Geostatistical Analysis of GPS Trajectory Data: Space-Time Densities

Tomislav Hengl¹, E. Emiel van Loon, Judy Shamoun-Baranes and Willem Bouten

Computational Geo-Ecology (CGE), Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, Netherlands

Abstract. Creation of density maps and estimation of home range is problematic for observations of animal movement at irregular intervals. We propose a technique to estimate space-time densities by separately modeling animal movement paths and velocities, both as continuous fields. First the length of trajectories for a given grid is derived; then the velocity of individual birds is interpolated using 3D kriging; finally the space-time density is calculated by dividing the density of trajectories (total length of lines per grid cell) by the aggregated velocity at that grid cell. The resulting map shows density of a species in both space and time, expressed in s/m² units. This length-by-velocity (LV) technique is illustrated using two case studies: (1) a synthetically generated dataset using the Lorenz model; and (2) GPS recordings of 14 individual birds of lesser black-backed gull (Larus Fuscus). The proposed technique is compared with kernel smoother – a technique commonly used to derive home range for species. The results of using a synthetic dataset proved that the LV method produces different outputs than kernel smoothing, especially if irregular observation intervals are used. The main advantages of the proposed technique over a kernel smoother are: (1) it is not sensitive to missing observations; (2) it is suited to analyze fly paths (e.g. it preserves information about velocities and directions), and (3) it allows the movement of birds (velocity, trajectory) to be modeled separately e.g. as function of environmental conditions, wind, day time and similar. The remaining research issues are development of methodology for selection of optimal grid size and optimal time interval between recordings.

Keywords: animal movement, 3D kriging, velocity, home range, space-time cube.

1. Introduction

In the last several years, the research group Computational Geo-Ecology has been involved in the development of the Netherlands Bird Avoidance Model with the main aim to predict the 3D + time distribution of birds in the Netherlands. These predictions are used as a decision support tool to improve flight safety in the Royal Netherlands Air Force (Shamoun-Baranes et al. 2007; Van Belle et al. 2007). A focus is now put on observing and modeling birds’ movement and their distribution using data from individually tracked birds.

The technique commonly used to estimate a home range of a species, and based on the tracks of individual animals, is the kernel smoother (Kernohan et al. 2001). The kernel smoother requires only point locations of the animal to generate a 2D probability density function (PDF). This expresses the probability of finding the animal at each location (x,y) in the space domain. So, by integrating this PDF (volume integration) one can define the area where the animal has spent e.g. 90% of its time. Such an area (other percentages can be used) is then called a home range. The unit is Time/Area or sometimes 1/Area if time is defined as a percentage of total time. The unit can also be dimensionless if the home range area is defined as the part of the total area of occurrence. Note also that home ranges derived in various software programs can differ largely, and analysts are often required to objectively determine home-range estimators, search radius, and input values of required parameters (Lawson and Rodgers 1997).

Recently, there has been an increasing interest to visualize and analyze animal trajectory data using geostatistics and various statistical operations implemented in a GIS (Austin et al., 2004). A review of

¹ Corresponding author; Tel.: +31-(0)20-5257379; fax: +31-(0)20-5257431
E-mail address: hengl@science.uva.nl

applications and techniques used to visualize the spatio-temporal observations can be found in Andrienko et al. (2003) and Pelekis et al. (2007). Laube et al. (2007) uses geostatistics to map local sinuosity of 54 homing pigeon trajectories. In their case study, the sampling rate was 1 s, so that even fine changes of the movement could be recorded.

In this paper we propose a novel technique to derive space-time densities based on the GPS trajectory data. Our objective is to accurately quantify areas where a species occurs most frequently and to understand how and why the velocity of birds changes in space and time. In the first case study, we use a synthetic trajectory data set in order to observe how the technique behaves in situations when the sampling in time is irregular. We then implement the technique on a real case study and discuss benefits and limitations of using the proposed technique. All computations are run in the R statistical computing environment using the gstat (geostatistics), adehabitat (habitat suitability and home range analysis) and splanc (point pattern analysis) packages. The processing steps are described in detail in Hengl et al. (2008); a copy of the scripts can be obtained by contacting the authors.

2. Theoretical concepts

The space-time density is a measure of how often a species is recorded in both time and space. Imagine a space-time cube where the geographical location \((x,y)\) represents the basis of the cube and the time is the third, vertical dimension (Fig. 1). The movement of each individual animal can be presented in this space-time cube as an inclining line (Pelekis et al. 2007). If such a line is accurately mapped, we can then derive the length of lines (the \(z\)-axis of the space-time cube) at each grid cell in the geographical space \((\Delta x, \Delta y)\). This is the concept of the space-time density and is closely connected with the concept of home range. It can be shown that time-dimension length can also be derived as the inverse of the velocity between the two point locations. So that the space-time density can be estimated by using only time-length and velocities:

\[
\hat{d}_{xyz}(s_{B0}) = \frac{\hat{d}_L(s_{B0})}{\hat{v}(s_{B0})} \quad \left[ \frac{\text{m}^2}{\text{m/s}} \right] = \left[ \frac{\text{s}}{\text{m}^2} \right]
\]

where \(d_L\) is the total length of trajectory over a specific grid node \((\Delta l / A)\) in \(\text{m/m}^2\), and \(v\) is the velocity of an animal at grid cell of interest \(s_{B0}\) (Fig. 1). The final output maps can be interpreted as follows – high values indicate areas where higher number of animals spend more time; low values indicate areas where there is either less animals or they move faster over the area.

![Space-time density estimation](image)

Fig. 1: Basic concepts of space-time density estimation: point recordings and assumed path (left); the first 150 points of the Lorenz model data set (see Sec. 3.1.) displayed in a space-time cube (right). See text for more explanation.

In practice, space-time densities can be derived by following these six steps:

- **Generate trajectories** from point recordings (sort point recordings by time and convert them to lines);
- **Estimate the suitable grid size** (support size) given the spatial density of observations and the size of the area;
- **Derive total length of trajectories** (total length of lines per grid cell);
Interpolate velocities over the whole grid of interest using 3D kriging and aggregate the values to obtain the average velocity per grid node;

Derive space-time density by dividing the length of lines per area by average velocity (Eq. 1.1)

If the point recordings are sampled using equal intervals (\( \Delta t \)), and if there are no missing observations, then velocity is inversely proportional to density of points:

\[
v(s_{B0}) = \frac{\Delta l(s_{B0})}{t(s_{B0})} = \frac{\Delta l(s_{B0})}{\sum_{j=0}^{m} n_j(s_{B0}) \cdot \Delta t}
\]

and a simple point-density estimation, assuming that the search area is equal to grid cell, will produce the same result as Eq. (1.1):

\[
\hat{d}_{jxy}(s_{B0}) = \frac{d_j(s_{B0}) \cdot \sum_{j=0}^{m} n_j(s_{B0}) \cdot \Delta t}{\Delta l(s_{B0})} = \frac{\Delta l(s_{B0}) \cdot \sum_{j=0}^{m} n_j(s_{B0}) \cdot \Delta t}{A(s_{B0}) \cdot \Delta l(s_{B0})} = \frac{\sum_{j=0}^{m} n_j(s_{B0}) \cdot \Delta t}{A(s_{B0})} \quad \left[ \text{s/m}^2 \right]
\]

The advantage of using Eq. (1.1) instead of Eq. (1.3), however, is that in the case of Eq. (1.1) also irregular observations can be used to estimate the densities. In addition, Eq. (1.1) allows us to model the trajectories and velocities separately, i.e. by using auxiliary information. In ecology, animal velocity is an effect of various environmental conditions (e.g. wind, temperature, season, elevation, obstacles etc.). If we are able to model velocity and actual animal paths more precisely by using auxiliary maps instead of by just running a kernel smoother on the point recordings, then we should be able to produce more accurate maps of space-time densities for a given support size in the space domain. This is further demonstrated using two case studies.

3. Case studies

3.1. Synthetically generated trajectory

We first test our technique\(^1\) using a synthetically generated trajectory data set (shown in Fig. 2). This butterfly-shape path is the partial (2D) output from the system of three differential equations known as the “Lorenz model”. The model was developed by Lorenz in 1963 as a minimal model for thermal convection in the atmosphere. For a detailed discussion of this model see e.g. Hairer et al. (2000). It should be noted that, in the interpretation of the Lorenz model, \(x, y, \) and \(z\) are not spatial coordinates, so that the trajectories in \(xyz\)-space do not correspond to paths in the convecting fluid model. Still, this synthetic dataset is ideal for our study since in this data set velocities are varying a lot over space, while at each time instant velocity is known via the differential equations. The Lorenz model dataset consists of 7201 points sampled at regular interval of four seconds. To each point, four variables are attached: \(x, y, t\) coordinates and velocity.

Because geostatistical analysis requires a discretized geographical space (grid topology), we first need to select a suitable grid resolution that will offer a maximum information content and that will fit the objective of the spatial analysis. Obviously, the finer the resolution the higher the accuracy of depicting fine paths. On the other hand, if the size of the grid is too fine, then there are not enough observations in all grids – the output of analysis will become noisy – and this will make it harder to visually interpret the image. Recall that we need to have more points falling within the same grid node, so that we can accurately determine the length of path and density of points. To satisfy both requirements, we propose the following geometric principle: we first derive the distances to nearest neighbour and then select the third quantile as the grid size. This means that for about 3/4 of the dataset there will be at least two point recordings within a grid node (Fig. 1). Following this empirical rule, we estimate a suitable grid size for this synthetic trajectory to be 0.34 units, which gives 109 by 145 grids (or 15,805 in total).

\(^1\) In further text, we will refer to it as the **Length-by-Velocity** (LV) technique.
Once we have estimated the grid of interest, we can derive the length of trajectory lines within each grid node, e.g. by using the segment density command in the ILWIS GIS (Fig. 2b). This shows that the animal visits certain paths more frequently than others and moves faster over certain areas than others. We further proceed with interpolating the velocities. Note that there could be a difference between flight velocity over the same grid node, but at different times. For example, a bird could be flying over the same grid node with the direction of wind or against it. By ignoring this aspect of animal movement, we would obtain a variogram with high nugget variation at zero (geographical) distances. Two velocities over the same grid node with opposite vectors are, in fact, most often fairly distant in the time-domain (Fig. 1, right). To avoid this problem, we use the 3D kriging method as implemented in the gstat package. This will
produce a range of maps (time-slices) for different time periods, so that the position in the time domain is also taken into account. After interpolating velocities for various time periods, the output maps can be aggregated over the spatial domain. The final map of average velocities for the Lorenz model shows that the object moves faster at straight paths and at the outer edge of the flying path (Fig. 2c).

Fig. 3 shows a comparison of the space-time densities derived using LV technique and kernel smoothing, first for the whole data set, and then by randomly taking out 50% of observations. The loss of spatial detail (anisotropic shapes, curves etc.), when kernel smoother is used, is evident. This happens even if the sampling density is relatively high (Fig. 3a). On the other hand, relatively long sampling interval causes inner circles to disappear; mainly because the flight paths are not estimated accurately. A limitation of the kernel smoother is that the extracted densities are proportional to the sampling interval, so that the density maps in Fig. 3(a) and Fig. 3(b) cannot be compared directly. In the case of deriving densities based on LV technique, the density values should always be at the same scale because we estimate the lengths of trajectories and the velocity as continuous fields. The densities for lower sampling density will, of course, in average be somewhat lower, because the estimated total length of lines will gradually decrease (underestimation of the trajectory lengths).

The final statistical comparison of the densities derived using two methods shows that they do produce similar images – correlation coefficient between the two maps in Fig. 3(a) is 0.82 (number of grid cells = 15,805). The kernel smoother seems to have more problems in depicting anisotropic movement patterns; estimation of paths can be difficult if the animal changes direction frequently. Differences between the two techniques grow \((r=0.79)\) with more irregular sampling intervals (Fig. 3b) and geometrically anisotropic movements.

### 3.2. GPS recordings of gulls

We now turn to real (i.e. non-artificial) data – the GPS recordings of the lesser black backed gull \((Larus fuscus)^2\), a migratory gull species. A total of 14 individual birds were selected. To each bird, a GPS receiver and ARGOS transmitter have been attached, and the GPS readings of position were recorded. For majority of birds, altitude and velocity were automatically recorded as well. Unfortunately, the velocity was only recorded by part of the sensors (with many erroneous values). Moreover, this velocity is instantaneous velocity: not an accurate measurement of the average velocity over a 2 hour time-span at all. The birds were released between 25 May and 15 June 2007 at their breeding colony in Vlieland. For this studies data collected until 24th of October 2007 was used.

The objective of this tracking study is to monitor the movement of gulls in the airspace of interest to cooperating air forces for several months and try to answer questions such as: where do gulls spend most of time? what is their home range? how do they travel between core areas? are there clear travel routes? are the routes used by just one individual or the sampled population? where do they forage and rest? are they active at night? does wind and environmental conditions (urban areas, landscape) influence their movement (Shamoun-Baranes et al. 2007)? how do we distinguish between foraging/exploratory movement and migration etc.

The dataset consists of a total of 9234 readings. To be able to calculate with the time-dimension, the time readings were first converted to the cumulative hours from the beginning of year 1970. There were many missing GPS velocity readings so we decided to also estimate the velocity between the two GPS readings based on the difference in locations and time. For the gulls, we can only sensibly use the distance between locations to estimate velocity, and with these large intervals it may be more sensible to call these estimates ‘displacements’ rather than vector ‘velocity’. Note also that the LBG species can cross large areas of Europe within few days (Fig. 4a).

We proceed with using environmental predictors to detect areas of potential distribution of a species, i.e. the Habitat Suitability Index (HSI). This map can be derived by running Ecological Niche Factor Analysis as implemented in the adehabitat package. The resulting map (Fig. 4b) shows that of the distribution of LBG birds can be closely connected with the environmental conditions: the birds seem to systematically avoid mountain chains and big urban areas. We proceed with interpolating the velocities over the whole area of interest. The HSI map shows not much benefit for interpolation of velocities as it explains only 10% of variability.

---

2 Further in text referred to as LBG species.
This means that the velocity at point locations is rather poorly estimated, mainly because the time interval between observations is relatively long (>2 hours). In addition, the distribution of velocities is heavily skewed – most of the birds spend longer time at one location (the breeding season) and then move very fast while migrating over sea. This also reflects on the variogram, which shows rather high nugget (58% of sill variation). Consequently, kriging heavily smoothes local high values of velocities. We were still able to interpolate velocities for six time slices, and then aggregate them to produce a map of mean velocities over the whole period. This shows that, as expected, birds move faster over ocean and areas of low habitat suitability (see also Hengl, et al. 2008). Note that 3D kriging with a real case study is much more computationally demanding and can take up to several hours on a standard PC.

The final map showing the space-time densities can be seen in Fig. 4c. If you compare this map with the result of running a kernel smoother (Fig. 4d), you will notice that several conclusions discussed previously with the synthetic case study also apply to real case studies: the kernel smoother seems to be less accurate in depicting actual flying paths; the LV technique offer more possibilities to include various explanatory variables to explain behavior of animals than kernel smoothing.

4. Discussion and conclusions

In this article, we have described and tested a novel technique to map space-time densities by combining
GIS-based spatial analysis and spatio-temporal geostatistics. The preliminary results show that the LV technique is suited for analysis of trajectory data and can be used to map spatio-temporal densities of species. The results of using a synthetic dataset have shown that separate modelling of velocities and paths (as continuous fields) is promising for several reasons: (1) spatial detail obtained is higher than with kernel smoothing; (2) unlike kernel smoother, the proposed technique is more ‘immune’ on the sampling interval; (3) the result of analysis are absolute estimates of the density, which allows us to compare outputs of analysis for different sampling intervals (not possible with kernel smoother).

The real case study (lesser black backed gull) confirms that the proposed technique offers more possibilities for the spatio-temporal analysis of animal movement than kernel smoothing – by separating modelling of velocities and paths, we are able to include various auxiliary maps, e.g. environmental conditions, topography and proximity of urban areas, so that the final output becomes even more detailed. However, because our estimation of velocities was less successful, we do not actually know if the map in Fig. 4 left is more accurate than kernel smoothing. This is something that will further need to be tested using case studies with validation sets.

Obvious limitations of this case study were relatively low number of individual birds and inaccurately estimated velocities (instead of being able to use the GPS-measurements of velocities). All this happened because the time between the GPS readings was long, velocity is poorly estimated by this GPS receiver and there were many missing observations. The time support of the point readings, in fact, is really small (say 1 second), but the problem is the complete interval to which it applies is rather long (between 1 and 4 hours). In addition, there are many missing readings, which complicate the accurate analysis even more. These are problems which are not easy to solve because on the one hand the technology (GPS receivers) limits the amount of readings that we can receive and on the other hand we would like to have enough readings to represent the whole season or at least the period of highest migration.

Further research is needed to design methodology to objectively select optimal average time interval between recordings, to consider also the direction of the vector into calculation of the velocity fields, and evaluate benefits and limitations of this technique for various case studies. We also see a great opportunity to improve the reconstruction of the trajectories by combining movement models, such as the Brownian bridge movement model (Horne et al. 2007), and geostatistics. With the rapid development of lighter and more durable GPS receivers, we expect that such type of analysis will be increasingly useful, both to ecologists and governmental agencies.

5. Acknowledgements:

The gull data was provided by Bruno Ens (SOVON, the Netherlands) and Michael Exo (Institute of Avian Research, Germany). This project is made possible in part by the European Space Agency FlySafe initiative.

6. References


3 Note that also modifications of kernel smoothing exist that can used to deal with irregular sampling intervals. For more details see Katajisto et al. (2006) and Horne et al. (2007).


