

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

Jocilene Otilia Costa¹, Maria Alice Prudêncio Jacques², Elisabete Fraga Freitas³, Paulo António Alves Pereira⁴

¹Universidade do Minho, Departamento de Engenharia Civil, 4800-058 Azurém, Guimarães, Portugal

email: jocilene.mt@gmail.com

²Universidade de Brasília, Departamento de Engenharia Civil, Brasília, Brasil

³Universidade do Minho, Departamento de Engenharia Civil, 4800-058 Azurém, Guimarães, Portugal

⁴Universidade 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 modelo de previsão de acidentes (MPA), especificamente colisões. Foram analisadas frequência de colisões, tráfego médio diário anual (TMDA) e características geométricas de 88 segmentos de 200m fora de áreas urbanas, inseridos em estrada nacionais de duas faixas de sentidos opostos. O procedimento de equações de estimação generalizadas – EEG foi aplicado a seis bases de dados distintas. A estrutura de correção predominante nos modelos foi a permutável. Os principais fatores contribuintes para a ocorrência das colisões foram TMDA, sinuosidade vertical e densidade de pontos de acesso. O MPA aceitável identificado foi o de período cumulativo 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].

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 MPA como o verdadeiro efeito de uma mudança incremental em um recurso de estrada associado 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 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 de um 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 específicas considerações estatísticas para o desenvolvimento

do modelo ([4]). Uma análise exaustiva dos dados e as 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 modelagem impostas pelas bases de dados com vários valores de zero acidentes tem causado a utilização de diferentes abordagens de modelação estatística para o desenvolvimento de MPA, que não são fáceis para justificar a partir de uma perspectiva da engenharia de tráfego ([6] e [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 traineis horizontais. Embora os fatores relacionados com o condutor e o veículo serem reconhecidos elementos que afetam a segurança viária, bases de dados relativas a estes fatores nem sempre são facilmente disponíveis. 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, MPAs são normalmente desenvolvidos utilizando modelos de regressão de Poisson e binomial negativo. Este último é uma extensão do modelo de Poisson, que assume que o parâmetro de Poisson (λ) segue uma distribuição de probabilidades gama com parâmetro de forma k , o valor da média μ e o parâmetro de escala k/μ [8]. A forma do MPA geralmente consiste no produto das medidas de exposição multiplicado por um termo exponencial relacionado com as outras variáveis explicativas. O procedimento do modelo linear generalizado (MLG) permite a estimativa de coeficientes para acidentes.

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 ([9]). Este último processo não será considerado no presente trabalho.

O procedimento EEG foi desenvolvido por [10] como uma extensão do MLG para a análise de dados longitudinais, quando o foco principal da análise é a dependência da resposta do modelo para as 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 (iv) a variância é uma função de probabilidade ([11]).

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 parâmetros 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 parâmetros e valores distintos das medidas de qualidade do ajuste para os modelos alternativos correspondentes. Em [12] 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 parâmetros 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 aqueles obtidos para as estimativas MLG correspondentes, o que pode causar algumas variáveis explicativas deixarem de ser significativas quando os dados longitudinais são considerados.

O objetivo principal deste estudo é identificar os fatores que contribuem para a frequência de acidentes relacionados a todos os tipos de colisões ocorridas nos segmentos rodoviários nas estradas nacionais portuguesas de duas vias e sentidos opostos. 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 serão avaliados através do procedimento de equações de estimação generalizadas (EEG) para a modelagem 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 pelo fato de que estas estradas nacionais servem 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 200 m de comprimento inseridos nas estradas nacionais de duas vias de sentidos opostos da região Norte de Portugal. Os dados disponíveis são para os anos de 1999 a 2010. As estradas nacionais e os respetivos trechos selecionados são apresentados no Quadro 1.

Quadro 1: Segmentos rodoviários de 200 m considerados no desenvolvimento de MPA's

Estrada Nacional	Trecho	CP	CE ^(a)	NSII ^(b)	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

^(a) Considera somente extensões de duas faixas não localizadas em áreas urbanas

^(b) Segmentos de 200 metros 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 análise de pontos negros de segurança das estradas é baseada em segmentos rodoviários de 200 m de comprimento ([13]). Portanto, os trechos selecionados foram divididos em segmentos de 200 m 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 a 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) mais de duas vias em um dos sentidos de tráfego; (ii) inseridos em áreas urbanas; e (iii) conter interseções ou 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 da ocorrência de acidentes do tipo colisões para este estudo foram fornecidos pela ANSR e abrangem o período de 1999 a 2010. A ANSR mantém uma base de dados com informações coletadas 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 de 200 m de comprimento, é formada por 1056 observações.

2.2 Dados de tráfego

Os dados de tráfego considerados para o desenvolvimento do modelo foram o tráfego médio diário anual (TMDA) por segmento rodoviário, considerando cada ano dos doze anos da base de dados de colisões (1999 – 2010).

O TMDA foi estimado para cada ano com base nas contagens de tráfego disponibilizadas pelo órgão Estradas de Portugal – EP, as bases de dados ([14], [15], [16], [17], [18] e [19]) para os anos 2003, 2004, 2005, 2006, 2007 e 2008 e as contagens realizadas em 2011 para 31 locais cobrindo os segmentos rodoviários selecionados e os postos de contagem do EP neste estudo. A contagem de 2011 foi então utilizada para definir o TMDA para 2011. Considerando os dados disponíveis para 2003, 2004, 2005, 2006, 2007, 2008 e 2010, o TMDA para todos os anos no período do estudo foi estimado de acordo com o método proposto por [20].

2.3 Características geométricas

Para o objetivo deste estudo, as seguintes características geométricas de cada segmento foram consideradas:

- Largura da faixa (L)
- Largura média da berma (B)
- Desobstrução lateral (DL)
- Proporção do comprimento da reta no trecho (PCR)
- Sinuosidade em planta (SH): $SH = \frac{\sum(|\Omega_i|)}{CS}$, onde Ω_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 Δ_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 coletados no campo e algumas estatísticas relacionadas com os valores observados para os segmentos de 200 m de comprimento são apresentados no Quadro 2. É importante ressaltar que essas características foram tratadas como variáveis explicativas iniciais para a frequência de colisões observada para cada segmento de estrada.

Quadro 2: Estatísticas descritivas das características geométricas dos segmentos

Variável	Unidade	Média	Desvio padrão	Mínimo	Máximo
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	degree/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/segment	0,15	0,35	0,00	1,00
DPA	nº/segment	4,01	2,93	0,00	15,00

3 DESENVOLVIMENTO DOS MODELOS DE PREVISÃO DE ACIDENTES DO TIPO COLISÕES

Os dados de frequência de acidentes apresentam correlação temporal quando os mesmos elementos rodoviários (segmentos ou interseções) são observados ao longo do tempo. Neste caso, cada elemento rodoviário é um grupo de análise e as observações repetidas para diferentes anos (dados longitudinais) são os elementos do grupo. Os dados longitudinais são referentes aos acidentes do tipo colisão (dados do painel). Os grupos são independentes um do outro.

Ao considerar a estimativa de acidentes para um dado elemento rodoviário (segmento ou interseção), representada por Y_{mt} , ao longo do período de tempo t , é considerada a seguinte equação:

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

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 ;
 α_t = parâmetro do modelo relacionado com o período de tempo t , e
 γ_i , β_i = parâmetros do modelo constante para todo t .

O modelo de previsão de acidentes do tipo colisões 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 da base de dados (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 longitudinais de tráfego 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 em um dado segmento de estrada por diferentes períodos de tempo. As estruturas de correlação foram propostas por [10] e são razoavelmente bem detalhadas no trabalho de [4].

3.1 Formulação do modelo

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

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

a partir da qual a versão MLG é derivada:

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

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
 α , γ , β_i = parâmetros do modelo a serem estimados.

No processo de modelação seguiu-se uma eliminação *backward* partindo do TMDA e todas as variáveis selecionadas (apresentadas no Quadro 2), testando as variáveis uma a uma para significância estatística e excluindo as menos significativas. O modelo final para cada período de tempo observado, que considerou três estruturas de correlação prevista 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 colisões definidas pelas combinações de tempo, 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 precisa ser compatível com a expectativa do ponto de vista da engenharia de tráfego.

3.2 Avaliação do modelo

Dois elementos estatísticos foram considerados para a análise da qualidade do ajuste de cada modelo gerado: o teste dos resíduos acumulados (*CURE plot*) e o critério informação de Akaike (AIC) no procedimento EEG, que é chamado de critério de informação quase-verossimilhança (QIC). Com o QIC foram avaliadas as estruturas de correlação e com o CURE plot foi selecionado o melhor modelo global.

O CURE plot considera a diferença entre o número de colisões observadas e previstas (o residual) como o elemento de base para julgar o ajuste do modelo ([3]).

O CURE plot permite o exame dos resíduos acumulados considerando a variável de interesse, que é o *Volume_{mi}* (TMDA observado no segmento *m* no tempo *t*) para o presente estudo. Para este exame, os resíduos *N* (*N* = número de segmentos × número de observações por segmento) são inicialmente classificados em ordem crescente da variável de interesse. Eles são, então, 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 formados por limites aceitáveis para os resíduos acumulados, que são 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

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

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

Em [21] é proposta uma modificação para o AIC no procedimento EEG. A modificação foi desenvolvida para resolver um problema de seleção do modelo no procedimento EEG sobre o tipo de correlação entre as observações em um determinado grupo (trabalhando a estrutura de correlação). A modificação consiste em utilizar a quasi-verossimilhança construída a partir das equações de estimação (QIC), utilizando o modelo de independência de trabalho e qualquer estrutura de correlação geral de trabalho no procedimento EEG. 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 200 m 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 esta estrutura de correlação pode-se afirmar que as correlações entre quaisquer duas observações dentro de um grupo são constantes. Ou seja, a correlação independência, que permite que os dados longitudinais sejam tratados como registos independentes com o auxílio do procedimento MLG básico, não é adequada para a base de dados corrente. No entanto, para a base de dados de 6 anos a melhor estrutura de correlação foi a independente.

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

Período de tempo	Parâmetros	Estrutura de Correlação								
		Independente			Permutável			Autoregressiva		
		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	0,026	0,0105	0,0132	0,026	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 para a calibração do modelo pode evitar a geração de um modelo aceitável. Para o caso dos segmentos de 200 m de comprimento, quando todas as observações anuais são agrupadas (12 anos), o volume de tráfego não é uma variável explicativa, 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 é possível verificar que dentre as variáveis estudadas, 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 (coeficientes com sinal positivo). Por exemplo, os resultados mostram que densidade de pontos de acesso maior pode ter um efeito negativo sobre a segurança viária. Como avaliação final dos modelos considerados aceitáveis, o *CURE plot* para cada caso foi desenvolvido (ver Fig.1).

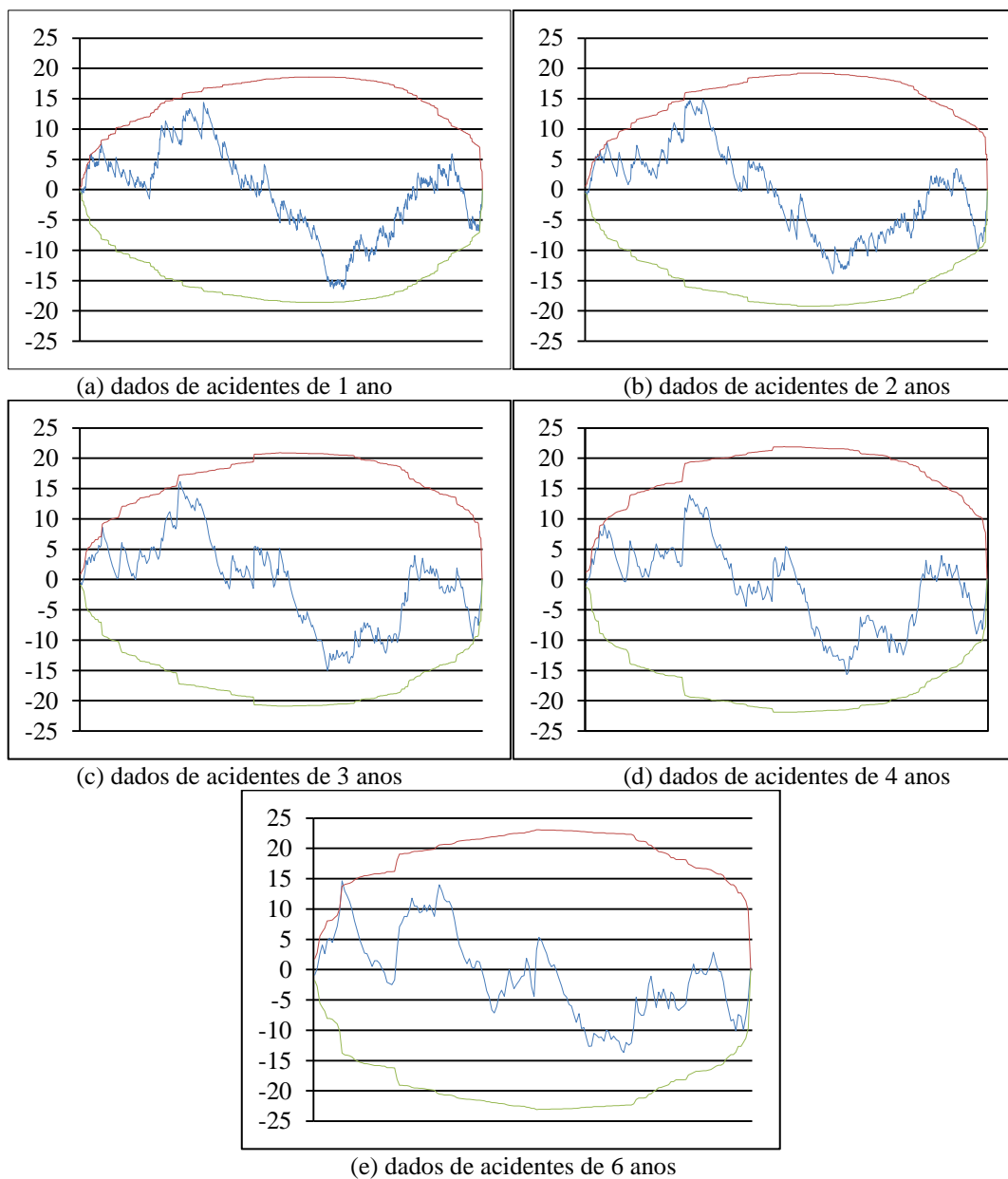


Fig.1. CURE plot para modelos desenvolvidos de segmentos (200 m)

Na Fig.1 é possível observar que para os modelos de dados de 2 anos, 3 anos e 4 anos de acidentes, 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. Devido a isso, o modelo de 4 anos pode ser considerado como o melhor modelo.

4 CONCLUSÕES

O objetivo do presente estudo foi a identificação dos principais fatores que contribuem para ocorrência das colisões rodoviárias para os segmentos de 200 m de comprimento de algumas das estradas nacionais de duas faixas de sentidos opostos localizados na região Norte de Portugal. A importância deste trabalho é contribuir para a promoção da segurança no sistema nacional rodoviário do norte Português, que atende muitas cidades e zonas industriais.

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 200 m 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 pelo órgão Estradas Portuguesas – EP. As características geométricas de cada segmento rodoviário foram recolhidas diretamente no campo ou calculadas com suporte dos dados coletados nas estradas nacionais.

Para as bases de dados estudadas, os resultados mostraram que para tamanhos de amostra superiores a 176 observações, os modelos foram capazes de capturar os mesmos fatores significativos que contribuem para as frequências de colisões observadas. Esses fatores foram o volume de tráfego (TMDA), a sinuosidade vertical e a densidade de pontos de acesso. Como estes fatores resultam em coeficientes positivos nos modelos, que são aceitáveis, quando elas aumentam, é razoável esperar que a frequência de colisões também aumente.

Outro ponto a destacar é que a aplicação do procedimento EEG mostrou que, para os segmentos de 200 m de comprimento, os dados de colisões observados (dois ou mais) 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 na base de dados seja investigado antes do desenvolvimento do modelo de previsão, quando cada localização (segmento ou interseção) é observada ao longo de períodos de tempo diferentes.

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

5 AGRADECIMENTOS

Os autores agradecem à Português 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: *18th 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. 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.
10. Liang, K.-Y., S. L. Zeger. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, Vol. 73, pp. 13–22, 1986.
11. 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.
12. 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.
13. Autoridade Nacional de Segurança Rodoviária – ANSR, Sinistralidade Rodoviária. *Observatório de Segurança Rodoviária*, Lisboa, 2009.
14. 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.
15. 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.
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. 2005.
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. 2006.
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. 2007.
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. 2008.
20. Silva, A. R. da; 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
21. Pan, W. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics* 57, pp. 120-125, 2001.