

# **Estudo da intensidade espacial de roubos de veículos na cidade de São Paulo utilizando modelos para processos pontuais**

**Rafael Agostinho Ferreira <sup>1</sup>, Rodrigo Ferreira de Abreu <sup>2</sup>, João Domingos Scalon <sup>3</sup>**

## **1 Introdução**

Temas relacionados à segurança pública tem sido bastante discutido nos últimos anos, dado um grande aumento na criminalidade no Brasil ao longo dos anos. Dentro desse contexto, diversos tipos de casos podem ser classificados como crime, seja contra uma pessoa ou contra um patrimônio. Exemplos comuns são tráfico, homicídio, porte de arma de fogo, entre outros. No entanto, dentre todas estas causas, roubos de patrimônios e, mais precisamente, roubos de veículos são os tipos de crimes mais praticados nos últimos anos, de acordo com Fórum Brasileiro de Segurança Pública (2018).

De acordo com o Anuário Brasileiro de Segurança Pública publicado em 2018, no Brasil houve um total 273.339 registros de roubos de veículos durante o ano de 2016. Em 2017, esse número aumentou para 276.371 ocorrências. Tais quantidades segregadas por estados revelam que o estado de São Paulo apresentou o maior número de ocorrências em relação aos demais, com 77.948 e 67.964 registros em 2016 e 2017 respectivamente (FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA, 2018).

Um estudo estatístico sobre a distribuição espacial das ocorrências criminais de roubos de veículos podem ser de grande valia tanto na esfera pública, visando a elaboração de políticas públicas para o aprimoramento da segurança na sociedade, quanto na esfera privada, voltado para criação de produtos atuariais relacionados à seguro de automóveis. Isto é, a informação georreferenciada dos roubos pode auxiliar seguradoras a proporem diferentes faixas de valores de prêmios para seus produtos de acordo com o risco espacial envolvido. Este estudo pode ser feito usando técnicas de processos pontuais espaciais.

Assim, o objetivo deste trabalho é modelar a intensidade das ocorrências de roubos de veículos registradas na cidade de São Paulo durante o ano de 2018, a partir do modelo de Cox log-Gaussiano. O estudo foi realizado a partir de uma comparação dos modelos segregados pelo período do dia: pela madrugada, de manhã, à tarde e à noite. Além disso, foi realizado um estudo sobre áreas de risco para cada período a partir do modelo ajustado, baseado em simulações Monte Carlo via Cadeias de Markov.

---

<sup>1</sup>Universidade Federal de Lavras. e-mail: *rafaelferreiratc@hotmail.com*

<sup>2</sup>Universidade Federal de Lavras. e-mail: *rodrigo10mega@hotmail.com*

<sup>3</sup>Universidade Federal de Lavras. e-mail: *scalon@des.ufla.br*

## 2 Materiais e métodos

### 2.1 Metodologia

Os processos pontuais espaciais são definidos por um processo estocástico que governa a geração de eventos aleatórios sobre uma região espacial  $\mathbf{W}$ , definida no plano  $\mathbb{R}^2$ . Sua realização consiste em um conjunto irregular, contável e finito de eventos sobre  $\mathbf{W}$ , resultantes de um modelo particular (WALLER; GOTWAY, 2004).

Um processo de Cox é uma generalização dos modelos de Poisson não homogêneos, definidos como processos de Poisson duplamente estocásticos. Tal processo consiste em caracterizar a função de intensidade dos eventos como uma realização de outro mecanismo estocástico, o processo de intensidade  $\Lambda(\mathbf{s})$ . A distribuição dos eventos de um processo de Cox, condicionado à uma realização do processo de intensidade, se torna um processo de Poisson com função de intensidade  $\lambda(\mathbf{s})$  (MØLLER; SYVERSVEEN; WAAGEPELTERSEN, 1998). Processos de Cox são amplamente utilizados para a caracterização de modelos de intensidade, uma vez que, na maioria dos casos, não se conhece a forma estrutural da intensidade. Deste modo, um modelo probabilístico é atribuído à esta medida a fim de caracterizar a incerteza quanto à sua forma.

Uma particularidade associada à essa classe de modelos são os processos de Cox log-Gaussianos, que consiste em descrever o processo de intensidade em função de um processo Gaussiano latente  $Z(\mathbf{s})$ . A relação entre  $\Lambda$  e  $Z$  ocorre por meio de uma função de ligação exponencial, isto é,  $\Lambda(\mathbf{s}) = \exp\{Z(\mathbf{s})\}$ . Uma representação do modelo proposto por Diggle, Rowlingson e Su (2005), que descreve a intensidade estocástica do processo de Cox log-Gaussiano é dado por

$$\Lambda(\mathbf{s}) = \lambda(\mathbf{s}) R(\mathbf{s}), \quad (1)$$

que fatora o processo de intensidade em dois termos, sendo  $\lambda(\mathbf{s})$  descrita como uma componente determinística que representa a intensidade global do processo pontual e  $R(\mathbf{s})$  uma componente estocástica residual com média unitária, o qual assume a forma de  $\exp\{Z(\mathbf{s})\}$ , sendo  $Z(\mathbf{s})$  é um processo Gaussiano espacial estacionário e isotrópico, com média zero e função de correlação  $c(u; \phi)$  controlado pelo parâmetro  $\phi$ .

Para a estimação do termo determinístico,  $\lambda(\mathbf{s})$ , foi utilizado um método não paramétrico de alisamento por kernel, a fim de se obter uma estimativa suavizada para a intensidade espacial. Seu estimador é dado por

$$\hat{\lambda}(\mathbf{s}_0) = \frac{1}{\delta(\mathbf{s}_0)} \sum_{i=1}^n \frac{1}{h^2} k\left(\frac{\mathbf{s}_0 - \mathbf{s}_i}{h}\right), \quad (2)$$

em que  $\mathbf{s}_0$  é um ponto definido no espaço o qual deseja-se estimar a intensidade,  $\mathbf{s}_i$  representa a localização espacial do  $i$ -ésimo evento observado,  $k(\cdot)$  é a função kernel,  $h$  é

descrito como uma largura de banda que controla o alisamento da kernel na superfície espacial e  $\delta(\mathbf{s}_0)$  é um fator para correção de efeito de borda.

O método de mínimos contrastes foi utilizado no processo de estimação dos parâmetros associado à componente estocástica  $R(\mathbf{s})$ , isto é, a variância  $\sigma_Z^2$  e o parâmetro  $\phi$  relacionado ao modelo de correlação espacial do processo  $Z(\mathbf{s})$  (DAVIES; HAZELTON, 2013). Este método tem por objetivo obter estimativas para  $\sigma_Z^2$  e  $\phi$  que minimizam a diferença quadrática entre duas funções, uma empírica (baseada nos dados observados) e uma teórica (baseado em um modelo paramétrico). Estas funções descrevem o comportamento de dependência espacial do processo pontual, o qual está diretamente relacionada as propriedades do processo Gaussiano. Para o critério de minimização de  $\hat{\sigma}_Z^2$  e  $\hat{\phi}$  entre as funções empírica e teórica, foi utilizada a Função  $K$  de Ripley, proposta por Ripley (1976), sendo

$$D(\sigma^2, \phi) = \int_{u_0}^{u_{max}} \{v[\hat{K}(u)] - v[K_{\sigma^2, \phi}(u)]\}^2 du, \quad (3)$$

o critério de contraste, em que  $u_0$  e  $u_{max}$  são os limites inferior e superior do critério de contraste,  $v[\cdot]$  é alguma transformação feita em  $K$  e  $\hat{K}$ .

A fim de detectar regiões na área de estudo em que o risco de roubo seja elevado, foi realizada uma análise sob o comportamento da componente residual. Este estudo é feito com a finalidade de verificar a existência de regiões com grandes variações em torno da média  $\lambda(\mathbf{s})$ . Desta forma, regiões que possuem uma grande variabilidade podem representar possíveis áreas de risco de furtos de veículos.

Para isto, foi obtido amostras de uma distribuição aproximada de  $R(\mathbf{s})$ , condicionado aos eventos observados, a partir de métodos Monte Carlo via Cadeias de Markov (MCMC). O algoritmo utilizado neste estudo foi o algoritmo de Langevin-Hastings.

Uma vez obtida amostras da distribuição de  $R(\mathbf{s})$ , uma probabilidade aproximada de  $R(\mathbf{s})$  exceder um determinado limiar  $\epsilon$  foi medida, tomando a proporção das amostras obtidas de  $\hat{R}(\mathbf{s})$  o qual excedeu este valor  $\epsilon$  em uma determinada região.

## 2.2 Conjunto de dados

O conjunto de dados utilizados neste trabalho corresponde a localizações georreferenciada de ocorrências de roubos de veículos ocorridos na cidade de São Paulo - SP, durante o ano de 2018.

A base de dados foi obtida a partir do site da Secretaria de Segurança Pública do estado de São Paulo, que consiste em informações coletadas a partir de boletins de ocorrência registrados através do Registro Digital de Ocorrências, da Delegacia Eletrônica e da Polícia Militar.

Para a caracterização do evento espacial foi considerada a informação georreferenciada, em latitude e longitude, do local em que o roubo do veículo aconteceu. No entanto, as

coordenadas foram reprojatadas para o sistema de coordenadas UTM (em metros), a fim de que cálculos baseados em distâncias pudessem ser feitos.

Todo o processo de análise dos dados foi realizado utilizando o *software* estatística R (R CORE TEAM, 2019). Para a realização do ajuste do modelo de Cox log-Gaussiano, as análises foram feitas com o auxílio das bibliotecas *spatstat* (BADDELEY; RUBAK; TURNER, 2015) e *lgcp* (TAYLOR et al., 2013), que lidam com ferramentas exploratórias e de modelagem para configurações pontuais no espaço e espaço-tempo.

### 3 Resultados

Com relação a estimação da componente determinística, observa-se pela Figura 1 que a intensidade espacial das ocorrências de roubos na cidade de São Paulo tende a se intensificar ao longo do dia, sendo o período da noite o momento de maior intensidade. No entanto, essa intensidade se mostra não homogênea no município. Observa-se que a zona leste de São Paulo apresentou altos níveis de intensidade em relação às demais regiões. Uma possível justificativa para esse resultado é que, além da grande massa populacional se concentrar nesta área, algumas regiões são consideradas regiões mais carentes, o que reflete em uma intensidade elevada.

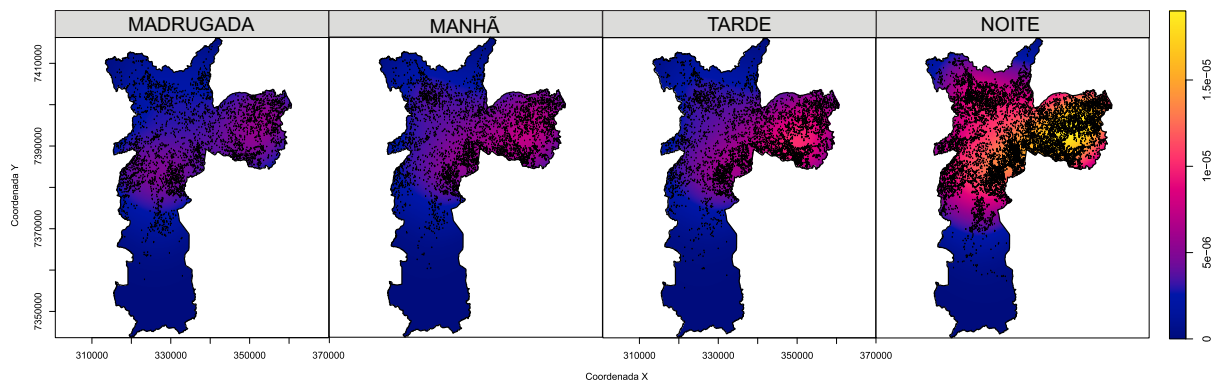


Figura 1: Estimativa de intensidade por Kernel de acordo com a periodicidade do dia.

Fonte: Do autor (2019).

Em relação aos parâmetros do modelo de covariância, foram obtidos, pelo método de mínimos contrastes, estimativas para  $\phi$  iguais à 1057,35, 1477,08, 1687,31 e 1728,41 e estimativas para  $\sigma_Z^2$  iguais a 1,45, 1,32, 1,51 e 1,19 para os períodos da madrugada, manhã, tarde e noite, respectivamente. Além disso, a Função  $K$  foi utilizada, para cada um dos períodos, como função descritora da dependência espacial do processo pontual de Cox. Em relação ao modelo teórico da correlação espacial do processo Gaussiano, utilizou-se a função exponencial.

Para a predição, foi utilizado uma simulação de 150.000 iterações, no qual as 50.000 primeiras foram queimadas (período de *burn-in*), obtendo-se, assim, uma taxa de aceitação

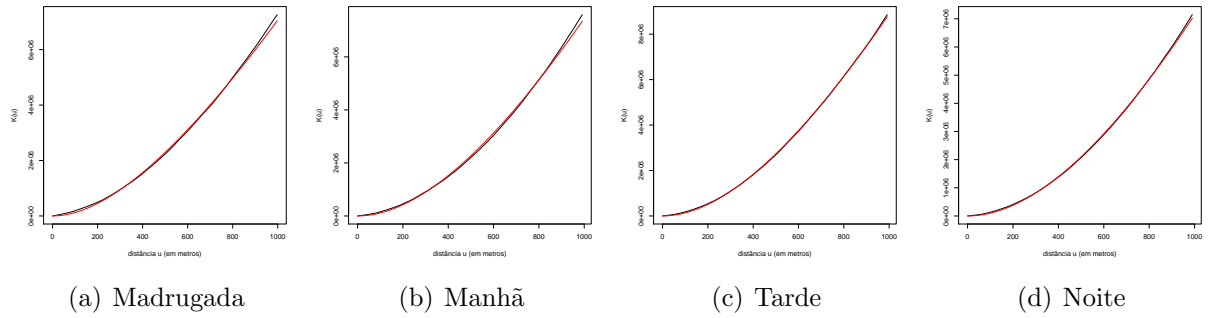


Figura 2: Função  $K$  estimada pelo método de mínimos contrastes para cada período.  
Fonte: Do autor (2019).

de 0,574. Com relação a precisão desta predição, foi observado o desvio-padrão de  $R(s)$ . Além disso, observou-se erros não elevados, o que revela um bom ajuste do modelo e da predição.

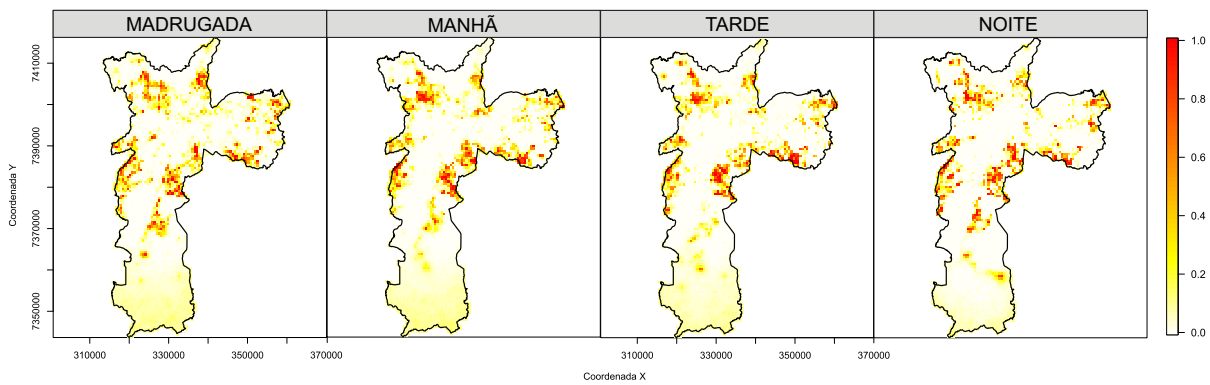


Figura 3: Mapas de risco de roubos de veículos de acordo com o período do dia.  
Fonte: Do autor (2019).

A Figura 3 apresenta os mapas de probabilidade de excedência obtidos para os quatro períodos do dia. Observou-se algumas regiões da cidade de São Paulo com altas probabilidades, o que caracteriza um alto risco de roubo de veículos. Estas regiões se concentram nas áreas periféricas da cidade, abrangendo alguns bairros que são considerados de baixa renda. Foram elas: Lajeado, Cidade Tiradentes, Jardim Helena, Iguatemi, Itaim Paulista e São Rafael, ambos situados na zona leste de São Paulo. Em relação à zona sul, observou-se que os bairros Marsilac, Jardim Ângela, Parelheiros, Grajaú também apresentaram altas probabilidades. Já na zona oeste de São Paulo, os bairros Jaguará, Jaraguá, São Domingos, Pirituba e Perus também receberam altas probabilidades.

## 4 Considerações finais

O modelo de Cox log-Gaussiano se mostrou como um bom modelo para a detecção de regiões onde o risco de ocorrer um roubo de um veículo é elevado na cidade de São Paulo. No entanto, a mesma técnica pode ser utilizada para estudos de riscos em outras localizações, podendo, assim, contribuir para a tomada de decisões por parte dos órgãos competentes a partir do conhecimento do comportamento espacial das ocorrências.

## Agradecimentos

Os autores agradecem a CAPES e a Fapemig pelo apoio financeiro.

## Referências

- BADDELEY, A.; RUBAK, E.; TURNER, R. *Spatial point patterns: Methodology and applications with R*. Florida: Chapman and Hall/CRC, 2015. 810p.
- DAVIES, T. M.; HAZELTON, M. L. Assessing minimum contrast parameter estimation for spatial and spatiotemporal log-Gaussian Cox processes. *Statistica Neerlandica*, v. 67, n. 4, p. 355–389, 2013.
- DIGGLE, P. J.; ROWLINGSON, B.; SU, T. li. Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics*, v. 16, n. 5, p. 423–434, 2005.
- FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. *Anuário Brasileiro de Segurança Pública*, n. 12, 2018. ISSN 1983-7364.
- MØLLER, J.; SYVERSVEEN, A. R.; WAAGEPETERSEN, R. P. Log-Gaussian Cox processes. *Scandinavian journal of statistics*, Wiley Online Library, v. 25, n. 3, p. 451–482, 1998.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2019. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org>.
- RIPLEY, B. D. The second-order analysis of stationary point processes. *Journal of applied probability*, JSTOR, v. 13, n. 2, p. 255–266, jun. 1976.
- TAYLOR, B. M. et al. lgcp - An R Package for inference with Spatio-Temporal Log-Gaussian Cox Processes. *Journal of Statistical Software*, v. 52, n. 4, jan 2013.
- WALLER, L. A.; GOTWAY, C. A. *Applied Spatial Statistics for Public Health Data*. Hoboken: John Wiley & Sons, 2004. v. 368.