

**MÉTODOS PARA ANÁLISE DE INDEPENDÊNCIA
ENTRE MARCAS E PONTOS EM PROCESSOS
PONTUAIS MARCADOS**

RICARDO ALVES DE OLINDA

2008

RICARDO ALVES DE OLINDA

**MÉTODOS PARA ANÁLISE DE INDEPENDÊNCIA ENTRE MARCAS E
PONTOS EM PROCESSOS PONTUAIS MARCADOS**

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

Orientador
Prof. Dr. João Domingos Scalon

LAVRAS
MINAS GERAIS -BRASIL
2008

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Olinda, Ricardo Alves de.

Métodos para análise de independência entre marcas e pontos em processos pontuais marcados / Ricardo Alves de Olinda. – Lavras : UFLA, 2008.

76 p. : il.

Dissertação (Mestrado) – Universidade Federal de Lavras, 2008.

Orientador: João Domingos Scalon.

Bibliografia.

1. Dependência. 2. CSR. 3. Compósito. 4. Marcadores. I. Universidade Federal de Lavras. II. Título.

CDD – 519.282

RICARDO ALVES DE OLINDA

**MÉTODOS PARA ANÁLISE DE INDEPENDÊNCIA ENTRE MARCAS E
PONTOS EM PROCESSOS PONTUAIS MARCADOS**

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

APROVADA em 23 de dezembro de 2008.

Prof. Dr. Ednaldo Carvalho Guimarães UFU

Prof. Dr. Marcelo Silva de Oliveira UFLA

Prof. Dr. Renato Ribeiro de Lima UFLA

Prof. Dr. João Domingos Scalon
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

Dedico esta vitória:

A DEUS

*força maior de todo o ser humano, por ter
me dado tudo que eu sempre precisei.*

*Aos meus pais, Roberto e Marizete,
aos quais tenho muito respeito.*

À minha esposa, Andreza.

À meu filho, Gabriel.

*”Se permanecerdes em mim, e as minhas
palavras permanecerem em vós, pedireis
o que quiserdes, e vos será feito.”*

(João; 15:7)

AGRADECIMENTOS

A DEUS nosso pai e criador, por ter me dado tudo que eu sempre precisei.

Ao professor João Domingos Scalon, pelo apoio, paciência em me orientar no desenvolvimento deste trabalho.

Aos funcionários do Departamento de Ciências Exatas: Josi, Edila, Joyce, Maria, Selminha e Vânia, pela simpatia e boa vontade no atendimento.

A todos os Professores da Pós-Graduação em Estatística e Experimentação Agropecuária da UFLA, especialmente aos Professores Lucas Monteiro, Daniel Furtado, Augusto Ramalho, Renato Ribeiro, Devanil e Júlio.

Aos professores João Gil de Luna e Juarez Fernandes do Centro de Ciências Tecnológicas (CCT) da UEPB, pela confiança, orientação durante a graduação e pelo incentivo à carreira acadêmica.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, pela oportunidade da realização deste curso.

À CAPES, pela bolsa de estudos, essencial para a realização deste trabalho.

Aos três Mosqueteiros "um por todos e todos por um", Altemir, Paulo e Edcarlos, pela amizade e companheirismo durante este percurso de tempo.

Ao grupo de pesquisa em Estatística Espacial em nome do professor Marcelo Oliveira, Gerson, João Marcos e Naje, pelas proveitosas sugestões que tivemos em vários encontros.

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

Aos amigos de turma, Patrícia, Augusto, Tânia, Ana Paula, Denise, Richardson, Isabel, Stephânia, Iron, pela força, troca de conhecimento e amizade.

Ao meu amigo Eustáquio, a minha amiga Francisca e a Dalvinha, pelo apoio,

carinho, principalmente, pela sincera amizade durante todo este tempo.

Aos meus irmãos Ronnie e Roberta e toda a minha família (Tios, Tias, Primos e Primas), por todo apoio e incentivo.

Aos meus sobrinhos Carol, Rennan e Ana Livia por trazerem tanta alegria a nossa casa.

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	ii
RESUMO	iv
ABSTRACT	v
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Um pouco da história	3
2.2 Estatística espacial	4
2.3 Processos pontuais	5
2.3.1 Propriedades de primeira e segunda ordem	7
2.3.2 Processo estacionário e isotrópico	9
2.3.3 Completa aleatoriedade espacial (<i>CSR</i>)	11
2.4 Análise de processos pontuais	13
2.4.1 Função <i>F</i>	15
2.4.2 Função <i>G</i>	17
2.4.3 Função <i>J</i>	18
2.4.4 Função <i>K</i>	19
2.4.5 Teste de hipótese por meio da <i>CSR</i>	23
2.5 Análise de processos pontuais marcados	24
2.5.1 Função de correlação marcada	27
2.5.2 Estimando a função de correlação marcada	28
2.5.3 Teste de independência para função de correlação marcada	32
2.5.4 Processos pontuais marcados na literatura	32
3 MATERIAL E MÉTODOS	35

3.1 Materiais	35
3.2 Análise exploratória	36
3.3 Análise exploratória espacial	37
3.4 Função de correlação marcada	39
3.4.1 Teste de hipótese para independência entre marcas e pontos	39
3.4.2 Análise gráfica para testar independência entre marcas e pontos	40
3.5 Software	42
4 RESULTADOS E DISCUSSÃO	43
4.1 Análise do material compósito	43
4.1.1 Análise exploratória	43
4.1.2 Análise exploratória espacial	45
4.1.3 Função de correlação marcada	52
4.2 Análise dos marcadores genéticos	55
4.2.1 Análise exploratória espacial	55
4.2.2 Função de correlação marcada	58
4.2.3 Considerações finais	60
5 CONCLUSÃO	61
REFERÊNCIAS BIBLIOGRÁFICAS	62
APÊNDICE	67

LISTA DE TABELAS

1	Estatística descritiva da variável diâmetro.	44
---	--	----

LISTA DE FIGURAS

1	Três padrões de pontos simulados: aleatório (esquerda), regular (centro), agrupado (direita).	6
2	Estimador de intensidade pontual do processo em toda região de estudo (Kernel).	14
3	Região de estudo (pontos pretos) de área finita, fixada e conhecida, onde ocorreu um fenômeno de interesse.	16
4	Idéia gráfica da função K.	22
5	Idéia gráfica da função de correlação marcada.	31
6	Distribuição das 730 partículas de cerâmica na matriz de alumínio na amostra A; distribuição das 937 partículas de cerâmica na matriz de alumínio na amostra B; distribuição das 1234 partículas de cerâmica na matriz de alumínio na amostra C. Os círculos de cada partícula são proporcionais aos tamanhos das partículas. Área das amostras: $512 \times 767 \text{ pixels}^2$	44
7	Histogramas referente aos diâmetros das partículas por meio das amostras A, B e C.	44
8	Distribuição das partículas na base de alumínio por meio do suavizador de Kernel, em que a escala (lado direito) representa a intensidade de pontos por unidade de área em toda a região de estudo.	46
9	Envelope de simulação por meio da função G referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).	47

10	Envelope de simulação por meio da função F referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).	48
11	Envelope de simulação por meio da função J referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).	49
12	Envelope de simulação por meio da função K referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).	51
13	Envelope de simulação para os diâmetros das partículas referente as amostras: A, B e C, por meio da função de correlação marcada; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).	54
14	Distribuição das árvores na região de estudos por meio do suavizador Kernel, em que a escala (lado direito) representa a intensidade de pontos por unidade de área em toda a região de estudo.	55
15	Envelope de simulação das funções: F , G , J , K ; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).	57
16	Envelope de simulação da função de correlação marcada referente aos marcadores genéticos: $\alpha - Est1$, $\alpha - Est2$, $Po1$ e $Po2$; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).	59

RESUMO

Olinda, Ricardo Alves de. **Métodos para análise de independência entre marcas e pontos em processos pontuais marcados.** 2008. 76 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG. *

Muitos fenômenos são expressos por meio de ocorrências identificadas como pontos localizados no espaço e são denominados processos pontuais. Quando considera-se uma variável aleatória (marca) associada a cada ponto, tem-se um fenômeno denominado de processo pontual marcado. O objetivo deste trabalho é propor metodologias baseadas na função de correlação marcada e em métodos de Monte Carlo para testar a hipótese de independência entre a localização e a marca do evento. Os métodos propostos neste trabalho são aplicados em duas situações distintas: distribuição espacial de partículas de cerâmicas incrustadas em uma base de alumínio de um material compósito (marcas contínuas) e distribuição espacial de características genéticas em árvores (marcas categóricas). Primeiramente, explora-se a completa aleatoriedade espacial *CSR* das distribuições usando as conhecidas funções *G*, *F*, *J*, *K*. A seguir, os novos métodos desenvolvidos neste trabalho são usados para testar a hipótese de independência entre marcas e pontos. Os resultados mostram que a hipótese de *CSR* é rejeitada em favor da hipótese de regularidade para a distribuição espacial das partículas e em favor de agrupamento para a distribuição espacial das árvores. Os resultados também mostram evidências estatísticas para a rejeição da hipótese de independência entre a localização e o diâmetro das partículas e para a aceitação da hipótese de independência entre a localização e a característica genética das árvores. Conclui-se por meio dos resultados obtidos que as metodologias baseadas nos métodos de Monte Carlo e na função de correlação marcada podem ser poderosas ferramentas para a análise de processos pontuais marcados.

* **Orientador:** João Domingos Scalon - UFLA.

ABSTRACT

Olinda, Ricardo Alves de. **Method for analysis of independence between marks and points in marked point processes.** 2008. 76 p. Dissertation (Master of Statistics and Agricultural Experimentation) Federal University of Lavras, Lavras, MG. *

Many phenomena are stated through events identified as points located in space and they are named point processes. When a random variable (mark) is associated to each point, then the process is called marked point process. The aim of this work is to propose methodologies based on both the mark correlation function and Monte Carlo methods for testing the hypothesis of independence between the localization and the mark of the event. The considered methods are applied to two different situations: spatial distribution of particles held in a composite material (continuous marks) and spatial distribution of genetic characteristics of trees (categorical marks). Firstly, it is explored the complete spatial randomness *CSR* of the distributions using the well known *G*, *F*, *J*, and *K* functions. Then, the new methods developed in this work are used to test the hypothesis of independence between marks and points. The results show that the hypothesis of *CSR* is rejected in favor of regularity for the distribution of particles and in favor of aggregation for the distribution of trees. The results also show statistical evidences for both rejecting the hypothesis of independence between the localization and the diameter of the particles and accepting the hypothesis of independence between the localization and the genetic characteristic of the trees. In conclusion, the results show that the methodologies based on both Monte Carlo methods and mark correlation function may be powerful tools to the analysis of marked point processes.

* **Adviser:** João Domingos Scalon - UFLA.

1 INTRODUÇÃO

Muitos fenômenos são expressos por meio de ocorrências identificadas como pontos localizados no espaço, denominados processos pontuais. São exemplos: localização de crimes, ocorrências de doenças e localização de espécies vegetais. Observa-se que o objeto de interesse é a própria localização espacial dos eventos em estudo, testando hipóteses sobre o padrão observado, ou seja, os pontos estão distribuídos de forma aleatória, os pontos apresentam-se em aglomerados ou os pontos estão regularmente distribuídos. Os métodos de análise estatística de dados oriundos de processos pontuais estão bem desenvolvidos e descritos em detalhes em diversas referências.

Considerando-se que tem-se uma variável aleatória (marca) associada a cada ponto localizado no espaço, tem-se um fenômeno denominado processo pontual marcado. São exemplos de marcas: tipo de crime, tipo de doenças e altura de espécies vegetais. O objetivo básico da análise de processos pontuais marcados é verificar se existe dependência entre marcas e pontos. Por exemplo: pode-se estar interessado se dois tipos de crimes são independentes ou dependentes (apresentam interação ou repulsão). Outro exemplo: Será que árvores mais próximas são mais altas? Para detectar dependência entre as marcas e a localização do evento diversas metodologias têm sido propostas, tais como, as funções E , F , G , J , K , V e correlação marcada.

Apesar da função de correlação marcada ser uma das funções mais utilizadas na análise de processos pontuais marcados, pouco se conhece sobre métodos inferenciais utilizando esta função. Assim, conduziu-se este estudo, tendo como principal objetivo propor metodologias baseadas na função de correlação marcada, por meio de simulações Monte Carlo, para testar a hipótese de independência entre

marcas e pontos, em marcas contínuas e categóricas e comparar com as metodologias utilizadas por Scalon et al. (1998) e Louzada (2003). Os métodos propostos serão aplicados em duas situações científicas distintas, envolvendo análise de processos pontuais marcados.

Na primeira situação, a análise estatística espacial será conduzida em três amostras metalográficas planas de um material compósito de partículas de cerâmicas. O interesse da análise consiste no fato de que existem evidências de que a quebra superficial associada às partículas agrupadas parece prevalecer mais do que aquelas associadas a materiais compósitos menos agrupados. Assim, a caracterização quantitativa da distribuição espacial das partículas e da dependência entre os tamanhos das partículas (marca) e a localização (ponto) na região de estudo é de interesse, não só para uma compreensão melhor da relação entre inclusões e o comportamento mecânico de materiais, mas também para um melhor controle da produção de materiais compósitos. Observa-se que as marcas desses dados são contínuas. Na segunda situação, a análise estatística espacial será conduzida na distribuição de marcadores genéticos em árvores de uma área florestal. O interesse da análise estatística é determinar se os marcadores genéticos são determinados pela proximidade das árvores. Observa-se que as marcas destes dados são marcas categóricas.

As análises dos dois processos pontuais marcados serão procedidas de uma análise exploratória espacial utilizando as funções F , G , J e K por meio do teste da hipótese nula de completa aleatoriedade espacial da distribuição dos eventos (partículas e árvores), não considerando as marcas. Nas duas situações testa-se a hipótese nula de independência entre marcas e pontos.

Espera-se obter resultados que possam contribuir para o avanço do entendimento da análise de dados oriundos de processos pontuais marcados.

2 REFERENCIAL TEÓRICO

2.1 Um pouco da história

Um exemplo pioneiro, onde intuitivamente se incorporou a categoria espaço às análises realizadas, foi realizado no século *XIX* por John Snow. Em 1854 ocorria em Londres uma das várias epidemias de cólera trazidas das Índias. Pouco se sabia então sobre os mecanismos causais da doença. Duas vertentes científicas procuravam explicá-la: uma relacionando-a aos mananciais, concentrados nas regiões baixas e pantanosas da cidade, outra à ingestão de água insalubre. Estudos posteriores confirmaram esta hipótese, corroborada por outras informações tais como a localização do ponto de captação de água desta bomba à jusante (rio abaixo) da cidade, em local onde a concentração de dejetos, inclusive de pacientes coléricos, era máxima.

A variabilidade em fenômenos espaciais vem sendo uma das preocupações de pesquisadores, praticamente desde o início do século. Conforme Vieira (2000), em 1910 Smith estudou a disposição de parcelas no campo, em experimentos de rendimento de variedade de milho. Em 1913 Montgomery preocupado com o efeito de nitrogênio no rendimento de trigo, fez um experimento com 224 parcelas, nas quais ele mediu o rendimento de grãos. Numa tentativa de encontrar uma maneira única de analisar uma vasta coleção de dados. A partir de 1930, Fisher, na conhecida estação experimental em Rothamsted (Inglaterra), adotou as técnicas de casualização e repetição, com melhores conhecimentos de funções de distribuição que levaram à adoção de amostragem ao acaso, desprezando assim as coordenadas geográficas do ponto amostral, fazendo com que experimentos descritos acima não tenham tido continuidade no tempo. Só em 1951, na África do Sul, Krige constatou a presença da dependência espacial trabalhando com dados de concentração de

ouro. Por volta de 1963, Matheron baseando-se nessas observações, desenvolveu uma teoria, a qual ele chamou de Teoria das Variáveis Regionalizadas. Em 1970 Tobler criou um conceito chave na compreensão e análise dos fenômenos espaciais que é a dependência espacial, a qual, é chamada de primeira lei da geografia: "*todas as coisas são parecidas, no entanto, coisas mais próximas se parecem mais que coisas mais distantes*". Generalizando, pode-se afirmar que a maior parte das ocorrências, sejam estas naturais ou sociais, apresentam entre si uma relação que depende da distância. Em 1977, Ripley propôs vários métodos baseados em distâncias, que são ferramentas fundamentais nas análises de padrões pontuais, por meio desses métodos Ripley propôs estimadores não viesados que corrigem os efeitos de bordas. A partir daí, diversos pesquisadores, tais como: Cressie (1993), Bailey & Gatrell (1995), Stoyan & Penttinen (2000), Diggle (2003), Baddeley & Turner (2005), Guan (2006), Van Lieshout (2006), Lloyd (2007) e Illian et al. (2008), desenvolveram métodos com o objetivo de se estudar, não unicamente os processos pontuais, mas as características associadas aos pontos, que atualmente são chamados de processos pontuais marcados.

2.2 Estatística espacial

A análise espacial pode ser definida como uma técnica que busca descrever os padrões existentes nos dados espaciais e estabelecer, preferencialmente, de forma quantitativa, os relacionamentos entre as diversas variáveis geográficas. Segundo Avelar (2008), os objetivos da estatística espacial são: descrição cuidadosa e precisa de eventos no espaço geográfico (incluindo a descrição de padrões) e exploração sistemática do padrão dos eventos e de sua associação no espaço e predição. A taxonomia mais utilizada para descrever o problema da análise espacial considera três tipos de dados conforme pode ser visto em Bailey & Gatrell (1995),

Cressie (1993) e Câmara et al. (2002), a saber: processos pontuais, superfícies contínuas e áreas com contagens.

Processos pontuais são fenômenos expressos por meio de ocorrências identificadas como pontos localizados no espaço. São exemplos desse tipo de dados a localização da ocorrência de casos de doenças e a localização de indivíduos de uma determinada espécie.

Superfícies contínuas são fenômenos que se distribuem continuamente em uma região. Usualmente, esses tipos de dados são resultantes de levantamentos de recursos naturais e que incluem mapas geológicos, topográficos e ecológicos. Um exemplo desse tipo de dados são medidas da concentração de um elemento químico no solo.

Áreas com contagens são fenômenos associados aos dados de levantamentos populacionais, como censos, e que, originariamente, referem-se a indivíduos localizados em pontos específicos no espaço. Normalmente, esses dados são agregados em unidades de análises, usualmente delimitadas por polígonos fechados, tais como: setores censitários, municípios e microrregiões.

A partir da divisão anterior, observa-se que os problemas de análise espacial lidam principalmente com dados ambientais, com dados socioeconômicos, entre outros. Em todos os casos, a análise espacial é composta por um conjunto de procedimentos encadeados, cuja finalidade é a escolha de um modelo inferencial que considere explicitamente os relacionamentos espaciais presentes no fenômeno.

2.3 Processos pontuais

Conforme Ripley (1981), uma característica importante de um processo pontual espacial é a localização dos pontos e a relação entre eles, verificando se os eventos observados apresentam algum tipo de padrão sistemático, ao invés de

estarem distribuídos aleatoriamente. A visão da estatística em processos pontuais é de analisar a estrutura geométrica de padrões formados por objetos que são distribuídos aleatoriamente no espaço *uni*, *bi* ou *tri*-dimensional. Assim, como a análise de dados espaciais é normalmente abordada?

Uma análise exploratória e uma representação gráfica do padrão espacial são muito úteis, sendo um passo preliminar para o entendimento de suas propriedades. A inspeção visual provê uma caracterização qualitativa do tipo de padrão, até mesmo se condições bastante vagas são usadas na descrição inicial.

Uma característica importante de um padrão espacial é a localização dos pontos e a relação entre eles (aleatório, regular e agrupado), conforme é apresentado na Figura 1.

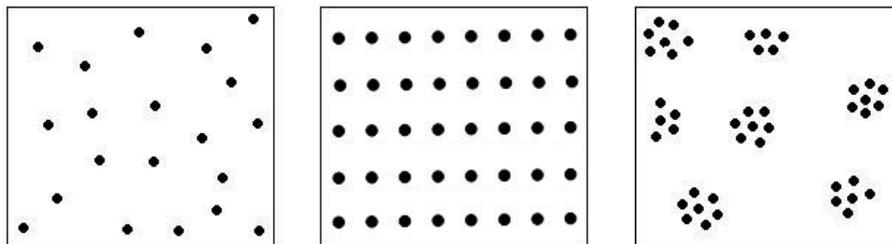


FIGURA 1: Três padrões de pontos simulados: aleatório (esquerda), regular (centro), agrupado (direita).

Câmara et al. (2002) afirmaram que essas técnicas permitem descrever a distribuição das variáveis em estudo, identificando-se observações atípicas (outliers) não só em relação ao tipo de distribuição, mas também em relação aos vizinhos, buscando-se a existência de padrões na distribuição espacial. Por meio desses procedimentos é possível estabelecer hipóteses sobre as observações, de forma a selecionar o melhor modelo suportado pelos dados.

Segundo Illian et al. (2008), os processos pontuais são modelos matemáticos que descrevem o arranjo de objetos que são agrupados, regulares e aleatórios dis-

tribuindo-se no plano ou no espaço, cuja localização foi gerada por um processo estocástico.

Segundo Waller & Gotway (2004), um processo estocástico é um modelo probabilístico definido por uma coleção de variáveis aleatórias, sendo $\{Z(\mathbf{t})/\mathbf{t} \in T\}$ indexado por um conjunto T . É conveniente considerar T como sendo um subconjunto de números reais; sendo na maioria dos casos cada T uma medida semelhante acontecendo em um tempo ou lugar diferente (por exemplo: o número de pessoas em uma agência postal, em uma fila no i -ésimo tempo, ou a quantidade de chuva no i -ésimo local).

Conforme Waller & Gotway (2004), um processo pontual descreve um processo estocástico como uma variável aleatória que representa o local de um evento no espaço. Uma realização desse processo é uma coleção de locais gerado sobre um modelo de processo pontual espacial, isto é, uma representação da realização de um conjunto de dados que é resultado de um modelo particular (observado ou simulado). Para sua caracterização, esse processo estocástico pode ser descrito em termos de propriedades de primeira e segunda ordem.

2.3.1 Propriedades de primeira e segunda ordem

Segundo Câmara et al. (2002), propriedade de primeira ordem, consideradas globais ou de larga escala, correspondem a variações no valor médio do processo no espaço. Nesse caso, estamos interessados na intensidade do processo, isto é, no número de eventos por unidade de área. Para Gatrell et al. (1995), as propriedades de segunda ordem, denominados locais ou de pequena escala, representam a dependência espacial do processo, proveniente da estrutura de correlação espacial. Para verificar a dependência espacial, procura-se estimar o relacionamento entre pares de eventos (por unidade de área) no espaço, o que corresponde a uma aproxi-

mação do cálculo da covariância entre as variáveis aleatórias que representam cada evento.

Conforme Perry et al. (2006), na prática, freqüentemente surge um composto (mistura) das propriedades de primeira e segunda ordem, ocasionando a independência nas variáveis aleatórias representadas por um processo estocástico espacial, precisando ser substituído por alguma alternativa simples na estrutura de covariância.

Considere um conjunto de pontos $N = \{x_1, x_2, \dots\}$ numa determinada região A onde ocorreram eventos. O processo pontual é modelado considerando-se sub-regiões S em A por meio de sua esperança $E[N(S)]$ e a covariância amostral $C[N(S_i), N(S_j)]$, em que $N(S)$ denota o número de eventos em S . Sendo o objetivo da análise, estimar as localizações prováveis de ocorrência de determinados eventos, essas estatísticas devem ser inferidas considerando o valor limite da quantidade de eventos por área. Esse valor limite corresponde à esperança de $N(S)$ para uma pequena região ds em torno do ponto x , quando essa tende a zero. Essa esperança é denominada intensidade (propriedade de primeira ordem), sendo definida como:

$$\lambda(x) = \lim_{|ds| \rightarrow 0} \left\{ \frac{E[N(ds)]}{|ds|} \right\}. \quad (2.1)$$

Propriedades de segunda ordem podem ser definidas da mesma forma, considerando a intensidade conjunta $\lambda(x_i, x_j)$ entre duas regiões infinitesimais $|ds_i|$ e $|ds_j|$ que contém os pontos x_i e x_j .

$$\lambda(ds_i, ds_j) = \lim_{|ds_i|, |ds_j| \rightarrow 0} \left\{ \frac{C[N(ds_i), N(ds_j)]}{|ds_i| \cdot |ds_j|} \right\}. \quad (2.2)$$

2.3.2 Processo estacionário e isotrópico

Quando o processo é estacionário, ou seja, as propriedades estatísticas da variável independem de sua localização absoluta, portanto, a média e a variância são constantes em qualquer sub-área e a covariância entre dois pontos quaisquer depende somente de sua localização relativa conseqüentemente, $\lambda(x)$ é uma constante, ou $\lambda(x) = \lambda$.

Segundo Lloyd (2007), o termo estacionariedade é freqüentemente usado nos resultados de algum processo que tenha propriedades semelhantes em todos os locais na região de interesse. Se além de estacionário, também é isotrópico, ou seja, a covariância depende somente da distância euclidiana entre os pontos e não da direção entre eles, $\lambda(x_i, x_j)$ se reduz à $\lambda(|h|)$ sendo $|h|$ a distância entre os dois pontos. Quando o processo é não estacionário, ou seja, a intensidade média varia na região A , a modelagem da estrutura de dependência $\lambda(x_i, x_j)$ incorpora a variação de $\lambda(x)$. A maior parte das técnicas de análise de distribuição de pontos supõe, explícita ou implicitamente, um comportamento estacionário e isotrópico do processo aleatório subjacente aos eventos analisados.

Segundo Illian et al. (2008), um processo pontual N é chamado de estacionário se N e o processo pontual transladado N_x possuem a mesma distribuição por meio de todas as translações de x . Isto é escrito como:

$$N \equiv N_x, \quad (2.3)$$

em que N_x é o processo pontual resultante de uma translação de todos os pontos de N pelo mesmo vetor x : se $N = \{x_1, x_2, \dots\}$, então $N_x = \{x_1 + x, x_2 + x, \dots\}$

as expressões tem a mesma distribuição significando:

$$\begin{aligned} P[N(S_1) = n_1, \dots, N(S_k) = n_k] &= P[N_x(S_1) = n_1, \dots, N_x(S_k) = n_k] \\ &= P[N(S_1 - x) = n_1, \dots, N(S_k - x) = n_k], \end{aligned}$$

em que a primeira igualdade na segunda linha mantêm-se inalterável porque $N_x(S) = N(S - x)$ para todo S e x , em que $S - x = \{y - x : y \in S\}$ é o conjunto S fixado pelo vetor $-x$. O número de pontos resultando e fixando um conjunto de observações S , e contando o número de pontos do processo transladado N_x é igual ao número de pontos obtidos ao fixar o processo pontual, e transladando o conjunto de observações na direção oposta. Da equação (2.3), subentende-se que:

$$N(S) \equiv N(S_x), \quad \text{para todo } S \text{ e } x, \quad (2.4)$$

isto é, o número de pontos em S e o conjunto transladado S_i , tem a mesma distribuição. Isto tem uma importante consequência a fim de caracterizar a distribuição de um processo pontual.

Conforme Illian et al. (2008), o conceito de isotropia é análogo ao de estacionariedade. Ao invés de translações por meio de vetores, as rotações ao redor da origem são consideradas. No caso do plano uma rotação é descrita por um ângulo θ entre 0° e 360° . Se $x = (\xi, \eta)$ é um ponto no \mathbb{R}^2 tendo as coordenadas ξ e η , então o movimento executado por um ponto em torno de seu próprio eixo $R_\theta x$ possui coordenadas dada por:

$$\begin{aligned} \xi_\theta &= \xi \cos \theta + \eta \operatorname{sen} \theta \quad e \\ \eta_\theta &= -\xi \operatorname{sen} \theta + \eta \cos \theta. \end{aligned}$$

Um processo N é considerado isotrópico se:

$$N = \{x_1, x_2, \dots\} \quad e \quad R_\theta N = \{R_\theta x_1, R_\theta x_2, \dots\}.$$

Tendo a mesma distribuição para todo θ . Um processo pontual que é isotrópico e estacionário é chamado de movimento invariante.

2.3.3 Completa aleatoriedade espacial (CSR)

Diggle (2003) afirma que a análise estatística dos padrões de distribuições de pontos requer um modelo teórico de referência base para o desenvolvimento de métodos formais que checam a significância dos resultados exploratórios. O modelo teórico mais simples é conhecido como completa aleatoriedade espacial (ou *complete spatial randomness* - CSR), que corresponde a uma situação onde é igualmente provável que um evento aconteça em qualquer local dentro da área de estudo, embora os locais sejam de outros eventos. Quer-se dizer que eventos seguem uma distribuição uniforme sobre a área de estudo, e são independentes um do outro. Conforme Waller & Gotway (2004), usa-se o termo uniforme no sentido de seguir uma distribuição de probabilidade uniforme pela área de estudo, não no sentido de "uniformemente" disperso na área. No modelo de CSR, considera-se que as ocorrências em cada sub-área sejam não-correlacionadas, estando-se associadas à mesma distribuição de probabilidade de Poisson.

Segundo Lima (2005), para explicar a associação da distribuição de Poisson com a ocorrência aleatória de pontos em uma determinada área, pode-se considerar um raciocínio semelhante ao utilizado na obtenção da distribuição de Poisson a partir da distribuição Binomial e considera-se $n \rightarrow \infty$ e $p \rightarrow 0$ de modo que $np \rightarrow \lambda$ e seja constante.

Considere uma determinada área A , em alguma unidade de área conveniente, onde se têm x pontos localizados aleatoriamente e, portanto, independentes. Assim, tem-se uma intensidade de $\lambda = \frac{x}{A}$ pontos por unidade de área.

Suponha que a área A seja composta por n sub-regiões extremamente pequenas, de tal maneira que a probabilidade de que cada uma dessas sub-regiões contenha mais de um ponto seja desprezível. Suponha ainda que, cada um desses x pontos tenham a mesma probabilidade p de estar localizado em qualquer uma das n sub-regiões. Portanto, pode-se assumir que têm-se x sucessos obtidos em n realizações de um ensaio e, sendo X a variável aleatória que representa esse número de sucessos, $X \sim Bin(n, p)$, ou seja,

$$p(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n. \quad (2.5)$$

Sendo $\lambda = np$, o qual representa o número esperado de sucessos por unidade de espaço, ou seja, a intensidade ou o número de pontos por unidade de área, como sendo constante e substituindo $p = \frac{\lambda}{n}$ na equação (2.5) tem-se:

$$\begin{aligned} p(X = x) &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{n(n-1) \cdot (n-2) \cdot \dots \cdot (n-x+1) \cdot (n-x)!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{\lambda^x (n-1) \cdot (n-2) \cdot \dots \cdot (n-x+1)}{x! n^{x-1}} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

Note que,

$$\begin{aligned} \frac{(n-1) \cdot (n-2) \cdot \dots \cdot (n-x+1)}{n^{x-1}} &= \frac{(n-1)}{n} \cdot \frac{(n-2)}{n} \dots \frac{(n-x+1)}{n} \\ &= \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{x-1}{n}\right) \end{aligned}$$

e, assim,

$$P(X=x) = \frac{\lambda^x}{x!} \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdot \left(1 - \frac{x-1}{n}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-x}.$$

Considerando $n \rightarrow \infty$, o que estaria relacionado a um número grande de sub-regiões, o termo $\left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}$, pela própria definição do número e e os termos $\left(1 - \frac{a}{n}\right) \rightarrow 1$, em que $a = 1, 2, \dots, x-1$ e $a = \lambda$. Assim,

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \dots, n,$$

que corresponde à distribuição de Poisson.

Intuitivamente, pode-se considerar que a posição dos eventos é independente e de que os eventos têm igual probabilidade de ocorrência em toda a região A . A suposição de independência é conveniente, tornando as inferências estatísticas mais simples. Entretanto, modelos que envolvem dependência estatística são frequentemente mais realistas (Cressie, 1993).

2.4 Análise de processos pontuais

Para conduzir tal análise, primeiramente, tem-se que caracterizar a distribuição espacial da ocorrência dos indivíduos em termos de estimação de propriedades de primeira e segunda ordem (Cressie, 1993). Para a análise das propriedades de

primeira ordem utiliza-se algum estimador de intensidade pontual do processo (ex. Kernel).

O estimador de intensidade de Kernel é uma alternativa simples, com objetivo de analisar o comportamento de padrões de pontos e estimar a intensidade pontual do processo em toda a região de estudo. Segundo Bailey & Gatrell (1995), o estimador de Kernel foi desenvolvido afim de obter uma estimativa "suave" de uma densidade de probabilidade univariada ou multivariada de observações amostrais.

Segundo Câmara et al. (2002), essa função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse; conforme é apresentado na Figura 2.

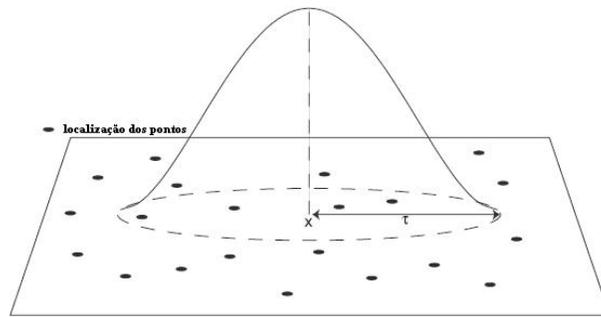


FIGURA 2: Estimador de intensidade pontual do processo em toda região de estudo (Kernel).

Conforme Bailey & Gatrell (1995), x_i representa uma localização genérica cujo valor se deseja estimar nas localizações de n eventos, ou seja, x_1, \dots, x_n em uma região A . O estimador de intensidade é calculado a partir dos m eventos $\{x_1, \dots, x_{1+m-1}\}$ contidos num raio de tamanho τ em torno de x_i e na distância h entre a posição e a i -ésima amostra, a partir de funções cuja a formula geral é:

$$\hat{\lambda}_t(x) = \frac{1}{\tau^2} \sum_{i=1}^n k\left(\frac{h(x_i, x)}{\tau}\right), \quad h(x_i, x) \leq \tau.$$

Esse estimador é chamado de estimador de intensidade de Kernel e seus parâmetros básicos são: (a) um raio de influência ($\tau \geq 0$) que define a vizinhança do ponto a ser interpolado e controla o "suavização" da superfície gerada; (b) uma função de suavização do fenômeno. Segundo Câmara et al. (2002), o raio de influência define a área centrada no ponto de estimação x_i que indica quantos eventos x_i contribuem afim de estimar a função intensidade λ . Conforme Illian et al. (2008), adotando-se um raio pequeno, esse raio gera uma superfície muito descontínua; se for grande, a superfície poderá ficar muito suavizada. No caso da função de interpolação $k(\cdot)$, é comum usar funções de terceira ou quarta ordem, como:

$$k(h) = \frac{3}{\pi}(1 - h^2),$$

em que h representa a distância euclidiana entre a localização, em que deseja-se calcular a função e o evento observado. Segundo Câmara et al. (2002), com o uso desta função de quarta ordem, o estimador de intensidade pode ser expresso por:

$$\hat{\lambda}_t(x) = \sum_{h_i \leq \tau} \frac{3}{\pi\tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2.$$

O estimador de intensidade é útil afim de fornecer uma visão geral da distribuição de primeira ordem dos eventos.

As técnicas mais utilizadas para estimar as propriedades de segunda ordem são baseadas nas funções empíricas F , G , J e K , as quais serão descritas a seguir.

2.4.1 Função F

A função F , introduzida por Ripley (1977), é a função de distribuição acumulada da distância entre um ponto aleatoriamente escolhido e o evento mais pró-

ximo. O mais simples estimador para a função F pode ser obtido pela equação:

$$\hat{F}(x) = \lambda^{-1} \sum_{i=1}^m I_x(x_i), \quad (2.6)$$

em que m é o número de pontos arbitrários e aleatórios no mapa, x_i representa a distância do i -ésimo ponto aleatório para o mais próximo dos n eventos no mapa analisado e $I_x(x_i)$ é uma função indicadora igual a 1 quando x_i é menor ou igual a x , e 0 caso contrário.

Entretanto, o estimador apresentado na equação (2.6) é viesado, uma vez que, para uma mesma distância, pontos aleatórios localizados nas bordas do mapa terão uma probabilidade menor de ter um evento próximo do que os pontos localizados no centro do mapa. Esse efeito ocorre porque, na prática, trabalha-se com regiões de área finita enquanto teoricamente a área é considerada infinita. Assim, a distância entre um ponto aleatório e um evento que se encontra fora do mapa é ignorada, mesmo que esta esteja menor que qualquer distância entre um ponto aleatório e um evento localizado dentro do mapa; conforme é apresentado na Figura 3.

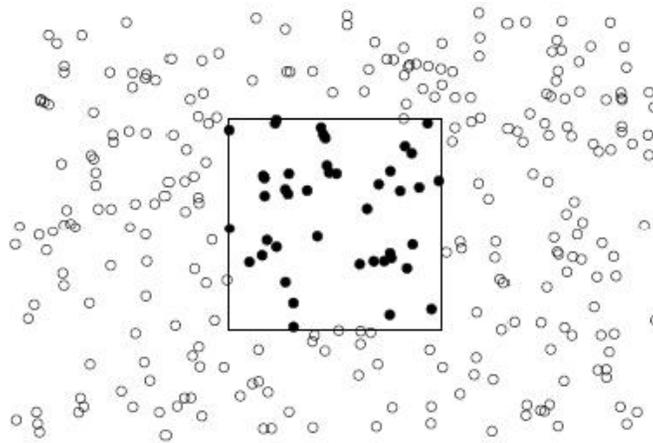


FIGURA 3: Região de estudo (pontos pretos) de área finita, fixada e conhecida, onde ocorreu um fenômeno de interesse.

Um estimador não viesado proposto por Ripley (1977), que corrige o efeito de bordas é dado pela equação:

$$\hat{F}(x) = \frac{\sum_{i=1}^n I_x(x_i, r_i)}{\sum_{i=1}^n I_x(r_i)}, \quad x > 0, \quad (2.7)$$

em que r_i é a distância de um evento até o ponto mais próximo na borda do mapa e $I_x(x_i, r_i)$ é uma função indicadora igual a 1, quando x_i é menor ou igual a x e r_i é maior ou igual a x e 0, caso contrário.

Diggle (2003), mostra que a função F teórica para um processo de Poisson homogêneo, com área infinita, é dada pela equação:

$$F(x) = 1 - \exp(-\lambda\pi x^2), \quad x > 0. \quad (2.8)$$

2.4.2 Função G

A função G definida por Ripley (1977), conhecida na literatura como o método do vizinho mais próximo, é uma função de distribuição acumulada da distância entre um evento e o vizinho mais próximo. O mais simples estimador para a função G é definido por:

$$\hat{G}(y) = \lambda^{-1} \sum_{i=1}^n I_y(y_i), \quad (2.9)$$

em que, n é o número de pontos arbitrário e aleatório no mapa y_i é a distância do i -ésimo ponto aleatório para o mais próximo dos n eventos analisado no mapa e $I_y(y_i)$ é uma função indicadora igual a 1 quando y_i é menor ou igual a y e 0 caso contrário.

Um estimador não viesado proposto por Ripley (1977) que corrige o efeito de

bordas é dado pela equação:

$$\hat{G}(y) = \frac{\sum_{i=1}^n I_y(y_i, r_i)}{\sum_{i=1}^n I_y(r_i)}, \quad y > 0, \quad (2.10)$$

em que r_i é a distância de um evento até o mais próximo na borda do mapa, $I_y(y_i, r_i)$ é uma função indicadora.

Desde que os eventos sejam independentemente distribuídos, a função de distribuição G aproximada seria dada por um processo de Poisson homogêneo análogo à da função F . Para tal, basta apenas trocar a distância x pela distância y na equação (2.8). Conforme afirma Venables & Ripley (2000), sobre a hipótese de *CSR* as funções F e G são iguais.

2.4.3 Função J

A função J proposta por Van Lieshout & Baddeley (1996), é a razão entre a função de distribuição acumulada da distância entre um evento e o vizinho mais próximo (função G), e a função de distribuição acumulada da distância entre um ponto aleatoriamente escolhido e o evento mais próximo (função F), na qual, compara as distâncias entre os eventos por meio de distâncias fixadas de algum ponto amostral. A função J é definida por:

$$J(z) = \frac{1 - G(z)}{1 - F(z)}, \quad (2.11)$$

para todo $z > 0$ tal que $F(z) < 1$.

Segundo Van Lieshout & Baddeley (1996), por meio de uma grande classe de processos pontuais, a função J é constante para valores z maiores que a distância efetiva de interações no espaço. Conseqüentemente, J pode ser utilizada

para grandes distâncias de interações espaciais. Além disso, a função J é capaz de avaliar explicitamente vários modelos estocásticos, de forma que isto poderia ser usado diretamente para estimação de parâmetros.

Segundo Thönnies & Van Lieshout (1999), uma simples análise gráfica da função J estimada fornece precisamente informações do tipo e veemência de interações entre eventos.

Conforme Baddeley et al. (2000), sob a hipótese de *CSR*, do processo de Poisson homogêneo ($G \equiv F$) tem-se que $J(z) \equiv 1$. Uma explicação intuitiva é que o viés devido aos efeitos de borda é aproximadamente igual para a estimativa de $1 - F(z)$ e $1 - G(z)$, de forma que esses vieses aproximadamente cancele a relação do estimador de J . Na prática, o grande problema nas análises de processos pontuais é a restrição por meio da fronteira de uma determinada região de estudos A . Como consequência, a estimação das funções empíricas são baseadas na média de várias distâncias no processo pontual, segue-se que, o estimador não corrigido de J pode ser usado para uma avaliação direta da divergência no processo de Poisson homogêneo, comportando-se bem sobre a suposição de *CSR*, algo não possível com as estimativas não corrigidas de F e G .

Segundo Van Lieshout & Baddeley (1996), valores de $J(z)$ menores ou maiores que 1 (um), por meio de uma particular distância z , indicam agrupamentos ou regularidades, respectivamente. Bedford & Van Den Berg (1997), afirmam que é possível encontrar um processo de Poisson heterogêneo para qual $J(z) = 1$.

2.4.4 Função K

Diggle (2003), mostra que a função F apresenta grande poder para detectar agrupamentos de eventos, enquanto a função G é mais poderosa para detectar eventos regularmente distribuídos. Assim, recomenda-se usar uma função que seja

igualmente poderosa para detectar os dois padrões em diferentes escalas, como a função K , proposta por Ripley (1977). Formalmente pode-se definir a função K como:

$$K(t) = \frac{2\pi}{\lambda^2} \int_0^t \lambda^2(x) x dx.$$

A medida de primeira ordem é representada por uma função $\lambda_1(x)$, que é a função de intensidade de primeira ordem. Ela indica a intensidade do processo na localização x . No caso do plano \mathfrak{R}^2 , uma localização é definida por um sistema de coordenadas (x, y)

Conforme Anjos (2004), ao gerar-se um padrão num plano, a intensidade de primeira ordem do processo pode ser interpretada como a densidade do padrão numa localização particular, na definição da função K :

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

ou seja, assumi-se que a intensidade de primeira ordem do processo que gerou o padrão espacial é constante e igual a λ em toda área de estudo. Essa pressuposição é chamada de pressuposição de estacionariedade do processo.

Conforme Capretz (2004), a função K , ao se combinar as funções de intensidade de primeira ordem e segunda ordem, resulta numa função descritiva do padrão espacial com interpretações mais diretas. Assim,

$$K(t) = \lambda^{-1} E [N_0(t)], \quad (2.12)$$

em que $N_0(t)$ é o número de eventos mais próximos dentro da distância t de um evento arbitrário e λ é a intensidade.

Quando se estuda o padrão espacial numa determinada área A , a aleatoriedade do padrão é definida pela *CSR*; isto é, o processo que gerou o padrão é aleatório em qualquer escala espacial.

Na *CSR*, as funções de intensidade são constantes, isto é,

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

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

em que $\|x - y\|$, é a distância euclidiana entre os pontos x e y .

Como consequência, a função K sobre a hipótese de *CSR* será:

$$K(t) = \frac{2\pi}{\lambda^2} \int_0^t \lambda_2(x) x dx$$

$$K(t) = \frac{2\pi}{\lambda^2} \int_0^t \lambda^2 x dx$$

$$K(t) = \pi t^2. \quad (2.13)$$

Portanto, a função K pode ser facilmente utilizada para testar a hipótese de *CSR*. O mais simples e natural estimador de $K(t)$ que corrige o efeito de bordas é dado por:

$$\hat{K}(t) = \frac{|A|}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{I_t(u_{ij})}{w_{ij}}, \quad t > 0, \quad (2.14)$$

em que u_{ij} é a distância entre os eventos i e j ($i \neq j$), $I_t(u_{ij})$ é uma função

indicadora igual a 1 quando w_{ij} é menor que a distância t e 0, caso contrário, n é o número de eventos no mapa analisado com área $|A|$ e w_{ij} é um fator de correção que representa a proporção da circunferência ao redor de um evento i , passando sobre o evento j que está dentro de $|A|$.

Para entender esta função, imagine que um evento é "visitado" e em volta dele é construído um conjunto de círculos concêntricos. O número acumulado de eventos dentro de cada uma destas distâncias é computado. Muitos outros eventos são similarmente visitados e o número acumulado de eventos dentro desta distância de raio t em volta de todos os eventos estima $K(t)$, como está apresentado na Figura 4.

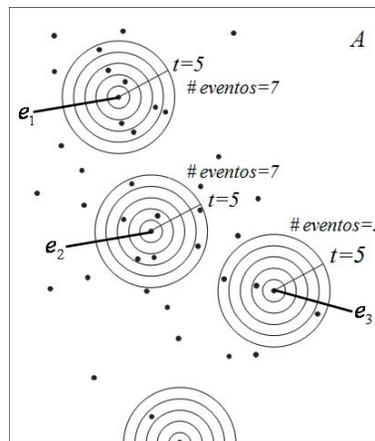


FIGURA 4: Idéia gráfica da função K.

Se a função K for comparada com outros métodos freqüentemente utilizados na literatura, comprovamos seu potencial de uso e sua robustez estatística. Métodos baseados em distâncias ou área, que são os mais comuns, carregam em sua essência "simplificações" da realidade, uma vez que tratam apenas da simples contagem de pontos dentro de uma região de estudo ou da procura pelo primeiro vizinho mais próximo da mesma espécie (Ripley, 1978). Desse modo, muita infor-

mação é perdida nas análises espaciais, uma vez que tais métodos só conseguem detectar o padrão espacial naquela escala particular em que os dados foram coletados, referindo-se à distância entre os pontos (Schlather, 2001).

Por meio de um processo aleatório, $K(t)$ seria igual a πt^2 . Uma forma de comparar a estimativa K de um conjunto de dados observados com πt^2 seria por meio de uma função auxiliar L , a fim de facilitar a interpretação gráfica. O estimador da função L é dado por:

$$\hat{L}(t) = \sqrt{\frac{\hat{K}(t)}{\pi}}. \quad (2.15)$$

2.4.5 Teste de hipótese por meio da CSR

Na simulação estocástica, os procedimentos reproduzem imagens de uma determinada função aleatória por meio de realizações equiprováveis do modelo no processo estocástico estabelecido. Cada realização também chamada de imagem estocástica reflete as propriedades consideradas no modelo da função teórica considerada. Geralmente as realizações devem distinguir os dados e reproduzir uma função de distribuição acumulada univariada qualquer, e a estrutura de covariância espacial.

As funções empíricas apresentadas anteriormente podem ser utilizadas de duas formas, gráfico e teste formal, a fim de testar a hipótese nula de que os eventos seguem uma distribuição de Poisson homogênea, ou seja, que os eventos estão aleatoriamente distribuídos no plano. Esses métodos estão descritos em Scalón et al. (2003).

Conforme Câmara et al. (2002), os gráficos com os resultados dessas funções podem ser usados como um método exploratório a fim de se verificar evidências de interação entre os eventos.

Bailey & Gatrell (1995), afirmam que, se o gráfico apresentar um crescimento rápido para pequenos valores de distância, esta situação aponta para interação entre os eventos caracterizando agrupamentos nestas escalas. Se o gráfico apresentar valores pequenos no seu início, e só crescer rapidamente para valores maiores de distância, esta situação aponta para uma distribuição mais regular. Conforme Baddeley & Turner (2005), em situações práticas, as interações entre os eventos existem, se e só se, houver uma proximidade física pequena.

A análise de vizinhança pode ser usada como método formal afim de se comparar estatisticamente a distribuição dos eventos observados com o que se esperaria sob a hipótese de *CSR* (Lloyd, 2007). Esta metodologia consiste em se criar envelopes de simulação por meio da distribuição *CSR*, para se acessar a significância dos desvios.

2.5 Análise de processos pontuais marcados

Na análise de processos pontuais marcados, o interesse não está apenas na ocorrência de determinado tipo de evento, mas em determinadas características. Por exemplo: em ecologia além de se verificar o comportamento espacial das espécies (aquáticas ou terrestres), pode-se estar interessado nas características associadas aos eventos, tais como: sexo, altura, classe de idade, etc. Em estudos florestais, como por exemplo: Penttinen et al. (1992), Gavrikov & Stoyan (1995) e Stoyan & Penttinen (2000), o interesse não estava apenas no comportamento espacial do evento em estudo (árvores), mas nas características presentes em cada árvore, tais como: diâmetro, altura, número de brotos após a renovação das árvores, etc. O objetivo básico da análise de processos pontuais marcados é verificar se existe dependência entre marcas e pontos.

Os processos pontuais marcados diferem dos dados de superfície e de área

pelo fato de que os pontos e as marcas não vêm associados à realização de uma variável aleatória (variável resposta). O interesse aqui é a informação a respeito das características dos pontos. Particularmente, tem-se interesse em responder se o ponto tem alguma relação com as marcas (características de cada ponto).

Segundo Penttinen et al. (1992), um processo pontual marcado é um processo estocástico $\Psi = \{[x_i, m(x_i)]\}$ que gera os pontos e as marcas associadas aos eventos no plano e m_i são as marcas associadas a esses pontos. Uma realização de Ψ é um conjunto de posições juntamente com as marcas associadas. Segundo Baddeley & Turner (2005), um processo pontual marcado Ψ é a medida da dependência entre marcas e pontos, numa distância euclidiana de raio h . Stoyan & Stoyan (1994), definem um processo pontual marcado por meio de um processo estacionário e isotrópico, em que x_i denota a localização das observações e m_i é o valor real da marca medida em x_i .

Para Illian et al. (2008), as suposições de um processo estacionário e isotrópico são consideradas iguais ao processo pontual sem marcas, entretanto uma definição formal de um processo pontual marcado estacionário também envolve as marcas.

Considere um processo Ψ junto com um processo transladado, denotado por Ψ_i e definido da seguinte forma: se $\Psi = \{[x_1, m(x_1)], [x_2, m(x_2)], \dots\}$, então:

$$\Psi_i = \{[x_1, m(x_1)], [x_2, m(x_2)], \dots\}.$$

Isto significa que no processo pontual marcado transladado, as marcas são fixas, mas os pontos são transladados. Consistindo na seguinte definição: um processo pontual marcado Ψ é estacionário se, e somente se,

$$\Psi \equiv \Psi_i \quad \text{para todo } x \in \mathfrak{R}^d,$$

isto é, o processo pontual marcado Ψ e o processo pontual transladado por x tem a mesma distribuição. A definição de um processo pontual isotrópico é a mesma definida na seção 2.3.2.

Conforme Gavrikov & Stoyan (1995), na prática, apenas realizações de Ψ são observadas, ou seja, um conjunto de posições com marcas associadas em uma área de estudo $A(W)$. Assim, afim de realizar-se uma análise estatística espacial assume-se que o processo pontual marcado seja estacionário e/ou isotrópico (Gavrikov & Stoyan, 1995; Illian et al., 2008). Esses conceitos estão relacionados aos efeitos de primeira e segunda ordem.

Efeito de primeira ordem é a média do processo no espaço, enquanto efeito de segunda ordem é a covariância entre os pontos x_i e x_j . O processo Ψ é considerado estacionário se os efeitos de primeira e segunda ordem são constantes em toda a região estudada, ou seja, não existe tendência. O processo Ψ é considerado isotrópico se além de estacionário, a covariância depende apenas da distância euclidiana entre os pontos e não da direção entre eles. Um processo pontual Ψ que é isotrópico e estacionário é chamado de movimento invariante. Veja Illian et al. (2008) e Stoyan et al. (1995) por meio de um tratamento mais teórico sobre essas suposições. Para análise de processos pontuais marcados, assumi-se que o processo Ψ seja, pelo menos, estacionário (Wälder & Wälder, 2008).

Descrevendo-se um processo pontual marcado, deve-se analisar as suas propriedades de primeira e segunda ordem. Uma maneira simples de analisar as propriedades de primeira ordem é estimar a intensidade do processo pontual em toda a região de estudo, ou seja, $\hat{\lambda} = n \times A(W)^{-1}$. Analisando-se as propriedades de segunda ordem deve-se utilizar alguma medida que não somente descreva a variabilidade dos pontos mas também a variabilidade das marcas, de tal forma que possa descrever a correlação entre as marcas e os pontos em diferentes escalas

(Walder & Stoyan, 1996). A funo de correlao marcada, apresentada a seguir,  uma dessas medidas.

2.5.1 Funo de correlao marcada

A correlao espacial das marcas no processo pontual marcado pode ser descrito usando-se uma funo teste $f(m_i, m_j)$ por meio de duas marcas m_i e m_j . Essa funo teste caracteriza a correlao entre as marcas m_i e m_j em diferentes pontos, condicionada a uma distncia euclidiana h entre os pontos. Baddeley & Turner (2005) e Walder & Stoyan (1996) sugerem as seguintes funo teste: Para marcas com valores reais contnuos usa-se $f(m_i, m_j) = m_i \times m_j$ ou, alternativamente, $f(m_i, m_j) = 0.5 \times (m_i - m_j)^2$. Para marcas categricas (padres de pontos mltiplos) usa-se $f(m_i, m_j) = I\{m_i = m_j\}$, em que I  um indicador valendo 1 quando a igualdade  satisfeita e zero, caso contrrio.

Em geral, as marcas so nmeros aleatrios positivos e, conseqentemente, a funo teste tambm  aleatria. As funo testes apresentam interpretaes diretas. No caso de marcas contnuas, marcas com valores grandes acarretam grandes valores na funo teste, enquanto marcas com pequenos valores acarretam funo com pequenos valores. Interpretao anloga pode ser feita para marcas discretas. Assim, a funo teste indica, de uma maneira natural, a correlao entre as marcas condicionada a uma particular distncia. Conforme Goulard et al. (1995), em determinadas situaes, a funo teste pode conduzir a interpretaes similares a marca de variograma, que  bem conhecido na Geoestatstica (Walder & Stoyan 1996, Schlather et al. 2004).

Gavrikov & Stoyan (1995) afirmam que, em um processo de localizaes no marcadas, as propriedades de primeira e segunda ordem resumem-se a intensidade λ , como sendo o nmero mdio de pontos por unidade de rea e a distribuio m_i

de marcas dando valores aos respectivos pontos, conseqüentemente, a marca média μ é freqüentemente usada. Penttinen et al.(1992), afirmaram que estas quantidades dadas anteriormente, são propriedades unicamente de interesses individuais, sendo λds , a probabilidade de encontrar um ponto num conjunto infinitesimal de área ds e μ as características médias por meio de um ponto aleatoriamente escolhido, enquanto m_i da a distribuição das marcas.

A função teste pode ser normalizada pela média, ou esperança, das marcas μ consideradas. Isto significa que, $E(m_i) = \mu$ para todo $i = 1, \dots, n$. Assim, pode-se definir a função de correlação marcada como sendo:

$$\rho_f(h) = \frac{E(f(m_i, m_j) \times h)}{\mu^2}, \quad (2.16)$$

em que o numerador da equação (2.16), é a média condicionada a distância h entre dois pontos, ou seja, $h = |x_i - x_j|$ a condição é que há um processo pontual subjacente ambos em 0 e x_i , em que x_i é algum ponto fixado na distância h de origem 0. O denominador remove os efeitos de escalas dados por marcas de diferentes ordens, uma vez que, a função teste é um número aleatório. Segundo Wälder & Wälder (2008), a função teste $f(m_i, m_j)$, sempre conduz a valores positivos por meio da função de correlação marcada.

2.5.2 Estimando a função de correlação marcada

Conforme Gavrikov & Stoyan (1995), a função de correlação marcada apresentada na equação (2.16), deve ser estimada, porque a distribuição condicional das marcas não são conhecidas, ou seja, na prática sem fazer suposições não pode-se calcular $E(f(m_i, m_j) \times h)$ teoricamente. O valor estimado de (2.16), é dado

por:

$$\hat{\rho}_f(h) = \frac{\sum_{i=1}^n \sum_{j=1}^n f(m_i, m_j)}{n(h) - \mu^2}, \quad (2.17)$$

em que, $n(h)$ é o número de pares de pontos (x_i, x_j) com distância h . segundo Wälder & Wälder (2008), esse estimador apresenta irregularidades, que lembram um diagrama de colunas. Além disso, dificilmente obtem-se pares de pontos com a mesma distância h . Assim, recomenda-se usar algum tipo de suavizador para o estimador da função de correlação marcada.

Um suavizador bem conhecido na literatura e recomendado por Penttinen et al. (1992) é o estimador Epanechnikov de Kernel dado por:

$$e_\delta(h) = \begin{cases} \frac{3}{4\delta} \left(1 - \frac{h^2}{\delta^2}\right), & \text{se } -\delta \leq h \leq \delta \\ 0, & \text{c.c.} \end{cases},$$

em que, δ é um parâmetro adicional que define o grau de "suavidade" da função. Quanto maior este valor mais suave será a função obtida. A escolha de δ é sempre um problema ao usar o suavizador de Kernel. Nesse trabalho estaremos seguindo Penttinen et al. (1992) que sugere um valor próximo de $0,15 \times \hat{\lambda}^{-1/2}$.

Um outro problema com o estimador da função de correlação marcada é que o mesmo pode ser influenciado pelo efeito de bordas, uma vez que a área de observação é finita e, portanto, pontos fora desta área não são considerados. Assim, recomenda-se usar algum dos procedimentos descritos em Stoyan et al. (1995) e Illian et al. (2008) para corrigir o efeito de bordas.

Ripley (1977) propôs um dos corretores de bordas mais utilizados na literatura por meio de regiões retangulares, denominada de correção isotrópica que será utilizado neste trabalho. Nessa correção considera-se w_{ij} como sendo o inverso da

proporção dentro da área $W(A)$ de um círculo centrado em x_i e de raio de distância $|x_i - x_j|$ com borda passando por x_j . Assim, quanto mais afastado da borda estiver o ponto x_i menor será o efeito.

Usando o estimador Epanechnikov de Kernel, obtém-se um estimador não viesado por meio da função de correlação marcada que corrige o efeito de bordas (Stoyan et al., 1987), dado pela equação:

$$\hat{\rho}_f(h) = \frac{1}{\mu^2} \sum_{i=1}^n \sum_{j=1}^n f(m_i, m_j) \times A_{ij}, \quad (2.18)$$

em que:

$$A_{ij} = \frac{e_{\delta}(h - |s_i - s_j|)}{W(A_i \cap A_j)}.$$

Note que, os pontos s_1, \dots, s_n ficam situados dentro da região de estudos A . Observa-se que a origem desta região é dada por:

$(\min \{x_i/i = 1, \dots, n\}, \min \{y_i/i = 1, \dots, n\})$ com $s_i = (x_i, y_i); i = 1, \dots, n$.

Segundo Penttinen et al. (1992), o resultado gráfico da função de correlação marcada pode ser usada como método exploratório para verificar se existe correlação entre marcas e pontos.

Baddeley & Turner (2005), afirmam que $\rho_f(h)$ não é a "correlação" usualmente utilizada na Estatística, podendo levar a qualquer valor real não negativo. Se as marcas forem não correlacionadas tem-se $\hat{\rho}_f(h) = 1$. Isto ocorre quando o numerador da equação (2.18) não depende da distância entre dois pontos, significando que as marcas m_i nos pontos x_i são independentes e identicamente distribuídas. Se $\hat{\rho}_f(h) \neq 1$ pode-se ter duas situações de dependência entre marcas e pontos. A primeira situação ocorre quando por meio de uma particular distância h , a cada par de pontos, tem-se as menores marcas associadas a esses pontos (ou marcas

diferentes para o caso discreto). Neste caso obtém-se $\hat{\rho}_f(h) < 1$, o que caracteriza uma repulsão entre as marcas. A segunda situação ocorre quando $\hat{\rho}_f(h) > 1$, ou seja, por meio de uma particular distância h , a cada par de pontos, tem-se as maiores marcas associadas a esses pontos (ou marcas iguais para o caso discreto); conforme é apresentado na Figura 5.

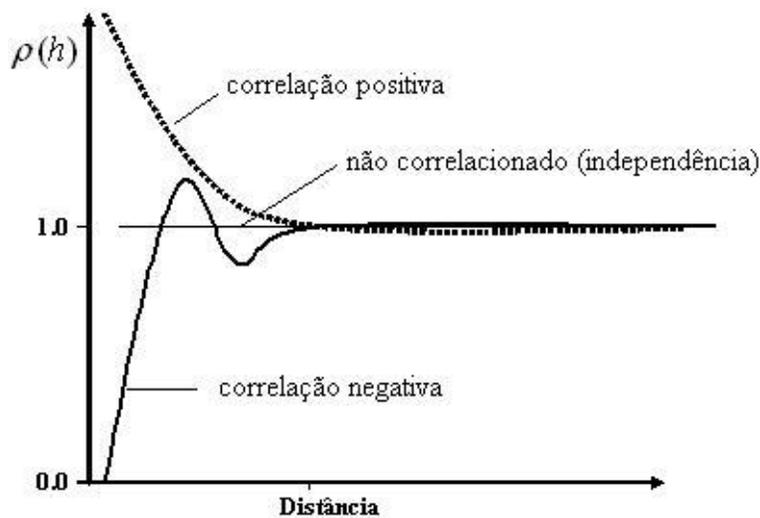


FIGURA 5: Idéia gráfica da função de correlação marcada.

Conforme Gavrikov & Stoyan (1995), a função de correlação marcada tem todos os seus valores entre zero e infinito, por meio de distâncias muito grandes $\rho_f(h)$ tende para 1 (um).

Conforme Schlather (2001), cada uma das várias medidas estatísticas mencionadas anteriormente, tem um valor esperado teórico por meio de certos tipos de pontos ordenados seguindo uma distribuição de Poisson homogênea.

Evidentemente, o problema com a análise exploratória gráfica descrita anteriormente é determinar qual o valor de $\hat{\rho}_f(h)$ para que se possa considerar que exista uma dependência entre marcas e pontos. Gavrikov & Stoyan (1995) e Illian et al. (2008) sugerem que este problema pode ser contornado usando-se computação in-

tensiva para construção de gráficos com envelopes que delimitam a hipótese nula. Entretanto, na literatura não encontra-se testes estatísticos formais para testar a hipótese nula de independência entre marcas e pontos. Esses métodos estatísticos são apresentados nas próximas seções

2.5.3 Teste de independência para função de correlação marcada

Illian et al. (2008) afirmam que, sob a hipótese de independência entre marcas e pontos, o objetivo básico é descobrir se existe uma correlação positiva ou uma correlação negativa entre as marcas e pontos. Esses mesmos autores sugerem que este objetivo básico pode ser alcançado de maneira análoga ao processo pontual não marcado. Entretanto, os autores não esclarecem de que forma esses procedimentos podem ser feitos.

Assim, nesse trabalho propõe-se métodos para testar a hipótese nula de independência entre marcas e pontos conforme serão descritos detalhadamente nas seções 3.4.1 e 3.4.2.

2.5.4 Processos pontuais marcados na literatura

A utilização de processos pontuais marcados, vem aumentando consideravelmente nos últimos anos. O crescente interesse na sua utilização é consequência das suas vantagens em relação aos métodos de processos pontuais não marcados. Ainda que, a maioria dos estudos tenham sido aplicados a análises florestais, novos artigos na literatura empregando estas funções nas áreas epidemiológicas e industriais, começam a tornar-se mais frequentes.

Goulard et al. (1995), utilizaram a função de correlação marcada, afim de estudar a dependência espacial entre os diâmetros em duas diferentes alturas do Castanheiro-Doce no Centro da França (região de Limousin), os autores consta-

taram uma correlação positiva entre as distâncias e uma relação de independência com os diâmetros ($2m$). Os autores salientaram a importância das condições de fertilidade do solo.

Nicotra (1998) investigou o padrão espacial e a estrutura sexual de *Siparuna Grandiflora* em uma floresta tropical da Costa Rica. O autor utilizou a função K bivariada, não encontrando evidências de inibição ou segregação entre os sexos, apenas padrões gerais agregados para a população como um todo.

Ogata (1998) e Schoenberg (2003) usaram métodos paramétricos a fim de detectar evidências contra a hipótese de independência entre marcas e pontos, em estudos sismológicos.

Schlather et al. (2004), propuseram as funções V e E e um teste formal de Monte Carlo, a fim de detectar a dependência entre marcas e pontos de uma classe estacionária de marcas Gaussianas.

Baddeley & Turner (2005), estudaram a dependência espacial dos diâmetros de 584 pinheiros Longleaf, na Georgia, Estados Unidos, encontrando uma correlação negativa em distâncias menores que 15 metros e um processo completamente aleatório nas distâncias acima de 15 metros.

Guan (2006) propôs duas novas funções, as funções K e G marcada, utilizando os dados de *Gus pearson*, em uma floresta de Pinheiros no Arizona, Estados Unidos, investigando a dinâmica de crescimento das árvores. Para ambas as funções, o referido autor desenvolveu testes gráficos bem como testes de aproximações formais, afirmando que a principal vantagem destas funções são as poucas especificações para os parâmetros, facilitando as estimativas para um conjunto de dados.

Van Lieshout (2006) propôs a função J marcada, estudando-se o padrão espacial da dependência entre 126 Pinheiros com seus respectivos diâmetros em uma

floresta na Finlândia. A autora detectou uma correlação positiva em distâncias de dois a quatro metros.

Mattfeldt et al. (2006), usaram a função de correlação marcada a fim de verificar a dependência espacial existente entre os diâmetros dos vasos capilares e o câncer de próstata, os autores concluíram que o fornecimento de oxigênio no tecido por meio dos vasos capilares inibe, consideravelmente, o aparecimento de câncer de próstata, desde que os diâmetros dos vasos capilares estejam distribuídos de forma dependente nas glândulas da próstata.

3 MATERIAL E MÉTODOS

3.1 Materiais

A análise estatística da distribuição espacial das partículas, foi realizada em três amostras metalográficas planas de um material compósito de partículas de cerâmicas incrustadas em uma base de alumínio, produzido pelo Departamento de Engenharia de Materiais, University of Sheffield, Reino Unido, onde as partículas apresentavam uma fração de volume igual a 11%.

As três amostras metalográficas foram extraídas utilizando microscopia óptica, que permitiu um ajustamento totalmente automático, o posicionamento e o scaneamento das amostras.

As imagens digitais bi-dimensionais foram analisadas pelo uso de processamento de imagens para extrair as coordenadas (centros) de cada partícula dentro das imagens. As imagens reais eram de 512×767 pixels. Maiores detalhes em Scalon et al. (2003).

A análise estatística dos marcadores genéticos de árvores, foram obtidos de uma população de *Copaifera langsdorffi* Desf., popularmente conhecida por copaíba ou óleo copaíba, localizada no campus da Universidade Federal de Lavras (UFLA), Lavras, MG. Trata-se de um fragmento florestal que possui uma área de 5,83 ha. Esse fragmento está localizado nas coordenadas $21^{\circ} 14'40''S$, e $44^{\circ} 57'50''W$, em uma altitude média de 925m.

Os dados foram analisados com base em folhas coletadas de 20 árvores de população de copaíba, na qual os indivíduos amostrados foram considerados adultos. Então, a amostra coletada foi submetida à caracterização genética da espécie.

Toda a análise laboratorial foi realizada no Laboratório de Melhoramento Florestal e Biologia Molecular do Departamento de Ciências Florestais da Universi-

dade Federal de Lavras. Apresentam-se os resultados referentes aos marcadores $\alpha - Est - 1$, $\alpha - Est - 2$, $Po1$ e $Po2$, os quais estão dispostos em Louzada (2003).

3.2 Análise exploratória

Realizou-se uma representação gráfica dos dados no sistema de coordenadas (x, y) , obtendo-se uma inspeção visual do comportamento espacial das partículas de cerâmica na base de alumínio, utilizando-se o estimador de intensidade de Kernel descrito por Bailey & Gatrell (1995). Realizou-se uma inspeção gráfica verificando-se o padrão espacial dos diâmetros das partículas, que estão associados a cada ponto amostral. Realizou-se estas análises com objetivo de verificar a caracterização qualitativa e qual tipo de padrão espacial (aleatório, regular e agrupado) os dados apresentam. Para os dados de marcadores genéticos, realizou-se apenas uma representação gráfica dos dados no sistema de coordenadas (x, y) , utilizando-se o estimador de intensidade de Kernel descrito por Bailey & Gatrell (1995).

Construíram-se gráficos de histograma, para os diâmetros das partículas, por ser bastante completo no sentido de informar, a variabilidade do fenômeno espacial. Foram também encontrados os primeiros quatro momentos amostrais centrado na média dos diâmetros nos n pontos amostrais, e as quantidades derivadas deles apresentadas em tabelas, com o objetivo da realização do teste de assimetria e curtose assintótico.

Os momentos amostrais foram encontrados com o uso da fórmula apresentada por Mood et al. (1974):

$$m_r = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{n},$$

em que n é o número de diâmetros existente em cada amostra, $(X_i - \bar{X})$ é o desvio de cada estimativa em relação à média e m_r o r -ésimo momento amostral. Depois de encontrarmos os momentos amostrais foram calculadas as quantidades de interesse $g_1 = m_3/m_2^{3/2}$ para avaliar a simetria e $g_2 = (m_4/m_2^2) - 3$ para avaliar o excesso de curtose. Testes foram realizados tendo em vista que $g_1 \sim N(0, (6/n)^{1/2})$ e $g_2 \sim N(0, (24/n)^{1/2})$, conforme Snedecor & Cochran (1980).

Em razão do fato dos marcadores genéticos da copaíba possuírem marcas categóricas, não realizou-se estas análises.

3.3 Análise exploratória espacial

Para os dados referentes ao material compósito de partículas de cerâmica incrustadas na base de alumínio, utilizou-se as funções F , G e J estabelecendo-se as distâncias máximas e mínimas compreendidas entre 0 e 20 pixels, levando-se em consideração os estimadores não viesados, que corrigem para o efeito de bordas definidos nas equações (2.7), (2.10), (2.11) respectivamente. Para as funções G e F estabeleceu-se a escala máxima e mínima no eixo y de 0 a 1, e a escala máxima e mínima de 0,6 a 1,4 por meio da função J .

Utilizando-se a função K , estabeleceu-se a distância mínima de 0 pixels e a distância máxima de 20 pixels, levando-se em consideração o estimador não viesado (Ripley, 1977) dado pela equação (2.14), que corrige para o efeito de bordas. Para uma melhor visualização das características de CSR , utilizou-se a forma transformada da função K , dada pela equação (2.15).

Diggle (2003) afirma que as distâncias de entrada compreendidas entre um valor mínimo e máximo utilizadas em cada função, são especificadas pelo pesquisador, podendo ser no máximo 25% da região de estudos.

Comparou-se a distribuição dos eventos observados criando-se envelopes de

simulação por meio da *CSR*, realizando-se 1000 simulações Monte Carlo. Calculando-se os envelopes superior e inferior e a construção do teste contra a hipótese de *CSR* conforme descrito em Scalon et al. (2003), onde o intervalo de integração são as distâncias máximas e mínimas definidas em cada função.

Diggle (2003) usou 100 simulações Monte Carlo para acessar a significância dos desvios em relação a hipótese de *CSR*, no presente trabalho, utilizou-se 1000 simulações devido às facilidades computacionais.

Para os dados referentes às árvores de copaíba, utilizando-se as funções F , G e J estabeleceu-se distâncias máximas e mínimas compreendidas entre 0 e 20 metros, levando-se em consideração os estimadores não viesados, que corrigem para o efeito de bordas definidos nas equações (2.7), (2.10), (2.11) respectivamente. Para as funções G e F estabeleceu-se a escala máxima e mínima no eixo y de 0 a 0,8, e a escala, máxima e mínima de 0 a 2 por meio da função J .

Utilizando-se a função K , estabeleceu-se a distância mínima de 0 metros e a distância máxima de 20 metros, levando-se em consideração o estimador não viesado (Ripley, 1977) dado pela equação (2.15), que corrige o efeito de bordas. Afim de uma melhor visualização das características de *CSR*, utilizou-se a forma transformada da função K , dada pela equação (2.16).

De modo análogo aos dados referentes ao material compósito realizou-se a construção dos envelopes de simulação e o teste contra hipótese de *CSR* para os dados das árvores de copaíba.

3.4 Função de correlação marcada

Para os dados das partículas incrustadas na base de alumínio, decorrente do fato das marcas serem contínuas escolheu-se a função teste dada por:

$$f(m_i, m_j) = m_i \times m_j.$$

Para os dados dos marcadores genéticos, onde suas respectivas marcas são categóricas utilizou-se a função teste dada por:

$$f(m_i, m_j) = I\{m_i = m_j\}.$$

Para a função de correlação marcada utilizou-se o estimador dado na equação (2.18), que corrige o efeito de bordas (Stoyan et al., 1987).

3.4.1 Teste de hipótese para independência entre marcas e pontos

Para construção do teste contra a hipótese de independência entre marcas e pontos sugere-se o seguinte procedimento:

i) Define-se u_i como sendo a medida da discrepância entre $\hat{\rho}_i(h)$ e $\rho(h)$ sobre toda distância h , dada por:

$$u_i = \int_0^{h_0} \{\hat{\rho}_i(h) - \rho(h)\}^2 dh; \quad (3.1)$$

em que $\hat{\rho}_i(h)$ é a função de correlação marcada estimada, $\rho(h)$ é a função de correlação marcada teórica, sob a hipótese de independência entre marcas e pontos, e o intervalo de integração ($0 \leq h \leq h_0$), é a distância máxima e mínima especificada pelo pesquisador. Observa-se que pode-se utilizar outras medidas de discrepância,

como por exemplo as medidas de Kolmogorov-Smirnov e Pearson entre outras.

ii) Calcula-se u_0 , em que u_0 é a medida para o processo pontual marcado observado.

iii) Calcula-se $u_i : i = 1, 2, \dots, s$, em que s é o total de simulações e u_i é a medida para os processos simulados sob a hipótese de independência entre marcas e pontos.

iv) Ordena-se u_i .

v) Se obtivermos u_0 em um dos extremos da distribuição dos $u_i : i = 1, 2, \dots, s$, então existem evidências estatísticas para rejeitar a hipótese de independência entre marcas e pontos. Suponha que, $u_{(1)} = u_{(j)}$ para algum $j \in \{1, \dots, s\}$, então rejeita-se a hipótese de independência se, $p\text{-valor} = ((s + 1 - j)/s) \leq \alpha$, em que α é o nível de significância do teste.

Para simular configurações de processos pontuais marcados sob a hipótese de independência entre marcas e pontos, Illian et al. (2008) sugere que pode-se utilizar alguma forma de reamostragem. Nesse trabalho, as localizações dos eventos permanecem fixas obedecendo-se as posições da configuração observada e as marcas são permutadas aleatoriamente nessas posições. Esse procedimento de simulação, denominado de *random labelling* e utilizado por Diggle (2003), para simular processos pontuais bivariados.

3.4.2 Análise gráfica para testar independência entre marcas e pontos

O teste de hipótese proposto anteriormente não indica se a rejeição da hipótese nula a favor da dependência positiva ou negativa entre marcas e pontos. Para determinar a direção da rejeição da hipótese nula, sugere-se a construção de envelopes de simulação. Para tal, propõem-se o seguinte procedimento:

i) Calcula-se $\hat{\rho}_1(h)$ a função de correlação marcada para uma configuração do

processo pontual marcado observado.

ii) Calcula-se a função de correlação marcada $\hat{\rho}_i(h)$: $i = 2, 3, \dots, s$, para cada $(s - 1)$ configuração, sob a hipótese de independência entre marcas e pontos.

iii) Calcula-se os envelopes de simulação superiores e inferiores como sendo:

$$\begin{aligned} U(h) &= \max \{ \hat{\rho}_i(h) \}, i = 1, \dots, s. \\ L(h) &= \min \{ \hat{\rho}_i(h) \}, i = 1, \dots, s. \end{aligned} \quad (3.2)$$

iv) Coloca-se em um gráfico os envelopes de simulações e $\hat{\rho}_1(h)$ contra a distância h . Esse gráfico apresenta a propriedade:

$$P \{ \hat{\rho}_1(h) > U(h) \} = P \{ \hat{\rho}_1(h) < L(h) \} = s^{-1}.$$

v) Interpreta-se o gráfico da seguinte maneira: se o gráfico de $\hat{\rho}_1(h)$ estiver acima do limite superior $U(h)$, então rejeita-se a hipótese de independência entre marcas em pontos em favor da correlação positiva. Se o gráfico de $\hat{\rho}_1(h)$ estiver abaixo do limite inferior $L(h)$, então rejeita-se a hipótese de independência entre marcas em pontos em favor da correlação negativa.

Para simular configurações de processos pontuais marcados sob a hipótese de independência entre marcas e pontos utiliza-se o mesmo procedimento utilizado na seção 3.4.1.

De acordo com as metodologias abordadas nesta seção, por meio da função de correlação marcada, realizou-se 1000 simulações Monte Carlo, adotando-se a distância máxima e mínima de 0 a 200 pixels para os dados do material compósito na base de alumínio. Adotou-se a distância máxima e mínima de 0 a 80 metros para os dados dos marcadores genéticos.

No eixo y fixou-se a escala de 0 a 1,5 para os diâmetros das partículas e de 0 a 2,5 para os marcadores genéticos. Calculando-se o teste de hipótese nula de

independência entre marcas e pontos, por meio da equação (3.1), variando-se o intervalo de integração para os marcadores genéticos de 0 a 80 metros e para as partículas de cerâmica de 0 a 200 pixels . Assumindo-se que a função de distribuição teórica $\rho(h)$, para independência entre marcas e pontos é igual a 1 (um), ou seja $\rho_f(h) \equiv 1$, significando que o processo pontual marcado é completamente aleatório.

3.5 Software

As análises utilizando as funções F , G , K , J e a função de correlação marcada e os respectivos envelopes de simulação (intervalo de confiança) foram realizados utilizando-se o software R (R Development Core Team, 2008), utilizando o pacote spatstat, Baddeley & Turner (2005).

4 RESULTADOS E DISCUSSÃO

4.1 Análise do material compósito

4.1.1 Análise exploratória

Inicia-se a análise exploratória do material compósito com algumas análises descritivas simples. As representações gráficas dos diâmetros das partículas de cerâmica na matriz de alumínio estão apresentados na Figura 6, onde observa-se que as configurações das três amostras metalográficas apresentam diferentes formas e tamanhos. Todas as medidas são dadas em pixels, em que um pixel equivale a $0,375\mu m$ e um μm é igual a $10^{-6}m$ (metros). Visualmente não percebe-se se partículas de mesmo tamanho estão mais próximas, ou seja, se há alguma evidência da dependência espacial entre as marcas e os pontos. Pode-se observar ainda que as três amostras apresentam uma distribuição assimétrica à direita conforme mostra os histogramas da Figura 7. Verifica-se o mesmo resultado utilizando-se o teste de assimetria e do excesso de curtose apresentado na Tabela 1, ou seja, os dados estão sendo influenciados pela não normalidade (Snedecor & Cochran, 1980). Observa-se também que as médias dos diâmetros nas três amostras apresentam-se bem próximos, pode ser, devido ao fato das amostras serem coletadas em partes distintas do mesmo material. Sendo assim, o engenheiro de materiais poderia verificar se a fabricação do material compósito é homogêneo. Deve-se salientar, porém, que os métodos utilizados na estatística espacial não exigem normalidade dos dados para serem aplicados (Ripley, 1981).

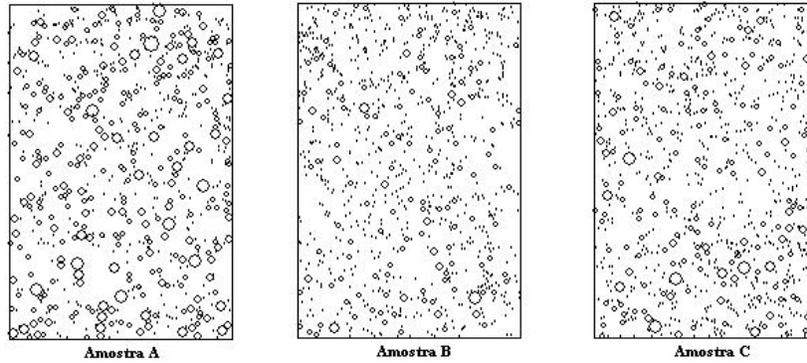


FIGURA 6: Distribuição das 730 partículas de cerâmica na matriz de alumínio na amostra A; distribuição das 937 partículas de cerâmica na matriz de alumínio na amostra B; distribuição das 1234 partículas de cerâmica na matriz de alumínio na amostra C. Os círculos de cada partícula são proporcionais aos tamanhos das partículas. Área das amostras: $512 \times 767 \text{ pixels}^2$.

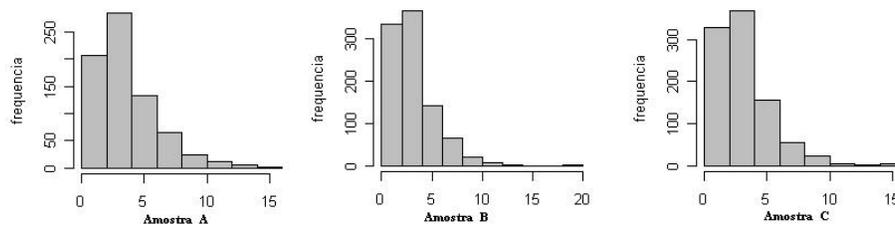


FIGURA 7: Histogramas referente aos diâmetros das partículas por meio das amostras A, B e C.

TABELA 1: Estatística descritiva da variável diâmetro.

Amostras	Média	Variância	máx	min	g_1	g_2
A	3,65	5,23	14,14	1,38	1,46	5,21
B	3,21	4,09	18,26	1,38	1,83	8,25
C	3,28	4,52	15,94	1,38	1,99	8,67

4.1.2 Análise exploratória espacial

O primeiro passo da análise espacial é verificar as propriedades de primeira ordem. Para tal, analisa-se a intensidade pontual das partículas na base de alumínio por meio do estimador de Kernel apresentado na seção 2.4. Analisando-se a Figura 8, percebe-se que as amostras A, B e C apresentam áreas onde a concentração de partículas é considerável, formando-se pequenos agrupamentos, todavia, visualmente é difícil perceber se as amostras apresentam algum tipo de padrão sistemático (agrupamento ou regularidade). De acordo com Ripley (1981), esta análise é apenas um passo preliminar para entender que tipo de padrão espacial seguem-se os dados. Esse fato só poderá ser consumado após a investigação por meio das distribuições empíricas mencionadas nesse texto.

Continuando a análise espacial investiga-se a hipótese de *CSR* das partículas empregando-se as funções F , G , J e K descritas na seção 2.4. Investigando-se a relação espacial das partículas na base de alumínio empregando-se a função G , Percebe-se nas amostras A, B e C apresentadas na Figura 9, a ausência de partículas nas distâncias menores que 4 pixels, esta situação aponta para uma distribuição mais regular, que está de acordo com Scaloni et al. (2003). Para distâncias acima de 5 pixels, a função de distribuição estimada encontra-se dentro do envelope de simulação. Utilizando-se a função F Por meio da Figura 10, percebe-se que a função estimada encontra-se bem próxima da função teórica nas amostras A e C. Verifica-se que na amostra B para distâncias acima de 10 pixels a função estimada encontra-se abaixo da função teórica. Esta situação aponta para uma distribuição mais regular. Utilizando-se o estimador não corrigido dado pela função J , observa-se que, para distâncias menores que 5 pixels nas amostras A, B e C, apresentadas na Figura 11, a função estimada encontra-se acima da função teórica. Verificando-se uma tendência a regularidade.

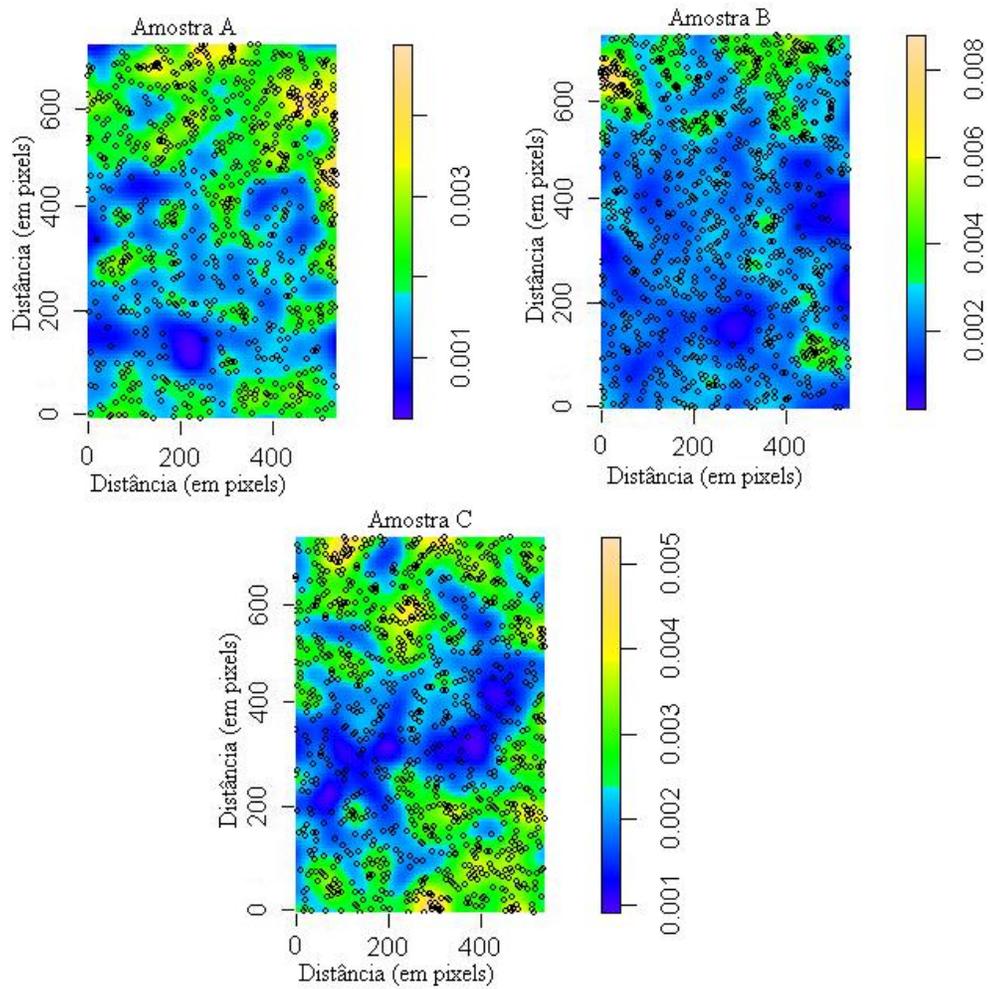


FIGURA 8: Distribuição das partículas na base de alumínio por meio do suavizador de Kernel, em que a escala (lado direito) representa a intensidade de pontos por unidade de área em toda a região de estudo.

Observa-se que a função de distribuição estimada apresenta valores fora do envelope de simulação nas amostras B e C, para distâncias acima de 10 pixels.

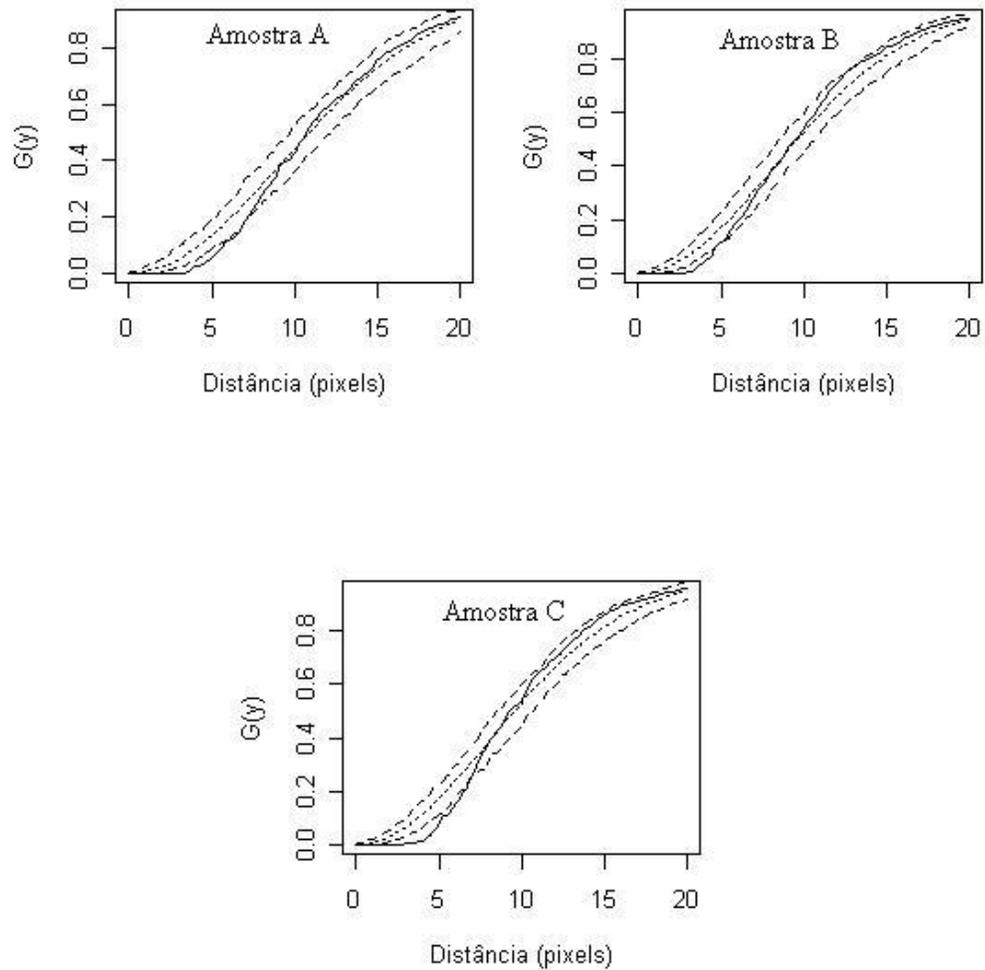


FIGURA 9: Envelope de simulação por meio da função G referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).

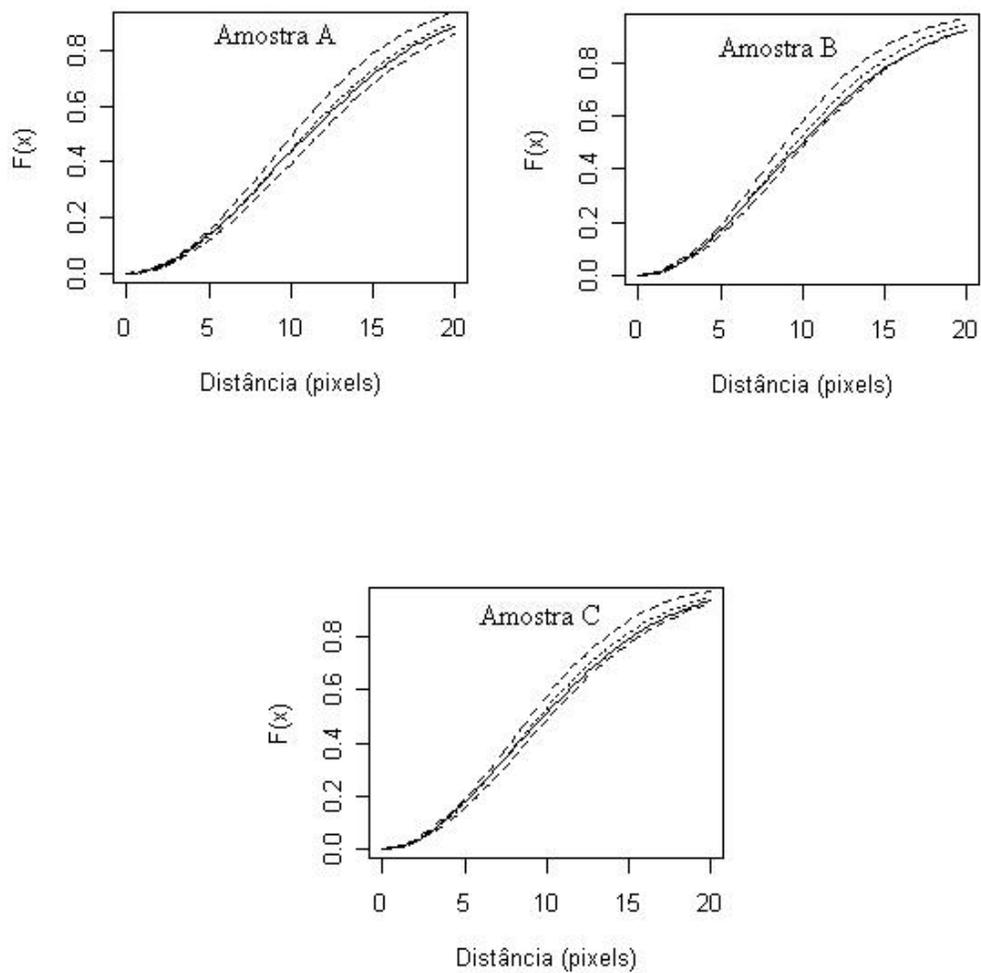


FIGURA 10: Envelope de simulação por meio da função F referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).

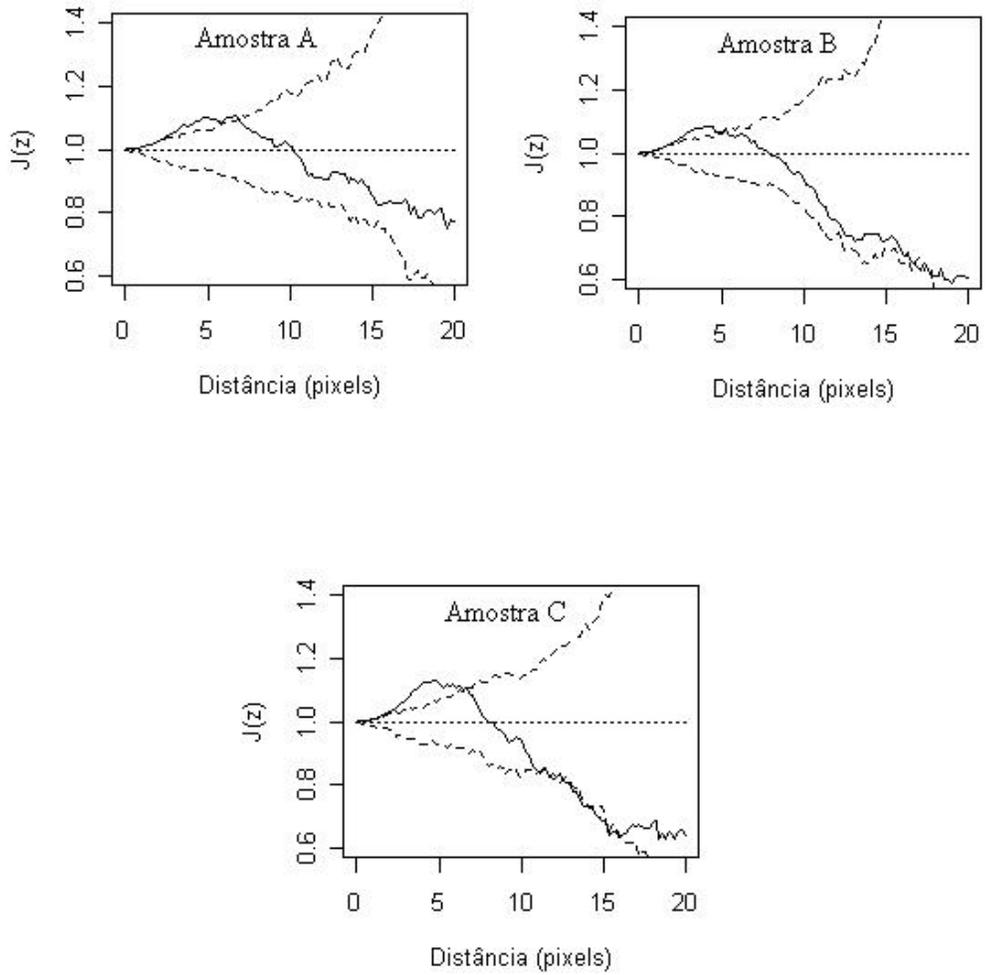


FIGURA 11: Envelope de simulação por meio da função J referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).

As funções F , G e J são métodos baseados em distâncias ou área, que são os mais comuns na literatura, e por conseqüência, carregam em sua essência simplificações da realidade, uma vez que tratam apenas da simples contagem de pontos dentro de uma região de estudo ou da procura pelo primeiro vizinho mais próximo da mesma espécie (Ripley, 1978). Desse modo, muita informação é perdida nas análises espaciais, uma vez que tais métodos só conseguem detectar o padrão espacial naquela escala particular em que os dados foram coletados, referindo-se à distância entre os pontos (Schlather, 2001). Assim, uma função descritiva espacial que leva-se em consideração uma extensa seqüência de escalas é a função K (Diggle, 2003).

Na Figura 12, pode-se observar que existe uma ausência de partículas dentro de uma distância aproximada de até 5 pixels nas amostras A, B e C. Esta ausência de partículas caracteriza-se uma distribuição mais regular das partículas até essa distância. Nas distâncias acima de 5 pixels, a função K estimada está contida no envelope de simulação. Apenas na amostra B, na distância de aproximadamente 10 a 20 pixels, a função estimada encontra-se acima do limite superior do envelope. Portanto, pode-se considerar que as partículas apresentam distribuição aleatória nas três amostras a partir de uma distância de 5 pixels. A regularidade da distribuição das partículas observadas por meio de pequenas distâncias pode ser devido ao fato de termos utilizado apenas o centro das partículas para fazer a análise (Scalon et al., 2003). Evidentemente, duas partículas não podem ocupar a mesma posição, o que poderia explicar a regularidade. Assim, a função K representa uma ferramenta natural e valiosa para a descrição de um processo espacial uma vez que fornece um resumo eficaz da dependência espacial sobre uma extensa seqüência de escalas. Sendo assim, a função K é igualmente poderosa para detectar os dois padrões em diferentes escalas (Diggle, 2003).

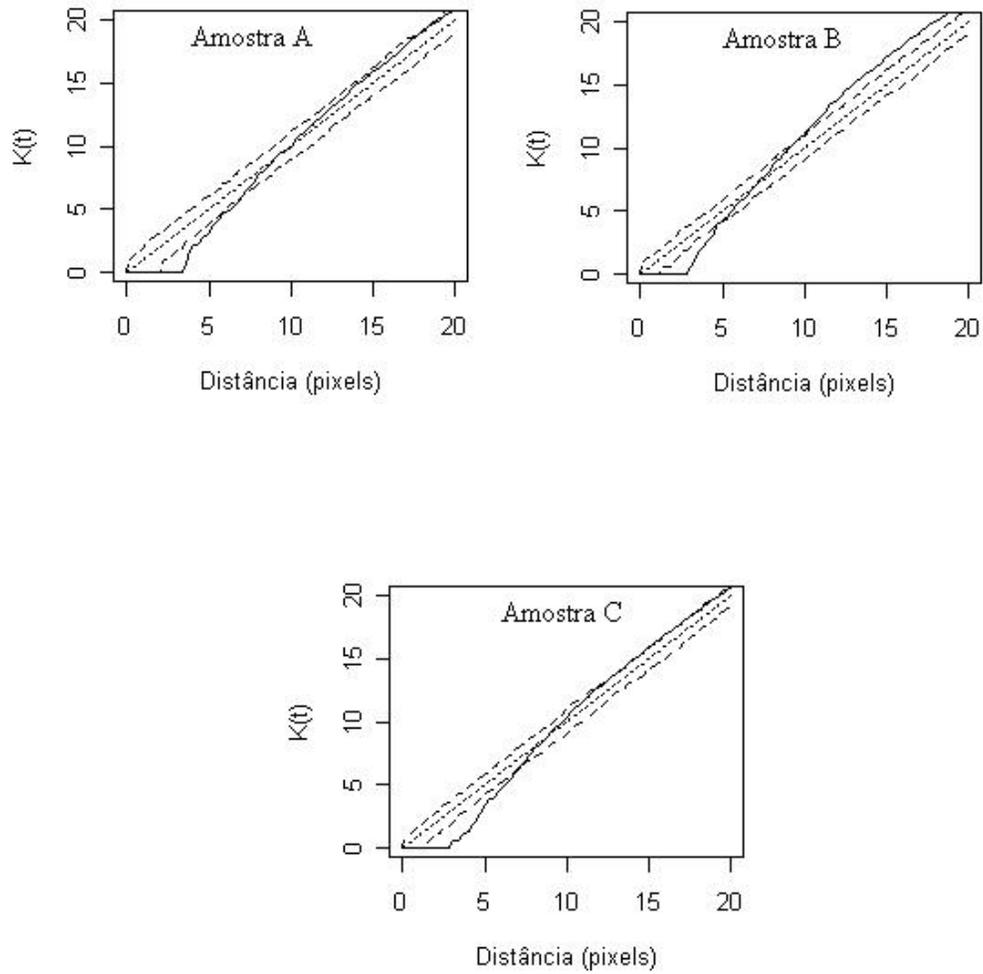


FIGURA 12: Envelope de simulação por meio da função K referente as amostras: A, B e C; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).

4.1.3 Função de correlação marcada

Iniciando-se a análise da dependência espacial existente entre as partículas do material compósito e os diâmetros, aplicando-se o teste estatístico formal da seção (3.4.1) afim de testar a hipótese nula de independência entre as localizações e os tamanhos das partículas. Utilizando-se a função teste para dados contínuos na função de correlação marcada, obteve-se ($u_1 = 4,22$; $p = 0,001$) para a amostra A, ($u_1 = 3,59$; $p = 0,001$) para amostra B e ($u_1 = 4,15$; $p = 0,001$) para a amostra C, indicando fortes evidências estatísticas para rejeitar a hipótese nula de independência entre marcas e pontos, ou seja, há uma dependência espacial entre os diâmetros das partículas com seus respectivos pontos na matriz de alumínio.

O teste de hipótese não indica se a rejeição da hipótese nula é a favor da atração ou repulsão das marcas entre os pontos. Para determinar a direção da rejeição da hipótese nula, foram construídos os envelopes de simulação descritos na seção (3.4.2). Utilizando-se os mesmos parâmetros do teste formal foram obtidos os gráficos apresentados na Figura 13.

Observa-se por meio da Figura 13, que as funções observadas ficaram abaixo do limite inferior do envelope de simulação por meio de distâncias, aproximadamente, menores que 25 pixels nas três amostras. Para distâncias acima de 25 pixels a função estimada esta contida no envelope de simulação. Assim, até a distância de 25 pixels observa-se repulsão entre as localizações e os tamanhos das partículas, enquanto para distâncias acima de 25 pixels, existe independência. Isto significa o fato óbvio que somente ocorrerão partículas de cerâmica muito próximas se seus diâmetros forem pequenos.

Sobre o ponto de vista da Engenharia de Materiais esses resultados indicam que pode-se até estar ocorrendo alguns agrupamentos de pequenas partículas, mas estas estão sempre distribuídas entre as grandes partículas. Em outras palavras, na formação do material compósito, as pequenas partículas estão ocupando os espaços deixados pelas grandes partículas.

Os resultados obtidos nesse trabalho corroboram com os resultados obtidos por Stoyan & Wincek (1991) e Scalon (1998). Stoyan & Wincek (1991) utilizaram a função de correlação marcada e detectaram o mesmo comportamento na distribuição espacial de partículas de prata em um substrato de ferro. Os autores fizeram uma análise exploratória subjetiva do gráfico da função de correlação marcada afim de chegar seus resultados. Scalon (1998) analisou o mesmo material utilizado nesse trabalho e teve que classificar as partículas em dois tamanhos diferentes para utilizar a função K -bivariada descrita em Diggle (2003). O autor utilizou o logaritmo da média do tamanho das partículas como ponto de corte para classificar as partículas. Assim, partículas com tamanho acima da média foram classificadas como grandes, caso contrário, de pequenas. A grande vantagem dos métodos apresentados nesse trabalho sobre a função K -bivariada é que não há necessidade de fazer uma classificação arbitrária de cada partícula. Conforme Scalon (1998) observou, a análise de interação entre localização e o tamanho de partículas usando a função K -bivariada em marcas contínuas, categorizadas artificialmente, é dependente do ponto corte.

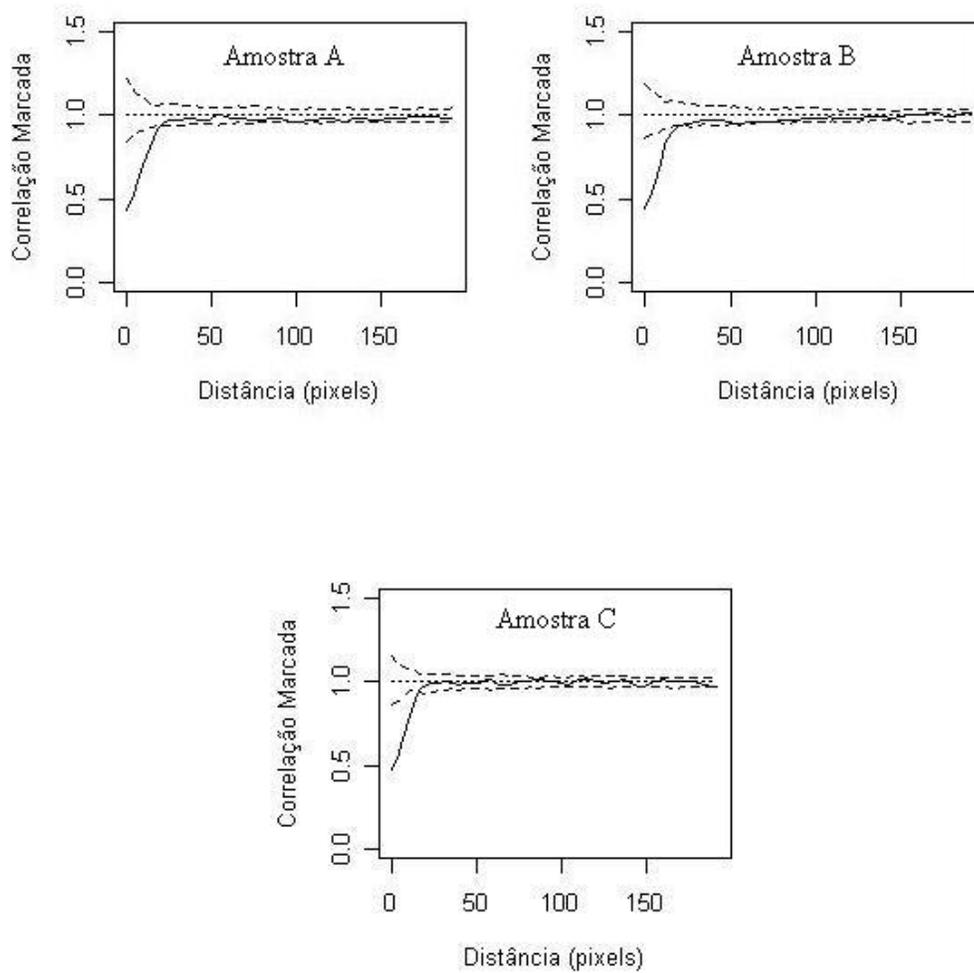


FIGURA 13: Envelope de simulação para os diâmetros das partículas referente as amostras: A, B e C, por meio da função de correlação marcada; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).

4.2 Análise dos marcadores genéticos

4.2.1 Análise exploratória espacial

O primeiro passo da análise espacial foi verificar as propriedades de primeira ordem. Para tal, analisa-se a intensidade pontual das árvores por meio do estimador de Kernel apresentado na seção 2.4. Dessa forma, observa-se na Figura 14, que apesar de ter poucos pontos amostrais, percebe-se que a região apresenta uma concentração considerável de árvores. Em áreas de florestas nativas há sempre uma seleção natural das espécies, formando agrupamentos por meio de um conjunto de árvores que têm a mesma essência (Goulard et al., 1995). Esta dependência espacial pode ocorrer devido à competição de luz e nutrientes entre árvores vizinhas da mesma espécie (Guan, 2006). Todavia, visualmente é difícil perceber se há algum tipo de padrão sistemático (Câmara et al., 2002). De acordo com Ripley (1981), isso é apenas um passo preliminar para entender que tipo de padrão espacial apresentam-se os dados.

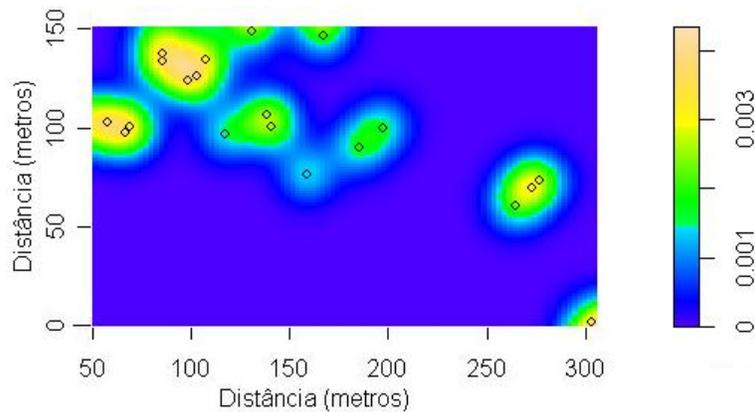


FIGURA 14: Distribuição das árvores na região de estudos por meio do suavizador Kernel, em que a escala (lado direito) representa a intensidade de pontos por unidade de área em toda a região de estudo.

Continuando a análise espacial investigando-se a relação espacial das árvores de copaíba distribuídas na região de estudo, utilizando-se o método do vizinho mais próximo, a função G . Conforme é apresentado na Figura 15, observa-se uma tendência a regularidade por meio de distâncias menores que 4 metros. Verifica-se que, distâncias acima de 5 metros a função de distribuição estimada encontra-se fora do envelope de simulação, portanto há indícios para se rejeitar a hipótese de CSR , em favor do agrupamento das árvores de copaíba.

Levando-se em consideração a função de distribuição acumulada da distância entre um ponto aleatoriamente escolhido e o evento mais próximo função F , conforme é apresentado na Figura 15. Percebe-se que a função estimada esta contida no envelope de simulação, portanto não há indícios para se rejeitar a hipótese de CSR , o resultado obtido por meio desses dados utilizando-se a função F , vai contra o que seria esperado por Diggle (2003), onde o autor afirmou que a função F apresenta grande poder para detectar agrupamentos de eventos.

Verifica-se, por meio da função J , que a função estimada encontra-se fora do envelope de simulação. Esta situação indica a existência de agrupamento para distâncias acima de 5 metros. Percebe-se que a função J não detectou nenhum padrão sistemático nas distâncias compreendidas entre 0 e 5 metros.

Utilizando-se a função K , percebe-se que a função estimada encontra-se abaixo da função teórica, nas distâncias de 0 a aproximadamente 4,5 metros, indicando um padrão de regularidade. Todavia, distâncias acima de 5 metros a função estimada esta contida no envelope de simulação, no entanto, encontra-se acima da função teórica, prescrevendo um padrão de agrupamento.

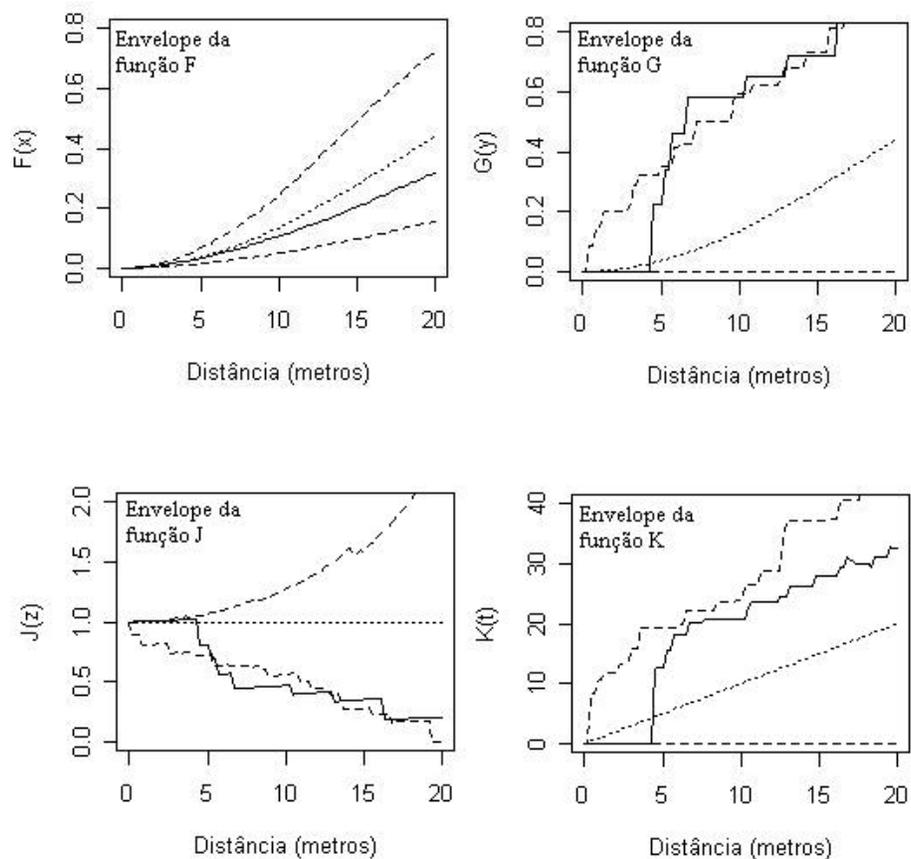


FIGURA 15: Envelope de simulação das funções: F , G , J , K ; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).

4.2.2 Função de correlação marcada

A análise espacial das características genéticas das árvores foi conduzida, inicialmente, aplicando o teste estatístico formal da seção (3.4.1) para testar a hipótese nula de independência entre a localização da árvore e a característica genética. Utilizando-se a função teste para dados categóricos na função de correlação marcada, obteve-se um ($u_1 = 7,80$, $p = 0,805$) para a característica genética do loco $\alpha - Est - 1$, ($u_1 = 19,51$, $p = 0,677$) para a característica genética do loco $\alpha - Est - 2$, ($u_1 = 13,67$, $p = 0,500$) para a característica do loco $Po1$ e ($u_1 = 12,53$, $p = 0,495$) para a característica do loco $Po2$. Estes resultados indicam que a evidências para não se rejeitar a hipótese nula de independência nos quatro locos e, portanto, detectando uma independência entre a localização da árvore e sua característica genética. Esses resultados podem ser visualizados na Figura 16 que mostra as funções de correlação marcada observada dentro dos limites do envelope de simulação para todo o intervalo de 0 a 80 metros.

O resultado obtido nesse trabalho vai contra o que seria esperado do ponto de vista da genética de populações, uma vez que, raramente ocorre independência entre a posição das árvores e suas características genéticas (Sokal & Jacquez, 1991). Entretanto, esse resultado pode ser explicado por características típicas de marcadores genéticos que poderiam ter contaminado os dados, tais como: distribuição geográfica altamente irregular, mensurações inacuradas, áreas mal definidas e outliers (Piazza et al., 1981). Entretanto, observa-se que o resultado obtido no presente trabalho corrobora o resultado obtido por Louzada (2003) que utilizou o índice de correlação de Moran para análise espacial do mesmo conjunto de dados.

Deve-se ressaltar que o índice de Moran foi desenvolvido para análise de outro tipo de dado espacial que é a contagem por área, ou *lattice* (Anselin, 2003). Nesse tipo de dado espacial, os mesmos são agregados em unidades de análise, usual-

mente delimitadas por polígonos fechados. A soma das áreas de todos os polígonos fornece a área total da análise. Parece claro que o uso de técnicas por meio de processos pontuais marcados é mais apropriado para os dados das características genéticas em árvores do que o índice de Moran.

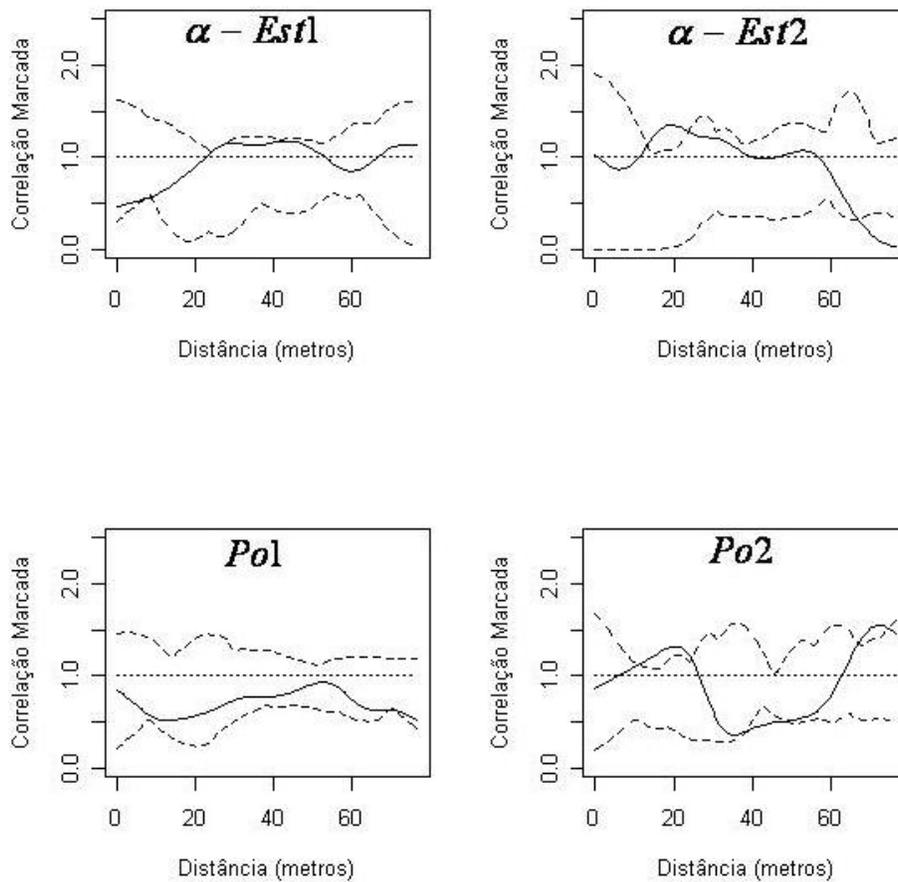


FIGURA 16: Envelope de simulação da função de correlação marcada referente aos marcadores genéticos: $\alpha - Est1$, $\alpha - Est2$, $Po1$ e $Po2$; limite superior (linha tracejada); limite inferior (linha tracejada); função teórica (linha pontilhada); função estimada (linha contínua).

4.2.3 Considerações finais

O estudo mostrou que a função de correlação marcada parece ser uma simples porém efetiva ferramenta para análise desse tipo de dado. Esta função compartilha do mesmo problema das funções V e E (Schlather et al., 2004) que é requerer um suavizador por meio de suas estimativas. Acredita-se que o uso incorreto de um parâmetro de suavização venha a influenciar os resultados das análises e, portanto, metodologias que não requerem este tipo de artifício como as funções J marcada (Van Lieshout, 2006) e G e K com marcas (Guan, 2006) poderiam ser mais apropriadas. Entretanto, este problema pode ser facilmente contornado com a escolha de um parâmetro suavizador conforme proposto por Penttinen et al. (1992). O ideal seria fazer uma comparação desses procedimentos utilizando-se simulações em diversas situações experimentais. Não encontra-se na literatura trabalhos com estas características. Outra vantagem da função de correlação marcada é a sua disponibilidade no pacote Spatstat (Baddeley & Turner, 2005) do software R, o que torna simples a implementação dos métodos apresentados. Os métodos gráficos e testes formais, propostos para análise de independência entre marcas e pontos são baseados nos métodos desenvolvidos por Ripley (1977) e disseminados por Diggle (2003) para a análise de processos pontuais não marcados. A novidade está em utilizar esses métodos, de maneira explícita, na função de correlação marcada. Por meio de simulação (não apresentada nesse trabalho) pode-se verificar que os métodos levam a aceitação da hipótese nula em mais de 99% dos casos.

Espera-se que este trabalho possa contribuir para a disseminação do uso dos métodos de estatísticas espaciais descritivas, para as funções empíricas F , G , J e K para testar a hipótese de CSR em processos pontuais e a função de correlação marcada para a análise de dados espaciais provenientes de processos pontuais marcados.

5 CONCLUSÃO

A partir dos métodos apresentados nesta dissertação chega-se as seguintes conclusões:

Conclui-se por meio dos resultados obtidos que as metodologias baseadas nos métodos de Monte Carlo e na função de correlação marcada podem ser poderosas ferramentas para a análise de processos pontuais marcados.

A hipótese de independência entre a localização e os diâmetros das partículas foi rejeitada. A hipótese de independência entre as árvores e os marcadores genéticos não foi rejeitada.

A hipótese de *CSR* da distribuição das partículas foi rejeitada em favor da hipótese de regularidade. A hipótese de *CSR* das árvores de copaíba foi rejeitada em favor da hipótese de agrupamento.

REFERÊNCIAS BIBLIOGRÁFICAS

ANJOS, A.; MAZZA, M. C. M.; SANTOS, A. C. M. C.; DELFINI, L. T. Análise do padrão de distribuição espacial da araucária (*Araucaria angustifolia*) em algumas áreas do Estado do Paraná, utilizando a função K de Ripley. **Scientia Forestalis**, Piracicaba, v. 66, p. 38-45, dez. 2004.

ANSELIN, L. **GeoDa 0.9 user's guide**. Urbana-Champaign: Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, 2003. 115p.

AVELAR, M. B. L. **Análise da agregação espacial do bicho-mineiro do cafeeiro (*Leucoptera coffeella*) (Guérin Mèneville & Perrottet, 1842) (*Lepidoptera: Lyonetiidae*) em lavoura cafeeira (*coffea arábica* L. orgânica em formação.)**. 2008. 66p. Dissertação (Mestrado em estatística e experimentação Agropecuárias) - Universidade Federal de Lavras, MG.

BADDELEY A. J.; TURNER R. spatstat: an R Package for Analyzing Spatial Point Patterns. **Journal of Statistical Software**, New York, v. 12, n. 6, Jun. 2005.

BADDELEY, A. J.; KERSCHER, M.; SCHLADITZ, K.; SCOTT, B. T. Estimating the *J* function without edge correction. **Statistica Neerlandica**, Oxford, v. 54, n. 3, p. 315-328, Nov. 2000.

BEDFORD, T.; VAN DEN BERG, J. A remark on the Van Lieshout and Baddeley J-function for point processes. **Advances in Applied Probability**, Sheffield, v. 29, n. 1, p. 19-25, Jan. 1997.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific. New York: Imperial College, 1995. 409p.

CÂMARA, G.; MONTEIRO, A.M.; FUCKS S. D.; CARVALHO M. S. **Introdução à ciência da geoinformação**. São José dos Campos: INPE, 2002. 186p.

CAPRETZ, R. L. **Análise dos padrões espaciais de árvores em quatro formações florestais do estado de São Paulo, através de análises de segunda ordem, como a função K de Ripley**. 2004. 93p. Dissertação (Mestrado) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba.

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

DIGGLE, P. J. **Statistical analysis of spatial point patterns**. London: Arnold, 2003. 153p.

FISHER, R.A **The Genetical Theory of Natural Selection**. Oxfor: Oxford University, 1930. Chap. 6.

FORTIN, M. J.; DALE, M. **Spatial analysis: a guide for ecologists**. Cambridge: Cambridge University Press, 2005. 365p.

GATRELL, A. C.; BAILEY, T. C.; DIGGLE, P. J.; ROWLINGSON, B. S. Spatial point pattern analysis and its application in geographical epidemiology. **Institute of the Institute of British Geographers**, London, v. 21, n. 1, p. 256-274, Mar. 1995.

GAVRIKOV, V.; STOYAN, D. The use of marked point processes in ecological and environmental forest studies. **Environmental and Ecological Statistics**, New York, v. 2, n. 2, p. 331-344, Jun. 1995.

GOULARD, M.; PAGÈS, L.; CABANETTES, A. Marked point process: using correlation functions to explore a spatial data set. **Biometrics**, Washington, v. 37, n. 3, p. 837-853, Sep. 1995.

GUAN, Y. Tests for independence between marks and points of a marked point process. **Biometrics**, Washington, v. 62, n. 1, p. 126-134, Mar. 2006.

HARRIS, J. A. Practical universality of field heterogeneity as a factor influencing plot yields. **Journal of agricultural research**, Washington, v.19, n.7, p. 279-314, Sep. 1920.

KRIGE, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Chemistry, Metal. and Mining Society of South Africa**, Johannesburg, v.52 n.6, p.119-139, 1951.

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

LIMA, R. R. **Modelagem espaço-temporal para dados de incidência de doenças em plantas**. 2005. 149p. Tese (Doutorado em agronomia/Estatística e Experimentação Agronomica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, SP.

- LOUZADA, J. M. **Avaliação e aplicação de testes para a detecção de autocorrelação espacial utilizando dados de marcadores genéticos.** 2003. 103p. Dissertação (Mestrado em estatística e experimentação Agropecuárias) - Universidade Federal de Lavras, MG.
- LLOYD, C.D. **Local models for spatial analysis.** New York: University Beefast, 2007. 250p.
- MATHERON, G. Principles of geostatistics. **Economic mGeology**, New York, v. 58, p.1246-1266, 1963.
- MATTFELDT, T.; ECKEL, S.; FLEISCHER, F.; SCHMIDT, V. statistical analysis of reduced pair correlation functions of capillaries in the prostate gland. **Journal of microscopy**, Oxford, v. 223, n. 1, p.107-119 , Jul. 2006.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics.** New York: J. Wiley, 1974. 564p.
- MONTGOMERY, E. G. **Experiments in wheat breeding: experimental error in the nursery and variation in nitrogen and yield.** Washington: Washington, U.S. Deptmanet. Agriculture, 1913. 61p.
- NICOTRA, A. B. Sex ratio and spatial distribution of siparuna grandiflora, a tropical dioecious shrub. **Oecologia**, Heidelberg, v. 115, p. 102-113, Sep. 1998.
- OGATA, Y. Space-time point process models for earthquake occurrence. **Journal of the American Statistical Association**, New York, v. 93, n. 444, p. 9-27, Dec. 1998.
- PERRY, G. L. W.; MILLER, B. P.; ENRIGHT, N. J. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. **Plant ecology**, Dordrecht, NL, v. 187, n. 1, p.59-82, Jan./Jun. 2006.
- PIAZZA, A.; MENOZZI, P.; CAVALLI-SFORZA, L. The making and testing of geographic gene-frequence maps. **Biometrics**, Washington, v. 37, n. 3, p. 635-659, Sep. 1981.
- PENTTINEN, A.; STOYAN, D.; HENTTONEN, H. M. Marked point processes in forest statistics. Forest Science. **Society of American Foresters**, Lawrence, v. 38, n. 3, p. 806-824, Sep. 1992.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org>>.

RIPLEY, B. D. **Spatial statistics**. London, J. Wiley, 1981. 340p.

RIPLEY, B. D. Modelling spatial patterns. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Oxford, v. 39, n. 2, p. 172-212, Jun. 1977. RIPLEY, B. D. Spectral analysis and the analysis of pattern in plant communities. **Journal of Ecology**, Oxford, v. 66, n. 5, p. 965-981, Sep./Out. 1978.

SCALON, J.D., FIELLE, N. R. J., STILLMAN, E., ATCKINSON, H. Spatial pattern analysis of second-phase particles in composite materials. **Materials science & engineering**, Lausanne, v. 356, n. 1-2, p. 245-257, Sep. 2003.

SCALON, J. D. **Spatial and size distributions of particles in composite materials**. 1998. 318p. Thesis (Ph.D) -University of Sheffield, Sheffield.

SCHLATHER M. On the second - order characteristics of marked point processes. **the Royal Statistical Society: Series B (Statistical Methodology)**, Oxford, v. 7, n. 1, p.9-117, Dec. 2001.

SCHLATHER, M.; RIBEIRO, P. J.; DIGGLE, P. J. Detecting dependence between marks and locations of marked point processes. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Oxford, v. 66, n. 1, p. 79-93, Feb. 2004.

SCHOENBERG, F. P. Multidimensional residual analysis of point process models for earthquake occurrences. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Oxford, v. 65, n. 4, p. 789-795, Nov. 2003.

SMITH, S. The faunal succession of the upper Bernician. **Transactions / Natural History Society of Northumberland, Durham and Newcastle-Upon-Tyne**, v. 24, p. 591-645, 1910.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical methods**. 7. ed. Ames: Iowa State University, 1980. 507p.

SOKAL, R. R.; JACQUEZ, G. M. Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis. **Evolution**. Washington, v.45, p.152-168, 1991.

- STOYAN, D.; PENTTINEN A. Recent applications of point process methods in forestry statistics. **Statistical Science**, Jyvaskyla, v. 15, n. 1, p. 61-78, 2000.
- STOYAN, D.; STOYAN, H. **Fractals: random shapes and point fields**. New York: J. Wiley, 1994. 389p.
- STOYAN, D.; KENDALL, W. S.; MECKE, J. **Stochastic geometry and its applications**. Berlin: Akademie - Verlag, 1987. 156p.
- STOYAN, D.; WINCECK K. Spatial correlation in metal structures and their analysis. **Materials Characterization**. New York, v.26, p.167-176, 1991.
- TOBLER, W.R. A computer movie simulating urban growth in the Detroit region. **Economic Geography**, Worcester, v. 46, n.1, p.234-40, 1970.
- THÖNNES, E.; VAN LIESHOUT, M. C. A comparative study on the power of Van Lieshout and Baddeley's "Stochastic geometry". [S.I.]: Department of Mathematical Statistics Chalmers University of Technology, 1999. Disponível em: < <http://citeseerx.ist.psu.edu/viewdoc/summarydoi=10.1.1.43.9107>>. Acesso em: 3 out. 2008.
- WALLER, L. A.; GOTWAY, C. A. **Applied spatial statistics for public health data**. New Jersey: J. Wiley. 2004, 641p. (Series in probability and statistics).
- WÄLDER K.; WÄLDER O. Analysing interaction effects in forests using the mark correlation function. *IForest*. v. 4, n. 1, p. 365-372, Feb. 2008. Disponível em: < <http://www.sisef.it/iforest/pdf/Walder449.pdf>>. Acesso em: 28 Feb. 2008.
- WÄLDER O.; STOYAN, D. On variograms in point process statistics. **Biometrical**, v. 38, p. 395-905, Feb. 1996.
- VAN LIESHOUT, M. N. M. **A J-function for marked point patterns**. Tokyo: Institute of Statistical Mathematics, 2006.
- VAN LIESHOUT, M. N. M.; BADDELEY, A. J. A nonparametric measure of spatial interaction in point patterns. **Statistica Neerlandica**, Cambridge, v. 50, p. 344-361, 1996.
- VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. New York: Springer, 2000. 461p.
- VIEIRA, S. R.; NOVAIS, R. F.; ALVARES, V. H.; SCHAEFFER, C. E. G. R. Geoestatística em estudos de variabilidade espacial do solo. **Tópicos em ciência do solo**. Viçosa: Sociedade Brasileira de Ciência do Solo, p.1-54, 2000.

APÊNDICE

PROGRAMA A	Programa utilizado para simulação da função de correlação marcada para marcas contínuas.	68
PROGRAMA B	Programa utilizado para simulação da função de correlação marcada para marcas categóricas.	70
PROGRAMA C	Programa utilizado para o teste de hipótese da função de correlação marcada para marcas contínuas.	72
PROGRAMA D	Programa utilizado para o teste de hipótese da função de correlação marcada para marcas categóricas.	73
PROGRAMA E	Processo de simulação para as funções G , F , K e J	75
PROGRAMA F	Análise exploratória dos dados.	76

PROGRAMA A: Programa utilizado para simulação da função de correlação marcada para marcas contínuas.

```
# Reconhecendo os dados

library(spatstat) #carregando o pacote spatstat
library(splanx) #carregando o pacote splanx
amostral<- read.table(C:/Alumi/IMG04.txt,h=T)
attach(amostral)
alumínio1<- ppp(amostral$x,amostral$y,c(0,511),
               c(0,768))
marcas1<- c(amostral$z)
Marcas1<- ppp(amostral$x, amostral$y, c(0,511),
              c(0,768), marks= marcas1)

#Simulação da função de correlação marcada
#para marcas contínuas

markcor<- function(event=alumínio1, nsim = 1000){
#Esta função implementa as simulações Monte Carlo
#a fim de criar e plotar os envelopes de
#simulação para testar a hipótese de
#independência entre marcas e pontos de
#um processo pontual marcado onde
# event: é o objeto da classe ppp que contém as
#coordenadas dos eventos e as marcas
#nsim: número de simulações Monte Carlo
  r<- seq(0, 0.25*max(event$x,event$y), length=100)
  area<- max(event>window$yrange)*max(event>window$xrange)
  lengt<- length(r)
  np<- length(event$x)
  y<- seq(0,1.5,length=lengt)
  mmin<- min(event$marks)
  mmax<- max(event$marks)
#obtendo a função de correlação marcada por
#meio do processo pontual marcado
ko<- markcorr(event, f = function(m1,m2) { m1 * m2
  }, correction= 'Ripley',r=r, method= 'density')
#efetuando o processo de simulação
```

```

hold<- matrix(0, nsim, lengt)
  for(i in (1:nsim)) {
    X<- rpoispp(np/area, win=owin(c(0,max(event$window
      $xrange)), c(0,max(event$window$yrange))))
    spts<- X %mark% runif(X$n, min=mmin, max=mmax)
    k <- markcorr(spts, f = function(m1,m2) { m1 * m2 },
correction='Ripley', r=r, method='density')
    hold[i, ] <- k$iso
  }
#observe que cada fileira de linha contém a
#função de distribuição empírica para uma única
#simulação de CSR no quadrado unitário
  mn<- apply(hold,2,mean)
  up<- apply(hold,2,max)
  dow<- apply(hold,2,min)
#criando o gráfico
  plot(r, y, type = 'n', xlab = 'Distância',
  ylab = 'Correlação Marcada')
  lines(r, ko$iso)
  lines(r, ko$theo, lty=3)
  lines(r, up, lty='2', col='red')
  lines(r, dow, lty='2', col='green')
  }
mar<- markcor(Marcas1)

```

PROGRAMA B: Programa utilizado para simulação da função de correlação marcada para marcas categóricas.

```
# Reconhecendo os dados

library(spatstat) #carregando o pacote spatstat
library(splancs) #carregando o pacote splancs
pontos <- read.table('C:/Alelos/Alelos.txt',h=T)
attach(pontos)
copaiba<- ppp(pontos$x,pontos$y,c(40,310 ),c(0,160))
marca<- c(pontos$Est2)
marca <- factor(marca)
M <- markcorr(Marca, function(m1,m2) {m1==m2},
correction='translate', method='density',
kernel='epanechnikov')
Marca<- ppp(pontos$x, pontos$y, c(40,310 ), c(0,160),
marks<- marca)

#Simulação da função de correlação marcada
#para marcas categóricas

mark<- function(event=Marca, nsim = 1000){
#Esta função implementa as simulações Monte Carlo
#a fim de criar e plotar os envelopes de simulação
#para testar a hipótese de independência entre marcas
#e pontos de um processo pontual marcado
#onde event: é o objeto da
#classe ppp que contém as coordenadas
#dos eventos e as marcas
#nsim: número de simulações Monte Carlo
  r<- seq(0, 0.25*max(event$x,event$y),
length=100)
  area<- max(event>window$yrange)*max(event$
window$xrange)
  lengt<- length(r)
  np<- length(event$x)
  y<- seq(0,1.5,length=lengt)
#obtendo a função de correlação marcada por
#meio do processo pontual marcado
  ko<- markcorr(event, f = function(m1,m2) {m1==m2
```

```

    }, correction= 'Ripley', r=r, method= 'density')
#efetuando o processo de simulação
hold<- matrix(0, nsim, lengt)
for(i in (1:nsim)) {
X<- rpoispp(np/area, win=owin(c(0,max(event$window
$ xrange)), c(0,max(event$window$yrange))))
Y<- sample(event$marks, length(event$x), replace=T)
Y<- factor (Y)
XX<- ppp(event$x, event$y, c(0,max(event$window
$ xrange)),
c(0,max(event$window$yrange)), marks=Y)
XX<- ppp(X$x, X$y, c(0,max(event$window$xrange)),
c(0,max(event$window$yrange)), marks=Y)
spts<- as.ppp(XX)
k <- markcorr(spts, f = function(m1,m2) { m1==m2
}, correction='Ripley', r=r, method='density')
hold[i, ] <- k$iso
#observe que cada fileira de linha contém a
#função de distribuição empírica para uma única
#simulação de CSR no quadrado unitário
mn<- apply(hold,2,mean)
up<- apply(hold,2,max)
dow<- apply(hold,2,min)
#criando o gráfico
plot(r, y, type = 'n', xlab = 'Distância (metros)
', ylab = 'Correlação Marcada')
lines(r, ko$iso)
lines(r, ko$theo, lty=3, col='red')
lines(r, up, lty=2, col='green')
lines(r, dow, lty=2, col='blue')
}
mar<- mark(Marca)

```

PROGRAMA C : Programa utilizado para o teste de hipótese da função de correlação marcada para marcas contínuas.

```

#Teste de hipótese para a função de correlação marcada
por meio da marcas contínuas

  r<- seq(0, 0.25*max(event$x,event$y), length=100)
  area<- max(event>window$yrange)*max(event>window
          $xrange)
  lengt<- length(r)
  np<- length(event$x)
  y<- seq(0,1.5,length=lengt)
  mmin<- min(event$marks)
  mmax<- max(event$marks)
#obtendo a área da integral por meio do processo
#pontual marcado observado
ko<- markcorr(event, f = function(m1,m2) { m1 * m2
      }, correction= 'Ripley',r=r, method= 'density')
  D<- (ko$iso-ko$theo)^2
  nteg <- 0
  lent <- length(r)
  step<-max(r)/lent
  integ<-0
steplength <- step * 2
for(count in (2:(lent - 1))) {
  integ <- integ + (D[count - 1] + 4 * D[count] +
      D[count+1])* steplength
  }
  integ <- integ/6
for(isim in (2:nsim)) {
  cat('doing simulation', isim, '\n')
  X<- rpoispp(np/area, win=owin(c(0,max(event$
window$xrange)),c(0,max(event>window$yrange))))
  spts<- X %mark% runif(X$n, min=mmin, max=mmax)
  k <- markcorr(spts, f = function(m1,m2) {m1 * m2},
      correction='Ripley', r=r, method='density')
  t <- (k - 1)^2
  teg <- 0
for(count in (2:(lent - 1))) {

```

```

teg <- teg + (t[count - 1] + 4 * t[count] +
             t[count + 1]) * steplength
}
teg <- teg/6
integ <- c(integ, teg)
}
p <- rank(integ)[1]
pvalue <- (nsim + 1 - p)/nsim
list(D = D, rank_D = p, p_value
     = pvalue)}
t<- marktest(Marcas1)

```

PROGRAMA D : Programa utilizado para o teste de hipótese da função de correlação marcada para marcas categóricas.

#Teste de hipótese para a função de correlação marcada por meio de marcas categóricas

```

r<- seq(0, 0.25*max(event$x,event$y), length=100)
area<- max(event$window$yrange)*max(event$window
    $xrange)
lengt<- length(r)
np<- length(event$x)
y<- seq(0,2,length=lengt)
#obtendo a área da integral por meio do processo
#pontual marcado observado
ko<- markcorr(event, f = function(m1,m2) { m1 == m2 },
    ),correction='Ripley', r=r, method='density')
D<-(ko$iso-ko$theo)^2
nteg <- 0
lent <- length(r)
step<-max(r)/lent
integ<-0
steplength <- step * 2
for(count in (2:(lent - 1))) {
    integ <- integ + (D[count - 1] + 4 * D[count] +
        D[count+1])* steplength
}
integ <- integ/6

```

```

for(isim in (2:nsim)) {
  X<- rpoispp(np/area, win=owin(c(0,max(event$
window$xrange)),c(0,max(event>window$yrange))))
  xx<- event$x
  yy<- event$y
  Y<-sample(event$marks, length(event$x),replace=T)
  Y<-factor(Y)
  X<- ppp(event$x,event$y,c(0,max(event>window$xrange)
,c(0,max(event>window$yrange)),marks=Y)
  spts<- as.ppp(X)
  k <- markcorr(spts, f = function(m1,m2) {m1 = m2},
    correction='Ripley', r=r, method='density')
  t <- (k$iso - k$theo)^2
  teg <- 0
  for(count in (2:(lent - 1))) {
    teg <- teg + (t[count - 1] + 4 * t[count] +
      t[count + 1]) * steplength
  }
  teg <- teg/6
  integ <- c(integ, teg)
  }
  p <- rank(integ)[1]
  pvalue <- (nsim + 1 - p)/nsim
  list(integ=integ, est = integ[1], rank_
    D = p, p_value
    = pvalue)}
t<- markttest(Marcas1)

```

PROGRAMA E : Processo de simulação para as funções G , F , K e J .

```
# SIMULAÇÃO POR MEIO DA FUNÇÃO G
r<- seq(0,20,length=100)
G<- envelope(alumíniol, Gest, nsim = 1000,
r=r,correction='rs')
plot(G,main='envelope da função G',xlab='Distância',
ylab='G(y)', xlim=c(0,20), ylim=c(0.0,1.0))

# SIMULAÇÃO POR MEIO DA FUNÇÃO F
r<- seq(0,20,length=100)
F<- envelope(alumíniol, Fest, nsim = 1000,
r=r,correction='rs')
plot(F,main='envelope da função F',xlab='Distância',
ylab='F(x)', xlim=c(0,20), ylim=c(0.0,1.0))

# SIMULAÇÃO POR MEIO DA FUNÇÃO J
r<- seq(0,20,length=100)
J<- envelope(alumíniol, Jest, nsim = 1000, r=r)
plot(J, main='envelope da função J',xlab='Distância',
ylab='J(z)', xlim=c(0,20), ylim=c(0.6,1.4))

# SIMULAÇÃO POR MEIO DA FUNÇÃO K
r<- seq(0,30,length=100)
K<- envelope(alumíniol, Lest, nsim = 1000, r=r,
correction='border')
plot(K,main='envelope da função K',xlab='Distância',
ylab='L(t)')
```

PROGRAMA F : Análise exploratória dos dados.

```
# Análise descritiva dos diâmetros das partículas
marcas1<- c(amostral$z)
Marcas1<- ppp(amostral$x, amostral$y, c(0,511),
             c(0,768), marks= marcas1)
plot(Marcas1)
hist(amostral$z,
     nc<- 8,
     right<- F,
     main<- 'Histograma da amostra A',
     xlab<- 'Diâmetro das partículas',
     ylab<- 'frequência',
     col<- '8')
M<- mean(amostral$z) # média
V<- var(amostral$z) # variância
DP<-sd(amostral$z) # desvio padrão
CV<- sd(amostral$z)/mean(amostral$z)*100
# coeficiente de variação
M; V; DP; CV

# Calculo do coeficiente de assimetria e curtose
m2<- (sum((amostral$z-mean(amostral$z))^2))/length(
     amostral$z) # primeiro momento amostral
m2<- (sum((amostral$z-mean(amostral$z))^2))/
     length(amostral$z) #segundo momento amostral
m3<- (sum((amostral$z-mean(amostral$z))^3))/length(
     amostral$z) #terceiro momento amostral
m4<- (sum((amostral$z-mean(amostral$z))^4))/length(
     amostral$z) #quarto momento amostral
m2; m3; m4
g1<- m3/(m2^(3/2)) # coeficiente de assimetria
curtose<- m4/(m2^2) # coeficiente de curtose
g1; curtose
g1.tab<- qnorm(0.95,0,(6/945)^1/2)
             # coeficiente de curtose
curtose.tab<- qnorm(0.95,0,(24/945)^1/2)
             # excesso de curtose
tabelado g1.tab*100; curtose.tab*100
```