

## CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ

Rio de Janeiro - RJ - Brasil

ISSN impresso 1413-9022 / ISSN on-line 2317-4536 - v. 35, p. 01 - 15, 2013

# ANÁLISE DE CONGLOMERADOS DE ACIDENTES DE TRÂNSITO UTILIZANDO GOOGLE MAPS E ESTATÍSTICA ESPACIAL

Marcos Antônio da Cunha Santos  
Universidade Federal de Minas Gerais  
msantos@est.ufmg.br

Marcelo Azevedo Costa  
Universidade Federal de Minas Gerais  
macosta.est@gmail.com

Marcos Oliveira Prates  
Universidade Federal de Minas Gerais  
marcosop@est.ufmg.br

### Resumo

*Este artigo apresenta novas abordagens para a análise de acidentes de trânsito direcionada a estruturas de arruamentos. O método proposto é uma extensão do modelo geo-estatístico de varredura temporal porém considera eventos pontuais espalhados ao longo de um arruamento. Coordenadas geográficas de eventos pontuais são inicialmente mapeadas em uma única dimensão, obtendo-se a linha georeferenciada do arruamento de interesse. Nesta dimensão, os eventos pontuais são identificados pelas suas distâncias relativas a um ponto de origem no arruamento. Uma janela unidimensional e de dimensão variável realiza uma varredura para identificar trechos nos quais a taxa de incidência de acidentes de trânsito é maior que a esperada. Inferência estatística é realizada a partir de simulações de Monte Carlo. As metodologias foram implementadas no software R com uso de interfaces gráficas e mapas de arruamentos obtidos a partir do ambiente Google maps.*

**Palavras-chave:** Acidentes de trânsito; Detecção de conglomerados; Estatística de varredura espacial; Método scan.

## 1. Introdução

Registros de acidentes de trânsito podem ser associados às coordenadas geográficas do local exato do acidente. A partir das coordenadas espaciais dos eventos é possível delinear regiões onde a incidência dos eventos é maior que o esperado. Neste contexto, a análise de conglomerados espaciais (DIGGLE, 1983) é uma metodologia que permite identificar estas regiões críticas e avaliar se a hipótese de que os eventos observados estão distribuídos espacialmente de modo homogêneo sobre a região em estudo deve ser aceita.

Dentre as técnicas estatísticas de análise de conglomerados espaciais, a estatística de varredura espacial, proposta por Kulldorff (1997), permite a identificação de conglomerados espaciais significativos a partir de eventos pontuais ou eventos de áreas. Esta metodologia tornou-se muito popular em diversas áreas do conhecimento, como mostra Costa e Kulldorff (2009). Como consequência, novas abordagens têm sido propostas, com extensões para detecção de conglomerados puramente temporais ou espaço-temporais (KULLDORFF *et al.* 1998; KULLDORFF, 2001; KULLDORFF *et al.*, 2005). Algumas novas extensões exploram variações na geometria espacial e espaço-temporal da janela de varredura (ALM, 1997; KULLDORFF *et al.*, 2006; DUCZMAL & ASSUNÇÃO, 2004; COSTA *et al.*, 2012).

Neste trabalho apresentamos uma nova variação do método de estatística de varredura espacial, desenvolvida para análise de eventos pontuais cuja ocorrência é restrita a estruturas de arruamentos. Esta é a situação típica em análises de acidentes de trânsito. Análises de conglomerados considerando somente os aspectos espaciais de dados de trânsito podem ser encontradas na literatura (HUANG *et al.*, 2009). Entretanto, uma restrição em uma análise puramente espacial é que não é possível identificar localmente, isto é, ao longo de um arruamento específico, as regiões de alta ou baixa incidência de eventos. Isto porque um conglomerado de eventos puramente espacial em duas dimensões pode abranger diversos arruamentos, sem que seja possível uma diferenciação com relação à contribuição dos eventos de cada arruamento separadamente (Figura 1). Como consequência, um trecho de um arruamento, em técnicas de análise que não consideram a geometria dos arruamentos, poderá ser caracterizado como crítico simplesmente porque na análise do conglomerado não foi

possível distinguir eventos de arruamentos distintos. É o caso, por exemplo, de uma rua ou segmento paralelo a uma avenida com alta incidência de eventos pontuais.

Metodologias estatísticas para a identificação e caracterização de trechos críticos de ruas e avenidas é assunto de interesse e pode auxiliar os órgãos responsáveis pelo trânsito na criação de políticas de restrição, como uso de radares, alterações de sinalização e outras possíveis melhorias.

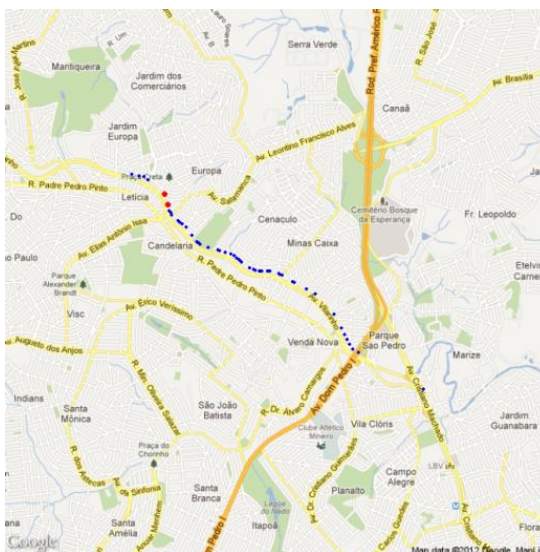
Figura 1 - Eventos localizados dentro do círculo e fora do arruamento de interesse são considerados no método de detecção de conglomerados espaciais, o que não permite uma análise por arruamento.



A metodologia apresentada neste trabalho foi desenvolvida a partir um projeto de pesquisa desenvolvido no Centro de Estudos de Criminalidade e Segurança da UFMG (CRISP), com participação da Empresa de Transporte e Trânsito de Belo Horizonte (BHTRANS). Neste projeto, dados de acidentes de trânsito ocorridos no período de 2004 a 2010 foram georeferenciados. Uma plataforma para consulta, visualização e análises de dados em ambiente do software R foi desenvolvida. Esta plataforma, denominada RBHTrans, possibilita ao usuário a consulta total ou parcial da base de dados. A partir dos dados selecionados, é possível a utilização de funcionalidades como análise de mapas de kernel, moda espacial, análise descritiva de eventos por arruamentos e uso de um método de estatística de varredura linear, que foi denominado *street scan*. A plataforma utiliza os pacotes RgoogleMaps e Rgooglevis, que possibilitam o acesso online a mapas da plataforma Google maps e permitem exportar atributos georeferenciados para visualização em browsers, como o Google Chrome ou Mozilla Firefox. Utilizando a plataforma desenvolvida com uso destes pacotes R, o usuário pode realizar análises de arruamentos e visualizar os dados

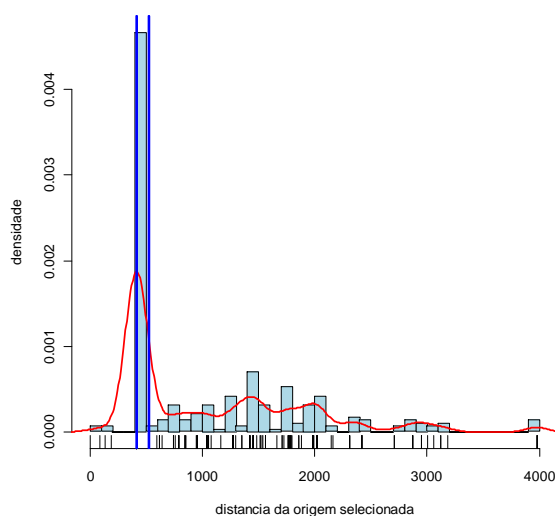
georeferenciados de acidentes sobrepostos a mapas do tipo satélite com utilização inclusive do ambiente Street View, recurso disponível a partir do Google maps. Como exemplo, a Figura 2 mostra as coordenadas de um conjunto de acidentes de trânsito ocorridos ao longo da Avenida Vilarinho, em Belo Horizonte (MG), durante o ano de 2010.

Figura 2 - Eventos pontuais ao longo de um arruamento de Belo Horizonte. Em vermelho eventos pertencentes a um conglomerado de alta incidência.



Fonte: Centro de Estudos de Criminalidade e Segurança Pública da UFMG (CRISP)

Figura 3 - Histograma da densidade de eventos ao longo de um arruamento. As linhas verticais delimitam um conglomerado atípico detectado pelo método *street scan*.



Fonte: Centro de Estudos de Criminalidade e Segurança Pública da UFMG (CRISP)

A Figura 3 mostra um histograma obtido para a densidade de eventos ao longo do arruamento. As linhas verticais delimitam um conglomerado atípico, detectado pelo método proposto neste trabalho, descrito na seção a seguir.

## 2. A Estatística de Varredura Unidimensional

Nesta seção descreveremos o método de análise estatística de conglomerados utilizando uma varredura unidimensional. Seja  $s_i = (x_i, y_i)$  o par de coordenadas espaciais atribuído ao  $i$ -ésimo evento pontual. Na abordagem utilizada neste trabalho, um arruamento ou corredor de tráfego é considerado unidimensional, sendo representado em um mapa como um segmento de linha. Considere o conjunto  $S = \{s_1, s_2, \dots, s_N\}$  formado pelos eventos de um mesmo arruamento e ordenado de acordo com as distâncias, medida ao longo do arruamento, a partir de um evento inicial  $s_1$  do arruamento.

Deste modo  $s_1$  e  $s_N$  são o primeiro e o último evento observado, respectivamente, do arruamento e  $N$  é número total de eventos. O objetivo é detectar trechos entre  $s_1$  e  $s_N$  com valores atípicos de intensidade de ocorrência de eventos. Seja  $d_{k,k+1}^*$ ,  $k \in \{1, \dots, (N-1)\}$ , a distância euclidiana entre os eventos consecutivos  $s_k, s_{k+1} \in S$ . Deste modo, a distância entre o evento inicial  $s_1$  e o  $i$ -ésimo evento pode ser escrita como (1):

$$d_{1,i} \approx \sum_{k=1}^{i-1} d_{k,k+1}^*. \quad (1)$$

Esta aproximação é adequada em situações onde a distância entre pontos consecutivos ao longo de um arruamento é pequena. A Figura 4 mostra a distribuição de eventos pontuais ao longo de um arruamento hipotético, considerando as distâncias relativas de cada evento ao ponto de origem  $s_1$ .

Para construção de um teste estatístico para localização de trechos de um arruamento com aglomerados atípicos, considere a hipótese nula de que os  $N-2$  eventos localizados entre os pontos  $s_1$  e  $s_N$  ocorram de forma homogênea no arruamento. Sob esta hipótese, a intensidade de eventos,  $\lambda_{H_0}$ , ao longo do arruamento é dada por (2):

$$\lambda_{H_0} = \frac{N-2}{d_{1,N}}. \quad (2)$$

Desta forma, o número de casos ocorridos ao longo de um trecho de comprimento  $\Delta d < d_{1,N}$  pode ser modelado como a variável de Poisson  $Y_{\Delta d} \sim \text{Poisson}(\mu_{\Delta d} = \lambda_{H_0} \Delta d)$ . É de particular interesse identificar automaticamente trechos ao longo do arruamento que apresentem uma intensidade maior do que a esperada. Para isto propomos o seguinte teste de razão de verossimilhança: seja  $l$  uma janela de comprimento variável  $0 < l < d_{1,N}$ . Seja  $c_l$  o número observado de eventos na janela  $l$  e  $\mu_l$  o número esperado de casos ao longo de  $l$ . Sob a hipótese nula,  $\mu_l = l \cdot \lambda_{H_0}$ . A estatística do teste da razão de verossimilhança é dada por (3):

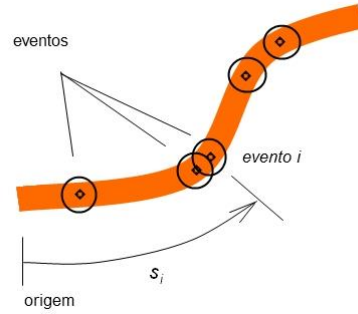
$$\hat{\kappa} = \kappa(\hat{l}) = \sup_l \left( \frac{c_l}{\mu_l} \right)^{c_l} \left( \frac{N - c_l - 2}{N - \mu_l - 2} \right)^{N - c_l - 2} \quad (3)$$

A partir da Equação (3), dado um arruamento, é possível identificar se um trecho  $\hat{l}$  apresenta incidência de eventos maior ou menor do que a esperada. Caso seja de interesse identificar somente trechos de alta incidência, deve ser aplicada a restrição  $c_l > \mu_l$ .

Para avaliar o valor observado da estatística de teste em relação à hipótese nula em um teste de nível  $\alpha$ , são realizadas  $m$  simulações Monte Carlo, de acordo com o seguinte procedimento:

1. Para  $1 \leq j \leq m$ , simule um padrão de pontos ao longo do arruamento, de acordo com a hipótese nula. Para isto, gere eventos pontuais de acordo com um processo de Poisson de intensidade  $\lambda_{H_0}$  ao longo de um trecho de comprimento  $d_{1,N}$ . Calcule a estatística da razão de verossimilhança  $\hat{\kappa}_j$ .
2. Calcule a estatística da razão de verossimilhança  $\hat{\kappa}_0$  a partir dos dados observados.
3. Rejeite a hipótese nula se o valor da estatística de teste  $\hat{\kappa}_0$  estiver acima do percentil  $100(1 - \alpha)\%$  dos valores simulados  $\hat{\kappa}_1, \hat{\kappa}_2, \dots, \hat{\kappa}_m$ ; caso contrário aceite a hipótese nula.

Caso a hipótese nula seja rejeitada em um trecho  $\hat{l}$ , há indícios de que trata-se de um trecho crítico e portanto o trecho deve ser foco de atenção.

Figura 4 - Distância do  $i$ -ésimo evento à origem, em um arruamento.

### 3. A Estatística de Varredura Unidimensional com eventos esparsos

O modelo descrito na seção anterior é adequado quando a distância mínima entre dois eventos consecutivos (distância euclidiana) em um arruamento é aproximadamente igual à distância entre estes dois eventos medida ao longo do arruamento. Deste modo, o modelo descrito na seção anterior é útil em arruamentos aproximadamente retos ou com curvas pouco acentuadas. Em arruamentos que apresentam curvas acentuadas, o procedimento é adequado somente se houver um grande número de eventos próximos.

Em várias situações, no entanto, os arruamentos de interesse apresentam curvas acentuadas e um número de eventos não muito próximos, o que pode impedir o cálculo correto das distâncias entre os eventos ao longo do arruamento. Nestes casos, um procedimento computacional alternativo é considerar as coordenadas de pontos do arruamento, além das coordenadas dos eventos.

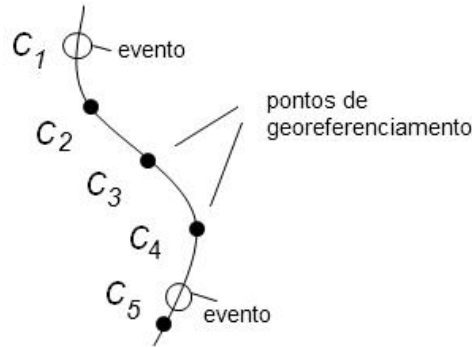
Seja  $R_i$  um par de coordenadas espaciais dadas por  $R_i = (x_i, y_i)$ , atribuídas a um  $i$ -ésimo ponto espacial de um arruamento, com  $i \in \{1, \dots, N_r\}$ , onde  $N_r$  é o número total de pontos conhecidos no arruamento. Seja  $S_i = (x_i, y_i)$  um par de coordenadas espaciais atribuídas ao  $i$ -ésimo evento pontual espacial,  $i \in \{1, \dots, N_s\}$ , onde  $N_s$  é o número total de eventos pontuais no arruamento. Dessa forma temos os conjuntos de coordenadas  $\mathbf{R} = \{R_1, R_2, \dots, R_{N_r}\}$  e  $\mathbf{S} = \{S_1, S_2, \dots, S_{N_s}\}$  associados ao arruamento. Considere o conjunto  $\mathbf{C} = \mathbf{R} \cup \mathbf{S}$  formado por  $N = N_r + N_s$  elementos. Considere o conjunto  $\mathbf{C}$  ordenado de tal modo que, para  $m \leq n$ ,  $d_{C_m} \leq d_{C_n}$ , onde  $d_{C_j}$  é a distância entre o ponto de origem e o ponto de coordenadas  $C_j$  medida ao longo do arruamento. Seja  $d_{C_k, C_{k+1}}^*$  a distância euclidiana entre dois elementos consecutivos do conjunto  $\mathbf{C}$ . A Figura 5 ilustra o conjunto de pontos de um conjunto  $\mathbf{C}$  para um arruamento. A distância

entre o evento inicial  $s_1$  e o  $i$ -ésimo evento medida ao longo do arruamento pode ser obtida como (4):

$$d_{s_1, s_i} \approx \sum_{k=1}^{N-1} d_{c_k, c_{k+1}}^* I_k I_{k+1} \quad (4)$$

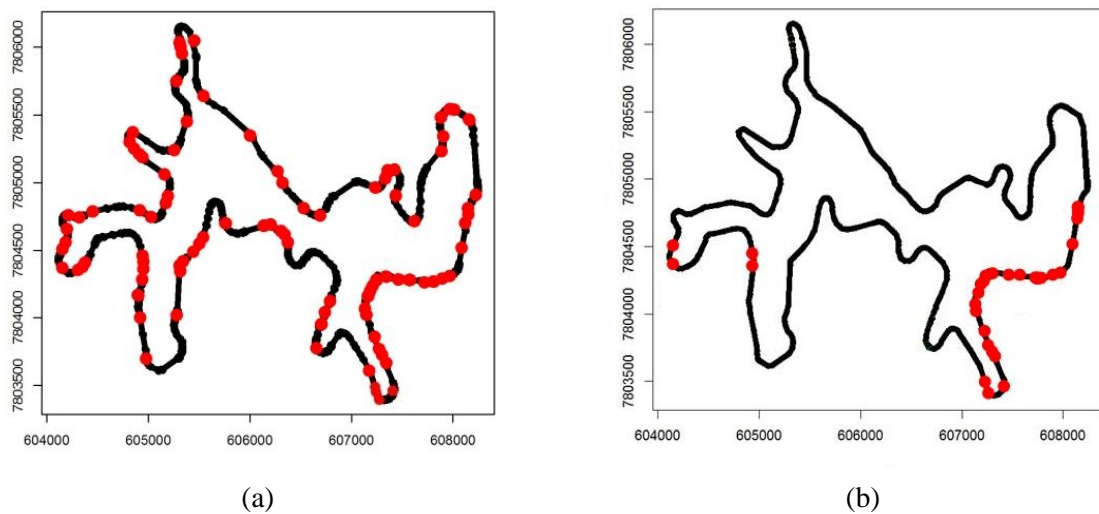
onde  $I_k = 1$ , se  $d_{s_1} \leq d_{c_k} \leq d_{s_i}$  e  $I_k = 0$ , caso contrário.

Figura 5 – Conjunto de pontos de georeferenciamento e eventos em um arruamento.



A Figura 6a mostra o entorno da Lagoa da Pampulha (Av. Otacílio Negrão de Lima), em Belo Horizonte, Minas Gerais. O contorno da avenida foi obtido a partir de dados disponíveis para georeferenciamento. Um conjunto com 211 ocorrências, registradas durante o ano de 2011, foi analisado pelo método proposto. A Figura 6b mostra os aglomerados mais significativos, após análise.

Figura 6 – Entorno da Lagoa da Pampulha. (a) Pontos para georreferência (preto) e ocorrências de acidentes de trânsito (vermelho). (b) Pontos para georreferência e eventos de aglomerados significativos.



Fonte: Centro de Estudos de Criminalidade e Segurança Pública da UFMG (CRISP)



#### 4. O modelo Bernoulli

O modelo de Poisson, apresentado anteriormente, supõe que a intensidade de eventos é constante ao longo de um arruamento. Na prática, um arruamento pode apresentar diferentes intensidades ao longo de seu comprimento. Na ausência de informações detalhadas da intensidade de eventos ou do fluxo de veículos ao longo do arruamento, pode-se comparar a intensidade dos eventos para diferentes períodos de tempo. Como exemplo, é possível comparar a intensidade dos eventos ao longo do ano de 2011 com relação ao ano anterior, 2010. Para isto pode ser utilizado o modelo de Bernoulli.

Seja  $c_{l,t}$  o número de eventos observados ao longo da janela  $l$  no intervalo de tempo  $[0, t]$  e  $c_{l,t+\Delta t}$  o número de eventos observados ao longo de  $l$  no intervalo de tempo  $[t, t + \Delta t]$ , onde  $\Delta t$  é um intervalo usual (mes, ano). Por simplicidade e sem perda de generalidade, considere  $\Delta t = 1$ . Seja  $n_l = c_{l,t} + c_{l,t+1}$  o total de eventos observados ao longo de  $l$ ,  $C_t = \sum_l c_{l,t}$  o total de eventos observados no tempo  $t$  e  $N = \sum_l C_t$  o número total de eventos observados no arruamento de interesse nos períodos  $t$  e  $t+1$ . Utilizando o modelo de Bernoulli, o trecho crítico  $\hat{l}$  é estimado a partir da maximização da função de verossimilhança (5):

$$L(\hat{l}) = \sup_{l, p > r} p^{c_{l,t}} (1-p)^{(n_l - c_{l,t})} r^{(C_t - c_{l,t})} (1-r)^{(N - n_l - C_t + c_{l,t})} \quad (5)$$

onde  $\hat{p} = c_{l,t}/n_l$  e  $\hat{r} = (C_t - c_{l,t})/(N - n_l)$ .

Semelhante ao modelo de Poisson, a inferência estatística é realizada a partir de simulações de Monte Carlo. Neste caso, sob a hipótese da não ocorrência de conglomerados espaciais ao longo do arruamento (hipótese nula), as coordenadas dos eventos pontuais podem ser aleatoriamente atribuídas aos instantes de tempo  $t$  e  $t+1$ . Dessa forma, a distribuição empírica da estatística de teste em (3) pode ser obtida através de simulações.

#### 5. Implementação Computacional

A estatística de varredura unidimensional para os modelos de Poisson e Bernoulli foi implementada no software R com uso dos pacotes *RgoogleMaps* e *googleVis*. O pacote *RgoogleMaps* (LOECHER, 2010) possibilita a importação de

imagens do ambiente *Google maps* para o software R. As imagens são importadas no formato png (*Portable Network Graphics*) e utilizadas como plano de fundo onde é possível a sobreposição de pontos, linhas e polígonos em uma estrutura de camadas. No pacote desenvolvido, as funcionalidades implementadas podem ser obtidas através de sequências simples de comandos no ambiente do software R. Por exemplo, a importação de mapas e sobreposição de imagem do Google Maps pode ser obtida com os comandos:

```
R> MyMape <- GetMap.bbox(lonR, latR, center, size = c(640, 640),  
  destfile = "MyTile.png",...)  
  
R> PlotOnStaticMap(MyMap, lat, lon, destfile, ...)
```

Neste caso, os valores *lonR* e *latR* são os limites de longitude e latitude do mapa a ser gerado, *center* é o parâmetro de centralidade do mapa (opcional), *size* é a resolução da imagem e *destfile* é o nome do arquivo de destino da imagem. No comando *PlotOnStaticMap*, *lat* e *lon* são os vetores de pontos a serem sobrepostos na imagem *MyMap*.

O pacote *googleVis* (GESMANN & DE CASTILLO, 2011) possibilita a exportação de dados em HTML utilizando recursos do Google Visualisation API. Utilizando a funcionalidade *gvisMap()* é possível visualizar dados pontuais utilizando diretamente a plataforma Google maps, a partir de um browser. Neste ambiente, a funcionalidade *street view* do *Google maps* pode ser utilizada para visualizar os dados ao longo do arruamento.

A metodologia de varredura unidimensional foi implementada na função *street\_scan()* do pacote *RBHTrans*. O procedimento de simulação de Monte Carlo, que apresenta grande custo computacional, foi implementado em linguagem C e incorporada ao ambiente R na forma de uma *dll* (*Dynamic-link library*). Foram criadas interfaces gráficas para a seleção de variáveis do banco de dados a serem analisados. O pacote final, assim contruído, incorpora as análises de varredura espacial descritas nas seções anteriores e permite a obtenção, a partir do conjunto de dados selecionados, de (a) análise da intensidade de eventos em arruamentos, (b) mapas de kernel, (c) moda espacial, (d) análise de séries temporais, (e) street scan e (f) visualização e dados. A base de dados, as funcionalidades implementadas, a *dll* e as rotinas de interface gráfica foram

encapsuladas em um único pacote denominado RBHTrans. Todas as funcionalidades são disponibilizadas no ambiente R a partir do comando:

```
R> require(RBHTrans)
```

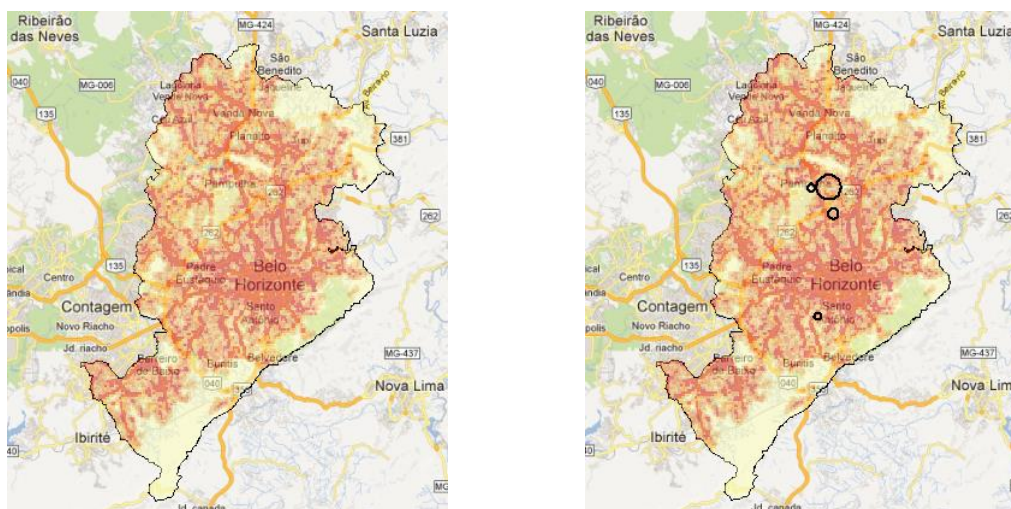
Na sequência, o usuário pode acessar interfaces gráficas do ambiente a partir de comando simples como `monta_banco()` e `escolhe_funcao()`. A figura 7 exemplifica o uso do recurso Street View, na visualização de uma região escolhida após análise do arruamento efetuada com os métodos implementados no ambiente pelo pacote RBHTrans. Neste exemplo, com uso do recurso Street View é possível a visualização de regiões que estão associadas aos conglomerados de acidentes de trânsito detectados pelos métodos de varredura descritos nas seções anteriores. A visualização é obtida utilizando-se a função `gvisMap()` do pacote `googleVis`.

Figura 7 – Um exemplo de visualização, no ambiente Google maps, de arruamentos em uma região de interesse obtida por análise de dados no ambiente do software R.



A Figura 8a mostra o mapa de intensidade de pontos para os eventos pontuais de acidentes de trânsito referentes aos anos de 2010 e 2011, na região metropolitana de Belo Horizonte. É nítida a intensidade de eventos nos principais arruamentos do município, que conduzem à área central do município. Na Figura 8b são apresentados os conglomerados mais críticos, ou seja, as regiões de maior intensidade de eventos no ano de 2011 com relação ao ano de 2010. Para esta análise foi utilizado o modelo espacial de Bernoulli.

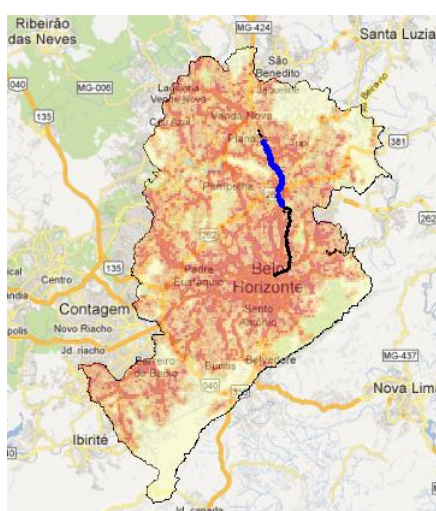
Figura 8. Mapa de intensidade de eventos pontuais (a) e conglomerados identificados utilizando o modelo espacial de Bernoulli (b).



Fonte: Centro de Estudos de Criminalidade e Segurança Pública da UFMG (CRISP)

A Figura 9 apresenta a análise de conglomerados da Avenida Cristiano Machado, em Belo Horizonte. Foi utilizando o modelo Bernoulli com dados de 2010 e 2011. Os resultados mostram que a análise utilizando somente os dados do arruamento permite identificar conglomerados diferentes daqueles encontrados quando a análise espacial é realizada considerando-se todos os eventos pontuais da região de interesse.

Figura 9. Conglomerado de maior incidência (linha azul) no ano de 2011, em comparação aos eventos pontuais do ano de 2010, considerando os dados da Avenida Cristiano Machado (linha preta e linha azul).



Fonte: Centro de Estudos de Criminalidade e Segurança Pública da UFMG (CRISP)

Comparando as figuras 8b e 9, percebe-se que os conglomerados não são sobrepostos. Ou seja, a análise focada a um único arruamento permite detectar trechos onde há evidência de um aumento excessivo do número de eventos pontuais quando comparado com os eventos do ano anterior. Outro aspecto interessante da análise é que o conglomerado encontrado no arruamento está presente em um trecho distante da área central. Ou seja, próximo ao centro é esperada uma intensidade maior de eventos, indiferente do ano da análise. Por outro lado, o trecho identificado apresenta um aumento da intensidade de eventos, conforme detectado pelo modelo Bernoulli.

## **6. Discussão e Conclusão**

Este trabalho apresenta novos métodos de varredura unidimensional desenvolvidos especificamente para detecção de conglomerados de acidentes de trânsito em arruamentos. O primeiro método utiliza o modelo estatístico de Poisson e supõe uma taxa de incidência homogênea ao longo do arruamento. Este modelo é utilizado para identificar trechos de arruamentos onde a intensidade de eventos é estatisticamente maior do que taxa esperada. O segundo modelo utiliza a distribuição de Bernoulli e é utilizado na detecção de conglomerados onde a taxa depende da incidência de eventos em um instante de tempo anterior. Neste último caso, o método permite identificar trechos do arruamento onde o número atual de eventos é significativamente superior ao número de eventos observados no mesmo trecho, em um instante de tempo anterior.

Ambos os métodos foram incorporados em um ambiente com interface gráfica que permite a análise dos eventos e dos conglomerados detectados utilizando recursos do Google maps. Dessa forma, o usuário pode visualizar remotamente o local do acidente de trânsito com grande riqueza de detalhes, além da possibilidade de análises descritivas e espaciais. Trabalhos futuros têm como objetivo agregar informações de tráfego de veículos e pedestres na estimativa de intensidade de eventos, sob a hipótese nula.

## **Agradecimentos**

Os autores agradecem à FAPEMIG, CNPq e CAPES pelo apoio financeiro.

## Referências

- ALM, S. E. On the distributions of scan statistics of a two dimensional Poisson process, **Advances in Applied Probability**, vol. 29, pages 1–18, 1997.
- COSTA, M. A.; KULLDORFF, M. In Scan statistics: methods and applications. **Birkhäuser: Statistics for Industry and Technology**, p. 129–52, 2009.
- COSTA, M. A.; ASSUNÇÃO, R. A.; KULLDORFF, M. Constrained spanning tree algorithms for irregularly-shaped spatial clustering. **Computational Statistics and Data Analysis**. vol. 56, p.1771–1783, 2012.
- DIGGLE, P. J. **Statistical Analysis of Spatial Point Patterns**, Academic-Press, London, 1983.
- DUCZMAL, L.; ASSUNÇÃO, R. A. Simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, **Computational Statistics and Data Analysis**, vol. 45, p. 269–286, 2004.
- GESMANN, M.; DE CASTILLO, D. Using the Google Visualisation API with R. **The R Journal**. vol. 3, n. 2, pages 40–44, 2011.
- HUANG, L.; STINCHCOMB, D. G.; PICKLE, L. W.; DILL, J. Identifying clusters of active transportation using spatial scan statistics. **American Journal of Preventive Medicine**. vol. 37, n. 2, pages 157–166, 2009.
- KULLDORFF, M. A spatial scan statistic. **Communications in Statistics: Theory and Methods**, vol. 26, pages 1481–1496, 1997.
- KULLDORFF, M.; ATHAS, W.; FEUER, E.; MILLER, B.; KEY, C. Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos. **American Journal of Public Health**, vol. 88, pages 1377–1380, 1998.
- KULLDORFF, M. Prospective time-periodic geographical disease surveillance using a scan statistic. **Journal of the Royal Statistical Society**, vol. A164, pages 61–72, 2001.
- KULLDORFF, M.; HEFFERNAN, R.; HARTMAN, J.; ASSUNÇÃO, R. M.; MOSTASHARI, F. A space-time permutation scan statistic for the early detection of disease outbreaks. **PLoS Medicine**, vol. 2, pages 216–224, 2005.
- KULLDORFF, M.; HUANG, L.; PICKLE, L.; DUCZMAL, L. An elliptic spatial scan statistic. **Statistics in Medicine**, vol. 25, pages 3929–3943, 2006.
- LOECHER, M. Plotting on Google Static Maps in R. **Technical Report**. 2010.
- R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria. Disponível em: < <http://www.R-project.org> >. Acesso em: 2013. 2013.

## TRAFFIC ACCIDENT CLUSTER ANALYSIS USING GOOGLE MAPS AND SPATIAL STATISTICS

### Abstract

*This paper presents a new approach for cluster detection of spatial point patterns which are restricted to street networks. The proposed method is an extension of the temporal scan statistic which is applied to spatial line segments. Geographical coordinates of points are initially mapped into a one dimension geographical structure, which is the geo-coded line of the street of interest. In this dimension, the events are identified by their relative distances to a point of origin. A one-dimensional varying scanning window identifies portions of the street where the incidence rate of car accidents is higher than the expected. Statistical inference is obtained using Monte Carlo simulations. The methodology was implemented in the R software and provides a friendly graphical user interfaces. The software provides online interface with Google maps.*

**Key-words:** *Detection of clusters; Traffic accidents; Spatial scan statistic, Scan method.*