Simulating detection-censored movement records for home range analysis planning

Lyndsie S. Wszola⁎, Victoria L. Simonsen, Lucia Corral, Christopher J. Chizinski, Joseph J. Fontaine

Nebraska Cooperative Fish and Wildlife Research Unit, and School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA
School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA
U.S. Geological Survey—Nebraska Cooperative Fish and Wildlife Research Unit, and School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

ARTICLE INFO

Keywords:
Detection probability
Ecological simulation
Home range asymptote
Home range bias
Home range estimation

ABSTRACT

Home range estimation is an important analytical method in applied spatial ecology, yet best practices for addressing the effects of spatial variation in detection probability on home range estimates remain elusive. We introduce the R package “DiagnoseHR,” simulation tools for assessing how variation in detection probability arising from landscape, animal behavior, and methodological processes affects home range inference. We demonstrate the utility of simulation methods for home range analysis planning by comparing bias arising from three home range estimation methods under multiple detection scenarios. We simulated correlated random walks in three landscapes that varied in detection probability and constructed home ranges from locations filtered through a range of sampling protocols. Home range estimates were less biased by reduced detection probability when sampling effort was increased, but the effects of sampling day distribution were minimal. Like others, we found that kernel density estimates were the least affected by variation in detection probability, while minimum convex polygons were most affected. Our results illustrate the value of quantifying uncertainty in home range estimates and suggest that field biologists working in environments with low detection may wish to weight sample-size greater than concerns about temporal autocorrelation when designing sampling protocols.

1. Introduction

Patterns of animal spatial distribution convey important information about how animals perceive and interact with their environments (Bowman et al., 2002; Lima, 2002; Sih, 2005). Thus, documenting and explaining when and where animals are found has profound implications for understanding population and community dynamics and consequences for the conservation and management of species and their habitats (Schofield et al., 2010). Technological advances, such as global positioning system (GPS) tags that determine animal locations using satellites, have improved our ability to document animal locations and movements (Mills et al., 2006). However, studies using methods such as color bands, camera traps, passive integrated transponder tags, and even small GPS tags deployed in rugged environments, continue to be subject to bias when the ability to conclude that an animal is occupying a sampled space given its presence, i.e. detection probability, co-variates with landscape features that influence the animal’s spatial decisions (Rettie and McLoughlin, 1999; Camp et al., 2016; Vance et al., 2017). Furthermore, movement datasets collected using methods such as traditional radio telemetry continue to be used for longitudinal studies and conservation planning, and a need exists to quantify uncertainty when integrating such analyses with emerging methods (Stackhouse, 2012).

Because landscape-driven variation in detection probability can undermine inference and even create spurious results (Kessel et al., 2013), detection probability has received considerable attention in the spatial ecology literature (e.g., Aarts et al., 2008; Frair et al., 2004, Royle et al., 2014), and analytical methods that explicitly address detection issues are increasingly common (Golding et al., 2017; Kéry and Royle, 2015; MacKenzie and Royle, 2005). Despite a consensus on the importance of detection probability to spatial ecological inference, addressing the effects of detection probability in practice remains a challenge for investigators planning field protocols and choosing analytical methods.

Recent analytical improvements have provided methods to account for detection probability by imputing missing locations, weighting observations by detection probability prior to rendering a home range

⁎ Corresponding author at: 508 Hardin Hall, 3310 Holdrege St, Lincoln, NE 68583, USA.
E-mail address: lyndsie.wszola@huskers.unl.edu (L.S. Wszola).
https://doi.org/10.1016/j.ecolmodel.2018.10.017
Received 15 October 2018; Accepted 19 October 2018
Available online 24 December 2018
0304-3800/ © 2018 Elsevier B.V. All rights reserved.
estimate, or weighting observations to reduce temporal autocorrelation (Fleming et al., 2018; Horne et al., 2007). Too often, however, detection probability is unknown and the effects of imputation and weighting likely vary according to research questions, sampling protocols, and analytical methods (Fieberg and Börger, 2012). It is difficult to quantify sampling limitations after data have been collected, especially in long-term telemetry datasets commonly used by natural resource agencies and other researchers seeking a practical answer to an applied conservation problem. Moreover, a need exists for tools that facilitate robust field sampling protocols and informed decisions about which home range estimation method to use for a given question in a given system. Investigators must be able to identify if a dataset contains sampling bias or variation in detection probability prior to applying any correction method, and select the analysis method that most effectively marginalizes the effects of varying detection in a given research scenario. Herein, we introduce the R (R statistical environment; R Core Team, 2018) package DiagnoseHR. DiagnoseHR implements simulation methods for creating detection-censored animal occurrence records reflecting a wide range of landscape-driven detection probability distributions, animal movement patterns, and sampling protocols.

We illustrate the utility of simulating detection-censored datasets for comparing spatial analysis methods using a case study of home range size estimation. Ecologists estimate animal home ranges, the space used while foraging, mating, and caring for young within a defined time period (Burt, 1943; Spencer, 2012), by inputting a set of relocations from a marked individual into an algorithm that uses the spatial and sometimes temporal relationships between locations to elucidate the animal’s use of space (Worton, 1987). Home ranges are assumed to be an asymptotic function of relocation sample size when an animal demonstrates some degree of site fidelity (Powell and Mitchell, 2012). The outcome of home range estimation varies by modeling approach, and is sensitive to sampling issues both within (e.g., sampling protocols, fix schedules) and beyond (e.g., landscape features) investigator control (Fig. 1; Gula and Theueurkauf, 2013; Harris et al., 1990). The effects of limited and biased relocation samples are an issue of considerable concern in the home range literature (Boulanger and White, 1990; Laver and Kelly, 2008; Lyons et al., 2013; Signer and Balkenhol, 2015; Wilson et al., 2018). Although telemetry sampling protocols, GPS fix schedules, and post-collection data processing methods traditionally optimize sample size and independence (Otis and White, 1999; Seaman et al., 1999), heterogeneity in detection probability across a landscape is often not explicitly addressed.

Using DiagnoseHR, we simulate an attempt to account for the role of varying detection probability in a home range analysis of radio telemetry data. We simulate the telemetry data collection process and assess how detection probability, sampling design and statistical methods affect home range inference. We focus on home range size, the total area used by an animal. Home range size, one of the most commonly calculated and compared home range metrics, is often applied in conservation planning, incurring potentially important consequences when estimators are biased. We simulated animal movement patterns under multiple detection scenarios and compare the performance of three established home range estimators under varying field protocol and landscape constraints.

2. Materials and methods

2.1. Calculation: simulation approach

The problem of attempting to record locations of marked animals under varying detection regimes is widespread across ecological field research (Millspaugh and Marzluff, 2001). Variation in home range analysis methods and reporting complicate the interpretation of differences in home range sizes between systems (Laver and Kelly, 2008), a common objective of home range studies (e.g., Anderson et al., 2005). Recent advancements in home range analysis have consequently focused on making methods more transparent and results more reproducible (Signer and Balkenhol, 2015). We seek to further the development of transparent and reproducible approaches by providing a tool for simulating animal movement datasets and quantifying home range estimate uncertainty.

DiagnoseHR is currently available as a package hosted on github and may be installed using the R package devtools (Wickham et al., 2018). DiagnoseHR facilitates the creation of simulated animal movement datasets for field planning and home range method selection by giving the user control over the landscape, animal behavior, and sampling processes. The simulation proceeds in four stages: landscape formation, population initiation, animal movement, and sampling. The resulting simulated datasets are then available for field planning and methodological comparison. DiagnoseHR currently supports home range estimate evaluation for a limited number of home range methods, with support for more being continually added. Details on DiagnoseHR are given below, and a full tutorial that will be updated as the package grows is available at https://github.com/lsw5077/DiagnoseHR.

2.1.1. Landscape formation and population initiation

The user initiates the simulation by constructing a landscape matrix wherein each cell is assigned a cell-specific detection probability that constrains an observer’s ability to record the presence of a simulated organism given its occurrence in the cell. The function make_world() creates a grid with x and y dimensions of the user’s specification. Each grid cell is assigned a detection probability drawn from either a random uniform distribution with user-defined maximum and minimum, or a beta distribution with user-defined shape parameters. The world is stored as a data frame, an R data storage object that can hold multiple data types, for use in the next steps. The user initiates the simulated population through the populate_world() function by specifying the number of organisms in the population, the world to populate (the object created and stored using make_world()), the maximum number of organisms that may share a cell, and whether initial cells should be assigned to simulated organisms with or without replacement. If a user has a landscape with X and Y coordinates and known detection probability, they may upload it into the simulation as a four-column data frame with columns “x”, “y”, a unique “cell_id,” and “cell_prob,” the cell-specific detection probability.

2.1.2. Animal movement simulation

Users initiate and record correlated random walks defined by the organism’s home range size, a total number of steps, step directions, distance from home range center, and site fidelity using the function move_critters(). The move_critters() function initiates and records correlated random walks for each simulated organism around the simulation space. The correlated random walk is defined by the world in which the walk occurs, the specified population created by populate_world(), whether the world is to be open or closed (i.e. whether animals can leave the simulation space), the home range size in square units, and direction parameters for the correlated random walk. If the user specifies home range type as “fixed,” (i.e. the same for all individuals in the population), then each individual is assigned the specified home range size. If the user specifies home range type as “random,” each individual draws a home range size from a gamma distribution with shape equal to the specified home range size and scale of 1 to simulate home range size variation that may be present in a population. Organisms make the number of movements specified in the “Nsteps” argument by selecting a cell from the set of available cells given their home range sizes. If the world is open, simulated organisms may move off the simulation space and become undetectable. If the world is closed, simulated organisms will only make movements to cells with coordinates greater than or equal to the minimum X and Y coordinates of the simulation space and smaller than or equal to the maximum X and Y coordinates of the simulation space. Users may also specify an optional site fidelity between 0 and 1 that controls how likely
organisms are to turn back toward their home range centers as they move farther away. Organisms with a high site fidelity value will be more likely to turn back with each additional distance unit away from their home range center than organisms with a low site fidelity value.

The direction of each step in the correlated random walk is determined in part by a Wrapped Cauchy distribution and step length is determined using a Weibull distribution, as recommended by Morales et al. (2004) for simulating GPS track movement data. At each step, the organism determines a step direction and length depending on the size of the home range and its position relative to the home range center.

Fig. 1. Conceptual diagram of the home range inference process. Inference of animal home range size is a function of animal (a, c), and investigator (b, d) decisions. Sampling design (b) determines which locations (a) are available to the investigator, and the animal’s choice of habitat features (c) may determine the detectability of each location. Choice of home range estimator (d) further influences inferences drawn about home range size.
The organism first draws a potential step direction from a Wrapped Cauchy distribution with user-specified \( \mu \), the mean direction taking any value on the unit circle in radians (0 to 2\( \pi \)), and \( r \), the dispersion parameter varying between 0 and 1. For example, a Wrapped Cauchy distribution with any \( \mu \) and a \( r \) of 0 would produce completely random trajectories, and a Wrapped Cauchy distribution with a \( \mu \) of 3.14 and a \( r \) of 1 would produce only trajectories of 3.14 radians. From a biological perspective, the wrapped Cauchy distribution allows users to describe a central tendency and variation in random walk behavior, providing a wide variety of underlying animal movement patterns. Step length is determined by drawing a random number from a Weibull distribution that with a shape that defaults to 2 and scale by default equal to the radius of the user-specified home range. However, the user may specify any shape and scale parameters for the Weibull distribution using the arguments “\(\text{wei_shape}\)” and “\(\text{wei_scale}\).”

To ensure that users may control the degree to which the simulated organisms move around a central location reflecting a home range, we developed a functional relationship between the organism’s distance from its home range center and the probability the organism’s next step would be directed toward its home range center \( (P_c; \text{equation } 1) \). This relationship is given as:

\[
P_c = \frac{\exp(sf + \frac{d}{r})}{1 + \exp(sf + \frac{d}{r})}
\]

Equation 1: DiagnoseHR calculates organisms’ probability of turning back toward their home range center as a function of site fidelity and distance from home range center.

Where \( sf \) is the organism’s site fidelity, a number between 0 (no site fidelity) and 1 (high site fidelity), \( d \) is organism’s distance from home range center if it makes the Weibull-determined step length in the Wrapped Cauchy-determined potential direction, and \( r \) is the radius of the user-specified home range size. Assuming a finite home range size, the probability of turning toward the home range center increases with increasing distance from home range center, and site fidelity defaults to 1. If \( P_c \) is greater than a random number drawn from a random uniform distribution between 0 and 1, the organism will turn back toward its home range center. If \( P_c \) is not greater than the random uniform draw, the organism will continue in its Wrapped Cauchy-determined direction. Because the tools provided are open source, users may readily alter the \( P_c \) curve equation as needed.

### 2.1.3. Sampling

DiagnoseHR allows the user to sample the simulated movement record using multiple sampling schemes reflecting realistic field planning constraints in situations where relocations are limited by observer time (e.g., color bands, radio telemetry) or limited automated relocation ability (e.g., for very small GPS tags with limited battery life or memory). The function \( \text{sample_world()} \) allows the user to sample the correlated random walks created using the \( \text{move\_critters()} \) function within the detection constraints of the user’s simulated world and defined sampling protocols. The process of detecting an animal given its presence includes stochasticity that we endeavor to reproduce by including stochastic variation in the landscape and the simulated organisms’ behavior.

DiagnoseHR provides tools for assessing the stochastic effects of individual organism behavior on detectability by drawing an individual-specific detection probability for each organism at each step from a random uniform distribution between zero and one. When the observer attempts to sample an occupied cell, they are successful when the cell-specific detection probability exceeds the organism-specific detection probability. The \( \text{sample\_world()} \) function samples a user-defined detection world, “\(\text{world}\)” movement record, “\(\text{walk}\)” number of cells per time step, “\(\text{n.cells}\)” a number of time steps to sample, “\(\text{sample\_steps}\)” whether to sample with replacement within a time step, “\(\text{replace\_step}\)” and whether to sample with replacement in the entire sampling simulation, “\(\text{replace\_world}\)”.

A simulated relocation is designated as detected when the user’s sampling protocol samples an occupied cell whose cell detection probability exceeds the organism’s detection probability and not detected when the cell is not sampled, or the organism’s detection probability exceeds the cell’s detection probability. \( \text{sample\_world()} \) generates a list containing all cells sampled, all animal movements and the detection/sampling-censored movement record. By comparing the results of home ranges calculated with all locations the animal occupied, and only those detected under sampling constraints, the investigator may assess how different sampling schemes and estimators are likely to affect inferences made for a specific question or application in light of detection and sampling constraints.

### 2.2. Other functions: estimator performance and uncertainty

We propose that simulated datasets may be a useful planning tool and that investigators conducting home range analyses may consider three supplementary methods to quantify uncertainty in home range estimates: asymptote assessment, sensitivity plots, and leverage plots. The concept of the home range as an asymptotic function of relocation sample size has an intuitive ecological significance, but it is not immediately apparent when a given home range estimate reaches an asymptote. Indeed, a home range estimate that does not reach an asymptote may still be ecologically relevant, but the interpretation will be different. For example, a non-asymptotic home range may provide information about space use, but not give a minimum viable space requirement for conservation. We follow previous suggestions (e.g., Laver and Kelly, 2008) that home range analysis should include an assessment of whether a home range estimate is asymptotic. Additionally, variation in detection probability may result in a dataset where low-detection spaces are under-represented and a relatively few relocations can have an outsized impact on home range size estimates. Quantifying the importance of each relocation to a home range estimate therefore has value for identifying potential bias and assessing the outcomes of home range analysis. DiagnoseHR provides utility for creating asymptote plots, sensitivity plots, and leverage plots that facilitate assessment of the importance of sample size and individual relocations on the reliability of estimates. The functions described below are currently compatible with the minimum convex polygon, local convex hull, and kernel utilization density methods from package adehabitatHR (Calenge and Fortmann-Roe, 2017), and will be continually expanded to include new methods. Because DiagnoseHR is an open source R package hosted on github, users may also modify package source code to experiment with other home range methods and contribute their expansions.

The function \( \text{hrAsym()} \) helps users determine whether a home range estimate reaches an asymptote as the number of relocations collected increases. \( \text{hrAsym()} \) recalculate each home range estimate iteratively, adding relocations to the initial minimum subsample as specified by the package adehabitatHR (five for MCP, 10 for LCH and KDE). \( \text{hrAsym()} \) then plots the resulting home range size curve to provide a visual assessment of when or whether home range size converges on an asymptote.

Sensitivity plots facilitate discussion of uncertainty in estimated home range size by employing a “jackknife” method to demonstrate how a home range estimate changes when individual relocations are iteratively removed. By iteratively recalculating home ranges, leaving out one relocation at a time and plotting relocation ID on the x-axis and home range size on the y-axis, users may assess how sensitive the estimate is to the addition or removal of a specific point. A home range estimate that is sensitive to the addition or removal of relocation will result in a sensitivity plot with larger and more frequent fluctuations in estimated home range size, while a less sensitive estimate will appear “smoother.” We emphasize that the definition of a “large” fluctuation depends entirely on the ecology of the study species and the intended
application of the analysis. The function hrDiag() in package DiagnoseHR creates an sensitivity plot for each individual in a relocation dataset and returns a list, “out,” into the global environment. Out contains a data frame of re-estimated home ranges, “results,” and a summary table, “result.tab.” result.tab displays the minimum, maximum, mean, standard error, 25th and 50th percentiles of each individual’s re-estimated home range size. A more sensitive home range will exhibit a relatively larger range between maximum and minimum estimated home range size.

DiagnoseHR further quantifies error in home range size estimation resulting from detection or sampling processes by calculating bias as:

$$\text{Bias} = \frac{\text{Observed Home Range Size} - \text{Expected Home Range Size}}{\text{Expected Home Range Size}}$$

Equation 2: DiagnoseHR calculates bias as the proportional difference between observed and expected home ranges, relative to expected home range.

Where the “expected” home range is the home range calculated using all possible locations and the selected home range estimator, and the “observed” home range is the home range calculated in the same estimator using the locations collected in a given detection scenario and sampling scheme. For example, a bias of 0.50 represents a 50% over-estimate of home range size and a bias of -0.50 represents a 50% under-estimate of home range size. Because it is difficult to compare the sensitivity of differently-sized home range estimates, we suggest assessing bias as a scaled value, or leverage, calculated as:

$$L_i = \frac{b_i}{\max(b)}$$

Equation 3: DiagnoseHR calculates leverage of each relocation as the bias resulting from removing that relocation divided by the maximum bias created by removing any relocation in the relocation set.

Where $L_i$ is the leverage of the $i$th point, $b_i$ is the bias of the home range measurement when point $i$ is removed, and $\max(b)$ is the maximum bias of any location in the relocation dataset. When individual relocation leverages are displayed in a histogram, the resulting plot demonstrates the scaled relative importance of each relocation to the home range estimate. A home range in which many points contribute similarly to the home range size estimate will have a leverage histogram with many observations clustered around 0, but if a few points are having a larger effect, the observations will be distributed away from zero toward the extreme ends of the leverage histogram. Whereas the sensitivity plot will vary according to the magnitude of the home range, leverage is unaffected by the overall size of a home range. Users may therefore compare the sensitivity of variously sized home range estimates by using the sensitivity plot in conjunction with the leverage plot.$hrDiag()$ calculates leverage for each individual relocation in a relocation dataset and returns one leverage histogram for each individual in the dataset.

2.3. Case study

We simulated an observer making home range inferences from traditional radio telemetry data collected under three detection regimes reflecting variation in detection across populations and environments (Fig. 2). We initiated one simulated organism per iteration, on simulation spaces containing 1.44 million cells, with detection probabilities determined by detection scenario. Cells in the uniform “high-detection” scenario were assigned detection probabilities drawn from a random uniform distribution from 0.7 to 0.9. Cells in the “random” scenario received a detection probability drawn from a random uniform ranging from 0 to 1.0, and cells in the uniform “low-detection” scenario received a detection probability drawn from a random uniform of 0.1 to 0.3. The probabilities of each cell were redrawn at the beginning of each iteration, but remained constant during each movement trial.

We simulated 1000 correlated random walks around the simulated landscape under each detection scenario, running four iterations in parallel. Each iteration had an independently drawn random seed (L’Ecuyer-CMRG) using the foreach package (Microsoft & West S., 2017). At the start of each iteration, an organism was generated at a random home range center within the landscape matrix and given a home range with radius equal to 492 cells. The animal then made daily movements between cells within its home range over a 120-day study period, creating 120 opportunities for it to be observed. The direction of the daily movements was determined using a Wrapped Cauchy distribution with $\rho = 0$ and $\mu = 0$, and $PC$ was calculated using a modified version of the more general function given above to create an exponentially increasing probability of turning toward the home range center as the animal’s distance from home range center increased:

$$PC = \frac{0.999}{1 + e^{(x - 0.1)(-0.40(\rho/\mu))}}$$

Equation 4: The function relating simulated organism distance from home range center and probability of turning around in the case study.

For the case study, we divided the 120-day study period into 10, 12-day sampling periods. Within each sampling period, days were designated as either a “sampling day,” when the collection of daily movements was attempted, or an “off day,” when the collection of daily movements was not attempted. To examine the effects of investigator sampling effort, we used six different sampling frequencies: one, two, three, four, six, or twelve days out of each 12-day sample period. To address the effects of temporal autocorrelation, an issue of some concern in the spatial ecology literature, we applied two sampling schemes to each of our six sampling frequencies. The first sampling scheme used was an evenly spaced schedule (hereafter “sequential”), wherein observations were attempted every one, two, three, four, six, or twelve days to create an even sampling schedule across each of the sampling periods. We also employed a grouped sampling scheme (hereafter “clustered”), wherein all observations (one, two, three, four, six, or 12) were attempted on consecutive days starting with the first day of the sampling period. For example, in a four-out-of-12-days sequential sampling schedule, days one, four, seven, and 10 would be sampled out of every 12-day sampling period up to 120 days, whereas in a four-out-of-12-days clustered sampling schedule, days one, two, three, and four out of every 12-day sampling period would be sampled up to 120 days. We thus created iterations with identical sample sizes but differing potential for temporal autocorrelation.

We used the simulated relocation datasets collected under each of the detection scenarios to estimate home range size under each of the 12 sampling schemes using three home range methods: Minimum Convex Polygon (MCP), Kernel Density Estimation (KDE), and Local Convex Hulls (LCH) in the R Package adehabitatHR (Calenge and Fortmann-Roe, 2017). MCP draws the smallest convex polygon possible by connecting exterior locations around an animal’s recorded relocations (Burt, 1943; Hayne, 1947). MCP represents the simplest method of constructing home ranges and is prone to both underestimating and over-estimating home range size (Börger et al., 2006). Calenge and Fortmann-Roe (2017) suggest removing a small number of outliers in MCP analyses to mitigate the over-estimation of home range size, so we constructed MCP estimates using 90% of the relocations in each iteration.

KDE estimates the distribution of individual space use across a home range by pairing a bivariate kernel function to each relocation (Worton, 1987, 1989, 1995). The kernel functions are averaged with the kernel functions of the neighboring relocations to create a utilization distribution (Van Winkle, 1975). Isopleths, polygons that contain a given percentage of relocations, within the range of 50–90% are less biased by sample size (Börger et al., 2006), so we used a 90% isopleth to determine the area of individuals’ home ranges for comparison with the results of the other two methods. We used the default ad-hoc smoothing method, which assumes that the kernel is bivariate normal (Calenge and Fortmann-Roe, 2017).
L.S. Wszola et al.

Ecological Modelling 392 (2019) 268–278

Fig. 2. Case study detection probability scenarios. Detection probability in the high-detection scenario was drawn at each iteration from a random uniform distribution of 0.7–0.9. Detection probability in the random scenario was drawn from a random uniform distribution of 0–1, and detection probability in the low-detection scenario was drawn at each iteration from a random uniform distribution of 0.1–0.3. The example matrices pictured are 25 × 25 sections of the 1200 × 1200 simulation spaces.

LCH examines space use by connecting either a specific number of locations or locations within a set distance of the root location to form hulls, minimum convex polygons that comprise parts of a home range, then merging overlapping or adjacent hulls together to represent a home range (Getz and Wilmers, 2004). We used an adaptive LCH method where the sum of the distances between locations within the hull and the root location had to be less than the specified distance α (the greatest distance between two locations). The adaptive LCH allowed the number of locations within a hull to fluctuate based upon the density of locations, creating smaller hulls in areas of high use and larger hulls in areas of low use (Getz et al., 2007). We recorded the rate at which home ranges were successfully estimated under each simulation permutation (i.e., the rate at which the home range estimation function produced a home range estimate without encountering a fatal error).

We assessed the effects of sampling schedule, detection scenario, and home range estimator by fitting one generalized linear model with a Gaussian error distribution in package lme4 in the R statistical computing environment (R 3.5.1; Bates et al., 2017; Bolker et al., 2009). Because we were interested in assessing the relative contributions of home range estimator choice, detection scenario, and sampling schedule to home range bias regardless of sign, we centered and scaled on 0 the absolute value of home range bias measurements. The model related scaled absolute bias of home range estimates to home range estimator (MCP, KDE, LCH) and sampling scheme (i.e., clustered or sequential) as well as an interaction between detection scenario and scaled number of days sampled. We assessed model fit using pseudo-R² (Zhang, 2018).

3. Results

3.1. Diagnostic plots

The sensitivity plot produced by the hrDiag() function for sample data from the low detection scenario was more jagged than the one produced using sample data from the high detection scenario (Fig. 3a-b). In the leverage plots, values were largely clustered around 0 when detection was high and distributed from 0 when detection was low (Fig. 3b-c). The asymptote assessment generated by hrAsym() exhibited an asymptote under high detection conditions (Fig. 3d) but not under low detection conditions (Fig. 5e).

3.2. Case study

We ran 108,000 simulation iterations that resulted in 93,633 successful home range calculations. As expected, the number of relocations successfully detected was positively related to sampling effort and scenario detection probability. For a given sampling effort, the high detection scenario yielded the greatest number of relocations, followed by the random scenario, and the low detection scenario yielded the fewest relocations. For example, under the minimum sampling effort (i.e., 10 sampling days) the mean (± SE) number of observations collected was 1.95 (± 0.02) for the low detection scenario, 4.95 (± 0.02) for the random scenario, and 7.98 (± 0.02) for the high detection scenarios. Similarly, the mean number of observations collected under the maximum sampling intensity (i.e., 120 sampling days) was 23.86 (± 0.06), 59.97 (± 0.07), and 96.15 (± 0.06), for low, random and high detection scenarios respectively. Detection scenario also affected the rate at which home ranges were successfully calculated. The high-detection scenario achieved 100% success at 20 sampling days, the random scenario achieved 100% success at 40 days, and the low detection scenario did not achieve 100% success until all 120 days were sampled. Sampling scheme (i.e., clustered or sequential sampling) did not affect the number of observations collected, or the proportion of home ranges that were successfully estimated. The KDE estimator generated the lowest mean absolute bias (24.62% ± 0.14%; Fig. 4a-b), followed by the LCH estimator (39.75% ± 0.13%; Fig. 4c-d) and the MCP estimator (50.30% ± 0.14%, Fig. 4e-f). When we accounted for positive and negative bias, the distinction was more pronounced, with KDE displaying a mean bias of 7.50% ± 0.20%, whereas LCH (−37.70% ± 0.15%) and MCP (50.00% ± 0.14%) systematically underestimated home range size under every detection scenario.

The model examining causes of variation in bias under varying sampling designs and detection constraints explained 47.50% of the variance in scaled absolute bias. Bias was negatively associated with increasing sampling effort and detection probability (Table 1). Bias was greater for LCH and MCP estimates than for KDE estimates, and while the effect of sampling scheme was statistically significant, sequential sampling did not appreciably change predicted bias (Fig. 5). The interaction between detection scenario and sampling days was significant, with bias decreasing slightly more for each additional sampling day in the high and random detection scenarios than in the low detection scenario.

4. Discussion

Home range estimation and other spatial ecology analytical methods are continually improving through technological and modeling advances (e.g., Fieberg and Börger, 2012; Horne et al., 2007; Laver and Kelly, 2008; Signer and Balkenhol, 2015). Nevertheless, detection effects persist for methods limited by observer time and expertise, like radio telemetry and color band relocation (e.g., Silva et al., 2018), for cryptic species like many herpetofauna (Corović et al., 2018), and for GPS studies in which bias co-varies with animal habitat preferences (Ironside et al., 2017). The pervasive yet unpredictable effects of variation in detection probability necessitate tools for planning field protocols and choosing modeling methods to fit a study’s detection and methodological conditions. DiagnoseHR facilitates the creation of
simulated datasets reflecting a wide range of detection conditions, animal movement patterns, sampling schedules, and home range estimation methods to facilitate efficient field protocols and effective analyses.

We simulated movement data and the home range estimation process to assess how sampling effort interacted with detection probability to influence home range estimate bias for three home range estimators and two sampling schemes under three detection scenarios. Home ranges estimated for the same simulated individual under multiple detection scenarios and methods produced drastically different results even when sampling effort was similar (Figs. 4–5). Home range estimate bias was negatively associated with increasing sampling effort under all detection scenarios, but decreased slightly more for each unit
increase in sampling effort under the random and high detection scenarios when compared to the low detection scenario (Table 1, Fig. 5). The case study results demonstrate that landscape-driven variation in detection probability is likely to confound comparisons of space use among populations of wide-ranging species, assessments of variation among individuals within the same population, and may even create biased estimates of space use within individual home ranges. Our findings are congruent with the expansive body of home range literature suggesting that additional relocations increase the precision of home range estimates (e.g., Seaman et al., 1999; Swihart and Slade, 1985; Horne et al., 2007; Fieberg, 2007; reviewed in Frair et al., 2010). DiagnoseHR additionally demonstrates how landscape variation in detection probability interacts with variation in sampling effort to create variation in home range estimates between systems. Our results emphasize the need for investigators to report the rate of location success and failure, for home range analysis and potentially for other analyses.
such as spatial capture-recapture models of dispersal distances (Ergon et al., 2014; Schaub et al., 2014).

Whereas the interaction of detection probability and sampling effort had a strong effect on the accuracy of home range estimates, the effects of sampling scheme (i.e. whether samples were clustered or sequential), were minimal (Table 1). The greater relative importance of detection and sampling effort compared to the importance of sampling scheme suggests that the tradeoff between sample independence and sample size in low-detection field scenarios may be best addressed by weighting sample size higher than independence when drafting field protocols, as has been suggested by others (e.g., Otis and White, 1999; Fieberg, 2007). The case study was designed to reflect traditional radio telemetry studies where collecting relocations is often labor-intensive and field conditions or the temporal bounds of the life stage of interest limit the duration of studies. The case study therefore simulated tradeoffs between sample independence and sample size, an issue with application beyond radio telemetry studies. Indeed, DiagnoseHR may facilitate planning for spatial ecology methods with different inherent tradeoffs. Notably, studies with small GPS units must make tradeoffs between battery life, sample size, and temporal autocorrelation. The simulation tools provided by diagnoseHR may readily facilitate planning for such scenarios by producing detection scenarios corresponding to real detection landscapes, and sampling schemes calibrated to reflect different GPS battery life-sample size-autocorrelation tradeoffs.

In addition to providing tools for understanding the effects of variation in sampling design and detection, DiagnoseHR facilitated comparison of home range estimators in the context of varying detection scenarios. KDE exhibited the smallest mean bias and smallest mean absolute bias under all detection scenarios, while MCP and LCH estimators systematically underestimated home range size (Fig. 4). The finding that MCP underestimated home range size is perhaps counter-intuitive given the potential for MCP to enclose areas not used by an animal, but aligns with previous studies (e.g., Girard et al., 2002) suggesting that MCP tends to underestimate home range size when sampling effort is high. Our finding that KDE is less subject to detection-imposed bias than other methods also corresponds with previous work (e.g., Börger et al., 2006; Seaman et al., 1999; Seaman and Powell, 1996). KDE may represent a conservative option when the purpose of home range estimation is to provide ecological inference and guidance for conservation actions. Systematic underestimation of home range size may lead to skewed density estimates and misinformed conservation planning. Given sufficient funding, conserving too large a space is unlikely to cause an undesirable conservation outcome, whereas conserving too little space based on a systematic underestimate of home range size could render conservation actions ineffective.

Even if conservation planning is not the objective of a study, exploring the behavior of home range estimates can add depth to discussions of home ranges’ ecological significance and provide insights on when to apply detection-mitigating analytical techniques (e.g., Frair et al., 2004, 2010; Horne et al., 2007; Katajisto and Moilanen, 2006). Using the function hrDiag(), users may examine the sensitivity of home range estimates and compare home range estimates using medians and quantiles. If sensitivity and leverage plots reveal that few points have an exaggerated impact on the home range estimate, the investigator may subject those points to increased scrutiny to assess whether they arise from equipment or investigator error, or whether they perhaps result from detection bias. For example, if the removal of a few relocations drastically alters a home range estimate and the relocations occur in the same vegetation community, or are isolated by a particular landscape feature, those relocations may warrant further scrutiny, as their effects could indicate a detection issue. Users may understand or mitigate estimate sensitivity by incorporating models of detection probability into their analyses and applying methods such as Brownian Bridge kernel utilization density that impute an animal’s likely path between observations (Horne et al., 2007). The classical definition of a home range assumes that home range size approaches an asymptote as the number of relocations increases (Burt, 1943). Recent work has questioned whether a home range must be asymptotic to be useful for ecological inference or conservation planning (Powell and Mitchell, 2012). Indeed, the question of how long an individual’s home range takes to reach an asymptote, or whether it asymptotes at all can provide information about how individuals with apparently similar home range sizes may differ in space use (Laver and Kelly, 2008). Using the hrAsym() function, it is possible to assess whether an estimate is indeed asymptotic by iteratively adding relocations. The resulting diagnostic plots can facilitate effective comparison between individual and population estimates of home ranges.

We used simulation tools from the R package DiagnoseHR to examine the interacting effects of detection probability and sampling effort on home range estimation using varying sampling protocols and estimators. Our findings are largely congruent with previous studies of detection, home range estimation method, and sampling intensity in isolation, demonstrating that simulated detection-censored movement records have utility for planning and methodological comparison. In the case study presented here, detection probabilities were randomly distributed in space. Users may readily apply the tools in DiagnoseHR to ask more nuanced questions about how the spatial distribution of detection probabilities affects home range inference and other spatial analysis methods. For example, an investigator could ask how the size of a home range affects bias given different detection scenarios, or how different assumptions in the correlated random walk affect size estimation. Users may study the effects of spatial variation in detection by uploading a simulated world with deliberate patches of high and low detection probability, or by manipulating the degree of heterogeneity in detection probability. Future studies could additionally ask how the degree of site fidelity affects bias in home range estimates by manipulating the site fidelity values in different detection scenarios. Understanding the behavior of estimators under varying conditions will continue to have value, particularly for longitudinal studies and reassessment of long-term datasets, such as those maintained by natural resource agencies.

Investigators seeking to conduct home range analyses may choose from an ever-increasing array of estimation methods (e.g., Lyons et al., 2013; Wilson et al., 2018). Simulating test datasets offers investigators a way to assess tradeoffs among estimators while designing field studies. Moreover, simulating detection-censored occurrence datasets has the potential to provide utility beyond home range analysis. DiagnoseHR’s detection cells and sampling intervals may symbolize any real increment of space or time per the user’s needs. Datasets generated by DiagnoseHR may therefore have utility for planning and evaluating other spatial ecology methods, such as ecological niche modeling or mark-recapture abundance estimation, in which investigators seek to understand the distribution of organisms. Because DiagnoseHR is an open source R package, we and others may continually add new methods to the sensitivity functions and use test datasets to evaluate emerging spatial ecology analytical methods. The supplementary home range diagnostic methods we present begin to fill a practical gap by empowering investigators to plan field studies with an analysis in mind, evaluate the validity of home range estimate assumptions, and assess estimate sensitivity when comparing home range estimates from different systems.

5. Conclusions

Ultimately, our results demonstrate that home range analyses should be conducted in response to clearly defined scientific questions in the context of a well-understood system. The methods we provide facilitate the strategic planning of animal movement studies and the open discussion of home range results. Moreover, uncertainty in home range estimates should not be ignored or used as a justification to discard home range analyses. Rather, uncertainty in home range analyses, when explicitly discussed, has the potential to add value and
depth to home range analyses and facilitate the application of home range analysis as a tool for ecological inference and conservation. Our approach of empowering investigators to simulate animal movement, landscape variation, and investigator behavior processes is an advancement in animal movement simulation that facilitates assessment of tradeoffs in home range data collection and analysis.

Authors’ contributions

LW, VS, LC and CC conceived the ideas and designed the animal movement and simulation diagnostic functions. All authors contributed critically to the drafts and gave final approval for publication.

Funding

This research was funded by Federal Aid in Wildlife Restoration Project W-98-R, administered by the Nebraska Game and Parks Commission. CJC was supported by Hatch funds through the Agricultural Research Division at the University of Nebraska-Lincoln and from Federal Aid in Wildlife Restoration project W-120-T, administered by the NGPC.

Declarations of interest

None.

Data accessibility

The home range simulation was developed by C. Chizinski, L.Wszola, V. Simonsen, and L. Corral. The simulation code used in this analysis is available from: https://github.com/chizinski/ HomeRangeSimulations. The R package DiagnoseHR was developed by L.Wszola, V.Simonsen, and C. Chizinski, and is available for install via devtools from https://github.com/lw5077/DiagnoseHR.R. All data and code require R. R Studio is recommended. Although these data have been processed successfully, no warranty expressed or implied is made regarding the display or utility of the data on any other system or for general or scientific purposes, nor shall the act of distribution constitute any such warranty. The USGS or the U.S. Government shall not be held liable for improper or incorrect use of the data described and (or) contained herein.

Acknowledgements

We thank L. Messinger for the conversation which inspired this work, PWB for personal observations concerning the detection of small mammals, and the members of the SNR R Group for troubleshooting the home range simulation R code. Any use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The Nebraska Cooperative Fish and Wildlife Research Unit is supported by a cooperative agreement among the U.S. Geological Survey, the Nebraska Game and Parks Commission, the University of Nebraska, the U.S. Fish and Wildlife Service, and the Wildlife Management Institute.

References


