



**HAL**  
open science

## Risk analysis on level crossings using a causal Bayesian network based approach

Ci Liang, Mohamed Ghazel, Olivier Cazier, El Miloudi El Koursi

► **To cite this version:**

Ci Liang, Mohamed Ghazel, Olivier Cazier, El Miloudi El Koursi. Risk analysis on level crossings using a causal Bayesian network based approach. WCTR 2016 - World Conference on Transport Research, Jul 2016, Shanghai, China. pp2167-2181, 10.1016/j.trpro.2017.05.418 . hal-02046035

**HAL Id: hal-02046035**

**<https://hal.science/hal-02046035v1>**

Submitted on 4 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016

# Risk analysis on level crossings using a causal Bayesian network based approach

Ci Liang<sup>a,b,c,\*</sup>, Mohamed Ghazel<sup>b,a,c</sup>, Olivier Cazier<sup>d,a</sup>, El-Miloudi El-Koursi<sup>b,a,c</sup>

<sup>a</sup> FCS Railenium, Valenciennes, France

<sup>b</sup> IFSTTAR-COSYS/ESTAS, Lille-Villeneuve d'Ascq, France

<sup>c</sup> University Lille 1, Lille-Villeneuve d'Ascq, France

<sup>d</sup> SNCF Réseau, Paris, France

---

## Abstract

Safety is a core issue in railway operation. In particular, Level Crossing (LX) safety is one of the most critical issues that railway stakeholders need to deal with. Accidents at European LXs account for about one-third of the entire railway accidents. They result in more than 300 deaths every year in Europe. However, due to non-deterministic causes, complex operation background and the lack of thorough statistical analysis based on detailed accident/incident data, the risk assessment of LXs remains a challenging task. In this paper, a general approach of Causal Statistic Risk Assessment based on hierarchical Causal Bayesian Networks (CSRA-CBN) is developed to analyze the various impacting factors which may cause accidents, and identify the factors which contribute most to the accidents at LXs, thus allowing for risk quantification. The detailed statistical analysis is carried out firstly according to the accident/incident database, then the CBN risk model is established based on the statistical results. In order to validate the effectiveness of this approach, we apply the CSRA-CBN approach on the basis of the accident data from SNCF, the French national railway operator. The CBN model allows for quantifying the causal relationships between the risk of accident occurring and the impacting factors considered. Moreover, the hierarchical structured modeling offers interesting benefits in terms of clarity, which makes it possible to highlight the complex factors influenced by a mass of parameters and identify the factors that contribute most to LX accidents. In addition, the main output of such a modeling system is conducive to improving safety at LXs, meanwhile, allowing for efficiently focusing on the effort/budget to make LXs safer.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

*Keywords:* Level crossings, causal statistic risk assessment, Bayesian networks, statistical analysis, railway safety;

---

\* Corresponding author. IFSTTAR, Lille-Villeneuve d'Ascq, 20 Rue Élisée Reclus, BP 70317 F-59666 Villeneuve d'Ascq Cedex.

Tel.: +33(0)320438310.

E-mail address: ci.liang@railenium.eu

2214-241X © 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

## 1. Introduction

### 1.1 Literature review on level crossing study

Accidents at railway level crossings (LXs) often give rise to serious material and human damage and seriously hamper railway safety reputation, although the majority of accidents are caused by road user violations. LX safety is one of the most critical issues in the railway field which needs to be improved urgently (Ghazel, 2009; Mekki et al., 2012; Liu et al., 2016). In 2012, there were more than 118,000 LXs in the 28 countries of the European Union (E.U.) which corresponds to an average of 5 LXs per 10 line-km (ERA, 2014). Accidents at European LXs account for about one-third of the entire railway accidents. They result in more than 300 deaths every year in Europe (Liu et al., 2016). In some European countries, accidents at LXs account for up to 50% of railway accidents (Ghazel and El-Koursi, 2014; Evans, 2011b). In the entire E.U. area, the overall number of fatalities per fatal accident in railways from 1990 to 2009 is 4.10, with no apparent long-term change over time (Evans, 2011a). In France, the railway network shows more than 15,000 LXs for 30,000 km of railway lines, which are crossed daily by 16 million vehicles on average, and around 13,000 LXs show heavy road and railway traffic (SNCF Réseau, 2011). In 2013, € 32 million was spent by the State, SNCF (the French national railway operator) and local authorities, to improve LX safety, and about € 36 million was invested in 2014. In recent years, a list of 300 prioritized LXs has been established and updated every year by SNCF Réseau, the French national railway infrastructure manager. Despite numerous measures taken to improve LX safety, SNCF Réseau counted 100 collisions at LXs leading to 25 deaths in 2014. This was half the total number of collisions at LXs a decade ago, but still too many (SNCF Réseau, 2015). If we want to significantly reduce the accidents and cost at LXs, it is necessary to develop further safety solutions to reach the most effective result.

In the literature pertaining to LX safety, a number of works can be found. In recent years, a systems analysis framework (Leveson, 2011; Read, 2016; Wilson, 2014) and a psychological schema theory (Salmon et al., 2013; Stanton, 2011) have been used to analyze the contributory factors underlying the accidents occurring at LXs. The study presented by Salmon et al. (2013) described a collision between a loaded semi-trailer truck and a train, occurring in North Victoria, caused by the truck driver crossing the LX while the barriers are closed. According to the investigation of OCI, the Office of the Chief Investigator, the truck driver in this study was not aware of the train and the activated state of the level crossing until it was too late to stop. A study conducted by Davey et al. (2008) discussed the intentional violation of vehicle drivers crossing LXs, particularly focusing on vehicle driver complacency due to the high level of familiarity. Tey et al. (2011) conducted an experiment to measure vehicle driver response to LXs equipped with stop signs (passive), flashing lights and half barriers with flashing lights (active) respectively. In this study, the vehicle driver responses result from both the field survey and a driving simulator.

Based on the above investigations, the main available studies dealing with LX safety have tended to take a qualitative approach to understand the potential reasons causing accidents at LXs. They employed surveys, interviews, focus group methods or driving simulators, rather than collecting real data. However, the aforementioned approaches show some limits. For instance, the reaction of vehicle drivers in simulation scenarios could differ from that in reality, due to the different levels of feeling of danger. Although those qualitative approaches are beneficial to explore the potential factors causing accidents, they cannot quantify the contribution degree of these factors. Thereby, quantitative approaches are indispensable if we want to understand the impacting factors thoroughly and enable the identification of practical design and improvement recommendations to prevent accidents at LXs.

### 1.2 Literature review on risk modeling

Current risk analysis methodologies are required to deal with increasingly complex systems with a large amount of configuration parameters. Therefore, such methodologies shall be equipped with the following characteristics:

- Strong modeling ability
- Easy to specify a risk scenario or a system
- High computational efficiency

In the domain of risk assessment, various methodologies are adopted for modeling and analyzing the systems or scenarios. Due to the combination of qualitative and quantitative analysis, the Fault Tree Analysis (FTA) developed by H.A. Watson at Bell Laboratories has been widely used for risk analysis since 1962 (Ericson and Li, 1999). FTA is a deductive and top-down method aimed at analyzing the effects of initiating faults and events on a complex system and offering the designer an intuitive high-level abstraction of the system, which can also be efficiently applied through cooperating with techniques such as Binary Decision Diagrams (Boudali and Dugan, 2005). Compared with Failure Mode and Effects Analysis (FMEA), which is an inductive and bottom-up analysis method aimed at analyzing the effects of single component or function failures on equipment or subsystems, FTA is good at showing how resistant a system is to single or multiple initiating faults. However, one obvious disadvantage of FTA is that it is not clear on failure mechanism, since the causal relationship between events is not a simple YES or NO (1 or 0). Therefore, FTA is prone to missing the possible initiating faults. In addition, traditional static fault trees cannot handle the sequential interaction and functional dependencies between components. Consequently, it is necessary to develop dynamic methodologies to overcome these weaknesses. Markov Chains (MCs) and their extensions have been mainly used for modeling complex dynamic system behavior and dependability analysis of dynamic systems. Two-state Markov switching multinomial logit models are introduced to explain unpredictable, unidentified or unobservable risk factors in road safety by Malyshkina and Mannering (2009). Although MCs can elaborate the statistical state transition of different variables, they cannot formalize causal relationships between various events.

Afterwards, risk analysis based on formal modeling flourished. In order to compare the effectiveness of two main Automatic Protection Systems (APSs) at LXs: two-half-barrier APS and four-half-barrier APS, Generalized Stochastic Petri Nets (GSPNs) were used to analyze the aleatory fluctuations of various parameters involved in the dynamics within the LX area by Ghazel and El-Koursi (2014). Over the last few years, Bayesian network (BN), a method of reasoning using probabilities, has been an increasingly popular method used for risk analysis of safety-critical systems or large and complicated dynamic systems (Chemweno et al., 2015). To obtain proper and effective risk control, risk planning should be performed based on risk causality which can provide more risk information for decision making. Thus, a model using BNs with causality constraints (BNCC) for risk analysis was proposed by Hu et al. (2013). The BN technique shows the flexibility of modeling, strong modeling power, high computational efficiency and, most importantly, the outstanding advantage involving causality analysis of both forward inference and reverse inference (Weber et al., 2012).

Therefore, based on the above investigation of risk analysis methodologies, a general approach of Causal Statistic Risk Assessment based on hierarchical Causal Bayesian Network (CSRA-CBN) is developed in this article to solve the risk assessment issue of LXs. We make a thorough statistical analysis on the accident/incident data involving LXs from SNCF and use the statistical results as the input source of CBN, then quantify the risk level of various potential factors and identify the factors which contribute most to the accidents at LXs, as well as their combined impact on LX safety.

The rest of this paper is organized as follows: In Section 2, the theory of probability which is the basic principle of CBN is stated firstly, then the integrated framework of CSRA-CBN is elaborated. In Section 3, we apply the CSRA-CBN approach on the LXs in France based on the accident/incident data from SNCF to validate the effectiveness of this approach. Statistical analysis is performed, risk model is established and simulation results are discussed in this section. Here, the GeNIe is used as a CBN modeling tool to formalize the causal relationships between accidents and the considered parameters. Finally, some concluding remarks and further directions are given in Section 4.

## **2. Causal statistic risk assessment based on causal Bayesian networks**

### *2.1. Causal Bayesian networks*

In the concept of Socio-technical systems (Leveson, 2011), potential hazards which may lead to accidents, including equipment failures, human errors, and other uncertain factors, such as environment aspects. In fact, the causality between the accidents and these uncertain factors exists, see Fig.1. Causality plays an important role in the process of inference. A functional intelligent identification model shall have the ability of making inference based on the causal knowledge.

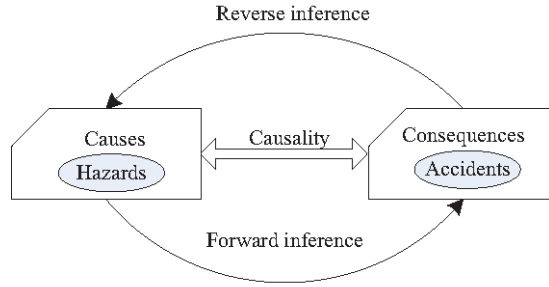


Fig. 1. Reasoning between hazards and accidents.

Causal Bayesian Network (CBN) employed to model causal relationships is a graphical model which consists of a structure and a set of parameters (Yehezkel and Lerner, 2009).  $CBN = (P, G)$ , where  $P$  represents the parameters of prior (conditional) probabilities that quantify the arcs and  $G$  represents the model structure.  $G = (V, A)$ , which is a Directed Acyclic Graph (DAG) comprised of nodes  $V$ s and arcs  $A$ s. Nodes  $V$ s represent random variables and directed arcs  $A$ s between pairs of nodes represent dependencies between the  $V$ s (Langseth, 2007).

CBN is developed on the theory of probability. Assume that there is a set of mutually exclusive events  $B_1, B_2, \dots, B_n$  and event  $A$ , in addition,  $P(A)$  can be expressed as follows:

$$P(A) = \sum_{i=1}^n P(B_i)P(A | B_i) \quad (1)$$

According to Bayes formula:

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)} \quad (2)$$

Therefore, Eq. (2) can be converted into:

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{P(A)} \quad (3)$$

$P(B_i)$  is the Prior Probability,  $P(B_i | A)$  is the posterior (conditional) probability. When evidence of the nodes is given in a CBN, the posterior probability can be calculated.

For any set of random variables in a CBN, the joint distribution can be calculated through conditional probabilities using the chain rule shown as follows:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{v=1, v \in V}^n P(X_v = x_v | X_{v+1} = x_{v+1}, \dots, X_n = x_n) \quad (4)$$

Due to the conditional independence,  $X_v$  only relates to its parent nodes  $Pa(X_v)$  and is independent of other nodes. So the last equation can be written as:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{v=1, v \in V}^n P(X_v = x_v | Pa(X_v)) \quad (5)$$

For more details about CBN, the reader can refer to a tutorial book of Bayesian networks edited by Pourret et al. (2008).

## 2.2. CSRA-CBN Framework

A paradigm for modeling should view an influential network not merely as passive parsimonious codes for storing factual knowledge, also as a computational architecture for reasoning about the knowledge. The CSRA-CBN is developed based on this principle. It means that the links in the network should be treated as the only pathways and activation units that direct and propel the flow of data in the process of querying and updating causal knowledge.

The framework of the CSRA-CBN approach can be illustrated in Fig. 2. The process of applying the CSRA-CBN approach consists in following five stages:

1) Identifying the risk scenario that we wish to focus on;

Before performing risk assessment, in order to set our research target, we must make a clear definition of the risk scenario boundary, and concentrate on this defined risk scenario to ensure that the follow-up study does not deviate from our original intention.

2) Collecting accident/incident data related to the defined risk scenario;

For predicting accident probability, assessing risk level and making improvement measures, accident/incident data related to the defined risk scenario need to be collected through monitoring the risk scenario. These data should be recorded in a dedicated database and as the source of the statistical analysis in step 3).

3) Developing statistical analysis on the accident/incident data;

Statistical analysis on the accident/incident data aims to identify impacting factors, further explore their statistic characteristics. The outcomes of statistical analysis should be as the importing evidence of CBN risk model.

4) Establishing the CBN risk model based on the statistical analysis;

In this step, we need to build the CBN model structure with respect to the causal relationships between identified impact factors and accidents, then employ the statistical results as the input evidence of the risk model.

5) Assessing risk level based on risk models;

In this step, accident probability will be quantified and risk level will be assessed. Moreover, impacting factors which contribute most to the entire risk we concerned will be identified to provide improvement measures for reducing the risk concerned and the cost brought by accidents.

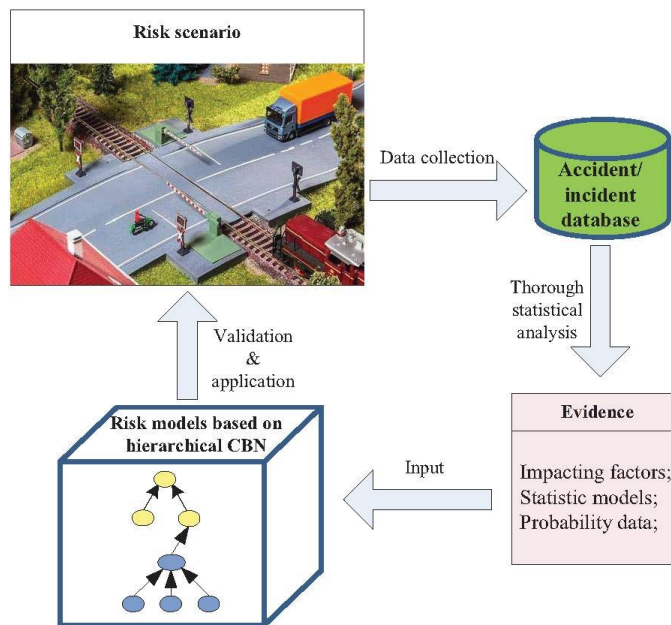


Fig. 2. Frame diagram of the CSRA-CBN approach.

In the fourth stage, the hierarchical network is adopted when establishing the risk model. The causal relationships between top nodes and intermediate nodes which cause the top nodes directly are established in the first layer of CBNs. In the second layer, the direct causal relationships between intermediate nodes and their root nodes are established. There is no indirect causality existing in any layer. Moreover, the network structure in each layer is single connected (SC), such as “Start” topology or “Tree” topology, which means one and only one path exists between any two nodes (without regard to the arc direction).

### 3. Application and validation of CSRA-CBN

In this section, the CSRA-CBN approach is applied on the LXs in France based on the accident/incident data collected by SNCF to assess the risk level and the probability of accidents occurring with regard to the various impacting factors taken into account, and further identify the important factors which contribute most to accidents occurring.

#### 3.1. Risk scenario identification and data collection

There are four types of LXs in France (SNCF, 2015) according to the data of SNCF:

- a) SAL4: Automated LXs with four half barriers and flashing lights;
- b) SAL2: Automated LXs with two half barriers and flashing lights;
- c) SAL0: Automated LXs with flashing lights but without barriers;
- d) Crossbuck LXs, without automatic signaling.

As shown in Table 1, SAL2 (more than 10,000) is the most widely used type of LX in France. Moreover, more than 4,000 accidents at SAL2 contributed most to the total number of accidents at LXs from 1974 to 2014.

Table 1. Accidents at different types of LXs in France from 1974 to 2014.

Type of LX	Number	# Accident
SAL4	>600	>400
SAL2	>10,000	>4,000
SAL0	>60	>50
Crossbuck LX	>3,000	>700

For the above reasons, “road users crossing French SAL2 LXs when trains are coming” is identified as the risk scenario. Detailed LX accident/incident data from 1974 to 2014 is collected by SNCF and input into a special database which we can access. Thereby, one analysis will consider the accident/incident data pertaining to the 40 years.

#### 3.2. Statistical analysis

According to the statistics of SNCF, geographical region, transport mode and traffic moment are essential factors which should be considered when improving the LX safety. Nevertheless, these factors were not thoroughly analyzed by the previous works. In this subsection statistical analysis with respect to traffic moment, different kinds of transport mode and different geographical regions is performed to establish the potential relationships between these principal impacting factors and accidents, and explore their distribution characteristics.

There are 21 geographical administrative regions in mainland France. Accidents which are caused by main types of transport mode: 1) motorized vehicle (MV), 2) pedestrian or bicycle (PB), are considered respectively to allow for making statistical analysis in different regions. The number of train-MV collisions, train-PB collisions and the number of SAL2 in 21 regions are presented respectively in Fig. 3.

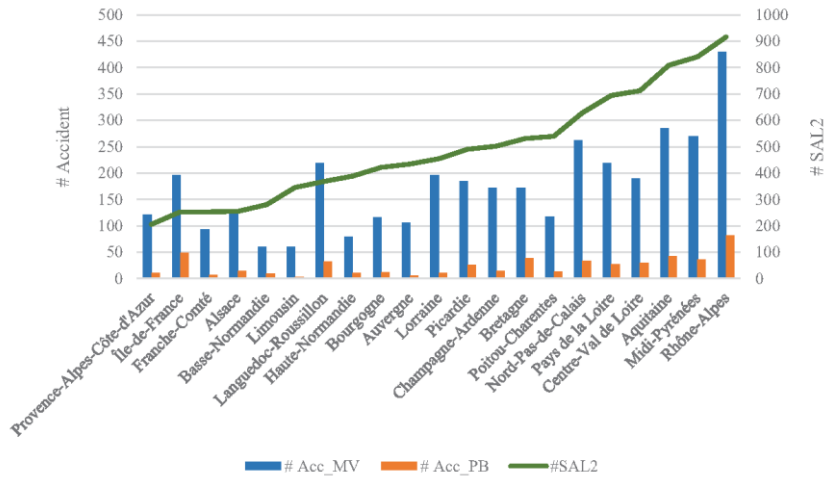


Fig. 3. Accidents caused by different transport modes at SAL2 in 21 French regions from 1974 to 2014.

### 3.2.1. General risk analysis

The general frequency of accidents occurring at each SAL2 per year are used to represent the general risk level involving total accidents, MV accidents, and PB accidents in different regions during the last 40 years. We can calculate the general frequency through the Eq. (6) in Definition 1:

- Definition 1

$$F_{G_i} = \frac{Nb\_acc_i}{Nb\_SAL2_i \times Nb\_year}, i = 1, 2, \dots, 21; \quad (6)$$

where  $F_{G_i}$  represents the general frequency in  $i^{th}$  region;  $Nb\_acc_i$  represents the number of accidents occurring in  $i^{th}$  region,  $Nb\_SAL2_i$  represents the number of SAL2 in  $i^{th}$  region, and  $Nb\_year$  represents the number of years of the considered period. Three kinds of general frequency related to total accidents, MV accidents, and PB accidents will be calculated. Correspondingly, the number of total accidents, MV accidents, and PB accidents are presented as  $Nb\_acc_i$  respectively when calculating these three kinds of general frequency.

Now that these three kinds of general frequency in each region are determined, maps of French regions with the frequency values presented are generated to show the frequency distribution in different regions. As shown in Fig. 4a, the general frequency value of total accidents in the red region (greater than 0.02) is the highest. The general frequency value of total accidents in the orange region (between 0.02 and 0.01) is the second highest, and the general frequency values of total accidents in the green region (less than 0.01) is the lowest. When we analyze the frequency figures in detail, we find that the risk is most serious in Île-de-France with a frequency of more than 0.02; Languedoc-Roussillon takes the second place with the frequency of about 0.017 followed by Provence-Alpes-Côte-d'Azur with the frequency of about 0.016. On the other hand, Limousin has the lowest risk with the frequency of about 0.005. Haute-Normandie and Basse-Normandie occupy the second and the third places of lowest risk successively. Turning to Fig. 3, the number of accidents occurring in Île-de-France is not the highest, but, when considering the corresponding number of LXs, this leads to the most serious risk. Conversely, although accidents in Rhône-Alpes are the most with the number of 521, the risk is not the highest due to a large number of LXs in this region. Limousin, Haute-Normandie and Basse-Normandie have significantly fewer accidents than any other region, less than 100 in 40 years. That is why they have the lowest risk.



In Fig. 4b and Fig. 4c, the general frequency of accidents caused by motorized vehicles, pedestrians and bicycles in different regions is shown respectively. Considering the accidents caused by motorized vehicles, the distribution of frequency in different regions in Fig. 4b is relatively consistent with the distribution shown in Fig. 4a. The only exception is Champagne-Ardenne. However, in Fig. 4c, as for the accidents caused by pedestrians and bicycles, the distributions of frequency in different regions are very different from the distribution shown in Fig. 4a.

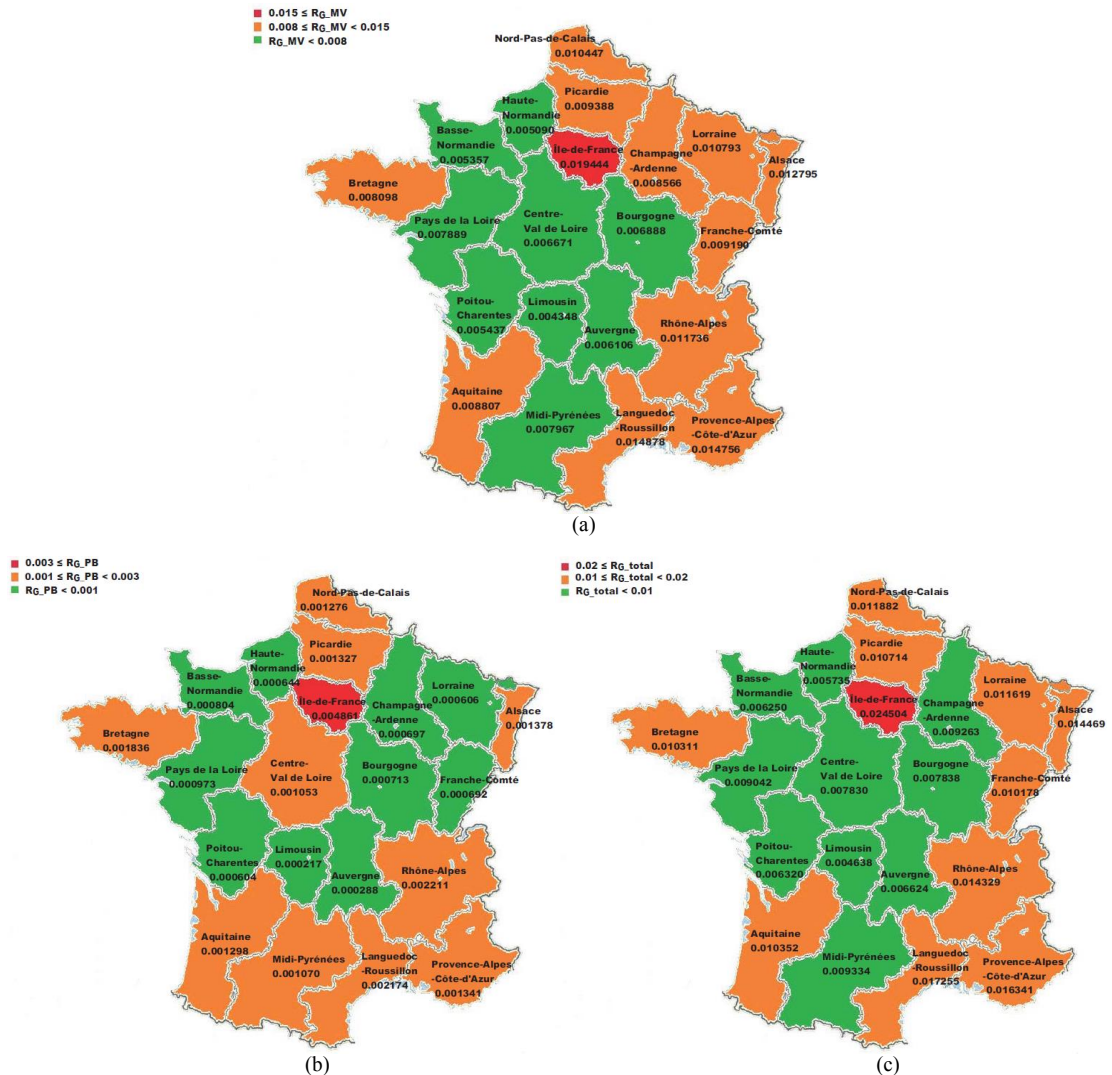


Fig. 4. (a) Frequency of total accidents distributing in different regions; (b) Frequency of MV accidents distributing in different regions; (c) Frequency of PB accidents distributing in different regions.

By analyzing the results shown in Fig. 4 and Fig. 5, we conjecture that the main mode causing accidents at SAL2 is the motorized vehicle. Moreover, as the frequency of accidents involving motorized vehicle collisions increases, the total accident frequency in the overwhelming majority of regions increases accordingly, as illustrated in Fig. 5. Further discussion and validation will be elaborated in section 3.3.

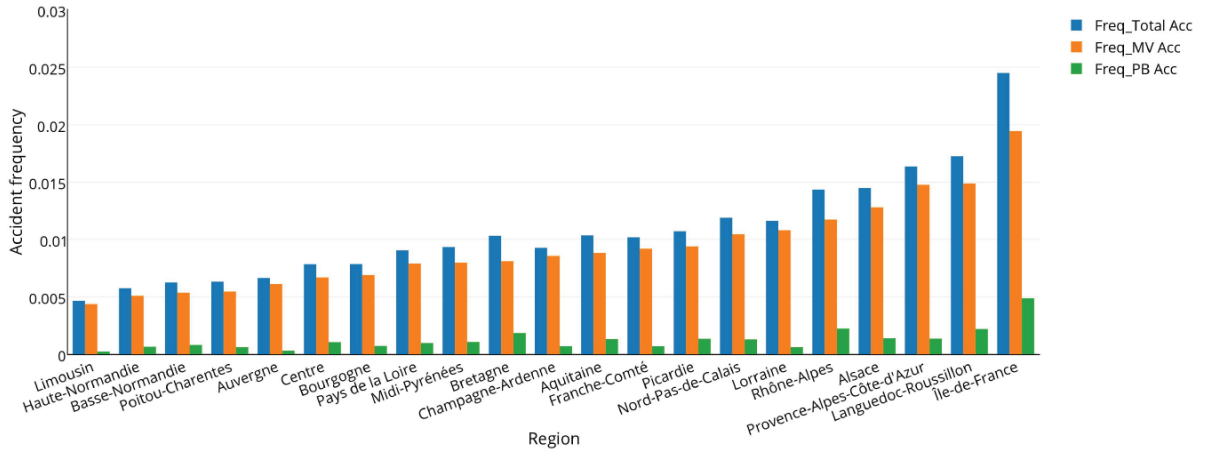


Fig. 5. Comprehensive frequency of accident in different regions.

### 3.2.2. Normalized risk analysis based on the traffic moment

As mentioned in section 3.2.1, the motorized vehicle is the main transport mode causing accidents at SAL2. In this section, we will focus on the impact of traffic moment on the whole risk level related to MV accidents. We suspect that this parameter is one of the main parameters impacting LX risk level. Indeed, traffic moment gives the combined traffic (train / MV) at the LX and is defined as follows:

- Definition 2

$$\text{Traffic moment} = \text{road traffic frequency} \times \text{railway traffic frequency} \quad (7)$$

In this section, we use “ $M$ ” for short to denote the traffic moment. In order to analyze the impact of  $M$ , we use the square root of  $M$  at SAL2 to normalize the frequency of accidents caused by train-MV collisions, which can represent the normalized frequency in different regions. The formula we adopt to establish the normalized frequency is the following:

- Definition 3

$$F_{N_i} = \frac{\sum_{j=1}^{Nb\_SAL2\_reg_i} \frac{Nb\_acc\_MV_j}{\sqrt{M_j}}}{Nb\_SAL2\_reg_i \times Nb\_year}, i = 1, 2, \dots, 21; \quad (8)$$

where  $F_{N_i}$  represents the normalized frequency of accidents caused by train-MV collisions in  $i^{th}$  region;  $Nb\_acc\_MV_j$  represents the number of accidents caused by train-MV collisions at  $j^{th}$  SAL2 in  $i^{th}$  region;  $M_j$  represents the moment at the  $j^{th}$  SAL2 in  $i^{th}$  region;  $Nb\_SAL2\_reg_i$  represents the number of SAL2 in  $i^{th}$  region, and  $Nb\_year$  represents the number of years of the considered period.

As shown in Fig. 6, the normalized frequency caused by train-MV collisions occurring at SAL2 LXs in each region is computed. Unlike Fig. 4b, Fig. 6 illustrates that the risk level is the highest in Île-de-France with the normalized frequency of  $7.31 \times 10^{-4}$ ; Alsace takes the second place followed by Languedoc-Roussillon and Provence-Alpes-Côte-d'Azur. Through detailed analysis, we find that the general frequency of MV accidents in these 4 regions (cf. Fig. 4a) is also the highest.

For the purpose of further analyzing the impact of  $M$  on the risk level in each region, SAL2 LXs are classified by the category of  $M$ . In order to make the number of SAL2 in each  $M$  group to be as far as possible similar to any other group in each region, the moment groups have been defined in such a way that the number of SAL2 LXs in every group belongs to the interval [102, 155]. For example, in Table 2, there are three  $M$  categories in Auvergne, which are “ $0 \leq M < 750$ ”, “ $750 \leq M < 5,000$ ”, “ $5,000 \leq M \leq 401,850$ ”, with the corresponding numbers of SAL2: 143, 144, 147, respectively. Besides, 401,850 is the maximum  $M$  in this region. In this way, we can make risk analysis with regard to these SAL2 groups, thus making it possible to highlight the risk level related to different categories of  $M$ . Logically, if the same number of MV accidents or more MV accidents occur per year at the SAL2 with small  $M$  than at the SAL2 with large  $M$ , this indicates that the SAL2 with small  $M$  has higher risk level than the SAL2 with large  $M$ .

The normalized frequency of collisions regarding the different  $M$  categories in each region is shown in Table 2. The  $M$  categories marked in red color have the most serious risk in various regions. We find that SAL2 LXs in the category of smallest  $M$  have the highest risk in 86% of the regions, except in Auvergne, Champagne-Ardenne and Nord-Pas-de-Calais where the SAL2 LXs with the second smallest  $M$  have the highest risk. On the other hand, SAL2 LXs in the category of largest  $M$  have the lowest risk in 81% of the regions. These statistical results prove the logical inference mentioned above that the SAL2 with small  $M$  has higher risk level than the SAL2 with large  $M$ , when considering the ratio of the number of MV accidents occurring per year at the SAL2 to the traffic moment of the SAL2. Further discussion and validation will be elaborated in section 3.3.

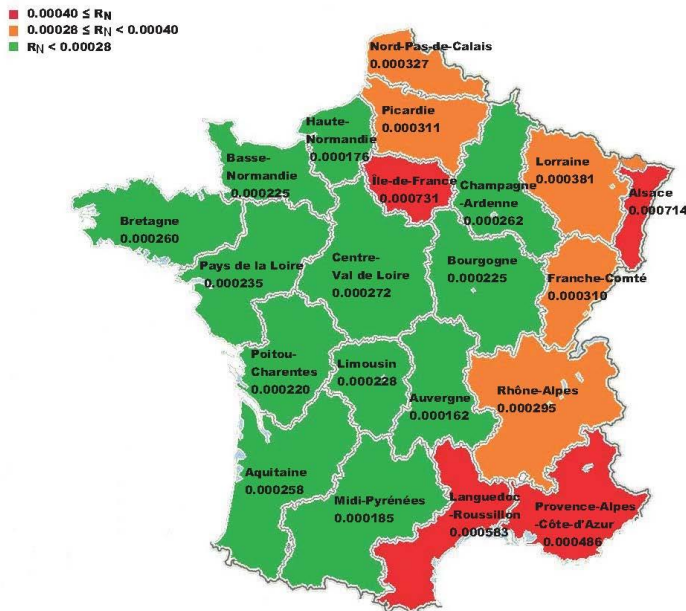


Fig. 6. Normalized frequency of MV accidents distributing in different regions.

Table 2. Normalized frequency of MV accidents at SAL2 in each region according to the category of “moment”.

Normalized frequency of MV accidents at SAL2 in each region according to the category of “moment”						
Auvergne	$M [0, 750)$	$M [750, 5,000)$	$M [5,000, 401,850]$			
	0.000049	0.000100	0.000078			
Haute-Normandie	$M [0, 500)$	$M [500, 4,500)$	$M [4,500, 382,788]$			
	0.000190	0.000089	0.000072			
Midi-Pyrénées	$M [0, 460)$	$M [460, 1,150)$	$M [1,150, 3,000)$	$M [3,000, 9,000)$	$M [9,000, 35,000)$	$M [35,000, 1169,782]$
	0.000202	0.000068		0.000066		

			0.000072		0.000095	0.000065	
<b>Basse-Normandie</b>	$M [0, 2,100)$ 0.000147	$M [2,100, 555,600]$ 0.000067					
<b>Poitou-Charentes</b>	$M [0, 500)$ 0.000302	$M [500, 2,200)$ 0.000064	$M [2,200, 10,500)$ 0.000098	$M [10,500, 1169,782]$ 0.000078			
<b>Bourgogne</b>	$M [0, 1,000)$ 0.000224	$M [1,000, 8,000)$ 0.000130	$M [8,000, 720,000]$ 0.000061				
<b>Limousin</b>	$M [0, 700)$ 0.000269	$M [700, 5,000)$ 0.000058	$M [5,000, 337,524]$ 0.000055				
<b>Pays de la Loire</b>	$M [0, 500)$ 0.000175	$M [500, 1,400)$ 0.000121	$M [1,400, 3,900)$ 0.000111	$M [3,900, 15,000)$ 0.000122	$M [15,000, 815,400]$ 0.000062		
<b>Champagne-Ardenne</b>	$M [0, 420)$ 0.000129	$M [420, 1,800)$ 0.000285	$M [1,800, 8,000)$ 0.000099	$M [8,000, 256,410]$ 0.000185			
<b>Centre-Val de Loire</b>	$M [0, 350)$ 0.000201	$M [350, 1,400)$ 0.000121	$M [1,400, 4,300)$ 0.000092	$M [4,300, 18,000)$ 0.000087	$M [18,000, 688,320]$ 0.000059		
<b>Aquitaine</b>	$M [0, 600)$ 0.000162	$M [600, 1,600)$ 0.000135	$M [1,600, 3,800)$ 0.000095	$M [3,800, 11,000)$ 0.000093	$M [11,000, 40,000)$ 0.000091	$M [40,000, 585,644]$ 0.000055	
<b>Bretagne</b>	$M [0, 800)$ 0.000324	$M [800, 4,200)$ 0.000117	$M [4,200, 18,000)$ 0.000068	$M [18,000, 2244,000]$ 0.000044			
<b>Rhône-Alpes</b>	$M [0, 320)$ 0.000297	$M [320, 1,400)$ 0.000122	$M [1,400, 3,600)$ 0.000106	$M [3,600, 11,000)$ 0.000101	$M [11,000, 35,000)$ 0.000080	$M [35,000, 100,000)$ 0.000054	$M [100,000, 1121,088]$ 0.000047
<b>Franche-Comté</b>	$M [0, 2,500)$ 0.000143	$M [2,500, 271,936]$ 0.000101					
<b>Picardie</b>	$M [0, 600)$ 0.000314	$M [600, 2,550)$ 0.000135	$M [2,550, 14,000)$ 0.000127	$M [14,000, 817,000]$ 0.000101			
<b>Nord-Pas-de-Calais</b>	$M [0, 450)$ 0.000079	$M [450, 3,000)$ 0.000203	$M [3,000, 10,000)$ 0.000082	$M [10,000, 45,000)$ 0.000063	$M [45,000, 1477,229]$ 0.000059		
<b>Lorraine</b>	$M [0, 1,350)$ 0.000250	$M [1,350, 8,000)$ 0.000122	$M [8,000, 402,500]$ 0.000099				
<b>Provence-Alpes-Côte-d'Azur</b>	$M [0, 5,700)$ 0.000212	$M [5,700, 921,250]$ 0.000094					
<b>Île-de-France</b>	$M [0, 11,500)$ 0.000285	$M [11,500, 752,854]$ 0.000099					
<b>Alsace</b>	$M [0, 12,000)$ 0.000513	$M [12,000, 767,088]$ 0.000113					
<b>Languedoc-Roussillon</b>	$M [0, 1,250)$ 0.000442	$M [1,250, 7,200)$ 0.000134	$M [7,200, 1428,000]$ 0.000152				

$M [a, b)$ :  $a \leq \text{moment} < b$ ;  $M [c, d]$ :  $c \leq \text{moment} \leq d$ .

### 3.3. Risk model establishment and discussion

#### 3.3.1. Risk model establishment

In this section, we establish the CBN risk model based on the analysis made so far, particularly the statistical results in section 3.2 to assess the risk level related to geographical region, transport mode and traffic moment. The “GeNie” software tool is used as the CBN modeling mean.

Firstly, the structure of risk model is established with respect to the causal relationships between the impacting factors and accidents. As shown in Fig. 7, the initial hierarchical model contains two layers: Sub-layer 1 is the sub-model in the “Risk\_MV” node and Sub-layer 2 is the sub-model in the “Risk\_PB” node. States of moment group are defined as “M\_Region” (e.g., the “M\_Alsace” node) of Sub-layer 1; states of MV accidents and PB accidents occurring in different regions are defined in region nodes of Sub-layer 1 and Sub-layer 2 (e.g., the “Alsace” node), respectively.

Assume that all nodes only have two states in CBNs. If one node has  $n$  parent nodes, the combinations of states in this node will be up to  $2^{(n+1)}$ . In our model, some parent nodes contain more than two states. This means there are more than  $2^{22}$  combinations in “Risk\_MV” and “Risk\_PB” nodes. Moreover, more than half of the combinations are invalid (some cases do not exist in reality, e.g.,  $risk_i$  only corresponds to the risk value in region  $i$  and it cannot exist when regarding more than two different regions). Thereby, in order to avoid the state space explosion (SSE), we optimize the structure of the model and eliminate the invalid combinations of states, then input evidence into the model based on the statistics in Section 3.2. As shown in Fig. 8, “Part 3” is used to replace “Part 1” in Fig. 7 and “Part 4” is used to replace “Part 2” in Fig. 7. Table 3 offers a summary of the states in each node of the optimized risk model. In order to avoid the SSE problem, we determine 3 groups of moment which contain all the SAL2 LXs and the number of SAL2 LXs in each group is almost equal.

Table 3. States of each node in the optimized risk model.

Nodes	States
<b>Accidents_MV_region</b> <b>Accidents_PB_region</b>	Île_de_France 1, Languedoc_Roussillon 2, Provence_Alpes 3, Alsace 4, Rhone_Alpes 5, Lorraine 6, Nord_Pas_de_Calais 7, Picardie 8, Franche_Comte 9, Aquitaine 10, Champagne_Ardenne 11, Bretagne 12, Midi_Pyrenees 13, Pays_de_la_Loire 14, Bourgogne 15, Centre_Val_de_Loire 16, Auvergne 17, Poitou_Charentes 18, Basse_Normandie 19, Haute_Normandie 20, Limousin 21;
<b>Moment</b>	M 0_1485 (moment between 0 and 1485), M 1485_10500 (moment between 1485 and 10500), M 10500_244000 (moment between 10500 and 244000);
<b>Risk_MV</b> <b>Risk_PB</b> <b>Risk_Accidents</b>	False, True;

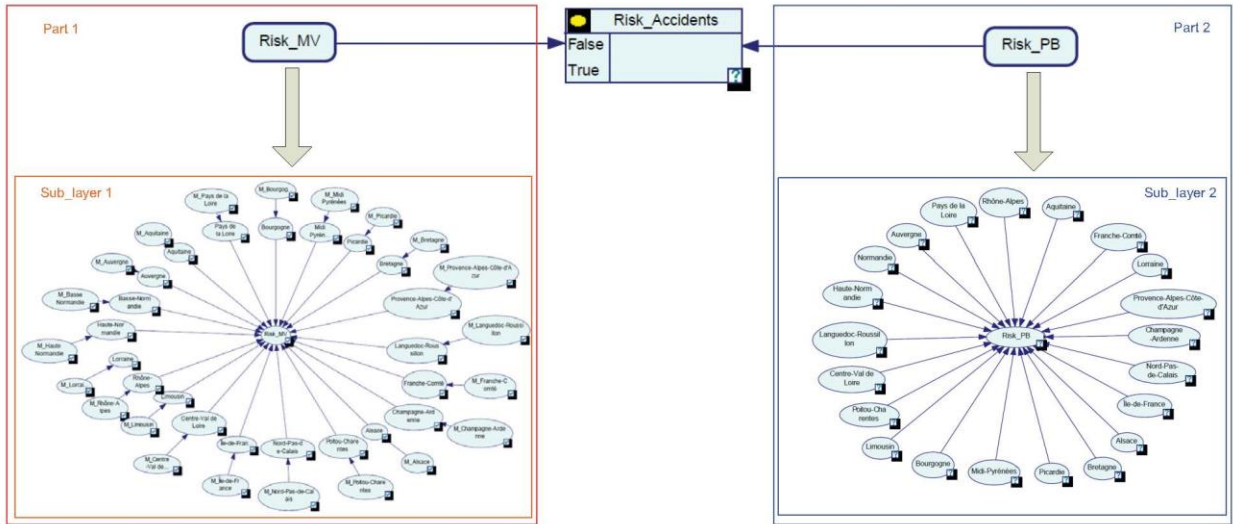


Fig. 7. The initial hierarchical risk model.

### 3.3.2. Accident prediction and risk level assessment

We can estimate the probability of accidents occurring with regard to the impact of these factors interacting with each other on accidents. As shown in Fig. 9, the quantified probability is obtained through forward inference. The estimated probability of accidents occurring impacted by the interaction of geographical region, transport mode and traffic moment is lower than 0.0065, and this result is consistent with the statistics developed by SNCF.

Subsequently, “Risk\_Accidents = True” is configured as the target evidence. In this way, we can assess the risk level of various impacting factors through reverse inference. Detailed results are offered in Fig. 10. We can find that when accidents have already occurred, the risk of accidents caused by motorized vehicles contributes 98% to the overall risk. However, the risk of accidents caused by pedestrians and bicycles only contributes 2% to the overall risk. Moreover, in terms of geographical region, the normalized risk level of MV accidents in Île-de-France is the highest, which induces 14% of the entire MV risk. The risk of MV accidents in Alsace takes the second place followed by Languedoc-Roussillon and Provence-Alpes-Côte-d’Azur. In addition, the risk level of PB accidents is also the highest in Île-de-France. These outcomes are consistent with the statistics in Section 3.2. Paris is indeed included in Île-de-France. In addition, Paris holds 6 main railway stations, among which the north railway station “Gare du Nord” is one of the most important railway stations in the world. This explains at least partially the fact that the risk of accidents occurring is the highest in Île-de-France.

In terms of moment group, the smallest moment group “M 0\_1485” has the highest risk level due to inducing 61% of the risk caused by motorized vehicles and the largest moment group “M 10500\_244000” has the lowest risk level for only inducing 19% of the risk caused by motorized vehicles. This result also strongly supports the logic inference mentioned in Section 3.2.2 that the SAL2 with small moment has higher risk level than the SAL2 with large moment when considering the ratio of the number of MV accidents occurring per year at the SAL2 to the moment value of the SAL2.

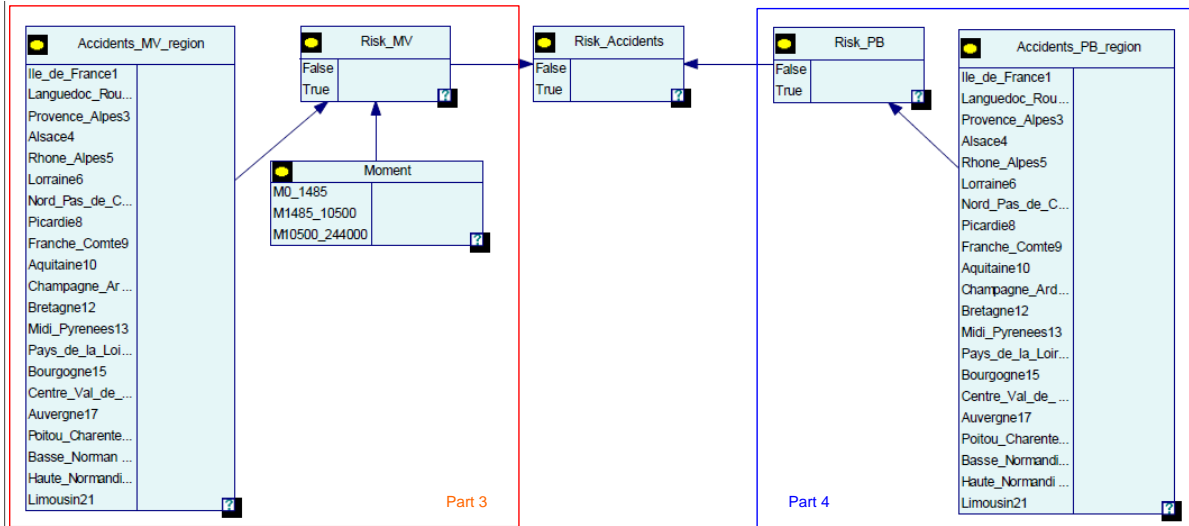


Fig. 8. The optimized risk model.

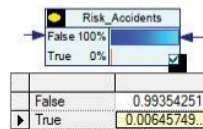


Fig. 9. Estimated probability of accidents occurring.

Furthermore, the sensitivity analysis is performed to elaborate the sensitivity for “Risk\_Accidents = True” related to various combination conditions. Through forward running the model, the probability of “Risk\_Accidents = True” is obtained as  $6.45749 \times 10^{-3}$ . Here, we use  $P(\text{Risk\_Accidents} = \text{True})$  and  $P(\text{Risk\_MV} = \text{False})$  to represent the probability of “Risk\_Accidents = True” and “Risk\_MV = False”, respectively. In this case, the current value  $V_c$  of  $P(\text{Risk\_MV} = \text{False} \mid \text{Accidents\_MV\_region} = \text{Île-de-France 1}, \text{Moment} = \text{M 0\_1485})$  is 0.95668.

The Sensitivity Tornado Diagram (STD) is offered by Fig. 11. The horizontal bars are viewed as two parts according to the central vertical axis  $P(\text{Risk\_Accidents} = \text{True}) = 6.45749 \times 10^{-3}$ .  $P(\text{Risk\_Accidents} = \text{True})$  in the green part is lower than  $6.45749 \times 10^{-3}$  and  $P(\text{Risk\_Accidents} = \text{True})$  in the red part is higher than  $6.45749 \times 10^{-3}$ . As the value range of impacting parameter is extended to the interval  $[V_c - 0.1V_c, V_c + 0.1V_c]$  (setting the parameter spread degree as 0.1), the values of  $P(\text{Risk\_Accidents} = \text{True})$  belong to an interval  $[V_{min}, V_{max}]$  and distribute with respect to the values of impacting parameter in the whole range  $[V_c - 0.1V_c, V_c + 0.1V_c]$ , accordingly.

The STD highlights the greatest contributor “Risk\_MV = False | Accidents\_MV\_region = Île-de-France 1, Moment = M 0\_1485” (the combination of various impacting factors) to the overall risk, that is no MV accidents occurring under the condition of “Accidents\_MV\_region = Île-de-France 1, Moment = M 0\_1485”. Here, the value range of  $P(\text{Risk\_MV} = \text{False} \mid \text{Accidents\_MV\_region} = \text{Île-de-France 1}, \text{Moment} = \text{M 0\_1485})$  is in the interval  $[0.86101, 1]$ , the value of  $P(\text{Risk\_Accidents} = \text{True})$  belongs to the interval  $[0.00586, 0.00778]$  distributing with respect to the values of  $P(\text{Risk\_MV} = \text{False} \mid \text{Accidents\_MV\_region} = \text{Île-de-France 1}, \text{Moment} = \text{M 0\_1485})$  in the whole range  $[0.86101, 1]$ . The second greatest contributor is “ $P(\text{Risk\_MV} = \text{False} \mid \text{Accidents\_MV\_region} = \text{Alsace 4}, \text{Moment} = \text{M 0\_1485})$ ” (i.e., no MV accident occurring in the moment group M 0\_1485 of Alsace);  $P(\text{Risk\_MV} = \text{False} \mid \text{Accidents\_MV\_region} = \text{Provence\_Alpes 3}, \text{Moment} = \text{M 0\_1485})$ , that is, no MV accident occurring in the moment group M 0\_1485 of Provence-Alpes-Côte-d’Azur, and  $P(\text{Risk\_MV} = \text{False} \mid \text{Accidents\_MV\_region} = \text{Languedoc\_Roussillon 2}, \text{Moment} = \text{M 0\_1485})$ , that is, no MV accident occurring in the moment group “M 0\_1485” of Languedoc-Roussillon, almost have the same degree of influence on the overall risk and take the third and fourth places. Besides, the overall risk is not very sensitive to other impacting parameters.

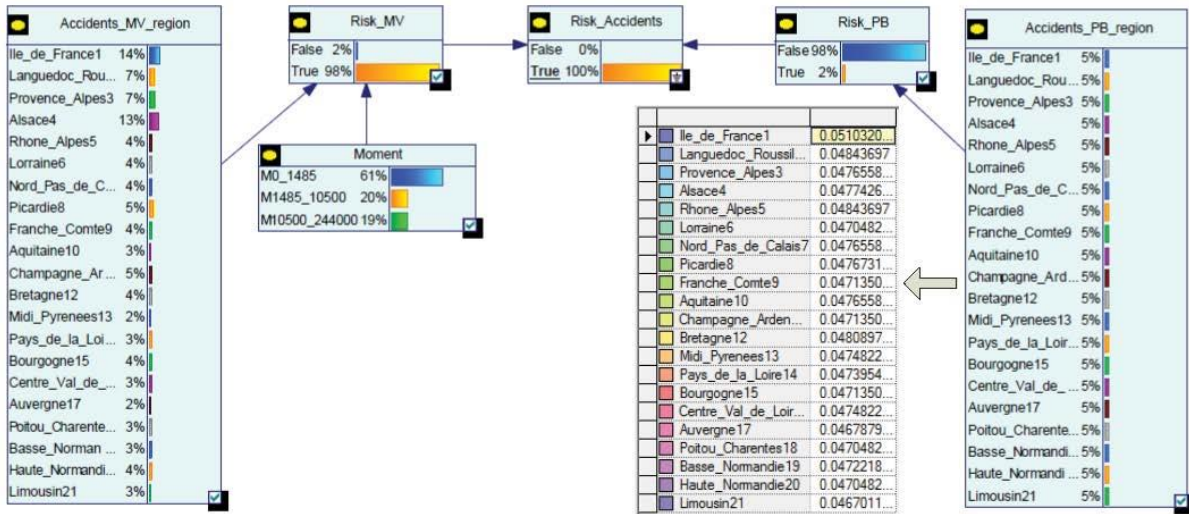


Fig. 10. Risk level assessment of various impacting factors.

The above analysis result is logical since the main transport mode causing accidents is MV and the probability of “no MV accidents occurring” is very high (after all, MV accidents occurring is a small probability event in reality). A small change in the probability of “Risk\_MV = False” can cause a dramatic change in the probability of “Risk\_Accidents = True”. This analysis result further proves that, if we want to decrease the risk at SAL2 LXs significantly, MV accidents shall be reduced considerably. In particular, more attention shall be paid to MV accidents related to the smallest moment groups. In reverse, all the above outcomes have verified the validity of the risk model by comparing with the statistics in Section 3.2.

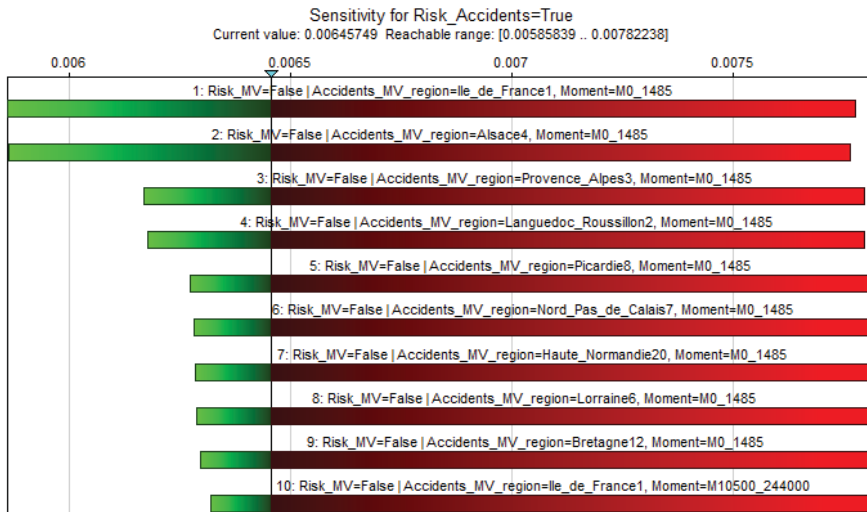


Fig. 11. Sensitivity Tornado Diagram of “Risk Accidents = True”.



#### 4. Conclusions

The contributions of the present study are as follows: an integrated risk assessment approach called CSRA-CBN is developed. In the framework of CSRA-CBN, the risk scenario is firstly identified. Then, thorough statistical analysis is performed based on the accident/incident data of the risk scenario. Finally, the risk model is established based on the statistical analysis to quantify the probability of accidents occurring and the contribution of various impacting factors, thus allowing for determining the important contributors to the overall risk.

The application of CSRA-CBN to investigate French LX safety allows us to identify the main LX risk factors and quantify their respective contributions to the overall LX risk. Thanks to the CSRA-CBN approach, various significant factors pertaining LX safety, i.e., the involved road transport mode, geographical regions and the traffic moment, have been analyzed meticulously. More specifically, the statistical analysis on general frequency of the entire LX accidents, train-MV LX accidents and train-PB LX accidents in different French regions is performed. Then, further investigation is carried out to analyze the train-MV LX accident frequency normalized by the square root of “traffic moment” according to various “traffic moment” categories in different regions. Afterwards, the risk model is established to assess the risk level influenced by the geographical region, the transport mode and the traffic moment, based on the causal relationships between these impacting factors and accidents occurring and the aforementioned statistical results. The sensitivity analysis based on the risk model is performed to highlight the greatest contributor to the overall risk at LXs. The greatest contributor is identified among the combinations of various impacting factors in different situations. The effectiveness of CSRA-CBN approach is validated by the obtained statistical results, the risk model outputs and the sensitivity analysis outcomes of the whole risk model.

It is worth noticing that the CSRA-CBN is a general approach which can be applied on various risk assessment studies. In future works, we will consider further impacting factors and bring into play Bayesian risk models to quantify the causal relationships between the factors and accidents occurring, and assess their impact on the risk level. The first version of risk model will be enriched while integrating further factors according to real data and causality analysis that we will perform on the basis of fault events.

#### Acknowledgements

This work has been in the framework of “MORIPAN project: Modèle de Risque pour les Passages à Niveau” within the Railenium technological research institute, in cooperation with the National Society of French Railways Networks (SNCF Réseau) and the French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR).

#### References

- Ghazel, M., 2009. Using stochastic Petri nets for level-crossing collision risk assessment. *IEEE Trans. on Intelligent Transportation Systems* 10 (4), 668-677.
- Mekki, A., Ghazel, M., Toguyeni, A., 2012. Validation of a new functional design of automatic protection systems at level crossings with model-checking techniques. *IEEE Trans. on Intelligent Transportation Systems* 13 (2), 714-723.
- Liu, B., Ghazel, M., Toguyeni, A., 2016. Model-Based Diagnosis of Multi-Track Level Crossing Plants. *IEEE Trans. on Intelligent Transportation Systems* 17 (2), 546-556.
- European Railway Agency (ERA), 2014. Railway safety performance in the European Union. 9(2) Agency Regulation, 881/2004/EC.
- Ghazel, M., El-Koursi, E.-M., 2014. Two-half-barrier level crossings versus four-half-barrier level crossings: a comparative risk analysis study. *IEEE Trans. on Intelligent Transportation Systems*, 15(3), 1123-1133.
- Evans, A.W., 2011. Fatal accidents at railway level crossings in Great Britain 1946–2009. *Accid. Anal. Prev.* 43(5), 1837-1845.
- SNCF Réseau, 2011. World Conference of Road Safety at Level Crossings (Journée Mondiale de Sécurité Routière aux Passages à Niveau), France. From <http://www.planetoscope.com/automobile/1271-nombre-de-collisions-aux-passages-a-niveau-en-france.html>
- SNCF Réseau, 2015. 8th National Conference of Road Safety at Level Crossings (8ème Journée Nationale de Sécurité Routière aux Passages à Niveau). From <http://www.sncf-reseau.fr/fr/dossier-de-presse-8eme-journee-nationale-de-securite-routiere-aux-passages-a-niveau>
- Leveson, N.G., 2011. *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press, Cambridge.
- Read, G.J., Salmon, P.M., Lenné, M.G., Stanton, N.A., 2016. Walking the line: Understanding pedestrian behaviour and risk at rail level crossings with cognitive work analysis. *Applied ergonomics* 53, 209-227.
- Wilson, J.R., 2014. Fundamentals of systems ergonomics/human factors. *Applied Ergonomics* 45, 5-13.

- Salmon, P.M., Read, G.J., Stanton, N.A., Lenné, M.G., 2013. The crash at Kerang: Investigating systemic and psychological factors leading to unintentional non-compliance at rail level crossings. *Accident Analysis & Prevention* 50, 1278-1288.
- Stanton, N.A., Walker, G.H., 2011. Exploring the psychological factors involved in the Ladbroke Grove rail accident. *Accident Analysis & Prevention* 43 (3), 1117-1127.
- Davey, J., Wallace, A., Stenson, N., Freeman, J., 2008. The experiences and perceptions of heavy vehicle drivers and train drivers of dangers at railway level crossings. *Accident Analysis & Prevention* 40 (3), 1217-1222.
- Tey, L.S., Ferreira, L., Wallace, A., 2011. Measuring driver responses at railway level crossings. *Accident Analysis & Prevention* 43 (6), 2134-2141.
- Ericson, C. A., Li, C., 1999. Fault Tree Analysis, Proc. 17th International Systems Safety Conference. Orlando, Florida, 1-9.
- Boudali, H., Dugan, J.B., 2005. A discrete-time Bayesian network reliability modeling and analysis framework. *Reliability Engineering and System Safety*, 87 (3), 337-349.
- Malyskhina, N. V., Mannering, F. L., 2009. Markov switching multinomial logic model: an application to accident-injury severities. *Accid. Anal. Prev.* 41 (4), 829-838.
- Ghazel, M., El-Koursi, E.-M., 2014. Two-half-barrier level crossings versus four-half-barrier level crossings: a comparative risk analysis study. *IEEE Trans.on Intelligent Transportation Systems*, 15(3), 1123-1133.
- Chemweno, P., Pintelon, L., Van Horenbeek, A., Muchiri, P., 2015. Development of a risk assessment selection methodology for asset maintenance decision making: An analytic network process (ANP) approach. *International Journal of Production Economics*, 170, 663-676.
- Hu, Y., Zhang, X., Ngai, E. W. T., Cai, R., Liu, M., 2013. Software project risk analysis using Bayesian networks with causality constraints. *Decision Support Systems*, 56, 439-449.
- Weber, P., Medina-Oliva, G., Simon, C., Iung, B., 2012. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, 25(4), 671-682.
- Yehezkel, R., Lerner B., 2009. Bayesian network structure learning by recursive autonomy identification. *Journal of Machine Learning Research*, 10, 1527-1570.
- Pourret, O., Naïm, P., Marcot, B., 2008. *Bayesian networks: a practical guide to applications*. John Wiley & Sons.
- Langseth, H., Portinale, L., 2007. Bayesian networks in reliability. *Reliab. Eng. Syst. Saf*, 92, 92-108.
- SNCF, 2015. Research on the material of level crossing in 2014, France.