Analysing spatial data of different accuracy: the case of greater horseshoe bats foraging

Fabio Bontadina[1,2] and Beat Naef-Daenzer[2]

[1] SWILD – Urban Ecology & Wildlife Research, Wuhrstrasse 12, CH – 8003 Zurich, Switzerland

[2] Swiss Ornithological Institute, CH – 6204 Sempach, Switzerland

Keywords

Rhinolophus ferrumequinum, utilisation distribution, radio-location error, radio-tracking, kernel-estimation, accuracy, foraging

Bontadina, F. & Naef-Daenzer, B. (in press): Analysing spatial data of different accuracy: the case of greater horseshoe bats foraging. *In* Le Maho, Y. (ed.). Wildlife – Biotelemetry. CNRS-CEPE, Strasbourg.

Abstract

Studying the habitat use of highly mobile, fast moving animal species by radio-tracking is particularly difficult. Often only a part of the locations reach the high spatial accuracy aimed for, especially where parts of the study area are not accessible to the observers, or if the accuracy of location is activity-dependent. As a result, excluding data of lower precision may result in biased information on home-range size and use.

We present a method to analyse sets of location data of different accuracy by kernel estimation. Each location is included into the estimation according to its class of accuracy.

For this purpose, the accuracy of locations has to be estimated in the field using an estimation of signal quality and topographic circumstances. In the given example three categories are defined on the base of test bearings with known position of the transmitter. Test-bearings are taken under conditions simulating those encountered during the field survey. The average deviation of the radiolocation from the real position is estimated for the three categories of data. In the kernel procedure, the smoothing parameter of each location is denoted by the standard deviation of the bearing error within each accuracy category. In this way the resulting density matrix accounts for differences in location accuracy. Calculations are made with a module of the software GRID.

The option to join together data of different accuracy improves the results of home-range analyses. To include locations of low accuracy allows areas of high utilisation density to be detected if large samples are available. As an illustration of the method, we present an example from a radio-tracking study in greater horseshoe bats (*Rhinolophus ferrum-equinum***).**

Introduction

Studying the range use of animals by radio-tracking is particularly difficult if the species is highly mobile and the observers cannot follow the tagged animal throughout the home-range. Very often homing-in on the signal (White & Garrott, 1990) to obtain visual observations is not possible. Thus the collection of data has to be based on the cross-triangulation of signal bearings. As a result the accuracy of locations may differ between parts of the study area (e.g. due to topographical features) or depending on the animal's behaviour (e.g. moving or resting). In topographically difficult areas, fixed stations for radio-tracking often give unreliable results. Therefore the locations have to be collected by two persons taking bearings from different places. Due to local conditions, the distance of the transmitter from the observers, and the behaviour of the animal, the estimated locations obtained by cross-triangulation often differ from the unknown true position of the tracked animal.

A measure of this error must be reported and included in the analysis, as pointed out by Saltz and White (1990) and Saltz (1995). Often, only parts of the spatial data reach the high level of accuracy aimed for. On the other hand additional data of lower accuracy may provide useful information. Up to now the usual way of dealing with data of different quality was to exclude locations exceeding a maximum tolerance, and to treat all remaining locations as if they were of the same quality. This procedure, however, has the disadvantage that a part of the information obtained from very accurate locations is reduced.

When radio-tracking greater horseshoe bats (_Rhinolophus ferrum-equinum,_ **Schreber 1774), these problems make it difficult to assess home range use and selection of foraging sites. We studied the species in an alpine valley where large parts of the study area were not accessible to observers (particularly at night). Therefore, the accuracy of locations varied largely depending on site and activity of the animals.**

In order to make optimal use of the information contained in the data we developed a method to include data of different location accuracy into a

kernel estimation. Herein we discuss the advantages and limitations of this approach using simulations and field data.

Materials and Methods

Home-range estimators

The following method to analyse spatial data with different accuracy is based on a kernel estimation (Bowman 1985, Worton 1989). Kernel procedures allow patterns of locations to be transformed into a matrix of utilisation distribution. This utilisation distribution is a two-dimensional (relative) probability distribution of the estimated locations (overview in Silverman 1986, Worton 1989, Naef-Daenzer 1993a, b).

The transformation of a pattern of locations (points) into an utilisation density distribution (values) can be described as follows: For each intersection of a superimposed grid, the utilisation density is calculated on the base of the distance from the grid intersection to the locations. The kernel-function and the smoothing parameter h determine the extent to which a location at distance x contributes to the estimate. In the bivariate normal kernel estimation, locations are weighed according to a normal distribution of defined variance. In a three-dimensional graph of the density matrix, one single location is represented as a small Gauss-bump, the width of which is determined by the smoothing parameter (Silverman 1986, Worton 1989, 1995). The final utilisation distribution can be considered as the 'sum' of the bumps of all locations.

Based on the kernel density estimations, contour lines which include a percentage of the total volume (e.g. 20%, 80%) are calculated and used to identify core areas and total home-range size. On one hand the result of home range estimations depends on the sample size. On the other hand the kernel algorithm and the smoothing parameter greatly affect the results of the estimation. In particular the estimate of the total home range is not independent of the selection of the kernel parameters.

Locational observations of different accuracy provide the same information unit, but the information is spread over a different area. The mean distance between estimated and actual position describes the accu-

racy of the observation (Saltz 1995). If there is evidence that the deviations from the 'true' location (of the tracked animal) vary randomly, the location error is expected to be bivariate normally distributed and is described by the standard deviation (SD) of the distribution.

In order to combine data of different accuracy we have to include the error measure of the estimated locations into the smoothing parameter. This adaptation takes into account that the distribution of each information unit is spread over different areas depending on the accuracy. As the information of one Gauss-bump always represents one observation of the animal, the maximum density (the top of the bump) of one location of low accuracy is lower than that of a location with high accuracy, whereas the volume covered by the different bumps is constant (fig. 1) .

[fig. 1]

If the accuracy of each data point is known, a kernel algorithm can account for the different location accuracy by using the location error as a variable to adapt the smoothing parameter for each location within the process. The calculations are made with a module of the software GRID (Naef-Daenzer 1993a). Core areas and home range areas were computed as percentage of total density volume. The contour lines are isoplets and represent similar density of utilisation.

<u>Study area and data collection</u>

25 greater horseshoe bats (*Rhinolophus ferrumequinum*) were radio-tagged and tracked in an alpine valley in the Grisons, Switzerland, from May to October, 1993 (Bontadina et al. 1995, 1997, unpubl. data). Greater horseshoe bats emerge at dusk and forage for some hours before they return to night roosts. Flying rapidly, they reach their foraging areas at distances of up to 7 km from their day roost. The position sensing transmitters (Holohil Ltd, Canada) allowed us to determine whether bats were flying or hanging. When aerial hawking the bats moved rapidly from one place to another over a large area. However, when perch hunting, the bats stayed at the same place for longer periods.

We tracked the bats using TRX-1000S (Wildlife Materials, USA) and modified YEASU FT-290 receivers (adapted by Karl Wagener, Germany) with hand-held H-aerials. The location of the tagged bats were

recorded in 5 minute intervals throughout the night by triangulation of the signal direction. Two field workers co-ordinated their simultaneous bearings using trigger signals from Casio DB-31 watches. Hand-held FM-radios were used to remain in contact with one another.

A category of accuracy was assigned to each location. In a first step, three categories of accuracy, notably "high", "medium" or "low", were used to categorise the locations in the field. These represented situations where estimated locations were supposed to lie within 50, 100 and 250m respectively of their true location. This estimation was made on the basis of a evaluation of the transmitter signal, the intensity of the signal and the estimated distance from each observer to the animal, as well as taking into account the possible influences of environmental conditions. In parallel with the data collection a field test was carried out with a transmitter being moved around in a foraging area by a colleague. Transmitters of known position were located and the actual accuracy of the three categories of locations was estimated using these control data.

Results

If the accuracy of each data point is known, a kernel algorithm can account for the different location accuracy by using the location error as a variable to adapt the smoothing parameter for each location within the process.

Simulations

By using a simulated sample of data, figure 2 demonstrates that core areas may be obtained by using either a few locations of high accuracy or a large sample of locations of low accuracy. We used two samples of the same distribution. Firstly, a data pattern of 100 locations (fig. 2, A) was analysed by a kernel estimation with low and high smoothing parameters. The resulting contour lines (B, C) differ in respect to shape and core area. In B, the kernel estimation with smaller smoothing parameter (according to a higher accuracy of the locations) reveals a significantly pronounced resolution and includes a core area, in contrast to C. However, a large sample may partly compensate for the loss in profile, as

demonstrated with the doubled data sample (fig. 2, D). In that case, the contour lines reveal again a core area.

[fig. 2]

When defining contour lines it must be taken into account that one location of high accuracy does not result in the same density pattern as one of low accuracy. Outliers of low accuracy can be excluded from a home range estimate by appropriate selection of contour levels. If the lowest contour line is above the maximum density of a single location, isolated locations are excluded. On the other hand, several inaccurate locations at the same position give the same maximum density as one of high precision. In the example of figure 2 the lowest contour line is the line with relative density values of more than one (larger than the maximum of one location).

For the comparison of areas it is important to notice that the estimated areas differ when using data of different accuracy and therefore different smoothing factors. In our simulation, if a core area is defined at a level of 50% of total density, then the area is 1.7 units, analysed with a smoothing factor of 50. If data of, for example half the accuracy is used and therefore the smoothing factor is set twice as large at 100, the area estimate is only 0.7 (fig. 3). Inversely, if the total home range area is estimated by 95% of total density, the kernel analysis with data of low accuracy (smoothing factor 100) reveals an area of 10.2 in distinction to 6.3 when using the same data but with smoothing factor of 50 (high accuracy).

[fig. 3]

The simulation shows, that areas of high location density are underestimated when using data of lower accuracy. In contrast, total home range size - represented by a contour line of low location density- is overestimated if the proportion of inaccurate locations is high. Therefore estimates of home range or activity areas increase when using data with lower accuracy because the information spreads over a larger area (fig. 3).

Field data

A total of 148 test bearings were available to analyse the accuracy of the three categories of locations, denoted „high“, „medium“ or „low“, respectively. They had their center not significantly different from the “true” center, the standard deviations of the normally distributed location errors were 44, 86 and 162 meters, respectively. The deviation was approximately normally distributed around the actual location with a mean not significantly different from the “true” center.

In the field study of foraging greater horseshoe bats 1331 locations were collected. Of these data only 28% reached the accuracy of the 50m category, another 31% reached an estimated accuracy of 100m and the last 41% were found within the 250m accuracy class.

37% of the locations of bats which were perch hunting, and therefore stayed at one place for a longer period, could be assigned to the highest accuracy category. But only 20% of the locations of individuals aerial hawking could be classified as such (Chi2=45.8, p<0.001). This demonstrates how the accuracy of bearings can depend on behaviour.

In addition, the reached accuracy may depend also on the topographic situation in the field. The distribution of the pooled locations of 7 greater horseshoe bats in figure 4A shows a clear spatial aggregation. The locations at point G are situated in a steep gorge, which was not accessible to the observers. Therefore only locations with low accuracy (grey or white dots) could be taken. In other areas, it was possible to track the bat very accurately, because it did not move much (black dots).

Three core areas result from the calculations, if only the 38 locations of high accuracy are included (fig. 4B). Another core area in the gorge appears, if all 148 locations are analysed according to the lowest accuracy class (fig. 4C), but one former core are disappears. The resolution is lowered as much that the two core areas merge. The resulting larger areas include also areas with low observed utilisation by foraging bats.

As a solution we include all data according to their accuracy in the kernel analysis. This procedure preserves the three former core areas based on the locations with high accuracy. But it also contains the core area in the gorge (point G), including many locations of low accuracy. Despite the fact that more core areas are present, their total area is

smaller than without including the different accuracies in the calculations.
By including all data at the appropriate level of accuracy, the results of the kernel estimation could be improved to a great extent. As a consequence, the spatial result would be biased, if data of lower accuracy is refused.
[fig. 4]

Discussion

The option to include locations of different accuracy into a kernel estimation has the advantage that full information of all gathered locations is used without a loss in detail where accurate locations were possible. Excluding data of lower precision can influence the results substantially as shown in the examples above, especially, when radio-tracked animals stay in topographically difficult areas and the observers of a radio-tracking team cannot approach the animal close enough. Ignoring the data of low quality (the 250m class) would result in a highly incorrect representation of the activity range of horseshoe bats. This in turn might potentially lead to the wrong conclusions in respect of the behavioural and conservation ecology of the species.
Furthermore areas where a particular behaviour is dominant may be underestimated, e.g. it is generally not possible to obtain accurate locations along travelling routes, where the animals move very fast from one place to another. The inclusion of data of low accuracy makes it possible to detect and retain areas over which animals move rapidly.
Certainly, no statistical method can correct for an estimation error inherent in the data set. For this reason the resolution of the data cannot be enhanced. At least, the present method avoids a loss of information when data of high precision must be combined with data of lower accuracy. With regard to the detection of core areas of activity, the proportion of locations which reach a given level of accuracy is of great importance. Low accuracy of single locations lead to a increased variance in their spatial distribution. Thus, aggregations of points appear only if the sample size is large enough. Therefore, in order to represent core areas in

the data set, an increase in the bearing error should be compensated by a large sample size.

In our study the deviations from the true position of the transmitter were approximately normally distributed as seen in the test. This is not the case when locations are taken by triangulation data from fixed stations. In this case the bearing errors depend on the position of the transmitter relative to the antenna stations and are not analogous to a bivariate normal distribution (White and Garrott 1990). In such cases it would be possible to replace the bivariate kernel with another kernel shaped according to the error distribution, for example a polygon or an ellipse.

One problem which has to be solved when using locations of different accuracy is how to choose an adequate smoothing parameter for the kernel estimation. The smoothing parameter determines the shape of the density distribution. Several authors propose approaches to determine the parameter based on the statistical properties of the location set. Worton (1989) proposes a cross validation method and Sain et al. (1994) use a biased cross validation-method for choosing an appropriate smoothing parameter. Tufto et al. (1996) provide a solution to the problem of correcting discretisation errors for some cases where least square cross-validation otherwise would crash. All these methods define the smoothing parameter based on the variance of the points in x and y direction. In the case of an extensive data sample or if the locational error is very small, the cross-validation procedure may propose a smoothing parameter smaller than the achieved SD of the bearings. In this cases the resolution of the analysis can be improved by using the value for the class of bearings with the lowest accuracy. The other classes must than be proportionally fitted.

Depending on the purposes of a study, e.g. the analysis of core areas or the estimation of home ranges, different smoothing parameters may give the best results (Wray et al. 1992). Therefore there does not appear to be a single best way of determining the smoothing parameter (see discussion in Silverman 1986). Although some objectivity in choosing the value of the parameters would be desirable, the best estimation is that which gives the most convincing information on the biological system. Optimal selection of the parameters of a kernel estimation still depends on the aims

of the study. To allow comparisons of area estimates we recommend to use additionally standard parameters.

Acknowledgements

We would like to thank the members of the ASHG, A. Beck, S. Gloor, Th. Hotz, M. Lutz & E. Mühlethaler for the possibility to use data from the study on greater horseshoe bats carried out in a alpine valley of Grisons, Switzerland. Thanks to L. Duvergé, S. Gloor and D. Hegglin for fruitful discussions and for critics on earlier drafts of the manuscript. The study on greater horseshoe bats was generously financially supported by the Swiss Agency for the Environment, Forests and Landscape BUWAL, and the Berthold Suhner-Foundation.

Fig. 1: Three locations of different accuracy are shown in the density matrix. The information of the location of highest accuracy is concentrated on a small area and the maximum density reaches the highest value. The more inaccurate the location is, the larger the area is where an animal could possibly have been present, and the lower is the density it contributes to the analysis.

Fig. 2: The same generated distribution: A) pattern of 100 locations, B) these locations are analysed as having an accuracy of 50 units, C) the same locations analysed as having a low accuracy of 100 units and D) twice as much locations from the same distribution analysed as having a low accuracy. The side of one square equals 1000 units.

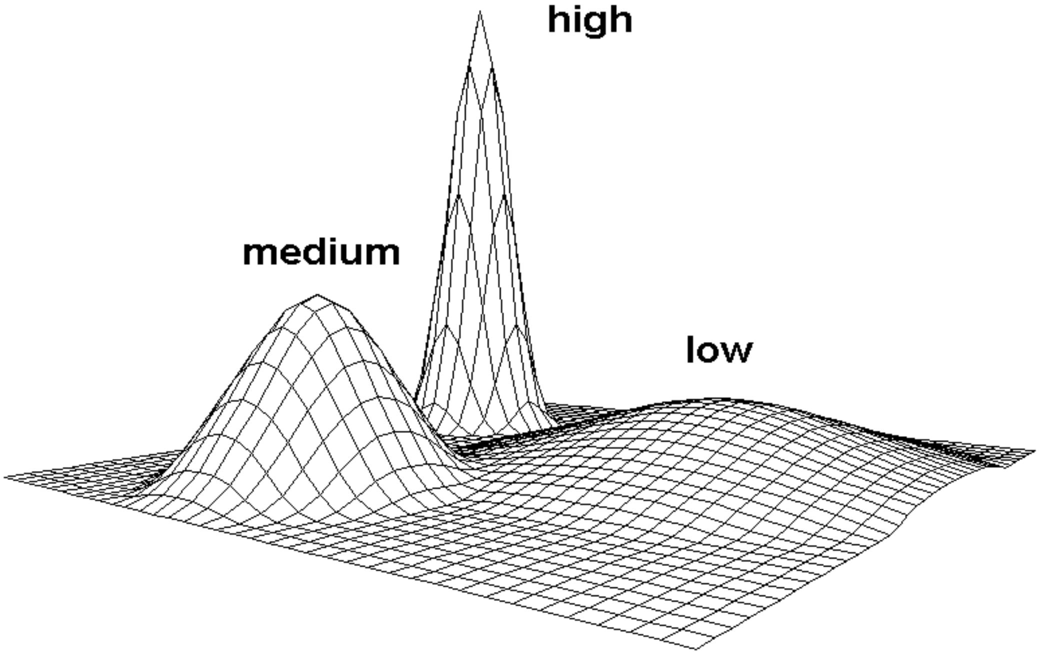
More locations of lower accuracy can compensate for peak areas. However the estimated areas become larger when using data with lower accuracy.

Fig. 3: Influence of accuracy on range estimates. Based on 200 generated points from a normal distribution with standard deviation (SD) of 1 unit, areas were calculated according to be of "high" and "low" accuracy (0.05 and 0.1 unit, respectively). Locations of the accuracy class „high" reveal larger core areas and smaller home range estimates than locations of "low" accuracy.

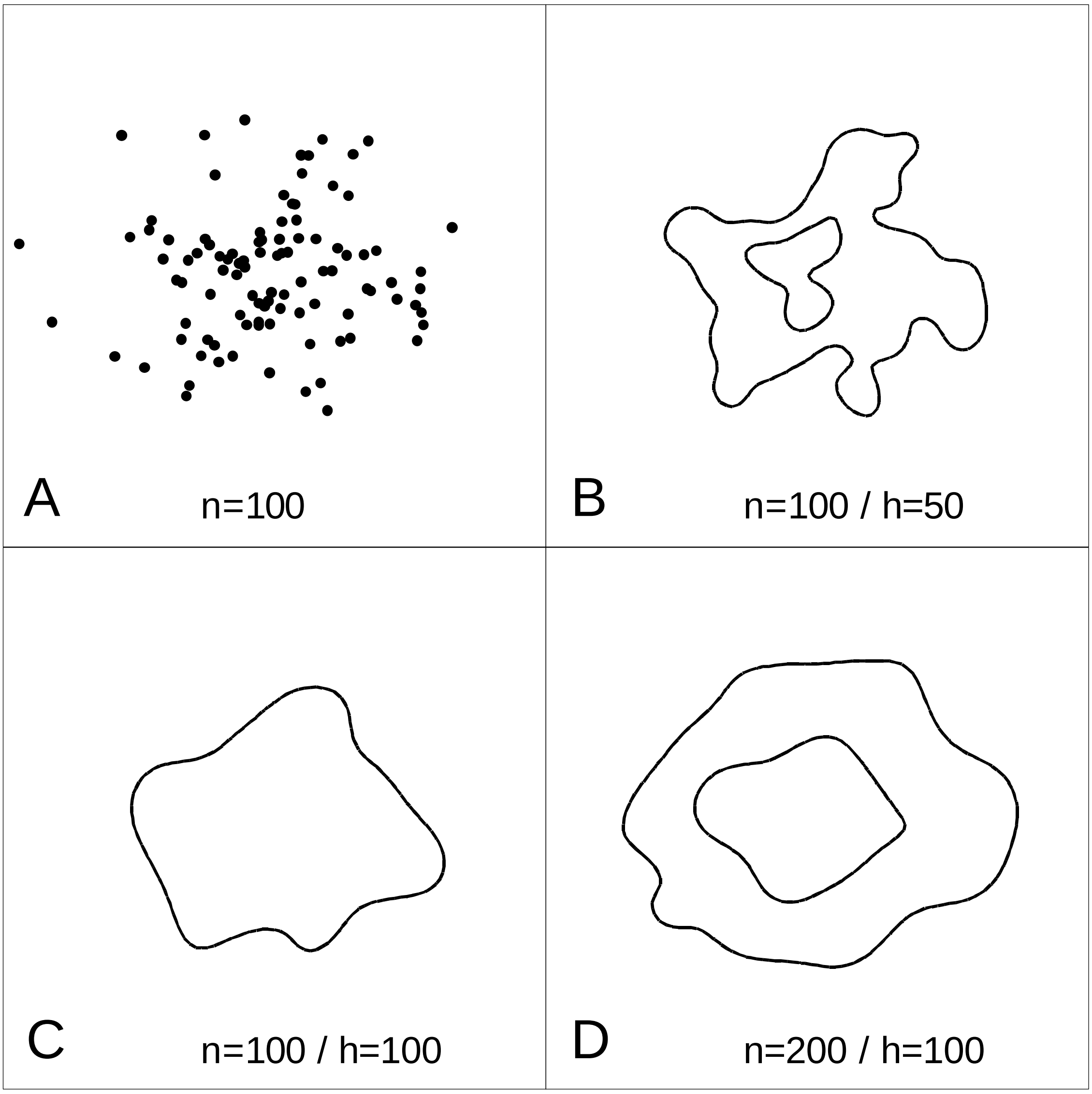
Fig. 4: Home range and core areas of seven greater horseshoe bats analysed by kernel estimations with locations of different accuracy. Contour lines are drawn at 80% and 30% density volume for home range and core areas, respectively. A) Distribution of the estimated locations of seven bats by radio-tracking. Locations estimated with „high" accuracy are marked black, those with „medium" or „low" accuracy are grey and white, respectively. B) Only the 38 locations of the „high" accuracy class are used for the kernel estimation. C) All 148 locations are used according to the lowest accuracy class. D) All locations were included in the kernel analysis according to

their accuracy. This result includes in high resolution all former areas with high utilisation density and the core area in the gorge (point G), where mainly data of low accuracy was available.

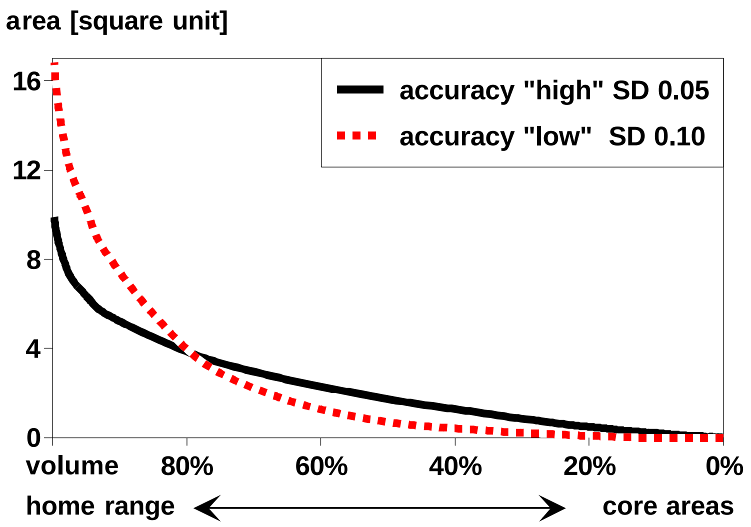
[fig. 1]



[fig. 2]



[fig. 3]


area [square unit]
16
12
8
4
0
accuracy "high" SD 0.05
accuracy "low" SD 0.10
volume
80%
60%
40%
20%
0%
home range
core areas

[fig. 4]