Correcting for missing and irregular data in home-range estimation

Article *in* Ecological Applications · February 2018

DOI: 10.1002/eap.1704

Christen H. Fleming
University of Maryland, College Park

Peter Leimgruber
Smithsonian Institution

Thomas Mueller
Senckenberg Biodiversität und Klima - Forschungszentrum

Dejid Nandintsetseg
Goethe University Frankfurt am Main and Biodiversität und Klima - Forschungsze…

Ecological Applications, 28(4), 2018, pp. 1003–1010

Correcting for missing and irregular data in home-range estimation

C. H. Fleming,[1,2,10] D. Sheldon,[3,4] W. F. Fagan,[2] P. Leimgruber,[1] T. Mueller,[5,6] D. Nandintsetseg,[5,6]
M. J. Noonan,[1] K. A. Olson,[1,7] E. Setyawan,[8,9] A. Sianipar,[10] and J. M. Calabrese[1,2]

[1] *Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, Virginia 22630 USA*
[2] *Department of Biology, University of Maryland College Park, College Park, Maryland 20742 USA*
[3] *College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, Massachusetts 01003-9264 USA*
[4] *Department of Computer Science, Mount Holyoke College, South Hadley, Massachusetts 01075 USA*
[5] *Senckenberg Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt (Main), Germany*
[6] *Department of Biological Sciences, Goethe University, Max-von-Laue-Straße 9, 60438 Frankfurt (Main), Germany*
[7] *Wildlife Conservation Society, Mongolia Program, 201 San Business Center, Amar Street 29, Small Ring Road, Sukhbaatar District, Post 20A, Box-21, Ulaanbaatar, Mongolia*
[8] *Manta Trust-Indonesian Manta Project, Badung, Bali 80361 Indonesia*
[9] *Institute for Marine and Antarctic Studies, University of Tasmania, Launceston, Tasmania 7250 Australia*
[10] *Conservation International Indonesia, Marine Program, Jalan Pejaten Barat 16A, Kemang, Jakarta, DKI Jakarta 12550 Indonesia*

Abstract. Home-range estimation is an important application of animal tracking data that is frequently complicated by autocorrelation, sampling irregularity, and small effective sample sizes. We introduce a novel, optimal weighting method that accounts for temporal sampling bias in autocorrelated tracking data. This method corrects for irregular and missing data, such that oversampled times are downweighted and undersampled times are upweighted to minimize error in the home-range estimate. We also introduce computationally efficient algorithms that make this method feasible with large data sets. Generally speaking, there are three situations where weight optimization improves the accuracy of home-range estimates: with marine data, where the sampling schedule is highly irregular, with duty cycled data, where the sampling schedule changes during the observation period, and when a small number of home-range crossings are observed, making the beginning and end times more independent and informative than the intermediate times. Using both simulated data and empirical examples including reef manta ray, Mongolian gazelle, and African buffalo, optimal weighting is shown to reduce the error and increase the spatial resolution of home-range estimates. With a conveniently packaged and computationally efficient software implementation, this method broadens the array of data sets with which accurate space-use assessments can be made.

Key words: animal tracking data; autocorrelation; home range; irregular sampling; kernel density estimation; marine tracking data; utilization distribution.

Introduction

Accurate home-range estimates are important for conservation and wildlife management interests, as they inform animal space use requirements (Burt 1943, Hayne 1949, Powell and Mitchell 2012). Home-range estimation is a deceptively difficult mathematical and statistical problem that involves the estimation of probability density functions (Winkle 1975) from sparse samples of autocorrelated movement processes (Swihart and Slade 1985, Hansteen et al. 1997, Fleming et al. 2015). In particular, home-range estimation is subject to a number of biases that can result in underestimation of home-range area. First, using a range estimator that assumes independent data (e.g., conventional kernel density estimator [KDE]) on autocorrelated data results in underestimation that increases with the degree of autocorrelation in the data (Swihart and Slade 1985, Fleming et al. 2015). Second, estimates based on data whose duration is not much longer than the average range crossing time (i.e., data with small effective sample size for home-range estimation) can have estimator-specific finite sample size biases (Fleming and Calabrese 2016). In this paper, we demonstrate that biases can also arise from under- or over-sampling particular times in a tracking data set.

The main sources of bias in home-range estimation, though long acknowledged in some cases (Swihart and Slade 1985, Hansteen et al. 1997), have only recently been addressed. First, the distinction has been made between range and occurrence distributions, with their different properties, whereas previously both were referred to interchangeably as utilization distributions (Fleming et al. 2015, 2016; Horne et al., *in press*). Second, the Gaussian reference function (GRF) KDE bandwidth optimizer has been generalized to handle autocorrelated data, resulting in the first autocorrelated kernel density estimator (AKDE; Fleming et al. 2015). Third, a small-sample-size area correction was developed for GRF-based KDE and AKDE estimators, denoted KDE_C and AKDE_C (Fleming and Calabrese 2016). Here, we consider the problem of samples that are biased with respect to the movement data's autocorrelation structure, in that some times are better sampled than others. This issue is of practical concern for researchers with irregular or missing data, which is a ubiquitous problem in movement ecology. Moreover, the resolution of this issue also provides mild accuracy improvements when data are autocorrelated

Manuscript received 24 July 2017; revised 13 November 2017; accepted 2 January 2018. Corresponding Editor: Devin S. Johnson.
[10] E-mail: flemingc@si.edu

and effective sample sizes are small (i.e., when few home-range crossings are observed) even if the data are evenly sampled.

Heterogeneity in sampling times is a common problem in movement ecology that can arise from multiple sources. In cases where tracking devices are programmed to collect regular observations, missing data can be caused by device malfunction or poor signal reception at particular times or places (Frair et al. 2004, Horne et al. 2007, Lewis et al. 2007). Examples of the latter include areas of high topographic relief, dense vegetation cover and canopy, diving behavior in aquatic animals (Johnson et al. 2008), and burrowing in terrestrial animals (Noonan et al. 2015). Intentional deviations from even sampling, such as duty-cycling or acceleration-informed sampling, are also frequently encountered in the literature (Brown et al. 2012). Home-range estimators that do not account for heterogeneous sampling in time will tend to predict that animals spend more time in areas that have better sampling coverage, even if differences in coverage only result from sampling issues and do not reflect underlying differences in movement (Katajisto and Moilanen 2006, Fieberg 2007). Ideally, oversampled and undersampled times and locations should receive lower and higher weights in kernel density estimation, respectively, so that temporal sampling bias is corrected for and probability density is neither overestimated nor underestimated. Here, we derive AKDE weights from first principles that account for temporal sampling bias and weight against oversampled times.

While bandwidth optimization in the presence of autocorrelation has only recently been addressed by AKDE (Fleming et al. 2015), there has been some related work on weighted kernel density estimation to account for temporal sampling bias. Most notably, Hines et al. (2005) optimized their KDE weights for mean estimation, under the assumption of an Ornstein–Uhlenbeck movement process (Uhlenbeck and Ornstein 1930), while optimizing their bandwidth under the assumption of an independently sampled process. More specifically, they applied the minimum variance unbiased (MVU) mean estimator weights (w_i from the formula $\bar{x} = \sum_{i=1}^{n} w_i x(t_i)$) to weighted KDE, arguing that mean estimation and KDE are similar. Unfortunately, for more general movement processes, MVU mean estimator weights can be negative at times, which can result in uninterpretable negative probabilities when applied to KDE. Katajisto and Moilanen (2006) introduced a temporal bandwidth method to weight data against oversampling, but they did not optimize their bandwidth for statistical estimation and left it as a user-tuned parameter. Fieberg (2007) weighted data by sampling interval, which is asymptotically equivalent to the method of Hines et al. (2005), and studied the problematic effects of biased sampling design on home-range estimation.

The novel contributions of this manuscript not only include the derivation of statistically optimal weight relations, whereby our kernel-density weights and bandwidth are simultaneously optimized for accurate density estimation, but also computational derivations important for numerically implementing weight optimization on large tracking data sets. As we show, AKDE weight optimization is essentially a quadratic programming (QP) problem. Generic QP algorithms, including the opensource R package `quadprog` (Turlach and Weingessel 2013), are too computationally prohibitive for large data sets. We therefore introduce fast algorithms specialized to our task that scale efficiently with data quantity.

We apply our new methods to simulated examples and real data sets featuring GPS collar malfunction on an African buffalo, naturally irregular marine data on a reef manta ray, and small effective sample size on a Mongolian gazelle.

METHODS

For a sample of n, q-dimensional locations $\mathbf{r}(t_i)$ at times t_i, a weighted kernel density estimate can be represented

$$\hat{p}(\mathbf{r}) = \sum_{i=1}^{n} w(t_i)\kappa(\mathbf{r} - \mathbf{r}(t_i)), \qquad \sum_{i=1}^{n} w(t_i) = 1 \tag{1}$$

where $w(t)$ denotes the weight at time t and κ denotes the kernel, which we choose to be Gaussian with covariance $\boldsymbol{\sigma}_\mathrm{B}$—the bandwidth matrix. Both the bandwidth matrix and the weight vector $\mathbf{w}$, with $w_i = w(t_i)$, will be optimized to minimize the mean integrated square error

$$\mathrm{MISE}(\boldsymbol{\sigma}_\mathrm{B}, \mathbf{w}) = \left\langle \int d^q\mathbf{r} |p(\mathbf{r}) - \hat{p}(\mathbf{r})|^2 \right\rangle \tag{2}$$

where $\int d^q\mathbf{r}$ denotes the q-dimensional volume integral and $\langle \cdots \rangle$ denotes the expectation value with respect to the distribution of the data, $\mathbf{r}(t_i)$, which may be autocorrelated. As the true density function $p(\mathbf{r})$ is unknown and the KDE method is non-parametric, the MISE (Eq. 2) must be approximated, and different approximations correspond to different KDE methods, all being asymptotically optimal (Silverman 1986, Izenman 1991, Turlach 1993). In Fleming et al. (2015) the MISE was derived for autocorrelated and possibly non-stationary data, with uniform weights and under the Gaussian reference function (GRF) approximation, where the true density $p(\mathbf{r})$ is taken to be normal in MISE (Eq. 2) as a first-order approximation (Silverman 1986). These results readily generalize to the case of non-uniform weights, by simply carrying the weights in Eq. 1 through the derivation, and so most of our effort will involve their optimization and study. For a stationary movement process, the autocorrelated GRF MISE can be represented as the quadratic form (Fleming et al. 2015; Eq. 2).

$$\mathrm{MISE}(\boldsymbol{\sigma}_\mathrm{B}, \mathbf{w}) = \frac{1}{(2\pi)^{\frac{q}{2}}} \left(\mathbf{w}^\mathrm{T}\mathbf{G}(\boldsymbol{\sigma}_\mathrm{B})\mathbf{w} - \frac{2}{\sqrt{\det(2\boldsymbol{\sigma}_0 + \boldsymbol{\sigma}_\mathrm{B})}} + \frac{1}{\sqrt{\det(2\boldsymbol{\sigma}_0)}} \right) \tag{3}$$

$$G_{ij}(\boldsymbol{\sigma}_\mathrm{B}) \equiv \frac{1}{\sqrt{\det\big(2\boldsymbol{\gamma}(t_i - t_j) + 2\boldsymbol{\sigma}_\mathrm{B}\big)}} \tag{4}$$

where $\boldsymbol{\gamma}(\tau) = \boldsymbol{\sigma}_0 - \boldsymbol{\sigma}(\tau)$ is the semi-variance function, $\boldsymbol{\sigma}(\tau) = \mathrm{COV}[\mathbf{r}(t+\tau), \mathbf{r}(t)]$ is the autocorrelation function, and $\boldsymbol{\sigma}_0 = \boldsymbol{\sigma}(0)$ is the covariance. The non-stationary bandwidth relations generalize with similar ease, and also result in a quadratic form on the weights (Appendix S.1).

Fleming et al. (2015) have already shown that uniformly weighted AKDE reduces to KDE in the limit of an independent and identically distributed (IID) process. In Appendix S.2.2, we show that the optimal weights for IID processes are uniform weights. This is an important result because it shows that optimally weighted AKDE reduces to conventional KDE when the data are not autocorrelated and there is no temporal bias to correct. Temporal bias cannot exist in IID data, because no time is any different from another. Finally, in Appendix S.2.3 we show that uniform weighting is asymptotically optimal for an evenly sampled process with stationary autocorrelations. This confirms our intuition that even sampling is generally ideal (Börger et al. 2006), while also informing us that optimal weighting can improve statistical efficiency when the data are autocorrelated and the effective sample size is small.

Numerical considerations

In practice, we perform a nested optimization of MISE (Eq. 3), where for each value of the bandwidth $\boldsymbol{\sigma}_\mathrm{B}$ we tabulate $\mathbf{G}(\boldsymbol{\sigma}_\mathrm{B})$ and then optimize the weights via quadratic programming (QP; Bertsekas 1999). The nested quadratic optimization problem is defined

$$\min_{\mathbf{w}} \mathbf{w}^\mathrm{T}\mathbf{G}(\boldsymbol{\sigma}_\mathrm{B})\mathbf{w}, \qquad \mathbf{1}^\mathrm{T}\mathbf{w} = 1, \qquad \mathbf{w} \geq \mathbf{0}$$

$$\sum_{i=1}^{n} w_i = 1, \qquad w_i \geq 0 \tag{5}$$

where $\mathbf{1}$ is the vector of ones and $\mathbf{0}$ is the vector of zeros. Problem (Eq. 5) can be solved by QP algorithms much faster than by a more general nonlinear optimizer, which will be burdened by the high dimensionality of $\mathbf{w}$. However, generic QP algorithms still have an $\mathcal{O}(n^3)$ computational cost, meaning that 10 times the amount of data will take 10^3 = 1,000 times longer to analyze. In Appendix S.3 we derive a novel preconditioned conjugate-gradient method, which is only $\mathcal{O}(n \log n)$ for stationary processes, meaning that 10 times the amount of data will take little more than 10 times longer to analyze. This method of weight optimization is implemented in R package ctmm (v0.3.4 and later, available on CRAN), via the akde() method with argument weights=TRUE, whereas weights=FALSE produces the uniformly weighted AKDE (Fleming and Calabrese 2015, Calabrese et al. 2016).

RESULTS

Empirical examples

In our first empirical example, we consider an African buffalo (*Syncerus caffer*, Cross et al. 2016) named "Pepper" that was monitored for 8.5 months in Kruger National Park, South Africa. In Calabrese et al. (2016) and Péron et al. (2016), it was noted that Pepper has a highly irregular sampling schedule with an artefactual daily periodicity. Moreover, early on in Pepper's timeseries, her average sampling interval abruptly changed from 1 h to 2 h per fix, likely due to device malfunction. Gaps in Pepper's data range from hours to days, and in total 71% of the hourly observations are missing. We contrast the uniformly weighted and optimally weighted AKDE$_\mathrm{C}$ range estimates for Pepper in Fig. 1 and Appendix S4: Fig. S.1. The designation of heavily used areas is noticeably distorted by uniform weighting, which gives exaggerated importance to the areas where Pepper's sampling interval happened to be 1 h and the optimally weighted estimate results in a 28% reduction in the estimated MISE. In this case, the oversampled times fell outside of the bulk of the data, causing the uniformly weighted estimate to produce a (50%) core home-range area 32% larger than that of the more accurate estimator.

Computation time for Pepper's 1,725 locations with the slow $\mathcal{O}(n^3)$ algorithm was 207 s, while the fast $\mathcal{O}(n \log n)$ was 46 s, both on an i5 CPU with 16 GB of RAM. This represents a sample size where direct matrix techniques are not prohibitive. However, if we consider a similar data set, but with 10 × 1,725 = 17,250 locations, then the slow algorithm will take ~2.4 d to process and require considerably more RAM, while our fast algorithm will only take ~7.5 min. We note that relocation data sets wtih sample sizes in the tens of thousands or more are becomming increasingly common as battery technology improves (Kays et al. 2015).

In our second empirical example, we consider data on a female reef manta ray (*Manta alfredi*, Setyawan and Sianipar 2018) monitored for 2.5 months in Komodo National

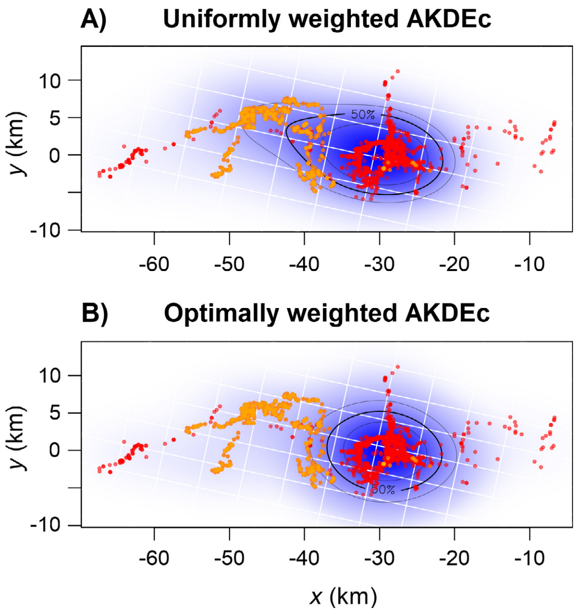

FIG. 1. African buffalo autocorrelated kernel density estimator, corrected for sample size (AKDE$_\mathrm{C}$) range distributions with (A) uniform weights and (B) optimal weights. The initial period of (oversampled) hourly times is contrasted in orange. The thick black contours denote the point estimate of the 50% (core) home-range contour, while the gray contours denote 95% confidence intervals on the magnitude of the core home-range area. Gridlines depict the scale and orientation of the optimal bandwidth, and thus the estimate's resolution. As the sampling was fairly irregular for this buffalo, and the effective sample size was relatively small ($n_\mathrm{A} \sim 13$), weight optimization reduced the estimated mean integrated square error (MISE) by 28%. Optimal weighting deemphasizes oversampled areas (orange) and better emphasizes undersampled areas (red).

Park, Indonesia. GPS fixes were programmed to be hourly, but were only obtainable if the manta ray happened to be at the surface at a scheduled fix time. As is typical in marine data, the realized sampling schedule is highly irregular and observed gaps frequently spanned days. We demonstrate our new method on this data set in Fig. 2 and Appendix S4: Fig. S.2, where we show that uniform weighting does not give appropriate importance to poorly sampled areas of use, where the manta ray spent more time at depth. In this case, undersampled times fell outside of the bulk of the data, causing the uniformly weighted estimate to produce a 95% home-range area 15% smaller than that of the more accurate estimator, while the 80% home-range area was 28% smaller.

In our final empirical example, we consider a Mongolian gazelle (*Procapra gutturosa*, Fleming et al. 2018) sampled at 23-h intervals for 17 months. Mongolian gazelles are nomadic herbivores that cross their nomadic range on seasonal scales (Fleming et al. 2014*a, b*), which here produced effective sample sizes of ~4 for the mean and ~10 for the area, meaning that the same quality of range estimate could be produced with approximately 10 independently sampled locations, if such data were possible to obtain. In this regime of small effective sample size where few range crossings were observed, optimal weighting produced an estimated 5% reduction in error by upweighting the end times, which we highlight in Fig. 3 and Appendix S4: Fig. S.3.

Simulation examples

Duty cycling.—Duty cycling is the practice of periodically suspending or reducing the sampling rate in an attempt to collect longer periods of (intermittently) high-resolution data, given constraints on battery life. Here we examine the effect of duty cycling and show that naive weighting schemes are not optimal. For example, if we sample the data at 6-h and 24-h intervals, each 24-h sample is not necessarily worth four 6-h samples, in terms of their relative weight. As shown in Appendix S.2.2, if the data are sampled coarsely enough to be uncorrelated (i.e., sampling interval $\gg$ range crossing time), there can be no temporal sampling bias and the optimal weights are uniform, regardless of the relative sampling intervals. On the other hand, if we have a burst of n locations, very close together in time, that all contain similar information due to strong autocorrelation, then we expect them to be each worth a factor of $1/n$ less than independently sampled data. Real data, however, will typically exist in an intermediate regime, and real data will have finite duration. This leads to two results, both demonstrated in Fig. 4, where two sampling intervals are employed in series. First, at intermediate timescales, the average optimal weights are between the two above-discussed extremes. Second, there is a small-sample-size effect on the end times. Because the initial and final times are less correlated with the remainder of the data, they are more informative and should receive higher weighting. This effect is why uniform weights are not optimal at small effective sample sizes for autocorrelated data, because they do not leverage this autocorrelation structure in the data. A similar phenomenon happens with the MVU mean estimator weights (Hines et al. 2005).

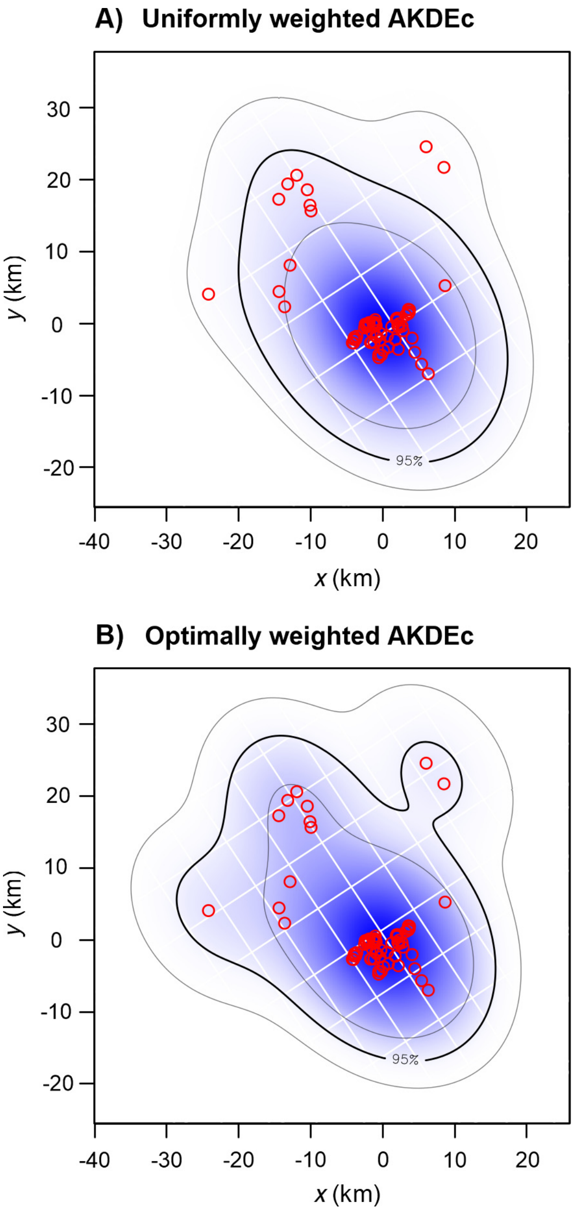

FIG. 2. Reef manta ray AKDE$_C$ range distributions with (A) uniform weights and (B) optimal weights. The thick black contours denote the point estimate of the 95% home-range contour, while the gray contours denote 95% confidence intervals on the magnitude of the 95% home-range area. Gridlines depict the scale and orientation of the optimal bandwidth, and thus the estimate's resolution. As GPS location fixes were only obtained when the manta ray surfaced at scheduled observation times, these data are highly irregular and weight optimization reduced the estimated MISE by 33%. Optimal weighting places significantly more importance on sparsely sample regions of the manta's home range.

Habitat-related signal loss.—Conventionally, autocorrelation has been viewed as a nuisance of the data that complicates analysis and reduces statistical efficiency. However, autocorrelation offers certain advantages if its information content can be leveraged. In Fig. 5, we demonstrate with simulation how optimizing the weights against temporal sampling bias can have the side effect of correcting for sampling biases due to differential signal loss in unknown

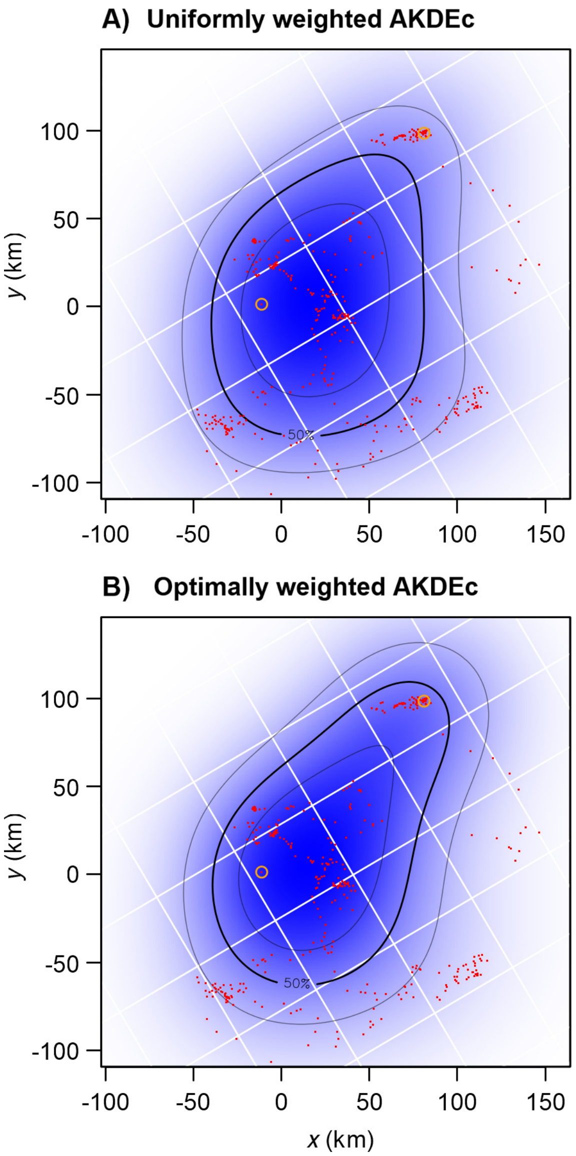

Fig. 3. Mongolian gazelle AKDE$_C$ range distributions with (A) uniform weights and (B) optimal weights. The initial and final times are highlighted with orange circles. The thick black contours denote the point estimate of the 50% range contour, while the gray contours denote 95% confidence intervals on the 50% range area. Gridlines depict the scale and orientation of the optimal bandwidth, and thus the estimate's resolution. The effective sample size is fairly small here ($n_A \sim 10$) and so optimal weighting reduced the estimated MISE by 5%, largely by placing more weight on the initial and final times to better predict past and future space use.

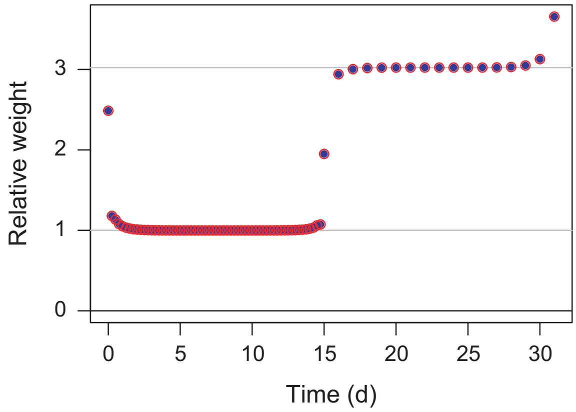

Fig. 4. Optimal weights for simulated tracking data where, at half way through, the schedule abruptly switches from a 6-h sampling interval to a 24-h sampling interval. Optimization was performed with both fast (blue) and slow (red) algorithms described in Appendix S.3, here demonstrating their equivalence. The median weights for each sampling frequency are depicted by a gray line, to make clear that the optimal weighting does *not* make one 24-h location worth four 6-h locations, even on average. Furthermore, there are noticeable deviations from the average weights at the beginning and end times, where the end locations are more independent and thus more informative.

habitats. For instance, when considering a species that prefers edge habitats between grasslands and forests, signal loss will be worse within the forests and therefore conventional estimators will overestimate grassland use and underestimate forest use (Horne et al. 2007). However, this simulated example assumes that habitat differences affect signal loss, but not movement behavior. In the real world, habitat differences that are pronounced enough to affect signal loss would also likely affect movement behavior. Disentangling these two effects would require both non-stationary home range estimation and weight optimization (see Appendix S.1), and appropriate non-stationary movement models that capture the habitat-specific behaviors. The core method of AKDE is fully non-stationary (Fleming et al. 2015, Appendix B.2), and non-stationary weight relations are given in Appendix S.1. The development of appropriate non-stationary models and related estimators is ongoing (Blackwell et al. 2015, Breed et al. 2017).

Discussion

Sampling irregularities in animal tracking data can significantly influence statistical analyses and subsequent ecological inference (White and Garrott 1986, Frair et al. 2004, 2010, DeCesare et al. 2005). We have derived a statistically rigorous weighting method to account for temporal sampling bias in home-range estimation. Optimal weighting can correct for sampling irregularities, missing data, and duty-cycled sampling schedules. Weight optimization can also result in mild improvements for evenly sampled autocorrelated data, when the effective sample size is small (where few home-range crossings were observed). As the combination of autocorrelation and data irregularity are nearly unavoidable aspects of animal tracking data, it is important to have sound statistical methods that can handle these complications. Our method differs from ordinary weighted kernel density estimation (Silverman 1986) in that (1) the weights are optimized simultaneously with the bandwidth to minimize error in the resulting density estimate and (2) autocorrelation is fully taken into account. The method derived here generalizes and improves the AKDE method of Fleming et al. (2015), reduces to conventional KDE in the limit of IID data and can still be followed by the small-sample-size bias correction of Fleming and Calabrese (2016). Moreover, we have assembled and derived computationally efficient algorithms to make this weight optimization feasible for the large data sets that are becoming the norm in

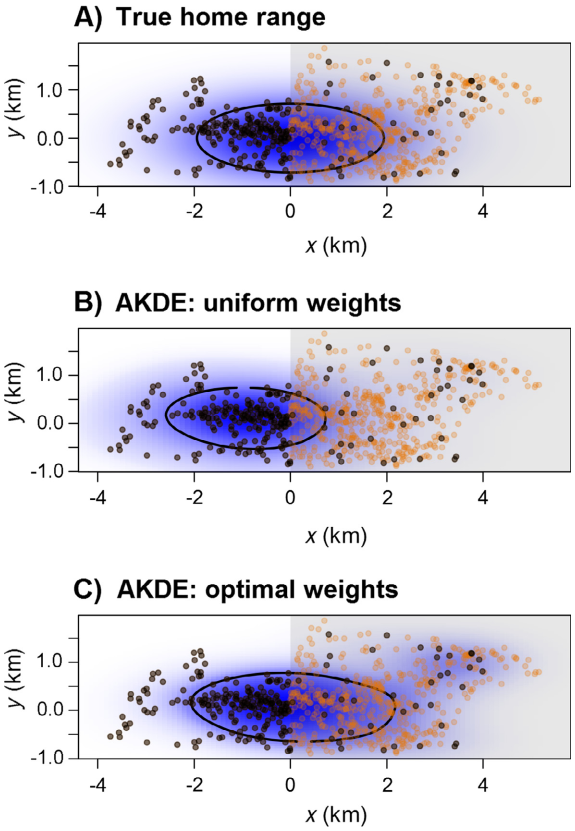

FIG. 5. Simulated movement data (black circles) with perfect GPS reception in the left habitat (white) and 90% signal loss in the right habitat (gray), with orange points being lost observations. The underlying movement process is the same in both habitat types. Panel A depicts the true 50% core home range, which straddles the two habitats. In panel B, a uniformly weighted estimator is used, which is biased toward areas of better signal reception. Finally, in panel C, the sparser data are upweighted to account for irregular sampling caused by signal loss. Neither of the estimators in panels B and C is aware of the two habitats, while estimator C accounts for the habitat-related signal loss by leveraging autocorrelation structure in the data.

movement ecology (Kays et al. 2015). These weighting methods and fast algorithms are implemented in the R package `ctmm` (Fleming and Calabrese 2015, Calabrese et al. 2016), which is available via the CRAN repository.

As weight optimization is performed simultaneously with bandwidth optimization, this method reduces the density estimate's MISE while increasing its resolution by allowing for a smaller optimal bandwidth. Whether or not uniformly weighted density estimators under- or overestimate the home-range area relative to optimally weighted estimators is entirely situational (e.g., Figs. 1 and 2). More broadly, optimal weighting improves the quality of home-range estimates for some of the most difficult, yet common types of problems encountered in animal tracking data (i.e., small effective sample size or irregular sampling).

We carefully note that, although this method can correct for temporal sampling bias, optimal weighting does not place (non-optimized) irregularly sampled data on equal footing with evenly sampled data, all else being equal. Given a known movement model, one could optimize the sampling schedule to have more density near gaps and end points and, in that case, such data would actually be worth more than regularly sampled data. However, in general, home-range estimation is more efficient with long periods of evenly sampled data. Home ranges are resolved by the number of range crossing events observed, as opposed to the number of fixes (Fleming and Calabrese 2016). As far as home-range estimation is concerned, increased sampling rates yield marginal gains (Fleming et al. 2015: Appendix B.4).

In contrast to range distributions, occurrence distributions become more detailed with increasing sampling rate, because they estimate where the individual was located during the observation period (Fleming et al. 2015, 2016). However, because occurrence distributions target the observation period, they do not predict future space use. While coverage areas of the range distribution are the frequently used areas, and are sampling independent, coverage areas of the occurrence distribution are proportional to our ignorance as to where the individual traveled during the observation period, and are inherently sampling dependent. In the limit of continuous sampling with zero telemetry error, the occurrence area is zero at all quantiles, because we have perfect knowledge of the animal's trajectory. When misapplied to estimating the home range, occurrence distribution estimators generally have a downward bias.

While our weights, $w(t)$, are allowed to vary in time, we left our bandwidth matrix, $\boldsymbol{\sigma}_{\mathrm{B}}$, constant for stationary processes. In contrast, Hines et al. (2005) introduced an ad hoc adjustment to the bandwidth matrix, $\boldsymbol{\sigma}_{\mathrm{B}}(t) \propto w(t)^{1/3}$, scaling it with the cube root of their weights. While it is not clear how this adjustment impacts statistical efficiency, whether positively or negatively, the idea of allowing the bandwidth to vary in time is worth considering. More accurately, one could attempt to optimize both time-dependent weights and time-dependent bandwidths, possibly with a relationship between the two enforced. Unfortunately, this optimization problem is numerically intractable because the dependence of the MISE on bandwidth is highly nonlinear (Appendix S.1). A promising avenue for future research would therefore be investigating if some simple weight–bandwidth optimization relations could be made computationally feasible and whether or not they have any substantial impact on statistical efficiency.

Finally, we note that, although applied here to the tracking records of individuals, the statistical and computational ideas in this study are necessary for future research into species and population range estimation with autocorrelated tracking data. Combining multiple tracking data sets, possibly also with survey data, into a single population estimate is of interest in many fields where tracking data gives more exhaustive and less biased spatial coverage than survey and transect data (Lindberg and Walker 2007, Fieberg et al. 2010). However, different tracks may contain different sampling schedules and correspond to different movement characteristics. Therefore, weight optimization across individuals is also critical in deriving optimal estimates of population distributions in the presence of autocorrelation.

ACKNOWLEDGMENTS

This work was supported by the U.S. NSF Advances in Biological Informatics Program (ABI-1458748 to J. M. Calabrese). M. J.

Noonan was supported by a Smithsonian Institution CGPS Grant. T. Mueller, D. Nandintsetseg, and K. A. Olson were funded by the Robert Bosch Foundation. We thank Conservation International-Indonesia and the SEAA Aquarium in Singapore for sponsoring the manta satellite tag used in this study, and the Indonesian Ministry of Marine Affairs and Fisheries for their support and permitting for the manta tagging program. We also thank the Komodo National Park Authority and Greg Heighes and Dive Komodo (and especially dive guides Fabi and Wai) for their able assistance in our manta tagging efforts. We thank Paul Cross for providing the African buffalo data. The buffalo data collection was supported by the USNSF and NIH Ecology of Infectious Disease Program (DEB-0090323 to W. M. Getz).

Literature Cited

Bertsekas, D. P. 1999. Nonlinear programming. Athena Scientific, Belmont, Massachusetts, USA.

Blackwell, P. G., M. Niu, M. S. Lambert, and S. D. LaPoint. 2015. Exact Bayesian inference for animal movement in continuous time. Methods in Ecology and Evolution 7:184–195.

Börger, L., N. Franconi, G. De Michele, A. Gantz, F. Meschi, A. Manica, S. Lovari, and T. Coulson. 2006. Effects of sampling regime on the mean and variance of home range size estimates. Journal of Animal Ecology 75:1393–1405.

Breed, G. A., E. A. Golson, and M. T. Tinker. 2017. Predicting animal home-range structure and transitions using a multistate Ornstein-Uhlenbeck biased random walk. Ecology 98:32–47.

Brown, D. D., S. LaPoint, R. Kays, W. Heidrich, F. Kümmeth, and M. Wikelski. 2012. Accelerometer-informed GPS telemetry: Reducing the trade-off between resolution and longevity. Wildlife Society Bulletin 36:139–146.

Burt, W. H. 1943. Territoriality and home range concepts as applied to mammals. Journal of Mammalogy 24:346–352.

Calabrese, J. M., C. H. Fleming, and E. Gurarie. 2016. ctmm: An R package for analyzing animal relocation data as a continuous-time stochastic process. Methods in Ecology and Evolution 7:1124–1132.

Cross, P. C., J. A. Bowers, C. T. Hay, J. Wolhuter, P. Buss, M. Hofmeyr, J. T. D. Toit, and W. M. Getz. 2016. Data from: Nonparameteric kernel methods for constructing home ranges and utilization distributions. Movebank Data Repository. https://doi.org/10.5441/001/1.j900f88t

DeCesare, N. J., J. R. Squires, and J. A. Kolbe. 2005. Effect of forest canopy on GPS-based movement data. Wildlife Society Bulletin (1973–2006) 33:935–941.

Fieberg, J. 2007. Utilization distribution estimation using weighted kernel density estimators. Journal of Wildlife Management 71:1669–1675.

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

Fleming, C. H., and J. M. Calabrese. 2015. ctmm: continuous-time movement modeling. https://CRAN.R-project.org/package=ctmm

Fleming, C. H., and J. M. Calabrese. 2016. A new kernel-density estimator for accurate home-range and species-range area estimation. Methods in Ecology and Evolution 8:571–579.

Fleming, C. H., J. M. Calabrese, T. Mueller, K. A. Olson, P. Leimgruber, and W. F. Fagan. 2014*a*. From fine-scale foraging to home ranges: A semi-variance approach to identifying movement modes across spatiotemporal scales. American Naturalist 183: E154–E167.

Fleming, C. H., J. M. Calabrese, T. Mueller, K. A. Olson, P. Leimgruber, and W. F. Fagan. 2014*b*. Non-Markovian maximum likelihood estimation of autocorrelated movement processes. Methods in Ecology and Evolution 5:462–472.

Fleming, C. H., W. F. Fagan, T. Mueller, K. A. Olson, P. Leimgruber, and J. M. Calabrese. 2015. Rigorous home-range estimation with movement data: A new autocorrelated kernel-density estimator. Ecology 96:1182–1188.

Fleming, C. H., W. F. Fagan, T. Mueller, K. A. Olson, P. Leimgruber, and J. M. Calabrese. 2016. Estimating where and how animals travel: An optimal framework for path reconstruction from autocorrelated tracking data. Ecology 97:576–582.

Fleming, C. H., et al. 2018. Data from: Correcting for missing and irregular data in home-range estimation. Dryad Digital Repository. https://doi.org/10.5061/dryad.n42h0

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

Frair, J. L., J. Fieberg, M. Hebblewhite, F. Cagnacci, N. J. DeCesare, and L. Pedrotti. 2010. Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. Philosophical Transactions of the Royal Society of London B: Biological Sciences 365:2187–2200.

Hansteen, T. L., H. P. Adreassen, and R. A. Ims. 1997. Effects of spatiotemporal scale on autocorrelation and home range estimators. Journal of Wildlife Management 61:280–290.

Hayne, D. W. 1949. Calculation of size of home range. Journal of Mammalogy 30:1–18.

Hines, W. G. S., R. J. O'Hara Hines, B. Pond, and M. E. Obbard. 2005. Allowing for redundancy and environmental effects in estimates of home range utilization distributions. Environmetrics 16:33–50.

Horne, J. S., E. O. Garton, and K. A. Sager-Fradkin. 2007. Correcting home-range models for observation bias. Journal of Wildlife Management 71:996–1001.

Horne, J. S., J. Fieberg, L. Börger, J. L. Rachlow, J. M. Calabrese, and C. H. Fleming. *In press*. Animal home ranges: concepts, uses and estimation. *In* D. Murray and B. K.Sandercock, editors. Population ecology in practice. Wiley-Blackwell, Hoboken, NJ, USA.

Izenman, A. J. 1991. Recent developments in nonparametric density estimation. Journal of the American Statistical Association 86:205–224.

Johnson, D. S., J. M. London, M.-A. Lea, and J. W. Durban. 2008. Continuous-time correlated random walk model for animal telemetry data. Ecology 89:1208–1215.

Katajisto, J., and A. Moilanen. 2006. Kernel-based home range method for data with irregular sampling intervals. Ecological Modelling 194:405–413.

Kays, R., M. C. Crofoot, W. Jetz, and M. Wikelski. 2015. Terrestrial animal tracking as an eye on life and planet. Science 348:aaa2478.

Lewis, J. S., J. L. Rachlow, E. O. Garton, and L. A. Vierling. 2007. Effects of habitat on GPS collar performance: using data screening to reduce location error. Journal of Applied Ecology 44:663–671.

Lindberg, M. S., and J. Walker. 2007. Satellite telemetry in avian research and management: Sample size considerations. Journal of Wildlife Management 71:1002–1009.

Noonan, M. J., A. Markham, C. Newman, N. Trigoni, C. D. Buesching, S. A. Ellwood, and D. W. Macdonald. 2015. A new magneto-inductive tracking technique to uncover subterranean activity: What do animals do underground? Methods in Ecology and Evolution 6:510–520.

Péron, G., C. H. Fleming, R. C. de Paula, and J. M. Calabrese. 2016. Uncovering periodic patterns of space use in animal tracking data with periodograms, including a new algorithm for the Lomb-Scargle periodogram and improved randomization tests. Movement Ecology 4:19.

Powell, R. A., and M. S. Mitchell. 2012. What is a home range? Journal of Mammalogy 93:948–958.

Setyawan, E., and A. Sianipar. 2018. Data from: Correcting for missing and irregular data in home-range estimation. Movebank Data Repository. https://doi.org/10.5441/001/1.3gj67c2k

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman & Hall, Boca Raton, Florida, USA.

Swihart, R. K., and N. A. Slade. 1985. Testing for independence of observations in animal movements. Ecology 66:1176–1184.

Turlach, B. A. 1993. Bandwidth selection in kernel density estimation: A review. *In* CORE and Institut de Statistique. https://doi.org/10.1.1.44.6770

Turlach, B. A., and A. Weingessel. 2013. quadprog: functions to solve quadratic programming problems. https://CRAN.R-project.org/package=quadprog

Uhlenbeck, G. E., and L. S. Ornstein. 1930. On the theory of the Brownian motion. Physical Review 36:823–841.

White, G. C., and R. A. Garrott. 1986. Effects of biotelemetry triangulation error on detecting habitat selection. Journal of Wildlife Management 50:509–513.

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

SUPPORTING INFORMATION

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.1704/full

DATA AVAILABILITY

African buffalo data in Movebank Data Repository: https://doi.org/10.5441/001/1.j900f88t. Manta ray data in Movebank Data Repository: https://doi.org/10.5441/001/1.3gj67c2k. Mongolian gazelle data in Dryad Digital Repository: https://doi.org/10.5061/dryad.n42h0.