UNIVERSITY OF CANTABRIA
Department of Applied Mathematics and Computational Sciences
39071 Santander, Spain

Doctoral Thesis

A Bayesian Network Approach for Probabilistic Safety Analysis of Traffic networks

Análisis Probabilista de Seguridad de Redes de Tráfico Basado en Redes Bayesianas

Author:

Elena Mora Villazán

Directors:

Enrique Castillo Ron
Zacarías Grande Andrade

Santander, October 2017

To my parents to whom I owe everything.

Once you stop learning, you start dying.
Albert Einstein

Acknowledgments

The result of this Thesis would not be possible without Enrique Castillo. I will be eternally grateful to him for helping me in an unconditional way, working hard but always with enthusiasm and passion. He has transmitted to me his effort and dedication in everything he does and I feel very fortunate, both professionally and personally, to have had the opportunity to work with him. I would like also thank to Zacarías Grande for his many contributions and great effort employed in the project.

To the Cantabria Government, Gesvican, the University of Cantabria, and the Spanish Royal Academy of Engineering for making possible to perform a probabilistic safety assessment of the Cantabrian highways and roads. Special thanks to the Spanish Royal Academy of Engineering for providing me the necessary means and economic support to work with Enrique Castillo.

To the Department of Applied Mathematics and Computational Sciences for allowing me to start teaching in the University of Cantabria, and in particular, to the people who have been there to help me when I needed.

To Alan O'Connor and María Nogal, who gave me the opportunity to work within one of their projects, learn a different way of working and meet a new culture. I am also grateful to the Trinity College for letting me use its facilities and for the attention I received.

Of course, I must thank to my family. They have been always with me, even when we were kilometers away, and helped me at any circumstance. They taught me their values and principles, to work hard and get up if I fall, and to treat with respect and education to anyone.

Should not be fair to forget Álvaro who helped me at the weak moments giving his unconditional support and patience.

Finally, to God for enable me have my family, friends and lead me in life.

Contents

ABSTRACT

This thesis consists mainly of two parts. The first one is based on a review of concepts and models that have been useful to carry out this study. The second is the one that covers the majority of the work, where a new model of probabilistic analysis for the study of road safety based on Bayesian Networks is presented.

Bayesian networks use the probabilistic structure of a multidimensional random variable based on an acyclic graph and a set of conditional probabilities to perform a probabilistic safety analysis. Due to the great advantages that they present with respect to other methods used, such as regression or failure tree methods, they have been used in the last years in many different fields: artificial intelligence, tunneling processes, biomedicine, nuclear plants, and more recently on railway lines. What is intended in this thesis is to perform a safety analysis on roads so this first part focuses on models of Bayesian networks applied to road safety. In this way, models with different purposes are presented: a) predicting the frequency with which accidents of different types occur, b) classifying traffic accidents according to their severity, c) analyzing and preventing accidents, and d) safety.

On the other hand, without the knowledge of certain aspects on graphs would not be possible a correct construction of a model of Bayesian networks. For this reason and thus to be able to clarify concepts and methods used, different explanatory illustrations are presented in which variables used later in the proposed model are utilized.

The second part presents a new model for the study of road safety using a probabilistic analysis based on Bayesian networks. The problem of the safety analysis is an undoubtedly random problem, since practically all the variables that intervene in the same one are random. This requires evaluating probabilities of occurrence of events and frequencies associated with different intervening elements. A first problem arises when representing the dependencies between the variables. Fault trees have some important limitations, including failure to easily represent common causes of failure. The tree structure, i.e. with open branches, does not allow closing them, which would be necessary to reproduce the common causes without replicating the corresponding variables. In contrast, Bayesian networks do not have this limitation and allow their reproduction without the need to replicate these variables. Another notable advantage is that, when using directed graphs, they can be closed and the joint probability of all variables determined by the conditional probabilities of each node given by their parents. In addition, these networks can reproduce any dependency structure without producing incompatibilities, which can occur when the definition of joint probability is made with arbitrary conditional distributions. Another great advantage is that all of the above conditional probabilities can be defined independently. Finally, it should be noted that there are very powerful and even free software packages, which have already been tried and tested, and which allow the implementation of these computer structures without any extra effort. All these advantages are what have led to choosing the Bayesian networks as the optimal model to solve this problem.

The Bayesian network model reproduces not only all the existing elements on the road but the driver's behaviour when he is driving through it. Each element contributes a set of variables according to their type. For example, curves include their radius, length, direction, etc., the signals include their state, the driver's decisions upon seeing them, the associated speeds, the distances between signals, and so on. Due to the great importance of human error in the field of safety, modeling variables associated with driver's behaviour are introduced, such as driver's tiredness and attention; as well as the type of driver or decision of the adopted speed or the presence of a signal. All variables are considered as random and their dependencies are reproduced by the Bayesian network, so that any set of probabilities can be calculated through a process of forward marginalization. The sets of conditional probabilities of the variables, given their parents, are established by means of closed formulas that allow to quantify the Bayesian network. To reduce the complexity of the problem, we propose to use a method that divides the Bayesian network into small parts, such that the complexity of the problem becomes linear in the number of elements. This is crucial to deal with real cases where the number of variables can be measured in thousands.

The probability of incidents related to the different road sections is calculated according to an equivalent number of severe incidents, so that the most critical elements can be identified and ranked in order of importance. This allows to obtain very relevant information to improve the safety and to save time and money in the measures that are necessary to adopt to improve certain roads. In addition, when an accident occurs, the Bayesian network can help identify its causes through a process of inference propagation backwards.

Different examples of Spanish roads, A-67, N-611 and CA-182, are represented to expose the operation of the model that arises. In particular, a detailed study of the regional network in Cantabria, CA-131, CA-132 and CA-142, is carried out by means of this new method.

RESUMEN

Esta tesis consta principalmente de dos partes. La primera de ellas se basa en una revisión de conceptos y modelos que han sido de utilidad para poder llevar a cabo este estudio. La segunda es la que abarca la mayor parte, y en ella se expone un nuevo modelo de análisis probabilista para el estudio de la seguridad de carreteras basado en Redes bayesianas.

Las redes bayesianas emplean la estructura probabilística de una variable aleatoria multidimensional basada en un gráfico acíclico y un conjunto de probabilidades condicionales para realizar el análisis probabilístico de seguridad. Debido a las grandes ventajas que presentan frente a otros métodos empleados, como son los métodos de regresión o los árboles de fallo, se han empezado a utilizar en los últimos años en muy diversos campos: en inteligencia artificial, procesos de excavación de túneles, biomedicina, plantas nucleares, y más recientemente en líneas ferroviarias. Lo que se pretende en esta tesis es realizar un análisis de seguridad en carreteras por lo que esta primera parte se centra en modelos de redes bayesianas aplicados a la seguridad vial. De esta forma, se presentan modelos con diferentes finalidades: a) predecir la frecuencia con la que ocurren accidentes de diferente tipo, b) clasificar accidentes de tráfico en función de su severidad, c) analizar y prevenir accidentes, d) desarrollar evaluaciones de la seguridad, etc.

Por otro lado, sin el conocimiento de ciertos aspectos sobre grafos no sería posible una correcta construcción de un modelo de redes bayesianas. Por este motivo y así poder clarificar conceptos y métodos empleados se presentan diferentes ilustraciones explicativas en las que se emplean variables utilizadas posteriormente en el modelo propuesto.

En la segunda parte se presenta un nuevo modelo para el estudio de la seguridad de carreteras que emplea un análisis probabilístico basado en redes bayesianas. El problema del análisis de seguridad es un problema indudablemente aleatorio, ya que prácticamente todas las variables que intervienen en el mismo lo son. Ello requiere evaluar probabilidades de ocurrencia de sucesos y frecuencias asociadas a los diferentes elementos intervinientes. Un primer problema surge a la hora de representar las dependencias entre las variables. Los árboles de fallos, tienen algunas limitaciones importantes, entre las que destaca la

de no poder representar fácilmente las causas comunes de fallos. La estructura de árbol, es decir con ramas abiertas, no permite cerrarlas, lo que sería necesario para reproducir las causas comunes sin replicar las variables correspondientes. Por el contrario, las redes bayesianas no tienen esta limitación y permiten su reproducción sin necesidad de replicar estas variables. Otra ventaja notable consiste en que, al utilizar grafos dirigidos, pueden cerrarse y definir la probabilidad conjunta de todas las variables mediante las probabilidades condicionales de cada nodo dados sus padres. Además, estas redes permiten reproducir cualquier estructura de dependencia sin producir incompatibilidades, que pueden ocurrir cuando la definición de la probabilidad conjunta se hace con distribuciones condicionadas arbitrarias. Otra gran ventaja es que todas las probabilidades condicionales anteriores pueden definirse independientemente. Finalmente, hay que indicar que existen paquetes de software muy potentes e incluso gratuitos, que ya han sido muy probados y que permiten implementar estas estructuras en ordenador sin grandes esfuerzos adicionales. Todas estas ventajas son las que han conducido a elegir las redes bayesianas como modelo óptimo para resolver este problema.

En el modelo de red bayesiana se reproducen todos los elementos existentes en la carretera y que el conductor se encuentra cuando circula por ella. Cada elemento contribuye con un conjunto de variables en función de su tipo. Por ejemplo, las curvas incluyen su radio, longitud, sentido, etc, las señales incluyen su estado, las decisiones del conductor al verlas, las velocidades asociadas, las distancias entre señales, etc, y así ocurre con cada uno de los elementos o ítems establecidos. Debido a la gran importancia del error humano en el ámbito de la seguridad, se introducen en la modelización variables asociadas al comportamiento de los conductores, como son el cansancio y atención del conductor; así como el tipo de conductor o decisión de la velocidad adoptada o ante la presencia de una señal. Todas las variables son consideradas como aleatorias y sus dependencias son reproducidas por la red bayesiana, de tal manera que cualquier conjunto de probabilidades pueda ser calculado mediante un proceso de marginalización hacia adelante. Los conjuntos de probabilidades condicionales de las variables, dados sus padres, se establecen mediante fórmulas cerradas que permiten cuantificar la red bayesiana. Para reducir la complejidad del problema, se plantea utilizar un método que divide la red bayesiana en pequeñas partes, tales que la complejidad del problema se convierte en lineal en el número de elementos. Esto es crucial para tratar con casos reales en las que el número de variables puede medirse en miles.

La probabilidad de incidentes relacionados con los diferentes tramos de carreteras se calculan en función de un número equivalente de incidentes severos, de manera que se puedan identificar los elementos más críticos y clasificarlos por orden de importancia. Esto permite obtener una información muy relevante para mejorar la seguridad y ahorrar tiempo y dinero en las medidas que son necesarias adoptar para mejorar determinadas carreteras. Además, cuando ocurre un accidente, la red bayesiana puede ayudar a identificar sus causas por medio de un proceso de propagación de inferencia hacia atrás.

Diferentes ejemplos de carreteras españolas, A-67, N-611 y CA-182, son representados para exponer el funcionamiento del modelo que se plantea. En particular, se realiza un detallado estudio de las carreteras de la red autonómica de Cantabria, CA-131, CA-132 y CA-142, por medio de este nuevo método de análisis de seguridad de carreteras.

OBJECTIVES OF THIS THESIS

The main objectives of this thesis are the following:

1. **PART I. INTRODUCTION TO BAYESIAN NETWORKS.**

   Before starting research in a given field, we need to investigate whether or not the research topic has been analyzed before and if the answer is positive, what solutions to the problem being analyzed have been given by other authors.

   The resolution of the problem of probabilistic analysis of road and highway safety requires the use of powerful tools that allow the reproduction of multidimensional random variables. Thus, Bayesian networks have been selected as the most powerful existing tools to work with these variables.

   In this part of the thesis we analyze, on one hand, some existing models that can be found in the existing literature and, on the other hand, some tools that are needed to work with Bayesian networks.

   (a) **State of the art in safety analysis of highways and roads, especially using Bayesian networks.**

      i. **Perform a literature review.** To perform a literature review of Bayesian networks used in probabilistic safety analysis of roads in order to identify the main contributions and the main pending problems to identify lines of present and future research.

      ii. **Identify the most adequate approaches for probabilistic safety analysis (PSA) of roads.** Since different approaches have been used in safety analysis of roads, we aim at identifying the most convenient ones and those who must not be used.

   (b) **Identification of existing problems deserving a research analysis.**

      i. **Discover possible representation deficienciess.** Some existing models have some representational problems. Our aim is to find possible ways of overcoming these deficiencies.

      ii. **Identify the different types of variables used in safety analysis.** Since several Bayesian networks have been already used to solve safety problems in the trasportation field, they constitute a good field for inspiration when deciding which variables should be included in a PSA.

iii. **Definition of conditional probabilities.** The definition of conditional probabilities is difficult and complex. Thus, new methods to define these probabilities are needed. If possible closed formulas should be investigated.

iv. **Human factor.** It is well known that the human factor is the most important factor playing a role in safety. Thus, how human factor is considered in safety analysis must be investigated.

v. **Analyze the required changes in the existing models for probabilistic safety analysis in order to satisfy reasonable safety levels and simplify its use.** This requires some changes in the models used that need to be analyzed and solved.

vi. **Discover the problems caused by an excess of memory and CPU requirements.** Some existing models have serious problems when the size of the problem is of medium or large size. Our aim is to find possible ways of reducing the complexity of the problem without a loss of quality and rigor.

2. **PART II. PROPOSED MODEL FOR PROBABILISTIC SAFETY ANALYSIS OF TRAFFIC NETWORKS.**

   Probabilistic safety assessments are currently used in nuclear power plants and repeated every five years to update them to the possible changes. However, in the road and highway industry they are not common. Since safety is a necessary requirement in engineering works, and the number of yearly casualties in the transportation field is very high, it appears very convenient to incorporate these techniques to it in order to reduce substantially the social damage produced.

   (a) **Bayesian network structure.**

   i. **Identify new variables.** Once a list of variables has been selected based on other models we need to investigate the need of missing and mainly other variables relevant to safety.

   ii. **Identify dependence structures.** The Bayesian network probabilistic structure is based on acyclic graphs. A correct definition of this graph is crucial in a good designed Bayesian network. We aim at providing adequate graphs.

   iii. **Automatic definition of conditional probabilities.** New methods to define these probabilities in an automated form are needed. If possible closed formulas should be investigated.

   (b) **Required changes in the existing models for probabilistic safety analysis.**

i. **Proposal of some methods to discover the causes of incidents.** One of the common problems in safety analysis is the identification of the causes of incidents. This thesis aims to provide some methods in this direction.

ii. **Reduce complexity.** Bayesian network methods increase complexity when the number of variables involved increase, so that the memory and the CPU capabilities can be exhausted. Thus, there is a need for a solution to this problem. In particular the non-linear increase of the complexity with the number of variables involved need to be dealt with. In this thesis, this problem is analyzed.

iii. **Identification of the riskiest elements in the highway.** The safety of a line is the combination of many different elements that add risk to provide the total line risk. Identification of the riskiest elements in the line and sorting them is the basis for an efficient treatment and a plan of corrective actions. We aim at the identification of those elements.

iv. **Identification of the circumstances causing the riskiest incidents.** To improve safety in a highway, it is not enough to identify the riskiest locations, but the set of circumstances leading to severe incidents. If these circumstances are not or wrongly identified the corrective actions will not be effective and a waste of money can be produced. Identification of these circumstances is an important aim of this work.

v. **Human factor integration.** The human factor needs to be not only analyzed but incorporated as one more factor. In addition, the interaction with all other variables must be considered carefully.

ORIGINAL CONTRIBUTIONS OF THIS THESIS

The main contributions of this thesis are:

1. **PART I. INTRODUCTION TO BAYESIAN NETWORKS.**

   In this part of the thesis existing models for safety analysis of roads have been analyzed and discussed.

   (a) **State of the art in safety analysis of highways and roads, especially using Bayesian networks.**

      i. **Literature review.** A literature review of Bayesian networks used in probabilistic safety analysis of roads has been done and its main contributions and some pending problems identified.

      ii. **Identify the most adequate approaches for probabilistic safety analysis (PSA) of roads.** As a consequence of the analysis done, Bayesian networks have been identified and selected as the most powerful tools for performing a PSA of railway lines.

   (b) **Identification of existing problems deserving a research analysis.**

      i. **Discover possible representation deficienciess.** Since event and fault trees present representational problems with common causes, which are very important and frequent in road safety, Bayesian networks have been used.

      ii. **Identify the different types of variables used in safety analysis.** A selection of the most relevant variables playing a role in safety has been done, taken into account what other authors have previously used and based on other considerations.

      iii. **Definition of conditional probabilities.** An important contribution of the thesis is the use of closed formulas to define conditional probabilities.

      iv. **Human factor.** Human factor has been considered, for the first time, in an integrated form with the rest of variables.

      v. **Analyze the required changes in the existing models for probabilistic safety analysis in order to satisfy reasonable safety levels**

and simplify its use. The deficiencies observed in other methods have been resolved in several forms and the process of assigning probabilities has been simplified.

vi. **Discover the problems caused by an excess of memory and CPU requirements.** The partitioning technique proposed in this thesis makes possible to solve the PSA in a reasonable amount of time. In fact only small subnetworks are handled sequentially, which implies a substantial reduction in memory and CPU requirements. The use of a new methodology, which is linear in the number of variables is an important achievement.

2. **PART II. PROPOSED MODEL FOR PROBABILISTIC SAFETY ANALYSIS OF TRAFFIC NETWORKS.**

   We have proposed that probabilistic safety assessments must be used systematically to evaluate the road safety.

   (a) **Bayesian network structure.**

      i. **Identify new variables.** The list of variables needed to investigate road safety has been identified and analyzed.

      ii. **Identify dependence structures.** The set of parents for each variable has been suggested and discussed.

      iii. **Automatic definition of conditional probabilities.** New methods to define these probabilities in an automated form have been developed. To our knowledge, a systematic use of closed formulas is completely new.

   (b) **Required changes in the existing models for probabilistic safety analysis.**

      i. **Proposal of some methods to discover the causes of incidents.** A backward analysis methodology has ben proposed to discover the causes of incidents. It is based on the properties and main representation advantages of Bayesian networks.

      ii. **Reduce complexity.** Thanks to a partitioning technique, the non-linear increase of the complexity with the number of variables involved has been replaced by a linear increase, which implies a very important achievement.

      iii. **Identification of the riskiest elements in the highways or roads.** The proposed method allow us to identify the riskiest locations in the roads and to sort their importance in order to facilitate and optimize the corrective actions.

      iv. **Identification of the circumstances causing the riskiest incidents.** A novel technique to identify the most common causes or circumstances under which the different incidents occur has been developed. It is very relevant from the point of view of the practice.

v. **Human factor integration.** The human factor has been incorporated as one more factor. In addition, the interaction with all other variables has been considered and analyzed.

(c) **Practical application to real roads.**

i. **Validation of the methodology.** The proposed methodology has been applied to real roads, which proves the suitability and convenience of the proposed methods.

List of Figures

List of Tables

Part I

INTRODUCTION TO BAYESIAN NETWORKS

Chapter 1

State of the art on Bayesian networks

Contents

1.1 Introduction

Before starting, it is necessary to clarify that Bayesian methods have little to do with Bayesian network (BN) models. The first group of models works with a prior distribution that accounts for the previous knowledge about the problem being dealt with and, based on sample data, calculates a posterior distribution that subsumes both items of knowledge and is used to perform the probabilistic safety analyses (PSA). In the case of Bayesian networks no priors or posteriors are used, but the probabilistic structure of a multidimensional random variable, based on an acyclic graph and a set of conditional probabilities (one per variable), is built to perform the PSA. Some important examples of the first group of models are: a probabilistic finite element model for updating structural systems by Sun and Betti (2015), a real-time system identification by Yuen and Mu (2015), a Bayesian model for the analysis of external corrosion in underground pipelines by Wang et al. (2015) or the application to natural hazard risk assessment in Graf et al. (2009). Relevant methods applied to traffic problems in this group can be seen in Li et al. (2008), Pawlovich et al. (2006), Hauer et al. (2002), Persaud and Lyon (2007) or Scherb et al. (2015). The second group of models have been used to find solutions to many problems, such as Larrañaga and Moral (2011) in artificial intelligence, Bielza et al. (2011) and Bielza et al. (2015), in classification problems. Other examples are: Bayesian network model to assess wildfire consequences of Papakosta and Straub (2013) or the dynamic Bayesian network model for probabilistic modeling of tunnel excavation processes proposed by Spackova and Straub (2015). Recently, PSA models based on Bayesian networks have also been extended to railway lines (see for example, Castillo et al. (2016b), Castillo et al. (2016a), Castillo et al. (2016c)) and they have been proved to provide not only an important information about safety but a useful quantification of the probabilities of occurrence of undesired incidents.

On the other hand, in order to improve traffic safety, the location of the black spots in highways and roads should be precisely identified (see, for example, Elvik (2007, 2008), so that some corrections could be implemented. Sometimes these corrections are low cost and easy to implement but in other cases they require important investments. Thus, not only the frequency and severity of incidents but their costs must be considered in a serious analysis. A wide collection of works concerning traffic safety can be found in the existing literature, however, this Thesis focuses on those studies based on Bayesian networks (BNs) as the most adequate statistical model, due to its incredible power to reproduce multidimensional random variables (see Castillo et al. (1997)). Bayesian networks permit integrating all the relevant items of the road in the same model. Mahboob (2014) in his Ph. D. thesis demonstrated that classical methods such as fault and event tree analysis are not the most adequate techniques to analyze complex systems involving multi-dependencies between system variables, especially when common causes are present, as it is the case of railways and highways, and Bobbio et al. (2001) compared Bayesian networks (BNs) with fault trees (FTs), demonstrating that any FT can be mapped into a BN and that the corresponding inference techniques of BN can be used to obtain the reliability of any event in the FT. In addition, they expose that some advantages can be obtained at (a) the modeling level because some restrictive assumptions of FTs can be removed and various kinds of dependencies (including uncertainty) among components can be incorporated, and (b) at the analysis level because a general diagnostic analysis, allowing both forward and backward analysis, can be performed.

Because of all of this, the attention is concentrated here on Bayesian network models used in traffic safety analysis and next some of them are described.

1.2 Traffic accident causality analysis in Jilin

Hongguo et al. (2010) analyse the traffic accident causality with Bayesian networks by developing a model consisting of two parts: (a) a structure learning component and (b) a parameter learning component.

To this aim, a training data on traffic accidents occurred in the main expressways in the Jilin province during 2003-2006 is used. The sample consists of 3019 data items. The data types and data items used in the analysis are shown in Table 1.1.

Table 1.1: Data types and data items used in the analysis of the traffic accidents occurred in the main expressways in the Jilin province during 2003-2006.

Data type	**Data items**
Environment factors	Terrain, Weather, Traffic control
Road factors	Road type, Pavement type, Road alignment, Type of intersection and road section, Road cross-section, Road condition
Traffic accident	Cause of accident, Accident form, Accident type, Number of deaths, Number of serious injuries, Number of light injuries, Property damage

The learning process of the Bayesian network has been done using the K2 algorithm, and consists of two stages: the first consists of learning the structure, that is, the directed acyclic graph (DAG) and the second is the parametric structure in which the conditional probability tables are learnt.

The result of the learning process is shown in Figure 1.1, where it can be seen that not all items have been selected as relevant variables (compare the items in the second column of Table 1.1 with the variables in Figure 1.1).

More precisely, the ten nodes in the DAG represent the following ten variables proposed for this model: terrain, road type, road cross-section, cause of accident, accident form, accident type, number of deaths, number of serious injuries, number of light injuries and property damage. The considered sets of values for each of the 10 variables are indicated in Table 1.2.

The Bayesian network structure indicated in Figure 1.1 shows that the conditional probability tables to be learnt are:

$$P(T), P(R_t|T), P(R_{cs}|R_t), P(A_f|R_t), P(A_t|R_t), P(C_a|R_{cs}),$$

$$P(N_d|A_t), P(N_{si}|A_t), P(P_d|A_t), P(N_{li}|A_t).$$

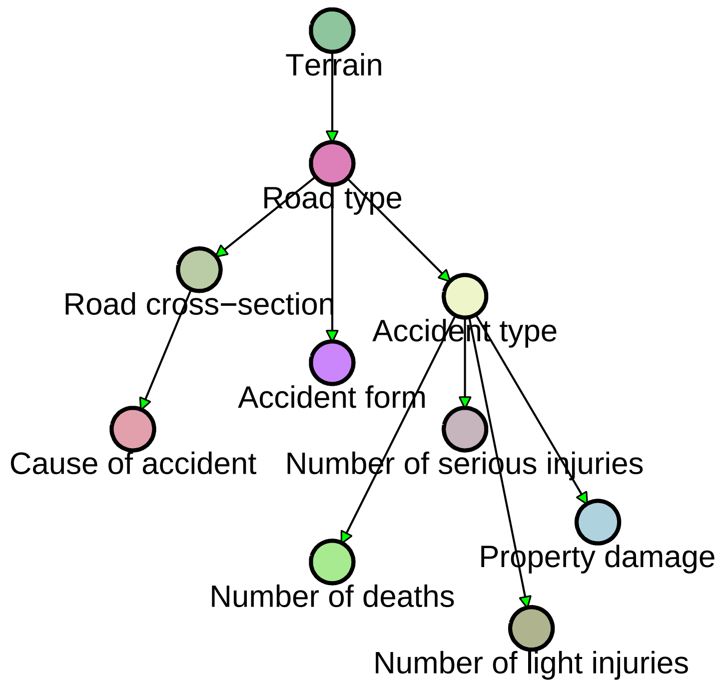


Figure 1.1: DAG resulting from the application of the K2 algorithm to the available data obtained from the main expressways in the Jilin province during 2003-2006.

Some examples of these tables can be seen in Hongguo et al. (2010).

It is noted that this model provides the probabilities of the different accident types for the whole traffic network but it is not applicable to a given traffic route or to evaluate the risk associated with each location.

1.3 Study of the impacts of different factors on accidents in Katowice

Krol (2014) studies the impact of different factors on the frequency distribution of streets road accidents in Katowice, especially in the streets where most traffic events take place.

The team of Decision Systems Laboratory at the University of Pittsburg developes the Genie package to implement a Bayesian network model with additional elements supporting the diagnostics and decision-making. They use the data contained in the original XML files of System of Recording the Accidents and Collisions (SEWIK) provided by the Police Headquarter and containing the data on the road incidents and the EM (Expectation-Maximization) algorithm in order to adjust the model parameters.

The study covers the period of two years, 2011 and 2012, and the streets under inves-

Table 1.2: Sets of values considered for the different variables in the model of Hongguo et al.

Terrain (T)	Plains, Hills, Mountain
Road type (R_t)	Expressway, Class I highway, Class II highway, Class III highway, Class IV highway, Sub-standard road, Urban road
Road cross-section (R_{cs})	Road divided by lanes and directions, Road divided by lanes, Road divided by directions, Lane-direction mixed
Cause of accident (C_a)	Vehicle breakdown, Violation of motor vehicle, Violation of non-motor vehicle, passenger or pedestrian, Unexpected reasons, Others
Accident form (A_f)	Front collision, Side collision, Rear collision, Scraping in the opposite direction, Scraping in the same direction, Rolling over, Rolling, Hitting a stationary object, Hitting a stationary vehicle, Falling, Catching fire, Others
Accident type (A_t)	Injury, Property damage, Fatal
Number of deaths (N_d)	0, 1, 2, 3, 4 or more
Number of serious injuries (N_{si})	0, 1, 2, 3 or more
Number of light injuries (N_{li})	0, 1, 2, 3 or more
Property damage (P_d)	$0,\ 0-1,000,\ 1,000-10,000,\ 10,000-100,000,$ more than $100,000$

tigation are: Rodzieskiego, Chorzowska, Murckowska, Kociuszki, Mikoowska, Korfantego, Grnolska and Bocheskiego,whose characteristics are shown in Table 1.3. The variables considered are the following: time of the year, time of the day, day of the weak, type of road, pavement condition, type of area and geometry of the road, location of the incident, its nature and consequences. In order to represent the relationship between them the Bayesian network structure shown in Figure 1.2 is proposed.

After carrying out the process of learning of the Bayesian network, the study focuses on some changes in distributions of probabilities that are significant. To this end, they set the status of some random variables and then observe the changes in the probability distributions of the states of other variables representing them as bar graphs. In a first stage it is based on results depending on the time of the day, the season and the day of the week. In a second stage, weather conditions (rain, fog, clouds and falling snow) are also included.

Finally the incorporation of the traffic volume is proposed in order to obtain better

Table 1.3: Characteristics of the streets under investigation.

Street	**Incidents**	**Length [m]**	**Characteristics**
Rodzieskiego	852	3400	S86 expressway with additional parallel carriageways
Chorzowska	657	4810	Dual carriageway with crossroads and pedestrian crossings, DK 79
Murckowska	416	2400	Terrain, Dual carriageway, collision free crossroads, 3 lanes, S86
Kociuszki	391	9800	Initially a one-way street in a dense urban area, then dual carriageway with crossroads and pedestrian crossings
Mikoowska	252	2340	Dual carriageway with crossroads and pedestrian crossings
Korfantego	224	3940	Dual carriageway with crossroads and pedestrian crossings
Grnolska	209	13000	A4 motorway
Bocheskiego	144	2000	Dual carriageway with crossroads and pedestrian crossings

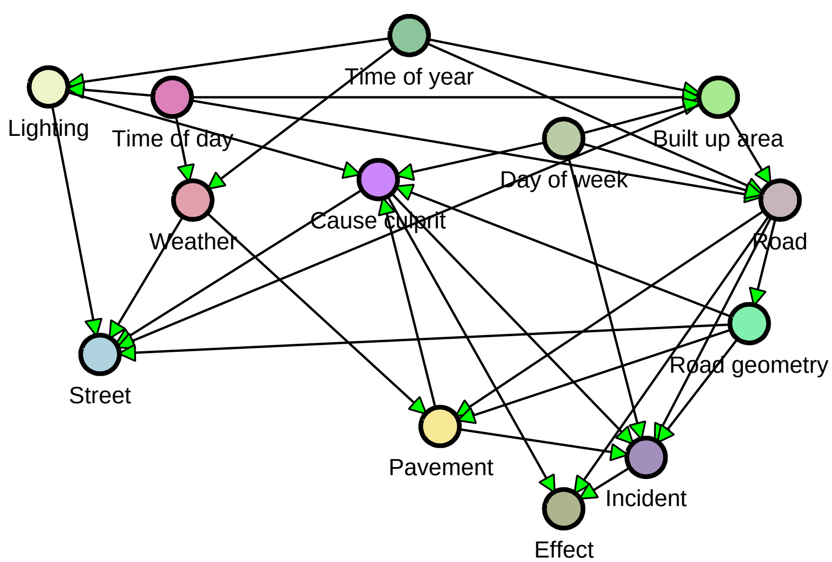


Figure 1.2: Bayesian network structure (Krol).

Table 1.4: % Traffic accidents regarding to number of vehicles and % injuries and fatalities concerning population in the years 1997 to 2006 occurred in Iran.

Year	Population	Vehicles	Traffic accidents per 100 vehicles	% Injuries per 100 inhabitants	% Fatalities per 100 inhabitants
1997	61897000	6494004	2.47	0.11	0.02
1998	62640000	6804939	2.94	0.13	0.02
1999	63392000	7115874	3.45	0.14	0.02
2000	64153000	7426809	3.94	0.17	0.03
2001	64922000	7737744	4.39	0.18	0.03
2002	65701000	8485285	5.28	0.25	0.03
2003	66490000	10364637	5.35	0.33	0.04
2004	67478000	12323989	5.07	0.36	0.04
2005	68467000	14283341	4.61	0.40	0.04
2006	70473000	16242693	3.96	0.39	0.04

results.

1.4 Prediction of vehicle traffic accidents in Iran

Alizadeh et al. (2014) provide a model for analysis and prediction of vehicle traffic accidents using Bayesian networks.

The possible injuries caused by accidents are divided into three categories: minor injuries, serious injuries and deaths.

This study is realised with accidents data from Iran and the variables are selected based on knowledge of the potential causes of accidents, accident reports, as well as variables derived from the literature review. They use for the model the following variables: traffic volume, weather conditions, day of week, road type, age, sex, impact the type of accident and severity of damage.

Table 1.4 shows the the number of traffic accidents and deaths in the years 1997 to 2006 occurred in Iran classified by injuries or deaths and in Figure 1.3 the model DAG is shown.

1.5 Traffic analysis to predict accidents in China

Lin et al. (2011) establish a Bayesian Network traffic analysis to make probability prediction and accident diagnosis. The learning process contains two components: the expert knowledge and the K2 algorithm, first, dividing the nodes into groups of high similarity

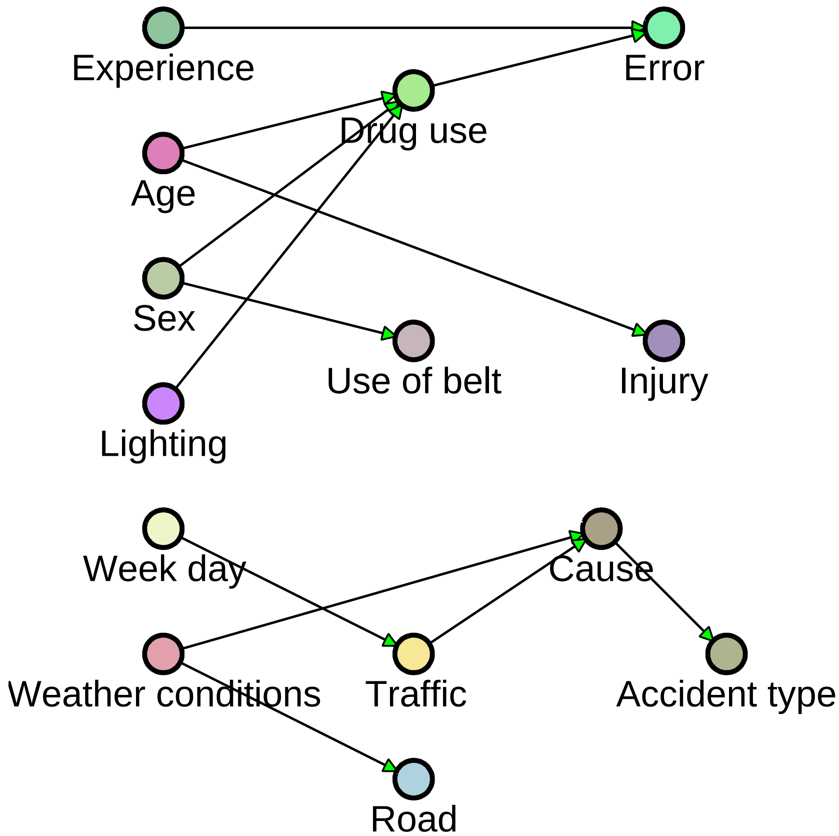


Figure 1.3: DAG for the Alizadeh et al. model.

with each other and then reordering the nodes by the expert experience. The data are obtained from the Road traffic accidents information collection project list of the ministry.

The flow chart of Figure 1.4 serves as explanation of the process of Bayesian networks modeling, which using the list of variables shown in Table 1.5, gives as result the following Bayesian network structure represented in Figure 1.5.

They indicate that the precision of the model could be better joining more factors to value.

1.6 Performing a transportation safety assessment

Zhang and Shi (2015) propose a method for transportation safety assessment of a mountainous freeway using a Bayesian network.

The analysis is based mainly on the driver's state, the road condition and the environmental condition. For the model they select eight observable variables and four latent variables to establish relations with the direct cause and the basic reason of accidents.

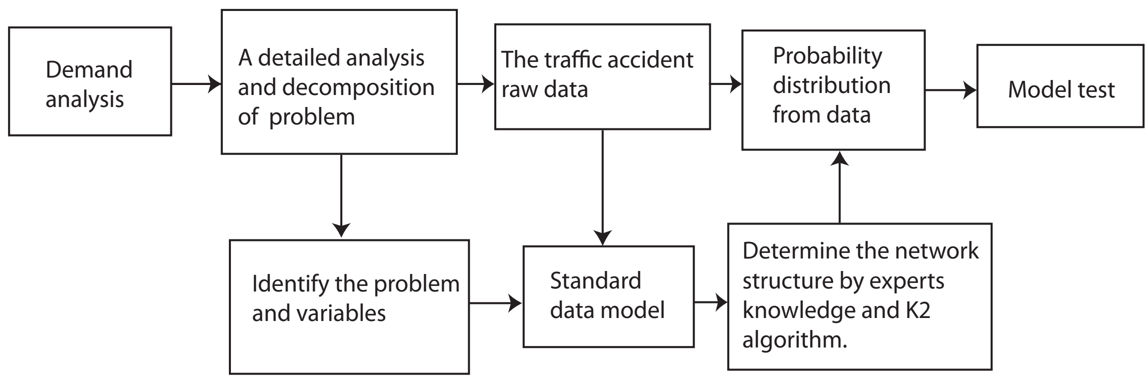


Figure 1.4: Bayesian network structure for the Lin et al. (2011) model.

Table 1.5: Sets of values considered for the different variables in the Lin et al. model.

Vehicle type (t0)	0-Truck, 1-Car
Visibility (t1)	0-Within 50 m, 1-Between 50-100 m, 2-Between 100-200 m, 3- Over 200 m
Lighting condition (t2)	0- Daytime, 1-Night with street lamp, 2-Night without street lamp
Road condition (t3)	0-Flat road, 1-Other
Weather (t4)	0-Fine, 1-Rain, 2-Cloudy
TRaffic condition (t5)	0-Crowded, 1-Uncrowded
Horizontal curve radius (t6)	0-Within 500 m, 1-Between 500-600 m, 2-Over 600 m
Gradient (t7)	0-Between 0-50 % 1-Between 50-100 % 2-Between 100-200 %
Accident type (A)	0-Unhappen, 1-Happen
Pilot tensity (P0)(Hidden)	0-Low, 1-Middle, 2-High
Road alignment reasonable degree (P1)(Hidden)	0-Low, 1-Middle, 2-High

The eight observable variables are: lighting condition (S1), average daily traffic (S2), horizontal curve radius (S3), longitudinal grade (S4), radius of the vertical curve (S5), sight distance (S6), fault casualty number (S7) and weather condition (S8). The four latent variables are the driver's state (B1), road condition (B2), environmental condition (B3)

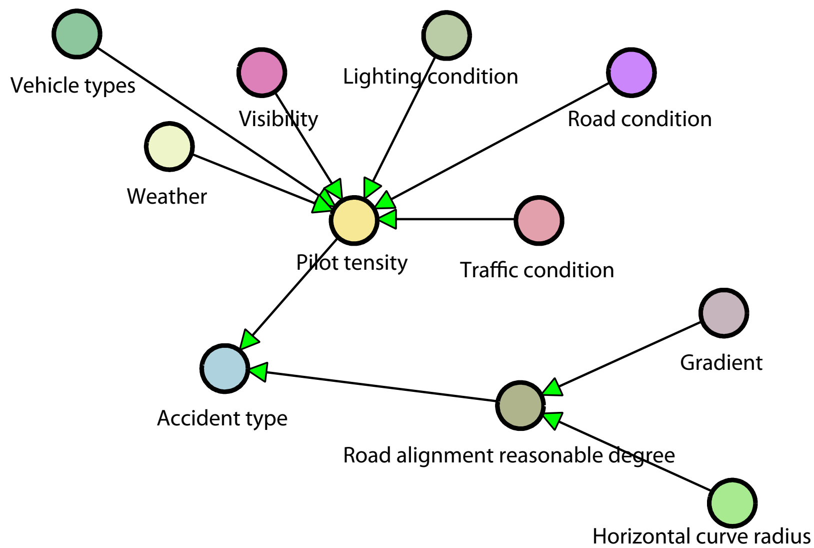


Figure 1.5: Bayesian network modeling for the Lin et al. (2011) model.

and accident type (A).

The qualitative structure and the conditional probabilities are defined by fifteen experts divided in three groups and the structure of the Bayesian network proposed is shown in Figure 1.6. In Figure 1.7 it is represented after removing the expert's opinion.

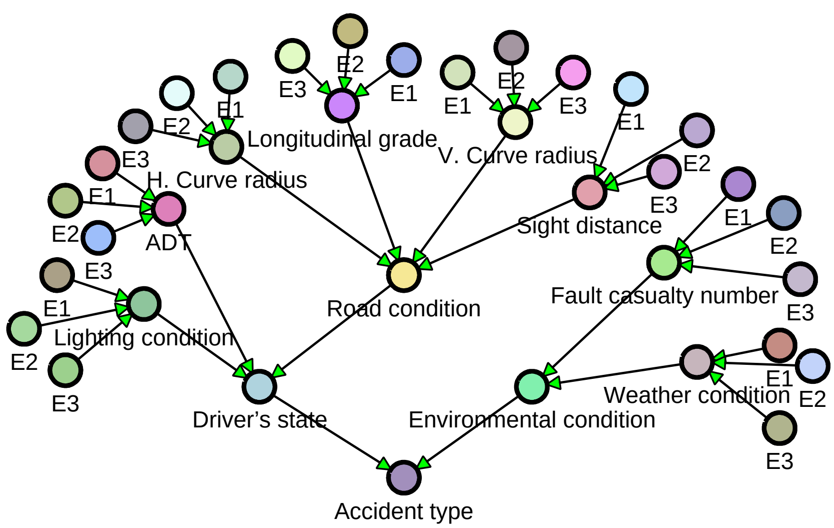


Figure 1.6: Bayesian network structure including expert's opinion.

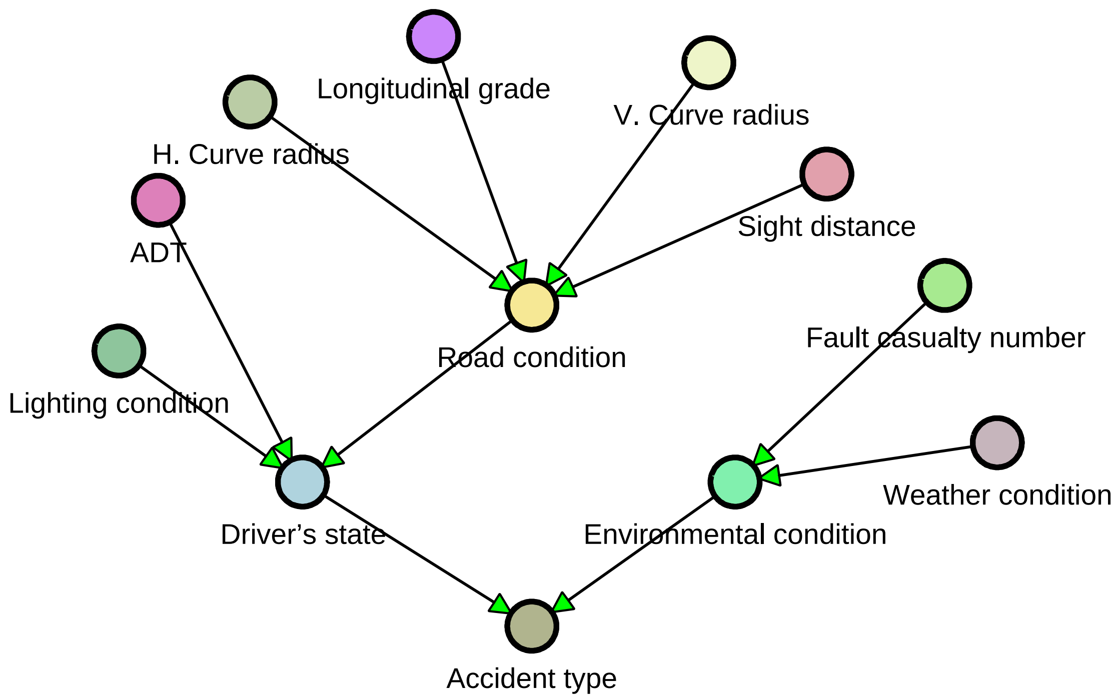


Figure 1.7: Bayesian network structure after removing the expert's opinion.

Table 1.6 shows the interval values taken into account in order to divide the sensitivity extent of the index and the values to estimate $P(Si|Emn)$.

Table 1.6: Interval values from the experts.

Degree of sensitivity significance	$P(Si\|Emn)$ **(High)**	$P(Si\|Emn)$ **(Low)**
Very significant	1.0-0.8 (0.9)	0.2-0.0 (0.1)
Significant	0.8-0.6 (0.7)	0.4-0.2 (0.3)
Potentially significant	0.6-0.4 (0.5)	0.6-0.4 (0.5)
Low significant	0.4-0.2 (0.3)	0.8-0.6 (0.7)
Very low significant	0.2-0.0 (0.1)	1.0-0.8 (0.9)

They do the reasoning for the model and use the following criteria shown in Table 1.7 to evaluate the safety of mountains freeway using the probability P(A=Yes).

1.7 Comparison between BN and Regresion models

Zong et al. (2013) compare Bayesian networks and Regression models for prediction of

Table 1.7: Criteria of the mountains freeway safety.

P(A= Yes)	Safety grade	P(A= Yes)	Safety grade
<0.1	1	0.1-0.3	2
0.3-0.5	3	>0.5	4

traffic accident severities. To this end, three indicators, number of fatalities, number of injuries and property damage, are investigated respectively with both techniques.

The data set are obtained from police-reported traffic accident records for the Jilin province, China, in 2010. It consists of 2,246 cases and contains information regarding to accident severity, accident characteristics, vehicle characteristics, environmental factors and road conditions. With these data, they analyse 17 variables as indicated in Table 1.8.

Table 1.8: Factors and variables analysed in the study of the traffic accident severities in the Jilin province during 2010.

Factors	**Variables**
Accident severity	Number of fatalities, number of injuries, Property damage
Accident characteristics	Time of day, Location-Motor vehicle lanes, Location-Crosswalk, Location-Regular road section, Location-Intersection
Vehicle characteristics	Motorcycle involved, Bus or truck involved, Vehicle condition
Environmental factors	Weather condition, Visibility distance
Roadway characteristics	Pavement condition, Roadway surface condition, Road geometrics, Traffic signal control

In order to get the structure of the severity prediction Bayesian network, the K2 algorithm is used. In the case of the study with Regression models, Ordered probit model is employed in forecasting of number of injuries and property damage while for number of fatalities a Binary Logit model is used.

In the case of the study with Regression models, Ordered probit model is employed in forecasting of number of injuries and property damage while for number of fatalities a Binary Logit model is used.

By comparing the two models, using Mean Absolute Percentage Error (MAPE) and Hit ratio, they conclude that the Bayesian network models are better in accident severity prediction than Regression models regarding modeling accuracy and they point out that there are also differences regarding the interactions between the variables in the models.

1.8 Classification of traffic accidents by severity

de Oña et al. (2011) study the validity of Bayesian networks to classify traffic accidents according to their injury severity, by building three Bayesian networks using three different score metrics based on the hill climbing search algorithm. Accident data are obtained from the Spanish General Traffic Directorate (DGT) for rural highways in Granada, Spain, for three years (from 2003 to 2005) and a sample of 1,536 accident real data set is used. Variables describing conditions, such as injury severity, roadway information, weather, accident, and drivers characteristics, are included. More precisely, the eighteen following variables are analyzed: accident type, age, atmospheric factors, cause, day, gender, lane width, lighting, month, number of injuries, occupants involved, paved shoulder, pavement width, pavement markings, shoulder type, sight distance, time and vehicles involved. In order to evaluate the performance of the three Bayesian networks several indicators are employed: accuracy, sensitivity, specificity, the Harmonic Mean of Sensitivity and Specificity (HMSS), the Receiver Operating Characteristic Curve (ROC) Area, the Most Probable Explanation (MPE) and the complexity (or total number of arcs). After that, inference is used to identify the values of the variables that are associated with killed or severely injured in traffic accidents on Spanish rural highways.

1.9 Accident prediction on Swiss highways

Deublein et al. (2015) use a Bayesian network model to predict accident on Swiss highways, divided in two groups of road types, open roads (including bridges) and tunnels (including galleries). The model predicts the number of accidents with light personal injury, severe and fatal injury accidents on a given highway segment and can be used to identify segments with a high expected number of accidents. All data are provided by the Federal Road Office (FEDRO) from 3 years (2010, 2011 and 2012), and all infrastructure and transportation information to be used to divide the network into segments as well as all accident data are associated to georeferenced points.

The methodology employed is the proposed by Deublein (2013) which is composed by 5 steps: (a) define homogeneous segments in which the values of the considered variables remain constant, (b) calculate posterior accident rates, (c) perform multivariate regression analysis to include not only different risk indicators (shown in Table 1.9) into the regression equation, but also several target variables (the accident rates of accidents with no more than light injuries (LINJ), accidents with no more than heavy personal injuries, (SINJ), and accidents with fatalities (FAT)), (d) develop the Bayesian network, where the qualitative structure (represented in Figure 1.8) is fixed and identical for all accident severities, and the conditional probability tables are built using a regression model, and (e) determine the expected number of accidents with the posterior Bayesian network to estimate frequencies and with the exposure of the segments to estimate number of accidents per year.

Table 1.9: Risk-indicator variables in the model proposed by Deublein et al.

Variables	Description
AADT	Annual average daily traffic based on the 2010 traffic model (ARE, 2010)
HGV	Percentage of heavy traffic (heavy good vehicles) with respect to the AADT based on the 2010 traffic model (ARE, 2010)
RAD	Middle curve radius based on authors calculations
GRAD+	Amount of the average slope (up) based on the authors
GRAD-	Amount of the average slope (down) based on the authors
SPEED	Signalled maximum speed
LANES	Number of lanes in each direction
EVEN	Road surface texture: evenness in the longitudinal direction, according to VSS standard SN 640 925b (VSS, 2003). In this case, a mark between 0 and 2 is good, between 2 and 4 is sufficient, and 5 is poor surface condition.
ROUGH	Road surface texture: grip, according to VSS standard SN 640 925b (VSS, 2003). In this case, a mark between 0 and 2 is good, between 2 and 4 is sufficient, and 5 is poor surface condition.
TYPE	Distinction between open road including bridges, and tunnel including galleries

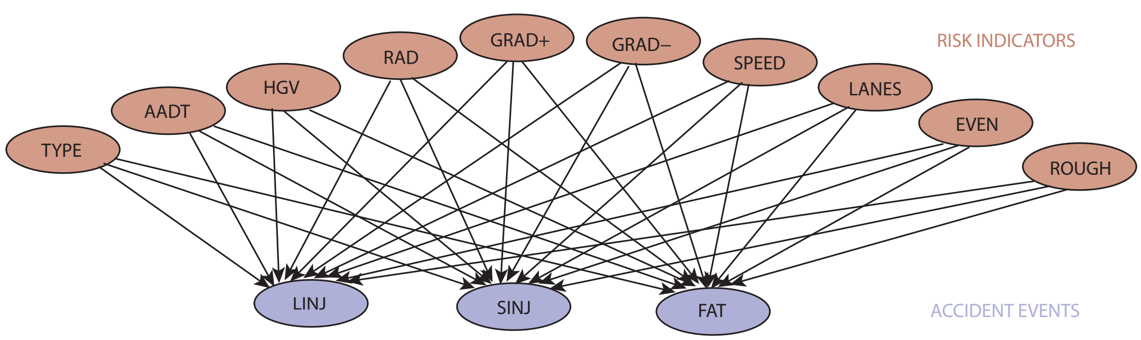


Figure 1.8: Bayesian network Deublein et al. model structure.

1.10 Safety analysis of expressways in China

Wang et al. (2014) develop a Bayesian network model for safety analysis for expressways in China. 949 accidents occurred on a section of freeway in mountain areas from 2009 to 2012 are analyzed and eight variables included within four influence factors: (a) driver characteristics, (b) highway characteristics, (c) vehicle characteristics, and (d) atmospheric characteristics, are used to build the Bayesian network model using the Netica software. Table 1.10 shows each of the variables with their possible values and the structure of the Bayesian network is represented in Figure 1.9.

Table 1.10: Variables with their values proposed by Wang et al. model.

Variables	Values
Accident type	Fatality, injury and property
Lighting	Dusk and daylight
Weather	Bright, cloudy, fog, rainy, snow, and other
Age	(18-24, 25-64, 65-inf)
Experience	(0-1, 2-5, 6-10, 11-inf)
Sex	Male and female
Cause	Fatigue driving, not according to stipulations, unsuitable safety distance, overspeed, wrong overtaking, improper operation, and other
Vehicle type	Large cars and small cars

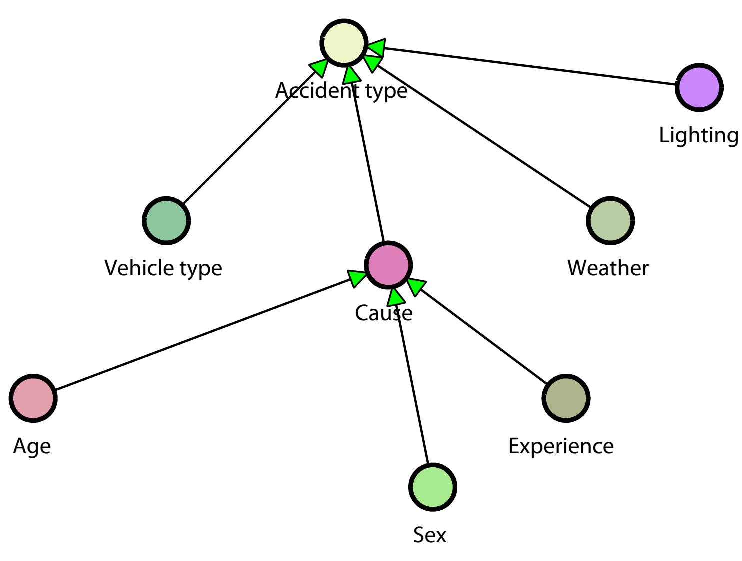


Figure 1.9: Bayesian network Wang et al. model structure.

In this study three types of accidents are considered, fatality, injury, and property, and a sensivity analysis is performed to determine the main cause of the occurrence of accidents on expressway in China, resulting the four main causes: cause, experience, weather and lighting.

1.11 Modeling highway traffic safety in Nigeria

Mbakwe (2011) in his Ph. D. thesis presents a model for highway traffic safety in Nigeria using BNs. In order to combat road safety problems it is intended to identify the major contributing factors, including poor road condition, road obstruction, anemic use of traffic control devices, driving under the influence, aggressive/reckless driving, mechanical failure, and driver fatigue. Since Nigeria does not have a reliable and comprehensive database of traffic incidents and fatalities, the Delphi technique is employed to generate the required data. An important conclusion derived from this work is that the Nigerian traffic safety would improve significantly if the existing laws and policies are enforced, even at a very moderate level.

1.12 Predictive accident modeling in Canada

Chen (2014) classifies the main causes of highway incidents into: external environment, operational environment, driver, and vehicle conditions and proposes a Bayesian Networks model to predict and continuously update the likelihood of highway incidents, by considering a set of 28 variables or risk factors from the four principal causes. Incidents data for the development of this accident predictive model are provided from Transport Canada's National Collision Database (NCDB), in particular the information of 293 highway accidents with fatalities is collected, in a period of 11 years (from 1999 to 2010).

Furthermore, they integrate the model with a Safety Instrumented System (SIS) and simulate 10 scenarios with different specific states of variables to predict the probability of fatal accident occurrence, in order to demonstrate how the integration of the BNs model with SIS is.

Chapter 2

Some concepts on graphs

Contents

2.1 Introduction

In this Chapter basic concepts needed to build the Bayesian networks that comprise the model proposed in this Thesis are shown. For a better understanding of the procedure to follow, several definitions are given, as well as the description of different algorithms that may interest us at some time. All examples shown are applied to the variables (described and their relations established in Chapter 3 and 4) from which our initial Bayesian network starts.

2.2 Triangulated graphs

Definition 1 (Chord of a loop.) *A chord is a link between two nodes in a loop that is not contained in the loop.*

In Figure 2.1 several chords in different loops are represented. Thus, the links $Vis - S$, $W - D$, and $Vt - D$, are chords in the loop $Vis - W - Vt - S - D - Vis$. A chord breaks the loop and decomposes it into smaller loops, $Vis - W - Vt - S - Vis$ and

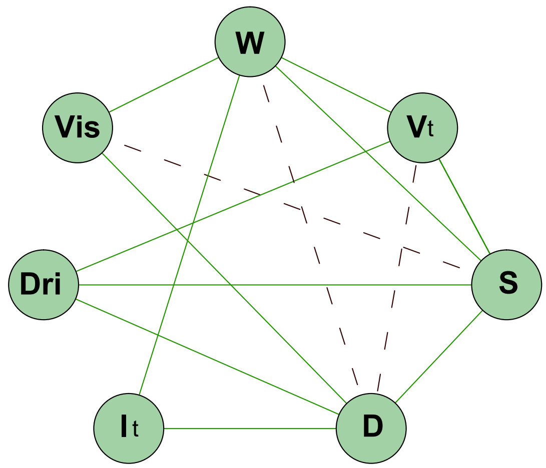


Figure 2.1: Example of several chords in different loops.

$Vis-S-D-Vis$ for example in the case of the $Vis-S$ chord, $W-Vt-S-D-W$ and $W-D-Vis-W$ for the $W-D$ chord, and, $Vt-S-D-Vt$ and $Vt-D-Vis-W-Vt$ for the $Vt-D$ chord. On the other hand, it can be seen that these links are also chords in other loops such as $W-D$ link which is a chord in the loop composed by the nodes $W-S-D-It-W$ and $Vt-D$ in the loop $Vt-S-D-Dri-Vt$.

In the same way, Figure 2.2 illustrates other representation of chords in loops. $Vis-Vt$ is a chord in the loop $Vis-W-Vt-S-D-Vis$ and, $W-D$ and $Vis-S$, in the loop $Vis-W-S-D-Vis$ for example.

Given its structure, loops of length 3 are the only loops that can have no chords. Therefore, these are the smallest elements in which a loop can be decomposed by the incorporation of chords in the graph. Loops of length 3 are called triangles.

Definition 2 Triangulated graph. *An undirected graph is said to be triangulated, or chordal, if every loop of length four or more has at least one chord.*

If a graph is not triangulated, it is possible to convert it into triangles by adding chords that divide the loops. Since a loop can be broken in different ways with a chord, triangulation is not unique, as shown in the example below, in which, the two graphs shown in Figure 2.3 correspond to two different triangulations where all loops of length four or more have at least one chord.

It is important to note that triangulating a graph does not consist in dividing it into triangles. For example, the two graphs of Figure 2.3 are triangulated and, therefore, do not need the addition of extra edges, such as the links $Vis-It$ and $Vt-D$.

Figure 2.4 shows a triangulated graph on the right hand and the same graph but not triangulated on the left hand side because a loop of length four, $Vis-W-S-D-Vis$, does not have a chord.

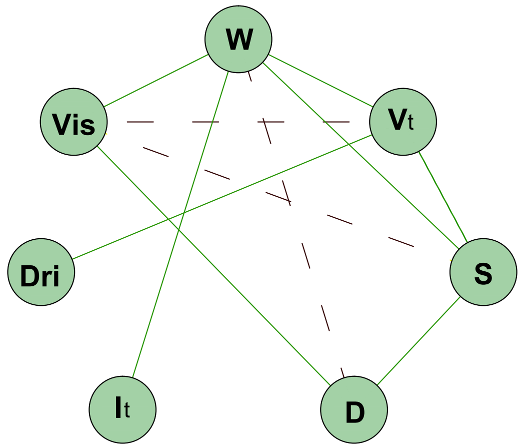


Figure 2.2: Illustration of different chords in loops.

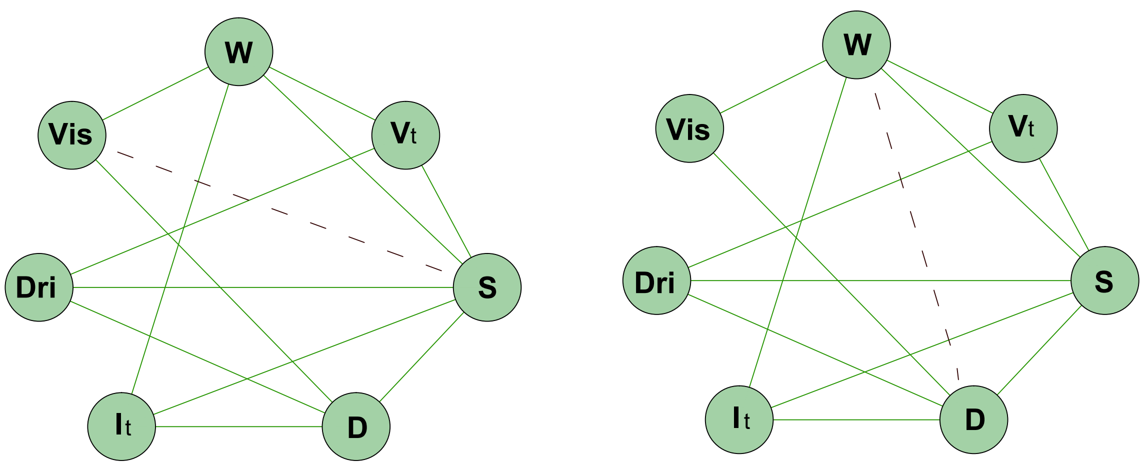


Figure 2.3: A triangulated graph in two different ways where all loops of length four or more have at least one chord.

In order to preserve as much as possible the original topology of the graph in the triangulation process, it is important to add as few chords as possible. In this sense, a triangulation is said to be minimal if it contains a minimum number of chords below which to triangulate the original graph is not possible. For this purpose, a simple algorithm called *Maximum cardinality search algorithm* is introduced, defining before some necessary concepts.

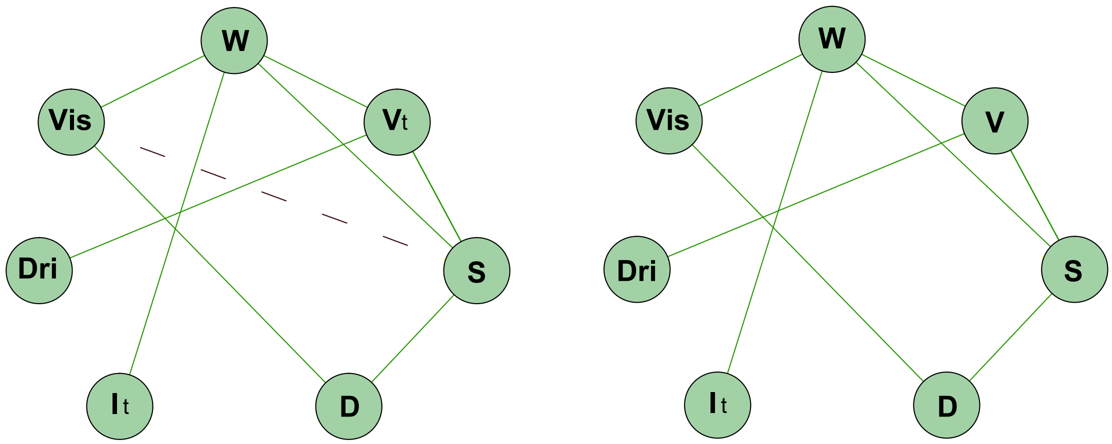


Figure 2.4: A triangulated graph on the right hand side and not triangulated on the left.

2.3 Perfect numbering

Definition 3 Perfect numbering *A given numbering, α, of the nodes of a graph is called a perfect numbering if the subset of nodes $Bnd(\alpha(i)) \cap \{\alpha(1), \ldots, \alpha(i-1)\}$ is complete*[1] *for $i = 2, \ldots, n$.*

Figure 2.5 shows a perfect numbering since it verifies the following conditions:

Being the nodes of the graph: $\alpha(1) = Vis$, $\alpha(2) = W$, $\alpha(3) = D$, $\alpha(4) = S$, $\alpha(5) = It$, $\alpha(6) = Vt$, and $\alpha(7) = Dri$

- For $i = 2$, $Bnd(\alpha(2)) \cap \{\alpha(1)\} = Vnd(W) \cap \{Vis\} = \{Vis, It, D, S, Vt\} \cap \{W\} = \{W\}$, which is a trivially complete set.

- For $i = 3$, $Bnd(\alpha(3)) \cap \{\alpha(1), \alpha(2)\} = \{It, Dri, Vis, W, S\} \cap \{Vis, W\} = \{Vis, W\}$ is complete, since the link $Vis - W$ is contained in the graph.

- For $i = 4$, $Bnd(\alpha(4)) \cap \{\alpha(1), \alpha(2), \alpha(3)\} = \{D, It, Dri, W, Vt\} \cap \{Vis, W, D\} = \{W, D\}$ is also complete.

  Similarly, for $i = 5, \ldots, 7$, it can be verified that
  $$Bnd(\alpha(i)) \cap \{\alpha(1), \ldots, \alpha(i-1)\}$$
  is complete and thus, α is a perfect numbering.

The perfect numbering is not unique. For example, Figure 2.6 shows another perfect numbering for the same previous graph and Figure 2.7 shows other graph with two different perfect numbering. On the other hand, there are also graphs that do not admit any perfect numbering.

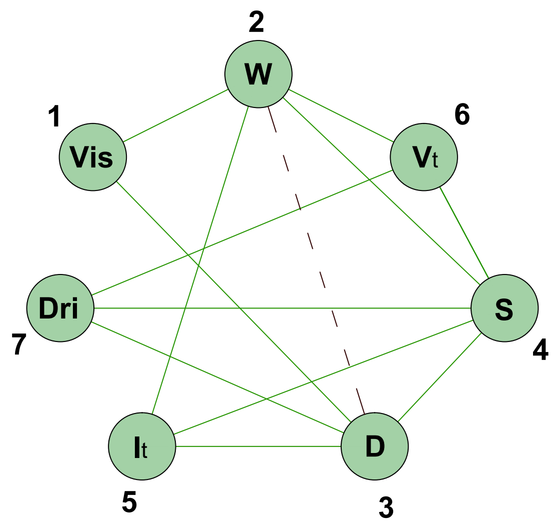


Figure 2.5: Example of a graph with a perfect numbering.

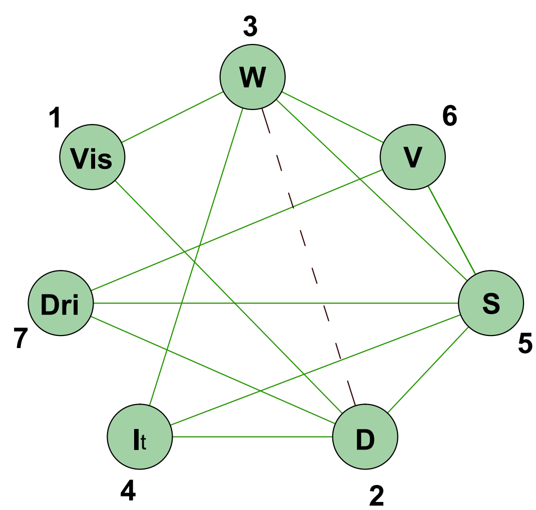


Figure 2.6: Example of a graph with a different perfect numbering.

Theorem 1 Triangulation and perfect numbering. *An undirected graph admits a perfect numbering if and only if it is triangulated.*

The *Maximum Cardinality Search algorithm* gives a numbering for the nodes of an undirected graph which is a perfect numbering only if the original graph is a triangulated graph.

[1]A subset of nodes is called complete if there is a link between every pair of nodes.

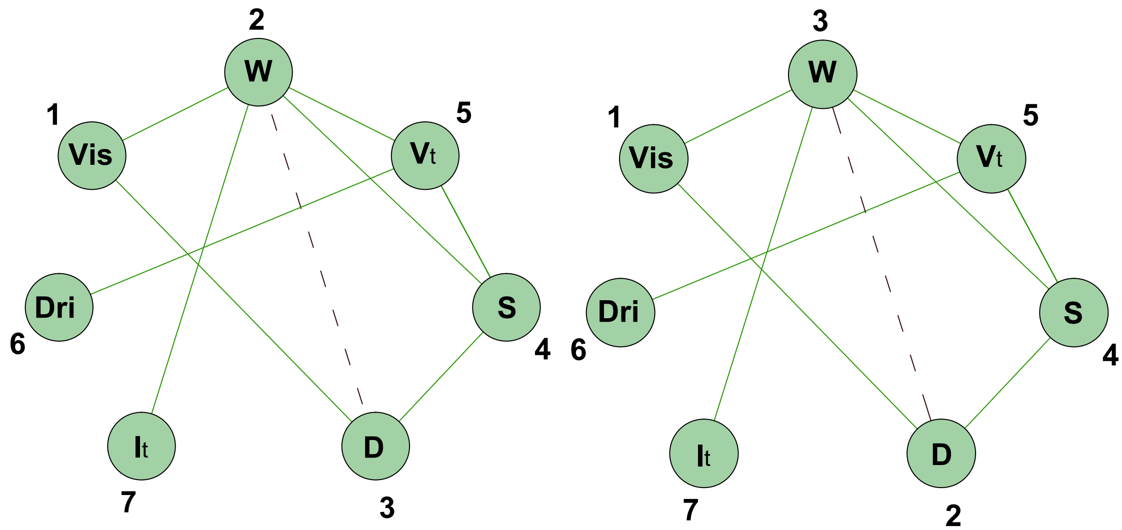


Figure 2.7: Graph with two different perfect numberings.

Algorithm 1 Maximum Cardinality Search Algorithm (MCS).

- **Input:** *An undirected graph $G = (X, L)$ and an initial node X_i.*
- **Output:** *A numbering α of the nodes in X.*

1. *Initialization: Assign the first number in the numbering to the initial node X_i, that is, $\alpha(1) = X_i$.*
2. *Repeat the iteration step for $i = 2, \ldots, n$.*
3. *Iteration i: In the ith iteration an unnumbered node X_k with the maximum number of numbered neighbors is chosen to be numbered, that is, $\alpha(i) = X_k$. Ties are broken arbitrarily.*

Theorem 2 MCS numbering. *Every numbering of the nodes of a triangulated graph obtained by the MCS algorithm is a perfect numbering.*

Algorithm 2 Maximum Cardinality Search Algorithm (MCS).

- **Input:** *An undirected graph $G = (X, L)$ and an initial node X_i.*
- **Output:** *A fill-in L', such that, $G' = (X, L \cup L')$ is a triangulated graph.*

Initialization Steps:

1. *Initially, the fill-in is empty, that is, $L' = \phi$.*

2. *Let* $i = 1$ *and assign the first number in the numbering to the initial node* X_i*, that is,* $\alpha(1) = X_i$.

Iteration Steps:

3. *An unnumbered node* X_k *with a maximum number of numbered neighbors is assigned label* i*,* $\alpha(i) = X_k$.

4. *If* $Nbr(X_k) \cap \{\alpha(1), \ldots, \alpha(i-1)\}$ *is not complete, add to* L' *the necessary links to make this set complete and go to Step 2; otherwise, go to Step 5.*

5. *If* $i = n$*, then stop; otherwise, let* $i = i + 1$ *and go to Step 3.*

Example 1 *Example of Maximum Cardinality Search Fill-In.*

The algorithm is applied to the graph in Figure 2.8 which is undirected and not triangulated. The node Vis *is chosen as the initial node.*

The steps followed are described below and schematically represented in Figure 2.9.

- *Step 1:* $L' = \phi$.
- *Step 2: Let* $i = 1$ *and* $\alpha(1) = Vis$.
- *Step 3: Nodes* D *and* W *have one and all other have no labeled neighbors. The node* D *is chosen and is labeled 2 (*$\alpha(2) = D$*).*
- *Step 4: The previous labeled neighbors form a complete set. Therefore, it is not necessary to include any link in* L'.
- *Step 5: Since* $i \neq n$*, it is increased by one and Step 3 starts again.*
- *Steps* $3-5$*: Nodes* W*,* It *and* S *have the same number of labeled neighbors (one) and the others nodes do not have. The tie is undone by choosing the node* W *as* $\alpha(3)$*.* $Vnd(\alpha(3)) \cap \{\alpha(1), \alpha(2)\} = \{W\} \cap \{Vis, D\}$ *is not complete. Thus, it is necessary to add the link* $W - D$ *to L'.*
- *Steps* $3-5$*: The nodes with maximum number of neighbors now are* It *and* S*.* It *is chosen as* $\alpha(4)$*. The previously numbered neighbors in this case form a complete set, so to add any link to* L' *is not necessary and the algorithm continues.*
- *Steps* $3-5$*: Following a similar process, the nodes* S*,* V *and* Dri *are labeled 5, 6 and 7, respectively.*

The resulting graph $G' = (X, L \cup L')$ *is now a triangulated graph, and the final numbering is a perfect numbering.*

Depending on the choice of the initial node and how the ties are undone, it is possible to obtain several triangulations of the same graph (see Fig 2.3).

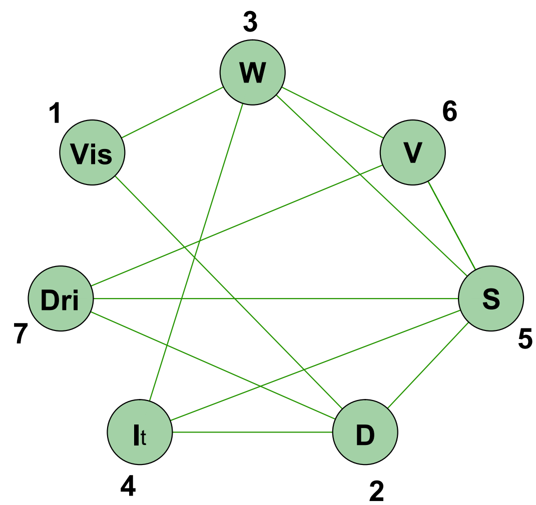


Figure 2.8: Undirected and non-triangulated graph.

2.4 Running intersection property and Chain of cliques

Definition 4 Running intersection property. *An ordering of the cliques of an undirected graph* $(C_1, \ldots, C_m)$ *is said to satisfy the running intersection property if the set* $C_i \cap (C_1 \cup \ldots \cup C_{i-1})$ *is contained in at least one of the cliques* $\{C_1, \ldots, C_{i-1}\}$*, for all* $i = 1, \ldots, m$.

This property states that the cliques of a graph can be ordered in such a way that the nodes that are common to a given clique and the union of the preceding cliques are also contained in at least one of the preceding cliques. An ordered sequence of cliques satisfying the running intersection property is referred to as a *chain of cliques*. Some undirected graphs have no chain of cliques, yet other undirected graphs have more than one chain of cliques. The following theorem characterizes the graphs with at least one chain of cliques.

Theorem 3 Chain of cliques. *An undirected graph has an associated chain of cliques if and only if it is triangulated.*

Algorithm 3 Generating a Chain of Cliques

- **Input:** *A triangulated undirected graph* $G = (X, L)$.
- **Output:** *A chain of cliques* $(C_1, \ldots, C_m)$ *associated with* G.

1. *Initialization: Choose any node to serve as an initial node, then use Algorithm 1 to obtain a perfect numbering of the nodes,* $X_1, \ldots, X_n$.

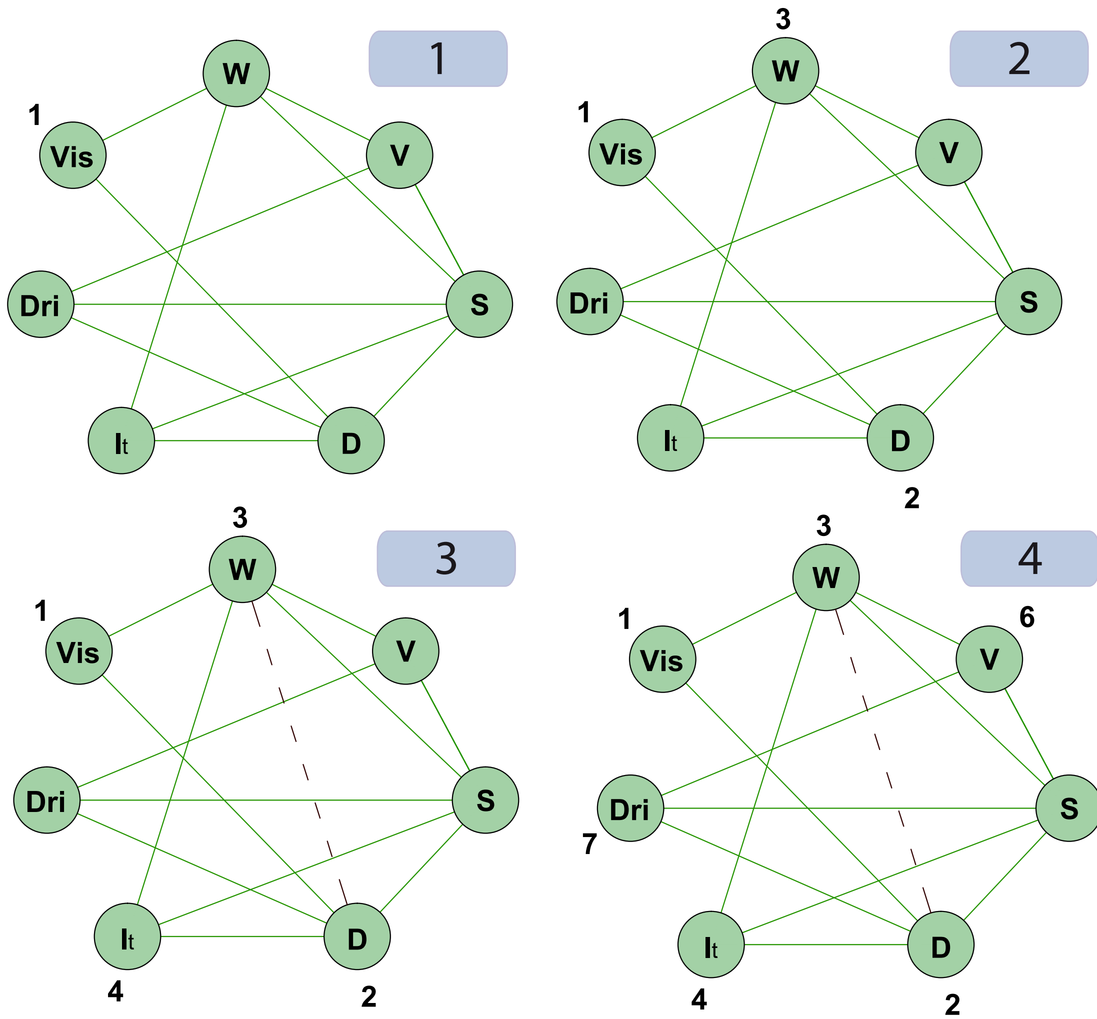


Figure 2.9: Schematic representation of the steps followed for a perfect numbering according to the Algorithm for Maximum Cardinality Search Fill-In.

2. *Identify the cliques of the graph, C.*

3. *Assign to each clique the largest perfect number of its nodes.*

4. *Order the cliques, $(C_1, \ldots, C_m)$, in ascending order according to their assigned numbers (break ties arbitrarily).*

Example 2 *Example of chain of cliques.*

The algorithm 3 is applied to generate a chain of cliques associated with the triangulated graph of perfect numbering given in Figure 2.5. The node $V$is has been considered as the initial node and in Figure 2.10 the cliques of the graph are represented. They are:

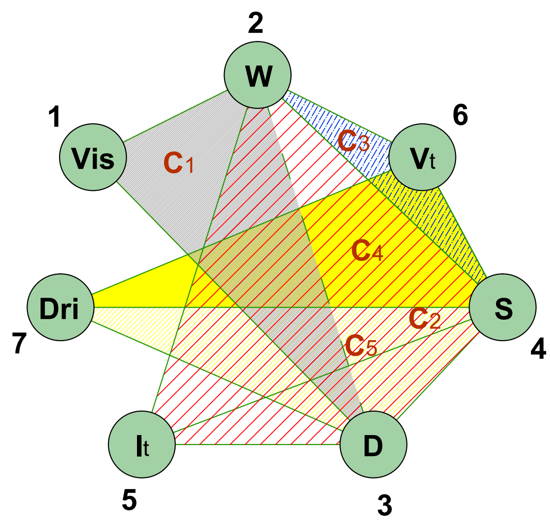


Figure 2.10: Example of a chain of cliques.

$C_1 = \{Vis, W, D\}$, $C_2 = \{W, D, S, It\}$, $C_3 = \{W, S, Vt\}$, $C_4 = \{S, Vt, Dri\}$, *and* $C_5 = \{D, S, Dri\}$.

Next, the largest perfect number of its nodes is assigned to each clique. In this case the number corresponding to the clique C_1 *is* 3, *for* C_2 *the number is 5, 6 for the clique* C_3, *and 7 for the cliques* C_4 *and* C_5. *From clique 1 to 3 they are already ordered in ascending order according to their assigned numbers and for the two last cliques it is possible to choose it. Therefore,* $(C_1, \ldots, C_5)$ *is a chain of cliques for this graph.*

Example 3 *Example of chain of cliques.*

In the same way as in the previous example it is proceeded with the triangulated graph which perfect numbering is shown on the left side of Figure 2.7. Here the cliques of the graph (see Figure 2.11) are: $C_1 = \{Vis, W, D\}$, $C_2 = \{W, D, S\}$, $C_3 = \{W, S, Vt\}$, $C_4 = \{Vt, Dri\}$, *and* $C_5 = \{W, It\}$.

The number assigned to the clique C_1 *is 3, being 4, 5, 6 and 7 for the cliques* C_2, C_3, C_4, *and* C_5, *respectively. Since all the cliques are ordered in ascending order according to their assigned numbers, the subset of cliques* $(C_1, \ldots, C_5)$ *is a chain of cliques for this graph.*

2.5 Cluster graphs

Definition 5 Cluster. *A set of nodes of a graph is called a cluster.*

Definition 6 Cluster graph associated with a graph. *Given a graph* $G = (X, L)$ *and a set of clusters of* X, $C = \{C_1, \ldots, C_m\}$, *such that* $X = C_1 \cup \ldots \cup C_m$, *then the*

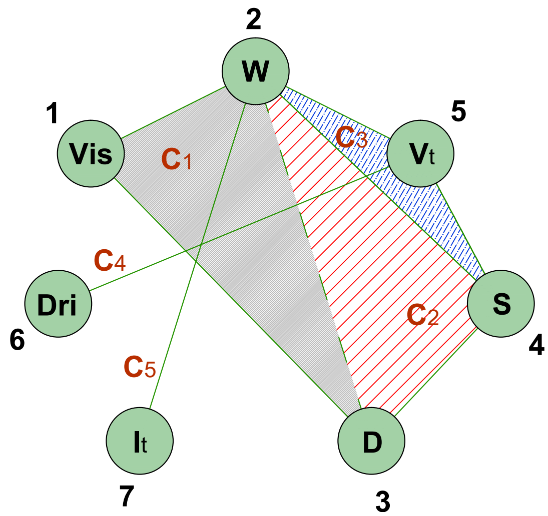


Figure 2.11: Example of a chain of cliques.

graph $G' = (C, L')$ is called a cluster graph of G if L' contains only links between clusters containing common nodes, that is, $(C_i, C_j) \in L' \;\Rightarrow\; C_i \cap C_j \neq \phi$.

Definition 7 Clique graph. *A cluster graph is called a clique graph if its clusters are the cliques of the associated graph.*

Definition 8 Join, or junction, graph. *A clique graph associated with an undirected graph is called a join or junction graph if it contains all the possible links joining two cliques with a common node.*

Definition 9 Join or junction tree. *A clique graph is called a join or junction tree if it is a tree and if every node that belongs to two clusters also belongs to every cluster in the path between them.*

Theorem 4 Join tree. *An undirected graph has a join tree if and only if it is triangulated.*

Algorithm 4 Generating a Join Tree

- **INPUT:** *A triangulated undirected graph $G = (X, L)$.*
- **OUTPUT:** *A join tree $G' = (C, L')$ associated with G.*

1. *Initialization:* Obtain a chain of cliques $(C_1, \ldots, C_m)$ of graph G, using the algorithm 3.

2. For each clique $C_i \in C$, choose from $\{C_1, \ldots, C_{i-1}\}$ a clique C_k with maximum number of common nodes and add the link $C_i - C_k$ to L' (initially empty). Break ties arbitrarily.

Definition 10 Family tree. *A family tree of a directed graph D is a join tree of some undirected graph G associated with D in which the family of every node is contained in at least one cluster.*

Algorithm 5 Generating a Family Tree

- **Input:** *A directed graph $D = (X, L)$.*
- **Output:** *A family tree $G' = (C, L')$ associated with D.*

1. *Moralize the directed graph.*
2. *Triangulate the resulting undirected graph using the triangulation algorithm.*
3. *Apply the join tree algorithm to obtain a join tree of the resulting graph.*

Example 4 *Example of a family tree.*

In order to generate a family tree applying the algorithm 5, the directed graph of Figure 2.12 is considered. The families of its nodes are: $\{W\}$, $\{W, Vis\}$, $\{W, It\}$, $\{Vt, Dri\}$, $\{Vis, D\}$, $\{W, Vt\}$, and $\{W, Vt, D, S\}$.

To moralize the graph, the parents of S, in this case W, V and D, must be joined by a link. W and D are already linked in the initial figure, in such a way that only the $W - D$ edge should be added (see Figure 2.13).

The resulting moral graph is already triangulated. Otherwise, the algorithm 2 would be used for its triangulation. Figure 2.14 shows the chain of cliques associated to the initial directed graph with its corresponding perfect numbering.

Finally, the family tree can be obtained by applying the algorithm 4 and it is shown in Figure 2.15.

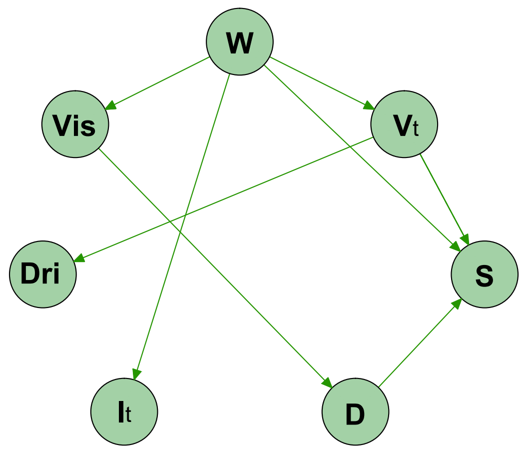


Figure 2.12: Initial directed graph to generate a family tree.

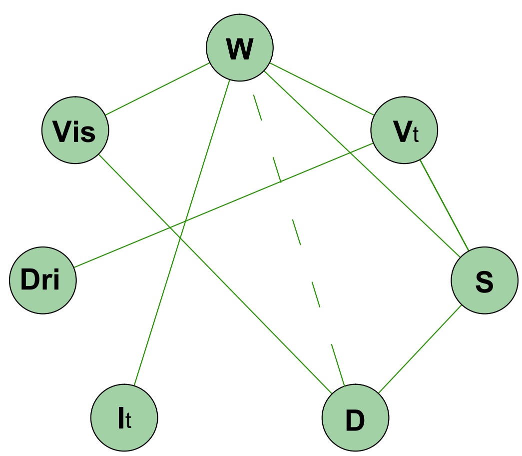


Figure 2.13: Directed graph after moralization.

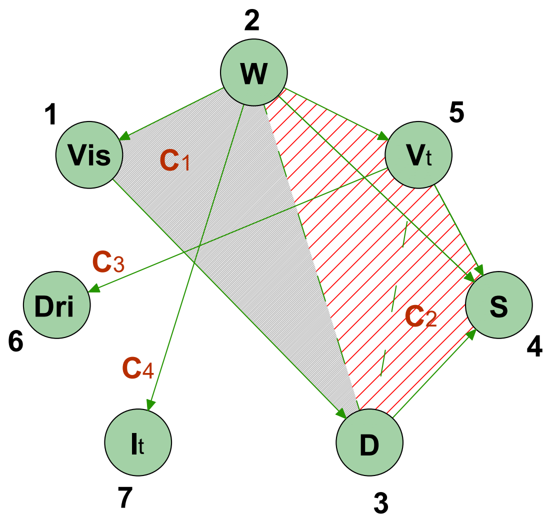


Figure 2.14: Illustration of the chain of cliques of the graph considered with the corresponding triangulation and perfect numbering.

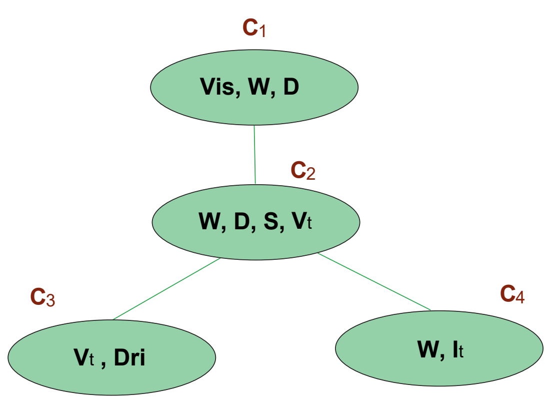


Figure 2.15: Illustration of the corresponding family tree.

Part II

PROPOSED MODEL FOR PROBABILISTIC SAFETY ANALYSIS OF TRAFFIC NETWORKS

Chapter 3

Model presentation

Contents

3.1 Introduction

One of the main concerns of our society today is the large number of accidents that occur on roads. Thus, the studies and improvements on safety traffic analysis on roads are considered necessary and useful. This work focuses on probabilistic safety analysis (PSA) as it is the best technique to asses the safety level of a system. It has been widely used for nuclear power plants and in recent years it has also been employed for railway lines (see for example Castillo et al. (2016b) and Castillo et al. (2016a)). On the other hand, it is known that Bayesian networks (BNs) have a great power to reproduce multidimensional random variables and present numerous advantages with respect to other existing methods such as fault or event trees. Nevertheless, there are few analysis of traffic infrastructures based on them.

A probabilistic safety analysis methodology based on Bayesian networks models for the probabilistic safety assessment (PSA) of highways and roads is proposed in this Thesis. Next, the proposed model following this methodology is introduced and described.

Essentially, there are two main methods to define the qualitative structure of a Bayesian network (BN):

1. *Subjective or experience based method.* The first method is based on our knowledge of the problem, i.e. we decide the direct dependencies among the variables based on the expert knowledge of the problem.

2. *Automatic or objective construction method.* The second method automatically learns the Bayesian network structure based on data.

In order to build the model the first method is used since the knowledge of the direct dependence among variables is sufficiently well known and because the existing data are not the most adequate to provide this information.

The construction of the Bayesian network can be based on the following steps:

1. *Step 1. Selection of the items that have a relevant influence on traffic safety.* This means to analyze the highway or road, for example by recording a video from its start to its end, in order to identify all possible items causing incidents related to traffic safety (light signals, stop signals, pavement failures, intersections, roundabouts, tunnels, acceleration or deceleration lines, and curves).

2. *Step 2. Identification of the variables related to these items.* For example, driver's tiredness, driver's attention, speed, traffic lights states, weather conditions, and traffic intensities.

3. *Step 3. Identification of the direct dependencies among the variables involved.* This means a careful study and analysis of the different variables to identify dependencies and direct relations between pairs of variables.

4. *Step 4. Definition of the conditional probability tables.* We need to define the conditional probability of each variable given their parent variables for each possible combination. If possible, closed formulas must be used to facilitate the analysis and implementation of the Bayesian network.

Once these four steps are covered, all the information needed to evaluate the probability of any possible univariate or multivariate event is available. This is the practical power of Bayesian networks.

The following two sections present the first steps needed to build the Bayesian network: a) the selection of the most relevant items that have influence on traffic safety being likely to cause incidents, and b) the identification of the variables related to these items.

3.2 Items

The identification of all possible items causing incidents related to traffic safety such as light signals, stop signals, pavement failures, intersections, roundabouts, tunnels, and acceleration or deceleration lines, is the first required step to build the Bayesian network. For that, the recording of a video is analyzed along the highway or road. The selection of the items considered of having a relevant influence on traffic safety to be included in the proposed model are presented in Table 3.1, where the names are self-explanatory.

The considered items could lead to different hazardous situations with various severity levels depending on the circumstances of the road at any time, as for example:

Table 3.1: List of all sub-Bayesian network types and associated item types considered in the model.

Subnetwork	**Item**	**Subnetwork**	**Item**
Initial	Initial	Single incident	CurveIn
Signal	Stop		LateralEntry
	Yield		AccelerationLane
	SpeedLimit		Intersection
	SpeedLimitTemp		RoundAbout
	GradeCrossing		Overpass
	PedestrianCrossing		Underpass
	OvertakingIn		ViaductIn
Parameter value change	TrafficChange		ViaductOut
	CurveOut		TunnelIn
	SlopeIn		TunnelOut
	SlopeOut	Warning signal	PermanentWarning
	RoadTypeChange		DistractingWarning
	Continuous		TemporalWarning
	ContinuousOff		OvertakingOut
	WeatherChange	Traffic light	TrafficLight
	WeatherModifOFF	Segment without signal	SegmentWSignals

1. In a segment without signal:
   - Hit with elements on the track.
   - Collision with crossing animals.
   - Crash with other vehicles.
   - Accidents with pedestrian and cyclists.
   - Cuttings and slope stability.
   - Pavement faults.
2. In located points which are divided by types:
   - In lateral entries:

    - Improper exits or entries of vehicles.
    - Road obstruction due to animals.
    - Lack of visibility.
  - In access lanes:
    - Inadequate road incorporations.
    - Driver's distractions.
  - In tunnels:
    - Material detachments.
    - Visibility losses at the entries or exits.
    - Water or snow accumulation.
  - In viaducts:
    - Differential settlements.
    - Wind gusts.
  - In curves:
    - Speed excess.
    - Lack of visibility.
    - Vehicle out of control.
  - In slopes:
    - Visibility losses.
    - Speed differences among users.
  - In intersections:
    - Stop sign trespassing.
    - Speed excess and distraction.

3. In sections of the highway or road where a decision by the driver is needed:
  - Before a stop sign
    - Incorrect decisions.
    - Inadequate signal location.
  - Before a speed limit sign
    - Speed excess.
    - Incorrect decision or inhibition.
  - Before a traffic signal
    - Traffic light trespassing.
    - Incorrect location.

3.3 Variables

As indicated, after the items related to traffic safety have been selected, the next step to define properly the BN model is to identify all the variables which are relevant to the problem being studied. In the present model, the following variables have been considered:

1. *D: Driver's attention.* As it has been noted previously, the human error should not be forgotten because it emerges as one of the most probable road accident cause. This variable represents the driver's attention levels, which in this model are simplified to three states:

   - *Distracted.* In this state the driver is assumed to be unaware of conduction, thus he/she will react with no action at all.
   - *Attentive.* In this state the driver is assumed to be paying attention to the road and it is expected that the required actions will be correct with a high probability, but otherwise the actions will be incorrect (erroneous).
   - *Alert.* In this situation the driver is assumed to react with no error at all.

2. *T: Driver's tiredness.* Since the driver is subject to an increase of tiredness with driving time, a variable is needed to analyze how the driver's tiredness changes when travelling along the road. This element is modeled as a hidden variable, which can modify the driver's attention D nodes. We consider the tiredness increases the probability of errors using a variable $\exp\left(-\frac{t}{t_0}^{2}\right)$, where t_0 is a reference time parameter and t is the driving time.

3. *Sd: Driver's speed decision.* This variable represent the driver's action in cases in which the driver must adjust the driving speed. The assumed values of this variable are:

   - *Correct.* The driver adapts properly the driving speed.
   - *Error I.* The driver fails to adjust the speed to the correct speed because he/she is distracted.
   - *Error II.* The driver fails to adjust the speed to the correct speed due to other reasons.

4. *Dri: Driver type.* This variable refers to the driver's quality, and can take the following values: *professional, experienced, standard* and *bad.*

5. *It: Traffic Intensity.* The incidents that occur along a road are deeply influenced by the traffic intensity. Therefore this information should be modeled as a variable. In this network it is represented by means of the It node, which determines the traffic

frequency along each road segment. The possible values of this variable are *slight traffic, medium traffic* and *heavy traffic*.

6. *Vis: Visualising quality.* This variable refers to the driver's visibility quality, and can take the following values: *good, medium* and *bad*.

7. *Vt: Vehicle type.* The type of traveling vehicles influences significantly the frequencies of incident occurrences along a road, therefore this information should also be modeled as a variable. In this network this variable is represented by V_t with possible values: *heavy vehicle, car* and *motorbike*.

8. *S: Speed.* Speed is a very relevant variable that is represented as a variable which indicates the circulating speed at a road segment. We represent it as a discrete variable with a finite set of values, which can change with the item location.

9. *V: Vehicle failure.* This variable considers the possibility of some vehicle failures and takes values: *no failure, minor failure, medium failure* and *severe failure*.

10. *P: Pavement failure.* This variable reproduces possible incident occurrences because of track deficiencies, such as bump, berm or lane deficiencies, and some undesired incident occurrences, such as the obstruction of the road by stones or material of cuttings, fallen trees, and blocking animals. The values of this variable are: *no failure, minor failure, medium failure* and *severe failure*. This variable is strongly influenced by road characteristics and depends on the environment factors which surround the road.

11. *Co: Collision.* This variable represents some incident occurrences because of the interaction with other vehicles that circulate along the road in the same or contrary direction and can lead to important incidents, such as frontal or lateral collisions. Its values are: *no failure, minor failure, medium failure* and *severe failure*. This variable is clearly influenced by the traffic intensity of the road.

12. *W: Weather.* The weather alters the driver's attention, conditions and tiredness evolution, as well as track and environment behaviour. Therefore it is necessary to consider and to model it as a variable with values: *fair, medium, bad* and *very bad*.

13. *SS: Signal State.* The traffic lights define the availability to trespass a concrete point along the line, therefore its state might influence the traffic behaviour and the incident occurrences. This variable is represented with three possible values: *free*, and *not free*.

14. *AS: Driver's decision at a traffic light.* This variable represents the driver's action at a traffic light, the possible values of this variable are:

    - *Correct.* The driver takes the action that determines the traffic light.

- *Error I.* The driver takes an action contrary to what it is determined by the traffic light because he/she is not aware of the mistake.
- *Error II.* The driver takes an action contrary to what it is determined by the traffic light voluntarily but she/he is aware of the signal state.

15. *DS: Driver's decision at a signal.* This variable represents the driver's action at a mandatory signal, e.g. Stop, yield street or speed limit signals. The possible values of this variable are:
    - *Correct.* The driver takes the correct action.
    - *Error.* The driver takes the wrong action.
16. *TF: Technical failure.* This variable represents a possible failure of the signal. The possible values of this variable are:
    - *Yes.* There is a technical failure.
    - *No.* There is no technical failure.
17. *I: Incident.* This variable includes all possible incidents that may happen along the track at a unique segment location and has values: *no incident, minor incident, medium incident* and *severe incident.*

The notation used for the variables involved in the model and the corresponding sets of values are shown schematically in Table 3.2.

Table 3.2: List of variable types with their definitions and possible values.

VARIABLE	DEFINITION	VALUES
D: Driver's attention	This variable represents the driver's attention	Distracted, Attentive, Alert
T: Driver's tiredness	Measures the driver's tiredness	A positive value increasing with driving time
Sd: Driver's speed decision	It represents the driver's action in cases in which the driver must adjust the driving speed	Correct, Error I, Error II
Dri: Driver type	This variable refers to the driver's quality	Professional, Experienced, Standard, Bad
It: Traffic Intensity	This variable refers to the traffic intensity	Slight traffic, Medium traffic and Heavy traffic
Vis: Visualising quality	This variable refers to the driver's visibility quality	Good, Medium and Bad
Vt: Vehicle type	It refers to the vehicle type	Heavy vehicle, Car and Motorbike
S: Speed	Speed	Set of positive real values
V: Vehicle failure	This variable considers the possibility of some vehicle failures	No failure, Minor failure, Medium failure and Severe failure
P: Pavement failure	This variable reproduces possible incident occurrences because of track deficiencies and some undesired incident occurrences	No failure, Minor failure, Medium failure and Severe failure
Co: Collision	This variable represents some incident occurrences because of the interaction with other vehicles that circulate along the road in the same or contrary direction	No failure, Minor failure, Medium failure and Severe failure
W: Weather	This variable represents the type of weather	Fair, Medium, Bad, and Very bad
SS: Signal State	This variable represents the signal state	Free, Not free
AS: Driver's decision at a traffic light	This variable represents the driver's decision at a traffic light	Correct, Error I, Error II
DS: Driver's decision at a signal	This variable represents the driver's action at a mandatory signal, e.g. Stop, yield street or speed limit signals	Correct, Error
TF: Technical failure	This variable represents a possible failure of the signal.	Yes, No
I: Incident	This variable includes all possible incidents that may happen along the track at a unique segment location	No incident, Minor incident, Medium incident, Severe incident

Chapter 4

Sub-Bayesian networks

Contents

4.1 Introduction

Once the items and the variables have been defined, the structure of the Markovian-Bayesian network should be designed. The previously defined elements can be modelled as 7 different sub-Bayesian networks with a particular structure and associated variables with their corresponding conditional probabilities.

The subnetworks components associated with the different items are described and shown below, where the green nodes, joined by a dashed blue square represent the nodes of the subnetwork analyzed in each case, and the grey nodes are those nodes of other subnetworks to which they are connected. The links of the subnetwork are indicated as green arrows and other links appear as dashed arrows. The node types appear in the center of the node circles that contain the corresponding type names and all the variables have associated initial marginal or conditional probabilities to be defined later.

4.2 Initial sub-Bayesian network

At the beginning of a road an initial subnetwork must be used so this sub-Bayesian network reproduces the situation when a user starts the driving and contains seven nodes that represent the following variables: a) the weather conditions W, b) the vehicle type Vt, c) the driver type Dri, d) the traffic intensity It, e) the visibility conditions Vis, f) the driver's attention D, and g) the speed S. This is illustrated in Figure 4.1.

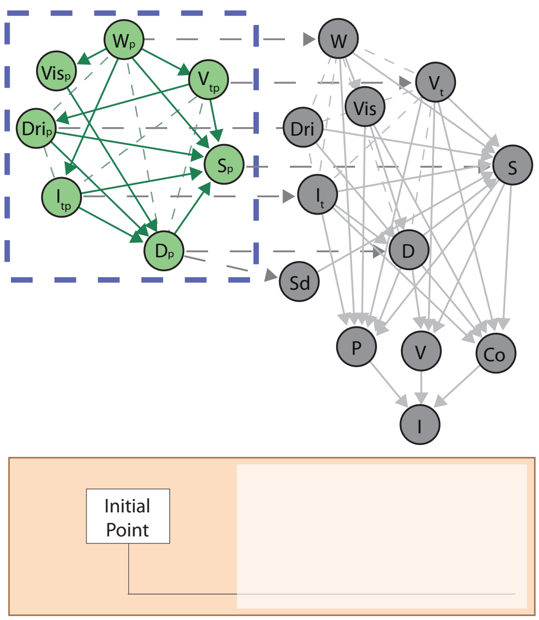


Figure 4.1: Initial sub-Bayesian network.

The item associated in this case is the initial location.

4.3 Segment without signs sub-Bayesian network

To consider the possibility of incidents due not to particular locations but to continuous elements in the road segment, a segment without signs sub-Bayesian network must be located between consecutive items. It contains twelve nodes: a) the weather conditions W, b) the vehicle type Vt, c) the driver type Dri, d) the traffic intensity It, e) the visibility conditions Vis, f) the driver's attention D, g) the speed S, h) the driver's speed decision Sd, i) the vehicle failure V, j) the pavement failure P, k) the collision variable

Co, and l) the incident variable I, as shown in Figure 4.2, and is associated with segment without sign items.

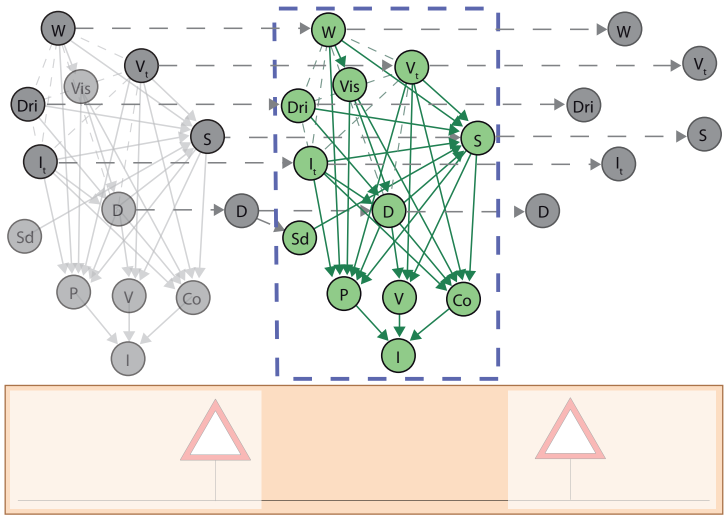


Figure 4.2: Segment without signs sub-Bayesian network.

Despite of the fact that the driver's attention affects significantly the incident occurrences and their level, even when drivers are completely alert, other types of incidents can take place at segments without signals. In fact, vehicle and pavement failures can cause accidents, and some collisions among vehicles in the same or contrary directions can occur. This is what variables V, P and Co take into account.

4.4 Concentrated incident sub-Bayesian network

This structure represents cases in which a concentrated incident can occur. This subnetwork contains only the I variable, as can be seen in Figure 4.3 and includes the following items: a) intersection, b) roundabout, c) tunnels, d) viaducts, e) underpasses, f) overpasses, g) lateral entries, and h) acceleration lanes.

4.5 Sign sub-Bayesian network

At locations where some action subject to error is required this structure is used. This subnetwork has associated the following items: a) yield, b) stop, c) grade crossing, d) pedestrian crossing, e) speed limit, and f) prohibition signs, and contains five variables: a) driver's attention D, b) driver's decision at a a signal Ds, c) speed S, d) technical failure TF, and e) incident I. This is represented in Figure 4.4.

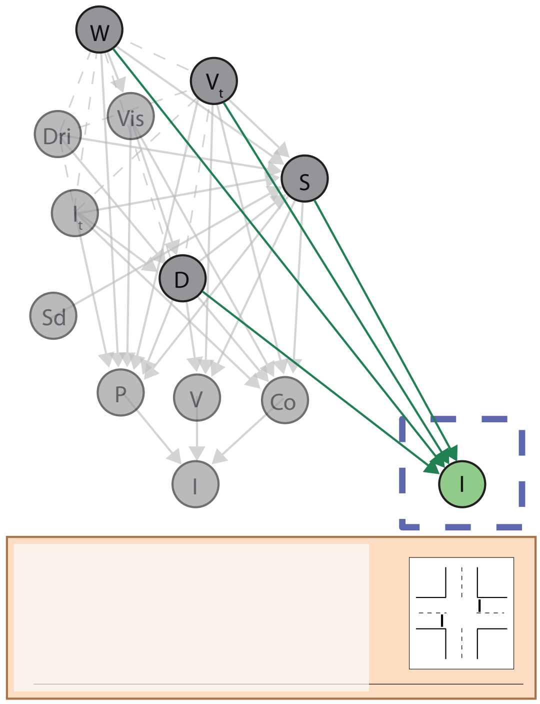


Figure 4.3: Single incident sub-Bayesian network.

4.6 Traffic signal sub-Bayesian network

This subnetwork is used for traffic lights, where some decision subject to error is needed. The only associated item is the traffic light and it contains six variables: a) driver's attention D, b) driver's decision at a a signal Ds, c) speed S, d) technical failure TF, e) signal state SS, and f) incident I, as illustrated in Figure 4.5.

4.7 Warning signal sub-Bayesian network

The only variable that appears in this subnetwork is the driver's attention D because this structure is used to reproduce the driver changes of attention due to warning signals. The following items are included: a) permanent warning signals, b) temporal warning signals, c) distracting warning signals (billboards), and d) end of bans, and Figure 4.6 shows this structure.

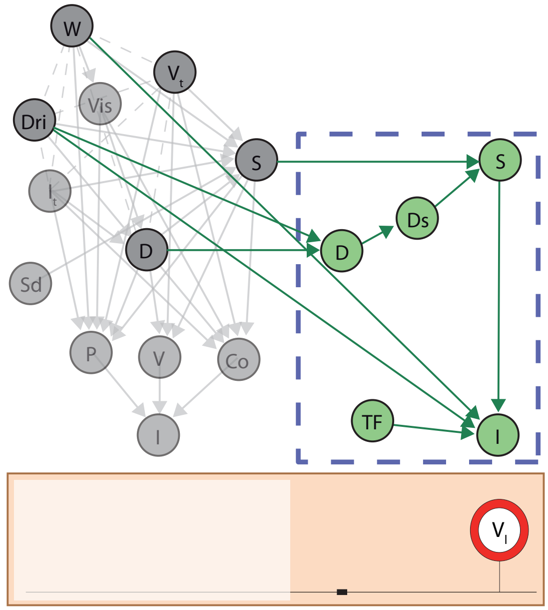


Figure 4.4: Generic signal sub-Bayesian network.

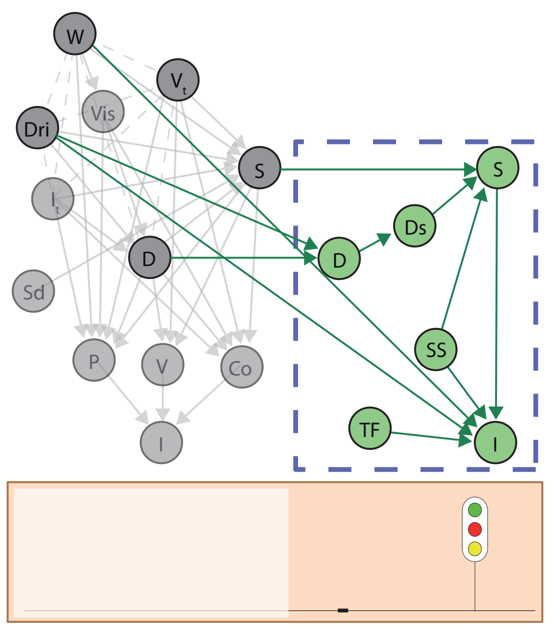


Figure 4.5: Traffic light sub-Bayesian network.

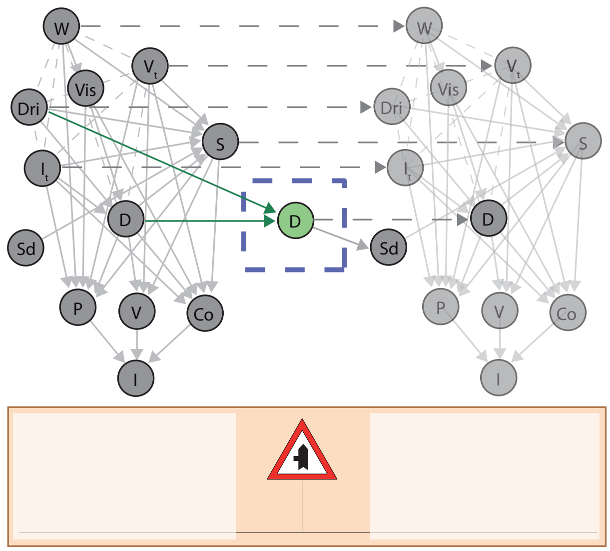


Figure 4.6: Warning signal sub-Bayesian network.

4.8 Parameter value change sub-Bayesian network

This structure is used to reproduce the parameter changes of track, environment or traffic characteristics among others. It includes the items: a) weather change, b) traffic intensity change, c) road quality change, and d) terrain change, and does not add any variable. This subnetwork can be seen in Figure 4.7.

4.9 Overtaking sub-Bayesian network

In the case of overtaking signals a first vertical signal of forbidden overtaking is usually found and after a certain section considered, a signal of ending this prohibition. Figure 4.8 illustrates this representation showing the case of two overtaking subnetworks associated with two no overtaking signs with all their variables (the second and fourth group of variables) and the neighbours subnetworks (first, third and fifth).

It must be noted that there are two types of subnetworks, those associated with segments without signs, where the risk is distributed along its length, and those associated with signs, where the risk is associated with their locations.

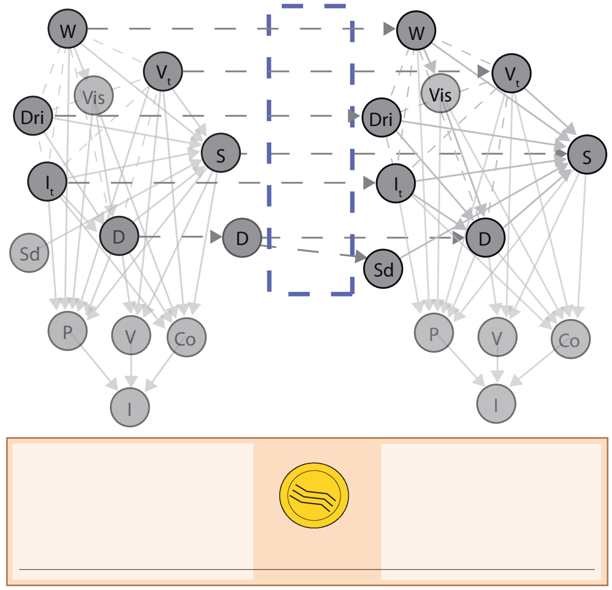


Figure 4.7: Parameter value change sub-Bayesian network.

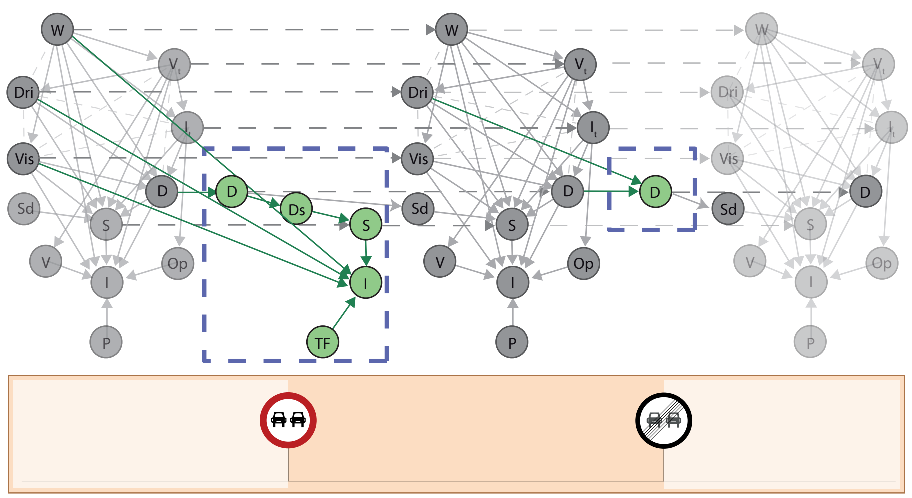


Figure 4.8: Overtaking subnetwork associated with the two no overtaking signs with all their variables (the second and fourth group of variables) and the neighbours subnetworks (first, third and fifth).

4.10 Insertion of the sub-Bayesian networks into the general network

Any road considered would be represented by different subnetworks along its length that they would be included within the general network to study the road as a whole. Figure

4.9 illustrates how the different subnetworks are integrated into the general network. The right subnetworks in this figure correspond to the locations where signs are located, and the left sub-networks correspond to the segments without signs, that is, where the risk is a function of the segment length.

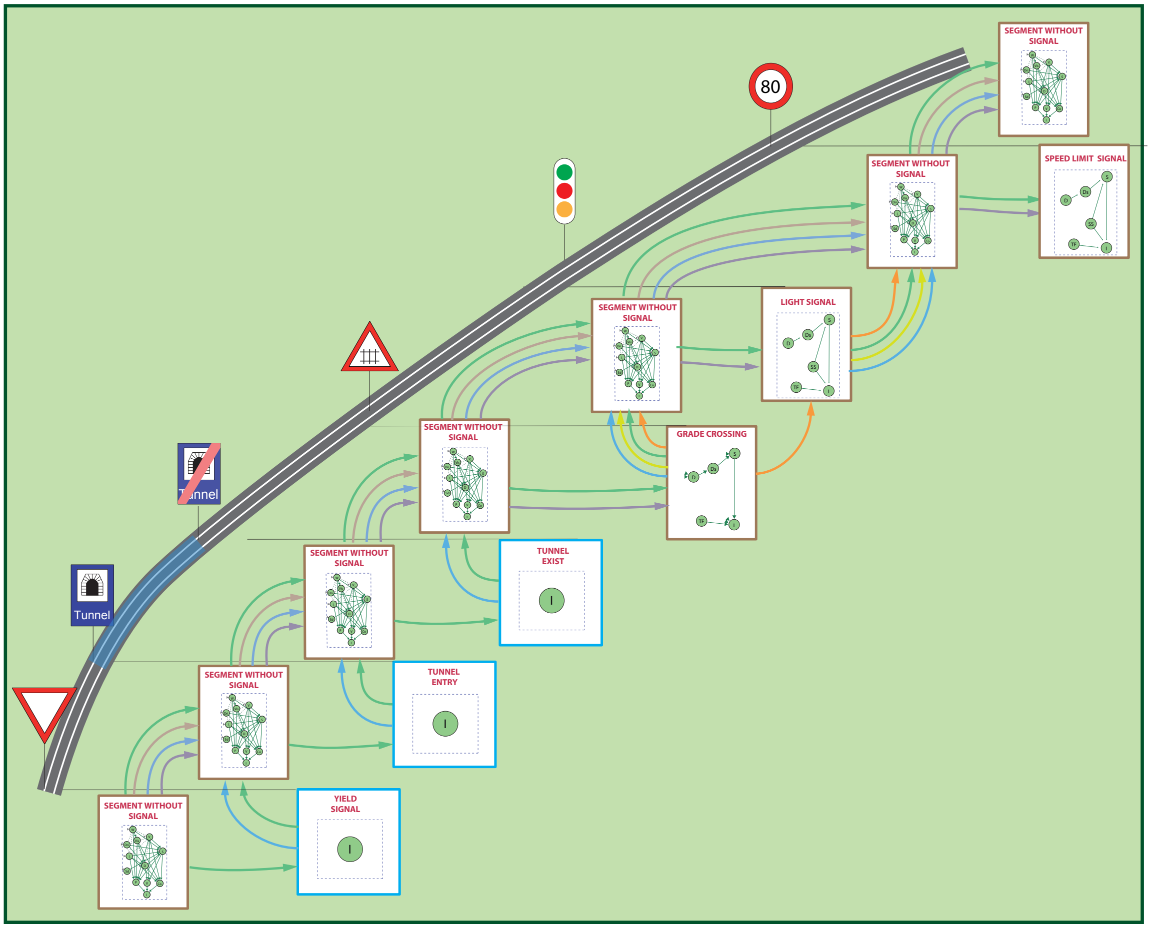


Figure 4.9: Illustration of how the different sub-networks are integrated into the general network. They include the segment without signs sub-networks (the left ones) and other subnetworks (the right ones).

Chapter 5

Conditional probability tables. Closed formulas

Contents

5.1 Introduction

One difficult step in the Bayesian network construction process is the definition of the conditional probability tables. It is necessary to provide the conditional probabilities tables of each node given their parents. As discussed below, this can be facilitated if closed formulas can be derived. Some examples that show how this can be done are provided in Section 5.3.

The different assumptions which must be made and the analysis of the conditional probability tables of the Bayesian network model are explained next.

5.2 Conditional probabilities of the initial subnetwork

In this section the assumptions made for the conditional probabilities of the nodes given their parents for the initial subnetwork shown in Figure 4.1 are discussed.

5.2.1 Conditional probability of node W weather

Since the weather node W has no parents, its conditional probability coincides with its marginal probability, which is assumed to be a multinomial probability with possible values: *'fair'*, *'medium'*, *'bad'* and *'very bad'*. The elements of this marginal probability are parameters to be given. However, in order to facilitate its definition, a set of non-negative values can be given instead, which indicate the relative frequencies of each possible value of variable W. These values are automatically normalized to obtain a valid probability.

5.2.2 Conditional probability of node V_t vehicle type

Since node V_t has one parent variable, W, it is necessary to define the conditional probabilities $P(V_t|W)$, of V_t given all values of W, which are assumed to be multinomial variables with possible values: *'heavy vehicle'*, *'car'* and *'motorbike'*. To this end the conditional probability $q(v) = P(V_t = v|W = \text{'}fair')$ is defined, for all values of v, for the case of *'fair'* weather. To give more flexibility, we allow for a true probability (non-negative values adding up to one) or simply non-negative values (relative frequencies), which are adequately normalized. Next, to define $P(V_t = v|W = w)$ for $w \neq \text{'}fair'$, that is, for other weather values different from 'fair', non-negative factors $f_h(w)$ and $f_m(w)$ are used, for heavy vehicles and motorbikes, respectively, which are the correction factors to obtain the relative frequencies of these vehicles for all weather types w. Consequently, the conditional probability becomes the matrix in Table 5.1. Note that $f_h(\textit{`fair'}) = 1$ and $f_m(\textit{`fair'}) = 1$ when defined or after adequate normalization.

Table 5.1: Conditional probability $P(V_t = v|W = w)$.

Conditional probability $P(V_t = v\|W = w)$			
	v		
w	'Heavy'	'Car'	'Motorbyke'
'Fair'	$q(1)f_h(1)$	$1 - q(1)f_h(1) - q(3)f_m(1)$	$q(3)f_m(1)$
'Medium'	$q(1)f_h(2)$	$1 - q(1)f_h(2) - q(3)f_m(2)$	$q(3)f_m(2)$
'Bad'	$q(1)f_h(3)$	$1 - q(1)f_h(3) - q(3)f_m(3)$	$q(3)f_m(3)$
'Very bad'	$q(1)f_h(4)$	$1 - q(1)f_h(4) - q(3)f_m(4)$	$q(3)f_m(4)$

5.2.3 Conditional probability of node Dri driver type

Since node Dri has one parent variable, V_t, it is necessary to define the conditional probabilities $P(Dri|V_t)$, of Dri given all values of V_t, which are multinomial variables. Thus, we have

$$p_{dv} = P(Dri = d|V_t = v),$$

where p_{dv} are parameters to be given.

As in the case of node W, this matrix can be replaced by a matrix of non-negative numbers, referring to the relative frequencies, because it can be adequately normalized to become a valid conditional probability.

5.2.4 Conditional probability of node I_t traffic intensity

Since node I_t has one parent variable, W, it is necessary to define the conditional probabilities $P(I_t = i|W = w)$, of I_t given all possible values of W. To this end, the I_t is assumed to have a normal density $N(\alpha(w)\mu, \alpha(w)\sigma)$, where μ and σ are the mean intensity and its standard deviation, respectively, for the case $W = w = \text{`}fair'$, that is for $\alpha(1) = 1$. To simplify, the normal variable is discretized using two intensity threshold values It_m and It_h, respectively, and a multinomial variable with three possible values for the traffic intensity It: *'slight'* ($It \leq It_m$), *'medium'* ($It_m < It \leq It_h$) and *'heavy'* ($It > It_h$), is considered. Thus, the conditional probability $P(I_t = i|W = w)$ becomes the one in Table 5.2 where $\Phi()$ is the cdf of the standard normal variable,

$$\begin{aligned} &\mu_{mw} = \alpha(w)\mu; \quad \sigma_{mw} = \alpha(w)\sigma; \\ &w = \textit{`fair', `medium', `bad'} \text{ and } \textit{`very bad'} \end{aligned} \tag{5.1}$$

and the $\alpha(w)$ are non-negative correction factors to consider the effect of weather type w.

Table 5.2: Conditional probability $P(I_t = i|W = w)$.

Conditional probability $P(I_t = i\|W = w)$			
	i		
w	'Slight'	'Medium'	'Heavy'
'Fair'	$\Phi\left(\frac{It_m - \mu_{m1}}{\sigma_{m1}}\right)$	$\Phi\left(\frac{It_h - \mu_{m1}}{\sigma_{m1}}\right) - \Phi\left(\frac{It_m - \mu_{m1}}{\sigma_{m1}}\right)$	$1 - \Phi\left(\frac{It_h - \mu_{m1}}{\sigma_{m1}}\right)$
'Medium'	$\Phi\left(\frac{It_m - \mu_{m2}}{\sigma_{m2}}\right)$	$\Phi\left(\frac{It_h - \mu_{m2}}{\sigma_{m2}}\right) - \Phi\left(\frac{It_m - \mu_{m2}}{\sigma_{m2}}\right)$	$1 - \Phi\left(\frac{It_h - \mu_{m2}}{\sigma_{m2}}\right)$
'Bad'	$\Phi\left(\frac{It_m - \mu_{m3}}{\sigma_{m3}}\right)$	$\Phi\left(\frac{It_h - \mu_{m3}}{\sigma_{m3}}\right) - \Phi\left(\frac{It_m - \mu_{m3}}{\sigma_{m3}}\right)$	$1 - \Phi\left(\frac{It_h - \mu_{m3}}{\sigma_{m3}}\right)$
'Very bad'	$\Phi\left(\frac{It_m - \mu_{m4}}{\sigma_{m4}}\right)$	$\Phi\left(\frac{It_h - \mu_{m4}}{\sigma_{m4}}\right) - \Phi\left(\frac{It_m - \mu_{m4}}{\sigma_{m4}}\right)$	$1 - \Phi\left(\frac{It_h - \mu_{m4}}{\sigma_{m4}}\right)$

5.2.5 Conditional probability of node Vis visibility

Since node Vis has one parent variable, W, it is necessary to define the conditional probabilities $P(Vis = v|W = w)$, of Vis given all values of W, which are assumed multinomial variables with possible values: *'good'*, *'medium'* and *'bad'*. However, in order to facilitate its definition, a set of non-negative values can be given instead, which indicate the relative frequencies of each possible value of variable W. These values are later normalized to obtain a valid probability.

5.2.6 Conditional probability of node D driver's attention

Since node D has three parent variables, Dri, It, Vis, it is necessary to define the conditional probabilities $P(D = a|Dri = b, It = c, Vis = d)$ for each possible combination of the conditioning variables. To this end, it is assumed initially that D is a normal random variable $N(\mu, 1)$ for a reference combination of Dri, It, Vis, which has been selected as a *'standard'* driver, a *'medium'* intensity and a *'medium'* visibility. Given p_d and p_a, the probabilities of the driver being *'distracted'* and *'attentive'*, respectively, it is possible to obtain two threshold values th_d and th_a for D as

$$\Phi(th_i - \mu) = p_i \Leftrightarrow th_i = \mu + \Phi^{-1}(p_i); \quad i \in \{d, a\}. \tag{5.2}$$

These two threshold values, th_d and th_a, allow us to discretize the variable D as a multinomial variable with possible values *'distracted'* ($D \leq th_d$), *'attentive'* ($th_d < D \leq th_a$) and *'alert'* ($D > th_a$).

It is also assumed that other combinations of variables Dri, It, Vis modify the threshold reference values, th_d and th_a, by means of dimensionless correction factors $\alpha(b)$, $\beta(c)$ and $\gamma(d)$ of the driver's type (b), the intensity (c) and the visibility level (d), respectively, that is, the new threshold values become $th_d\alpha(b)\beta(c)\gamma(d)$ and $th_a\alpha(b)\beta(c)\gamma(d)$, respectively.

With this, the conditional probability $P(D = a|Dri = b, It = c, Vis = d)$ becomes:

$$(\delta_{1a} - \delta_{2a})\Phi(th_d\alpha(b)\beta(c)\gamma(d) - \mu) + (\delta_{2a} - \delta_{3a})\Phi(th_a\alpha(b)\beta(c)\gamma(d) - \mu) + \delta_{3a}. \quad (5.3)$$

Note that, due to the chosen reference combination, we must have $\alpha(\text{'}standard\text{'}) = 1$, $\beta(\text{'}medium\text{'}) = 1$ and $\gamma(\text{'}medium\text{'}) = 1$; otherwise, the correction factors must be normalized by dividing them by the adequate values.

5.2.7 Conditional probability of node S speed

Since node S has five parent variables, W, Vt, Dri, It and D, it is necessary to define the conditional probabilities $P(S|W, Vt, Dri, It, D)$ for any combination of the conditioning variables.

To this end, it is initially assumed that the S variable is a gamma distribution $G(k, \theta)$ with mode:

$$S_{mode}(b, c, d, e) = (k - 1)\theta = S_{max}\alpha(b)\beta(c)\gamma(d)\rho(e), \quad (5.4)$$

where $\alpha(b), \beta(c), \gamma(d)$ and $\rho(e)$ are factors to take into account the effect of weather W, vehicle type Vt, driver's type Dri and intensity It, respectively.

Next, the S variable is discretized as a multinomial with given values $s(1), s(2), \ldots, s(a^*)$, where a^* is the last index of the speed variable S, and it is obtained the following conditional probability $p(a, b, c, d, e, f) = P(S = a|W = b, Vt = c, Dri = d, It = e, D = f)$:

$$\begin{aligned} p(a, b, c, d, e, f) \quad &= \quad (1 - \delta_{aa^*})F_{G(1+S_{mode}(b,c,d,e)/\theta,\theta)}\left(\frac{s(a) + s(a+1)}{2}\right) \\ &\quad -(1 - \delta_{a1})F_{G(1+S_{mode}(b,c,d,e)/\theta,\theta)}\left(\frac{s(a-1) + s(a)}{2}\right) + \delta_{aa^*}, \quad (5.5) \end{aligned}$$

where $F_{G(a,b)}(x)$ is the cdf of the gamma random variable.

5.3 Conditional probabilities of intermediate subnetworks without signals

In this section the assumptions made for the conditional probabilities of the nodes given their parents for intermediate subnetworks without signals, shown in Figure 4.2, are discussed. It is possible to see how due to the large number of possible combinations this is a complicated process. For this reason, the use of closed formulas which is explained below is proposed.

A general way to define the conditional probabilities of the form $P(A|B_1, B_2, \ldots, B_k)$ is introduced, where $\mathcal{P}(A) \equiv \{B_1, B_2, \ldots, B_k\}$ is the set of parents of node A.

We look for valid formulas of the probabilities $p_{ab_1b_2\ldots b_k} = P(a|b_1, b_2, \ldots, b_k)$, where the lower case letters refer to the values of the corresponding upper case letters, that is, they must be non-negative and the sum of the probabilities of all values of the son for any combination of parent values must be one. When these sets of probabilities are generated automatically, care should be taken to check these conditions.

Here it is proposed a formula of the form:

$$p_{a,b_1,b_2,\ldots,b_k} = \sum_{j_1} q_{j_1}(\theta_{j_1}) \left[\sum_{j_2} q_{j_1,j_2}(\theta_{j_1,j_2}) \left[\ldots \right.\right.$$

$$\left.\left.\left[\sum_{j_{s_{r-1}}} q_{j_1,\ldots,j_{s_{r-1}}}(\theta_{j_1,\ldots,j_{s_{r-1}}}) \left[\sum_{j_{s_r}}^{n_a} q_{j_1,\ldots,j_{s_r}}(\theta_{j_1,\ldots,j_{s_r}}) \delta_{aj_{s_r}} \right]\right]\right]\right], \tag{5.6}$$

where the right hand side term is a sum of products, the index j_{s_r}, which depends on the summand being considered, refers to the last factor in each summand, n_a is the cardinal of the set of values of node A, $\theta_{j_1,j_2,\ldots,j_{s_t}}$ are vectors of parameters, and all $q_{j_1,j_2,\ldots,j_{s_t}}; t = 1, 2, \ldots, r$ are non-negative valued functions of a subset of parents of node A, that is, whose arguments are a subset of $\{b_1, b_2, \ldots, b_k\}$, and they must satisfy

$$\sum_{j_t} q_{j_1,\ldots,j_t}(\theta_{j_1,\ldots,j_t}) = 1; \quad \forall j_1, j_2, \ldots, j_t; t = 1, 2, \ldots, s_r. \tag{5.7}$$

Note that the last and only the last summation contains the son values a. This guarantees that $p_{a,b_1,b_2,\ldots,b_k}$ is a valid conditional probability table, because all terms are non-negative, there are no minus signs and we have

$$\sum_{a=1}^{n_a} p_{a,b_1,b_2,\ldots,b_k} = \sum_{a=1}^{n_a} \sum_{j_1} q_{j_1}(\theta_{j_1}) \left[\sum_{j_2} q_{j_1,j_2}(\theta_{j_1,j_2} \left[\ldots \right.\right.$$

$$\left.\left.\left[\sum_{j_{s_{r-1}}} q_{j_1,\ldots,j_{s_{r-1}}}(\theta_{j_1,\ldots,j_{s_{r-1}}}) \left[\sum_{j_{s_r}}^{n_a} q_{j_1,\ldots,j_{s_r}}(\theta_{j_1,\ldots,j_{s_r}}) \delta_{aj_{s_r}} \right]\right]\right]\right]$$

$$= \sum_{j_1} q_{j_1}(\theta_{j_1}) \left[\sum_{j_2} q_{j_1,j_2}(\theta_{j_1,j_2}) \left[\ldots \left[\sum_{j_{s_{r-1}}} q_{j_1,\ldots,j_{s_{r-1}}}(\theta_{j_1,\ldots,j_{s_{r-1}}}) \right.\right.\right.$$

$$\left.\left.\left.\left[\sum_{j_{s_r}}^{n_a} q_{j_1,\ldots,j_{s_r}}(\theta_{j_1,\ldots,j_{s_r}}) \sum_{a=1}^{n_a} \delta_{aj_{s_r}} \right]\right]\right]\right] = 1, \tag{5.8}$$

where it has been taken into account that $\sum_{a=1}^{n_a} \delta_{aj_{s_r}} = 1$ and the set of constraints (5.7).

Next, several examples that clarify the meaning of the $q_{j_1,\ldots,j_{s_r}}$ functions are given. Figures 5.2 to 5.8 will show these functions for the particular cases of driver's speed decision, vehicle failure, collision, pavement failure, curves, and traffic lights nodes, respectively.

5.3.1 Conditional probability of node W weather

Since the node W has a parent Wp, its conditional probability $P(W = w|Wp = wp)$ must be obtained for all values of Wp. It is assumed that W is a multinomial variable with possible values: 'good', 'medium', 'bad', and 'very bad' and the conditional probability is calculated with the rates in Table 5.3, where α_1 and α_2 are two factors which are different from zero when a weather deterioration occurs and whose values are the factors that should be used for one and two stages of deterioration, respectively, β_1 and β_2 are two factors that are different from zero when a weather improvement takes place and whose values are the factors that should be used for one and two stages of improvement, respectively, and the parameters λ are the different rates.

Table 5.3: Rates used to calculate the conditional probability $P(W = w|Wp = wp)$.

Rates used for conditional probability $P(W = w\|Wp = wp)$				
	w			
wp	'Fair'	'Medium'	'Bad'	'Very bad'
'Fair'	0	$\alpha_1\lambda_2$	$\alpha_2\lambda_3$	0
'Medium'	$\beta_1\lambda_1$	0	$\alpha_1\lambda_3$	$\alpha_2\lambda_4$
'Bad'	$\beta_2\lambda_1$	$\beta_1\lambda_2$	0	$\alpha_1\lambda_4$
'Very bad'	0	$\beta_2\lambda_2$	$\beta_1\lambda_3$	0

5.3.2 Conditional probability of node V_t vehicle type, Dri driver type, and I_t traffic intensity

Since the type of vehicle V_t, the type of driver Dri and the traffic intensity I_t have a parent V_{tp}, $Drip$ and I_{tp}, respectively, their conditional probabilities $P(V_t|V_{tp})$, $P(Dri|Drip)$, and $P(I_t|I_{tp})$, must be calculated for all values of the conditioning variables. Since these variables do not change within a link, the type of vehicle, the type of driver, and the traffic intensity are the same along the segment. This means that the conditional probability tables $P(V_t|V_{tp})$, $P(Dri|Drip)$, and $P(I_t|I_{tp})$ coincide with the identity matrix.

5.3.3 Conditional probability of node Vis visibility

Since node Vis has one parent variable, W, it is necessary to define the conditional probabilities $P(Vis = v|W = w)$, of Vis given all values of W, which are assumed multinomial variables with possible values: *'fair', 'medium'* and *'bad'*. However, in order to facilitate its definition, a set of non-negative values can be given instead, which indicate the relative frequencies of each possible value of variable W. These values are later normalized to obtain a valid probability.

5.3.4 Conditional probability of node D driver's attention

Since any node D at any intermediate subnetwork has four parent variables, D_p, Dri, It, Vis, the conditional probabilities $P(D = a|D_p = b, Dri = c, It = d, Vis = e)$ for each possible combination of the conditioning variables need to be defined.

Before analyzing the conditional probability, the Markov model for the driver's attention is discussed.

Markov's model for driver's attention

In this subsection how the driver's attention changes with time or traveled distance is discussed.

In order to reproduce the driver and the vehicle behaviour at a segment without signals, discrete or continuous Markov processes can be used (see references Benjamin and Cornell (1970), Doob (1953) or Kijima (1997)). In the present model the continuous case is only considered, because it seems to be more appropriate.

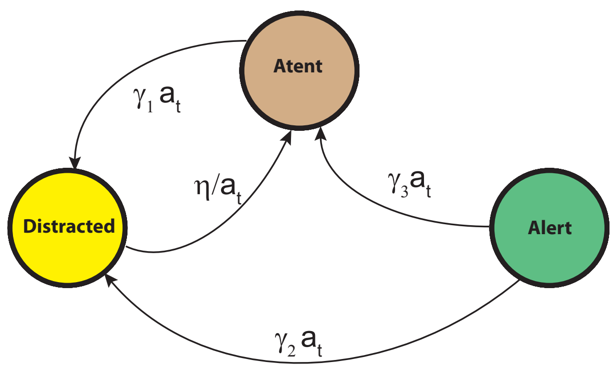


Figure 5.1: Markov model of driver's attention states: Markov model illustrating the transitions among different attention levels and the corresponding probabilities.

In this case we have the differential equation associated with a standard Markov model (see Figure 5.1):

$$\begin{pmatrix} p'_0(t^*, t - t_0) \\ p'_1(t^*, t - t_0) \\ p'_2(t^*, t - t_0) \end{pmatrix} = \begin{pmatrix} -\eta/a_{t^*} & \gamma_1 a_{t^*} & \gamma_2 a_{t^*} \\ \eta/a_{t^*} & -\gamma_1 a_{t^*} & \gamma_3 a_{t^*} \\ 0 & 0 & -(\gamma_2 + \gamma_3) a_{t^*} \end{pmatrix} \times \begin{pmatrix} p_0(t^*, t - t_0) \\ p_1(t^*, t - t_0) \\ p_2(t^*, t - t_0) \end{pmatrix}, \qquad (5.9)$$

where $p_0(t^*, t)$, $p_1(t^*, t)$ and $p_2(t^*, t)$ are the probabilities associated with the three driver's states (distracted, attentive and alert), t, t_0 and t^* are a given time, the segment starting time and the segment center time, respectively. η/a_{t^*} is the time rate of recovering attention when the driver is distracted after a time t^* from the beginning of the trip,

$\gamma_i a_{t^*}; i = 1, 2$ are the time rates of becoming distracted when being attentive and alert, respectively, and $\gamma_3 a_{t^*}$ is the time rate of becoming attentive when the driver is alert, which leads to the model

$$\begin{pmatrix} p_0(t^*, t - t_0) \\ p_1(t^*, t - t_0) \\ p_2(t^*, t - t_0) \end{pmatrix} = M_\ell(t^*, t - t_0; \eta, \boldsymbol{\gamma}_1, \gamma_2, \gamma_3) \begin{pmatrix} p_0^0 \\ p_1^0 \\ p_2^0 \end{pmatrix}, \tag{5.10}$$

where ℓ is the segment length and $p_i^0 = p_i(t^*, 0); \quad i = 1, 2, 3.$

Formula (5.10) permits us to calculate how the probabilities of each of the three states (distracted, attentive and alert) evolve with time $t - t_0$.

Thus, the conditional probabilities for node D given node D_p are given by $p_{a,b} = P(D = a|D_p = b)$:

$$\begin{pmatrix} \delta_{a,1} & \delta_{a,2} & \delta_{a,3} \end{pmatrix} M_\ell(t^*, t_{end} - t_0; \eta, \boldsymbol{\gamma}_1, \gamma_2, \gamma_3) \begin{pmatrix} \delta_{b,1} \\ \delta_{b,2} \\ \delta_{b,3} \end{pmatrix}, \tag{5.11}$$

where t_{end} is the segment exit time.

The $p_{a,b}$ probabilities allow us to calculate the marginal probabilities p_d and p_a of the driver being *'distracted'* and *'attentive'*, respectively.

Next, the threshold values th_d and th_a associated with p_d and p_a are calculated, that is

$$th_i = \alpha(b)\beta(c)\gamma(d) + \Phi^{-1}(p_i); \quad i \in \{d, a\}, \tag{5.12}$$

Finally, the conditional probability

$$p(a, b, c, d, e) = P(D = a|D_p = b, Dri = c, It = d, Vis = e)$$

becomes:

$$\begin{aligned} p(a, b, c, d, e) &= (\delta_{1a} - \delta_{2a})\Phi(th_d\alpha(b)\beta(c)\gamma(d) - \mu) \\ &\quad + (\delta_{2a} - \delta_{3a})\Phi(th_a\alpha(b)\beta(c)\gamma(d) - \mu) + \delta_{3a}. \end{aligned} \tag{5.13}$$

5.3.5 Conditional probability of node Sd driver's speed decision

Since the driver's speed decision node Sd has one parent D_p, its conditional probability $P(Sd|D_p)$ must be defined for all values of D_p. For that, the following assumptions are made:

1. There is no action if the driver is in the *'distracted'* state, that is, the error is of type I.

2. If the driver is in the *'attentive'* state, there is a probability τa_t of making an error and this error will be of type I with probability κ and of type II, otherwise, where a_t is the factor used to correct the probability τ of making an error due to driver's tiredness.

3. There is no error in the decision if the driver is in the *'alert'* state.

With these assumptions, the conditional probability $P(Sd = a|D = b)$ becomes as given in Table 5.4. This probability is also represented in Figure 5.8 by means of a closed formula as explained previously. Here it can be seen that the conditional probability is of the form (5.6) and, the q functions can be identified and Formula (5.14) obtained, by simply adding the contributions of all branches.

Table 5.4: Conditional probability $P(Sd = a|D_p = b)$.

Conditional probability $P(Sd = a\|D_p = b)$			
	a		
b	'correct'	'Error I'	'Error II'
'distracted'	0	1	0
'attentive'	$1 - \tau a_t$	$\tau a_t \kappa$	$\tau a_t(1 - \kappa)$
'alert'	1	0	0

$$P(Sd = a|D = b) = \delta_{b1}\delta_{a2} + \delta_{b2}\left[\tau\left(\kappa\delta_{a2} + (1 - \kappa)\delta_{a3}\right) + (1 - \tau)\delta_{a1}\right] + \delta_{b3}\delta_{a1}. \quad (5.14)$$

5.3.6 Conditional probability of node V vehicle failure

Since node V has three parent variables, Vt, D and S, the conditional probabilities $P(V = a|Vt = b, D = c, S = d)$ for any combination of the conditioning variables need to be defined. To this end, first the probability Z of a vehicle failure is calculated assuming that there is a failure rate z_t per unit length, which depends on the track type, and also on the variables vehicle type Vt and driver's attention D, that is, the probability of failure Z becomes:

$$Z_{b,c} = L z_t \gamma_{Vt}(b)\gamma_D(c), \quad (5.15)$$

where L is the distance between the previous and the actual locations, z_t is a failure rate dependent on the track type, $\gamma_{Vt}(b)$ and $\gamma_D(c)$ are correction factors to consider the vehicle type Vt and the driver's attention D, respectively.

Since given that a vehicle failure has occurred, the severity of the incident depends on its speed. Thus, the following different speed cases are considered next: low ($S_p < 30$

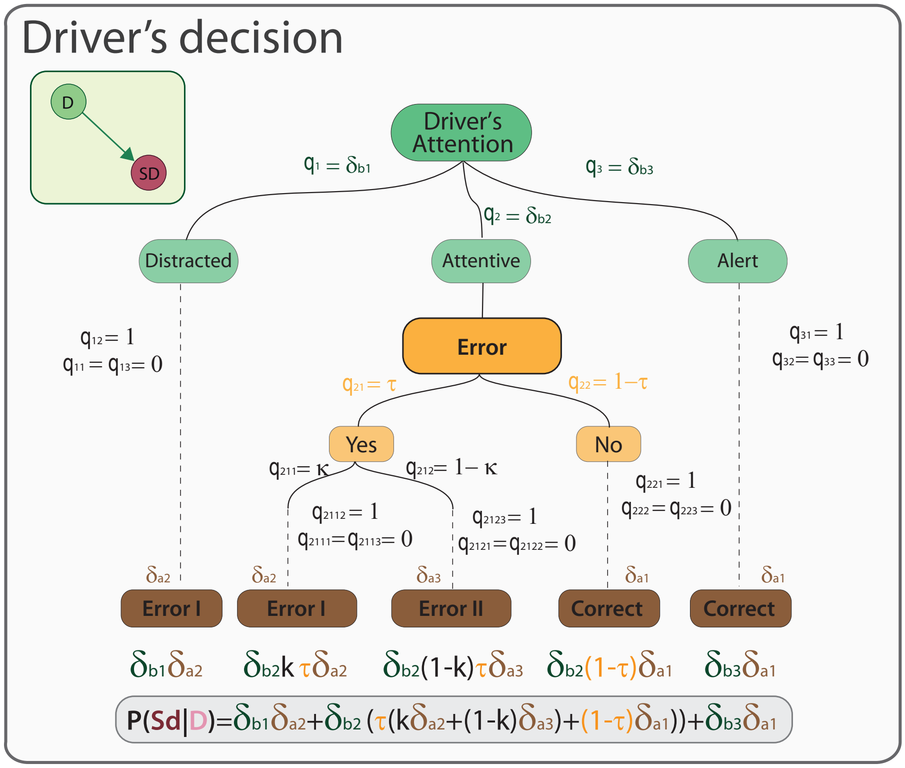


Figure 5.2: Illustration of the closed formula for the driver's speed decision conditional probability table.

km/h), medium ($30 \leq S_p < 55$ km/h), high ($55 \leq S_p < 80$ km/h) and very high ($S_p \geq 80$ km/h), where the following threshold values for the speed S have been selected:

$$th(1) = 30; \quad th(2) = 55; \quad th(3) = 80; \quad th(4) = \infty. \tag{5.16}$$

Next, it is assumed that the speed S is a normal random variable with mean the actual speed $S = a$ corrected by a factor $\rho_D(c)$ and a coefficient of variation c_v, that is, a $N(S\rho_D(c), c_v S\rho_D(c))$.

With this, the conditional probabilities $p(a, b, c, d)$=$P(V = a|Vt = b, D = c, S = d)$ become:

$$\begin{aligned} p(a,b,c,d) \quad &= \quad Z_{b,c}\delta_{a1}\Phi\left(\frac{th(a) - d\rho_D(c)}{c_v d\rho_D(c)}\right) + Z_{b,c}\delta_{aa^*}\left(1 - \Phi\left(\frac{th(a-1) - d\rho_D(c)}{c_v d\rho_D(c)}\right)\right) \\ &\quad +Z_{b,c}(1-\delta_{aa^*}-\delta_{a1})\Phi\left(\frac{th(a)-d\rho_D(c)}{c_v d\rho_D(c)}\right) \\ &\quad -Z_{b,c}(1-\delta_{aa^*}-\delta_{a1})\Phi\left(\frac{th(a-1)-d\rho_D(c)}{c_v d\rho_D(c)}\right) + (1 - Z_{b,c})\delta_{a1}. \end{aligned} \tag{5.17}$$

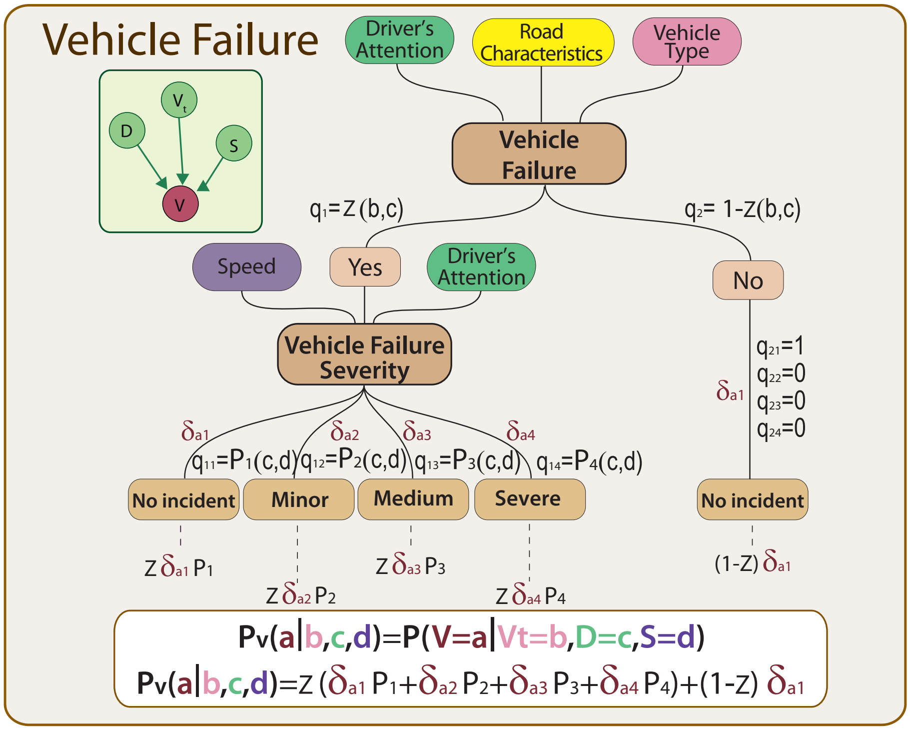


Figure 5.3: Illustration of the closed formula for the vehicle failure conditional probability table.

In the same way as in the previous subsection, the conditional probability of occurrence of a vehicle failure can be expressed in the form (5.6). In Figure 5.3 the q functions can be identified and Formula (5.18) can be obtained by simply adding the contributions of all branches.

$$\begin{aligned} P(V=a|Vt=b,D=c,S=d) &= Z(\delta_{a1}p_1+\delta_{a2}p_2+\delta_{a3}p_3+\delta_{a4}p_4) \\ &\quad +(1-Z)\delta_{a1}. \end{aligned} \tag{5.18}$$

5.3.7 Conditional probability of node Co collision

Since node Co has five parent variables, Vt, It, Vis, D and S, it is necessary to define the conditional probabilities $P(P=a|Vt=b, It=c, Vis=d, D=e, S=f)$ for any combination of the conditioning variables.

To this end, the probability Z of a collision is calculated first, assuming that there is a collision rate u_t per unit length, which depends on the track type, and on the variables traffic intensity It, visibility Vis and driver's attention D, and the speed excess over the maximum speed at the site, that is, the probability of failure Z is assumed to be of the form

$$Z(c,d,e,f) = Lu_t h_{It}(c) h_{Vis}(d) h_D(e) h_{ex}(f), \tag{5.19}$$

where L is the segment length, u_t is a failure rate dependent on the track type, $h_{It}(c)$, $h_{Vis}(d)$, $h_D(e)$ and $h_{ex}(f)$ are correction factors to consider the traffic intensity It, visibility Vis, driver's attention D and the speed excess, respectively.

Since given that a collision failure has occurred, the severity of the incident depends on its speed. Thus, as before, the following different speed cases are considered next: low ($S_p < 30$ km/h), medium ($30 \leq S_p < 55$ km/h) high ($55 \leq S_p < 80$ km/h) and very high ($S_p \geq 80$ km/h), where the following threshold values for the speed $S = f$ have been selected:

$$th(1) = 30; \quad th(2) = 55; \quad th(3) = 80; \quad th(4) = \infty. \tag{5.20}$$

Next, it is assumed that the speed S is a normal random variable with mean the actual speed $S = f$ corrected by a factor $w_{Vt}(b)$, which depends on the vehicle type $Vt = b$ and a coefficient of variation c_v, that is, a $N(fw_{Vt}(b), c_v fw_{Vt}(b))$.

With this, the conditional probabilities $p(a,b,c,d,e,f)$= $P(P = a|Vt = b, It = c, Vis = d, D = e, S = f)$ become (see figure 5.4):

$$Z(\delta_{a1}p_1 + \delta_{a2}p_2 + \delta_{a3}p_3 + \delta_{a4}p_4) + (1 - Z)\delta_{a1}, \tag{5.21}$$

where

$$p_1 = \Phi\left(\frac{th(a) - fw_{Vt}(b)}{c_v fw_{Vt}(b)}\right) \tag{5.22}$$

$$p_4 = 1 - \Phi\left(\frac{th(a-1) - fw_{Vt}(b)}{c_v fw_{Vt}(b)}\right) \tag{5.23}$$

$$p_a = \Phi\left(\frac{th(a) - fw_{Vt}(b)}{c_v fw_{Vt}(b)}\right) - \Phi\left(\frac{th(a-1) - fw_{Vt}(b)}{c_v fw_{Vt}(b)}\right) \tag{5.24}$$

for $a = 1, 2$.

Note again that the conditional probability (5.21) is of the form (5.6). In Figure 5.4 the q functions can be identified and Formula (5.21) can be obtained by simply adding the contributions of all branches.

5.3.8 Conditional probability of node P pavement failure

This node P has six parent variables, W, Vt, It, Vis, D, and S. In this way, the conditional probabilities $P(P = a|W = b, Vt = c, It = d, Vis = e, D = f, S = g)$ for any combination of the conditioning variables need to be defined. For this, the probability Z of pavement failure is calculated first, assuming that there is a failure rate d_t per unit length, which depends on the track type, and also on the variables weather W, traffic intensity It, visibility Vis, and driver's attention D, that is, the probability of failure Z becomes:

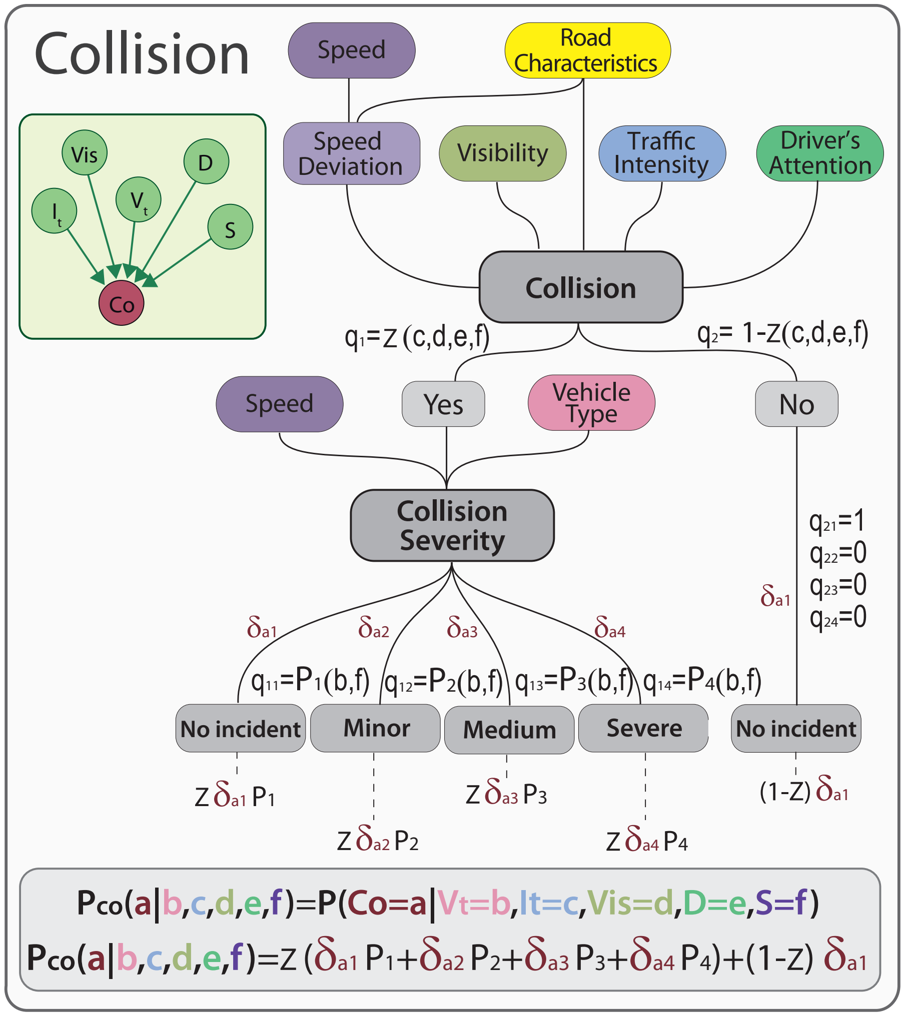


Figure 5.4: Illustration of the closed formula for the collision conditional probability table.

$$Z_{b,d,e,f} = Ld_t n_W(b) n_{It}(d) n_{Vis}(e) n_D(f), \tag{5.25}$$

where L is the distance between the previous and the actual locations, d_t is a failure rate dependent on the track type, $n_W(b)$, $n_{It}(d)$, $n_{Vis}(e)$, and $n_D(f)$ are correction factors to consider the weather W, the traffic intensity It, the visibility Vis, and the driver's attention D, respectively.

Since given that a pavement failure has occurred, the severity of the incident depends on its speed. In this way, as in the previous cases, the following different speed cases are considered: low ($S_p < 30$ km/h), medium ($30 \leq S_p < 55$ km/h), high ($55 \leq S_p < 80$ km/h) and very high ($S_p \geq 80$ km/h), where the following threshold values for the speed

S have been selected:

$$th(1) = 30; \quad th(2) = 55; \quad th(3) = 80; \quad th(4) = \infty. \tag{5.26}$$

Then, it is assumed that the speed S is a normal random variable with mean the actual speed $S = g$ corrected by a factor $w_{Vt}(b)$, which depends on the vehicle type $Vt = b$ and a coefficient of variation c_v, that is, a $N(fw_{Vt}(b), c_v fw_{Vt}(b))$.

With this, the conditional probabilities $p(a, b, c, d, e, f, g)=$
$P(P = a|W = b, Vt = c, It = d, Vis = e, D = f, S = g)$ become (see figure 5.5):

$$Z(\delta_{a1}p_1 + \delta_{a2}p_2 + \delta_{a3}p_3 + \delta_{a4}p_4) + (1 - Z)\delta_{a1}, \tag{5.27}$$

where

$$p_1 = \Phi\left(\frac{th(a) - fw_{Vt}(c)}{c_v fw_{Vt}(c)}\right) \tag{5.28}$$

$$p_4 = 1 - \Phi\left(\frac{th(a-1) - fw_{Vt}(c)}{c_v fw_{Vt}(c)}\right) \tag{5.29}$$

$$p_a = \Phi\left(\frac{th(a) - fw_{Vt}(c)}{c_v fw_{Vt}(c)}\right) - \Phi\left(\frac{th(a-1) - fw_{Vt}(c)}{c_v fw_{Vt}(c)}\right) \tag{5.30}$$

for $a = 1, 2$.

Being the conditional probability of the form (5.6) and being identified the q functions in Figure 5.5, Formula (5.27) can be obtained by adding the contributions of all branches.

5.3.9 Conditional probability of node incident I at a curve

In this case we have the conditional probability $P(I = a|W = b, Vt = c, D = d, S = e)$. However, to calculate these probabilities it is necessary to analize the curve stability problem first.

Curve stability

When vehicles circulate along curves they are subject to several types of incidents. The forces acting on the vehicles are the centrifugal force, F_c, its own weight, mg, the forces normal to the wheels, N_1 and N_2 and the friction forces, transversal to the wheel contacts with the pavement, R_1 and R_2 (see Figure 5.6). To increase safety at curves cambers are used, that is, banked roads.

The sliding and the overturning of the vehicle are considered as possible failure modes. The sliding critical condition (see figure 5.6) is given by:

$$v^2 = rg(p \pm f_t), \tag{5.31}$$

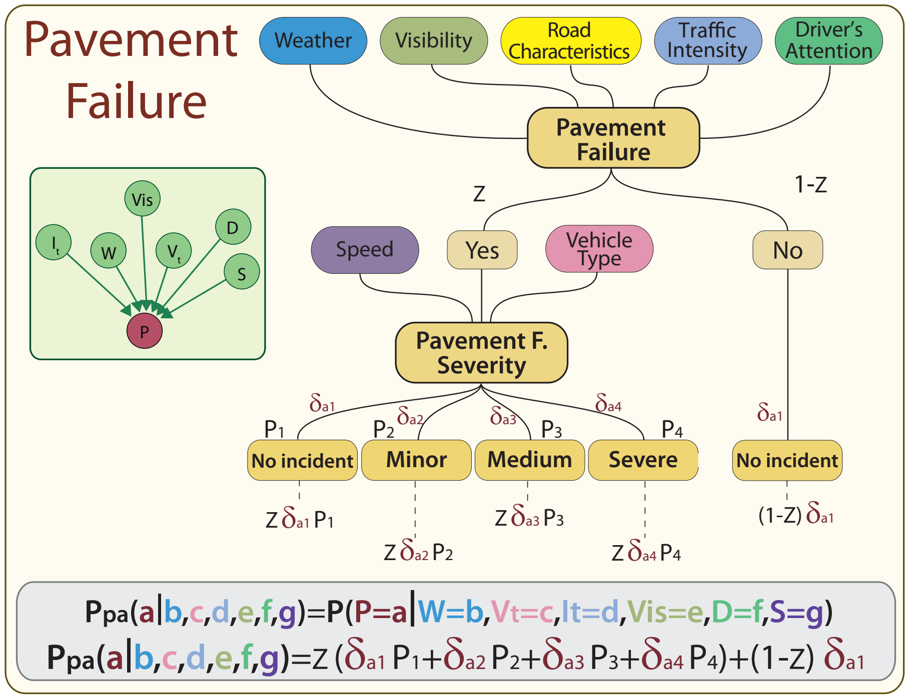


Figure 5.5: Illustration of the closed formula for the pavement failure conditional probability table.

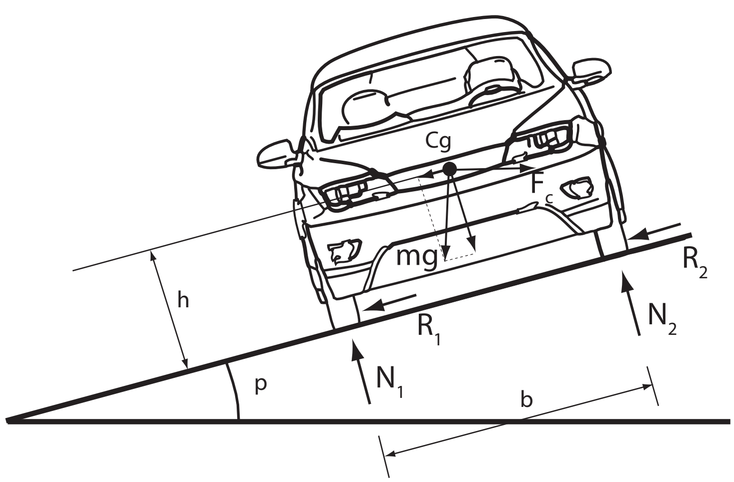


Figure 5.6: Illustration of all forces acting on a vehicle at a curve.

where v is the vehicle speed, r is the curve radius, g is the gravity acceleration, f_t is the friction coefficient, p is the slope and the plus and minus signs refer to outward or inward directions of the slide.

The overturning critical condition (see figure 5.6) is given by:

$$\frac{b}{2}(N_2 - N_1) = \pm h(R_1 + R_2) = \pm f_t h(N_1 + N_2), \tag{5.32}$$

where b is the width of the track, h is the height of the vehicle gravity center with respect to the pavement and the plus and minus signs refer to outward or inward directions of the overturning.

In fact, there are four types of failure modes:

1. *Outward slide:* When the vehicle slides in the outward direction of the curve due to insufficient friction or slope, to a small curve radius or to a high vehicle speed. This failure occurs when
$$v^2 > rg(p + f_t). \tag{5.33}$$

2. *Inward slide:* When the vehicle slides in the inward direction of the curve due to insufficient friction, to a high slope or to a low vehicle speed. This failure occurs when
$$v^2 < rg(p - f_t). \tag{5.34}$$

3. *Outward overturning:* When the vehicle overturns in the outward direction of the curve due to insufficient width of the track or to a high location of the gravity center. This failure occurs when $N_1 = 0$, that is, when
$$f_t > \frac{b}{2h}. \tag{5.35}$$

4. *Inward overturning:* When the vehicle overturns in the inward direction of the curve due to high width of the track or to a high location of the gravity center. This failure occurs when $N_2 = 0$, that is, when
$$f_t > \frac{b}{2h}. \tag{5.36}$$

The design curve radius and the camber values as a function of the speed were defined using the Spanish code (3.1-IC).

Next, the critical speed v_{sl} against outside sliding is calculated, given by (5.31):

$$v_{sl} = \sqrt{rg(p + f_t)}. \tag{5.37}$$

If the actual speed v is smaller than v_{sl} sliding does not occur and a severity level 0 is considered, otherwise, the severity of the sliding is classified depending on the ratio $(v/v_{sl})^2$. If $1 < (v/v_{sl})^2 \leq 1.5$ a severity level 1 is considered, if $1.5 < (v/v_{sl})^2 \leq 2$ the severity level is 2 and if $(v/v_{sl})^2 > 2$ the severity level is 3.

Assuming a sliding at severity level ℓ, there can be an incident, with probability pp_ℓ, or not, with probability $1-pp_\ell$.

Next, to consider the effect of the vehicle type variable Vt on the incident severity, it is assumed that the severity is a normal random variable

$$Sev \sim N((v-v_{sl})f_{Vt(c)}, (v-v_{sl})f_{Vt(c)c_v(c)}),$$

that is, with mean equal to the speed excess over the critical one v_{sl} and a coefficient of variation $c_v(c)$, and we associate light, medium and severe incident levels with the intervals $Sev < 20Km/h$, $20 \leq Sev < 45Km/h$ and $Sev \geq 45Km/h$, respectively.

Finally, in the case of no sliding, the case of a driver's distraction leading to an incident with probability pp is considered too. Similarly, it is assumed that the severity is a normal random variable $Sev \sim N(vf_{Vt(c)}, vf_{Vt(c)}c_v(c))$ and we associate light, medium and severe incident levels to $Sev < 20Km/h$, $20 \leq Sev < 45Km/h$ and $Sev \geq 45Km/h$, respectively.

With all this, the conditional probability

$$P(I_c = a|W = b, Vt = c, D = d, S = e)$$

becomes:

$$\begin{aligned}
&\delta_{slid}\left((1-pp)\delta_{a1} + pp\left(\delta_{a2}\Phi\left(\frac{s(a-1)-(v-v_{sl})f_{Vt(c)}}{(v-v_{sl})f_{Vt(c)}c_v(c)}\right)\right.\right.\\
&\left.\left.+(1-\delta_{a1}-\delta_{a2})\left(\Phi\left(\frac{s(a-1)-(v-v_{sl})f_{Vt(c)}}{(v-v_{sl})f_{Vt(c)}c_v(c)}\right)-\Phi\left(\frac{s(a-2)-(v-v_{sl})f_{Vt(c)}}{(v-v_{sl})f_{Vt(c)}c_v(c)}\right)\right)\right)\right)\\
&+(1-\delta_{slid})\left((1-\delta_{d1})\delta_{a1} + \delta_{d1}\left((1-pp)\delta_{a1} + pp\left(\delta_{a2}\Phi\left(\frac{s(a-1)-vf_{Vt(c)}}{vf_{Vt(c)}c_v(c)}\right)\right.\right.\right.\\
&\left.\left.\left.+(1-\delta_{a1}-\delta_{a2})\left(\Phi\left(\frac{s(a-1)-vf_{Vt(c)}}{vf_{Vt(c)}c_v(c)}\right)-\Phi\left(\frac{s(a-2)-vf_{Vt(c)}}{vf_{Vt(c)}c_v(c)}\right)\right)\right)\right)\right).
\end{aligned} \tag{5.38}$$

In Figure 5.7 the formula for this conditional probability and how it is obtained are illustrated, where δ_{ex} is one if there is speed excess and zero, otherwise, α_1 and α_2 are the probabilities of an incident at the curve under no speed excess and speed excess, respectively, p_2, p_3 and p_4 are the probabilities of occurrence of a light, medium and severe incident, respectively, under an speed excess and once the incident is produced and p'_2, p'_3 and p'_4 are the probabilities of occurrence of a light, medium and severe incident, respectively, under no speed excess and once the incident is produced.

The probabilities α_1 and α_2 are calculated by the formula

$$\alpha_1(b,d,e) = \rho_1(b,d,e), \tag{5.39}$$

$$\alpha_2(b,d,e) = \rho_1(b,d,e) + \beta\rho_2(b,d,e)\left(1+\frac{v_{ex}(b,d,e)}{v_{crit}(b)}\right)^\gamma, \tag{5.40}$$

where ρ_1 and ρ_2 are the probabilities of incident at the curve under no speed excess and speed excess, respectively, and β and γ are two parameters to consider the relative difference of the speed excess v_{ex} with respect to the critical speed at the curve v_{crit}.

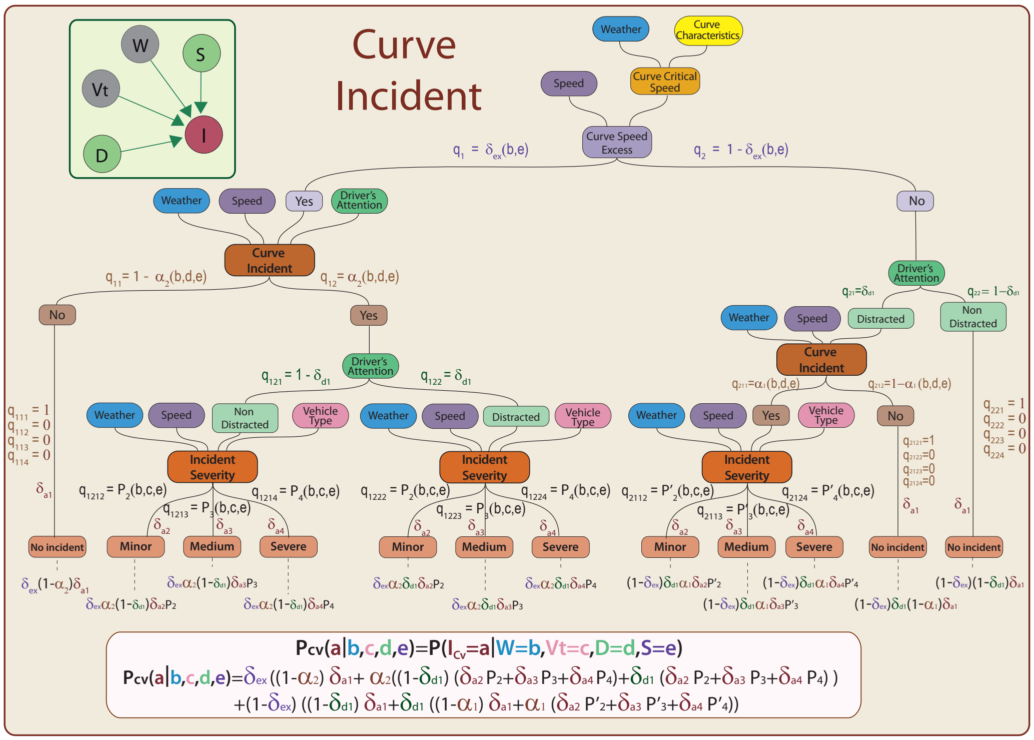


Figure 5.7: Illustration of the closed formula for an incident at a curve conditional probability table.

5.3.10 Conditional probability of node I_{t_s} incident at a traffic signal

Since node Is in some cases has five parent variables, W, Dri, SS, S and Tf, we need to define the conditional probabilities $P(Is = a|W = b, Dri = c, SS = d, S = e, Tf = f)$ for any combination of the conditioning variable.

This incident node includes the traffic light, stop, yield, grade and pedestrian crossing, for which we define a target speed $S_{target}(b, c, f)$ and an incident probability $\alpha_1(b, c, f)$, both dependent on the type and state of signal, the weather (b), the driver's type (c) and the possibility of a technical failure (f). For example, the target speed for a stop signal is zero and in a free light signal is the maximum speed of the road. Similarly, the incident probabilities for these two signals are the probabilities of having an incident when they are not respected.

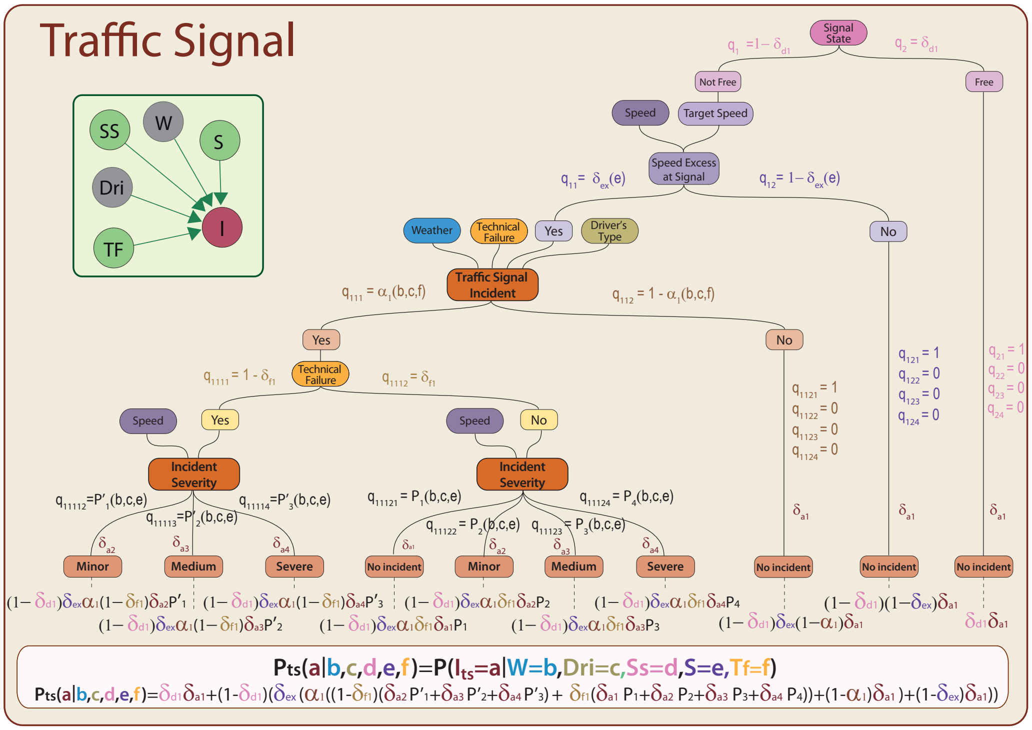


Figure 5.8: Illustration of the closed formula for an incident at a light signal conditional probability table.

Since the incident occurrence and its severity are related to the speed excess over the target speed S_{target} at the signal location and depends on the weather W and the driver type Dri, this speed is calculated and modified S^*_{target} as follows:

$$S^*_{ex}(b,c,e) = abs(e - S_{target}) f_W(b) f_{Dri}(c), \tag{5.41}$$

where $f_W(b)f$ and $f_{Dri}(c)$ are correction factors due to the W and Dri variables.

Then, it is assumed that if the signal is violated, the probability of an incident is α_1, and that the corrected excess speed with respect to the target is a normal random variable $N(S^*_{ex}, c_v S^*_{ex})$, where c_v is a coefficient of variation which depends on the S^*_{ex} level. In order to determine the severity of the incidence, three threshold values for the S^*_{ex}: $th = (10, 30, 60, \infty)$ km/h are considered, so that there is no incident if $S^*_{ex} \leq 10$ km/h, a light incident if $10 < S^*_{ex} \leq 30$ km/h, a medium incident if $30 < S^*_{ex} \leq 60$ km/h and a severe incident if $S^*_{ex} \geq 60$ km/h. This allows us to calculate the probabilities of I_{t_s} using the cdf of the normal distribution if no technical failure occurs (δ_{f_1}).

If a technical failure occurs $(1 - \delta_{f_1})$, we consider that an incident occurs and then the above probabilities are truncated, that is, the probability of $S^*_{ex} \leq 10$ km/h is distributed among the other cases.

With this, the conditional probabilities

$$P(Is = a|W = b, Dri = c, SS = d, S = e, Tf = f)$$

denoted $p(a, b, c, d, e, f)$ become (see figure 5.8):

$$\begin{aligned}&\delta_{d1}\delta_{a1} + (1 - \delta_{d1})\left(\delta_{ex}\left(\alpha_1\left(\delta_{f1}\left(\delta_{a1}p_1 + \delta_{a2}p_2 + \delta_{a3}p_3 + \delta_{a4}p_4\right)\right.\right.\right.\\&\left.\left.\left.+(1-\delta_{f1})\left(\delta_{a1}p'_1 + \delta_{a2}p'_2 + \delta_{a3}p'_3 + \delta_{a4}p'_4\right)\right) + (1 - \alpha_1)\delta_{a1}\right) + (1 - \delta_{ex})\delta_{a1}\right),\end{aligned} \tag{5.42}$$

where

$$th^*(a) = \frac{th(a) - th(1)}{1 - th(1)} \tag{5.43}$$

$$p_1 = \Phi\left(\frac{th(1) - S^*_{ex}}{c_v S^*_{ex}}\right) \tag{5.44}$$

$$p_4 = 1 - \Phi\left(\frac{th(a-1) - S^*_{ex}}{c_v S^*_{ex}}\right) \tag{5.45}$$

$$p'_4 = \left(1 - \Phi\left(\frac{th^*(a-1) - S^*_{ex}}{c_v S^*_{ex}}\right)\right) \tag{5.46}$$

and for $a = 2, 3$ we have

$$p_a = \left(\Phi\left(\frac{th(a) - S^*_{ex}}{c_v S^*_{ex}}\right) - \Phi\left(\frac{th(a-1) - S^*_{ex}}{c_v S^*_{ex}}\right)\right) \tag{5.47}$$

$$p'_a = \left(\Phi\left(\frac{th^*(a) - S^*_{ex}}{c_v S^*_{ex}}\right) - \Phi\left(\frac{th^*(a-1) - S^*_{ex}}{c_v S^*_{ex}}\right)\right) \tag{5.48}$$

Chapter 6

Network partition and representation

Contents

6.1 Introduction

As it has been seen, each and every one of the elements involved in the safety of the road are reproduced in the model. Each element involves a series of variables that generate subnetworks that are assembled to define the global Bayesian network that constitutes the model of probabilistic safety of the highway or road. The analysis of the dependencies between the variables is very important, especially the behaviour of the driver. In this way, a structure that allows to quantify the frequency of occurrence of incidents along the road considering all its elements together is obtained. In Figure 6.1 the dependence relations among the variables are represented.

With this methodology, whose most relevant steps are shown in Figure 6.2, it is possible to perform a joint scanner of the road (examples of scanners of twoo different segments of a road can be seen in Figure 6.3), identify the points where accidents are most likely

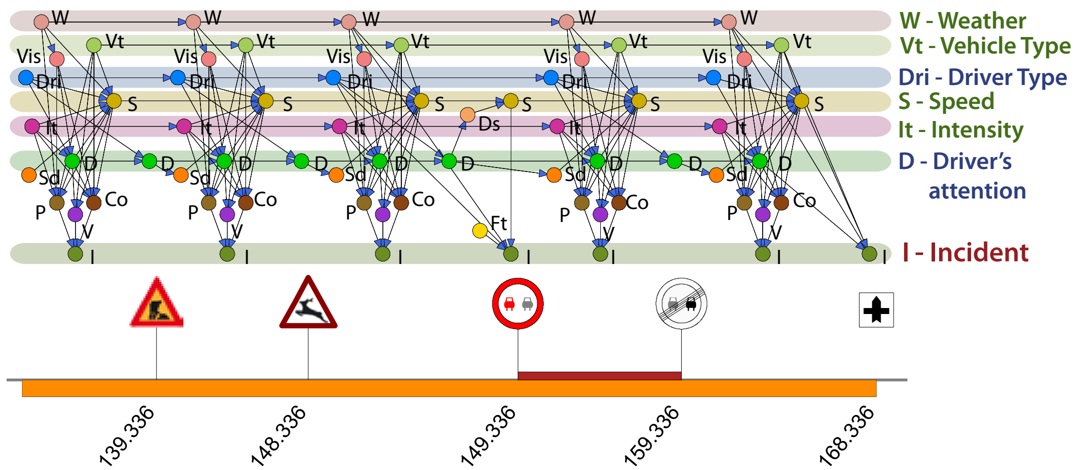


Figure 6.1: Example of how the dependence of variables is reproduced in the Bayesian network.

and most dangerous of the road, quantify the risks and obtain the list of the most likely incidents. In this way, special attention can be given to the critical points to avoid them in order to improve safety and optimize improvement actions, defining the solutions and analyzing their benefits.

The analysis of the hazardous points of roads is done through the probabilistic risk analysis, using the concept of expected number of equivalent severe incidents (ENSI), in which taking as reference the associated social costs it is considered that: 230 minor incidents are equivalent to 1 severe incident and 6.4 medium incidents amounted to 1 severe incident (these values are represented in Figure 6.4). ENSI values are used to determine the most risky points of the road and to evaluate their level of safety. The points that generate a major risk by individually producing an elevated ENSI can lead to severe incidents with a high frequency. It is estimated that points with an ENSI greater than $1e-9$ should be improved, and those where the probability of severe incidents exceeds the values $1e-7$, $1e-6$ or $1e-5$ should be remedied with an increasing degree of celerity.

The results are obtained for the whole road, however, the line representation has been divided into short segments, with the aim of obtaining a more detailed information. Furthermore, for the propagation of uncertainty, a specific partition of the Bayesian network is made, which is explained in Section 6.2. In Section 6.3 the software development of the model is described and the provided information explained and shown in more detail.

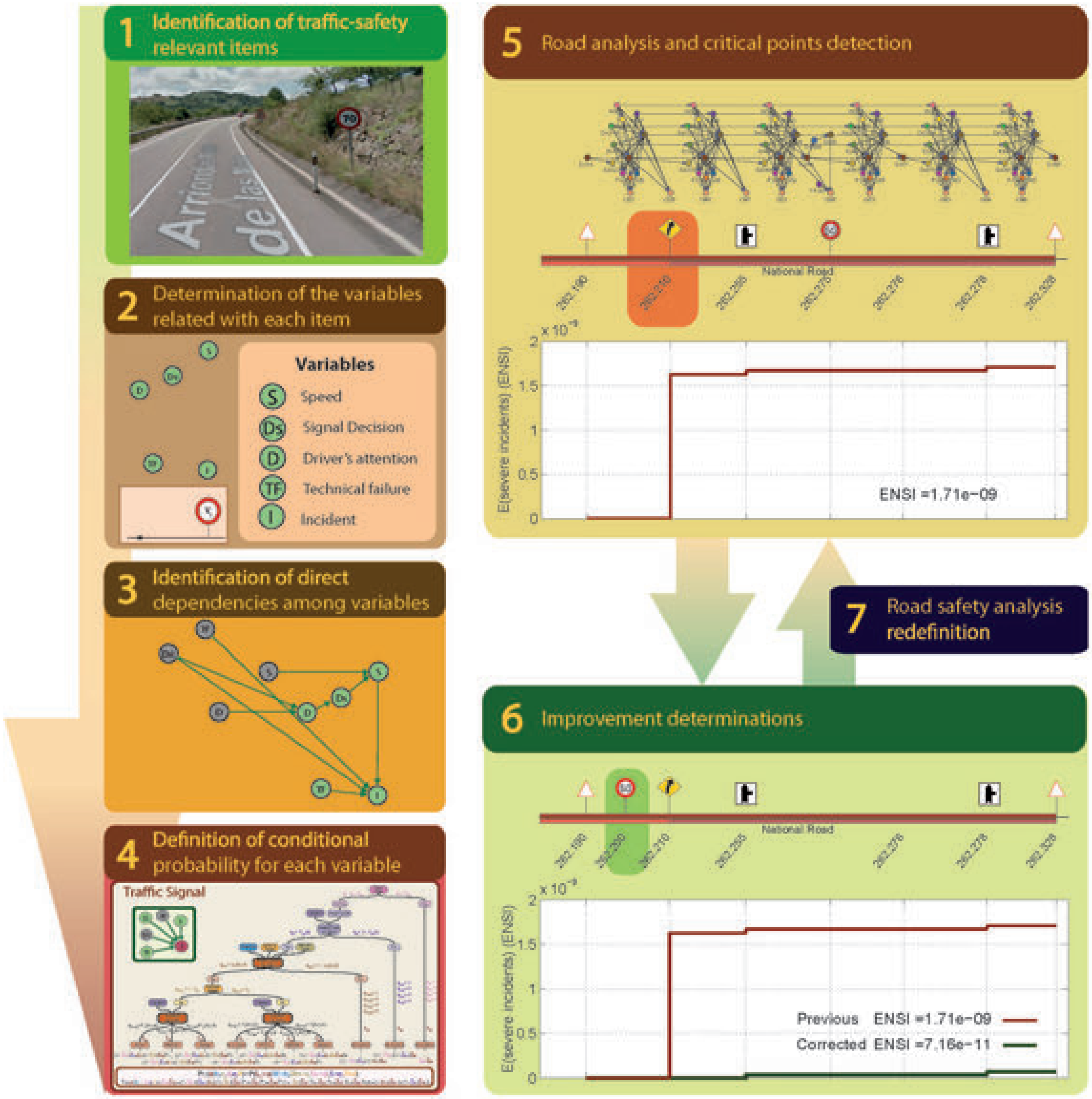


Figure 6.2: Main steps of the proposed methodology.

6.2 Network partition

Since Bayesian networks (BNs) associated with real cases imply a very high number of variables (many thousands), it is necessary to reduce the complexity of the calculations; otherwise, the memory and the CPU requirements will exhaust the computer capacity. Thus, some solution to this problem must be found if the proposed method is to be applicable to very large networks.

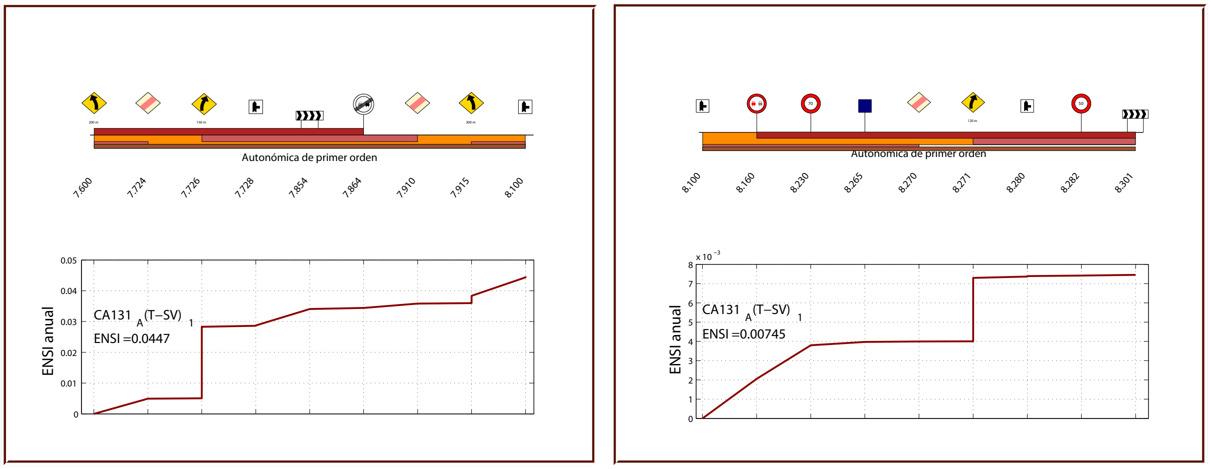


Figure 6.3: Examples of scanners of different segments of a road.

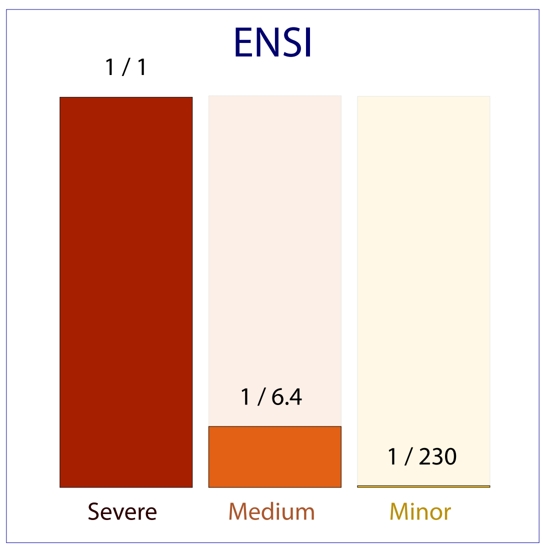


Figure 6.4: Illustration of how the expected number of equivalent severe incidents (ENSI) is obtained.

To solve this problem, the Bayesian network (BN) is partitioned into a series of, as small as possible subnetworks, such that the results of the computations are not modified and 100 % equivalent. This is the aim of the method described below. In fact the subnetworks will contain a very reduced number of variables (in the neighbourhood of 20 variables each).

Figure 6.5 shows an example, in which the acyclic graph of the BN of a segment of a road is given (upper plot). It has been partitioned into three subnetworks (lower plot), denoted A, B and C, that can be identified by their different background colors.

The partitions are selected based on the conditional independence property shown in Figure 6.5, where it can be seen that the nodes in set C are independent of the nodes of set A given the nodes in B, because any path from set A to set C passes by set B in the moral graph of the set $A \cup B \cup C$ and its ancestors (see Castillo et al. (1997)). This means that the variables in B contain all the information the variables in A have on the variables in C. Consequently, variables in A are not needed to get information on the variables in C given the variables in B. This is some kind of a Markov property, which is not an assumption but a consequence. In fact, the partitioning technique is based on finding parts of the network satisfying this assumption.

Let $\{A_1, A_2, \ldots, A_n\}$ be one of such a sequence and let $\{A_k^1, A_k^2\}$ be a partition of the nodes in subnetwork A_k. We look for partitions such that the nodes of A_{k+1} become independent of the nodes of $A_1, A_2, \ldots, A_{k-1}$ and A_k^1 given the nodes of A_k^2. This implies that the information on the nodes in A_{k+1} contained in the nodes in $A_1, A_2, \ldots, A_{k-1}$ and A_k^1 is already contained in the nodes of A_k^2. The set of nodes in A_k^2 is called *a separator* of the initial BN and they must be included in the two adjacent subnetworks. Then, the selected partition is composed by the sequence of Bayesian subnetworks:

$$A_1^* = A_1, A_2^* = \left\{A_1^2, A_2\right\}, \ldots, A_n^* = \left\{A_{n-1}^2, A_n\right\}.$$

It can be easily shown that Bayesian subnetwork A_{k+1} is independent of the nodes in BN $\{A_1, A_2, \ldots, A_k^1\}$, given the nodes in the separator A_k^2 for $k = 1, 2, \ldots, n-1$. This can be done using the graphical method given in Castillo et al. (1997) pages 181–184.

The lower plot in Figure 6.5 shows the partitions associated with the BN in the upper plot in Figure 6.5 and the separator (middle figure). Note that the separator nodes have been duplicated and included in both subnetworks because they must belong to both subnetworks to transmit the required information throughout the separator nodes. The corresponding subnetworks are shown with all its nodes and links. In particular, A and C contain nodes $W271$ to $I300$ and $D301$ to $D314$, respectively, and the separator B contains nodes $W289$ to $It292$, $D294$ and $S296$.

A corresponds to the two subnetworks A_1^* and A_2^* associated with the first partition. Note that the six artificial nodes in $B = A_1^2$ have been duplicated from the previous subnetworks and added to the second subnetwork. We note that some software packages provide only marginals of nodes involved in a clique.

The selected partition is not arbitrary at all. The key property for a partition to be valid is to contain a set of separators (subsets of nodes) such that the conditional probability of the posterior nodes becomes independent on the previous nodes given the separator subset. Consequently, the separator subset and the partitions have been selected to satisfy this condition.

Note that the separator B when marrying the parents with common children becomes a complete set, that is, all pairs of nodes are connected, as indicated in the lowest B set, where the discontinuous lines correspond to married parents. This means that the joint

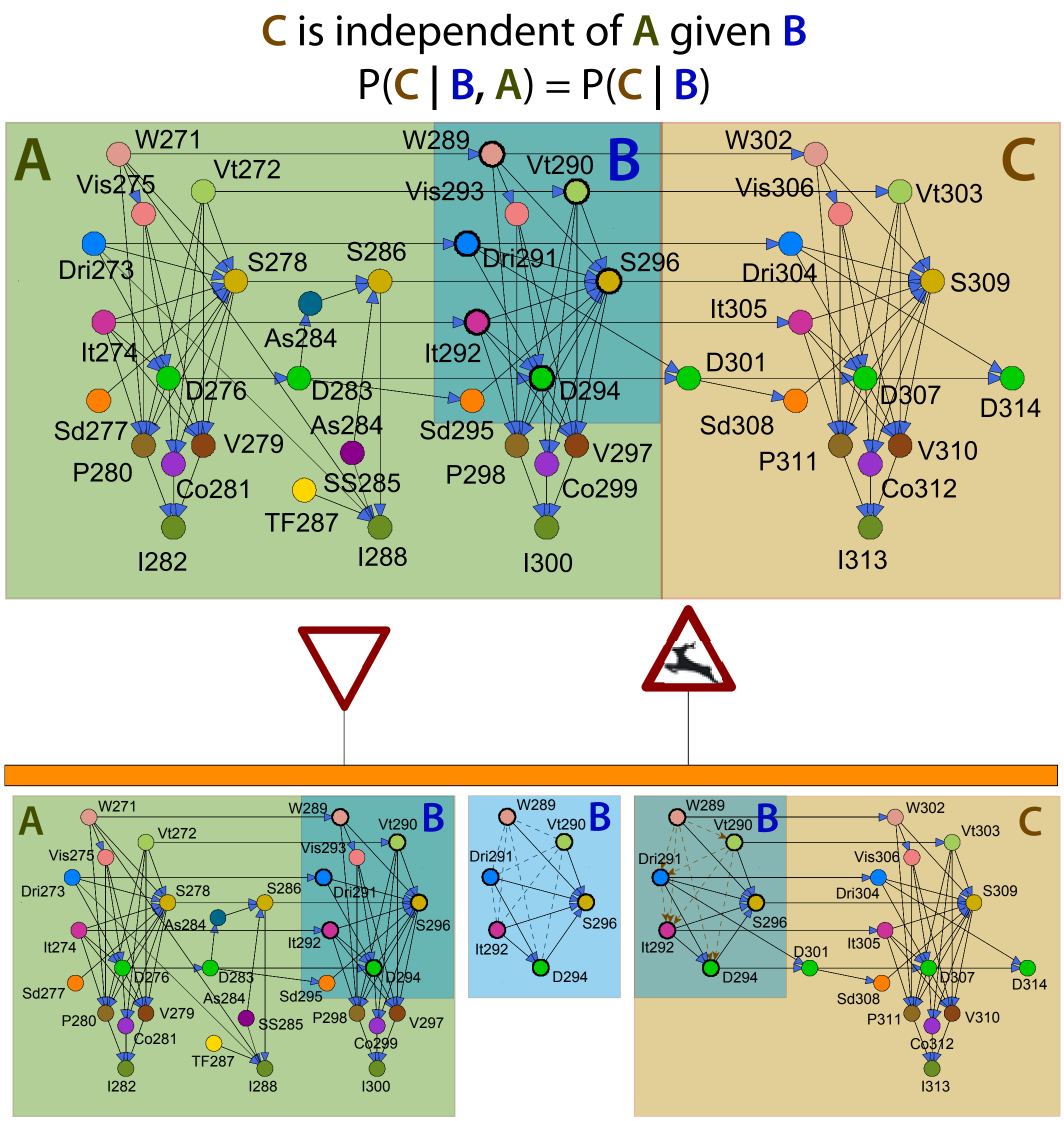


Figure 6.5: Illustration of how a BN can be partitioned into a sequence of Bayesian subnetworks to obtain the marginal probabilities (forward process).

probability of the separator nodes can be obtained from the first partition by the BN software.

In addition, the artificial links $W289-D294$, $Vt290-D294$, $W289-It292$, $Vt290-It292$, $Dri291-It292$, $W289-Dri291$, $Vt290-Dri291$ and $W289-Vt290$ have been added to the separator in A_1^2 in the right partition for it to become a clique, so that any joint probability, in particular the one obtained from the left partition can be assigned to the clique.

Using this partition methodology and its important independence properties, the marginal densities of the nodes in the initial BN can be obtained using the following process:

Step 1. Obtain the joint density of the separator. The joint density of the separator $B = A_k^2$ is obtained using some standard methods for BNs. Since the separator has been forced to be a complete set, its joint probability function $P(W, Vt, Dri, It, D, S)$ can be easily obtained without a high computational cost. This is not true for joint probabilities of nodes not in a complete set.

Step 2. Obtain the conditional probabilities of the nodes in the separator set. For the separator with nodes W, Vt, Dri, It, D and S, the conditional probabilities $P(S|D, It, Dri, Vt, W)$, $P(D|It, Dri, Vt, W)$, $P(It|Dri, Vt, W)$, $P(Dri|Vt, W)$, $P(Vt|W)$, and $P(W)$ are calculated based on the previous joint probability using the conditional probability definition, that is,

$$P(Vt|W) = \frac{P(Vt, W)}{P(W)};\ P(Dri|Vt, W) = \frac{P(Dri, Vt, W)}{P(Vt, W)} \dots \tag{6.1}$$

These probabilities are needed to be transferred to the following partitions.

Step 3. Build the BN A_k^*. The BN A_k is built considering not only all its nodes and links but the separator nodes and some extra links to allow for reproducing the dependence structure in the separator node, which is exported from the previous sub-network. The added links are used to convert the separator into a clique (saturated set) so that any joint probability can be passed to A_k from A_{k-1}.

This partition procedure leads to a computation time linear in the length of the road or highway. Consequently, the CPU times are reduced substantially. This means that some small cases (with 7,000 variables) requiring some hours of CPU, could be calculated in a few minutes and this time reduction is even more impressive for larger networks.

6.3 Software development

The computer software developed by our group to implement the model proposed has been written in Matlab with calls to Latex, JavaBayes and BNT software.

It requires as data input a sequential description of all the items encountered when travelling along the highway.

Each line of the code refers to one single item, which can be any of the item types described in Table 3.1. Apart from the item type it is necessary to provide its location (Kilometer Point, KP) and in some cases some extra information (see examples below).

In addition to the highway description, the parameter values used in the calculations need to be given. Most of them are parameters used to calculate the conditional probability matrices of each node given its parents. The list of parameters is not provided here because it is a long list and their estimation and validation are explained in the next chapter.

The program has been done taking into account the desired information and in such a way that it is easy to understand it. Thus, the detailed scheme of the appearance received for a fraction of a road is represented in Figure 6.6, where A) is the segment acyclic graph of the Bayesian network, B) the graphical representation of traffic signs and track elements, C) the segment characteristics, and D) a cumulated risk chart.

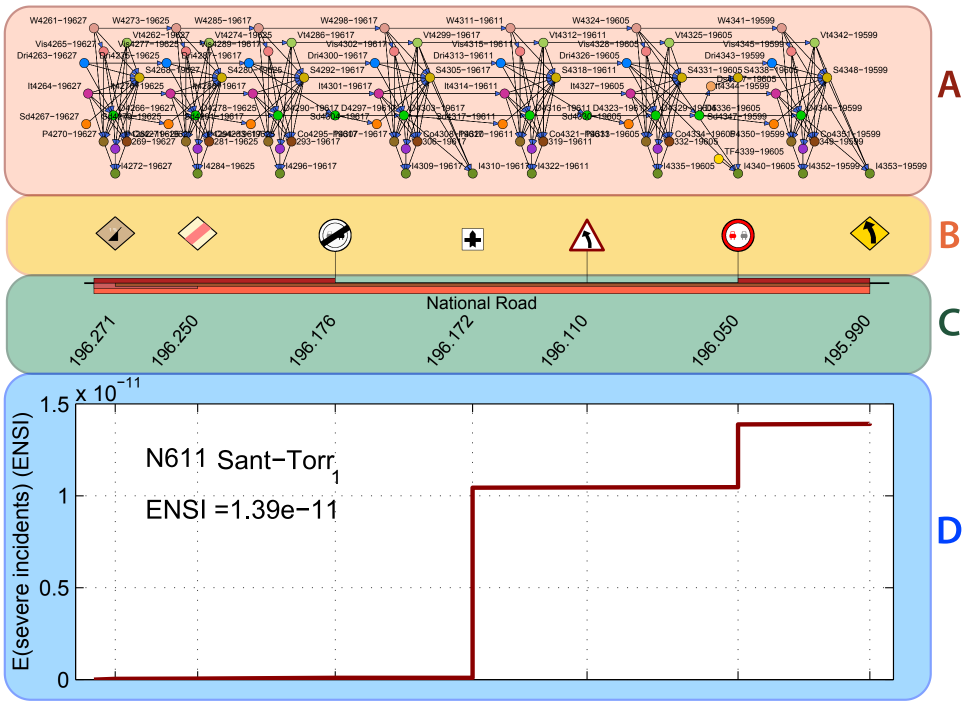


Figure 6.6: Example of the information supplied by the computer program.

The computer program developed is able to solve the following problems:

1. *Checks the given data for errors.* The code scans the lines looking for errors or inconsistencies (for example a 'TunnelIn' line without a 'TunnelOut' line or vice versa).

2. *Builds the acyclic graph of the Bayesian network.* This means that all variables involved in the problem are constructed and the direct dependencies are identified, that is, all directed links between two nodes.

3. *Builds the conditional probability tables.* These tables are built using the closed formulas derived for these conditional probabilities, using the directed connections of the acyclic graph and the provided parameter values.

4. *Calculates the incident probabilities.* This means that it calculates the marginal probabilities of all nodes of type 'incident'.

5. *Evaluates the expected number of equivalent severe incidents (ENSI) of incidence nodes.* The ENSI values are evaluated for all nodes of type incident.

6. *Provides a table of the ENSI frequencies for all items.* Obtains a table with the ENSI frequencies of all items in their KP sequence.

7. *Sorts ENSI values by importance.* Provides a table with the ENSI frequencies of all items in decreasing order of importance.

8. *Provides the sorted list of expected number of incidents.* Provides the expected number of incidents sorted by total associated ENSI, in decreasing order.

9. *Plots the segments Bayesian subnetworks.* Returns a plot (in pdf) for each segment of the highway or road corresponding to a given number of items. This plot includes: (a) the acyclic subgraph associated with the segment, (b) a graphical representation of the highway or road, showing the different items represented by their icons (signal icons, etc.) and the corresponding KP locations, and (c) a plot of the cumulated ENSI, showing graphically where the unsafest items are (see the examples below).

10. *Gives the 'JavaBayes' code.* A set of files are given with the code necessary to run the 'JavaBayes' software of the whole Bayesian network model or each of the segments. This means that we have direct access to the possible values of any node and to the conditional probability table of any node given its parents. In addition, backward analysis can be done, by providing any type of evidence and recalculating the corresponding marginal probabilities of any set of nodes.

11. *Gives the 'BN' code.* A set of files are given with the code necessary to run the 'BN' matlab software of the whole Bayesian network model or any Bayesian subnetwork of the segments.

12. *Provides a report file.* A pdf file with all the plots and tables above indicated is given.

The following examples provide some figures where is possible to see how the program works and what can be obtained from it.

6.3.1 Tabular information

With the aim of clarifying the information that can be obtained from the program, several tables with a brief explanation of what is shown are presented, thus, in the following examples it is explained exactly what is being represented.

For location of the most risky items a plan site of the road with the 5 most critical points for Natinal road N-611 is given in Figure 6.7, with Table 6.1 in which the hazardous rank, KP, item type, node and ENSI values of each item are represented.

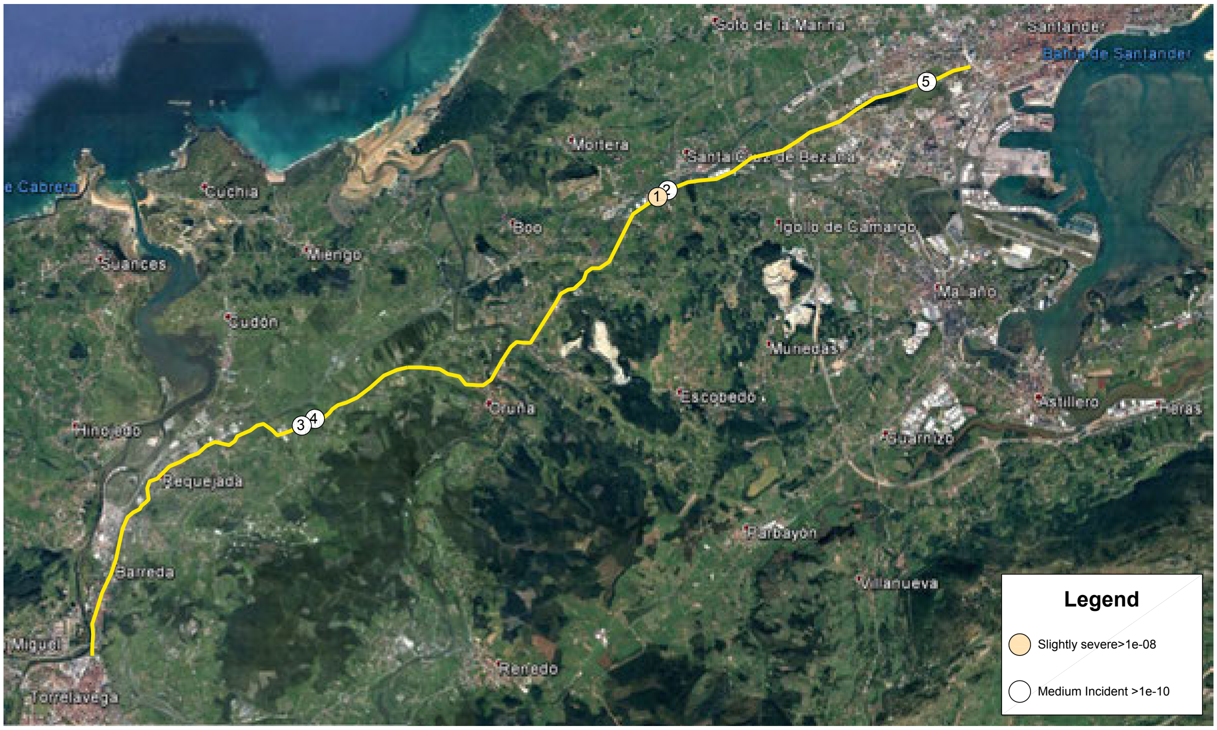


Figure 6.7: Plan site of the National road N-611 with the 5 most critical points.

Table 6.1: Representation of the hazardous rank, KP, item type, node and ENSI values of each item.

Rank	Item	Item name	KP	Node	Local ENSI
1	341	CurveIn	199.300	I4613-199300Cv	1.01e-08
2	346	CurveIn	199.500	I4676-199500Cv	8.9e-09
3	163	CurveIn	191.380	I2245-191380Cv	6.83e-09
4	173	CurveIn	191.630	I2370-191630Cv	4.61e-09
5	472	Yield	204.229	I6397-204229Yl	2.67e-09

Table 6.2 provides the information of the elements within its segment. It represents the following information of each item, sorted by location order, item number, KP, name, incident node identification and their corresponding local and accumulated ENSI values.

Once the segments of the whole line are displayed individually, Table 6.3 is generated, which presents the information of all items sorted by location and represents the joint information of the above individual segment tables.

With the aim of facilitating the safety assessment and management, Table 6.4 represents only the items whose local risk (local ENSI) excesses a given threshold. The items are still sorted by location.

Table 6.2: Information of the elements with their local and accumulated ENSI values within its segment.

				ENSI	
Item	**KP**	**Item name**	**Node**	**Local**	**Cumulated**
1	208.000	Initial	Start	0	0
2	207.995	SegWSignals	I19-207995S	7.39e-14	7.39e-14
		TrafficLight	I25-207995TL	1.2e-12	1.28e-12
3	207.935	SegWSignals	I37-207935S	3.68e-13	1.64e-12
		LateralEntry	I38-207935LE	7.46e-11	7.63e-11

Table 6.3: All items sorted by location representing the joint information of the previous individual segment tables.

				ENSI	
Item	**KP**	**Item name**	**Node**	**Local**	**Cumulated**
1	208.000	Initial	Start	0	0
2	207.995	SegWSignals	I19-207995S	7.39e-14	7.39e-14
		TrafficLight	I25-207995TL	1.2e-12	1.28e-12
3	207.935	SegWSignals	I37-207935S	3.68e-13	1.64e-12
		LateralEntry	I38-207935LE	7.46e-11	7.63e-11
4	207.915	SegWSignals	I50-207915S	1.4e-13	7.64e-11
		Intersection	I51-207915Int	2.68e-11	1.03e-10

In Table 6.5 the riskiest items sorted by risk level (the first is the most hazardous point) are given. It shows: the item rank, the item number, the item name, the item KP, the node and its ENSI value.

In addition, the total potential ENSI values per year considering the existing traffic are generated. This is an estimate of the potential number of severe incidents per year, at each location along the road.

With the purpose of obtaining a brief assessment of all the elements along a road, Table 6.6 represents the cumulated ENSI by type of incident, with the numbers of times that they appear in the network. They are sorted by decreasing risk level order.

Table 6.4: Items whose local risk (local ENSI) excesses a given threshold sorted by location.

				ENSI	
Item	**KP**	**Item name**	**Node**	**Local**	**Cumulated**
5	207.900	CurveIn	I64-207900Cv	6.63e-09	6.73e-09
24	207.160	Intersection	I329-207160Int	1.37e-10	7.22e-09
55	205.354	Yield	I774-205354Yl	1.25e-09	8.53e-09
62	205.235	Yield	I878-205235Yl	1.66e-09	1.03e-08
69	205.005	LateralEntry	I974-205005LE	1.12e-10	1.06e-08

Table 6.5: Riskiest items sorted by risk level and total potential ENSI values per year.

					ENSI	ENSI per year
Rank	**Item**	**Item name**	**KP**	**Node**	**Local**	**Local**
1	440	CurveIn	190.390	I5963-190390Cv	3.44e-07	0.565
2	5	CurveIn	207.900	I64-207900Cv	6.63e-09	0.0109
3	449	Yield	189.945	I6083-189945Yl	3.46e-09	0.00568
4	89	Yield	204.567	I1256-204567Yl	2.7e-09	0.00444
5	177	Yield	201.724	I2489-201724Yl	2.59e-09	0.00425

Table 6.6: Cumulated ENSI by type of incident.

Incident types	**Frequency**	**Local ENSI**
Curve	35	3.52e-07
Yield	19	3.04e-08
Lateral Entry	79	6.9e-09
Intersection	97	1.07e-09
Roundabout	15	9.45e-10
Pedestrian Crossing	19	1.92e-10
Overpass	6	8.89e-11
Overtaking	13	3.86e-11
Underpass	4	3.67e-11
Traffic Light	29	2.84e-11
Speed Limit	36	1.73e-11
SegWSignal	585	1.39e-11
Viaduct	2	9.81e-12

6.4 Examples of application

6.4.1 A-67 Highway

In this example a A-67 highway segment of 27.7 Km, from KP 105.3 to KP 133.0, with an Average Daily Traffic (ADT) of $10,068$ vehicles is considered. A brief example of how the items are introduced in the program is shown below.

```
InitialRT='Highway';
IniMaxSpeed=120;
ItLink=[10068, 1850];
initialVt=[0.0044,0.8288,0.1669];

LinkTrip={{'Initial', 133.00 },
    {'DistractingWarning', 132.995, 'Autovia','Good'},
    {'Yield', 132.873, 0, 'Good'},
    {'DistractingWarning', 132.705, 'River', 'Good' },
    {'ViaductIn', 132.700, 0, 'Rio izarilla'},
    {'ViaductOut', 132.690, 0, 'Rio izarilla'},
    {'DistractingWarning', 132.400, 'Information', 'Good'},
    {'PermanentWarning',132.330, 'Animals','Good'},
    {'PermanentWarning', 132.260, 'Straight', 'Good'},
    ...
    {'DistractingWarning' 106.150, 'Mountain','Good','Palen'},
    {'CurveIn', 106.100, 675, 'R'},
    {'Overpass', 105.840, 'R',0},
    {'CurveOut', 105.300, 675}
    };
```

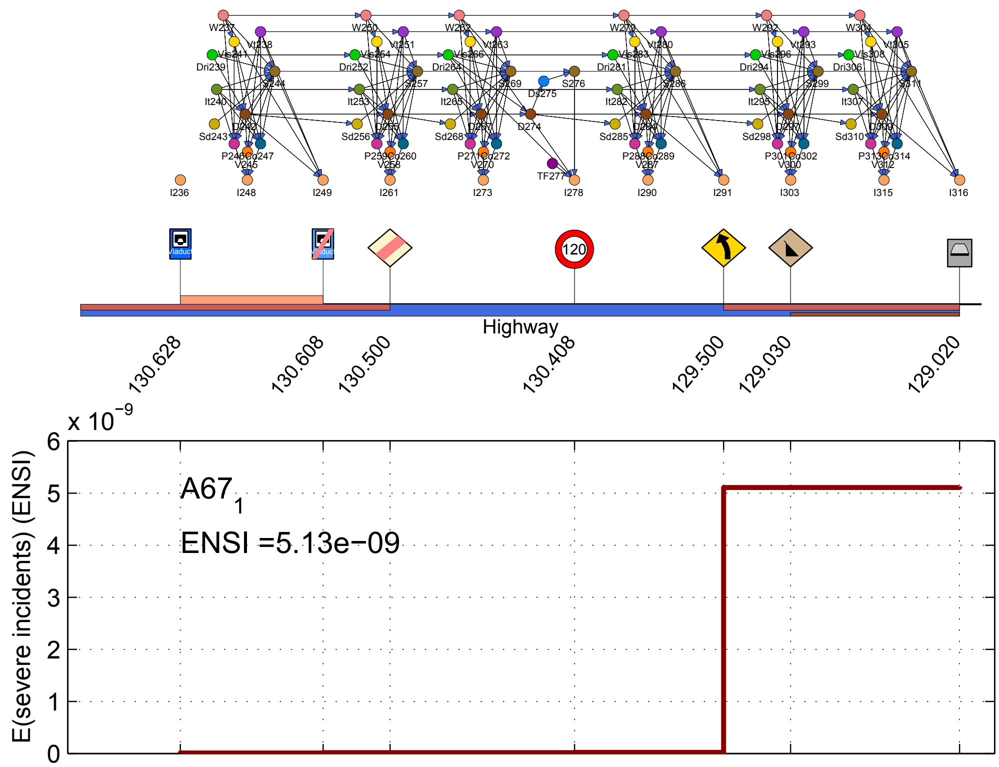
Figures 6.8 and 6.9 show: (a) the plots of the two acyclic sub-graphs corresponding to two segments in their upper parts; (b) the graphical representation of the highway or road segments, in their intermediate parts, and (c) the plots of the cumulated ENSI frequencies in their lowest parts. In addition, a table with all the items, their KP locations, their names and associated node names and the ENSI values (local and cumulated) associated with each item of the segments are provided.

The relative importance of the different items can be easily identified by comparing the discontinuities (jumps) of the graphs.

In order to improve safety at KP 129.500 and KP 125.590 curves, the speed limit at KP 130.408 has been changed from 120 to 100 km/h and added a speed limit signal (100 Km/h) at KP 125.690, respectively, as shown in Figure 6.10. The resulting improvements in the reduction of the ENSI values have been boldfaced in the last column of Table 6.7 which shows the item numbers and names, the corresponding KP locations, the node names and the local ENSI values of the initial (previous) and corrected line (after performing these two changes). It can be also easily seen that 'curves', 'acceleration lanes' and a 'yield' sign are the most critical items and the beneficial effect of these small changes can be appreciated since after correction no ENSI exceeds the value 1×10^{-9}.

Table 6.8 gives the same information as Table 6.7, but showing the annual frequencies, that is, considering the average yearly traffic.

Finally, it is indicated that this Bayesian network has 129 items and $1,704$ variables. The required CPU time to build the Bayesian network, calculate all incident probabilities,

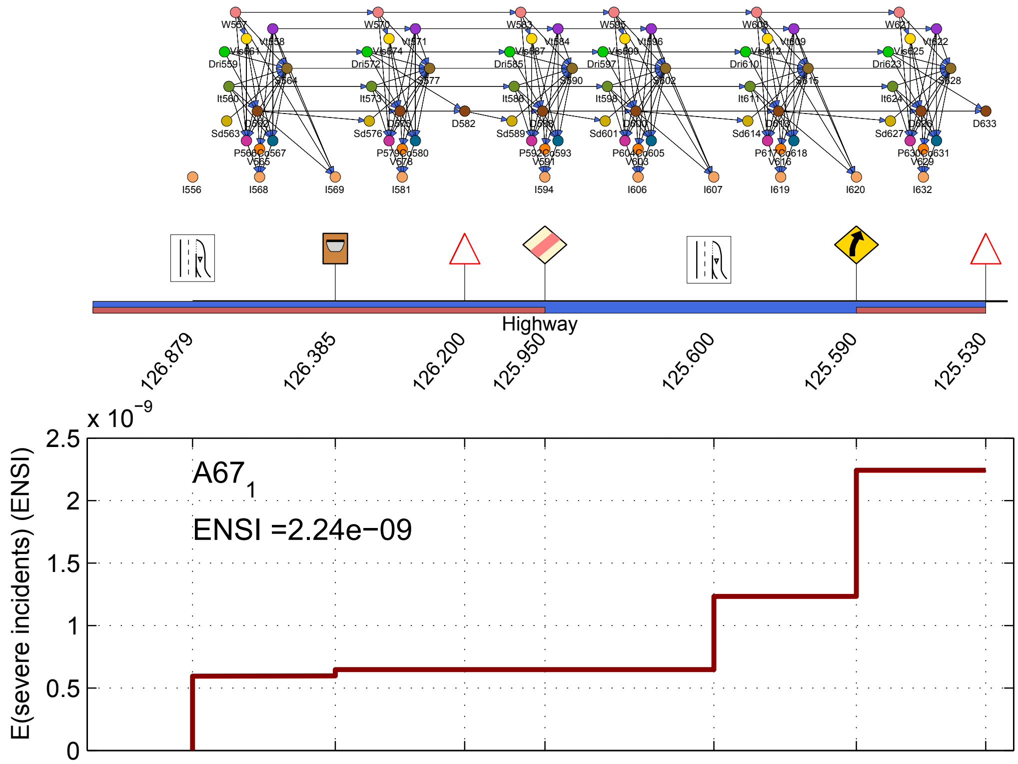


Item	KP	Item name	Node	ENSI Local	ENSI Cumulated
18	130.628	ViaductIn	I236-VI	5.95e-12	6.6e-10
19	130.608	SegWSignals	I248-SCr	2.98e-14	6.6e-10
		ViaductOut	I249-VO	5.96e-12	6.66e-10
20	130.500	SegWSignals	I261-SCr	1.54e-13	6.66e-10
21	130.408	SegWSignals	I273-S	1.3e-13	6.67e-10
		SpeedLimit	I278-Sl	6.55e-12	6.73e-10
22	129.500	SegWSignals	I290-S	9.46e-13	6.74e-10
		CurveIn	I291-Cv	5.08e-09	5.76e-09
23	129.030	SegWSignals	I303-SCr	5.79e-13	5.76e-09
24	129.020	SegWSignals	I315-SSC	1.36e-14	5.76e-09
		Overpass	I316-Ov	2.56e-11	5.78e-09

Figure 6.8: A-67 Highway example. Probabilities of the incident nodes between nodes 236 and 316.

plot all figures and write the report was 177 sec with an Intel(R) Core TM i7-4712HQ CPU @ 2.30GHz and 8.00 GB of RAM memory.

Thanks to the partitioning technique mentioned, the required CPU time is linear in the number of variables, with a mean number of 13 variables per item. Finally, note that



				ENSI	
Item	**KP**	**Item name**	**Node**	**Local**	**Cumulated**
42	126.879	AcelerationLane	I556-AL	5.95e-10	6.4e-09
43	126.385	SegWSignals	I568-SCr	4.12e-13	6.4e-09
		Underpass	I569-Un	5.02e-11	6.45e-09
44	126.200	SegWSignals	I581-SCr	1.76e-13	6.45e-09
45	125.950	SegWSignals	I594-SCr	1.74e-13	6.45e-09
46	125.600	SegWSignals	I606-S	3.13e-13	6.45e-09
		AcelerationLane	I607-AL	5.85e-10	7.03e-09
47	125.590	SegWSignals	I619-S	9.97e-15	7.03e-09
		CurveIn	I620-Cv	1.01e-09	8.05e-09
48	125.530	SegWSignals	I632-SCr	5.99e-14	8.05e-09

Figure 6.9: A-67 Highway example. Probabilities of the incident nodes between nodes 556 and 633.

changing conditions implies only changes in the parameter values but not in the number of variables.

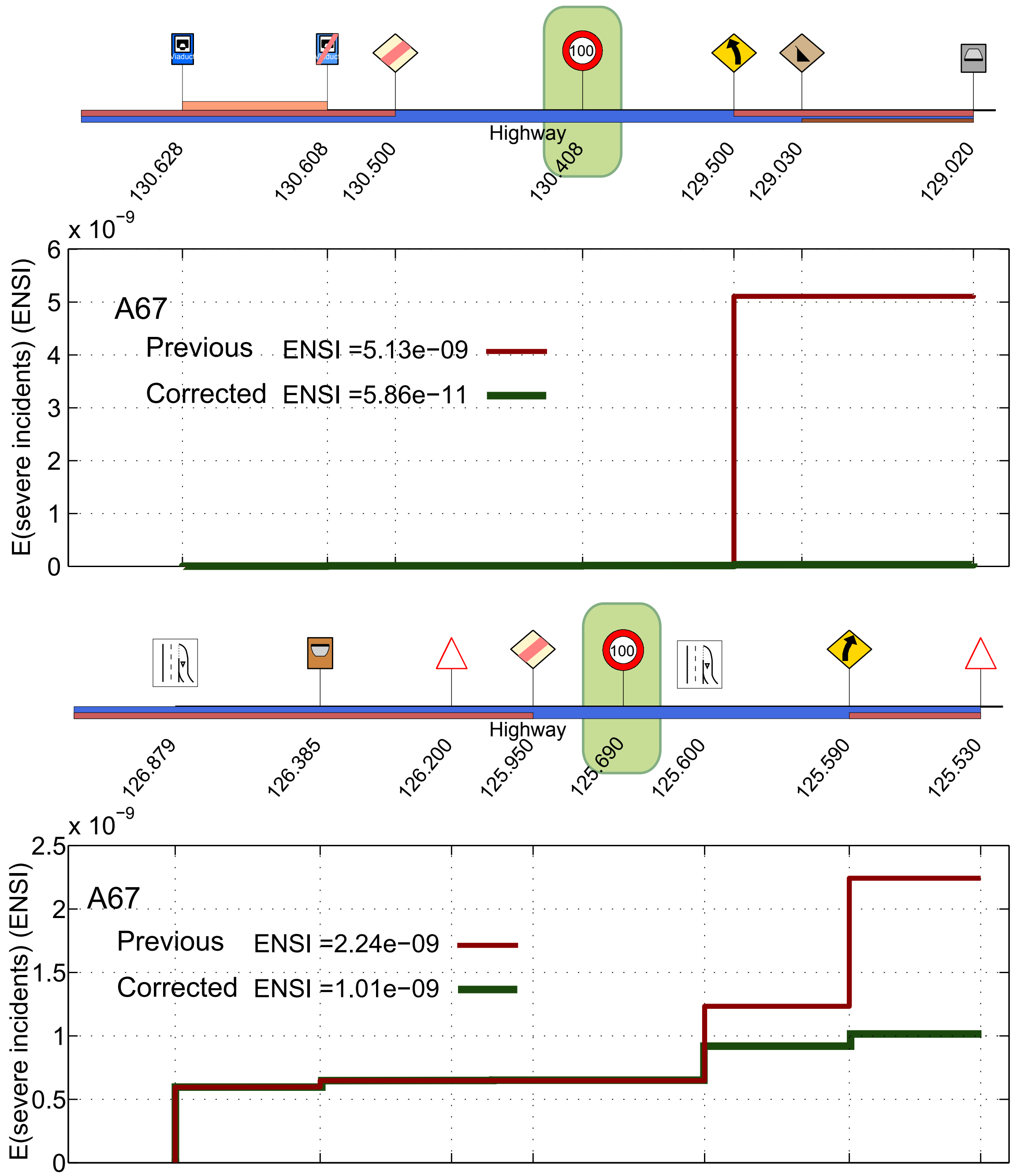


Figure 6.10: Improved safety in A-67 Highway example.

Table 6.7: A-67 Highway example. Sorted list of the items sorted by ENSI, showing the corresponding locations and their associated ENSI values.

					Local ENSI	
Rank	**Item**	**Item name**	**KP**	**Node**	**Previous**	**Corrected**
1	22	CurveIn	129.500	I291-Cv	5.08e-09	**1.52e-11**
2	47	CurveIn	125.590	I620-Cv	1.01e-09	**9.48e-11**
3	91	AcelerationLane	117.250	I1192-AL	6.75e-10	**6.26e-10**
4	68	AcelerationLane	122.085	I895-AL	6.56e-10	6.56e-10
5	97	AcelerationLane	116.240	I1268-AL	6.46e-10	6.45e-10
6	42	AcelerationLane	126.879	I556-AL	5.95e-10	5.95e-10
7	46	AcelerationLane	125.600	I607-AL	5.85e-10	**2.65e-10**
8	60	AcelerationLane	123.190	I791-AL	5.79e-10	5.79e-10
9	3	Yield	132.873	I37-Yl	5.6e-10	5.6e-10
10	93	CurveIn	117.100	I1217-Cv	4.03e-10	**3.28e-10**

Table 6.8: A-67 Highway example. Sorted critical list by annual ENSI values, of incident items with the corresponding KP and nodes.

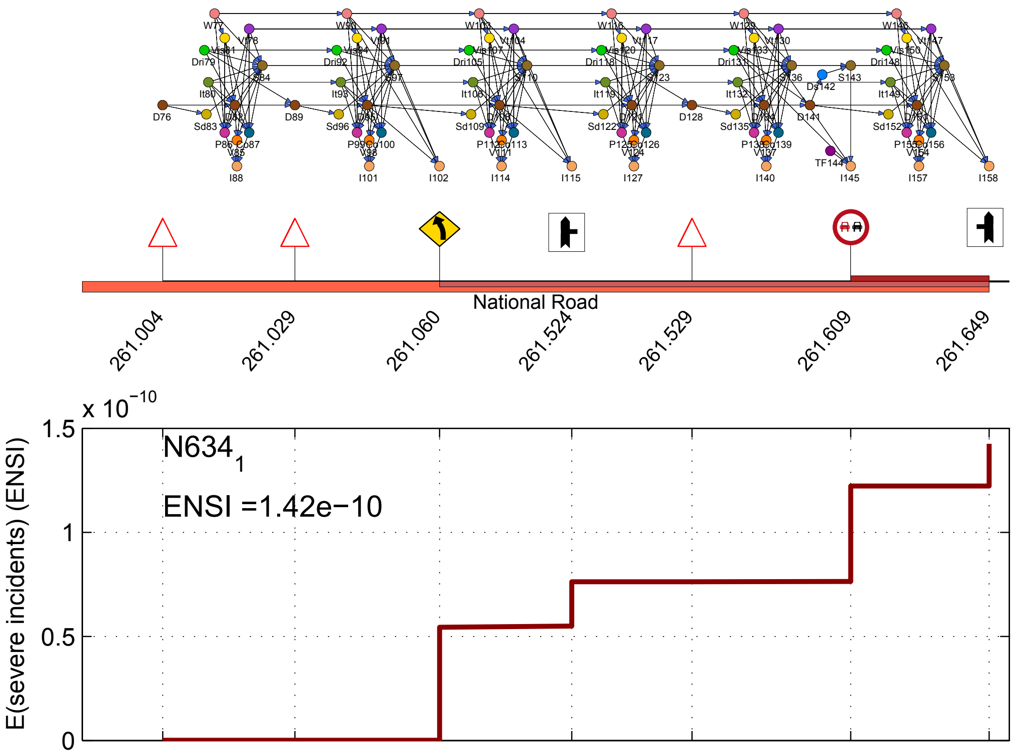
					ENSI per year	
Rank	**Item**	**Item name**	**KP**	**Node**	**Previous**	**Corrected**
1	22	CurveIn	129.500	I291-Cv	0.0187	**5.58e-05**
2	47	CurveIn	125.590	I620-Cv	0.00371	**0.000348**
3	91	AcelerationLane	117.250	I1192-AL	0.00248	**0.0023**
4	68	AcelerationLane	122.085	I895-AL	0.00241	0.00241
5	97	AcelerationLane	116.240	I1268-AL	0.00237	0.00237
6	42	AcelerationLane	126.879	I556-AL	0.00219	0.00219
7	46	AcelerationLane	125.600	I607-AL	0.00215	**0.000975**
8	60	AcelerationLane	123.190	I791-AL	0.00213	0.00213
9	3	Yield	132.873	I37-Yl	0.00206	0.00206
10	93	CurveIn	117.100	I1217-Cv	0.00148	**0.0012**

6.4.2 N-634 National road

In this example the N-634 National road from KP 260.945 to KP 264.870, with a daily mean intensity of 2,145 vehicles is analyzed.

Figures 6.11, 6.12 and 6.13 show for the N-634 road, the same information already described in Figures 6.8 and 6.9 for the A-67 highway. The relative importance of the different items can be identified by comparing the discontinuities of the graphs.

In a similar way, in Tables 6.9 and 6.10 the same information as in Tables 6.7 and 6.8,

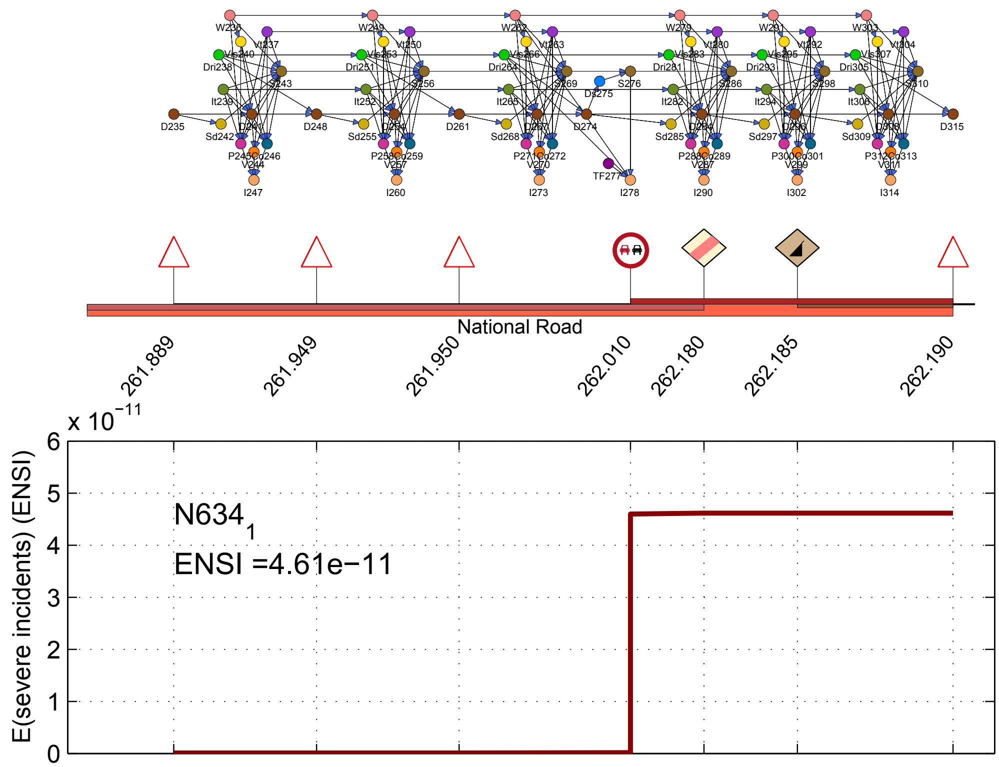


Item	KP	Item name	Node	ENSI Local	ENSI Cumulated
7	261.029	SegWSignals	I88-S	2.7e-14	5.35e-11
8	261.060	SegWSignals	I101-S	2.93e-14	5.36e-11
		CurveIn	I102-Cv	5.42e-11	1.08e-10
9	261.524	SegWSignals	I114-SCr	6.36e-13	1.08e-10
		LateralEntry	I115-LE	2.13e-11	1.3e-10
10	261.529	SegWSignals	I127-SCr	1.05e-14	1.3e-10
11	261.609	SegWSignals	I140-SCr	8.67e-14	1.3e-10
		OvertakingIn	I145-Oti	4.59e-11	1.76e-10
12	261.649	SegWSignals	I157-SOC	4.06e-14	1.76e-10
		LateralEntry	I158-LE	2.03e-11	1.96e-10

Figure 6.11: N-634 National road example. Probabilities of the incident nodes between nodes 76 and 158.

respectively, is given. It can be easily seen that a ‘curve’ and a ‘lateral entry’ (see Table 6.9) are the most critical items.

In order to improve safety at the KP 262.210 (curve) and KP 262.416 (lateral entry), a speed limit signal (40 Km/h) at KP 262.184 and other with limit 50 Km/h at KP 262.275 have been added, as shown in Figure 6.14. The resulting improvements in the

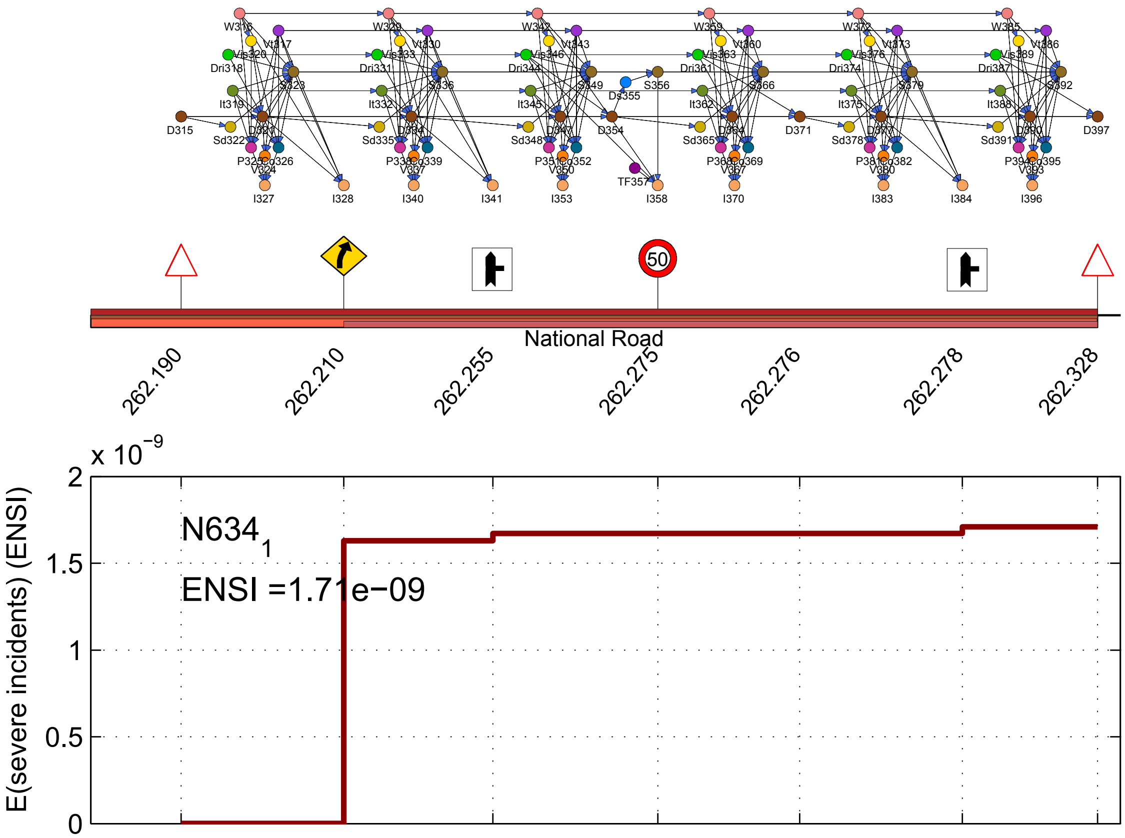


Item	KP	Item name	Node	ENSI Local	ENSI Cumulated
19	261.949	SegWSignals	I247-SCr	5.38e-14	2.2e-10
20	261.950	SegWSignals	I260-SCr	9.03e-16	2.2e-10
21	262.010	SegWSignals	I273-SCr	4.57e-14	2.2e-10
		OvertakingIn	I278-Oti	4.58e-11	2.66e-10
22	262.180	SegWSignals	I290-SOC	1.74e-13	2.66e-10
23	262.185	SegWSignals	I302-SOv	8.69e-15	2.66e-10
24	262.190	SegWSignals	I314-SOS	1.1e-14	2.66e-10

Figure 6.12: N-634 National road example. Probabilities of the incident nodes between nodes 235 and 315.

reduction of the ENSI values have been boldfaced in the last column of Table 6.10, where the beneficial effect of this change at the curve but small for the case of the lateral entry can be appreciated. This means that the action must be done on the vehicles using the lateral entry, for example, by adding a stop or yield signal.

Finally, note that this Bayesian network has 76 items and 992 variables. The required CPU time to build the Bayesian network, calculate all incident probabilities, plot all figures and write the report was 102 sec with the same computer.



				ENSI	
Item	**KP**	**Item name**	**Node**	**Local**	**Cumulated**
25	262.210	SegWSignals	I327-SOS	2.58e-14	2.66e-10
		CurveIn	I328-Cv	1.63e-09	1.9e-09
26	262.255	SegWSignals	I340-Sall	5.94e-14	1.9e-09
		LateralEntry	I341-LE	4.07e-11	1.94e-09
27	262.275	SegWSignals	I353-Sall	3.1e-14	1.94e-09
		SpeedLimit	I358-Sl	1.92e-13	1.94e-09
28	262.276	SegWSignals	I370-Sall	2.38e-16	1.94e-09
29	262.278	SegWSignals	I383-Sall	8.93e-16	1.94e-09
		LateralEntry	I384-LE	3.85e-11	1.98e-09
30	262.328	SegWSignals	I396-Sall	1.72e-14	1.98e-09

Figure 6.13: N-634 National road example. Probabilities of the incident nodes between nodes 315 and 397.

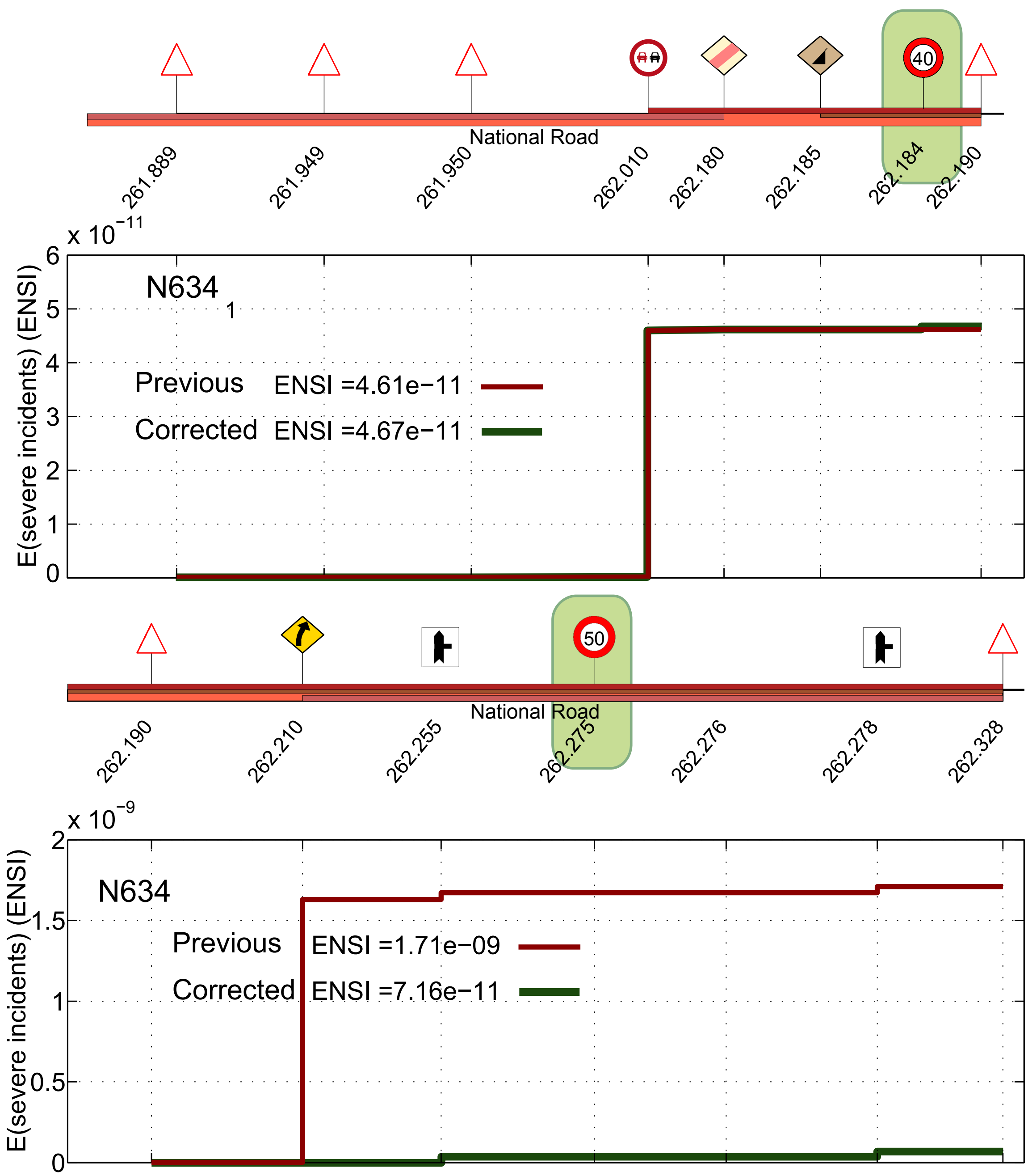


Figure 6.14: Illustration of the safety improvements produced by the corrections.

Table 6.9: N-634 National road example. Sorted critical list by local ENSI values of incident items with the corresponding KP and nodes.

					Local ENSI	
Rank	**Item**	**Item name**	**KP**	**Node**	**Previous**	**Corrected**
1	25	CurveIn	262.210	I328-Cv	1.63e-09	**1.87e-14**
2	35	LateralEntry	262.416	I464-LE	1.12e-10	**9.92e-11**
3	8	CurveIn	261.060	I102-Cv	5.42e-11	5.42e-11
4	11	OvertakingIn	261.609	I145-Oti	4.59e-11	4.59e-11
5	21	OvertakingIn	262.010	I278-Oti	4.58e-11	4.58e-11
6	26	LateralEntry	262.255	I341-LE	4.07e-11	4.07e-11
7	49	LateralEntry	263.710	I643-LE	3.87e-11	3.87e-11
8	29	LateralEntry	262.278	I384-LE	3.85e-11	3.85e-11
9	50	OvertakingIn	263.750	I660-Oti	3.42e-11	1.43e-11
10	2	SpeedLimit	260.955	I24-Sl	3.28e-11	3.28e-11

Table 6.10: N-634 National road example. Sorted critical list by annual ENSI values of incident items with the corresponding KP and nodes.

					ENSI per year	
Rank	**Item**	**Item name**	**KP**	**Node**	**Previous**	**Corrected**
1	25	CurveIn	262.210	I328-Cv	**0.00128**	**1.46e-8**
2	35	LateralEntry	262.416	I464-LE	**8.75e-05**	**7.77e-05**
3	8	CurveIn	261.060	I102-Cv	4.24e-05	4.24e-05
4	11	OvertakingIn	261.609	I145-Oti	3.59e-05	3.59e-05
5	21	OvertakingIn	262.010	I278-Oti	3.59e-05	3.59e-05
6	26	LateralEntry	262.255	I341-LE	3.19e-05	3.19e-05
7	49	LateralEntry	263.710	I643-LE	3.03e-05	3.03e-05
8	29	LateralEntry	262.278	I384-LE	3.01e-05	3.01e-05
9	50	OvertakingIn	263.750	I660-Oti	2.67e-05	1.12e-05
10	2	SpeedLimit	260.955	I24-Sl	2.57e-05	2.57e-05

Chapter 7

Parameter estimation and validation

Contents

7.1 Introduction

One of the most critical parts of the proposed model is the parameter estimation and validation procedure. Since the behaviour of drivers is very difficult if not impossible to know exactly, some of the parameters cannot be estimated with precision because they are related to human errors. Similarly, other parameters, such as those related to slope and embankment stability are very costly to be estimated. Finally, some parameters cannot be estimated because they are very unfrequent and there are no data.

Thus, to obtain some reliable estimates the collaboration of miscellaneous groups of experts is required to improve the power, the credibility of the results and the efficiency of the method. As a positive part the long experience in nuclear power plants can be highlighted, which has allowed us to agree on the frequency values of some very unfrequent events, that cannot be estimated from data because they do not exist.

A validation procedure will require to measure data as speeds at different locations, weather conditions at different locations, frequencies of driver's errors, driver types, traffic intensities, visibilities at different locations, frequencies of vehicle failures and locations, pavement condition at different locations, frequencies of red and green at traffic signals, and accident frequencies.

It is also necessary to test the resulting incident frequencies with observed values. However, where there are no sufficient data, for the comparison only global and general data such as mean number of incidents at curves, intersections, and lateral entries have been used from the roads of our study and other roads too. In this way, several parameters have been undergoing changes as it has been advanced in the model and having more data.

For the moment, the results provided by the proposed model have shown that the method can detect some critical locations in highway and conventional roads by selecting parameter values by common sense, experience in road traffic and a global validation that is based on fitting the incident observations with the model predictions. In particular, the model facilitates the comparison of different incident types and different locations and allows important savings in maintenance operations and accident reduction policies.

7.2 Parameter estimation

The construction of the Bayesian networks model has required the estimation of a large number of parameters. Most of them are parameters used to calculate the conditional probability matrices of each node given its parents and all have been determined by our working group, thanks to previous engineering knowledge and common sense, the help of works and opinions of other experts, and the experience and knowledge that have been applied before in nuclear power plants.

In this Thesis the whole list of parameters is not included since it is an extensive list in which to see the values of each parameter out of context would not be very useful.

Therefore, specifying certain parameters when they are named or appear in different formulas of this document has been considered more convenient. In this way it is much more practical and understandable.

The estimation process has led to different steps to the current results. These results obtained by the presented model can be considered acceptable for the analysis of road safety due to the validation made and shown graphically in the following section.

7.3 Validation

Three roads analyzed with this model are presented in a graphical form to show the validation of the obtained results. For that, the observed accident rate on these roads between the years 2006 and 2016 (data that have been provided to us and later have been represented graphically), as well as the potential accidents, those that would be expected to occur (obtained with the Bayesian networks model) have been represented in the layout with the purpose of comparing the similarity of the results and performing an acceptable validation.

7.3.1 Observed accident rate and potential accidents in the Autonomic road CA-131.

A summary of the most relevant information on the observed accident rate of the Autonomic road CA-131 is shown in Table 7.1, which gives the types of accidents recorded, their severity (severe (Sev) medium (Med) or minor (Min)) and their frequency. Headings A, D, N and S refer to ascending, descending, north and south directions, respectively.

Note that this table is only a summary of all accidents produced.

Table 7.1: List of possible types of accidents, directions and frequencies by severity (severe (Sev), medium (Med) and minor (Min)) in the Autonomic road CA-131.

	A			D			N			S			Total		
Accident type	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**
Run over animals	0	0	23	0	0	15	0	0	3	0	0	0	0	0	41
Run over pedestrians	0	1	3	0	2	1	0	0	0	0	0	0	0	3	4
Collision of vehicles with obstacle in road: another object or material	0	0	3	0	0	3	0	0	0	0	0	0	0	0	6
Collision of vehicles with obstacle in road: parked or broken vehicle	0	0	1	0	0	1	0	0	1	0	0	1	0	0	4
Collision of vehicles underway: reach	1	0	28	0	0	16	0	0	0	0	0	0	1	0	44
Collision of vehicles underway: head on collision	0	0	3	0	1	3	0	0	0	0	1	4	0	2	10
Collision of vehicles underway: front-to-side collision	1	3	19	0	2	11	0	0	2	1	0	6	2	5	38
Collision of vehicles underway: lateral	0	0	12	0	0	6	0	0	0	0	0	3	0	0	21
Collision of vehicles underway: multiple	0	0	13	0	0	11	0	0	1	0	0	0	0	0	25
Other type of accident	0	1	3	0	1	7	0	0	0	0	0	1	0	2	11
Crash on the right	0	0	18	0	0	16	0	0	0	0	0	0	0	0	34
Crash on the right with free fall	0	0	3	0	0	1	0	0	0	0	0	0	0	0	4
Crash on the right with overturning	0	0	6	0	0	4	0	0	0	0	0	0	0	0	10
Crash on the right plain land	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2
Crash on the right, others	0	0	3	0	0	3	0	0	0	0	0	0	0	0	6
Crash on the right, other crash type	0	0	3	0	0	2	0	0	1	0	0	0	0	0	6
Crash on the left	1	0	1	0	0	4	0	0	0	0	0	0	1	0	5
Crash on the left into kerb	0	0	2	0	0	2	0	0	0	0	0	0	0	0	4
Crash on the left into building	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4
Crash on the left into tree or post	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Crash on the left, others	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Crash on the left, other crash type	0	0	8	0	0	6	0	0	1	0	0	0	0	0	15
Overturning	0	0	2	0	0	1	0	0	0	0	0	0	0	0	3

In order to analyze the results in a much faster and more effective way, the trajectory and the accident rate corresponding to the two directions (ascending and descending) have been graphically plotted (see lower part of Figure 7.1), differentiating between minor (Min), medium (Med) or severe (Sev) incidents, according to the classification provided. The points where incidents have been observed have been indicated with circles whose diameter is proportional to the square root of their frequency and whose colors indicate the severity of the incidents (light yellow in minor, orange, medium and red, severe). They significantly facilitates the identification of the most dangerous points and the sections of greater concentration of accidents that require an immediate action.

Figure 7.1: At the top, the graphical representation of the incident prediction for the Autonomic road CA-131 is shown, indicating the frequencies of level incidents (yellow), medium (orange) and severe (red) by circles of diameter proportional to the square root of the frequency of the incidents at said points. It is also shown, in the lower part, the observed accidents so that their similarity can be observed.

In addition, the observed accident rate is used to compare the potential accidents resulting from the proposed Bayesian network model and to identify possible discrepancies, which would motivate possible improvements to be introduced in the model.

The accidents observed between 2006 and 2016 includes 4 severe incidents, 12 medium and 299 minor, making a total of 315 incidents. The observation leads to the identification of nine different sections that will be described in detail in the chapter of the study of real cases.

The purpose of this section is to show the similarity between the observed accidents for a period of 10 years and the potential accidents. Thus, both are shown together in Figure 7.1. At the top of Figure the potential accidents obtained with the program of Bayesian networks can be appreciated and in its lower part the observed accidents. To make the subsequent comparison, the same stretches that are used to describe the observed accident rate can be identified in the potential accidents, being possible to see how the similarity that they present serves to proceed with the validation of the results.

Table 7.2 shows the different types of incidents, sorted by their frequency in decreasing order, and these are differentiated by level of severity (severe, medium and minor), as well as totals.

This table shows that incidents on road are the most frequent (78.45), followed by those caused by collisions at intersections (75.67) and road (65.50), run over animals (53.54) and at the lateral entries (34.86). It also indicates that the severe incidents expected are 5.57, the medium 21.18 and the minor 319.03, with a total of 345.78 incidents.

Table 7.2: Potential incidents by type: List of points with potential incidents in $CA - 131$, discretized by type of incident.

	Severity			
Accident Type	**Severe**	**Medium**	**Minor**	**Total**
Incident on road	1.20	3.94	73.32	78.45
Intersection	0.00	3.11	72.56	75.67
Collision	2.02	1.03	62.45	65.50
Run over animals	2.36	4.28	46.91	53.54
Lateral entry	0.00	1.01	33.85	34.86
Curve	0.00	2.09	17.11	19.21
Pedestrian crossing	0.00	4.54	2.48	7.03
Roundabout	0.00	1.17	4.65	5.83
Traffic Light	0.00	0.00	3.40	3.40
Run over pedestrians	0.00	0.00	2.30	2.30
Total	5.57	21.18	319.03	345.78

7.3.2 Observed accident rate and potential accidents in the Autonomic road CA-132.

In the same way as in the previous case it is proceeded with the Autonomic road CA-132. Figure 7.2 shows the similarity between the accidents observed between the years 2006 and 2016 and the potential accidents predicted by the model, in addition to pointing out some information about the type of incidents that appear without going into detail since as discussed above, this will be analyzed in another chapter.

Table 7.3 shows a summary of the most relevant information of the observed accident rate on this road giving the types of accidents recorded, their severity (severe (Sev) medium (Med) or minor (Min)) and their frequency.

The graphical representation of these data is given in Figure 7.2, which on the right side shows the places where the incidents have occurred, their frequency, indicated by the size of the circles, with a radius proportional to the square root of their frequency, and their severity, indicated by the color (light color in minor incidents, orange, in medium and red, in severe). There were not severe incidents, 11 medium, and 164 minor, that is, a total of 175 incidents. Three different sections can be distinguished.

On the left of Figure 7.2 the accident prediction obtained with the Bayesian network

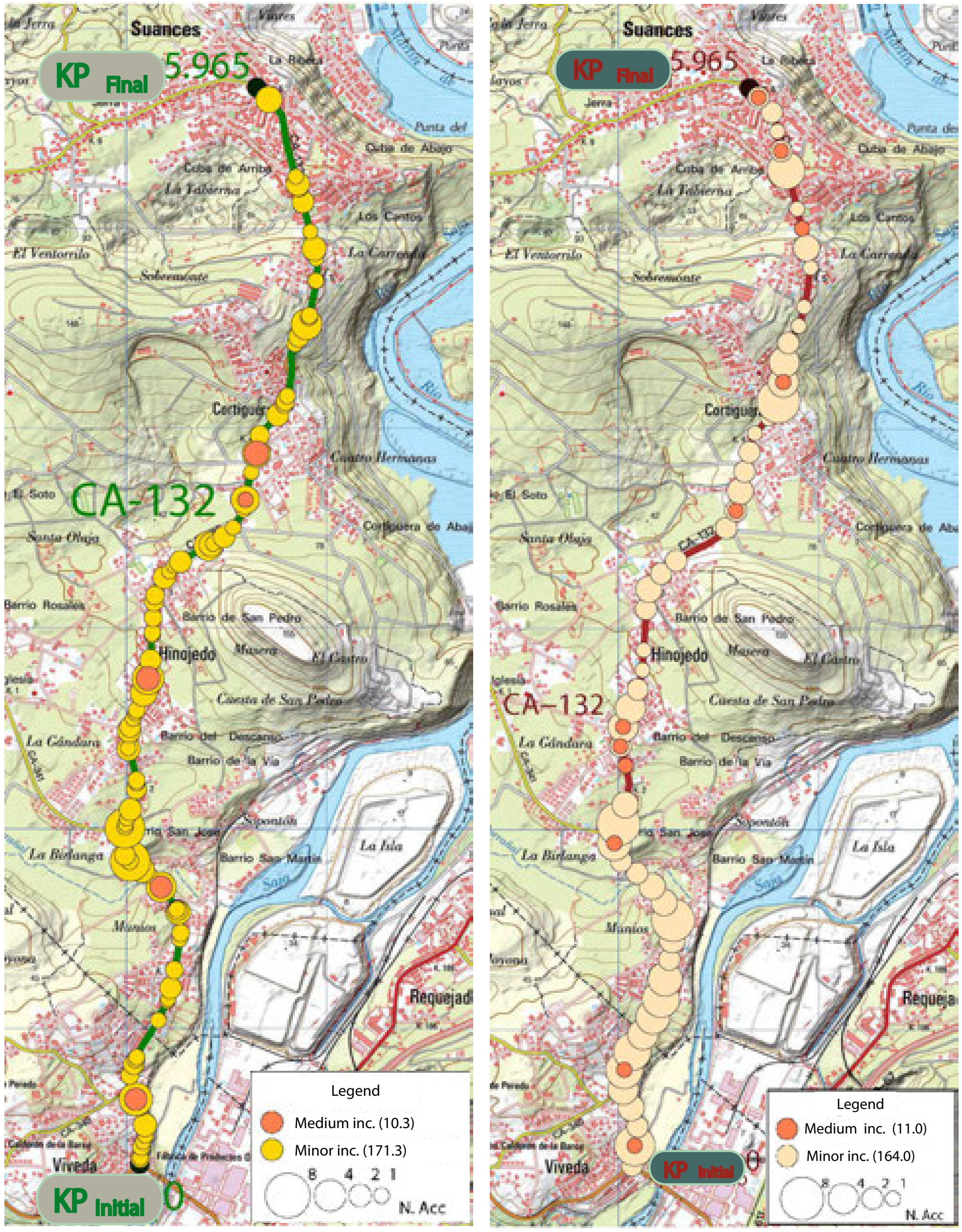


Figure 7.2: On the left it is the graphical representation of the incident prediction for the CA-132, which shows the frequencies of level incidents (yellow), medium (orange) and severe (red) by means of circles of proportional diameter to the square root of the frequency of the incidents at said points. It is also shown, on the right side, the observed accidents so that their similarity can be observed.

program is shown.

Table 7.4 shows the different types of incidents, sorted by their frequency in decreasing

Table 7.3: List of possible types of accidents, directions and frequencies by severity (severe (Sev), medium (Med) and minor (Min)) in the Autonomic road CA-132.

Accident type	A			D			N			S			Total		
	Sev	Med	Min	Sev	Med	Min	Sev	Med	Min	Sev	Med	Min	Sev	Med	Min
-	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Run over animals	0	0	3	0	0	5	0	0	0	0	0	0	0	0	8
Run over pedestrians	0	1	1	0	3	2	0	0	0	0	0	0	0	4	3
Collision of vehicles with obstacle in road: another object or material	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Collision of vehicles underway: reach	0	1	36	0	2	18	0	0	0	0	0	0	0	3	54
Collision of vehicles underway: head on collision	0	1	1	0	1	1	0	0	0	0	2	5	0	4	7
Collision of vehicles underway: front-to-side collision	0	0	15	0	0	7	0	0	1	0	0	3	0	0	26
Collision of vehicles underway: lateral	0	0	4	0	0	5	0	0	0	0	0	1	0	0	10
Collision of vehicles underway: multiple	0	0	16	0	0	11	0	0	0	0	0	1	0	0	28
Other type of accident	0	0	1	0	0	0	0	0	1	0	0	0	0	0	2
Crash on the right	0	0	9	0	0	3	0	0	0	0	0	0	0	0	12
Crash on the right, others	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Crash on the right, other crash type	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Crash on the left into building	0	0	7	0	0	1	0	0	0	0	0	0	0	0	8
Crash on the left, other crash type	0	0	2	0	0	2	0	0	1	0	0	0	0	0	5

order, and these are differentiated by level of severity (severe, medium and minor), as well as totals.

This table shows that incidents at intersections are the most frequent (83.26), followed by those caused by collisions (46.39), incidents on roads (35.69) and at lateral entries (18.25). It also indicates that severe incidents are not expected and that 6.03 medium and 184.90 minor are expected, with a total of 190.92 incidents.

Table 7.4: Potential incidents by type: List of points with potential incidents in $CA-132$, discretized by type of incident.

	Severity			
Accident type	**Severe**	**Medium**	**Minor**	**Total**
Intersection	0.00	6.03	77.24	83.26
Collision	0.00	0.00	46.39	46.39
Incident on road	0.00	0.00	35.69	35.69
Lateral entries	0.00	0.00	18.25	18.25
Run over animals	0.00	0.00	4.11	4.11
Traffic Light	0.00	0.00	2.22	2.22
Run over pedestrians	0.00	0.00	1.00	1.00
Total	0.00	6.03	184.90	190.92

7.3.3 Observed accident rate and potential accidents in the Autonomic road CA-142.

Proceeding in the same way on the Autonomic road CA-142, the following data stand out.

Table 7.5: List of possible types of accidents, directions and frequencies by severity (severe (Sev), medium (Med) and minor (Min) in the Autonomic road CA-142.

	A			D			N			S			Total		
Accident type	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**	**Sev**	**Med**	**Min**
Run over flocks	0	0	1	0	0	2	0	0	0	0	0	0	0	0	3
Run over animals	0	0	36	0	0	34	0	0	3	0	0	3	0	0	76
Run over pedestrians	0	0	5	0	2	2	0	0	0	0	1	0	0	3	7
Collision of vehicles with obstacle in road: another object or material	0	0	7	0	0	8	0	0	1	0	0	1	0	0	17
Collision of vehicles with obstacle in road: parked or broken vehicle	0	0	0	0	0	7	0	0	1	0	0	0	0	0	8
Collision of vehicles underway: reach	0	1	45	0	2	51	0	0	1	0	0	3	0	3	100
Collision of vehicles underway: head on collision	0	1	3	1	0	2	0	0	2	1	2	9	2	3	16
Collision of vehicles underway: front-to-side collision	0	3	30	0	1	34	0	0	4	1	3	17	1	7	85
Collision of vehicles underway: lateral	0	0	21	0	0	18	0	0	2	0	0	15	0	0	56
Collision of vehicles underway: multiple	0	0	11	0	1	10	0	0	1	0	0	4	0	1	26
Other type of accident	0	0	6	0	0	4	0	0	4	0	0	4	0	0	18
Crash on the right	0	2	25	0	4	24	0	0	0	0	0	0	0	6	49
Crash on the right with free fall	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Crash on the right with overturning	0	0	5	0	0	5	0	0	0	0	0	0	0	0	10
Crash on the right plain land	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1
Crash on the right, others	0	0	1	0	0	3	0	0	0	0	0	0	0	0	4
Crash on the right, other crash type	0	0	7	0	0	1	0	0	0	0	0	1	0	0	9
Crash on the left	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Crash on the left with kerb	0	1	4	0	1	8	0	0	0	0	0	0	0	2	12
Crash on the left into building	0	1	6	0	0	9	0	0	0	0	0	0	0	1	15
Crash on the left into tree or post	0	0	2	0	0	3	0	0	0	0	0	1	0	0	6
Crash on the left with free fall	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Crash on the left with overturning	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Crash on the left, others	0	1	2	0	0	0	0	0	1	0	0	0	0	1	3
Crash on the left, other crash type	0	1	11	0	1	12	0	0	0	0	0	0	0	2	23
Overturning	0	1	2	0	0	3	0	0	0	0	0	1	0	1	6

The most relevant information provided on the observed accident rate is in Table 7.5, which gives the types of accidents recorded, their severity (severe (Sev), medium (Med) or minor (Min)) and their frequency.

These same data are represented graphically in Figure 7.3, which on the right side the places where the incidents have occurred, their frequency, indicated by the size of the circles with a radius proportional to the square root of its frequency, and their severity, indicated by the color (in light color are the minor incidents, in orange, the medium and in red, the severe) are shown. Between these years 3 severe incidents ocurred and there were 30 medium and 552 medium. Thus, on the road can be identified 7 sections. On the left, the prediction of potential accidents obtained with the Bayesian networks program is shown, being possible to distinguish the same 7 sections.

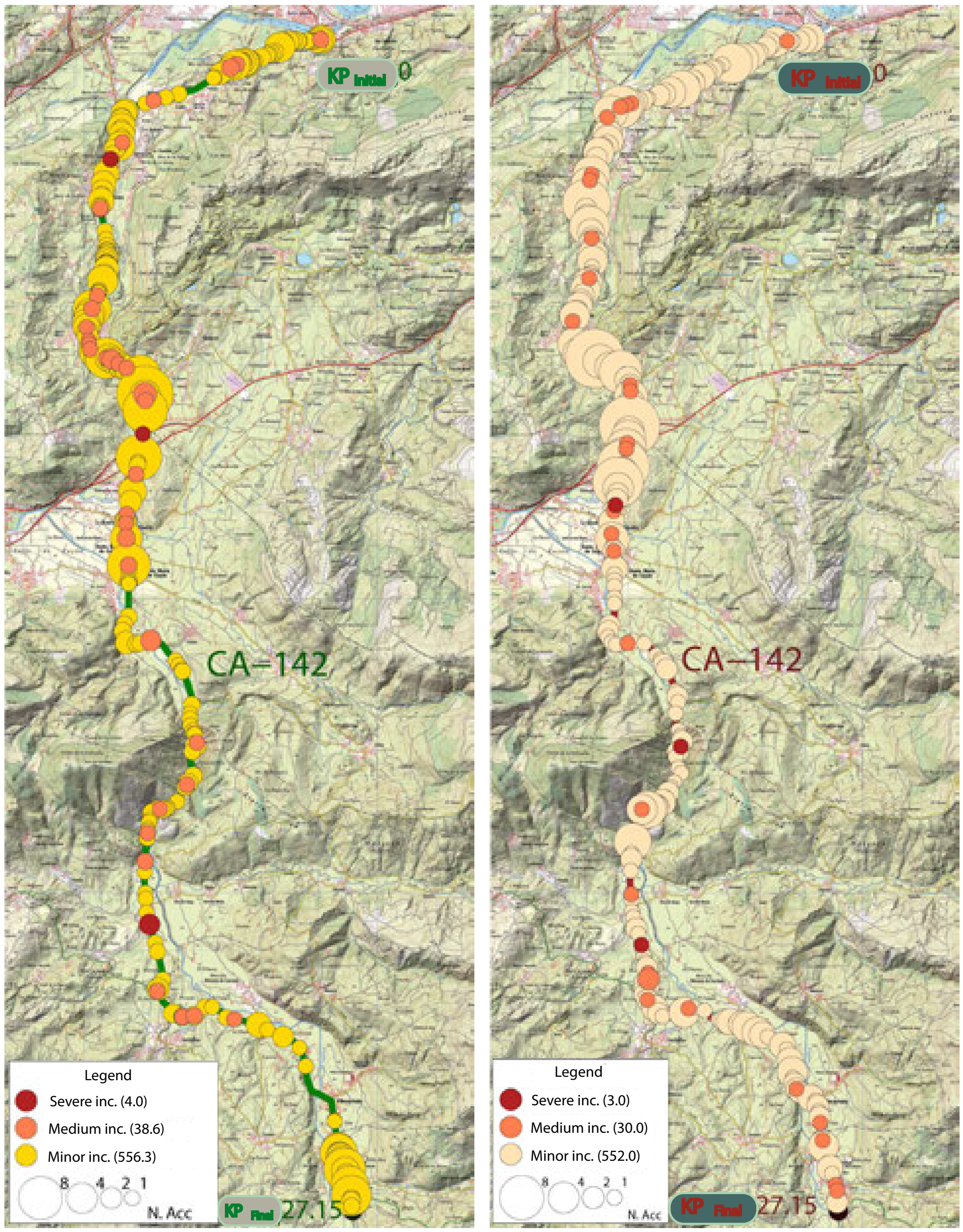


Figure 7.3: On the left hand side the graphical representation of the accident prediction for the CA-142 is, showing the frequencies of level incidents (yellow), medium (orange) and severe (red) by circles of diameter proportional to the square root of the frequency of the incidents at said points. On the right side it is also shown the observed accidents to be able to observe their similarity.

Table 7.6 shows the different types of incidents, sorted by their frequency in decreasing order, and these are differentiated by level of severity (severe, medium and minor), as well as totals.

Table 7.6: Potential incidents by type: List of points with potential incidents in $CA-142$, discretized by type of incident.

	Severity			
Accident type	**Severe**	**Medium**	**Min**	**Total**
Collision	0.00	3.21	140.42	143.63
Incident on road	0.00	4.07	126.68	130.75
Lateral entries	0.00	5.71	80.19	85.89
Intersection	0.00	3.99	66.76	70.75
Run over animals	0.00	2.87	62.23	65.10
Curve	0.00	2.04	20.55	22.59
Run over pedestrians	0.00	0.00	1.00	1.00
Total	0.00	21.88	497.83	519.71

This Table shows that the incidents due to collisions are the most frequent (143.63), followed by incidents on road (130.75) and those at lateral entries (85.89) and at intersections (70.75). It also indicates that severe incidents are not expected, 21.88 medium and 497.83 minor, with a total of 519.71 incidents.

With all this data and Figure 7.3, the similarity between the observed accident rate on the road in the years analyzed and the prediction of potential accidents can be seen, considering acceptable for validation the results obtained.

Chapter 8

Sensitivity analysis

Contents

8.1 Sensitivity analysis

Apart from providing some estimates of the probabilities of occurrence of relevant undesired events, to analyze how sensitive and/or robust the estimates are to the assumed values of the parameters is very important. This can be done using well known techniques such as sensitivity analysis. For a general view of some sensitivity analysis methods see Sobol (2001), Saltelli (2002), Saltelli et al. (2004), or Castillo et al.(2006a) and for some particular applications to Civil engineering see, for example, Castillo et al. (2004, 2009), and Kala and Puklicky (2009).

In order to know the way in which the parameters can affect the model and how they vary, to perform a sensitivity analysis is very useful .

There are several options to develop this type of analysis. Below some of them are listed:

1. *Global sensitivity analysis.* This provides a global view of how the parameter affects the result for a wide range of values or combination of values with other parameters (see Saltelli et al. (2004)).

2. *Local sensitivity analysis.* This method provides the local derivatives of the result with respect to the parameter (see Castillo et al. (2006)).

3. *Discrete sensitivity analysis.* In this case a discrete set of values for the parameter is selected and the calculations are done for each of them.

4. *Continuous sensitivity analysis.* Here a continuous range of values for the parameter is selected and a curve is given showing the result for all parameter values in the given interval.

In this Thesis the Continuous sensitivity analysis is performed. For that, $10 \leq n \leq 40$ equally spaced values are taken for the parameter and a spline is interpolated to plot the curves. In addition, splines that conserve the increasing or decreasing character of the base points are used, since it is a very important property to maintain in the interpolations.

Be noted that this fourth option is much more complicated than the second one because the derivative implies an infinitesimal increment in the parameter and this means that the actual values of all variables or nodes in the network are known. However, when finite changes are produced, the values of all the variables must be determined again. For example, one increment in the speed limit value of a given signal implies changes in the speeds of many variables located after such a signal until another speed limit change occurs. This means that analyzing the set of all variables or nodes changing their conditional probability tables is difficult and costly.

Thus, the partitioning technique explained in Section 6.2 acquires a relevant role in the sensitivity analysis because only a small portion of the BN need to be used in the calculations. More precisely, it is necessary to recalculate a subnetwork starting at the location of the parameter with respect to which the sensitivity is calculated and ending at the node or variable whose sensitivity is looked for. This implies a very important reduction in the evaluation time.

The partition technique for the sensitivity analysis performed can be optimized by using the following process:

1. *Step 1. Decide the parameters whose sensitivity is looked for.* This means selecting the set of parameters with respect to which a sensitivity analysis is wanted to be performed.

2. *Step 2. Select the partitions adequately.* Since there is interest in using small partitions, the partitions can be selected to include both the parameters and the variables whose sensitivity is desired.

3. *Step 3. Perform the probabilistic safety analysis of the BN storing the relevant information for sensitivity analysis.* The information of the relevant subnetworks to be used later must be stored (in our software the BNT and the JavaBayes subnetworks are stored).

4. *Step 4. Perform the sensitivity analysis using only the subnetworks.* This can be done with an independent computer program without the need of repeating the probabilistic safety analysis of the BN again.

The following is an example of a Spanish secondary road in which this method has been used to develop a sensitivity analysis of some parameters of the model.

8.2 Example of application: CA-182 secondary road

The case of the Spanish secondary road CA-182 between KP 8.501 and KP 15.250 with a daily mean intensity of 558 vehicles is considered in this example and it is shown in Figure 8.1.

Figure 8.1: CA-182 secondary road trace from KP 8.150 to 15.400, showing the three riskiest points.

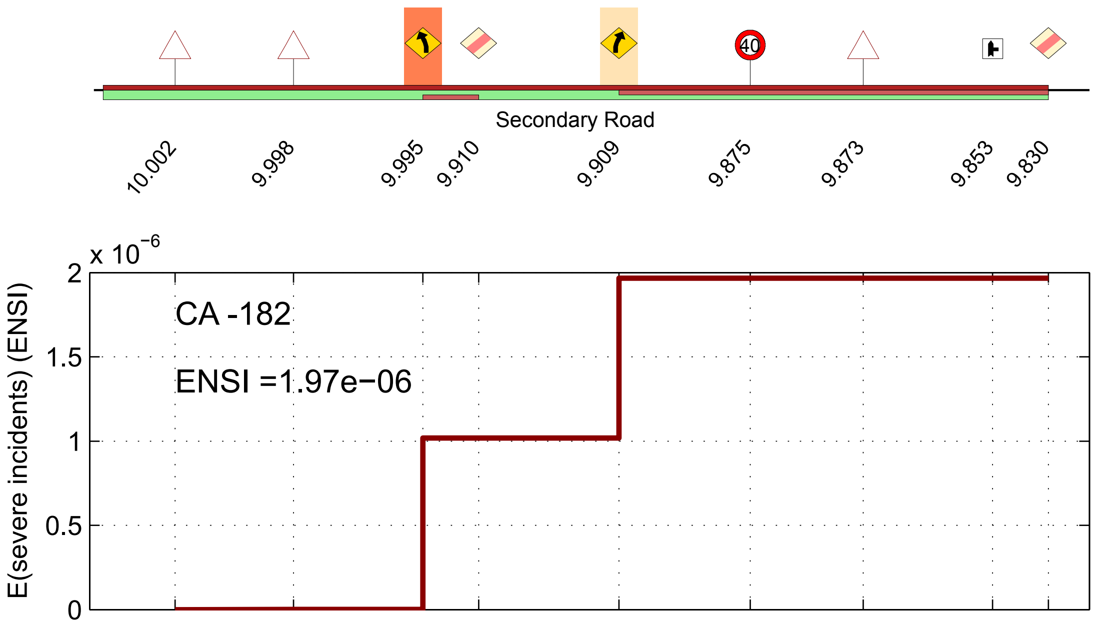
Table 8.1 shows the critical list of items of the road with the largest ENSI (Expected Number of equivalent Severe Incidents) values and the corresponding KP, incident nodes and local probabilities, where three curves can be identified as the main responsible for the risks at KP 9.995, KP 9.909 and KP 13.200. The plant location of these curves can be seen in Figure 8.1 at points 1, 2 and 3 respectively, distinguishing each point by colors according to the severity obtained in the ENSI values.

Table 8.1: Critical list of items of the $CA - 182$ road with the largest ENSI values and the corresponding KP, incident nodes and local probabilities.

Rank	Item	Item name	KP	Node	Local ENSI
1	90	CurveIn	9.995	I1202-9995Cv	1.02e-06
2	92	CurveIn	9.909	I1227-9909Cv	9.5e-07
3	35	CurveIn	13.200	I466-13200Cv	8.89e-08

$ENSI^*$ Sorted list of Expected Number of Severe Incidents

Figures 8.2 and 8.3 show the graphical representation of two road segments (specifically the segments in which the three highest risk curves are included), in their upper parts, and the plots of the cumulated ENSI frequencies in their lower parts. In addition, a table with the ENSI values associated with each item of the segments is provided at the figure bottom.

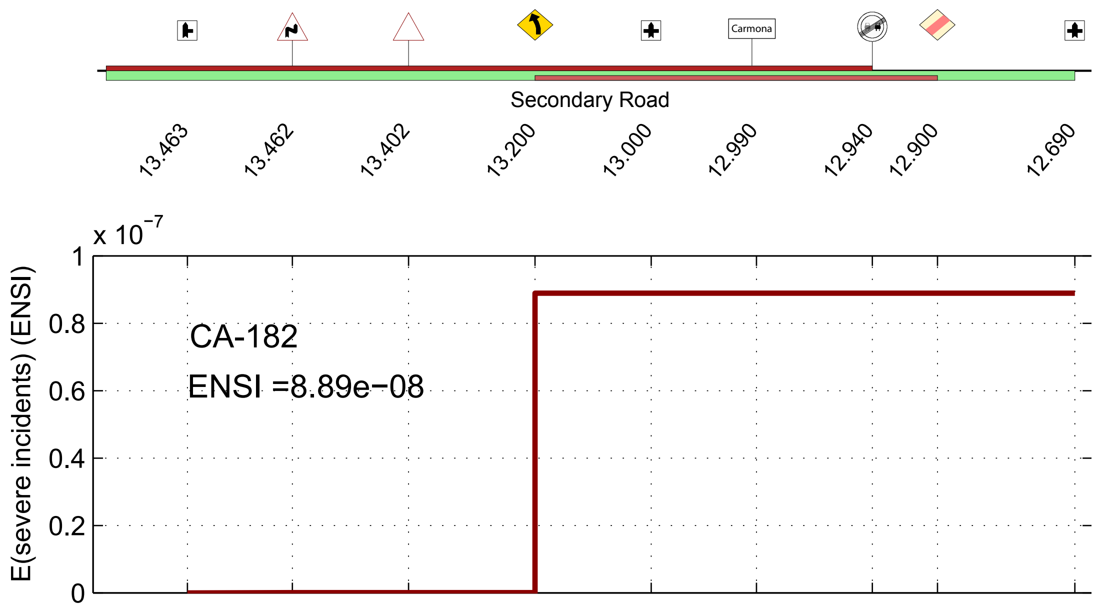


Item	KP	Item name	Node	Probability Local	Probability Cumulated
88	10.002	SegWSignals	I1175-10002SOv	4.94e-13	1.03e-07
89	9.998	SegWSignals	I1188-9998SOv	4.78e-13	1.03e-07
90	9.995	SegWSignals	I1201-9995SOv	2.99e-13	1.03e-07
		CurveIn	I1202-9995Cv	**1.02e-06**	1.12e-06
91	9.910	SegWSignals	I1214-9910SOC	7.24e-11	1.12e-06
92	9.909	SegWSignals	I1226-9909SOv	1.69e-13	1.12e-06
		CurveIn	I1227-9909Cv	**9.5e-07**	2.07e-06
93	9.875	SegWSignals	I1239-9875SOC	9.99e-11	2.07e-06
94	9.873	SegWSignals	I1254-9873SOC	3.53e-14	2.07e-06
95	9.853	SegWSignals	I1267-9853SOC	8.02e-16	2.07e-06
		LateralEntry	I1268-9853LE	3.57e-12	2.07e-06
96	9.830	SegWSignals	I1280-9830SOC	9.95e-16	2.07e-06

Figure 8.2: Cumulative ENSI graphs for the most risky zone of CA-182 road, showing that two curves are responsible for the high ENSI values.

The relative risks associated with the different items can be identified (the two riskiest curves in Figure 8.2 and the third one in Figure 8.3) by comparing the discontinuities of the graphs.

Figure 8.4 shows the cumulative ENSI graphs corresponding to a segment in which signs are not the main causes of risk, but segment without signs. They correspond to the KP 10.884 to KP 10.005, where the slope of the graph is the highest.

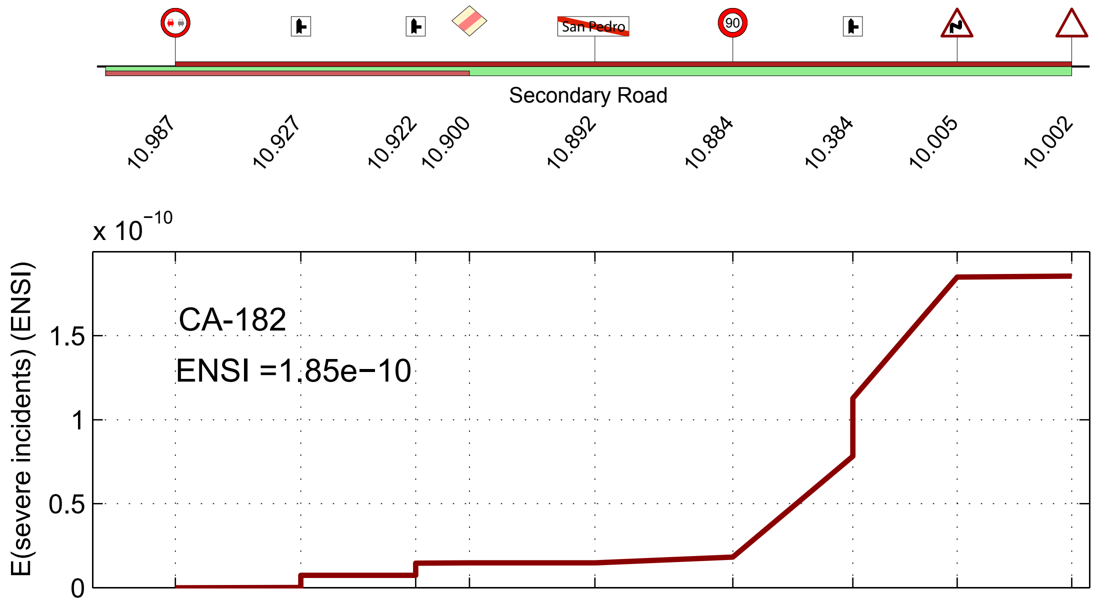


Item	KP	Item name	Node	Probability Local	Probability Cumulated
32	13.463	SegWSignals	I426-13463SOv	2.68e-14	4.64e-10
		LateralEntry	I427-13463LE	1.19e-11	4.76e-10
33	13.462	SegWSignals	I439-13462SOv	5.99e-15	4.76e-10
34	13.402	SegWSignals	I452-13402SOv	1.97e-13	4.76e-10
35	13.200	SegWSignals	I465-13200SOv	7.81e-13	4.77e-10
		CurveIn	I466-13200Cv	**8.89e-08**	8.93e-08
36	13.000	SegWSignals	I478-13000SOC	7.73e-12	8.93e-08
		Intersection	I479-13000Int	3.11e-11	8.94e-08
37	12.990	SegWSignals	I491-12990SOC	4.41e-13	8.94e-08
38	12.940	SegWSignals	I504-12940SOC	3.04e-12	8.94e-08
39	12.900	SegWSignals	I517-12900SCr	1.65e-12	8.94e-08
40	12.690	SegWSignals	I529-12690S	1.39e-12	8.94e-08
		Intersection	I530-12690Int	1.65e-11	8.94e-08

Figure 8.3: Cumulative ENSI graphs showing the ENSI values due to the curve at KP 13.200.

Table 8.2 shows the ENSI values associated with the groups of identical items, such as curves, intersections, and underpasses, so that we can have an idea of the influence of these types of items on the global safety of the road. It is clear that curves are the type of items leading to the highest risks in this road segment.

To solve the problem with the two curves at KP 9.995 and KP 9.909, it has been



				Probability	
Item	**KP**	**Item name**	**Node**	**Local**	**Cumulated**
80	10.987	SegWSignals	I1070-10987SCr	1.8e-12	1.03e-07
81	10.927	SegWSignals	I1083-10927SOC	7.65e-14	1.03e-07
		LateralEntry	I1084-10927LE	7.27e-12	1.03e-07
82	10.922	SegWSignals	I1096-10922SOC	6.69e-15	1.03e-07
		LateralEntry	I1097-10922LE	7.28e-12	1.03e-07
83	10.900	SegWSignals	I1109-10900SOC	3.25e-14	1.03e-07
84	10.892	SegWSignals	I1121-10892SOv	3.54e-15	1.03e-07
85	10.884	SegWSignals	I1134-10884SOv	3.51e-12	1.03e-07
86	10.384	SegWSignals	I1149-10384SOv	6.01e-11	1.03e-07
		LateralEntry	I1150-10384LE	**3.44e-11**	1.03e-07
87	10.005	SegWSignals	I1162-10005SOv	7.23e-11	1.03e-07
88	10.002	SegWSignals	I1175-10002SOv	4.94e-13	1.03e-07

Figure 8.4: Cumulative ENSI graph corresponding to a segment in which segment without signs are the main causes of risk (KP 10.884 to KP 10.005).

considered to move the 40 km/h speed limit sign from KP 9.875 to KP 10.200 and to reduce the speed limit to 70 km/h at KP 10.884, thereby significantly reducing the risk of a severe incident.

Figure 8.5, which shows a comparison of the previous and the corrected plots illustrates the important improvement produced with this correction.

When a new evaluation of risks with these changes is performed, Figure 8.6 is obtained,

Table 8.2: Grouped incident list: List of items of CA-182 road with their frequencies and total ENSI values.

Incident types	Frequency	Local ENSI
Curve	17	2.1e-06
SegWSignal	129	7e-10
Lateral Entry	25	5.49e-10
Intersection	11	3.82e-10
Underpass	1	8.07e-12
Pedestrian Crossing	4	2.86e-16
$ENSI^{*}$ Expected Number of Severe Incidents		

Figure 8.5: Illustration of the improvements after correcting the speed limit signs.

in which the most risky locations after the corrections are shown. Note that the two previous curves are not listed and only the third curve remains with an ENSI value larger than 10^{-9}.

Figure 8.6: CA-182 secondary road trace from KP 8.150 to KP 15.400, showing the most risky locations after making the corrections.

Finally, to illustrate the sensitivity analysis method, the example in Figure 8.1 is considered. It is a simple example consisting of two curves, one at KP 9.995 with radius 80 m and another at KP 9.909 with radius 90 m with a previous speed limit sign of 90 km/h at KP 10.884.

The sequence of traffic signs of this example and the associated KPs of their locations are indicated in the upper plot in Figure 8.2 and the bottom of the plot shows the cumulated ENSI values. Finally, the lower part is a table with all the relevant information, showing that the ENSI at the first curve is 1.02×10^{-6} and the ENSI at the second curve is 9.5×10^{-7}.

As indicated in Figure 5.7 the more relevant parameters analyzing a curve incident are, the relative to the characteristics of the curve, in this case the curve speed limit at KP 10.884 and the curve radius, and the parameters ρ_1, ρ_2, β and γ in Formula (5.40). To illustrate the sensitivity analysis, the ENSI values when these parameters are changed one by one, are calculated .

Figure 8.7 shows these changes. These plots permit us to study the changes due to parameters ρ_1, ρ_2, β and γ in Equation (5.40), being ρ_1 and ρ_2 the probabilities of incident at the curve under no speed excess and speed excess, respectively, and β and γ two parameters to consider the relative difference of the speed excess v_{ex} with respect to the critical speed at the curve v_{crit}. In addition, the sensitivities with respect to the curve radius and the speed limit are calculated. These are non-linear which means that the changes in the ENSI value can be immediately recalculated when one of the parameters is modified without any recalculation and that the changes due to small changes can be easily estimated in several parameters with simple calculations based on the gradients.

The knowledge of these sensitivity curves is very useful because they provide the necessary corrections to obtain reasonable ENSI values. For example, it can be seen that an increase of the curve radius from 80 m to 120 m implies a reduction of the ENSI value

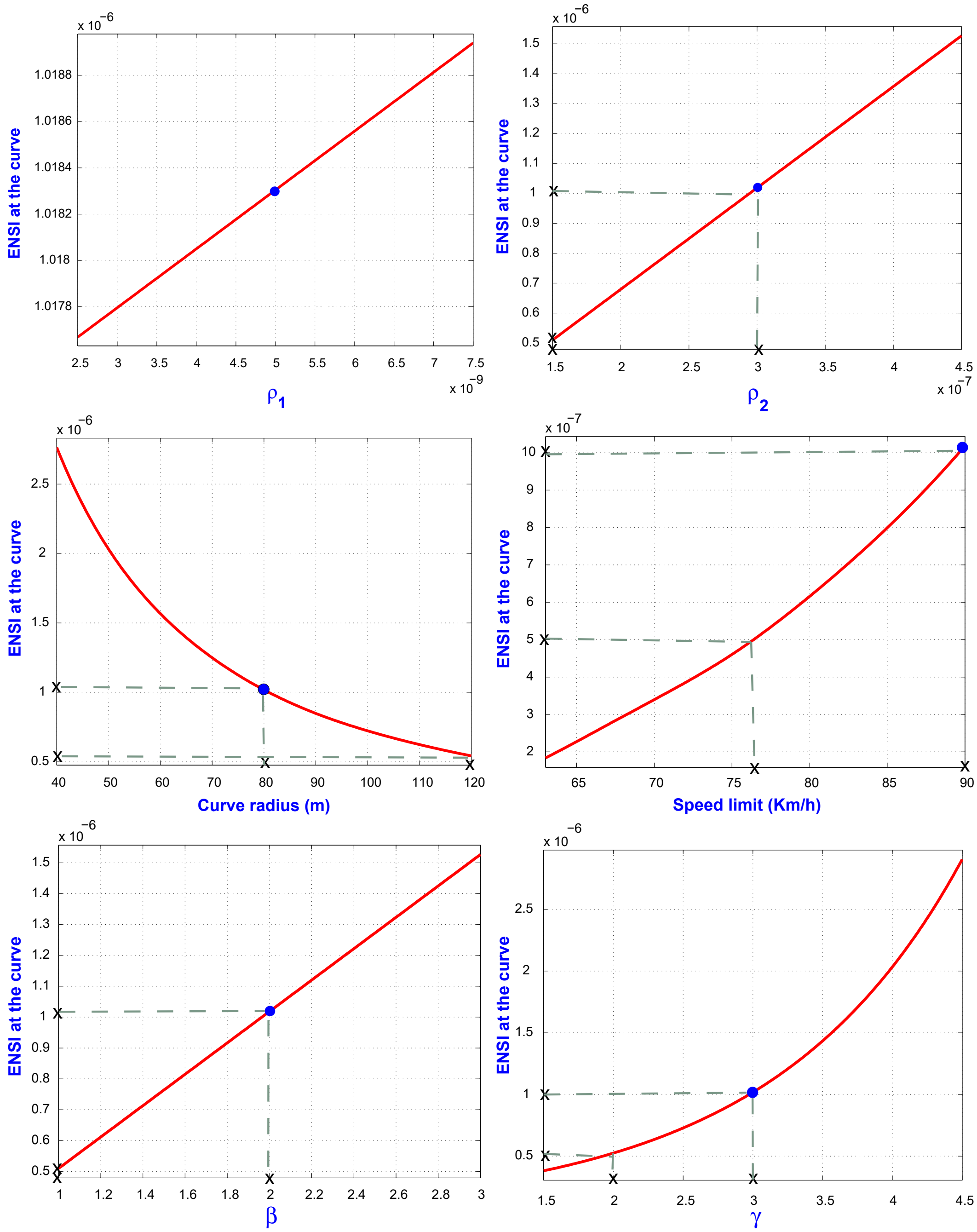


Figure 8.7: Upper plots: sensitivities of the ENSI at the curve with respect to parameters ρ_1 and ρ_2 in Equation (5.40). Medium plots: sensitivities of the ENSI at the curve with respect to curve radius and speed limit. Lower plots: sensitivities of the ENSI at the curve with respect to β and γ in Equation (5.40).

to one half. Similarly, a reduction of the speed limit from 90 Km/h to 77 km/h implies a reduction of the ENSI value to another half (see medium plots of Figure 8.7).

A change in ρ_1 implies almost no reduction and a reduction of ρ_2 from 3 to 1.5 implies a reduction of one half of ENSI values (see upper plots of Figure 8.7). If β changes from 2 to 1 the ENSI value is reduced in one half and when there is a reduction of γ from 3 to 2 the same reduction of ENSI value is produced (see lower plots of Figure 8.7). It is also noted that ENSI values change linearly with ρ_1, ρ_2 and β, while the changes with the curve radius, the speed limit and γ are non linear.

Chapter 9

Prognosticate and causes of incidents

Contents

9.1 Prognosticate incident occurrence

Once the most critical locations of the road or highway have been identified, it is important to obtain the most important combinations of variable values, that is, the circumstances leading to severe incidents. In other words, the characteristics of the incidents at given locations are wanted to be known, that is, we would like to know answers to questions such as, for example, did the incidents occur in good or bad weather?, were the vehicles involved cars, motorcycles or trucks?, were the drivers attentive or distracted? and, was the road traffic intensity low or high?

To answer these and similar questions multi-dimensional marginal probabilities of selected subsets of variables must be worked with.

Of special interest are the joint probabilities of the subset of variables involved in the conditional probability tables. For example, consider the case of an incident I at a curve and its conditional probability $P(I|W, Vt, D, S)$, where W, Vt, D, S refer to weather, vehicle type, driver's attention and speed, respectively. Since

$$P(I|W, Vt, D, S) = \frac{P(I, W, Vt, D, S)}{P(W, Vt, D, S)}, \tag{9.1}$$

the joint probability of (I, W, Vt, D, S) is obtained as

$$P(I, W, Vt, D, S) = P(I|W, Vt, D, S)P(W, Vt, D, S), \tag{9.2}$$

from which the ENSI values can be achieved.

It is noted that calculation of the joint probability (9.2) is not a problem because it involves members of a family (a child I and its parents W, Vt, D and S).

If the $ENSI(W, Vt, D, S)$ values are sorted in decreasing order of magnitude, the combinations of the random variables W, Vt, D and S leading to the most probable severe incidents can be identified. This permits, not only to enumerate the most relevant sequences of events with serious hazards and to quantify their probabilities of occurrence, but to have a picture of the most frequent situations leading to those incidents and consequently to address the corrective actions to the most frequent causes of each incident.

Thanks to the inference engine of the BN, it is possible to determine the most likely combinations of variables that give rise to each incident analyzed. To illustrate this, an example of a Spanish road is exposed below.

9.1.1 Example of application: Prognosticate of incidents on the National road N-611

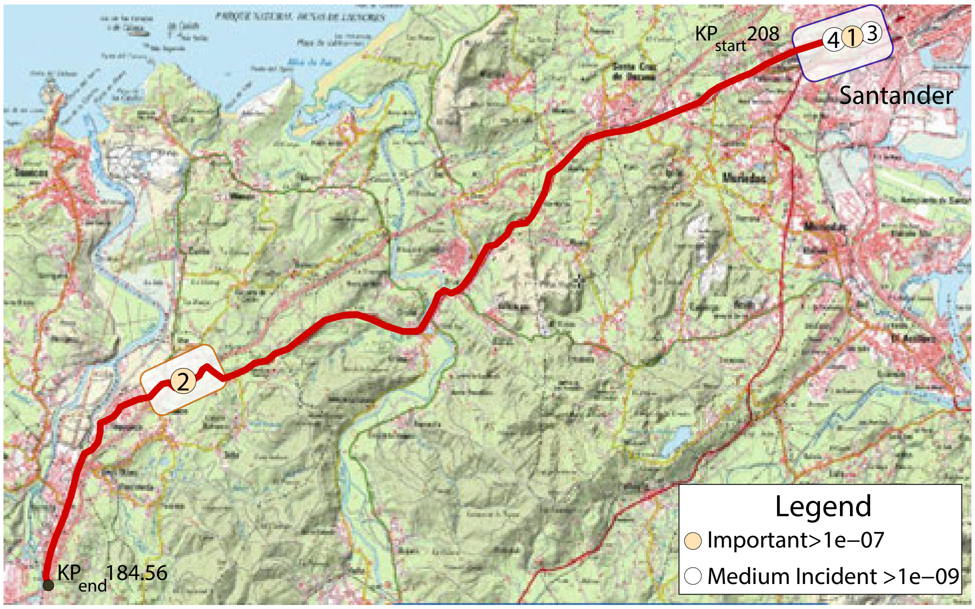
This illustrative example of the Spanish National road N-611 from Santander to Torrelavega, from KP 208.000 to KP 184.560, with total length of 23.440 km (see Figure 9.1), in addition to showing results obtained with the presented methodology, explains the part of the prognosticate incident occurrence in a practical way. It has been modelled with 586 items and 7917 variables. Figure 9.1 shows the road plant where the positions of the most hazardous locations and black spots are represented, highlighting two areas (rectangles) which represent the main critical segments of the road. The lower part of this figure shows the riskiest item locations, three curves at KP 207.550, KP 190.390 and KP 207.900 with ENSI values of $2.05e-07$, $1.96e-07$ and $6.66e-08$, respectively[1], and a T intersection located at KP 207.160 with ENSI $= 1.42e-08$, together with their associated yearly ENSI values, 0.337, 0.332, 0.109 and 0.0234, respectively[2].

The critical list table of the road which represents the ENSI values corresponding to a single trip of the road and which contains the list of the items sorted by their incident risk levels with their corresponding KP, node identifications and local probabilities is represented in Table 9.1. Note that only the items whose ENSI values exceed $1.0e-9$ are listed and the most hazardous locations of the road can be immediately identified. It can be seen that the first four most critical items correspond to the three curves and the T intersection mentioned above. The remaining items in the list correspond to curves, lateral entries, yield signs and intersections.

Table 9.2 contains the list of ENSI values grouped by item types and sorted by decreasing order of ENSI values. It reveals that the 35 curves, the 79 lateral entries, the 97 intersections and the 19 yield signs are those contributing more to the N-611 road risks.

[1]This refers to the expected number of severe incidents each time we travel this road.

[2]This refers to the expected number of severe incidents suffered by all vehicles using this road during one year.



Rank	Item	Item name	KP	Node	Local Probability	ENSI per year
1	14	**CurveIn**	**207.550**	I185-207550Cv	**2.05e-07**	**0.337**
2	440	**CurveIn**	**190.390**	I5876-190390Cv	**1.96e-07**	**0.322**
3	5	**CurveIn**	**207.900**	I64-207900Cv	**6.66e-08**	**0.109**
4	24	**Intersection**	**207.160**	I329-207160Int	**1.42e-08**	**0.0234**

Figure 9.1: Plant of the N-611 road, from Santander to Torrelavega, showing the four riskiest locations.

To display the results of the analysis the model builds some figures such as Figure 9.2, 9.3 and 9.4, which show the acyclic graph, the schematic representations of the road segments, the graph of the cumulated ENSI values, and also include tables at the bottom of the figures with the details of the item locations and their ENSI values, local and cumulated for the segment considered.

Figure 9.2 shows, in particular, the acyclic graph corresponding to the segment ranging between KP 207.770 and KP 207.500 and the corresponding road items.

Figure 9.3 also represents the segment ranging between KP 207.770 and KP 207.500 and the corresponding road items, in which the most hazardous point, the curve of radius 240 m at KP 207.550 appear, with a table at the bottom of the figure with the details of the item locations and their ENSI values, local and cumulated for this segment. Finally, Figure 9.4 represents the same information for the most hazardous intersection, a T intersection at KP 207.160.

This collection of figures, graphs and tables for all segments provides a detailed profile of the road, which permits a clear and precise probabilistic assessment of its safety to be performed.

Table 9.1: National road N-611. Critical list of items sorted by their incident risk levels with their corresponding KP, nodes and local ENSI values.

Rank	**Item**	**Item name**	**KP**	**Node**	**ENSI Local**
1	14	**CurveIn**	**207.550**	I185-207550Cv	**2.05e-07**
2	440	**CurveIn**	**190.390**	I5876-190390Cv	**1.96e-07**
3	5	**CurveIn**	**207.900**	I64-207900Cv	**6.66e-08**
4	24	**Intersection**	**207.160**	I329-207160Int	**1.42e-08**
5	467	CurveIn	189.520	I6248-189520Cv	7.5e-09
6	430	CurveIn	190.690	I5748-190690Cv	5.75e-09
7	408	CurveIn	191.510	I5467-191510Cv	3.71e-09
8	449	Yield	189.945	I5996-189945Yl	3.57e-09
9	6	LateralEntry	207.895	I77-207895LE	2.68e-09
10	3	LateralEntry	207.935	I38-207935LE	2.61e-09
11	361	LateralEntry	193.162	I4871-193162LE	2.41e-09
12	177	Yield	201.724	I2456-201724Yl	2.17e-09
13	127	Yield	203.451	I1767-203451Yl	2.17e-09
14	89	Yield	204.567	I1234-204567Yl	2.11e-09
15	7	LateralEntry	207.785	I90-207785LE	2.06e-09
16	548	Intersection	186.068	I7364-186068Int	1.91e-09
17	299	LateralEntry	196.665	I4078-196665LE	1.79e-09
18	349	LateralEntry	193.770	I4717-193770LE	1.79e-09
19	115	Intersection	203.854	I1604-203854Int	1.58e-09
20	495	LateralEntry	188.654	I6626-188654LE	1.53e-09
21	294	LateralEntry	197.060	I4013-197060LE	1.47e-09
22	166	LateralEntry	202.185	I2302-202185LE	1.44e-09
23	62	Yield	205.235	I860-205235Yl	1.43e-09
24	293	LateralEntry	197.140	I4000-197140LE	1.41e-09
25	448	LateralEntry	189.985	I5979-189985LE	1.4e-09
26	81	Intersection	204.732	I1124-204732Int	1.3e-09
27	490	LateralEntry	188.949	I6562-188949LE	1.23e-09
28	120	LateralEntry	203.716	I1674-203716LE	1.17e-09
29	164	LateralEntry	202.315	I2276-202315LE	1.17e-09
30	21	LateralEntry	207.255	I280-207255LE	1.1e-09
31	19	LateralEntry	207.350	I254-207350LE	1.07e-09
32	34	Intersection	206.565	I465-206565Int	1.02e-09

$ENSI^*$ Expected Number of Severe Incidents

Table 9.2: National road N-611. Grouped incident list of items with their frequencies and total ENSI values.

Incident types	Frequency	Local ENSI
Curve	35	4.87e-07
Lateral Entry	79	5.15e-08
Intersection	97	3.5e-08
Yield	19	2.02e-08
SegWSignal	585	1.58e-09
Traffic Light	29	9.28e-10
Roundabout	15	1.21e-10
Overpass	6	8.76e-11
Pedestrian Crossing	19	7.09e-11
Underpass	4	2.06e-11
Viaduct	2	9.87e-12
$ENSI^{*}$ Expected Number of Severe Incidents		

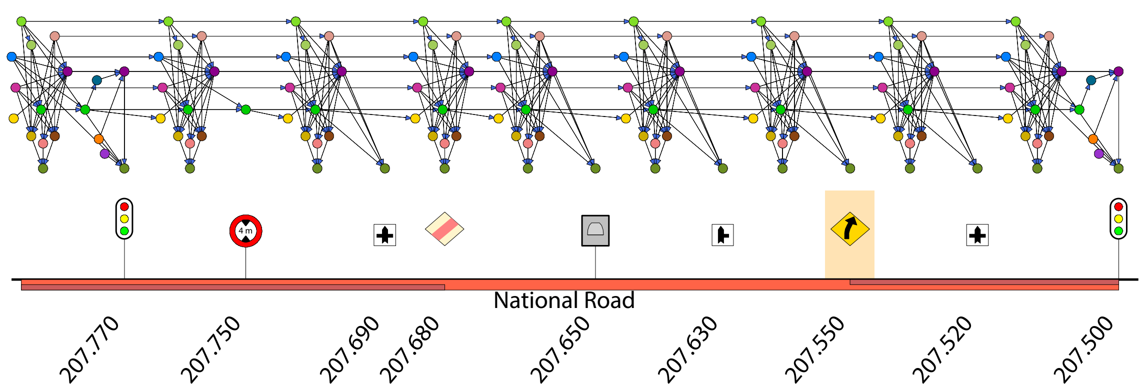
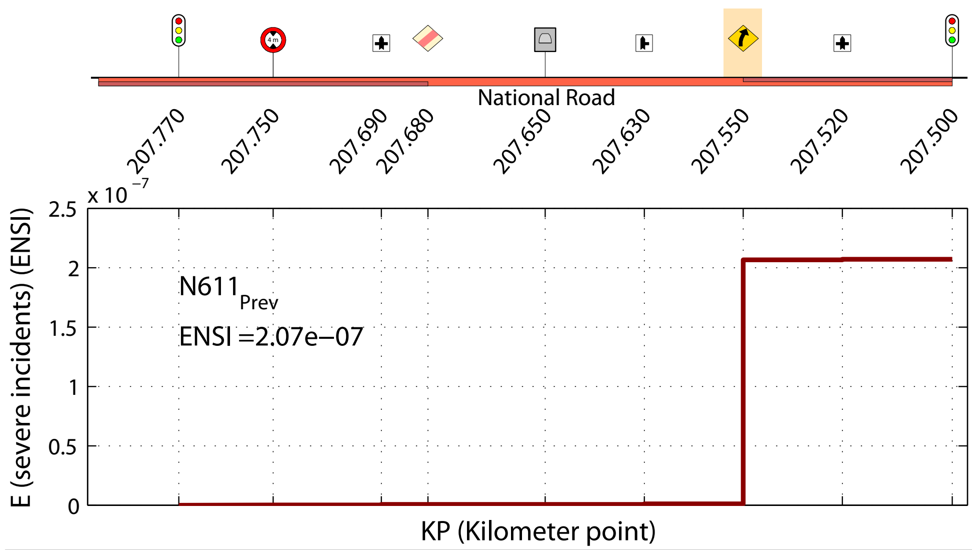
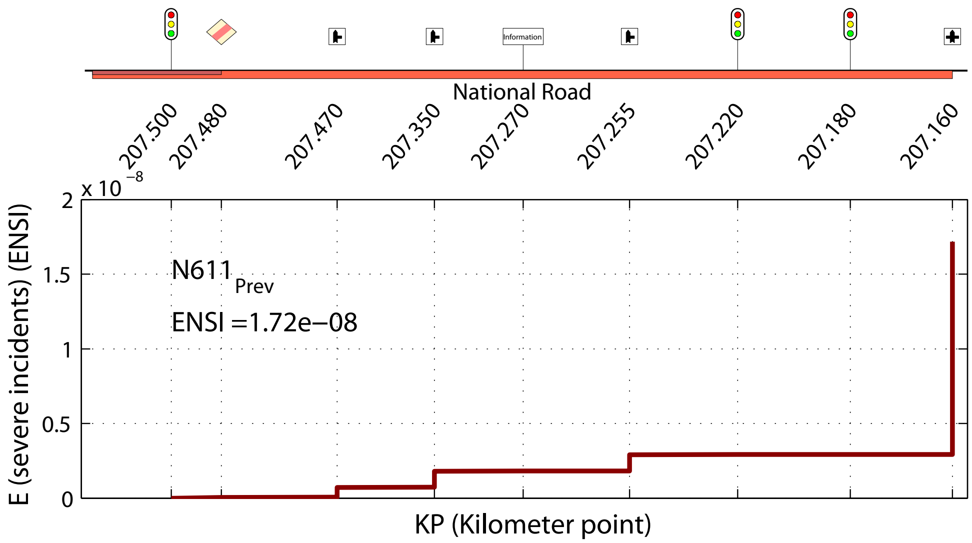


Figure 9.2: National road N-611. Acyclic graph corresponding to segment ranging between KP 207.770 and KP 207.500.



				Probability	
Item	**KP**	**Item name**	**Node**	**Local**	**Cumulated**
8	207.770	SegWSignals	I102-207770SCr	3.28e-11	7.5e-08
		TrafficLight	I108-207770TL	2.52e-12	7.5e-08
9	207.750	SegWSignals	I120-207750SCr	4.27e-11	7.51e-08
10	207.690	SegWSignals	I133-207690SCr	2.52e-11	7.51e-08
		Intersection	I134-207690Int	4.56e-10	7.56e-08
11	207.680	SegWSignals	I146-207680SCr	4.49e-12	7.56e-08
12	207.650	SegWSignals	I158-207650S	2.98e-12	7.56e-08
		Overpass	I159-207650Ov	4.78e-12	7.56e-08
13	207.630	SegWSignals	I171-207630S	2.02e-12	7.56e-08
		LateralEntry	I172-207630LE	5.35e-10	7.61e-08
14	207.550	SegWSignals	I184-207550S	9.5e-12	7.61e-08
		CurveIn	I185-207550Cv	**2.05e-07**	2.82e-07
15	207.520	SegWSignals	I197-207520SCr	2.8e-11	2.82e-07
		Intersection	I198-207520Int	5.05e-10	2.82e-07
16	207.500	SegWSignals	I210-207500SCr	1.87e-11	2.82e-07
		TrafficLight	I216-207500TL	4.62e-13	2.82e-07

Figure 9.3: National road N-611. Data corresponding to segment ranging between KP 207.770 and KP 207.500.



Item	KP	Item name	Node	Probability Local	Probability Cumulated
16	207.500	SegWSignals	I210-207500SCr	1.87e-11	2.82e-07
		TrafficLight	I216-207500TL	4.62e-13	2.82e-07
17	207.480	SegWSignals	I228-207480SCr	5.39e-11	2.82e-07
18	207.470	SegWSignals	I240-207470S	3.29e-12	2.82e-07
		LateralEntry	I241-207470LE	6.64e-10	2.83e-07
19	207.350	SegWSignals	I253-207350S	1.17e-11	2.83e-07
		LateralEntry	I254-207350LE	1.07e-09	2.84e-07
20	207.270	SegWSignals	I266-207270S	9.05e-12	2.84e-07
21	207.255	SegWSignals	I279-207255S	2.26e-12	2.84e-07
		LateralEntry	I280-207255LE	1.1e-09	2.85e-07
22	207.220	SegWSignals	I292-207220S	5.82e-12	2.85e-07
		TrafficLight	I298-207220TL	6.24e-13	2.85e-07
23	207.180	SegWSignals	I310-207180S	5.88e-12	2.85e-07
		TrafficLight	I316-207180TL	2.03e-13	2.85e-07
24	207.160	SegWSignals	I328-207160S	5.25e-12	2.85e-07
		Intersection	I329-207160Int	**1.42e-08**	2.99e-07

Figure 9.4: National road N-611. Data corresponding to segment ranging between KP 207.500 and KP 207.160.

In order to show how the prognosticate incident occurrence is employed by the proposed model, the most critical point of this illustrative example, a curve of 240 m at KP 207.550 with a speed limit of 90 km/h (road maximum speed), is analyzed. The point is represented in plant as point 1 in Figure 9.1 where the risk of a severe incident at this location results in 2.05e-07 ENSI per trip[3] or 0.337 ENSI per year[4] for all circulating vehicles and it can be appreciated in Figure 9.3.

When the most likely combinations of variable values that give rise to incidents at the curve located at KP 207.550 are evaluated, Table 9.3 which represents the most frequent combinations of values for the variables weather W, vehicle type Vt, driver's attention D, and speed S, that lead to incidents at the curve with frequencies indicated by the ENSI values and the contributions of each combination given in % is obtained.

It is quite surprising that the most likely combinations, that is, 55.822 of the times, correspond to rainy days (medium weather) and not to bad or very bad weather, for example snow, and with drivers attentive or alert to driving and not distracted. This is due to the fact that bad and very bad weather are much less likely than good and medium weather. Additionally, under these weather conditions drivers would not circulate unless necessary, which implies a low traffic intensity and a small driver distraction probability.

Since incidents imply a complex evaluation of all combinations of values of a large set of variables, engineers have not a correct intuition about which combination of variable values leads to more frequent incidents. Thus, the Bayesian network inference engine becomes crucial to help us identify these combinations and proceed consequently to choose the adequate actions or interventions to improve safety when needed. Discovering the real causes of incidents and their associated frequencies is very relevant to reduce risk and save maintenance costs. If it is believed, for example, that incidents are caused mainly by novice drivers with bad weather when traffic intensity is high, the actions will be oriented to them instead of addressing the corrections to the most frequent combinations leading to a reduction in the efficiency of the correction procedure. All this shows the relevance of a correct identification of combinations of events leading to incidents.

If the severity of the incident and the probability per trip instead of the ENSI values are included in the table, a different picture of the problem can be obtained. By analyzing the severity of the potential incidents, frequent occurrence not only of severe but medium or minor incidents can be detected too. This is the case of Table 9.4 in which the 5th case corresponds to a frequent potential medium incident. Note that the occurrence probability of the 5th case is larger than those of the 3rd and 4th cases, but the opposite is true for the ENSI values.

Since a speed excess problem at the curve located at KP 207.550 has been identified, it is decided to change the speed limit sign at KP 207.800 from $90km/h$ to $60km/h$. A recalculation of probabilities shows that the ENSI value at the curve located at KP 207.550 drops considerably down to 7.41e-11, see Figure 9.5, and its most likely causes

[3]That is, two incidents every 10,000,000 trips

[4]This means one severe incident every three years

Table 9.3: ENSI Table. Combination of main variable values contributing to the total ENSI value of incident at the riskiest curve Node (ENSI=2.05454e-07).

n	Weather	Veh. type	Attention	Speed	ENSI	% ENSI
1	**medium**	**car**	**attentive**	**108**	1.15e-07	55.82
2	**medium**	**car**	**attentive**	**90**	5.29e-08	25.77
3	**medium**	**car**	**alert**	**108**	1.64e-08	8.00
4	**medium**	**car**	**alert**	**90**	8.71e-09	4.24
5	medium	car	attentive	121.5	3.11e-09	1.51
6	medium	heavy	attentive	90	3.07e-09	1.50
7	bad	car	attentive	90	2.06e-09	1.00
8	medium	heavy	alert	90	1.19e-09	0.58
9	medium	motorcycle	attentive	108	1.19e-09	0.58
10	bad	car	alert	90	5.88e-10	0.29

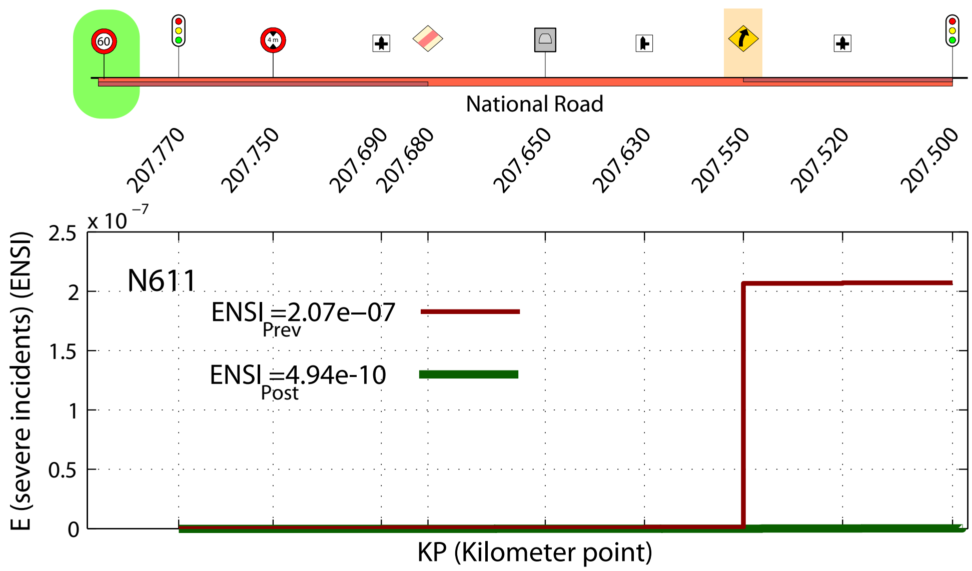
Table 9.4: Combination of main variable values contributing to the total ENSI value at the riskiest curve Node (ENSI=2.05454e-07) including the severity of incident and the probability per trip.

n	Weather	Veh. type	Attention	Speed	Incident	Probability	% ENSI
1	medium	car	attentive	108	severe	1.14e-07	55.71
2	medium	car	attentive	90	severe	4.97e-08	24.17
3	medium	car	alert	108	severe	1.64e-08	7.98
4	medium	car	alert	90	severe	8.17e-09	3.98
5	medium	car	attentive	90	**medium**	**3.27e-08**	1.59
6	medium	car	attentive	121.5	severe	3.11e-09	1.51
7	medium	heavy	attentive	90	severe	2.95e-09	1.44
8	bad	car	attentive	90	severe	2.02e-09	0.98
9	medium	motorcycle	attentive	108	severe	1.19e-09	0.58
10	medium	heavy	alert	90	severe	1.15e-09	0.56
11	bad	car	alert	90	severe	5.78e-10	0.28
12	medium	car	alert	90	medium	5.38e-09	0.26

also change significantly, as shown in Table 9.5.

The curious result is that now severe incidents are mainly caused by driver distraction, unfavorable weather conditions, and medium and minor incidents correspond to attentive drivers[5]. Hence, with this intervention (a speed limitation to 60 km/h) the risk associated with the curve becomes reasonable because:

[5]It can be seen that with the previous speed limit incidents cannot be due to drivers, while with the new speed limit reduction, incidents are caused by drivers (speed and distraction).



Item	KP	Item name	Node	Probability Local	Probability Cumulated
8	207.785	SegWSignals	I104-207785SCr	7.41e-12	2.21e-09
		LateralEntry	I105-207785LE	7.42e-10	2.95e-09
9	207.770	SegWSignals	I117-207770SCr	1e-12	2.95e-09
		TrafficLight	I123-207770TL	9.62e-13	2.95e-09
10	207.750	SegWSignals	I135-207750SCr	3.99e-11	2.99e-09
11	207.690	SegWSignals	I148-207690SCr	8.93e-13	2.99e-09
		Intersection	I149-207690Int	8.51e-11	3.08e-09
12	207.680	SegWSignals	I161-207680SCr	1.51e-13	3.08e-09
13	207.650	SegWSignals	I173-207650S	1.01e-13	3.08e-09
		Overpass	I174-207650Ov	4.75e-12	3.09e-09
14	207.630	SegWSignals	I186-207630S	7.15e-14	3.09e-09
		LateralEntry	I187-207630LE	1.95e-10	3.28e-09
15	207.550	SegWSignals	I199-207550S	3.28e-13	3.28e-09
		CurveIn	I200-207550Cv	**7.41e-11**	3.35e-09
16	207.520	SegWSignals	I212-207520SCr	9.91e-13	3.36e-09
		Intersection	I213-207520Int	9.04e-11	3.45e-09

Figure 9.5: Data corresponding to segment ranging between KP 207.770 and KP 207.500 after installing a speed limit sign (60 km/h) at KP 207.800.

- Global ENSI has dropped from 2.05e-07 to 7.41e-11 (2770 times safer).
- Severe incidents have decreased significantly, emerging only medium and minor in-

Table 9.5: Combination of main variable values contributing to the total ENSI value at the riskiest curve Node (ENSI=7.41208e-11) after installing a speed limit sign at KP 207.800.

n	Weather	Veh. type	Attention	Speed	Incident	Probability	% ENSI
1	medium	car	attentive	81	**medium**	2.47e-10	33.29
2	medium	car	**distracted**	90	severe	2.02e-11	27.25
3	medium	car	**distracted**	81	severe	1.9e-11	25.60
4	medium	car	attentive	81	**minor**	1.31e-09	3.53
5	medium	car	attentive	90	severe	1.1e-12	1.48
6	bad	car	distracted	90	severe	9.23e-13	1.25
7	medium	car	alert	81	medium	6.97e-12	0.94
8	medium	motorcycle	attentive	81	medium	6.03e-12	0.81
9	fair	car	distracted	72	severe	5.55e-13	0.75
10	bad	car	distracted	81	severe	5.13e-13	0.69
11	fair	car	distracted	81	severe	4.1e-13	0.55

cidents.

- Severe incidents originate only from distracted drivers.

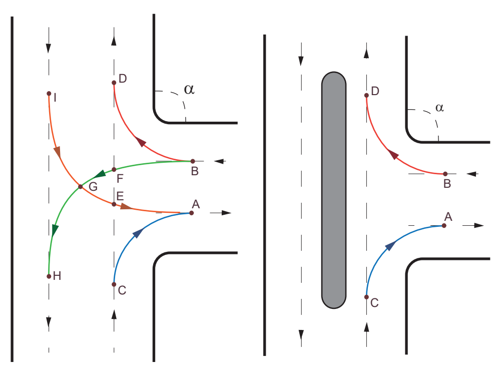


Figure 9.6: Possible conflict point in a T intersection (left) and T with median intersection (right).

Next, the T intersection at KP 207.160 (see Figure 9.4) is considered and the associated most likely causes are shown in Table 9.6. Unexpectedly, in this case the accidents frequently occur with good and slightly rainy days (fair and medium weather), with attentive or alert car drivers and high and very low speeds. The low speed vehicles are those turning at the intersection. Hence, the conclusions that can be extracted from the analysis made are that severe incidents can occur due to (a) lateral or frontal collisions of high speed vehicles circulating across the main road with those entering or leaving the

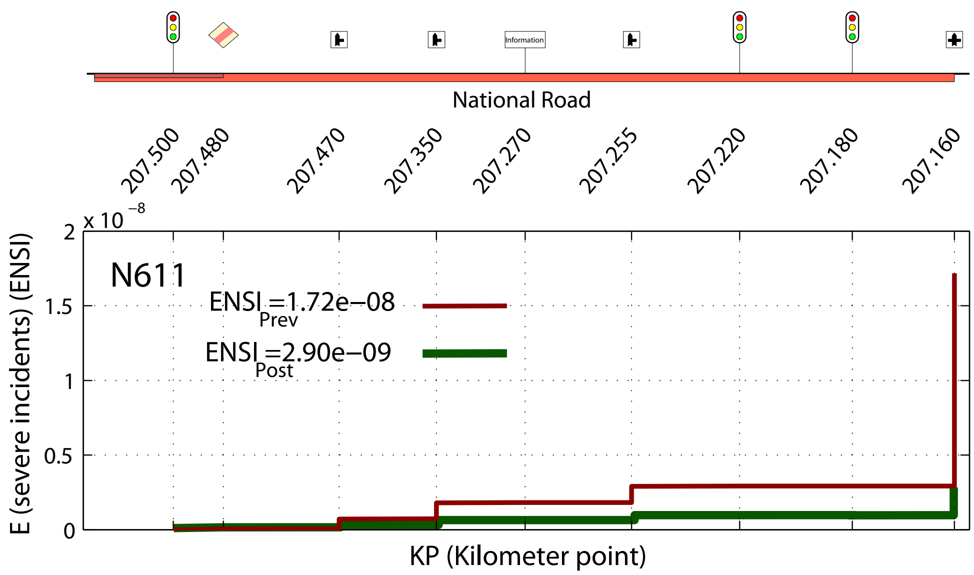
Table 9.6: Combination of main variable values contributing to the total ENSI value at the riskiest intersection Node (ENSI=1.42428e-08).

n	Weather	Veh. type	Attention	Speed	ENSI	% ENSI
1	**fair**	**car**	**alert**	**120**	1.85e-09	12.97
2	**fair**	**car**	**attentive**	**120**	1.51e-09	10.60
3	**medium**	**car**	**alert**	**36**	1.49e-09	10.49
4	**medium**	**car**	**alert**	**120**	1.47e-09	10.32
5	**fair**	**car**	**alert**	**36**	1.32e-09	9.27
6	medium	car	attentive	120	9.63e-10	6.76
7	medium	car	attentive	36	9.58e-10	6.73
8	fair	car	attentive	36	8.62e-10	6.05
9	bad	car	alert	36	3.16e-10	2.22
10	bad	car	alert	120	3.09e-10	2.17
11	very bad	car	alert	36	2.68e-10	1.88
12	very bad	car	alert	120	2.62e-10	1.84
13	fair	heavy	alert	36	2.25e-10	1.58
14	fair	heavy	alert	120	2.2e-10	1.54
15	fair	car	alert	99.4987	1.82e-10	1.28
16	medium	heavy	alert	36	1.73e-10	1.22
17	medium	car	alert	99.4987	1.71e-10	1.20
18	medium	heavy	alert	120	1.69e-10	1.19
19	fair	car	attentive	99.4987	1.43e-10	1.00
20	bad	car	attentive	36	1.1e-10	0.77
21	medium	car	attentive	99.4987	1.09e-10	0.76
22	bad	car	attentive	120	1.07e-10	0.75
23	fair	heavy	attentive	36	9.14e-11	0.64
24	fair	heavy	attentive	120	8.93e-11	0.63
25	medium	heavy	attentive	36	5.74e-11	0.40

main road or (b) rear collisions with vehicles circulating too slowly when incorporating or getting off the secondary road.

After an analysis of this situation it is concluded that one way of reducing the risk at the intersection is to reduce the speed and change the intersection type, from a T intersection to a T with median, which limits considerably the traffic switch, as shown in Figure 9.6. The result is that the risk decreases from $1.42e-08$ to $1.85e-09$ (7.68 times safer) as it can be seen in Figure 9.7.

The most likely causes after the correction are shown in Table 9.7, from which it can be concluded that incident occurrences decrease due to a speed reduction, but the incidents still take place by speed excesses with attentive or alert drivers and in not bad weather



				Probability	
item	**KP**	**item name**	**node**	**Local**	**Cumulated**
17	207.500	SegWSignals	I225-207500SCr	7.2e-13	3.45e-09
		TrafficLight	I231-207500TL	4.21e-13	3.45e-09
18	207.480	SegWSignals	I243-207480SCr	6.13e-11	3.51e-09
19	207.470	SegWSignals	I255-207470S	2.69e-12	3.51e-09
		LateralEntry	I256-207470LE	1.88e-10	3.7e-09
20	207.350	SegWSignals	I268-207350S	5.46e-13	3.7e-09
		LateralEntry	I269-207350LE	3.9e-10	4.09e-09
21	207.270	SegWSignals	I281-207270S	3.92e-13	4.09e-09
22	207.255	SegWSignals	I294-207255S	9.44e-14	4.09e-09
		LateralEntry	I295-207255LE	4e-10	4.49e-09
23	207.220	SegWSignals	I307-207220S	2.55e-13	4.49e-09
		TrafficLight	I313-207220TL	6.91e-13	4.49e-09
24	207.180	SegWSignals	I325-207180S	1.54e-12	4.49e-09
		TrafficLight	I331-207180TL	1.76e-13	4.49e-09
25	207.160	SegWSignals	I343-207160S	6.18e-12	4.5e-09
		Intersection	I344-207160Int	**1.85e-09**	6.34e-09

Figure 9.7: Data corresponding to segment including the riskiest intersection (KP 207.160) after improvements.

days.

In summary, this tool could be very useful for determining the conditions under which potential accidents occur and for organizing corrective actions in a rational and efficient

Table 9.7: Combination of main variable values contributing to the total ENSI value at the riskiest intersection Node (ENSI=1.84514e-09) after improvements.

n	Weather	Veh. type	Attention	Speed	ENSI	% ENSI
1	**fair**	car	alert	**107**	2.19e-10	11.89
2	**fair**	car	attentive	**107**	2.15e-10	11.65
3	**medium**	car	alert	**107**	2.01e-10	10.87
4	**medium**	car	attentive	**107**	1.59e-10	8.64
5	fair	car	alert	24	1.3e-10	7.04
6	fair	car	attentive	24	1.28e-10	6.95
7	medium	car	alert	24	1.18e-10	6.38
8	medium	car	attentive	24	9.35e-11	5.07
9	fair	car	attentive	90	5.98e-11	3.24
10	fair	car	alert	90	5.97e-11	3.23
11	medium	car	alert	90	5.34e-11	2.89
12	bad	car	alert	107	4.32e-11	2.34
13	medium	car	attentive	90	4.24e-11	2.30
14	very bad	car	alert	107	3.68e-11	2.00
15	fair	heavy	alert	107	3.03e-11	1.64
16	bad	car	alert	24	2.53e-11	1.37
17	medium	heavy	alert	107	2.36e-11	1.28
18	very bad	car	alert	24	2.2e-11	1.19
19	bad	car	attentive	107	1.84e-11	1.00
20	fair	heavy	alert	24	1.78e-11	0.96
21	fair	heavy	attentive	107	1.56e-11	0.84
22	fair	car	alert	72	1.4e-11	0.76
23	fair	car	attentive	72	1.39e-11	0.75
24	medium	heavy	alert	24	1.38e-11	0.75
25	bad	car	alert	90	1.15e-11	0.62
26	bad	car	attentive	24	1.08e-11	0.59
27	very bad	car	alert	90	9.99e-12	0.54
28	medium	heavy	attentive	107	9.82e-12	0.53
29	medium	car	alert	72	9.5e-12	0.51
30	fair	heavy	attentive	24	9.13e-12	0.50
31	medium	car	attentive	72	8.14e-12	0.44
32	fair	heavy	alert	90	8.07e-12	0.44

manner.

9.2 Analysis of the causes of accidents by backward analysis

Another helpful possibility offered by the Bayesian network is backward inference. When an event (for example, a severe accident with specific characteristics) occurs at a certain place, the network modifies the probabilities of all variables and allows us to identify its causes, which would be difficult to detect without the help of this tool.

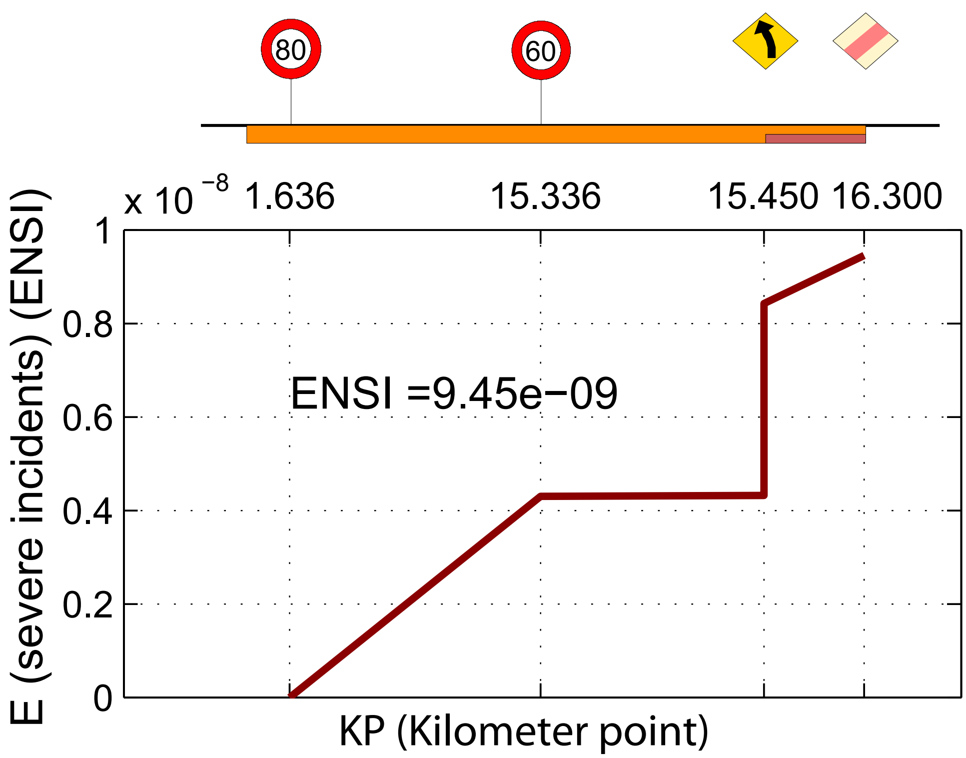


Figure 9.8: Segment of road to illustrate the backward propagation to determine the causes of accidents.

A simple segment represented in Figure 9.8 is used to show how it works. The ENSI value associated with the curve at KP 15.450 is 4.1×10^{-9}, showing that the probability of an incident at this location is very low (tables of conditional probability are shown in Figure 9.9). However, if a severe incident with a motorcycle and on a rainy day (medium weather) is observed at the curve, the most likely circumstances causing the incident are the following (see Figure 9.10):

- A vehicle circulating at an excessive speed (121.5 km/h) at the curve.
- A bad or novice driver.

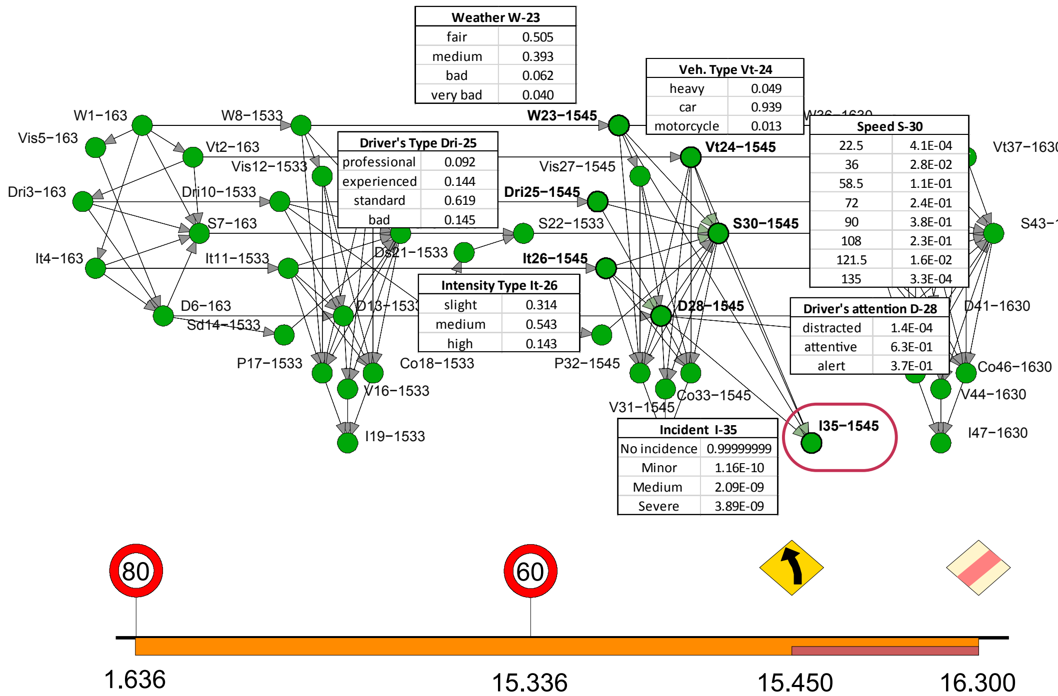


Figure 9.9: Tables of conditional probability before the evidence of an incident.

- An attentive driver.
- A low traffic intensity.

A bad driver causing the severe incident due to an excessive speed and a low traffic intensity may have been expected perhaps because it allows for higher speed, but it is not clear that an attentive driver's condition, for example, would have been expected. Consequently, the Bayesian network inference engine plays an important role in identifying these situations.

With this small example it is possible to see how the tables of conditional probability are recalculated after evidence of the severe incident with specific characteristics and these values are different from the initial (see Figure 9.9 and 9.10).

If, in addition, a severe incident with a motorcycle, on a rainy day (medium weather) and an excessive speed is identified now, the most likely circumstances become (see Figure 9.11):

- A bad (dare) driver.
- An attentive driver.
- A low traffic intensity.

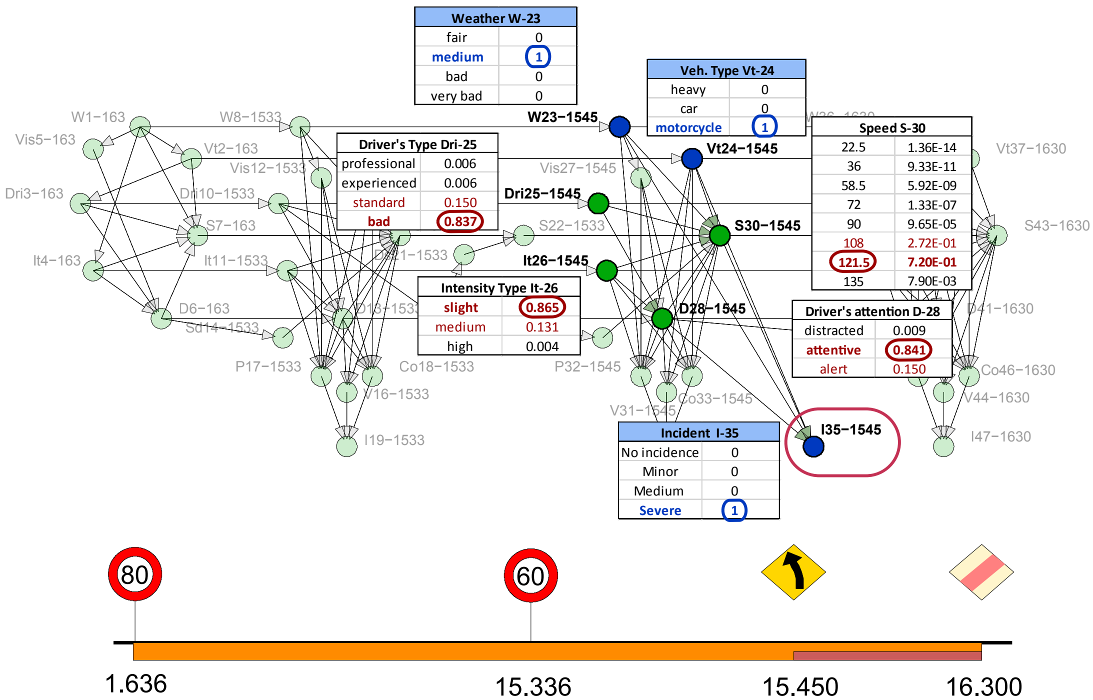


Figure 9.10: Tables of conditional probability recalculated after evidence of a severe incident of a motorcycle and a medium weather.

Finally, if the severe incident involves a motorcycle running at an excessive speed on a good day, the conclusions would be different because the most likely events would be (see Figure 9.12):

- A bad driver.
- A low traffic intensity.
- The almost certainty that the driver was distracted.

This type of analysis can only be done if the possibility of having access to the sub-networks such as the one indicated in Figure 9.9 and 9.10 are available. The software package that has been developed, in addition to providing a complete report with the acyclic graphs, a road graphical description and the accumulated ENSI values and their locations, provides the code for the JavaBayes of each of the sub-Bayesian network in which the initial Bayesian network is partitioned. This means that these analyses can be done at wish. Providing this information is considered an important original contribution and a relevant change for what information a safety report must contain. This extends safety analysis into a new direction.

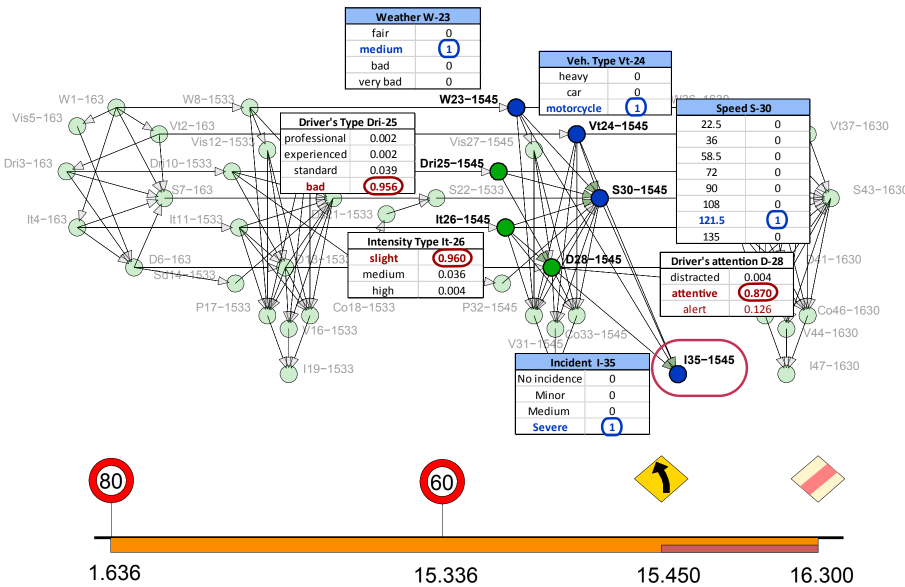


Figure 9.11: Tables of conditional probability recalculated after evidence of a severe incident of a motorcycle one day with medium weather and a speed of 121.5 km/h.

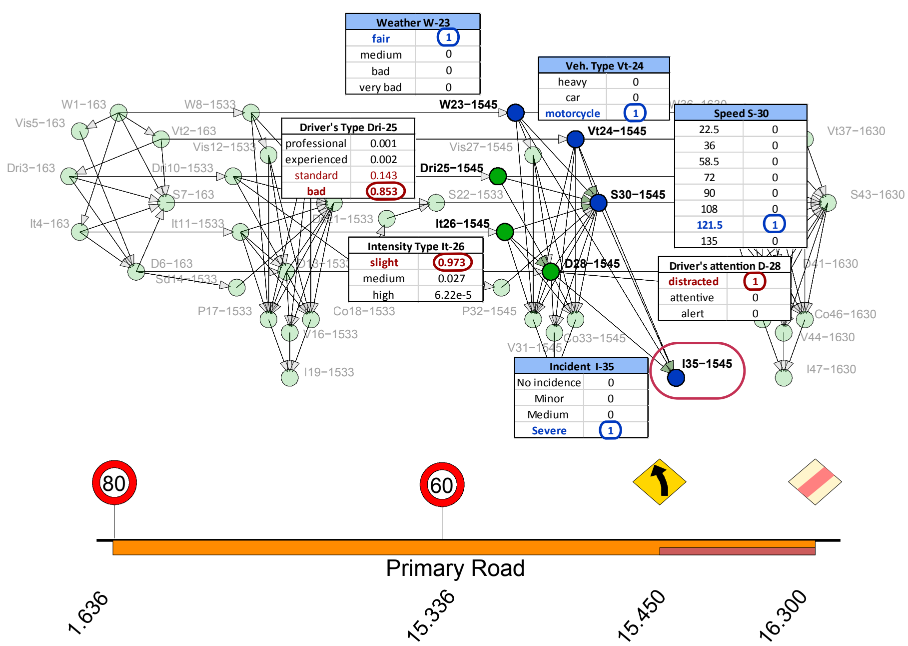


Figure 9.12: Tables of conditional probability recalculated after evidence of a severe incident of a motorcycle one day with good weather and a speed of 121.5 km/h.

Chapter 10

Learning the Bayesian network model

Contents

10.1 Introduction

In this section the problem of parametric learning in the Bayesian network model is adressed. There are two types of learning in probability models, namely, *structural learning* and *parametric learning*. Structural learning is concerned with learning the acyclic graph whose topology can be used to determine the qualitative relationships among a set of variables. More precisely, it is concerned with identifying which links must be included in the acyclic graph and which ones must not be incorporated. Bayesian networks usually contain many conditional independence relationships which lead to a substantial reduction (parsimony) in the number of links of the graph compared with other more general models. Contrary, parametric learning is concerned with estimating the conditional probability tables once the acyclic graph is given. This is the only learning type that has been considered in this chapter.

The likelihood function of the Bayesian network can be written as

$$L(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_n) = \prod_{k=1}^{n} P_k(x_k|\mathcal{P}^k; \theta_k), \tag{10.1}$$

where θ_k are vectors of parameters, in fact, those associated with node k, $\mathcal{P}^k$ refers to the parents of node X_k and $P_k(x_k|\mathcal{P}^k; \theta_k)$ refers to the conditional probability of node X_k given its parents.

Then, the likelihood of an i.i.d. sample of size m is given by

$$\begin{aligned} L &= \prod_{j=1}^{m} L(x_{1j}, x_{2j}, \ldots, x_{nj}; \theta_1, \theta_2, \ldots, \theta_n) \\ &= \prod_{j=1}^{m} \prod_{k=1}^{n} P_k(x_{kj}|\mathcal{P}_j^k; \theta_k), \end{aligned} \tag{10.2}$$

and the log-likelihood of the sample can be written as

$$\begin{aligned} \log L &= \sum_{j=1}^{m} \log L(x_{1j}, x_{2j}, \ldots, x_{nj}; \theta_1, \theta_2, \ldots, \theta_n) \\ &= \sum_{j=1}^{m} \sum_{k=1}^{n} \log P_k(x_{kj}|\mathcal{P}_j^k; \theta_k) \\ &= \sum_{k=1}^{n} \left(\sum_{j=1}^{m} \log P_k(x_{kj}|\mathcal{P}_j^k; \theta_k) \right) \\ &= \sum_{k=1}^{n} \sum_{j=1}^{m} \sum_{y_{jn}=1}^{l} \log P_k(x_{kj}|y_{jk}; \theta_k) \end{aligned} \tag{10.3}$$

where $\mathcal{P}_j^k$ refers to the parent values of the j-th sample data point and y_{jn} is the set of values of $\mathcal{P}_j^k$.

Equation (10.3) demonstrates that the log-likelihood of the sample is the sum of the log-likelihoods associated with the conditional probabilities of all nodes. Since different conditional distributions refer to different situations, to simplify, it can be assumed that the parameter vectors $\theta_1, \theta_2, \ldots \theta_n$ appear each in only one conditional probability. Then, maximization of the log-likelihood of the sample is equivalent to maximization of each of the sample likelihoods associated with each conditional probability and the problem is separable. This has a high practical relevance from the point of view of complexity and CPU time required. The importance of maximizing conditional likelihoods was demonstrated by Grossman and Domingos (2004).

However, defining the conditional probability associated with any node X_k means that it is necessary to provide a large number ℓ of conditional probabilities. This ℓ is the number of possible combinations of the parent $\mathcal{P}^k$ values for each node X_k, but the important fact is that they can be given independently.

10.2 Categorical conditional probability tables

Before explaining the categorical conditional probability tables of the model, a brief explanation of concepts needed for its understanding is made.

10.2.1 The categorical distribution

The categorical distribution $cat(p_1, p_2, \ldots, p_s)$, which is a generalization of the Bernoulli distribution is a discrete distribution with probability mass function defined by

$$P(X = i) = p_i; \quad i = 1, 2, \ldots, s, \tag{10.4}$$

where $\sum_{i=1}^{s} p_i = 1$. It is used when an experiment with only s possible events, which are mutually exclusive, is dealt with.

10.2.2 Natural conjugate of a categorical distribution

Bayesian statisticians often work with conjugate priors, which are parametric families of distributions such that their associated priors and posteriors belong to the same families. The practical rationale is that the posterior parameters can be easily obtained from the prior parameters and the sample. The parameters of the conjugate family, are referred to as *hyperparameters*.

Arnold et al. (1993) and Arnold et al. (1996) show that, in addition to the classical conjugate families given by DeGroot and Schervish (2002), many others are possible and they characterize the most general family of conjugate distributions for exponential families, which include the categorical, multinomial and Dirichlet distributions.

The conjugate distribution of the categorical distribution is the Dirichlet distribution, this means that if the prior distribution of a categorical family $Cat(p_1, p_2, \ldots, p_k)$ is Dirichlet $D(\theta_1, \theta_2, \ldots, \theta_k)$, then the posterior distribution is also a Dirichlet $D(\theta'_1, \theta'_2, \ldots, \theta'_k)$, where

$$\theta'_i = \theta_i + n_i, \tag{10.5}$$

being n_i the number of observations with value i in a sample of size $n = \sum_{i=1}^{k} n_i$.

The posterior predictive probability mass function is given by

$$P(X = i|D) = \frac{\theta'_i}{\sum_{i=1}^{k} \theta'_i} = \frac{\theta_i + n_i}{\sum_{i=1}^{k} \theta_i + n}, \tag{10.6}$$

where D refers to data.

It is noted that when n goes to infinity it is had

$$\lim_{n\to\infty} P(X=i|D) = \lim_{n\to\infty} \frac{\theta_i + n_i}{\sum\limits_{i=1}^{k} \theta_i + n} = \frac{n_i}{n}, \tag{10.7}$$

that is, when there is a lot of information, the prior information has no influence and the predictive values approach the observed ones.

10.2.3 Learning the categorical conditional probabilities

Here how the Bayesian method above to learn all categorical conditionals can be used is shown. It is assumed that it is started with a prior set of conditional probabilities and that information from passing vehicles is obtained sequentially. To illustrate the learning and updating method for the conditional probability table $P(V_1|V_2,\ldots,V_\ell)$ of node V_1 when one observation is available, it is necessary to distinguish between three cases:

1. *The observation is complete.* This means that all variables $V_1 = a, V_2 = b, \ldots, V_\ell = c$ involved in a conditional probability table have been observed. In this case, only the conditional probability table with parents and values $V_2 = b, \ldots, V_\ell = c$ are updated and the observation to the particular observed value $V_1 = a$ and no observation to $V_1 \neq a$ are assigned, that is, the corresponding conditional probability $p(a|b,\ldots,c) = P(V_1 = a|V_2 = b,\ldots,V_\ell = c)$ is updated, using the formula

$$p(x|b,\ldots,c) = \begin{cases} \dfrac{\theta_{a;b,\ldots,c}+1}{\sum\limits_a \theta_{a;b,\ldots,c}+1} & if \quad x=a \\ \dfrac{\theta_{a;b,\ldots,c}}{\sum\limits_a \theta_{a;b,\ldots,c}+1} & if \quad x\neq a \end{cases} \quad \forall b,\ldots,c. \tag{10.8}$$

Note that only conditional probability tables with parents and values $V_2 = b, \ldots, V_\ell = c$ are updated and that each table requires updating $\sum\limits_a 1$ values. In fact, only the parameter $\theta_{a;b,\ldots,c}$ needs to be updated, using the formula

$$\theta_{a;b,\ldots,c} = \theta_{a;b,\ldots,c} + 1. \tag{10.9}$$

Once the parameters are updated, the conditional probabilities become

$$p(a|b,\ldots,c) = \frac{\theta_{a;b,\ldots,c}}{\sum\limits_a \theta_{a;b,\ldots,c}}. \tag{10.10}$$

2. *The observation is incomplete but V_1 is observed.* Let $|Uobs|$, $|Obs|$ and $|V_1|$ be the numbers of possible combinations of values for the unobserved variables, the observed variables and V_1, respectively. Since in this case there are some variables in the set $V_2, \ldots, V_\ell$ that have not been observed, it is necessary to consider the joint probability mass $q(b, \ldots, c)$ of $V_2, \ldots, V_\ell$ and modify Formula (10.8) as

$$p(x|b, \ldots, c) = \begin{cases} \dfrac{\theta_{a;b,\ldots,c} + \beta_{Uobs}}{\sum\limits_a \theta_{a;b,\ldots,c} + 1} & if \quad x = a \\ \dfrac{\theta_{a;b,\ldots,c}}{\sum\limits_a \theta_{a;b,\ldots,c} + 1} & if \quad x \neq a \end{cases} \quad \forall Uobs, \tag{10.11}$$

where

$$\beta_{Uobs} = P(Uobs|Obs) = \frac{q(b, \ldots, c)}{\sum\limits_{S_{Uobs}} q(b, \ldots, c)}, \tag{10.12}$$

being S_{Obs} and S_{Uobs} the subsets of $\{b, \ldots, c\}$ associated with the observed and unobserved variables, respectively.

In this case, it is necessary to update $|Uobs|$ conditional probability tables and for each table, $|V_1|$ probabilities. However, if only the parameters are updated, it is necessary to update only $|Uobs|$ parameters.

It is noted that the joint probability $q(b, \ldots, c)$ can be easily obtained because the parents of a node are contained in a family of the Bayesian network.

3. *The observations are incomplete and V_1 is not observed.* In this case nothing is updated because the parameters and the conditional probability tables do not change.

10.2.4 Categorical conditional probability tables of the model

Since it is worked with categorical data, each conditional probability table is categorical. Though arbitrary sets of conditional probabilities cannot be given independently because they can lead to inconsistencies (see Arnold et al. (2001)), the set of conditional probabilities used by BN always leads to compatible models. All this means that they can be learned independently without compatibility concerns.

A formula for the conditional probability $p_{a,b_1,b_2,\ldots,b_k}$ of a child node X given its parents is proposed as follows

$$p_{a,b_1,b_2,\ldots,b_k} = \sum_{j_1} q_{j_1}(\theta_{j_1}) \left[\sum_{j_2} q_{j_1,j_2}(\theta_{j_1,j_2}) \left[\ldots \right.\right.$$
$$\left.\left.\left[\sum_{j_{s_{r-1}}} q_{j_1,\ldots,j_{s_{r-1}}}(\theta_{j_1,\ldots,j_{s_{r-1}}}) \left[\sum_{j_{s_r}}^{n_a} q_{j_1,\ldots,j_{s_r}}(\theta_{j_1,\ldots,j_{s_r}}) \delta_{a j_{s_r}} \right]\right]\right]\right], \tag{10.13}$$

where δ is the Kronecker's delta, a refers to the child node X, and $b_1, b_2, \ldots, b_k$ to its parents, and the right hand side term is a sum of products, the index j_{s_r}, which depends on the summand being considered, refers to the last factor in each summand, n_a is the cardinal of the set of values of the child node X, $\theta_{j_1,j_2,\ldots,j_{s_t}}$ are vectors of parameters, and all $q_{j_1,j_2,\ldots,j_{s_t}}; t = 1, 2, \ldots, r$ are non-negative valued functions of a subset of parents of the child node X, that is, whose arguments are a subset of $\{b_1, b_2, \ldots, b_k\}$, and they must satisfy

$$\sum_{j_t} q_{j_1,\ldots,j_t}(\theta_{j_1,\ldots,j_t}) = 1; \quad \forall j_1, j_2, \ldots, j_t; t = 1, 2, \ldots, s_r. \tag{10.14}$$

Note that the last, and only the last summation, contains the child values a. This guarantees that $p_{a,b_1,b_2,\ldots,b_k}$ is a valid conditional probability table, because all terms are non-negative, there are no minus signs and they add up to one, that is, it is had

$$\begin{aligned}\sum_{a=1}^{n_a} p_{a,b_1,b_2,\ldots,b_k} &= \sum_{a=1}^{n_a}\sum_{j_1} q_{j_1}(\theta_{j_1}) \left[\sum_{j_2} q_{j_1,j_2}(\theta_{j_1,j_2} [\ldots\right. \\ &\left[\sum_{j_{s_{r-1}}} q_{j_1,\ldots,j_{s_{r-1}}}(\theta_{j_1,\ldots,j_{s_{r-1}}})\left[\sum_{j_{s_r}}^{n_a} q_{j_1,\ldots,j_{s_r}}(\theta_{j_1,\ldots,j_{s_r}})\delta_{aj_{s_r}}\right]\right]\Bigg]\Bigg] \\ &= \sum_{j_1} q_{j_1}(\theta_{j_1})\left[\sum_{j_2} q_{j_1,j_2}(\theta_{j_1,j_2})\left[\ldots\left[\sum_{j_{s_{r-1}}} q_{j_1,\ldots,j_{s_{r-1}}}(\theta_{j_1,\ldots,j_{s_{r-1}}})\right.\right.\right. \\ &\left[\sum_{j_{s_r}}^{n_a} q_{j_1,\ldots,j_{s_r}}(\theta_{j_1,\ldots,j_{s_r}})\sum_{a=1}^{n_a}\delta_{aj_{s_r}}\right]\Bigg]\Bigg]\Bigg] = 1,\end{aligned} \tag{10.15}$$

where it has been taken into account that $\sum_{a=1}^{n_a} \delta_{aj_{s_r}} = 1$ and the set of constraints (10.14).

For each conditional probability function some particular interesting cases can result.

Saturated models. All $q_{j_1,j_2,\ldots,j_r}$ functions are the identity function, that is, they correspond to direct parameters. In this case all degrees of freedom are exploited and all the conditional probabilities $P(X_K|\mathcal{P}^k)$ are independent leading to the most general possible model. Since they must satisfy constraint (10.14), a model that depends on $(s_k - 1) \prod_{r\in\mathcal{P}^k} s_r$ independent parameters, which can be denoted $\theta^k_{a;\mathcal{P}^k}; a = 1, 2, \ldots, s_k; \forall\mathcal{P}^k$ results. These models incorporate to the likelihood function factor terms of the form $\left(\theta^k_{x_k;\mathcal{P}^k}\right)^{n^k_{x_k;\mathcal{P}^k}}$, where $n^k_{x_k;\mathcal{P}^k}$ is the number of observed vehicles in the sample such that $X_k = x_k$ and the parent values of X_k in the sample are $\mathcal{P}^k$.

Marginal models. All $q_{j_1,j_2,\ldots,j_r}$ functions are independent on the parent variables. In this case the conditional probabilities degenerate to marginal probabilities and the likelihood of the Bayesian network incorporates factors of the form

$$\left(\theta^t\right)^{n^k_{t;\mathcal{P}^k}} \text{ and } \left(1-\sum_{r=1}^{s_k-1}\theta^r\right)^{n^k_{s_k;\mathcal{P}^k}}; \quad t < s_k.$$

Intermediate models. Some of the $q_{j_1,j_2,\dots,j_r}$ functions contain non-identity functions $h_j(\tilde{\mathcal{P}}^k;\boldsymbol{\theta}^k_j); j = 1, 2, \dots, t$ of one or several parameters, where $\tilde{\mathcal{P}}^k$ is a subset of the parents of node X_k and $\boldsymbol{\theta}^k$ is a vector of parameters. These models incorporate to the likelihood function factor terms of the form

$$\left(h_j(\tilde{\mathcal{P}}^k;\boldsymbol{\theta}^k{}_j)\right)^{n^k_{x_k;\mathcal{P}^k}} \text{ and } \left(1-\sum_j h_j(\tilde{\mathcal{P}}^k;\boldsymbol{\theta}^k{}_j)\right)^{n^k_{x_k;\mathcal{P}^k}}.$$

It is advanced that in all these three cases the parameters involved in these subsets of factors can be learned independently. For example, by the maximum likelihood or Bayesian conjugate methods (see Persaud and Lyon (2007)). See also Mu and Yuen (2016).

Next, several examples that clarify the meaning of the $q_{j_1,j_2,\dots,j_{s_r}}$ functions are given, but before, in Table 10.1 a list of some of the conditional probability tables that are used in our model and the number of parents which are observable and unobservable are provided, together with the number of total conditional probability tables and the number of tables and parameters to be updated. It can be seen that some tables cannot be learned because they involve non observable childs, such as those for nodes D (driver's attention), Sd (Driver's speed decision), Dri (driver type), Ds (Driver's decision at a sign), V (vehicle failure) and Tf (thechnical failure). Other tables can be learned with complete data, such as It, Vis, Vt, W and SS, and some that are partially observable with incomplete data, such as S, P, Co and I.

Example 5 (Conditional probabilities for the collision node) *The conditional probability $p(a, b, c, d, e, f) = P(P = a|Vt = b, It = c, Vis = d, D = e, S = f)$ of a collision node can be*

$$\begin{aligned} p(a,b,c,d,e,f) = & z(c,d,e,f;\theta_z)(\delta_{a2}p_2(b,f;\theta_p)+\delta_{a3}p_3(b,f;\theta_p) \\ & +\delta_{a4}p_4(b,f;\theta_p))+(1-z(c,d,e,f;\theta_z))\delta_{a1}, \end{aligned} \tag{10.16}$$

where z is the collision probability and p_1 to p_4 are the probabilities of no incident, minor, medium and severe incident, respectively. Thus, the contribution of the collision node to the log-likelihood of the sample is given by

$$\delta_{a1}\sum_{\{1.cdef\}}\log(1-z(c,d,e,f;\theta_z))$$

Table 10.1: List of nodes showing observable and unobservable parents, number of possible values of the child node, number of different conditional probability tables, and numbers of conditional probability tables and parameters to be updated.

Node	Parents		Child values	Total	Updating	
	Observable	Unobservable		# of tables	# of tables	Parameters
D	It, Vis	D_p, Dri	3	108	12	36
Sd		D	3	3	-	-
Dri		Dri_p	4	4	-	-
It	It_p		3	3	1	3
Vis	W		3	4	1	4
Vt	Vt_p		3	3	1	3
S	Sp, Vt, W, It	Dri, Sd, D	8	10368	36	288
V	Vt, S	D	4	72	-	-
P	It, W, Vis, Vt, S	D	4	2592	3	12
Co	It, Vis, Vt, S	D	4	648	3	12
W	W_p		4	4	1	4
SS			3	1	1	3
Ds		D	2	3	-	-
Tf			2	1	-	-
I	P, Co	V	4	64	4	16
I	W, Vt, S	D	4	384	3	12

$$
\begin{aligned}
&+\sum_{\{..cdef\}} \log z(c,d,e,f;\theta_z) \left(\delta_{a2} \sum_{\{2b\ldots f\}} \log p_2(b,f;\theta_p) \right. \\
&\left. +\delta_{a3} \sum_{\{3b\ldots f\}} \log p_3(b,f;\theta_p) + \delta_{a4} \sum_{\{4b\ldots f\}} \log p_4(b,f;\theta_p) \right), \qquad (10.17)
\end{aligned}
$$

where the summations are extended over the indices $\{abcdef\}$ using subsets of indices of the type $\{..cdef\}$, where the dots refer to all values of the corresponding index and the individual letters to the particular values.

This means that the log-likelihood can be divided into independent summands where different subsets of parameters can be included and then, estimated independently by maximizing the corresponding summands separately. This implies a substantial reduction in the computation time of the likelihood function.

For example, in expression (10.17) the first two summands and the last four summands can be considered independently to estimate the parameters θ_z contained in term z and the parameters θ_p contained in p_2 to p_4, respectively.

Example 6 (Conditional probabilities at a decision node) *Consider for example the conditional probability of node Sd Driver's decision, given the driver's attention (see Figure 10.1)*

$$P(Sd = a|D = b) \quad = \quad \delta_{b1}\delta_{a2} + \delta_{b2}\left[\tau\left(\kappa\delta_{a2} + (1-\kappa)\delta_{a3}\right)\right.$$

$$+(1-\tau)\delta_{a1}] + \delta_{b3}\delta_{a1}, \tag{10.18}$$

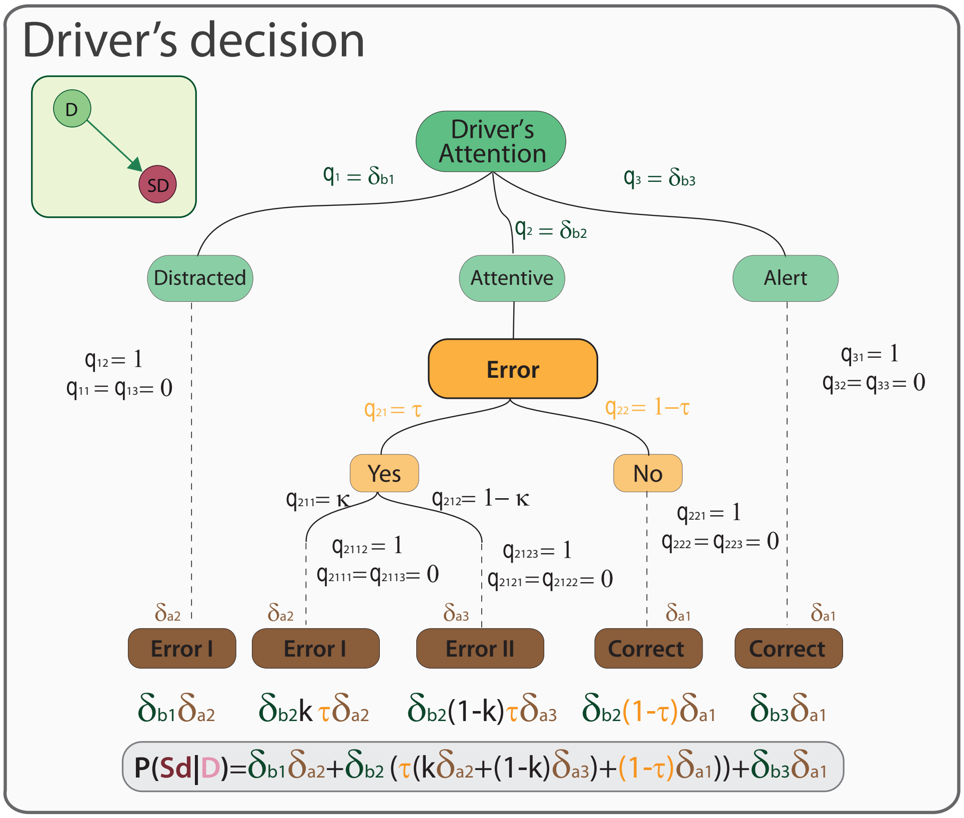


Figure 10.1: Illustration of the q functions and how the conditional probability is obtained for node Sd.

where τ is the probability of the driver to make an error and κ is the probability of ignoring that a decision must be made (Error I) once the error has occurred.

In this case the contribution of the conditional probability of node Sd Driver's decision to the likelihood of the sample is (see (10.18))

$$\sum_{\{.2\}} \log \tau + \sum_{\{12\}} \log(1-\tau) + \sum_{\{22\}} \log \kappa + \sum_{\{32\}} \log(1-\kappa). \tag{10.19}$$

And then, the maximum likelihood estimates of parameters τ and κ are the proportions of sample values in which an error is made and those in which this error is Error I, respectively.

10.3 Learning the saturated model

Let $\{X_1, X_2, \dots, X_s\}$ be the set of variables of the Bayesian network, which are denoted by capital letters. Consider a sample of m vehicles circulating by the road and let $\{x_{1i}, x_{2i}, \dots, x_{si}\}; i = 1, 2, \dots, m$ be the sample values, where lower case letters have been used to refer to the particular values of the corresponding variables in the sample.

Let $\theta^k_{X_k;\mathcal{P}^k} = P(X_k|\mathcal{P}^k)$ be the conditional probabilities of the node X_k given their parents $\mathcal{P}^k$, which are considered the parameters of the Bayesian network to be estimated. For the parameters to be a valid conditional probability they must satisfy the constraints

$$\sum_{a=1}^{s_k} \theta^k_{a;\mathcal{P}^k} = 1; \quad \forall \mathcal{P}^k, \tag{10.20}$$

where s_k is the number of possible values of node X_k.

Then, the likelihood of the sample is given by

$$L(\mathbf{x};\boldsymbol{\theta}) = \prod_{i=1}^{m} \prod_{\mathcal{P}^k} \prod_{k=1}^{s} \prod_{x_k=1}^{s_k} \theta^k_{x_{ki};\mathcal{P}^k_i} = \prod_{\mathcal{P}^k} \prod_{k=1}^{s} \prod_{x_k=1}^{s_k} \left(\theta^k_{x_k;\mathcal{P}^k}\right)^{n^k_{x_k;\mathcal{P}^k}}, \tag{10.21}$$

where x_{ki} and $\mathcal{P}^k_i$ are the values of the node X_k and its parents in sample data i, and $n^k_{x_k;\mathcal{P}^k}$ is the number of observed vehicles in the sample such that $X_k = x_k$ and the parent values of X_k in the sample are $\mathcal{P}^k$.

The log-likelihood of the sample becomes

$$\begin{aligned} \log L(\mathbf{x};\boldsymbol{\theta}) &= \sum_{\mathcal{P}^k} \sum_{k=1}^{s} \sum_{x_k=1}^{s_k} n^k_{x_k;\mathcal{P}^k} \log \theta^k_{x_k;\mathcal{P}^k} \\ &= \sum_{k=1}^{s} \sum_{\mathcal{P}^k} \left(\sum_{x_k=1}^{s_k} n^k_{x_k;\mathcal{P}^k} \log \theta^k_{x_k;\mathcal{P}^k} \right). \end{aligned} \tag{10.22}$$

Equation (10.22) reveals that the maximization of the log-likelihood function is equivalent to the maximization of the summands corresponding to the different conditional probabilities of the nodes one by one and separately. More precisely, it is necessary to maximize

$$\sum_{x_k=1}^{s_k} n^k_{x_k;\mathcal{P}^k} \log \theta^k_{x_k;\mathcal{P}^k}; \quad \forall \mathcal{P}^k, \quad \forall k = 1, 2, \dots, n, \tag{10.23}$$

which implies a very important reduction in complexity and CPU time.

In addition, given that the parameters must satisfy the constraints in (10.20), the maximization of the expressions in (10.23) leads to

$$\hat{\theta}^k_{x_k;\mathcal{P}^k} = \frac{n^k_{x_k;\mathcal{P}^k}}{\sum\limits_{x_k=1}^{s_k} n^k_{x_k;\mathcal{P}^k}}; \quad \forall \mathcal{P}^k, \quad \forall k = 1, 2, \dots, n, \tag{10.24}$$

which are the well known classical estimates, that is, the sample proportions.

If Bayesian estimated associated with the Dirichlet conjugate distributions are considered, the following expression can be obtained:

$$\hat{\theta}^k_{x_k;\mathcal{P}^k} = \frac{n^k_{x_k;\mathcal{P}^k} + n^{k0}_{x_k;\mathcal{P}^k}}{\sum\limits_{x_k=1}^{s_k} n^k_{x_k;\mathcal{P}^k} + N_0}; \quad \forall \mathcal{P}^k, \quad \forall k = 1, 2, \ldots, n, \tag{10.25}$$

where $n^{k0}_{x_k;\mathcal{P}^k}$ are the prior parameters and $N_0 = \sum\limits_{x_k=1}^{s_k} n^{k0}_{x_k;\mathcal{P}^k}$.

If there are no observable parents, Formula (10.25) must be replaced by the following formula

$$\begin{aligned}\hat{\theta}^k_{x_k;\mathcal{P}^k,\hat{\mathcal{P}}^k} &= \frac{n^k_{x_k;\mathcal{P}^k} P(\bar{\mathcal{P}}^k) + n^{k0}_{x_k;\mathcal{P}^k}}{\sum\limits_{\bar{\mathcal{P}}^k} \sum\limits_{x_k=1}^{s_k} n^k_{x_k;\mathcal{P}^k} P(\bar{\mathcal{P}}^k) + N_0}; \\ & \qquad \forall \mathcal{P}^k, \quad \forall k = 1, 2, \ldots, n, \end{aligned} \tag{10.26}$$

where now $\mathcal{P}^k$ and $\bar{\mathcal{P}}^k$ refer to the subsets of parents which are observed and unobserved, respectively, and $P(\bar{\mathcal{P}}^k)$ is the joint probability of the unobserved parents of node X_k, which can be easily obtained from the X_k-parents clique.

It is clear from expression (10.26) that the effect of the prior information $n^{k0}_{x_k;\mathcal{P}^k}$ and N_0 on the parameter estimates $\hat{\theta}^k_{x_k;\mathcal{P}^k,\hat{\mathcal{P}}^k}$ of the observable nodes becomes negligible when the sample size is large. However, this can take place only for very large sample sizes if the true values of the parameter is very small.

There are important particular cases in which the functions $q_{j_1,j_2,\ldots,j_{s_t}}$ depends only on single parameters. In this case, terms of the following form result

$$\sum_{j_1,j_2,\ldots,j_t} \log \theta_{j_1,j_2,\ldots,j_t}, \tag{10.27}$$

with

$$\sum_{j_1,j_2,\ldots,j_t} \theta_{j_1,j_2,\ldots,j_t} = 1, \tag{10.28}$$

which implies that one parameter, say the last one, is one minus the sum of the rest of parameters in this set.

In this case, the maximum likelihood estimate of the parameter is the observed frequency in the sample, so that closed and simple formulas result for the parameter estimates.

Chapter 11

Real cases studies

Contents

11.1 Introduction

This chapter presents the most relevant data obtained when applying the model proposed in this Thesis for the study of three specific cases of Autonomic roads in Cantabria, Spain. These are the Autonomic road CA-131, which goes from Torrelavega to the National Road N-634 near San Vicente de la Barquera, the road CA-132 that connects Viveda with Suances and CA-142 between Astillero and Selaya. Firstly, the accident data observed on the three roads are analyzed. For that, the statistics of accidents corresponding to the period 2006-2016, that the Council of Public Works of Cantabria has provided us have been used. For a more detailed analysis of specific accidents, other data are available, such as the moment of occurrence of the incident, type of intersection, direction of movement, type of vehicle, luminosity, pavement condition (dry and wet), atmospheric factors (good weather, rain, and fog), circulation (fluid and dense), type of road, that serve to analyze the specific causes of each one. This allow the establishment of accidents and the division of these into different sections. In addition, the causes of these incidents have been identified by comparing the environment in which they occurred with the characteristics of each of them. Since the interest of this Thesis is to see the value of the model and the results that can be obtained by it, it will not go into detail in this aspect. Simply add, as discussed in Chapter 7, that the same division of sections has been used for the validation of the results when analyzing the prediction of the accidents. Thus, by applying the Bayesian networks model, the prediction of potential incidents is obtained. Conclusions regarding to global safety and local safety are made. From each road the 20 riskiest points are indicated and illustrative examples are given of some of the points particularly showing their characteristics.

11.2 Preliminary considerations to the analysis of the results

In order to carry out a complete safety analysis of the roads considered, it is necessary to make a detailed study of them with the help of the information provided by the probabilistic safety analysis resulting from the Bayesian networks model. This allows us to analyze three important aspects:

- **Risk of the road per trip.** Obtaining the cumulative values of ENSI, every time the road is traveled and per kilometer, allows to assess the intrinsic risk of the road in general and per unit of length, and therefore to evaluate its level of safety regardless of the number of vehicles that run through it. This is of interest to users, who may know the risks they can have each time they use it. To do this the cumulative probabilities of incidents when traveling the road in each direction are determined. In the case of a large difference, it could be due to design errors, to an incorrect or unbalanced road signaling or to the fact that the two directions are clearly different with respect to their conditions, what happens in some cases, for example, in the case of strong slopes. The cumulative overall probabilities of incidents considering both directions are done by adding the risks of the two directions.

- **Risk associated with the road and its users per year.** This will be of interest to those responsible for traffic safety in general, as they should consider all users of the road in question. In this way the annual ENSI is analyzed, which is obtained by multiplying the ENSI per trip of each one of its sections by the corresponding number of users and adding the contributions of all its sections. There are points that, although having a lower probability of individual ENSI (for a trip), result in a high annual ENSI due to the high number of users that circulate through it, or also those that, having a lower ENSI value, have a high frequency of minor or medium incidents. It is therefore appropriate to focus improvements on those points whose annual ENSI values are the highest, since when acting at those points, the potential reduction in the number of incidents and their severity after the improvements will be greater and, therefore,the investment made will be more profitable.

- **Causes or typical circumstances of each incident.** After locating those points with elevated ENSI, it is possible to determine the causes or the type of circumstances that give rise to such potential incidents. Thanks to this evaluation, the solutions to these can be accurately defined, since the most probable causes of the incidents will be known.

These methods, applied to each road and each direction, allow to generate a precise description of the global safety and will serve to detect the riskiest points, determining the conditions or situations more likely to have potential incidents, in order to avoid them. In this case, the joint risk has been considered in both directions and the 20 points with the highest ENSI per year have been identified.

11.3 Autonomic road CA-131

The first thing that has been done for the safety study of this road is to analyze the observed data of accidents in the period 2006 to 2016. For this, the data provided have been represented in values of ENSI, in the layout of the road. The points where incidents have been observed have been indicated with circles whose diameter is proportional to the square root of their frequency and their colors indicate the severity of the incidents (light yellow in minor, orange, medium and red, severe). This information can be seen in Figure 11.1 where the top shows the accident rate observed, and at the bottom, the road has been divided into sections of similar accidentality.

The same division of sections of the route has been used to make the prediction of points that have the greatest potential for occurring accidents. The following is a brief description of each segment and possible incidents that may occur.

1. Section between KP 0.000 and KP 0.800. It is between Barreda and the KP 0.800 located near Viveda. It is a segment with medium expected incidents in ten years only in the roundabout at the starting point (1.17) and especially at the pedestrian crossing at KP 0.045 (4.54), at the lateral entry at KP 0.215 (1.01) and collision on road in the area of KP 1.028 (1.03).

2. From KP 0.800 to KP 11.000. KP 0.800 is located in Viveda and KP 11.000 after the intersection with the CA-920. In this section, severe incidents due to collisions on road and run over animals can be predicted at KP 7.902 (1.00) and KP 8.635 (1.00), respectively. Others medium potential incidents occur at the intersection at KP 2.295 (1.11), KP 5.372 (1.74) and KP 7.266 (1.03) because of collisions on road, the curve at KP 7.902 (1.04) and run over animals near KP 6.390 (1.04) and KP 9.126 (1.21).

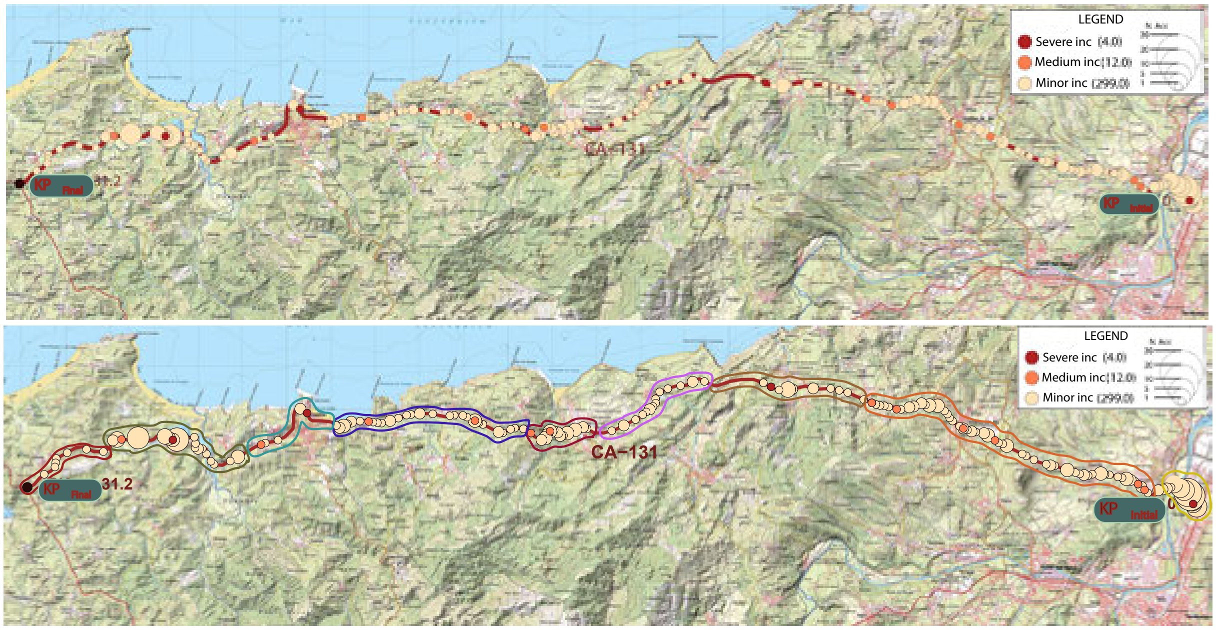


Figure 11.1: At the top the graphical representation of the observed accident rate on the Autonomic road CA-131, which shows the frequencies of minor accidents (light yellow color), medium (orange color) and severe (red color) and the diameter of the circles is proportional to the square root of their frequency, is shown. At the bottom, the road has been divided into sections of similar accidents.

Figure 11.2: Potential accidents of the Autonomic road CA-131.

3. The next section goes between KP 11.000, after the intersection with the Autonomic road CA-920 and Volao (KP 12.800). There is a low accident rate here, the incidents are minor and principally referred to run over animals, collisions on road and some curves.

4. The segment between KP 12.800 and KP 15.000 goes from Volao to Cóbreces. The prediction includes only severe incidents at the curve located at KP 14.030 (1.05), being the rest minor incidents of low frequency.

5. The section between KP 15.000 and KP 18.000 takes place between Cóbreces and Paderna. Severe incidents are expected near KP 15.917 due to run over animals (1.36) and collisions on road (1.20), medium because of run over animals (1.00) in the area of KP 15.160. The rest are minor incidents of low frequency.

6. Paderna and Casasola villages are located between KP 18.000 and KP 21.300. The only severe prediction of this section corresponds to collisions on road (1.02) at the zone of KP 18.651. There are also predictions of medium incidents at KP 20.073, where there is an intersection (4.06) and in the area of KP 20.485 where incidents are collisions on road (1.17). The rest are minor incidents, highlighting three intersections, with minor severity, at KP 20.443 (4.65), KP 20.073 (4.055), and KP 19.696 (2.875), and a lateral entry at KP 18.2548 (2.65).

7. Section between KP 21.300 and KP 24.000. It includes Casasola and the entrance to Rubárcena. A low accident rate without anything outstanding can be predicted, except the possibility of running over animals.

8. Section between KP 24.000 and KP 27.000. It is from Rubárcena to the Capitan River. It has a low risk. There are minor incidents at the intersections at KP 24.444 (2.14) and KP 25.376 (2.46) and at the lateral entry at KP 26.750 (2.71) and also the possibility of incidents due to run over animals.

9. The final segment includes the kilometers between KP 27.000 and KP 31.200 which correspond to Capitan River and the intersection at KP 31.200 with the National road N-634. It is a low risk section in which intersections can be highlighted at KP 27.210 (3.22), KP 29.953 (2.62) and KP 30.920 (2.10). Furthermore, there is a possibility of incidents due to run over animals.

11.3.1 Road safety analysis

After performing the probabilistic safety analysis of the Autonomic road CA-131 in both directions, the following conclusions regarding to the safety of the road can be drawn.

Global risk

When studying the cumulative risk, an annual global ENSI of 0.613 is obtained upwards and 0.418 downwards, resulting a total anual ENSI value of 1.031. Because this road has a length of 31.05 km, the value per km in ascending direction is 0.020, 0.013 in downward direction and 0.033 the global value. From these values the following conclusions can be derived:

- The risk in the upward direction differs slightly from the downward direction. A clear negative difference will be observed in comparison with the downward direction below.

- The overall values of anual ENSI per km, which determine the level of safety of the road, do not have a value higher than 0.0924. Thus, the road can be considered to have a reasonable level of safety.

Local risk

Although the road is considered to have an acceptable level of safety at a global level, 209 road points are identified (120 upwards and 89 downwards, coinciding many of them in the same KP) as potential points to improve their safety.

As it can be seen in Table 11.1, most incidents would be located at curves or along certain segments of the road, but it should also be noted that there are points with a significant ENSI value at roundabouts, intersections or, particularly, at a pedestrian crossing on the first part of the road.

Table 11.1: Table of the riskiest points of the $CA-131$ road divided in upwarding direction and downward, with their probability of incidents sorted by level of risk, with their corresponding KP, nodes and ENSI local values.

Rank	Item	Item type	KP	Node	ENSI Local	Severity Severe	Medium	Minor
				Ascending direction				
1	5	PedestrianCrossing	0.045	I71-45Pc	0.0756	0.00433	0.454	0.0499
2	11	CurveIn	0.132	I159-132Cv	0.0412	0.0344	0.0374	0.208
3	2	RoundAbout	0.001	I22-1Rd	0.03	0.0157	0.0822	0.34
4	267	CurveIn	7.726	I4051-7726Cv	0.0232	0.0195	0.0216	0.069
5	14	LateralEntry	0.215	I203-215LE	0.0189	0.0037	0.0776	0.718
6	595	Intersection	20.073	I9057-20073Int	0.0104	0.00466	0.0309	0.204
7	265	CurveIn	7.600	I4022-7600Cv	0.00892	0.00661	0.0134	0.0494
8	261	CurveIn	7.491	I3963-7491Cv	0.00817	0.00582	0.0137	0.0488
9	392	RunoverAnimals	12.779	A5951-12779SSC	0.00647	0.00484	0.00949	0.0339
10	397	CurveIn	12.891	I6029-12891Cv	0 .00631	0.00543	0.00504	0.0209
11	34	RoundAbout	0.823	I513-823Rd	0.00606	0.000179	0.0351	0.0923
12	567	CurveIn	18.996	I8634-18996Cv	0.00557	0.0029	0.0141	0.105
13	383	RunoverAnimals	12.390	A5819-12390Sall	0.00522	0.00392	0.00763	0.0268
14	13	Collision	0.210	Co187-210Sall	0.00488	0.000215	0.0235	0.228
15	383	Collision	12.390	Co5821-12390Sall	0.00487	0.00452	0.000772	0.0527
16	16	TrafficLight	0.250	I237-250TL	0.00477	0.00357	0.00383	0.137
17	392	Collision	12.779	Co5953-12779SSC	0.0047	0.00435	0.000826	0.0511
18	17	CurveIn	0.255	I252-255Cv	0.00463	0.00343	0.00419	0.124
19	437	RunoverAnimals	14.531	A6625-14531Sall	0.00422	0.00316	0.00616	0.0218
20	389	CurveIn	12.495	I5910-12495Cv	0.00414	0.0034	0.0041	0.0248
21	399	RunoverAnimals	13.160	A6055-13160Sall	0.00414	0.00311	0.00605	0.0213

Continued on next page

Table 11.1 – continued from the previous page

Rank	Item	Item type	KP	Node	ENSI Local	Severity Severe	Medium	Minor
22	602	Intersection	20.443	I9161-20443Int	0.004	0.000112	0.0182	0.238
				Downward direction				
1	647	CurveIn	7.910	I9829-7910Cv	0.0174	0.0141	0.0198	0.0649
2	895	Intersection	0.250	I13595-250Int	0.0146	2.42e-06	0.0494	1.59
3	822	CurveIn	2.200	I12496-2200Cv	0.0137	0.0106	0.0182	0.0563
4	651	CurveIn	7.724	I9888-7724Cv	0.0133	0.0111	0.013	0.0455
5	481	CurveIn	14.030	I7315-14030Cv	0.0127	0.00997	0.0164	0.0441
6	658	CurveIn	7.450	I9991-7450Cv	0.0114	0.00823	0.0189	0.049
7	544	CurveIn	11.480	I8251-11480Cv	0.00948	0.00777	0.0102	0.0274
8	512	CurveIn	13.160	I7781-13160Cv	0.00948	0.00837	0.00642	0.0248
9	328	Intersection	20.073	I4962-20073Int	0.00841	0.00243	0.0327	0.201
10	654	CurveIn	7.595	I9932-7595Cv	0.00828	0.00609	0.0127	0.0471
11	322	Intersection	20.443	I4872-20443Int	0.00606	0.000564	0.0289	0.226
12	528	Collision	12.161	Co8015-12161Sall	0.00573	0.00537	0.000722	0.0562
13	516	CurveIn	12.870	I7840-12870Cv	0.00533	0.00433	0.00559	0.0288
14	815	Intersection	2.295	I12393-2295Int	0.00521	0.000658	0.0266	0.0922
15	523	CurveIn	12.405	I7944-12405Cv	0.00455	0.0035	0.00601	0.0267
16	477	Collision	14.292	Co7254-14292SSC	0.0042	0.00398	0.000353	0.0383
17	520	Collision	12.545	Co7898-12545Sall	0.00409	0.00386	0.000365	0.0394
18	533	Collision	11.840	Co8087-11840Sall	0.00409	0.00383	0.000551	0.0414
19	528	RunoverAnimals	12.161	A8013-12161Sall	0.00403	0.00193	0.0124	0.0382

$ENSI^*$ Expected number of equivalent severe incidents
Frequency of severe incident > 9.2
Frequency of medium incident > 0.92
Frequency of minor incident > 0.092

11.3.2 Specific points where safety must be improved

In order to reduce the likelihood of severe incidents, and thus the global risk of the road, it is considered necessary to act on the 20 riskiest points, considering the joint risk of both directions, as shown in Table 11.2 and represented in Figure 11.3.

Next, two of these specific points are illustrated. They are analyzed and their characteristics and the most probable circumstances that give rise to such high potential risk are defined.

Figure 11.3: Riskiest points of the CA-131 road.

Table 11.2: Riskiest points of the $CA-131$ road considering the two directions.

Rank	Item type	KP	Annual ENSI Total	Annual ENSI Ascending	Annual ENSI Downward
1	PedestrianCrossing	0.045	0.078	0.076	0.002
2	CurveIn	0.132-0.210	0.041	0.041	3.57E-08
3	CurveIn	7.726-7.910	0.041	0.023	0.017
4	RoundAbout	0.001	0.031	0.030	0.001
5	CurveIn	7.600-7.724	0.022	0.009	0.013
6	LateralEntry	0.215	0.020	0.019	0.001
7	Intersection	0.25	0.019	0.005	0.015
8	Intersection	20.073	0.019	0.010	0.008
9	CurveIn	7.491-7.595	0.016	0.008	0.008
10	CurveIn	12.891-13.160	0.016	0.006	0.009
11	CurveIn	2.100-2.200	0.014	2.30E-07	0.014
12	CurveIn	13.899-14.03	0.013	1.47E-06	0.013
13	CurveIn	7.291-7.450	0.011	3.80E-07	0.011
14	Intersection	20.443	0.010	0.004	0.006
15	CurveIn	11.250-11.48	0.010	4.46E-05	0.009
16	RunoverAnimals	12.545-12.779	0.009	0.007	0.003
17	CurveIn	18.996-19.260	0.008	0.006	0.002
18	RoundAbout	0.823	0.007	0.006	0.001
19	Intersection	19.696	0.003	3.86E-04	0.003
20	Intersection	17.815	0.003	0.001	0.001
21	Intersection	1.229	0.003	0.001	0.001
22	Intersection	27.21	0.002	0.001	0.001
23	Intersection	25.376	0.002	0.001	0.001
24	LateralEntry	26.75	0.001	0.001	0.001
25	Intersection	5.84	0.001	0.001	3.15E-04

Point 1: Pedestrian crossing at KP 0.045

These are incidents that may occur at the pedestrian crossing located at KP 0.045, shown in Figure 11.4. To improve the knowledge about these types of incidents, the most probable circumstances or situations that lead to incidents at this point are determined. Table 11.3 shows these circumstances provided by the Bayesian networks model and the following conclusions can be drawn for this case:

1. In spite of being a pedestrian crossing, the incidents are mainly caused not by run over pedestrians but by collision or by reaching between vehicles that are giving way to pedestrians and/or are in the process of continuing driving, in days of good weather or slightly rainy, mainly by standard conductors and at relatively high speeds.

2. Due to the high heavy vehicles traffic or to an unexpected speed reduction, these collisions can lead to incidents of greater severity due to the fact that most of them are medium incidents.

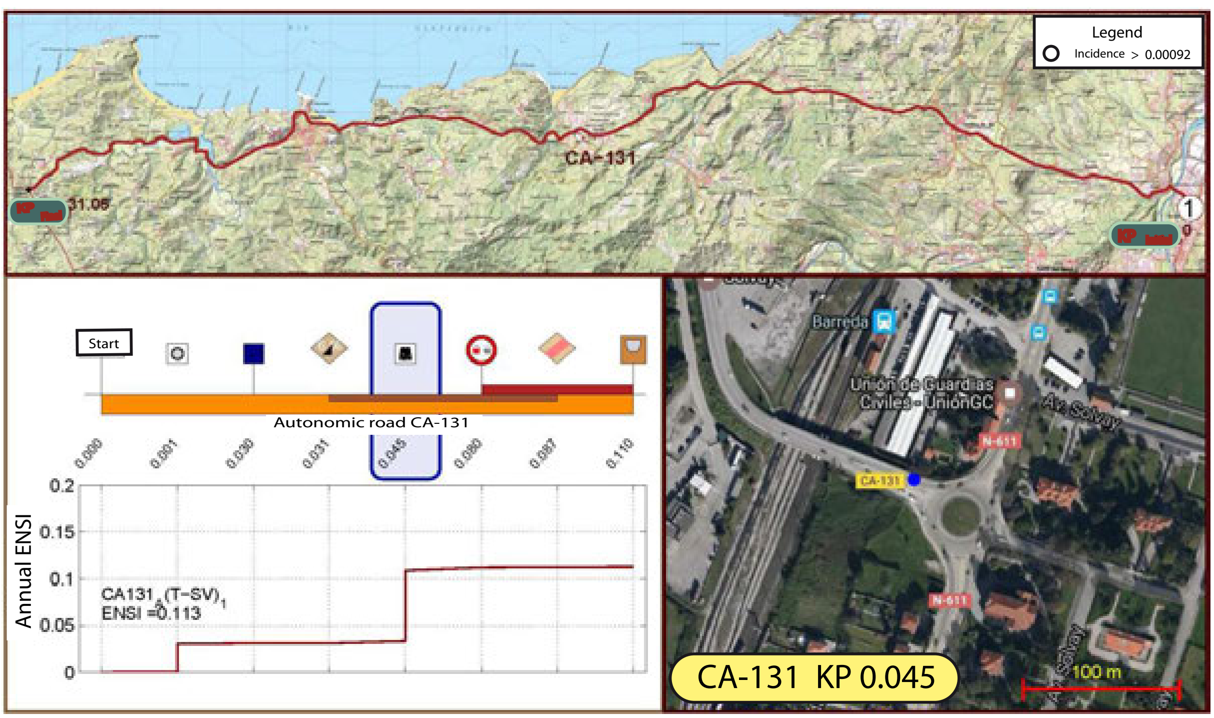


Figure 11.4: Detail of the incident at the riskiest point (pedestrian crossing at KP 0.045).

Table 11.3: Combination of main variable values contributing to the total ENSI value at the riskiest point (pedestrian crossing at KP 0.045).

n	Weather	Driver type	Tech. failure	Speed	Incident	Probability	% ENSI
1	Fair	Standard	No	66	Medium	2.04e-08	14.61
2	Fair	Standard	No	54	Medium	1.79e-08	12.82
3	Medium	Standard	No	66	Medium	1.67e-08	12.00
4	Medium	Standard	No	54	Medium	1.47e-08	10.53
5	Fair	Bad	No	66	Medium	7.13e-09	5.11

Point 2: Curve between KP 0.132 and KP 0.210.

These are the potential incidents that can occur at the curve of 125 m radius between KP 0.132 and KP 0.210, which is shown in Figure 11.5. Firstly, it should be noted that the greatest risk, see Table 11.4, comes from the upward direction, thus, it is necessary to focus mainly on this.

In order to determine the circumstances that give rise to such potential incidents, Table 11.4 separates the ENSI into the main components. According to this it can be said that:

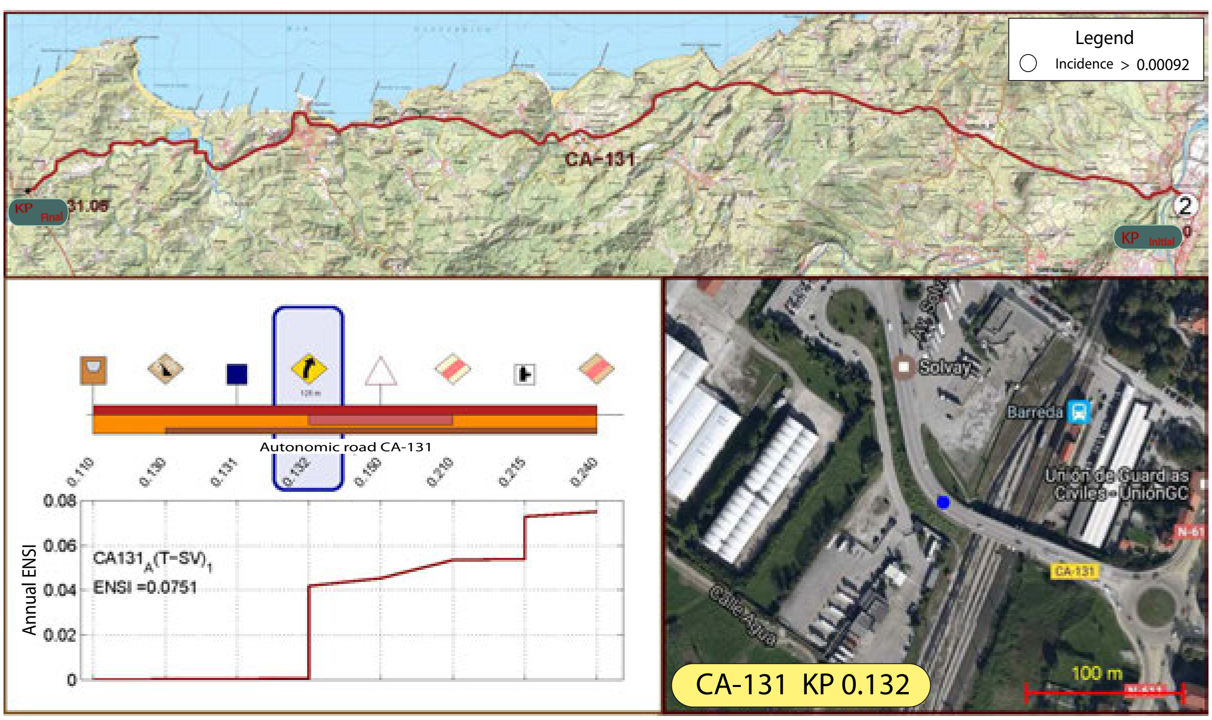


Figure 11.5: Detail of the incident at the riskiest curve (at KP 0.132).

Table 11.4: Main variable combination values contributing to the total ENSI value of incident at the riskiest curve (KP 0.132).

n	Weather	Veh. type	Attention	Speed	Incident	Probability	% ENSI
1	Medium	Car	Alert	90	Severe	3.52e-09	29.67
2	Medium	Car	Alert	108	Severe	2.17e-09	18.27
3	Medium	Car	Attentive	90	Severe	1.7e-09	14.32
4	Medium	Car	Attentive	108	Severe	1.13e-09	9.53
5	Medium	Car	Alert	90	Medium	5.72e-09	7.53
6	Medium	Car	Attentive	90	Medium	2.76e-09	3.64

1. These incidents occur mainly on rainy, foggy or drizzle days.

2. There are potentially medium or severe incidents, due to cars driving with speed excess which, despite not being distracted, due to a loss of adhesion to the pavement, there are cars running off the road.

11.4 Autonomic road CA-132

In this section the prediction of potential accidents in the Autonomic road CA-132 which links Viveda with Suances and has a length of 5.95 km is analyzed.

The same sections that were used previously for dividing the observed accidents (represented on the left side of Figure 11.6 are analyzed below for the potential accidents (see Figure 11.6).

1. Section between KP 0.000 and KP 1.000. It is between Viveda and Munios. With respect to the prediction of medium accidents of this section the intersection located at KP 0.380 (2.57) can be highlighted. The rest corresponds to predictions of minor incidents, such as those produced at the lateral entries located at KP 0.250 (4.69), KP 0.478 (2.95) and KP 0.932 (2.29) and at the intersections of KP 0.380 (14.99) and KP 0.057 (3.80).

2. The following section includes from the area of Munios, KP 1.000, to La Carreada, KP 5.000. Here the prediction of medium accidents includes only the intersection located at KP 1.863 (2.43). The other potential accidents are minor, being among them the intersections at KP 1.488 (4.59), KP 1.673 (8.22), KP 1.863 (11.52), KP 3.867 (2.82), KP 3.970 (2.97), KP 4.252 (2.37) and KP 4.390 (2.27), susceptible mainly of collisions and incidents on the road due to an excess of speed.

3. The final segment is between KP 5.000 and KP 5.95. It runs between La Carreada and Suances. In this section, the intersection located at KP 6.005 is highlighted with 1.03 expected medium incidents and 3.27 minor in 10 years.

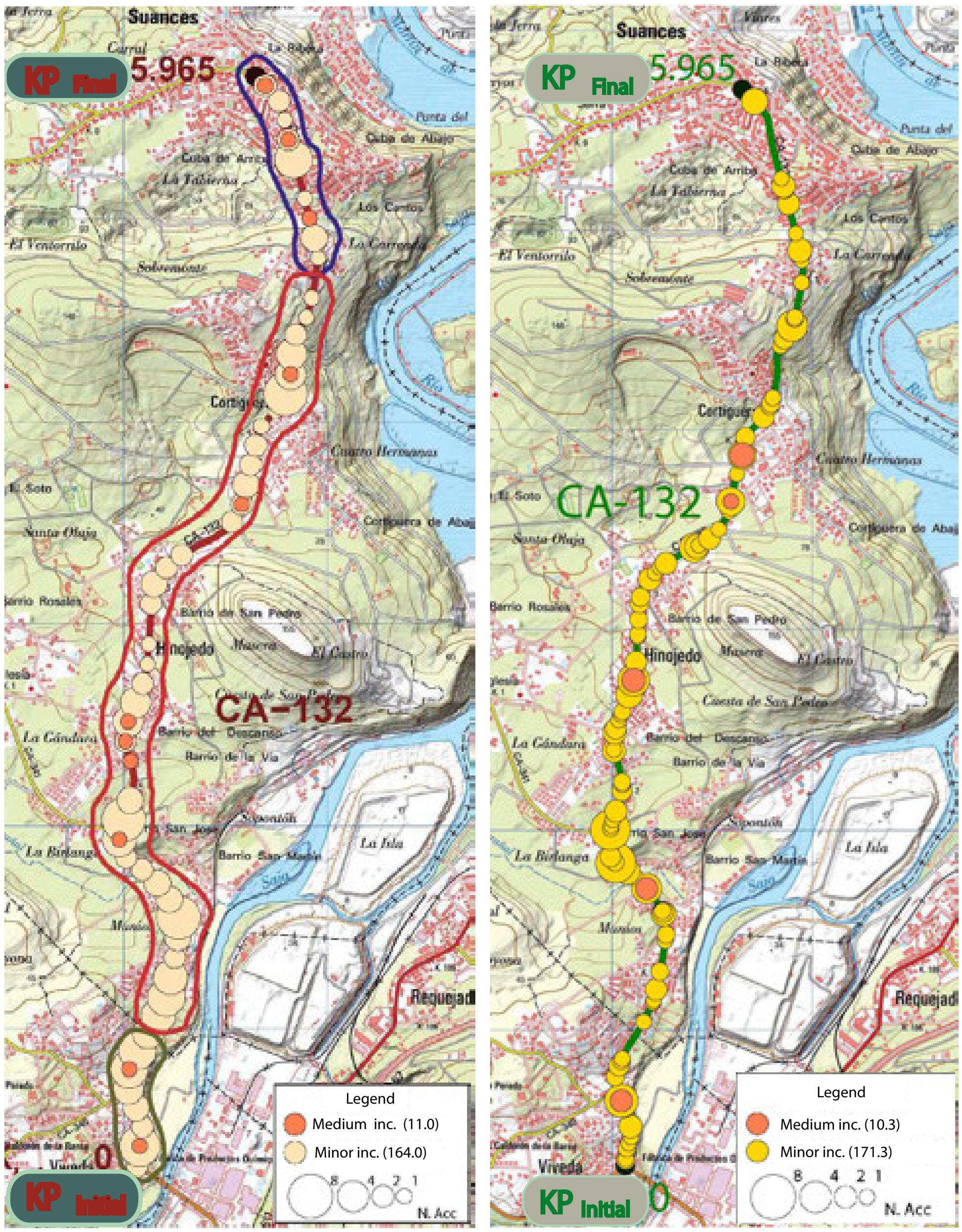


Figure 11.6: On the left the different sections in which the observed accidents of the road CA-132 have been divided are shown. On the right the graphical representation of the incident prediction for the same road is given, which shows the frequencies of level incidents (yellow), medium (orange) and severe (red) by means of circles of proportional diameter to the square root of the frequency of the incidents at said points.

11.4.1 Road safety analysis

The following conclusions concerning the road safety of the Autonomic road CA-132 can be made after performing the probabilistic safety analysis in both directions.

Global risk

When studying the accumulated risk, an annual global ENSI of 0.104 is obtained in the ascending direction and 0.111 in the downward direction, resulting in a total annual ENSI value of 0.215. Since this road has a length of 6.005 km, the value of the ENSI per km turns out to be 0.017 upwarding, 0.018 downwarding and 0.036 in total. Thus, it can be said that:

- The risks in the two directions are similar, so that there are no notable differences in one direction and its opposite.
- The global annual ENSI per km values do not reach a value greater than 0.092, which is the threshold chosen, therefore considering that the road CA-132 has a good safety level.

Local risk

At the global level the road is considered to have an acceptable level of safety, however, 21 points of the road (11 in the ascending direction and 10 in the descending direction) are indicated as potential points to be improved. Many of them coincide in KP for being collisions at intersections in opposite directions.

As it can be seen in Table 11.5, the main problem of this road is the potential number of collisions at intersections, which are much higher than any other type of incidents, due to the high circulation of vehicles on this road and the confluence of drivers of this and those circulating by the different localities.

Table 11.5: Table of the riskiest points of the $CA-132$ road divided in ascending and descending directions, with their probability of incidents sorted by level of risk, with their corresponding KP, nodes and ENSI local values.

Rank	Item	Item type	KP	Node	ENSI Local	Severity Severe	Severity Medium	Severity Minor
Ascending direction								
1	22	Intersection	0.380	I322-380Int	0.0104	0.000149	0.0454	0.718
2	109	Intersection	1.863	I1627-1863Int	0.00866	0.000133	0.0378	0.603
3	102	Intersection	1.673	I1524-1673Int	0.00658	0.000116	0.0301	0.406
4	87	Intersection	1.488	I1298-1488Int	0.00603	0.000141	0.0311	0.237
5	236	Intersection	3.970	I3533-3970Int	0.00471	2.58e-05	0.0262	0.139

Continued on next page

Table 11.5 – continued from the previous page

Rank	Item	Item type	KP	Node	ENSI Local	Severity Severe	Medium	Minor
6	130	Intersection	2.190	I1949-2190Int	0.00465	7.15e-05	0.0203	0.322
7	6	Intersection	0.057	I80-57Int	0.00401	5.78e-05	0.0175	0.279
8	215	Intersection	3.651	I3215-3651Int	0.00307	4.42e-05	0.013	0.227
9	7	TrafficLight	0.071	I100-71TL	0.00294	0.00131	0.00901	0.0498
10	141	Intersection	2.320	I2115-2320Int	0.00293	7.11e-05	0.0154	0.107
11	145	Intersection	2.390	I2174-2390Int	0.00288	7.01e-05	0.0151	0.103
Downward direction								
1	359	Intersection	0.380	I5396-380Int	0.0144	0.000355	0.068	0.781
2	273	Intersection	1.863	I4106-1863Int	0.0135	0.000627	0.0673	0.55
3	3	Intersection	6.005	I37-6005Int	0.00687	0.000227	0.0334	0.326
4	280	Intersection	1.673	I4210-1673Int	0.00595	9.12e-05	0.0259	0.416
5	295	Intersection	1.488	I4435-1488Int	0.00572	0.000134	0.0296	0.222
6	252	Intersection	2.190	I3785-2190Int	0.00525	8.12e-05	0.0236	0.342
7	140	Intersection	3.970	I2113-3970Int	0.00389	8.34e-05	0.0199	0.158
8	241	Intersection	2.320	I3619-2320Int	0.00325	7.69e-05	0.0172	0.11
9	162	Intersection	3.651	I2444-3651Int	0.00304	4.37e-05	0.0133	0.21
10	234	Intersection	2.390	I3514-2390Int	0.00275	5.89e-05	0.0141	0.112

$ENSI^*$ Expected number of equivalent severe incidents
Frequency of severe incident > 9.2
Frequency of medium incident > 0.92
Frequency of minor incident > 0.092

11.4.2 Specific points where safety must be improved

It is considered necessary to act on the 20 riskiest points considering the two directions (shown in Table 11.6 and represented in Figure 11.7), in order to reduce the likelihood of severe incidents, and thus the global risk of the road. As an example, the riskiest point is presented below, with its particular incident characteristics and conditioning factors for it to occur.

Table 11.6: Riskiest points of the $CA - 132$ road considering the two directions.

Rank	Item type	KP	Annual ENSI Total	Annual ENSI Ascending	Downward
1	Intersection	0.38	0.025	0.010	0.014
2	Intersection	1.863	0.022	0.009	0.014
3	Intersection	1.673	0.013	0.007	0.006

Continued on next page

Table 11.6 – continued from the previous page

Rank	Item type	KP	Annual ENSI Total	Annual ENSI Ascending	Annual ENSI Downward
4	Intersection	1.488	0.012	0.006	0.006
5	Intersection	2.19	0.010	0.005	0.005
6	Intersection	3.97	0.009	0.005	0.004
7	Intersection	6.005	0.007	2.88e-05	0.007
8	Intersection	2.32	0.006	0.003	0.003
9	Intersection	3.651	0.006	0.003	0.003
10	Intersection	2.39	0.006	0.003	0.003
11	Intersection	0.057	0.005	0.004	0.001
12	Intersection	1.434	0.004	0.002	0.002
13	LateralEntry	0.25	0.004	0.002	0.002
14	Intersection	0.996	0.004	0.002	0.002
15	Intersection	3.867	0.004	0.002	0.002
16	Intersection	4.252	0.003	0.002	0.002
17	Intersection	3.126	0.003	0.002	0.001
18	Intersection	3.052	0.002	0.001	0.001
19	Intersection	1.93	0.002	0.001	0.001
20	Intersection	3.171	0.002	0.001	0.001
21	TrafficLight	0.071	0.001	0.001	7.78e-05
22	Intersection	4.063	0.001	0.001	0.001
23	LateralEntry	0.478	0.001	0.001	0.001
24	Collision	1.294-1.356	0.001	2.36e-04	3.80e-04
25	Collision	1.742-1.809	0.001	2.62e-04	3.39e-04

Point 1: Intersection at KP 0.380

These are incidents that can occur at the "T" intersection located at KP 0.380, shown in Figure 11.8. Table 11.7 shows the circumstances that can cause these potential incidents in this location:

1. Front or front-to-side collisions, between those vehicles that are using the intersection and those that have been circulating along the road, occurring in adverse climatology and their consequences can become medium or severe incidents.

2. Incidents between vehicles operating at the intersection, such as reach collision and small front collisions, which, although more frequent, are less severe and, when they occur on days with a good or medium weather, they allow drivers to react as much as possible, to avoid an incident of worse consequences.

Figure 11.7: Riskiest points of the CA-132 road.

11.5 Autonomic road CA-142

The Autonomic road CA-142 links El Astillero with Selaya and has a length of 27.150 km. On the left side of Figure 11.9 the accidents observed are represented on the road layout divided in different sections. This same division is used to analyze the prediction of potential accidents (represented on the right of Figure 11.9) which can be described as follows:

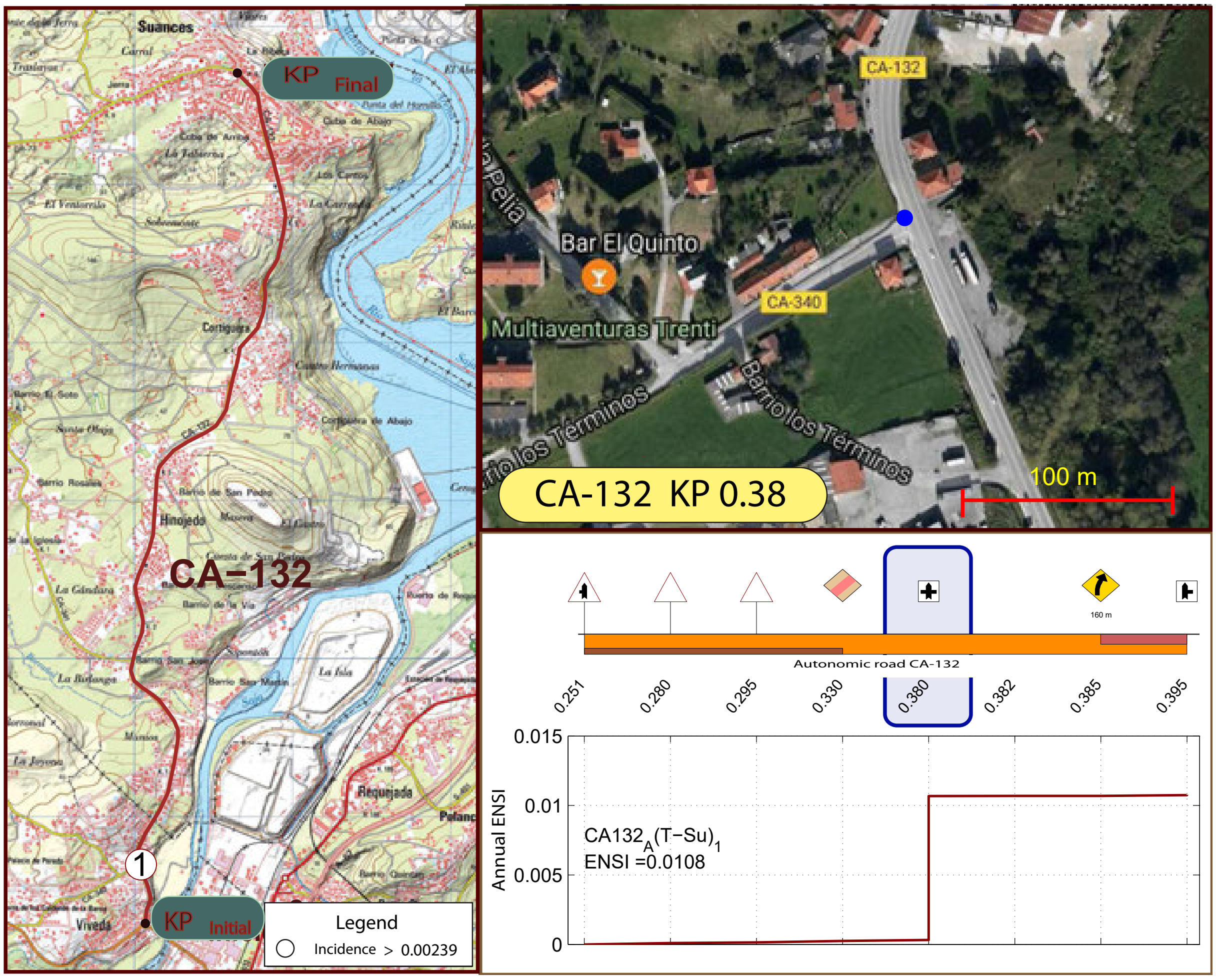


Figure 11.8: Detail of the incident at the riskiest point of the Autonomic road CA-132 (intersection at KP 0.380).

11.5.1 Road safety analysis

As before, the road has been divided into several sections.

1. Section between KP 0.000 and KP 3.300. It takes place between El Astillero at its intersection with the National road N-635 to the area of Bofetán. The riskiest point is found at the intersection located at KP 0.098 (1.3), being the rest of the predictions minor incidents especially at the intersections located at KP 1.940 (2.12) and KP 3.190 (2.16). In addition there are several lateral entries of lower severity that together with the collisions on the road, can lead to various incidents.

2. The segment from KP 3.300 to KP 7.000 runs between the area of Bofetán and the main entrance of Parque de Cabárceno. The area between KP 3.300 and KP 4.000 can be highlighted for its severity and collision risks. The lateral

Table 11.7: Combination of main variable values contributing to the total ENSI value at the riskiest point of the Autonomic road CA-132 (intersection at KP 0.380).

n	Weather	Veh. type	Attention	Speed	Probability	% ENSI
1	Bad	Car	Alert	24	5.32e-10	16.38
2	Very bad	Car	Alert	24	3.93e-10	12.11
3	Medium	Car	Alert	24	3.46e-10	10.67
4	Very bad	Car	Alert	15	2.27e-10	7.01
5	Fair	Heavy	Alert	24	2.24e-10	6.90
6	Medium	Heavy	Alert	24	1.84e-10	5.66
7	Fair	Car	Alert	48	1.7e-10	5.23
8	Medium	Car	Alert	39	1.44e-10	4.43
9	Medium	Car	Alert	48	1.29e-10	3.98
10	Fair	Car	Alert	39	1.11e-10	3.42

entries at KP 3.500 (2.03), KP 5.169 (2.31), KP 6.128 (2.10), KP 6.483 (2.32), KP 6.790 (2.68), and KP 6.800 (3.05) as well as the intersections at KP 3.990 (4.64), KP 4.240 (2.71), KP 4.553 (2.27), KP 5.850 (6.85) and KP 6.556 (3.69) can be pointed as potential locations of possible minor incidents, and in addition, diverse collisions on road can be produced too.

3. The next segment goes from Parque de Cabárceno, KP 7.000, to Santa María de Cayón, KP 12.300, crossing the road A-42. The intersection at KP 7.362 (2.70), the lateral entries at KP 8.791 (1.05), KP 9.470 (3.66), KP 9.587 (2.41) and KP 11.057 (2.61), collisions at KP 8.342 (1.13), incidents on road at KP 11.230 (1.02), and run over animals near KP 12.071 (1.00) are the riskiest potential incidents. Incidents at KP 7.065 (9.78) and KP 7.362 (11.87), collisions at KP 8.013 (2.67), KP 8.158 (3.07), KP 9.126 (2.19), KP 9.232 (2.14), at the whole segment between KP 9.683 and KP 9.817 (more than 30), the lateral entries at KP 8.660 (2.74), KP 8.791 (3.78), KP 9.17 (5.59), KP 9.325 (2.25) and KP 9.47 (27.36), and incidents caused by the presence of curves, speed excesses and risky overtakings can be highlighted as candidate to possible minor incidents.

4. Section between KP 12.300 and KP 18.500. It is between Santa María de Cayón and El Pindio. According to the prediction, medium incidents are due to the presence of a curve at KP 14.181 (1.02), possible incidents on road at KP 16.492 (1.00) and collisions at KP 17.895 (1.04). The rest of potential incidents are minor and related to run over animals, curves and speed excesses.

5. Section between El Pindio, KP 18.500, and Santibáñez, KP 22.300. The expected medium incidents here are related to run over animals at KP 19.746 (1.86), incidents on road around KP 20.613 (1.01) and a lateral entry at KP 21.635 (1.00). As minor incidents, the intersections at KP 21.085 (2.29) and KP 21.453 (2.37), the lateral

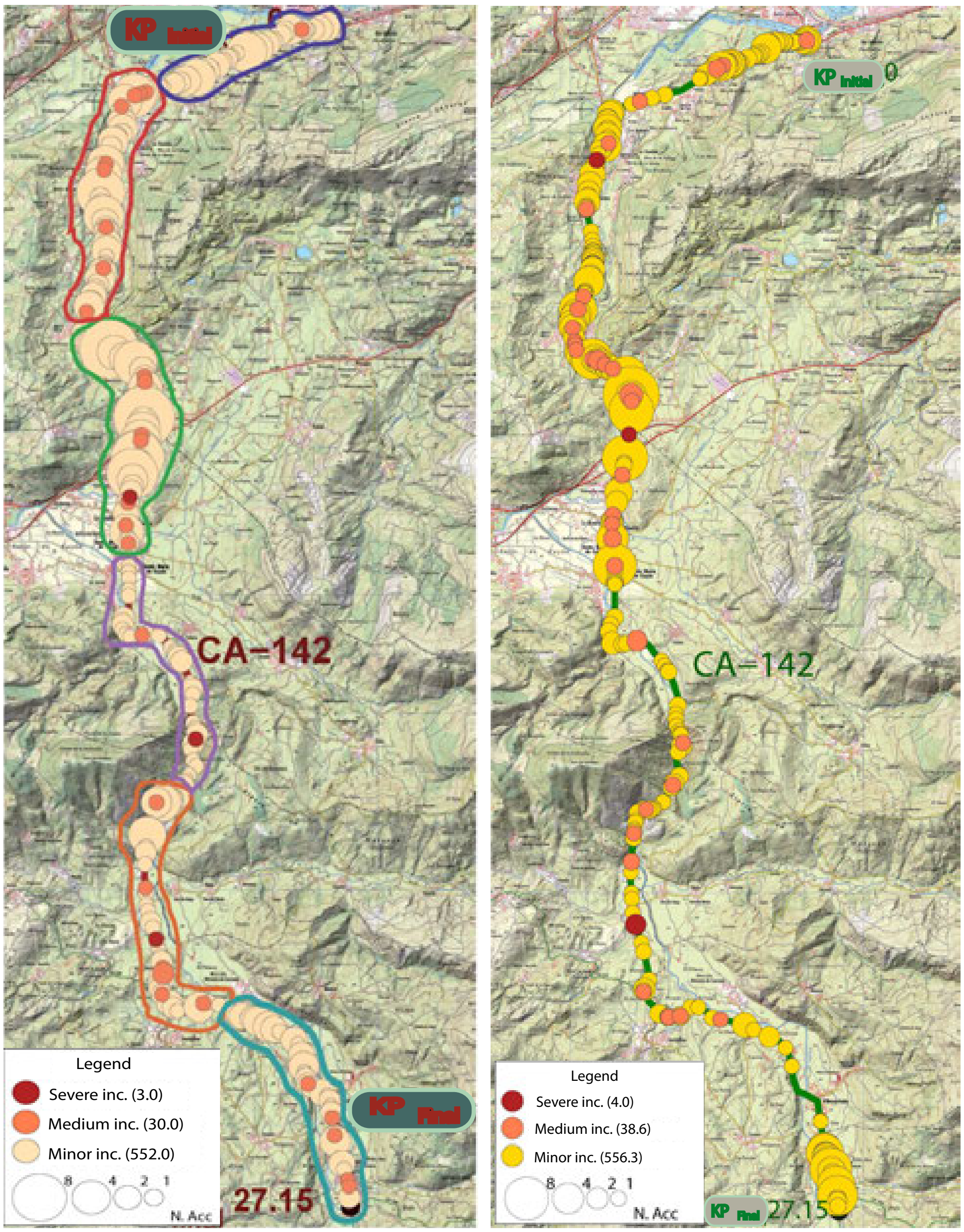


Figure 11.9: On the left hand side the different sections in which the observed accidents of the CA-142 road have been divided are shown. On the right the graphical representation of the incident prediction for the same road is given, which shows the frequencies of level incidents (yellow), medium (orange) and severe (red) by means of circles of proportional diameter to the square root of the frequency of the incidents at the corresponding points.

entry at KP 21.635 (3.01) and possible run over animals, collisions and incidents on road, the presence of some risky curves and speed excess can be pointed out.

6. Section between KP 22.300 and KP 23.700. It goes from Santibáñez to La Granja. The risk is minor in this section. Only the intersection at KP 23.205 (3.55) must be highlighted. Run over animals and possible speed excess can also lead to risks producing collisions.

7. The final section is between KP 23.700 (La Granja) and KP 27.150 (Selaya). The riskiest point corresponds to the curve at KP 27.043 (1.01). Among the incidents of lower risk, the intersection at KP 24.056 (2.82) and possible collisions on road near KP 27.036 deserve to be mentioned.

Once the probabilistic analysis has been carried out in the two directions of the Autonomic road CA-142 the following conclusions are obtained:

Global risk

When studying the cumulative risk, an annual global ENSI value of 0.387 is obtained for the ascending direction and 0.477 for the descending one, resulting a total value of 0.866. This road has a length of 27.045 km, thus, the value of annual ENSI per km is 0.014 upwards, 0.018 downwards and 0.032 in total. From these values it can be concluded that:

- The risk in the upward direction differs slightly from the one in the downward direction, being greater in the downward direction than in the opposite direction.
- The global values of annual ENSI per km, which determine the safety level of the road, have values much smaller than 0.152; thus, this road can be considered to have a good level of safety.

Local risk

Although the road at the global level is considered to have an acceptable level of safety, 98 points of the road have been identified (44 upwards and 54 downwards coinciding many of them at KP) that are potential points to be improved.

As it can be seen in Table 11.8, most accidents will be located at curves or along sections of the road, because of run over animals or incidents with other users of the road. However, it should be noted that there are also points with a significant ENSI value, identified at intersections, lateral entries, or with special relevance at a pedestrian crossing, at the first part of the road.

Table 11.8: Table of the riskiest points of the $CA-142$ road divided in the upward and downward directions, with their probability of incidents sorted by risk level, with their corresponding KP, nodes and local ENSI values.

					ENSI		Severity	
Rank	**Item**	**Item type**	**KP**	**Node**	**Local**	**Severe**	**Medium**	**Minor**
			Ascending direction					
1	400	LateralEntry	9.470	I6021-9470LE	0.0404	0.00107	0.216	1.3
2	3	PedestrianCrossing	0.025	I42-25Pc	0.0158	0.000894	0.0949	0.0104
3	8	Intersection	0.098	I117-98Int	0.0114	0.00374	0.0429	0.209
4	328	Intersection	7.362	I4950-7362Int	0.0103	0.000176	0.0473	0.627
5	582	CurveIn	14.181	I8797-14181Cv	0.00729	0.0051	0.0127	0.046
6	318	Intersection	7.065	I4800-7065Int	0.00692	9.44e-05	0.0301	0.491
7	592	CurveIn	14.472	I8941-14472Cv	0.00655	0.00442	0.0122	0.0488
8	588	CurveIn	14.340	I8884-14340Cv	0.00636	0.00431	0.0118	0.0477
9	884	Intersection	24.056	I13270-24056Int	0.00589	0.00167	0.023	0.144
10	877	CurveIn	23.501	I13167-23501Cv	0.00549	0.00429	0.00724	0.0164
11	359	CurveIn	8.373	I5413-8373Cv	0.00487	0.00327	0.00737	0.104
12	578	CurveIn	14.001	I8740-14001Cv	0.0048	0.00322	0.00886	0.0439
13	559	CurveIn	13.354	I8458-13354Cv	0.00475	0.00317	0.00887	0.0441
14	696	CurveIn	18.655	I10461-18655Cv	0.00466	0.00353	0.00672	0.0187
15	376	LateralEntry	8.791	I5662-8791LE	0.00463	0.000708	0.0199	0.188
16	301	Intersection	6.556	I4543-6556Int	0.0043	0.000242	0.0209	0.183
17	278	Intersection	5.850	I4197-5850Int	0.00388	3.94e-05	0.0147	0.355
18	820	CurveIn	21.780	I12329-21780Cv	0.00333	0.00221	0.00592	0.0438
19	496	Intersection	11.602	I7505-11602Int	0.00299	0.000252	0.0156	0.0687
20	372	LateralEntry	8.660	I5603-8660LE	0.00275	0.00045	0.0122	0.0913
21	400	Incident	9.470	P6018-9470Ssl	0.00267	0.000684	0.01	0.0979
22	366	Collision	8.545	Co5514-8545Sall	0.00266	0.000471	0.0105	0.125
23	867	Intersection	23.205	I13019-23205Int	0.00264	5.11e-05	0.0117	0.176
24	13	CurveIn	0.173	I190-173Cv	0.00262	0.0012	0.0073	0.0653
25	351	Collision	8.211	Co5294-8211Sall	0.00253	0.000446	0.00992	0.122
26	404	LateralEntry	9.587	I6079-9587LE	0.00252	0.000386	0.011	0.0951
27	375	CurveIn	8.745	I5647-8745Cv	0.00242	0.0021	0.00171	0.0137
28	351	Incident	8.211	P5293-8211Sall	0.00223	0.000473	0.00804	0.116
29	347	Collision	8.075	Co5236-8075Sall	0.00219	0.000381	0.00856	0.108
30	332	Incident	7.458	P5009-7458Sall	0.00206	0.000801	0.00759	0.0159
31	333	Collision	7.529	Co5025-7529Sall	0.00203	0.000277	0.00839	0.101
32	368	CurveIn	8.615	I5545-8615Cv	0.00199	0.00166	0.0018	0.0101
33	347	Incident	8.075	P5235-8075Sall	0.00195	0.000416	0.00704	0.1
34	185	Intersection	3.990	I2779-3990Int	0.00193	2.23e-07	0.00609	0.225
35	818	LateralEntry	21.635	I12299-21635LE	0.0019	8e-05	0.00797	0.133

Continued on next page

Table 11.8 – continued from the previous page

					ENSI		Severity	
Rank	**Item**	**Item type**	**KP**	**Node**	**Local**	**Severe**	**Medium**	**Minor**
36	366	Incident	8.545	P5513-8545Sall	0.00187	0.000393	0.00673	0.0977
37	400	Collision	9.470	Co6019-9470Ssl	0.00181	0.000895	0.000206	0.202
38	450	Intersection	10.352	I6792-10352Int	0.00165	1.09e-06	0.00852	0.0734
39	553	CurveIn	13.070	I8371-13070Cv	0.00164	0.00131	0.0017	0.0154
40	333	Incident	7.529	P5024-7529Sall	0.00162	0.000343	0.00584	0.0836
41	350	Collision	8.133	Co5279-8133Sall	0.00159	0.000275	0.00625	0.078
42	681	CurveIn	18.046	I10241-18046Cv	0.00158	0.000811	0.00425	0.0242
43	511	Intersection	11.922	I7736-11922Int	0.00155	3.93e-07	0.00662	0.119
44	391	LateralEntry	9.170	I5887-9170LE	0.00152	3.17e-07	0.00192	0.281
				Downward direction				
1	636	LateralEntry	9.470	I9559-9470LE	0.0349	0.00526	0.15	1.44
2	4	CurveIn	27.043	I50-27043Cv	0.0301	0.0243	0.0326	0.154
3	706	Intersection	7.362	I10596-7362Int	0.0177	0.000405	0.0952	0.56
4	667	CurveIn	8.742	I10023-8742Cv	0.0123	0.0109	0.00818	0.036
5	455	CurveIn	13.930	I6783-13930Cv	0.00985	0.00728	0.015	0.0544
6	240	CurveIn	20.862	I3587-20862Cv	0.00771	0.00506	0.0155	0.0541
7	717	Intersection	7.065	I10761-7065Int	0.00764	0.000118	0.0346	0.488
8	447	CurveIn	14.275	I6667-14275Cv	0.00737	0.00507	0.0134	0.0469
9	689	Collision	8.085	Co10343-8085Sall	0.00608	0.00111	0.0238	0.288
10	440	CurveIn	14.446	I6567-14446Cv	0.00606	0.00408	0.0114	0.0466
11	436	CurveIn	14.572	I6510-14572Cv	0.006	0.00409	0.011	0.0451
12	450	CurveIn	14.111	I6711-14111Cv	0.00591	0.00403	0.0106	0.0498
13	138	Intersection	24.056	I2079-24056Int	0.0057	0.00163	0.0222	0.138
14	673	CurveIn	8.550	I10110-8550Cv	0.00516	0.0035	0.00768	0.106
15	663	LateralEntry	8.791	I9964-8791LE	0.00501	0.000732	0.0221	0.19
16	625	Collision	9.763	Co9396-9763SSC	0.00483	4.76e-10	1.95e-06	1.11
17	692	Collision	7.940	Co10386-7940Sall	0.00439	0.000754	0.0173	0.214
18	697	CurveIn	7.602	I10461-7602Cv	0.00412	0.00272	0.00618	0.1
19	732	Intersection	6.556	I10987-6556Int	0.00411	0.000223	0.0197	0.186
20	402	CurveIn	15.957	I6005-15957Cv	0.00398	0.00293	0.00617	0.0205
21	693	Collision	7.789	Co10400-7789SOS	0.00396	0.000718	0.0153	0.198
22	182	CurveIn	22.415	I2726-22415Cv	0.00376	0.00245	0.00723	0.0422
23	627	Collision	9.728	Co9425-9728Ssl	0.00324	7.97e-11	1.22e-06	0.745
24	689	Incident	8.085	P10342-8085Sall	0.00312	0.000668	0.0112	0.16
25	756	Intersection	5.850	I11357-5850Int	0.00301	3.58e-07	0.0101	0.331
26	146	CurveIn	23.632	I2198-23632Cv	0.00296	0.0021	0.00511	0.0127
27	657	Incident	8.920	P9874-8920Sall	0.0027	0.00069	0.0101	0.0997
28	693	Incident	7.789	P10399-7789SOS	0.00265	0.000571	0.00956	0.134

Continued on next page

Table 11.8 – continued from the previous page

Rank	Item	Item type	KP	Node	ENSI Local	Severity Severe	Medium	Minor
29	692	Incident	7.940	P10385-7940Sall	0.00261	0.000558	0.00942	0.134
30	632	LateralEntry	9.587	I9501-9587LE	0.00242	0.000365	0.0104	0.0997
31	686	Collision	8.230	Co10300-8230Sall	0.00242	0.000426	0.00962	0.113
32	668	Collision	8.660	Co10036-8660SSC	0.00236	0.000398	0.00941	0.114
33	668	LateralEntry	8.660	I10038-8660LE	0.00233	0.000357	0.00999	0.0945
34	991	CurveIn	1.080	I14922-1080Cv	0.00229	0.000956	0.00547	0.11
35	724	LateralEntry	6.800	I10867-6800LE	0.0022	1.47e-05	0.00898	0.18
36	158	Intersection	23.205	I2377-23205Int	0.00216	2.39e-05	0.00868	0.179
37	845	Intersection	3.990	I12708-3990Int	0.00207	2.58e-07	0.00658	0.239
38	683	CurveIn	8.321	I10257-8321Cv	0.00201	0.00168	0.00183	0.0104
39	632	Collision	9.587	Co9499-9587Ssl	0.00197	0.000179	0.00809	0.121
40	523	Intersection	11.922	I7815-11922Int	0.00195	1.09e-06	0.00904	0.123
41	684	Collision	8.278	Co10270-8278Sall	0.00189	0.000358	0.00735	0.0875
42	624	Collision	9.806	Co9381-9806SSC	0.00187	5.36e-12	6.87e-12	0.431
43	676	Collision	8.440	Co10153-8440Sall	0.00186	0.000266	0.0077	0.0909
44	683	CurveIn	23.480149	I2242-23480Cv	0.00186	0.00153	0.00181	0.0117
45	659	CurveIn	8.852	I9905-8852Cv	0.00182	0.00155	0.00144	0.0111
46	681	Collision	8.375	Co10226-8375Sall	0.0018	0.000315	0.00722	0.0827
47	636	Collision	9.470	Co9557-9470Ssl	0.0018	0.000218	0.00709	0.108
48	683	Collision	8.321	Co10255-8321SOS	0.00162	0.000274	0.0065	0.076
49	588	Intersection	10.352	I8824-10352Int	0.0016	1.06e-06	0.00827	0.0713
50	627	LateralEntry	9.728	I9427-9728LE	0.00155	0.00035	0.00738	0.0108
51	657	Collision	8.920	Co9875-8920Sall	0.00154	0.000758	0.000175	0.174
52	629	Collision	9.694	Co9454-9694Ssl	0.00154	1.45e-06	0.00552	0.155
53	664	Collision	8.745	Co9977-8745SSC	0.00153	0.000279	0.00598	0.0735
54	689	RunoverAnimals	8.085	A10341-8085Sall	0.00152	0.000391	0.00525	0.0701

$ENSI^*$ Expected number of equivalent severe incidents
Frequency of severe incident > 9.2
Frequency of medium incident > 0.92
Frequency of minor incident > 0.092

11.5.2 Specific points where safety must be improved

In the same way as done in the other two roads, it has been considered necessary to act on the 20 riskiest points taking into account the two directions, which are shown in Table 11.9 and represented in Figure 11.10, to reduce the likelihood of severe incidents, and thus the global road risk. Next, the riskiest point of this road, a lateral entry, is analyzed in detail.

Table 11.9: Riskiest points of the $CA - 142$ road considering the two directions.

Rank	Item type	KP	Annual ENSI Total	Annual ENSI Ascending	Downward
				Continued on next page	
1	LateralEntry	9.47	0.075	0.040	0.035
2	CurveIn	27.015-27.043	0.030	1.46e-05	0.0301
3	Intersection	7.362	0.028	0.010	0.018
4	PedestrianCrossing	0.025	0.016	0.016	0.000
5	CurveIn	14.181-14.275	0.015	0.007	0.007
6	Intersection	7.065	0.015	0.007	0.008
7	CurveIn	8.615-8.742	0.014	0.002	0.012
8	CurveIn	14.472-14.572	0.013	0.007	0.006
9	CurveIn	14.34-14.446	0.012	0.006	0.006
10	Intersection	0.098	0.012	0.011	0.001
11	Intersection	24.056	0.012	0.006	0.006
12	CurveIn	14.001-14.111	0.011	0.005	0.006
13	CurveIn	8.373-8.550	0.010	0.005	0.005
14	CurveIn	13.820-13.93	0.010	1.32e-05	0.00985
15	LateralEntry	8.791	0.010	0.005	0.005
16	CurveIn	23.501-23.632	0.008	0.005	0.003
17	Intersection	6.556	0.008	0.004	0.004
18	CurveIn	20.730-20.862	0.008	1.38e-05	0.00771
19	Intersection	5.85	0.007	0.004	0.003
20	Collision	7.94-8.060	0.006	0.002033	0.00439
21	Collision	8.085-8.220	0.006	2.69e-04	0.00608
22	LateralEntry	8.66	0.005	0.003	0.002
23	Collision	9.763-9.806	0.005	2.53e-05	0.00483
24	CurveIn	13.354-13.660	0.005	0.005	1.18e-05
25	CurveIn	18.655-18.797	0.005	0.005	2.63e-08

Point 1: Lateral entry at KP 9.47

In this section we analyze the accidents that can occur at the confluence of the lateral entry located at KP 9.47, which is shown in Figure 11.11. To determine the circumstances leading to these potential accidents it is possible to extract from Table 11.10 the information concerning the following types of incidents:

1. Collisions between vehicles of all types (heavy, car and motorcycle) that access or leave the lateral entry and collide side or front-to-side with vehicles that circulate along the main road. These incidents do not occur under special climatic conditions and their consequences can become severe. It is convenient to note that it may mainly be caused by an incorrect action of vehicles entering or exiting the road

Table 11.10: Combination of main variable values contributing to the total ENSI value at the riskiest point of the Autonomic road CA-142 (lateral entry at KP 9.47).

n	**Weather**	**Veh. type**	**Attention**	**Speed**	**Incident**	**Probability**	**% ENSI**
1	Medium	Car	Alert	45	Medium	2.7e-08	29.52
2	Good	Car	Alert	45	Medium	9.57e-09	10.47
3	Very bad	Car	Alert	45	Medium	8.99e-09	9.83
4	Bad	Car	Alert	45	Medium	8.49e-09	9.28
5	Medium	Motorcycle	Alert	45	Medium	4.81e-09	5.26
6	Good	Heavy	Alert	45	Medium	3.55e-09	3.88
7	Medium	Car	Attentive	45	Medium	2.86e-09	3.13
8	Medium	Heavy	Alert	45	Medium	2.78e-09	3.04
9	Good	Car	Alert	67.5	Minor	9.2e-08	2.80
10	Medium	Car	Alert	67.5	Minor	7.97e-08	2.43

CA-142. It may also be attributable to the drivers disruption driving through the CA-142 road, but this event is less likely.

2. There may also be caused by collisions of lower severity but with a greater probability of occurrence, which due to favorable climatic conditions allow drivers to react and avoid more serious consequences in front-to-side or side collisions, or reach collisions because of the occasional dismissal of a driver. This type of accident has minor consequences and can occur in days with good or humid weather.

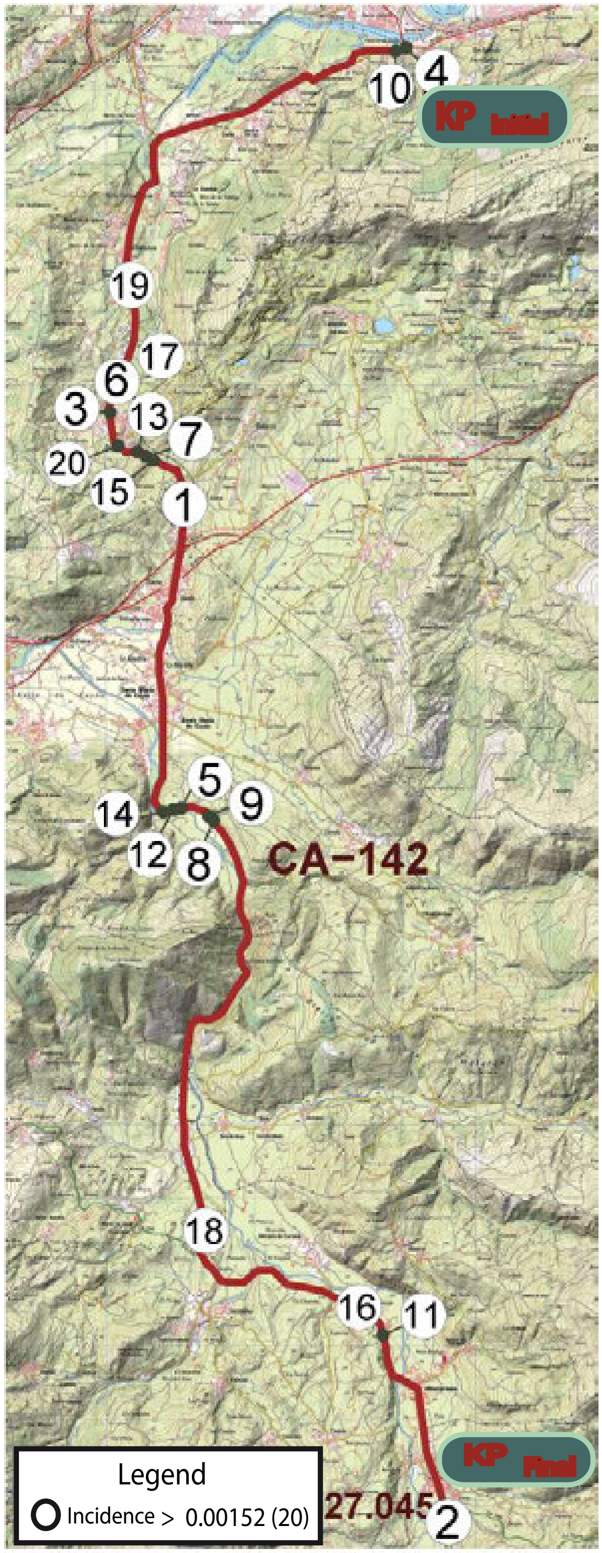


Figure 11.10: Riskiest points of the CA-142 road.

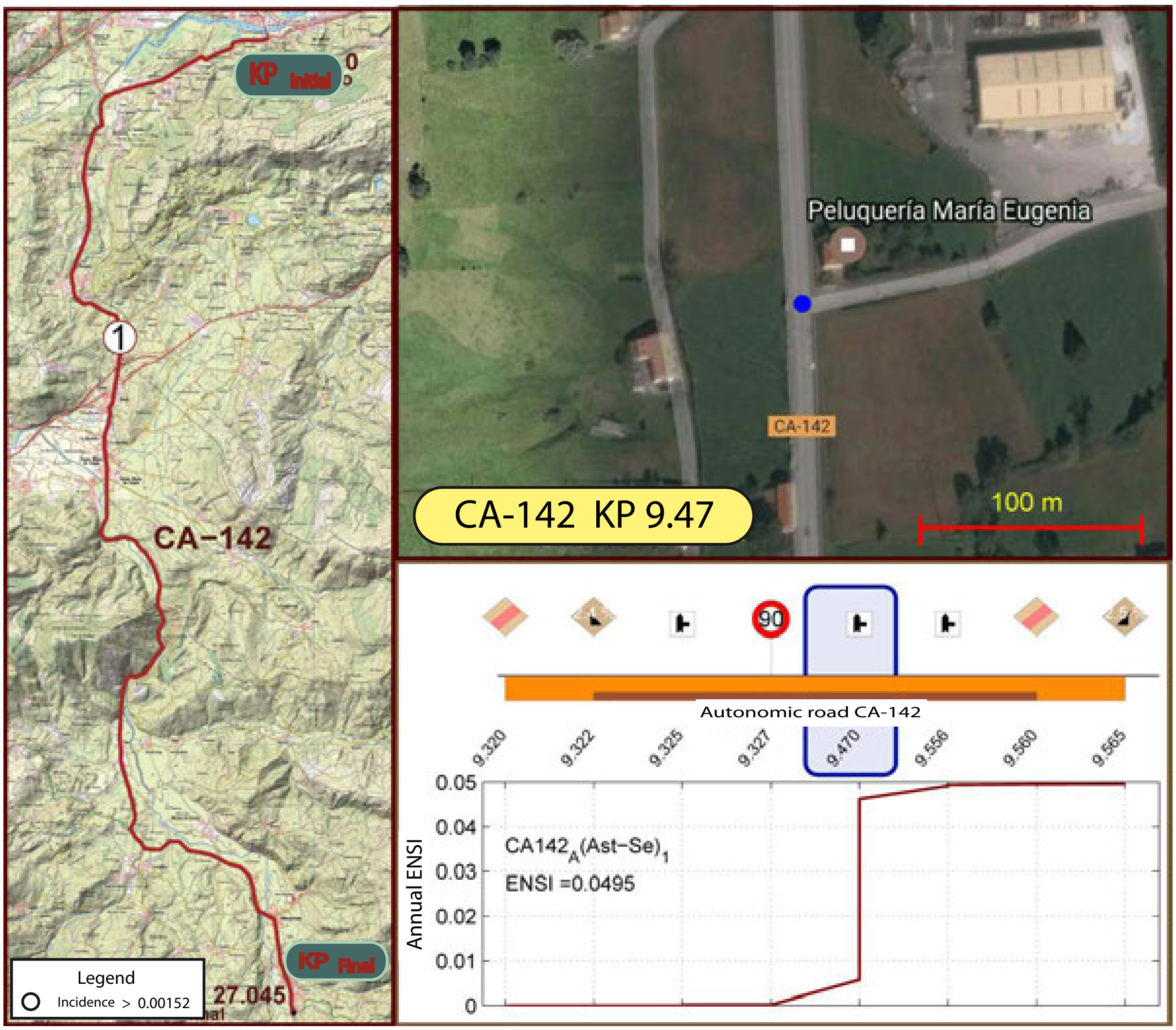


Figure 11.11: Characteristics of the incident at the riskiest point of the Autonomic CA-142 road (lateral entry at KP 9.47).

Part III

CONCLUSIONS AND PUBLICATIONS

Chapter 12

Conclusions

Contents

12.1 Conclusions and future work (in English)

The following conclusions can be drawn from the content of this thesis:

1. Bayesian network models provide an important tool and appear to be the most adequate tools to reproduce and to perform a probabilistic safety assessment of highways and roads. Clearly, they are more powerful than fault and event trees used in nuclear power plants, especially when common causes are present, which is the case (for example, driver's attention, weather, and speed, are some common causes). This means that Bayesian networks can help to improve the safety of highways and roads.

2. The acyclic graph defining the network can be easily constructed by reproducing all the items found when travelling the road or highway. This means that all possible items encountered by the cars and drivers along the road and related to safety must and can efficiently be incorporated into the analysis. A video taken from the cars becomes an essential and obvious tool to this purpose.

3. The proposed model permits: (a) to include all the variables involved in the problem, (b) to reproduce their qualitative dependencies and (c) to quantify the probabilities of any possible marginal or conditional probabilities. This implies reproducing the probabilistic structure of the associated multivariate random variable associated with the highway or road segments. More important, the model permits to quantify

any probability no matter it refers to univariate or multivariate marginals or to conditional probabilities.

4. The construction of the nodes (variables) and the structure of the Bayesian network is very natural because all the items encountered when the vehicles travel along the road are reproduced. A simple list of items can be given for a computer program to build the acyclic graph associated with the Bayesian network automatically. The closed form formulas given in this thesis permit us adding the conditional probabilities in a simple way. Consequently, the practical application of the proposed methodology to real cases is possible and relatively simple.

5. The application of the proposed methodology to real examples with 76 and 129 items and 992 and 1704 variables proves that the method can be applied to very large segments of highways and shows the power of the method to identify sequences of events leading to severe incidents and to quantify their probabilities.

6. Some of the particular examples analyzed show that the method is able to identify the most relevant incident causes and quantify their probabilities of occurrence.

7. As shown with the detailed examples included in this thesis, the conditional probability tables required to quantify the Bayesian network can be defined by closed formulas in general. In particular it has been proven that the resulting conditional probabilities are valid. The proposed method of definition of these tables using trees permits a better understanding of these probabilities and a more organized way of analyzing the different combinations of values of the variables involved and the possibility of automating the obtention of closed form expressions for them.

8. The proposed partitioning technique permits reducing the CPU time drastically. More precisely, using this technique the required CPU time increases linearly with the number of variables instead of the non-linear character of the alternative methods. This makes the proposed method valid to be applied in real cases, where the number of variables can be several thousands. With this technique we can reduce CPU times of hours to minutes.

9. A sensitivity analysis can be easily done with the help of the partitioning technique. This implies that the sensitivity analysis can be done with small subnetworks, that is, in a very reduced time. Range sensitivity analysis plots provide much more information than local sensitivity analysis. More precisely, a sensitivity analysis of a parameter A with respect to another parameter B provides the exact changes of parameter A when the values of parameter B change within a selected interval, while local sensitivity gives only the partial derivative of parameter A with respect to B at the design point. This has important implications from a practical point of

view. In particular it is very important to calibrate the parameters of the Bayesian network.

10. The real examples discussed show that the method can identify relevant incidents, quantify their probabilities of occurrence and find explanations about their causes and a picture of the most frequent situations in which these incidents take place. Once they have been identified, some changes to reduce the associated risks can be done and evaluated from the point of view of safety.

11. The parameter estimation and calibration processes are the most critical parts of the proposed model. To this aim, the collaboration of miscellaneous groups of experts is needed to improve the quality, the credibility of the results and the efficiency of the method. A serious probabilistic safety assessment must put a great effort on this parameter estimation and calibration process.

12. The main limitations of the proposed method come from the difficulty in estimating rare events and the high number of parameters that need to be estimated for each particular item in the road. This takes time and experience.

13. The possibilities of a backward analysis offered by Bayesian networks permits us to investigate the causes of given incidents by forcing them to have probability one (their occurrence) and a recalculation of the probabilities of the remaining events, which can also be observed up to we obtain a satisfactory explanation of the occurrence of the target event. The main advantage of this process consists in that events are quantified and can be sorted by their probabilities.

14. The Bayesian network model can also identify the combination of variable values that contribute more frequently to given events. In other words the circumstances under which critical events occur can be identified. This permits us to orient the adequate corrective actions to them and avoid loss of resources that will be produced if we orient these actions in the wrong direction.

15. Some learning techniques have also been given that allow us considering the expert-group based Bayesian network as a prior that is later corrected when data become available. Conjugate Bayesian methods appear to be the key star in this methodology.

16. We end by mentioning some future work that could include: (a) an improvement of the formulas for obtaining the conditional probabilities, (b) automatic learning and updating techniques, for example, methods based on Bayesian categorical and Dirichlet families conjugate methods or log-linear models, and (c) application to many other real lines that deserve a careful probabilistic safety analysis.

12.2 Conclusiones y trabajo futuro (in Spanish)

Del contenido de esta tesis pueden extraerse las siguientes conclusiones:

1. Los modelos de redes bayesianas constituyen una herramienta importante y parecen ser las herramientas más adecuadas para reproducir y realizar una evaluación de seguridad probabilística de carreteras y autopistas. Es evidente que son mucho más potentes que los árboles de fallos y de sucesos que se utilizan en las centrales nucleares, especialmente cuando hay causas comunes, como es el caso, por ejemplo, de la atención del conductor, el tiempo atmosférico, la velocidad, etc. Esto significa que las redes bayesianas pueden ayudar a mejorar la seguridad de las autopistas y carreteras.

2. El gráfico acíclico que define la red se puede construir fácilmente reproduciendo todos los elementos que se encuentran al recorrer la carretera o autopista. Esto significa que todos los elementos encontrados por los vehículos y los conductores al recorrer la carretera y relacionados con la seguridad deben y pueden ser incorporados eficientemente al análisis. Un vídeo tomado desde los vehículos se convierte en una herramienta esencial y obvia para este propósito.

3. El modelo propuesto permite: (a) incluir todas las variables involucradas en el problema, (b) reproducir sus dependencias cualitativas y (c) cuantificar las probabilidades de cualesquiera probabilidades marginales o condicionales. Esto implica reproducir la estructura probabilística de la variable aleatoria multivariada asociada a los segmentos de carretera o autopista. Más importante aún, el modelo permite cuantificar cualquier probabilidad, independientemente de si se trata de marginales univariadas o ultivariadas o de probabilidades condicionales.

4. La construcción de los nodos (variables) y la estructura de la red bayesiana es muy natural porque se reproducen todos los elementos encontrados cuando los vehículos viajan por la carretera. Una simple lista de elementos basta para que un programa de ordenador construya el gráfico acíclico asociado a la red bayesiana automáticamente. Las fórmulas de forma cerrada dadas en el trabajo permiten agregar las probabilidades condicionales de una manera automática y simple. En consecuencia, la aplicación práctica de la metodología propuesta a casos reales es posible y relativamente sencilla.

5. La aplicación de la metodología propuesta a ejemplos reales con 76 y 129 items y 992 y 1704 variables demuestra que el método puede aplicarse a segmentos muy grandes de carreteras y muestra la potencia del método para identificar secuencias de eventos que conducen a incidentes graves, cuantificando además sus probabilidades.

6. Algunos de los ejemplos particulares analizados en este documento muestran que el método es capaz de identificar las causas de incidentes más relevantes y cuantificar sus probabilidades de ocurrencia.

7. Como se muestra con los ejemplos detallados incluídos en este documento, las tablas de probabilidad condicional requeridas para cuantificar la red bayesiana pueden definirse por fórmulas cerradas en general. En particular, se ha probado que las probabilidades condicionales resultantes son válidas. El método propuesto de definición de estas tablas mediante árboles permite una mejor comprensión de estas probabilidades y una forma más organizada de analizar las diferentes combinaciones de valores de las variables implicadas y la posibilidad de automatizar la obtención de las mismas.

8. La técnica de partición propuesta permite reducir drásticamente el tiempo de CPU. Más precisamente, utilizando esta técnica, el tiempo de CPU requerido aumenta linealmente con el número de variables en lugar del carácter no lineal de los métodos alternativos. Esto hace que el método propuesto sea válido para ser aplicado en casos reales, donde el número de variables puede llegar a varios miles. Con esta técnica se puede reducir los tiempos de CPU de horas a minutos.

9. Se puede hacer fácilmente un análisis de sensibilidad con la ayuda de la técnica de partición. Esto implica que el análisis de sensibilidad se puede realizar con pequeñas subredes, es decir, en un tiempo muy reducido. Las gráficas de análisis de sensibilidad de rango proporcionan mucha más información que el análisis de sensibilidad local. Más precisamente, un análisis de sensibilidad de un parámetro A con respecto a otro parámetro B proporciona los cambios exactos del parámetro A cuando los valores del parámetro B cambian dentro de un intervalo seleccionado, mientras que la sensibilidad local da sólo la derivada parcial del parámetro A con respecto a B en el punto de diseño. Esto tiene implicaciones importantes desde un punto de vista práctico. En particular, es muy importante calibrar los parámetros de la red bayesiana.

10. Los ejemplos reales discutidos en este documento muestran que el método puede identificar incidentes relevantes, cuantificar sus probabilidades de ocurrencia y encontrar explicaciones sobre sus causas y una imagen de las situaciones más frecuentes en las que ocurren estos incidentes. Una vez que se han identificado, algunos posibles cambios para mejorar la seguridad, pueden ser hechos y evaluados de nuevo para comprobar que ya se alcanzan los niveles de fiabilidad deseados.

11. La estimación de parámetros y los procesos de calibración son las partes más críticas del modelo propuesto. Para ello, se necesita la colaboración de diversos grupos de expertos para mejorar la calidad, la credibilidad de los resultados y la eficiencia

de los métodos propuestos. Una evaluación de seguridad probabilística seria debe poner un gran esfuerzo en este proceso de estimación de parámetros y calibración.

12. Las principales limitaciones del método propuesto vienen de la dificultad de estimar los eventos raros y el alto número de parámetros que deben estimarse para cada ítem particular en la carretera. Esto requiere tiempo y experiencia.

13. Las posibilidades de un análisis retroactivo ofrecido por las redes bayesianas permiten investigar las causas de incidentes dados forzándolos a tener una probabilidad uno (su ocurrencia) y recalculando las probabilidades de los sucesos restantes, que también pueden observarse hasta obtener una explicación satisfactoria de la ocurrencia del evento objetivo. La principal ventaja de este proceso consiste en que las probabilidades de ocurrencia de los eventos son cuantificados y pueden ser ordenados por sus probabilidades respectivas.

14. El modelo de red bayesiano también puede identificar la combinación de valores variables que contribuyen con mayor frecuencia a eventos dados. En otras palabras, las circunstancias bajo las cuales ocurren los eventos críticos pueden ser identificadas. Esto permite orientar las acciones correctivas adecuadas y evitar la pérdida de recursos que se producirán si se orientan estas acciones en la dirección equivocada.

15. También se han dado algunas técnicas de aprendizaje que nos permiten considerar la red bayesiana basada en un grupo de expertos como una información "a priori", que luego se corrige cuando los datos están disponibles. Los métodos bayesianos conjugados parecen ser la estrella clave en esta metodología.

16. Se concluye mencionando algunos trabajos futuros que podrían incluir: (a) una mejora de las fórmulas para obtener las probabilidades condicionales, (b) desarrollo de técnicas automáticas de aprendizaje y actualización, por ejemplo, métodos basados en métodos categóricos Bayesianos y métodos conjugados de familias de Dirichlet o modelos log-lineales, y (c) aplicación a muchas otras líneas reales que merecen un cuidadoso análisis probabilístico de seguridad.

Chapter 13

Publications

In this chapter the main publications in journals and congresses resulting from this Thesis are included:

1. Grande, Z., Castillo, E., Mora, E., and Lo, H. K. Highway and Road Probabilistic Safety Assessment Based on Bayesian Network Models. *Computer Aided Civil And Infrastructure Engineering*, 32, 379-396 (2017).

2. Castillo, E., Grande, Z., Mora, E., Lo, H. K., and Xu, X. Complexity Reduction and Sensitivity Analysis in Road Probabilistic Safety Assessment Bayesian Network Models. *Computer Aided Civil And Infrastructure Engineering*, 32, 546-561 (2017).

3. Castillo, E., Grande, Z., Mora, E., Xu, X., and Lo, H. K. Proactive, Backward Analysis and Learning in Road Probabilistic Bayesian Network Models. *Computer Aided Civil And Infrastructure Engineering*, 32, 820-835 (2017).

4. Mora, E., Nogal, M., Grande, Z., OConnor, A., and Castillo, E. Reduction of Risk of Accident Througth Self-explaining Roads. Analysis of Design Consistency Measures, (In revision).

5. Mora, E., Nogal, M., Grande, Z., OConnor, A., and Castillo, E. Improving Risk of Incidents in Roads in Works by Safety Probabilistic Bayesian Networks Methodology, (In process).

6. Mora, E., Grande, Z., and Castillo, E. Bayesian Networks Probabilistic safety Analysis of Highways and Roads. *Computer Aided Systems Theory- Eurocast 2017.Springer LNCS*.

7. Mora, E., Nogal, M., Grande, Z., OConnor, A., and Castillo, E. A critical analysis of consistency measures for self-explaining roads. *Proceedings of 7th Transport Research Arena TRA 2018*.

Bibliography

Alizadeh, S. S., Mortazavi, S. B., and Sepehri, M. M. (2014). Prediction of vehicle traffic accidents using Bayesian networks. *Scientific journal of pure and applied sciences*, 3(6):356–362.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (1993). Conjugate exponential family priors for exponential family likelihoons. *Statistics*, 25:71–77.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (1996). Specification of distributions by combinations of marginal and conditional distributions. *Statistics and Probability Letters*, 26:153–157.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (2001). Conditionally specified distributions: An introduction (with comments and a rejoinder by the authors). *Statist. Sci.*, 16(3):249–274.

Benjamin, J. R. and Cornell, C. A. (1970). *Probability statistics and decision for civil engineers*. McGraw-Hill, New York.

Bielza, C., Li, G., and Larrañaga, P. (2011). Multi-dimensional classification with bayesian networks. *International Journal of Approximate Reasoning*, 52:705–727.

Bielza, C., Moral, S., and Salmerón, A. (2015). Recent advances in probabilistic graphical models. *International Journal of Intelligent Systems*, 30:207–208.

Bobbio, A., Portinale, L., Minichino, M., and Ciancamerla, E. (2001). Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliability Engineering and System Safety*, 71(3):249–260.

Castillo, E., Conejo, A. J., Mínguez, R., and Castillo, C. (2006). A closed formula for local sensitivity analysis in mathematical programming. *Engineering Optimization*, 38(1):93–112.

Castillo, E., Gallego, I., Ureña, J. M., and Coronado, J. M. C. (2009). Timetabling optimization of a single railway track line with sensitivity analysis. *Top*, 17(2):256–287.

Castillo, E., Grande, Z., and Calviño, A. (2016a). A Markovian-Bayesian Network for risk analysis of High Speed and Conventional Railway Lines Integrating Human Errors. *Computer Aided Civil And Infrastructure Engineering*, 31:193–218.

Castillo, E., Grande, Z., and Calviño, A. (2016b). Bayesian networks based probabilistic risk analysis for railway lines. *Computer Aided Civil and Infrastructure Engineering*, 31:681–700.

Castillo, E., Grande, Z., Calviño, A., Nogal, M., and O'Connor, A. (2016c). Probabilistic Safety Analysis of High Speed Railway Lines including Human Errors. In Pombo, J., editor, *Proc. Third Intern. Conf. on Railway Technology: Research, Development and Maintenance*, number Paper 73, Stirlingshire, UK. Civil-Comp Press.

Castillo, E., Gutierrez, J. M., and Hadi, A. S. (1997). *Expert Systems and Probabilistic Network Models*. Springer Verlag, New York.

Castillo, E., Hadi, A. S., Conejo, A., and Fernández-Canteli, A. (2004). A general method for local sensitivity analysis with application to regression models and other optimization problems. *Technometrics*, 46(4):430–444.

Chen, D. (2014). *Predictive accident modeling for highway transportation system using Bayesian networks*. Ms. thesis, Memorial University of Newfoundland.

de Oña, J., Mujalli, R. O., and Calvo, F. J. (2011). Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. *Accident Analysis and Prevention*, 43(1):402–411.

DeGroot, M. H. and Schervish, M. J. (2002). *Probability and Statistics*. Addison Wesley, 3rd edition.

Deublein, M., Schubert, M., Adey, B. T., and de Soto Garcia, B. (2015). A Bayesian network model to predict accidents on Swiss highways. *Infrastructure Asset Management*, 2(4):145–158.

Deublein, M. K. (2013). *Roadway accident risk prediction based on Bayesian probabilistic networks*. Ph. {d}., ETH Zürich, Zürich.

Doob, J. L. (1953). *Stochastic Processes*. New York: John Wiley and Sons.

Elvik, R. (2007). State-of-the-art approaches to road accident black spot management and safety analysis of road networks. Technical Report 883, Institute of Transport Economics, Norwegian Center for Transport Research.

Elvik, R. (2008). A survey of operational definitions of hazardous road locations in some european countries. *Accident Analysis & Prevention*, 40(6):1830 – 1835.

Graf, M., Nishijima, K., and Faber, M. H. (2009). Bayesian updating in natural hazard risk assessment. *Australian Journal of Structural Engineering*, 9(1):34–35.

Grossman, D. and Domingos, P. (2004). Learning bayesian network classifiers by maximizing conditional likelihood. In Brodley, C., editor, *Proceedings of the 21st International Conference on Machine Learning*, pages 1–8, Banff, Canada. ACM digital library.

Hauer, E., Harwood, D., Council, F., and Griffith, M. (2002). Estimating Safety by the Empirical Bayes Method: A Tutorial. *Transportation Research Record: Journal of the Transportation Research Board*, 1784:126–131.

Hongguo, X., Huiyong, Z., and Fang, Z. (2010). Bayesian Network-Based Road Traffic Accident Causality Analysis. In *Information Engineering (ICIE), 2010 WASE International Conference on*, volume 3, pages 413–417.

Kala, Z. and Puklicky, L. (2009). Variance-Based Methods for Sensitivity Analysis in Civil Engineering. In Yuan, Y., Cui, J., and Mang, H. A. E., editors, *Proceedings of the International Symposium on Computational Structural Engineering*, pages 991–997, Shanghai, China. Springer.

Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. Cambridge: Chapman & Hall, 1 edition.

Krol, A. (2014). Application of the Bayesian network to identify the correlations in the circumstances of the road accidents for selected streets in Katowice. *Logistyka*, 4:2947–2957.

Larrañaga, P. and Moral, S. (2011). Probabilistic graphical models in artificial intelligence. *Applied Soft Computing*, 11(2):1511–1528.

Li, W., Carriquiry, A., Pawlovich, M., and Welch, T. (2008). The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. *Accident Analysis & Prevention*, 40(4):1531–1542.

Lin, F., Jiang, Y., Xu, Z. X., and Dai, L. (2011). The Analysis and Prevent in Traffic Accidents Based on Bayesian Network. In *Emerging Engineering Approaches and Applications*, volume 1 of *Advanced Engineering Forum*, pages 21–25. Trans Tech Publications.

Mahboob, Q. (2014). *A Bayesian Network Methodology for Railway Risk, Safety and Decision Support*. Ph. {d}. thesis, Technische {U}niversität {D}resden, Dresden.

Mbakwe, A. C. (2011). *Modeling highway traffic safety in Nigeria using Bayesian network*. Ph. {d}. thesis, Morgan State University, Baltimore.

Mu, H.-Q. and Yuen, K.-V. (2016). Ground motion prediction equation development by heterogeneous Bayesian learning. *Computer-Aided Civil and Infrastructure Engineering*, 31(10):761–776.

Papakosta, P. and Straub, D. (2013). A Bayesian network approach to assessing wildfire consequences. In *11th International Conference on Structural Safety & Reliability ICOSSAR*, New York. Columbia University.

Pawlovich, M., Li, W., Carriquiry, A., and Welch, T. (2006). Experience with Road Diet Measures: Use of Bayesian Approach to Assess Impacts on Crash Frequencies and Crash Rates. *TRR: Journal of the Transportation R. B.*, 1953:163–171.

Persaud, B. and Lyon, C. (2007). Empirical Bayes before-after safety studies: Lessons learned from two decades of experience and future directions. *Accident Analysis & Prevention*, 39(3):546–555.

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis*, 22(3):1–12.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley and Sons, New York.

Scherb, A., Garré, L., and Straub, D. (2015). Probabilistic Risk Assessment of Infrastructure Networks Subjected to Hurricanes. In *12th International Conference on Applications of Statistics and Probability in Civil Engineering*, Vancouver. ICASP1.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(a):271–280.

Spackova, O. and Straub, D. (2015). Dynamic Bayesian network for probabilistic modeling of tunnel excavation processes. *Computer-Aided Civil and Infrastructure Engineering*, 28(1):1–21.

Sun, H. and Betti, R. (2015). A Hybrid Optimization Algorithm with Bayesian Inference for Probabilistic Model Updating. *Computer-Aided Civil and Infrastructure Engineering*, 30(8):602–619.

Wang, H., Yajima, A., Liang, R. Y., and Castaneda, H. (2015). Modeling of External Corrosion in Underground Pipelines Based on the Integration of Markov Chain Monte Carlo Techniques and Clustered Inspection Data. *Computer-Aided Civil and Infrastructure Engineering*, 30(4):300–316.

Wang, L., Lu, H.-p., Zheng, Y., and Qian, Z. (2014). Safety analysis for expressway based on Bayesian network: a case study in China. *Computer modelling and new technologies*, 18(12C):438–444.

Yuen, K. V. and Mu, H. Q. (2015). Real-Time System Identification: An Algorithm for Simultaneous Model Class Selection and Parametric Identification. *Computer-Aided Civil and Infrastructure Engineering*, 30(10):785–801.

Zhang, K. and Shi, P. (2015). Transportation Security Assessment Method for a Mountainous Freeway Using a Bayesian Network. In *ICTE 2015*, chapter 368, pages 2891–2896. ASCE.

Zong, F., Xu, H., and Zhang, H. (2013). Prediction for Traffic Accident Severity: Comparing the Bayesian Network and Regression Models. *Mathematical Problems in Engineering*, 2013:9.