



Universidade de Brasília  
Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Ecologia

ALMIR PICANÇO DE FIGUEIREDO

**Novos métodos em Ecologia de Estradas: Correção da heterogeneidade espacial na análise de agregação de atropelamentos de fauna e definição da suficiência amostral**

BRASÍLIA  
2017

**Novos métodos em Ecologia de Estradas: Correção da heterogeneidade espacial na análise de agregação de atropelamentos de fauna e definição da suficiência amostral**

Almir Picanço de Figueiredo

Dissertação apresentada ao Programa de Pós-Graduação em Ecologia da Universidade de Brasília como parte dos requisitos necessários para a obtenção do título de Mestre em Ecologia.

Orientadora: Ludmilla M. S. Aguiar

BRASÍLIA  
2017

*Se avexe não  
Amanhã pode acontecer tudo  
Inclusive nada*

*Se avexe não  
A lagarta rasteja até o dia  
Em que cria asas*

*Se avexe não  
Toda caminhada começa  
No primeiro passo  
A natureza não tem pressa  
Segue seu compasso  
Inexoravelmente chega lá*

*Se avexe não  
Observe quem vai subindo a ladeira  
Seja princesa ou seja lavadeira  
Pra ir mais alto vai ter que suar*

*(Natureza das Coisas - Flávio José)*

**Ao meu filho, Benjamin. Para que o mundo que ele vê seja sempre melhor do que o meu**

## AGRADECIMENTOS

Agradeço sempre aos meus pais Ademir e Romana por tudo, por formarem meu caráter e por me fazer capaz de me desafiar e superar meus desafios

Agradeço aos meus irmãos Ademir, Marcy e Neto pelo amor, pelo apoio que sempre nos demos, por terem me ajudado a crescer, até hoje inclusive.

Aos meus professores, da Escolinha Pingo de Mel, da Escola Tenente Rêgo Barros, do IB-USP e do PPG de Ecologia da UnB.. Em nome deles cito os professores Samuel e Davi Eduardo que me inspiraram a escolher o estudo da vida como profissão.

Ao pessoal do Rodofauna, Rodrigo, Leandro, Felipe, Marina, Javier, Cecília e Carol, pelo grande apoio nas coletas de campo, pelas boas conversas durante as coletas e pelos debates durante as análises de dados.

Aos amigos da Coordenação de Fauna do Ibram, Ana Nira, Elenize, João Bosco, Thiago, Marina, Rodrigo (de novo) e Fernanda, por formarem o melhor setor para se trabalhar e por me ajudarem a aguentar o peso no mestrado, quebrando vários galhos no trabalho.

Aos amigos da Ecologia da Unb, Marcela, Jéssica, Elba, Marília, Sara, Danilo, Laura, Nayara, Dariane, Camila, Carla, Bárbara, Tarcísio, Silvia, Marcos, Vicente e os que esqueci de citar mas que em diversos momentos deste 02 anos e meio me deram muito apoio, seja com a experiência ou com a juventude me inspiraram a persistir.

A todos os professores do Programa de Pós-Graduação em Ecologia da UnB por todo o aprendizado que obtive e por terem me dado tempo e solidariedade quando eu precisei.

À professora Ludmilla, por ter me aceitado como aluno e ter confiado em mim desde o início mesmo com o oceano Atlântico de distância entre nós. E por terminar todas as nossas conversas deixando sempre claro que confiava em mim.

Ao Ibram por manter um programa de monitoramento de fauna atropelada há tantos anos, sendo o ÚNICO órgão ambiental do Brasil a ter um programa permanente deste tema, mostrando a consciência quanto à relevância do impacto das rodovias sobre a fauna.

E por último, por que o melhor sempre fica para o final, quero agradecer muito, mas muito mesmo a minha esposa Giovanna, que trabalhou muito mais do que eu, durante esses longos 02 anos e meio, para que o resto da minha vida funcionasse enquanto eu me dedicava ao mestrado. Te amo!!! Nunca conseguiria concluir o mestrado sem você!!!

## SUMÁRIO

LISTA DE FIGURAS .....	viii
LISTA DE TABELAS .....	x
RESUMO GERAL.....	1
Palavras chave .....	1
KEY WORDS .....	2
INTRODUÇÃO GERAL .....	3
Agregação espacial na Ecologia de Estradas.....	4
O problema da não homogeneidade na estatística espacial de pontos .....	8
Suficiência amostral em Ecologia de Estradas .....	10
CAPÍTULO 01 - DEALING WITH INHOMOGENEITIES IN ROAD ECOLOGY: A NEW HOTSPOT ANALYSIS .....	17
ABSTRACT .....	18
2. MATERIAL AND METHODS.....	22
2.1. Simulated study .....	22
2.2. Real data study.....	26
3. RESULTS .....	29
3.1 Simulated study .....	29
3.1 Real data study.....	34
4. DISCUSSION.....	37
6. ACKNOWLEDGEMENTS.....	41
7 BIBLIOGRAPHY .....	42
CAPÍTULO 02 – DEFINIÇÃO DA CURVA DE PRECISÃO PARA ANÁLISE DA SUFICIÊNCIA AMOSTRAL EM ECOLOGIA DE ESTRADAS .....	48
RESUMO.....	48
ABSTRACT .....	49
INTRODUÇÃO.....	51
MÉTODOS.....	53
Coleta de dados.....	53

Análise de dados .....	55
RESULTADOS .....	59
DISCUSSÃO .....	64
CONCLUSÃO.....	67
REFERÊNCIAS .....	68
CONSIDERAÇÕES FINAIS.....	72
ANEXO I - SCRIPTS DA PROGRAMAÇÃO REALIZADA PARA A DISSERTAÇÃO	
ANEXO II – LISTA DAS ESPÉCIES	

## LISTA DE FIGURAS

### INTRODUÇÃO GERAL

Figura 1 Diferença entre distribuição de eventos sobre um espaço Euclidiano (a) e sobre ou ao longo de uma rede linear (b). Fonte: Okabe & Sugihara (2012) ..... 5

### CAPÍTULO 01 - DEALING WITH INHOMOGENEITIES IN ROAD ECOLOGY: A NEW HOTSPOT ANALISYS

Figure 1. The large green area is the Águas Emendadas Ecological Station – ESECAE, in Brasília, Federal District of Brazil. .... 27

Figure 2. All graphs show the effects of changes in density and in distribution noise types on the three methods results. Graphs A and B show the results considering all scales of analysis and graphs C and D show those considering only the most efficient analyses for each method, that is, a radius of 200 m for the Hotspot 2D formula and Malo’s method and a radius of 200 m with a window of 2,800 m for the Windowned method. For all graphs, the red triangles, green squares and black circles represent the Hotspot 2D formula, Malo’s method and the Windowned method, respectively ..... 32

Figure 3. Results of the hotspot analysis using the Windowned methodology (graph A), Hotspot 2D (graph B) and Malo’s method (graph C)..... 34

Figure 4. The columns represent the results obtained using the Windowned method, Hotspot 2D and Malo’s method, respectively. Each of the lines represents the removal of a highway from the study area, and in the downward direction, we have the removal of BR-020, followed by DF-345, DF-205 and finally DF-128. The triangles represent original hotspots that were not classified as such by the analysis with the removal of a highway, and the circles are the points not originally identified as hotspotsbut have been classified as such in the analyses..... 36

### CAPÍTULO 02 – DEFINIÇÃO DA CURVA DE PRECISÃO PARA ANÁLISE DA SUFICIÊNCIA AMOSTRAL EM ECOLOGIA DE ESTRADAS

Figura 1 A grande área preenchida em verde é a Estação Ecológica Águas Emendadas – ESECAE. ....	54
Figura 2. Análise da correlação entre a taxa de Precisão na análise de um em um subconjunto e o logaritmo da sua abundância, a linhas em vermelho representam a regressão para o Q50 e as linhas em azul são para os Q2,5 e Q97,5.....	61
Figura 3 Curva de Precisão na indicação de <i>hotspots</i> de atropelamento de fauna para os quatro grupos estudados. As linhas sólidas, pontilhadas e tracejadas representam, respectivamente, os cenários Q2.5, Q50 e Q97.5.....	62
Figura 4 Valores da Precisão de indicação de <i>hotspots</i> para o grupo “Todos” se a curva de precisão fosse elaborada em momentos diferentes da coleta de dados. Os números no eixo x representam o tamanho da amostra no momento da confecção de cada curva. As linhas sólidas, pontilhadas e tracejadas representam, respectivamente, os cenários Q2.5, Q50 e Q97.5, e a linha sólida vermelha indica o valor de precisão igual a 0.7. ....	63

## LISTA DE TABELAS

### CAPÍTULO 01 - DEALING WITH INHOMOGENEITIES IN ROAD ECOLOGY: A NEW HOTSPOT ANALYSIS

Table 1 Number of hotspots that ‘appeared’ - False Positive Hotspot -and ‘disappeared’ - False Negative Hotspot - due to the withdrawal of a highway in the analysis of hotspots.....	35
---	----

### CAPÍTULO 02 – DEFINIÇÃO DA CURVA DE PRECISÃO PARA ANÁLISE DA SUFICIÊNCIA AMOSTRAL EM ECOLOGIA DE ESTRADAS

Tabela 1 Padrões de divisões do total de dados para criação dos subconjuntos .....	56
Tabela 2 Conjuntos de dados utilizados para avaliar a construção da curva de precisão de tempos diferentes da coleta de dados.....	56
Tabela 3 Estimativa da precisão atual da indicação de <i>hotspots</i> de atropelamento de fauna em consideração ao número de registros de cada grupo.....	62

## RESUMO GERAL

O atropelamento de fauna é considerado por diversos autores como a principal causa direta de morte de animais na natureza. No entanto, as intervenções necessárias para mitigar este efeito negativo das rodovias são geralmente onerosas e por isso é preciso ter confiabilidade na proposição de locais para as mesmas. As análises de agregação de atropelamento usadas em Ecologia de Estrada não corrigem o efeito da heterogeneidade de densidade de primeira ordem gerando uma autocorrelação espacial entre os atropelamentos maior do que a real. O primeiro capítulo desta dissertação apresenta um método de correção, denominado *Windowed Method*, o qual pondera os resultados obtidos em um raio de análise por uma janela de observação com menor heterogeneidade que a área total de estudo. O método proposto apresentou menores taxas de erro de classificação de *hotspot* e sofreu menos influência da heterogeneidade de distribuição dos eventos, quando comparado com dois outros métodos usados em Ecologia de Estradas. Esta dissertação também abordou a deficiência de uma análise de suficiência amostral para estudos de Ecologia de Estradas. No segundo capítulo desta dissertação foi verificado que existe relação positiva entre a precisão de classificação de *hotspot* e o tamanho da amostra. Por meio de simulações por reamostragem *Bootstrap* e por extrapolações utilizando Regressão Quantílica, esta relação foi utilizada para construção de uma Curva de Precisão com a qual é possível identificar o tamanho da amostra desejada para se atingir um grau de precisão determinado pelo pesquisador. Com este método a suficiência amostral pode ser determinada pelo acúmulo de registro permitindo protocolos de coletas diversos, permitindo ao pesquisador variar a velocidade de busca por carcaças e frequência de campanhas de coletas, conforme sua conveniência.

## PALAVRAS CHAVE

Agregação Virtual, Atropelamento, Heterogeneidade, Suficiência Amostral

## **ABSTRACT**

Roadkills is considered by several authors as the main direct cause of death of animals in nature. However, the interventions needed to mitigate this negative effect of the highways are usually onerous and therefore it is necessary to have reliability in proposing sites for them. The roadkill aggregation analyzes used in Road Ecology do not correct the effect of first order density heterogeneity by generating spatial autocorrelation between events greater than the actual is. The first chapter of this dissertation presents a method of correction, called Windowed Method, which weighs the results obtained in a radius of analysis by an observation window with less heterogeneity than the total area of study. The proposed method presented lower hotspot classification error rates and was less influenced by the heterogeneity of event distribution when compared to two other methods used in Road Ecology. This dissertation also addressed the deficiency of a sample adequacy analysis for Road Ecology studies. In the second chapter of this dissertation it was verified that there is a positive relationship between the hotspot classification accuracy and the sample size. By means of Bootstrap resampling simulations and extrapolations using Quantile Regression, this relation was used to construct a Precision Curve with which it is possible to identify the desired sample size to reach a degree of precision determined by the researcher. With this method the sampling sufficiency can be determined by the accumulation of record allowing diverse collection protocols, it permits researchers to vary both the speed of carcasses search as the frequency of collections campaigns, according to their convenience.

## **KEY WORDS**

Heterogeneity, Roadkill, Sample adequacy, Virtual Aggregation.

## INTRODUÇÃO GERAL

O termo "ecologia da estrada" foi proposto por Richard T. T. Forman em 1998 (Coffin 2007). O termo se refere a um assunto de investigação ecológica com base na evidência de que as estradas exercem efeitos sobre componentes, processos e estruturas do ecossistema, e que as causas desses efeitos são tanto relacionadas à engenharia quanto ao planejamento do uso da terra e à política de transportes (Coffin 2007). Estes efeitos incluem a degradação do habitat (Jaeger & Fahrig 2004, Taylor & Goldingay 2012), poluição química e sonora, erosão e sedimentação dos corpos hídricos (Trombulak & Frissell 2000), mudança no comportamento de algumas espécies (Blackwell *et al.* 2014, Lima *et al.* 2014, DeVault *et al.* 2015) e a dispersão de espécies exóticas (Trombulak & Frissell 2000). No entanto, a colisão com automóveis é a principal fonte de mortalidade direta em algumas populações animais (Trombulak & Frissell 2000, Gibbs & Shriver 2002, Glista *et al.* 2008).

As estratégias para mitigação do impacto dos atropelamentos de fauna são normalmente muito caras (Huijser *et al.* 2009, Mountrakis & Gunson 2009, Santos *et al.* 2015), não sendo viável a implantação ao longo de toda rodovia estudada. Por isso, para identificar trechos da rodovia onde existe maior probabilidade de sua ocorrência e dessa forma obter maior benefício para as populações de fauna impactadas, a partir da aplicação dos recursos disponíveis, é necessário conhecer o padrão espacial da distribuição de atropelamentos (Huijser *et al.* 2009, Polak *et al.* 2014, Santos *et al.* 2015).

A análise do padrão de distribuição de atropelamentos em Ecologia de Estradas pode ser dividida em dois grupos distintos: o primeiro grupo é representado por estudos que procuram modelar o padrão de eventos em função de variáveis de paisagem ou de tráfego (ex. Nielsen *et al.* 2003, Grovenburg *et al.* 2008, Roger & Ramp 2009, Grilo *et*

al. 2014, Snow *et al.* 2014), e o segundo grupo por aqueles que visam identificar áreas geográficas com incidência de atropelamento maior do que o esperado ao acaso, ou seja, os *hotspots* de atropelamento (Clevenger *et al.* 2003, Ramp *et al.* 2005, Gomes *et al.* 2008, Seo *et al.* 2013, Skórka *et al.* 2015).

#### *Agregação espacial na Ecologia de Estradas*

O primeiro trabalho que demonstrou que atropelamentos de fauna não ocorrem aleatoriamente foi realizado por Puglisi *et al.* (1974). Nesse trabalho foi observado que a quantidade de atropelamentos de cervídeos era menor em locais onde as cercas no entorno das rodovias estavam a mais de 23 metros de áreas arborizadas. Porém, a presença de vegetação somente tinha relação com atropelamentos nos locais onde não havia cercas. Entretanto, o primeiro trabalho a descrever e medir o padrão espacial de agregação ocorreu em 2003, no qual os autores apresentaram uma adaptação da função K de Ripley (Ripley 1976) para eventos pontuais que ocorrem sobre rodovias (Clevenger *et al.* 2003). A adaptação proposta por Clevenger *et al.* (2003) ocorreu de forma independente e praticamente simultânea ao trabalho de Okabe & Yamada (2001), que propôs adaptação semelhante para eventos denominados por eles de *eventos sobre ou ao longo de redes lineares* ou *eventos em rede* (*network events*).

Para análise das propriedades de distribuição de eventos em rede é importante compreender a diferença entre os dois tipos de processos. Uma mesma distribuição de eventos pode ser considerada não randômica (figura 1-a) e randômica (figura 1-b) quando os eventos são avaliados sobre uma área ou sobre uma rede linear. Os pontos selecionados no plano (figura 1-a) inevitavelmente formam agregações. Isto ocorre pelo fato destes eventos terem sido aleatoriamente plotados somente sobre as linhas da rede (figura 1-b).

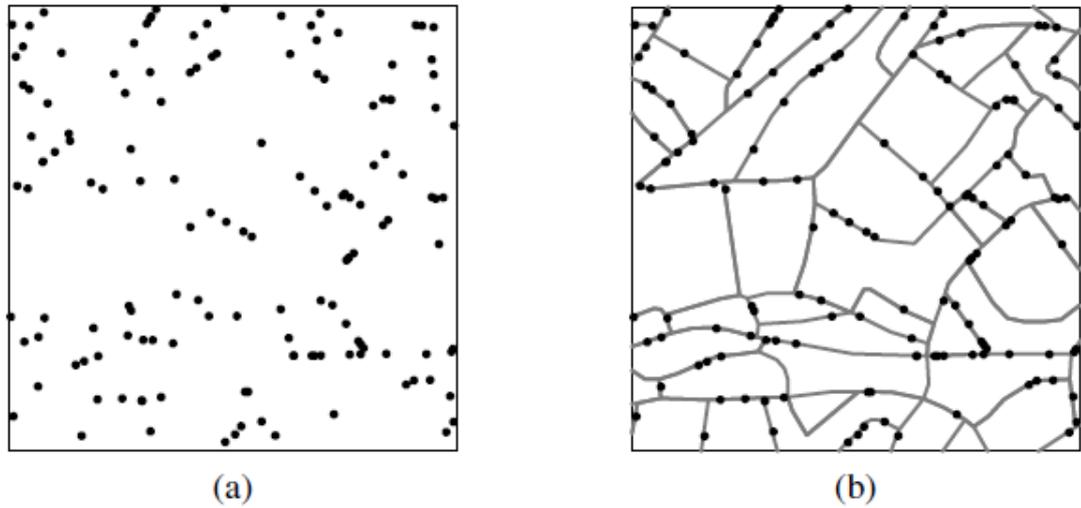


Figura 1 Diferença entre distribuição de eventos sobre um espaço Euclidiano (a) e sobre ou ao longo de uma rede linear (b). Fonte: Okabe & Sugihara (2012)

A análise da não aleatoriedade da agregação de eventos compara os resultados observados com aqueles obtidos por aleatorizações e estes eventos simulados podem ocorrer em qualquer localização do espaço euclidiano em questão. Assim sendo, ao se comparar os eventos que somente ocorreram sobre determinadas linhas com os aleatorizados sobre todo o plano, o valor de K não representaria a realidade de dependência entre os eventos (Okabe & Sugihara, 2012). A correção deste problema é o uso do caminho mais curto sobre a rede, ao invés de distância Euclidiana, para medir a distância entre dois eventos na aplicação da função K (Okabe & Yamada, 2001). Esta correção foi denominada Função K para Rede (*Network K Function*) – fórmula (1) – e os autores denominaram a função tradicionalmente utilizada como Função K planar (*Planar K Function*) para distinguir ambas fórmulas.

$$\hat{K}_L(r) = \frac{L}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n 1\{d_L(x_i, x_j) \leq r\} \quad (1)$$

Onde  $L$  significa o comprimento total da rede linear estudada,  $r$  é o raio da circunferência de análises e  $d_L(x_i, x_j)$  significa a menor distância entre  $i$  e  $j$ , que tem valor de 1 se for menor ou igual ao raio  $r$  e valor 0 se for maior.

Um outro estudo comparou os resultados obtidos com a aplicação das funções  $K$  para redes e planar sobre o padrão de acidentes de trânsito (Yamada & Thill 2004). Além disso, incluíram nesta comparação uma metodologia intermediária: o uso da distância euclidiana entre os eventos, conforme a função planar, comparando com eventos aleatorizados sobre a rede em análise, conforme a função  $K$  para redes. Neste estudo, os autores identificaram uma superestimativa de dependência entre eventos tanto para metodologia intermediária, como para a função  $K$  planar - esta última apresentando maior diferença.

Steenberghen *et al.*(2010) denominaram como redes bi-dimensionais (2D Network) aquelas redes cujos movimentos não estão restritos aos trechos lineares da rede. Para estes casos Okabe & Sugihara (2012) ressaltam que mesmo os eventos ocorrendo sobre, ou ao longo de redes, o uso do caminho mais curto sobre a rede não é o mais adequado para medir a distância entre pares de eventos.

O atropelamento de fauna é um exemplo de processo pontual que ocorre em rede 2D, visto que apesar dos veículos trafegarem exclusivamente nos trechos lineares das redes (rodovias) os animais se deslocam em todo o plano no qual a rede está disposta. Esta característica repercute na adequada aplicação das funções de estatística espacial para pontos, inclusive as adaptadas para redes. Coelho *et al.*(2008) propuseram a adaptação para a função  $K$ , na qual as distâncias são analisadas no espaço euclidiano, mas as aleatorizações para o modelo nulo somente selecionam pontos sobre a rede, analisada da mesma forma que Yamada & Thill (2004) haviam teorizado como metodologia intermediária entre as funções linear e planar. A função  $K$  examina se uma

determinada distribuição de pontos difere de uma distribuição aleatória. Porém, não revela a localização das agregações dentro da distribuição (Steenberghen *et al.* 2010). Em Ecologia de Estrada, alguns autores usam essa função – adaptada por Okabe & Yamada (2001) e depois por Coelho *et al.*(2008) – como uma etapa anterior na identificação de pontos com taxas de acidentes superiores aos esperados, denominado *hotspots* de atropelamento (ex. Ramp *et al.* 2005, Mountrakis & Gunson 2009, Coelho *et al.* 2012). A análise de *Hotspot* 2D, proposta por Coelho *et al.*(2012) é uma adaptação da função K que usa janelas de varredura para testar se a intensidade dos pontos dentro de uma destas janelas é significativamente maior do que esperado ao acaso (Coelho *et al.* 2014).

Tanto a função K como a análise de *Hotspot* 2D testam a completa aleatoriedade espacial. Os testes de função K avaliam se os pontos exibem agregação ou dispersão, em vez de independência, enquanto a análise de *Hotspot* 2D assume que os pontos são independentes e testam se existem regiões com maior intensidade que o esperado (Coelho *et al.* 2014).

No entanto, outros métodos de identificação de *hotspots* foram desenvolvidos ao longo dos anos. Entre eles podemos citar o método Malo, um dos mais utilizados para definir *hotspots* de atropelamento (ex. Malo *et al.* 2004, Grilo *et al.* 2009, Santos *et al.* 2015). Uma vez que o método Malo exigiria menores extensões de mitigação para evitar um maior número de mortes, Gomes *et al.* (2008) sugeriram que o método Malo deveria ser o preferido para a identificação de *hotspot*, em comparação com os outros quatro métodos: regressão logística binária (BLR), análise do fator de nicho ecológico (ENFA), estimativa da densidade do Kernel (KDE) e agrupamento hierárquico pelo vizinho mais próximo (NNHC).

O método Malo, assim como o *Hotspot 2D*, compara a quantidade de registros em uma janela de análise contra o esperado por uma distribuição aleatória de Poisson. As diferenças entre os dois métodos são: 1) as janelas do método Malo não se sobrepõem, ao contrário das 'janelas deslizantes' usadas no *Hotspot 2D*; 2) O método *Hotspot 2D* usa o raio em torno de um ponto fixo e distância euclidiana para registrar como regra de contagem, enquanto o método Malo usa distâncias de estrada; 3) O *Hotspot 2D* corrige os dados registrados pelos trechos rodoviários presentes no raio de análise, o que não é necessário no método Malo, pois todos os trechos de análise possuem a mesma extensão rodoviária. Assim, a principal semelhança entre estes dois métodos é a utilização de um modelo nulo, baseado em distribuição de Poisson, para identificar quais trechos das rodovias apresentam valores superiores ao esperado, com base na completa aleatoriedade espacial dos eventos. Além disso, ambos os métodos consideram como premissa para suas aplicações a homogeneidade da distribuição dos eventos com padrões pontuais.

#### *O problema da não homogeneidade na estatística espacial de pontos*

As distribuições espaciais de eventos pontuais são classificadas em função dos efeitos de primeira ordem, que determinam a intensidade dos eventos, e os de segunda ordem, que determinam a interação entre eventos. Quanto a intensidade de eventos, ela pode ser constante em toda área de estudo, ou variar em função do tamanho da vizinhança estudada. Quanto à interação entre eventos, ela pode ser nula, quando a ocorrência de um evento independe da ocorrência de um outro; homogênea, quando a regra de dependência entre dois eventos é a mesma em toda a região; ou heterogênea, quando existe dependência entre pontos e esta regra de agregação ou dispersão é diferente para diferentes locais na região de estudo (Wiegand & Moloney 2014).

A combinação destes efeitos gera seis (6) classes de distribuição (Wiegand & Moloney 2014). No entanto, esta dissertação abordará duas classificações, denominadas: “Processo pontual homogêneo com interações” e “Processo pontual heterogêneo de primeira ordem”. Um processo pontual homogêneo com interações é a classe alvo da maioria das análises de padrões pontuais. Segundo esta distribuição, um ponto pode ocorrer em qualquer lugar na área de estudo com a mesma probabilidade (intensidade constante). Além disso, por esta distribuição, existe interação entre os eventos, e a regra de interação é a mesma em qualquer local da área de estudo (interação homogênea). Para estas análises, utiliza-se a distribuição de Poisson como modelo nulo. Já um processo pontual heterogêneo de primeira ordem também apresenta interação homogênea entre os eventos, mas nesse caso, os fatores externos exercem influência na intensidade de eventos. Ou seja, a probabilidade de ocorrência de um evento não é a mesma para diferentes locais da área de estudo (Wiegand & Moloney, 2014).

Uma das principais questões a ser respondida por análises de padrões pontuais é se a dependência ambiental contribui ou não para uma distribuição observada. A seleção da técnica analítica é importante. Análises de padrões heterogêneos com ferramentas desenhadas para padrões homogêneos podem gerar resultados distorcidos (Wiegand & Moloney 2014). Tanto as adaptações da Função K utilizadas em Ecologia de Estradas (Clevenger *et al.* 2003, Coelho *et al.* 2012) como o método de Malo (Malo *et al.* 2004) não levam em consideração a influência da heterogeneidade ambiental no padrão de distribuição de atropelamentos de fauna. Neste sentido, Baddeley *et al.* (2000) propuseram uma correção para a aplicação da função K em processos pontuais não estacionários. Nesta correção, a densidade local de cada ponto dentro da vizinhança seria usada para ponderar o valor de K. De forma semelhante à relatada por Baddeley *et al.* (2000) para a função K de Ripley, Ang *et al.* (2012) indicam que a adaptação

proposta por Okabe & Yamada (2001) precisa ser corrigida para distribuições não homogêneas de eventos. Dentro desse escopo, o capítulo 1 desta dissertação apresenta uma proposta de correção metodológica para o método *Hotspot 2D* com a utilização de janelas de observação, de forma que a densidade local seja ponderada pela densidade de uma vizinhança com maior homogeneidade do que toda a área de estudo. O referido capítulo está formatado como artigo científico submetido à apreciação para publicação.

### *Suficiência amostral em Ecologia de Estradas*

Quando uma amostra é tomada de um universo amostral, não é possível saber se o estado de um atributo obtido coincide com o seu estado verdadeiro (Pillar 2004). Porém, quanto maior o tamanho da amostra, maior é a chance de obter novas amostras que indiquem as mesmas conclusões (Pillar 2004). O estado de um dado atributo obtido a partir da amostra evolui e atinge estabilidade a medida que o número de unidades amostrais na amostra aumenta. O incremento de unidades amostrais implica em alterações relativamente menores no valor do atributo considerado (Pillar 2004). Assim, o tamanho suficiente da amostra será aquele em que o atributo da amostra atinge estabilidade (Pillar 2004).

A precisão desejada pode ser definida atribuindo-se a quantidade de erro tolerável nas estimativas da amostra (Cochran 1977). Esta quantidade deve ser determinada de acordo com os objetivos para os quais os resultados da amostra são utilizados (Cochran 1977). Para definir as necessidades mínimas de dados para a precisão adequada, deve-se entender tanto a taxa em que a precisão aumenta com o incremento de dados, como a máxima precisão possível pelo método (Stockwell & Peterson 2002).

A curva "número de espécies *versus* número de unidades amostrais" é usada para indicar a suficiência de amostragem em ecologia de comunidades, mas quaisquer outros atributos, simples ou complexos (e.g., medidas de diversidade), poderiam também ser considerados nessas curvas (Pillar 2004). No entanto, em Ecologia de Estradas não se conhece a relação entre o tamanho da amostra e as respostas que se quer obter com a sua análise.

Duas importantes fontes de erro amostral nos estudos de ecologia de estradas são amplamente conhecidas na literatura especializada: o tempo em que a carcaça fica disponível nas rodovias para ser registrada (Slater 2002, Teixeira *et al.* 2013a), e a ineficiência de detecção das carcaças nas rodovias, intrínseco ao método de coleta utilizado (Teixeira *et al.* 2013a, Santos *et al.* 2016). Todos estes autores indicam a necessidade de se corrigir estatisticamente as taxas de atropelamento. Entretanto, não foram encontrados estudos que demonstrem o impacto da correção da taxa de atropelamento na distribuição espacial da ocorrência dos mesmos.

Os estudos em Ecologia de Estradas apresentam alto custo associado à coleta de dados, pois a baixa eficiência, intrínsecas às técnicas de coletas de dados de atropelamento (ver Slater 2002, Teixeira *et al.* 2013, Santos *et al.* 2016) resulta na necessidade de se realizar estudos com alta frequência de coletas, e com longo tempo de duração. Neste sentido, o segundo capítulo desta dissertação utiliza o método, proposto no primeiro capítulo, para estabelecer um procedimento que avalie a relação entre a precisão na identificação dos *hotspots* com o tamanho da amostra, visando definir a suficiência amostral em estudos de agregação de atropelamentos de fauna.

## Referências

- Ang Q.W., A. Baddeley & G. Nair (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics* 39: 591–617.
- Baddeley A.J., J. Møller & R. Waagepetersen (2000). Non and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54: 329–350.
- Blackwell B.F., T.W. Seamans & T.L. Devault (2014). White-tailed deer response to vehicle approach : evidence of unclear and present danger. *PLoS One* 9: e109988. [doi:10.1371/journal.pone.0109988](https://doi.org/10.1371/journal.pone.0109988).
- Clevenger A.P., B. Chruszcz & K.E. Gunson (2003). Spatial patterns and factors influencing small vertebrate fauna road-kill aggregations. *Biological Conservation* 109: 15–26.
- Cochran W.G. (1977). *Sampling Techniques*. Third Edition. John Wiley and Sons Ltd.
- Coelho A.V.P., I.P. Coelho, F.Z. Teixeira & A. Kindel (2014). Siriema: road mortality software. User's guide. NERF/UFRGS, Porto Alegre, Brasil.
- Coelho I.P., A. Kindel & A.V.P. Coelho (2008). Roadkills of vertebrate species on two highways through the Atlantic Forest Biosphere Reserve, southern Brazil. *European Journal of Wildlife Research* 54: 689–699.
- Coelho I.P., F.Z. Teixeira, A.V.P. Coelho & A. Kindel (2012). Anuran road-kills neighboring a peri-urban reserve in the Atlantic Forest, Brazil. *Journal of Environmental Management* 112: 17–26.
- DeVault T.L., B.F. Blackwell, T.W. Seamans, S.L. Lima & E. Fernández-Juricic (2015). Speed kills: Ineffective avian escape responses to oncoming vehicles.

- Proceedings of Royal Society B, p. 282. <http://dx.doi.org/10.1098/rspb.2014.2188>.
- Forman R.T.T. & L.E. Alexander (1998). Roads and their major ecological effects. *Annual Review of Ecology and Systematics* 29: 207–231.
- Gibbs J.P. & W. G. Shriver (2002). Estimating the effects of road mortality on turtle populations. *Conservation Biology* 16: 1647–1652.
- Glista D.J., T.L. DeVault & J.A. DeWoody (2008). Vertebrate road mortality predominantly impacts amphibians. *Herpetological Conservation and Biology* 3: 77–87.
- Gomes L., C. Grilo, C. Silva & A. Mira (2008). Identification methods and deterministic factors of owl roadkill hotspot locations in mediterranean landscapes. *Ecological Research* 24: 355–370.
- Grilo C., J.A. Bissonette & M. Santos-Reis (2009). Spatial–temporal patterns in mediterranean carnivore road casualties: consequences for mitigation. *Biological Conservation* 142: 301–313.
- Grilo C., D. Reto, J. Filipe, F. Ascensão & E. Revilla (2014). Understanding the mechanisms behind road effects: linking occurrence with road mortality in owls. *Animal Conservation* 17: 555–564.
- Grovenburg T.W., J.A. Jenks, K.L. Monteith, D.H. Galster, R.J. Schauer, W.W. Morlock & J. A. Delger (2008). Factors affecting road mortality of whitetailed deer in eastern South Dakota. *Human-Wildlife Interactions* 2: 48–59.
- Huijser M. P., J. W. Duffield, A. P. Clevenger, R. J. Ament & P. T. McGowen (2009). Cost–benefit analyses of mitigation measures aimed at reducing collisions with large ungulates in the United States and Canada; a decision support tool. *Ecology and Society* 14(2): 15.
- Jaeger J.A.G. & L. Fahrig (2004). Effects of road fencing on population persistence.

- Conservation Biology 18: 1651–1657.
- Lima S.L., B.F. Blackwell, T.L. Devault & E. Fernández-Juricic (2014). Animal reactions to oncoming vehicles: a conceptual review. *Biological Reviews* 10.1111/brv.12093.
- Malo J.E., F. Suárez & A. Díez (2004). Can we mitigate animal–vehicle accidents using predictive models? *Journal of Applied Ecology* 41: 701–710.
- Mountrakis G. & K. Gunson (2009). Multi-scale spatiotemporal analyses of moose–vehicle collisions: a case study in northern Vermont. *International Journal of Geographical Information Science* 23: 1389–1412.
- Nielsen C.K., R.G. Anderson & M.D. Grund (2003). Landscape influences on deer–vehicle accident areas in an urban environment. *Journal of Wildlife Management* 67: 46.
- Okabe, A. & K. Sugihara (2012). *Spatial analysis along networks: Statistical and computational methods*. John Wiley and Sons Ltd.
- Okabe A. & I. Yamada (2001). The K-Function Method on a network and its computational implementation. *Geographical Analysis* 33: 270–290.
- Pillar V. D. P. (2004). *Suficiência amostral. Amostragem em Limnologia*. pp. 25–43. RIMA, São Carlos.
- Puglisi M. J., J.S. Lindzey & E.D. Bellis (1974). Factors associated with highway mortality of white-tailed deer. *The Journal of Wildlife Management* 38: 799–807.
- Ramp D., J. Caldwell, K.A. Edwards, D. Warton & D. B. Croft (2005). Modelling of wildlife fatality hotspots along the Snowy Mountain Highway in New South Wales, Australia. *Biological Conservation* 126: 474–490.
- Ripley B. D. (1976). The Second-Order Analysis of stationary point processes. *Journal of Applied Probability* 13: 255–266.

- Roger E. & D. Ramp (2009). Incorporating habitat use in models of fauna fatalities on roads. *Diversity and Distributions* 15: 222–231.
- Santos S. M., J.T. Marques, A. Lourenço, D. Medinas, A.M. Barbosa, P. Beja & A. Mira (2015). Sampling effects on the identification of roadkill hotspots: Implications for survey design. *Journal of Environmental Management* 162: 87–95.
- Santos R.A.L., S.M. Santos, M. Santos-Reis, A. Picanço de Figueiredo, A. Bager, L.M.S. Aguiar & F. Ascensão (2016). Carcass persistence and detectability: Reducing the uncertainty surrounding wildlife-vehicle collision surveys. *PLoS One* 11: 1–15.
- Seo C., J. H. Thorne, T. Choi, H. Kwon & C.H. Park (2013). Disentangling roadkill: the influence of landscape and season on cumulative vertebrate mortality in South Korea. *Landscape and Ecological Engineering* 11: 87–99.
- Skórka P., M. Lenda, D. Moroń, R. Martyka, P. Tryjanowski, W. J. Sutherland, D. Moron, R. Martyka, P. Tryjanowski & W. J. Sutherland (2015). Biodiversity collision blackspots in Poland: Separation causality from stochasticity in roadkills of butterflies. *Biological Conservation* 187: 154–163.
- Slater F.M. (2002). An assessment of wildlife road casualties – the potential discrepancy between numbers counted and numbers killed. *Web Ecology* 3: 33–42.
- Snow N.P., D. M. Williams & W. F. Porter (2014). A landscape-based approach for delineating hotspots of wildlife-vehicle collisions. *Landscape Ecology* 29: 817–829.
- Steenberghen T., K. Aerts & I. Thomas (2010). Spatial clustering of events on a network. *Journal of Transport Geography* 18: 411–418.
- Stockwell D.R.B. & A. T. Peterson (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1–13.

- Taylor B.D. & R.L. Goldingay (2012). Restoring connectivity in landscapes fragmented by major roads: A case study using wooden poles as 'stepping stones' for gliding mammals. *Restoration Ecology* 20: 671–678.
- Teixeira F.Z., A.V.P. Coelho, I.B. Esperandio & A. Kindel (2013). Vertebrate road mortality estimates: Effects of sampling methods and carcass removal. *Biological Conservation* 157: 317–323.
- Trombulak S.C. & C.A. Frissell. 2000. Review of ecological effects of roads on terrestrial and aquatic communities. *Conservation Biology* 14: 18–30.
- Wiegand T. & K.A. Moloney. 2014. *Handbook of spatial point-pattern analysis in ecology*. CRC Press, Boca Raton.
- Yamada I. & J.C. Thill. 2004. Comparison of planar and network K-functions in traffic accident analysis. *Journal of Transport Geography* 12: 149–158

## CAPÍTULO 01 - DEALING WITH INHOMOGENEITIES IN ROAD ECOLOGY: A NEW HOTSPOT ANALYSIS

Este capítulo está formatado de acordo com o periódico *Methods in Ecology and Evolution*.

***Running title:*** *Dealing with inhomogeneities in road ecology*

***Dealing with inhomogeneities in road ecology: a new hotspot analysis***

Almir Picanço de Figueiredo<sup>a,b,c,\*</sup>, Ludmilla Moura de Souza Aguiar<sup>a,b</sup>

<sup>a</sup>Curso de Pós-Graduação em Ecologia, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF, Brazil

<sup>b</sup>Laboratório de Biologia e Conservação de Morcegos, Departamento de Zoologia, Universidade de Brasília, Brasília, DF, Brazil

<sup>c</sup>Instituto Brasília Ambiental (IBRAM), Brasília, DF, Brazil

\*Corresponding author. E-mail address: almir.ibram@gmail.com

*ABSTRACT*

Determining roadkill hotspots is essential in identifying mitigation measures, although some studies note uncertainties regarding their use. Nevertheless, the methods usually used in road ecology infer homogeneity in notoriously heterogeneous distributions. A heterogeneous density of points may cause what the recent literature has called "virtual aggregation." Thus, the purpose of this study was to evaluate whether the study of roadkill hotspots in subdivisions of a highway of interest is less sensitive to the first-order effects of an inhomogeneous event distribution. To address this problem, we propose an adaptation of the Hotspot 2D method that weights the value found in each analysis radius 'r' by that found in an observation window with a radius 'w' of sufficient size to represent a minimally homogeneous window. We applied Hotspot 2D, Malo's method and the proposed method, called Windowed Hotspot, to twenty different types of simulated point patterns and also, to an experiment using four different roads with the removal of data from one highway at a time. In the first experiment, we observed that the Windowed method has more accuracy in hotspot identification and less sensitivity to the regional intensity distribution of events. In the second experiment, we observed that the intensity of events along one road can determine the classification of hotspots for another, even both being distant and having different characteristics, for all methods. The second experiment reveals the need to broaden the studies of heterogeneous null models or corrections in the metrics used in road ecology. The foremost benefit of this new method is its capability to identify smaller stretches with intensities greater than expected in the interior of a zone with a high intensity of roadkill, which improves the process of choosing sites for mitigation.

**Keywords:** Virtual Aggregation, Hotspot 2D, Roadkill, heterogeneity.

## 1. INTRODUCTION

A collision between a wild animal and a vehicle is a process comparable to other point processes, such as traffic accidents, crimes, and cases of epidemic diseases or extreme weather events. Knowing the aggregation patterns of these events is one of the main challenges for researchers of this area (Miller & Han, 2009, Liu et al., 2015). To determine roadkill aggregation, defined in road ecology studies as hotspots, it is essential to develop a more robust approach for predictive models of wildlife-vehicle-collisions, and to define mitigation measures (e.g., Beaudry et al. 2008, Malo et al. 2004). It is also necessary to develop a more robust approach to distributions and patterns when combined with predictive models (Ramp et al. 2005).

However, hotspots may not be indicated when past mortality reduces surrounding populations (Eberhardt et al. 2013) or when affected specimens are from spatially separated populations (Teixeira et al. 2017). Additionally, the variations in hotspot locations over time (Mountrakis & Gunson 2009, Seo et al. 2013, Barrientos & Plaza 2016, Seiler et al. 2016) and reductions in sampling efforts (Costa et al. 2015, Santos et al. 2015) could give rise to uncertainty regarding the use of hotspots for the selection of mitigation sites. Notwithstanding, hotspots can be reliable about their stability over time when larger study scales are used (Lima Santos et al. 2017).

To use aggregation patterns in predicting future events, one must have confidence in the robustness of the analytical method. Methods that infer the homogeneity of the studied spaces are inappropriate in analyzing non-stationary sets of points (Marcon & Puech 2009). These methods consider as second-order effects (conditional density) the mere reflection of a first-order effect (heterogeneity), indicating a stronger positive autocorrelation than exists (Schiffers et al., 2008).

Although methods for dealing with heterogeneity in spatial point pattern analysis has been developed recently (e.g., Baddeley, Møller & Waagepetersen, 2000, Wiegand & Moloney, 2004, 2014, Schiffers *et al.*, 2008, Ang, Baddeley & Nair, 2012), allowing for the exploration of certain inhomogeneous point pattern classes, no adaptation has proposed hotspot identification of inhomogeneous events distributed on, or along, a linear network. There are four possibilities for dealing with inhomogeneity: (1) we can ignore it; (2) we can avoid it by selecting observation windows that omit heterogeneities; (3) we can adapt the null model to model the heterogeneity explicitly; and (4) we may factor out the effect of the heterogeneity by modifying the summary statistics (Wiegand & Moloney, 2014).

Therefore, the aim of this study was to propose a methodological correction that minimizes the effect of heterogeneity in the identification of roadkill hotspots through the selection of observation windows, called the Windowed method. To do so, we tested the following hypotheses: (a) the proposed correction presents a lower rate of classification errors of a region as a hotspot or not when compared to others methods used in road ecology studies and (b) the proposed correction is less sensitive to the non-homogeneity of the distribution of wildlife-vehicle-collisions.

To test the first hypothesis, we simulated data sets for 20 scenarios with varying densities and heterogenic patterns of roadkills distribution, in which there are well-defined 'true' hotspots. Then, a test was performed to determine which method was better in distinguishing them. We expected to find the methodology with less sensitivity to the heterogeneity of the event distribution greater accuracy in the hotspot identification.

For the second hypothesis, we performed two experiments. The first, with the simulated data, we selected and compared the best results obtained in each of the three

methods according to the density of the events and according to the different patterns of heterogeneity. We expected that the accuracy of the Windowed methodology would not vary according to the patterns of heterogeneity nor according to the density of events. In the second experiment we used real data and removing four different road stretches, one at a time, we created the different patterns of heterogeneity. In this case, we expected that the classification error, when compared to the result obtained for the highway without the removal of the sections, would be smaller for the Windowed methodology.

## 2. MATERIAL AND METHODS

### 2.1. Simulated study

#### 2.1.1. Creating the dataset

We created the datasets based on different simulations of heterogenic probabilities of occurrences. First, one real highway was georeferenced, and the line generated (figures 1-A) was divided into stretches of approximately 100 meters, totaling 408 stretches. Then, six notable clusters were predefined (black stretches on figure 1-A), varying in length and density, simulating a feature that promotes an accumulation of roadkills at a higher rate than expected, like a stretch without fences on a fenced road. The values of length and density were determined as follows: two hotspots of 300 meters in length and a density of 0.18 roadkill/m; two of 700 meters and 0.12 roadkill/m; and two of 1,000 meters with 0.1 roadkill/m. As the road was approximately 41 km in length, this represents almost 10% of the road length.

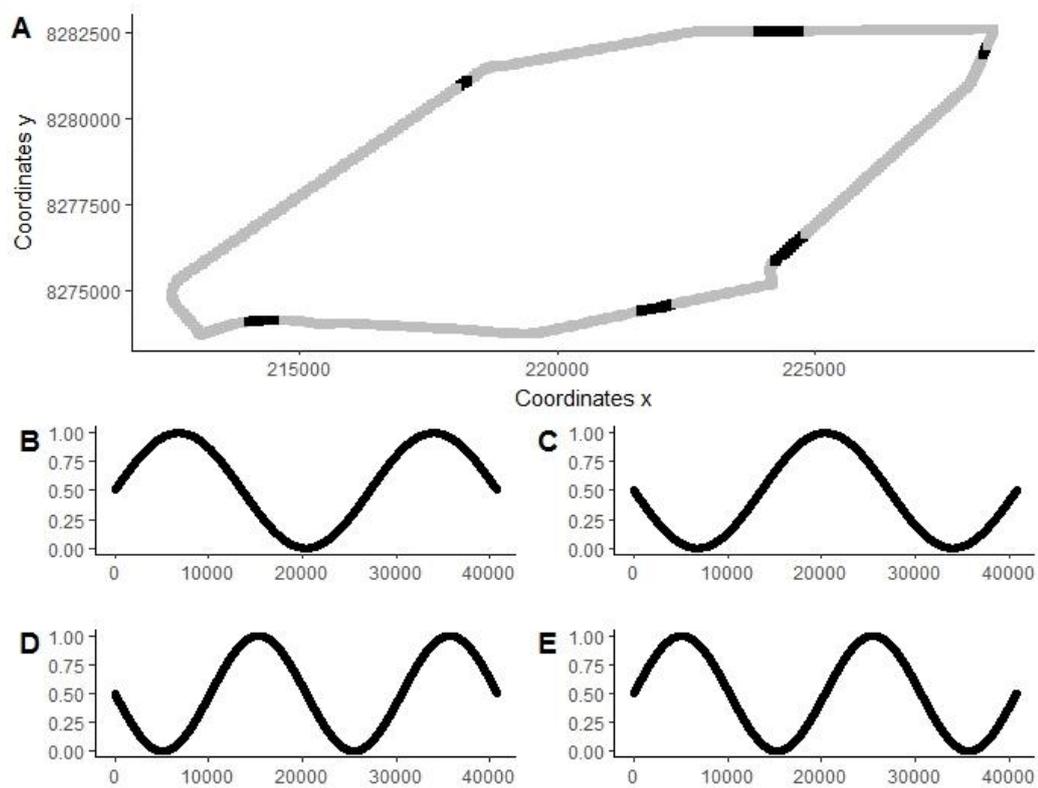


Figure 1. In graph A, the black sections represent the stretches defined as hotspots. In figures B to E, each peak represents higher probabilities of occurrence of events at the relative position on the road, while the curved valley simulates a low probability of occurrence of events in this region.

We performed the simulation of the points where roadkills occurred with the overlapping of two distributions. Initially, we performed a Poisson distribution only for the pre-determined hotspots. In this step, almost five hundred events were created using the hotspot densities as the lambda. As the other stretches had a density equal to zero, at this stage, all events fell into the hotspots.

In the second step, we performed four different Poisson distribution to simulate events that do not occur, following the forces of attraction that explain the hotspots. Thus, they are called 'noise events' or 'noise,' and the distribution that generated them was a 'noise type' distribution. Four different types of inhomogeneous event

distributions, with a total density of 0.06 roadkill/m (approximately 2,500 events) were created to simulate the noise distributions. We use a new Poisson distribution. However, at this time using the propensity distribution created by four different sinusoidal curves as the lambda (figure 1-B to 1-E). Some events also occurred within the hotspots, in this second distribution. They increased their final densities. As the probability of receiving new events varies according to the type of sinusoidal curve, the same hotspot presents different densities according to the kind of simulated heterogenic distribution. Each of the four Poisson distributions generated in the second step was superimposed on the distribution generated in the first step, thus creating the four simulated spatial patterns. The four patterns have the same locations as the hotspots. However, each has a different type of noise.

After the creation of the four patterns of distribution of runoffs, we split each in one-third, one-fourth, and one-fifth. Thus, five different densities were set for each of the four distribution patterns, totaling 20 scenarios. Each of the 20 event distributions created 21 Bootstrap samples, by re-sampling the events with replacement. Thus, this experiment used 420 different simulations of roadkill patterns, but with the same expected hotspot positions.

### *2.1.2. Testing the accuracy of hotspot identification*

The method proposed here adjusts the formula of Coelho et al. (2012, 2014) utilizing the same windows logic suggested by Wiegand & Moloney (2014) to the cluster identification analysis for the correction of first-order heterogeneity in CSR rejection analyses. According to them, we can test second-order homogeneity in a reasonably straightforward way, by selecting several windows, within which the point distribution appears to be approximately homogeneous. The estimated values of the

functions within the different windows should be essentially the same and should only vary due to stochastic fluctuations.

We used the Hotspot 2D formula for the Windowed method to find an aggregation of events. These events were concerning to regularly distributed points 'i', both for the local neighbourhood using a radius 'r' –  $H(r)$ , and for the observation window using a radius 'w' –  $H(w)$ . Both areas were centred on each point 'i'. The  $H(r)$  value for each point 'i' was divided by the equivalent  $H(w)$  to get the value of  $H(rw)$ . We compared the observed results to expected values for one hundred Poisson distributions. The observed values for a point 'i' greater than the expected 0.975th percentile value classifies this point as a hotspot.

To find the best ratio  $r/w$  for the new algorithm, 19 radii between 200 and 2,000, separated by 100-m increments, were used as 'r'. The windows 'w' were defined between 1,000 m and 4,000 m, with values increasing by 200. All values of 'r' and 'w' were combined, and when ratio  $r/w$  was equivalent to 1.33 or less, the results were excluded, totalling 314 combinations. Exclusions were made for ecological and mathematical reasons. First, because a window only slightly larger than the radius 'r' under analysis does not present an external pattern that can be compared. Moreover, this small difference increases the probability of having values of  $r/w$  close to 1 and is considered an unsuccessful analysis effort by a large number of falsely high  $H(rw)$  values. Additionally, the biggest value for 'w' was 4,000 because huge windows may not reflect relatively homogeneous areas.

The Hotspot 2D formula was applied to the data similarly, without the quotient between 'r' and 'w' radii, hence generating fewer analyses (19 per simulation). Malo's method was applied using the same scales used for Hotspot 2D.

To assess the accuracy of each method, we constructed confusion matrices between the observed hotspot pattern and its respective ‘true’ pattern. From those matrices, we extracted the values of the total error rate. The lower the total error rate, the higher the accuracy of the analysis. The total error rate is the ratio between all misclassifications and all possible ratings. For defining roadkill hotspots, the false positive classification (Type Error I) means that a point is identified as a hotspot when it is not one. The false negative classification (Type Error II) is when a point is not properly identified as a hotspot by the analysis. Both errors are considered equally negative since false positives can direct conservation efforts to site with smaller demand, while false negatives may fail to mitigate an impact in a place where it is demanded.

With the new method, The best result obtained for each radius ‘r’ was used to compare the effect of variations in the noise type and the density. The variable controls of those analyses were the radius ‘r’, noise type and density. Finally, for the radius ‘r’ we selected the results with the lowest mean total error rate for each method, and their accuracies were compared by noise type and density.

For all comparisons, confidence intervals were calculated with 1,000 iterations using bootstrap method, and we considered a significant difference if the 95% confidence intervals did not overlap among compared factors.

## *2.2. Real data study*

### *2.2.1. Study area*

We conducted the surveys along four roads (total 40 km), in Brasília, Federal District of Brazil, located in the central-western region. Surveys included four-lane (BR-020, 11 km), two-lane (DF-345 and DF-128, 10 km each) and dirt roads (DF-205,

09 km). These road sections delimit a protected area, named Águas Emendadas-ESECAE Ecological Station (10,000 ha), which is recognised by UNESCO as a core area of the Cerrado Biosphere Reserve (Figure 02).

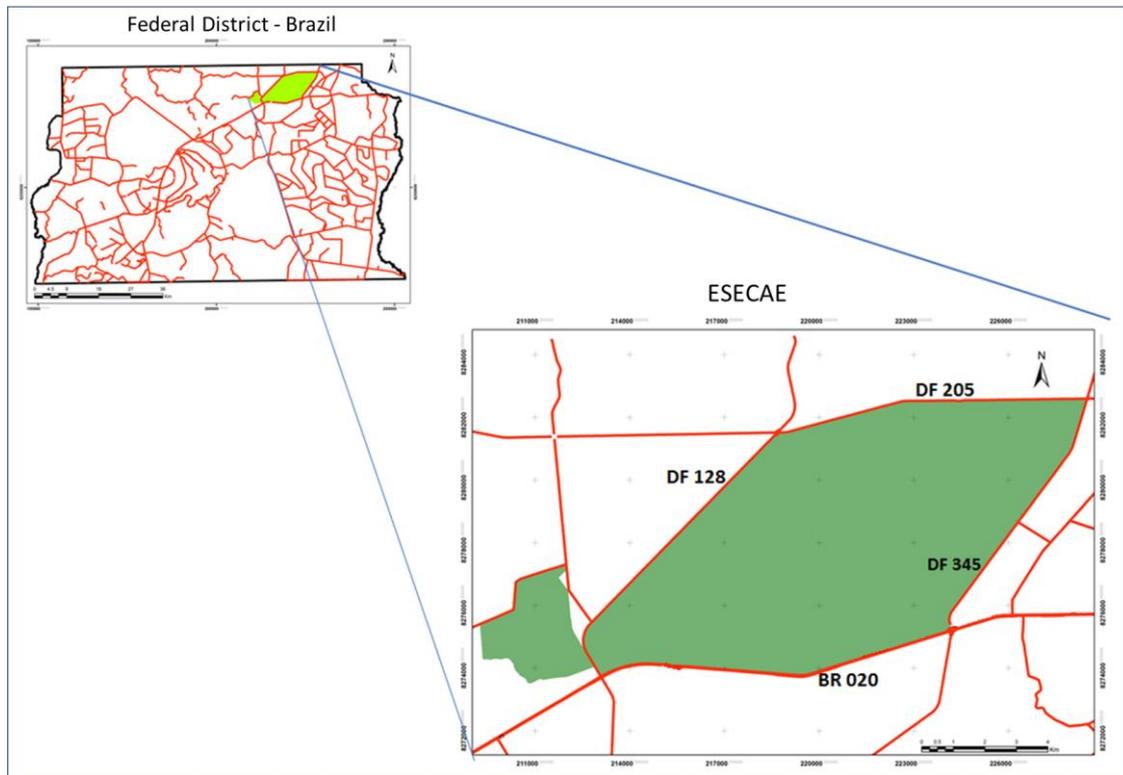


Figure 1. The large green area is the Águas Emendadas Ecological Station – ESECAE, in Brasília, Federal District of Brazil.

### 2.2.2. Data Collection

Sampling campaigns were conducted twice a week, between April 2010 and March 2015, by a team of three observers in a car at 50 km/h on average. The animals found were photographed, and their geographical coordinates were registered using a hand-held GPS (Garmin eTrex Touch 35) with five m-accuracy. Animals were immediately removed from the road to avoid possible recounting. For this case study,

hotspot analyses were performed considering all the events without taxonomic or functional distinction.

### *2.2.3. Data analysis*

We performed the hotspot analysis using both the Hotspot 2D formula and Malo's method, with a 200 m radius. For the application of the new method, the result obtained for the radius of 200 m was weighted in windows of 2,800 m. For all methods, the null models were formed by 1000 random Poisson distributions.

To analyse the influence of the heterogeneity of the distribution on identified hotspots we remove each of the four highways that compose the study area from the analysis, one at a time, disregarding their extensions and number of events that occurred there.

The aggregation profile observed on the non-removed sections was compared with that obtained for the same stretches in the control analysis. That is, the analysis performed without removal of any road.

A confusion matrix was also used to compare the results before and after the removal of each highway using the 'appearance' – False Positive Classifications – or 'disappearance' – False Negative Classification – of hotspots, as well as by the geographical distribution of classification divergences.

For all experiments analyses were performed with R 3.3.1 (R Core Team 2016), and the main packages used for those functions were Spatstat (Baddeley & Turner 2005, Baddeley et al. 2015), Fields (Nychka et al. 2015) and SDMTTools (VanDerWal et al. 2014).

### 3. RESULTS

#### 3.1 Simulated study

We performed 420 hotspot analyses for each of the sample units combinations of 'r' and 'w' for the Windowed method, and 'r' for the Hotspot 2D and Malo's methods, considering the combination of noise types and density, with 21 simulations conducted for each.

According to figure 3, the lowest error rates occurred for smaller radii. However, when considering each radius separately, the error rates decrease as the observation window radius increases until reaching a certain radius for which the larger the 'w' radius, the greater the error rate. The most efficient combination was a radius 'r' of 200 m and a window 'w' of 2,800 m (total error rate: 0.074 CI: [0.072;0.076]).

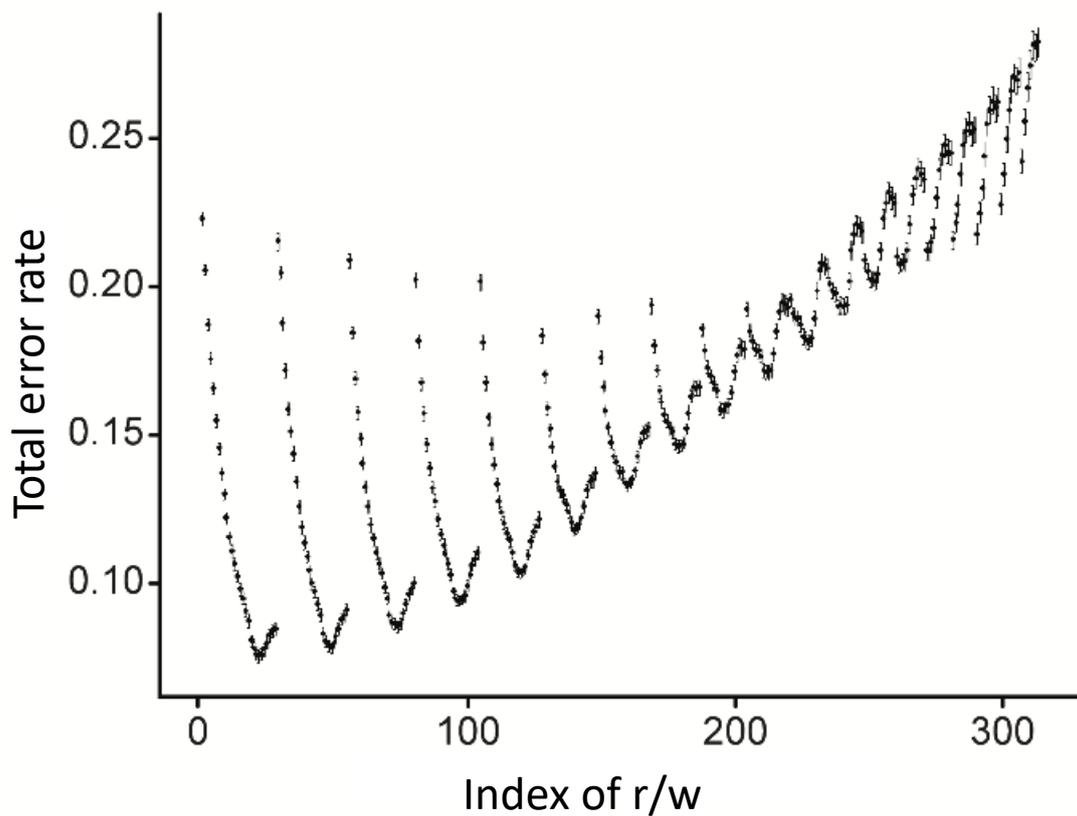


Figure 3. Results obtained from each division of radii ‘ $r$ ’ and windows ‘ $w$ ’ considering the 20 scenarios of noise type and density. These graphics were formed by different series of queued points. Each sequence of points represents a radius ‘ $r$ ’, and each point, present in each row, represents the average obtained by bootstrap for each combination of this radius with a window ‘ $w$ ’, totalling 3.

The results obtained for each combination of noise type and density is illustrated in Figure 4. For this analysis and the others in sequence, only the best result obtained for each radius for the new method was selected. The error rates from the new methodology are smaller than those from Hotspot 2D and Malo’s method. However, these values tend to be closer with decreasing density (graphs on the bottom line of the figure 4).

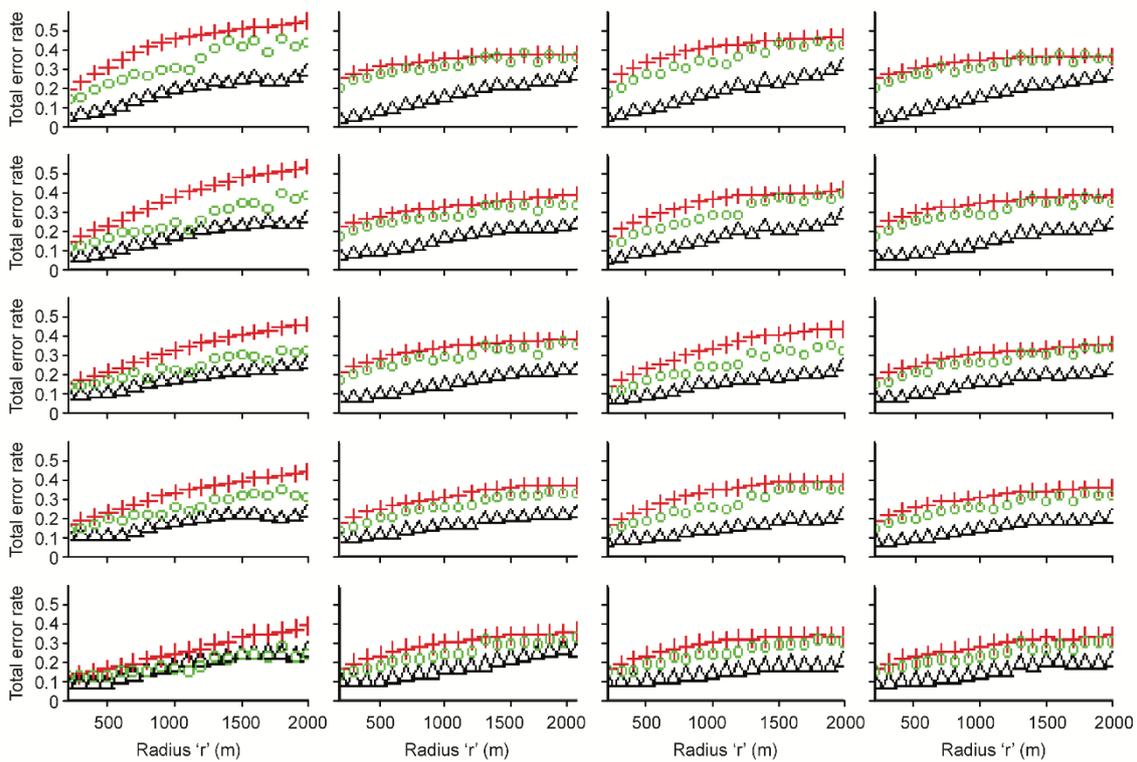


Figure 4 In this figure, each chart column represents one of the four types of noise illustrated in figure 1 B-E, following the same sequence from the first to the fourth column. The lines represent densities, with the density decreasing in the downward direction. The red crosses represent the total error rate values obtained with the Hotspot 2D formula for each radius, the green circles are the results of Malo's method, and the black triangles are values achieved with the new methodology.

For Hotspot 2D and Malo's methodologies, the results reveal a slight inversely proportional trend between the density of events and the accuracy of the method. The Windowned method does not show much variation in its accuracy with changes in density despite acting efficiently in all density grades (figure 5-A). Regardless of whether the new formula generates better results for all noise types (figure 5-B), the noise types caused different effects in all methods. Considering only the best results for each formula (figure 5-C), the variation in density promotes an opposite effect for the

new method and the other two methodologies. For the Hotspot 2D and Malo's methods, shorter densities present greater accuracy, whereas for the new method, the higher the density of events, the lower the error rate was. At least, considering only the best analysis scales for all type of noise, the Total error rate for Windowed Method was smaller than the others (Figure 5-D). Besides that, despite the type of noise, the new method results were significantly the same, while the accuracies of the Hotspot 2D and Malo's methods vary according to the type of noise (Figure 5-D).

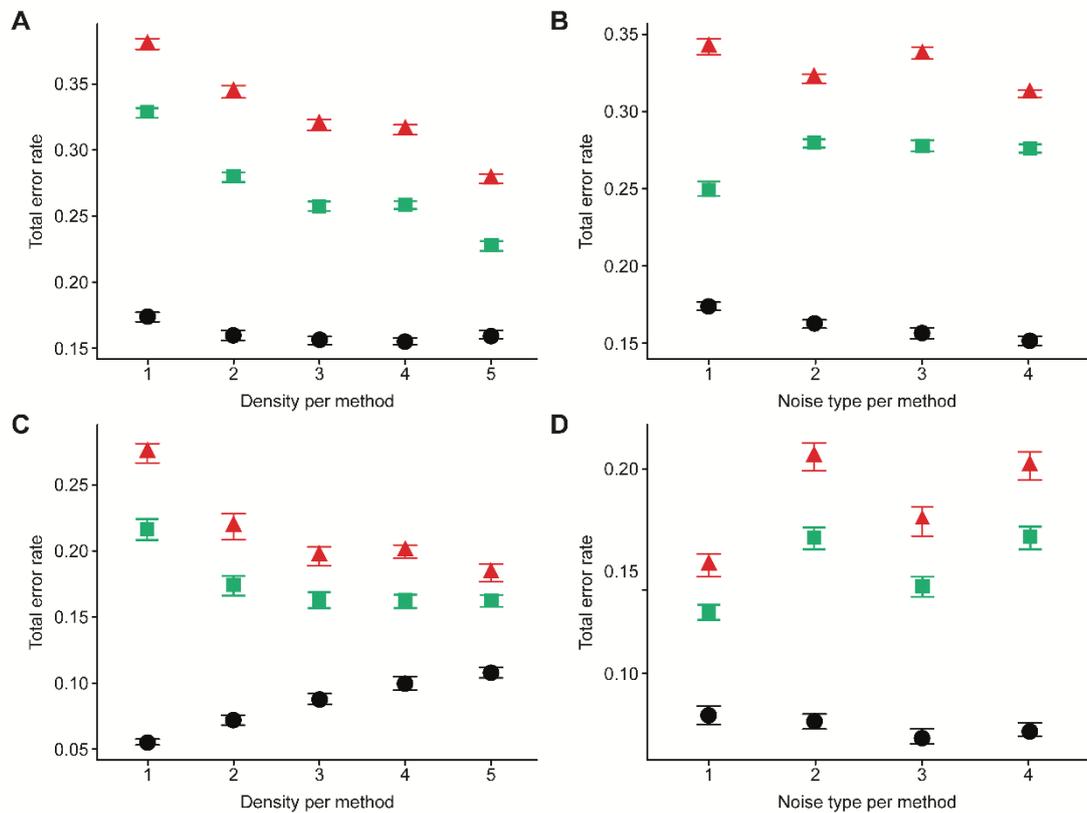


Figure 2. All graphs show the effects of changes in density and in distribution noise types on the three methods results. Graphs A and B show the results considering all scales of analysis and graphs C and D show those considering only the most efficient analyses for each method, that is, a radius of 200 m for the Hotspot 2D formula and Malo's method and a radius of 200 m with a window of 2,800 m for the Windowed

method. For all graphs, the red triangles, green squares and black circles represent the Hotspot 2D formula, Malo's method and the Windowed method, respectively

### 3.1 Real data study

During the five years of collection, we recorded the carcasses of 1,251 birds, 370 reptiles, 282 mammals and 150 amphibians, totaling 2,053 records. Of these records, we found 53% on the BR-020, 35% on the DF-128, 10% on the DF-345 and 2% on the DF-205.

The total length of the sections indicated as hotspots (black stretches in figure 6) for Malo's method – 7.6 km – was much smaller than that indicated by the other two methods, both of which identified 16.9 km as hotspots (Figure 6). However, the new methodology presented 30 stretches with an average extension of  $0.563 \text{ km} \pm 0.438 \text{ km}$ , while Hotspot 2D indicated 11 hotspots with a mean extension of  $1.537 \text{ km} \pm 1.891 \text{ km}$ , and Malo identified 14 stretches with an average length of  $0.542 \text{ km} \pm 0.946 \text{ km}$ .

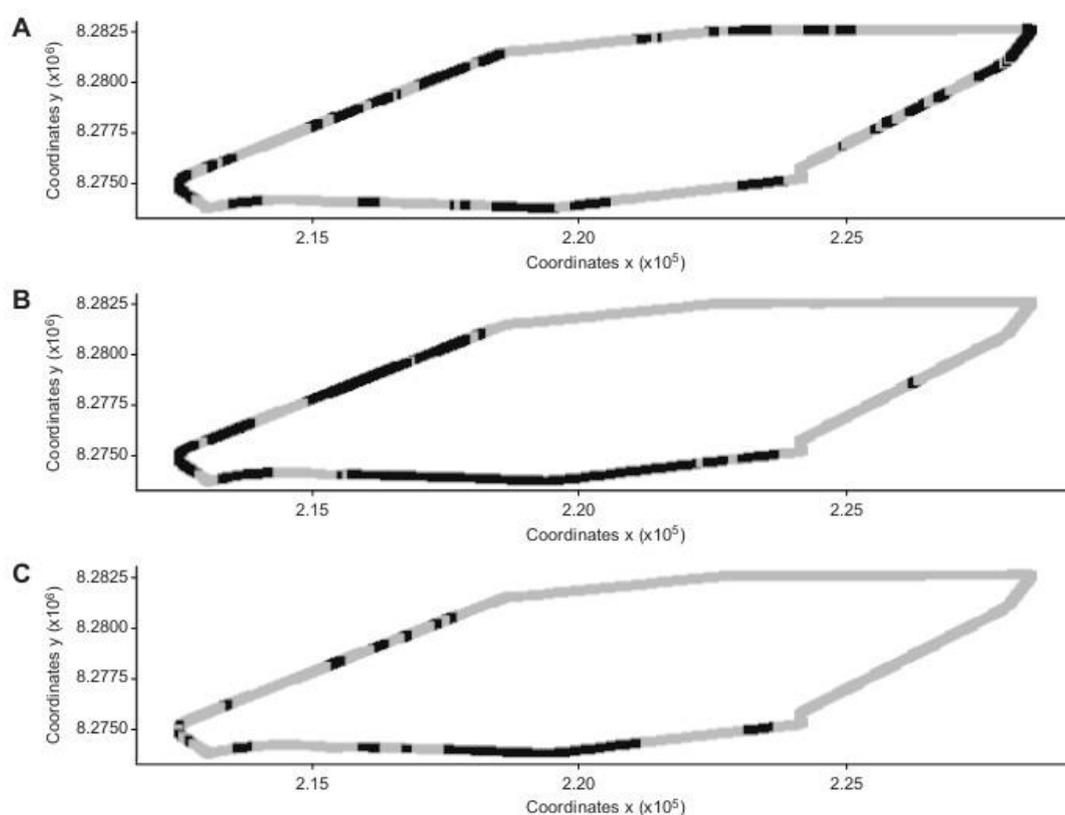


Figure 3. Results of the hotspot analysis using the Windowed methodology (graph A), Hotspot 2D (graph B) and Malo's method (graph C).

The hotspots proposed by the Hotspot 2D and Malo's methods are concentrated in two highways (BR-020 and DF-128), and both indicate a large continuous stretch as a hotspot on BR-020, with 6.7 km for Hotspot 2D and 3.8 km for Malo. Meanwhile, the new methodology indicates significant aggregations in all the highways, and the biggest continuous stretch is 1.8 km in length.

According to figure 5 B, almost all of the BR-020 highway is considered to be a hotspot. However, refining the methodology with window weighting, it is possible to verify that, within this large region susceptible to mitigation, there are areas whose density is significantly higher than the already high density of the stretch. Thus, justifying the identification of these points for proposed mitigation measures.

In the experiment with highway removal, all methods show variation in the correct classification of a stretch as a hotspot (table 2). For the Malo and Hotspot 2D methods the false positives only occurred when the BR-020 and DF-128 roads were removed, and the opposite occurred for the DF-205 and 345 highways.

Table 1. Number of hotspots that 'appeared' - False Positive Hotspot -and 'disappeared' - False Negative Hotspot - due to the withdrawal of a highway in the analysis of hotspots.

	False Negative Hotspots				False Positive Hotspots			
	I	II	III	IV	I	II	III	IV
Proposed	1	4	19	2	17	18	6	20
Hotspot 2D	0	10	24	0	39	0	0	11
Malo's	0	10	18	0	32	0	0	6

*I-Dataset analyzed without BR-020; II - Dataset analyzed without DF-345; III - Dataset analyzed without DF-205; and IV - Dataset analyzed without DF-128.*

However, for the Windowed method, most of the classification errors occurred near the edges of the road stretches contiguous to the removed highway. For the other methods, the number of errors was much higher when BR-020 was removed, and the classification error pattern became more marked. The removal of BR-020 and DF-128 resulted in the appearance of false hotspots, while the removal of DF-205 and DF-345 resulted in the disappearance of true hotspots (Figure 6).

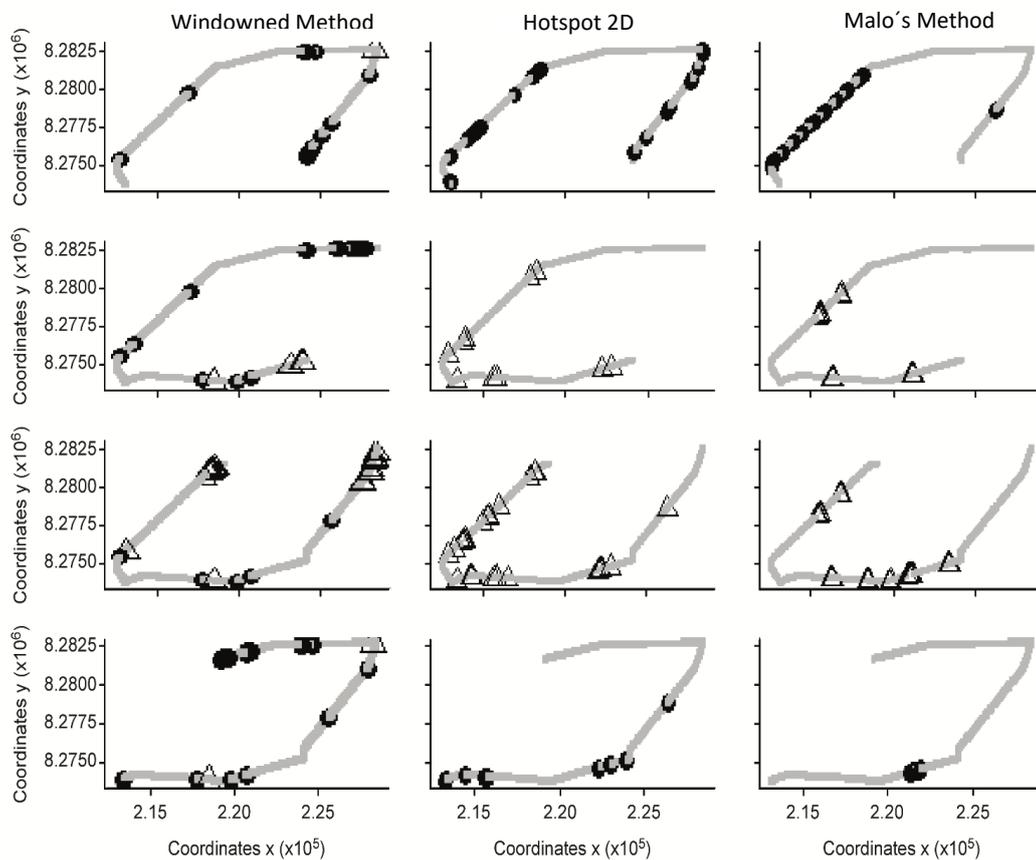


Figure 4. The columns represent the results obtained using the Windowed method, Hotspot 2D and Malo's method, respectively. Each of the lines represents the removal of a highway from the study area, and in the downward direction, we have the removal of BR-020, followed by DF-345, DF-205 and finally DF-128. The triangles represent original hotspots that were not classified as such by the analysis with the

removal of a highway, and the circles are the points not originally identified as hotspots but have been classified as such in the analyses

#### *4. DISCUSSION*

Our results from the simulated study indicate that regardless of the method used and the scenarios of the event distributions the Windowed methodology application is less sensitive to the point pattern distribution first-order heterogeneity than the other methods analyzed (figure 3). Moreover, the smaller the analysis radius, the lower the sensitivity of the method was to the influence of the first-order effects of the distribution of point events on the identification of second-order processes. This situation goes against the common use of Ripley's K function as a tool to identify radii for which the aggregation is nonrandom. In this study, the study radius for hotspot definition can be determined (Mountrakis & Gunson 2009, Danks & Porter 2010, Coelho et al. 2012, Langen et al. 2012, Teixeira et al. 2013).

It is important to note that Ripley's K statistic has been subject to many questions (e.g., Wiegand & Moloney 2004, 2014, Schiffers et al. 2008, Marcon & Puech 2009), mainly related to the so-called virtual aggregation effect. Thus, the use of Ripley's K statistic may lead to choices that reflect virtual aggregation in hotspot analysis.

Other relevant information that can be obtained from the simulated study is the effect of density on method accuracy. It is expected that increased sampling effort results in greater method efficiency. The results obtained with the Windowed method support this assertion. However, the accuracies of the Hotspot 2D and Malo's methods were higher with decreasing event density. The effect of the differentiation between the

noise types is also important to note. Regardless of the kind of noise tested, the Windowed method experienced little variation in the accuracy of its results. The fact that Malo's and Hotspot 2D methods presented oscillations in accuracy with changes in the distribution profile shows that those methodologies are more sensitive to first-order heterogeneities. As each highway has a different profile, it is not possible to determine which type of highway the formulas will be more or less efficient.

The real data study presents a large difference in the classifications generated by the different methodologies. The Hotspot 2D formula generated a hotspot pattern that reflects more the effects of the landscape on roadkills than punctual processes. The vast stretches of hotspots generate uncertainty about the ideal locations for implementing mitigation measures. Malo's method is more accurate than Hotspot 2D, indicating small stretches as hotspots, but both identify significant aggregations only on the roads that present greater regional intensities in event occurrences.

Unlike the other methods, the method proposed in this article indicated significant points of aggregation along the four highways, making it possible to verify points with a higher probability of being hit on highways with low densities of events. Also, the map presents another significant finding: according to figure 5 B, almost all of the BR-020 highway is considered to be a hotspot. However, when refining the methodology with window weighting, it is possible to check that, within this broad region susceptible to mitigation, there are areas where density is significantly higher than the previous high density of the stretch, justifying the identification of these points for proposed mitigation measures.

The results obtained by the one-road removal experiment verify the lower sensitivity of the Windowed method to variations in the type of data distribution. For both Malo and Hotspot 2D, the removal of the higher-density road (simulations 'I' of

table 2) increases the number of hotspot classifications. The increase occurs because the null model takes into account the ratio between the number of individuals and road length, removing a considerable number of events. The lambda value tends to decrease, reducing the threshold value for classification as a hotspot, even though there is no variation in the records of this section. This effect is also observed when the removal of a small-density highway causes the threshold value to increase, causing some points to cease to be significant aggregations, even though there has been no change in the number of records at those points.

Notwithstanding these peculiarities, the three methodologies presented changes in the classification of stretches as hotspots. The explanation for these changes in all cases is the null model used. When applying a Poisson distribution for all highways, even with notorious variations in roadkill density, a road with a high density of events determines the classification of the other roads, even in some cases when there is independence between the registers. Borda-de-Agua et al. (2016) questioned the use of the Poisson distribution as a null model owing to the high probability of type II errors.

There are situations in which hotspots may not be the best identifier of sites for mitigation. For instance, where the road segments are independent, where there are differences in traffic intensity, and when road-killed animals on the different segments represent individuals from different populations (Eberhardt et al., 2013, Teixeira et al., 2017). The authors conclude that due to the effects of past mortality, the relationship between road mortality and population abundance in the surroundings may be better for informing mitigation priorities than identifying roadkill hotspots alone. Thus, all these studies reinforce the need to consider the different ways in which the characteristics of the landscape can determine the first-order effects on roadkill aggregation. At some points identified by this methodology as hotspots, the absolute incidence of roadkill is

much lower than that at other non-identified points. This fact does not mean that mitigation priority should be given to points with fewer records, though these points may help to compose aggregation models based on the surrounding environmental characteristics.

Considering the four possibilities for dealing with the non-homogeneities in the point distribution of events proposed by Wiegand & Moloney (2014), state of the art in road ecology still using the alternative of ignoring this characteristic in analyses of roadkills aggregation. This paper begins the debate on the application of other alternatives, proposing a methodological change to avoid or reduce the effects of heterogeneity by using observation windows.

Although this article goes beyond the first option, this approach should be better by using a more rigorous method for delineating homogeneous sub areas. Our results minimized part of the heterogeneity effect in the pattern, but the windows used still reflect a weak signal of heterogeneity. The subdivision of a studied highway into minimally homogeneous segments can be carried out in several ways. Using one or more landscape and traffic characteristics that show a significant correlation with the trampling of the species of interest. Using cluster identification techniques, such as a method based on the similarity between neighbors called Shared Nearest Neighbour (SNN), created by Jarvis & Patrick (1973) and improved upon by Ertöz, Steinbach & Kumar (2003). The algorithm of this method finds the nearest neighbors of each data point and then redefines the similarity between pairs of points regarding how many nearest neighbors the two points share.

The third and fourth alternatives proposed by Wiegand & Moloney (2014) can be tested in future research, including new methods for window definition, the

proposition of null models with heterogeneous Poisson distributions, or changes in the metrics used in road ecology, as proposed in Ang, Baddeley & Nair (2012).

### *5. CONCLUSION*

This paper introduces the debate on the application of other alternatives, proposing a methodological change to avoid or diminish the effects of heterogeneity by using observation windows. The Windowed method can be used as a way to minimize the effect of the first-order heterogeneity of a distribution of point events on or along a linear network. The main utility of this method is its ability to identify smaller stretches with intensities greater than expected in the interior of a zone with a high intensity of roadkill, which improves the process of choosing sites for mitigation.

The method should not be applied in a unique way for the proposal of mitigation measures, since the identification of hotspots in areas with low roadkill density may not always indicate a mitigable intensity. However, the identification of these points, which do not usually appear in traditional hotspot analyses, can improve the proposition of landscape-based models.

### *6. ACKNOWLEDGEMENTS*

We thank the Postgraduate Programme in Ecology of the University of Brasília – UnB for providing time and understanding when requested by A.P.F. We thank Lorena Ribeiro and Rodrigo Santos for discussions and suggestions during the conception of this study. We thank also all members of Rodofauna Project of Environmental Brasília Institute (IBRAM) for their help in collecting data. LMS Aguiar thanks CNPq for her research scholarship. All authors declare no conflicts of interest.

## 7 BIBLIOGRAPHY

- Ang, Q. W., Baddeley, A., &Nair, G. (2012). Geometrically corrected second order analysis of events on a Linear Network, with applications to ecology and criminology. *Scandinavian Journal of Statistics*, **39**, 591–617. doi:10.1111/j.1467-9469.2011.00752.x.
- Baddeley, A. J., Møller, J., &Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, **54**, 329–350. doi:10.1111/1467-9574.00144.
- Baddeley, A., Rubak, E., &Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. CRC Press, Boca Raton, FL.
- Baddeley, A., &Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, **12**. doi:10.18637/jss.v012.i06.
- Barrientos, R., &Plaza, M. (2016). Road-kill hot spots can change over the time, variables explaining them do not. *International Conference on Ecology and Transportation, Programme and Abstracts*. (ed. É.Guinard), p. 288. CEREMA, Lyon, France.
- Beaudry, F., DeMaynadier, P. G., &Hunter, M. L. (2008). Identifying road mortality threat at multiple spatial scales for semi-aquatic turtles. *Biological Conservation*, **141**, 2550–2563. doi:10.1016/j.biocon.2008.07.016.
- Borda-de-Agua, L., Ascensão, F., Barrientos, R. & Pereira, H.M. (2016). On the identification of high mortality rate hotspots. *International Conference on*

*Ecology and Transportation, Programme and Abstracts* (ed É. Guinard), p. 286.  
CEREMA, Lyon, France.

Coelho, A. V. P., Coelho, I. P., Teixeira, F. Z., & Kindel, A. (2014). *Siriema: Road Mortality Software. User's Guide*. NERF/UFRGS, Porto Alegre, Brasil.

Coelho, I. P., Teixeira, F. Z., Colombo, P., Coelho, A. V. P., & Kindel, A. (2012). Anuran road-kills neighboring a peri-urban reserve in the Atlantic Forest, Brazil. *Journal of Environmental Management*, **112**, 17–26. doi:10.1016/j.jenvman.2012.07.004.

Costa, A. S., Ascensão, F., & Bager, A. (2015). Mixed sampling protocols improve the cost-effectiveness of roadkill surveys. *Biodiversity and Conservation*, **24**, 2953–2965. doi:10.1007/s10531-015-0988-3.

Danks, Z. D., & Porter, W. F. (2010). Temporal, spatial, and landscape habitat characteristics of moose–vehicle collisions in western Maine. *Journal of Wildlife Management*, **74**, 1229–1241. doi:10.2193/2008-358.

Eberhardt, E., Mitchell, S., & Fahrig, L. (2013). Road kill hotspots do not effectively indicate mitigation locations when past road kill has depressed populations. *Journal of Wildlife Management*, **77**, 1353–1359. doi:10.1002/jwmg.592.

Ertöz, L., Steinbach, M., & Kumar, V. (2003). Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. *Proceedings of Second SIAM International Conference on Data Mining*, pp. 47–59.

- Grilo, C., Bissonette, J. A., & Santos-Reis, M. (2009). Spatial-temporal patterns in Mediterranean carnivore road casualties: Consequences for mitigation. *Biological Conservation*, **142**, 301–313. doi:10.1016/j.biocon.2008.10.026.
- Jarvis, R. A., & Patrick, E. A. (1973). Clustering using a similarity measure based on shared near neighbors. *IEEE Transactions on Computers*, C-22, 1025–1034. doi:10.1109/T-C.1973.223640.
- Langen, T. A., Gunson, K. E., Scheiner, C. A., & Boulerice, J. T. (2012). Road mortality in freshwater turtles: Identifying causes of spatial patterns to optimize road planning and mitigation. *Biodiversity and Conservation*, **21**, 3017–3034. doi:10.1007/s10531-012-0352-9.
- Liu, Q., Li, Z., Deng, M., Tang, J., & Mei, X. (2015). Modeling the effect of scale on clustering of spatial points. *Computers, Environment and Urban Systems*, **52**, 81–92. doi:10.1016/j.compenvurbsys.2015.03.006.
- Malo, J. E., Suárez, F., & Díez, A. (2004). Can we mitigate animal-vehicle accidents using predictive models? *Journal of Applied Ecology*, **41**, 701–710. doi:10.1111/j.0021-8901.2004.00929.x.
- Marcon, E., & Puech, F. (2009). *Generalizing Ripley 's K function to inhomogeneous populations*. HAL archives-ouvertes.fr, HAL Id: halshs-00372631.
- Miller, H. J., & Han, J. (2009). Geographic Data Mining and Knowledge Discovery an Overview. *Geographic Data Mining and Knowledge Discovery* (eds H.J. Miller, J. Han), pp. 1–26. CRC Press, Boca Raton, FL.

- Mountrakis, G., & Gunson, K. (2009). Multi-scale spatiotemporal analyses of moose–vehicle collisions: A case study in northern Vermont. *International Journal of Geographical Information Science*, **23**, 1389–1412. doi:10.1080/13658810802406132.
- Nychka, D., Furrer, R., Paige, J., & Sain, S. (2015). *Fields: Tools for Spatial Data*. doi:10.5065/D6W957CT.
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>.
- Ramp, D., Caldwell, J., Edwards, K. A., Warton, D., & Croft, D. B. (2005). Modelling of wildlife fatality hotspots along the Snowy Mountain Highway in New South Wales, Australia. *Biological Conservation*, **126**, 474–490. doi:10.1016/j.biocon.2005.07.001.
- Santos, R. A., Santos, S. M., Santos-Reis, M., Picanço de Figueiredo, A., Bager, A., Aguiar, L. M., & Ascensão, F. (2016). Carcass persistence and detectability: Reducing the uncertainty surrounding wildlife-vehicle collision surveys. *PLOS One*, **11**, e0165608. doi:10.1371/journal.pone.0165608.
- Santos, R. A., Ascensão, F., Ribeiro, M. L., Bager, A., Santos-Reis, M., & Aguiar, L. M. S. (2017). Assessing the consistency of hotspot and hot-moment patterns of wildlife road mortality over time. *Perspectives in Ecology and Conservation*, **15**, 56–60. doi:10.1016/j.pecon.2017.03.003.
- Santos, S. M., Marques, J. T., Lourenço, A., Medinas, D., Barbosa, A. M., Beja, P., & Mira, A. (2015). Sampling effects on the identification of roadkill hotspots:

Implications for survey design. *Journal of Environmental Management*, **162**, 87–95. doi:10.1016/j.jenvman.2015.07.037.

Schiffers, K., Schurr, F. M., Tielbörger, K., Urbach, C., Moloney, K., & Jeltsch, F. (2008). Dealing with virtual aggregation - A new index for analysing heterogeneous point patterns. *Ecography*, **31**, 545–555. doi:10.1111/j.0906-7590.2008.05374.x.

Seiler, A., Sjölund, M., Andrasik, R., Rosell, C., Torrellas, M., Sedonik, J., Bíl, M., & Jägerbrand, A. (2016). Are animal-vehicle collisions a random event? – Analysis of the spatial distribution of accident data. *International Conference on Ecology and Transportation, Programme and Abstracts*. (ed. É. Guinard), p. 290. CEREMA.

Seo, C., Thorne, J. H., Choi, T., Kwon, H., & Park, C.-H. (2015). Disentangling roadkill: The influence of landscape and season on cumulative vertebrate mortality in South Korea. *Landscape and Ecological Engineering*, **11**, 87–99. doi:10.1007/s11355-013-0239-2.

Teixeira, F. Z., Coelho, I. P., Esperandio, I. B., Oliveira, N. R., Peter, F. P., Dornelles, S. S., Delazeri, N. R., Tavares, M., Martins, M. B., & Kindel, A. (2013). Are road-kill hotspots coincident among different vertebrate groups? *Oecologia Australis*, **17**, 36–47. doi:10.4257/oeco.2013.1701.04.

Teixeira, F. Z., Kindel, A., Hartz, S. M., Mitchell, S., & Fahrig, L. (2017). When road-kill hotspots do not indicate the best sites for road-kill mitigation. *Journal of Applied Ecology*. doi:10.1111/1365-2664.12870

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L., & Storlie, C. (2014). *SDMTools: Species Distribution Modelling Tools: Tools for Processing Data Associated with Species Distribution Modelling Exercises*. – R package ver. 1.1-20, <  
[http://CRAN.R-project.org/package = SDMTools](http://CRAN.R-project.org/package=SDMTools) >.

Wiegand, T., & Moloney, K. A. (2004). Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, **104**, 209–229. doi:10.1111/j.0030-1299.2004.12497.x.

Wiegand, T., & Moloney, K. A. (2014). *Handbook of Spatial Point-Pattern Analysis in Ecology*. CRC Press, Boca Raton, FL.

## **CAPÍTULO 02 – DEFINIÇÃO DA CURVA DE PRECISÃO PARA ANÁLISE DA SUFICIÊNCIA AMOSTRAL EM ECOLOGIA DE ESTRADAS**

### *RESUMO*

Em uma pesquisa científica, a amostra deve ter o tamanho suficiente para validar os resultados sem representar, no entanto, um desperdício de recurso e tempo. Para este capítulo testou-se a hipótese de que a precisão da análise de agregação de atropelamentos de fauna com o tamanho da amostra pode ser utilizada para estabelecer a suficiência amostral. Os valores de precisão de classificação de *hotspot* foram obtidos comparando-se vinte e um conjuntos de dados simulados, a partir de reamostragens com reposição de registros reais de atropelamentos, com cada um de seus subconjuntos, resultados de divisões de três a dez vezes do conjunto da amostra simulada. Utilizou-se regressão quantílica entre os valores de precisão e o logaritmo dos tamanhos dos subconjuntos dos quais as precisões foram obtidas. Nesta regressão foram simulados os quantis 2,5%, 50% e 97,5% para representar três cenários de grau de eficiência do método de classificação de *hotspots*, do menos ao mais eficiente. Os resultados demonstraram que existe correlação positiva entre a precisão de classificação de *hotspots* com o tamanho da amostra e que esta relação é válida para diferentes formas de se agrupar os resultados. Além disso, pode ser utilizada como forma de avaliar a precisão com o acúmulo de registros até que se atinja o valor de precisão desejado. A Curva de Precisão criada para os dados observados indicou que para um grau de precisão de 70%, mesmo no cenário pessimista, a suficiência amostral já havia sido obtida, indicando excesso de amostragem. Com este método, o momento para finalizar as coletas depende do acúmulo de registros, independentemente do esforço amostral.

Este método dá robustez às proposições de medidas mitigadoras, uma vez que garante o grau de precisão dos locais indicados.

**Palavras-chave:** coleta de dados, licenciamento ambiental, medidas mitigadoras, regressão quantílica, rodovias, suficiência amostral.

#### *ABSTRACT*

In an efficient scientific research, the sample must be large enough to validate the results without representing, however, a waste of resources and time. For this chapter, the hypothesis that the relationship between precision of the roadkills aggregation analysis and the sample size can be used to establish sample adequacy was tested. The hotspot classification precision values were obtained by comparing twenty-one simulated data set, by resampling with replacement of real roadkill records, with each of their subsets, results of three to ten-fold divisions of the simulated sample set. Quantile Regression was used between the precision values and the logarithm of the sizes of the subsets from which the precisions were obtained. In this regression the quantiles 2.5%, 50% and 97.5% were simulated to represent three scenarios of classifying hotspots method efficiency degree, from least to most efficient. The results showed that there is a positive correlation between the accuracy of classification of hotspots with the size of the sample and that this relation is valid for different ways of grouping the results. In addition, it can be used as a way of evaluating the precision with the register accumulation until the desired precision value is reached. The Precision Curve created for the observed data indicated that for a 70% precision degree, even in the pessimistic scenario, the sample adequacy had already been obtained, indicating an excess of sampling. With this method, the moment to finalize the collections depends on the accumulation of records, regardless the sample effort. This method gives robustness

to mitigating measures propositions, since it guarantees the precision degree of the indicated places.

**Key words:** data collection, environmental licensing, highways, mitigation measures, regression quantiles, sample sufficiency.

## INTRODUÇÃO

A definição do tamanho amostral é uma decisão muito importante em qualquer planejamento para uma pesquisa científica, pois amostras muito grandes representam desperdício de recursos, enquanto que muito pequenas diminuem a aplicabilidade e robustez dos resultados (Cochran 1977).

Dois vieses reduzem a eficiência das coletas de dados em Ecologia de Estradas: o tempo de permanência das carcaças nas rodovias e a capacidade de detecção destas carcaças pelo pesquisador (Slater 2002, Santos *et al.* 2011, 2016, Teixeira *et al.* 2013, Collinson *et al.* 2014). Estes fatores interferem no tempo de duração necessária do experimento para se atingir um tamanho amostral suficiente para as análises pretendidas. Para diminuir o efeito do tempo de permanência das carcaças nas rodovias, é recomendado curtos períodos de tempo entre as campanhas, no máximo uma semana, se o objetivo do estudo é a identificação da riqueza de espécies afetadas (Bager & da Rosa 2011). No entanto, para a redução da probabilidade de falsa indicação de *hotspots* de atropelamento recomendam-se coletas diárias ou com um dia de intervalo (Santos *et al.* 2015).

A correção do viés de detecção pode se dar pela redução da velocidade de deslocamento durante a busca pelas carcaças. Alguns estudos são realizados à pé (Attademo *et al.* 2011, Coelho *et al.* 2012, Skórka *et al.* 2015) ou de bicicleta (Garriga *et al.* 2012, Eberhardt *et al.* 2013, Garrah *et al.* 2015). Porém, em função da necessidade de se realizar grandes deslocamentos para os registros dos atropelamentos, a grande maioria dos estudos é realizada em automóveis.

O cenário apresentado indica que as pesquisas em Ecologia de Estradas deveriam ocorrer de forma sistemática, com frequência diária ou com um dia de

intervalo, e com deslocamentos com velocidade não superior a 20km/h (Collinson *et al.* 2014). No entanto, as formas de registro, frequência de coletas de dados e duração das amostragens têm grande variação entre os pesquisadores. Encontra-se na literatura estudos com coletas não sistemáticas (Mountrakis & Gunson 2009) ou coletas sistemáticas que variaram entre 80 dias consecutivos de bicicleta (Eberhardt *et al.* 2013), três anos de monitoramento semanal a pé (Skórka *et al.* 2015) ou mesmo em cinco anos de monitoramento com carro, realizadas parte diariamente e parte a cada dois dias (Seo *et al.* 2013).

Normalmente os protocolos de pesquisa são definidos com base nos custos envolvidos e na capacidade operacional (Cochran, 1977). Nos casos em que o protocolo ideal é inviável, o pesquisador pode optar por duas decisões: ou se abandona os esforços até conseguir mais recursos, ou assume-se a redução do tamanho amostral, reduzindo a precisão das análises (Cochran, 1977). Neste sentido, conhecer a relação entre o tamanho da amostra e a precisão dos resultados que ela oferece é uma das importantes informações a se obter para decidir pelo término do esforço de coleta de dados.

A pesquisa sobre protocolos mais eficientes em Ecologia de Estradas ainda é bastante escassa, com exceção de alguns trabalhos que relacionam o custo e benefício para conservação (Costa *et al.* 2015, Santos *et al.* 2017), o efeito da variação do esforço amostral na riqueza das espécies afetadas (da Rosa & Bager 2012, Collinson *et al.* 2014) ou na classificação dos *hotspots* (Santos *et al.* 2015). Entretanto, não foram encontrados estudos em Ecologia de Estradas que definam a relação entre o tamanho da amostra – independente do protocolo de coleta – e a eficiência desses dados na proposição de locais preferenciais para instalação de medidas de mitigação.

Snow et al. (2015) testaram o efeito da sub-amostragem nas análises estatísticas de correlação entre variáveis ambientais e probabilidade de atropelamento de grandes ungulados e concluíram que a sub-amostragem (até 70%, no caso estudado) não deveria impedir o uso dos modelos estatísticos de atropelamento para proposição de medidas de mitigação. Neste sentido, pode-se afirmar que 30% da amostra total estudada seria suficiente para as decisões baseadas nas análises dos dados. Assim conhecer o tamanho da amostra que permite realizar análises com segurança é importante para tornar mais eficiente a pesquisa em Ecologia de Estradas.

Desta forma, o objetivo deste trabalho é propor um mecanismo de análise da relação entre a precisão da classificação dos *hotspots* e o tamanho da amostra utilizada para esta classificação, visando auxiliar na definição da suficiência amostral de pesquisa de atropelamento de fauna. Para isso, foi identificada a relação entre a precisão na indicação de *hotspots* de atropelamento entre subconjuntos de dados e os respectivos conjuntos de dados simulados dos quais o subconjunto fora retirado. Além disso, foi feita a avaliação da aplicabilidade do método de construção da curva de precisão no decorrer do tempo, para de avaliar se a curva de precisão pode ser utilizada na definição da suficiência amostral.

## *MÉTODOS*

### *Coleta de dados*

As coletas de aniamis atropelados foram realizadas em Brasília, no Distrito Federal, no Brasil, ao longo de quatro rodovias (total de 40 km): uma pista duplicada (BR-020, 11 km), duas rodovias pavimentadas de mão simples (DF-345 e DF-128, 10 km cada) e uma estrada não pavimentada (DF-205, 09 km). Essas rodovias delimitam Estação Ecológica de Águas Emendadas-ESECAE (10 000 ha), Unidade de

Conservação reconhecida pela UNESCO como área central da Reserva da Biosfera do Cerrado (Figura 1).

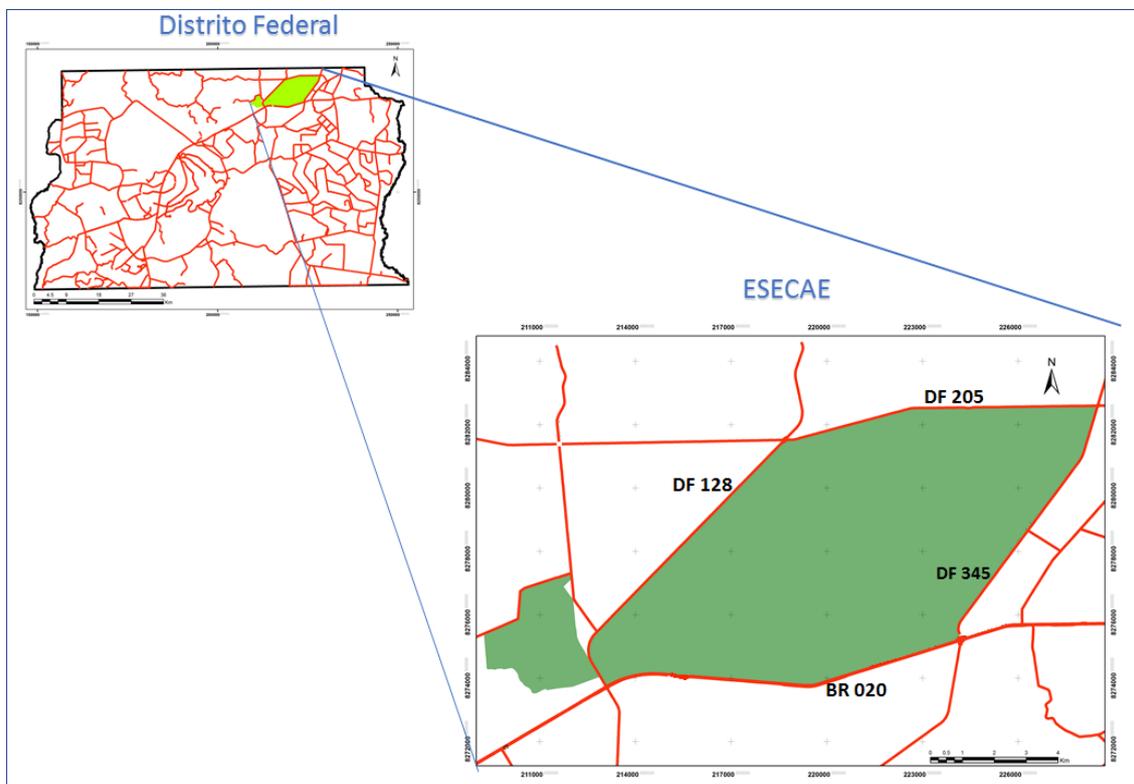


Figura 2. A grande área preenchida em verde é a Estação Ecológica Águas Emendadas – ESECAE.

As campanhas de amostragem foram realizadas duas vezes por semana, entre abril de 2010 e março de 2015, por uma equipe de três observadores em um carro a 50 km/h. Os animais encontrados foram fotografados, tiveram suas coordenadas geográficas registradas usando um GPS, com precisão de 5 m. Os animais registrados foram identificados no menor nível taxonômico possível e imediatamente removidos da pista para evitar possíveis recontagens.

Os registros foram organizados em uma tabela na qual cada linha representa uma ocorrência. Cada registro de atropelamento tem informação espacial (Coordenadas Geográficas) e temporal (dia de registro). No entanto, para cada data de coleta de dados

em que não ocorreu nenhum registro foi inserida uma linha na tabela, para não perder a informação temporal da ausência de registros.

#### *Análise de dados*

##### *Simulação de conjuntos e subconjuntos de dados*

Inicialmente, os dados registrados durante os cinco anos de coletas foram reamostrados 21 vezes, por meio de sorteio com reposição de cada registro (para este sorteio utilizou-se o número de identificação de cada linha da tabela de dados brutos). Para esta simulação foram mantidos os dias de coleta em que não ocorreram registros de atropelamento, o que torna os novos conjuntos de dados mais realistas. Após a reamostragem, cada conjunto de dados simulados foi organizado por classe: aves, répteis e mamíferos. Além de um grupo formado por todos os registros denominado “Todos”, que inclui também os anfíbios e animais domésticos. O agrupamento dos resultados foi feito desta maneira para utilizar conjuntos de dados que originalmente apresentam tamanhos distintos, sendo dois grupos com muitos registros, 2.345 eventos registrados para o grupo “Todos” e 1.271 para aves e outros dois com poucos registros, sendo 370 para répteis e 282 para mamíferos. Os anfíbios não foram analisados separadamente pois o número de indivíduos registrados (150) foi muito pequeno para o método, gerando muito erros, como diversas divisões por zero. No entanto, optou-se por incluir os anfíbios, assim como os animais domésticos, no grupo “Todos” para que este grupo refletisse uma maior diversidade do uso das rodovias, em contraste aos demais grupos que apresentam menor diversidade destes usos.

Para criar subconjuntos de diversos tamanhos foram realizados oito padrões de divisão temporal com os dados de coleta dos cinco anos (484 coletas). As divisões não resultaram em sobreposição de períodos para evitar autocorrelação dos resultados. A divisão foi feita da seguinte forma: no primeiro, o tempo total foi dividido em três

períodos semelhantes, no segundo, foram quatro períodos semelhantes e assim sucessivamente até o oitavo com dez períodos semelhantes, conforme tabela 1.

Tabela 1. Padrões de divisões do total de dados para criação dos subconjuntos

Nº de divisões	nº de campanhas por subconjunto
3	2 subconjuntos (161) e 1 subconjunto (162)
4	4 subconjuntos (121)
5	4 subconjuntos (97) e 1 subconjunto (96)
6	5 subconjuntos (81) e 1 subconjunto (79)
7	6 subconjuntos (69) e 1 subconjunto (70)
8	4 subconjuntos (61) e 4 subconjunto (60)
9	8 subconjuntos (54) e 1 subconjunto (52)
10	6 subconjuntos (48) e 4 subconjunto (49)

Para avaliar se a curva proposta neste trabalho teria aplicabilidade durante o acúmulo de dados, aplicou-se a mesma metodologia para subconjuntos de dados que representam diferentes momentos durante os cinco anos de coletas de dados. A tabela 2 indica o período e o número de campanhas de coleta de dados que cada subconjunto representa, bem como o número de atropelamentos registrados realizados para cada período.

Tabela 2. Conjuntos de dados utilizados para avaliar a construção da curva de precisão de tempos diferentes da coleta de dados

Período de coletas	nº de campanhas	nº de registros
01/04/2010 - 17/03/2011	94	450
01/04/2010 - 18/07/2011	128	621
01/04/2010 - 05/03/2012	187	1024
01/04/2010 - 15/10/2012	247	1234
01/04/2010 - 09/05/2013	302	1563
01/04/2010 - 23/01/2014	371	1856
01/04/2010 - 22/09/2014	435	2073
01/04/2010 - 30/04/2015	484	2345

### *Análise de hotspot de atropelamento*

O algoritmo de identificações de *hotspot* usado neste artigo é baseado no método *Hotspot 2D* (Coelho et al., 2012, 2014). A fórmula *Hotspot 2D* foi utilizada para encontrar a agregação de eventos em relação aos pontos "i" distribuídos regularmente tanto para uma vizinhança usando um raio 'r' -  $H(r)$  - como para uma janela de observação, maior que a vizinhança, usando um raio 'w' -  $H(w)$ , ambas as áreas centradas em cada ponto 'i'. O valor  $H(r)$  para cada ponto 'i' foi dividido pelo equivalente  $H(w)$  para obter o valor de  $H(r/w)$ .

Os resultados observados foram comparados com um modelo nulo formado por 100 distribuições de Poisson e os valores observados para um ponto "i" maior que o percentile 97,5% classificam esse ponto como *hotspot*. Para este experimento foram utilizados os valores de 'r' igual a 200m e 'w' igual a 2.800m.

### *Taxa de Precisão*

Para cada um dos quatro grupos usados neste experimento foram gerados 1.134 subconjuntos e cada subconjunto foi comparado com a simulação que lhe deu origem. Desta comparação foi obtido o valor de precisão entre a classificação de *hotspot* gerada pelo subconjunto em relação ao conjunto de referência. O valor de precisão representa o percentual dos *hotspots* propostos pela análise do subconjunto que coincidem com os *hotspots* indicados pela análise do conjunto total de dados do qual o subconjunto foi retirado. Para esta análise utiliza-se a premissa de que o conjunto total de dados gera a melhor classificação de *hotspot* possível.

### *Correlação e Regressão Quantílica*

Inicialmente foram realizados os testes de Correlação de Pearson para avaliar se existe e qual o sentido da relação entre as variáveis Precisão e Tamanho da amostra. Foram gerados gráficos de dispersão entre os valores de precisão obtidos em função do logaritmo do tamanho das amostras e realizada análise de resíduos para investigar a adequabilidade de um modelo de regressão. Para todas as distribuições testadas, o pressuposto de homocedasticidade foi excluído. Assim, os estimadores obtidos por meio do método dos mínimos quadrados deixaram de ser adequados.

Para situações como a dos dados deste trabalho, Koenker & Basset Jr (1978) propuseram uma abordagem mais geral, a Regressão Quantílica (RQ), que se baseia no método dos erros absolutos ponderados. Nesta metodologia é realizada uma ponderação na minimização dos erros para se estimar os diversos quantis de interesse. O uso de tal modelo torna os resultados mais robustos ao observar a resposta de cada quantil e utiliza a mediana condicional como medida de tendência central (ver Cade & Noon 2003). A partir das regressões quantílicas é possível observar a resposta de cada quantil, enquanto na regressão por Mínimos Quadrados Ordinários (MQO) há apenas uma reta de regressão em torno da média, podendo ofuscar importantes efeitos distributivos (Marioni *et al.* 2006). A Regressão Quantílica pode ser utilizada para construir previsões e intervalos de tolerância sem a necessidade de assumir erros de distribuição paramétrica e sem especificar o quanto as variâncias heterogênicas influenciam nas mudanças das médias (Cade & Noon 2003).

Assim, o modelo de regressão quantílica foi utilizado para os resultados das análises de precisão. Para tal, foram estimados os quantis 2.5%, 50% e 97,5%. Estes valores representam três cenários relativos à eficiência do método de classificação de *hotspot* utilizado para construção da curva: 1) Cenário pessimista – Quantil 2.5% (Q2.5)

- Outras variáveis não consideradas neste teste implicam na baixa eficiência da aplicação do método de classificação de *hotspots*, 2) Cenário intermediário – Quantil 50% (Q50) – Representa um grau intermediário entre o cenário pessimista e otimista no que se refere à influência de variáveis não consideradas neste estudo e na aplicação do método de classificação de *hotspots*: 3) Cenário otimista – Quantil 97.5% (Q97.5) - Outras variáveis não consideradas não geram efeito significativo na aplicação do método de classificação de *hotspots*, de forma que ele apresenta alta eficiência. Os coeficientes das retas para os três quantis foram obtidos e com eles se estabeleceu a Curva de Precisão para cada grupo de análise. Todas as análises foram realizadas no programa R (R Core Team, 2017), com o pacote *quantreg* (Koenker 2017). Com este pacote foi possível gerar as análises de regressão quantílicas bem como fazer a predição dos valores de precisão com base no tamanho da amostra original de cada um dos grupos para se estimar a precisão atual da proposição de *hotspot*.

## RESULTADOS

A média do número de registros simulados para o grupo “Todos” foi de  $2.345 \pm 05$ , para Aves foi de  $1.265 \pm 28$ , para Répteis foi  $370 \pm 13$  e para Mamíferos  $288 \pm 18$ . A variação no tamanho dos grupos se deu pelo fato da reamostragem ser feita incluindo dias de coletas nos quais não houveram registros, assim, para alguns conjuntos a simulação resultou em maior sorteio destes dias do que para outros. Os quatro grupos analisados apresentaram correlação positiva entre a Taxa de Precisão e a abundâncias das amostras: para o grupo “Todos” ( $p:0.636$ ,  $p\text{-value}<0.05$ ); para Aves ( $p:0.628$ ,  $p\text{-value}<0.05$ ); para Répteis ( $p:0.459$ ,  $p\text{-value}<0.05$ ) e Mamíferos ( $p:0.578$ ,  $p\text{-value}<0.05$ ). A força de correlação, usando o guia que Evans (1996), foi considerada moderada para os grupo Répteis e Mamíferos e forte para Aves e para o grupo “Todos”.

A figura 2 ilustra os gráficos de dispersão dos valores da Precisão encontrada em função do logaritmo da abundância das amostras. Para Mamíferos os coeficientes angulares das três retas de regressão são significativamente semelhantes (Q2.5 –  $0.15 \pm 0.011$ ; Q50 –  $0.152 \pm 0.007$ ; Q97.5 –  $0.151 \pm 0.017$ ). Para os demais grupos, os coeficientes angulares no cenário Q97.5 foram menores que os demais. Para os grupos “Todos” e Aves, os coeficientes angulares nos cenários Q2.5 e Q50 foram estatisticamente semelhantes e ambos maiores que o Q97.5, sendo para “Todos” (Q2.5 –  $0.141 \pm 0.013$ ; Q50 –  $0.127 \pm 0.005$ ; Q97.5 –  $0.109 \pm 0.015$ ) e para Aves (Q2.5 –  $0.147 \pm 0.012$ ; Q50 –  $0.141 \pm 0.007$ ; Q97.5 –  $0.082 \pm 0.02$ ). Para répteis os três cenários apresentaram coeficientes angulares distintos, sendo o Q2.5 maior que os demais e o Q97.5 menor que os demais (Q2.5 –  $0.159 \pm 0.017$ ; Q50 –  $0.102 \pm 0.006$ ; Q97.5 –  $0.041 \pm 0.015$ ).

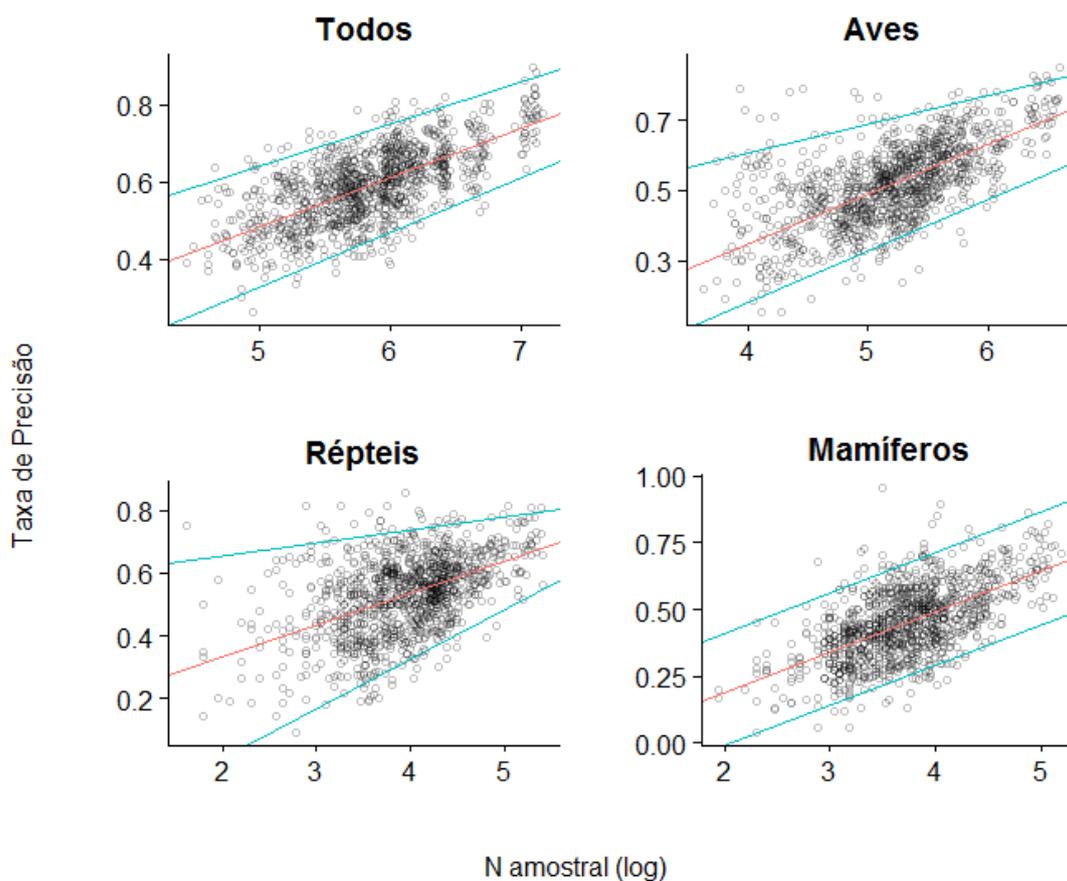


Figura 3. Análise da correlação entre a taxa de Precisão na análise de um em um subconjunto e o logaritmo da sua abundância, a linhas em vermelho representam a regressão para o Q50 e as linhas em azul são para os Q2,5 e Q97,5.

A figura 3 ilustra graficamente as Curvas de Precisão para cada grupo estudado considerando os três cenários. É possível verificar, para répteis, uma convergência entre os três cenários quanto ao valor de precisão em relação ao tamanho da amostra. Para aves pode-se observar que esta convergência também ocorre de forma mais branda entre os Q50 e Q97,5.

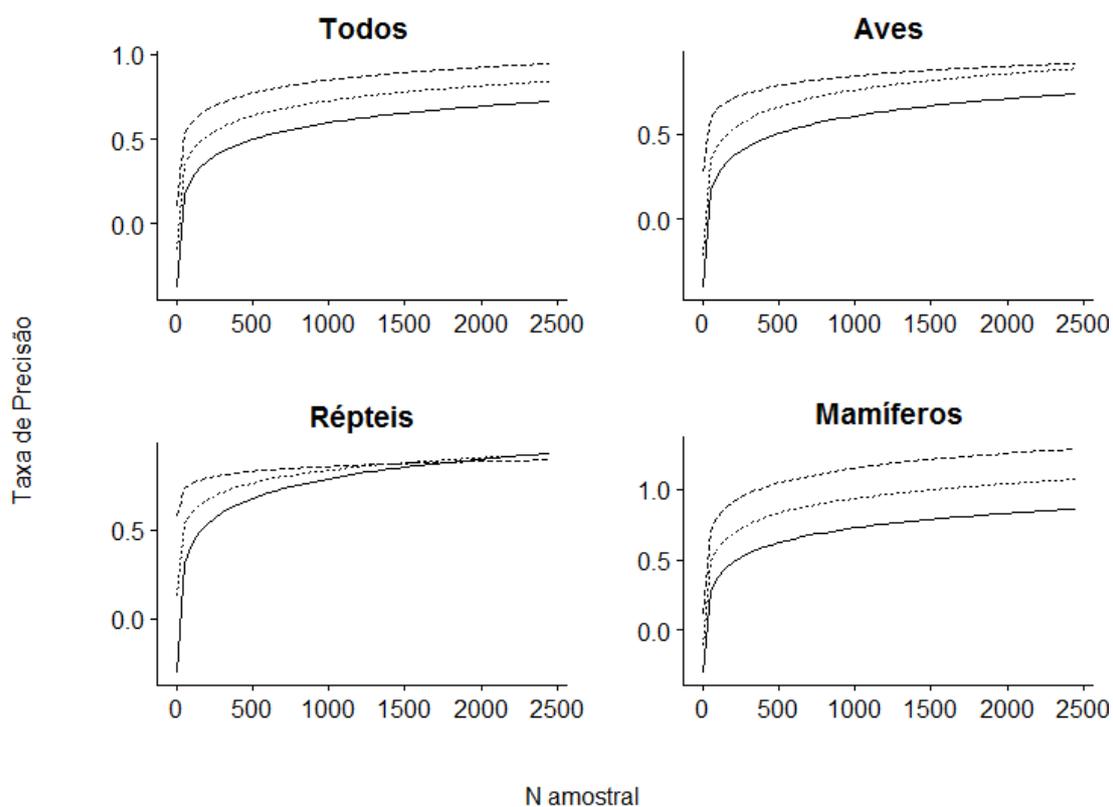


Figura 4. Curva de Precisão na indicação de *hotspots* de atropelamento de fauna para os quatro grupos estudados. As linhas sólidas, pontilhadas e tracejadas representam, respectivamente, os cenários Q2.5, Q50 e Q97.5.

A tabela 3 ilustra os resultados do cálculo da precisão na classificação de *hotspots* para o número de registros de cada grupo.

Tabela 3 Estimativa da precisão atual da indicação de *hotspots* de atropelamento de fauna em consideração ao número de registros de cada grupo.

Grupo	Nº de registros	Q2.5	Q50	Q97.5
Todos	2.345	0,717	0,835	0,942
Aves	1.271	0,643	0,793	0,863
Répteis	370	0,630	0,733	0,815
Mamíferos	282	0,540	0,744	0,962

A figura 4 indica a evolução da precisão da análise do grupo “Todos” no decorrer do tempo. É possível verificar que com exceção das duas primeiras análises,

existe uma tendência à precisão aumentar com o acúmulo de dados. Considerando 70% de precisão como limite aceitável, a partir de amostra inferior a 450 já se atingiria este limite para o cenário Q97,5. Para o cenário intermediário este limite seria alcançado com 1.234 indivíduos e, se fosse utilizado o cenário mais pessimista, a eficiência da análise ultrapassaria o limite proposto após o acúmulo de 2073 registros.

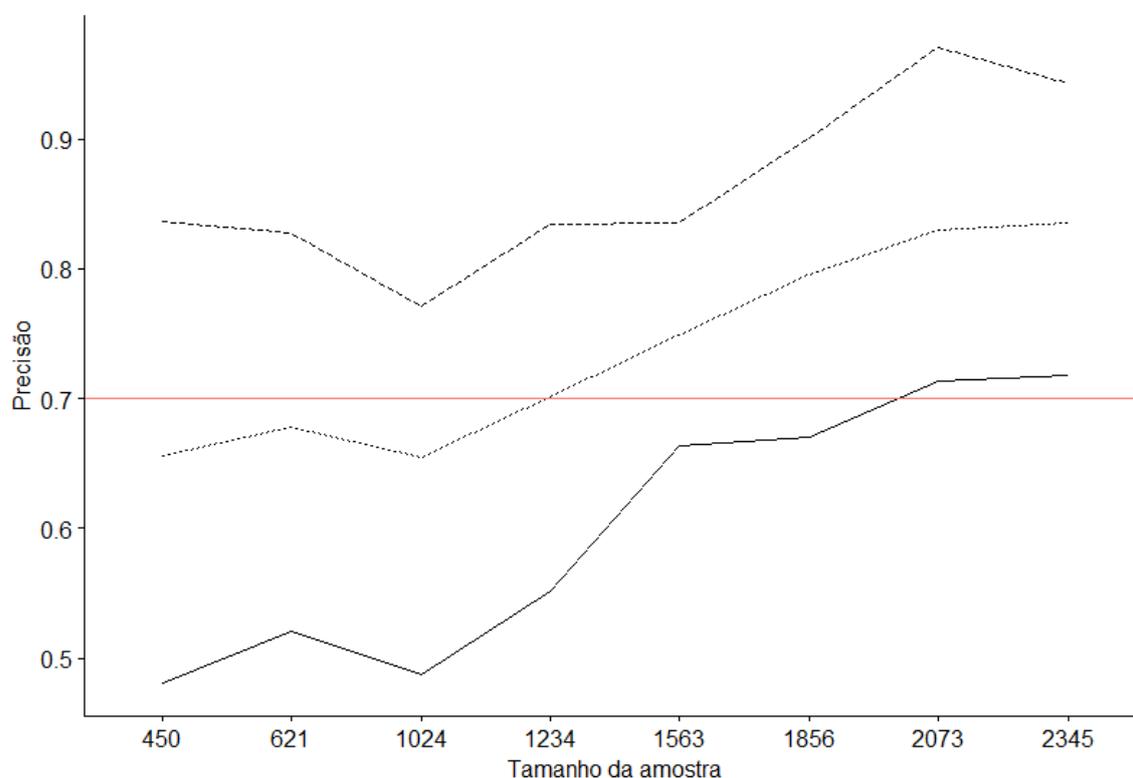


Figura 5. Valores da Precisão de indicação de *hotspots* para o grupo “Todos” se a curva de precisão fosse elaborada em momentos diferentes da coleta de dados. Os números no eixo x representam o tamanho da amostra no momento da confecção de cada curva. As linhas sólidas, pontilhadas e tracejadas representam, respectivamente, os cenários Q2.5, Q50 e Q97.5, e a linha sólida vermelha indica o valor de precisão igual a 0.7.

## DISCUSSÃO

Observamos que os resultados deste experimento indicaram que a precisão na classificação dos *hotspots* é positivamente relacionada com o tamanho da amostra utilizada para esta classificação. Esta correlação permite a identificação da amostra suficiente para que a identificação dos *hotspots* de atropelamento tenha um grau de precisão desejada pelo pesquisador para que se possa propor medidas de mitigação. Pesquisas que avaliaram a persistência dos *hotspots* no decorrer do tempo (Mountrakis & Gunson 2009, Santos *et al.* 2017) ou a diminuição na precisão em função da diminuição da frequência de coletas (Santos *et al.* 2015) podem ter a redução do tamanho das amostras em subconjuntos como explicação para as variações espaciais no perfil de agregações.

A divisão temporal do conjunto de atropelamentos pode ser entendida como a redução do tamanho de uma unidade amostral. Unidades amostrais menores tendem a ser mais variáveis entre si e quanto maior a variância, maior é o número de unidades amostrais necessárias para uma mesma precisão da estimativa de uma média (Cochran 1977). Para corrigir este problema, no presente trabalho, foram realizadas simulações com reamostragens com reposição, aumentando o número de unidades amostrais para ampliar a precisão da estimativa.

Além de apresentar a taxa de crescimento da precisão em função do aumento da amostra permitindo a estimativa da precisão dos dados já coletados, a Curva de Precisão permite realizar algumas inferências. O afunilamento da curva observado para aves e , de forma mais acentuada, para répteis, significa a redução da variância da precisão da forma escolhida para a análise dos *hotspots* ( método *Windowed Hotspot*, com as escalas  $r=200m$  e  $w=2.800m$ ).

De acordo com Cochran (1977), unidades amostrais menores tendem a ser mais variáveis entre si e quanto maior a variância. Assim, esperava-se que a variância dos resultados de precisão obtidos entre subconjuntos de tamanhos semelhantes fosse menor, quanto maior fosse o tamanho destes subconjuntos. Esta variância foi avaliada pela comparação dos coeficientes angulares das retas das regressões quantílicas, sendo a convergência entre os cenários otimista e pessimista o indicativo de que o aumento tamanho da amostra causa redução da variância. Para o grupo Mamíferos os coeficientes angulares foram estatisticamente semelhantes ( $Q2.5 - 0.15 \pm 0.011$  e  $Q97.5 - 0.151 \pm 0.017$ ), indicando que o aumento do tamanho da amostra não implica na redução da variância. Para os demais grupos, a reta  $Q97.5$  apresenta ângulo menor que a  $Q2.5$  indicando convergência entre elas. Observa-se, no entanto, que quanto maior a diferença entre os valores dos coeficientes angulares, mais evidente é a convergência. Para o grupo “Todos” a convergência é sutil, visto que a diferença entre os coeficientes é muito pequena, sendo  $0.141 \pm 0.013$  para  $Q2.5$  e  $0.109 \pm 0.015$  para  $Q97.5$ . Para aves estes coeficientes foram um pouco mais distantes,  $0.147 \pm 0.012$  para  $Q2.5$  e  $0.082 \pm 0.02$  para  $Q97.5$ , enquanto que para répteis a diferença foi ainda maior:  $0.159 \pm 0.017$  para  $Q2.5$  e  $0.041 \pm 0.015$  para  $Q97.5$ .

Os grupos “Todos” e aves, que são os grupos com maior quantidade de eventos apresentaram resultados semelhantes. Foram os maiores valores de correlação entre precisão e tamanho da amostra, e apresentaram, mesmo que sutilmente, diminuição da variância das análises de precisão com o aumento do tamanho da amostra. No entanto, os dois grupos que tiveram a menor quantidade de registro, , répteis e mamíferos, foram bem diferentes no que se refere ao efeito do tamanho da amostra na variância das análises de precisão. O resultado obtido para répteis pode ser explicado pela quantidade de eventos de atropelamento do grupo dentro de regiões consideradas como *hotspot*.

Este valor pode ser grande o suficiente, ao ponto do incremento no tamanho dos subconjuntos aumentar a chance destes eventos terem sido selecionados. Ou ainda, a quantidade de *hotspots* pode ser pequena ao ponto das amostras pequenas não terem necessidade em identifica-los. Porém, o aumento da amostra faz com que esses poucos *hotspots* se estabeleçam mais rapidamente quando comparado com os demais grupos estudados.

A construção de diferentes curvas para o grupo “Todos”, em tempos diferentes da coleta de dados, simula uma situação real de testar a suficiência amostral no decorrer das coletas para avaliar o momento de finalizá-las. Os resultados obtidos demonstram que a metodologia de cálculo da curva é confiável para ser usada para esta decisão, uma vez que quando considerado o mesmo grupo de análise, o aumento da amostra resulta em aumento da precisão. Se o limite aceitável de precisão fosse 70% poderia se concluir que, independente do cenário considerado, a amostra utilizada neste trabalho estaria em uma situação de desperdício de recursos, pois com amostras menores e com sete meses de antecedência ao término das coletas, a suficiência amostral já havia sido alcançada.

Verificar que a precisão aumentou com o acúmulo de dados para o grupo “Todos” também aumenta a relevância dos valores de precisão obtidos para os mamíferos e répteis. Os altos valores identificados para estes grupos, mesmo com amostras muito menores que os demais, não são apenas uma decorrência matemática do método. Estes resultados podem corresponder a uma maior probabilidade destes grupos apresentarem maior especificidade no uso do ambiente no entorno das rodovias, ao ponto de um número significativamente menor de indivíduos registrados já indicar alta precisão dos *hotspots* de atropelamento.

Este trabalho, além de apresentar esta inovação para o protocolo de estudos, também demonstra uma forma de utilização de técnicas de simulação com

reamostragem e *Bootstrap*, além da extrapolação com o uso da regressão quantílica, para predição de eficiência de qualquer tamanho amostral desejado.

A Curva de Precisão apresentada neste trabalho não é uma metodologia rígida. Os valores para os quantis da regressão quantílica podem ser alterados, conforme o entendimento do pesquisador. Além disso, a regra para finalizar as coletas também pode variar, podendo-se utilizar os valores médios ou os extremos, no entanto, para isso, é preciso conhecer a eficiência da metodologia utilizada para definição dos *hotspots*. Neste estudo, para se obter um resultado bastante conservador no que se refere à eficiência da análise de *hotspot*, considerou-se um grau de eficiência muito baixo para o cenário pessimista.

Finalmente, ao se analisar a tabela com os valores de precisão referente ao conjunto de registros reais para cada grupo, é possível observar que atualmente os quatro grupos apresentam uma confiança na proposição dos *hotspots* superior a 70% no cenário intermediário. Mesmo no cenário mais pessimista, com exceção dos mamíferos, esses valores seriam superiores a 60%.

## CONCLUSÃO

Esse estudo demonstrou que o aumento da precisão ocorre de forma logarítmica em função do aumento do tamanho da amostra, possibilitando o uso desta relação de forma análoga à curva do coletor a fim de orientar o pesquisador sobre a qualidade dos dados coletados para se tomar a decisão de finalizar a coleta em estudos de atropelamento de fauna em estradas.

Além disso, inferir a precisão das análises oriundas de uma determinada amostra de atropelamentos de fauna diminui a incerteza sobre a proposição de medidas mitigadoras. Permitindo, também, que se estabeleça valores de precisão elevados como

critério para determinadas intervenções. Tendo em vista que as medidas mitigatórias, normalmente, são onerosas e com longa vida útil, o modelo de análise proposto neste trabalho aumenta o poder de negociação para implantação de tais medidas.

É importante continuar a exploração acadêmica de melhorias nos protocolos de amostragem e de análises de dados. Futuros estudos devem buscar avaliar se a própria curva apresenta dinâmica em relação ao tamanho do conjunto de dados referência para simulações, ou se o formato da curva ou dos limites de tolerância podem guardar alguma relação com características biológicas do grupo estudado, ou ainda se esta metodologia de análise é aplicável para outras métricas de identificação de *hotspots* de atropelamento, como os métodos de Malo e *Hotspot 2D*.

#### REFERÊNCIAS

- Attademo, A.M., Peltzer, P.M., Lajmanovich, R.C., Elberg, G., Junges, C., Sanchez, L.C. & Bassó, A. (2011). Wildlife vertebrate mortality in roads from Santa Fe Province, Argentina. *Revista Mexicana de Biodiversidad*, **82**, 915–925.
- Bager, A. & da Rosa, C.A. (2011). Influence of sampling effort on the estimated richness of road-killed vertebrate wildlife. *Environmental Management*, **47**, 851–858.
- Cade, B.S. & Noon, B.R. (2003). A gentle introduction to quantile regression for ecologists In a nutshell : *Frontiers in Ecology and the Environment*, **1**, 412–420.
- Cochran, W.G. (1977). *Sampling Techniques*, Thirdn. John Wiley and Sons Ltd.
- Coelho, I.P., Teixeira, F.Z., Colombo, P., Coelho, A.V.P. & Kindel, A. (2012). Anuran road-kills neighboring a peri-urban reserve in the Atlantic Forest, Brazil. *Journal of Environmental Management*, **112**, 17–26.
- Collinson, W.J., Parker, D.M., Bernard, R.T., Reilly, B.K. & Davies-Mostert, H.T.

- (2014). Wildlife road traffic accidents: a standardized protocol for counting flattened fauna. *Ecol Evol*, **4**, 3060–3071.
- Costa, A.S., Ascensão, F. & Bager, A. (2015). Mixed sampling protocols improve the cost-effectiveness of roadkill surveys. *Biodiversity and Conservation*, **24**, 2953–2965.
- Eberhardt, E., Mitchell, S. & Fahrig, L. (2013). Road kill hotspots do not effectively indicate mitigation locations when past road kill has depressed populations. *Journal of Wildlife Management*, **77**, 1353–1359.
- Evans, J. D. (1996) Straightforward statistics for the behavioral sciences. Brooks/Cole Publishing; Pacific Grove, Calif
- Garrah, E., Danby, R.K., Eberhardt, E., Cunnington, G.M. & Mitchell, S. (2015). Hot Spots and Hot Times: Wildlife Road Mortality in a Regional Conservation Corridor. *Environmental Management*, **56**, 874–889.
- Garriga, N., Santos, X., Montori, A., Richter-Boix, A., Franch, M. & Llorente, G.A. (2012). Are protected areas truly protected? The impact of road traffic on vertebrate fauna. *Biodiversity and Conservation*, **21**, 2761–2774.
- Koenker, R. (2017). quantreg: Quantile Regression. *R package version 5.33*, <https://CRAN.R-project.org/package=quantreg>.
- Koenker, R. & Basset Jr, G. (1978). Regression Quantiles. *Econometrica*, **46**, 33–50.
- Marioni, S., Vale, V.D.A., Salgueiro, F. & Freguglia, S. (2006). Uma Aplicação de Regressão Quantílica para Dados em Painel do PIB e do Pronaf. *Revista de Economia e Sociologia Rural*, **54**, 221–242.
- Mountrakis, G. & Gunson, K. (2009). Multi-scale spatiotemporal analyses of moose–vehicle collisions: a case study in northern Vermont. *International Journal of Geographical Information Science*, **23**, 1389–1412.

- da Rosa, C.A. & Bager, A. (2012). Seasonality and habitat types affect roadkill of neotropical birds. *Journal of Environmental Management*, **97**, 1–5.
- Santos, S.M., Carvalho, F. & Mira, A. (2011). How long do the dead survive on the road? Carcass persistence probability and implications for road-kill monitoring surveys. *PLoS One*, **6**, e25383.
- Santos, S.M., Marques, J.T., Lourenço, A., Medinas, D., Barbosa, A.M., Beja, P. & Mira, A. (2015). Sampling effects on the identification of roadkill hotspots: Implications for survey design. *Journal of Environmental Management*, **162**, 87–95.
- Santos, R.A.L., Santos, S.M., Santos-Reis, M., Picanço de Figueiredo, A., Bager, A., Aguiar, L.M. de S. & Ascensão, F. (2016). Carcass Persistence and Detectability: Reducing the Uncertainty Surrounding Wildlife-Vehicle Collision Surveys. *PLoS One*, **11**, 1–15.
- Santos, R.A. L., Ascensao, F., Ribeiro, M.L., Bager, A., Santos-Reis, M. & Aguiar, L.M.S. (2017). Assessing the consistency of hotspot and hot- moment patterns of wildlife road mortality over time. *Perspectives in Ecology and Conservation*, <http://dx.doi.org/10.1016/j.pecon.2017.03.003>.
- Seo, C., Thorne, J.H., Choi, T., Kwon, H. & Park, C.-H. (2013). Disentangling roadkill: the influence of landscape and season on cumulative vertebrate mortality in South Korea. *Landscape and Ecological Engineering*, **11**, 87–99.
- Skórka, P., Lenda, M., Moroń, D., Martyka, R., Tryjanowski, P., Sutherland, W.J., Moron, D., Martyka, R., Tryjanowski, P. & Sutherland, W.J. (2015). Biodiversity collision blackspots in Poland: Separation causality from stochasticity in roadkills of butterflies. *Biological Conservation*, **187**, 154–163.
- Slater, F.M. (2002). An assessment of wildlife road casualties – the potential

discrepancy between numbers counted and numbers killed. *Web Ecology*, **3**, 33–42.

Teixeira, F.Z., Coelho, A.V.P., Esperandio, I.B. & Kindel, A. (2013). Vertebrate road mortality estimates: Effects of sampling methods and carcass removal. *Biological Conservation*, **157**, 317–323.

## CONSIDERAÇÕES FINAIS

Nesta dissertação foram apresentadas duas propostas de soluções para dois problemas metodológicos. A primeira proposta é um método de análise que diminui o efeito da não homogeneidade de primeira ordem na utilização de uma métrica que apresenta como premissa a homogeneidade. A quebra desta premissa deve ser considerada análoga à utilização de ANOVA para dados não oriundos de uma distribuição normal ou não homocedásticos. Por isso, o método que utiliza janelas de análise, para diminuir a heterogeneidade dos fatores relacionados ao atropelamento deve ser explorado e desenvolvido. Este método indicou um melhor desempenho na identificação de *hotspots* em dados artificialmente controlados e também demonstrou menor sensibilidade à variações de densidade regional de atropelamento.

A construção de métodos de análises necessita de constante revisão por diversos pesquisadores conhecedores do tema, assim, o primeiro capítulo desta dissertação apresenta um desafio aos estudiosos em Ecologia de Estradas para torná-lo cada vez mais robusto.

A segunda proposta está ligada à resposta para a pergunta que todos os pesquisadores de qualquer área da ciência fazem: já posso parar de coletar? As análises de suficiência amostral são bastante utilizadas de forma rotineira em vários campos da ciência, mesmo com algumas críticas para algumas delas. No método proposto, o momento para finalizar a coleta vira uma consequência da capacidade de coletar dados que tragam taxas de precisão previamente entendidas como razoáveis. Neste sentido, a forma e velocidade de deslocamento na procura das carcaças, a frequência de amostragens, ou o período total em que as amostragens serão realizadas ficam a critério ou à capacidade operacional do pesquisador.

O tempo necessário de execução de um diagnóstico ambiental ou de um programa de monitoramento é a principal dúvida tanto para analistas ambientais, ao se aprovar ou não um Estudo de Impacto Ambiental, como para os realizadores dos estudos e principalmente para os gestores da política de transporte, que não querem esperar o tempo necessário para se conhecer o efeito que se deseja mitigar. Quando o pesquisador não consegue responder para o tomador de decisão qual é o ‘tempo necessário’, qualquer tempo se torna suficiente. Assim, o foco do segundo capítulo foi trazer à tona a necessidade de se estabelecer um método científico mas aplicável para que todos os envolvidos nas análises dos efeitos negativos de uma rodovia possam conhecer o “tempo adequado” de trabalho.

É importante verificar que as propostas oriundas dos dois capítulos se complementam, pois, a primeira tem por objetivo aumentar a confiança de que locais indicados para mitigação dos atropelamentos de fauna, enquanto a segunda proposta busca acelerar o processo de mitigação ao propor uma forma de reconhecer a atual precisão quanto à indicação destes locais até se atingir a precisão desejada.

# ANEXO I - SCRIPTS DA PROGRAMAÇÃO REALIZADA PARA A DISSERTAÇÃO

## Capítulo 01

### A 1 - LINHAS DE COMANDO PARA EXTRAIR TRAÇADO DE RODOVIA A PARTIR DE UM ARQUIVO \*.KML E CRIAR SEGMENTOS PARA ANÁLISE DE *HOTSPOT*

As linhas de comando a seguir extraem o traçado de qualquer rodovia obtidas a partir do Google Earth (arquivo .kml) e cria o objeto “rodovia” que é utilizado na função de análise de *hotspot* “s.hot”.

```
library("rgdal", "spatstat")  
lyr <- ogrListLayers('rodovia_esecae.kml')  
road2 <- readOGR ("rodovia_esecae.kml", layer=lyr)  
road2b <- spTransform(road2, CRS("+init=epsg:31983"))  
ESECAE.road <- as(road2b, "psp")  
pont.rod.esecae<-pointsOnLines(ESECAE.road,eps=1,shortok = FALSE)  
rodovia<-as.data.frame(pont.rod.esecae)  
colnames(rodovia)<-c('x','y')
```

As linhas de comando a seguir cria os segmentos para análises de *hotspot* e o objeto “pont.rod” que são os pontos médios destes segmentos, nos quais são centrados os raios de análise utilizados para análise de *hotspot* pela função “s.hot”.

```
window.rod <- riplas(rodovia,shape = "rectangle")  
window.rod <- dilation.owin(window.rod,200)  
rod.ppp <- ppp(rodovia$x,rodovia$y>window.rod)  
pontos.para.linnet <- ppp(rodovia$x[seq(1,length(rodovia$x),by=100)],  
rodovia$y[seq(1,length(rodovia$y),by=100)],window=window.rod)
```

```
edge <- cbind(1:(pontos.para.linnet$n-1),2:pontos.para.linnet$n)
linnet.rod <- linnet(pontos.para.linnet,edges = edge)
rod.lines <- as.psp(linnet.rod)
pont.rod.ppp<-midpoints.psp(rod.lines)
pont.rod<-cbind(x=pont.rod.ppp$x,y=pont.rod.ppp$y)
```

## A 2 - LINHAS DE COMANDO PARA SIMULAR A DISTRIBUIÇÃO DE ATROPELAMENTOS DE FAUNA, COM *HOTSPOTS* PREDETERMINADOS E QUATRO TIPOS DE RUÍDOS

As linhas de comando a seguir determinam os 06 trechos preestabelecidos como *hotspots*, bem como suas extensões e o lambda (densidade) de cada um. O objeto “lista.de.trechos” é utilizado a diante para simulação dos eventos de atropelamento.

```
trechos.modelo<-c(11,123, 196, 242, 310, 89)
fim.trechos.modelo<-trechos.modelo+c(6,9,2,9,2,6)
trechos.modelo<-mapply(function(x,y) x:y,trechos.modelo,fim.trechos.modelo)
lambda.por.trecho<-c(0.12,0.1,0.18,0.1,0.18,0.12)
lista.de.trechos<-lapply(trechos.modelo, function(x) rod.lines[x])
```

As linhas de comando a seguir simulam os 04 perfis de probabilidade de ocorrência de um atropelamento, os objetos “prob.areas.het”, “prob.areas.het2”, “prob.areas.het3” e “prob.areas.het4” são utilizados a diante para simulação dos eventos de atropelamento.

```
prob.areas.het<-{sin(seq(-pi*2,pi,length.out = nrow(rodovia)))+abs(min(sin(seq(-
pi*2,pi,length.out = nrow(rodovia))))))/2
prob.areas.het2<-{sin(seq(pi*4,pi,length.out =
nrow(rodovia)))+abs(min(sin(seq(pi*4,pi,length.out = nrow(rodovia))))))/2
prob.areas.het3<-{sin(seq(-pi*3,pi,length.out =
nrow(rodovia)))+abs(min(sin(seq(pi*4,pi,length.out = nrow(rodovia))))))/2
prob.areas.het4<-{sin(seq(pi*5,pi,length.out =
nrow(rodovia)))+abs(min(sin(seq(pi*4,pi,length.out = nrow(rodovia))))))/2
```

As linhas de comando a seguir usam os valores das densidades de cada *hotspot* para simular uma distribuição de Poisson para cada um dos 06 trechos.

```

eventos.agreg<-mapply(function(x,y)
                        rpoisppOnLines(x,y),lambda.por.trecho,lista.de.trechos,SIMPLIFY = F)
eventos.agreg<-superimpose(eventos.agreg[[1]],
                            eventos.agreg[[2]],
                            eventos.agreg[[3]],
                            eventos.agreg[[4]],
                            eventos.agreg[[5]],
                            eventos.agreg[[6]])

```

As linhas de comando a seguir simulam 04 perfis de distribuição heterogênea de atropelamentos utilizando 04 perfis diferentes de probabilidade de ocorrência dos eventos no decorrer da rodovia.

```

eventos.disp<-rpoisppOnLines(0.06,rod.lines)
eventos.disp.het1<-rodovia[sample(1:nrow(rodovia),eventos.disp$n,prob=prob.areas.het),]
eventos.disp.het1<-ppp(eventos.disp.het1$x,eventos.disp.het1$y>window.rod)
eventos.modelo.het1<-superimpose(eventos.agreg,eventos.disp.het1)
pontos.het1<-coords(eventos.modelo.het1)
eventos.disp.het2<-rodovia[sample(1:nrow(rodovia),eventos.disp$n,prob=prob.areas.het2),]
eventos.disp.het2<-ppp(eventos.disp.het2$x,eventos.disp.het2$y>window.rod)
eventos.modelo.het2<-superimpose(eventos.agreg,eventos.disp.het2)
pontos.het2<-coords(eventos.modelo.het2)
eventos.disp.het3<-rodovia[sample(1:nrow(rodovia),eventos.disp$n,prob=prob.areas.het3),]
eventos.disp.het3<-ppp(eventos.disp.het3$x,eventos.disp.het3$y>window.rod)
eventos.modelo.het3<-superimpose(eventos.agreg,eventos.disp.het3)
pontos.het3<-coords(eventos.modelo.het3)
eventos.disp.het4<-rodovia[sample(1:nrow(rodovia),eventos.disp$n,prob=prob.areas.het4),]
eventos.disp.het4<-ppp(eventos.disp.het4$x,eventos.disp.het4$y>window.rod)
eventos.modelo.het4<-superimpose(eventos.agreg,eventos.disp.het4)
pontos.het4<-coords(eventos.modelo.het4)

```

```
lista.pontos.het<-list(pontos.het1, pontos.het2, pontos.het3, pontos.het4)
```

O comando a seguir cria 05 densidades totais de eventos para cada um dos tipos de ruído. Para cada tipo de ruído são mantidos os valores totais e são geradas reduções pela metade, por um terço, por um quarto e por um quinto.

```
combine<-expand.grid(1:5,1:4)
```

```
dados.sim<-apply(combine,1,function(x)
```

```
lista.pontos.het[[x[2]]][sample(1:nrow(lista.pontos.het[[x[2]]]),  
nrow(lista.pontos.het[[x[2]]])/x[1]),)]
```

### A 3 - LINHAS DE COMANDO PARA ANÁLISE DE *HOTSPOT* – HOTSPOT 2D

Função “s.hot” conta todos os “pontos” que ocorrem à uma distância “rs” de pontos regularmente distribuídos (“pont.rod”) na “rodovia” de interesse para o estudo.

```
s.hot<-function(pontos,rs,rodovia,pont.rod){  
  s1<- spam_rdist(pontos,pont.rod,rs)  
  s2<-aggregate(s1@entries,list(s1@colindices),function(x) length(x))[,2]  
  s3<-table(spam_rdist(rodovia[seq(1,nrow(rodovia),10),],pont.rod,rs)@colindices)  
  s4<-vector(mode='numeric',length = nrow(pont.rod))  
  s4[sort(unique(spam2spind(s1)$ind[,2]))]<-s2  
  k.s<-as.matrix((s4/(s3*10))*2*rs)  
  return(k.s)  
}
```

Para a análise completa de *hotspot*, pelo método *Hotspot 2D*, aplica-se esta fórmula tanto para a distribuição desejada, com o raio escolhido para análise, como para cada uma das aleatorizações geradas para o modelo nulo. Os valores observados que forem superiores ao percentil 97,5% dos valores para cada raio do modelo nulo, são convertidos em valor igual 1 e os demais, valor igual a 0, criando assim uma sequência binária de presença ou ausência de *hotspots*.

#### A 4 - LINHAS DE COMANDO PARA ANÁLISE DE *HOTSPOT* – MÉTODO WINDOWED *HOTSPOT* E *HOTSPOT* 2D

As linhas de comando a seguir são utilizadas para a análise de *hotspot* utilizando o método Windowed *Hotspot*. Do mesmo script se retiram os resultados para o método *Hotspot* 2D, neste último caso utilizam-se somente os valores observados para os raios ‘r’, enquanto que no método Windowed *Hotspot* estes valores para ‘r’, são divididos pelos valores encontrados para as janelas de observação ‘w’. O objeto “value.bin” contém todos os perfis binários de *hotspots* observados para os dados simulados. Este objeto é utilizado para identificar a taxa total de erros em relação ao perfil pré-determinado de *hotspots*.

```
r<-seq(200,2000,100)
w<-seq(1000,4000,200)
r.unique<-unique(c(r,w))
rr<-t(combn(r.unique,2))
rw<-t(combn(factor(r.unique),2))
pos<-which(rr[,1]<=2000&rr[,2]>1.33*rr[,1])
rw<-cbind(as.numeric(rw[,1]), as.numeric(rw[,2]))
rw<-rw[pos,]
rr<-rr[pos,]
nsim<-100
testing<-lapply(dados.sim[[1]],function(z)
  rply(nsim,coords(rpoisppOnLines(nrow(z)/nrow(rodovia),linnet.rod))))
road.aleat<-lapply(testing,function(a) lapply(r.unique, function(z)
  sapply(a, function(x) s.hot(x,z,rodovia,pont.rod))))
r.aleat.divid<-lapply(road.aleat,function(z) aply(rw,1,function(x) z[[x[1]]]/z[[x[2]]]))
```

```

r.aleat.max<-lapply(r.aleat.divid, function(a) lapply(a,function(z)
  apply(z,1,function(x) quantile(x,0.975,na.rm = T))
))
testing2<-lapply(dados.sim, function(x) rply(21,x[sample(1:nrow(x),nrow(x),replace = T),]))
aggreg.data<-lapply(testing2, function(a) lapply(r.unique, function(x)
  sapply(a, function(y) s.hot(y,x,rodovia,pont.rod))))
aggreg.data.divid<-lapply(aggreg.data, function(z) alply(rw,1,function(x) z[[x[1]]]/z[[x[2]]]))
value.bin<- mapply(function(a,b) mapply(function(y,z)
  apply(y,2,function(x) ifelse(x>z,1,0)),a,b,SIMPLIFY =
  F),aggreg.data.divid,r.aleat.max,SIMPLIFY = F)

```

As linhas de comando a seguir são os resultados somente para o método *Hotspot 2D*

```

aleat.old<-lapply(road.aleat,function(x) x[which(r.unique%in%r)])
r.aleat.max.r<-lapply(aleat.old, function(a) lapply(a,function(y)
aggreg.old<-lapply(aggreg.data, function(x) x[which(r.unique%in%r)])
value.bin.r<- mapply(function(a,b) mapply(function(y,z) apply(y,2,function(x)
  ifelse(x>z,1,0)),a,b,SIMPLIFY = F),aggreg.old,r.aleat.max.r,SIMPLIFY = F)

```

## A 5 - LINHAS DE COMANDO PARA ANÁLISE DE *HOTSPOT* – MÉTODO MALO

```
seg.malo <- lapply(r,function(r)
  ppp(rodovia$x[seq(1,length(rodovia$x),by=r)],
  rodovia$y[seq(1,length(rodovia$y),by=r)], window=window.rod))
edge.malo <- lapply(seg.malo, function(x) cbind(1:(x$n-1),2:x$n))
linnet.malo <- mapply(function(x,y) linnet(x,edges=y),seg.malo,edge.malo, SIMPLIFY = F)
rod.lines.malo <- lapply(linnet.malo,as.psp)
obs.malo<-lapply(testing2, function(c) lapply(rod.lines.malo, function(b)
  lapply(c, function(a) table(project2segment(ppp(a$x,a$y>window =
  window.rod),b)$mapXY))))
correct.obs.malo<-rply(20,lapply(rod.lines.malo, function(c) rply(21,rep(0,c$n))))
for(i in 1:20){
  for(j in 1:19){
    for(k in 1:21){
names(correct.obs.malo[[i]][j][k])<-as.character(1:rod.lines.malo[[j]]$n)
correct.obs.malo[[i]][j][k][names(correct.obs.malo[[i]][j][k])%in%names(obs.malo[[i]][j][k])]
<-obs.malo[[i]][j][k]
    }
  }
}
obs.malo<-lapply(correct.obs.malo, function(a) lapply(a, function(b) sapply(b,function(x) x*1)))
lim.malo<-lapply(testing, function(c) lapply(rod.lines.malo, function(b)
  lapply(c, function(a) table(project2segment(ppp(a$x,a$y>window =
  window.rod),b)$mapXY))))
correct.lim.malo<-rply(5,lapply(rod.lines.malo, function(c) rply(100,rep(0,c$n))))
for(i in 1:5){
  for(j in 1:19){
    for(k in 1:100){
names(correct.lim.malo[[i]][j][k])<-as.character(1:rod.lines.malo[[j]]$n)
```

```

correct.lim.malo[[i]][j][k][names(correct.lim.malo[[i]][j][k])%in%names(lim.malo[[i]][j][k])]
k]]<-lim.malo[[i]][j][k]
}
}
}

```

```

lim.malo<-lapply(correct.lim.malo, function(a) lapply(a, function(b) sapply(b,function(x) x*1)))
treshould.malo<-lapply(lim.malo,function(a) lapply(a, function(b)
      apply(b,1,function(x) quantile(x,0.975))))
treshould.malo<-rep(treshould.malo,4)
bin.malo<-mapply(function(w,y) mapply(function(a,b)
      apply(a,2,function(x) ifelse(x>b,1,0)),w,y,SIMPLIFY =
      F),obs.malo,treshould.malo,SIMPLIFY = F)
corresp.malo.rod<-lapply(rod.lines.malo, function(x)
      project2segment(pont.rod.ppp,x)$mapXY)
bin.malo.corres<-lapply(bin.malo, function(z) mapply(function(w,y)
      apply(w,2,function(x) as.vector(x[y])),z,corresp.malo.rod,SIMPLIFY = F))

```

## A 6 - LINHAS DE COMANDO PARA ANÁLISE DA TAXA DE ERRO TOTAL

```
tabela.confusion.rw<-lapply(value.bin, function(s) lapply(s, function(z)
      aply(z,2,function(x) confusion.matrix(x,data.bin))))
erro<-unlist(lapply(tabela.confusion.rw, function(s)
      lapply(s, function(z) unlist(lapply(z, function(x) (x[1,2]+x[2,1])/sum(x)))))))
tabela.confusion.r<-lapply(value.bin.r, function(s) lapply(s, function(z)
      aply(z,2,function(x) confusion.matrix(x,data.bin))))
erro.r<-unlist(lapply(tabela.confusion.r,function(s)
      lapply(s, function(z) unlist(lapply(z, function(x) (x[1,2]+x[2,1])/sum(x)))))))
conf.malo <-lapply(bin.malo.corres, function(w) lapply(w, function(z)
      aply(z,2,function(x)confusion.matrix(x,data.bin))))
erro.malo<-unlist(lapply(conf.malo, function(x) lapply(x, function(y)
      lapply(y, function(w) (w[1,2]+w[2,1])/sum(w))))))
```

## A 7 - LINHAS DE COMANDO PARA IDENTIFICAÇÃO DAS RAZÕES R/W MAIS EFICIENTES NO MÉTODO WINDOWED *HOTSPOT*

Nas linhas de comando a seguir os perfis binários observados e esperados não comparados. O objeto “tab.final” contém os valores da taxa de erro total de todas as 314 razões r/w testadas para cada uma das 420 unidades amostrais (que são as 21 simulações de cada um dos 20 cenários).

```
tab.final<-expand.grid(1:21,1:nrow(rw),1:5,1:4)
tab.final<-data.frame(noise=factor(tab.final[,4]),
                     density=factor(tab.final[,3]),
                     radius=factor(rr[tab.final[,2],1]),
                     subwindow=factor(rr[tab.final[,2],2]),
                     comb.rw=factor((1:nrow(rr))[tab.final[,2]]),
                     ratio.rw=factor(round(rr[tab.final[,2],2]/rr[tab.final[,2],1],1)))
tab.final<-cbind(tab.final,erro=erro)
tab.final<-cbind(formula=factor('new'),tab.final)
```

Linhas de comando para cálculo de valores médios das taxas de erros total e intervalos de confiança, para cada razão r/w testada, utilizando Bootstrap. O objeto “d.rw” é um data frame contendo os valores médios e os intervalos de confiança da taxa total de erros para cada uma das 314 razões r/w testadas.

```
split.comb.rw<-split(tab.final,tab.final$comb.rw)
results.rw<-lapply(split.comb.rw,function(x)
                  boot(data=x$erro,function(y,i) mean(y[i]),R=1000))
ci.rw<-lapply(results.rw,function(x) boot.ci(x,type=c('bca')))
tab.ci.rw<-t(sapply(ci.rw,function(x) cbind(x$bca[4],x$bca[5])))
mean.erro.rw<-sapply(split.comb.rw,function(x) mean(x$erro))
x.rw<-as.numeric(names(split.comb.rw))
```

```
d.rw <- data.frame(  
  x = x.rw  
  , y = mean.erro.rw  
  , ci.inf = tab.ci.rw[,1]  
  , ci.sup = tab.ci.rw[,2]  
)
```

## A 8 - LINHAS DE COMANDO PARA COMPARAÇÕES ENTRE OS TRÊS MÉTODOS PARA OS 20 CENÁRIOS

As linhas de comando a seguir selecionam somente os valores referentes à razão r/w mais eficientes para cada escala r de análise com o método *Windowed Hotspot*. Essa seleção é necessária para que o objeto “tab.etic.new” apresente dimensões compatíveis com as tabelas “tab.final.r” e “tab.malo”.

```
new.tab<-cbind(rr,d.rw)
split.radius.nt<-split(new.tab,new.tab$`1`)
tab.etic<-as.data.frame(t(mapply(function(x) x[x[,6]==min(x[,6])],split.radius.nt)))
tab.etic.new<-tab.final[which(tab.final$comb.rw%in%unlist(tab.etic$x)),]
tab.final.r<-expand.grid(1:21,1:19,1:5,1:4)
tab.final.r<-data.frame(noise=factor(tab.final.r[,4]),
                        density=factor(tab.final.r[,3]),
                        radius=factor(r[tab.final.r[,2]])
                        )
tab.final.r<-cbind(tab.final.r, erro=erro.r)
tab.final.r<-cbind(formula=factor('old'),tab.final.r)
tab.malo<-cbind(tab.final.r,erro.malo)
tab.malo$formula<-'malo'
tab.malo$erro<-erro.malo
tab.malo<-tab.malo[,1:5]

tab.etic.final<-rbind(tab.etic.new[,c(1:4,8)],tab.final.r,tab.malo)
splitn.d.malo.2<-split(tab.etic.final,list(tab.etic.final$noise,tab.etic.final$density))
splitn.d.malo.20<-lapply(splitn.d.malo.2, function(x) split(x,list(x$formula,x$radius)))
list.resultsn.d.malo20<-lapply(splitn.d.malo.20, function(y) lapply(y, function(x)
                        boot(x$erro,function(y,i) mean(y[i]),R=1000)))
list.tab.cin.d.malo<-lapply(list.resultsn.d.malo20, function(y)
```

```
t(sapply(y, function(x) quantile(x$t,c(0.025,0.975))))  
list.mean.erron.d.malo<-lapply(list.resultsn.d.malo20, function(y) sapply(y, function(x) x$t0))  
list.xn.d.malo<-rep(seq(200,2000,100),each=3)  
list.dn.d.malo <- mapply(function(a,b)  
  data.frame(  
    x=rep(seq(200,2000,100),each=3)  
    ,y = a  
    , ci.inf = b[,1]  
    , ci.sup = b[,2]  
    , formula = rep(c('new','old','malo'),19)  
  ),list.mean.erron.d.malo,list.tab.cin.d.malo,SIMPLIFY = F)
```

## A 9 - LINHAS DE COMANDO PARA COMPARAÇÕES ENTRE OS TRÊS MÉTODOS EM FUNÇÃO DA DENSIDADE

```
tab.efic.malo<-tab.malo[which(tab.malo$radius==200),]
tab.efic.r<-tab.final.r[which(tab.final.r$radius==200),]
tab.efic<-tab.final[which(tab.final$comb.rw==22),]
tab.efic.fim<-rbind(tab.efic[, -c(4:7)], tab.efic.r[, -4], tab.efic.malo[, -4])
formula.dens.m2<-split(tab.efic.fim, list(tab.efic.fim$formula, tab.efic.fim$density))
results.m2<-lapply(formula.dens.m2, function(x)
  boot(data=x$erro, function(y,i) mean(y[i]), R=1000))
tab.ci.m2<-t(sapply(results.m2, function(x) quantile(x$t, c(0.025, 0.975))))
mean.erro.m2<-sapply(results.m2, function(x) x$t0)
x.m2<-paste0('Dens.', rep(1:5, each=3))
d.m2 = data.frame(
  x=x.m2
  ,y = mean.erro.m2
  , ci.inf = tab.ci.m2[,1]
  , ci.sup = tab.ci.m2[,2]
  , formula = factor(rep(c('new', 'old', 'malo'), 5))
)
tab.efic.malo<-tab.malo[which(tab.malo$radius==200),]
tab.efic.r<-tab.final.r[which(tab.final.r$radius==200),]
tab.efic<-tab.final[which(tab.final$comb.rw==22),]
tab.efic.fim<-rbind(tab.efic[, -c(4:7)], tab.efic.r[, -4], tab.efic.malo[, -4])

formula.dens.m2<-split(tab.efic.fim, list(tab.efic.fim$formula, tab.efic.fim$density))
results.m2<-lapply(formula.dens.m2, function(x)
  boot(data=x$erro, function(y,i) mean(y[i]), R=1000))
tab.ci.m2<-t(sapply(results.m2, function(x) quantile(x$t, c(0.025, 0.975))))
mean.erro.m2<-sapply(results.m2, function(x) x$t0)
x.m2<-paste0('Dens.', rep(1:5, each=3))
```

```
d.m2 = data.frame(  
  x=x.m2  
  ,y = mean.erro.m2  
  , ci.inf = tab.ci.m2[,1]  
  , ci.sup = tab.ci.m2[,2]  
  , formula = factor(rep(c('new','old','malo'),5))  
)
```

## A 10 - LINHAS DE COMANDO PARA COMPARAÇÕES ENTRE OS TRÊS MÉTODOS EM FUNÇÃO DO TIPO DE RUÍDO

```
formula.noise.malo2<-split(tab.efic.fim,list(tab.efic.fim$formula,tab.efic.fim$noise))
results.n.m2<-lapply(formula.noise.malo2, function(x)
    boot(data=x$erro,function(y,i) mean(y[i]),R=1000))
tab.ci.n.m2<-t(sapply(results.n.m2, function(x) quantile(x$t,c(0.025,0.975))))
mean.erro.n.m2<-sapply(results.n.m2,function(x) x$t0)
x.n.m2<-paste0('Noise',rep(1:4,each=3))
d.n.m2 = data.frame(
    x=x.n.m2
    ,y = mean.erro.n.m2
    , ci.inf = tab.ci.n.m2[,1]
    , ci.sup = tab.ci.n.m2[,2]
    , formula = factor(rep(c('new','old','malo'),4))
)
```

```
tab.efic.malo<-tab.malo[which(tab.malo$radius==200),]
tab.efic.r<-tab.final.r[which(tab.final.r$radius==200),]
tab.efic<-tab.final[which(tab.final$comb.rw==22),]
tab.efic.fim<-rbind(tab.efic[,c(4:7)],tab.efic.r[,4],tab.efic.malo[,4])
```

```
formula.noise.malo2<-split(tab.efic.fim,list(tab.efic.fim$formula,tab.efic.fim$noise))
results.n.m2<-lapply(formula.noise.malo2, function(x)
    boot(data=x$erro,function(y,i) mean(y[i]),R=1000))
tab.ci.n.m2<-t(sapply(results.n.m2, function(x) quantile(x$t,c(0.025,0.975))))
mean.erro.n.m2<-sapply(results.n.m2,function(x) x$t0)
x.n.m2<-paste0('Noise',rep(1:4,each=3))
d.n.m2 = data.frame(
    x=x.n.m2
    ,y = mean.erro.n.m2
```

```
, ci.inf = tab.ci.n.m2[,1]  
, ci.sup = tab.ci.n.m2[,2]  
, formula = factor(rep(c('new', 'old', 'malo'), 4))  
)
```

## A 11 - LINHAS DE COMANDO PARA ANÁLISES DE *HOTSPOTS*

As linhas de comando a seguir constroem os perfis binários de presença e ausência de *hotspots* para os três métodos utilizados no capítulo 01 desta dissertação. Os objetos criados foram denominados “baseline.bin.old”, “baseline.bin.new” e “bin.malo.corresp.200”.

```
real.data<- read.csv2("~/your_data.csv",header = T,row.names = 1,;"")
real.data<-cbind(x=real.data$x,y=real.data$y)
real.aleat<-rply(1000,coords(rpoisppOnLines(nrow(real.data)/nrow(rodovia),linnet.rod)))

baseline.200<-s.hot(real.data,rs=200,rodovia,pont.rod)
baseline.2800<-s.hot(real.data,rs=2800,rodovia,pont.rod)
baseline.rw<-baseline.200/baseline.2800

baseline.lim.200<-sapply(real.aleat, function(x) s.hot(x,rs=200,rodovia,pont.rod))
baseline.lim.200<-quantile(baseline.lim.200,0.975)
baseline.lim.2800<-sapply(real.aleat, function(x) s.hot(x,rs=2800,rodovia,pont.rod))
baseline.lim.2800<-quantile(baseline.lim.2800,0.975)
baseline.lim.rw<-baseline.lim.200/baseline.lim.2800

baseline.bin.old<-ifelse(baseline.200-baseline.lim.200>0,1,0)
baseline.bin.new<-ifelse(baseline.rw-baseline.lim.rw>0,1,0)

road.2.plot.new<-data.frame(pont.rod3,baseline.bin.new)
road.2.plot.old<-data.frame(pont.rod3,baseline.bin.old)

seg.malo <- ppp(rodovia$x[seq(1,length(rodovia$x),by=200)],
               rodovia$y[seq(1,length(rodovia$y),by=200)],window=window.rod)
edge.malo <- cbind(1:(seg.malo$n-1),2:seg.malo$n)
```

```

linnet.malo <- linnet(seg.malo,edges = edge.malo)
rod.lines.malo <- as.psp(linnet.malo)
n.seg.malo<-project2segment(ppp(whole.data[,1],whole.data[,2],
      window = window.rod),rod.lines.malo)
obs.malo<-table(n.seg.malo$mapXY)
correct.obs.malo<-rep(0,rod.lines.malo$n)
names(correct.obs.malo)<-as.character(1:rod.lines.malo$n)
correct.obs.malo[names(correct.obs.malo)%in%names(obs.malo)]<-obs.malo
aleat.malo<-rply(100,rpoisppOnLines(nrow(whole.data)/nrow(rodovia),linnet.malo))
lim.malo<-sapply(aleat.malo, function(x) table(project2segment(x,rod.lines.malo)$mapXY))
correct.lim.malo<-rply(100,rep(0,rod.lines.malo$n))

for(k in 1:100){
  names(correct.lim.malo[[k]])<-as.character(1:rod.lines.malo$n)
  correct.lim.malo[[k]][names(correct.lim.malo[[k]])%in%names(lim.malo[[k]])]<-lim.malo[[k]]
}

lim.malo<-sapply(correct.lim.malo, function(b) b*1)
treshould.malo<-apply(lim.malo,1,function(x) quantile(x,0.975))
bin.malo<-ifelse(correct.obs.malo>treshould.malo,1,0)
corresp.malo.rod.200<-project2segment(ppp(x=pont.rod[,1],y=pont.rod[,2],
      window = window.rod),rod.lines.malo)$mapXY
bin.malo.corresp.200<-as.vector(bin.malo[corresp.malo.rod.200])

```

## A 12 – LINHAS DE COMANDO PARA COMPARAÇÃO DE RESULTADOS DE ANÁLISE DE *HOTSPOT* APÓS A EXCLUSÃO DE TRECHOS DE RODOVIAS.

```
animals.ppp<-ppp(x=real.data[,1],y=real.data[,2],window = window.rod)
animals.segment<-project2segment(animals.ppp,rod.lines)
cords.anim.seg<-coords(animals.segment$Xproj)
anim.seg<-data.frame(cords.anim.seg,seg= animals.segment$mapXY)
seq.remove.10k.lin<-list(1:115,116:204,205:299,300:408)
anim.menos.10k<-lapply(seq.remove.10k.lin,function(x)
                        anim.seg[which(anim.seg$seg%in%c(x)==F),])
real.data3<-lapply(anim.menos.10k, function(x) x[,1:2])
pont.rod3<-lapply(seq.remove.10k.lin,function(x) pont.rod[-c(x),])
linnet.rod3<-lapply(seq.remove.10k.lin,function(x)
                    thinNetwork(linnet.rod,retainededges = seq(1,408,1)[-sort(x)]))
ESECAE.road3<- lapply(linnet.rod3,function(x) as.psp.linnet(x))
rodovia3<-lapply(ESECAE.road3,function(x) pointsOnLines(x,eps=1,shortok = FALSE))
rodovia3.pts<-lapply(rodovia3,coords)

real.aleat3<-mapply(function(x,y,z)
                    rlpoly(100,coords(rpoisppOnLines(nrow(x)/nrow(y),z))),
                    real.data3,rodovia3.pts,linnet.rod3,SIMPLIFY = F)

baseline.200.3<-mapply(function(x,y,z) s.hot(x,rs=200,y,z),
                    real.data3,rodovia3.pts,pont.rod3,SIMPLIFY = F)
baseline.200.3<-sapply(baseline.200.3,function(x) x*1)
baseline.2800.3<-mapply(function(x,y,z) s.hot(x,rs=2800,y,z),
                    real.data3,rodovia3.pts,pont.rod3,SIMPLIFY = F)
baseline.2800.3<-sapply(baseline.2800.3,function(x) x*1)
baseline.rw.3<-mapply(function(a,b) a/b,baseline.200.3,baseline.2800.3)
```

```

baseline.lim.200.3<-mapply(function(a,b,c) sapply(a, function(x) s.hot(x,rs=200,b,c)),
      real.aleat3,rodovia3.pts,pont.rod3,SIMPLIFY = F)
baseline.lim.200.3s<-sapply(baseline.lim.200.3,function(x) quantile(x,0.975))
baseline.lim.2800.3<-mapply(function(a,b,c) sapply(a, function(x) s.hot(x,rs=2800,b,c)),
      real.aleat3,rodovia3.pts,pont.rod3,SIMPLIFY = F)
baseline.lim.2800.3s<-sapply(baseline.lim.2800.3,function(x) quantile(x,0.975))
baseline.lim.rw.3<-mapply(function(a,b) a/b,baseline.lim.200.3s,baseline.lim.2800.3s)

baseline.bin.old.3<-mapply(function(a,b) ifelse(a-b>0,1,0),baseline.200.3,baseline.lim.200.3s)
baseline.bin.new.3<-mapply(function(a,b) ifelse(a-b>0,1,0),baseline.rw.3,baseline.lim.rw.3)

baseline.bin.new<-as.vector(baseline.bin.new)
baseline.bin.old<-as.vector(baseline.bin.old)

baseline.teste.old<-lapply(seq.remove.10k.lin, function(x) baseline.bin.old[-c(sort(x))])
baseline.teste.new<-lapply(seq.remove.10k.lin, function(x) baseline.bin.new[-c(sort(x))])

seq.remove.10k.malo<-list(1:57,58:102,103:150,151:204)
linnet.malo4 <- lapply(seq.remove.10k.malo,function(x)
      thinNetwork(linnet.malo,retainededges = seq(1,204,1)[-c(x)])
rod.lines.malo4 <- lapply(linnet.malo4,as.psp)
obs.malo4<-lapply(seq.remove.10k.malo,function(x) correct.obs.malo[-c(x)])
aleat.malo4<-lapply(1:4,function(x)
      rply(100,rpoisppOnLines(sum(obs.malo4[[x]])/nrow(rodovia3.pts[[x]]),
      linnet.malo4[[x]])))
lim.malo4<-lapply(1:4, function(z) lapply(aleat.malo4[[z]],
      function(x) table(project2segment(x,rod.lines.malo4[[z]])$mapXY)))
correct.lim.malo4<-lapply(1:4, function(x) rply(1000,rep(0,rod.lines.malo4[[x]]$n)))

for(i in 1:4){

```

```

for(k in 1:1000){
  names(correct.lim.malo4[[i]][[k]])<-as.character(1:rod.lines.malo4[[i]]$n)
  correct.lim.malo4[[i]][[k]][names(correct.lim.malo4[[i]][[k]])%in%names(lim.malo4[[i]][[k]])]<-
lim.malo4[[i]][[k]]
}
}

lim.malo4<-lapply(1:4,function(x) sapply(correct.lim.malo4[[x]], function(b) b*1))
treshould.malo4<-lapply(1:4,function(z) apply(lim.malo4[[z]],1,function(x) quantile(x,0.975)))
bin.malo4<-lapply(1:4,function(x) ifelse(obs.malo4[[x]]>treshould.malo4[[x]],1,0))
corresp.malo.rod.200.4<-lapply(1:4, function(a)
  project2segment(ppp(x=pont.rod3[[a]][,1],y=pont.rod3[[a]][,2],
  window = window.rod),rod.lines.malo4[[a]])$mapXY)
bin.malo.corres.200.4<-lapply(1:4, function(x)
  as.vector(bin.malo4[[x]][corresp.malo.rod.200.4[[x]]]))
conf.malo4<-sapply(1:4,function(x) confusion.matrix(bin.malo.corres.200.4[[x]],
  bin.malo.corres.200[-c(seq.remove.10k.lin[[x]])]))
nnn<-lapply(1:4, function(x)
  as.vector(paste0('x',baseline.teste.new[[x]],baseline.bin.new.3[[x]])))
ooo<-lapply(1:4, function(x)
  as.vector(paste0('x',baseline.teste.old[[x]],baseline.bin.old.3[[x]])))
mmm<-lapply(1:4, function(x)
  as.vector(paste0('x',bin.malo.corres.200[-
  c(seq.remove.10k.lin[[x])],bin.malo.corres.200.4[[x]])))

road.2.plot.new.4<-lapply(1:4, function(x)
  data.frame(pr[[x]],bin=factor(nnn[[x]],levels=c('x00','x01','x10','x11'))))
road.2.plot.old.4<-lapply(1:4, function(x)
  data.frame(pr[[x]],bin=factor(ooo[[x]],levels=c('x00','x01','x10','x11'))))
road.2.plot.malo.4<-lapply(1:4, function(x)
  data.frame(pr[[x]],bin=factor(mmm[[x]],levels=c('x00','x01','x10','x11'))))

```



## Capítulo 2

A 13 –

As linhas e comando a seguir foram utilizadas para criar os 21 conjuntos simulados de dados, a partir de um conjunto de dados reais, bom como seus 54 subconjuntos. Para este anexo foram apresentadas as linhas de comando utilizadas para aves, no entanto pode se fazer para qualquer agrupamento de dados desejado. Após a criação dos conjuntos e subconjuntos foram analisados os *hotspots* para cada um deles.

```
real.data<- read.csv2("~/your.data.csv",header = T,row.names = 1,";")
id_times<-lapply(2:10, function(x) cut(real.data$id_mes,x,F))
lista.real.data<-lapply(id_times, function(x) cbind(real.data,id_times=x))
lista.real.data.esecae<-lapply(lista.real.data, function(x) x[x$UC=='ESECAE',])
sample.boot<-rply(21,sample(1:nrow(lista.real.data.esecae[[1]]),
                          nrow(lista.real.data.esecae[[1]]),T))
lista.real.data.boot<-lapply(lista.real.data.esecae, function(y)
                             lapply( sample.boot, function(x) y[x,]))
lista.real.data.boot<-lapply(lista.real.data.boot,function(x)
                             lapply(x, function(y) y[y$Especie!="Nada_observado",]))
lista.real.data.esecae<-lapply(lista.real.data.esecae, function(x)
                             x[x$Especie!='Nada_observado',])
lista.total.aves.sim<-lapply(lista.real.data.boot[[1]], function(w) w[w$Classe=='Ave',])
lista.total.aves.sim<-lapply(lista.total.aves.sim, function(e) data.frame(x=e$x,y=e$y))
lista.total.aves.sim.200<-sapply(lista.total.aves.sim, function(x)
                                s.hot(x,rs=200,rodovia,pont.rod))
lista.total.aves.sim.2800<-sapply(lista.total.aves.sim, function(x)
                                s.hot(x,rs=2800,rodovia,pont.rod))
lista.total.aves.sim.rw<- lista.total.aves.sim.200/lista.total.aves.sim.2800

lista.total.aves.sim.aleat<-lapply(lista.total.aves.sim, function(x)
```

```

        rply(100,coords(rpoisppOnLines(nrow(x)/nrow(rodovia),
        linnet.rod))))
lista.total.aves.sim.lim.200.a<-lapply(lista.total.aves.sim.aleat, function(y)
        sapply(y, function(x) s.hot(x,rs=200,rodovia,pont.rod)))
lista.total.aves.sim.lim.200<-sapply(lista.total.aves.sim.lim.200.a,function(y)
        apply(y,1, function(z) quantile(z,0.975)))
lista.total.aves.sim.lim.2800.a<-lapply(lista.total.aves.sim.aleat, function(y)
        sapply(y, function(x) s.hot(x,rs=2800,rodovia,pont.rod)))
lista.total.aves.sim.lim.2800<-sapply(lista.total.aves.sim.lim.2800.a,function(y)
        apply(y,1, function(z) quantile(z,0.975)))
lista.total.aves.sim.lim.rw<-lista.total.aves.sim.lim.200/lista.total.aves.sim.lim.2800

lista.total.aves.sim.bin<- ifelse((lista.total.aves.sim.rw-lista.total.aves.sim.lim.rw)>0,1,0)
lista.total.aves.sim.bin<-apply(lista.total.aves.sim.bin,2,function(x) replace(x,is.na(x),0))

- As linhas de comando a seguir criam os subconjuntos para cada um dos conjuntos
simulados e faz as análises de hotspots para cada um.

lista.dados.aves.sim<-lapply(lista.real.data.boot, function(z)
        lapply(z, function(w) w[w$Classe=='Ave',]))
lista.dados.aves.sim<-lapply(lista.dados.aves.sim, function(z)
        lapply(z, function(w) data.frame(t=w$id_times,x=w$x,y=w$y)))
split.dados.aves.sim<-lapply(lista.dados.aves.sim, function(z) lapply(z, function(x) split(x,x$t)))
lista.times.aves.sim<-lapply(split.dados.aves.sim, function(e) lapply(e, function(w)
        lapply(w, function(z) cbind(x=z$x,y=z$y))))
lista.times.aves.sim.200<-lapply(lista.times.aves.sim, function(s) lapply(s,
        function(z) lapply(z, function(x)
        s.hot(x,rs=200,rodovia,pont.rod))))
lista.times.aves.sim.2800<-lapply(lista.times.aves.sim, function(s) lapply(s,
        function(z) lapply(z, function(x) s.hot(x,rs=2800,rodovia,pont.rod))))
lista.times.aves.sim.rw<-mapply(function(x,y) mapply(function(z,w)

```

```

mapply(function(a,b) a/b,z,w), x,y,SIMPLIFY = F),
lista.times.aves.sim.200,lista.times.aves.sim.2800,SIMPLIFY = F)
lista.times.aves.sim.bin<-mapply(function(a,b) lapply(a, function(x) ifelse(x-b>0,1,0)),
lista.times.aves.sim.rw,lista.times.aves.obs.lim.rw,SIMPLIFY = F)
lista.times.aves.sim.bin<-lapply(lista.times.aves.sim.bin, function(w)
lapply(w, function(y) apply(y,2,function(x) replace(x,is.na(x),0))))

```

## A 14 – ANÁLISE DA PRECISÃO DE *HOTSPOTS* PELO TAMANHO DA AMOSTRA

As linhas de comando a seguir calculam a Precisão de todos os subconjuntos com cada um dos conjuntos simulados que os originaram.

```

a.aves<-lapply(lista.times.aves.sim.bin, function(y)
mapply(function(a,b) alply(a,2,function(x)
confusion.matrix(x,lista.total.aves.sim.bin[,b])),y,1:21,SIMPLIFY=F))
a1.aves<-lapply(a.aves, function(x) unlist(lapply(x, function(y) lapply(y, function(w) w[2,2]))))
a2.aves<-lapply(a.aves, function(x) unlist(lapply(x, function(y) lapply(y, function(w)
sum(w[,2]))))))
todos.aves.boot.b<-a3.aves<-mapply(function(x,y) x/y,a1.aves,a2.aves)

b.aves<-lapply(lista.times.aves.sim, function(w)
unlist(lapply(w, function(z) sapply(z, function(x) nrow(x)))))
c.aves<-data.frame(p=unlist(a3.aves),n=unlist(b.aves))
cor.test(c.aves$n,c.aves$p)

```

- Linhas de comando para prever a precisão de qualquer tamanho de amostra com base na regressão quantílica realizada

```

library(quantreg)
rq.aves<-rq(p~log(n),data=c.aves,tau=c(0.025,0.5,0.975))

```

```
sum.aves<-summary(rq.aves)
```

```
n.df <- data.frame(n=seq(1,2500,50))
```

```
f.aves <- predict(rq.aves, newdata=n.df, type = 'percentile', stepfun = FALSE)
```

- Linha de comando para reverter a correção logarítmica e assim ilustrar graficamente a curva de precisão

```
pr.aves <- as.data.frame(exp(1)^cbind(x=log(n.df), f.aves))
```

A 15

- Para o segundo experimento do capítulo 02 utilizou-se as linhas de comando presente em A 13 e A 14. Entretanto, para gerar os conjuntos e subconjuntos, ao invés de se agrupar por Classe (no exemplo utilizou-se “Aves”) agrupa-se por tempo, ou seja repete-se a análise para cada conjunto de dados acumulados desde o início das coletas até cada data pré-determinada. Para cada agrupamento, que representa um período temporal, foi feita a predição de precisão conforme as linhas de comando presente em A 14, obtendo-se para cada grupo, um valor de precisão, podendo-se assim realizar a ilustração gráfica conforme desejado.

## ANEXO II – LISTA DAS ESPÉCIES

Classe	Ordem	Família	Espécie	Total
Amphibia	Anura	Bufonidae	Rhinella schneideri	9
			Rhinella sp	93
		Hylidae	Hypsiboas albopunctatus	1
			Scinax sp	1
		Leptodactylidae	Leptodactylus labyrinthicus	5
			Leptodactylus latrans	5
			Leptodactylus ochraceus	4
			Leptodactylus sp	1
	não identificado	19		
	Microhylidae	não identificado	1	
não identificado	não identificado	7		
Gymnophiona	Caecilidae	Siphonops paulensis	4	
Aves	Accipitriformes	Accipitridae	Gampsonyx swainsonii	1
			Geranoaetus albicaudatus	1
			Heterospizias meridionalis	1
			Rupornis magnirostris	3
	Anseriforme	Anatidae	Anas platyrhynchos domesticus	3
	Apodiformes	Apodidae	Streptoprocne zonaris	1
			Tachornis squamata	1
		Trochilidae	Amazilia fimbriata	9
			Colibri serrirostris	8
			Eupetomena macroura	4
			não identificado	9
	Polytmus theresiae	1		
	Caprimulgiformes	Caprimulgidae	Antrostomus rufus	3
			Hydropsalis climacocerca	6
			Hydropsalis torquata	1
			Nannochordeiles pusillus	4
			não identificado	15
	Nyctidromus albicollis	7		
	Cariamiformes	Cariamidae	Cariama cristata	4
	Cathartiformes	Cathartidae	Coragyps atratus	1
	Charadriiformes	Charadriidae	Vanellus chilensis	1
	Columbiformes	Columbidae	Columba livia	31
Columbina picui			1	
Columbina sp			2	
Columbina talpacoti			9	
não identificado			3	
Patagioenas sp			6	

	<b>Coraciiformes</b>	Alcedinidae	Chloroceryle amazona	2
	<b>Cuculiformes</b>	Cuculidae	Crotophaga ani	25
			Guira guira	21
	<b>Falconiformes</b>	Falconidae	Caracara plancus	5
			Falco sparverius cearae	3
			Fannia femoralis	1
		não identificado	não identificado	3
	<b>Galbuliformes</b>	Bucconidae	Nystalus chacuru	4
	<b>Galliforme</b>	Phasianidae	Gallus gallus domesticus	14
	<b>não identificado</b>	não identificado	não identificado	59
	<b>Passeriformes</b>	Furnariidae	Furnarius rufus	3
			Phacellodomus ruber	2
			Phacellodomus rufifrons	4
		Hirundinidae	Alopochelidon fucata	2
		Icteridae	Gnorimopsar chopi	2
		Melanopareiidae	Melanopareia torquata	1
		Mimidae	Mimus saturninus	5
		não identificado	não identificado	242
		Passeridae	Passer domesticus	2
		Thamnophilidae	Thamnophilus doliatus	2
		Thraupidae	Ammodramus humeralis	9
			Coryphospingus cucullatus	1
			Coryphospingus pileatus	10
			Emberizoides herbicola	8
			não identificado	4
			Nemosia pileata	1
			Neothraupis fasciata	1
Saltatricula atricollis			1	
Sicalis flaveola			2	
Sporophila caerulescens			8	
Sporophila nigricollis			8	
Tangara sayaca			4	
Volatinia jacarina			428	
Zonotrichia capensis			2	
Troglodytidae		Troglodytes musculus	4	
Turdidae		Turdus amaurochalinus	7	
		Turdus leucomelas	1	
	Turdus rufiventris	2		
Tyrannidae	Elaenia chiriquensis	13		
	Machetornis rixosa	13		
	não identificado	1		
	Pitangus sulphuratus	2		
	Tyrannus melancholicus	5		
	Tyrannus savana	40		

			Xolmis cinereus	1
		Vireonidae	Cyclarhis gujanensis	2
	<b>Piciforme</b>	Picidae	Colaptes campestris	5
	<b>Psittaciformes</b>	Psittacidae	Amazona aestiva	1
			Brotogeris chiriri	4
			Eupsittula aurea	2
			não identificado	1
	<b>Strigiforme</b>	Strigidae	Aegolius harrisii	4
			Asio clamator	18
			Athene cunicularia	39
			Glaucidium brasilianum	1
			Megascops choliba	8
			não identificado	2
		Tytonidae	Tyto furcata	33
	<b>Tinamiforme</b>	Tinamidae	Crypturellus parvirostris	15
			não identificado	2
			Nothura maculosa	8
Rhynchotus rufescens			12	
<b>Mammalia</b>	<b>Artiodactyla</b>	Cervidae	Mazama gouazoubira	1
	<b>Carnivora</b>	Canidae	Canis lupus familiaris	163
			Cerdocyon thous	43
			Chrysocyon brachyurus	1
			Lycalopex vetulus	6
		Felidae	Felis catus	106
			Leopardus tigrinus	1
			não identificado	1
			Puma concolor	2
		Mephitidae	Conepatus semistriatus	21
		Mustelidae	Galictis cuja	10
	Procyonidae	Procyon cancrivorus	8	
	<b>Chiroptera</b>	Molossidae	Molossops sp	1
			não identificado	4
		não identificado	não identificado	43
		Phyllostomidae	Artibeus sp	2
			Glossophaga sp	11
			não identificado	9
			Platyrrhinus sp	2
	Sturnira sp		1	
<b>Cingulata</b>	Dasypodidae	Dasypus novemcinctus	3	
		Dasypus septemcinctus	5	
		Dasypus sp.	1	
		Euphractus sexcinctus	4	
		não identificado	1	
		não identificado	não identificado	1

	<b>Didelphimorphia</b>	Didelphidae	Didelphis albiventris	38
	<b>Lagomorpha</b>	Leporidae	Sylvilagus brasiliensis	1
	<b>não identificado</b>	não identificado	não identificado	8
	<b>Perissodactyla</b>	Equidae	Equus ferus caballus	2
	<b>Pilosa</b>	Myrmecophagidae	Myrmecophaga tridactyla	1
	<b>Primates</b>	Atelidae	Alouatta caraya	1
		Cebidae	Callithrix penicillata	14
			Sapajus libidinosus	1
	<b>Rodentia</b>	Cricetidae	Calomys tener	8
			não identificado	20
			Necromys lasiurus	6
		Erethizontidae	Coendou prehensilis	1
	Hydrochaeridae	Hydrochoerus hydrochaeris	1	
	<b>Reptilia</b>	<b>Squamata</b>	Amphisbaenidae	Amphisbaena alba
Boidae			Boa constrictor	39
			Epicrates cenchria	10
Colubridae			Chironius flavolineatus	3
			Clelia sp.	1
			não identificado	1
			Spilotes pullatus	2
			Tantilla melanocephala	1
Dipsadidae			Apostolepis albicollaris	1
			Boiruna maculata	4
			Erythrolamprus aesculapii	6
			não identificado	5
			Oxyrhopus guibei	20
			Oxyrhopus sp	29
			Oxyrhopus trigeminus	1
			Phalotris nasutus	1
			Philodryas agassizii	1
			Philodryas nattereri	1
			Philodryas olfersii	3
			Philodryas patagoniensis	5
			Philodryas sp	22
Pseudoboa nigra			8	
Sibynomorphus mikanii			22	
não identificado			não identificado	19
Polychrotidae			Polychrus acutirostris	8
Teiidae			Ameiva ameiva	8
			Cnemidophorus sp.	1
Tropiduridae			Tropidurus sp.	5
Viperidae	Bothrops moojeni	1		
	Bothrops sp.	3		
	Crotalus durissus	45		

		Xenodon merremii	2	
		Xenodon sp	1	
	<b>Testudines</b>	Chelidae	Phrynops geoffroanus	11
		Testudinidae	não identificado	1