

Correcting Home-Range Models for Observation Bias

JON S. HORNE,¹ *University of Idaho, Department of Fish and Wildlife, Moscow, ID 83844, USA*

EDWARD O. GARTON, *University of Idaho, Department of Fish and Wildlife, Moscow, ID 83844, USA*

KIMBERLY A. SAGER-FRADKIN, *University of Idaho, Department of Fish and Wildlife, Moscow, ID 83844, USA*

ABSTRACT Home-range models implicitly assume equal observation rates across the study area. Because this assumption is frequently violated, we describe methods for correcting home-range models for observation bias. We suggest corrections for 3 general types of home-range models including those for which parameters are estimated using least-squares theory, models utilizing maximum likelihood for parameter estimation, and models based on kernel smoothing techniques. When applied to mule deer (*Odocoileus hemionus*) location data, we found that uncorrected estimates of the utilization distribution were biased low by as much as 18.4% and biased high by 19.2% when compared to corrected estimates. Because the magnitude of bias is related to several factors, future research should determine the relative influence of each of these factors on home-range bias. (JOURNAL OF WILDLIFE MANAGEMENT 71(3):996–1001; 2007)

DOI: 10.2193/2005-678

KEY WORDS biotelemetry, bivariate normal, home range, kernel, Global Positioning System, maximum likelihood, nonresponse, observation bias, telemetry.

Accurate depiction of animal home ranges is a significant component of wildlife research and management. Home ranges have been used to gain insight into important resources (Marzluff et al. 2004), predator–prey relationships (Norbury et al. 1998), and competition (Minta 1992), as well as social pressures and mating systems (Conner et al. 1999). Because direct observation of the home range is usually impossible or impractical, home ranges are typically modeled from a sample of observations used to estimate the probability that the animal occurred in an area during a specified period of time (i.e., utilization distribution).

The sample of observations may be obtained using a variety of techniques, including direct observation of marked individuals, as well as the more common techniques of radio- and Global Positioning System (GPS) telemetry. Because the utilization distribution is estimated from a sample, its accuracy can be affected by statistical characteristics of the sample including independence of locations, sample size, measurement error, and observation rate (Samuel and Fuller 1994, Garton et al. 2001, Kernohan et al. 2001).

The most prominent sampling concerns in the wildlife literature have been serial correlation and sample size, and several authors have suggested appropriate sampling designs for dealing with these problems (Samuel and Fuller 1994, Otis and White 1999, Garton et al. 2001, Kernohan et al. 2001). Two sampling concerns that have received comparably less attention are measurement error and observation bias. Measurement error occurs when the true location of the animal is different from the estimated location, with the general result being decreased power of subsequent analyses (White and Garrott 1990). Although procedures for mitigating measurement error of telemetry data have received some attention (e.g., Samuel and Kenow 1992, Anderson-Sprecher 1994), likely a more significant problem for estimating the utilization distribution is observation bias.

Current home-range models implicitly assume an equal probability of obtaining a location, given one is attempted, in every part of the animal's home range. However, numerous studies have shown a marked difference in these probabilities (e.g., Moen et al. 1996, Johnson et al. 1998, D'Eon et al. 2002). Although the vast majority of studies have shown observation bias for GPS telemetry, it is likely that other methods for gathering location data may be biased as well (Rodgers 2001). When home-range models fail to incorporate observation bias, the resulting estimates and inferences based on these estimates will be biased. Therefore, we describe techniques for explicitly incorporating the probability of obtaining a location into many home-range models so that researchers can correct home-range models for observation bias and determine the relative impact of this bias on home-range estimates.

CORRECTING HOME-RANGE MODELS

Observation bias occurs when an attempt to gather location data on an animal is unsuccessful. Failure to obtain a location may result from equipment failure, position of the radiotransmitting antenna, or position of the GPS-receiving antenna. In addition to these factors, researchers have shown certain physiographic characteristics of the animal's location to affect observation rates (Dussault et al. 1999, D'Eon et al. 2002, Frair et al. 2004). Field studies have shown substantial variation in observation rates across study sites, ranging from 13% to 100% (Moen et al. 1996, Obbard et al. 1998, D'Eon et al. 2002, Frair et al. 2004). Because observation rates are often related to physiographic characteristics, researchers have successfully used spatially explicit covariates (θ) to model the observation rates across a study site [$\hat{p}(x, y)|\theta$], where \hat{p} is the estimated probability of successfully collecting a location at (x, y) , given one was attempted (e.g., Johnson et al. 1998, D'Eon et al. 2002, Frair et al. 2004). These estimated observation rates (\hat{p}) serve as the basis for correcting home-range models for observation bias.

Two approaches for correcting observation bias include

¹ E-mail: horn8262@uidaho.edu

weighting and imputation (Cochran 1983). Weighting methods use only the data that were collected and correct observation bias by assigning a weight or inflation factor to each unit in the sample. Imputation methods correct observation bias by inserting values for missed measurements. In general, researchers use weighting methods for unit nonresponse (i.e., all data are missing from the intended sample unit), whereas researchers use imputation for item nonresponse (i.e., some measurements are missing from the intended sample unit; Little 1986). Because nonresponse in the context of home-range estimation deals with complete loss of data (i.e., both x and y coordinates) from a sample unit, we describe methods for correcting home-range models based on weighting.

We suggest corrections for 3 general types of home-range models. The first is parametric models in which researchers estimate parameters using least squares, including the bivariate circular normal (Calhoun and Casby 1958) and the generalized bivariate normal (Jennrich and Turner 1969). The second is parametric home-range models in which researchers estimate parameters using maximum likelihood, such as the 2-mode bivariate normal mix (Don and Rennolls 1983, Horne and Garton 2006a) and the bivariate exponential power (Horne and Garton 2006a). The last type includes nonparametric models based on kernel smoothing techniques (Worton 1989).

Correcting home-range models is relatively straightforward once researchers have constructed an appropriate model of observation rates [i.e., $\hat{p}(x, y)$] in the study area (see Johnson et al. 1998, D'Eon et al. 2002, Frair et al. 2004 for examples). Researchers assign each individual location i a weight w_i computed as the reciprocal of the probability of being included in the sample $w_i = 1/\hat{p}_i$ (Lohr 1999:265).

For the bivariate normal home range, researchers estimate the bias-corrected model using a weighted mean $\hat{\mu}_w$ and covariance matrix \hat{S}_w :

$$\hat{\mu}_w = \left(\sum_{i=1}^n w_i \right)^{-1} \times \sum_{i=1}^n w_i \mathbf{x}_i,$$

$$\hat{S}_w = \left[\left(\sum_{i=1}^n w_i \right) \left(1 - \frac{1}{n} \right) \right]^{-1} \times \sum_{i=1}^n w_i (\mathbf{x}_i - \hat{\mu}_w)(\mathbf{x}_i - \hat{\mu}_w)',$$

where \mathbf{x} is a vector of x, y coordinates from n locations (West 1979). Researchers estimate the circular normal home range similarly with equal variances in the x and y dimensions.

Researchers use a weighted likelihood function to correct home-range models that use maximum likelihood to estimate parameters. Under simple random sampling with no observation bias (i.e., all N locations are fully observed), researchers would maximize the log-likelihood function

$$L = \sum_{i=1}^N \ln[\hat{f}(\mathbf{x}_i)]$$

to find parameter estimates where \hat{f} is a probability density function describing the home range. When there is

observation bias, a subset of locations (n) is observed from N . Researchers can obtain an estimate of the complete data log-likelihood, using only the observed locations, by weighting them inversely proportional to their probability of being included in the sample (Lawless et al. 1999:420–421):

$$\tilde{L} = \sum_{i=1}^n w_i \ln[\hat{f}(x_i)].$$

Because \tilde{L} is an estimate of the log-likelihood, it is generally referred to as pseudolikelihood. Researchers find parameter estimates by maximizing this weighted pseudolikelihood function.

Researchers can correct kernel home-range models for nonresponse bias by increasing the volume of individual kernels by w_i (Jones 1991). Thus, the general weighted kernel density estimate is

$$\hat{f}_{\text{kernel}}(x) = \left(\sum_{i=1}^n w_i \right)^{-1} \times \sum_{i=1}^n w_i K_h(x - X_i),$$

where K is a symmetric probability density function. When researchers use a circular bivariate normal for K , the weighted kernel estimate is

$$\hat{f}_{\text{kernel}}(x, y) = \left(\sum_{i=1}^n w_i \right)^{-1} \times \sum_{i=1}^n \frac{w_i}{2\pi h^2} \times \exp\left(-\frac{d_i^2}{2h^2}\right),$$

where d_i is the distance of the i^{th} observation from the x, y -coordinate and h is the smoothing parameter. We suggest using likelihood cross-validation (CVh) to choose h (Horne and Garton 2006b), with the kernel estimate at each hold-out location constructed using a weighted kernel.

EXAMPLES

To illustrate the effect of observation bias on estimates of the utilization distribution, we first defined a true distribution of space use as a 1-dimensional normal distribution with mean = 6 and standard deviation = 2. We generated 2 sets of locations where $n = 131$. The first represented an unbiased sample (i.e., no observation bias). For the second, we created a biased sample by simulating the effect of different location-acquisition probabilities in different parts of the sample space. For the range $3 \leq x \leq 4$, the probability of obtaining a location was 1 (i.e., high success rate for obtaining a location); for the range $5.5 \leq x \leq 6.5$, the probability was 0.33 (i.e., low success rate); and the remainder received a probability of 0.8. We used a one-dimensional fixed kernel with CVh choice of smoothing parameter to estimate the true distribution with 1) unbiased data, 2) biased data and an uncorrected kernel, and 3) biased data and a weighted kernel. We chose the normal distribution and the parameter values for convenience and to reflect typical success rates and sample sizes.

The kernel estimate using the unbiased data provided an estimate similar to the true distribution, although we expected some discrepancies because we are using a finite sample to estimate the true distribution (Fig. 1). However,

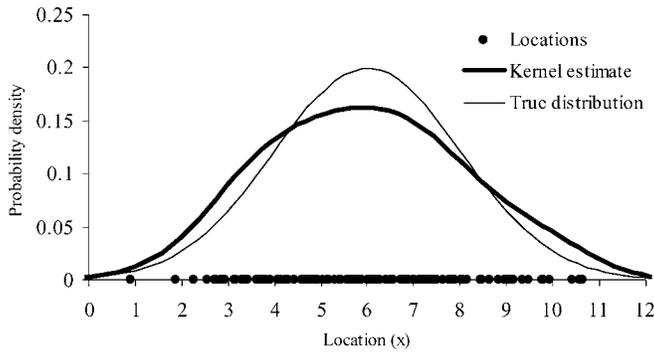


Figure 1. Fixed-kernel density estimate of a one-dimensional normal distribution with mean = 6 and standard deviation = 2.

when we used data biased by varying levels of acquisition probability, the kernel estimate was substantially different from when we used unbiased data (Fig. 2). When we corrected the kernel model for observation bias by weighting individual kernels, the estimate using the biased sample was nearly identical to the estimate using the unbiased sample (Fig. 2).

As a second example, we estimated the utilization distribution of a mule deer (*Odocoileus hemionus*) at Starkey Experimental Forest in LaGrande, Oregon, USA, using 3 home-range models (i.e., bivariate normal, 2-mode bivariate normal mix, and fixed kernel). For each home-range model, we estimated 1) an uncorrected utilization distribution (i.e., assumed equal observation rates across the animal's range) and 2) a bias-corrected distribution (i.e., used weightings to correct for unequal observation rates). We obtained weightings from a model of observation rates developed by Johnson et al. (1998), with individual weightings (w_i) ranging from 1.34 to 2.34 (location data and weightings for mule deer 930318D03 can be obtained from <http://www.fs.fed.us/pnw/starkey/>).

We quantified the proportional difference in the estimated probability-of-occurrence between bias-corrected ($\hat{f}(x, y)_c$) and uncorrected ($\hat{f}(x, y)_u$) home ranges using *Relative Bias*(x, y) = $(\hat{f}(x, y)_u - \hat{f}(x, y)_c) / \hat{f}(x, y)_c$.

Positive values of *Relative Bias*(x, y) indicate areas where the uncorrected model overestimated the probability of occurrence, while negative values indicate underestimates.

All home-range models showed differences between estimates of the utilization distribution uncorrected for observation bias and those corrected for spatially varying observation rates. However, the amount of difference varied spatially across the home range and was dependent on home-range model (Fig. 3). Of the 3 home-range models we considered, the bivariate normal model was least affected by unequal observation rates, with the uncorrected model relative bias ranging from -0.25% to 10%. The 2-mode bivariate mix and fixed-kernel home-range models had similar ranges of relative bias ranging from -14.2% to 19.2% and -18.4% to 16.8%, respectively.

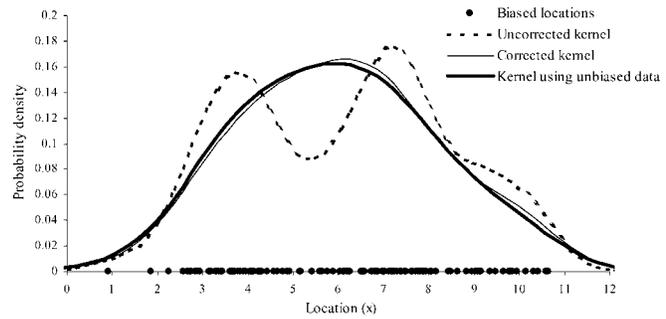


Figure 2. Corrected and uncorrected kernel density estimates used to illustrate the effect of observation bias on estimates of the utilization distribution.

DISCUSSION

Habitats used by wildlife are typically heterogeneous in terms of terrain and vegetation attributes. Because these habitat attributes affect our ability to collect biotelemetry data, observation rates are frequently unequal across the landscape. This results in a systematic omission of locations in portions of the home range and has the potential to bias subsequent analyses of space use. Recognizing this problem, several researchers have sought to correct resource selection analyses for observation bias by adjusting estimates of habitat use (e.g., Johnson et al. 1998, D'Eon 2003, Frair et al. 2004). Models of the utilization distribution are also based on estimates of space use, and our results demonstrate that they can be biased by unequal observation rates as well. The methods presented here, based on weighting individual locations, provide the first approach to correct current home-range models for observation bias.

If home-range estimates are not corrected for observation bias, subsequent analyses and inferences based on these estimates have the potential to be biased. For example, Marzluff et al. (2004) recently suggested methods to determine resource selection by animals based directly on estimates of the utilization distribution. Their analysis assumed that habitats with many locations per area are used more frequently than habitats with few locations. However, a relatively dense cluster of locations may be the result of high observation rates in that area and not necessarily a higher frequency of use. Without correcting models of the utilization distribution for unequal observation rates, researchers would be unable to differentiate between areas of actual high use versus areas of low use with high observation rates and vice versa.

Although there is potential for incorrect inference based on biased home ranges, the frequency at which this occurs will depend on the magnitude of the difference between biased and unbiased home-range estimates. In addition to our mule deer example, we estimated corrected and uncorrected home ranges for several other mule deer, elk (*Cervus elaphus*), and black bear (*Ursus americanus*). These preliminary investigations suggested that the magnitude of the difference depends, at least in part, on 1) the magnitude of the difference in observation rates across the home range, 2) the extent to which areas of similar probabilities are

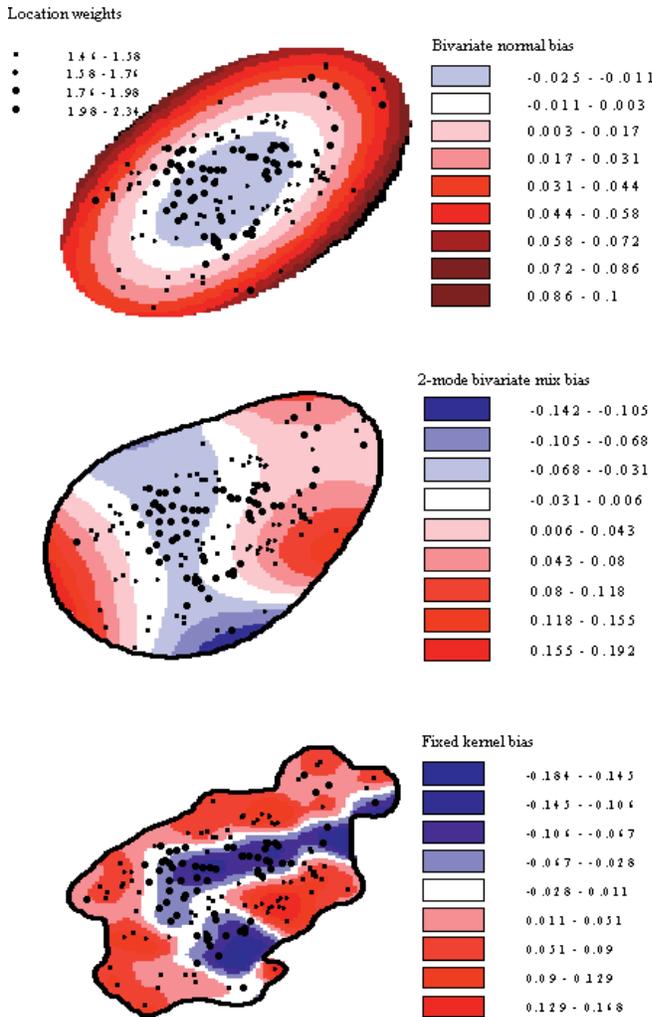


Figure 3. Proportional difference between uncorrected home-range models and models corrected for observation bias. Outer boundary represents 98% cumulative probability contour of corrected home range.

aggregated in space, 3) the home-range model, and 4) sample size. Future research should seek to determine the relative effect of each of these factors on the magnitude of the difference between uncorrected and corrected home-range models. Until this research is conducted, we warn of certain conditions that will lead to greater bias.

First, home-range estimates will have greater bias if there are areas with very low observation rates ($p \leq 0.5$), as well as areas with high observation rates ($p \approx 1$). Approximately equal observation rates, even if they are low, will not lead to a substantial difference between corrected and uncorrected home-range models. Several studies have developed models of observation rates for a variety of environmental conditions (see section on correcting home-range models). We suggest researchers review these and other studies documenting observation rates, with similar equipment, to determine if their study area is likely to have a wide range of observation rates.

Second, greater bias will occur if areas of low observation rates are aggregated in one part of the home range while areas of high observation rates are aggregated in other parts.

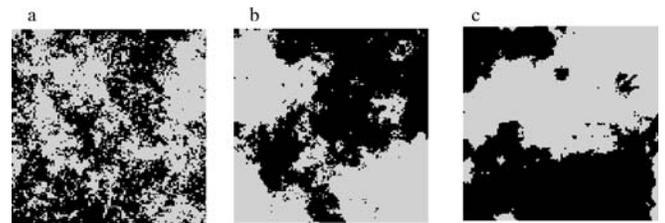


Figure 4. Three example landscapes (a-c) depicting the interaction between home-range extent (i.e., boundary of landscape) and patchiness of areas with low (gray) and high (black) observation rates. Home-range estimates become more biased as patches of similar observation rates become more aggregated (a to c).

The degree to which observation rates are aggregated within a home range depends on the extent of the home range relative to the patchiness of areas with similar observation rates. Home-range estimates will be more biased as patches of similar observation rates become more aggregated (Fig. 4). We found that black bear home ranges in Olympic National Park, Washington, USA, were generally quite large relative to the extent of areas with similar observation rates. Therefore, substantially different observation rates were not aggregated in large, continuous parts of the home range. The result was locations with large weights were interspersed with locations with small weights. In contrast, mule deer home ranges were generally smaller and thus incorporated relatively large, more continuous, areas of similar observation rates. In this case, we would expect observation bias to have a greater effect on mule deer home ranges than on black bear home ranges.

The role of home-range model and sample size in determining the magnitude of the difference between biased and unbiased home-range estimates is less predictable. Different home-range models will likely be more sensitive to observation bias than others. In our mule deer example, the bivariate normal was least affected by observation bias compared to the other 2 home-range models, especially if we excluded the outer contours. However, we caution against extrapolating this result to other situations and suggest future research should address the susceptibility of different home-range models to observation bias. As for the effect of sample size, the method we outlined is based on weighting observed locations. If observation rates are low enough in certain areas and sample size is inadequate, these areas may be void of locations, in which case the corrections we suggested would fail to compensate for missed locations.

The bias corrections we suggest require an a priori model of observation rates for the study area. However, developing these models may require a substantial investment of resources. Therefore, it is important to determine the susceptibility of each particular study to observation bias. Studies conducted in areas with extensive habitat heterogeneity including dense vegetation; open areas and steep topography are prone to more observation bias than sites with relatively flat terrain and homogenous vegetation.

If researchers suspect that significant observation bias will affect their results, it would be prudent to conduct a pilot

study to estimate the magnitude of bias and its effects on subsequent analyses. Researchers can either develop a model of observation rates specific to their study site or use a surrogate model developed with similar equipment under similar environmental conditions. Researchers can then use these models to assign weights to animal locations. By applying the corrections we present, researchers can determine the relative effect of observation bias on their analyses.

To this point, we have discussed unequal observation rates under the premise that they were extrinsic to the researcher's control. However, the methods we introduce can be applied to intentional differences in observation rates. For example, often researchers are more likely to collect radiotelemetry data during daylight hours as opposed to 24-hour sampling (Beyer and Haufler 1994). This type of sampling would bias estimates of the utilization distribution high in areas used during diurnal periods and low during nocturnal periods. By weighting locations collected under different sampling intensities, researchers could correct these unequal observation rates using the methods we present. For example, if a study was designed to spend half the effort collecting locations at night as during the day, then an unbiased utilization distribution could be estimated by weighting nocturnal locations twice as much as diurnal locations. More generally, this approach could potentially be applied to a variety of sampling designs by relating sampling effort to the probability of obtaining a location (see Cochran 1977).

MANAGEMENT IMPLICATIONS

Wildlife management decisions are often based on estimates of animal space use, and much work has been devoted to developing new techniques and statistical models that more accurately depict an animal's true use of space. However, if researchers use estimates of space use that are biased by unequal observation rates, this might lead to erroneous management decisions. Thus, it is critical that researchers identify the level of observation bias in their study and mitigate any substantial bias using the methods we presented.

ACKNOWLEDGMENTS

The United States Forest Service and the Starkey Experimental Forest and Range, LaGrande, Oregon, provided mule deer data. We thank J. L. Rachlow, K. J. Jenkins, and 3 anonymous reviewers for providing valuable comments on the manuscript. We thank J. G. Kie at the Pacific Northwest Research Station, United States Forest Service, and the Idaho Department of Fish and Game for funding.

LITERATURE CITED

Anderson-Sprecher, R. 1994. Robust estimates of wildlife location using telemetry data. *Biometrics* 50:406–416.

Beyer, D. E., Jr., and J. B. Haufler. 1994. Diurnal versus 24-hour sampling of habitat use. *Journal of Wildlife Management* 58:178–180.

Calhoun, J. B., and J. U. Casby. 1958. Calculation of home range and density of small mammals. U.S. Public Health Service Public Health Monograph 55, Washington, D.C., USA.

Cochran, W. G. 1977. Sampling techniques. Third edition. John Wiley & Sons, New York, New York, USA.

Cochran, W. G. 1983. Historical perspective. Pages 11–25 in W. G. Madow, I. Olkin, and D. B. Rubin, editors. *Incomplete data in sample surveys. Volume 2: theory and bibliographies*. Academic Press, New York, New York, USA.

Conner, M., B. Plowman, B. D. Leopold, and C. Lovell. 1999. Influence of time-in-residence on home range and habitat use of bobcats. *Journal of Wildlife Management* 63:261–269.

D'Eon, R. G. 2003. Effects of stationary GPS fix-rate bias on habitat-selection analyses. *Journal of Wildlife Management* 67:858–863.

D'Eon, R. G., R. Serrouya, G. Smith, and C. O. Kochanny. 2002. GPS radiotelemetry error and bias in mountainous terrain. *Wildlife Society Bulletin* 30:430–439.

Don, B. A. C., and K. Rennolls. 1983. A home range model incorporating biological attraction points. *Journal of Animal Ecology* 52:69–81.

Dussault, C., R. Courtois, J. P. Ouellet, and J. Huot. 1999. Evaluation of GPS telemetry collar performance for habitat studies in the boreal forest. *Wildlife Society Bulletin* 27:965–972.

Frair, J. L., S. E. Nielsen, E. H. Merrill, S. R. Lele, M. S. Boyce, R. H. M. Munro, G. B. Stenhouse, and H. L. Beyer. 2004. Removing GPS collar bias in habitat selection studies. *Journal of Applied Ecology* 41:201–212.

Garton, E. O., M. J. Wisdom, F. A. Leban, and B. K. Johnson. 2001. Experimental design for radiotelemetry studies. Pages 15–42 in J. J. Millspaugh and J. M. Marzluff, editors. *Radio tracking and animal populations*. Academic Press, New York, New York, USA.

Horne, J. S., and E. O. Garton. 2006a. Selecting the best home range model: an information theoretic approach. *Ecology* 87:1146–1152.

Horne, J. S., and E. O. Garton. 2006b. Likelihood cross-validation versus least squares-cross-validation for choosing the smoothing parameter in kernel home-range analysis. *Journal of Wildlife Management* 70:641–648.

Jennrich, R. I., and F. B. Turner. 1969. Measurements of non-circular home range. *Journal of Theoretical Biology* 22:227–237.

Johnson, B. K., A. A. Ager, S. L. Findholt, M. J. Wisdom, D. B. Marx, J. W. Kern, and L. D. Bryant. 1998. Mitigating spatial differences in observation rate of automated telemetry systems. *Journal of Wildlife Management* 62:958–967.

Jones, M. C. 1991. Kernel density estimation for length biased data. *Biometrika* 78:511–519.

Kernohan, B. J., R. A. Gitzen, and J. J. Millspaugh. 2001. Analysis of animal space use and movements. Pages 125–166 in J. J. Millspaugh and J. M. Marzluff, editors. *Radio tracking and animal populations*. Academic Press, New York, New York, USA.

Lawless, J. F., J. D. Kalbfleisch, and C. J. Wild. 1999. Semiparametric methods for response-selective and missing data problems in regression. *Journal of the Royal Statistical Society, Series B* 61:413–438.

Little, R. J. A. 1986. Survey nonresponse adjustments for estimates of means. *International Statistical Review* 54:139–157.

Lohr, S. L. 1999. Sampling: design and analysis. Brooks/Cole, Pacific Grove, California, USA.

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

Minta, S. C. 1992. Tests of spatial and temporal interaction among animals. *Ecological Applications* 2:178–188.

Moen, R., J. Pastor, Y. Cohen, and C. C. Schwartz. 1996. Effects of moose movement and habitat use on GPS collar performance. *Journal of Wildlife Management* 60:659–668.

Norbury, G. L., D. C. Norbury, and R. P. Heyward. 1998. Behavioral responses of two predator species to sudden declines in primary prey. *Journal of Wildlife Management* 62:45–58.

Obbard, M. E., B. A. Pond, and A. Perera. 1998. Preliminary evaluation of GPS collars for analysis of habitat use and activity patterns of black bears. *Ursus* 10:209–217.

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.

Rodgers, A. R. 2001. Recent telemetry technology. Pages 79–121 in J. J. Millspaugh and J. M. Marzluff, editors. *Radio tracking and animal populations*. Academic Press, New York, New York, USA.

Samuel, M. D., and M. R. Fuller. 1994. Wildlife radiotelemetry. Pages 370–418 in T. A. Bookhout, editor. *Research and management techniques for wildlife and habitats*. Fifth edition. The Wildlife Society, Bethesda, Maryland, USA.

Samuel, M. D., and K. P. Kenow. 1992. Evaluating habitat selection with radio-telemetry triangulation error. *Journal of Wildlife Management* 56: 725–734.

West, D. H. D. 1979. Updating mean and variance estimates: an improved method. *Communications of the Association for Computing Machinery* 22:532–535.

White, G. C., and R. A. Garrott. 1990. *Analysis of wildlife radio-tracking data*. Academic Press, San Diego, California, USA.

Worton, B. J. 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70:164–168.

Associate Editor: Alpizar-Jara.