Proceedings of the Sixth International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 7, Paper 8.1
Civil-Comp Press, Edinburgh, United Kingdom, 2024
ISSN: 2753-3239, doi: 10.4203/ccc.7.8.1

Maintenance Applications of a Machine Learning Model of Rail Defects

O. Vo Van[1] and V. Laurent[2]

[1]Direction Technologies, Innovation et Projets Groupe SNCF, France

[2]Centre Borelli, ENS Paris-Saclay, Université Paris-Saclay, France

Abstract

Rolling contact fatigue is the first cause of rail degradation. The phenomenon is deemed random yet depending on many exogenous parameters. Numerical physical modeling has shown its limits and data modeling had to be introduced. In this paper, one proposes to show diverse applications of a hybrid model using data and physics in an industrial context. The data architecture linking physical model, random forest classification model and survival forest model is quickly exposed. Then some indicators are proposed to eval the model performance. Three applications are then developed based on this mixed-model, phenomenon understanding - closely linked with model validation by expert knowledge -, predictive maintenance scheduling by rail grinding, and prescriptive maintenance illustrated by fixing the time delay before the first visit by ultrasound engines.

Keywords: rolling contact fatigue, rail, predictive maintenance, random forest, prescriptive maintenance

1 Introduction

Repeated passage of train wheels over the rail, with a small contact patch and high hydrostatic pressure, can lead to the development of cracks on the rail's surface or subsurface. This phenomenon, known as rolling contact fatigue (**RCF**), is the primary cause of rail degradation. Damages due to RCF, such as squats on the contact table and head-checks on the contact flange, can propagate and lead to rail fracture. The occurrence of rail fractures could have economic (delayed trains, maintenance costs) and passenger safety (derailment) implications. To avoid this event, the French National Railway Company (SNCF) applies a strict method with both corrective and scheduled maintenance strategies which have a high cost for the company.

Improvement on rail manufacturing, track monitoring and maintenance policies have led to drastically reduce the number of broken rails in the network. Over the last five years, less than a hundred of breaks appeared each year over the thirty thousand kilometres of track and thus, twice of rail. Most of breaks appear in non-commercial track and thus are generally not part of the maintained part of the network. Despite each defect caused by RCF can lead to a rail break, this is an event too rare to attempt to predict it. While the currently deployed ultrasonic monitoring system offers good results for detecting cracks before break, it is not sensitive enough to activate preventive maintenance such as rail grinding. The dreaded event thus slid from rail break to defect initiation.

Fatigue crack initiation is difficult to predict based on physical models. On one hand, the phenomenon is sensitive to a large amount of parameters, such as material crystallography, train load or dynamic stiffness of the track, which are difficult to consider altogether [1,2]. On the other hand, all these parameters are strongly variable. Most of these sizes are not systematically measured, such as the contact patch, friction conditions or local wear, which leads the inherently random fatigue phenomenon, to appear even more random. Physical models thus show their limits for maintenance applications.

To tackle this health monitoring problem, data modelling seems more promising. Several recent studies have shown its performance [3] and the possibility to consider physics in these models [4]. The present paper develops applications based on a model previously presented [5], which has been improved by complementary data and models such as developed in the next section. Three examples of applications are presented and then discussed. The first is a sensitivity analysis to both validate the model with expert knowledge and question the impact of various exogenous parameters that have been added. The second concerns prescriptive maintenance, illustrated by defining the first time an ultrasound monitoring vehicle should check rail defects. The last application is on predictive maintenance, for planning the grinding work on the most risky segments of the network.

2 Data & Method

In this section data used and their structure are briefly explained. Most of these data have been introduced in a previous paper [5] and well described in [3]. The machine learning model has also been introduced in [6] and is thus shortly developed here.

2.1 Data

Three different types of data have been identified among the available raw data.

- **Static properties**: Rail linear mass, rail grade, rail age, new or reused, curve inverse radius, declivity, sleepers types and space, tunnels, substructure type (ballast or slab track), rail joint. Ambient temperature can also be aggregated as a static property as shown in [7];

- **Time Events**: Inspection data (Rail defect first discover) and maintenance work (rail and/or track renewal, rail local replacement, rail grinding, ballast tamping);

- **Usage data**:Train load, train type (regional train, high speed train, freight), train velocity and acceleration, traffic.

Data are aggregated on rail segments with a fixed 108m-length. This value is motivated by the regular 36m-length of the rail provided by manufacturers and welded in workshop or on track in case of continuous welded rails (CWR). Taking a congruent of the latter is, on average, equivalent to having the same number of welds or of rail joints per segment.

Properties such as linear mass of the rail are averaged if they change along the segment and punctual informations such as presence of switch & crossing, level crossing or rail insulated joints are counted. Other informations such as frequency of acceleration of trains are not available and thus indicated as missing values. Models are then adapted consequently.

With almost 100 000 km of rails (two per track), this segmentation leads to around 1 million segments, which is then increased by the time segmentation.

Time event can be considered in different manners. The rail defect is the supervising feature and is used to define the failure mode of the rail segment, namely the type of defect observed. The considered defects are specific to each studies and detailed in the next section. Maintenance work are separated in two types. The renewal or local replacement consist of a new (or reused) rail that did not live the history of previous rail. The rail is thus considered as another segment than the one before this work and leads to a time segmentation. The last considered works are grinding and tamping. In this paper, they will be considered as static properties of each rail segment by counting the number of time each kind of work is done.

2.2 Models

Two types of data-based models are used and can be combined in these studies, scoring and survival.

The scoring model is a random forest classifier parametrised with 20 trees and 15 max depth, as developed in [5]. This model has been compared to other usual classification models and gives, with XGBOOST the best AUC performance (Area Under Curve), the chosen indicator based on the Receiver Operating Characteristic (ROC) curves. Random Forest is then preferred over XGBOOST for its better explainability.

The scoring consists on computing $\hat{\eta}(x)$, which is an evaluation of

$$\eta(x) = \mathbb{P}(Y = 1|X = x) \tag{1}$$

where Y is the state of the segment, 0 if there is no damage, 1 if at least one damage appear, and x the properties associated with the rail segment. This evaluation does not indicate any time horizon at which to expect a defect and is only used for quantifying a risk and compare it to other segments. It is thus good to prioritize works, but not to quantify the optimal amount of work or periodicity.

The survival forest model, extensively detailed in [6] is an adaptation of decision trees for Left-Truncated Right-Censored (LTRC) data developped in [8]. Most of rail segments have more than 30 years and defaults records have been generalised only in 2014 and justify the need of a left-truncated model. On the other side, less than 2% of rail segments experience recorded defects and right-censored model is thus mandatory. The performance of survival models are followed by temporal AUC, which is the AUC computed in classification for different time horizon.

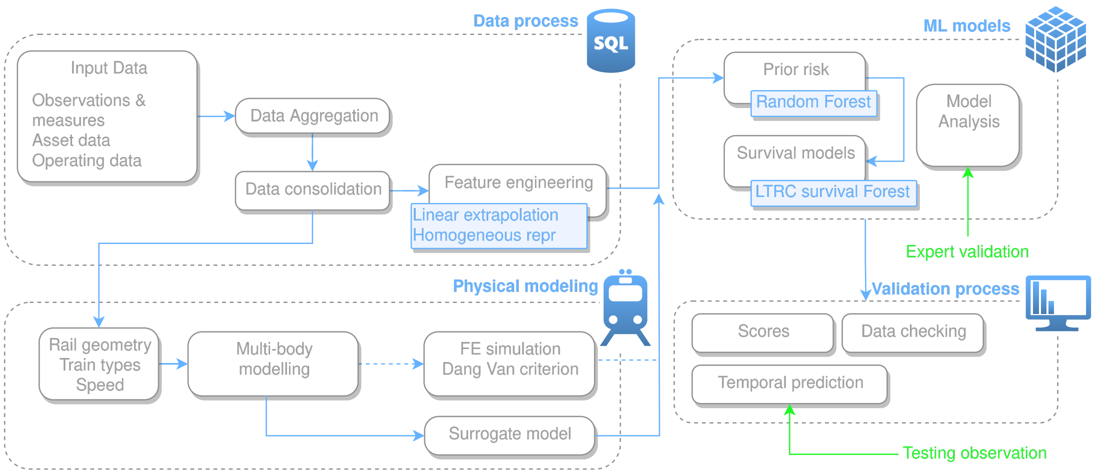

Figure 1: architecture diagram

Physical models are also implemented in the global architecture as shown in figure 1 to aggregate data such as track geometry to produce a more representative indicator for data models such as fatigue index or average contact force. In practice these physical model did not yet enhance the data model performance, mainly because the data set is already large enough.

3 Applications

Model exploitation is here illustrated with three different examples. As each of them requests a different configuration of the model, its performance is evaluated systematically.

3.1 Phenomenon understanding

Objectives

The first application is to get new knowledge on the parameters and combined parameters that favour defect initiation. It helps to prioritize developments in physics modeling and can reveal weak signals that sould be considered for new design.

The example developped here is the one discussed in [7]. Ambient temperature has been recorded at different points of the country and used to interpolate the temperature around the track with minimum and maximum temperature each day during six years. The corresponding fatigue damage has been computed by applying Rainflow counting and Lemaitre & Chaboche cumulative law. A yearly temperature damage is then associated to each rail segment depending on it's geographical position. In this study, only RCF on the contact table are kept, namely, squats.

Model configuration and performance

The objective is to understand and quantify the impact of this added feature to fatigue phenomena. Survival models are thus not required and one can fully exploit the good explainability of random forests. The number of kept features has to be minimal and features have to be uncorrelated. Only 8 features were thus kept, with an AUC dropping from 0.77 with all features to 0.7. Results are thus consistent.

Results and discussions

Several indicators were calculated to define importance of variables, such as Gini Index, Permutation Importance or Mean Decrease Impurity or SHappley Additive exPlaination, the last being shown on figure 2. Most of them led to approximately the same conclusion, the proposed aggregation of cumulative temperature damage is relevant and has a significant impact on squat initiation.

Temperature is empirically considered influent only in extremes. High temperature increase buckling risk and low temperature favor rail break. But that break generally

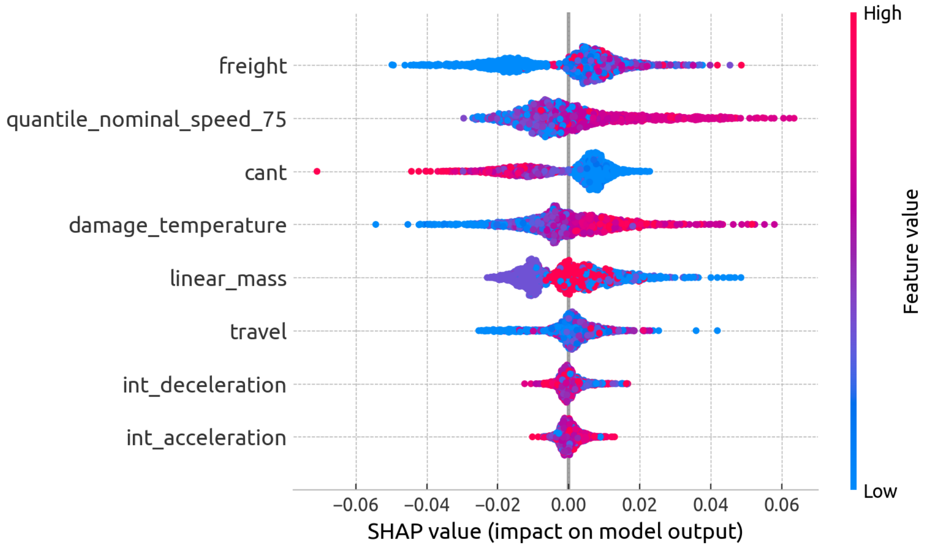

Figure 2: Individual explanation for SHAP importance.

occur on already initiated cracks with a significant propagation. This behavior can not be extended to the initiation part as shown on Figure 3. Daily variation represents around 90% of the cumulative damage, compared to seasonal variation. This observation thus changes the possible levers to reduce RCF defects.

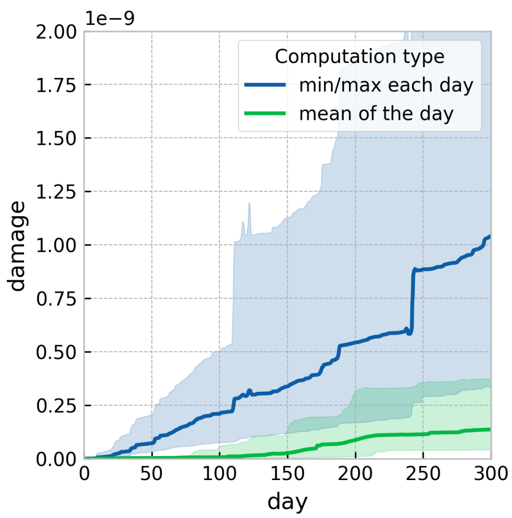

Figure 3: Evolution of damage against day considering variations during the day(blue) or not (green).

3.2 Prescriptive maintenance application

Objectives

The second application proposed is for prescriptive maintenance. The example here is to question the empirically fixed period before the first monitoring of rail cracks using an ultrasonic measurement vehicle. The current prescriptions are based on UIC Class, which is an aggregation of two indicators, the maximum speed limit of the track and the cumulative load called tonnage seen by the track every day [9]. Low index of UIC corresponds to more critical cases such as high load or high speed. As these class define a more global way to monitor, maintain and manage all the infrastructure, from track to catenary, these class are difficult to change. Nonetheless, they can be questioned by the model.

Model configuration and performance

For survival analysis, the machine learning model is not needed. Data structuration is still mandatory to be able to compute the survival curves based on the Kaplan Meier or *product-limit* estimator,

$$\hat{S}(t) = \prod_{i:t_i<=t} \left(1 - \frac{d_i}{n_i}\right) \tag{2}$$

with t_i the distinct times of defects detection, d_i the number of defects that happened at time t_i, and n_i the number of segments without defect up to time t_i. The defects considered here are all defects deemed be detectable by ultrasonic vehicles. The value of $\hat{S}(t)$ is calculated for each UIC group, and the confidence interval represented by the colored shades is the exponential Greenwood [10] confidence interval.

The current classification is then compared to the model. In this case, all defects appearing before the start of the test date are used for learning. Classification of most critical segments is then computed based on random forests over 2 years. Both classifications methods are then compared using the time-AUC as performance indicator.

Results and discussions

Figure 4 shows the survival curves of each UIC groups in different colors. Stars along these curves correspond to the period after which the first visit by a vehicle equipped with ultrasonic measurement has to be planned. After that period, a cyclic period is planned but this cycle is not questioned here since the risk to evaluate is then the probability of rail break, after crack initiation and propagation, which is not in the scope of the presented model.

One observes that relative positioning of groups is globally coherent. Two particular cases can however be noticed. First, UIC 2 survival curve has a specific behavior compared to other, which can be attributed to the fact that the group, mainly composed of all high speed lines and some very dense lines, is the only one for which the cyclic preventive grinding has been strictly applied over the whole time window of the observed defects. The survival curve thus becomes more flat. The second observation is for UIC 7 to 9 for which the survival curves go under the other UIC while keeping their relative position. This behavior is attributed to the reuse of undamaged old rail from other UIC groups to replace some degradated sections. Maintenance policy has thus a direct impact on calculated survival curves, showing their relevance.

As of the empirically defined periods indicated by stars, most of them are in the interval $[0.97 - 0.99]$ of the survival curve, with a higher risk for higher index of UIC. As the crack propagation phase is also directly linked with the total load passing over the rail, the risk of break is thus increased for lower index of UIC. Previously defined periods are thus adapted and do not need to be modified.

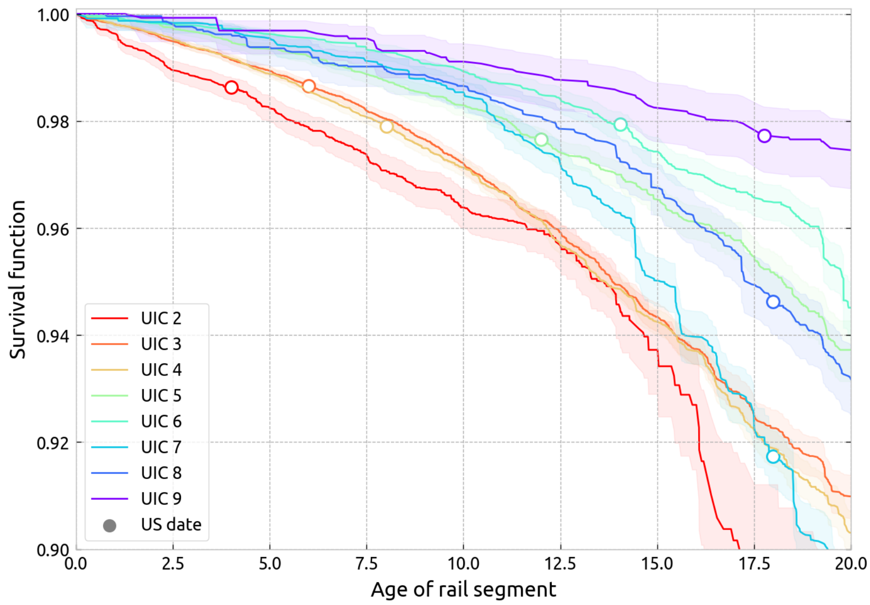

Figure 4: Survival functions of each UIC groups regarding the RCF defect initiation.

Figure 5 shows the time-AUC of both classification methods, UIC-based or ML-based. As expected, ML-based classification outperforms UIC-based one with an AUC over 0.8 for the 24 months of evaluation. But this classification does not consider practical constraints and highly segments a continuous track, which is not manageable in monitoring planification. With an AUC always over 0.67, the UIC class based period definition is thus quite relevant despite not being optimal. Again, changing this empirically defined monitoring rule would need deeper studies, considering track continuity and vehicle route, that might not lead to a significant improvement in the obtained results.

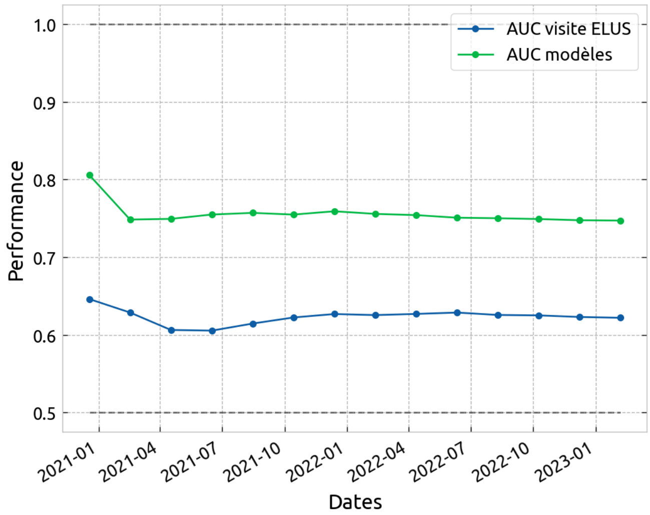

Figure 5: Comparison of time-AUC performance between the UIC based categorization and an optimal ML-based categorization.

3.3 Predictive maintenance application

Objectives

Historically, the dreaded event in the rail is its break. Researches in understanding fatigue and monitoring the phenomenon led to a good crack detection drastically reducing the rail break risk and making the crack initiation the new dreaded event. In this context, the new goal is to reduce the corrective maintenance such as rail replacement, renewal or corrective grinding, by preventive maintenance, mainly by grinding. The current method is to plan systematic maintenance by cyclic grinding. The goal here is to improve this maintenance by defining priority and time for each rail segment.

Model configuration and performance

Rail grinding is a maintenance work that concerns only RCF. The three defects considered are thus the shelling, head checking and squats, which represent together more than 75% of rail defects. All available features are taken since explainability is not mandatory here.

The LTRC random forest is used here to predict survival curves of each segments. Data from all lines are used for the training until the date at which the prediction starts. As illustrated in the previous section, a threshold has to be fixed to define the date at which it is most probable to observe the first defect. This threshold is fixed so as to have a false positive (or false alarm) rate of 20%.

The model is then evaluated on one specific line selected to illustrate the application. The prediction performance is evaluated in two ways, the confusion matrix and the time-AUC.

Results and discussions

Figure 6 shows the typical results produced for a single track on a line of around $500km$ length. Predictions were generated every four months over two years. On the left, one observes that for the first month, the curve is irregular, due to the small amount of new recorded defects. The performance is very steady over these 2 years with an AUC around 0.88, computed specifically for that track. The line was selected for its good performance, mainly due to a long length and good coverage of data in space and time. To illustrate in practice how it could be handled, the confusion matrix at the last time step, t=2022-12-31, and shown on the right. For readability, the number is given in kilometer of rail. If one plans to grind $108km$ of rail by this time, 83% of the rail that would have experienced defect initiation during this period would be saved by grinding. Compared to a typical systematic grinding every two years, $389km$ of the track do not need that periodicity, which thus represent a significant maintenance reduction.

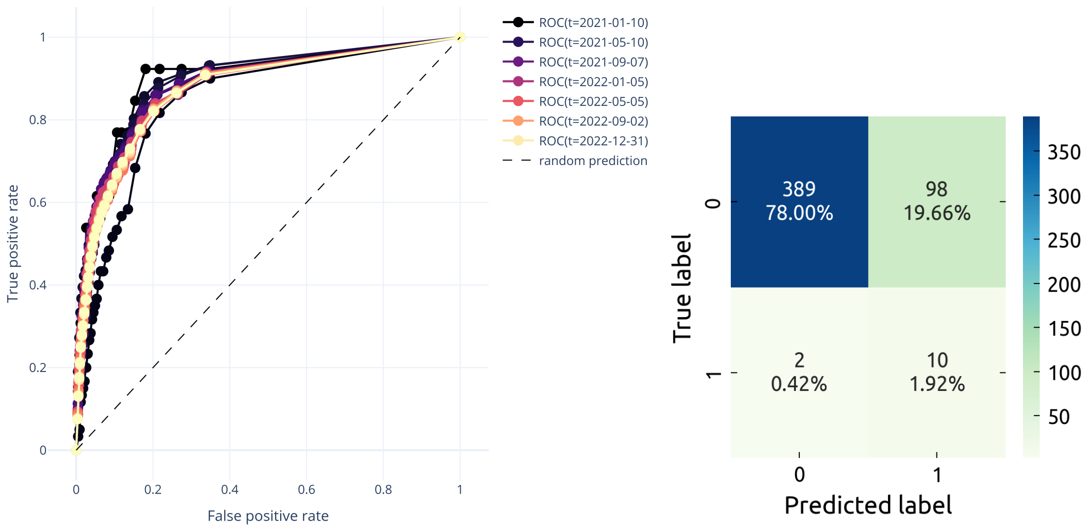

Figure 6: ROC curve (left) and confusion matrix (right)

4 Conclusions & Perspectives

A supervised data-based model has been introduced with its architecture. Its flexibility in the choice of supervising features and combination of classification or survival model has been illustrated through three different applications.

- **Phenomenon understanding**, with the impact of daily temperature variations,
- **Prescriptive maintenance**, by defining the rule of a first monitoring,
- **Predictive maintenance**, by prioritizing the order of preventive rail grinding.

The current models and it's architecture has thus already several applications. Nevertheless, some improvement with new challenges can be tackled. First, representation of time events should be improved on the survival model. As an example, the way of how to consider grinding works and make it appear in the survival curves is not clears.

Moreover, we have seen that physical model did not bring significant gain, except for temperature consideration. It has however a role to play if extrapolation is considered. For now, the classification nature of random tree does not allow a robust extrapolation and some other methods have to be investigated, such as transfer learning. This improvement would unlock new applications that can be useful if innovations are introduced in the railway system.

References

[1] D. Li, E.T. Selig, ”Wheel/Track Dynamic Interaction: Track Substructure Perspective”, Vehicle System Dynamics, 24:sup1, 183-196, 1995

[2] J.W. Ringsberg, ”Life prediction of rolling contact fatigue crack initiation”, International Journal of Fatigue, 23(7), 575-586, 2001

[3] X. Wang, X. Liu, Z. Bian, ”A machine learning based methodology for broken rail prediction on freight railroads: A case study in the United States”, Construction and Building Materials, 346, 2022.

[4] F. Ghofrani, A. Pathak, R. Mohammadi, A. Aref, Q. He, ”Predicting rail defect frequency: An integrated approach using fatigue modeling and data analytics”, Computer-Aided Civil and Infrastructure Engineering, 35(2), 101-115, 2020.

[5] V. Laurent, O. Vo-Van, M. Mougeot, J.-M. Ghidaglia, ”Machine Learning Based Prediction of Fatigue Events in Railway Rails”, 15th WCEAM Proceedings, 455-466, 2022

[6] V. Laurent, O. Vo-Van, ”Survival forest for left-truncated right-censored data”, IPOL, Preprint submitted January 2023

[7] O. Vo-Van, V. Laurent, ”Impact of daily and seasonal temperature variation on rolling contact fatigue damage in the rail”, Fatigue Design 2023, Senlis, France

[8] W. Fu and J.-S. Simonoff, Survival trees for left-truncated and right-censored data, with application to time-varying covariate data, Biostatistics, 18 (2017), pp. 352–369.

[9] UIC, Railway application - Classification of lines for the purpose of track maintenance, leaflets-irs-table, reference number 714.

[10] Hosmer Jr, D. W., Lemeshow, S., May, S. (2011). Applied survival analysis: regression modeling of time-to-event data. John Wiley & Sons.