



**FRANCIELLA MARQUES DA COSTA**

**DISTRIBUIÇÃO E ESTRUTURA DE  
DEPENDÊNCIA ESPACIAL EM DADOS DE  
ÁREA**

**LAVRAS - MG**

**2014**

**FRANCIELLA MARQUES DA COSTA**

**DISTRIBUIÇÃO E ESTRUTURA DE DEPENDÊNCIA ESPACIAL EM  
DADOS DE ÁREA**

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

Orientador

Dr. João Domingos Scalon

**LAVRAS - MG**

**2014**

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e  
Serviços da Biblioteca Universitária da UFLA**

Costa, Franciella Marques da.

Distribuição e estrutura de dependência espacial em dados de  
área / Franciella Marques da Costa. – Lavras : UFLA, 2014.

113 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2014.

Orientador: João Domingos Scalon.

Bibliografia.

1. Índice de Moran. 2. Correlograma. 3. Testes de hipótese. 4.  
Teste do Syrjala. 5. Bicho-mineiro do cafeeiro. I. Universidade  
Federal de Lavras. II. Título.

CDD – 519.2

**FRANCIELLA MARQUES DA COSTA**

**DISTRIBUIÇÃO E ESTRUTURA DE DEPENDÊNCIA ESPACIAL EM  
DOS DE ÁREA**

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

APROVADA em 25 de julho de 2014.

Dr. Ednaldo Carvalho Guimarães	UFU
Dr. José Márcio de Mello	UFLA
Dr. Marcelo Silva de Oliveira	UFLA
Dr. Mauricio Sergio Zacarias	EMBRAPA-CAFÉ

Dr. João Domingos Scalon  
Orientador

**LAVRAS - MG**

**2014**

Aos meus pais, Cleuza e Orlando;  
meus irmãos, Kelle e Orlando Júnior;  
meus sobrinhos, Amanda e Murilo;  
minha avó, Irene e, *in memoriam*, aos  
avós, Joaquim, Maria Antônia e José e  
meu esposo Ronivon.

**Dedico**

## AGRADECIMENTOS

À Deus, que sempre iluminou e guiou a minha vida, dando-me forças para realizar meus sonhos.

Aos meus pais, Cleuza e Orlando, pelo amor e apoio a mim dedicados.

Aos meus irmãos, Kelle e Orlando Júnior, pelo incentivo e amizade.

Aos meus sobrinhos, Amanda e Murilo, por trazerem tanta alegria para minha vida.

À avó, Irene e, *in memoriam*, aos avós, Joaquim, Maria Antônia e José pela atenção e carinho.

Ao meu esposo, Ronivon, pela compreensão e apoio incondicional para a realização deste sonho.

Ao professor Dr. João Domingos Scalon, pela orientação deste trabalho.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, pela oportunidade de cursar o Doutorado.

Aos professores e demais funcionários do Departamento de Ciências Exatas - UFLA.

Aos membros da banca examinadora, pelas sugestões e contribuições.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela concessão da bolsa de estudos.

À Universidade Federal de Uberlândia/Faculdade de Ciências Integradas do Pontal por autorizar o afastamento integral para terminar o doutorado.

Ao professor Dr. José Laércio Dorício da Universidade Federal de Uberlândia/Faculdade de Ciências Integradas do Pontal pela disponibilidade em ajudar com a parte de simulação.

A todos aqueles, que de alguma forma, acreditaram e contribuíram para a realização deste sonho, meu sincero agradecimento.

## RESUMO GERAL

A análise da estrutura de dependência espacial de uma variável aleatória é de extrema importância em diversas áreas de conhecimento tais como ecologia, epidemiologia, geografia e ciências do solo, entre outras. Assim, o objetivo principal desta tese é apresentar métodos estatísticos para analisar a estrutura de dependência espacial em dados de área. Neste sentido, esta tese apresenta novos métodos para verificar não somente a significância estatística de um correlograma, mas também para testar se duas populações, distribuídas no mesmo espaço geográfico, apresentam a mesma estrutura de dependência espacial. Os métodos propostos são avaliados em casos típicos de dependência espacial e, sempre que o método permite, são realizadas análises do poder do teste e do erro do tipo I. Os métodos propostos também são aplicados em dados reais relacionados ao bicho-mineiro do cafeeiro (*Leucoptera coffeella*) em uma plantação de café orgânico. Os resultados obtidos permitem afirmar que os métodos propostos podem ser úteis para atingir os objetivos para os quais foram desenvolvidos.

**Palavras-chave:** Índice de Moran. Correlograma. Testes de hipótese. Aleatorização. Envelope de simulação. Teste do Syrjala. Bicho-mineiro do cafeeiro.

## GENERAL ABSTRACT

The analysis of spatial dependence structure of a random variable is of paramount importance in various areas of knowledge such as ecology, epidemiology, geography and soil sciences, among other. Thus, the main aim of this thesis is to present statistical methods to analyze the structure of spatial dependence in areal data. Therefore, this thesis presents new methods to verify not only the statistical significance of a correlogram, but also to test whether two populations, distributed in the same geographical space, have the same spatial dependence structure. The proposed methods are evaluated on typical cases of spatial dependence and, where the method allows, it is carried out analysis of both test power and type I error. The methods are also applied to real data related to the coffee leaf miner (*Leucoptera coffeella*). The results allow us to affirm that the proposed methods can be useful for achieving the aims for which they were developed.

**Keywords:** Moran's Index. Correlogram. Hypotheses tests. Randomization. Envelope simulation. Syrjala'test. Coffee leaf miner.

## LISTA DE FIGURAS

### PRIMEIRA PARTE

Figura 1 Esquema do semivariograma experimental..... 25

### SEGUNDA PARTE - ARTIGOS

#### ARTIGO 1

Figura 1 Gráfico da precipitação, média mensal de folhas minadas, minas predadas e vespas no ano de 2006 ..... 42

Figura 2 Número de folhas minadas nos 35 pontos de amostragem para os períodos de seca (a) e chuva (b) no ano de 2006. As áreas dos círculos são proporcionais às frequências absolutas observadas. A menor área corresponde à frequência absoluta igual a 5, enquanto a maior área corresponde à frequência absoluta igual a 43..... 43

Figura 3 Número de minas predadas nos 35 pontos de amostragem para os períodos de seca (a) e chuva (b) no ano de 2006. A menor frequência absoluta é igual a 1 e a maior frequência absoluta é igual a 36. O símbolo + indica que não houve ocorrência no ponto de amostragem correspondente..... 43

Figura 4 Número de folhas minadas (a) e vespas (b) nos 35 pontos de amostragem para o período seco. A menor frequência absoluta é igual a 8 e a maior frequência absoluta é igual a 43. O símbolo + indica que não houve ocorrências nos pontos de amostragem correspondentes ..... 44

Figura 5 Número de minas predadas (a) e vespas (b) nos 35 pontos de amostragem para o período seco. A menor frequência absoluta é igual a 1 e a maior frequência absoluta é igual a 30. O símbolo +

	indica que não houve ocorrências nos pontos de amostragem correspondentes .....	45
Figura 6	Número de folhas minadas (a) e vespas (b) nos 35 pontos de amostragem para o mês de setembro. A menor frequência absoluta é igual a 5 e a maior frequência absoluta é igual a 10. O símbolo + indica que não houve ocorrências nos pontos de amostragem correspondentes .....	45
Figura 7	Número de minas predadas (a) e vespas (b) nos 35 pontos de amostragem para o mês de setembro. A menor frequência absoluta é igual a 7 e a maior frequência absoluta é igual a 19. O símbolo + indica que não houve ocorrências nos pontos de amostragem correspondentes .....	46

## ARTIGO 2

Figura 1	Porcentagem de rejeição de $H_0$ para os testes C (vermelho) e K (azul), seguindo o modelo de dependência espacial $Sph(x, 80, 40)$ com valores para o efeito pepita ( <i>eixo x</i> ) iguais a 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71, 76.....	63
Figura 2	Porcentagem de rejeição de $H_0$ para os testes C (vermelho) e K (azul), seguindo o modelo de dependência espacial $Sph(x, 80, 60)$ com valores para o efeito pepita ( <i>eixo x</i> ) iguais a 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71, 76.....	64
Figura 3	Configuração espacial de contagens sob a suposição de forte dependência espacial. Dados simulados seguindo o modelo de dependência espacial $Sph(1, 80, 60)$ .....	66
Figura 4	Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho).....	67

Figura 5	Configuração espacial de contagens sob a hipótese de moderada dependência espacial. Dados simulados seguindo o modelo de dependência espacial $Sph(41,80,60)$ .....	67
Figura 6	Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho). .....	68
Figura 7	Configuração espacial de contagens sob a hipótese de fraca dependência espacial. Dados simulados seguindo o modelo de dependência espacial $Sph(76,80,60)$ .....	68
Figura 8	Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho). .....	69
Figura 9	Configuração espacial de contagens sob a suposição de completa aleatoriedade espacial.....	69
Figura 10	Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho). .....	70
Figura 11	Contagens referentes ao número de minas predadas para as 35 áreas na plantação de café orgânico no mês de setembro de 2006....	71
Figura 12	Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho).....	72

### **ARTIGO 3**

Figura 1	Correlogramas da população 1 (vermelho) e população 2 (azul) obtidos com a configuração $Sph(3,60,40)$ considerando as variabilidades radiais iguais a 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.....	86
Figura 2	Porcentagem de rejeição de $H_0$ com as configurações seguindo o modelo de dependência espacial $Sph(3,60,40)$ para os testes C (vermelho) e K (azul) .....	88
Figura 3	Porcentagem de rejeição de $H_0$ com as configurações seguindo o modelo de dependência espacial $Sph(3,60,60)$ para os testes C (vermelho) e K (azul) .....	88

Figura 4	Porcentagem de rejeição de $H_0$ com as configurações seguindo o modelo de dependência espacial $Sph(30,60,40)$ para os testes C (vermelho) e K (azul) .....	89
Figura 5	Porcentagem de rejeição de $H_0$ com as configurações seguindo o modelo de dependência espacial $Sph(30,60,60)$ para os testes C (vermelho) e K (azul) .....	90
Figura 6	Porcentagem de rejeição de $H_0$ com as configurações seguindo o modelo de dependência espacial $Sph(55,60,40)$ para os testes C (vermelho) e K (azul) .....	91
Figura 7	Porcentagem de rejeição de $H_0$ com as configurações seguindo o modelo de dependência espacial $Sph(55,60,60)$ para os testes C (vermelho) e K (azul) .....	91
Figura 8	Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial $Sph(3,60,60)$ e uma variabilidade radial de 1. ....	93
Figura 9	Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial $Sph(3,60,60)$ e uma variabilidade radial de 10. ....	94
Figura 10	Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial $Sph(55,60,60)$ e uma variabilidade radial de 1. ....	95
Figura 11	Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial $Sph(55,60,60)$ e uma variabilidade radial de 10. ....	96

Figura 12 Contagens referentes ao número de folhas minadas (a) e vespas predadoras (b) para as 35 áreas na plantação de café orgânico no mês de setembro de 2006 .....	98
Figura 13 Correlogramas do número de folhas minadas (vermelho) e vespas predadoras (azul) (a) e envelope simulado com 999 permutações (b) .....	99

## LISTA DE TABELAS

### SEGUNDA PARTE - ARTIGOS

#### ARTIGO 1

Tabela 1	Estatística de teste e valores-p para o teste da hipótese nula de que as distribuições espaciais das duas variáveis no ano de 2006 são iguais .....	47
----------	---	----

#### ARTIGO 2

Tabela 1	Configuração utilizada para simulação.....	61
Tabela 2	Proporção de rejeição de $H_0$ sendo $H_0$ verdadeira.....	63

#### ARTIGO 3

Tabela 1	Configurações utilizadas para simulação.....	84
----------	--	----

## SUMÁRIO

<b>PRIMEIRA PARTE</b>	
<b>1</b>	<b>INTRODUÇÃO</b> ..... 15
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> ..... 18
<b>2.1</b>	<b>Estatística espacial</b> ..... 18
<b>2.2</b>	<b>Autocorrelação espacial</b> ..... 20
<b>2.2.1</b>	<b>Matriz de proximidade espacial</b> ..... 20
<b>2.2.2</b>	<b>Índice global de Moran</b> ..... 21
<b>2.2.3</b>	<b>Testes para autocorrelação espacial</b> ..... 22
<b>2.2.4</b>	<b>Correlograma</b> ..... 23
<b>2.2.5</b>	<b>Semivariograma</b> ..... 24
<b>2.3</b>	<b>Teste de aleatorização</b> ..... 27
<b>2.4</b>	<b>Bicho-mineiro do cafeeiro</b> ..... 28
	<b>REFERÊNCIAS</b> ..... 30
<b>SEGUNDA PARTE - ARTIGOS</b>	
	<b>ARTIGO 1 Análise estatística das distribuições espaciais do bicho-mineiro do cafeeiro e das vespas predadoras</b> ..... 33
	<b>ARTIGO 2 Testes para verificar a significância estatística de correlogramas</b> ..... 51
	<b>ARTIGO 3 Avaliação da estrutura de dependência espacial entre dois processos estocásticos</b> ..... 75
	<b>CONCLUSÕES GERAIS</b> ..... 103
	<b>ANEXOS</b> ..... 104

## **PRIMEIRA PARTE**

### **1 INTRODUÇÃO**

A distribuição espacial de uma população refere-se à forma com que os indivíduos dessa população estão localizados em um determinado espaço. Eles podem estar distribuídos, por exemplo, de forma aleatória ou em agrupamentos. A análise da distribuição espacial de uma população em uma região é de muita importância para diversas áreas do conhecimento, tais como ecologia e agronomia, entre outras, pois pode indicar a maneira como indivíduos de uma população se comportam entre si. Por exemplo, dentro de uma plantação será que a presa e o predador apresentam a mesma distribuição espacial ou será que as distribuições espaciais dos indivíduos dessas espécies sofrem modificações ao longo do tempo?

O fato de duas populações, que dividem o mesmo território, apresentarem a mesma distribuição espacial, não implica que essas populações possuem a mesma estrutura de dependência espacial. Assim, duas populações podem apresentar indivíduos que estão distribuídos em agrupamentos localizados nos mesmos locais. Entretanto, não somente a força, mas também a distância da dependência espacial, que faz esses indivíduos estarem agrupados pode ser completamente diferente nas duas populações. Por exemplo, uma população apresenta indivíduos (presas) agregados no sul da região, mas essa dependência é fraca, atingindo uma distância de quinhentos metros, enquanto a outra população apresenta indivíduos (predadores) agregados no mesmo local, mas essa dependência pode ser forte, atingindo uma distância máxima de duzentos metros.

O contrário também pode ocorrer, ou seja, duas populações que dividem o mesmo território podem apresentar distribuições espaciais completamente

distintas, mas possuírem a mesma estrutura de dependência espacial. Por exemplo, uma população apresenta indivíduos que estão distribuídos em agrupamentos localizados no Sul da região, enquanto a outra população apresenta indivíduos que estão distribuídos em agrupamentos localizados no Norte da região. Entretanto, a força e a distância da dependência espacial, que fazem os indivíduos dessas populações estarem agrupados, podem ser iguais nas duas populações.

Assim, para obter uma melhor compreensão dos fenômenos espaciais de populações que dividem o mesmo espaço é necessário caracterizar, não somente a distribuição espacial dos indivíduos dessas populações, mas também a dependência espacial existente entre esses indivíduos.

Syrjala (1996) propôs um teste estatístico para verificar se a distribuição espacial de duas populações que dividem o mesmo espaço geográfico é a mesma. Esse teste é apresentado e aplicado no primeiro artigo desta tese.

Existem disponíveis na literatura muitos indicadores para realizar a análise de estrutura de dependência espacial, mais especificamente, análise de autocorrelação espacial em dados de área, sendo o índice global de Moran o mais popular e, portanto, um dos mais utilizados nesse tipo de estudo. Em alguns casos pode-se fazer uso do correlograma, que é uma representação gráfica bastante útil, pois permite analisar o comportamento da autocorrelação espacial, pelo índice de Moran, em função das distâncias entre as áreas vizinhas. Apesar de existirem propostas de testes estatísticos para fazer essa análise do correlograma, em geral, essa análise é realizada por inspeção visual. No segundo artigo desta tese, apresentam-se alguns métodos estatísticos para análise da dependência espacial em dados de área, tendo como base a análise da significância estatística de um correlograma.

Nesta tese também são propostos métodos para realizar uma análise comparativa da estrutura de dependência espacial de duas populações que

dividem o mesmo espaço geográfico. Esses métodos são extensões dos métodos desenvolvidos no segundo artigo e são apresentados no terceiro artigo.

Os métodos propostos nesta tese, para análise da estrutura de dependência espacial, são baseados em estatísticas de teste que não apresentam distribuições amostrais conhecidas sob a hipótese nula e, portanto, métodos de computação intensiva são utilizados para obter a significância estatística desses métodos.

Sempre que possível, os métodos apresentados nos três artigos desta tese são analisados quanto ao poder, ao erro do tipo I e em situações típicas. Todos os métodos são aplicados em dados relacionados ao bicho-mineiro do cafeeiro (*Leucoptera coffeella*), tendo em vista a importância de conhecer a distribuição e a estrutura de dependência espacial dessa praga no contexto da produção orgânica de café.

## 2 REFERENCIAL TEÓRICO

### 2.1 Estatística espacial

Um dos pressupostos para o uso de diversos métodos estatísticos para análise de dados é que as realizações da variável aleatória sejam independentes entre si. Porém, diversos fenômenos apresentam dependência espacial, o que impede a utilização de vários desses métodos. Para situações de dependência espacial podem-se utilizar métodos da estatística espacial para análise de dados. Tais métodos utilizam as coordenadas geográficas de forma explícita na análise dos dados (CRESSIE, 1993).

Segundo Assunção (2001) a estatística espacial é uma área da estatística que estuda metodologias para coleta, descrição, visualização e análise de dados, associados a posições geográficas, que possam ser modelados como processos estocásticos. Os processos estocásticos são definidos como um conjunto de variáveis aleatórias  $\{Z(s): s \in D\}$  em que  $D$  é um subconjunto de  $\mathbb{R}^2$  ou  $\mathbb{R}^3$ .

Na estatística espacial, um conceito de extrema importância é a estacionariedade. De acordo com Guimarães (2004), um processo estocástico é dito estacionário se todos os momentos estatísticos são constantes e independentes da origem no espaço. Se o processo estocástico é estacionário de ordem  $K$ , temos:

- $E[Z(s)] = m_1(s)$  é constante para  $\forall s \in D$
- $E[Z^2(s)] = m_2(s)$  é constante para  $\forall s \in D$
- ⋮
- $E[Z^k(s)] = m_k(s)$  é constante para  $\forall s \in D$

Se um processo estocástico é estacionário de ordem  $K$ , então o processo será estacionário para todas as ordens inferiores  $K$ . Para a aplicação de alguns métodos da estatística espacial, exige-se apenas a estacionariedade de segunda

ordem. Segundo Guimarães (2004), um processo estocástico é considerado estacionário de segunda ordem se:

- $E[Z(s)] = m_1(s)$  é constante para  $\forall s \in D$ , ou seja, a média do processo estocástico é constante independente da origem no espaço.
- $E[Z^2(s)] = m_2(s)$  é constante para  $\forall s \in D$ . Isto implica que a variância do processo estocástico é constante, independente da origem no espaço e que para cada par  $\{Z(s), Z(s+h)\}$  a função de covariância existe e depende apenas da distância  $h$ .

Os dados espaciais podem ser classificados, de uma maneira geral, em três tipos (CRESSIE, 1993): Processos pontuais; Superfície aleatória (geoestatística) e Dados de áreas.

Processos pontuais são fenômenos em que se conhece a localização exata dos eventos. O conjunto de dados consiste de uma série de coordenadas geográficas dos eventos em estudo, como por exemplo, locais de ocorrência de crimes, residências de indivíduos doentes, entre outros. Para processos pontuais a variável de interesse é a localização do evento.

Superfícies contínuas (geoestatística) são fenômenos caracterizados pela continuidade espacial da variável aleatória, como por exemplo, a concentração de um minério no solo.

Dados de áreas são caracterizados por eventos agregados por regiões poligonais (bairros, municípios, etc.). Neste caso não se conhece a localização exata do evento, apenas o número de ocorrências por região, como por exemplo, o número de casos de homicídio por município.

Em geral, existem métodos de análise específicos para cada tipo de dados, mas em determinadas situações um mesmo método pode ser aplicado em diferentes tipos de dados. Os métodos estatísticos propostos neste trabalho são desenvolvidos, exclusivamente, para análise de dados de áreas.

## 2.2 Autocorrelação espacial

A dependência espacial é um conceito de extrema importância para a compreensão dos fenômenos espaciais e pode ser avaliada por meio da autocorrelação espacial. A autocorrelação espacial refere-se à correlação espacial entre realizações de uma mesma variável aleatória medida em locais distintos do espaço. O índice global de Moran é uma das medidas de autocorrelação espacial mais utilizada para dados de áreas. Esse índice possibilita estimar o quanto o valor observado de uma variável aleatória em uma determinada localização é dependente dos valores desta mesma variável nas localizações vizinhas (MORAN, 1950).

O cálculo do índice de Moran depende da definição da matriz de proximidade espacial, descrita a seguir.

### 2.2.1 Matriz de proximidade espacial

A matriz de proximidade espacial também é conhecida como matriz de vizinhança. Segundo Waller e Gotway (2004), o elemento  $w_{ij}$  da matriz de proximidade espacial  $W_{N \times N}$  quantifica a dependência espacial entre as regiões  $i$  e  $j$  e a matriz  $W$  define a estrutura de vizinhança para toda a área em estudo. Existem vários critérios para construir a matriz de proximidade espacial, tais como:

- $w_{ij} = 1$  se as regiões  $i$  e  $j$  compartilham uma fronteira,  $w_{ij} = 0$  caso contrário;
- $w_{ij} = 1$  se o centroide da região  $j$  é um dos  $q$  mais próximos do centroide da região  $i$ ,  $w_{ij} = 0$  caso contrário;

- $w_{ij} = 1$  se  $d_{ij} < d$  em que  $d$  é uma distância definida pelo pesquisador e  $d_{ij}$  é a distância entre os centroides das regiões  $i$  e  $j$ ,  $w_{ij} = 0$  caso contrário;
- $w_{ij} = \frac{l_{ij}}{l_i}$  em que  $l_{ij}$  é o comprimento da fronteira comum entre as regiões  $i$  e  $j$  e  $l_i$  é o perímetro da região  $i$ .

Pode-se normalizar as linhas da matriz de proximidade espacial. Para tanto basta dividir cada elemento da matriz pela soma de todos os elementos da linha e assim  $w_i = \sum_{j=1}^N w_{ij} = 1$  (CÂMARA et al., 2004).

### 2.2.2 Índice global de Moran

O Índice global de Moran (MORAN, 1950) é uma medida de autocorrelação espacial que fornece uma medida geral da autocorrelação espacial existente em um conjunto de dados, sendo utilizado para dados de áreas.

Para o cálculo do índice de Moran, os dados devem ser estacionários de segunda ordem (MORAN, 1950), sendo calculado, considerando apenas o primeiro vizinho, pela expressão:

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i=1}^N (Y_i - \bar{Y})^2) (\sum_{i=1}^N \sum_{j=1}^N w_{ij})} \quad (1)$$

em que,  $N$  é o número de áreas,  $Y_i$  é o valor do atributo na área  $i$ ,  $Y_j$  é o valor do atributo na área  $j$ ,  $\bar{Y}$  é a média do atributo e  $w_{ij}$  são os elementos da matriz de proximidade espacial.

De acordo com Upton e Fingleton (1985) o índice global de Moran pode assumir qualquer valor real, entretanto, na maioria dos casos pertence ao intervalo  $-1 \leq I \leq 1$ .

Segundo Waller e Gotway (2004), o índice global de Moran será negativo quando as áreas vizinhas possuírem valores diferentes. Quando as áreas vizinhas tiverem valores próximos, o índice global de Moran vai apresentar um valor positivo.

### 2.2.3 Testes para autocorrelação espacial

Após calcular o índice de Moran, pode-se aplicar um teste de hipótese para verificar se a autocorrelação espacial é estatisticamente significativa. A hipótese de independência espacial pode ser testada utilizando duas abordagens: a distribuição aproximada do índice de Moran ou o teste de permutação aleatória (BAILEY; GATRELL, 1995).

De acordo com Bailey e Gatrell (1995), se houver um número moderado de áreas observadas e supondo que  $y_i$  são observações de uma variável aleatória  $Y_i$  com distribuição normal e que  $Y_i$  e  $Y_j$  são espacialmente independentes para  $i \neq j$ , ou seja, assumindo que as variáveis aleatórias são independentes e identicamente distribuídas com distribuição normal, a distribuição amostral do Índice de Moran ( $I$ ) é aproximadamente normal. Segundo Waller e Gotway (2004), a média e a variância da distribuição amostral do índice de Moran são dadas por:

$$E(I) = -\frac{1}{(N-1)} \quad (2)$$

$$VAR(I) = \frac{N^2 S_1 - N S_2 + 3 S_0^2}{(N+1)(N-1) S_0^2} - \left(\frac{1}{N-1}\right)^2 \quad (3)$$

em que  $S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$ ,  $S_1 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (w_{ij} + w_{ji})^2$  e  $S_2 = \sum_{i=1}^N (w_{i+} + w_{+j})^2$  em que  $w_{i+} = \sum_{j=1}^N w_{ij}$  e  $w_{+j} = \sum_{i=1}^N w_{ij}$ .

Segundo Waller e Gotway (2004), sob a suposição de normalidade dos dados, a estatística de teste para a hipótese nula de independência espacial é dada por:

$$Z = \frac{I - E(I)}{\sqrt{VAR(I)}} \quad (4)$$

Para testar a hipótese nula, basta comparar a estatística de teste apresentada na expressão (4) com a distribuição normal padrão, como em qualquer teste estatístico clássico.

Outra abordagem, sem pressuposto em relação à distribuição dos dados e ao número de áreas observadas, é o teste de permutação aleatória. Seja  $N$  o número de áreas e  $y_i$  o valor observado para cada área. Então será possível realizar  $N!$  permutações no conjunto de dados original. Calcula-se o índice de Moran para cada permutação dos dados obtendo assim a distribuição empírica de  $I$ . Se o índice de Moran, obtido para o conjunto de dados observado, estiver em um extremo da distribuição empírica então se rejeita a hipótese nula, ou seja, a autocorrelação espacial é estatisticamente significativa (BAILEY; GATRELL, 1995).

#### 2.2.4 Correlograma

O índice global de Moran pode ser estimado para cada ordem de vizinhança (*lag*) pela expressão dada por (BAILEY; GATRELL, 1995):

$$I^{(k)} = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij}^{(k)} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i=1}^N (Y_i - \bar{Y})^2) (\sum_{i=1}^N \sum_{j=1}^N w_{ij}^{(k)})} \quad (5)$$

em que  $w_{ij}^{(k)}$  são os elementos da matriz de proximidade espacial para a ordem de vizinhança  $k$ .

A representação gráfica do índice de Moran para cada ordem de vizinhança  $k$  em função de  $k$  é denominado de correlograma (BAILEY; GATRELL, 1995). A inspeção visual do correlograma permite conhecer o comportamento da autocorrelação espacial para diferentes ordens de vizinhança. Além da inspeção visual, pode-se testar a significância estatística de um correlograma. Propostas de testes para significância estatística de um correlograma podem ser vistos em Oden (1984) e no Artigo II desta tese.

### 2.2.5 Semivariograma

Outra maneira de avaliar a dependência espacial, particularmente para superfícies contínuas, é através da semivariância. Sob a suposição que o processo estocástico  $Z(s)$  atenda a hipótese intrínseca, ou seja, para todo vetor  $h$ , o incremento  $[Z(s) - Z(s + h)]$  admite esperança matemática e variância finita independentes da posição  $s$ , mas dependendo apenas do valor de  $h$ , pode-se obter a função de semivariância.

A função de semivariância é uma função matemática intrínseca, definida como a metade da variância das diferenças entre todos os pares de pontos presentes na área estudada, separados por uma distância  $h$  conforme a equação (6). Portanto, essa função representa o grau de dependência espacial entre duas observações da função aleatória  $Z(s)$  (CRESSIE, 1993).

$$\gamma(h) = \frac{1}{2} \{VAR[Z(s) - Z(s + h)]\} \quad (6)$$

O estimador de Matheron para a semivariância é dado por

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{s_i \text{ e } s_j \text{ tal que } |s_i - s_j| = h} [z(s_i) - z(s_i + h)]^2 \quad (7)$$

em que  $\hat{\gamma}(h)$  é o estimador da semivariância para uma distância  $h$ ;  $N(h)$  representa o número de pares de valores medidos  $(Z(s_i), Z(s_i + h))$  separados por um vetor de distância  $h$  (CRESSIE, 1993).

O gráfico de  $\hat{\gamma}(h)$  versus  $h$  representa o semivariograma, que permite obter a estimativa do valor de semivariância para as diferentes combinações de pares de pontos e assim analisar o grau de dependência espacial da variável estudada e definir os parâmetros necessários para a estimativa de suas características em locais não amostrados (GUIMARÃES, 2004). Na Figura 1, está representado um semivariograma e seus parâmetros. A representação da Figura 1 é próxima ao ideal, ou seja, à medida que  $h$  aumenta a semivariância  $\hat{\gamma}(h)$  também aumenta, pois é esperado que quanto mais próximos estiverem os pontos amostrados, maior será a semelhança entre eles e, portanto, menor a semivariância.

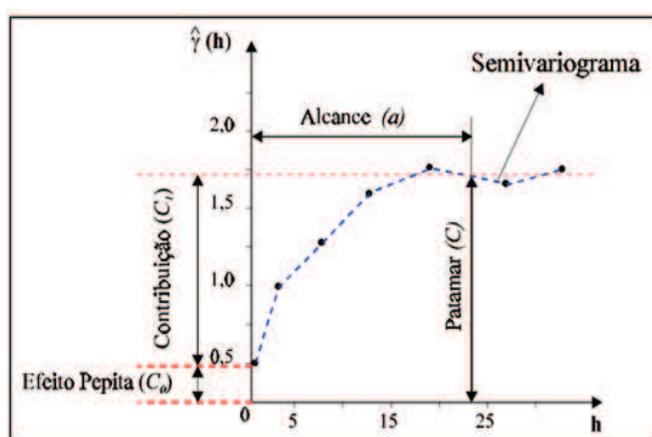


Figura 1 Esquema do semivariograma experimental.

Fonte: Camargo (1998)

A interpretação dos parâmetros da Figura 1, segundo Camargo (1998) e Guimarães (2004), é a seguinte:

Alcance ( $a$ ): corresponde ao valor da distância  $h$  a partir da qual a semivariância  $\gamma(h)$  se torna aproximadamente constante. É a distância a partir da qual as amostras passam a ser independentes, sem correlações espaciais, ou seja, a variação média entre duas observações não é mais função da distância entre elas. Então, o alcance do semivariograma separa o campo estruturado (amostras correlacionadas) do campo aleatório (amostras independentes).

Efeito Pepita ( $C_0$ ): é o valor que representa a descontinuidade na origem do semivariograma, que é justificada por vários fatores como erros de amostragem, erros de medição, erros de análises de laboratório, etc.

Patamar ( $C$ ) : representa o nível de variabilidade em que o semivariograma se estabiliza, ou seja, é o valor constante da semivariância  $\gamma(h)$  que corresponde ao seu alcance.

Contribuição ( $C_1$ ): é a diferença entre o patamar e o efeito pepita, correspondendo à variação espacial dos dados.

O grau de dependência espacial pode ser medido pela relação entre o efeito pepita e o patamar, chamado coeficiente de efeito pepita, dado por:

$$IDE = \frac{C_0}{C_0 + C_1} \quad (8)$$

O valor encontrado no coeficiente da equação (8) pode ser utilizado para classificar o grau de dependência espacial do atributo em estudo, como forte, se for menor que 0,25; moderado, se os valores estiverem entre 0,25 e 0,75 ; e fraca dependência espacial, se o valor for maior que 0,75 (CAMBARDELLA et al., 1994).

### 2.3 Teste de aleatorização

Quando a distribuição amostral da estatística de um teste proposto não é conhecida, o método usual da estatística clássica para conduzir testes de hipótese não pode ser utilizado. Dessa forma, para conduzir o teste de hipótese, pode-se utilizar o método de computação intensiva denominado de aleatorização.

A ideia de um teste de aleatorização é bastante simples e consiste em considerar uma estatística de teste  $T$  para testar a hipótese nula e obter a estatística de teste para o conjunto de dados observados denominada  $T_0$ . Em seguida, permuta-se o conjunto de dados e calcula  $T$  para cada novo arranjo, obtido após a permutação dos dados, obtendo assim a distribuição amostral da estatística de teste  $T$  sob a hipótese nula. Se a hipótese nula é verdadeira, todos os possíveis arranjos para o conjunto de dados têm a mesma probabilidade de ocorrer. Logo, a ordem do conjunto de dados observados é uma das ordens equiprováveis, sendo assim,  $T_0$  deve aparecer como um valor típico da distribuição amostral de  $T$ . Se  $T_0$  estiver no extremo da distribuição de  $T$  então se rejeita a hipótese nula. O *valor - p* é obtido pela proporção dos valores de  $T$  que são maiores ou iguais à estatística de teste  $T_0$  (MANLY, 1994).

Em situações em que há uma grande quantidade de valores observados, considerar todas as permutações no conjunto de dados seria inviável. Neste caso, pode-se utilizar um teste de aleatorização aproximado, em que não serão realizadas todas as permutações possíveis.

A aleatorização é uma técnica muito utilizada na análise espacial como pode ser vista nos livros de Bailey e Gatrell (1995) e Waller e Gotway (2004), nos trabalhos de Chapman et al. (2009), Cornuault et al. (2013), Hedger et al. (2004), Ma et al. (2012), Seidel (2012), Swain e Wade (2003), Syrjala (1996) e nos artigos desta tese.

## 2.4 Bicho-mineiro do cafeeiro

O bicho-mineiro do cafeeiro (*Leucoptera coffeella* (Guérin-Mèneville & Perrottet, 1842) (Lepidoptera: Lyonetiidae)) é uma praga que causa preocupação ao produtor de café brasileiro. O bicho-mineiro é de origem africana, foi detectada no Brasil a partir de 1851 e recebeu esse nome devido a suas lagartas minarem as folhas de café (SOUZA; REIS; RIGITANO, 1998). O bicho-mineiro ataca somente o cafeeiro (REIS; SOUZA, 1986).

A mariposa (adulto) do bicho-mineiro mede 6,5 mm de envergadura é branco-prateada com asas franjadas (REIS; SOUZA; ZACARIAS, 2006). Os ovos são depositados na parte superior das folhas do cafeeiro e as lagartas provenientes destes ovos penetram na folha formando as minas onde vivem até seu desenvolvimento completo (REIS; SOUZA; ZACARIAS, 2006).

As mariposas iniciam suas atividades no final da tarde e começo do anoitecer, enquanto durante o dia elas ficam sob as folhas dos cafeeiros (SOUZA; REIS; RIGITANO, 1998). Segundo Reis, Souza e Zacarias (2006), as maiores infestações do bicho-mineiro do cafeeiro são observadas nos períodos secos do ano.

O combate ao bicho-mineiro geralmente é realizado pelo controle químico. Porém esse tipo de controle é restringido ou proibido em se tratando de cultivos orgânicos. Assim o controle biológico é de extrema importância para cultivos orgânicos. O controle biológico é realizado por meio de parasitoides (*microhimenópteros*) que parasitam e matam as lagartas do bicho-mineiro, por vespas predadoras que retiram as lagartas de dentro das minas e as comem e por patógenos (bactérias e fungos) que são causadores de doenças nas lagartas (SOUZA; REIS; RIGITANO, 1998). De acordo com Reis e Souza (1986), os predadores são os agentes mais eficientes para o controle biológico, em que todos os predadores identificados são da ordem *Hymenoptera* e da família

*Vespidae*. Em condições experimentais foi possível mudar ninhos de vespas predadoras para lugares próximos do cafezal, porém na prática esse controle ainda tem sido realizado em condições naturais.

## REFERÊNCIAS

ASSUNÇÃO, R. M. **Estatística espacial com aplicações em epidemiologia, economia e sociologia**. São Carlos: UFSCar, 2001. 136 p.

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific & Technical, 1995. 413 p.

CÂMARA, G. et al. Análise espacial de áreas. In: \_\_\_\_\_. **Análise espacial de dados geográficos**. São José dos Campos: INPE, 2004. cap. 5. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf>>. Acesso em: 1 ago. 2013.

CAMARGO, E. C. G. Geoprocessamento para projetos ambientais. In: \_\_\_\_\_. **Geostatística: fundamentos e aplicações**. São José dos Campos: INPE, 1998. cap. 5. Disponível em: <[http://www.dpi.inpe.br/gilberto/tutoriais/gis\\_ambiente/5geoest.pdf](http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf)>. Acesso em: 1 ago. 2013.

CAMBARDELLA, C. A. et al. Field-scale variability of soil properties in Central Iowa soils. **Soil Science Society America Journal**, Madison, v. 58, n. 5, p. 1501-1511, Sept./Oct. 1994.

CHAPMAN, A. V. et al. Dispersal of *Trichogramma ostrinia* (Hymenoptera: Trichogrammatidae) in potato field. **Environmental Entomology**, College Park, v. 38, n. 3, p. 677-685, June 2009.

CORNUAULT, J. et al. The role of ecology in the geographical separation of blood parasites infecting an insular bird. **Journal of Biogeography**, Oxford, v. 40, n. 7, p. 1313-1323, July 2013.

CRESSIE, N. A. C. **Statistics for spatial data**. New York: J. Wiley, 1993. 928 p.

GUIMARÃES, E. C. **Geoestatística básica e aplicada**. Uberlândia: UFU, 2004. 77 p. Apostila.

HEDGER, R. et al. Analysis of the spatial distributions of mature cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) abundance in the North Sea (1980-1999) using generalized additive models. **Fisheries Research**, New York, v. 70, n. 1, p. 17-25, Nov. 2004.

MA, Z. et al. Use of localized descriptive statistics for exploring the spatial pattern changes of bird species richness at multiple scales. **Applied Geography**, Oxford, v. 32, n. 2, p. 185-194, Mar. 2012.

MANLY, B. F. J. **Randomization and Monte Carlo methods in biology**. London: Chapman & Hall, 1994. 281 p.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, London, v. 37, p. 17-23, 1950.

ODEN, N. L. Assessing the significance of a Spatial Correlogram. **Geographical Analysis**, Columbus, v. 16, p. 1-16, 1984.

REIS, P. R.; SOUZA, J. C. de. Pragas do cafeeiro. In: RENA, A. B. et al. (Ed.). **Cultura do cafeeiro: fatores que afetam a produtividade**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1986. p. 323-378.

REIS, P. R.; SOUZA, J. C. de; ZACARIAS, M. S. Alerta para o bicho-mineiro. **Cultivar**, Pelotas, v. 8, n. 84, p. 13-16, abr. 2006.

SEIDEL, E. J. **Novas contribuições para avaliação e descrição da estrutura de dependência espacial em geoestatística**. 2012. 145 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2012.

SOUZA, J. C.; REIS, P. R.; RIGITANO, R. L. O. **Bicho mineiro do cafeiro: biologia, danos e manejo integrado**. 2. ed. Belo Horizonte: EPAMIG, 1998. 48 p. (Boletim Técnico, 54).

SWAIN, D. P.; WADE, E. J. Spatial distribution of catch and effort in a fishery for snow crab (*Chionoecetes opilio*): tests of predictions of the ideal free distribution. **Canadian Journal of Fishery and Aquatical Science**, Ottawa, v. 60, n. 8, p. 897-909, Aug. 2003.

SYRJALA, S. E. A statistical test for a difference between the spatial distributions of two populations. **Ecology**, Durham, v. 77, n. 1, p. 75-80, Jan. 1996.

UPTON, G. J. G.; FINGLETON, B. **Spatial data analysis by example: point pattern and quantitative data**. New York: J. Wiley, 1985. v. 1, 422 p.

WALLER, L. A.; GOTWAY, C. A. **Applied spatial statistics for public health data**. New Jersey: J. Wiley, 2004. 519 p.

**SEGUNDA PARTE - ARTIGOS****ARTIGO 1: ANÁLISE ESTATÍSTICA DAS DISTRIBUIÇÕES ESPACIAIS DO BICHO-MINEIRO DO CAFEEIRO E DAS VESPAS PREDADORAS**

Artigo redigido conforme as normas da *Revista Coffee Science* - versão aceita para publicação

Franciella Marques da Costa<sup>1</sup>, Gabriella de Freitas Alves<sup>2</sup>  
João Domingos Scalon<sup>3</sup>, Mauricio Sergio Zacarias<sup>4</sup>

**RESUMO:** O cafeeiro está sujeito ao ataque de pragas que podem causar prejuízos significativos ao produtor. A praga que mais preocupa os cafeicultores brasileiros é o bicho-mineiro do cafeeiro (*Lepidoptera: Lyonetiidae*). Uma maneira de efetuar o controle biológico do bicho-mineiro é através de vespas predadoras (*Hymenoptera: Vespidae*). Pesquisadores afirmam que para realizar esse tipo de controle é importante conhecer a distribuição espacial do bicho-mineiro e das vespas. Também existem evidências que a distribuição espacial dessas espécies pode ser influenciada pelas condições climáticas. Este trabalho teve por objetivo testar se existe diferença entre as distribuições espaciais de

---

<sup>1</sup> Universidade Federal de Uberlândia/UFU - Faculdade de Ciências Integradas do Pontal/FACIP - Rua 20, 1600 – Bairro Tupã - 38.304-402 - Ituiutaba - MG - franciella@pontal.ufu.br

<sup>2</sup> Universidade Federal de Uberlândia/UFU - Faculdade de Ciências Integradas do Pontal/FACIP - Rua 20, 1600 – Bairro Tupã - 38.304-402 - Ituiutaba - MG - gabriellaalves@pontal.ufu.br

<sup>3</sup> Universidade Federal de Lavras/UFLA - Departamento de Ciências Exatas/DEX - Cx. P. 3037 - 37200-000 - Lavras - MG - scalon@dex.ufla.br

<sup>4</sup> Embrapa Café - EPAMIG/EcoCentro, Campus da UFLA s/n, Cx. P. 176 - 37200-000 - Lavras – MG - mauricio.zacarias@embrapa.br

folhas minadas, minas predadas e vespas em diferentes situações climáticas. Os dados utilizados foram obtidos no ano de 2006 em um hectare de plantação de café orgânico, situado no município de Santo Antônio do Amparo-MG. Para testar a hipótese de igualdade entre duas distribuições espaciais foi aplicado o teste do Syrjala. Os resultados obtidos foram: não houve diferença na distribuição espacial de folhas minadas e de minas predadas entre o período seco e chuvoso, folhas minadas e vespas apresentaram mesma distribuição espacial tanto no período seco quanto no mês de maior infestação da praga, as distribuições de minas predadas e vespas foram diferentes tanto no período seco quanto no mês de maior infestação da praga. Observa-se que não houve alteração na distribuição espacial do bicho-mineiro do período chuvoso para o seco. A igualdade na distribuição espacial de folhas minadas e vespas predadoras, sugere que as vespas procuram naturalmente o bicho-mineiro.

**Termos para indexação:** Teste Syrjala, controle biológico, café orgânico.

#### **STATISTICAL COMPARISON OF THE SPATIAL DISTRIBUTION OF BOTH COFFEE LEAF MINER AND PREDATORY WASPS**

**ABSTRACT:** The coffee plant is subject to pests that can cause considerable damage to the producer. The pest that concerns most the Brazilian coffee growers is the coffee leaf miner (Lepidoptera: Lyonetiidae). One way to make the biological control of the coffee leaf miner is through predatory wasps (Hymenoptera: Vespidae). Researchers claim that for performing this type of control is important to know the spatial distribution in the coffee plantation of both the coffee leaf miner and the predatory wasp. There is also evidence that the spatial distribution of these species can be influenced by weather conditions. The aim of this work was to test whether or not there is a difference between the

spatial distributions of mined leaves, predated mines and predatory wasps in different weather conditions. Data were obtained in 2006 from one hectare of organic coffee production, located in the county of Santo Antônio do Amparo, Brazil. To test the hypothesis of equality between two spatial distributions it was applied the Syrjala's test. The results were: there was no difference between the spatial distribution of both mined leaves and predated mines during the dry and rainy seasons; mined leaves and predatory wasps showed the same spatial distribution for both in the dry and in the month of peak of the pest infestation; spatial distributions of predated mines and predatory wasps were different for both in the dry and in the month of greatest pest infestation. It is observed that there was no change in the spatial distribution of the leaf miner from the rain to the dry season. Equality in the spatial distribution of mined leaves and predatory wasps, suggests that wasps naturally seek the leaf miner.

**Index terms:** Syrjala's test, biological control, organic coffee.

## 1 INTRODUÇÃO

A cafeicultura é uma atividade de grande importância para a economia brasileira, sendo o Brasil o maior produtor e exportador mundial de café. De acordo com a terceira estimativa de produção de café (arábica e conilon) divulgada pela Companhia Nacional de Abastecimento - (CONAB, 2013) a colheita na safra de 2013, no Brasil, deverá ser de 47,54 milhões de sacas de 60 quilos de café beneficiado. Se comparada com a produção anterior, esta estimativa representa uma redução de 6,46%. Houve uma redução maior na produção de café arábica, devido principalmente a baixa bienalidade na maioria das áreas de café arábica e as chuvas bastante irregulares aliado às altas temperaturas. Segundo Mendonça et al. (2011), a bienalidade é a variação de

anos com alta e baixa produção, sendo mais intensa no café arábica, embora também ocorra no café conilon.

Minas Gerais é o estado brasileiro que possui a maior área plantada com café e lideram a sua produção no Brasil. Segundo o terceiro levantamento para a safra de 2013, de café beneficiado, realizado pela CONAB (2013), Minas Gerais possui 199.581,0 hectares de área em formação e 1.037.348,0 hectares de área em produção, totalizando uma produção de 26.158,0 sacas de café, das quais 25.872,0 são da espécie arábica. Embora a cafeicultura continue com grande potencial de crescimento, também há fatores limitantes, como as pragas, que têm sido um dos principais problemas que afetam a produtividade dessas lavouras. De uma forma geral, a praga que mais preocupa os cafeicultores brasileiros é o bicho-mineiro do cafeeiro, *Leucoptera coffeella* (Guérin-Mèneville & Perrottet, 1842) (Lepidoptera: Lyonetiidae).

O bicho-mineiro do cafeeiro em sua fase adulta é uma mariposa com 6,5 mm de envergadura de coloração branco-prateada. A mariposa deposita seus ovos na parte superior das folhas e as lagartas ao eclodirem passam diretamente para o interior das folhas, formando as lesões ou minas. Tais lesões provocam uma redução na área foliar e queda de folhas o que leva a uma diminuição na fotossíntese ocasionando uma queda de produção (REIS; SOUZA; ZACARIAS, 2006).

O combate ao bicho-mineiro é normalmente realizado por meio do controle químico. Em cultivos orgânicos essa prática é restringida ou proibida, tornando o controle biológico uma tática chave. De acordo com Reis, Souza e Zacarias (2006), o controle biológico se dá por meio de parasitóides (microhimenópteros), vespas predadoras e patógenos.

Segundo Reis, Souza e Zacarias (2006), vespas predadoras (Hymenoptera:Vespidae) são consideradas os principais agentes no controle biológico do bicho-mineiro. Entretanto, os ninhos de vespas, muitas vezes, são

destruídos pelos trabalhadores rurais, pois as mesmas são agressivas, o que pode causar acidentes. Assim, para o controle biológico é de extrema importância preservar os ninhos das vespas em locais próximos aos cafezais.

Dentro do contexto do controle biológico do bicho-mineiro, tendo em vista a produção orgânica, Scalon et al. (2011a, 2011b) afirmam que o conhecimento do padrão da distribuição espacial da praga em cultivos de café pode ser um importante aliado do produtor. Esses autores também sugerem que possa existir uma relação entre a localização do café e a presença do bicho-mineiro, uma vez que é possível que algumas regiões da cafeicultura ou do cafezal possam combinar, de alguma maneira, as melhores condições para o desenvolvimento da praga.

Sabe-se também que a ocorrência de maiores ou menores infestações do bicho-mineiro está muito relacionada a fatores climáticos tais como a temperatura e a umidade relativa do ar (REIS; SOUZA; ZACARIAS, 2006). Uma questão básica que ainda não foi respondida é como a distribuição espacial da ocorrência de bicho-mineiro se comporta em duas situações climáticas distintas (períodos de seca e chuva). Também seria importante verificar se existe diferença entre as distribuições espaciais do bicho-mineiro e das vespas predadoras.

Syrjala (1996) propôs um teste que permite verificar estatisticamente a diferença existente entre as distribuições espaciais de duas populações. Esse teste tem sido utilizado com sucesso para testar se existe diferença na distribuição espacial de espécies de presas e predadores (SWAIN; WADE, 2003), para avaliar a distribuição de uma espécie em diferentes períodos de tempo (HEDGER et al., 2004, MA et al., 2012), para caracterizar distribuições espaciais de pragas em tomateiros e, assim, auxiliar o controle biológico (CHAPMAN et al., 2009) e para caracterizar distribuição espacial de parasitas de aves (CORNUAULT et al., 2013), para citar apenas alguns. No melhor do

nosso conhecimento, o método proposto por Syrjala (1996) ainda não foi aplicado para avaliar distribuições espaciais de presas e predadores encontrados na cafeicultura.

O objetivo deste trabalho é utilizar a metodologia proposta por Syrjala (1996) para analisar as distribuições espaciais do bicho-mineiro do cafeeiro e das vespas predadoras. Para tanto, será verificado se existe diferença entre as distribuições espaciais de folhas minadas e também de minas predadas em duas situações climáticas distintas. Será verificado ainda a existência de diferença entre a distribuição espacial de folhas minadas e vespas e minas predadas e vespas no período seco do ano de 2006. Será analisado também se existe diferença entre folhas minadas e vespas e minas predadas e vespas no mês de setembro, que foi o mês de maior infestação de bicho-mineiro no ano.

## **2 MATERIAL E MÉTODOS**

### **Área Experimental**

Neste trabalho foram analisados os dados obtidos de um hectare de produção de café orgânico da Cultivar Catucaí, situado no município de Santo Antônio do Amparo-MG. A coleta de dados foi realizada, mensalmente, de janeiro à dezembro de 2006, em 35 pontos de amostragem cujas coordenadas, próximas de uma grade regular, foram obtidas por GPS. Em cada ponto de amostragem foram coletadas dez folhas de cafeeiro no terço médio da planta para contagem e identificação de folhas minadas e minas predadas. A contagem das folhas minadas e minas predadas foi realizado no laboratório da Epamig-CTSM/EcoCentro, localizado em Lavras-MG. Os dados consistiram da quantidade total de folhas minadas em cada ponto de amostragem.

Para a captura das vespas predadoras instalou-se nos 35 pontos de

amostragem armadilhas adesivas amarelas retangulares (24 x 9,5 cm), que ficaram penduradas por uma semana em estacas na altura da planta. Depois de capturadas as vespas foram todas identificadas no laboratório da Epamig-CTSM/EcoCentro, em Lavras, MG.

A partir das informações pluviométricas da região do estudo no ano de 2006 foi possível classificar os meses de janeiro a março e outubro a dezembro como período chuvoso do ano, e os meses de abril a setembro como o período seco no ano. O número de folhas minadas e minas predadas para cada um dos períodos foi obtido, respectivamente, a partir da soma do número de folhas minadas e minas predadas, por ponto de amostragem, nos meses considerados. Para comparação das distribuições espaciais das folhas minadas e das vespas e minas predadas e vespas utilizaram-se a contagem de folhas minadas, minas predadas e vespas para o período seco e para o mês de maior infestação do bicho-mineiro, que foi setembro de 2006. Mesmo o mês de setembro sendo um dos meses do período seco, realizou-se análises utilizando as contagens de folhas minadas, minas predadas e vespas para o mesmo, devido ao fato de este ser o mês de maior infestação do ano de 2006.

### **Teste para a diferença entre duas distribuições espaciais**

Neste teste proposto por Syrjala (1996), a hipótese nula ( $H_0$ ) a ser testada é que a distribuição espacial das duas populações é a mesma, enquanto a hipótese alternativa ( $H_1$ ) é que existe alguma diferença entre as duas distribuições espaciais das populações em estudo.

Para conduzir o teste, considere um retângulo sobre a área em estudo e um sistema de coordenadas cartesianas sobreposto a este retângulo com a origem em um canto arbitrário do retângulo. Sejam  $(x_k, y_k)$  e  $d_i(x_k, y_k)$ , respectivamente, as coordenadas do  $k$ -ésimo local de amostragem e a densidade

da amostra no  $k$ -ésimo local de amostragem da  $i$ -ésima população, em que  $\{k = 1, \dots, K; i = 1, 2\}$ .

Com o objetivo de remover o efeito dos tamanhos das populações as densidades observadas (contagens), em cada local de amostragem, são normalizadas da seguinte forma,

$$\gamma_i(x_k, y_k) = \frac{d_i(x_k, y_k)}{D_i},$$

em que,

$$D_i = \sum_{k=1}^K d_i(x_k, y_k)$$

Somando todas as densidades observadas normalizadas, cuja localização  $(x, y)$  é tal que  $x \leq x_k$  e  $y \leq y_k$ , obtém-se a função de distribuição acumulada dada por

$$\Gamma_i(x_k, y_k) = \sum_{\forall x \leq x_k, \forall y \leq y_k} \gamma_i(x, y). \quad (1)$$

Syrjala (1996) propõe a estatística para testar a hipótese nula como sendo

$$\Psi = \sum_{k=1}^K [\Gamma_1(x_k, y_k) - \Gamma_2(x_k, y_k)]^2. \quad (2)$$

Considerando que  $\Psi$  depende da origem do sistema de coordenadas dentro da área e considerando que esta área é retangular e que a origem pode ser em qualquer um dos cantos do retângulo, a estatística do teste pode ser obtida pela média dos valores obtidos para cada um dos cantos da área retangular (SYRJALA, 1996). Assim, as equações (1) e (2) podem ser reescritas, respectivamente, como

$$\Gamma_i(x_{c,k}, y_{c,k}) = \sum_{\forall x \leq x_{c,k}, \forall y \leq y_{c,k}} \gamma_i(x, y) \quad (3)$$

e

$$\Psi_c = \sum_{k=1}^k [F_1(x_{c,k}, y_{c,k}) - F_2(x_{c,k}, y_{c,k})]^2 \quad (4)$$

em que,  $c = \{1,2,3,4\}$  identifica o canto da área tomado como origem dos eixos coordenados,  $(x_{c,k}, y_{c,k})$  é a coordenada do  $k$ -ésimo local de amostragem localizada em relação a origem no canto  $c$  e  $F_i(x_{c,k}, y_{c,k})$  é a  $i$ -ésima função de distribuição acumulada obtida com a coordenada do local amostrado em relação ao canto  $c$ .

Assim, a estatística do teste é dada por (SYRJALA, 1996)

$$\Psi = \frac{1}{4} \sum_{c=1}^4 \Psi_c. \quad (5)$$

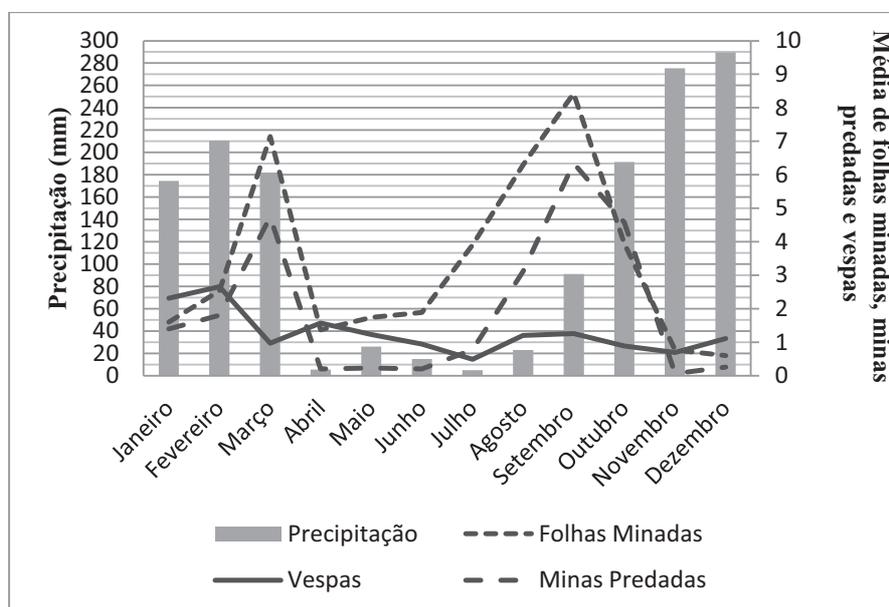
O estimador  $\Psi$  não tem distribuição amostral conhecida e, portanto, valor-p da estatística da equação (5) pode ser obtido através de aleatorização. Sob a hipótese nula, a distribuição da estatística de teste é construída por meio do cálculo do valor da estatística de teste para  $n$  permutações do conjunto de dados. O valor-p é obtido pela proporção dos valores obtidos em (5), nas  $n$  permutações, que são maiores ou iguais a estatística de teste observada (SYRJALA, 1996). Neste trabalho foram utilizadas 1000 permutações.

Todos os cálculos foram realizados utilizando o *software* R (R DEVELOPMENT CORE TEAM, 2013) conjuntamente com sua biblioteca ECESPA (DE LA CRUZ, 2008).

### 3 RESULTADOS E DISCUSSÃO

A precipitação pluviométrica mensal, as médias mensais do número de folhas minadas, minas predadas e vespas para os meses de janeiro de 2006 a dezembro de 2006 são representadas na Figura 1. Pode-se observar que os picos de folhas minadas para o ano de 2006 ocorreram nos meses de março e setembro. O pico de setembro já era esperado por este ser um mês seco, o que

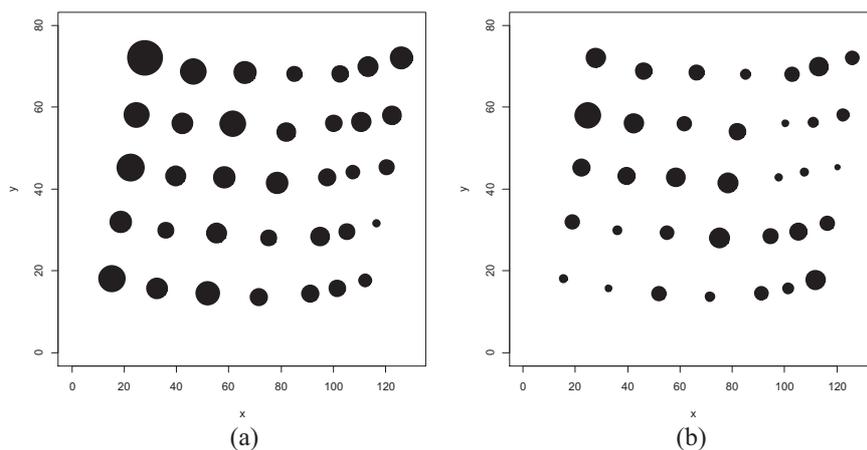
influencia o aparecimento do bicho-mineiro. Segundo Reis, Souza e Zacarias (2006) também pode ocorrer um aumento na população desta praga nos meses de março e abril devido à presença de veranicos, o que deve ter ocasionado o aumento no número de folhas minadas no mês de março. Observa-se ainda que os picos de minas predadas no ano de 2006 ocorreram em março e setembro, coincidindo com os picos de folhas minadas.



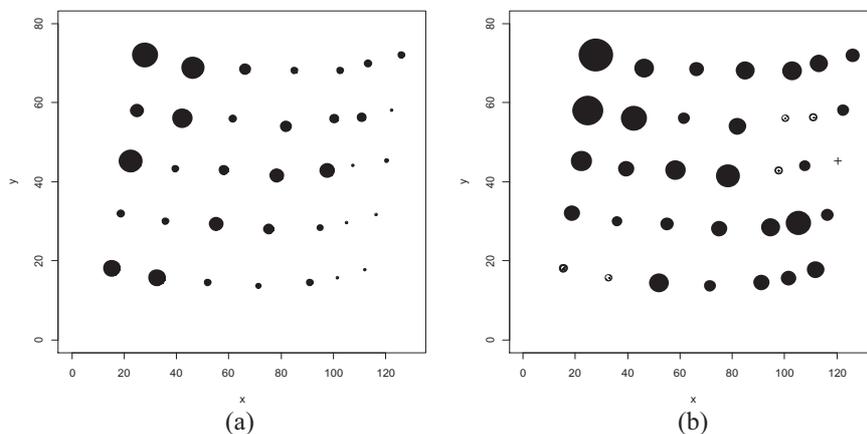
**FIGURA 1:** Gráfico da precipitação, média mensal de folhas minadas, minas predadas e vespas no ano de 2006.

Podemos observar na figura 2 o número de folhas minadas e na figura 3 o número de minas predadas, nos 35 locais de amostragem para o período de seca (a) e de chuva (b) do ano de 2006. Como esperado, o número de folhas minadas foi maior no período de seca, visto que as chuvas influenciam negativamente no aparecimento da praga. Já, com relação às minas predadas ocorreu o inverso. No caso dessa última variável, as vespas podem necessitar da precipitação como recurso de água, diminuindo suas atividades ou mesmo

apresentando redução populacional na seca. Outra possibilidade, não abordada no presente estudo, pode ser que as baixas temperaturas venham a reduzir o período de atividade de forrageamento diário.

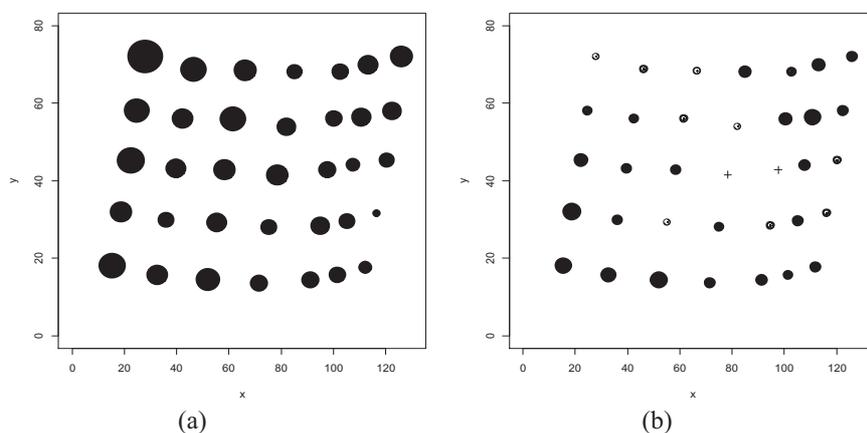


**FIGURA 2:** Número de folhas minadas nos 35 pontos de amostragem para os períodos de seca (a) e chuva (b) no ano de 2006. As áreas dos círculos são proporcionais às frequências absolutas observadas. A menor área corresponde à frequência absoluta igual a 5, enquanto a maior área corresponde à frequência absoluta igual a 43.

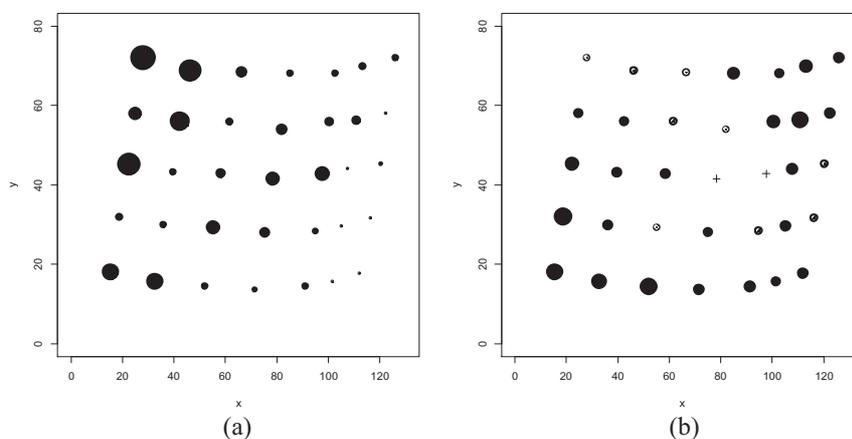


**FIGURA 3:** Número de minas predadas nos 35 pontos de amostragem para os períodos de seca (a) e chuva (b) no ano de 2006. A menor frequência absoluta é igual a 1 e a maior frequência absoluta é igual a 36. O símbolo + indica que não houve ocorrência no ponto de amostragem correspondente.

Ao se compararem os números de folhas minadas e o número de minas predadas com a ocorrência das vespas (figuras 4 e 5, respectivamente), podemos observar que os números de folhas minadas e minas predadas são bem maiores que a quantidade de vespas. As vespas apresentam grande mobilidade e podem alcançar diferentes distâncias para encontrar seu alimento, sendo isto uma provável causa dessa não coincidência. Embora em ambas as situações sejam claras as diferenças nas quantidades, não é possível tirar conclusões a respeito da existência de diferença entre as distribuições espaciais.

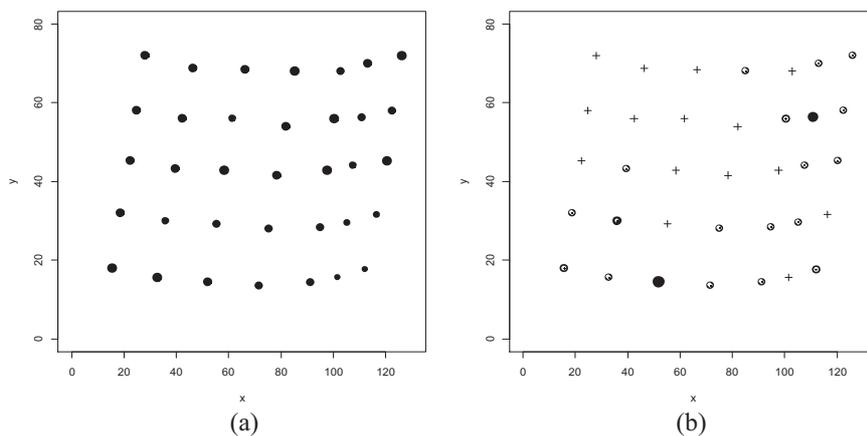


**FIGURA 4:** Número de folhas minadas (a) e vespas (b) nos 35 pontos de amostragem para o período seco. A menor frequência absoluta é igual a 8 e a maior frequência absoluta é igual a 43. O símbolo + indica que não houve ocorrências nos pontos de amostragem correspondentes.

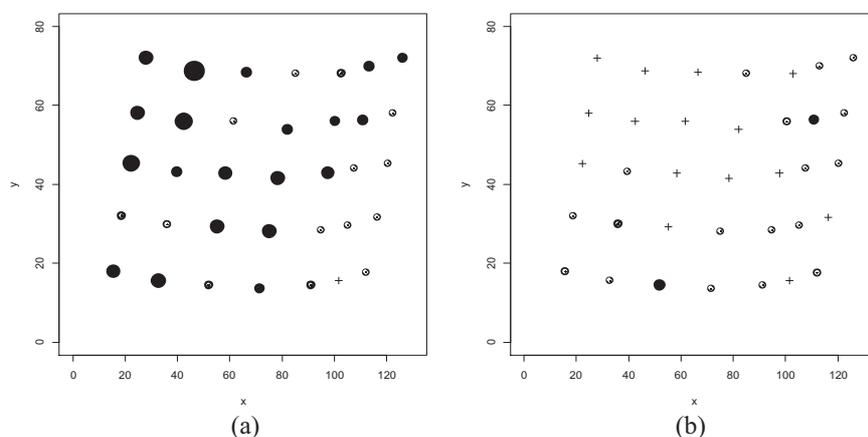


**FIGURA 5:** Número de minas predadas (a) e vespas (b) nos 35 pontos de amostragem para o período seco. A menor frequência absoluta é igual a 1 e a maior frequência absoluta é igual a 30. O símbolo + indica que não houve ocorrências nos pontos de amostragem correspondentes.

Confrontando-se o número de folhas minadas e o número de minas predadas com o número de vespas capturadas (Figuras 6 e 7, respectivamente) para o mês de setembro (mês de maior infestação de bicho-mineiro no ano de 2006), observa-se resultado semelhante ao do período seco como um todo (figuras 4 e 5).



**FIGURA 6:** Número de folhas minadas (a) e vespas (b) nos 35 pontos de amostragem para o mês de setembro. A menor frequência absoluta é igual a 5 e a maior frequência absoluta é igual a 10. O símbolo + indica que não houve ocorrências nos pontos de amostragem correspondentes.



**FIGURA 7:** Número de minas predadas (a) e vespas (b) nos 35 pontos de amostragem para o mês de setembro. A menor frequência absoluta é igual a 7 e a maior frequência absoluta é igual a 19. O símbolo + indica que não houve ocorrências nos pontos de amostragem correspondentes.

Como se trata de um grande número de espécies de vespas (MARAFELLI et al., 2007), provavelmente uma ou mais espécies ocorrem em maiores quantidades em períodos diferentes do ano. Algumas espécies podem preferir (ou serem mais eficientes) o período seco para o forrageamento, enquanto outras podem preferir o chuvoso.

A inspeção visual das figuras de 2 a 7 não permite identificar claramente se as distribuições espaciais são iguais ou distintas. Para uma caracterização estatística sobre a igualdade das distribuições espaciais em questão, utilizou-se o teste de Syrjala (1996) que foi apresentado anteriormente e cujos resultados estão apresentados na Tabela 1.

**Tabela 1:** Estatística de teste e valores-p para o teste da hipótese nula de que as distribuições espaciais das duas variáveis no ano de 2006 são iguais.

Variáveis	Estatística ( $\Psi$ )	Valor - p
Folhas minadas no período chuvoso e seco	0,02	0,31
Minas predadas no período chuvoso e seco	0,13	0,16
Folhas minadas e vespas predadoras no período seco	0,12	0,06
Minas predadas e vespas predadoras no período seco	0,39	0,01
Folhas minadas e vespas predadoras em setembro	0,37	0,15
Minas predadas e vespas predadoras em setembro	0,83	0,04

A tabela 1 mostra que o teste de Syrjala levou a não rejeição da hipótese nula de que as distribuições espaciais são iguais, ao nível de 5% de significância, para as seguintes comparações: folhas minadas no período chuvoso e seco; minas predadas no período chuvoso e seco; folhas minadas e vespas predadoras no período seco e, finalmente, folhas minadas e vespas predadoras no mês de maior infestação do bicho-mineiro (setembro).

Embora as chuvas influenciam negativamente o aparecimento da praga, pode-se observar que a maneira como o bicho-mineiro se distribui no cafezal não se altera do período chuvoso para o período seco. A igualdade na distribuição espacial de folhas minadas e vespas predadoras pode sugerir que as vespas predadoras procuram naturalmente os locais onde se encontram o bicho-mineiro. Sendo assim, pode-se pensar em manter ninhos de vespas em locais próximos ou dentro dos cafezais não sendo necessário transferir ninhos de vespas para os locais onde há maior concentração de bicho-mineiro.

A tabela 1 mostra que o teste de Syrjala levou a rejeição da hipótese nula de que as distribuições espaciais são iguais, ao nível de 5% de significância, para

a comparação entre minas predadas e vespas predadoras no período seco e para a comparação entre minas predadas e vespas predadoras no mês de setembro, ou seja, o mês de maior infestação do bicho-mineiro.

No período de 2004 à 2006, Marafeli et al. (2007) já haviam sido identificadas 14 espécies de vespas predadoras no mesmo local, sendo que apenas três espécies apresentaram correlação positiva significativa com o número de minas predadas.

Isso pode explicar a diferença no tipo de distribuição espacial obtido, visto que as demais espécies de vespas podem estar no agroecossistema cafeeiro por outras fontes de alimento ou mesmo pelo bicho-mineiro não ser o seu único alimento forrageado.

#### 4 CONCLUSÕES

Por meio da aplicação do teste do Syrjala, pode-se observar que a distribuição espacial do bicho-mineiro no cafezal não se altera do período chuvoso para o período seco. Observou-se também a igualdade na distribuição espacial de folhas minadas e vespas predadoras, o que pode sugerir que as vespas predadoras procuram naturalmente os locais onde se encontram o bicho-mineiro. Espera-se que os resultados obtidos neste trabalho possa contribuir para um melhor entendimento da distribuição espacial do bicho-mineiro e das vespas predadoras em cafezais de produção orgânica.

#### 5 REFERÊNCIAS

CHAPMAN, A. V. et al. Dispersal of *Trichogramma ostrinae* (Hymenoptera: Trichogrammatidae) in potato field. **Environmental Entomology**, College Park, v. 38, n. 3, p. 677-685, 2009.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira café safra 2013 terceira estimativa, setembro/2013.**

Disponível em:

<[http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13\\_09\\_09\\_15\\_34\\_48\\_boletim\\_cafe\\_-\\_setembro\\_2013.pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13_09_09_15_34_48_boletim_cafe_-_setembro_2013.pdf)>. Acesso em: 10 dez. 2013.

CORNUAULT, J. et al. The role of ecology in the geographical separation of blood parasites infecting an insular bird. **Journal of Biogeography**, Oxford, v. 40, n. 7, p. 1313-1323, 2013.

DE LA CRUZ, M. Metodos para analizar datos puntuales. In: MAESTRE, F. T.; ESCUDERO, A.; BONET, A. (Ed.). **Introduccion al analisis espacial de datos em ecologia y ciencias ambientales: metodos y aplicaciones**. Madrid: Asociacion Espanola de Ecologia Terrestre; Universidad Rey Juan Carlos y Caja de Ahorros del Mediterraneo, 2008. p. 76-127.

HEDGER, R. et al. Analysis of the spatial distributions of mature cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) abundance in the North Sea (1980-1999) using generalized additive models. **Fisheries Research**, New York, v. 70, p. 17-25, 2004.

MA, Z. et al. Use of localized descriptive statistics for exploring the spatial pattern changes of bird species richness at multiple scales. **Applied Geography**, Oxford, v. 32, p. 185-194, 2012.

MARAFELI, P. P. et al. Ocorrência e identificação de vespas predadoras (Hymenoptera: Vespidae) em cafezal orgânico em formação (*Coffea arabica* L.) e sua relação com a predação do bicho mineiro, *Leucoptera coffeella* (Guér.-Mènev., 1942) (Lepidoptera: Lyonetiidae). In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 5., 2007, Águas de Lindóia. **Anais**. Instituto Agrônomo de Campinas. Águas de Lindóia, 2007. p. 1-4.

MENDONÇA, R. F. et al. Abordagem sobre a bienalidade de produção em plantas de café. **Enciclopédia Biosfera**, Goiânia, v. 7, n. 13, p. 1-9, 2011.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 out. 2013.

REIS, P. R.; SOUZA, J. C.; ZACARIAS, M. S. Alerta para o bicho-mineiro. **Cultivar**, Pelotas, v. 8, p. 13-16, 2006.

SCALON, J. D. et al. Spatial and temporal dynamics of coffee-leaf-miner and predatory wasps in organic coffee field in formation. **Ciência Rural**, Santa Maria, v. 41, n. 4, p. 646-652, 2011a.

SCALON, J. D. et al. Spatial distribution of the coffee-leaf-miner (*Leucoptera coffeella* (Guérin-Mèneville&Perrottet, 1842)) in an organic coffee (*Coffea Arabica* L.) field in formation. **Coffee Science**, Lavras, v. 6, n. 3, p. 226-232, set./dez. 2011b.

SWAIN, D. P.; WADE, E. J. Spatial distribution of catch and effort in a fishery for snow crab (*Chionoecetes opilio*): tests of predictions of the ideal free distribution. **Canadian Journal of Fishery and Aquatical Science**, Ottawa, v. 60, p. 897-909, 2003.

SYRJALA, S. E. A statistical test for a difference between the spatial distributions of two populations. **Ecology**, Durham, v. 77, n. 1, p. 75-80, 1996.

## ARTIGO 2: TESTES PARA VERIFICAR A SIGNIFICÂNCIA ESTATÍSTICA DE CORRELOGRAMAS

Franciella Marques da Costa<sup>1</sup>, João Domingos Scalon<sup>2</sup>

**RESUMO:** Uma questão fundamental no estudo de fenômenos espaciais é a caracterização da dependência espacial. O índice global de Moran é uma medida amplamente utilizada para estimar a autocorrelação espacial para dados de áreas. O correlograma de Moran pode ser utilizado para avaliar a dependência espacial considerando várias ordens de vizinhança. O objetivo deste trabalho é propor testes estatísticos para determinar a significância de um correlograma, ou seja, testar a hipótese nula de inexistência de autocorrelação espacial de uma variável aleatória. Para tal, constrói-se o correlograma considerando: vizinhança de primeira ordem, diferentes distâncias e estatísticas de testes baseadas nas distâncias de Cramer-von Mises e Kolmogorov. Análises do poder dos testes e do erro do tipo I mostraram que os testes propostos são adequados. Também foi proposto um envelope de simulação para verificar se um correlograma é estatisticamente significativo. Os métodos propostos foram aplicados para análise da autocorrelação espacial do número de minas predadas do bicho-mineiro do cafeeiro.

**Palavras-chave:** Índice de Moran. Aleatorização. Teste de hipótese. Café orgânico.

---

<sup>1</sup>Universidade Federal de Uberlândia/UFU - Faculdade de Ciências Integradas do Pontal/FACIP - Rua 20, 1600 – Bairro Tupã - 38.304-402 - Ituiutaba - MG - franciella@pontal.ufu.br

<sup>2</sup>Universidade Federal de Lavras/UFLA - Departamento de Ciências Exatas/DEX - Cx. P. 3037 - 37200-000 - Lavras - MG - scalon@dex.ufla.br

**ABSTRACT:** A key issue in the study of spatial phenomena is the characterization of the spatial dependence. The global Moran's index is a measure widely used to estimate the spatial autocorrelation for areal data. The Moran's correlogram can be used to evaluate the spatial dependence considering several orders of neighborhood. The aim of this work is to propose statistical tests to determine the significance of a correlogram, that is, testing the null hypothesis of no spatial autocorrelation of a random variable. Therefore, the correlogram is constructed considering: first-order neighbors, distances at different scales and statistical tests based on both Cramer-von Mises and Kolmogorov distances. Power tests and type I error analyzes were performed showed that the proposed tests are adequate. It has also been proposed an envelope simulation to verify that a correlogram is statistically significant. The proposed methods were applied to analyze the spatial autocorrelation of the number of preyed mines of the coffee leaf-miner.

**Keywords:** Moran's Index. Randomization. Hypothesis test. Organic coffee.

## 1 INTRODUÇÃO

Os fenômenos naturais apresentam, em geral, algum grau de dependência espacial. Desse modo, uma questão fundamental no desenvolvimento de análises envolvendo tais fenômenos é a caracterização dessa dependência espacial.

O índice global de Moran é uma medida amplamente utilizada para estimar a autocorrelação espacial para dados de áreas. Esse índice retorna um único valor que representa a autocorrelação espacial global para toda a região em estudo, sendo de fundamental importância para a caracterização da dependência espacial da região como um todo (MORAN, 1950).

Após calcular o índice de Moran, pode-se aplicar um teste de hipótese para verificar se a autocorrelação espacial é estatisticamente significativa. A hipótese nula de independência espacial pode ser testada utilizando duas abordagens: a distribuição aproximada do índice de Moran ou o teste de permutação aleatória, conforme descritas em Bailey e Gatrell (1995).

Caso exista o interesse em analisar a autocorrelação espacial para diferentes ordens de vizinhança pode-se usar o correlograma, ou seja, a representação gráfica do índice de Moran calculado para diferentes ordens de vizinhança  $k$  em função de  $k$ .

Em geral, a interpretação do correlograma é feita de forma visual e, portanto, subjetiva, o que não é a melhor maneira sob o ponto de vista científico. Entretanto, existem algumas propostas estatísticas para testar a significância estatística de um correlograma conforme podem ser vistas em Oden (1984). Uma maneira de testar a significância estatística de um correlograma é utilizando o critério de Bonferroni, em que o correlograma é considerado estatisticamente significativo ao nível de significância  $\alpha$ , quando, pelo menos, um dos coeficientes de Moran for estatisticamente significativo, a um nível de significância  $\alpha/k$ , sendo  $k$  o número de classes consideradas na obtenção do correlograma.

Com o intuito de obter informações a respeito da dependência espacial de um conjunto de dados para diferentes escalas, seria interessante construir o correlograma considerando apenas o vizinho de primeira ordem, porém alterando as distâncias consideradas nessa primeira ordem. Dessa forma pode-se avaliar o comportamento da dependência espacial à medida que se aumenta a distância envolvida, ou seja, como se dá o "*enfraquecimento*" da dependência espacial.

O objetivo deste trabalho é propor dois testes de hipótese e um envelope de simulação para testar a significância estatística de um correlograma obtido a

partir do índice global de Moran considerando o primeiro vizinho, ou seja, testar a hipótese nula de inexistência de autocorrelação espacial de uma variável aleatória para diferentes distâncias. Os métodos propostos neste trabalho serão submetidos ao estudo de poder e do erro do tipo I e aplicados a um conjunto de dados referentes ao bicho-mineiro do cafeeiro (*Leucoptera coffeella*).

## 2 MATERIAL E MÉTODOS

### 2.1 Índice global de Moran e correlograma

Considerando que o processo estocástico é estacionário de segunda ordem, o índice global de Moran (1950) para o primeiro vizinho pode ser calculado por

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i=1}^N (Y_i - \bar{Y})^2) (\sum_{i=1}^N \sum_{j=1}^N w_{ij})} \quad (1)$$

em que,  $N$  é o número de áreas,  $Y_i$  é o valor do atributo na área  $i$ ,  $Y_j$  é o valor do atributo na área  $j$ ,  $\bar{Y}$  é a média do atributo e  $w_{ij}$  são os elementos da matriz de proximidade espacial.

Neste trabalho será utilizada a matriz de proximidade espacial dada por  $w_{ij} = 1$  se  $d_{ij} < d$  em que  $d$  é uma distância definida pelo pesquisador e  $d_{ij}$  é a distância euclidiana entre os centroides das regiões  $i$  e  $j$  e  $w_{ij} = 0$  caso contrário.

O índice global de Moran mede a autocorrelação (dependência) espacial através de um único valor que, geralmente, assume valores no intervalo  $-1 \leq I \leq 1$  (UPTON; FINGLETON, 1985; WALLER; GOTWAY, 2004).

A interpretação do índice global de Moran é bastante simples. Se  $I$  for igual a zero, não existe autocorrelação espacial, caso contrário, existe autocorrelação espacial. Se  $I$  for positivo significa que as regiões vizinhas possuem valores semelhantes, enquanto se  $I$  for negativo implica que as regiões vizinhas tendem a ter valores diferentes. Quanto mais próximo de  $|1|$ , mais forte é a autocorrelação espacial (UPTON; FINGLETON, 1985; WALLER; GOTWAY, 2004).

Considerando que a região de estudo apresenta um número moderado de áreas e sob a suposição de que os dados possuem distribuição normal, pode-se testar a significância estatística do índice de Moran utilizando aproximação normal. Nos casos em que essas condições não são satisfeitas, pode-se utilizar um teste de permutação aleatória (BAILEY; GATRELL, 1995).

Segundo Bailey e Gatrell (1995), a representação gráfica do índice global de Moran, obtida para diferentes ordens de vizinhança  $k$ , em função de  $k$  é chamada de correlograma.

Alternativamente, como propõe este trabalho, o correlograma pode ser obtido a partir do índice global de Moran, calculado para diferentes distâncias, considerando sempre o vizinho de primeira ordem.

## **2.2 Testes de hipótese para verificar a significância estatística de correlogramas**

Os testes propostos têm por objetivo verificar a significância global de um correlograma. As hipóteses consideradas nos testes são:

$$\begin{cases} H_0: \nexists \text{ autocorrelação espacial nas distâncias analisadas} \\ H_1: \exists \text{ autocorrelação em, pelo menos, uma distância} \end{cases}$$

Para testar a hipótese nula, o primeiro passo é calcular todas as distâncias euclidianas entre os centroides de cada área. Em seguida, obtém-se a matriz de proximidade espacial. Neste trabalho, será utilizada a matriz de proximidade espacial dada por  $w_{ij} = 1$  se  $d_{ij} < d$  sendo  $d_{ij}$  a distância euclidiana entre os centroides das regiões  $i$  e  $j$ ,  $w_{ij} = 0$  caso contrário e  $d$  é uma distância definida pelo pesquisador.

A seguir, o índice global de Moran é calculado para diferentes distâncias ( $d$ ), para  $min < d < max$ , em que  $min$  = distância mínima entre os centroides e  $max$  = distância máxima entre os centroides. Finalmente, constrói-se o correlograma, ou seja, o gráfico dos índices globais de Moran em função das distâncias ( $d$ ).

Para testar a significância estatística do correlograma, propõem-se duas estatísticas de teste baseadas nas distâncias de Cramer-von Mises e de Kolmogorov.

A estatística de teste baseada na distância de Cramer-von Mises é dada por

$$C = \int (I(d) - E(I))^2 d(d), \quad (2)$$

em que  $I(d)$  é o índice global de Moran para cada distância e  $E(I)$  é o valor esperado para o índice de Moran na completa ausência de autocorrelação espacial, ou seja,  $E(I) = -1/(N - 1)$  para qualquer distância  $d$ .

A integral da equação (2) não pode ser calculada analiticamente, mas pode-se obtê-la aproximadamente por algum método numérico. Neste trabalho, utiliza-se a Regra de Simpson.

A estatística de teste baseada na distância de Kolmogorov é dada por

$$K = \max |I(d) - E(I)|, \quad (3)$$

em que  $I(d)$  e  $E(I)$  foram definidos anteriormente na equação (2).

Como as duas estatísticas de teste não possuem distribuições amostrais conhecidas, será utilizado o método de computação intensiva de aleatorização para obter a distribuição amostral das estatísticas de teste sob a hipótese nula. Por aleatorização, entende-se embaralhar os dados entre as áreas da região. O valor  $-p$  dos testes estatísticos é obtido pela proporção dos valores das estatísticas, obtidas por meio das aleatorizações, que são maiores ou iguais à estatística de teste calculada com os dados observados (MANLY, 1994).

### 2.3 Envelope Simulado

Os testes abordados anteriormente não identificam as distâncias em que o correlograma foi estatisticamente significativo. Para esse propósito, pode-se construir um envelope simulado. Bailey e Gatrell (1995) e Diggle (2003) descrevem a construção de envelopes de simulação para configurações pontuais, enquanto Diggle e Ribeiro Junior (2007) os utiliza para avaliar a dependência espacial em dados geoestatísticos. A mesma ideia será utilizada para testar a hipótese nula de que o correlograma não é estatisticamente significativo. O procedimento para a construção de envelopes simulados é descrito pelos seguintes passos:

- Constroem-se os correlogramas  $I(d)$  (correlograma para a população em estudo) e  $I'(d)$  (correlograma sob ausência de correlação espacial), e  $d$  são as distâncias consideradas;
- Obtém-se a diferença  $I_d(d) = I(d) - I'(d)$  para os dados observados;
- Permutam-se os dados observados  $m$  vezes e calcula-se  $I_i(d) = I_i(d) - I'_i(d)$  para  $i = 1, 2, 3, \dots, m$ ;
- Define-se o limite superior e inferior do envelope simulado como sendo:

$$I_S(d) = \max \{I_i(d)\}, i = 1, 2, 3, \dots, m$$

$$I_I(d) = \min \{I_i(d)\}, i = 1, 2, 3, \dots, m$$

em que  $I_S(d)$  e  $I_I(d)$  são, respectivamente, o limite superior e o limite inferior do envelope simulado.

- Obtém-se a representação gráfica com os valores  $I_d(d)$ ,  $I_S(d)$  e  $I_I(d)$  em função das distâncias  $d$ .

De acordo com Bailey e Gatrell (1995) e Diggle (2003), neste procedimento é válida a seguinte propriedade:

$$P\{I_d(d) > I_S(d)\} = P\{I_d(d) < I_I(d)\} = \frac{1}{m+1} \quad (4)$$

Vale ressaltar que, assim como nos testes apresentados na Seção 2.2, as distâncias ( $d$ ) devem ser definidas pelo pesquisador. Obtidos o envelope de simulação sua interpretação é bastante simples. O correlograma será estatisticamente significativo se a linha de  $I_d(d)$  não estiver totalmente contida entre o limite superior e o limite inferior do envelope simulado. Caso contrário o correlograma não é estatisticamente significativo.

#### 2.4. Estudo do erro do tipo I e do poder dos testes

Com o objetivo de avaliar o desempenho do teste  $C$  e do teste  $K$ , propostos neste trabalho, foi realizado um estudo por meio de simulação para estudar o poder e o erro do tipo I de cada teste.

Considerou-se diferentes situações de estruturas de dependências espaciais (fraca, moderada e forte) e dados sob completa aleatoriedade espacial (independência). Para simular dados com dependência espacial, utilizou-se a função *grf* da biblioteca *geoR* (DIGGLE; RIBEIRO JUNIOR, 2007; RIBEIRO

JUNIOR; DIGGLE, 2001) do *software* R (R CORE TEAM, 2014), utilizando um processo Gaussiano estacionário.

Utilizou-se a função de correlação esférica com patamar igual a 80 e diferentes valores de efeito pepita (1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71 e 76) e alcance (40 e 60). Para obter maiores detalhes sobre a função de correlação esférica e seus parâmetros ver Diggle e Ribeiro Junior (2007).

Para cada situação, foi simulada uma realização com 100 observações nas coordenadas dos cruzamentos de uma grade regular, em que as coordenadas dos cruzamentos representam os centroides de cada área. Foi considerada uma região com 10 x 10 unidades de áreas quadradas.

Deve-se observar que à medida que se aumenta o efeito pepita, mantendo os outros parâmetros da função de correlação esférica constantes, caminha-se de uma forte dependência espacial para uma fraca dependência espacial. Para medir esse grau de dependência espacial, pode-se utilizar o índice proposto por Cambardella et al. (1994), que é dado pela expressão:

$$IDE = \frac{C_0}{C_0 + C_1} \quad (5)$$

em que  $C_0$  é o efeito pepita e  $C_0 + C_1$  é o patamar.

O índice de Cambardella assume valores entre 0 e 1. Se o valor do índice IDE for menor que 0,25 significa dependência espacial forte, índice entre 0,25 e 0,75 caracteriza dependência espacial moderada e valores maiores que 0,75 indica dependência espacial fraca (CAMBARDELLA et al., 1994).

Para simular dados sob independência espacial, considerou-se uma região dividida em 10 x 10 unidades de área em uma grade regular, em que as coordenadas dos cruzamentos da grade regular representam os centroides de cada área. As observações foram obtidas de uma distribuição uniforme, simuladas utilizando a função *runif* do *software* R (R CORE TEAM, 2014).

Geralmente os dados de áreas são obtidos por meio de contagens, sendo assim, trabalhou-se com a parte inteira de cada valor simulado tanto pela função *grf* quanto pela função *runif*.

A Tabela 1 mostra as configurações utilizadas no processo de simulação considerando o parâmetro alcance da função de correlação esférica igual a 40. Considerou-se o mesmo plano de simulação mudando apenas o parâmetro alcance para 60. Considerando os parâmetros patamar e feito pepita fixo, a classificação quanto ao grau de dependência espacial para o alcance 40 e 60 é a mesma, visto que o índice proposto por Cambardella et al. (1994) não utiliza o parâmetro alcance. Além das configurações descritas na Tabela 1 foram simulados experimentos sob completa aleatoriedade espacial.

Foram simulados 1.000 experimentos para cada situação considerada. Para os dados sob completa aleatoriedade espacial ( $H_0$  verdadeira) obteve-se a proporção de vezes em que os testes indicaram a rejeição de  $H_0$ , ou seja, cometeu-se o erro do tipo I. Para os dados com dependência espacial ( $H_0$  falsa) calculou-se a proporção de vezes em que os testes rejeitaram  $H_0$ , ou seja, obteve-se o poder do teste.

Tabela 1 Configuração utilizada para simulação.

Nomenclatura	Efeito pepita	Patamar	Alcance	Grau de dependência espacial
Sph(1,80,40)	1	80	40	forte
Sph(6,80,40)	6	80	40	forte
Sph(11,80,40)	11	80	40	forte
Sph(16,80,40)	16	80	40	forte
Sph(21,80,40)	21	80	40	moderada
Sph(26,80,40)	26	80	40	moderada
Sph(31,80,40)	31	80	40	moderada
Sph(36,80,40)	36	80	40	moderada
Sph(41,80,40)	41	80	40	moderada
Sph(46,80,40)	46	80	40	moderada
Sph(51,80,40)	51	80	40	moderada
Sph(56,80,40)	56	80	40	moderada
Sph(61,80,40)	61	80	40	fraca
Sph(66,80,40)	66	80	40	fraca
Sph(71,80,40)	71	80	40	fraca
Sph(76,80,40)	76	80	40	fraca

## 2.5 Aplicação a dados reais

O conjunto de dados utilizados neste trabalho são provenientes de um hectare de produção de café orgânico da Cultivar Catucaí, situado no município de Santo Antônio do Amparo-MG. As observações referentes às minas predadas foram obtidas em 35 locais de amostragem, em que cada local representa o centro de uma área. A contagem das minas predadas foi realizada no laboratório da Epamig-CTSM/EcoCentro, em Lavras – MG. Foram utilizadas as observações referentes ao mês de setembro (mês de maior infestação do bicho-mineiro) de 2006.

## 2.6 Software

As análises apresentadas neste trabalho foram realizadas utilizando o *software* R (R CORE TEAM, 2014), usando as bibliotecas *geoR* (DIGGLE; RIBEIRO JUNIOR, 2007; RIBEIRO JUNIOR; DIGGLE, 2001) e *ape* (PARADIS; CLAUDE; STRIMMER, 2004), respectivamente, para a simulação de dados e cálculo do índice global de Moran.

# 3 RESULTADOS E DISCUSSÃO

## 3.1 Erro do tipo I e poder

Foram simulados 1.000 experimentos sob a hipótese de completa aleatoriedade espacial ( $H_0$  verdadeira) e aplicado os testes  $C$  e  $K$  para cada situação. Os resultados podem ser observados na Tabela 2.

Tabela 2 Proporção de rejeição de  $H_0$  sendo  $H_0$  verdadeira

Testes Propostos	Nível de Significância		
	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,1$
Teste $C$	0,042	0,095	0,109
Teste $K$	0,001	0,019	0,037

A proporção de rejeição de  $H_0$ , sendo  $H_0$  verdadeira, utilizando o teste  $C$  foi maior do que o nível de significância considerado ( $\alpha = 1\%$ ,  $\alpha = 5\%$ ,  $\alpha = 10\%$ ), enquanto que com a aplicação do teste  $K$  obteve-se uma proporção menor do que o nível de significância.

Na Figura 1, são apresentados os resultados do estudo do poder dos testes propostos, considerando o alcance igual a 40.

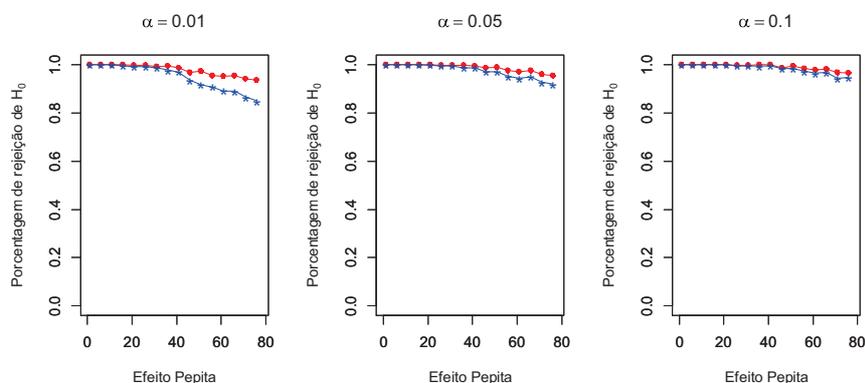


Figura 1 Porcentagem de rejeição de  $H_0$  para os testes  $C$  (vermelho) e  $K$  (azul), seguindo o modelo de dependência espacial  $Sph(x, 80, 40)$  com valores para o efeito pepita (*eixo x*) iguais a 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71 e 76.

Pode-se observar na Figura 1 que ao nível de significância de 1%, o teste  $C$  obteve proporção de rejeição de  $H_0$  igual a 1 para efeito pepita iguais a 1, 6 e 11. À medida que se aumentou o efeito pepita, enfraquecendo a dependência

espacial, houve uma redução muito pequena nos valores da proporção de rejeição de  $H_0$ , chegando a 0,934, para efeito pepita igual a 76. Para o teste  $K$  obteve-se proporção de rejeição de  $H_0$  igual a 1, para efeito pepita igual a 1 e proporção igual a 0,849 para efeito pepita igual a 76.

Considerando um nível de significância de 5%, com a aplicação do teste  $C$  tem-se a proporção de rejeição de  $H_0$  igual a 1, para efeito pepita iguais a 1, 6 e 11 e proporção igual a 0,953 para efeito pepita igual a 76. Considerando o teste  $K$  obteve-se proporção igual a 1 para efeito pepita iguais a 1, 6 e 11 e valor igual a 0,918 para efeito pepita igual a 76.

Para um nível de significância igual a 10%, tem-se proporções iguais a 1, para efeito pepita iguais a 1, 6, 11 e 21 para ambos os testes. Para efeito pepita igual a 76, obteve-se proporções iguais a 0,964 e 0,945 para o teste  $C$  e  $K$ , respectivamente.

Os resultados do estudo do poder dos testes propostos (considerando o parâmetro alcance igual a 60) são apresentados na Figura 2.

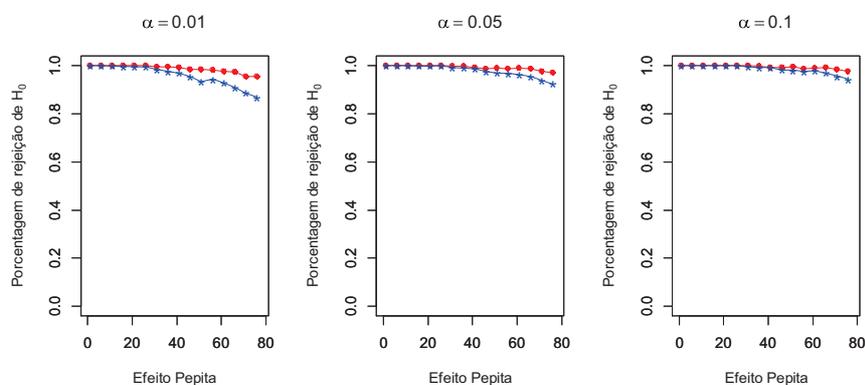


Figura 2 Porcentagem de rejeição de  $H_0$  para os testes  $C$  (vermelho) e  $K$  (azul), seguindo o modelo de dependência espacial  $Sph(x, 80, 60)$  com valores para o efeito pepita (*eixo x*) iguais a 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71 e 76.

Pode-se observar na Figura 2 que para o nível de significância de 1%, considerando o efeito pepita igual a 1, 6 e 21, obteve-se com a aplicação do teste

$C$  valor igual a 1 para a proporção de rejeição de  $H_0$  e 0,954 para efeito pepita igual a 76. Para o teste  $K$ , tem-se proporção igual a 1 para efeito pepita igual a 1 e 6 e 0,868 para efeito pepita igual a 76.

Ao nível de significância de 5%, com o uso do teste  $C$  tem-se proporção igual a 1 para efeito pepita iguais a 1, 6, 11, 16, 21 e 26, com efeito pepita 76 obteve-se proporção igual a 0,969. Para o teste  $K$ , obteve-se proporção igual a 1 para efeito pepita iguais a 1 e 6 e proporção igual a 0,922 para efeito pepita igual a 76.

Obteve-se, ao nível de significância de 10%, com a aplicação do teste  $C$  proporções iguais a 1, para efeito pepita iguais a 1, 6, 11, 16, 21, 26 e 31, e para efeito pepita igual a 76, tem-se proporção igual a 0,975. Para o teste  $K$ , tem-se proporção igual a 1, para efeito pepita igual a 1, 6, 11 e 26, e para efeito pepita igual a 76, obteve-se 0,942.

Para os níveis de significância 1%, 5% e 10% as proporções de rejeição de  $H_0$  para  $H_0$  falsa, atingidas pelo teste  $C$ , foram maiores ou iguais às obtidas pelo teste  $K$ , ou seja, o poder do teste  $C$  foi maior ou igual ao poder do teste  $K$  em todas as situações consideradas. Nos cenários considerados, os testes  $C$  e  $K$  obtiveram taxa de erro do tipo I próximas do nível de significância, ou seja, de um modo geral, pode-se dizer que os testes apresentaram um controle adequado do erro tipo I. Os resultados obtidos permitem afirmar que os testes propostos possuem bom desempenho e podem ser úteis para atingir os objetivos para os quais foram desenvolvidos. Sendo o teste  $C$  melhor do que o teste  $K$ , visto que o teste  $C$  apresentou poder igual, ou superior, ao teste  $K$ .

### **3.2 Envelopes simulados para situações típicas**

Para aplicação da metodologia proposta, utilizou-se 999 permutações no conjunto de dados e 25 distâncias. Os dados mostrados na Figura 3 foram

simulados sob a suposição de forte dependência espacial. Pode-se observar na Figura 3 os centroides das 100 áreas na região em estudo, com as respectivas contagens.

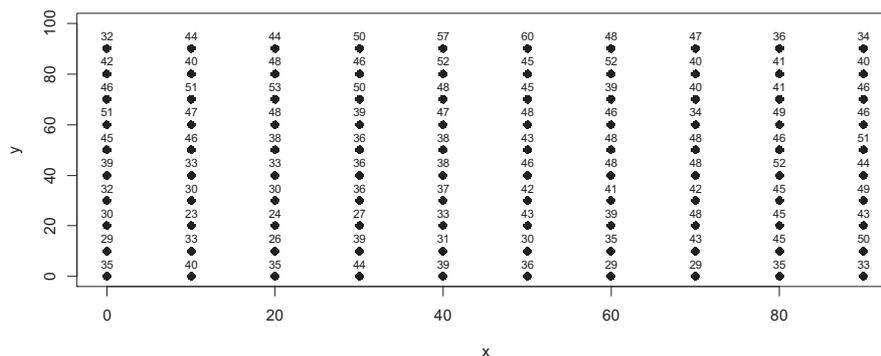


Figura 3 Configuração espacial de contagens sob a suposição de forte dependência espacial. Dados simulados seguindo o modelo de dependência espacial  $Sph(1, 80, 60)$ .

A Figura 4 apresenta o envelope simulado (azul) com a diferença entre os dois correlogramas (vermelho).

A análise da Figura 4 permite afirmar que o correlograma é estatisticamente significativo, pois a diferença observada entre os dois correlogramas extrapolam os limites superior e inferior para determinadas distâncias.

Com a aplicação dos testes, obteve-se  $C = 6,701$  ( $p = 0,001$ ) e  $K = 0,625$  ( $p = 0,001$ ), levando a rejeição de  $H_0$  ao nível de significância de 1%.

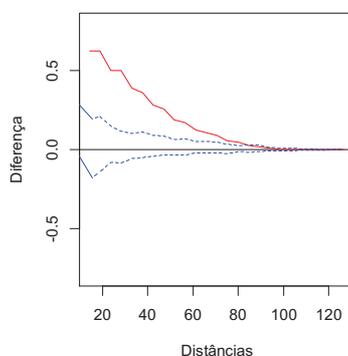


Figura 4 Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho).

Os dados apresentados na Figura 5 foram simulados considerando moderada dependência espacial. Os centroides e as respectivas contagens referentes a cada área são apresentados na Figura 5.

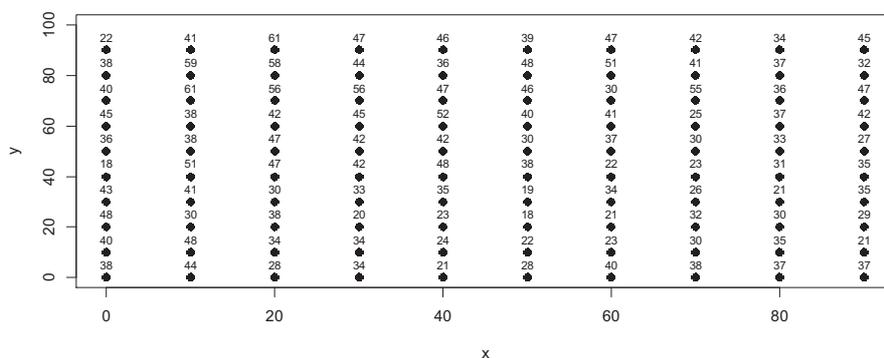


Figura 5 Configuração espacial de contagens sob a hipótese de moderada dependência espacial. Dados simulados seguindo o modelo de dependência espacial  $Sph(41,80,60)$ .

A partir do envelope simulado, apresentado na Figura 6, pode-se dizer que o correlograma é estatisticamente significativo, visto que a diferença observada entre os dois correlogramas não está totalmente contida no envelope simulado. Com a aplicação dos testes propostos, obteve-se  $C = 4,897$  ( $p =$

0,001) e  $K = 0,481$  ( $p = 0,001$ ), levando a rejeição de  $H_0$  ao nível de significância de 1%.

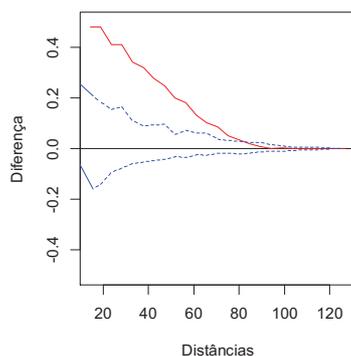


Figura 6 Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho).

A Figura 7 mostra os centroides e as contagens referentes a cada área, em que os dados foram simulados sob a suposição de fraca dependência espacial.

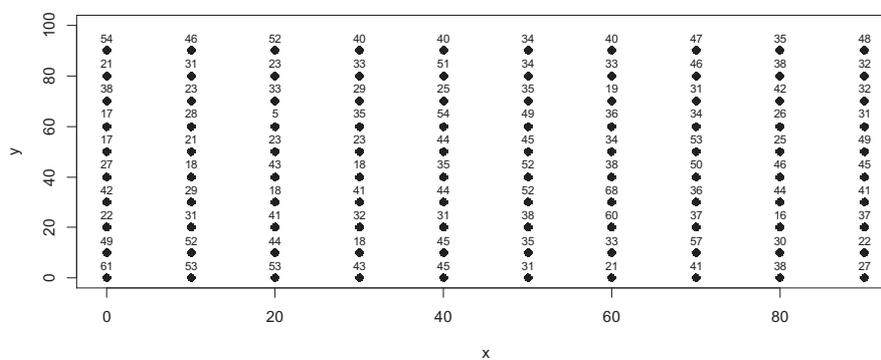


Figura 7 Configuração espacial de contagens sob a hipótese de fraca dependência espacial. Dados simulados seguindo o modelo de dependência espacial  $Sph(76, 80, 60)$ .

A representação gráfica mostrada na Figura 8 permite concluir que o correlograma é estatisticamente significativo, pois a diferença observada entre os dois correlogramas não está totalmente contida no envelope simulado. Com a

aplicação dos testes, tem-se  $C = 0,695$  ( $p = 0,001$ ) e  $K = 0,256$  ( $p = 0,001$ ), assim rejeita-se  $H_0$  ao nível de significância de 1%.

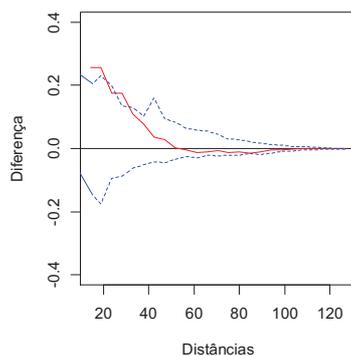


Figura 8 Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho).

A Figura 9 apresenta uma região com dados simulados considerando ausência de dependência espacial, em que os pontos representam os centroides das áreas e os números representam as contagens de cada área.

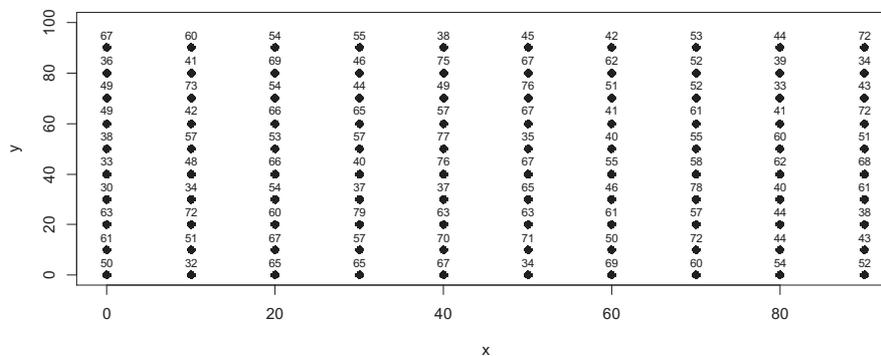


Figura 9 Configuração espacial de contagens sob a suposição de completa aleatoriedade espacial.

A Figura 10 mostra que a diferença observada entre os dois correlogramas está totalmente contida no envelope simulado confirmando que o correlogram não é estatisticamente significativo. Obteve-se  $C = 0,009$  ( $p =$

0,982) para o teste  $C$  e  $K = 0,035$  ( $p = 0,963$ ). Assim, não se rejeita a hipótese nula  $H_0$ , ao nível de significância de 1%.

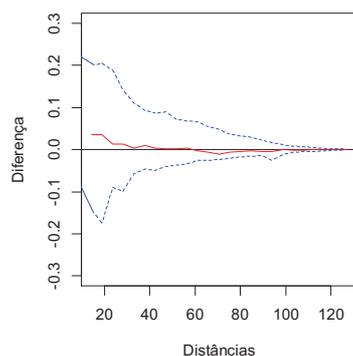


Figura 10 Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho).

Em todas as situações consideradas, os resultados obtidos com a utilização do envelope de simulação e a aplicação do teste  $C$  e do teste  $K$  levaram às mesmas conclusões.

As metodologias propostas rejeitaram  $H_0$ , ao nível de significância de 1%, em todos os exemplos que apresentavam dependência espacial (forte, moderada e fraca) e não rejeitou  $H_0$  no exemplo em que não havia dependência espacial.

### 3.3 Análise das minas predadas

Uma das pragas que mais preocupa os produtores de café no Brasil é o bicho-mineiro do cafeeiro (*Leucoptera coffeella*), pois uma infestação dessa praga, principalmente nos períodos secos, causa sérios prejuízos aos produtores (REIS; SOUZA; ZACARIAS, 2006).

Em se tratando de cultivos orgânicos, o controle biológico é extremamente importante. Os agentes mais eficientes no controle biológico são

os predadores, todos os predadores identificados são da ordem *Hymenoptera* e da família *Vespidae* (REIS; SOUZA, 1986). As vespas predadoras rasgam a epiderme das folhas no local das minas (lesões) e retiram as lagartas do bicho-mineiro, essas lagartas servem de alimento para as vespas (SOUZA; REIS; RIGITANO, 1998).

A Figura 11 mostra os centroides representando as 35 áreas na plantação de café. Cada número associado ao centroide indica as contagens referentes a cada área.

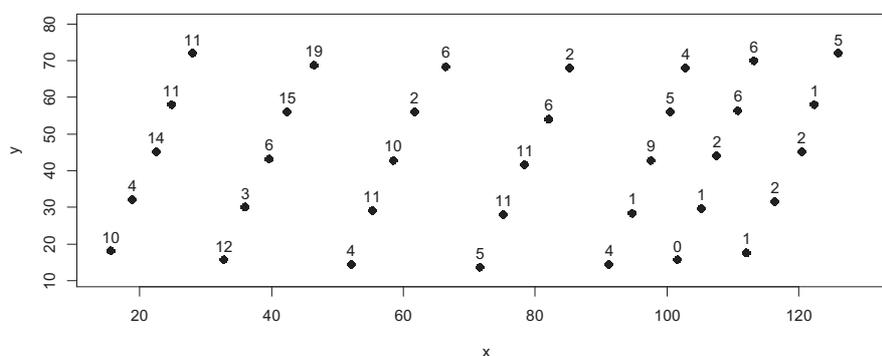


Figura 11 Contagens referentes ao número de minas predadas para as 35 áreas na plantação de café orgânico no mês de setembro de 2006.

O envelope simulado apresentado na Figura 12 mostra que para várias distâncias a diferença entre os dois correlogramas ficou fora do envelope simulado e, portanto, indicando que o correlograma é estatisticamente significativo.

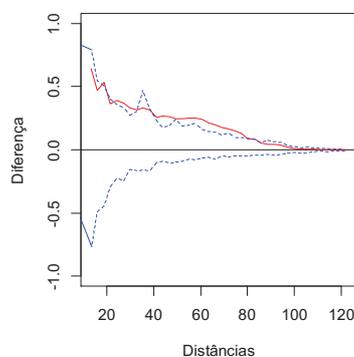


Figura 12 Envelope simulado (azul) com 999 permutações e a diferença entre os dois correlogramas (vermelho).

Com a aplicação dos testes de hipóteses tem-se  $C = 5,836$  ( $p = 0,001$ ) e  $K = 0,639$  ( $p = 0,026$ ), assim rejeita-se  $H_0$ , ao nível de significância de 5%, ou seja, a variável número de minas predadas exhibe autocorrelação espacial.

#### 4 CONCLUSÕES

A partir da análise de poder e erro do tipo I pode-se dizer que os testes propostos apresentaram bom desempenho nas situações avaliadas. Os estudos realizados com o envelope de simulação indicam que os mesmos são adequados para verificar a existência de autocorrelação espacial. Os métodos foram capazes de detectar a presença de dependência espacial no número de minas predadas em várias escalas de distâncias. Para uma melhor avaliação dos métodos propostos, sugere-se que em trabalhos futuros sejam realizados estudos sobre o poder, erro do tipo I e envelopes considerando outros cenários.

#### 5 REFERÊNCIAS

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific & Technical, 1995. 413 p.

CAMBARDELLA, C. A. et al. Field-scale variability of soil properties in Central Iowa soils. **Soil Science Society America Journal**, Madison, v. 58, n. 5, p. 1501-1511, Sept./Oct. 1994.

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

DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. **Model based geostatistics**. New York: Springer, 2007. 243 p.

MANLY, B. F. J. **Randomization and Monte Carlo methods in biology**. London: Chapman & Hall, 1994. 281 p.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, London, v. 37, p. 17-23, 1950.

ODEN, N. L. Assessing the significance of a Spatial Correlogram. **Geographical Analysis**, Columbus, v. 16, p. 1-16, 1984.

PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: analyses of phylogenetics and evolution in R language. **Bioinformatics**, Oxford, v. 20, n. 2, p. 289-290, Jan. 2004.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <<http://www.R-project.org/>>. Acesso em: 01 maio 2014.

REIS, P. R.; SOUZA, J. C. de. Pragas do cafeeiro. In: RENA, A. B. et al. (Ed.). **Cultura do cafeeiro: fatores que afetam a produtividade**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1986. p. 323-378.

REIS, P. R.; SOUZA, J. C. de; ZACARIAS, M. S. Alerta para o bicho-mineiro. **Cultivar**, Pelotas, v. 8, n. 84, p. 13-16, abr. 2006.

RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. GeoR: a package for geostatistical analysis. **R-News**, New York, v. 1, n. 2, p. 15-18, June 2001.

SOUZA, J. C.; REIS, P. R.; RIGITANO, R. L. O. **Bicho mineiro do cafeeiro: biologia, danos e manejo integrado**. 2. ed. Belo Horizonte: EPAMIG, 1998. 48 p. (Boletim Técnico, 54).

UPTON, G. J. G.; FINGLETON, B. **Spatial data analysis by example: point pattern and quantitative data**. New York: J. Wiley, 1985. v. 1, 422 p.

WALLER, L. A.; GOTWAY, C. A. **Applied spatial statistics for public health data**. New Jersey: J. Wiley, 2004. 519 p.

### ARTIGO 3: AVALIAÇÃO DA ESTRUTURA DE DEPENDÊNCIA ESPACIAL ENTRE DOIS PROCESSOS ESTOCÁSTICOS

Franciella Marques da Costa<sup>1</sup>      João Domingos Scalon<sup>2</sup>

**RESUMO:** O objetivo deste trabalho é propor dois testes de hipótese e um procedimento gráfico para comparar correlogramas e, conseqüentemente, verificar se a estrutura de dependência espacial de duas populações é a mesma. Os métodos foram testados em dados simulados e os resultados obtidos mostraram sua validade para comparar populações com diversas estruturas de dependência espacial. Os métodos também foram aplicados para testar a hipótese nula que o número de folhas minadas pelo bicho-mineiro do cafeeiro (*Leucoptera coffeella*) apresenta a mesma dependência espacial do número de vespas predadoras no mês de maior infestação da praga. Os métodos levaram a aceitação da hipótese nula.

**Palavras-chave:** Índice de Moran. Dependência espacial. Correlograma, Controle biológico. Bicho-mineiro do cafeeiro.

**ABSTRACT:** The aim of this work is to propose two hypothesis tests and a graphic procedure to compare correlograms and, consequently, verify that the structure of spatial dependence in two populations is the same. The methods were tested in simulated data and the results showed their validity for comparing

---

<sup>1</sup>Universidade Federal de Uberlândia/UFU - Faculdade de Ciências Integradas do Pontal/FACIP - Rua 20, 1600 – Bairro Tupã - 38.304-402 - Ituiutaba - MG - franciella@pontal.ufu.br

<sup>2</sup>Universidade Federal de Lavras/UFLA - Departamento de Ciências Exatas/DEX - Cx. P. 3037 - 37200-000 - Lavras - MG - scalon@dex.ufla.br

populations with many structures of spatial dependence. The methods were also applied to test the null hypothesis that the number of mined leaves by the coffee-leaf-miner (*Leucoptera coffeella*) has the same spatial dependence of the number of predatory wasps in the month of greatest infestation of the pest. The methods led to acceptance of the null hypothesis.

**Keywords:** Moran's index. Spatial dependence. Correlogram, Biological Control, Coffee leaf miner.

## 1 INTRODUÇÃO

A autocorrelação espacial descreve o grau de sincronismo que variáveis exibem dentro de um espaço geográfico e seu estudo é importante para avaliar vários fenômenos como, por exemplo, aqueles que ocorrem em ecologia (KOENIG; KNOPS, 1998; LEGENDRE, 1993).

A determinação da distância que uma espécie apresenta autocorrelação espacial em dados de área é, normalmente, realizada através do correlograma, ou seja, a representação gráfica da autocorrelação espacial da variável aleatória, considerando a ordem de vizinhança, conforme pode ser visto em diversos trabalhos como nos de Bailey e Gatrell (1995), Oden (1984) e Waller e Gotway (2004).

Apesar do grande uso da análise de autocorrelação espacial, o que se observa é o pouco desenvolvimento de métodos estatísticos para obter uma melhor caracterização da dependência espacial (THOMPSON et al., 1996). Apesar de existir algumas propostas de testes de significância, como pode ser visto no trabalho de Oden (1984), em geral, essa caracterização é feita quase que, exclusivamente, através de inspeção visual do correlograma.

Em outras áreas da estatística espacial, como configurações pontuais e geoestatística, a análise da estrutura da dependência espacial encontra-se muito mais desenvolvida. Nessas áreas, a significância estatística da dependência espacial é obtida através, não somente de testes estatísticos com distribuições amostrais conhecidas, mas também de métodos de computação intensiva como permutação e Monte Carlo (DIGGLE, 2003; DIGGLE; RIBEIRO JUNIOR, 2007).

A avaliação de que duas espécies, ocupando o mesmo espaço geográfico, apresentem a mesma estrutura de dependência espacial é uma análise de muita importância em diversas áreas, como por exemplo, no controle biológico de pragas, sendo importante caracterizar a estrutura de dependência espacial da praga e do seu predador. Syrjala (1996) apresenta um teste para determinar se duas espécies, ocupando o mesmo espaço geográfico, apresentam a mesma distribuição espacial. Deve-se observar que o método proposto por Syrjala (1996) não avalia e/ou compara estruturas de dependência (correlação) espacial existente entre as espécies.

Diante do exposto, este trabalho tem como objetivo propor dois testes de hipótese e um procedimento gráfico para verificar se a estrutura de dependência espacial de duas espécies é a mesma, desde que essas espécies convivam no mesmo espaço geográfico. Os métodos propostos são baseados no correlograma do índice de Moran, computação intensiva e nas medidas de distâncias de Cramer-von Mises e Kolmogorov.

Os métodos serão aplicados em dados típicos simulados para mostrar o seus funcionamentos em diversas situações. Os métodos também serão utilizados para verificar se a estrutura de dependência espacial do número de folhas minadas pelo bicho-mineiro é a mesma das vespas predadoras no mês de maior infestação da praga. Esse tipo de procedimento ainda não foi empregado na análise da estrutura de dependência espacial em dados de área nem utilizado

como metodologia para avaliar correlação espacial da ocorrência de presa e predador em plantações orgânicas de café.

## 2 MATERIAL E MÉTODOS

### 2.1 Índice de Moran e correlograma

A metodologia proposta neste trabalho utiliza o índice global de Moran (MORAN, 1950) que é uma medida de autocorrelação espacial amplamente utilizada para dados de áreas. Esse índice fornece um único valor que representa uma medida global da associação espacial dos dados na região de interesse (WALLER; GOTWAY, 2004).

O índice global de Moran, considerando apenas o primeiro vizinho (primeira ordem), é dado pela expressão

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i=1}^N (Y_i - \bar{Y})^2) (\sum_{i=1}^N \sum_{j=1}^N w_{ij})}, \quad (1)$$

em que  $N$  é o número de áreas,  $Y_i$  é o valor do atributo na área  $i$ ,  $Y_j$  é o valor do atributo na área  $j$ ,  $\bar{Y}$  é a média do atributo e  $w_{ij}$  são os elementos da matriz de proximidade espacial.

Existem várias maneiras de construir a matriz de proximidade espacial conforme pode ser visto em Bailey e Gatrell (1995) e Waller e Gotway (2004). Neste trabalho, adota-se  $w_{ij} = 1$  se  $d_{ij} < d$  em que  $d$  é uma distância definida pelo pesquisador e  $d_{ij}$  é a distância euclidiana entre os centroides das regiões  $i$  e  $j$  e  $w_{ij} = 0$ , caso contrário.

O índice global de Moran pode assumir qualquer valor no domínio dos reais, entretanto, na maioria dos casos fica restrito entre -1 e 1 (UPTON;

FINGLETON, 1985). Valores positivos indicam que áreas vizinhas apresentam valores semelhantes, enquanto índices negativos indicam que áreas vizinhas possuem valores diferentes. Valores próximos de zero indicam ausência de dependência espacial.

Para determinar a significância estatística do índice de Moran, é preciso associá-lo a uma distribuição estatística. Upton e Fingleton (1985) mostram que, sob a suposição de normalidade dos dados, o índice de Moran apresenta distribuição Normal. Assim, pode-se testar a hipótese nula de independência espacial, relacionando uma estatística de teste à distribuição normal. Outra possibilidade, sem pressupostos em relação à distribuição dos dados, é utilizar um teste de pseudos-significância baseado em aleatorização (BAILEY; GATRELL, 1995; UPTON; FINGLETON, 1985; WALLER; GOTWAY, 2004).

O índice global de Moran pode ser calculado considerando os vizinhos de maior ordem (vizinhos dos vizinhos). A representação gráfica da autocorrelação espacial medida pelo índice de Moran para a mesma variável considerando cada ordem de vizinhança é chamada de correlograma (ODEN, 1984).

Como o objetivo deste trabalho é analisar o comportamento da autocorrelação espacial, à medida que se aumenta a distância, propõem-se uma alteração na obtenção do correlograma baseado no índice de Moran. Assim, ao invés de calcular o índice para cada ordem de vizinhança, o mesmo será sempre calculado para os vizinhos de primeira ordem para diferentes distâncias. Por exemplo, considerando as distâncias ( $d$ ) de  $0 - d_1, 0 - d_2, \dots, 0 - d_{max}$ . Neste trabalho, a representação gráfica para esta situação também será denominada de correlograma.

## 2.2 Testes de hipótese para igualdade entre duas estruturas de autocorrelação espacial

Os testes propostos são baseados na comparação dos correlogramas obtidos para cada população em estudo e têm por objetivo verificar a igualdade da estrutura de dependência espacial de duas populações.

A hipótese nula ( $H_0$ ) a ser testada é que a estrutura de dependência espacial é igual nas duas populações, enquanto que a hipótese alternativa ( $H_1$ ) é que existe alguma diferença na estrutura de dependência das duas populações.

Sejam  $P = (p_x, p_y)$  e  $Q = (q_x, q_y)$  as coordenadas dos centroides das áreas em estudo. Devem-se calcular as distâncias euclidianas entre os centroides das regiões  $i$  e  $j$ , obtendo assim a matriz simétrica de distâncias euclidianas dada por

$$A = \begin{bmatrix} 0 & d_{12} & d_{13} & \dots & d_{1N} \\ d_{21} & 0 & d_{23} & \dots & d_{2N} \\ d_{31} & d_{32} & 0 & \dots & d_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \dots & 0 \end{bmatrix},$$

em que,  $d_{12}$  é a distância euclidiana entre as áreas 1 e 2 e  $N$  é o número de áreas.

Considerando  $a$  = distância mínima entre os centroides e  $b$  = distância máxima entre os centroides, calcula-se o índice global de Moran  $I(d)$  para as diferentes distâncias em que  $a < d < b$ . As distâncias ( $d$ ) devem ser definidas pelo pesquisador.

Para o cálculo do índice de Moran, será utilizada a matriz de proximidade espacial em que  $w_{ij} = 1$  se  $d_{ij} < d$ , sendo que  $d$  é uma distância

definida pelo pesquisador e  $d_{ij}$  é a distância euclidiana entre os centroides das regiões  $i$  e  $j$ ,  $w_{ij} = 0$  caso contrário.

Definida a matriz de proximidade espacial, o índice global de Moran é calculado para a população 1 ( $I^{(1)}(d)$ ) e para a população 2 ( $I^{(2)}(d)$ ). Construindo um gráfico dos índices globais de Moran ( $I(d)$ ), em função das distâncias ( $d$ ), obtêm-se o correlograma.

Para verificar se as estruturas de dependência espacial entre as duas populações são iguais, deve-se escolher estatísticas de teste que consigam medir a diferença existente entre os correlogramas das duas populações. Neste caso, pode-se utilizar duas estatísticas de teste.

A primeira estatística de teste, baseada na distância de Cramer-von Mises, é dada por:

$$\Omega = \int \left( I^{(1)}(d) - I^{(2)}(d) \right)^2 d(d). \quad (2)$$

Como não é possível calcular essa integral analiticamente, pode-se obtê-la, aproximadamente, usando algum método numérico. Neste trabalho, utiliza-se a Regra de Simpson.

A segunda estatística de teste é baseada na distância de Kolmogorov dada por

$$\Phi = \max |I^{(1)}(d) - I^{(2)}(d)|. \quad (3)$$

As duas estatísticas de teste não apresentam distribuições amostrais conhecidas, o que impede o uso de metodologias usuais da estatística clássica para conduzir os testes de hipótese. Assim, para conduzir os testes de hipóteses, utilizando as duas estatísticas apresentadas anteriormente, recorrem-se ao

método de computação intensiva de aleatorização. Nesse sentido, sob a hipótese nula, a distribuição amostral da estatística de teste é construída a partir do cálculo do valor da estatística de teste para cada permutação do conjunto de dados. O valor- $p$  é obtido pela proporção dos valores das estatísticas de teste, obtidas na aleatorização que são maiores ou iguais à estatística de teste observada.

### 2.3 Envelopes simulados

Além dos testes de hipóteses apresentados na seção anterior, pode-se utilizar um método gráfico, baseado em envelopes de simulação, para testar a hipótese nula de que a estrutura de dependência espacial seja a mesma para as duas populações.

Para testar a hipótese nula de que a estrutura de dependência espacial das duas populações seja igual, adapta-se o procedimento proposto por Diggle (2003) para a construção de envelopes de simulação para testar a hipótese de completa aleatoriedade espacial em configurações pontuais. Assim, o procedimento para a construção de envelopes simulados é dado pelos seguintes passos:

- Obtém-se os correlogramas  $I_o^{(1)}(d)$  e  $I_o^{(2)}(d)$  para as duas populações em estudo, o índice "o" significa que os cálculos foram efetuados a partir dos dados observados, os expoentes (1) e (2) indicam a população considerada e  $d$  são as distâncias consideradas;
- Calcula-se a diferença  $I_o(d) = I_o^{(1)}(d) - I_o^{(2)}(d)$  para os dados observados;
- Permuta-se os dados  $m$  vezes, em seguida calcula-se  $I_i(d) = I_i^{(1)}(d) - I_i^{(2)}(d)$  para  $i = 1, 2, 3, \dots, m$ ;
- Obtém-se o limite superior e inferior do envelope simulado como sendo:

$$I_S(d) = \max \{I_i(d)\}, i = 1, 2, 3, \dots, m$$

$$I_I(d) = \min \{I_i(d)\}, i = 1, 2, 3, \dots, m$$

em que  $I_S(d)$  e  $I_I(d)$  são, respectivamente, o limite superior e o limite inferior do envelope simulado.

- Constrói-se um gráfico com os valores  $I_o(d)$ ,  $I_S(d)$  e  $I_I(d)$  contra as distâncias  $d$ .

Diggle (2003) mostra que esse procedimento apresenta a propriedade  $P\{I_o(d) > I_S(d)\} = P\{I_o(d) < I_S(d)\} = \frac{1}{m+1}$ .

A interpretação do método gráfico é muito simples. Se a linha de  $I_o(d)$  estiver entre o limite superior e o limite inferior do envelope simulado, existe evidências de que a estrutura de dependência espacial das duas populações é a mesma. Se a linha  $I_o(d)$  não estiver totalmente dentro dos limites do envelope simulado então a estrutura de dependência espacial das duas populações é diferente.

#### 2.4 Estudo de casos com dados simulados

Com o objetivo de verificar o comportamento dos métodos propostos, foi realizado um estudo, por meio de simulação, de alguns casos típicos de estrutura de dependência espacial (fraca, moderada e forte) em uma grade regular de 100 pontos, representando os centroides de cada uma das 100 áreas da região com área de 10 x 10 unidades de área. Para tal, foi utilizada a função *grf* da biblioteca *geoR* (RIBEIRO JUNIOR; DIGGLE, 2001), do *software* R (R CORE TEAM, 2014).

A população 1 foi composta de uma realização de 100 observações nas coordenadas dos cruzamentos da grade regular (representando os centroides de cada área) utilizando um processo Gaussiano estacionário. Para a estrutura de

dependência espacial, utilizou-se a função de correlação esférica com patamar igual a 60 e diferentes valores de efeito pepita e alcance.

Para a classificação quanto ao grau de dependência espacial foi utilizado o índice proposto por Cambardella et al. (1994), que é dado pela expressão

$$IDE = \left( \frac{C_0}{C_0 + C_1} \right) \times 100, \text{ em que } C_0 \text{ é o efeito pepita e } C_0 + C_1 \text{ é o patamar.}$$

O índice de Cambardella assume valores entre 0 e 100%. Se o valor do índice for menor que 25% tem-se dependência espacial forte, índice entre 25% e 75% caracteriza dependência espacial moderada e valores maiores que 75% significa dependência espacial fraca (CAMBARDELLA et al., 1994).

O plano do trabalho de simulação da população 1 é apresentado na Tabela 1.

Tabela 1 Configurações utilizadas para simulação.

<b>Nomenclatura</b>	<b>Modelo</b>	<b>Efeito pepita</b>	<b>Patamar</b>	<b>Alcance</b>	<b>Grau de dependência espacial</b>
Sph(3,60,40)	Esférico	3	60	40	Forte
Sph(3,60,60)	Esférico	3	60	60	Forte
Sph(30,60,40)	Esférico	30	60	40	Moderada
Sph(30,60,60)	Esférico	30	60	60	Moderada
Sph(55,60,40)	Esférico	55	60	40	Fraca
Sph(55,60,60)	Esférico	55	60	60	Fraca

Para obter os dados da segunda população, utilizou-se os dados da primeira população e adicionou-se um termo aleatório que segue uma distribuição normal com média zero e variância como um parâmetro que controla a distância entre os dois correlogramas. Neste trabalho, a variância da distribuição normal gerada é denominada de variabilidade radial. Deve-se observar que à medida que se aumenta a variabilidade radial, sendo essa acrescentada em um conjunto de dados com dependência espacial, o resultado é um conjunto de dados que tende à completa independência (aleatoriedade) espacial e, conseqüentemente, o seu correlograma tende a se afastar do correlograma da População 1.

## **2.5 Estudo de caso com dados reais**

Os dados utilizados neste trabalho são provenientes do mês de setembro (mês de maior infestação do bicho-mineiro), do ano de 2006, de um hectare de produção de café orgânico da Cultivar Catucaí, localizado no município de Santo Antônio do Amparo – MG. A coleta de dados foi realizada em 35 pontos de amostragem (representando centroides de áreas). Em cada ponto de amostragem foi determinado o número de folhas minadas em 10 folhas do cafeeiro e o número de vespas predadoras em armadilha adesiva. A contagem das folhas minadas e das vespas foi realizada no laboratório da Epamig-CTSM/EcoCentro, em Lavras – MG.

## **2.6 Software**

Todas as análises foram realizadas utilizando funções construídas utilizando o *software* R (R CORE TEAM, 2014), juntamente com as bibliotecas *geoR* (DIGGLE; RIBEIRO JUNIOR, 2007; RIBEIRO JUNIOR; DIGGLE, 2001) e *ape* (PARADIS; CLAUDE; STRIMMER, 2004).

### 3 RESULTADOS E DISCUSSÃO

#### 3.1.1 Simulação com forte dependência espacial

A Figura 1 mostra um exemplo do correlograma da população 1 obtidos com a configuração  $Sph(3, 60, 60)$  e os correlogramas da população 2 considerando as variabilidades radiais iguais a 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9 e 10. Pode-se observar na Figura 1 que à medida que se aumenta a variabilidade radial aumenta-se a diferença entre os correlogramas das duas populações.

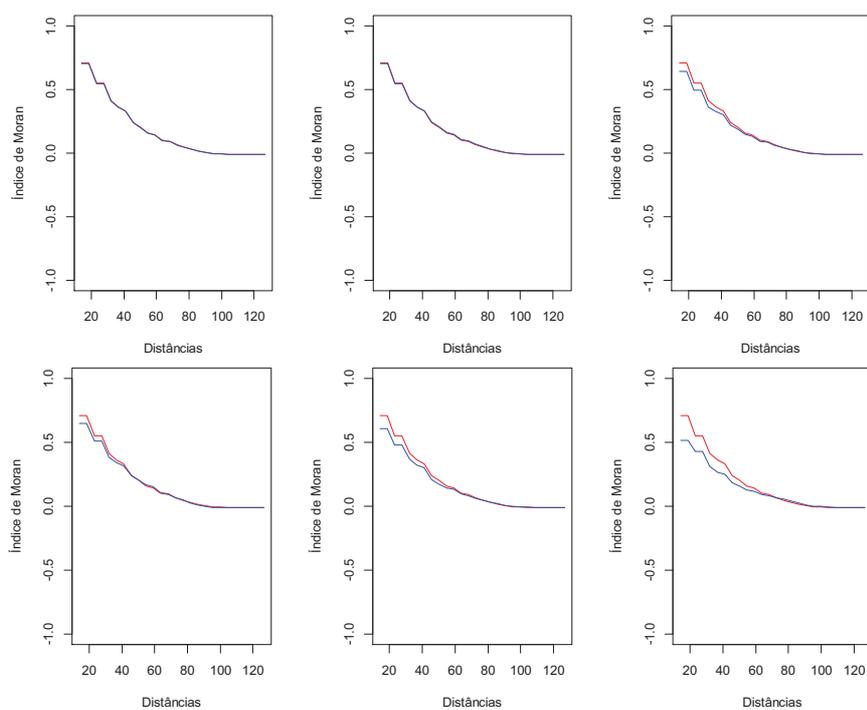


Figura 1 Correlogramas da população 1 (vermelho) e população 2 (azul) obtidos com a configuração  $Sph(3,60,60)$  considerando as variabilidades radiais iguais a 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. (continua)

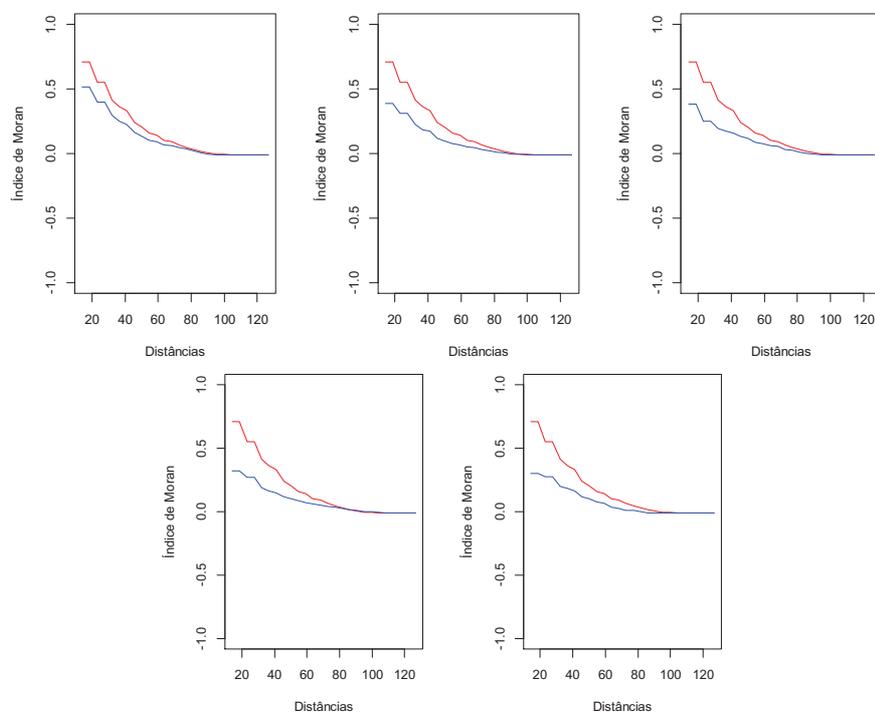


Figura 1 (continuação) Correlogramas da população 1 (vermelho) e população 2 (azul) obtidos com a configuração  $Sph(3,60,60)$  considerando as variabilidades radiais iguais a 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Para cada variabilidade radial considerada, foi gerada 100 amostras e aplicado os testes  $C$  e  $K$  para cada amostra. As Figuras 2 e 3 mostram respectivamente, a porcentagem de rejeição de  $H_0$  para as configurações  $Sph(3,60,40)$  e  $Sph(3,60,60)$ , ao nível de significância  $\alpha = 1\%$ ,  $\alpha = 5\%$  e  $\alpha = 10\%$ .

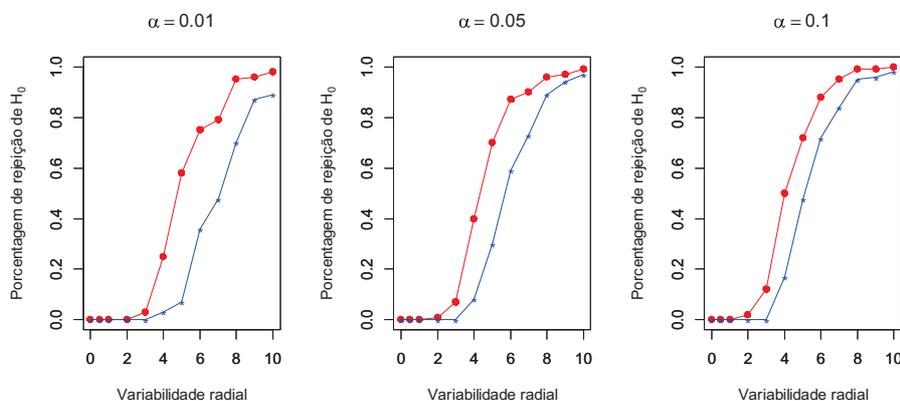


Figura 2 Porcentagem de rejeição de  $H_0$  com as configurações seguindo o modelo de dependência espacial  $Sph(3,60,40)$  para os testes C (vermelho) e K (azul).

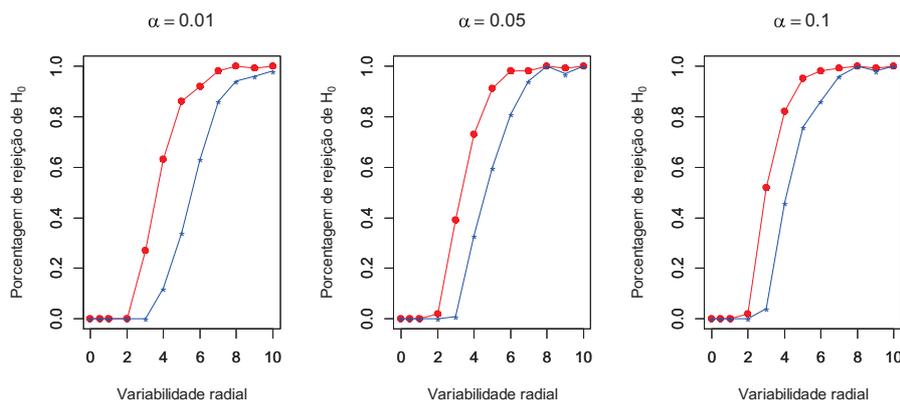


Figura 3 Porcentagem de rejeição de  $H_0$  com as configurações seguindo o modelo de dependência espacial  $Sph(3,60,60)$  para os testes C (vermelho) e K (azul).

De modo geral, a porcentagem de rejeição de  $H_0$  aumentou à medida que se aumentou a variabilidade radial, ou seja à medida que os correlogramas vão se distanciando. A porcentagem de rejeição de  $H_0$  do teste  $C$  foi maior ou igual a do teste  $K$  para os dois cenários analisados e considerando o nível de significância  $\alpha = 1\%$ ,  $\alpha = 5\%$  e  $\alpha = 10\%$ . Esse comportamento das estatísticas de teste já foi observado por Diggle (2003) quando utilizadas no contexto de configurações pontuais.

Com exceção do teste  $K$  para  $Sph(30,60,40)$ , ao nível de significância de 1%, em todas as outras situações para uma variabilidade radial igual a 10 a proporção de rejeição de  $H_0$  com a configuração  $Sph(3,60,40)$  foi próxima de 1 ou exatamente 1.

Observa-se que as proporções de rejeição de  $H_0$ , considerando o parâmetro alcance igual a 60, foram maiores ou iguais do que as proporções em relação ao alcance de 40, o que indica que o alcance da dependência espacial pode influenciar os resultados do teste.

### 3.1.2 Simulação com moderada dependência espacial

As porcentagens de rejeição de  $H_0$  com a configuração  $Sph(30,60,40)$  e  $Sph(30,60,60)$  são apresentadas na Figura 4 e na Figura 5, respectivamente.

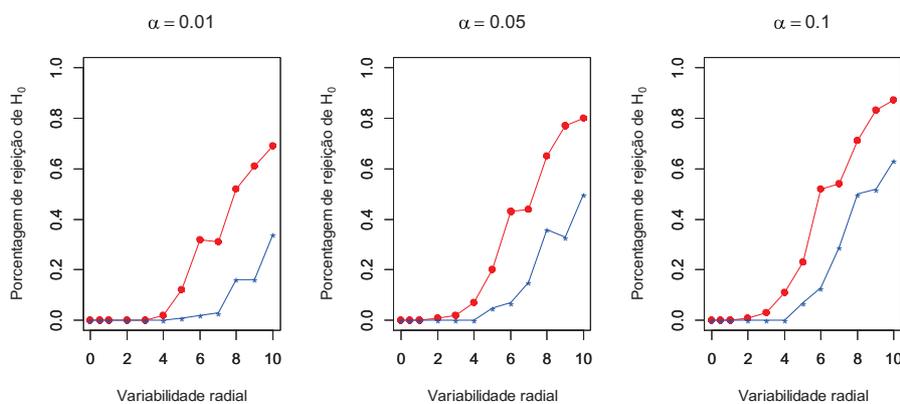


Figura 4 Porcentagem de rejeição de  $H_0$  com as configurações seguindo o modelo de dependência espacial  $Sph(30,60,40)$  para os testes C (vermelho) e K (azul).

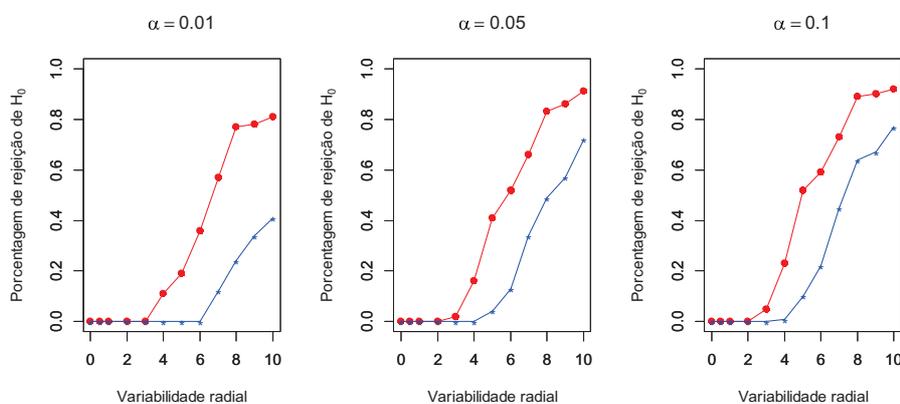


Figura 5 Porcentagem de rejeição de  $H_0$  com as configurações seguindo o modelo de dependência espacial  $Sph(30,60,60)$  para os testes C (vermelho) e K (azul).

Observa-se que as proporções de rejeição de  $H_0$  foram menores em relação às configurações com a presença de forte dependência espacial, discutidas anteriormente, com exceção de situações que ambos foram iguais a zero. As proporções de rejeição de  $H_0$  para o teste C foram maiores quando comparadas às obtidas com a aplicação do teste K em ambas as situações analisadas. Nenhum dos testes, em caso de moderada dependência espacial, atingiu a proporção 1 até a variabilidade radial 10. Os resultados mostraram que os testes são mais “poderosos” sob a suposição de forte dependência espacial do que em cenários de moderada dependência espacial.

### 3.1.3. Simulação com fraca dependência espacial

As Figuras 6 e 7 mostram a porcentagem de rejeição de  $H_0$  com a configuração  $Sph(55,60,40)$  e  $Sph(55,60,60)$ , respectivamente.

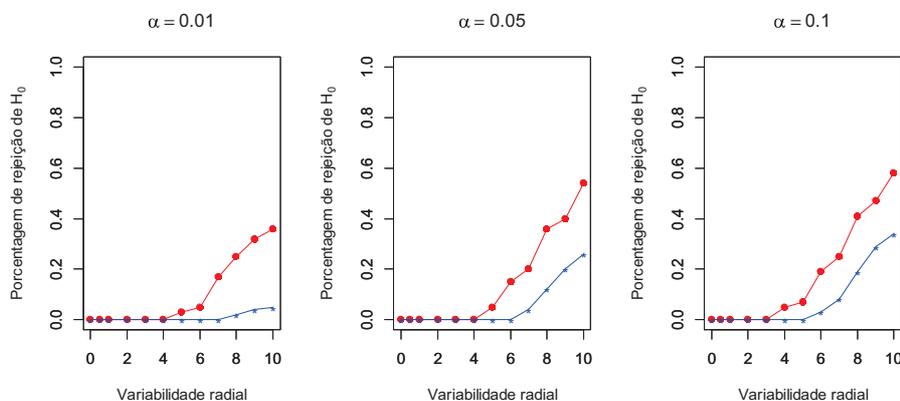


Figura 6 Porcentagem de rejeição de  $H_0$  com com as configurações seguindo o modelo de dependência espacial  $Sph(55,60,40)$  para os testes C (vermelho) e K (azul).

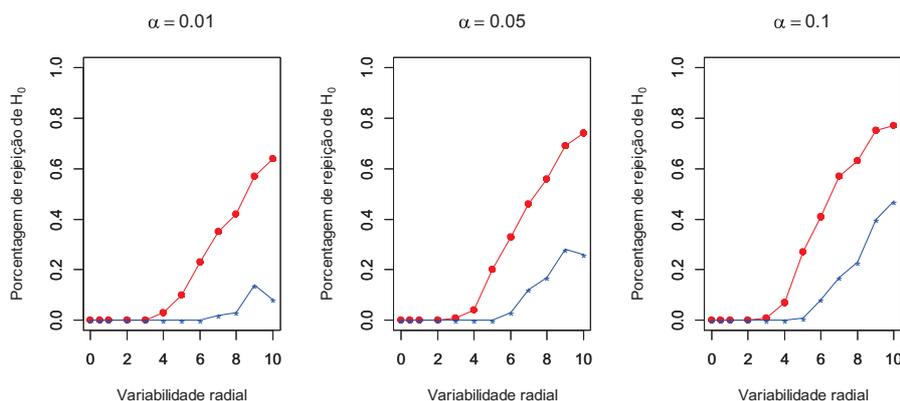


Figura 7 Porcentagem de rejeição de  $H_0$  com com as configurações seguindo o modelo de dependência espacial  $Sph(55,60,60)$  para os testes C (vermelho) e K (azul).

Em relação aos cenários analisados anteriormente, as proporções de rejeição de  $H_0$  são menores. Com a situação de fraca dependência os testes não atingiram 100% de rejeição de  $H_0$ . Com os resultados apresentados, pode-se concluir que o desempenho dos testes com fraca dependência é inferior aos casos de forte e moderada dependência espacial.

As estatísticas de teste, quando a variabilidade radial for igual a zero (as curvas estiverem sobrepostas), será igual a zero para ambos os testes propostos. O valor-p é obtido pela proporção dos valores das estatísticas de teste que são maiores ou iguais à estatística de teste observada. As estatísticas do teste  $C$  e  $K$  serão sempre maiores ou iguais a zero. Assim sempre que os correlogramas estiverem sobrepostos, têm-se  $\text{valor} - p = 1$  para qualquer situação.

Vale ressaltar que não se podem tirar conclusões a cerca de erro do tipo I ou poder do teste baseado nos exemplos apresentados, visto que foram gerados dados de uma população e, com base na amostra desta população, obteve-se a segunda população, apenas adicionado um termo aleatório a cada observação.

Observou-se que sobre moderada e fraca dependência espacial os testes não atingiram 100% de rejeição de  $H_0$ , porém deve-se ter cuidado ao interpretar esses resultados. Por exemplo, considere a população 1 gerada sobre fraca dependência espacial e a população 2 obtida utilizando uma variabilidade radial igual a 10. Neste caso, a distância entre os dois correlogramas será pequena. Se a população 1 for gerada sobre forte dependência espacial e a população 2 obtida utilizando uma variabilidade espacial igual a 10, a distância entre os dois correlogramas será bem maior do que no caso de fraca dependência espacial. Assim, as distâncias entre os dois correlogramas não será a mesma, mesmo considerando a mesma variabilidade radial.

### **3.1.4 Envelopes simulados para casos típicos**

#### **3.1.4.1 Forte dependência espacial**

A Figura 8 apresenta os dois correlogramas (a) e o envelope simulado (b) para a diferença dos dois correlogramas em uma situação de forte

dependência espacial e, portanto, ocorrendo área muito pequena entre os dois correlogramas.

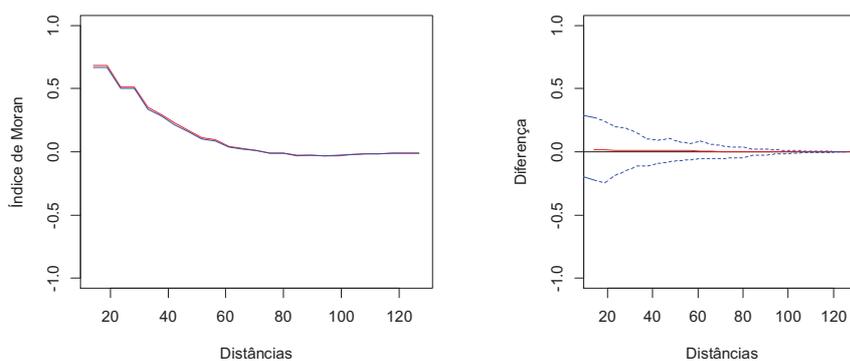


Figura 8 Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial  $Sph(3,60,60)$  e uma variabilidade radial de 1.

Para o teste  $K$ , obteve-se  $\Phi = 0,019$  ( $p = 1$ ) e para o teste  $C$  tem-se  $\Omega = 0,007$  ( $p = 1$ ). Em ambos os testes, não rejeita-se  $H_0$ , ao nível de significância de 1%, concluindo que a estrutura de dependência espacial é a mesma para as duas espécies. Como a diferença observada entre os dois correlogramas está totalmente contida no envelope simulado conclui-se que a estrutura de dependência espacial das duas populações é a mesma. Os três métodos propostos levaram à mesma conclusão.

Observa-se na Figura 9 os dois correlogramas (a) e o envelope simulado para a diferença entre os dois correlogramas (b), sob forte dependência espacial, porém com uma área entre os dois correlogramas um pouco maior do que aquela observada na Figura 8.

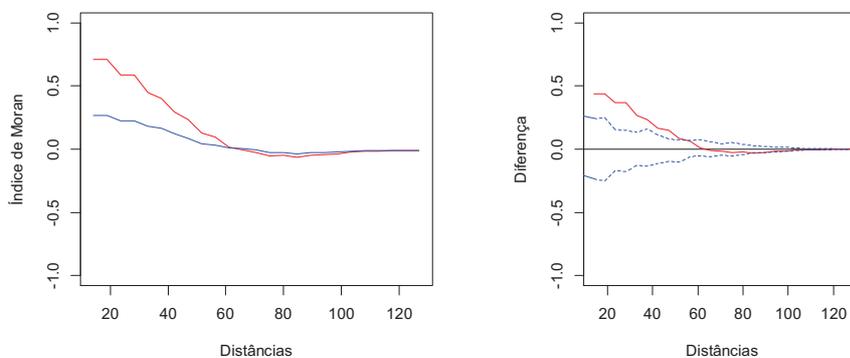


Figura 9 Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial  $Sph(3,60,60)$  e uma variabilidade radial de 10.

Para o teste  $K$ , tem-se  $\Phi = 0,440$  ( $p = 0,001$ ) e para o teste  $C$  tem-se  $\Omega = 3,080$  ( $p = 0,001$ ), assim rejeita-se  $H_0$  ao nível de significância de 1%, logo a estrutura de dependência espacial é diferente para as duas populações. O resultado obtido por meio do envelope simulado corrobora os resultados obtidos com a aplicação dos testes  $C$  e  $K$ .

#### 3.1.4.2 Fraca dependência espacial

Na Figura 10, têm-se os dois correlogramas (a) e o envelope simulado para a diferença entre os dois correlogramas (b), para uma situação de fraca dependência espacial, em que a área entre os dois correlogramas é muito pequena.

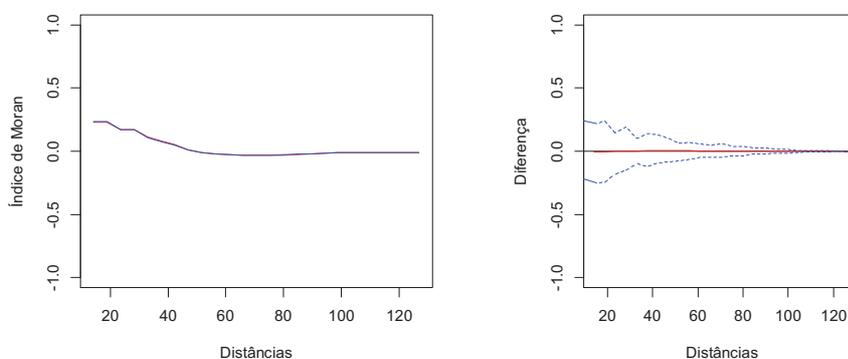


Figura 10 Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial  $Sph(55,60,60)$  e uma variabilidade radial de 1.

Aplicando o teste  $C$  e o teste  $K$  obteve-se  $\Phi = 0,005$  ( $p = 1$ ) e para o teste  $C$  tem-se  $\Omega = 0,0005$  ( $p = 1$ ), assim não se rejeita  $H_0$  ao nível de significância de 1%, ou seja, a estrutura de dependência espacial é igual. O resultado obtido por meio do envelope simulado corrobora os resultados obtidos com a aplicação dos testes  $C$  e  $K$ .

Na Figura 11, têm-se os dois correlogramas (a) e o envelope simulado para a diferença entre os mesmos (b), para uma situação de fraca dependência espacial e variabilidade radial igual a 10.

Obteve-se  $\Phi = 0,073$  ( $p = 0,667$ ) e  $\Omega = 0,063$  ( $p = 0,502$ ). Assim com o teste  $K$  não se rejeita  $H_0$  ao nível de significância de 1%, e com o teste  $C$  não rejeita-se  $H_0$ . O resultado obtido por meio do envelope simulado corrobora o resultado obtido com a aplicação dos testes  $C$  e  $K$ .

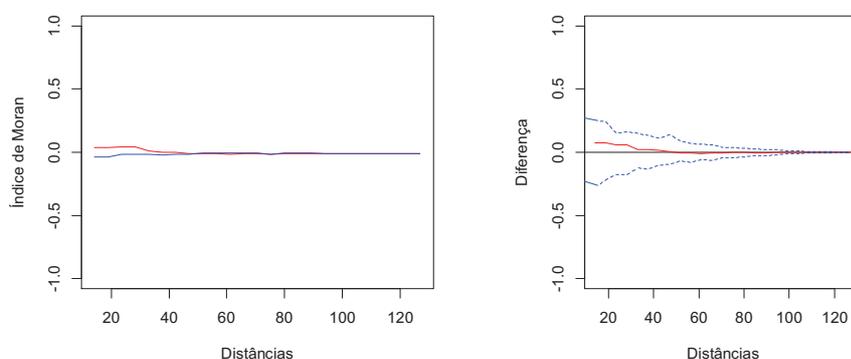


Figura 11 Correlograma (a) e envelope simulado com 999 permutações (b) ambos obtidos a partir de dados simulados com as configurações seguindo o modelo de dependência espacial  $Sph(55,60,60)$  e uma variabilidade radial de 10.

### 3.2. Análise das folhas minadas e vespas predadoras

O Brasil é o maior produtor e exportador de café do mundo, sendo a cafeicultura de extrema importância para a economia brasileira. De acordo com o levantamento divulgado pela Companhia Nacional de Abastecimento (CONAB, 2014), a previsão de safra de café (arábica e robusta) no Brasil para o ano de 2014 é de 44,57 milhões de sacas de 60 kg de café beneficiado, observando uma redução de 9,33%, o que representa 4,58 milhões de sacas, em relação a produção do ciclo anterior.

A infestação de pragas é um dos problemas que mais prejudica a produção de café, causando enormes prejuízos aos cafeicultores e às exportações brasileiras. O bicho-mineiro do cafeeiro (*Leucoptera coffeella*) é uma das pragas que mais preocupa os produtores de café no Brasil (REIS; SOUZA, 1986; REIS; SOUZA; ZACARIAS, 2006).

Apesar das chuvas e da umidade relativa do ar exercer um controle natural da população da praga (REIS; SOUZA; ZACARIAS, 2006), é usual

controlar o bicho-mineiro por meio do controle químico ou biológico, sendo o controle biológico o único permitido em cultivos orgânicos. De acordo com Souza, Reis e Rigitano (1998), o controle biológico pode ser conduzido por meio de: (a) Parasitoides (*microhimenópteros*) que parasitam e matam as lagartas do bicho-mineiro; (b) Vespas predadoras de várias espécies que rasgam a epiderme da folha e comem as lagartas do bicho-mineiro e (c) Patógenos (bactérias e fungos) que causam doenças nas lagartas.

Segundo Reis e Souza (1986), o controle de pragas utilizando predadores é realizado somente em condições naturais, mas em trabalhos experimentais, foi possível mudar ninhos de vespas para locais próximos a plantações de café. Sendo assim, sabendo que as folhas minadas e as vespas predadoras possuem a mesma estrutura de dependência espacial pode-se pensar em transferir ninhos de vespas para os locais onde há focos de infestação do bicho-mineiro, assim possibilitando um controle mais eficiente desta praga. Deve-se observar a quantidade de vespas suficientes para o controle do bicho-mineiro.

O conhecimento da estrutura de dependência espacial do bicho-mineiro e de suas vespas predadoras pode fornecer informações importantes para o produtor que visa uma produção de café orgânica ou uma redução no uso de inseticidas. Nesse sentido, Scalon et al. (2011b) utilizaram métodos da geoestatística (semivariograma) para mostrar que o número de folhas minadas, para o mês de maior infestação do bicho-mineiro, apresenta efeito pepita puro, o que caracteriza uma ausência total de dependência espacial dessa variável no cafezal. Entretanto, Scalon et al. (2011a, 2011b) argumentam que é necessário novos estudos, envolvendo novas metodologias estatísticas, para uma melhor caracterização da dependência espacial do bicho-mineiro e da vespa predadora. Assim, este trabalho estuda, não somente o comportamento da dependência espacial dessas espécies em função da distância, mas também avalia se essas

espécies possuem a mesma estrutura de dependência espacial.

A Figura 12 apresenta os centroides representado as 35 áreas na região de estudo (cafezal). Cada número associado ao centroide indica as contagens referentes a cada área.

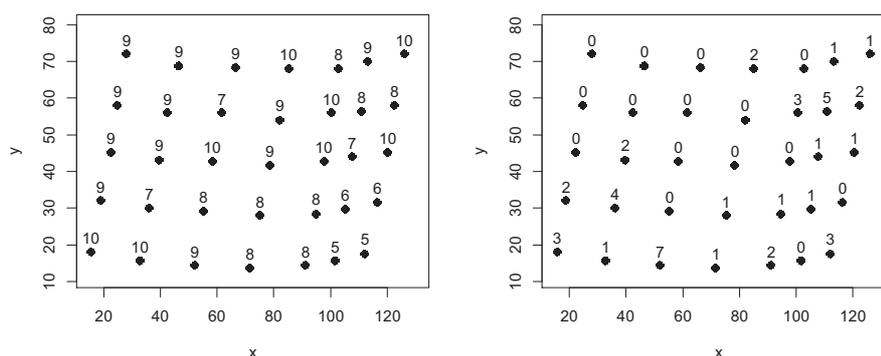


Figura 12 Contagens referentes ao número de folhas minadas (a) e vespas predadoras (b) para as 35 áreas na plantação de café orgânico no mês de setembro de 2006.

A Figura 13(a) apresenta os correlogramas do número de folhas minadas e vespas predadoras, em que é possível observar que existe uma diferença entre as duas espécies no que se refere à estrutura de dependência espacial. Entretanto, é necessário verificar se essas diferenças observadas visualmente representam uma diferença em estruturas de dependência espacial que seja estatisticamente significativa.

Para verificar se existe uma diferença, estatisticamente significativa, entre os dois correlogramas, utiliza-se a metodologia proposta neste trabalho. Para a aplicação dos testes considerou-se 999 permutações no conjunto de dados observados e 40 distâncias.

Para o teste  $C$  obteve-se  $\Omega = 1,540$  ( $p = 0,016$ ), logo levando a não rejeição de  $H_0$ , ao nível de significância de 1%. Assim, a estrutura de dependência espacial das folhas minadas é igual à estrutura de dependência espacial das vespas predadoras para o mês de maior infestação do bicho-mineiro.

Para o teste K obteve-se  $\Phi = 0,572$  ( $p = 0,13$ ) e, assim, não se rejeita  $H_0$ , concluindo que a estrutura de dependência espacial é a mesma para as duas espécies.

Observa-se na Figura 13 (b) que a diferença entre os dois correlogramas observados encontra-se dentro do envelope de simulação. Este comportamento mostra que a estrutura de dependência espacial entre o número de folhas minadas e vespas é a mesma, ao nível de significância de 1%, o que corrobora os resultados obtidos pelo teste *C* e *K*.

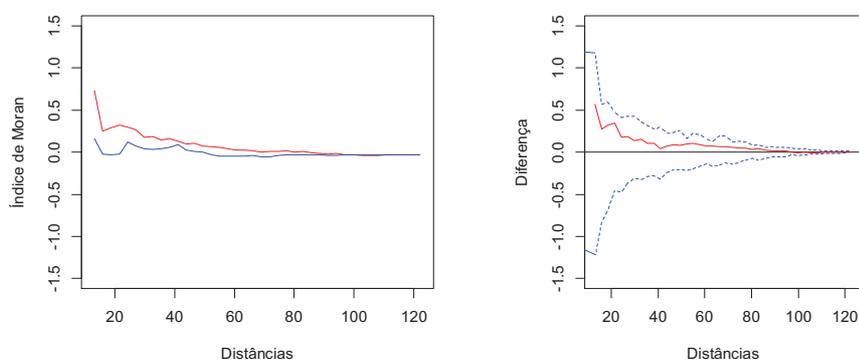


Figura 13 Correlogramas do número de folhas minadas (vermelho) e vespas predadoras (azul) (a) e envelope simulado com 999 permutações (b).

#### 4 CONCLUSÕES

Os métodos são capazes de testar a hipótese nula de que duas populações apresentam a mesma estrutura de dependência espacial em diferentes cenários. As metodologias propostas foram capazes de mostrar que as estruturas de dependência espacial das folhas minadas e das vespas predadoras são iguais para o mês de maior infestação do bicho-mineiro do cafeeiro.

## 5 REFERÊNCIAS

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific & Technical, 1995. 413 p.

CAMBARDELLA, C. A. et al. Field-scale variability of soil properties in Central Iowa soils. **Soil Science Society America Journal**, Madison, v. 58, n. 5, p. 1501-1511, Sept./Oct. 1994.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira: café safra 2014, segundo levantamento, maio/2014**. Disponível em:  
<[http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14\\_05\\_20\\_08\\_49\\_17\\_boletim\\_maio-2014.pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14_05_20_08_49_17_boletim_maio-2014.pdf)>. Acesso em: 24 jun. 2014.

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

DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. **Model based geostatistics**. New York: Springer, 2007. 243 p.

KOENIG, W. D.; KNOPS, J. M. H. Testing for spatial autocorrelation in ecological studies. **Ecography**, Copenhagen, v. 21, n. 4, p. 423-429, Aug. 1998.

LEGENDRE, P. Spatial autocorrelation: trouble or new paradigm? **Ecology**, Durham, v. 74, n. 6, p. 1659-1673, Sept. 1993.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, London, v. 37, p. 17-23, 1950.

ODEN, N. L. Assessing the significance of a Spatial Correlogram. **Geographical Analysis**, Columbus, v. 16, p. 1-16, 1984.

PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: analyses of phylogenetics and evolution in R language. **Bioinformatics**, Oxford, v. 20, n. 2, p. 289-290, Jan. 2004.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <<http://www.R-project.org/>>. Acesso em: 01 maio 2014.

REIS, P. R.; SOUZA, J. C. de. Pragas do cafeeiro. In: RENA, A. B. et al. (Ed.). **Cultura do cafeeiro: fatores que afetam a produtividade**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1986. p. 323-378.

REIS, P. R.; SOUZA, J. C. de; ZACARIAS, M. S. Alerta para o bicho-mineiro. **Cultivar**, Pelotas, v. 8, n. 84, p. 13-16, abr. 2006.

RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. GeoR: a package for geostatistical analysis. **R-News**, New York, v. 1, n. 2, p. 15-18, June 2001.

SCALON, J. D. et al. Spatial and temporal dynamics of coffee-leaf-miner and predatory wasps in organic coffee field in formation. **Ciência Rural**, Santa Maria, v. 41, n. 4, p. 646-652, abr. 2011a.

SCALON, J. D. et al. Spatial distribution of the coffee-leaf-miner (*Leucoptera coffeella* (Guérin-Mèneville&Perrottet, 1842)) in an organic coffee (*Coffea Arabica* L.) field in formation. **Coffee Science**, Lavras, v. 6, n. 3, p. 226-232, Sept./Dec. 2011b.

SOUZA, J. C.; REIS, P. R.; RIGITANO, R. L. O. **Bicho mineiro do cafeeiro: biologia, danos e manejo integrado**. 2. ed. Belo Horizonte: EPAMIG, 1998. 48 p. (Boletim Técnico, 54).

SYRJALA, S. E. A statistical test for a difference between the spatial distributions of two populations. **Ecology**, Durham, v. 77, n. 1, p. 75-80, Jan. 1996.

THOMSON, J. D. et al. Untangling multiple factors in spatial distributions: lilies, gophers, and rocks. **Ecology**, Durham, v. 77, n. 6, p. 1698-1715, Sept. 1996.

UPTON, G. J. G.; FINGLETON, B. **Spatial data analysis by example: point pattern and quantitative data**. New York: J. Wiley, 1985. v. 1, 410 p.

WALLER, L. A.; GOTWAY, C. A. **Applied spatial statistics for public health data**. New Jersey: J. Wiley, 2004. 519 p.

### 3 CONCLUSÕES GERAIS

No primeiro artigo foi apresentado o teste do Syrjala para verificar se duas populações apresentam a mesma distribuição espacial. Os resultados mostraram a efetividade do teste para comparar variáveis aleatórias de diversas populações associadas à praga bicho-mineiro do cafeeiro e seu predador natural, as vespas.

No segundo artigo foram propostos métodos para analisar a significância estatística de correlogramas. A partir dos resultados da análise de poder e do erro do tipo I pode-se dizer que os métodos propostos apresentaram bom desempenho nas situações avaliadas. A aplicação dos métodos em dados reais mostrou que os mesmos foram capazes de detectar a presença de dependência espacial, em várias escalas de distâncias, no número de minas do bicho-mineiro do cafeeiro que foram predadas pelas vespas.

No terceiro artigo foram propostos métodos para testar se duas populações apresentam a mesma estrutura de dependência espacial. Os resultados mostraram que os métodos propostos apresentaram resultados satisfatórios quando avaliados em casos típicos de dependência. Os resultados obtidos em dados reais permitiram concluir que a estrutura de dependência espacial da população de folhas minadas pelo bicho-mineiro do cafeeiro é a mesma da população de vespas predadoras.

## ANEXOS

Nestes anexos são apresentadas as rotinas com, os comandos do *software* R, utilizadas na tese. Nos anexos 1 e 2 constam as rotinas usadas no artigo 2, enquanto os anexos 3 e 4 constam as rotinas usadas no artigo 3.

**ANEXO 1:** Rotina com os comandos do software R para aplicação dos testes para verificar a significância de um correlograma.

---

```
library(ape)
moranFR=function(coords, var1, lengt, nsim ){
  # - coords: coordenadas (x, y)
  # - var1: variável referente a primeira população
  # - nsim: número de permutações
  # - lengt: número de distâncias
  # - N é o números de áreas
  x=coords$x
  y=coords$y
  dados.dists<- as.matrix(dist(cbind(x, y)))
  mini<- min(dados.dists[dados.dists>11])
  maxi<- max(dados.dists)
  inte<- round((maxi-mini)/lengt,digit=1)
  #
  # Obtenção do Índice de Moran para as diferentes distâncias (Correlograma)
  #
  dist <- seq(mini,maxi,by=inte)
  resu1 <- numeric(0)
  resu2 <- numeric(0)
  for(i in 1:length(dist)){
    dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[i])
```

```

resu1[i] <- Moran.I(var1, dados.dists.bin)$observed
resu2[i] <- -1/(N-1)
}
#
# Estatística de teste usando a distância de Cramer-Von Mises
#
lent<- length(dist)
D<-(resu1-resu2)^2
integ<-0
for(count in (2:(lent - 2))) {
  st=dist[count]-dist[count-1]
  integ <- integ + (D[count - 1] + 4 * D[count] + D[count + 1]) * st
}
integ <- integ/6
#
# Estatística de teste usando a distância de Kolmogorov
#
DK<-max(abs(resu1-resu2))
#
# Obtenção do valor-p
#
for(i in (2:nsim)) {
  resu11 <- numeric(0)
  resu22 <- numeric(0)
  for(i in 1:length(dist)){
    var11 <- sample(var1)
    dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[i])
    resu11[i] <- Moran.I(var11, dados.dists.bin)$observed
    resu22[i] <- -1/(N-1)
  }
#

```

```

lent <- length(dist)
t <- (resu11 - resu22)^2
teg <- 0
for(count in (2:(lent - 2))) {
  st1=dist[count]-dist[count-1]
  teg <- teg + (t[count - 1] + 4 * t[count] + t[count + 1]) * st1
}
teg <- teg/6
integ <- c(integ, teg)
#
test<-max(abs(resu11-resu22))
DK<-c(DK,test)
}
# Gráfico dos Correlogramas
y <- seq(-1,1,length.out=lent)
plot(dist,y,xlab="Distâncias", ylab="Índice de Moran", type="n")
lines(dist,resu1,col="red")
lines(dist,resu2,col="blue")
#
valork <- DK[DK >= DK[1]]
pvaluek <- length(valork) /nsim
valori <- integ[integ >= integ[1]]
pvaluei <- length(valori) /nsim
list(Kolmogorov=DK, est_Kolm=DK[1], p_valueKolm=pvaluek, Cramer=integ,
est_cramer=integ[1], p_valueCramer=pvaluei)
}

```

**ANEXO 2:** Rotina com os comandos do software R para a construção dos envelopes de simulação para testar a significância de um correlograma.

---

```
graphFR=function(coords, var1, lengt, nsim ){
  # - coords: coordenadas (x, y)
  # - var1: variável referente a primeira população
  # - nsim: número de permutações
  # - lengt: número de distâncias
  # - N é o número de áreas

  x=coords$x
  y=coords$y
  nareas=length(x)*length(y)
  dados.dists<- as.matrix(dist(cbind(x, y)))
  mini<- min(dados.dists[dados.dists>11])
  maxi<- max(dados.dists)
  inte<- round((maxi-mini)/lengt,digit=1)
  #
  dist <- seq(mini,maxi,by=inte)
  resu1 <- numeric(0)
  resu2 <- numeric(0)
  for(i in 1:length(dist)){
    dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[i])
    resu1[i] <- Moran.I(var1, dados.dists.bin)$observed
    resu2[i] <- -1/(N-1)
  }
  #
  D<-(resu1-resu2)
  #
  #
  hold<-matrix(0, nsim,length(dist))
  for(i in (1:nsim)) {
```

```
resu11 <- numeric(0)
resu22 <- numeric(0)
for(j in 1:length(dist)){
  var11 <- sample(var1)
  dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[j])
  resu11[j] <- Moran.I(var11, dados.dists.bin)$observed
  resu22[j] <- -1/(N-1)
}
hold[i, ] <- resu11 - resu22
}
#
up <- apply(hold, 2, max)
low <- apply(hold, 2, min)
#
# Gráfico do envelope de simulação
#
par(pty = "s")
y <- seq(-1, 1, length.out = length(dist))
plot(dist, y, xlab = "Distâncias", ylab = "Diferença", type = "n")
abline(h = 0)
lines(dist, D, col = "red")
lines(dist, up, lty = 2, col = "blue")
lines(dist, low, lty = 2, col = "blue")
}
```

**ANEXO 3:** Rotina com os comandos do software R para aplicação dos testes para a diferença entre as estruturas de dependência espacial de duas populações.

---

```

library(ape)
moranFR=function(coords, var1, var2, lengt, nsim ){
  # - coords: coordenadas (x, y)
  # - var1: variável referente a primeira população
  # - var2: variável referente a segunda população
  # - nsim: número de permutações
  # - lengt: número de distâncias
  x=coords$x
  y=coords$y
  dados.dists<- as.matrix(dist(cbind(x, y)))
  mini<- min(dados.dists[dados.dists>11])
  maxi<- max(dados.dists)
  inte<- round((maxi-mini)/lengt,digit=1)
  #
  # Obtenção do Índice de Moran para as diferentes distâncias (Correlograma)
  #
  dist <- seq(mini,maxi,by=inte)
  resu1 <- numeric(0)
  resu2 <- numeric(0)
  for(i in 1:length(dist)){
    dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[i])
    resu1[i] <- Moran.I(var1, dados.dists.bin)$observed
    resu2[i] <- Moran.I(var2, dados.dists.bin)$observed
  }
  #
  # Estatística de teste usando a distância de Cramer-Von Mises
  #
  lent<- length(dist)

```

```

D<-(resu1-resu2)^2
integ<-0
for(count in (2:(lent - 2))) {
  st=dist[count]-dist[count-1]
  integ <- integ + (D[count - 1] + 4 * D[count] + D[count + 1]) *
    st
}
integ <- integ/6
#
# Estatística de teste usando a distância de Kolmogorov
#
DK<-max(abs(resu1-resu2))
#
# Obtenção do valor-p
#
for(i in (2:nsim)) {
  resu11 <- numeric(0)
  resu22 <- numeric(0)
  for(i in 1:length(dist)){
    var11 <- sample(var1)
    var22 <- sample(var2)
    dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[i])
    resu11[i] <- Moran.I(var11, dados.dists.bin)$observed
    resu22[i] <- Moran.I(var22, dados.dists.bin)$observed
  }
#
#
lent <- length(dist)
t <- (resu11 - resu22)^2
teg <- 0
for(count in (2:(lent - 2))) {

```

```

st1=dist[count]-dist[count-1]
teg <- teg + (t[count - 1] + 4 * t[count] + t[count + 1]) *
  st1
}
teg <- teg/6
integ <- c(integ, teg)
#
#
test<-max(abs(resu11-resu22))
DK<-c(DK,test)
}
# Gráfico dos Correlogramas
y<-seq(-1,1,length.out=lent)
plot(dist,y,xlab="Distâncias", ylab="Índice de Moran", type="n")
lines(dist,resu1,col="red")
lines(dist,resu2,col="blue")
#
#
valork <- DK[DK >= DK[1]]
pvaluek <- length(valork) /nsim
valori <- integ[integ >= integ[1]]
pvaluei <- length(valori) /nsim
list(Kolmogorov=DK, est_Kolm=DK[1], p_valueKolm=pvaluek, Cramer=integ,
est_cramer=integ[1], p_valueCramer=pvaluei)
}

```

**ANEXO 4:** Rotina com os comandos do software R para a construção dos envelopes de simulação para a diferença entre os correlogramas.

---

```
graphFR=function(coords, var1, var2, lengt, nsim ){
  # - coords: coordenadas (x, y)
  # - var1: variável referente a primeira população
  # - var2: variável referente a segunda população
  # - nsim: número de permutações
  # - lengt: número de distâncias
  x=coords$x
  y=coords$y
  nareas=length(x)*length(y)
  dados.dists<- as.matrix(dist(cbind(x, y)))
  mini<- min(dados.dists[dados.dists>11])
  maxi<- max(dados.dists)
  inte<- round((maxi-mini)/lengt,digit=1)
  #
  dist <- seq(mini,maxi,by=inte)
  resu1 <- numeric(0)
  resu2 <- numeric(0)
  for(i in 1:length(dist)){
    dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[i])
    resu1[i] <- Moran.I(var1, dados.dists.bin)$observed
    resu2[i] <- Moran.I(var2, dados.dists.bin)$observed
  }
  #
  D<-(resu1-resu2)
  #
  hold<-matrix(0, nsim,length(dist))
  for(i in (1:nsim)) {
    resu11 <- numeric(0)
```

```
resu22 <- numeric(0)
for(j in 1:length(dist)){
  var11 <- sample(var1)
  var22 <- sample(var2)
  dados.dists.bin <- (dados.dists > 0 & dados.dists <= dist[j])
  resu11[j] <- Moran.I(var11, dados.dists.bin)$observed
  resu22[j] <- Moran.I(var22, dados.dists.bin)$observed
}
hold[i, ] <- resu11 - resu22
}
#
up <- apply(hold, 2, max)
low <- apply(hold, 2, min)
#
# Gráfico do envelope de simulação
#
par(pty = "s")
y <- seq(-1, 1, length.out = length(dist))
plot(dist, y, xlab = "Distâncias", ylab = "Diferença", type = "n")
abline(h = 0)
lines(dist, D, col = "red")
lines(dist, up, lty = 2, col = "blue")
lines(dist, low, lty = 2, col = "blue")
}
```