

Mapeamento do risco de homicídio com base na co-krigeagem binomial e simulação: um estudo de caso para São Paulo, Brasil

Mapping of homicide risk with binomial cokriging and simulation: a case study in São Paulo, Brazil

Eduardo Celso Gerbi Camargo ¹
Suzana Druck ²
Antônio Miguel Vieira Monteiro ¹
Corina Costa Freitas ¹
Gilberto Câmara ¹

Abstract

The impact of violence on the population's health profile in large Brazilian cities is a real concern and has been incorporated into the country's public health agenda. Although violent crimes affect the entire population, they are distributed unequally in the city. In this context, analytical tools to evaluate the risk and spatial distribution of homicide are important for surveillance and prevention. The current article is a contribution in this direction. A geostatistical methodology is used to estimate and map homicide risk, and a new semivariogram estimator is presented and its application is evaluated in the city of São Paulo, Brazil, for the period 2002 to 2004. Homicide risk scenarios are generated using binomial cokriging and stochastic simulation procedures, allowing the demarcation of areas with greater or lesser risk that can be used to study spatial determinants of violence and improve the understanding of problems through analysis of differences across urban territories.

Homicide; Spatial Analysis; Risk Assessment

Introdução

O impacto da violência sobre o perfil de saúde da população nos grandes centros urbanos brasileiros é uma realidade e tem sido incorporado na agenda das ações de saúde. O homicídio é a principal causa de mortes violentas dentro do capítulo das causas externas, o que é sustentado pela análise dos registros encontrados no Sistema de Informações sobre Mortalidade (SIM) do Departamento de Informática do SUS (DATASUS) ¹. Somente a cidade de São Paulo, Brasil, registrou para o período de 1998 a 2002 mais de 31 mil mortes por homicídio, correspondendo em torno de 57% do total de ocorrências da região metropolitana. Embora o crescimento dos crimes contra a vida seja sentido por toda a população, a sua distribuição pela cidade ocorre de modo desigual. Nesse contexto, ferramentas de análise que permitam produzir uma avaliação do risco de homicídio e de sua distribuição espacial potencializam os meios de vigilância e, conseqüentemente, possibilitam fornecer informações importantes para o desenho de políticas de promoção da saúde considerando novas estratégias de controle e prevenção.

A reflexão brasileira sobre o impacto da violência sobre a saúde, mesmo que ainda incipiente, avançou muito nos últimos 12 anos ². Os novos métodos, que têm incorporado o espaço e o tempo como variáveis de análise, com base na possibilidade de localização para os dados de

¹ Divisão de Processamento de Imagens, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brasil.

² Empresa Brasileira de Pesquisa Agropecuária, Brasília, Brasil.

Correspondência

E. C. G. Camargo
Divisão de Processamento de Imagens, Instituto Nacional de Pesquisas Espaciais.
Av. dos Astronautas 1758,
São José dos Campos, SP
12227-010, Brasil.
eduardo@dpi.inpe.br

homicídio, sistematizados por meio da classificação “causas externas”^{3,4,5,6,7,8}, vêm contribuindo de maneira consistente e firme para melhorar a fase de diagnóstico de situações e estabelecer o avanço necessário, em bases operativas, para o apoio aos serviços de saúde e ao planejamento das ações. Visando a contribuir nessa direção, este trabalho apresenta uma alternativa para a modelagem do risco de homicídio. A base teórica do modelo empregado se baseia em uma abordagem geoestatística e tem sua aplicação avaliada na cidade de São Paulo para o período de 2002 a 2004, baseando-se nas taxas de homicídio observadas nos 96 distritos que compõem o município.

A aplicação da abordagem geoestatística possibilita estimar o risco de homicídios de forma contínua com base nas taxas de ocorrência de homicídio agregadas por área. Isso conduz para uma visão alternativa de como esta “epidemia” se espalha por toda a cidade, que pode ser empregada para orientar desde a formulação de políticas sociais até o planejamento de intervenções, informando as ações de monitoração, vigilância e controle. Diferente, por exemplo, dos métodos Bayesianos empíricos (global e local)^{9,10,11}, ou então, da aplicação dos indicadores de Moran & Geary^{12,13}, que resumidamente consideram a distribuição do fenômeno investigado homogêneo dentro de cada área componente da região de estudo, ou seja, mudanças importantes só ocorrem nas fronteiras destas áreas. Tais métodos são úteis também para este tipo de análise, mas diferem extremamente quanto às hipóteses subjacentes a respeito do padrão espacial e da distribuição dos valores do risco. Uma outra possibilidade da abordagem geoestatística é o emprego de procedimentos de simulação condicionada. Isso permite a construção de cenários para o risco de homicídios em um nível de probabilidade desejado, possibilitando a delimitação de áreas com maior ou menor risco, que podem ser usados para estudar condicionantes espaciais da violência e auxiliar para um melhor entendimento dos problemas segundo a análise de suas diferenças ao longo dos territórios das cidades.

A implementação tradicional dos métodos geoestatísticos, no entanto, não acomodam a heterocedasticidade das taxas de doenças e contagens; isto é, a variância em cada lugar assume valores em função do tamanho da população¹⁴. Então, alternativas para o estimador de semivariograma e algoritmos de krigeagem apresentados por Matheron necessitam ser desenvolvidos para considerar a natureza específica dos dados de saúde.

A literatura geoestatística apresenta algumas soluções para o problema da não estacionariedade da variância causada pela variação do tama-

nho da população em risco, as quais podem ser vistas em Cressie¹⁵, Kelsall & Wakefield¹⁶, Berke¹⁷ e Goovaerts et al.¹⁸. Neste trabalho, seguimos a proposta inicial de Lajaunie¹⁹, que desenvolve a função semivariograma do risco. Posteriormente, Oliver et al.²⁰ apresentam uma alternativa para o estimador de semivariograma do risco e aplicam conjuntamente com a co-krigeagem binomial para estimar o risco de câncer em crianças no centro-oeste da Inglaterra.

Materiais e métodos

Parte dos algoritmos geoestatísticos necessários para modelagem do risco de homicídio na cidade de São Paulo foram implementados no programa MATLAB-6.5 (Mathworks, Natick, Estados Unidos; <http://www.mathworks.com>) e utilizados conjuntamente com a biblioteca de programas geoestatísticos, *Geostatistical Software Library* (GSLIB)²¹. Os resultados obtidos foram exportados para o Sistema de Processamento de Informações Georreferenciadas (SPRING)²², que conta com bases cartográficas dos distritos censitários do município, para visualização e produção de mapas e cenários do risco de homicídio. Seguindo a recomendação de vários estudos sobre esquemas de cores em mapas^{23,24}, um esquema de cor *double-ended* foi utilizado: um gradiente de vermelho é usado para indicar áreas de alto risco, maior que a média, e um de azul para indicar áreas de baixo risco.

Área de estudo

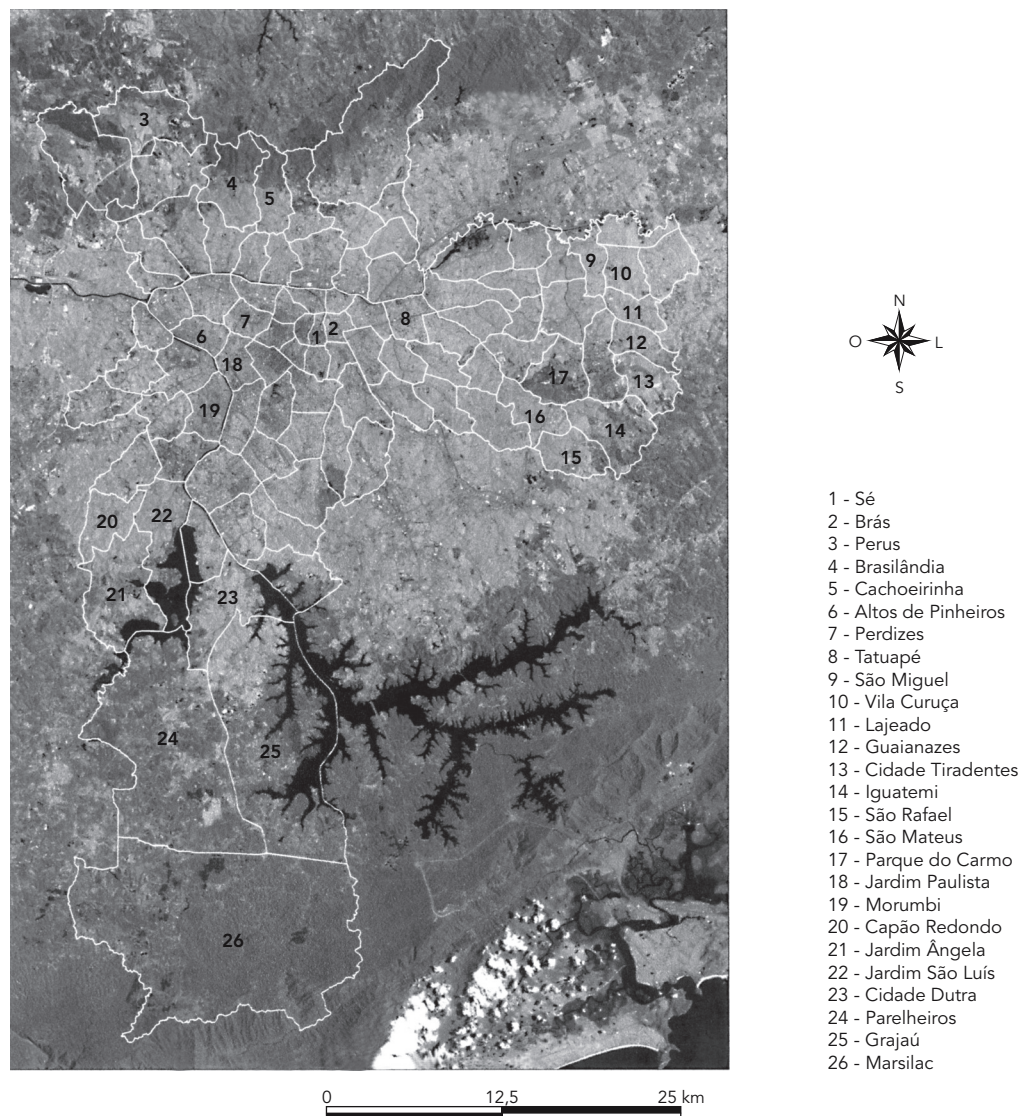
A região de estudo refere-se à cidade de São Paulo, Figura 1, que ocupa uma área de apenas 0,02% do território brasileiro. Segundo o *Censo Demográfico* de 2000 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE; <http://www.ibge.gov.br>)²⁵, o município apresenta alguns números que o coloca entre as maiores cidades do mundo, tais como: área total de 1.524km², população de aproximadamente 10,5 milhões de habitantes, divisão territorial composta por 96 distritos, densidade populacional 6.885 habitantes/km², número de ruas e avenidas igual a 48.840, bairros e vilas perfazendo 1.544. Sua localização está compreendida entre as seguintes coordenadas UTM: na direção Leste-oeste de 310.854m a 362.716m, e na direção Norte-sul de 7.418.926m a 7.340.836m.

Os dados

Os registros dos homicídios foram disponibilizados pelo Programa de Aprimoramento das Infor-

Figura 1

Mapa da região de estudo com destaque dos 96 distritos da cidade de São Paulo, Brasil.



mações de Mortalidade (PROAIM) ²⁶ no Município de São Paulo. Os registros processados referem-se aos óbitos de residentes e que ocorreram no município. Por razões confidenciais, esses dados são agregados por distritos e disponibilizados na forma de contagens (número de ocorrências). Os dados de contagem foram utilizados conjuntamente com dados de pessoas (número de habitantes por área), sendo possível calcular as taxas de óbitos por 100 mil habitantes. Por fim, essas taxas foram associadas aos centróides dos

distritos, constituindo a informação observada do fenômeno em estudo.

Método

- O modelo

Considere uma região de estudo A , composta de N áreas. Associado a cada área tem-se o número de casos de ocorrência de homicídio e a população em risco, denotados por $l(\mathbf{u}_i)$ e $n(\mathbf{u}_i)$,

respectivamente, em que o vetor de coordenadas espaciais $\mathbf{u}_i = (x_i, y_i)$, $i = 1, \dots, N$, refere-se geograficamente ao centróide da i -ésima área. A variável aleatória $Z(\mathbf{u}_i)$ é denominada de taxa de homicídio e $z(\mathbf{u}_i)$, estabelecida pela razão $l(\mathbf{u}_i)/n(\mathbf{u}_i)$, é uma das possíveis realizações da variável aleatória $Z(\mathbf{u}_i)$.

Supõe-se que existe um risco de homicídio, denotado por $R(\mathbf{u})$, decorrente de um processo estocástico $\{R(\mathbf{u}), \mathbf{u} \in A, A \subset \mathbb{R}^2\}$ e que todo indivíduo em A está a ele submetido. Portanto, $R(\mathbf{u})$ é uma variável aleatória contínua e espacialmente correlacionada, cujos valores não são diretamente observados.

Nesse ponto, é importante notar que há dois suportes geográficos distintos. Um refere-se às áreas geográficas que compõem a região de estudo A , para as quais existem valores observados $z(\mathbf{u}_i)$ da variável aleatória $Z(\mathbf{u}_i)$. O outro se refere à natureza do processo investigado, isto é, uma superfície contínua de $R(\mathbf{u})$ que se deseja estimar. Uma discussão mais ampla da combinação de informações espaciais residindo em suportes geográficos distintos pode ser vista em Waller & Gotway²⁷. Para estabelecer o relacionamento entre taxa de homicídio e o risco subjacente a ser inferido é necessário que ambas as informações estejam sobre o mesmo suporte geográfico. Para isso, denota-se por $R(\mathbf{u}_i)$, $i = 1, \dots, N$, o risco médio associado ao centróide da i -ésima área (distrito).

Supõe-se também que os diferentes casos de mortes por homicídio ocorrem independentemente quando o risco é fixado (em outras palavras, o risco é a única fonte de correlação entre os casos). Assim, as taxas de ocorrência de homicídio são variáveis com distribuição¹⁹:

$$Z(\mathbf{u}_i) | R \sim \frac{1}{n(\mathbf{u}_i)} \text{Bi}[r(\mathbf{u}_i), n(\mathbf{u}_i)] \quad (1),$$

em que: Bi representa uma distribuição binomial.

- **Estimador proposto para o semivariograma do risco**

Com base no modelo expresso na equação (1), Oliver et al.²⁰ apresentam uma alternativa para o estimador de semivariograma do risco, o qual foi concebido para casos em que as áreas componentes da região de estudo são relativamente homogêneas (em área e população), como foi o estudo de caso apresentado em seu trabalho.

Para regiões de estudos que apresentam áreas heterogêneas (agregação de grupos sociais distintos, diferenças em população e área), que é a realidade das grandes cidades brasileiras, a alternativa apresentada por Oliver et al.²⁰ pode ser em alguns casos limitada, porque propõe modelar um processo que supõe ser homogêneo,

com média e variância do risco constantes sobre toda a região de estudo, quando na realidade pode não ser. O fenômeno em estudo pode apresentar tendências, isto é, zonas de baixo e alto riscos em localizações específicas dentro da região de estudo. Além disso, a instabilidade que se observa nos dados de taxa, decorrente de áreas com populações pequenas^{28,29}, pode interferir diretamente na estimativa da estrutura de correlação espacial do risco, conduzindo em análises errôneas sobre a variabilidade espacial do fenômeno em estudo.

Para contornar os problemas mencionados acima, duas modificações na estimação dos parâmetros que compõem o semivariograma do risco são propostas: a primeira é incorporar o impacto do tamanho da população na estimação do semivariograma do risco. A segunda é a incorporação de médias e variâncias zonais para tratar o problema de tendência que se observa nos dados. Para tal, a área em estudo é dividida inicialmente em W zonas de risco, supostamente homogêneas.

O estimador proposto do semivariograma de risco é denotado por $\hat{\gamma}_{(\mathbf{u}_i, \mathbf{u}_j)}^{*R}$, levando à seguinte notação:

$$\hat{\gamma}_{(\mathbf{u}_i, \mathbf{u}_j)}^{*R} = \hat{\gamma}_{(\mathbf{u}_i, \mathbf{u}_j)}^{*Z} - \frac{1}{2} [\hat{\mu}_w(1 - \hat{\mu}_w) - \hat{\sigma}_{R_w}^2] \left[\frac{n(\mathbf{u}_i) + n(\mathbf{u}_j)}{n(\mathbf{u}_i) n(\mathbf{u}_j)} \right] \quad (2),$$

em que:

$$\hat{\gamma}_{(\mathbf{u}_i, \mathbf{u}_j)}^{*Z} = \frac{1}{2 \sum_{i,j=1}^{M(\mathbf{u}_i, \mathbf{u}_j)} n(\mathbf{u}_i) n(\mathbf{u}_j)} \sum_{i,j=1}^{M(\mathbf{u}_i, \mathbf{u}_j)} \{n(\mathbf{u}_i) n(\mathbf{u}_j) [z(\mathbf{u}_i) - z(\mathbf{u}_j)]^2\}$$

é denominado estimador de semivariograma empírico ponderado pela população, em que $M(\mathbf{u}_i, \mathbf{u}_j)$ refere-se ao número total de pares de pontos disponíveis para uma certa distância de análise (*lag*).

$$\hat{\mu}_w^* = \frac{1}{n_w} \sum_{i=1}^{n_w} z(\mathbf{u}_i) \quad \text{e} \quad \hat{\sigma}_{R_w}^{*2} = \sum_{i=1}^{n_w} \frac{[z(\mathbf{u}_i) - \hat{\mu}_w^*]^2}{n_w - 1},$$

são escolhidos adequadamente dentre as W médias e variâncias zonais do risco, dependendo das posições \mathbf{u}_i e \mathbf{u}_j ; em que $w = 1, \dots, W$ representa as zonas de risco e n_w refere-se ao número de observações $z(\mathbf{u}_i)$ contidos em w . O critério para estabelecer as zonas de risco e suas respectivas médias e variâncias não é único, depende da informação disponível e da tendência observada nos dados. Mais adiante na seção de resultados é descrito em detalhes o critério utilizado neste trabalho para a definição de tais zonas de risco, e como suas respectivas médias e variâncias foram empregadas no cálculo de $\hat{\gamma}_{(\mathbf{u}_i, \mathbf{u}_j)}^{*R}$ para o estabelecimento da estrutura de correlação espacial do risco de homicídio.

$$\left[\frac{n(\mathbf{u}_i) + n(\mathbf{u}_j)}{n(\mathbf{u}_i) n(\mathbf{u}_j)} \right] = \left[\frac{n(\mathbf{u}_i) + n(\mathbf{u}_i + \mathbf{h})}{n(\mathbf{u}_i) n(\mathbf{u}_i + \mathbf{h})} \right]:$$

refere-se à média de todos os pares de localizações envolvidos no cálculo do vetor distância \mathbf{h} , calculado da seguinte forma:

$$\frac{1}{M(\mathbf{h})} \sum_{i=1}^{M(\mathbf{h})} \frac{n(\mathbf{u}_i) + n(\mathbf{u}_i + \mathbf{h})}{n(\mathbf{u}_i) n(\mathbf{u}_i + \mathbf{h})},$$

em que $M(\mathbf{h}) = M(\mathbf{u}_i, \mathbf{u}_j)$ conforme definido anteriormente.

- **Co-krigeagem binomial**

A co-krigeagem binomial é um interpolador linear, que é empregado para estimar o risco em uma localização qualquer dentro da área de estudo, denotada por \mathbf{u}_0 , a partir de uma combinação linear de K taxas vizinhas $z(\mathbf{u}_i)$, da seguinte forma²⁰:

$$\hat{R}(\mathbf{u}_0) = \sum_{i=1}^K \lambda_i(\mathbf{u}_i) z(\mathbf{u}_i) \quad (3),$$

em que: K refere-se ao número de centróides considerados no cálculo de $\hat{R}(\mathbf{u}_0)$. Em geral, somente os centróides mais próximos de \mathbf{u}_0 são considerados, e λ_i é o peso atribuído à i -ésima observação $z(\mathbf{u}_i)$.

O diferencial dos interpoladores geoestatísticos para outros métodos de interpolação é a forma com que os pesos são calculados. Por exemplo, no caso do interpolador determinístico de média simples os pesos são todos iguais a $1/K$ (K = número total de amostras). Outro exemplo, no interpolador determinístico baseado no inverso do quadrado das distâncias, os pesos são calculados em função do inverso do quadrado da distância entre os valores observados e a localização a ser estimada. No caso da co-krigeagem, os pesos λ_i são calculados usando-se um sistema de equações^{19,20,30}, que considera as covariâncias diretas entre $Z(\mathbf{u}_i)$ e $Z(\mathbf{u}_j)$, e as covariâncias cruzadas entre $Z(\mathbf{u}_i)$ e o risco $R(\mathbf{u}_0)$; e duas propriedades básicas: não tendenciosidade e variância mínima.

O resultado da aplicação do método de co-krigeagem binomial, em geral, é uma superfície que representa a média da distribuição espacial do risco, mostrando suas principais tendências, ou então, a média do risco em uma localização específica dentro da região de estudo.

- **Avaliação da estrutura de correlação espacial do risco**

A qualidade e a precisão dos mapas gerados pelo interpolador de co-krigeagem binomial depende substancialmente da estrutura de correlação espacial do risco estimada. Nesse sentido, antes de aplicar a co-krigeagem binomial é importante avaliar se a estrutura estimada é adequada ou não. Neste trabalho, isso é realizado por meio de um estudo de simulação. A

idéia é gerar um conjunto de mapas simulados da variável aleatória $Z(\mathbf{u}_i)$, os quais representam as possíveis realizações do que poderia ser a informação observada. Para cada mapa simulado, o estimador proposto para o semivariograma do risco é aplicado, constituindo-se desta forma a distribuição simulada para o semivariograma do risco. Finalmente, uma análise é conduzida para verificar o comportamento da estrutura de correlação espacial do risco estimada *versus* a simulada, conforme apresentada mais adiante na seção de resultados.

- **Construção de cenários do risco de homicídio**

Conforme mencionado anteriormente, a co-krigeagem binomial quando aplicada fornece estimativas da média da distribuição do risco de homicídio. No entanto, existem outras situações, normalmente para fins de planejamento, em que se exige avaliar a probabilidade do campo aleatório em estudo exceder um dado valor de corte (em outras palavras, o que se pretende é avaliar quantos daqueles valores no espaço excedem simultaneamente um dado valor limite e qual a probabilidade de isto acontecer). Para satisfazer essa necessidade, procedimentos de simulação seqüencial condicionada devem ser empregados.

Neste trabalho a construção de cenários para o risco de homicídio é estabelecida por simulação seqüencial condicionada por indicação (SSCI), um tipo de simulação condicionada não-paramétrica, conforme descrito em Goovaerts³¹ e outras literaturas correlatas. A razão dessa escolha é que esse procedimento de simulação não impõe nenhum tipo de distribuição de probabilidade *a priori* para a variável aleatória $R(\mathbf{u})$. Inicialmente o risco é estimado próximo dos centróides das áreas componentes da região de estudo por co-krigeagem binomial. Em seguida, aplica-se a SSCI para gerar um conjunto de realizações equiprováveis do campo aleatório $R(\mathbf{u})$ (em outras palavras, para gerar um conjunto de mapas simulados do risco de homicídios). Tomando o conjunto de realizações simuladas, tem-se, para cada localização \mathbf{u} da malha do mapa espacial, um conjunto de valores simulados $r(\mathbf{u})$ da variável aleatória $R(\mathbf{u})$. A partir desse conjunto, a função de distribuição acumulada de $R(\mathbf{u})$ é estabelecida, possibilitando o cálculo de vários valores de cortes e, posteriormente, a construção de cenários para o risco de homicídio. Por exemplo, valores de cortes obtidos de decís 0,1 e 0,9 da distribuição acumulada de $R(\mathbf{u})$ podem representar, de acordo com o fenômeno em estudo, cenários otimistas e pessimistas, respectivamente.

Resultados e discussão

Análise preliminar dos dados

Uma análise preliminar usando-se estatísticas descritivas das taxas de homicídios, agregadas aos 96 distritos da cidade de São Paulo, mostra que a taxa média de homicídios por 100 mil habitantes decresce no período 2002-2004. Em 2002 a taxa média de homicídios registrada na cidade de São Paulo foi da ordem de 41 por 100 mil habitantes, contra uma taxa de 38 homicídios por 100 mil habitantes em 2003 e 30 por 100 mil habitantes em 2004, portanto um decréscimo da ordem de 27%. Entretanto, essas estimativas ainda são extremamente elevadas quando comparadas às de outras cidades de países desenvolvidos. Por exemplo, no triênio 2002-2004, a taxa média de homicídio em Nova York por 100 mil habitantes foi da ordem de 8,77; em Tóquio, uma das mais baixas do mundo, 1,24; e Londres 2,59. Observa-se também uma diminuição nas variâncias estimadas no mesmo período, indicando um provável decréscimo das diferenças entre os distritos com valores extremos de taxas de homicídio. Os

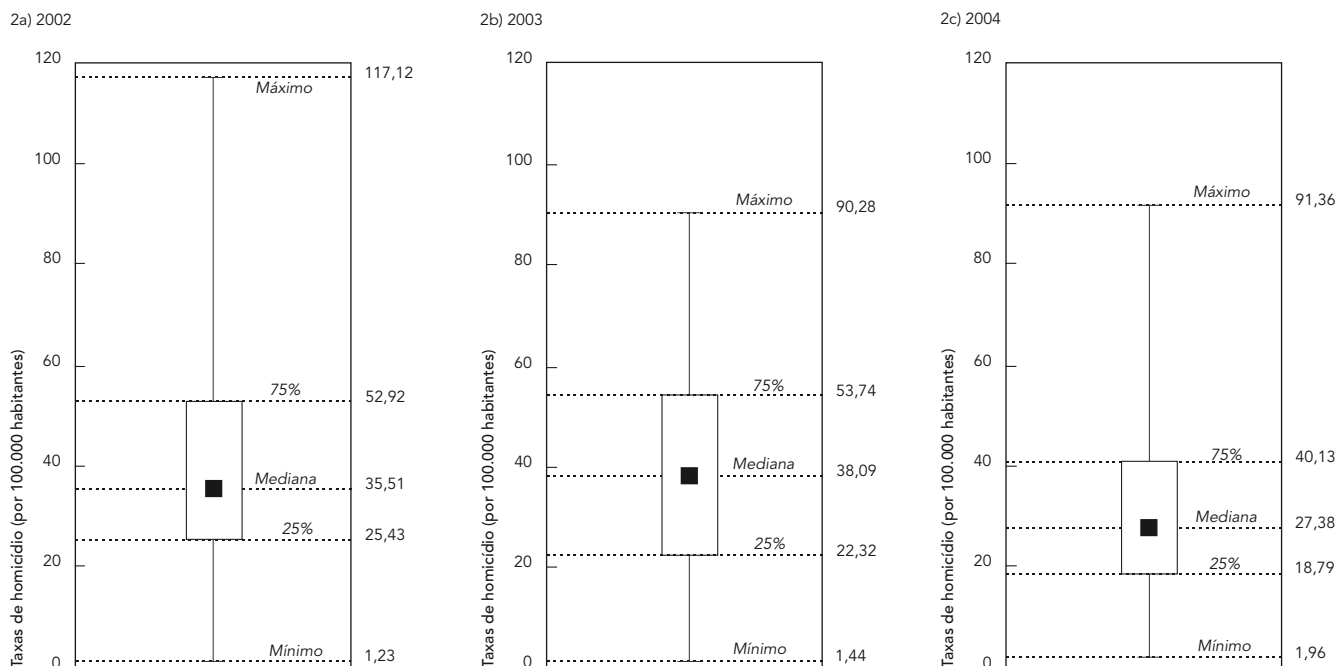
resultados dessa análise são apresentados nas Figuras 2a, 2b e 2c.

Para verificar a dependência da variância das taxas de homicídios ao tamanho da população, a análise é conduzida por meio de um gráfico que mostra os valores das taxas de homicídios *versus* o número de habitantes para os distritos da cidade de São Paulo. Para cada ano investigado, os resultados obtidos mostram que quanto menor o tamanho da população maior a variabilidade da taxa observada. Conforme vários estudos^{28,29}, esse comportamento é típico de taxas em geral. Desse modo, as flutuações extremas, isto é, os valores mais altos e mais baixos, ocorrem nas áreas de menores populações, sem ter associação com os riscos associados àquelas áreas. Isso tende a produzir um efeito, denominado de “efeito funil”, o qual está realçado por linhas pontilhadas nas Figuras 3a, 3b e 3c.

Para explorar a variação da tendência espacial dos dados utilizou-se o método denominado Média Móvel Espacial, conforme descrito em Bailey³². Resumidamente, esse método reduz a variabilidade espacial, pois tende a produzir resultados com menor flutuação que os dados

Figura 2

Estatísticas descritivas das taxas de homicídio.



originais, fornecendo uma visão das grandes tendências do fenômeno em estudo. Analisando as distribuições dos homicídios pelos distritos da capital, conforme mostradas nas Figuras 4a, 4b e 4c, verifica-se facilmente que o fenômeno investigado apresenta um forte gradiente centro-periferia.

Finalizando esta primeira análise, deve ser ressaltado que os problemas detectados, a instabilidade das taxas e sua tendência, devem ser adequadamente tratados pelo método de modelagem escolhido, a fim de se evitar falsas estimativas do risco de homicídios.

Definição das zonas de risco

Inicialmente cabe lembrar que a definição das zonas de risco faz-se necessária, uma vez que a configuração heterogênea da região de estudo apresenta médias e variâncias zonais distintas. Neste trabalho, para cada ano investigado, zonas de risco foram estabelecidas empiricamente baseando-se em um valor de corte da distribuição acumulada das taxas de ocorrência de homicídio, como sendo o percentil de 75%. Esse valor de corte reflete a configuração de tendência observada nos dados, possibilitando a produção de dois estratos de risco diferenciado, em que 72 distritos em torno da região central foram classificados como de baixo risco e 24 do centro e periferia como de alto risco. A Tabela 1 sumariza as principais características dessas duas zonas de risco.

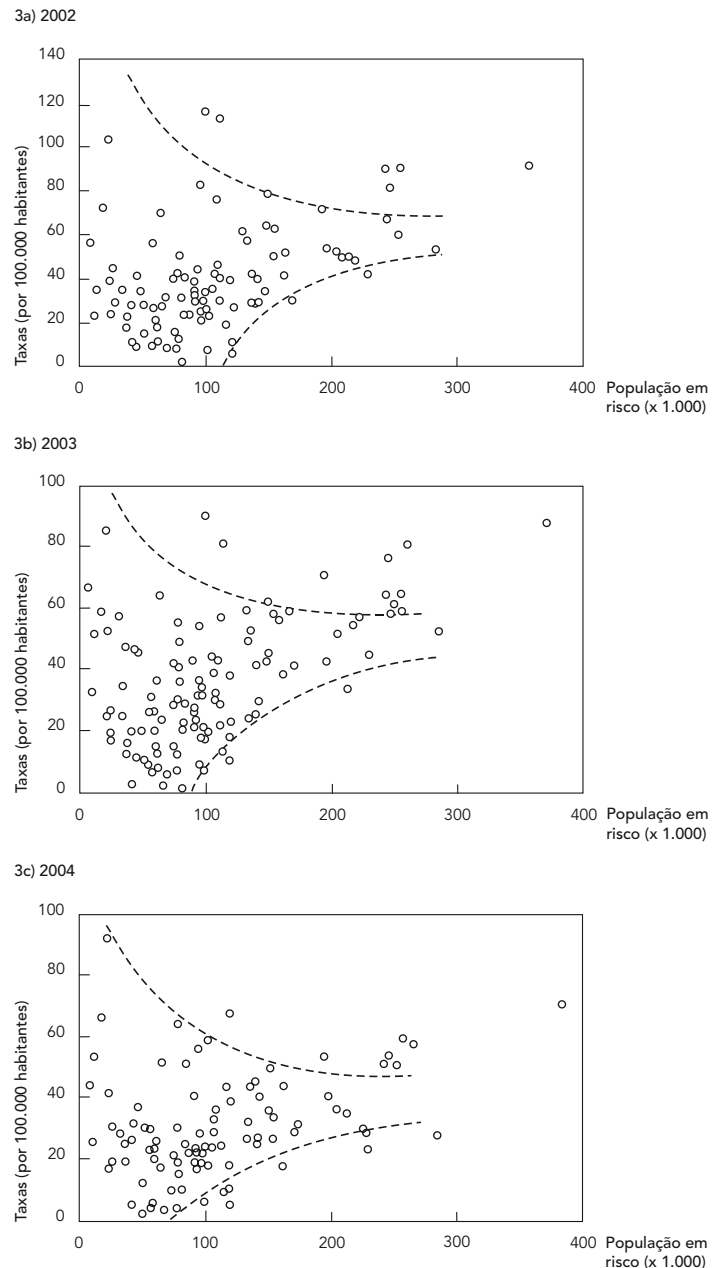
Estrutura de correlação espacial do risco de homicídio

Baseando-se na estratificação da região de estudo em duas zonas de risco, o estimador proposto para o semivariograma do risco foi empregado da seguinte forma: observou-se empiricamente que até uma certa distância d de análise ($d \cong 20\text{km}$) a maior parte dos pares de pontos, com localizações u_i e u_j , pertencem à zona de baixo risco. Assim, para distâncias de análise menor ou igual a d , empregou-se no cálculo de $\hat{\gamma}_{(u_i, u_j)}^{*R}$ a média e a variância decorrentes da zona de baixo risco. Por outro lado, para distâncias de análise maior que d , observou-se grande influência de pontos que pertencem à zona de alto risco sobre áreas de baixo risco. Nesse caso, para se evitar a superestimação do semivariograma do risco a partir da distância d , empregou-se a média e a variância decorrentes da zona de alto risco.

As estruturas de correlação espacial do risco de homicídio estimadas na cidade de São Paulo, de 2002 a 2004, segundo o critério estabelecido acima, são apresentadas nas Figuras 5a, 5b e 5c. Nessas figuras, a linha sólida de cor preta repre-

Figura 3

Instabilidade das taxas de homicídio.



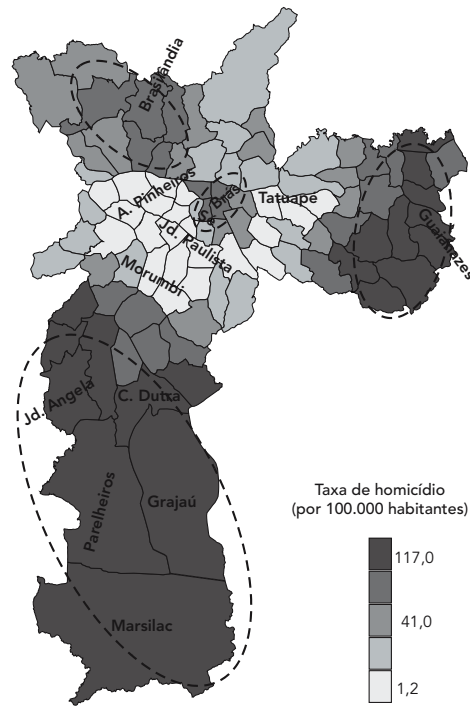
senta o modelo teórico de semivariograma, com seus respectivos parâmetros: efeito pepita (C_0), patamar (C), contribuição (C_1) e alcance (a).

Para cada período investigado a variabilidade espacial do risco de homicídio é descrita por um modelo esférico (o que apresentou melhor

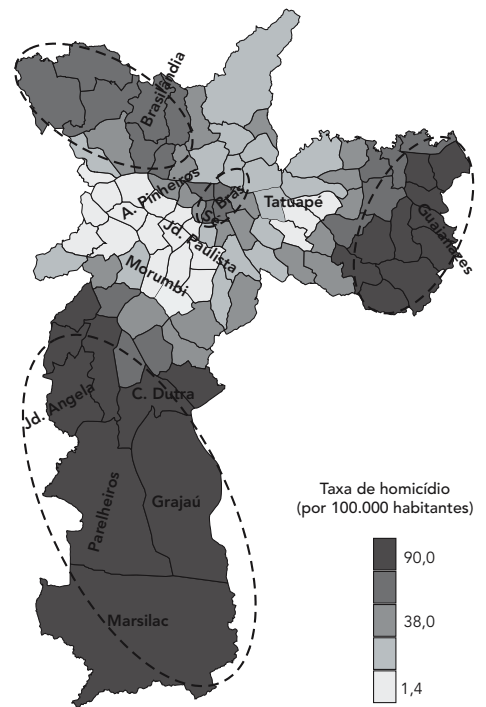
Figura 4

Tendência do risco de homicídio com realce das zonas de alto risco.

4a) 2002



4b) 2003



4c) 2004

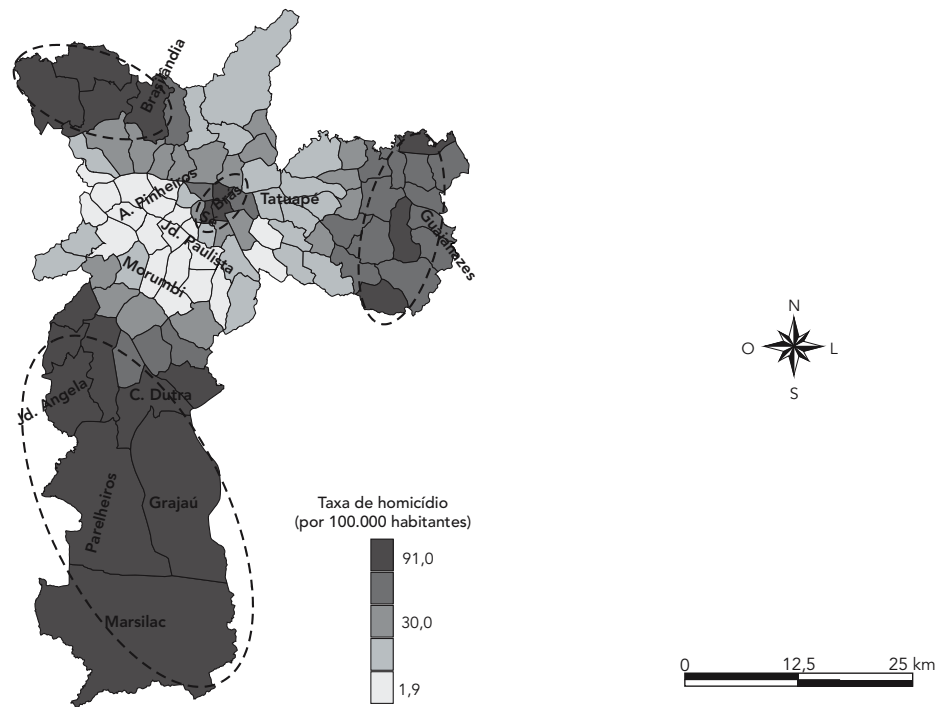


Tabela 1

Síntese da classificação das zonas de risco (critério: baixo risco \geq percentil 75%) dos 96 distritos da cidade de São Paulo, Brasil, de 2002 a 2004.

Taxas/100mil	Baixo risco (72 distritos: em torno do centro)			Alto risco (24 distritos: centro e periferia)		
	2002	2003	2004	2002	2003	2004
Mínima	1,2	1,4	1,9	53,3	54,0	40,2
Média	29,8	28,8	22,4	74,8	65,3	54,0
Máxima	52,2	52,5	40,0	117,1	90,2	91,3
Desvio-padrão	12,5	11,2	9,6	18,6	14,6	11,5

ajuste ao semivariograma do risco estimado). Essas estruturas revelam que o risco subjacente de homicídio pode estar autocorrelacionado numa distância [alcance (a)] de até aproximadamente 25km. Observa-se que para pequenas distâncias de análise o risco de homicídio apresenta pequena dissimilaridade; isto é, semivariogramas com valores baixos. A medida em que se aumenta a distância de análise essa dissimilaridade tende a crescer, até se estabilizar por volta de 25km. A partir dessa distância o fenômeno investigado não apresenta mais correlação espacial. Quando observamos as dimensões da área de estudo (largura \cong 52km e altura \cong 78km – vide Figura 1), a literatura geoestatística sugere como distância máxima de análise a metade da menor dimensão (largura ou altura) que compõe a área de estudo, neste caso em torno de 26km. Portanto, as estruturas obtidas são condizentes com as dimensões da área de estudo, com a geometria de amostragem e também coerentes sob a ótica da geoestatística convencional.

Um outro fato que deve ser observado é que o estimador proposto produz uma redução substancial do efeito pepita (C_0), quase zero. Esse efeito representa a componente caótica da estrutura de correlação espacial. Quanto menor for a proporção do efeito pepita em relação ao patamar (C) do semivariograma do risco mais forte é a dependência espacial do fenômeno investigado ³³, e portanto maior a sua continuidade espacial.

Para concluir esta análise, os resultados de 100 simulações mostraram que as estruturas de correlação espacial do risco de homicídio estimadas por $\hat{\gamma}^R(u_i, u_j)$ representam de forma razoável a média da distribuição simulada do semivariograma do risco, conforme apresentadas nas Figuras 5a, 5b e 5c. Os resultados dessas simulações reforçam a importância da estimação dos parâmetros para o semivariograma do risco, e neste caso, apontam que o estimador proposto pode ser adequado à análise do fenômeno investigado.

Estimação da superfície do risco de homicídio

Após a análise e seleção do modelo de estrutura de correlação espacial, estimativas do risco de homicídio foram estabelecidas por co-krigeagem binomial. Optou-se por calculá-las em várias localizações dispostas numa grade regular (ou malha espacial), com resolução de 150m na direção norte-sul e 98m na direção leste-oeste. A escolha dessa resolução foi apenas para manter a mesma relação de aspecto da área de estudo (norte-sul \cong 78km/leste-oeste \cong 52km) e proporcionar grades densas de valores (520 linhas por 530 colunas).

As superfícies geradas representam a evolução média da distribuição do risco de homicídio na cidade de São Paulo, de 2002 a 2004. Trata-se de um resultado suavizado que nos fornece uma primeira visão das principais tendências do fenômeno investigado. Os resultados são apresentados nas Figuras 6a, 6b e 6c.

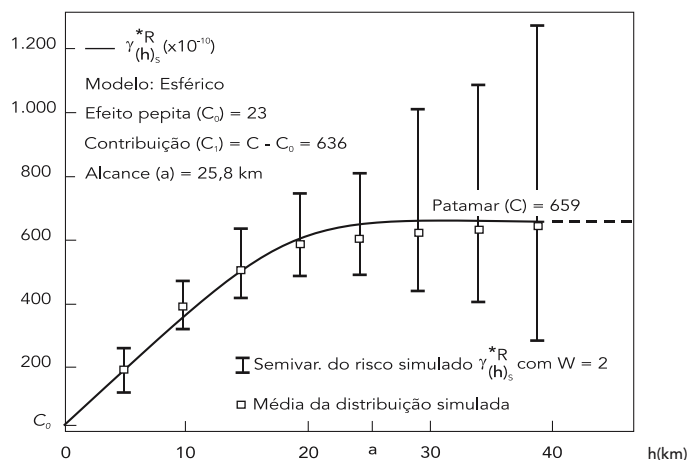
Quando analisamos a evolução do padrão espacial do risco de homicídio na cidade de São Paulo, de 2002 a 2004, verifica-se facilmente que algumas áreas concentram grande parte dos homicídios, enquanto em outras as mortes ocorrem em menores proporções (vide Figura 1 para identificação dos distritos).

Partindo da região central da cidade, observa-se a ocorrência de um foco de homicídios, também denominado “de área quente” (em inglês *hot zones* ou *hot spot*) ³⁴, que abarca principalmente os distritos da Sé, Brás e imediações. Conforme estudo recente ³⁵, boa parte dessa região é caracterizada por prédios comerciais, hotéis, restaurantes, cinemas e hospitais, são lugares que concentram muitas pessoas durante o dia. Essas áreas possuem ou estão localizadas próximas a terminais de transportes (metrô e ônibus) e têm uma atividade noturna intensa em bares, casas de bingos, casas de prostituição (saunas); além disto, está rodeada de cortiços e concentram um grande mercado de drogas. Um fato que chama a

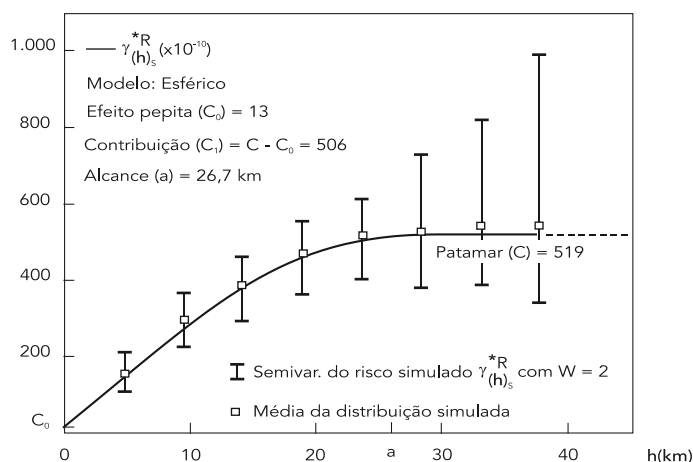
Figura 5

Estruturas de correlações espaciais do risco de homicídio estimadas versus simuladas.

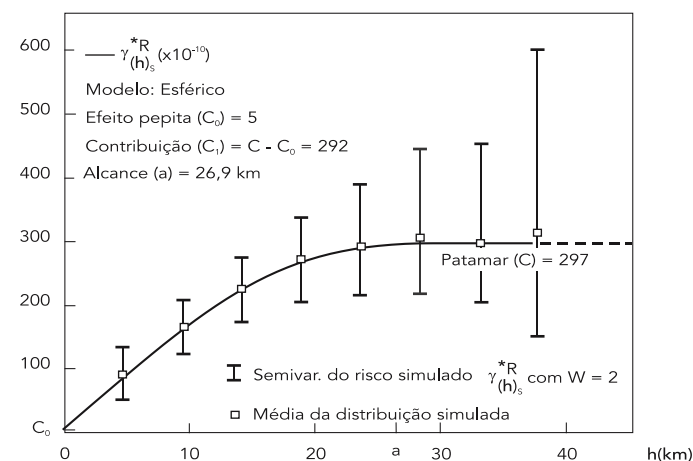
5a) 2002



5b) 2003



5c) 2004



atenção nos mapas é que essa área praticamente se mantém estável de 2002 a 2004. Para se ter uma idéia aproximada do risco nessa região, calculou-se, para cada ano, a média do risco de homicídio sobre os distritos envolvidos. Em 2002, essa área apontava para um risco médio da ordem de 128 mortes por 100 mil habitantes, passando para cerca de 116 mortes por 100 mil habitantes em 2003, e depois para quase 124 por 100 mil habitantes em 2004, portanto uma redução de apenas 3,1% em relação a 2002.

Deslocando-se do centro em direção às periferias, observa-se que o *hot spot* central vai se dissolvendo, até que surge uma área significativa que engloba vários distritos que são menos vulneráveis à criminalidade. Boa parte dessa área é dotada de melhor infra-estrutura, nível econômico mais elevado e qualidade de vida melhor se comparada com o resto da cidade. Como exemplo, os distritos de Jardim Paulista, Alto de Pinheiros, Morumbi, Tatuapé e Perdizes. Somente essas áreas tiveram seus índices reduzidos de 2002 para 2004 da ordem 32%, cerca de dez vezes mais que no centro da cidade. Um fato que chamou atenção, vide Figura 6, refere-se ao distrito do Morumbi. Em 2002 apresentava um índice médio da ordem de 29,29, em 2003 ao invés de diminuir, aumenta para 37,76 e, em 2004 cai para 22,85. Uma possível explicação para essa flutuação é porque ali convivem extremos de riqueza e pobreza (favela de Heliópolis, a maior da capital com aproximadamente 100 mil habitantes).

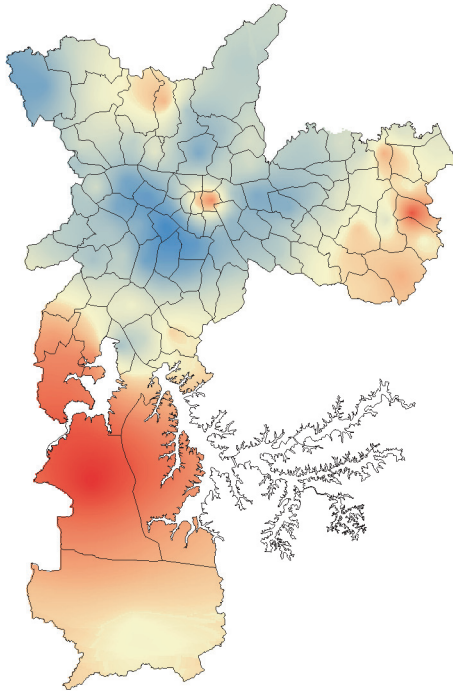
Na medida em que se avança das áreas de baixo risco em direção às periferias, estas vão se dissolvendo até que surgem três áreas de alto risco, uma localizada na periferia sul da cidade e as outras duas nas periferias leste e norte, respectivamente. Essa concentração espacial da violência nas regiões periféricas da cidade certamente não se deve a "maus fluidos" provenientes do subsolo.

A periferia sul teoricamente é considerada área de proteção ambiental por englobar as represas de Guarapiranga e Billings. Distritos como Capão Redondo, Jardim Ângela, Jardim São Luís, Cidade Dutra, Parelheiros, Grajaú e Marsilac são considerados lugares de altíssimo risco do município; boa parte da população que não tem para onde ir se desloca para esta região. São distritos que possuem grandes densidades populacionais com ocupação totalmente desordenada e irregular nas bordas das represas, falta infra-estrutura básica, alta proporção de pessoas jovens (de 10 a 25 anos) e uma grande parcela de adultos com baixo grau de escolaridade. Somente em Parelheiros e Grajaú a população cresceu quase 40 mil pessoas, de 2002 para 2004, acompanhada de um índice que oscilou entre 90 a 60 homicí-

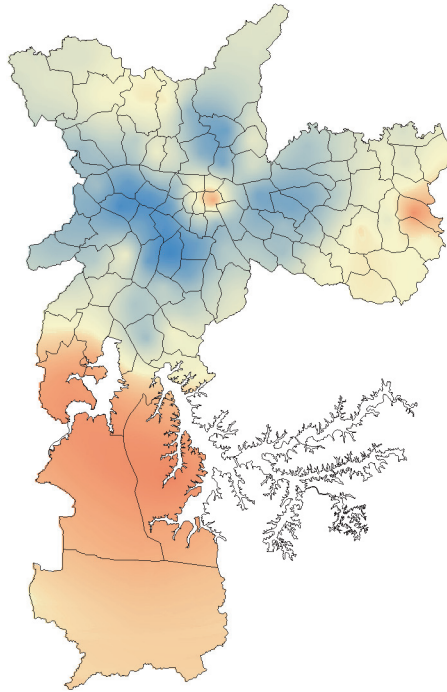
Figura 6

Evolução da média da distribuição do risco de homicídio na cidade de São Paulo, Brasil.

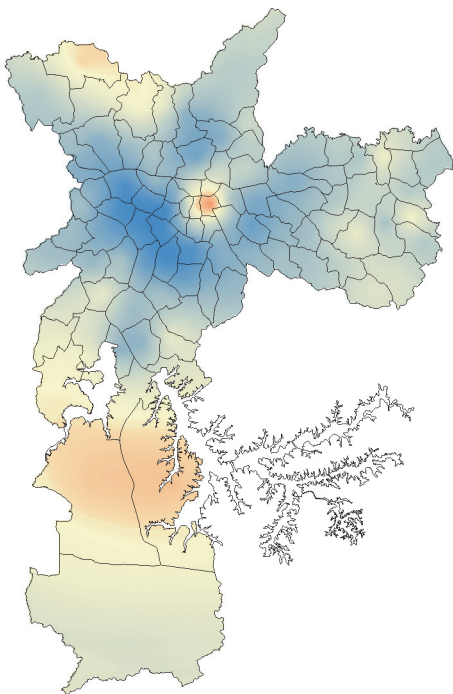
6a) 2002



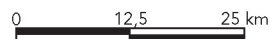
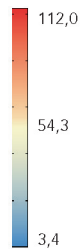
6b) 2003



6c) 2004



Risco por 100.000 habitantes



dios por 100 mil habitantes. Apesar desse quadro trágico, houve uma melhora de 2002 para 2004. Observe nos mapas, Figuras 6a, 6b e 6c, que o *hot spot* sofre uma atenuação gradual. No entanto, essas áreas ainda são de alto risco quando comparadas a outros locais da cidade. Por exemplo, em 2004 Parelheiros apresentava um risco médio da ordem de 61 mortes por 100 mil habitantes, cerca de 13 vezes maior que no Jardim Paulista (4,7), em Grajaú o risco médio era em torno de 62 mortes por 100 mil habitantes, oito vezes maior em relação a Alto de Pinheiros (7,7), ou então, no Jardim Ângela, 56 por 100 mil habitantes, cerca de nove vezes maior que em Perdizes (6,2).

A periferia leste da cidade concentra aproximadamente 20% da população do município (são quase 2 milhões de habitantes). É uma região marcada pela presença de loteamentos precários e favelas, baixa renda familiar, falta de infra-estrutura urbana etc. É uma área também estigmatizada pela violência, que se espalha por cerca de dez distritos. Conforme ilustra a Figura 6a, em 2002 observou-se a ocorrência de três *hot spots*: o primeiro que englobava parte dos distritos de São Miguel e Vila Curuçá, cujo risco médio de homicídios foi da ordem de 60 mortes por 100 mil habitantes; o segundo, um pouco maior, que englobava boa parte dos distritos de Lajeado, Guaianazes, Cidade Tiradentes e José Bonifácio, com risco médio de homicídios da ordem de 66 mortes por 100 mil habitantes; e o terceiro, que abarcava parte dos distritos de Parque do Carmo, Iguatemi, São Rafael e São Mateus, apresentava risco médio de homicídios próximo de 63 mortes por 100 mil habitantes. Somente em 2002, esses distritos atingiram um total de aproximadamente 640 mortes por 100 mil habitantes. No entanto, de 2002 para 2004, conforme ilustram os mapas das Figuras 6a, 6b e 6c, houve uma queda acentuada da violência nessa região. A queda da média do risco de homicídios de 2002 para 2003 foi da ordem 10,77%, e de 2002 para 2004 de aproximadamente 40,4%.

Na periferia norte da cidade constatou-se, em 2002, a presença de um *hot spot* que englobava parte dos distritos de Cachoeirinha e Brasilândia. Esses dois distritos juntos possuíam uma aglomeração populacional da ordem de 400 mil habitantes. Parte dessa região é considerada área de conservação e recuperação ambiental e tem características similares às outras regiões periféricas da cidade. Um fato que chamou atenção nos mapas, Figuras 6a, 6b e 6c foi o deslocamento do foco de homicídios nessa região. Inicialmente, em 2002, estava concentrado entre os distritos de Cachoeirinha e Brasilândia, atingindo em 2004 o distrito de Perus. Em 2002 esse distrito apresen-

tava um risco da ordem 40 mortes por 100 mil habitantes, passando para cerca de 56 por 100 mil habitantes em 2004, um aumento razoável, da ordem de 40%.

Cenários do risco de homicídio

A construção de cenários para o risco de homicídio foi estabelecida aplicando-se o procedimento SSCI, para dois níveis de probabilidades: a 20% e a 80%.

As Figuras 7a, 7b e 7c apresentam a evolução de cenários do risco de homicídio de 2002 a 2004, respectivamente, para valores de cortes obtidos do segundo decil da função de distribuição acumulada do risco $R(\mathbf{u})$. Referem-se a representações de cenários otimistas, em que cada localização \mathbf{u} da malha do mapa espacial é excedida por 80% dos valores simulados. Nesse caso, as áreas do mapa que apresentam valores elevados indicam que o risco de homicídio nessas localizações é muito alto, mesmo em cenários otimistas. Dessa forma, a periferia e o centro da cidade ainda permanecem como áreas de alto risco, quando comparados às outras localidades da cidade.

A situação adversa dos cenários otimistas são os cenários pessimistas. As Figuras 7d, 7e e 7f apresentam a evolução de cenários do risco de homicídio de 2002 a 2004, respectivamente, para valores de cortes obtidos do oitavo decil da função de distribuição acumulada de $R(\mathbf{u})$. Referem-se a representações de cenários pessimistas, em que cada localização \mathbf{u} da malha do mapa espacial é excedida apenas por 20% dos valores simulados. Nesse caso, as áreas do mapa que apresentam valores baixos indicam que o risco de homicídio nessas localizações é baixo, mesmo em cenários pessimistas. Uma forma alternativa de visualizar os resultados apresentados na Figura 7 é apresentar nos mapas somente as áreas consideradas de alto risco.

Esta análise teve por objetivo apresentar o uso da simulação seqüencial condicionada não-paramétrica para a construção de cenários do risco de homicídio na cidade de São Paulo, no triênio 2002-2004. Os resultados apresentados elucidam diferentes características do campo aleatório investigado e podem auxiliar o planejador que, orientado pelos seus objetivos, tem a possibilidade de escolher cenários mais adequados ao trabalho a ser executado.

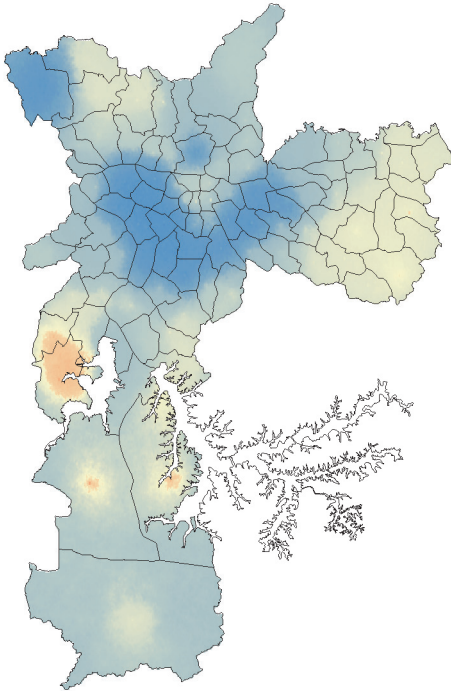
Comentários finais

Este artigo apresentou uma metodologia geostatística para estimação e mapeamento da dis-

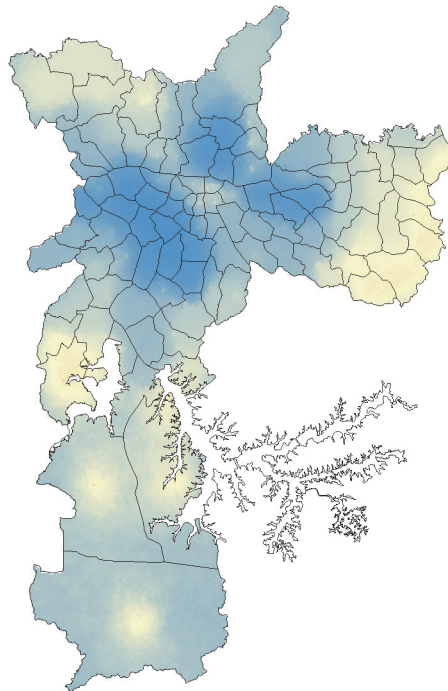
Figura 7

Evolução de cenários do risco de homicídio.

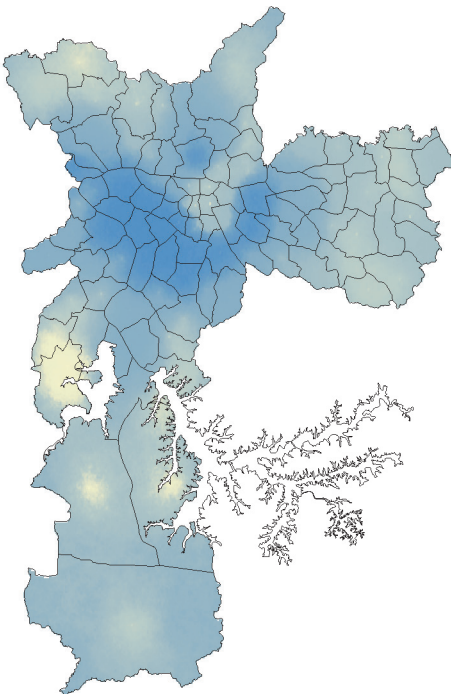
7a) Cenário otimista, 2002



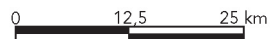
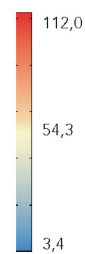
7b) Cenário otimista, 2003



7c) Cenário otimista, 2004



Risco por 100.000 habitantes

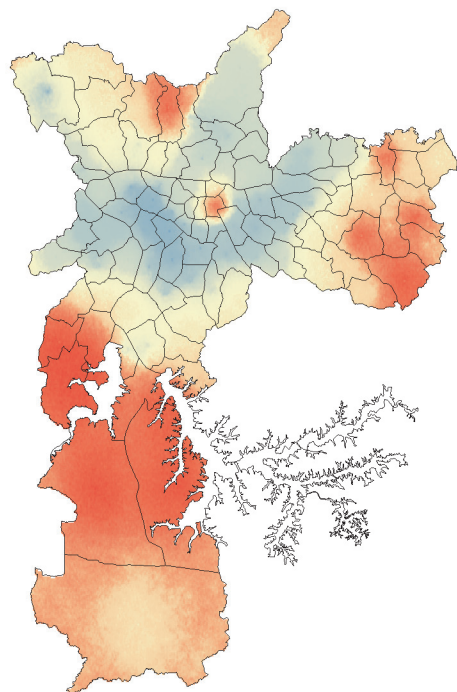


(continua)

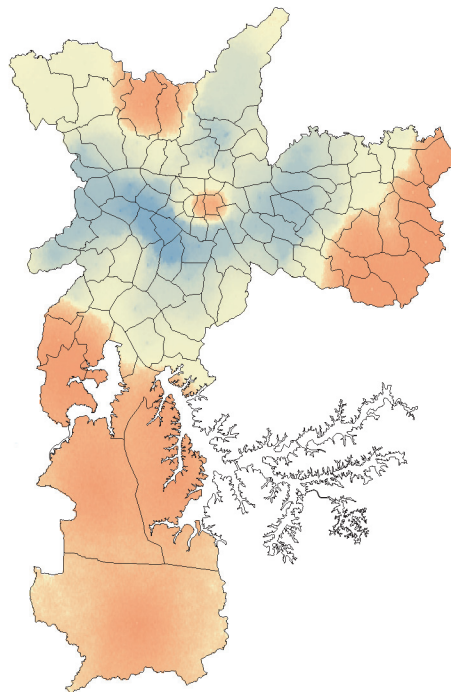
Figura 7 (continuação)

Evolução de cenários do risco de homicídio.

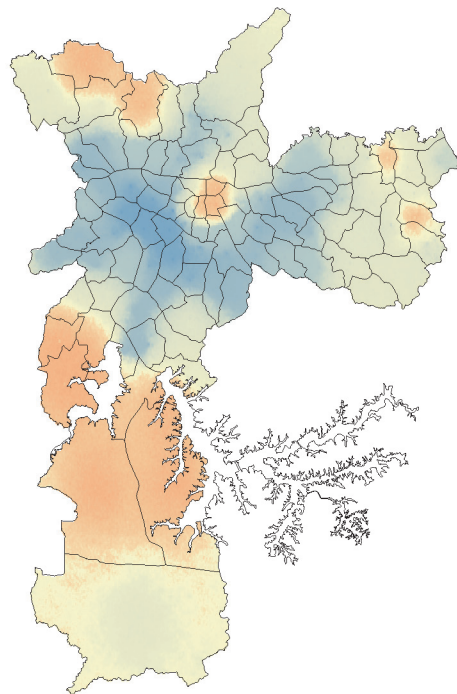
7d) Cenário pessimista, 2002



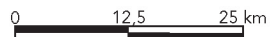
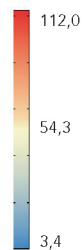
7e) Cenário passimista, 2003



7f) Cenário pessimista, 2004



Risco por 100.000 habitantes



tribuição espacial do risco de homicídio na cidade de São Paulo durante os anos de 2002 a 2004, sob o enfoque de um novo estimador para o semivariograma do risco. Os resultados apresentados são condizentes com estudos recentes^{7,35} e a estrutura urbana da cidade, não à guisa de conclusão, mas como uma forma alternativa de auxiliar na compreensão da distribuição espacial do risco de homicídio e poder contribuir para os meios de vigilância e prevenção. Embora a explicação da distribuição do fenômeno

investigado seja complexa, existem diversos fatores que podem contribuir para explicar esta distribuição, tais como: índice de concentração de renda; taxa de evasão escolar; diferenças nas taxas de desemprego entre os jovens; distribuição de equipamentos públicos como escolas, parques, hospitais e outros. A incorporação de alguns desses fatores no procedimento de modelagem apresentado pode tornar as estimativas do risco mais precisas, o que nos motiva à investigação deste tema em trabalhos futuros.

Resumo

O impacto da violência sobre o perfil de saúde da população nos grandes centros urbanos brasileiros é uma realidade e tem sido incorporado na agenda das ações de saúde. Embora o crescimento dos crimes contra a vida seja sentido por toda a população, a sua distribuição pela cidade ocorre de modo desigual. Nesse contexto, ferramentas de análise que permitam produzir uma avaliação do risco de homicídio e de sua distribuição espacial potencializam os meios de vigilância e prevenção. Este artigo oferece uma contribuição nessa direção. Uma metodologia geoestatística é empregada para a estimação e mapeamento do risco de homicídio, sob o enfoque de um novo estimador de semivariograma, e tem sua aplicação avaliada na cidade de São Paulo, Brasil, para o período de 2002 a 2004. Cenários para o risco de homicídio são gerados com base na co-krigeagem binomial e de procedimentos de simulação estocástica, possibilitando a delimitação de áreas com maior ou menor risco, que podem ser usados para estudar condicionantes espaciais da violência e auxiliar para um melhor entendimento dos problemas por meio da análise de suas diferenças ao longo dos territórios das cidades.

Homicídio; Análise Espacial; Medição de Risco

Colaboradores

E. C. G. Camargo foi o principal responsável pela realização da pesquisa e redação do trabalho. S. Druck, A. M. V. Monteiro, C. C. Freitas e G. Câmara foram responsáveis pela orientação e revisão do trabalho.

Agradecimentos

À Divisão de Processamento de Imagens (DPI) do Instituto Nacional de Pesquisas Espaciais (INPE) pela infraestrutura cedida para o desenvolvimento do trabalho, pelo apoio e incentivo. Ao amigo Dr. Paulo Justiniano Ribeiro Junior, do Departamento de Estatística da Universidade Federal do Paraná, pelo apoio e ensinamentos. Ao Programa de Aprimoramento das Informações de Mortalidade (PRO-AIM) no Município de São Paulo pela disponibilidade dos dados.

Referências

1. Departamento de Informática do SUS. Sistema de informações sobre mortalidade. <http://www.datasus.gov.br/catalogo/sim.htm> (acessado em 07/Mar/2006).
2. Minayo MCS. Violência social sob a perspectiva da saúde pública. *Cad Saúde Pública* 1994; 10 Suppl 1:S7-18.
3. Cruz OG. Homicídios no Estado do Rio de Janeiro: análise da distribuição espacial e sua evolução [Dissertação de Mestrado]. São Paulo: Faculdade de Saúde Pública, Universidade de São Paulo; 1996.
4. Carvalho MS. Aplicação de métodos de análise espacial na caracterização de áreas de risco à saúde [Tese de Doutorado]. Rio de Janeiro: Instituto Alberto Luiz Coimbra de Pós-graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro; 1997.
5. Santos SM, Barcellos C, Carvalho MS, Flores R. Detecção de aglomerados espaciais de óbitos por causas violentas em Porto Alegre, Rio Grande do Sul, Brasil, 1996. *Cad Saúde Pública* 2001; 17:1141-51.
6. Beato CCF. Determinantes da criminalidade em Minas Gerais. *Rev Bras Ciênc Soc* 1998; 13:74-89.
7. Nery MB, Monteiro AMV. Análise intra-urbana dos homicídios dolosos no Município de São Paulo. In: XV Encontro Nacional de Estudos Populacionais. http://www.abep.org.br/usuario/GerenciaNavega.caophp?caderno_id=504&nivel=2&texto_id=2988 (acessado em 25/Set/2006).
8. Lima MLC, Ximenes RAA, Souza ER, Luna CF, Albuquerque MFPM. Spatial analysis of socioeconomic determinants of homicide in Brazil. *Rev Saúde Pública* 2005; 39:176-82.
9. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. *Appl Stat* 1991; 40:283-94.
10. Martuzzi M, Elliott P. Empirical Bayes estimation of small area prevalence of non-rare conditions. *Stat Med* 1996; 15:1867-73.
11. Clayton DG, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; 43:671-81.
12. Anselin L. Local indicators of spatial association – LISA. *Geogr Anal* 1995; 27:91-115.
13. Câmara G, Druck S, Carvalho MS, Monteiro AMV, Camargo ECG, Felgueira CA, et al. Análise espacial de dados geográficos. Planaltina: Embrapa Cerrados; 2004.
14. Pickle LW. Spatial analysis of disease. In: Beam C, editor. *Biostatistical applications in cancer research*. Boston: Kluwer Academic Publishers; 2002. p. 113-50.
15. Cressie N. *Statistics for spatial data*. New York: Wiley; 1993.
16. Kelsall J, Wakefield J. Modelling spatial variation in disease risk: a geostatistical approach. *J Am Stat Assoc* 2002; 97:692-701.
17. Berke O. Exploratory disease mapping: kriging the spatial risk function from regional count data. *Int J Health Geogr* 2004; 3:18.
18. Goovaerts P, Jacquez GM, Greiling D. Exploring scale-dependent correlations between cancer mortality rates using factorial kriging and population-weighted semivariograms. *Geogr Anal* 2005; 37:152-82.
19. Lajaunie C. Local risk estimation for a rare non-contagious disease based on observed frequencies. Fontainebleau: Center de Géostatistique, École des Mines de Paris; 1991.
20. Oliver MA, Webster R, Lajaunie C, Muir KR, Parkes SE, Cameron AH, et al. Binomial cokriging for estimating and mapping the risk of childhood cancer. *IMA J Math Appl Med Biol* 1998; 15:279-97.
21. Deutsch CV, Journel AG. *GSLIB: Geostatistical Software Library and user's guide*. New York: Oxford University Press; 1998.
22. Câmara G, Souza RCM, Freitas UM, Garrido J. SPRING: integrating remote sensing and GIS by object-oriented data modeling. *Comput Graph* 1996; 20:395-403.
23. Grauman DJ, Tarone RE, Devesa SS, Fraumeni Jr. JF. Alternate ranging methods for cancer mortality maps. *J Natl Cancer Inst* 2000; 92:534-43.
24. Brewer CA, Pickle L. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Ann Assoc Am Geogr* 2002; 92:662-81.
25. Instituto Brasileiro de Geografia e Estatística. Censo demográfico de 2000. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2000.
26. Programa de Aprimoramento das Informações de Mortalidade no Município de São Paulo. Dados de homicídios. <http://www2.prefeitura.sp.gov.br/cgi/deftohtm.exe?secretarias/saude/TABNET/SIM/obito.def> (acessado em 07/Mar/2006).
27. Waller LA, Gotway CA. *Spatial clustering of health events: regional count data. Applied spatial statistics for public health data*. New Jersey: John Wiley and Sons; 2004. p. 200-71.
28. Assunção RM, Barreto SM, Guerra HL, Sakurai E. Mapas de taxas epidemiológicas: uma abordagem Bayesiana. *Cad Saúde Pública* 1998; 14:713-23.
29. Beato CC, Assunção RM, Santos MC, Espírito-Santo LE, Saporì LF, Batitucci E, et al. Criminalidade violenta em Minas Gerais – 1986 a 1997. <http://www.crisp.ufmg.br/cvmg.htm> (acessado em 10/Abr/2006).
30. Webster R, Oliver MA, Muir KR, Mann JR. Kriging the local risk of a rare disease from a register of diagnoses. *Geogr Anal* 1994; 26:168-85.
31. Goovaerts P. *Geostatistics for natural resources evaluation*. New York: Oxford University Press; 1997.
32. Bailey TC, Gatrell AC. *Interactive spatial data analysis*. New York: Wiley; 1995.
33. Chilès JP, Delfiner P. *Geostatistics: modeling spatial uncertainty*. New York: Wiley; 1999.
34. Harries K. *Mapping crime: principles and practice*. Washington DC: National Institute of Justice, U.S. Department of Justice; 1999.
35. Ceccato V, Haining R, Kahn T. The geography of homicide in São Paulo, Brazil. *Environ Plan A* 2007; 39:1632-53.

Recebido em 17/Ago/2006

Versão final reapresentada em 09/Mar/2007

Aprovado em 10/Dez/2007