

ANÁLISE ESTATÍSTICA E DA DISTRIBUIÇÃO ESPACIAL DOS ACIDENTES DE TRÂNSITO NA ZONA SUL DA CIDADE DO RIO DE JANEIRO

M.M.Rocha e C.D.Nassi

RESUMO

O Brasil é atualmente um dos maiores produtores mundiais de veículos, constando como um dos cinco países com maior índice de acidentes de transportes terrestres. O presente artigo emprega técnicas de análise espacial na visualização, exploração e modelagem de acidentes de trânsito na zona sul do Rio de Janeiro para os anos de 2008 a 2011. Empregam-se os mapas coropléticos e de densidade de Kernel na visualização dos acidentes de trânsito e na análise exploratória espacial geram-se os mapas coropléticos a partir dos *Moran scatterplots* com a finalidade de detectar aglomerados espaciais de acidentes. Por fim, aplicam-se técnicas de análise espacial na geração de variáveis explicativas a serem utilizadas em uma modelagem de regressão múltipla. Obteve-se como variáveis mais explicativas a quantidade de trechos de ruas e o comprimento de ruas do tipo arterial primária presentes nas zona de tráfego.

1 INTRODUÇÃO

Atualmente, o Brasil consta como um dos maiores produtores de veículos do mundo, com uma produção que vem crescendo ano após ano. Por outro lado, o Brasil consta como o 5º país com maior quantidade de acidentes de trânsito no mundo, o que vem causando grandes prejuízos materiais e sociais ao país.

Tendo em vista a importância do assunto, a Organização das Nações Unidas elegeu o período de 2011-2020 como sendo a Década de Ação pelo Trânsito Seguro, na qual governos de todo o mundo se comprometem a tomar novas medidas para prevenir os acidentes no trânsito. Em 2011, o governo brasileiro lançou o Projeto Vida no Trânsito buscando, dentre outros objetivos, identificar fatores de riscos e grupos de vítimas mais vulneráveis a acidentes de transportes terrestres (ATT).

O presente artigo procura trazer recursos de análise espacial na compreensão da distribuição espacial dos acidentes na cidade do Rio de Janeiro, bem como na geração de novas variáveis explicativas que poderão ser empregadas na modelagem dos acidentes de trânsito.

2 ANÁLISE ESPACIAL DE ACIDENTES DE TRÂNSITO

Os estudos sobre os acidentes de trânsito tradicionalmente costumam associar as ocorrências dos mesmos às informações contidas nos boletins de ocorrência policial, tais como características das vias onde ocorreu o acidente, dia da semana, horário, características dos envolvidos nos acidentes, como sexo, idade e se estavam alcoolizados

ou não. Tais informações costumam ser armazenadas em planilhas ou tabelas de banco de dados e posteriormente empregadas em modelagens estatísticas.

Nos últimos anos, vem-se empregando técnicas de análise espacial na visualização e detecção de regiões com grande concentração de acidentes de trânsito (Anderson, 2009), bem como na modelagem de acidentes. Neste caso, vem-se empregando tais técnicas na geração de variáveis explicativas, como é o caso daquelas associadas às características das vias tais como comprimento e número de interseções dentro de determinadas regiões de interesse (Miranda-Moreno *et al*, 2011; Marshall *et al*, 2010; Noland *et al*, 2004). Rifaat *et al*. (2011), por exemplo, mostrou o efeito da distribuição espacial das vias na gravidade de acidentes na cidade de Calgary (Canadá), onde dividiu os padrões de ruas em quatro categorias: grade (*grid-iron*), paralela fragmentada (*fragmented parallel*), paralela deformada (*warped parallel*), curvilínea (*loops and lollipops*) e padrões mistos (*mixed patterns*), representados visualmente pelas Fig 1a, 1b, 1c, 1d e 1e, respectivamente. A abordagem espacial também pode ser empregada modelagem dos acidentes, como é o caso de Ha e Thill (2011), os quais empregaram um modelo de regressão SAR (*Spatial Autoregressive*) na análise de áreas de risco de acidentes.

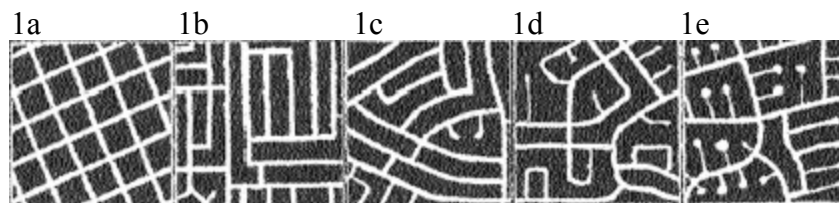


Fig 1 Classificação do padrão de ruas

Fonte: Southworth e Ben-Joseph (2003) *apud* Rifaat *et al*. (2011).

2.1 Análise espacial

A análise espacial apresenta três componentes: a visualização, a análise exploratória dos dados espaciais e a modelagem. A visualização mostra os padrões de interesse, a análise exploratória identifica os padrões e a modelagem explica os padrões de interesse (Anselin, 2002). Para melhor compreender a distribuição espacial dos acidentes de trânsito serão geradas dois tipos de visualização: mapas coropléticos e mapas produzidos a partir da estimativa da densidade de Kernel. A análise exploratória será realizada por meio da visualização do Moran *scatterplot* e a modelagem a partir de utilização de uma regressão linear multivariada.

2.2.1 Estimativa da densidade de Kernel

A estimativa de densidade de Kernel corresponde à estimativa da intensidade de ocorrência do processo pontual em toda a região de estudo, que é feita pelo ajuste de uma função bidimensional a cada um dos eventos considerados, onde se associa a intensidade dos eventos por unidade de área a uma superfície bidimensional (Druck *et al*, 2004). Esta função realiza uma contagem de todos os pontos em uma região de influência, atribuindo-se um peso a cada ponto dentro da região em função de sua distância em relação ao evento considerado. Os parâmetros básicos do estimador Kernel, são o raio de influência τ ($\tau \geq 0$) que define a vizinhança do ponto a ser interpolado e controla a "suavização" da superfície gerada e uma função de estimação com propriedades de suavização do fenômeno (Fig 2).

O raio de influência define a área centrada no ponto de estimação s_i que indica quantos

eventos contribuem para a estimativa da função intensidade. Um raio muito pequeno irá gerar uma superfície muito descontínua; se for grande demais, a superfície poderá ficar muito suave.

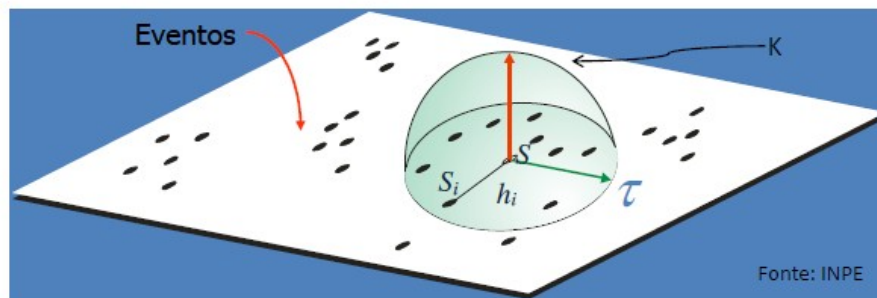


Fig 2 Determinação da estimativa de Kernel

Fonte: Druck et al (2004).

Supondo que s_1, s_2, \dots, s_i são localizações de n eventos observados em uma região A e que s represente uma localização genérica cujo valor queremos estimar, a forma geral da função densidade de Kernel estimada para qualquer ponto s_i é igual a:

$$k\left(\frac{(s - s_i)}{\tau}\right) \quad (1)$$

Onde:

$$\int k(s_i) ds_i \quad (2)$$

A função de Kernel k é comumente escolhida de tal modo que seja unimodal com valor máximo em 0.

No momento em que se for representar espacialmente o estimador de Kernel em meio digital, costuma-se adotar a estrutura de dados matricial, onde o tamanho de cada célula (pixel) pode ser escolhido de acordo com a escala de trabalho. Nesse caso, o valor do estimador de intensidade de Kernel passa então a ser associado a cada pixel do arquivo matricial.

2.2.2 Mapa coroplético

O mapa coroplético é um tipo de representação de mapa temático no qual se associa diferentes tons de cores para representar a intensidade do fenômeno em estudo, comumente empregada para a representação de feições de áreas (Dent, 1985). O critério adotado na construção da legenda no contexto desta pesquisa será o do quartil, o qual agrupa em uma classe da legenda a mesma quantidade de zonas de tráfego. Aqui serão construídos dois tipos de mapas coropléticos das zonas de tráfego: um representando os valores da densidade de acidentes nas zonas de tráfego e outro com a visualização do Moran *scatterplot*.

O Moran *scatterplot* compara o valor do desvio do valor da variável em relação à média global com o valor da média móvel espacial. O desvio da média global está representada pela Equação 3.

$$Z_i = y_i - y_{med} \quad (3)$$

Sendo y_i o valor da variável, no caso acidentes de trânsito, em cada uma das zonas de tráfego e y_{med} a média aritmética da quantidade de acidentes de todas as zonas de tráfego.

A média móvel espacial em para cada região i pode ser representada pela Equação 4 (Bailey e Gatrell, 1995).

$$\hat{\mu}_i = \frac{\sum_{j=1}^n w_{i,j} y_j}{\sum_{j=1}^n w_{i,j}} \quad (4)$$

Onde $i=1, 2, \dots, n$, y_i é o valor do atributo na área i , n é o número de áreas e w_{ij} é o elemento da matriz de vizinhança que representa as relações existentes entre os polígonos das zonas de tráfego. Costuma-se adotar o valor de 1 quando existe relação entre os polígonos e 0 quando não existe relação entre os mesmos. Pode-se adotar diversos critérios para a construção da matriz de proximidade. No contexto desta pesquisa, adotou-se o critério da rainha, onde se atribui valores diferentes de 0 quando os polígonos possuem lados ou vértices em comum e 0 quando não se tem lados nem vértices comuns.

O Moran *scatterplot* apresenta os resultados em quatro classes: Alto-alto, Alto-baixo, Baixo-baixo e Baixo-alto nos quadrantes de I a IV, respectivamente, referindo-se ao valores de Z_i e WZ_i , nessa ordem (Anselin, 1996). A Fig 3 mostra o Moran *scatterplot* da densidade de acidentes nas zonas de tráfego no ano de 2011, onde cada um dos pontos representa uma zona de tráfego. Tomando-se um ponto no quadrante I (lado superior direito) como exemplo, pode-se afirmar que possui densidade de acidentes acima da média global e acima da média local.

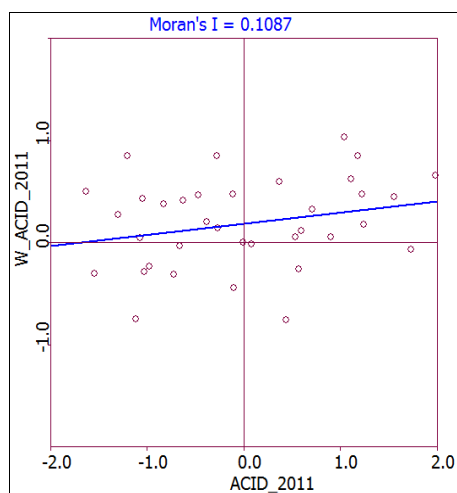


Fig 3 Moran scatterplot dos acidentes em 2011

3 ÁREA DE TRABALHO

A área de trabalho utilizada na pesquisa foi a região da Zona Sul da cidade do Rio de

Janeiro, onde se adotou uma divisão da mesma em 36 (trinta e seis) zonas de tráfego, estas produzidas pela Secretaria Estadual de Transportes do Rio de Janeiro. Os bairros e as zonas de tráfego neles contidas seguem conforme a Fig 4.

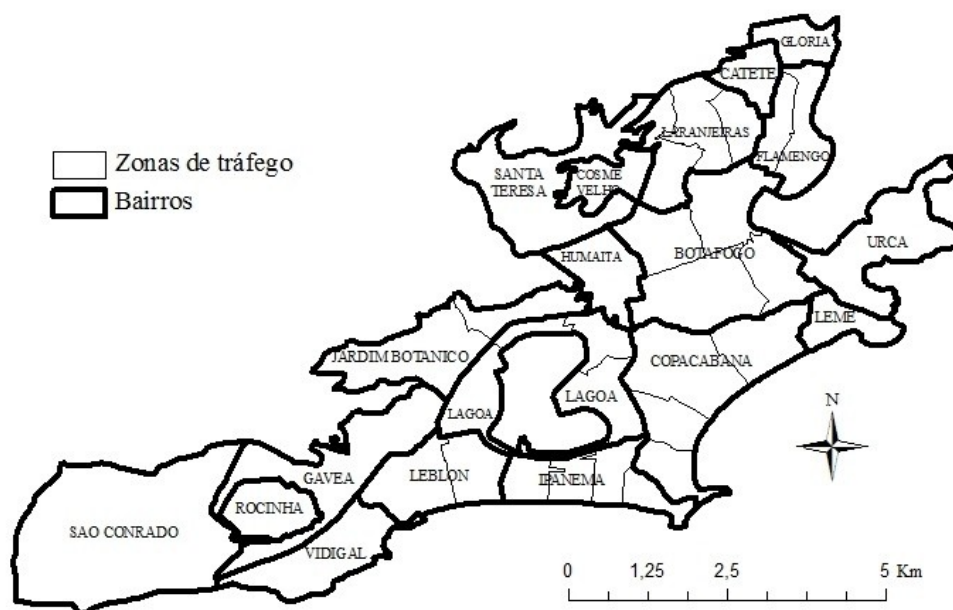


Fig 4 Bairros da Zona Sul do Rio de Janeiro

4 DADOS E PROGRAMAS COMPUTACIONAIS EMPREGADOS

4.1 Programas computacionais

Os programas computacionais empregados na análise exploratória espacial foram o aplicativo de SIG ArcGIS10, o Terraview e o GeoDa; na geração de variáveis explicativas, mais especificamente de característica das vias, foram os aplicativos ArcGIS10, *Google Maps* e *Google Street Map* e o programa empregado na modelagem estatística e seleção de variáveis foram o programa estatístico R.

4.2 Dados de acidente

Os dados de acidente aqui empregados são dados de acidentes de trânsito registrados pela Polícia Militar do Estado do Rio de Janeiro no período de 2008 a 2011, obtidos por meio da chamada de emergência 190, sendo eles do seguinte tipo: colisão entre veículos com e sem vítima, tombamento de veículo com e sem vítima, capotamento com e sem vítima e colisão entre veículos e pedestres (atropelamento) com e sem vítima.

4.3 Variáveis explicativas

As variáveis explicativas aqui empregadas estão divididas em dois grupos: as variáveis estatísticas e as variáveis associadas às características das vias. A informação de referência utilizada na obtenção destas variáveis foi a malha de ruas da cidade do Rio de Janeiro na escala de 1:10000, produzida pelo Instituto Pereira Passos (IPP), órgão da Prefeitura do Rio de Janeiro encarregado de prover as informações cartográficas desta cidade. As variáveis estatísticas foram associadas à malha do setor censitário de 2010 e, em seguida, por meio de operações espaciais, associou-se as informações contidas nos polígonos dos

setores censitários aos polígonos das zonas de tráfego.

4.3.1 Variáveis estatísticas

As variáveis estatísticas estão associadas a fatores demográficos e socioeconômicos e foram obtidas a partir do censo demográfico do IBGE de 2010, agregados no nível de setor censitário, englobando informações sobre os domicílios, tais como quantidade de moradores, número de banheiros, tipo de coleta de lixo, fornecimento de energia elétrica, etc., bem como informações sobre os moradores como faixa etária e nível de escolaridade.

4.3.2 Variáveis associadas às características das vias

Adotou-se como variáveis associadas às características das vias, a contagem da quantidade de interseções e de trechos de ruas, a largura e o comprimento das vias dentro de cada zona de tráfego.

Às larguras das vias não se atribui um valor em unidade de comprimento e sim valores de acordo com a quantidade de faixas de rolamento. Dessa forma, atribuíram-se valores entre 1 e 5 para cada trecho de via, com intervalos de 0,5 para os casos em que se tivesse um estacionamento na lateral da via. Por exemplo, quando uma via que possui originalmente 2 (duas) faixas de rolamento, mas com uma delas sendo empregada como estacionamento, adotou-se o valor de 1,5 para a quantidade de faixas de rolamento. Tal procedimento foi adotado por se verificar que, na prática, grande parte das vias nessa situação comporte duas fileiras de carros de passeio com pequena distância entre si.

A partir dos comprimentos dos trechos das vias, obteve-se o comprimento total das vias situadas dentro de cada zona de tráfego e o comprimento das vias separadas em cada hierarquia adotada pela CET-Rio, a qual classifica as vias em expressas, estruturais, arteriais primárias, arteriais secundárias, coletoras e locais. As vias estruturais são as vias que estabelecem ligações rápidas para o tráfego de passagem exclusivo e compõem o eixo principal do sistema viário urbano. As arteriais primárias fazem as ligações entre os centros de alcance metropolitanos, bem como as ligações entre estes e as vias estruturais. As vias secundárias fazem as ligações entre os centros de alcance municipal e destes com os centros de alcance metropolitano e também com as vias estruturais e arteriais primárias. As vias coletoras fazem a coleta e a distribuição de tráfego interno aos bairros, alimentando o sistema arterial e as vias locais englobam todas as ruas utilizadas para o acesso direto às residências, comércio ou indústrias, com tráfego exclusivamente local.

Gerou-se a partir da largura das vias, as variáveis largura média das vias ponderado pelo comprimento total das vias ($Larg_{med}$) e o comprimento médio das vias ponderado pela largura das mesmas ($Comp_{med}$), constantes nas Equações 5 e 6, respectivamente.

$$Larg_{med} = \frac{1.Comp_1 + 1,5.Comp_{1,5} + \dots + n.Comp_n}{Comp_{total}} \quad (5)$$

Onde n está entre 1 e 5, com intervalos de 0,5. $Comp_1$ representa o somatório do comprimento da via cujo número de faixas de rolamento é 1 e $Comp_{total}$ o comprimento total das vias dentro da zona de tráfego.

$$Comp_{med} = \frac{1.Comp_1 + 1,5.Comp_{1,5} + \dots + n.Comp_n}{22,5} \quad (6)$$

Onde i está entre 1 e 5, com intervalos de 0,5. O valor de 22,5 representa o somatório de todos os diferentes tipos de faixa de rolamento considerados.

Gerou-se também as seguintes variáveis: razão entre o comprimento das vias e o número de vias, razão entre o número de interseções e a área da zona de tráfego e a razão entre a quantidade de trechos de vias e o número de interseções.

5 METODOLOGIA EMPREGADA

As etapas seguidas, desde a preparação dos dados até a seleção de variáveis e a construção de um modelo de regressão, serão descritas a seguir.

5.1 Preparação dos dados geoespaciais e dos dados estatísticos

Conforme mencionado, as medições relacionadas às características das vias foram obtidas a partir da malha de ruas do IPP. Estes inicialmente estavam no formato de *Computed Aided Design* (CAD) e sem a preparação para ser utilizado em SIG. Converteu-se para o formato *shapefile*, utilizado no ArcGIS e editou-se o arquivo de tal forma que em cada interseção de via fosse associado um ponto e que houvesse somente uma linha ligando tais interseções. Transformou-se o arquivo do formato espaguete para o formato topológico, gerando-se nós e arestas no lugar de pontos e linhas.

Fez-se a ligação do arquivo da malha do setor censitário da cidade do Rio de Janeiro com as tabelas contendo as informações estatísticas no ArcGIS. Por meio de uma operação espacial, relacionou-se os dados estatísticos dos setores censitários às zonas de tráfego que continham os centróides dos polígonos dos setores censitários. Desta forma, a cada polígono de zona de tráfego passou-se a ter o somatório dos valores de cada uma das variáveis obtidas do censo de cada um dos setores censitários.

5.2 Geração dos dados no SIG

Obteve-se o número de interseções de vias dentro de cada zona de tráfego por meio de uma operação espacial, na qual se somou o número de nós contidos em cada polígono e associou-se às zonas de tráfego. Desta quantidade de interseções, foram descontados aqueles valores de nós que estavam na extremidade dos segmentos.

O número de faixas de circulação de cada uma das vias foi obtido a partir da observação das imagens de alta resolução do *Google Maps*. Quando não era possível visualizar a via a partir do satélite, principalmente pela existência de árvores, recorreu-se ao aplicativo *Google Street View*. O número de vias era então associado à tabela de atributos da malha de ruas no ArcGIS. Por meio de operações de álgebra de mapas, obteve-se o somatório dos comprimentos das vias de cada tipo e a quantidade de trechos de ruas, contidos em cada zona de tráfego. Por meio de operações algébricas na própria tabela de atributos das zonas de tráfego, obtiveram-se as variáveis largura média das vias ponderado pelo comprimento total das vias, comprimento médio das vias ponderado pela largura das mesmas, razão entre o comprimento das vias e o número de vias, razão entre o número de interseções pela área da zona de tráfego e a razão entre a quantidade de trechos de vias e o número de interseções. A hierarquia das vias foi associada diretamente ao arquivo de atributos da zona

de tráfego a partir da informação da CET-Rio.

5.3 Análise exploratória espacial

A análise exploratória espacial foi feita inicialmente gerando-se os mapas de Kernel dos acidentes para todos os anos a partir das localizações pontuais dos acidentes, com a finalidade de se verificar a existência de mudança considerável na distribuição da intensidade da variável ao longo dos anos de estudo. Somou-se as informações de acidentes contidos nos polígonos das zonas de tráfego e gerou-se os mapas temáticos dos acidentes para se verificar a distribuição espacial dos acidentes por zonas de tráfego que apresentam, bem como a sua evolução ao longo do tempo. Construiu-se os mapas coropléticos com a visualização do Moran *scatterplot* para verificar o quão os valores dos acidentes em cada zona de tráfego estão acima ou abaixo da média de toda a área de trabalho ou de sua vizinhança e assim detectar aglomerados espaciais.

5.4 Modelagem de regressão

Aplicou-se uma modelagem de regressão linear multivariada com o objetivo de verificar dentre as variáveis utilizadas na pesquisa, aquelas que seriam as mais explicativas dos acidentes de trânsito e o quão as mesmas seriam explicativas do fenômeno.

Aplicou-se o índice de Moran sobre a variável densidade de acidentes e sobre os resíduos da regressão linear com o objetivo de verificar se a variável dependente apresenta autocorrelação espacial e o quão essa dependência espacial afeta a regressão linear. Escolheu-se como variável dependente a densidade de acidentes no ano de 2011 por serem os mais recentes e apresentarem comportamento similar aos dos anos de 2008 e 2009.

6 RESULTADOS

A partir da localização dos acidentes, gerou-se as superfícies de Kernel dos acidentes da Zona Sul do Rio de Janeiro, no período de 2008 a 2011, conforme pode ser visto na Fig 5. O mapa de Kernel possui a vantagem de se ter uma visão mais contínua do evento pontual em análise, no caso os acidentes de trânsito.

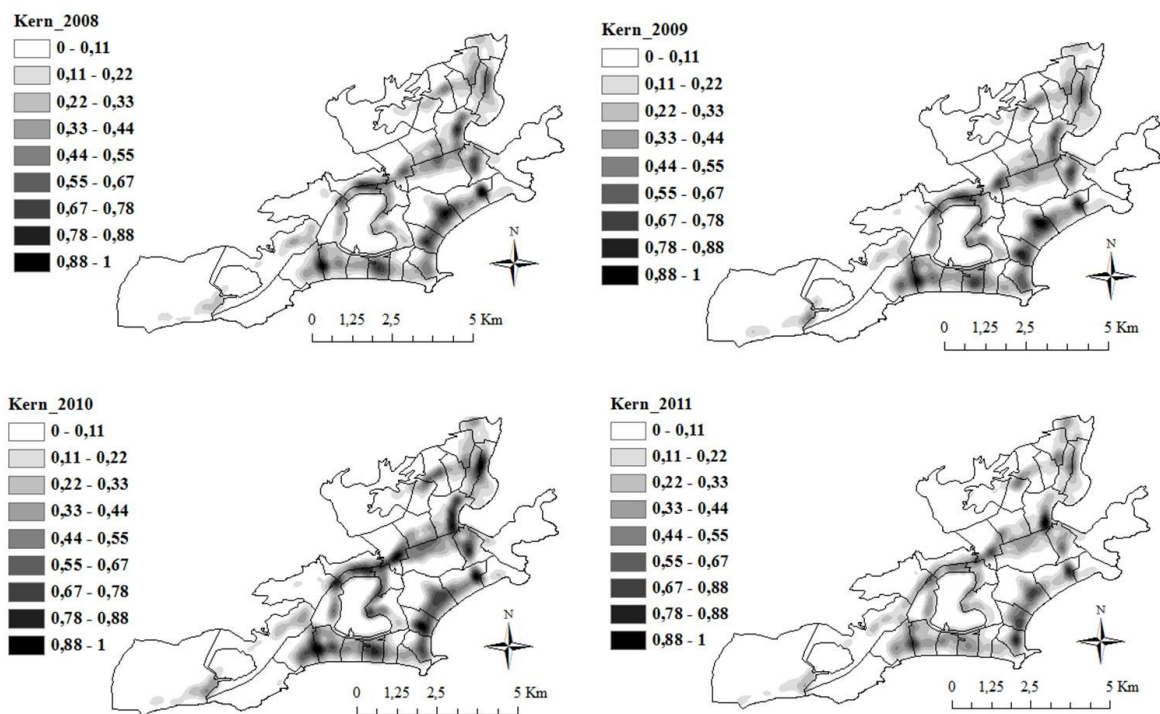


Fig 5 Mapas de Kernel dos acidentes da Zona Sul do Rio de Janeiro

Os mapas coropléticos dos acidentes nos anos de 2008 a 2011 obtidos da agregação dos dados pontuais dos acidentes nas áreas das zonas de tráfego estão apresentados na Fig 6, cujas legendas foram construídas pelo critério dos quartis.

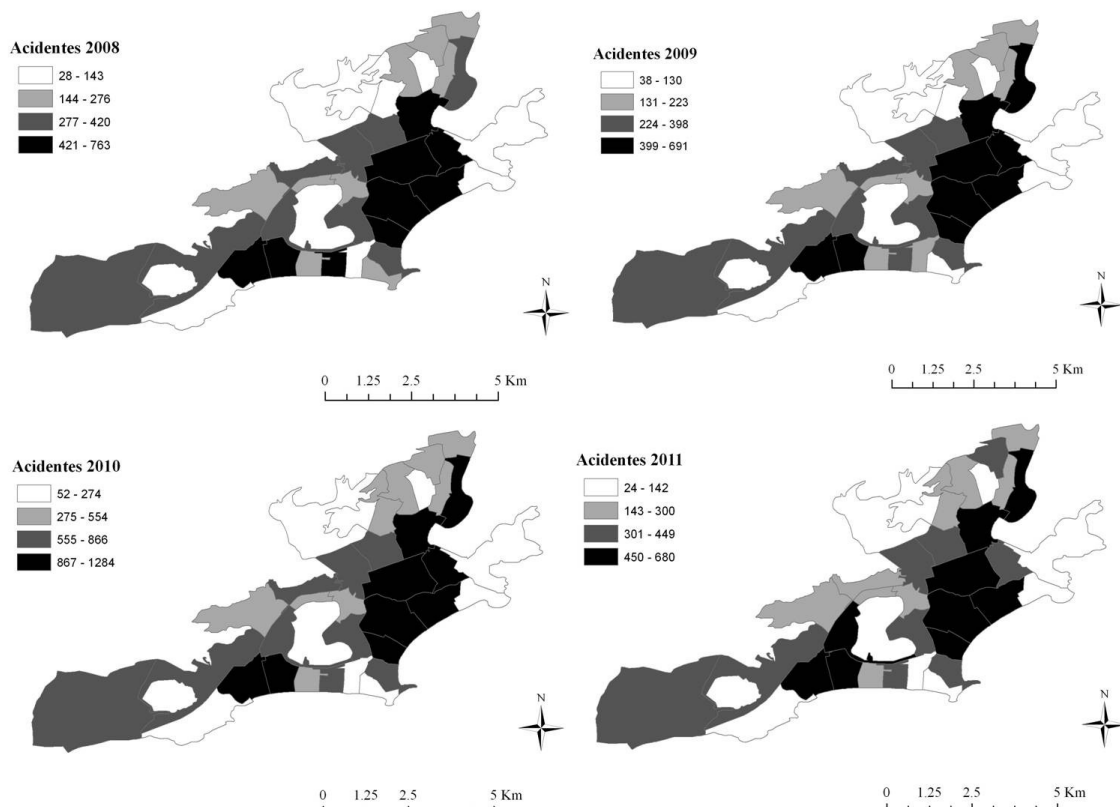


Fig 6 Mapas coropléticos dos acidentes nos anos de 2008 a 2011

Os mapas com a visualização do *scatterplot* para os anos de 2008 a 2011 podem ser vistos na Fig 7.

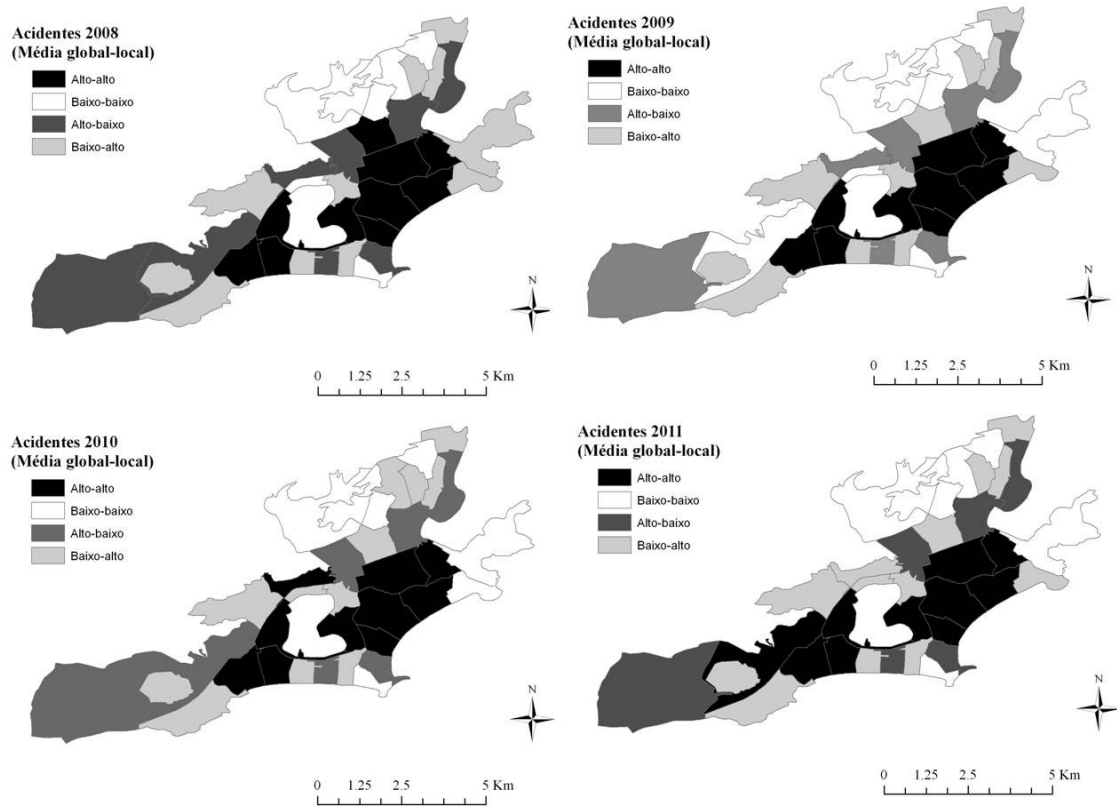


Fig 7 Visualização do *scatterplot* dos acidentes no período 2008-2011

Após a aplicação do modelo de regressão linear utilizando-se as variáveis citadas anteriormente, obteve-se como variáveis mais representativas com nível de confiança de 95% a quantidade de trechos de ruas e o comprimento de ruas do tipo arterial primária presentes em cada zona de tráfego. As variáveis domicílios particulares permanentes com três banheiros de uso exclusivo dos moradores e largura média das vias ponderado pelo comprimento total das vias apresentaram correlação com a variável dependente. No entanto, apresentam alta correlação com o somatório do número de trechos de ruas e o comprimento de ruas do tipo arterial primária. O valor do R^2 ajustado do modelo ficou em torno de 0,69.

Após a aplicação do índice de Moran com o objetivo de se verificar a dependência espacial do resíduo da regressão linear, obteve-se para ambos um valor próximo de 0, o que indica independência espacial. Costuma-se verificar a dependência espacial dos resíduos para identificar a necessidade ou não de se utilizar um modelo de regressão espacial, sendo este utilizado quando se obtém valores do índice próximos de 1.

7 CONCLUSÕES

A partir da observação das superfícies de Kernel, verifica-se a existência de regiões com maior incidência de acidentes em Ipanema, Leblon, Jardim Botânico, Humaitá, Copacabana, Flamengo e Botafogo. No caso de Copacabana, identifica-se três regiões bem

distintas e no de Botafogo estão mais concentrados mais à direita do bairro, região esta mais próxima da praia de Botafogo e do shopping Rio Sul. No caso do Flamengo, estão concentrados na praia do Flamengo e no aterro deste nome.

Observando-se os mapas coropléticos dos acidentes, verifica-se que o ano de 2010 apresenta quantidade de acidente bem superior aos demais anos. Os bairros mais evidenciados nesta representação são os de Leblon, Copacabana, Botafogo e Flamengo. No ano de 2008 aparece também em destaque uma parte do bairro de Ipanema e em 2011 aparece evidenciado o bairro da Lagoa.

A partir da observação da visualização do Moran *scatterplot*, pode-se confirmar a existência de aglomerados de regiões com maiores índices de acidente em Copacabana, Leblon e parte dos bairros de Botafogo, do Flamengo e da Lagoa. O bairro de Santa Teresa é que apresenta o maior aglomerado de regiões de baixo índice de acidente.

A grande importância da variável comprimento de ruas do tipo arterial primária e largura média das vias ponderadas pelo comprimento total das vias mostra que os bairros com ruas mais largas tendem a apresentar maior quantidade de acidentes. A variável quantidade de trechos de ruas mostra a tendência a haver mais acidentes em regiões com maior quantidade de interseções. As variáveis número de domicílios e número de moradores não apresentaram correlação com os acidentes, o que pode mostrar que muitos dos acidentes possam estar envolvendo pessoas que trabalhem ou estejam de passagem pela região.

8 REFERÊNCIAS

Anderson. T.K. (2009) Kernel density estimation and K-means clustering to profile road accident hotspots. **Accident Analysis and Prevention**, v. 41, p. 359-364.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association, *in* M. Fischer, H. Scholten, D. Unwin (eds), **Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences**, Taylor and Francis, London.

Anselin, L. (2002). **Mapping and Analysis for Spatial Social Science**. Disponível em <http://www.csiss.org/aboutus/presentations/files/anselin_aaa.pdf>.

Bailey, T.C. e A.C. Gatrell (1995). **Interactive Spatial Data Analysis**, John Wiley & Sons, New York.

Dent, B.D (1985). **Principles of Thematic Map Design**. Addison-Wesley Publishing Company.

Druck, S., M.S. Carvalho, G, Câmara, A.V.M. Monteiro (eds) (2004). **Análise Espacial de Dados Geográficos**. Brasília, EMBRAPA, 2004.

Ha, H.H and J.C. Thill (2011). Analysis of traffic hazard intensity: A spatial epidemiology case study of urban pedestrians. **Computers, Environment and Urban Systems**, v. 35, p. 230-240.

Marshall, W. E. e N.W.Garrick (2010) Does street network design affect traffic safety?

Accident Analysis and Prevention, In press.

Miranda-Moreno, L.F., P. Morency, A. M. El-Geneidy (2011) The link between built environment, pedestrian activity and pedestrian - vehicle collision occurrence at signalized intersections. **Accident Analysis and Prevention**, In press.

Noland, R. B. e M.A.Quddus. (2004) A spatially disaggregate analysis of road casualties in England. **Accident Analysis and Prevention**, v. 36, p. 973-984.

Rifaat, S.M., R. Tay; A. de Barros (2011) Effect of street pattern on the severity of crashes involving vulnerable road users. **Accident Analysis and Prevention**, v. 43, p. 276-283.

Southworth, M.e E. Ben-Joseph (2003) **Streets and the Shaping of Towns and Cities**. Island Press, Washington, DC.