

Application of random effects to the study of resource selection by animals

CAMERON S. GILLIES*‡, MARK HEBBLEWHITE*, SCOTT E. NIELSEN*,
MEG A. KRAWCHUK†, CAMERON L. ALDRIDGE*§, JACQUELINE L.
FRAIR*, D. JOANNE SAHER†, CAMERON E. STEVENS* and
CHRISTOPHER L. JERDE*

*Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9, Canada; and

†Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada T6G 2H1, Canada

Summary

1. Resource selection estimated by logistic regression is used increasingly in studies to identify critical resources for animal populations and to predict species occurrence.

2. Most frequently, individual animals are monitored and pooled to estimate population-level effects without regard to group or individual-level variation. Pooling assumes that both observations and their errors are independent, and resource selection is constant given individual variation in resource availability.

3. Although researchers have identified ways to minimize autocorrelation, variation between individuals caused by differences in selection or available resources, including functional responses in resource selection, have not been well addressed.

4. Here we review random-effects models and their application to resource selection modelling to overcome these common limitations. We present a simple case study of an analysis of resource selection by grizzly bears in the foothills of the Canadian Rocky Mountains with and without random effects.

5. Both categorical and continuous variables in the grizzly bear model differed in interpretation, both in statistical significance and coefficient sign, depending on how a random effect was included. We used a simulation approach to clarify the application of random effects under three common situations for telemetry studies: (a) discrepancies in sample sizes among individuals; (b) differences among individuals in selection where availability is constant; and (c) differences in availability with and without a functional response in resource selection.

6. We found that random intercepts accounted for unbalanced sample designs, and models with random intercepts and coefficients improved model fit given the variation in selection among individuals and functional responses in selection. Our empirical example and simulations demonstrate how including random effects in resource selection models can aid interpretation and address difficult assumptions limiting their generality. This approach will allow researchers to appropriately estimate marginal (population) and conditional (individual) responses, and account for complex grouping, unbalanced sample designs and autocorrelation.

Key-words: functional response, grizzly bear, habitat selection, random effects, resource selection function.

Journal of Animal Ecology (2006) **75**, 887–898

doi: 10.1111/j.1365-2656.2006.01106.x

Correspondence: Cameron S. Gillies, PO Box 750, Windermere, BC V0B 2L0, Canada. Tel: 1 250 342 8640; Fax: 1 250 342 8644; E-mail: cam.gillies@telus.net

‡The order of authors following S. Nielsen was drawn at random.

§Current address: US Geological Survey, 2150 Centre Avenue, Building C, Fort Collins, CO 80526–8118, USA.

Introduction

Resource selection by animals is an important determinant of fitness and is a focus of many ecological studies (Franklin *et al.* 2000). A common approach for examining species occurrence and habitat selection in the

ecological literature is resource selection functions (RSF; Manly *et al.* 2002). RSF models are attractive to ecologists because they provide quantitative, spatially explicit, predictive models for animal occurrence (e.g. Mladenoff *et al.* 1995; Johnson, Seip & Boyce 2004). RSF models are commonly developed by comparing habitat characteristics at sites that were used by animals to those that were potentially available (RSF; Manly *et al.* 2002). Model coefficients are estimated using logistic regression, which assumes independence among observations (Hosmer & Lemeshow 2000). While independence is feasible in some RSF designs, recent reviews emphasize that most studies fail to satisfy this assumption (Morrison 2001). Autocorrelation among observations produces incorrect variance estimates (Otis & White 1999) and an increased Type I error rate (Leban *et al.* 2001). To avoid pseudoreplication (Hurlbert 1984), researchers often rarify data to achieve independence (Swihart & Slade 1985), resulting in an unfortunate loss of information (McNay & Bunnell 1994).

There have been two general solutions for non-independence among observations in resource selection studies. The first is compositional analysis (Aebischer, Robertson & Kenward 1993), in which individual animals are identified as the unit of replication. Unfortunately, compositional analysis is limited by increased Type I error rates from rare habitats (Bingham & Brennan 2004). In addition, it cannot accommodate continuous covariates or interaction terms when comparing among individuals, nor Poisson, binomial or other dependent variable structures. A second solution is the Huber–White sandwich variance estimator, which can be used to calculate robust standard errors without affecting coefficient estimates (White 1982; Newey & West 1987; Pendergast *et al.* 1996). However, because unbalanced numbers of locations among individuals are common in telemetry studies, coefficients will be biased toward the most sampled individuals (Follmann & Lambert 1989). Therefore, in the presence of an unbalanced design, variance inflators provide only a partial solution to non-independence.

Mysterud & Ims (1998) discuss an additional difficulty in studies of resource selection that has yet to be addressed comprehensively. They demonstrated how use of a resource might differ contingent upon the availability of that resource, which they define as a functional response in resource selection. If animals require a particular amount of a given resource, they may show strong selection for it when scarce but avoid it when it is abundant. Although Mysterud & Ims (1998) criticized the assumption that selection is independent of availability, a flexible treatment of functional responses has not been attempted.

The dual problems of non-independence and functional responses in resource selection can be addressed through the application of random effects to RSF models. Random effects are applied widely in cohort, survival and other hierarchical designs where individuals or groups are sampled repeatedly (e.g. Natarajan & McCulloch 1999; Burnham & White 2002; Krawchuk & Taylor 2003).

Random effects can accommodate non-independence within groups, such as samples within individuals or individuals within populations (Breslow & Clayton 1993). Although Aebischer *et al.* (1993) first suggested using random effects in resource selection studies, few have incorporated random effects into resource selection or species distribution models in general (see reviews in Rushton, Ormerod, & Kerby 2004; Guisan & Thuiller 2005). Recent developments of generalized linear mixed models extend random-effect designs to binomial responses (Breslow & Clayton 1993; Skrandal & Rabe-Hesketh 2004) and, thus, to modelling resource selection.

In this paper, we first provide a brief overview of random-effects models and introduce their application to resource-selection modelling. We then illustrate the application of random-effects models to a case study of grizzly bear (*Ursus arctos* L.) resource selection in the Canadian Rocky Mountain Foothills (Nielsen *et al.* 2002). We consider grizzly bear resource selection for simple categorical and continuous covariates, and compare fixed-effects (without random effects) RSF models to those with random effects for the intercept, categorical and continuous variables. To aid in our interpretation of random effects in this empirical example, we simulated data for three common scenarios where random effects are included in RSF models: (1) balanced vs. unbalanced samples; (2) differences in selection among individuals for a continuous or categorical covariate where availability is constant; and (3) availability varying among individuals and selection is either constant or follows a functional response. We conclude with a discussion of how the inclusion of random effects can control for common limitations in resource selection studies and yield more robust ecological insights.

A BRIEF OVERVIEW OF RANDOM EFFECTS

Following from their first exposition in ANOVA-type models (e.g. Bennington & Thayne 1994), a variable is considered random when the investigator has not controlled explicitly for levels of the variable in the experimental design, but has chosen a random sample of levels from the population (Neter *et al.* 1996). An example would be individual red deer (*Cervus elaphus* L.) within a population where levels of individual variation (e.g. age) were not fixed but assumed to be representative of the population. By including a random effect for individuals, individual variability is identified explicitly and the scope of inference can be extended to the entire population (Neter *et al.* 1996).

In addition to providing valid population-level inferences, random effects are often invoked to control for correlations among samples. For example, a particular response variable (e.g. telemetry locations) may be correlated within particular strata; for example, within a group (individual deer) or hierarchical association (deer within herds). This unobserved heterogeneity within levels could produce pseudoreplicated samples (Hurlbert 1984) that lack independence, even after controlling for

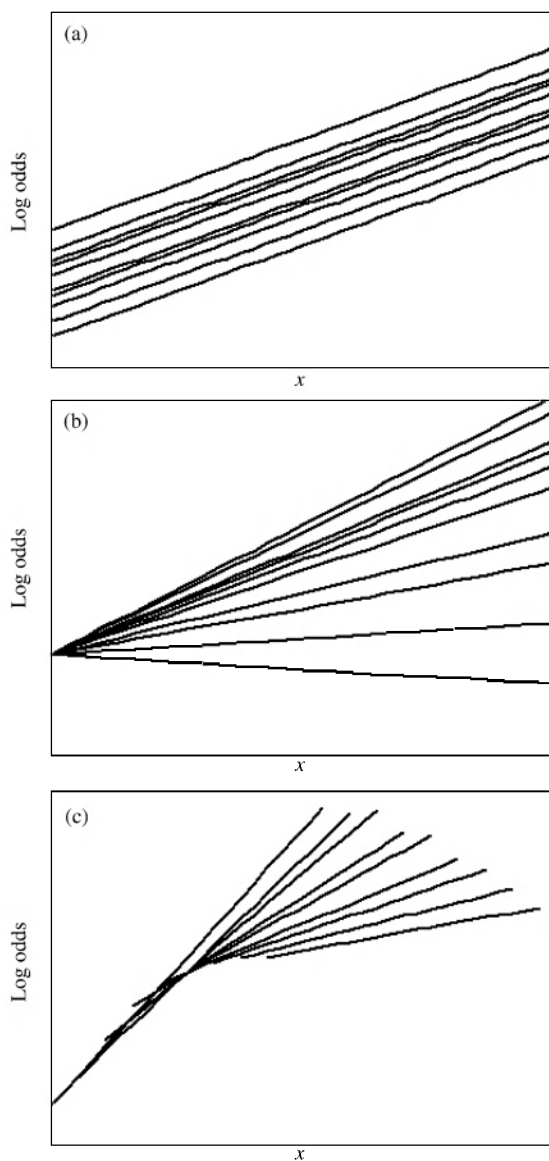


Fig. 1. Conceptual plot of the use of a resource unit along a gradient of a continuous covariate x for individuals having random intercepts (a), random coefficients (b), or a functional response to the availability of x (c).

the fixed effects of covariates (Skrondal & Rabe-Hesketh 2004). Parameter estimates from such fixed-effects models will often be biased (Skrondal & Rabe-Hesketh 2004). An added benefit of random-effect models is to allow group-level specific estimates for a response, known as the conditional estimate. In comparison, the overall model estimate is known as the marginal, or population-level estimator for a particular response variable (Breslow & Clayton 1993; Begg & Parides 2003; Skrondal & Rabe-Hesketh 2004).

In addition to accounting for within-strata variation, random effects can be used to control for unbalanced designs in the number of observations among individuals or groups (Bennington & Thayne 1994). Without a random intercept for individuals with unbalanced data, sample size differences may influence model coefficients. By accounting for these relationships among samples,

including correlation or sampling design-related issues, random effects provide more robust ecological inferences (Pendergast *et al.* 1996).

Random effects can be added to fixed-effects regression models, including RSF models, in two ways. Random intercepts allow the intercept or magnitude of the response to vary among groups (Fig. 1a), whereas the inclusion of random coefficients allows the effect of covariates to vary among groups (Fig. 1b) (Begg & Parides 2003; Skrondal & Rabe-Hesketh 2004). In RSF models, random intercepts influence overall prevalence which, as we illustrate below, often arises because of unbalanced samples (Fig. 1a). Random coefficients can be included when there is variation in individual animal, group, etc. responses to a particular covariate (Fig. 1b). Random-effects models can easily accommodate two or more levels, e.g. samples from individual deer within herds within populations, or wolves (*Canis lupus* L.) within packs. When a model contains both random- and fixed-effects, it is termed a mixed-effect model. Functional responses in selection might be accommodated through the combination of a random intercept and random coefficient (Fig. 1c).

Assumptions of random-effects models include (1) correlations within groups are constant over time unless modelled explicitly; (2) the random effects are normally distributed with a zero mean and unknown variance components; and (3) the variance-covariance structure is specified correctly (Breslow & Clayton 1993; Skrondal & Rabe-Hesketh 2004). The most common structure is compound symmetric, which considers covariance among all responses of an individual to be constant (Skrondal & Rabe-Hesketh 2004). For time-series data, an autoregressive structure could be useful (Pinheiro & Bates 2000). More complex structures could include average, lagged, factor, unrestricted and hybrid correlation structures that are beyond our purview (see Pinheiro & Bates 2000 for more detailed information).

Materials and methods

INCLUDING RANDOM EFFECTS IN RSF MODELS

Following Manly *et al.* (2002: 100), we use a typical fixed-effects exponential RSF:

$$\hat{w}(x) = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_n x_n) \quad \text{eqn 1}$$

with covariates x_n , and $\hat{\beta}_n$ are the coefficients (parameters) estimated from logistic regression (Manly *et al.* 2002). Commonly, the intercept $\hat{\beta}_0$ is dropped from the RSF formulation as discussed and justified by Manly *et al.* (2002); however, because we will be using random intercepts, we include $\hat{\beta}_0$ in the expression for $\hat{w}(x)$.

Coefficients for the random-intercept and random-effect RSF model are estimated using logistic regression by a generalized linear mixed-effects logit model (Skrondal & Rabe-Hesketh 2004). The conditional mean

of Y given x , $\pi(x)$ follows the standard logistic regression notation presented, discussed and reviewed by Hosmer & Lemeshow (2000). In our example, we consider a two-level random-effect model, where observations $i = 1 \dots n$ are clustered within strata $j = 1 \dots m$; for example, locations within individuals. For a random-intercept model, the logit model, $g(x)$, is estimated for location i for grizzly bear j :

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_n x_{nij} + \gamma_{0j}, \quad \text{eqn 2}$$

where x_n are covariates with fixed regression coefficients β_n , β_0 is the mean intercept, and γ_{0j} is the random intercept, which is the difference between the mean intercept β_0 for all groups and the intercept for group j (Skronal & Rabe-Hesketh 2004: 51–54). Note that γ_{0j} is the random effect in eqn 2 and all preceding terms represent the normal fixed effects as in eqn 1. Here and throughout, we assume that random effects are normally distributed, as is common in mixed-effects modelling (Hosmer & Lemeshow 2000; Skronal & Rabe-Hesketh 2004). However, this assumption should be investigated using exploratory data analysis and plots of the residuals.

For the model with a random intercept and a random coefficient, the RSF coefficients are estimated following:

$$g(x) = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_n x_{nij} + \gamma_{ij} x_{ij} + \gamma_{0j} \quad \text{eqn 3}$$

where notation follows from eqn 2 with the addition of $\gamma_{ij} x_{ij}$ where γ_{ij} is the random coefficient of covariate x_n for group j . Models with a random coefficient include a random intercept because a random coefficient forces variation in the intercept (Skronal & Rabe-Hesketh 2004).

Recent advances in maximum likelihood theory have made implementing random effects in generalized linear models easier in many statistical packages. For STATA, the standard function is GLLAMM, reviewed by Skronal & Rabe-Hesketh (2004), available at <http://www.gllamm.org>. For SAS, the standard procedure is GLIMMIX, from <http://support.sas.com/tnd/app/papers/glimmix.pdf>. For S-PLUS AND R, standard functions include glme and glmmPQL, and glmmML, glmm, respectively (Pinheiro & Bates 2000).

APPLICATION OF RANDOM EFFECTS TO GRIZZLY BEAR RESOURCE SELECTION

Grizzly (brown) bears are a species of conservation concern across the circumpolar north, and as a result their resource selection patterns have frequently been the subject of applied research (e.g. McLellan & Hovey 2001; Nielsen *et al.* 2002). To explore how random effects can influence RSF models, we reanalysed a grizzly bear global positioning system (GPS) radiotelemetry data set from Nielsen *et al.* (2002) in the Eastern slopes of Alberta's Canadian Rocky Mountain Foothills. To minimize complications in seasonal variation in habitat use we focus on only the late summer and autumn period (1 August to

Table 1. Number of GPS telemetry locations per grizzly bear used in random effect resource selection function (RSF) models

Bear ID	No. of GPS locations	Sex
G2	493	F
G3	227	F
G4	388	F
G5	98	M
G6	92	M
G8	89	M
G10	149	F
G16	441	F
G20	494	F
Total	2471	

denning). In total, 2471 use locations from three adult male and six adult female bears during 1999 were used from a 5332 km² study area. Samples were unbalanced, varying from 89 to 494 observations per individual (Table 1). Availability was defined for each animal by drawing 1000 random locations from 100% minimum convex polygon (MCP) home ranges (ranging in size from 383 to 1588 km²), thus the measure of availability was unique to each animal. As such, our analysis corresponded to analysing resource selection at the third-order scale of selection (Johnson 1980). For each used and available location, two environmental variables were queried from a geographic information system (GIS): open habitats [a categorical landcover variable from Franklin *et al.* (2001) identifying the location as either open = 1 or forested = 0] and elevation (in 100 m units). A more detailed description of the study design, data and study area can be found in Nielsen *et al.* (2002).

We estimated grizzly bear RSF models using four approaches. We first used fixed-effects logistic regression to estimate the coefficients of the RSF in eqn 1, which we refer to as the naive RSF model. Secondly, we evaluated a common method used to account for autocorrelation within individuals, namely, by employing the Huber–White sandwich variance estimator (*sensu* Nielsen *et al.* 2002) within a fixed-effects logit model. Finally, we compare these two models to the RSF models derived from a random-intercept model (eqn 2), and models with a random intercept and random coefficient (eqn 3) for either open habitat or elevation.

Random-effect models were estimated using the GLLAMM procedure with adaptive quadrature (Rabe-Hesketh, Pickles & Skronal 2001; Skronal & Rabe-Hesketh 2004) in STATA version 8.2 (StataCorp 2003) and a compound symmetric covariance structure, which assumes that all samples within a group are, on average, equally correlated (Skronal & Rabe-Hesketh 2004). Conditional coefficient estimates for each individual were produced using the GLLAPRED procedure (Rabe-Hesketh *et al.* 2001). Model selection for models with random effects is complicated because the intended scope of inference, conditional or marginal, influences the derivation of information theoretic metrics such as the consistent Aikake information criterion (cAIC)

developed specifically for application to such models (Burnham & Anderson 2002; Burnham & White 2002). See Vaida & Blanchard (2005) for details of model selection with random effects; herein, we do not consider model selection for random effects further. We focus instead on evaluating changes to model fit based on log-likelihoods, $\log(\mathcal{L})$, marginal coefficient estimates, their standard errors (SE) and the variance of the random effects.

Understanding random effects in resource selection studies: simulated examples

To provide insight into interpreting RSF models with random effects, we simulated data under three common sampling designs. We designed our simulation following the grizzly bear data, generating used and available points, and estimated the coefficients for RSF models following eqns 1–3 above. Due to the computational time required to solve mixed models using conventional software, our study was a demonstration using a single simulation for each of the scenarios considered, not a statistical simulation study with 1000s of iterations to reveal inferential bounds of random effects in RSF models (*sensu* Burnham & White 2002).

Simulating use–availability data

Using STATA version 8.2 (StataCorp 2003), we simulated data with a logit function of the form $\pi(x) = e^{g(x)} / (1 + e^{g(x)})$ because it allowed us to generate used (1) and unused (0) points, based on the simulation selection function $g(x)$. We retained only simulated use (1) points and generated an independent random sample of available points. The linear function, $g(x)$, of the parameters is provided for each example discussed below. Our set of covariates (fixed effects) included one standardized continuous variable, elevation (x_1) and one categorical variable, open habitat (x_2). Unless otherwise noted, all elevations were standardized to be uniformly available over a range of 0–2 for x_1 , and the two categories of x_2 , open and closed canopy, were equally prevalent. For each analysis, we randomly selected 500 available points for each individual from its range of available elevations and from the available habitat types. We simulated population-level resource selection producing a distribution of used points that selected higher elevations (higher values of x_1) and selected open habitat, with 61% of used points being in open habitat. A copy of our simulation and analysis code for STATA version 8.2 is available from the senior author, and our simulations were verified independently (M. Taper, personal communication, Montana State University).

EXAMPLE 1: FIXED EFFECTS FOR BALANCED AND UNBALANCED DESIGNS

$$\text{Model 1: } g(x) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$$

β_0 is the intercept, β_1 and β_2 are the coefficients on the variables x_1 and x_2 , respectively (i designates the

observation while j designates the group). In all three examples $\beta_0 = -0.5$, $\beta_1 = 1$, and $\beta_2 = 1$. In this example, selection was invariant across individuals for both elevation and open habitat, and animals had the same availability. For our balanced design we observed 20 individuals ($j = 1 \dots 20$) with 100 observations each ($i = 1 \dots 2000$). For the unbalanced design, the number of observations per individual was drawn from a normal distribution ($\mu = 100, \sigma = 40$). No random effect was used in the generation of these example data.

EXAMPLE 2: DIFFERENCES IN SELECTION AMONG INDIVIDUALS USING A RANDOM EFFECT

$$\text{Model 2a: } g(x) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \gamma_1 x_{1ij} + \gamma_{0j}$$

$$\text{Model 2b: } g(x) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \gamma_2 x_{2ij} + \gamma_{0j}$$

For model 2a, γ_1 was drawn from a normal distribution ($\mu = 0, \sigma = 2$) for each individual j , while γ_2 for model 2b was drawn from a normal distribution ($\mu = 0, \sigma = 1$) for each individual j . The gamma (γ) terms are random effects that add differences in selection among individual animals (as in eqns 2, 3). We considered differences in selection among animals for either elevation or open habitat across the same range of availability, with balanced samples among individuals (Fig. 1b).

EXAMPLE 3: DIFFERENCES IN AVAILABILITY AND FUNCTIONAL RESPONSES AMONG INDIVIDUALS

$$\text{Model 3a: } g(x) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$$

$$\text{Model 3b: } g(x) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \gamma_1 x_{1ij} + \gamma_{0j}$$

$$\text{Model 3c: } g(x) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij}$$

$$\text{Model 3d: } g(x) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \gamma_2 x_{2ij} + \gamma_{0j}$$

We hypothesized that availability and the corresponding selection function could differ among individuals in two ways. Individuals with differences in availability could exhibit the same selection despite differences in availability (Fig. 1a). Alternately, selection could change with availability for each individual, with the population exhibiting a functional response (see Fig. 1c). Model 3a uses a fixed-effects model but the range of available elevations ($x_{1,j}$) for each j individual was different. All individuals had the same selection. Model 3b uses the same shifts in the range of available as in model 3a but the strength of selection (represented as $\gamma_{1,j}$) for higher elevations by an individual (j) was inversely related to

Table 2. Grizzly bear RSF models with (a) fixed effects, (b) fixed effects with cluster, (c) mixed effects with random intercepts and (d) mixed effects fitting a random intercept and coefficients, with parameter estimates (β_i) and standard errors (SE). Elevation is in 100 m intervals, and open habitat is a categorical covariate (open = 1 or forested = 0). Log-likelihood reflects model fit. The variance estimates represent the variance in the random intercept (Int.) or coefficient (Coef.)

Model structure	Log-likelihood	Elevation x_1			Open habitat x_2			Variance	
		β_i	SE	<i>P</i>	β_i	SE	<i>P</i>	Int.	Coef.
(1) Logistic	-5902.4	-0.005	0.008	0.503	0.572	0.050	<0.001	-	-
(2) Logistic with cluster	-5902.4	-0.005	0.089	0.951	0.572	0.288	0.046	-	-
(3) Logistic with random intercept	-5555.7	0.023	0.012	0.065	0.477	0.052	<0.001	0.47	-
(4) Logistic with random intercept and random x_1	-5426.2	0.026	0.033	0.439	0.417	0.055	<0.001	19.4	0.047
(5) Logistic with random intercept and random x_2	-5499.1	0.041	0.013	0.001	0.431	0.145	0.028	0.761	0.300

the shift in $x_{1,i,j}$. This produced stronger selection for higher elevations when the mean elevation available to that individual was low and weaker selection when the mean elevation available was high. This reflects a situation where bears living at lower elevations show strong selection for higher elevations within their home range, whereas bears living in high mountain areas do not exhibit selection for high elevations areas because these areas may be unproductive high alpine areas. Models 3c and 3d mirrored those above but for the categorical open habitat covariate. In both scenarios, the availability of the two resource categories differed among individuals. The prevalence of open habitat ranged from 7% to 84% and in Model 3d, selection for open habitat was related to its prevalence such that selection increased as open habitat declined in prevalence and selection decreased when open habitat was more prevalent. This type of functional response to open habitat could occur if grizzly bears were obtaining most of their forage in this open habitat so they would exhibit strong selection for this habitat when it is rare, but much weaker selection for this habitat when it, and the forage it contains, is abundant.

STATISTICAL ANALYSES

The statistical analyses of our simulated data were the same as for the grizzly bear data, but we used XTLOGIT in STATA version 8.2 (StataCorp 2003) to solve models with only a random intercept.

Results

GRIZZLY BEAR RSF

Model coefficients, their standard errors and random effect variances are presented in Table 2. The 'naive' RSF model indicated that relative probability of use was higher in open habitats and declined at higher elevations (Table 2). Instead of reducing variance by clustering on individual bears, the Huber-White variance estimator (cluster) increased the standard error on the coefficients for both open habitat and elevation

(Table 2). The addition of a random intercept improved model fit substantially and changed the magnitude of the coefficients with the coefficient for elevation becoming marginally significant and, notably, changing sign (Table 2). In the model with a random coefficient for open habitat, model fit improved again, and the coefficient for elevation changed markedly from being negative and non-significant to being positive and highly significant (Table 2). The model with a random coefficient for elevation exhibited similar results to the model with only a random intercept with relatively large variance in the random intercept (Table 2). Conditional estimates for selection for elevation for individual grizzly bears (Fig. 2) confirms the absence of functional responses or more complex patterns in selection, yet reveals clearly how much individual variation in selection for elevation occurs. Clearly, the variability in coefficients and their significance yields differing conclusions depending on the model used. In most of the models, managers would conclude that elevation is not

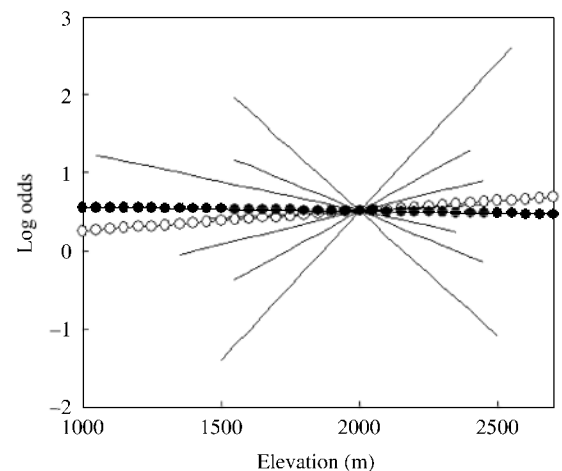


Fig. 2. Conditional estimates of the relative predicted probability of use as a logit function for individual grizzly bear selection for elevation (thin lines), the marginal population estimate (connected white dots), and the traditional fixed-effects logit model estimate (connected black dots) for a grizzly bear RSF model with a random coefficient for elevation.

an important variable, but its effect becomes very strong once a random coefficient for open habitat is added to the model. The model with the random coefficient for elevation is, however, a much better fit to the data, measured by the $\log(\mathcal{L})$.

In these data, individual bears had differing sample sizes of used points, differing home ranges and hence differing availability, and they appear to have differing selection for both elevation and the open habitat variable. It is not clear, however, which of these individual differences are exerting the greatest influence in the random effects models.

SIMULATIONS

Balanced vs. unbalanced designs

When simulated data contained no variation in resource selection among individuals and the design was balanced across individuals, as expected, the inclusion of a random effect did not improve model fit (Table 3a). Log-likelihoods [$\log(\mathcal{L})$] and coefficient estimates were stable across all modelling approaches. As expected, there was very little variation in the intercept and coefficient estimates for models that included respective random variables. In contrast, when the design was unbalanced across individuals (a range of 31–181 use points per individual), model fit was improved with the inclusion of a random intercept (Table 3b, Fig. 3a). All three mixed-effect models resulted in a similar decrease in $\log(\mathcal{L})$ compared to the fixed effect logistic model. Coefficients and standard errors were fairly robust across all models, with coefficients in mixed-effect models deviating only slightly from the fixed-effect logistic model. In the unbalanced design, where the model included both a random intercept and coefficient, the variance in the random intercept was much larger than the variance in the random coefficient,

when compared relative to the coefficient estimate, suggesting that individuals vary primarily in their intercept. In both balanced and unbalanced designs, clustering on individuals using the Huber–White sandwich estimator resulted in decreased standard errors for both the continuous and categorical coefficients (Table 3).

Differences in selection

Simulations introduced variation among individuals in selection for either elevation or open habitat when availability was constant (Table 4). Adding a random intercept did not affect model fit or coefficient estimates. The Huber–White sandwich estimator (clustering) inflated standard errors for simulated random individual coefficients (Table 4a,b). However, clustering deflated the standard error associated with open habitat where individuals varied only in their response to elevation (Table 4a, Fig. 3b) and for elevation where individuals varied only in their selection for open habitat (Table 4b). Models including a random intercept and coefficient (Table 4a,b) resulted in different parameter estimates and standard errors relative to the fixed-effect logistic models, as would be expected based on the substantial variance estimated in the random effect (Table 4). Changes in standard errors and coefficients were seen predominantly in the covariate for which we simulated individual variation.

Differences in availability and functional responses

Adding random effects for data with a differing range of available elevations (Table 5a), improved model fit [$\log(\mathcal{L})$] and altered β_1 and its standard error compared to the fixed-effect logistic model. Adding random effects had no influence on the model fit or parameters when there was differing availability of the two habitat types among individuals (Table 5c). In contrast, with a functional

Table 3. Parameter estimates and standard errors for logistic regression models on data simulated with no correlation structure among individuals. Models fitted to the data were (1) fixed effects (2) fixed effects with cluster (see text) (3) mixed effects fitting a random intercept (4) mixed effects fitting a random intercept and random coefficient for elevation (x_1), or (5) a random coefficient for open habitat (x_2). Elevation is a continuous covariate and habitat type is a categorical covariate. Estimates are shown for (a) a balanced design (100 used and 500 available points for each of 20 animals) and (b) an unbalanced design (31–181 used and 500 available points for each of 20 animals)

Model structure	Log-likelihood	Variable x_1		Variable x_2		Variance	
		β_i	SE	β_i	SE	Int.	Coef.
(a) Balanced design							
(1) Logistic	–5303	0.434	0.043	0.513	0.051		
(2) Logistic with cluster	–5303	0.434	0.033	0.513	0.039		
(3) Logistic with random intercept	–5303	0.434	0.043	0.513	0.051	0.000	
(4) Logistic with random intercept and random x_1	–5303	0.434	0.043	0.513	0.051	0.000	0.000
(5) Logistic with random intercept and random x_2	–5303	0.434	0.043	0.513	0.051	0.000	0.000
(b) Unbalanced design							
(1) Logistic	–5257	0.463	0.044	0.490	0.051		
(2) Logistic with cluster	–5257	0.463	0.043	0.490	0.043		
(3) Logistic with random intercept	–5168	0.467	0.044	0.494	0.051	0.568	
(4) Logistic with random intercept and random x_1	–5167	0.467	0.047	0.496	0.051	0.165	0.000
(5) Logistic with random intercept and random x_2	–5166	0.469	0.044	0.522	0.058	0.214	0.006

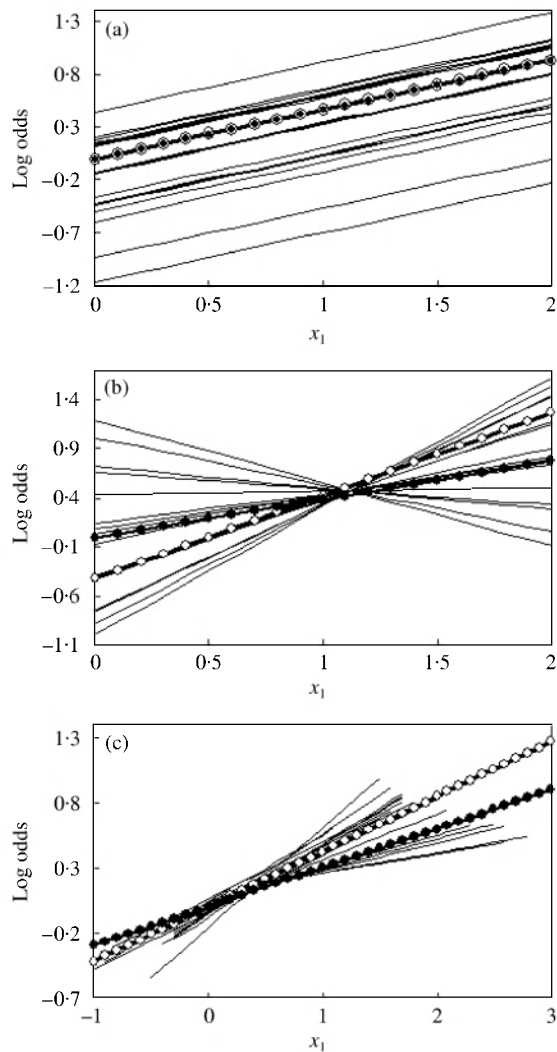


Fig. 3. Conditional estimates of the relative predicted probability of use as a logit function for simulated individuals (thin lines), the marginal population estimate (connected white dots) and the traditional fixed-effect logistic model estimates (connected black dots) for individuals having differing samples sizes and a model with a random intercept (a), differing selection and a random coefficient for elevation (x_1) (b), and a functional response to elevation (x_1) with a random coefficient for elevation (c).

response in resource selection for either elevation or open habitat (Tables 5b,d, Fig. 3c), incorporation of a random intercept and random coefficient improved model fit. Parameter estimates changed significantly for the variables that were simulated to have functional responses. Again, clustering inflated standard errors for variables that included random variation and deflated standard errors for variables simulated without individual variation.

Discussion

Our empirical and simulated examples demonstrate the utility and need for the application of random effects for estimating population-level responses in studies of

resource selection. The analysis of the grizzly bear telemetry data demonstrated that inferences from resource selection models can change with the addition of random effects, suggesting important group level correlation that would otherwise be overlooked. For example, the strength of grizzly bear selection for elevation varied greatly depending on whether a random coefficient for open habitat was included in the model (Table 2). Model fit was greatly improved with the addition of random effects, suggesting that random effects have merit in grizzly bears RSF models, and conditional estimates of selection (Fig. 2) for elevation illustrates wide individual variation in this trait. The greatest improvement in model fit came from the addition of a random intercept (Table 2), which our simulations revealed could compensate for the widely unbalanced samples among bears (Tables 1 and 3). Further improvements in model fit to the grizzly bear data with the addition of random coefficients combined with the results from the simulations illustrates that there appear to be differences among individual bears in their selection for these two variables. While a functional response could be conceivable for elevation, conditional estimates from Fig. 2 clearly illustrated that the pattern was a result of variation in selection, not a functional response.

Where sample sizes are balanced among individuals and animals respond to resources in a similar way we found, as expected, random effects to be unnecessary for estimating coefficients for an RSF model. However, for unbalanced designs, including a random intercept provides an alternative to compositional analyses (Aebischer *et al.* 1993) or rarefaction of data (Swihart & Slade 1985). The individual animal is accounted for as the sample unit, and the predicted probability of use for the population is independent of the sampling intensity for individuals (Table 1). In the grizzly bear data, three bears had roughly five times as many locations as three other bears, which would normally result in those bears having five times the influence on model coefficients (Table 1). Using a random intercept alone to account for this imbalance changed the direction of the response to elevation and the coefficient changed from being non-significant to being marginally significant, and dramatically improved model fit. Use of the Huber–White variance estimator to generate ‘robust’ standard errors would have concluded that the selection for open habitat was only marginally significant, a conclusion quite different from the one drawn from the model with a random intercept.

Our results suggest that using the Huber–White variance estimator (White 1982; Pendergast *et al.* 1996) may help to identify correlation structure among individuals. In our simulated balanced design case with no correlation structure among individuals (Table 1a), standard errors estimated with the Huber–White estimator (clustering) decreased relative to the fixed effect logistic model. In the simulation, where variation was induced among individuals in their selection, and in the grizzly bear example (Table 2), clustering inflated standard errors.

Table 4. Parameter estimates and standard errors for logistic regression models on data simulated with (a) individual variation in their response to elevation (x_1) and (b) individual variation in their response to the open habitat (x_2). Availability was constant across all individuals. Models fitted to the data were the same as for Tables 2 and 3

Model structure	Log-likelihood	Variable x_1		Variable x_2		Variance	
		β_i	SE	β_i	SE	Int.	Coef
(a) Random individual coefficients for x_1							
(1) Logistic	-5331	0.387	0.043	0.414	0.050		
(2) Logistic with cluster	-5331	0.387	0.150	0.414	0.036		
(3) Logistic with random intercept	-5331	0.387	0.043	0.414	0.050	0.000	
(4) Logistic with random intercept and random x_1	-5250	0.423	0.150	0.426	0.051	0.514	0.408
(5) Logistic with random intercept and random x_2	-5331	0.387	0.043	0.414	0.050	0.000	0.000
(b) Random individual coefficients for x_2							
(1) Logistic	-5326	0.420	0.043	0.404	0.050		
(2) Logistic with cluster	-5326	0.420	0.040	0.404	0.158		
(3) Logistic with random intercept	-5326	0.420	0.043	0.404	0.050	0.000	
(4) Logistic with random intercept and random x_1	-5326	0.420	0.043	0.404	0.050	0.000	0.000
(5) Logistic with random intercept and random x_2	-5261	0.428	0.044	0.460	0.165	0.199	0.488

Table 5. Parameter estimates and standard errors for logistic regression models on data simulated with differing availabilities among individuals and with and without a functional response to these changes in availability. (a) constant response to elevation (x_1) with changing availability, (b) changing response to elevation (x_1) with changing availability, (c) constant response to open habitat (x_2) with changing availability and (d) changing response to open habitat (x_2) with changing availability. Simulation models fitted to the data were the same as for Table 4

Model structure	Log-likelihood	Variable x_1		Variable x_2		Variance	
		β_i	SE	β_i	SE	Int.	Coef
(a) Constant selection, changing availability of x_1							
(1) Logistic	-5335	0.251	0.033	0.465	0.050		
(2) Logistic with cluster	-5335	0.251	0.037	0.465	0.046		
(3) Logistic with random intercept	-5335	0.251	0.033	0.465	0.050	0.000	
(4) Logistic with random intercept and random x_1	-5333	0.310	0.052	0.465	0.050	0.000	0.007
(5) Logistic with random intercept and random x_2	-5334	0.301	0.049	0.465	0.055	0.001	0.010
(b) Differing selection as a function of changing availability of x_1							
(1) Logistic	-5317	0.299	0.033	0.494	0.050		
(2) Logistic with cluster	-5317	0.299	0.056	0.494	0.042		
(3) Logistic with random intercept	-5317	0.299	0.033	0.494	0.050	0.000	
(4) Logistic with random intercept and random x_1	-5305	0.423	0.066	0.496	0.051	0.013	0.038
(5) Logistic with random intercept and random x_2	-5312	0.417	0.048	0.496	0.054	0.046	0.008
(c) Constant selection, changing availability of x_2							
(1) Logistic	-5317	0.468	0.044	0.391	0.049		
(2) Logistic with cluster	-5317	0.468	0.042	0.391	0.042		
(3) Logistic with random intercept	-5317	0.468	0.044	0.391	0.049	0.000	
(4) Logistic with random intercept and random x_1	-5317	0.468	0.044	0.391	0.049	0.000	0.000
(5) Logistic with random intercept and random x_2	-5317	0.468	0.044	0.391	0.049	0.000	0.000
(d) Differing selection as a function of changing availability of x_2							
(1) Logistic	-5253	0.436	0.044	0.717	0.051		
(2) Logistic with cluster	-5253	0.436	0.035	0.717	0.176		
(3) Logistic with random intercept	-5253	0.436	0.044	0.717	0.051	0.000	
(4) Logistic with random intercept and random x_1	-5253	0.436	0.044	0.723	0.055	0.000	0.000
(5) Logistic with random intercept and random x_2	-5154	0.430	0.044	0.741	0.191	0.150	0.679

Thus, clustering may have utility as a diagnostic, directing researchers to where random effects may be necessary. Further work is needed to verify these preliminary suggestions.

We considered only one level of nesting in our simulated examples. In the presence of multiple hierarchies, random effects become even more important (Ten

Have, Kunselman & Tran 1999; Begg & Parides 2003). For example, individuals can be nested within herds, which are nested themselves in river basins, or subpopulations. Studies of resource selection of social animals in such settings have suffered from an inability to accommodate multiple levels of nesting (Garshelis 2000; Morrison 2001). The most important consideration, however, is

that including a random effect in studies with inherent hierarchical structure ensures that the marginal population inferences of the resultant RSF will be valid (Cam *et al.* 2002; Cooch, Cam & Link 2002), and will provide appropriate conditional (group) level inferences (e.g. Fig. 2). Although we focused on marginal effects (population-level) here, mixed-effect models provide a powerful approach for examining evolutionary processes and questions related to the fitness consequences of individual-level variation in studies of resource selection (Franklin *et al.* 2000). For example, relying on marginal (population) inferences for an endangered species may hide important conditional (subpopulation or individual) differences that could have important implications for conservation. Conditional estimates of resource selection could be used to identify which subpopulations to focus conservation efforts upon.

Our simulated examples demonstrate that random intercepts can correct for unbalanced designs, but balanced use–availability designs may require both a random intercept and coefficient to detect individual variation in selection. Simulations in other fields (Ten Have *et al.* 1999) draw similar conclusions regarding the importance of random coefficients. Many wildlife studies thus far, however, have focused upon the inclusion of a random intercept without incorporation of random coefficients (Cam *et al.* 2002; Franklin, Anderson & Burnham 2002; Boyce, Irwin & Barker 2005). We caution that in resource selection studies with use–availability designs, including only a random intercept will only account for differences in sample sizes but not for differences in selection among individuals. In our unbalanced simulation, adding a random coefficient in addition to a random intercept decreased the random intercept variance and improved model fit slightly. This was probably a result of the random coefficient explaining some of the variance in the random intercept (Skrondal & Rabe-Hesketh 2004) and accounting for slight differences in the coefficient due to the random generation of used and available points.

Perhaps the most compelling argument for considering random coefficients in RSF models comes from the ability of random coefficients to model functional responses (Myerud & Ims 1998). Myerud & Ims (1998) provide a simple framework for assessing functional responses in examples with two habitat types (e.g. Osko *et al.* 2004). However, available resources are often more than two categories or continuous, and Myerud & Ims (1998) concluded by urging future studies to consider generalizations of the logit model. Our results suggest that inclusion of a random intercept and coefficient provides a useful generalization. To our knowledge, this is the first demonstration of an approach to model functional responses effectively in resource selection. As a guide in using random effects to uncover functional responses, we offer the following suggestions. The isolation of functional responses in continuous covariates may require a multifaceted approach. Consider that after we simulated a functional response in elevation (Table 3b), we improved model fit over the

fixed-effect model by including a random coefficient for open habitat rather than elevation. This is an example of conditionality between the model intercept and the categorical covariate coefficient. When the coefficient for the continuous covariate (elevation) is altered, individual intercepts are altered (see Fig. 3b), having an effect on the categorical variable (habitat type), because the effect of habitat type = 0 is absorbed by the intercept. Even so, results in Table 3b indicate a functional response in elevation given the magnitude of change in model fit. Thus, we believe that measures of model fit will be critical to assessing where functional responses in RSF occur when there is no a priori decision to consider particular random effects (see also Greenland 2000).

Critical to modelling a functional response in resource selection is identification of a resource type that is limiting in a trade-off situation (Myerud & Ims 1998). Without a trade-off, constant selection (Fig. 1a) will be possible (e.g. a constant proportion of habitat in a home range). However, as in the grey squirrel (*Sciurus carolinensis* Gmelin) example in Myerud & Ims (1998), grey squirrels made a trade-off once the amount of cropland increased beyond some threshold (Fig. 1 of Myerud & Ims 1998), showing avoidance for cropland once availability of cropland exceeded 30% of a squirrel's home range. Often, however, ecologists will be faced with the problem of identifying for which covariate the random effect or functional response occurs. Simple approaches include graphical examination of conditional effects (e.g. Figure 2), and dividing animals into two groups for preliminary RSF modelling (Osko *et al.* 2004).

An additional challenge for researchers is that in a RSF design as the number of available points increases, the magnitude of the $\log(\mathcal{L})$ also increases (unpublished data), and therefore model selection using AIC or similar likelihood approaches may be sensitive to the choice of the number of available points. Finding a way to use information-theoretic approaches in RSF studies and in mixed-effect models is an issue that deserves future attention.

Conclusions

Animal data often possess nested or grouped data structures, and inclusion of random effects in resource selection and species distribution models will accommodate such data structures, yielding more robust inference. Random effects improve our ability to account for differences in selection or sample size among individuals or groups and their inclusion can affect the conclusions drawn. Conditional inferences from these mixed effect models will allow researchers to make group-specific inferences, with obvious applications to endangered species management and other conservation applications where individual level variation is important. By including random coefficients, the assumption that selection patterns remain constant as availability changes need no longer restrict the development and application of RSF models. We believe that relaxation of this

requirement will provide increasingly flexible and powerful resource selection models that allow extrapolation beyond study area borders with increasing biological realism, efficiency and validity. Given the success of existing resource selection modelling approaches in natural resource management, we believe specification of the functional response will increase the utility of these models to ecology and conservation.

Acknowledgements

We thank M. Boyce, S. Cumming, S. Lele, M. Lewis, E. Merrill, C. Paszkowski, F. Schmiegelow and C. St Clair for supervisory support. We thank M. Taper for a thorough statistical review, and E. Bayne, M. Boyce, L. McDonald, D. Strickland, C. St Clair and two anonymous reviewers for comments that greatly improved the manuscript. We kindly thank G. Stenhouse from the Foothills Model Forest Grizzly Bear Project (http://www.fmf.ca/pa_GB.html) for providing the grizzly bear data. Financial support for the authors during the time this manuscript was prepared was provided by Canon-National Parks Science Scholarship for the Americas (M.H.), Alberta Conservation Association (J.L.F.), Izaak Walton Killiam Memorial Predoctoral Scholarship (C.L.A.), Prairie Adaptation Research Collaborative, Province of Alberta Graduate Fellowship (M.A.K.) and the Province of Alberta Postgraduate Scholarship and Weyerhaeuser Company (D.J.S.).

References

- Aebischer, N.J., Robertson, P.A. & Kenward, R.E. (1993) Compositional analysis of habitat use from animal radio-tracking data. *Ecology*, **74**, 1313–1325.
- Begg, M.D. & Parides, M.K. (2003) Separation of individual-level and cluster-level covariate effects in regression of correlated data. *Statistics in Medicine*, **22**, 2591–2602.
- Bennington, C.C. & Thayne, W.V. (1994) Use and misuse of mixed-model analysis of variance in ecological studies. *Ecology*, **75**, 717–722.
- Bingham, R.L. & Brennan, L.A. (2004) Comparison of type I error rates for statistical analyses of resource selection. *Journal of Wildlife Management*, **68**, 206–212.
- Boyce, M.S., Irwin, L.L. & Barker, R. (2005) Demographic meta-analysis: synthesizing vital rates for spotted owls. *Journal of Applied Ecology*, **42**, 38–49.
- Breslow, N.E. & Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multi-Model Inference*. Springer-Verlag, New York.
- Burnham, K.P. & White, G.C. (2002) Evaluation of some random effects methodology applicable to bird ringing data. *Journal of Applied Statistics*, **29**, 245–264.
- Cam, E., Link, W.A., Cooch, E.G., Monnat, J.Y. & Danchin, E. (2002) Individual covariation in life-history traits: seeing the trees despite the forest. *American Naturalist*, **159**, 96–105.
- Cooch, E.G., Cam, E. & Link, W.A. (2002) Occam's shadow: levels of analysis in evolutionary ecology – where to next? *Journal of Applied Statistics*, **29**, 19–48.
- Follmann, D.A. & Lambert, D. (1989) Generalized logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, **84**, 295–301.

- Franklin, A.B., Anderson, D.R. & Burnham, K.P. (2002) Estimation of long-term trends and variation in avian survival probabilities using random effects models. *Journal of Applied Statistics*, **29**, 267–287.
- Franklin, A.B., Anderson, D.R., Gutierrez, R.J. & Burnham, K.P. (2000) Climate, habitat quality, and fitness in northern spotted owl populations in northwestern California. *Ecological Monographs*, **70**, 539–590.
- Franklin, S.E., Stenhouse, G.B., Hansen, M.J., Popplewell, C.C., Dechka, J.A. & Peddle, D.R. (2001) An integrated decision tree approach (IDTA) to mapping landcover using satellite remote sensing in support of grizzly bear habitat analysis in the Alberta Yellowhead Ecosystem. *Canadian Journal of Remote Sensing*, **27**, 579–592.
- Garshelis, D.L. (2000) Delusions in habitat evaluation: measuring use, selection, and importance. *Research Techniques in Animal Ecology: Controversies and Consequences* (eds L. Boitani & T.L. Fuller), pp. 111–154. Columbia University Press, New York.
- Greenland, S. (2000) When should epidemiologic regressions use random coefficients? *Biometrics*, **56**, 915–921.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Hosmer, D.W. & Lemeshow, S. (2000) *Applied Logistic Regression*. John Wiley and Sons, New York.
- Hurlbert, S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, **54**, 187–211.
- Johnson, C.J., Seip, D.R. & Boyce, M.S. (2004) A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. *Journal of Applied Ecology*, **41**, 238–251.
- Johnson, D.H. (1980) The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, **61**, 65–71.
- Krawchuk, M.A. & Taylor, P.D. (2003) Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos*, **103**, 153–161.
- Leban, F.A., Wisdom, M.J., Garton, E.O., Johnson, B.K. & Kie, J.G. (2001) Effect of sample size on the performance of resource selection analyses. *Radio Tracking and Wildlife Populations* (eds J.J. Millspaugh & J.M. Marzluff), pp. 291–307. Academic Press, New York.
- Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2002) *Resource Selection by Animals: Statistical Analysis and Design for Field Studies*, 2nd edn. Kluwer, Boston.
- McLellan, B.N. & Hovey, F.W. (2001) Habitats selected by grizzly bears in a multiple use landscape. *Journal of Wildlife Management*, **65**, 92–99.
- McNay, R.S. & Bunnell, F.L. (1994) Characterizing independence of observations in movements of Columbian black-tailed deer. *Journal of Wildlife Management*, **58**, 422–429.
- Mladenoff, D.J., Sickley, T.A., Haight, R.G. & Wydeven, A.P. (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the northern great-lakes region. *Conservation Biology*, **9**, 279–294.
- Morrison, M.L. (2001) Invited paper: a proposed research emphasis to overcome the limits of wildlife-habitat relationship studies. *Journal of Wildlife Management*, **65**, 613–623.
- Myerud, A. & Ims, R.A. (1998) Functional responses in habitat use: availability influences relative use in trade-off situations. *Ecology*, **79**, 1435–1441.
- Natarajan, R. & McCulloch, C.E. (1999) Modeling heterogeneity in nest survival data. *Biometrics*, **55**, 553–559.
- Neter, J., Kutner, M.H., Wasserman, W., Nachtsheim, C.J. & Neter, J. (1996) *Applied Linear Statistical Models*, 4th edn. McGraw-Hill Publishers, New York.

- Newey, W.K. & West, K.D. (1987) A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.
- Nielsen, S.E., Boyce, M.S., Stenhouse, G.B. & Munro, R.H.M. (2002) Modeling grizzly bear habitats in the Yellowhead ecosystem of Alberta: taking autocorrelation seriously. *Ursus*, **13**, 45–56.
- Osko, T.J., Hiltz, M.N., Hudson, R.J. & Wasel, S.M. (2004) Moose habitat preferences in response to changing availability. *Journal of Wildlife Management*, **68**, 576–584.
- Otis, D.L. & White, G.C. (1999) Autocorrelation of location estimates and the analysis of radiotracking data. *Journal of Wildlife Management*, **63**, 1039–1044.
- Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. & Fisher, M.R. (1996) A survey of methods of analyzing clustered binary response data. *International Statistics Review*, **64**, 89–118.
- Pinheiro, J.C. & Bates, D.M. (2000) *Mixed Effects Models in and s-plus*. Springer-Verlag, New York.
- Rabe-Hesketh, S., Pickles, A. & Skrondal, A. (2001) *GLLAMM Manual*. Department of Biostatistics and Computing, Institute of Psychiatry, Kings College, University of London, London.
- Rushton, S.P., Ormerod, S.J. & Kerby, G. (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**, 193–200.
- Skrondal, A. & Rabe-Hesketh, S. (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall, New York.
- StataCorp (2003) *Stata Statistical Software*, Release 8.0. Stata Corporation, College Station, Texas, USA.
- Swihart, R.K. & Slade, N.A. (1985) Testing for independence of observations in animal movements. *Ecology*, **66**, 1176–1184.
- Ten Have, T.R., Kunselman, A.R. & Tran, L. (1999) A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Statistics in Medicine*, **18**, 947–960.
- Vaida, F. & Blanchard, S. (2005) Conditional Akaike criteria for mixed models. *Biometrika*, **92**, 351–370.
- White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–26.

Received 25 October 2005; accepted 6 March 2006