Mini-review

Autocorrelation-informed home range estimation: a review and practical guide

Inês Silva[1,2], Christen H. Fleming[3,4], Michael J. Noonan[5], Jesse Alston[1,2],
Cody Folta[4], William F. Fagan[4], Justin M. Calabrese[1,2,4,6*]

[1] Center for Advanced Systems Understanding (CASUS), 02826, Görlitz, Germany
[2] Helmholtz-Zentrum Dresden-Rossendorf (HZDR), 01328, Dresden, Germany
[3] Smithsonian National Zoo and Conservation Biology Institute, Virginia, USA
[4] Department of Biology, University of Maryland, College Park, Maryland, USA
[5] Department of Biology, University of British Columbia Okanagan, British Columbia, Canada
[6] Helmholtz Centre for Environmental Research—UFZ, 01328, Leipzig, Germany

* Corresponding author: j.calabrese@hzdr.de

Running head: Guide to autocorrelated home range estimation

Abstract

1. Modern tracking devices allow for the collection of high-volume animal tracking data at improved sampling rates over VHF radiotelemetry. Home range estimation is a key output from these tracking datasets, but the inherent properties of animal movement can lead traditional statistical methods to under- or overestimate home range areas.

2. The Autocorrelated Kernel Density Estimation (AKDE) family of estimators were designed to be statistically efficient while explicitly dealing with the complexities of modern movement data: autocorrelation, small sample sizes, and missing or irregularly sampled data. Although each of these estimators has been described in separate technical papers, here we review how

these estimators work and provide a user-friendly guide on how they may be combined to reduce multiple biases simultaneously.

3. We describe the magnitude of the improvements offered by these estimators and their impact on home range area estimates, using both empirical case studies and simulations, contrasting their computational costs.

4. Finally, we provide guidelines for researchers to choose among alternative estimators and an `R` script to facilitate the application and interpretation of AKDE home range estimates.

Keywords: home range, tracking data, telemetry, kernel density estimation, movement process

Introduction

An animal's *home range* was first defined in Burt (1943) as "*the area traversed by the individual in its normal activities of food gathering, mating, and caring for young*". Although this definition does not provide a mathematical description or statistical method for estimation, it highlights how behavior drives animal movement: areas selected by individual animals are usually distinct from the larger areas an animal could explore given their movement abilities. Early translations into a statistical definition include quantifying an animal's probability of using a given location (*i.e.*, utilization distribution; Jennrich & Turner, 1969; Worton, 1989). The concept of *home range* has been redefined by many authors over the years (Harris et al., 1990); here, we follow the definition of *home range* as the area repeatedly used throughout an animal's lifetime for all its normal behaviors and activities, excluding occasional exploratory excursions outside of home range boundaries. The characteristic temporal stability of a home range also highlights additional concepts: *range residency*, defined as the tendency of an animal to remain within its home range; and timescale parameters that quantify the weakness of this tendency, including the *home range crossing timescale* (τ), defined as the average time required for an animal to cross the linear extent of its home range.

Home range area estimates are used to inform conservation practitioners and wildlife managers about protected area sizes and to advocate for conservation policy changes (Bartoń et al., 2019; Lambertucci et al., 2014; Linnell et al., 1997). It is thus crucial to provide a reliable and statistically robust metric that is comparable across individuals, species, and sites. Natural landscapes are becoming increasingly fragmented (Curtis et al., 2018; Hansen et al., 2020), imposing new challenges at local, regional, and global scales, and unreliable estimations may hinder area-based conservation. Reliable estimates of home ranges, however, have proven to be deceptively difficult to achieve, and have occupied generations of ecologists (Fieberg & Börger, 2012; Horne et al., 2020; Jennrich & Turner, 1969; Worton, 1989). The inherent properties of animal tracking data create unique analytical challenges. Specifically, animal movement data frequently feature some combination of autocorrelation, small sample sizes, missing observations, or irregular sampling, and home range estimators that are not designed to handle these issues can both under- and overestimate the sizes of home ranges.

Though many home range estimators exist (Horne et al., 2020), Autocorrelated Kernel Density Estimation (AKDE) was the first to explicitly account for temporal autocorrelation in the data (Fleming et al., 2015). Since its introduction, AKDE has grown into a family of related techniques, each aimed at mitigating a different source of bias that can affect home range estimates, including *unmodeled autocorrelation* (Hemson et al., 2005; Kie et al., 2010; Swihart & Slade, 1997), *oversmoothing* (Seaman & Powell, 1996; Worton, 1995), *autocorrelation estimation bias* (Cressie, 2015), and *unrepresentative sampling in time* (Frair et al., 2004; Horne, Garton, & Kimberly A. Sager-Fradkin, 2007; Katajisto & Moilanen, 2006). These biases are mitigated, respectively, by the original AKDE (Fleming et al., 2015), the area-corrected AKDE (Fleming & Calabrese, 2017), the perturbative Hybrid Residual Maximum Likelihood parameter estimation and parametric bootstrapping (Fleming et al., 2019), and weighted AKDE (Fleming et al., 2018). Residual Maximum Likelihood (REML) is a form of maximum likelihood estimation that reduces biases in variance/covariance estimation.

AKDE and associated corrections have been shown to outperform traditional home range estimators across species, degrees of autocorrelation, and sample size (Noonan et al., 2019). The `ctmm` workflow also allows researchers to partially account for the location errors associated with their tracking datasets (Fleming et al., 2021). These methods can be run using the programming language `R` (www.r-project.org) and the `ctmm` or `amt` packages (Calabrese et al., 2016; Signer & Fieberg, 2021), or the `ctmmweb` graphical user interface (https://ctmm.shinyapps.io/ctmmweb; Calabrese et al., 2021). In addition to offering flexible and open-source tools for home range estimation, these software programs allow easy documentation and implementation of new methods by sharing code and workflows. Such reproducible methods can increase reliability and transparency in ecology (Alston & Rick, 2020; Culina et al., 2020; Powers & Hampton, 2019; Signer & Fieberg, 2021).

Because movement data often violates multiple assumptions of traditional methods, the individual methodological advances offered by the AKDE family of home range estimators can and often should be combined. The costs and benefits of each estimator have previously been described in separate technical papers, so in this paper, we bring all of these estimation methods together in one document. We describe their effects on the quality of home range estimates, both in isolation and in combination, while evaluating how sample size interacts with multiple different sources of bias. We use tracking data from African buffalo (*Syncerus caffer*; Cross et al., 2009), lowland tapir (*Tapirus terrestris*; Fleming et al., 2019), and jaguar (*Panthera onca*; Morato et al., 2018) as empirical case studies to guide researchers through the application and value of these analyses. Finally, we use simulations to show the improvements offered by combining these techniques and demonstrate their application in real-world problems. We conclude by giving clear guidance on how ecologists can choose among these alternatives to best achieve their study goals. We hope that this review provides a practical guide to *why* and *how* to use AKDE methods to estimate home ranges that will be useful for both researchers and practitioners who are unfamiliar with these methods.

Sources of bias and mitigation measures

Many biases, including most that affect home range estimates, are exacerbated by small sample sizes. Conversely, large sample sizes in modern tracking datasets are typically achieved through higher sampling frequencies, which exacerbate autocorrelation. Autocorrelation is a general statistical property of variables measured across geographic and temporal space (Dale & Fortin, 2002; Legendre, 1993), as observations sampled more closely in space or time tend to be more similar. In these conditions, it is thus important to distinguish between two different measures of sample size: *absolute sample size* (n) and *effective sample size* (N). *Absolute sample size* is simply the total number of observations in a dataset. More relevant for home range estimation, however, is the *effective sample size*. Specifically, the amount of information available to home range estimators is governed not simply by the total number of observations, but by the number of *range crossings* that occurred during the observation period (*i.e.*, how many times an animal traversed the linear extent of its home range). The *effective sample size* can be roughly estimated as T/τ, where T is the temporal duration of the tracking dataset, and τ is the average *home range crossing time parameter*. Increasing sampling frequency leads to larger *absolute sample sizes*, but does not increase the *effective sample size* commensurately. For autocorrelated data, the *effective sample size* is necessarily smaller than the *absolute sample size* and, very frequently in practice, orders of magnitude smaller (Fleming et al., 2019). In contrast, small *absolute sample sizes* commonly occur in VHF tracking data but are becoming rarer in modern GPS tracking data.

We now describe each source of bias and the mitigation measure available to correct it, highlighting the difference each correction makes with real data from multiple case studies. We present the bias sources in order of their general importance, from the largest bias to the smallest. Note that this ranking refers to the typical magnitude of each type of bias, but the order may be different under some conditions.

Bias I: Unmodeled autocorrelation

Traditional home range estimators such as minimum convex polygons (MCPs) and kernel density estimators (KDEs) assume *independently and identically distributed* (IID) data. When these techniques came into common use in the 1980s, the sheer difficulty of obtaining VHF location fixes ensured that the time interval between successive observations was typically long enough for most of the autocorrelation among observations to have decayed (Swihart & Slade, 1997; Worton, 1989). The IID assumption at the heart of these techniques was therefore usually satisfied by VHF-quality data (Harris et al., 1990). The situation began to change with the arrival of new technologies, most notably GPS tracking systems (Rempel et al., 1995), which now routinely feature large volumes of data with much more frequent temporal sampling than is feasible for VHF-based animal tracking. As autocorrelation arises from observations sampled closely in time also being located closely in space, increasing sampling frequencies inevitably leads to more strongly autocorrelated tracking data (De Solla et al., 1999). Automated, high-sampling frequency tracking data has undoubtedly revolutionized movement ecology (Kays et al., 2015), but these advances have broken the armistice between the statistical assumptions of traditional home range estimators and the reality of the datasets now used to study animal movement (Boyce et al., 2010).

Specifically, feeding autocorrelated data into a home range estimator based on the IID assumption yields negatively biased estimates (Noonan et al., 2019). Autocorrelation-induced underestimation of home range areas is particularly pronounced when the *effective sample size* is small. In the recent comparative study of Noonan *et al*. (2019), 368 out of 369 tracking datasets featured strong autocorrelation, and roughly half were also plagued by small *effective sample size*. In these conditions, conventional estimators—such as MCPs, KDEs, and local convex hull polygons—underestimate home range areas by a factor of ~2 to 13 (on average), depending on the method and bandwidth optimizer, which is what determines how tightly KDEs conform to the data.

Accordingly, published estimates featuring these traditional methods may severely underestimate animal space-use requirements, hindering conservation and management decisions.

Mitigation Measure I: AKDE

Fortunately, it is not autocorrelation *per se* that causes errors in home range estimation, but rather autocorrelation that is statistically 'unmodeled' (Calabrese et al., 2021). Home range estimators that account for autocorrelation can therefore avoid the biases and violated assumptions of traditional methods. Autocorrelated Kernel Density Estimation (AKDE) explicitly requires a movement model that accounts for the autocorrelation in the tracking data (**Figure 1**) and then estimates the home range while conditioned on the same movement model (Fleming et al., 2015). This model is identified via formal model selection among a range of plausible alternatives, currently spanning from uncorrelated data (IID), correlated positions but uncorrelated velocities (Ornstein-Uhlenbeck or OU; Uhlenbeck & Ornstein, 1930), to correlated positions and correlated velocities (Ornstein-Uhlenbeck foraging process or OUF; Fleming et al., 2014). In this framework, IID is both a candidate model and one limit of a continuum of possibilities, rather than an *a priori* assumption. These models are ranked based on Akaike's Information Criterion adjusted for small sample sizes (AICc) by default, though the `ctmm` package also offers AIC, Bayesian Information Criterion (BIC), *Leave-One-Out Cross-Validation* (LOOCV), and Half-Sample Cross-Validation (HSCV).

Ad hoc measures such as data thinning (Harris et al., 1990; Rooney et al., 1998) are not necessary, as AKDE allows model assumptions to conform as closely as possible to empirical reality, instead of coercing the data to fit a model with unrealistic assumptions. Feeding IID data into AKDE will not have any adverse effects, as it will simply result in a conventional KDE estimate. This workflow also allows reliable confidence intervals to be determined for home range area estimates, which historically have not been applied to home range estimates. This measure of confidence is fundamental for any statistical estimate (Pawitan, 2001), increasing the comparability of AKDE and its relevance for biogeographical and conservation applications.

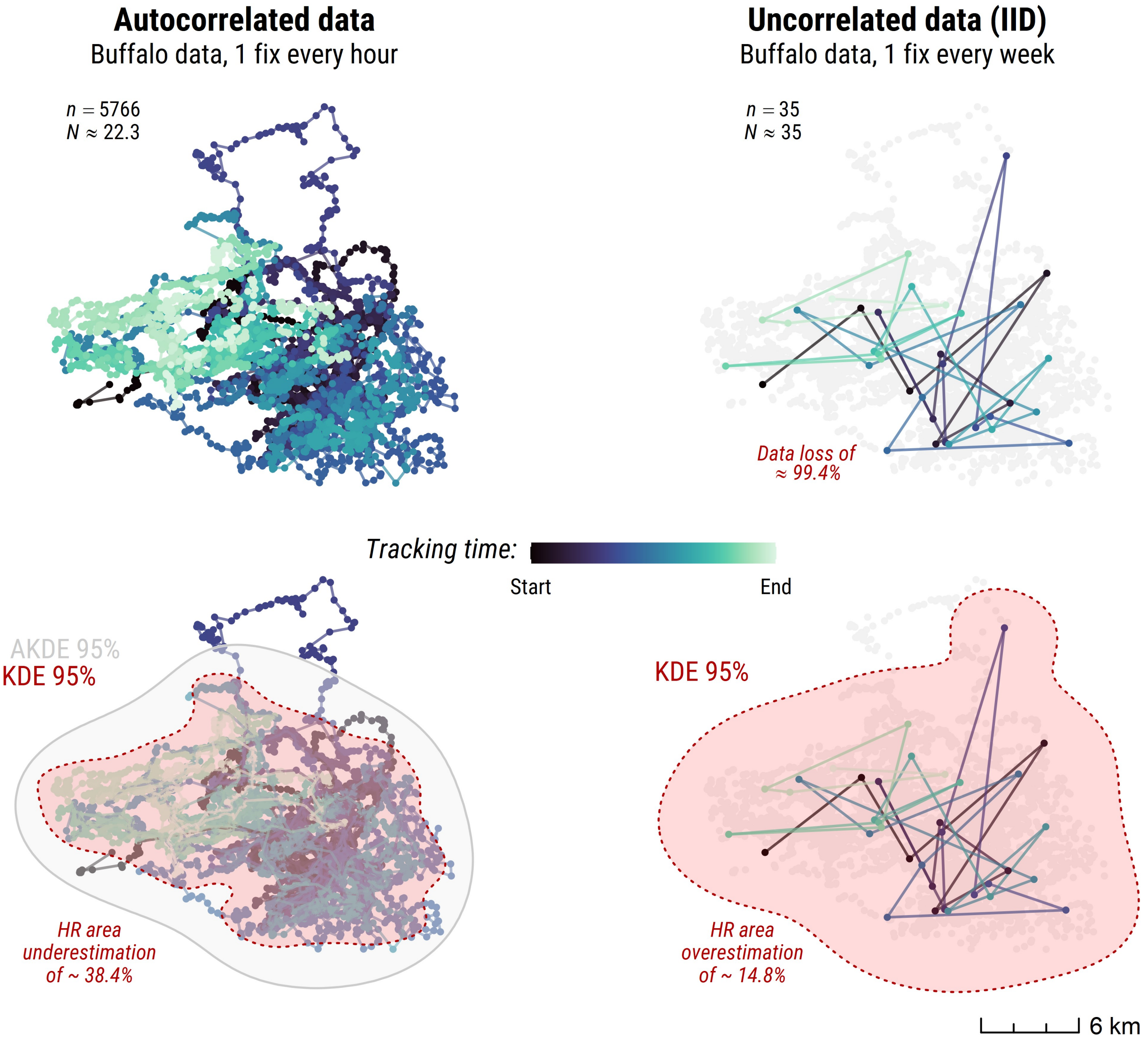


Figure 1. An example of autocorrelated data (individual six from the African buffalo dataset, available within the ctmm package), and the same data when it achieves independence (IID) after data thinning (from one fix per hour to one fix per week). We calculated the 95% contour of an Autocorrelated Kernel Density Estimation (AKDE) and a Gaussian reference function KDE (GRF-KDE). Displayed errors correspond to % bias of full dataset KDE and subset KDE against full dataset AKDE. N: effective sample size, n: absolute sample size.

Bias II: Oversmoothing

Kernel density estimators are best-in-class tools for estimating unknown probability distributions and are used in this capacity across the sciences (Chen, 2017; Silverman, 1986; Wang et al., 2013). In the context of tracking data, KDEs estimate the probability distribution of locations, which is then used to estimate the area of a home range (Powell, 2000; Worton, 1989). Typically, ecologists are more interested in this area estimate than in the distribution itself.

Even when we account for autocorrelation (AKDE), kernel density estimators based on the Gaussian reference function (GRF) remain biased owing to the natural tendency of the GRF approximation to oversmooth (yielding a more spread-out distribution). This bias is estimator-specific, and may be either positive or negative (Kie et al., 2010; Worton, 1995): for GRF-KDEs—such as AKDE and h_{ref} (Silverman, 1986)—this bias is positive and, all else being equal, leads to an overestimated home range (Seaman & Powell, 1996). Importantly, for estimators that do not account for autocorrelation, like h_{ref} but unlike AKDE, this positive bias can be masked by the often stronger negative bias caused by unmodeled autocorrelation. For KDEs based on least-squares cross-validation, h_{LSCV}, this bias is typically negative (Blundell et al., 2001; Hemson et al., 2005) and exacerbates the autocorrelation-induced underestimation of home range areas.

Mitigation Measure II: KDE$_\mathrm{C}$ or AKDE$_\mathrm{C}$

Fleming and Calabrese (2017) derived an improved KDE by calculating the bias in area estimation under a GRF approximation and applying a correction in an area-based coordinate system. By pulling the contours of the location distribution estimate inward towards the data without distorting its shape, this correction removes the tendency of GRF-based methods (including AKDE) to overestimate the area of home ranges, particularly at small *effective sample sizes* (**Figure 2**). Formally correcting the density function estimate allows us to calculate a more reliable home range area and confidence intervals. This correction can be applied to both conventional and autocorrelated GRF-KDEs (then termed KDE_c and AKDE_c, respectively), and is the default method within the `ctmm` package. As this source of bias is estimator-specific, the mitigation must also be estimator-specific, so this correction cannot be applied to non-GRF KDE approaches such as h_{LSCV}.

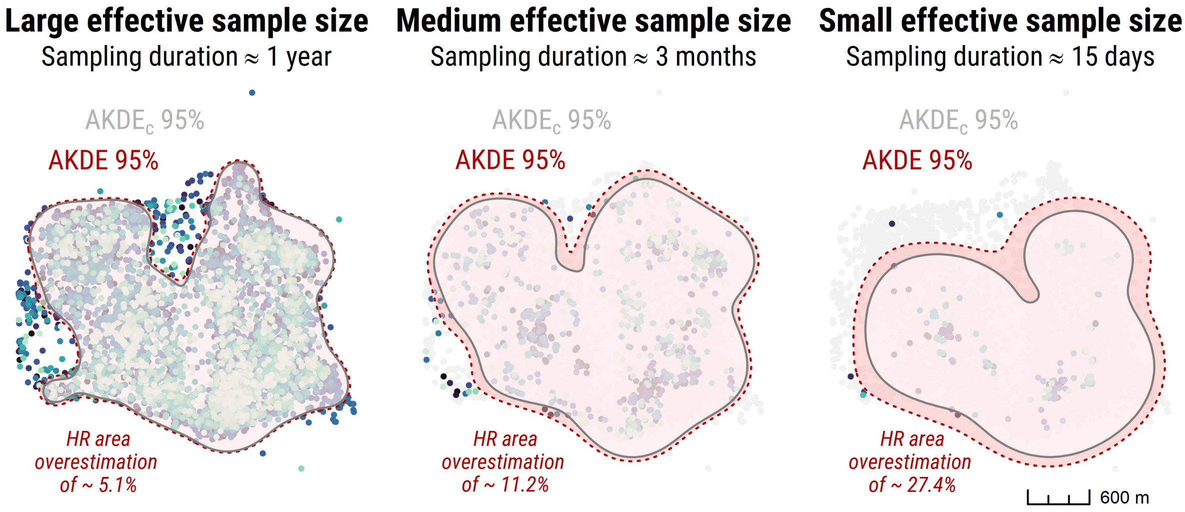


Figure 2. Autocorrelated Kernel Density Estimation (AKDE) and area-corrected AKDE (AKDE_c) calculated for one individual from the lowland tapir tracking dataset with: large *effective sample size* ($N \approx$ 1,566), medium *effective sample size* ($N \approx$ 261), and small *effective sample size* ($N \approx 30$). Displayed errors correspond to % bias of AKDE against AKDE_c of the same individual. Note that for large N values the estimates from AKDE and AKDE_c overlap considerably.

Bias III: Autocorrelation estimation bias

The main advantage of AKDE is that it accounts for the autocorrelated structure of animal movement data; for optimal performance, we need to estimate this autocorrelation correctly. Maximum Likelihood (ML) estimation is the standard approach to fitting movement models to animal tracking data (Horne, Garton, Krone, et al., 2007; Michelot et al., 2016) due to its versatility, widespread use, and relatively good performance (Pawitan, 2001). However, ML performs best at large sample sizes, while parameters related to variances and covariances tend to be underestimated in small sample size conditions (Cressie, 2015). As variance-associated parameters are closely related to home range size, their underestimation propagates into underestimated home range areas (Noonan et al., 2019).

Mitigation Measure III: pHREML and parametric bootstrapping

Residual Maximum Likelihood (REML) estimation is often used to improve (co)variance parameter estimation with small sample sizes, but it can perform poorly for the class of movement models on which AKDE depends (Fleming et al., 2019). To mitigate the small sample size bias in autocorrelation model parameter estimates, Fleming *et al.* (2019) developed a series of REML-based estimators that focus on small *effective sample sizes* (perturbative REML; pREML), small *absolute sample sizes* (Hybrid REML; HREML), or both *small absolute* and small *effective sample sizes* (perturbative Hybrid REML; pHREML). We focus on pHREML here (**Figure 3**) as it is the most broadly applicable of these methods and has no serious disadvantages relative to the others, because it combines the bias correction of REML and the stability of ML. It is currently the default parameter estimation method in the `ctmm` package.

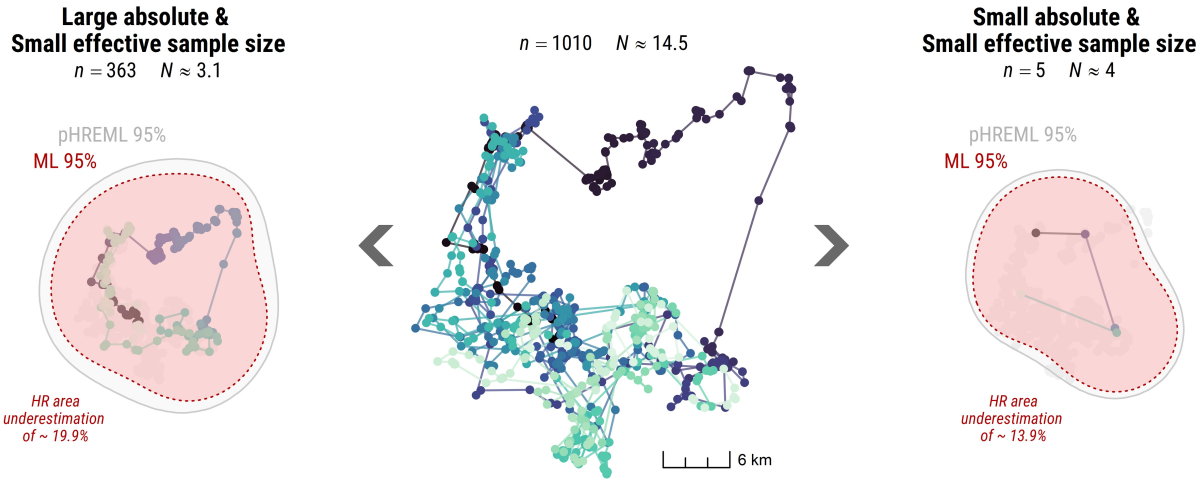


Figure 3. AKDE_C calculated with Maximum Likelihood (ML) and with perturbative Hybrid REML (pHREML) for an individual within the jaguar dataset, showcasing its effect on large absolute but small effective sample size (reduce to a sampling duration of 3 *months*: n = 362 locations, N ≈ 3.1), and both small absolute and small effective sample size (3 months thinned to n = 5 locations, N ≈ 4). Displayed errors correspond to % bias of ML-fitted AKDE against pHREML-fitted AKDE.

The parametric bootstrap method (Efron, 1982) is another standard solution for the biases caused by ML estimation and can be applied on top of REML-based estimations to further reduce biases. In extreme cases where *effective sample sizes* are ~5 or less, parametric bootstrapping may result in substantial improvements. However, the high computational cost incurred by bootstrapped pHREML (**Supplementary File 1**), coupled with the usually modest improvements it provides, reinforce its use only as a last resort.

Bias IV: Unrepresentative sampling in time

From a statistical perspective, evenly spaced temporal sampling of tracking data ensures the widest possible range of analytical options. In practice, however, many real-world issues can lead to animal locations being sampled irregularly in time: duty-cycling tags to avoid wasting battery during periods of inactivity, acceleration-informed sampling, device malfunction, habitat-related signal loss, and many other causes (DeCesare et al., 2005; Frair et al., 2004; Horne, Garton, & Kimberly A. Sager-Fradkin, 2007). When unaccounted for, such cases can yield biased datasets, causing area estimates associated with over-sampled portions of home ranges to be too large and those associated with under-sampled parts of home ranges to be too small (Fieberg, 2007). There is no guarantee that these contrasting biases cancel each other out, so the overall home range area estimate may be either positively or negatively biased.

Mitigation Measure IV: wAKDE

Weighted AKDE (or wAKDE) corrects for unrepresentative sampling in time (Fleming et al. 2018) through the larger bias addressed is *where* the area is distributed: it optimally upweights observations that occur during under-sampled times, while optimally downweighting observations occurring during over-sampled times. In IID data, optimal weights are uniform (*i.e.*, there is no temporal sampling bias, as all times are equally important) so there is no advantage to weighting.

For autocorrelated data with highly irregular sampling, however, the difference between weighted and unweighted AKDE can be considerable (**Figure 4**).

In practice, very few tracking datasets are perfectly regular, so it is essential to handle data irregularity appropriately. Missing data equate to a loss of information, and these errors can propagate into biases in habitat selection or area-based conservation outputs (Frair et al., 2004). For example, areas with good satellite reception (*e.g.*, open flat landscapes) may appear over-used even when animals did not spend more time in them compared to areas with poorer reception. Shifting sampling schedules (based on behavioral or seasonal patterns) is a common strategy employed in animal tracking projects, due to the trade-off between sampling intensity and battery life (Brown et al., 2012); in these circumstances, weight optimization via wAKDE is critical for comparisons between individuals or populations.

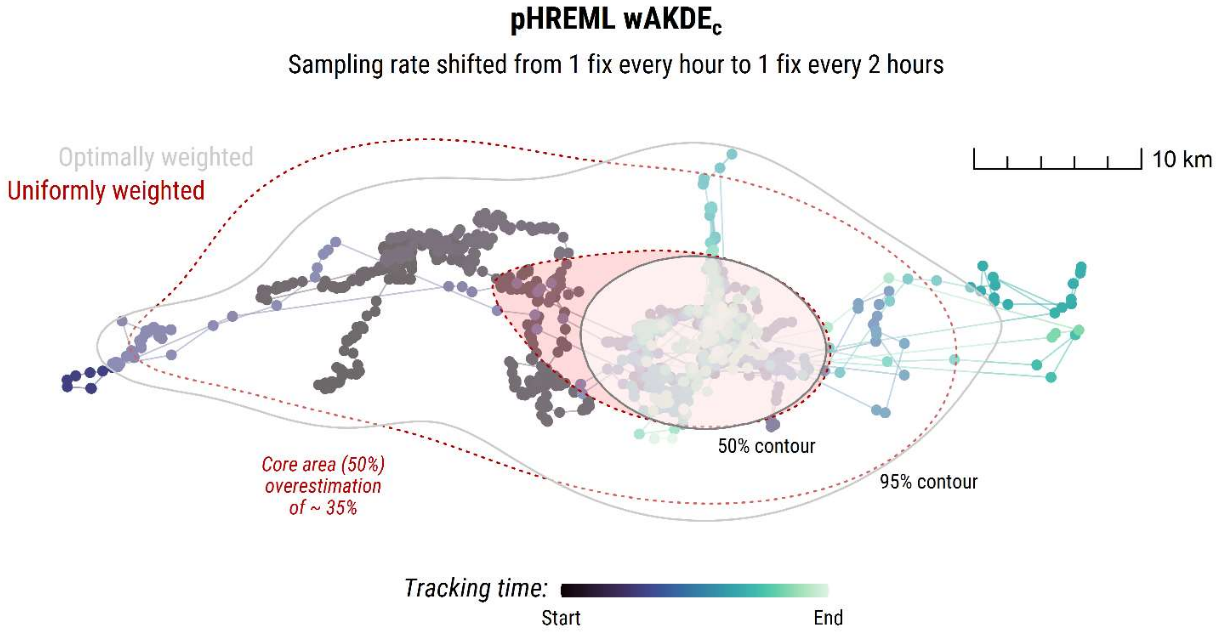


Figure 4. A uniformly weighted AKDEc and an optimally weighted AKDEc (wAKDEc), calculated from an individual from the African buffalo dataset with an irregular sampling schedule likely due to a device malfunction (nicknamed "Pepper"; available within the ctmm package). Displayed errors correspond to % bias of AKDEc core area (50%) against wAKDEc core area (50%).

Combination of mitigation measures

In practice, different sources of bias frequently occur together in the same datasets. This is a key reason why home ranges are so difficult to estimate accurately. However, the mitigation measures described above can be implemented simultaneously when necessary to combat multiple biases. For example, if a tracking dataset features autocorrelation, small *effective sample size*, and irregular temporal sampling, we can use pHREML to estimate and select the underlying movement model, and then pass the selected model to an optimally weighted area-corrected AKDE (w*AKDE$_c$*) to properly estimate home range area. The default settings in the `ctmm` package have been carefully chosen to balance performance against computational cost, so in this example, only optimal weighting would need to be manually selected by the user. Default values and alternative options are discussed in more detail in the `ctmm` documentation (Fleming & Calabrese, 2021).

To quantify the level of improvement offered by each mitigation measure and to explore the tradeoff between accuracy and computational cost, we performed a detailed simulation study. Our simulations are based on an OUF movement model, which features both correlated velocities (*i.e.*, directional persistence), correlated positions, and restricted space use. We chose the OUF model because it was the most frequently selected across all empirical GPS datasets in the Noonan *et al.* (2019) study, with 240 out of 369 datasets. We set both the directional persistence and range crossing timescales to one day, and varied the duration of the simulated datasets from 1 to 4,096 days in a doubling series, sampled hourly (except for bootstrapped pHREML wAKDE$_c$ which was significantly more computationally intensive and impractical to simulate over the whole sampling duration). This setup results in *effective sample sizes* that approximate the duration of each simulated dataset. We then sequentially fit home range estimators in the `ctmm` package to each simulated dataset in the following order: KDE, AKDE, AKDE$_c$, pHREML AKDE$_c$, pHREML wAKDE$_c$, and bootstrapped pHREML wAKDE$_c$. This represents a progression from no bias corrections (KDE) through all possible bias corrections applied simultaneously, in order of the typical importance of the corrections. Each simulation was repeated 400 times. We calculated bias as the 95% area estimate of the method in question divided by the exact expectation value of the true 95% area (under the model from which the data were simulated), while the computational cost was the time the simulation took to complete in seconds. All simulations were performed in the `R` environment (version 3.5.2; R Core Team 2018) using the `ctmm` package (version 0.5.2; Calabrese *et al.* 2016) and conducted on the University of Maryland High Performance Cluster.

Table 1. Mean improvement (%) in area estimation for each AKDE method compared to baseline KDE, over small (N < 32), medium (32 > N < 512), and large effective sample sizes (N > 512). Numbers in red denote underperformance.

	Improvement over KDE		
Methods	**Small *N***	**Medium *N***	**Large *N***
AKDE	51.3%	-31.3%	-78.4%
AKDE$_c$	42.6%	48.3%	36.4%
pHREML AKDE$_c$	59.8%	52.9%	40.4%
pHREML wAKDE$_c$	59.4%	52.9%	43.1%
Bootstrapped pHREML wAKDE$_c$	72.0%	NA	NA

Compared to conventional KDE, the original AKDE offered clear advantages for small *effective sample sizes,* but failed to improve area estimation for medium and large *effective sample sizes* (*N* > 32; **Table 1**). By solving the oversmoothing bias, AKDE$_c$ improved over KDE for all effective sample sizes. The next technique, pHREML-fitted AKDE$_c$ (`ctmm` default settings) further improved over conventional KDE and all previous measures, and stabilized the closest to 0% relative error after only eight sampling days (**Figure 5a**). Additional mitigation measures do lead to an increasingly higher computational cost: for the full sampling duration (4,096 days), pHREML-fitted wAKDE$_c$ ran on average 2.7 times longer than the original AKDE, and 230 times longer than a conventional KDE (**Figure 5b**, **Supplementary File 1**). With an Intel i7 3.9GHz processor using a single core, and an hourly tracking dataset collected for a year, this could correspond to an increase from a few seconds to approximately 45 minutes. However, unlike AKDE, conventional KDE does not run any autocorrelation model selection, or numerical optimization of parameter estimates.

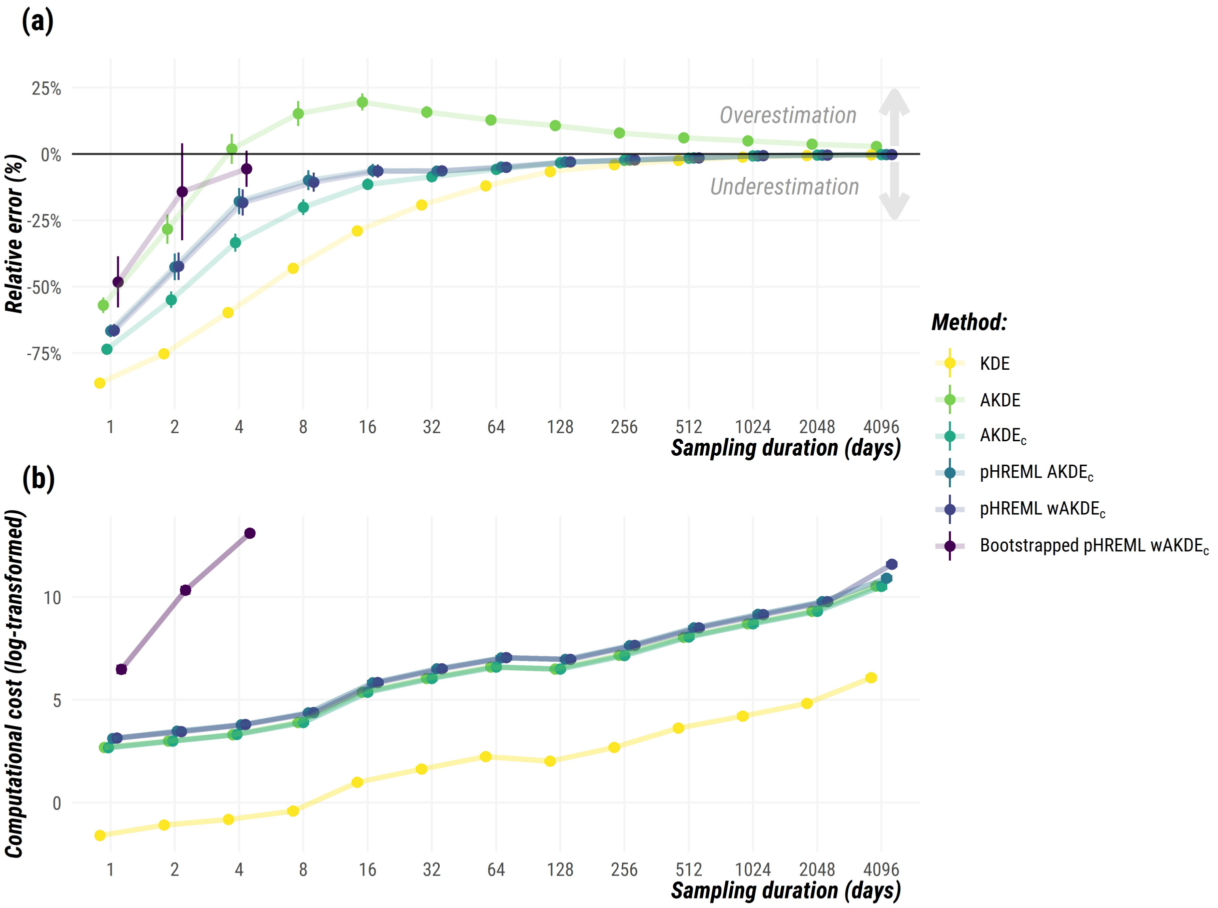


Figure 5. (a) Mean relative error (%) of the home range area estimation and **(b)** computational cost (log-transformed) for each method, by sampling duration (in days). Based on simulations (repeated 400 times) where the position and velocity autocorrelations were kept constant at 1 day and 1 hour, respectively. Error bars represent 95% confidence intervals. AKDE and AKDEc have almost overlapping performances, as well as pHREML-fitted AKDEc and pHREML wAKDEc.

Discussion

The techniques presented in this paper represent a family of home range estimators starting with conventional GRF-KDE and progressing through a series of estimation methods designed to mitigate bias arising when the core assumption of IID data is not met. These methods are implemented with efficient computational algorithms that work with both small and large animal tracking datasets. We have brought these techniques together in a single document to demonstrate when each correction is applicable, the degree to which home range estimates can be improved, and when and how they can be combined to handle the unique quirks of each tracking dataset to yield accurate home range estimates.

The AKDE family of estimators are all implemented in the `ctmm` `R` package (Calabrese et al., 2016), so we provide an annotated `R` script in the supplementary material of this paper to guide users through the applications of these techniques (**Supplementary File 2**). The current default settings are pHREML, for estimating movement model parameters, and (A)KDE$_c$, for estimating home ranges. The decision between KDE$_c$ and AKDE$_c$ is determined using model selection, and dependent on whether the data are independently distributed or autocorrelated, respectively. We recommend that users keep pHREML and (A)KDE$_c$ as the default settings and especially caution against changing these settings for any *effective sample sizes* below 20. When working with legacy data where small *effective sample sizes* are a serious concern, but additional data collection is not an option, bootstrapped pHREML may be used as a method of last resort to obtain the best possible home range estimates. However, due to the high computational cost of using pHREML, users should decide on a bias threshold (typically > 1–5% with an initial pHREML estimate) before applying this measure (see **Supplementary File 2** for how to determine this bias threshold). Finally, wAKDE$_c$ can account for temporal sampling bias (*i.e.*, missing or irregular tracking data), but is switched off by default due to its considerable computational cost (approximately 200 times longer than KDE).

Most mitigation measures we have discussed here become increasingly valuable at small sample sizes (absolute and/or effective), allowing researchers to maximize the value of legacy datasets or to handle situations where larger *effective sample sizes* are impossible to obtain (*e.g.*, on smaller animals for which battery size limits the temporal duration of GPS lifespans). Nevertheless, the application of these analytical methods should not replace careful study design before data collection, tailored to address specific research questions (Fieberg & Börger, 2012). When the main goal of a study is home range estimation, the sampling duration should be many times larger than the average *range crossing time* of the focal species. This ensures that the *effective sample size* will be large enough to facilitate reliable estimation of the area of home ranges. Achieving this goal may require researchers to decrease the sampling rate of their devices to save battery power, although small *effective sample sizes* may be inevitable even in these conditions due to battery constraints, specific ranging behaviors, or short lifespans.

Although AKDEs provide reliable home range area estimations in the conditions presented in this manuscript, there are scenarios in which they fail. A known issue of KDE methods is that their estimates can spill over discontinuities, such as hard boundaries (*e.g.*, coastal lines and water sources for terrestrial species, land for aquatic species) or three-dimensional features (*e.g.*, steep terrain, impermeable barriers), and have difficulties resolving narrow movement corridors (Guo et al., 2019; Péron, 2019; Silverman, 1986; Worton, 1995). The positive bias from boundary spillover is likely less influential than the negative bias due to unmodeled autocorrelation; nevertheless, it is possible to correct for hard boundaries by following the workflow presented in appendix 3 of Noonan *et al.* (2019). Kernel density methods also fail to adequately resolve non-stationary behavior and nomadism (Lichti & Swihart, 2011; Nandintsetseg et al., 2019), as nomadic species lack site fidelity to movement pathways or key sites (*e.g.*, breeding or wintering areas). Addressing non-stationarity requires home range estimates that accommodate multiple centers and allowing for variation in use patterns (Breed et al., 2017). In addition, a misspecified model due to migratory behaviors will affect the accuracy of AKDE area outputs due to the stationary movement models being leveraged (OU, OUF, IID). However, if an animal is not range resident, then the data are not appropriate for any home range estimation method.

Moving forward, we hope to address two remaining challenges in home range estimation: location error and resource selection (which includes boundary interactions). Home range estimation is not as sensitive to location error as fine-scale quantities, such as speed estimation (Noonan et al., 2019). However, any biological inferences can become compromised if location errors are comparable to the relevant movement scales. At present, locations errors can be partially accounted for in the autocorrelation modeling stage of the `ctmm` workflow (Fleming et al., 2021). First, the autocorrelation and bandwidth calculations are error-informed, which mitigates various biases in autocorrelation and bandwidth estimates that would otherwise occur had location error been ignored. Second, location estimates are fed through a Kalman smoother before kernel placement, which counteracts overdispersal. However, there are two adjustments that could provide further improvements. First, more erroneous location estimates could be optimally downweighted in the sense of Fleming *et al.* (2018), which would increase statistical efficiency. Second, instead of employing a Kalman smoother that is only consistent with normal distributions, the data could be nonlinearly smoothed via iteration. Finally, the inclusion of resource selection parameters into KDE has been considered by Guo *et al.* (2019) and Péron (2019). Resource selection can include the respecting of hard boundaries, as well as softer habitat preferences. The influence of resource selection on AKDE can be considered at various stages, including autocorrelation modeling, bandwidth optimization, and kernel shape, and will likely take multiple research efforts to fully implement in a general use software solution.

Only by estimating home ranges in a comparable way across sampling schedules, study designs, and behavioral idiosyncrasies can wildlife researchers provide wildlife managers and practitioners with accurate information for conservation planning and land-use decision-making. Movement ecology has reached an inflection point where it is no longer possible to ignore autocorrelation: using autocorrelated tracking datasets with estimators that assume IID data will result in underestimated home range areas (Noonan et al. 2019). Although further technological advances will only increase the amount of autocorrelation present in tracking data, autocorrelation is often still present even in VHF data and should not be overlooked. We have provided guidelines to obtain accurate home range area estimates with the AKDE family of home range estimators which, in their current form, provide the most reliable and flexible solution for home range area estimation. These methods were explicitly designed to work synergistically, eliminating discrepancies between empirical reality and estimator assumptions that drive home range under- or overestimation with conventional techniques. Furthermore, these techniques can be implemented with open source software and code (Calabrese et al., 2016, 2021), and new movement processes can be easily added into the AKDE workflow as they are developed. This flexibility “future proofs” the AKDE family of analyses by allowing it to be tailored to new datasets, movement behaviors, and species as necessary.

Author contributions

CHF, MJN, WFF and JMC conceived the ideas. CF conducted the simulations. JMC and IS led the writing of the manuscript. All authors contributed critically to the draft and gave final approval for publication.

Conflict of Interest statement

The authors declare no conflict of interest.

Data availability statement

All empirical datasets used in the manuscript are currently openly accessible: the African buffalo tracking data are archived in the MoveBank Data Repository (Cross et al., 2016) and partially included in the `ctmm` package (Fleming & Calabrese, 2021); lowland tapir tracking data are archived in the Dryad Digital Repository (Fleming et al., 2019); jaguar tracking data are available as a data paper (Morato et al., 2018) and partially included in the `ctmm` package (Fleming & Calabrese, 2021). Simulation data and the code required to reproduce the `R` tutorial are available on GitHub (https://github.com/ecoisilva/AKDE_minireview).

Acknowledgements

This work was partially funded by the Center of Advanced Systems Understanding (CASUS), which is financed by Germany's Federal Ministry of Education and Research (BMBF) and by the Saxon Ministry for Science, Culture and Tourism (SMWK) with tax funds on the basis of the budget approved by the Saxon State Parliament. WFF, CHF, and JMC were supported by NSF IIBR 1915347.

References

Alston, J. M., & Rick, J. A. (2020). A Beginner's Guide to Conducting Reproducible Research. *The Bulletin of the Ecological Society of America*.

Bartoń, K. A., Zwijacz-Kozica, T., Zięba, F., Sergiel, A., & Selva, N. (2019). Bears without borders: Long-distance movement in human-dominated landscapes. *Global Ecology and Conservation*, *17*, e00541.

Blundell, G. M., Maier, J. A., & Debevec, E. M. (2001). Linear home ranges: Effects of smoothing, sample size, and autocorrelation on kernel estimates. *Ecological Monographs*, *71*(3), 469–489.

Boyce, M. S., Pitt, J., Northrup, J. M., Morehouse, A. T., Knopff, K. H., Cristescu, B., & Stenhouse, G. B. (2010). Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2213–2219.

Breed, G. A., Golson, E. A., & Tinker, M. T. (2017). Predicting animal home-range structure and transitions using a multistate Ornstein-Uhlenbeck biased random walk. *Ecology*, *98*(1), 32–47.

Brown, D. D., LaPoint, S., Kays, R., Heidrich, W., Kümmeth, F., & Wikelski, M. (2012). Accelerometer-informed GPS telemetry: Reducing the trade-off between resolution and longevity. *Wildlife Society Bulletin*, *36*(1), 139–146. https://doi.org/10.1002/wsb.111

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

Calabrese, J. M., Fleming, C. H., & Gurarie, E. (2016). ctmm: An r package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution*, *7*(9), 1124–1132.

Calabrese, J. M., Fleming, C. H., Noonan, M. J., & Dong, X. (2021). ctmmweb: A graphical user interface for autocorrelation-informed home range estimation. *Wildlife Society Bulletin*, *2021*, 1–8. https://doi.org/10.1002/wsb.1154

Chen, Y.-C. (2017). A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, *1*(1), 161–187.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

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

Cross, P. C., Heisey, D. M., Bowers, J. A., Hay, C. T., Wolhuter, J., Buss, P., Hofmeyr, M., Michel, A. L., Bengis, R. G., & Bird, T. L. F. (2009). Disease, predation and demography: Assessing the impacts of bovine tuberculosis on African buffalo by monitoring at individual and population levels. *Journal of Applied Ecology*, *46*(2), 467–475.

Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLoS Biology*, *18*(7), e3000763.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science*, *361*(6407), 1108–1111.

Dale, M. R., & Fortin, M.-J. (2002). Spatial autocorrelation and statistical tests in ecology. *Ecoscience*, *9*(2), 162–167.

De Solla, S. R., Bonduriansky, R., & Brooks, R. J. (1999). Eliminating autocorrelation reduces biological relevance of home range estimates. *Journal of Animal Ecology*, *68*(2), 221–234.

DeCesare, N. J., Squires, J. R., & Kolbe, J. A. (2005). Effect of forest canopy on GPS-based movement data. *Wildlife Society Bulletin*, *33*(3), 935–941.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM.

Fieberg, J. (2007). Utilization distribution estimation using weighted kernel density estimators. *The Journal of Wildlife Management*, *71*(5), 1669–1675.

Fieberg, J., & Börger, L. (2012). Could you please phrase “home range” as a question? *Journal of Mammalogy*, *93*(4), 890–902.

Fleming, C. H., & Calabrese, J. M. (2017). A new kernel density estimator for accurate home-range and species-range area estimation. *Methods in Ecology and Evolution*, *8*(5), 571–579.

Fleming, C. H., & Calabrese, J. M. (2021). *Ctmm: Continuous-Time Movement Modeling. R package version 0.6.0*. (R package version 0.6.0) [Computer software]. https://CRAN.R-project.org/package=ctmm

Fleming, C. H., Calabrese, J. M., Mueller, T., Olson, K. A., Leimgruber, P., & Fagan, W. F. (2014). From fine-scale foraging to home ranges: A semivariance approach to identifying movement modes across spatiotemporal scales. *The American Naturalist*, *183*(5), E154–E167.

Fleming, C. H., Drescher-Lehman, J., Noonan, M. J., Akre, T. S. B., Brown, D. J., Cochrane, M. M., Dejid, N., DeNicola, V., DePerno, C. S., Dunlop, J. N., Gould, N. P., Harrison, A.-L., Hollins, J., Ishii, H., Kaneko, Y., Kays, R., Killen, S. S., Koeck, B., Lambertucci, S. A., … Calabrese, J. M. (2021). *A comprehensive framework for handling location error in animal tracking data* (p. 2020.06.12.130195). https://doi.org/10.1101/2020.06.12.130195

Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., & Calabrese, J. M. (2015). Rigorous home range estimation with movement data: A new autocorrelated kernel density estimator. *Ecology*, *96*(5), 1182–1188.

Fleming, C. H., Noonan, M. J., Medici, E. P., & Calabrese, J. M. (2019). Overcoming the challenge of small effective sample sizes in home-range estimation. *Methods in Ecology and Evolution*, *10*(10), 1679–1689.

Fleming, C. H., Sheldon, D., Fagan, W. F., Leimgruber, P., Mueller, T., Nandintsetseg, D., Noonan, M. J., Olson, K. A., Setyawan, E., & Sianipar, A. (2018). Correcting for missing and irregular data in home-range estimation. *Ecological Applications*, *28*(4), 1003–1010.

Frair, J. L., Nielsen, S. E., Merrill, E. H., Lele, S. R., Boyce, M. S., Munro, R. H., Stenhouse, G. B., & Beyer, H. L. (2004). Removing GPS collar bias in habitat selection studies. *Journal of Applied Ecology*, *41*(2), 201–212.

Guo, J., Du, S., Ma, Z., Huo, H., & Peng, G. (2019). A Model for Animal Home Range Estimation Based on the Active Learning Method. *ISPRS International Journal of Geo-Information*, *8*(11), 490.

Hansen, M. C., Wang, L., Song, X.-P., Tyukavina, A., Turubanova, S., Potapov, P. V., & Stehman, S. V. (2020). The fate of tropical forest fragments. *Science Advances*, *6*(11), eaax8574.

Harris, S., Cresswell, W. J., Forde, P. G., Trewhella, W. J., Woollard, T., & Wray, S. (1990). Home-range analysis using radio-tracking data–a review of problems and techniques particularly as applied to the study of mammals. *Mammal Review*, *20*(2-3), 97–123.

Hemson, G., Johnson, P., South, A., Kenward, R., Ripley, R., & MACDONALD, D. (2005). Are kernels the mustard? Data from global positioning system (GPS) collars suggests problems for kernel home-range analyses with least-squares cross-validation. *Journal of Animal Ecology*, *74*(3), 455–463.

Horne, J. S., Fieberg, J., Börger, L., Rachlow, J. L., Calabrese, J. M., & Fleming, C. H. (2020). Animal Home Ranges. *Population Ecology in Practice*.

Horne, J. S., Garton, E. O., & Kimberly A. Sager-Fradkin. (2007). Correcting home-range models for observation bias. *The Journal of Wildlife Management*, *71*(3), 996–1001.

Horne, J. S., Garton, E. O., Krone, S. M., & Lewis, J. S. (2007). Analyzing animal movements using Brownian bridges. *Ecology*, *88*(9), 2354–2363.

Jennrich, R., & Turner, F. (1969). Measurement of non-circular home range. *Journal of Theoretical Biology*, *22*(2), 227–237.

Katajisto, J., & Moilanen, A. (2006). Kernel-based home range method for data with irregular sampling intervals. *Ecological Modelling*, *194*(4), 405–413.

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*(6240).

Kie, J. G., Matthiopoulos, J., Fieberg, J., Powell, R. A., Cagnacci, F., Mitchell, M. S., Gaillard, J.-M., & Moorcroft, P. R. (2010). The home-range concept: Are traditional estimators still relevant with modern telemetry technology? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2221–2231.

Lambertucci, S. A., Alarcón, P. A., Hiraldo, F., Sanchez-Zapata, J. A., Blanco, G., & Donázar, J. A. (2014). Apex scavenger movements call for transboundary conservation policies. *Biological Conservation*, *170*, 145–150.

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, *74*(6), 1659–1673.

Lichti, N. I., & Swihart, R. K. (2011). Estimating utilization distributions with kernel versus local convex hull methods. *The Journal of Wildlife Management*, *75*(2), 413–422.

Linnell, J. D., Aanes, R., Swenson, J. E., Odden, J., & Smith, M. E. (1997). Translocation of carnivores as a method for managing problem animals: A review. *Biodiversity & Conservation*, *6*(9), 1245–1257.

Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., Jiang, N., & Halpern, B. S. (2017). Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*, *1*(6), 1–7.

Michelot, T., Langrock, R., & Patterson, T. A. (2016). MoveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, *7*(11), 1308–1315.

Morato, R. G., Thompson, J. J., Paviolo, A., de La Torre, J. A., Lima, F., McBride Jr, R. T., Paula, R. C., Cullen Jr, L., Silveira, L., & Kantek, D. L. (2018). *Jaguar movement database: A GPS-based movement dataset of an apex predator in the neotropics*. Wiley Online Library.

Nandintsetseg, D., Bracis, C., Olson, K. A., Böhning-Gaese, K., Calabrese, J. M., Chimeddorj, B., Fagan, W. F., Fleming, C. H., Heiner, M., & Kaczensky, P. (2019). Challenges in the conservation of wide-ranging nomadic species. *Journal of Applied Ecology*, *56*(8), 1916–1926.

Noonan, M. J., Tucker, M. A., Fleming, C. H., Akre, T. S., Alberts, S. C., Ali, A. H., Altmann, J., Antunes, P. C., Belant, J. L., & Beyer, D. (2019). A comprehensive analysis of autocorrelation and bias in home range estimation. *Ecological Monographs*, *89*(2), e01344.

Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood*. Oxford University Press.

Péron, G. (2019). Modified home range kernel density estimators that take environmental interactions into account. *Movement Ecology*, *7*(1), 1–8.

Powell, R. A. (2000). Animal home ranges and territories and home range estimators. *Research Techniques in Animal Ecology: Controversies and Consequences*, *442*, 65–110.

Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological Applications*, *29*(1), e01822. https://doi.org/10.1002/eap.1822

Rempel, R. S., Rodgers, A. R., & Abraham, K. F. (1995). Performance of a GPS animal location system under boreal forest canopy. *The Journal of Wildlife Management*, 543–551.

Rooney, S. M., Wolfe, A., & Hayden, T. J. (1998). Autocorrelated data in telemetry studies: Time to independence and the problem of behavioural effects. *Mammal Review*, *28*(2), 89–98.

Seaman, D. E., & Powell, R. A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*, *77*(7), 2075–2085.

Signer, J., & Fieberg, J. R. (2021). A fresh look at an old concept: Home-range estimation in a tidy world. *PeerJ*, *9*, e11031.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.

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

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, *36*(5), 823.

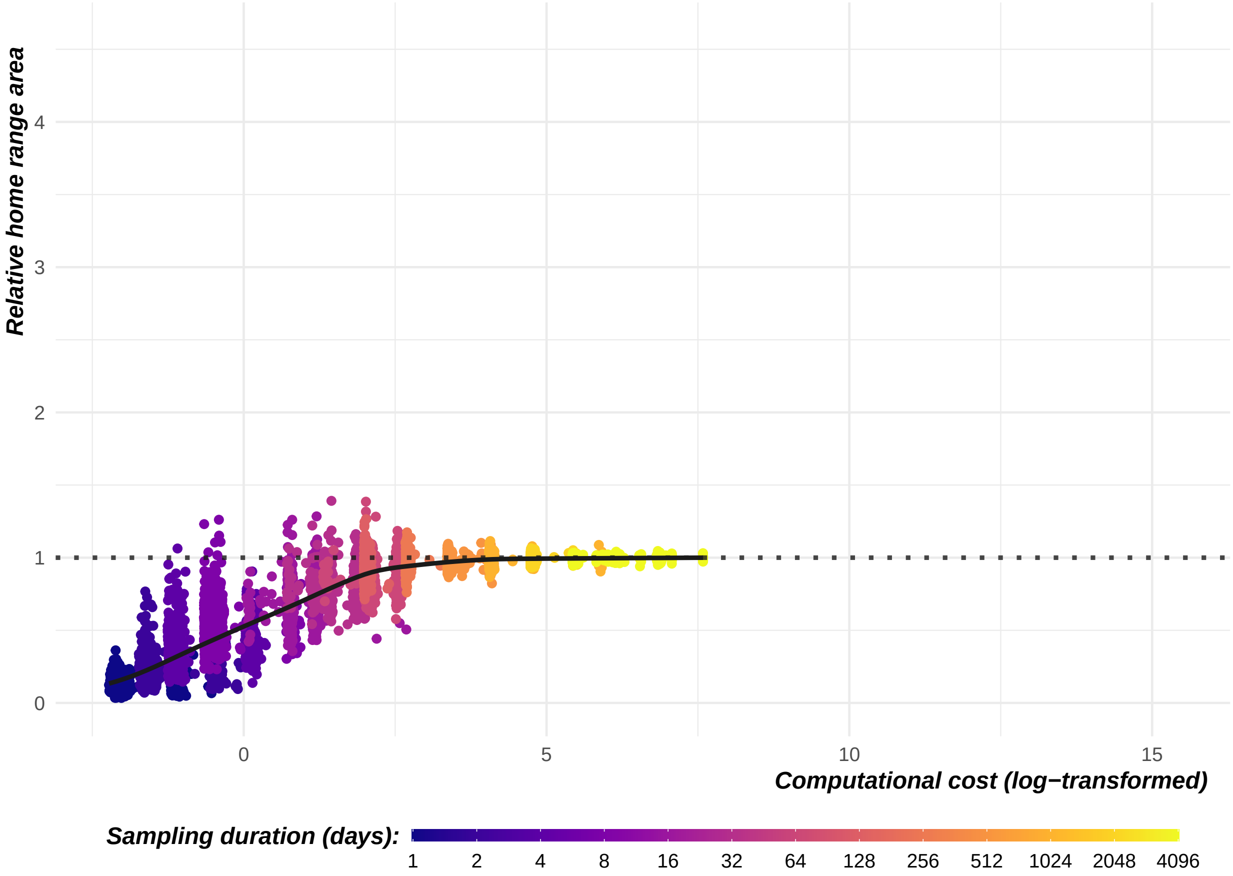
Wang, S., Wang, J., & Chung, F. (2013). Kernel density estimation, kernel methods, and fast learning in large data sets. *IEEE Transactions on Cybernetics*, *44*(1), 1–20.

Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, *70*(1), 164–168.

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

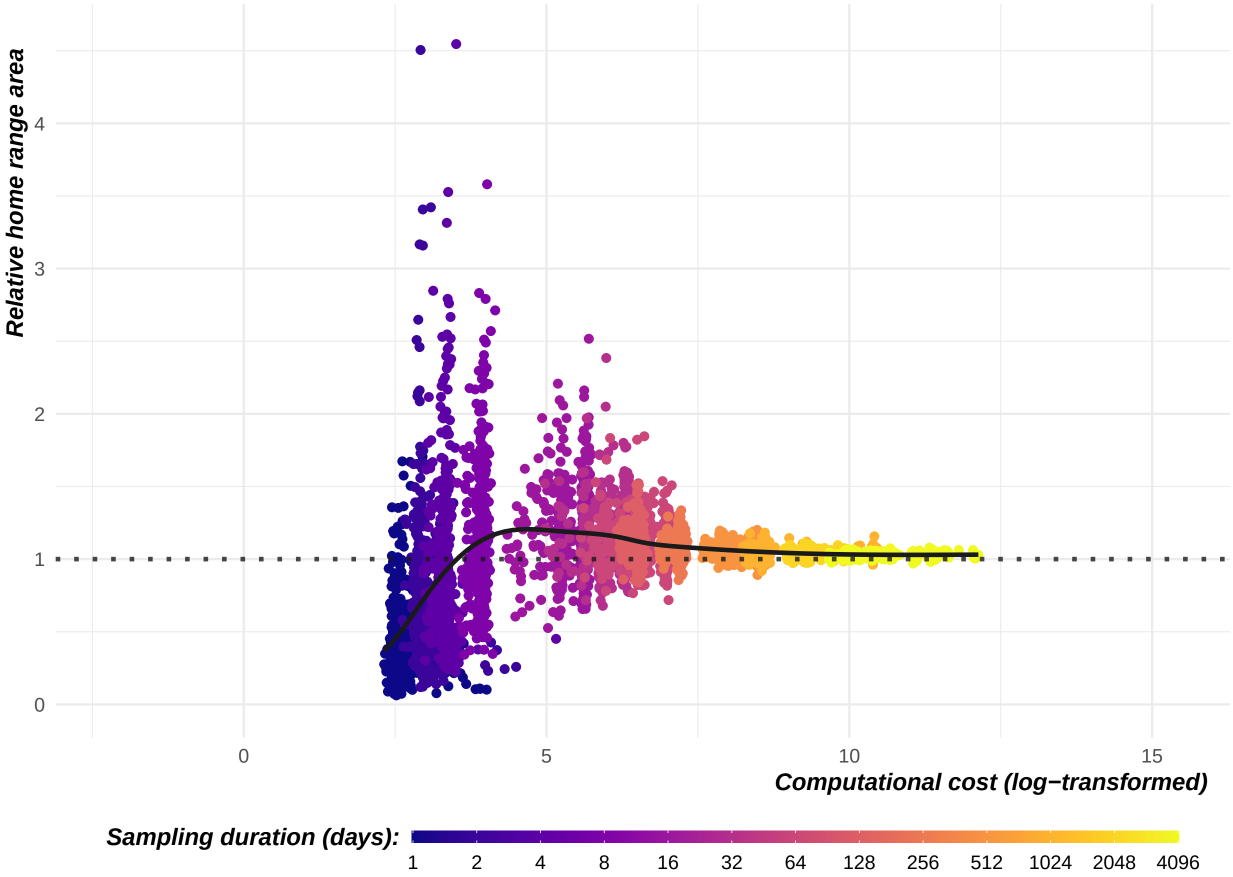
Supplementary file 1: Computational costs

KDE



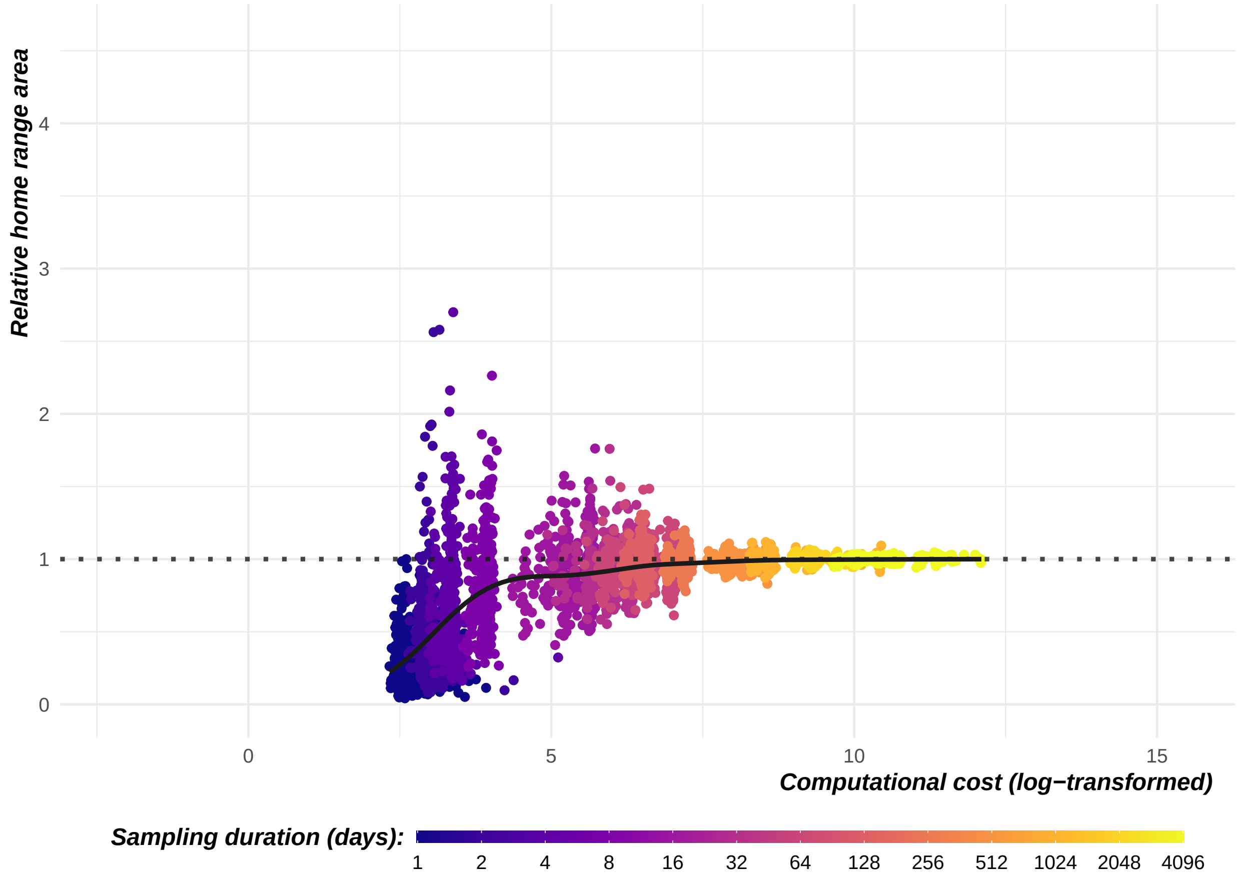
Supplementary figure 1. Relative home range area, *i.e.*, where the true home range area is scaled to 1, versus computational cost (in logarithmic scale) for the **Kernel Density Estimation (KDE)** method. Based on simulations (repeated 400 times) of different sampling durations (1 to 4,096 days).

AKDE



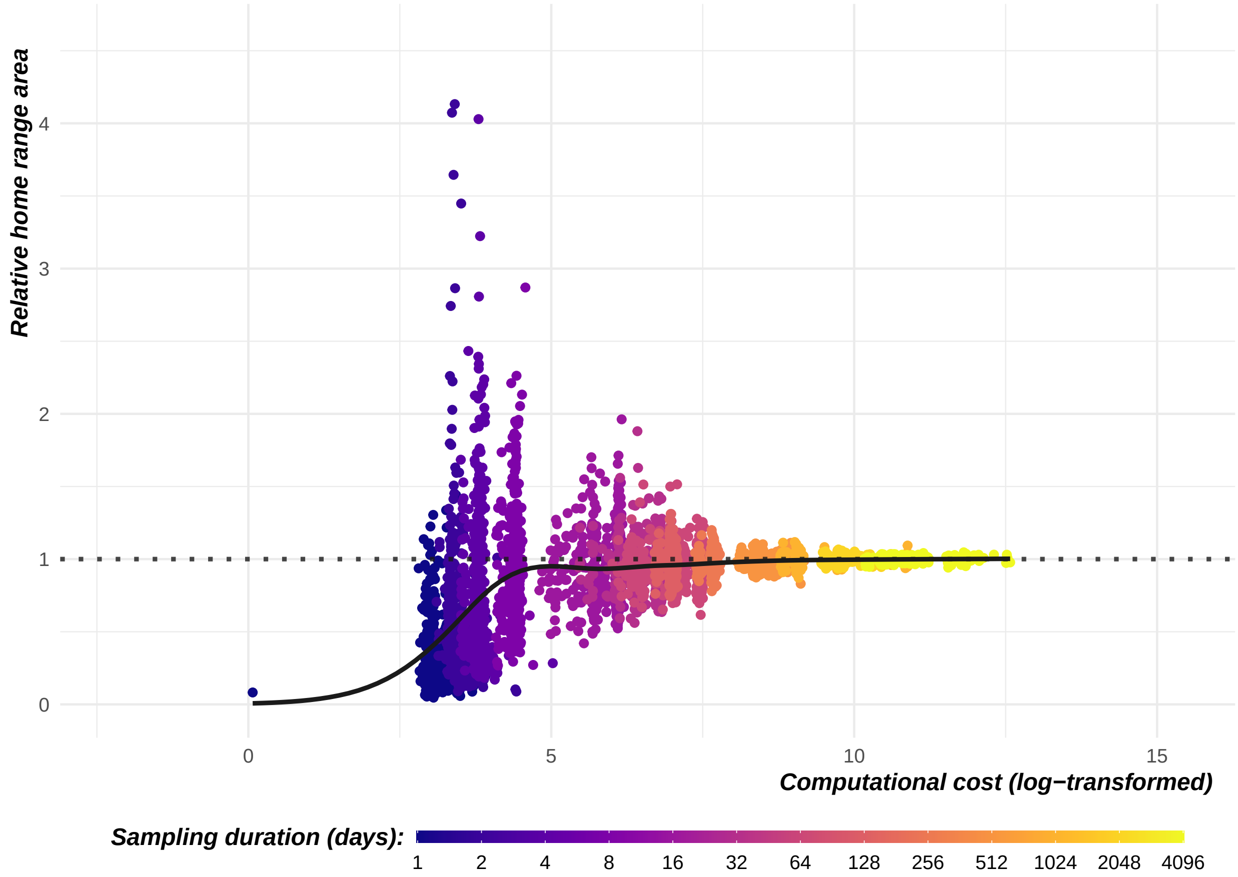
Supplementary figure 2. Relative home range area, *i.e.*, where the true home range area is scaled to 1, versus computational cost (in logarithmic scale) for the **Autocorrelated Kernel Density Estimation (AKDE)** method. Based on simulations (repeated 400 times) of different sampling durations (1 to 4,096 days).

$AKDE_c$



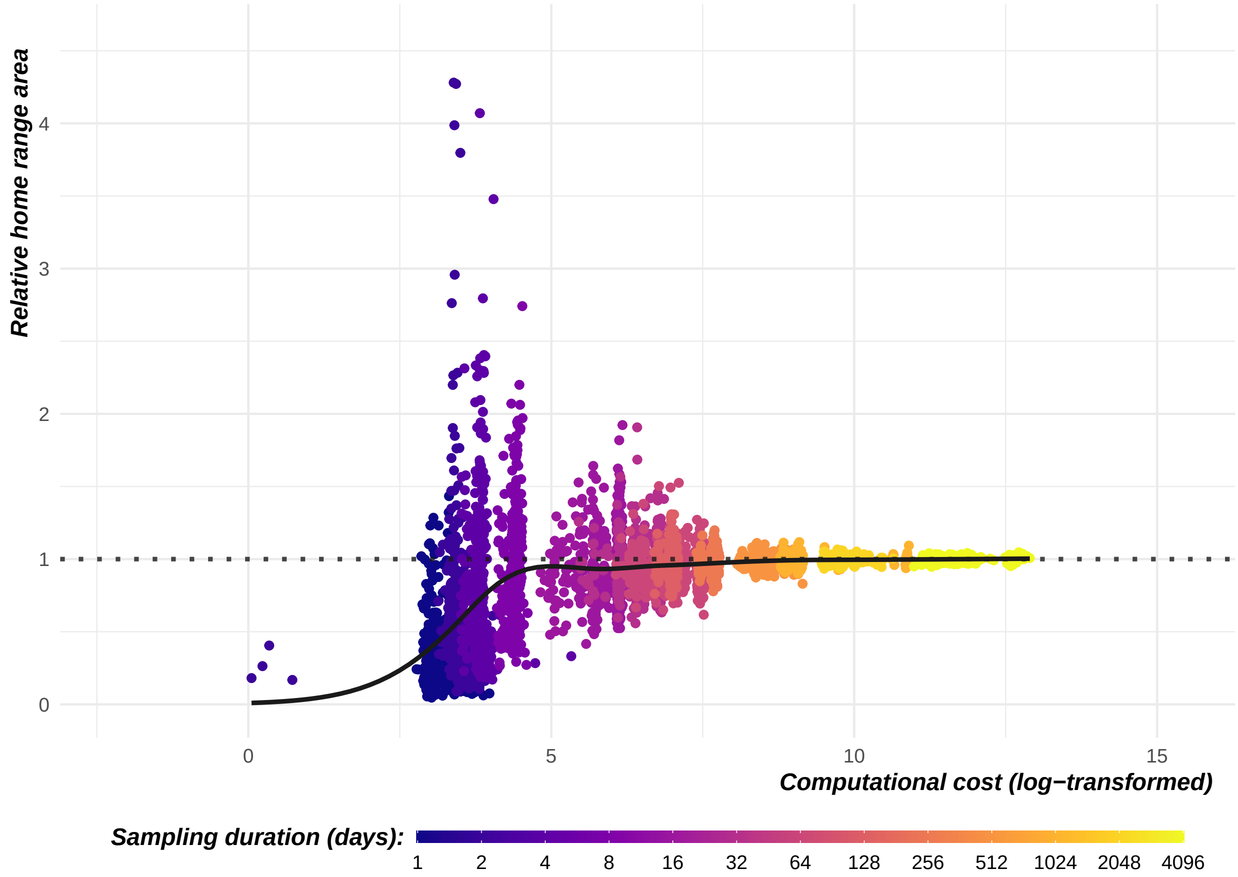
Supplementary figure 3. Relative home range area, *i.e.*, where the true home range area is scaled to 1, versus computational cost (in logarithmic scale) for the **area-corrected Autocorrelated Kernel Density Estimation ($\mathbf{AKDE_c}$)** method. Based on simulations (repeated 400 times) of different sampling durations (1 to 4,096 days).

pHREML AKDE$_c$



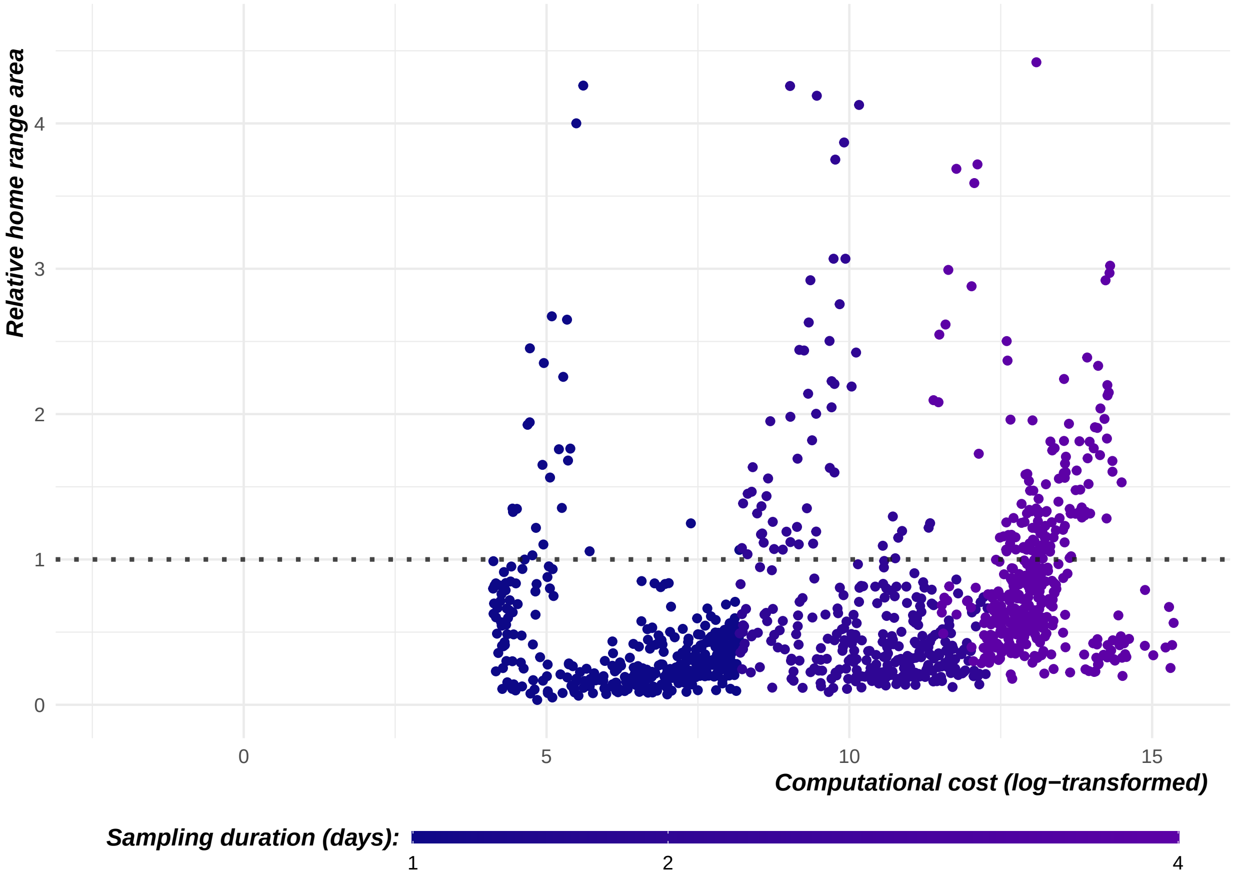
Supplementary figure 4. Relative home range area, *i.e.*, where the true home range area is scaled to 1, versus computational cost (in logarithmic scale) for the **area-corrected Autocorrelated Kernel Density Estimation** method fitted with **perturbative Hybrid REML (pHREML AKDE$_c$)**. Based on simulations (repeated 400 times) of different sampling durations (1 to 4,096 days).

pHREML wAKDE$_c$



Supplementary figure 5. Relative home range area, *i.e.*, where the true home range area is scaled to 1, versus computational cost (in logarithmic scale) for the **weighted, area-corrected, Autocorrelated Kernel Density Estimation** method fitted with **perturbative Hybrid REML (pHREML wAKDE$_c$)**. Based on simulations (repeated 400 times) of different sampling durations (1 to 4,096 days).

Bootstrapped pHREML wAKDE$_c$



Supplementary figure 6. Relative home range area, *i.e.*, where the true home range area is scaled to 1, versus computational cost (in logarithmic scale) for the **boostrapped, weighted, area-corrected Autocorrelated Kernel Density Estimation** method fitted with **perturbative Hybrid REML (Bootstrapped pHREML wAKDE$_c$)**. Based on simulations (repeated 400 times) of different sampling durations (1 to 4 days).

Supplementary file 2: Autocorrelation-informed home range estimation with the ctmm R package

Introduction

Home range estimation is a key output from animal tracking datasets, but the inherent properties of animal movement can lead traditional methods to under- or overestimated their size. **Autocorrelated Kernel Density Estimation (AKDE)** methods were designed to be statistically efficient while explicitly dealing with the complexities and biases of modern movement data, such as *autocorrelation*, *small sample sizes*, and *missing or irregularly sampled data*.

This tutorial is a companion piece to our manuscript *"Autocorrelation-informed home range estimation: a review and practical guide"*. The **AKDE** family of home range estimators will be run using **R software** (https://www.r-project.org/) and the `ctmm` package (Calabrese et al., 2016). For any definitions, check the main manuscript or the *Glossary*. The techniques and mitigation measures available in this package include:

Method	**When to run?**	**What does it do?**	**R function**
AKDE	Tracking data is autocorrelated	Estimates range distributions from autocorrelated data, by conditioning on an autocorrelation model.	`akde(…, debias = FALSE)`
AKDEc	If using GRF-based KDEs (such as AKDE)	Removes the tendency of Gaussian reference function (GRF) methods to overestimate the area of home ranges.	`akde(…, debias = TRUE)`
pHREML	Small (absolute and effective) sample sizes	Improves upon ML and REML autocorrelation estimation, mitigating small sample size biases.	`ctmm.select(…, method = "pHREML")`
wAKDEc	Irregular sampling schedules or missing data	Upweights observations that occur during under-sampled times, while downweighting those occurring during over-sampled times.	`akde(…, weights = TRUE)`
Parametric bootstrap	Extremely small effective sample size	Calculates and corrects for autocorrelation estimation biases, by simulating from an approximate sampling distribution.	`ctmm.boot(...)`

AKDE$_{\text{c}}$ and pHREML are default arguments within the `akde()` and `ctmm.select()` functions, respectively. Both measures will run automatically if arguments `debias` and `method` are left unspecified. For most situations, we recommend keeping both of these arguments as the default.

```
# Installing & loading package:
install.packages("ctmm")
library(ctmm)
```

We provide a guide to **home range estimation** using the following workflow:

- **Step 1.** – Formatting and loading an animal tracking dataset;
- **Step 2.** – Checking for the *range residency* assumption;
- **Step 3.** – Selecting the best-fit movement model through *model selection*;
- **Step 4.** – Feeding a movement model into the *home range estimator*;
- **Step 5.** – Evaluating additional *biases*, applying *mitigation measures.*

Data Preparation

We will use two datasets, both available within the `ctmm` package: African buffalos (*Syncerus caffer*), and Mongolian gazelles (*Procapra gutturosa*). Information on the data collection protocol is available in Cross *et al.* (2009) and Fleming *et al.* (2014). The `ctmm` package requires data to conform to Movebank naming conventions (https://www.movebank.org/node/2381). We recommend uploading your data to Movebank (http://www.movebank.org/) as this will facilitate data preparation, and ensure that your data are correctly formatted for `ctmm`. If needed, Movebank allows you to keep your data private.

We will focus on the simplest data structure:

- `animal ID` or `ID` — An individual identifier for each animal tracked;
- `timestamp` or `t` — The date and time corresponding to a sensor measurement;
  - **Example**: 2021-01-01 18:31:00.000
    - **Format**: yyyy-MM-dd HH:mm:ss.SSS
- `longitude` or `x` — The geographic longitude of the location as estimated by the sensor.
  - **Example**: -121.1761111
    - **Units**: decimal degrees, WGS84 reference system.
- `latitude` or `y` — The geographic latitude of the location as estimated by the sensor;
  - **Example**: -41.0982423
    - **Units**: decimal degrees, WGS84 reference system.

Location can also be described as UTM locations instead of latitude/longitude. In this case, you should provide `UTM easting`, `UTM northing`, and `UTM zone`. For all terms and conventions, please see the full vocabulary list here: http://vocab.nerc.ac.uk/collection/MVB/current/.

Step 1. – Formatting and loading an animal tracking dataset

You can import data into R through the `read.table()` or `read.csv()` functions; make sure to navigate to the appropriate folder or working directory. For example:

```
# First, list all files in your current directory:
list.files() # make sure this leads to the location of your file
```

```
# Second, load the correct file:
animal0_longlat <- read.csv(here("example_data_longlat.csv"))
head(animal0_longlat)
```

```
##        ID        timestamp longitude latitude
## 1 animal0 2005-07-14 05:35  30.59648 65.24774
## 2 animal0 2005-07-14 07:35  30.53256 65.27085
## 3 animal0 2005-07-14 08:34  30.52398 65.26446
## 4 animal0 2005-07-14 09:35  30.52848 65.25948
## 5 animal0 2005-07-14 10:35  30.53382 65.26643
## 6 animal0 2005-07-14 11:34  30.49970 65.27868
```

```
animal0_utm <- read.csv(here("example_data_utm.csv"))
head(animal0_utm)
```

```
##        ID        timestamp UTM.Easting UTM.Northing UTM.zone
## 1 animal0 2005-07-14 05:35    387730.0      7238204      36N
## 2 animal0 2005-07-14 07:35    384846.3      7240894      36N
## 3 animal0 2005-07-14 08:34    384418.1      7240197      36N
## 4 animal0 2005-07-14 09:35    384606.5      7239634      36N
## 5 animal0 2005-07-14 10:35    384885.8      7240399      36N
## 6 animal0 2005-07-14 11:34    383347.9      7241826      36N
```

```
# Finally, convert to telemetry object:
animal0a <- as.telemetry(animal0_longlat)
```

```
## Minimum sampling interval of 59 minutes in animal0
```

```
animal0b <- as.telemetry(animal0_utm)
```

```
## Geocentric coordinates not found. Looking for UTM coordinates.

## Minimum sampling interval of 59 minutes in animal0
```

```
# if left unspecified, as.telemetry() will assume timezone = UTC
```

Both these files represent the same individual, with either *longitude/latitude*, or UTM coordinates (*easting*, *northing*, and *UTM zone*). The `as.telemetry()` function will immediately identify the columns if they are correctly named, then output the minimum sampling interval for each individual in the dataset. In this example, `animal0` has a minimum sampling interval of 59 minutes.

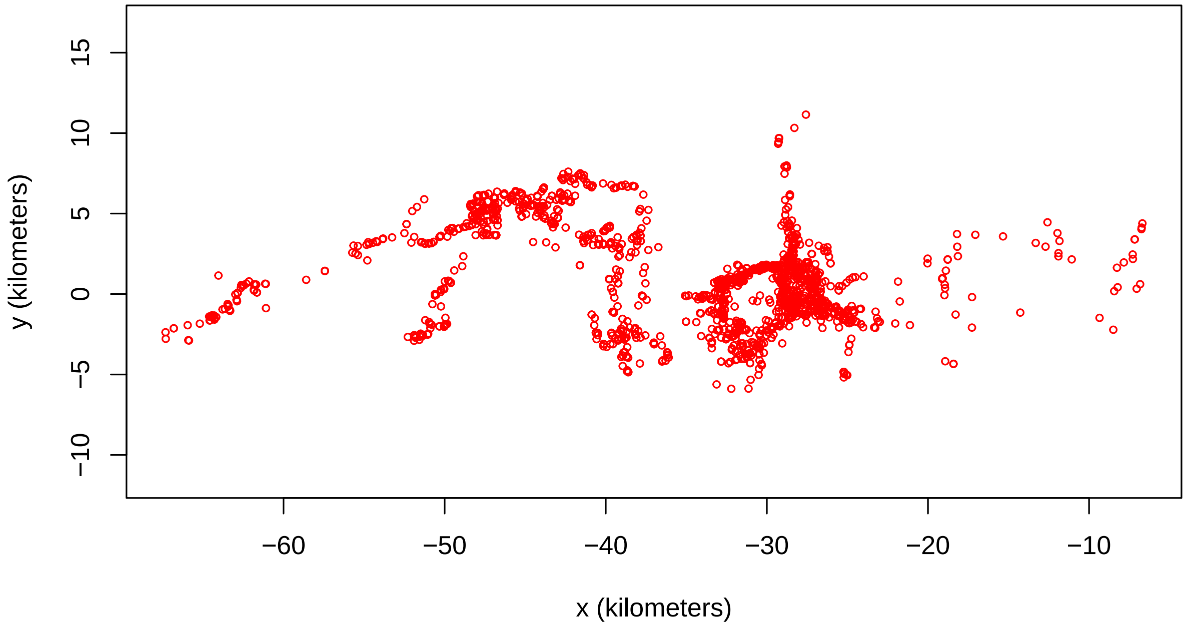
1.1. Buffalo tracking data

For this tutorial, we will use data already prepared into a list of `telemetry` objects. We can load it directly from the `ctmm` R package with the `data()` function:

```
data("buffalo")
animal1_buffalo <- buffalo[[4]] # selecting individual number 4
head(animal1_buffalo)
```

```
##                 timestamp longitude  latitude          t         x         y
## 17517 2006-04-25 05:09:00  31.73749 -24.19705 1145941740 -51803.35 -2715.663
## 17518 2006-04-25 06:09:00  31.73653 -24.19929 1145945340 -51569.29 -2845.660
## 17519 2006-04-25 07:09:00  31.73946 -24.20100 1145948940 -51340.72 -2576.353
## 17520 2006-04-25 08:09:00  31.73987 -24.20092 1145952540 -51344.11 -2533.788
## 17521 2006-04-25 10:09:00  31.74086 -24.20365 1145959740 -51029.45 -2474.771
## 17522 2006-04-25 11:09:00  31.74098 -24.20370 1145963340 -51022.23 -2463.655
```

```
# Plotting locations:
plot(animal1_buffalo)
```



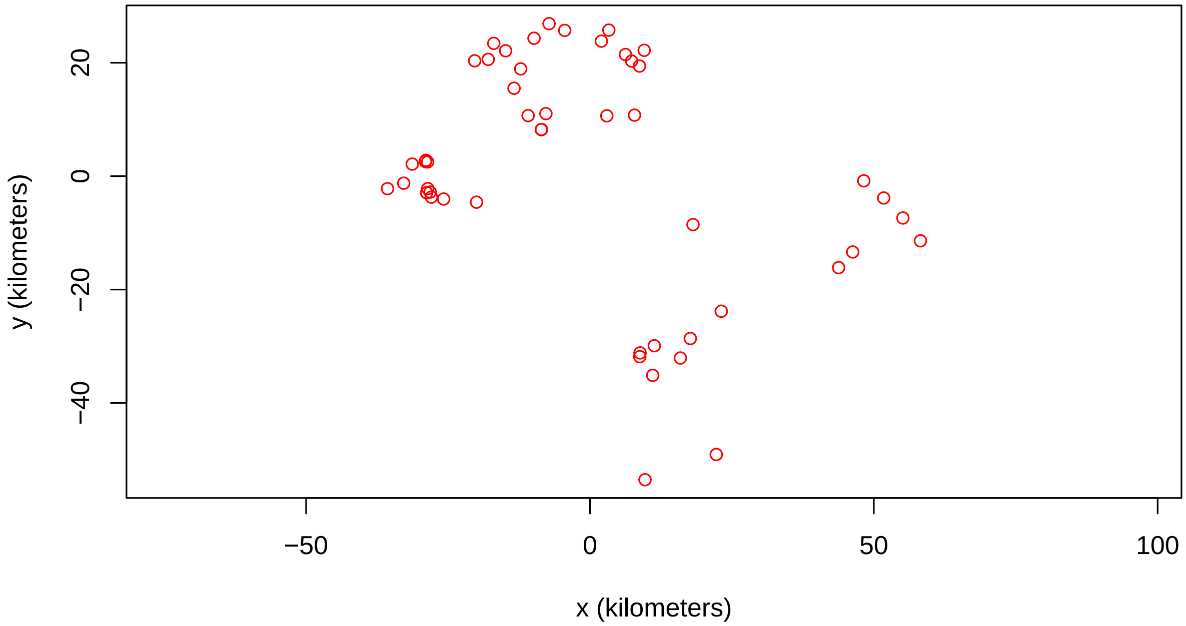
This dataset showcases an irregular sampling schedule: the buffalo nicknamed *"Pepper"* had a sampling rate shift from one fix every hour to one fix every two hours. We will use this dataset to highlight data irregularity and the **wAKDE** mitigation measure.

1.2. Gazelle tracking data

```
data("gazelle")
animal2_gazelle <- gazelle[[11]] # selecting individual number 11
head(animal2_gazelle)
```

```
##              x          y       t
## 2742  18152.70  -8539.799       0
## 2743  15931.16 -32069.788 1306800
## 2744  17678.84 -28632.329 1396800
## 2745  23135.50 -23820.789 1486800
## 2746 -20310.78  20348.792 2419200
## 2747 -17920.31  20598.668 2509200
```

```
# Plotting locations:
plot(animal2_gazelle)
```



Mongolian gazelles have a **home range crossing time** of a few months, and with a maximum longevity around 10 years, it is impossible to get a considerable **effective sample size** no matter the study duration (Fleming et al., 2019). We will use this dataset to highlight how to check **effective sample size** and apply the **parametric bootstrap** mitigation.

Data Analysis

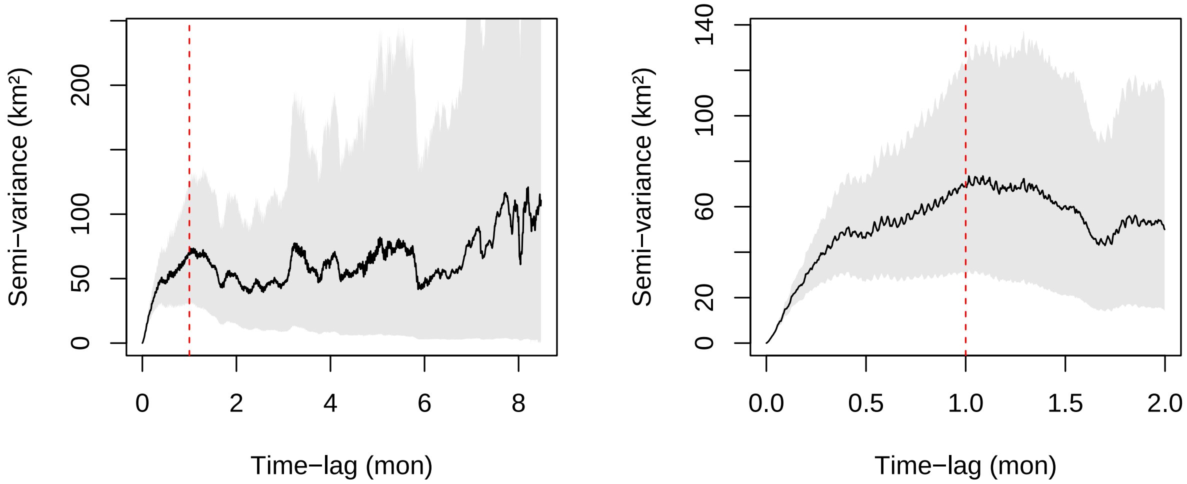
Step 2. – Checking for the *range residency* assumption

First, we want to check if our first tracking dataset (`animal1_buffalo`) can be used for home range estimation by checking for **range residency**. To achieve this, we calculate the **semi-variance function (SVF)**, and visualize it through the `variogram()` function.

Variograms are an unbiased way to visualize *autocorrelation* structure, representing the average square displacement (y-axis) over a specific time lag (x-axis). To facilitate interpretation, we have the **SVF** of `animal1_buffalo` zoomed out (right) to showcase all time lags and (left) zoomed in to showcase time lags up to two months:

```
level <- 0.95 # we want to display 95% confidence intervals
xlim <- c(0,2 %#% "month") # to create a window of 2 months

SVF <- variogram(animal1_buffalo)
par(mfrow = c(1,2))
plot(SVF, fraction = 1, level = level)
abline(v = 1, col = "red", lty = 2) # adding a line at 1 month
plot(SVF, xlim = xlim, level = level)
abline(v = 1, col = "red", lty = 2)
```



We can see that the variogram flattens (*i.e.*, reaches an asymptote) after approximately **1 month** (red line). This also indicates at how coarse the timeseries needs to be to assume independence (no autocorrelation), and corresponds to when traditional methods —such as **minimum convex polygons (MCPs)** and **Kernel Density Estimators (KDEs)**— could be applied without violating their assumptions.

Step 3. – Selecting the best-fit movement model through *model selection*

It is necessary to choose a home range estimator that accounts for the autocorrelated structure of the data, now that we see that it is **not** independently and identically distributed (non-IID). We need to test what movement model may explain the autocorrelated structure of our tracking data. We can run different movement processes with **maximum likelihood (ML)** or other parameter estimators, such as **perturbative Hybrid REML (pHREML)**. To facilitate further comparisons, we will run both ML and pHREML with the `ctmm.select` function.

```
# Calculate an automated model guesstimate:
GUESS1 <- ctmm.guess(animal1_buffalo, interactive = FALSE)
```

```r
# Automated model selection, starting from GUESS:
FIT1_ML <- ctmm.select(animal1_buffalo, GUESS1, method = 'ML')
FIT1_pHREML <- ctmm.select(animal1_buffalo, GUESS1, method = 'pHREML')
## reminder: it will default to pHREML if no method is specified.
```

```r
summary(FIT1_ML)
```

```
## $name
## [1] "OUF anisotropic"
##
## $DOF
##      mean      area     speed
##  11.24296  20.01326 747.72143
##
## $CI
##                                 low       est       high
## area (square kilometers) 431.446013 706.20245 1047.55828
## t[position] (days)         6.775273  12.18604   21.91786
## t[velocity] (minutes)     31.861346  35.81413   40.25730
## speed (kilometers/day)    15.948352  16.54124   17.13385
```

```r
summary(FIT1_pHREML)
```

```
## $name
## [1] "OUF anisotropic"
##
## $DOF
##      mean      area     speed
##  10.34454  15.65772 746.09401
##
## $CI
##                                 low       est       high
## area (square kilometers) 439.064822 773.50766 1201.09155
## t[position] (days)         6.664052  13.36206   26.79221
## t[velocity] (minutes)     31.807736  35.75689   40.19636
## speed (kilometers/day)    15.949646  16.54326   17.13658
```

Within these summaries, `$name` provides the selected best-fit model, `$DOF` provides information on the degrees of freedom (where `$DOF["area"]` corresponds to the **effective sample size** of the home-range area estimate), and `$CI` are the parameter outputs (area, position autocorrelation timescale, velocity autocorrelation timescale, and speed).

The typical pool of candidate models includes isotropic (when diffusion is the same in every direction; symmetrical) and anisotropic (when diffusion varies with direction; asymmetrical) variants. The automated model selection shows that *OUF anisotropic* (anisotropic Ornstein-Uhlenbeck foraging process) is our best-fit model. This movement process features a home range, correlated positions, and correlated velocities. To check the full model selection table, we can run the following command:

```
FIT1_pHREML_verbose <- ctmm.select(animal1_buffalo, GUESS1, verbose = TRUE)
summary(FIT1_pHREML_verbose)
```

	ΔAICc	ΔRMSPE (km)	DOF[area]
OUF anisotropic	0.0000	2.224045	15.657669
OU anisotropic	295.0898	2.634793	9.638283
OUF isotropic	330.2236	3.564231	11.204035
OUf anisotropic	1890.8878	0.000000	333.620712

By adding the argument `verbose = TRUE` we have access to the model selection table. By default, model selection is based on *Akaike's Information Criterion adjusted for small sample sizes* (AICc). The `ctmm` package also offers BIC, LOOCV, and HSCV. LOOCV seems to work slightly better for very small datasets, but we recommend AICc for the majority of datasets.

Step 4. – Feeding a movement model into the *home range estimator*

Now we can fit this movement process into the `akde()` function, and estimate the home range of `animal1_buffalo`. This function currently defaults to the **area-corrected AKDE**, or **AKDEc** (Fleming & Calabrese 2017):

```
# Run an area-corrected AKDE:
UD1_ML <- akde(animal1_buffalo, FIT1_ML)
UD1_pHREML <- akde(animal1_buffalo, FIT1_pHREML)
```

```
summary(UD1_pHREML)$CI # home range area estimation
```

```
##                                low      est     high
## area (square kilometers) 429.6573 756.9343 1175.357
```

We have calculated our home range for `animal1_buffalo`, resulting in an estimation of 757 km^2 (with 95% confidence intervals: 430–1,175 km^2).

Step 5. – Evaluating additional *biases*, applying *mitigation measures*

5.1. Buffalo tracking data

```
summary(UD1_pHREML)$DOF["area"] # effective sample size of animal1
```

```
##     area
## 15.65772
```

```
nrow(animal1_buffalo) # absolute sample size
```

```
## [1] 1725
```

Our output here also reveals more information regarding our dataset: the **effective sample size (N)** and the **absolute sample size (n)**. We can return this measure with the `summary` function: in our case, the N for `animal1_buffalo` is 15.7. Comparatively, our **absolute sample size** is easy to output, as it is the total number of observations within our dataset (n = 1,725).

As mentioned earlier, `animal1_buffalo` had a device malfunction that led GPS fixes to shift from one fix per hour, to one fix every two hours. As such, this individual is particularly suited for a **weighted AKDEc** (or **wAKDEc**), so we can re-run the function with weights set to `TRUE`:

```r
UD1w_pHREML <- akde(animal1_buffalo, FIT1_pHREML, weights = TRUE)
```

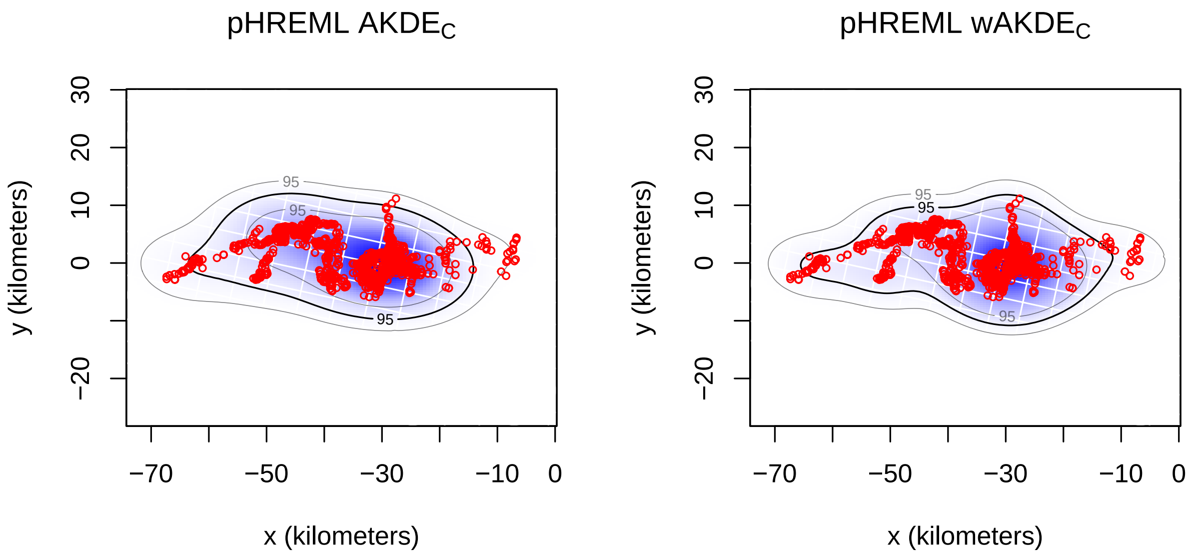
```r
summary(UD1w_pHREML)$CI # home range area estimation (weighted)
```

```
##                                 low      est     high
## area (square kilometers) 432.2234 761.4551 1182.376
```

Our new home range area estimation for `animal1_buffalo` is 761 km^2 (with 95% confidence intervals: 432–1,182 km^2). We can now plot our home range estimate for `animal1_buffalo`:

```r
# Creating an extent that includes both UDs at the 95% CI level:
EXT <- extent(list(UD1_ML, UD1_pHREML, UD1w_pHREML), level = 0.95)

# Plotting pHREML (with and without weights) side-by-side:
par(mfrow = c(1,2))
plot(animal1_buffalo, UD = UD1_pHREML, ext = EXT)
title(expression("pHREML AKDE"["C"]))
plot(animal1_buffalo, UD = UD1w_pHREML, ext = EXT)
title(expression("pHREML wAKDE"["C"]))
```



For `animal1_buffalo`, the difference between model parameter estimators is not substantial; we only have a ~5.7% AKDE area underestimation by ML compared to pHREML. However, the data fits the spatial locations much better.

```
( 1 - summary(UD1_ML)$CI[1,2] / summary(UD1w_pHREML)$CI[1,2] ) * 100
```

```
## [1] 5.742588
```

5.2. Gazelle tracking data

We can also check the difference with `animal2_gazelle`'s tracking data, where the small **effective sample size** issue is clearer:

```
GUESS2 <- ctmm.guess(animal2_gazelle, interactive = FALSE)

FIT2_ML <- ctmm.select(animal2_gazelle, GUESS2, method = 'ML')
FIT2_pHREML <- ctmm.select(animal2_gazelle, GUESS2, method = 'pHREML')

UD2_ML <- akde(animal2_gazelle, FIT2_ML)
UD2_pHREML <- akde(animal2_gazelle, FIT2_pHREML)
```

With `animal2_gazelle`, we have a more substantial area underestimation by ML compared to pHREML (~15.2%). We can also see that our *effective sample size* is only 4.5, with an *absolute sample size* of 49 ($N \ll n$).

```
( 1 - summary(UD2_ML)$CI[1,2] / summary(UD2_pHREML)$CI[1,2] ) * 100
```

```
## [1] 15.19245
```

```
summary(UD2_pHREML)$DOF["area"] # effective sample size
```

```
##     area
## 4.528926
```

```
nrow(animal2_gazelle) # absolute sample size
```

```
## [1] 49
```

At this point, we have selected a movement process, fed it into a home range area estimation with different model parameter estimators, and corrected for irregular sampling rates. With **small effective sample sizes**, it is important to see if **parametric bootstrapping** may be worth it to further reduce our estimation error. In order to do so, we can check the expected order of bias from pHREML:

```
# Expected order of pHREML bias:
1/summary(FIT2_pHREML)$DOF['area']^2
```

```
##       area
```

```
## 0.04875392
```

The bias is currently $\mathcal{O}(5\%)$ ("in the order of" 5%). As such, we will run parametric bootstrapping for `animal2_gazelle`. The relative error target is 1% by default (`argument error = 0.01`), but can be adjusted if necessary.

```
start_time <- Sys.time() # start recording running time
BOOT <- ctmm.boot(animal2_gazelle, FIT2_pHREML, trace = 2)
## note: this function incurs substantial computational cost, may take hours.
( total_time <- Sys.time() - start_time ) # output running time
```

```
summary(BOOT)
```

```
## $name
## [1] "OUF isotropic"
##
## $DOF
##      mean      area     speed
##  3.196957  3.530735 10.000358
##
## $CI
##                                  low          est         high
## area (square kilometers) 3429.704883 14089.021613 32137.455223
## t[position] (months)        0.000000     1.781242     3.758819
## t[velocity] (hours)         2.248230    10.481973    48.870349
## speed (kilometers/day)      5.590052     8.072361    10.551247
```

```
1/summary(BOOT)$DOF['area']^3 # expected order of bias
```

```
##       area
## 0.02271981
```

We can see that the expected order of bias was reduced to 2.3%, which is comparable to the numerical error target of 1%. To reduce the numerical error further, we would need to change the default relative error target of `ctmm.boot`, but the computational cost would continue to increase, and the comparably large statistical bias (2%) would remain.

Now we will calculate the **AKDEc** based on the estimated parameters, and plot the home range of `animal2_gazelle`. Because of small **effective sample size**, we set optimal weights to `TRUE` for improved statistical efficiency:

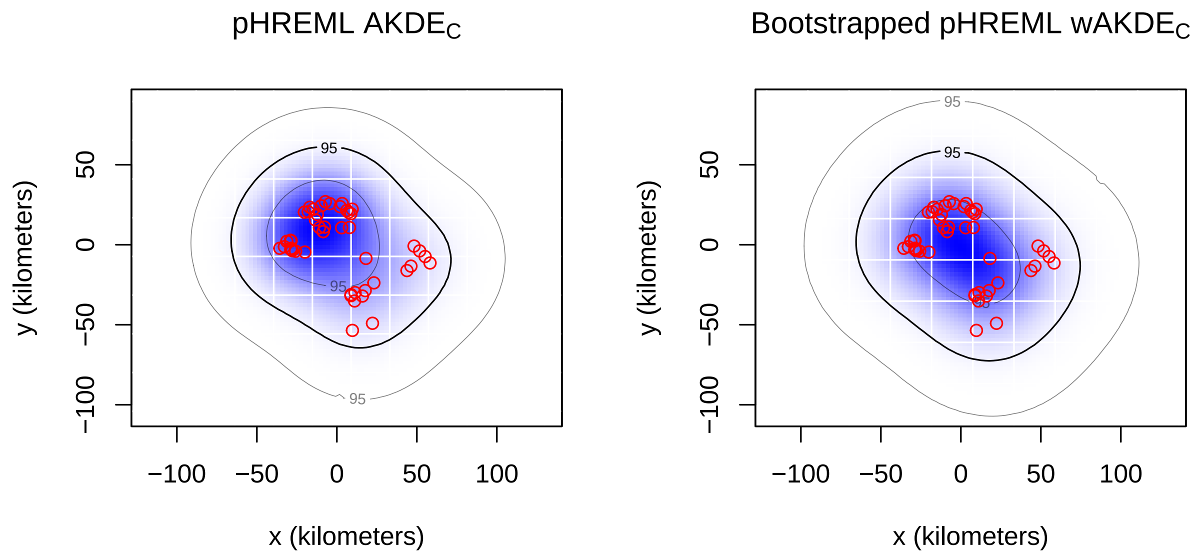
```
UD2_bpHREML <- akde(animal2_gazelle, BOOT, weights = TRUE)
summary(UD2_bpHREML)$CI
```

```
##                               low      est     high
## area (square kilometers) 3231.504 13274.82 30280.25
```

Finally, we have calculated our home range for `animal2_gazelle`, with an estimated area of 13,274 square kilometers (with 95% confidence intervals: 3,231–30,280 km^2). Our uncertainty with `animal2_gazelle` is substantially higher than with `animal1_buffalo`, as expected due to the small **effective sample size**.

```
# Creating an extent that includes both UDs at the 95% CI level:
EXT <- extent(list(UD2_pHREML, UD2_bpHREML), level = 0.95)

# Plotting pHREML and bootstrapped-pHREML side-by-side:
par(mfrow = c(1,2))
plot(animal2_gazelle, UD = UD2_pHREML, ext = EXT)
title(expression("pHREML AKDE"["C"]))
plot(animal2_gazelle, UD = UD2_bpHREML, ext = EXT)
title(expression("Bootstrapped pHREML wAKDE"["C"]))
```



The results presented here were generated with `R` version `4.1.1`, and `ctmm` version `0.6.0`.

Glossary

- *Home range*: the area repeatedly used throughout an animal's lifetime for all its normal behaviors and activities, excluding occasional exploratory moves.

- *Range residency*: the tendency of an animal to remain within their home range.

- *Home range crossing time*: the time required for an animal to cross the linear extent of its home range.

- *Absolute sample size* (n): the observations in a dataset.

– *Effective sample size* (N): number of range crossings that occurred during the observation period. Can be roughly estimated by dividing the duration of the tracking dataset by the average *home range crossing time* parameter.

References

Calabrese, J. M., Fleming, C. H., & Gurarie, E. (2016). Ctmm: An r package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution*, *7*(9), 1124–1132.

Cross, P. C., Heisey, D. M., Bowers, J. A., Hay, C. T., Wolhuter, J., Buss, P., Hofmeyr, M., Michel, A. L., Bengis, R. G., & Bird, T. L. F. (2009). Disease, predation and demography: Assessing the impacts of bovine tuberculosis on African buffalo by monitoring at individual and population levels. *Journal of Applied Ecology*, *46*(2), 467–475.

Fleming, C. H., Calabrese, J. M., Mueller, T., Olson, K. A., Leimgruber, P., & Fagan, W. F. (2014). From fine-scale foraging to home ranges: A semivariance approach to identifying movement modes across spatiotemporal scales. *The American Naturalist*, *183*(5), E154–E167.

Fleming, C. H., Noonan, M. J., Medici, E. P., & Calabrese, J. M. (2019). Overcoming the challenge of small effective sample sizes in home-range estimation. *Methods in Ecology and Evolution*, *10*(10), 1679–1689.