MODELING ROAD ACCIDENTS USING COMPOUND ROAD ENVIRONMENTS

Ana Fernandes
Researcher in Civil Engineering, CESUR – Center for Urban and Regional Systems, Technical University of Lisbon
Lisbon, Portugal, e-mail: acapote@civil.ist.utl.pt

José Neves
Assistant Professor in Civil Engineering, CESUR – Center for Urban and Regional Systems, Technical University of Lisbon
Lisbon, Portugal, e-mail: neves@civil.ist.utl.pt

Submitted to the 3rd International Conference on Road Safety and Simulation,
September 14-16, 2011, Indianapolis, USA

ABSTRACT

The main purpose of this paper is to present a methodology of using compound road environments in modeling of road accidents which was developed in a PhD study. The technique presented involves evaluating the influence of pavement surface properties in road accidents.

Nowadays, the most common method of studying the influence of certain road features in accidents is to consider uniform road segments characterized by a unique feature. However, when an accident is related to the infrastructure, its cause is usually not a single road characteristic but rather a combination of characteristics. Therefore, the methodology presented here looks at the road as a complete environment, thereby overcoming the limitations inherent in considering only uniform road sections.

The proposed methodology consists of: (i) dividing a sample of roads into segments; (ii) grouping them into quite homogeneous road environments using cluster analysis, considering characteristics like traffic, road geometry and weather conditions; (iii) identifying the influence of skid resistance and texture depth on road accidents in each environment by using generalized linear models; and (iv) validating the results by doing simulations in virtual scenarios with different pavement properties using *PC-Crash* software to reproduce vehicle crashes.

In terms of implementing pavement maintenance programs, these results are of primary importance to determining threshold levels of skid resistance and texture depth. In terms of scientific knowledge, the cluster analysis used to identify road environments is an innovative and valid alternative way of choosing segments to be used in accident prediction models as it considers compound road environments instead of uniform ones.

Keywords: accidents, road environments, cluster analysis, generalized linear models, pavements

INTRODUCTION

Numerical modeling is a common tool for estimating the frequency of road accidents. Various models that have been intensively tested and validated are available in the literature (Persaud & Dzbik, 1993; Maher & Summersgill, 1996; Mountain et al., 1996; Cardoso, 1996; Abdel-Aty & Radwan, 2000; Wang et al., 2006; Caliendo et al., 2007). These numerical models are useful for estimating the expected number of accidents based on variables related to traffic, road geometry and road environment. In addition, it would also be desirable if these models could estimate the effect of changing one or more of those variables on the expected number of accidents. Adjustment of models to estimate the frequency of accidents is based on historical accident data and on the characteristics of experimental sections selected from the road network.

The occurrence of accidents is a typical case that cannot be modeled as continuous data using a normal regression. Generalized linear models (GLZ), first presented in 1972 by Nelder and Wedderburn and later developed by McCullagh & Nelder (1983), are considered the most suitable models for determining relationships between accidents and characteristics of traffic and road geometry (Maher & Summersgill, 1996; Cardoso, 1996; Lebaye, 1997; Wood, 2002; Greibe, 2003).

Modeling of accidents is often based on uniform road sections, but this is an important constraint. In fact, a study focused on a single characteristic of a road segment (circular or straight alignments, width of lanes, shoulders properties) is very limiting because it does not consider the influence of other possible variables of the road environment (Cardoso, 1996). The most appropriate technique is to consider compound road environments characterized by similar properties. In order to obtain these types of road environments, cluster analysis can be a useful statistical tool. This technique is suitable for classifying and recognizing objects and grouping them according to similar characteristics. Cluster analysis has the advantage of grouping the objects without having to set criteria for inclusion in a given group beforehand.

The most important surface properties related to pavement adherence are skid resistance and texture depth. With respect to the influence of road infrastructure on accidents modeling, different research studies have been conducted in order to evaluate the influence of adherence, as measured by skid resistance and texture depth, on accident risk (Rizemberg et al., 1976; Yerpez & Ferrandez, 1986; Roe et al., 1991; Ferrandez, 1993; Gothié, 2000; Carney & Styles, 2005). In general, the results clearly confirm that traffic safety depends on these surface pavement properties. For example, an increase in the accident rate is normally observed when the pavement surface shows low skid resistance values. However, the tendencies observed are strongly dependent on the road environment. This means that the relationships vary from one case study to another, and it is not possible to establish a fully defined relationship (Patte, 2005). Taking this into account, levels of adherence in quality control should not be always the same. Different categories of roads belonging to different regions will obviously have different surface properties. The relative importance of the pavement characteristics of road sections with different traffic volumes, road geometry and weather conditions must therefore be assessed.

In the specific case of the influence of skid resistance on accident risk, the biggest challenge is to achieve the best relationship with other road characteristics, such as traffic flow and geometrical design. Accident risk tends to be higher when braking forces and/or lateral forces are unusually high: as in the case of collisions at intersections; accidents on curves (Ferrandez, 1993). Some researchers have proposed linear functions between skid resistance and the risk of accident (Noyce et al., 2005; Murad, 2006), but others authors believe that non-linear functions are more suitable. Accident risk is usually expressed as occurrences per million vehicles.km (/Mvkm), where occurrences may be victims or accidents with and/or without victims. A good review of studies conducted in Europe o the influence of skid resistance on accident risk was compiled by Wallman & Astrom (2001).

The present work also addresses the importance of numerical modeling of road accidents, in this case using a new methodology based on the concept of compound road environments. The construction of these types of road environments was based on cluster analysis. This is the first research study using cluster analysis for this objective. Application of this methodology has demonstrated that it is possible to achieve a more realistic approach to the multiple facets of road infrastructure that could affect the occurrence of accidents. Skid resistance and texture depth were also identified as the most important surface pavement characteristics in accident risk analysis. Modeling of road accidents was based on generalized linear models and the results obtained were validated in virtual scenarios using the *PC-Crash* software program.

NEW METHODOLOGICAL APPROACH

The proposed methodology is presented in Figure 1 and basically consists of:

(1) Dividing a sample of roads into segments
(2) Grouping them into quite homogeneous road environments using cluster analysis, taking into consideration characteristics like traffic, road geometry and weather conditions
(3) Identifying the influence of a specific feature of the infrastructure on road accidents in each environment by using generalized linear models.
(4) Finally, validating the results obtained in the previous step by doing simulations in virtual scenarios by using software that reproduces vehicle crashes.

In the first phase, a sample of roads must be chosen for the study. This selection should be done randomly or using a non-random sampling method, depending on the data availability. This phase is important because the sample selected must be representative of the roads from the larger road network being studied. In fact, to avoid a biased sample, different types of roads (primary and secondary infrastructures) should be chosen to ensure a varied distribution throughout the road network and to include accidents of different origins. During this phase, good characterization of road segments is desirable. Information about traffic, the presence of intersections, urban characteristics, road geometry, and weather and pavement conditions should be collected.

The second phase involves defining road environments. To separate sample into different road environments with distinct traffic characteristics, road geometry and weather and pavement conditions, a cluster analysis must be performed by using the most appropriate criteria.

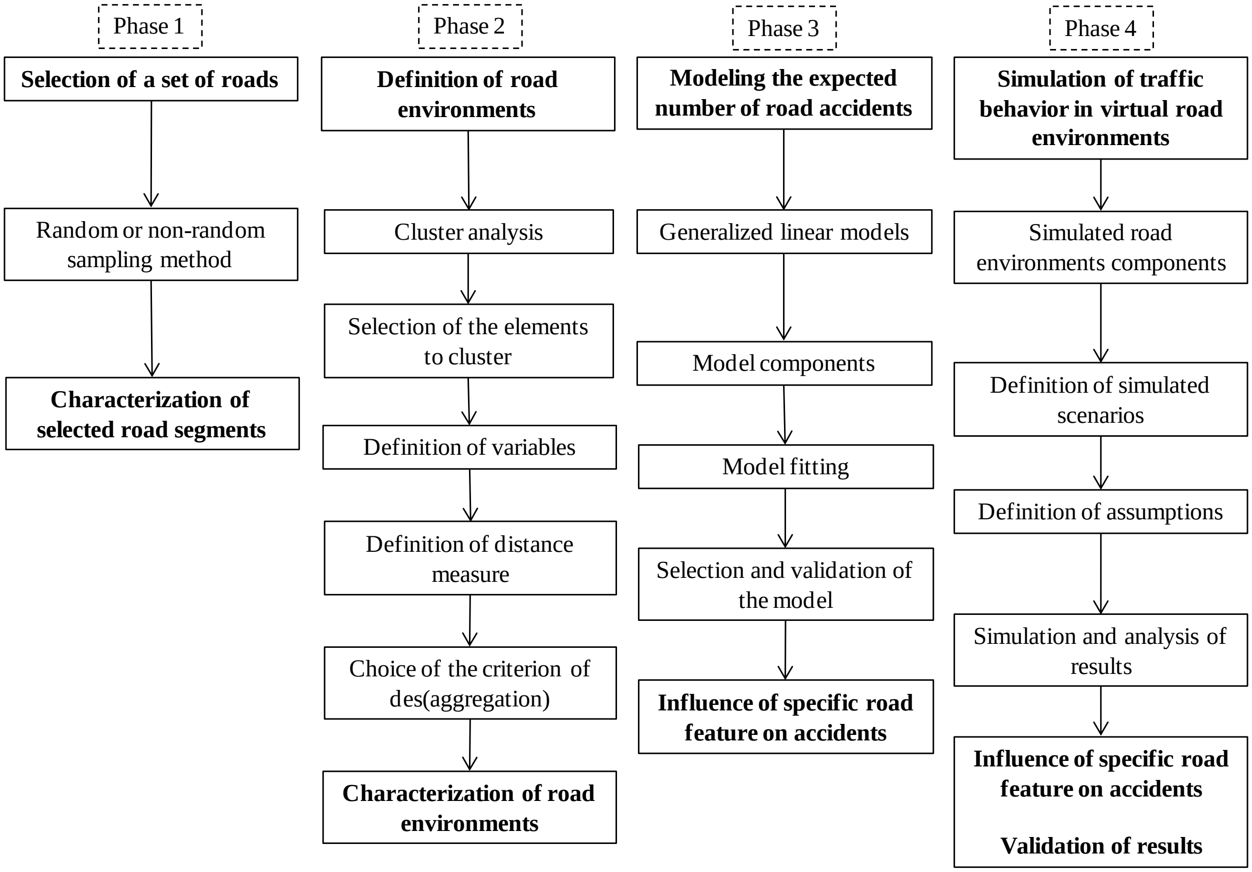

Figure 1 Proposed methodology

The modeling of the expected number of road accidents, in the third phase, consists of using generalized linear models in order to assess the influence of the specific road feature on accidents. In each case, the selected variable should be the most representative of the feature under analysis. The generalized linear models should be calibrated assuming that, in each road environment, the segments are homogeneous.

Taking into account that the statistical analysis and the database obtained from the selected roads present some limitations, the fourth phase is based on simulation of traffic behavior in virtual road environments. A software program designed to reconstruct accidents is a useful tool in order to overcome some of the difficulties inherent in other processes of simulation or observation under real conditions and to validate and complement results from the previous modeling phase.

This methodology will be illustrated by the example presented in the next section.

EVALUATION OF COMPOUND ROAD ENVIRONMENTS

Application of the new methodological approach presented in this paper consists of evaluating the influence of pavement surface properties in road accidents. This work is a small part of a PhD study done at the Technical University of Lisbon (Fernandes, 2010).

In this chapter, the selection and characterization of a sample of roads is presented, as well as the definition of compound road environments using cluster analysis.

Selection of roads

Roads were selected by applying a sequential type of non-random sampling method, which was considered the most appropriate for the existing conditions despite the disadvantages of non-random sampling. The Portuguese Road Administration only had complete data available for some roads, making it impossible to randomize the set of roads comprising the Portuguese road network. The final set was adjusted according to the information available in the Portuguese Road Administration database and comprises eight roads (A to H), spanning a total length of 254 km.

To avoid a biased sample, roads belonging to different categories (primary and secondary networks) were chosen to ensure a varied geographical distribution throughout the country and include good and bad levels of pavement conditions and accidents.

Characterization of the pavement surface was obtained from a pavement conditions survey conducted in 1999 by the Portuguese Road Administration. The parameters chosen to represent pavement surface conditions were skid resistance (represented by the coefficient of friction) and texture depth. Coefficient of friction (CAT) and texture depth (AAE) were measured using a SCRIM and a laser-based texture meter device, respectively. The International Friction Index (IFI) is a parameter that takes into account both skid resistance and texture depth, and its value is independent of the device that was used to obtain the measures. For the SCRIM, the formula that relates the IFI with CAT and AAE measurements is presented in Equations 1 and 2, where CAT is measured with SCRIM at 60 km/h. An analysis of the IFI, CAT and AAE is presented in Table 1.

$$IFI=-0{,}0141+0{,}875\times CAT\times e^{\left(\frac{-39{,}5}{S_p}\right)} \tag{1}$$

$$S_p=17{,}63+93\times AAE \tag{2}$$

Table 1 Analysis of IFI, CAT and AAE by road

Road	IFI			CAT			AAE (mm)		
	Mean	Percentile 85	Standard Deviation	Mean	Percentile 85	Standard Deviation	Mean	Percentile 85	Standard Deviation
A	17	20	5.7	34	37	8.3	0.584	0.688	0.1193
B	32	35	3.8	58	64	5.9	0.729	0.789	0.0698
C	38	42	3.3	72	78	4.7	0.650	0.736	0.0821
D	31	33	2.5	61	63	3.4	0.586	0.636	0.0508
E	20	25	3.5	40	48	6.7	0.576	0.642	0.0627
F	35	37	1.8	68	71	2.3	0.622	0.704	0.0712
G	36	39	3.9	67	70	4.1	0.700	0.784	0.1573
H	37	39	2.9	69	72	2.9	0.675	0.752	0.1099

Accident data include only accidents with victims between 1997 and 2002. This information was obtained from the Directorate-General for Traffic. However, even with a standardized form filled out by police at the time of the accident, there are often inconsistencies and a lack of information, making data analysis difficult. Figure 2 shows some information about accidents on different roads.

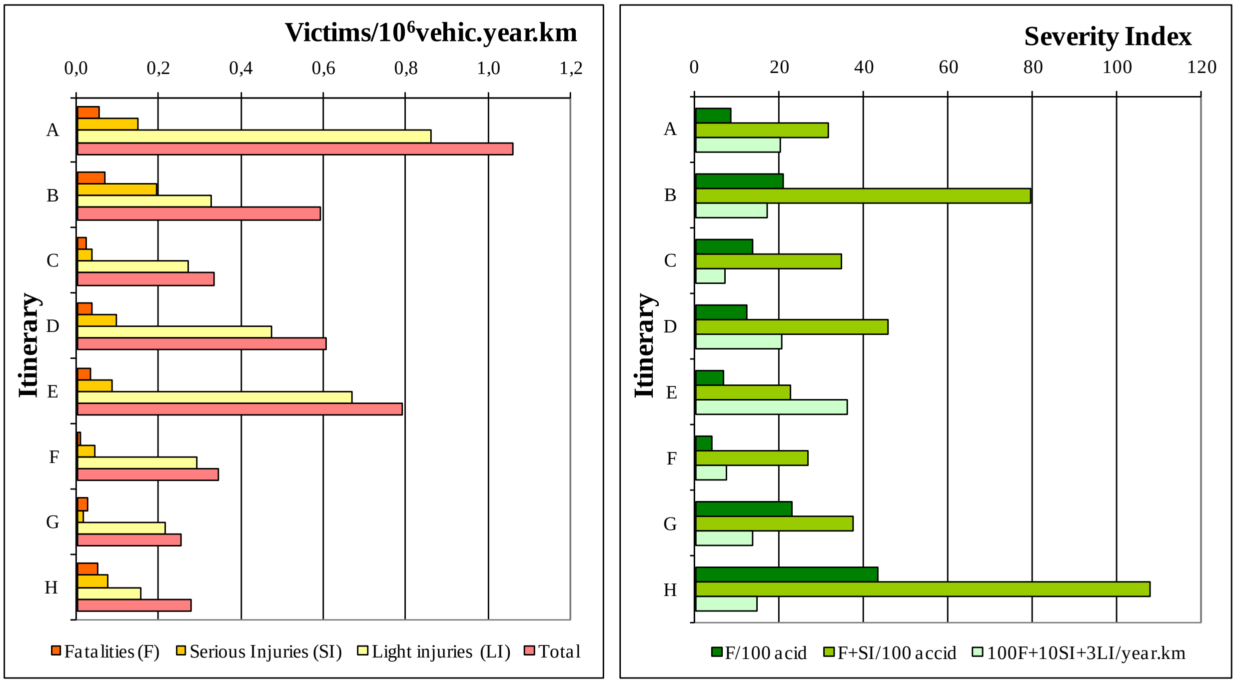

Figure 2 Accident analysis by road

In order to model the frequency of accidents as a function of pavement surface properties, the roads were divided into 1000-meter-long segments. If the sampling unit had been smaller, there would have been a risk of most segments showing zero victims, resulting in too many zeros in the sample. If the sampling unit had been larger, the heterogeneity in each segment would be greater, making it more difficult to define them.

Characterization of road segments

Analyzing the influence of pavement on road accidents without considering other explanatory variables is almost impossible. In addition to skid resistance and texture depth, information about traffic, the presence of intersections, urban characteristics, road geometry and weather conditions was also collected. This information was obtained from the Portuguese Road Administration and the Hydro Resources Information System.

The variables were carefully chosen to closely reflect reality and to appropriately characterize the road environments (RE). For choosing the final group of variables, the correlation matrix among them was calculated. Some of the variables showed significant correlation, which would negatively affect the analysis if all of them were considered simultaneously. It was decided to select representative variables of traffic, road geometry and weather condition with lower inter-variable correlation and with the highest correlation with the accident variables. The final set of selected variables is presented in Table 2.

Table 2 Selected variables to characterize road environments

Variables		Mean	Standard deviation	Coefficient of variation
Percentage of heavy traffic (%)	%H_TRAF	9.7	5.55	0.574
Average speed (km/h)	AV_SP	86.1	4.50	0.052
85th percentile speed (km/h)	SP_85	93.6	3.79	0.041
Percentage of a segment’s stretch in intersections	%EXT_I	11.7	19.36	1.653
Percentage of a segment’s stretch in urban zones	%EXT_UZ	7.4	23.96	3.228
Curved stretch extension (m)	EXT_C	320.9	264.23	0.823
Class of curvature [0 to 4]	CL_C	1.8	1.47	0.816
Class of longitudinal gradient [0 to 1]	CL_G	0.9	0.43	1.086
Annual precipitation (mm)	A_PREC	880.6	461.22	0.524

Definition of compound road environments

Some road environments require higher levels of skid resistance. To isolate different road environments with distinct traffic characteristics, road geometry and weather conditions, a cluster analysis was undertaken using STATISTICA 6.0 software.

Prior standardization of the variables was done to prevent the weight of those with the highest value and most dispersion from being reflected in the measure of similarity used in the cluster analysis.

The number of groups was determined through a dendogram produced by the hierarchical technique (Figure 3). This information was then used in the optimization technique, from which the composition of the final groups was obtained. In the hierarchical technique, the Ward criterion was selected as the criterion of (dis)aggregation of individuals. In applying the optimization technique, iterative partitive k-means clustering was the method used.

Since the cluster analysis was done with a set of 254 elements (segments), Figure 3 cannot show the distribution of those segments into groups. However, there is a clear trend to form between four and seven groups. The final result was a set of seven distinct road environments, a solution which presented better, statistically significant results (Table 3). Table 4 summarizes information about each cluster necessary to classify each road environment and Table 5 maps the distribution of each road segment in each cluster, where the letter represents the road and the number represents the segment.

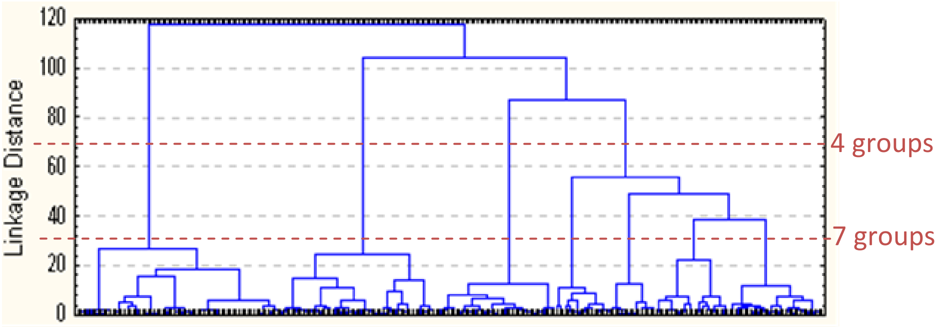

Figure 3 Dendogram produced by the Hierarchical Technique / Ward Criterion

Table 3 Variance analysis

Variables	SSB	Df	SSW	Df	F	Significance (*p-value*)
%H_TRAF	195.7324	6	57.2676	247	140.7018	0.000000
AV_SP	214.6450	6	38.3550	247	230.3798	0.000000
SP_85	206.4890	6	46.5110	247	182.7625	0.000000
%EXT_I	136.3774	6	116.6226	247	48.1399	0.000000
%EXT_UZ	192.6142	6	60.3858	247	131.3105	0.000000
EXT_C	119.1264	6	133.8736	247	36.6319	0.000000
CL_C	128.6011	6	124.3989	247	42.5573	0.000000
CL_G	112.7746	6	140.2254	247	33.1078	0.000000
A_PREC	227.1807	6	25.8193	247	362.2206	0.000000

where

SSB	Sum of squares between (groups)
SSW	Sum of squares within (groups)
Df	Degrees of freedom
F	Statistical Test F

Table 4 Means of variables for each cluster

Variables	Cluster 1 (RE1)	Cluster 2 (RE2)	Cluster 3 (RE3)	Cluster 4 (RE4)	Cluster 5 (RE5)	Cluster 6 (RE6)	Cluster 7 (RE7)	Mean	Standard Deviation
%H_TRAF	26%	4%	8%	7%	10%	9%	9%	10%	6%
AV_SP (km/h)	81	81	84	83	89	94	85	86	5
SP_85 (km/h)	90	90	94	92	91	101	93	94	4
%EXT_I	7.9%	29.2%	2.1%	3.8%	10.4%	7.3%	50.3%	11.7%	19.4%
%EXT_UZ	23%	87%	1%	1%	0%	0%	0%	7%	24%
EXT_C (m)	320	220	41	466	486	471	270	321	264
CL_C	2.3	1.8	0.1	2.2	3.0	2.3	1.7	1.8	1.5
CL_G	0.3	0.0	0.1	0.3	0.8	0.6	0.2	0.4	0.4
A_PREC (mm)	1058	735	591	693	1669	490	722	881	461
Nº of segments	19	15	63	38	55	39	25		

Table 5 Distribution of each road segment in each cluster

Cluster 1: 19 segments	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
	A15	A16	A17	A18	A19									
Cluster 2: 15 segments	E9	E15	E22	E23	E24	E29	E30	E31	E32	E33	E34	E37	E38	E39
	C1													
Cluster 3: 63 segments	E19	E36	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
	D13	B4	B5	B10	B20	B21	C2	C3	C4	C6	C10	C13	C14	C15
	C16	C17	C18	C19	C20	C22	C23	C24	C25	C26	C27	C28	C29	C32
	C33	C34	C35	C36	C37	C38	C39	C40	F2	F3	F5	F6	F10	F11
	F14	F15	F16	F17	F18	F22	F24							
Cluster 4: 38 segments	E1	E2	E3	E4	E5	E6	E7	E8	E10	E12	E13	E16	E17	E20
	E26	E27	E28	E35	B3	B7	B8	B12	B19	C7	C8	C9	C11	C30
	C31	C41	C43	F1	F7	F8	F9	F12	F21	F23				
Cluster 5: 55 segments	B1	B13	B14	B17	G1	G2	G3	G4	G5	G6	G7	G8	G10	G11
	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	G25
	G26	G27	G28	G29	G30	G31	G32	G33	G34	G35	G36	G37	G38	G39
	G40	G41	G42	G43	G44	G45	G46	G47	G48	G49	G50	G51	G52	
Cluster 6: 39 segments	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
	H15	H16	H17	H19	H20	H21	H22	H23	H24	H25	H26	H27	H28	H29
	H30	H31	H32	H33	H34	H35	H36	H37	H38	H39	H40			
Cluster 7: 25 segments	E11	E14	E18	E21	E25	D14	D15	B2	B6	B9	B11	B15	B16	B18
	C5	C12	C21	C42	F4	F13	F19	F20	F25	H18	G9			

With cluster analysis, one starts with the assumption that the characteristics of each element in each cluster are homogeneous. However, while this may naturally occur in some variables, it is not true for all of them. In the final seven groups, some features clearly differentiate the cluster, while others are less important and show some variation (although there is less variation within groups than between groups).

From Table 4, we can conclude the following:

- RE1 segments differ from the others with their very high percentage of heavy traffic and a significant proportion of their stretch in urban zones;
- RE2 is characterized by segments with an extremely high percentage of their stretch in urban zones, a low percentage of heavy traffic and segments with no longitudinal gradient;
- RE3 segments are mostly straight, with low average annual precipitation;
- RE4 is characterized by segments with 50% of their stretch curved;
- RE5 is characterized by very high precipitation and segments with 50% of their stretch curved, a longitudinal gradient and a small radius of curvature;
- RE6 is characterized mainly by very low precipitation and speeds above the acceptable speed, and by segments with 50% of their stretch curved with a longitudinal gradient;
- RE7 is characterized by a heavy presence of intersections in rural areas.

MODELING ROAD ACCIDENTS

Modeling the expected number of road accidents (Naccid/km) was achieved by using generalized linear models in order to assess the influence of pavement surface properties on road accidents. The IFI was chosen to represent the pavement surface. The models were calibrated considering that, in each road environment, segments present homogeneous traffic characteristics, road geometry and weather conditions, and regression was done with only one explanatory variable, the IFI (RE_IFI).

The traffic volume was introduced in the model as an "offset variable" to represent exposure to risk. To do this, a variable representative of the total traffic accumulated during the period of analysis, $TRAF_{ACUM}$, was created. Equation 3 represents the overall model, where β_1 is the regression coefficient associated with variable IFI and β_0 is the independent term.

$$N_{accid}/km = TRAF_{ACUM_i} \times exp\left(\beta_0 + \beta_1 \times IFI_i\right) \qquad (3)$$

Over-dispersion, which is common in this type of analysis, influenced the choice of regression method. Whenever over-dispersion was not negligible, the "over-dispersed" extra-Poisson regression proved most appropriate for modeling the frequency of accidents. However, in cases where over-dispersion was negligible, the Poisson regression was employed. The over-dispersion problem was evaluated using the Lagrange Multiplier Test for the parameter K of the Negative Binomial Distribution (H_0: K = 0). If the test is not significant, then over-dispersion is not a problem for the set of data and Poisson regression is adequate (Table 6).

The model was adjusted using historical accident data and the maximum likelihood method was used for calibration. The Wald statistical test was used to evaluate the statistical significance of the estimated coefficients.

Table 7 shows models that were calibrated for each road environment. Figure 4 compares the observed number of accidents per km between 1997 and 2002 and the respective numbers modeled by RE_IFI. Figure 5 shows absolute residuals.

Table 6 Lagrange Multiplier Test results

			p-value	
Road Environment	z	$K < 0$	$K > 0$	$K \neq 0$
RE1_IFI	1.865	0.969	0.031	0.062
RE2_IFI	1.406	0.92	0.08	0.16
RE3_IFI	1.461	0.928	0.072	0.144
RE4_IFI	1.614	0.947	0.053	0.107
RE5_IFI	1.196	0.884	0.116	0.232
RE6_IFI	0.722	0.765	0.235	0.47
RE7_IFI	1.129	0.87	0.13	0.259

Table 7 Calibrated models

Road Environment	Models – Expected number of accidents per km
RE1_IFI	$Naccid/km=TRAF_{ACUM_i}\times exp\ (-13{,}33159143-0{,}07200512\times IFI_i)$
RE2_IFI	$Naccid/km=TRAF_{ACUM_i}\times exp\ (-13{,}30480979-0{,}06142634\times IFI_i)$
RE3_IFI	$Naccid/km=TRAF_{ACUM_i}\times exp\ (-15{,}03645919)$
RE4_IFI	$Naccid/km=TRAF_{ACUM_i}\times exp\ (-13{,}15179749-0{,}06462487\times IFI_i)$
RE5_IFI	$Naccid/km=TRAF_{ACUM_i}\times exp\ (-14{,}61692487-0{,}03868385\times IFI_i)$
RE6_IFI	$Naccid/km=TRAF_{ACUM_i}\times exp\ (-10{,}54165889-0{,}14862363\times IFI_i)$
RE7_IFI	$Naccid/km=TRAF_{ACUM_i}\times exp\ (-13{,}10895401-0{,}06231246\times IFI_i)$

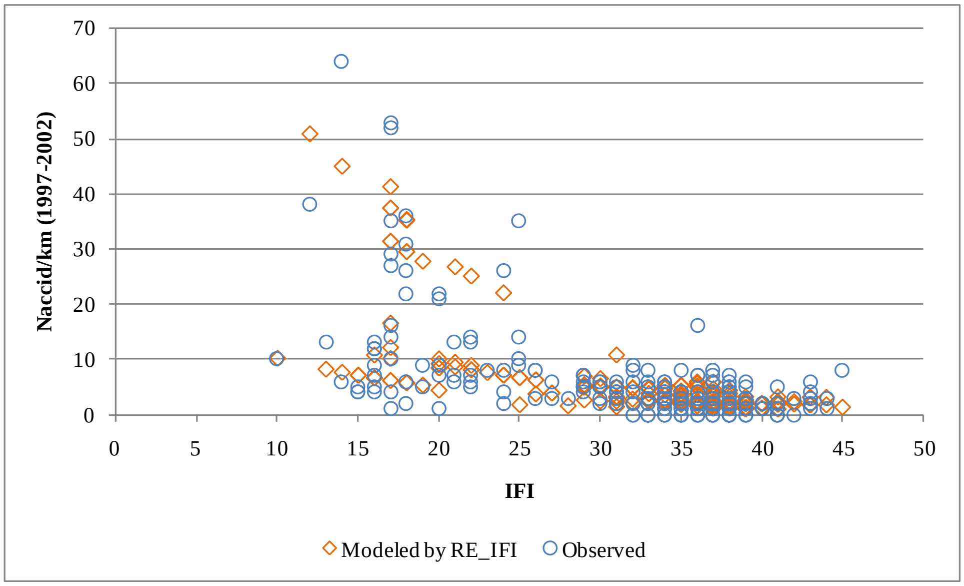

Figure 4 Comparison between the observed number of accidents and the respective numbers modeled by RE_IFI

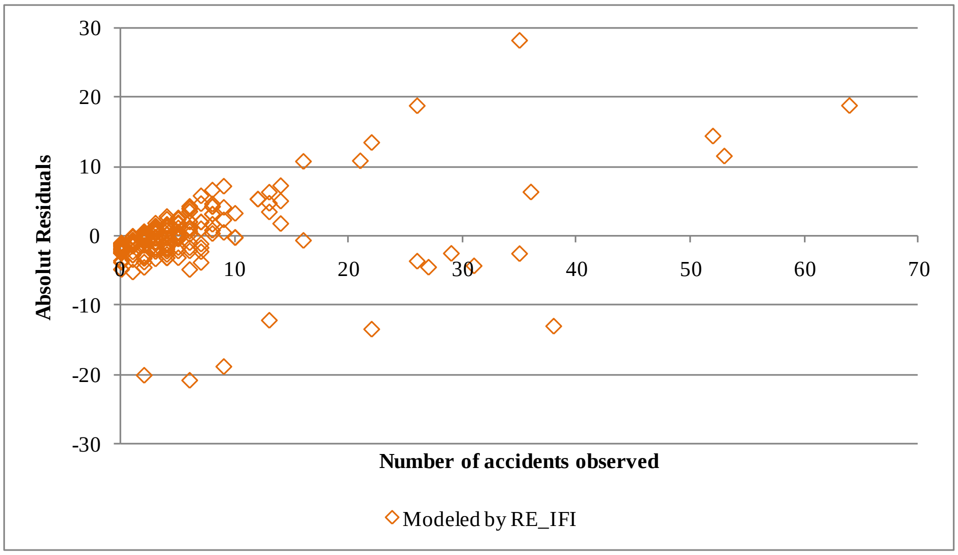

Figure 5 Absolute residuals

DISCUSSION OF RESULTS

Analysis of the modeling results

The main purpose of the PhD study that is presented in part in this paper was to establish skid resistance and texture depth threshold values based on safety criteria to be implemented in pavement maintenance programs. In order to do that, the influence of surface characteristics on accident occurrence was evaluated by analyzing the coefficients associated with the explanatory variable IFI and measuring the impact that a change (δ) in IFI produces in the expected number of accidents, ($E[Y_i]$, as presented in Equation 4. Table 8 represents the analysis of statistical significance of the estimated coefficient β_1.

$$E[Y_i|IFI_i+\delta]=E[Y_i|IFI_i]\times(exp(\delta\beta_1)) \quad (4)$$

Table 8 Statistical significance of β_1, the regression coefficient associated with variable IFI

Model	Coefficient β_1	Stand. Dev.	Significance (*p-value*)	Confidence interval (95%)	ρ^2	R^2
RE1_IFI	-0.07200512	0.0346	0.037	[-0.140; -0.004]	0.093	≈ 0.3
RE2_IFI	-0.06142634	0.0385	0.111	[-0.137; 0.014]	0.090	≈ 0.3
RE3_IFI	-0.008	0.0142	0.595*	[-0.035; 0.020]	0.002	-
RE4_IFI	-0.06462487	0.0253	0.010	[-0.114; -0.015]	0.144	≈ 0.4
RE5_IFI	-0.03868385	0.0220	0.078	[-0.082; 0.004]	0.013	-
RE6_IFI	-0.14862363	0.0496	0.003	[-0.246; -0.051]	0.059	≈ 0.2
RE7_IFI	-0.06231246	0.0179	0.001	[-0.097; -0.027]	0.217	≈ 0.5

*Not significant

From the analysis of the results, we were able to conclude that there are, basically, three environments (E_i) where the pavement properties significantly, and distinctively, influence the occurrence of accidents:

- E_1: Rural environment with a heavy presence of urban characteristics (e.g., urban zones and intersections) – RE1 and RE2;
- E_2: Environment characterized by a considerable predominance of intersections in a rural environment – RE7;
- E_3: Environment with curved segments, high longitudinal gradients and average speed higher than the tolerable speed – RE6.

Figure 6 represents accident risk (Naccid/vehic.km) as a function of IFI. Here there are some noticeable differences between the three environments. Clearly in E_3, when IFI falls below 30, there is a sharp increase in accident risk, reaching unacceptable levels for IFI values below 22. In environments with urban characteristics (E_1) and intersections (E_2), where braking maneuvers are often necessary, the permissible IFI values fall to 20 and 25, respectively, while for lower values a strong increase in accident risk is expected.

Figure 7 shows how CAT and AAE threshold values were calculated as a function of accident risk. In E_2, levels of skid resistance below 45 and texture depth of less than 0.4 mm are not recommended.

The International Friction Index, coefficient of friction (measured with SCRIM at 60 km/h) and texture depth threshold values (minimum value) for the three environments were established based on these results. As the UK Highways Agency (2004) recommends, the values should not be set too low. Therefore, safety values were also set with an eye towards preventive intervention. These safety values translate into an increase over the minimum values of 0.1 mm for texture depth, 10 units for the coefficient of friction and 3 to 5 units for the IFI (Table 9).

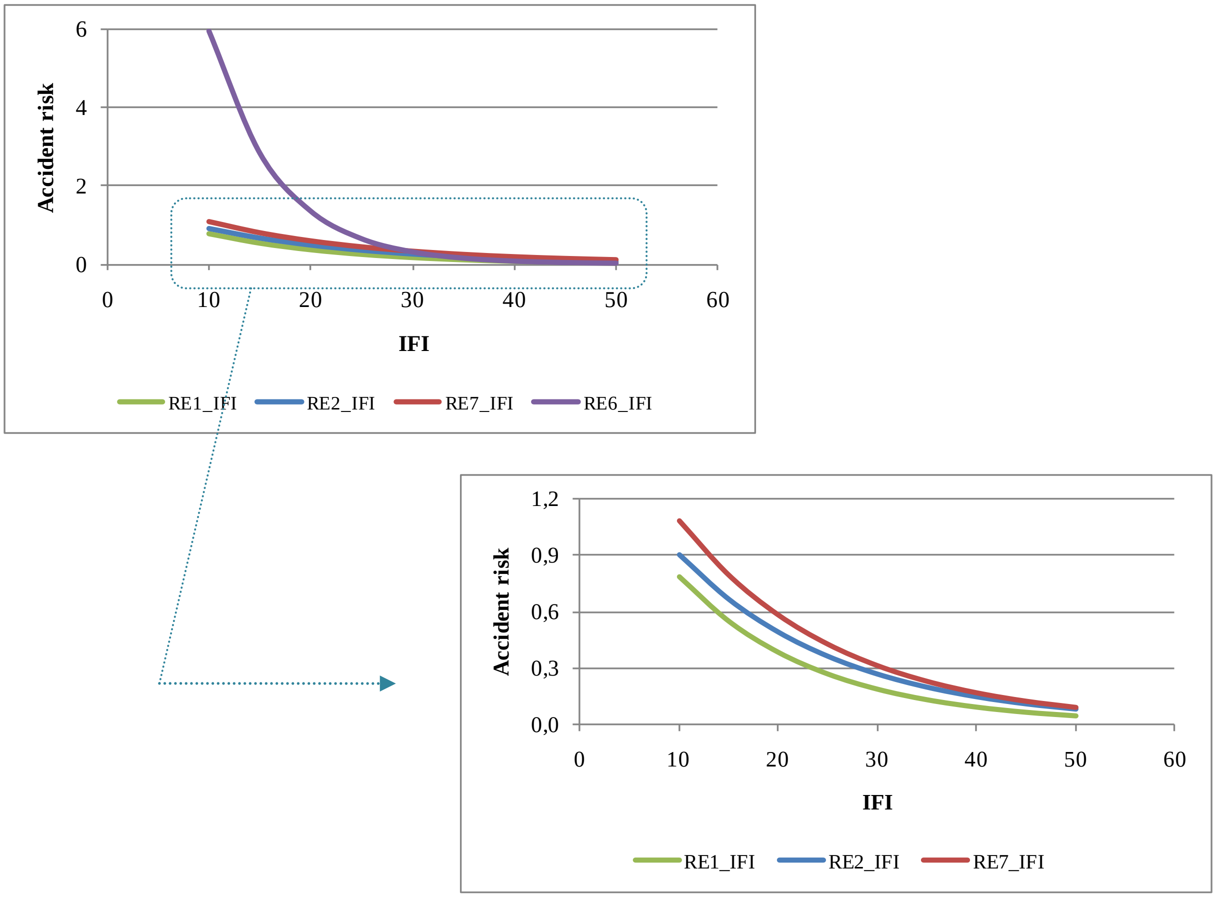

Figure 6 Accident risk and IFI

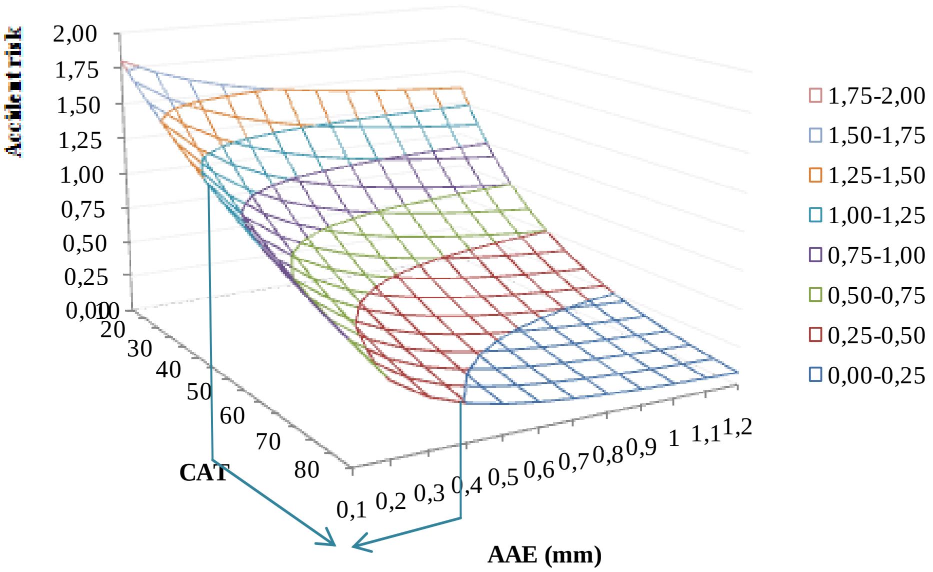

Figure 7 Accident risk, CAT and AAE in E_2

Table 9 Threshold values for IFI, CAT and AAE

	Minimum Values / Safety Values		
	IFI	CAT	AAE (mm)
E_1	20 / 25	40 / 50	0.4 / 0.5
E_2	25 / 28	45 / 55	0.4 / 0.5
E_3	30 / 33	50 / 60	0.5 / 0.6

Modeling difficulties

With respect to the modeling of the expected number of road accidents, difficulties identified by other authors (Ferrandez, 1993) took place primarily during the calibration process (e.g., problems with statistical significance, over-dispersion and reliability of accident data).

As for statistical significance, accidents are considered rare events, which results in an excess of zeros in the sample, which constitutes a problem for modeling.

The sample size (254 segments) proved to be too small. The small size was due to the difficulty of finding available information on all the variables used in the study. This sample was divided into seven road environments, leading to smaller data sets, which prevented successful adaptation of negative binomial regression models, which work poorly with small samples.

The quality of accident data is one of the most crucial points of the modeling. Some accident records showed inconsistencies.

Associating a measure of skid resistance with an accident becomes very difficult when some years have elapsed between the pavement condition survey and the occurrence of an accident. To overcome this problem, in this study, it was assumed that the measurements taken in 1999 represented the conditions observed between 1997 and 2002, which is, in fact, an approximation rather than the actual truth.

Prior to executing the complex modeling process, every precaution was taken to ensure that the final output would be able to be used to calculate the benefits associated with improving pavement surface properties to reduce the expected number of accidents.

Simulation with PC-Crash

Reconstruction of road accidents is essential for an understanding of the factors that gave rise to them. The *PC-Crash* program has proven to be an effective tool in helping experts in accident reconstruction by simulating the movement and collision of vehicles. The use of *PC-Crash* in this study was important for assessing the influence of surface properties on vehicular movement, and for validating and complementing results from the previous modeling step. This kind of software was chosen because of difficulties inherent in other processes of simulation or observation in real-life conditions.

The characteristics of the segments selected for simulation are similar to those of the road environments in which they operate. Collision and skidding are the most frequent types of accidents in the segments under study. To simulate the maneuvers that cause them, traffic safety was evaluated in three possible scenarios: (i) taking curves, (ii) braking on a straight section and (iii) braking on a curved section. The simulations were performed by varying the speed and the conditions of pavement friction. To consider the sensitivity of coefficient of friction to speed, the *PC-Crash* Wet Friction sequence was adopted. The results validate and support the minimum values identified in the previous section:

- Taking curves is significantly affected by vehicle speed, curvature radius and the coefficient of transverse friction. Skidding occurs when a dangerous combination of factors is observed: high speed (>100 km/h), low friction values (<25 for 60 km/h) and a curvature radius of less than 500 m;
- On straight sections, stopping distances are greater in segments with smaller longitudinal gradients. With friction values less than 45 (measured at 60 km/h), there is a slight change of direction during the braking maneuver, especially for speeds above 90 km/h. For values of approximately 25, there is a trajectory deviation for low speeds and skidding for speeds above 90 km/h;
- On curves, the vehicle can only be safely immobilized with high friction values.

CONCLUSIONS

In the literature, the cluster analysis used to identify different road environments is presented as an innovative and valid alternative for choosing the segments to be used in road accident prediction models. This methodology has the major advantage of taking into account compound road environments (characterized by traffic, road geometry, weather conditions, etc.), thus countering the tendency to consider uniform segments, which is cited as a limitation of other models.

However, this approach also has some weaknesses. In addition to requiring careful definition of variables to be used in defining the road environment, the groups formed are not, in most cases, completely homogeneous and there is some variation of characteristics within the same group, even if within-group variation is less than between-group variation.

Results show that road environments where braking maneuvers (E_1 and E_2) are more common or those with small radii of curvature and high speeds (E_3) require higher skid resistance and texture depth levels.

The Portuguese Highways Agency recently recognized the importance of research studies to support the development of maintenance programs for surface characteristics to be incorporated into pavement management systems. This work aims to contribute a set of values for skid resistance and texture depth maintenance established based on safety criteria.

This work also constitutes a further scientific attempt to establish relationships between functional characteristics of pavement and road accidents by using a set of roads selected from the Portuguese road network.

REFERENCES

Abdel-Aty, M. A., & Radwan, E. A. (2000). “Modeling traffic accident occurrence and involvement”, *Accident Analysis and Prevention*, 633-642.

Caliendo, C., Guida, M., & Parisi, A. (2007). “A crash-prediction model for multilane roads”, *Accident Analysis and Prevention*, 657-670.

Cardoso, J. L. (1996). “Assessment of the relationships between road characteristics, speed and road accidents. Application to two-lane roads with two ways”, PhD Thesis in Civil Engineering. Lisbon, Portugal. Technical University of Lisbon (in Portuguese).

Carney, P., & Styles, E. (2005). “A pilot study of relationship between macrotexture and crash occurrence”. Victoria, Australia, ARRB Transport Research.

Fernandes, A. (2010). “Maintenance programmes for pavement surface characteristics based on road safety criteria”, PhD Thesis in Civil Engineering. Lisbon, Portugal. Technical University of Lisbon (in Portuguese).

Ferrandez, F. (1993). “Analyse des accidents. Infrastructure et securité”. *Bulletin de liaison dês Laboratoires des Ponts et Chaussées*, 19-26.

Gothié, M. (2000). “Apport à la securité routiére des caractéristiques de surface des chaussées”, *Bulletin des Laboratoires des Ponts et Chaussées,* 5-12.

Greibe, P. (2003). “Accident prediction models for urban roads”, *Accident Analysis and Prevention,* 273-285.

Highways Agency (2004). “HD 28/04: Skid resistance. Design Manual for Roads and Bridges”, Vol 7, Section 3. London, UK. Highways Agency.

Lebaye, Y. (1997). “Modélisation de l'insecurité routière en relation avec l'infrastructure. Mise en place d'une méthodologie d'application du modèle linéaire généralisé”. CETE Normandie-Centre.

MacCullagh, P., & Nelder, J. (1983). “*Generalized Linear Models”.* London, Chapman & Hall.

Maher, M., & Summersgill, I. (1996). “A comprehensive methodology for the fitting of predictive accident models”, *Accident Analysis and Prevention,* 281–296.

Mountain, L., Bachir, F., & Jarrett, D. (1996). “Accident prediction models for roads with minor Junctions”, *Accident Analysis and Prevention,* 695-707.

Murad, M. M. (2006). “Modeling asphalt pavement friction and wet-pavement traffic accidents for two-lane rural highways”. *International Journal of Pavements*.

Nelder, J. A., & Wedderburn, R. W. (1972). “Generalized linear models”, *Journal of the Royal Statistical Society,* 370-384.

Noyce, D. A., Bahia, H. U., Yambó, J. M., & Kim, G. (2005). “Incorporating road safety into pavement management: maximizing asphalt pavement surface friction for road safety improvements”. Midwest Regional University Transportation Center.

Patte, L. (2005). “Accidents par perte d'adhérence: relation adhérence-sécurité routière et analyse préalable à l'intervention”, *Bulletin des Laboratoires des Ponts et Chaussées,* 169- 178.

Persaud, B., & Dzbik, L. (1993). “Accident prediction models for freeways”. *Transportation Research Record,* 55-60.

Rizemberg, R. L., Burchet, J. L., & Napier, C. T. (1976). “Accidents on rural interstate and parkway roads and their relation to pavement friction”, *Transportation Research Board*.

Roe, P., Webster, D., & West, G. (1991). “The relation between the surface texture of roads and accidents”. London, Transportation Road Research Laboratory.

Wallman, C. G., & Astrom, H. (2001). “Friction measurement methods and the correlation between road friction and traffic safety. A literature review”. Linkoping, Swedish National Road and Transport Research Institute (VTI).

Wang, X., Abdel-Aty, M., & Brady, P. A. (2006). “Crash estimation at signalized intersections: significant factors and temporal effect”, *TRB 2006 Annual Meeting*. Washington.

Wood, G. R. (2002). “Generalised linear accident models and goodness of fit testing”, *Accident Analysis and Prevention,* 417-427.

Yerpez, J., & Ferrandez, F. (1986). “Caractéristiques routières et sécurité. Reconaissance de la contribution des facteurs route dans la génèse des accidents”. Arcueil, France: INRETS.