

INVESTIGAÇÃO DE NOVAS FORMAS FUNCIONAIS PARA MELHORIA DA APLICABILIDADE DOS MODELOS DE ESTIMATIVA DE FREQUÊNCIA DE ACIDENTES

Carlos Roque¹, João Lourenço Cardoso²

¹ Laboratório Nacional de Engenharia Civil, Departamento de Transportes. Núcleo de Planeamento, Tráfego e Segurança., Av do Brasil 101, 1700-066 Lisboa, Portugal

email: croque@lnec.pt

² Laboratório Nacional de Engenharia Civil, Departamento de Transportes. Núcleo de Planeamento, Tráfego e Segurança, Av do Brasil 101, 1700-066 Lisboa, Portugal

Sumário

Em segurança rodoviária, a modelação da sinistralidade é desenvolvida com o propósito de efetuar uma estimativa realista da frequência esperada de acidentes ou de vítimas, em função de um conjunto de variáveis explicativas associadas a um determinado trecho do sistema de transporte (por exemplo, referentes ao tráfego e às características da estrada e da sua área adjacente à faixa de rodagem).

O desenvolvimento de modelos de estimativa da frequência esperada de acidentes (MEFA) consiste no estabelecimento de relações entre variáveis explicativas e o número de acidentes ou de vítimas, através de funções matemáticas. Este procedimento permite quantificar a variação da sinistralidade associada à alteração de cada variável explicativa considerada.

Apesar da sua relevância no desenvolvimento dos MEFA, a procura de uma forma funcional apropriada para estes modelos, relacionando frequência de acidentes e volume de tráfego, é tema escassamente abordado na bibliografia existente. Trata-se, no entanto, de um aspeto fundamental, para melhorar quer a representação de explicações causais quer a aplicabilidade prática dos modelos.

No caso da análise de dados relativos a despistes, a aplicação de modelos de Binomial Negativa com diferentes formas funcionais permite representações mais realistas das relações subjacentes entre o número de despistes e o volume de tráfego.

Nesta comunicação é justificada a noção de que a modelação de despistes com uma forma funcional que represente a relação entre a frequência de despistes e o volume de tráfego mediante curva convexa mostra melhor qualidade de ajuste do que a forma funcional tradicional, monotonamente variada, correspondendo ao pressuposto de a frequência de despistes aumentar continuamente com o volume de tráfego diário.

Palavras-chave: Modelos de estimativa de frequência de acidentes; Despiste; Forma funcional; Intervalo de confiança; Nível de serviço.

1 INTRODUÇÃO

Em segurança rodoviária, a modelação da sinistralidade é desenvolvida com o propósito de efetuar uma estimativa realista da frequência esperada de acidentes ou de vítimas, em função de um conjunto de variáveis explicativas associadas a um determinado trecho do sistema de transporte (por exemplo, referentes ao tráfego e às características da estrada e da sua área adjacente à faixa de rodagem).

O desenvolvimento de modelos de estimativa da frequência esperada de acidentes (MEFA) consiste no estabelecimento de relações entre variáveis explicativas e o número de acidentes ou de vítimas, através de funções matemáticas. Este procedimento permite quantificar a variação da sinistralidade associada à alteração de cada variável explicativa considerada.

Apesar da sua relevância no desenvolvimento dos MEFA, a procura de uma forma funcional apropriada para estes modelos, relacionando frequência de acidentes e volume de tráfego, é tema escassamente abordado na bibliografia existente. Trata-se, no entanto, de um aspeto fundamental, para melhorar quer a representação de explicações causais quer a aplicabilidade prática dos modelos.

Tal como sugerido por Lord *et al.* [1], pode ser preferível começar por desenvolver modelos que tenham em consideração o processo de produção de acidentes em vez de os esforços serem, desde o início, dirigidos para a melhoria do ajuste matemático de um modelo cuja forma geral não espelhe aquele processo.

No caso dos despistes, evidências empíricas e considerações lógicas levam à convicção de que a relação entre a frequência esperada de acidentes e o volume de tráfego não é linear nem monótona. Para volumes de tráfego baixos pode esperar-se que o número de despistes por unidade de tempo seja proporcional ao volume de tráfego. Mas, à medida que aumenta o volume de tráfego, torna-se cada vez mais difícil não colidir com outro veículo. Assim, no caso dos despistes, não é expectável manter a referida proporcionalidade para volumes de tráfego mais elevados. No congestionamento, os despistes são impossíveis, exceto para condições de aderência ao pavimento muito reduzidas (por exemplo, na presença de gelo ou neve no pavimento).

Esta relação decorre das diferenças comportamentais dos condutores em função do nível de volume de tráfego (reduzido, elevado, ou congestionado), permite inferir que a frequência de despistes dependerá do estado do sistema de tráfego rodoviário (fluxo, velocidade e densidade) no tempo e no espaço [2].

Assim, no caso da análise de dados relativos a despistes, a aplicação de modelos de Binomial Negativa com diferentes formas funcionais permite representações mais realistas das relações subjacentes entre o número de despistes e o volume de tráfego.

A análise reportada na presente comunicação demonstra que, para o caso particular dos despistes, a forma funcional de um MEFA – no que se refere ao volume de tráfego diário – influencia a qualidade das estimativas obtidas. Assim, analisam-se as diferenças potenciais entre modelos com duas formas funcionais distintas, usando para o efeito diversas medidas da qualidade de ajustamento, nomeadamente o Critério de Informação de Akaike (AIC), o G^2 , o desvio médio do modelo (MAD) e o Índice de Elvik. A qualidade obtida na modelação foi ainda avaliada através do método CURE aplicado à variável referente ao tráfego médio diário.

Nesta comunicação é justificada a noção de que a modelação de despistes com uma forma funcional que represente a relação entre a frequência de despistes e o volume de tráfego mediante curva convexa mostra melhor qualidade de ajuste do que a forma funcional tradicional, monotonamente variada, correspondendo ao pressuposto de a frequência de despistes aumentar continuamente com o volume de tráfego diário.

2 METODOLOGIA

Neste âmbito, diversos autores têm vindo a identificar um conjunto de questões relacionadas com os dados e com aspetos metodológicos como potenciais fontes de erro na definição de modelos estatísticos, podendo originar estimativas desajustadas das frequências de acidentes. Essas questões foram recentemente sintetizadas por Lord e Mannering [3]

O desenvolvimento dos MEFA descritos foi efetuado através da utilização das técnicas de Modelação Linear Generalizada (aplicando o método da máxima verosimilhança), mais concretamente através do modelo de Binomial Negativa. A adoção deste tipo de modelo ficou a dever-se ao facto de todas as amostras apresentarem sobre-dispersão.

O modelo de Binomial Negativa (ou Poisson-gama) é uma extensão do modelo de regressão de Poisson adaptado à possível existência de sobre-dispersão nos dados. O modelo de Binomial Negativa e de Poisson seguem uma distribuição de probabilidade Gama [3]. O modelo de Binomial Negativa é derivado re-escrevendo o parâmetro de Poisson para cada observação i num determinado intervalo de tempo como:

$$\lambda_i = \exp(\beta X_i + \varepsilon_i) \quad (1)$$

onde $\exp(\varepsilon_i)$ é um termo de erro aleatório que segue uma distribuição de probabilidade gama com média 1 e variância α . A soma deste termo permite que a variância divirja da média de acordo com a seguinte expressão [3]:

$$VAR[y_i] = E[y_i] [1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2 \quad (2)$$

A aplicação de modelos de Binomial Negativa com diferentes formas funcionais na análise de dados relativos a despistes, permite explorar as relações subjacentes entre a frequência de ocorrência de despistes e o fluxo de tráfego. Assim, para examinar essas relações, nos referidos modelos foi utilizado um conjunto selecionado de trechos de estradas de faixa de rodagem dupla da Rede Rodoviária Nacional (RRN). Foi ainda usado um estudo de simulação de Monte Carlo (*Monte Carlo simulation study*) para avaliar o seu desempenho genérico. Este estudo é descrito detalhadamente em [4].

Os modelos foram desenvolvidos utilizando o *software R* (versão 2.15.0) [5], tendo sido utilizadas as aplicações “*MASS*” e “*car*”.

A seleção das variáveis explicativas a incluir nos modelos foi realizada recorrendo aos métodos de regressão passo a passo (*stepwise regression*). O nível de significância escolhido foi de 5%.

As variáveis explicativas *TMD* e *Extensão* (por não terem valores nulos) foram transformadas para a escala logarítmica. Estas transformações permitem reduzir os desvios e minimizar a heterocedasticidade das variáveis [6].

2.1 Formas funcionais

No ajuste dos MEFA foram consideradas duas formas funcionais alternativas, embora usando as mesmas variáveis explicativas: o Tráfego Médio Diário (*TMD*), que corresponde à média do tráfego médio diário durante o período considerado (quatro anos), e a extensão do trecho (*Extensão*).

Na primeira forma funcional considerada, doravante designada Forma Funcional 1 (FF1), o número esperado de despistes por unidade de tempo para o segmento i (λ_i) é função de duas variáveis explicativas: Tráfego Médio Diário no trecho de estrada i , TMD_i , e comprimento (em quilómetros) do trecho de estrada i ($Extensão_i$). Na equação (1), os efeitos do volume de tráfego e do comprimento dos trechos são modelados em termos de elasticidade, isto é, uma potência, β_1 ou β_2 , à qual o TMD_i ou a $Extensão_i$ são elevados. Embora este modelo possa sofrer de enviesamento por variáveis omissas, a avaliação empírica realizada fornece, ainda assim, informações úteis para o desenvolvimento de melhores MEFA.

O número esperado de despistes que ocorre num trecho de estrada i (λ_i) pode ser calculado com a equação:

$$\lambda_i = \beta_0 TMD_i^{\beta_1} Extensão_i^{\beta_2} \quad (3)$$

Onde, β_0 é um coeficiente geral (a estimar), β_1 é o coeficiente associado ao TMD (a estimar) e β_2 é o coeficiente associado à *Extensão* (a estimar).

É geralmente aceite que, à medida que o volume de tráfego aumenta, o número total de acidentes aumenta com uma taxa decrescente (ver Fig.1.a). Em MEFA, esta relação caracteriza-se por um coeficiente para a variável volume de tráfego (β_1) inferior a 1.

Na segunda forma funcional considerada, a relação entre o número esperado de despistes, λ_i , e o TMD, é descrita através de uma modificação do modelo de Ricker [8], o qual foi inicialmente utilizado em investigação na área da ecologia, onde a densidade populacional de animais e plantas não evidencia, normalmente, uma tendência indefinidamente crescente, revelando, pelo contrário, máximos locais no espaço da variável explicativa. Este modelo é importante para estimar taxas de sobrevivência.

Aplicado à estimativa de frequência de acidentes, o modelo de Ricker é expresso pela equação genérica:

$$\lambda_i = \beta_0 TMD_i e^{-\beta_1 TMD_i} \quad (4)$$

Graficamente, esta é uma curva que passa pela origem, que apresenta um máximo e que diminui assintoticamente para zero. No modelo Ricker, que não é monotonamente crescente, a cada valor do número de acidentes diferente do máximo correspondem dois valores de TMD.

O modelo modificado de Ricker na segunda forma funcional – doravante designada Forma Funcional 2 (FF2) – é calculado juntando ao referido modelo o parâmetro β_1 , afetando a variável TMD, e o parâmetro β_3 , associado à variável explicativa “extensão do trecho” (*Extensão*):

$$\lambda_i = \beta_0 TMD_i^{\beta_1} e^{\beta_2 \frac{TMD}{10\,000}} Extensão_i^{\beta_3} \quad (5)$$

Onde, λ_i corresponde ao número esperado de despistes que ocorre num trecho de estrada i , TMD_i corresponde ao Tráfego Médio Diário no trecho de estrada i , $Extensão_i$ refere-se ao comprimento (em quilômetros) do trecho de estrada i , β_0 é um coeficiente (a estimar), β_1 e β_2 são coeficientes associados ao TMD (a estimar) e β_3 é um coeficiente associado a *Extensão* (a estimar). Neste caso, é utilizado TMD/10000 para evitar uma eventual sobrecarga (*overflow*) no processo de estimação. De acordo com as considerações lógicas acima descritas, β_1 deve ser maior que 1 e β_2 deve ser negativo (ver Fig.1.b).

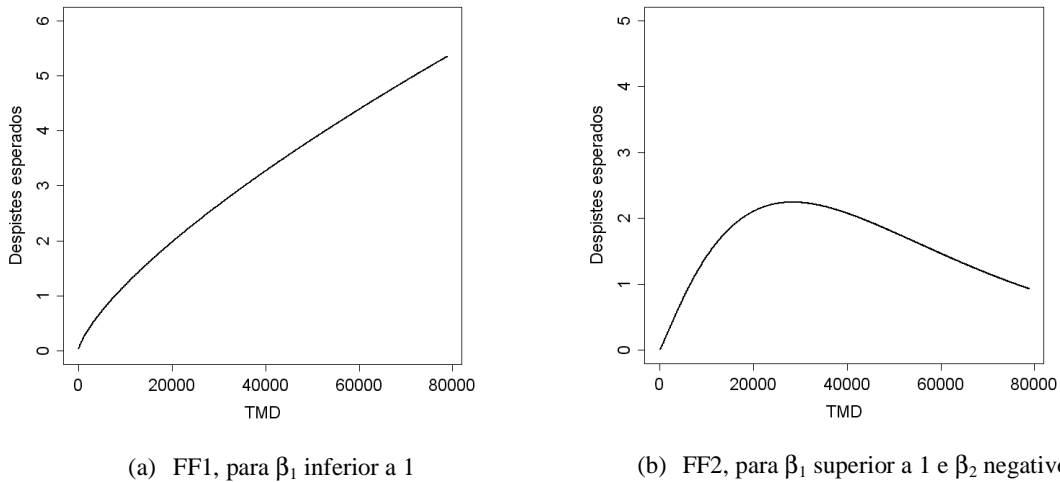


Fig.1. Formas funcionais consideradas: FF1 (a) e FF2 (b).

Com esta forma funcional e para um determinado trecho de estrada, o comportamento da variável de resposta à variação do TMD (que reflete o aumento do volume de tráfego no referido trecho) vai ao encontro das considerações descritas na Introdução, apresentando uma configuração em forma de U invertido. A frequência de despistes aumenta com o TMD até atingir um máximo; a partir daí, para níveis de TMD superiores àquele a que corresponde esse número máximo de despistes, o número de despistes diminui com o aumento do TMD.

2.2 Qualidade de ajustamento

No âmbito do presente trabalho foram utilizadas diversas estatísticas de ajuste para avaliar a qualidade obtida na modelação. Algumas destas medidas da qualidade de ajustamento permitem explicar o desempenho dos modelos em termos de poder explicativo, usando o desvio médio do modelo (MAD), o R_p^2 e o Índice de Elvik para esse efeito. Outras servem exclusivamente para efetuar comparações no desenvolvimento do modelo: o Critério de Informação de Akaike (AIC) e o Somatório dos desvios do modelo (G^2).

Foi ainda utilizado o método CURE (*CUmulative REsiduals*), originalmente definido por Hauer e Bamfo [9], para analisar o ajuste dos MEFA à variável explicativa TMD, ao longo do intervalo de variação desta variável.

2.3 Estimação de intervalos de confiança e de previsão

Paralelamente à avaliação da qualidade do ajuste conseguido com os MEFA, interessa dispor de intervalos de confiança para os parâmetros dos modelos e de intervalos de previsão para os resultados da sua aplicação na estimativa da sinistralidade de um trecho. Wood [10] propôs um método que permite estimar intervalos de confiança para o número esperado de acidentes λ , bem como intervalos de previsão para m e y , designados nível de segurança (*safety*) e número de acidentes (observados) respetivamente. De acordo com Hauer [2], a segurança de uma entidade, m , corresponde ao número esperado de acidentes que ocorrerá nessa mesma entidade durante um determinado intervalo de tempo, sendo também designada como média Gama (*Gamma mean*). Estes intervalos fornecem informação sobre a amplitude da variação destas quantidades.

No Quadro 1 encontram-se as equações para o cálculo dos intervalos de confiança e de previsão de 95% para modelos de Binomial Negativa. Nesta tabela, η_i é o logaritmo da resposta média estimada λ_i para o segmento i , enquanto ϕ é o inverso do parâmetro de dispersão (*inverse dispersion parameter*) estimado no desenvolvimento do modelo.

Quadro 1 – Intervalos de confiança e de previsão de 95% [10].

Parâmetro	Intervalos
λ	$\left[\frac{\hat{\lambda}_i}{e^{1.96} \sqrt{\text{Var}(\hat{\eta}_i)}}, \hat{\lambda}_i e^{1.96} \sqrt{\text{Var}(\hat{\eta}_i)} \right]$
m	$\left[\max \left\{ 0, \hat{\lambda}_i - 1.96 \sqrt{\hat{\lambda}_i^2 \text{Var}(\hat{\eta}_i) + \frac{\hat{\lambda}_i^2 \text{Var}(\hat{\eta}_i) + \hat{\lambda}_i^2}{\phi}} \right\}, \hat{\lambda}_i + 1.96 \sqrt{\hat{\lambda}_i^2 \text{Var}(\hat{\eta}_i) + \frac{\hat{\lambda}_i^2 \text{Var}(\hat{\eta}_i) + \hat{\lambda}_i^2}{\phi}} \right]$
y	$\left[0, \left[\hat{\lambda}_i + \sqrt{19} \sqrt{\hat{\lambda}_i^2 \text{Var}(\hat{\eta}_i) + \frac{\hat{\lambda}_i^2 \text{Var}(\hat{\eta}_i) + \hat{\lambda}_i^2}{\phi}} + \hat{\lambda}_i \right] \right]$

O cálculo dos intervalos de confiança para λ_i , e dos intervalos de previsão para m e y permite detetar diferenças significativas nos MEFA usando formas funcionais alternativas. Para esse efeito são utilizadas as equações constantes do Quadro 1. A variância do logaritmo de λ_i para o segmento i , $\text{Var}(\eta_i)$, foi calculada utilizando $XI^{-1}XT$, onde I^{-1} representa a matriz de covariâncias (*variance-covariance matrix*) e X corresponde à matriz que contém os valores observados. As matrizes de covariâncias foram calculadas recorrendo à aplicação R .

3 MODELOS DESENVOLVIDOS

Neste capítulo são apresentados os MEFA desenvolvidos para trechos unidireccionais de auto-estrada da RRN. Os modelos aqui descritos são apresentados de forma detalhada em [4].

3.1 Descrição dos dados

Foram apenas considerados acidentes com vítimas (mortos, feridos graves e feridos ligeiros) envolvendo um único veículo (no caso do atual BEAV correspondentes a “Despistes”), fora da zona dos nós. Assim, para efeitos de modelação da sinistralidade, os dados analisados neste estudo foram recolhidos a partir de 796 trechos unidireccionais, relativos a acidentes ocorridos no período compreendido entre 2007 e 2010.

No Quadro 2 apresentam-se resumos com estatísticas relevantes para as variáveis explicativas e para o número de despistes recolhidos durante o período considerado (quatro anos).

Quadro 2 – Estatísticas selecionadas para trechos unidirecionais em estradas com dupla faixa de rodagem.

Variável	Mínimo	Máximo	Média (desvio padrão)	Total
Extensão do trecho (km)	0,31	33,28	5,66 (5,52)	4913,32
TMD (veículos/dia)	1056	104778	16838,37 (17192,6)	-
Número de despistes	0	76	5,86 (7,48)	5085,00

3.2 Resultados dos modelos

Verificou-se um melhor ajuste nos modelos desagregados pelo número de vias (duas e três ou quatro vias). No Quadro 3 são apresentadas as estimativas dos coeficientes dos MEFA obtidos, respectivamente com a FF1 e a FF2. No mesmo quadro são ainda apresentadas as estatísticas de ajuste utilizadas para avaliar a qualidade dos MEFA desenvolvidos. O modelo com FF2 para trechos de autoestrada com duas vias por faixa apresenta, por um lado, valores de AIC, G^2 e MAD inferiores, e, por outro, valores para o índice de Elvik e para o R^2 ajustado superiores, demonstrando claramente um melhor ajuste estatístico do que o modelo com FF1.

Quadro 3 – Estimativa dos coeficientes dos MEFA e medidas da qualidade de ajustamento

Forma funcional	Nº de vias	$\ln(\beta_0)$	β_1	β_2	β_3	AIC	G^2	MAD	Elvik's index
FF1	2	-6.493	0.724	0.874	-	2943.0	1732.32	2.661	0.572
	3 ou 4	-8.080	0.880	1.035	-	1312.1	857.59	3.499	0.699
FF2	2	-9.936	1.161	-0.410	0.847	2923.4	1674.07	2.625	0.600
	3 ou 4	-	-	-	-	-	-	-	-

Na Fig.2 são apresentados os gráficos CURE relativos às duas formas funcionais – gráficos referentes à FF1 e FF2 nos lados esquerdo e direito, respectivamente. O gráfico CURE para o MEFA com FF2 apresenta um ajuste claramente superior ao do modelo com FF1. Ao contrário do que acontece com o MEFA com FF2, o modelo com FF1 é significativamente enviesado ao longo do conjunto de valores da variável TMD, excede na larga maioria dos valores a banda contida entre $\pm 2\sigma^*$, não oscila em torno da origem e termina afastada do valor nulo.

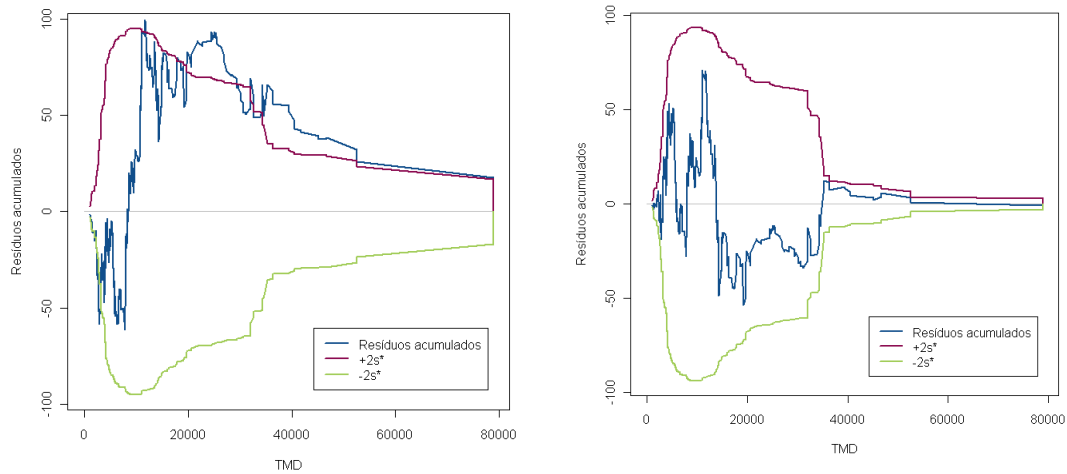


Fig.2. Comparação dos MEFA com FF1 (à esquerda) e FF2 (à direita) para trechos de um quilómetro e duas vias por faixa.

A Fig.3. apresenta os intervalos de previsão de 95% para a média Gama (m) e para o número de despistes (y). Esta figura demonstra que a largura do intervalo de previsão para a média Gama (m) do MEFA com FF2 é claramente inferior à do modelo com FF1, para valores de TMD superiores a 30 000 veículos/dia. Esta figura apresenta ainda os intervalos de previsão de 95% para o número de despistes (y) de MEFA com FF1 e FF2. Também para estes intervalos as diferenças são bastante significativas, apresentando o MEFA com FF2 um intervalo de previsão significativamente mais estreito em, pelo menos, dois terços da banda de valores que a variável explicativa TMD pode tomar.

Não foi possível calibrar um modelo com FF2 para trechos de dupla faixa de rodagem com 3 ou mais vias em cada sentido. Para averiguar os motivos desta impossibilidade usou-se o MEFA com FF1 para trechos de duas vias por faixa de rodagem para relacionar o TMD com o nível de serviço na hora de ponta, admitindo que nesta flui 10% do TMD . Para avaliação do nível de serviço nesta hora foi utilizada a metodologia do *Highway Capacity Manual* [11]. Admitiram-se os seguintes pressupostos para trechos de autoestrada com duas vias por faixa de rodagem: largura das vias = 3.75 m; desobstrução lateral: 2.90 m; velocidade em regime livre base = 120 km/h; tráfego habitual; volume em hora de ponta numa direção: 10% do TMD ; percentagem de veículos pesados = 2%; Fator de Ponta Horária = 0.92; densidade de pontos de convergência: 1 ponto/km; e terreno plano.

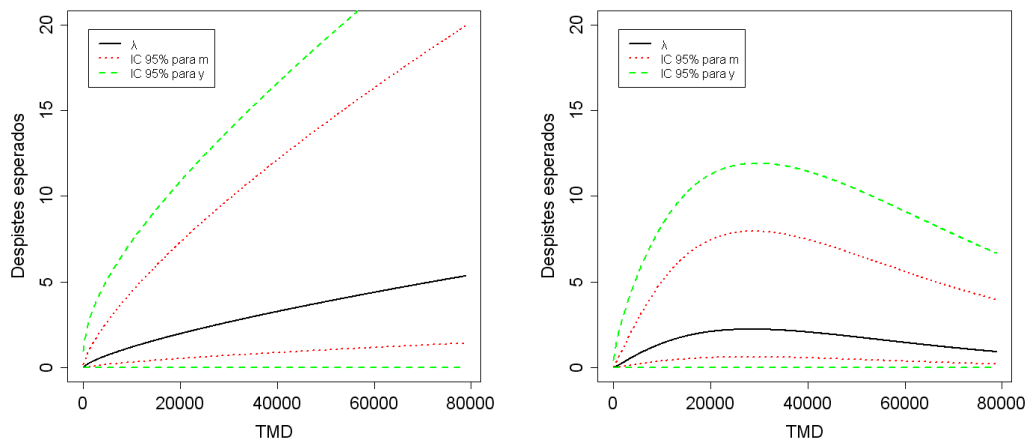


Fig.3. Intervalo de previsão de 95% para a média Gama (m) e para o número de despistes (y) de MEFD com FF1 (à esquerda) e FF2 (à direita) para trechos de um quilómetro e duas vias por faixa.

Os resultados do cálculo dos níveis de serviço (NS) em função do TMD foram sobrepostos à Equação (5) relativa ao MEFA com FF2 em trechos de um quilómetro com duas vias por sentido, sendo os mesmos apresentados na Fig.4 juntamente com o número observado de despistes por quilómetro relativo às observações que serviram de base ao referido MEFA.

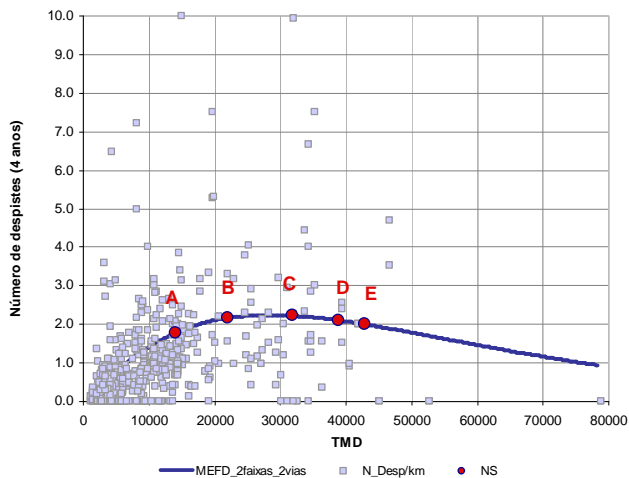


Fig.4. Relação entre o número esperado de despistes, o TMD e os níveis de serviço para trechos de um quilómetro de estradas de dupla faixa de rodagem e duas vias por sentido

Na referida figura é possível observar que o máximo do número esperado de despistes é atingido no NS C (próximo dos 30 000 veículos por dia) e que a partir desse ponto o número esperado de despistes começa a decrescer até se atingir a capacidade (NS E), próximo dos 42 500 veículos por dia. O TMD máximo registado no conjunto das observações é de, aproximadamente, 42 500 veículos por dia.

Verifica-se mais uma vez que, nesta forma funcional, os valores da variável de resposta em resultado da variação do TMD se ajustam às considerações descritas em 2.1. Destaca-se ainda o facto de a variável de resposta atingir

o seu máximo no NS C, correspondendo a uma concentração entre 11 e 16 veículos ligeiros de passageiros por quilómetro por via [11]; ou seja condições de circulação ainda satisfatórias e correspondendo a escoamento estável.

No caso dos trechos de três ou quatro vias por sentido, a capacidade deverá ser aproximadamente uma vez e meia a duas vezes superior à capacidade dos trechos de duas vias. Para ajustar uma equação com FF2 nestas condições seria necessário ter nestes trechos condições de circulação degradadas em parte significativa do tempo, ou seja, um conjunto significativo de observações com valores de *TMD* elevados (entre 45 000 e 160 000 veículos por dia). Na Fig.5 observa-se que, em trechos com três ou quatro vias por sentido, são escassas as observações nestas condições, sendo o *TMD* máximo observado de aproximadamente 105 000 veículos por dia. Ainda assim, constata-se que o número observado de despistes por quilómetro decresce para valores de *TMD* superiores a 60 000 veículos por dia, o que leva a crer que dispor de um maior número de observações com *TMD* elevado seria possível ajustar com sucesso MEFA com FF2. No entanto, este tipo de modelo apresenta um problema adicional, que é o de agregar trechos com capacidades e NS diferentes, o que também dificulta o ajuste de MEFA com FF2. Por exemplo, na Fig.5 constata-se a existência de intervalos de valores de *TMD* aos quais correspondem níveis de serviço distintos consoante as faixas tenham três ou quatro vias (no cálculo dos NS em trechos de 3 e 4 vias foram mantidos os pressupostos da figura anterior, com exceção do número de vias).

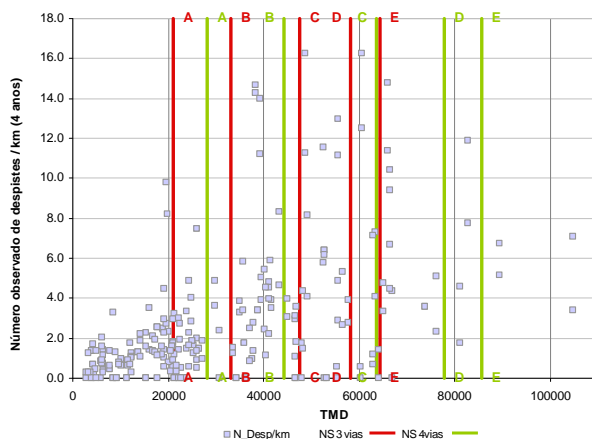


Fig.5. Relação entre o número observado de despistes, o TMD e os níveis de serviço para trechos de estradas de dupla faixa de rodagem e três ou quatro vias por sentido.

4 CONCLUSÕES

Nos modelos apresentados na presente comunicação foram utilizadas as técnicas de modelação linear generalizada, recorrendo para esse efeito ao modelo de Binomial Negativa. Nestes, foram testadas diferentes formas funcionais contribuindo-se para um melhor conhecimento da relação entre o processo de despiste envolvendo a AAFR e o volume de tráfego médio diário.

A análise dos modelos com formas funcionais alternativas permitiu, de forma sumária, verificar que os gráficos CURE relativos a modelos com FF2 mostraram sistematicamente melhor qualidade de ajuste, quando comparados com os modelos com FF1. Por outro lado, há uma diferença clara entre modelos com FF2 e modelos com FF1 no que se refere aos intervalos de confiança estimados para a média Gama e para o número de despistes. Globalmente, um modelo com FF1 tem intervalos de confiança mais amplos do que um modelo com FF2. Finalmente, é de realçar que, para além de intervalos de confiança mais estreitos obtidos com modelos com FF2, esta forma funcional gera números esperados de despistes e intervalos de confiança para a média Gama (*m*) mais próximos das “médias verdadeiras” e dos “intervalos verdadeiros”, respetivamente.

Em sùmula, é justificada a noção de que a modelação de despistes com uma forma funcional que represente a relação entre a frequência de despistes e o volume de tráfego mediante curva convexa mostra melhor qualidade de ajuste do que uma forma funcional assente no pressuposto de uma frequência de despistes monotonamente crescente com o volume de tráfego.

Uma das tarefas indispensáveis no processo de modelação consiste em verificar se a relação entre as variáveis e a distribuição ou modelo é lógica ou teoricamente correta, a designada "qualidade lógica" [2]. A FF2 demonstra essa "qualidade lógica", desde logo pelo princípio de "zero acidentes para volume de tráfego nulo" e por permitir que a relação entre os despistes e o volume de tráfego médio diário não seja monótona. De facto, o número de despistes aumenta, atinge um pico e diminui à medida que aumenta o TMD, o que implica, por um lado, que frequências de despiste semelhantes possam ocorrer sob condições de tráfego diferentes e, por outro, que em alguns casos o número de despistes possa aumentar quando o TMD diminui.

5 AGRADECIMENTOS

Os autores agradecem o apoio financeiro fornecido pela Fundação para a Ciência e a Tecnologia (FCT) através da bolsa de doutoramento atribuída a Carlos Roque, com a referência SFRH/BD/82228/2011.

6 REFERÊNCIAS

- 1 D. Lord, , S. Washington, , J. Ivan, Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory, *Accident Analysis and Prevention* 37(1), 35–46, 2005.
- 2 E. Hauer, Observational Before-After Studies in Road Safety, *Pergamon*, 1997.
- 3 D. Lord, F. Mannering, The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives, *Transportation Research Part A* 44 , 291–305, 2010.
- 4 C. Roque, J.L. Cardoso, Investigating the relationship between run-off-the-road crash frequency and traffic flow through different functional forms, *Accident Analysis & Prevention*, Volume 63, Pages 121-132, 2014.
- 5 R Development Core Team. R: A language and environment for statistical computing. R *Foundation for Statistical Computing*, Vienna, Austria, 2011.
- 6 C. Roque, Critérios de segurança para a área adjacente à faixa de rodagem na Rede Rodoviária Nacional, Tese de Doutoramento em Sistemas de Transportes, *Instituto Superior Técnico*, Lisboa, 2013.
- 7 AASHTO. Highway Safety Manual 1st Edition, *American Association of State Highway and Transportation Officials*, Washington, DC. 2010.
- 8 W. E. Ricker, Stock and recruitment. *Journal of the Fisheries Research Board of Canada* 11, 559–623, 1954.
- 9 E. Hauer, E.; J. Bamfo, Two tools for finding what function links the dependent variable to the explanatory variables. *Proceedings of the ICTCT 1997 Conference*, Lund, Sweden, 1997.
- 10 G.R.Wood, Confidence and prediction intervals for generalized linear accident models. *Accident Analysis and Prevention* Vol. 37 (2), 267–273, 2005.
- 11 TRB. Special Report 209: Highway Capacity Manual. *Transportation Research Board, National Research Council*, Washington, D.C., 2000.