

Lucas da Cunha Godoy

Testes de Associação Espacial Entre Dois Tipos de Polígonos.

Belo Horizonte

2019, Abril

Lucas da Cunha Godoy

Testes de Associação Espacial Entre Dois Tipos de Polígonos.

Universidade Federal de Minas Gerais – UFMG

Instituto de Ciências Exatas

Programa de Pós-Graduação

Orientador: Renato Martins Assunção

Belo Horizonte

2019, Abril

Resumo

É comum em estatística espacial o interesse em testar hipóteses relacionadas à associação espacial entre dois padrões espaciais. Muitas vezes, as entidades espaciais que melhor representam os dados são polígonos. Contudo a complexidade e falta de técnicas desenvolvidas neste campo força pesquisadores a simplificarem tais polígonos por pontos, como os seus centróides. Neste trabalho, foram desenvolvidas técnicas para testar hipóteses relacionadas a padrões de polígonos. Estas técnicas tiram vantagem de toda a informação presente nas entidades espaciais de interesse. A metodologia aqui desenvolvida tem como ponto central o *toroidal shift*. A utilização deste método traz consigo desafios que fizeram com que outros autores o desconsiderassem, mesmo sendo ele uma melhor alternativa de modelo nulo. Também foi realizado um estudo de simulação para comprovar a eficácia da metodologia desenvolvida e comparar diferentes estatísticas de teste. Por fim, as técnicas foram aplicadas a dados reais referentes a diferentes espécies de gramas curtas no estado do Colorado.

Palavras-chaves: Padrões de Polígonos Espaciais, Associação Espacial, Testes de Monte Carlo Condicionais, Toroidal Shift.

Abstract

It is common in spatial statistics the interest in testing hypotheses related to the spatial association between two spatial patterns. Often the spatial entities that best represent the data are polygons. However, the complexity and lack of techniques developed in this field forces researchers to simplify such polygons by points, as their centroids. In this work, techniques were developed to test hypotheses related to polygons patterns. These techniques take advantage of all the information present in these spatial entities. The methodology developed here is centered on the *toroidal shift*. The use of this method brings challenges that caused other authors to disregard it, even though it is a better null model alternative. A simulation study was also carried out to prove the efficacy of the developed methodology and to compare different test statistics. Finally, the techniques were applied to real data regarding different species of short grams in the state of Colorado.

Key-words: Spatial Polygons Patterns, Spatial Association, Conditional Monte Carlo Tests, Toroidal Shift.

Lista de ilustrações

Figura 1 – <i>Bouteloua gracilis</i> (em cinza) e <i>Buchloe dactyloides</i> em diferentes anos em Nunn, Colorado, USA.	10
Figura 2 – a) Dados simulados de maneira independente. Os centróides do padrão de polígonos brancos seguem um processo de Matérn, enquanto os centróides do padrão cinza foram simulados de acordo com um processo Poisson homogêneo; b) Envelope da função K cruzada adaptada para polígonos; c) Envelope da função K cruzada onde os polígonos foram simplificados pelos seus centróides.	12
Figura 3 – Exemplo de uma região retangular envolvida em um torus.	17
Figura 4 – Região \mathcal{D}^* , composta por \mathcal{D} e suas cópias.	18
Figura 5 – a) Translação aleatória; b) Dados Observados.	19
Figura 6 – a) : Dados simulados; b) : Correção toroidal aplicada para calcular as estatísticas referentes aos polígonos brancos em relação aos cinzas; e c) : Correção para mensurar estatísticas dos polígonos cinza com respeito aos brancos.	26
Figura 7 – (a) : Area de guarda utilizada para computar distâncias (ou interseção de area) do padrão de polígonos cinzas em relação ao branco; (b) : Area de guarda utilizada para computar distâncias (ou interseção de area) do padrão de polígonos brancos em relação aos polígonos cinzas	27
Figura 8 – Representação gráfica do fator de ajuste.	28
Figura 9 – Exemplos de padrões de polígonos simulados com relação de repulsão.	30
Figura 10 – Exemplos de dados simulados com relação de atração e diferentes estruturas marginais.	31
Figura 11 – Erro do tipo I estimado para diferentes funções e distâncias.	33
Figura 12 – Painel que ilustra o poder do teste no cenário de atração associado à diferentes estatísticas de teste utilizando diferentes distâncias e funções para descrever a segunda ordem dos processos.	35
Figura 13 – Painel que ilustra o poder do teste no cenário de repulsão associado à diferentes estatísticas de teste utilizando diferentes distâncias e funções para descrever a segunda ordem dos processos.	36
Figura 14 – Comparando métodos de simulação de dados sob a hipótese nula no cenário de atração.	37
Figura 15 – Comparando métodos de simulação de dados sob a hipótese nula no cenário de repulsão.	38
Figura 16 – Exemplos de tipos de gramas observadas nos dados de Chu et al. (2013).	39

Figura 17 – Localização dos quadrats de acordo com o tipo de pasto e ocorrência de pastagem.	40
Figura 18 – a) : Histograma das áreas das plantas; b) : Histograma do número de polígonos das plantas.	42
Figura 19 – Histograma dos p valores calculados de acordo com os Testes de Monte Carlo.	43

Lista de tabelas

Tabela 1 – Cenários simulados para estimar o poder do teste.	32
Tabela 2 – Número de testes realizados por espécie de planta.	41
Tabela 3 – Número de testes realizados por tipo de pastagem; Frequências absolutas e relativas.	41
Tabela 4 – Número de testes realizados por tipo de pasto.	42
Tabela 5 – Número de testes realizados por ano; Frequências absolutas e relativas.	42
Tabela 6 – Tabela da Anova para os efeitos fixos do modelo.	44
Tabela 7 – Comparações múltiplas entre as espécies segundo o Teste de Tukey, com nível de significância de 5%.	44
Tabela 8 – Comparações múltiplas entre os tipos de ocorrência de pastagem segundo o Teste de Tukey, com nível de significância de 5%.	45

Sumário

1	INTRODUÇÃO	9
2	TRABALHOS RELACIONADOS	14
3	UM TESTE MONTE CARLO CONDICIONAL	16
3.1	Hipóteses	16
3.2	O Toroidal Shift	16
3.3	O que vem sendo feito na prática	17
3.4	A distância de Hausdorff e uma pequena adaptação do Toroidal Shift	19
3.5	Equívocos envolvendo a adaptação do Toroidal Shift para polígonos	20
3.6	Propriedades de Segunda Ordem de Processos Espaciais Marcados	21
3.7	Estatísticas de Teste	23
3.7.1	Envelopes ponto-a-ponto	23
3.7.2	Goodness of Fit ou Deviation Tests	24
3.8	Correções de Borda	25
3.8.1	Correção Toroidal	26
3.8.2	Área de Guarda	27
3.8.3	Fator de Ajuste	27
4	ESTUDO DE SIMULAÇÃO	29
4.1	Simulando Padrões de Polígonos	29
4.1.1	Padrões com Relação Espacial de Repulsão	29
4.1.2	Padrões com Relação Espacial de Atração	30
4.1.3	Padrões com Relação Espacial de Independência	31
4.2	Poder do Teste	31
5	RESULTADOS	33
5.1	Erro do Tipo I	33
5.2	Análise do Poder	34
5.2.1	Estatísticas de Teste	34
5.2.2	Modelos Nulos	37
6	ESTUDO DE CASO	39
6.1	Modelo Misto	43
7	CONCLUSÕES	46

Referências 47

1 Introdução

Manchas ecológicas possuem um importante papel em ecologia de paisagens. Elas podem representar tanto conjuntos de arbustos no deserto, quanto clarões em florestas causados pelo desmatamento ou até recifes no oceano. Todos estes exemplos de ecossistemas que podem ser representados por manchas ecológicas têm em comum o fato de serem homogêneos, terem limites bem definidos e diferirem significativamente do ambiente ao seu redor, conforme [McGarigal \(2014\)](#). Também vale ressaltar que, dependendo da escala na qual estas manchas estão sendo estudadas, elas podem ser grandes, pequenas e apresentar diversas variações em relação a sua forma. Dado que elas podem possuir diferentes formas e configurações espaciais, as mesmas podem ser representadas matematicamente por diferentes polígonos localizados na região de estudo [Cullen et al. \(2001\)](#); [A. Jansen et al. \(2008\)](#). Outros exemplos em ecologia nos quais polígonos são objetos adequados para representar as entidades espaciais associadas aos dados são os *home ranges*, ou áreas de vida de diferentes animais [Worton \(1987\)](#); [Schofield et al. \(2010\)](#).

Na Figura 1 são apresentados exemplos de manchas ecológicas referentes a duas diferentes espécies de gramas comuns na América do Norte, a *Bouteloa gracilis* e *Buchloe dactyloides*. A imagem traz dados referentes a estas duas espécies coletados ao longo de diferentes anos. Os dados estão localizados em *quadrats* de 1m². *Quadrats* são quadrados contíguos utilizados como unidades amostrais em experimentos na área de ecologia ([Fortin et al. \(2014\)](#)). As duas espécies alcançam até 30 centímetros de altura sendo que ambas são plantas de estação quente. Estes dados foram coletados em um experimento de [Guo et al. \(2013\)](#) localizado no Central Plains Experimental Range, em Nunn no estado do Colorado nos Estados Unidos. O experimento durou 14 anos, indo de 1997 até 2010.

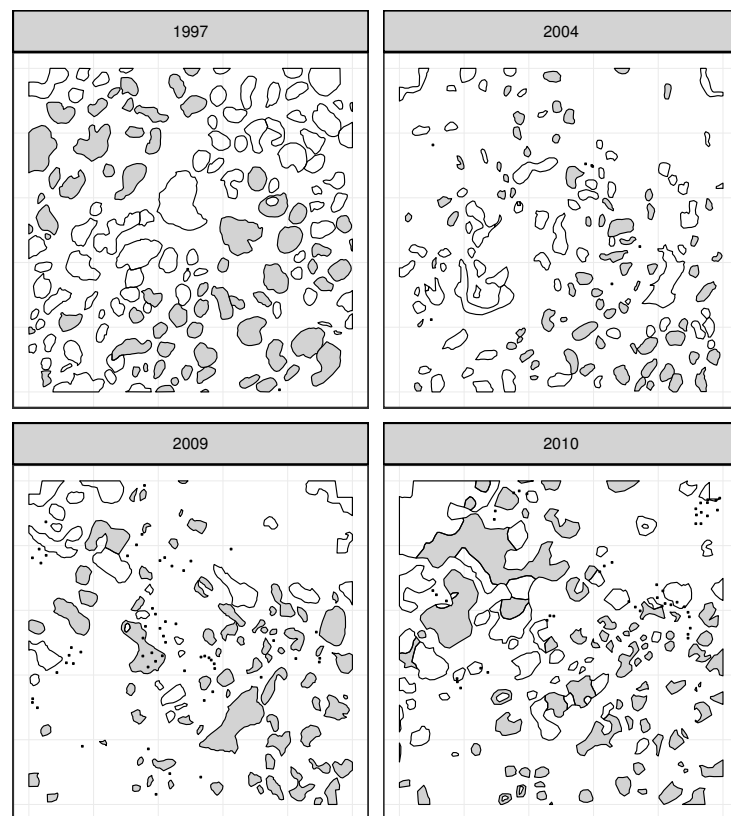


Figura 1 – *Bouteloua gracilis* (em cinza) e *Buchloe dactyloides* em diferentes anos em Nunn, Colorado, USA.

Este trabalho tem como principal objetivo fornecer uma ferramenta estatística capaz de responder uma questão comum em estatística espacial mas que ainda possui poucos trabalhos quando se trabalha com dois padrões de polígonos. Esta questão é testar se dois padrões espaciais são independentes ou apresentam algum tipo de associação. A evidência de interação espacial entre dois padrões levará analistas a formular novas hipóteses sobre a influência na ocorrência de padrões do tipo 1 na ocorrência, ou ausência, de padrões do tipo 2. Duas hipóteses alternativas que são candidatas naturais às hipóteses de independências entre padrões são as hipóteses de atração e repulsão. Estas hipóteses indicam que a ocorrência de um tipo de polígonos em uma certa região faz com que a probabilidade de ocorrência de eventos do outro tipo cresça ou decresça, respectivamente. Eventualmente, relações diferentes de independência entre padrões de polígonos podem estar associadas a outros fatores que não sejam a interação entre estes objetos. Por exemplo, no caso das gramas apresentado na Figura 1, se estes dados não pertencessem à um estudo controlado, e fosse encontrada uma relação de atração entre as espécies, esta relação poderia estar associada ao fato de que ambas necessitam do calor para se desenvolver e, em uma área onde os dados foram coletados, pode ocorrer uma menor incidência de luz solar devida à presença de uma árvore. Com isso, as plantas tenderiam a se aglomerar em uma mesma região não pelo fato de que existe uma relação de atração entre os dois padrões de

polígonos, mas por causa de um fator externo que esteja influenciando o desfecho.

Não se trata de um evento raro pesquisadores interessados neste tipo de problemas adotarem a solução de simplificar a complexidade espacial dos dados com a redução de cada mancha poligonal à um ponto, tal como o seu centroide. Este tipo de simplificação costuma ser justificado pelo vasto número de estudos aplicados que utilizam técnicas de processos pontuais para testar hipóteses praticamente equivalentes, mas onde as unidades amostrais em questão são bem representadas por pontos. Como exemplos da literatura os trabalhos [Prentice and Werger \(1985\)](#), [Leemans \(1991\)](#), [Getzin et al. \(2006\)](#), e [Hao et al. \(2017\)](#), reduziram manchas a pontos nos seus mapas. Este tipo de abordagem só é adequada e justificável quando os polígonos são tão pequenos que uma representação dos mesmos por pontos não perderia nenhuma informação relevante. Contudo, a abordagem em questão é incomum na maioria das aplicações. Principalmente naquelas quais os polígonos que representam os padrões apresentam variações de tamanho e forma, ou quando os pesquisadores estão interessados em efeitos de pequena escala. Na Figura 2 são apresentados dois padrões de polígonos simulados (rótulo **a**), bem como dois envelopes de funções K cruzadas associadas a estes padrões. Estes envelopes são utilizados para fins de análise descritiva, sua interpretação é baseada na relação entre o padrão da linha preta sólida e das linhas cinzas pontilhadas. Em geral, quando a linha preta está acima da linha cinza superior, assume-se que existem evidências que os dois processos possuem uma relação de atração. Enquanto no caso em que esta linha preta aparece abaixo da linha cinza inferior, as evidências amostrais indicam uma possível relação de repulsão entre os padrões. O primeiro dos dois envelopes, rotulado pela letra **b** foi construído utilizando técnicas que levam em conta a forma e o tamanho dos polígonos presentes nos dois padrões. Já o envelope rotulado pela letra **c**, ainda na Figura 2, foi construído simplificando todos polígonos pelos seus centróides. Note que, o segundo envelope indica evidências de que possa existir uma atração entre os dois padrões de polígonos. Outro detalhe deste envelope é que o mesmo aparenta ter uma maior na variação das linhas pretas ao longo do eixo x em relação ao envelope baseado em áreas.

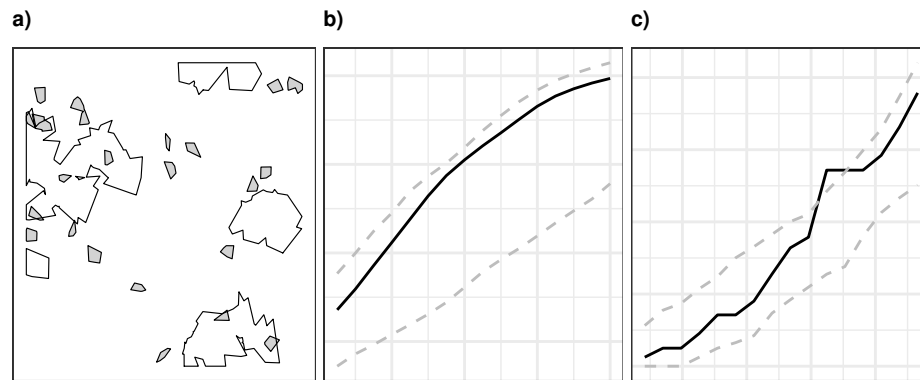


Figura 2 – a) Dados simulados de maneira independente. Os centróides do padrão de polígonos brancos seguem um processo de Matérn, enquanto os centróides do padrão cinza foram simulados de acordo com um processo Poisson homogêneo; b) Envelope da função K cruzada adaptada para polígonos; c) Envelope da função K cruzada onde os polígonos foram simplificados pelos seus centróides.

Embora outros autores, principalmente no campo da ecologia, tenham tentado evitar reduzir polígonos à pontos (Wiegand et al. (2006), Nuske et al. (2009), Guo et al. (2013)), estudos sobre tal metodologia são escassos. Conseqüentemente, são poucas as estratégias publicadas para trabalhar com padrões de polígonos. Este trabalho visa preencher esta lacuna metodológica presente na literatura através da construção de testes de hipótese de Monte Carlo baseados em adaptações do modelo nulo *toroidal shift* de Lotwick and Silverman (1982). Também foram propostas extensões da função K cruzada de Ripley (Diggle (2003), Lotwick and Silverman (1982)) que utilizam a distância de Hausdorff (Dubuisson and Jain (1994)) ou uma função baseada no buffer de polígonos ao invés da distância euclidiana entre fronteiras para quantificar uma medida de distância entre polígonos. Além disso, são utilizadas as ideias de Loosmore and Ford (2006), Grabarnik et al. (2011), Baddeley et al. (2014) e Myllymäki et al. (2017) a respeito da utilização de funções como a K cruzada de Ripley (Diggle (2003)) como estatística de teste.

Correções de efeitos de borda são um tema importante quando se analisa padrões espaciais. Estes métodos são baseados na ideia de que determinadas estatísticas necessitam de correções quando os padrões estudados possuem eventos próximos das bordas da região de estudo. Ripley (1979) estudou estes métodos em detalhes e propôs diversas correções que, por sua vez, também foram extendidas para o contexto de padrões de polígonos neste trabalho.

As maiores dificuldades associadas a adaptação do método de Lotwick and Silverman (1982), originalmente proposto para processos pontuais, para padrões de polígonos é causada pela eventual quebra de polígonos em subpolígonos. Isto implica em dificuldades tanto conceituais quanto computacionais, o que motivou autores como Wiegand et al. (2006) e Guo et al. (2013) a não utilizar esta metodologia. As soluções aqui propostas envolvem a utilização de identificadores únicos para cada polígono de modo que, caso um polígono

seja quebrado em mais de um pedaço, estes novos polígonos gerados pela quebra recebem o mesmo indicador único e, por consequência, de certa forma, serão considerados como sendo o mesmo polígono.

Além destas contribuições metodológicas, neste trabalho também foi desenvolvido um estudo de simulação onde são comparados testes de Monte Carlo utilizando diferentes estatísticas de testes e diferentes modelos nulos. As comparações foram feitas no que diz respeito ao erro do tipo I e ao poder dos testes.

Por fim, foi feita uma análise dos dados de [Chu et al. \(2013\)](#) sobre a interação entre plantas sob diferentes condições.

2 Trabalhos Relacionados

Em estatística diversos autores já abordaram processos pontuais multitypos. Neste contexto [Lotwick and Silverman \(1982\)](#) propôs um Teste de Monte Carlo baseado na função K cruzada de Ripley, veja [Diggle \(2003\)](#) para exemplos da utilização desta função, para verificar se dois processos pontuais possuem algum tipo de associação espacial.

[Loosmore and Ford \(2006\)](#), [Grabarnik et al. \(2011\)](#) e [Baddeley et al. \(2014\)](#), discutiram detalhadamente o uso de funções como a K de Ripley, a L de Besag e a G, como estatísticas de teste para testes de hipóteses de Monte Carlo. Todos estes autores alertaram que a utilização destas funções avaliadas em um conjunto de distâncias não são estatísticas de teste adequadas sem o devido tratamento, pois sua utilização pode acarretar em um aumento exponencial no erro do tipo I associado aos testes em questão. Como solução, eles utilizaram *deviation* e *goodness of fit tests*. [Myllymäki et al. \(2017\)](#), estendeu estes métodos para o contexto de processos pontuais multitypos.

Os processos pontuais multitypos também têm sido objeto de estudo de pesquisadores da ciência da computação, onde alguns autores utilizam métodos de *data mining* para lidar com este tipo de problema. [Huang et al. \(2004\)](#) definiu o problema de encontrar relações entre padrões de pontos como *collocation pattern*, bem como introduziu uma abordagem baseada em uma medida chamada *Participation Index*, que, por sua vez, utiliza as mesmas informações da função K cruzada de Ripley em uma determinada distância fixa para o seu cálculo. Ainda de acordo com este autor, *collocation pattern* é o processo de encontrar subconjuntos de processos pontuais que tendem ocorrer conjuntamente com uma frequência maior ou menor que o usual. Quando falam em subconjuntos de processos pontuais, os autores querem dizer, por exemplo, dois processos com rótulos distintos. A distinção entre *collocated patterns* é feita através de *collocation rules*, que são regras baseadas no *participation index*.

Uma extensão do *participation index* chamada *utility participation index* foi proposta por [Wang et al. \(2017\)](#). Sua diferença em relação ao indicador antecessor se dá pelo fato desta medida utilizar covariáveis como espécies de pesos para o *participation index*. [Lu et al. \(2017\)](#) explorou o problema em que dois padrões de pontos parecem possuir uma relação de atração mas na verdade são *pares competitivos*.

Ainda na área da ciência da computação, [Wang et al. \(2013\)](#) acentuou a conexão entre a metodologia utilizada em *collocation patterns* e estatística, desenvolvendo um método para localizar subregiões em um espaço contínuo nas quais dois padrões de pontos tem uma relação mais (ou menos) intensa de atração. Neste mesmo trabalho também foram propostos testes de hipóteses, baseados em estatísticas *scan*, para testar a significância

destas subregiões.

Estes métodos, propostos por pesquisadores da ciência da computação, são apropriados para casos em que tanto a quantidade de observações quanto o número de processos pontuais, ou diferentes padrões, associados ao problema em questão são de alta dimensionalidade. Contudo, existem limitações metodológicas associadas aos *collocation patterns*. Alguns exemplos destas limitações são que estes métodos, com exceção de algumas partes da metodologia de Wang et al. (2013), são determinísticos. Além disso, a utilização destas metodologias envolvem a entrada por parte do usuário de um *threshold* para a distância máxima que dois pontos podem apresentar para serem considerados *collocated*. A imputação de um *threshold* arbitrário não é trivial e pode ser uma fonte de erro em estudos aplicados.

Quando os objetos de interesse são polígonos, a maioria dos estudos que abordaram este tipo de problema são oriundos da ecologia. Dentre estes estudos, Wiegand et al. (2006) propôs uma extensão para esta metodologia compatível com dados representados por polígonos. Sua metodologia converte estes polígonos em *rasters* com o objetivo de reduzir a complexidade do problema. A desvantagem deste método é ignorar interseções entre polígonos, o que ocasiona perda de informação, que em muitos casos podem ser valiosas. Além disso, este autor utilizou métodos de simulação da hipótese nula que não preservam as estruturas espaciais marginais dos processos espaciais.

Nuske et al. (2009) utilizou estratégias similares as de Wiegand et al. (2006) mas para um problema no qual apenas um padrão de polígonos é observado, tendo como objetivo determinar se este padrão de polígonos é distribuído de forma aleatória no espaço ou se existe alguma associação espacial entre polígonos do mesmo tipo. Nuske et al. (2009) também estendeu a *pair correlation function* utilizada por Wiegand et al. (2006) para adequá-la aos padrões de polígonos.

Guo et al. (2013) estudou o problema quando deseja-se testar a associação espacial entre um padrão de polígonos e outro de pontos. A exemplo de Wiegand et al. (2006), este autor utilizou um modelo nulo que não preserva as estruturas marginais dos processos.

3 Um Teste Monte Carlo Condicional

Testes de Monte Carlo para dados espaciais tem sido utilizados por um longo período de tempo e já foram estudados exhaustivamente em diversos livros e artigos, tais como [Lotwick and Silverman \(1982\)](#) e [Manly \(2006\)](#). Estes testes podem ser definidos, de maneira simplista como: dada uma estatística de teste $\{T_i : i = 1, \dots, n\}$, simular um número $n - 1$ suficientemente grande de dados sob a hipótese nula que deseja-se testar e então calcular a estatística T escolhida para cada um destes conjuntos de dados simulados indexados por i , bem como para os dados observados. Estes cálculos permitem que uma distribuição de probabilidade empírica da estatística T seja construída sob a hipótese nula previamente definida. A seguir, esta distribuição empírica é utilizada para calcular as probabilidades $P(T \geq T_{obs})$ e $P(T \leq T_{obs})$. Se alguma destas duas probabilidades for menor que um nível de significância α , então a hipótese nula é rejeitada.

3.1 Hipóteses

Dados dois conjuntos (ou padrões) de polígonos, onde cada polígono é identificado por um Identificador Único e um rótulo indicando sua classe, o interesse é testar se um padrão de polígonos atrai ou repele o outro. Para testar este tipo de hipótese, a hipótese nula foi formulada de modo que os polígonos de um padrão e do outro são espacialmente independentes. Isto é, a presença ou ausência de um polígono do tipo A , não altera a distribuição espacial dos polígonos do tipo B .

Portanto, dada esta hipótese nula e os objetivos deste teste, temos duas hipóteses alternativas potenciais. E são elas:

- (i) Os padrões possuem uma relação de atração;
- (ii) Os padrões possuem uma relação de repulsão.

Basicamente, ao rejeitar H_0 , estamos aceitando, com um nível de significância α , que os polígonos dos dois padrões observados não são espacialmente independentes. Para determinar se a relação espacial entre estes polígonos está associada ao item (i) ou ao item (ii), deve-se analisar a estatística de teste construída e utilizada para o teste em questão.

3.2 O Toroidal Shift

[Lotwick and Silverman \(1982\)](#) propuseram uma abordagem inteligente para simular dois processos pontuais, observados em uma região retangular, sob a hipótese de

independência espacial entre os mesmos. Para ilustrar o método, considere dois processos pontuais arbitrários chamados de Processo A e B , respectivamente, e observados na região retangular \mathcal{D} . Então, a proposta dos autores pode ser descrita como segue.

1. Transforme a região \mathcal{D} em um torus (Figura 3);
2. Mantenha um dos dois processos pontuais fixos. (Por hora, consideremos que o processo escolhido para ser mantido fixo é o indexado como A);
3. Gere um novo processo aplicando uma translação aleatória no processo do B ;
4. Repita o passo anterior por um número de vezes suficientemente grande.

Então, para cada iteração deste método, será gerado um processo pontual multitypo onde a estrutura espacial conjunta é de independência, enquanto as estruturas marginais são as mesmas dos processos originais.

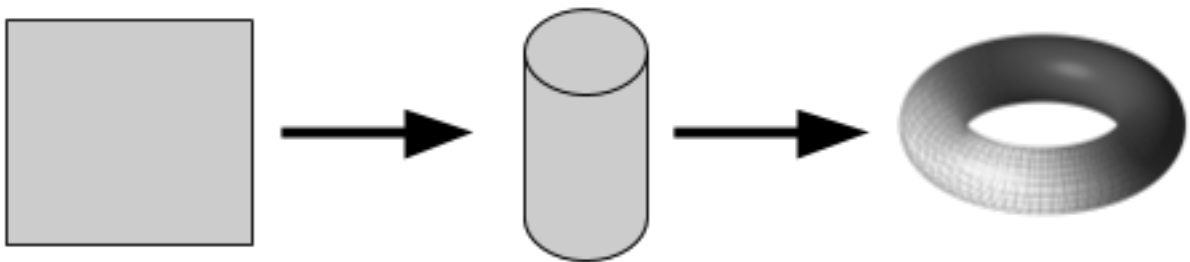


Figura 3 – Exemplo de uma região retangular envolvida em um torus.

A grande vantagem deste método é dada pela completa preservação das estruturas espaciais marginais de cada um dos processos. Esta característica implica em testes mais poderosos do que aqueles nos quais tais estruturas não são preservadas.

3.3 O que vem sendo feito na prática

O método desenvolvido por [Lotwick and Silverman \(1982\)](#) se tornou muito popular tanto em estudos na própria área de estatística quando em um contexto aplicado em outras áreas. Com essa popularização, uma alternativa mais simples para imitar o *Toroidal Shift* tem sido utilizada em diversos softwares, como o *spatstat*¹ de [Baddeley and Turner \(2005\)](#). Embora esta simplificação venha sendo utilizada no contexto de processos pontuais espaciais, ela será descrita considerando dois padrões espaciais de polígonos.

Considere dois padrões de polígonos espaciais, suponha também que uma estatística de teste já tenha sido definida e que estes polígonos estão localizados em uma região

¹ Atualmente, o *spatstat* é o ambiente mais utilizado na análise estatística de processos pontuais espaciais.

quadrada ou retangular \mathcal{D} . Estes padrões de polígonos serão rotulados como padrões A e B . O objetivo aqui consiste em simular dados sob a hipótese de que A e B são espacialmente independentes, visando testar se existe alguma associação espacial entre estes padrões.

Para adaptar o *Toroidal Shift* de Lotwick and Silverman (1982), uma primeira mudança no algoritmo deve ser feita no seu primeiro passo. Devida a complexidade da mudança de coordenadas necessária para envolver a região de estudo em um torus, um método equivalente é composto pela criação de três cópias da região de estudo, dando origem à uma nova região que será chamada de \mathcal{D}^* . Note que esta região é formada por um grid de quatro regiões retangulares, onde a 4ª região é a região \mathcal{D} (Figura 4).

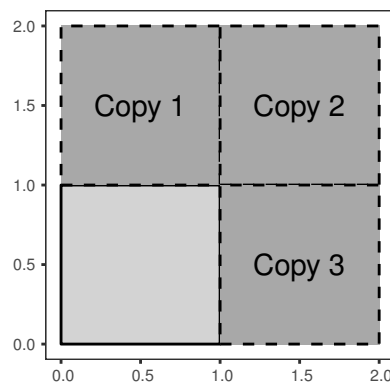


Figura 4 – Região \mathcal{D}^* , composta por \mathcal{D} e suas cópias.

Criada esta nova região, um dos padrões de polígonos observados necessita ser aleatoriamente escolhido para que sejam criadas três cópias do mesmo, da mesma maneira que a região \mathcal{D} . Feito isso, uma translação aleatória da região de interesse dará origem à uma nova região. Esta mesma translação deve ser aplicada ao padrão de polígonos que não foi selecionado no passo anterior, ou seja, aquele que não foi copiado três vezes previamente. Na Figura 5, os rótulos A, B, C e D são responsáveis por ilustrar como os padrões de polígonos que ficarão fixos são observados em uma região aleatoriamente transladada. Isto é equivalente a calcular a interseção entre a região \mathcal{D}^* e os polígonos nela contidos com os com a nova região gerada pela translação aleatória. Note que, polígonos do padrão fixo podem ser quebrados ao calcular as translações aleatórias.

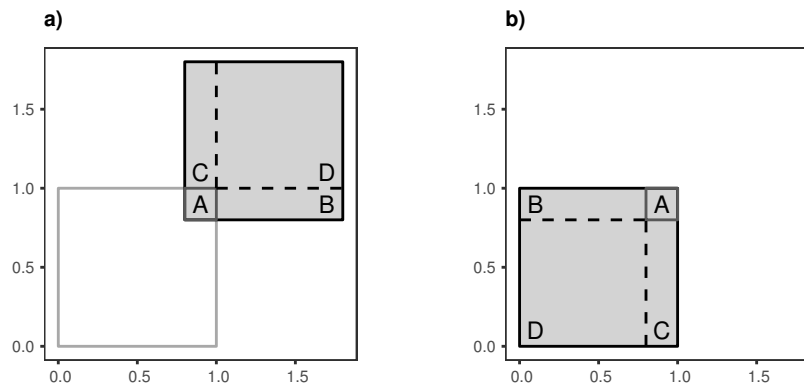


Figura 5 – a) Translação aleatória; b) Dados Observados.

Mesmo que esta abordagem imite o *Toroidal Shift*, ela não o faz perfeitamente. Existe um detalhe muito importante que não deve ser esquecido. Quando os objetos de estudo forem padrões de polígonos, e não de pontos, este método pode quebrar polígonos, conforme apontado anteriormente. Em outras palavras, no passo do algoritmo no qual se calcula a interseção entre a região transladada de forma aleatória com a região \mathcal{D}^* , um mesmo polígono pode ser dividido entre os dois extremos da nova região. Entretanto, estes polígonos divididos que aparecem nos dois extremos de uma região, serão tratados como um único polígono através do seu identificador único. Considerar um polígono dividido como um único polígono é um detalhe muito importante nesta metodologia, pois o caso contrário introduziria um viés na distribuição de qualquer estatística de teste baseada em contagens de polígonos. Como consequência direta, este viés faria com que a mesma estatística calculada para os dados observados não fosse comparável com a distribuição empírica construída.

3.4 A distância de Hausdorff e uma pequena adaptação do Toroidal Shift

A distância de *Hausdorff* é um tipo de distância muito utilizada em algoritmos de correspondência de objetos e reconhecimento de imagens. A utilização nestas áreas se dá por a mesma ser uma medida de distância mais apropriada para lidar com objetos diferentes de pontos, dentre os quais estão incluídos polígonos. Veja [Huttenlocher et al. \(1993\)](#); [Dubuisson and Jain \(1994\)](#); [Jesorsky et al. \(2001\)](#) para exemplos.

Esta métrica entre dois polígonos pode ser definida de maneira simples como sendo a distância máxima de um polígono para o ponto mais próximo do outro. Dada esta definição, é fácil ver que a distância de Hausdorff não é apropriada para o caso descrito na seção anterior, onde os polígonos podem ser divididos. Pois, mesmo considerando os identificadores únicos, estatísticas baseadas nestas distâncias não seriam comparáveis com

as calculadas nos dados originais, tornando assim testes de hipóteses baseados nestas estatísticas inválidos.

Afim de remediar este problema, uma pequena alteração na adaptação *Toroidal Shift* que viabiliza o uso da distância de *Hausdorff* consiste em, ao invés de calcular a interseção dos polígonos transladados aleatoriamente com a região \mathcal{D}^* , todos os polígonos que possuem qualquer área de interseção com a área transladada devem ser mantidos, conservando também os identificadores únicos de cada um dos polígonos. Desta maneira, aqueles polígonos que seriam quebrados agora podem aparecer até quatro vezes em uma mesma região aleatoriamente transladada.

3.5 Equívocos envolvendo a adaptação do Toroidal Shift para polígonos

Wiegand et al. (2006) criticou a utilização do *Toroidal Shift*, proposto por Lotwick and Silverman (1982), na construção de testes de Monte Carlo para a análise da associação espacial entre padrões de polígonos. A argumentação destes autores é fundada no viés introduzido pelo quebra de polígonos ocasionada pela técnica. Contudo, como já foi descrito neste trabalho, o nome *Toroidal Shift* foi dado baseando-se na transformação de uma região retangular de interesse, e todos os objetos nela contidos, em um torus. Uma consequência direta desta transformação é que nenhum objeto seria dividido nesta metodologia, isso porque o torus é uma superfície contínua. Entretanto, a simplificação do método ilustrada na subseção 3.3 se tornou mais popular que a metodologia original. O problema desta popularização é que a mesma fez com que os pesquisadores que a utilizam esquecessem daquilo que sua metodologia estava tentando reproduzir. Este mal entendido conduziu diversos autores à criticarem a aplicação do método de Lotwick and Silverman (1982) em geometrias diferentes de pontos. Contudo, a sutil modificação de incluir um identificador único para os polígonos implica em uma adaptação muito mais razoável do *Toroidal Shift*.

Autores como Wiegand et al. (2006) e Guo et al. (2013), utilizaram outros modelos nulos baseando-se na ideia de que o viés introduzido pela quebra de polígonos ocasionada pelo *Toroidal Shift* tornaria a metodologia inválida. O modelo nulo denominado *Antecedent Condition* utilizado por ambos autores foi introduzido por Wiegand et al. (2006). A aleatorização deste modelo é feita de modo que um dos padrões de polígonos é mantido fixo enquanto são aplicadas diferentes translações e rotações em cada um dos polígonos do padrão remanescente. Após a aplicação destas transformações nos polígonos, uma amostra é aceita se todos os polígonos estiverem completamente contidos dentro da região de interesse. Note que, esta metodologia acarreta em uma perda de informação, tendo em vista que a estrutura espacial marginal de um dos padrões de polígonos não é respeitada.

3.6 Propriedades de Segunda Ordem de Processos Espaciais Marcados

Quando o objeto de estudo de pesquisadores é representado por processos espaciais marcados², a estrutura de variabilidade destes processos é um tópico de extrema relevância, sendo ela, usualmente, utilizada para formular e testar hipóteses associadas à estes processos.

Dentre as diversas maneiras de mensurar este tipo de estrutura no contexto de processos pontuais, o método mais popular consiste em utilizar a função K cruzada de Ripley ou a sua versão com variância estabilizada chamada L cruzada de Besag (Ripley (1981)).

Neste trabalho são propostas três extensões da função K cruzada para a sua utilização com padrões de polígonos ao invés de pontos. A primeira, e mais simples, destas três extensões possui sua formulação exatamente igual à função aplicada para padrões de pontos substituindo apenas a distância entre pontos pela distância entre fronteiras de polígonos. A definição formal é dada por

Definição 1 *Seja $\mathbf{1}\{\cdot\}$ a função indicadora e d_{ij} a distância entre fronteiras do i -ésimo polígono do tipo A para o j -ésimo polígono do tipo B . Então a função K cruzada baseada em distâncias para polígonos é definida como*

$$\tilde{K}_{AB}(r) = \tilde{K}_{BA}(r) = (n_A n_B)^{-1} \mathcal{A}(\mathcal{D}) \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \mathbf{1}\{d_{ij} \leq r\}, \quad (3.1)$$

onde n_A e n_B são o número de polígonos em cada padrão e $\mathcal{A}(\cdot)$ é uma função que retorna a área de algum polígono ou conjunto de polígonos.

A grande limitação da função 1 se dá pelo fato de que a função trata como equivalentes no termo d_{ij} dois pares de polígonos que apresentam diferentes interseções. Isto é, digamos que o polígono 1 do tipo A tem sua fronteira encostada na fronteira do polígono 1 do tipo B , enquanto sua área de interseção com o segundo polígono do tipo B é muito grande. Neste caso, $d_{11} = d_{12} = 0$ e, portanto, a função não distingue diferentes graus de atração entre polígonos quando os mesmos possuem áreas de interseção. Esta característica indesejável pode ser contornada através da utilização de um outro tipo de distância que não a euclidiana, a distância de *Hausdorff*, definida na subseção 3.4. A adaptação da função K cruzada para a utilização da distância de *Hausdorff* requer algumas manipulações algébricas, tendo em vista que, neste tipo específico de distância, d_{ij} não necessariamente é igual a d_{ji} . Após estas manipulações, a função K cruzada baseada na distância de *Hausdorff* é dada por:

² *Processos Espaciais Marcados* são processos espaciais observados em uma mesma região mas que possuem diferentes *labels* ou grupos, tais como os padrões de polígonos estudados neste trabalho.

Definição 2 Seja $\mathbf{1}\{\cdot\}$ a função indicadora e d_{ij} a distância de Hausdorff entre o i -ésimo polígono do tipo A para o j -ésimo polígono do tipo B . Então,

$$\tilde{K}_{AB}(r) = \tilde{K}_{BA}(r) = (n_A n_B)^{-1} \mathcal{A}(\mathcal{D}) \times \frac{n_B \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \mathbf{1}\{d_{ij} \leq r\} + n_A \sum_{j=1}^{n_B} \sum_{i=1}^{n_A} \mathbf{1}\{d_{ji} \leq r\}}{n_A + n_B}. \quad (3.2)$$

Apesar da distância de Hausdorff ser robusta e tirar grande proveito da estrutura geométrica dos polígonos, a função K cruzada baseada nessa distância ainda pode obter mais informações a respeito dos padrões de polígonos que estão sendo estudados. Como estamos falando em polígonos, faz muito sentido utilizarmos a área destes objetos e a proporção da região de estudo ocupada pelos mesmos. Além desta alteração, é razoável substituir as medidas de distância por outra abordagem que considere tanto as áreas quanto a sua interseção com outros polígonos. Assim sendo, uma função candidata para substituir a distância entre dois polígonos é dada por:

Definição 3 Seja $\mathcal{B}\{A, r\}$ uma função que aplica um *buffer*³ de comprimento r , no polígono A e $\mathcal{A}(\cdot)$ a função que retorna a área de um polígono. Então,

$$b(A, B, r) = \mathcal{A}(B \cap \mathcal{B}\{A, r/2\}). \quad (3.3)$$

Definida esta função auxiliar, a função K cruzada baseada em áreas pode ser definida como:

Definição 4 A Função K Cruzada Baseada em Áreas é definida como

$$\hat{K}_{AB}(r) = \hat{K}_{BA}(r) = (\mathcal{A}(A)\mathcal{A}(B))^{-1} \mathcal{A}(\mathcal{D}) \times \frac{\mathcal{A}(A) \times b(A, B, r) + \mathcal{A}(B) \times b(B, A, r)}{(\mathcal{A}(A) + \mathcal{A}(B))}. \quad (3.4)$$

Note que, utilizando esta função, o padrão de polígonos que ocupar uma maior proporção da área de estudo terá mais peso no cálculo desta estatística. Nas outras duas adaptações da Função K Cruzada de Ripley aqui apresentadas, o padrão mais influente era aquele com um maior número de polígonos.

Outro detalhe importante a respeito destas funções adaptadas é que, para adaptar a função L Cruzada de Besag, aplica-se a seguinte transformação nas funções K : $L_{AB}(r) = \sqrt{\hat{K}_{AB}(r)/\pi}$. O grande benefício desta transformação é que, segundo Diggle (2003), ela estabiliza a variância da função K Cruzada.

³ Um *buffer* de um polígono representa uma expansão de mesmo comprimento em todas as direções.

3.7 Estatísticas de Teste

A primeira estatística de teste proposta para o problema é uma média da distância entre fronteiras de um polígono para o vizinho mais próximo do outro padrão. Embora esta estatística já tenha sido definida em [Jacquez \(1995\)](#), aqui ela será chamada de *Polygon Spatial Association Measure* e, por simplicidade, também será referida como *PSAM*.

Definição 5 A *Polygon Spatial Association Measure* é definida como:

$$m_{i,j} = m_{j,i} = \frac{\sum_{i=1}^{n_A} (\min_j(d_{i,j})) + \sum_{j=1}^{n_B} (\min_i(d_{i,j}))}{n_A + n_B}. \quad (3.5)$$

Uma outra extensão natural, e mais apropriada para o nosso problema, dessa mesma estatística é construída através da substituição da distância euclidiana pela distância de *Hausdorff*.

3.7.1 Envelopes ponto-a-ponto

Pointwise Envelopes foram propostos no contexto de Processos Pontuais Espaciais univariados por [Ripley \(1977\)](#) e popularizados na área de ecologia pelo trabalho de [Kenkel \(1988\)](#). A semelhança desta abordagem com a metodologia proposta neste texto é baseada no fato de que, a exemplo deste trabalho, as propriedades de segunda ordem dos processos pontuais são reduzidas à funções que possuem a mesma estrutura e as mesmas propriedades das funções K aqui utilizadas.

Este tipo de técnica fornece várias estatísticas de teste ao invés de apenas uma, como usual. Estas várias estatísticas, no nosso caso, correspondem à uma das funções K cruzadas avaliadas em um conjunto discreto de distâncias R . Ainda que pareça uma boa estratégia e tenha sido muito utilizada na literatura de processos pontuais, esta metodologia não é eficaz para testes de hipóteses de Monte Carlo. A ineficácia de testes do tipo *Pointwise Envelopes* se dá pelo fato de que, usualmente, eles consistem em avaliar a função K escolhida em um mesmo conjunto de distâncias tanto nos dados observados quanto nos simulados via o modelo nulo de Monte Carlo escolhido e então construir uma distribuição de probabilidade empírica para a função em cada distância. Feito isso, rejeita-se a H_0 ⁴ se em alguma destas distâncias nas quais a função foi avaliada o valor observado na amostra estiver acima ou abaixo dos percentis $\% \alpha$ e $\%(1 - \alpha)$, respectivamente, destas distribuições empíricas.

[Loosmore and Ford \(2006\)](#) estudou este problema trazendo a tona as complicações causadas por tal metodologia, que equivale a utilização diversas estatísticas de teste, rejeitando H_0 se alguma delas rejeitar esta hipótese. Isto acarreta em um aumento

⁴ Hipótese Nula

exponencial do *erro do tipo I*⁵. Caso estas estatísticas fornecidas pelas avaliações da função K em diversas distâncias não fossem correlacionadas, métodos de correção tais como o de *Bonferroni* poderiam ser utilizados para produção de testes com tamanho α , a exemplo de outros métodos estatísticos como testes de comparações múltiplas utilizados em planejamentos de experimentos.

Autores renomados como [Diggle \(2003\)](#); [Baddeley et al. \(2014\)](#), alertaram em seus trabalhos que os *Pointwise Envelopes* devem ser evitados para testes de hipóteses. Contudo, ambos recomendam a utilização desta técnica para realização da análise exploratória dos dados.

3.7.2 Goodness of Fit ou Deviation Tests

Embora [Diggle \(2003\)](#) tenha recomendado o uso de estatísticas que resumam funções do tipo K para testar hipóteses a respeito das estruturas de variabilidade de processos pontuais, tanto univariados quanto marcados, tal metodologia ganhou a devida atenção apenas com o trabalho de [Loosmore and Ford \(2006\)](#). Estes autores apontaram a falha metodológica ocasionada pela utilização da função K avaliada em diversas distâncias como estatística de teste. A estatística de teste, referida aqui como *Integral Measure (IM)*, proposta por [Loosmore and Ford \(2006\)](#) para resumir uma função K e, conseqüentemente, fornecer um teste de hipóteses com o erro do tipo I controlado é definida na equação 3.6.

$$u_i = \int_R (H_i(r) - \bar{H}_i(r))^2 \delta r, \quad (3.6)$$

onde a função $H_i(r)$ pode ser substituída por qualquer uma das funções (K ou L) aqui propostas e $\bar{H}_i(r) = \frac{1}{n-1} \sum_{j \neq i} H_j(r)$. Note que esta estatística é análoga à utilizada no teste de qualidade de ajuste de Cramér-von Misses, utilizado para comparar funções de distribuição acumuladas que, assim como as funções do tipo K e L, são funções monótonas não decrescentes. A semelhança com este tipo de teste faz com que testes de hipóteses baseados na metodologia de [Loosmore and Ford \(2006\)](#) sejam chamados de *Goodness of Fit Tests*.

Seguindo a mesma linha de raciocínio, outros autores ([Grabarnik et al. \(2011\)](#); [Baddeley et al. \(2014\)](#); [Myllymäki et al. \(2017\)](#)) utilizaram como estatística de resumo para as funções mencionadas na seção 3.6 uma estatística baseada no máximo desvio absoluto entre uma curva observada e a curva média. Esta estatística, chamada *Maximum Absolute Deviation* (ou *MAD*), que se assemelha à utilizada no teste Kolmogorov-Smirnov de *goodness of fit* é apresentada na equação 3.7, onde $H_i(r)$ e $\bar{H}_i(r)$ são análogos aos apresentados na equação 3.6.

$$u_i = \max_{r \in R} |H_i(r) - \bar{H}_i(r)|. \quad (3.7)$$

⁵ $P(\text{Rejeitar } H_0 | H_0 \text{ Verdadeira})$

Extensões das estatísticas de resumo apresentadas acima que são robustas aos casos em que as funções responsáveis por descrever as propriedades de segunda ordem dos processos espaciais apresentam heterogeneidade de variâncias ou assimetria foram propostas por Myllymäki et al. (2017). No primeiro caso a correção é simples e consiste em dividir, para cada r , as estatísticas por $\sqrt{Var(H(r))}$. As versões studentizadas destas estatísticas são apresentadas nas equações 3.8 e 3.9 e são denominadas *Studentized Integral Measure* e *Studentized Maximum Absolute Deviation*, respectivamente.

$$u_i = \int_R \frac{(H_i(r) - \bar{H}_i(r))^2}{\sqrt{Var(H(r))}} \delta r, \quad (3.8)$$

$$u_i = \max_{r \in R} \left| \frac{H_i(r) - \bar{H}_i(r)}{\sqrt{Var(H(r))}} \right|. \quad (3.9)$$

Nos casos em que as funções K (ou L) cruzadas apresentam distribuições assimétricas para cada $r \in R$, Myllymäki et al. (2017) sugeriu a utilização de correções *Quantil Direcionais*. Estas correções consistem em dividir as estatísticas nas quais $H_i(r) \geq \bar{H}_i(r)$ pela distância absoluta entre o percentil $1 - \alpha/2$ de $H(r)$ e $\bar{H}(r)$. Enquanto nos casos em que $H_i(r) < \bar{H}_i(r)$, o divisor é definido pela distância absoluta entre $\bar{H}(r)$ e o percentil $\alpha/2$ de $H(r)$. A nomenclatura utilizada para essas estatísticas corrigidas é *Directional Quantile Integral Measure* e *Directional Quantile Maximum Absolute Deviation*, e as mesmas são apresentadas nas equações 3.10 e 3.11, respectivamente.

$$u_i = \int_R \left(\mathbf{1}_{sup}\{H_i(r)\} \frac{(H_i(r) - \bar{H}_i(r))^2}{|H_{1-\alpha/2}(r) - \bar{H}_i(r)|} + \mathbf{1}_{inf}\{H_i(r)\} \frac{(H_i(r) - \bar{H}_i(r))^2}{|\bar{H}_i(r) - H_{\alpha/2}(r)|} \right) \delta r, \quad (3.10)$$

$$u_i = \max_{r \in R} \left[\mathbf{1}_{sup}\{H_i(r)\} \left| \frac{H_i(r) - \bar{H}_i(r)}{H_{1-\alpha/2}(r) - \bar{H}_i(r)} \right| + \mathbf{1}_{inf}\{H_i(r)\} \left| \frac{H_i(r) - \bar{H}_i(r)}{\bar{H}_i(r) - H_{\alpha/2}(r)} \right| \right]. \quad (3.11)$$

Onde, $H_{\alpha/2}(r)$ e $H_{1-\alpha/2}(r)$ representam os percentis da distribuição empírica de $H(r)$, $\mathbf{1}_{sup}\{H_i(r)\} = \mathbf{1}\{H_i(r) \geq \bar{H}_i(r)\}$ e $\mathbf{1}_{inf}\{H_i(r)\} = \mathbf{1}\{H_i(r) < \bar{H}_i(r)\}$.

3.8 Correções de Borda

Ripley (1979) pesquisou sobre testes de "aleatoriedade" (*Complete Spatial Randomness*) e correções de borda no contexto de processos pontuais espaciais univariados, propondo diversas correções para diferentes estatísticas de teste, dentre elas, a tão falada Função K de Ripley. As correções de borda foram baseadas na ideia de que os dados são observados em uma região $D \subset \mathcal{D}$ onde, por alguma razão, não foi possível coletá-los em toda a região \mathcal{D} . Portanto, estatísticas que estimem a área ou o número de polígonos dentro de uma distância fixa, tais como as utilizadas neste estudo, precisam de um tratamento especial quando eventos são observados muito próximos das fronteiras da região de estudo D . Este tipo de comportamento observado nos dados tem reflexo nas estatísticas utilizadas, fazendo com que estas subestimem tanto contagens quanto áreas dentro de uma determinada distância.

Nas próximas subseções serão listadas três correções de bordas compatíveis com os dados utilizados nesta dissertação.

3.8.1 Correção Toroidal

A Correção Toroidal se assemelha ao *Toroidal Shift*, pois cria uma nova região baseada em cópias da região de estudo com o objetivo de reproduzir um toróide. Toróides são utilizados por serem regiões sem fronteiras. O benefício trazido por isso é que podem ser aplicadas infinitas translações à um objeto em um mesmo sentido sem que o mesmo seja transladado para uma região onde não foi possível fazer a coleta de dados. Na Figura 6, é apresentado um exemplo da correção toroidal sendo aplicada em um conjunto de dados simulado. Note que, são criadas oito cópias da região de estudo nas quais um dos padrões também é copiado, enquanto o remanescente permanece fixo, sendo observado apenas na região central, que corresponde a região de estudo original. Após a criação destas cópias as estatísticas de teste baseadas em distâncias ou *buffer* de polígonos são calculadas de modo que, por exemplo, $d_{i,j}$ seja mensurada de uma observação i contida na região central para observações j contidas nas regiões observadas ao redor da região original. Estas adequações fazem com que a função K cruzada seja reescrita de uma maneira similar à sua versão adequada para a utilização da distância de *Hausdorff*. Por tanto a função K apresentada na equação 1 é reescrita como segue:

$$\tilde{K}_{AB}(r) = \tilde{K}_{BA}(r) = (n_A n_B)^{-1} \mathcal{A}(\mathcal{D}) \times \frac{n_B \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} I(d_{ij} \leq r) + n_A \sum_{j=1}^{n_B} \sum_{i=1}^{n_A} I(d_{ji} \leq r)}{n_A + n_B}. \quad (3.12)$$

Note que as fórmulas das funções K cruzadas que são baseadas na distância de *Hausdorff* e em áreas, apresentadas respectivamente nas equações 2 e 4 permanecem inalteradas quando esta correção de borda é utilizada.

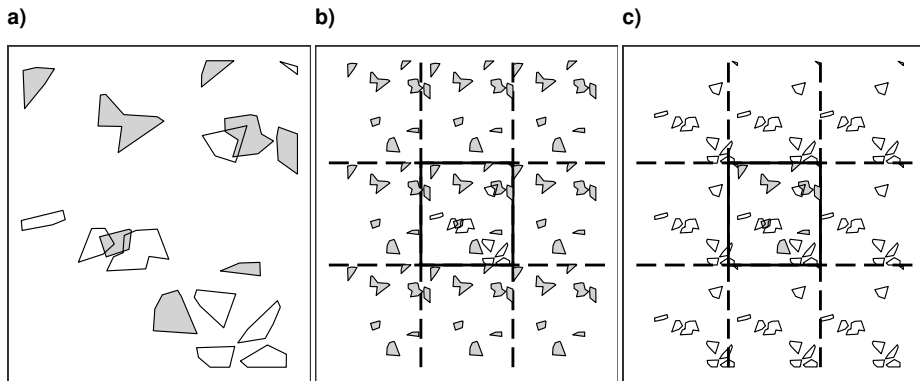


Figura 6 – **a)**: Dados simulados; **b)**: Correção toroidal aplicada para calcular as estatísticas referentes aos polígonos brancos em relação aos cinzas; e **c)**: Correção para mensurar estatísticas dos polígonos cinza com respeito aos brancos.

3.8.2 Área de Guarda

Outro método utilizado para remediar os problemas causados pela proximidade de unidades amostrais com as bordas da região de estudo é chamado Área de Guarda. Este método, como indica o seu nome, cria uma área de guarda, que se trata de uma espécie de cinturão, onde os dados são descartados para computar estatísticas. Para exemplificar este método, considere uma região de estudo \mathcal{D} composta por $\mathcal{C} \cup \mathcal{E}$. De modo que esta partição da região é formada por uma região com o mesmo shape da região original mas em uma escala menor. E a diferença entre a região original, representada por \mathcal{D} , e sua versão reduzida \mathcal{C} formam a região \mathcal{E} . Que também pode ser vista como a região \mathcal{C} após a aplicação de um *buffer* e a exclusão dela mesma. Feito isso, as estatísticas são calculadas de maneira similar à correção toroidal. Isto é, considerando a Figura 7, para o cálculo das distâncias do padrão cinza em relação ao branco, os polígonos brancos que ficam na região da guarda são desconsiderados. Enquanto para os cálculos das distâncias dos polígonos cinzas em relação aos brancos o mesmo procedimento é aplicado. Com isso, a função K cruzada recebe as mesmas modificações aplicadas na utilizada na correção toroidal. Um detalhe importante a respeito deste método é que ele tende a ser mais conservador em relação aos outros, por perder dados. Além disso, a definição do tamanho da área de guarda não é trivial e impacta diretamente nos resultados dos testes utilizados com estatísticas que são construídas utilizando esta correção.

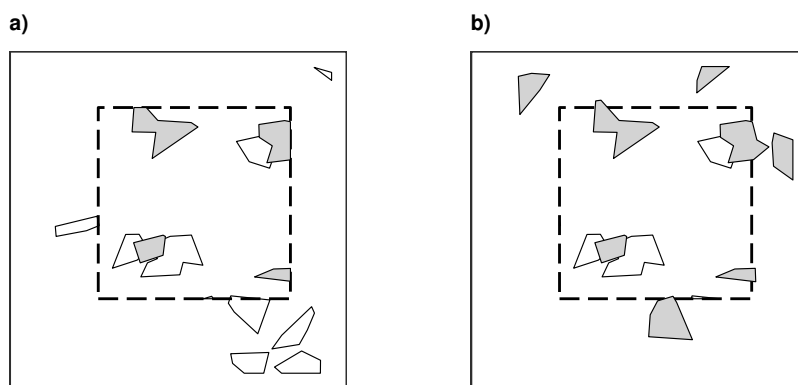


Figura 7 – (a): Área de guarda utilizada para computar distâncias (ou interseção de área) do padrão de polígonos cinzas em relação ao branco; (b): Área de guarda utilizada para computar distâncias (ou interseção de área) do padrão de polígonos brancos em relação aos polígonos cinzas

3.8.3 Fator de Ajuste

O Fator de Ajuste é um método simples de correção de efeitos de bordas, também é o mais utilizado na literatura. Esta abordagem baseia-se na ideia de que o número esperado de polígonos (ou o percentual esperado de área ocupado pelos mesmos) é o mesmo dentro e fora da região de estudo. Por tanto, quando calcula-se estatísticas baseadas, por exemplo, na contagem de polígonos dentro distância r do polígono i para polígonos j , multiplica-se

esta contagem por um peso w_{ij} que representa 1 sobre a proporção do polígono i com um buffer de $r/2$ que esta dentro da região de estudo. Veja Figura 8 e a equação 3.13, que exemplifica o calculo da função K cruzada utilizando este método de ajuste.

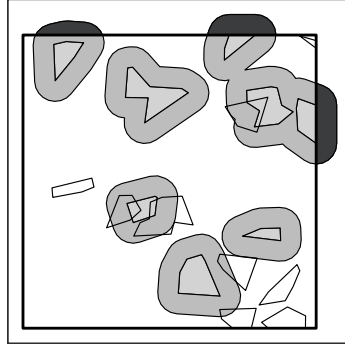


Figura 8 – Representação gráfica do fator de ajuste.

$$\tilde{K}_{AB}(r) = \tilde{K}_{BA}(r) = (n_A n_B)^{-1} \mathcal{A}(\mathcal{D}) \times \frac{n_B \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} w_{ij} I(d_{ij} \leq r) + n_A \sum_{j=1}^{n_B} \sum_{i=1}^{n_A} w_{ji} I(d_{ji} \leq r)}{n_A + n_B}. \quad (3.13)$$

4 Estudo de Simulação

Os objetivos principais deste estudo de simulação são comparar o poder e o erro do tipo I associado aos testes realizados com diferentes estatísticas bem como avaliar o poder associado com diferentes modelos nulos, isto é, diferentes maneiras de simular dados sob a hipótese nula.

4.1 Simulando Padrões de Polígonos

Afim de acessar as propriedades descritas no início do capítulo, foram desenvolvidos métodos para simular padrões de polígonos com diferentes estruturas espaciais. Todos os métodos aqui propostos seguem um mesmo padrão, primeiro são simulados os centróides dos polígonos com alguma estrutura espacial pré-definida e então são gerados polígonos associados a cada um dos centróides. Para ilustrar como os polígonos são simulados, considere que já foram gerados os centróides de todos eles. Então, para cada centróide, deve ser gerado um círculo de raio r^* centrado neste centróide. O próximo passo consiste em gerar n^* pontos uniformemente distribuídos neste círculo gerado e, por fim, calcular o menor polígono convexo que contém estes pontos (Eddy (1977)). Note que o parâmetro r^* controla o tamanho médio dos polígonos, enquanto n^* é responsável por definir a forma. Nenhum destes dois parâmetros precisam ser fixos, eles podem ser variáveis aleatórias. Outra possibilidade é que estejamos simulando polígonos tais que as formas e o tamanho variam de acordo com o padrão ao qual cada polígono pertence.

4.1.1 Padrões com Relação Espacial de Repulsão

Para a simulação de padrões de polígonos com relação espacial de repulsão (Figura 9), foram utilizados processos *Hardcore*, definidos a seguir.

Definição 6 *Dois processos pontuais espaciais A e B são considerados Processos Pontuais Hardcore Bivariados se, para cada dois pares de pontos pertencentes a padrões distintos, a distância entre os mesmos não é menor que uma distância fixa d . Esta distância é chamada Distância Hardcore.*

Dada a definição 6, a simulação de padrões de polígonos com repulsão espacial pode ser feita em dois passos, e são eles:

1. Gerar os centróides de acordo com um Processo Hardcore Bivariado;
2. Criar um polígono associado com cada um dos centróides gerados no passo anterior.

Note que a verdadeira *distância de hardcore*, que será denotada por θ , associada aos padrões de polígonos simulados é desconhecida e contida no intervalo $[d_{hc} - r_A^* - r_B^*, d_{hc}]$. Onde d_{hc} é a *distância de hardcore* utilizada para simular os centróides dos polígonos, enquanto r_A^* e r_B^* são os raios utilizados para gerar os polígonos associados aos centróides dos padrões A e B , respectivamente. Outro detalhe importante é que, quanto maior for θ mais repulsão existirá entre os diferentes padrões. Por outro lado, quanto mais próximo este parâmetro estiver de 0, os padrões de polígonos associados a este parâmetro mais se aproximarão de dois padrões espacialmente independentes.

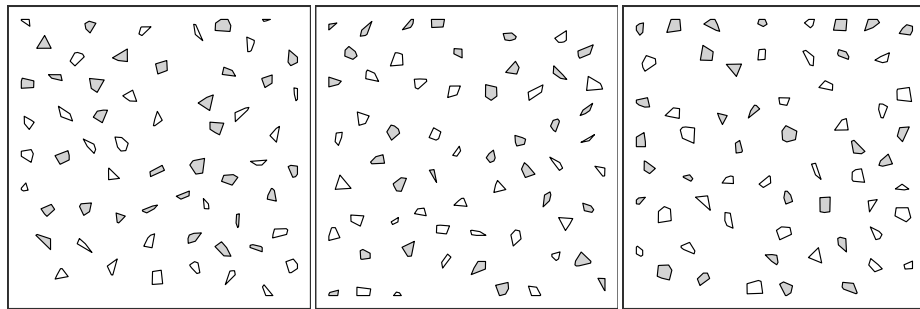


Figura 9 – Exemplos de padrões de polígonos simulados com relação de repulsão.

4.1.2 Padrões com Relação Espacial de Atração

Neste trabalho, a simulação de conjuntos de polígonos com relação de atração (Figura 10) são simulados de forma sequencial, gerando primeiro os centróides de um dos padrões, digamos padrão A , e então, para simular cada centróide relacionado ao outro padrão (que chamaremos de padrão B), um dos centróides de A é aleatoriamente escolhido com probabilidade $1/n_A$, onde n_A representa o número de polígonos associados ao padrão A . Posteriormente, para cada um dos centróides do padrão B , são gerados em círculos de raios r_{att} e centrados nos pontos de A aleatoriamente escolhidos. Feito isso, é simulado um ponto dentro de cada um destes círculos e, por fim, estes pontos são atribuídos como centróides dos polígonos do tipo B .

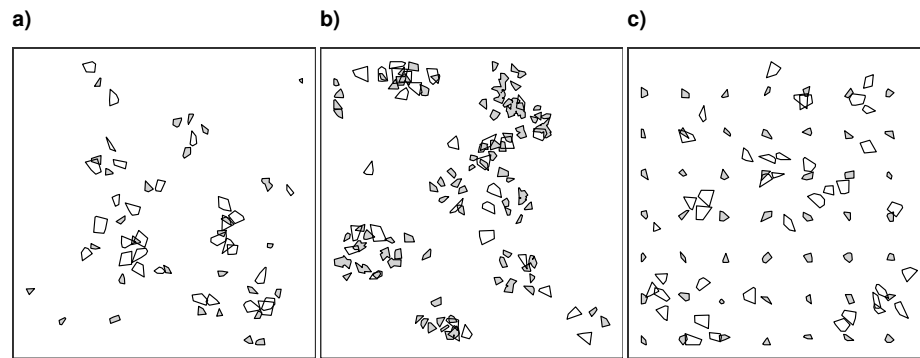


Figura 10 – Exemplos de dados simulados com relação de atração e diferentes estruturas marginais.

4.1.3 Padrões com Relação Espacial de Independência

A simulação de padrões independentes é o caso mais simples dentre os três apresentados nesta seção e consiste em gerar os centróides de cada um dos padrões de maneira independente e então criar os polígonos conforme indicado no início da seção. Este tipo de simulação será utilizada para estimar o erro do tipo I associado com as estatísticas de teste.

4.2 Poder do Teste

Um estudo de simulações de Monte Carlo foi realizado para conhecer estimativas empíricas do poder dos testes de Monte Carlo associados à cada uma das estatísticas de teste aqui propostas. Além disso, diferentes maneiras de simular a hipótese nula foram comparadas com respeito ao poder do teste. A adaptação proposta para o *toroidal shift* foi comparada com o modelo nulo utilizado por [Wiegand et al. \(2006\)](#) e [Guo et al. \(2013\)](#).

Na Tabela 1 são descritos os cenários utilizados para verificar o poder dos testes. Para cada um destes cenários foram simulados 1000 conjuntos de dados contendo dois padrões de polígonos observados em uma região quadrada com dimensões $(0, 1) \times (0, 1)$. Os parâmetros associados a forma e ao tamanho dos polígonos simulados foram 4 e 0.02, respectivamente. Posteriormente, para cada cenário simulado, os testes de Monte Carlo foram calculados utilizando 99 simulações e diferentes estatísticas de teste. Uma variável binária associada com cada estatística de teste (ou modelo nulo) foi criada de modo que a mesma assume o valor 1 quando a hipótese nula for rejeitada com um nível de significância $\alpha = 5\%$ e 0 no caso contrário. Conseqüentemente, o poder do teste associado a cada estatística de teste ou modelo nulo dentro de cada cenário de simulação é dada pela média aritmética desta variável binária.

Tabela 1 – Cenários simulados para estimar o poder do teste.

Parâmetro	n_A	n_B	Relação
$\phi \in [0, 0.8]$	50	50	Atração
$\theta \in [0, 0.0798]$	50	50	Repulsão

Note que enquanto altos valores de θ indicam uma relação severa de repulsão entre os padrões de polígonos, baixos valores de ϕ estão associados com uma forte atração entre estes padrões.

5 Resultados

5.1 Erro do Tipo I

Além dos resultados apresentados na Tabela 1, 1000 conjuntos de dados onde os padrões de polígonos eram espacialmente independentes foram simulados. Após a simulação destes dados, foram calculados testes de hipóteses usando cada uma das estatísticas descritas na seção 3.7, com o objetivo de estimar o erro do tipo aproximado destas estatísticas. Os resultados são apresentados na Figura 11 através de um painel onde, suas colunas indicam se a estatística foi calculada baseando-se na distância euclidiana, de Hausdorff ou em áreas. Enquanto as linhas deste painel representam qual função resumo das propriedades de segunda ordem foi utilizada, isto é, a função K cruzada ou sua versão com variância estabilizada L cruzada. Em cada um dos gráficos contido no painel, o eixo x representa o erro do tipo I, já o eixo y é responsável por discriminar as diferentes estatísticas de teste. Note que, para acomodar a estatística de teste $PSAM$ nestes gráficos, a mesma foi incluída nos gráficos associados a função K cruzada. Outro detalhe importante é que esta estatística não possui uma versão baseada em áreas e, por esta razão, a linha que representa esta estatística nos gráficos baseados em áreas está vazia. Ainda sobre os gráficos, o ponto representa a estimativa pontual do erro do tipo I e a linha que cruza o ponto representa o intervalo de confiança com confiança de 95% assumindo normalidade assintótica. Por fim, as linhas verticais pontilhadas indicam o erro do tipo I nominal.

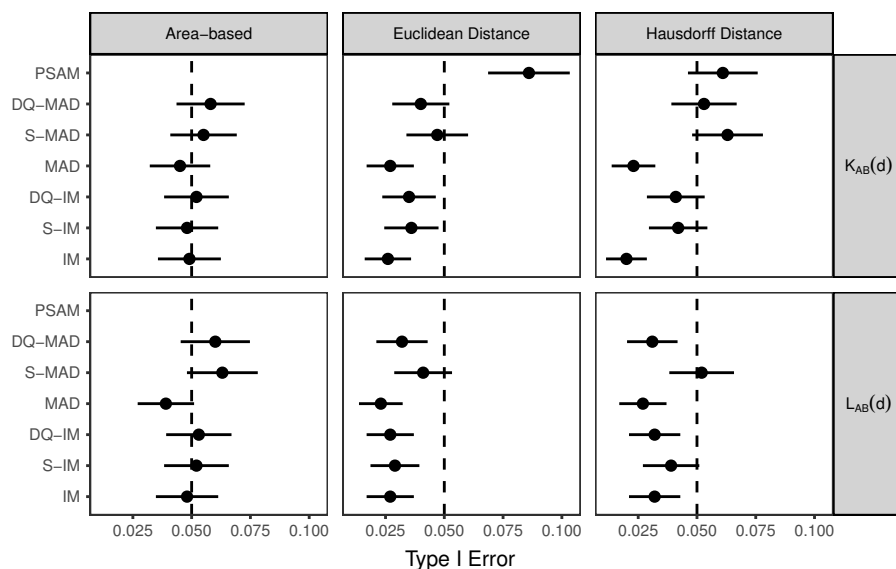


Figura 11 – Erro do tipo I estimado para diferentes funções e distâncias.

Todos os intervalos de confiança para o erro do tipo I associados com estatísticas de

teste que resumiam funções baseadas em áreas contiveram o nível de significância utilizado nos testes. Este é o resultado ideal e esperado quando testes de hipóteses são construídos. Outro detalhe importante a respeito destes resultados é que os testes realizados utilizando como estatística de teste a *Studentized Maximum Absolute Deviation* contiveram o erro do tipo I esperado, que era de 5%. Testes conduzidos utilizando a estatística *Studentized Integral Measure* apresentaram intervalos de confiança que contiveram o valor esperado do erro do tipo I tanto quando foi utilizada a função K ou L cruzadas baseadas na distância de *Hausdorff*. Os testes baseados nas outras estatísticas de teste apresentaram um erro do tipo I menor do que o nominal. Embora este resultado soe como favorável, este comportamento pode indicar testes muito conservadores e deve ser investigado com mais detalhes em trabalhos futuros.

5.2 Análise do Poder

5.2.1 Estatísticas de Teste

Para cada configuração apresentada na Tabela 1 foram simulados 1000 conjuntos de dados com o intuito de aplicar testes de Monte Carlo utilizando diferentes estatísticas de teste com um nível de significância $\alpha = 5\%$ à eles. Feito isso, estimativas empíricas do poder do teste de cada situação foram calculadas, bem como seus intervalos de confiança com 95% de confiança assumindo normalidade assintótica.

As estimativas empíricas do poder dos testes e seus respectivos intervalos de confiança para cada uma das estatísticas de teste são apresentadas nas Figuras 12 e 13, para os cenários de atração e repulsão, respectivamente. Nestas figuras, são apresentados painéis similares ao apresentado na Seção 5.1. A diferença é que aqui as linhas destes painéis representam o que era representado por colunas no painel da seção anterior. Enquanto as estatísticas de teste são representadas pelas colunas. Dentro de cada gráfico deste painel são apresentadas as curvas do poder estimadas para cada um desses cenários utilizando ou a função K cruzada (linhas sólidas cinzas) ou a função L cruzada (linhas pretas pontilhadas). As pequenas barras verticais são responsáveis por indicar os intervalos de 95% de confiança para as estimativas do poder. O eixo y de ambos os painéis representa o poder do teste, enquanto o eixo x é responsável por exibir os valores de: ϕ (atração) e θ (repulsão) para os quais o poder foi estimado. É importante resaltar que no eixo x do painel apresentado na Figura 12, ϕ está sendo apresentado em uma escala reversa, pois quanto menor seu valor, maior o poder do teste. Esta mudança de escala foi feita para manter um padrão em todos os painéis. Além disso, assim como no painel apresentado na Figura 11, a estatística *PSAM* não tem uma versão baseada em áreas.

No caso em que os padrões de polígonos possuem estrutura conjunta de atração, apresentado na Figura 12, as estatísticas baseadas na distância de *Hausdorff* obtiveram

melhores performances de acordo com o poder do teste. Outra inferência importante que este gráfico permite fazer é que testes baseados na função K cruzada em conjunto com a estatística de resumo *Maximum Absolute Deviation* tem uma performance ruim se comparados às outras estatísticas. Este problema é remediado utilizando ou a função L cruzada ou suas versões padronizada ou com correção de assimetria. Logo, pode-se inferir que esta baixa performance deve estar associada à grande variabilidade da função K cruzada, tendo em vista que, as três correções citadas lidam com este problema. Outro detalhe importante a respeito destes testes é a queda de performance em regiões do espaço de ϕ que indicam atração muito severa quando são utilizadas as funções K ou L em conjunto com a distância euclidiana. Mesmo que esta queda aparente ser atípica, ela pode ser atribuída a grande limitação deste tipo de distância que se mostra inadequada em distinguir pares de polígonos que possuem diferentes áreas de interseção. Neste tipo de distância, dois polígonos que se tocam possuem a mesma distância euclidiana que dois polígonos que se interseccionam completamente. Por fim, ainda que nenhuma estatística de teste domine completamente as outras no cenário de atração, aconselha-se a utilização de estatísticas de resumo baseadas na função L cruzada e na distância de *Hausdorff* para detectar este tipo de interação entre polígonos.

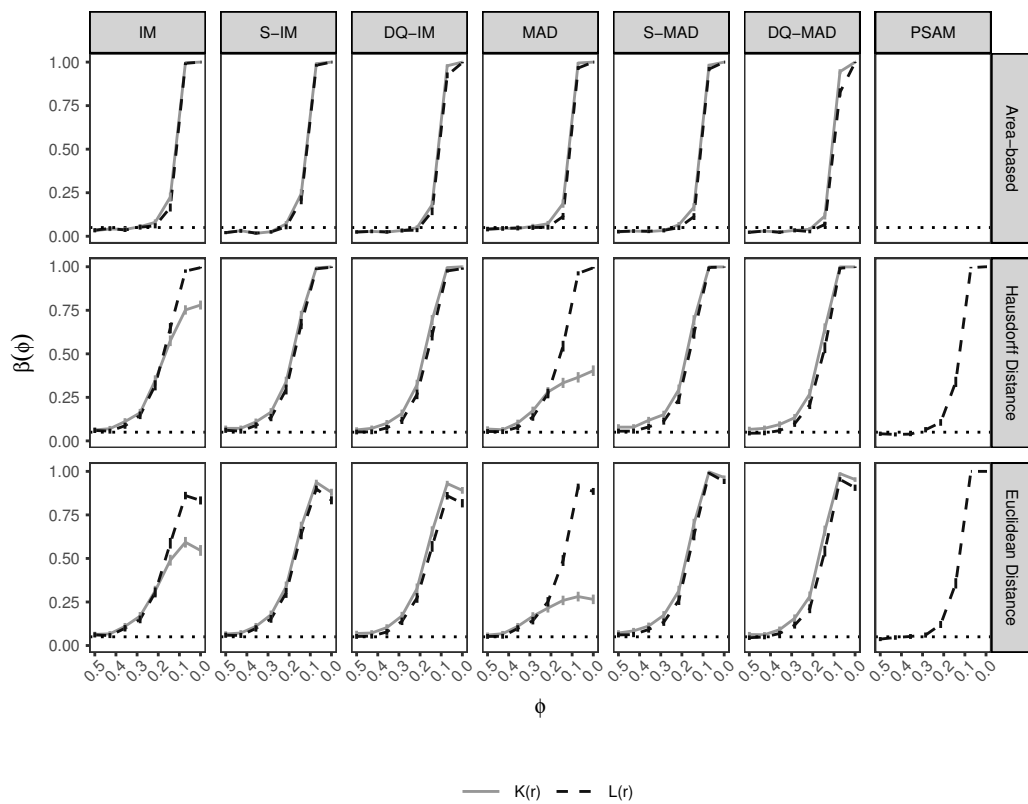


Figura 12 – Painel que ilustra o poder do teste no cenário de atração associado à diferentes estatísticas de teste utilizando diferentes distâncias e funções para descrever a segunda ordem dos processos.

Diferentemente do cenário de atração, o cenário de repulsão (Figura 13) sugere que testes onde as funções que descrevem as propriedades de segunda ordem dos processos espaciais são calculadas baseadas em áreas são consideravelmente mais poderosas que as outras. Entre as estatísticas baseadas em distâncias, a *PSAM* obteve uma melhor performance. Contudo, esta superioridade deve estar associada à baixa complexidade das estruturas conjuntas simuladas. Espera-se que, a medida que a complexidade destas estruturas aumente esta superioridade diminua.

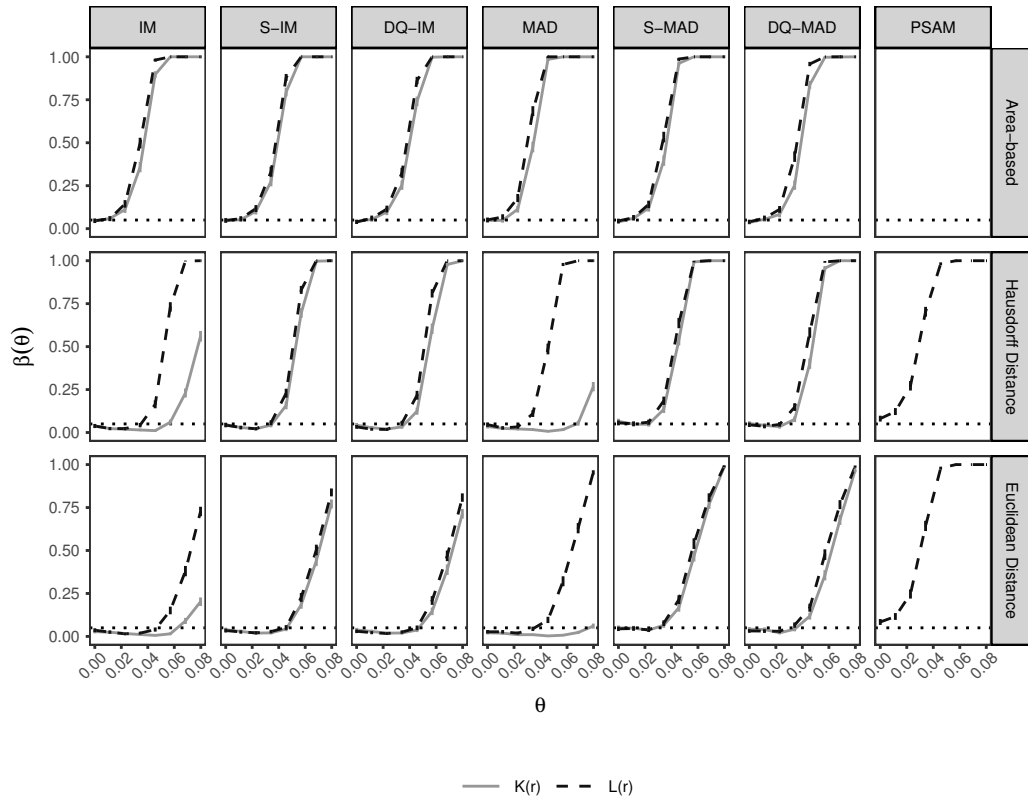


Figura 13 – Painel que ilustra o poder do teste no cenário de repulsão associado à diferentes estatísticas de teste utilizando diferentes distâncias e funções para descrever a segunda ordem dos processos.

Um padrão interessante em ambos os painéis é que, tanto a estatística *Integral Measure* quanto a *Maximum Absolute Deviation* apresentaram uma performance muito baixa em relação ao poder do teste quanto utilizadas com a função *K* cruzada baseada nas distâncias euclidianas e de *Hausdorff*. Este comportamento não se repete nas funções baseadas em áreas e, apesar de não termos um resultado teórico, isto pode ser um indício de que a variabilidade da função *K* baseada em áreas é menor. Outra possibilidade é que esta versão da função *K* consegue detectar de maneira mais acurada as estruturas conjuntas dos processos.

5.2.2 Modelos Nulos

Assim como na subseção anterior, foram simulados 1000 conjuntos de dados para cada configuração de parâmetros apresentada na Tabela 1. Nesta subseção o objetivo é comparar a performance no poder do teste do modelo nulo proposto neste trabalho com o *Antecedent Condition*, introduzido por Wiegand et al. (2006). A estatística de teste utilizada para esta comparação foi a *Studentized Integral Measure* baseada na função L cruzada e distância de *Hausdorff*, o nível de significância dos testes foi de 5%. A Figura 14 apresenta o poder associado à cada um destes modelos e seus respectivos intervalos de confiança com 95% de confiança no cenário de atração, enquanto a figura 15 mostra os resultados análogos para o cenário de repulsão.

De acordo com a Figura 14, a performance dos testes calculados com diferentes modelos nulos nos cenários de atração é similar quando esta atração é moderada. Entretanto, com o crescimento do grau de atração entre os padrões de polígonos, o *Toroidal Shift* se mostra um modelo nulo mais poderoso e, por fim, na zona onde a atração é severa ambos modelos se equivalem.

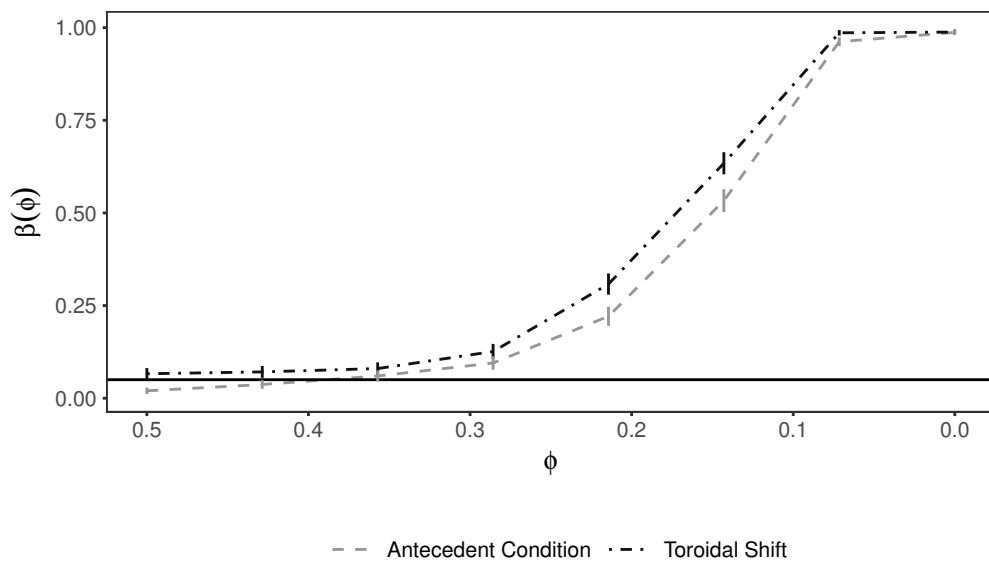


Figura 14 – Comparando métodos de simulação de dados sob a hipótese nula no cenário de atração.

No cenário de repulsão, apresentado na Figura 15, ambos modelos nulos foram equivalentes em relação ao poder do teste. Entretanto, tal equivalência é devida ao fato de que as estruturas marginais dos processos simulados são muito simples. Espera-se que, a medida que a complexidade, tanto das estruturas marginais quanto das estruturas conjuntas, aumente o *Toroidal Shift* se mostre um modelo nulo mais poderoso do que o *Antecedent Condition*. Pois, o *Toroidal Shift* é um modelo nulo que conserva completamente as estruturas marginais dos processos.

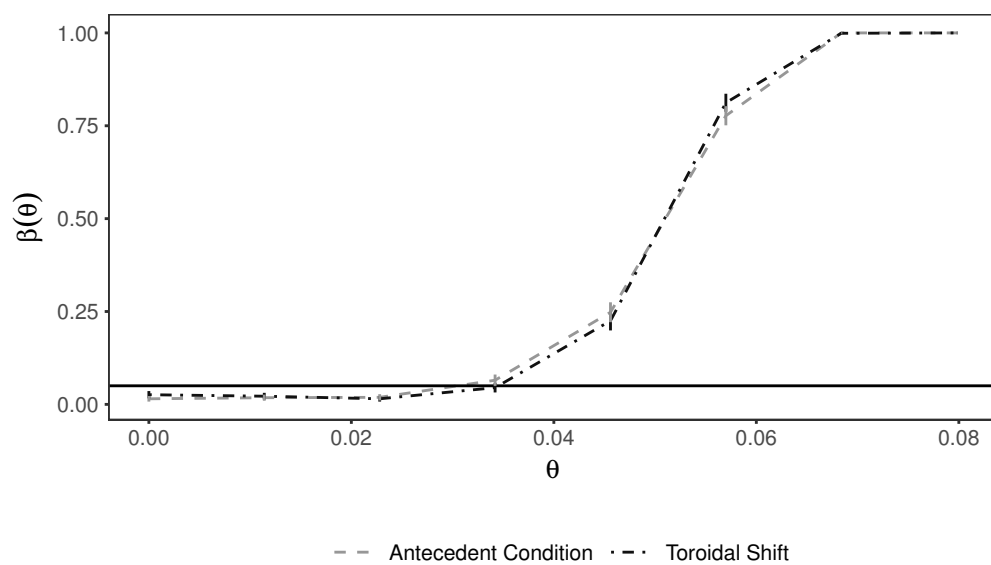


Figura 15 – Comparando métodos de simulação de dados sob a hipótese nula no cenário de repulsão.

6 Estudo de caso

Chu et al. (2013) realizou um experimento no *Central Plains Experimental Range*, em Nunn, Colorado, Estados Unidos. Um pantógrafo foi utilizado para coletar dados de localização e shape de gramas curtas da América do Norte em 24 *quadrats* permanentes de 1m² entre os anos de 1997 e 2010, provendo dados de qualidade que tornam possíveis estudos demográficos sobre este tipo de planta. Além destes estudos demográficos, este conjunto de dados, segundo os autores, também é ideal para analisar interações entre diferentes tipos de gramas.

Nas análises feitas aqui, foram utilizadas 22 espécies de gramas. Dentre elas estão *Bouteloua gracilis*, *Sitanion hystrix*, *Carex spp.* e *Stipa comata*, ilustradas na Figura 16. Outro detalhe importante, é que foram excluídos da análise dados de polígonos associados à cogumelos, fezes de vaca, gramas mistas e não identificadas.

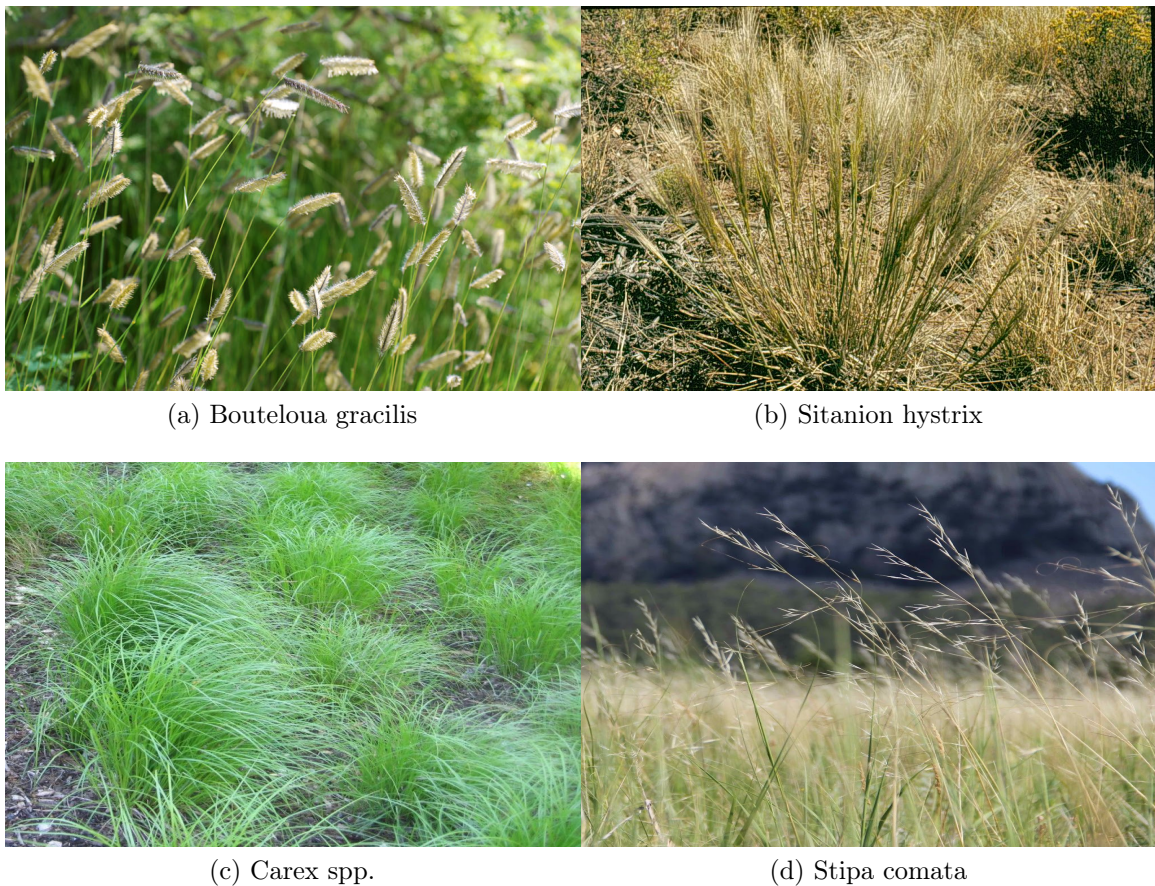


Figura 16 – Exemplos de tipos de gramas observadas nos dados de Chu et al. (2013).

O desenho experimental do estudo de Chu et al. (2013) foi feito de modo que suas unidades amostrais, i.e. os *quadrats*, fossem expostas a diferentes condições em relação à

dois fatores: o tipo pasto e a ocorrência de pastagem. Foram observados 6 tipos diferentes de pastos identificados pelos códigos: 19, 11, 24, 7, 5a, 5b. No que diz respeito à pastagem, o estudo foi feito em quatro grupos de áreas. O primeiro grupo é composto por áreas onde havia pastagem de pecuária tanto antes quanto depois de 1991, este grupo é codificado como *gzgz*; o segundo grupo é formado pelas áreas em que a pastagem ocorria antes de 1991 mas não após este ano, o código associado à este grupo é *gzun*; O penúltimo grupo é caracterizado por áreas nas quais a pastagem só começou a ocorrer a partir de 1991 e sua codificação é *ungz*; Por fim, o último grupo que é codificado por *unun* e é composto por áreas nas quais não havia pastagem nem antes e nem depois de 1991. Na Figura 17 está presente o georreferenciamento dos *quadrats*, onde cada um deles foi colorido de acordo com a ocorrência da pastagem, enquanto o seu shape é definido de acordo com o tipo de pasto.

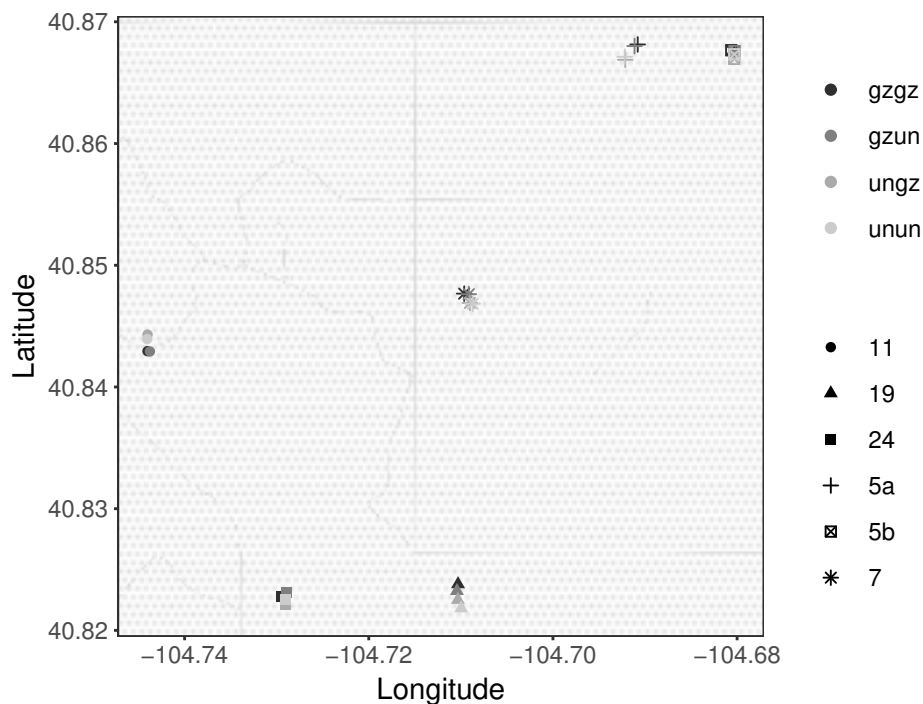


Figura 17 – Localização dos quadrats de acordo com o tipo de pasto e ocorrência de pastagem.

Afim de examinar fatores determinantes nas relações entre pares de espécies de gramas, foram aplicados os testes de hipóteses de Monte Carlo em cada uma das possíveis combinações entre duas espécies de plantas dentro de cada um dos *quadrats* com exceção dos dados observados nos anos de 2003 e 2009, resultando em um total de 1731 testes. Estes testes foram aplicados utilizando as seguintes especificações: 99 simulações, função L cruzada baseada em áreas e, como estatística de resumo, a *Integral Measure*. A distribuição

do número de testes por espécie de planta, pastagem, tipo de pasto e ano são ilustradas nas Tabelas 2, 3, 4 e 5, respectivamente.

Tabela 2 – Número de testes realizados por espécie de planta.

Espécie	Frequência	Frequência Relativa
Bouteloua gracilis	765	22.1%
Carex spp.	668	19.3%
Aristida longiseta	296	8.6%
Sitanion hystrix	295	8.5%
Stipa comata	290	8.4%
Artemisia frigida	275	7.9%
Buchloe dactyloides	270	7.8%
Sporoboulos cryptandrus	232	6.7%
Ant hill	78	2.3%
Schedonnardus paniculatus	68	2.0%
Eriogonum effusum	55	1.6%
Muhlenbergia torreyi	48	1.4%
BOGR/BUDA	45	1.3%
Ceratoides lanata	17	0.5%
Carex eleocharis	13	0.4%
Carex filifolia	11	0.3%
Aristida spp.	9	0.3%
BOGR/CAEL	9	0.3%
Artemisia dracuncululus	5	0.1%
BOGR/BUDA/CAEL	5	0.1%
Monroa squarrosa	5	0.1%
Artemisia ludoviciana	3	0.09%

Tabela 3 – Número de testes realizados por tipo de pastagem; Frequências absolutas e relativas.

Pastagem	Frequência	Frequência Relativa
ungz	578	33.4%
gzun	455	26.3%
gzzz	372	21.5%
unun	326	18.8%

Tabela 4 – Número de testes realizados por tipo de pasto.

Tipo de Pasto	Frequência	Frequência Relativa
24	557	32.2%
7	346	20.0%
19	336	19.4%
11	210	12.1%
5b	203	11.7%
5a	79	4.6%

Tabela 5 – Número de testes realizados por ano; Frequências absolutas e relativas.

Ano	Frequência	Frequência Relativa
1997	129	7.5%
1998	254	14.7%
1999	236	13.6%
2000	109	6.3%
2001	239	13.8%
2002	142	8.2%
2004	102	5.9%
2005	120	6.9%
2006	70	4.0%
2007	162	9.4%
2008	74	4.3%
2010	94	5.4%

Além destas características, foram computadas para cada espécie de planta a proporção de área da região de estudo ocupada pelas mesmas e o número de polígonos associados à essas plantas observados em cada *quadrat* e em cada ano. Os histogramas associados à estas medidas estão na Figura 18.

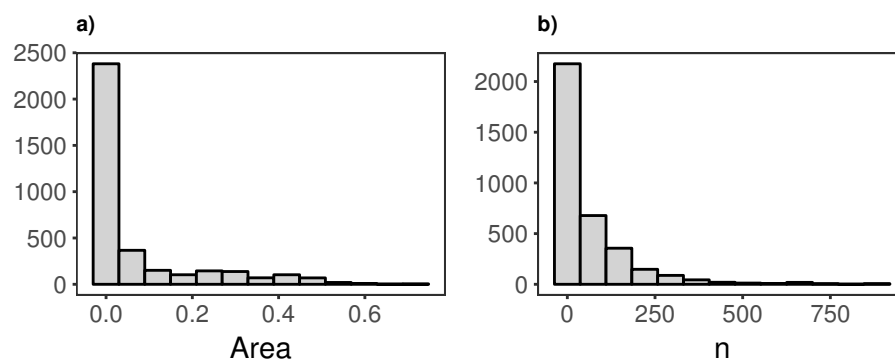


Figura 18 – **a)**: Histograma das áreas das plantas; **b)**: Histograma do número de polígonos das plantas.

A distribuição dos p-valores dos testes de associação espacial entre os pares de espécies é ilustrada através do histograma apresentado na Figura 19. No momento da realização de cada um dos testes de hipóteses foi criada uma variável categórica que recebe três valores, repulsão, atração ou independência/indeterminado. A atribuição destes valores é baseada no *Pointwise Envelope* da seguinte maneira: se na maioria das distâncias o valor de função escolhida para caracterizar as propriedades de segunda ordem do processo estivesse abaixo do percentil 5% da distribuição empírica destes valores, então a relação é considerada de repulsão. Caso os valores observados estivessem acima do percentil 95%, a variável recebe o valor atração. Caso contrário, recebe independência ou indeterminado. Esta abordagem foi adotada pela impraticabilidade de fazer uma análise descritiva dos envelopes para cada um dos 1731 testes realizados. A distribuição das relações ficou a seguinte: em 23.6% dos testes a relação empírica encontrada por este método foi de repulsão, em 6.8% atração e, por fim, 69.6% independência.

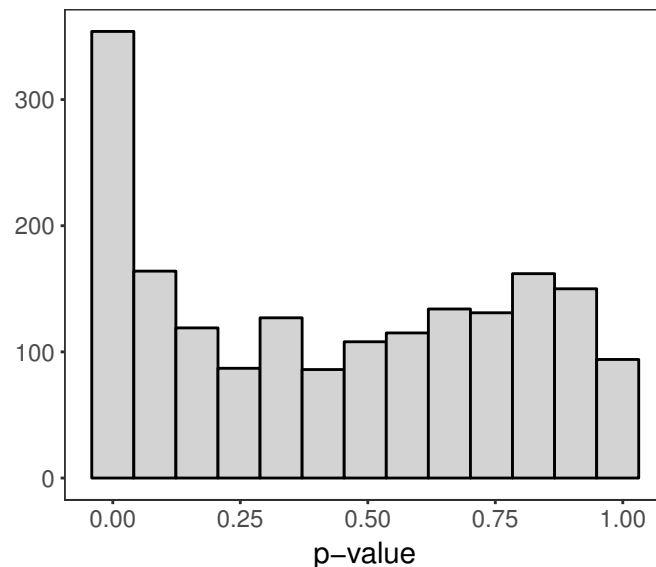


Figura 19 – Histograma dos p valores calculados de acordo com os Testes de Monte Carlo.

6.1 Modelo Misto

Com o objetivo de inferir se alguma espécie de planta tem uma maior propensão a não ser independente das outras, foi ajustado um modelos misto. Antes do ajuste do modelo foram desconsideradas plantas para as quais foram calculados menos de 200 testes, com o objetivo de prevenir uma possível heterogeneidade de variâncias causada pelo desbalanceamento. Com esta redução amostral o número de testes considerados no modelo misto passou a ser 1374 testes.

O modelo ajustado tem como variável resposta o *p-valor* dos testes realizados entre os pares de plantas. As variáveis explicativas foram a *Espécie*, a *Ocorrência de Pastagem*

e um intercepto aleatório para o *Tipo de Pasto*. Dado que os *quadrats* são identificados através combinação do tipo de pasto e da ocorrência de pastagem, esta configuração de modelo foi construída de modo que o efeito de quadrat fosse isolado. Note que, como são pares de espécies o número de observações é correspondente a duas vezes o número de testes realizados, que é 2748.

Na Tabela 6 estão os resultados referentes aos efeitos fixos do modelo. Através da mesma, é possível inferir que pelo menos uma das espécies apresentou um p-valor médio diferente das demais. A mesma conclusão vale para a variável ocorrência de pastagem.

Tabela 6 – Tabela da Anova para os efeitos fixos do modelo.

Coeficiente	GL	p-valor
Espécie	7	<0.0001
Ocorrência de Pastagem	3	0.0191

As comparações múltiplas realizadas utilizando um nível de significância $\alpha = 5\%$, ilustradas na Tabela 7, feitas através do Teste de Tukey para o valor médio marginal dos p-valores de cada uma das espécies, indicam que a espécie *Bouteloua gracilis* apresenta um p-valor médio significativamente menor que mais de 50% das espécies consideradas no estudo. Foi observado, através de uma análise descritiva, que de 117 testes realizados utilizando esta espécie, 98% deles indicaram uma relação de rejeição com a outras espécies utilizadas nos testes. Portanto, existem fortes indícios de que esse tipo de planta tenha tendência a repelir outras espécies de plantas dos seus arredores.

Tabela 7 – Comparações múltiplas entre as espécies segundo o Teste de Tukey, com nível de significância de 5%.

Espécie	Média	EP	Grupo
<i>Bouteloua gracilis</i>	0.314	0.0181	A
<i>Buchloe dactyloides</i>	0.363	0.0255	AB
<i>Stipa comata</i>	0.378	0.0248	AB
<i>Sporobolus cryptandrus</i>	0.394	0.0267	BC
<i>Carex spp.</i>	0.456	0.0187	CD
<i>Sitanion hystrix</i>	0.471	0.0245	CD
<i>Artemisia frigida</i>	0.472	0.0254	CD
<i>Aristida longiseta</i>	0.492	0.0247	D

Quando comparados os p-valores entre os tipos de ocorrências de pastagem, verificou-se que, segundo o Teste de Tukey com 5% de significância, existe diferença entre o p-valor médio da não ocorrência de pastagem (codificada por *unun*) com o *gzun* (ocorrência de pastagem apenas antes de 1991). Este não é um resultado de interpretação trivial.

Tabela 8 – Comparações múltiplas entre os tipos de ocorrência de pastagem segundo o Teste de Tukey, com nível de significância de 5%.

Pastagem	Média	EP	Grupo
unun	0.388	0.0193	A
gzgz	0.417	0.0191	AB
ungz	0.420	0.0179	AB
gzun	0.446	0.0186	B

7 Conclusões

Padrões de polígonos possuem, sem dúvidas, vasta aplicação em diversas áreas da ciência, dentre as quais, se destaca a ecologia. Contudo, a complexidade destas estruturas de dados aliada a falta de metodologias estatísticas adequadas desencorajam pesquisadores à utilizá-los como entidades espaciais que representam seus dados. Fazendo com que em muitas vezes se abdique de informações relevantes, como as formas e tamanhos dos polígonos, para realizar análises. Buscando preencher esta lacuna metodológica, foram desenvolvidos testes de hipóteses Monte Carlo baseados na literatura de processos pontuais espaciais. A principal ferramenta para a construção destes testes foi o *toroidal shift*. Entretanto a utilização deste modelo nulo com padrões de polígonos apresenta um obstáculo até então não contornado na literatura, a quebra de polígonos. Este problema foi contornado através da utilização de identificadores únicos para cada polígono.

Além de um modelo nulo adequado, foram desenvolvidas extensões das funções K cruzada de Ripley e L de Besag, tradicionalmente utilizadas no contexto de processos pontuais. As correções de borda, também utilizadas em processos pontuais, tiveram adaptações para padrões de polígonos desenvolvidas neste trabalho, assim como as *deviation tests* e *goodness of fit tests* que atraíram recentemente a atenção da comunidade estatística no campo de processos pontuais.

Em um estudo de simulação, verificou-se que com exceção de testes que utilizaram estatísticas baseadas na distância euclideana, os testes propostos neste estudo eram de fato de tamanho α . Em outras palavras, possuem o erro do tipo I controlado. Também foi possível mostrar, empiricamente, que quando o assunto é o poder dos testes, estatísticas baseadas em áreas ou na distância de *Hausdorff* apresentam testes mais poderosos do que aquelas baseadas na distância euclideana.

Diferentes modelos nulos também foram comparados no estudo de simulação aqui realizado, tendo sido o modelo nulo proposto neste trabalho sempre melhor, ou pelo menos equivalente, que seu concorrente. Espera-se que, à medida que as estruturas marginais dos processos de interesse se tornem mais complexas, o nosso modelo nulo mostre-se mais superior em relação ao outro modelo aqui considerado. Isso se deve ao fato do *toroidal shift* preservar completamente as estruturas espaciais marginais dos padrões ao gerar dados sob a hipótese nula.

Por fim, ilustrou-se a utilidade e valia deste tipo de metodologia através de dados de espécies de gramíneas curtas em uma estação experimental localizada em Nunn, no estado americano do Colorado.

Referências

- A. Jansen, P., S. A. Bohlman, C. X. Garzon-Lopez, H. Olf, H. C. Muller-Landau, and S. Joseph Wright
2008. Large-scale spatial variation in palm fruit abundance across a tropical moist forest estimated from high-resolution aerial photographs. *Ecography*, 31(1):33–42. [9](#)
- Baddeley, A., P. J. Diggle, A. Hardegen, T. Lawrence, R. K. Milne, and G. Nair
2014. On tests of spatial pattern based on simulation envelopes. *Ecological Monographs*, 84(3):477–489. [12](#), [14](#), [24](#)
- Baddeley, A. and R. Turner
2005. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42. [17](#)
- Chu, C., J. Norman, R. Flynn, N. Kaplan, W. K. Lauenroth, and P. B. Adler
2013. Cover, density, and demographics of shortgrass steppe plants mapped 1997–2010 in permanent grazed and ungrazed quadrats. *Ecology*, 94(6):1435–1435. [4](#), [13](#), [39](#)
- Cullen, L., E. Bodmer, and C. Valladares-Padua
2001. Ecological consequences of hunting in atlantic forest patches, são paulo, brazil. *Oryx*, 35(2):137–144. [9](#)
- Diggle, P.
2003. *Statistical Analysis of Spatial Point Patterns*, Mathematics in biology. Arnold. [12](#), [14](#), [22](#), [24](#)
- Dubuisson, M.-P. and A. K. Jain
1994. A modified hausdorff distance for object matching. In *Proceedings of 12th international conference on pattern recognition*, Pp. 566–568. IEEE. [12](#), [19](#)
- Eddy, W. F.
1977. A new convex hull algorithm for planar sets. *ACM Trans. Math. Softw.*, 3(4):398–403. [29](#)
- Fortin, M.-J., M. R. Dale, and J. M. Ver Hoef
2014. Spatial analysis in ecology. *Wiley StatsRef: Statistics Reference Online*, Pp. 1–13. [9](#)
- Getzin, S., C. Dean, F. He, J. A. Trofymow, K. Wiegand, and T. Wiegand
2006. Spatial patterns and competition of tree species in a douglas-fir chronosequence on vancouver island. *Ecography*, 29(5):671–682. [11](#)

- Grabarnik, P., M. Myllymäki, and D. Stoyan
2011. Correct testing of mark independence for marked point patterns. *Ecological Modelling*, 222(23-24):3888–3894. [12](#), [14](#), [24](#)
- Guo, L., S. Du, R. Haining, and L. Zhang
2013. Global and local indicators of spatial association between points and polygons: a study of land use change. *International Journal of Applied Earth Observation and Geoinformation*, 21:384–396. [9](#), [12](#), [15](#), [20](#), [31](#)
- Hao, H.-M., Z. Huang, R. Lu, C. Jia, Y. Liu, B.-R. Liu, and G.-L. Wu
2017. Patches structure succession based on spatial point pattern features in semi-arid ecosystems of the water-wind erosion crisscross region. *Global ecology and conservation*, 12:158–165. [11](#)
- Huang, Y., S. Shekhar, and H. Xiong
2004. Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and data engineering*, 16(12):1472–1485. [14](#)
- Huttenlocher, D. P., G. A. Klanderman, and W. J. Rucklidge
1993. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863. [19](#)
- Jacquez, G. M.
1995. The map comparison problem: Tests for the overlap of geographic boundaries. *Statistics in Medicine*, 14(21-22):2343–2361. [23](#)
- Jesorsky, O., K. J. Kirchberg, and R. W. Frischholz
2001. Robust face detection using the hausdorff distance. In *International conference on audio-and video-based biometric person authentication*, Pp. 90–95. Springer. [19](#)
- Kenkel, N.
1988. Pattern of self-thinning in jack pine: testing the random mortality hypothesis. *Ecology*, 69(4):1017–1024. [23](#)
- Leemans, R.
1991. Canopy gaps and establishment patterns of spruce (*picea abies* (l.) karst.) in two old-growth coniferous forests in central sweden. *Vegetatio*, 93(2):157–165. [11](#)
- Loosmore, N. B. and E. D. Ford
2006. Statistical inference using the g or k point pattern spatial statistics. *Ecology*, 87(8):1925–1931. [12](#), [14](#), [23](#), [24](#)
- Lotwick, H. and B. Silverman
1982. Methods for analysing spatial processes of several types of points. *Journal of the Royal Statistical Society. Series B (Methodological)*, Pp. 406–413. [12](#), [14](#), [16](#), [17](#), [18](#), [20](#)

- Lu, J., L. Wang, Y. Fang, and M. Li
2017. Mining competitive pairs hidden in co-location patterns from dynamic spatial databases. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Pp. 467–480. Springer. 14
- Manly, B. F.
2006. *Randomization, bootstrap and Monte Carlo methods in biology*, volume 70. CRC press. 16
- McGarigal, K.
2014. Landscape pattern metrics. *Wiley StatsRef: Statistics Reference Online*. 9
- Myllymäki, M., T. Mrkvička, P. Grabarnik, H. Seijo, and U. Hahn
2017. Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):381–404. 12, 14, 24, 25
- Nuske, R. S., S. Sprauer, and J. Saborowski
2009. Adapting the pair-correlation function for analysing the spatial distribution of canopy gaps. *Forest Ecology and Management*, 259(1):107–116. 12, 15
- Prentice, I. and M. Werger
1985. Clump spacing in a desert dwarf shrub community. *Vegetatio*, 63(3):133–139. 11
- Ripley, B. D.
1977. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, Pp. 172–212. 23
- Ripley, B. D.
1979. Tests of randomness' for spatial point patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, Pp. 368–374. 12, 25
- Ripley, B. D.
1981. Spatial statistics. 21
- Schofield, G., V. J. Hobson, M. K. Lilley, K. A. Katselidis, C. M. Bishop, P. Brown, and G. C. Hays
2010. Inter-annual variability in the home range of breeding turtles: implications for current and future conservation management. *Biological Conservation*, 143(3):722–730. 9
- Wang, L., W. Jiang, H. Chen, and Y. Fang
2017. Efficiently mining high utility co-location patterns from spatial data sets with instance-specific utilities. In *International Conference on Database Systems for Advanced Applications*, Pp. 458–474. Springer. 14

Wang, S., Y. Huang, and X. S. Wang

2013. Regional co-locations of arbitrary shapes. In *International Symposium on Spatial and Temporal Databases*, Pp. 19–37. Springer. [14](#), [15](#)

Wiegand, T., W. D. Kissling, P. A. Cipriotti, and M. R. Aguiar

2006. Extending point pattern analysis for objects of finite size and irregular shape. *Journal of Ecology*, 94(4):825–837. [12](#), [15](#), [20](#), [31](#), [37](#)

Worton, B.

1987. A review of models of home range for animal movement. *Ecological modelling*, 38(3-4):277–298. [9](#)