Journal of Animal Ecology

British Ecological Society

Journal of Animal Ecology 2013 doi: 10.1111/1365-2656.12080

LOCATION-ONLY AND USE-AVAILABILITY DATA

Reconciling resource utilization and resource selection functions

Mevin B. Hooten[1]*, Ephraim M. Hanks[2], Devin S. Johnson[3] and Mat W. Alldredge[4]

[1]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Departments of Fish, Wildlife & Conservation Biology and Statistics, Colorado State University, Fort Collins, CO, USA;* [2]*Department of Statistics, Colorado State University, Fort Collins, CO, USA;* [3]*National Marine Mammal Laboratory, National Oceanic and Atmospheric Administration, Seattle, WA, USA; and* [4]*Colorado Parks and Wildlife, Fort Collins, CO, USA*

Summary

1. Analyses based on utilization distributions (UDs) have been ubiquitous in animal space use studies, largely because they are computationally straightforward and relatively easy to employ. Conventional applications of resource utilization functions (RUFs) suggest that estimates of UDs can be used as response variables in a regression involving spatial covariates of interest.

2. It has been claimed that contemporary implementations of RUFs can yield inference about resource selection, although to our knowledge, an explicit connection has not been described.

3. We explore the relationships between RUFs and resource selection functions from a heuristic and simulation perspective. We investigate several sources of potential bias in the estimation of resource selection coefficients using RUFs (e.g. the spatial covariance modelling that is often used in RUF analyses).

4. Our findings illustrate that RUFs can, in fact, serve as approximations to RSFs and are capable of providing inference about resource selection, but only with some modification and under specific circumstances.

5. Using real telemetry data as an example, we provide guidance on which methods for estimating resource selection may be more appropriate and in which situations. In general, if telemetry data are assumed to arise as a point process, then RSF methods may be preferable to RUFs; however, modified RUFs may provide less biased parameter estimates when the data are subject to location error.

Key-words: kernel density estimation, space use, spatial statistics, utilization distribution

Introduction

Resource utilization function (RUF) analyses (Marzluff *et al*. 2004) are widely employed in the study of animal space use and enjoy the advantages of being relatively intuitive and comparatively easy to implement. Based on the estimation of an individual-based 'utilization distribution' (UD; e.g. Millspaugh *et al*. 2006), RUF analyses are commonly intended to obtain inference about the relationship between an animal or population's use of space and the underlying environmental niche. This desired inference is a critical component in the field of ecology (Krebs 1978). In linking the UD to the underlying environment, RUF analyses go beyond that of home range and core area (e.g. Wilson *et al*. 2010) estimation and also relate to resource selection analyses, where desired inference pertains to whether the use of resources is disproportionate to those available (Manly *et al*. 2002). One potential advantage of the RUF approaches is that they may improve selection inference when the telemetery data are subject to measurement error (Millspaugh *et al*. 2006).

Resource utilization function analyses hold an appeal because of their simplicity, but the specific connection in how they relate to resource selection functions (RSFs) has not been described. In this paper, we attempt to reconcile RUF analysis with RSF analysis and examine several sources of potential bias in doing so. We begin by

*Correspondence author. E-mail: Mevin.Hooten@colostate.edu

describing the RUF and RSF analyses as they are traditionally employed. We then explore several potential sources of bias that could affect RUF analyses and use simulation to demonstrate our findings. Finally, we suggest a few simple diagnostics that could be employed in selection analyses and illustrate them using a real data set pertaining to the spatial ecology of mountain lions (*Puma concolor*) in Colorado, USA.

RESOURCE UTILIZATION FUNCTIONS

The conventional perspective in animal space use studies is that the UD is a spatial probability distribution that gives rise to a spatial point process (i.e. the observed telemetry locations). That is, one assumes there is a surface over a spatial domain ($\mathcal{S}$) of interest that specifies the likelihood (f) an animal will occur at any given location ($\mathbf{s}$) in the domain. Thus, for a finite set of times at which an animal's location is observed, say $t = 1,\ldots,T$, we have a statistical model for location where $\mathbf{s}_t \sim f(\mathbf{s})$, for $\mathbf{s}_t \in \mathcal{S}$.

The RUF procedure outlined by Marzluff *et al.* (2004) assumes that the probability distribution f (i.e. the UD) then depends on the underlying environment $\mathbf{X}$ (i.e. $f(\mathbf{s}) \equiv f(\mathbf{s}|\mathbf{X},\boldsymbol{\beta})$) and adopts a two-stage estimation approach for the coefficients $\boldsymbol{\beta}$. The first step in the analysis is concerned with estimating the UD (with say, $\hat{f}$), while the second stage links the UD to a set of underlying covariates $\mathbf{X}$.

To estimate the UD, a wide variety of density estimation techniques can be employed to find $\hat{f}$ based on the telemetry data ($\mathbf{s}_t$); however, we will focus on kernel density estimation (KDE), because (i) this is a commonly applied technique familiar to many animal ecologists and (ii) Marzluff *et al.* (2004) employed this approach in their seminal paper on the topic. It should be noted, however, that many of the following results would apply to RUFs based on any form of UD estimation technique.

In KDE, one takes a nonparametric approach to estimating f whereby for any location of interest $\mathbf{c} = (c_1, c_2)\prime$ in the spatial domain $\mathcal{S}$, the estimate of the UD is as follows:

$$\hat{f}(\mathbf{c}) = \frac{\sum_{t=1}^{T} k((c_1 - s_{1,t})/b_1)k((c_2 - s_{2,t})/b_2)}{Tb_1b_2} \qquad \text{eqn 1}$$

where $\mathbf{s}_t = (s_{1,t}, s_{2,t})\prime$, k represents the kernel (which we assume to be Gaussian) and the parameters b_1 and b_2 are bandwidth parameters that control the diffuseness of the kernel (Venables & Ripley 2002, Chapter 5). There are various ways to choose the bandwidth parameters, and these are well described in the literature (e.g. Silverman 1986). In practice, the UD, $f(\mathbf{c}_i)$, is estimated for a large but finite set of points (or grid cells, $i = 1,\ldots,m$) in the spatial domain $\mathcal{S}$ for the purposes of graphical display or further use in a RUF model.

Consider, as an illustration, the situation where there is a single covariate of interest x and telemetry locations are simulated from $f(\mathbf{s}|x, \beta_0, \beta_1)$ (Fig. 1). In this case, the coefficients were chosen to provide a positive relationship between the covariate and the UD (i.e. $\beta_1 > 0$, where β_0 only has an effect on the total number of observed telemetry locations T). Figure 1 depicts a large-scale spatial pattern in the covariate where the telemetry data are constrained to the unit square region shown; this constraint serves as the 'home range' and could take any shape, but the rectangular shape is used here for display purposes only. We will show that the spatial pattern in the covariate, which is only a function of the spatial arrangement of the landscape, will prove to be an important factor in the spatially explicit models that follow.

A conventional RUF analysis typically proceeds by fitting a linear model with $\hat{f}(\mathbf{c}_i)$ as the response variable and $\mathbf{x}(\mathbf{c}_i)$, a $p \times 1$ vector, representing the covariates (i.e. environmental resources) at location $\mathbf{c}_i$. That is, the second stage of the RUF analysis for an individual involves fitting the regression model:

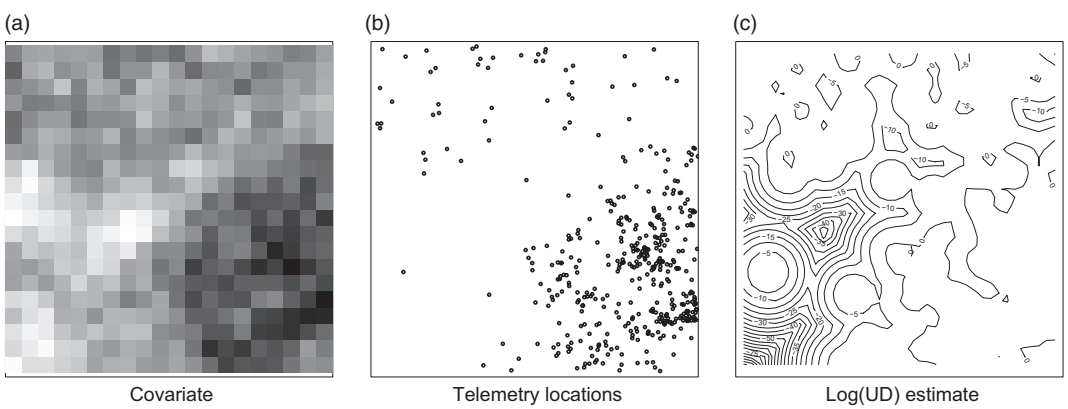

Fig. 1. (a) Spatial covariate x, (b) simulated telemetry locations $\mathbf{s}_t$, for $t = 1,\ldots,400$, and (c) the log transformed KDE representing the estimated UD based on the simulated data.

$$\hat{f}(\mathbf{c}_i) = \beta_0 + \mathbf{x}(\mathbf{c}_i)'\boldsymbol{\beta} + \varepsilon_i, \qquad \text{eqn 2}$$

for i=1,…,m and $\varepsilon_i \sim N(0, \sigma^2)$, where the regression coefficients $\boldsymbol{\beta}$ control the linear relationship between the environmental covariates and the UD, and β_0 corresponds to an intercept parameter that is not typically interpreted.

At the individual level, RUF analysis provides inference about the regression coefficients $\boldsymbol{\beta}$ in terms of significance and possibly subset selection, thereby illuminating the potential environmental influences on space use. In a population-level analysis, where telemetry data exist for multiple individuals (say, $\mathbf{s}_{j,t}$ for $j = 1,\ldots,J$ individuals) one would index the regression coefficients $\boldsymbol{\beta}_j$ such that they are labelled for each individual. Then, the focus shifts towards the expectation or variance in coefficient estimates $\hat{\boldsymbol{\beta}}_j$ among individuals; for example, we may be interested in learning about $\boldsymbol{\mu}_\beta = E(\hat{\boldsymbol{\beta}}_j)$ for all $j = 1,\ldots,J$ animals. In this latter case, the individual becomes the sample unit and the sample size J most heavily influences the uncertainty concerning $\boldsymbol{\mu}_\beta$.

In implementing the RUF approach described previously, Marzluff *et al.* (2004) wisely noticed that there may be lurking forms of dependence in the regression errors ε_i. They posited that such forms of dependence might arise from the smoothing induced by the KDE approach for estimating the UD (eqn 1) [in addition to other possible sources of latent autocorrelation such as missing covariates in eqn (2)]. Marzluff *et al.* (2004) propose a geostatistical approach (Cressie 1993) that involves modelling the covariance structure between the errors ε_i in a spatially explicit manner. A simple geostatistical model for the RUF analysis is the exponential spatial model given by:

$$\text{cov}(\varepsilon_i, \varepsilon_l) = \sigma_e^2 + \sigma_s^2 \exp\left(-\frac{||\mathbf{c}_i - \mathbf{c}_l||}{\phi}\right), \qquad \text{eqn 3}$$

where the numerator in the exponential refers to the Euclidean distance between cell i and cell l, and the denominator ϕ is a range parameter that controls the decay in the spatial structure of $\boldsymbol{\varepsilon}$ with distance. The two variance components σ_e^2 (nugget) and σ_s^2 (sill) account for the variance associated with a non-spatially structured and spatially structured source of error, respectively. In matrix notation, the model for the errors is then often expressed as $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_m)\prime$ and the $(i,l)^{\text{th}}$ element of the covariance matrix $\boldsymbol{\Sigma}$ is equal to (3). Often, the covariance matrix is written as $\boldsymbol{\Sigma} = \sigma_e^2\mathbf{I} + \sigma_s^2\mathbf{R}(\phi)$.

The conventional procedure used to fit geostatistical models to continuous spatial data involves a multi-step process of first (i) fitting the linear regression model assuming independent errors, then (ii) characterizing the spatial structure in the residuals using variogram estimation (Cressie 1993), and finally (iii) using generalized (or weighted) least squares (GLS) to estimate the regression coefficients ($\boldsymbol{\beta}$) while taking into account the correlated errors. Other approaches such as maximum likelihood can also be used, but for simplicity, we retain the GLS method in our simulations.

RESOURCE SELECTION FUNCTIONS

Resource selection is the differential use of resources given those resources available. In describing the conventional approach for estimating RSFs (e.g. Manly *et al.* 2002; Johnson *et al.* 2006), we note that most recent applications of RSFs take a weighed distribution approach where the probability distribution of use $f_u(\mathbf{x})$ can be expressed as an updated distribution of availability $f_a(\mathbf{x})$ given the RSF $g(\mathbf{x},\boldsymbol{\beta})$ which is usually expressed in an exponential form as $g(\mathbf{x},\boldsymbol{\beta}) = \exp(\mathbf{x}'\boldsymbol{\beta})$ (although other functional forms are possible, e.g., Lele & Keim 2006). This equivalence between use and the updated version of availability can be written as:

$$f_u(\mathbf{x}) = \frac{g(\mathbf{x},\boldsymbol{\beta}) f_a(\mathbf{x})}{\int g(\mathbf{v},\boldsymbol{\beta}) f_a(\mathbf{v}) d\mathbf{v}}, \qquad \text{eqn 4}$$

because the distribution of use $f_u(\mathbf{x})$ is not observed directly, a maximum likelihood approach can be taken to maximize a product over the right-hand-side of eqn (4) with respect to $\boldsymbol{\beta}$:

$$\prod_{t=1}^{T} \frac{g(\mathbf{x}(\mathbf{s}_t),\boldsymbol{\beta}) f_a(\mathbf{x}(\mathbf{s}_t))}{\int g(\mathbf{v},\boldsymbol{\beta}) f_a(\mathbf{v}) d\mathbf{v}}. \qquad \text{eqn 5}$$

Various tricks can be employed to maximize (eqn 5) without having to analytically solve the integral in the denominator (e.g. Johnson *et al.* 2006; Lele 2009). The most common approach involves taking a 'background' sample (sometimes referred to as an availability sample) of locations from $\mathcal{S}$ and labelling those as zeros in a binary response vector with the ones corresponding to the observed telemetry locations. A logistic regression is then fit to the binary data using the covariates at all of the used and available locations. Under certain conditions, the parameter estimates $\hat{\boldsymbol{\beta}}$ have been shown to be equivalent to those obtained by maximizing (eqn 5). Incidentally, Warton and Shepherd (2010) and Aarts *et al.* (2012) have recently shown that maximizing (eqn 5) is equivalent to maximizing the likelihood of an inhomogeneous spatial point process for the purpose of estimating $\boldsymbol{\beta}$. Furthermore, Aarts *et al.* (2012) show that the required maximization can be achieved using a Poisson generalized linear model (GLM), with an offset term corresponding to availability.

To fit the Poisson GLM, one bins the telemetry locations into a large set of grid cells spanning the spatial domain $\mathcal{S}$, and the resulting response variable $y(\mathbf{c}_i)$ (for $i = 1,\ldots,m$ grid cells) consists of cell counts where the model is expressed as $y(\mathbf{c}_i) \sim \text{Pois}(\lambda(\mathbf{c}_i)a(\mathbf{c}_i))$, and a log link is used to model the intensities $\lambda(\mathbf{c}_i)$:

$$\log(\lambda(\mathbf{c}_i)) = \beta_0 + \mathbf{x}(\mathbf{c}_i)'\boldsymbol{\beta}, \qquad \text{eqn 6}$$

where if the availability weights $a(\mathbf{c}_i)$ are all equal (i.e. even availability within the region $\mathcal{S}$), then this procedure becomes a regular Poisson log-linear regression of the cell counts on the covariates without weights. In what follows, we set all $a(\mathbf{c}_i) = 1$; however, if $a(\mathbf{c}_i)$ are set to be the area of the grid cells, then $\lambda(\mathbf{c}_i)$ can be interpreted as the average number points per unit area.

Reconciling RUFs and RSFs

From one perspective, some might argue that the big difference between the RUF and RSF analyses is that a UD (i.e. $\hat{f}$) is estimated prior to fitting the RUF model, whereas in the RSF approach, the UD is implicitly estimated as a function of the spatial covariates (i.e. $\hat{\lambda} = \exp(\mathbf{x}'\hat{\boldsymbol{\beta}})$) based on the data directly. On the other hand, even though it may not be obvious, the Poisson regression employed to fit the RSF is also estimating the UD first as a 2-D spatial histogram (i.e. $y(\mathbf{c}_i)$, for $i = 1, \ldots, m$) at the scale of the underlying grid. In this sense, the grain size (i.e. $a(\mathbf{c}_i)$) of the cells in the grid over which the telemetry locations are summed is equivalent to the bandwidth parameters in the KDE for the RUF approach. That is, if $a(\mathbf{c}_i)$ increases, then $y(\mathbf{c}_i)$ becomes a smoother process over $\mathcal{S}$ (similar to increasing the bandwidth in the KDE). In both cases, as the smoothness in the estimated point process density increases, it yields a more biased density estimate; however, it also decreases the variance; therefore, the choice in the amount of smoothing to apply involves some notion of optimality.

Perhaps, a bigger concern is how the RUF fitting procedure affects the estimation of selection coefficients $\boldsymbol{\beta}$, as these coefficients are typically our main focus. When population-level inference is desired, some have argued that the uncertainty associated with our knowledge of $\boldsymbol{\beta}_j$ for individual animals $j = 1, \ldots, J$, is a minor concern compared with the sample size of individuals J (e.g. Otis & White 1999); for this reason, we focus only on bias in the estimation of $\boldsymbol{\beta}_j$ at the individual-level herein. That is, individual-level bias will have the biggest and most dubious effect on population-level inference when the number of telemetered individuals is large; thus, it is our focus here.

In an examination of RUFs and RSFs, we discovered the following important differences between methods when used to estimate resource selection:

1. The use of $\hat{f}$ instead of $\log(\hat{f})$ in conventional RUFs.
2. The characterization of availability via the choice of $\mathcal{S}$.
3. The marginal smoothing induced by the UD KDE.
4. The pattern of covariates in the spatial RUF.
5. The possibility of location error in telemetry data.

We discuss each of these items in turn, providing some insight into how they play a role in the estimation of resource selection, and we also suggest some modifications for reconciling RUFs and RSFs.

THE USE OF $\hat{f}$ INSTEAD OF LOG$(\hat{f})$ IN CONVENTIONAL RUFS

Based on the assumptions of the Poisson point process model, the density f and intensity λ are related by: $f(\mathbf{c}_i) = \lambda(\mathbf{c}_i)/\int_{\mathcal{S}} \lambda(\mathbf{s}) d\mathbf{s}$, where the denominator is the expected number of points in the study area $\mathcal{S}$. Thus, the Poisson intensities $\lambda(\mathbf{c}_i)$ governing the point process (i.e. telemetry data) are proportional to the densities $f(\mathbf{c}_i)$ being modelled in the RUF analysis. That is,

$$\lambda(\mathbf{c}_i) = \text{const} \times f(\mathbf{c}_i), \qquad \text{eqn 7}$$

where the 'const' term is related to the number of telemetry locations T in the data set. Thus, because of the two-stage fitting procedure in the RUF analysis, it would be considered an approximation to the RSF analysis if the log transformation was applied to the estimated density function $\hat{f}(\mathbf{c}_i)$ (at least in terms of estimating $\boldsymbol{\beta}$). That is, if the second stage (eqn 2) of the RUF model was modified such that

$$\log(\hat{f}(\mathbf{c}_i)) = \beta_0 + \mathbf{x}(\mathbf{c}_i)'\boldsymbol{\beta} + \varepsilon_i, \qquad \text{eqn 8}$$

where β_0 implicitly includes '-log(const),' then the main difference between the RSF (eqn 6) and the RUF (eqn 8) would be the Poisson instead of Gaussian error, respectively. Furthermore, from a practical perspective, the log transformation expands the support of the response variable in the RUF model (eqn 8) from the positive to the real numbers. Thus, in the remainder of the article, we refer to eqn (8) as the RUF model and examine its properties.

THE CHARACTERIZATION OF AVAILABILITY VIA THE CHOICE OF $\mathcal{S}$

Recall that resource selection is the degree of use given resource availability. If RUFs are approximations to RSFs, then how does availability play a role in RUF analyses? A surprising amount of variation in the estimation of $\boldsymbol{\beta}$ can be observed by simply changing how the background sample is taken. This background sample provides a Monte Carlo approximation of the integral in the weighed distribution (eqn 4, and associated point process model), and the spatial extent of the integral ($\mathcal{S}$) is what controls availability in the RSF under the assumption of uniform availability in that region. Millspaugh *et al.* (2006) recommend defining $\mathcal{S}$ based on the UD itself. We agree that areas outside of the natural availability to the individual animal should not be considered in RSF analyses. The region of potential space use or home range is typically thought to be a function of external and/or internal biological forces either constraining (e.g. territorial behaviour) or attracting (e.g. central place foragers) movement. Thus, assuming uniform availability over $\mathcal{S}$, both RSF and RUF analyses account for

availability simply by limiting the spatial support of the response variable in the model in question. If one takes a Poisson GLM approach to fitting a RSF, then the extent of the grid over which the telemetry locations are counted acts as the spatial support in the model and, in the case of the RUF (eqn 8), it is the grid over which the UD is estimated. Given that overly conservative availability extents can cause a dramatic bias in the results, the recommendation by Millspaugh *et al.* (2006) to use a large isopleth of the estimated UD is sensible.

THE MARGINAL SMOOTHING INDUCED BY THE UD KDE

As Marzluff *et al.* (2004) point out, there is an inherent marginal (i.e. not explicitly considering the covariates) smoothing that is induced in the estimation (eqn 1) of the point process density $\hat{f}$ based on the telemetry locations $\mathbf{s}_i$. It is not easy to see how this smoothing manifests itself when the log UD (eqn 8) is used as a response variable in the RUF model because of the complex nature of the KDE procedure. However, we can write out a heuristically similar model that is based on smoothing the response variable directly. In this case, to simplify the notation, let y represent a non-smoothed representation of the log UD, then suppose the log UD is generated as $\mathbf{y} \sim N(\mathbf{X\beta}, \sigma^2\mathbf{I})$. Now, if we apply a linear smoother to the log UD ($\mathbf{Wy}$) that is based on a weighing of the $\mathbf{y}$ at all locations, then using the properties of a multivariate normal distribution, we have the correct model for the smoothed log UD:

$$\log(\hat{\mathbf{f}}) \approx \mathbf{Wy} \sim N(\mathbf{WX\beta}, \sigma^2\mathbf{WW}'). \qquad \text{eqn 9}$$

Using similar notation, the RUF model (eqn 8) is akin to $\mathbf{Wy} \sim N(\mathbf{X\beta}, \sigma^2\mathbf{I})$. Thus, if the log UD is obtained via marginal smoothing of the point process, the RUF model (eqn 8) is misspecified. In fact, a more appropriate specification would be similar to that presented in eqn (9). The problem is that we do not know the exact form of the smoother matrix $\mathbf{W}$, and it will vary with the choice of marginal density estimator.

The effect of using the misspecified RUF model (eqn 8) on the estimation of the selection coefficients $\boldsymbol{\beta}$ is that the smoothing operator will be applied to the log UD but not to the mean field $\mathbf{X\beta}$, hence inducing a bias in $\hat{\boldsymbol{\beta}}$. This implies that regardless of whether ordinary least squares (OLS) or GLS is used to estimate $\boldsymbol{\beta}$, we will obtain biased selection coefficients. A possible remedy for this situation, because the exact form of $\mathbf{W}$ is unknown, is to try to induce a similar operator on $\mathbf{X\beta}$ by simply smoothing the covariates $\mathbf{X}$ before fitting the model; this yields the model

$$\log(\hat{f}) \sim N(\text{smooth}(\mathbf{X})\boldsymbol{\beta}, \sigma^2\mathbf{I}). \qquad \text{eqn 10}$$

This would not yield the correct model (eqn 9), but it would be an improvement. If the *post hoc* smoother $\text{smooth}(\mathbf{X}) = \tilde{\mathbf{W}}\mathbf{X}$ could be written as a linear smoother, then it could also be easily employed in the covariance structure yielding the model $\log(\hat{\mathbf{f}}) \sim N(\tilde{\mathbf{W}}\mathbf{X\beta}, \sigma^2\tilde{\mathbf{W}}\tilde{\mathbf{W}}')$. Alternatively, one could assume that a second-order covariance matrix estimated from the data in the geostatistical sense would serve as an approximation. This latter modification would yield:

$$\log(\hat{\mathbf{f}}) \sim N(\text{smooth}(\mathbf{X})\boldsymbol{\beta}, \boldsymbol{\Sigma}), \qquad \text{eqn 11}$$

a model quite similar to the spatial RUF proposed by Marzluff *et al.* (2004), but with covariates smoothed to the same degree as the log UDs.

THE PATTERN OF COVARIATES IN THE SPATIAL RUF

The previous sections show, at least heuristically, how a modified version of the RUF analysis could be considered as an approximation of the RSF analysis. Continuing the example using our simulated data from Fig. 1, we fit the linear model in eqn (8) assuming independent errors and then estimated the variogram and modelled it using the exponential form of spatial structure (eqn 3) previously discussed. The resulting variogram fit (Fig. 2) indicated that residual autocorrelation exists in our data even though it was simulated based on the relationship with the covariate alone. As Marzluff *et al.* (2004) suggest, this residual autocorrelation is likely due to the smoothing induced by the KDE of the UD. Because latent spatial autocorrelation exists in our simulated data, we would be wise to account for it so that we may obtain accurate inference about the parameters $\boldsymbol{\beta}$ in the RUF.

We make a slight modification to the specification of the spatially explicit RUF model such that, using matrix notation, we now have:

$$\begin{aligned} \log(\hat{f}) &= \beta_0 + \mathbf{X\beta} + \boldsymbol{\varepsilon}, \\ &= \beta_0 + \mathbf{X\beta} + \mathbf{Hz} + \boldsymbol{\eta}, \end{aligned} \qquad \text{eqn 12}$$

where each of the vectors is concatenated over all cells in $\mathcal{S}$, and the original error vector $\boldsymbol{\varepsilon} = (\varepsilon(c_1), \ldots, \varepsilon(c_m))'$ is

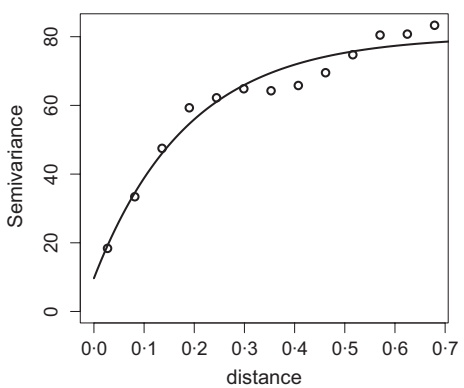

Fig. 2. Semi-variogram (points) and weighted least squares fit (line) of the exponential covariance model (eqn 3) resulting from the residuals of the linear regression using our simulated data.

now split into two pieces $\boldsymbol{\varepsilon} = \mathbf{Hz}+\boldsymbol{\eta}$; the first (i.e. $\mathbf{Hz}$) controlling the spatial dependence and the second (i.e. $\boldsymbol{\eta}$) accounting for any unstructured error. In fact, the spatially correlated errors arise from a normal distribution $\mathbf{Hz} \sim N(\mathbf{0}, (\tau\mathbf{Q})^{-1})$ where the precision matrix $\tau\mathbf{Q}$ is the inverse of the former covariance matrix (i.e. $(\sigma_s^2\mathbf{R})^{-1} = \tau\mathbf{Q}$) and the unstructured errors $\boldsymbol{\eta}$ are independent and identically normal such that $\boldsymbol{\eta} \sim N(\mathbf{0}, \sigma_e^2\mathbf{I})$. This reparameterization makes it easier to illustrate how second-order spatial dependence can impose a bias on the estimates of $\boldsymbol{\beta}$.

From eqn (12), it is apparent that the model contains two sets of covariates (i.e. $\mathbf{X}$ and $\mathbf{H}$). This implies that the covariates (i.e. columns) in $\mathbf{H}$ are spatial maps that may influence the log UD depending on a new set of regression coefficients $\mathbf{z}$. It can be shown that these 'spatial maps', acting as unobserved covariates, are actually eigenvectors of the aforementioned $\mathbf{Q}$ in the precision matrix, where $\mathbf{Q}=\mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$ (Clayton, 1993; Paciorek 2010). In other words, the spatial structure imposed by the geostatistical model (eqn 3) implies that there are an entire set of covariates in our model aside from those measured environmental variables $\mathbf{X}$! The parameters $\mathbf{z}$ then act as regression coefficients that control the relative importance of the latent covariates in $\mathbf{H}$ for predicting the log UD. Further, it can be shown that $\mathbf{z} \sim N(\mathbf{0}, (\tau\boldsymbol{\Lambda})^{-1})$ are random effects, where $\boldsymbol{\Lambda}$ is the diagonal eigenvalue matrix resulting from the spectral decomposition of $\mathbf{Q}$. The subtle but important consequence of having additional covariates $\mathbf{H}$ in the model is that they may be collinear with the known environmental covariates $\mathbf{X}$. This is potentially a big problem that is well described in the statistical literature (e.g. Clayton, 1993; Reich *et al*. 2006; Hodges & Reich 2010; Paciorek, 2010), although has received little attention in the ecological literature.

In our continued example with the simulated data shown in Fig. 1, we have computed the implied spatial covariate matrix $\mathbf{H}$ based on the variogram fit in Fig. 2 and illustrate the correlation with our covariate x using a few of the most important eigenvectors in $\mathbf{H}$ (Fig. 3). These three eigenvectors represent the second, third and fourth most important spatial patterns implied by the autocorrelation (Fig. 2) in the residuals of our simulated data. As implied spatial covariates in $\mathbf{H}$, each indicates an absolute correlation of approximately 0·5 with our simulated covariate x.

Several potential modifications have been suggested to alleviate the bias induced by collinearity between the covariates and spatially correlated errors (e.g. Reich *et al*. 2006; Hodges & Reich 2010; Hughes & Haran 2013); however, each of them would 'correct' the bias in the selection coefficients such that it is exactly equal to the non-spatial model fit using OLS. The suggested modifications (i.e. spatially restricted regression) have the additional effect of appropriately adjusting the variance of the estimators, but because we are primarily concerned with the bias here, we refer the interested reader to the cited literature herein.

THE EFFECT OF LOCATION ERROR IN TELEMETRY DATA

Hepinstall et al. (2004) hint that RUF methods were developed as an *ad hoc* procedure for fitting point process models. Before it was recognized that resource selection parameters could be estimated using readily available GLM fitting software, the required integration in the point process likelihood (eqn 4) made it challenging to fit point process models directly. If RSF methods are now just as accessible as RUFs, given their relationship, then what, if anything, do we gain when analysing telemetry data using the RUF approach? In this light, Millspaugh *et al*. (2006) claim that the RUFs are better able to handle measurement error in the telemetry data (i.e. with less bias). It seems reasonable that the marginal smoothing would help account for noise in the data; thus, using simulation, we evaluate this claim in the following section.

Data analysis

SIMULATION STUDY

We constructed a large simulation study to empirically verify the differences among the various methods for

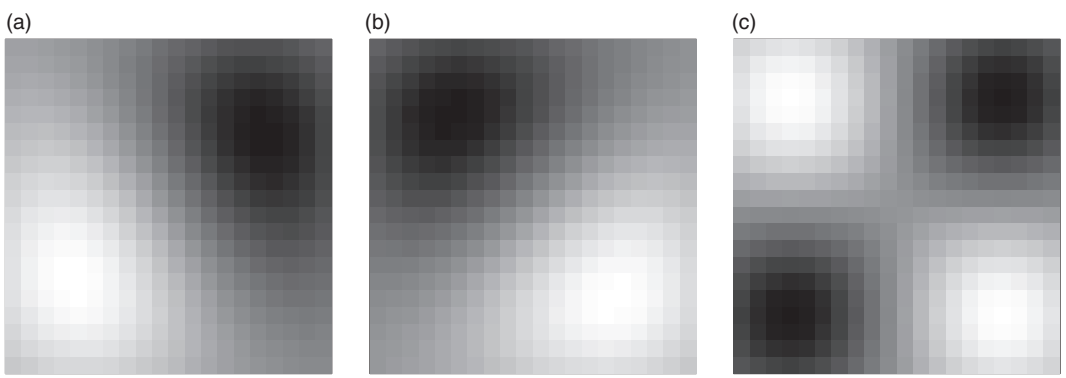

Fig. 3. (a) Second most important eigenvector in $\mathbf{H}$; correlation with $\mathbf{x}$ is 0·503 (b) third most important eigenvector in $\mathbf{H}$; correlation with $\mathbf{x}$ is −0·504, and (c) fourth most important eigenvector in $\mathbf{H}$; correlation with $\mathbf{x}$ is 0·468.

estimating resource selection. In doing so, we used a range of covariates (scaled to have mean zero and variance 1 on a 20×20 regular grid) from small-scale to large-scale, we varied the sample size from 25 to 400 independent telemetry points resulting from the intensity surface defined by $\lambda = e^{\beta_0+\beta_1 x}$, we varied the bandwidth in the KDE of the UD, and we used 3 different levels of measurement error by adding Gaussian noise to the simulated telemetry locations (with a variance of 0, 0·05 and 0·1, respectively). Further, we used a range of selection coefficient values from 0 to 2 and compared each of the following estimation procedures where the modified RUFs incorporate a degree of covariate smoothing that best improves the model fit (via R^2), and when we use the term 'spatial' here we are referring to the explicit spatially structured covariance version of the model:

1. PGLM: Poisson GLM described in Section Resource selection functions and eqn (6).
2. NSRUF: non-spatial RUF described in Section The use of $\hat{f}$ instead of $\log(\hat{f})$ in conventional RUFs and eqn (8) assuming independent and identically distributed errors ε_i.
3. SRUF: spatial RUF described in Section The use of $\hat{f}$ instead of $\log(\hat{f})$ in conventional RUFs and eqn (8) assuming correlated errors following the model (eqn 3).
4. NSMRUF: non-spatial modified RUF described in Section The marginal smoothing induced by the UD KDE and eqn (10) assuming independent and identically distributed errors ε_i.
5. SMRUF: spatial modified RUF described in Section The marginal smoothing induced by the UD KDE and eqn (11) assuming correlated errors following the model (eqn 3).

All analyses were carried out using the R Statistical Computing Environment (R Core Team, 2012; with R functions 'glm, 'variog' and 'variofit' from the 'geoR' package; Ribeiro & Diggle 2001). A subset of the results from the simulation study is presented in Table 1. The biases reported in Table 1 were approximated using 1000 simulations of point processes with a sample size of 100, $\beta_1 = 1$, large-scale covariates (i.e. range of spatial structure was approximately two- thirds of the maximum distance in the spatial domain), the plug-in bandwidth for the KDE estimate of the UD and over three different levels of location error in the data (i.e. none, small and moderate). To maintain the same sample size in each realization of the point process, we used an inflated value for β_0 and then thinned the simulated points. An alternative simulation approach would be to choose β_0 such that the desired sample size was merely the expected number of points (i.e. $E(T) = \int_S \lambda(\mathbf{s})d\mathbf{s}$), but this would not maintain a constant sample size across simulations.

The results presented in Table 1 hold generally across the full range of simulations performed and are representative of a broad range of scenarios. Overall, the most obvious pattern we notice is that the SRUF is the most biased method for estimating resource selection across all scenarios; we attribute this to two sources of bias: the marginal smoothing of the UD and the potential spatial confounding. The NSMRUF performs the best across all scenarios shown in Table 1; however, it was not unbiased in all simulations (not shown here), but it was always the second best method compared with the PGLM. In cases where there is measurement error, the NSMRUF and SMRUF do quite well in terms of bias, although no method stays completely unbiased when location error is present. The SRUF and SMRUF appear to pick up an additional source of bias that does not effect the non-spatial estimation procedures. Based on the literature and the high degree of correlation with the second-order spatial error (which was nearly always greater than 0·6, indicating collinearity), we suspect this additional bias may be caused by spatial confounding (e.g. Hodges & Reich 2010). Overall, these simulations support the arguments made in Section Reconciling RUFs and RSFs concerning the differences between methods and possible sources of bias.

Table 1. The first three columns display the results of the simulation study showing the bias incurred when estimating the resource selection coefficient β_1 using each of the methods under varying amounts of location error. The small and large location error corresponds to an additive symmetrical error with standard deviation of $1/20^{th}$ and $1/10^{th}$ of the maximum distance in the spatial domain, respectively. Bias values close to zero indicate the method is relatively unbiased for estimating resource selection. The last column shows the resource selection parameter estimates under the different methods. It is important to note that the last column displays the estimates themselves, whereas the previous three columns represent bias values

	Bias			Estimate
	Amount of location error			
	None	Small	Moderate	Mountain lion $\hat{\beta}_1$
PGLM	−0·008	−0·246	−0·402	−0·242
NSRUF	−0·288	−0·343	−0·428	−0·109
SRUF	−0·931	−0·943	−0·963	0·001
NSMRUF	−0·007	−0·077	−0·182	−0·317
SMRUF	−0·103	−0·175	−0·361	−0·145

MOUNTAIN LION DATA

To illustrate a diagnostic approach for performing a resource selection analysis using non-simulated data, we consider an individual mountain lion and a single covariate. The telemetry data are comprised of global positioning system (GPS) locations at a fairly regular fix interval of approximately 3 h in an ongoing Colorado Parks and Wildlife (CPW) monitoring effort. We focused on a single individual (# AF50, an adult female) to demonstrate a potential diagnostic procedure for determining the best resource selection approach for inference. We thinned the

original data, keeping only those points greater than 10 days apart to alleviate any concerns due to temporal autocorrelation (e.g. Swihart & Slade 1985). We used the topographical covariate of solar exposure (i.e. modified Beers' aspect transform; Beers 1966) for the analyses as this is a potentially important resource on the Colorado Front Range for these large carnivores (Fig. 4).

In using each of the methods to estimate resource selection, we found great variability in the point estimates for the parameter β_1 (Table 1; far right column). The SRUF demonstrated similar results as in our simulations, estimating β_1 far from any of the other estimates (and positive), while the spatial modified RUF estimate for β_1 seemed to improve (but was still not equivalent with the other methods). There did appear to be a slight difference between the estimates using the NSMRUF and the PGLM. Both performed well in our simulations where no location error was present. Given that these data were GPS telemetry locations, we would not expect location inaccuracy at the scale of our covariates. However, because the relationship between exposure and the UD may not be causal (which is not explored in our simulations), there may be missing covariates that could help explain resource selection. As a substitute for measurement error, this type of misspecification error may be accounted for in the RUF (but not in the RSF), although this is mostly speculative. It should be noted that the NSMRUF, SMRUF and the PGLM indicate that there is a negative effect of exposure on selection, implying that this individual is selecting for more protected aspects.

Both functions are trivial to estimate, but with the added smoothing for the covariates in the NSMRUF and SMRUF, the GLM is slightly more straightforward. It is clear that using the spatial models for this data set is not advised due to the additional bias. Further simulation based on the exposure covariate and a range of β_1 values encompassing those estimated here could provide additional guidance as to whether the RUF or RSF provides less biased resource selection inference. However, in this scenario, with no obvious source of measurement error, we would choose the RSF point process model (i.e. PGLM) for inference.

Conclusion

In an examination of the properties of both RUFs and RSFs, we find that generally the RSF is preferred because it is slightly easier to implement and yields unbiased inference about selection coefficients when no measurement error exists in the telemetry data. However, we note that when there is location uncertainty in the data, a modified version of the RUF can outperform the traditional RSF in terms of less bias in the estimation of selection coefficients. This advantage was mentioned by Millspaugh *et al*. (2006) but was not demonstrated, nor was the RUF reconciled with the RSF to provide inference about the same coefficients.

The residuals resulting from RUF models will typically indicate latent spatial autocorrelation, and normally, it would be a good idea to account for this; however, when using large-scale covariates, there is a high likelihood of multi-collinearity between the covariates and second-order spatial structure (i.e. spatial eigenvectors) inducing a bias in the resource selection coefficients. Thus, the spatial RUFs do not seem to provide valid inference about resource selection, at least in the conventional sense and in the range of scenarios we simulated.

Overall, it is evident that the original RUFs do, in fact, attempt to model some form of resource selection but that the coefficients obtained could not be expected to be comparable with those arising from fitting an RSF without some modification. Perhaps, the biggest finding we offer, aside from the potential spatial confounding induced by the second-order structure in the SRUF and SMRUF, is that the RUF approach can be modified (NSMRUF) such that it is a better estimator of resource selection (in terms of bias) than the traditional RSF when the data are subject to measurement error. This may be valuable in the analysis of VHF or ARGOS satellite telemetry data (which usually have more location uncertainty than GPS data). Although we have focused on simpler models herein, an

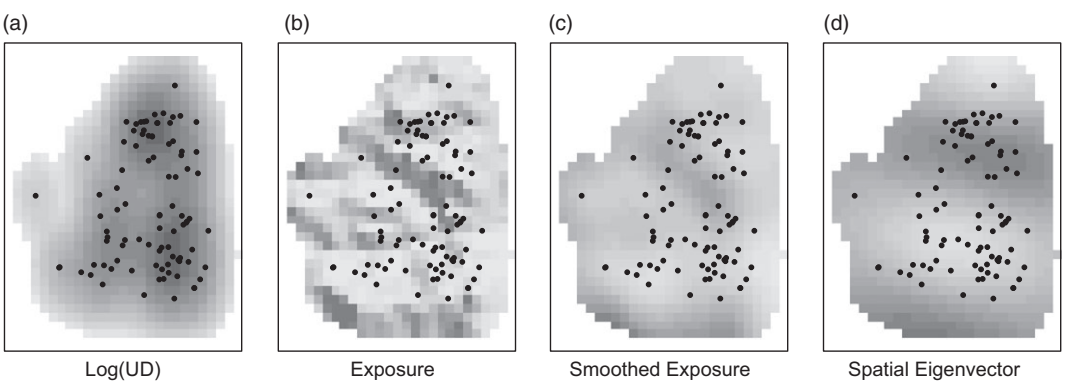

Fig. 4. The (a) mountain lion KDE log UD, (b) spatial covariate x (exposure), (c) smoothed covariate, and (d) spatial eigenvector that correlates most strongly with the covariate (correlation: $-0{\cdot}5$).

alternative framework could be constructed to explicitly account for any measurement error when making inference about resource selection (e.g. Johnson *et al*. 2008).

Finally, as a reminder, we note that the general RUF approach requires a two-stage procedure where the 'response' variable (i.e. the KDE) is first estimated using the original data and then it is statistically linked to covariates in a second-stage analysis. Like with all two-stage analyses, a potential shortcoming of the approach is that the uncertainty associated with the estimated density surface in the first stage is not accomodated in the second-stage analysis. One way to remedy this would be to employ either a bootstrapping, data augmentation or multiple imputation procedure to help account for any uncertainty in the KDE; however, at that point, one could argue that the RUF method may have lost its simple and straightforward appeal.

Acknowledgements

Funding for this project was provided by Colorado Parks and Wildlife (#1201). The use of trade names or products does not constitute endorsement by the U.S. Government.

References

Aarts, G., Fieberg, J. & Matthiopoulos, J. (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, **3**, 177–187.

Beers, T., Dress, P. & Wensel, L. (1966) Aspect transformation in site productivity research. *Journal of Forestry*, **64**, 691–692.

Clayton, D., Bernardinelli, L. & Montomoli, C. (1993) Spatial correlation in ecological analysis. *International Journal of Epidemiology*, **22**, 1193–1202.

Cressie, N.A.C. (1993) *Statistics for Spatial Data*. John Wiley & Sons, Inc., New York, USA.

Hepinstall, J.A., Marzluff, J.M., Handcock, M.S. & Hurvitz, P. (2004) Incorporating resource utilization distributions into the study of resource selection: dealing with spatial autocorrelation. *Resource Selection Methods and Applications* (ed. S. Huzurbazar), pp. 12–19. Omnipress, Madison, WI.

Hodges, J.S. & Reich, B.J. (2010) Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, **64**, 325–334.

Hughes, J. & Haran, M. (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, **75**, 139–159.

Johnson, C.J., Nielson, S.E., Merrill, E.H., McDonald, T.L. & Boyce, M.S. (2006) Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. *Journal of Wildlife Management*, **70**, 347–357.

Johnson, D.S., Thomas, D.L., Ver Hoef, J.M. & Christ, A. (2008) A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, **64**, 968–976.

Krebs, C. (1978) *Ecology: The Experimental Analysis of Distribution and Abundance*. Harper & Row Publishers Inc., New York, USA.

Lele, S.R. & Keim, J.L. (2006) Weighted distributions and estimation of resource selection probability functions. *Ecology*, **87**, 3021–3028.

Lele, S.R. (2009) A new method for estimation of resource selection probability function. *The Journal of Wildlife Management*, **71**, 122–127.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2002) *Resource Selection by Animals*. Kluwer Academic Publishers, Dordrecht.

Marzluff, J.M., Millspaugh, J.J., Hurvitz, P. & Handcock, M.S. (2004) Relating resources to a probabilistic measure of space use: forest fragments and stellar's jays. *Ecology*, **85**, 1411–1427.

Millspaugh, J.J., Nielson, R.M., McDonald, L.L., Marzluff, J.M., Gitzen, R.A., Rittenhouse, C.D., Hubbard, M.W. & Sheriff, S.L. (2006) Analysis of resource selection using utilization distributions. *Journal of Wildlife Management*, **70**, 384–395.

Otis, D.L. & White, G.C. (1999) Autocorrelation of location estimates and the analysis of radiotracking data. *Journal of Wildlife Management*, **63**, 1039–1044.

Paciorek, C.J. (2010) The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, **25**, 107–125.

R Core Team. (2012) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reich, B.J., Hodges, J.S. & Zadnik, V. (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, **62**, 1197–1206.

Ribeiro, P.J. & Diggle, P.J. (2001) geoR: a package for geostatistical analysis. *R-NEWS*, **1**, 15–18.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Swihart, R.K. & Slade, N.A. (1985) Testing for independence of observations in animal movements. *Ecology*, **66**, 1176–1184.

Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*, 4th edn. Springer, New York.

Warton, D.I. & Shepherd, L.C. (2010) Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *The Annals of Applied Statistics*, **4**, 1383–1402.

Wilson, R.R., Hooten, M.B., Strobel, B.N. & Shivik, J.A. (2010) Accounting for individuals, uncertainty, and multi-scale clustering in core area estimation. *Journal of Wildlife Management*, **74**, 1343–1352.

Received 9 August 2012; accepted 26 February 2013
Handling Editor: Wayne Thogmartin