Análises espaciais

Valéria Forni Martins

Pesquisadora Colaboradora do Depto. Biologia Vegetal, UNICAMP

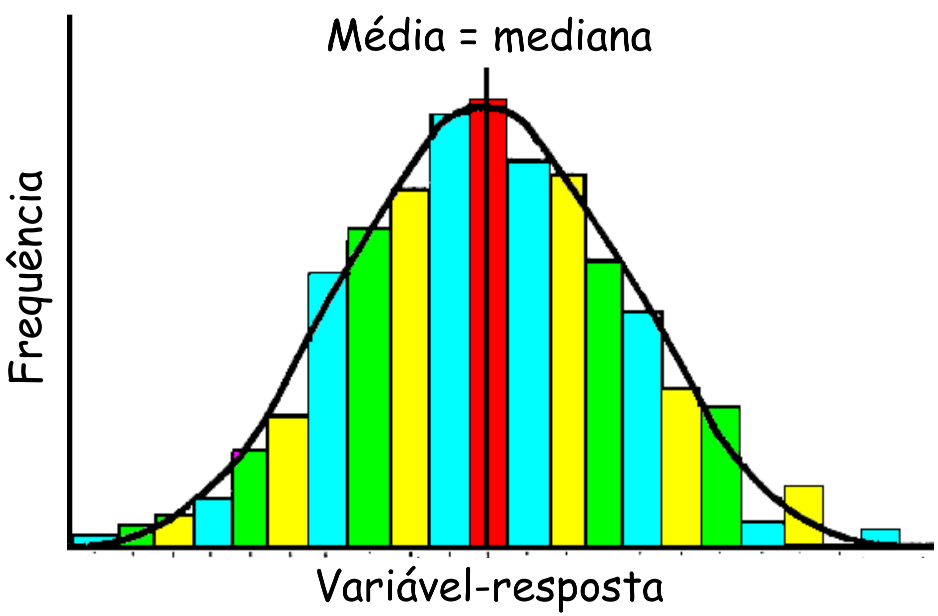
E-mail: valeriafm@gmail.com

Introdução às análises espaciais

- Por que análises espaciais e não estatística convencional (paramétrica e não-paramétrica)?
- O uso da estatística convencional tem requisitos:

- Amostras independentes: a medida da variável-resposta (parâmetro) de uma amostra não deve ser influenciada pela medida da variável-resposta de outra amostra.

- Dados com distribuição normal (estatística paramétrica).



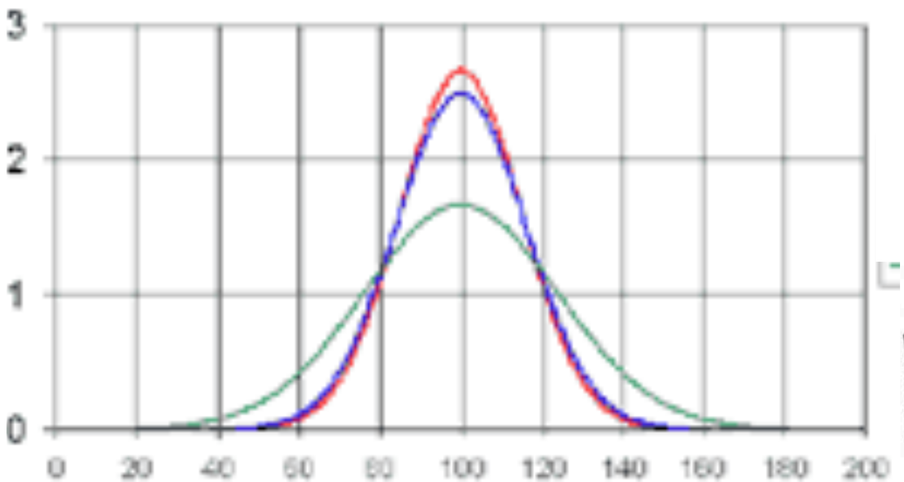
A média é onde mais se concentram os dados da distribução; é o ponto de equilíbrio das frequências no histograma. $\overline{x} = \frac{\sum \text{valor}}{n}$

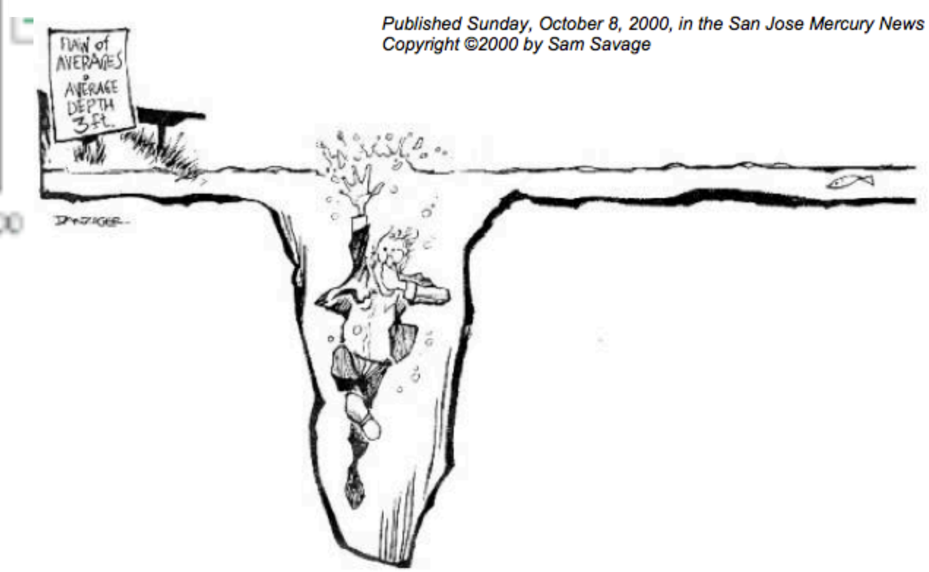
A mediana divide os valores da variável-resposta em duas partes iguais; é o valor central dos dados ordenados.

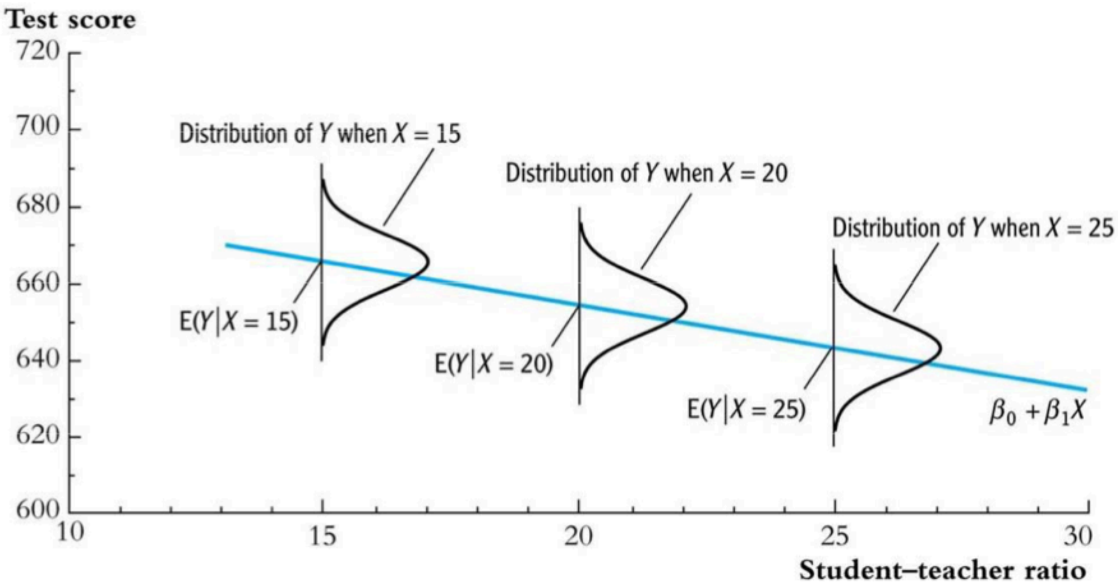
- Homocedasticidade dos dados: a variância dos dados deve ser semelhante para as diferentes amostras (estatística paramétrica).

• A variância é uma medida de dispersão da média.

$$\sigma^2 = \frac{\sum (\text{valor-média})^2}{n}$$

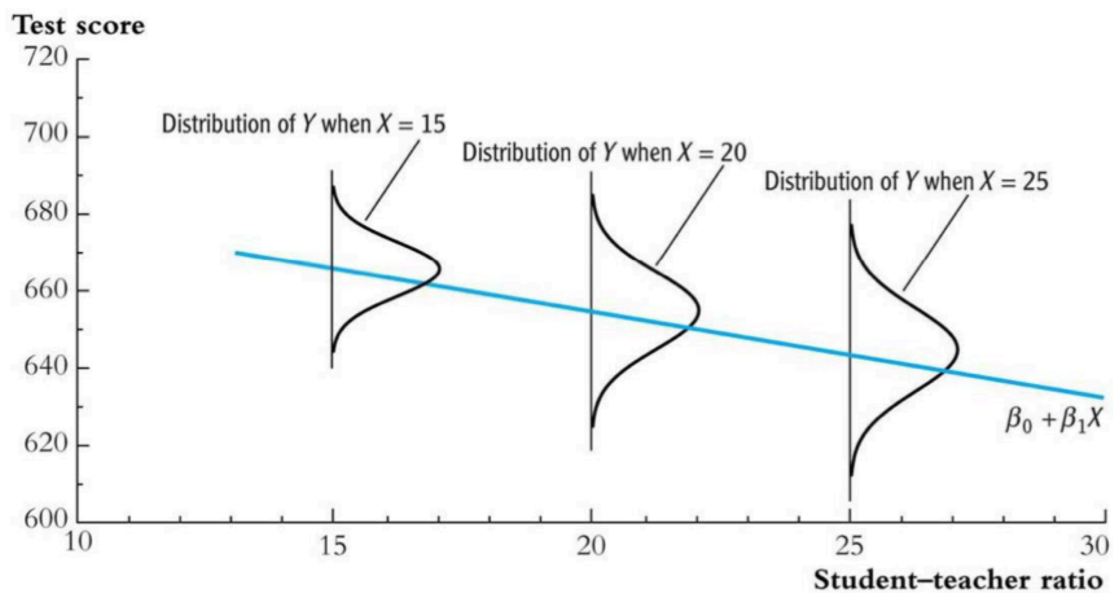






Homocedasticidade

Heterocedasticidade



- Coeficiente de dispersão: mede a variação da densidade de indivíduos no espaço. $CD = \frac{\sigma^2}{\bar{x}}$

  - Quando σ^2 é alta, CD é maior: população mais agregada conforme se aumenta CD.
  - Quando σ^2 é baixa, CD tende a 0: população com estrutura regular.
  - Quando $\sigma^2 = \bar{x}$, a distribuição do nº de indivíduos em cada unidade amostral é igual e CD = 1: população com estrutura aleatória.
  - CD vai de 0 a ∞: não se pode fazer comparações entre áreas, mas sim dizer se populações são mais ou menos agregadas.

- Índice de Morisita: mede a concentração de indivíduos no espaço, independente da densidade. $I_y = \frac{n\sum(x)(x-1)}{(N)(N-1)}$

n = nº de unidades amostrais
x= nº de indivíduos/unidade amostral
N = nº total de indivíduos no universo amostral

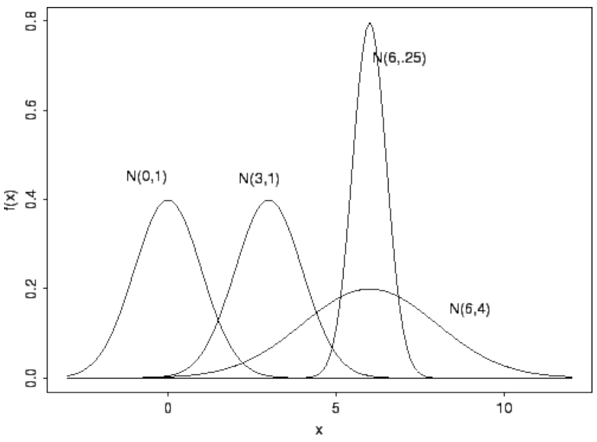
- I_y vai de 0 a 1: permite comparações entre áreas.
- I_y mais próximo de 1: população com estrutura agregada.
- I_y mais próximo de 0: população com estrutura regular.
- I_y intermediário: população com estrutura aleatória.

- CD e índice de Morisita requerem que as amostras sejam independentes: desuso.

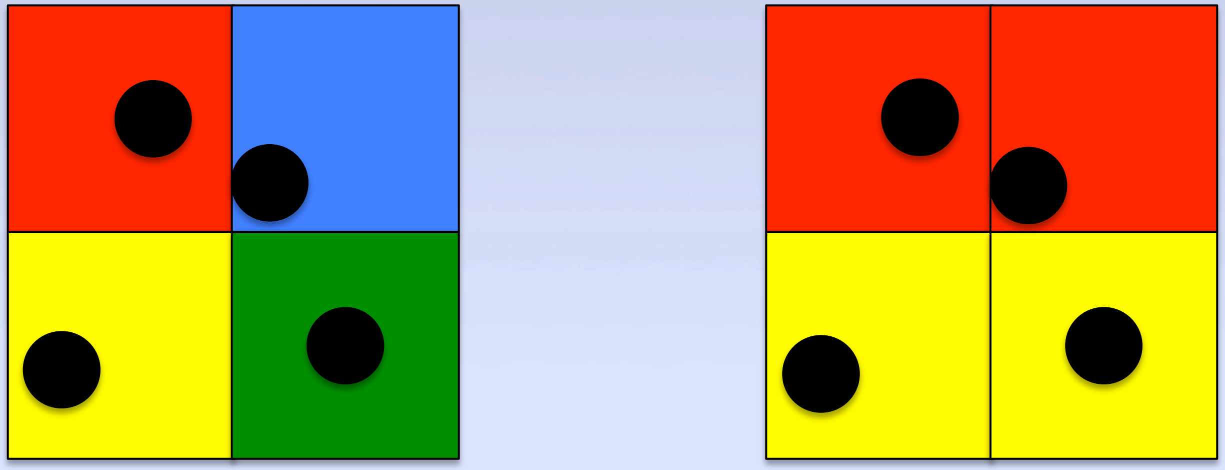
• O problema da autocorrelação espacial: amostras mais próximas no espaço são mais semelhantes (ou mais diferentes) do que o esperado; diz-se que a variável é autocorrelacionada, espacialmente estruturada ou regionalizada.

- Pseudorreplicação e não amostras independentes: não se cumpre o primeiro requisito da estatística convencional.

  - Replicação é a medida da variável-resposta em várias unidades amostrais; é importante para a estimativa da dispersão dos dados em torno de uma medida central (média ou mediana).



- Pseudorreplicação ocorre quando há mais de uma medida da variável-resposta na mesma unidade amostral.



- Infla os graus de liberdade.
    - Graus de liberdade (g.l. ou d.f. em inglês): tamanho da amostra menos o nº de parâmetros avaliados; estima o nº de categorias independentes em um teste estatístico.
- Teste dá resultado significativo não quando não deveria.

- Sistema de amostragem ao acaso ou sistematizado não resolve o problema da autocorrelação espacial: zona de influência do padrão pode ser maior do que a distância entre as unidades amostrais.

• Os dados espacialmente autocorrelacionados podem ser em contagens (quantos objetos em uma unidade amostral) ou em padrão de pontos (coordenadas x,y de todas as unidades amostrais presentes dentro de um universo amostral).

Dados em contagens

• Correlograma: gráfico de correlação entre o nº de indivíduos de uma unidade amostral e o nº de indivíduos da unidade amostral vizinha, unidade amostral mais longe, etc, em função da distância entre as unidades amostrais.

- Em quais classes de distância e até qual distância há autocorrelação espacial.

- Índice I de Moran é o mais utilizado.

- Valores ≥ 0,2 geralmente são significativos.
- Tamanho das classes de distância.
- Correlograma não distingue bem diferentes distribuições.
- Não identifica tamanho dos agrupamentos
- Descreve apenas um padrão.

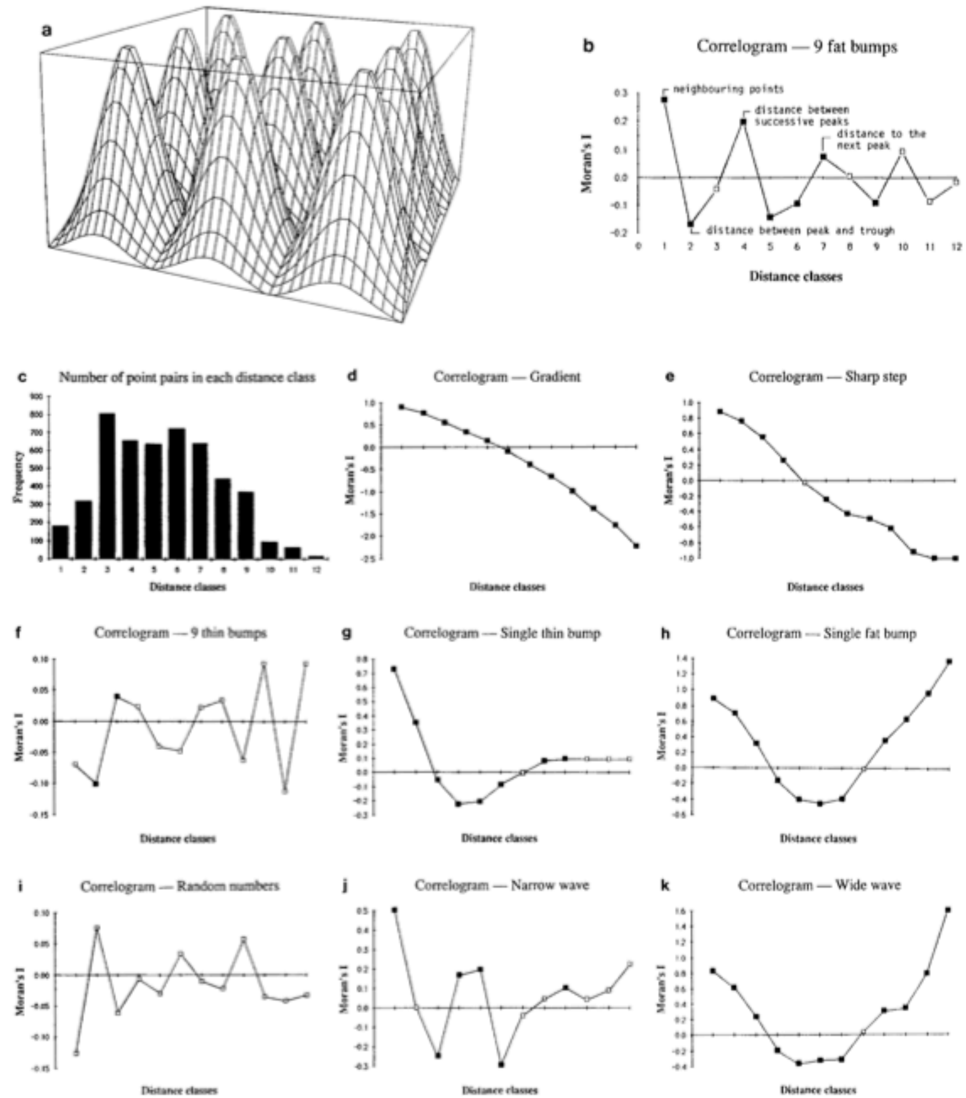


Fig. 1. All-directional spatial correlograms of artificial structures (see text). (a) depicts the structure analysed by the correlogram in (b). (c) displays the number of distances (between pairs of points) in each distance class, for all the correlograms in this figure. In the correlograms (b, d–k), black squares represent significant values at the $\alpha = 5\%$ level, before applying the Bonferroni correction to test the overall significance of the correlograms; white squares are non-significant values.

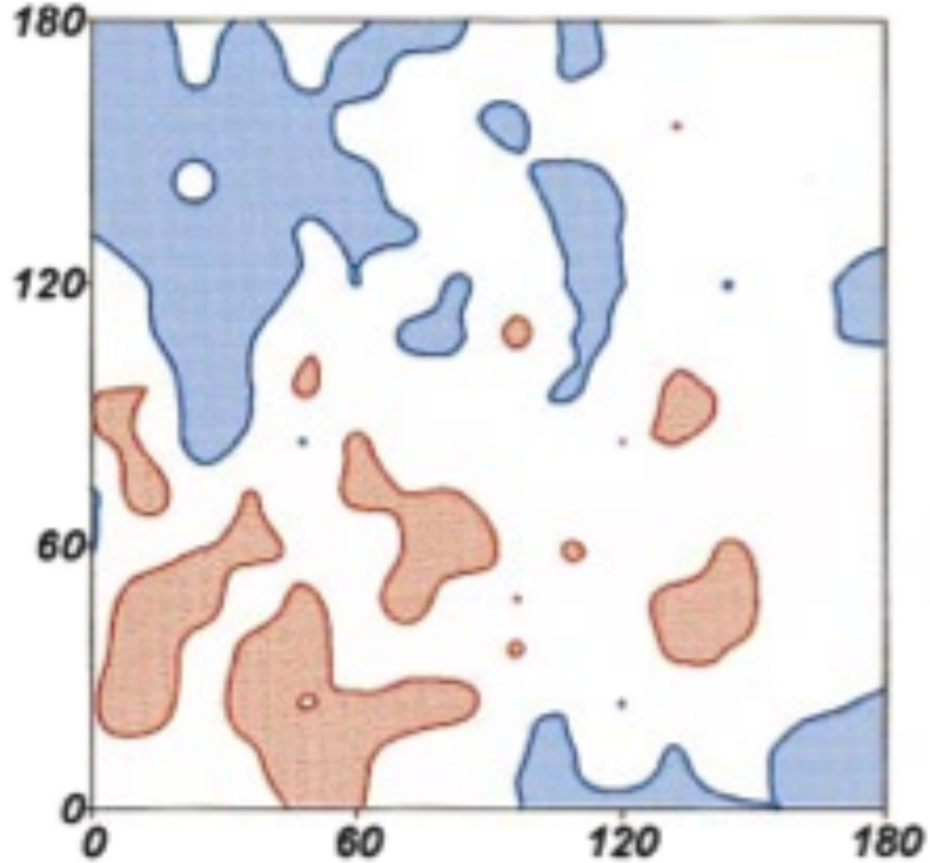
Legendre & Fortín 1989

- SADIE (Spatial Analysis by Distance IndicEs):

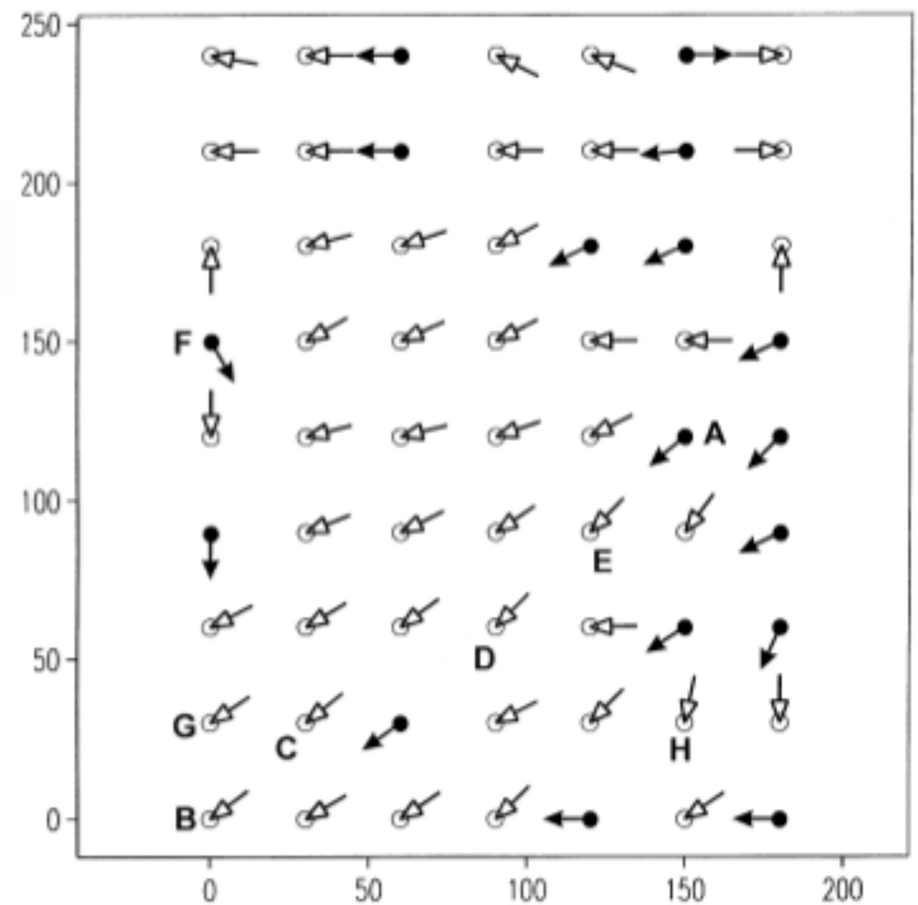
  - Mais adequada porque as contagens de densidade de indivíduos (valores inteiros) não são contínuas, mas sim discretas; geralmente formam manchas com grandes quantidades de indivíduos e outras com nenhum indivíduo; são extremamente dinâmicas e geralmente caracterizadas por grupos isolados, os quais podem atuar como metapopulações com variados graus de dispersão entre manchas.

  - Identifica agregações com contagens altas que estão próximas (*patch*), ou com contagens baixas (*gap*).

  - Esforço mínimo dos indivíduos da população para se moverem até uma distribuição completamente regular (= abundância em todas as unidades amostrais): D (distância mínima).



Perry *et al.* 1997



Winder *et al.* 2001

- A estrutura espacial é quantificada pela permutação das contagens entre as unidades amostrais: fornece a hipótese nula de contagens distribuídas aleatoriamente ao mesmo tempo em que mantém precisamente as propriedades numéricas dos dados.

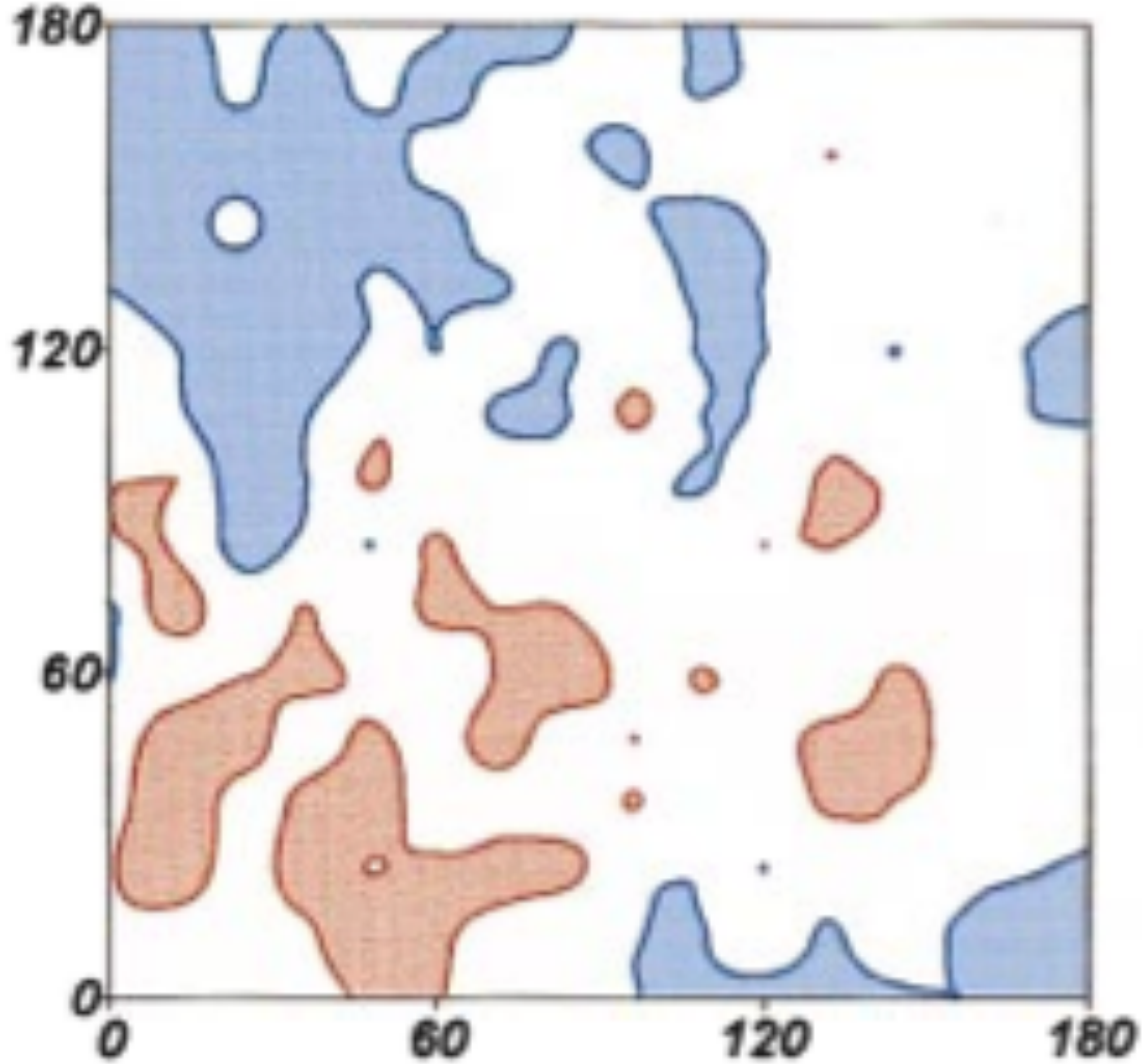
  - H_0 traduz a ausência do efeito que se quer verificar.
  - H_1 (hipótese alternativa) traduz o que se quer investigar.

- D/valor médio de centenas de aleatorizações = I_a (índice de agregação); I_a = 1 indica contagens aleatórias; I_a > 1 indica contagens agregadas (de *patches* ou *gaps*).

- No entanto, I_a não indica se a agregação decorre de *patches* ou de *gaps*:

- v_i: índice de agregação que mede o grau em que cada unidade amostral contribui para a agregação como parte de um grupo de unidades doadoras de um *patch*; valores > 1,5 indicam que a unidade amostral pertence a um *patch*; valores próximos de 1 indicam distribuição aleatória de uma unidade amostral em relação às suas vizinhas.
- v_j: índice correspondente para unidades que são parte de um grupo recebedor de um *gap*; tem valores negativos por convenção; valores < -1,5 indicam que a unidade amostral pertence a um *gap*; valores próximos de -1 indicam distribuição aleatória de uma unidade amostral em relação às suas vizinhas.
- v_i e v_j são estandardizados e adimensionais: permitem comparações.

- Ao contrário do índice I de Moran, identifica *patches* e *gaps*.



• Para se comparar a distribuição de duas variáveis no espaço, pode-se utilizar a SADIE.

- Compara os índices de agregação dos dois conjuntos de dados ao invés de correlacionar as contagens das duas variáveis: mantém intrinsecamente o padrão espacial de cada variável.

- Retira deliberadamente o peso de contagens muito altas ou muito baixas que ocorrem isoladas no espaço e com índices de agregação próximos de 1 (aleatório): X (medida de associação local) não é afetada por este ruído, aumentando o poder do teste em detectar associações significativas.

- X varia de +1 (associação espacial completa) a -1 (dissociação espacial completa), com 0 indicando independência espacial.

- Associações são testadas por permutações: não há uma distribuição prevista e os resultados podem ser comparados entre populações e desenhos amostrais com diferentes estruturas espaciais.

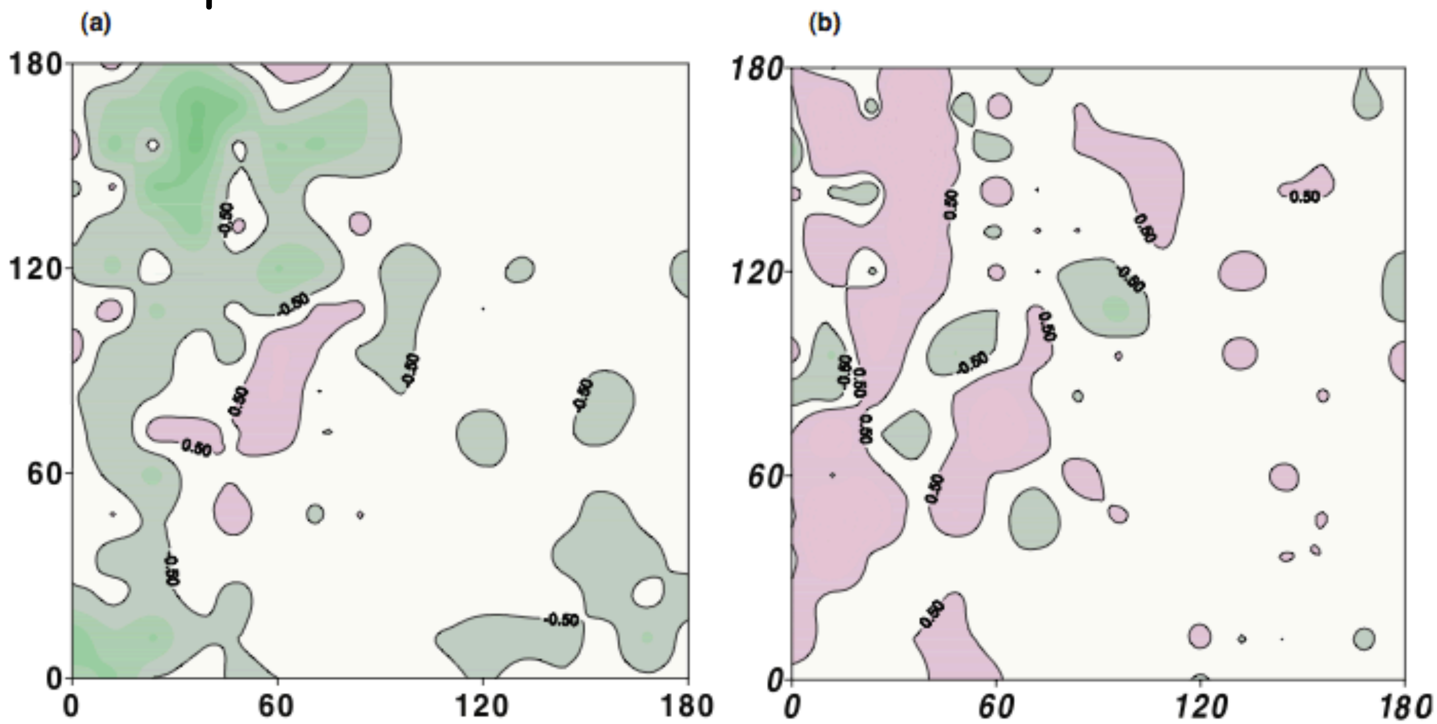
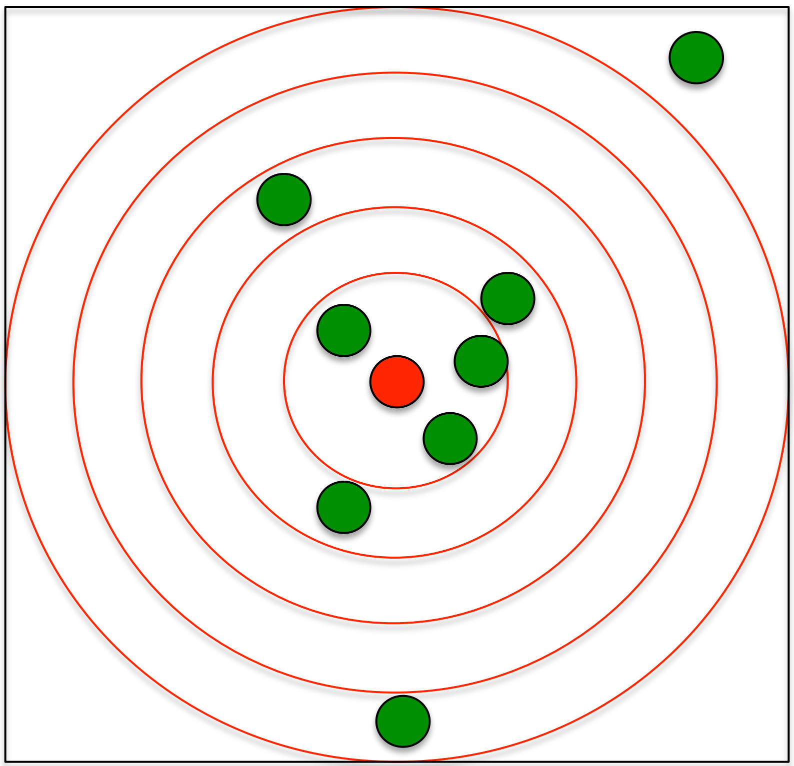


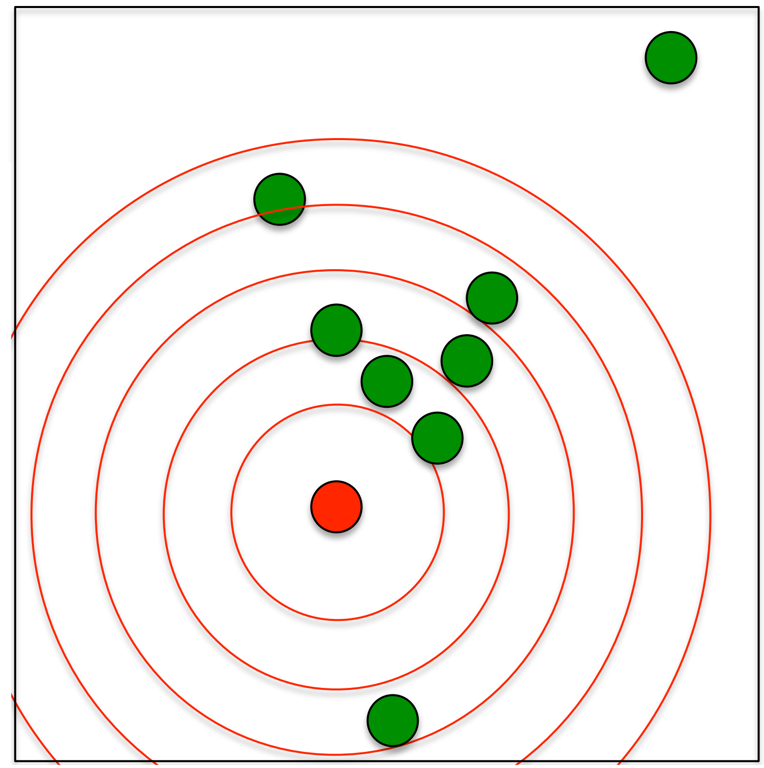
Figure 2 Spatio-temporal association between predator and prey. Association and dissociation are represented in violet and green, respectively. (a) Contour map of local association between *P. melanarius* on 7 June (occasion 2) and *M. dirhodum* on 21 June (occasion 3) shows overall negative dissociation (X = –0.41, $P < 0.001$) with considerable spatial structure. Axes indicate coordinates in metres. (b) As (a) but for both *P. melanarius* and *M. dirhodum* on 21 June (occasion 3), showing overall positive association (X = 0.29, $P < 0.001$) with considerable spatial structure in similar areas.

Winder *et al.* 2001

Dados em padrão de pontos

• Função K de Ripley: conta quantos indivíduos há em círculos em torno de uma planta focal; os círculos começam com um raio pequeno e vão até um raio que inclui toda a área de estudo; faz-se uma média do nº de indivíduos nas classes de distâncias em torno de todas as plantas focais da população.





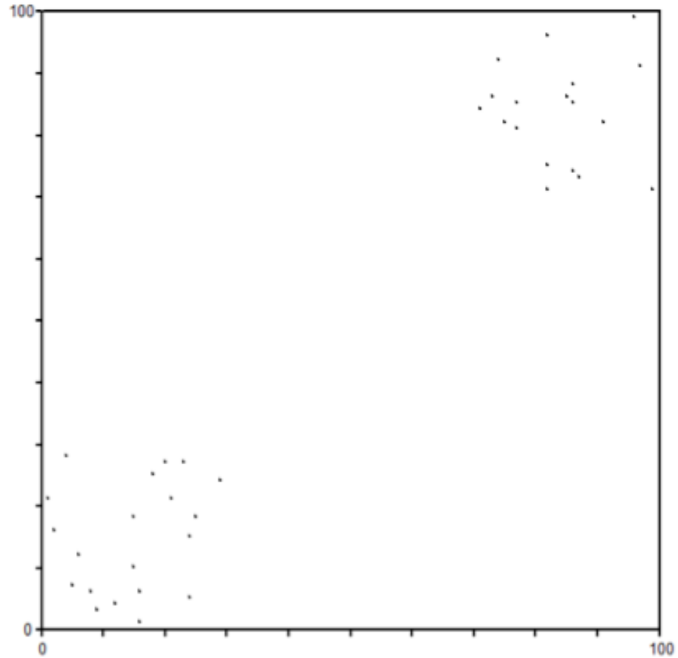
- A distribuição é cumulativa e representa o nº esperado de vizinhos em um círculo de raio t centrado em uma planta arbitrária dividido pela intensidade λ do padrão dos pontos na área de estudo.

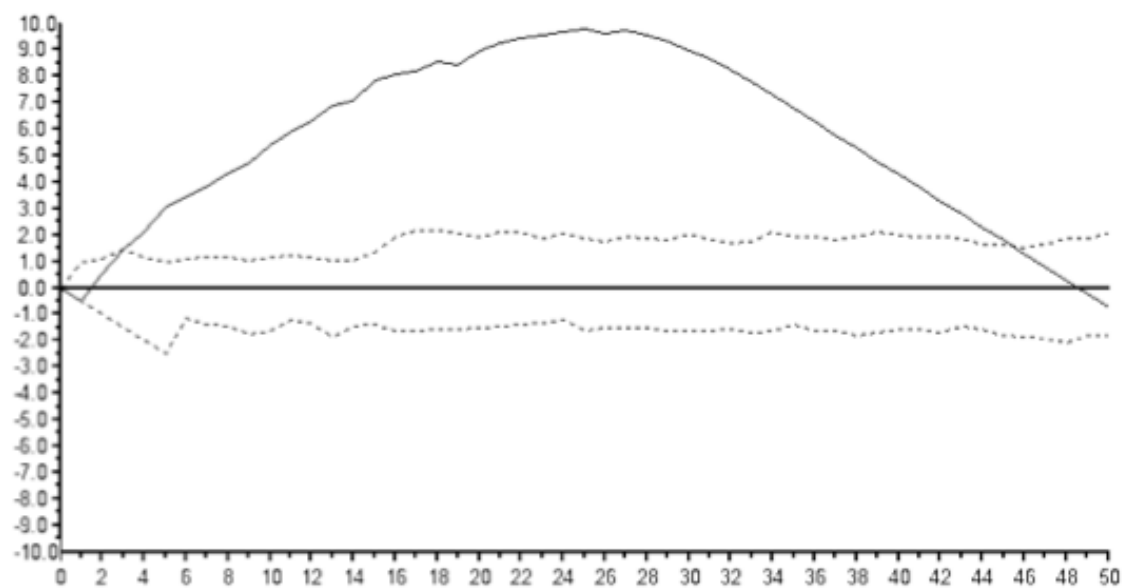
- Descreve a característica dos pontos em uma amplitude de escalas de distância: detecta padrões mistos.

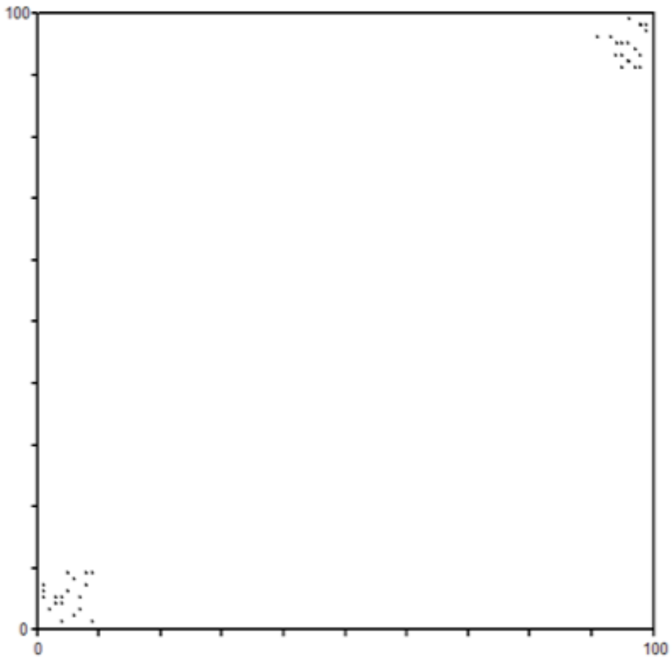
- Por ter uma distribuição cumulativa, não distingue claramente os efeitos a maiores distâncias dos efeitos a distâncias mais curtas a partir do ponto focal.

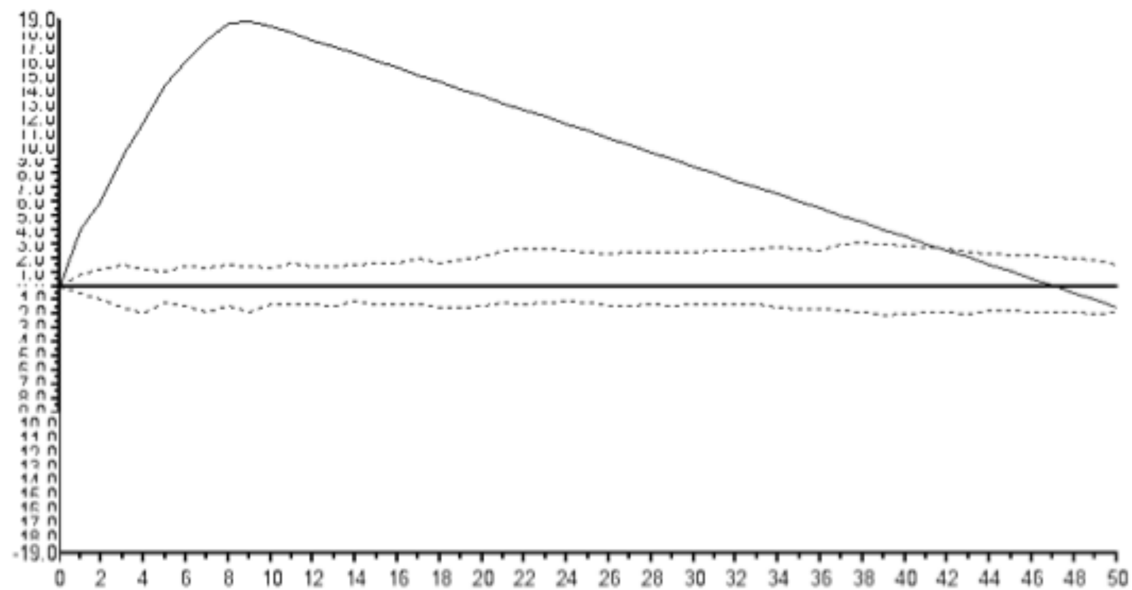
- São feitas simulações de Monte Carlo utilizando-se o modelo nulo de distribuição aleatória dos pontos (*CSR*).

- Envelope de confiança: intervalo de estimativas prováveis de um parâmetro; para α = 5%, envelope de confiança = 95%, ou seja, 95% das estimativas do parâmetro estão dentro do envelope, 2,5% das estimativas estão abaixo do envelope (significativo) e 2,5% das estimativas estão acima do envelope (significativo).

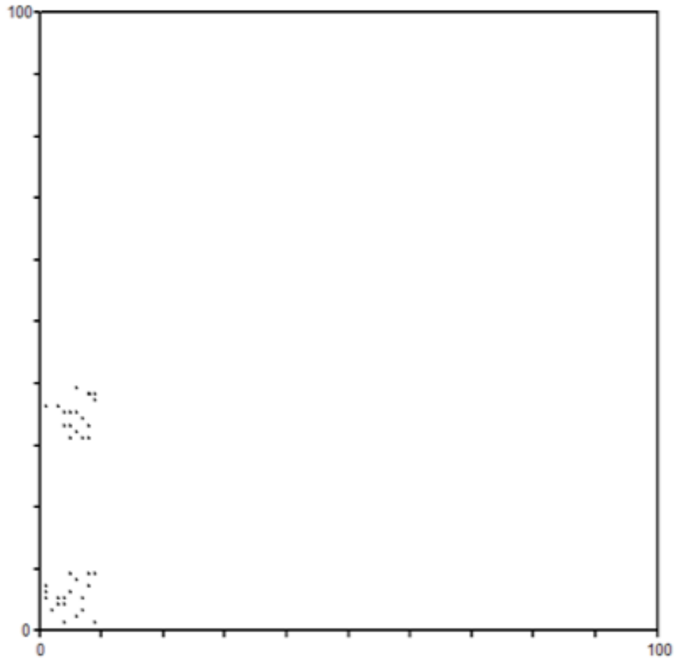

100
0
0
100

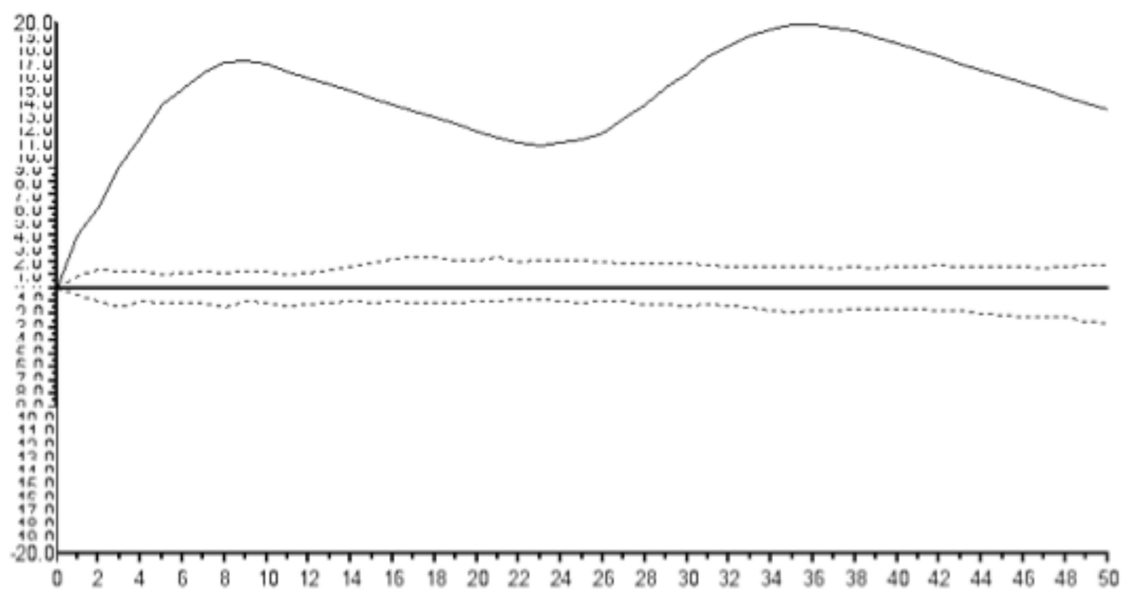

10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
-7.0
-8.0
-9.0
-10.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50


100
0
0
100


19.0
-19.0
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50

Análises espaciais de dados em padrão de pontos.: função K de Ripley





- *Pair-correlation function* (*g*) :

  - Substitui os círculos da função K por anéis: nº esperado de pontos em uma distância r a partir de um ponto arbitrário dividido pela intensidade λ do padrão dos pontos na área de estudo.

  - Função não-cumulativa como a função K: isola classes de distâncias específicas e detecta padrões mistos sem confundir efeitos a diferentes distâncias do ponto focal

  - É uma função de probabilidade de densidade, podendo ser interpretada como densidade de vizinhança: mais intuitivo do que uma função cumulativa.

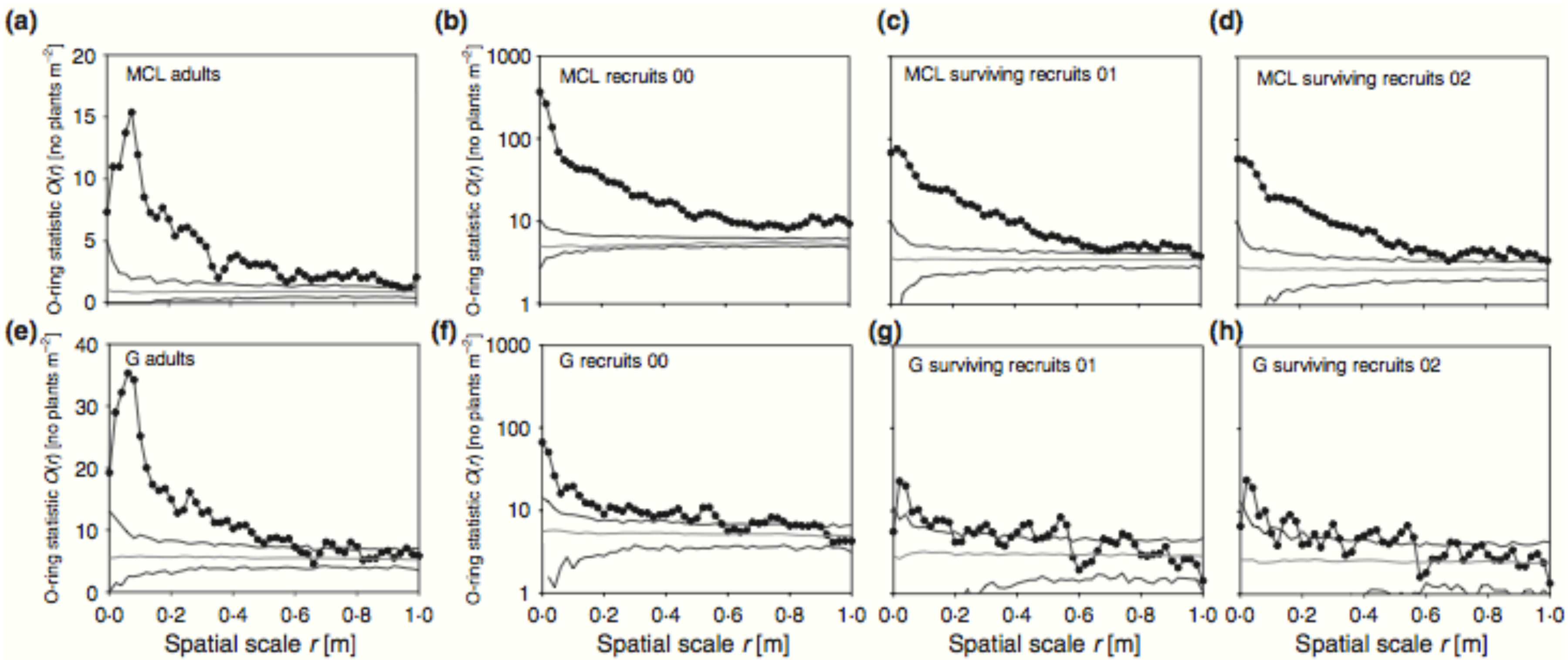


Fig. 1. Analysis of the univariate spatial patterns of adults and recruits (00: recruits present in 2000, 01: surviving recruits in 2001; 02: surviving recruits in 2002) under the two contrasting management treatments, grazing (G) and mowing plus clearing (MCL). The data were contrasted to the null model of complete spatial randomness. The O-ring statistic $O(r)$ giving the density of plants at distance r away from a typical plant of the pattern (dots), simulation envelopes (black solid line) being the 5th lowest and highest values of the O-ring statistic taken from the 199 simulations of null model, and the average O-ring statistic under the null model (grey solid line). Note the logarithmic scale of the y-axis for the recruit patterns (b–d and f–h). The patterns are aggregated if the O-ring statistic is above the simulation envelopes.

Jacquemyn *et al.* 2010

- Para se comparar a distribuição de duas variáveis no espaço, pode-se utilizar a função K de Ripley bivariada:

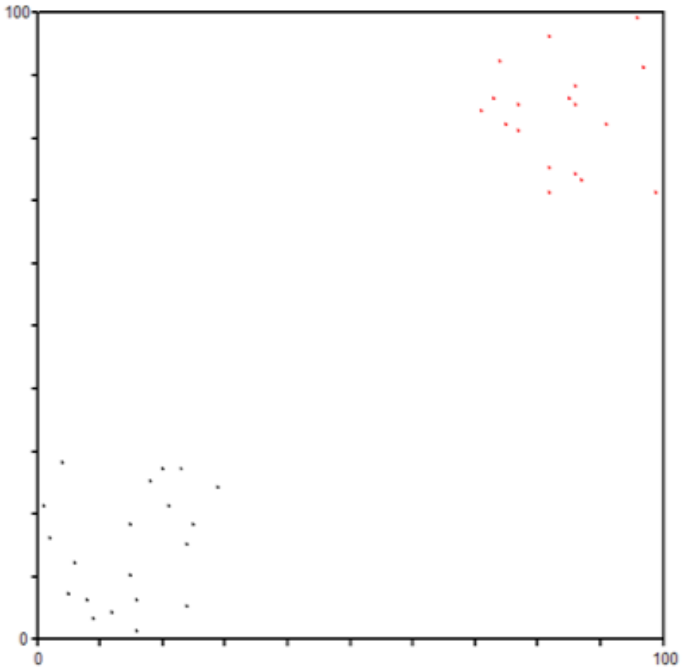
  - Calcula a média esperada do nº de indivíduos de um tipo dentro das classes de distância de um indivíduo arbitrário de outro tipo.

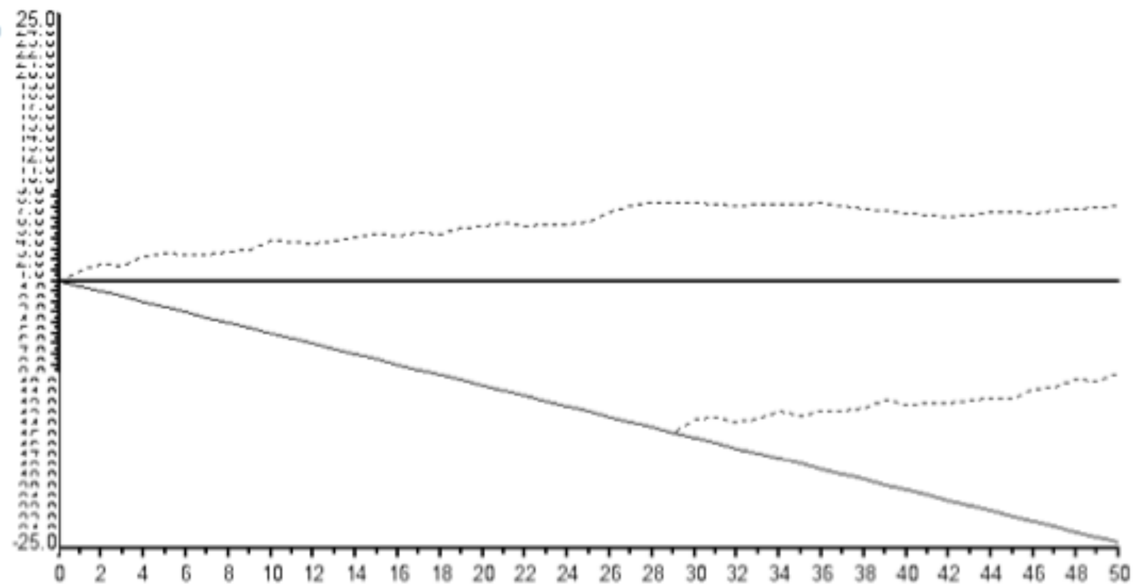
  - Atração entre as variáveis: valores estão acima do envelope de confiança.

  - Repulsão entre as variáveis: valores estão abaixo do envelope de confiança.

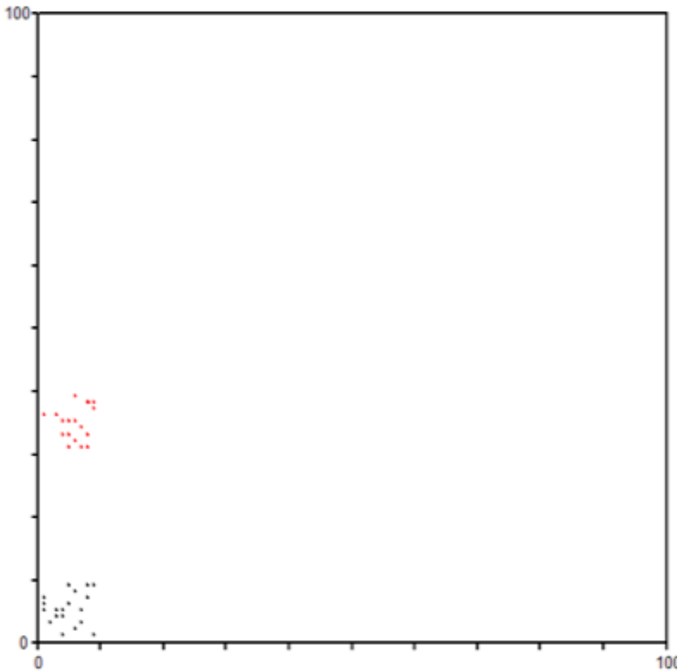
  - Independência entre as variáveis: valores estão dentro do envelope de confiança.

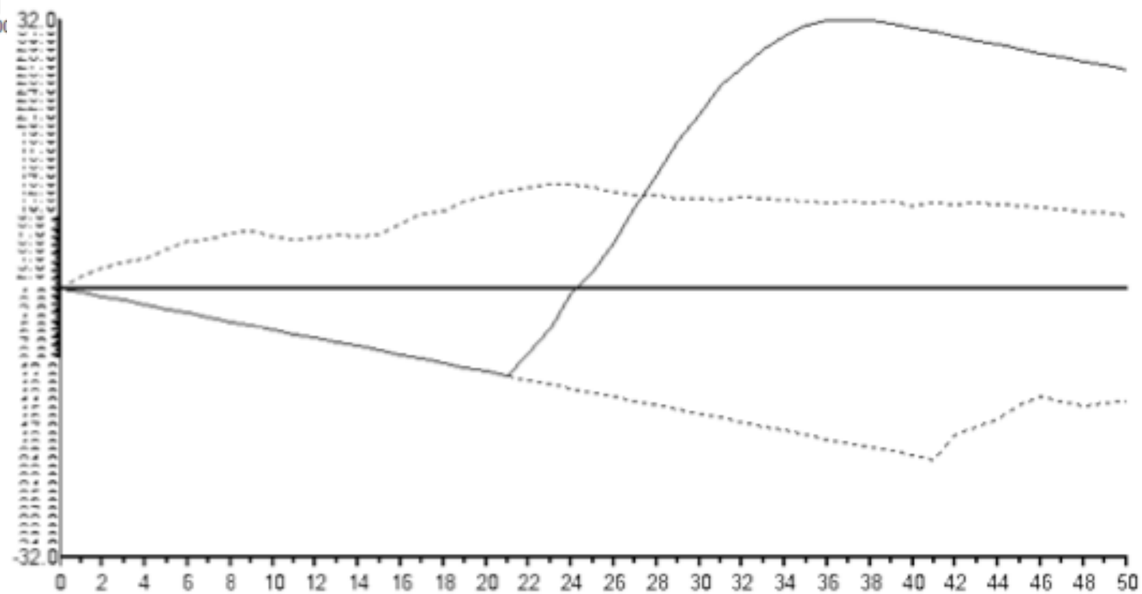
Análises espaciais de dados em padrão de pontos.: função K de Ripley bivariada





Análises espaciais de dados em padrão de pontos.: função K de Ripley bivariada





• Para se comparar a distribuição de duas variáveis no espaço, também se pode utilizar a *pair-correlation function* bivariada.

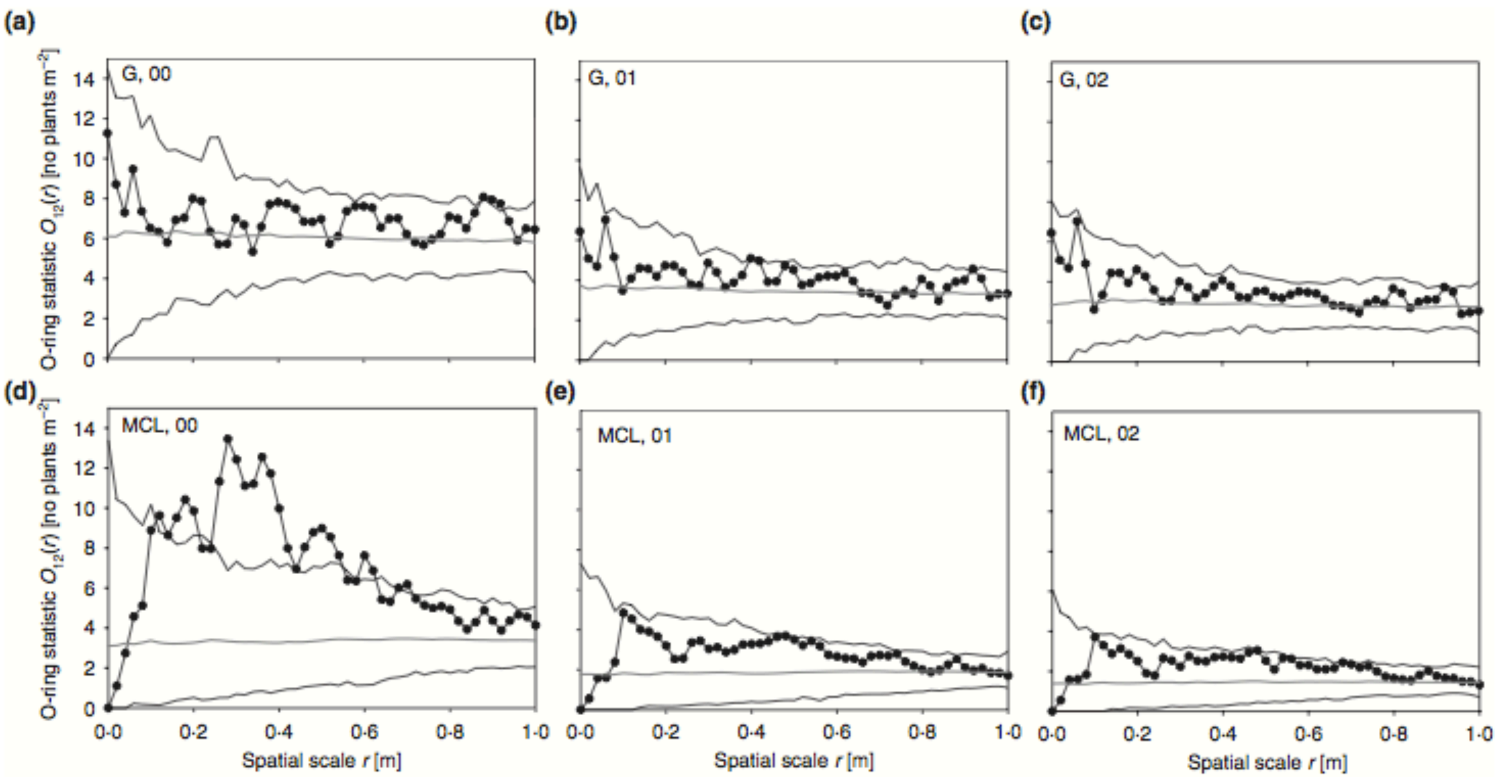


Fig. 2. Analysis of the recruit-adults relationship. Test of independence between the pattern of recruits in years 2000, 2001, and 2002 and the adult pattern under the two contrasting management treatments grazing (G) and mowing plus clearing (MCL). The data were compared to toroidal shift as null model. The bivariate O-ring statistic $O_{12}(r)$ giving the density of recruits at distance r away from a typical adults plant (dots), simulation envelopes (black solid line) being the 5th lowest and highest values of the $O_{12}(r)$ taken from the 199 simulations of null model, and the average $O_{12}(r)$ under the null model (grey solid line). Recruits are significantly aggregated around adults if the $O_{12}(r)$ is above the simulation envelopes.

Jacquemyn *et al.* 2010