

# FUNÇÕES L HOMOGÊNEA E NÃO HOMOGÊNEA EM TESTES DE HIPÓTESE DE COMPLETA ALEATORIEDADE ESPACIAL

Sidleide Barbosa de Sousa<sup>1</sup>, João Domingos Scalon<sup>2</sup>

**Resumo:** *O objetivo principal deste trabalho é investigar o poder de testes contra a hipótese de completa aleatoriedade espacial em configurações pontuais simuladas que apresentam não homogeneidade e dependência. Os testes utilizam as funções L homogênea e não homogênea na estatística de Kolmogorov-Smirnov. Os resultados mostram que a função L homogênea não deveria ser usada em configurações suspeitas de serem não homogêneas e que a função L não homogênea não deveria ser usada em configurações homogêneas. Os métodos apresentados neste trabalho são aplicados em uma configuração pontual não homogênea real e os resultados obtidos corroboram com os resultados dos dados simulados.*

**Palavras-chave:** *Estatística espacial, Processo de Poisson, Função K de Ripley, Intensidade kernel, Monte Carlo, Teste de hipótese.*

**Abstract:** *The main aim of this work is to investigate the power of tests against the null hypothesis of complete spatial randomness in simulated point patterns that present both inhomogeneity and dependence. The tests use both the homogenous and inhomogeneous L functions on a Kolmogorov-Smirnov statistic. The results show that the homogeneous L function should not be used in patterns that is know or suspect to be spatially inhomogeneous and that the inhomogeneous L function should not be used in homogeneous patterns. The methods presented in this paper are applied in an actual spatially inhomogeneous pattern and the results corroborate the results from simulated data sets.*

**Keywords:** *Spatial statistics, Poisson process, Ripley's K function, Intensity kernel, Monte Carlo, Hypothesis test.*

## 1 Introdução

Muitos fenômenos podem ser expressos através de ocorrências identificadas como pontos (coordenadas) localizados no espaço (plano) e são denominados de processos pontuais. Exemplos desse tipo de processos são: localização de indivíduos com uma determinada doença, posição de partículas em um material composto e localização de plantas de uma espécie vegetal. Observa-se que nesses exemplos o interesse é a própria localização espacial das ocorrências em estudo. O objetivo básico da análise desse tipo de dados é estudar a distribuição espacial desses pontos testando a hipótese nula de que a configuração pontual observada apresenta completa aleatoriedade espacial na região em estudo. Caso a hipótese nula seja rejeitada pode-se continuar a

---

<sup>1</sup>DEX - UFLA. e-mail: sbarbosas1987@yahoo.com.br

<sup>2</sup>DEX - UFLA. e-mail: scalon@dex.ufla.br

análise no sentido de verificar se a configuração apresenta pontos que exibam alguma forma de tendência ou interação (DIGGLE, 2003; CRESSIE, 1991).

Existem diversas funções (ex.  $F, G, J, K$  e  $L$ ) propostas para analisar se a configuração espacial observada apresenta interação entre os pontos, sendo que a função  $L$  é uma das mais conhecidas (DIGGLE, 2003). A aplicação da função  $L$  assume que a configuração pontual é uma realização de um processo pontual homogêneo (sem tendência), isto é, a intensidade dos pontos é a mesma em todos os lugares (RIPLEY, 1977, 1981). Entretanto, essa suposição nem sempre pode ser atendida na prática. Caso o processo pontual seja não homogêneo é necessário utilizar métodos que consideram a não homogeneidade do processo para análise da interação entre os pontos. Dentre esses métodos, a função  $L$  não homogênea, proposta por Baddeley et al., (2000), é uma das mais utilizadas.

A função  $L$  não homogênea têm sido utilizada com frequência na análise de configurações pontuais em que exista suspeita de não ser observada a homogeneidade espacial (BADDELEY e TURNER, 2005). Entretanto, não existem estudos que mostram o poder dessa função em testes contra a hipótese de completa aleatoriedade espacial em configurações que exibam diferentes formas de tendência e interação espacial.

O principal objetivo deste trabalho é avaliar o poder das funções  $L$  homogênea e não homogênea em um teste, baseado na distância de Kolmogorov-Smirnov, contra a hipótese de completa aleatoriedade espacial em configurações espaciais de pontos que apresentam diferentes níveis de não homogeneidade e diferentes tipos de interação espacial. Outro objetivo é aplicar os métodos descritos neste trabalho em uma configuração pontual não homogênea. Para tal, utilizou-se as coordenadas dos centros de partículas de bronze localizadas em uma seção plana longitudinal de um filtro fabricado pelo método de sinterização.

## 2 Simulação das configurações

Na análise do poder são utilizadas realizações de configurações que seguem os seguintes padrões de configurações espaciais: agrupamentos, regularidade e tendência. As configurações com agrupamentos e regularidade são simuladas considerando a existência de homogeneidade espacial. As configurações com tendência são simuladas para não exibir interação entre eventos, apenas não homogeneidade em uma única direção.

As realizações com agrupamentos são geradas para apresentar 100 pontos distribuídos dentro da área unitária. Para simular uma configuração com agrupamento, é utilizado o algoritmo de *Matern* que é um caso especial do processo de *Neyman-Scott*. Assim, é gerado uma configuração pontual utilizando um processo Poisson homogêneo com uma intensidade conhecida em que o número de pontos corresponde ao número de agrupamentos. São simuladas configurações que exibem 2, 5, 10 e 20 agrupamentos. A seguir, são gerados novos pontos dentro de discos de raios conhecidos com centros formados pelos pontos gerados anteriormente. O raio do disco é conhecido como variabilidade radial. São simuladas configurações com variabilidades radiais iguais a 0,05, 0,10, 0,15, 0,20, 0,25, 0,30, 0,35, 0,40, 0,45 e 0,50. Deve-se observar que quanto menor a variabilidade radial maior a força da interação (agrupamento). Deve-se observar também que o posicionamento desses pontos dentro do disco é uniforme, ou seja, a posição independente da distância do ponto ao centro. O conjunto de todos os 100 pontos gerados nos dois passos da simulação dentro da área é uma realização de um processo estacionário com agrupamentos de *Matern* (MATERN, 1986; WAAGEPETERSEN, 2007).

Com o procedimento descrito acima são geradas um total de 40 configurações, representando todas as combinações do número de agrupamentos e variabilidades radiais. Para gerar essas realizações é utilizada a função *rMatClust* da biblioteca *spatstat* (BADDELEY e TURNER, 2005) disponível no *software R* (R CORE TEAM, 2016).

As realizações das configurações exibindo regularidade são geradas para apresentarem 100 pontos distribuídos dentro da área unitária. Para simular uma configuração com regularidade é utilizado o algoritmo *Simple Sequential Inhibition (SSI)* sugerido por Diggle (2003). Esse

algoritmo começa com uma região vazia e, então, pontos são adicionados um por um, onde cada ponto é gerado por um processo de Poisson homogêneo. Se o próximo ponto gerado estiver dentro de uma distância mínima estabelecida de qualquer outro ponto já existente na região, o ponto é rejeitado e outro ponto é gerado aleatoriamente. Assim que o número de pontos que se deseja é alcançado o algoritmo é encerrado. Deve-se observar que quanto maior a distância mínima, maior é a regularidade da configuração pontual.

São simuladas configurações que exibem regularidade com as distâncias mínimas entre os pontos de 0,010, 0,012, 0,014, 0,016, 0,018, 0,020, 0,030, 0,040, 0,050 e 0,060. Assim, são geradas um total de 10 configurações apresentando regularidade. Para gerar essas realizações é utilizada a função *rSSI* da biblioteca *spatstat* (BADDELEY e TURNER, 2005) disponível no *software R* (R CORE TEAM, 2016).

As realizações de configurações contendo 100 pontos em uma área unitária com tendência em uma direção são geradas utilizando um processo de Poisson não homogêneo com intensidade  $\lambda(x) = 300 * \exp(-\beta * latitude)$ . Deve-se observar que a intensidade aumenta na direção oeste-leste do eixo  $x$ . Assim, quanto maior o valor do parâmetro  $\beta$  maior a intensidade de pontos na direção leste.

São simuladas configurações com os seguintes parâmetros  $\beta$ : 0, 1, 2, 3, 4, e 5. Para obter essas realizações é utilizada a função *rpoispp* da biblioteca *spatstat* (BADDELEY e TURNER, 2005) disponível no *software R* (R CORE TEAM, 2016).

## 2.1 Testes de hipóteses

Tem-se o interesse em apresentar métodos para testar a hipótese nula de que a configuração pontual exibe completa aleatoriedade espacial, ou seja, se a configuração pontual é uma realização de um processo de Poisson homogêneo com intensidade constante e conhecida. Consideram que a hipótese nula  $H_0$  representa o processo de Poisson homogêneo, enquanto a hipótese alternativa  $H_1$  representa um processo estocástico desconhecido que pode apresentar tendência (não estacionariedade) e/ou algum tipo de interação (ex. aglomerados ou regularidade). Dessa forma  $H_1$  é complementar da hipótese nula  $H_0$  (DIGGLE, 2003).

Os testes contra a hipótese nula de completa aleatoriedade espacial são aplicados para cada uma das realizações simuladas utilizando as funções  $L$  homogênea e não homogênea que são apresentadas a seguir.

A função  $K$ , proposta por Ripley (1977), descreve a propriedade de segunda ordem do padrão espacial de um processo pontual em diferentes escalas. Para um processo pontual homogêneo, a função  $K$  é definida por:

$K(h) = \lambda^{-1} E(\text{número de eventos contidos dentro de uma distância } h \text{ a partir de um evento arbitrário}).$

Para um processo aleatório sob a suposição de CAE, tem-se que  $K(h) = \pi h^2$ . Para um processo pontual com aglomerados temos que  $K(h) > \pi h^2$ , ou seja, os pontos têm, em média, mais eventos do que o esperado para uma configuração CAE, para uma distância  $h$ , enquanto para um processo regular tem-se que  $K(h) < \pi h^2$ , ou seja, os pontos têm, em média, menos pontos do que esperado sob a suposição de CAE, para uma distância  $h$  (CRESSIE, 1993).

Um estimador da função  $K$  que corrige o efeito de bordas, proposto por Ripley (1977), é dado por:

$$\hat{K}(h) = \frac{|A|}{n^2} \sum_i^n \sum_{j, i \neq j}^n \frac{I_h(d_{ij})}{w_{ij}}, \quad (1)$$

em que  $|A|$  é a área da região,  $n$  é o número de eventos observados,  $d_{ij}$  é a distância euclidiana entre os eventos  $i$  e  $j$ ,  $i \neq j$ ,  $I_h(d_{ij})$  é uma função indicadora cuja valor é igual a 1 se  $(d_{ij}) \leq h$  e 0 caso contrário e  $w_{ij}$  é o fator de correção do efeito de borda que representa a proporção da circunferência do círculo centrado no evento  $i$  passando sobre o evento  $j$  pertencente a região de estudo.

Besag e Diggle (1977) fizeram uma proposta de um estimador modificado, denominado de função  $L$ , que é dado por:

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}}. \quad (2)$$

A função  $L$  é mais fácil de interpretar do que a função  $K$  pois a linha obtida sob a hipótese de nulidade de CAE é uma reta com 45 graus para todas as distância  $h$ . Para configurações com agrupamentos, a função  $L$  está acima da linha reta, enquanto as configurações com regularidades apresentam função  $L$  abaixo da linha reta. Outra vantagem do estimador da função  $L$  é que a variância é mais estável do que o estimador da função  $K$  (BESAG; DIGGLE, 1977).

A função  $K$  é baseada na suposição de que o processo pontual espacial é estacionário e isotrópico. Se o processo pontual espacial tiver indícios de que não é homogêneo, ou seja, tem intensidade não constante, a função  $K$  homogênea não deveria ser utilizada. Existem várias funções que podem ser utilizadas para caracterizar interação em processos pontuais não homogêneos. Dentre essas funções, a mais utilizada é a função  $K$  não homogênea.

A função  $K$  não homogênea é uma generalização da função  $K$  homogênea e seu estimador, com a correção do efeito de borda, foi proposto por Baddeley et al., (2000) como sendo:

$$\hat{K}_{inhom}(h) = \frac{1}{|A|} \sum_i^n \sum_{i \neq j}^n \frac{w_{ij}^{-1} I_h(d_{ij})}{\hat{\lambda}(x_i) \hat{\lambda}(x_j)}, h \geq 0, \quad (3)$$

em que  $|A|$  é a área da região,  $n$  é o número de eventos observados,  $d_{ij}$  é a distância euclidiana entre os eventos  $i$  e  $j$ ,  $i \neq j$ ,  $I_h(d_{ij})$  é uma função indicadora cuja valor é igual a 1 se  $(d_{ij}) \leq h$  e 0 caso contrário,  $w_{ij}$  é o fator de correção do efeito de borda que representa a proporção da circunferência do círculo centrado no evento  $i$  passando sobre o evento  $j$  pertencente a região de estudo e  $\hat{\lambda}(x_i)$ ,  $\hat{\lambda}(x_j)$ , são os valores estimados da função intensidade no ponto  $x_i$  e  $x_j$ , respectivamente.

Baddeley et al. (2000) mostram que para um processo pontual espacial não estacionário a função  $K$  não homogênea é exatamente igual a função  $K$  homogênea usual, ou seja,  $K_{inhom}(h) = \pi h^2$ .

O estimador da função  $L$  não homogênea é obtido da mesma forma que na função  $L$  homogênea, ou seja,

$$\hat{L}_{inhom}(h) = \sqrt{\frac{\hat{K}_{inhom}(h)}{\pi}}. \quad (4)$$

Baddeley et al. (2000) afirmam que esse estimador também pode ser utilizado para linearizar e estabilizar a variância da função  $K$  não homogênea.

A maneira de construir um teste de hipótese formal contra a hipótese nula de completa aleatoriedade espacial é definir uma medida de discrepância entre a estimativa da função  $L$  homogênea (ou  $L$  não homogênea) e a função  $L$  teórica para todas as distâncias. Diggle (2003) sugere utilizar duas medidas de discrepância: Kolmogorov-Smirnov e Cramer-von Mises. Neste trabalho utiliza-se a estatística de teste baseada na medida de discrepância de Kolmogorov-Smirnov que é dada por

$$L_K(i) = \max_{0 \leq h \leq h_0} |\hat{L}(h) - L(h)|. \quad (5)$$

Para obter a estatística de teste de Kolmogorov-Smirnov considera-se uma distância máxima  $h_0 = 0,25$ , seguindo o que recomenda Diggle (2003). Para obter-se o *valor - p* dos testes de hipótese utiliza-se 100 simulações de Monte Carlo seguindo o processo de Poisson homogêneo com intensidade  $\lambda = 100$  e o processo de Poisson não homogêneo com intensidade  $\lambda(x) = 300 * \exp(-\beta * latitude)$ , também seguindo o recomendado por Diggle (2003) e Baddeley et al. (2000).

## 2.2 Análise do poder dos testes

São geradas 1000 configurações para cada realização dos modelos descritos anteriormente, exibindo as diferentes formas de não homogeneidade, regularidade e agrupamentos. Para cada uma dessas configurações é aplicado os testes de Kolmogorov-Smirnov usando as funções  $L$  homogênea e não homogênea. Aplica-se um nível de significância de 0,05 para rejeitar a hipótese nula de completa aleatoriedade espacial. O número de vezes que o teste rejeita a hipótese nula corretamente é considerado como uma medida do poder do teste.

## 3 Aplicação

A base de dados utilizados foi obtida por Bernhardt, Meyer-Olbersleben e Kieback (1997) e está disponível na biblioteca *spatstat* (BADDELEY e TURNER, 2005) disponível no *software R* (R CORE TEAM, 2016). Os dados são formados por 678 coordenadas cartesianas de partículas de bronze localizadas em uma secção diagonal retangular plana de 18 x 7 mm, com área de 126  $mm^2$ , de um filtro fabricado pelo método de sinterização. O processo de fabricação do filtro introduz uma não homogeneidade na distribuição das partículas de bronze no filtro. A presença de não homogeneidade das partículas é importante para que o filtro cumpra com a sua finalidade. Assim, o interesse do fabricante é caracterizar a distribuição espacial das partículas de bronze no filtro.

Inicialmente é feita uma representação gráfica da posição das partículas no plano. A seguir, é explorado as propriedades de primeira ordem utilizando o estimador de intensidade da kernel. Segundo Bailey e Gatrell (1995), a estimação por kernel foi desenvolvida para obter uma estimativa *suave* de uma densidade de probabilidade univariada a partir das observações amostrais. Seja  $x$  uma localização em qualquer região  $A$ , cujo valor da intensidade se deseja estimar. O estimador de intensidade kernel para  $\lambda$  em  $x$  é dado por:

$$\hat{\lambda}_\tau(x) = \frac{1}{\delta_\tau(x)} \sum_{i=1}^n \frac{1}{\tau^2} S\left(\frac{d(x, x_i)}{\tau}\right), d(x, x_i) \leq \tau. \quad (6)$$

Dessa forma tem-se o estimador de kernel para um raio de influência ( $\tau \geq 0$ ), também conhecido como largura de banda, e que determina a quantidade de suavização. Deve-se observar que se o raio for pequeno, será gerada uma superfície com muitos picos e se esse raio for grande a superfície ficará mais suavizada. Vários métodos têm sido propostos para a escolha ideal da largura de banda. Porém ainda não foi definido qual o melhor. Por exemplo, Diggle (2003) sugere usar um critério baseado na minimização do erro quadrático médio do estimador de kernel.

A função de interpolação  $S(\cdot)$  é conhecida também como função kernel a qual deve ser simétrica em relação à origem, em que  $d(x, x_i)$  é a distância entre o ponto avaliado  $x$  e cada evento observado  $x_i$ . Diversas funções podem ser utilizadas, sendo a Gaussiana a mais comum. A distribuição normal pondera os pontos dentro do círculo de forma que os pontos mais próximos têm maior peso comparados com os mais afastados. Dessa forma, é representada por:

$$S(h) = \frac{1}{2\pi\tau} e^{-\frac{h^2}{2\tau^2}},$$

em que  $h$  representa a distância euclidiana entre o ponto e o evento e  $\tau$  é a largura da banda.

O fator  $\delta_\tau(x)$  é uma correção de borda, que representa a proporção do volume sob o kernel encontrado em  $x$  que está contido na região  $A$ . Dessa forma pode ser obtido por

$$\delta_\tau(x) = \int_A \frac{1}{\tau^2} S\left(\frac{x-u}{\tau}\right) dx. \quad (7)$$

Dessa forma, utilizou-se uma função kernel gaussiana com raio de influência  $\tau$  igual a 3, obtido por validação cruzada proposto por Diggle (2003).

Para testar a hipótese nula de que as partículas de bronze segue um processo de Poisson homogêneo dentro do filtro, utiliza a função  $L$  não homogênea tanto para construir os envelopes de simulação quanto para obter a estatística de teste de Kolmogorov-Smirnov. Nas duas situações utiliza-se uma distância máxima  $h_0$  igual a 4 mm e 100 simulação de Monte Carlo sob a hipótese de completa aleatoriedade espacial.

A hipótese de completa aleatoriedade espacial também pode ser testada através dos envelopes de simulação superiores e inferiores que são construídos a partir de  $s$  simulações independentes do processo de Poisson homogêneo com intensidade constante conhecida e  $n$  pontos dentro de uma região  $A$ . O máximo e o mínimo da função  $L$  homogênea (ou  $L$  não homogênea) das configurações simuladas são representados por:

$$\begin{aligned} LS(h) &= \max_{i=1,\dots,s} \{\hat{L}_i(h)\} \\ LI(h) &= \min_{i=1,\dots,s} \{\hat{L}_i(h)\}. \end{aligned}$$

A interpretação dos envelopes de simulação é bastante simples. Se a curva da estimativa da função  $L$  homogênea (ou  $L$  não homogênea) ficar dentro dos envelopes de simulação existe evidências estatísticas para aceitar a hipótese nula de CAE. Se a curva da estimativa da função  $L$  homogênea (ou  $L$  não homogênea) ficar fora dos envelopes de simulação existe evidências estatísticas de que o processo exibe interação (dependência) entre os eventos e, portanto, evidências para rejeitar a hipótese nula de CAE. Além disso, se a curva da função  $L$  homogênea (ou  $L$  não homogênea) estimada ficar acima do envelope superior existe indícios de que a configuração pontual apresenta agrupamentos. Caso a curva da função  $L$  homogênea (ou  $L$  não homogênea) estimada ficar abaixo do envelope inferior existe indícios de que a configuração pontual apresenta regularidade.

### 3.1 Software

Para conduzir todas as análises foram construídas funções utilizando a biblioteca *spatstat* (BADDELEY e TURNER, 2005) disponível no *software R* (R CORE TEAM, 2016).

## 4 Poder dos testes em configurações homogêneas com interação para agrupamentos

Na análise de processos pontuais espaciais tem-se interesse em verificar se a configuração espacial observada foi gerada por um processo estocástico espacial que exibe completa aleatoriedade espacial, ou seja, a configuração espacial é uma realização de um processo de Poisson homogêneo com intensidade conhecida e constante. Para testar a hipótese nula de completa aleatoriedade espacial pode-se utilizar diversos métodos que exploram as propriedades de primeira e segunda ordem. A propriedade de primeira ordem (relacionada com a tendência do processo) pode ser analisada através de estimadores locais de intensidade do tipo kernel, enquanto a propriedade de segunda ordem (relacionada com a interação entre os pontos do processo) pode ser analisada pelas funções  $G$ ,  $F$ ,  $J$ ,  $K$  e  $L$  (DIGGLE, 2003).

A função  $L$  não homogênea têm sido utilizada com frequência na análise de configurações pontuais em que exista suspeita de não ser observada a homogeneidade espacial. Entretanto, não encontra-se na literatura estudos que mostram o poder dessa função em testes contra a hipótese de completa aleatoriedade espacial em configurações que exibam diferentes formas de tendência e interação espacial.

Na Tabela 1 tem-se os resultados da análise do poder dos testes baseados nas funções  $L$  homogênea e não homogênea para configurações homogêneas com agrupamentos. Observa-se

que o aumento do número de agrupamentos e do desvio padrão implica em uma redução do poder dos dois testes. As duas funções não apresentam comportamentos diferentes quando uma configuração apresenta pequeno número de agrupamentos e desvio padrão pequeno, ou seja, uma configuração de agrupamentos bem definidos. À medida que se aumenta o desvio padrão, a função  $L$  não homogênea tende a considerar que a configuração apresenta uma completa aleatoriedade espacial. Já para configurações com grande número de agrupamentos e com grande desvio padrão, a função  $L$  homogênea demonstra ser mais poderosa.

As Figuras de 1 a 4 apresentam as curvas do poder dos testes de Kolmogorov-Smirnov baseados nas funções  $L$  homogênea e não homogênea considerando configurações com 2, 5, 10 e 20 agrupamentos. Pode-se observar nessas figuras que o poder dos testes diminui quando o desvio padrão dos agrupamentos aumenta, ou seja, o poder diminui a medida que a configuração se aproxima da completa aleatoriedade espacial. Observa-se também que o teste baseado na função  $L$  homogênea tende a ser mais poderoso do que os teste baseado na função  $L$  não homogênea.

As funções  $L$  homogênea e não homogênea tem por objetivo analisar o efeito de segunda ordem com a finalidade de detectar se o processo exibe interação na presença de homogeneidade espacial. Observa-se que em configurações com agrupamentos bem definidos os testes baseados nas duas funções  $L$  detectam perfeitamente a existência de interação, porém, quando o desvio padrão aumenta e, conseqüentemente, as configurações caminham para a completa aleatoriedade espacial o teste baseado na função  $L$  homogênea tende a ser mais poderoso.

Deve-se observar que o estimador da função  $L$  não homogênea utiliza uma ponderação dada pelas intensidades locais estimadas e, conseqüentemente, reduz o efeito do agrupamento da configuração, pois o mesmo é tratado como uma tendência. Este fato pode explicar o baixo poder do teste que utiliza essa função para detectar agrupamentos em configurações que sabidamente apresentam interação para a formação de agrupamentos de pontos.

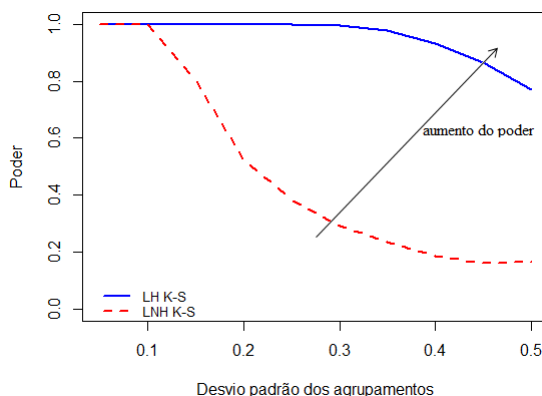


Figura 1: Poder dos testes de Kolmogorov-Smirnov em configurações com dois agrupamentos, em que LH K-S é o teste baseado na função  $L$  homogênea e LNH K-S é o teste baseado na função  $L$  não homogênea.

Os resultados obtidos neste estudo corroboram os resultados obtidos por Baddeley et al. (2000) em alguns casos típicos de configurações pontuais. Esses autores afirmam que a função  $L$  não homogênea é especialmente designada para processos não homogêneos e ao ser utilizada em processos homogêneos pode-se cometer o erro de se tomar uma decisão que não seja verdadeira. Porém na literatura não existem trabalhos que mostram o poder do teste baseado na função  $L$  não homogênea. Myles et al. (1995), Diggle (1979) e Ripley (1981) confirmam que testes baseados na função  $L$  homogênea tendem a apresentar alto poder para detectar agrupamentos em processo pontuais homogêneos.

Tabela 1: Poder dos testes de Kolmogorov-Smirnov baseados nas funções  $L$  homogênea ( $L$ ) e não homogênea ( $L_{inhom}$ ) em configurações homogêneas com agrupamentos.

Número agrupamento	$\sigma$	Teste da função $L$	Teste da função $L_{inhom}$
02	0,05	1,000	1,000
02	0,10	1,000	1,000
02	0,15	1,000	0,803
02	0,20	1,000	0,526
02	0,25	0,998	0,385
02	0,30	0,994	0,291
02	0,35	0,978	0,235
02	0,40	0,933	0,187
02	0,45	0,864	0,164
02	0,50	0,771	0,165
05	0,05	1,000	1,000
05	0,10	1,000	0,999
05	0,15	1,000	0,584
05	0,20	0,998	0,200
05	0,25	0,982	0,098
05	0,30	0,924	0,043
05	0,35	0,789	0,024
05	0,40	0,589	0,010
05	0,45	0,473	0,006
05	0,50	0,340	0,006
10	0,05	1,000	1,000
10	0,10	1,000	0,941
10	0,15	0,999	0,178
10	0,20	0,954	0,024
10	0,25	0,827	0,003
10	0,30	0,633	0,002
10	0,35	0,395	0,000
10	0,40	0,258	0,000
10	0,45	0,145	0,000
10	0,50	0,088	0,000
20	0,05	1,000	1,000
20	0,10	1,000	0,406
20	0,15	0,928	0,002
20	0,20	0,643	0,000
20	0,25	0,408	0,000
20	0,30	0,181	0,000
20	0,35	0,094	0,000
20	0,40	0,026	0,000
20	0,45	0,008	0,000
20	0,50	0,001	0,000

## 5 Poder dos testes em configurações homogêneas com interação para regularidade

A Tabela 2 apresenta os resultados da análise do poder dos testes baseados nas funções  $L$  homogênea e não homogênea para configurações homogêneas com interação entre os pontos para regularidade. Observa-se na Tabela 2 que o aumento da distância mínima entre os eventos



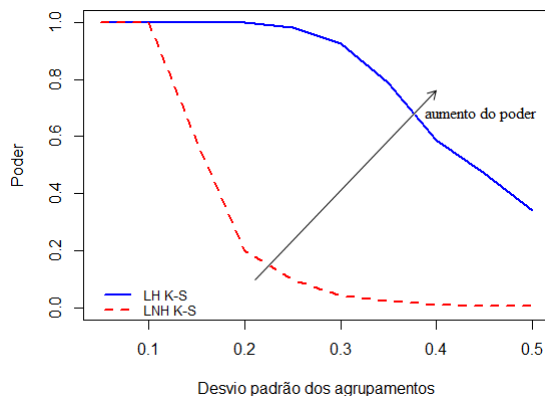


Figura 2: Poder dos testes de Kolmogorov-Smirnov em configurações com cinco agrupamentos, em que LH K-S é o teste baseado na função  $L$  homogênea e LNH K-S é o teste baseado na função  $L$  não homogênea.

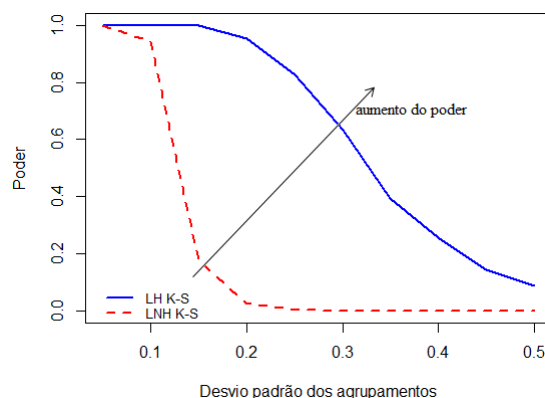


Figura 3: Poder dos testes de Kolmogorov-Smirnov em configurações com dez agrupamentos, em que LH K-S é o teste baseado na função  $L$  homogênea e LNH K-S é o teste baseado na função  $L$  não homogênea.

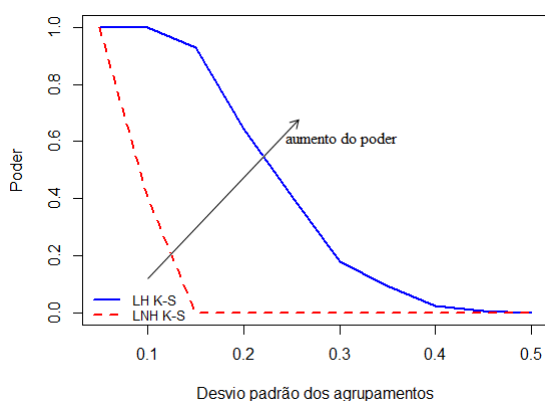


Figura 4: Poder dos testes de Kolmogorov-Smirnov em configurações com vinte agrupamentos, em que LH K-S é o teste baseado na função  $L$  homogênea e LNH K-S é o teste baseado na função  $L$  não homogênea.

implica no aumento do poder dos dois testes. As duas funções não apresentam diferenças significantes quanto ao poder para todas as distâncias mínimas entre os pontos avaliadas.

Tabela 2: Poder dos testes de Kolmogorov-Smirnov baseados nas funções  $L$  homogênea ( $L$ ) e não homogênea ( $L_{inhom}$ ) em configurações homogêneas com regularidade.

Distância	Teste da função $L$	Teste da função $L_{inhom}$
0,010	0,141	0,107
0,012	0,238	0,182
0,014	0,430	0,350
0,016	0,805	0,610
0,018	0,994	0,996
0,020	1,000	1,000
0,030	1,000	1,000
0,040	1,000	1,000
0,050	1,000	1,000
0,060	1,000	1,000

A Figura 5 apresenta as curvas do poder dos testes de Kolmogorov-Smirnov baseados nas funções  $L$  homogênea e não homogênea considerando configurações com regularidade. Pode-se observar nessa figura que o poder dos dois testes diminui quando a distância mínima entre os pontos diminui, ou seja, os testes perdem poder a medida que a configuração se aproxima da completa aleatoriedade espacial.

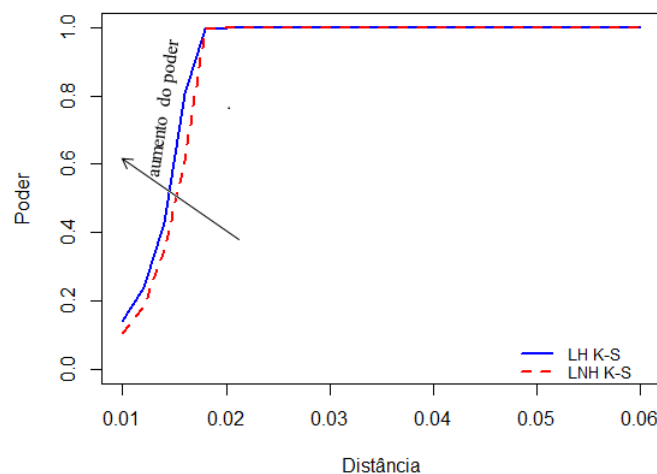


Figura 5: Poder dos testes de Kolmogorov-Smirnov em configurações com regularidade, em que LH K-S é o teste baseado na função  $L$  homogênea e LNH K-S é o teste baseado na função  $L$  não homogênea.

Observa-se também na Figura 5 que as duas funções conduzem a testes que apresentam praticamente o mesmo poder para qualquer distância mínima entre os pontos. Assim, o fato do estimador da função  $L$  não homogênea utilizar uma ponderação dada pelas intensidades estimadas não altera o poder do teste em configurações com regularidade. Isso pode ser explicado pelo fato de que em configurações regulares, a intensidade é praticamente constante em toda a área de estudo.

### 5.1 Poder dos testes em configurações não homogêneas

A Tabela 3 apresenta os resultados da análise do poder dos testes de Kolmogorov-Smirnov baseados nas funções  $L$  homogênea e não homogênea para configurações não homogêneas e sem interação entre os pontos. Observa-se na Tabela 3 que o aumento da intensidade implica no

aumento do poder dos dois testes, principalmente, no teste baseado na função  $L$  não homogênea.

Tabela 3: Poder dos testes de Kolmogorov-Smirnov baseados nas funções  $L$  homogênea e ( $L$ ) não homogênea ( $L_{inhom}$ ) em configurações não homogêneas sem interação entre os pontos.

Parâmetro beta da função intensidade	Teste da função $L$	Teste da função $L_{inhom}$
0	0,046	0,037
1	0,646	0,021
2	0,993	0,042
3	1,000	0,151
4	1,000	0,345
5	1,000	0,395

A Figura 6 apresenta as curvas do poder dos testes de Kolmogorov-Smirnov baseados nas funções  $L$  homogênea e não homogênea considerando configurações com tendência. Pode-se observar nessas figuras que o poder dos dois testes diminui quando o parâmetro beta que especifica a força da tendência diminui, ou seja, os testes perdem poder a medida que a configuração se aproxima da completa aleatoriedade espacial.

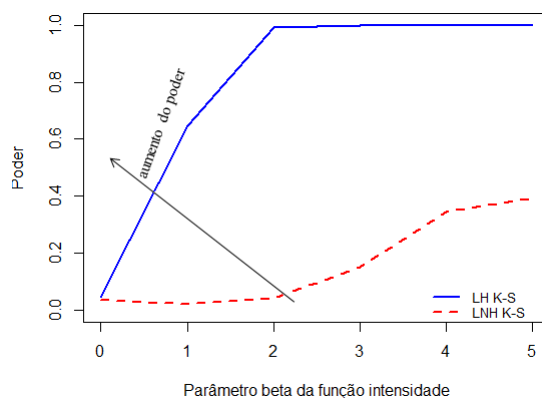


Figura 6: Poder dos testes de Kolmogorov-Smirnov em configurações com tendência espacial (não homogeneidade), em que LH K-S é o teste baseado na função  $L$  homogênea e LNH K-S é o teste baseado na função  $L$  não homogênea.

As funções  $L$  homogênea e não homogênea tem por objetivo analisar o efeito de segunda ordem com a finalidade de detectar se o processo exibe interação na presença de homogeneidade espacial. Nas configurações simuladas nesta seção ocorre exatamente o contrário, ou seja, as configurações não exibem interação, mas exibem tendência. Como o teste baseado na função  $L$  homogênea considera que o processo pontual seja homogêneo, é natural que este teste identifique interação para agrupamentos sempre que a intensidade aumentar, mesmo que o aumento dessa intensidade seja devido a algum efeito de primeira ordem e não devido a existência de interação entre os pontos. Isso explica porque o teste baseado na função  $L$  homogênea tenha apresentado um maior poder do que o teste baseado na função  $L$  não homogênea. Assim, neste caso, o poder desse teste deve ser analisado com cautela. O baixo poder da função  $L$  não homogênea em configurações não homogêneas que não exibem interação é um bom diferencial para esse teste confirmando que deveria ser usado sempre que existir a suspeita de não homogeneidade espacial.

## 6 Aplicação a dados reais que apresentam configuração não homogênea

No controle de processos industriais pode ser necessário utilizar análise de processos pontuais para avaliar se um determinado produto atende as especificações estabelecidas para que esse produto possa ser utilizado com êxito. Por exemplo, na fabricação de filtros é comum utilizar partículas metálicas dentro de um outro material para melhorar o processo de filtragem (MEYER-OLBERSLE-BEN e KIEBACK, 1997).

Este é o caso da fabricação de filtros de bronze pelo método de sinterização, em que partículas de bronze são pulverizadas sobre um meio poroso. Espera-se que as partículas menores tendam a se depositarem na parte inferior do meio poroso, enquanto as partículas maiores tendem a ficar em suspensão na parte superior do filtro (veja Figura 7). Esse comportamento no processo de fabricação fornece uma característica ideal para esse tipo de filtro. Assim, o fabricante está interessado em saber se a distribuição das partículas foi efetuada de maneira correta no processo de produção do filtro, ou seja, se a distribuição de partículas de bronze no filtro foi feita de forma não homogênea (HAHN et al., 2003).

Para analisar a distribuição de partículas de bronze no filtro, o fabricante corta longitudinalmente um filtro e recorta uma área plana retangular de  $18 \times 7$  mm, com área de  $126 \text{ mm}^2$  e identifica a posição e o raio de todas as partículas no plano conforme mostra à Figura 7 (MEYER-OLBERSLEBEN e KIEBACK, 1997).

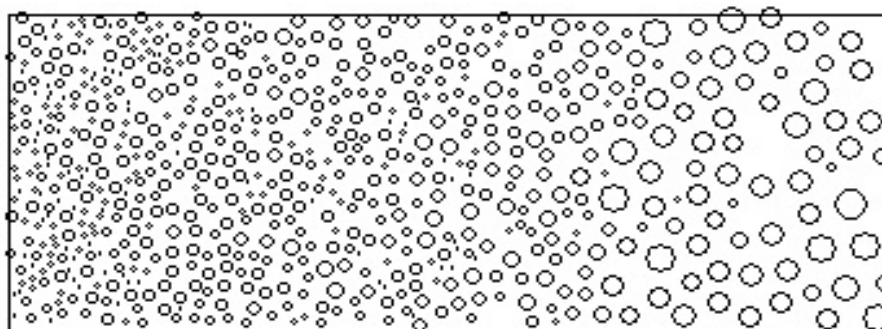


Figura 7: Distribuição das 678 partículas de bronze em uma área plana de  $126 \text{ mm}^2$ . Os tamanhos da partículas são proporcional ao seu raio

A distribuição espacial dos centros das 678 partículas no plano é apresentada na Figura 8. A Figura 8 mostra evidências que configuração de pontos exibe não homogeneidade (tendência). Para uma melhor caracterização dessa tendência utiliza-se o estimador de intensidade local de kernel. A análise da imagem suavizada da intensidade kernel é um passo preliminar com intuito de compreender que tipo de processo pontual espacial segue os dados (RIPLEY, 1981). Uma imagem da intensidade das partículas na área está apresentada na Figura 9.

Pode-se observar na Figura 9 que a intensidade estimada está entre uma e catorze partículas por  $\text{mm}^2$ . Além disso, a intensidade aumenta na direção leste-oeste. Na parte oeste da área, a intensidade estimada média é, aproximadamente, o triplo da intensidade estimada na parte leste da área do filtro. Estas informações são importantes sob o ponto de vista prático, pois possibilita ao fabricante avaliar o processo de produção do filtro. Além disso, o gradiente observado nas intensidades estimadas fornecem indícios suficientes de que o processo estocástico espacial que gera a posição das partículas é não homogêneo.

A análise da intensidade espacial é importante para direcionar as próximas etapas da análise. Tendo em vista os resultados obtidos na seção anterior mostrou que não é recomendável usar a função  $L$  homogênea na análise da interação em configurações não homogêneas, utiliza-se nas próximas etapas da análise métodos baseados na função  $L$  não homogênea.

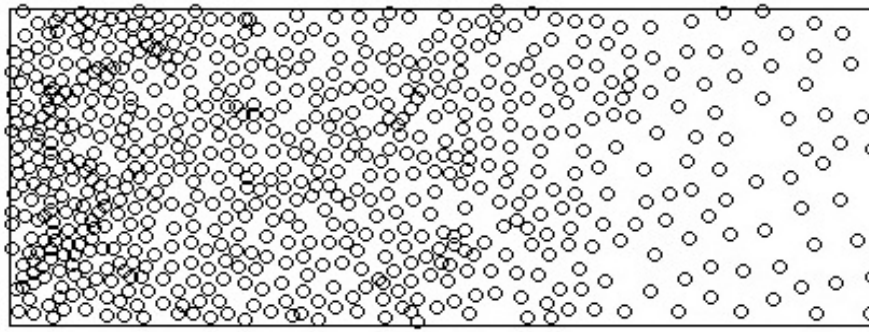


Figura 8: Configuração pontual dos centros das 678 partículas de bronze em uma área de  $126 \text{ mm}^2$ .

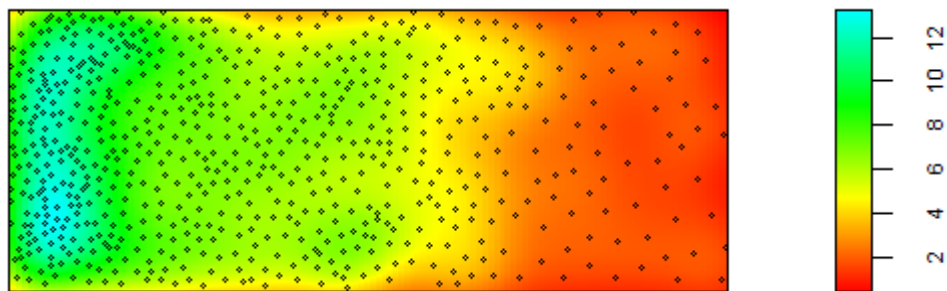


Figura 9: Estimativas kernel da intensidade das partículas de bronze no área de  $126 \text{ mm}^2$ .

Primeiramente, é conduzido o teste de Kolmogorov-Smirnov com a função  $L$  não homogênea para testar a hipótese nula de que as partículas de bronze exibe completa aleatoriedade espacial. Os resultados, obtidos foram de 0,1972 para a estatística do teste e 0.01 para o *valor - p*, portanto se rejeita a hipótese nula com nível de significância de 5%.

O teste formal não mostra a direção (regularidade ou agrupamento) que a hipótese nula foi rejeitada. Assim, foram construídos os envelopes de simulação que são apresentados na Figura 10.

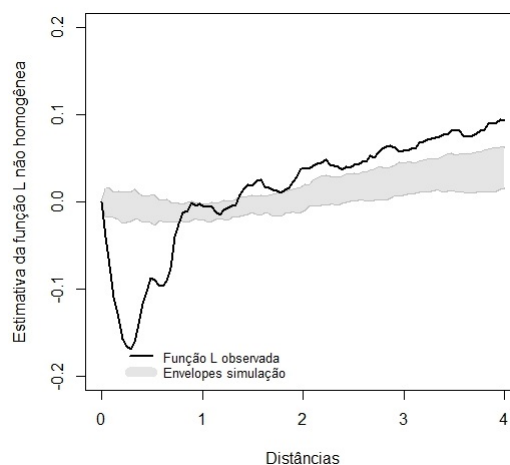


Figura 10: Estimativas da função  $L$  não homogênea ( $L_{inhom}$ ), representada pela linha cheia e envelopes de simulação sob a hipótese de completa aleatoriedade espacial com 99% de confiança (área em cinza).

Na Figura 10 observa-se que existe uma distância mínima entre as partículas indicando regularidade até uma distância de 0,8 (mm). Essa regularidade pode ser explicada pelos tamanhos das partículas que não foram considerados na análise. A partir da distância de dois (mm) as partículas tendem formar agrupamentos. Esse comportamento pode ser resultado da atuação de dois processos estocásticos para explicar a interação entre as partículas de bronze no filtro: regularidade para pequenas escalas e agrupamento para escalas maiores. Como a função  $L$  não homogênea (corrige) o efeito de primeira ordem, pode-se afirmar que existe interação entre as partículas de bronze.

Para uma análise mais objetiva da configuração espacial das partículas de bronze seria necessário ajustar um modelo que tenha parâmetros que representam as duas propriedades (primeira e segunda ordem). Alguns modelos que apresentam essas características são descritos em Baddeley (2008), Diggle (2003) e Illian et al. (2008).

## 7 Conclusões

Os resultados mostraram que a função  $L$  homogênea não deveria ser usada em configurações suspeita de serem não homogêneas e que a função  $L$  não homogênea não deveria ser usada em configurações homogêneas. A análise espacial dos dados reais mostrou que a distribuição das partículas de bronze apresentam não homogeneidade (tendência). Os métodos baseados na função  $L$  não homogênea detectou a presença de interação (regularidade e agrupamento) entre as partículas.

## Referências

- [1] BADDELEY, A. J.; MØLLER, J.; WAAGEPETERSEN, R. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. **Statistica Neerlandica**, Hoboken, v. 54, n. 3, p. 329-350, 2000.
- [2] BADDELEY, A; TURNER, R. Spatstat: an R package for analyzing spatial point patterns. **Journal of Statistical Software**, Los Angeles, v. 12, n. 6, p. 1-42, 2005.
- [3] BADDELEY, A. Analysing spatial point patterns in R. **Workshop Notes**, Copyright CSIRO, Australia, February, 2008. Disponível em: <http://www2.stat.unibo.it/bruno/dati%20spaziali%200809/spatialpointpatterns.pdf>. Acesso em: 11 Jun. 2015.
- [4] BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. New York: Longman Scientific e Technical, p. 413, 1995.
- [5] BESAG, J.; DIGGLE, P. J. Simple Monte Carlo tests for spatial pattern, **Journal of the Royal Statistical Society Series. C (Applied Statistics)**, Hoboken, v. 26, n. 3, p. 327-333, 1977.
- [6] CRESSIE, N. A. C. **Statistics for spatial data**. New York: John Wiley e Sons, p. 900, 1991.
- [7] DIGGLE, P.J. On parameters estimation and goodnessoffit testing for spatial point patterns. **Biometrics**, Washington, v. 35, n. 1, p. 87-101, 1979.
- [8] DIGGLE, P. J. **Statistical analysis of spatial point patterns**. 2. ed. London: Hodder Arnold Publishers, p. 159, 2003.
- [9] HAHN, U.; JENSEN, E.B.; VAN LIESHOUT, M.-C.; NIELSEN, L.S. Inhomogeneous spatial point processes by location-dependent scaling, **Advances in Applied Probability**, Sheffield, p. 319-336, 2003.

- [10] MATERN, B. **Spatial Variation**. Lecture Notes in Statistics, Springer-Verlag, New York, v. 36, 1986.
- [11] MYLES, J. P.; FLENLEY, N. R. J.; ATKINSON, H. V.; JONES, H. Statistical tests for clustering of second phases in composite materials. **Philosophical Magazine A**, London, v. 72, n. 2, p. 515-528, 1995.
- [12] R CORE TEAM. **R: a language and environment for statistical computing**. Vienna, Austria, 2016. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org/>. Acesso em: 11 janeiro 2016.
- [13] RIPLEY, B. D. Modelling Spatial Patterns. **Journal of the Royal Statistical Society. Series B (Methodological)**, Hoboken, v. 39, n. 2, p. 172- 212, 1977.
- [14] RIPLEY, B. D. **Spatial Statistical**. New York: John Wiley e Sons, p. 183, 1981.
- [15] WAAGEPETERSEN, R. P. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. **Biometrics**, v. 63, n. 1, p. 252-258, 2007.