

Original Research Paper

A scaled line-based kernel density estimator for the retrieval of utilization distributions and home ranges from GPS movement tracks

Stefan Steiniger*† and Andrew J. S. Hunter†

† Department of Geomatics Engineering, University of Calgary, 2500 University Drive N.W.,
Calgary, Alberta, Canada T2N 1N4

*Corresponding author. Email: ssteinig@ucalgary.ca

Abstract Utilization distributions (UDs) can be used to describe the intensity with which an animal or human has used a certain geographical location. Within the domain of wildlife ecology, a density distribution model represents one way to describe an animals' home range. Several methods have been developed to derive UD, and subsequently home ranges. Most of these methods, e.g. Kernel Density Estimation (KDE), and Local Convex Hull methods, have been developed with point-based datasets in mind, and do not utilize additional information that comes with GPS-based tracking data (e.g., temporal information). To employ such additional information we extend the point-based KDE approach to work with sequential GPS-point tracks, the outcome of which is a line-based KDE. We first describe the design criteria for the line-KDE algorithm. Then we introduce the basic modelling approach and its refinement through the introduction of a scaling function. This scaling function modifies the utilization distribution so that a bone-like probability distribution for a single GPS track segment is obtained. Finally we compare the estimated utilization distributions and home ranges for two datasets derived using our line-KDE approach with those obtained using the point-KDE and Brownian Bridge (BB) approaches. Advantages of the line-based KDE by design are (i) a better representation of utilization density near GPS points when compared against the BB approach, and (ii) the ability to model and retain movement corridors when compared against point-KDE.

Keywords line-based kernel density estimation, movement analysis, utilization distribution, home range

1. Introduction

Wildlife ecology research has frequently been interested in understanding where an animal has been, or will be in the future. These questions are also important within the domain of location-based services (LBS), where the subject matter is typically human movement. For wildlife ecologists representing 'space utilization' may be beneficial for answering questions about why an animal spends time at a certain location, e.g. searching for food, sleeping, mating, etc., and hence useful for building behavioural models, or identifying natural reserves for endangered species (Powell, 2000; Kie et al., 2010; Stenhouse and Munro, 2000). For researchers and providers of location-based services such space utilization models can aid the

identification of daily travel patterns that can then be used to provide personalized (push & pull) services to a user (Steiniger et al., 2006; Nanni et al., 2008), such as information on traffic conditions between home and work place, or advertisements on specials by shops along a users daily commute.

A *Utilization distribution* (UD) can be derived from occurrence data, i.e. sightings or tracking information, and can be used to describe the intensity that an animal or human uses a certain geographical location. In wildlife ecology UD's are traditionally derived with probabilistic approaches, such as Kernel Density Estimation (KDE) applied to point sets that describe animal sightings (Harris et al., 1990; Worton, 1995; Powell, 2000). The KDE method generates a 2 dimensional (2D) probability surface represented by a grid of regular spaced locations (cells), which describes the probability (cell value) of an object being at a particular location (cell) in any part of its *home range* (Powell, 2000). The term "home range" is used here follows Burt's (1943) definition, "[... *the*] area traversed by the individual in its normal activities of food gathering, mating, and caring for young. Occasional sallies outside of the area, perhaps exploratory in nature, should not be considered as home range."

To account for occasional sallies, only a certain probability of use is considered when deriving a home range. For instance in the literature a home range is often considered to be the area contained within the 95 % contour (i.e. isopleth) calculated as 95 % of the volume of the probability surface (Powell, 2000). However, Kie et al. (2010) point out that home range and utilization distribution are often used interchangeably despite subtle differences in their definitions.

Besides the use of KDE to derive utilization distributions for home range analysis (Worton, 1989; Seaman and Powell, 1996) other methods for generating UD's have been developed relatively recently. These include the Local-Convex Hull (LoCoH) approach by Getz and co-workers (Getz and Wilmers, 2004; Getz et al., 2007), and the Brownian Bridge approach developed by Bullard (1991), which was applied and refined by Horne et al. (2007) and Calenge (2006) for use in animal home range analysis. Furthermore a Geo-Ellipse approach has been presented by Downs (2010) – but not with an application to animal home range estimation. However, Long and Nelson (2012) describe a very similar ellipse method for the calculation of home ranges and apply it to simulated tracking data. Other frequently used approaches for home range estimation, such as the Minimum Convex Hull/Polygon (MCP, see Burgman and Fox, 2003; Nielsen et al., 2008), the characteristic hull (Downs and Horner, 2009), parametric models (see Boulanger and White, 1990; Powell, 2000), or buffer-based approaches (Steiniger and Hunter, 2012), are able to generate home ranges described by one or several polygons, but do not generate utilization distributions. However, surrogates for density surfaces can be derived for some of these approaches. Getz et al. (2007) show that local convex hulls can be used to generate isopleths, as have Steiniger and Hunter (2012) for buffer-based approaches.

It is important to note that KDE, MCP, LoCoH and parametric approaches have been designed for and applied to location datasets derived using traditional observation methods, such as VHF telemetry and sightings. These point datasets typically contain small numbers of observations (50-300) and additional information, such as the exact time of recording, the direction of travel, or velocity of the moving object, may not be available, or is sparse.

The introduction of GPS to the wildlife community for tracking has enabled researchers to acquire datasets containing several thousands of points per field campaign. This has led to new research questions. A question that has been addressed to some extent, is: How do existing estimators perform with these new datasets? I.e., it is necessary to test existing estimators to determine if assumptions are still met, and if they are able to handle large volumes of data (but see Hemson et al., 2005; Börger et al., 2006; Huck et al., 2008). Another question is: How can estimators utilize additional information available via GPS tracking?

Two recently developed utilization density estimators that consider the time an observation is recorded are the previously mentioned Brownian Bridge approach (Horne et al., 2007) and the Geo-Ellipse approach (Downs, 2010; Long and Nelson, 2012). The Geo-Ellipse approach has been presented only very recently, i.e. it has not been tested with a diversity of LBS or animal location datasets yet. The Brownian Bridge approach has undergone some testing that has highlighted advantages (Horne et al., 2007), and weaknesses (Huck et al., 2008, and our tests in Section 4). Although the BB approach is promising, we have concerns that the basic assumptions underlying the approach are derived from uncertainty modeling, as BB fundamentally describes the probable location of a particle experiencing Brownian motion. Hence, the BB model was not designed originally to model space utilization. A consequence this is that the expected probability for the bridge is zero at each GPS point (see Fig. 2 later). Despite Bullards (1991) effort to overcome this issue by adding a kernel to the bridge endpoints (e.g. the GPS points), the parameterization of the BB remains challenging if the uncertainty model is to be used to model space utilization or probability of use.

The aim of this article is to present a third estimator for GPS datasets. As such, our estimator for utilization distributions extends the point-based kernel density estimator. However, the estimator adopts concepts from the Brownian bridge and geo-ellipse approaches, i.e. it makes use of GPS-track segments as the base-object. After introducing the design specification in the next section we outline how the line-based Kernel Density Estimator (line-KDE) works, and later compare results for a test dataset with results obtained for the BB and point-KDE approaches.

2. Design Specification for the Utilization Density Estimator

The utilization distribution/density estimator has been designed with certain data and applications in mind, in particular, GPS-tracking data and animal home range analysis. All in all four different sets of criteria were identified:

A - GPS track vs. (Road) Networks – The estimator should model densities from tracking data of the following type: GPS points are connected as temporal sequences to form a chain. Under- and overpasses of that chain can exist, but the chain will not contain junctions at which the track divides into two or more chains. This specification is important since it does not guarantee that our estimator can be applied to planar networks, such as road networks. In contrast the kernel density estimator developed by Okabe et al. (2009) for road networks requires a planar network where no under- or over-passes exists, i.e. lines do not intersect except at junctions. Since road networks have junctions, Okabe et al. (2009) model two cases of estimator behaviour at junctions, a discontinuous and a continuous case (see also Produit et al., 2010). This is not necessary for our use case. Further, Okabe et al. (2009) also define that the network is not directed. In contrast, GPS track data are. Based on the specifications we developed a small reference dataset shown in Fig. 2, image A.

With respect to the model developed we also required that the model conforms to the following requirements: (i) the density/utilization value at a node where two line segments touch is the same as for a track starting or end point, (ii) density values add-up at crossings of line segments, (iii) the peak of the density for two subsequent line segments connecting in a acute angle should be at the connecting point of the two line segments, and (iv) if two line-segments within the bandwidth (i.e. kernel size) run parallel to each other (no matter what direction both have), then density values are added as well. These rules informed the creation of our reference dataset (see Fig. 2, image A).

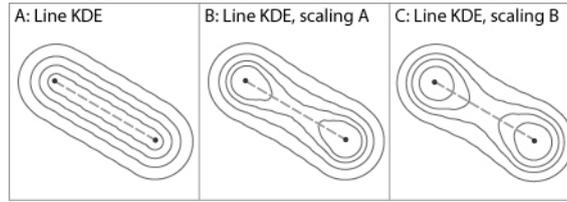


Fig. 1 Probability (25, 50, 75, 95 %) of utilization contours for line-KDE and scaled line-KDE for a single line segment – scaling A: $sf \in [\sqrt{0.5}, 1.0]$ and B: $sf \in [0.5, 1.0]$. For a description of the scaling variants see Section 3.3.

B - Ecological considerations – Since the application focus of the estimator is animal home range analysis we define that: (i) the estimator must produce a density/utilization surface, so that different values of probability of use, i.e. contours, can be derived, (ii) the estimator should allow a user to identify / model travel corridors (Rosenberg et al., 1997; Mabry and Barrett 2002; Bennett, 2003) as they are in our opinion an important part of the home range and home range analysis. This criterion has been introduced since point-based KDE is not able to uncover information on corridors if the GPS point sampling frequency is inappropriate, i.e. if the sampling period is too large (e.g. one or two sampling locations per day). (iii) The estimator should be able to respect the shape of the point dataset, i.e. contours created from the density grid should resemble the shape of the point-distribution as outlined in Downs and Horner (2007, 2008). We note here, that we have not yet tested if this criterion is fulfilled – but we assume that our model will; in particular when the kernel bandwidth is chosen based on the average daily travel distance (see below)

C - Uncertainty vs. utilization - The estimator should not model uncertainty, as the Brownian Bridge approach does, but utilization and probability of occurrence. Subsequently: (i) GPS points will not yield a zero density/probability value but a value similar to that obtained with point-KDE, and (ii) the density/probability value along the line segment will decrease from the end points toward the (geometric) center of the segment. As a result the contours generated from the utilization function for a single line segment should resemble a *bone*-like shape instead of the convex buffer shape that is obtained when buffering a straight line in a GIS (compare Fig. 1, image A — where the outline forms a convex hull similar to a GIS buffer operation, versus Fig. 1 images B & C where the outline narrows towards the center of the line segment resulting in a bone-like shape).

D - Use of additional (time) information – To determine the (bandwidth-) parameter of the estimator it would be beneficial to utilize the time information of the GPS points. This is the approach adopted by Horne et al. (2007) for their Brownian bridge implementation and Downs’ (2010) geo-ellipse method. We have found that the average daily travel distance is also a promising approach for determining bandwidth (see Steiniger and Hunter, 2012).

3. A Scaled Line-based Kernel Density Estimator

In the following we will describe the basic approach that was developed for the calculation of utilization densities from GPS tracking data. The estimator is based on the well-known point-based KDE approach to respect specification *B-i*, the creation of a utilization surface. Extending the KDE approach to work on lines as base objects, instead of points, should ensure that specification *B-ii*, modeling corridors, and *C-i*, non-zero density values at GPS point locations, are met. We then introduce a scaling function used to achieve a bone-like shape for contours of the utilization function of a single line segment (specification *C-ii*). Finally, we will address options for parameter selection (specification *D*).

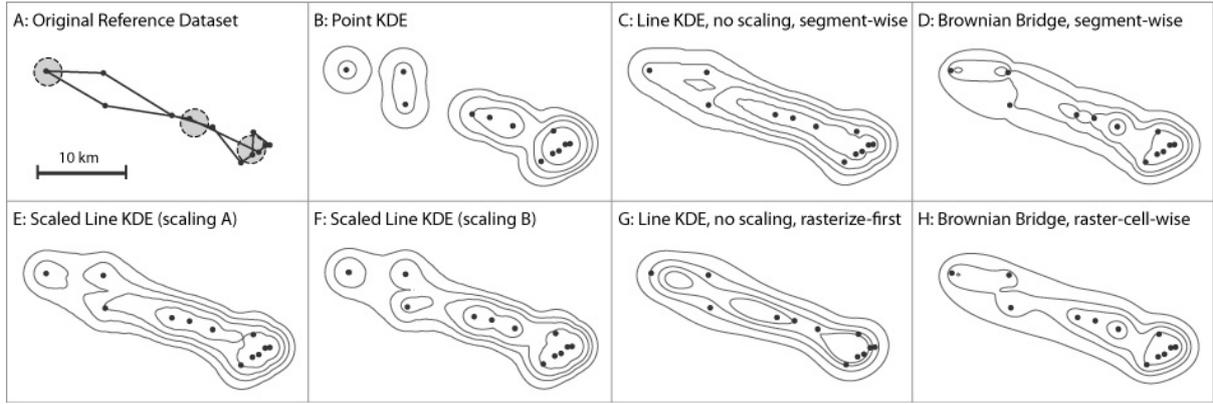


Fig. 2 Reference point dataset and 25, 50, 75 and 95 percent contours for probability of utilization with point-KDE, line-KDE (rasterize-first and segment version), scaled line-KDE (scaling A and B), and Brownian Bridge. For a description of the line-KDE versions and scaling variants see Sections 3.2 and 3.3. The dashed circles in image A mark important configurations, i.e. i - acute angle, ii - parallel tracks, iii - crossing tracks.

3.1 Line-based Kernel Density

The line-based kernel density estimator (line-KDE) for GPS tracks is fundamentally built on the point-based KDE. The basic equation that describes the bivariate point-based kernel density estimate is (Silverman, 1986 pg. 76; Worton, 1989):

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^2} \sum_{i=1}^n K \left\{ \frac{1}{h}(\mathbf{x} - X_i) \right\} \quad (1)$$

where n is the number of randomly sampled observations $X_1 \dots X_n$, h is the bandwidth parameter, and K is the kernel function. Density estimation can be considered a continuous form of binning data, similar to that done during histogram estimation. To obtain a 2D kernel density estimate for a point dataset in praxis a 2D kernel function K of a particular shape and with certain properties is chosen (e.g. a Gaussian function; see Silverman, 1986; and Scott, 1992, for a range of functions and properties). A raster with a user-defined cell size s is created that covers all observations, whereby s is smaller than h . Then the kernel function is placed over each observation X_i . The 2D kernel function is scaled according to the bandwidth h . For each grid cell x_j that is within the window defined by h the (weighted) kernel value $\hat{f}(x_j)$ is calculated with respect to the distance d_{ij} to the observation X_i (i.e. $d_{ji} = x_j - X_i$); with weights previously assigned to each X_i . Typically a weight of 1 is given to tracking data, but the weight may differ in cases where X_i represents a set of points. If a raster cell already contains a density value $\hat{f}(x_j)$, from an earlier kernel calculation of a close by observation point $X_{k \neq i}$, then the new value $\hat{f}_n(x_j)$ is added to the existing value, i.e. $\hat{f}(x_j) = \hat{f}(x_j) + \hat{f}_n(x_j)$.

The line-KDE is based on the previously described point-based approach such that each line segment $l_{i,i+1}$ (with $i = 1 \dots n$ GPS points p_i) of the GPS track is transformed into single points $p_r^{temp(i,i+1)}$ (i.e. rasterized with respect to cell size s), and then the point-KDE algorithm is applied to each of those points $p_r^{temp(i,i+1)}$. To avoid density values being added together from consecutive line segments the maximum density value between an existing cell value $\hat{f}(x_j)$ and its newly calculated value $\hat{f}_n(x_j)$ is chosen, i.e. $\hat{f}(x_j) = \max \left\{ \hat{f}(x_j), \hat{f}_n(x_j) \right\}$.

In principle there are two approaches that can be used to implement the line-KDE approach for GPS tracks:

- *rasterize-first* - the complete track of segments $l_{1...n}$ is rasterized and the point-KDE is calculated; or
- *segment-wise* - each segment $l_{i,i+1}$ is rasterized individually, and the raster $r_{i,i+1}$ is calculated using the point-KDE. Then the rasters $r_{i...n}(l_{i,i+1})$ are added.

In the latter case when the rasters $r_{i,i+1}$ and $r_{i+1,i+2}$ are added it must be ensured that the density values derived for the observation point X_{i+1} , which connects the segments $l_{i,i+1}$ and $l_{i+1,i+2}$, do not aggregate (specification *A-i*). This can be achieved by subtracting the density raster of the previous line segment $r_{i,i+1}$ from $r_{i+1,i+2}$ before adding $r_{i+1,i+2}$ to the final density raster.

With respect to specification *A-iii* we note that the rasterize-first approach will yield a translated center of maximum density for two subsequent line segments $l_{i,i+1}$ and $l_{i+1,i+2}$ that connect with an acute angle, i.e. the maximum density value will not be at the GPS point X_{i+1} that is connecting both track segments (compare Figure 2, images C and G). From visual inspection the rasterize-first approach appears to be applied in ESRI's ArcGIS product. However, both approaches, i.e. segment-wise and rasterize-first, ensure that densities are added when line-segments of the GPS track cross or run parallel to each other (specifications *A-ii* and *A-iv*). We finally note that this approach does not result in a normalization of the volume. Hence, probabilities cannot directly be measured from the raster, but pseudo probability contours can be derived.

3.2 Scaling approach

In design specification *C-ii* we defined that density values for a single line segment should decrease towards the geometric middle of the line segment, since we think that this models the probability of space utilization more accurately. Contours generated from a utilization density grid derived for a single line segment should resemble a bone-like shape. To retain a bone-like shape we apply a scaling function to the kernel K that was initially developed by Caspary and Scheuring (1993) for the line error-band model (see also Shi et al., 1999, for a similar line uncertainty model). The equation for the positional error σ_{P_i} for a line is according to Caspary and Scheuring (1993, pg. 108):

$$\sigma_{P_i} = \sqrt{2\sigma_{x_i}} \quad \text{with} \quad \sigma_{x_i}^2 = \left(1 - \frac{2l_i}{l} + \frac{2l_i^2}{l^2}\right) \sigma^2 \quad (2)$$

with σ being the initial error (e.g. from digitizing a line), l_i the length of the rasterized line between the line segment end points, and l the length of the line segment. The error function σ_{P_i} can be used as scaling function sf in our model to modify (i.e. decrease) the probability of object observation in a location. When setting $\sigma = 1.0$ and defining $sf_{x_i}^A = \sqrt{\sigma_{x_i}}$, we obtain a scaling range $sf^A \in [\sqrt{0.5}, 1]$. An alternative scaling is achieved by setting $sf_{x_i}^B = \sigma_{x_i}$, resulting in the range $sf^B \in [0.5, 1]$. The derived scaling value sf is then used to adjust the calculated density value $\hat{f}(x_j)$ for each rasterized point $p_r^{temp(i,i+1)}$ of the line segment $l_{i,i+1}$, i.e. $\hat{f}(x_j) = \hat{f}(x_j) \cdot sf(p_r^{temp(i,i+1)})$.

Currently scaling is only applied to the segment-wise line KDE described above, and not for the rasterize-first approach. In the latter approach and with the current processing strategy it is not possible to store configuration information to distinguish between segment crossings and endpoints. Subsequently it is not clear if the new and old density values of a raster cell should be added, or if the maximum of new and old values should be taken. It is easy to see from the contours in Figure 1 that scaling by sf^A is less influential than for sf^B , with version B

having a more pronounced bone-like shape. Resulting contours for the line-KDE with and without scaling for the reference dataset are shown in Figure 2.

3.3 Parameterization

The scaled line-KDE approach described in the previous two sub-sections requires three parameters to be defined by the user. The choice of the kernel function K (i), and (ii) the selection of the scaling function (sf^A or sf^B) can be considered of minor importance in comparison with (iii) the choice of the KDE bandwidth parameter h . Changing the kernel function has limited effect as has been pointed out by Silverman (1986, pgs. 43 and 86), and was also apparent in our own experiments.

Besides the option to choose the window width h based on expert knowledge, several automatic methods have been proposed to calculate h for the point-based KDE. The article by Kie et al. (2010) includes a review on these methods. The most prominent automated methods for choosing h are the *Reference* method (with $h_{ref} = \sigma_{xy} \cdot n^{-1/6}$, Silverman, 1986), and the *Least-Squares Cross Validation* method (h_{LSCV}) that is based on stepwise minimization of an error criterion (Worton, 1995; Sheather, 2004). As our line-KDE approach is essentially built on the point-based KDE, using h_{ref} is one option for choosing the bandwidth. However, having additional information on the recording time for each GPS point allows the analyst to derive average travel distance of an animal per day, for example. Hence, in our experiments we used the median value of all average daily travel distances (of one GPS track) as bandwidth parameter h_{mdb} , in addition to h_{ref} .

3.4 Implementation

We implemented the scaled line-KDE approach in the free and open source GIS software OpenJUMP (Steiniger and Hay, 2009). The Sextante toolbox was coupled with OpenJUMP (Olaya, 2008) to deliver the basic point-KDE and raster processing functionality. The current implementation contains only one kernel function at this time, the biweight kernel described in Silverman (1986:76, eq. 4.5). We note that the kernel implementation operates without scaling for volume/probability. However, other kernels can be easily added, since they are already available for the point-based KDE. A special home-range analysis edition, OpenJUMP HoRAE (Steiniger and Hunter, 2012), which contains the scaled line-based KDE and other home range analysis method is available from http://gisciencegroup.ucalgary.ca/wiki/OpenJUMP_HoRAE.

4. Comparison with other Density Estimators

4.1 Estimators

To evaluate the performance of the line-based KDE implemented, we compared the estimator with point-based KDE, and the Brownian Bridge (BB) estimator. We chose these two as we view them as benchmarks with respect to our design criteria, which included among others the ability to model travel corridors (for KDE), and to not resemble a pure uncertainty model, as BB does. A point-based KDE function is already implemented in the Sextante toolbox, but was modified to work with the tracking data (e.g. calculation of grid bounds and determination of h_{ref}). The BB implementation we used is a port of the version implemented by Calenge (2006) in the R *adehabitat* package. However, similar to the line-based KDE we implemented two BB versions, one resembling a segment-wise processing approach and the other using a cell-based approach. To determine the two BB parameters σ_1 and σ_2 we used/ported the method developed by Horne et al. (2007). In our experiments σ_1 is estimated by keeping σ_2 fixed with 30m, whereby the value of 30m reflects the estimated GPS location

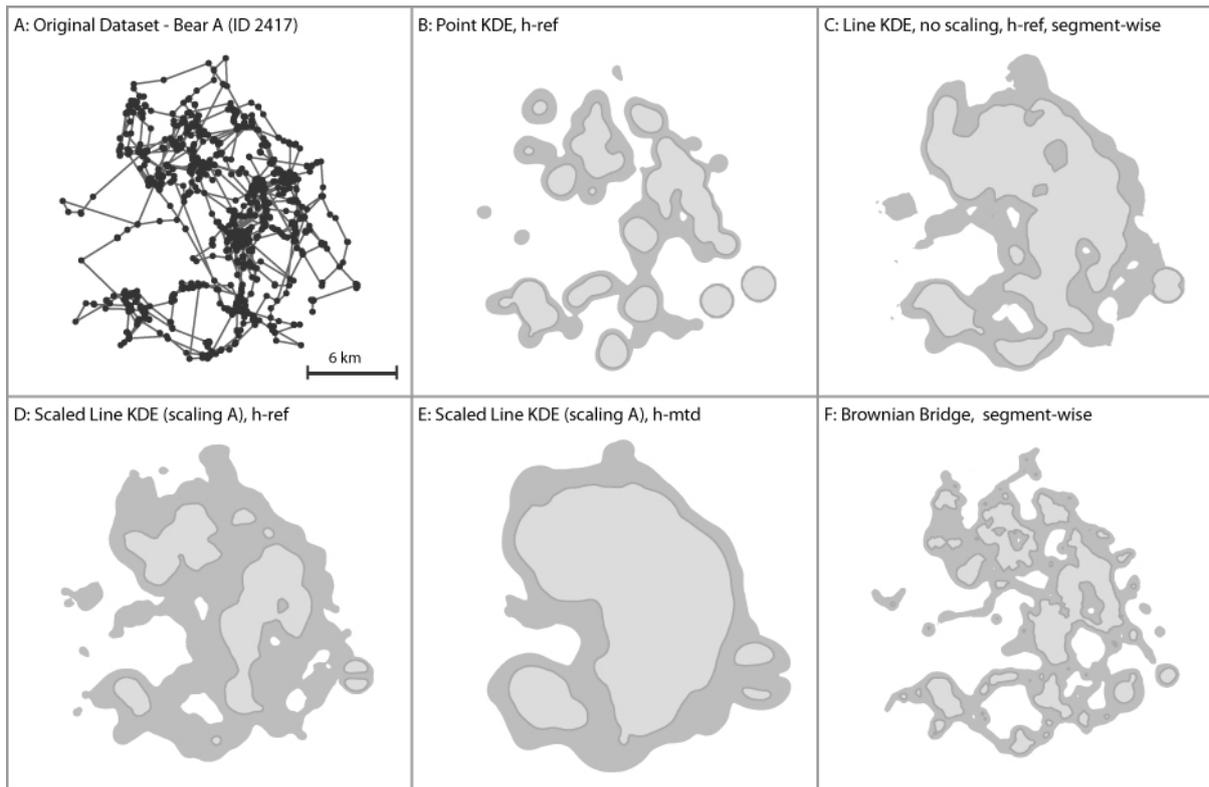


Fig. 3 The results for different utilization distribution estimators, visualized by the 95% probability of use area (dark grey) and the core area (light grey) for bear A. Parameters: KDE $h_{ref}=1240m$, $h_{mtd}=2600m$; Brownian Bridge $\sigma_1=8.30m$, $\sigma_2=30.0m$.

error (95 % confidence interval). We note that for the BB method parameter estimation difficulties arise from numerical problems, i.e. we have not been able to calculate e^n for $n > 710$ with a 32-bit java machine due to precision limits. Hence, the estimated value of σ_1 may be incorrect, and problems in the calculations may lead to different results than expected, given the mathematical model.

To aid the comparison, we also calculated the probability that determines the core of a home range, representing areas of higher utilization. The method used is based on the evaluation of home range area for the x percentile contour (with $x \in [2.5\%, 100\%]$ in 2.5% steps), and is outlined in Harris et al. (1990). The criterion used to determine the core probability was adopted from Seaman and Powell (1990) and Powell (2000) — choosing the longest distance from the diagonal line of the area-probability plot. Seaman and Powell’s (1990) core calculation approach differs from Harris et al. (1990) in that it is cell-based and not contour-based. In our experiments both methods often gave different results. Hence, for the comparison we used only the contour-based method.

4.2 Datasets

In addition to the artificial reference dataset that was used to control if our design criteria were met we also tested the implementation using two real datasets. We obtained GPS collar data from the Alberta Foothills Research Institute from two grizzly bears. Bear A was female and resides in the foothills of the Rocky Mountains within Alberta, whereas bear B was male and resides generally at higher elevations within the Rocky Mountains. As such, travel paths and home range extents should be different. It is expected that the mountain grizzly will tend

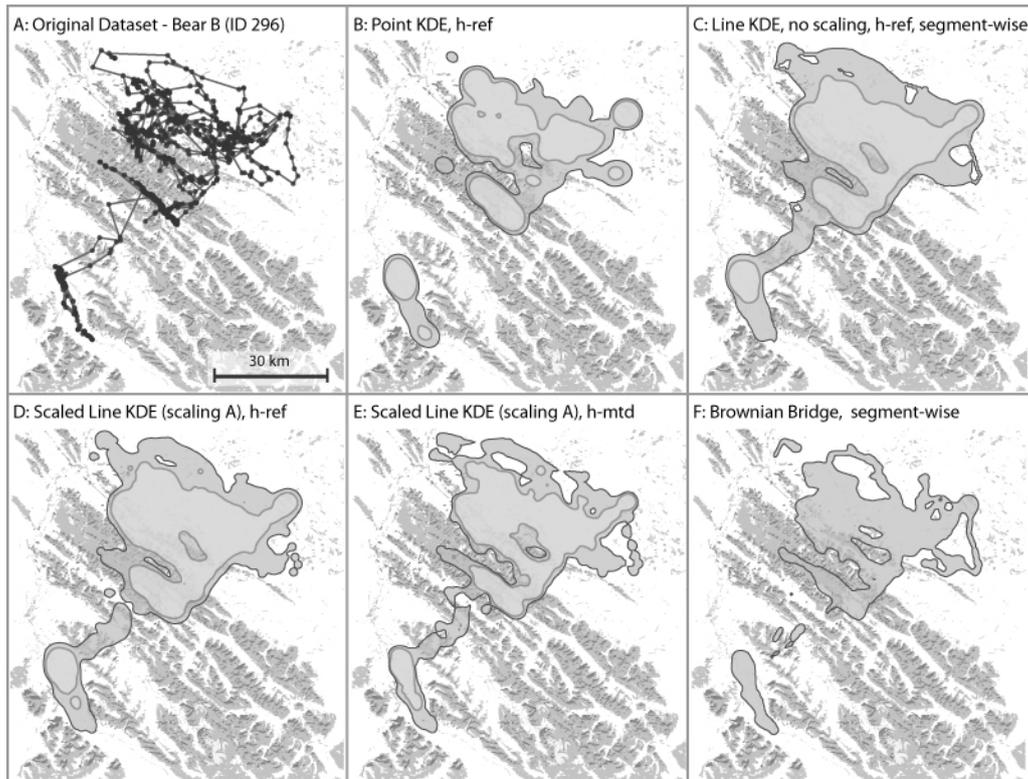


Fig. 4 The results for different utilization distribution estimators, visualized by the 95% probability of use area (dark grey) and the core area (light grey) for bear B. The background shows unvegetated areas, indicating steep mountain slopes. For the Brownian Bridge no core contour is shown, since the calculated utilization probability for the core was 95% as well. Parameters: KDE $h_{ref}=4900m$, $h_{mtd}=3250m$; Brownian Bridge $\sigma_1=19.05m$, $\sigma_2=30.0m$.

to travel along valleys. Corridors that connect those valleys may be important from a behavioural analysis perspective.

The dataset for bear A contained 2,215 points, and was recorded over 139 days from June to November during 2007. The dataset for bear B contained 1,525 GPS locations that were acquired over 194 days from April to November 2001. Both datasets include the time and date that a GPS location was recorded, and therefore the point locations can be ordered. For the Brownian Bridge method the time information was transformed into seconds of the year.

4.3 Results

The comparison was performed more or less qualitatively and visually, focussing on features related to shape, such as (i) shape complexity (patches, holes, smoothness of the outline) and (ii) area of the home range polygon. We also noted if (iii) travel corridors were observed, and (iv) how long calculation took (Table 1). The results for the reference dataset are displayed in Figure 2. Figures 3 and 4 show the results for the grizzly bear datasets. For those we have displayed the 95 percent contour of utilization density (a commonly accepted value in the literature; Laver and Kelly, 2008) and the core contour of the home range. For bear B we also displayed unvegetated areas (in grey) that reflect steep mountain terrain. This allowed us to assess if the bear traveled along valleys, and to visually compare how well the estimators model travel paths.

Complexity of home range shape – When looking at the results for both datasets the shapes calculated using the Brownian Bridge (BB) approach are rougher, or alternatively, more detailed. The point-based and line-based KDE approaches tended to result in less detailed

outlines when compared with the BB method. For the line-based approaches the bandwidth parameter, i.e. h_{ref} or h_{mtd} , determined the level of smoothing of the home range contours. For grizzly bear A the BB results appear too detailed (observe the convoluted core contours) and produced a higher number of patches and holes for the 95% contour (Table 1). The point-KDE approach looks somewhat spotted as well. For bear B in the mountains, the BB contours seem to adapt best to the topography and indicate that travel paths are in the valleys. Unfortunately, for the BB a core contour could not be identified for further assessment — we identified the 97.7 % contour as the core contour, which was larger than the chosen 95% probability to delineate the home range. However, also the line-KDE with bandwidth h_{mtd} appeared to adapt well to the mountain ranges, which we believe is a result of the smaller bandwidth when compared to the other KDE parameterizations. Interestingly, the results for the line-KDE with and without scaling are not that different for these datasets. If one compares only the 95% contour for bear A, and not the core contour, then the differences are negligible. Hence, the scaling has little influence on both datasets; however, it does affect the smoothness of the outlines and the calculation of the core area.

Table 1. Experimental Results.

method	parameters	processing time ⁴	core probability	Area [km ²]	patches/holes	corridors observed
Bear A ¹						
Point-KDE	$h_{ref}=1240\text{m}$	1 sec	80%	142	9/1	No
Line-KDE(s ²)	$h_{ref}=1240\text{m}$	11min	75%	230	7/6	Yes
S-LKDE A ³	$h_{ref}=1240\text{m}$	13min	72.5%	229	6/6	Yes
S-LKDE A	$h_{mtd}=2600\text{m}$	13min	77.5%	292	1/0	Yes
BB (s ²)	$\sigma_1=8.3\text{m}, \sigma_2=30\text{m}$	21min	70%	164	9/14	Yes
Bear B ¹						
Point-KDE	$h_{ref}=4900\text{m}$	1 sec	82.5%	1416	4/1	No
Line-KDE(s ²)	$h_{ref}=4900\text{m}$	11min	75%	2000	1/9	Yes
S-LKDE A ³	$h_{ref}=4900\text{m}$	14min	77.5%	1974	6/2	Yes
S-LKDE A	$h_{mtd}=3250\text{m}$	17min	77.5%	1670	6/4	Yes
BB (s ²)	$\sigma_1=19.05\text{m}, \sigma_2=30\text{m}$	31min	97.5%	1182	8/8	Partly

¹) Raster cell size for bear A: 200m, and bear B: 400m. ²) s: segment-wise processing. ³) S-LKDE A: Scaled Line KDE-with segment-wise processing and scaling A [0.707...1.0]. ⁴) The algorithms were run on a Dell XPS Laptop with a Intel Core 2 Duo CPU 2.2 GHz, with Java 1.6 and within the Eclipse development environment.

Modelling travel corridors – For all three datasets it is evident that the point-KDE does not observe travel corridors. In contrast the line-KDE method models travel corridors fairly well (see Figure 4, bear B, for the best example). For the same dataset it can be observed that those corridors are not necessarily connected with other parts of the home range, i.e. several home range polygons were calculated. Similar to the line-KDE the BB approach was able to model travel corridors (Figures 3 and 4). However, not all corridors could be modelled. This is observed in the results for the mountain bear B where three small polygons indicate a possible corridor between the two main home range patches. We note in addition that the effect of the zero density values at GPS locations for the BB method is particularly visible on single-track segments that may be corridors. The contour lines in Figure 2-D indicate zero density values, a result of the underlying uncertainty concept.

Area of home range – Considering travel corridors as part of the home range has an effect on the calculated area. Hence, the area for the home ranges calculated with the line-KDE

approaches was always larger than the area for the point-KDE method (Table 1). Although the BB approach models corridors in part, we obtained different results for the two bear datasets. For bear A the BB-based home range area is larger than the point-KDE area, whereas for bear B the BB area is smaller than the point-KDE area.

Calculation times – We observed remarkable differences in calculation times for the three utilization density estimators. Clearly the fastest approach is the point-KDE taking on average only one to two seconds to produce a grid for 1,500-2,000 GPS locations (Table 1). The segment-wise line-KDE approaches required 10 to 20 minute run times. In contrast, the rasterize-first line-KDE was comparably fast, taking three seconds (data not shown). However, we wish to note that the rasterize-first version does not allow for scaling, and that density centers may move away from point locations (compare Figure 2, images C and G). Finally, the calculation of the segment-wise BB approach took between 20 to 30 minutes and the cell-based BB version was the slowest with over an hour of processing time (data not shown). Interestingly the Brownian Bridge processing in R using the *adehabitat* package was much faster than our java implementation. However, from testing with our software it is clear that advantages in modelling have to be weight against the costs of processing time for large datasets.

5. Discussion and Conclusions

The tests and comparisons have shown that the designed line-based KDE algorithms produce the expected utilization density grids with respect to the GPS point configurations. Subsequently the expected home range estimates are obtained. In particular (i) travel-corridors are now included in the derivation of a home range, a limitation of point-based KDE, and (ii) GPS locations do not generate zero utilization density values, a limitation of the Brownian Bridge model. However, we also note that introducing a scaling of the kernel along the line/track-segment had little effect with our two bear datasets when considering the 95% contour.

After we drafted this manuscript Benhamou and Cornelis (2010) published an article on a new moving kernel-based algorithm, like ours. However, the details of their specification and implementation are different. For instance they advocate a geo-ellipse-like shape and include constraints to the home range boundary that stem from topographic features (e.g. a river). Whereas experiments performed by us have shown as well a need to include topographic constraints in home range modeling; the choice of a segments UD shape is clearly different. We advocate the choice of a bone-like shape, which originates from error models, while others advocate the geo-ellipse shape, stemming from time-geography. It would be interesting to know what shape biologists and wildlife ecologists would chose, and why.

Benhamou and Cornelis (2010) noted that a potential weakness of their method comes from its lack of a standardized procedure for setting minimum and maximum values of the smoothing parameter/bandwidth. When we tested our new algorithm we proposed a new option to determine the kernel bandwidth, h_{mtd} , which is calculated as the median of the average travel distance per day. The experiments on two real datasets show that this value could be smaller or larger than h_{ref} . Based on our limited tests we recommend the use of h_{mtd} if the value is smaller than h_{ref} , since it has been noted that h_{ref} may return a too large value for clumped data – i.e. often produces oversmoothed density estimates (see Sheather, 2004; Kie et al., 2010). If h_{mtd} is larger than h_{ref} , then we recommend using h_{ref} to retain more detail.

As we only undertook a limited set of experiments to show that the algorithm works it is now necessary to test the robustness of the algorithm and to determine its application range (Rykiel, 1996). This testing should be done in particular with respect to (i) other types of GPS movement data, and (ii) for other species, since different modes of movement will produce different types of GPS point location patterns (Turchin 1998). Finally, a additional need, besides the determination of the application range of the algorithm, is to find ways to optimize

processing time. Because waiting 20 minutes for each calculation to be finished will discourage experimentation with the parameter settings, which we believe will limit insight into the data. All code and functions presented in this article can be freely downloaded, modified, and distributed in accordance with the GPL software license (Steiniger and Hay, 2009). Hence, we invite others to test and improve our algorithms.

Acknowledgements

Parts of this work have been accomplished under Phase IV GEOIDE project, #03. We are grateful for funding. We also thank the Foothills Research Institute, Alberta, for providing data.

References

- Benhamou, S., Cornélis, D., 2010. Incorporating movement behaviour and barriers to improve kernel home range space use estimates. *J. Wildl. Manage.* 74, 1353-1360.
- Bennett, A.F., 2003. Linkages in the Landscape: The Role of Corridors and Connectivity in Wildlife Conservation. Conservation Forest Ecosystems Series No. 1, second ed., IUCN, Gland.
- Boulanger, J.G., White, G.C., 1990. A comparison of home-range estimators using Monte Carlo simulation. *J. Wildl. Manage.* 54, 310-315.
- Börger, L., Franconi, N., De Michele, G., et al., 2006. Effects of sampling regime on the mean and variance of home range size estimates. *J. Anim. Ecol.* 75, 1393-1405.
- Bullard, F., 1991. Estimating the Home Range of an Animal: a Brownian Bridge Approach. MSc thesis. University of North Carolina at Chapel Hill.
- Burgman, M.A., Fox, J.C., 2003. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Anim. Conserv.* 6, 19-28.
- Burt, W.H., 1943. Territoriality and home range concepts as applied to mammals. *J. Mammal.* 24, 346-352.
- Calenge, C., 2006. The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals. *Ecol. Modell.* 197, 516-519
- Caspary, W., Scheuring, R., 1993. Positional accuracy in spatial databases. *Comput. Environ. Urban. Syst.* 17, 103-110.
- Downs, J.A., 2010. Time-geographic density estimation for moving point objects, in: Fabrikant, S.I., Reichenbacher, T., van Kreveld, M., Schlieder, C. (Eds.), *GIScience 2010*, LNCS 6292, Springer, Berlin, pp. 16-26.
- Downs, J.A., Horner, M.W., 2007. Network-based home range analysis using Delaunay triangulation. *ISVD '07*, IEEE Comp. Soc., Washington, pp. 255-259.
- Downs, J.A., Horner, M.W., 2008. Effects of point pattern shape on home-range estimates. *J. Wildl. Manage.* 72, 1813-1818.
- Downs, J.A., Horner, M.W., 2009. A characteristic-hull based method for home range estimation. *Transactions in GIS* 13, 527-537.
- Getz, W.M., Wilmers, C.C., 2004. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography* 27, 489-505.
- Getz, W.M., Fortmann-Roe, S., Cross, P.C., et al., 2007. LoCoH: Nonparametric kernel methods for constructing home ranges and utilization distributions. *PloS ONE* 2, e207
- Hemson, G., Johnson, P., South, A., et al. 2005. Are kernels the mustard? Data from global positioning system (GPS) collars suggests problems for kernel home-range analysis with least-squares cross-validation. *J. Anim. Ecol.* 74, 455-463.

- Huck, M., Davison, J., Roper, T.J., 2008. Comparison of two sampling protocols and four home-range estimators using radio-tracking data from urban badgers *Meles meles*. *Wildl. Biol.* 14, 467-477.
- Harris, S., Cresswell, W.J., Forde, P.G., et al., 1990. Home-range analysis using radio-tracking data – a review of problems and techniques particularly as applied to the study of mammals. *Mammal Rev.* 20, 97-123.
- Horne, S.J., Garton, E.O., Krone, S.M., Lewis, J.S., 2007. Analyzing animal movements using Brownian Bridges. *Ecology* 88, 2354-2363.
- Kie, J.G., Matthiopoulos, J., Fieberg, J., et al., 2010. The home-range concept: are traditional estimators still relevant with modern telemetry technology? *Philos. Trans. R. Soc. B* 365, 2221-2231.
- Laver, P.N., Kelly, M.J., 2008. A critical review of home range studies. *J. Wildl. Manage.* 72, 290-298.
- Long, J.A., Nelson, T.A., 2012. Time geography and wildlife home range delineation. *J. Wildl. Manage.* 76, 407-413.
- Mabry, K.E., Barrett, G.W., 2002. Effects of corridors on home range sizes and interpatch movements of three small mammal species. *Land. Ecol.* 17, 629-636.
- Nanni, M., Kuijpers, B., Körner, C., May, M., Pedreschi, D., 2008. Spatiotemporal data mining, in: Giannotti, F., Giannotti, G., Pedreschi, D. (Eds.), *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*, Springer, Heidelberg, pp. 267–296.
- Nielsen, E.B., Pedersen, S., Linnell, J.D.C., 2008. Can minimum convex polygon home ranges be used to draw biological meaningful conclusions? *Ecol. Res.* 23, 635-639.
- Okabe, A., Satoh, T., Sugihara, K., 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geog. Inf. Sci.* 23, 7-32.
- Olaya, V., 2008. SEXTANTE, a free platform for geospatial analysis. Available from: <https://forge.osor.eu/docman/view.php/13/74/ArticleForOsGEOJournal.pdf>
- Powell, R.A., 2000. Animal home ranges and territories and home range estimators, in: Boitani, L., Fuller, T.K. (Eds.), *Research Techniques in Animal Ecology*, Columbia University Press, New York, pp. 65-110.
- Produit, T., Lachance-Bernard, N., Strano, E., Porta, S., Joost, S., 2010. A network based kernel density estimator applied to Barcelona economic activities, in: Taniar, D., et al. (Eds.), *ICCSA 2010, LNCS 6016*, Springer, Heidelberg, pp. 32-45.
- Rosenberg, D.K., Noon, B.R., Meslow, E.C., 1997. Biological corridors: form, function and efficacy. *BioScience* 47, 677-687.
- Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecol. Modell.* 90, 229-244.
- Scott, D.W., 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York.
- Seaman, D.E., Powell, R.A., 1990. Identifying patterns and intensity of home range use. *Int. Conf. Bear Res. and Manage.* 8, 243-249.
- Seaman, D.E., Powell, R.A., 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology* 77, 2075-2085.
- Sheather, S.J., 2004. Density estimation. *Stat. Sci.* 19, 588-597.
- Shi, W.Z., Ehlers, M., Tempfli, K., 1999. Analytical modelling of positional and thematic uncertainties in the integration of remote sensing and geographical information systems. *Transactions in GIS* 3, 119-136.
- Stenhouse, G.B., Munro, R.H.M., 2000. *Foothills Model Forest Grizzly Bear Research Project 1999 Annual Report*.
- Steiniger, S., Hay, G.J., 2009. Free and open source geographic information tools for landscape ecology. *Ecol. Inform.* 4, 183-195.

- Steiniger, S., Hunter, A.J.S., 2012. OpenJUMP HoRAE – A free GIS and toolbox for home range analysis. *Wildlife Society Bulletin* 36, 600-608.
- Steiniger, S., Neun, M., Edwardes, A., 2006. *Lecture Notes: Foundations of Location Based Services*. Department of Geography, University of Zurich.
- Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Turchin, P., 1998. *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sinauer Associates, Sunderland, Massachusetts.
- Worton, B.J., 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70, 164-168.
- Worton, B.J., 1995. Using Monte Carlo simulation to evaluate kernel-based home range estimators. *J. Wildl. Manage.* 59, 794-800.