



SIDCLEIDE BARBOSA DE SOUSA

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

LAVRAS - MG

2016

SIDCLEIDE BARBOSA DE SOUSA

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

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador
Dr. João Domingos Scalon

**LAVRAS - MG
2016**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Sousa, Sidcleide Barbosa de.

Funções L homogênea e não homogênea em testes de hipótese de completa aleatoriedade espacial / Sidcleide Barbosa de Sousa. – Lavras: UFLA, 2016.

70 p. : il.

Dissertação (mestrado acadêmico)–Universidade Federal de Lavras, 2016.

Orientador(a): João Domingos Scalon.

Bibliografia.

1. Estatística espacial. 2. Processo de Poisson. 3. Função K de Ripley. 4. Intensidade kernel. 5. Monte Carlo. 6. Teste de hipótese. I. Universidade Federal de Lavras. II. Título.

SIDCLEIDE BARBOSA DE SOUSA

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

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 22 de Janeiro de 2016.

Dr. Denismar Alves Nogueira

UNIFAL-MG

Dr. Renato Ribeiro de Lima

UFLA-MG

Dr. João Domingos Scalon
Orientador

LAVRAS - MG

2016

DEDICO

*A DEUS o autor da minha VIDA
Aos meus pais ROSALIA e SEVERINO
Aos meus irmãos em especial a SIVONEIDE que
colaboraram com compreensão e amor nessa jornada
Aos meus sobrinhos em especial DEIVISSON
Obrigado por vocês fazerem parte da minha vida
Por que família é tudo.*

AGRADECIMENTO

Em primeiro lugar agradeço a DEUS, por ter me dado força e pela vitória a mim concedida.

Ao meu orientador, Professor Dr. João Domingos Scalon, pela orientação, incentivo, confiança, dedicação e paciência durante todo o desenvolvimento desse trabalho.

A todos os Professores da Pós-Graduação em Estatística e Experimentação Agropecuária da UFLA.

A Universidade Federal de Lavras e ao Departamento de Ciência Exatas, pela realização desse curso.

A CAPES pela concessão da bolsa de mestrado para realização desse trabalho.

A toda a minha família: aos meus pais Rosalia e Severino, aos meus irmãos, em especial a Sivoneide e aos meus sobrinhos, em especial Deivisson Barbosa que é o querido da Titia, pelo apoio, pelo amor incondicional em todos os momentos, fazendo-se sempre presente ao longo do curso mesmo estando distante. A todos vocês o meu muito obrigado, eu os amo.

A todos os amigos e colegas de pós-graduação em Estatística e Experimentação Agropecuária.

A todos os funcionários do DEX pelo serviços prestados.

Às meninas que dividiram o apartamento comigo, Janaína e Kelly, obrigado por tudo.

A todos que contribuíram, direta ou indiretamente, meu eterno agradecimento.

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 nesta dissertação 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 dissertation 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.

LISTA DE FIGURAS

Figura 1	Tipos básicos de realizações de configurações pontuais: aleatório (A), regular (B) e aglomerado (C).	16
Figura 2	Realização de um processo de Poisson homogêneo em uma área unitária com intensidade igual a 100.	20
Figura 3	Realização de um processo de Poisson não homogêneo no quadrado unitário com intensidade $\lambda(\mathbf{x}) = 300 * \exp(-3 * x)$ com 100 pontos.	21
Figura 4	Representação gráfica do estimador kernel de intensidade local.	22
Figura 5	Intensidades locais estimadas por uma função kernel em uma realização do processo de Poisson homogêneo no quadrado unitário com 100. A escala representa a intensidade, ou seja, número de eventos por unidade de área.	24
Figura 6	Intensidades locais estimadas por uma função kernel em uma realização de um processo de Poisson não homogêneo com intensidade $\lambda(x) = 300 * \exp(-3 * latitude)$ no quadrado unitário com 100 pontos. A escala representa a intensidade, ou seja, número de eventos por unidade de área.	25
Figura 7	Ideia gráfica do estimador da função K	27
Figura 8	Funções K estimada (linha cheia) e teórica sob a hipótese de completa aleatoriedade espacial (linha tracejada) para a realização de um processo de Poisson homogêneo.	27
Figura 9	Funções L estimada (linha cheia) e teórica sob a hipótese de completa aleatoriedade espacial (linha tracejada) para a realização de um processo de Poisson homogêneo.	28
Figura 10	Envelopes de simulação seguindo o Processo de Poisson Homogêneo (área cinza) com a estimativa da função L homogênea (linha cheia) da configuração da Figura 1 (B).	31
Figura 11	Poder dos testes de Kolmogorov-Smirnov em configurações com dois agrupamentos, em que LH K-S é o teste baseado na função	

	<i>L</i> homogênea e LNH K-S é o teste baseado na função <i>L</i> não homogênea.	41
Figura 12	Poder dos testes de Kolmogorov-Smirnov em configurações com cinco agrupamentos, em que LH K-S é o teste baseado na função <i>L</i> homogênea e LNH K-S é o teste baseado na função <i>L</i> não homogênea.	41
Figura 13	Poder dos testes de Kolmogorov-Smirnov em configurações com dez agrupamentos, em que LH K-S é o teste baseado na função <i>L</i> homogênea e LNH K-S é o teste baseado na função <i>L</i> não homogênea.	42
Figura 14	Poder dos testes de Kolmogorov-Smirnov em configurações com vinte agrupamentos, em que LH K-S é o teste baseado na função <i>L</i> homogênea e LNH K-S é o teste baseado na função <i>L</i> não homogênea.	42
Figura 15	Poder dos testes de Kolmogorov-Smirnov em configurações com regularidade, em que LH K-S é o teste baseado na função <i>L</i> homogênea e LNH K-S é o teste baseado na função <i>L</i> não homogênea.	44
Figura 16	Poder dos testes de Kolmogorov-Smirnov em configurações com tendência espacial (não homogeneidade), em que LH K-S é o teste baseado na função <i>L</i> homogênea e LNH K-S é o teste baseado na função <i>L</i> não homogênea.	45
Figura 17	Distribuição das 678 partículas de bronze em uma área plana de 126 mm^2 . Os tamanhos da partículas são proporcional ao seu raio	47
Figura 18	Configuração pontual dos centros das 678 partículas de bronze em uma área de 126 mm^2	47
Figura 19	Estimativas kernel da intensidade das partículas de bronze no área de 126 mm^2	48
Figura 20	Estimativas da função <i>L</i> 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).	49

LISTA DE TABELAS

Tabela 1	Poder dos testes de Kolmogorov-Smirnov baseados nas funções L homogênea (L) e não homogênea (L_{inhom}) em configurações homogêneas com agrupamentos.	40
Tabela 2	Poder dos testes de Kolmogorov-Smirnov baseados nas funções L homogênea (L) e não homogênea (L_{inhom}) em configurações homogêneas com regularidade.	43
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.	45

SUMÁRIO

1	INTRODUÇÃO	12
2	REFERENCIAL TEÓRICO	14
2.1	Estatística espacial	14
2.2	Processos pontuais espaciais	15
2.2.1	Propriedade de primeira ordem	16
2.2.2	Propriedade de segunda ordem	17
2.2.3	Completa aleatoriedade espacial (CAE)	18
2.3	Processos de Poisson no espaço	19
2.3.1	Processo de Poisson homogêneo	19
2.3.2	Processo de Poisson não homogêneo	20
2.4	Análise de processos pontuais	21
2.4.1	Estimação por kernel	22
2.4.2	Função K para processos pontuais homogêneos	24
2.4.3	Função K para processos pontuais não homogêneos	28
2.5	Métodos para testar a hipótese de CAE	29
2.6	Método gráfico para testar a hipótese de CAE	30
2.7	Teste formal contra a hipótese nula de CAE	31
2.8	Poder do teste e erros do tipo I e II	32
3	MATERIAIS E MÉTODOS	34
3.1	Simulação das configurações	34
3.2	Testes de hipóteses	35
3.3	Análise do poder dos testes	36
3.4	Aplicação	36
3.5	Software	37
4	RESULTADOS E DISCUSSÃO	38
4.1	Poder dos testes em configurações homogêneas com interação para regularidade e agrupamentos	38
4.2	Poder dos testes em configurações não homogêneas	44
4.3	Aplicação a dados reais que apresentam configuração não homogênea	46
5	CONCLUSÕES	50
	REFERÊNCIAS	51
	ANEXO	54

1 INTRODUÇÃO

Analisar dados de fenômenos que ocorrem no espaço é um grande desafio em diversas áreas do conhecimento tais como saúde, geologia, agronomia, etc.

Alguns fenômenos podem ser expressos apenas através de ocorrências identificadas como pontos (coordenadas) localizados no espaço (plano) e denominados processos pontuais. Exemplos desse tipo de processos são: locais de ocorrência de crimes, localização de indivíduos com uma determinada doença 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 análise no sentido de verificar se a configuração apresenta pontos que exibam alguma forma de tendência ou interação.

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. 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. 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 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. 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 desta dissertação 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 do presente trabalho é aplicar os métodos descritos nesta dissertação em uma configuração pontual não homogênea. Para tal, utilizou-se as coordenadas dos centros de partículas de bronze localizadas em uma secção plana longitudinal de um filtro fabricado pelo método de sinterização.

2 REFERENCIAL TEÓRICO

2.1 Estatística espacial

A análise espacial está relacionada com métodos estatísticos em que a localização espacial do fenômeno em estudo desempenha um papel explícito na análise dos dados. Segundo Ripley (1976), a estatística espacial é definida como uma técnica que busca descrever os padrões existentes nos dados espaciais, com estudos e modelagem de processo que se expressam por meio de uma distribuição no espaço. A diferença entre a estatística clássica e a estatística espacial é que na estatística espacial a posição espacial do dado é explicitamente incluída no modelo.

Para Cressie (1993), a estatística espacial está relacionada com a natureza estocástica da observação, sendo assim, os fenômenos espaciais podem ser analisados conceitualmente em três categorias, de acordo com a natureza dos mesmos: dados de processos pontuais, dados de superfícies contínuas e dados de área.

Dados de Processos Pontuais: nesses tipos de dados, os fenômenos são expressos por meio de ocorrências identificadas como pontos localizados no espaço. Sendo assim, um dos interesses do pesquisador é analisar a intensidade da ocorrência de eventos e avaliar a existência de interação (dependência) entre os pontos (aglomerados, regular e aleatório). Existem vários exemplos de fenômenos que podem ser analisados sob a perspectiva de processos pontuais tais como: localização de indivíduos de uma determinada espécie, ocorrência de crimes e casos de doenças (DIGGLE, 2003).

Dados de Superfície Contínua (Geoestatística): esses tipos de dados representam fenômenos que se distribuem continuamente em uma região. Os fenômenos naturais tais como: temperatura e pressão atmosférica são características que podem ser observadas e medidas em qualquer parte da superfície terrestre. Esses tipos de dados são resultados, em geral, de levantamentos de recursos naturais, incluindo mapas geológicos, topográficos e ecológicos. Concentração de certa medida de elemento químico no solo é um exemplo desse tipo de dado (BAILEY; GATRELL, 1995).

Dados de Áreas: nesse caso, os fenômenos são associados a dados de lev-

antamentos populacionais, tais como: censos e estatística da saúde. Assim, pode-se dizer que esses dados são agregados em unidades de análises, normalmente delimitado por polígonos fechados. Como exemplos tem-se o número de casos de dengue nos diversos bairros de uma cidade, taxas de mortalidade e natalidade de determinadas regiões. Dessa forma, o interesse do pesquisador é detectar padrões subjacentes ou quantificar o grau de associação espacial na região dispondo de medições individuais associadas a cada sub-região (BAILEY; GATRELL, 1995).

Ao definir as três categorias da análise espacial, pode-se dizer que a principal finalidade é a escolha do modelo inferencial que explique os relacionamentos espaciais existentes nos fenômenos em estudos. Geralmente, o processo de análise é realizado utilizando-se gráficos e mapas, que identificam os padrões de dependência espacial no fenômeno em estudo. Neste trabalho concentra-se na análise de processos pontuais no espaço.

2.2 Processos pontuais espaciais

Segundo Diggle (2003), processos pontuais são definidos estatisticamente como sendo um conjunto de pontos irregularmente distribuídos no espaço, no qual a localização e o número de realizações foram gerados por mecanismos estocásticos. O conjunto das localizações espaciais observadas $Z = \{x_1, \dots, x_n\}$, é chamada de configuração pontual, em que x_i representa a coordenada da localização do i -ésimo evento, para distingui-la de pontos arbitrários no plano. Segundo Câmara et al. (2004), o termo "*evento*" refere-se a qualquer fenômeno localizado no espaço dentro da escala de investigação que possa estar associado a uma representação pontual.

Na verificação do padrão de ocorrência do fenômeno na área de estudo, essencialmente, são considerados três tipos básicos de configurações pontuais: aleatório, regular (uniforme) e aglomerado (as ocorrências estão concentradas em regiões no espaço de forma mais agregada do que seria esperado em uma configuração aleatória). Na Figura 1 mostra-se uma realização simulada de cada um desses diferentes tipos de possibilidades.

A Figura 1A mostra-se uma área quadrada de $32,49m^2$ com 65 mudas

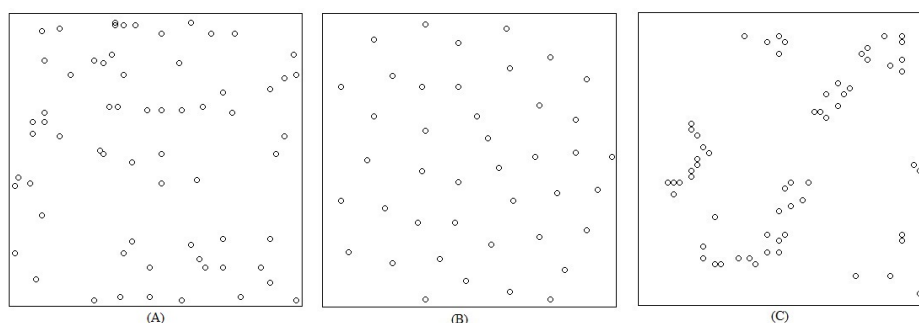


Figura 1 Tipos básicos de realizações de configurações pontuais: aleatório (A), regular (B) e aglomerado (C).

de pinheiro japoneses que se distribuem no espaço de maneira completamente ao acaso. Na Figura 1B observa-se 42 núcleos de células biológicas com uma configuração pontual regular. Pode-se observar a existência de uma distância mínima entre os núcleos, ou seja, a distribuição da ocorrência dos pontos na área tende a ser uniforme. A Figura 1C apresenta a distribuição com agrupamentos de 62 mudas de uma espécie de sequoia em uma área quadrada de $529m^2$. Essas configurações pontuais são utilizadas por diversos autores (DIGGLE, 2003; BADDELEY; TURNER, 2005) como modelos canônicos de configurações pontuais que exibem, respectivamente, padrões de completa aleatoriedade espacial, regularidade e agrupamentos de eventos.

Segundo Cressie (1993), Baddeley, Møller e Waagepetersen (2000) e Diggle (2003), um processo estocástico pontual pode ser descrito em termos das propriedades de primeira e segunda ordem.

2.2.1 Propriedade de primeira ordem

Segundo Câmara et al. (2002) e Diggle (2003), a propriedade de primeira ordem, de escala global ou de larga escala, correspondem às variações no valor médio do processo pontual no espaço. Dessa forma, o real interesse é na intensidade do processo, ou seja, no número de eventos por unidade de área.

Geralmente a propriedade de primeira ordem é representada por $\lambda(x)$, que é uma função de intensidade. Essa função indica a intensidade do processo numa

certa localização x , sendo essa localização definida por um sistema de coordenadas simples $x = (\textit{latitude}_i, \textit{longitude}_j)$, no plano R^2 .

Considera-se um conjunto de pontos $Z = \{x_1, x_2, \dots, x_n\}$ em uma determinada região de a área $|A|$. O processo pontual é modelado considerando X como uma sub-região de A e dx é uma pequena região em torno do ponto x .

Diggle (2003) descreve a propriedade de primeira ordem de um processo espacial de pontos pela função intensidade de primeira ordem que é dada por:

$$\lambda_1(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{E[Z(dx)]}{|dx|} \right\}, \quad (1)$$

em que $E[Z(dx)]$ denota o valor esperado do número de eventos em dx , dx é uma região infinitesimal em torno do ponto $x = (x_i, x_j)$ e $|dx|$ é a área dessa região.

Para um processo pontual estacionário a intensidade é invariante sob a translação e $\lambda(x)$ assume um valor constante $\lambda = \frac{E[A]}{|A|}$, ou seja, é o número médio de eventos por unidade de área.

2.2.2 Propriedade de segunda ordem

A propriedade de segunda ordem são denominada de local ou de pequena escala e é resultado da dependência (interação) ente os pontos no processo pontual espacial. O efeito de segunda ordem pode ser estimado pelo relacionamento de pares de eventos, o que é semelhante ao cálculo da covariância de variáveis aleatórias. O efeito de segunda ordem pode ser descrito pela função intensidade de segunda ordem que é uma medida de estrutura de dependência entre as localizações x_i e x_j (DIGGLE, 2003).

A propriedade de segunda ordem é definida da mesma forma que a propriedade de primeira ordem. Aqui considera-se a intensidade conjunta $\lambda(x_i, x_j)$ entre duas regiões infinitesimais $|dx_i|$ e $|dx_j|$ que contém os pontos x_i e x_j .

Diggle (2003) descreve a propriedade de segunda ordem de um processo espacial de pontos pela função de intensidade de segunda ordem dada por:

$$\lambda(x_i, x_j) = \lim_{|dx_i|, |dx_j| \rightarrow 0} \left\{ \frac{E[Z(dx_i)Z(dx_j)]}{|dx_i||dx_j|} \right\}, \quad (2)$$

em que o processo pontual espacial é dito estacionário numa região A se $\lambda(x)$ é uma constante, ou seja,

$$\lambda_1(x) = \lambda \quad \forall x \in A.$$

Se o processo pontual espacial é também isotrópico, tem-se.

$$\lambda_2(x_i, x_j) = \lambda_2(h) \quad \forall x_i, x_j \in A,$$

em que $h = x_i - x_j$ é a distância entre os pontos.

Um processo é estacionário, se a média e a variância forem constantes, ou seja, o número esperado de eventos em uma localização arbitrária é constante em qualquer sub-área e a dependência entre os eventos em duas localizações quaisquer depende apenas do vetor de distância h e não das localizações específicas x_i e x_j (CÂMARA et al., 2004; DIGGLE, 2003).

O processo será isotrópico se for estacionário e a dependência for em função da distância entre os eventos e não da direção. Assim, se o processo é estacionário e também isotrópico, a covariância desse processo vai depender da distância entre os pontos e não da direção entre eles (DIGGLE, 2003).

2.2.3 Completa aleatoriedade espacial (CAE)

A análise estatística dos padrões de pontos no espaço necessita de um modelo teórico de referência. O modelo mais simples é aquele que corresponde a uma situação em que é igualmente provável que um evento aconteça em qualquer local na área de estudo. Esse modelo é conhecido como modelo de Completa Aleatoriedade Espacial - CAE (*Complete Spatial Randomness - CSR*). Teoricamente falando, esse modelo corresponde ao processo pontual de Poisson homogêneo, ou seja, as posições dos eventos são independentes e apresentam igual probabilidade de ocorrer em toda a área de estudo (DIGGLE, 2003).

Cressie (1993) observa que a suposição de independência é conveniente pois torna a inferência estatística mais simples. No entanto modelos que envolvem a dependência estatística são frequentemente mais realistas.

2.3 Processos de Poisson no espaço

2.3.1 Processo de Poisson homogêneo

Segundo Illian et al. (2008), o processo de Poisson homogêneo é o modelo mais importante e simples dos processos pontuais, pois esse modelo representa uma configuração aleatória no espaço e sem efeitos de primeira (tendência) e segunda (interação) ordem.

Diggle (2003) descreve que o processo de Poisson homogêneo em uma região de área $|A|$ é caracterizado por algumas propriedades:

PP1 - Para algum $\lambda > 0$ e qualquer região plana finita A , $N(A)$ corresponde ao número de eventos em A , que segue uma distribuição de Poisson com média $\lambda(|A|)$, em que $|A|$ refere-se a área de A .

PP2 - Dado $N(A) = n$, os n eventos em A formam uma amostra aleatória independente com distribuição uniforme em A .

O parâmetro λ é uma constante e refere-se à intensidade do processo, isto é, o número esperado de pontos por unidade de área. Com isso,

$$E[N(A)] = \lambda|A|.$$

Segundo Diggle (2003), para simular uma realização de um processo de Poisson em uma região A , dado um valor de $N(A)$, se faz necessário gerar pontos independentes e uniformemente distribuídos sobre a região A . Na Figura 2 tem-se uma configuração simulada (realização) de um processo de Poisson homogêneo com intensidade 100 sobre uma área unitária.

Apesar do processo de Poisson homogêneo ser útil ao servir como hipótese nula nos testes de completa aleatoriedade espacial, o mesmo raramente pode ser considerado como um modelo plausível para a maioria das situações práticas. Mesmo que não se tenha interação espacial entre os eventos, raras são as vezes que se tem homogeneidade na intensidade. Assim, assumindo que os eventos sejam independentes e com a intensidade $\lambda(x)$ variando no espaço, uma configuração de pontos espaciais pode ser modelada através do processo de Poisson não homogêneo.

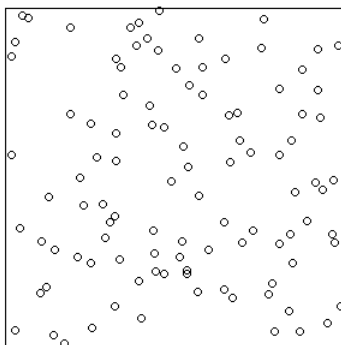


Figura 2 Realização de um processo de Poisson homogêneo em uma área unitária com intensidade igual a 100.

2.3.2 Processo de Poisson não homogêneo

O processo de Poisson não homogêneo é uma generalização do processo de Poisson homogêneo sob a condição de que o processo pontual é não estacionário. Neste caso, a intensidade do processo de Poisson homogêneo é substituída por uma função de intensidade $\lambda(x)$, na qual o valor varia de acordo com a localização x (BIVAND; PEBESMA; GOMES-RUBIO, 2008).

Diggle (2003) descreve que o processo de Poisson não homogêneo apresenta as seguintes propriedades:

PPNH1 - $N(A)$ tem distribuição de Poisson com média $\int_A \lambda(x)dx$.

A função de intensidade de primeira ordem do processo é uma função determinística da localização espacial, com cada ponto dentro da área de estudos e o número total de eventos na sub-região X com $X \subset A$. Dessa forma:

$$N(A) = \int_A \lambda(x)dx$$

Para quaisquer conjunto finitos disjuntos $A_1, A_2, A_3, \dots, A_k$, os números $N(A_1), N(A_2), \dots, N(A_k)$ de eventos são variáveis aleatórias independentes.

Para o número de eventos $N(A)$ em qualquer conjunto A for uma distribuição de Poisson com intensidade $\lambda(x)$ e média,

$$\int_A \lambda(x) dx,$$

isto é

$$P\{N(A) = n\} = \frac{1}{n!} \exp\left\{-\int_A \lambda(x) dx\right\} \left[\int_A \lambda(x) dx\right]^n$$

PPNH2 - Os n eventos em A ($N(A) = n$) formam uma amostra aleatória independente e identicamente distribuída com função densidade de probabilidade $f(x) = \frac{\lambda(x)}{I}$, em que $I = \int_A \lambda(x) dx$.

Na Figura 3 pode-se observar uma configuração simulada de um processo de Poisson não homogêneo com 100 pontos em uma região de área unitária.

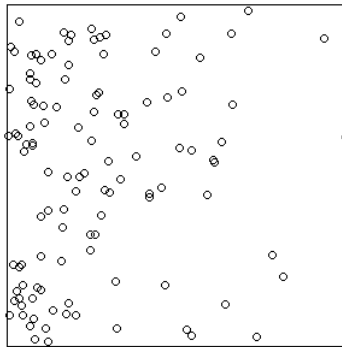


Figura 3 Realização de um processo de Poisson não homogêneo no quadrado unitário com intensidade $\lambda(\mathbf{x}) = 300 * \exp(-3 * x)$ com 100 pontos.

2.4 Análise de processos pontuais

Cressie (1993) afirma que a análise de configurações pontuais pode ser conduzida em termos da estimação de parâmetros que caracterizam as propriedades de primeira e segunda ordem. As propriedades de primeira ordem podem ser analisadas através de estimativas da intensidade do processo utilizando estimadores do tipo kernel. Essas estimativas podem fornecer indícios sobre a estacionariedade

(homogeneidade) do processo Pontual. Para obter os estimadores do efeito de segunda ordem para análise da interação entre os pontos do processo pontual, pode-se utilizar diversas funções tais como a função L homogênea e a função L não homogênea que serão abordadas nesta dissertação.

2.4.1 Estimação por kernel

Um dos primeiros procedimentos na análise de configurações de pontos espaciais é estimar a intensidade local na região de estudo. Neste caso, o estimador kernel é um dos mais utilizados, conforme pode ser visto em (CRESSIE, 1993; DIGGLE, 2003; BADDELEY; TURNER, 2005).

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. Câmara et al. (2004) afirma que o estimador de intensidade de kernel realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância entre cada ponto e uma localização de interesse, isto pode ser visto na Figura 4.

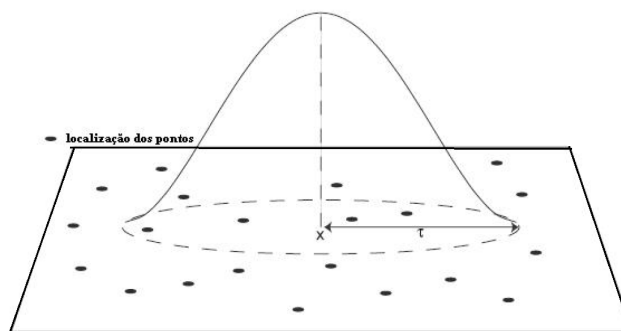


Figura 4 Representação gráfica do estimador kernel de intensidade local.

Fonte: Frade (2014)

Seja x uma localização em qualquer região A , cujo valor da intensidade se deseja estimar. O estimador de intensidade kernel para λ em x é dado por:

$$\hat{\lambda}_\tau(x) = \frac{1}{\delta_\tau}(x) \sum_{i=1}^n \frac{1}{\tau^2} S\left(\frac{d(x, x_i)}{\tau}\right), d(x, x_i) \leq \tau. \quad (3)$$

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 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 τ 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

$$\delta_\tau(x) = \int_A \frac{1}{\tau^2} S\left(\frac{x-u}{\tau}\right) dx. \quad (4)$$

Na Figura 5 pode-se observar uma configuração simulada de um processo de Poisson homogêneo com 100 pontos utilizando o $\tau = 0,3$ em um quadrado de área unitária.

Pode-se observar na Figura 5 que as intensidades estimadas por meio do

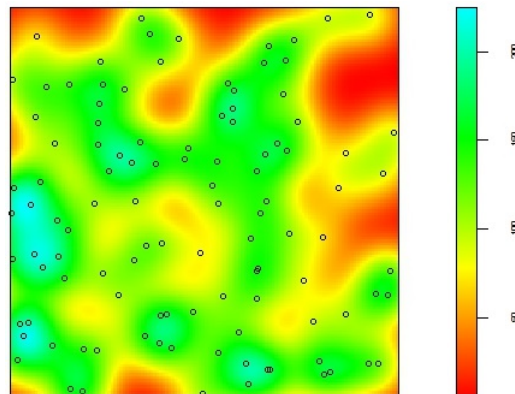


Figura 5 Intensidades locais estimadas por uma função kernel em uma realização do processo de Poisson homogêneo no quadrado unitário com 100. A escala representa a intensidade, ou seja, número de eventos por unidade de área.

estimador kernel flutuam em torno da intensidade do processo de Poisson homogêneo, sugerindo que o processo é estacionário.

Na Figura 6 pode-se observar uma configuração simulada de um processo de Poisson não homogêneo com 100 pontos utilizando o $\tau = 0,3$ na área quadrada unitária.

Pode-se verificar na Figura 6 que a intensidade local da configuração pontual aumenta na direção oeste-leste do eixo "x", o que caracteriza uma tendência típica de processos não estacionários.

2.4.2 Função K para processos pontuais homogêneos

A função K , proposta por Ripley (1977), descreve as propriedades 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(número de eventos contidos dentro de uma distância h a partir de um evento arbitrário).

Sob a suposição de estacionariedade, as funções de intensidade são con-

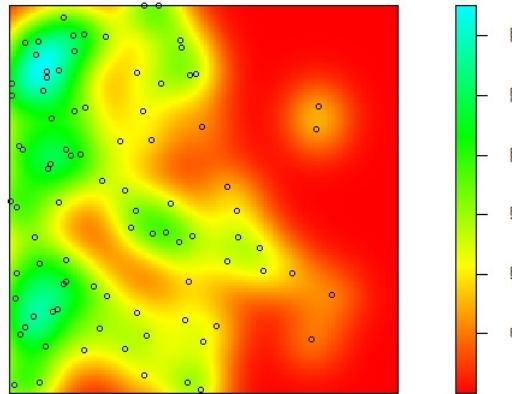


Figura 6 Intensidades locais estimadas por uma função kernel em uma realização de um processo de Poisson não homogêneo com intensidade $\lambda(x) = 300 * \exp(-3 * latitude)$ no quadrado unitário com 100 pontos. A escala representa a intensidade, ou seja, número de eventos por unidade de área.

stantes. Com isso,

$$\lambda_1(x) \equiv \lambda$$

e

$$\lambda_2(x_i, x_j) = \lambda_2(\|x_i - x_j\|) = \lambda_2(x) \equiv \lambda^2$$

em que $\|x_i - x_j\|$ é a distância euclidiana entre os pontos x_i e x_j .

A função K para um processo pontual estacionário, sob a hipótese de CAE, é definida por Cressie (1993) como:

$$\begin{aligned} K(h) &= \frac{2\pi}{\lambda^2} \int_0^h \lambda_2(x) x dx. \\ &= \frac{2\pi}{\lambda^2} \int_0^h \lambda^2 x dx. \\ &= \pi h^2. \end{aligned}$$

Assim, a função K pode ser utilizada para testar a hipótese nula de CAE,

em que λ representa a intensidade do processo na localização x . Para um processo aleatório, 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 o esperado sob a suposição de CAE, para uma distância h .

Um fato importante da função K na análise de configurações de pontos é que ela permite fazer inferências sobre amplos limites de escalas h sem restringir a distância de vizinhos mais próximos, como ocorre com as funções F, G e J (CRESSIE, 1993).

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

$$\hat{K}(h) = \frac{|A|}{n^2} \sum_i^n \sum_{j, i \neq j}^n \frac{I_h(d_{ij})}{w_{ij}}, \quad (5)$$

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.

Para entender melhor o que ocorre no estimador da função K , tem-se que um círculo de raio h centrado em um evento. Em torno desse evento é construído um círculo e uma contagem do número de eventos dentro de cada círculo é realizada. Todos os outros eventos são, similarmente, visitados e a contagem do número de eventos na distância de raio h em volta de todos os eventos é o estimador do numerador da função K , como mostra na Figura 7.

Para cada círculo de raio h existe uma distância d_{ij} que mede as distâncias entre os eventos desse círculo. Não se tem regra para a escolha da distância h . A escolha vai depender do objetivo estabelecido pelo pesquisador e pela área de estudo.

O gráfico da função K é uma parábola (ver Figura 8). Por isso, Besag e Diggle (1977) fizeram uma proposta de um estimador modificado, denominado de

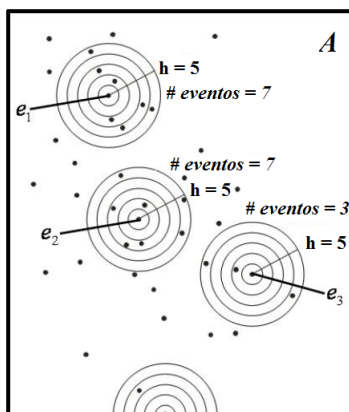


Figura 7 Ideia gráfica do estimador da função K .

Fonte: Baily e Gatrell (1995)

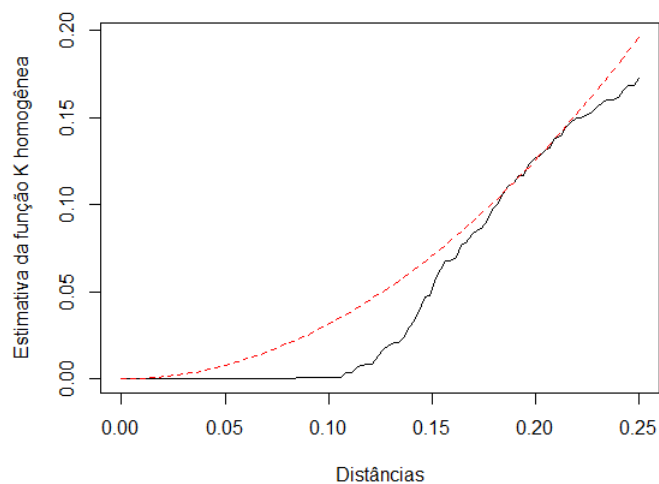


Figura 8 Funções K estimada (linha cheia) e teórica sob a hipótese de completa aleatoriedade espacial (linha tracejada) para a realização de um processo de Poisson homogêneo.

função L , que é dado por:

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}}. \quad (6)$$

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 distâncias 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 Figura 9 mostra as funções L estimada e teórica, sob a hipótese de completa aleatoriedade espacial, para uma realização de um processo de Poisson homogêneo.

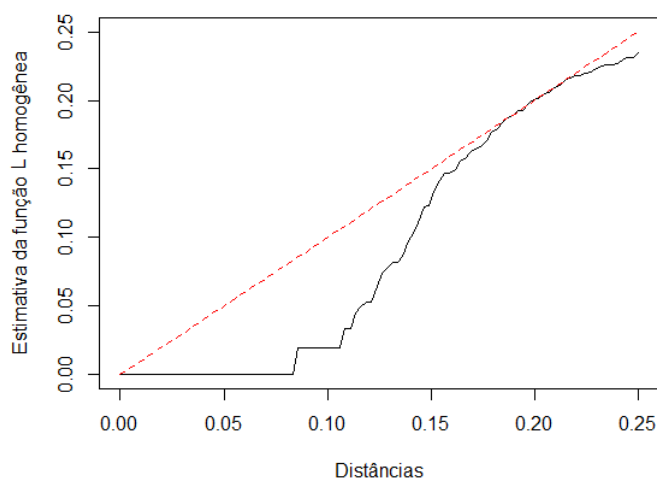


Figura 9 Funções L estimada (linha cheia) e teórica sob a hipótese de completa aleatoriedade espacial (linha tracejada) para a realização de um processo de Poisson homogêneo.

2.4.3 Função K para processos pontuais não homogêneos

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 a 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; MØLLER; WAAGEPETERSEN, 2000) como sendo:

$$\hat{K}_{inhom}(h) = \frac{1}{|A|} \sum_i^n \sum_{i \neq j}^n \frac{w_{ij}^{-1} I_h(d_{ij})}{\hat{\lambda}(x_i) \hat{\lambda}(x_j)}, h \geq 0, \quad (7)$$

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, Møller e Waagepetersen (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, igual a πh^2 .

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

$$\hat{L}_{inhom}(h) = \sqrt{\frac{\hat{K}_{inhom}(h)}{\pi}}. \quad (8)$$

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

2.5 Métodos para testar a hipótese de CAE

Nesta dissertação 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.

Os métodos apresentados nesta dissertação 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).

Para testar a hipótese nula de que a configuração pontual é uma realização de um processo de Poisson homogêneo pode-se utilizar as funções L homogênea e não homogênea. Essas funções podem ser utilizadas de duas formas para conduzir esse teste: graficamente e teste formal. Nos dois casos utiliza-se o método de Monte Carlo.

2.6 Método gráfico para testar a hipótese de CAE

Segundo Diggle (2003) é possível utilizar o estimador da função L homogênea (ou L não homogênea) como método para comparar estatisticamente a configuração pontual dos eventos observados com o que se esperaria sob a hipótese de CAE. Para tal, são construídos envelopes de simulação sob a suposição de CAE. Os envelopes de simulação superiores e inferiores são construídos a partir de s simulações independentes sob a hipótese de completa aleatoriedade espacial (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 do gráfico 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. Para Câmara et al. (2002), os resultados dos gráficos podem ser usados como métodos exploratórios para verificar evidências de interação entre os pontos.

Na Figura 10 tem-se uma análise gráfica com a configuração pontual descrito na Figura 1(B). Pode-se observar que a estimativa da função L fica abaixo do limite inferior do envelope de simulação e, portanto, rejeitando a hipótese nula de CAE na direção de regularidade.

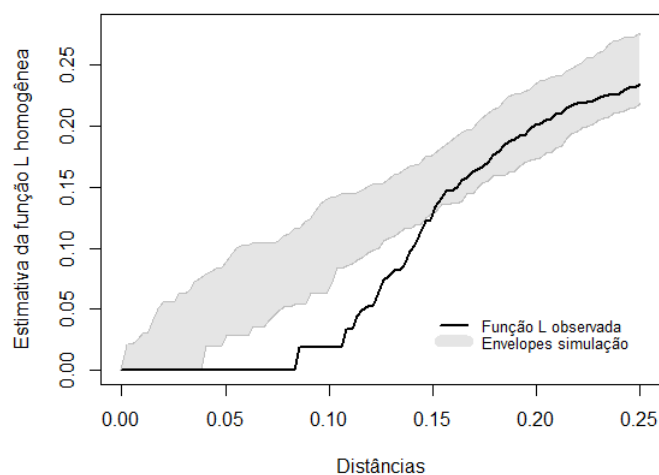


Figura 10 Envelopes de simulação seguindo o Processo de Poisson Homogêneo (área cinza) com a estimativa da função L homogênea (linha cheia) da configuração da Figura 1 (B).

2.7 Teste formal contra a hipótese nula de CAE

Uma 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 discrepân-

cia: Kolmogorov-Smirnov e Cramer-von Mises. A estatística de teste utilizando a medida de discrepância de Kolmogorov-Smirnov é dada por

$$L_K(i) = \max_{0 \leq h \leq h_0} |\hat{L}(h) - L(h)|. \quad (9)$$

Pode-se observar que a estatística de Kolmogorov-Smirnov mede a diferença máxima entre os valores observados e os valores esperados sob a hipótese nula de CAE dentro do intervalo de distância que vai de 0 a h_0 . Neste caso, observe que os valores esperados de $L(h)$ é o próprio vetor de distâncias (0 a h_0).

A estatística de teste não apresenta distribuição amostral conhecida e, portanto, Diggle (2003) sugere utilizar o Método de Monte Carlo para obter a significância estatística. Primeiramente, calcula-se $L_K(1)$ para a configuração observada. A seguir, as estatísticas são obtidas para $i = 2, \dots, s$, realizações sob a suposição de completa aleatoriedade espacial. Se ao ordenar $L_K(i)$, obtiver $L_K(1)$ em um dos extremos da distribuição, há evidências estatísticas para se rejeitar a hipótese nula de completa aleatoriedade espacial. Mais formalmente, se $L_K(1)$ é o k -ésimo maior valor, então se rejeita a hipótese nula de CAE caso o valor- p empírico $P = (s + 1 - k)/s$ seja inferior ou igual a um nível de significância α pré especificado.

2.8 Poder do teste e erros do tipo I e II

Se a hipótese H_0 for verdadeira e não rejeitada ou falsa e rejeitada a decisão estará correta. Entretanto, ao conduzir um teste de hipótese, o pesquisador pode cometer dois tipos de erro: o tipo I e II (MOOD; GRAYBILL; BOES, 1974).

O erro do tipo I ocorre ao rejeitar-se a hipótese H_0 quando ela é verdadeira. A probabilidade de se cometer esse tipo de erro pode ser controlada pelo pesquisador e é definida por.

$$P[\text{erro tipo I}] = P[\text{rejeitar } H_0 \text{ dado } H_0 \text{ verdadeira}] = \alpha.$$

O erro do tipo II ocorre ao não rejeitar a hipótese H_0 quando ela é falsa. Se a hipótese H_0 for rejeitada sendo verdadeira ou não rejeitada sendo falsa, a decisão estará errada. A probabilidade de se cometer esse tipo de erro não pode ser diretamente controlável pelo pesquisador e é definida por.

$P[\text{erro tipo II}] = P[\text{não rejeitar } H_0 \text{ dado } H_0 \text{ falsa}] = \beta.$

Ao conduzir um teste de hipótese é de interesse avaliar o poder do teste definido como sendo a probabilidade de se rejeitar a hipótese nula quando ela é falsa, ou seja, $1 - \beta$. Esse valor depende dos parâmetros que estão sendo testados, em que β é a probabilidade de se cometer o erro do tipo II (MOOD; GRAYBILL; BOES, 1974).

As características probabilísticas do poder de um teste são descritas em forma de uma função que é associada a cada valor de θ a $\pi(\Theta)$ de rejeitar H_0 . $\pi(\Theta)$ é a função de poder do teste e RC é a região crítica para testar H_0 contra H_1 é definida por:

$$\pi(\Theta) = P[X \in RC/\Theta], \forall \theta \in \Theta$$

A função poder de um teste é utilizada para verificar se o teste é adequado ou para comparar dois ou mais testes. Assim, uma função de poder seria $\pi(\Theta) = 0$ em que θ satisfaz H_0 e $\pi(\Theta) = 1$ em que θ satisfaz H_1 .

3 MATERIAIS E MÉTODOS

3.1 Simulação das configurações

No trabalho de 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. 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, primeiramente, 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, 2015).

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, 2015).

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, 2015).

3.2 Testes de hipóteses

Os testes contra a hipótese nula de completa aleatoriedade espacial descritos nas seções (2.4.2 e 2.4.3) são aplicados para cada uma das realizações simuladas. Para obter a estatística de teste de Kolmogorov-Smirnov descrita na seção (2.7) considera-se uma distância máxima $h_0 = 0,25$, seguindo o que re-

comenda 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, Møller e Waagepetersen (2000).

3.3 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.

3.4 Aplicação

A base de dados utilizados neste trabalho 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, 2015).

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 descrito na seção (2.4.1). Para tal, utilizou-se uma função kernel gaussiana com raio de influência τ igual a 3 é 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.

3.5 Software

Para conduzir todas as análises foram construídas funções (ver anexo) utilizando a biblioteca *spatstat* (BADDELEY e TURNER, 2005) disponível no *software R* (R CORE TEAM, 2015).

4 RESULTADOS E DISCUSSÃO

4.1 Poder dos testes em configurações homogêneas com interação para regularidade e agrupamentos

Uma configuração espacial de pontos deve ser compreendido como uma realização de um processo estocástico pontual espacial conforme pode ser visto em detalhes em Bailey e Gatrell (1995), Cressie (1993) e Diggle (2003). Na análise de processos pontuais espaciais tem-se interesse em verificar se a configuração espacial observada foi gerada por um processo estocástico espacial que exhibe 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).

Dentre as diversas funções propostas para analisar se a configuração espacial observada apresenta interação entre os pontos, a função K , proposta por Ripley (1977) e sua forma linearizada, a função L , proposta por Besag e Diggle (1977) são as mais conhecidas. A aplicação das funções K e L assumem 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. 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, Møller e Waagepetersen (2000) é uma das mais utilizadas. Essas funções podem ser utilizadas em métodos gráficos e testes formais de hipótese

A função L não homogênea têm sido utilizada com frequência na análise de configurações pontuais em que exista suspeita de não ser observada a homo-

geneidade 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 na Tabela 1 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 11 a 14 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

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	σ	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

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.

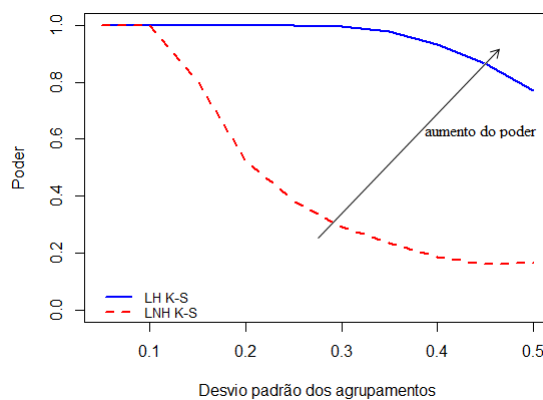


Figura 11 Poder dos testes de Kolmogorov-Smirnov em configurações com dois agrupamentos, em que LH K-S é o teste baseado na função L homogênea e LNH K-S é o teste baseado na função L não homogênea.

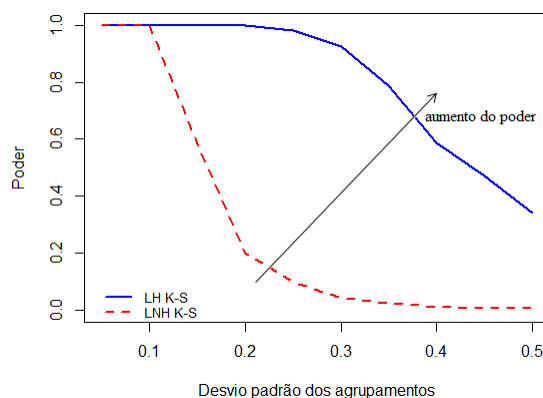


Figura 12 Poder dos testes de Kolmogorov-Smirnov em configurações com cinco agrupamentos, em que LH K-S é o teste baseado na função L homogênea e LNH K-S é o teste baseado na função L não homogênea.

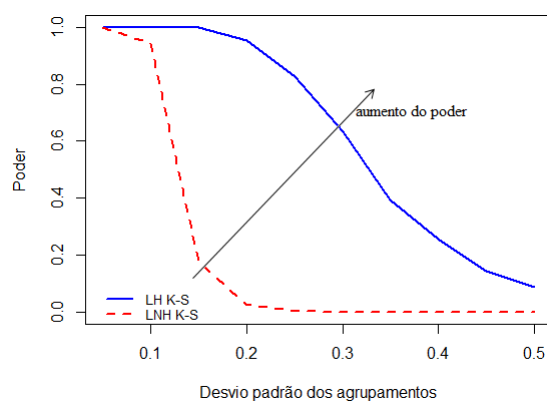


Figura 13 Poder dos testes de Kolmogorov-Smirnov em configurações com dez agrupamentos, em que LH K-S é o teste baseado na função L homogênea e LNH K-S é o teste baseado na função L não homogênea.

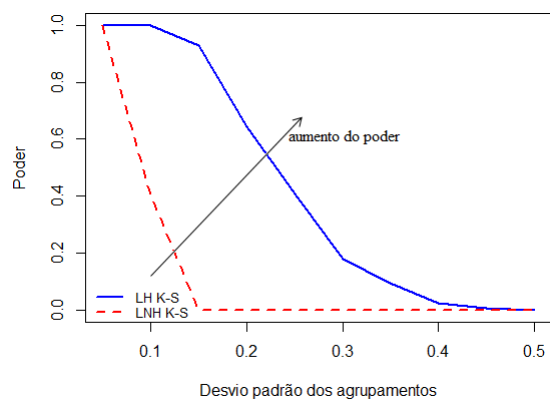


Figura 14 Poder dos testes de Kolmogorov-Smirnov em configurações com vinte agrupamentos, em que LH K-S é o teste baseado na função L homogênea e LNH K-S é o teste baseado na função L não homogênea.

Os resultados obtidos neste estudo corroboram por Baddeley, Møller e Waagepetersen (2000). 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.

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 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 15 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.

Observa-se também na Figura 15 que as duas funções conduzem a testes que apresentam praticamente o mesmo poder para qualquer distância mínima en-

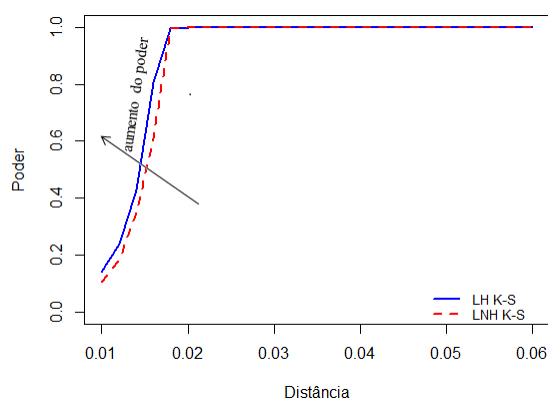


Figura 15 Poder dos testes de Kolmogorov-Smirnov em configurações com regularidade, em que LH K-S é o teste baseado na função L homogênea e LNH K-S é o teste baseado na função L não homogênea.

tre 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.

4.2 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.

A Figura 16 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 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

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

da completa aleatoriedade espacial.

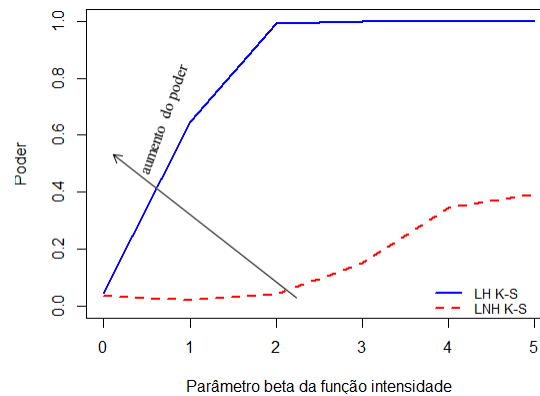


Figura 16 Poder dos testes de Kolmogorov-Smirnov em configurações com tendência espacial (não homogeneidade), em que LH K-S é o teste baseado na função L homogênea e LNH K-S é o teste baseado na função L não homogênea.

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 iden-

tifique 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.

4.3 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-OLBERSLEBEN 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 17). 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 x 7 mm, com área de 126 mm^2 e identifica a posição e o raio de todas as partículas no plano conforme mostra à Figura 17 (MEYER-OLBERSLEBEN e KIEBACK, 1997).

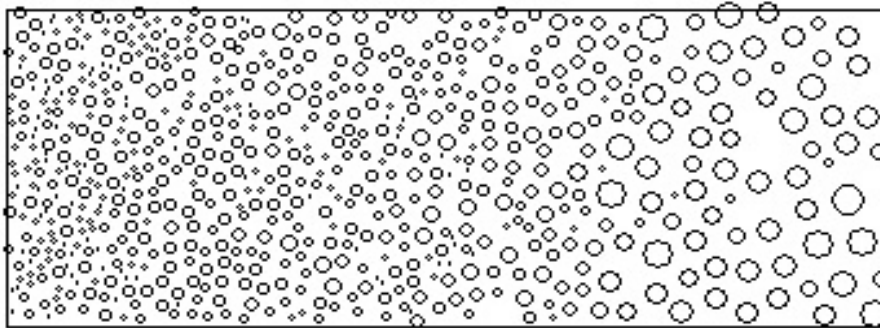


Figura 17 Distribuição das 678 partículas de bronze em uma área plana de 126 mm^2 . Os tamanhos da partículas são proporcional ao seu raio

A distribuição espacial dos centros das 678 partículas no plano é apresentada na Figura 18. A Figura 18 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 (descrito na seção 1.4.1). 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 19.

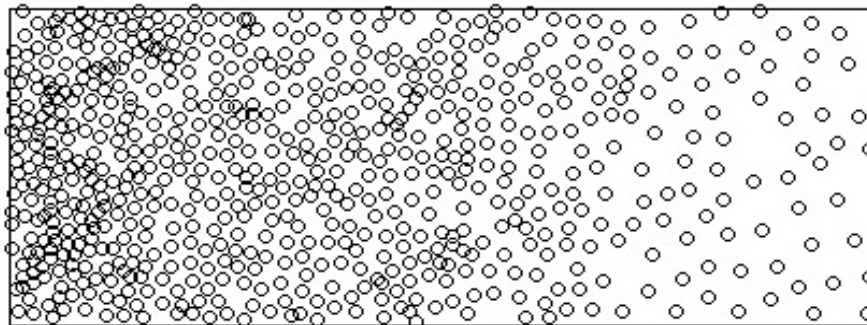


Figura 18 Configuração pontual dos centros das 678 partículas de bronze em uma área de 126 mm^2 .

Pode-se observar na Figura 19 que a intensidade estimada está entre uma

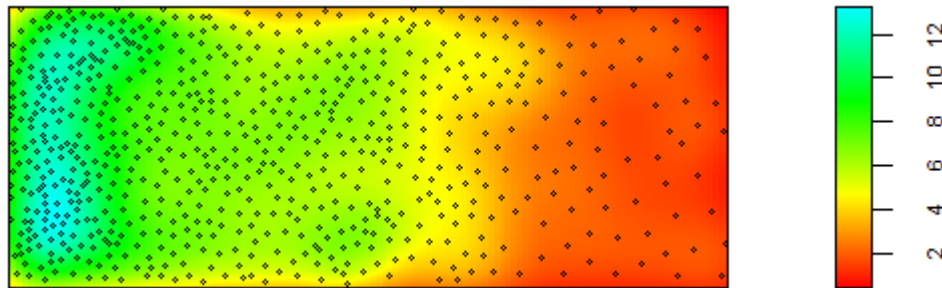


Figura 19 Estimativas kernel da intensidade das partículas de bronze no área de 126 mm^2 .

e catorze partículas por 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.

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 20.

Na Figura 20 observa-se que existe uma distância mínima entre as partícu-

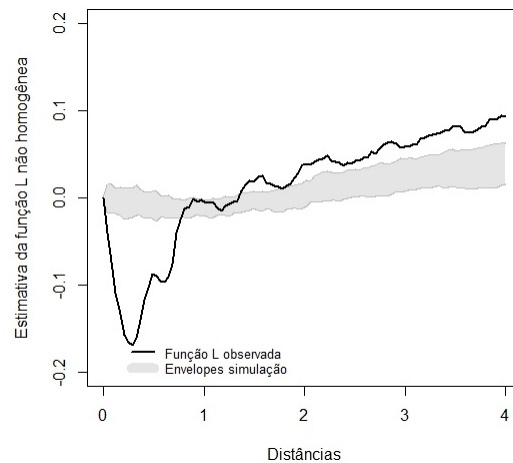


Figura 20 Estimativas da função L não homogênea (L_{inhom}), representada pela linha cheia e envelopes de simulação sob a hipótese de completa aleatoriedade espacial com 99% de confiança (área em cinza).

las 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).

5 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

- 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.
- 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.
- 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.
- BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. New York: Longman Scientific e Technical, p. 413, 1995.
- 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.
- BERNHARDT, R.; MEYER-OLBERSLEBEN, F.; KIEBACK, B. **Fundamental investigation on the preparation of gradient structures by sedimentation of different powder fractions under gravity**. Proc. of the 4th Int. Conf. On Composite Engineering, July 6-12, Hawaii, Ed. David Hui, p. 147-148, 1997.
- BIVAND, R. S.; PEBESMA, E. J.; GÓMES-RUBIO, V. **Applied Spatial Data Analysis with R**. New York: Springer, p. 374, 2008.
- CÂMARA, G.; MONTEIRO, A. M.; DRUCK, S.; CARVALHO, M. S.

Introdução à ciência da geoinformação. São José dos Campos: INPE, p. 186, 2002.

CÂMARA, G.; DRUCK, S.; CARVALHO, M. S.; MONTEIRO, A. V. M.
Análise Espacial de Dados Geográficos. Planaltina: EMBRAPA, p. 208, 2004.

CRESSIE, N. A. C. **Statistics for spatial data.** New York: John Wiley e Sons, p. 900, 1993.

DIGGLE, P.J. On parameters estimation and goodnessofit testing for spatial point patterns. **Biometrics**, Washington, v. 35, n. 1, p. 87-101, 1979.

DIGGLE, P. J. **Statistical analysis of spatial point patterns.** 2. ed. London: Hodder Arnold Publishers, p. 159, 2003.

FRADE, D. D. R. **Relações entre fatores ambientais e espécies florestais por metodologias de processos pontuais.** Dissertação de mestrado em Estatística e Experimentação Agronômica. Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz". Piracicaba-SP, 2014.

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.

ILLIAN, J.; PENTTINEN, A.; STOYAN, H.; STOYAN, D. **Statistical analysis and modelling of spatial point patterns.** New York: John Wiley & Sons, p. 534, 2008.

MATERN, B. **Spatial Variation.** Lecture Notes in Statistics, Springer-Verlag, New York, v. 36, 1986.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. Singapore: McGraw-Hill, p. 564, 1974.

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.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria, 2015. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em: 11 maio 2015.

RIPLEY, B. D. The Second-Order Analysis of Stationary Point Processes. **Journal of Applied Probability**, Sheffield, v. 13, n. 2, p. 255-266, 1976.

RIPLEY, B. D. Modelling Spatial Patterns. **Journal of the Royal Statistical Society. Series B (Methodological)**, Hoboken, v. 39, n. 2, p. 172- 212, 1977.

RIPLEY, B. D. **Spatial Statistical**. New York: John Wiley e Sons, p. 183, 1981.

WAAGEPETERSEN, R. P. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. **Biometrics**, v. 63, n. 1, p. 252-258, 2007.

ANEXO

PROGRAMA 1- Função `Lest.test.cluster`

Esta função é utilizada dentro da função `power.Lest.cluster`.

Esta função implementa as simulações de Monte Carlo para testar um padrão contra a hipótese de CAE em uma configuração homogênea com agrupamentos, usando a função `L` e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
Lest.test.cluster<- function(event, nsim = 100){
# A partir dos valores
fd<- 0.25*max(pairdist(event))
rr<- seq(0, fd, length.out = 100)
theo<- pi*rr^2; np<- length(event$x)
lent<- length(rr); h<- fd-rr[1]
W<- as.owin(event); densi <- np/area.owin(W)
# Obter a função L observada
go<- Lest(event, r=rr, correction="iso")
# Obtenhe a est. de Kolmogorov para o padrão observado
D<- max(abs(go$iso - theo))
# Obter a estatística de padrão simulado sob a CAE
for(isim in (2:nsim)) {
  spts<- runifpoint(np, win=W)
  g<- Lest(spts, r=rr, correction="iso")
  #It uses default for sigma (nrd0)
  testK<- max(abs(g$iso - theo));D <- c(D, testK)
# Obter a estatística e o valor-p
pk <- rank(D)[1]; pvaluek <- (nsim + 1 - pk)/nsim
# Obter os resultados
list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 2- Função `power.Lest.cluster`

Esta função implementa simulações de Monte Carlo para obter o poder de um teste contra a hipótese de CAE de um processo de Poisson homogêneo de ponto espacial simulado em uma configuração com agrupamento. Padrões de pontos tem área igual a um com 100 pontos. Usando a função L e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

pv: número de simulação para obter o poder

p: Ponto de corte (nível de significância)

```
Power.Lest.cluster<- function(p, pv){
# Parâmetros de ajuste
clusters<-rep(c(2,5,10,20), each=10)
# número de agrupamento
sigma<-rep(c(0.05,0.10,0.15,0.20,0.25,0.30,0.35,
0.40,0.45,0.50),times=4)# variabilidade radial
pontos<-rep(c(100),each=40)
eventos<-cbind(clusters, sigma, pontos)
# Simulação de agrupamento
contK<-rep(0,nrow(eventos))
# Poder
for (i in 1:nrow(eventos)){
  for (j in 1:pv){
    ze<-rMatClust(eventos[i,1],eventos[i,2],
eventos[i,3])
    a<-Lest.test.cluster(event=ze)
    ptesK<-a$Pvalue_Kolmo
    if(ptesK<p) contK[i]<-contK[i] + 1 }}
result<-cbind(eventos, contK)
list(resultados = result)}
```


PROGRAMA 3- Função `Linhom.test.cluster`

Esta função é utilizada dentro da função `power.Linhom.cluster`.

Esta função implementa as simulações de Monte Carlo para testar um padrão contra a hipótese de CAE em uma configuração homogênea com agrupamento, usando a função L não homogênea e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatstat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
Linhom.test.cluster<- function(event, nsim = 100){
# A partir dos valores
fd <- 0.25*max(pairdist(event))
rr <- seq(0, fd, length.out = 100)
theo <- pi*rr^2; np<- length(event$x)
lent <- length(rr); h<- fd-rr[1]
W <- as.owin(event); densi<- np/area.owin(W)
# Obter a função $L_{inhom}$ observada
go <- Linhom(event, r=rr, correction="iso")
# It uses default for sigma (nrd0)
# Obtenhe a estatística de Kolmogorov
# para o padrão observado
D <- max(abs(go$iso - theo))
# Obter a estatística de padrão simulado sob a CAE
for(isim in (2:nsim)) {
  spts<- runifpoint(np, win=W)
  g<- Linhom(spts, r=rr, correction="iso")
  #It uses default for sigma (nrd0)
  testK<- max(abs(g$iso - theo)); D<- c(D, testK)}
# Obter a estatística e o valor-p
pk<- rank(D)[1]; pvaluek<- (nsim + 1 - pk)/nsim
# Obter os resultados
list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 4- Função `power.Linhom.cluster`

Esta função implementa simulações de Monte Carlo para obter o poder de um teste contra a hipótese de CAE de um processo de Poisson homogêneo de ponto espacial simulado em uma configuração com agrupamento. Padrões de pontos tem área igual a um com 100 pontos. Usando a função L não homogênea e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

pv: número de simulação para obter o poder

p: Ponto de corte (nível de significância)

```
Power.Linhom.cluster<- function(p, pv){
# Parâmetros de ajuste
clusters<-rep(c(2,5,10,20), each=10)
# números de agrupamentos
sigma<-rep(c(0.05,0.10,0.15,0.20,0.25,0.30,0.35,
0.40,0.45,0.50),times=4)# variabilidade radial
pontos<-rep(c(100),each=40)
eventos<-cbind(clusters, sigma, pontos)
# Simulação de agrupamento
contK<-rep(0,nrow(eventos))
# Poder
for (i in 1:nrow(eventos)){
  for (j in 1:pv){
    ze<-rMatClust(eventos[i,1],eventos[i,2],
eventos[i,3])
    a<-Linhom.test.cluster(event=ze)
    ptesK<-a$Pvalue_Kolmo
    if(ptesK<p) contK[i]<-contK[i] + 1 }}
result<-cbind(eventos, contK)
list(resultados = result)}
```

PROGRAMA 5- Função `Lest.test.SSI`

Esta função é utilizada dentro da função `power.Lest.SSI`.

Esta função implementa as simulações de Monte Carlo para testar um padrão contra a hipótese de CAE em uma configuração homogênea com regularidade, usando a função `L` e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
Lest.test.SSI<- function(event, nsim = 100){
# A partir dos valores
fd<- 0.25*max(pairdist(event))
rr<- seq(0, fd, length.out = 100)
theo<- pi*rr^2; np<- length(event$x)
lent<- length(rr); h<- fd-rr[1]
W<- as.owin(event); densi<- np/area.owin(W)
# Obter a função L observada
go<- Lest(event, r=rr, correction="iso")
# Obtenha a est. de Kolmogorov para o padrão obs.
D<- max(abs(go$iso - theo))
# Obter a estatística de padrão simulado sob a CAE
for(isim in (2:nsim)) {
  spts<- runifpoint(np, win=W)
  g<- Lest(spts, r=rr, correction="iso")
  #It uses default for sigma (nrd0)
  testK<- max(abs(g$iso - theo))
  D<- c(D, testK) }
# Obter a estatística e o valor-p
pk<- rank(D)[1]; pvaluek<- (nsim + 1 - pk)/nsim
# Obter os resultados
list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 6- Função `power.Lest.SSI`

Esta função implementa simulações de Monte Carlo para obter o poder de um teste contra a hipótese de CAE de um processo de Poisson homogêneo de ponto espacial simulado em uma configuração com regularidade. Padrões de pontos tem área igual a um com 100 pontos. Usando a função L e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatstat**

Os argumentos desta função são:

pv: número de simulação para obter o poder

p: Ponto de corte (nível de significância)

```
Power.Lest.SSI<- function(p, pv) {
  # Parâmetros de ajuste
  dist<-c(0.01, 0.012, 0.014, 0.016, 0.018, 0.020, 0.030,
  0.040,0.050,0.060)# Distância mínima entre os pontos
  pontos<-rep(c(100),each=10)
  eventos<-cbind(dist, pontos)
  contK<-rep(0,nrow(eventos))
  # Poder
  for(i in 1:nrow(eventos)){
    for(j in 1:pv){
      X<-rSSI(r=eventos[i,1],n=eventos[i,2])
      a<-Lest.test.SSI(event=X)
      ptesK<-a$Pvalue_Kolmo
      if(ptesK<p) contK[i]<-contK[i] +1
    }
  }
  result<-cbind(eventos, contK)
  list(resultados = result)
}
```

PROGRAMA 7- Função `Linhom.test.SSI`

Esta função é utilizada dentro da função `power.Linhom.SSI`.

Esta função implementa as simulações de Monte Carlo para testar um padrão contra a hipótese de CAE em uma configuração com regularidade, usando a função L não homogênea e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
Linhom.test.SSI<- function(event, nsim = 100){
#   A partir de valores
fd<- 0.25*max(pairdist(event))
rr<- seq(0, fd, length.out = 100)
theo<- pi*rr^2; np<- length(event$x)
lent<- length(rr); h<- fd-rr[1]
W<- as.owin(event); densi<- np/area.owin(W)
#   Obter a função  $L_{\{inhom\}}$  observada
go<- Linhom(event, r=rr, correction="iso")
#It uses default for sigma (nrd0)
#   Obtenhe a est. de Kolmogorov para o padrão obs.
D<- max(abs(go$iso - theo))
#   Obter a estatística de padrão simulado sob a CAE
for(isim in (2:nsim)) {
  spts<- runifpoint(np, win=W)
  g<- Linhom(spts, r=rr, correction="iso")
  #It uses default for sigma (nrd0)
  testK<- max(abs(g$iso - theo))
  D<- c(D, testK)}
#   Obter a estatística e o valor-p
pk<- rank(D)[1]; pvaluek<- (nsim + 1 - pk)/nsim
#   Obter os resultados
list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 8- Função `power.Linhom.SSI`

Esta função implementa simulações de Monte Carlo para obter o poder de um teste contra a hipótese de CAE de um processo de Poisson homogêneo de ponto espacial simulado em uma configuração com agrupamento. Padrões de pontos tem área igual a um com 100 pontos. Usando a função L não homogênea e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

pv: número de simulação para obter o poder

p: Ponto de corte (nível de significância)

```
Power.Linhom.SSI<- function(p, pv){
# Parâmetro de ajuste
dist<-c(0.01, 0.012, 0.014, 0.016,
0.018, 0.020, 0.030, 0.040,
0.050,0.060)# Distância mínima entre os pontos
pontos<-rep(c(100),each=10)
eventos<-cbind(dist, pontos)
contK<-rep(0,nrow(eventos))
# Poder
for(i in 1:nrow(eventos)){
for(j in 1:pv){
X<-rSSI(r=eventos[i,1],n=eventos[i,2])
a<-Linhom.test.SSI(event=X)
ptesK<-a$Pvalue_Kolmo
if(ptesK<p) contK[i]<-contK[i] +1}}
result<-cbind(eventos, contK)
list(resultados = result)}
```

PROGRAMA 9- Função `Lest.test`

Esta função é utilizada dentro da função `power.Lest`.

Esta função implementa as simulações de Monte Carlo para testar um padrão contra a hipótese de CAE em uma configuração não homogênea com tendência, usando a função `L` e a medida de discrepância máxima de Kolmogorov-Smirnov.

Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
Lest.test<- function(event, nsim = 100){
# A partir dos valores
fd<- 0.25*max(pairdist(event))
rr<- seq(0, fd, length.out = 100)
theo<- pi*rr^2; np<- length(event$x)
lent<- length(rr); h<- fd-rr[1]
W<- as.owin(event); densi<- np/area.owin(W)
# Obter a função L observada
go<- Lest(event, r=rr, correction="iso")
# Obtenha a est. de Kolmogorov para o padrão obs.
D<- max(abs(go$iso - theo))
# Obter a estatística de padrão simulado sob a CAE
for(isim in (2:nsim)) {
  spts<- runifpoint(np, win=W)
  g<- Lest(spts, r=rr, correction="iso")
  #It uses default for sigma (nrd0)
  testK<- max(abs(g$iso - theo))
  D<- c(D, testK)}
# Obter a estatística e o valor-p
pk<- rank(D)[1]
pvaluek<- (nsim + 1 - pk)/nsim
# Obter os resultados
list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 10- Programa `power.Lest`

Esta função implementa simulações de Monte Carlo para obter o poder de um teste contra a hipótese de CAE de um processo de Poisson não homogêneo de ponto espacial simulado. Padrões de pontos tem área igual a um com intensidade igual a $\lambda = \text{function}(x,y)100*\exp(\beta*x)$, aproximadamente 100 pontos. Usando a função L e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

pv: número de simulação para obter o poder

p: Ponto de corte (nível de significância)

```
Power.Lest<- function(p, pv){
# definindo o parâmetro beta da função intensidade
beta<-c(0,1,2,3,4,5)
lb = length(beta)
contK<-rep(0, lb)
# Poder
for(i in 1:lb){
  cat("Doing simulation for intensity
parameter equal to", beta[i], "\n")
  for(j in 1:pv){
    ze = function(x,y){300*exp(-beta[i]*x)}
    X = rpoispp(ze)
    a = Linho.test(event=X)
    ptesK = a$Pvalue_Kolmo
    if(ptesK < p) contK[i]<- contK[i] +1}}
result<-cbind(beta, contK)
list(resultados = result)
}
```


PROGRAMA 11- Função `Linhom.test`

Esta função é utilizada dentro da função `power.Linhom`.

Esta função implementa as simulações de Monte Carlo para testar um padrão contra a hipótese de CAE em uma configuração não homogênea com tendência, usando a função `L` não homogênea e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatstat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
Linhom.test<- function(event, nsim = 100){
#   A partir de valores
fd<- 0.25*max(pairdist(event))
rr<- seq(0, fd, length.out = 100)
theo<- pi*rr^2;np<- length(event$x)
lent<- length(rr);h<- fd-rr[1]
W<- as.owin(event);densi<- np/area.owin(W)
#   Obter a função $L_{inhom}$ observada
go<- Linhom(event, r=rr, correction="iso")
#It uses default for sigma (nrd0)
#   Obtenhe a est. de Kolmogorov para o padrão observado
D<- max(abs(go$iso - theo))
#   Obter a estatística de padrão simulado sob a CAE
for(isim in (2:nsim)) {
  spts<- runifpoint(np, win=W)
  g<- Linhom(spts, r=rr, correction="iso")
  #It uses default for sigma (nrd0)
  testK<- max(abs(g$iso - theo))
  D<- c(D, testK)}
#   Obter a estatística e o valor-p
pk<- rank(D)[1];pvaluek<- (nsim + 1 - pk)/nsim
#   Obter os resultados
list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 12- Função `power.Linhom`

Esta função implementa simulações de Monte Carlo para obter o poder de um teste contra a hipótese de CAE de um processo de Poisson não homogêneo de ponto espacial simulado. Padrões de pontos tem área unitária com intensidade igual a $\lambda = \text{function}(x,y)100 \cdot \exp(\beta \cdot x)$, aproximadamente 100 pontos. Usando a função L não homogênea e a medida de discrepância máxima de Kolmogorov-Smirnov. Esta função necessita carregar o pacote **spatistat**

Os argumentos desta função são:

pv: número de simulação para obter o poder

p: Ponto de corte (nível de significância)

```
Power.Linhom<- function(p, pv){
# definindo o parâmetro beta da função intensidade
beta<-c(0,1,2,3,4,5)
lb = length(beta)
contK<-rep(0, lb)
# Poder
for(i in 1:lb){
  cat("Doing simulation for intensity
parameter equal to",
beta[i], "\n")
  for(j in 1:pv){
ze = function(x,y){300*exp(-beta[i]*x)}
X = rpoispp(ze)
a = Linho.test(event=X)
ptesK = a$Pvalue_Kolmo
if(ptesK < p) contK[i]<- contK[i] +1}}
result<-cbind(beta, contK)
list(resultados = result)}
```

PROGRAMA 13- Função `kolmogorov.test`

Esta função implementa simulações de Monte-Carlo para testar um padrão de pontos contra a hipótese de completa aleatoriedade espacial usando a função L homogênea e a medida de discrepância Kolmogorov-Smirnov com distância máxima. Esta função necessita carregar o pacote **spatstat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
kolmogorov.test=function(event=
unmark(bronzefilter), nsim = 100){
  # Obtendo valores
  fd=4
  rr<- seq(0, fd, length.out = 50)
  np<- length(event$x)
  W<- as.owin(event)
  # Obter a função L observada
  go<- Lest(event, r=rr, correction="iso")
  # Obter a estatística de Kolmogorov
  D<- max(abs(go$iso - go$theo))
  # Obter a estatística de padrão simulado contra CAE
  for(isim in (2:nsim)) {
    spts<- runifpoint(np, win=W)
    g<- Lest(spts, r=rr, correction="iso")
    testK<- max(abs(g$iso - g$theo))
    D<- c(D, testK)}
  # Obter estatística e valor-p
  pk<- rank(D)[1]; pvaluek<- (nsim + 1 - pk)/nsim
  # Obter resultados
  list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 14- Função `kolmogorov.test`

Esta função implementa simulações de Monte-Carlo para testar um padrão de pontos contra a hipótese de completa aleatoriedade espacial usando a função L não homogênea e a medida de discrepância Kolmogorov-Smirnov com distância máxima. Esta função necessita carregar o pacote **spatstat**

Os argumentos desta função são:

`event`: é o objeto de classe `ppp` que contém as coordenadas dos pontos

`nsim`: número de simulação Monte Carlo

```
kolmogorov.test=function(event=
unmark(bronzefilter), nsim = 100){
  # Obtendo valores
  fd=4
  rr<- seq(0, fd, length.out = 50)
  np<- length(event$x)
  W<- as.owin(event)
  # Obter a função L não homogênea observada
  go<- Linhom(event, r=rr, correction="iso")
  # Obter a estatística de Kolmogorov
  D<- max(abs(go$iso - go$theo))
  # Obter a est. de padrão simulado contra CAE
  for(isim in (2:nsim)) {
    spts<- runifpoint(np, win=W)
    g<- Linhom(spts, r=rr, correction="iso")
    testK<- max(abs(g$iso - g$theo))
    D<- c(D, testK)}
  # Obter estatística e valor-p
  pk<- rank(D)[1]
  pvaluek<- (nsim + 1 - pk)/nsim
  # Obter resultados
  list(Komogorov=D[1], Pvalue_Kolmo = pvaluek)}
```

PROGRAMA 15- Função `kest.env`

Esta função implementa simulações de Monte-Carlo e cria envelopes de simulação para testar a hipótese de completa aleatoriedade espacial usando a função L homogênea.

```

kest.env<-function(dados, simul=99){
  y<-seq(-0.2,0.5, length.out=100)
  r<-seq(0, 7,length.out=100)
  E=envelope(dados, Kest, r=r, nsim=simul)
  plot(E$r, y, type = 'n', main="",
  xlab="Distâncias",
  ylab="Estimativa da função L homogênea")
  lines(r, sqrt(E$lo/pi) - r, lty="solid",
  col="grey", lwd=2)
  lines(r, sqrt(E$hi/pi) - r, lty="solid",
  col="grey", lwd=2)
  polygon(c(r, rev(r)), c(sqrt(E$hi/pi)-r,
  rev(sqrt(E$lo/pi)-r)),
  col = "grey90", border = NA)
  lines(r, sqrt(E$obs/pi)-r, lty="solid",
  col="black", lwd=2)
  legend(max(E$r)/25, -0.5,
  c("Função L observada","Envelopes simulação"),
  lty=c("solid","solid"), lwd=c(2,10),
  col=c("black","grey90"), bty="n", cex=0.8)}

```

PROGRAMA 16- Função kinhom.env

Esta função implementa simulações de Monte-Carlo e cria envelopes de simulação para testar a hipótese de completa aleatoriedade espacial usando a função L não homogênea.

```
kinhom.env<-function(dados, simul=99){
  y<-seq(-0.2,0.7, length.out=100)
  r<-seq(0, 7,length.out=100)
  E=envelope(dados, Kinhom, r=r, nsim=simul)
  plot(E$r, y, type = 'n', main="", x
  lab="Distâncias",
  ylab="Estimativa da função L não homogênea")
  lines(r, sqrt(E$lo/pi) - r, lty="solid",
  col="grey", lwd=2)
  lines(r, sqrt(E$hi/pi) - r, lty="solid",
  col="grey", lwd=2)
  polygon(c(r, rev(r)), c(sqrt(E$hi/pi)-r,
  rev(sqrt(E$lo/pi)-r)),
  col = "grey90", border = NA)
  lines(r, sqrt(E$obs/pi)-r, lty="solid",
  col="black", lwd=2)
  legend(max(E$r)/25, -0.5,
  c("Função L não homogênea observada",
  "Envelopes simulação"),
  lty=c("solid","solid"), lwd=c(2,10),
  col=c("black","grey90"), bty="n", cex=0.8)}
```