

Impact of Temporally Downsampling Movement Data on Interpretation

Laleh Moussavi^{a,*}, Aidan Slingsby^a, Gennady Andrienko^{a,b}

^a *giCentre, City, University of London, UK – laleh.moussavi@city.ac.uk, a.slingsby@city.ac.uk, gennady.andrienko.1@city.ac.uk*

^b *Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), Germany*

* Corresponding author

Keywords: Movement Analysis, Visual Analysis, Animal Tracking, Episodic Data

Abstract:

The advances in locational technologies have resulted in a huge amount of geo-spatial data. The abundance of such data has led to the research area of movement analysis (Andrienko et al., 2013). This research area investigates the surrounding condition of a moving object to understand the factors and context under which a specific movement takes place (Parent, et al., 2013). One of the applications is in studying animals' movement to understand their interaction with each other and with the environment (Slingsby and Emiel, 2017). Researchers have developed a variety of methods that enable them to consider the context when analysing trajectories. Most of these methods use interpolation to reconstruct the object's trajectories for unmeasured locations and times. This approach is acceptable for analysing movement data with fine spatial and temporal resolution (quasi-continuous); however, many movement data types do not allow interpolation, including data with large or irregular spatial and temporal gaps (episodic) (Andrienko et al., 2012) (Chen et al., 2015). The reason is that when the measurements are irregular or the temporal gaps are large, an accurate estimation of trajectories and their geometry is not possible. Hence in episodic movement data, estimation of object's positions between the measured positions (interpolation) is not valid. Episodic data are very common. They are usually produced by event-based and location-based data collection methods. They may also be produced by time-based methods when the position measurements cannot be recorded sufficiently frequently (Andrienko et al., 2013).

In this research, we aim to understand the differences between quasi-continuous and episodic movement data. For this purpose, we started with a quasi-continuous dataset and then performed downsampling and investigated its impact. We used the juvenile lesser black-backed gulls' dataset¹ as our case study to investigate the effects of using temporally coarser data on contextual data visual analysis. The dataset has 271807 records for 50 birds in a time period of six months (July to December 2020), approximately one record per bird every 20 minutes. We performed the following steps²:

First, we drew monthly trajectory plots with the original gulls' dataset for each bird.

Second, we simulated episodic movement data by downsampling the original dataset to 2 hours gaps (16.7%) and plotted the monthly trajectories for each bird using the episodic dataset.

Third, we visually compared these trajectory plots to understand their behaviours in different months and explore the possibility of information loss during data degradation. The following are examples of information that was preserved in both plot types: I. In July, almost all gulls stayed in the same place where they were tagged and released. Most of them stayed there during August, too, and started their journey in September. II. They eventually moved to the south to places such as France, Spain, and Morocco, either directly or in multiple steps. III. They prefer to have their long stay stops near the water, either by a pool, sea, or golf. IV. They fly long distances for a couple of days and then stay in one place for a month or more. Figure 1 shows four sample trajectories of a bird during two different months, using the original and the degraded data. Some of the above explanations could be seen in the figure.

In contrast to high-level information that is preserved in both plot types, there is some detailed information visible in the original, but missing in the degraded plots (Figure 1, A vs B). These are information related to activities that their occurrence takes less than 2 hours time, such as going for a short food hunting, having a short rest in between flights, and starting or finishing their daily movements.

In the **Fourth** step, we quantitatively investigated the effects of coarser data on trajectories.

- A) We *downsampled* the dataset and computed the spatial differences between the original and this synthetic dataset. we randomly dropped P% (e.g., 20%) of the data and kept (100-P)% (e.g., 80%) of the data. In order to simulate event-based episodic movement data, we randomly dropped or kept each record independently from other records.

¹ Data source: https://github.com/inbo/bird-tracking/tree/master/data/processed/LBBG_JUVENILE

² Our code is available at: https://github.com/LalehMoussavi/gulls_movement_downsampling

- B) We used linear *interpolation* to interpolate the dropped values.
 C) We used the geodesic great circle method to compute the distance between the original positional records and the synthetic/episodic records.

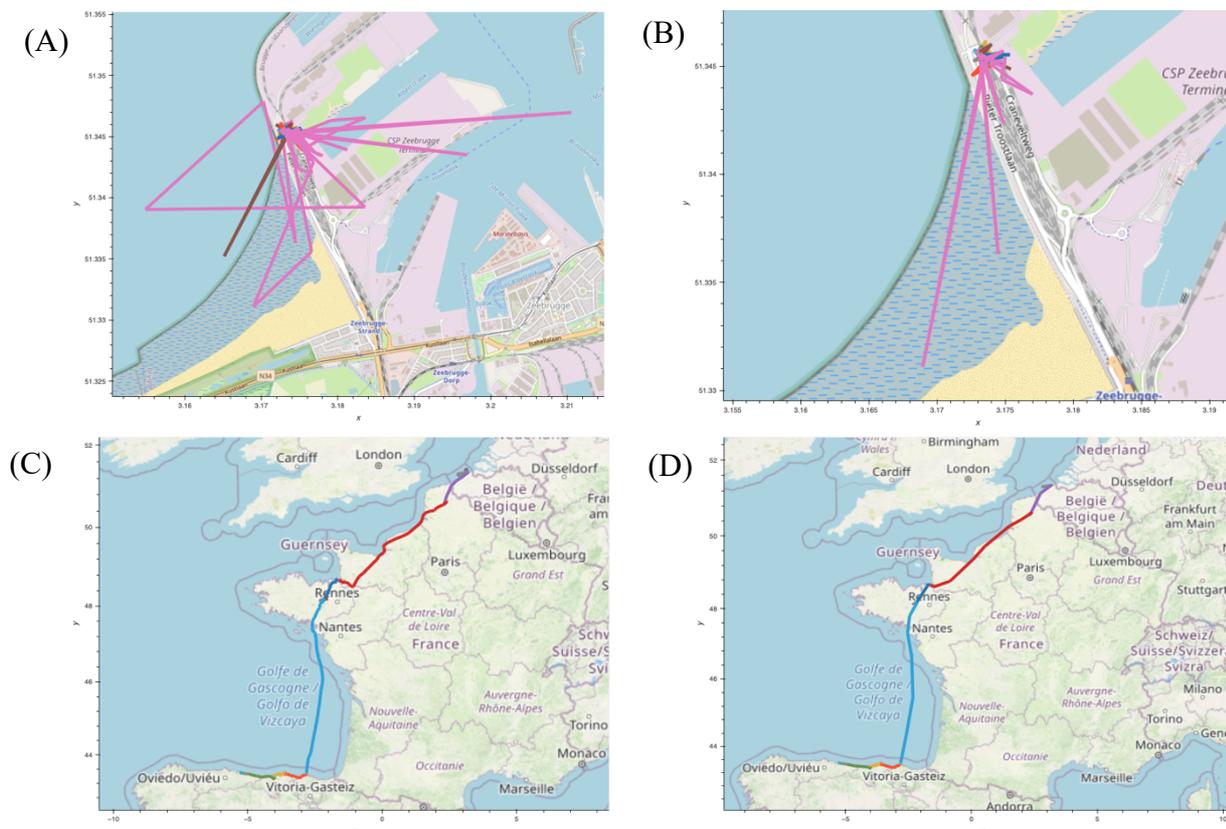


Figure 1. Trajectories of a bird during July (A and B) and September (C and D). In each row, the left image shows the bird's trajectory using the original dataset (20 minutes time gap) and the right image shows it using the degraded data (120 minutes time gap).

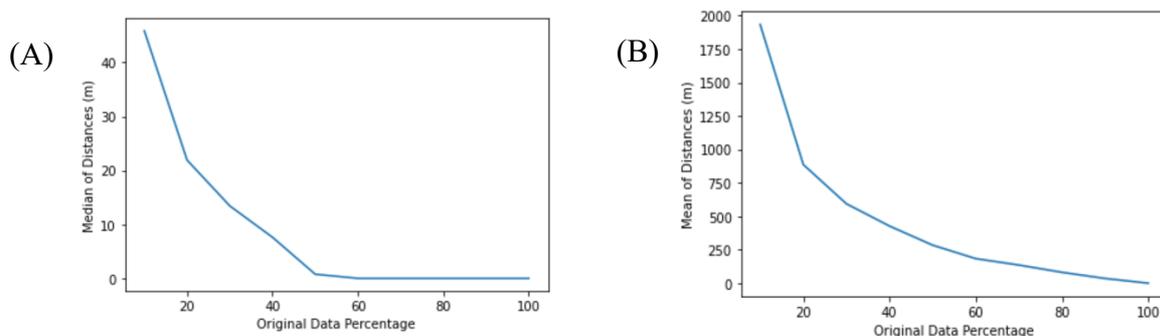


Figure 2 Change in A) Median and B) Mean of distances by different percentage of original data.

Figure 2 show the median and mean of these distances on a range of original data percentage. In figure 2A, as we get more portion from the original data, the median of distances decreases. With more than 50% of the original records, the median of distances becomes 0, which is because most data stay untouched in the episodic dataset. In figure 2B, the average distance is quite high when we have only a small amount of the original dataset (e.g., around 2000 meters when we have only 10% of the original data). We could say that the average distances are quite sensitive to outlier values and are not a good representative if we want to show the spatial difference on a map.

In this study, we investigated the impact of temporally downsampling movement data on its interpretation. We conclude that it would be possible to rely on interpolation when we have quasi-continuous data. However, depending on the application, we might be losing important information when the data is episodic. In the future, we are going to analyse real-world episodic movement data such as human movement data by developing novel methods that are robust to the missing data.

References:

Andrienko, Gennady, et al. *Visual analytics of movement*. Springer Science & Business Media, 2013.

Andrienko, Natalia, et al. "Visual analytics for understanding spatial situations from episodic movement data." *KI-Künstliche Intelligenz* 26.3 (2012): 241-251.

Chen, Siming, et al. "Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data." *IEEE transactions on visualization and computer graphics* 22.1 (2015): 270-279.

Parent, Christine, et al. "Semantic trajectories modeling and analysis." *ACM Computing Surveys (CSUR)* 45.4 (2013): 1-32.

Slingsby, Aidan, and Emiel van Loon. "Visual characterisation of temporal occupancy for movement ecology." (2017).