Research Article

Bandwidth Selection for Fixed-Kernel Analysis of Animal Utilization Distributions

ROBERT A. GITZEN,[1,2] *College of Forest Resources, University of Washington, Seattle, WA 98195, USA*

JOSHUA J. MILLSPAUGH, *Department of Fisheries and Wildlife Sciences, University of Missouri, Columbia, MO 65211, USA*

BRIAN J. KERNOHAN, *Forest Capital Partners, LLC, Monmouth, OR 97361, USA*

Abstract

In studies of animal space use, researchers often use kernel-based techniques for estimating the size of an animal's home range and its utilization distribution from radiotracking data. However, the kernel estimator is highly sensitive to the bandwidth value used. Previous ecological studies recommended least-squares cross-validation (LSCV) as the default bandwidth selection method, but some statisticians consider this technique inferior to newer methods. We used simulations to compare the performance of the scaling LSCV and reference approaches to plug-in and solve-the-equation (STE) bandwidth methods. We generated samples of 20, 50, and 150 points from mixtures of 2, 4, and 16 bivariate normal distributions. We selected the ranges of potential variances for these distributions to create 4 distribution types with varied levels of clumping to simulate the diversity of location patterns expected from radiotracking data. For most distribution types, plug-in and STE methods performed as well or better than LSCV in % absolute error of home-range size estimates and overlap of estimated and true utilization distributions. Although the relative differences usually were small, the plug-in and STE approaches provide good alternatives to LSCV. However, LSCV performed better with distribution types composed entirely of tight clumps of points. The reference bandwidth performed poorly for most distributions. Surprisingly, it often had the lowest absolute error at outer contours for distributions consisting of a single very tight cluster surrounded by more dispersed points. Although our results demonstrate the utility of plug-in and STE approaches, no method was best across all distributions. Rather, choice of a bandwidth selection method may vary depending on the study goals, sample size, and patterns of space use by the study species. In general, we recommend plug-in and STE approaches for estimating relatively smooth outer contours. The LSCV approach is better at identifying tight clumps, including areas of peak use, although risk of LSCV failure also increases when a distribution has a very tight cluster of points. When planning to use kernel methods, researchers should consider these factors to make preliminary decisions about the bandwidth method expected to be most appropriate in their study. (JOURNAL OF WILDLIFE MANAGEMENT 70(5):1334–1344; 2006)

Wildlife studies frequently seek to estimate home-range size and other aspects of animal space use based on locations obtained with radiotelemetry. Compared to previous home-range estimators such as the minimum convex polygon (MCP) and bivariate-normal methods, kernel-density estimation (Silverman 1986) offers lower bias (Worton 1995, Seaman and Powell 1996, Swihart and Slade 1997) and higher flexibility in handling complex location patterns (Worton 1989, Seaman et al. 1999). Moreover, the kernel method directly provides an estimated probability density function (Silverman 1986) that corresponds to an animal's utilization distribution (UD; Van Winkle 1975, Worton 1987, Kernohan et al. 2001). The UD measures the intensity or probability of use throughout an animal's home range, allowing calculation of home-range area within any desired probability contour. In addition, recent studies have recommended the use of kernel-estimated UDs to measure joint space use of multiple animals (Seidel 1992, Millspaugh et al. 2004, Fieberg and Kochanny 2005) and to evaluate resource selection (Marzluff et al. 2001, 2004), expanding the utility of the kernel method for animal-movement studies.

However, the kernel approach is highly sensitive to the choice of bandwidth values (smoothing parameters; Silverman 1986, Seaman et al. 1999, Kernohan et al. 2001). The

bandwidth determines the relationship between the distance of a used location from an evaluation point on the landscape and the contribution of the location to the density estimate at that point. An excessively large bandwidth value produces an overly smooth estimated UD, which overestimates home-range size (Kernohan et al. 2001). Conversely, with very small bandwidths, the UD may fragment into numerous components, producing negatively biased estimates. Biologists desire an objective method of bandwidth selection that they can automatically apply to numerous sets of locations and that usually provides the most accurate estimates of the parameter of interest.

Previous ecological studies (Worton 1995, Seaman and Powell 1996, Seaman et al. 1999, Kenward et al. 2001, Gitzen and Millspaugh 2003) evaluated 3 general methods for selecting the bandwidth values: the normal or reference method (REF), the scaled-REF approach in which the REF-selected value (h_{REF}) is multiplied by a fixed proportion (e.g., $0.8 \times h_{REF}$), and the least-squares cross-validation (LSCV) method. The REF method performs poorly except for unimodal distributions (Seaman et al. 1999, Gitzen and Millspaugh 2003). The scaled-REF approach depends on ad hoc choice of the appropriate scaling factor (Worton 1995). In contrast, LSCV avoids such potentially subjective choices, and has much lower bias than REF when used with complex UDs. Therefore, ecological studies and many home-range-estimation software programs recommend LSCV as the default bandwidth

[1] *E-mail: gitzenr@missouri.edu*
[2] *Present address: Department of Fisheries and Wildlife Sciences, University of Missouri, Columbia, MO 65211, USA*

method (Seaman and Powell 1996, Kie et al. 1996, Hooge and Eichenlaub 1997, Seaman et al. 1998, 1999).

In contrast, some statisticians consider plug-in and solve-the-equation (STE) techniques to be far better default selection methods than LSCV (Wand and Jones 1995, Jones et al. 1996a,b). The LSCV method has high sampling variability (Wand and Jones 1995) and a tendency to choose low bandwidth values that produce a fragmented UD (Kie et al. 1996, Blundell et al. 2001, Kernohan et al. 2001). Plug-in and STE methods reduce sampling variability. However, they often oversmooth slightly (Wand and Jones 1994, Loader 1999a). If the true distribution is highly fragmented, seemingly low values of h_{LSCV} may reflect the distribution accurately, whereas the newer methods are likely to choose large bandwidths that smooth over the true narrow peaks (Loader 1999b). Therefore, Loader (1999a) argues that LSCV should be the preferred bandwidth method. Space-use studies have used plug-in and STE methods (Roloff et al. 2001, Amstrup et al. 2004), but more extensive comparisons with LSCV are needed, particularly given the disagreement among statisticians regarding which approaches are best.

We compared the performance of plug-in and STE approaches to LSCV and REF. Our purpose was to determine how these methods compare across various distributions and sample sizes. Using simulations, we assessed the performance of these bandwidth methods for fixed-kernel estimation of space use with sample sizes typical of Very High Frequency (VHF) transmitter studies. We compared how well each method estimated home-range size at the 0.95 and 0.25 probability contours and examined the general overlap of estimated and true distributions.

Methods

Simulations

Similar to previous studies (Boulanger and White 1990, Seaman and Powell 1996, Seaman et al. 1999), we used mixtures of bivariate normal distributions to model animal location patterns. By varying the degree of clumping of simulated locations, we created distribution types that would approximate a range of real UDs. Our simulations followed a $4 \times 3 \times 3$ factorial design, with factors of distribution type (general, partially clumped, all clumped, nest tree), number of component normals (2, 4, 16), and sample size (20, 50, 150).

For each bivariate normal distribution in a mixture, we chose the x- and y-variance parameters separately from a uniform distribution with a range that depended on the distribution type. By modifying this range, we created 4 distribution types intended to represent animals that 1) use a home range with relatively broad areas of higher activity (*general*), 2) forage broadly but have a few patches of much higher use (*partially clumped*), 3) focus intensively on localized patches without spending much time in between these patches (*all clumped*), or 4) use a broad home range, but with a single small patch of highly concentrated use (*nest tree*).

After examining preliminary point sets generated from bivariate normal mixtures, we established the following ranges for the variances to produce these 4 general distribution types (Fig. 1). For the general distribution
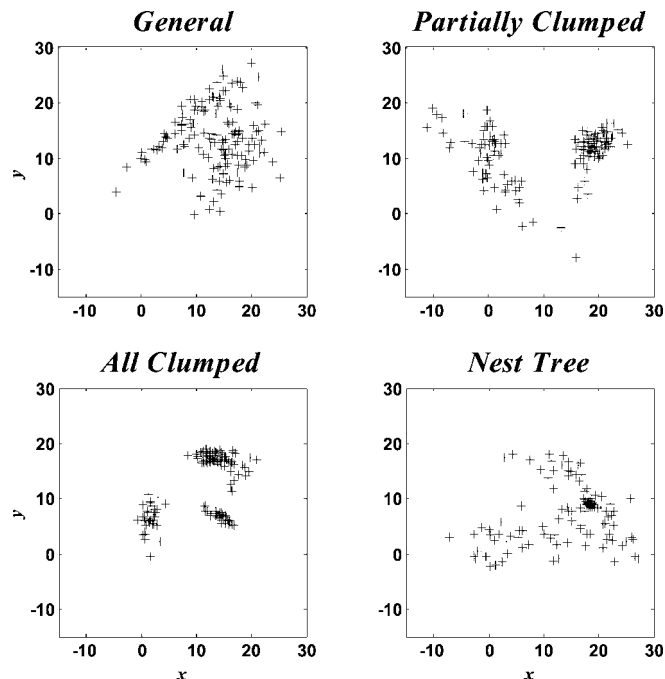
Figure 1. Examples of point patterns from 4 types of mixture distributions used for bandwidth-comparison simulations. Each plot shows 150 points generated from mixtures of 4 bivariate normal distributions for each distribution type.

type, we randomly chose the variance for each coordinate in each normal in the mixture between 1 and 36, as in Seaman et al. (1999). For the partially clumped type, we chose half of the normals in the mixture as in the general type, and we chose the variances for the other normals in the interval (0.5, 5). For the all-clumped type, we chose all normals in the mixture with variances in (0.5, 5). Finally, for the nest-tree type, we chose one bivariate normal with variances in (0.1, 0.4) and we chose the other bivariate normals as in the general class. For all types, we used a correlation value generated from the uniform distribution in (−1, 1) to calculate the covariance $\sigma_{x,y}$ for each bivariate normal. For the mixing proportions in the general, partially clumped, and all-clumped distribution types, we chose values from the uniform distribution between 0 and 1, and then rescaled these values to produce mixing proportions that summed to 1. For the last type, we assigned the nest-tree cluster a mixing proportion of 0.33, and the other bivariate normals in the mixture had randomly chosen proportions constrained to sum to 0.67. For all types, we chose the x- and y-mean parameters separately for each bivariate normal from the uniform distribution between 0 and 20.

The number of component normals partially determined the complexity of the mixtures, producing relatively simple (mixtures of 2 normals) or complex (4 and 16 normals) distributions. We chose the 3 focal sample sizes to determine whether the relative performance of the bandwidth methods changed between levels considered to be inadequate (20), minimally adequate (50), and large (150), respectively, for kernel estimators in home-range studies that use VHF transmitters (Seaman et al. 1999). We rounded point locations generated from each distribution to

4 decimal places to mimic the degree of discretization present in real data sets. We examined estimated locations of elk (*Cervus elaphus*) in South Dakota (Millspaugh et al. 2000) to decide on the degree of rounding.

Researchers apply home-range estimators to a diversity of species and location patterns. Therefore, we focused on comparative bandwidth performance across a large number of distributions with randomly chosen parameters, rather than exploring a more limited set of distributions in detail. We performed 500 replications per combination of distribution type, number of normals, and sample size. For each replication, we selected a new mixture distribution, generated a random sample of points from this distribution, and calculated bandwidth values. Overall, we generated 18,000 utilization distributions.

Kernel and Bandwidth Calculations
We limited our comparisons to a subset of the numerous general variations of the kernel method. Here, we briefly describe some options and the specific forms we chose. First, the bandwidth can remain constant for all data points (fixed kernel), or vary depending on the number of nearby observations (adaptive kernel). We focused on the fixed-kernel method, which generally has lower bias at outer contours of the home range and better surface overlap (Seaman et al. 1999), at least compared to the form of the variable-kernel approach implemented by Seaman and Powell (1996). Second, for bivariate data the bandwidth can be composed of a single value, which applies the same amount of smoothing to each coordinate; 2 values to smooth differently along the *x*- and *y*-axes; or 3 values (a bandwidth matrix) to allow smoothing along rotated coordinate axes (Wand and Jones 1995). We considered only bandwidths composed of separate values for the *x*- and *y*-coordinates, partially for computational ease. Third, different kernel functions are available, but are likely to perform similarly. We used a normal kernel for all calculations (Silverman 1986:43).

Several bandwidth selection methods, including LSCV, plug-in, and REF, have both univariate and bivariate (or generally multivariate) versions. The STE method has only a univariate version. For bivariate data, an alternative strategy to using the bivariate version of these methods is to apply the univariate version separately to the *x*- and *y*-data vectors to produce the bandwidth vector (h_x, h_y). Because STE performs well for univariate data (Wand and Jones 1995, Jones et al. 1996a, Venables and Ripley 2002), applying it separately to the *x*- and *y*-vectors may be a viable strategy for bivariate data (e.g., Roloff et al. 2001). We considered STE and both the bivariate plug-in (PI_{2d}) and the univariate plug-in (PI_{1d}). We did not include the univariate LSCV approach because its performance is similar to or worse than bivariate LSCV approaches (Gitzen and Millspaugh 2003).

All bandwidth methods we considered estimate the value that minimizes the discrepancy between the estimated and true distribution. The plug-in approach "minimizes the [error] function theoretically and then estimates this minimizing value directly" (Bowman and Azzalini 1997:35). The equation for the minimum contains functions of the unknown density. The plug-in approach uses pilot bandwidths to estimate these function values, and the estimates are "plugged in" to the equation for the ideal bandwidth. The STE method follows a strategy similar to the plug-in method. It assumes that the bandwidths used to estimate a function of the unknown density can be scaled as a fixed proportion of the ideal bandwidth (Venables and Ripley 2002), which is estimated with an iterative loop.

We used a 2-stage approach for the plug-in and STE calculations (Appendix). The univariate plug-in (PI_{1d}) and STE methods followed the steps of Wand and Jones (1995). Ćwik and Koronacki (1997) noted that a univariate bandwidth method requires modification when applied separately to each vector of a multidimensional data set. For bivariate data, they raised each coordinate bandwidth to the 0.833 power. Our trial simulations indicated that for both PI_{1d} and STE, this transformation increased overlap of true and estimated distributions, and we used this approach for all simulations. The bivariate plug-in method (PI_{2d}) followed the steps of Wand and Jones (1994).

In addition to PI_{2d}, PI_{1d}, and STE, we calculated the reference bandwidth (REF) and 2 variants of LSCV. The reference approach estimates the optimal bandwidth by assuming data come from a normal distribution, producing an estimate based simply on the sample size and standard deviation for each coordinate of the observations. We calculated REF as in Seaman and Powell (1996): $h_{REF} = n^{-0.167} \times (\sigma_x, \sigma_y) = (h_x, h_y)$, where *n* equals the number of locations. The LSCV method uses a resampling, cross-validation approach to estimate the bandwidth that minimizes error between true and estimated distributions. We calculated the LSCV bandwidth with the scaling approach (Silverman 1986, Wand and Jones 1993) used by Seaman and Powell (1996). We standardized data for both coordinates to have unit variances, we minimized a score function based on a single value of *h*, and we multiplied *h* by the unscaled standard deviations for each coordinate vector to produce (h_x, h_y). For home-range estimation, this approach is similar to several other LSCV variants (Gitzen and Millspaugh 2003).

The LSCV approach minimizes a score function within a suitable interval, typically with the upper and lower bounds of the search interval set to fixed proportions of h_{REF} for the data set (Silverman 1986). In some cases, the score function may descend smoothly to the lower bound of the search interval, indicating failure of the LSCV method (unless the lower bound was too high). The score function may have multiple local minima, and statisticians recommend using the largest local minimum (i.e., the largest bandwidth at which there is a local min. in the score function) instead of the global minimum (Wand and Jones 1995). However, previous simulations indicated that this issue is unimportant for distributions without strongly clumped components (Gitzen and Millspaugh 2003). We included both the global ($LSCV_{gm}$) and largest local minimum ($LSCV_{llm}$) for comparison across the wider range of point patterns

considered here. We looked for the global minimum in the score function by directly searching over 300 possible bandwidth values equally spaced in the search interval ($0.1 \times h_{REF}$, $2 \times h_{REF}$). Although relatively inefficient compared to algorithms such as the 'golden search' method used in home-range software (e.g., Hooge and Eichenlaub 1997), our strategy was suitable for automatic application to thousands of simulated data sets and facilitated the search for local minima at bandwidths larger than the global minimum. For each score value, we calculated the difference $score_{(i)} - score_{(i-1)}$. We took the largest bandwidth value for which this difference was negative as the largest local minimum. When the minimum value of the score function was at a boundary of the search interval, we classified this as failure of the LSCV method, but set the bandwidth equal to h_{REF} for both $LSCV_{gm}$ and $LSCV_{llm}$.

We wrote functions to calculate bandwidth values and perform all other simulation steps in MATLAB 5.3 (MathWorks 1999). We used Beardah and Baxter's (1995) and Bowman and Azzalini's (1997) bandwidth functions as references. We used a function from Beardah and Baxter (1995) to calculate kernel-density estimates.

Bandwidth Comparisons

For each replication, we calculated the true mixture density and the estimated fixed-kernel density for each bandwidth value across a fine-scale grid (231×231, or 53,361 grid points, for the general, partially clumped, and all-clumped distributions; 251×251 for the nest-tree type). For all distributions, the evaluation grid covered the square region within the interval ($-20 \leq x \leq 40$, $-20 \leq y \leq 40$). Trial simulations indicated that the 231×231 grid was adequate for calculating density values at peaks in the true UD for all distributions, except the nest-tree type. We needed a finer grid to ensure complete coverage of the narrow nest-tree peak in the UD. To compare how bandwidths performed at outer and inner UD contours, we calculated the areas within the 0.95 and 0.25 probability contours for the true distribution and for the distributions estimated from the 6 bandwidth methods. We focused on these contours because researchers commonly use the 0.95 contour for calculating home-range size, while the 0.25 contour encompassed portions of the home range with highest use.

For each replicate, we calculated absolute percent error between the true and estimated areas for each contour level, defined as $100 \times$ [absolute value (estimated home-range size − true home-range size)]/true size. Both the accuracy and precision of the estimator affected this metric. For the 0.95 contour, we also calculated percent relative error, defined as [$100 \times$ (estimated size − true size)]/true size. To measure the overall fit of estimated and true distributions, we used the Volume of Intersection Index (Seidel 1992, Millspaugh et al. 2000, 2004):

$$VI = 100\% \times \int_{-20}^{40} \int_{-20}^{40} \min\left[f_1(x,y), \hat{f}_2(x,y)\right] dx\, dy,$$

where f_1 is the true mixture distribution and $\hat{f}_2$ is the estimated distribution. The VI statistic is bounded between

0 (i.e., no overlap of the distributions) and 100 (complete overlap). We used the VI Index because it gives an intuitive, readily interpretable summary of both the overall performance of a bandwidth method and the practical importance of differences among bandwidth methods in distribution overlap. To obtain the VI Index value, we estimated the shared probability mass (volume of overlap) within each cell of the evaluation grid, and then summed these volumes. The limits of the evaluation grid (-20, 40) contained 0.995–1.00 of each UD's volume. Including areas with little or no use does not change the VI Index because these areas contribute nothing to the shared volume of overlap.

Many simulation runs contained a few extremely high error values. Therefore, we reported medians for absolute error values at the 0.95 and 0.25 probability contours, and used bootstrapping (Efron and Tibshirani 1993) to estimate standard errors of medians. We calculated 10,000 bootstrap samples in S-PLUS 2000 (Mathsoft 1999) from the 500 replicates in each combination of design factors. For relative error at the 0.95 probability contour, we calculated medians by distribution type and sample size, pooled across the number-of-normals factor. The VI scores were relatively symmetric without values close to 0 and 100, and we reported means and observed standard errors of means for this measure.

For LSCV, we counted how frequently there was a local minimum larger than the global minimum. For replicates in which we found a larger local minimum, we computed the average pairwise difference ($LSCV_{gm} - LSCV_{llm}$) in absolute error and VI scores. We used means rather than medians for these differences because their distribution was relatively symmetric and without extreme values. Because $LSCV_{gm}$ and $LSCV_{llm}$ were rarely different, and because home-range software programs likely use the global minimum approach, we presented only results for $LSCV_{gm}$ in comparisons with other bandwidth methods.

Results
Overall Patterns
For all bandwidth methods, absolute error increased from the 0.95 to 0.25 contours and with decreasing sample size (Figs. 2, 3). Absolute error increased as the tightness of clustering for some components of each mixture increased (i.e., from general to partially clumped to all-clumped and nest-tree types). For example, when averaged across all combinations of sample sizes and number of normals, the median absolute error ranged from 23% (general) to 71% (nest-tree) for LSCV, and from 18% (general) to 48% (all-clumped) for PI_{2d}. Absolute error was high at the 0.25 contour for the nest-tree type, with median absolute error values exceeding 1,000% for REF. Distribution complexity (no. of normals) had minor effects on bandwidth performance for the general and partially clumped types, with median absolute errors for LSCV, plug-in, and STE methods varying by 1–5% among the complexity levels. For the all-clumped and nest-tree types, absolute error of these bandwidths varied by 4–45% among complexity levels.

All bandwidths produced positively biased estimates of home-range size at the 0.95 contour, except LSCV for the
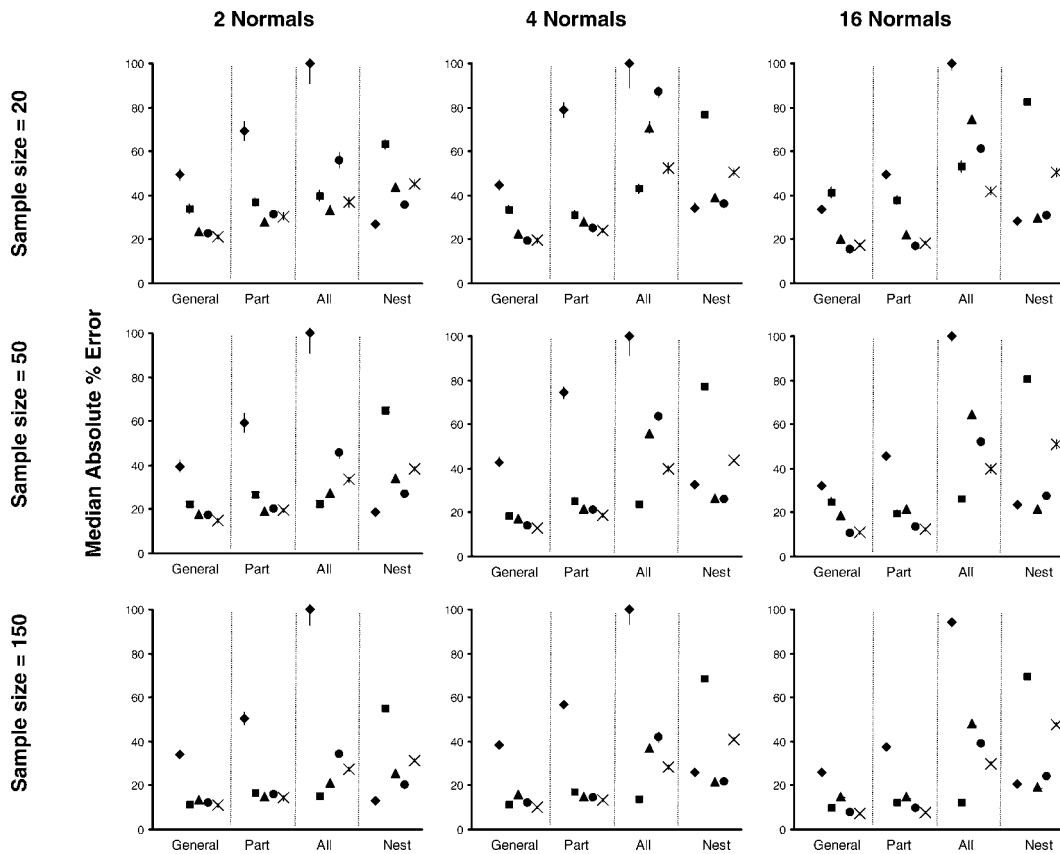
Figure 2. Median (±1 SE) absolute percent error in home-range size estimates for the 0.95 probability contour by bandwidth method, distribution type (general, partially clumped, all-clumped, nest-tree), number of normals in the mixture distributions (2, 4, 16), and sample size (20, 50, 150). Values are from 500 mixture distributions for each combination of distribution type, number of normals, and sample size. For each replicate, we calculated absolute percent error as absolute value [(estimated size − true size)/true size × 100]. We bootstrapped each set of 500 replicates with 10,000 re-samples to estimate standard errors of the medians. Median error values exceeding 100% are set to 100%. Where error bars are not visible, they are smaller than the symbol for the median. Symbols: ♦ = Reference method; ■ = Least-squares cross-validation, global minimum; ▲ = Bivariate plug-in; ● = Univariate plug-in; × = Solve-the-equation.

partially clumped type, STE with a sample size of 20 for the general and partially clumped types, and LSCV, plug-in methods, and STE for the nest-tree type (Table 1). Relative and absolute error did not decrease uniformly with increasing sample size, particularly for the plug-in and STE approaches. For the general and partially clumped distribution types, median relative errors of plug-in and STE methods increased by 1–10% when sample sizes increased from 20–150. However, absolute error of these methods decreased by 8–20% with this increase in sample size. The average overlap of estimated with true distributions, as measured by the VI scores, increased consistently by 6–7% with each increase in sample size (i.e., from 20 to 50 and from 50 to 150; Fig. 4). Overlap scores were high, with average scores between 60–80% for the general, partially clumped, and all-clumped distributions. Overlap decreased as clumping increased, with the nest-tree distribution having average VI scores for each bandwidth method that were lower by 14–25% than corresponding VI scores for the general type.

All Bandwidths

The distribution type substantially affected the comparative performance of the bandwidth methods we evaluated. For the general and partially clumped types at the 0.95 contour,

the plug-in and STE bandwidths produced median absolute error values 5–10% lower than LSCV on average (Fig. 2). At the 0.25 contour, plug-in and STE methods had absolute error values that were 5–15% lower than LSCV for the general type but 6–30% higher than LSCV for the partially clumped distributions (Fig. 3). At sample sizes of 50 and 150 with the all-clumped distributions, LSCV had median absolute error values that were 8–25% lower on average than the plug-in and STE methods at the 0.95 contour, and 14–39% lower at the 0.25 contour. At the 0.95 contour of the nest-tree distributions, LSCV produced median absolute error values 27–46% higher than other methods. At the 0.25 contour of the nest-tree type, LSCV had high overall median absolute error (140%), but plug-in and STE methods had absolute error values 3–5 times higher. In contrast, REF had the lowest median absolute errors for the nest-tree type at the 0.95 contour, but produced error values of 600–3,700% at the 0.25 contour. For other distribution types, REF had median absolute and relative error values that were at least twice as high as other bandwidth methods at the 0.95 contour (Figs. 2, 3).

For overlap scores, the relative rankings of all bandwidth methods were similar to the patterns for outer home-range contours, but the magnitudes of differences were much
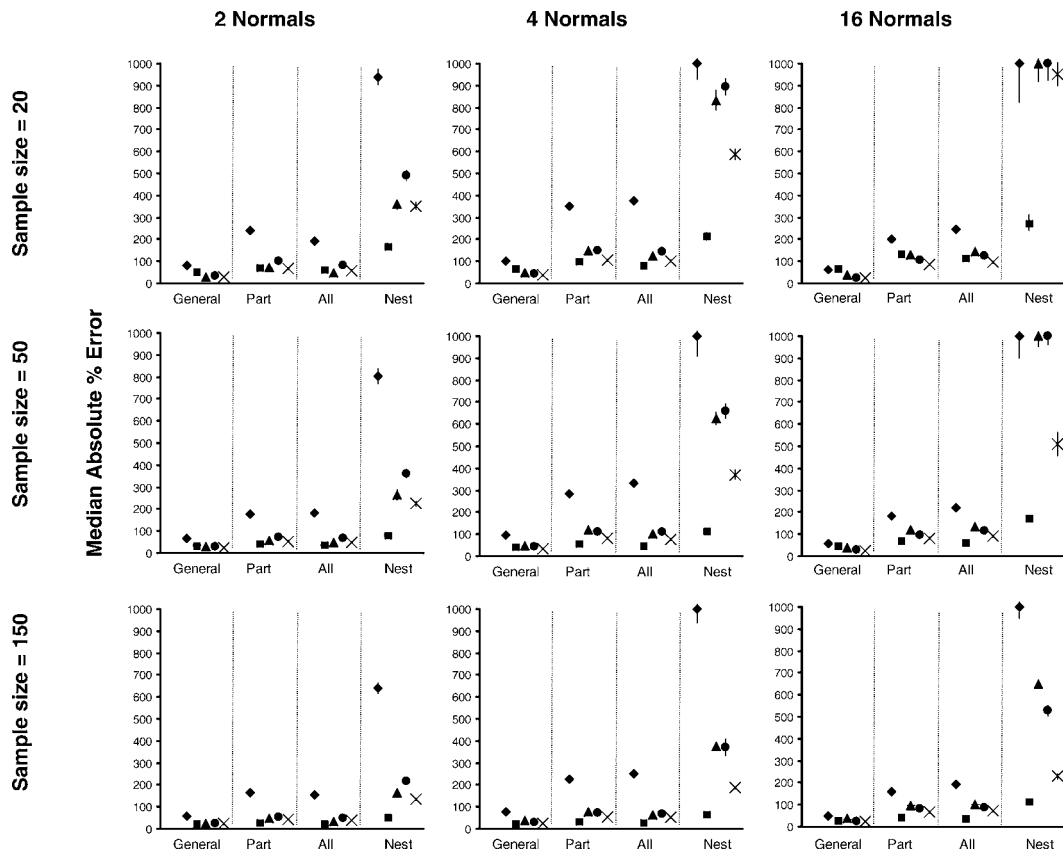
Figure 3. Median (±1 SE) absolute percent error in home-range size estimates for the 0.25 probability contour by bandwidth method, distribution type, number of normals in the mixture distributions, and sample size. Medians exceeding 1,000% are set to 1,000%. Symbols: ◆ = Reference method; ■ = Least-squares cross-validation, global minimum; ▲ = Bivariate plug-in; ● = Univariate plug-in; × = Solve-the-equation. See Figure 2 for additional information.

smaller with the VI Index (Fig. 4). The overall similarity among the bandwidth methods in overall overlap also was indicated by the relatively low variability of the 6 VI scores (i.e., one score for each bandwidth method) for each replicate. For example, across all 18,000 replicates, the average within-replicate coefficient of variation was 74% for absolute error at the 0.95 contour and 76% at the 0.25 contour, but only 7% for the VI Index. Average differences in VI scores between LSCV versus plug-in and STE methods were only 1–2% for the general, partially clumped, and all-clumped distributions. However, for the nest-tree type, LSCV had VI scores 5–8% lower than the plug-in and STE methods. For the general distribution type, the REF method had VI scores only 1–3% lower, on average, then the other methods. With partially and all-clumped distributions, REF did worse, producing average VI scores that were 3–16% lower than the LSCV, plug-in, and STE methods.

Least-Squares Cross-Validation

Both LSCV options failed at the same rates, 13% for the nest-tree distribution (572/4,500, pooled across other factors), but only 0.3% (41/13,500) for the other distributions. Failure rate increased with increasing sample size, with 92, 192, and 329 failures for sample sizes of 20, 50, and 150, respectively (6,000 replicates per sample size). Similarly, failure rate increased with more complex distributions,

with 71, 192, and 350 failures for mixtures of 2, 4, and 16 normals (6,000 replicates each).

A local minimum larger than the global minimum was present rarely (0.4% of simulations, 78/18,000). A larger local minimum occurred at similar rates across the distribution types (summed across all sample sizes and no. of normals: 26 cases for general, 20 for partially clumped, 17 for all-

Table 1. Median percent relative error[a] for 0.95 probability-contour home-range size estimates by bandwidth method, distribution type, and sample size (n).

Distribution type	n	REF[b]	LSVG$_{gm}$	PI$_{2d}$	PI$_{1d}$	STE
General	20	41.1	26.2	10.1	3.7	−3.5
	50	37.3	15.1	15.1	7.6	3.6
	150	31.2	7.1	14.4	9.2	6.7
Partially clumped	20	62.2	0.7	10.2	8.6	−4.6
	50	56.8	−7.4	14.9	11.7	1.8
	150	45.0	−10.0	11.3	8.6	2.6
All-clumped	20	150.9	40.7	58.3	67.0	42.8
	50	133.8	17.8	49.6	53.0	38.0
	150	111.6	10.5	37.4	38.4	28.5
Nest-tree	20	13.1	−74.5	−34.9	−30.2	−46.6
	50	20.0	−74.1	−26.3	−25.3	−44.2
	150	19.6	−63.6	−22.1	−21.6	−40.6

[a] Results summarize 1,500 replications for each combination of distribution type and sample size.
[b] REF = reference method; LSVG$_{gm}$ = least-squares cross-validation, global minimum; PI$_{2d}$ = bivariate plug-in; PI$_{1d}$ = univariate plug-in; STE = solve-the-equation.
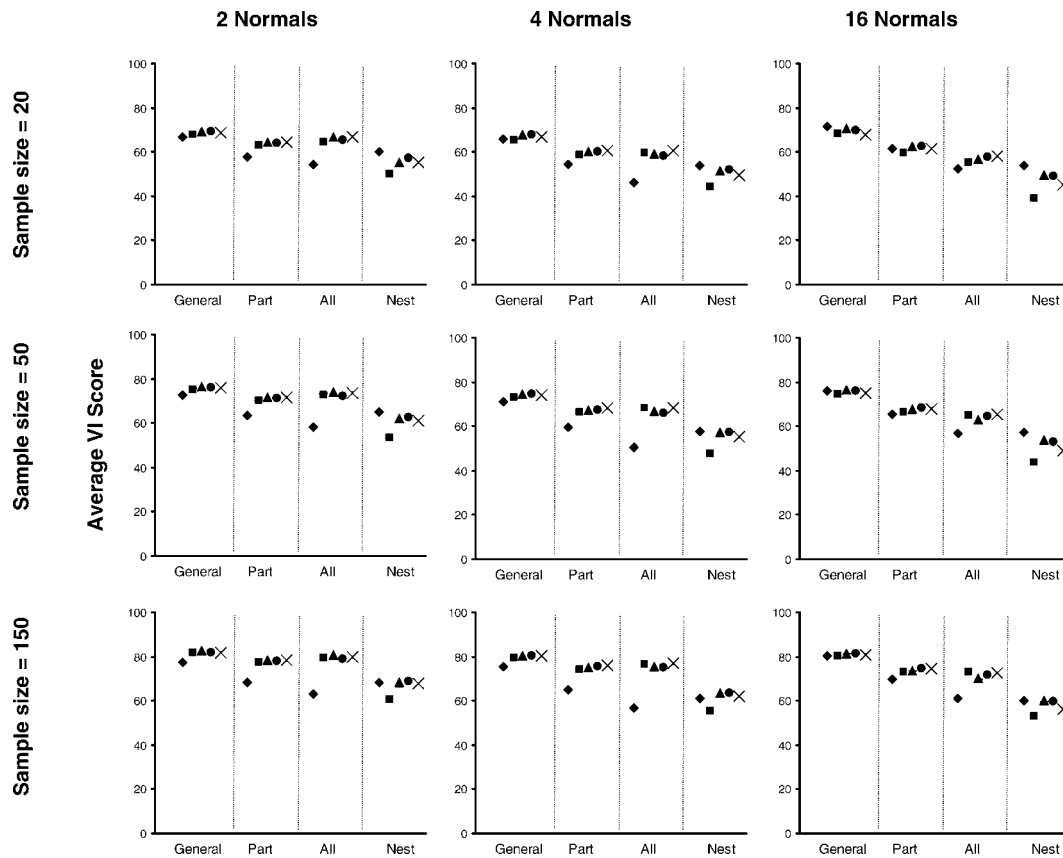
Figure 4. Average Volume of Intersection scores measuring overlap of estimated with true distributions by bandwidth method, distribution type, number of normals in the mixture distributions, and sample size. A score of 100 would indicate perfect overlap of the estimated and true distributions. Standard error bars are too small to display (average SE = 0.3, range = 0.1–0.7). Symbols: ♦ = Reference method; ■ = Least-squares cross-validation, global minimum; ▲ = Bivariate plug-in; ● = Univariate plug-in; × = Solve-the-equation.

clumped, 15 for nest-tree) and more frequently with complex distributions (2 normals: 21 cases; 4 normals: 16 cases; 16 normals: 41 cases). We found a larger local minimum mainly at small sample sizes (sample size of 20: 61 cases; 50: 15 cases; 150: 2 cases). When a larger local minimum was present, it produced worse estimates at the 0.95 home-range contour for the 78 simulations in which the 2 differed (average difference in absolute error, global − largest local min., = −31%, SE = 8), but better overall distribution overlap (average difference in VI scores = −6.5%, SE = 1.2).

Discussion

Our results indicate that plug-in and STE methods perform as well as or better than LSCV except with distributions composed of multiple tight clumps. The plug-in and STE methods perform much better than REF except with UDs dominated by a single peak area of high use. Researchers frequently should consider using these newer methods in kernel analyses, and home-range software packages should make these methods available. However, overall differences between these methods and LSCV are much lower than we expected. Moreover, no bandwidth method that we considered outperforms the other methods in all situations, at least for sample sizes typical of VHF-telemetry studies.

An important advantage of plug-in and STE methods is that they avoid some complications of LSCV. The LSCV method often chooses low bandwidth values, causing the

estimated UD to break into numerous fragments even when the true distribution is much smoother (Kie et al. 1996, Blundell et al. 2001, Kernohan et al. 2001). The LSCV method sometimes does not provide meaningful bandwidths. For some data sets, particularly when discretization is significant and numerous data points have the same or nearly the same values, the LSCV score function may have a minimum only at a bandwidth value of zero (Silverman 1986, Hooge and Eichenlaub 1997, Seaman et al. 1998), indicating that the method has failed. Moreover, our results probably underestimate this advantage; failure of LSCV may occur more frequently with real data than in our simulations (Tufto et al. 1996, Seaman et al. 1998, Linders 2000, Blundell et al. 2001). When LSCV fails for numerous location sets, the analyst must choose an ad hoc or subjective solution or use methods with higher bias. For example, the researcher may use h_{REF}, an arbitrary scaling of h_{REF} (Worton 1995), add random noise to identical locations to get a sensible LSCV answer (Tufto et al. 1996), or switch to another, potentially inferior, home-range estimation technique such as the minimum convex polygon method. Plug-in and STE approaches provide a valuable alternative in this situation.

Comparative performance of bandwidth methods depended on the distribution type and, to a lesser extent, on the distribution complexity and sample size. In many radio-tracking studies, biologists planning to use kernel methods should choose a bandwidth selection method a priori based on

expected characteristics of the animal's space use and the goals of the study. For example, if the species under investigation forages widely across much of its home range, the researcher should use plug-in or STE methods. These methods are the best default choice when a researcher wishes to identify smooth outer boundaries. For example, in studies that examine resource selection within the general home-range area (use vs. availability), the researcher seeks to delimit a broader area that includes patches of high and low use. In addition, the tendency of the newer methods to oversmooth slightly throughout the distribution makes them intuitively appropriate if there is relatively high error associated with all estimated locations. For example, in mountainous terrain, radiotelemetry locations estimated by triangulation may have large error polygons (White and Garrot 1990:54, 69), and moderate oversmoothing seems appropriate.

In contrast, if the focal species spends most of its time visiting small patches of resources and the study focuses on the area of these clumps, LSCV methods should be the default choice if kernel estimation is used. However, even this recommendation must be qualified. As the degree of clumping increases and clusters begin to contain multiple nearly identical locations, LSCV will fail more frequently. Researchers need further simulations to compare LSCV-based kernel estimators with linkage methods (Kenward et al. 2001) for examining use of discrete patches and core areas. Kenward et al. (2001) noted the risk of circular verdicts against linkage estimators in simulation studies. Data generated from smooth, continuous distributions may better match the statistical models used by kernel-density estimators. When UDs have rigid boundaries or edges between used and unavailable areas, or holes of unusable areas surrounded by used patches, unmodified fixed-kernel estimators may estimate high space use in unavailable areas (Getz and Wilmers 2004). However, we obtained realistic distributions even with bivariate normal mixture distributions. For example, the point patterns we generated could have relatively sharp boundaries between simulated used and unused areas (Fig. 1). Implementation of boundary kernels that handle edge effects (Müller and Stadmüller 1999) or recently developed variable-kernel approaches (Sain 2002) may extend the flexibility of kernel approaches for space-use analyses.

For species that forage widely but spend much of their time at a single den or nest, the reference bandwidth may be the best choice for estimating home-range size at outer probability contours if the nest-tree locations are included in the analysis. The dominant feature of our nest-tree distribution was a single sharp peak, which probably drove the good performance of REF at outer contours of this type. If an animal uses more than one den or nest tree in different parts of its home range, the resulting distribution will have 2 or more sharp peaks, and REF would perform poorly compared to plug-in or LSCV methods.

Further complications arise when animals within a single study show different degrees of clustering in their locations. For example, in many polygynous species during the breeding season, females may focus activity around the nest, but adult males may prospect widely for mates. In this case, use of a single bandwidth method could produce good estimates for one sex but biased estimates for the other sex.

If study animals show such disparate patterns of movements, we recommend the use of different bandwidth methods to produce the best estimate for each study sub-population. Within the same study, different bandwidth methods could be appropriate for different study components. Even with a tightly clumped pattern of locations, use of plug-in or STE methods may be appropriate to identify broader areas for examining resource selection, while LSCV could produce the best estimates of home-range size and identify discrete patches of high use.

Differences among bandwidths are most important when estimating home-range size. Choice of bandwidth method usually had minor effects on VI scores, a pattern noted for REF and LSCV by Seidel (1992) when she developed the VI Index. Similarly, surface fit was similar for REF and LSCV with both fixed and adaptive kernels in simulations of Seaman et al. (1999). In our simulations, all bandwidth methods produced estimated distributions that matched most of the true UD's volume, as indicated by the high VI scores.

This study focused on broad-scale comparisons of bandwidth methods across a large number of randomly selected distributions. Additional insight would be gained from detailed examination of how bandwidths perform in different parts of each distribution (e.g., at the UD peak) and for specific distributions with purposely selected parameters (e.g., Jones et al. 1996a, Loader 1999b). Detailed examination could help explain some counter-intuitive patterns we observed. For example, bandwidth performance often was better with mixtures of 16 normals than with 4 normals. Mixtures of 16 normals often may have been less complex, particularly in cases where the means of many of the normals were close together and produced a single broader peak.

Our results apply to VHF-based telemetry studies in which a large sample size consists of 150 locations per animal. In contrast, studies using Global Positioning System collars and satellite telemetry routinely can collect thousands of observations per animal. We recommend additional examination of bandwidth selection methods and broader kernel-analysis options in the context of such data-rich studies. With such large data sets, efficient computation is essential (Kern et al. 2003). Reasonably, researchers might assume that choice of bandwidth method would be unimportant at such large sample sizes if all methods converged on an estimate of the ideal bandwidth. However, Amstrup et al. (2004) described poor performance of LSCV when applied to thousands of observations in a satellite-telemetry study of polar bears (*Ursus maritimus*). As the number of locations increases, clusters of observations at the same or nearly the same locations are likely. In this situation, the tendency of LSCV to fail may be exacerbated at large sample size. For example, the LSCV method failed most frequently at the highest sample size we evaluated (150). Most of these failures occurred with the nest-tree distribution, which would have a large number of nearly identical points in the nest-tree cluster. Similarly, we expect that the REF method would not improve greatly at high sample sizes. Assuming that the true UD was multimodal, a high number of observations would delineate the multiple peaks of use more clearly. Counter-intuitively, the advantage of plug-in and STE methods may increase at such high sample sizes.

We expect that our comparative rankings of the bandwidth methods would carry over to the adaptive-kernel method implemented by Seaman and Powell (1996). This approach uses pilot fixed-kernel bandwidth estimates (Silverman 1986) that are then adapted to account for local point distributions. Choice of bandwidth method used for estimating these pilot bandwidths might be less important than in fixed-kernel estimation. However, LSCV performed much better than REF for both fixed and adaptive kernels in Seaman et al. (1999). We recommend additional research to examine whether patterns we observed hold true for this and other forms of a variable-kernel approach (Silverman 1986).

Several refinements of the plug-in and STE methods could increase their accuracy and precision. We considered only a 2-stage approach for each method; adding a third stage likely would enhance performance at the cost of additional computing time. For calculating pilot bandwidths, we used scale estimates (standard deviation for the univariate versions, covariance matrix for the bivariate plug-in), that outlier locations influence strongly. Researchers should evaluate a more robust scale estimate (Wand and Jones 1994). Finally, calculation of 3 bandwidth values to allow smoothing away from the coordinate axes could provide additional improvements (Duong and Hazelton 2003, Amstrup et al. 2004). Combined with more fundamental options, such as the choice between fixed and adaptive kernels, and other subtle options, such as the choice of the kernel function, these considerations lead to a bewildering number of variations on the kernel method. As with variants of the LSCV method (Gitzen and Millspaugh 2003), many of these options may have little effect on average estimator performance. However, even relatively unimportant variations still may contribute to variability of estimates among home-range studies and software packages.

Faced with choosing among these options and dealing with bandwidth selection uncertainty, a researcher might consider switching to a simpler method, such as the MCP approach. In most cases, this decision is not justifiable. Previous techniques such as the MCP method may be useful for some studies depending on the objectives, but they generally have higher bias, lower comparability, and greater sample-size dependence than any form of the kernel method discussed above (Kernohan et al. 2001).

Management Implications

When using the kernel approach to summarize animal space use, ecologists should be aware that bandwidth selection could have a large effect on estimates and the resulting biological conclusions. The goal of having a single automatic selection method that works best in any situation probably is not obtainable. Rather, ecologists should recognize that there is high uncertainty in selecting the appropriate bandwidth value. They should carefully consider their analysis options before conducting studies (Kenward et al. 2001, Kernohan et al. 2001), and they should chose a specific bandwidth method a priori based on study goals and the expected distribution patterns of locations. Moreover, in a study that uses the kernel method, investigators should examine whether the biological conclusions of their studies change depending on whether LSCV or one of the plug-in or STE bandwidth methods is used.

Acknowledgments

Literature Cited

Amstrup, S. C., T. L. McDonald, and G. M. Durner. 2004. Using satellite radiotelemetry data to delineate and manage wildlife populations. Wildlife Society Bulletin 32:661–679.

Beardah, C. C., and J. J. Baxter. 1995. MATLAB routines for kernel density estimation and the graphical representation of archaeological data. Anelecta Prehistorica Leidensia 28, Leiden University, Rapenburg, the Netherlands.

Blundell, G. M., J. A. K. Maier, and E. M. Debevec. 2001. Linear home ranges: effects of smoothing, sample size, and autocorrelation on kernel estimates. Ecological Monographs 71:469–489.

Boulanger, J. G., and G. C. White. 1990. A comparison of home-range estimators using Monte Carlo simulation. Journal of Wildlife Management 54:310–315.

Bowman, A. W., and A. Azzalini. 1997. Applied smoothing techniques for data analysis: the kernel approach with S-PLUS illustrations. Oxford University, New York, New York, USA.

Ćwik, J., and J. Koronacki. 1997. A combined adaptive-mixtures/plug-in estimator of multivariate probability densities. Computational Statistics and Data Analysis 26:199–218.

Duong, T., and M. L. Hazelton. 2003. Plug-in bandwidth matrices for bivariate kernel density estimation. Journal of Nonparametric Statistics 15:17–30.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, London, United Kingdom.

Fieberg, J., and C. O. Kochanny. 2005. Quantifying home-range overlap: the importance of the utilization distribution. Journal of Wildlife Management 69:1346–1359.

Getz, W. M., and C. C. Wilmers. 2004. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. Ecography 27:489–505.

Gitzen, R. A., and J. J. Millspaugh. 2003. Comparison of least-squares cross-validation bandwidth options for kernel home-range estimation. Wildlife Society Bulletin 31:823–831.

Hooge, P. N., and B. Eichenlaub. 1997. Animal movement extension to Arcview: version 1.1. U.S. Geological Survey, Alaska Biological Science Center, Anchorage, USA.

Jones, M. C., J. S. Marron, and S. J. Sheather. 1996a. A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91:401–407.

Jones, M. C., J. S. Marron, and S. J. Sheather. 1996b. Progress in data-based bandwidth selection for kernel density estimation. Computational Statistics 11:337–381.

Kenward, R. E., R. T. Clarke, K. H. Hodder, and S. S. Walls. 2001. Density and linkage estimators of home range: nearest-neighbor clustering defines multinuclear cores. Ecology 82:1905–1920.

Kern, J. W., T. L. McDonald, S. C. Amstrup, G. M. Durner, and W. P. Erickson. 2003. Using the bootstrap and fast Fourier transform to estimate confidence intervals of 2D kernel densities. Environmental and Ecological Statistics 10:405–418.

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. Radiotracking and animal populations. Academic, San Diego, California, USA.

Kie, J. G., J. A. Baldwin, and C. J. Evans. 1996. CALHOME: a program for estimating animal home ranges. Wildlife Society Bulletin 24:342–344.

Linders, M. J. 2000. Spatial ecology of the western gray squirrel (*Sciurus griseus*) in Washington: the interaction of season, habitat and home range. Thesis, University of Washington, Seattle, USA.

Loader, C. R. 1999*a*. Bandwidth selection: classical or plug-in? Annals of Statistics 27:415–438.

Loader, C. R. 1999*b*. Local regression and likelihood. Springer-Verlag, New York, New York, USA.

Marzluff, J. M., S. T. Knick, and J. J. Millspaugh. 2001. High-tech behavioral ecology: modeling the distribution of animal activities to better understand wildlife space use and resource selection. Pages 309–326 *in* J. J. Millspaugh and J. M. Marzluff, editors. Radiotracking and animal populations. Academic, San Diego, California, USA.

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

Mathsoft. 1999. S-PLUS 2000 Professional Release 2. Mathsoft, Seattle, Washington, USA.

MathWorks. 1999. MATLAB: the language of technical computing. Version 5.3. The MathWorks, Natick, Massachusetts, USA.

Millspaugh, J. J., G. C. Brundige, R. A. Gitzen, and K. J. Raedeke. 2000. Elk and hunter space-use sharing in South Dakota. Journal of Wildlife Management 64:994–1003.

Millspaugh, J. J., R. A. Gitzen, B. J. Kernohan, M. A. Larson, and C. L. Clay. 2004. Comparability of three analytical techniques to assess joint space use. Wildlife Society Bulletin 32:148–157.

Müller, H. G., and U. Stadmüller. 1999. Multivariate boundary kernels and a continuous least squares principle. Journal of the Royal Statistical Society B 61:439–458.

Roloff, G. J., J. J. Millspaugh, R. A. Gitzen, and G. C. Brundige. 2001. Validation tests of a spatially explicit habitat effectiveness model for Rocky Mountain elk. Journal of Wildlife Management 65:899–914.

Sain, S. R. 2002. Multivariate locally adaptive density estimation. Computational Statistics and Data Analysis 39:165–186.

Seaman, D. E., B. Griffith, and R. A. Powell. 1998. KERNELHR: a program for estimating animal home ranges. Wildlife Society Bulletin 26:95–100.

Seaman, D. E., J. J. Millspaugh, B. J. Kernohan, G. C. Brundige, K. J. Raedeke, and R. A. Gitzen. 1999. Effects of sample size on kernel home range estimates. Journal of Wildlife Management 63:739–747.

Seaman, D. E., and R. A. Powell. 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. Ecology 77: 2075–2085.

Seidel, K. D. 1992. Statistical properties and applications of a new measure of joint space use for wildlife. Thesis, University of Washington, Seattle, USA.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman and Hall, London, United Kingdom.

Swihart, R. K., and N. A. Slade. 1997. On testing for independence of animal movements. Journal of Agricultural, Biological, and Environmental Statistics 2:48–63.

Tufto, J., R. Andersen, and J. Linnell. 1996. Habitat use and ecological correlates of home range size in a small cervid: the roe deer. Journal of Animal Ecology 65:715–724.

Van Winkle, W. 1975. Comparison of several probabilistic home-range models. Journal of Wildlife Management 39:118–123.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Fourth edition. Springer-Verlag, New York, New York, USA.

Wand, M. P., and M. C. Jones. 1993. Comparison of smoothing parameterizations in bivariate kernel density estimation. Journal of the American Statistical Association 88:520–528.

Wand, M. P., and M. C. Jones. 1994. Multivariate plug-in bandwidth selection. Computational Statistics 9:97–116.

Wand, M. P., and M. C. Jones. 1995. Kernel smoothing. Chapman and Hall, London, United Kingdom.

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

Worton, B. J. 1987. A review of models of home range for animal movement. Ecological Modelling 38:277–298.

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

Worton, B. J. 1995. Using Monte Carlo simulation to evaluate kernel-based home range estimators. Journal of Wildlife Management 59: 794–800.

Appendix: Summary of Plug-In and Solve-the-Equation Bandwidth Calculations

Overview

Both plug-in and STE approaches use an estimator for the ideal (error-minimizing) bandwidth. This bandwidth depends on a density functional, a function of the true density and its derivative. These methods calculate this functional with a kernel estimator that requires a suitable bandwidth. This bandwidth in turn depends on a higher derivative functional. A kernel estimator calculates this functional, but this estimator requires yet another suitable bandwidth value. Researchers could repeat these steps for multiple stages; at each stage, the bandwidth depends on a higher derivative functional with a more preliminary bandwidth estimate. The 2-stage PI and STE approaches calculate a preliminary bandwidth based on the sample standard deviation (for univariate approaches) or covariance matrix (for the bivariate plug-in approach). This preliminary bandwidth is plugged into a functional that allows estimation of the first-stage bandwidth, which in turn is plugged into a lower-derivative functional. This estimates a second-stage bandwidth. A functional uses this bandwidth to estimate the final bandwidth. The plug-in methods calculate this final bandwidth analytically. In the STE approach, the final bandwidth is still a function of itself, but a simple iterative search finds the final bandwidth.

Univariate Plug-In

For the univariate plug-in approach, we followed the steps of Wand and Jones (1995:72) for the 2-stage approach with a normal kernel function, applying the steps separately to the vectors of x- and y-coordinates of each set of simulated locations. The plug-in and STE methods required estimates of multiple kernel functionals of the form $\psi_i = \int f^{(i)}(x) f(x) dx$, where $f^{(i)}(x)$ is the ith derivative of the true density $f(x)$. First, we estimated ψ_8 with a simple estimate based on the bandwidth that would be optimal for normal data, producing $\hat{\psi}_8^{NS} = \frac{105}{32\pi^{0.5}(\hat{\sigma})^9}$, where $\hat{\sigma}$ is the square root of the variance of the data vector, a scale estimate. We used this normal-scale functional to estimate the first-stage bandwidth:

$$g_1 = \left[\frac{-2K^{(6)}(0)}{\hat{\psi}_8^{NS} n} \right]^{1/9}$$

where $K^{(6)}(0)$ is the 6th derivative of the standard normal density function evaluated at 0, and n is the number of locations. Next, we used g_1 to estimate $\hat{\psi}_6 = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_{g_1}^{(6)}(x_i - x_j)$. We calculated this estimate as the sixth derivative of the normal kernel function K evaluated at each pairwise difference among the x- or y-vector elements. This was a kernel estimator with bandwidth g_1. We estimated the optimal second-stage bandwidth:

$$g_2 = \left[\frac{-2K^{(4)}(0)}{\hat{\psi}_6 n} \right]^{1/7}.$$

We used this bandwidth to estimate the functional $\hat{\psi}_4 =$

$n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_{g2}^{(4)}(x_i - x_j)$. Finally, we calculated the plug-in bandwidth estimate:

$$h_{PI} = \left[\frac{\frac{1}{2\sqrt{\pi}}}{\hat{\psi}_4 n} \right]^{0.20}.$$

Because we were applying this univariate method to each coordinate of bivariate data, we transformed this estimate, using $h'_{PI} = (h_{PI})^{5/6}$.

Solve-the-Equation

For the STE approach, we first calculated preliminary estimates of 2 functionals (Wand and Jones 1995:74–75): $\hat{\psi}_8^{NS} = \frac{105}{32\pi^{0.5}(\hat{\sigma})^9}$ and $\hat{\psi}_6^{NS} = \frac{-15}{16\pi^{0.5}(\hat{\sigma})^7}$. We then calculated intermediate bandwidths, again with a normal kernel:

$$g_1 = \left[\frac{-2K^{(4)}(0)}{\hat{\psi}_6^{NS} n} \right]^{1/7}$$

and

$$g_2 = \left[\frac{-2K^{(6)}(0)}{\hat{\psi}_8^{NS} n} \right]^{1/9}.$$

We estimated the functionals ψ_4 and ψ_6 with bandwidths g_1 and g_2 following the same steps used for the univariate plug-in approach. The remaining step required a function of the final bandwidth h:

$$\gamma(h) = \left[\frac{2K^{(4)}(0)\hat{\psi}_4}{-(\hat{\psi}_6)\frac{1}{2\sqrt{\pi}}} \right]^{1/7} h^{5/7}.$$

The bandwidth h was the solution to the function

$$h_{STE} = \left[\frac{\frac{1}{2\sqrt{\pi}}}{\left(n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} K_{\gamma(h_{STE})}^{(4)}(x_i - x_j) \right)} \right]^{0.20}.$$

We used an iterative search to find h_{STE}, with the search halting when successive iterations changed the estimated bandwidth by < 1%. As with the univariate plug-in, we transformed this estimate, using $h'_{STE} = (h_{STE})^{5/6}$.

Bivariate Plug-In

The bivariate plug-in method followed the same steps as the univariate approach. However, the bivariate kernel functionals were more complex because they required partial derivates of the density and standard bivariate normal kernel. We used the following steps (Wand and Jones 1994):

1) We standardized the bivariate location data with original standard deviation ($\hat{\sigma}_x$, $\hat{\sigma}_y$) to have $\hat{\sigma} = 1$ for both coordinates. We calculated the correlation coefficient between the standardized x- and y-vectors, ρ.

2) We calculated quick estimates of kernel functionals $\psi_{8,0}^{NS}$, $\psi_{0,8}^{NS}$, $\psi_{6,2}^{NS}$, $\psi_{2,6}^{NS}$, $\psi_{4,4}^{NS}$ based on the normal-scale approach:

$$\psi_{m_1,m_2}^{NS} = \frac{(-1)^{(m_1+m_2)/2} \lambda_{m_1 m_2}}{2^{(m_1+m_2+4)/2} \pi (1 - \rho^2)^{(m_1+m_2+1)/2}}$$

where

$$\lambda_{2r,2s} = \frac{(2r)!(2s)!}{2^{r+s}} \sum_{j=0}^{\min(r,s)} \frac{(2\rho)^{2j}}{(r-j)!(s-j)!(2j)!}.$$

3) We used the normal-scale functionals to calculate the stage-1 bandwidth values ($a_{6,0}$, $a_{0,6}$, $a_{4,2}$, $a_{2,4}$):

$$a_{m_1,m_2} = \left[\frac{-2K^{(m_1,m_2)}(0)}{\left(\psi_{(m_1+2,m_2)}^{NS} + \psi_{(m_1,m_2+2)}^{NS} \right) n} \right]^{\frac{1}{(4+m_1+m_2)}}$$

where

$$K^{(m_1,m_2)}(0) = (-1)^{m_1+m_2} \psi_{m_1,m_2}^{NS}(0.5).$$

4) We used these bandwidths to estimate kernel functionals $\psi_{6,0}$, $\psi_{0,6}$, $\psi_{4,2}$, $\psi_{2,4}$: $\hat{\psi}m_1, m_2 = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_{a_{m_1,m_2}}^{(m_1,m_2)}(x_i - x_j, y_i - y_j)$. This formula denotes that we plugged all pair-wise differences among the standardized set of locations into the (m_1, m_2) partial derivative of the bivariate normal kernel function evaluated with bandwidth a_{m_1,m_2}.

5) We used these functional estimates to calculate the stage-2 bandwidth values ($a_{4,0}$, $a_{0,4}$, $a_{2,2}$) and then functional estimates $\psi_{4,0}$, $\psi_{0,4}$, and $\psi_{2,2}$, with the equations used in 3) and 4), using ψ_{m_1,m_2} in place of ψ_{m_1,m_2}^{NS}.

6) We plugged these functional estimates into the equation for (h_1, h_2):

$$h_1 = \left[\frac{(\psi_{0,4})^{0.75}(\frac{1}{4\pi})}{(\psi_{4,0})^{0.75} \left[(\psi_{4,0})^{0.5}(\psi_{0,4})^{0.5} + \psi_{2,2} \right] n} \right]^{\frac{1}{6}}$$

and

$$h_2 = h_1 (\psi_{4,0}/\psi_{0,4})^{0.25}.$$

Finally, we transformed these values back to the original scale of the location data: $(h_1, h_2) = (\hat{\sigma}_x h_1, \hat{\sigma}_y h_2)$.

Associate Editor: Strickland.