Ecology, 87(12), 2006, pp. 3021–3028

WEIGHTED DISTRIBUTIONS AND ESTIMATION OF RESOURCE SELECTION PROBABILITY FUNCTIONS

Subhash R. Lele[1,3] and Jonah L. Keim[2]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1 Canada*
[2]*AXYS Environmental Consulting, Suite 203, 4208 97th Street, Edmonton, Alberta T6E 5Z9 Canada*

Abstract. Understanding how organisms selectively use resources is essential for designing wildlife management strategies. The probability that an individual uses a given resource, as characterized by environmental factors, can be quantified in terms of the resource selection probability function (RSPF). The present literature on the topic has claimed that, except when both used and unused sites are known, the RSPF is non-estimable and that only a function proportional to RSPF, namely, the resource selection function (RSF) can be estimated. This paper describes a close connection between the estimation of the RSPF and the estimation of the weight function in the theory of weighted distributions. This connection can be used to obtain fully efficient, maximum likelihood estimators of the resource selection probability function under commonly used survey designs in wildlife management. The method is illustrated using GPS collar data for mountain goats (*Oreamnos americanus* de Blainville 1816) in northwest British Columbia, Canada.

Key words: GPS collars; mountain goats; resource selection functions; simulated likelihood; telemetry data; use-available study design; used–unused study design.

Introduction

"When resources are used disproportionately to their availability, use is said to be selective" (Manly et al. 2002:15). Understanding the differential selection of resources by animals is an essential component of conservation biology, wildlife management and applied ecology (Boyce and McDonald 1999). Two common tools used for gathering such understanding are the resource selection probability function (RSPF) and the resource selection function (RSF). There are many excellent sources that describe the fundamental concepts and variety of applications of these tools. For example, Manly et al. (2002) discuss the statistical and ecological underpinnings of the resource selection by animals. Additionally, a recent edited volume (Huzurbazar 2003) and a special section of the *Journal of Wildlife Management* (March 2006) dedicated to the study of resource selection functions provides further evidence for the importance and use of resource selection functions in conservation biology, wildlife management, and other applied ecological studies. Given the wide availability of such literature, we concentrate on the problems and issues related to the statistical inference for the RSPF. Many studies that try to infer differential selection of resources by animals rely on sequential animal locations (Manly et al. 2002). A common assumption is that if an animal is present at a location, then the habitat at that location is being used. Furthermore, if an animal is present within a particular habitat disproportionate to the availability of that habitat, there is differential selection. Generally, information on the availability of different habitat types is obtained through biological surveys or information in geographic information systems (GIS).

The RSPF is a function that gives the probability that a particular resource, as characterized by a combination of environmental variables, will be used by an individual animal. Manly et al. (2002) discuss various models such as the exponential, logistic, probit and the log-log link models for RSPFs. They note that, under the sampling protocol where used locations and unused locations are observed, one can fit any of these RSPFs. However, if unused locations are unknown and available locations are randomly sampled, standard logistic regression procedures can be used to estimate parameters of the exponential RSPF, except its intercept parameter. This procedure provides *relative* probabilities of use, relative to a reference location, provided exponential RSPF is appropriate. Manly et al. (2002) term this relative

Manuscript received 28 February 2006; revised 24 April 2006; accepted 2 May 2006; final version received 12 June 2006. Corresponding Editor: G. M. Henebry.

[3] E-mail: slele@ualberta.ca

probability function a resource selection function (RSF). Specifically, if one reference location has a probability of use of 0.1 and another location has a probability of use of 0.5, then this procedure provides us with the information that the second location is five times more likely to be used than the first location. However, it does not offer any information on the absolute probabilities associated with these two locations. Thus, if the reference location has a probability of use of 0.001 and the second location has a probability of use of 0.005; the relative probability calculations will describe the second location as five times more likely to be used than the reference location; irrespective of their absolute probabilities.

The simplicity of estimating RSFs has made the exponential form of the RSPF popular in practice. However, as noted by Manly et al. (2002:Eq. 5.9), the exponential RSPF constrains the parameter values so that the exponent of the function is negative for all values of the covariates. Such constraints prompted Keating and Cherry (2004) to express reservations about the usefulness of the exponential form of the RSPF. Furthermore, Keating and Cherry (2004) recognized the difference between the use-available study design and the case-control study design. Studies have demonstrated that if some of the "controls" are in reality "cases," the simple logistic regression procedure leads to biased estimators (Lancaster and Imbens 1996). In the context of resource selection, the random sample of available resource units contains both "used" and "unused" units. Therefore, the estimation procedure in Manly et al. (2002) can potentially lead to biased estimators. Following the terminology of Lancaster and Imbens (1996), Keating and Cherry (2004) call this a contamination problem. They suggest that if the amount of contamination is small, the potential bias in the estimation of RSF is likely to be small as well.

Johnson et al. (2006), in response to Keating and Cherry (2004), argued that the use-available study design is properly formulated in terms of weighted distributions (Patil and Rao 1978). They proposed an alternative method based on logistic discriminant analysis (Seber 1984) and showed that, under the use-available sampling design (sampling protocol A, sampling design I; Manly et al. 2002), all parameters of the exponential RSPF, except for the intercept, can be estimated using standard logistic regression.

The purpose of this paper is to extend the ideas in Johnson et al. (2006). We demonstrate that parametric forms other than the exponential RSPF allow estimation of absolute probabilities. This method provides flexibility in modeling and generates absolute probabilities as opposed to relative probabilities. Estimating the absolute probabilities under use-available sampling designs will be a major advantage in the analysis of commonly collected survey and radio-collar data in ecology and wildlife management (McDonald and McDonald 2002). We describe in detail the simulated maximum likelihood estimation method (Robert and Casella 1999) for estimation of parameters for any RSPF, removing the restriction of employing only the exponential RSPF in the analysis of use-available data. Furthermore, we extend the method to practically important cases of location dependent and home range dependent distributions of available resources. We illustrate our method using GPS collar location data on mountain goats in northwest British Columbia. In addition to providing estimates and confidence intervals for the parameters, we show that the Logistic RSPF obtains a better fit to the data than the exponential RSPF. The Logistic model not only fits the data better but also provides the absolute probability of use rather than the relative probability of use. This underlines the importance of using RSPFs that are different from the exponential RSPF.

Resource Selection Probability Function and Weighted Distributions

We start with the most straightforward situation of the used–unused design. We assume that the use of the resource is nondestructive. Since the resource use is nondestructive, a particular location may potentially be visited repeatedly. Suppose we have a sample of size N from the study area. Suppose that, for each sample point, we know whether it was used or unused. Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ denote the vector of environmental covariates representing resources that may be used by animals. Let us denote the data by $(Y_i, \mathbf{X}_i)$, $i = 1, 2, \ldots, N$ where $Y_i = 1$ if the ith sample point is used, $Y_i = 0$ if that sample point is unused and $\mathbf{X}_i$ are the set of environmental covariates associated with that location. To study how the environmental covariates affect the probability of use, one models $P(Y = 1 \mid \mathbf{X} = \mathbf{x}) = \pi(\mathbf{x}, \beta)$ where $\pi(\mathbf{x}, \beta)$ is any function such that $0 \leq \pi(\mathbf{x}, \beta) \leq 1$ for all possible values of $\mathbf{x}$ and β. This probability function is called the resource selection probability function (RSPF). Under the used–unused design, the maximum likelihood estimation of the parameters β can be carried out by maximizing the log-likelihood function: $\Sigma_{i=1}^{N} \{Y_i \log \pi(x_i, \beta) + (1 - Y_i)\log(1 - \pi(x_i, \beta)\}$. Under the use-available study design, the situation is somewhat different. In this case, we know only those sample points where $Y_i = 1$. The goal of the analysis, however, remains the same: to estimate the resource selection probability function $\pi(\mathbf{x}, \beta)$ or equivalently to estimate the parameters β. To do this, we need to make an additional assumption. We assume that the covariate vectors $\mathbf{X}_i$ are a random sample from some multivariate distribution $f^{\mathrm{A}}(\mathbf{x})$. Provided this assumption is reasonable, standard probability argument leads to the following result:

$$f^{\mathrm{U}}(\mathbf{X} = \mathbf{x} | Y = 1; \beta) = \frac{P(Y = 1 | \mathbf{X} = \mathbf{x}; \beta) f^{\mathrm{A}}(\mathbf{x})}{P(Y = 1)}$$

$$= \frac{\pi(\mathbf{x}, \beta) f^{\mathrm{A}}(\mathbf{x})}{\int \pi(\mathbf{x}, \beta) f^{\mathrm{A}}(\mathbf{x}) d\mathbf{x}} = \frac{\pi(\mathbf{x}, \beta) f^{\mathrm{A}}(\mathbf{x})}{P(\beta)}.$$

This distribution is known as the weighted distribution

(Patil and Rao 1978). When use-available study design is implemented, we can use the likelihood based on this weighted distribution to estimate the parameters β. Before discussing the estimation procedure, it is important to establish the identifiability of the parameters. Following the results in Gilbert et al. (1999), all parameters in the RSPF $\pi(\mathbf{x}, \beta)$ are identifiable if for $\beta \neq \theta$, $[\pi(\mathbf{x}, \beta)]/[\pi(\mathbf{x}, \theta)] \neq K$ for all values of **x** and any constant K. That is, no two RSPFs are exactly proportional to each other. In precise mathematical terms, if $\beta \neq \theta$, then $\max_x |\pi(\mathbf{x}, \beta) - K\pi(\mathbf{x}, \theta)| > 0$ for all $K > 0$.

Most standard forms of RSPFs such as the logistic, probit, and the log-log link (Eqs. 5.1, 5.2, and 5.3 in Manly et al. 2002) satisfy the identifiability condition as long as not all covariates are categorical. Hence if one uses any of these models, one can estimate absolute probability of use. However, there are two common situations where the second condition is not satisfied and only relative probabilities are estimable. The first case occurs when all covariates are categorical. In this case, similar to the Logistic regression case, one of the categories is considered as a reference category (Hosmer and Lemeshow 1989:48). The relative probability of selection, relative to the reference category is estimable, but absolute probability of selection is not. Similarly, for the exponential RSPF, $\exp(\beta_0 + \beta_1 X)$ where $\beta_0 + \beta_1 X < 0$ for all parameter values and for all covariates (Eq. 5.9 in Manly et al. 2002), only the relative probabilities, relative to some reference location, are estimable but absolute probabilities are not (for a detailed proof of non-identifiability in these two cases, see the Appendix). The limitation of being able to estimate only relative probabilities, along with the constraints on the permissible values of the parameters, may limit the usefulness of the exponential RSPF (Keating and Cherry 2004). However, logistic, probit, and log–log RSPF put no such constraints and allow estimation of absolute probability of selection.

We have now established that when data are collected under the "use-available" study design (sampling protocol A, design I described on page 15 of Manly et al. [2002]), the distribution of covariates of the "used" sites $f^{\mathrm{U}}(\mathbf{x}; \beta)$ can be written as a weighted distribution (Patil and Rao 1978, Johnson et al. 2006):

$$f^{\mathrm{U}}(\mathbf{x}; \beta) = \frac{\pi(\mathbf{x}; \beta) f^{\mathrm{A}}(\mathbf{x})}{P(\beta)}$$

where $P(\beta) = E[\pi(\mathbf{X}; \beta)] = \int \pi(\mathbf{x}; \beta) f^{\mathrm{A}}(\mathbf{x})\, d\mathbf{x}$ and $f^{\mathrm{A}}(\mathbf{x})$ denotes the distribution of the covariates in the available population.

We are interested in estimating parameters β in the function $\pi(\mathbf{x}; \beta)$ given a random sample from the distribution $f^{\mathrm{U}}(x; \beta)$. This can be achieved using the method of simulated maximum likelihood (Robert and Casella 1999). The exact description of the method follows. In addition to the two conditions discussed earlier, we make assumptions A1–A8 and B1–B6 described on pages 12–14 of Manly et al. (2002).

Let $s_1, s_2, \ldots, s_n$ denote telemetry locations. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ denote environmental covariates at these locations. Based on these observations, the log-likelihood function can be written as: $l(\boldsymbol{\beta}; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \Sigma_{i=1}^{n} \{\log \pi(\mathbf{x}_i; \boldsymbol{\beta}) - \log P(\boldsymbol{\beta}) + \log f^{\mathrm{A}}(x_i)\}$. The probability density function $f^{\mathrm{A}}(\mathbf{x})$ is independent of the parameter β, and it can be ignored when maximizing the log-likelihood function. Furthermore, the probability density function $f^{\mathrm{A}}(\mathbf{x})$ is not known in an analytical form, hence the form of $P(\beta)$ is not known analytically as well. However, since we can sample observations from the availability distribution by randomly choosing points from the study area and observing the environmental variables at those sampled locations, we can obtain a Monte-Carlo estimate of $P(\beta)$ for any fixed value of β. Thus, one can obtain a Monte-Carlo estimate of the log-likelihood function (ignoring the terms independent of β) using

$$\hat{l}(\beta; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} \left\{ \log \pi(\mathbf{x}_i; \beta) - \log \left[\frac{1}{B} \sum_{j=1}^{B} \pi(\mathbf{x}_j^*; \beta) \right] \right\} \tag{1}$$

where $\mathbf{x}_j^*, j = 1, 2, \ldots, B$ is a simple random sample with replacement from the distribution $f^{\mathrm{A}}(\mathbf{x})$. Provided the value of B is large, $\geq$10 000 so as to ignore the Monte-Carlo error, one can apply standard numerical optimization techniques to maximize the function in Eq. 1 and obtain the maximum likelihood estimator of β. These estimators, being maximum likelihood estimators, are fully efficient, consistent, and asymptotically normal. The asymptotic standard errors and confidence intervals can be computed using the inverse of the matrix of the second derivatives. Such a matrix is readily available from numerical optimization routines. The usual methods of model selection such as the Akaike information criterion (AIC), its variants and likelihood ratio test of hypotheses are applicable without any modification.

Location- and Home-Range-Dependent Distribution of the Available Resources

So far, we have assumed that the distribution of available resources is identical for all animal locations. In practice, that assumption is not always tenable. Resources closer to used locations are more accessible than resources farther in space. Similarly, if individual animals have specific home ranges, the distribution of the available resources is different for each individual. A common practice is to define a buffer around the used location and assume only resources within this buffer or a specified home range are available (McClean et al. 1998 and references therein). This study design is claimed to be akin to the matched case-control design in epidemiology (Hosmer and Lemeshow 1989: Chapter 7, Compton et al. 2002) and conditional likelihood method is used for estimation. Hence, it is claimed that under the exponential RSPF, a standard logistic regression package can be used to analyze this type of data. However, the effects of contamination and overlap (Keating and Cherry 2004, Johnson et al. 2006) are likely to have even more serious consequences for this procedure. As pointed out by Johnson et al. (2006), the use-available study design is properly modeled as a weighted distribution and not as a case-control study. When approached from this perspective, the issue of contamination becomes irrelevant. The logistic discrimination method (Johnson et al. 2006) can be extended to deal with location and home range dependent distribution of the available resources. Such an extension leads to a stratified logistic regression and not a conditional logistic regression method (Appendix). In stratified logistic regression, the number of parameters increases at the same rate as the number of used locations resulting in biased and inconsistent estimators (Hosmer and Lemeshow 1989:Chapter 7). However, a simple adaptation of simulated likelihood facilitates the use of the non-exponential RSPFs. These are more flexible than the exponential RSPF and provide consistent estimators of absolute probabilities as opposed to relative probabilities.

Location-dependent availability.—Since the distribution of available resources is different for each used location, the weighted distribution for used locations varies with location. The weighted distribution corresponding to location s_i can be written as

$$f_i^{\mathrm{U}}(\mathbf{x}, s_i; \beta) = \frac{\pi(\mathbf{x}; \beta) f_i^{\mathrm{A}}(x; \mathbf{s}_i)}{P_i(\beta)}$$

where $P_i(\beta) = \int \pi(\mathbf{x}; \beta) f_i^{\mathrm{A}}(x; \mathbf{s}_i)\, d\mathbf{x}$. The notation $f_i^{\mathrm{A}}(x, s_i)$ denotes the distribution of resources available for location s_i. These may correspond to all resources within a certain distance of location s_i. The log-likelihood function can be written as

$$l(\boldsymbol{\beta}; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} [\log \pi(\mathbf{x}_i; \boldsymbol{\beta}) - \log P_i(\boldsymbol{\beta})].$$

Using the notation of the previous section, the corresponding simulated log-likelihood function is

$$\hat{l}(\beta; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} \left\{ \log\pi(\mathbf{x}_i; \beta) - \log\left[\frac{1}{B}\sum_{j=1}^{B} \pi(\mathbf{x}_j^*; \beta) w(d_{ij^*})\right]\right\}$$

where $w(d_{ij^*}) = 1$ if $d_{ij^*} < \delta$ and $w(d_{ij^*}) = 0$ if $d_{ij^*} \geq \delta$, for a circular buffer of radius δ. This function is maximized with respect to β to obtain the maximum likelihood estimators. The size of the buffer is usually fixed based on biological considerations such as the maximum travel distance within a certain amount of time.

Home-range-dependent availability.—Similar to the previous case, the application of the Logistic discriminant function approach (Johnson et al. 2006) leads to a stratified logistic regression model where each individual constitutes a stratum. The resultant estimators are generally biased and inefficient. However, the method of simulated likelihood provides consistent and efficient estimators.

Since the distribution of available resources is different from individual to individual, the weighted distribution for used locations also varies from individual to individual. The weighted distribution corresponding to the locations used by the ith individual can be written as

$$f_i^{\mathrm{U}}(\mathbf{x}; \beta) = \frac{\pi(\mathbf{x}; \beta) f_i^{\mathrm{A}}(\mathbf{x})}{P_i(\beta)}$$

where $P_i(\beta) = \int \pi(\mathbf{x}; \beta) f_i^{\mathrm{A}}(\underline{x})\; d\mathbf{x}$. The notation $f_i^{\mathrm{A}}(x)$ denotes distribution of resources available within the home range of the ith individual. Suppose there are I individuals in the sample. Let $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in_i})$ be the data corresponding to the ith individual. The log-likelihood function then can be written as

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} [\log \pi(\mathbf{x}_{ij}; \boldsymbol{\beta}) - \log P_i(\boldsymbol{\beta})].$$

Using the notation of the previous section, the corresponding simulated log-likelihood function is

$$\hat{l}(\beta) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left\{ \log \pi(\mathbf{x}_{ij}; \beta) - \log\left[\frac{1}{B}\sum_{k=1}^{B} \pi(\mathbf{x}_{ik}^*; \beta)\right]\right\}$$

where x_{ik}^*, $k = 1, 2, \ldots, B$ is a random sample from the distribution of the available resources within the home range of the ith individual. The maximum likelihood estimator of β is obtained by maximizing this function.

Statistical Properties: A Simulation Study

We now study the statistical properties of the simulated maximum likelihood estimator using simulations. In a simulation study, used as well as unused sites are known. Hence we can compare the performance of estimator based on the use-available study design using weighted distribution formulation with the estimator based on the "used–unused" study design. We consider the logistic RSPF:

$$\pi(\mathbf{x}; \beta) = \frac{\exp(\mathbf{x}\beta)}{1 + \exp(\mathbf{x}\beta)}$$

so that logistic regression will be an appropriate method for the used–unused data. However, note that the weighted distribution based method is not restricted to this functional form, one can choose any function that takes values in the range (0, 1) and satisfies the identifiability condition described earlier. Consider a hypothetical landscape, where corresponding to each location there are two environmental covariates. For simulation purposes, we assume that these covariates follow a bivariate Normal distribution with mean vector 0 and identity matrix as the covariance matrix. Let β_T denote the value of the regression parameters under which simulations are conducted. We assume that the

TABLE 1. Comparison of the logistic regression (used–unused design) and simulated maximum likelihood (use-only design) estimators.

	$N = 500$		$N = 1000$		$N = 2000$	
Statistical summary	Mean	SD	Mean	SD	Mean	SD
Logistic regression (used–unused)						
β_0	2.474	0.234	2.451	0.163	2.439	0.113
β_1	2.264	0.248	2.249	0.168	2.234	0.119
β_2	1.423	0.194	1.418	0.138	1.409	0.094
Simulated maximum likelihood (used only)						
β_0	2.716	1.292	2.551	0.6378	2.484	0.451
β_1	2.450	0.909	2.326	0.445	2.266	0.312
β_2	1.539	0.582	1.464	0.323	1.433	0.218

Notes: True values of the parameters are $\beta_0 = 2.434932$, $\beta_1 = 2.229461$, $\beta_2 = 1.407078$. Sample size N is the number of sites.

resources at a location can be used repeatedly, that is, the use is nondestructive. The steps in the simulations are:

1) Randomly select a location from the hypothetical landscape. Let $\mathbf{x}$ denote the habitat covariates corresponding to the selected location.

2) This location is used with probability

$$\pi(\mathbf{x}; \beta_T) = \frac{\exp(\mathbf{x}\beta_T)}{1 + \exp(\mathbf{x}\beta_T)}$$

and with probability

$$1 - \pi(\mathbf{x}; \beta_T) = \frac{1}{1 + \exp(\mathbf{x}\beta_T)}$$

it is not used. The used locations are indexed by 1 and the unused locations are indexed by 0.

3) Repeat steps 1 and 2, for N number of times. At the end of step 3, we have N locations with their covariates along with the index of whether they were used or unused.

4) Apply the logistic regression to estimate the parameters β based on the used–unused data.

To mimic the "use-available" study design, we now ignore information on all those locations that were unused. We only consider information on the used locations. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ denote the covariates at the "used" locations.

5) Select a simple random sample with replacement of size $B = 10\,000$ locations from the hypothetical landscape. Let the covariates at these random locations be denoted by $\underset{\sim}{x}_j^*$, $j = 1, 2, \ldots, 10\,000$.

6) Maximize

$$\hat{l}(\beta; \underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n) = \sum_{i=1}^{n} \left\{ \log \pi(\underset{\sim}{x}_i; \beta) - \log \left[\frac{1}{B} \sum_{j=1}^{B} \pi(\underset{\sim}{x}_i^*; \beta) \right] \right\}$$

with respect to β to obtain the simulated maximum likelihood estimator based on used-available study design.

7) Repeat steps 1–5, $S = 1000$ number of times.

To mimic small, medium, and large sample sizes, we considered $N = 500$, 1000, and 2000. In Table 1, we report the mean and the standard deviation of the sampling distribution of the estimators based on the two methods. As expected, when used–unused data are available, logistic regression method works well. More importantly, the simulated maximum likelihood method based on the weighted distribution and the used-available study design also works very well, providing estimators of all parameters that are nearly unbiased. The variance of these estimators is somewhat larger than the ones that are based on used–unused sampling design. This is to be expected given that the use-available study design contains less information than the used–unused study design. Notice also that as the sample size increases, the bias and the variance of the simulated maximum likelihood estimators decreases, illustrating the consistency property of these estimators.

EXAMPLE: ANALYSIS OF MOUNTAIN GOAT (*OREAMNOS AMERICANUS* DE BLAINEVILLE 1816) TELEMETRY DATA IN NORTHWEST BRITISH COLUMBIA

In this section, we provide an example analysis using the logistic RSPF. This example does not constitute a full-fledged ecological analysis of the data per se. Analysis detailing data exploration, graphical methods for model construction, effect of the consideration of home-range- and location-dependent available resources, and so on is provided elsewhere (S. R. Lele and J. L. Keim, *unpublished manuscript*). The purpose of this analysis is to show that our approach provides sensible answers for these data. Furthermore, this example illustrates that a non-exponential RSPF may provide a better fit to the data than the commonly used exponential RSF, underlining the importance of the methodological extension described in this paper.

Mountain goat telemetry data were collected in the Coast Mountains of northwest British Columbia, Canada (58°48′–59°12′ N, 133°18′–133°48′ W). The study area is located approximately 60–100 km east of the Pacific Ocean at the Alaska panhandle. In general, the area contains a transition of environmental conditions between a drier colder climate found in the interior of northwestern British Columbia and a warmer more

Table 2. Log-likelihood values for various models.

Model	Log-likelihood
exp(HEAT)	4212.84
sin(slope) + sin(slope)^2	5069.99
ET	7419.25
sin(slope) + sin(slope)^2 + ET	7428.85
exp(HEAT) + ET	9166.84

Notes: A model with a larger log-likelihood value is considered to provide a better fit. Key to variables: ET, access to escape terrain; HEAT, heat load index.

humid climate found in coastal southeastern Alaska. Topographic relief in the area is variable; ranging in elevation from 300 to 2500 m above sea level. In general, alpine tundra occurs above 1400 m above sea level. Winter temperatures can drop to −40°C with snow accumulations often reaching depths greater than 5 m in some areas. Estimates of mountain goat density of 0.45 mountain goats/km^2 have been reported within areas of this geographic range.

We used 6337 animal locations collected by global positioning system (GPS) radio collars (Lotek GPS 2000 model collars; Lotek Wireless, Newmarket, Ontario, Canada) from 10 mountain goats (seven females and three males). We used winter (defined as 1 January to 20 April of the calendar year) data from 2000, 2001, and 2002. The GPS collars were programmed to attempt acquisition of GPS locations six times per day at four-hour intervals (fix rate of 81%).

In this analysis, we considered only topographical covariates: elevation, slope, heat load index (HEAT), and access to escape terrain (ET; see Plate 1). The heat load index (McCune and Keon 2002) is a function of the latitude, longitude, slope and aspect of a given location. For mountain goats, access to steeper terrain is critical to avoid predation. We define access to escape terrain as the distance to the nearest 45°–60° slope. The log-likelihood values for different models are provided in Table 2. These indicate that ET is the most important topographical covariate and that heat load index is the next best covariate. Other covariates such as elevation and slope were statistically insignificant. The final model we use is

$$\pi(X;\beta) = \frac{\exp[\beta_0 + \beta_1 \exp(\mathrm{HEAT}) + \beta_2 \mathrm{ET}]}{1 + \exp[\beta_0 + \beta_1 \exp(\mathrm{HEAT}) + \beta_2 \mathrm{ET}]}.$$

The parameter estimates along with their standard errors are provided in Table 3. These standard errors were obtained from the inverse of the estimated Fisher Information (Hessian) matrix. All covariates are significantly different from zero. This analysis shows that mountain goats prefer habitats with a high heat load index and in close proximity to escape terrain. To provide a comparison to the standard method of analysis, we also fit the exponential RSF model to this data. As discussed earlier, the intercept parameter of the exponential RSPF is non-identifiable. The Bayesian information criterion (BIC) value (Burnham and Anderson 2002) for the fitted logistic RSPF is −18 307.43, whereas for the exponential RSF (with the same covariates), it is −18 170.11. This indicates that the logistic RSPF model is a better descriptor of the data than the exponential RSPF model. In addition, the logistic RSPF model provides absolute probabilities of use whereas the exponential RSF model provides information only on relative probability of use.

Discussion

Earlier, we defined the concept of RSPF in the context of used–unused study design. The concept of weighted distribution was proposed as a solution for estimation of the RSPF when data on used locations only are available. The explanation of RSPF in the framework of used–unused data and binary regression, although commonly proposed in the literature (Manly et al. 2002), is somewhat confusing. The nature of the telemetry data is such that we never know which locations were visited but not used. Any location that is visited is implicitly assumed to have been used. Thus, given enough time, eventually most of the study area gets visited and one could, albeit incorrectly, infer that probability of use is 1 for every type of habitat. However, it is clear that if a particular type of habitat is visited more often than some other habitat, we should say that such a habitat is used preferentially. There is an alternative explanation of the concept of RSPF that reflects this thinking directly. This explanation is closely related to the concept of response dependent probability sampling in the theory of survey sampling (Godambe 2002) where sampling units, depending on their characteristics, have differential probability of being reported in the sample. In the survey sampling situation, if the sampling is done with replacement, the observed sample follows a weighted distribution with weights corresponding to the reporting probabilities (Godambe 2002). The derivation of the weighted distribution in the context of RSPF as described in our paper is identical to the one provided in the survey sampling context in Godambe (2002). In the RSPF context, one can imagine an animal as a sampler who is selecting samples from the population of available resources. A unit is used, or in other words, reported in the sample, with some unknown probability $\pi(\underline{x};\ \beta)$. In the survey sampling context, the reporting probability is completely known and based on the observed sample, one wants to infer about the unknown

Table 3. Estimates and standard errors for the parameters in the resource selection probability function model.

Parameters	Estimated value	SE
Intercept	−4.990	0.086
exp(HEAT)	2.166	0.064
ET	−0.019	0.0004

Note: The resource selection probability function model, with HEAT as the heat load index and ET as the access to escape terrain, is $\pi(X;\beta) = \{\exp[\beta_0 + \beta_1\exp(\mathrm{HEAT}) + \beta_2\mathrm{ET}]\}/\{1 + \exp[\beta_0 + \beta_1\exp(\mathrm{HEAT}) + \beta_2\mathrm{ET}]\}$.

PLATE 1. Photograph of mountain goats in escape terrain (northwest British Columbia, February 2006). Photo credit: J. L. Keim.

population. In the context of RSPF, the reporting probability is unknown and based on the observed sample and the known population distribution $f^{A}(\underline{x})$, we want to infer about the unknown reporting probability. The concept of "available units" in RSPF context corresponds to the concept of "population" in the sampling theory; the "used units" correspond to the observed sample in the sampling theory and "resource selection probability" corresponds to the "reporting probability." The methodology developed in this paper shows when and how the reporting probabilities can be estimated using the observed sample and the known population. In this paper we assume that sampling is with replacement and that reporting probability is a function of the location specific habitat characteristics. The analogy between response-dependent probability sampling and RSPF estimation also helps indicate how one might include spatial correlation in the analysis of telemetry data. We simply need to infer about the spatial sampling plan using the known population and the observed sample. One can use mixed models to extend this methodology to account for individual variation and herd effect. It is also important to take into account measurement error in the GPS location data (Friar et al. 2003). Such extensions will be discussed elsewhere.

Spatial extent of the pixel or the spatial unit is another important issue. Change in the pixel size affects the distribution of the available resources. In the terminology of sampling theory, changing the pixel size changes what is considered as the population from which the sample is drawn. Similar to the sampling theory, in RSPF inferences are conditional on what are considered the available units. The problem of the modifiable area unit problem (MAUP) is as relevant to RSPF as it is to spatial statistics. It relates to the issue of what is the right scale to study a particular phenomenon. We are not aware of any general solution to this problem.

Relative probability maps based on estimated RSFs are routinely used in applied ecology and wildlife management (Manly et al. 2002, McDonald and McDonald 2002, Johnson et al. 2005). However, absolute probabilities of use are more desirable than the relative probabilities (Keating and Cherry 2004; M. Boyce, *personal communication*) and are a more powerful tool for managers. In heavily altered habitats, knowledge of absolute probabilities is critical as some of the "relatively good" habitats according to RSF may actually be not so desirable in terms of absolute probabilities (RSPF).

The methodology developed in this paper provides applied ecologists with a tool to estimate the absolute probability of use under the use-available study design. Furthermore, it removes the restriction of exclusively using the exponential RSPF and expands the applicable class of models that includes models such as logistic, log–log, and probit link among others. This increases the usefulness and applicability of telemetry data in scientific understanding and decision making.

The example data and the computer code written in R (R Development Core Team 2005) used for the data analysis is available in the Appendix. A user-friendly version of the code that facilitates use of link functions other than the logistic link, model selection, and mixed-model inference is under preparation.

Acknowledgments

We are grateful to Chris Jerde, Jacqui Frair, Sophie Czetwertynski, Mark Boyce, Chris Johnson, Trent MacDonald, and Scott Nielson for useful discussions. The British Columbia Ministry of Environment provided the data. We thank Karen Diemert, Brian Fuhr, and Len Vanderstar for their support. We are also grateful for many useful suggestions made by the referees.

Literature Cited

Boyce, M., and L. McDonald. 1999. Relating populations to habitats using resource selection function. Trends in Ecology and Evolution **14**:268–272.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and inference: a practical information theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

Compton, B. W., J. M. Rhymer, and M. McCollough. 2002. Habitat selection by wood turtles (*Celmmys insculpta*): an application of paired logistic regression. Ecology **83**:833–843.

Friar, J., S. Nielson, E. Merrill, S. R. Lele, M. Boyce, R. Munroe, G. Stenhouse, and H. Beyer. 2003. Removing GPS collar bias in habitat selection studies. Journal of Applied Ecology **41**:201–212.

Gilbert, P., S. Lele, and Y. Vardi. 1999. Semiparametric inference for selection bias models with application to AIDS vaccine trials. Biometrika **86**:27–43.

Godambe, V. P. 2002. Utilizing survey framework in scientific investigations. Sankhya, Series A **64**:268–281.

Hosmer, D. W., and S. Lemeshow. 1989. Applied logistic regression. John Wiley and Sons, New York, New York, USA.

Huzurbazar, S. V., editor. 2003. Resource selection methods and applications. Omnipress, Madison, Wisconsin, USA.

Johnson, C. J., M. S. Boyce, R. L. Case, H. D. Cluff, F. J. Gau, A. Gunn, and R. Mulders. 2005. Cumulative effects of human developments on arctic wildlife. Wildlife Monographs 160.

Johnson, C. J., S. E. Nielsen, E. H. Merrill, T. L. McDonald, and M. S. Boyce. 2006. Resource selection functions based on use-availability data: theoretical motivation and evaluation method. Journal of Wildlife Management **70**(20):347–357.

Keating, K. A., and S. Cherry. 2004. Use and interpretation of logistic regression in habitat selection studies. Journal of Wildlife Management **68**:774–789.

Lancaster, T., and G. Imbens. 1996. Case-control studies with contaminated controls. Journal of Econometrics **71**:145–160.

Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. McDonald, and W. P. Erickson. 2002. Resource selection by animals: statistical analysis and design for field studies. Second edition. Kluwer Press, Boston, Massachusetts, USA.

McClean, S. A., M. A. Rumble, R. M. King, and W. L. Baker. 1998. Evaluation of resource selection methods with different definitions of availability. Journal of Wildlife Management **62**:793–801.

McCune, B., and D. Keon. 2002. Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science **13**:603–606.

McDonald, T. L., and L. L. McDonald. 2002. A new ecological risk assessment procedure using resource selection models and geographic information systems. Wildlife Society Bulletin **30**:1015–1021.

Patil, G. P., and C. R. Rao. 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics **34**:179–189.

Robert, C. P., and G. Casella. 1999. Monte Carlo statistical methods. Springer-Verlag, New York, New York, USA.

R Development Core Team. 2005. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Seber, G. A. F. 1984. Multivariate observations. John Wiley and Sons, New York, New York, USA.

APPENDIX

An alternative derivation of Johnson et al. (2006) method and proof of nonidentifiability of the intercept parameter for categorical covariates (*Ecological Archives* E087-181-A1).

SUPPLEMENT

Source code in R for estimating logistic resource selection probability function and the data set used in the paper (*Ecological Archives* E087-181-S1).