

# Model-Based Segmentation and Classification of Gull Trajectories

Maike Buchin and Stef Sijben

Ruhr-Universität Bochum, Germany {Maike.Buchin,Stef.Sijben}@rub.de

## 1 Introduction

We analyse the trajectories of two species of gulls. The dataset consists of the trajectories of 101 individuals breeding at the Belgian Coast over just more than 2 years (Stienen et al., 2016). Our goal is to analyse the migration and non-migration behaviour of these birds, in particular comparing these two behaviours across species and individuals. For this we employ model-based segmentation and classification (Alewijnsse et al., 2014b).

First we segment and classify the input trajectories. Next we partition the segments into migration and non-migration. We compare the resulting classification of these sets across species and individuals with respect to diffusion coefficient and speed. In the following we first outline our approach in Section 2 and present our experimental results in Section 3.

## 2 Methods

In this section we outline the model-based segmentation and classification we use for analysis, which were developed jointly with several co-authors (Alewijnsse et al., 2014b; Buchin & Sijben, 2015).

Previous work on trajectory segmentation in computational geometry has focussed on *criteria-based segmentation*, where each segment fulfills given spatio-temporal criteria. Here, an optimal segmentation is one with a minimal number of segments. For this setting several algorithms have been proposed. Buchin et al. (2011) developed a framework that computes a segmentation given a *decreasing monotone* criterion, that is a criterion which if it holds on a certain segment, also holds on every subsegment of that segment. In this framework, a segmentation can generally be computed in  $O(n \log n)$  time, where  $n$  is the number of sampling points. Although many natural criteria are monotone, not all are, and for this case Aronov et al. (2013) developed an algorithm that runs in  $\Theta(n^2)$  time. Recently, Alewijnsse et al. (2014a) proposed a framework that can efficiently handle both decreasing and increasing monotone (defined analogously to decreasing monotone) criteria in  $O(n \log n)$  time. Criteria-based segmentation can also be used for classification, by using multiple criteria, one for each class. This setting has been successfully applied to a data set of migrating geese (Buchin et al., 2012b). Recently, Sankararaman et al. (2013) proposed to

segment a trajectory by detecting similar subtrajectories<sup>1</sup>. In movement ecology, classification algorithms have been used to identify behavioural states from acceleration data (Whitney et al., 2010; Nathan et al., 2012; Shamoun-Baranes et al., 2012).

Criteria-based segmentation and classification partitions data based on pre-specified criteria. The motivation for this –and other movement analysis tasks– is in many cases to make inferences about the underlying movement process. In the light of this objective it seems only natural to take a more statistical perspective on these analysis tasks: As we describe in more detail below, trajectory segmentation and classification can be seen as fitting a parameterized movement model to the data.

Taking such an approach is essential when designing algorithms for applications –as in ecology– that use movement data in a statistical analysis. Therefore we now discuss movement models used in ecology. Movement models are used to infer a continuous motion from discrete samples of the movement path. In ecology, mostly random movement models (Humphries et al., 2010), like the Brownian bridge movement model (BBMM) (Bullard, 1999; Horne et al., 2007) and variants of it (Benhamou, 2011; Kranstauber et al., 2012), Lévy walks (de Jager et al., 2011) and behavioural change point analysis (BCPA) (Gurarie et al., 2009) are used. Recently, the BBMM has also been introduced to computational movement analysis (Buchin et al., 2012a).

In a parameterized movement model, a link  $l$ , i.e. the part of the trajectory between consecutive observations, has an associated log-likelihood function  $L_l(x)$  as a function of the model parameter  $x$ . The likelihood of a parameter value for a set of links  $B$  (e.g. a segment or trajectory) is given by  $L_B(x) = \sum_{l \in B} L_l(x)$ .

We use the term *partition* to refer to either a segmentation or classification, which additionally assigns a value  $x(P)$  of the model parameter to each part  $P$  (i.e. a segment or class). A partition  $\mathcal{P}$  has an associated log-likelihood, which is defined as

$$L_{\mathcal{P}} = \sum_{P \in \mathcal{P}} \sum_{e \in P} L_e(x(P)). \quad (1)$$

That is, the log-likelihood of each part  $P$  (i.e., segments or classes) is the sum over the log-likelihoods of its elements  $e$  (i.e., single links or segments), while the log-likelihoods of the parts are added to obtain the log-likelihood of  $\mathcal{P}$ .

We could now define an optimal partition as one that maximizes the log-likelihood, but then it would be optimal to put each link or trajectory into its own part, since that allows optimizing all the likelihoods separately. One solution is to compute the optimal partition with a fixed number of parts, but typically the number of parts is not known beforehand. To avoid overfitting the movement model, the Bayesian information criterion (BIC) is frequently used (Kranstauber et al., 2012; Gurarie et al., 2009).

To facilitate multi-scale analysis, we use a more general notion of an *information criterion* (IC) to define the optimal partition. An IC assigns a value to

---

<sup>1</sup> For this they introduce a new model for similarity. The term *model* is used in our paper in a different sense, namely as referring to statistical models.

each partition based on its likelihood and the complexity of the model (that is, the number of parts). In particular we consider ICs of the form

$$IC(\mathcal{P}) = -2L_{\mathcal{P}} + |\mathcal{P}| \cdot p, \quad (2)$$

where  $L_{\mathcal{P}}$  is the log-likelihood of the model instance and  $|\mathcal{P}|$  is the number of parts of the partition. The number  $p$  is a penalty factor for adding complexity to the model that counteracts overfitting.

There are several ways to choose a good information criterion, the two most commonly used being the Bayesian information criterion (BIC) and the Akaike information criterion (AIC). The BIC has a penalty  $p = \ln(n)$ , where  $n$  is the number of measured locations for segmentation or the number of trajectories for classification. The AIC has a constant penalty  $p = 2$ . Additionally, stability diagrams (Alewijjnse et al., 2014a) can be used to find a good penalty factor for the specific data set.

We now define an optimal segmentation or classification to be one that minimizes the value of the IC. That is, an *optimal segmentation*  $\mathcal{S}_{opt}$  of a trajectory  $\tau$  is the segmentation and selection of model parameters for the segments  $x(S_i)$ ,  $1 \leq i \leq |\mathcal{S}_{opt}|$ , that achieves the minimum value for the information criterion among all segmentations and parameter values for  $\tau$ . An *optimal classification*  $\mathcal{C}_{opt}$  of a set of trajectories  $\mathcal{T}$  is the classification and selection of model parameters for the classes  $x(C_i)$ ,  $1 \leq i \leq |\mathcal{C}_{opt}|$ , that achieves the minimum value for the information criterion among all classifications and parameter values for  $\mathcal{T}$ .

Alewijjnse et al. (2014b) developed an efficient algorithm for model-based segmentation of a trajectory and we presented an algorithm for model-based classification (Buchin & Sijben, 2015; Alewijjnse et al., 2014b). We implemented both algorithms in R for the case of the Brownian bridge movement model and using a discrete set of allowed diffusion coefficients.

### 3 Experiments and Results

We segmented each input trajectory and then classified the resulting set of segments using the methods described before. The penalty factors were tuned by hand, increasing the penalty factor for the BIC by a factor of 8 for segmentation and 16 for classification. The 98 input trajectories with a total of 2,485,399 observations were subdivided into 12,787 segments, which subsequently resulted in 19 classes.

Each segment was labelled either *migration* or *non-migration*, where migration is defined as the majority of the segment being outside the 95% home range as computed by the BBMM (Horne et al., 2007). Of all the segments, 12,498 were designated non-migration and 289 migration. Some examples of the labellings produced by this method are shown in Fig. 1. This method produces reasonably good results, although some segments that intuitively look like they should be migration are not labelled as such. A likely cause for this is that a segment has to be labelled as a whole, and apparently these segments are not entirely

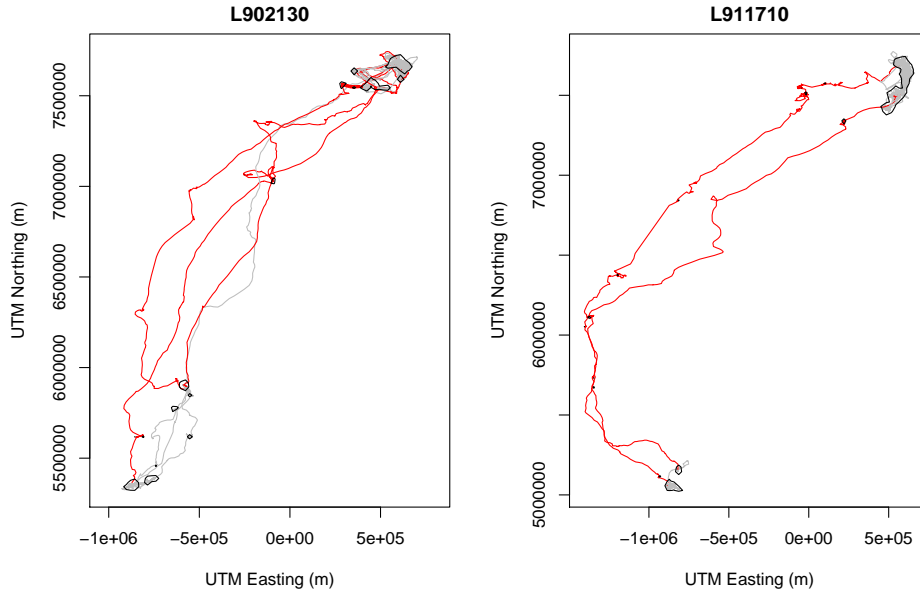


Fig. 1: Example trajectories, coloured by segment type. Migration in red, non-migration in grey. The contours of the 95% home ranges are shown in black.

homogeneous, including some stopovers and/or resting time at the start or destination. The labelling may be improved by further tuning of the parameters, i.e. the segmentation penalty factor, the level of the home range and the fraction that has to be outside.

The results were aggregated into a table containing the total duration of the segments assigned to each class, separated by animal ID and the type of trajectory.

The distribution of the segments over the classes are summarized for the two segment types in Fig. 2. There is a clear tendency for migration segments to have higher diffusion coefficients than non-migration segments, indicating less homogeneous motion during migration.

We divided the migration segments into spring and fall migration based on the date (March-August and September-February), but could detect no significant differences in the class distribution between those.

Fig. 3 summarizes the average speed in each of the classes. For each segment, the movement speed was sampled at regular intervals and the average speed for the segment computed from these. For each class, the average speeds of the segments in the class are summarized in a box plot. There is a clear correlation ( $\rho = 0.425$ ) between the average speed in a class and its diffusion coefficient.

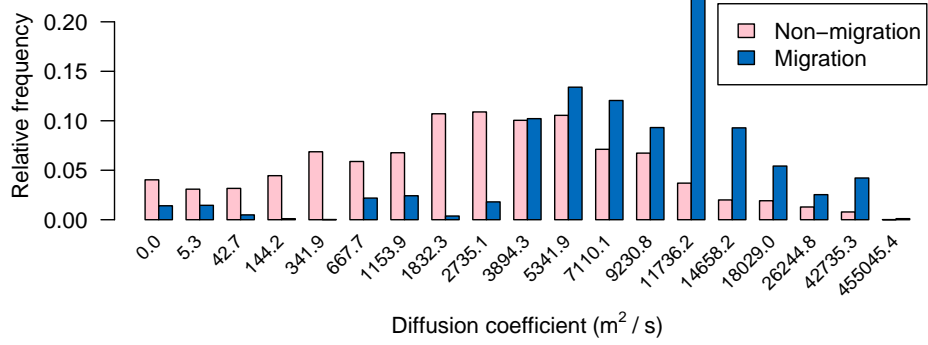


Fig. 2: Distribution of segments over classes for migration and non-migration.

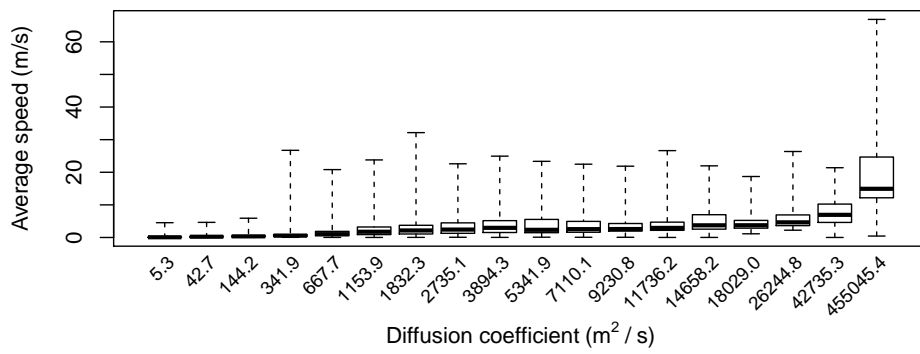


Fig. 3: Average speed for segments within each class.

## Bibliography

- Alewijnse, S. P. A., Buchin, K., Buchin, M., Kölzsch, A., Kruckenberg, H., & Westenberg, M. A. (2014a). A framework for trajectory segmentation by stable criteria. In *Proc. 22nd Internat. Conf. Advances in Geographic Information Systems (ACM GIS)*, (pp. 351–360). ACM.
- Alewijnse, S. P. A., Buchin, K., Buchin, M., Sijben, S., & Westenberg, M. A. (2014b). Model-based segmentation and classification of trajectories. Manuscript in preparation. Abstract appeared at EuroCG 2014.
- Aronov, B., Driemel, A., van Kreveld, M. J., Löffler, M., & Staals, F. (2013). Segmentation of trajectories for non-monotone criteria. In *Proc. 24th Annu. ACM-SIAM Sympos. Discrete Algorithms (SODA)*, (pp. 1897–1911).
- Benhamou, S. (2011). Dynamic approach to space and habitat use based on biased random bridges. *PloS one*, 6(1), e14592.
- Buchin, K., Sijben, S., Arseneau, T. J., & Willems, E. P. (2012a). Detecting movement patterns using Brownian bridges. In *Proc. 20th Internat. Conf. Advances in Geographic Information Systems (ACM GIS)*, (pp. 119–128).
- Buchin, M., Driemel, A., van Kreveld, M., & Sacristán, V. (2011). Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. *Journal of Spatial Information Science*, 3, 33–63.
- Buchin, M., Kruckenberg, H., & Kölzsch, A. (2012b). Segmenting trajectories based on movement states. In *Proc. 15th Internat. Sympos. Spatial Data Handling (SDH)*, (pp. 15–25). Springer-Verlag.
- Buchin, M., & Sijben, S. (2015). Model-based classification of trajectories. In K. Elbassioni, & K. Makino (Eds.) *26th International Symposium on Algorithms and Computation*, (pp. 393–403). Berlin, Heidelberg: Springer Berlin Heidelberg.  
URL [http://dx.doi.org/10.1007/978-3-662-48971-0\\_34](http://dx.doi.org/10.1007/978-3-662-48971-0_34)
- Bullard, F. (1999). *Estimating the Home Range of an Animal: A Brownian Bridge Approach*. Master's thesis, The University of North Carolina.
- de Jager, M., Weissing, F. J., Herman, P. M. J., Nolet, B. A., & van de Koppel, J. (2011). Lévy walks evolve through interaction between movement and environmental complexity. *Science*, 332, 1551–1553.
- Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, 12(5), 395–408.  
URL <http://dx.doi.org/10.1111/j.1461-0248.2009.01293.x>
- Horne, J., Garton, E., Krone, S., & Lewis, J. (2007). Analyzing animal movements using Brownian bridges. *Ecology*, 88(9), 2354–2363.
- Humphries, N. E., Queiroz, N., Dyer, J. R. M., Pade, N. G., Musyl, M. K., Schaefer, K. M., Fuller, D. W., Brunnschweiler, J. M., Doyle, T. K., Houghton, J. D. R., Hays, G. C., Jones, C. S., Noble, L. R., Wearmouth, V. J., Southall, E. J., & Sims, D. W. (2010). Environmental context explains Lévy and Brow-

- nian movement patterns of marine predators. *Nature*, *465*, 1066–1069.  
URL <http://dx.doi.org/10.1038/nature09116>
- Kranstauber, B., Kays, R., LaPoint, S., Wikelski, M., & Safi, K. (2012). A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, *81*(4), 738–746.
- Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., & Getz, W. M. (2012). Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. *The Journal of Experimental Biology*, *215*(6), 986–996.
- Sankararaman, S., Agarwal, P. K., Mølhave, T., Pan, J., & Boedihardjo, A. (2013). Model-driven matching and segmentation of trajectories. In *Proc. 21st Internat. Conf. Advances in Geographic Information Systems (ACM GIS)*. ACM.
- Shamoun-Baranes, J., Bom, R., van Loon, E. E., Ens, B. J., Oosterbeek, K., & Bouten, W. (2012). From sensor data to animal behaviour: An oystercatcher example. *PLoS ONE*, *7*(5), e37997.
- Stienen, E. W., Desmet, P., Aelterman, B., Courtens, W., Feys, S., Vanermen, N., Verstraete, H., de Walle, M. V., Deneudt, K., Hernandez, F., Houthoofd, R., Vanhoorne, B., Bouten, W., Buijs, R.-J., Kavelaars, M. M., Müller, W., Herman, D., Matheve, H., Sotillo, A., & Lens, L. (2016). Gps tracking data of lesser black-backed gulls and herring gulls breeding at the southern north sea coast. *ZooKeys*, *555*, 115–124.  
URL <http://dx.doi.org/10.3897/zookeys.555.6173>
- Whitney, N. M., Pratt Jr, H. L., Pratt, T. C., & Carrier, J. C. (2010). Identifying shark mating behaviour using three-dimensional acceleration loggers. *Endangered Species Research*, *10*, 71–82.