

## Acidentes de viação em interseções de três ramos: Estudo comparativo das formas funcionais do tráfego e das técnicas de modelação

Jocilene O. Costa<sup>1</sup>, Elisabete F. Freitas<sup>2,†</sup>, Paulo A. A. Pereira<sup>2</sup>

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

Maria Alice P. Jacques<sup>3</sup>

*Universidade de Brasília, Brasília, Brasil*

### RESUMO

Os acidentes de viação apresentam uma enorme carga social e económica, que é imposta à sociedade. Portanto, o conhecimento dos fatores que afetam a probabilidade de ocorrência dum acidente tem sido um domínio de investigação de há muitas décadas, o qual está longe de se esgotar. Assim, este trabalho estuda os acidentes ocorridos em 177 interseções de três ramos inseridas nas estradas nacionais – EN's da região norte de Portugal (2005-2010), sendo apresentada uma contribuição inicial para o desenvolvimento de modelo de previsão de acidentes (MPA), em que considera-se quatro formas funcionais diferentes do tráfego nas aproximações das interseções e difundidas no estudo da sinistralidade. As técnicas de modelação utilizadas foram: (i) Modelos Lineares Generalizados – MLG; (ii) Equações de Estimção Generalizadas – EEG e (iii) Modelos Zero Inflacionados (ZIP e ZINB), com distribuições de Poisson e binomial negativa para a componente do erro aleatório. Obteve-se duas bases de dados, uma geral e uma derivada, com os acidentes e o tráfego das interseções, sendo uma para dados longitudinais (1062 observações) e outra para dados agregados. Como resultado apenas as EEG e os MLG forneceram modelos válidos, tendo sido identificada a melhor forma funcional em ambos os casos.

### 1. INTRODUÇÃO

A nível mundial, quando se avalia os acidentes de viação conclui-se que a ocorrência destes eventos depende de múltiplas causas, as quais podem ser imediatas ou indiretas (Lord & Mannering, 2010). As causas imediatas limitam-se ao contexto do acidente e referem-se: (i) ao condutor; (ii) à estrada e seu ambiente viário; e (iii) ao veículo. As causas indiretas são aquelas de ordem ambiental, tecnológica ou institucional, que criam situações propícias aos

---

<sup>1</sup> Doutoranda

<sup>2</sup> Professor (a)

<sup>†</sup> Autora correspondente ([efreitas@civil.uminho.pt](mailto:efreitas@civil.uminho.pt))

<sup>3</sup> Investigadora colaboradora

acidentes ou agravam as suas consequências (Vogt & Bared, 1997; Lord & Persaud, 2000; Kim & Washington, 2006; Gomes, 2010).

No estudo da sinistralidade, o ajuste de equações matemáticas, através de métodos estatísticos, permite relacionar os dados sobre acidentes ou vítimas com uma série de variáveis explicativas (fatores contribuintes), dando origem aos modelos de previsão de acidentes. A principal vantagem deste tipo de modelos reside na forma expedita como se pode avaliar o efeito sobre o indicador de segurança associado às alterações de determinadas variáveis explicativas, ainda que a necessidade de informação detalhada seja elevada (Hauer, 1997).

A forma funcional dos modelos estabelece a relação entre a variável dependente e as variáveis explicativas e é uma parte crítica do processo da modelação dos acidentes de viação. A maioria dos modelos de dados de contagem assume que as variáveis explicativas influenciam as variáveis dependentes de alguma forma linear (Lord & Mannering, 2010). No entanto, na definição da forma funcional dum modelo de previsão de acidentes, importa destacar a contribuição da variável tráfego, visto que alguns investigadores constataram que o tráfego tem um efeito positivo decrescente na ocorrência de acidentes e defendem que o risco diminui por unidade de exposição, ou seja, é menor à medida que o tráfego aumenta (Hauer, 1997). Há também divergência de opinião quanto ao modo como o tráfego deve ser considerado na forma funcional dos modelos de previsão. Cardoso et al. (2005) afirmam que na escolha da forma funcional do modelo de previsão é preciso ter em conta que simplificações como a utilização de valores médios na estimação das relações entre o tráfego e os acidentes, em vez do tráfego real no instante da ocorrência do acidente, podem enviesar os resultados.

Diante do exposto, o objetivo deste trabalho é realizar uma investigação exploratória da aplicabilidade das formas funcionais que envolvem somente o tráfego, e que são amplamente difundidas no estudo da sinistralidade, para o caso das rodovias portuguesas. Esta análise será feita através do desenvolvimento de modelos de previsão dos acidentes, com vítimas, ocorridos nas interseções de três ramos localizadas nas estradas nacionais – EN's da região do Minho do Norte de Portugal (EN-14, EN-101 e EN-206), incorporando a variação temporal através do uso de técnicas de modelação diferentes, nomeadamente: (i) Modelos Lineares Generalizados – MLG; (ii) Equações de Estimação Generalizadas – EEG e (iii) Modelos Zero Inflacionados. A base de dados é formada pelos dados de tráfego e dos acidentes de 177 interseções. No desenvolvimento destes modelos serão estudadas quatro formas funcionais diferentes apresentadas por Lord & Park (2008). Relativamente aos fatores de análise da qualidade do ajuste dos modelos, serão utilizados a análise de resíduos (*CURE Plot*) e o  $R^2$  marginal. O principal objetivo da análise de resíduos é identificar casos para os quais as estimativas do modelo se distanciem muito dos valores observados.

## 2. METODOLOGIA

Na modelação estatística os modelos representam sempre de forma simplificada a realidade, a qual integra uma série de fatores considerados importantes para a sua descrição. As técnicas de modelação são um meio para a obtenção de modelos que reproduzam o mais fielmente possível a realidade, sendo, por isso mesmo, muito importante a adoção de técnicas adequadas à natureza dos dados a tratar. Desta forma, para o desenvolvimento do presente trabalho utilizou-se os Modelos Lineares Generalizados – MLG; as Equações de Estimação Generalizadas – EEG e os Modelos Zero Inflacionados (Modelo de Poisson Zero-Inflacionado – ZIP e Modelo Binomial Negativo Zero-Inflacionado – ZINB), com as suposições feitas para

o uso destas técnicas, sendo determinados fatores de avaliação da qualidade do ajuste para cada modelo.

Relativamente à dimensão temporal, Wang & Abdel-Aty (2006) destacam que os dados de frequência dos acidentes de viação apresentam correlação temporal quando os mesmos elementos viários (segmentos ou interseções) são observados ao longo do tempo. Neste caso, cada elemento viário 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 de viação (dados do painel). Os grupos são independentes entre si. No presente trabalho, as EEG foram utilizadas para a modelação dos dados onde essa situação foi verificada.

Portanto, a estimativa dos acidentes de viação para um dado elemento viário, representada por  $y_{mt}$ , ao longo do período de tempo  $t$ , é traduzida pela seguinte equação geral:

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

onde  $y_{mt}$  é o número esperado de acidentes para o elemento viário  $m$  no período de tempo  $t$ ;  $a_{i,mt}$ ,  $x_{j,mt}$  são as variáveis explicativas ( $i$  e  $j$ ) observadas no elemento viário  $m$  no tempo  $t$ ; e  $\alpha, \gamma_i, \beta_j$  são os parâmetros do modelo a serem estimados.

No uso das EEG, a identificação do tipo de estrutura de correlação dos dados longitudinais dos acidentes foi realizada testando as estruturas de correlação de trabalho desenvolvidas por Liang & Zeger (1986), nomeadamente: (i) independente; (ii) permutável; e (iii) autorregressiva. Estas estruturas permitem a especificação da correlação entre o número de acidentes observado numa dada interseção de três ramos por diferentes períodos de tempo. Assim, no ajuste dos modelos de previsão foi utilizada a seguinte expressão:

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

onde  $y_{mt}$  é o número esperado de acidentes para a interseção  $m$  no período de tempo  $t$ ;  $\text{Tráfego}_{mt}$  é função de variáveis relacionadas ao tráfego na interseção de três ramos  $m$  no período de tempo  $t$ ;  $x_{j,mt}$  é a variável explicativa  $i$  observada na interseção de três ramos  $m$  no período de tempo  $t$ ; e  $\alpha, \gamma, \beta_i$  são os parâmetros do modelo a serem estimados.

No entanto, para atingir o objetivo do presente estudo, utilizou-se a Equação 2 considerando somente a contribuição do tráfego, levando em conta diferentes configurações desta variável na direção principal e na direção secundária das interseções. Com isso, foram testadas quatro formas funcionais (Equações de 3 a 6) sugeridas por Lord & Park (2008).

$$\ln y_{mt} = \beta_0 + \beta_1 \times \ln(F_{1,mt} + F_{2,mt}) \quad (3)$$

$$\ln y_{mt} = \beta_0 + \beta_1 \times \ln(F_{1,mt}) + \beta_2 \times \ln(F_{2,mt}) \quad (4)$$

$$\ln(y_{mt}) = \beta_0 + \beta_1 \times \ln(F_{1,mt} \times F_{2,mt}) \quad (5)$$

$$\ln y_{mt} = \beta_0 + \beta_1 \times \ln(F_{1,mt} + F_{2,mt}) + \beta_2 \times \ln\left(\frac{F_{2,mt}}{F_{1,mt}}\right) \quad (6)$$

onde  $y_{mt}$  é o número esperado de acidentes na interseção  $m$  no ano  $t$ ;  $F_1$  é o TMDA na direção principal;  $F_2$  é o TMDA na direção secundária; e  $\beta_0, \beta_1, \beta_2$  são os parâmetros a serem estimados.

Assim, para os dados longitudinais foram utilizadas as equações de estimação generalizadas – EEG, enquanto no ajuste dos modelos para os acidentes ocorridos no período do estudo, seis anos (2005, 2006, 2007, 2008, 2009 e 2010), onde não existem dados longitudinais, foram utilizados os MLG, o ZIP e o ZINB.

Lord *et al.* (2005) destacam que as técnicas de modelação dos modelos ZIP e ZINB devem ser utilizadas numa base de dados com excesso de zeros e quando é possível afirmar que os elementos viários (interseções ou segmentos viários) que apresentam observações nulas podem ser tratados como elementos viários potencialmente seguros. Nesta modelação dos zeros inflacionados com ZIP e ZINB, portanto, assume-se que a ocorrência de zero acidentes em um elemento pode decorrer tanto do fato do elemento ser potencialmente seguro quanto da probabilidade de se ter zero acidentes em um elemento que não é potencialmente seguro. Isto é, o processo de ocorrência de acidentes é formado por dois estados que devem ser modelados separadamente (ver mais detalhe em Carson & Mannering, 2001). Portanto, no ajuste dos modelos ZIP e ZINB com as Equações de 3 a 6, tentou-se verificar a explicação das observações nulas de acidentes com a mesma forma funcional dos dados de contagem, como sugerem Carson & Mannering (2001) e Kumara & Chin (2003).

Para a escolha de um modelo de previsão que possa ser reconhecido como útil para a previsão de acidentes em interseções de três ramos de EN's portuguesas, aliou-se a validade do modelo sob o ponto de vista da engenharia de tráfego (um modelo em que o TMDA apresente contribuição positiva para segurança viária certamente não apresenta essa reconhecida utilidade) com o nível de significância desejável e possível, dadas as diversas limitações das bases de dados disponíveis. Para o presente trabalho foi considerado um nível de confiança no ajuste dos modelos de 95% (nível de significância igual a 5%).

O modelo de previsão, para cada forma funcional testada e para o nível de significância adotado, foi obtido com as três estruturas de correlação previstas para as EEG, com os MLG e com os Modelos Zero Inflacionados (ZIP e ZINB). Para a seleção dos fatores fez-se a análise do sinal dos parâmetros do modelo que precisa ser compatível com a expectativa do ponto de vista da engenharia de tráfego. Outra consideração importante é relativamente aos MLG, o ZIP e o ZINB, que apresentam o Critério de Informação de Akaike (*Akaike Information Criterion* – AIC) como uma ferramenta de comparação entre modelos originados de uma mesma base de dados, distinguindo os modelos que melhor explicam os dados dum fenómeno com o menor número de parâmetros, penalizando os modelos com elevado número de parâmetros. Akaike (1974) destaca que o modelo com o menor AIC é considerado o melhor, entre os modelos testados. No entanto, no presente estudo o AIC não foi efetivamente considerado dado que o foco não era a busca de um modelo parcimonioso, mas sim o teste da aplicabilidade de formas funcionais alternativas envolvendo parâmetros apenas ligados a variáveis que representam a exposição ao risco de acidentes.

Para a avaliação do ajuste dos modelos foram considerados três elementos estatísticos, nomeadamente: (i) critério de informação quasiverossimilhança – QIC; (ii) teste dos resíduos acumulados (*CURE Plot*); e (iii)  $R^2$  marginal.

## 2.1. Critério de informação quasiverossimilhança – QIC

Pan (2001) propôs um método de seleção de estrutura de correlação de trabalho para as EEG, mas com a suposição de que as medidas dos grupos possam ser correlacionadas. Este critério foi denominado critério de quasiverossimilhança sob o modelo de independência (*Quasi-likelihood under the Independence model Criterion* – QIC) e para a matriz de correlação  $R$  é definido da seguinte forma:

$$QIC(R) = -2Q(\hat{\beta}(R); I, D) + 2\text{traço}(\hat{\Omega}_I, \hat{V}_R) \quad (8)$$

onde  $Q$  é a quasiverossimilhança;  $\hat{\beta}(R)$  é o vetor de estimadores de quasiverossimilhança sob o modelo candidato com matriz de correlação  $R$ ;  $I$  é a matriz identidade;  $D$  são os dados observados;  $\hat{\Omega}_t$  é o inverso da estimativa da covariância do modelo baseado no pressuposto da matriz de correlação de trabalho independente; e  $\hat{V}(R)$  é o estimador de covariâncias robusto obtido através do modelo contendo a matriz de correlação  $R$ .

Segundo Pan (2001), o QIC é calculado através da comparação dum modelo com uma determinada estrutura de correlação de trabalho com aquele ajustado com a estrutura independente. Posteriormente, os valores de QIC obtidos para cada modelo são comparados. A melhor estrutura de correlação de trabalho dos dados será aquela com menor valor de QIC.

## 2.2. Análise dos resíduos acumulados

Um dos fatores de análise da qualidade do ajuste mais utilizada nos modelos de regressão é a análise de resíduos. Um resíduo pode ser definido como a distância entre o valor estimado e o seu correspondente valor observado da variável dependente (Cox & Snell, 1968). O principal objetivo da análise de resíduos é identificar casos para os quais as estimativas do modelo se distanciem muito dos valores observados e, também, verificar a aleatoriedade dos resíduos. Em estudos longitudinais é recomendado analisar o gráfico de dispersão dos resíduos *versus* período de tempo.

No estudo da modelação da sinistralidade recomenda-se o uso do *CURE Plot* (Hauer, 2004). Este elemento estatístico de análise considera a diferença entre o valor observado e o previsto (o resíduo) como a componente de base para julgar o ajuste do modelo e permite o exame dos resíduos acumulados considerando uma variável de interesse. Para este exame, os  $N$  resíduos são inicialmente classificados por 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 de zero. Além disso, o *CURE Plot* apresenta duas curvas adicionais formadas por limites aceitáveis para os resíduos acumulados.

## 2.3. $R^2$ marginal

Nos modelos ajustados com as EEG, os resíduos não são independentes, porque não há independência entre as observações. Logo, os métodos baseados na verossimilhança e as medidas de ajuste dos modelos de regressão linear precisam ser adaptados. Zheng (2000) introduziu uma extensão do  $R^2$  para modelação com as EEG com variável dependente contínua, binária, ou inteira, denominado  $R^2$  marginal. É necessário calcular os valores previstos pelo modelo para obter o valor do  $R^2$  marginal. Estes valores são comparados com os valores observados e divididos pela soma do quadrado dos desvios das observações em relação à média da variável dependente. O  $R^2$  marginal é dado por:

$$R^2_{marginal} = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^n (y_{it} - \hat{y}_{it})^2}{\sum_{t=1}^T \sum_{i=1}^n (y_{it} - \bar{y})^2} \quad (9)$$

onde  $y_{it}$  é o número de acidentes observado para a interseção  $i$  no período de tempo  $t$ ;  $\hat{y}_{it}$  é o número de acidentes esperado para a interseção  $i$  no período de tempo  $t$ ;  $\bar{y}$  é média marginal,

dada por:  $\bar{y} = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (y_{it})$ ; n é o número de interseções; e T é Número de períodos de tempo considerados para cada interseção.

A interpretação do  $R^2$  marginal é semelhante à do  $R^2$ ; indica quanto a variância da variável resposta é explicada pela variabilidade do modelo ajustado (Hardin & Hilbe, 2003). Além disso, o  $R^2$  marginal tem as mesmas propriedades que o  $R^2$  tem para os modelos de regressão, com a exceção de poder assumir valores inferiores a zero, quando o modelo ajustado produz estimativas piores do que o modelo somente com o intercepto.

Diante do exposto, na modelação dos acidentes de viação das interseções de três ramos foi utilizada uma sequência de estádios na aplicação das diferentes técnicas de modelação, sendo utilizado o *software* SAS (versão 9.3®) no ajuste destes modelos. Estes estádios são:

- ✓ Ajuste dos modelos com as diferentes técnicas de modelação selecionadas para as quatro formas funcionais envolvendo as variáveis de tráfego;
- ✓ Identificação dos modelos aceites em relação à contribuição do tráfego e do nível de confiança;
- ✓ Identificação do melhor modelo ajustado, com as EEG, em função da estrutura de correlação (permutável, independente ou autorregressiva) para cada forma funcional, definida com a utilização do QIC;
- ✓ Definição da melhor da forma funcional em função da avaliação do *CURE Plot* e do  $R^2$  marginal dos modelos que foram aceites nos estádios anteriores.

### 3. DESCRIÇÃO DOS DADOS

A seleção dos trechos das estradas nacionais – EN's deu-se pela facilidade logística, como também pela necessidade de melhorar a segurança viária na região do Minho servida por EN's que ligam uma alta concentração de cidades e zonas industriais. Os trechos rodoviários eleitos foram: (i) Braga-Famalicão (EN-14); (ii) Braga-Vila Verde (EN-101); (iii) Braga-Guimarães (EN-101); (iv) Guimarães-Felgueiras (EN-101); (v) Famalicão-Guimarães (EN-206); e (vi) Guimarães-Fafe (EN-206), totalizando 95,2 quilómetros de estrada.

Relativamente às interseções, destaca-se que o tipo de interseção selecionado para o presente estudo foi a de três ramos, também chamada de entroncamento ou interseção em T, por ser o tipo de interseção com maior incidência na rede rodoviária do estudo. Portanto, os critérios estabelecidos para identificação destas interseções de três ramos são apresentados a seguir, assim como, a distribuição nos trechos das 177 interseções de três ramos (Tabela 1).

1º Critério: distância mínima entre o centro das interseções de 76 metros;

2º Critério: interseções inseridas em EN's de uma faixa de rodagem com duas vias, uma em cada sentido;

3º Critério: interseções em que a direção secundária apresentava uma faixa de rodagem com duas vias, uma em cada sentido;

4º Critério: interseções inseridas em EN's, com e sem canalização de tráfego;

5º Critério: interseções com e sem canalização de tráfego na direção secundária.

Tabela 1 – Distribuição das interseções nas EN's estudadas

Trecho	Estrada	Interseções
Braga-Famalicão	EN-14	35
Braga-Guimarães	EN-101	30
Braga-Vila Verde	EN-101	17
Famalicão-Guimarães	EN-206	39
Guimarães-Felgueiras	EN-101	36
Guimarães-Fafe	EN-206	20
Total		177

### 3.1. Dados de acidentes

Os dados dos acidentes de viação utilizados neste estudo foram fornecidos pela Autoridade Nacional de Segurança Rodoviária – ANSR e abrange o período de 2005 a 2010, sendo aproveitados somente os acidentes com vítimas que ocorreram nas 177 interseções referidas na Tabela 1. 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 de viação. Nesta base é possível identificar os acidentes desagregados por tipo (colisão, despistes e atropelamentos). A variável resposta utilizada neste estudo são os acidentes com exclusão dos atropelamentos, sendo identificados 252 no período de estudo que apresentaram estatísticas descritivas presentes na Tabela 2. Na contabilização dos acidentes ocorridos em cada interseção considerara-se uma área de influência de 76 metros antes e depois do centro da interseção na direção principal (FHWA & TFHRC, 2003). Os acidentes ocorridos na direção secundária das interseções não foram considerados por indisponibilidade de dados das entidades gestoras.

Tabela 2 – Estatísticas descritivas dos acidentes de viação considerados no estudo

Dados	Nº de observações	Média	D. P.	Min	Max
Desagregados	1024	0,24	0,54	0	5
Agregados	177	1,42	1,48	0	8

### 3.2. Dados de tráfego

O tráfego médio diário anual – TMDA nas aproximações das interseções é um indicador de exposição ao risco de acidentes essencial e por isso obrigatoriamente presente nos modelos de previsão. Portanto, foram obtidas as séries históricas do TMDA das interseções (direção principal e secundária) no período de 2005 a 2010, totalizando 1062 observações. Os dados de tráfego foram estimados com base em contagens de tráfego disponibilizadas pela empresa Estradas Portuguesas – EP.

No processo da modelação dos acidentes foram definidas cinco variáveis de tráfego, que compõem as formas funcionais testadas no presente trabalho. Estas cinco variáveis consideradas e a simbologia de cada uma são apresentadas na Tabela 3 e as estatísticas descritivas do tráfego na direção principal e na direção secundária, são apresentadas na Tabela 4, visto que as formas funcionais testadas serão compostas por diferentes configurações do tráfego na direção principal e secundária de cada interseção.

Tabela 3 – Variáveis de tráfego utilizadas na modelação dos elementos viários

Abreviatura	Nomenclatura
TMDAP	Tráfego médio diário anual da direção principal
TMDAS	Tráfego médio diário anual na direção secundária
TMDAT	Soma do tráfego médio diário anual da direção principal e da direção secundária
TMDAPMS	Multiplicação do tráfego médio diário anual da direção principal pelo da direção secundária
TMDASDP	Divisão do tráfego médio diário anual da direção secundária pelo da direção principal

Tabela 4 – Estatísticas descritivas do tráfego nas 177 interseções

Variável	Média	D. P.	Mín	Máx
TMDAP	17135	8244	3160	42493
TMDAS	924	1419	15	8353

#### 4. MODELOS AJUSTADOS COM AS DIFERENTES TÉCNICAS DE MODELAÇÃO

Nesta seção são apresentados os modelos de previsão ajustados com os dados das 177 interseções de três ramos inseridas nas EN's 14, 101 e 206, sendo o foco da modelação o estudo de quatro formas funcionais diferentes para o tráfego da direção principal e da direção secundária, assim como, o uso de diferentes técnicas de modelação na definição da melhor forma funcional a ser utilizada numa modelação futura com demais variáveis. Estas formas funcionais são descritas na Tabela 5.

Tabela 5 – Formas funcionais testadas nos modelos de previsão

Nº	Formas funcionais	Observações
1	$\ln y_{mt} = \beta_0 + \beta_1 \times \ln(TMDAT_{mt})$	$TMDAT = (TMDAP + TMDAS)$
2	$\ln y_{mt} = \beta_0 + \beta_1 \times \ln(TMDAP_{mt}) + \beta_2 \times \ln(TMDAS_{mt})$	-
3	$\ln(y_{mt}) = \beta_0 + \beta_1 \times \ln(TMDAPMS_{mt})$	$TMDAPMS = (TMDAP \times TMDAS)$
4	$\ln y_{mt} = \beta_0 + \beta_1 \times \ln(TMDAT_{mt}) + \beta_2 \times \ln(TMDASDP_{mt})$	$TMDASDP = \left( \frac{TMDAS}{TMDAP} \right)$

$y_{mt}$ : Número esperado de acidentes no elemento m no ano t; TMDAP: TMDA na direção principal; TMDAS: TMDA na direção secundária.

##### 4.1. Modelos de previsão dos acidentes para interseções de três ramos

Nos modelos de previsão dos acidentes ajustados para as interseções de três ramos, foi considerado o período de estudo os anos de 2005 a 2010. O modelo de previsão foi ajustado para cada forma funcional com nível de significância de 5%. Para este ajuste, foram utilizadas as técnicas de modelação EEG, MLG, ZIP e ZINB, onde os dados anuais de acidentes e de tráfego descritos na Seção 3 foram considerados. A distribuição do erro dos modelos foi a binomial negativa no ajuste com MLG, EEG e ZINB, enquanto no ZIP foi utilizada a distribuição de Poisson.

Os resultados dos modelos de previsão dos acidentes são descritos na Tabela 6. No entanto, para definição da melhor forma funcional do modelo de previsão, analisam-se os elementos estatísticos e os *CURE Plot* dos modelos ajustados aceites. Nesta tabela pode ser observado que somente os modelos ajustados com MLG e EEG podem ser aceites para explicação dos acidentes, sendo que o mesmo não ocorreu com os modelos zero inflacionados (ZIP e ZINB). Nesses últimos modelos, observa-se que os coeficientes dos modelos, os quais referem-se às interseções onde a não ocorrência de acidentes justifica-se por uma condição potencialmente segura, não são estatisticamente significativos.

Nos modelos ajustados com EEG é possível observar que a melhor estrutura de correlação dos dados para as formas funcionais estudadas é a permutável, definida com a melhor estrutura de correlação de cada forma funcional (Tabela 6). Relativamente aos resíduos dos modelos ajustados com EEG, na Figura 1 verifica-se que todas as curvas dos resíduos acumulados dos *CURE Plot* oscilam em torno de zero, mas somente os modelos que utilizam as formas funcionais 2 (Figura 1.b) e 3 (Figura 1.c) não extrapolam os limites, inferior e superior. No entanto, o modelo ajustado com a forma funcional 2 apresentou os resíduos acumulados mais bem distribuídos em torno de zero, que associado ao fato de este modelo ter também apresentado um melhor  $R^2$  marginal, em torno de 3,3%, leva a concluir que para o período de estudo de seis anos esta forma funcional é a melhor.

Alusivamente aos modelos ajustados com MLG observa-se que os resíduos acumulados dos modelos, ilustrados na Figura 2 através do *CURE Plot*, oscilam em torno de zero, porém os resíduos acumulados da Figura 2.b apresentam-se melhor distribuídos em torno do zero. Portanto, conclui-se que o modelo ajustado com a forma funcional 2, que apresenta os resíduos acumulados mais bem distribuídos em torno de zero e também um  $R^2$



marginal melhor que a maioria, cerca de 14,5%, é o mais indicado para o período de estudo de seis anos do presente trabalho.

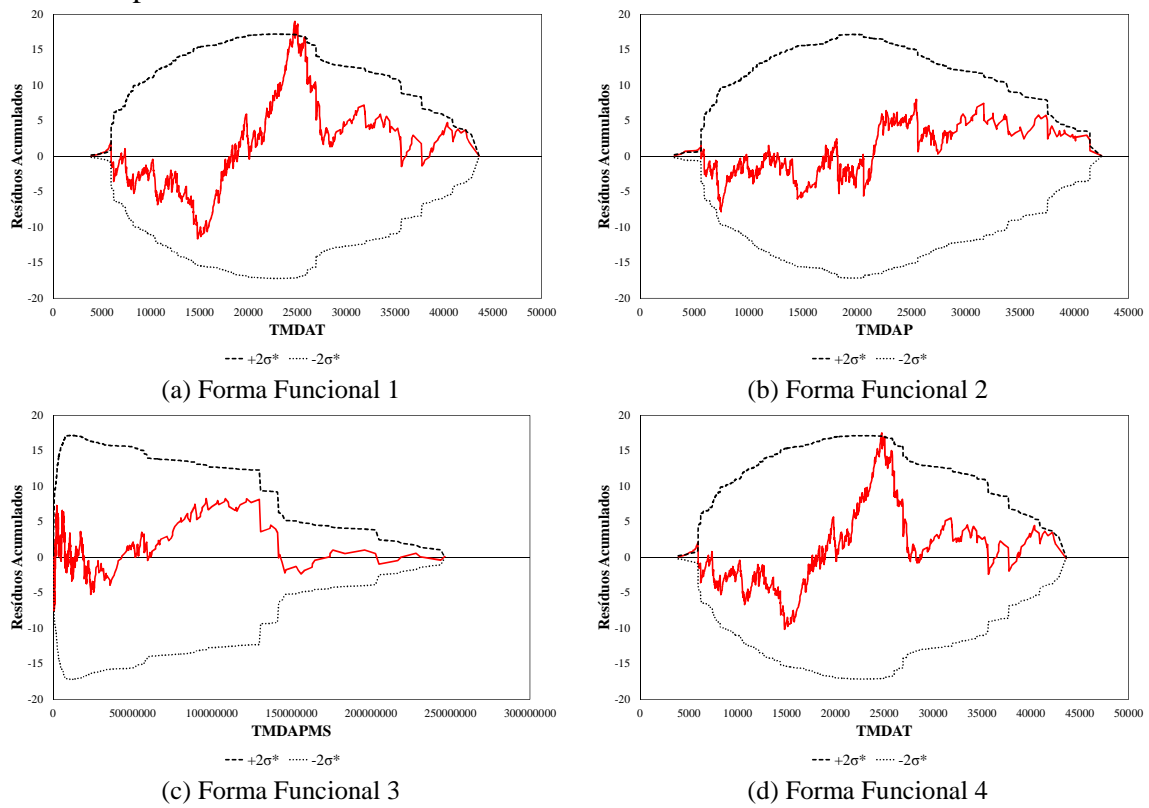


Figura 1 – CURE Plot dos modelos ajustados com EEG com a estrutura permutável

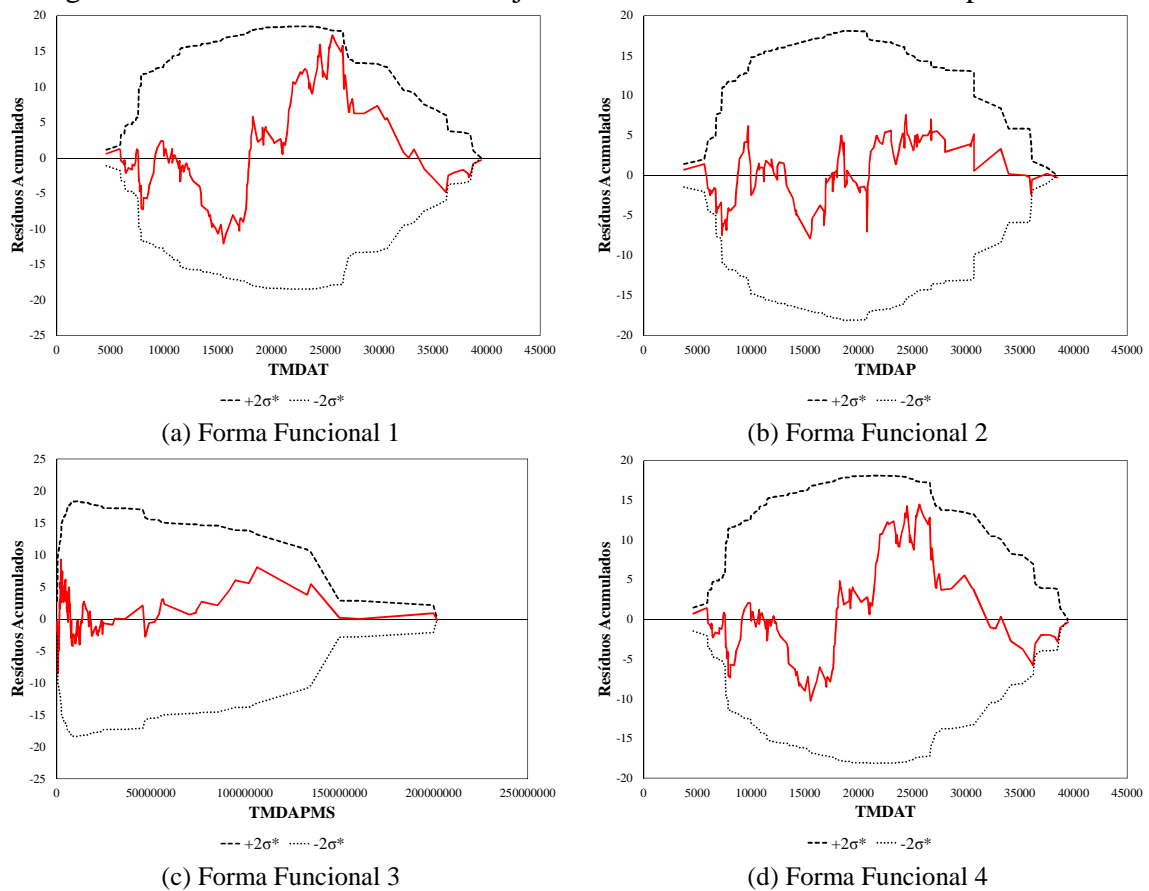


Figura 2 – CURE Plot dos modelos ajustados com MLG

Tabela 6 – Modelos de previsão ajustados com as 4 formas funcionais – FF diferentes e diferentes técnicas de modelação

FF	Parâmetros	EEG – Estrutura de Correlação									MLG			Zero inflacionados					
		Independente			Permutável			Autorregressiva			Binomial Negativa			Poisson			Binomial Negativa		
		Coef.	D. P.	p	Coef.	D. P.	p	Coef.	D. P.	p	Coef.	D. P.	p	Coef.	D. P.	p	Coef.	D. P.	p
1	Interceto	-8,47	1,57	<0,0001	-8,53	1,57	<0,0001	-8,47	1,57	<0,0001	-6,39	1,48	<,0001	-5,33	1,56	0,0006	-5,57	1,67	0,0008
	ln (TMDAT)	0,72	0,16	<0,0001	0,73	0,16	<0,0001	0,72	0,16	<0,0001	0,69	0,15	<,0001	0,60	0,16	0,0001	0,62	0,17	0,0002
	Interceto*	-	-	-	-	-	-	-	-	-	-	-	-	3,15	6,10	0,6054	2,78	7,42	0,7075
	ln (TMDAT)*	-	-	-	-	-	-	-	-	-	-	-	-	-0,49	0,62	0,4288	-0,48	0,75	0,5211
	QIC ou AIC	1173,49			1172,78			1173,29			555,62			556,45			557,69		
	R <sup>2</sup> marginal	0,0282			0,0282			0,0282			0,1179			-			-		
	2	Interceto	-7,78	1,52	<0,0001	-7,83	1,53	<0,0001	-7,78	1,53	<0,0001	-5,74	1,45	<,0001	-4,37	1,61	0,0067	-4,58	1,73
ln (TMDAP)		0,55	0,16	0,0008	0,55	0,16	0,0007	0,55	0,16	0,0008	0,53	0,15	0,0006	0,40	0,18	0,0234	0,42	0,20	0,0348
ln (TMDAS)		0,16	0,07	0,0152	0,16	0,07	0,015	0,16	0,07	0,0152	0,16	0,06	0,0039	0,16	0,07	0,0334	0,16	0,09	0,0822
Interceto*		-	-	-	-	-	-	-	-	-	-	-	-	4,95	6,84	0,4698	5,33	9,59	0,5785
ln (TMDAP)*		-	-	-	-	-	-	-	-	-	-	-	-	-0,66	0,85	0,4406	-0,77	1,39	0,5805
ln (TMDAS)*		-	-	-	-	-	-	-	-	-	-	-	-	-0,05	0,42	0,9032	0,02	0,75	0,984
QIC ou AIC		1177,45			1176,81			1177,31			553,10			555,61			557,10		
R <sup>2</sup> marginal	0,0334			0,0335			0,0335			0,1447			-			-			
3	Interceto	-5,10	0,85	<0,0001	-5,12	0,85	<0,0001	-5,10	0,85	<0,0001	-3,18	0,77	<,0001	-2,56	0,84	0,0022	-2,65	0,89	0,003
	ln (TMDAPMS)	0,23	0,05	<0,0001	0,23	0,05	<0,0001	0,23	0,05	<0,0001	0,22	0,05	<,0001	0,19	0,05	0,0001	0,20	0,05	0,0002
	Interceto*	-	-	-	-	-	-	-	-	-	-	-	-	1,34	3,54	0,7045	1,20	4,20	0,7749
	ln (TMDAPMS)*	-	-	-	-	-	-	-	-	-	-	-	-	-0,19	0,22	0,3928	-0,19	0,26	0,4616
	QIC ou AIC	1189,18			1188,87			1189,08			555,44			555,55			557,06		
	R <sup>2</sup> marginal	0,0287			0,0287			0,0287			0,1275			-			-		
4	Interceto	-8,01	1,56	<0,0001	-8,06	1,56	<0,0001	-8,01	1,56	<0,0001	-5,95	1,47	<,0001	-6,05	1,31	<,0001	-4,99	1,81	0,0057
	ln (TMDAT)	0,72	0,16	<0,0001	0,72	0,16	<0,0001	0,72	0,16	<0,0001	0,69	0,15	<,0001	0,70	0,13	<,0001	0,61	0,17	0,0004
	ln (TMDASDP)	0,13	0,06	0,0422	0,13	0,06	0,0423	0,13	0,06	0,0424	0,13	0,05	0,0193	0,13	0,05	0,0063	0,12	0,09	0,1799
	Interceto*	-	-	-	-	-	-	-	-	-	-	-	-	23,13	15,84	0,1443	3,55	9,18	0,6987
	ln (TMDAT)*	-	-	-	-	-	-	-	-	-	-	-	-	-5,20	136,22	0,9696	-0,58	0,82	0,4781
	ln (TMDASDP)*	-	-	-	-	-	-	-	-	-	-	-	-	-1,29	83,95	0,9878	-0,05	0,68	0,9403
	QIC ou AIC	1180,07			1179,48			1179,97			552,22			561,94			556,07		
R <sup>2</sup> marginal	0,0355			0,0355			0,0355			0,1541			-			-			

\*Coeficientes das observações nulas

## 4.2. Análises dos resultados da modelação

O resumo dos resultados dos modelos de previsão dos acidentes desenvolvidos para as interseções de três ramos e selecionados como os mais adequados para cada caso é apresentado a seguir. Na Tabela 6 são expostas as formas funcionais dos modelos com sua variável dependente, as técnicas de modelação (MLG ou EEG) e os elementos estatísticos de análise dos modelos (nível de significância, QIC, estrutura de correlação e  $R^2$  marginal). A investigação dos resíduos, feita através do *CURE Plot*, foi ilustrada nas Figuras 1.b e 2.b.

Relativamente à variável tráfego, representada por quatro diferentes formas funcionais, verifica-se que esta produziu um bom ajuste em todos os casos. A sua contribuição para a segurança viária foi sempre negativa, correspondendo a um incremento no número de acidentes com o aumento do tráfego. Em ambos os casos a forma funcional que apresentou melhor desempenho foi a forma funcional 2 da Tabela 5, confirmando o estudo realizado por Miaou & Lord (2003). Nos modelos de previsão os valores dos coeficientes das variáveis de tráfego foram divergentes conforme a combinação utilizada no ajuste e a técnica de modelação. Como pode ser observado na Tabela 7 os modelos finais apresentam um nível de significância de 5%.

Tabela 7 – Equações e estatísticas descritivas dos modelos finais das interseções em T

Técnica de Modelação	Equação do Modelo	$\alpha$	QIC ou AIC	Estrutura de Correlação	$R^2$ marginal
EEG	$y = e^{-7.83} \times (TMDAP)^{0.55} \times (TMDAS)^{0.16}$	5%	1176,81	Permutável	0,0335
MLG	$y = e^{-5.74} \times (TMDAP)^{0.53} \times (TMDAS)^{0.16}$	5%	553,10	-	0,1447

## 5. CONCLUSÕES

O objetivo principal do presente trabalho foi uma investigação exploratória das formas funcionais que envolvem somente o tráfego e que são amplamente difundidas no estudo da sinistralidade. Assim, este estudo foi feito através do desenvolvimento de modelos de previsão dos acidentes de viação ocorridos nas interseções de três ramos inseridas nas EN's de uma faixa de rodagem com duas vias de sentidos opostos localizadas na região norte de Portugal. Estas interseções estão inseridas em seis trechos, nomeadamente: (i) Braga-Famalicão (EN-14); (ii) Braga-Vila Verde (EN-101); (iii) Braga-Guimarães (EN-101); (iv) Guimarães-Felgueiras (EN-101); (v) Famalicão-Guimarães (EN-206); e (vi) Guimarães-Fafe (EN-206).

A principal contribuição deste trabalho foi proporcionar uma correta identificação da forma funcional envolvendo a variável de exposição (tráfego), para que sejam incorporados, posteriormente, os demais fatores contribuintes para a ocorrência de acidentes de viação. Destaca-se ainda que o estudo da sinistralidade desta zona portuguesa é pouca explorada, principalmente na promoção da segurança no sistema nacional rodoviário, que liga diversas cidades e zonas industriais.

Para o desenvolvimento dos modelos foram utilizados dados de acidentes com vítimas (colisões e despistes) de 2005 a 2010, fornecidos pela Autoridade Nacional de Segurança Rodoviária – ANSR, assim como, os dados de tráfego, que foram estimados com base em contagens de tráfego disponibilizadas pela empresa Estradas Portuguesas – EP. No processo de modelação utilizou-se diferentes técnicas de modelação também amplamente difundidas no estudo da sinistralidade, sendo estas técnicas: (i) Modelos Lineares Generalizados – MLG; (ii) Equações de Estimção Generalizadas – EEG e (iii) Modelos Zero Inflacionados (ZIP e

ZINB). As distribuições de Poisson e binomial negativa foram eleitas como a componente do erro aleatório dos modelos.

No ajuste dos modelos utilizou-se uma base de dados principal. Esta base de dados foi definida para modelação dos acidentes ocorridos na direção principal das interseções de três ramos em que foram observados os dados dos acidentes com vítimas e o tráfego (representado pelo tráfego médio diário anual – TMDA da direção principal e da direção secundária) para o período de estudo, que compreende os anos de 2005 a 2010. Também foi produzida uma base de dados derivada, composta pelo agrupamento dos dados dos seis anos.

Nos modelos de previsão obtidos, verificou-se que a variável de tráfego, representada pelo TMDA da direção principal e da direção secundária das interseções, produziu um bom ajuste em todos os casos e também com contribuição negativa para a segurança viária, correspondendo a um incremento no número de acidentes com o aumento do tráfego. Os valores dos coeficientes das variáveis de tráfego foram divergentes conforme a técnica de modelação utilizada no ajuste destes modelos.

Um ponto a destacar é que os modelos de previsão ajustados com zero inflacionados, ZIP e ZINB, não apresentaram resultados significativos, indicando que as observações nulas presentes na base de dados não é explicada pela variação do tráfego, visto que o uso destes modelos permite identificar os fatores que contribuem para as observações não nulas e as nulas, conforme Carson & Mannering (2001). Relativamente aos modelos ajustados com EEG, estes apresentaram resultados significativos, mas inferiores aos modelos ajustados com MLG, mostrando que para a situação estudada o uso de dados desagregados não apresenta melhor resultado que o obtido com os dados agregados.

Por fim, o estudo mostra que a análise em conjunto do *CURE Plot* e do  $R^2$  marginal dos modelos de previsão desenvolvidos para os acidentes considerados (colisões e despistes com vítimas) das interseções de três ramos indica que a forma funcional 2 da Tabela 5 apresenta melhores resultados (Tabela 6, Figura 1.b e Figura 2.b) para as duas técnicas de modelação válidas para a base de dados.

Portanto, considera-se que o presente trabalho estabelece uma contribuição importante para a melhoria da segurança rodoviária nas EN's da região do Minho no norte de Portugal, visto que os modelos de previsão desenvolvidos servem de base para o estudo da sinistralidade por parte dos responsáveis da gestão da infraestrutura rodoviária no sentido de fundamentar intervenções na mesma, direcionadas para a mitigação desta sinistralidade.

No entanto, é importante destacar algumas limitações do presente trabalho. A principal foi a impossibilidade de incluir no estudo os acidentes ocorridos na direção secundária das interseções, visto que na base de dados fornecida pela ANSR não foi possível obter a localização dos acidentes ocorridos na área de influência das interseções nessa direção. Isso ocorreu porque no boletim estatístico de acidentes de viação – BEAV esta informação não era preenchida em conformidade, ou seja, com um padrão de anotação dos endereços. Outra limitação foi o pequeno tamanho da amostra das interseções, em parte devido aos altos custos de recolha de dados e escassez de recursos logísticos e equipamentos. Já como recomendação merece destacar que num estudo posterior devem ser consideradas outras variáveis além das variáveis de exposição do tráfego, testando as formas funcionais aceites neste estudo para verificar se o acréscimo de novas variáveis permite o desenvolvimento de modelos com maior capacidade de explicação da frequência de acidentes observada nas interseções das EN's da região norte de Portugal. A realização de estudos similares referentes a interseções localizadas nas EN's de outras regiões de Portugal também é recomendada.

## AGRADECIMENTOS

Os autores agradecem o suporte financeiro concedido pela Fundação para a Ciência e a Tecnologia (FCT), Portugal, através do financiamento da bolsa de doutoramento SFRH/BD/62458/2009 e do projeto UI 4047 – 2014: PEst-OE/ECI/UI4047/2014.

## REFERÊNCIAS

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Cardoso, J., Eenink, R., Elvik, R., Janssen, T., Reurings, M., & Stefan, C. (2005). *Accident Prediction Models and Road safety Impact Assessment: a state-of-the-art*. European Union: RIPCORD-ISEREST.
- Carson, J., & Mannering, F. (2001). The effect of ice warning signs on accident frequencies and severities. *Accident Analysis and Prevention*, 33 (1), 99-109.
- Cox, D., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society*, 30(2), 248-275.
- FHWA - Federal Highway Administration; TFHRC - Turner-Fairbank Highway Research Center. (2003). *Crash Prediction Module (CPM) Engineer's Manual*. EUA: IHSDM.
- Gomes, S. V. (2010). *Avaliação da Influência da Infra-estrutura na Segurança Rodoviária em Meio Urbano*. Coimbra: Faculdade de Ciências e Tecnologia da Universidade de Coimbra.
- Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimating equations*. Chapman and Hall/CRC Press.
- Hauer, E. (1997). *Observational Before After Studies in Road Safety*. Oxford, UK: Pergamon Press, Elsevier Science Ltd.
- Hauer, E. (2004). Statistical Road Safety Modeling. *Transportation Research Record*, 1897, 81-87.
- Kim, D., & Washington, S. (2006). The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Accident Analysis and Prevention*, 38, 1094-1100.
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22.
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A*, 44, 291-305.
- Lord, D., & Park, P. Y. (2008). Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. *Accident Analysis & Prevention*, 40(4), 1441-1457.
- Lord, D., & Persaud, B. N. (2000). Accident Prediction Models With and Without Trend – Application of the Generalized Estimating Equations Procedure. *Transportation Research Record*, 1717, 102-108.
- Lord, D., Washington, S. P., & Ivan, J. N. (2005b). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*, 37, 35-46.

Miaou, S., & Lord, D. (2003). Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus Empirical Bayes. *Transportation Research Record*, 1840, 31–40.

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, 57, 120-125.

Vogt, A., & Bared, J. (1997). Accident models for two-lane rural segments and intersections. *Transportation Research Record*, 1635, 18-29.

Wang, X., & Abdel-Aty, M. (2006). Temporal and spatial analysis of rear-end crashes at signalized intersections. *Accident Analysis and Prevention*, 38, 1137-1150.

Zheng, B. (2000). Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine*, 19(10), 1265-1275.