

# Avaliação da dependência espacial na modelagem do desempenho da segurança viária em zonas de tráfego

Marcos José Timbó Lima Gomes<sup>1</sup>, Caio Assunção Torres<sup>2</sup> e Flávio José Craveiro Cunto<sup>3</sup>

**Resumo:** Uma técnica comumente usada no processo de modelagem do Desempenho da Segurança Viária (DSV), no nível de planejamento, são os Modelos Lineares Generalizados (MLG) assumindo a distribuição binomial negativa para os erros. Limitações dessa técnica, por não considerar os efeitos espaciais, têm sido contornadas com a utilização de modelos espaciais locais a partir de técnicas de regressão espacial como a Regressão de Poisson Geograficamente Ponderada (RPGP). Este trabalho tem por objetivo apresentar uma análise comparativa entre modelos de previsão de acidentes globais não espaciais e locais espaciais para a estimação do DSV agregado ao nível de zonas de tráfego em Fortaleza/CE. Foram calibrados modelos para a variável dependente acidentes totais e acidentes com vítimas e os resultados mostraram que os modelos RPGP apresentaram melhor desempenho que os MLG nas medidas de ajustes e na redução da autocorrelação espacial dos resíduos, sendo capazes de captar a heterogeneidade espacial da frequência dos acidentes.

*Palavras-chave:* modelagem da segurança viária, autocorrelação espacial, modelo espacial local.

**Abstract:** A common technique used in the modeling process of the Road Safety Performance at the planning level is the Generalized Linear Models (GLM) procedure with the assumption of negative binomial error distribution. A main limitation of this technique, which is the no consideration of spatial effects, has been overcome by the use of local spatial models such as the Geographically Weighted Poisson Regression (GWPR). This work aims to present a comparative analysis between non spatial global and spatial local accident prediction models focused to estimate the safety performance of traffic accident zones of Fortaleza city. Models were calibrated to the dependent variable total accidents and accidents with victims and the results showed that GWPR models performed better than GLM on measures of adjustment and the reduction of residual spatial autocorrelation, being able to capture the spatial heterogeneity in the frequency of accidents.

*Keywords:* road safety modeling, spatial autocorrelation, local spatial models.

## 1. INTRODUÇÃO

A segurança viária tem sido cada vez mais considerada como um componente necessário no processo de planejamento de transporte. Os estudos focados no planejamento do desempenho da segurança viária (DSV), segundo Ferreira e Couto (2012), podem ser efetivados ao nível de rede viária ou em áreas de acordo com o nível de agregação dos dados. Estes últimos são geralmente baseados em dados agregados ao nível de zonas de tráfego (ZT) e incluem informação sobre fatores de exposição, características da rede viária, uso do solo, fatores socioeconômicos e demográficos (Aguero-Valverde, 2013).

Uma técnica comum usada para a modelagem do DSV no nível de planejamento é a utilização dos Modelos Lineares Generalizados (MLG) com a hipótese de distribuição dos erros seguindo a distribuição binomial negativa. Assim, os coeficientes dos modelos são estimados para representar a relação média entre a variável dependente, normalmente o número de acidentes por zona de tráfego, e cada variável explicativa. Os parâmetros do modelo são estimados globalmente para toda a área de estudo com o pressuposto de que a relação entre a variável dependente e cada uma das variáveis independentes não varia ao longo da área geográfica (Hadayeghi *et al.*, 2010).

Conforme Matkan *et al.* (2011), a suposição de estacionariedade nos MLG, ou seja, relações de dependência constante sobre o espaço, pode esconder alguns aspectos espaciais que afetam os acidentes, como por exemplo a presença ostensiva de fiscalização ou qualidade do pavimento de forma diferenciada entre regiões da cidade. Dessa forma, a precisão dos modelos para descrever as relações entre os acidentes e as variáveis independentes pode ser questionada. É provável que, enquanto alguns fatores explicativos que têm forte poder de previsão sobre o número de acidentes em uma zona, podem não ser tão suficientemente fortes em outra.

Gomes *et al.* (2015) apresentaram uma análise exploratória das variáveis que contribuem na modelagem da segurança viária ao nível estratégico e tático de planejamento, utilizando os MLG para o município de Fortaleza assumindo distribuição binomial negativa. Os modelos de nível estratégico apresentaram resíduos com valores significativos para a dependência espacial, representado pelo Índice de Moran. Esse índice, medido em áreas vizinhas, pode variar de -1 e 1. Assim, quanto mais próximo de 1, maior a semelhança entre vizinhos. O valor 0 (zero) indica inexistência de dependência espacial e valores negativos indicam dessemelhança. Como recomendações os autores sugeriram a investigação da inserção do caráter espacial na modelagem do DSV a partir de modelos espaciais globais ou locais.

Portanto, para superar essas dificuldades dos MLG, que não captam a heterogeneidade espacial nas relações entre acidentes e as variáveis explicativas, recentemente alguns estudos utilizaram modelos locais espaciais a partir de técnicas de regressão geponderadas ou mais especificamente a Regressão de Poisson Geograficamente Ponderada (RPGP).

<sup>1</sup> Marcos José Timbó Lima Gomes, Departamento de Engenharia de Transportes, UFC. (timbo@det.ufc.br)

<sup>2</sup> Caio Assunção Torres, Departamento de Engenharia de Transportes, UFC. (caioatr@gmail.com)

<sup>3</sup> Flávio José Craveiro Cunto, Departamento de Engenharia de Transportes, UFC. (flaviocunto@det.ufc.br)

Manuscrito recebido em 03/04/2016 e aprovado para publicação em 25/09/2016

.Este artigo é parte de TRANSPORTES v. 24, n. 4, 2016. ISSN: 2237-1346 (online). DOI:10.4237/transportes.v24i4.1110

Assim, este trabalho tem por objetivo apresentar uma análise comparativa entre modelos de previsão de acidentes (MPA), em nível estratégico, dos tipos globais não espaciais e locais espaciais para a estimação do DSV agregados ao nível de zonas de tráfego. Destaca-se que os modelos testados neste estudo permitem a avaliação do DSV em cenários de longo prazo que envolvam mudanças nas características socioeconômicas e demográficas de uma região.

## 2. MODELAGEM ESPACIAL DO DSV

Esforços consideráveis têm sido realizados para desenvolver regressões nas quais a dependência espacial é considerada na modelagem do DSV (Aguero-Valverde, 2013 e Zeng e Huang, 2014). Dentre esses esforços tem-se a regressão espacial tratada de forma global e a regressão espacial tratada de forma local.

Se o fenômeno em análise apresentar forte estacionariedade espacial, modelos de regressão global são mais justificados, assim a estrutura espacial é capturada através de um único parâmetro que é adicionado ao modelo de regressão tradicional, dentre esses modelos destacam-se o *Spatial Error Model* (SEM) e *Spatial Auto Regressive* (SAR). No entanto, essa técnica assume uma estrutura de erro normalmente distribuído na calibração de modelos de regressão e tal suposição não é adequada para dados de contagem, como acidentes de trânsito (Hadayeghi *et al.*, 2003).

Quando a regressão global causa estimativas pouco precisas em alguns pontos, é preferível o uso do modelo de regressão local, este por sua vez analisa as relações entre as variáveis de forma específica para cada área de estudo, através de regressão geponderada. Dentre as técnicas existentes, uma que se adequa ao fenômeno de acidentes de trânsito é a Regressão de Poisson Geograficamente Ponderada desenvolvida por Nakaya *et al.* (2005) que propuseram essa técnica para relacionar o número de mortes (dado de contagem), em zonas, com variáveis socioeconômicas. Os resultados indicaram que houve variações espaciais significativas nas relações entre a variável dependente e algumas das variáveis independentes e que, conseqüentemente, a aplicação de modelos globais tradicionais renderia resultados não satisfatórios. O modelo RPGP proposto é definido pela Equação 1:

$$E(y|X_1, X_2, \dots, X_k) = \exp[\beta_0(u_i) + \beta_1(u_i)X_1 + \beta_2(u_i)X_2 + \dots + \beta_k(u_i)X_k] \quad (1)$$

em que  $E(y|X_1, X_2, \dots, X_k)$  é o valor esperado da variável resposta dado que um conjunto de  $k$  valores das variáveis explicativas é observado. Os  $\beta$ 's são os parâmetros que se deseja estimar em função da localização  $u_i = (u_{ix}, u_{iy})$  que caracteriza as coordenadas do ponto  $i$  (centróides da ZT) no espaço e representa um vetor de duas coordenadas dimensionais que descrevem a localização do elemento  $i$  com coordenadas  $x$  e  $y$ .

A ideia básica da regressão geponderada é que os dados observados próximos ao ponto  $i$  tenham mais influência sobre a estimativa do  $\beta_j(u_i)$ 's de que dados localizados mais longe de  $i$ . Esta influência em torno do ponto  $i$  é descrita por uma função de ponderação. O método RPGP tenta capturar a variação espacial ajustando um modelo de regressão em cada ponto individualmente, ponderando todas as

observações vizinhas por uma função da distância. A função espacial *kernel* e o tamanho da janela, isto é, a largura de banda, usado no processo de montagem do modelo tem um forte impacto nas estimativas dos coeficientes, sendo sua seleção um passo importante para a aplicação de RPGP (Hadayeghi *et al.*, 2010).

Segundo Nakaya *et al.*, (2005) os pesos das observações diminuem gradualmente à medida que a distância entre o local de regressão  $i$  e ponto no qual a observação de ordem  $j$  torna-se maior. Uma escolha clássica do *kernel* de ponderação é o *kernel* Gaussiano definido pela Equação 2.

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{\theta^2}\right) \quad (2)$$

em que  $w_{ij}$  é o valor do peso de uma observação em  $j$  para a estimação do coeficiente em  $i$  e  $d_{ij}$  é a distância euclidiana entre os pontos  $i$  e  $j$ . O parâmetro  $\theta$  (chamado de largura de banda) regula o tamanho do *kernel* e controla a taxa que o peso de um dado ponto  $i$  diminui à medida que se distancia do local do ponto de regressão em análise,  $j$ . Detalhes computacionais para os procedimentos de calibração da RPGP podem ser encontrados em Nakaya *et al.* (2005).

Hadayeghi *et al.* (2010) utilizaram a RPGP para investigar as variações espaciais na relação entre o número de acidentes zonais e potenciais variáveis explicativas de planejamento de transporte para a cidade de Toronto. Desenvolveram e compararam os MPA com RPGP e MLG, assumindo distribuição binomial negativa. Os resultados mostraram que os coeficientes de correlação linear dos modelos melhoraram com a utilização da RPGP, diminuindo assim os erros quadráticos médios dos mesmos em relação aos MLG.

No estudo de Toronto, os autores estimaram a frequência de acidentes totais e acidentes com vítimas, considerando dois tipos de modelos: os que utilizavam somente variáveis relativas às características da rede e do tráfego, e modelos considerando todas as variáveis de planejamento disponíveis, como características do uso do solo, variáveis socioeconômicas, demográficas, características da rede/tráfego. Embora estes últimos modelos considerem todas as variáveis de planejamento disponíveis, o seu potencial uso em outras regiões pode ser limitado.

Diferentemente do estudo de Hadayeghi *et al.* (2010), neste artigo pretende-se desenvolver e comparar modelos exclusivamente em nível estratégico, que estão relacionados a fatores que geram viagens no meio urbano, por isso foram considerados somente variáveis socioeconômicas e demográficas.

O uso de regressão de Poisson em conjunto com RGP, a RPGP, representa um avanço na calibração de modelos de avaliação do DSV em comparação com estudos anteriores que utilizaram a abordagem MLG para a calibração de tais modelos. Ressalta-se que, apesar de a distribuição binomial negativa frequentemente se adequar melhor aos resíduos dos modelos não espaciais, rotinas computacionais dedicadas à estimação de modelos RGP assumindo essa distribuição ainda estão restritas a pequenos grupos de pesquisa em estatística.

O aplicativo GWR4 utilizado neste estudo, não suporta a calibração desse tipo de modelo e como tal, os mesmos não serão calibrados.

### 3. PROCEDIMENTO METODOLÓGICO

Este estudo foi realizado a partir das seguintes etapas sequenciais: i) Análise e seleção das variáveis; ii) Desenvolvimento e avaliação dos modelos; iii) Cálculo de medidas de ajustes e comparação dos modelos; iv) Análise da dependência espacial.

#### 3.1. Análise e seleção das variáveis

Os dados desse estudo são os mesmos usados por Gomes *et al.* (2015) que se utilizaram de uma base georreferenciada de Fortaleza já com informações socioeconômicas atualizadas do Censo 2010 e agregadas em 126 zonas de tráfego, além de informações de acidentes de trânsito obtidas do Sistema de Informações de Acidentes de Trânsito de Fortaleza (SIAT/FOR). A variável dependente corresponde aos acidentes totais e acidentes com vítimas, referente ao ano de 2011 e agregados nas zonas de tráfego.

Nesse trabalho foram utilizadas somente as variáveis destinadas a modelos de nível estratégico, que são modelos cujo objetivo é avaliar políticas públicas para reduzir acidentes de trânsito, sendo as variáveis demográficas, de uso do solo, socioeconômicas e de acessibilidade as mais indicadas para o estudo. A escolha dos modelos no nível estratégico deve-se ao fato que foram os que apresentaram autocorrelação espacial dos resíduos conforme Gomes, *et al.* (2015).

Uma limitação dos dados utilizados é que algumas variáveis importantes não estão disponíveis, tais como volume de tráfego e outras não estão na sua forma pura. Por exemplo, renda está combinada com a variável “número de domicílios”. A Tabela 1 apresenta as estatísticas descritivas das variáveis utilizadas para a modelagem em nível estratégico.

Em seguida foram construídos os gráficos de dispersão e matrizes de correlação entre as variáveis e a partir destes resultados decidiu-se por agregar ou excluir algumas variáveis, com intuito de reduzir a multicolinearidade nas variáveis explicativas. Por exemplo, as duas variáveis de consumo de energia, separadas por tipo de setor, foram agregadas em virtude da alta correlação apresentada, denominada E\_TOT.

Além dessas, as variáveis populações por idade são altamente correlacionadas entre si e apresentam alta correlação com o número de domicílios com renda até três salários. Assim, decidiu-se por considerar apenas a variável domicílio por renda que caracteriza bem a realidade demográfica e de variação da renda entre zonas de tráfego em Fortaleza. A Tabela 2 apresenta a matriz de correlação entre todas as variáveis para o nível estratégico.

#### 3.2. Desenvolvimento e avaliação dos modelos

##### 3.2.1. Modelos Lineares Generalizados

Nos MPA com variáveis agregadas observa-se um direcionamento aos modelos lineares generalizados em virtude da variável resposta ser aleatória, discreta e positiva, além dos erros se apresentarem com distribuição compatível com a distribuição de Poisson ou binomial negativa (Hadayeghi *et al.*, 2003; Lord e Persaud, 2004; Tarko *et al.*, 2008). A formulação do modelo, como proposta na literatura, adotada neste estudo é apresentada na Equação 3.

$$E(y|x_1, x_2, \dots, x_k) = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad (3)$$

em que  $E(y|x_1, x_2, \dots, x_k)$  é o valor esperado da variável resposta dado que um conjunto de  $k$  valores das variáveis explicativas é observado e os  $\beta$ 's são as estimativas dos coeficientes para cada variável independente do modelo.

Gomes *et al.* (2015) desenvolveram MLG como ferramenta de análise exploratória das variáveis sem a intenção de estabelecer relação de causa e efeito. Realizaram um teste inicial para avaliar a distribuição mais adequada para o nível de dispersão encontrado nos dados. Os resultados indicaram que a distribuição binomial negativa representou melhor a superdispersão dos dados quando comparada com a distribuição de Poisson. Os parâmetros dos modelos foram estimados pelo método da máxima verossimilhança (Hauer, 2002; Lord, 2006). A Tabela 3 apresenta os coeficientes e erros padrões para os modelos MLG.

Tabela 1. Estatística descritiva das variáveis

Grupos	Variáveis	Descrição	Média	Mínimo	Máximo	Desvio Padrão
Variáveis respostas	ACT_11	Total de Acidentes	140,4	2,0	934,0	152,5
	ACV_11	Total de Acidentes com Vítimas	46,5	1,0	256,0	39,2
Variáveis explicativas	E_IND	Consumo de Energia – Indústria	12.576	19	105.442	14.096
	E_CO_SE	Consumo de Energia – Comércio e Serviço	271.096	6.414	4.017.135	517.957
	PO_0_19	População entre 0 e 19 anos	6.080	322	42.260	6.278
	PO_20_64	População entre 20 e 64 anos	11.854	736	68.335	9.957
	PO_M_64	População maior que 64 anos	1.267	81	4.622	866
	D_A_3SM	Domicílios com renda até 3 Salários	3.779	201	25.893	3.783
D_M_3SM	Domicílios com renda acima de 3 Salários	1.087	31	7.015	1.102	

Tabela 2. Correlações das variáveis de nível estratégico

	ACT_11	ACV_11	E_TOT	PO_0_19	PO_20_64	PO_M_64	D_A_3SM	D_M_SM
ACT_11	1,00							
ACV_11	0,90	1,00						
E_TOT	0,78	0,62	1,00					
PO_0_19	0,05	0,33	-0,01	1,00				
PO_20_64	0,18	0,43	0,08	0,98	1,00			
PO_M_64	0,46	0,60	0,28	0,73	0,85	1,00		
D_A_3SM	0,03	0,34	-0,04	0,99	0,97	0,74	1,00	
D_M_3SM	0,69	0,50	0,54	0,12	0,29	0,57	0,05	1,00

Segundo Gomes *et al.* (2015), os resultados dos MLG corroboram a hipótese de que zonas com maior densidade demográfica, com população de maior poder aquisitivo e maior intensidade de atividade econômica apresentam maior ocorrência de acidentes de trânsito totais e com vítima. O fato da variável “domicílios com renda até 3 salários” apresentar efeito significativo apenas nos modelos de acidentes com vítimas indica que zonas de menor poder aquisitivo apresentam uma eventual maior severidade dos acidentes que pode estar associada a maior intensidade de deslocamentos não motorizados ou estrutura viária precária.

### 3.2.2. Modelos de Regressão de Poisson Geograficamente Ponderados

Para estimar os modelos RPGP utilizou-se o *software GWR4* com função *kernel* do tipo fixo Gaussiano e a largura da banda obtida a partir da opção *Golden Section Method*, que otimiza a determinação do tamanho da largura da banda.

A largura de banda pode ser considerada como um parâmetro de suavização: quanto maior a banda, maior a suavização dos coeficientes locais; por sua vez, quanto menor a banda, mais heterogeneidade nas respostas se terá, pois menos observações serão usadas ao redor do ponto de calibragem. A largura de banda envolve um equilíbrio entre viés e variância, sendo que a escolha de uma largura de banda muito pequena leva a uma variância grande nas estimativas locais, enquanto que uma banda muito larga traz viés às estimativas locais (Fortheringham *et al.*, 2002).

Para a RPGP foram calibrados os mesmos modelos obtidos pelos MLG, visto que o objetivo é realizar a comparação entre eles. A Tabela 4 mostra os modelos de RPGP calibrados para as variáveis dependentes “acidentes totais” e “acidentes com vítimas” que apresentam um conjunto de parâmetros para cada variável independente em cada zona de tráfego, sendo mostrado os valores mínimo, máximo, 1º quartil, mediana e 3º quartil dos coeficientes.

A Tabela 4 também apresenta o valor do AIC (critério de informação de Akaike) para RPGP e para o modelo global de Poisson. O AIC pode ser aplicado na comparação entre possíveis modelos para um mesmo conjunto de dados.

Para modelos com elevado número de parâmetros a variância será baixa, entretanto o modelo será viesado, ou seja, não poderá ser utilizado em outra situação sem produzir grandes erros. Por outro lado, com poucos parâmetros o modelo terá um pequeno viés, porém com elevada variância. Procura-se então equilibrar viés/variância escolhendo o modelo com o menor AIC (Akaike, 1973). Nesse caso, observa-se que em comparação ao modelo global de Poisson, o RPGP possui um menor AIC em ambos os modelos calibrados.

Observando a coluna “Mediana” da Tabela 4 percebe-se que os coeficientes estimados no modelo geponderado são muito próximos aos valores determinados na regressão binomial negativa. Entretanto, observando as colunas “Mínimo” e “Máximo” da Tabela 4 nota-se que os sinais dos coeficientes das duas variáveis domicílio por renda podem ser negativos em algumas zonas de tráfego. As variáveis desses modelos estão diretamente relacionadas à produção e atração de viagens entre zonas, e são *proxy* para o fluxo de tráfego, portanto esperava-se um impacto positivo sobre o número total de acidentes.

No estudo de Hadayeghi *et al.* (2010), também foi observado sinais inesperados para algumas variáveis e conforme os autores Zhao *et al.* (2005), Wheeler e Calder (2007) e Almeida (2012), o problema com os sinais inesperados é comum nos modelos de regressão geponderados e existem duas razões possíveis para isso:

1. Um coeficiente local está correlacionado como outro ou com uma combinação linear de outros coeficientes locais. Não existe correlação entre as variáveis explicativas em nível global, mas pode existir localmente e atualmente não há no *software GWR4* uma ferramenta de diagnóstico disponível para examinar essa multicolinearidade do modelo, não podendo ser verificado.
2. Algumas variáveis podem ser menos significativas ou não significativas em todas ou em algumas zonas. No relatório de saída do *software GWR4* não se tem o resultado da significância do coeficiente, mas tem-se o valor da estatística *t* observada em cada zona de tráfego.

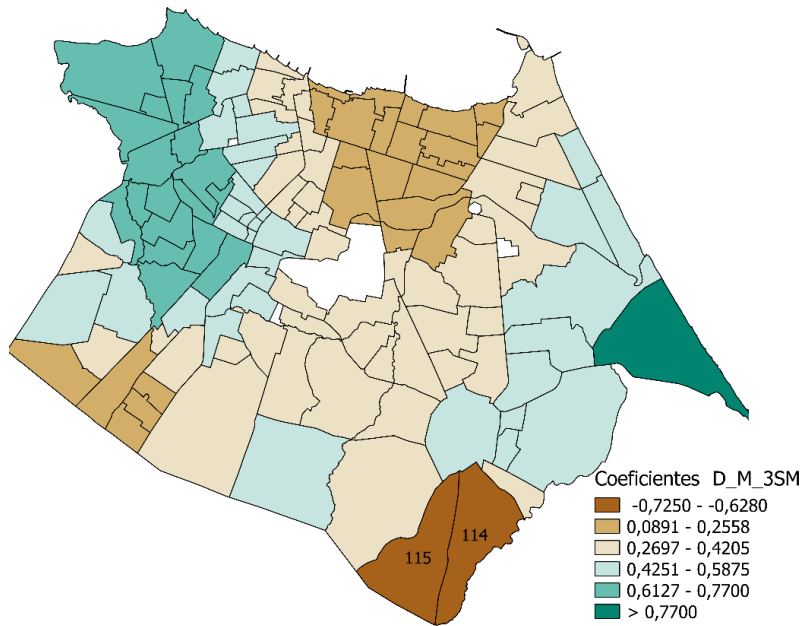
**Tabela 3.** Modelos de previsão de acidentes - MLG

Fonte: Adaptado de Gomes *et al.*, 2015

Parâmetro	Acidentes totais		Acidentes com vítimas	
	$\beta$	EP	$\beta$	EP
Constante	4,077	0,088	3,00	0,097
E_TOT ( $10^7 Wh$ )	0,601	0,014	0,495	0,012
D_A_3SM (1000 dom.)			0,078	0,015
D_M_3SM (1000 dom.)	0,427	0,067	0,251	0,061
AIC	1418,2		1151,7	

**Tabela 4.** Modelos RPGP para acidentes totais e acidentes com vítimas

Parâmetro	Acidentes Totais					Acidentes com Vítimas				
	Mín	Máx	1ºQuartil	Mediana	3ºQuartil	Mín	Máx	1ºQuartil	Mediana	3ºQuartil
Constante	2,66	5,13	3,79	4,11	4,46	1,33	3,50	2,82	3,05	3,21
E_TOT ( $10^7 Wh$ )	0,14	4,84	0,38	0,60	1,11	0,04	4,30	0,37	0,52	0,92
D_A_3SM (1000 dom.)						-0,0002	0,25	0,05	0,08	0,12
D_M_3SM (1000 dom.)	-0,72	1,06	0,28	0,39	0,52	-0,77	0,89	0,16	0,27	0,38
AIC RPGP	3.539,4					1.109,0				
AIC Global	7.382,4					2.017,8				



**Figura 1.** Variação espacial dos coeficientes significativos para a variável “domicílio com renda maior que 3SM” para o modelo acidentes totais

O teste-t para os parâmetros foi realizado para todos os coeficientes estimados com o objetivo de determinar em quais zonas as variáveis relacionadas a renda por domicílio eram significativas com 95% de confiança. A Figura 1 mostra a distribuição espacial dos coeficientes significativos para a variável “domicílios com renda maior que 3SM” para o modelo de acidentes totais.

Todos os coeficientes dessa variável foram significativos e pela Figura 1 observa-se que os coeficientes com sinal negativo ocorreram somente nas zonas de nº 114 e 115, localizadas no sul do município. Uma hipótese para estes sinais alterados pode ser o fato de se utilizou somente variáveis pertencente às zonas de tráfego, entretanto não existe uma variável que incorpore o tráfego de passagem.

Uma considerável extensão do limite entre as zonas de nº 114 e 115 corresponde a rodovia BR-116, que possui um intenso tráfego de veículos com origem ou destino na Região Metropolitana de Fortaleza e esse tráfego de passagem pode estar produzindo um elevado número de acidentes, que faz com que o modelo tente compensar com alteração dos sinais dos coeficientes.

As Figuras 2a e 2b mostram a variação espacial dos coeficientes para as variáveis “domicílio com renda até que 3SM” e “domicílio com renda maior que 3SM” do modelo acidentes com vítimas, respectivamente. Na Figura 2a, para a variável “domicílio com renda até 3SM”, a única zona que teve um coeficiente de sinal negativo e foi não significativa com 95% de confiança foi a zona 117. Já para a Figura 2b, que trata da variável “domicílio com renda maior que 3SM”, houve quatro zonas (28, 115, 125, 126) cujos coeficientes apresentaram sinais negativos e não foram significativos.

O que chama atenção é que a maioria das zonas que apresentaram coeficientes com sinais negativos e não significativos são zonas próximas de fronteiras, sendo uma hipó-

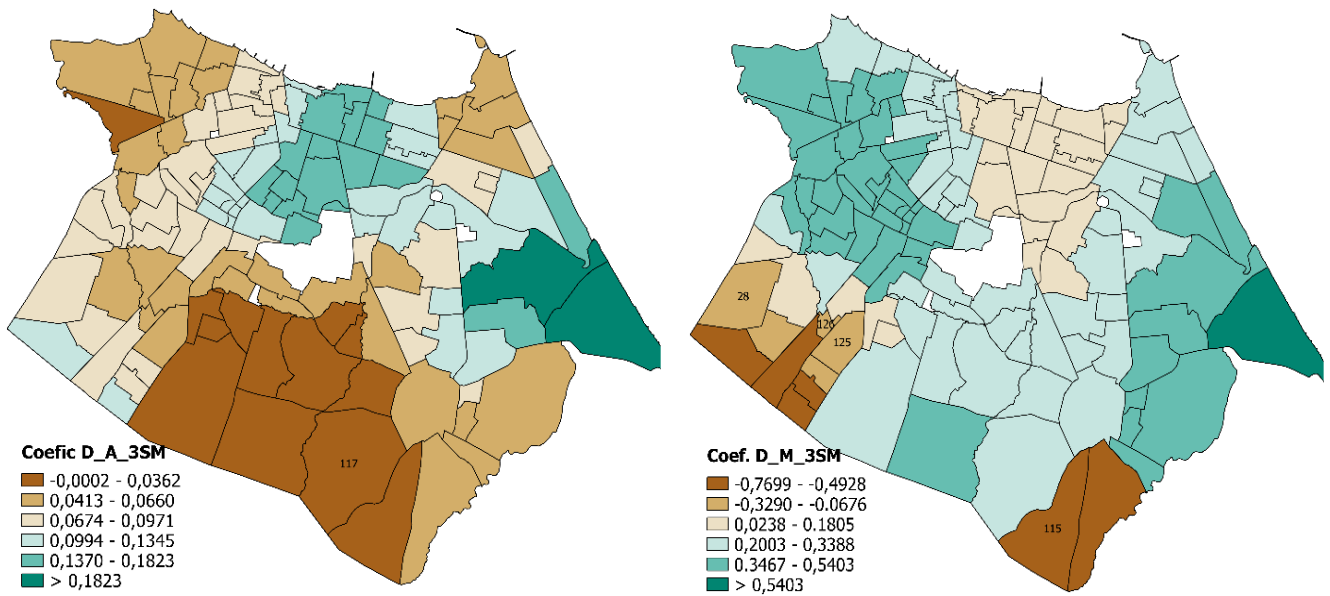
tese para a ocorrência desse fato a utilização do *kernel* espacial do tipo fixo. Segundo Fotheringham *et al.* (2002), a desvantagem da especificação de pesos com largura fixa de banda é que o número de observações dentro da subamostra pode variar consideravelmente e em algumas regiões, próximas da fronteira da área de estudo, a estimação dos coeficientes locais pode ser feita com poucas observações, acarretando um problema de eficiência dos coeficientes.

### 3.3. Medidas de Ajustes

Para quantificar e comparar os desempenhos de modelos desenvolvidos é recomendado a utilização de algumas métricas (Oh *et al.*, 2003). Dentre essas, destacam-se o coeficiente de correlação de Pearson ( $r$ ) e o Erro Quadrático Médio (EQM). O coeficiente de correlação de Pearson é uma medida da relação linear entre duas variáveis, neste caso a frequência de acidentes observada ( $Y_i$ ) e estimada ( $\hat{Y}_i$ ). Um coeficiente ( $r$ ) de valor igual a 1 indica que o modelo prediz perfeitamente os dados observados e é definida conforme a Equação 4.

$$r = \frac{\sum_{i=1}^n (Y_i - Y_{méd}) (\hat{Y}_i - \hat{Y}_{méd})}{\left[ \sum_{i=1}^n (Y_i - Y_{méd})^2 \cdot \sum_{i=1}^n (\hat{Y}_i - \hat{Y}_{méd})^2 \right]^{0,5}} \quad (4)$$

em que  $Y_{méd}$  é o valor médio da frequência de acidentes observados e o  $\hat{Y}_{méd}$  é o valor esperado da frequência de acidentes estimada pelos modelos. O EQM é a soma dos quadrados das diferenças entre as frequências de acidentes estimadas e observadas divididas pelo tamanho da amostra ( $n$ ), como é definido na Equação 5.



**Figura 2.** (a) Variação espacial dos coeficientes para a variável “domicílio com renda até que 3SM” para o modelo de acidentes com vítimas, (b) Variação espacial dos coeficientes para a variável “domicílio com renda maior que 3SM” para o modelo de acidentes com vítimas.

Quanto menor for os valores dessa métrica, melhor é a previsão do modelo e um valor 0 indica que o modelo prevê os dados observados perfeitamente.

$$EQM = \sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n} \quad (5)$$

As métricas foram utilizadas para avaliar o desempenho dos modelos RPGP em relação aos MLG com distribuição binomial negativa, medindo a qualidade de ajuste dos modelos. A Tabela 5 apresenta os valores dessas métricas para os modelos calibrados permitindo a avaliação objetiva dos seus ajustes aos dados. Os modelos de RPGP apresentaram um valor consideravelmente melhor para o coeficiente de correlação de Pearson e uma redução significativa no valor do EQM em relação ao MLG com binomial negativa, confirmando que a RPGP possui uma melhor performance em relação aos MLG.

**Tabela 5.** Comparação das medidas de ajuste para os modelos calibrados

Modelos	Acidentes Totais		Acidentes com Vítimas	
	r	EQM	r	EQM
Binominal Negativa	0,56	185.788	0,41	2.578
RPGP	0,90	4.462	0,86	399

### 3.4. Análise da Dependência Espacial

A análise comparativa final da eficiência dos modelos globais e locais foi realizada a partir da verificação da autocorrelação espacial dos resíduos dos modelos calibrados, calculando-se o Índice Global de Moran, para isso utilizou-se o *software* Geoda. A Tabela 6 mostra a comparação dos valores da autocorrelação espacial dos modelos calibrados.

Analisando os dados da Tabela 6, observa-se pelo valor-p que os modelos globais apresentaram valores da autocorrelação espacial significativos, como já identificado por Gomes, *et al.* (2015) e os modelos locais foram capazes de eliminar a dependência dos resíduos, visto que o índice de

Moran não foi significativo para o nível de confiança de 95%. Esses testes sugerem que a RPGP conseguiu reduzir a heterogeneidade espacial dos dados, produzindo resíduos com autocorrelação espacial reduzida em comparação aos MLG.

**Tabela 6.** Índice Global de Moran para os modelos calibrados

Modelos	Acidentes Totais		Acidentes com Vítimas	
	Índice de Moran	Valor-p	Índice de Moran	Valor-p
Binomial Negativa	0,216	0,005	0,206	0,002
RPGP	0,068	0,088	0,002	0,414

## 4. CONSIDERAÇÕES FINAIS

Nesse estudo foi realizada uma comparação de ajustes de modelos não espaciais do tipo MLG com erros seguindo a distribuição binomial negativa, com modelos espaciais locais do tipo Regressão de Poisson Geograficamente Ponderado (RPGP) aplicados ao desempenho da segurança viária no município de Fortaleza.

Os modelos calibrados, para determinar a frequência de acidentes totais e com vítimas, aplicam-se ao nível estratégico de planejamento e utilizaram variáveis explicativas do tipo socioeconômicas e demográficas. Embora os modelos RPGP tenham exibido alterações nos sinais dos coeficientes de algumas zonas de tráfego, as medidas de comparação desses modelos apresentaram melhor desempenho que os convencionais MLG.

Foi verificado também a autocorrelação espacial dos resíduos dos modelos globais e locais, visto que essa análise é uma etapa importante no diagnóstico do modelo e busca indícios de ruptura dos pressupostos de independência.

A análise dos resíduos mostrou que os modelos RPGP reduziram de forma significativa a dependência espacial e isso deve-se ao fato que os modelos locais foram capazes de captar a heterogeneidade espacial da frequência dos acidentes.

Outra vantagem dos modelos geponderados é que eles permitem que os resultados sejam mapeáveis, o que torna mais fácil a visualização e a busca de tendências ou diferenças nas variáveis explicativas.

Uma limitação deste trabalho é o uso de apenas um ano (2011) da base de dados de acidentes, além de um número limitado de variáveis utilizadas nos modelos e ausência de variáveis importantes nos bancos de dados existentes, como o fluxo veicular. Recomenda-se ampliar o espaço amostral do estudo para permitir a execução de procedimentos de validação que possam confirmar o bom desempenho dos modelos em um conjunto de dados independente. Outra limitação é que o *software GWR4* não fornece a significância dos parâmetros, sendo o teste de significância, neste estudo, realizado utilizando o teste *t* de *Student* clássico. Uma abordagem mais adequada para esse teste de significância, porém mais complexa, foi parcialmente resolvido por Silva e Fotheringham (2016) através de testes *t* múltiplos.

Um cuidado importante que o analista deve se preocupar é com uma análise exploratória das variáveis, anteriormente à construção dos modelos geponderados. Isso porque esses modelos são suscetíveis a problema de multicolinearidade, sendo a baixa correlação entre as variáveis explicativas, muito importantes para a estabilidade do modelo.

Apesar da abordagem promissora, a literatura aponta problemas com a regressão geponderada, recomendando que se deve ter cuidados com a interpretação dos resultados obtidos (Almeida, 2012). Alterações de sinais dos coeficientes nesta pesquisa pode ter ocorrido em decorrência de: i) número de observações reduzidas em zonas de fronteiras da área de estudo; ii) limitações dos modelos calibrados no que diz respeito ao tráfego de passagem por não ser capturado por nenhuma variável utilizada; iii) tamanho das zonas, uma vez que a heterogeneidade espacial dos acidentes não foi considerada na divisão das zonas.

Em relação as dificuldades ainda existentes em alguns municípios brasileiros, quando se trata do processo de coleta/processamento dos acidentes de trânsito e disponibilidade de informações básicas da oferta e demanda de transportes, houve nos últimos anos uma disseminação de dados espacializados pelas instituições públicas no Brasil. Com esses dados disponíveis, a incorporação dos efeitos espaciais pelos modelos do tipo RGP apresentam melhorias na estimação dos parâmetros e na avaliação do DSV em nível de planejamento estratégico. Tais modelos permitem avaliar o impacto de políticas de segurança viária a longo prazo de forma diferenciada para cada zona da cidade. Uma desvantagem desses modelos, entretanto, é que os mesmos não são transferíveis espacialmente, já que se tem um coeficiente para cada zona de uma região geográfica específica, sendo necessário que cada município desenvolva seus modelos.

Especificamente, uma desvantagem da Regressão de Poisson Geograficamente Ponderada é a limitação de não modelar dados com superdispersão, como o caso dos acidentes de trânsito, sendo mais adequada para isso a Regressão Binomial Negativa Geograficamente Ponderada – RBNGP. O tratamento matemático para a RBNGP apesar de formalmente desenvolvido por Silva e Rodrigues (2014), somente foi implementado em algoritmo do aplicativo SAS em 2016 (Silva e Rodrigues, 2016), não sendo possível a sua inclusão durante a execução desse experimento.

## AGRADECIMENTOS

Os autores agradecem o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de produtividade em pesquisa do terceiro autor, a Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pela bolsa de doutorado do primeiro autor e a Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) pela bolsa de mestrado do segundo autor.

## REFERÊNCIAS

- Aguero-Valverde, Jonathan, (2013). Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. *Accident Analysis & Prevention*, v. 50, p. 289-297. <http://dx.doi.org/10.1016/j.aap.2012.04.019>.
- Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In: B. N. PETROV and F. CSAKI, eds. *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, p. 267–281.
- Almeida, E. (2012). *Econometria Espacial Aplicada*. Editora Alínea, Campinas, SP.
- Ferreira, S. e Couto, A. (2012). Avaliação da Segurança Rodoviária em Fase de Planejamento: Modelo Estatístico de Resposta Qualitativa. *Transportes*, v.20, n.2, p. 48-56. DOI: 10.4237/transportes.v20i2.548.
- Fotheringham, A.S., Brunson, C., Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*. Wiley, New York.
- Gomes, M. J. T. L., Torres, C. A., Oliveira Neto, F. M. e Cunto, F. J. C. (2015). Análise exploratória para a modelagem da frequência de acidentes de trânsito agregados ao nível de zonas de tráfego. *Transportes*, v.23, n.4, p. 42-50. DOI: 10.14295/transportes.v23i4.927.
- Hadayeghi, A., Shalaby, A.S., Persaud, B. (2003). Macro-level accident prediction models for evaluating safety of urban transportation systems. *Transportation Research Record*, v. 1840, n. 1, p. 87-95. DOI: 10.3141/1840-10.
- Hadayeghi, A., Shalaby, A. and Persaud, B. (2010). Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accident Analysis and Prevention*, v. 42, p. 676-688. DOI: 10.1016/j.aap.2009.10.016.
- Hauer, E (2002). *Observational Before-after Studies in Road Safety*. (1a ed.). Pergamon. Toronto, Canadá.
- Lord, D. and Persaud, B. N. (2004). Estimating the safety performance of urban road transportation networks. *Accident Analysis and Prevention*, v. 36, n. 4, p. 609-620. DOI: 10.1016/S0001-4575(03)00069-1.
- Lord, D. (2006). Modeling motor vehicle crashes using poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis and Prevention*, v. 38, n. 4, p. 751-766. DOI: 10.1016/j.aap.2006.02.001.
- Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J. (2003). Validation of FHWA crash models for rural intersection lessons

learned. *Transportation Research Record*, v. 1840, p. 41–49. DOI: 10.3141/1840-05.

Matkan, A. A., Mohaymany, A. S., Mirbagheri, B., SHAHRI, M. (2011). Explorative spatial analysis of traffic accidents using GWPR model for urban safety planning. 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA.

Nakaya, T., Fotheringham, A.S., Brunson, C. e Charlton, M. (2005). Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine*, v.24, p. 2695–2717. DOI: 10.1002/sim.2129.

Silva, A. R. e Fotheringham, A. S. (2016). The Multiple Testing Issue in Geographically Weighted Regression. *Geographical Analysis*. v.48, p.233-247. DOI: 10.1111/gean.12084.

Silva, A. R. da, Rodrigues, T. C. V. (2014). Geographically weighted negative binomial regression-incorporating overdispersion. *Statistics and Computing*, v.24, p. 769-783. DOI 10.1007/s11222-013-9401-9.

Silva, A. R. da, Rodrigues, T. C. V. (2016). A SAS® macro for geographically weighted negative binomial regression. Disponível em: <<http://support.sas.com/resources/papers/proceedings16/8000-2016.pdf>>. Data de acesso: 01/06/2016.

Tarko, A. P., Inerowicz, M., RamOS, J. and Li, W. (2008). Tool with road-level crash prediction for transportation safety planning. *Transportation Research Record: Journal of Transportation Research Board*, v. 2083, n. 1, p. 16-25. DOI: 10.3141/2083-03.

Wheeler, D., Calder, C. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, v.9, p. 145–166. DOI: 10.1007/s10109-006-0040-y.

Zeng, Q. e Huang, H. (2014). Bayesian spatial joint modeling of traffic crashes on an urban road network. *Accident Analysis & Prevention*, v. 67, p. 105-112. DOI: 10.1016/j.aap.2014.02.018.

Zhao, F., Chow, L., LI, M., LIU, X., 2005. A Transit ridership model based on geographically weighted regression and service quality variables. Report DO97591, Lehman Center for Transportation Research, Department of Civil and Environmental Engineering, Florida International University.