Ecology, 94(7), 2013, pp. 1456–1463

Reports

Practical guidance on characterizing availability in resource selection functions under a use–availability design

Joseph M. Northrup,[1,5] Mevin B. Hooten,[1,2,3] Charles R. Anderson, Jr.,[4] and George Wittemyer[1]

[1]*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, 1474 Campus Delivery, Fort Collins, Colorado 80523 USA*

[2]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, 1474 Campus Delivery, Fort Collins, Colorado 80523 USA*

[3]*Colorado State University, Department of Statistics, Colorado State University, 1474 Campus Delivery, Fort Collins, Colorado 80523 USA*

[4]*Mammals Research Section Colorado Parks and Wildlife, 711 Independent Avenue, Grand Junction, Colorado 81505 USA*

Abstract. Habitat selection is a fundamental aspect of animal ecology, the understanding of which is critical to management and conservation. Global positioning system data from animals allow fine-scale assessments of habitat selection and typically are analyzed in a use–availability framework, whereby animal locations are contrasted with random locations (the availability sample). Although most use–availability methods are in fact spatial point process models, they often are fit using logistic regression. This framework offers numerous methodological challenges, for which the literature provides little guidance. Specifically, the size and spatial extent of the availability sample influences coefficient estimates potentially causing interpretational bias. We examined the influence of availability on statistical inference through simulations and analysis of serially correlated mule deer GPS data. Bias in estimates arose from incorrectly assessing and sampling the spatial extent of availability. Spatial autocorrelation in covariates, which is common for landscape characteristics, exacerbated the error in availability sampling leading to increased bias. These results have strong implications for habitat selection analyses using GPS data, which are increasingly prevalent in the literature. We recommend that researchers assess the sensitivity of their results to their availability sample and, where bias is likely, take care with interpretations and use cross validation to assess robustness.

Key words: autocorrelation; GPS radio telemetry; resource selection function, RSF; spatial point process; species distribution model; use–availability data; wildlife.

Introduction

Habitat selection is a behavioral process by which animals choose the most suitable locations in order to maximize fitness (Fretwell and Lucas 1969). Understanding the selection process can provide insight into population regulation, species interactions, and predator–prey dynamics (Morris 2003) and thus is fundamental to animal ecology. With advancements in global positioning systems (GPS), radio telemetry, and geographic information systems (GIS), the data required to examine habitat selection patterns of free-ranging animals are increasingly available, spurring a proliferation of recent studies on this topic.

The most common method for examining habitat selection patterns from GPS radio collar data is the resource selection function (RSF, see Table 1 [Manly et al. 2002, Johnson et al. 2006]). Resource selection functions typically are fit in a use–availability framework, whereby environmental covariates (e.g., elevation) at the locations where the animal was present (the used locations) are contrasted with covariates at random locations taken from an area deemed to be available for selection (the availability sample [Manly et al. 2002, Johnson et al. 2006]). Such methods are inherently based on models for spatial point processes (as are many species distribution models; e.g., Warton and Shepherd [2010]), however logistic regression, which asymptotically approximates a point process model (Johnson et al. 2006, Aarts et al. 2012), typically is used to estimate coefficients (but see Baddeley and Turner [2000], Lele and Keim [2006], Johnson et al. [2008], and Aarts et al. [2012] for alternate approaches). Logistic regression allows researchers to easily obtain inference on selection or avoidance of covariates and to generate maps for use in subsequent analysis (Boyce and McDonald 1999). Such methods have been used to examine numerous ecological processes and address important management questions, including the interplay between habitat and dispersal (Shafer et al. 2012), the presence of ecological traps (Northrup et al. 2012), and functional responses in wildlife interactions with anthropogenic development

Manuscript received 2 October 2012; revised 7 March 2013; accepted 12 March 2013. Corresponding Editor: B. D. Inouye.
[5] E-mail: joe.northrup@colostate.edu

Table 1. Terms used in resource selection function (RSF) analysis and their definitions, adapted from Manly et al. (2002), Johnson et al. (2006), Lele and Keim (2006), Beyer et al. (2010), and Aarts et al. (2012).

Term	Definition
Habitat	The set of biotic and abiotic factors characterizing the space an animal inhabits; in RSF analysis, a set of environmental covariates at discrete locations in space, meant to approximate these factors
Use	The exploitation of habitat to meet a real or perceived biological need; in RSF analysis, the presence of an animal at a location
Used distribution	The probability density functions for all animal locations over a specific time period; $f^{\mathrm{U}}(\mathbf{x})$ in the weighted distribution (Eq. 1)
Used sample	A measured subset of the used distribution
Availability	The amount and configuration of habitat over an area of interest
Availability distribution	The probability density function of all locations available to be selected over an area of interest; $f^{\mathrm{A}}(\mathbf{x})$ in the weighted distribution (Eq. 1)
Availability sample	A measured, user-defined subset of the availability distribution (used to approximate the integral in the weighted distribution; Eq. 1)
Selection	Use disproportionate to availability
Resource selection function (RSF)	Any function proportional to the probability of selection of habitat; $w(\mathbf{x}'\boldsymbol{\beta})$ in the weighted distribution.

Notes: In the definitions above, $\mathbf{x}$ is a vector of environmental covariates, with a corresponding vector of coefficients, $\boldsymbol{\beta}$.

(Hebblewhite and Merrill 2008, Matthiopoulos et al. 2011).

The relative ease of fitting RSFs has made them popular in animal ecology. However, these methods offer a number of methodological challenges (e.g., Aarts et al. 2008). In particular, the size and spatial extent of the availability sample can significantly influence coefficient estimates and subsequent inference (Boyce et al. 2003, Boyce 2006, Warton and Shepherd 2010). Despite this fact, there is a striking lack of robust guidance for choosing the availability sample and most applied studies likely are incorrectly sampling availability (Warton and Shepherd 2010). Here we illustrate the influence of the availability sample size and spatial extent on inference from RSFs under the most commonly used sampling designs, with the goal of offering robust guidance for practitioners. We first review pertinent literature regarding the availability sample and summarize recognized issues. We then illustrate the influence of the availability sample on coefficient estimates through simulations and an empirical analysis of GPS data from mule deer (*Odocoileus hemionus*), and provide guidance on how best to implement robust RSFs.

The use–availability framework and important considerations.—For RSFs fit under a use–availability design, the used locations are a realization from the used distribution $f^{\mathrm{U}}(\mathbf{x})$ (see Table 1), which can be written as a weighted version of the availability distribution $f^{\mathrm{A}}(\mathbf{x})$ (Johnson et al. 2006, Lele and Keim 2006, Hooten et al. 2013):

$$f^{\mathrm{U}}(\mathbf{x}) = \frac{w(\mathbf{x}'\boldsymbol{\beta})f^{\mathrm{A}}(\mathbf{x})}{\displaystyle\int w(\mathbf{x}'\boldsymbol{\beta})f^{\mathrm{A}}(\mathbf{x})d\mathbf{x}} \tag{1}$$

where $\mathbf{x}$ is a vector of environmental covariates, with a corresponding vector of coefficients, $\boldsymbol{\beta}$. In this weighted distribution (Eq. 1), $w(\mathbf{x}'\boldsymbol{\beta})$ is the RSF, and can be interpreted as how the animal selects habitat from $f^{\mathrm{A}}(\mathbf{x})$. The RSF can take a number of functional forms (e.g., probit, logistic [Lele 2009]); however Johnson et al. (2006) prove that, provided $w(\mathbf{x}'\boldsymbol{\beta})$ takes the exponential form [i.e. $w(\mathbf{x}'\boldsymbol{\beta}) = e^{\mathbf{x}'\boldsymbol{\beta}}$], logistic regression can be used to obtain unbiased estimates of $\boldsymbol{\beta}$. When using logistic regression, the RSF approximates a spatial point process model and can be interpreted as the expected number of used locations per unit area (Warton and Shepherd 2010, Aarts et al. 2012). Thus, Poisson regression also can be used to obtain unbiased estimates of $\boldsymbol{\beta}$ in Eq. 1, with the dependent variable being the number of used locations within a discrete spatial unit. The intercept in Poisson regression scales the RSF to the number of used locations, but as with logistic regression has no biological meaning (W. Fithian and T. Hastie, *unpublished manuscript*).

The purpose of the availability sample is to approximate the integral in the denominator of Eq. 1, and if this sample is too small then the point process model itself is poorly approximated and any inference drawn from the resulting coefficients is incorrect. In determining the size of the availability sample, it is the ratio of used to available locations that is of paramount importance, with larger ratios providing worse approximations (W. Fithian and T. Hastie, *unpublished manuscript*). While these factors imply that the availability sample should be as large as possible, there is a trade-off between size and computation time, with little guidance on optimal sample size. Manly et al. (2002) suggest sensitivity analyses be conducted to determine the sample size. Several studies have suggested that a minimum of 10 000 locations are required (Lele and Keim 2006, Lele 2009, Barbet-Massin et al. 2012), and Aarts et al. (2012) report that samples of 10 000 locations provide accurate estimates for data simulated from a single covariate. Both Warton and Shepherd (2010) and Aarts et al. (2012) also indicate that regular (as opposed to random) sampling of the availability space can reduce the sample needed to approximate the point process model. Likewise, W. Fithian and T. Hastie (*unpublished manuscript*) show that weighting the availability sample by an arbitrarily large value can accomplish the same. In addition, Barbet-Massin et al. (2012) suggest that the modeling framework (e.g., GLM, GAM, or machine learning methods) can influence the number of availability points needed. Despite these suggestions, ad hoc approaches to choosing the size of the availability sample appear to be the norm (e.g., 1 point/km^2 [Hebblewhite and Merrill 2008]), and likely under-sample availability, thus poorly approximating the integral in Eq. 1 (Warton and Shepherd 2010). However, it is unclear how such under-sampling influences coefficient estimates in a real-world example where researchers assess multiple correlated environmental factors across large landscapes and for multiple individuals.

As with the sample size, the spatial extent over which availability is drawn can substantially influence coefficient estimates and subsequent inference (Johnson 1980, Garshelis 2000, Boyce et al. 2003, Beyer et al. 2010). This extent depends on the scale of inference desired (i.e., first-, second-, third-, or fourth-order selection [Johnson 1980]), and the availability sample must match the scale of inference or there could be strong biases in the interpretation of coefficient estimates (Beyer et al. 2010). This issue has rarely been addressed explicitly from a methodological perspective (but see Beyer et al. 2010). Instead studies typically compare used locations to availability samples drawn across differing spatial extents (Johnson 1980, Boyce et al. 2003, Boyce 2006), and interpret differences in coefficients as the behavioral response of the animal to habitats at different scales. In most GPS studies, however, animal locations are not independent from one another (i.e., they are autocorrelated), which causes difficulties in inference from RSFs. With the exception of Johnson et al. (2008), the issue of autocorrelation in habitat selection studies only has been addressed in terms of model assumptions (i.e., independence of errors [Fieberg et al. 2010]). When animal locations are sampled at high resolution, the habitat available to be selected also is autocorrelated (Hooten et al. 2013), an issue that has been largely overlooked. Despite this autocorrelation, inference can be obtained at the desired scale through thinning of autocorrelated data, or accounting for autocorrelation explicitly in the model (Hooten et al. 2013). Without proper correction or thinning, comparing the used locations to a misinterpreted availability sample (i.e., areas that were not accessible to the animal) complicates the interpretation of coefficients. These coefficients likely represent some mix of a behavioral response to the environmental factors, and noise induced by the distribution of the covariates on the landscape and the movement of the animal (Beyer et al. 2010). The interaction between the spatial extent from which availability is drawn, autocorrelation in landscape covariates, and the availability sample size is of critical importance and has not been assessed.

Methods

We examined the influence of the size and spatial extent of the availability sample on RSF coefficient estimates. Using simulations, we first examined the most common scale of inference in the applied literature: selection of habitat within the home range (third-order selection [Johnson 1980]). Next we examined selection of habitat from within a buffer around each used location (third/fourth-order selection), again using simulation. We also examined the consequences of inaccurately assessing availability in both cases. Finally we examined these scales of selection in an analysis of GPS data from mule deer in the Piceance Basin, Colorado, USA. All analyses herein were conducted in the R statistical software (R Development Core Team 2012).

Third-order simulation.—We simulated used animal locations as an inhomogeneous Poisson spatial point process (IPP) on a true landscape in the Piceance Basin in northwestern Colorado. Locations were simulated as a function of a single environmental covariate (elevation) with $w(\mathbf{x}'\boldsymbol{\beta}) = e^{\beta_0+\beta_1 x}$ across a subset of the study area (here $\beta_1 = 2$, and we varied β_0 to achieve desired used sample sizes). We then drew 1 000 000 random locations across (1) the same spatial extent as the used locations (hereafter the "matched sample") and (2) an area greater than that from which use was simulated (hereafter the "mismatched sample"). The mismatched sample simulates a situation in which what was truly available to be selected by the animal is inaccurately assessed by the researcher. From the larger availability samples, we randomly drew smaller samples ranging in size from 100 to 50 000 (100, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000, 30 000, and 50 000) and fit RSFs using logistic regression. We repeated this process 500 times for three different ratios of used to available locations (80, 650, and 3500 used samples), and calculated the expectation of the coefficient estimator [$E(\hat{\beta}_1)$] and the 95% simulation envelope.

To assess the interaction between landscape heterogeneity, availability sample size, and spatial extent, we repeated the above analyses on simulated landscapes with varying levels of autocorrelation for a binary and a continuous covariate (see Appendix A). For the binary covariate, we varied the proportion of the landscape composed of that covariate. We simulated use and fit models as above (with $\beta_1 = 0.5$) for matched and mismatched availability. We calculated the coefficient estimator and 95% simulation envelope for two ratios of use to availability (600 and 6000 used samples, though only the former for the binary covariate).

Third/fourth-order simulation.—A common approach to characterizing availability in RSFs entails delineating a buffer around each used location, with the buffer radius determined by the movement of the animal (e.g., the mean Euclidean displacement between locations

[Boyce et al. 2003]), and assessing availability within each buffer. In this case, Eq. 1 is then modified such that

$$f_i^{\mathrm{U}}(\mathbf{x}) = \frac{w(\mathbf{x}'\boldsymbol{\beta})f_i^{\mathrm{A}}(\mathbf{x})}{\int w(\mathbf{x}'\boldsymbol{\beta})f_i^{\mathrm{A}}(\mathbf{x})d\mathbf{x}} \qquad (2)$$

where f_i^{A} is the availability distribution for point i. RSFs are fit using conditional logistic regression, with the used points matched to the available points within their respective buffers. To examine the influence of the size of the availability sample on coefficients estimated with this approach, we randomly placed 500 buffers with a 100 m radius (size was chosen arbitrarily) on landscapes simulated with different levels of autocorrelation. We then simulated use as an IPP within each buffer with $w(\mathbf{x}'\boldsymbol{\beta}) = e^{\beta_0+\beta_1 x}$ (a single point was then randomly selected to act as the used location). We then drew 1000 random locations within each buffer. From this sample we drew availability samples ranging from 1 to 500 points, repeating this process 500 times for each sample size, from which the expectation of the coefficient estimator and 95% simulation envelope were calculated. We repeated this process for a mismatched availability sample, drawn from within a 200-m buffer drawn around the same centroids.

Mule deer analysis.—We explored the above issues using an empirical data set from 53 female mule deer captured and fit with GPS radio collars set to attempt a fix once every 5 hours between 2008 and 2010 (C. R. Anderson, *unpublished data*). Though these data arise from a movement process, they are commonly used to fit RSFs, approximating a point process model, and thus all of the same issues apply. We fit RSFs in a use–availability framework separately for each deer, examining a suite of 14 environmental covariates expected to influence deer habitat selection based on preliminary analysis (Appendix B) and compared three approaches for sampling availability. The first two methods were based on home range estimates, where 100 000 random locations were drawn for each animal across both the 100% minimum convex polygon (MCP) and a polygon delineated by buffering all locations for each individual by the mean Euclidean displacement between locations (400 m), and combining these into a single polygon for each deer. These analyses provide inference at the third order of selection. Aside from controlling for differing availability, we made the assumption that that the GPS locations were independent, following the advice of Otis and White (1999). We next examined location-based availability for a limited number of individuals by buffering each use location by 400 m and drawing 1000 random locations within each buffer. For all analyses, we extracted and standardized ($[x - \bar{x}]/\sigma_x$) all continuous predictor covariates for every used and available location, and randomly selected subsets of the availability sample; for the MCP and buffered polygon, we selected samples ranging from 100 to 50 000 locations, and for the movement buffers between 5 and 500 locations per buffer. We fit RSFs to individual deer using either logistic regression or conditional logistic regression. We repeated this process 1000 times and recorded the expectation of the coefficient estimator and 95% intervals of the mean coefficient estimates (i.e., 95% quantiles of the group of all 1000 $\hat{\beta}$ from the model iterations; note these are not simulation envelopes). For a subset of individuals, we drew 5 000 000 random locations across their MCP and repeated this process, drawing availability samples ranging from 5000 to 1 000 000 locations.

Results

Simulations.—In all matched sample analyses examining third-order selection, with true or simulated covariates, coefficient estimates were unbiased and converged to an accurate value at availability samples of 10 000 or less (Fig. 1D–F and Appendix C). In the mismatched sample analysis, $E(\hat{\beta}_1)$ was consistently biased on the true landscape regardless of sample size and differed substantially between small and large availability samples (Appendix C). We note that in discussing bias throughout, we are not strictly discussing a statistical bias, as the model is accurately estimating coefficients for the given used and available samples, but rather a bias in inference, as results do not reflect the data-generating process at this order of selection. With a smaller used sample size, these issues were less pronounced. In both analyses, the simulation envelope was wider with fewer used samples (Fig. 1 and Appendix C). On simulated landscapes, autocorrelation substantially influenced both the bias and the size of the availability sample needed for convergence (Fig. 1). For the continuous covariate, when autocorrelation was weak, $E(\hat{\beta}_1)$ was unbiased and converged rapidly, but both bias and the size of the availability sample needed for convergence increased with autocorrelation. This bias is not directly a result of autocorrelation, but rather autocorrelation increases the degree of imbalance between the true and sampled availabilities in the mismatched sample analysis. Again, a larger availability sample was needed for convergence with larger ratios of use to availability and, in some cases, convergence was not reached even at very large sample sizes. For the binary covariate, coefficient estimates converged rapidly. With moderate autocorrelation, estimates were biased but the degree of bias depended on the proportion of the landscape composed of that covariate (Appendix A). Coefficient estimates from RSFs examining third/fourth-order selection converged to a stationary value at availability samples of 20–100 points per buffer and were unbiased for the matched sample analysis (Appendix C). With a mismatched sample, estimates were influenced by autocorrelation, though bias was only an issue at moderate levels of autocorrelation (Appendix C) and estimates converged at similar sample sizes as for the matched sample.

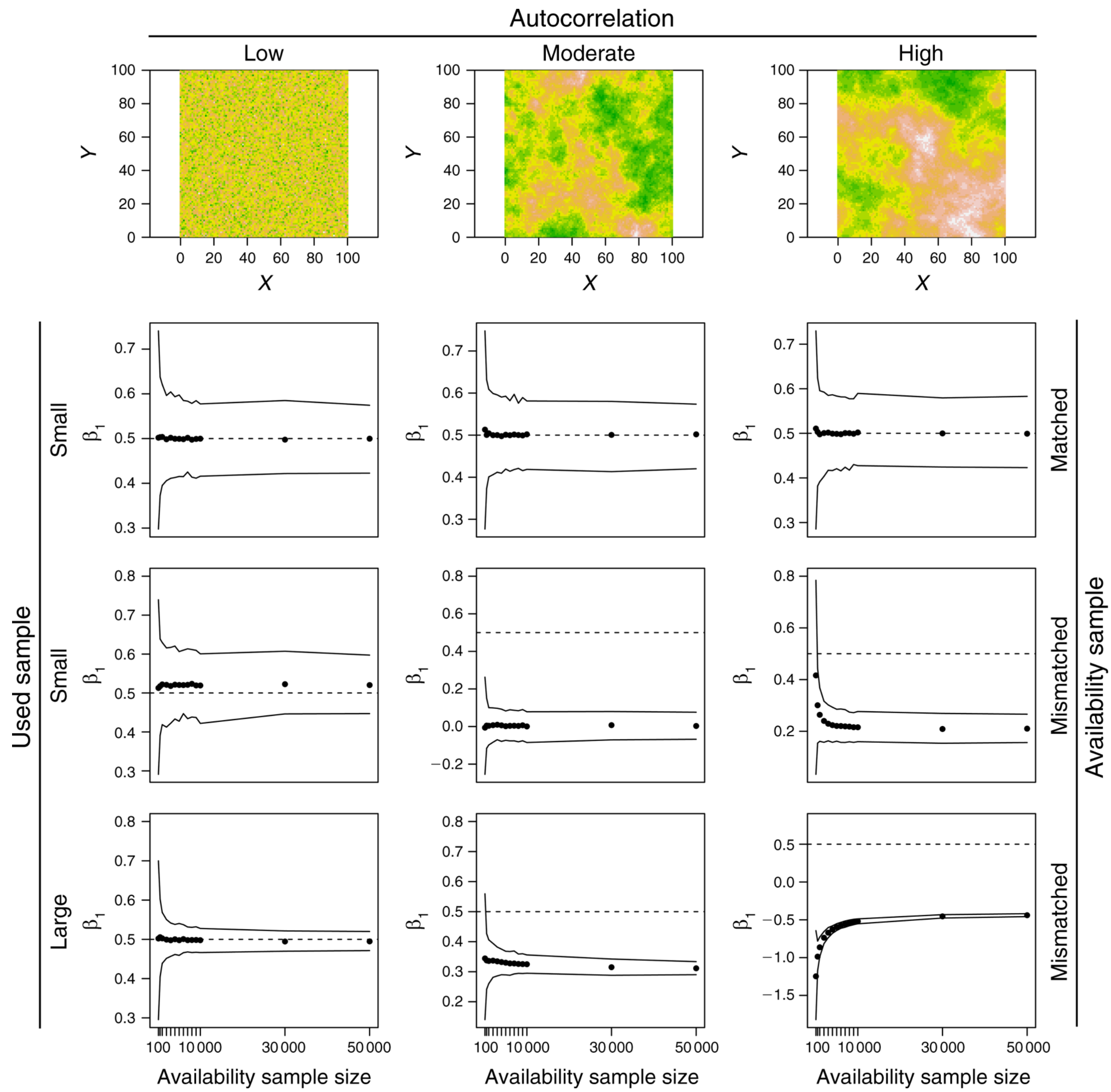

FIG. 1. Continuous landscape covariates simulated as a Gaussian random field with low (range parameter $\phi = 0.001$), moderate ($\phi = 10$), or high ($\phi = 100$) autocorrelation, and expectations of the coefficients (β_1, black points) and 95% simulation envelopes (solid lines) from 500 resource selection function (RSF) model iterations as a function of availability sample size, with matched or mismatched availability compared to small (600) or large (6000) used sample sizes. Dotted lines represent the value used for simulation. Models were fit with logistic regression in all cases.

Mule deer analysis.—Results varied substantially among individuals and among covariates within individuals. For many animals, coefficient estimates were highly variable at small availability samples, but appeared to converge to a consistent value at sample sizes ranging from 1000 to 10 000 locations, or higher (Fig. 2A). However, for many individual and covariate combinations, there were substantial differences between $E(\hat{\beta}_1)$ at small sample sizes and the value to which it eventually converged (Fig. 2B, C). For a few individuals, coefficient estimates did not converge until extraordinarily large availability samples were used (Fig. 2B). These patterns often were not consistent among covariates within the same individuals, and appeared to be a function of the individual and covariate combination (though for some individuals these issues persisted across covariates). In addition, these results were not consistent between availability samples drawn from the MCP and the buffered polygon. When examining third/fourth-order selection coefficient estimates were consistent at samples of 20 points per buffer or greater (Fig. 2D). We found no cases of extreme differences in $E(\hat{\beta}_1)$

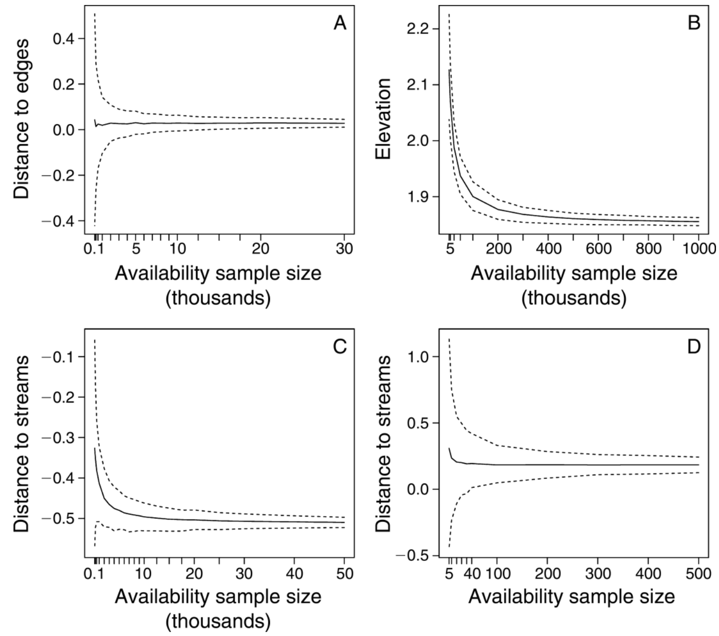

FIG. 2. Expectation of the coefficients (solid line) and upper and 95% quantiles of all $\hat{\beta}$ from 1000 RSF model iterations (dashed lines) as a function of availability sample size, for (A) distance to edges for deer 10, (B) elevation for deer 62, and (C, D) distance to streams for deer 2. In panel A, availability was drawn from the buffered polygon, for panels B and C it was drawn from the MCP, and for panel D it was drawn from buffers around each location. Models were fit with logistic regression for panels A–C and with conditional logistic regression for panel D.

between small and large availability samples as seen in the third-order analyses. In addition, the scale of the conditional analysis limited inference to those covariates that the deer interacted with locally, but reduced or eliminated our ability to make inference on interactions at a larger scale (e.g., broad avoidance of a covariate).

DISCUSSION

It has long been recognized that the definition of the availability sample is critical when estimating RSFs in a use–availability framework (Johnson 1980, Manly et al. 2002). However, to date there has been little formal assessment of how coefficient estimates are influenced by the size of this sample, with examinations of spatial extent set in a biological rather than a methodological context (but see Beyer et al. 2010). Thus, there is little guidance for researchers using these methods. Our results indicate that both factors must be carefully considered to avoid analytical and interpretive biases.

The availability sample must be large enough to avoid significant numerical integration error. If a sufficiently large sample is not used then the model does not accurately approximate a point process model, and any inference is compromised. However, a sufficient size is dependent on the animal, the covariates, the ratio of use to availability, and an accurate representation of what is available to the animal. In simulations with matched samples, coefficient estimates were similar at all availability sample sizes and relatively few locations were needed for estimates to converge ($<$10 000 third-order analysis, and $<$100 per buffer for third/fourth-order analysis). In simulations with a mismatched sample, more locations were needed for convergence in the third-order analysis, but the expectation of the coefficient estimators were biased at all sample sizes and differed substantially between small and large samples.

Attributes of the environmental covariates heavily influenced the interpretational bias of coefficient estimates, but these factors were related to the scale of inference. At the third order, bias was evident for covariates with moderate and high spatial autocorrelation. This issue was only present with moderate autocorrelation when examining the third/fourth order, with almost no bias at the highest levels of autocorrelation. Autocorrelation induces bias because a mismatch in true and sampled availability in geographic space leads to an imbalance in parameter space. Thus, the level of imbalance appears to result from an interaction between the autocorrelation structure and the extent over which availability is sampled. With the third/fourth order analysis the spatial extent is such that the imbalance was greatest at moderate levels of autocorrelation, likely relating to the size of the covariate patches relative to the extent of the availability sample. With increasing buffer sizes in this analysis, similar bias likely would occur at higher autocorrelation.

In the deer analysis, estimates often differed substantially between small and large availability samples, but more locations typically were needed for convergence than in simulations. The results of the deer analysis paired with those from the mismatched simulations point to a likely inaccurate assessment of what was available to the animal at the 3rd order, with unclear results for the third/fourth-order (i.e., neither the simulations nor the deer analysis exhibited large differences between coefficient estimates at small and large availability samples). Thus, it is possible that an interpretational bias resulted from incorrectly assessing what was available to be selected by the deer. Beyer et al.

(2010) suggest that in such cases the term "preference" should be used in place of "selection" to highlight that the behavioral process has not been captured. We agree that some differentiation is needed and our results provide some guidance for conditions that are likely to cause a mismatch between the scale of availability and the scale of desired inference (e.g., autocorrelation, and small ratios of use to availability; however we note that these results appear highly context and individual dependent). While third/fourth-order analyses appear to provide less bias between small and large availability samples, we caution that location based analyses can be more computationally intensive and limit inference regarding interactions that occur at a larger scale than that of the movement process (i.e., avoidance of covariates at the third order will not be captured). In addition, because the spatial extent of availability is reduced with this method, there can be little variation within certain environmental variables leading to high multicollinearity and an ill-posed model. More sophisticated methods for assessing selection and behavior exist that can address the issues described here, including movement-based RSFs that account for temporal autocorrelation (e.g., Johnson et al. 2008, Hooten et al. 2010, 2013), hierarchical methods providing robust population-level inference (Duchesne et al. 2010), and methods that explicitly account for the influence of availability (Matthiopoulos et al. 2011). We note that these methods require advanced statistical knowledge and do not guard against interpretational bias.

The results of our analyses highlight the myriad of issues that can influence coefficient estimates in RSF analysis, but the question of the degree to which inference is impacted remains. For studies that use RSFs to strictly draw inference from resulting coefficients, it seems clear that there is the potential for interpretational bias, likely exacerbated by high serial autocorrelation in telemetry locations. However, RSFs often are used solely to produce maps for subsequent analysis or for use in management (Boyce and McDonald 1999, Northrup et al. 2012, Shafer et al. 2012). Often, such maps are categorized into broad bins and cross validated or validated with other data (Johnson et al. 2006). In these cases, small biases might have little impact on the resulting map, particularly if validations indicate a highly predictive surface.

Practical guidance and conclusions.—While our results highlight numerous issues that can affect inference from RSF analyses, they also offer guidance:

1) Most critically, a sufficiently large availability sample must be used. If this sample is insufficient, then logistic regression does not approximate the point process model as intended, and no faith can be put in coefficient estimates. A sensitivity analysis of the availability sample size at the spatial extent of interest should be included in any RSF analysis. Such assessments could follow the methods presented here, and those suggested elsewhere (e.g., Manly et al. 2002, Warton and Shepherd 2010, Aarts et al. 2012) where multiple samples of varying sizes are tested until coefficient estimates converge.
2) Provided a sufficiently large sample will be used, how availability is drawn depends directly on the desired scale of inference. Once this is determined, accurately defining what is available to the animal and matching the scale of availability to the desired scale of inference is paramount in studies aimed at obtaining inference on selection behavior. Such definitions are difficult to obtain, thus, when examining serially autocorrelated GPS data, multiple scales of availability should be considered and knowledge of the system in question will be critical in interpreting responses across scales. However, we note that inference is likely prone to bias, which can vary across covariates relative to differences in autocorrelation structure, and coefficients might not represent the behavioral process (Beyer et al. 2010).
3) Where bias in inference is likely, behavioral interpretation should be avoided. In such cases, mapping applications validated with other data are still useful (e.g., Shafer et al. 2012).
4) Extremely large availability samples will be needed in some systems, which may add computing time, thus researchers will need to decide what level of consistency is desired, assess selection at a different scale, or identify and remove problem individuals (i.e., those for which convergence failed). Otherwise, methods such as regular sampling of availability, or weighting of the availability sample could be explored (Aarts et al. 2012; W. Fithian and T. Hastie, *unpublished manuscript*).

The fields of animal movement and habitat selection are evolving at a rapid pace due to vast improvements in data collection. Analyses of these data increasingly are being used in resource management decision making and planning, making robust analysis and inference critically important. With such an ever-evolving field that has potential societal implications, the need to continually assess methods and assumptions is paramount.

Acknowledgments

This work was supported by Colorado Parks and Wildlife, Bureau of Land Management, Colorado Mule Deer Association, Colorado Mule Deer Foundation, Colorado State Severance Tax Fund, EnCana Corporation, ExxonMobil Production Corporation, Federal Aid in Wildlife Restoration, Marathon Oil Corporation, Shell Petroleum, Williams Production LMT Corporation, and Piceance Basin land owners. We thank M. Rice, K. Logan, J. Matthiopoulos, and one anonymous reviewer for helpful comments on the manuscript. The use of trade names or products does not constitute endorsement by the U.S. Government.

Literature Cited

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

Aarts, G., M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos. 2008. Estimating space-use and habitat preference from wildlife telemetry data. Ecography 31:140–160.

Baddeley, A., and R. Turner. 2000. Practical maximum pseudolikelihood for spatial point patterns. Australian and New Zealand Journal of Statistics 42:283–322.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3:327–338.

Beyer, H. L., D. T. Haydon, J. M. Morales, J. L. Frair, M. Hebblewhite, M. Mitchell, and J. Matthiopoulos. 2010. The interpretation of habitat preference metrics under use-availability designs. Philosophical Transactions of the Royal Society B 365:2245–2254.

Boyce, M. S. 2006. Scale for resource selection functions. Diversity and Distributions 12:269–276.

Boyce, M. S., J. S. Mao, E. H. Merrill, D. Fortin, M. G. Turner, J. Fryxell, and P. Turchin. 2003. Scale and heterogeneity in habitat selection by elk in Yellowstone National Park. Ecoscience 10:421–431.

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

Duchesne, T., D. Fortin, and N. Courbin. 2010. Mixed conditional logistic regression for habitat selection studies. Journal of Animal Ecology 79:548–555.

Fieberg, J., J. Matthiopoulos, M. Hebblewhite, M. S. Boyce, and J. L. Frair. 2010. Correlation and studies of habitat selection: problem, red herring or opportunity? Philosophical Transactions of the Royal Society B 365:2233–2244.

Fretwell, S. D., and H. L. Lucas. 1969. On territorial behavior and other factors influencing habitat distribution in birds. Acta Biotheoretica 19:16–36.

Garshelis, D. L. 2000. Delusions in habitat evaluation: measuring use, selection and importance. Pages 111–164 *in* L. Boitani and T. K. Fuller, editors. Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York, New York, USA.

Hebblewhite, M., and E. Merrill. 2008. Modelling wildlife–human relationships for social species with mixed-effects resource selection models. Journal of Applied Ecology 45: 834–844.

Hooten, M. B., E. M. Hanks, D. S. Johnson, and M. W. Alldredge. 2013. Temporal variation and scale in movement-based resource selection functions. Statistical Methodology, *in press*.

Hooten, M. B., D. Johnson, E. Hanks, and J. Lowry. 2010. Agent-based inference for animal movement and selection. Journal of Agricultural, Biological, and Environmental Statistics 15:523–538.

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

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 61:65–71.

Johnson, D. S., D. L. Thomas, J. M. V. Hoef, and A. Christ. 2008. A general framework for the analysis of animal resource selection from telemetry data. Biometrics 64:968.

Lele, S. R. 2009. A new method for estimation of resource selection probability function. Journal of Wildlife Management 73:122–127.

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

Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. Mcdonald, and W. P. Erickson. 2002. Resource selection by animals: statistical design and analysis for field studies. Second edition. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Matthiopoulos, J., M. Hebblewhite, G. Aarts, and J. Fieberg. 2011. Generalized functional responses for species distributions. Ecology 92:583–589.

Morris, D. W. 2003. Toward an ecological synthesis: a case for habitat selection. Oecologia 136:1–13.

Northrup, J. M., G. B. Stenhouse, and M. S. Boyce. 2012. Agricultural lands as ecological traps for grizzly bears. Animal Conservation 15:369–377.

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

R Development Core Team. 2012. R: a language and environment for statistical computing. R Project for Statistical Computing, Vienna, Austria. www.r-project.org

Shafer, A. B. A., J. M. Northrup, K. White, M. S. Boyce, S. D. Côté, and D. Coltman. 2012. Habitat selection predicts genetic relatedness in an alpine ungulate. Ecology 93:1317–1329.

Warton, D. I., and L. C. Shepherd. 2010. Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. Annals of Applied Statistics 4:2203–2204.

Supplemental Material

Appendix A

Simulation of landscape covariates as Gaussian random fields (*Ecological Archives* E094-131-A1).

Appendix B

Environmental covariates used in resource selection function (RSF) modeling for mule deer (*Ecological Archives* E094-131-A2).

Appendix C

Results of basic simulations and location-based availability simulations (*Ecological Archives* E094-131-A3).

Supplement

R code used in simulations and .R data files used in empirical deer analysis presented in the paper (*Ecological Archives* E094-131-S1).