Correlation-based Arrangement of Time Series for Movement Analysis in Behavioural Ecology

Florian Mansmann[1], David Spretke[1], Halldor Janetzko[1], Bart Kranstauber[2], Kamran Safi[2]

[1]University of Konstanz, Germany
Emails: Florian.Mansmann@uni-konstanz.de, David.Spretke@uni-konstanz.de, Halldor.Janetzko@uni-konstanz.de

[2]Max Planck Institute for Ornithology, Radolfzell, Germany
Emails: ksafi@orn.mpg.de, kranstauber@orn.mpg.de

1. Introduction

The collection and analysis of animal movement data has the potential to address important questions about evolution, ecology, and global change. However, showing the "right" information to recognize dependencies between different measurements and to derive valuable information from these data remains a challenge. To simplify this process we propose a correlation-based arrangement of time series data collected from animal tracking sensors, or associated contextual data, to support domain experts in the hypothesis forming and validation phase of their data analysis process.

Our methodology is composed of a two-phase interactive process, in which first a particular pattern of interest is selected and all other time series are aligned according to the degree of similarity in the pattern they contain. Afterwards, based on correlation scores between all these patterns, a one-dimensional arrangement of these time series is calculated. Since this arrangement problem is NP complete, we revert to heuristic solvers of the Traveling Salesman Problem (TSP) for arranging more than 10 such time series.

2. Related Work

Visualization is essential for gaining understanding of spatiotemporal data, especially movement data, and their underlying phenomena. One established visualization technique for unprocessed movement data is the space-time cube (e.g. Kraak (2003)). This visualization technique, however, is only effective for a small number of trajectories since cluttering and overlapping of trajectories otherwise become an issue. For our work we rely on the features of Movebank (cf. Kranstauber et al. (2011)) to annotate animal tracking data with additional data such as weather conditions and build upon the Animal Ecology Explorer (A2E) system presented in Spretke et al. (2011) for visualization and interaction with this movement data.

As soon as more than two attributes in multivariate time series should be displayed simultaneously the question of ordering these dimensions arises. As described by Ankerst et al. (1998), this task is known to be NP-complete and describe methods for dimension reordering based on (partial) similarity measures and heuristics to arrange the attributes. In our case, the NP-completeness is challenging even though the number of attributes might be low, as we show one time series per attribute and per individual. The number of visualized time series is consequently the number of selected attributes multiplied by the number of chosen individuals.

3. Correlation-Based Arrangement

Looking for patterns and dependencies in the visualizations of all parameters of the tracked individual animals can be a very tedious task and the goal of this work is to provide a meaningful arrangement, which places the most similar time series near each other. Due to this dimension reordering we are able to reveal patterns and correlations among several attributes or between different animals. Note, that as soon as we take attributes of different animals into account we have to tackle several issues, namely a potential resampling of the time series to make values comparable, the temporal alignment and a resulting arrangement according to similarity.

3.1 Temporal Alignment

The first problem occurring is the asynchronous sampling of different animals. If we want to compare attributes, we have to align the time series and interpolate for time points with no measurements. After the user has chosen the *focus selection* either spatially by marking start and end points of the pattern on one trajectory on a map (see Figure 1) or temporally by marking a range in a line chart, we match all sampling points and rates by performing an artificial sampling. For this purpose the sampling rate of the time series in the focus selection is used to sample all other time series with a linear interpolation based on two consecutive measurements. As a result, all time series have matching sampling rates and time points.

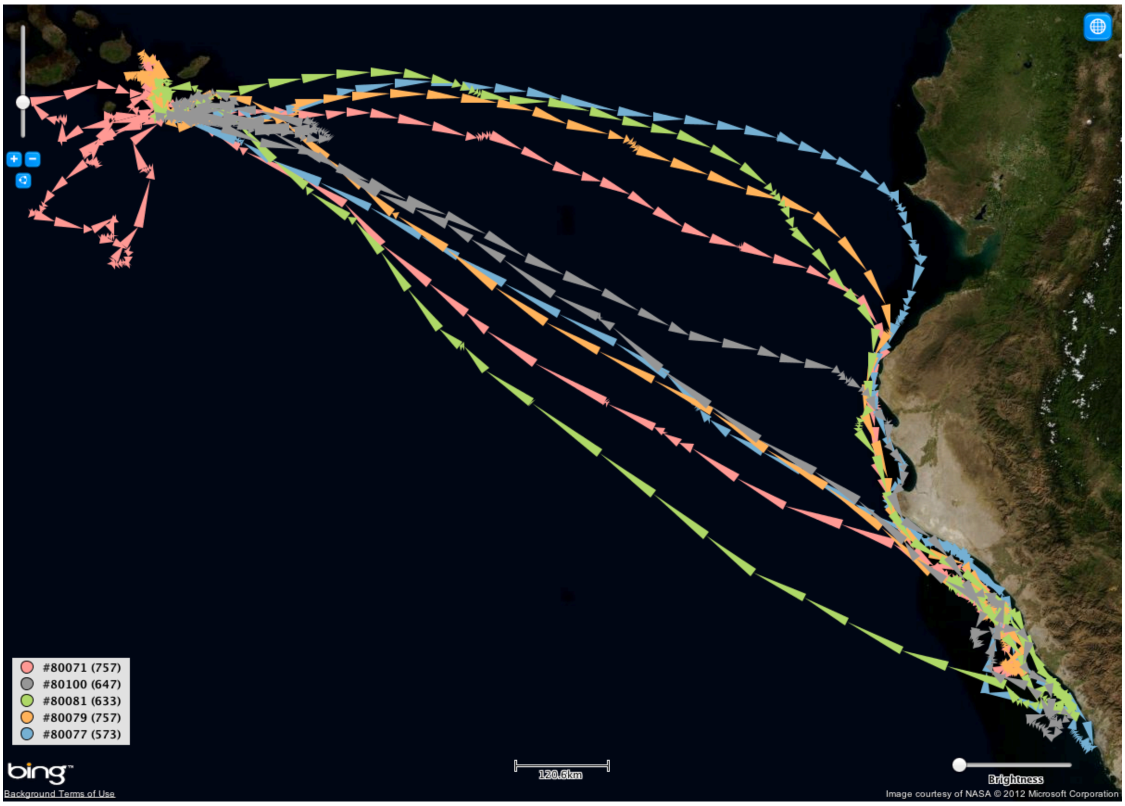

Figure 1. Map interface for choosing the *focus selection*. A pattern is selected by marking start and end points on one trajectory. As an alternative or for refinement this pattern can then be specified in details in the line chart.

The second issue occurring is the different time spans in which the animals were tracked. There are cases where animals in different years should be compared, or cases where animals should be aligned according to their behaviour, such as breeding or migration. This alignment in time can be either done manually or automatically. Our system provides an automatic alignment mode, which uses the user-given focus selection and aligns all other time series in the following manner. The focus selection describes a pattern, which should be concerned when aligning the time series. We look in all other time series for the region matching the focus selection best and align this region with the focus selection. The result of our approach is an alignment of time series matching a given time window best and helping the visual analysis process of animal movement data.

3.2 Linear Arrangement

In both cases, analysing several attributes of the same animal without temporal alignment or one attribute of different animals with a temporal alignment, we want to increase the visual salience of similarity of attributes. As the human visual system is best in finding similarities when the compared visualizations are placed nearest, we want to place similar time series near each other. We therefore calculate a pair-wise similarity score of all time series. For this calculation, we again use the user specified focus selection and compute the similarity based on linear correlation of the time spans of the focus selections. The similarity score is hereby defined as a distance measure, which is based on the Pearson correlation coefficient:

$$dist(v1, v2) = 1 - \left|corr(v_1, v_2)\right|$$

We assume in this equation that the values are already aligned and compute the similarity only for the values of the focus selection already stored in v_1 and v_2. This step can be best understood by regarding Figure 2. Here the aligned distance measures of respectively two birds are plotted in a scatterplot and the corresponding correlation value is shown on the bottom left side of the matrix.

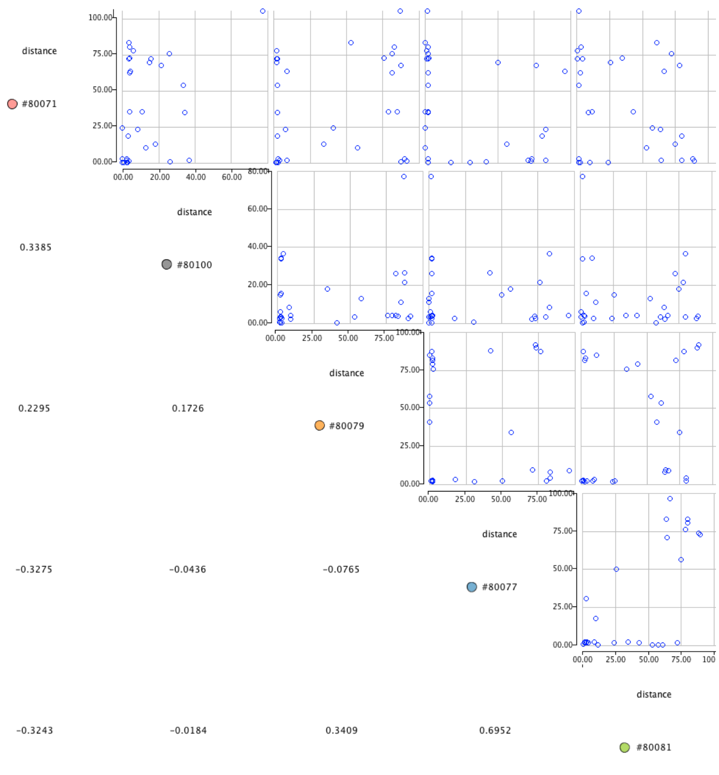

Figure 2. Correlation values and scatterplots for patterns (one per trajectory) that best match the spatially selected pattern.

After this step, we have for all pairs of attributes a distance value, which then can be fed into a solver for the Travelling Salesman Problem (TSP). The solution for the TSP will order all attributes in a way that the sum of the distances between two consecutive attributes is smallest. As the TSP is a circular trip, we cut the sequence of attributes at the location of the highest distance resulting in the first and the last attribute of our linear ordering of attributes. Note that the although the TSP is NP-complete, exact solutions for the alignment of up to nine or ten charts can be computed in a relatively small amount of time on a modern computer. Since the TSP is one of the most intensively studied problems in optimization, we can revert to reliable heuristic solutions, such as Lin et al. (1973), for aligning more than ten charts. Figure 3 shows the result for the correlation-based arrangement of five time-series.

4. Case Study: Migration Behavior of Galapagos Albatrosses

Galapagos albatrosses make long flights between the Galapagos Islands where they breed and the coastal waters of South America where they forage on the productive Humboldt Current. Tracking these birds revealed that most of them make a round trip where a northern route to the coast is taken, while they return on a southern route (cf. Fig. 1). This cannot be explained by a displacement by the Humboldt Current because it flows from South to North. In this example we want to focus on the different travel patterns of these birds during their flights from the mainland to the Galapagos Islands.

By marking the first return trip on the red trajectory (animal #80071), the pattern is automatically shown in the previously opened time series plots showing in this case the distances travelled between two location measurements taken at regular intervals of 1.5 hours (Fig 3). Since the other individuals started their return trips at different points in time, the patterns would only be comparable with a lot of cognitive effort. Starting the correlation-based arrangement automatically detects the most similar pattern in each of the other time series and aligns them. After that, a rearrangement of the plots is computed based on globally maximizing the correlation values between respectively two neighbouring patterns. Thus it becomes evident that individuals #80077 (blue) and #80081 (green), by appearing next to each other in the time series views shown in Figure 3, have most similar distribution of distances since their correlation value (0.6952) is very high (cf. matrix in Figure 2). In addition, one can see that in most tracks, especially #80081 (green) and #80079 (orange), the selected section seems to be consistently preceded by a 24-hour period of movement with irregular distances. Thus, the method allows directly picking specific trips or sections of trajectories, aligning them and comparing various attributes of associated time series.

5. Conclusions

In this work we presented an interactive correlation-based visual arrangement of time series data for exploration of movement data. After interactive selection of a pattern, the time series are first horizontally aligned according to the closest matching subsequence and then vertically placed next to each according to the strongest correlations. This scheme visually and computationally supports domain experts in their task to explore patterns animal movement behaviour.

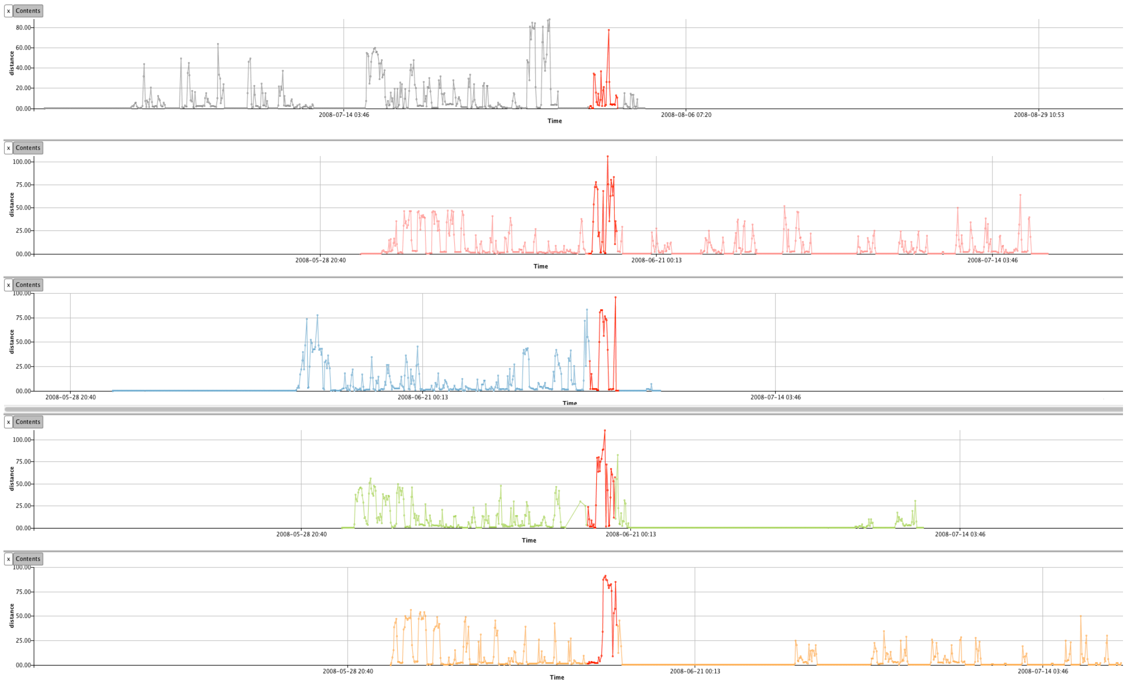

Figure 3. Vertically aligned and horizontally arranged distance time series. After marking the pattern in the red distance line chart of trajectory #80071, the correlation-based arrangement first finds and aligns the best matches in the time series of the other animals and then optimizes the sum of correlation values between respectively two adjacent charts through vertical rearrangements.

Acknowledgements

We thank Sebastian M. Cruz for sharing his tracking data. This work was funded by the German Research Foundation (DFG) within the projects "Visual Spatiotemporal Pattern Analysis of Movement and Event Data" (ViaMod) and "MoveBank Virtual Research Environment" (MoveVRE).

References

Ankerst M, Berchtold S, and Keim DA, 1998, *Similarity Clustering of Dimensions for an Enhanced Visualization of Multidimensional Data*. In Proceedings of IEEE Symposium on Information Visualization (InfoVis '98), 52–60.

Kraak MJ, 2003, *The Space-Time Cube Revisited from a Geovisualization Perspective.* Proceedings of the 21st International Cartographic Conference, 1988–1995, Durban, South-Africa.

Lin S and Kernighan B, 1973, *An effective heuristic algorithm for the traveling-salesman problem*. Operations research, 498–516.

Spretke D, Janetzko H, Mansmann F, Bak P, Kranstauber B, Davidson S and Mueller M, 2011, *Exploration through Enrichment: A Visual Analytics Approach for Animal Movement.* Proceedings of the 19th SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 421–424.

Kranstauber B, Cameron A, Weinzerl R, Fountain T, Tilak S, Wikelski M, and Kays R, 2011. The movebank data model for animal tracking. Environmental Modelling & Software, 26(6):834–835.