An original multi-nets approach for modelling and evaluating maintenance strategies

Olivier François & Laurent Bouillaut
French National Institute for Transportation and safety research (INRETS)
Laboratoire des Technologies Nouvelles, 2, rue de la Butte Verte
F-93166, Noisy-Le-Grand Cedex, France.

Stéphane Dubois
Régie Autonome des Transports Parisiens (RATP)
Dept. Des Equipements et des Systèmes du Transport, 40bis, rue Salengro
F-94724, Fontenay-sous-Bois Cedex, France.

Abstract

Reliability analysis is an integral part of system design and operating. Moreover, it can be an input to optimize maintenance policies. Recently, Dynamic Bayesian Networks (DBN) have been proved relevant to represent complex systems and perform reliability studies. The major drawback of this approach comes from the constraint on the sojourn times which are necessarily exponentially distributed, as in usual Markovian approaches. To avoid this constraint, a new formalism named Graphical Duration Models (GDM) was introduced[1]. This approach, based on semi-Markovian models, allows representing all kind of sojourn time distributions. Then, the degradation process of complex systems (multi-components, multi-states, eventually influenced by contextual variables) can be accurately modeled and thus, the related reliability indicators correctly estimated. With this generic approach (named VirMaLab, for Virtual Maintenance Laboratory) various industrial applications were developed, especially as decision support tools for the optimization of railway infrastructure maintenance strategies.

Keywords: Dynamic Bayesian Networks, Graphical Duration Models, Maintenance, Reliability, Degradation Process Modelling.

1. Introduction

In this paper, an extension of the commonly used VirMaLab formalism will be introduced. Indeed, this new application deals the broken rails prevention in an automation context for railway Paris metro lines. The final goal of the project is to evaluate and compare various diagnostic, maintenance and operating scenarios, in terms of availability, broken rails frequency… Due to the peak hour's constraints, the operator (RATP) needs to estimate, hour by hour its ability to detect broken rail. But, for many reasons (time computation, parameter accuracy, learning data…), the modeling of a rail degradation process with a one hour step is impossible.

To address this problem, a multi-nets model was developed, allowing a variable granularity in respect of the state of the rail. Usually, in VirMaLab applications, the all model infers with a constant step. Here, four models were introduced, with their own inference step fixed in accordance with the defect gravity (from one month for early inner rail cracks to one hour for broken rails) and their own set of diagnosis devices (all defects levels are not detected by the same appliances). Finally, the three first models emphasize the use of the preventive maintenance strategies on the availability of the network whereas the last model focuses on the corrective maintenance and evaluates, hour by hour, the response of the diagnosis system in terms of broken rail detection ability.
Parameters of these models are learnt by use of REX databases and/or expert advises. Then, the global model is validated by various experiments with the standard running, diagnosis and maintenance parameters. Receiving the validation of these first results by RATP experts, new sets of scenarios can be computed, evaluating the influence of any parameters.
To evaluate a given maintenance strategy, various indicators are analyzed, from annual numbers of broken rails and preventive maintenance actions to delays before broken rails detection and related number of missed trains (a broken rail induces the stop of the exploitation till the defect is consolidated. Then, the acceptable speed is strongly decreased up to the rail refurbishment).

Our goal is to model the influence of maintenance on the reliability and exploitation performance of Parisian metro railway of the RATP which is the major transit operator responsible for public transportation in Paris and its surroundings. The context of the application is the Parisian metro command
system renewal and decision makers will need to update existing maintenance policies. Our constraint is to build a system with a granularity of one hour as we need to evaluate the lost exploitation loop number and as number of exploitation trains vary with day time. To answer this problem we have build a multi-model containing four models with four different steps.

The paper is sectioned as follow. The next part will introduce the methodology of our approach. The third section will deal with the technical developments of the formalisms of Dynamic Bayesian Networks and of Graphical Duration Models. The forth part will detail our model. Then, we will conclude with results and future work.

2. Methodology

As we need a model with a hour step and as inference and simulation of probabilistic graphical models are complex, we have chosen to build a multi-model consisting in four different models concentrating on different tasks and with different steps.
These four step are motivated by the fact before the rail crack, it evolves between four normalized sizes of default. The first normalized abnormality of the rail is named X1 and represents a small default inside the structure of the rail. The second normalized abnormality is named X2 and affects the maintenance planning. That kind of default is often observed. The last normalized default is the crack which is named S and obviously the last possible state is the normal state named Ok.

Considering that classification, the first dynamical model evaluates the transition of the rail state from Ok to X1, and given the slow evolution of the rail, the step of this model is the month. The second dynamical model represent the evolution of the rail from the state X1 to the state X2 and is also based on a monthly step. The third dynamical model simulates the degradation of the rail from the state X2 to the crack S, its step is the week.
These three first models emphasize the role of the preventive maintenance strategies.
The last model (from the state S to the state Ok) is the one that emphasize the role of the corrective maintenance and evaluate hour by hour the response of the system to the crack until the rail replacement by a new component.
When using simulation of the model, the four models are waiting a transition to a new state to enable the following corresponding model. The originality of this work is the use of many dynamical Bayesian networks together with the use of Graphical Duration Model recently introduced in the literature by [3].
When considering exact inference on this multi model, as each model concentrates on different state of evolution of the rail, we consider that performing separate inference on each model even if there are links between these give a solution to the component life time estimation as we could add the estimated life time in the state Ok with the ones in the state X1, the state X2 and the state S. We could use that approximation because only rail in states X2 and S are renewable. Then, we need to estimate the percentage of rail that access the state S compare to those in state X2 which have been corrected by preventive maintenance to weight the above sum.

3. Technical developments

3.1 Dynamical Bayesian Networks

Bayesian networks [6] are a formalism of probabilistic reasoning used increasingly in decision aid, diagnosis and complex systems control [5, 10, 9].
Let $X = \{X_1,\ldots,X_n\}$ be a set of discrete random variables. A Discrete Bayesian network $B = < G,\theta >$ is defined by

a directed acyclic graph (dag) $G = < N,U >$ where N represents the set of nodes (one node for each variable) and U the set of edges
and parameters $\theta = [[\theta_{ijk}]]$, $1 \leq i \leq n$, $1 \leq j \leq q_i$, $1 \leq k \leq r_i$ the set of conditional probability tables of each node X_i knowing its parents' state set $Pa(X_i)$ (with r_i and q_i as respective cardinalities of X_i and $Pa(X_i)$).

Determination of θ and G is often based on expert knowledge, but several learning methods based on data have appeared.
Using BN is also particularly interesting because of the easiness for knowledge propagation through the network. Indeed, various inference algorithms allow computing the marginal distribution of any subset of variables. The most classical one relies on the use of a junction tree [7].
Finally, note that such modeling is able to represent dynamic systems (e.g. which contain variables with time dependant distributions) via the Dynamic Bayesian Network (DBN) solution [8].

In reliability analysis, one can be interested in modeling how a system changes from an up state to a down state over time. Most of the time, a modeling based on the DBN formalism was done [2].

The major drawback of this approach comes from the constraint on the state sojourn times which are necessarily exponentially distributed. Indeed, if the considered system follows (or is very close to) an exponentially distributed degradation process, this approach can be perfectly suitable. On the other hand, if the sojourn times are far from an exponential distribution, a Markovian modeling will be unable to take this fact into account and the modeling of the degradation process will be biased. In a reliability analysis, such inaccurate estimation can have strong consequences, notably if one wants to optimize parameters of maintenance policies based on reliability. This constraint can be solved by using Semi-Markov models which allow considering any kind of sojourn time distributions. One solution is introduced in the following section.

3.2 Graphical Duration Models

The Graphical Duration Model is a specific DBN, using semi-Markov models. The main idea is the introduction of remaining time variable into the graph that allows to model multi-state systems featuring complex sojourn times. Figure 1 shows a GDM in its DBN form.

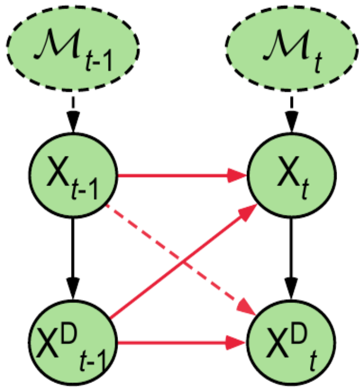

Figure 1. GDM in the form of a DBN.

The solid lines define the basic structure; dashed lines indicate optional items and red bold edges characterize dependencies between time slices. The model handles two kinds of variable:

(X_t), $1 \leq t \leq T$, represents the system state over a sequence of length T.

(X^D_t), $1 \leq t \leq T$, represents the remaining time before a system state modification (remaining sojourn time).

These variables are called duration variables. Optionally, it is possible to introduce a context description of the studied system by means of a prior graphical model M_t . It aims to define the distribution of a possible collection of context variables (covariates) $Z_t = (Z_{p,t})$, $1 \leq p \leq P$ (one variable at least) that works on variable state X_t and/or duration variable X^D_t . Besides, the DAG of a GDM shows that the current system state X_t depends on the previous system state X_{t-1}, the previous remaining duration X^D_{t-1} and, optionally, on contextual variables $Z_{n,t}$. On the other hand, the current duration variable X^D_t is dependent on the previous duration variable X^D_{t-1}, the current state X_t and, optionally, on the previous state X_{t-1} and some contextual variables $Z_{n,t}$. Consequently, the process (X_t) (resp. (X^D_t)) is not Markovian since

$$X_{t-1} \not\perp\!\!\!\perp X_{t+1} \mid X_t \quad \left(\text{resp. } X^D_{t-1} \not\perp\!\!\!\perp X^D_{t+1} \mid X^D_t\right).$$

Where the notation $A \perp\!\!\!\perp B$ means that variables A and B are statistically independent. On the other hand, the GDM structure leads to

$$(X_{t-1}, X^D_{t-1}) \perp\!\!\!\perp (X_{t+1}; X^D_{t+1}) \mid (X_t, X^D_t)$$

So, the set (X_t, X^D_t) engendered by a GDM is Markovian, despite (X_t) is not. The GDM generalizes the recent studies on discrete semi-Markovian processes [1].

On the practical point of view, this approach allows specifying arbitrary state sojourn time distributions by contrast with a classic Markovian framework in which all durations have to be exponentially distributed.
This modeling is therefore particularly interesting as soon as the question is to capture the behavior of a given system subjected to a particular context and a complex degradation distribution. More details on GDMs (quantitative description, optional context description) can be found in [3, 4].

3.3 Our multi-model

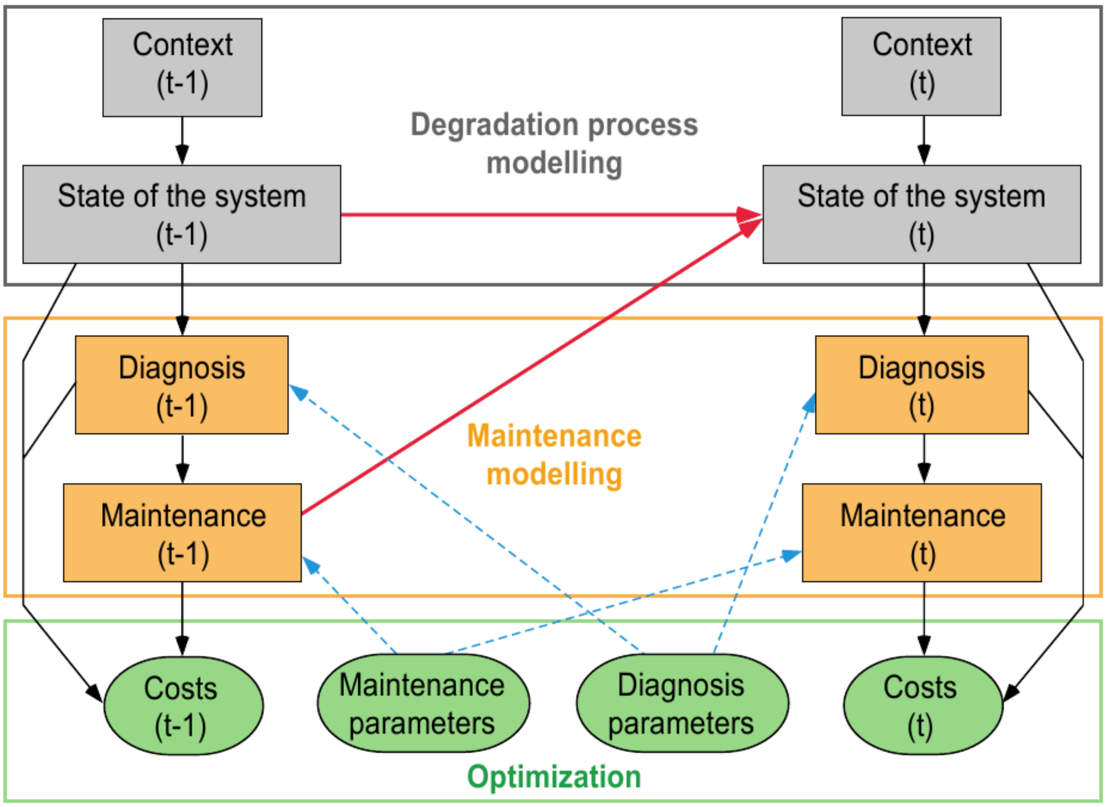

Figure 2. Generic model of a complex maintenance system.

A generic model using a Dynamic Bayesian Network of a degradation process together with a maintenance modeling could be seen in Figure 2. Let remark that

inputs as maintenance or diagnosis parameters could be added together with costs function outputs to initiate an optimization process if needed.
The link between states of the system at time t-1 and time t could be a simple Markov chain (exponential distributions) or a Graphical Duration model (generic discrete distributions). Obviously, As we can't model the sojourn time using exponential distribution, we have chosen to use a graphical duration model [4].

We have been guided for designing the model by the fact that a crack evolves in the rail structure following four normalized default. The first state is Ok, when the rail is alright. Then the first level of default, named X1, represents a crack that has less than 5 millimeters of length. The second size of default, named X2, represents a crack that has more than 5 millimeters of length. Even if the real evolution speed isn't really known as it depends on too many parameters. It is empirically known that bigger is the crack, more speed is its evolution. That's why discriminating the size of the default allows to model it in a better way. The last state, which is named S, represents a default that need immediate replacement of the rail.
That structure of evolution of cracks with different speed, and different maintenance policies associated, has leaded us to a multi-model. The first model represents the (slow) evolution from the normal state (Ok) to the first level of perceptible default X1. This model has a monthly step. The second model represents the degradation from the low level default X1 to the high level of default X2. This second model has also a monthly step as it is known that the evolution from X1 to X2 could take many years. For instance, safety policies admit that, after 3 years of classification in X1, a default is automatically over classed in X2. The third model represents the degradation of the rail from the state X2 (less than 3mm crack length) to the state BR (unsafe crack) and it uses a weekly step. These three first models concentrate on the evaluation of predictive maintenance as the rail is always considered as safe.
The last model concentrate on the efficiency of the corrective maintenance. It deal with a unsafe rail (BR) and hourly evaluate if the crack is detected by the different agent of detection. When it is done the rail is replaced, the cost is evaluated, and we get back to the first model with a new rail (Ok).

This first model, introduced in figure 3, deals with the rail’s preventive maintenance strategy. As a VirMaLab modelling, it is constituted of two blocks.
The first one describes the degradation process of the rail, using the GDM formalism (introduced in section 2.3.). The rail degradation can be influenced by several contextual variables such as the rolling stock (changing from on line to another), the curve radius (and if we consider the inner or outer rail) and the steel’s stiffness.

The second block of this model describes the diagnosis devices and the maintenance strategy. Three devices trigger periodic auscultations of the rails: The Ultra-sonic vehicle (*USV)*, walking survey teams *(WT)* and drivers (*Drv*, whom presence depends on the state of the traffic, with peak hours, night operating stops…). The modelling of the last device is a little more complex. Indeed, various Track Circuits technologies constitute the whole signalling network, with different failure rates, different sizes... Moreover, the analysis of RATP databases underlines that, during worm seasons, the rail dilatation keeps the electric contact of many cracks. All these variables have, therefore, to be taken into account in the final modelling.

All four diagnosis devices supply an estimation of the current state of the rail (integrating their own good detection and false alarm rates) that influences the maintenance decision. When a maintenance action is performed, it is assumed that the system turns to the *OK* state in a single iteration.

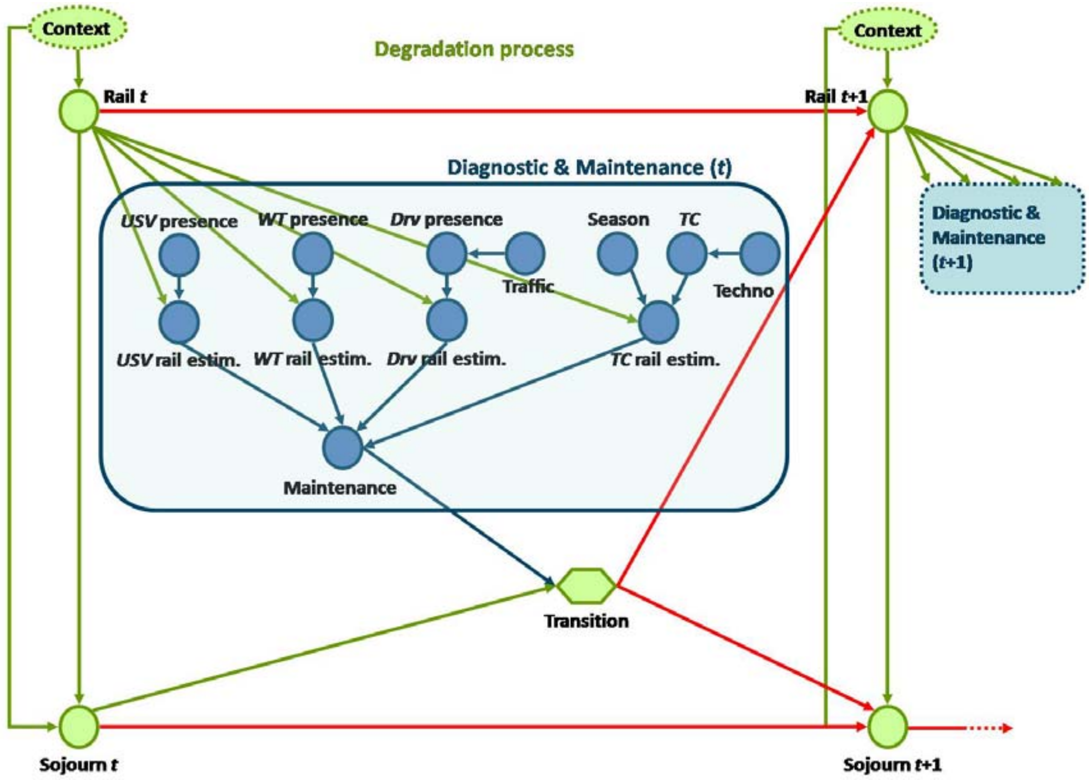

Figure 3. Structure of the *VirMaLab* model for the 3 first slices of the *StatAvaries* Multi-nets.

Four models inspired by the one represented in figure 3 are linked together in a dynamical process. The resulting model represented in figure 4 is a dynamic Bayesian multi-network that allows to simulate life time of the observed rail together with the maintenance policies that are developed on the rail network.

3.4 The final decision support software

To make easier the use of this multi-nets model for both maintenance operators and managers of the automation lines project, a friendly user interface was developed.
It allows determining the following parameters:

– The considered line (among the 11 iron contact RATP metro lines)
– The rail context: The whole line or only the in curve rails (eventually only the upper rail).
– The critical curve radius. It determines the set of curves on which a crack could have critical consequences in terms of passengers' safety.
– The rail quality. For different reasons, an operator can decide to change the iron stiffness. Consequently, the rail degradation process must be adapted.
– Rolling stock specifications: Running period, mean speed, length and axle load. These parameters influence the rail degradation speed and are also necessary to evaluate some final indicators.

- Diagnosis parameters: Good detection and false alarms rates, *USV* and *WT* auscultation periods, parameters of the *TC* technologies encountered on the considered metro line.
- Traffic periods. The user can define the night and running periods (usually, a metro line is operating 20 hours a day) and, in the operating period, 6 different temporal windows and their own train periods. Thus, the real traffic conditions of each line (but also hypothetical parameters that might be evaluated) can be modeled.

When all parameters are defined, the inference can begin. Due to the modelling complexity, the computation of an experiment can be quite long (around 2 hours). But the user can be sure that the *StatAvaries* tool provides the exact values of expected results since the inference of the modelling is based on an exact inference algorithm [7].
As an illustration, the next section will introduce results obtained in one of the scenarios investigated for RATP. The aim of this paper is not to list all results obtained during the study but to introduce the *VirLaLab* multi-nets extension, illustrated by one experimental example. For more information on some of the obtained results, readers can contact the authors.

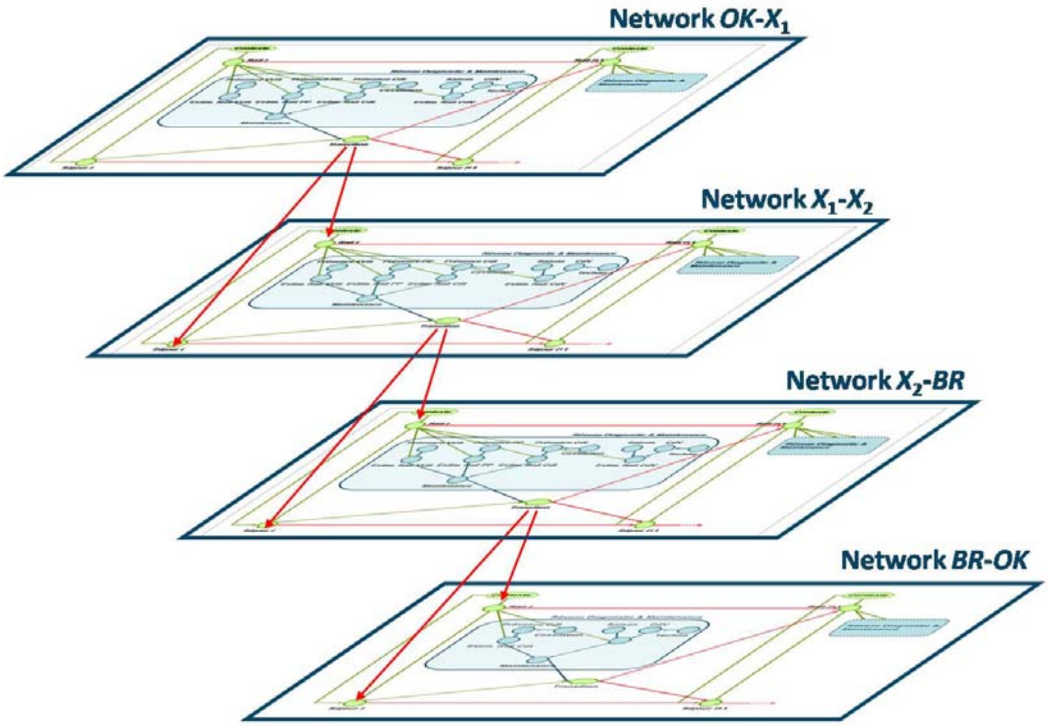

Figure 4. Multi-nets structure of the *VirMaLab* decision support tool *StatAvaries*.

3.5 Some experimental results

In this study, one of the considered scenarios deals with the influence of the *USV* auscultation period on maintenance actions and network's availability.

The figure 5 introduces some results of this experiment, obtained for line 7. For industrial reasons, exact values of indicators are deleted. Nevertheless, the interest of this picture lays in the dynamic of defects numbers.
For this experiment, the ultrasonic auscultation period was changed (the currently commonly used value is *To*), with three considered options: 2*To*, *To*/2 and *To*/6.
We can note that, as expected, the more frequently ultrasonic equipment sound the infrastructure, the more preventive actions will be planed. Early defects are therefore more easily diagnosed, and then, corrected before they turn to the critical state of broken rail.
Moreover, the gain in terms of broken rails is especially significant for the first simulations (*T*o/2) and, beyond, seems to decrease.

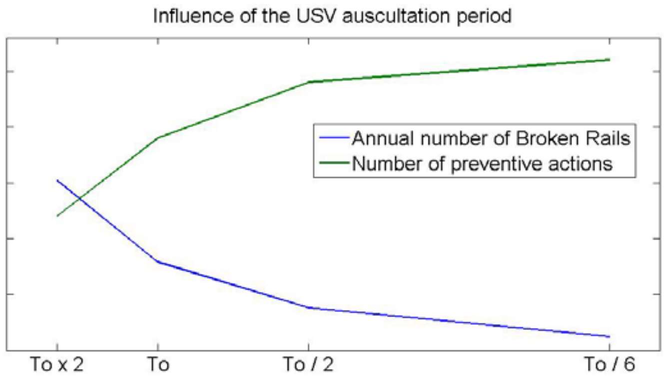

Figure 5. Influence of the *USV* period on rail's degradation.

In terms of network's availability, this experiment furnished a number of lost trains, balanced according to the day's period when the *BR* occurs (operating periods, peak hours, night...). Indeed, the model assumes that, when a rail breaks, the running in completely stopped during 45 minutes. This induces around 22 lost trains in peaks hours, 9 in early morning, 4 in late evening and none during the 4 hours 'night' period. Due to the rolling stock action on the upper rail (located in curves), the larger part of *BR* occurs in curves. Figure 9 introduces the influence of preventive maintenance on the localization of *BR* and on their number.

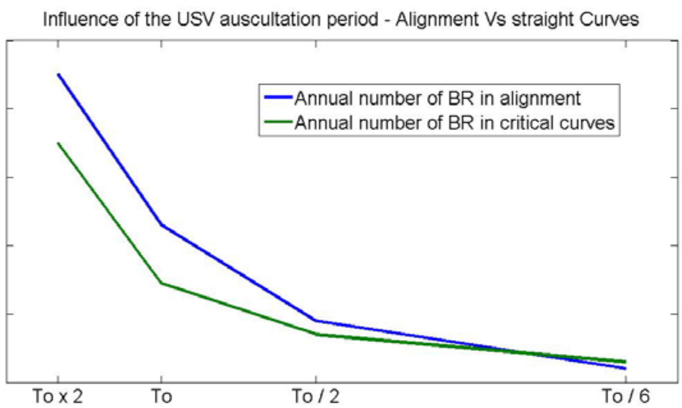

Figure 9. Influence of the *USV* period on rail's degradation for different curves.

We can note that, decreasing the *USV* auscultation period, the preventive maintenance improves (as introduced in figure 8). Then, more X_2 defaults can be detected which means less *BR*. As introduced previously, mechanical constrains on the rail are higher in curves. So, this context triggers a quicker degradation speed and a higher number of X_2 (with finally a higher number of *BR*). The improvement of preventive maintenance shows a decrease of ‘in *risky* curves’ *BR* numbers with finally a nearly equivalent behaviour of both contexts (Alignment and *risky* curves).

4. Conclusions and Future work

We have introduced an original Bayesian multi-network than uses different granularities. This multi-model could be used as a simulator to describe scenarios in order to extract indicators that represent the efficiency of maintenance policies. The focused application is dedicated for the prevention of broken rails, in a metro lines automation context.
The model is based on a generic approach named *VirMaLab* (Virtual Maintenance Laboratory) using the Dynamic Bayesian Network theory, with its modular approach. Thus, the proposed model can be divided in sub networks, eventually interconnected, describing the rail degradation process, the different diagnosis devices and, finally, the maintenance actions decision.
The originality of this work is that, if the application introduced in this paper deals with the railway infrastructure, the considered approach is generic and can easily be extended to all kind of maintenance processes modelling for determining Maintenance and/or Diagnosis optimal parameters.
Moreover, the use of Graphical Duration Models ensures an accurate degradation process modelling, whatever sojourns times distributions in all system’s states.
As an illustration of this generic approach, some results are introduced, focusing on the influence of *USV* auscultation period on annual broken rails and on their localization. It illustrates the ability of the approach to simulate all kinds of scenarios, modifying maintenance decisions, diagnosis parameters or running variables.
One last advantage of the introduced method leads in the fact that all new information (from database or expert advice) or modification of the diagnosis process can easily be taken into account to amend the modelling.
Finally, the integration of metaheuristics in the inference algorithm is actually in progress will furnish useful tool to determine, in respect of some predetermined criteria, the optimal diagnosis and/ or maintenance parameters.

References

[1] Barbu, V., Boussemart, M., Limnios, N., 2004. Discrete time semi-markov processes for reliability and survival analysis. Communication in Statistics - Theory and Methods 33 (11), 2833-2868.
[2] Bouillaut, L., Francois, O., Leray, P., Aknin, P., Dubois, S., 2008. Temporal bayesian network modeling maintenance strategies : Prevention of broken rails. In: Procedings of World Congress on Rail Research WCRR'08.
[3] Donat, R., Bouillaut, L., Aknin, P., Leray, P., 2008. A dynamic graphical model to represent complex survival distributions. Advances in Mathematical Modeling for Reliability ISBN: 978-1-58603-865-6, 17-24.

[4] Donat, R., Leray, P., Bouillaut, L., Aknin, P., 2009. Representation of discrete duration models by means of a specic dynamic bayesian network. Neurocomputing ISSN: 0925-2312, 1-12.
[5] Jensen, F., 1996. An introduction to Bayesian Networks. Taylor and Francis, London, United Kingdom.
[6] Kim, J., Pearl, J., 1987. Convice; a conversational inference consolidation engine. IEEE Trans. on Systems, Man and Cybernetics 17, 120{132.
[7] Lauritzen, S., Spiegelhalter, D., 1988. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society B 50 (2), 157-224.
[8] Murphy, K., 2002. Dynamic bayesian networks: Representation, inference and learning. Ph.D. thesis, University of california, Berkeley.
[9] Naïm, P., Wuillemin, P.-H., Leray, P., Pourret, O., Becker, A., 2007. Réseaux bayésiens. Eyrolles, ISBN : 2-212-11972-0, 3e edition.
[10] Pearl, J., 1998. Graphical models for probabilistic and causal reasoning. In: Gabbay, D. M., Smets, P. (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1: Quanti_ed Representation of Uncertainty and Imprecision. Kluwer Academic Publishers, Dordrecht, pp. 367-389.