Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator

C. H. FLEMING,1,2,5 W. F. FAGAN,2 T. MUeller,1,2,3,4 K. A. OLSON,1 P. LEMGRUBER,1 AND J. M. CALABRESE1

1Conservation Ecology Center, Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, Virginia 22630 USA
2Department of Biology, University of Maryland College Park, College Park, Maryland 20742 USA
3Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt, Germany
4Department of Biological Sciences, Goethe University Frankfurt, Max-von-Laue-Strasse 9, 60438 Frankfurt, Germany

Abstract. Quantifying animals' home ranges is a key problem in ecology and has important conservation and wildlife management applications. Kernel density estimation (KDE) is a workhorse technique for range delineation problems that is both statistically efficient and nonparametric. KDE assumes that the data are independent and identically distributed (IID). However, animal tracking data, which are routinely used as inputs to KDEs, are inherently autocorrelated and violate this key assumption. As we demonstrate, using realistically autocorrelated data in conventional KDEs results in grossly underestimated home ranges. We further show that the performance of conventional KDEs actually degrades as data quality improves, because autocorrelation strength increases as movement paths become more finely resolved. To remedy these flaws with the traditional KDE method, we derive an autocorrelated KDE (AKDE) from first principles to use autocorrelated data, making it perfectly suited for movement data sets. We illustrate the vastly improved performance of AKDE using analytical arguments, relocation data from Mongolian gazelles, and simulations based upon the gazelle's observed movement process. By yielding better minimum area estimates for threatened wildlife populations, we believe that future widespread use of AKDE will have significant impact on ecology and conservation biology.

Key words: autocorrelation; Brownian bridge; home range; kernel density; minimum convex polygon; Mongolian gazelle, Procapra gutturosa; tracking data; utilization distribution.

INTRODUCTION

Home range estimation, a critical statistical challenge, applies to areas of ecology ranging from theoretical ecology to wildlife management. Whether quantifying space use or designing conservation strategies, ecologists need to know what habitats an animal uses in terms of both location and extent. Animal tracking data increasingly constitute the key inputs into home range estimation procedures. Conventional methods of home range estimation largely fall into two camps: geometric techniques, such as the minimum convex polygon (MCP; Bekoff and Mech 1984, Fieberg and Borger 2012), that lack an underlying probabilistic model, and statistical techniques that were not developed for use with animal tracking data, such as kernel density estimators (KDEs; Worton 1989). While KDEs are the most efficient nonparametric estimators of probability density functions (PDFs), they are derived under the assumption of independent and identically distributed (IID) data, an assumption violated by autocorrelation and nonstationarity (Silverman 1986). When faced with realistic, autocorrelated movement data, KDEs have been observed (Swihart and Slade 1985, Hansteen et al. 1997) and proven (Fleming et al. 2014a) to underestimate home range area, often dramatically (Fig. 1). Common suggestions for dealing with autocorrelated location data include coarsening the sampling rate (Swihart and Slade 1985) and stratification across individuals (Otis and White 1999), but these types of adjustments are generally inefficient.

Autocorrelation means that an individual’s position, velocity, or acceleration measured at one point in time are statistically correlated with the same measurements in the past, and also implies that these correlations will carry on into the future. Autocorrelation is the rule, not the exception, in animal movement data. Autocorrelation can arise from diverse sources and may occur over multiple timescales in a single individual’s movement path. First, and most fundamentally, autocorrelation is an ineluctable consequence of the fact that animal movement is a continuous process. Uncorrelated location data would lack any degree of continuity, whereas real animals trace continuous paths through the environment and have continuous velocity and acceleration. Second, autocorrelation arises when individuals...
continue a particular movement behavior for an extended period of time or repeat certain behaviors such as revisiting the same foraging areas, dens, or nesting sites. Typically, correlations diminish as observations grow farther apart in time, but autocorrelations in movement data often persist over long time periods, e.g., months or years (McNay et al. 1994, Rooney et al. 1998, Boyce et al. 2010, Fleming et al. 2014a, b).

The conceptual definition of home range given by Burt (1943) lacks an objective, mathematical description that can be statistically estimated from data. We view relocation data as a sample of a much longer, continuous trajectory that is only one of many possible movement-path realizations of a continuous-time stochastic process. In this formalism, movement-path realizations that exhibit realistic behaviors and result in heavy use of the observed animal’s core areas are assigned higher probabilities than those that stray off into little-used regions. This definition thus operationalizes Burt’s intuition that rare excursions should not be included in the home range by down-weighting such excursions. We therefore define the home range area as a percent coverage region, usually taken to be 95%, of the probability distribution of all possible locations, as determined from the distribution of all possible paths (hereafter, range distribution). The range distribution addresses the lifetime space requirements of an animal and provides a metric that can be compared across individuals. Unfortunately, the range distribution is frequently conflated with the occurrence distribution, which does not quantify the home range but instead estimates where an animal was located during the observation period (see Appendix A for a detailed discussion of these two distributions, their differences, and their estimators). While both of these distributions are sometimes referred to as the utilization distribution, in what follows, we focus only on the range distribution, as it is closest to Burt’s original definition.

Natural statistical intuitions often fail in the presence of autocorrelation. In a random sample of $n$ indepen-
dent observations from an individual’s position distribution, on average, we expect 0.95 \( n \) observations to fall inside the estimated distribution’s 95% contour and 5% to fall outside. However, this is not the case for autocorrelated data, which contain less positional information than an equivalently sized sample of independent data (Fig. 1). For autocorrelated data, the proportion of sampled points that falls within a given contour of the range distribution depends entirely on the timespan and the strength of the autocorrelation. Consequently, to match the position information contained in an independent sample of a particular size, a larger and longer-term sample is needed for autocorrelated movement data. This explains why the conventional KDE will tend to yield underestimates on autocorrelated data and also why, when using such data, an individual’s estimated home range area tends to initially increase with sampling duration even if their movement process remains stationary (Girard et al. 2002). In this latter scenario, the underlying range distribution being estimated, and thus the true home range, has not necessarily changed, but it becomes more completely sampled and thus better resolved as the observation time increases.

What then are ecologists to do when faced with autocorrelated movement data of short duration, for which conventional KDE always underestimates the animal’s ultimate space use? Fortunately, the other side of the autocorrelation coin offers a solution. Autocorrelation implies relationships between past and future movements and can therefore be harnessed to make statistically rigorous predictions of future movement. Most current space-use estimators discard this information, but an estimator purpose-built for autocorrelated data could, in a mechanistic way, leverage the information to make better home range predictions (Börger et al. 2008). Previous home range estimators that account for autocorrelation have been extremely limited. Autocorrelated bivariate Gaussian density estimation (AGDE; Dunn and Gipson 1977, Fleming et al. 2014b) can incorporate realistic movement behaviors featuring strong, multiscale autocorrelations; however, AGDE typically estimates Gaussian range distributions, which will not work for many species. Mechanistic home range analysis (MHRA; Moorcroft and Lewis 2006, Moorcroft 2012, Potts and Lewis 2014) can provide more detailed range distributions, however current modeling efforts are limited to Markov processes (Appendix A: A.2), which cannot describe the continuous velocity motion revealed by modern ARGOS and GPS telemetry data (Johnson et al. 2008, Fleming et al. 2014a). Moreover, while Moorcroft and Barnett (2008) provide a fitting method that can account for Markovian autocorrelation, the traditional method of assuming independent observations remains in use (Bateman et al. 2014). Finally, while Brownian bridge density estimation (BBDE; Horne et al. 2007) is sometimes mistakenly employed as a home range estimator, its estimation target is actually the occurrence distribution, which does not quantify the home range (Appendix A: A.3).

We develop a new home range estimator that combines KDE’s flexibility and efficiency with AGDE’s ability to account for and leverage the information content of highly autocorrelated movement data. We formally re-derive the KDE explicitly assuming the data represent a sample from a nonstationary, autocorrelated, continuous movement process. The resulting autocorrelated KDE (hereafter AKDE) incorporates movement effects through the autocorrelation function (ACF), which can either be derived from a fitted movement model (as in Fleming et al. 2014a) or directly estimated from the data (Appendix C). We illustrate AKDE’s improved performance with both simulated data where the true home range area is known and an empirical example featuring Mongolian gazelles (Procapra gutturosa), which previous analyses have shown to exhibit highly autocorrelated movement. We outline the conditions under which the AKDE will outperform the classical KDE and show that our AKDE reduces to the standard KDE in the limit where autocorrelation vanishes and samples are truly independent. The AKDE is therefore a generalization of KDE.

**Kernel density estimation**

KDE proceeds by placing small kernels of smoothing bandwidth or covariance \( \sigma_B \) at each sampled location (Silverman 1986). The average of these kernels provides the estimate, \( \hat{\rho} \), of the PDF \( \rho \). The kernel’s shape matters little, but the bandwidth selection is paramount (Silverman 1986, Izenman 1991, Turlach 1993). The optimal bandwidth minimizes the mean integrated squared error (MISE) between \( \rho \) and its estimate \( \hat{\rho} \). The optimal \( \sigma_B \) will vary among data sets, but its calculation can be automated. KDE bandwidth optimization poses more problems than ordinary regression analysis. Several methods have been developed; plug-in and cross-validation approaches are the two most common (Silverman 1986, Turlach 1993). In general, plug-in methods tend to over-smooth the estimate, while cross-validation methods tend to under-smooth the estimate; adaptive-bandwidth methods tend to produce more detail in areas of high frequency compared to fixed-bandwidth methods (Silverman 1986, Izenman 1991, Turlach 1993).

To make our approach explicit, we first derive the bandwidth relations for the conventional KDE in Appendix B: Eq. B.1, using a Gaussian kernel and reference function. The MISE is

\[
\varepsilon(\sigma_B) = \frac{1}{(2\pi)^{d/2}} \left( \frac{1}{n} \frac{1}{\sqrt{\det(2\sigma_B)}} + \frac{n-1}{n} \frac{1}{\sqrt{\det(2\sigma_B + 2\sigma_0)}} \right) - \frac{2}{\sqrt{\det(2\sigma_0 + 2\sigma_B)}} + \frac{1}{\sqrt{\det(2\sigma_0)}}
\]

(1)
in terms of the covariance of the locations $\sigma_0$, number of spatial dimensions $q$, and number of data points $n$, and where $\sigma_0$ is the bandwidth. For uncorrelated data, this relation provides an asymptotically consistent estimate and an asymptotically optimal order of error. We choose to base our generalization to autocorrelated data on Eq. 1, and our AKDE derivation in Appendix B: Eq. B.2 follows the same sequence of steps while relaxing the assumption of IID data. We avoided the least-squares cross-validation approach, because cross-validation itself assumes independence, whereas the generalization of MISE Eq. 1 is conceptually straightforward. Moreover, the leave-one-out cross-validation method that is conventionally used in KDE-based home range estimation (Seaman and Powell 1996, Gitzen and Millspaugh 2003) is particularly vulnerable to autocorrelation. Because any single data point will be strongly correlated with adjacent data points, leaving only one data point out of an estimate will not yield any substantial change, which will then be falsely attributed to fit quality (also, see Hemson et al. 2005, for other issues).

The new material in Appendix B: Eq. B.2 is highly technical, but it allows us to derive a closed form expression for the MISE, whose minimum occurs at the optimal bandwidth. Given an estimate of the movement process’s mean and ACF, the optimal bandwidth can be calculated numerically. The AKDE converges to the conventional KDE when autocorrelation is absent, but outperforms it by a wide margin as autocorrelation increases (Appendix B: Eq. B.3). Subsequently, we consider the corresponding MISE for the special case of a stationary stochastic process, where temporal variation is absent from both the mean and the ACF governing animal movement, in which case the autocorrelated MISE is given by

$$
\varepsilon(\sigma_0) = \frac{1}{(2\pi)^q} \left( \frac{1}{n} \sum_{\tau} \frac{n(\tau)}{\sqrt{\det(2\gamma(\tau) + 2\sigma_0)}} \right) - \frac{2}{\sqrt{\det(2\sigma_0 + \sigma_0)}} + \frac{1}{\sqrt{\det(2\sigma_0)}}
$$

(2)

where $\gamma(\tau)$ is the semi-variance function (SVF: Appendix B: Eq. B.36) and $n(\tau)$ is the number of location pairs with time lag $\tau$ between them. To compare the uncorrelated and autocorrelated MISEs, the above sum over lags can be expressed as

$$
\frac{1}{n^2} \sum_{\tau} \frac{n(\tau)}{\sqrt{\det(2\gamma(\tau) + 2\sigma_0)}} = \frac{1}{n} \frac{1}{\sqrt{\det(2\sigma_0)}} + \frac{1}{n^2} \sum_{\tau \neq 0} \frac{n(\tau)}{\sqrt{\det(2\gamma(\tau) + 2\sigma_0)}}
$$

(3)

and so the second term here generalizes the second term of Eq. 1. In the limit of uncorrelated data, the SVF limits to the covariance, $\gamma(\tau) \rightarrow \sigma_0$, for $\tau \neq 0$, and the autocorrelated MISE reduces to the uncorrelated MISE. Therefore, there is no reason to prefer the uncorrelated MISE over the autocorrelated MISE, even if the autocorrelation is weak.

**Application**

In Fig. 2, we plot KDEs for the ranging behavior of one Mongolian gazelle monitored over a 1-yr period. We obtained a total of $n = 866$ relocations for this individual using an hourly sampling schedule with 5-h gaps every 20 h and 11-d gaps after every 5 d of sequential data. Fig. 2A uses the conventional bandwidth, which only applies to uncorrelated data, while Fig. 2B uses Eq. B.35 (in Appendix B), along with the ACF estimate calculated in Fleming et al. (2014b). The AKDE predicts a home range area of 50 000 – 94 000 km$^2$, with confidence intervals calculated according to data in Appendix B: Eq. B.3, while the conventional KDE estimate is only 19 000 – 20 000 km$^2$. Next, we restrict our analysis to the first half of the data to test how the two methods predict future space use, using conventional KDE in Fig. 2C and the AKDE in Fig. 2D. With half the data, AKDE predicts a home range area of 47 000 – 96 000 km$^2$, which is consistent with its better estimate derived from the full data, though the confidence intervals are slightly wider. In contrast, conventional KDE predicts a home range area of 9000 – 10 000 km$^2$, which differs substantially from its better estimate. Because of the quantity and resolution of the data, the range estimate using the conventional KDE approach falls tightly around the sampled data. Increasing the sampling frequency will further degrade the conventional estimate, causing the home range to split into numerous isolated areas of high utilization.

This behavior typifies the conventional KDE and leads some researchers (e.g., Swihart and Slade 1985) to advocate for nonstatistical measures (i.e., MCP) that at least appear visually reasonable. However, note that in both cases (Fig. 2A, C) the 95% home range area of the AKDE is much larger than the MCP would be. Mongolian gazelles are nomadic wanderers whose movements may involve gross displacements exceeding 1000 km/yr with little concordance among years (Olson et al. 2010, Mueller et al. 2011, Fleming et al. 2014b). Consequently, longer observation periods tend to show the gazelles using larger amounts of space, up to an asymptote set by the ACF’s details (Fleming et al. 2014a). The AKDE captures this important behavior, whereas conventional space-use estimates will miss it because they discard the information encoded in the ACF on the movement process’ long-run behavior.

Visually, AKDE estimates often look too large because they contain substantial areas where the focal individual was not directly observed (e.g., Figs. 1 and 2; Appendix B: Eq. B.3). However, as previously men-
tioned, statistical intuition derived from experience with independent data often fails to transfer to situations where data are autocorrelated. In particular, as we have demonstrated with simulation (Fig. 1; Appendix B: B.3) and shown mathematically (Appendix B: Eq. B.35), the home range area cannot be determined from the locations alone, so the intuitive expectation that a home range area should conform closely to the observed locations is often inappropriate. Instead, accurately characterizing the individual’s long-term space use requires leveraging the information that exists in the transitions between spatial locations (i.e., in the ACF). Because the data points are linked together by the animal’s movements, it is the time dependence in the data that ultimately provides insight into space usage.

**DISCUSSION**

We have introduced AKDE, a new kernel density estimator that properly takes autocorrelation into account so that it accurately estimates home ranges from animal movement data. Among KDE techniques, only our estimator is asymptotically consistent, with an asymptotically optimal order of error, when data are autocorrelated. As we have shown, the conventional KDE only provides a lower bound for the estimate of home range area (Appendix B: B.2.1), and it is only valid when the relocation data are sampled so coarsely that they appear uncorrelated in time and the data are recorded for a far longer time than the timescale over which autocorrelations persist. When the sampling interval is much shorter than the autocorrelation timescales, which is inevitably the case for finely sampled (e.g., hourly) movement data in ungulates and other
large animals, then the conventional KDE of home range area is too small. What is worse, using KDE with more continuously sampled locations will more severely underestimate an animal’s area requirements (see Appendix B: B.4 and Fleming et al. 2014a), which runs counter to the intuition that more tracking data will reveal more space-use detail. Moreover, if autocorrelations persist over timescales comparable to the observation period, then all conventional methods will underestimate space use.

Because of the aforementioned limits on easy visual assessment of autoco-related movement data, ordinary intuitions are significantly biased when applied to animal tracking data, and, unfortunately, these biases only increase with increasing sampling frequency. In the limit of continuously sampled data, the conventional KDE home range area vanishes, while the AKDE asymptotes to a well-defined estimate (Appendix B: B.4). Unless the sampling interval is much larger than the autocorrelation timescales, an increase in the sampling rate does not result in a proportional increase in the effective sample size (i.e., the information content). Importantly, the period of data needs to be at least as long as the home range crossing time, if not many times longer, for any home range estimation (Appendix C).

Conventional estimates of an individual’s home range size often increase as the observation period increases (Girard et al. 2002). In this case, the underlying range distribution can be the same, but some of the bias inherent in conventional home range estimates decreases as the observation period increases. Conventional estimators assume that locations are sampled independently from the animal’s range distribution; however, it can take a significant amount of time for an animal to journey through its range, and a brief observation window will not yield a representative sample. Moreover, in the conventional perspective it can be difficult to ascertain if the home range has stopped increasing, as there are no reliable confidence intervals to compare growth with, and the home range estimates themselves are temporally autocorrelated, making trends difficult to distinguish from errors. Bootstrap and cross-validation techniques may seem useful for this purpose, but they also generally assume a lack of autocorrelation in the data. Because our approach is rigorously built from first principles to account for autocorrelation, it provides accurate confidence intervals that can diagnose situations where the data are insufficient to provide a reasonable home range estimate. For our gazelle example with short observation periods (Fig. 2), instead of yielding a grossly underestimated home range with deceptively narrow confidence intervals, the AKDE returns a reasonable estimate with wide confidence intervals that appropriately reflect the estimate’s limited precision.

In contrast with Swihart and Slade (1985) and Hansteen et al. (1997), several simulation studies have demonstrated situations where autocorrelation is not problematic for conventional KDE (De Solla et al. 1999 and references in Fieberg 2007). In particular, De Solla et al. (1999) simulated a situation in which velocities were discontinuous and even the shortest sampling interval was comparable to the home range crossing time. As a result, De Solla et al. (1999) drew the premature conclusion that including more autoco-related data was generally better than coarsening the data to avoid autocorrelation. However, real movement data sampled with modern GPS technology will tend to feature much stronger and possibly much longer-lasting autocorrelations, as well as much finer sampling relative to the home range crossing time. Under these conditions autocorrelation’s negative effects on conventional home range estimates must be considered (McNay et al. 1994, Hansteen et al. 1997, Rooney et al. 1998, Boyce et al. 2010, Fleming et al. 2014a, b). Our results demonstrate how badly the conventional KDE will fail for different movement processes and different sampling schedules. Importantly, AKDE obviates the need for brute-force strategies to avoid autocorrelation, such as thinning the data or sampling in an intentionally coarse way. There is no need to try and separate home range problems into situations that need AKDE and those for which a standard KDE is appropriate, because when there is no autocorrelation, AKDE returns the same result as KDE.

As animal tracking technology improves and relocation data sets continue to increase in sampling frequency, both the highly autocorrelated nature of movement data and the need for techniques that can leverage autocorrelation will become more apparent. While our AKDE is asymptotically consistent and has an asymptotically optimal order of error as a nonparametric estimator, there is a vast program of conventional KDE research focused on important statistical goals such as reducing the proportionality constant of the asymptotic error’s leading-order term, reducing bias (Silverman 1986, Izenman 1991, Turlach 1993) and correcting for hard boundaries (Silverman 1986). Undoubtedly, some of these techniques could be translated to AKDE. Even in its current form, however, AKDE provides a rigorous and flexible solution for estimating animal space use from autocorrelated movement data. Finally, future parametric estimators derived from consideration of the biological mechanisms governing space use, such as MHRA (Appendix A: A.2), but developed for autocorrelated data have the potential for even more statistical efficiency. The advantage of nonparametric estimators like AKDE lies in their broader applicability and lack of requirement for a detailed understanding of the focal species.

ACKNOWLEDGMENTS

The project was funded by the U.S. National Science Foundation ABI 1062411. T. Mueller was funded by the Robert Bosch Foundation. C. H. Fleming was funded by a Smithsonian Institution postdoctoral fellowship.
LITERATURE CITED


SUPPLEMENTAL MATERIAL

Ecological Archives

Appendices A–C are available online: http://dx.doi.org/10.1890/14-2010.1.sm