Characterizing Bird Movement using Spectral Clustering

Raphaël Romero, Lucy Mitchell, Frederick Verbruggen, Luc Lens, and Jefrey Lijffijt

Ghent University, Belgium

Abstract. High-resolution animal telemetry data, increasingly available through GPS technologies, has proved very valuable to study animal behavior in relation to their environment. A key open question is: *How can cognitive measurements from laboratory studies be meaningfully related to behavioral patterns observed in the wild?* In this study, we analyze the movements of juvenile Herring Gulls tracked in Belgium from 2022 to 2025. We introduce a simple yet effective method inspired by spectral clustering to extract informative, individual-level behavioral features. We show that the eigenvalues of the Laplacian have a natural interpretation and can be use to assess the difference in movement patterns between individuals. Moreover, we introduce a method to quantify revisit patterns, which provides an additional view on similarities between individuals.

1 Introduction

Enabled by the recent advances in GPS technology, high-resolution animal telemetry data has become increasingly available in the past decades [3]. Analyzing this data has critical applications in species conservation and environmental preservation [8,7,9,11,12]. As a result, behavioral ecologists have increasingly relied on statistical analysis techniques such as HMMs, spatio-temporal point processes and differential equations, as surveyed in [8], with the goal of extracting both individual and population-level characteristic quantities.

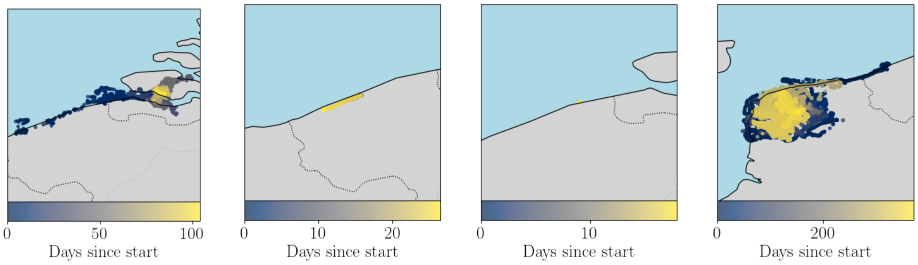
Telemetry data collected using currently available GPS devices presents several challenges. For example, solar-powered GPS units depend on light exposure, which can result in irregularly sampled location data over time. Moreover, the number of spatio-temporal measurements may vary greatly accross individuals.

In this paper, we focus on a specific sub-task: extracting features from animal trajectories such that the similarity between these features reflects underlying cognitive or behavioral similarities between individuals. We argue that an effective feature extractor should satisfy three key properties: (1) **Invariance** to trajectory transformations that do not carry behavioral meaning, (2) **Scalability** to large volumes of high-resolution trajectory data, and (3) **Direct interpretability** in terms of behavioral patterns.

Our work uses well established spectral clustering methodology [1,4,10,6] to a resampled version of the trajectories in order to extract features satisfying these properties.

2 Data

We consider 157 juvenile herring gulls (Larus argentatus) tracked in Belgium, with data collected from 2022 to 2025. Each bird's trajectory is represented by a set of observations (x_i, y_i, t_i), where (x_i, y_i) are the spatial coordinates and t_i is the timestamp of the i-th observation. Note that due to the nature of the tracking technology, the observations are irregularly sampled, meaning that the time intervals between observations can vary significantly. In Appendix A, we provide more specific information about the distribution of time gaps between observations for each individual.



(a) H911142 (b) H912505 (c) H911975 (d) H911566

Fig. 1: Trajectories of four juvenile herring gulls nearby coastal Belgium.

Objective The goal of this study is to analyze the movement patterns of birds by extracting features from their trajectories that can be linked to cognitive performance measured in controlled laboratory settings. The general strategy is to identify individual-level features that can discriminate between different "types" of birds, where type is defined by characteristics such as cognitive traits.

Strategy Our general strategy is to build features which are invariant to certain transformations of the trajectories that are not behaviorally meaningful. These include for instance differences in the number of recorded measurements and rigid transformations such as rotations and translations. In ecological terms, we aim to identify patterns of movement that may indicate different behavioral states, such as foraging, nesting, or migration. We hereby describe how these different states are expected to manifest in the data.

- **Foraging:** Birds may exhibit frequent changes in location, with short-term movements between feeding sites.
- **Nesting:** Birds may show more localized movements, with longer periods of inactivity or short movements around a fixed location.
- **Migration:** Birds may display long-distance movements with fewer stops, indicating a more directed travel pattern.

3 Method

Our proposed method consists in three main steps: (1) smoothing/averaging the trajectories, resulting in a discrete trajectory constructed from the irregularly sampled observations, (2) constructing for each bird a similarity matrix S which represents the similarity of timestamps of the discrete trajectories, and (3) a spectral analysis step where the spectral properties of the similarity matrix are computed and linked to behavioral patterns of the birds.

3.1 Trajectory Resampling

We start with a set of observations (x_m, y_m, t_m) for each bird, where $z_m \overset{\Delta}{=} (x_m, y_m)$ are the spatial coordinates and t_m is the timestamp of the m-th observation. We construct a discrete trajectory by averaging the observations over a fixed time interval (in our case we use days $\mathcal{I}_i$), resulting in a set of discrete points $\bar{z}_i = \big(x_i, y_i\big)$ for $i = 1, \ldots, N$, where N is the number of days in the trajectory.

$$\bar{z}_i = \frac{1}{|\mathcal{I}_i|} \sum_{\substack{m \\ t_m \in \mathcal{I}_i}} z_m$$

In case of missing data, i.e. if no observation is observed for one or several days, linear interpolation is used to impute the missing coordinate values. The choice of granularity should reflect the resolution at which we expect behaviorally discriminating features to emerge from the trajectories.

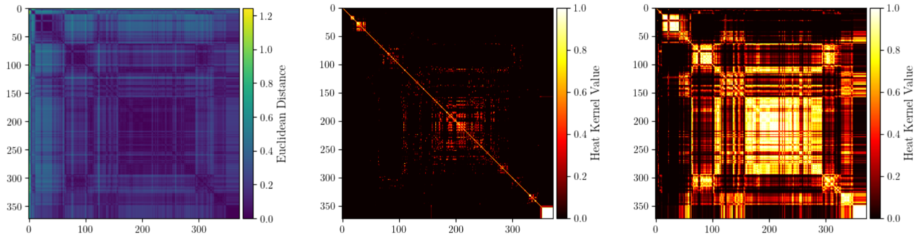
3.2 Similarity Matrix Construction

We construct a similarity matrix $S^{(k)}$ for each bird k based on its discrete trajectory, following two steps. In the first step, we compute the pairwise Euclidean distances $\Delta_{ij}^{(k)} = \left\| \bar{z}_i^{(k)} - \bar{z}_j^{(k)} \right\|$ between the bird's locations on different days, where $\bar{z}_i^{(k)}$ denotes the location of bird k on day i. Next, we apply the so-called *heat kernel* [1] to transform these distances into similarities:

$$S_{ij}^{(k)} = \exp\left(- \frac{\Delta_{ij}^{(k)^2}}{\sigma_k^2} \right),$$

where σ_k is a scale parameter that may be bird-specific or shared across individuals. The resulting similarity matrix $S^{(k)} \in \mathbb{R}^{N^{(k)} \times N^{(k)}}$ captures the pairwise similarity between days based on the spatial proximity of the measurements made on different days, where $N^{(k)}$ is the number of days with available observations for bird k. We choose σ_k to be the median distance between all pairs of days for bird k, effectively transforming all pairwise distances into a common scale of day-to-day similarity that can be compared across birds.

Figure 2 shows the day-to-day distance and corresponding similarity matrices for two values of σ_k. Smaller σ_k values emphasize local structure by assigning low



(a) Distance Matrix (b) Heat Kernel $\sigma = 0.01$ (c) Heat Kernel $\sigma = 0.1$

Fig. 2: Similarity matrix with different temperature parameters for individual H911566 which is observed for a total of 372 days.

similarity to distant days, while larger values capture more global patterns by smoothing over time. The resulting distance/similarity matrices are inherently invariant to translation and rotation of the birds, but still reveal interpretable behavioral structures in the trajectories.

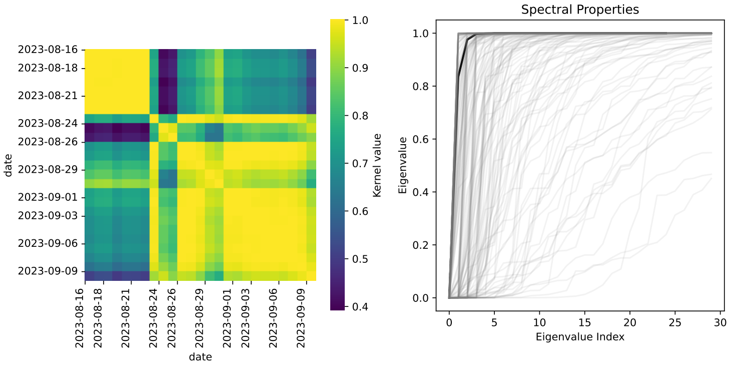
First these matrices are symmetric and have zeros along the diagonal (as the distance between a day and itself is zero). Moreover, diagonal blocks of near-zero values indicate that the bird remained in a single location for a sustained period—behavior that may correspond to **nesting**. Off-diagonal regions with similarly low values suggest that the bird revisited a previous location, as indicated by two time intervals during which the bird was spatially co-located, signaling a **revisit**. Conversely, diagonal patches with consistently high distance values reflect more **dynamic** behavior over the associated interval. Smaller such patches containing both low and large similarity values may typically be associated with **foraging** or **explorative** behavior, where the bird switches between a dynamic mode and static behavioral mode. In contrast, larger low similarity blocks indicate more consistent movement and are generally characteristic of **migration**.

3.3 Spectral Analysis

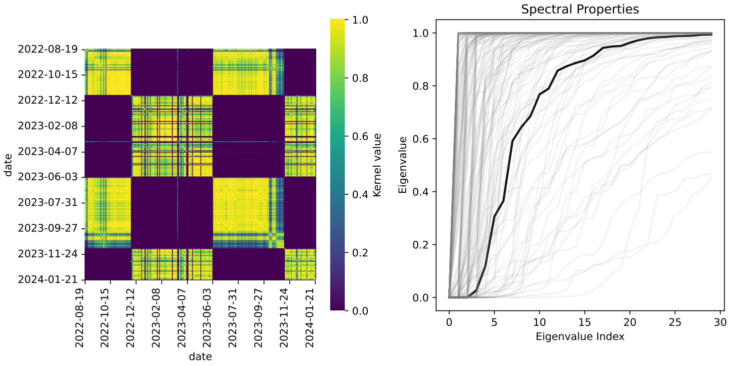
In order to automatically extract the structural patterns discussed above in a way that is comparable across individuals, we apply spectral clustering via the eigendecomposition of the normalized Laplacian matrix L, following [10,6,4]. For an individual k (suppressing the superscript for clarity), we define the diagonal degree matrix D with entries $D_{ii} = \sum_j S_{ij}$, representing how typical or frequently visited a location is on day i. Low values of D_{ii} indicate exploratory behavior, while high values reflect revisits or prolonged stays.

While the unnormalized Laplacian $L^{(\mathrm{un})} = D - S$ is a natural starting point, its eigenvalues depend on trajectory length, making it unsuitable for cross-individual comparisons. We therefore use the normalized Laplacian, defined as

$$L = I - D^{-1/2} S D^{-1/2}.$$



(a) H908641: relative to all the birds in the study (light gray line), this bird was relatively static overall, hence all the values in the kernel metrics are close to 1, and no clear structure emerges in the Laplacian spectrum. All the non-zero eigenvalues are close to 1



(b) H911140: This bird in contrast clearly visits several locations, with two main modes: a static mode between 2022-08-19 and 2022-12-12, then a more dynamic mode (leading to several high index eigenvalues) between 2022-12-12 and 2024-06-03, then another static mode between 2023-06-03 and 2024-09-27. This richer structure is reflected in the Normalized Laplacian spectrum (on the right), which shows several zero eigenvalues indicating two main sites, and various intermediate eigenvalues, indicating a more complex structure at multiple scale in the trajectory.

Fig. 3: Laplacian eigenvalues and raw kernel matrix for two birds. In the right-hand plot, the spectra of all individuals are shown as light gray lines, with the spectrum of the selected individual highlighted in black.

We apply the methods described above in order to visualize the spectral properties of each individual bird relative to the others. On Figure 3 we show the Laplacian spectrum and the raw kernel matrix for two birds, H908641 and H911140. The results show that the spectrum (i.e. set of eigenvalues) of the Laplacian matrix is a normalized indicator of how clustered the trajectory is. Notably, it reveals the number of stay points (i.e. locations where the bird stayed for a long time), which is the number of connected components of the similarity graph. As we will see later, the eigenvectors of the Laplacian matrix can be used to cluster the days into such sites, and to quantify revisitation patterns.

Spatio-Temporal Segmentation via Clustering As described in [5,10], while the eigenvalues of the normalized Laplacian L reflect the connectivity of the similarity graph S, its eigenvectors can be used for clustering. Specifically, we embed each day into a d-dimensional Euclidean space using the first d eigenvectors $\phi_1, \ldots, \phi_d$, forming the matrix $\Phi = [\phi_1 \mid \ldots \mid \phi_d] \in \mathbb{R}^{N \times d}$, where each row corresponds to a day. We then apply k-means clustering to the rows of Φ, assigning each day to one of k clusters. The number of clusters k can be chosen based on prior knowledge or heuristics such as the elbow method or silhouette score, resulting in a sequence of cluster labels $l_1, \ldots, l_N$.

3.4 Quantifying Revisitation Patterns

Revisitation patterns are a well-established source of insights into population ecology [2]. Notably, they also capture key cognitive traits, such as spatial memory. Building on the segmentation derived from the previous clustering step, we introduce a novel method for quantifying revisitation patterns in individual movement trajectories. Recall that the clustering step produces a sequence of state labels $l_1, \ldots, l_N$, where each label l_i represents the cluster assigned to day i. We now provide a general definition of a revisitation pattern:

Definition 1 (Reoccurrence). *Consider a sequence of state labels* $l_1, \ldots, l_N$*. A reoccurrence is said to occur if there exist indices* j, k, l *with* $j < k < l$ *such that*

$$l_j = l_l \quad \text{and} \quad l_k \neq l_j .$$

In other words, the individual returns to a previously visited state after having visited at least one different state in between.

Following this definition, the sequence of state label may be factorized as a set of patches

$$l_1, \ldots, l_N = \underbrace{l_1, \ldots, l_j}_{\text{patch 1}}, \underbrace{l_{j+1}, \ldots, l_k}_{\text{patch 2}}, \underbrace{l_{k+1}, \ldots, l_l}_{\text{patch 3}}, \ldots$$

where each patch corresponds to a contiguous subsequence of the state labels where the state remains constant. Based on this definition, we can define two additional quantities that characterize the reoccurrence behavior of the individual, namely the **gap** and the **dwell time**:

Definition 2 (Gap of a Reoccurrence). *Given a sequence of state labels* $l_1, \ldots, l_N$*, consider indices* j *and* l *with* $j < l$ *such that* $l_j = l_l$*. The gap of this reoccurrence is defined as the number of steps between these indices, i.e.,*

$$Gap = l - j - 1,$$

provided that for each intermediate index k *(with* $j < k < l$*) we have* $l_k \neq l_j$*.*

Definition 3 (Dwell Time). *For any occurrence of a state* l_j*, the dwell time is defined as the number of consecutive steps during which the individual remains in the state* l_j*. Formally, if* d *is the largest non-negative integer such that*

$$l_{j+1} = l_{j+2} = \cdots = l_{j+d} = l_j,$$

(with $j + d = N$ *or* $l_{j+d+1} \neq l_j$*), then* d *is the dwell time associated with the occurrence at index* j*.*

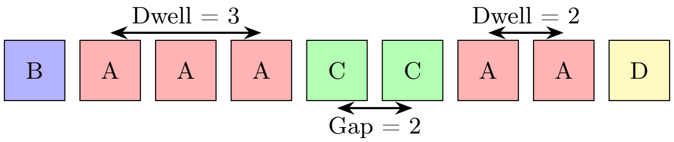
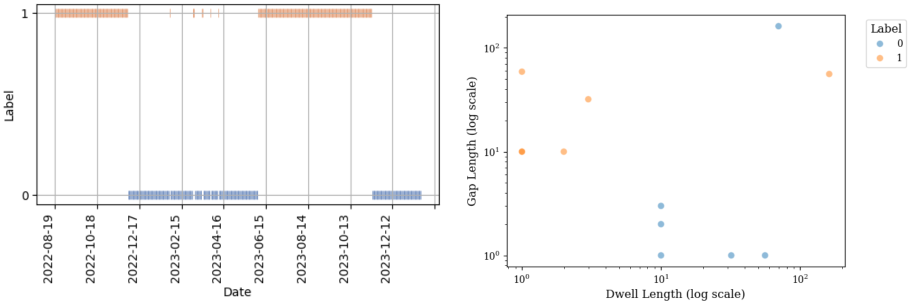


Fig. 4: Illustration of a reoccurrence of state A, showing the gap and dwell time.

These definitions allow to quantify revisit behavior in a way that is invariant to translations and rotations of the trajectory. We show the revisit statistics for bird H911140 in Figure 5a. To compare individuals based on their revisit patterns, we proceed as follows. First, we identify all revisit patterns for each bird. Then, we compute the magnitude of each revisit simply by summing the gap and dwell time:magnitude = gap + dwell time. For each individual, we select the revisit pattern with the highest magnitude and use its gap and dwell time to represent that individual. This approach yields a two-dimensional representation of revisit magnitude for each bird, where the x-axis corresponds to the maximum gap length and the y-axis corresponds to the maximum dwell time.

The resulting maximum dwell time and maximum revisit gap for each bird are shown in Figure 6. The scatter plot reveals that, for individuals tracked over a large number of days, the maximum gap length and maximum dwell time tend to align along a circular pattern. Points in the top right of the plot correspond to individuals that revisited a location after a long absence and remained there for an extended period. This pattern may suggest that these individuals possess strong spatial memory.



(a) Cluster Label for each day (b) Revisit Gap Length vs. Dwell Time

Fig. 5: Cluster label sequence and revisit statistics for bird H911140. The right panel shows gap vs. dwell time for the bird's revisits to different cluster labels. Notably, the top-right orange patch (label 1) represents a revisit with a long dwell time but a shorter gap, indicating a return to a frequently used, more stable location. In contrast, the blue patch on the bottom right (corresponding to the top right blue point on the scatter plot) corresponds to a revisit with a longer gap but shorter dwell time.

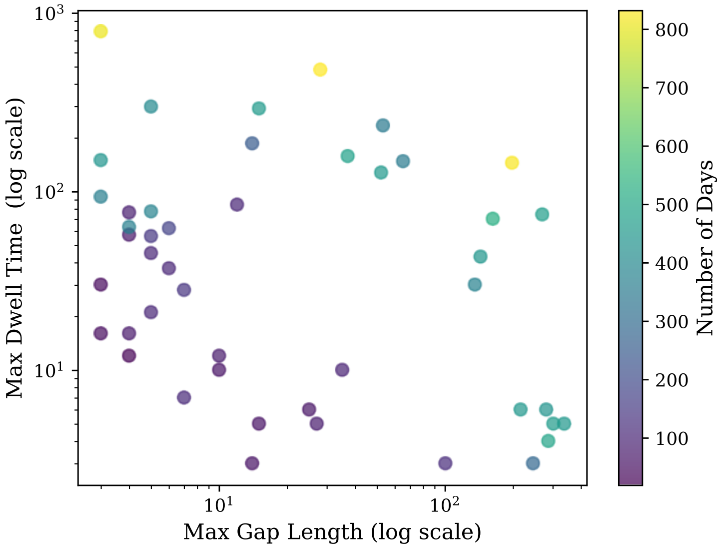


Fig. 6: Scatter plot of maximum revisits vs maximum dwell time for all bird.

4 Conclusion and Future Work

In this work, we have presented a simple method inspired by spectral clustering to extract behavioral features from bird trajectories. These features are invariant to the number of observations, scalable to large datasets, and directly interpretable in terms of behavioral patterns. We demonstrated that spectral clustering—through its use of eigenvalues and the resulting sequence of labels—offers a simple yet principled approach to deriving interpretable features. Future work will explore applications to other animal movement datasets and compare the extracted behavioral features with known cognitive attributes measured in laboratory settings. Additionally, a crucial next step is to quantitatively and qualitatively compare our method with other trajectory segmentation techniques, such as the widely used (H)DBSCAN, as well as with existing feature extraction methods, to validate the effectiveness of our approach.

Acknowledgments. This research was supported by the Special Research Fund (BOF) of Ghent University (BOF20/IBF/117), by the Flemish Government through the Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen programme, and by the Research Foundation – Flanders (FWO, project nos. G0F9816N, 3G042220, G073924N). It also received funding from the European Union (ERC, VIGILIA, grant agreement no. 101142229), the ERC Consolidator Grant (Horizon 2020 research and innovation programme, grant agreement no. 769595), and the Methusalem Project 01M00221 (Ghent University).

References

1. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation **15**(6), 1373–1396 (Jun 2003). `https://doi.org/10.1162/089976603321780317`
2. Bracis, C., Bildstein, K.L., Mueller, T.: Revisitation analysis uncovers spatio-temporal patterns in animal movement data. Ecography **41**(11), 1801–1811 (2018). `https://doi.org/10.1111/ecog.03618`
3. Cagnacci, F., Boitani, L., Powell, R.A., Boyce, M.S.: Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. Philosophical Transactions of the Royal Society B: Biological Sciences **365**(1550), 2157–2162 (Jul 2010). `https://doi.org/10.1098/rstb.2010.0107`
4. Chen, Y., Chi, Y., Fan, J., Ma, C.: Spectral Methods for Data Science: A Statistical Perspective. Foundations and Trends® in Machine Learning **14**(5), 566–806 (2021). `https://doi.org/10.1561/2200000079`
5. Chen, Y., Chi, Y., Fan, J., Ma, C.: Spectral Methods for Data Science: A Statistical Perspective. Foundations and Trends® in Machine Learning **14**(5), 566–806 (2021). `https://doi.org/10.1561/2200000079`
6. Chung, F.R.K.: Spectral Graph Theory. No. no. 92 in Regional Conference Series in Mathematics, Published for the Conference Board of the mathematical sciences by the American Mathematical Society, Providence, R.I (1997)

7. Getz, W.M.: A hierarchical path-segmentation movement ecology framework. Ecological Processes **11**(1), 56 (Sep 2022). `https://doi.org/10.1186/s13717-022-00399-5`
8. Hooten, M.B. (ed.): Animal Movement: Statistical Models for Animal Telemetry Data. CRC Press/Taylor & Francis Group, Boca Raton (2017). `https://doi.org/10.1201/9781315117744`
9. Janmaat, K.R.L., de Guinea, M., Collet, J., Byrne, R.W., Robira, B., van Loon, E., Jang, H., Biro, D., Ramos-Fernández, G., Ross, C., Presotto, A., Allritz, M., Alavi, S., Van Belle, S.: Using natural travel paths to infer and compare primate cognition in the wild. iScience **24**(4), 102343 (Apr 2021). `https://doi.org/10.1016/j.isci.2021.102343`
10. von Luxburg, U.: A Tutorial on Spectral Clustering (Nov 2007). `https://doi.org/10.48550/arXiv.0711.0189`
11. Mitchell, L.J.: The influence of environmental variation on individual foraging and habitat selection behaviour of the European nightjar
12. Patterson, T.A., Parton, A., Langrock, R., Blackwell, P.G., Thomas, L., King, R.: Statistical modelling of individual animal movement: An overview of key methods and a discussion of practical challenges. AStA Advances in Statistical Analysis **101**(4), 399–438 (Oct 2017). `https://doi.org/10.1007/s10182-017-0302-7`

A Analysis of Observation Time Gaps in the Data

As mentioned in the main text, the locations of the birds are recorded irregularly, resulting in time gaps between observations for each of the 172 birds in the dataset. For each bird, we compute the maximum number of days between two consecutive location observations. The distribution of these maximum gaps is shown in Figure 7. Most individuals exhibit gaps of fewer than 25 days (typically less than 10 days), although one outlier shows a maximum gap of 193 days.

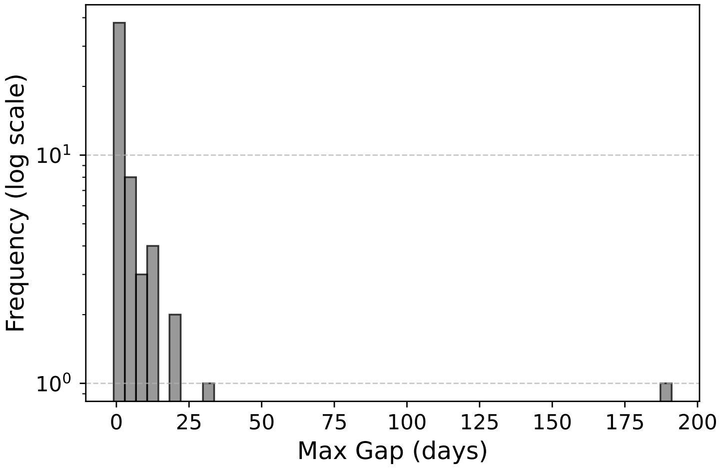


Fig. 7: Distribution of the maximum gap between two consecutive location observations.