Version 28. July 2010, unpublished document/white paper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 Geography, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4

§ 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 with what intensity an animal or human may use a certain geographical location within the environment it is living in. Such a density distribution model represents one way to describe and obtain an animals' home range in wildlife ecology. Several methods to derive UDs and subsequently home ranges have been developed, for instance Kernel Density Estimation (KDE) and Brownian Bridges (BB), two probabilistic approaches, and Local Convex Hull (LoCoH) methods. KDE and LoCoH have been developed with point-based datasets in mind that describe the observation of an animal, and hence don't utilize additional information that comes with GPS-based tracking data from collars. We have extended the point-based KDE approach to work with sequential GPS-point tracks, calling it a line-based KDE. We (i) introduce the basic-approach, (ii) refine it by introducing a scaling function to achieve a better model for the utilization function of space, and (iii) compare results of our approach with point-KDE and BB. Advantages of the line-based KDE by design are a better representation of utilization density near GPS points in comparison with the BB approach, and the ability to model and retain travel corridors in comparison with the point-KDE.

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

1. Introduction

For research in wildlife ecology but nowadays also for research in location-based services (LBS) it is of interest to precisely model where an animal or a human has been or will be. For wildlife ecologists such model of space use, or: 'space utilization', may be useful to answer why an animal spends time at a certain location, e.g. searching for food, sleeping, mating, etc., and hence to build behavioural models or outline 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 model allows to gain information on daily travel patterns and to provide personalized (push & pull) services to a smartphone user (Steiniger et al. 2006), such as information on traffic conditions between home and work place, or advertisements on specials by shops on the daily commuting route.

A *Utilization distribution* (UD) can be derived from occurrence data, i.e. sightings or tracking information, and can be used to describe with which intensity an animal or human may use a certain geographical location within the environment it is living in. In wildlife ecology UDs 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 2d probability grid of regular spaced locations x,y which describes the probability (grid value) of an animal being in any part of its *home range* (Powell 2000). The term "home range" is used here as in the definition of Burt (1943) as *"[... 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 the mentioned occasional sallies, only a certain probability of use is considered to derive the home range. For instance in the literature the home range is often considered to be the area contained within the 95% contour (i.e. isopleth) calculated from the probability/UD grid (Powell 2000). However, Kie et al. (2010) point out that home range and utilization distribution are often used interchangeably despite subtle differences in the definition.

Besides the use of KDE to derive utilization distributions for home range analysis (Worton 1989, Seaman and Powell 1996) other methods to generate UDs have been developed comparably recently. These are 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 has been 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 (yet). Other frequently used approaches to 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, are able to generate home ranges described by one or several polygons, but are not able to generate utilization distributions. However, density grids can be derived for such approaches too as Getz et al. (2007) show for the local convex hulls and own experiments show for buffer-based approaches.

It is important to know that KDE, MCP, LoCoH and parametric approaches have been designed with respect to and applied to location datasets derived with traditional observation methods, such as VHF telemetry and sightings. These point datasets contain often only small numbers of observation points (50-300) and additional information, such as the exact time of recording, the direction of travel and velocity of the moving animal, may not be available or is sparse. The advent of GPS collars for animals that allow to record such information and that deliver point datasets containing up to several thousands of points per recording period requires that at least two actions should (and have been) taken: First, the existing estimators need to be tested with these new datasets, i.e. tested if assumptions are still met and if they can handle the amount of data (but see Hemson et al. 2005, Börger et al. 2006, Huck et al. 2008). Second, new estimators should be developed that allow to utilize additionally available information, for instance by formulating new assumptions.

Two recently developed utilization density estimators that consider information on recording time are the previously mentioned Brownian Bridge approach (Horne et al. 2007) and the Geo-Ellipse approach (Downs 2010). 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 been tested already in a few works and showed advantages (Horne et al. 2007) but also weaknesses; for the latter see for instance Huck et al. 2008 and our tests in Section 4. Although the BB approach is very promising problematic from our perspective is that it is a probabilistic model for uncertainty modeling, describing the probable location of a particle in the Brownian motion model. Hence, it is not designed to model space utilization. A consequence is that the expected probability (/density) value for the bridge is zero at the location of GPS points (see Figure 2 later). Despite Bullards

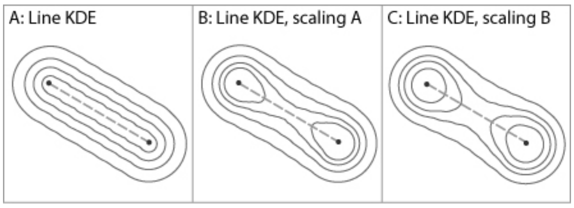

Fig. 1 Contours of 25, 50, 75, 95 percent probability of utilization for line-KDE and scaled line-KDE for a single line segment - scaling A: $sf \in$ [sqrt(0.5), 1.0] and B: $sf \in [0.5, 1.0]$. For a description of the scaling variants see Section 3.3.

(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 should be used to model space utilization or probability of use.

The aim of this article is to present a third estimator for GPS datasets. Thereby our estimator for utilization distributions will be built on the point-based kernel density estimator. The estimator should resemble a similar idea as the Brownian bridge and geo-ellipse estimators – i.e. to use not a single point but a GPS-track segment as base-object. But we will divert from the uncertainty model of the BB. After introducing the design specifications 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 design for certain data and applications in mind, in particular for GPS-tracking data and animal home range analysis. All in all four different sets of criteria where identified:

A - GPS track vs. (Road) Networks – The estimator should model densities for tracking data of the following type: GPS points will be connected to form a chain, whereby the sequence of the points is ordered according to the time of recording. Under- and overpasses of that chain can exists, but the chain will not contain junctions at which the track divides up into three 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 application case. Further, Okabe et al. (2009) also define that the network is not directional. In contrast, the GPS track data are - and can be assumed to be – directional, which is also utilized in the estimator design. Based on these specifications we developed a small reference dataset shown in Figure (2). With respect to our data and this reference dataset we define that: (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, and (iv) if two lines-segments within the bandwidth (i.e. kernel size) run parallel to each other (no matter what direction both have), then density values are added-up as well.

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 to identify/model travel corridors (Bennett 2003) as they are in our opinion a part of the home range. 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 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) middle 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 buffer shape that is obtained for buffering a straight line in a GIS (see Figure 1).

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 was for instance done in the Brownian bridge application by Horne et al. (2007) and the Geo-Ellipse approach by Downs (2010). A promising idea for parameterization is for instance the determination and use of the average distance traveled by the animal per day.

3. A Scaled Line-based Kernel Density Estimator

In the following we will describe the basic approach that we developed for the calculation of utilization densities for 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 known 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 the parameter selection (specification *D*).

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):

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

with n the number of randomly sampled observations $\mathrm{X}_{\mathrm{i}}...\mathrm{X}_{\mathrm{n}}$, the bandwidth parameter h, and the kernel function K. Such density estimate can be considered a continuous form of binning data as done in histogram calculations. To obtain a 2-d kernel density estimate for a point dataset in praxis a 2-d kernel function K of a particular shape and with certain properties is chosen (e.g. a Gaussian bell-shaped 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 2-d 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 $\hat{f}(\mathrm{x}_j)$ is calculated with respect to the distance d_{ij} to the observation X_i (i.e. $d_{ji} = \mathrm{x}_j - \mathrm{X}_i$); with weights previously assigned to each X_i. If a raster cell contains already 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 in such way that each line segment $l_{i,i+1}$ (with $i = 1...n$ GPS points p_i) of the GPS track is split-up 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 that density values are added along each single line segment $l_{i,i+1}$, and subsequently density values would reach a maximum in the geometric centre of the line, the maximum density value between the existing cell value $\hat{f}(x_j)$ and the 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 options to apply this line-KDE approach to a GPS track:

- *rasterize-first* - the complete track with segments $l_{1...n}$ is rasterized and the point-KDE is calculated; or
- *segment-wise* - each segment $l_{i,,i+1}$ is rasterized 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 needs to 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 add-up (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 seems 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 probability contours can be derived.

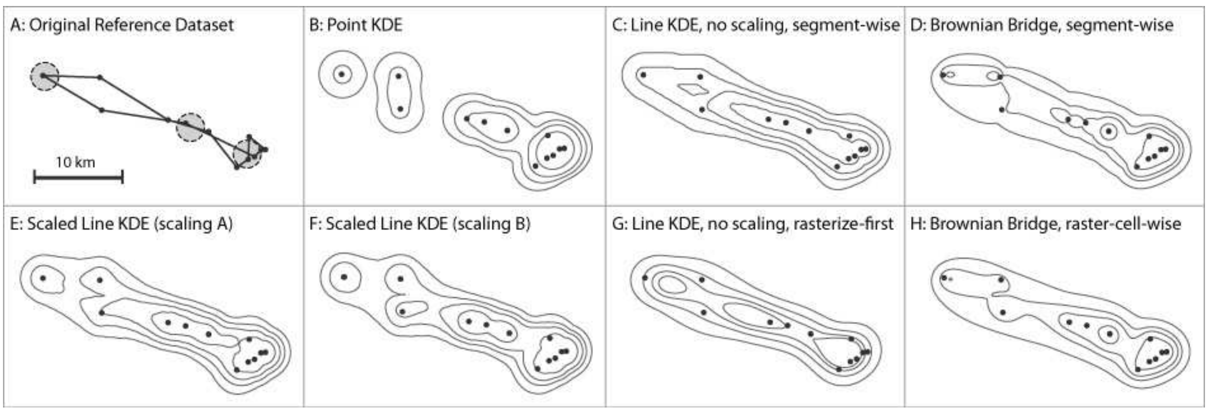

Fig. 2 Reference point dataset and 25, 50, 75 and 95 percent contours for probability of utilization with point–KDE, line-KDE (rasterize-first and segment version), scaled line-KDE (scaling A and B), and Brownian Bridge. For a description of the line-KDE versions and scaling variants see Sections 3.2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

3.2 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 better models the probability of space utilization. Contours generated from such 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):

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

with σ the initial error (e.g. from digitizing a line), l_i the position (distance) of the rasterization between the line segment end points, and l the length of the line segment. The error σ_{P_i} can be used as scaling function *sf*. When setting $\sigma = 1.0$ and defining $sf_{x_i}^A = \sqrt{\sigma_{x_i}}$, then we obtain a scaling range $sf^A \in \left[\sqrt{0.5}, 1\right]$. 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 values *sf* are multiplied with the calculated density value $\hat{f}(\mathrm{x}_j)$ for each rasterized point $p_r^{temp(i,i+1)}$ of the line segment $l_{i,i+1}$, i.e. $\hat{f}(\mathrm{x}_j) = \hat{f}(\mathrm{x}_j) \cdot sf(p_r^{temp(i,i+1)})$.

Currently the 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 new and old density value of a raster cell should be added, or if the maximum of new and old value should be taken. It is easy to see from the contours in Figure 1 that the scaling for sf^A is less strong 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 as of minor importance in comparison with (iii) the choice of the KDE bandwidth parameter *h*. That changing the kernel function has only small effects has been pointed out by Silverman (1986, pgs. 43 and 86) and was also apparent in own experiments that are to be described elsewhere.

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 mini 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 for instance to derive the average travel distance of the animal per day. Hence, in our experiments we used the median value of all average daily travel distances (of one GPS track) as bandwidth parameter h_{mdt} besides h_{ref}.

3.4 Implementation

We implemented the scaled line-KDE approach in the free and open source GIS OpenJUMP (Steiniger and Hay 2009). The Sextante toolbox that can be coupled with OpenJUMP (Olaya 2008) delivered the basic point-KDE and raster processing functionality. The current implementation contains only one kernel function so far, which is 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 HRE, which contains the scaled line-based KDE and other home range analysis methods (Steiniger and Hunter, unpublished) will be available from:
http://sourceforge.net/userapps/mediawiki/mentaer/index.php?title=Movement_Analysis.

4. Comparison with other Density Estimators

4.1 Estimators

To evaluate the performance of the developed line-based KDE we compared the estimator with point-based KDE and the Brownian Bridge (BB) estimator. We chose these two as they can be seen 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 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 GPS location error. We note that for the BB method and its 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 parameter σ_1 to be estimated may be faulty (to be seen by a graph with jumps), and problems in the calculations may lead to different results than for the mathematical model.

For 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 the area for the x percent contour (with $x \in [2.5\%, 100\%]$ in 2.5% steps) and is outlined in Harris et al. 1990. The criterion that determines the core probability is taken from the approach presented in 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 is different from Harris et al. 1990 in that it is cell-based and not contour-based. In our experiments both methods gave often different results.

4.2 Datasets

Besides the earlier mentioned artificial reference dataset that is used to control if our design criteria have been met we used also two real datasets. In particular we obtained GPS collar data from the Alberta Foothills Research Institute from two grizzly bears. Bear A was female and lives in the Alberta Foothills of the Rocky Mountains, whereas bear B was male and lives in the Rocky Mountains. Subsequently travel paths and sizes of the home ranges should be different. In particular the mountain grizzly will travel in valleys - and corridors that connect those valleys may be important for a behavioural analysis.

The dataset of bear A contained 2215 points and has been recorded over 139 days from June to November in the year 2007. The dataset of bear B contained 1525 GPS locations that have been recorded over 194 days from April to November 2001. Both datasets include the time of recording of the GPS locations and therefore the point locations can be ordered. For the Brownian Bridge method the time information needed to be transformed into seconds of the year.

4.3 Results

The comparison was performed more or less qualitative and visually, focussing on home range shape features, such as (i) the shape complexity (patches, holes, smoothness of the outline) and (ii) area of the home range polygon. We also controlled if (iii) travel corridors are observed, and (iv) measured calculation times (Table 1). The results for the reference dataset are displayed in Figure 2. The two Figures 3 and 4 show the results for the grizzly bear datasets. For those we display 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 additionally display in grey unvegetated areas that reflect steep mountain terrain. This allows to assess if the bear traveled in the 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 with the Brownian Bridge (BB) approach are less smooth, or conversely, most detailed. Less detailed outlines than the BB method produce the point-based and line-based KDE approaches. For the line-based approaches the bandwidth parameter, i.e. h_{ref} or h_{mtd}, determines the level of smoothing of the home range contours. For grizzly bear A the BB results seems to be too detailed, which is especially visible in the convolution of the core

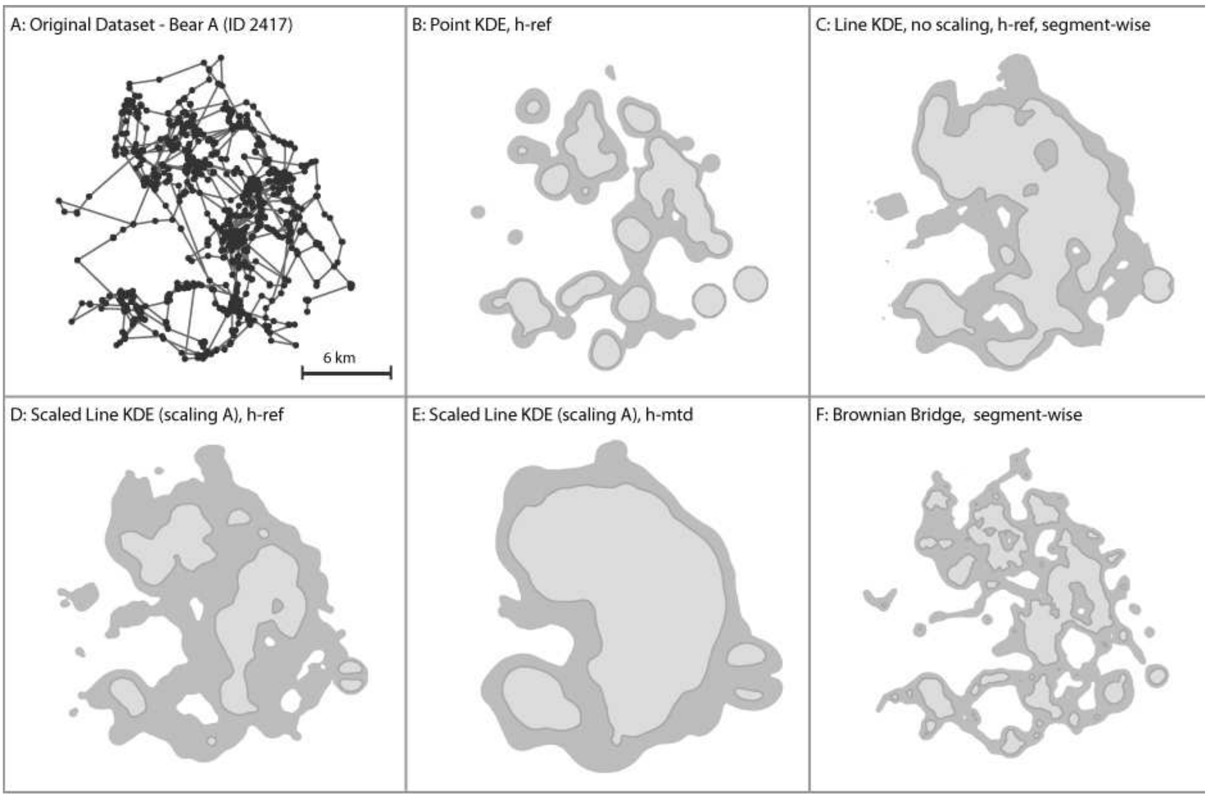

Fig. 3 The results for different utilization distribution estimators, visualized by the 95% probability of use area (dark grey) and the core area (light grey) for bear A. Parameters: KDE h_{ref}=1240m, h_{mtd}=2600m; Brownian Bridge σ_1=8.30m, σ_2= 30.0m.

contours and the number of patches and holes for the 95% contour (Table 1). The point-KDE approach looks slightly spotteds 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 for further assessment could not be utilized since we derived a core probability of 97.7% which was larger than the chosen 95% probability to delineate the home range. However, also the line-KDE with h_{mtd} seems to adapt well to the mountain ranges, which is a result of the smaller bandwidth in comparison to the other KDE parameterizations with h_{ref}. Interestingly the results for the line-KDE with and without scaling are not much different for these datasets. If one compares only the 95% contour for bear A, and not the core contour, then the differences are neglectable. Hence, the scaling has only little influence for both datasets, however, it affects the smoothness of the outlines and the calculation of the core area.

Modelling travel corridors – For all three datasets it is too see that the point-KDE does not observe travel corridors. In contrast the line-KDE models travel corridors fairly well, as it is to see best in the dataset of bear B (Figure 4). However, 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 are calculated. Similar to the line-KDE the BB approach is able to model travel corridors (Figures 3 and 4). But here as well only parts of the corridors may be modelled. This is observable in the results for the mountain bear B where three small polygons indicate a possible corridor between the two main home range patches. We additionally note here that the effect of the zero density values at GPS locations for the BB method is in particular visible on single track segments that may be corridors. The contour lines in Figure 2-D indicate such zero density values, a result of the underlying uncertainty concept.

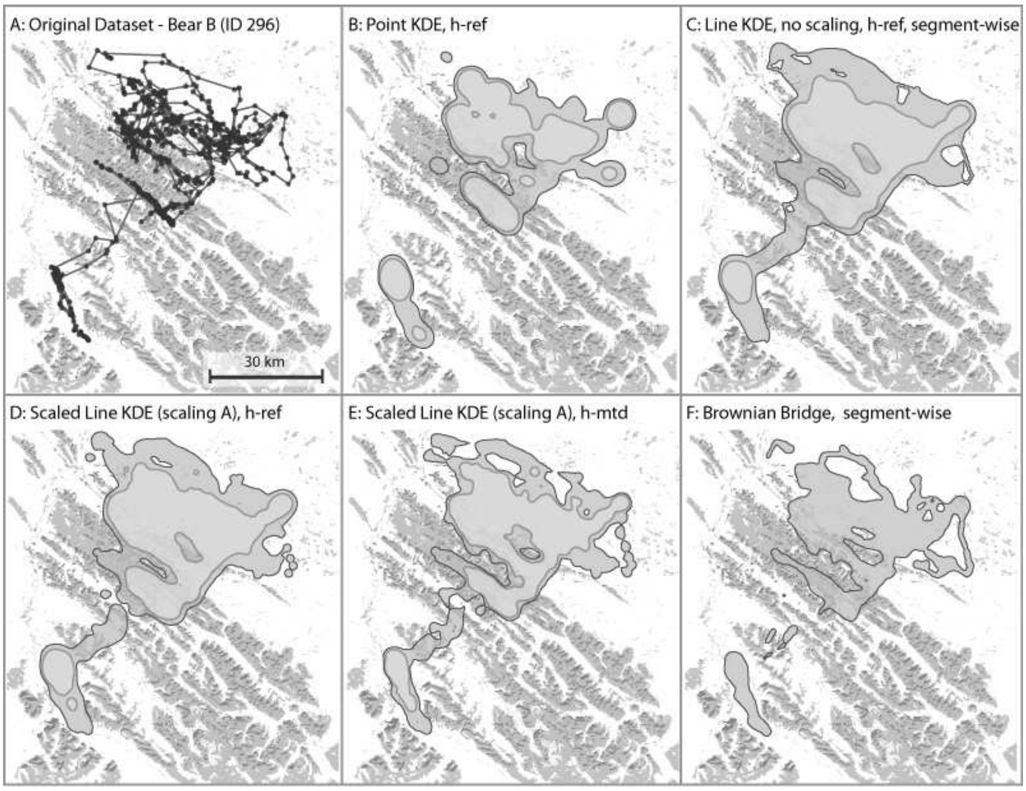

Fig. 4 The results for different utilization distribution estimators, visualized by the 95% probability of use area (dark grey) and the core area (light grey) for bear B. The background shows unvegetated areas, indicating steep mountain slopes. For the Brownian Bridge no core contour is shown, since the calculated utilization probability for the core was 95% as well. Parameters: KDE h_{ref}=4900m, h_{mtd}=3250m; Brownian Bridge σ_1=19.05m, σ_2= 30.0m.

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 has been always larger than the area for the point-KDE (Table 1). Although the BB approach partly models corridors too 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 – Remarkable are the differences in calculation times for the three utilization density estimators. Clearly the fastest approach is the point-KDE taking at average only 1-2 seconds to produce a grid for 1500-2000 GPS locations in our tests (Table 1). The segment-wise line-KDE approaches had running times between 10 to 20 minutes. In contrast the rasterize-first line-KDE was also comparably fast with 3 seconds (data not shown). But we remark that the rasterize-first version does not allow for scaling and that density centers may be moved 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. 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.

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}=1240m	1 sec	80%	142	9/1	No
Line-KDE(s[2])	h_{ref}=1240m	11min	75%	230	7/6	Yes
S-LKDE A[3]	h_{ref}=1240m	13min	72.5%	229	6/6	Yes
S-LKDE A	h_{mtd}=2600m	13min	77.5%	292	1/0	Yes
BB (s[2])	σ_1=8.3m, σ_2=30m	21min	70%	164	9/14	Yes
			Bear B[1]			
Point-KDE	h_{ref}=4900m	1 sec	82.5%	1416	4/1	No
Line-KDE(s[2])	h_{ref}=4900m	11min	75%	2000	1/9	Yes
S-LKDE A[3]	h_{ref}=4900m	14min	77.5%	1974	6/2	Yes
S-LKDE A	h_{mtd}=3250m	17min	77.5%	1670	6/4	Yes
BB (s[2])	σ_1=19.05m, σ_2=30m	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 where 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.

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, and subsequently generate the expected home range estimates. In particular (i) travel-corridors are now included in the home range derivation, a deficit of point-based KDE, and (ii) GPS locations don't receive zero utilization density values, a deficit of the Brownian Bridge model. However, we also noticed that introducing a scaling of the kernel along the line/track-segment had only little effects with our two bear datasets when considering the 95% contour.

Besides the presentation of a new algorithm we also proposed a new option to determine the bandwidth h_{mtd} of the kernel, which is calculated as the median of the average travel distance per day. The experiments on two real datasets show that this value can be smaller or larger than h_{ref}. Based on our limited tests we currently recommend to use 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 did only a limited set of experiments to show that the algorithm works it will be the next stage 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 types of animals than grizzly bears, since different modes of movement will produce different types of GPS point location patterns. Finally a further need, besides the determination of the application range of the algorithm, is to find ways to optimize the processing time. Because waiting 20 minutes for each calculation to be finished will discourage experimenting with the parameter settings to gain better 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

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(2), 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(5-6), 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(4), 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(2-3), 97-123.
Horne, S.J., Garton, E.O., Krone, S.M., Lewis, J.S., 2007. Analyzing animal movements using Brownian Bridges. Ecology 88(9), 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(1), 290-298.
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(1), 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, pp. 32-45.
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(7), 2075-2085.
Sheather, S.J., 2004. Density estimation. Stat. Sci. 19(4), 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(2), 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(4), 183-195.
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.

Worton, B.J., 1989. Kernel methods for estimating the utilization distribution in home-range studies. Ecology 70(1):164-168.
Worton, B.J., 1995. Using Monte Carlo simulation to evaluate kernel-based home range estimators. J. Wildl. Manage. 59(4), 794-800.