Horizon 2020
European Union Funding
for Research & Innovation

In2Rail

In2Rail

Project Title:	**INNOVATIVE INTELLIGENT RAIL**
Starting date:	01/05/2016
Duration in months:	36
Call (part) identifier:	H2020-MG-2014
Grant agreement no:	635900

Deliverable D9.4
Asset Status Forecasting and Feature Selection Methodologies

Due date of deliverable	30-10-2017
Actual submission date	24-10-2017
Organization name of lead contractor for this deliverable	LTU
Dissemination level	CO
Revision	Final

AUTHORS

		Details of contribution
Author(s)	**Luleå University of Technology (LTU)** Matti Rantatalo Adithya Thaduri Madhav Mishra Johan Odelius	Coordination of the Deliverable Scenario description
Contributor(s)	**Ansaldo STS (ASTS)** Carlo Dambra	WP9 coordinator
	Trafikverket TRV Arne Nissen Karl-Einar Jonsson Magnus Wahlborg Eric Neldemo	Scenario description, data sharing
	University of Genoa (UNIGE) Emanuele Fumeo Davide Anguita Luca Oneto	Scenario description, management of scenario
	Deutsches Zentrum für Luft- und Raumfahrt e.V. / German Aerospace Center (DLR) Jörn Groos Beate Dutschk Daniela Narezo Guzman Christian Rahmig	Scenario description, management of scenario
	Swedish Institute of Computer Science (SICS) Martin Joborn	Scenario description
	Evoleo Technologies (EVOLEO) Rodolfo Martins Pedro Ribeiro Saulo Sobreiro Hélder Fonseca	Scenario description, data sharing
	Universidade Do Porto (UPORTO) Rui Calçada Adriano Carvalho Rosaldo Rossetti Cecília Vale Pedro Montenegro José Neto Tiago Azevedo Cristiana Bonifácio	Scenario description, data sharing
	Virtual Vehicle (ViF) Bernd Luber	Scenario description, management of scenario
	Network Rail (NR) Nadia Hoodbhoy	Scenario description, data sharing
	Rete Ferroviaria Italiana (RFI) Renzo Canepa Simone Petralli	Scenario description, data sharing
	Strukton Rail (SR) Edin Hadzic Henk Samson	Scenario description, data sharing

Executive Summary

This work is complementary to the one described in the other documents outputted by Task 9.1 (i.e. D9.1 & D9.2), continues to Task 9.2 with nowcasting methodologies (with D9.3) and paves the way towards verification, validation and assessment of these methodologies (within D9.5). In D9.3, different scenarios were designed and developed for nowcasting of the railway assets. In D9.4, five scenarios are developed for forecasting of the railway assets.

The NR scenario will deal with the problem of assessing the impact of several different types of asset failures on the traffic by analyzing the relevant data related to train movements, assets and their failures, and delay attributions. The proposal of this forecasting scenario is to build data-driven models that can measure the impact of an asset failure by considering the current situation of the railway network, described by data.

In TRV/LTU forecasting scenario, the main objective is to forecast the status of S&C by calculating the probability of failure using different data sources for rerouting the trains. This was achieved by first predicting the standard deviation of track geometry in three panels of S&C using hybrid approach by estimating parameter model and particle filter based approach. Then, the remaining useful life and probability of failure was calculated considering the threshold limits for carrying out tamping. The obtained results are also useful for TMS for scheduling the maintenance (tamping) and speed reduction.

The main objective of the ViF/UPORTO/UNIGE scenario is to forecast the risk of derailment by fast calculation methods with sufficient prediction quality as a decision basis for TMS/Maintenance. The developed forecasting method includes the prediction of several input parameters in combination with controllable parameters. The forecasting results support TMS regarding speed reductions/closing the line and track geometry correction.

In SR/UNIGE scenario, aims at designing, implementing, testing and, at a future stage, validating a set of predictive models for forecasting purposes, based on data provided by SR about maintenance/repair actions and weather data. The problems that will be investigated (are correlation and influence of executed maintenance/repair actions on failures and influence of the weather conditions on failures. A set of predictive models able to forecast the probability of failures for an asset will be developed based on the data provided by SR about historical weather conditions, performed maintenance/repair actions and failures.

The SR/DLR scenario addresses the forecasting of potential switch failures to prevent train delays and to improve switch maintenance and repair procedures. A statistical process control approach based on the detection of unusual behavior was implemented and evaluated. By this approach, the model training is led with data from normal operation and no explicit knowledge about possible failure types and their characteristics is required beforehand. The forecasting results show that the approach is capable to detect variations due to emerging malfunctions in a very early stage of development depending of the type of failure.

TABLE OF CONTENTS

Abbreviations and acronyms

Abbreviation / Acronyms	Description
CI	Confidence Interval
CRISP-DM	Cross Industry Standard Process for Data Mining
DA	TRUST Delay Attribution
DAG	Delay Attribution Guide
FC	Forecasting
FMS	Fault Management System
Index "P"	Predicted/Forecasted values
Index "T"	True measurement values
KNMI	Royal Netherlands Meteorological Institute
KPI	Key Performing Indicator
KRLS	Kernel Regularized Least Squares
MGT	Million Gross Ton
NC	Nowcasting
NSE	Electrical point machine of NS type switch
ORE	Office for Research and Experiments
PCA	Principal Component Analysis
POSS®	Preventive Maintenance and Fault Diagnosis System
Q	Vertical wheel/rail force
RUL	Remaining Useful Life
S&C	Switching and Crossing
SDH	Standard Deviation Average
SPC	Statistical Process Control
SPE	Squared Prediction Error
Stanox	Station Number
TGD	Track geometry degradation
TMS	Train Management System
TOPS	Total Operations Processing System
TRUST	Train Running System on Total Operating Processing System
WTT	Working Timetable
Y	Lateral wheel/rail force

1. Background

The present document constitutes the first issue of Deliverable D9.4 “Asset status forecasting and feature selection methodologies” in the framework of the Project titled “Innovative Intelligent Rail” (Project Acronym: In2Rail; Grant Agreement No 635900). D9.4 is the report of part of the activities of Task 9.2 – “Asset status forecasting for TMS/dispatching system” related to asset status nowcasting methodologies.

The outcome of the activities described in D9.4 is an important tool for In2Rail research. The information and knowledge obtained from this document serves the Traffic Management Systems (TMS) for carrying out maintenance decision-making of future status of railway assets. Transforming this data into actionable knowledge is a key task that must be solved to exploit its fullest potential. The forecasting methodologies presented in this document will possibly change the shape of the TMS/maintenance decision support systems of the future that can be integrated into the operational management of railways and ways forward.

The document gives a defined in the context of In2Rail, it describes the forecasting scenarios that have been designed through the collaboration between all the partners involved in WP9, the proposed solutions and preliminary results. These scenarios are independent and showcase different approaches to forecast the future status of the railway assets.

It is important to highlight that this deliverable focuses on forecasting modelling approaches and on the related preliminary results achieved during laboratory tests. The extended concrete results of both nowcasting and forecasting will be presented in the upcoming Deliverable D9.5 “Nowcasting and Forecasting algorithms verification, evaluation and assessment report”.

The next steps in this work package WP9 is the evaluation and validation of the models, and the additional information regarding the nowcasting and forecasting scenarios will be provided in Deliverables 9.5.

2. Objective / Aim

The main goal of this document is to describe the work done in the context of the WP9 – "Nowcasting and Forecasting" of the In2Rail project regarding Task 9.2 "Asset status forecasting for TMS/dispatching system". The forecasting is defined as the process of exploiting past and present data to make deductions about the future.

This document describes the forecasting methodologies developed in the context of In2Rail WP9 as part of the activities of Task 9.2. The deliverable describes the work done on a per-scenario basis and the forecasting methodologies proposed in the Task 9.2 activities.

The objectives of the individual forecasting scenarios are defined below:

The objective of NR scenario will deal with the problem of assessing the impact of several different types of asset failures on the traffic by analyzing the relevant data related to train movements, assets and their failures, and delay attributions (i.e. delay effects).

The objective of TRV/LTU scenario is to forecast the status of the Switches and Crossing by probability of failure using different data sources for rerouting the trains.

The objective ViF/UPORTO/UNIGE Scenario is to forecast the risk of derailment by fast calculation methods with sufficient prediction quality as a decision basis for TMS/Maintenance.

The main objective of SR/UNIGE scenario is to forecast possible failures of assets based on the correlation of past asset failures and past weather conditions or maintenance actions, considering a set of different infrastructure assets selected as the most relevant ones from the TMS perspective

The objective of the SR/DLR-scenario is to forecast the status of the switches in order to improve the maintenance and the repair procedures. In order to achieve this goal, data mining methods will be exploited by relying on the historical data regarding the switches.

3. Forecast Scenarios

3.1 NR Scenario (*Theoretical)*

UNIGE and NR defined the main idea for a new scenario, whose description is included in this chapter. It has been agreed by all the partners of the WP9 that this work will be only described in a theoretical fashion, but no developments will be carried out in the context In2Rail due to limitations that cannot be solved within the time frame of the project. The idea will be eventually recovered in the next EU funded projects of the Shift2Rail initiative, in case the partners will be able to collaborate again.

Title	**Impact analysis of asset failures on traffic**
Organisations Involved	• Network Rail (NR) • University of Genoa (UNIGE)
Objective(s) of the scenario	The scenario will deal with the problem of assessing the impact of several different types of asset failures on the traffic by analyzing the relevant data related to train movements, assets and their failures, and delay attributions (i.e. delay effects). The analysis could be extended by integrating weather data, maintenance data and train characteristics (e.g. train composition and weight) to be associated with data about train movements.
Relationship with TMS and/or maintenance	The proposed solution could be used in real time by the TMS in order to immediately have a measure of the impact of an asset failure to the railway traffic / operations. Alternatively, the solution proposed could be used in an offline, simulated fashion by the maintenance department in order to prioritize maintenance interventions based on the real impact of a sudden asset failure to the railway traffic / operations. The final goal is not to replace the TMS-Decision-support-function but to improve it by including another decision support system which can provide useful and actionable information for the TMS.
Description of the scenario	It is usually difficult to measure the impact of an asset failure to the railway traffic and operations with expressive KPIs, which could be easily understood by rail traffic regulators. Sometimes, an asset failure has little or no impact on the circulation, but sometimes the complexity of railway networks, the variety of timetable with respect to the date and time in which a failure occurs, which is directly connected to the number of trains circulating on the railway network at a specific time, as well as many other factors cannot be easily taken into account. The proposal of this forecasting scenario is to build data-driven models that are able to measure the impact of an asset (or category of assets) failure by taking into account the current situation of the railway network, described by data. For this purpose, the main sources of historical information are records of train movements, assets failures and relationships between failures and their effects on the circulation, which are all available from NR databases. Eventually, other relevant information (e.g. weather data) will be evaluated for integration in data-driven models. The impact of the assets failures can be measured through different parameters, e.g. in delay minutes, numbers of trains affected and any resultant cancellations, and the like. This information are included in the Delay Attribution Records and will be used to study and assess the best combination of KPIs able to describe the impact of an asset failure in the most expressive and understandable way. Once the KPIs will be defined, data-driven models will be built by investigating the relationship between different ways of describing the current situation of a railway network with the KPIs related to the asset failure impact. The extraction and

	engineering of features will be of paramount importance in order to derive inputs for the data-driven models from the original datasets. For example, indicators of the number of trains travelling on the railway line and their actual delays, the number of tracks available, the scheduled trains that have to travel on the same specific line, and many others will be investigated and eventually included as inputs of the data-driven models. This work will be done at different aggregation levels with two goals. On the one hand, we want to make the problem treatable through machine learning techniques able to achieve a reasonable accuracy level. On the other hand, we would like to provide to the Infrastructure Managers a tool that can assess the impact of failures at different levels, starting from the most general one (e.g. the impact of failures on a specific area or station) to the most detailed one (e.g. the impact of that specific type of failure of that specific asset.
Data exploited for the scenario	Data already owned by Network Rail: • Train Running System on Total Operating Processing System (TRUST) train movements (same as the ones exploited for the old scenario); • ELLIPSE asset register; • Fault Management System (FMS) data, referring to asset failures and related information. Data whose availability has to be assessed: • weather data; • asset maintenance data; • train characteristics (e.g. train composition and weight).

Table 3.1: Tabular Description for Scenario by NR/UNIGE

3.1.1. Data Sources detailed description

The data sources already owned by Network Rail are described briefly below[1]:

- **ELLIPSE:** Mincom Ellipse, previously known as MIMS, is a tool owned by Network Rail and used for managing the infrastructure assets. Among its other functions, the system has an equipment register and records all the schedule maintenance tasks carried out on each asset;
- **FMS:** Fault Management System, is used by Network Rail to manage and report faults of the network infrastructure assets. Faults are logged on the system against equipment assets and allocated a unique identifier regardless of whether they caused train service delays/cancellations or not. FMS is configured as a series of operation systems in Network Rail territories, which are linked to a central data repository (FMS Central). Information about equipment location, asset type, asset ID, etc is sourced from the Mincom Ellipse and the Telecomm Equipment Database (EQUIP);
- **TRUST:** Train Running System on TOPS (Total Operations Processing System) is owned by Network Rail who uses it for monitoring train performance. TRUST monitors movements at particular Recording Points along the train's journey and

[1] For more details see the user manuals on Network Rail's ASD Online -

compares the actual time with the working timetable. The comparison between the actual time and the booked time provides a lateness value at that point. If a train loses 3 minutes, or more, or is cancelled at its starting point or between two consecutive Recording Points, the delay/cancellation is recorded in TRUST. A unique identification number is given to each incident and delays to individual trains are allocated to that incident, as applicable. The Recording Points in TRUST are identified through their STANOX code, which is a 5-digit number allocated to each rail location.

3.1.1.1. ELLIPSE asset register data

The ELLIPSE asset database consists of a table composed of the following columns:

- **Equip No:** internal equipment identification number;
- **EGI Code:** Equipment Group Identifier code;
- **ELR:** engineer's line reference;
- **Track ID:** track identifier;
- **Start Mileage:** start mileage of the location of the asset;
- **End Mileage**: end mileage of the location of the asset;
- **Route:** main route on the English railway network;
- **DU:** Delivery Unit;
- **Asset Desc 1:** first string describing the asset;
- **Asset Desc 2:** secondary string describing the asset;
- **Equipment Location Desc:** longer string describing the location of the asset;
- **Equipment Status:** code describing the status of the asset (can be either FM or PM);
- **Asset Class Code:** a code related to the asset class.

3.1.1.2. Fault Management System (FMS) data

The FSM database consists of the following information:

- **Failure Number:** internal failure identification code;
- **Date_Time:** date and time of the notification of the failure occurred;
- **Year:** derived from Date_Time field;
- **Period:** derived from Date_Time field, it expresses the time of the year by dividing it in 13 periods that are 4 weeks long;
- **MIMS ID:** Mincom Information Management System identifier;
- **ELR Ref:** engineer's line reference (see ELLIPSE data);
- **Place Name:** place where the asset failure occurred;
- **Equipment Description:** string describing the failed equipment;
- **Route:** main route on the English railway network;
- **Delivery Unit:** see ELLIPSE data description;
- **System Asset Type:** string describing the failed asset;
- **Suffix:** group/category to which the failed asset belongs;
- **Priority:** failure priority / severity level;

- **Fault/Incident:** Boolean value for differentiating failures from incidents;
- **Component Level 1:** group/category of level 1 to which the failed asset belongs, according to the NR internal asset hierarchy/taxonomy;
- **Component Level 2:** group/category of level 2 to which the failed asset belongs, according to the NR internal asset hierarchy/taxonomy;
- **TRUS Number:** TRUST identification number of the failure/incident. This field is populated only if the failure/incident is linked to one or more train delays, so to express the relationship between the two different datasets;
- **Failure Detail-Cause-Action:** string containing a detailed description of the failure, including the identified cause and the action carried out to solve the restore asset functionality. The three different strings are separated by different tags.

3.1.1.3. TRUST data

The TRUST data includes train movements records composed of the following information:

- **TrainID:** the identification number of a train. For example, taking the Train Id "541D23MU19":

54	1D23	MU	19
First two numbers of the departing station, e.g. 54311 → Kings X (London Kings Cross)	Train unique string	Type of train	Day of the month, e.g. 19/11/2015

- **Timing Event Name**: the type of event occurred, it can be [T, P, A, D, O] where:
  - T – Terminal (final planned destination of the service),
  - P – Passage,
  - A – Arrival,
  - D – Departure,
  - O – Origin;
- **WTT Datetime:** Working Timetable (WTT) date and time, i.e. the theoretical timetable plan;
- **Actual Datetime:** actual time in which the event occurred formatted as "HH24 DD-Mon-YYYY";
- **Direction:** direction of travel, there are only two values, UP (U) and DOWN (D). Indeed, in England & South Wales the direction of travel is described in relation to London (UP is towards London, DOWN is away). In Scotland, the direction is in relation to Edinburgh. The Valley Lines in Wales are UP valley and DOWN valley**;**
- **Section Start Location Name:** name of the location where section starts, e.g. "Leeds";
- **Section Start Location Code:** code of the location where section starts, e.g. "17132";
- **Section End Location Code:** for planned schedule purposes, this field will be blank because the timing point is represented as a single location, 'Section Start Location

XXX'. If there is a performance event resulting in delay between two timing points, the 'Section End Location XXX' will be populated;

- **Section End Location Name:** see "Section End Location Code";
- **Derived Route Name:** name of the NR route, e.g. "LNE". The UK Railway is split into geographically related operational routes. "LNE" relates to London North Eastern, broadly the East Coast Mainline and associated lines. Other routes2 include 'LNW' (London North West), Anglia, South East, Wales, Western, Wessex, Scotland.

3.1.1.4. Delay Incidents Reports

In the Delay Incidents (Historic Delay Attribution) dataset, each record is characterized by the following columns:

- **Financial Year and Period:** the "railway" period that the delay occurred in;
- **Date:** this is the date of the train within the system**;**
- **TrainID:** Performance Systems Strategy (PSS) does not contain UID codes, it contains trainids, which are unique within a railway period for a given route but not within a year. The same train in the timetable should have the same eight digit TrainID, while the last two digits are the day of the month**;**
- **Location codes:** these are Stanox (Station Number) codes – PSS works on TRUST (Train Running System) TOPS (Total Operations Processing System) recording locations, not timetable locations**;**
- **GBTT and WTT:** the GBTT (Great Britain Timetable) times are those that appear in the published timetable, and delays are calculated against;
- **TSC:** this is the train service code at the point where the delay occurred**;**
- **Service Group code:** this is the service group within the Schedule 8 performance regime (and on Real Time PPM – Performance and Punctuality – screens);
- **Operator:** this is the operator code (i.e. TOC – Train Operating Company – which ran the train);
- **English Day Type:** weekday, Saturday, Sunday, bank holiday, Christmas**;**
- **Applicable timetable flag:** if N the train is not in official performance records as it is a short-term replacement of a train plan – normally a reinstatement of part of a cancelled service;
- **Train schedule type / traction type/ trailing load / unit class:** these fields are not mandated or validated;
- **Incident number:** the TRUST Delay Attribution (DA) incident number (not unique without the create date);
- **Incident create date:** the date the incident was entered into the system**;**
- **Incident start/end date:** the date the system has the incident live (this is not the

[2] See https://www.networkrail.co.uk/structure-and-governance/our-routes/ for the complete routes diagram.

length of the incident on the ground);

- **Section code:** where the incident took place**;**
- **Network Rail location Manager:** the area of the country**;**
- **Responsible Manager:** who within the industry is responsible for the delay – all delays have responsibility for performance improvement purposes;
- **Incident Reason:** the Delay Attribution Guide (DAG) cause code for the incident**;**
- **Attribution Status:** All delays go through an acceptance process. Only once they are agreed, the linkage to the incident becomes official. "Disputed delays" normally mean that further investigation is ongoing about either the cause of the incident or the delay;
- **Incident equipment:** internal free form information**;**
- **Incident description:** short description of the incident for internal use**;**
- **Reactionary reason code:** again in the DAG. If no code is provided, the delay is primary (i.e. the delay is at the site of the incident), while if reactionary the delay is a later consequence of that incident**;**
- **Incident Responsible train**: which train initially caused the incident (if any);
- **Performance Event Code**: Whether the train has been delayed or cancelled. "A" and "M" denote delays, while "C", "D", "O", "P", "S" and "F" are cancellations:
  - C is a full cancellation,
  - D is a diversion,
  - F is a failure to stop,
  - S is a scheduled cancellation,
  - O/P are partial cancellations;
- **Start/end stanox:** the location of the delay (not the incident);
- **Event Datetime**: the time the train encountered the delay;
- **Pfpi minutes**: the "size" of the delay. Note that if the train is cancelled (see the Performance Event Code), a deemed delay minute is generated for internal usage in these circumstances, even if the train hasn't been delayed, so that including this delay in any calculation of delay is misleading**;**
- **Train ID responsible**: see Incident Responsible Train;
- **Train ID react**: if the delay is a reactionary delay, the system will try to capture the train that made this train late.

3.1.1.5. Historic delay attribution glossary

This dataset (represented as a table) includes the glossary of some of the columns of the Delay Incidents files, in particular of:

- **Service Group Code:** we have a table with:
  - Service Group Code – Affected,
  - Service Group Description – Affected;
- **Operator Name:** we have a table with:
  - Operator – Affected,

- Operator Name – Affected;

- **Train Service Code:** we have a table with:
  - Service Group Code – Affected,
  - TSC – Affected,
  - TSC Description – Affected;
- **Performance Event Code:** we have a table with:
  - Performance Event Code,
  - Performance Event Group,
  - Performance Event Name;
- **Reactionary Reason Code:** we have a table with:
  - Reactionary Reason Code,
  - Reactionary Reason Description,
  - Reactionary Reason Name;
- **Responsible Manager:** we have a table with:
  - Responsible Organisation NR-TOC/FOC Others,
  - Responsible Organisation,
  - Responsible Manager,
  - Responsible Manager Name,
  - Responsible Organisation Full Name,
  - Responsible Organisation Name;
- **Incident Reason:** we have a table with:
  - Incident Category,
  - Incident Category Description,
  - Incident Category Group Description,
  - Incident Category Super Group Code,
  - Incident JPIP Category,
  - Incident Reason,
  - Incident Reason Description,
  - Incident Reason Name;
- **Period Dates:** we have a table with:
  - Shortened Convention,
  - Financial Year & Period,
  - Min,
  - Max,
  - Day Count;
- **Stannox Codes:** we have a table with:
  - STANOX NO,
  - FULL NAME,
  - CRS CODE,
  - NR ROUTE.

This information is useful for different reasons. Firstly, it represents a reference for understanding all the possible values contained in the data about Delay Incidents (Historic Delay Attribution). Secondly, the glossary gives the possibility to map the data to the physical problems, since giving a context to simple numbers can be fundamental in order to understand the problem fully. Finally, this glossary may allow better normalizing the available data and checking for possible inconsistencies or errors that may cause problems during the analysis.

The application of this scenario in forecasting is because of following limitations:

- the quality of the results of the first NR scenario have been influenced by several factors, mainly related to data quality problems;

Restricting / changing the area of the English railway network under analysis (proposed at last progress meeting) did not solve the issues encountered.

3.2 TRV/LTU Scenario

3.2.1. Summary of Scenario "TRV/LTU" (D9.3)

In this scenario, the nowcasting of S&Cs can be estimated by using the state-of-the-art modelling of Non-Homogeneous Poisson Process and forecasting can be estimated by novel hybrid modelling. The inputs for both the modelling techniques are taken from available data sources, mainly, state of the railway network, asset utilization & maintenance and weather conditions. The specific inputs being; asset registry, failure and inspections, track geometry, loading conditions and weather. The outputs for both nowcasting and forecasting are the probability of failure and facilitates TMS for rerouting the trains as illustrated in Figure 3.1.

Title	**To nowcast and forecast the probability of failure of Switches and Crossings (S&C) for replanning for routing the trains in the railway network.**
Organisations Involved	• Trafikverket (TRV) • Luleå University of Technology (LTU)
Objective(s) of the scenario	The objectives of the scenario are to identify the track geometry condition in S&C, predict the time to reach the threshold and estimate time to restoration.
Relationship with TMS and/or maintenance	The asset nowcasting and forecasting of S&Cs will facilitate TMS for increasing traffic density due to reduced reactive maintenance demand and usage of infrastructure in future and providing statistics for effective traffic management. The TMS user also adjusts the production plan to take the uncertainty of the restoration time into account.
Description of the scenario	In this scenario, the nowcasting of S&Cs can be estimated by using the state-of-the-art modelling of Non-Homogeneous Poisson Process and forecasting can be estimated by novel hybrid modelling. The inputs for both the modelling techniques are taken from available data sources, mainly, state of the railway network, asset utilization & maintenance and weather conditions. The specific inputs being; asset registry, failure and inspections, track geometry, loading conditions and weather. The outputs for both nowcasting and forecasting are probability of failure and time to restoration.
Data exploited for the scenario	The following are the available data sources required to perform the scenario; BIS (Asset register), 0Felia (Failures), Bessy (Inspections), Optram (Track geometry), STEG (Train and load), DS-Analys (Interlocking) and SMHI (Weather). For forecasting, the standard deviation of short-range longitudinal level of both tracks from the Optram is considered.

Table 3.2: Tabular Description for Scenario by TRV/LTU

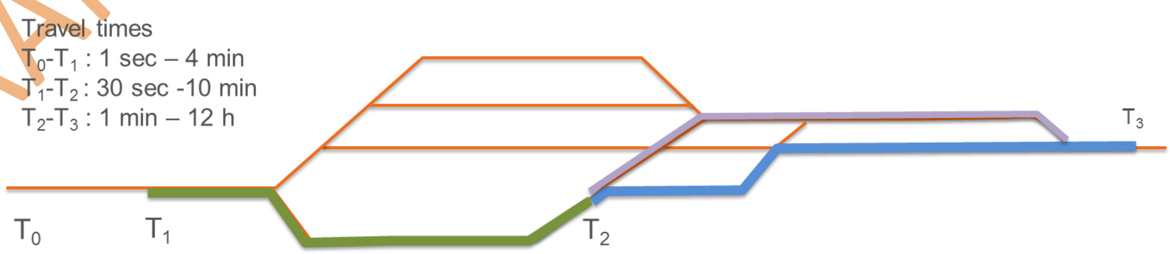


Figure 3.1: TRV/LTU Nowcasting and Forecasting Scenario (D9.3)

Note: The time span of forecasting varies after the time span of locked route to several minutes to days to months that depends on the requirements, length and planning of the train.

3.2.2. Data description for Forecasting scenario

The following are the available data sources required to perform the scenario; BIS (Asset register), 0Felia (Failures), Bessy (Inspections), Optram (Track geometry), STEG (Train and load), DS-Analys (Interlocking) and SMHI (Weather). For forecasting, the standard deviation of short-range longitudinal level of both tracks from the Optram is considered.

3.2.2.1. Input parameters

- Optram (Track Geometry):
  - S&C Type,
  - length of S&C,
  - Standard Deviation SDH = average values of left and right rails;
- STEG (Train and Load):
  - Track Disruption,
  - MGT per year,
  - number of trains,
  - speed of train.

3.2.2.2. Output parameters and its relation to TMS

- S&C predicted value:
  - Remaining Useful Life (RUL),
  - confidence level;
- possible TMS/maintenance actions:
  - speed reductions,
  - closing down a particular line / track section,
  - performing track geometry corrections like Tamping.

3.2.2.3. Uncertainties of the input parameters and forecast

The uncertainty presented in this scenario is mainly based on the model form approaches. Here the objective is set to capture the impact of forecasting and nowcasting achieved by hybrid models and compare to individual based model, i.e. physical based or data driven based. This impact reveals the amount of uncertainty for each individual approach and can be considered whether case the hybrid model increases the accuracy or provide similar result as individual modelling approached. Because of less number of data points, there is high uncertainty in the prediction.

- Track geometry:
  - uncertainties in data acquisition, processing and prediction of SDH;
- Load;
- Speed.

3.2.2.4. Ranking of the parameters

The prediction of the track geometry is mainly conducted based on the longitudinal level of the track. It has the highest priority for forecasting at this point. The other parameters being, in order, load (*MGT*), speed, traffic volume and weather conditions.

3.2.3. Methods for prediction

3.2.3.1. State of the art

Particle filter-based prognostic approach for railway track geometry

Track geometry degradation is a complex phenomenon that occurs due to loading and unloading traffic cycles and results in both elastic and plastic deformation. To predict the behavior, a particle filter based was developed by Mishra et al, 2016 [73]. The parameters of the system state are estimated by extrapolating from a prior state and noisy measurements from *SDH* of S&C. But, this approach requires an initial degradation model for describing the state transition. To predict the track geometry behavior of S&C, a model that describes the relation among the track geometry inputs, need to be selected or developed from the existing models from the literature. There are several existing models that can be taken from D6.3 "Dynamic Track Model". In this scenario, for initial implementation, an existing degradation model (ORE [72]) was used to predict the time for the next tamping action.

As shown in Figure 3.2, after the initial settlement, a gradual decrease in track quality occurs over time, followed by a rapid loss in quality due to the ageing of the track. The initial settlement is often not included in the 18-month prognostic since it occurs directly after tamping. The track geometry degradation model used in this study was proposed by ORE [72]. The equation below consists of two parts: the deterioration directly after maintenance or tamping r_0 and the traffic-induced degradation, which is dependent on the traffic volume *T*, dynamic axle load *Q* and speed *V* as shown in below equation.

$$\sigma = \sigma 0 + hT^{\alpha}(2Q)^{\beta}V^{\gamma}$$

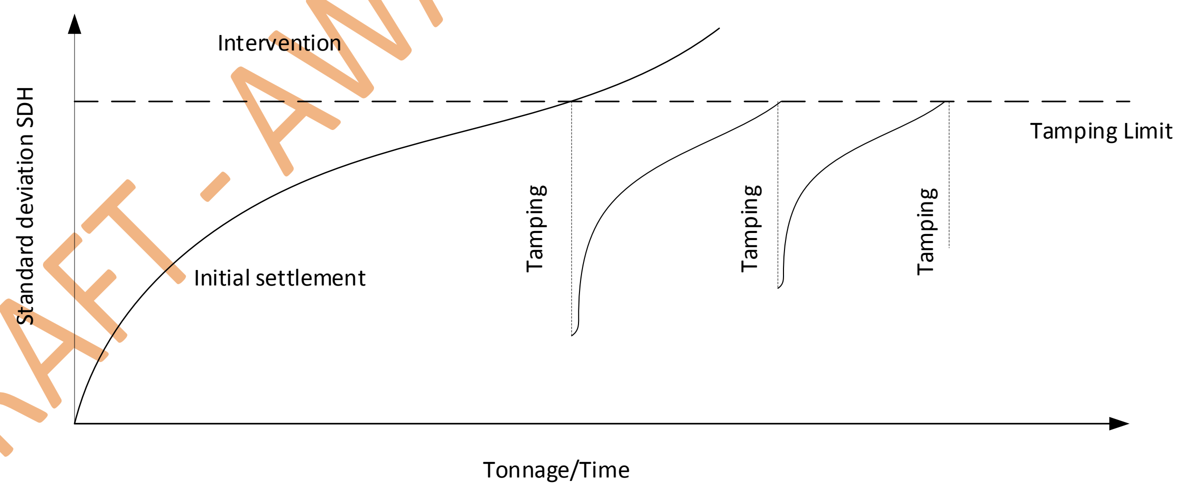


Figure 3.2: Typical diagram of track degradation and restoration over time

where *h* is a constant and the parameters *α*, *β* and *γ* are estimated from experimental data. From the historic record of the operational data, the remaining load (or traffic volume) can be predicted before reaching a threshold defined by experts. The *RUL* of the track can thus be calculated based on the induced load reaching a threshold.

3.2.3.2. Methodology and methods

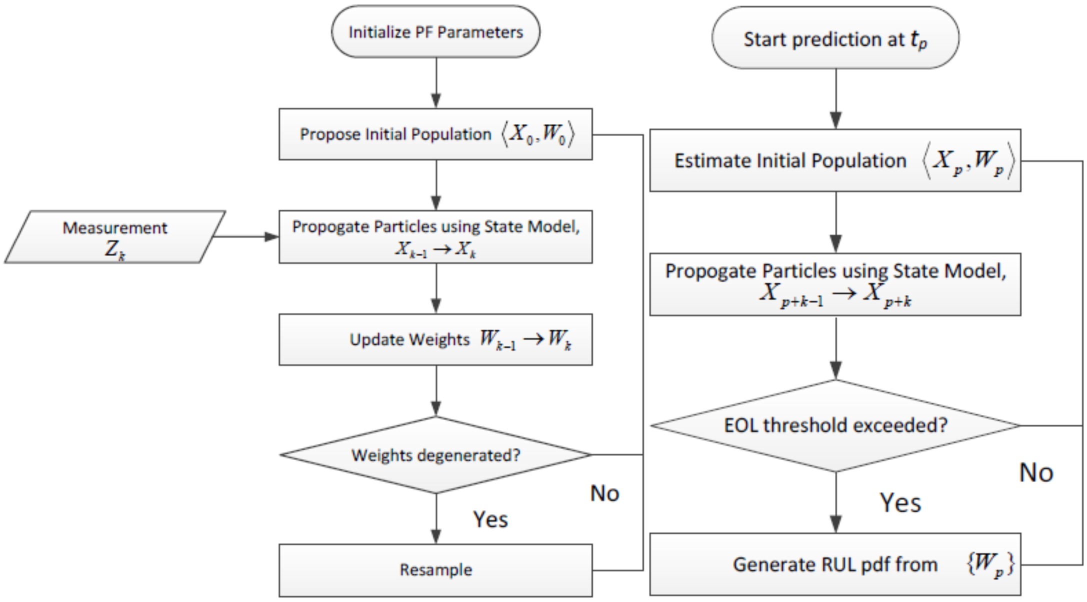


Figure 3.3: Estimation and prediction process of track geometry condition

This study considers the prognostic of the changing standard deviation of an S&C section. The deteriorated track geometry is represented by the calculated standard deviation of the changing track geometry. The first step in this study was to estimate the time when the standard deviation (*r*) will reach the maintenance limit of 1.8 mm (*true end-of-life*). This was done by fitting the ORE model to the complete set of track degradation data for the four S&Cs. The second step was to predict the track degradation for S&Cs 1 and 3 using the regression and particle filter methods. S&C No. 1 was chosen because it had different properties than the other S&Cs, and S&C No. 3 was chosen because it had more limited data than the other S&Cs. This four case studies illustrate how the data-driven regression-based approach and model-based approach perform when less data is available for estimating the *RUL*. The initial distribution of the ORE model parameters that were used in the particle filter method were assumed to be normally distributed, taking the mean and standard deviation of the estimated parameters for all 4 S&Cs. The results of the particle filter method and the regression method were then compared with the estimated true end-of-life.

3.2.4. Results

3.2.4.1. Proposed formalization

In the scenario description, the asset in the interest is the Switches and Crossings. In case of No Failure condition, the forecasting will provide information to TMS whether it is possible to lock the next train routes (colored in orange and purple) within the time interval as shown in Figure 3.4. Since there is no failure, it is safe to send the train to that route with a probability of failure for switch S21/S31. It is based on the probability of failure calculated from the particle-filter based approach of each of the switches and it is dependent on the type of switches, reliability, remaining useful life and supportability.

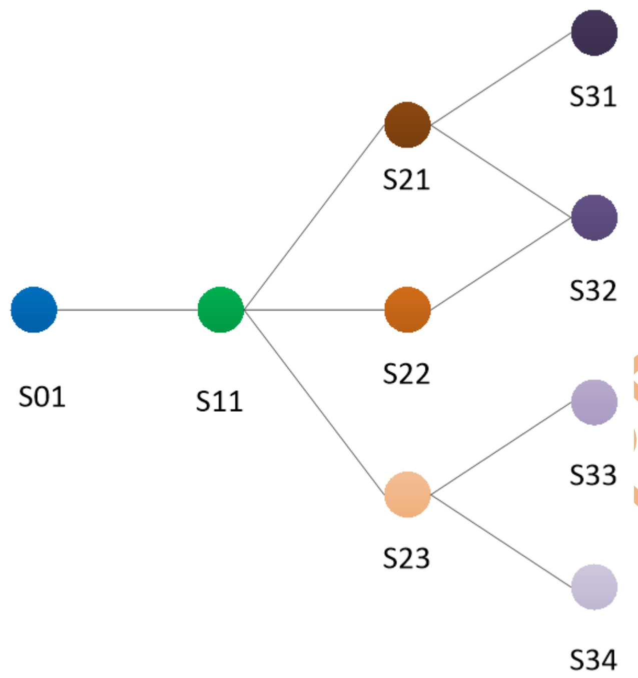


Figure 3.4: Switches and Crossings for forecasting graphical view

3.2.4.2. Proposed Solution

The assets chosen in this scenario, the switches and crossings, are the repairable systems. The probabilities can be estimated by using particle-filter based approach for rerouting the traffic by TMS. The developed models based on the particle-filter based approach are useful to predict the forecast for future condition. Using of this information can be used for following applications [65]:

- setting maintenance schedules;
- making provisions for spare parts;
- assuring suitable performance.

3.2.4.3. Preliminary Results

For the analysis carried out in forecasting, the short-range longitudinal level of both tracks from Optram database is considered. There are couple of reasons why short-wave longitudinal level statistics is used as indicator for maintenance modelling, planning and scheduling:

1. there are many scientific works/standards on longitudinal level to compare the result of the forecasting model with;
2. tamping is one of the largest railway maintenance works in terms of cost and possession, and the main trigger for tamping is longitudinal level status;

3. from historical record, longitudinal level changes faster than other geometry parameters except in few instances that can be considered as “black spot”. Settlement in the vertical direction is higher than other direction due to higher vertical force from moving trains;
4. short wave signal is used because it’s the signal component that shows the condition of the ballast.

Figure 3.5 shows the 18-months prognostic results for four S&C’s. The graph compares the regression method to the particle filter method for the predicted *RUL* 18 months prior to the estimated end-of-life. The 1.8 mm *SDH* threshold is indicated by a solid line, and the estimated end of life is marked with a square.

Figure 3.6 shows the histogram of the particle filter *RUL*. The histograms show the distributions of the 18-month prediction generated by the particle filter. For comparison, the probabilistic results cannot be produced by the regression method since the distribution deviates from a normal distribution. The prognostic performance of the particle filter and regression method after each measurement (*SDH*) are shown in Figure 3.7 for the four S&Cs. For S&C No. 1, 2, 4 the two methods show similar performance, but for S&C No. 3, the particle filter performs better for the long-term prognostics. The figures show the deviation of the prognostic compared to the elapsed time. The slope of the curve represents the remaining time to ‘‘failure” (*RUL*) for each point in time.

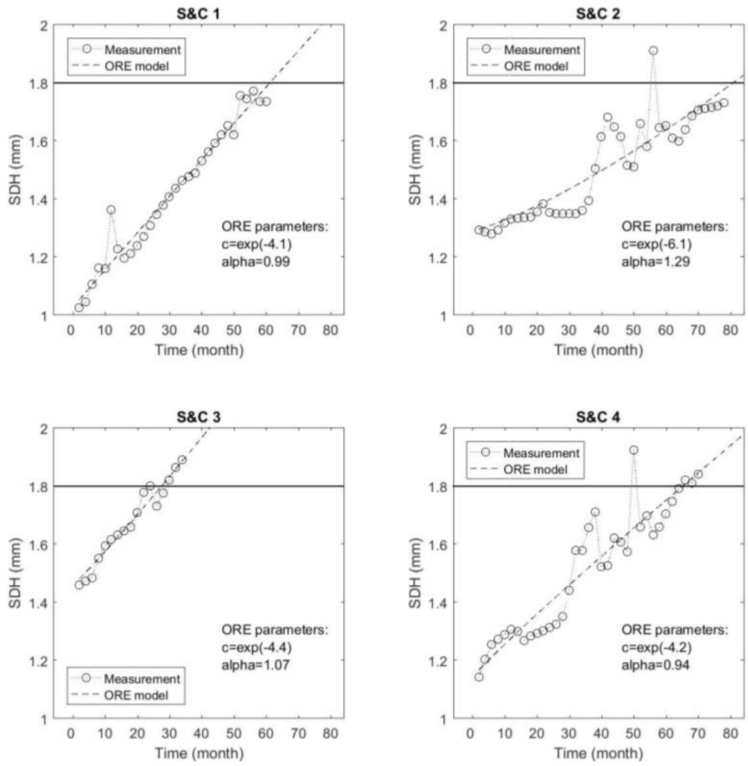


Figure 3.5: Prognostics performed 18 months prior to the true RUL for S&C No. 1 (a) S&C No. 3 (b), S&C No. 2 (c), and S&C No. 4 (d) using linear regression and particle filter state estimation

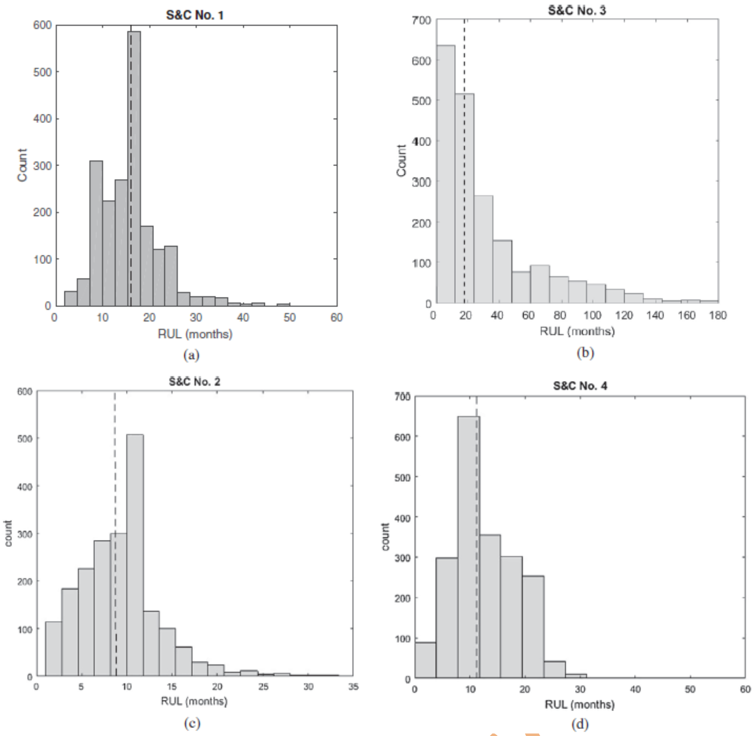


Figure 3.6: The particle distribution (grey bars) for the particle filter prognostic of the S&C No. 1 (a), S&C No. 3 (b), S&C No. 2 (c), and S&C No. 4 (d)The dashed line shows the median value, and the Y-axis shows the count

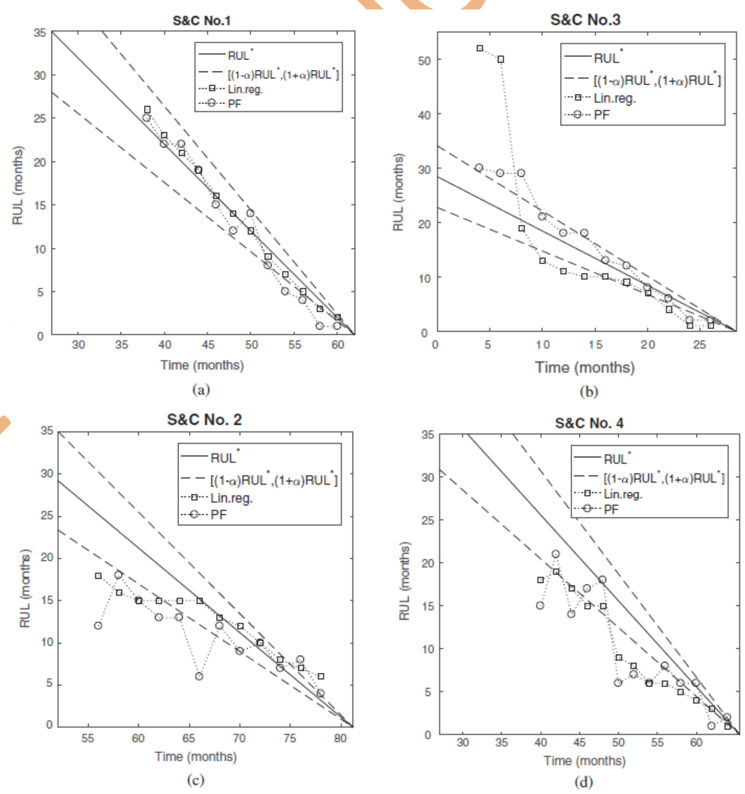


Figure 3.7: Comparison of the RUL predictions from regression and state estimation using particle filters for S&C No. 1 (a), S&C No. 3 (b), S&C No. 2 (c), and S&C No. 4 (d) The allowable error bound value a is set to 0.2

3.2.5. Analysis and Discussion

Forecasting of track geometry behavior is necessary to ensure the best performance of the S&C as it is the critical component. The degradation of the track may lead not only for the train derailment but also degradation of other subcomponents, the later still needs to be researched. Tamping is one of the important maintenance actions that is carried out on the ballasted railway track systems. This action will ensure the restoration of the degraded track geometry state to its working state. To increase the working condition of S&C, the tamping must be planned prior. This forecast, in general with TRV, is often made 18 months before the actual event and gives the expected time remaining before maintenance intervention is needed. By adopting a particle filter-based approach in this scenario, the forecast results include not only the most likely time of the failure but also the prediction distribution, which allows a risk-based maintenance decision to be made.

The true *RUL* is difficult to establish from the standard deviation of the track degradation data due to the fluctuations of the standard deviation over time. Therefore, the true *RUL* is generated using the complete set of data. In this study, four S&Cs of same type (EV-UIC60-760-1:15) from same track section 414 were used to estimate the distribution of the model parameters; the mean value and standard deviation were used to set the initial distributions. This is an important step for implementing the suggested method.

In Figure 3.7, S&C No. 3 shows that the particle filter methods have improved ability to predict the *RUL* when the amount of data is limited. When more data are available, such as for S&C No. 1, both methods show less prediction error. In the four example cases, the particle filter approach performs better than or as good as the regression method. This shows that the model is verified with the existing data. This can be further validated with time series data or data from another S&Cs.

By grouping S&Cs, with common degradation behaviors, the estimation of the model parameters could be improved with a reduced uncertainty in the prediction as a result. In this scenario, the traffic volume (million gross tons) and the axle load of the rolling stock could change over time. The load could also be associated with uncertainty based on the uncertainties in the cargo loads. This uncertainty can be represented in the load variable and can propagate along with all the other uncertainties and measurement noise to form a probabilistic result of the *RUL* prediction.

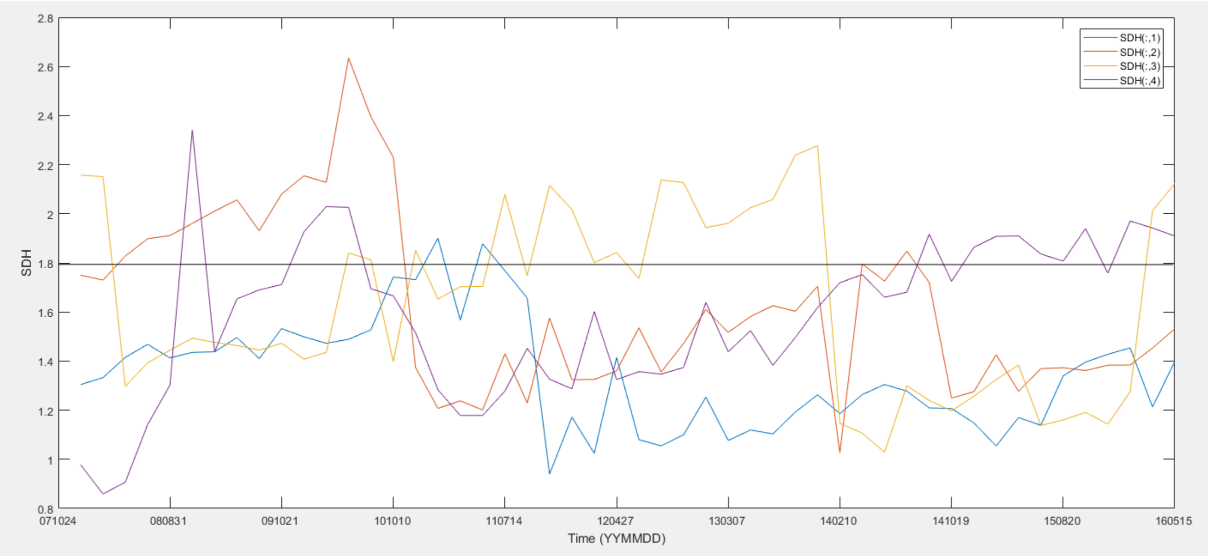


Figure 3.8: Track geometry behavior of S&Cs over the time

The track geometry behavior for the continuous tamping actions is illustrated in Figure 3.5. The track geometry behavior of the selected S&Cs is shown in Figure 3.8. After the tamping actions carried out on the individual S&Cs, the parameters of the model will be recalculated and learned from the previous behavior of S&C to predict the future condition. TMS can utilize the forecasting status of the track geometry to plan for future maintenance actions or speed restriction to increase the life or stop the train to that route. The work will be extended Shift2Rail for:

- modelling from the parametric estimation need to be selected from D6.3;
- including other parameters such as S&C type, age, measurement car, etc.;
- validated with different sets of data.

3.3 UPORTO/EVOLEO/IP Scenario

3.3.1. Summary of Scenario "UPORTO/EVOLEO/IP" (D9.3)

Title	**Study lateral and vertical wheel-rail contact forces of running trains for nowcasting and forecasting the risk of derailment**
Organisations Involved	• Infraestruturas de Portugal (IP) • University of Porto (UPORTO) • Evoleo Technologies (EVOLEO) • Virtual Vehicle (ViF) • University of Genoa (UNIGE)
Objective(s) of the scenario	This proposed NC/FC scenario develops around two objectives: • development of a methodology able to provide an indication of the risk of derailment of a running train by exploiting data-driven methodologies, based on information about the train, the track conditions and the weather; • assessment of the possibility to substitute physical white-box models of contact forces between the wheels of a running train and the rails with data-driven ones. In particular, this scenario focuses on developing a data-driven system able to estimate the lateral (Y) and vertical (Q) wheel-rail contact forces, as well as their ratio (Y/Q), which represent the most relevant parameters specified in the European Standard CSN EN 14363 (2016) as indicators of running safety for the assessment of the derailment risk. Moreover, this scenario aims at studying and evaluating the possibility of NC/FC the risk of derailment based on fast calculation data-driven methodologies with sufficient prediction quality. Indeed, several physical white-box models, able to estimate accurately Y and Q wheel-rail contact forces and validated on real data, are reported in literature. However, using these white-box models to simulate the behavior of a running train usually requires high computational resources and large amounts of time, which make them unsuitable for time-critical applications (e.g. for nowcasting purposes). For this reason, this scenario aims at substituting white-box models with data-driven ones, which instead are capable of shrinking into a powerful and lightning-fast model the information content hidden in data. Indeed, although data-driven models require large computational resources and time in order to be built, their usage is straightforward as the production of the output is a matter of simple calculations. A further advantage is that the final user of data-driven models (e.g. TMS) does not need the "train model" (Finite Elements or Multi Body System) and the corresponding simulation software.
Relationship with TMS and/or maintenance	Data-driven models could be used to provide useful information for supporting the decision process of TMS/maintenance actions, such as setting speed reductions, closing down a particular line or performing track geometry corrections, so to prevent derailments. More specifically, NC could provide to TMS the current risk of derailment for a specific vehicle for different vehicle speeds and for different wind speeds and directions, therefore expressing the risk of derailment as a function of the different inputs. Long-term FC, instead, could provide to maintenance the remaining time until a limit of risk of derailment is exceeded.
Description of the scenario	Since the physical modelling approach has some major drawbacks that have to be carefully considered (difficult construction, calibration and validation, and unsuitability for time-critical applications), data-driven models represent an appealing alternative because they can automatically infer the relationship between some real input data, which is representative of a certain system behaviour or phenomena, to some real output data, which is usually difficult to measure or that has to be predicted at future instants. Additionally, data-driven models are capable of shrinking the information content hidden in data into a powerful and lightning-

	fast model that can respond in near real-time to previously unseen inputs. In order to build data-driven models for the estimation and prediction of the lateral and vertical wheel-rail contact forces and their ratio Y/Q, it is necessary to collect data related to the main factors affecting the run of a train, namely the vehicle type (e.g. passenger trains, freight trains, etc.), the vehicle conditions (e.g. unbalanced loading, defected wheels, etc.), the track assets conditions (e.g. bridges, tunnels, etc.) and the operating/environmental conditions (e.g. running speed, wind, etc.). In the context of this scenario, a simplified approach is proposed, so to limit the complexity of the scenario, and because of the problem of the availability of data. The proposed approach considers a single type of train (i.e. the Alfa Pendular train), a particular line in Portugal (i.e. the Portuguese Railway Line between Porto and Lisbon, for which real track condition data is available), and the presence of wind as a single (but most representative for a train run) weather variable. Moreover, since real measurements of Y and Q forces are not available, the creation of the NC/FC models will be based on synthetic data of wheel-rail contact forces generated through the validated physical Finite Elements model of the Alfa Pendular train. This approach, if performed correctly, can be equally considered valid and meaningful. Furthermore, as soon as real data related to wheel-rail contact forces will be available, the application of this approach will be straightforward. The fundamental steps for the completion of this scenario are listed here below: 1. collecting/designing input data; 2. performing simulations with physical Finite Elements model of the Alfa Pendular train with input data of step (1), so to generate output data, namely Y and Q wheel-rail contact forces values; 3. building the data-driven model through state-of-the-art machine learning algorithms that can infer the relationship between the same input data of step (1), and lateral and vertical forces generated at step (2) through simulations. Once the data-driven models will be available, they could be used to provide useful information for supporting the decision process of TMS/maintenance actions.
Data exploited for the scenario	**Track data** The track data for the use in the NC scenario will be provided by IP, and consist of measured track irregularities and layout parameters (e.g. longitudinal level, alignment, curvature, etc.). **Vehicle data** The vehicle data that comprise the vehicle information (e.g. dynamic characteristics for the FE model) and the operational information (e.g. vehicle running speed, loading, tonnage) will be provided by IP and UPORTO. In particular, the vehicle model to be used in the NC/FC analysis will be the Alfa Pendular train, which is a passenger train whose dynamic characteristics have been calibrated by UPORTO in the last years. **Environmental (Wind) data** Since this scenario will be based on simulations of runs of a train under different wind conditions, at the first stage this kind of data will be simulated in order to apply in an easier way different wind conditions to the simulated running trains.

Table 3.3: Tabular Description for Scenario by IP/UPORTO/EVOLEO/ViF/UNIGE

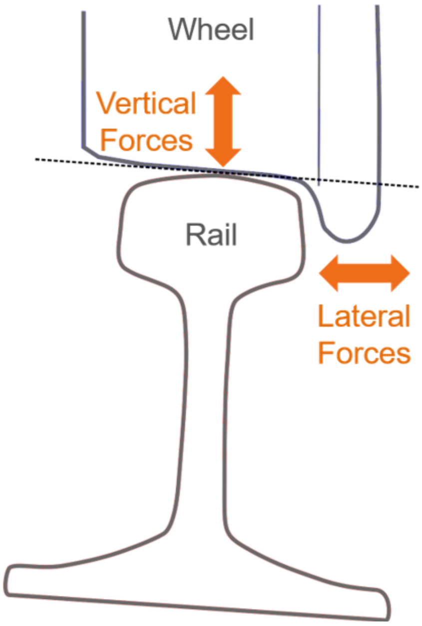


Figure 3.9: Wheel-rail lateral and vertical contact forces

The main question to answer within the In2Rail project: "Is it possible to FC the risk of derailment by the aim of fast calculation methods with sufficient prediction quality as a decision basis for TMS/maintenance?"

3.3.2 Data description for Forecasting scenario

The risk of derailment of a specific vehicle depends on several parameters and operating conditions of the vehicle, of the track and of the environment. To fulfil the requirements of a fast calculation time, it is necessary to limit the number of input parameters and to consider several parameters and operating conditions as constant over time. The following input parameters (current asset status and historical data) are at least necessary to predict the risk of derailment of a specific vehicle for this scenario:

- required historical input data along a specific track;
- additional historical data as an information (if available) but not used by the FC method.

The input parameters are explained in the following chapter.

3.3.2.1. Input parameters

- required historical input data along a specific track:
  - track geometry (alignment left/right rail, longitudinal level left/right rail, track gauge, twist and cross level),
  - track layout (curvature),
  - vehicle conditions (speed),
  - environmental conditions (wind speed and wind direction);
- additional historical data as an information (if available) but not used by the FC method:
  - track geometry (inclination, rail profiles),
  - track layout (cant),

- track structure (bridge – surrogate model parameters),
- vehicle type (geometry, masses/inertia, stiffness/damping characteristics, …),
- vehicle conditions (loading, wheel profiles, …),
- environmental conditions (try/wet, …).

With the constant need to increase the vehicle speed, the consideration of crosswinds as an input parameter is imperative for the risk assessment. Moreover, as stated by [66], in modern trains the leading vehicle is a relatively light motor coach rather than the former heavy locomotive, which represents a mass reduction, increasing instability to crosswinds.

In addition, wind-train interaction studies are mandatory for vehicle homologation and are one of the main concerns for the interoperability project of the European railway network [67],[68].

Furthermore, the climate change has been aggravating the phenomenon, amplifying its intensity most in areas alongshore.

Considering the statements above and the fact that the Scenario is part of the North Line, which is mostly in open fields near the coast, the wind is a key factor and an important input that has to be evaluated for TMS decisions.

The input parameter 'wind speed' plays an important role for Infraestructuras Portugal due to the considered coast line. This input parameter can be measured and also predicted by a meteorological service and is therefore used to proof the proposed FC methodology concept. The consideration of further (environmental) input parameters is discussed in Chapter 3.3.6.

3.3.2.2. Output parameters and its relation to TMS

The output parameter is the predicted risk of derailment of a specific vehicle along a specific track. According to the European Standard CSN EN 14363 (2016), the most relevant parameter as an indicator for the risk of derailment is the ratio between the lateral (Y) and vertical (Q) wheel-rail contact forces. The method is able to provide the following information levels based on these Y/Q-ratios as a function of the (controllable) vehicle speed for TMS:

- predicted (Y/Q)max for a complete track section;
- predicted (Y/Q)max for track subsections (e.g. with fixed length of 50 m);
- predicted (Y/Q) for every point along the track (e.g. 25 cm).

These results provide TMS/Maintenance a decision basis for e.g. the following actions:

- speed reductions;
- closing down a particular line / track section;
- performing track geometry corrections.

3.3.2.3. Uncertainties of the input parameters and forecast

The output parameter 'risk of derailment' is provided as a distribution function for each point (e.g. every 25 cm) along the track. The distribution function is a result of the prediction methodology which includes several uncertainties of the input parameters. The following list shows the considered uncertainties of these input parameters:

- track geometry:
  - uncertainty (degradation) will analysed and considered within the FC method. For short time prediction (e.g. one week in the future), the track geometry can be considered as constant;
- track layout:
  - it is considered as constant without uncertainties;
- vehicle conditions:
  - vehicle parameters are considered as constant for the FC method,
  - vehicle speed (controllable operating condition) is an independent FC parameter and is considered by different speed levels within the FC method;
- environmental conditions:
  - the forecast of wind speed and wind direction (from a meteorological service) provided as a distribution function is required as an input for the FC method. If there is only the information of the forecasted mean value of the wind speed is available, the user can assume value for the standard deviation of a normal distribution to consider the uncertainty. Details of the assumptions are described in the following document chapters.

Uncertainties of other parameters listed in Chapter 3.3.2.1 are assumed as constant over time and considered as input parameters with no uncertainties. This assumption is made to proof the concept of this FC method.

3.3.2.4. Ranking of the parameters

The ranking of the parameters is not applicable for this scenario.

3.3.3 Methods for prediction

3.3.3.1. State of the art

3.3.3.1.1. Data-Driven Model

The state of the art of data-driven models for prediction a future asset status is described in detail in Chapters 5.1 and 5.2 of deliverable D9.3 (especially in the subsection 5.2.1.2.1 "Time Series Forecast").

3.3.3.1.2. Physical Model

The following steps describe the principal application concept of the developed FC method for this scenario:

- definition of the desired prediction horizon for the future assets status;
- definition of suitable vehicle speed classes (controllable by TMS) for the specific track section;

- forecast of the input parameters of the track geometry by applying the ViF degradation model;
- forecast of the input parameters of the wind (speed and direction) by a meteorological service;
- generation of several forecasted input data sets to consider the input parameter uncertainties;
- application of the NC method with every forecasted input data set;
- post-processing of the predicted Y/Q ratio along the track to calculate a suitable decision basis for TMS/Maintenance.

Figure 3.10 shows an overview of the developed method.

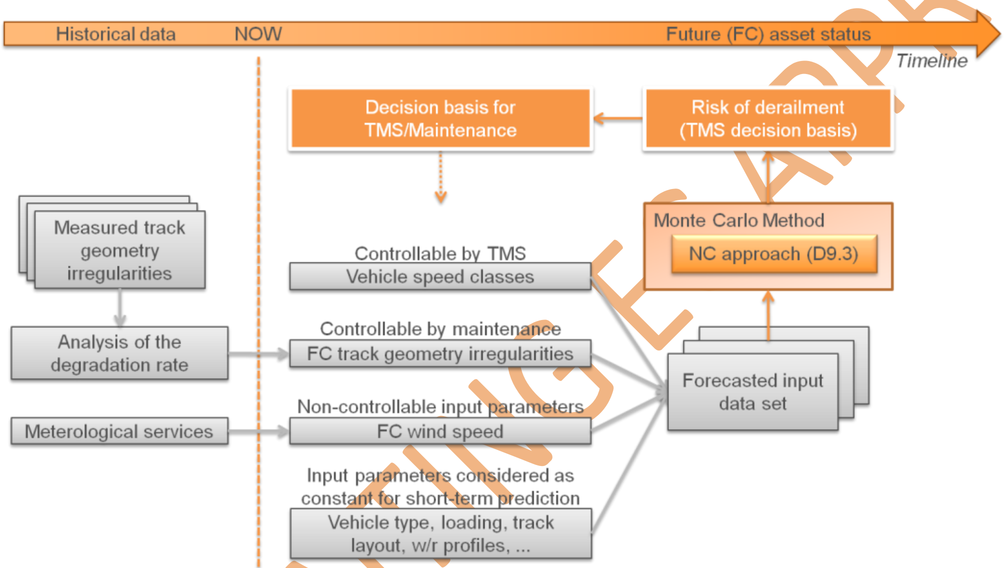


Figure 3.10: Methodology for UPORTO/EVOLEO/IP Scenario

3.3.3.2. ViF track geometry degradation model

One of the most challenging tasks within this method is the prediction of the input parameters. The track geometry has a significant influence on the risk of derailment. For short-term predictions, the degradation of the track geometry can generally be neglected. Usually, the track geometry is measured e.g. two times per year and this could lead to a leak even for short-term predictions. Therefore, a track geometry degradation model was developed at Virtual Vehicle in the past – called *ViF TGD* model. The concept of this model can also be applied at track geometry data of this scenario to proof the FC concept.

The basic idea of the ViF TGD model is the analysis of the degradation rate in the wavelength domain within several wavelength-bands. Figure 3.11 shows the analysis of the track geometry degradation for data with a measurement interval of two weeks (source data: not public and not provided by an In2Rail WP9 partner). These short measurement intervals provide the possibility to calculate the degradation rate with a high accuracy.

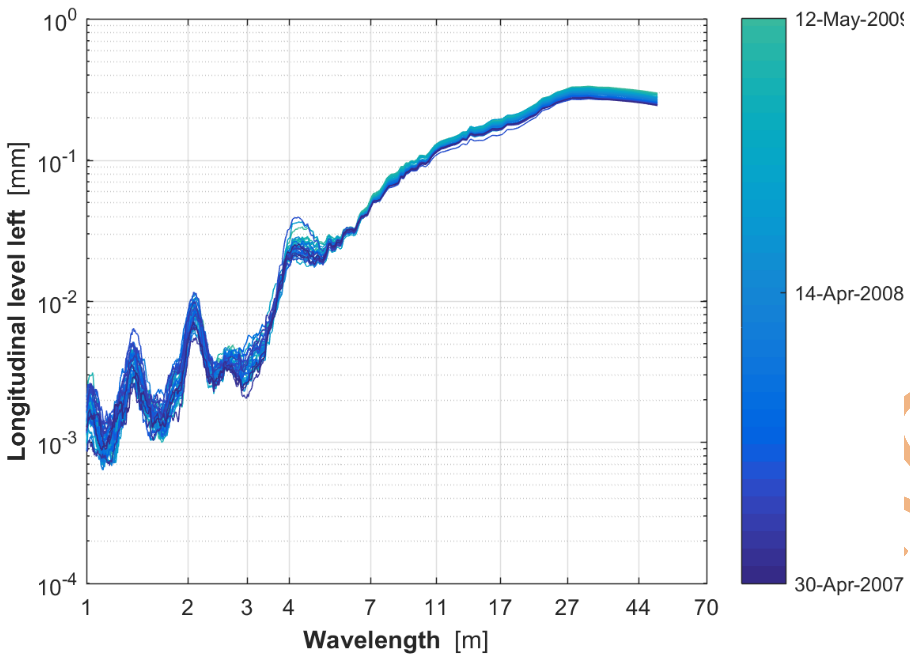


Figure 3.11: Wavelength spectrum of track geometry over time

The growth rate distribution within each wavelength band is calculated. Based on these results, a distribution function can be fitted (e.g. gamma distribution) and can be used for the prediction of the track geometry. As an example, Figure 3.12 shows the distribution for the wavelength band 17 – 27 m.

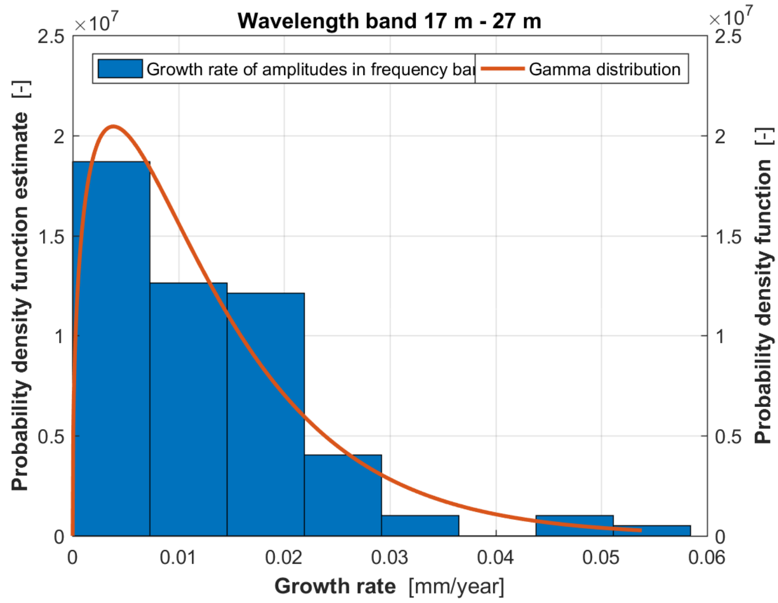


Figure 3.12: Growth rate distribution of track geometry of a specific wavelength band

Figure 3.13 shows a validation of the ViF TGD model by comparing the predicted track geometry (after 365 days) compared with available measurement data (after 364 days). It can be seen that the comparison between the predicted values of the longitudinal level left (zl_P) and the 'true' measurement values (zl_T) shows a high correlation coefficient of R^2 = 98%.

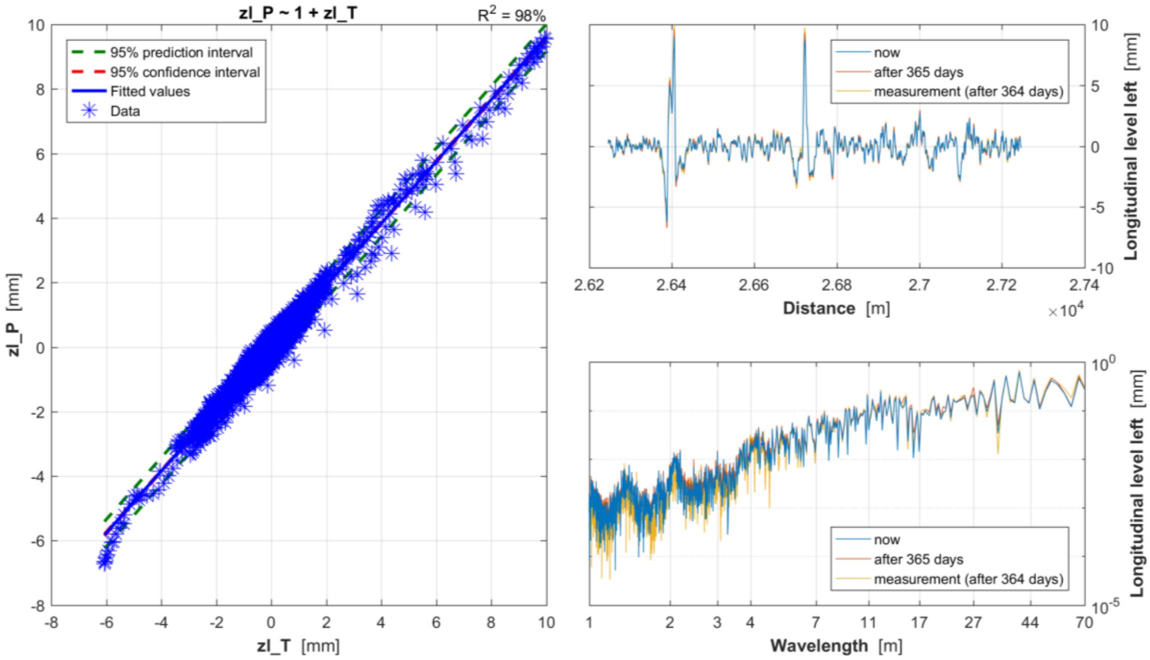


Figure 3.13: Validation plots of the ViF Track Geometry Degradation model

3.3.3.3. Assumptions for environmental conditions and controllable parameters

For short-term predictions, the environmental conditions for wind speed (and direction) can be delivered by a meteorological service. For long-term predictions, the distribution could be generated by historical meteorological data.

Figure 3.14 shows the distribution function for a wind speed with a mean value of v_Wind = 60 km/h. The probability distribution function is divided into an arbitrary number of classes (e.g. N = 15) and one sample is randomly selected for each class. This method guarantees that also the boundary classes of the distribution assumption are considered within the FC method.

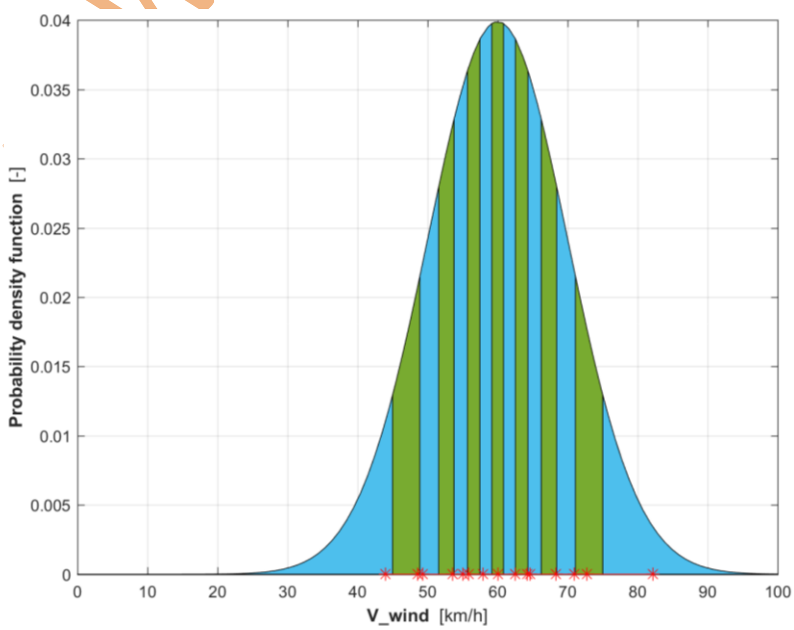


Figure 3.14: Non-controllable input parameter "wind speed" as a probability density function

As an example of a controllable parameter for TMS, the vehicle speed is divided in appropriate classes e.g. v_Vehicle = [140, 180, 220] km/h.

3.3.3.4. FC method based on predicted input parameters

The FC method provides the possibility to select the desired number of variations for each controllable parameter class. For this specific scenario example, the following controllable and uncontrollable input parameter samples for N = 15 variations are taken into account:

- vehicle speed (controllable): 140, 180 and 220 km/h ;
- track geometry (uncontrollable by TMS): 15 predicted data sets for longitudinal level left and right and alignment left and right (prediction horizon of 365 days) ;
- wind speed (uncontrollable): 15 samples out of a normal distribution with μ = 60 km/h, σ = 10 km/h.

For the proof of concept, further parameters (e.g. curvature, wheel/rail profiles, condition of friction, wind direction…) are considered as constant along the track as well as over time.

In the next step of the FC algorithm, the NC method described in the Deliverable D9.3 in Chpater 7.2.2 is applied at every input data set. The output of the method is the predicted Y/Q ratio variation along the track (every 25 cm) for each controllable input parameter.

3.3.3.5. Post-processing of the FC results

The predicted Y/Q ratio is evaluated according to the EN 14363 (see Deliverable D9.3 Chapter 11.3.2) to assess the risk of derailment. The FC method provides different aggregation levels as a decision basis for TMS:

- highest level of information: mean/median/max values of the EN 14363 evaluated Y/Q ratio variation at every track location (e.g. 25 cm) for each vehicle speed class ;
- medium level of information: 95% percentile values of the EN 14363 evaluated Y/Q ratio variation within predefined subsections (e.g. 50 m) ;
- minimum level of information: 95% percentile values of the EN 14363 evaluated Y/Q ratio variation for the complete considered track section (e.g. 3 km).

3.3.4 Results

Figure 3.15 shows the result of the EN 14363 evaluated Y/Q ratio along a specific subsection between 840 – 870 m. The colour represents the (controllable) vehicle speed levels. The min/mean/max values of the Y/Q ratio distribution are visualized.

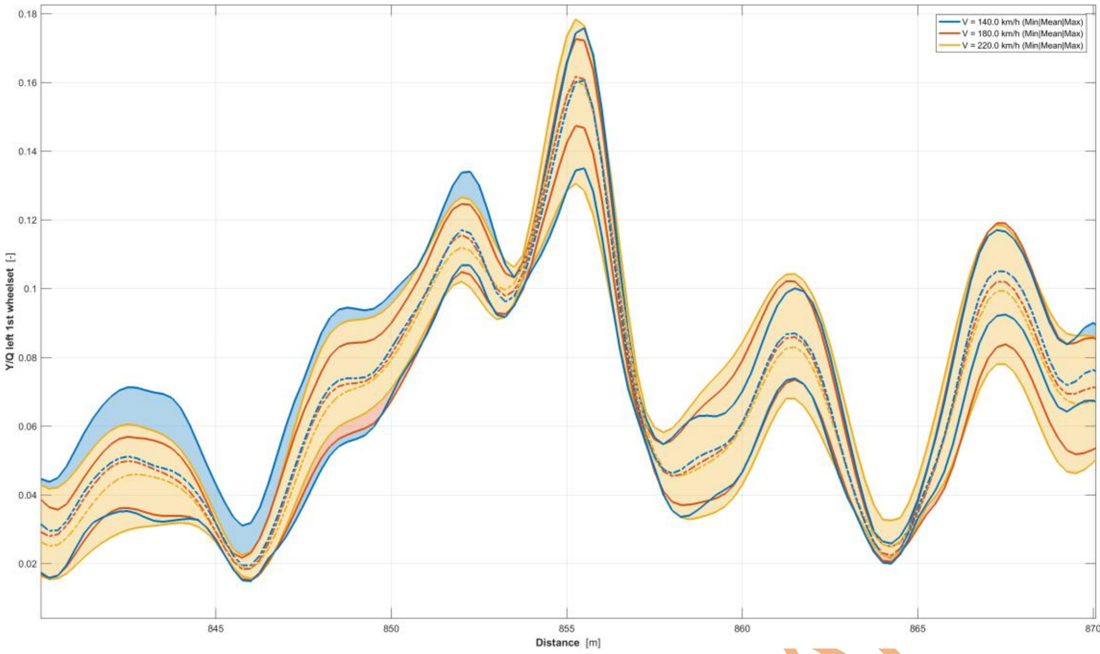


Figure 3.15: FC of Y/Q_2m ratio along the track – zoomed between position 840-870 m

For each subsection (50 m), the 95% percentile values of the Y/Q ratio variations are calculated. The result can be seen in the following figure. Additionally, the maximum values of the 95% percentile values of the Y/Q ratio variations of the complete track section are visualized in the figure legend.

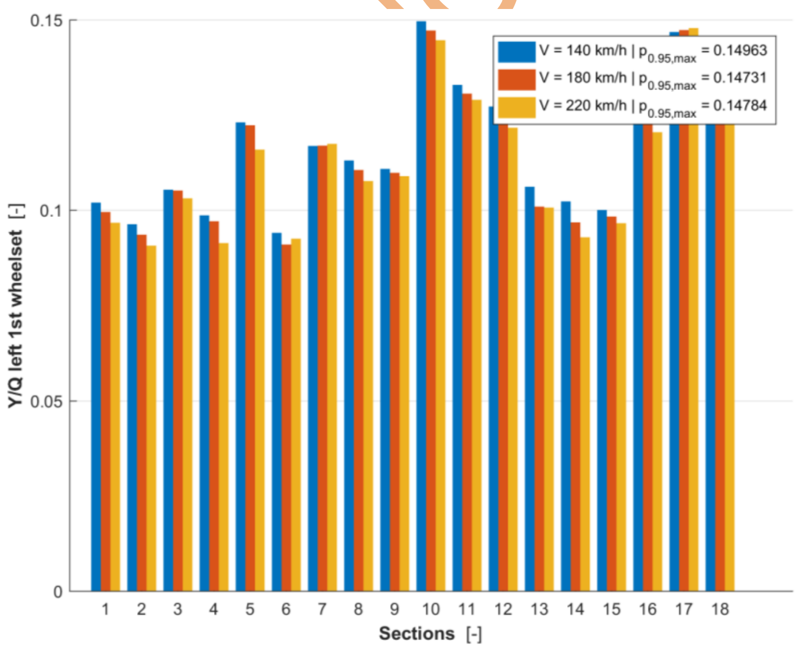


Figure 3.16: FC of Y/Q_2m ratio for each 50-m-subsection

3.3.5 Analysis

The considered scenario example is based on a vehicle running over a straight track. As expected, the limit value of the Y/Q ratio (0.8) was not reached. The low influence of the vehicle speed is also an effect of the straight track in combination with the selected varied input parameters.

Further investigations of this scenario should also consider a curved track as well as the variation of further input parameters e.g. wind direction to increase the variation of the Y/Q ratio.

3.3.6 Discussion

The proposed FC method provides TMS a fast-calculated decision basis regarding the risk of derailment along the track as a function of several input parameters. The advantages of this method can be summarized as followed:

- very short calculation time ;
- very high level of detail (risk of derailment for every track point – e.g. 25 cm) ;
- consideration of effects of input parameter uncertainties (including track geometry degradation which also provides a decision basis for maintenance actions) ;
- possibility to include wheel/rail force measurements in the training phase ;
- possibility to include a variation of more input parameters in the training phase.

The consideration of further input parameters can be done in the same way as shown by the already considered parameters. In a first step, the parameter has to be included as an input parameter of the NC method (e.g. estimated/measured data of the coefficient of friction). In a second step, this input parameter has to be predicted or assumed for the future (e.g. distribution function for the coefficient of friction at the prediction horizon). In a third step, the FC method considers this new parameter due to the updated input parameter set.

The prediction quality of the proposed FC method depends on several factors. The key is the estimation quality of the NC method proposed and described in the Deliverable D9.3 in Chapter 7.2.2. This NC method shows a high prediction quality for the investigated example data of a straight track (see Deliverable D9.3 in Chapter 7.2.3.3). As mentioned in the previous chapter, the results for curved tracks as well as for the variation of further input parameters e.g. wind direction should be analysed to access the prediction quality. Furthermore, the prediction quality also depends on the vehicle model quality which is used to generate the reference simulation data for the training phase. Nevertheless, if measurements of wheel/rail forces are available in combination with the according input parameter values, the FC method can be extended by considering this data.

3.4 SR/UNIGE Scenario

3.4.1 Summary of Scenario "SR/UNIGE"

In this scenario, two cases are investigated. For the sake of simplicity, this proposed scenario has been divided into two different parts, of which the first one is described in Table 3.4, and the second one is described in Table 3.5.

Title	**Prediction of time to restoration for different assets and different failures based on maintenance/repair reports**
Organisations Involved	• Strukton Rail Netherlands (SR) • University of Genoa (UNIGE)
Objective(s) of the scenario	An analysis will be carried out about the "time to restoration" (or "repair time") needed to restore the asset to a proper functional state after a specific failure occurred. The analysis will try to develop a forecasting methodology able to estimate in advance the precise repair time once a problem on an infrastructure asset arises. Therefore, this scenario aims at designing, implementing, testing and, at a future stage, validating a set of predictive models for forecasting purposes, based on data provided by SR about maintenance/repair actions and weather data.
Relationship with TMS and/or maintenance	Since in this case the output of data-driven models is the time required to complete a repair action, it could be used by the TMS to estimate the time the asset will be unavailable for exploitation. For example, TMS could exploit this information for planning and managing line possessions in an informed way. Estimation of the time to restoration for the less urgent (plannable repairs) incidents can be used by maintenance department for better estimation of the planning maintenance slots.
Description of the scenario	Every time an infrastructure asset is affected by a failure, it is clear that this will affect not only the single asset functional behaviour, but also the normal execution of railway operations. For this reason, the objective of the third analysis is to estimate the time to restoration for future (planned) and urgent maintenance actions by looking at the past maintenance reports, correlated to the different assets and different types of failures. The predictive models that will be designed will be able to exploit the knowledge enclosed into maintenance reports so to predict the time needed to complete a maintenance action over an asset in order to restore its functional status. Moreover, historical weather conditions data will be included in the analysis in order to take into account of the atmospheric factors affecting railway maintenance/repair operations (e.g. fog reducing visibility).
Data exploited for the scenario	**Weather condition:** data retrieved from the Royal Netherlands Meteorological Institute (KNMI) **Maintenance/repair actions:** historical datasets regarding the maintenance/repair activities (including their duration) will be provided by Strukton Rail. This data is collected by Strukton Rail but commissioned by the rail infrastructure manager. Data is originally stored in a Maintenance Management System. This information could be provided by Asset Manager. **Failures:** historical datasets regarding the recorded failures will be provided by Strukton Rail.

Table 3.4: Tabular Description for Scenario by SR/UNIGE (1)

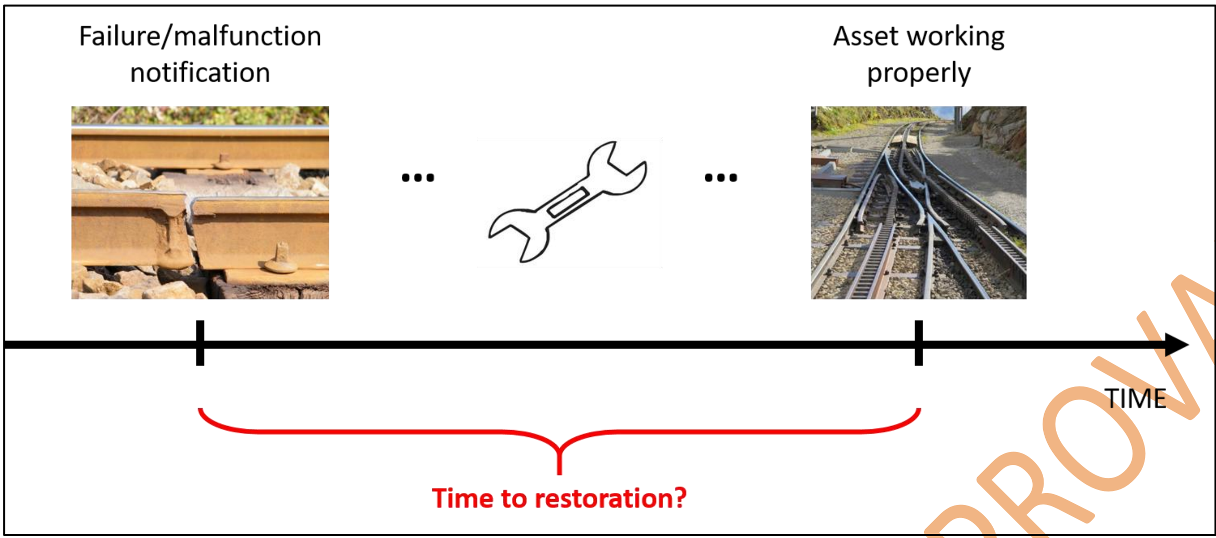


Figure 3.17: Pictorial representation of scenario by SR/UNIGE (1)

Title	**Data mining correlation and influence of maintenance/repair actions and weather conditions on railway assets**
Organisations Involved	• Strukton Rail Netherlands (SR) • University of Genoa (UNIGE)
Objective(s) of the scenario	The main objective of this scenario is to forecast possible failures of assets based on the correlation of past asset failures and past weather conditions or maintenance actions, in particular considering a set of different infrastructure assets selected as the most relevant ones from the TMS perspective (see Deliverable 9.1 of the In2Rail project). Therefore, this scenario aims at designing, implementing, testing and, at a future stage, validating a set of predictive models for forecasting purposes, based on data provided by SR about maintenance/repair actions, weather and failures. The problems that will be investigated (and for which one or more forecasting models will be developed) are: 1. Correlation and influence of executed maintenance/repair actions on failures. 2. Correlation and influence of the weather conditions on failures.
Relationship with TMS and/or maintenance	The information outputted by the models can be very useful because it could be used by the TMS to reroute trains through safer paths, minimizing the risks of any problem. Moreover, the same output could be used by the maintenance department in order to prevent possible problems and to schedule proper maintenance actions that could prevent additional or worst problems.
Description of the scenario	Every time an infrastructure asset is affected by a failure, it is clear that this will affect not only the single asset functional behaviour, but also the normal execution of railway operations. The functional behaviour of railway infrastructure assets degrades for many different reasons: age, extreme weather conditions, heavy loads, and the like. Additionally, problems can be introduced unknowingly by performing maintenance actions, for example by a simple human error or as a reaction of the system to changes made on an object. For these reasons, this scenario aims at investigating two among all the factors that might affect the degradation of assets, i.e. maintenance/repair actions and weather conditions, and at designing and developing new forecasting methodologies by exploiting data-driven predictive techniques. A set of predictive models able to forecast the probability of failures for a particular asset will be developed based on the data provided by SR about historical weather conditions, performed maintenance/repair actions and failures.
Data exploited for the scenario	**Weather condition:** data retrieved from the Royal Netherlands Meteorological Institute (KNMI) **Maintenance/repair actions:** historical datasets regarding the maintenance/repair activities (including their duration) will be provided by Strukton Rail. This data is collected by Strukton Rail but commissioned by the rail infrastructure manager. Data is originally stored in a Maintenance Management System. This information could be provided by Asset Manager. **Failures:** historical datasets regarding the recorded failures will be provided by Strukton Rail.

Table 3.5: Tabular Description for Scenario by SR/UNIGE (2)

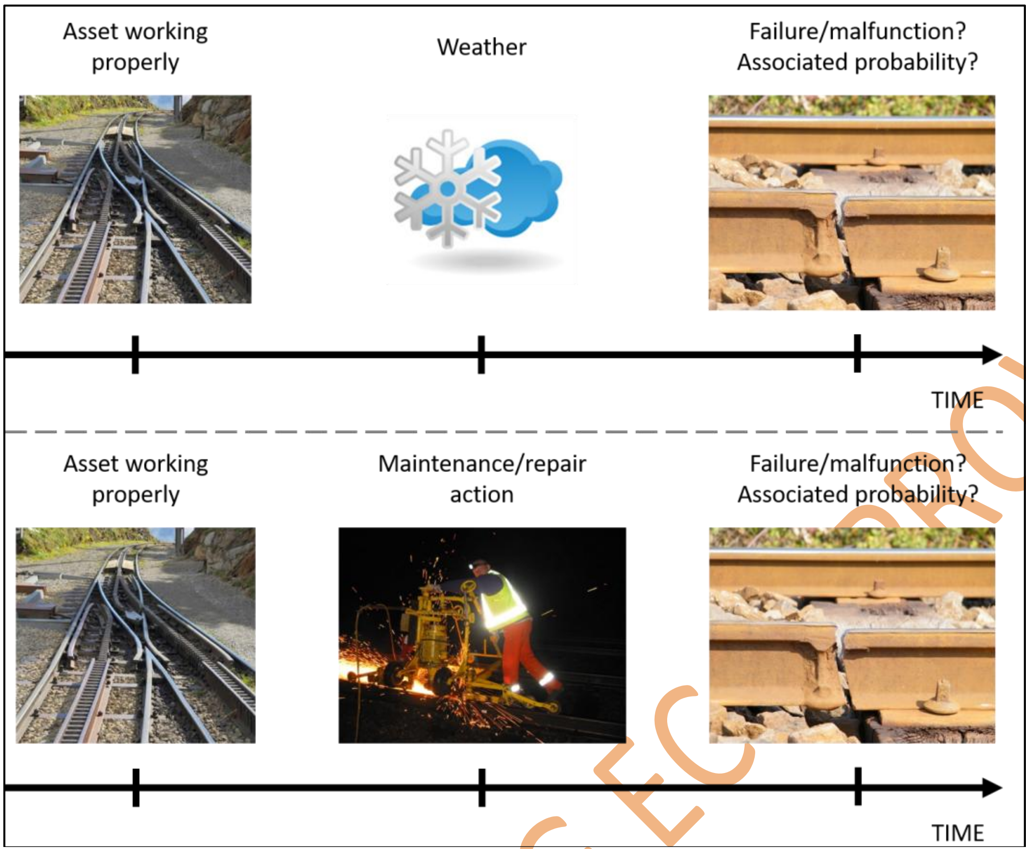


Figure 3.18: Pictorial representation of scenario by SR/UNIGE (2)

3.4.2 Data description for Forecasting scenario

3.4.2.1. Input parameters

For this scenario, mainly three datasets have been exploited:

- **weather condition**: data retrieved from the Royal Netherlands Meteorological Institute (KNMI) ;
- **maintenance/repair actions**: historical datasets regarding the maintenance/repair activities (including their duration) will be provided by Strukton Rail. This data is collected by Strukton Rail but commissioned the rail infrastructure manager. Data is originally stored in a Maintenance Management System. This information could be provided by Asset Manager ;
- **failures**: historical datasets regarding the recorded failures will be provided by Strukton Rail.

Moreover, a novel set of features have been extracted from the data included in the aforementioned datasets:

- temporal intervals features ;
- weather features ;
- past failures features ;
- open failures features ;
- Boolean value indicating whether past failures occurred on the specific asset or not ;
- time from last failure occurred.

All the aforementioned data and the related data sources are described in detail in the next chapters, as well as in the Chapter 6 (Appendix A1).

3.4.2.2. Output parameters and its relation to TMS

The output parameters of this analysis are predictions of the following four main time quantities of interest for new repair interventions:

- **Function Restoration Time**: amount of time needed to complete the intervention and free the railway line ;
- **Travel Time**: amount of time between the reception of the failure notification and the arrival of the mechanics on the location of the asset to be repaired ;
- **Response Time**: time needed by the mechanics to start operating on the asset from the moment in which the failure notification has been received ;
- **Repair Time**: time needed by the mechanics to perform the repair on the asset.

These quantities are strictly related to the maintenance/repair process identified by Strukton, which is described in detail in the next chapters.

These outputs could be used by the TMS to estimate the time the asset will be unavailable for exploitation. For example, TMS could exploit this information for planning and managing line possessions in an informed way. Estimation of the time to restoration for the less urgent (plannable repairs) incidents can be used by maintenance department for better estimation of the planning maintenance slots.

3.4.2.3. Uncertainties of the input parameters and forecast

In this project, we propose to use the Cross-validation as one of the most powerful tool in the context of model selection and error estimation (see [25] and [26] for a general overview of the topic, and references from [27] to [44] for more details). This technique allows assessing how the results of a statistical analysis will generalize to an independent data set, so to estimate how accurately a predictive model will perform in practice. Please refer to the Deliverable 9.3 – "Nowcasting methodologies" for a detailed explanation of the cross-validation technique.

3.4.2.4. Ranking of the parameters

Feature selection and ranking techniques [1] [2] [3] [4] [5] are exploited to assess the importance of each variable from the point of view of a prediction model. The goal of feature selection/ranking is three-fold: (a) improving the prediction performance of the predictive model, (b) providing faster and more cost-effective predictors, and (c) providing a better understanding of the underlying process that generated the data. The general idea related to these techniques has been already introduced in Deliverable 9.3.

For this scenario, a technique based on permutation test [69] [70] [71] (also called randomization test) has been exploited in order to quantify the importance of the feature (input) variables of a dataset by computing the sensitivity of a model to random permutations of feature values. The technique and the related results will be described further in the document.

3.4.3 Methods for prediction

In the context of In2Rail WP9, the goal was to take advantage of data-driven methodologies to build models that can then be used to deduct (predict) the future outputs of a particular system. In order to achieve these objectives, a well-defined methodology based on the famous CRISP-DM [67] [68] standard for data mining has been exploited. All the related information have been already presented in Deliverable 9.3 "Nowcasting methodologies".

For this scenario, we exploited a kernel method [16] belonging to the supervised learning framework [19] [20] that is able to tackle multivariate regression problems [13]. The method will be briefly described later in the chapter.

3.4.4 Results

This chapter reports the work done on the two scenarios proposed by SR/UNIGE (see Chapter 3.4.1) in the context of In2Rail WP9.

The first scenario aims at predicting time to restoration of new repair interventions by analysing records of maintenance/repair actions performed in response to failures detected by the Dutch infrastructure manager. SR shared with UNIGE several years of data, which has been combined with weather data in order to analyse the repair process by looking at the amounts of time required to complete each of its phases.

The analysis allowed developing a set of data-driven models for predicting the time to restoration (and other interesting quantities described later in the chapter) for different railway assets and failures. Moreover, it has been possible to measure the relevance of each input parameter considered in the estimation of the time to restoration by means of state-of-art feature ranking techniques.

Concerning the second scenario, the noise in data and the few examples of failures for some of the many different failure types made this task very difficult, so that it was not possible to complete the analysis proposed at the beginning of the project. However, the data available for the two scenarios is the same, although the objectives of the two analyses are different. For this reason, it has been decided to combine into a single analysis (as far as possible) both the tasks proposed by SR. Indeed, by demonstrating that it is possible to predict successfully the time to restoration based on information about past maintenance/repair actions and weather conditions, we also demonstrate that there is a correlation between these data and the time required to perform a maintenance/repair intervention.

Nonetheless, in order to estimate the significance of the different parameters on the time to restoration, a deeper analysis has been conducted. On the one hand, the performance of data-driven models has been evaluated by comparing their accuracy in presence and in absence of weather information. On the other hand, a state-of-the-art feature ranking methodology has been applied in order to estimate the relevance of each single input parameter of the models with the real outputs that the models aim at predicting.

This chapter describes in detail how the aforementioned important results have been achieved. In particular, Chapter 3.4.4.1 recalls the general problem more in details and formalizes its main characteristics. Chapter 3.4.4.2 describes the solution proposed for the forecasting of time to restoration and ranking the relevance of each input parameter, and the associated methodologies. Finally, Chapter 3.4.4.3 shows the preliminary results for this forecasting scenario.

3.4.4.1. Problem Formalization

3.4.4.1.1 The maintenance/repair process in details

Considering the objectives of the analysis described above, this work started from analysing the maintenance/repair process shown in Figure 3.19 and described by the available data. The records composing the maintenance/repair dataset report a number of variables related to each specific maintenance/repair action (see Figure 3.20): for example, the ID of the asset that shows a failure, the geographical location of the asset, the priority associated with the specific intervention, the type of failure, the technical department that took in charge the intervention, a series of timestamps related to the different phases of the intervention, and many others.

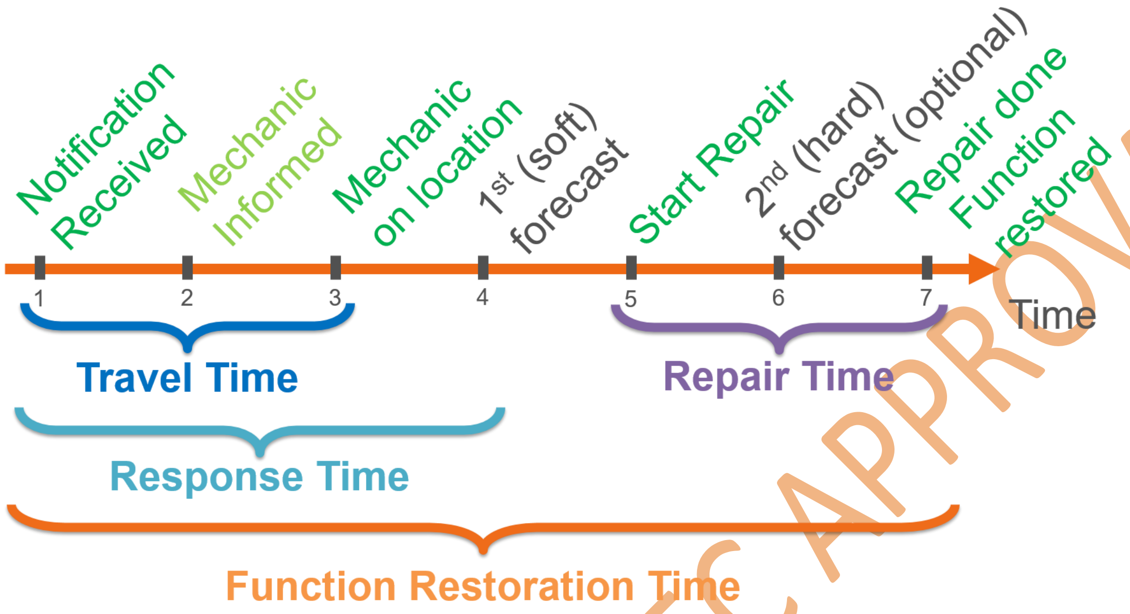


Figure 3.19: Maintenance/repair process timeline

Figure 3.20: Maintenance/Repair Report Contents

The set of fundamental steps for the maintenance/repair process timeline depicted in Figure 3.19 are described in Table 3.6. Analogously, the four interesting time quantities depicted in Figure 3.19 are described in Table 3.7. Each step has been associated with a number, and the duration of the interesting time quantities are defined in terms of step numbers in the first column of Table 3.7.

Step N.	Step Name	Description
1	Notification Received	At the very first step, Strukton receives a notification from the Infrastructure Manager that a failure on an asset has been detected. The notification includes several information, such as a priority level, the ID and location of the asset to be repaired, etc. Depending on the priority and the desired time to

Step N.	Step Name	Description
		restoration associated with the intervention, the rest of the repair/maintenance process is scheduled, therefore it can be carried out immediately (e.g. in case of emergency) or later (e.g. in case the failure is not critical).
2	Mechanic Informed	After the notification has been received, mechanics are informed that a failure on an asset has been reported. In case of high priority intervention, this step occurs immediately after the notification has been received, and the team is sent to the location of the asset. Otherwise, this step coincides with the scheduled time to start the intervention.
3	Mechanic on Location	This step records the timestamp in which mechanics arrives on the asset location, travelling from the closest headquarter. At this step, mechanics communicate that they reached the asset location and starts inspecting the asset for assessing its status.
4	1st (Soft) Forecast	Based on the first inspection of the asset, the mechanics perform a first "soft" forecast on the time needed to complete the intervention. The roughly estimated time to restoration is communicated to both the local headquarter and the infrastructure manager.
5	Start Repair	This step records the timestamp in which the mechanics start the repair intervention on the asset.
6	2nd (Hard) Forecast (optional)	Sometimes, the mechanics perform a second forecast on time to restoration, so to communicate a more accurate estimation to both local headquarters and IM. In this case, the IM can prepare the train operation in advance.
7	Repair Done – Function Restored	This step records the timestamp in which the mechanics completed the intervention, and the line has been freed so that train can travel again over it. Therefore, the complete function of the asset has been restored.

Table 3.6: Description of steps of maintenance/repair process

Duration	Interesting Time Quantity	Description
From step 1 to step 7	Function Restoration Time	This quantity describes the amount of time needed to complete the intervention and free the railway line from the moment in which the notification of failure of the asset has been received. It describes the time needed to complete the entire maintenance/repair process for a specific notification.
From step 1 to step 3	Travel Time	Amount of time between the reception of the notification and the arrival of the mechanics on the location of the asset to be repaired/maintained. In case of low priority interventions, it might not coincide with the time needed to travel from headquarters to the asset location.
From step 1 to step 4	Response Time	This quantity describes the time needed by the mechanics to start operating on the asset from the moment in which the notification has been received by Strukton. Analogously to Travel Time, this quantity might not coincide with the sum of the time needed to travel from headquarters to the asset location plus the time needed for the first inspection of the asset in case of low priority interventions.
From step 5 to step 7	Repair Time	This quantity represents the time needed by the mechanics to perform the repair on the asset. It comprehends the time needed to free the line after the intervention has been completed.

Table 3.7: Description of interesting time quantities for the maintenance/repair process

3.4.4.1.2 Problem Formalization based on available data

Considering the data described above, related to repair actions, the problem has been formalized as a multivariate regression problem, where each repair action represents a single observation of a process characterized by a set of steps ordered in time, and associated with specific assets, types of assets, types of failures reported, and the like. Simplistically, the goal is to analyse historical data about the repair actions, and try to estimate the time needed to perform the next action based on the time needed for similar past actions. The "similarities" are automatically estimated by machine learning algorithms by taking into account the available complementary information: the type of asset, the type of failures reported, the location of the asset, and the like.

The analysis is carried out by considering only the information that is available at each step. For example, the information collected by mechanics during the first inspection on the asset (carried out between step 3 and 4) cannot be exploited at the beginning of the repair process, therefore they cannot be used to perform a first estimation of the interesting time quantities.

Additionally, weather data have been integrated into the analysis, due to their paramount importance. Indeed, bad weather conditions such as rain or fog can noticeably affect the capability of mechanics to operate on the asset, and therefore taking them into account can enhance the predictive performance of data-driven models.

The repair action data are available from 2010 to 2015, while weather data start from 2000 and end on June 25th, 2014. The two datasets have been linked by applying the same methodology exploited for the RFI NC and FC scenario (described in "Deliverable 9.3 – Nowcasting methodologies"), i.e. the geographical locations of the assets have been correlated to the locations of the weather stations, so to find the closest one for which it is possible to extract the most accurate weather information related to each asset. The entire data refers to Netherlands, where Strukton is the responsible for asset maintenance.

The maintenance/repair action data refers to a set of railway assets that have been defined by the In2Rail WP9 partners in "Deliverable 9.1 – Asset Status Representation" as the most critical ones for railway operations. The list includes:

- bridge ;
- embankment ;
- level crossing ;
- line Section ;
- Signaling Lamp (standard, WIDO, WUBO, Bridge) ;
- switch ;
- switch-heating ;
- track.

3.4.4.1.3 Feature Extraction

For this dataset, a large feature engineering work has been carried out, resulting in more than 290 features added to the original dataset, based on the historical data available. The new features can be divided in five categories:

1. Time Intervals;
2. Weather;
3. Past Failures;
4. Open/Unresolved Failures;
5. Missing Values.

New features related to Time Intervals are extracted from the original data, which contain the timestamps of the steps of the maintenance/repair process (e.g. Failure Notification). The timestamps are converted to numerical quantities for analysis purposes by computing the time intervals between couples of events. Analogously to the original data, the new time intervals are available at different steps of the process, based on the last occurred event: for instance, the new feature "ToInform_Time", which is computed by subtracting the timestamp of the Failure Notification step to the one of the Mechanic Informed step, is only available after the Mechanic Informed timestamp has been recorded.

New weather features are extracted from the original weather data, including 39 different variables representing daily averages of weather measurements (mainly). The dataset includes information coming from different weather stations all over Netherlands, which are identified by a station ID and their geographical location. As previously mentioned, the geographical location of the assets has been correlated to the locations of the weather stations, so to find the closest one. The new features are organized in three blocks (composed of 39 features each) that represent the average of weather variables of the last 7, 30 and 90 days before the notification of failure is received by Strukton. Moreover, 3 new features have included from an external data source, i.e. a categorical value for the current season, and the time of dawn and sunset, and have been associated to each repair action.

A new group of 30 features has been extracted from the failures that occurred in the previous months/years before a certain notification is received. They can be divided into three subgroups, where the first one relates to the past failures for a specific asset. The second one refers to the past failures occurred in the geographical area identified by the same "geocode" (which is used by Strukton for internal purposes, such as determining the maintenance team that has to perform the intervention) of the asset failure under examination. The third group is again related to the past failures occurred in a certain geographical area, but these new features are extracted by considering the specific area surrounding the asset based on the latitude and longitude coordinates. These features are computed for different time horizons (namely 7, 30, and 365, 730 and 3650 days before the notification under examination), and for different types of failures. Finally, two more features are included in the new dataset, i.e. a Boolean value indicating whether a repair action has been already carried out in the past on the asset under examination, and the amount of time passed from the last repair action performed on that asset.

New features belonging to the group of open/unresolved failures features are extracted by considering all the maintenance/repair actions that still have to be completed at a certain time. In particular, these new features are generated by computing the number of open/unresolved failures at the different steps of the considered maintenance/repair action: for example, one subgroup of features relates to the number of all the open/unresolved failures between the Notification Received step and the Mechanic Informed step of the single maintenance/repair action under examination. Moreover, an additional level of granularity is added by looking only at the open/unresolved failures in charge to the same Technical Department included in the dataset. Finally, the features are again subdivided by priority code, for which 5 levels are indicated in the dataset.

Last but not least, a group of new features has been extracted by looking at missing values in the dataset. Instead of simply discarding values/records for which only partial information is available, which would lead to a loss of information, it has been decided to add a set of new features in order to highlight the unavailability of data, while inserting a numerical value in place of the missing one.

The next chapter further describes the proposed solution for this forecasting scenario.

3.4.4.2. Proposed Solution

As presented in Chapter 3.4.4.1.2, the problem of forecasting time to restoration has been formalized as a multivariate regression problem [47] [48]. Indeed, the problem includes multiple variables of interest (i.e. the four quantities to be predicted) and other possible correlated variables (i.e. information included in maintenance/repair reports, novel features extracted, and weather data). The goal is to find a solution able to model the link between the variables of interest, their past values (i.e. their history), and the other correlated variables. In other words, the resulting models should predict, with the highest possible accuracy, the Function Restoration Time, the Travel Time, the Response Time and the Repair Time for a new repair action to be performed in response to an asset failure. Figure 3.21 shows a simple graphical representation of the mapping of the problem into a multivariate regression one.

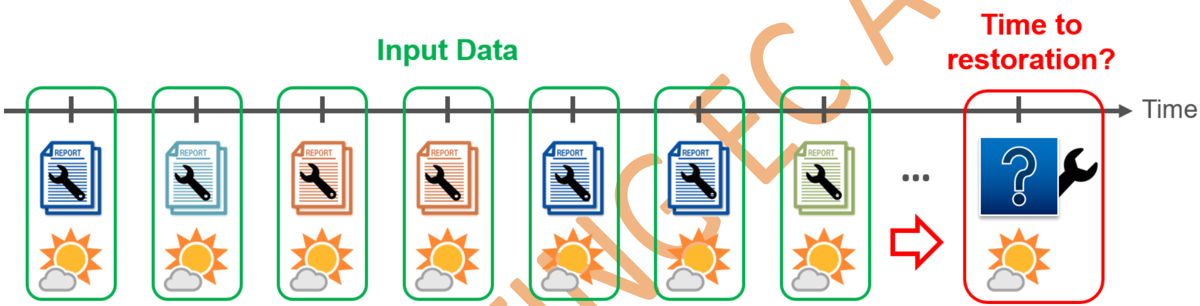


Figure 3.21: Simple graphical representation of forecasting time needed for repair actions. Although not included in this picture, actions could overlap and/or could be performed in close areas

By exploiting the aforementioned data, a dataset has been built and inputted to machine learning algorithms in order to build data-driven models, one for each of the 4 quantities to be predicted. The dataset included more than 40,000 rows, each composed of more than 350 input features and the 4 output features. Each row of the dataset can be treated as a different sample (i.e. a distinct event), because the information regarding the interaction between these different events is included in the novel extracted features. For instance, some of the new features deal with spatio-temporal correlations of events, and therefore give the possibility to the algorithms to take into account these factors while building the data-driven models.

In order to perform an exhaustive analysis, a different set of data-driven models have been built by taking into account only the repair action associated with the highest priority level. As it will be described in the next chapter, this led to different results, in particular to increased performances of the models, because high priority interventions follow a slightly different process with respect to the others.

Moreover, in order to assess the importance of weather data into the prediction of the interesting time quantities, the data-driven models have been built both with and without weather information.

Kernel Regularized Least Squares (KRLS) [49] [50] [51], a state-of-the-art machine learning algorithm belonging to the family of kernel methods (see Chapter 3.4.3), has been exploited for regression analysis. It is based on the concepts of Structural Risk Minimization, Mean Square Error and on the so-called "kernel trick". The Gaussian kernel has been used, since it enables learning every possible function [52] [53].

Finally, a technique based on permutation test [69] [70] [71] (also called randomization test) has been exploited in order to assess the importance of each variable for each of the setup tested (i.e. weather vs. no weather, and high priority vs. all the interventions). In data-driven modelling, these techniques are used in order to quantify the importance of the feature (input) variables of a dataset by computing the sensitivity of a model to random permutations of feature values. The intuition behind permutation tests is that if a feature is not useful for predicting an outcome, then altering or permuting its values will not result in a significant reduction in a model's performance. Results of this analysis have been included in the next chapter.

3.4.4.3. Preliminary Results

This chapter describes the laboratory tests and simulations performed in order to retrieve the preliminary results for the forecasting of time to restoration. It is important to notice that two different types of results are presented:

- Forecasting Performance, which relate to assessing the performance of the data-driven models in predicting the interesting time quantities through the exploitation of historical data. In other words, simulations have been performed in order to quantify the goodness of predictions with historical data by comparing what the models would have predicted and what really happened in the past ;
- Feature Ranking, which is a statistical analysis for assessing the significance of each input feature of a dataset for prediction. Therefore, the result is a set of lists (one for each simulation scenario) ranking the importance of variables.

The two types of tests, their setup, and the achieved preliminary results are reported in separate chapters.

3.4.4.3.1 Forecasting Performance

The general idea behind simulations for assessing the performance of data-driven models follows the one at the basis of the model selection procedures [25] [26]. In short, part of the available data ("training set") is used to build models, while the rest is exploited for performance evaluation ("test set").

In order to take advantage of the temporal order of data provided by Strukton, a so-called "online" modelling approach has been exploited. In particular, data has been further divided, so that the models have been trained firstly on the first years of data. Then, time evolution has been simulated by inputting to the models newer maintenance/repair action reports one by one in time order, while measuring models performances. Once a certain batch of data had been inputted to the models, it had been integrated into the training dataset and the models have been retrained, so to exploit the new information available. This procedure is shown graphically in Figure 3.22.

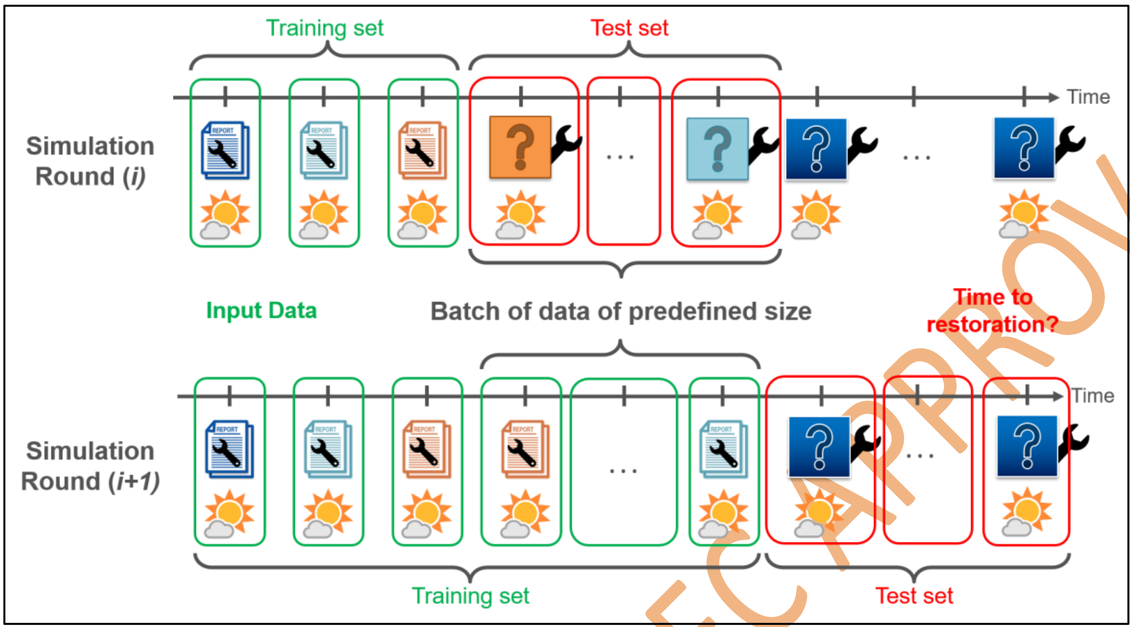


Figure 3.22: Simulation procedure for online update of data-driven models

The models have been trained in different "simulation scenarios", so that different settings have been tested and the performance of each scenario have been compared with the ones achieved in the other scenarios. Four scenarios have been defined, according to the availability of weather data (i.e. available vs. not available), and to the priority level associated with maintenance/repair interventions (i.e. high priority vs. all the priorities together). The results will be presented separately for the four scenarios. Table 3.8 shows in a concise way all the combinations that lead to the four different scenarios.

	Weather	**No Weather**
High Priority (Level 2)	W + HP	NoW + HP
All the priorities together	W + ALL	NoW + ALL

Table 3.8: Simulation scenarios based on weather information and priority levels

Finally, since the quantities to be predicted are four for each single maintenance/repair intervention (recalled in Table 3.9), four different data-driven models had to be developed, based on different input data depending on the ones available during the action under examination. This approach makes possible to guarantee always the best performance for data-driven models, and that each model considers all the possible available information.

Function Restoration Time	Repair Time
Response Time	Travel Time

Table 3.9: Quantities to be predicted

To sum up, many simulations have been carried out due to the different simulation scenarios, to the need of creating different models for the different quantities to be predicted, and to the online approach that is explained above. Therefore, the entire set of preliminary results related to Forecasting Performance (described in the next chapter) comprehend the 4-simulation scenario for each of the 4 models built for the 4 quantities to be predicted, resulting in a total number of setups of 16, for which preliminary results have been computed.

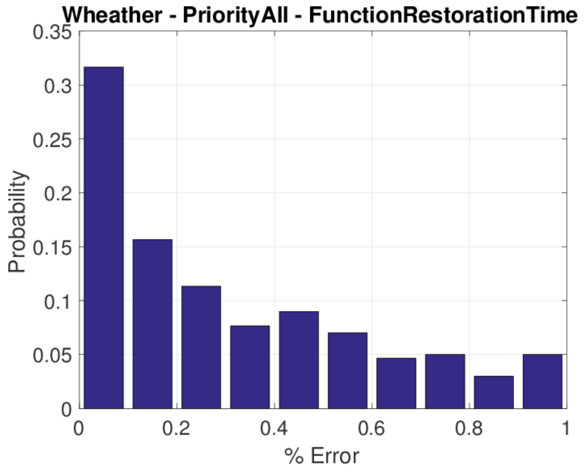
3.4.4.3.1.1 Description of Preliminary Results

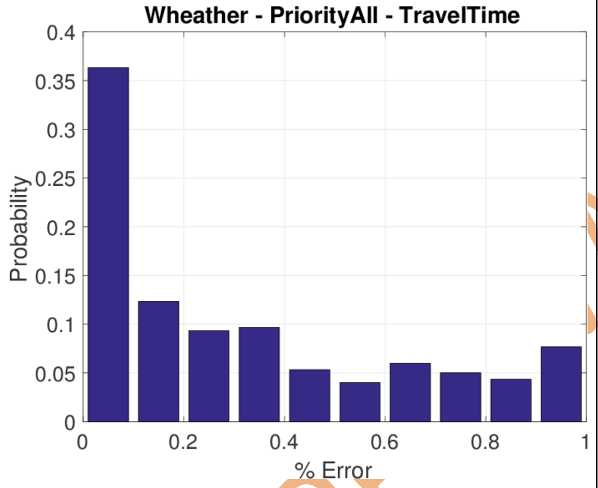
The preliminary results achieved for the problems investigated in this scenario are shown in the 16 graphs included in Table 3.10. These graphs are usually exploited in order to visualize and inspect the results of a regression analysis, and represent some of the requested forms of data visualization.

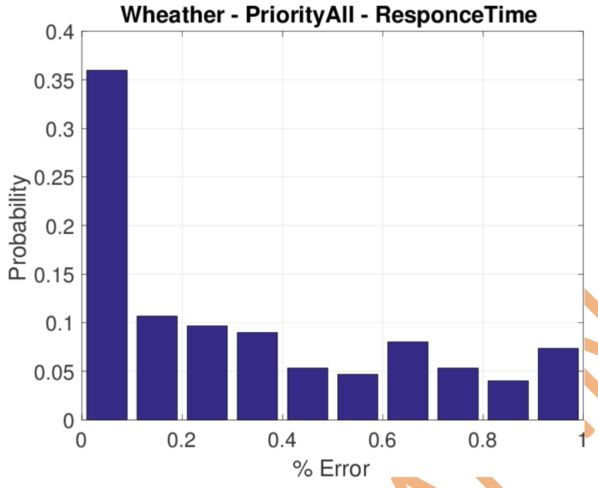
The Percentage Error Distribution graph is a histogram showing the distribution of the percentage error, which is defined as the difference, in percentage with respect to the true value, between the estimated/predicted values and the true values (e.g. if the true repair time is 7 hours, an error of 5% means that the estimated/predicted repair time can be 7 hours plus or minus 20 minutes). In particular, the histogram includes the values of percentage error on the x-axis, and the probability of occurrence of that particular value of the percentage of error on the y-axis. For this type of graph, in the ideal situation, the probability of having a percentage error equal to zero would be equal to one. Generally, the closer is the mode of the probability to zero, the better are the model performance. Moreover, if the probability of having a certain percentage error decreases as the percentage error increases in the graph, it is possible to state that the model captured the existence of some information inside data.

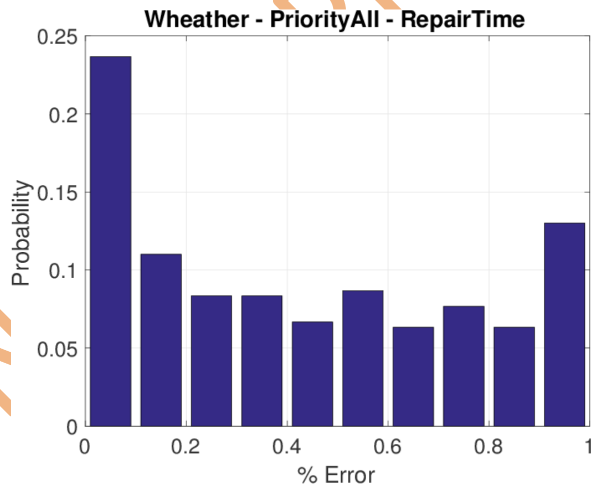
Forecasting Performance – Graphs of Preliminary Results

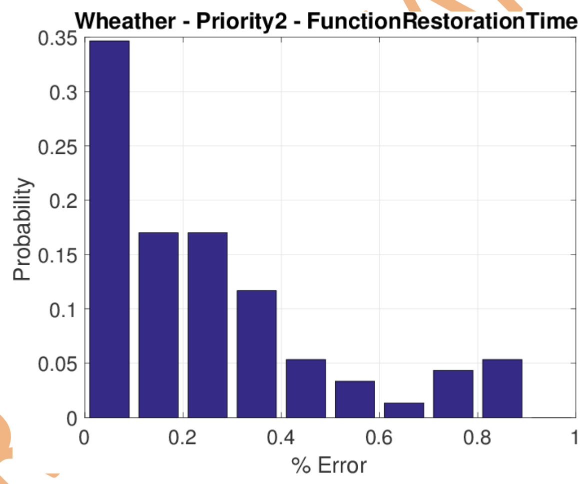
Forecasting Performance – Graphs of Preliminary Results

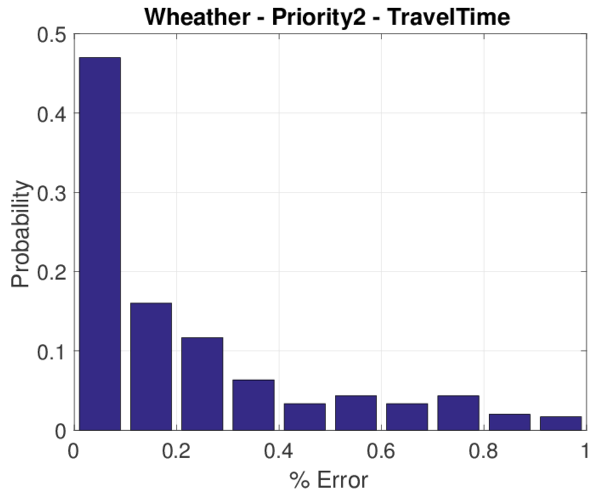




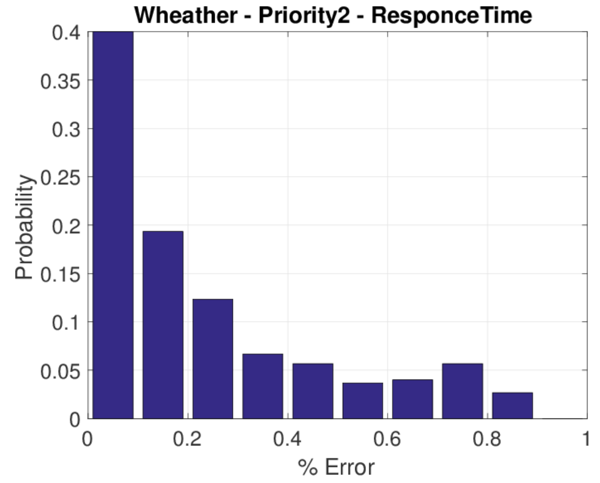


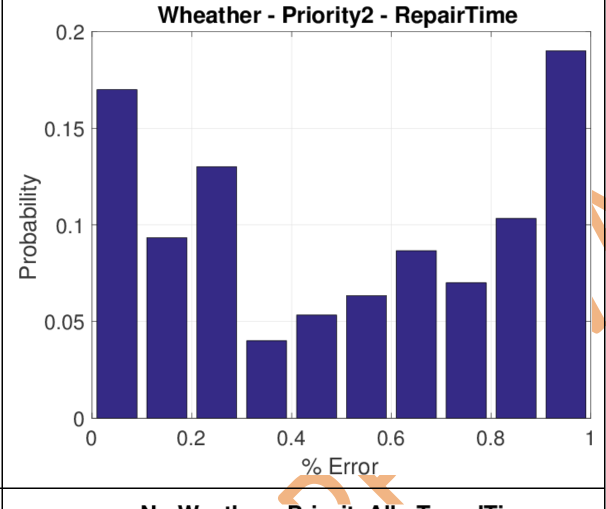


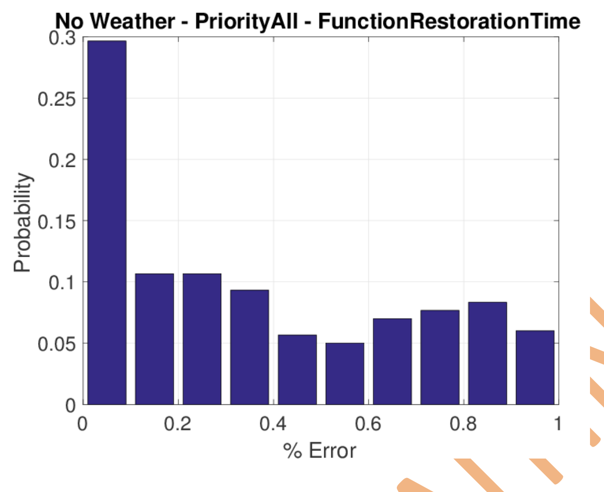


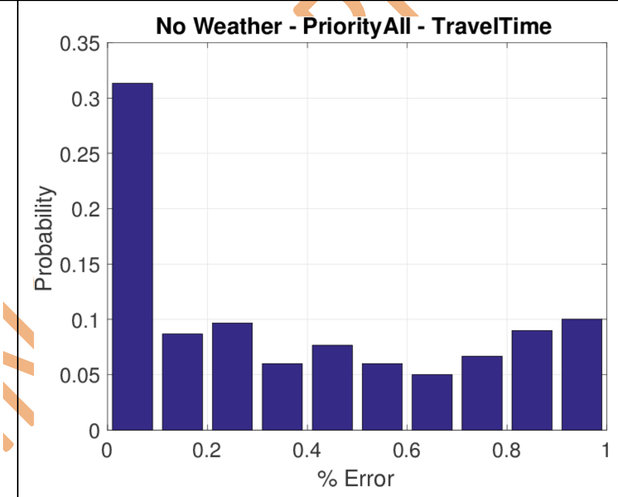


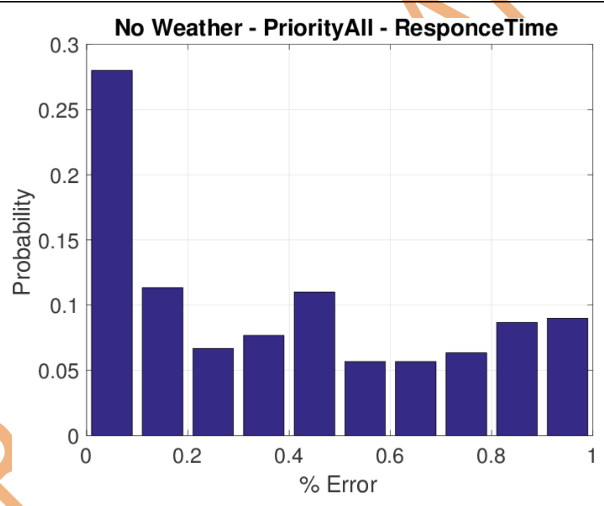
Forecasting Performance – Graphs of Preliminary Results

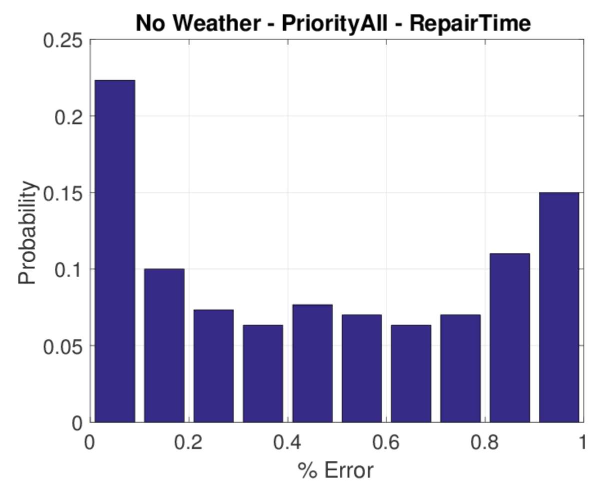












Forecasting Performance – Graphs of Preliminary Results

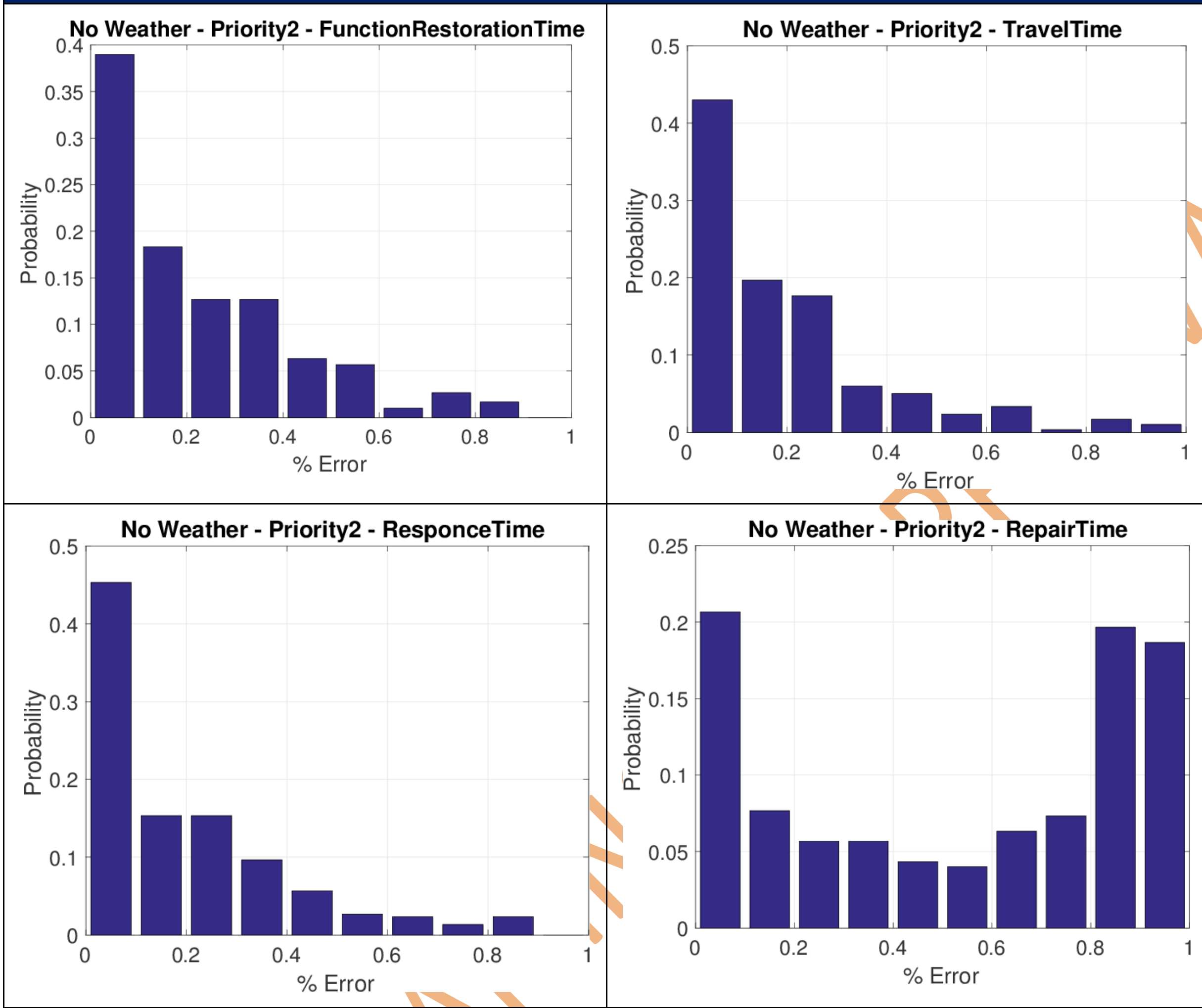


Table 3.10: Forecasting Performance – Graphs of Preliminary Results

3.4.4.3.2 Feature Ranking

In order to assess the relevance of each feature included in the available dataset (composed of both original and extracted features) for the prediction of the interested quantities, a technique based on permutation test [69] [70] [71] (also called randomization test) has been exploited. Tests have been conducted on the data-driven models built for the tests related to the assessment of the forecasting performance, therefore models for the same simulation scenarios and the same quantities have been used for this activity (see Table 3.8 and Table 3.9).

The result of this procedure is of tables that associate to each feature a P-value, from 0 to 1 (1 means highly relevant), which represents the importance of that particular feature for performing predictions for each specific simulation scenario defined and for each quantity to be predicted. This means that 16 different tables have been produced. Due to the large number of features examined, it has been decided to report here in this document only the top 10 most relevant features for each simulation scenario and for each quantity to be predicted, instead of the complete tables. Moreover, this chapter only includes four tables related to the prediction of Function Restoration Time, to be intended as exemplary tables, while the others have been included in appendix 6.1 (page 83). Finally, a detailed description of features (both original and extracted) included in the tables is available in Chapter 6 (Appendix A1).

P-value	Feature
1,000	Failure Type ID
0,602	Day average Temperature – averaged over 7 days
0,488	Global radiation – averaged over 30 days
0,409	Reference crop evaporation – averaged over 30 days
0,244	Minimum temperature – averaged over 7 days
0,198	Time from last failure
0,192	Maximum temperature – averaged over 7 days
0,177	Lowest hour-value of the air pressure reduced to sea level – averaged over 30 days
0,171	Hour in which maximum temperature was measured – averaged over 90 days
0,162	Reference crop evaporation – averaged over 7 days

Table 3.11: Weather – Priority 2 – Function Restoration Time

P-value	Feature
1,000	Failure Type ID
0,139	Day average overcast (coverage of the upper air into eighths, 9=sky is invisible) – averaged over 7 days
0,100	Day average Temperature – averaged over 7 days
0,094	Lowest hour-value of the air pressure reduced to sea level – averaged over 30 days
0,093	Reference crop evaporation – averaged over 7 days
0,082	TD_Open_failures_s1_in23_p4 → Open failures of priority 4 that are between mechanic informed and mechanic on location for the same Technical Department between notification and mechanic informed of current repair action
0,075	Global radiation – averaged over 30 days
0,067	Time from last failure
0,050	Hour in which minimum occurred sight was measured – averaged over 30 days
0,049	Minimum temperature – averaged over 7 days

Table 3.12: Weather – Priority All – Function Restoration Time

P-value	Feature
1,000	TD_Open_failures_s1_in23_p4 → Open failures of priority 4 that are between mechanic informed and mechanic on location for the same Technical Department between notification and mechanic informed of current repair action
0,986	Open_failures_s1_in23_p4 → Open failures of priority 4 that are between mechanic informed and mechanic on location between notification and mechanic informed of current repair action
0,922	Failure Type ID
0,674	Time from last failure
0,480	Open_failures_s1_in12_p5 → Open failures of priority 5 that are between notification and mechanic informed, between notification and mechanic informed of current repair action
0,443	TD_Open_failures_s1_in12_p5 → Open failures of priority 5 that are between notification and mechanic informed for the same Technical Department between notification and mechanic informed of current repair action
0,356	Time of sunset in the current day
0,311	Time of dawn in the current day
0,310	Year
0,247	Object ID

Table 3.13: No Weather – Priority 2 – Function Restoration Time

P-value	Feature
1,000	Past failures in a $1km^2$ circular area surrounding the failure under examination over the past 3650 days
0,806	Past failures in the area identified by the same "GeoCode"[3] over the last 3650 days
0,719	Weather Station identifier
0,679	Priority Code of the intervention
0,643	Past failures in a $1km^2$ circular area surrounding the failure under examination over the past 730 days
0,582	Past failures in a $1km^2$ circular area surrounding the failure under examination over the past 365 days
0,573	Past failures in the area identified by the same "GeoCode" over the last 30 days
0,420	Past failures in a $1km^2$ circular area surrounding the failure under examination over the past 30 days
0,411	Maximum Repair Time allowed for the intervention
0,346	X(Long)_Begin

Table 3.14: No Weather – Priority All – Function Restoration Time

[3] The "GeoCode" is one of the input features provided by Strukton. It is used to identify geographical areas based on the location of railway lines.

3.4.5 Analysis and Discussion

3.4.5.1. Comments on Forecasting Results

By looking at the graphs of Table 3.10, it is possible to draw the following considerations:

- the general results achieved for this scenario are very interesting and show that it is possible to perform quite accurate predictions of time to restoration and of most of the other relevant quantities for estimating the time needed to for a complete repair action, except for the Repair Time. It is important to take into account that original data are noisy, and many missing values have been encountered ;
- all the graphs, except for the ones related to Repair Time, show that the probability of the models to give a prediction showing a certain percentage error decreases as the percentage error increases, meaning that the models are able to extract some useful knowledge from the available data and so they are not randomly guessing ;
- the Repair Time is the most difficult quantity to predict. In all the four simulation scenarios, the performance of the associated data-driven models are not very satisfactory. Noise in the data is the most probable reason for this behaviour, since the data-driven models for Repair Time are exploiting some additional data with respect to the ones, for example, for predicting the Functional Restoration Time. Therefore, with the current data it is not possible to improve the performance of these models ;
- the data-driven models associated with the other three quantities, instead, show satisfactory performance on average. Among these, the Function Restoration Time is the most difficult to predict, while models for Travel Time and Response Time always show similar performances ;
- there is a relevant difference between simulation scenarios where all the priority levels are considered all together and the ones where only the repair actions with highest priority level 2 are included. Results show that data-driven models constantly achieve better performance in the second situation ;
- the inclusion of weather data increases the performance of data-driven models on average, although slightly.

3.4.5.2. Comments on Feature Ranking

By looking at the tables indicating the top most important variables for the prediction of the relevant quantities for each simulation scenario, it is possible to draw the following considerations:

- all the top ten tables based on the output of the permutation test include some of the variables generated through the feature extraction process previously described, meaning that this work had a great impact on the performance of the data-driven models ;
- concerning models predicting Function Restoration Time, it is possible to say that the

most important variables are both part of the original dataset and of the extracted features ;

- concerning the Travel Time models, it is clear from the tables that the most critical variables are related to the weather conditions and to the open failures, which seems a reasonable result. The same consideration holds for models predicting Response Time, which only slightly differs from the Travel Time ;
- models predicting Repair Time, instead, show that they take advantage of variables that are not available to the other models (i.e. variables that become available in the middle of the repair process and not at its beginning). For example, the Action_carried_out_code variable and the ToLocation_Time variable belong to this category ;
- P-values included in tables related to the "Priority 2" simulation scenarios are generally higher than the ones computed for "Priority All" scenarios, and this holds for the rest of the variables not included in this document. This behaviour reflects the fact that performance of data-driven models for the SR/UNIGE scenario are highly dependent on the quality of data, and that for "Priority 2" scenarios it is possible to achieve higher performance.

3.5. SR/DLR Scenario

3.5.1 Summary of Scenario by SR/DLR (D9.3)

In this scenario, the forecasting of the partial switch status is investigated based on findings gained in 9.1 Asset Status Nowcasting and described in D9.3. This proposed scenario is described in Table 3.4.

Title	**Forecasting switch status by mining monitoring data**
Organisations Involved	• Deutsches Zentrum für Luft- und Raumfahrt DLR e.V. / German Aerospace Center (DLR) • Strukton Rail Netherlands (SR)
Objective(s) of the scenario	The main objective of the scenario is to be able to forecast detectable anomalies related to switch condition by performing data analyses on the monitoring data. The basic information used for this task is the measured current of the switch engine needed for moving the switch blades from one position to the other (end-)position, although some additional available data (such as reported incidents, observed defects, repair times) are exploited. In Task 9.1 an exploratory data analysis was conducted to derive a classification for condition nowcasting based on a decision tree (see D 9.3). The experience gathered especially regarding the definition of informative features is input information for the switch failure detection model developed in WP6 and the switch status forecasting in Task 9.2. The result is a model which suitable to identify unusual behavior in an early stage of emerging failures based on historical data of normal operation to forecast failures of the switch. This information will be a valuable input for TMS in order to anticipate upcoming technical problems.
Relationship with TMS and/or maintenance	Most of rail infrastructure managers will indicate switches as very critical assets for its operation, because whenever the availability of a switch is compromised, it introduces numerous problems leading to unavailability of the train path and resulting in train delays, significant disturbances in the operation, increased fuel costs, crew expenses, maintenance and repair costs, and generally in a negative impact on reputation and revenues. In order to prevent this kind of undesirable situation and events, seeing the problems developing and being able to anticipate before they get critical, it would provide significant benefits for train operation in order to prevent problems before they occur. In case a maintenance crew is engaged in order to fix a growing problem in a switch, more information about the possible problem is necessary in order to successfully perform right maintenance and to fix the real problem or the cause of it.
Description of the scenario	The forecasting approach will exploit monitoring and meteorological data in order to gain knowledge to improve the switches availability. One monitoring device for switches is e. g. POSS® (Preventive Maintenance and Fault Diagnosis System) by Strukton, which records the measurement of the power consumption of the switch engine for each switch movement. Other useful data include maintenance data, failures data and possibly asset usage data (e. g. number of movements, number of trains passed over the switch, etc.). Forecasting will be performed by analyzing data in order to improve the general understanding of the relationships between power consumption and asset degradation, and to detect the differences in normal and fail-behavior. Gained knowledge will help further understanding of the information in data which can be used for predictability of the critical asset status.
Data exploited for	**Switch Monitoring data**

the scenario	The POSS® system can be used to monitor rail assets such as switches, train detection systems, level crossings, etc. The monitoring data are presented in a universal format and can be reviewed via the Internet. This data is stored in a POSS® database, which includes data from many (thousands) switches related to many years. This monitoring system has its own thresholds, managed by maintenance engineers or switch experts, which are also an input for the analysis. Scope of this monitoring system, and also therefore the scope of the analysis, is allocated at Switch Panel part (see D9.1 5.4.2. Asset sub-components) and mainly on the switch engine and the deviations caused in-/by construction in switch panel. For this scenario, the used data set is collected from a period of three years from 19 switches with a single electrical engine. These switches are of type NSE (a DC powered electrical engine). **Recorded Incidents** Historical datasets regarding the recorded failures on the same switches/points in same timeframe as the monitoring data. This data is provided by Strukton Rail. **Maintenance actions data** Historical datasets regarding the maintenance activities executed in the same timeframe on the same switches provided by Strukton Rail. This data is collected by Strukton Rail but commissioned by ProRail (Dutch rail infrastructure manager). Data is originally stored in a Maintenance Management System. **Usage/Load data** An additional dataset, the Load data is generated and provided by ProRail, the Asset Manager. This data is mainly related to the usage conditions of the assets, and includes information about the number of trains that passed over the switch, their weight, etc. **Weather condition:** Data retrieved from the Royal Netherlands Meteorological Institute (KNMI) More detailed descriptions of the POSS® system, NSE type switches and the available data sets are given in the appendix.

Table 3.15: Tabular Description for Scenario by SR/DLR

3.5.2 Data description for Forecasting scenario

3.5.2.1. Input parameters

The available data sets used for the input parameters are described in Table 3.4 and in the Appendix 6.1.

The main input for the data analysis is the current time series (acquired at 50 Hz) captured while switches are changing blade position. The input data is further enriched by the ambient temperature measured at the relay house (located in the order of 1 km away from the switch) at the time of the switch movement. Available information about maintenance actions (scheduled or due to failure) performed on the asset supports the data analysis specially to understand sudden changes in the power consumption and the general behavior of the asset.

It is foreseen from the work done in In2Rail WP9 and WP6 that more data describing relevant influences such as weather conditions (including rain, exposure to sun, etc.), maintenance (scheduled and reactive) as well as train operations must be included into the automated data analysis to further improve now- and forecasting for switches. The gathering, preprocessing and utilization of these data sources for the automated now- and forecasting will be addressed in the Shift2Rail project In2Smart.

In the following chapters, the data processing steps involved in the switch SPC failure detection model readily available in D6.4 is summarized. The analysis and discussion take as example the movements of switch 3076 in both directions.

3.5.2.1.1Feature selection and scaling

The capability and performance of the Statistical Process Control SPC model for failure detection (summarized description below, details in In2Rail D6.4) depends on the features extracted from the measured time series. The expert's knowledge about the asset objects normal behavior is crucial for feature engineering and selection/ranking. The feature set considered for the analysis presented in the following chapters is enlisted below:

- area under the curve (represents total power consumed for switch blade movement as well as locking and unlocking of the blades) ;
- maximum current ;
- median current ;
- Kurtosis ;
- Skewness ;
- duration of switch movement ;
- mean current value during switch blade movement ;
- standard deviation of current during switch blade movement.

Typically, the different types of features (e. g. max values vs. standard deviation values) extracted from time series vary in range. For example, typical maximum values of current measurements lay within 14 and 16 A, the current standard deviation during switch movement is below 0.5 A. The failure detection approach is sensitive to these differences in range/absolute value and requires the normalization of each feature vector (containing as many values as there are current curves) such that each scaled feature vector has zero mean and a standard deviation equal to one; this transformation is also known as centering and scaling (Kuhn & Johnson, 2016).

Additionally, the features extracted from the time series present a systematic temperature variation. For example, the total power consumption decreases when temperature increases due several different variables (length of the switchblades changing due the temperature in combination of adjustment, switch glides (different types), oil (if present) etc.), leading to a reduced friction between the moving parts. In order to account for the temperature dependency of the features, the centering and scaling transformation is separately applied to subsets of the features extracted from the current curves, which were measured at approximately the same temperature (subsets are defined in terms of 1 K temperature bins). The effectiveness of the scaling method is discussed in more detail in D6.4. In what follows we refer to the scaled features as features.

3.5.2.1.2 Output parameters of Statistical Process Control SPC model

Basic idea of the forecasting approach presented here is to use the output parameters (Hotelling's parameter T^2 and the Squared Prediction Error SPE) of the Statistical Process Control model (Böhm, Schenkendorf, & Lemmer, 2016) not only for switch failure detection (see In2Rail D6.4) but furthermore to create also alerts in an early stage of failure development. The used SPC output parameters and the SPC model are summarized and explained in more detail in the following chapters. In general, the SPC model reduces the information contained in all current curves into the two parameters T^2 and SPE quantifying and characterizing the deviation of the current curves under analysis from normal asset behavior.

3.5.2.2. Output parameters and its relation to TMS

Output of the model is information about the technical status of the switch in comparison to usual asset behaviour. At the current stage, the model is capable to create alerts as forecasts of unusual behaviour possibly developing towards a switch malfunction. The asset status can be described and translated into several stages/levels based on the extent of the deviations of the switch from normal behaviour. The levels can then be divided based on the urgency of the status based on the decision tree developed for switch status nowcasting (D9.3). Automated levelling must be implemented and will be further improved within Shift2Rail project In2Smart. Other levels of the translated asset status into the level of the relevant information are shown in Figure 3.23. For example, the second level of predicted asset status is the one where the asset urgently needs attention to stay functional. This information is relevant as an input for TMS operator because if there is no action put in motion, the switch is going to fail in which case the operation must be adapted to the situation. Other, less critical, levels of asset status will eventually lead to the worst-case scenario if the necessary countermeasure has been taken. The forecasted asset status in terms of alerts created in an early stage of failure development provide an opportunity to stop the degeneration of the asset and fix the asset before it reaches critical point/a malfunction.

CURRENT SITUATION		NOWCASTED ASSET STATUS			FORECASTED ASSET STATUS	
asset status as TMS input parameter (current situation)	effect		asset status after data Nowcasting data analysis	additional TMS-input parameters	Forecasted asset status	TMS-input parameters
Switch is not funcioning	constrains in asset usage; rerouting trains no train passage ...		Switch not responding Responding but jammed	additional information in order to fix the issue faster	Forecasted possible fail- status	chance to react before the asset fails ->no trainpath disturbances ->no train delays
Switch is functioning	normal usage of the asset until it fails	1	Very bad condition	can be barely used; maintenance crew engaged with urgency	Very bad condition	Provides information to TMS about possible unavailibility due the necessary maintenance; Gives additional information and triggers necessary maintenance actions in order to fix the switch and prevent future problems ->possible problems prevented
		2	Bad condition	can be used; maintenance crew informed	Bad condition	
		3	Minor condition issues	Input for maintenance plans in order to prevent further decaying	Minor condition issues	
		4	Good condition	normal operation	Good condition	

Figure 3.23: Example of how the asset status (now- or forecasted) can be used by TMS

3.5.2.3. Uncertainties of the input parameters and forecast

The inputs for TMS will be generated alerts of unusual asset behaviour. TMS can anticipate on this status based of the urgency. The urgency is depending on the estimated time when the forecasted status is getting critical. However, the switch can behave differently depending on external influences. If there is no certain way to determine the degradation speed, the forecasting of the status cannot be linked with an accurate timeframe. As switches are complex electro-mechanical systems with numerous different types of failure the reliable estimation of the remaining time to malfunction is not possible at the current stage of knowledge. A more detailed discussion of this issue is given in the following chapters.

A further type of uncertainty related to the complexity of switches is the 'flipping' status observed on switches in operation. This means the status which changes over a short period of time from normal – bad – normal without any maintenance actions conducted in the meantime. The switch status is determined or forecasted based on its behaviour represented by power consumption and patterns of normal behaviour from the past. The behaviour can be significantly influenced by external processes (e. g. obstacles blocking the blades) which lead to a 'bad status' but after some time the switch gets to 'normal' state. This irregularity be uncertainty and can eventually be interpreted as false negative forecasting as up to now not all relevant influences are known or represented by suitable parameters in the data set. To reduce this source of uncertainty known influences are considered during the feature scaling (see temperature-compensated feature scaling described above).

It is foreseen that the uncertainties of the failure forecasting in the current state of development will predominantly be related to incomplete modelling (not all relevant influences are/can be included by describing parameters/features yet and/or training data sets are not (yet) complete).

The obtained forecasting can furthermore only cover a subset of typical switch failures and the uncertainties remain unknown for the moment. Due to the described reasons, a valid quantification of the forecasting uncertainties is not yet possible. As part of the upcoming verification and validation work (D9.5) a careful evaluation of the obtained forecasting alerts will be conducted with the available data set to assess the forecasting performance. The forecasting approach developed in In2Rail WP9 will be further developed and improved in In2Smart WP8 by adding further not yet available data (sources) representing relevant influencing parameters. Furthermore, the analysis done in In2Rail revealed that the sampling rate of the current measurements (50 Hz) is too low to capture quite small deviations of the current curves getting now relevant for the anomaly detection for failure forecasting.

3.5.2.4. Ranking of the parameters

The features used for the forecasting are results of the work in WP6 and Task 9.1 and were further improved and extended in Task 9.2. The ranking and evaluation of features is mainly based on expert knowledge introduced to the process by exploratory data analysis (see D9.3). Furthermore, state-of-the-art methodologies such as Principal Component Analysis (PCA) are used to further analyse the available parameters and defined features. The applied methodology (Statistical Process Control SPC, see Chapter 3.5.4.2) includes furthermore a feature ranking based on robust PCA by itself.

3.5.3 Methods for prediction

3.5.3.1. State of the art

Automated fforecasting for switch asset status based on continuous switch current consumption (or other comparable measurements such as from force sensor at the switch blades) are not yet seen in daily and reliable operation. The main challenge is the complexity of railway switches as electro -mechanical systems with numerous types of failure modes. At the same time, significant effort is spent at research institutions and companies to develop corresponding forecasting models (a comprehensive overview is given in (Camci, Eker, Baskan, & Konur, 2016). The complexity of the failure modes (and their potential superposition at switches in operation) is the tremendous challenge for the development of especially physical degradation models. Even under well controlled laboratory conditions with simulated failure development physical models show up to now very poor performance (Camci, Eker, Baskan, & Konur, 2016). For these reasons the main focus is laid in the last years on the development of data-driven models based on historic data (e. g. (Camci, Eker, Baskan, & Konur, 2016)(Eker, Camci, & Kumar, 2010)(Letot, et al.)). Several feasibility studies have been conducted and are ongoing utilizing a wide range of sophisticated empirical statistical models and supervised machine learning approaches. Main advantage of the data-driven methods (especially of the supervised machine learning approaches) is that already today models with good apparent prediction performance can be derived for example data sets. Main remaining challenges are the problems of over-fitting for the example data sets under analysis, the creation of complete (containing all relevant types of switch failures) training data sets with correct labelling and the generalization of the derived models for a large number of switches. These data-driven models furthermore imply the availability of historic failure data for training. Up to now it is furthermore not proven that data-driven models trained with available historic failure data can reliably be applied to data from the same switch after a major maintenance or repair action. One major challenge for all data-driven models but especially for the empirical statistical models is the nonlinear and volatile representation of the development of numerous of the relevant switch failure types in the gathered measurement data. Stable linear or exponential trends suitable for modelling are not frequently observed for switches in operation. Furthermore, the superposition of different failure types at the same time occurs in daily operation and has severe impact on every type of forecasting models under research.

3.5.3.2. Methodology and methods

Significant effort was and is spent to implement switch failure prediction based on data-driven models derived with supervised learning on labelled training data sets. Also models with high prediction performance can be obtained by supervised learning such models are not yet in common operation due to challenges with the training sets as summarized above. Therefore, a different and complementing approach is investigated in In2Rail WP9. Here an approach not relying on a training data set with labelled switch failures will be implemented and evaluated for its failure prediction performance. Basic concept for the approach under investigation is to use the output of the switch failure detection model based on Statistical Process Control (SPC) developed in In2Rail WP6 to provide a short- to mid-term forecast of emerging switch failures based on time series forecasts for the output of the switch failure detection model.

3.5.4 Results

3.5.4.1. Problem Formalization

Every influence affecting the movement of the switch blades or the mechanics to move the switch blades can be observed by variances of the switch engine power consumption. Furthermore, some other influences exist which can affect the measured current such as defects at the electrical motor itself or at the power cables. Therefore, power consumption shows either the expected or not expected behaviour of the point covering relevant (but not all) types of typical switch failures. The expected behaviour is interpreted as normal behaviour due the current conditions. For example, in very cold weather conditions it is known that some components reach a state when they need more force to do what they must do. This can be detected by increased power consumption. So, the expected current graph in cold weather can look the same as a not normal current graph on a hot day. The problem of the classic monitoring system is that without analysis from data driven models the thresholds cannot distinguish expected from normal behaviour.

The current monitoring systems are not able to classify the power consumption and relate it to the asset status without the interpretation of the specialist/engineer. One main challenge for implementing automated now- and forecasting is the identification and consideration of all relevant influences and the corresponding data to derive and include the necessary describing parameters. In Task 9.1 careful exploratory data analysis was conducted to improve the understanding of the current consumption characteristics and to derive meaningful features from the current measurements for nowcasting. Cluster analyses were applied to start with the implementation of a status classification by a decision tree (see D9.3). One major advantage of the classification approach under development based on the decision tree is a high transparency and interpretability which is important to support the maintenance engineers by drawing decisions from the classification results.

The forecasting of the asset status is often based on linear models. More complex data-driven models for failure prediction are under development based on supervised learning. Supervised learning methods are nowadays powerful tools to derive models with high prediction performance if complete training sets are available. The creation of such complete labelled training sets for switch current consumption is a challenging task due to the large amount of different types of switch failures and significant variances in “normal” switch behaviour due to differences in the environmental conditions and technical realizations. The evaluation/validation of these derived prediction models for individual switches based on the available data sets is therefore quite challenging and the generalization of the derived models to the same switch after maintenance or to other switches of the same type is furthermore not solved. As a complement to the work on prediction models based on supervised learning a data-driven approach without labelled training set is implemented and evaluated here.

3.5.4.2. Methodology and methods / Proposed solution

Significant effort was and is spent into implementing switch failure prediction based on data-driven models on labelled training data sets. Even so models with high prediction performance can already be obtained by supervised learning for example data sets such models are not yet in common operation due to challenges with the training sets as summarized above. Therefore, a different and complementing approach is investigated in In2Rail WP9. Here an approach not relying on a training data set with labelled switch failures will be implemented and evaluated for its failure forecasting capability. The first-order purpose of the switch failure detection model based on Statistical Process Control (SPC) developed in In2Rail WP6 (see In2Rail D6.4) is to automatically generate alarms (a fault/malfunction is affecting the asset) not relying on a manual setting of object-specific thresholds and reference curves (state of the art of switch monitoring systems in operation).

Basic concept for the forecasting approach under investigation here is to use the output of the switch failure detection SPC model also to provide a short- to mid-term (several days to few weeks) forecast of emerging switch failures in terms of automatically creating alerts (Atienza, Ang, & Tang, 1997). Aim for the first step is the detection of unusual behaviour in a very early stage based on statistical rules applied to the output of the SPC model to create alerts for the maintenance engineers. By this approach focus is laid on switches with potential emerging malfunctions. For this approach, no historic failure data set and no knowledge about the underlying failure types and their typical degradation behaviour (if existent) are necessary. All types of failures which are affecting the monitored parameters (here current drawn by the electrical motor) can be detected and will create failure forecasts in terms of alerts. In a second step (beyond In2Rail) the SPC output can be utilized to automatically detect systematic variations (degradation development) if the existent failure type causes such a stable trend in the measured parameter(s).

The data-driven switch failure detection model is based on Statistical Process Control (SPC) which utilizes a robust PCA to transform a large amount of features/predictors to only two quantities, the Hotelling-Parameter T^2 and the Squared Prediction Error SPE, which are then used to identify anomalies in terms of unusual behaviour of the predictors .The failure detection is then realized by the detection of outliers or unexpected behaviour (e. g. systematic variations) in the SPC output parameters, especially the SPE value, based on statistical rules and confidence intervals. Main advantage of this approach is that no a priori knowledge of switch failures and their characteristics (a labelled training set) is necessary as failures are detected in terms of unusual behaviour of the predictors. A SPC model is built by utilizing data of normal operation, indeed failures have to be explicitly excluded before modelling. In this way, a SPC model can be trained within a short time of operation shortly after installing or maintaining a switch.

Nevertheless, if emerging switch failures are affecting power consumption both output parameters of the SPC model show systematic trends (e. g. instead of following a normal distribution) indicating unusual behaviour of increasing magnitude before the switch failure finally occurs. Basic concept of the approach investigated here is to identify these trends in an early stage and to forecast their further development to forecast emerging switch failures. The forecasting of a large variety of different types of switch failures is therefore “reduced” to a single parameter time series forecasting problem. The performance of the forecasting in terms of switch failures is mainly influenced by the completeness and quality of the feature set (including normalisation) as input of the SPC model. Relevant influences not represented within the feature set will negatively influence the forecasting performance and must be further investigated and handled.

3.5.5 Analysis

The input to train the SPC model are the features obtained from switch movements measured within a time frame in which no failures were reported, which is referred to as the training set. The analysis presented here focuses on switch 3076 and includes both directions of blade movement. The training set for switch 3076 consists of nearly all measured current curves between July 1st, 2014 and July 1st, 2015. Since this time frame contains current graphs representing unusual behavior (which did not cause a switch malfunction) an automated outlier removal based on two selection criteria were applied to the training data set before model building. One criterion is based on the total duration of the current curve and the other on the total power consumption. The selection criteria thresholds are derived from statistics of the training set and do not depend on any manually selected thresholds.

The output parameters T^2 and SPE of the model trained for switch 3076 in the movement direction 0 are shown in Figure 3.24 and Figure 3.25. The parameters based on the model trained in the movement direction 1 are shown in Figure 3.26 and Figure 3.27, respectively. Each dot represents a measured current curve. Dark black colored points belong to the training data set, light black dots are current curves outside the training data set. If the T^2 and SPE parameter values of the training set are normally distributed, confidence intervals for different probabilities around the mean value are calculated based on the standard deviation of the distribution. Horizontal lines indicate the mean value of the training set (blue), 68.2% CI (green), 95.4% CI (yellow) and 99.7 % CI (red) in the figures and help identify odd switch behavior. The confidence intervals are decisive for the success of the forecasting based on the SPC output, therefore it is foreseen that the validity of the normal distribution assumption needs to be further explored and tested, and other distribution types considered.

Twelve vertical lines in bright colors (two nearly overlap, thus only eleven can be visually be identified) indicate reported switch malfunctions (note there is none in the training set) and (thinner) vertical lines in light pink indicate maintenance actions. The maintenance actions consist of different activities and in many cases, are not related to Switch Panel part where influence of it can be detected.

As can be observed in the presented figures, the training set data points mostly lay within the 99.7% CI delimited by the red horizontal lines. Most of the data points outside the 99.7% CI are related to time windows in which the 12 malfunction incidents were reported. Forty-four maintenance actions of different nature (mechanical – e. g. performed on the tracks or the geometry, and signaling – e. g. performed on the contacts or the engine) and with different implications for the operation of the switch are indicated in these figures (light pink vertical lines). From these figures, it is observed that there are differences in the model output parameters between the blade moving directions. To link the failure reported incidents with the results of the SPC models (for both movement directions) a closer look into single incidents is necessary (see Figure 3.28 to Figure 3.31).

The shaded vertical colorful areas in Figure 3.28 to Figure 3.31 indicate the time between a switch malfunction being reported and being repaired: reported on December 26th 2013 and repair time of about 19 hours (purple); reported on January 3rd 2014 and the repair time of 2 hours and 10 minutes (brown); reported on February 27th 2014 and repair time of nearly 2 hours and 15 minutes (blue). The faint pink vertical shaded areas indicate all reported mechanical and/or signaling maintenance actions performed on the switch. There is no information available on the maintenance duration, thus the width of these shaded areas is set by default to 24 hours.

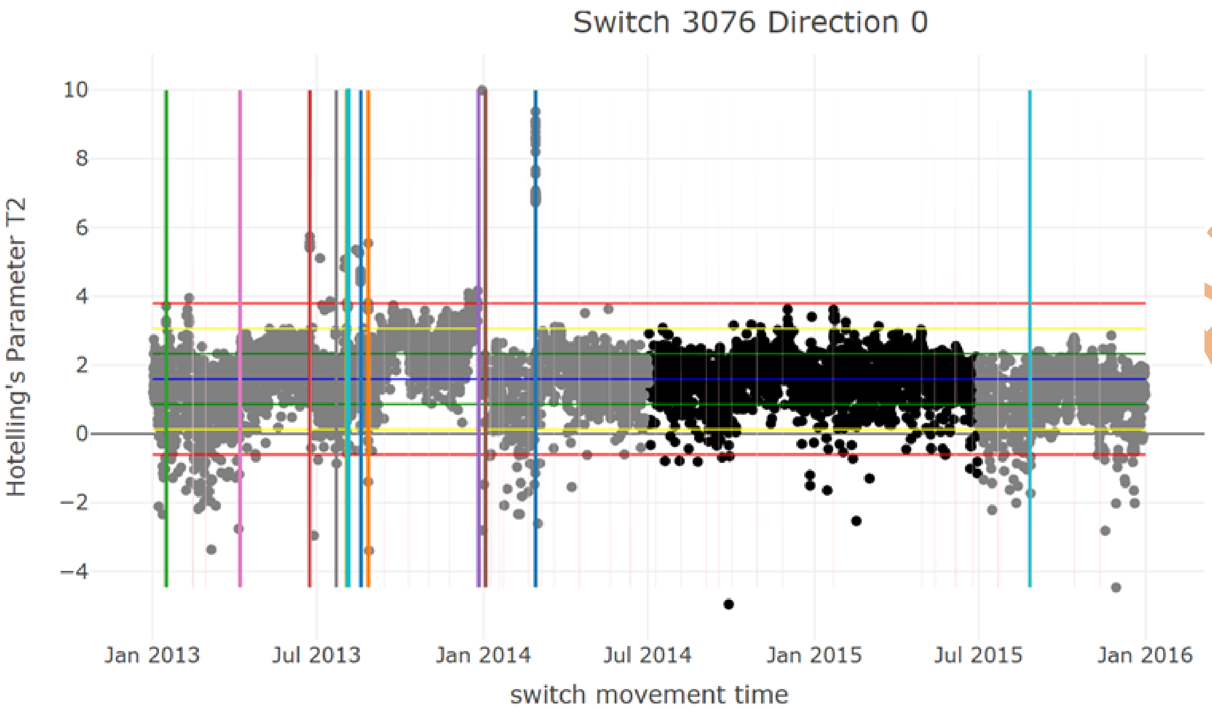


Figure 3.24: T2 parameter of switch 3076 and movement direction 0 throughout the acquisition time. Dark black colored points belong to the training data set, light black dots are current curves outside the training data set

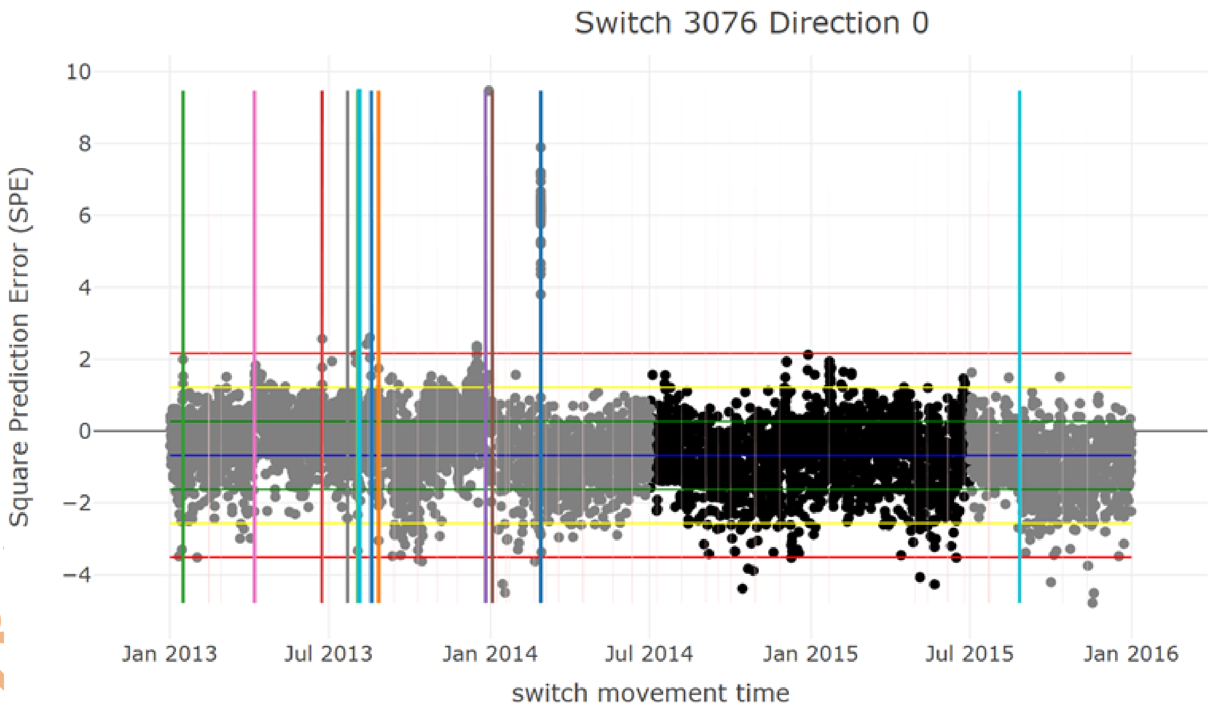


Figure 3.25: SPE parameter of switch 3076 and movement direction 0 throughout the acquisition time

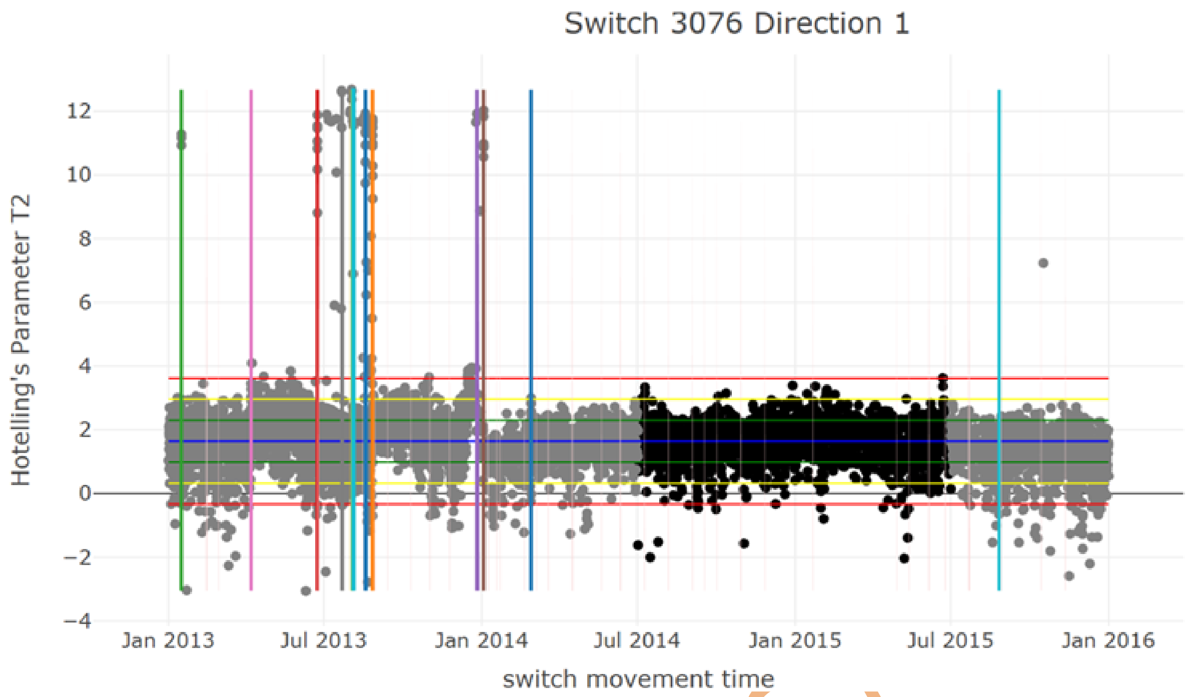


Figure 3.26: Same as Figure 3.24 but for movement direction 1

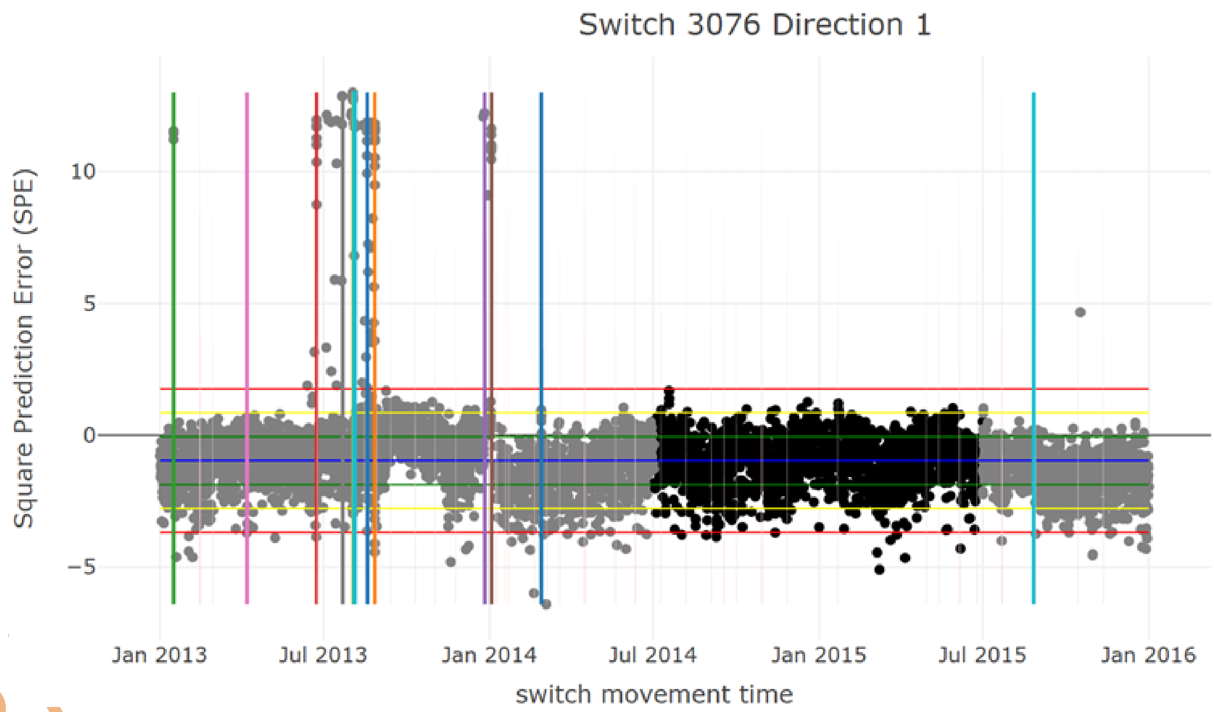


Figure 3.27: same as Figure 3.25 but for movement direction 1

The failure reported on December 26th, 2013 (in purple) was due to a rusting gear box. A water leak over a longer time caused the degradation of the gear box. Over a time of several weeks an increase of power consumption especially during switch blade movement can be observed. In this case first alerts could have been raised days before the malfunction finally occurred, based on the SPC model results. That is, in the 0-direction the T^2 and SPE values cross and exceed the 99% CI (so-called outliers) for the first time on December 15th since the previous big-mechanical maintenance action (on November 1st). In the 1-direction there are some T^2 and SPE outliers found after November 1st and before the failure, however there is no systematic trend around those outliers. What is striking is the systematic increase especially clear in the T^2 parameter evolution in both directions, which starts a few days before December 15th. This trend reflects the steadily increasing power consumption due to the degrading gear box. The model for direction 1 calculates extremely high parameter values one hour before the malfunction was reported. Clearly all these indications (outliers, systematic trends followed by extreme outliers) can be used to build statistical rules to be implemented for failure forecast.

After the 26th of December reported malfunction (purple) and its consecutive repair, the switch performed no movements during three consecutive days. The next movement after the repair was identified as an extreme outlier detected in both directions according to both SPC output parameters. From the movements that followed and that took place before the next reported failure on January 3rd (in brown), two are detected as extreme outliers by the model in the direction 0. This could indicate that the repair actions conducted for the December 26th malfunction (purple) did not fully solve all the mechanical/signalling problems that initiated in mid-December.

After the January 3rd malfunction (in brown) was reported to be repaired, the switch was not used for four days, until the switch engine was given maintenance on January 8th. After this maintenance action, the switch resumed its normal operation until February 27th, when the next malfunction (in blue) was reported. Even though there are a few “isolated” outliers in both directions and both parameters, no clear systematic trend can be identified. Nevertheless, in direction 0 the model identifies the failure through extreme outliers. This indicates that some types of failures, like the one that occurred on February 27th (blue) which was not a mechanical one, might not be able to be forecasted by condition monitoring based on current measurements, even though they are detected right when they occur. The malfunction on February 27th was caused by a burned electrical contact belonging to the control part of the switch. This type of failure is not detectable by the monitoring system.

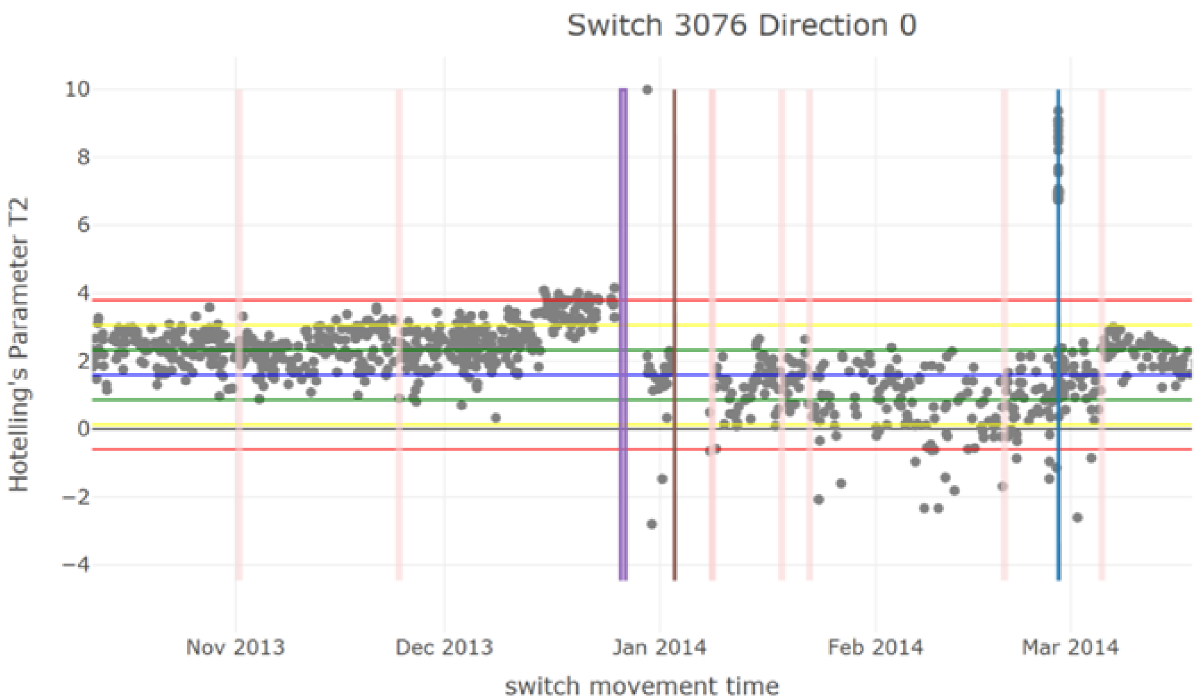


Figure 3.28: T2 parameter of switch 3076 and movement direction 0 between mid-October 2013 and mid-March 2014

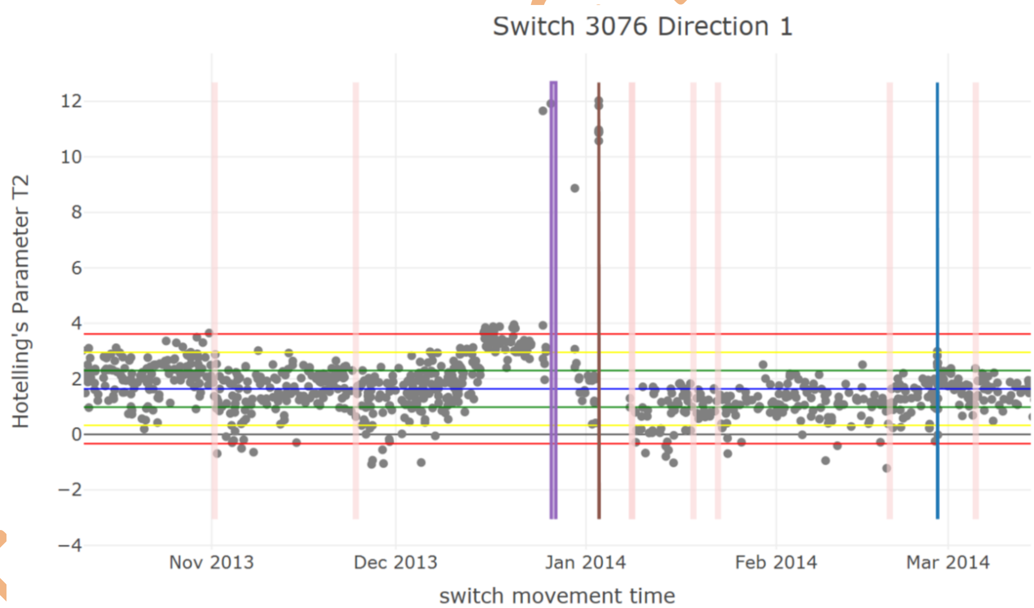


Figure 3.29: T2 parameter of switch 3076 (direction 1) between mid-October 2013 and mid-March 2014

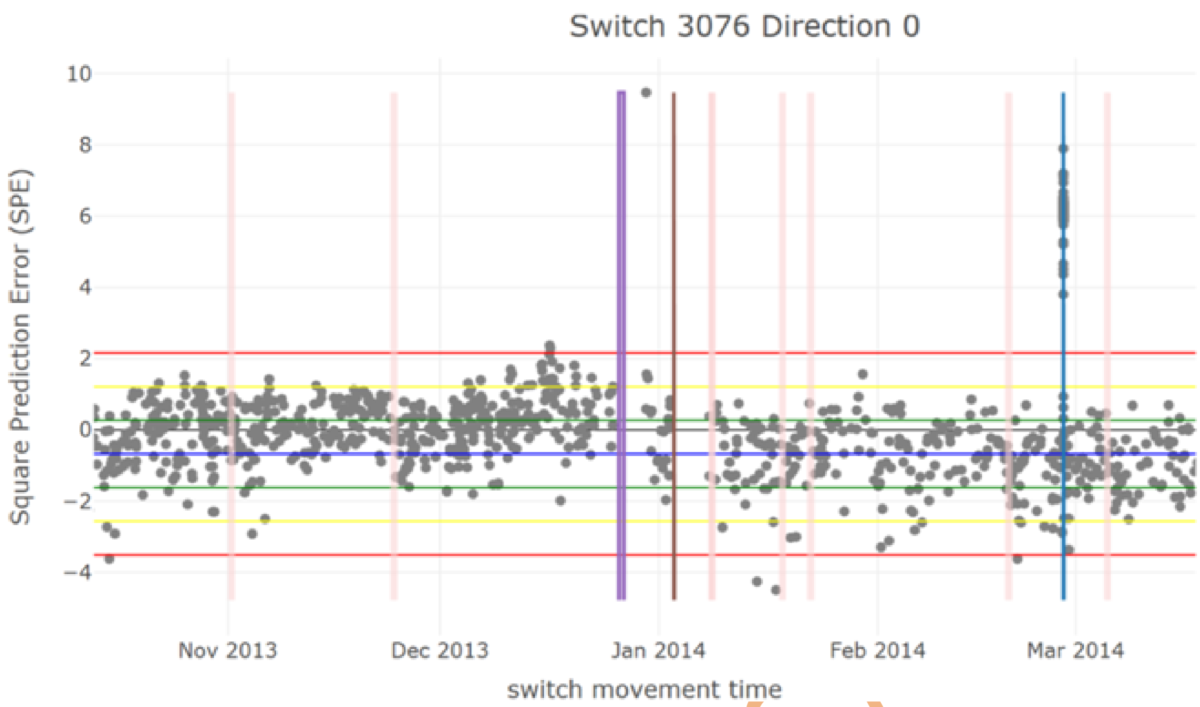


Figure 3.30: SPE of switch 3076 (direction 0) between mid-October 2013 and mid-March 2014

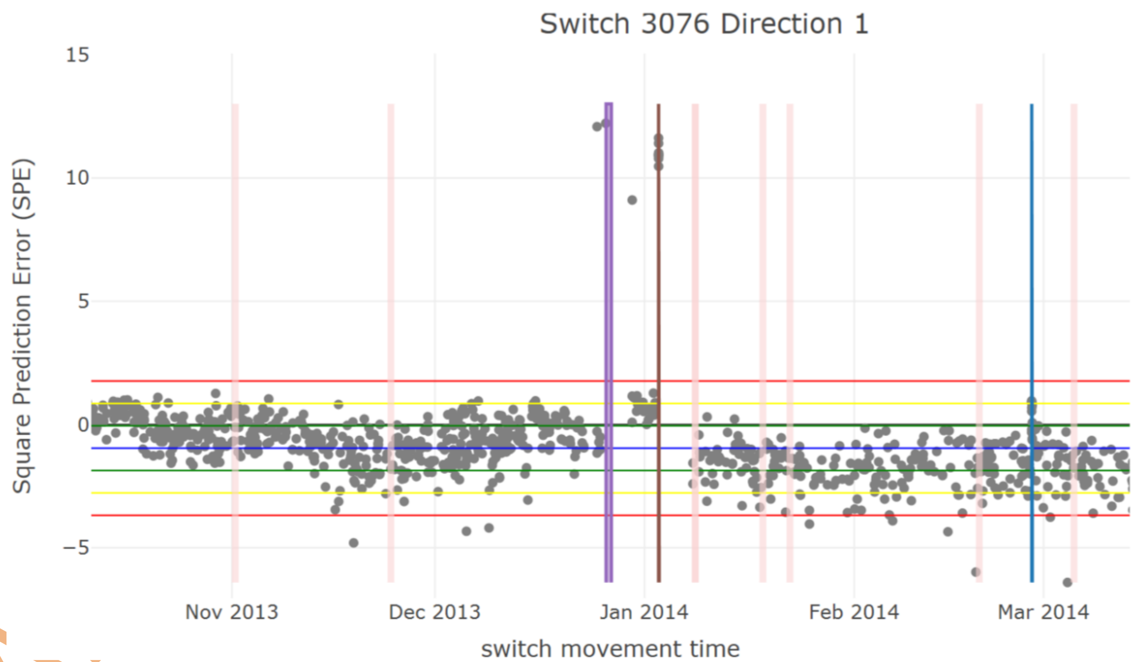


Figure 3.31: SPE of switch 3076 and (direction 1) between mid-October 2013 and mid-March 2014

3.5.6 Discussion

One significant advantage of the SPC model over the state of the art is that no manual (selected by experts) threshold setting for individual asset objects is necessary. It can be seen as an unsupervised black box model (no labelled abnormal behavior is necessary) based on historical data representing the asset object normal behavior (no failures detected in this period) to automatically detect unusual behavior. The detailed analysis of the SPC model output parameters shows that the approach is capable to provide valuable information for switch failure forecasting. The combination of both parameters (SPE and T2) reliably revealed unusual behaviour before final switch failures occurred. The performance is already outperforming human expert analysis by identifying unusual behaviour within the typical variations of normal behaviour due to environmental conditions. A forecasting in terms of generating alerts can be realized by generic statistical rules applied to the SPC model output. The far most of the up to now analysed switch failures do not cause systematic evolving variations of the power consumption. Therefore, the forecasting is limited in the current state of development to the generation of alerts without providing an estimated remaining time to malfunction for these failure types. Nevertheless, some failure types such as the analysed malfunction due to a rusting gear box are causing systematic variations of current consumption which are represented by systematic variations of the SPC output parameters. For these failure types, an estimation of the remaining time to malfunction may be realized by a trend detection and extrapolation and/or time series forecasting approaches.

The research conducted in In2Rail WP9 verified that this approach to overcome the difficulties related to the model building based on training sets with failure data provides valuable forecasting capabilities. Nevertheless, several issues must be addressed and further improved for the operationalization. Main point is the further development of the statistical rules to automatically and reliably generate alerts. This includes the implementation of trend detection and extrapolation to provide estimates of the remaining time to failure if systematic variations of the model output parameters are detected. To further improve the performance of the model to detect unusual behaviour more of the relevant environmental conditions (such as rain fall or exposure to the sun) should be included to the feature scaling. To do so additional data sources with informative parameters must be identified.

These open issues to further improve and operationalize this approach of switch failure/status forecasting will be addressed in the Shift2Rail project In2Smart.

4. Conclusion

This document represents the final report for part of the activities undertaken in the second task of WP9, Task 9.2 "Asset status forecasting for TMS/dispatching system", concluding at month 30 of the project. This deliverable provides the development of forecasting methodologies for the estimation of the status of a selected range of infrastructure railway assets playing a key role for TMS/maintenance.

As an output of all these activities, the WP9 partners have designed a set of methodologies based on data-driven approaches that provide nowcasts and forecasts of the status of the considered assets. It is worth pointing out that this document contains the description of each of the five forecasting scenarios as given below:

1. forecasting of delay attributions based on train movements records (proposed by NR) (theoretical);
2. forecasting of Switch & Crossing probability of failure (proposed by TRV and LTU);
3. derailment risk assessment through wheel-rail contact forces nowcasting (proposed by UPORTO, IP, EVOLEO and ViF);
4. prediction of time to restoration for different assets and different failures based on maintenance/repair reports (by SR and UNIGE);
5. Switch & Crossing asset status forecasting (proposed by SR and DLR).

The preliminary results achieved through laboratory testing are promising for each of the scenarios, except for the one proposed by NR. The conclusion of each of the scenario is mentioned below.

In NR scenario, it deals with the problem of assessing the impact of several different types of asset failures on the traffic by analyzing the relevant data related to train movements, assets and their failures, and delay attributions (i.e. delay effects). The analysis could be extended by integrating weather data, maintenance data and train characteristics (e.g. train composition and weight) to be associated with data about train movements. The proposed solution could be used in real time by the TMS in order to immediately have a measure of the impact of an asset failure to the railway traffic / operations.

In TRV/LTU scenario, the probabilities are estimated by considering the standard deviation and threshold for tamping of the track geometry of three panels of S&C. The developed models based on the particle-filter based approach are useful to predict the forecast for future condition. Three S&Cs were considered for training purposes for estimating the parameters of ORE model. The true RUL was predicted for all 4 S&Cs with uncertainties. It was also concluded that the true RUL will change according to the new data point and uncertainty will also change accordingly. This approach can be applied to remaining S&Cs within the same track section for validation of the developed model.

In ViF/UPORTO/UNIGE scenario, the potential of a fast calculation method to forecast the risk of derailment is shown. This is done by three main steps. First, the input parameters are forecasted according the desired prediction horizon. Therefore, the track geometry irregularities are forecasted by a so-called "ViF Track Geometry Degradation model". Furthermore, the wind speed distribution is assumed and vehicle speed classes are defined. Second, the forecasted parameter sets are used as input for the nowcasting method described in D9.3. Third, the forecasted risk of derailment values along the track are post processed and build the decision basis for TMS/Maintenance. It is also shown, that the method can be easily extended by further input parameters. If data of measurement wheelsets is available in the future, the method can be validated and optimised if necessary.

In SR/UNIGE scenario, the main objective is to forecast possible failures of assets based on the correlation of past asset failures and past weather conditions or maintenance actions, considering a set of different infrastructure assets selected as the most relevant ones from the TMS perspective. Kernel Regularized Least Squares and permutation tests were conducted to extract relevant features from the available database. The results showed the forecasting performance, which relate to assessing the performance of the data-driven models in predicting the interesting time quantities through the exploitation of historical data and feature ranking, which is a statistical analysis for assessing the significance of each input feature of a dataset for prediction.

In SR/DLR-scenario, it is shown that the new approach based on a Statistical Process Control (SPC) model is capable to provide valuable information for detection of abnormal behaviour and switch failure forecasting. The model output parameters reliably revealed unusual behaviour before final switch failures occurred if the underlying failure type affects the power consumption of the switch (the monitored parameter). The performance is already outperforming human expert analysis by identifying unusual behaviour within the typical variations of normal behaviour due to environmental conditions. A forecasting, in terms of generating alerts, can be realized by generic statistical rules applied to the SPC model output. The far most of the up to now analysed switch failures do not cause systematic evolving variations of the power consumption. Therefore, the forecasting is limited in the current state of development to the generation of alerts without providing an estimated remaining time to malfunction for these failure types. Nevertheless, some failure types such as the analysed malfunction due to a rusting gear box are causing systematic variations of current consumption which are represented by systematic variations of the SPC output parameters. For these failure types an estimation of the remaining time to malfunction may be realized by a trend detection and extrapolation and/or time series forecasting approaches. This and other identified open issues identified to further improve and operationalize this approach of switch failure/status forecasting will be addressed in the Shift2Rail project In2Smart. This will include a systematic coupling of this data-driven forecasting approach with the switch functional model developed in In2Rail WP6 to provide in-depth diagnostic information about the detected present or emerging failures.

The final validation of the preliminary results included in this deliverable will be carried out by the end of the In2Rail project (i.e. Month 36) and will be described in the Deliverable 9.5 – "Nowcasting and Forecasting algorithms verification, evaluation and assessment report".

5. References

[1] J. Gareth, D. Witten, T. Hastie, and R. Tibshirani. An introduction to statistical learning. New York: Springer, 2013.

[2] I. Guyon, and S. Elisseeff. An introduction to variable and feature selection. The Journal of Machine Learning Research 3 (2003): 1157-1182.

[3] H. Zou, and T. Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, no. 2 (2005): 301-320.

[4] K. Yoon, K. Yang, and C. Shahabi. Feature subset selection and feature ranking for multivariate time series. Knowledge and Data Engineering, IEEE Transactions on 17, no. 9 (2005): 1186-1198.

[5] Y. W. Chang, and C. J. Lin. Feature ranking using linear svm. Causation and Prediction Challenge Challenges in Machine Learning 2 (2008): 47.

[6] F. Takens "Detecting strange attractors in turbulence." In Dynamical systems and turbulence, Warwick 1980, pp. 366-381. Springer Berlin Heidelberg, 1981.

[7] C. Chatfield. The analysis of time series: an introduction. CRC press, 2013.

[8] H. Lutkepohl. New introduction to multiple time series analysis. Springer Science & Business Media, 2005

[9] D. C. Montgomery, L. A. Johnson, and J. S. Gardiner. Forecasting and time series analysis. McGraw-Hill Compa- nies, 1990.

[10] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. Time series analysis: forecasting and control. John Wiley & Sons, 2011.

[11] P. Bloomfield. Fourier analysis of time series: an introduction. John Wiley & Sons, 2004.

[12] J. D. Hamilton. Time series analysis. Princeton university press Princeton, 1994.

[13] F. Mosteller, J. W. Tukey. "Data analysis and regression: a second course in statistics." Addison-Wesley Series in Behavioral Science: Quantitative Methods (1977).

[14] T. G. Dietterich "Ensemble methods in machine learning." In International workshop on multiple classifier systems, pp. 1-15. Springer Berlin Heidelberg, 2000.

[15] C. M. Bishop. Neural networks for pattern recognition. Oxford university press, 1995.

[16] J. Shawe-Taylor, and C. Nello. Kernel methods for pattern analysis. Cambridge university press, 2004.

[17] T. Cover, and H. Peter "Nearest neighbor pattern classification." IEEE transactions on information theory 13, no. 1 (1967): 21-27.

[18] S. R. Safavian, and D. Landgrebe. "A survey of decision tree classifier methodology." (1990).

[19] V. N. Vapnik. Statistical learning theory. Vol. 1. New York: Wiley, 1998.

[20] C. C. Aggarwal. Data mining: the textbook. Springer, 2015.

[21] T. Hastie, R. Tibshirani, and Jerome Friedman. "Unsupervised learning." In The elements of statistical learning, pp. 485-585. Springer New York, 2009.

[22] A. K. Jain, M. N. Murty, P. J. Flynn (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

[23] V. Chandola, A. Banerjee, and V. Kumar. "Anomaly detection: A survey." ACM computing surveys (CSUR) 41, no. 3 (2009): 15.

[24] Vapnik, V. N. (1998). Statistical learning theory. Wiley New York.

[25] Anguita, D., Ghio, A., Oneto, L., & Ridella, S. (2012). In-sample and out-of-sample model selection and error estimation for support vector machines. IEEE Transactions on Neural Networks and Learning Systems, 23 (9), 1390-1406.

[26] Bartlett, P. L., Boucheron, S., & Lugosi, G. (2002). Model selection and error estimation. Machine Learning, 48 (1-3), 85-113.
[27] P. L. Bartlett, S. Mendelson, Rademacher and gaussian complexities: Risk bounds and structural results, Journal of Machine Learning Research 3 (2002) 463–482.
[28] P. L. Bartlett, O. Bousquet, S. Mendelson, Local rademacher complexities, Annals of Statistics 33 (4) (2005) 1497–1537.
[29] S. Floyd, M. Warmuth, Sample compression, learnability, and the vapnik-chervonenkis dimension, Machine learning 21 (3) (1995) 269– 304.
[30] J. Langford, D. McAllester, Computable shell decomposition bounds, Journal of Machine Learning Research 5 (2004) 529–547.
[31] O. Bousquet, A. Elisseeff, Stability and generalization, Journal of Machine Learning Research 2 (2002) 499–526.
[32] T. Poggio, R. Rifkin, S. Mukherjee, P. Niyogi, General conditions for predictivity in learning theory, Nature 428 (6981) (2004) 419–422.
[33] L. Oneto, A. Ghio, S. Ridella, D. Anguita, Fully empirical and data- dependent stability-based bounds, IEEE transactions on cybernetics 45 (9) (2015) 1913–1926.
[34] G. Lever, F. Laviolette, F. Shawe-Taylor, Tighter pac-bayes bounds through distribution-dependent priors, Theoretical Computer Science 473 (2013) 4–28.
[35] I. O. Tolstikhin, Y. Seldin, Pac-bayes-empirical-bernstein inequality, in: Neural Information Processing Systems, 2013.
[36] P. Germain, A. Lacasse, M. Laviolette, A. ahd Marchand, R. J. F., Risk bounds for the majority vote: From a pac-bayesian analysis to a learning algorithm, JMLR 16 (4) (2015) 787–860.
[37] L. B´egin, P. Germain, F. Laviolette, J. F. Roy, Pac-bayesian bounds based on the r´enyi divergence, in: International Conference on Artificial Intelligence and Statistics, 2016.
[38] S. Nitzan, J. Paroush, Optimal decision rules in uncertain dichotomous choice situations, International Economic Review 23 (2) (1982) 289–97.
[39] O. Catoni, Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning, Institute of Mathematical Statistics, 2007.
[40] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, A. Roth, Preserving statistical validity in adaptive data analysis, in: arXiv preprint arXiv:1411.2664 (v3), 2014.
[41] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, A. Roth, Preserving statistical validity in adaptive data analysis, in: Annual ACM Symposium on Theory of Computing, 2015.
[42] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, A. Roth, Generalization in adaptive data analysis and holdout reuse, in: arXiv preprint arXiv:1506.02629 (v2), 2015.
[43] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, A. Roth, Generalization in adaptive data analysis and holdout reuse, in: Neural Information Processing Systems, 2015.
[44] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, A. Roth, The reusable holdout: Preserving validity in adaptive data analysis, Science 349 (6248) (2015) 636–638.
[45] Takens, F. (1981). Detecting strange attractors in turbulence. Springer.
[46] Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. Physical Review Letters, 45 (9), 712.
[47] Vapnik, V. N. (1998). Statistical learning theory. Wiley New York.
[48] Shawe-Taylor, J., & Cristianini, N. (2004). Kernel methods for pattern analysis. Cambridge university press.

[49] L. Gyorfi, A distribution-free theory of nonparametric regression. Springer, 2002.
[50] D. Pollard, Empirical processes: theory and applications, in NSF-CBMS regional conference series in probability and statistics, 1990.
[51] A. Caponnetto and E. De Vito, Optimal rates for the regularized least-squares algorithm, Foundations of Computational Mathematics, vol. 7, no. 3, pp. 331–368, 2007.
[52] S. S. Keerthi and C.-J. Lin, Asymptotic behaviors of support vector machines with gaussian kernel, Neural computation, vol. 15, no. 7, pp. 1667–1689, 2003.
[53] L. Oneto, A. Ghio, S. Ridella, and D. Anguita, Support vector machines and strictly positive definite kernel: The regularization hyperparameter is more important than the kernel hyperparameters, in International Joint Conference on Neural Networks, 2015.
[54] Bisio, F., Gastaldo, P., Zunino, R., & Cambria, E. (2015). A learning scheme based on similarity functions for affective common-sense reasoning. In International joint conference on neural networks.
[55] Huang, G. B., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Transactions on Neural Networks, 17 (4), 879-892.
[56] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In IEEE international joint conference on neural networks.
[57] Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 42 (2), 513-529.
[58] G. Lever, F. Laviolette, and F. Shawe-Taylor, "Tighter pac-bayes bounds through distribution-dependent priors," Theoretical Computer Science, vol. 473, pp. 4–28, 2013.
[59] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," The Annals of Statistics, vol. 26, no. 5, pp. 1651–1686, 1998.
[60] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, Bayesian data analysis. Taylor & Francis, 2014.
[61] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.
[62] Kecman, P., & Goverde, R. M. P. (2015). Online data-driven adaptive prediction of train event times. IEEE Transactions on Intelligent Transportation Systems, 16 (1), 465-474.
[63] Basile, Olivier, Pierre Dehombreux, and Fouad Riane. "Identification of reliability models for non-repairable and repairable systems with small samples." In Proceedings of IMS2004 Conference on Advances in maintenance and modeling, simulation and intelligent monitoring of degradation, Arles, pp. 1-8. 2004.
[64] NIST, "Non-Homogeneous Poisson Process (NHPP) - Power Law". Itl.nist.gov. N.p., 2017. Web. 13 Feb. 2017.
[65] "Minitab Reliability and Survival Analysis". Pinzhi.org. N.p., 2017. Web. 13 Feb. 2017.
[66] Beyer, Mark A., and Douglas Laney. "The importance of 'big data': a definition." Stamford, CT: Gartner (2012): 2014-2018.
[67] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
[68] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). The CRISP-DM user guide. In 4th CRISP-DM SIG Workshop in Brussels in March (Vol. 1999).

[69] P. Good, Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer Science & Business Media, 2013.
[70] H. Deng, G. Runger, and E. Tuv, “Bias of importance measures for multi-valued attributes and solutions”, Artificial Neural Networks and Machine Learning–ICANN 2011, pp. 293–300, 2011.
[71] A. P. White and W. Z. Liu, “Technical note: Bias in information-based measures in decision tree induction”, Machine Learning, vol. 15, no. 3, pp. 321–329, 1994.
[72] ORE, D. (1988). 161 rp4: Dynamic vehicle/track interaction phenomena from point of view of track maintenance. Utrecht, September.
[73] Mishra, M., Odelius, J., Thaduri, A., Nissen, A., & Rantatalo, M. (2017). Particle filter-based prognostic approach for railway track geometry. Mechanical systems and signal processing, 96, 226-238.

6. Appendix 1

6.1 A1: Scenario by SR/UNIGE

Detailed Description of Available Data

For what concern the maintenance/repair reports dataset, we have the following information available for each failure:

Failure	Description
Priority_Code	A numerical code that indicates the urgency of the failure. This indicates when the failure has to be fixed. 2 means it has to be resolved 1:45 after it was reported
Geo-Code	The code of the Geographical location of the failed asset
Km-Location	Better localization of the failure location on the track
Failure_Type	generically classification of the (main) problem
Object-Type	The main type of the asset (like Switch, Track)
ObjectID	Unique ID of the asset which failed
Reference_number_SR	Reference ID (internal Database number)
Technical_Department	Department responsible for fixing the failure
DT_notification	**Date** and Time when failure was reported
DT_Mechanic_informed	Date and Time when the repair team was informed about failure
DT_Mechanic_on-location	Date and Time when the repair team was arrived on the site / location
DT_Starting-repair	Date and Time when the repair team started fixing the problem
DT_function-restored	Date and Time when the failure was fixed, and function of the asset restored
DT_Repair_wanted	Date and Time the problem to be fixed as notified by the train operator. Only for urgency 5
Year	The Year when the problem was reported (extracted from the date)
Month	The month when the problem was reported (extracted from the date)
ResponceTime(min)	Responce time in minutes; calcullated difference between [dt_notification] and [DT_Mechaninc_on-location]
RepairTime(min)	Responce time in minutes; calcullated difference between [DT_Mechaninc_on-location] and [dt_function-restored]
Function_restorationTime (min)	Total repair-time in minutes; calcullated difference between [dt_notification] and [DT_function_restored]
Part_Code	Failed part code / Failed part number
Part_description	Standardized description of the failed part
Action_carried_out_code-description	Description of the coded action (number) taken in order to solve the problem
Action_carried_out_code	Standardized number of the action taken (many numbers correspond to 1 description)
Failure_cause_main_group	Main group of the failure cause; like Mechanical problem (Infra) or vandalism (Others)
Failure_cause_code	Standardized number of the fail-cause
Failure_cause_Description	Failure cause; Standard text
GEO_Shape_Length	Length of the railway section
X(Long.)_Begin	Longitude of the starting point of the railway section
Y(Lat.)_Begin	Latitude of the starting point of the railway section
X(Long.)_Mid	Longitude of the middle point of the railway section
Y(Lat.)_Mid	Latitude of the middle point of the railway section
X(Long.)_End	Longitude of the end point of the railway section
Y(Lat.)_End	Latitude of the end point of the railway section

For what concern the weather data, for each weather station sample every day we have the following information available:

Weather data	Information
STN	Weather Station number
YYYYMMDD	Date (YYYY=Year; MM=month; DD=day)
DDVEC	Vector Average wind direction in degrees (360=north, 90=east; 180=south;270=west, 0=no wind/variable)
FHVEC	Vector Average wind speed (in 0.1m/s)
FG	Day Average wind speed (in 0.1m/s)
FHX	Highest hour average wind speed (in 0.1m/s)
FHXH	Hour in which FHX was measured
FHN	Lowest hour average wind speed (in 0.1m/s)
FHNH	Hour in which FHN was measured
FXX	Highest wind blast (in 0.1 m/s)
FXXH	Hour in which FXX was measured
TG	Day Average temperature (in 0.1 degrees Celsius)
TN	Minimum temperature (in 0.1 degrees Celsius)
TNH	Hour in which TN was measured
TX	Maximum temperature (in 0.1 degrees Celsius)
TXH	Hour in which TX was measured
T10N	Minimum temperature on 10 cm altitude above ground (in 0.1 degrees Celsius)
T10NH	6-hours window in which T10N was measured ; 6=0-6 UT, 12=6-12 UT, 18=12-18 UT, 24=18-24 UT
SQ	Sunshine duration (in 0.1 hour) calculated frm global radiation (-1 for <0.05 hour)
SP	Percentage of the maximum possible sunshine duration
Q	Global radiation(in J/cm2)
DR	Duration of the rainfall (in 0.1 hour)
RH	Daily sum of the rainfall (in 0.1 mm) (-1 for <0.05 mm)
RHX	Highest hour-sum of the rainfall (in 0.1 mm) (-1 for <0.05 mm)
RHXH	Hour in which RHX was measured
PG	Day-Average air pressure reduced to sea level (in 0.1 hPa) calculated from 24-hours value
PX	Highest hour-value of the air pressure reduced to sea level (in 0.1 hPa)
PXH	Hour in which PX was measured
PN	Lowest hour-value of the air pressure reduced to sea level (in 0.1 hPa)
PNH	Hour in which PN was measured
VVN	Minimum occurred sight; 0: <100 m, 1:100-200 m, 2:200-300 m,..., 49:4900-5000 m, 50:5-6 km, 56:6-7 km, 57:7-8 km,..., 79:29-30 km, 80:30-35 km, 81:35-40 km,..., 89: >70 km)
VVNH	Hour in which VVN was measured;
VVX	Maximum occurred sight; 0: <100 m, 1:100-200 m, 2:200-300 m,..., 49:4900-5000 m, 50:5-6 km, 56:6-7 km, 57:7-8 km,..., 79:29-30 km, 80:30-35 km, 81:35-40 km,..., 89: >70 km)
VVXH	Hour in which VVX was measured
NG	Day Average overcast (coverage of the upper air into eighths, 9=sky is invisible)
UG	Day Average relative humidity (in %)
UX	Maximum relative humidity (in %)
UXH	Hour in which UX was measured
UN	Minimum relatieve humidity (in %)
UNH	Hour in which UN was measured
EV24	Reference Crop Evaporation (Makkink) (in 0.1 mm)

Moreover, the weather station dataset give us information regarding:

- STN: Weather Station number ;
- Longitude ;
- Latitude ;
- Altitude.

Feature Extraction Notation

Concerning extracted features, the following rules for a common notation have been defined:

- Temporal Intervals features :
  - for each event timestamp included in the original data, the duration between events that are adjacent in time is computed;
- weather features :
  - the mean of weather values are computed for 7, 30 and 90 days before a maintenance/repair action is carried out,
  - for each weather feature, a suffix is appended: "_av7", "_av30" and "_av90",
  - "_av1" stands for the mean of the weather values for the current day in which the maintenance/repair intervention is carried out,
  - "CT_start", "CT_end", "Season" features have been added to the original features; they identify time of dawn and sunset, and season for the current day in which the maintenance/repair intervention is carried out ;
- past failures/malfunctions features :
  - Object_failures → past failures for a specific ObjectID,
  - Geo_failures → past failures in the area identified by the same GeoCode,
  - Area_failures → past failures in a 1km2 circular area surrounding the failure/malfunction under examination,
  - _7, _30, _365, _730, _3650 → strings added to identify the number of days considered before the current day in which the maintenance/repair intervention is carried out,
  - _FC → indicates that in the current feature only the failures/malfunctions related to the same Failure_cause_code are counted ;
- open Failures/malfunctions features
  - _s1, _s2, _s3, _s5 → strings added to identify that the number of open intervention is counted by only considering those that are active in a specific moment of the maintenance/repair process timeline. For instance, features where "_s5" is appended include all the actions that are still open between step 5 (Start Repair) and step 7 (Repair done – Function Restored) ,
  - _TD → indicates that in the current feature only the active failures/malfunctions in charge to the same Technical_Department are counted,
  - _in12, _in23, _in35, _in57 → these strings have a similar meaning to the first ones described (i.e. _s1, _s2, _s3, _s5), but they refer to the intervention which are currently active, and not to the one under examination,
  - _p1, _p2, _p4, _p5 → analogously, this strings are used to clarify that the active maintenance/repair actions are counted by filtering by priority level ;

- is_failed → if the maintenance/repair action has been already carried out on the same asset in the past ;
- time_from_last_fail → how long passed by the previous maintenance/repair action arried out on the same asset.

Feature Ranking for Repair Time modelling for each Simulation Scenario

Wheather - Priority2 - RepairTime	
P-value	**Feature**
1,000	STN
0,736	Object_Type
0,311	Failure_Type
0,299	FHN_av90
0,262	Day
0,247	TNH_av30
0,217	X(Long)_Begin
0,215	RH_av30
0,197	Action_carried_out_code
0,179	RHX_av30

Wheather - PriorityAll - RepairTime	
P-value	**Feature**
1,000	Reference_number_SR
0,823	TN_av90
0,433	Open_failures_s2_in12_p5
0,428	TD_Open_failures_s2_in12_p5
0,373	PNH_av7
0,330	Q_av1
0,271	SP_av1
0,241	SQ_av1
0,241	ToLocation_Time
0,206	TG_av90

No Weather - Priority2 - RepairTime	
P-value	**Feature**
1,000	STN
0,900	Object_Type
0,522	Action_carried_out_code
0,462	X(Long)_Mid
0,375	Reference_number_SR
0,355	Y(Lat)_Begin
0,353	T10N_av1
0,346	PX_av1
0,317	Y(Lat)_Mid
0,315	X(Long)_Begin

No Weather - PriorityAll - RepairTime	
P-value	**Feature**
1,000	T10N_av1
0,646	TXH_av1
0,368	DR_av1
0,314	PNH_av1
0,194	TX_av1
0,164	Day
0,118	TNH_av1
0,098	RHX_av1
0,089	Object_Type
0,075	TN_av1

Feature Ranking for Response Time modelling for each Simulation Scenario

Wheather - Priority2 - ResponseTime	
P-value	**Feature**
1,000	Reference_number_SR
0,356	TG_av7
0,345	Q_av30
0,182	EV24_av7
0,166	EV24_av30
0,165	NG_av7
0,153	time_from_last_fail
0,139	TN_av7
0,130	Open_failures_s1_in23_p4
0,126	FXXH_av30

Wheather - PriorityAll - ResponseTime	
P-value	**Feature**
1,000	Reference_number_SR
0,146	NG_av7
0,125	Q_av30
0,108	TG_av7
0,101	PN_av30
0,092	Open_failures_s1_in23_p4
0,091	EV24_av7
0,088	TD_Open_failures_s1_in23_p4
0,062	TN_av7
0,058	time_from_last_fail

No Weather - Priority2 - ResponseTime

P-value	Feature
1,000	EV24_av1
0,986	PX_av30
0,877	VVXH_av1
0,770	Reference_number_SR
0,522	CT_start
0,483	RHXH_av1
0,471	CT_end
0,462	RHX_av30
0,419	UXH_av1
0,366	ObjectID

No Weather - PriorityAll - ResponseTime	
P-value	**Feature**
1,000	PX_av1
0,782	T10N_av1
0,715	RHXH_av1
0,679	STN
0,658	Priority_Code
0,594	PG_av1
0,587	TNH_av1
0,387	RHX_av1
0,348	Max_Repair_Time
0,326	Day

Feature Ranking for Travel Time modelling for each Simulation Scenario

Wheather - Priority2 - TravelTime	
P-value	**Feature**
1,000	Reference_number_SR
0,275	TG_av7
0,256	Q_av30
0,225	EV24_av7
0,163	PN_av30
0,138	TD_Open_failures_s1_in23_p4
0,132	time_from_last_fail
0,124	Open_failures_s1_in23_p4
0,122	NG_av7
0,111	TN_av7

Wheather - PriorityAll - TravelTime	
P-value	**Feature**

1,000	Reference_number_SR
0,126	NG_av7
0,118	TG_av7
0,113	EV24_av7
0,096	PN_av30
0,094	Q_av30
0,081	TD_Open_failures_s1_in23_p4
0,077	Open_failures_s1_in23_p4
0,057	time_from_last_fail
0,045	TN_av7

No Weather - Priority2 - TravelTime	
P-value	**Feature**
1,000	EV24_av1
0,850	PX_av30
0,793	VVXH_av1
0,688	T10N_av1
0,635	Reference_number_SR
0,549	RHXH_av1
0,481	CT_start
0,478	CT_end
0,449	RHX_av30
0,444	UXH_av1

No Weather - PriorityAll - TravelTime	
P-value	**Feature**
1,000	PX_av1
0,800	T10N_av1
0,683	STN
0,656	RHXH_av1
0,595	Priority_Code
0,544	PG_av1
0,543	TNH_av1
0,374	RHX_av1
0,340	Day
0,328	X(Long)_Begin