

Review

Correlation and studies of habitat selection: problem, red herring or opportunity?

John Fieberg^{1,*}, Jason Matthiopoulos², Mark Hebblewhite³,
Mark S. Boyce⁴ and Jacqueline L. Frair⁵

¹*Biometrics Unit, Minnesota Department of Natural Resources, 5463-C W. Broadway, Forest Lake, MN 55434, USA*

²*NERC Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife KY16 8LB, UK*

³*Wildlife Biology Program, College of Forestry and Conservation, University of Montana, Missoula, MT 59812, USA*

⁴*Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9*

⁵*State University of New York, College of Environmental Science and Forestry, Syracuse, NY 13210, USA*

With the advent of new technologies, animal locations are being collected at ever finer spatio-temporal scales. We review analytical methods for dealing with correlated data in the context of resource selection, including post hoc variance inflation techniques, ‘two-stage’ approaches based on models fit to each individual, generalized estimating equations and hierarchical mixed-effects models. These methods are applicable to a wide range of correlated data problems, but can be difficult to apply and remain especially challenging for use–availability sampling designs because the correlation structure for combinations of used and available points are not likely to follow common parametric forms. We also review emerging approaches to studying habitat selection that use fine-scale temporal data to arrive at biologically based definitions of available habitat, while naturally accounting for autocorrelation by modelling animal movement between telemetry locations. Sophisticated analyses that explicitly model correlation rather than consider it a nuisance, like mixed effects and state-space models, offer potentially novel insights into the process of resource selection, but additional work is needed to make them more generally applicable to large datasets based on the use–availability designs. Until then, variance inflation techniques and two-stage approaches should offer pragmatic and flexible approaches to modelling correlated data.

Keywords: generalized estimating equation; generalized linear mixed model; hierarchical model; resource-selection function; telemetry; use–availability

1. INTRODUCTION

Wildlife telemetry data often are positively correlated in space and time. Correlation patterns may be simple, decreasing monotonically with temporal interval or Euclidean distance, or more complex, fluctuating with time interval or some non-Euclidean metric of distance. The causes of correlation in telemetry studies are varied and may be linked to physical and physiological limitations (e.g. constraints on the speed of the animal; Fancy & White 1987; Verwajen & Van Damme 2008), within animal behavioural processes (e.g. periodic feeding patterns or migration schedules; Cushman *et al.* 2005; Fryxell *et al.* 2008), between animal behavioural processes (e.g. sociality; Haydon *et al.* 2008) or extrinsic forces (e.g. seasonality in temperature or resource availability, landscape heterogeneity, precipitation or prey

abundance; Johnson *et al.* 2002; Wittemyer *et al.* 2008). Further, repeated observations on the same individuals, as is the case with radio-telemetry locations, are often assumed to give rise to constant within-group (or ‘exchangeable’) correlation structures. For example, analysts may assume a constant correlation between any two observations from the same individual (with independence among observations taken on different individuals). These assumptions easily generalize to hierarchically structured populations (e.g. one may assume a constant correlation among observations from the same individual, correlation to a lesser degree among observations from animals in the same herd and independence among observations from animals in different herds). We will commonly use the term ‘correlation’ to allow for these more general correlation structures, while reserving the term ‘autocorrelation’ to refer specifically to non-exchangeable within-animal correlation patterns (e.g. correlations that change as a function of temporal or spatial distance between observations).

* Author for correspondence (john.fieberg@state.mn.us).

One contribution of 15 to a Theme Issue ‘Challenges and opportunities of using GPS-based location data in animal ecology’.

Until recently, ecologists viewed correlation as a nuisance because it violates the assumption of independence in traditional statistical inference, hence producing deceptively low estimates of uncertainty, overfit models and spurious conclusions. The importance given to obtaining reliable measures of statistical significance, often at the expense of learning from biologically relevant patterns in the data (Johnson 1999), has led some authors to dub autocorrelation a 'red herring' (Diniz-Filho *et al.* (2003) and Fieberg (2007a) in the context of species richness and home-range studies, respectively). Autocorrelation has been shown to impact inferences from habitat-selection models (Nielsen *et al.* 2002), and the ability to collect fine-scale temporal data with global positioning system (GPS) radiotelemetry, combined with the ongoing challenges of scaling up individual-level differences to population-level inferences, suggests that correlation has the potential to divert attention away from studying biologically meaningful patterns in habitat-selection studies too. Ideally, methods need to be developed that exploit fine-scale temporal information without sacrificing the ability to make reliable inferences.

Several approaches have been suggested for modelling correlated data in the context of resource-selection studies, and we use that body of literature to frame this review. Recommended solutions span a range of techniques with broad applicability, but studies of resource selection also present unique challenges because they compare animal space-use to habitat availability (by various definitions; Manly *et al.* 2002). Data are often modelled by a binomial process, rather than a normal distribution of errors and correlation patterns of used and available points are unlikely to fit common parametric forms. As such, resource-selection studies provide an illuminating frame of reference for considering correlation issues and contemporary solutions.

There have been three main avenues for addressing correlation in the literature: (i) data censoring: collect data *a priori* (or rarify through subsampling) in a manner that attempts to satisfy the assumption of independence, (ii) variance inflation: a post hoc approach to adjust (inflate) standard errors after parameter estimation, or (iii) explicit modelling of correlation: change the statistical framework used for inference to explicitly account for correlation in the process. Early efforts to address serial correlation in telemetry studies used the concept of 'time to independence' (Swihart & Slade 1985; Rooney *et al.* 1998; Salvatori *et al.* 1999), defined either as the time lag beyond which autocorrelation drops below a critical value, or, alternatively, a time interval sufficient for the study animal to move anywhere in its home range. Researchers using very high frequency (VHF) radiotelemetry often temporally separated telemetry locations *a priori* by design (Laird 1987), or by subsampling (Hansteen *et al.* 1997). Not surprisingly, these approaches often led to unacceptable data losses. For example, McNay & Bunnell (1994) deleted approximately 90 per cent of their locations on Columbian black-tailed deer (*Odocoileus hemionus*) to achieve statistical independence. In the case of GPS

collars, this would be counterproductive because their great strength lies in finer characterization of animal behaviour in space and time. Although it is nearly always the case that estimator precision will increase with increasing sample size, gains in precision usually plateau at some level of sampling. Thus, when designing telemetry studies, costs associated with collecting higher frequency data (e.g. monitoring costs associated with additional VHF sampling or decreased battery life of GPS collars) should be considered relative to the benefits of collecting these additional data. However, once collected, it will be beneficial to use all available data and deal with the effects of correlation on statistical inference through variance inflation or explicit likelihood models (approaches (ii) and (iii) above). We therefore focus our review on these two broad approaches.

In this review, we first revisit the importance of obtaining a representative sample of locations and individuals (§2), and briefly discuss approaches that set the foundation of using the animal, rather than the telemetry location, as the unit of replication (§3) (Cagnacci *et al.* 2010; Hebblewhite & Haydon 2010). These sections set the conceptual stage for the more complicated, but flexible, regression approaches that we consider in depth in §4. We start by reviewing methods for first estimating models under the assumption of independence between locations, but then inflating variances using robust standard errors to account for the presence of correlation (Nielsen *et al.* 2002; Clark & Strevens 2008; Craiu *et al.* 2008). Next, we discuss extensions based on generalized estimating equations (GEEs) that, assuming non-independence, allow the use of other correlation structures during the estimation process in conjunction with robust standard errors for inference (Zeger *et al.* 1988; Fieberg *et al.* 2009; Koper & Manseau 2009). We then review two methods that explicitly model among-animal variability in their selection patterns: first, a two-stage approach that fits models separately to individual animals and then averages regression parameters across individuals to estimate population-level selection patterns (Glenn *et al.* 2004; Sawyer *et al.* 2006); and second, generalized linear mixed-effects models (GLMMs) that include random effects for individual animals as well as other social grouping structures (Gillies *et al.* 2006; Aarts *et al.* 2008; Hebblewhite & Merrill 2008). The latter are often referred to as multi-level or hierarchical models. Throughout, we highlight the challenges particular to the use–availability sampling design, which arise because the probability of use in the sample is dependent on sampling rates of used and available points and because common correlation structures (e.g. a first-order autoregressive model AR1) are unlikely to adequately describe clusters of used and available points (Park & Kim 2004; Craiu *et al.* 2008; Koper & Manseau 2009). Finally in (§5), we discuss recent developments that take advantage of autocorrelated data by integrating models of animal movement into the process of resource selection. These approaches are attractive because they provide biologically based definitions of available habitat, and have the potential to circumvent the problem

Table 1. Comparison of the advantages and disadvantages of correlated data regression modelling approaches to studying habitat selection.

estimator	advantages	disadvantages	citations & examples
variance inflation techniques, generalized estimating equations (GEEs) with working independence correlation structure	simple to apply	treats correlation as a nuisance; is not informative with respect to animal-specific response patterns (and their variation in the population)	Newey & West (1987), Nielsen <i>et al.</i> (2002), Clark & Strevens (2008) and Craiu <i>et al.</i> (2008)
GEEs with other working correlation structures	in theory, could be used to increase precision of parameter estimators compared to models using a working assumption of independence	leads to biased estimators when used with case–control or use–availability sampling designs (Park & Kim 2004; Craiu <i>et al.</i> 2008)	Koper & Manseau (2009); but see Craiu <i>et al.</i> (2008) for potential concerns
two-stage sampling	allows subject-specific inferences and variance decomposition between and within groups and can accommodate variable habitat selection responses among individuals. Computationally and statistically simple (with fewer assumptions than GLMMs). Individual model log-likelihoods additive making model selection at population-level possible	requires enough data from each individual for separate model fits. May overestimate error in the population-level responses if uncertainty in individual-level estimates are not used when estimating population-level parameters. Difficult to adapt to complicated (e.g. non-nested) variance structures. Problematic when covariates are not present in an individual	Davidian & Giltinan (1995), Glenn <i>et al.</i> (2004), Sawyer <i>et al.</i> (2006) and Murtaugh (2007)
generalized linear mixed-models (GLMMs)	allows subject-specific inferences and variance decomposition between and within groups with a single model fit. Accommodates hierarchical correlation structures	computationally demanding. Requires more assumptions than two-stage methods (e.g. specification of a parametric distribution for animal-specific selection coefficients and within-individual correlation structure). Specification of correlation structure problematic with use–availability designs	Skrondal & Rabe-Hesketh (2004), Gillies <i>et al.</i> (2006), Thomas <i>et al.</i> (2006), Aarts <i>et al.</i> (2008) and Hebblewhite & Merrill (2008)
movement-based models of availability	may account for autocorrelation by modelling independent transition probabilities (between locations). Uses a biologically based definition of availability. Integrates easily into a matched-case control logistic regression design	may not fully account for all correlation (but can be applied with robust s.e.s). Not appropriate for resource selection at larger spatial scales. More computationally demanding (but can be implemented with matched case–control logistic regression software with distance as a covariate)	Hjermann (2000); Rhodes <i>et al.</i> (2005), Johnson <i>et al.</i> (2008), Christ <i>et al.</i> (2008), Barnett & Moorcroft (2008), Moorcroft & Barnett (2008), Forester <i>et al.</i> (2009) and Nielson <i>et al.</i> (2009)

of autocorrelation by modelling transitions between locations. There are advantages and disadvantages associated with each method, which we take care to summarize so that analysts can make informed choices as to their modelling approach (table 1).

2. IMPORTANCE OF REPRESENTATIVE SAMPLES

GPS-based locations yield unbiased representations of habitat use and movements provided that missing fixes are rare or occur randomly in time and space (Fieberg 2007a). Assuming further that location data have been collected over a fixed and biologically meaningful

study period (e.g. fixed start and end dates defining a typical breeding season), then design-based inferential techniques developed for analysing survey data can provide useful estimates of animal-specific parameters and their uncertainty (Thompson 2002). For example, population-level parameters and their uncertainty can be estimated by treating the data as if they arise from a two-stage cluster sample (individuals serve as clusters, and telemetry points as within-cluster observations) as long as individuals marked with GPS devices are representative of the larger population (Clark & Strevens 2008).

In reality, many telemetry studies track non-random samples of individuals for varying lengths of time.

Moreover, GPS-based locations themselves may be habitat-biased owing to canopy or terrain interference with satellite signals (Frair *et al.* 2010). If fix rates are not high, and the probability of a successful fix depends on environmental covariates, then a model for the probability of detection (i.e. acquiring a GPS fix) can be used to fill in missing locations or statistically correct for sampling biases (Frair *et al.* 2004; Horne *et al.* 2007; Nielson *et al.* 2009). One approach, sample weighting, also may be used to correct for systematic deviations in sampling intensity during the course of a study (Fieberg 2007b). In some cases, it may be possible to adjust for sampling-related effects by including covariates that capture differences in animal-specific sampling regimes (e.g. Börger *et al.* (2006a,b) in the context of modelling the relationship between habitat metrics and the home-range size). In other cases, it may be necessary to assess the impact of non-representative sampling on parameter estimates and also recognize a more limited scope of inference (e.g. Fieberg *et al.* (2008) for a consideration of how non-representative individuals may influence estimates of migration probabilities from telemetry data). Wherever possible, the implications of non-representative sampling, or variation in sampling intensities and durations, should be addressed.

3. THE ANIMAL AS SAMPLE UNIT: FOUNDATIONS

Statistical tests of non-uniform use of resources at the population level usually should treat individual animals as the sampling units (Aebischer *et al.* 1993; Otis & White 1999). White & Garrott (1990) and Dasgupta & Alldredge (1998) formulated hypothesis tests of selection at the population level for categorically defined habitats based on the sum of and maximal animal-specific χ^2 statistics (comparing used and available proportions), respectively. Aebischer *et al.* (1993) suggested a MANOVA-based procedure (compositional analysis) involving ratios of proportions associated with used and available habitat types where individual animals were considered as the sample unit. These population-level tests focus on mean selection patterns (averaged across animals), but variability in selection among animals is usually of interest too. For categorically defined habitat types, Calenge & Dufour (2006) offer a multi-variate analysis approach to analysing habitat-selection ratios with this latter aim in mind. All of these methods solve problems associated with unbalanced sampling among individuals, and within-individual correlation, by focusing on animal-level summaries (i.e. proportions of used and available habitat types; Kneib *et al.* 2007). More commonly, habitat-selection studies aim to understand the importance of both continuous and categorical predictors (associated with gradients, linear features like edges or habitat patches), which require more flexible regression approaches. Nevertheless, these earlier studies clearly established the logic of considering variation among individual animals, rather than among telemetry points, when making population-level inferences.

4. REGRESSION APPROACHES TO STUDYING RESOURCE SELECTION

Habitat-selection analyses investigating both continuous and categorical predictors typically infer selection of resources or habitats by comparing measured characteristics (e.g. elevation, dominant vegetation) of 'used points', represented by telemetry locations, to those of random points that represent 'available' habitats (Johnson *et al.* 2006; Lele & Keim 2006; Aarts *et al.* 2008; Lele 2009). The random points might be chosen locally (e.g. within a buffer surrounding each observed location), within an animal's home range, or over a larger landscape (study area), with the interpretation of regression parameters being dependent upon the scale of availability chosen for inference (Johnson 1980; Boyce 2006; Beyer *et al.* 2010). Statistical methods developed for analysing binary data, e.g. logistic regression (Manly *et al.* 2002), are frequently used to analyse the combined used-available sample data; however, a multi-nomial or conditional-logit likelihood can also be used (Cooper & Millsbaugh 1999; Compton *et al.* 2002; Thomas *et al.* 2006; Craiu *et al.* 2008). We do not review the nuances of different habitat-selection designs here, but refer readers to Keating & Cherry (2004), Johnson *et al.* (2006), Lele & Keim (2006), Thomas & Taylor (2006) and Beyer *et al.* (2010).

(a) *Variance inflation and GEEs*

When marked animals are representative of a larger population, and sample sizes associated with these animals are essentially random with respect to the response of interest (i.e. the number of observed telemetry locations is not influenced by individual animal movements or their resource preferences), then unbiased estimators of population parameters often can be formulated by treating all (within- and between-subject) data as though they are independent. Although regression parameter estimators will be unbiased, associated estimates of uncertainty will be optimistic when data are assumed to be independent. Thus, robust standard errors that quantify variation among independent units, typically individuals, should be used for statistical inference (Williams 2000; Clark & Strevens 2008). This approach of combining a working assumption of independence for point estimates with robust (cluster level) standard errors can be generalized to allow for more complicated data dependencies. For example, robust variance estimators exist to deal with within-individual serial correlation (Newey & West 1987; Nielsen *et al.* 2002).

Rather than inflating the variances in a post hoc manner, the assumed correlation structure can be included in the regression model under a GEEs framework (Zeger *et al.* 1988). GEEs extend generalized linear models (e.g. logistic regression, Poisson regression) and quasi-likelihood estimators to correlated data problems, and require a model for the response mean ($E[Y_i | X_i]$ —typically assumed to be a linear function of covariates, X , and regression parameters, β , on a transformed scale), and for the response variance ($\text{var}[Y_i | X_i]$ —most often assumed

to be a function of the mean, similar to generalized linear models). For example, for binary data, we have the following:

$$\log\left(\frac{E[Y_i|X_i]}{1 - E[Y_i|X_i]}\right) = X_i\beta \quad (4.1a)$$

and

$$\text{var}[Y_i] = \sigma^2 E[Y_i|X_i](1 - E[Y_i|X_i]), \quad (4.1b)$$

where σ^2 is a scale factor that allows for overdispersion (or increased variance) relative to a standard Bernoulli random variable. In addition, one specifies a ‘working correlation structure’ that describes the dependence among observations within mutually independent clusters of observations. The clusters can consist of successive observations in time, a group of observations close in space or all observations within one individual. Common working correlation assumptions include independence, exchangeability (equal correlation among all observations from the same cluster) or autoregressive (AR1); multiple working correlations also may be compared as a sort of sensitivity analysis. Estimators of regression parameters will be asymptotically unbiased (i.e. as the number of clusters goes to infinity) if the model for $E[Y_i|X_i]$ is correct, even if the correlation structure is mis-specified (Diggle *et al.* 1994), however, estimator precision will improve if the working correlation structure captures salient characteristics in the data. Outcome-dependent sampling designs (e.g. case–control and use–availability designs in which the probability $[Y_{ij} = 1]$ depends on how the sample was drawn) require special care because the model for $E[Y_i|X_i]$ must account for the sampling design (Park & Kim 2004; Craiu *et al.* 2008; Forester *et al.* 2009). In particular, outcome-dependent sampling designs for binary data, when analysed with working correlations other than independence, often result in biased regression parameter estimators (Park & Kim 2004; Craiu *et al.* 2008). Thus, applications of GEEs to habitat-selection problems should typically assume an independent correlation structure. In any case, confidence intervals and hypothesis tests are formed using robust sandwich standard errors (Zeger *et al.* 1988; Williams 2000), which depend on the *a priori* specification of mutually independent clusters. Alternatively, Pan (2001) developed a quasi-likelihood-based information selection criterion that can be used for model selection.

Various approaches have been used to define clusters in resource-selection analyses. Clark & Strevens (2008) used a single set of available points sampled from a larger study area and treated used points from individuals as unique clusters while assigning each available point to its own cluster. Thus, available points were assumed to be mutually independent and also independent from used points, whereas used points were assumed to be correlated within individuals. In contrast, Fortin *et al.* (2005), Craiu *et al.* (2008) and Forester *et al.* (2009) matched each observed location to a separate set of randomly sampled available locations (to serve as ‘controls’), with availability defined by potential movement

trajectories from the last known location (see §5 for more details). To define independent clusters, Fortin *et al.* (2005) estimated animal-specific autocorrelation functions for observed step lengths (the Euclidean distance between paired locations). Their analysis indicated that animals moved independently from one another, and that steps of the same animal more than 14 lags apart (in this case 70 h) exhibited little correlation. Thus, Fortin *et al.* (2005) discarded sets of 15 observations, to yield a set of mutually independent clusters of observations. Forester *et al.* (2009) applied a similar approach but modelled the autocorrelation function of summed deviance residuals (summing residuals associated with both used and available points within each matched set) to determine independent clusters. They split their data into two sets of mutually independent clusters based on a time lag of 10 observations (or 50 h). Rather than discard one of these sets of observations, Forester *et al.* (2009) estimated regression parameters using the full dataset, and estimated standard errors as the average of robust sandwich standard errors applied separately to the two mutually independent datasets. Lastly, Craiu *et al.* (2008) analysed GPS data collected from plains bison (*Bison bison*), in which clusters of 48 hourly locations, themselves separated by 120 h, were collected for each animal. Similar to Fortin *et al.* (2005) and Forester *et al.* (2009), they assumed that these clusters of observations (within each animal), as well as observations from different animals, were mutually independent. They also tested their approach using simulated data, and found that sandwich standard errors performed well, whereas naive variance estimates that assumed independence were an order of magnitude too small.

(b) Two-stage methods for estimating population-level effects

As one of the simplest and most intuitive approaches to dealing with autocorrelation, regression models may be fit to data from individual animals and then averaged to determine population-level responses (e.g. Nielsen *et al.* 2002; Glenn *et al.* 2004; Sawyer *et al.* 2006). This two-stage approach provides an alternative to fitting random-effect models (§5) when sufficient data have been collected to allow efficient estimation of individual-specific regression parameters, β_i , where i indexes unique individuals (Davidian & Giltinan 1995). Typically, sample means and variances are used in the second stage, e.g. to characterize population means and variances in animal-specific selection coefficients (e.g. Sawyer *et al.* 2006); however, variance estimators will be biased high unless the sampling uncertainty associated with the β_i 's is accounted for in the estimation process (Davidian & Giltinan 1995). More complicated population models also can be fit to explore relationships between individual-level covariates (e.g. gender) and variation in the β_i 's (Davidian & Giltinan 1995). Although less sophisticated than mixed models (see §4c), the two-stage approach has several practical advantages (table 1). For example, it is not necessary to assume a distribution for the random effects, and

within-individual correlation patterns do not need to be correctly specified when estimating mean animal-specific parameters, $\bar{\beta} = E(\beta_i)$, provided that the animal-specific parameter estimators are themselves unbiased (Murtaugh 2007). Thus, two-stage approaches will produce unbiased estimators of $\bar{\beta}$ assuming individuals are independent.

Model selection may, at first, appear problematic for two-stage approaches. However, the log likelihoods for individual animal models will be additive if animals can be treated as independent (Davidian & Giltinan 1995). Thus, likelihood ratio tests may be used if sample sizes for each individual are large enough for the χ^2 approximation to the likelihood ratio statistic to be valid (e.g. Molenberghs & Verbeke 2005, p. 354). If the goal is to characterize typical selection patterns and their variability among animals, model selection performed at the population level makes more sense than determining different covariate structures for each individual (i.e. the coefficient pertaining to a particular variable may be approximately 0 for some animals, but not all). A potential drawback of the two-stage approach is that sufficient data must exist for fitting individual-level models (Davidian & Giltinan 1995). Dropping individuals that are followed for only a short period of time could make the sampled data less representative of the study population if, for example, sampling duration was biased by the type of individual. Estimation of population parameters also might become problematic if some individuals are never exposed to certain types of habitat. For example, animal-specific parameters associated with a factor (say, dominant vegetation type) will not be estimable unless each individual is exposed to all levels of the factor (i.e. all categories of dominant vegetation). However, lumping of similar categories (or lumping little used categories into the reference category) provides a reasonable, if not altogether satisfying, solution. On the other hand, focusing the analysis on individual unit-level summaries can help to clarify the amount of information in the data (Murtaugh 2007). For example, if only a single animal is exposed to a specific habitat type, then a two-stage approach will fail (i.e. regression parameters associated with this habitat type will not be estimable except for this animal), making it clear that all information regarding that type is derived from a single animal. It is unclear how mixed-effects models would perform in this situation (i.e. an extreme case of individually varying availability as opposed to variability in strength of selection).

(c) *Mixed-effects models*

Mixed-effects models offer a powerful approach to modelling correlated data under the assumption that latent or unmeasured characteristics associated with individuals (i.e. random effects) might induce correlation among repeated measurements on these individuals. Mixed-effects models have been growing in popularity among ecologists in recent decades (Bennington & Thayne 1994; Bolker *et al.* 2009), naturally leading to their consideration in habitat-selection

studies (Gillies *et al.* 2006; Thomas *et al.* 2006; Aarts *et al.* 2008; Fortin *et al.* 2009; Godvik *et al.* 2009). Gillies *et al.* (2006) suggested adding random intercepts to logistic regression models to account for correlation and unequal sample sizes in telemetry studies with use–availability sampling designs. They defined ‘individuals’ as combinations of used and available points with separate samples of available points drawn from within each individual animal’s home range. The application of mixed models in this case differs from most binary repeated-measures designs in that the response ($Y = 1$ for a used point, 0 for an available point) is not a true, repeatedly observed random variable, but rather a combination of data (used points) and some randomization process (generating available points). Whereas random intercepts in typical repeated binary measures studies provide a natural way to model individual heterogeneity in the propensity to have an event, with use–availability designs, the propensity of having an event is equivalent to the probability of a point being used (rather than random), something under control of the investigator.

Similar to challenges encountered when applying GEEs to outcome-dependent sampling designs, mixed models assume the likelihood of the observed data has been correctly specified, necessitating consideration of the sampling design (e.g. Neuhaus & Jewell 1990). To understand this challenge, consider that the correlation structure of the observed data will depend on how random points are generated (i.e. if they are selected in a buffer around each telemetry location, within individual animal home ranges, or within a larger study area; Clark & Strevens 2008). Although telemetry locations are themselves likely to be serially correlated, the correlation among pairs of used points is likely to differ from the correlation among used–available points. Thus, common correlation structures (exchangeability, serial AR1 correlation) are unlikely to adequately describe clusters of used and available points. As such, traditional GLMMs are likely to mis-specify the correlation structure, leading to biased estimators of regression parameters and their sampling variance (Litiere *et al.* 2008). For example, Koper & Manseau (2009) found that model-based estimates of uncertainty from mixed-effects models fitted to use–availability data were much smaller than robust sandwich estimates based on variance-inflation techniques in which the underlying parametric assumptions were relaxed.

Although random intercepts are commonly used to account for correlation (assuming an exchangeable within-animal correlation structure), more biologically meaningful models may be constructed by allowing the selection of various habitat features to vary by individual via random coefficients (Gillies *et al.* 2006), such as assumed when fitting separate regression models to individual animals in the two-stage approach. Random coefficient models have been applied using both a logistic regression framework (Gillies *et al.* 2006; Aarts *et al.* 2008; Hebblewhite & Merrill 2008; Godvik *et al.* 2009) and a multinomial

(discrete choice) likelihood framework (Thomas *et al.* 2006; Kneib *et al.* 2007; Fortin *et al.* 2009). Similar to two-stage approaches, random coefficients allow selection parameters to vary by individual (β_i), while also providing a framework to estimate average selection parameters (i.e. $\bar{\beta} = E(\beta_i)$). Most importantly, random coefficients from either individual animal fits or mixed models offer the potential to examine functional responses in resource selection (Mysterud & Ims 1998; Beyer *et al.* 2010). Following estimation, random (individual specific) coefficients might be plotted against overall availability of that resource (e.g. at the home range level) to observe how individual responses varied with individual exposure to levels of the resource. Hebblewhite & Merrill (2008) used this approach to show that wolf (*Canis lupus*) avoidance of human activity centres varied directly (but nonlinearly) with the overall level of human activity in wolf home ranges. More recently, Godvik *et al.* (2009) found that selection of pastures by red deer (*Cervus elaphus*) decreased with increasing availability.

The use of mixed-effects models in ecology is increasing, and they are appealing because they often match the way in which ecological data are structured (e.g. individuals sampled within different population segments). Mixed models have a number of advantages compared with two-stage approaches, but also some serious disadvantages (table 1). Mixed models can effectively use data from individuals having few data points, eliminating the problem of sampling inequities among individuals. Mixed models also simultaneously fit both individual- and population-level models, efficiently pooling information across individuals to estimate common parameters (e.g. fixed-effects parameters or common-variance parameters). Mixed-model regression parameter estimators should be more precise than GEEs or two-stage approaches if parametric assumptions regarding the random effects and within individual correlation structures are met. However, again, correctly specifying the correlation structure remains challenging with use–availability sampling designs that require grouping used and available points into independent clusters (Koper & Manseau 2009). Nonlinear mixed-effects models, including GLMMs, also can be computationally challenging to fit, requiring approximate likelihood techniques, numerical integration or alternatively Bayesian implementations using Markov chain Monte Carlo methods (Pinheiro & Bates 2000; Bolker *et al.* 2009). Although most current statistical software packages can fit these models, parameter estimates can be sensitive to the choice of method (e.g. numerical integration or approximation leading to penalized quasi-likelihood; Molenberghs & Verbeke 2005). For GLMMs using a logit link, Gaussian adaptive quadrature is advised, which is a computationally intensive and sophisticated numerical integration approach that sometimes can fail to converge (Rabe-Hesketh *et al.* 2005; Gillies *et al.* 2006). In general, mixed-effects models promise interesting ecological insights, but clearly require sophisticated technical skills or help from a statistician for proper application, especially given a use–available sampling design.

5. USING MOVEMENT MODELS TO DETERMINE LOCAL AVAILABILITY

So far, we have discussed statistical methods designed to alleviate the negative effects of correlation. Now, we turn our attention to the opportunities that arise from the collection of fine-scale spatio-temporal (i.e. auto-correlated) data. In particular, high-frequency telemetry data should facilitate more realistic models of animal movement (Cagnacci *et al.* 2010), including the ability to link movement rates (and directions) to fine-scale habitat features, memory and perception and individual behavioural states (see review by Patterson *et al.* 2008). Several papers have suggested the use of models of animal movement to constrain what is considered available to an individual when studying resource selection (Hjermann 2000; Matthiopoulos 2003; Rhodes *et al.* 2005; Johnson *et al.* 2008; Moorcroft & Barnett 2008). To demonstrate, following the notation of Johnson *et al.* (2008), let s_t represent the location of an animal at time t , $H_{t-1} = \{s_1, \dots, s_{t-1}\}$ be the history of locations prior to time t , $g_u(s_t|H_{t-1})$ be the conditional probability density for the realized location at time t (conditional on its past location history), $w(s_t|H_{t-1})$ represent the resource-selection function describing habitat preferences at time t (conditional on H_{t-1}) and $g_a(s_t|H_{t-1})$ be a model for movement in the absence of habitat selection (also conditional on H_{t-1}). This last function is used to describe an availability surface from which non-random selection of resources can be inferred, via the following model:

$$g_u(s_t|H_{t-1}) = K_t^{-1} w(s_t|H_{t-1}) g_a(s_t|H_{t-1}), \quad (5.1)$$

with $K_t = \int w(u|H_{t-1}) g_a(u|H_{t-1}) du$. The approach is conceptually similar to that of Lele & Keim (2006), based on weighted-distribution theory, with the distribution of available resources determined by constraints on animal movement (Johnson *et al.* 2008). By modelling transitions that are assumed to be conditionally independent (i.e. moves are independent after conditioning on past locations), this approach also has the potential to directly account for autocorrelation in the data.

Several recent applications have studied resource selection using a movement model framework to define availability, although inference regarding w (the resource-selection function) has been approached in different ways. The simplest approach is to assume g_a is known, in which case parameters in $w(\cdot)$ can be estimated by generating available points (using g_a) for comparison to known telemetry observations in a conditional logistic regression framework (Compton *et al.* 2002; Forester *et al.* 2009). Simulations in Forester *et al.* (2009) suggest one can guard against mis-specification of g_a to some extent by allowing w to be a nonlinear function of distance (e.g. modelled using regression splines; see also Aarts *et al.* 2008). This approach was first used with circular buffers around observed locations to determine availability (e.g. Arthur *et al.* 1996), which implies a rather strange model for g_a in which the probability of an observed step length increases with distance until one reaches the radius of the buffer, before dropping to 0 (outside

the buffer) (Rhodes *et al.* 2005). Rather than assume g_a to be known, Hjermann (2000) estimated g_a using observed step lengths (r) and a model for $\log(r^2)$ as a linear function of the $\log(\text{time between observations})$ and environmental variables (with w subsequently estimated using \hat{g}_a and the MANOVA approach of Aebischer *et al.* 1993). Fortin *et al.* (2005) used the empirical distribution of step lengths and turning angles to define g_a for use with conditional logistic regression. These last two approaches assume g_a can be directly observed. However, if the model given by equation (5.1) is correct, the observed distribution of step lengths (g_u) will depend on both g_a and w . Thus, parameters associated with the resource-independent movement (in g_a) should be estimated simultaneously with selection parameters (in w), e.g. using maximum likelihood. Rhodes *et al.* (2005) used a full likelihood approach with observations regularly spaced in time, assuming that step lengths were exponentially distributed; Johnson *et al.* (2008) extended the approach to non-regularly sampled observations by assuming (continuous) movement according to a bivariate Ornstein–Uhlenbeck process (Dunn & Gibson 1977). Nielson *et al.* (2009) showed how a full likelihood approach could be extended also to allow for missed GPS fixes, by adding a model of the probability of detection as a function of covariates (conditional on a site being used) to the likelihood in equation (5.1).

A potential problem with all of these approaches is that the conditioning is done with respect to the (relatively arbitrary) sampling time interval. The time between two successive observations is informative about which other locations the animals might have visited, but did not. However, the range of accessibility implied by animal speed and the time intervening between observations is not necessarily the same as the range of perception that the animals have of their environment. More than likely, animals perceive their environment at multiple scales, particularly if memory and intention are included in our definition of perception and thus equation (5.1) is likely to be a gross oversimplification of animal movement behaviour. In an attempt to overcome these limitations, Dalziel *et al.* (2008) used neural networks to fit a more flexible model (than implied by equation (5.1)) in which each used location was determined using a nonlinear function of the distance from the previous used location, a memory function (reflecting locations where the animal had previously been located), and habitat covariates, as well as interactions among these three components. In addition, movement models in which animals switch between different behavioural modes (e.g. short- and long-range movements) may better explain autocorrelation patterns in location data (Morales *et al.* 2004), and their incorporation into habitat-selection models may lead to more informative, temporally varying, measures of habitat selection (e.g. Forester *et al.* 2009).

Regardless of the approach taken, the inclusion of movement to resource-selection models adds a non-trivial level of complexity. Estimating parameters associated with movement and habitat selection simultaneously by full maximum likelihood can be computationally demanding. Although the process

can be sped up using the fast Fourier transform with a radially symmetric resource-independent movement model (Barnett & Moorcroft 2008), these approaches generally require approximating the study area on a relatively coarse two-dimensional grid (Hjermann 2000; Rhodes *et al.* 2005; Dalziel *et al.* 2008; Johnson *et al.* 2008; Forester *et al.* 2009). Whereas a full likelihood approach can be difficult to extend to multiple individuals in a mixed effect or GEE framework, a two-stage approach could be used for population-level inference (Nielson *et al.* 2009). Alternatively, the conditional logistic regression approach (with distance as a covariate to guard against mis-specification of g_a) can be applied with robust variance estimators (e.g. ‘sandwich’ standard errors) to guard against residual within-subject correlation, while estimating population patterns related to resource selection (e.g. Fortin *et al.* 2005; Craiu *et al.* 2008; Forester *et al.* 2009).

6. DISCUSSION

Data collected using GPS telemetry are clearly at odds with the traditional notion that telemetry studies should be designed to collect independent observations (Swihart & Slade 1985). Users of GPS telemetry data must embrace the complexities of correlated data and meet the challenge of analysing the additional information available from frequently collected locations without sacrificing robust statistical inference. On the basis of our review, we suggest two relatively simple approaches as being most generally accessible and satisfying for data analysts, GEEs with a working independence assumption and two-stage approaches involving models fit to individual animals. Other correlated data methods, including GEEs with more complicated correlation structures and GLMMs deserve further exploration, but require more sophisticated analytical skills to reliably return unbiased estimators with use-availability sampling designs.

The appeal of the mixed-models approach, and the largely parallel two-stage approach, is the ability to make both subject-specific and population-level inferences (Skrondal & Rabe-Hesketh 2004). Multi-level (hierarchical) mixed-effects models also might allow explicit tests of variance structure between and within groups of animals (Skrondal & Rabe-Hesketh 2004; Hebblewhite & Merrill 2008). These kinds of insights into the underlying variation driving resource selection are not possible using GEEs or variance inflation techniques that treat correlation as a nuisance. And it is exactly this kind of approach, striving to understand the contribution of individual variation in habitat selection to fitness, that ecologists are being urged to explore (Chesson 1978; Bolnick *et al.* 2007; Gaillard *et al.* 2010). Yet, the detailed data available from GPS telemetry often are acquired from a small set of animals, such that variance estimates might be derived from a few replicates only.

Although mixed-effects models allow the estimation of population-level parameters, it is important to clarify that these parameters are not equivalent to, and thus not directly comparable to, population-level

measures of selection obtained using GEEs or traditional logistic regression models under an independence assumption (Neuhaus *et al.* 1991; Molenberghs & Verbeke 2005, §16.3; Fieberg *et al.* 2009). Specifically, parameters estimated in typical generalized linear and nonlinear mixed-effects models reflect subject-specific mean response patterns as a function of covariates, whereas parameters in standard GEEs and traditional logistic regression approaches reflect changes in the population mean response pattern as a function of covariates. For example, regression parameters in a random-intercept logistic model applied to use–non-use data describe changes in the probability of use for a particular subject as one changes that subject's covariate values, and mean regression coefficients describe response profiles for a 'typical' subject (i.e. a subject with a random effect = 0). In contrast, regression parameters estimated in the GEE framework given by equations (4.1a) and (4.1b) reflect changes in the average probability of use in the population as a function of changes in covariates. With logistic-response models, the population response pattern estimated by a GEE will be attenuated relative to the average individual response pattern estimated by random-effects models, with the degree of attenuation dependent upon the amount of between-subject variability (Zeger *et al.* 1988; Molenberghs & Verbeke 2005, §16.3; Fieberg *et al.* 2009). Thus, it is commonly recommended that the choice of method (GEE or mixed-effect model) should be predicated upon the desired parameter interpretation (i.e. if one wants to characterize subject-specific or population-averaged response patterns; Davidian & Giltinan 1995; Heagerty 1999; Fieberg *et al.* 2009). On the other hand, population mean response patterns comparable to the GEE can be recovered post hoc from fitted mixed effects regression models by integrating over the random-effects distribution numerically or by using simulation (Molenberghs & Verbeke 2005, §16.3; Aarts *et al.* 2008; Fieberg *et al.* 2009), yet this is rarely done.

Ecologists frequently think of correlation as problematic. However, much can be learned by studying the causes and consequences of correlation in ecological datasets (Boyce *et al.* 2010). For example, patterns of correlation in raw data or model residuals, summarized using autocorrelation or partial autocorrelation functions, can suggest important behavioural mechanisms responsible for generating movements and hence habitat selection, and comparisons of these patterns across space, time or different species can provide additional insight into these mechanisms (e.g. Wittemyer *et al.* 2008; Boyce *et al.* 2010). This is particularly important for long-term, periodic patterns that could be, but are currently not, accommodated by existing autoregressive structures (e.g. AR1). The ability to collect fine-scale temporal data with GPS telemetry also opens the door to new ways to study habitat selection, e.g. by using an autocorrelated movement process to identify available habitats. By specifying an appropriate model likelihood for the observed locations in terms of conditionally independent transitions, it is possible to estimate parameters that describe movement, habitat selection and

detection processes from a single model (Nielson *et al.* 2009). The main drawback to this approach is that computation difficulties may impose a relatively coarse grid for analysis (e.g. modelling movement between centroids of the grid squares). A simpler and less computationally expensive alternative is to sample available points using an assumed movement model and then analyse the data using conditional logistic regression (Craiu *et al.* 2008; Forester *et al.* 2009). This latter approach will result in biased estimators if the assumed movement model is misspecified, but one can guard against mis-specification by including distance as a predictor in the selection model, especially if the effect of distance is modelled using a flexible approach (e.g. regression splines; Aarts *et al.* 2008; Forester *et al.* 2009). A further benefit of using a movement model to define available habitats is that this approach naturally accounts for autocorrelation if transitions between locations are independent (e.g. Diggle *et al.* 1994 for a review of the use of transition models with correlated data).

We do not claim to have covered all approaches suggested for addressing correlation in the analysis of habitat selection or animal movements. For example, some authors have suggested using a fixed-intercept for each individual animal in a habitat-selection model with use–availability designs (Manly *et al.* 2002). Practically, this approach would require including a large number of parameters, one for each animal as well as a large number of interaction parameters (between individuals and habitat covariates), to allow selection for a particular resource to vary among individuals. Critically, this approach restricts the scope of inference to the specific study animals rather than to the population from which they were sampled, and we have not seen this approach in widespread use. Thus, we chose to focus on what we considered the most general and flexible approaches here.

We conclude that correlation in general, and autocorrelation in particular, offers both challenges and opportunities for data analysts, particularly for GPS telemetry data and resource-selection studies. Our review indicates that further research into the use of GLMMs, and GEEs with non-independence working correlation structures is needed to clarify outstanding issues. The primary challenge with the use–available design remains that correlation patterns among used locations will depend on biology, whereas correlation among available locations depends more on the spatial variance of the covariates in question (e.g. Boyce *et al.* 2002). Depending on how sensitive GLMM models and GEEs are to mis-specification of correlation structure, this problem could be anything from trivial to severe, demanding additional research attention and sophisticated analytical skills. Nevertheless, consideration of hierarchical correlation structures using these approaches seems a particularly promising area of enquiry. Integration of animal movement and resource-selection models promises novel ecological insights, but remains computationally intensive and requires strong analytical skills, restricting these approaches at present to more theoretical rather than applied research problems. Given the complexity of the task, and uncertainty in the process of specifying

correlation structures, we suggest adopting a conservative approach, with biological guidance and cross-validation playing key roles in determining the importance of various predictors in habitat-based regression models (Boyce *et al.* 2002). The more generally accessible and transparent approaches—variance inflation factors and two-stage approaches—thus offer a pragmatic approach to accounting for correlation in habitat-selection studies. Regardless of the approach, we would be remiss not to recommend thorough validation of final habitat-selection models, because, in applied conservation, an assessment of the validity of predictions often is the most important consideration.

We thank the Edmund Mach Foundation for funding the GPS data analysis workshop in Trento, Italy, Fall 2008. We thank D. Haydon, J. Morales, D. Thomas and an anonymous reviewer for their critical reviews and G. Aarts, T. Duschesne, for helpful comments on a previous draft.

REFERENCES

- Aarts, G., MacKenzie, M., McConnell, B., Fedak, M. & Matthiopoulos, J. 2008 Estimating space-use and habitat preference from wildlife telemetry data. *Ecography* **31**, 140–160. (doi:10.1111/j.2007.0906-7590.05236.x)
- Aebischer, N. J., Robertson, P. A. & Kenward, R. E. 1993 Compositional analysis of habitat use from animal radio-tracking data. *Ecology* **74**, 1313–1325. (doi:10.2307/1940062)
- Arthur, S. M., Manly, B. F. J., McDonald, L. L. & Garner, G. W. 1996 Assessing habitat selection when availability changes. *Ecology* **77**, 215–227. (doi:10.2307/2265671)
- Barnett, A. H. & Moorcroft, P. R. 2008 Analytic steady-state space use patterns and rapid computations in mechanistic home range analysis. *J. Math. Biol.* **57**, 139–159. (doi:10.1007/s00285-007-0149-8)
- Bennington, C. C. & Thayne, W. V. 1994 Use and misuse of mixed-model analysis of variance in ecological studies. *Ecology* **75**, 717–722. (doi:10.2307/1941729)
- Beyer, H. L., Haydon, D. T., Morales, J. M., Frair, J. L., Hebblewhite, M., Mitchell, M. & Matthiopoulos, J. 2010 The interpretation of habitat preference metrics under use–availability designs. *Phil. Trans. R. Soc. B* **365**, 2245–2254. (doi:10.1098/rstb.2010.0083)
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. & White, S. S. 2009 Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127–135. (doi:10.1016/j.tree.2008.10.008)
- Bolnick, D. I., Svanbäck, R., Araujo, M. S. & Persson, L. 2007 Comparative support for the niche variation hypothesis that more generalized populations also are more heterogeneous. *Proc. Natl Acad. Sci. USA* **104**, 10 075–10 079. (doi:10.1073/pnas.0703743104)
- Börger, L., Franconi, N., De Michele, G., Gantz, A., Meschi, F., Manica, A., Lovari, S. & Coulson, T. 2006a Effects of sampling regime on the mean and variance of home range estimates. *J. Anim. Ecol.* **75**, 1393–1405. (doi:10.1111/j.1365-2656.2006.01164.x)
- Börger, L., Franconi, N., Ferretti, N., Meschi, F., De Michele, G., Gantz, A. & Coulson, T. 2006b An integrated approach to identify spatio-temporal and individual-level determinants of animal home range size. *Am. Nat.* **168**, 471–485. (doi:10.1086/507883)
- Boyce, M. S. 2006 Scale and resource selection functions. *Divers. Distrib.* **12**, 269–276. (doi:10.1111/j.1366-9516.2006.00243.x)
- Boyce, M. S., Vernier, P. R., Nielsen, S. E. & Schmiegelow, F. K. A. 2002 Evaluating resource selection functions. *Ecol. Model.* **157**, 281–300. (doi:10.1016/S0304-3800(02)00200-4)
- Boyce, M. S., Pitt, J., Northrup, J. M., Morehouse, A. T., Knopff, K. H., Cristescu, B. & Stenhouse, G. B. 2010 Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data. *Phil. Trans. R. Soc. B* **365**, 2213–2219. (doi:10.1098/rstb.2010.0080)
- Cagnacci, F., Boitani, L., Powell, R. A. & Boyce, M. S. 2010 Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Phil. Trans. R. Soc. B* **365**, 2157–2162. (doi:10.1098/rstb.2010.0107)
- Calenge, C. & Dufour, A. B. 2006 Eigenanalysis of selection ratios from animal radio-tracking data. *Ecology* **87**, 2349–2355. (doi:10.1890/0012-9658(2006)87[2349:EOSRFA]2.0.CO;2)
- Chesson, J. 1978 Measuring preference in selective predation. *Ecology* **59**, 211–215. (doi:10.2307/1936364)
- Christ, A., Ver Hoef, J. & Zimmermon, D. L. 2008 An animal movement model incorporating home range and habitat selection. *Environ. Ecol. Stat.* **15**, 27–38. (doi:10.1007/S10651-007-0036-x)
- Clark, R. G. & Strevens, T. C. 2008 Design and analysis of clustered, unmatched resource selection studies. *J. R. Stat. Soc. Ser. C* **57**, 535–551. (doi:10.1111/j.1467-9876.2008.00629.x)
- Compton, B. W., Rhymer, J. M. & McCollough, M. 2002 Habitat selection by wood turtles (*Clemmys insculpta*): an application of paired logistic regression. *Ecology* **83**, 833–843. (doi:10.1890/0012-9658(2002)083[0833:HSBWTC]2.0.CO;2)
- Cooper, A. B. & Millsbaugh, J. J. 1999 The application of discrete choice models to wildlife resource selection studies. *Ecology* **80**, 566–575. (doi:10.1890/0012-9658(1999)080[0566:TAODCM]2.0.CO;2)
- Craiu, R. V., Duchesne, T. & Fortin, D. 2008 Inference methods for the conditional logistic regression model with longitudinal data. *Biometr. J.* **50**, 97–109. (doi:10.1002/bimj.200610379)
- Cushman, S. A., Chase, M. & Griffin, C. 2005 Elephants in space and time. *Oikos* **109**, 331–341. (doi:10.1111/j.0030-1299.2005.13538.x)
- Dalziel, B. D., Morales, J. M. & Fryxell, J. M. 2008 Fitting probability distributions to animal movement trajectories: using artificial neural networks to link distance, resources, and memory. *Am. Nat.* **172**, 248–258. (doi:10.1086/589448)
- Dasgupta, N. & Alldredge, J. R. 1998 A multivariate chi-square analysis of resource selection data. *J. Agric. Biol. Environ. Stat.* **3**, 323–334. (doi:10.2307/1400586)
- Davidian, M. & Giltinan, D. M. 1995 *Nonlinear models for repeated measurement data*. New York, NY: Chapman and Hall.
- Diggle, P. J., Liang, K. Y. & Zeger, S. L. 1994 *Analysis of longitudinal data*. London, UK: Oxford University Press.
- Diniz-Filho, J. A. F., Bini, L. M. & Hawkins, B. A. 2003 Spatial autocorrelation and red herrings in geographical ecology. *Glob. Ecol. Biogeogr.* **12**, 53–64. (doi:10.1046/j.1466-822X.2003.00322.x)
- Dunn, J. & Gibson, P. 1977 Analysis of radio-telemetry data in studies of home range. *Biometrics* **33**, 85–101. (doi:10.2307/2529305)
- Fancy, S. G. & White, R. G. 1987 Energy expenditures for location by barren-ground caribou. *Can. J. Zool.* **65**, 122–128. (doi:10.1139/z87-018)

- Fieberg, J. 2007a Kernel density estimators of home range: smoothing and the autocorrelation red herring. *Ecology* **88**, 1059–1066. (doi:10.1890/06-0930)
- Fieberg, J. 2007b Utilization distribution estimation with weighted kernel density estimators. *J. Wildl. Manage.* **71**, 1669–1675. (doi:10.2193/2006-370)
- Fieberg, J., Kuehn, D. W. & DelGiudice, G. D. 2008 Understanding variations in autumn migration of northern white-tailed deer by long-term study. *J. Mammal.* **89**, 1529–1539. (doi:10.1644/07-MAMM-A-277.1)
- Fieberg, J., Rieger, R. H., Zicus, M. C. & Schildcrout, J. S. 2009 Regression modelling of correlated data in ecology: subject-specific and population averaged response patterns. *J. Appl. Ecol.* **46**, 1018–1025. (doi:10.1111/j.1365-2664.2009.01692.x)
- Forester, J. D., Im, H. K. & Rathouz, P. J. 2009 Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology* **90**, 3554–3565. (doi:10.1890/08-0874.1)
- Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T. & Mao, J. S. 2005 Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. *Ecology* **86**, 1320–1330. (doi:10.1890/04-0953)
- Fortin, D., Fortin, M.-E., Beyer, H. L., Duchesne, T. & Courant, S. 2009 Group-size-mediated habitat selection and group fusion-fission dynamics of bison under predation risk. *Ecology* **90**, 2480–2490. (doi:10.1890/08-0345.1)
- Frair, J. L., Nielsen, S. E., Merrill, E. H., Lele, S. R., Boyce, M. S., Munro, R. H. M., Stenhouse, G. B. & Beyer, H. L. 2004 Removing GPS collar bias in habitat selection studies. *J. Appl. Ecol.* **41**, 201–212. (doi:10.1111/j.0021-8901.2004.00902.x)
- Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J. & Pedrotti, L. 2010 Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Phil. Trans. R. Soc. B* **365**, 2187–2200. (doi:10.1098/rstb.2010.0084)
- Fryxell, J. M., Hazell, M., Börger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T. & Rosatte, R. C. 2008 Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proc. Natl Acad. Sci. USA* **105**, 19 114–19 119. (doi:10.1073/pnas.0801737105)
- Gaillard, J.-M., Hebblewhite, M., Loison, A., Fuller, M., Powell, R., Basille, M. & Van Moorter, B. 2010 Habitat-performance relationships: finding the right metric at a given spatial scale. *Phil. Trans. R. Soc. B* **365**, 2255–2265. (doi:10.1098/rstb.2010.0085)
- Gillies, C. S., Hebblewhite, M., Nielsen, S. E., Krawchuk, M. A., Aldridge, C. L., Frair, J. L., Saher, J., Stevens, C. E. & Jerde, C. L. 2006 Application of random effects to the study of resource selection by animals. *J. Anim. Ecol.* **75**, 887–898. (doi:10.1111/j.1365-2656.2006.01106.x)
- Glenn, E. M., Hansen, M. C. & Anthony, R. G. 2004 Spotted owl home range and habitat use in young forests of western Oregon. *J. Wildl. Manage.* **68**, 33–50. (doi:10.2193/0022-541X(2004)068[0033:SOH AHUJ]2.0.CO;2)
- Godvik, I. M. R., Low, L. E., Vik, J. O., Veiberg, V., Langvatn, R. & Mysterud, A. 2009 Temporal scales, trade-offs, and functional responses in red deer habitat selection. *Ecology* **90**, 699–710. (doi:10.1890/08-0576.1)
- Hansteen, T. L., Andreassen, H. P. & Ims, R. A. 1997 Effects of spatiotemporal scale on autocorrelation and home range estimators. *J. Wildl. Manage.* **61**, 280–290. (doi:10.2307/3802583)
- Haydon, D. T., Morales, J. M., Yott, A., Jenkins, D. A., Rosatte, R. & Fryxell, J. M. 2008 Socially informed random walks: incorporating group dynamics into models of population spread and growth. *Proc. R. Soc. B* **275**, 1101–1109. (doi:10.1098/rspb.2007.1688)
- Heagerty, P. J. 1999 Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* **55**, 688–698. (doi:10.1111/j.0006-341X.1999.00688.x)
- Hebblewhite, M. & Haydon, D. T. 2010 Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Phil. Trans. R. Soc. B* **365**, 2303–2312. (doi:10.1098/rstb.2010.0087)
- Hebblewhite, M. & Merrill, E. H. 2008 Modelling wildlife-human relationships for social species with mixed-effects resource selection models. *J. Appl. Ecol.* **45**, 834–844. (doi:10.1111/j.1365-2664.2008.01466.x)
- Hjermann, D. O. 2000 Analyzing habitat selection in animals without well-defined home ranges. *Ecology* **81**, 1462–1468. (doi:10.1890/0012-9658(2000)081[1462:AHSLAW]2.0.CO;2)
- Horne, J. S., Garton, E. O. & Sager-Fradkin, K. A. 2007 Correcting home-range models for observation bias. *J. Wildl. Manage.* **71**, 996–1001. (doi:10.2193/2005-678)
- Johnson, D. H. 1980 The comparison of usage and availability measurements for evaluating resource preference. *Ecology* **61**, 65–71. (doi:10.2307/1937156)
- Johnson, D. H. 1999 The insignificance of statistical significance testing. *J. Wildl. Manage.* **63**, 763–772. (doi:10.2307/3802789)
- Johnson, C. J., Parker, K. L., Heard, D. C. & Gillingham, M. P. 2002 Movement parameters of ungulates and scale-specific responses to the environment. *J. Anim. Ecol.* **71**, 225–235. (doi:10.1046/j.1365-2656.2002.00595.x)
- Johnson, C. J., Nielsen, S. E., Merrill, E. H., McDonald, T. L. & Boyce, M. S. 2006 Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. *J. Wildl. Manage.* **70**, 347–357. (doi:10.2193/0022-541X(2006)70[347:RSFBOU]2.0.CO;2)
- Johnson, D. S., Thomas, D. L., Ver Hoef, J. M. & Christ, A. 2008 A general framework for the analysis of animal resource selection from telemetry data. *Biometrics* **64**, 968–976. (doi:10.1111/j.1541-0420.2007.00943.x)
- Keating, K. A. & Cherry, S. 2004 Use and interpretation of logistic regression in habitat selection studies. *J. Wildl. Manage.* **68**, 774–789. (doi:10.2193/0022-541X(2004)068[0774:UAIOLR]2.0.CO;2)
- Kneib, T., Knauer, F. & Küchenhoff, H. 2007 *A general approach for the analysis of habitat selection*. Technical Report Number 001 2007, Department of Statistics, University of Munich, Germany. See <http://epub.ub.uni-muenchen.de/2052/>.
- Koper, N. & Manseau, M. L. 2009 Generalized estimating equations and generalized linear mixed-effects models for modeling resource selection. *J. Appl. Ecol.* **46**, 590–599. (doi:10.1111/j.1365-2664.2009.01642.x)
- Laird, H. 1987 Estimating the location of the focal center in red squirrel home ranges. *Ecology* **68**, 1092–1101. (doi:10.2307/1938381)
- Lele, S. R. 2009 A new method for estimation of resource selection probability function. *J. Wildl. Manage.* **73**, 122–127. (doi:10.2193/2007-535)
- Lele, S. R. & Keim, J. L. 2006 Weighted distributions and estimation of resource selection probability functions. *Ecology* **87**, 3021–3028. (doi:10.1890/0012-9658(2006)87[3021:WDAEOR]2.0.CO;2)
- Litiere, S., Alonso, A. & Molenberghs, G. 2008 The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Stat. Med.* **27**, 3125–3144. (doi:10.1002/sim.3157)
- Manly, B. F. J., McDonald, L. L., Thomas, D., McDonald, T. L. & Erickson, W. 2002 *Resource selection by animals:*

- statistical design and analysis for field studies. Boston, MA: Kluwer Academic Publishers.
- Matthiopoulos, J. 2003 The use of space by animals as a function of accessibility and preference. *Ecol. Model.* **159**, 239–268. (doi:10.1016/S0304-3800(02)00293-4)
- McNay, R. S. & Bunnell, F. L. 1994 Characterizing independence of observations in movements of Columbian black-tailed deer. *J. Wildl. Manage.* **58**, 422–429. (doi:10.2307/3809312)
- Molenberghs, G. & Verbeke, G. 2005 *Models for discrete longitudinal data*. New York, NY: Springer.
- Moorcroft, P. R. & Barnett, A. 2008 Mechanistic home range models and resource selection analysis: a reconciliation and unification. *Ecology* **89**, 1112–1119. (doi:10.1890/06-1985.1)
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E. & Fryxell, J. M. 2004 Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology* **85**, 2436–2445. (doi:10.1890/03-0269)
- Murtaugh, P. A. 2007 Simplicity and complexity in ecological data analysis. *Ecology* **88**, 56–62. (doi:10.1890/0012-9658(2007)88[56:SACIED]2.0.CO;2)
- Mysterud, A. & Ims, R. A. 1998 Functional responses in habitat use: availability influences relative use in trade-off situations. *Ecology* **79**, 1435–1441. (doi:10.1890/0012-9658(1998)079[1435:FRIHUA]2.0.CO;2)
- Neuhauser, J. M. & Jewell, N. P. 1990 The effect of retrospective sampling on binary regression models for clustered data. *Biometrics* **46**, 977–990. (doi:10.2307/2532442)
- Neuhauser, J. M., Kalbfleisch, J. D. & Hauck, W. 1991 A comparison of cluster-specific and population-averaged approaches for analyzing correlated data. *Int. Stat. Rev.* **59**, 25–35. (doi:10.2307/1403572)
- Newey, W. K. & West, K. D. 1987 A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708. (doi:10.2307/1913610)
- Nielsen, S. E., Boyce, M. S., Stenhouse, G. B. & Munro, R. H. M. 2002 Modeling grizzly bear habitats in the Yellowstone ecosystem of Alberta: taking autocorrelation seriously. *Ursus* **13**, 45–56.
- Nielson, R. M., Manly, B. F. J., McDonald, L. L., Sawyer, H. & McDonald, T. 2009 Estimating habitat selection when GPS fix success is less than 100%. *Ecology* **77**, 215–227. (doi:10.1890/08-1562.1)
- Otis, D. L. & White, G. C. 1999 Autocorrelation of location estimates and the analysis of radiotracking data. *J. Wildl. Manage.* **63**, 1039–1044. (doi:10.2307/3802819)
- Pan, W. 2001 Akaike's information criterion in generalized estimating equations. *Biometrics* **57**, 120–125. (doi:10.1111/j.0006-341X.2001.00120.x)
- Park, E. & Kim, Y. 2004 Analysis of longitudinal data in case-control studies. *Biometrika* **91**, 321–330. (doi:10.1093/biomet/91.2.321)
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. 2008 State-space models of individual animal movement. *Trends Ecol. Evol.* **23**, 87–94. (doi:10.1016/j.tree.2007.10.009)
- Pinheiro, J. C. & Bates, D. M. 2000 *Mixed-effects models in S and S-Plus*. New York, NY: Springer.
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. 2005 Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *J. Econ.* **128**, 301–323. (doi:10.1016/j.jeconom.2004.08.017)
- Rhodes, J. R., McAlpine, C. A., Lunney, D. & Possingham, H. P. 2005 A spatially explicit habitat selection model incorporating home range behavior. *Ecology* **86**, 1199–1205. (doi:10.1890/04-0912)
- Rooney, S. M., Wolfe, A. & Hayden, T. J. 1998 Autocorrelated data in telemetry studies: time to independence and the problem of behavioural effects. *Mammal Rev.* **28**, 89–98. (doi:10.1046/j.1365-2907.1998.00028.x)
- Salvatori, V., Skidmore, A. K., Corsi, F. & van der Meer, F. 1999 Estimating temporal independence of radio-telemetry data on animal activity. *J. Theor. Biol.* **198**, 567–574. (doi:10.1006/jtbi.1999.0936)
- Sawyer, H., Nielsen, R. M., Lindzey, F. & McDonald, L. L. 2006 Winter habitat selection of mule deer before and during development of a natural gas field. *J. Wildl. Manage.* **70**, 396–403. (doi:10.2193/0022-541X(2006)70[396:WHSOMD]2.0.CO;2)
- Skrondal, A. & Rabe-Hesketh, S. 2004 *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. New York, NY: Chapman and Hall.
- Swihart, R. K. & Slade, N. 1985 Testing for independence of observations in animal movements. *Ecology* **66**, 1176–1184. (doi:10.2307/1939170)
- Thomas, D. L. & Taylor, E. J. 2006 Study designs and tests for comparing resource use and availability II. *J. Wildl. Manage.* **70**, 324–336. (doi:10.2193/0022-541X(2006)70[324:SDATFC]2.0.CO;2)
- Thomas, D. L., Johnson, D. & Griffith, B. 2006 A Bayesian random effects discrete-choice model for resource selection: population-level selection inference. *J. Wildl. Manage.* **70**, 404–412. (doi:10.2193/0022-541X(2006)70[404:ABREDM]2.0.CO;2)
- Thompson, S. K. 2002 *Sampling*, 2nd edn. New York, NY: Wiley.
- Verwajen, D. & Van Damme, R. 2008 Foraging mode and locomotor capacities in Lacertidae. *Amphibia Reptilia* **29**, 197–206. (doi:10.1163/156853808784124965)
- White, G. C. & Garrott, R. A. 1990 *Analysis of wildlife radio-tracking data*. San Diego, CA: Academic Press.
- Williams, R. L. 2000 A note on robust-variance estimation for cluster-correlated data. *Biometrics* **56**, 645–646. (doi:10.1111/j.0006-341X.2000.00645.x)
- Wittemyer, G., Polansky, L., Douglas-Hamilton, I. & Getz, W. M. 2008 Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. *Proc. Natl Acad. Sci. USA* **105**, 19 108–19 113. (doi:10.1073/pnas.0801744105)
- Zeger, S. L., Liang, K.-Y. & Albert, P. S. 1988 Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–1060. (doi:10.2307/2531734)