigma_b(n))$.

Failure rates: By means of failure rates, the expected time until failure and the expected failures per time can be calculated. Adversely, they give no information about the condition development. In [7], failure rates with exponential and Weibull distributions are used to define optimal maintenance strategies with Petri nets.

Stochastic process: The deterioration process is often described by a stochastic process (like the Gamma-process). With it, for each time interval the expected deterioration can be calculated from which the expected time until failure can be determined. In [21], the deterioration is given with the Gamma-Process $X_{t.t\geq 0}$ which has the following properties: $X_t$has independent increments, thus the

deterioration rate of each time interval is independent from the past deterioration rate, and for $s \geq 0$ and $t \geq 0$, $X_{t+s} - X_t$ has a Gamma distribution with the parameters αs and β (densitiy)

$$f_{X_{t+s}-X_t}(x) = \frac{1}{\Gamma(\alpha s)} \beta^{\alpha s} x^{\alpha s - 1} e^{-\beta x} I_{\{x \geq 0\}}$$

where $\alpha, \beta > 0$. The resulting Gamma process has a constant mean degradation rate $\frac{\alpha}{\beta}$ and a constant degradation variance rate $\frac{\alpha}{\beta^2}$. Figure 20 shows an example of this Gamma process.

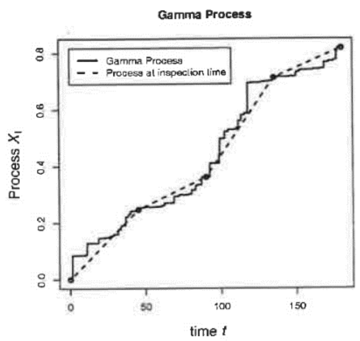

Figure 20. Gamma process as Used in [21]

Markov chains: In road and rail, deterioration depends on different parameters and can be divided in a set of condition states (like "no maintenance required", "small repair suggested", "speed restrictions required" …). The Markov chain gives the probability to change the condition state in a given time interval. (see, e.g. [22], [23] and [24]). In [25], a Markov chain to determine the optimal strategy for the use of ultrasonic inspection cars is presented. There are different failure modes and there is a certain probability to not detect a failure with the inspection car. Aim is, to determine the optimal inspection interval and the optimal waiting time between detection and repair. In Figure 21, the corresponding Markov chain is shown.

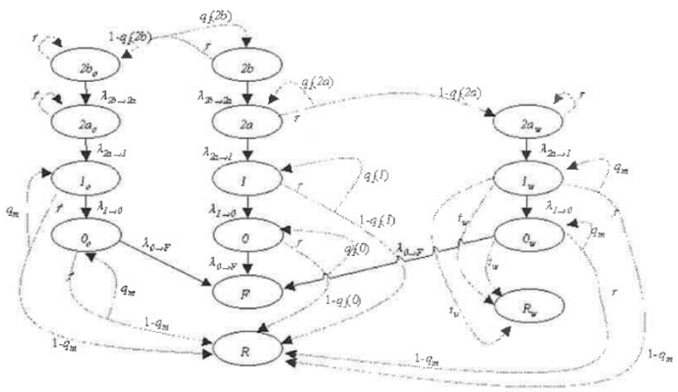

Figure 21. Marko Chain as Used in [25]

For modelling the alerts, Markov chains will be a good choice. The discrete condition states should be the exceeding of thresholds which are linked to required maintenance interventions (consists of maintenance tasks and associated to costs, working duration and resource requirements), restrictions in service and/or risks. At planning time, the probability distribution of the current states of the sections are evaluated. It isn't necessary to create an alert if the section stay in optimal condition until the end of the planning horizon. Otherwise, the alert will be afflicted with transition probabilities for each time. Transition probability is the probability to change from one state to another in a certain time-period. For example, this information can be derived from simulation.

In Figure 22 an example for a Markov chain for maintenance planning is shown. Based on expertise or regulations, four degradation levels are defined. The first one is “small deterioration” which can lead to a small preventive maintenance activity, for example if there is a good opportunity for maintenance. The second degradation level is called “visible deterioration” and should be resolved with the small preventive activity. The next degradation level is “great deterioration” and results in speed reduction and requires corrective maintenance. The last degradation level is break down, then then track is closed and corrective maintenance is needed. The transition probabilities $p(t)$ show that a direct transition from “small deterioration” to “great deterioration” is possible. That means, in some cases the degradation level “visible deterioration” does not occur.

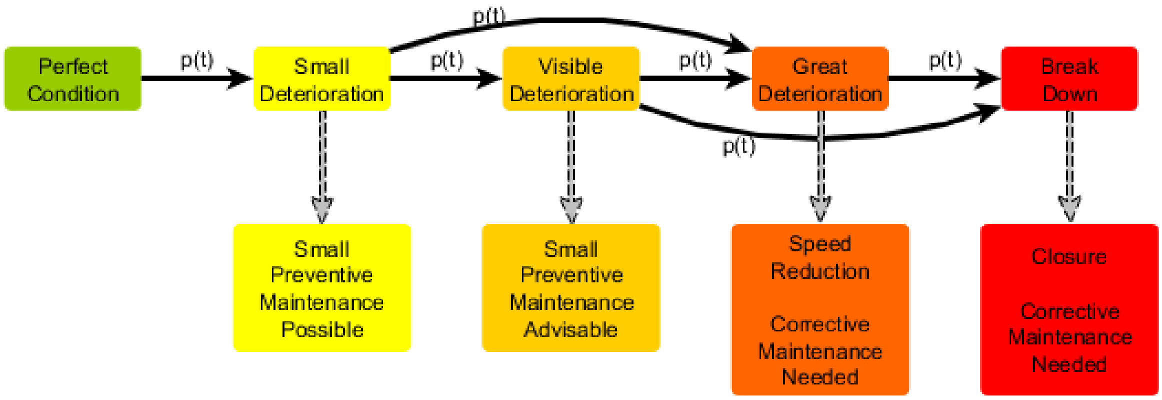

FIGURE 22. MARKOV CHAIN FOR MAINTENANCE PLANNING

The Markov chain is a sequence of stochastic variables X_t from a set of states $S = \{S_1 \dots S_n\}$ and with a (time-dependent) transition matrix $M(t) \in \mathbb{R}^{n \times n}$ whereby $m_{ij}(t) = P(X_t = S_j | X_{t-1} = S_i\}$ is the probability to be on time t in state S_j when the system was in time $t-1$ in state S_i. Additional, each state is associated with an intervention (see 5.2.3), thus equipped with a working duration, a vector of resources $r \in \mathbb{R}^{m_r}$ and a vector of costs $c \in \mathbb{R}^{m_c}$.

The advantages of Markov chains are:

- enfold all data for planning (degradation levels are associated with tasks, costs, resources, working duration, risks; transition probabilities allows prediction about future track condition)
- useful for statistical analysis
- adaptable to the most deterioration models (transition probabilities can be derived e.g. by simulation)
- adaptable to different parameters and thresholds by customized definition of degradation levels
- easy to handle, low storage effort

Besides the probabilities for condition development, also information about the earliest and latest useful maintenance time are helpful. For example, the latest allowed time can be the time when the probability for a critical condition is above a certain threshold (e.g. 1%).

Some condition developments cannot be predicted, which leads to unexpected events and unplanned maintenance. To have a plan that is able to react on changes and to integrate urgent and reactive maintenance. Such unexpected events will be simulated and the robustness of the plan against it will

be calculated. Therefore, a list of unexpected events with probability for occurring, urgency and needed maintenance activity has to be developed.

5.2.3 Stochastic Interventions

Maintenance planning on all levels is related to ***interventions***. Interventions are a set of maintenance activities or tasks associated with working duration, resource requirements (e.g. men power, machine hours) and costs (e.g. material costs, possession costs). They are estimations based on the maintenance activity and in reality only predictions as each maintenance job is a little bit different. They can also be modeled as stochastic variables.

The working duration enfolds the time needed for working on the track and is base for calculating the track possession costs. The time for traveling is not a part of the working duration, because of track possession must not book for traveling.

The resource requirement is an m_r-dimensional vector $r \in \mathbb{R}^{m_r}$ which defines the necessary units of resources to carry out the tasks. Required material is not necessarily a resource, because in the most cases it is not considered as being limited. However, in operational planning material can be limited because it has to be ordered in advance. All limited requirements like men power and machine hours are resources. By calculating resource requirement one part of the requirement results from traveling to the track segment, since as a worker or a machine travels to a track section, it cannot be used by another intervention. The second part of the resources is the requirement of the task, thus for the real working done on the track.

The costs often consist of three parts: working costs, travel costs and service costs. But it depends on the application and the operator. Therewith the costs should be modeled as m_c-dimensional vector $c \in \mathbb{R}^{m_c}$. Working costs enfolds the costs of the workers, machines and material, but not the costs for going to the track section or leaving the track section. The traveling costs are estimated from the distance between the maintenance station and the track section as well as the travel effort (thus the number of workers, machines and units of material). The service costs are costs resulting from restrictions in service, e.g. costs for track possession.

In maintenance planning nothing is really known in advance. Nearly every part of maintenance planning based on assumptions and estimations. Hence, the costs, resource requirements and time needed for the different maintenance interventions can also be formulated as stochastic values. Even the maintenance activity to be carried out can be of stochastic nature by itself, for example if the measurement data suggests a certain failure but the maintenance crew find another one that requires another maintenance tasks, or if preventive maintenance shows the need of corrective activities. It can be useful to integrate these uncertainty in the planning process by defining resource requirements, costs and time for maintenance or even the intervention itself as stochastic variables. E.g. by defining

$$\sum_{selected\ alerts\ a} r_a \leq R$$

with $r_a : \Omega \to \mathbb{R}$ is a stochastic variable for the resource requirement of alert a over the probability space $(\Omega, \mathcal{F}, P)$ with $\Omega = [r^1_{min}, r^1_{max}] \times \ldots \times [r^{m_r}_{min}, r^{m_r}_{max}] \in \mathbb{R}^{m_r}$, thus the resource requirement is a random vector in $\mathbb{R}^{m_r}$.

Another possibility is to define

$$\sum_{selected\ alerts\ a} r(I_a) \leq R$$

with $I_a : \Omega \to \mathbb{R}$ is a stochastic variable for the intervention linked to alert *a* and $r\left(I_a^k\right) \in \mathbb{R}^{m_r}$ are the deterministic resources required for a certain intervention I_a^k. Normally, $\Omega = \{I_a^1, \dots I_a^k\}$ is a (small) set of possible interventions I_a^k.

Also, the resources available to the operator change over time, crew members are in holyday or sick at home, machines go off. It is conceivable to model also the planning constraints as stochastic variables. E.g.

$$\sum_{selected\ alerts\ a} r_a \leq R$$

with $R : \Omega \to \mathbb{R}$ is a stochastic variable for the resources available over the probability space $(\Omega, \mathcal{F}, P)$ with $\Omega = \left[R_{min}^1, R_{max}^1\right] \times \dots \times \left[R_{min}^{m_r}, R_{max}^{m_r}\right] \in \mathbb{R}^{m_r}$, thus the resources available is a random vector in $\mathbb{R}^{m_r}$. Also, r_a can be stochastic variables.

5.2.4 Objectives

The main objective for the company is to ensure the availibilty of the system and the safety of the users by minimal effort.

The objective functions will base on internal KPIs, like deferred alerts, expected costs (user costs, maintenance costs, infrastructure costs …), risk evaluation, robustness measures, service impact; and depends on the focus of the operator and the kind of planning problem. If possible, also robustness measures will be integrated, for example the mean value, median, quantil or interquantil range of optimised KPIs. This statistic measure will be genrated by using our robust mantenance planning aproach.

Possible terms of the objective function are:

- $\min \mathbb{E}(\text{maintenance costs})$ … minimise expected costs
- $\min \inf(C \mid P(\text{maintenance costs} < C) > 0.9)$ … minimise the cost value that is not exceeded with a probability of 90%
- $\min \inf(R \mid P(\text{risk evaluation} < R) > 0.95)$ … minimise the risk value that is not exceeded with a probability of 95%
- $\max \mathbb{E}(\text{availibility})$ … maximise expected availibility

5.2.5 Risk-based Optimisation

By applying INFRALERT's concept, maintenance planning should generally lead to a reduction of the risk in its most basic meaning, i.e. as the (positive or negative) effects on the objectives as consequences of uncertainty. Thus, one feature of the concept explained in this subsection is to apply an approach to ***risk-based optimisation***.

Such an approach involves to implement suitable mechanisms for risk protection (or reduction), so-called hedging. Hedging in the first place opts for building robustness against uncertainty. But since robustness refers to the avoidance of negative impacts of uncertainty only, in a more optimistic setting risk-based optimisation could also be designed and applied in order to even profit from uncertainty.

Such an effect is called ***antifragility*** and can be interpreted as the strictly logical counterpart of fragility, i.e. the suffering from uncertainty, whereas robustness is the neutral state of independence from uncertainty [26]. Antifragility could be achieved by opportunistic planning, which allows for using additional actions ("opportunities"), but requires some degree of risk affinity from the decision-maker.

Three approaches are possible to integrate risk-based optimisation when actually designing concrete mathematical models for maintenance planning on strategic, tactical or operational level: implicit, explicit and semi-explicit:

Implicitly, the underlying mathematical concepts for optimisation under uncertainty described above contribute to risk-based optimisation, since they inherently use probabilistic information regarding (future) asset condition, cost, resources or any other relevant uncertain input parameter. The probabilistic modelling of constraints, variables and objective values as done in an "optimisation under uncertainty"-environment naturally leads to the consideration of risks (in the general meaning of deviation due to uncertainty). One reason is that these models avoid the use of point estimation for parameter values and thus produce solutions prepared to handle uncertainty.

An ***explicit*** approach is based on the definition of concrete risk indicators, e.g. robustness measures, and their integration into optimisation models, which as usually can be achieved by a mathematical formulation either as constraints or objective values. Appropriate measures for the evaluation of risk or robustness strongly depend on the planning problem itself, typical examples are as follows:

- "classical" risk evaluation related to undesirable events, e.g. failure and breakdown probabilities, availability and reliability of components and subsystems which can be derived from historical maintenance and interventions activities
- statistical indicators expressing the aggregated infrastructure performance, e.g. considering the overall capacity, punctuality of service: such a rather complex indicator is useful to summarise risk in a system (as consequences of uncertain behaviour) in an abstract way
- statistical values derived from the range and frequencies of optimised "regular" objective values like costs: addressing the robustness of solutions and the planning procedure itself, i.e. to estimate the influence of input uncertainty to the resulting output in the specific situation
- flexibility as a measure for the ability of solutions to recover from "bad" events that cause adaptions: this is related to risk in a rather indirect way, as it assumes that inflexible solutions, that are not easy to adapt to new information, carry a higher risk of being prone to uncertainty (fragile)

Through the robust maintenance planning as solution approach (described in 5.3) it is possible to control such risk indicators, and thus to explicitly search for solutions in a risk-based manner.

A ***semi-explicit*** approach is done via the specific use of results and output of risk assessment in different planning levels. As described in Section 4.2.2, quantitative risk assessment methods provide information about probabilities and consequences of identified risks, normally by evaluating scenarios, by using historical information or by merging opinions from different experts. In doing so it is again possible to derive probability distribution functions for failures and effects instead of simple point estimates. These distributions provide a more valuable base for further processing in maintenance planning. For instance, maintenance alerts generated for tactical planning by analysing actual and predicted conditions will get prioritised using results from risk assessment. These priorities reflect the risks associated with alerts: minor alerts will have a low risk when getting deferred in the tactical maintenance schedule, but major alerts shall be resolved in the near future. In this way, results from risk assessment indirectly influence decision-making in maintenance planning, leading to a semi-

explicit approach. Other examples could be given for strategic planning, where results from risk analysis will be used to evaluate maintenance strategies in terms of their long-term behaviour and affect to the adjustment of strategy parameters.

An important factor in risk-based optimisation is the ***risk attitude*** of decision-maker is important, i.e. whether a risk aversion or risk affinity predominates the planning decisions. Of course, this is not solely dependent on individual attitude, but also on requirements of the considered situation, e.g. in a case with regulatory restrictions due to safety it is forbidden to take any risk at all. However, risk attitude has an influence on the problem modelling and should be included when designing risk-based optimisation models, for instance in form of adjustable parameters whereever constraints or objectives related to risk have been introduced.

5.3 Solution Approach

5.3.1 Robust Maintenance Planning

As shown, the produced maintenance plans should be robust against uncertainties to reduce risks. Robustness means that the plan is stable in the unknown future, thus can handle uncertain developments and unexpected changes. To prove the robustness of a plan, different possible future scenarios are generated and played through. Thereby, small adaptions of the plan are allowed because in reality small adjustments are also done. For this purpose, the so-called ***Monte Carlo Rollout*** approach can be used:

Monte Carlo Rollout combines ideas from Rollout algorithms for combinatorial optimisation [27] and the Monte Carlo Tree Search from game theory [28]. Basic elements of the Monte Carlo Rollout method are: a simple heuristic $\mathcal{H}$ that is capable to generate "good" solutions to the given problem based on current information, and a stochastic model to simulate the uncertain future. Both are combined to take a look into the future and to estimate their effects on current decisions.

The Monte Carlo Rollout method as shown in Figure 23 works as follows: Initially, a set of different alternative solutions is generated (the yellow dots). Each of these alternatives is proven and evaluated by a number of Monte Carlo rollouts. In each rollout another future scenario is “played” in terms of a two-player game. Thereby the stochastic model is used to simulate random events (moves of the “random player”, small orange dots), and the changed situation is solved using the base heuristic $\mathcal{H}$ (moves of the “decision maker“, green dots). The two players move alternative until the end of the game or a predefined number of steps (the “depth”) is reached. The outcome of each scenario is evaluated (see the small blue triangles), and the solution quality of the alternative is determined, e.g. by averaging scenario evaluations (large blue triangle). After all the best alternative is chosen, being a high-quality solution additionally equipped with high robustness.

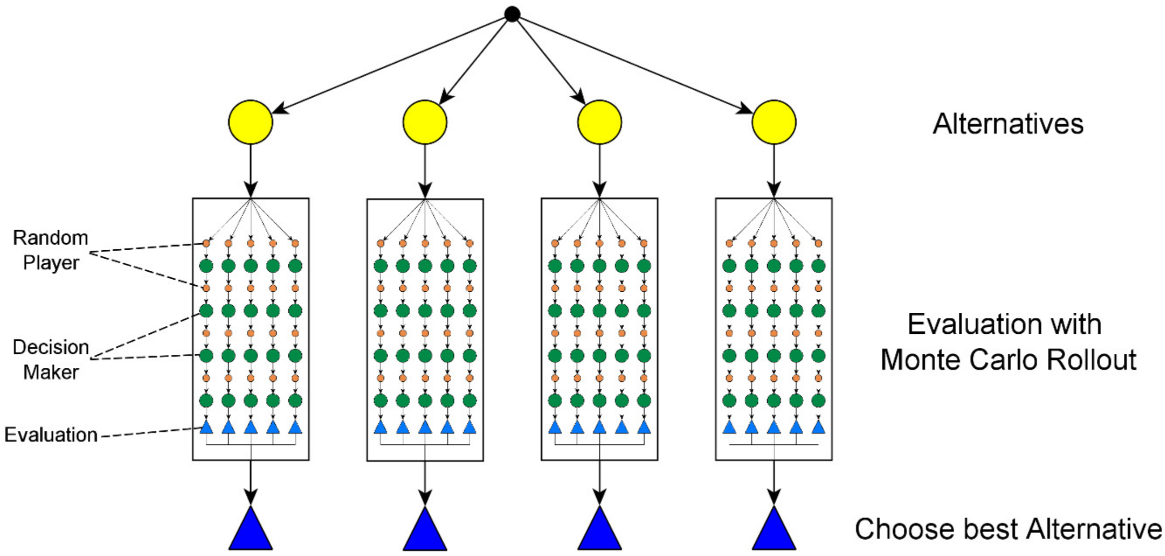

FIGURE 23. THE MONTE CARLO ROLLOUT METHOD

In tactical planning, alerts are afflicted with stochastic information (Markov chain) to represent the uncertain track condition and the uncertain need for maintenance. Furthermore, there is a stochastic model representing the appearance of unpredicted reactive maintenance activities because this activities are not known in advance and has to be executed promptly. In the simulation process different scenarios are generated. In each scenario, a certain outcome of each random variable is assumed. Based on the resulting deterministic maintenance activities, the plan is adjusted by the base heuristic and different robustness measures are calculated as scenario evaluation, e.g. number of critical alerts or number of deferred activities. The entirety of all the scenarios shows the robustness of the maintenance plan.

In dynamic planning, the planning horizon is short and the alerts are nearly deterministic. There are only small uncertainties regarding working time, material requirements; but there are also uncertainties regarding operational constraints like staff changes or available resources. To prove the robustness of the short-term dynamic plan against short-term changes in the operational constraints, the impact of such changes on the plan are analysed and by means of a base heuristic possible adjustments of the plan are shown, e.g. the deference of an uncritical preventive activity.

Also in strategic planning robustness is important. The information about failure and deterioration rates, effects of maintenance etc. is not exact, rather it is an estimation. Because of that, the sensitivity of the maintenance policies against variations in the input information will be analysed. Furthermore, the used RAMS and LCC parameters will be of stochastic nature to indicate the uncertain input. With this it could be analysed how changes in the models and stochastic parameters affect maintenance planning. For instance, failure models can be varied in order to determine the influence on RAMS and LCC. Maintenance strategies that are robust against changes in the failure model can be found and should be preferred in maintenance planning. Sometimes, it can be appropriate to use a safe preventive strategy instead of a cost effective but doubtful predictive approach.

5.3.2 Nested and Adaptive Planning

The developed planning concept follows a nested and adaptive approach: dynamic, tactical and strategic planning levels will get linked together, providing feedback to each other in terms of input and results.

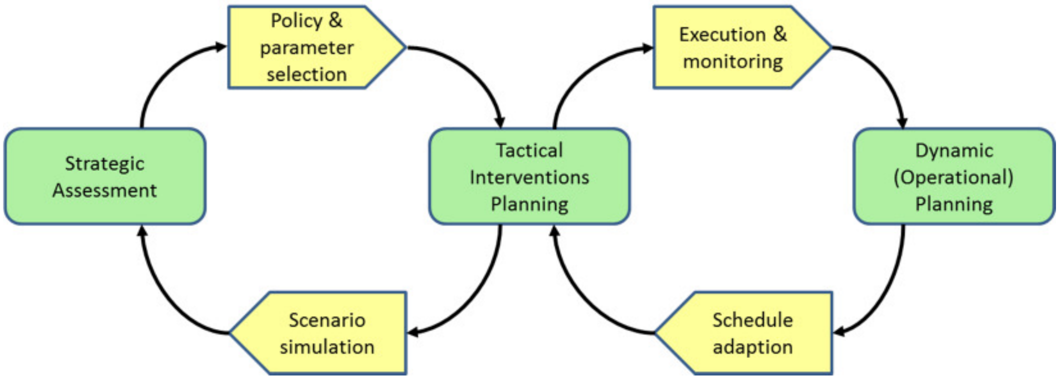

Figure 24. Nested and Adaptive Approach to Maintenance Planning

At operational level, robust maintenance schedules which are forwarded from tactical planning have to be adapted dynamically, thus preserving flexibility. Tactical planning then re-allocates activities due to new information gathered on the operational level, thereby providing new inputs. On the strategic level, maintenance schedules determined by tactical planning will be applied in long-term scenario simulation where strategic new construction activities are also evaluated. As a result, parameters of maintenance policies will be derived that influence tactical planning. In this way, a nested planning loop is being formed, see Figure 24.

The adaptiveness of the concept follows from the fact that planning at each level is done in a rolling time horizon: Decisions derived from a solution in a certain planning step (at a certain point in time) will be input in the next step, and could be revised and adapted if possible as a reaction to new information that becomes available. Indeed, this is one of the main features of the underlying mathematical concepts for optimisation under uncertainty, as described in Section 0.

5.4 Uncertainty Management

The planning concept is based on the use of probabilistic information, which is induced by the numerous uncertainties involved in maintenance management. As described in the above sections, mathematical concepts for optimisation under uncertainty are applied. Such optimisation models and solution procedures use probabilistic information which describes the uncertainty in several ways - as random variables and parameters, stochastic restrictions, objectives. And therefore they rely on the validity and dependability of this information used: starting with the correct failure rates, the deterioration model and the risk assessment, and ending with realistic cost values, resource utilisation and time needed for scheduled activities. All this information has a high influence to the maintenance planning approach and to its results. For example, if the deterioration process of an asset is underestimated due to unreliable data, then maintenance activities will not be planned appropriately in time. This could lead to disturbances in the maintenance plans once the "real" deterioration development becomes visible, and are associated to risks for the maintenance and infrastructure operator.

From this follows that it is reasonable to carefully think about the handling and use of probabilistic information. In this sense, an ***uncertainty management*** is introduced to the planning concept, which affects the overall process "how to deal with information about uncertainty". Uncertainty management is not about the uncertain planning methods itself - which is handled via "optimisation under uncertainty"-techniques and the nested approach - but separately from the planning process applies "offline" tools to analyse the process and to control the information retrieval.

In this context, four main questions arise:

- How to measure uncertainty in input variables?
- How does uncertainty propagate from input to output?
- What is the value of probabilistic information when used for planning?
- How to control the continous validity and how to update information (if necessary)?

Each of these questions will be examined in one of the following subsections 5.4.2 to 5.4.5, together with their particular meaning in maintenance planning.

The issue of uncertainty management is related to what can be named the "subjectivity of randomness": Since it is clear that probabilistic information on uncertainty is never purely objective, but rather occurs as a mix from information-based (e.g. historical data, physical knowledge) and expert-based data retrieval, it always contains a subjective component. The decision-makers "degree of belief" or "degree of knowledge" in probabilistic information determines the importance of uncertainty management.

We start this section with a short introduction to uncertainty analysis, a review of basic terms, tasks and techniques in this area.

5.4.1 Uncertainty Quantification and Sensitivity Analysis

For decision-making under uncertainty it is necessary to analyse the nature and degree of uncertainty in the parameters and variables used, since otherwise it is not possible to evaluate the validity of observations and models that represent the knowledge base to handle these problems. Uncertainty analysis is the sub-category of mathematical statistics and modelling that provides the models and techniques to (statistically) quantify uncertainties in the relevant parameters and to determine the mutual influences and dependencies.

The standard object of investigation in uncertainty quantification is as follows: Some input, which is afflicted with uncertainty, is processed by a mathematical or computational model, function or system, which is typically an abstraction of a real-world process and thus incomplete, returning an output. The uncertainty in input is being propagated through the system to the output, due to system's uncertainty it can even be amplified.

Besides the general distinction between uncertainty inherent to input parameters and the uncertainty introduced by incomplete models, it is usual to categorise uncertainty (in input or output variables) as follows:

- ***Aleatoric*** or ***statistical uncertainty***: uncertain vairables whose exact values ***cannot be*** measured or determined with certainty because there exists some stochastic variability which causes the values to differ from one realisation to the other; the so-called "unknown unknown"
- ***Epistemic*** or ***systematic uncertainty***: uncertain variables that ***could be*** determined exactly but will not because the system's or measurement's incompleteness, because certain effects will be neglected by the model or data is ignored; the so-called "known unknown"

We also remember the distinction between ***stochastic*** and ***dynamic*** uncertainty as made in Section 0.

Uncertainty analysis is basically concerned with the following three types of tasks:

Forward propagation aims to quantify the combined effect of the input and system uncertainty on the output variation, thus how uncertainty is propagated from input to output. The quantification can be expressed in numerous ways: Let the relationship between output variables y and input x be formally described by $y = f(x)$. The absolute error Δf as a function of the error Δx could be of interest, together with the relative error $\Delta f / f$. Most commonly the propagation is defined in terms of the output's standard deviation σ_f or variance σ_f^2, also as a function of the input's standard deviation σ_x (seen as vector of the standard deviation σ_i of the single variables x_i). In case of correlated variables x_i and x_j, their covariances ρ_{ij} have to be taken into account. As a simple example, for the linear combination

$$y = f(x_1, x_2) = ax_1 + bx_2$$

the propagated variance as an expression for the forward uncertainty can be calculated

$$\sigma_f^2 = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho_{12}.$$

Of course, for complex systems and relationships f the forward propagation normally cannot be derived analytically. Commonly used methodologies reach from probabilistic approaches like simulation-based methods, local and functional expansion, most probable point to non-probabilistic approaches like interval analysis and fuzzy theory.

Inverse quantification tries to deduce the other way around: Given varying output as a result of a computational system using some input parameters, the question here is to estimate and quantify the influence of the system as well as the input. Mathematically spoken, this involves two sub-problems: The ***bias correction*** quantifies the uncertainty coming from the system, as reflected in the deviation between the observed (or measured, experimental) output f_e and the "modelled" output according to the system f_m. The task is to find an estimation for the so-called discrepancy or bias function $\delta(x)$ in the formula

$$f_e(x) = f_m(x) + \delta(x) + \epsilon$$

where ϵ denotes the experimental uncertainty. The second sub-problem is the ***parameter calibration*** and aims to estimate the uncertain or unknown input parameter θ from the modelled system f_m, find a best-fitting θ^* (or a probability distribution for it) that solves

$$f_e(x) = f_m(x, \theta^*) + \epsilon.$$

It is clear that in most practical situation both sub-problems occur in combination, thus the general task in inverse quantification is to find $\delta(x)$ and θ^* in

$$f_e(x) = f_m(x, \theta^*) + \delta(x) + \epsilon.$$

The most basic approach to tackle such problems is related to regression analysis. Some more specific solution approaches for inverse quantification with bias and parameter calibration exists under the framework of Bayesian statistics.

Sensitivity analysis, although normally considered a field of research of its own, in some sense can be seen as a type of uncertainty quantification, too: Sensitivity analysis makes use of re-calculated outputs under varying, alternative inputs in order to "simulate" and then to quantify the influences of uncertainty. By doing so in a selective way, it tries to understand how much each input variable contributes to the output uncertainty. As can already be seen from this brief description, there is an overlap with forward propagation, which also manifests in the methodologies and techniques used. Views in literature are not consistent in this respect, sometimes, sensitivity analysis is seen as part of uncertainty quantification, sometimes as a subsequent step.

In the following four subsections we describe the tasks related to uncertainty analysis in the context of INFRALERT's planning concept and how they contribute to uncertainty management.

5.4.2 INPUT UNCERTAINTY

A natural first step in uncertainty management - after identifying sources of uncertain input - is to somehow express the amount or level of uncertainty in terms of numerical values. For this, several measures have to be defined, based on the information available to quantify a specific uncertainty.

Stochastic uncertainty in input parameters can appear in form of intervals or ranges where values are varying in, which have to be determined. Then, the length of the interval can be used as a measure for the variability, maybe as an absolute or a relative value. If reasonable, a measure can be derived from the observed frequencies of the values' distribution within the intervals, e.g. as weighted sum representing any statistical parameter like moments or quantils. In the best case, assumptions on the underlying probability distribution with its various parameters are available that serve as uncertainty measure.

For input which is subject to dynamic uncertainty a variety of measures for the "degree of dynamics" is imaginable, depending on the concrete situation. For instance, in the literature related to the field of online optimisation such measures are used to evaluate the strength of an adversary. These definitions are orientated on the rate of occurrences of events or new information, the "size" of events, "amount" of information or something similar expressing the influence of the dynamically occurring information on the current situation.

After defining reasonable measures for input uncertainty a regular assessment of the obtained values has to be made. The common understanding here is that high input uncertainty has a somehow "bad" influence on the maintenance planning process, since it causes more disturbances or the effects of disturbances are more drastic. Or when seen from the other side: In order to be able to cope with high input uncertainty the planning process has to be more robust or flexible.

Of course, input uncertainty cannot be assessed spearately, but has to be considered in the context of the process, i.e. in relation to its influence on output or its handling in robust planning algorithms. These tasks of uncertainty management will be discussed in the next sections (5.4.3 and 5.4.4). However, even input measures itself are useful for several purposes, for example: to identify the relevant sources of input uncertainty, to compare and classify uncertain input, to find dependencies amongst uncertainties by using combined measures. To make use of this, the measures have to be monitored and updated regularly in the framework (see below 5.4.5), so as to see changes and possible trends in the evolution of input.

5.4.3 Output uncertainty

The reduction of output uncertainty or - from an optimistic perspective - their proper "exploitation" is the main issue of optimisation under uncertainty, which is the base of the planning concept. To this end, output uncertainty has to be measured numerically in order to perceive the effects of planning. This task is related to the approaches to quantify the uncertainty propagation from input to output parameters as described above, see 5.4.1. But in the specific context of the planning concept it is not applied in the usual way of the standard approach, where output uncertainty develops through the propagation scheme:

uncertain input → incompete model → uncertain output

Rather, it has to be taken into account that the algorithms used in the planning process act as a kind of "corrective" or "compensator" buffering the uncertainty, i.e. the scheme now is:

uncertain input → incompete model → corrective algorithm → uncertain output

Thus, measuring output uncertainty in our context means measuring the ability of the planning process to deal with uncertainty, which allows to recover from changes or disturbances via adaptions and corrections of the solution produced. Again, input uncertainty together with incompleteness of the model leads to an output uncertainty, but the planning concept is designed to be able to react and to reduce these effects. For the definition of proper measures two different directions can be followed, where the assessment is focussed on: robustness and flexibility.

When measuring ***robustness*** one implicitly assumes that the objective of the planning algorithms is to produce solutions that are stable under varying input, and thus provide a similar quality of output, even if the situation changes. Here, the classical, direct way of measuring output uncertainty by evaluating the results in terms of the objective function is applied, e.g. by calculating statistical parameters from their variations. Such a robustness measure clearly involves an assessment of the algorithm's recovering ability mentioned above, since this will be reflected in the final results: The assumption is that "poor" algorithms behave "poorly" under uncertainty, thus deliver "un-robust" values.

Flexibilty measures provide a rather indirect way of assessing output uncertainty. The idea behind this approach is that solutions with high flexibility are easily adaptable to new situations without loosing their quality (with respect to the objective function). This is particularly true for planning environments where there are costs associated to the adaption of solutions, e.g. by introducing recourse functions into the objective as done in models for multi-stage stochastic programming (see 5.1.3). Flexibility looks at the structure of the solutions produced, and evaluates certain features of it which e.g. allow for switching decisions or avoid early "big decisions". For the definition of flexibility measures it is helpful to imagine the planning process as a decision tree, and to consider the possible recovery paths in the course of planning.

It depends on the process and the application behind which kind of measure is more relevant for assessing output uncertainty - robustness or flexibilty. However, both approaches can be defined and used independently. It is worth to mention that the calculation of the measures itself is not a simple and straight-forward exercise, since their definitions can be rather complicated and could involve in-depth analyses of the planning process. Thus, it is a challenging task also to set up algorithmic procedures that calculate measures from real data.

For the purpose of uncertainty management, the measuring of output uncertainty offers the possibility to analyse its dependencies from different influencing factors:

- From quantified input uncertainty: The understanding of dependencies between input and output uncertainties helps to identify most relevant sources and to select the ones to focus on in the planning concept.
- From structural features of the problem: Each planning problem shows its own characteristics, which are reflected in certain structural features to be formalised accordingly (e.g. interlacing of decisions, existence of traps, dominance, mobility restrictions etc.), and which then are measurable in concrete problem instances. The knowledge of such features and their influence on output uncertainty helps to select the proper solution approaches.
- From algorithm features: As mentioned above, also the algorithm itself used for handling uncertainties in planning process determines the level of output uncertainty. Again, insights gained from this analysis helps in the algorithm design phase, e.g. to create algorithms that introduce robustness respectively flexibility in planning.

5.4.4 Value of Probabilistic Information

In this section, the application of measures for input and output uncertainty together with their dependency analyses is addressed. The aim is to briefly sketch how the considerations from the above sections 5.4.2 and 5.4.3 can be exploited to evaluate the use of probabilistic information in the planning concept.

The issue of uncertainty management considered here is to provide recommendations or guidance on how to use probabilistic information. Typical questions that arise in this context are: Which probabilistic information to use? How to process them algorithmically?

The first question appears in practical situations where higher quality or more accurate information could be available for maintenance planning due to refined models for prediction, diagnosis, risk assessment, LCC analysis etc. Evenly, it could be the case that models could produce better results when using more detailed and complex data, which are available thanks to evolved monitoring, measuring or data capture technologies. The information then is still of probabilistic nature, but could provide more insight into the underlying process. Of course, better information comes to the price of higher effort, which normally calls for the application of cost-benefit analysis (CBA). In this case, an CBA could support decision-makers in selecting technologies and models to introduce enhanced quality of probabilistic information.

The gathered probabilistic information has to be processed accordingly by "intelligent" algorithms which are specifically designed for this purpose. The underlying hypothesis is, that the use of probabilistic information helps to deal with uncertainty, for instance by reducing their negative effects or even by exploiting arising opportunities. Optimisation algorithms for robust planning cause an additional effort, both in implementation as well as in computation. In particular, there are many algorithmic approaches where the computational effort depends on how much probabilistic information is used. Typical examples arise when scenario sampling techniques are applied, and algorithm parameters like lookahead depth and width (sample size) have to be configured. Again, a natural question to be considered can be raised: What is the benefit (in terms of solution quality) of enhanced costs (in terms of computational effort, e.g. induced by certain algorithm parameters)? Former research on the Monte Carlo Rollout approach provided interesting results regarding the dependencies between lookahead depth and solution quality, proving the existence of somehow surprising effects called the "lookahead pathology" [29]. This clearly is an evidence for the importance of examining the value of probabilistic information in a given situation, since it is not always clear that "more information" automatically means "better results". Rather, CBA-like investigations can be made

in order to do a (relative) comparison between different algorithm designs or configurations. By defining a "stupid" algorithm which uses no probabilistic information at all and making a comparison against this one it is also possible to determine an absolute value, showing the true effect of using probabilistic information.

Besides these algorithm-related analyses it is also possible to identify maintenance planning problems where probabilistic information provides added value compared to others, where subsequently the focus in algorithm development and information retrieval could be set on.

5.4.5 Monitoring and Updating

Algorithms to cope with uncertainty rely on the validity and applicability of probabilistic information, e.g. on accuracy of predictions, assumptions on stochastic models, relevance of historical data etc. Such assumptions could be wrong from the beginning, but they could also shift in the course of time because of some trends not considered duly. This problem in principle applies to all kind of model parameters, but is of particular interest for uncertain information, since the "degree of belief" is much lower due to its subjectivity. Therefore, a sensible uncertainty management has to address this issue accordingly by implementing a mechanism which has to monitor the validity of information related to uncertainty and to update if necessary. Calculated maintenance plans and predictions where they are based on should be compared with the real development. Then, deviations which originate from invalid information can be recognised, critical assumptions can be reviewed, thresholds can be adapted, inspection intervals can be adjusted and so on.

This monitoring and updating mechanism is not restricted to information related to a single planning problem: Through the nested and adaptive approach all information is connected, decisions made at one planning level or in one step based on wrong assumptions will be corrected by algorithms, either at the next level or in the next step. If this correction happens more often or more extreme this clearly is a hint for uncertainty management that some information or models inappropriate.

To illustrate the possibilities of this mechanism we give some examples from different building blocks in the planning concept:

- In failure and deterioration analysis, it should be checked whether the proposed failures rates are conform with the real failure; whether the assumed deterioration processes fits to the real deterioration processes of the assets; whether the improvement through maintenance is as supposed.
- In strategic planning, it should be checked whether the policies lead to the estimated maintenance effort; whether the RAMS targets are met in real and maybe with which additional corrective maintenance effort; whether the expected LCC correspond to the actual costs.
- In tactical planning, it should be checked whether the plan reaches the required robustnuss, thus whether dynamic planning leads to large changes of the tactical plan or not; whether the operational constraints are realistic.
- In dynamic planning the maintenance schedule and the executed maintenance should be checked to compare assumed resource and time effort with the real efforts

6 Framework for Smart Operation and Maintenance Decision Support

6.1 Smart O&M Decision Support in the eIMS

The smart decision support tools to be developed will be a part of the automated data processing chain of the INFRALERT eIMS. Since maintenance and intervention planning is the end point of this chain there is high demand for interactivity with the user of the system. To assure a high acceptance and usability of the planning tools, the following ***framework*** has been designed to integrate ***smart decision support*** with existing procedures and ambient building blocks discussed in Section 4. The resulting framework will be general enough to be easily adapted and applied to a wide range of maintenance planning scenarios. It provides the basis for the development of specific optimisation models following the condition- and risk-based planning concept specified in Section 5.

6.1.1 Overall Framework

In compliance with existing practices and standards, maintenance planning is separated in three levels: strategic, tactical and dynamic planning. To be able to integrate existing tools, methods and approaches, the framework has a modular structure.

In Figure 25, the overall maintenance management framework is shown. Boxes represent tools or tasks and clouds represent data and information. The colours stand for different responsibilities: blue for data storage and tasks done by the operator; green for asset data analysis and evaluation; and orange for decision support tools.

The maintenance management framework starts with the asset data base (1). Based on these information, failure analysis (2) will be undertaken. This enfolds the detection of failure modes, root-cause analysis, risk assessment, analysing the deterioration process of components and so on. The outcome of these activities are failure models, risks, models describing the deterioration process and models describing the restoration process of maintenance. This data is used for strategic planning (3) and for nowcasting and forecasting (4).

In strategic planning (3), the assets are grouped; for example regarding asset type, geometric characterisations, or traffic volume. For each asset group, the best strategy or the best mix of strategies is determined. Input of strategic planning are failure rates, deterioration model and maintenance models. The selected strategies have to meet RAMS-targets or given KPIs and have to minimise LCC. Therefore, strategic planning is connected to a RAMS/LCC-tool. Operational constraints, like annual budget, existing machines and men power available, should also be considered to ensure that the strategies could be realised. Output of strategic planning is a collection of strategies for the asset groups.

For tactical and dynamic planning, the current and future track condition has to be determined by nowcasting and forecasting (4). This is done based on measurement data and the deterioration and failure models. In the alert & interventions management tools (5) is determined, which maintenance activities are necessary in the medium-term and in the short-term. Thereby, for each asset is proven whether the strategy requires maintenance. This means for preventive maintained assets that time period will be exceed or the usage trigger reaches the threshold; for predictive or condition-based maintained assets that the condition reaches now or in future the threshold value; and for reactive maintenance activities that a defect was detected. The list of alerts can be reduced by combining asset

with homogeneous interventions in order to carry out the intervention in a longer distance. The interventions are accumulated with e.g. level of severity, condition probabilities, a time interval for execution, costs and resource requirement. They are planned on two levels, in a medium-term time horizon in tactical planning and in short-term in dynamic planning.

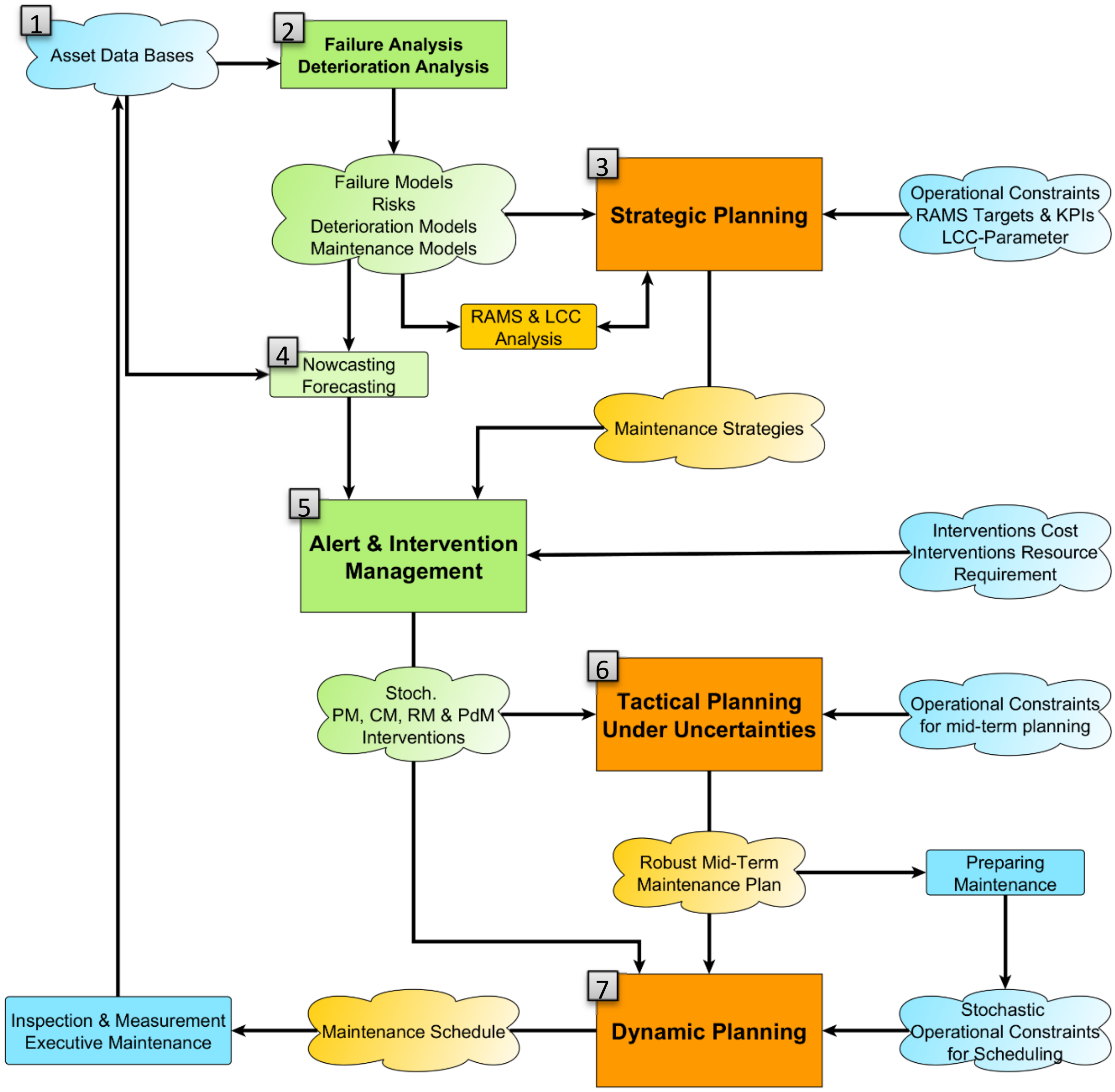

FIGURE 25. THE OVERALL MAINTENANCE MANAGEMENT FRAMEWORK

In tactical planning (6), the alerts or interventions will be selected, combined and allocated to time intervals. Based on the resulting coarse tactical plan, the operator can order material, book machines or track possession and plan holydays. The tactical plan will also be the input of dynamic planning (7). There, the selected and allocated maintenance activities, as the case afflicted with new condition information, and newly added and urgent activities will be scheduled in detail. Thereby, the changeable operational constraint like men power, machines available, material available, etc. will be considered.

Eventually, maintenance is executed according to the schedule, assets are inspected and new measurement data is generated. This results in new information in the asset data base and with it new information in the whole planning framework.

6.1.2 Strategic Level

In Figure 26, the part of strategic planning is shown more detailed as in the general framework. Input for strategic planning are on the one hand the failure models and maintenance models including the description of the deterioration process, failure rates and an estimation of the condition improvement caused by the different possible interventions. On the other hand, a set of possible strategies for the different assets has to be formulated. This set depends amongst others on the kind of asset, installed monitoring systems, available machines, existing knowledge by the maintenance crews and reliable deterioration models at hand. Furthermore, in a lot of applications an annual budget and/or resource capacities have to be considered in planning.

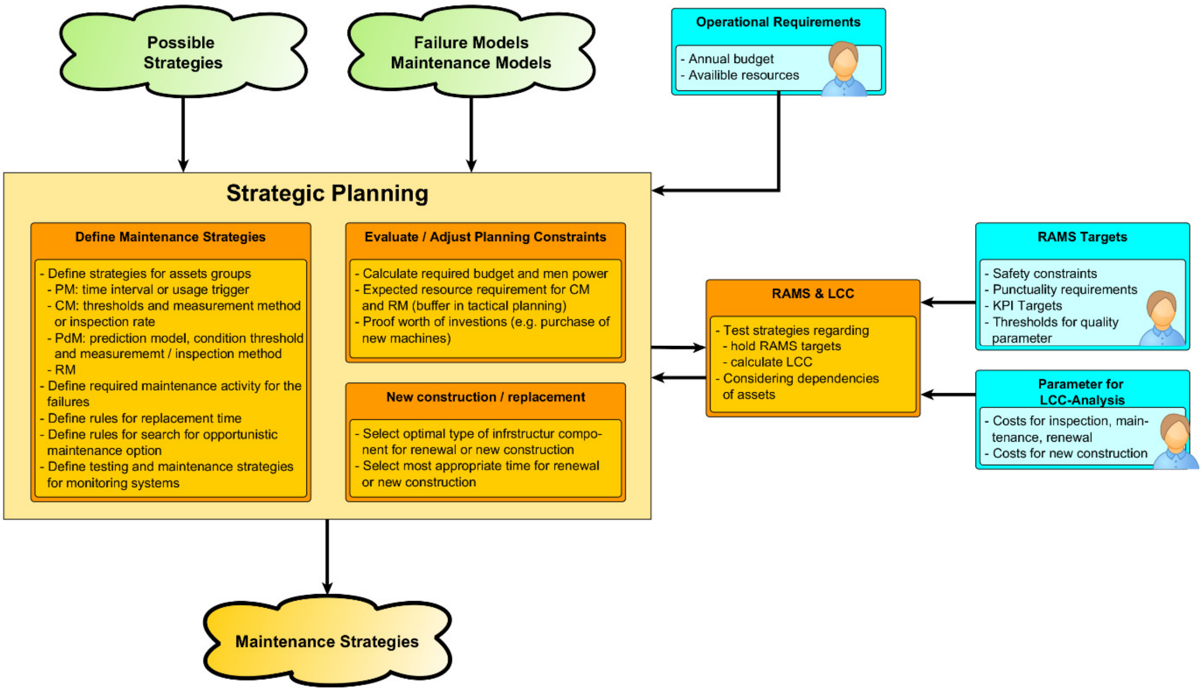

Figure 26. Framework for Strategic Planning

Aim of strategic planning is the definition of maintenance policies in order to support tactical and dynamic planning as well as to ensure cost effective maintenance by reconcilable risks. To reach this aim, a RAMS&LCC tool will be used to evaluate the life cycle costs, the risks and the system functionality in terms of reliability and availability. From all the policies that hold the RAMS targets defined by the operator, the strategy with the smallest LCC should be chosen.

Another parts of strategic planning is to evaluate or adjust of planning constraints. It could be proven, which budget and resources are necessary to meet the aimed RAMS targets. It could be estimated, which resources are required by condition-based and reactive maintenance to define suitable buffers in tactical planning. Also the worth of interventions, e.g. the purchase of a new machine or installation of further monitoring systems, can be rated.

Beside this, strategic planning can also support decisions in new construction and replacement: for different kinds of asset, the maintenance strategies can be analysed and the asset type whit the best maintainability can be identified. Also the time, when replacement is more cost effective since further maintenance can be determined.

The main output of strategic planning is a set of maintenance policies. In general, a maintenance policy may be defined as a decision rule which establishes the sequence of maintenance actions to be undertaken. That means, for each kind of asset the strategy, thus preventive, predictive, condition-based or reactive maintenance, is defined and the corresponding trigger values are fixed.

6.1.3 Tactical and Dynamic Level

In Figure 27, the part of tactical and dynamic planning is shown more detailed. Both planning stages base on so called alerts and interventions. In the alert & interventions management tools is determined, which maintenance activities are necessary in the medium-term and in the short-term. Thereby, for each asset is proven whether the strategy requires maintenance. That means:

- For assets with preventive maintenance the time since the last intervention has to be compared to the time trigger or the usage trigger has to be compared with the expected usage in the next time. When the trigger will be reached during the planning period, a corresponding intervention is created. This can be afflicted with an execution interval, thus an earliest and a latest feasible execution time, some combinations and a list of activities with associated costs and resources.
- For assets with predictive alerts the appropriate condition parameters or quality measures has to be forecasted in order to predict further condition / quality. When the trigger can be reached during the planning period, a corresponding intervention is created. This can be afflicted with some possibilities for combination, a stochastic model for deterioration or the time of exceeding the trigger value, a due-date, a list of stochastic and/or time-dependent activities with costs, risks and resources, as well as a stochastic risk evaluation.
- For assets with condition-based or reactive maintenance the current condition is analysed in nowcasting. If the current condition exceeds the trigger or shows a breakdown, a short-term alert with urgency and due-date, an assessment of service impact, risk evaluation and a list of maintenance activities afflicted with costs and resources will be created.
- Opportunistic interventions are a kind of preventive maintenance. If on a section an intervention is generated, than is there a possibility to do further maintenance on this section without additional disturbing of the service by possessing this section. Into consideration comes preventive, predictive and corrective activities. If the need of maintenance is foreseen or nearby, it can be suitable to prepone the activity and combine it with another, needed intervention.

In tactical planning, the alerts or interventions generated by the alert & intervention management tool will be planned in medium-term. This enfolds the selection, combination and allocation to time intervals. This plan should be cost efficient, should reduce track possession, as well as it has to hold risk constraints and resource restrictions. To create a robust plan, the future will be simulated. This consists of defining a lot of random future scenarios with new occurring alerts and a certain track condition development. For each scenario, dynamic planning will be used to simulate maintenance execution. From the output of the simulation, a set of internal KPIs will be calculated, e.g. the average costs, the quantiles of costs, the probability for irreconcilable risks. Based on the internal KPIs, the plan will be improved until a robust plan with a good cost-risk-balance is found.

Based on the resulting coarse tactical plan, the operator can order material, book machines or track possession and plan holidays.

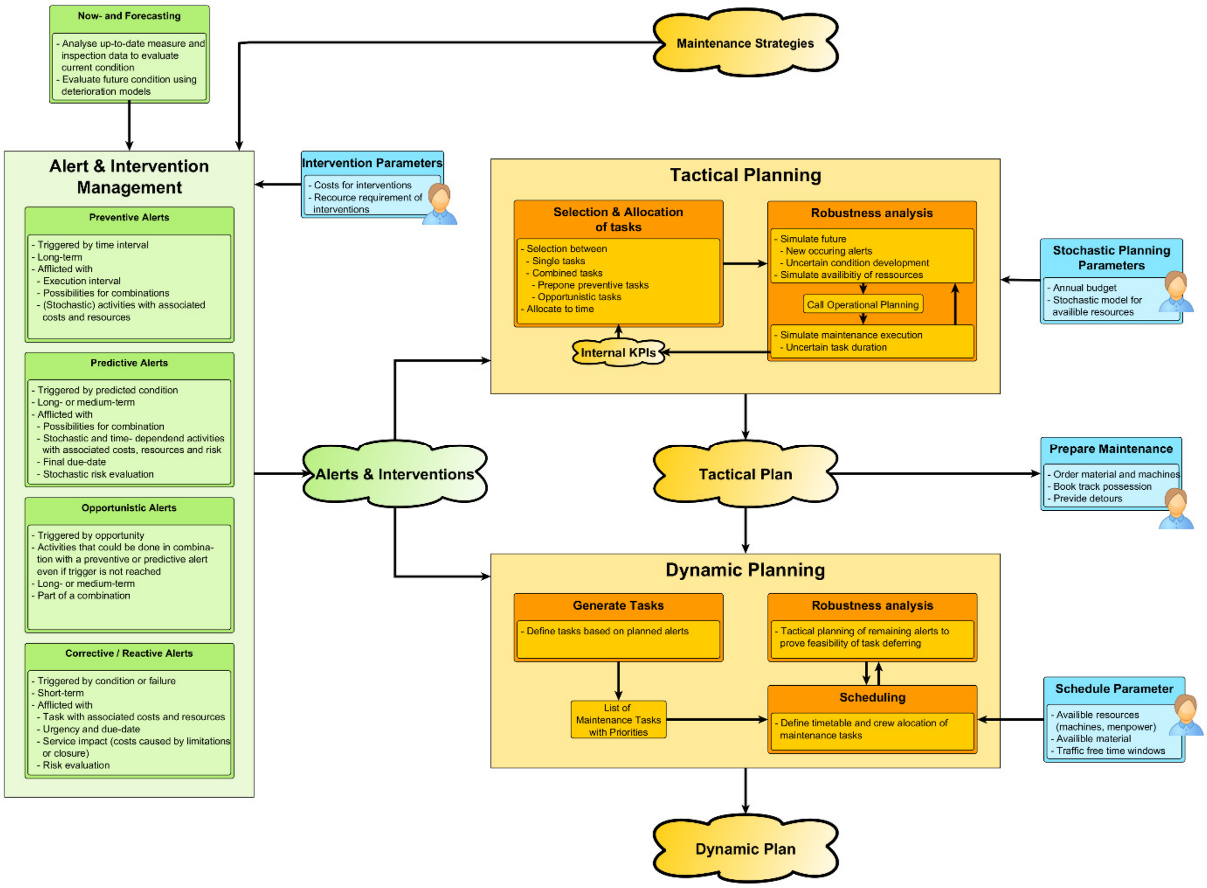

Figure 27. Framework for Tactical and Dynamic Planning

In dynamic planning, the selected and allocated maintenance activities, as the case afflicted with new condition information, and newly added and urgent activities will be scheduled in detail. Thereby, the changeable operational constraint like men power, machines available, material available, etc. will be considered. Dynamic planning starts with analysing the preplanned and new alerts and intervention in order to generate aprioriesed list of maintenance tasks. These tasks will be scheduled, for each task the crew and time for execution are defined. Thereby should be considered that the time needed to execute a certain maintenance task is only an assumption and that delays can occur. During the robustness analyse is proven, whether all urgent tasks can be executed before a critical stage and tasks, which will be deferred in case of delays, can be executed in the next dynamic planning period.

The output of dynamic planning is a timetable for each crew and a remark, which task can be skipped in case of delay.

6.2 Workflow Diagram

This chapter and the next one are devoted to the more system-oriented aspects of the smart decision support framework and show how the eIMS system will support the decision-making process.

As it is common in system architecture, we demonstrate the system in development in two aspects. Both of them are called behavioural diagrams in UML terminology, because both diagram types gives information about the dynamics of the would-be systems.

- In this chapter we provide ***workflow diagrams*** (also called activity diagrams, see Figure 28 and Figure 29) to demonstrate how and in what order system processes will be made available to the user and how groups of system functions will follow each other. Workflow diagrams give a user-oriented view of the system, they model the would-be processes (both computational and organisational ones) and gives an overall view of flow of control within the would-be system.

- ***Interaction diagrams*** (see Figure 30 and Figure 31), on the other hand, give a much more technical view of the system. They demonstrate how system components and processes interoperate and how (and in what order) messages are propagated between the different components of the system. Interaction diagrams will be shown in the next chapter.

As it can also be seen in 6.1.1, functionality related to decision support can be separated into two parts: strategic planning cooperates strongly with RAMS&LCC computation, whereas tactical and dynamic planning are related to each other and both are closely depend on alert management. Therefore both workflow and interaction diagrams are split into two parts with one pair of diagrams showing strategic planning and related functions and the other pair showing the strongly connected alert management/tactical planning feature set.

6.3 Interaction Diagram

As mentioned in the introduction of previous chapter, this section shows the interaction between main components of the system. Though the description of system architecture is out of the scope of this document, a brief explanation is necessary to understand the following diagrams: eIMS will be a so called 3-tier cloud-based system with a rich internet ***presentation layer*** to facilitate user interaction, a Java-based middleware to host ***business logic*** (that is, ontology, computations, expert-based features) and a ***persistence layer*** (database) called ***Data Farm*** to store all the data necessary for expert-based computations. To support definition integrity and seamless cooperation of expert-based features, also partial results (having been computed by one toolkit for the use of another one) will also be stored in the Data Farm.

The two diagrams in Figure 30 and Figure 31 demonstrate how this architecture will support the strategic and tactical/dynamic planning processes, respectively. One can see that the communication between different toolkit elements always takes place through the business logic layer that guarantees that provides definition integrity and controlled communication for the individual toolkit operations. It is also shown that results of operations are stored in the Data Farm, thus every time one toolkit performs a computation or user interaction, the outcomes of the operation can be made available to other toolkits, as well.

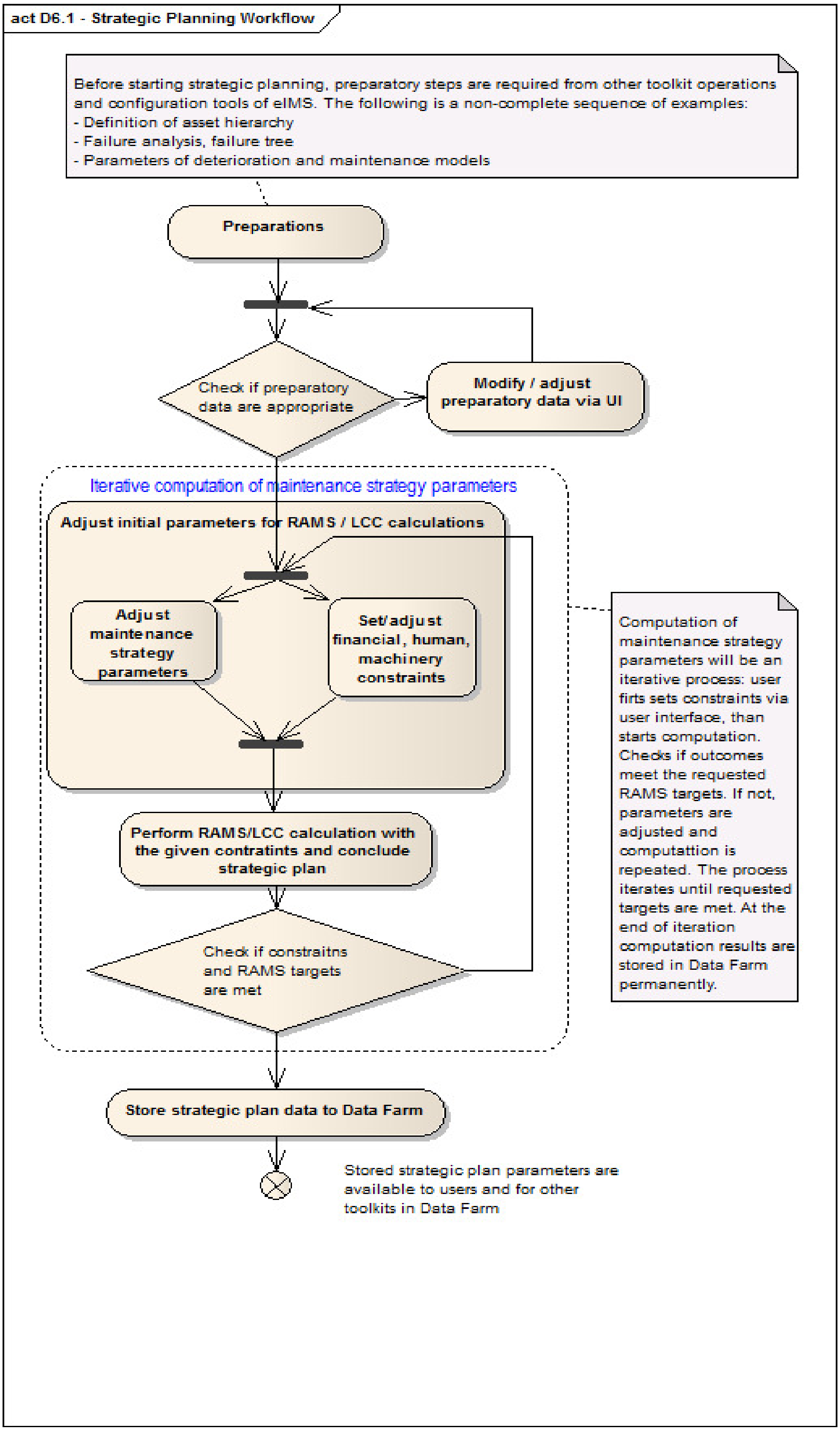

FIGURE 28. STRATEGIC PLANNING WORKFLOW DIAGRAM

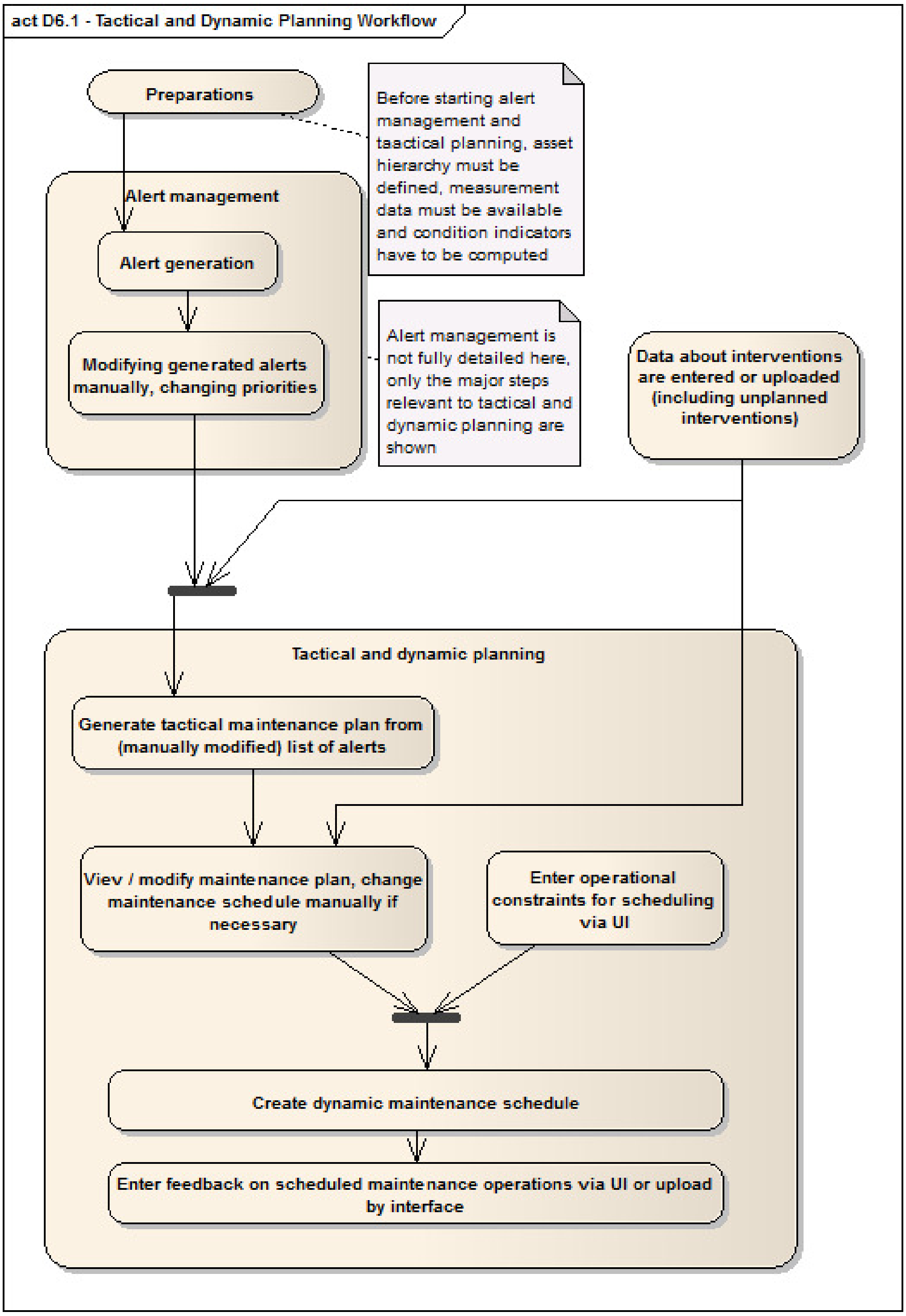

FIGURE 29. TACTICAL AND DYNAMIC PLANNING WORKFLOW

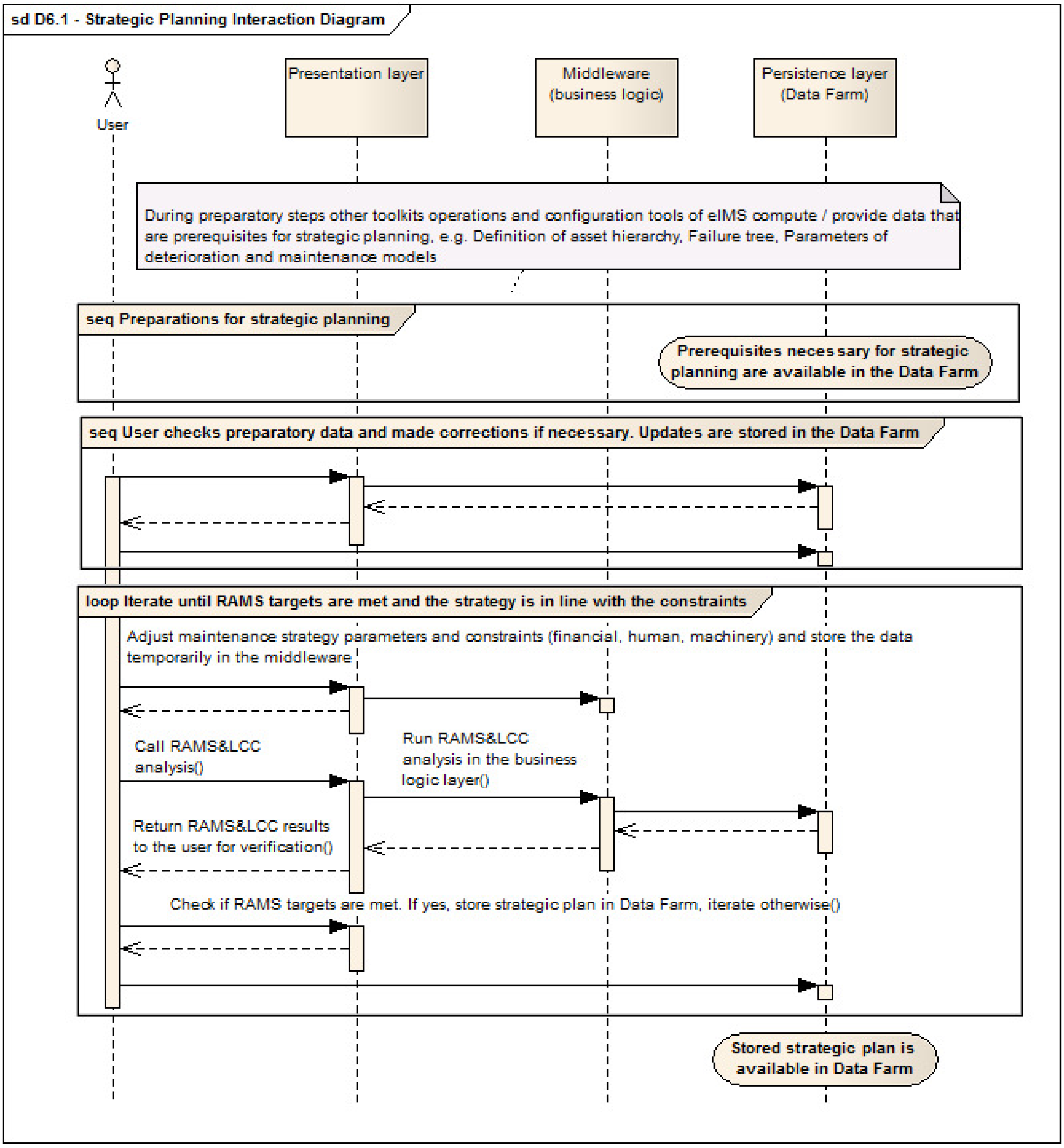

Figure 30. Strategic Planning Interaction Diagram

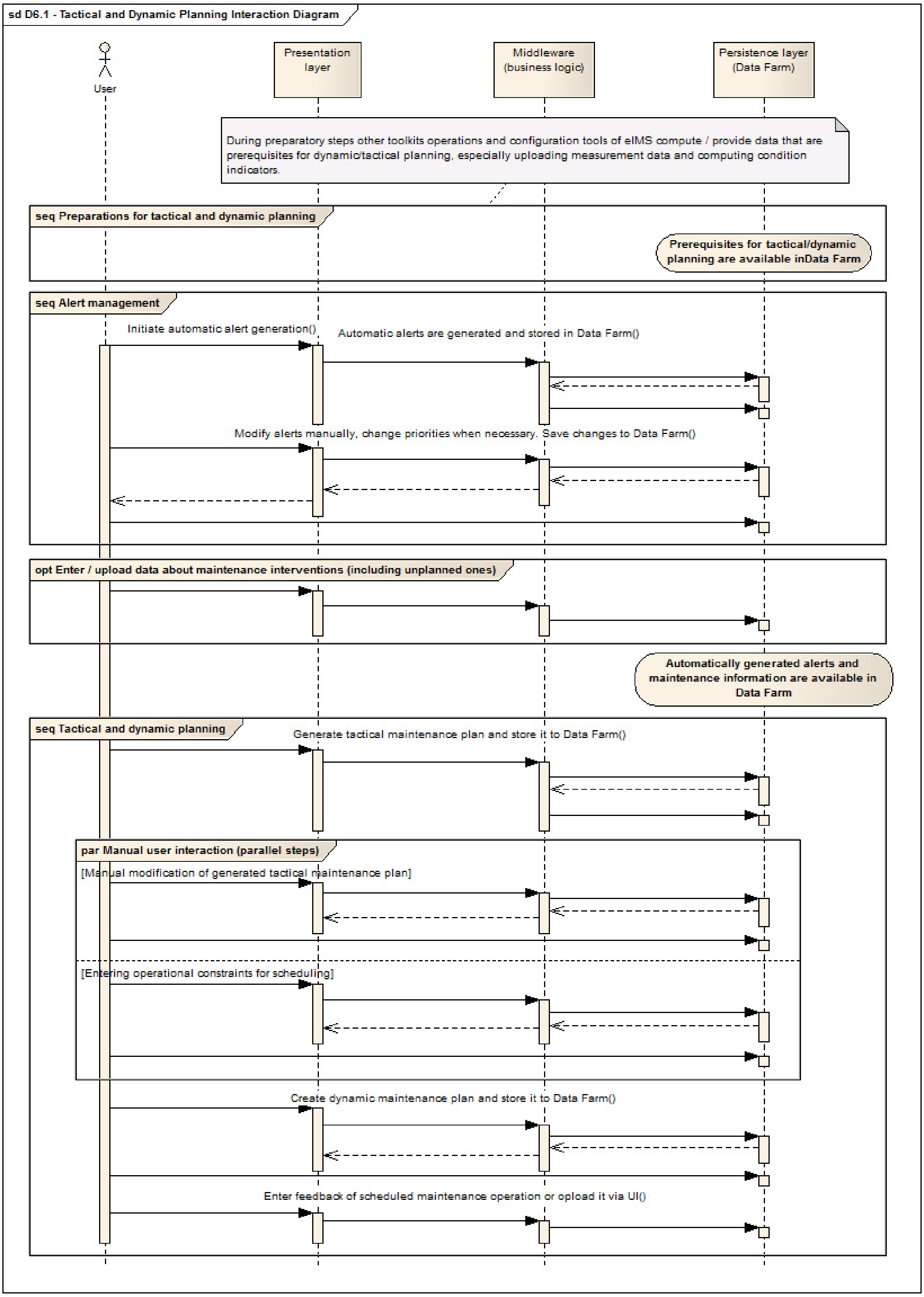

Figure 31. Tactical and Dynamic Planning Interaction Diagram

7 Conclusions

In this deliverable a condition- and risk-based concept for maintenance planning has been defined and integrated into a framework for smart decision support, specifying the workflow between building blocks on strategic, tactical and operational planning level. The planning concept uses sound mathematical approaches to optimisation under uncertainty, and has its background on existing maintenance policies and planning principles, which have been reviewed in this document.

The main conclusions that can be drawn for future research and develepment are as follows:

- Existing approaches to maintenance management, reaching from policies like condition-based, preventive and predictive maintenance to planning principles like risk-based, reliability-centred, opportunistic, are suited for a restricted use case only and have to be combined, extended and adapted in order to contribute to a whole-system planning for a transport infrastructure.
- The separation of the overall maintenance planning process into three planning levels - strategic, tactical and operational - is reasonable when made on the basis of the level of detail considered in each task as well as on the planning decision which are concerned.
- A sound definition of the output coming from ambient building blocks which has to be used for planning - namely RAMS and LCC models, risk assessment, nowcasting and forecasting, alert management - is an important aspect to be carefully elaborated.
- The use of mathematical models and concepts for optimisation under uncertainty - amongst others robust optimisation, probabilistic contraint programming, Markov decision processes - is an added value when designing a planning concept for maintenance decision-making.
- To solve mathematical optimisation models defined in this way, scenario generation and sampling is a technique to be considered, e.g. in solutions approaches like Monte Carlo Rollout. The proper use of sampling techniques is critical in order to overcome challenges accrued with rare events that regularly occur in asset management processes.
- Three differing approaches to risk-based optimisation have been identified - implicit, explicit and semi-explicit - and proposed to be integrated into the concept according to their suitability.
- The presented theoretical background, mathematical modelling and solution approaches provide the means for an effective and efficient use of probabilistic information in infrastructure maintenance and interventions planning, as declared as on of the main innovative features of the INFRALERT project.
- The necessity of a sound management of information related to uncertainty has been justified, which mainly has to be concerned with the quantification of uncertainties and the value of probabilistic information, but also with a monitoring and updating mechanism.

8 Glossary of terms

Asset	The physical transportation infrastructure (e.g. travel way, structures, etc.); more generally can include the full range of resources capable of producing value-added for an agency (e.g. human resources, equipment, materials, financial capacity, real state, corporate information, etc.)
Asset condition	Measure of an asset's physical state as affected by deterioration and past maintenance and repair; can be expressed in terms of damage present (e.g. amount or percentage of cracking), an agency defined or standard scale (e.g. condition states 1 through 5; or good, fair, poor); often used in conjunction with 'performance' when described in the context of performance-based processes.
Asset Management	A strategic approach to managing transportation infrastructure. It focusses on business processes for resource allocation and utilisation with the objective of better decision making based upon quality information and well defined objectives.
Data	Raw or partially processed observations, measurements, facts, figures, statistics, records, etc. collected by an agency.
Decision	Determination of a course of action or selection of an option from available choices.
Decision support	The use of information (e.g. from management systems, other analytic tools, or estimates and studies by staff) to help understand the consequences of decisions.
eIMS	The expert-based Infrastructure Management System (eIMS) is a cloud-based system which hosts the expert-based toolkits developed by INFRALERT and includes all the necessary integration and communication layers.
Interventions	Synonym of actions, it denotes any maintenance activities, such as inspection, repair, renewal or new construction of infrastructure assets, or any operational change, such as lane/line closure, speed restrictions, traffic timetables, diversions, etc.
Life cycle	A length of time that spans the stages of asset construction, operation, maintenance, rehabilitation, and reconstruction or disposal/abandonment; when associated with analyses, refers to a length of time sufficient to span these several stages and to capture the costs, benefits, and long-term performance impacts of different investment options.
Maintenance	Program of activities to enable a transportation system to continue to perform at its intended level; comprises a range of services in preservation, cleaning, replacing worn or failed components, periodic or unscheduled repairs and upkeep, user services (incident response, hazardous materials response), snow and ice control, and servicing of traffic devices and aids; does not add to structural or operational capacity of an existing facility.
Monitoring	Collecting and processing condition and performance data and related data (e.g., traffic usage) to understand the current status of the transportation system, identify problem areas, gauge improvements resulting from investments, and track progress toward performance targets; provides a feedback mechanism for resource allocation and utilization decisions.

Preventive maintenance	Proactive maintenance approach that is applied while the asset is still in good condition; extends asset life by preventing the onset or growth (propagation) of distress.
Infrastructure Manager	Any public body or undertaking responsible in particular for establishing and maintaining road or railway infrastructure, as well as for operating the control and safety systems. The functions of the infrastructure manager on a network or a part of a network may be allocated to different bodies or undertakings.
Renewal	All activities involved in replacing a infrastructure part or object by a same or similar type of infrastructure part or object. It is capitalised at the time it is carried out, and then depreciated.
Nowcasting	Nowcasting methods are used to identify faults that will lead to failure within a few hours; this is done for safety and to extend remaining useful life (RUL).
Forecasting	Forecasting can be useful to assess the condition of a linear asset for the remaining useful life in the long run.
Failure	Departure of a component's functionality targets from specification. Termination of the ability of an entity to perform a required function under specified conditions.
Failure mode	Predicted or observed consequences of a failure in a component or system in relation to the operation conditions at the time of failure. Effect by which a failure is observed.

9 References

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

[2] David J. Smith, Reliability, Maintainability and Risk 8th Edition: Practical Methods for Engineers including Reliability Centred Maintenance and Safety-Related Systems, Butterworth-Heinemann, Oxford, 2011

[3] UNE-EN 50126-1:2010. Railway applications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS). Part 1: Basic requirements and generic process. British Standards (BSI), 2010

[4] NTNU.edu, Reliability Data Sources - ROSS - NTNU, www.ntnu.edu/ross/info/data, accessed 05.04.2016

[5] J.J.A. van den Breemer, RAMS and LCC in the design process of infrastructural construction projects : an implementation case, Civil Engineering and Management MSc (60026), University of Twente, The Netherlands, 2009

[6] Al-Jibouri, S., Ogink, G., Proposed model for integrating RAMS method in the design process in construction. Architectural Engineering and Design Management, 5(4), 179–92, 2009

[7] Patra, A.P., Maintenance decision support models for railway infrastructure using RAMS & LCC analyses (Doctoral thesis), LTU, 2009

[8] INNOTRACK consortium, Development of cost-effective high performance track infrastructure for heavy and light rail systems, available at: cordis.europa.eu/project/rcn/81513_en.html, 2005

[9] International Standard ISO 31000:2009(E): Risk management - Principles and guidelines, 2009

[10] Meyer, T., Reniers, G., Engineering Risk Management, De Gruyter Textbook, 2013

[11] International Standard IEC/FDIS 31010:2009(E): Risk management - Risk assessment techniques, 2009

[12] P.K. Marhavilas, D. Koulouriotis, V. Gemeni, Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000-2009, Journal of Loss Prevention in the Process Industries 24, 2011

[13] P. Chemweno, L. Pintelon, A. van Horenbeek, P. Muchiri, Development of a risk assessment selection methodology for asses maintenance decision making: An analytic network process (ANP) approach, International Journal Production Economics 170, 2015

[14] V. Cozzani, G. Antonioni, G. Spadoni, Quantitative assessment of domino scenarios by a GIS-based software tool, Journal of Loss Prevention in the Process Industries 19(5), 2006

[15] P. Trucco, M. Cavallin, A quantitative approach to clinical risk assessment: the CREA method, Safety Science 44(6), 2006

[16] E.R. Vaidogas, First step towards preventing losses due to mechanical damage from abnormal actions: knowledge-based forecasting the actions, , Journal of Loss Prevention in the Process Industries 19(5), 2006

[17] N.S. Arunraj, J. Maiti, Risk-based maintenance - Techniques and applications, Journal of Hazardous Materials 142, 2007

[18] N.N. Taleb, The Black Swan: The Impact of the Highly Improbable (2nd ed.), New York, 2010

[19] A.B. Owen, Monte Carlo theory, methods and examples, 2013

[20] L.M. Quiroga, E. Schnieder, A heuristic approach to railway track mainenance scheduling, Computers in Railways XII, 2010

[21] C. Meier-Hirmer, A. Senée, G. Riboulet, F. Sourget, M. Roussignol, A decision support system for track maintenance, WIT Transactions on The Built Environment, Vol. 88, 2006

[22] Y. Shafahi, R. Hakhamaneshi, "Application of a Maintenance Management Model for Iranian Railways Based on the Markov Chain and Probabilistix Dynamic Planning, Structure and Infrastructure Engineering, 2010

[23] N. Lyngby, P. Hokstad, J. Vatn, RAMS Management of Railway Tracks, Handbook of Perfomrance Engineering, Springer, 2008

[24] D. Prescott, J. Andrews, Modelling Maintenance in Railway Infrastructure Management, Reliability and Maintainability Symposium (RAMS), 2013

[25] L. Podofillini, E. Zio, J. Vatn, Risk-informed optimisation of railway tracks inspection and maintenance procedures, Reliability Engineering and System Safety 91, 2006

[26] N.N. Taleb, Antifragile: Things That Gain from Disorder, New York, 2012

[27] Bertsekas, D. & Castanon, D. A.: Rollout Algorithms for Stochastic Scheduling Problems, Journal of Heuristics, 5, 89-108, 1999

[28] Chaslot, G.; Bakkes, S.; Szita, I. & Spronck, P.: Monte-Carlo Tree Search: A New Framework for Game AI, Proceedings of the Fourth Artificial Intelligence and Interactive Digital Entertainment Conference,216-217, 2008

[29] Holfeld, D.; Simroth, A.: Monte Carlo Rollout method for optimization under uncertainty, 15th International Conference on Civil, Structural and Environmental Engineering Computing CC2015, Civil-Comp Press, 2015

10 Appendices

Appendix 1: Best practices of risk assessment methods and techniques in the field of oil and gas