

# ACIDENTES RODOVIÁRIOS DO TIPO COLISÕES NAS ESTRADAS NACIONAIS NA REGIÃO DO MINHO: MODELOS DE PREVISÃO PARA SEGMENTOS RODOVIÁRIOS

Jocilene Otilia Costa<sup>1</sup>, Maria Alice Prudêncio Jacques<sup>2</sup>, Elisabete Fraga Freitas<sup>3</sup>, Paulo António Alves Pereira<sup>4</sup>

<sup>1</sup>Universidade do Minho, Departamento de Engenharia Civil, 4800-058 Azurém, Guimarães, Portugal

email: jocilene.mt@gmail.com

<sup>2</sup>Universidade de Brasília, Departamento de Engenharia Civil, Brasília, Brasil

<sup>3</sup>Universidade do Minho, Departamento de Engenharia Civil, 4800-058 Azurém, Guimarães, Portugal

<sup>4</sup>Universidade do Minho, Departamento de Engenharia Civil, 4800-058 Azurém, Guimarães, Portugal

---

## Sumário

*Este artigo estuda as colisões ocorridas na região norte portuguesa (1999-2010). Apresenta-se uma contribuição inicial para o desenvolvimento dum modelo de previsão de acidentes (MPA). Foram analisados os parâmetros: frequência de colisões, tráfego e características geométricas, de 88 segmentos de 200m fora de áreas urbanas, inseridos em estrada nacionais de duas vias (uma em cada sentido). O procedimento de modelação, baseado em equações de estimação generalizadas (EEG), foi aplicado a cinco bases de dados distintas (uma geral e 4 derivadas). A estrutura de correlação predominante nos modelos foi a permutável. Os principais fatores contribuintes para a ocorrência de colisões foram TMDA, sinuosidade vertical e densidade de pontos de acesso. O MPA aceitável identificado considera um período cumulativo de acidentes de quatro anos.*

---

**Palavras-chave:** Modelos de previsão de acidentes; equações lineares generalizadas; colisões.

## 1 INTRODUÇÃO

O número crescente de acidentes de trânsito em estradas rurais e urbanas criou a necessidade de desenvolver estratégias para ajudar os gestores rodoviários a reduzir essas ocorrências. Devidamente registados e analisados os dados relativos a acidentes rodoviários permitem a identificação das áreas ou dos locais onde as medidas de segurança têm um maior potencial para bom desempenho e eficácia [1]. A definição de medidas de segurança eficazes, visando a redução dos acidentes de viação em um dado elemento rodoviário (segmento ou interseção), é beneficiada pelo conhecimento a respeito das estimativas da frequência esperada de acidentes para cada elemento rodoviário, juntamente com a contribuição do elemento e das suas características operacionais para o desempenho de segurança esperado. As abordagens comumente utilizadas para fornecer estimativas da frequência esperada de acidentes, atuais e futuras, para segmentos de vias ou interseções consideram dados históricos de acidentes, modelos estatísticos baseados em análise de regressão, estudos antes-depois e avaliação de especialistas [2].

Os modelos de previsão de acidentes (MPAs) são ferramentas importantes na promoção da segurança viária em diferentes elementos rodoviários. Eles podem fornecer estimativas precisas para a frequência total de acidentes para uma localização por unidade de tempo, que é geralmente uma função do tráfego da estrada e das características geométricas. No entanto, de acordo com [2] e [3], a interpretação de um coeficiente dum MPA como sendo o verdadeiro efeito de uma mudança incremental de uma variável associada, geralmente não é satisfatória. Estes autores relatam que esta situação pode ocorrer devido a problemas, nomeadamente: (i) a causa-efeito assumida entre algumas características viárias e os acidentes pode não ser sempre verdadeira, (ii) a presença de uma forte correlação entre as variáveis independentes do modelo; e (iii) a falta de variáveis explicativas importantes no modelo, o que faz com que o coeficiente de uma ou mais variáveis do modelo, em vez de seu próprio efeito, representem também o efeito de variáveis omitidas. Mesmo com estas limitações, a análise dos sinais e identificação da importância relativa das variáveis independentes apresentadas no modelo pode levar a conhecimentos importantes sobre os principais fatores que contribuem para a segurança viária.

O desenvolvimento dum MPA baseia-se na disponibilidade de dados de acidentes, que são aleatórios, não-negativos, discretos e altamente dispersos. Além disso, em alguns casos, os dados disponíveis podem apresentar

correlações temporais ou espaciais, que impõem considerações estatísticas específicas para o desenvolvimento do modelo [4]. Uma análise exaustiva dos dados e a explicitação das questões metodológicas sobre o desenvolvimento de abordagens analíticas para estudar os fatores relacionados a acidentes de trânsito podem ser encontradas no trabalho apresentado por [5]. Além disso, algumas dificuldades de modelação impostas pelas bases de dados com vários valores de zero acidentes deram origem à utilização de diferentes abordagens de modelação estatística para o desenvolvimento de MPA, difíceis justificar no âmbito da engenharia de tráfego [6-7].

Os modelos regularmente utilizados para fornecer estimativas de frequências de acidentes em estradas rurais de duas vias geralmente consideram apenas a estrada e os fatores de trânsito. Alguns modelos abordam um determinado segmento e outros lidam especificamente com curvas e trechos horizontais. Embora os fatores relacionados com o condutor e o veículo afetem a segurança viária, nem sempre se dispõe de dados relativos a estes fatores. A falta de consideração desses fatores (falta de uma ou mais importantes variáveis explicativas) pode levar a problemas relacionados com a interpretação dos coeficientes do MPA [3].

Devido às características intrínsecas dos dados de acidentes, os MPAs são normalmente desenvolvidos utilizando modelos de regressão de Poisson e modelos binomiais negativos. Estes últimos são uma extensão do modelo de Poisson, que assume que o parâmetro de Poisson ( $\lambda$ ) segue uma distribuição de probabilidades gama com parâmetro de forma  $k$ , o valor da média  $\mu$  e o parâmetro de escala  $k/\mu$  [8]. A forma do MPA geralmente consiste no produto das variáveis por um termo exponencial.

Em Portugal, recentemente foram realizados diversos estudos na área de segurança rodoviária, em que foram ajustados modelos de estimativa da frequência de acidentes com o procedimento MLG nas zonas urbanas [9-10].

Para o caso de dados longitudinais, os procedimentos para estimar os parâmetros são derivados a partir do procedimento MLG. Para saídas não-gaussianas, os procedimentos podem ser separados em um modelo de família marginal, que são as equações de estimação generalizadas (EEG), como por exemplo, uma família de efeitos aleatórios, um modelo linear generalizado misto (MLGM), que é uma família mais complexa de modelos [11]. Este último processo não será considerado no presente trabalho.

O procedimento EEG foi desenvolvido por [12] como uma extensão do MLG para a análise de dados longitudinais, quando o foco da análise é a dependência da resposta do modelo das variáveis explicativas. O procedimento EEG pode ser aplicado para as variáveis de resposta tanto gaussianas como não-gaussianas e é um método geral de análise de dados em agrupamento, onde são verdadeiras: (i) observações dentro de um agrupamento podem ser correlacionadas; (ii) observações em grupos separados são independentes; e (iii) a variância é uma função de probabilidade [13].

Quanto à correlação entre as observações em um determinado grupo, o procedimento EEG permite diferentes opções, incluindo a condição de não-correlação. Em [4] é apresentada uma aplicação de EEG para dados agrupados longitudinalmente e espacialmente, e também é analisado o impacto das diferentes opções para modelar a correlação temporal e espacial dentro dos grupos. Eles mostraram que, para a escolha não-correlação (observações independentes) os valores dos coeficientes são iguais aos produzidos pelo MLG, mas os erros padrão são diferentes, conforme o esperado. Os autores também mostraram que as considerações de diferentes correlações produzem diferentes conjuntos de valores dos coeficientes e valores distintos das medidas de qualidade do ajuste para os modelos alternativos correspondentes. Em [14] foram comparados os resultados do EEG com observações independentes a partir dos procedimentos MLG ao longo de uma base de dados e foi observado o seguinte: (i) os coeficientes iguais produzidos pelos dois procedimentos são explicados pelo uso da mesma base de dados e pelo fato da base de dados não ter valores em falta; e (ii) os erros padrão para as estimativas EEG são muito maiores do que os obtidos para as estimativas MLG correspondentes, o que pode implicar que algumas variáveis explicativas deixem de ser significativas quando são considerados os dados longitudinais.

O objetivo principal deste estudo é identificar os fatores que contribuem para a frequência de acidentes relacionados com todos os tipos de colisões ocorridas nos segmentos rodoviários nas estradas nacionais portuguesas de uma faixa de rodagem com uma via em cada sentido. Além disso, este estudo visa analisar o impacto das estruturas das diferentes bases de dados no tempo sobre os fatores identificados. O tipo de correlação entre os dados relativos a cada segmento da estrada também será avaliada através do procedimento de equações de estimação generalizadas (EEG) para a modelação dos dados longitudinais. A importância de estudar

as colisões na região norte de Portugal é justificada pela necessidade de melhorar a segurança viária nesta região servida por estradas nacionais com uma alta concentração de cidades e zonas industriais.

## 2 DESCRIÇÃO DOS DADOS

A base de dados inclui o número de acidentes, volume de tráfego e características geométricas de 88 segmentos rodoviários de 200m de comprimento inseridos nas estradas nacionais de duas vias de sentidos opostos da região Norte de Portugal. Os dados disponíveis compreendem o período de 1999 a 2010. As estradas nacionais e os respetivos trechos selecionados são apresentados no Quadro 1.

Quadro 1: Segmentos rodoviários de 200m considerados no desenvolvimento de MPAs

Estrada Nacional	Trecho	CP	CE <sup>(a)</sup>	NSII <sup>(b)</sup>	NSS (200m)
NR - 14	Braga-Famalicão	22,20	17,60	75	13
NR - 101	Braga-Guimarães	18,00	17,60	66	22
	Braga-Vila Verde	8,80	8,80	39	5
	Guimarães-Felgueiras	20,40	17,80	61	28
NR - 206	Famalicão-Guimarães	19,20	19,20	81	15
	Guimarães-Fafe	6,40	6,40	27	5
Total		95,00	89,40	349	88

<sup>(a)</sup> Considera somente extensões de duas faixas não localizadas em áreas urbanas

<sup>(b)</sup> Segmentos de 200metros de comprimento com presença ou influência de interseções

CP: Comprimento Total (em km); CE: Comprimento Estudado (em km); NSII: Número de segmentos influenciados por interseções; NSS: Número de segmentos selecionados (200m)

Em Portugal, segundo a Autoridade Nacional de Segurança Rodoviária (ANSR), a identificação de pontos negros nas estradas é baseada em segmentos rodoviários de 200m de comprimento [15]. Portanto, os trechos selecionados foram divididos em segmentos de 200m de comprimento, para os quais as características geométricas, o volume de tráfego (expresso em tráfego anual médio diário – TMDA) e o número de colisões entre os anos de 1999 e 2010 foram registados. Alguns segmentos destes trechos não foram incluídos na amostra estudada, já que apresentaram uma ou mais características que não se encaixam no objetivo do presente estudo, nomeadamente: (i) têm mais de duas vias em um dos sentidos de tráfego; (ii) estão inseridos em áreas urbanas; e (iii) contêm interseções ou a influência de interseções (com estradas de acesso a cidades ou com alternativas para o sistema de via expressa nacional).

### 2.1 Dados de acidentes do tipo colisões

Os dados dos acidentes utilizados neste estudo foram fornecidos pela ANSR e abrangem o período de 1999 a 2010. A ANSR mantém uma base de dados com informações recolhidas a partir do Boletim Estatístico de Acidentes de Viação – BEAV, que é preenchido no local da ocorrência do acidente. Assim, a base de dados inicial, que tem 12 observações para cada um dos 88 segmentos, é formada por 1056 observações e com um total de 303 colisões e uma média de 0,29 por observação.

### 2.2 Dados de tráfego

Os dados de tráfego considerados para o desenvolvimento do modelo foi o TMDA por segmento rodoviário e por ano no período considerado (doze anos).

O TMDA do ano de 2011 para os 31 locais cobrindo os segmentos rodoviários selecionados foi estimado com base na metodologia utilizada pela Estradas de Portugal – EP. A EP disponibilizou dados de contagem que foram realizadas em seus postos nos anos de 2000, 2003 e 2005. A empresa também disponibilizou os dados de TMDA por estrada para os anos de 2003, 2004, 2005, 2006, 2007 e 2008 [16-21]. No entanto, como não foi possível

aceder-se à metodologia utilizada na EP, seguiu-se a metodologia proposta por [22], por se considerar adequada as condições deste trabalho.

### 2.3 Características geométricas

As características geométricas consideradas neste estudo foram:

- Largura da via (L)
- Largura média da berma (B)
- Desobstrução lateral (DL)
- Proporção do comprimento da reta no trecho (PCR)
- Sinuosidade em planta (SH):  $SH = \frac{\sum (\cos i)}{CS}$ , onde  $i$  é a curvatura do alinhamento da estrada para curva horizontal  $i$  (em graus) e CS é o comprimento do segmento considerado;
- Proporção do comprimento do trainel no trecho (PCT)
  - Sinuosidade em perfil (SV):  $SV = \frac{\sum (\Delta i)}{CS}$ , onde  $\Delta i$  é a diferença algébrica em percentagem observada na curva vertical  $i$  e CS é o comprimento do segmento considerado;
- Densidade de travessias pedonais (DTP): a qual é definida como o número de travessias de peões por segmento; e
- Densidade de pontos de acesso (DPA): que é calculada como o número de acessos às propriedades particulares (e/ou para estradas secundárias sem saída) por segmento.

Os dados geométricos foram recolhidos no campo e algumas estatísticas relacionadas com os valores observados para os segmentos de 200m de comprimento são apresentadas no Quadro 2, incluindo o TMDA e as colisões dos 88 segmentos. É importante ressaltar que os dados geométricos foram tratados como variáveis explicativas iniciais para a frequência de colisões observada para cada segmento de estrada.

Quadro 2: Estatísticas descritivas do TMDA e das características geométricas dos segmentos

Variável	Unidade	Média	Desvio padrão	Mínimo	Máximo
TMDA	veículos/segmento	12935,66	6326,05	2164,83	32857,43
L	m	3,66	0,18	3,05	4,23
B	m	0,74	0,56	0,10	3,45
DL	m	2,32	1,14	0,60	5,68
PCR		0,62	0,26	0,04	1,00
SH	%/m	300,01	280,46	0,00	1491,50
PCT		0,85	0,31	0,00	1,00
SV	%/m	4,63	8,20	0,00	25,35
DTP	nº/segmento	0,15	0,35	0,00	1,00
DPA	nº/segmento	4,01	2,93	0,00	15,00

As bases de dados utilizadas para o desenvolvimento dos modelos foram compostas por uma base geral, em que foram considerados os dados de colisão, o TMDA e as características geométricas dos segmentos dos 12 anos estudados e com as informações desta base geral. A partir desta base foram formadas bases derivadas considerando diferentes formas de agregação do tempo (2 anos, 3 anos, 4 anos e 6 anos). A base de dados geral apresentou 303 colisões, sendo que o número máximo de colisões por segmento foi 3. Para a base que agrega 2 anos, o valor máximo foi 5, enquanto que para as bases de 3 e 4 anos foi de 8 e para a de 6 anos foi de 10. Em todas as bases o valor mínimo foi nulo. As agregações do tempo foram feitas com o objetivo de reduzir a quantidade segmentos com a ocorrência de zero acidentes, o que pode trazer dificuldades no ajuste do modelo (ver artigo [4]). No entanto, com esta agregação é reduzido o número de observações disponíveis para o ajuste dos modelos.

### 3 DESENVOLVIMENTO DOS MODELOS DE PREVISÃO DE ACIDENTES DO TIPO COLISÃO

Os dados de frequência de acidentes apresentam uma correlação temporal quando os mesmos elementos rodoviários (segmentos ou interseções) são observados ao longo do tempo. Neste caso, cada elemento rodoviário constitui um grupo de análise e as observações repetidas para diferentes anos (dados longitudinais) são os elementos do grupo. Os grupos são independentes uns dos outros.

A forma geral do modelo de estimativa de acidentes para um dado elemento rodoviário (segmento ou interseção), representada por  $Y_{mt}$ , ao longo do período de tempo  $t$ , é a seguinte:

$$E(y_{mt}) = \alpha_t \left\{ \left[ \prod_i (a_{i,mt})^{\gamma_i} \right] e^{\sum_j \beta_j x_{j,mt}} \right\} \quad (1)$$

onde:  $E(Y_{mt})$  = número esperado de acidentes para o elemento  $m$  no período de tempo  $t$ ;

$a_{i,mt}$ ,  $x_{j,mt}$  = variáveis explicativas ( $i$  e  $j$ ) observadas no elemento  $m$  no tempo  $t$ ;

$\alpha_t$  = parâmetro do modelo relacionado com o período de tempo  $t$ , e

$\gamma_i$ ,  $\beta_i$  = parâmetros do modelo constante para todo  $t$ .

O modelo de previsão de acidentes do tipo colisão para os segmentos rodoviários de estradas nacionais de duas vias de sentidos opostos foi desenvolvido utilizando as equações de estimação generalizadas (EEG). No caso do modelo para o número total de colisões ocorridas nos 12 anos, como não existem dados longitudinais, foi utilizado o procedimento de modelo linear generalizado (MLG) básico para obtenção do modelo.

A verificação da presença e do tipo da estrutura de correlação dos dados de acidentes foi realizada testando as estruturas fornecidas pelo procedimento EEG, nomeadamente: (i) independente; (ii) permutável; e (iii) autoregressiva. Estas estruturas permitem a especificação da correlação entre o número de colisões observadas num dado segmento de estrada por diferentes períodos de tempo. As estruturas de correlação foram propostas por [12] e estão razoavelmente bem detalhadas no trabalho de [4].

#### 3.1 Formulação do modelo

A expressão geral de todos os modelos ajustados é dada por:

$$E(y_{mt}) = \alpha \left\{ [Volume_{mt}]^{\gamma} e^{\sum_j \beta_j x_{j,mt}} \right\} \quad (2)$$

a partir da qual a versão MLG é derivada:

$$\ln[E(y_{mt})] = \ln[\alpha] + \gamma \times \ln[Volume_{mt}] + \sum_j \beta_j x_{j,mt} \quad (3)$$

onde  $E(y_{mt})$  = número esperado de acidentes para o segmento  $m$  no período de tempo  $t$ ;

$Volume_{mt}$  = TMDA observado no segmento  $m$  no período de tempo  $t$ ;

$x_{j,mt}$  = valor da variável explicativa  $i$  observada no segmento  $m$  no período de tempo  $t$ ; e

$\alpha$ ,  $\gamma$ ,  $\beta_i$  = parâmetros do modelo a serem estimados.

No processo de modelação seguiu-se a técnica *backward*. Partiu-se do TMDA e de todas as outras variáveis estudadas (apresentadas no Quadro 2), e excluindo-se sucessivamente as menos significativas. O modelo final para cada período de tempo observado, que considerou três estruturas de correlação previstas pelo procedimento

EEG, apresenta apenas as variáveis explicativas que são estatisticamente significativas ao nível de significância de 5%.

A identificação dos fatores que afetam a frequência de acidentes do tipo colisão por agregação temporal, baseou-se no modelo que apresentou melhor ajuste aos dados. Outra consideração importante para a seleção do modelo foi a análise do sinal dos parâmetros do modelo, que deve ser compatível com o expectável do ponto de vista da engenharia de tráfego.

### 3.2 Avaliação do modelo

No procedimento de EEG são comumente utilizados dois elementos para o estudo do ajuste dos modelos, nomeadamente: o teste dos resíduos acumulados (*CURE plot*) e o critério de informação de Akaike (AIC) no EEG, que é chamado de critério de informação quasi-verossimilhança (QIC) [4, 14]). Os autores referem que com o QIC são avaliadas as estruturas de correlação e com o *CURE plot* seleciona-se o melhor modelo, destacando que as técnicas de análise do procedimento MLG não são recomendáveis ao EEG.

O *CURE plot* considera a diferença entre o número de colisões observadas e previstas (os resíduos) como o elemento de base para julgar o ajuste do modelo [3]. Este elemento permite também o exame dos resíduos acumulados considerando a variável de interesse, que é o  $Volume_{,mt}$  (TMDA observado no segmento  $m$  no tempo  $t$ ) para o presente estudo. Para este exame, os resíduos  $N$  ( $N = \text{número de segmentos} \times \text{número de observações por segmento}$ ) são inicialmente classificados por ordem crescente da variável de interesse, numerados de 1, 2, ...,  $z$ , ...,  $N$ . O resíduo acumulado para cada  $z$  é calculado como a soma dos resíduos de 1 a  $z$ . Um bom ajuste significa que os resíduos acumulados no *CURE plot* oscilam em torno do valor zero.

Além disso, o *CURE plot* apresenta duas curvas adicionais que representam limites considerados aceitáveis para os resíduos acumulados, obtidos da seguinte forma:

- a) Calcula-se o quadrado do resíduo para cada um dos resíduos ordenados ( $N$ );
- b) Calcula-se  $\hat{\sigma}^2(z)$ , como a soma destes resíduos quadrados de 1 a  $z$ ;
- c) Calcule-se

$$\sigma^*(z) = \hat{\sigma}(z) \sqrt{1 - \frac{\hat{\sigma}^2(z)}{\hat{\sigma}(N)}} \quad (4)$$

- d) Adiciona-se os limites de  $\pm 2\sigma^*(z)$  ao *CURE plot*.

Em [23] é proposta uma modificação para o AIC no procedimento EEG. A modificação foi desenvolvida para resolver o problema de seleção do modelo no procedimento EEG relacionado com o tipo de correlação entre as observações num determinado grupo (trabalhando a estrutura de correlação). A modificação consiste em utilizar a quasi-verossimilhança construída a partir das equações de estimação (QIC). A estrutura de correlação para ser adotada é a estrutura que produz o menor QIC.

### 3.3 Análises dos modelos desenvolvidos

Os principais resultados dos modelos obtidos no estudo dos segmentos de 200m de comprimento são apresentados no Quadro 3. Pode-se observar que de acordo com o parâmetro QIC, os modelos gerados das bases de dados de 1 ano, 2 anos, 3 anos e 4 anos apresentaram a estrutura de correlação permutável como a que melhor se adapta aos dados longitudinais. Com base nesta estrutura pode-se afirmar que as correlações entre quaisquer duas observações dentro de um grupo são constantes. Isto significa que a estrutura independente, que permite que os dados longitudinais sejam tratados como registos independentes pelo procedimento MLG básico, não é adequada. No entanto, para a base de dados de 6 anos a melhor estrutura de correlação foi a independente.

Quadro 3: Modelos ajustados para segmentos rodoviários de 200m de comprimento

Período de tempo	Parâmetros	Estrutura de Correlação								
		Independente			Permutável			Autorregressiva		
		Coef.	D. P.	p-valor	Coef.	D. P.	p-valor	Coef.	D. P.	p-valor
1 ano	Interceto	4,8025	1,4471	0,0009	4,9809	1,4535	0,0006	4,8008	1,4433	0,0009
	ln (TMDA)	0,3232	0,1522	0,0337	0,3425	0,1531	0,0253	0,3233	0,1518	0,0332
	DPA	0,0851	0,0245	0,0005	0,0845	0,0248	0,0007	0,0851	0,0243	0,0005
	SV	0,0260	0,0105	0,0132	0,0260	0,0104	0,0128	0,0259	0,0105	0,0139
	QIC	1325,7302			1324,6263			1328,4642		
	Número de observações na base de dados = 1056									
2 anos	Interceto	4,3785	1,5122	0,0038	4,5967	1,5141	0,0024	4,3862	1,5023	0,0035
	ln (TMDA)	0,3552	0,1591	0,0256	0,3787	0,1595	0,0176	0,3569	0,1579	0,0238
	DPA	0,0809	0,0255	0,0015	0,0804	0,0258	0,0019	0,0816	0,0252	0,0012
	SV	0,0251	0,0105	0,0162	0,0252	0,0104	0,0156	0,0254	0,0105	0,015
	QIC	913,9351			913,201			926,8214		
	Número de observações na base de dados = 528									
3 anos	Interceto	3,8846	1,5136	0,0103	4,0142	1,5162	0,0081	3,836	1,5103	0,0111
	ln (TMDA)	0,3458	0,1591	0,0298	0,3597	0,1595	0,0241	0,3416	0,1587	0,0313
	DPA	0,0811	0,0253	0,0014	0,0809	0,0255	0,0015	0,0821	0,0257	0,0014
	SV	0,025	0,0105	0,0173	0,025	0,0105	0,0171	0,0251	0,0106	0,0177
	QIC	648,3483			647,8261			660,1411		
	Número de observações na base de dados = 352									
4 anos	Interceto	3,628	1,509	0,0162	3,6949	1,5085	0,0143	3,6471	1,5099	0,0157
	ln (TMDA)	0,3492	0,1587	0,0277	0,3564	0,1586	0,0247	0,3512	0,1586	0,0268
	DPA	0,0811	0,0254	0,0014	0,081	0,0255	0,0015	0,0814	0,0255	0,0014
	SV	0,0247	0,0106	0,0194	0,0247	0,0106	0,0193	0,0249	0,0106	0,0186
	QIC	475,3136			474,9322			476,7039		
	Número de observações na base de dados = 264									
6 anos	Interceto	3,3217	1,5259	0,0295	3,3205	1,5259	0,0296	3,3205	1,5259	0,0296
	ln (TMDA)	0,3597	0,1603	0,0248	0,3596	0,1603	0,0248	0,3596	0,1603	0,0248
	DPA	0,081	0,0255	0,0015	0,081	0,0255	0,0015	0,081	0,0255	0,0015
	SV	0,0247	0,0106	0,0199	0,0247	0,0106	0,0199	0,0247	0,0106	0,0199
	QIC	219,9681			219,9716			219,9716		
	Número de observações na base de dados = 176									
12 anos (MLG)	Interceto	0,7469	0,1572	(<,0001)						
	DPA	0,0834	0,0267	0,0018						
	SV	0,0237	0,0096	0,0132						
	AIC	397,9361								
	Número de observações na base de dados = 88									

Outro ponto importante é que, como esperado, um pequeno número de observações considerado na modelação pode impedir o ajuste aceitável do modelo. No caso do modelo cujas observações anuais foram agrupadas no período de 12 anos, o volume de tráfego deixou de ser uma variável explicativa, o que é contra intuitivo do ponto de vista da engenharia de tráfego.

Considerando apenas os modelos aceitáveis do ponto de vista da engenharia de tráfego e da estatística, verificou-se que os principais fatores que contribuem para a frequência de colisões são o TMDA, a sinuosidade vertical e a densidade de pontos de acesso. Todas estas variáveis têm um impacto positivo sobre a variável dependente (parâmetros com sinal positivo). Por exemplo, os resultados mostram que uma maior densidade de pontos de acesso pode ter um efeito negativo sobre a segurança viária. Para apoiar a seleção final dos modelos considerados aceitáveis, foi desenvolvido o *CURE plot* para cada caso (ver Fig.1).

Na Fig.1 é possível observar que para os modelos de dados agrupados em 2 anos, 3 anos e 4 anos, o *CURE plot* oscila em torno de 0 e não cruza os limites superiores e inferiores, portanto, os modelos podem ser considerados aceitáveis. No entanto, os resíduos acumulados para os modelos de 2 anos e 3 anos apresentam mais pontos perto do limite superior. por esta razão, o modelo de 4 anos pode ser considerado como o melhor modelo.

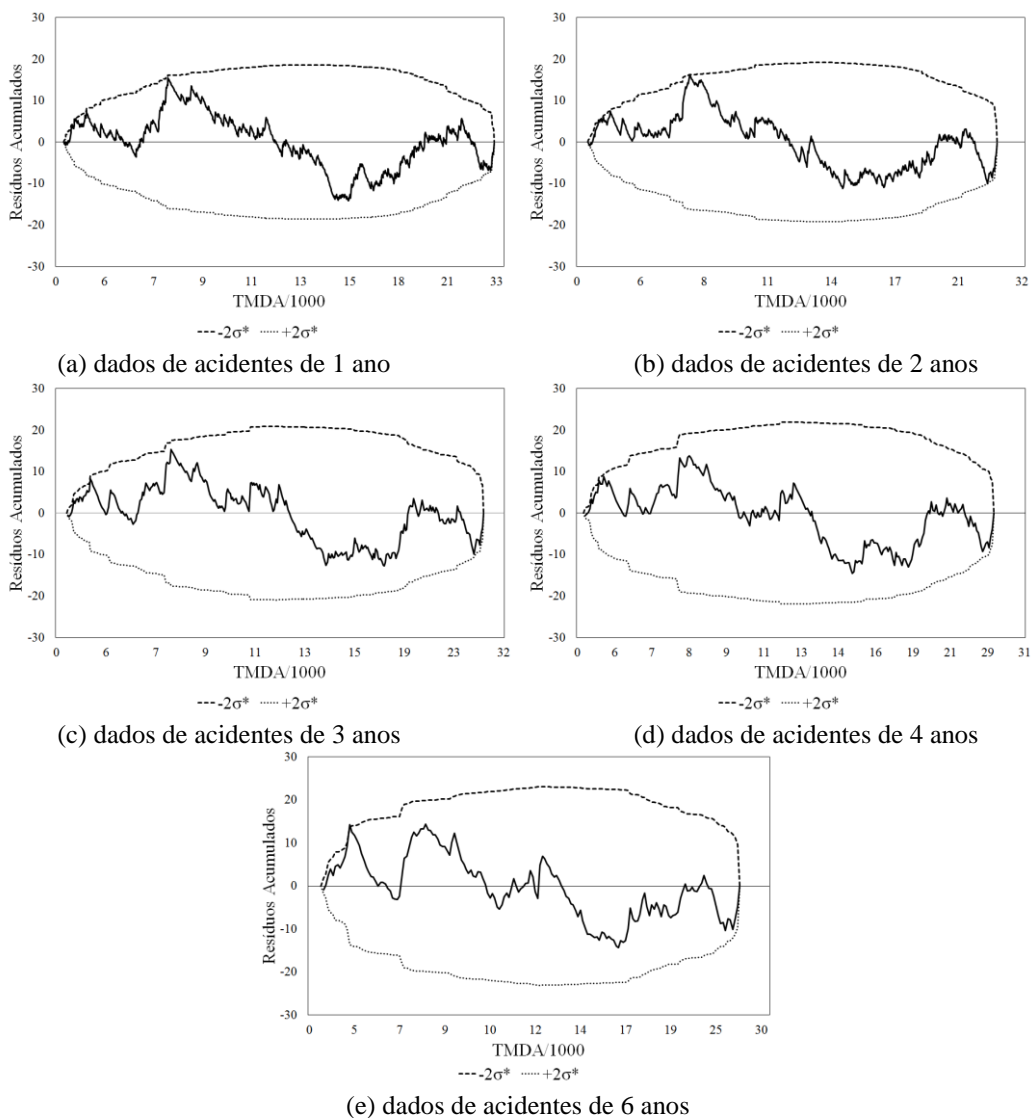


Fig.1. CURE plot para modelos desenvolvidos de segmentos (200m)



#### 4 CONCLUSÕES

O objetivo do presente estudo foi a identificação dos principais fatores que contribuem para ocorrência das colisões rodoviárias para os segmentos de 200m de comprimento de algumas das estradas nacionais de uma faixa de rodagem com uma via em cada sentido, localizadas na região Norte de Portugal. Este trabalho pretende contribuir para a promoção da segurança do sistema rodoviário particularmente do norte de Portugal, que serve muitas cidades e zonas industriais importantes.

A base de dados inicial considerada para este estudo foi formada pela frequência de colisões, o tráfego médio diário anual (TMDA) e as características geométricas de 88 segmentos de 200m de comprimento nos anos de 1999 a 2010. Esta base de dados contém 1056 registos de dados, dos quais 816 têm zero acidentes anuais. Os dados sobre acidentes foram fornecidos pela Autoridade Nacional de Segurança Rodoviária – ANSR, enquanto os dados de volume foram estimados com base em contagens de tráfego disponibilizadas pela Estradas de Portugal – EP. As características geométricas de cada segmento rodoviário foram recolhidas diretamente no campo ou calculadas a partir dos dados recolhidos nas estradas nacionais.

O baixo número de colisões ocorridas no período de estudo dificultou o ajuste dos modelos. Os fatores significativos que integram os modelos são o volume de tráfego (TMDA), a sinuosidade vertical e a densidade de pontos de acesso. Verificou-se que os valores de  $p$  do TMDA são superiores aos dos demais fatores, devido à variação limitada do TMDA nos segmentos estudados por se encontrarem em grande parte seguidos.

A aplicação do procedimento EEG mostrou que os dados observados referentes aos acidentes do tipo colisão, para dois ou mais anos de agregação são efetivamente correlacionados. A estrutura de correlação correspondente foi a permutável, exceto para a base de dados de 6 anos. Sendo assim, é sugerido que o tipo de correlação entre as observações seja investigado antes da modelação.

Finalmente, o estudo mostra que a base de dados com segmentos rodoviários de 200m de comprimento e com os dados agrupados por um período de 4 anos produz um modelo de previsão de colisões aceitável, de acordo com a análise do *CURE plot*.

#### 5 AGRADECIMENTOS

Os autores agradecem à Fundação para a Ciência e a Tecnologia (FCT) pelo apoio dado através da concessão da bolsa de doutoramento SFRH/BD/62458/2009.

#### REFERÊNCIAS

1. Thomas, P., Morris, A., Otte, D. & Breen, J. Real-world Accident Data-Coordinated Methodologies for Data Collection to Improve Vehicle and Road Safety. In: *18<sup>th</sup> International Technical Conference on the Enhanced Safety of Vehicles*, Japan, May 19-22, 2003.
2. Harwood, D.W., Council, F.M., Hauer, E., Hughes, W. E., Vogt, A. Prediction of expected safety performance of rural two-lane highways. *Publication No. FHWA-RD-99-207. FHWA – Federal Highway Administration*, 2000.
3. Hauer, E. Statistical Road Safety Modeling. *Transportation Research Record No. 1897*, pp. 81-87, 2004.
4. Wang, X., Abdel-Aty, M. Temporal and spatial analysis of rear-end crashes at signalized intersections. *Accident Analysis and Prevention* 38, pp. 1137-1150, 2008.
5. Lord, D., Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A*, 44, pp. 291-305, 2010.
6. Lord, D., Washington, S.P., Ivan, J.N., Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 37, pp. 35-46, 2005.
7. Lord, D., Washington, S.P., Ivan, J.N. Further notes on the application of zero-inflated models in highway safety. *Accident Analysis and Prevention* 39, pp. 53-57, 2007.

8. Sawalha, Z., Sayed, T. Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*. Vol. 33, No. 9, pp. 1115-1124, 2006.
9. Ferreira, S. *A Segurança Rodoviária no Processo de Planeamento de Redes de Transporte em Meio Urbano*. Dissertação para obtenção de grau de Doutor, Faculdade de Engenharia, Universidade do Porto, 2010.
10. Gomes, S. *Avaliação da Influência da Infra-estrutura na Segurança Rodoviária em Meio Urbano*, Dissertação para obtenção de grau de Doutor, Faculdade de Ciências e Tecnologia, Universidade de Coimbra. 2010
11. Vangeneugden, T., Molenberghs, G., Verbeke, G., Demétrio, C.G.B. Marginal correlation from an extended random-effects model for repeated and overdispersed counts. *Journal of Applied Statistics*, Vol. 38, No.2, pp. 215-232, 2011.
12. Liang, K.-Y., S. L. Zeger. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, Vol. 73, pp. 13–22, 1986.
13. Halekoh, U., Højsgaard, S., Yan, J. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, Volume 15, Issue 2, pp. 1-11, 2006.
14. Lord, D., Persaud, B.N., Accident Prediction Models With and Without Trend – Application of the Generalized Estimating Equations Procedure. *Transportation Research Record 1717*, pp. 102-108, 2000.
15. Autoridade Nacional de Segurança Rodoviária – ANSR, Sinistralidade Rodoviária. *Observatório de Segurança Rodoviária*, Lisboa, 2009.
16. EP – Estradas de Portugal. Relatório de Tráfego na Rede Rodoviária de Portugal. <http://telematica.estradasdeportugal.pt/pls/alqueva/f?p=105:35:0::NO:RP::> Acessado em: Novembro de 2012. 2003.
17. EP – Estradas de Portugal. Relatório de Tráfego na Rede Rodoviária de Portugal. <http://telematica.estradasdeportugal.pt/pls/alqueva/f?p=105:35:0::NO:RP::> Acessado em: Novembro de 2012. 2004.
18. EP – Estradas de Portugal. Relatório de Tráfego na Rede Rodoviária de Portugal. <http://telematica.estradasdeportugal.pt/pls/alqueva/f?p=105:35:0::NO:RP::> Acessado em: Novembro de 2012. 2005.
19. EP – Estradas de Portugal. Relatório de Tráfego na Rede Rodoviária de Portugal. <http://telematica.estradasdeportugal.pt/pls/alqueva/f?p=105:35:0::NO:RP::> Acessado em: Novembro de 2012. 2006.
20. EP – Estradas de Portugal. Relatório de Tráfego na Rede Rodoviária de Portugal. <http://telematica.estradasdeportugal.pt/pls/alqueva/f?p=105:35:0::NO:RP::> Acessado em: Novembro de 2012. 2007.
21. EP – Estradas de Portugal. Relatório de Tráfego na Rede Rodoviária de Portugal. <http://telematica.estradasdeportugal.pt/pls/alqueva/f?p=105:35:0::NO:RP::> Acessado em: Novembro de 2012. 2008.
22. Silva, A. R.; Araújo, C. E. F.; Rocha, C. H.. Previsão da Demanda de Passageiros no Eixo de Oportunidades Taguatinga-Ceilândia. In: Associação Nacional de Pesquisa e Ensino em Transportes, 2006, Brasília. *Panorama Nacional da Pesquisa em Transportes*. Rio de Janeiro: ANPET, v. 1. p. 467-478, 2006
23. Pan, W. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics* 57, pp. 120-125, 2001.