GfÖ Ecological Society of Germany, Austria and Switzerland

Basic and Applied Ecology 34 (2019) 25–35

Basic and Applied Ecology

www.elsevier.com/locate/baae

On the identification of mortality hotspots in linear infrastructures

Luís Borda-de-Água[a,b,*], Fernando Ascensão[a,b], Manuel Sapage[c], Rafael Barrientos[a,b], Henrique M. Pereira[a,b,d]

[a]*Theoretical Ecology and Biodiversity Group, and Infraestruturas de Portugal Biodiversity Chair, CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos Genéticos, Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal*

[b]*CEABN/InBio, Centro de Ecologia Aplicada "Professor Baeta Neves", Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal*

[c]*cE3c — Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal*

[d]*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*

Received 8 June 2018; accepted 10 November 2018
Available online 16 November 2018

Abstract

One of the main tasks when dealing with the impacts of infrastructures on wildlife is to identify hotspots of high mortality so one can devise and implement mitigation measures. A common strategy to identify hotspots is to divide an infrastructure into several segments and determine when the number of collisions in a segment is above a given threshold, reflecting a desired significance level that is obtained assuming a probability distribution for the number of collisions, which is often the Poisson distribution. The problem with this approach, when applied to each segment individually, is that the probability of identifying false hotspots (Type I error) is potentially high. The way to solve this problem is to recognize that it requires multiple testing corrections or a Bayesian approach. Here, we apply three different methods that implement the required corrections to the identification of hotspots: (*i*) the familywise error rate correction, (*ii*) the false discovery rate, and (*iii*) a Bayesian hierarchical procedure. We illustrate the application of these methods with data on two bird species collected on a road in Brazil. The proposed methods provide practitioners with procedures that are reliable and simple to use in real situations and, in addition, can reflect a practitioner's concerns towards identifying false positive or missing true hotspots. Although one may argue that an overly cautionary approach (reducing the probability of type I error) may be beneficial from a biological conservation perspective, it may lead to a waste of resources and, probably worse, it may raise doubts about the methodology adopted and the credibility of those suggesting it.

Keywords: Bayesian hierarchical model; False discovery rate correction; Familywise error rate correction; Hotspot; Spatial auto-correlation

*Corresponding author at: CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos Genéticos, Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.

E-mail address: lbagua@gmail.com (L. Borda-de-Água).

https://doi.org/10.1016/j.baae.2018.11.001

Introduction

Linear infrastructures, such as power lines, roads or railways, are among the most ubiquitous man-made features worldwide, and are known to have negative impacts on natural habitats and ecosystems (van der Ree, Smith, & Grilo, 2015; Borda-de-Água, Barrientos, Beja & Pereira 2017).

A main concern when evaluating the impact of linear infrastructures on wild animal populations is the identification of sections with high mortality (Quinn, Alexander, Heck, & Chernoff, 2011; Gunson & Teixeira, 2015). Sections with a large number of animal collisions ("mortality hotspots") are generally prioritized when mitigation measures are to be taken, because this is where such measures are more likely to be cost-effective in reducing the excessive animal mortality or in increasing human safety (Gunson & Teixeira, 2015; Barrientos, Alonso, Ponce, & Palacin, 2011). Mitigation measures are often expensive (Barrientos et al., 2011), therefore, a judicious identification of the sections that represent a more serious threat to wildlife populations may lead to significant economic savings while still achieving the goal of reducing the impacts of the infrastructure. On the other hand, mitigating false hotspots results in a waste of resources and may undermine the credibility of conservation efforts. Our objective here is to develop procedures that compromise between minimizing the erroneous identification of false hotspots and the correct identification of true cases of high mortality.

There is a vast body of literature using different methods to identify hotspots, such as, generalized linear models (Gomes, Grilo, Silva, & Mira, 2009), ecological niche factor analysis (Hirzel, Hausser, Chessel, & Perrin, 2002), kernel density estimation (Gitman & Levine, 1970; Bíl, Andrašik, & Janoška, 2013; Bíl, Andrašik, Svoboda, & Sedoník, 2016), nearest neighbour hierarchical clustering (Levine, 2004), distance-based approaches (Kolowski & Nielsen, 2008), and methods based on modelling the number of collisions in a segment assuming a Poisson distribution (Malo, Suárez, & Díez, 2004; Grilo, Ascensão, Santos-Reis, & Bissonette, 2011; Planillo, Kramer-Schadt, & Malo, 2015; Santos et al., 2015). The latter methods based on the Poisson distribution (or any other), if applied without additional information, are likely to lead to the identification of hotspots that are mere statistical artefacts. This happens because researchers do not correct for multiple testing when performing simultaneous tests of hypotheses (e.g. Gelman, Hill, & Yajima, 2012).

We first illustrate the consequences of not correcting for multiple testing and then describe different approaches to deal with the problem. We chose two frequentist methods, the familywise error rate correction (FWER) and the false discovery rate (FDR), because they are easy to implement and their interpretation is straightforward, and chose a Bayesian method because it forces the explicit statement of the assumptions and it has the potential to be extended to a wide range of situations. Similar versions of the Bayesian models we develop here have been applied extensively in traffic accident research and the knowledge gained there can easily be extended to our purposes (e.g., Heydecker & Wu, 2001; Ahmed, Huang, Abdel-Aty, & Guevara, 2011).

For simplicity, throughout the text we assume that the detectability of carcasses is perfect, as other works have addressed the effects of detectability and persistence on mortality estimates but not on hotspot identification bias (e.g. Ponce, Alonso, Argandona, García Fernandez, & Carrasco, 2010; Santos, Carvalho, & Mira, 2011). Therefore, we concentrate solely on the statistical aspects of identifying mortality hotspots given a certain distribution of collisions. Moreover, we assume that the data has been collected in such a way that there is only information on the number of collisions per segment, that is, the exact location of the collisions is unknown, as often occurs in road surveys (Gunson, Clevenger, Ford, Bissonette, & Hardy, 2009).

Finally, this paper deals solely with the statistical aspects of identifying hotspots, but these methods do not substitute studies on the road mortality effect on population viability, nor the discussion on the societal choices underlying the need to identify hotspots. These choices are implicitly or explicitly incorporated in the thresholds adopted in each method. However, we highlight that the Bayesian methods we present force the researcher to explicitly state the perceived costs associated with missing hotspots or identifying false ones, thus they help state and confront different strategies to address road mortality.

This paper does not follow the typical structure of papers in ecology. First, we illustrate the problems that can arise when performing multiple tests. Then we present three different procedures that address the problems arising from multiple tests: the first two use a frequentist approach and the third develops a Bayesian approach. We exemplify the application of these methods using data on road mortality of two bird species, the burrowing owl (*Athene cunicularia*) and the blue–black grassquit (*Volatinia jacarina*). We finish by discussing the merits of the three approaches.

Multiple hypothesis tests and the need for corrections

When assessing which segments of a linear infrastructure are hotspots, it is often assumed that the number of collisions in each segment follows a Poisson distribution and that this distribution describes equally well the collisions in all segments (e.g. Malo et al., 2004; Grilo et al., 2011; Planillo et al., 2015; Santos et al., 2015; Costa, Ascensão, & Bager, 2015; Garriga, Franch, Santos, Montori, & Llorente, 2017). Accordingly, the probability of n occurrences pertaining to a segment of size l, in a linear infrastructure of length L, is described by a Poisson distribution

$$P(n) = \frac{e^{-\lambda l}(\lambda l)^n}{n!} \quad (1)$$

where λ is the rate of collisions per unit of length estimated during a certain period of time.

If collisions are uniformly randomly distributed along the infrastructures there should be no hotspots, however, the occasional appearance of regions with a high concentration of collisions is to be expected due to the inherent stochasticity of the process. Naturally, one wants to safeguard against the identification of such spurious occurrences as hotspots. However, if we use Eq. (1) to establish a threshold to test whether a segment is a hotspot, and then applying such a test to all segments simultaneously, we are likely to obtain a large number of false hotspots. For instance, assume that $\lambda = 0.635$ collisions km^{-1} and define that a segment with 3 or more collisions is a hotspot. Using Eq. (1) this leads to the probability $P(n \geq 3) \approx 3\%$ (Malo et al., 2004). In the typical terminology of hypothesis testing, the probability of a Type I error is 3%. This may look a small probability, but if the infrastructure has a large number of segments and we apply the test to all the segments, the probability of incorrectly identifying segments as hotspots increases considerably. For example, if a road consists of 100 segments we expect that on average 3 segments (i.e., 3%) will be incorrectly identified; we further elaborate on this example in Appendix A in Supplementary material.

Correcting for multiple testing

The familywise error rate correction (FWER)

Statisticians have for long considered several corrections when conducting multiple tests. One of them is the Dunn–Šidák correction (Sokal & Rohlf, 1995). In our case, it is important to distinguish the probability of identifying incorrectly one or more segments in the entire infrastructure, which we want to be smaller than or equal to P_L, and the probability of incorrectly identifying one segment, which we want to be smaller than or equal to P_S. With these definitions Dunn–Šidák correction is given by

$$P_S \leq 1 - (1 - P_L)^{1/M} \tag{2}$$

where M is the total number of segments. In most cases Eq. (2) can be approximated by the Bonferroni correction:

$$P_S \leq \frac{P_L}{M} \tag{3}$$

(see Appendix A in Supplementary material). For example, if $P_L = 0.05$ and $M = 100$, then we should use $P_S = 0.0005$ at the segment level.

The false discovery rate (FDR)

The FWER correction reduces the probability of identifying false positive hotspots (Type I error) at the expense of increasing the probability of missing true hotspots (false negatives, Type II error). If we are also concerned with Type II errors then, under the frequentist approach, we can consider the FDR method.

The FDR procedure was introduced by Benjamini and Hochberg (1995). These authors suggested a method that consists of sorting all p-values from M hypotheses tested in ascending order and then determine whether a k-ordered p-values (denoted p_{value}) obeys to

$$P_{value} > \frac{kP_S}{M} \tag{4}$$

where P_S is a threshold imposed by the researcher. If a p_{value} obeys to (4) then it is declared non-significant or, in the present context, it is not a hotspot. (See Appendix A in Supplementary material for more details.)

For instance, consider again an infrastructure with $M = 100$ segments, and a threshold $P_S = 0.027$. Calculate the p_{value} for each segment, and order these values from the smallest to the largest; call this ordered vector P. Suppose that the p_{value} of the segment in position $k = 15$ of the vector P is 0.136. Inequality (4) leads to $kP_S/M = 15 \times 0.027/100 = 0.0039$, hence, because $p_{\text{value}} = 0.136 > 0.0039$, we conclude that this segment is not a hotspot. Assume now a segment with $p_{\text{value}} = 5.3 \times 10^{-5}$ corresponding to position $k = 1$ of vector P. Then $kP_S/M = 1 \times 0.027/100 = 2.7 \times 10^{-4}$. Because $p_{\text{value}} = 5.3 \times 10^{-5} < 2.7 \times 10^{-4}$, we identify the segment as a hotspot. In the Appendix B in Supplementary material we provide an R (R Core Team, 2016) function that implements this procedure.

Bayesian hierarchical models (BHM)

The analysis of a dataset under the Bayesian paradigm consists of combining the information contained in the data, that is, the likelihood given a parametric model, $P(D|M)$, with prior information, $P(M)$, using Bayes Rule (e.g. Gelman et al., 2014). The resulting distribution is the posterior, $P(M|D)$,

$$P(M|D) \propto P(D|M)\, P(M),$$

from where all conclusions are derived. In our case, the posterior of interest is the distribution of the parameter λ of the Poisson distribution (Eq. (1)) given the observed number of collisions, n, in each segment. As we will see, under the Bayesian framework multiple testing corrections are taken into account implicitly.

We start by assuming that the number of collisions, n_i, in a segment i is a random variable with a Poisson distribution with parameter λ_i,

$$N_i|\lambda_i \sim \text{Poisson}(\lambda_i).$$

This is an important departure from the previous methods, because now each segment has its own λ_i, while before a single parameter λ applied to all segments. Next we assume that the set of parameters λ_i are realizations of another distribution, called a prior distribution. For the models developed here the parameters of the prior distribution will have probability distributions, the so-called hyperpriors, and these will also have associated distributions, and so on, i.e., there is a hierarchy of distributions.

Assuming no spatial autocorrelation, we develop two hierarchical models: the Poisson-lognormal model and a Poisson-gamma model. This is because when the rate of fatalities λ is small, a Poisson-lognormal is to be preferred to the Poisson-gamma but for larger λ the latter offers some analytical advantages and performs better (Lord & Miranda-Moreno, 2008).

For the Poisson-lognormal model the prior for λ_i is

$$\lambda_i = \exp(\varepsilon_i),$$

and the hyperprior for the parameter ε_i is a normal distribution, such as

$$\varepsilon_i \mid \phi \sim \text{Normal}(0, \phi)$$

where ϕ is the precision (the inverse of the variance). For the hyperparameter, ϕ, we assume

$$\phi \mid c_1, c_2 \sim \text{Gamma}(c_1, c_2),$$

where c_1, and c_2 are constants.

The development of the Poisson-gamma model is similar, except that for λ_i we assume a gamma distribution instead of a lognormal. Thus,

$$\lambda_i \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta),$$

where α and β are the shape and scale parameters of the gamma distribution. The last step is to specify the distributions of the hyperparameters α and β,

$$\alpha \mid a_1, a_2 \sim \text{Gamma}(a_1, a_2),$$

$$\beta \mid b_1, b_2 \sim \text{Gamma}(b_1, b_2)$$

where a_1, a_2, b_1, and b_2 are constants. Fig. 1A illustrates the structure of the hierarchical model for the Poisson-gamma model and Fig. 1B for the Poisson-lognormal model; Appendix C in Supplementary material provides codes for the implementation of the above two models using the software package JAGS (Plummer, 2003).

Until now we have disregarded spatial autocorrelation among the segments in order to provide very simple methods that give a first correction to the more serious problems of not considering multiple testing. Introducing spatial autocorrelation means we take into account the positions of the segments relative to each other when determining their relative influence. Because of the mathematical problems of dealing with a multivariate gamma distribution (Zhang, Matthaiou, Karagiannidis, & Dai, 2016), we consider only the Poisson-lognormal model.

We introduce spatial dependence in the Poisson-lognormal model by changing expression $\varepsilon_i \mid \phi \sim \text{Normal}(0, \phi)$ to (Banerjee, Carlin, & Gelfand, 2014)

$$\varepsilon_i | \delta, \phi, \varepsilon_j, i \neq j \sim \text{Normal}(\delta \sum_{j=1}^{M} b_{ij}\varepsilon_j, \phi)$$

where $0 \leq \delta \leq 1$ controls the overall spatial dependence and $0 \leq b_{ij} \leq 1$ controls the dependence between segments i and j. The term $\delta \sum_{j=1}^{M} b_{ij}\varepsilon_j$ adds to the formulation the influence of the neighbourhood of segment i on the estimation of its ε_i and, hence, of λ_i. For instance, if $b_{ij} = 0$ site j is not in the immediate neighbourhood of site i and it does not have a direct influence, if $b_{ij} \geq 0$ then site j influences site i. The sum of all influences for one segment must sum up to one, so each b_{ij} is equal to the inverse of the number of immediate neighbours of i. If $\delta = 0$, ε_i does not take into account spatial autocorrelation.

It is more convenient to use the above expression in a matricial form (see Jin, Bradley & Sudipto (2005) for details):

$$\varepsilon \mid \phi, \delta, \ \mathbf{W} \sim \text{Multivariate Normal}(\mathbf{0}, \phi\,(\mathbf{D} - \delta\mathbf{W})),$$

where $\phi(\mathbf{D} - \delta\mathbf{W})$ is the precision matrix (the inverse of the variance matrix). Vector $\boldsymbol{\varepsilon}$ is of length M and contains $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M$, and $\mathbf{0}$ is a vector of zeroes, also of length M. $\mathbf{W}$ is an adjacency matrix of size $M \times M$ where $w_{ij} = 1$ if i is adjacent to j and $w_{ij} = 0$ otherwise. $\mathbf{D}$ is a vector of size M where each element i is the sum of the elements of row i in matrix $\mathbf{W}$. This model is known as the conditionally autoregressive (CAR) model (Jin et al., 2005).

Finally, to complete the models we need to specify the distributions of the hyperparameters ϕ and δ. We assume

$$\phi \mid c_1, c_2 \sim \text{Gamma}(c_1, c_2),$$

$$\delta \sim \text{Uniform}(0,\ 1),$$

where c_1, c_2 are constants that define the hyperpriors, and may be set by the researcher or be based on previous related studies; Fig. 1C illustrates this model graphically and Appendix D in Supplementary material provides codes in RStan (Stan Development Team, 2016a) for its implementation. Please refer to Box 1 for a summary of the steps to build a hierarchical Bayesian model.

The important point is that the hierarchical models implicitly take into account corrections when performing multiple tests since the posterior of λ_i of segment i is calculated based on information from all segments, and not only from segment i. When we add spatial autocorrelation, we explicitly take into account the position of the segment i relatively to the others. The overall result is that the value of a given λ_i is moved towards the mean of its distribution and reduces the size of the confidence intervals, thus reducing the number of significant results (Gelman et al., 2012).

We test whether a segment is a hotspot using a decision-theoretic approach as in Miranda-Moreno, Labbe and Fu (2007). These authors described three procedures, but we use only the so-called Bayesian test with weights, for three main reasons: first, it is the most straightforward to apply; second, it allows to state explicitly the researchers' concerns regarding incurring in Type I or II errors; and, third, for a reasonable choice of parameters it gives results that are intermediate to those of the other two methods.

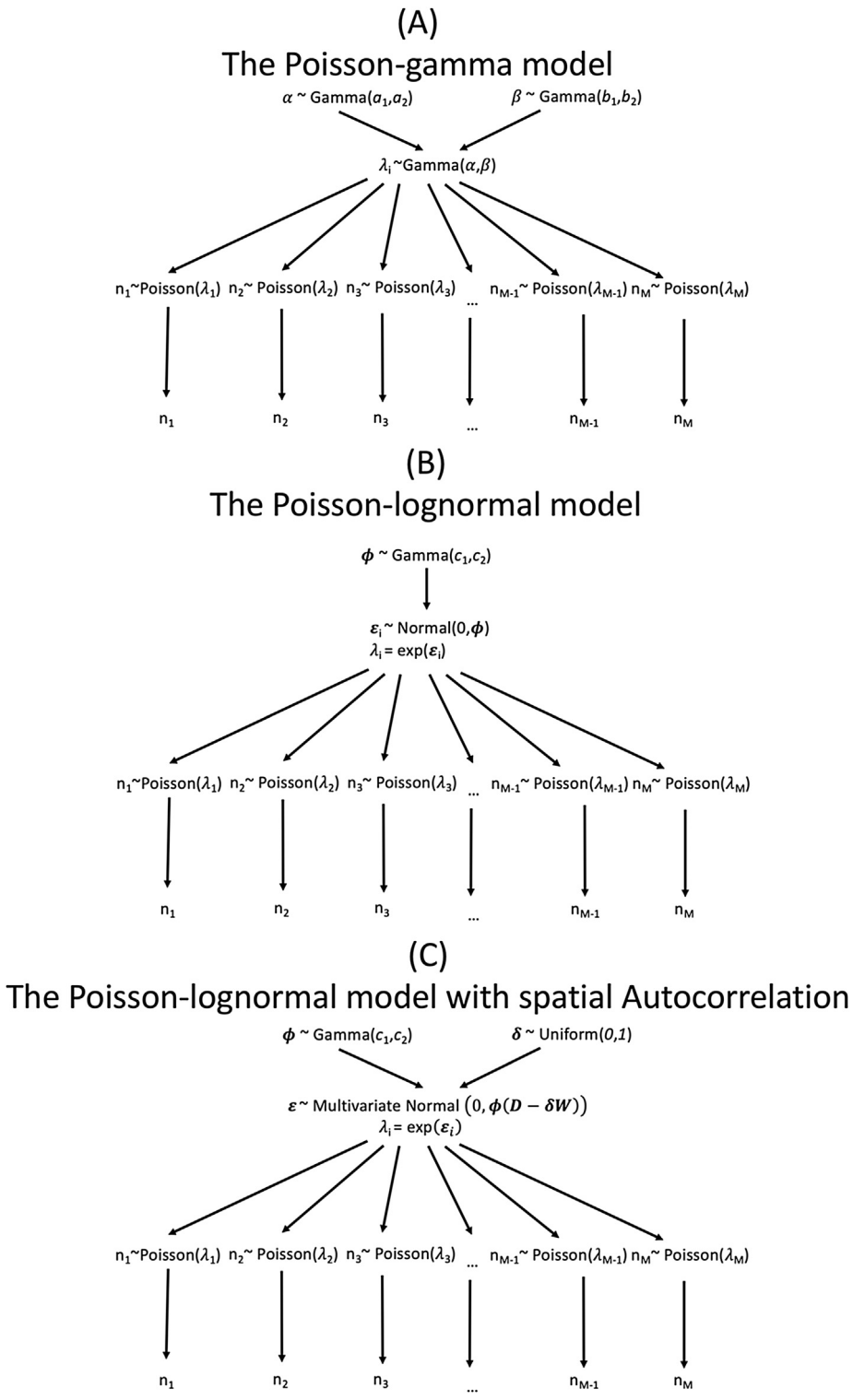

Fig. 1. Graphical representation of the Bayesian hierarchical models. Plots (A) and (B) show the Poisson-lognormal hierarchical and Poisson-gamma models, respectively, without spatial autocorrelation. Plot (C) shows the Poisson-lognormal hierarchical model with spatial autocorrelation. In the latter model, the term $(\boldsymbol{D} - \delta \boldsymbol{W})$ implements the spatial autocorrelation.

The purpose of the Bayesian test with weights approach is to minimize a loss function that takes into account the costs of making a Type I error, C_I, and the costs of making a Type II error, C_{II}. The costs C_I and C_{II}, or their relative values, can be obtained by expert judgment or can reflect social perceptions associated with different choices. Therefore, the specification of C_I and C_{II} allows us to state explicitly whether we are more

Box 1: The Bayesian approach step-by-step

We summarize here the several steps to perform the Bayesian analysis. For simplicity we divide them into two sets of procedures.

Building a hierarchical Bayesian model:

1) Identify a sampling distribution to model the number of collisions in each segment (in our case a Poisson distribution);
2) Establish prior distributions for the parameters of the distribution of the number of collisions in each segment (in our case a prior to model the λ_i), and if spatial autocorrelation is present then introduce a formulation for it;
3) Choose the distributions of the (hyper) parameters of the prior distribution.

Performing hypothesis tests:

1) Define a threshold t_S;
2) Assign the costs C_I and C_{II} and calculate the threshold t_C;
3) Based on the posterior of λ_i calculate the probability $P(\lambda_i > t_S \mid n_i)$;
4) Assess whether $P(\lambda_i > t_S \mid n_i) \geq t_C$. If so, the segment is a hotspot, otherwise it is not.

concerned with missing hotspots or incorrectly identifying segments as hotspots.

For our analyses we need to establish two thresholds, t_S, and t_C, the former establishing whether or not a segment is a hotspot, and the latter defining the probability of that segment being a hotspot. Accordingly, for a given t_S and using the terminology H_{0i} for the null hypothesis and H_{1i} for the alternative hypothesis, we have

$$\begin{cases} H_{0i} : \lambda_i \leq t_s & \text{Not a hotspot} \\ H_{1i} : \lambda_i > t_s & \text{Hotspot} \end{cases}$$

Ideally, the choice of t_S should be based on expert judgement or some other criteria that reflect our knowledge of the population dynamics and the risks posed by the infrastructure to its viability. Here, assuming that there is no information on the abundance or the dynamics of the targeted population (as it often happens), we estimate t_S from the mean and standard deviation of the number of collisions n obtained directly from the raw data (Miranda-Moreno et al., 2007):

$$t_S = \text{mean}(n_i) + z_S \,\text{standard deviation}(n_i). \quad (5)$$

With this formulation the decision is now in terms of z_S, which defines how many standard deviations t_S is away from the mean.

Once t_S (or z_S) has been defined we can calculate the probabilities of the null and alternative hypotheses based on the posterior distribution of λ_i. Thus, the probability of a segment being a hotspot is $P(H_{1i} \mid n_i) = P(\lambda_i > t_S \mid n_i)$, knowing that an incorrect classification of a false positive has cost C_{II}. Equally, the probability of not being a hotspot is $P(H_{0i} \mid n_i) = P(\lambda_i \leq t_S \mid n_i)$, and an incorrect classification of false negative has cost C_I. The cost of a correct identification is zero. The objective is to minimize costs and it can be shown (Berger, 1985) that this is achieved when

$$P(H_{1i}|n_i) = P(\lambda_i > t_S|n_i) \geq \frac{C_I}{C_I + C_{II}} \quad (6)$$

The ratio $C_I/(C_I + C_{II}) = t_C$ defines the threshold value that rejects the null hypothesis and defines whether a segment is a hotspot. We summarize these steps in Box 1.

Table 1. Number of hotspots identified using different methods. "Without corrections" correspond to the results obtained without taking into account multiple testing corrections and assuming that the number of collisions in each segment follows a Poisson distribution and a significance level of 0.05. "FWER" stands for the "familywise error rate correction" method, "FDR" for the "false discovery rate" method, and "S.A." for "spatial autocorrelation".

Method	Burrowing owl (*Athene cunicularia*)	Blue–black grassquit (*Volatinia jacarina*)
Without corrections	10	10
FWER	1	6
FDR	3	9
Bayesian without S.A.		
$C_I = 1$; $C_{II} = 2$	6	6
$C_I = 1$; $C_{II} = 1$	3	2
$C_I = 2$; $C_{II} = 1$	1	2
Bayesian with S.A.		
$C_I = 1$; $C_{II} = 2$	4	6
$C_I = 1$; $C_{II} = 1$	4	4
$C_I = 2$; $C_{II} = 1$	2	1

Two case studies

We illustrate the application of the previous methods with roadkill data on two bird species, the burrowing owl (*A. cunicularia*, with $n = 40$ collisions) and the blue–black grassquit (*V. jacarina*, with $n = 475$ collisions) collected in Brazil along a 50-km road; each segment is 1 km, thus $M = 50$ (see Santos et al. (2016) for details). We use these two data sets because these two species exhibit very different number of casualties, thus allowing us to illustrate the two different Bayesian models developed without spatial autocorrelation.

Without corrections, assuming that the number of collisions in each segment follows a Poisson distribution and a significance level of 0.05, the collision threshold for the burrowing owl is $n_T = 2$ and for the blue–black grassquit is $n_T = 15$ leading, coincidentally, to 10 hotspots for each species (Figs. 2 A and 3 A, and Table 1).

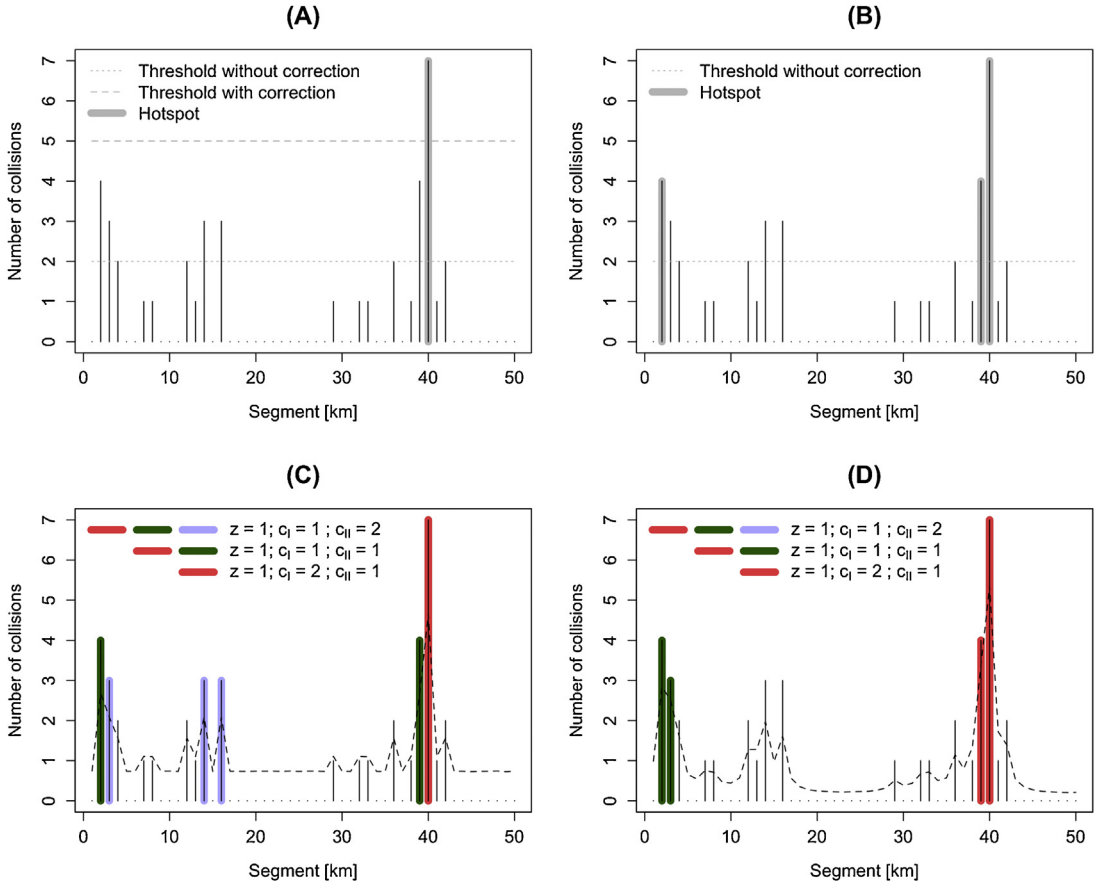

Fig. 2. Results for the burrowing owl (*Athene cunicularia*). The black vertical lines depict the number of collisions along the 50 1-km road segments. Plot (A): the grey dotted line shows the threshold when no corrections are applied (leading to 10 hotspots), and the grey dashed line shows the threshold after application of the familywise error rate correction (leading to 1 hotspot) identified by the vertical grey bar. Plot (B): The application of the false discovery rate correction leads to the identification of 3 hotspots identified by the vertical grey bars. Plot (C): The application of the Bayesian hierarchical Poisson-lognormal model without spatial autocorrelation leads to the hotspots identified by vertical coloured bars: in red are the hotspots identified when $C_I = 2$ and $C_{II} = 1$, in red and green when $C_I = 1$ and $C_{II} = 1$, and in red, green and blue when $C_I = 1$ and $C_{II} = 2$. The broken line corresponds to the average value of the posterior of each λ_i (the mean rate of collisions per unit of length estimated during a certain period of time). Plot (D): The results of the Bayesian hierarchical model with spatial autocorrelation, using the same convention for the hotspots as in plot C. Notice that in both plots (C) and (D) that. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The familywise error rate correction (FWER)

Assuming $P_L = 0.05$, the FWER correction leads to $P_S \approx 0.001$ (Eq. (2) or to its simplified version Eq. (3)). The corresponding number of collision thresholds are $n_T = 5$ for the burrowing owl and $n_T = 20$ for the blue–black grassquit, thus we identify 1 hotspot for the burrowing owl and 6 hotspots for the blue–black grassquit (Figs. 2 A and 3 A, and Table 1).

The false discovery rate correction (FDR)

Applying the FDR correction with $P_S = 0.05$ (Eq. (4)), we identified 3 and 9 hotspots for the burrowing owl and the blue–black grassquit, respectively (Figs. 2 B and 3 B, and Table 1). Not surprisingly, the FDR correction identifies more hotspots than the FWER correction, because the FDR reduces the probability of false negatives (Type II errors).

Bayesian hierarchical modelling (BHM)

The Bayesian analysis requires that we assume values for z_S, C_I and C_{II}, to establish thresholds according to expressions (5) and (6). We considered three combinations for C_I and C_{II}: ($C_I = 1$, $C_{II} = 2$), ($C_I = 1$, $C_{II} = 1$), and ($C_I = 2$, $C_{II} = 1$) (Table 1). For example, the combination ($C_I = 1$, $C_{II} = 2$) means that the cost of missing a hotspot is twice as large as that of identifying a hotspot incorrectly and leads to $t_C = 0.333$ (see Eq. (6)). For the value of z_S we estimated the mean and standard deviation for λ from the data and then estimated z_S (Eq. (5)) such that the threshold would be close to 0.05. These led to $z_S = 1$ for the burrowing owl and $z_S = 1.5$ for the blue–black grassquit.

Concerning the hyperpriors, because we had no extra information on these parameters, we tried to have non-informative distributions. We assessed the impact of three different gamma distributions: Gamma(0.01,0.01), Gamma(0.1,1),

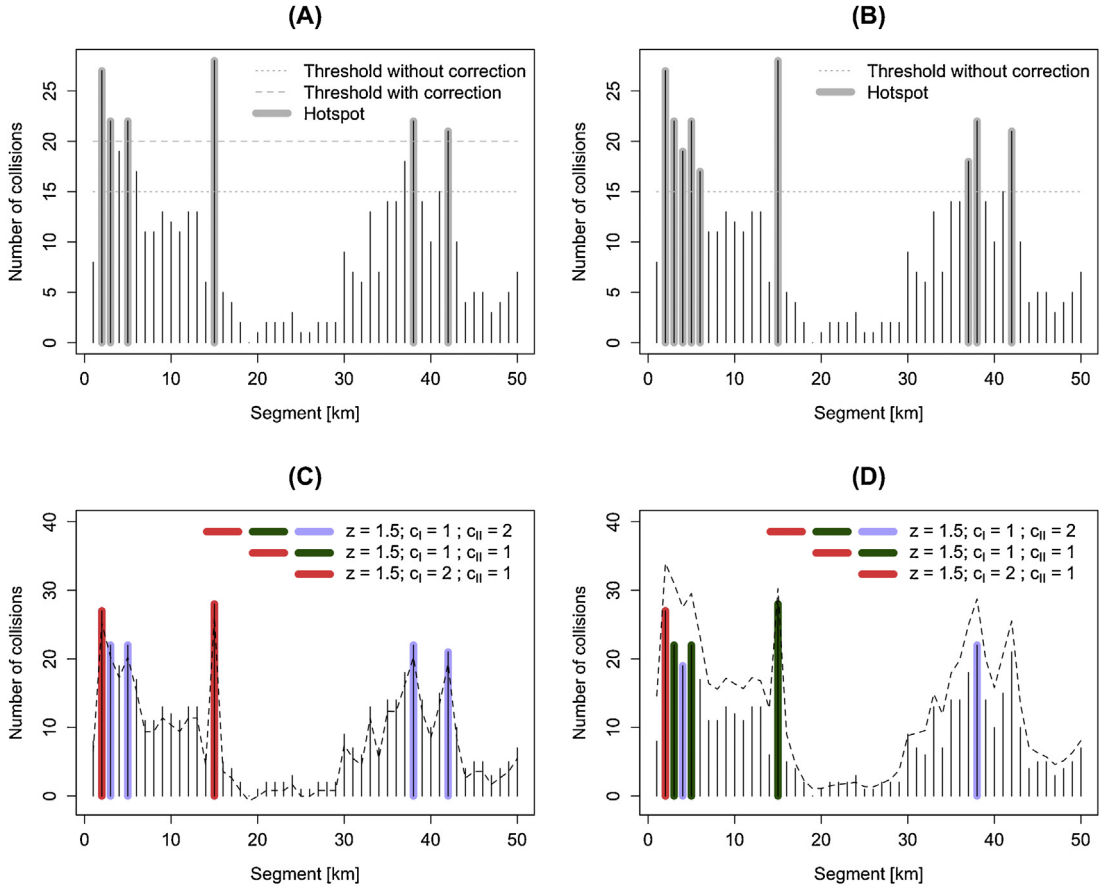

Fig. 3. Results for the blue–black grassquit (*Volatinia jacarina*). The black vertical lines depict the number of collisions along the 50 1-km road segments. Plot A: the grey dotted line shows the threshold when no corrections are applied (leading to 10 hotspots), and the grey dashed line shows the threshold after application of the familywise error rate correction (leading to 6 hotspots) identified by the vertical grey bar. Plot B: The application of the false discovery rate correction leads to the identification of 9 hotspots identified by the vertical grey bars. Plot C: The application of the Bayesian hierarchical Poisson-gamma model without spatial autocorrelation leads to the hotspots identified by vertical coloured bars: in red are the hotspots identified when $C_I = 2$ and $C_{II} = 1$, in red and green when $C_I = 1$ and $C_{II} = 1$, and in red, green and blue when $C_I = 1$ and $C_{II} = 2$. The broken line corresponds to the average value of the posterior of each λ_i (the mean rate of collisions per unit of length estimated during a certain period of time). Plot D: The results of the Bayesian hierarchical model with spatial autocorrelation, using the same convention for the hotspots as in plot C. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and Gamma(1,1). Although tests revealed that the choice of prior did not affect the conclusions significantly (results not shown), we used the Gamma(1,1) because it has a more flat distribution for the range of λ observed, thus closer to a non-informative distribution.

Because the models are easier, we start the Bayesian analyses assuming no spatial autocorrelation. We ran the analysis in the R environment and JAGS using 10,000 burn in iterations, and collected information on the parameters of interest using 50,000 iterations. As expected, the number of hotspots depends on the relative values of C_I and C_{II} and they are within the range of the values obtained with the two previous corrections (Figs. 2 C and 3 C and Table 1). If we were averse to missing hotspots ($C_I = 1, C_{II} = 2$), we would have identified 6 hotspots for both the burrowing owl and the blue–black grassquit. At the other extreme, if we are concerned about incorrectly identifying segments as hotspots ($C_I = 2, C_{II} = 1$) we could consider only 1 hotspot for the burrowing owl and 2 for the blue–black grassquit.

A simple test with the autocorrelation function (not shown) revealed that both data sets exhibit spatial autocorrelation, thus we re-analysed the data with the Bayesian CAR model. We ran the analyses in the R environment and RStan using four chains with 1000 burn in iterations, and collected information on the parameters of interest using 1000 iterations of each chain. We used a delta sampler control of 0.99. This sampler will make the program run slower but it can avoid divergent transitions after warmup (Stan Development Team, 2016a). To run the model efficiently in RStan we also used a sparse CAR representation (Stan Development Team, 2016b). As before, the number of hotspots depends on the relative values of C_I and C_{II} and they are within the range of the values obtained previously (Figs. 2 D and 3 D and Table 1). Adopting a policy of being averse to missing

hotspots ($C_I = 1, C_{II} = 2$), we identified 4 hotspots for the burrowing owl and 6 for the blue–black grassquit. At the other extreme, being concerned about incorrectly identifying segments as hotspots ($C_I = 2, C_{II} = 1$), we identified 2 hotspots for the burrowing owl and only 1 for the blue–black grassquit.

The neighbourhood of a segment plays a role when spatial autocorrelation is introduced. A segment identified as a hotspot in an analysis without spatial autocorrelation may turn out not to be a hotspot when spatial autocorrelation is taken into account if casualty counts are low in the neighbouring segments. Conversely, a segment not identified as a hotspot under the assumption of no spatial autocorrelation, may become a hotspot if the surrounding segments have a high count of occurrences. For instance, for the blue–black grassquit without spatial autocorrelation, Fig. 3C, hotspots that are not surrounded by other hotspots are common. On the other hand, segments surrounded hotspots not identified as hotspots in the analysis without spatial autocorrelation, are identified as hotspots when we consider spatial autocorrelation, Fig. 3D. Although the Bayesian analysis always takes into account the information on all segments, therefore implicitly taking into account the problem of multiple testing, when we include spatial autocorrelation the information on the neighbouring segments is also relevant to decide whether or not a segment is a hotspot.

Discussion

Our work was motivated by the high probability of identifying false hotspots (that is, of making Type I errors) inherent to methods that do not take into account corrections for multiple-testing, a common practice in road ecology research (e.g. Malo et al., 2004; Grilo et al., 2011; Planillo et al., 2015; Santos et al., 2015; Costa et al., 2015; Garriga et al., 2017), and we suggest three different methods to deal with this problem.

Besides the profound differences between frequentist and Bayesian schools, we do not see any method as being inherently better than the others. Ultimately, the choice of the procedure reflects the set of concerns and priorities of the researcher. For example, the FDR procedure should be used if a researcher is concerned about missing true hotspots (Type II error). On the other hand, if the concern is the probability of detecting false hotspots (Type I error) the researcher should use the FWER procedure. Nevertheless, our preference goes to the Bayesian approach because it allows a very clear specification of the researcher's concerns about making Type I or II errors, and, furthermore, it can take spatial autocorrelation into account. Since the two frequentist methods do not take into account spatial autocorrelation, we may say that they are "wrong" when spatial autocorrelation is present.

The frequentist methods exposed here are clearly easier to apply. Therefore, these are the methods we are most likely to apply if we just plan to check if a previous identification of hotspots is likely to have omitted any corrections for multiple testing, or if someone reports that in studies conducted at different times hotspots keep changing their locations. However, we would recommend using a Bayesian approach to identify hotspots. This may take longer to implement, but it is a time investment worth considering given its advantages, in particular, the possibility of stating explicitly the perceived costs of giving more or less weight to the costs of missing hotspots associated with different policies.

Furthermore, the Bayesian hierarchical models can easily incorporate explanatory variables that reflect the characteristics of the roads or of the surrounding environment, thereby leading to a better understanding of the causal relationships underlying mortality hotspots (e.g., Bartonička, Andrášik, Dul'a, Sedoník, & Bíl, 2018). To this end, the experience gained in the field of human accidents and prevention can be invaluable (e.g. Ahmed et al., 2011; Li, Carriquiry, Pawlovich, & Welch, 2008; Huang & Abdel-Aty, 2010). It should be noted that our approach assumed that there was no precise information on the location of the collisions and, instead, the number of collisions was given per segment. This is often the case for data on collisions of birds with power lines (F. Moreira, personal communication), but can also occur with road data (Gunson et al., 2009). If the precise location of the collisions is known, then other methods can be applied (e.g. Bíl et al., 2013, 2016).

The methods discussed here deal with the statistical aspects of identifying hotspots, yet they do not substitute the discussion on the societal choices underlying the need to identify hotspots or to adopt measures that mitigate their impacts. These choices are incorporated in the thresholds adopted in each method, and are explicitly stated under the Bayesian approach. From a strict conservation perspective, the important question is not whether individuals are killed by an infrastructure, but how such additional (non-natural) mortality impacts the viability of a population or of a species, or how much the depletion of a species affects the entire community (Borda-de-Água, Grilo, & Pereira, 2014). For instance, for the two species analysed here we observed a much larger number of casualties among the blue–black grassquit than among the burrowing owl. However, this is likely to reflect solely the higher abundance of one species relatively to the other, in particular, in the vicinity of roads, but what our analysis could not ascertain is how much the road mortality affects the viability of these populations. Both species have wide distributions and their conservation status is "least concern" (IUCN Species Survival Commission, 2000), but it is nevertheless possible that the levels of road mortality observed could endanger the local populations.

Some may argue that for conservation purposes it is preferable to identify false hotspots than to miss real ones. The problem with such a view is that efforts aimed at mitigating the negative impacts of an infrastructure are unlikely to wield the desired results.

In particular, if segments identified as hotspots are false positive hotspots then their locations are likely to continuously shift over time in a rather arbitrary way. This

undermines the credibility of conservation efforts in the long term and results in a waste of resources, including economic and manpower. In turn, this may discourage well-intentioned individuals or organizations due to the lack of clear results.

Authors contributions

LBA conceived the paper. All the authors contributed to the analysis of the results and the writing of the paper.

Acknowledgements

We thank Rodrigo Lima Augusto for providing the road mortality datasets and Giovani Silva, Paulo Soares and Juan Malo for reading an earlier version of this paper. This article is a result of the project NORTE-01-0145-FEDER-000007, supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). LBA, FA and RB were funded by Infraestruturas de Portugal Biodiversity Chair. FA was also supported by a FCT postdoctoral grant (SFRH/BPD/115968/2016). MS was funded by a PhD grant from Fundacão para a Ciência e a Tecnologia (FCT), Portugal (ref. PD/BD/128349/2017). All sources of funding are acknowledged in the manuscript, and the authors declare no direct financial benefits from its publication. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.baae.2018.11.001.

References

Ahmed, M., Huang, H., Abdel-Aty, M., & Guevara, B. (2011). Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. *Accident Analysis & Prevention*, *43*, 1581–1589.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Boca Raton: CRC Press.

Barrientos, R., Alonso, J. C., Ponce, C., & Palacin, C. (2011). Meta-analysis of the effectiveness of marked wire in reducing avian collisions with power lines. *Conservation Biology*, *25*, 893–903.

Bartonička, T., Andrášik, R., Dul'a, M., Sedoník, J., & Bíl, M. (2018). Identification of local factors causing clustering of animal-vehicle collisions. *The Journal of Wildlife Management*, *82*, 940–947.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Bíl, M., Andrašik, R., & Janoška, Z. (2013). Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis and Prevention*, *55*, 265–273.

Bíl, M., Andrašik, R., Svoboda, T., & Sedoník, J. (2016). The KDE+ software: A tool for effective identification and ranking of animal-vehicle collision hotspots along networks. *Landscape Ecology*, *31*, 231–237.

Borda-de-Água, L., Grilo, C., & Pereira, H. M. (2014). Modeling the impact of road mortality on barn owl (*Tyto alba*) populations using age-structured models. *Ecological Modelling*, *276*, 29–37.

Borda-de-Água, L., Barrientos, R., Beja, P., & Pereira, H. M. (2017). *Railway ecology*. Cham: Springer.

Costa, A. S., Ascensão, F., & Bager, A. (2015). Mixed sampling protocols improve the cost-effectiveness of roadkill surveys. *Biodiversity and Conservation*, *24*, 2953–2965.

Garriga, N., Franch, M., Santos, X., Montori, A., & Llorente, G. A. (2017). Seasonal variation in vertebrate traffic casualties and its implications for mitigation measures. *Landscape and Urban Planning*, *157*, 36–44.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton: Chapman & Hall/CRC.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*, 189–211.

Gitman, I., & Levine, M. D. (1970). An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique. *IEEE Transactions on Computers C*, *19*, 583–593.

Gomes, L., Grilo, C., Silva, C., & Mira, A. (2009). Identification methods and deterministic factors of owl roadkill hotspot locations in Mediterranean landscapes. *Ecological Research*, *24*, 355–370.

Grilo, C., Ascensão, F., Santos-Reis, M., & Bissonette, J. A. (2011). Do well-connected landscapes promote road-related mortality? *European Journal of Wildlife Researc*, *57*, 707–716.

Gunson, K. E., Clevenger, A., Ford, A., Bissonette, J. A., & Hardy, A. (2009). A comparison of data sets varying in spatial accuracy used to predict the occurrence of wildlife-vehicle collisions. *Environmental Management*, *44*, 268–277.

Gunson, K. E., & Teixeira, F. Z. (2015). Road–wildlife mitigation planning can be improved by identifying the patterns and processes associated with wildlife-vehicle collisions. In R. van der Ree, D. J. Smith, & C. Grilo (Eds.), *Handbook of road ecology* (pp. 101–109). Sussex: John Wiley & Sons.

Heydecker, B. G., & Wu, J. (2001). Identification of sites for road accident remedial work by Bayesian statistical methods: An example of uncertain inference. *Advances in Engineering Software*, *32*, 859–869.

Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, *83*, 2027–2036.

Huang, H., & Abdel-Aty, M. (2010). Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis & Prevention*, *42*, 1556–1565.

IUCN Species Survival Commission. (2000). *IUCN Red List of threatened species*. Gland, Switzerland: World Conservation Union. http://www.iucnredlist.org. (Accessed 8 June 2018)

Jin, X., Bradley, P. C., & Sudipto, B. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, *61*, 950–961.

Kolowski, J. M., & Nielsen, C. K. (2008). Using Penrose distance to identify potential risk of wildlife–vehicle collisions. *Biological Conservation*, *141*, 1119–1128.

Levine, N. (2004). *CrimeStat III: A spatial statistics program for the analysis of crime incident locations (version 3.0)*. Washington, DC: Ned Levine & Associates.

Li, W., Carriquiry, A., Pawlovich, M., & Welch, T. (2008). The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. *Accident Analysis & Prevention*, *40*, 1531–1542.

Lord, D., & Miranda-Moreno, L. F. (2008). Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. *Safety Science*, *46*, 751–770.

Malo, J. E., Suárez, F., & Díez, A. (2004). Can we mitigate animal–vehicle accidents using predictive models? *Journal of Applied Ecology*, *41*, 701–710.

Miranda-Moreno, L. F., Labbe, A., & Fu, L. (2007). Bayesian multiple testing procedures for hotspot identification. *Accident Analysis & Prevention*, *39*, 1192–1201.

Planillo, A., Kramer-Schadt, S., & Malo, J. E. (2015). Transport infrastructure shapes foraging habitat in a raptor community. *PLoS One*, *10*, e0118604.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*.

Ponce, C., Alonso, J. C., Argandona, G., García Fernandez, A., & Carrasco, M. (2010). Carcass removal by scavengers and search accuracy affect bird mortality estimates at power lines. *Animal Conservation*, *13*, 603–612.

Quinn, M., Alexander, S., Heck, N., & Chernoff, G. (2011). Identification of bird collision hotspots along transmission power lines in Alberta: An expert-based geographic information system (GIS) approach. *Journal of Environmental Informatics*, *18*, 12–21.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

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*, 1–12.

Santos, S. M., Marques, J. T., Lourenço, A., Medinas, D., Barbosa, A. M., Beja, P., et al. (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., de Figueiredo, A. P., Bager, A., Aguiar, L. M., et al. (2016). Carcass persistence and detectability: reducing the uncertainty surrounding wildlife-vehicle collision surveys. *PLoS One*, *11*, e0165608.

Sokal, R. R., & Rohlf, F. J. (1995). *Biometry: The principles and practice of statistics in biological research* (3rd ed.). New York: W.H. Freeman and Company.

Stan Development Team. (2016a). *Stan modeling language users guide and reference manual, version 2.14.0.*. http://mc-stan.org/documentation/case-studies/mbjoseph-CARStan.html. (Accessed 8 June 2018)

Stan Development Team. (2016b). *RStan: The R interface to Stan, R package version 2.14.1.*. http://mc-stan.org. (Accessed 8 June 2018)

van der Ree, R., Smith, D. J., & Grilo, C. (2015). *Handbook of road ecology*. pp. 2015. Sussex: John Wiley & Sons.

Zhang, J., Matthaiou, M., Karagiannidis, G. K., & Dai, L. (2016). On the multivariate gamma–gamma distribution with arbitrary correlation and applications in wireless communications. *IEEE Transactions on Vehicular Technology*, *65*, 3834–3840.