

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

Marcelo Azevedo Costa

Departamento de Engenharia de Produção – UFMG
Av. Presidente Antônio Carlos, 6627, Belo Horizonte – MG – Brazil
e-mail: macosta.est@gmail.com

Marcos Oliveira Prates

Departamento de Estatística – UFMG
e-mail: marcosop@est.ufmg.br

Marcos Antônio da Cunha Santos

Departamento de Estatística – UFMG
e-mail: msantos@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 mas, considera eventos pontuais espalhados ao longo de um arruamento. Dessa forma, coordenadas geográficas de eventos pontuais são inicialmente mapeadas em uma única dimensão, que é a linha georeferenciada do arruamento de interesse. Nesta dimensão, os eventos pontuais são identificados pelas suas distâncias relativas a um ponto de origem. Uma janela unidimensional e de dimensão variável realiza a varredura no arruamento, procurando identificar trechos nos quais a taxa de incidência de acidentes de trânsito é maior que a esperada. Inferência estatística é obtida a partir de simulações de Monte Carlo. As metodologias foram implementadas no software R e utilizam interfaces gráficas e mapas de arruamentos obtidos a partir de interfaces com o ambiente Google maps.

PALAVRAS CHAVE: análise de conglomerados, estatística espacial, google maps. **Área principal:** L&T - Logística e Transportes

ABSTRACT

This paper presents new approaches for cluster detection of spatial point patterns which are restricted to street networks. The proposed method is an extension of the temporal scan statistic. The geographical coordinates of the 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.

KEYWORDS. cluster analysis, spatial statistics, google maps.

Main area: L&T - Logistics and Transportation

1. Introdução

Registros de acidentes de trânsito podem associar coordenadas geográficas ao local exato no qual o evento ocorreu. A partir das coordenadas 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, automaticamente, a região crítica e avaliar a hipótese de que os eventos poderiam ocorrer de forma homogênea sobre a região em estudo.

Dentre as técnicas estatísticas de análise de conglomerados espaciais, a estatística de varredura espacial, proposta por Kulldorff (1997), permite a identificação de conglomerados espaciais a partir de eventos pontuais ou eventos de áreas. Esta metodologia se tornou muito popular, em diversas áreas do conhecimento, como demonstra Costa e Kulldorff (2009). Como consequência, novas abordagens tem sido propostas, como extensões para detecção de conglomerados puramente temporais ou espaço-temporal [Kulldorff et al. 1998; Kulldorff, 2001; Kulldorff et al., 2005]. Além de novas metodologias que exploram variações na geometria espacial e espaço-temporal da janela de varredura [Alm, 1997; Kulldorff, 2006; Duczmal and Assunção, 2004; Costa et al., 2012].

Este trabalho apresenta uma nova variação da estatística de varredura desenvolvida, a princípio, para a análise de eventos pontuais cuja ocorrência é restrita a estruturas de arruamentos. Análises de conglomerados puramente espaciais aplicados a dados de trânsito podem ser encontrados na literatura [Huang et al., 2009]. Entretanto, uma análise puramente espacial não permite identificar localmente, isto é, ao longo de um arruamento específico, regiões de alta ou maior incidência de eventos pontuais. Por um lado, um conglomerado puramente espacial poderá abranger diversos arruamentos, sem que haja qualquer diferenciação com relação à contribuição dos eventos de cada arruamento (Figura 1). Como consequência, um trecho de um arruamento poderá ser caracterizado como crítico simplesmente porque a análise de conglomerado não faz distinção quanto a esta característica. É o caso, por exemplo, de um arruamento paralelo à uma avenida que apresenta alta incidência de eventos pontuais. Em particular, a caracterização de trechos críticos de ruas e avenidas permitirá aos órgãos responsáveis a criação de políticas de restrição como radares e melhorias de sinalização.

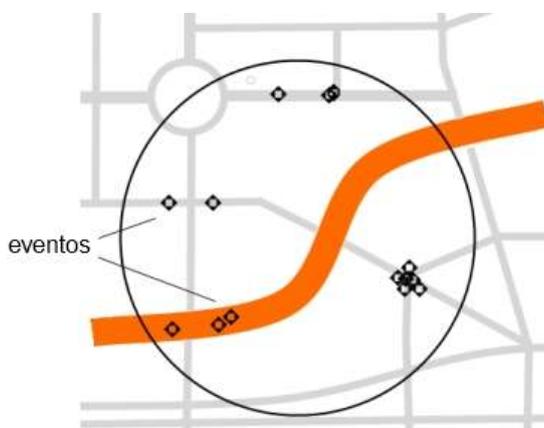


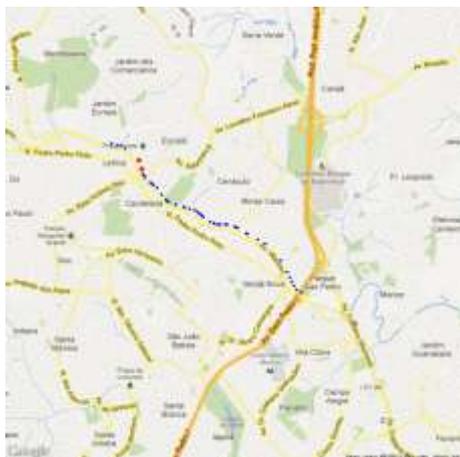
Figura 1. Limitação da técnica varredura espacial, encontrada na literatura, quando aplicada a arruamentos. Eventos fora do corredor de interesse são considerados, impedindo uma análise detalhada por corredor.

A metodologia apresentada foi desenvolvida a partir um projeto de pesquisa desenvolvido no Centro de Estudos de Criminalidade e Segurança da UFMG (CRISP) com

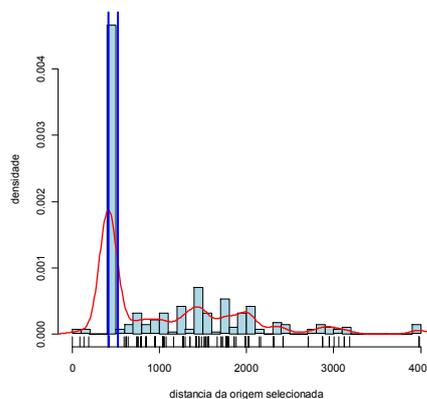
participação da Empresa de Transporte e Trânsito de Belo Horizonte (BHTRANS). Utilizando dados georeferenciados provenientes de acidentes de trânsito ocorridos no período de 2004 a 2010, foi desenvolvida uma plataforma para consulta, visualização e análises de dados em ambiente R. A plataforma, denominada RBHTrans possibilita ao usuário a consulta total ou parcial da base de dados e, a partir dos dados selecionados, disponibiliza funcionalidades de análise de mapas de kernel, moda espacial, análise descritiva de eventos de arruamentos e a estatística de varredura linear, denominada street scan. A plataforma utiliza os pacotes RgoogleMaps e Rgooglevis que possibilitam o acesso online a mapas da plataforma Google maps, além da possibilidade de exportar atributos georeferenciados para visualização em ambiente browser, como o Google Chrome ou Mozilla Firefox. Utilizando esta plataforma, o usuário pode realizar análises de arruamentos e visualizar os dados georeferenciados de acidentes sobrepostos a mapas de arruamento, satélite, ou mesmo visualizações utilizando o ambiente street view do Google maps.

2. A Estatística de Varredura Unidimensional

Seja s_i um par de coordenadas espaciais, $s_i = (x_i, y_i)$, atribuídas ao i -ésimo evento pontual espacial. Seja também $i \in \{1, \dots, N\}$, onde N é o número total de eventos pontuais em um determinado arruamento. Como exemplo, a figura 2.a mostra as coordenadas relativas a acidentes de trânsito ocorridos ao longo da Avenida Vilarinho em Belo Horizonte/MG, no ano de 2010.



(a) Eventos pontuais ao longo de um arruamento. Em vermelho estão indicados os eventos pertencentes a um conglomerado de alta incidência.



(b) Histograma da densidade de eventos pontuais. As linhas verticais delimitam um conglomerado atípico detectado pelo método descrito.

Figura 2. Eventos pontuais localizados ao longo da Avenida Vilarinho, Belo Horizonte, MG.

Para delimitar o arruamento de interesse, definem-se os eventos s_1 e s_N como o primeiro evento e o último evento, respectivamente, no arruamento. O interesse é detectar trechos entre os pontos s_1 e s_N que apresentam altas ou baixas intensidades de eventos pontuais significativos em termos estatísticos. Seja $d_{1,N}$ a distância euclidiana, no arruamento de interesse, entre os eventos s_1 e s_N e $d_{i,i+1}$ a distância entre dois pontos consecutivos s_i e s_{i+1} , tal que não exista nenhum outro ponto entre estes dois eventos. Considerando-se que os pontos consecutivos ao longo do arruamento são muito próximos, temos

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

Como consequência, a distância $d_{1,i}$ pode ser definida como a distância, ao longo do arruamento, do ponto s_1 para o ponto s_i . A Figura 3 mostra a distribuição de eventos pontuais ao longo de um arruamento e as distâncias relativas de cada evento ao ponto de origem, s_1 .

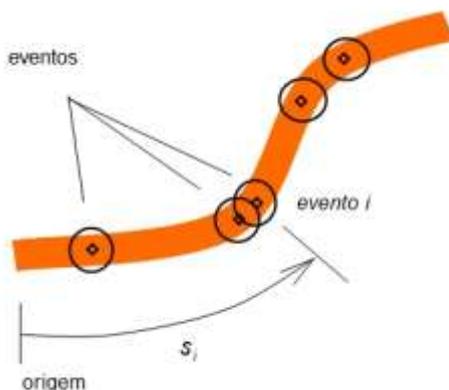


Figura 3. Distância dos i-ésimo evento à origem em um corredor.

Seja agora definida a hipótese nula de que os $N - 2$ eventos localizados entre os pontos s_1 e s_N ocorram de forma homogênea no trecho. Sob esta hipótese, a intensidade de eventos, λ_{H_0} , ao longo do trecho do arruamento é definida por:

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

Sob esta hipótese, o número de casos ocorridos ao longo de um trecho de comprimento Δd ($\Delta d < d_{1,N}$) pode ser modelado por variável de Poisson, $Y_{\Delta d} \sim Poisson(\mu_{\Delta d} = \lambda_{H_0} \cdot \Delta d)$.

É de particular interesse identificar automaticamente trechos ao longo do arruamento que apresentam uma intensidade maior de ocorrência de eventos do que a esperada sob a hipótese nula. Para isto, propomos o seguinte teste de razão de verossimilhança: seja l uma janela de comprimento variável, tal que $0 < l < d_{1,N}$. Seja c_l o número observado de eventos ao longo de l e μ_l o número esperado de casos ao longo de l sob a hipótese nula, ou $\mu_l = d_l \cdot \lambda_{H_0}$. A estatística do teste da razão de verossimilhança é obtida na forma:

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

A partir da Equação 3, é possível identificar o trecho \hat{l} que apresenta a maior ou menor incidência de eventos. Caso seja de interesse identificar somente trechos de alta

incidência, deve ser aplicada a restrição $c_l > \mu_l$. Para avaliar a significância do valor observado da estatística de teste em relação à hipótese nula, pode ser utilizada uma simulação de Monte Carlo, de acordo com o seguinte algoritmo:

1. Faça S simulações independentes com $N - 2$ eventos pontuais homogeneamente distribuídos ao longo de $s_{1,N}$ em cada simulação.
2. Para cada simulação calcule a estatística da razão de verossimilhança e forme o conjunto de valores $\hat{\kappa}_1, \dots, \hat{\kappa}_S$, onde $\hat{\kappa}_i$ é a estatística de teste obtida na i -ésima simulação.
3. Caso o valor observado da estatística de teste esteja acima do percentil $100(1 - \alpha)\%$ dos valores simulados, rejeita-se a hipótese nula. Neste caso pode-se dizer que o trecho \hat{l} detectado é crítico.

2.1 O modelo Bernoulli

O modelo de Poisson, apresentado anteriormente, supõe que a intensidade de eventos é constante ao longo do arruamento. Na prática, o 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 as 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 tempo t e $c_{l,t+1}$ o número de eventos observados ao longo de l no tempo $t+1$. 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 $N = \sum_t C_t$ 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:

$$L(\hat{l}) = \sup_{l, p > r} p^{c_{l,t}} (1-p)^{(n_l - c_{l,t})} r^{(C_t - c_{l,t})} (1-r)^{(N - C_t + c_{l,t})} \quad (4)$$

onde $\hat{p} = c_{l,t}/n_l$ e $\hat{r} = (C_t - c_{l,t})/(N - n_l)$. Semelhante ao modelo de Poisson, 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, 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, definida na equação 4, pode ser gerada.

4. Implementação Computacional

A estatística de varredura unidimensional para os modelos de Poisson e Bernoulli foi implementada no software R e utiliza os 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. A importação de mapas e sobreposição da imagem é obtida a partir da seguinte sequência de comandos:

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

onde *lonR* e *latR* são os limites de longitude e latitude do mapa a ser obtido, *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 and 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, como ilustrado na figura 4. Neste ambiente, a funcionalidade *street view* do *Google maps* pode ser utilizada para visualizar os dados ao longo do arruamento.



Figura 4. Visualização de dados pontuais no ambiente Google maps, utilizando a funcionalidade *gvisMap()* do pacote *googleVis*. Utilizando o ambiente *street view* é possível visualizar as coordenadas de conglomerados de acidentes de trânsito ao longo do arruamento de interesse.

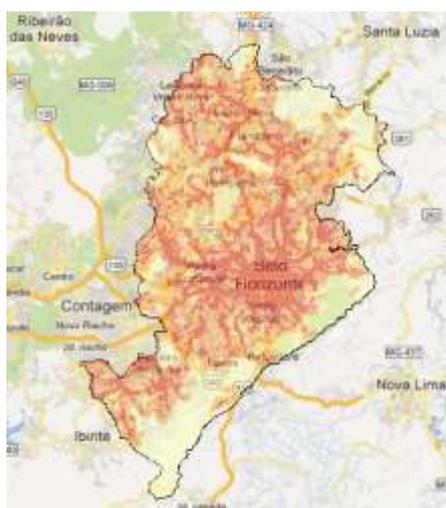
A metodologia de varredura unidimensional foi implementada na funcionalidade *street_scan()*. O procedimento de simulação de Monte Carlo, que apresenta grande custo computacional, foi implementado em linguagem C e incorporada ao ambiente R na forma de uma *dll* (*Dynamic-link library*) denominada *varredura.dll*. Foram criadas interfaces gráficas para a seleção de atributos do banco de dados bem como a seleção de parâmetros para as funcionalidades: (a) análise da intensidade de eventos em arruamentos, (b) mapa 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**. Dessa forma, todas as funcionalidades propostas são disponibilizadas a partir do comando:

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

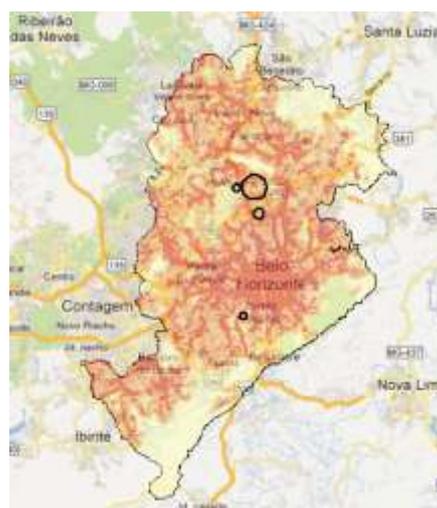
Na sequência, o usuário pode acessar as interfaces gráficas do ambiente a partir dos comandos: `monta_banco()` e `escolhe_funcao()`.

5. Exemplo de aplicação

A figura 5.a 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 corredores do município, que conduzem à área central do município. Na figura 5.b 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.



(a) Mapa de intensidade de pontos para os eventos referentes aos anos de 2010 e 2011



(b) Conglomerados de maior intensidade no ano de 2011 em relação aos eventos pontuais do ano de 2010.

Figura 5. Mapa de intensidade de eventos pontuais (a) e conglomerados identificados utilizando o modelo espacial de Bernoulli (b).

A figura 6 apresenta a análise de conglomerados utilizando a metodologia de arruamento, modelo Bernoulli, para os dados da Avenida Cristiano Machado. Foram utilizados os dados dos anos 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 todos os eventos pontuais da região de interesse. Comparando as figuras 5.b e 6, percebe-se que os conglomerados não são sobrepostos. Ou seja, a análise focada a um único arruamento permite detectar trechos do arruamento 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 esperado 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

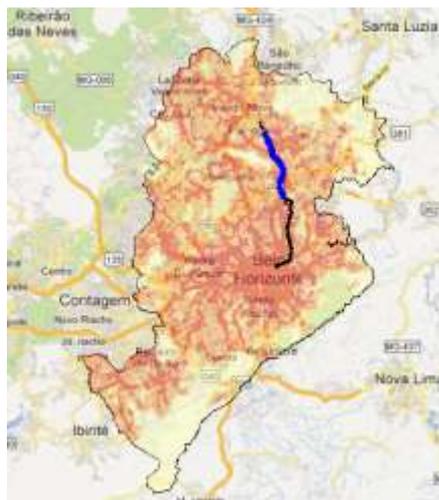


Figura 6. Conglomerado de maior incidência (linha azul) no ano de 2011 em comparação aos eventos pontuais do ano de 2010, considerando somente os dados da Avenida Cristiano Machado (linha preta e linha azul).

5. 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 inicialmente supõe que a taxa de incidência é 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 utilizar 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 caso, o método permite identificar trechos do arruamento onde o número atual de eventos é estatisticamente superior do que o 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 disponibilidade de análises puramente 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.** (1997). On the distributions of scan statistics of a two dimensional Poisson process, *Advances in Applied Probability*, vol. 29, pages 1–18.
- Costa, M. A. and Kulldorff, M.** (2009). In *Scan statistics: methods and applications*. Birkhäuser: Statistics for Industry and Technology, pages 129–52 [chapter 6].
- Costa, M. A. and Assunção, R. A. and Kulldorff, M.** (2012). Constrained spanning tree algorithms for irregularly-shaped spatial clustering. *Computational Statistics and Data Analysis*. vol. 56, pages 1771–1783.
- Duczmal, L. and Assunção, R. A.** (2004). Simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, *Computational Statistics and Data Analysis*, vol. 45, pages 269–286.
- Gesmann, Markus and de Castillo, Diego** (2011). Using the Google Visualisation API with R. *The R Journal*. vol. 3, n. 2, pages 40–44.
- Huang, L. and Stinchcomb, D. G. and Pickle, L. W. and Dill, J.** (2009). Identifying clusters of active transportation using spatial scan statistics. **American Journal of Preventive Medicine**. vol. 37, n. 2, pages 157–166.
- Kulldorff, M.** (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, vol. 26, pages 1481–1496.
- Kulldorff, M. and Athas, W. and Feuer, E. and Miller, B. and Key, C.** (1998). Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos. *American Journal of Public Health*, vol. 88, pages 1377–1380.
- Kulldorff, M.** (2001). Prospective time-periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society*, vol. A164, pages 61–72.
- Kulldorff, M. and Heffernan, R. and Hartman, J. and Assunção, R. M. and Mostashari, F.** (2005). A space-time permutation scan statistic for the early detection of disease outbreaks. *PLoS Medicine*, vol. 2, pages 216–224.
- Kulldorff, M. and Huang, L. and Pickle, L. and Duczmal, L.** (2006). An elliptic spatial scan statistic. *Statistics in Medicine*, vol. 25, pages 3929–3943.
- Loecher, Markus** (2010). *Plotting on Google Static Maps in R*. Technical Report.
- Diggle, P. J.** (1983). *Statistical Analysis of Spatial Point Patterns*.