A decision support system for track maintenance

C. Meier-Hirmer[1,2], A. Senée[1], G. Riboulet[1], F. Sourget[1] & M. Roussignol[2]
[1] *SNCF, Paris, France*
[2] *Marne-la-Vallée University, France*

Abstract

For any railway infrastructure manager, track maintenance is one of the most important tasks to perform. This mission requires a regular follow-up of geometric track quality and monitoring of intervention execution. Among other tools, TIMON, a computer application using track geometry measurements and maintenance operation data, is currently in use in France. This application illustrates the evolution of track quality indicators.

We present a possible extension of this tool: a systematic analysis of the available data by applying a stochastic model for the track geometry deterioration is proposed. The use of environmental variables in order to identify the significance of their influence on track degradation is introduced.

The stochastic model has two advantages: on the one hand it enables one to take into account the intrinsic variability of the degradation phenomenon and on the other, significant environmental variables are included into the model. The work presented here aims to classify sections of the high-speed railway network with respect to similar degradation behaviour. Two different graphical representations that could ease decision-making are proposed.

Keywords: track degradation, stochastic process, maintenance scheduling, classification, decision support tool.

1 Introduction

In order to control track geometry, periodic levelling measurements are carried out by a measurement vehicle called Mauzin. The inspection step depends on the UIC group (indicator of the line importance). This study refers to high speed lines

where information is available every 45 days. Track geometry defects that can be detected by the Mauzin vehicle are presented in fig. 1. To each defect corresponds a specific measure.
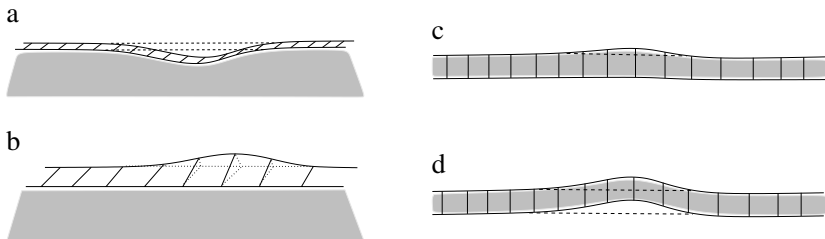

Figure 1: The different degradations of track geometry. a: longitudinal levelling, lateral view; b: transversal levelling, lateral view; c: gauge, top view; d: lining, top view. To simplify, only one selective defect is illustrated. The synthesised measures summarise the quality over 200 meters or one kilometre with respect to the type of degradation.

Using these measurements, indicators of track quality over 200 m or 1 km are calculated. The indicator over 1 km has been used for 30 years. Owing to new computer technology, an indicator over 200 m was developed at a later date. This new indicator is more detailed; it allows to better separate track sections containing switches and expansion joints from 'normal' sections. For example, using the older indicator the whole kilometre containing a switch has higher measurements. Using the 200 m indicator, only one of the five 200 m sections of the concerned kilometre has increased measurements. In order to maintain the history of the complete database, the 1 km indicator is still in use. There are five types of track quality: NL (longitudinal levelling), NT (transversal levelling), D (lining), Ec (gauge) and vEc (gauge variation). This article presents the use of statistical methods and their utility for track maintenance scheduling. For confidentiality reasons, the values of the significant environmental variables and the location of the different degradation classes are not given. The research presented in this article is based on the longitudinal levelling.

The article is built up as follows: section 2 describes an existing computer application TIMON (Traitement Informatique du Mauzin et des Operations de Nivellement: Computer processing of the Mauzin data and levelling maintenance) that is used at the SNCF for track degradation surveillance and for decision-making concerning major maintenance interventions. This tool displays information on the track geometry inspections as well as on maintenance operations carried out for each line.

Section 3 describes the data used in our study in addition to the NL indicator (3.1). This section also gives a short introduction to the statistical methods used to estimate the track degradation rate and to generate classes by cutting the high

speed line into fractions having similar degradation (3.2). We apply these methods to real data and present the results (3.3).

Section 4 shows two possible types of diagrams that can be used to include the results of the analysis presented in section 3.3 into the existing decision-making tool for track maintenance TIMON.

The article ends with a discussion on proposed improvements that could ease decision-making.

2 The existing decision-making tool: TIMON

In order to display and to treat data coming from the Mauzin measurement vehicle, a computer application named TIMON was developed by the infrastructure information system department. For other decision-aiding applications see Le Bihan [1]. The synthesized indicators mentioned in section 1 are calculated after every inspection and are included in the TIMON database. This database contains, among other things, all synthesised Mauzin indicators ever measured on high speed lines. The inspections on these lines are usually carried out every 45 days. We are particularly interested in the NL indicator (*cf.* explanation in section 1) as it is often used for assessing the necessity of major interventions.

In the top part of fig. 2 we can see different interventions carried out associated with the length of the concerned section. At the upper right-hand side switches or bridges are indicated. For example, the selected section, which is displayed in fig. 2 includes a bridge. fig. 2 shows also the development of the NL indicator. We have chosen the 1 km section, but the same information is available for the 200 m sections. It can be seen from the graph that the tamping carried out during the first trimester 2004 reduced the value of the NL indicator. A special feature of this type of display is its aid in predicting the exceeding of thresholds. It is possible to select NL measurements and to conduct a linear regression on these points. The date at which the threshold is likely to be exceeded, is also calculated.
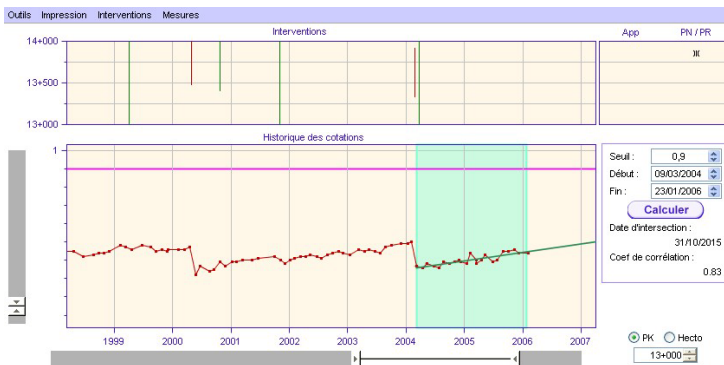

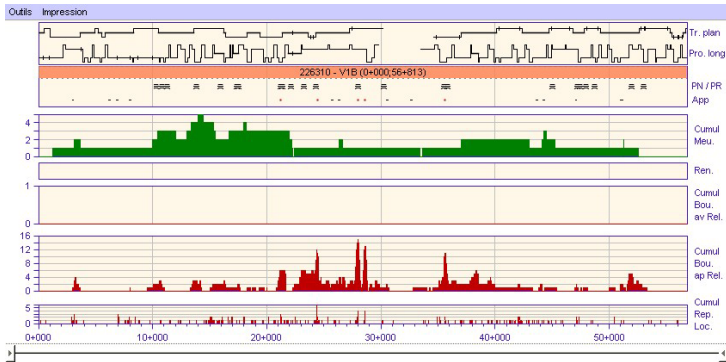Figure 2: The TIMON application: NL development.

Figure 3: The TIMON application: cumulated interventions.

The work presented in this article aims to ameliorate this calculation by including the uncertainty of the track development. We are therefore using a stochastic process for modelling. Our aim is not only to display the track characteristics but to include them directly in the prediction of track development. TIMON is able to display all types of track interventions (tamping, grinding, track lifting, ballast and rail replacement). Information on track characteristics, like curves or switches, is also given.

Fig. 3 shows one of the possible screens: the cumulated intervention display. The number of grinding and tamping interventions is calculated. With respect to tampings, it is even possible to separate the number of tampings before and after a track lifting or ballast replacement. As this screen gives supplementary information about track characteristics, it is possible to verify if there is a correlation between these environmental variables and the number of interventions.

It is also possible to compare the measurements of selected inspections for a whole line or a part of a line. Other features include the search for extreme synthesised indicator values and the computation of the mean degradation over a selected section including or excluding bridges, tunnels, switches and expansion joints.

The display of real Mauzin measurements (in addition to the synthesised indicators) and the acceleration measurements is under development. A prototype, which allows in addition a combined analysis of these indicators, is tested at the moment.

3 Track degradation study

3.1 Available data

The statistical analyses are based on track geometry measurements of French high speed lines. We use the synthesised indicator NL for this study. Data are available for every high speed line since its opening. Inspections are usually carried out

every 45 days. In order to explain the degradation difference between different track sections, we introduce environmental variables (covariates). The following information is incorporated into the study:

- ballast: type, height of sub-ballast;
- ballast renewal;
- maximum speed;
- UIC group;
- curves and weighted average of curvature;
- cant and cant deficiency;
- ascending and descending gradients;
- combination of curves and gradients;
- tonnage in 2004 and cumulated tonnage since the last ballast renewal;
- climate: temperature, precipitation and frost;
- sleepers: type and year of fabrication;
- age of the line;
- rail: type and treatment;
- tunnels and bridges;
- expansion joints and switches.

The climate is only available for every administrative district. These two covariates are therefore highly correlated. The information about the climate may be too coarse to give coherent results.

The NL indicator and the environmental variables are merged into two databases; one with a 200 m, the other one with a 1 km resolution. There is no specific treatment for the synthesised indicator, as the information is already available for every 200 m and 1 km. In order to get the same splitting for the covariates, the following rule is used: the predominant value with respect to the length is attached to every 200 m or 1 km section. Tunnels and bridges are entered if they are longer than 200 m. Expansion joints and switches are considered if they exist on the concerned section. Altogether we have two databases with 31 categorical and 21 continuous covariates.

3.2 Methods

3.2.1 The Gamma process

The goal is to propose an estimate of the track degradation rate incorporating the uncertainty about track evolution. The estimate should be based on the NL indicator, not just on adjacent measurements between two interventions. The TIMON application can only treat the latter analysis (*cf.* 2). We use the stochastic Gamma process for modelling. The Gamma process $X_{t,t\geq0}$ has the following properties:

- X_t has independent increments,
- for $s \geq 0$ and $t \geq 0$, $X_{t+s} - X_t$ has a Gamma distribution with the parameters αs and β. The density is given by

$$f_{X_{t+s}-X_t}(x) = \frac{1}{\Gamma(\alpha s)} \, \beta^{\alpha s} \, x^{\alpha s-1} \, e^{-\beta x} \, \mathbb{I}_{\{x\geq0\}}.$$

where $\alpha > 0, \beta > 0$. Therefore the Gamma process has a constant mean degradation rate ($\frac{\alpha}{\beta}$) and a constant degradation variance rate ($\frac{\alpha}{\beta^2}$).
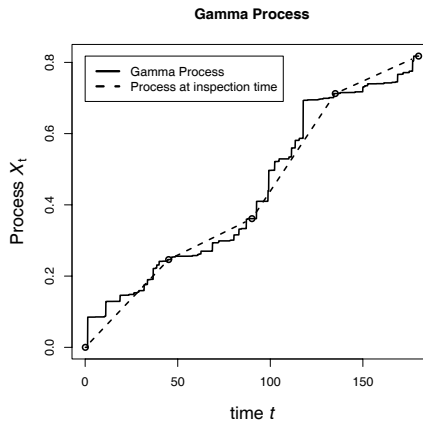

Figure 4: Trajectory of the Gamma process and the information known at inspection times.

Fig. 4 shows the trajectories of the Gamma process and explains the information loss due to the periodic inspections. Only continuous surveillance could provide an observation of the exact trajectory. The slope of the dashed line is approximately $\frac{\alpha}{\beta}$. The variation of the solid line from this mean degradation is related to the quantity $\frac{\alpha}{\beta^2}$. The properties of the Gamma process comply with SNCF expert judgement and it was verified that it gives satisfactory results. Simulations of the process development were performed and compared to real data.

Real NL indicator data is used to estimate α and β for every 200 m or 1 km unit of all French high speed lines. Values ($\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$) are then calculated.

3.2.2 Classification method

The method used to explain the track degradation based on environmental variables is called CART (Classification and regression trees). This method was introduced by Breiman *et al.* [2]. A feasibility study was conducted by Senée [3] and gave reasonable results. We decided, therefore, to retain the CART method.

Regression trees can be used to predict the value of a continuous output variable based on one or more input variables. The input variables can be a mixture of continuous and categorical variables. The "regression tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then slitting it up further on each of the branches" [4]. The process works as follows:

1. The whole data is located in one class.
2. For every variable and for every possible split, the algorithm calculates the sum of the squared deviations, which is an indicator of the homogeneity of

the partitioning. The partition minimising this criterion is kept.

3. The algorithm repeats point number two, above, with the two new branches.

This process continues until one class is created for every observation. "However, this would not make much sense since one would likely end up with a tree structure that is as complex and 'tedious' as the original data" [5]. In our study we use two stop criteria:

- the complexity level of the tree is predefined regarding the length of the concerned railway line,
- splitting is not realised if a partition regroups less than five observations for the 1 km sections and ten for the 200 m sections.

The use of regression trees has the following advantages: "Tree methods are nonparametric and nonlinear. The final results of using tree methods for regression can be summarised in a series of logical 'if-then' conditions (tree nodes). Therefore, there is no implicit assumption that the underlying relationships between the predictor variables and the dependent variable are linear, follow some specific non-linear link function, or that they are even monotonic in nature" [5].

The following section shows the results of this method applied to our data. "In most cases, the interpretation of results summarised in a tree is very simple" [5]. In comparison with other methods, tree methods "can often yield a much simpler 'model' for explaining why observations are classified or predicted in a particular manner" [5].

3.3 Results

The methods described in section 3 are used to estimate the mean degradation rate $\frac{\alpha}{\beta}$ and the degradation variance rate $\frac{\alpha}{\beta^2}$ every 200 m or 1 km, respectively. The computation is done using the R software [6]. Classification trees are constructed for the two objective values ($\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$) and for the 200 m and 1 km sections. We thus dispose of four trees per line. Trees based on the data of all French high speed lines are also constructed. The computation was done using SAS® Enterprise Miner[TM] [7]. For confidentially reasons we cannot give the values of the separating variables.

Fig. 5 (a) shows the results of the mean degradation rate for the 1 km resolution. The variable TOTTON has an important impact on the degradation rate. TOTTON is the total tonnage of the line since the opening. Another variable that separates well the sections in two classes is the variable expansion joints.

Fig. 5 (b) presents the results of the mean degradation rate for the 200 m resolution. Again we find the covariate TOTTON. Another covariate which divides the line sections into two groups is CRV+GRAD. This covariate indicates where there are significant curves and at the same time significant gradients. Both trees divide the line into three degradation groups. In the following section (4) we propose graphical representations of the three classes.

The trees per line give interesting results. It is possible to identify covariates that were judged so far not to be the most influencing ones, such as the combination of curves and gradients. But even in the case where the trees confirm expert

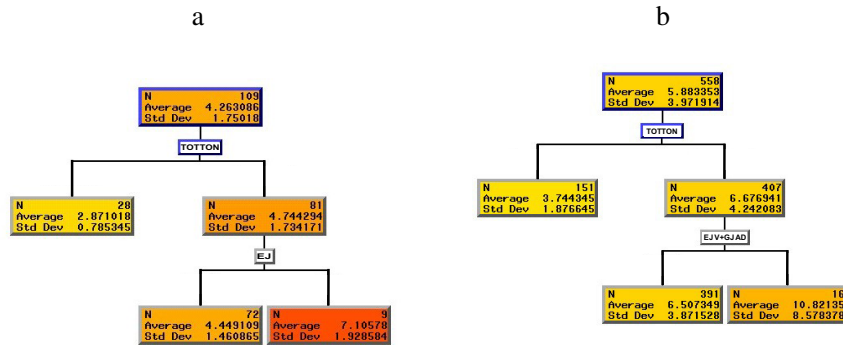a   b

Figure 5: Classification tree for a high speed line; (a) 1 km resolution (b) 200 m resolution.

knowledge (influence of speed and tonnage) they are able to rearrange these covariates according to their importance for every specific line.

The global trees, based on the data of all high speed lines, bring out covariates which are very different between the lines and whose importance was emphasised by the experts already before the analysis (*e.g.* tonnage and age). The trees for every single line, however, show some interesting results. The covariates specific to every line will be investigated by track maintenance experts. There are some lines where the 200 m and the 1 km trees gave similar results like in the case we presented here; for others, the results are very different. This can be explained by the enhanced accuracy of the 200 m indicator and by the slight difference in the CART algorithm (minimal class size of five or ten observations; *cf.* 3.2.2). A similar phenomenon exists for the mean degradation rate trees and the degradation variance trees.

4 Improvements of the existing tool

The following pictures are screen shots coming from a graphical display application for infrastructure data developed at the Innovation and Research Department of the SNCF. In order to propose improvements of the maintenance surveillance tool TIMON, we added our results manually.

The first proposal is to use a map of France to visualise the different degradation classes (*cf.* fig. 6). This design takes account of the mental model principles formulated by Wickens *et al.* [8] for interface design. One of the principles is that the design display should look like the object that it is trying to illustrate. Using this type of display, it is easy to locate the zones having good or bad track degradation behaviour.

Another representation is shown in fig. 7. The advantage of this type of display is the fact that it is based on the long-existing track layout plans. Maintenance experts

are used to this type of illustration, already incorporated in the existing application TIMON. Some of the covariates used for the construction of the trees are available on the track layout plans (*e.g.* tunnels and bridges). It is therefore possible to better understand the partitioning created by the algorithm. From an ergonomic point of view, this type of representation respects the memory principle. Information similar to the information delivered by the existing tool, is displayed in a similar way. Thus, the different displays are consistent in its information projection.
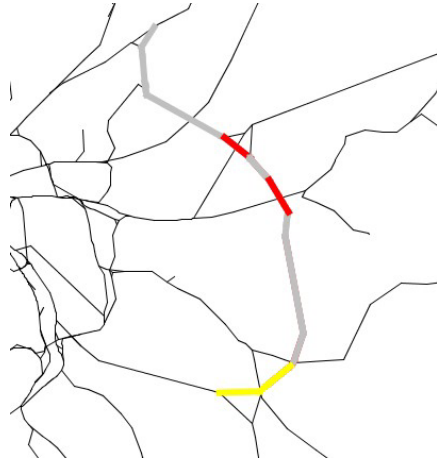

Figure 6: One possible presentation of the track degradation classification results is the plot on a map of France. Displaying the results in this way facilitates the geographical localisation of line parts having characteristics leading to good or bad geometric behaviour. Every colour corresponds to a degradation class.
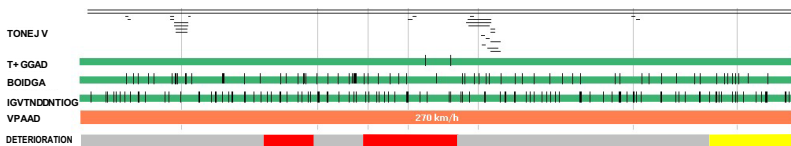

Figure 7: A second presentation of the track degradation classification results. The line is represented by a vertical line. Displaying the results in this way has the advantage that additional covariate information, such as height of the track or the presence of bridges, can be indicated together with the degradation class.

5 Conclusion

Our study uses environmental variables in order to explain track degradation and to create classes of homogeneous track quality. The chosen CART method seems appropriate for the task to accomplish. The classification trees show the importance of covariates so far not used for maintenance scheduling and sort the covariates known by the experts according to their influence for every line. Based on these results, we propose improvements of an existing track maintenance surveillance tool that could ease major operation allocation.

There are two major perspectives for the study. In the first place, we want to include even more covariates that could be important for the track degradation. A larger inspection history and more precise covariate information could enhance the analysis. For example, the climate data are too coarse. Efforts are currently being made to obtain more detailed information. In the second place we want to insert the results into the TIMON application. One of the graphical representation of the results has to be chosen and the corresponding code has to be developed. The results of the trees will be carefully analysed before incorporating them into the decision-making tool and before using them for maintenance scheduling.

Acknowledgements

References

[1] Le Bihan, A., Évolutions de la maintenance de la voie à la SNCF. *Congrès International Voie*, 1997.
[2] Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J., *Classification and regression trees*. Chapman & Hall, 1984.
[3] Senée, A., *Analyse des indicateurs des défauts de géométrie de la voie ferroviaire*. Master's thesis, ENSAI (École Nationale de Statistique et de l'Analyse de l'Information), 2005.
[4] Resampling Stats, http://www.resample.com/xlminer/help/rtree/rtree_intro.htm
[5] Statistica, http://www.statsoft.com/textbook/stcart.html
[6] R project, http://www.r-project.org
[7] SAS Enterprise Miner, http://www.sas.com/technologies/analytics/ datamining/miner
[8] Wickens, C.D., Gordon, S.E. & Liu, Y., *Introduction to Human Factors Engineering*. Addison-Wesley-Longman, 1997.