

## ANÁLISE DA ESTRUTURA DE DEPENDÊNCIA ESPACIAL EM DADOS DE ÁREAS

Franciella Marques da COSTA<sup>1</sup>  
João Domingos SCALON<sup>2</sup>

- RESUMO: A análise da estrutura de dependência espacial em dados de área é de extrema importância em diversas áreas de conhecimento científico. Assim, o objetivo deste artigo é apresentar métodos estatísticos para verificar a significância estatística de um correlograma e se duas populações, distribuídas na mesma região, apresentam a mesma estrutura de dependência espacial. 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 bichomineiro 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; autocorrelação; aleatorização; envelopes de simulação; cafeicultura.

### 1 Introdução

A análise da estrutura de dependência espacial de uma determinada espécie, bem como a avaliação de que duas espécies, ocupando o mesmo espaço geográfico, apresentem a mesma estrutura de dependência espacial é uma atividade muito importante em diversas áreas do conhecimento científico, como por exemplo, no controle biológico de pragas, onde é importante caracterizar a estrutura de dependência espacial da praga e do seu predador (KOENIG; KNOPS, 1998; LEGENDRE, 1993).

Em geral, a dependência espacial é analisada utilizando coeficientes de autocorrelação que descrevem o grau de sincronismo que variáveis exibem dentro de um espaço geográfico. 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 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

---

<sup>1</sup> Universidade Federal de Uberlândia - UFU, Faculdade de Ciências Integradas do Pontal - FACIP, CEP: 38.304-402, Ituiutaba, Minas Gerais, Brasil. E-mail: [franciella@pontal.ufu.br](mailto:franciella@pontal.ufu.br)

<sup>2</sup> Universidade Federal de Lavras - UFLA, Departamento de Ciências Exatas - DEX, CEP: 37200-000, Lavras, Minas Gerais, Brasil, E-mail: [scalon@dex.ufla.br](mailto:scalon@dex.ufla.br)

estatística da dependência espacial (THOMSON et al., 1996). Apesar de existir algumas propostas de testes de significância, como pode ser visto em Oden (1984), em geral, essa caracterização é feita quase que, exclusivamente, através de inspeção visual do correlograma.

Diante do exposto, este trabalho tem como objetivo propor dois testes de hipótese e um procedimento gráfico para verificar a significância estatística de um correlograma. Este trabalho também propõe a extensão desses métodos para que se possa 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, métodos de computação intensiva e nas medidas de distâncias de Cramer-von Mises e Kolmogorov.

## 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. Evidentemente, qualquer outro índice global de associação espacial (e.g. Geary) para análise da dependência espacial para dados de área pode ser utilizado.

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, em que  $w_{ij} = 1$  se  $d_{ij} < d$  e  $w_{ij} = 0$ , caso contrário, em que  $d$  é uma distância definida pelo pesquisador e  $d_{ij}$  é a distância euclidiana entre os centros de massa das regiões  $i$  e  $j$ .

O índice global de Moran pode assumir qualquer valor no domínio dos reais, entretanto, na maioria dos casos fica restrito entre -1 e 1. 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, obtido para diferentes ordens de vizinhança  $k$ , em função de  $k$ , é chamado de correlograma. 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 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 (ODEN, 1984).

Com o intuito de obter informações a respeito da dependência espacial de um conjunto de dados para diferentes escalas, este trabalho propõe a construção do correlograma a partir do índice global de Moran, calculado para diferentes distâncias, considerando sempre o vizinho de 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.

## 2.2 Testes de hipótese para significância estatística de correlogramas

A hipótese nula ( $H_0$ ) a ser testada é que a população não apresenta autocorrelação espacial nas distâncias analisadas, enquanto que a hipótese alternativa ( $H_1$ ) é que a população apresenta autocorrelação espacial em, pelo menos, uma das distâncias avaliadas.

Para testar a hipótese nula ( $H_0$ ), primeiramente, o índice global de Moran é calculado para diferentes distâncias ( $d$ ), para  $\min < d < \max$ , em que  $\min$  é a distância mínima entre os centros de massa das áreas e  $\max$  é a distância máxima entre os centros de massa das áreas. Depois, constrói-se o correlograma, ou seja, gráfico dos índices globais de Moran em função das distâncias ( $d$ ) e, finalmente, define-se as 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

$$C = \int_0^{d_0} (I(d) - E(I))^2 \delta d, \quad (2)$$

em que  $d_0$  é a distância máxima entre dois centros de massa,  $I(d)$  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 utilizada 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  $-p$  do teste estatístico é obtido pela proporção dos valores das estatísticas, obtidas por meio das aleatorização, que são maiores ou iguais a estatística de teste calculada com os dados observados (MANLY, 1994).

### 2.3 Envelopes simulados para verificar significância estatística de um correlograma

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:

1. Constroem-se os correlogramas  $I(d)$  (correlograma para a população em estudo) e  $I'(d)$  (correlograma sob ausência de correlação espacial);
2. Obtém-se a diferença  $I_1(d) = I(d) - I'(d)$  para os dados observados;
3. Permutam-se os dados observados  $m$  vezes e calcula-se  $I_i(d) = I_i(d) - I'_i(d)$  para  $i = 2, 3, \dots, m$ ;
4. Define-se o limite superior e inferior do envelope simulado como sendo:
 
$$U(d) = \max\{I_i(d)\}, \quad i = 2, 3, \dots, m \quad \text{e}$$

$$L(d) = \min\{I_i(d)\}, \quad i = 2, 3, \dots, m;$$
5. Constrói-se um gráfico com os valores  $I_1(d)$ ,  $U(d)$  e  $L(d)$  contra as distâncias  $d$ .

De acordo com Bailey e Gatrell (1995) e Diggle (2003), neste procedimento é válida a propriedade  $P\{I_1(d) > U(d)\} = P\{I_1(d) < L(d)\} = \frac{1}{m+1}$ .

Obtidos o envelope de simulação sua interpretação é bastante simples. O correlograma não será estatisticamente significativo se a linha de  $I_1(d)$  estiver totalmente contida entre o limite superior e o limite inferior do envelope simulado. Caso contrário, se  $I_1(d)$  estiver acima de  $U(d)$ , para algum  $d$ , o correlograma é considerado estatisticamente significativo em favor da dependência espacial positiva, enquanto se  $I_1(d)$  estiver abaixo de  $L(d)$ , para algum  $d$ , o correlograma é considerado estatisticamente significativo em favor da dependência espacial negativa.

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

Os testes de hipótese para igualdade entre duas estruturas de autocorrelação espacial, apresentados nesta seção são uma extensão dos testes apresentados na Seção 2.2.

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.

Nas estatísticas de teste utilizadas na Seção 2.2 utiliza-se o correlograma da população em estudo e o correlograma na ausência completa de autocorrelação espacial. Nas estatísticas de teste propostas nesta seção utilizam-se os correlogramas das duas populações em estudo.

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

$$C^* = \int_0^{d_0} \left( I^{(1)}(d) - I^{(2)}(d) \right)^2 \delta d. \quad (4)$$

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

$$K^* = \max |I^{(1)}(d) - I^{(2)}(d)|, \quad (5)$$

em que,  $d_0$  é a distância máxima entre dois centros de massa. Nas duas estatísticas de teste,  $I^{(1)}(d)$  e  $I^{(2)}(d)$  são respectivamente, o correlograma referente a população 1 e a população 2.

Para conduzir os testes de hipóteses, utilizando as duas estatísticas apresentadas anteriormente, recorrem-se a aleatorização, conforme apresentada anteriormente.

## 2.5 Envelopes simulados para verificar a igualdade entre duas estruturas de autocorrelação espacial

O procedimento para a construção de envelopes simulados, seguindo o que foi apresentado anteriormente, é dado pelos seguintes passos:

1. Obtêm-se os correlogramas  $I^{(1)}(d)$  e  $I^{(2)}(d)$  para as duas populações em estudo;
2. Calcula-se a diferença  $I_1(d) = I^{(1)}(d) - I^{(2)}(d)$  para os dados observados;
3. Permuta-se os dados  $m$  vezes, em seguida calcula-se  $I_i(d) = I_i^{(1)}(d) - I_i^{(2)}(d)$  para  $i = 2, 3, \dots, m$ ;
4. Obtém-se o limite superior  $U(d)$  e inferior  $L(d)$  dos envelopes simulados como sendo:
 
$$U(d) = \max\{I_i(d)\}, \quad i = 2, 3, \dots, m \quad \text{e}$$

$$L(d) = \min\{I_i(d)\}, \quad i = 2, 3, \dots, m;$$
5. Constrói-se um gráfico com os valores  $I_1(d)$ ,  $U(d)$  e  $L(d)$  contra as distâncias  $d$ .

Se a linha de  $I_1(d)$  estiver entre o limite superior e o limite inferior dos envelopes simulados, existe evidências de que a estrutura de dependência espacial das duas populações é a mesma. Se a linha  $I_1(d)$  não estiver totalmente dentro dos limites do envelope simulado, para alguma distância  $d$ , então a estrutura de dependência espacial das duas populações é considerada diferente para aquela distância.

## 2.6. Estudo do erro do tipo I e do poder dos testes de inexistência de autocorrelação espacial.

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 centros de massa de cada área. Foi considerada uma região com 10x10 unidades de áreas quadradas.

Deve-se observar que a 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 por  $IDE = \frac{C_0}{C_0 + C_1}$ , 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 10x10 unidades de área em uma grade regular, em que as coordenadas dos cruzamentos da grade regular representam os centros de massa de cada área.

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 efeito pepita fixos, 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.

Tabela 1 - Configurações utilizadas 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

Foram simulados 1000 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.

## 2.7 Estudos de casos típicos de duas estruturas de dependência espacial

Com o objetivo de verificar o comportamento dos métodos propostos para comparar as estruturas de dependência espacial de duas populações (seção 2.4), foi realizado um estudo, por meio de simulação, considerando uma estrutura de dependência espacial forte, em uma grade regular de 100 pontos, representando os centros de massa de cada uma das 100 áreas da região com área de  $10 \times 10$  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 centros de massa 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, efeito pepita igual a 3 e alcance igual a 40 e 60. Para a classificação quanto ao grau de dependência espacial foi utilizado o índice proposto por Cambardella et al. (1994) apresentado anteriormente.

Para obter os dados da segunda população, utilizou-se os dados da primeira população e adicionou-se um termo aleatório seguindo 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 espacial radial. Deve-se observar que à medida que se aumenta a variabilidade espacial radial, sendo essa acrescentada em um conjunto de dados com dependência espacial, o resultado é um conjunto de dados que tende à completa independência espacial (ISAAKS; SRIVASTAVA, 1989) e, conseqüentemente, o seu correlograma tende a se afastar do correlograma da população 1.

## 2.8 Softwares

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

## 3 Resultados e discussão

### 3.1 Erro do tipo I e poder dos testes de correlogramas

Foram simulados 1000 experimentos sob a hipótese de completa aleatoriedade espacial ( $H_0$  verdadeira) e aplicados 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, patamar igual a 80 e efeito pepita variável.

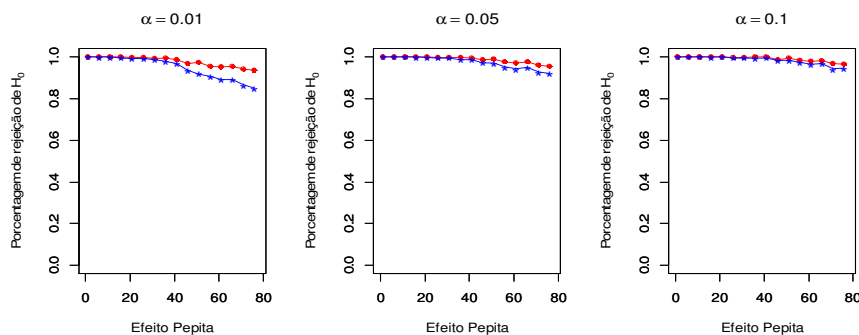


Figura 1 - Porcentagem de rejeição de  $H_0$  para os testes C (vermelho) e K (azul), seguindo o modelo de dependência espacial  $Sph(x, 80, 40)$  com valores para o efeito pepita (eixo  $x$ ) iguais a 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71 e 76.

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.



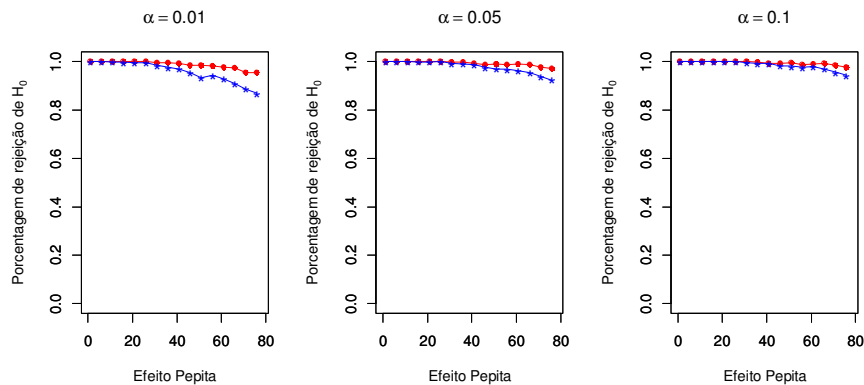


Figura 2 - Porcentagem de rejeição de  $H_0$  para os testes  $C$  (vermelho) e  $K$  (azul), seguindo o modelo de dependência espacial  $Sph(x, 80, 60)$  com valores para o efeito pepita (*eixo x*) iguais a 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71 e 76.

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 de dependência espacial

Para testar a hipótese nula de que o correlograma de uma situação típica de dependência espacial não é estatisticamente significativa, considerou-se uma região dividida em 100 áreas unitárias (em metros). Os correlogramas foram construídos

considerando uma distância máxima de 120 metros. Os testes de hipóteses e os envelopes simulados foram obtidos a partir de 999 permutações.

Sob a suposição que os dados de área apresentam forte dependência espacial,  $Sph(1, 80, 60)$ , obteve-se  $C = 6,701 (p = 0,001)$  e  $K = 0,625 (p = 0,001)$ , levando a rejeição da hipótese nula de inexistência de autocorrelação espacial. Analisando a Figura 3, observa-se que o correlograma observado fica acima do limite superior dos envelopes simulados e, portanto, rejeitando a hipótese de inexistência de correlação, em favor da correlação espacial positiva, para distâncias inferiores a 80 metros.

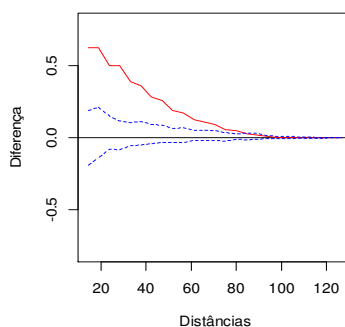


Figura 3 - Correlograma observado (vermelho) para dados de área com forte dependência espacial,  $Sph(1, 80, 60)$  e envelopes simulados (azul) com 999 permutações.

Sob a suposição que os dados de área apresentam moderada dependência espacial,  $Sph(41, 80, 60)$ , obteve-se  $C = 4,897 (p = 0,001)$  e  $K = 0,481 (p = 0,001)$ , levando a rejeição da hipótese nula de inexistência de autocorrelação espacial. Analisando a Figura 4, observa-se que o correlograma observado fica acima do limite superior dos envelopes simulados e, portanto, rejeitando a hipótese de inexistência de correlação, em favor da correlação espacial positiva, para distâncias inferiores a 80 metros.

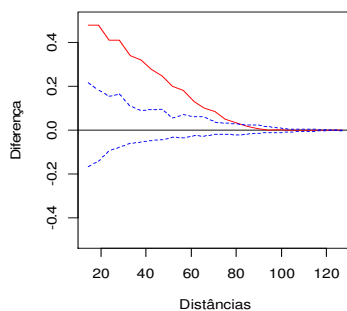


Figura 4 - Correlograma observado (vermelho) para dados de área com moderada dependência espacial,  $Sph(41, 80, 60)$  e envelopes simulados (azul) com 999 permutações.

Sob a suposição que os dados de área apresentam fraca dependência espacial,  $Sph(76, 80, 60)$ , obteve-se  $C = 0,695 (p = 0,001)$  e  $K = 0,256 (p = 0,001)$ , levando

a rejeição da hipótese nula de inexistência de autocorrelação espacial. Analisando a Figura 5, observa-se que o correlograma observado fica acima do limite superior dos envelopes simulados apenas para pequenas escalas e, portanto, a hipótese de inexistência de correlação, em favor da correlação espacial positiva, é rejeitada para distâncias inferiores a 30 metros.

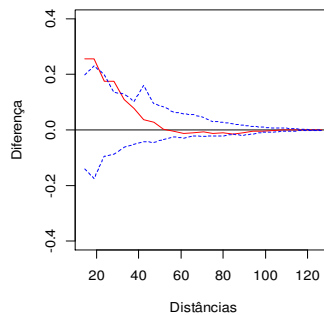


Figura 5 - Correlograma observado (vermelho) para dados de área com fraca dependência espacial,  $Sph(76, 80, 60)$  e envelopes simulados (azul) com 999 permutações.

Sob a suposição que os dados de área não apresentam dependência espacial, obteve-se  $C = 0,009$  ( $p = 0,982$ ) e  $K = 0,035$  ( $p = 0,963$ ), levando a aceitação da hipótese nula de inexistência de autocorrelação espacial. Analisando a Figura 6, observa-se que o correlograma observado está dentro dos limites dos envelopes simulados para todas as distâncias analisadas.

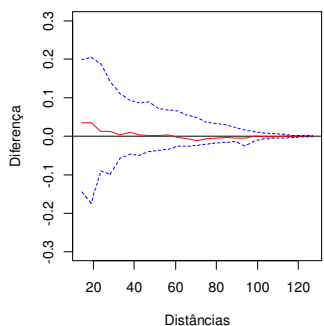


Figura 6 - Correlograma observado (vermelho) para dados de área com ausência de dependência espacial,  $Sph(76, 80, 60)$  e envelopes simulados (azul) com 999 permutações.

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. Pode-se observar ainda que os métodos propostos rejeitaram a hipótese nula de ausência de correlação espacial em todos os casos típicos que apresentavam algum tipo

de dependência espacial (forte, moderada e fraca) e não rejeitou a hipótese nula no único caso com ausência de dependência espacial.

### 3.3 Estudos de casos típicos da análise de duas estruturas de dependência espacial

A Figura 7 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 7 que à medida que se aumenta a variabilidade radial e, portanto, faz a população 2 caminhar no sentido da ausência de dependência espacial, aumenta-se a diferença entre os correlogramas das duas populações.

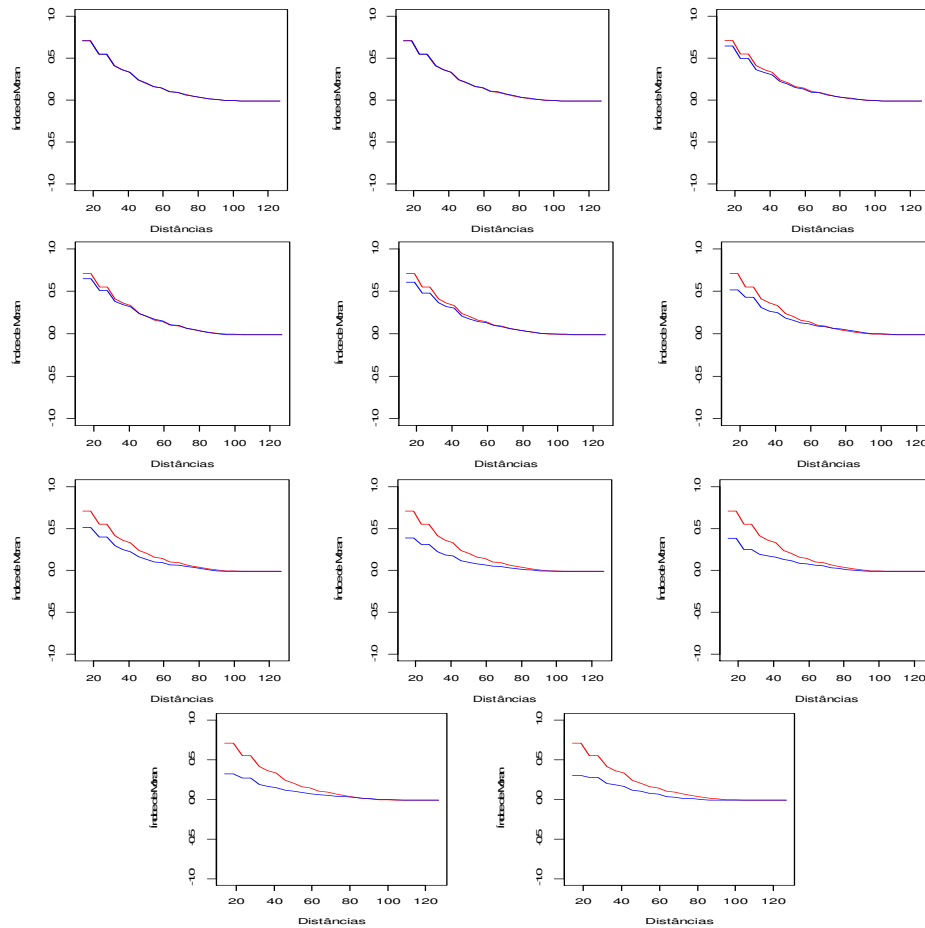


Figura 7 - Correlogramas da população 1 (vermelho) e população 2 (azul) obtidos com a configuração  $Sph(3, 60, 60)$  considerando as variabilidades radiais iguais a 0,5; 1; 2; 3; 4; 5; 6; 7; 8; 9 e 10, respectivamente.

Para cada variabilidade radial considerada foram gerados 100 casos típicos de dados de área seguindo o modelo Sph(3, 60, 60) e, portanto, exibindo forte dependência espacial. Para cada um desses casos foram aplicados os testes para a hipótese nula de igualdade entre as duas estruturas de autocorrelação espacial apresentados na seção 2.4. A Figura 8 mostra a porcentagem de casos típicos que tiveram a hipótese nula rejeitada aos níveis de significância iguais a 1%, 5% e 10%.

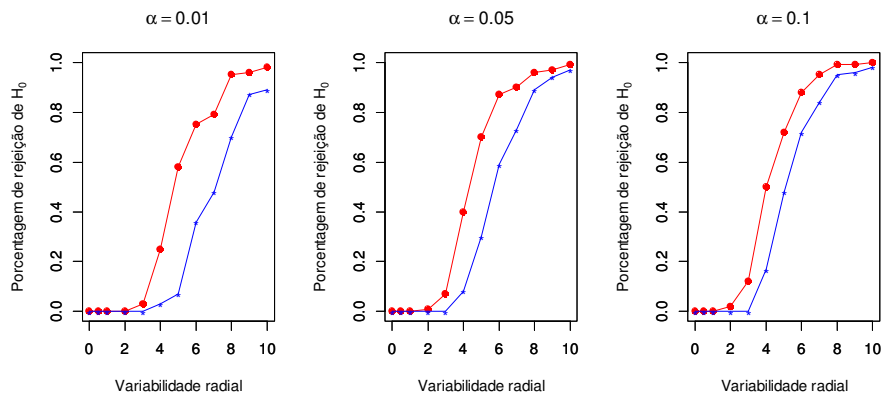


Figura 8 - Porcentagem de casos em que a hipótese nula de igualdade entre duas estruturas de dependência espacial foi rejeitada, para diferentes variabilidades radiais e níveis de significância, pelos testes  $C^*$  (vermelho) e  $K^*$  (azul).

De modo geral, pode-se observar na Figura 8 que a porcentagem de rejeição da hipótese nula de igualdade entre as duas estruturas de dependência espacial aumentou à medida que se aumentou a variabilidade radial, ou seja, à medida que os correlogramas vão se distanciando o teste tende a considerar as duas populações com estruturas de dependência espacial diferentes. Esse comportamento é observado nas duas estatísticas de teste. A porcentagem de rejeição da hipótese nula quando se usa a estatística de Cramer-von Mises ( $C^*$ ) foi maior ou igual que a porcentagem de rejeição observada para a estatística de Kolmogorov ( $K^*$ ) para todas as variabilidades radiais e níveis de significância analisados. Esse tipo de comportamento dessas estatísticas de testes já foi observado por Diggle (2003) quando utilizadas para testar a hipótese nula de completa aleatoriedade espacial em configurações pontuais.

### 3.4 Análises de casos típicos de comparação entre duas estruturas de dependência espacial

Para testar a hipótese nula de que duas populações apresentam a mesma estrutura de dependência espacial, considerou-se uma região dividida em 100 áreas unitárias (em metros). Os correlogramas das duas populações foram construídos considerando uma distância máxima de 120 metros. Os testes de hipóteses e os envelopes simulados foram obtidos a partir de 999 permutações.

Considerando que a população 1 apresenta forte dependência espacial com dados obtidos a partir do modelo Sph(3, 60, 60) e que a população 2 também apresenta forte dependência espacial com dados obtidos a partir do mesmo modelo da população 1,

entretanto com o acréscimo de uma variabilidade radial espacial igual a 1, obteve-se  $K^* = 0,019$  ( $p = 1$ ) e  $C^* = 0,007$  ( $p = 1$ ). Esses resultados levam a aceitação da hipótese nula que as duas populações apresentam a mesma estrutura de dependência espacial.

Analisando a Figura 9 (esquerda), observa-se que os dois correlogramas estão sobrepostos e que a diferença existente entre eles é praticamente zero para todas as distâncias (Figura 9 - direita). Assim, o correlograma da diferença fica dentro dos envelopes de simulação e, conseqüentemente, levando a aceitação da hipótese nula de que as duas populações apresentam a mesma estrutura de dependência espacial.

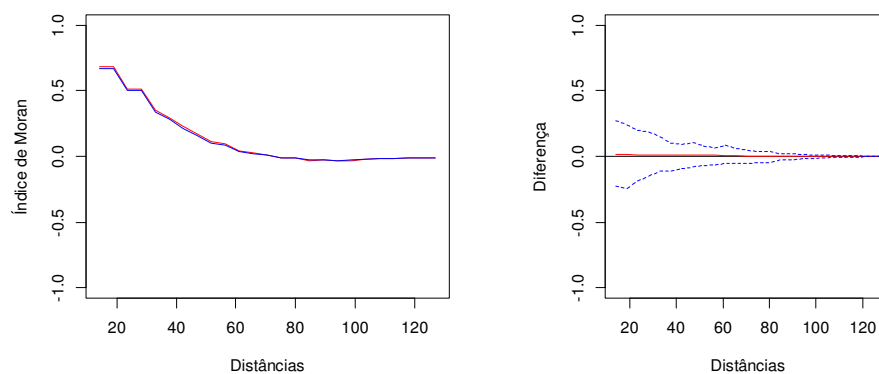


Figura 9 - Esquerda - Correlogramas da população seguindo o modelo de dependência espacial  $Sph(3, 60, 60)$  (vermelho) e seguindo o mesmo modelo adicionando uma variabilidade radial espacial igual a 1 (azul). Direita - Correlograma da diferença (vermelho) com os envelopes de simulação (linhas tracejadas em azul) sob a suposição que as duas populações apresentam a mesma estrutura de dependência espacial.

Considerando que a população 1 apresenta forte dependência espacial com dados obtidos a partir do modelo  $Sph(3, 60, 60)$  e que a população 2 apresenta uma dependência espacial fraca com dados obtidos a partir do mesmo modelo da população 1, entretanto com o acréscimo de uma variabilidade radial espacial igual a 10, obteve-se  $K^* = 0,44$  ( $p = 0,001$ ) e  $C^* = 3,08$  ( $p = 0,001$ ). Esses resultados levam a rejeição enfática da hipótese nula que as duas populações apresentam a mesma estrutura de dependência espacial.

Analisando a Figura 10 (esquerda), observa-se que os dois correlogramas estão separados até a distância de 60 metros e que a diferença existente entre eles é estatisticamente significativa até a distância de 50 metros (Figura 10 - direita). Assim, o correlograma da diferença fica acima do limite superior dos envelopes de simulação e, conseqüentemente, levando a rejeição da hipótese nula de que as duas populações apresentam a mesma estrutura de dependência espacial.

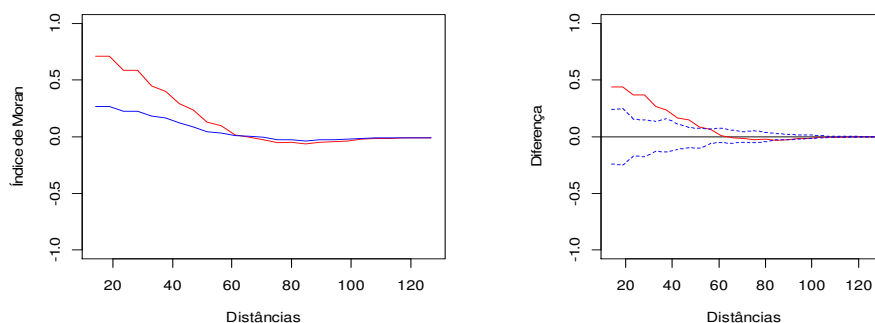


Figura 10 - Esquerda - Correlogramas da população seguindo o modelo de dependência espacial  $Sph(3,60,60)$  (vermelho) e seguindo o mesmo modelo adicionando uma variabilidade radial espacial igual a 10 (azul). Direita - Correlograma da diferença (vermelho) com os envelopes de simulação (linhas tracejadas em azul) sob a suposição que as duas populações apresentam a mesma estrutura de dependência espacial.

### 3.5 Estudo de caso

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 do bicho-mineiro do cafeeiro pode ser realizado através de predadores naturais como as vespas (*Hymenoptera Vespidae*) (REIS; SOUZA, 1986).

As vespas predadoras rasgam a epiderme das folhas no local das minas e retiram as lagartas do bicho-mineiro que servem de alimento para as vespas (SOUZA; REIS; RIGITANO, 1998).

Para o efetivo controle biológico do bicho-mineiro do cafeeiro, os pesquisadores precisam entender como essas duas espécies se distribuem espacialmente no cafezal. Para tal, foi conduzido um experimento em um hectare de produção de café orgânico da Cultivar Catucaí, situado no município de Santo Antônio do Amparo, MG. Essa região foi dividida em 35 áreas de amostragem e em cada uma dessas áreas foi coletado o número de minas predadas em 10 folhas selecionadas aleatoriamente dos cafeeiros e o número de vespas predadoras coletadas em armadilhas adesivas. Foram utilizados os dados referentes apenas ao mês de setembro de 2006 (mês de maior infestação do bicho-mineiro).

A Figura 11 apresenta os centros de massa das 35 áreas na região de estudo. Cada número associado ao centro de massa da área indica as contagens de bicho-mineiro (esquerdo) e vespas (direito) referentes a cada área.

Para conduzir a análise da estrutura de dependência espacial das duas populações foram construídos correlogramas considerando uma distância máxima de 120 metros. Os testes de hipóteses e os envelopes simulados foram obtidos a partir de 999 permutações.

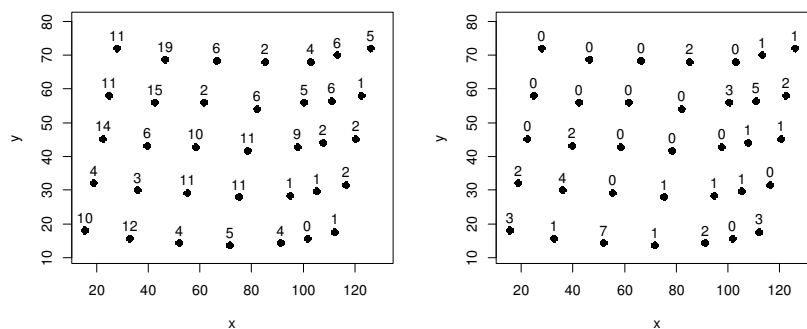


Figura 11 - Contagens referentes ao número de minas predadas (esquerdo) e vespas predadoras (direito) para as 35 áreas na plantação de café orgânico.

Aplicando os testes de hipóteses para a ausência de autocorrelação espacial das minas predadas obteve-se  $C = 5,836$  ( $p = 0,001$ ) e  $K = 0,639$  ( $p = 0,017$ ), rejeitou-se a hipótese nula e, conseqüentemente, pode-se afirmar que o número de minas predadas exibe autocorrelação espacial. Aplicando os mesmos testes de hipóteses para o número de vespas predadoras obteve-se  $C = 0,228$  ( $p = 0,617$ ) e  $K = 0,189$  ( $p = 0,673$ ), levando a aceitação da hipótese nula e, conseqüentemente, pode-se afirmar que o número de vespas predadoras não exibe autocorrelação espacial.

Analisando a Figura 12 - esquerda, observa-se que o correlograma observado do número de minas predadas fica sobreposto ao limite superior dos envelopes simulados e, portanto, mostrando evidências que a hipótese de inexistência de correlação deve ser rejeitada, em favor da correlação espacial positiva, para distâncias inferiores a 90 metros. Por outro lado, o correlograma observado do número de vespas predadoras fica dentro dos envelopes simulados (Figura 12 - direita) e, portanto, mostrando que a hipótese nula de inexistência de correlação deve ser aceita.

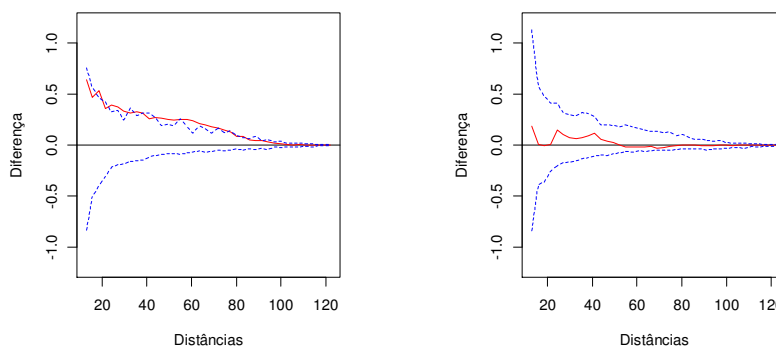


Figura 12 - Correlogramas observados (vermelho) para o número de minas predadas (esquerda) e número de vespas predadoras (direita) com os respectivos envelopes simulados (azul) com 999 permutações.



As análises anteriores parecem indicar que as duas espécies apresentam estruturas de dependência espacial diferentes. Para verificar essa suspeita, analisa-se a diferença entre os dois correlogramas. Para tal, aplica-se os testes de hipótese e constrói-se os envelopes de simulação apresentados nos itens 2.4 e 2.5, respectivamente, considerando uma distância máxima de 120 metros e 999 permutações.

Os testes de hipótese apresentaram resultados conflitantes. Enquanto o teste baseado na estatística de Kolmogorov levou a aceitação da hipótese nula ( $K^* = 0,537$ ,  $p = 0,157$ ), indicando que as duas populações apresentam a mesma estrutura de dependência espacial, o teste baseado na estatística de Cramer-von Mises levou a rejeição da hipótese nula ( $C^*=4,725$ ,  $p = 0,001$ ), indicando que as duas populações apresentam estrutura de dependência espacial diferentes.

Analisando a Figura 13 (esquerda), observa-se que os dois correlogramas estão ligeiramente separados até a distância de 100 metros. Entretanto, a diferença existente entre eles é ligeiramente superior ao limite superior dos envelopes de simulação entre as distâncias de 55 e 75 metros (Figura 13 - direita), indicando que para essas distâncias a diferença é estatisticamente significativa e, conseqüentemente, levando a rejeição da hipótese nula de que as duas populações apresentam a mesma estrutura de dependência espacial.

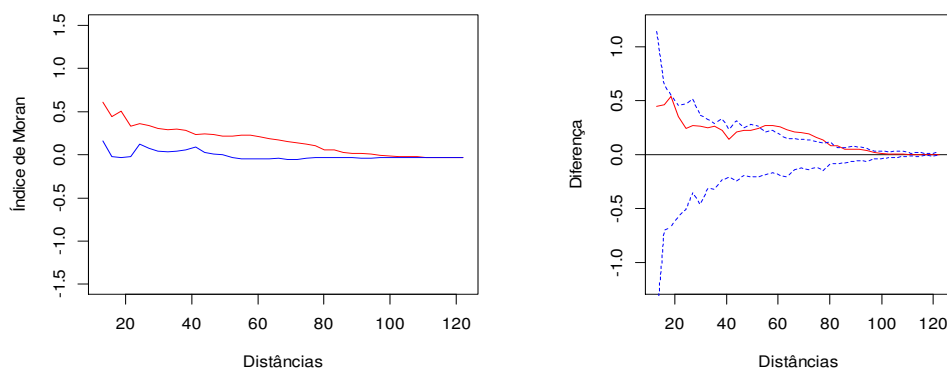


Figura 13 - Esquerda - Correlogramas das populações de minas predadas (vermelho) e vespas predadoras (azul). Direita - Correlograma da diferença (vermelho) com os envelopes de simulação (linhas tracejadas em azul) sob a suposição que as duas populações apresentam a mesma estrutura de dependência espacial.

## Conclusão

Os resultados obtidos mostraram que os métodos propostos permitem não somente analisar a dependência espacial de uma população, mas também verificar se duas populações, distribuídas na mesma região, apresentam a mesma estrutura de dependência espacial. Os métodos propostos foram aplicados a dados relacionados ao bicho mineiro do cafeeiro. Os resultados mostraram que o número de minas predadas exibe autocorrelação espacial, enquanto o número de vespas predadoras do bicho mineiro não exibe

autocorrelação espacial. O teste baseado na estatística de Cramer-von Mises e o envelope de simulação indicaram que as duas populações apresentam estruturas de dependência espacial diferentes. Já o teste baseado na estatística de Kolmogorov indicou que as duas populações apresentam a mesma estrutura de dependência espacial. Entretanto, para uma melhor avaliação dos métodos propostos sugere-se maiores estudos considerando outros cenários de dependência espacial.

### **Agradecimentos**

Os autores agradecem ao Dr. Mauricio Sergio Zacarias (EMBRAPA/Café) pelos dados sobre o bicho-mineiro do cafeeiro e a CAPES pelo apoio financeiro.

COSTA, F. M.; SCALON, J. D. Analysis of the spatial dependence structure in a real data. *Rev. Bras. Biom.* São Paulo, v.33, n.2, p.197-215, 2015.

*ABSTRACT: The analysis of spatial dependence structure in a real data is of utmost importance in many areas of scientific knowledge. The aim of this paper is to present statistical methods to verify both the statistical significance of a correlogram and whether two populations, distributed in the same region, have the same spatial dependence structure. Whenever the method allows analysis of test power and type I error. The proposed methods are also applied on real data related to the coffee leaf miner (Leucoptera coffeella) from an organic coffee plantation. The results allow us to affirm that the proposed methods can be useful for achieving the aims for which they were developed.*

*KEYWORDS: Moran index; autocorrelation; randomization; envelopes simulation; Coffee production.*

### **Referências**

- BAILEY, T. C.; GATRELL, A. C. *Interactive spatial data analysis*. Essex: Longman Scientific & Technical, 1995. 413p.
- 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, 1994.
- DIGGLE, P. J. *Statistical analysis of spatial point patterns*. London: Arnold, 2003. 153p.
- DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. *Model based geostatistics*. New York: Springer, 2007. 243p.
- ISAAKS, E. H.; SRIVASTAVA, R. M. *Applied geostatistics*. Oxford: Oxford University, 1989. 560p.
- KOENIG, W. D.; KNOPS, J. M. H. Testing for spatial autocorrelation in ecological studies. *Ecography*, Copenhagen, v. 21, n. 4, p. 423-429, 1998.
- LEGENDRE, P. Spatial autocorrelation: trouble or new paradigm? *Ecology*, Durham, v.74, n.6, p.1659-1673, 1993.
- MANLY, B. F. J. *Randomization and Monte Carlo methods in biology*. London: Chapman & Hall, 1994. 281p.

- 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: 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, 2006.
- RIBEIRO JÚNIOR, P. J.; DIGGLE, P. J. GeoR: a package for geostatistical analysis. *R-News*, New York, v.1, n.2, p.15-18, 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).
- THOMSON, J. D. et al. Untangling multiple factors in spatial distributions: lilies, gophers, and rocks. *Ecology*, Durham, v.77, n.6, p.1698-1715, 1996.
- UPTON, G. J. G.; FINGLETON, B. *Spatial data analysis by example: point pattern and quantitative data*. New York: J. Wiley, 1985. v.1, 410p.
- WALLER, L. A.; GOTWAY, C. A. *Applied spatial statistics for public health data*. New Jersey: J. Wiley, 2004. 519p.

Recebido em 24.11.2014

Aprovado após revisão em 26.04.2015