

FATBIRD: A Tool for Flight and Trajectories Analyses of Birds

Daniyal Kazempour*, Anna Beer*, Friederike Herzog**, Daniel Kaltenthaler*,
Johannes-Y. Lohrer*, Thomas Seidl*

*Ludwig-Maximilians-Universität, Institut für Informatik,
Lehrstuhl für Datenbanksysteme und Data Mining, Munich, Germany
Email: {kazempour,beer,kaltenthaler,lohrer,seidl}@dbs.ifi.lmu.de

**Bavarian Society for the Protection of Birds (LBV), Hilpoltstein, Germany
Email: friederike.herzog@lbv.de

Abstract—Analyzing flyways of birds is one approach ornithologists pursue e.g. to be able to detect potential risks during the animal's migration. But this analysis is not trivial and the functionalities of existing supporting tools are neither perfect nor all-encompassing. In this paper, we introduce our new FATBIRD Tool, which not only visualizes flyways or arbitrary trajectories, but also helps the researchers in several aspects of the analysis. Similarities between all trajectories of the individual birds are calculated via Dynamic Time Warping distances, which is to the best of our knowledge the first usage in this field and delivers promising results. We show the functionalities of our tool on a use case based on real data of a GPS/GSM telemetry study of Eurasian curlews of the “Bavarian Society for the Protection of Birds”. The similarities are shown in an intuitively understandable heat map colored distance matrix as well as a hierarchical clustering dendrogram. The clustering of all data points is performed and shown, and the data can be filtered by several parameters. With that, potential stop-over and wintering areas can be detected very fast and easily. After having obtained the similarities and differences of the trajectories in an automatic way, the researchers can focus on the biological reasons of the generated results of the FATBIRD Tool. These can lead to a better understanding of e.g. why certain birds die on their flyways and thus to new approaches to develop optimized conservation measures for the specific species.

Index Terms—Clustering, Distance Models, Dynamic Time Warping, FATBIRD Tool, Ornithology, Trajectories,

I. INTRODUCTION

Winning new insights from gathered data is striven for by many research fields, but there is no general approach which can be used for all data or for all use cases.

In ornithology, a common way to gather such data is to equip birds with high-tech transmitters to track enriched geo-data, as it is done by e.g. the *Bavarian Society for the Protection of Birds* (LBV)¹, one of the largest species and biotope protection associations in Bavaria, Germany [1]. The LBV tracks Eurasian curlews (*Numenius aquuata*) over the whole annual cycle, as can be seen in the map in Fig. 1. The smaller the bird, the smaller the transmitter has to be. This leads to some restrictions regarding emerging data: Data is

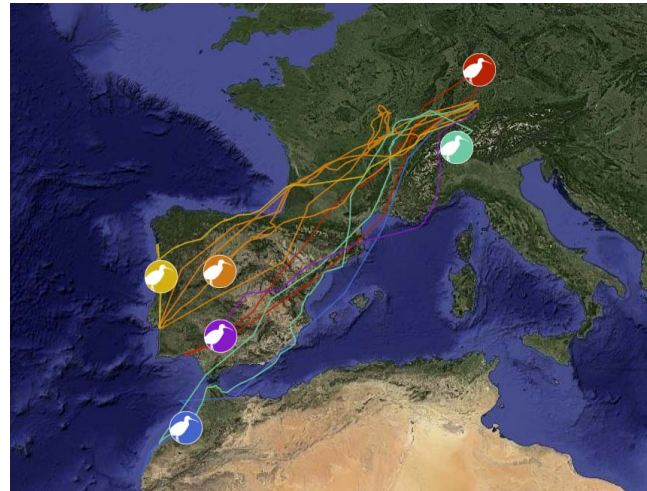


Fig. 1. The extracted trajectories of six different individuals of Eurasian curlews from the GPS/GSM telemetry data from 2016 to 2018. The home ranges in Bavaria, Germany, are hidden due to protection reasons. At www.lbv.de/brachvogel-karte it is possible to view the live map of the curlews. Map data © 2018 GeoBasis-DE/BKG (©2009), Google Pictures © 2018 TerraMetrics.

only retrieved in fixed frequencies, so e.g. data of the tracked bird sampled every 30 minutes containing various parameters, like position, height, temperature, etc. From that, biologists want to obtain information with the purpose of approaching a manifold of questions, like e.g. why some birds die on their flyway to the wintering grounds or back, and where are the most dangers in their annual migration cycle, which may lead to their population decline. On basis of this data the biologists want to prove existing and develop continuing species conservation measures.

Currently researchers use different methods to analyze migration patterns, depending on which issue is in the focus of their interest, for example the determination of stop-over areas of the birds.

¹Landesbund für Vogelschutz in Bayern e.V., <http://www.lbv.de>

Our new **Flight and Trajectories Analyses of Birds** (FATBIRD) Tool does not only visualize the gathered data in an appealing and clear manner, it also allows researchers to discover information interactively, easy, and fast which would else be hidden from view.

The main contributions of the FATBIRD Tool, that we describe in this paper, are:

- Automated computation of similarity between flyways using Dynamic Time Warping (DTW)
- Intuitively understandable visualization of similarities between flyways through heat map coloring
- Different visualization of flyways
- Clustering of flyways of birds
- Possibility to select interesting data to treat it separately more detailed

With these functionalities ornithologists are provided with a new and easy-to-use tool to examine high dimensional gathered, enriched geo-data.

We formulate the scope and targeted functionalities in Section II and give an overview over existing related work in Section III. We introduce the FATBIRD Tool and go into detail regarding the visualization concepts and the computation of the flyway similarities in Section IV. The use case for which our expandable and adaptable tool was created, the flyways of the Eurasian curlews, is treated in Section V. Finally, we conclude our work and give vital directions for future work and potential targets for extending the FATBIRD Tool in Section VI.

Annotation: The visualizations of the presented FATBIRD Tool is able to display cartographical material in the background of the trajectories. However, because we are analyzing real data in our use case, we deliberately forego this representation in this publication to protect the birds by keeping the exact positions of the breeding places unknown, and therefore undisturbed.

II. PROBLEM FORMULATION

According to Hahn et al. [2], there are 2.1 billion birds on migration between Europe and Africa each year. With the rising amount of data gathered by telemetry studies, the task to analyze this enormous data sets gets more challenging. For biologists, who want to check the data for some specific questions, a short and clear way to firstly sort the data into “points on the move” and “points during overwintering/resting” would be very helpful. Furthermore, it is important to consider all aspects of the available data to be able to analyze the overall context of the situation.

The analysis of telemetry data is already common for several years. Especially the home ranges of birds are already explored in several telemetry projects. For example, the migration route of birds is tracked to answer a set of questions: Where are the home ranges of the birds in the wintering areas in Africa and the breeding areas in Europe? [3] Which is the direction of the continued flight from a home range? [4] What is the timing of the migration of birds, dependent on their age? [5].

Also, the routes between the home ranges is an important aspect that has to be considered during the analysis. There are already studies that analyze the routes, e.g. to compare the death rate on different general routes [6]. To make such statements it is sufficient to consider the general routes of the birds – it is not necessary to consider the detailed differences within these routes. However, a very detailed view on the similarities, respectively differences of migration routes of individual birds may help to identify local key issues on the specific migration routes, which may influence e.g. the fitness of each bird.

But also, on the technical side there are some challenges: The visualization of the flyways has to be in a clear and plain manner while providing as much useful information as possible. The researchers should intuitively recognize what is shown and be able to investigate details of their choice further. Similarities between different flyways should be easy to read, which led us to use a heat map matrix showing all pairwise similarities between the flyways of particular birds.

We answer the question of how similarity between those trajectories is defined with the Dynamic Time Warping Distance. Breeding, wintering, and stop-over areas should be highlighted automatically. The exemplary positions should be interpolated linearly. This is necessary because of technical reasons: Due to transmitter and connection issues the period of time between two data points may vary from minutes to hours, sometimes days.

III. RELATED WORK

Regarding the technical aspect of our paper it boils down to the problem statement of computing distances or similarities between a set of trajectories with the final purpose of detecting clusters of similar trajectories. For computing the distance between trajectories there exists a rich body of literature.

Besse [7] provided a brief review in their work regarding trajectory distance models. To the best of our knowledge, there exist four basic types of distance models on a high level, as it can be seen in Fig. 2. First, we have the network-constrained models, which are not suitable for our domain as this model requires detailed knowledge about the paths. As for the Euclidean model, it fails to consider shifts of the trajectories, matching each residue of a path one-to-one yielding unrealistically high distance values. It is feasible for comparing subtrajectories of same length, but not on entire trajectories. In contrast warping distances provide the properties for matching locations of different indexes with different locations. These methods aim to find an optimal alignment with using e.g. dynamic computing, given a cost between matched positions.

Wang [8] proposed in their work a finer granular categorization of warping-based distance models. As in Fig. 2 we have dynamic time warping-based measures like DTW and PDTW. Further we have edit distance-based measures like EDR and ERP and as a more recent member, the longest common subsequence-based measures (LCSS). All of the here mentioned warping-based models have in their origin form

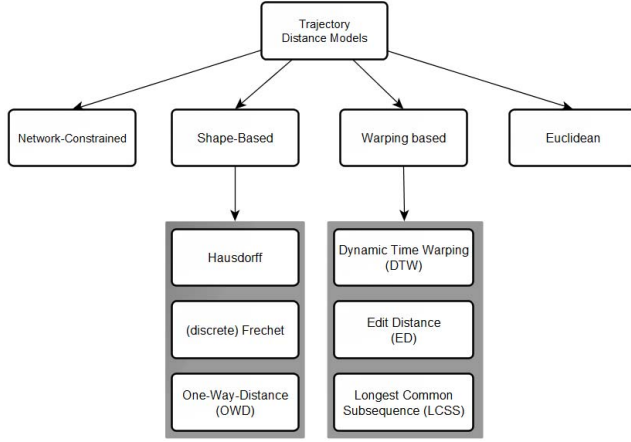


Fig. 2. Overview on various trajectory distance models.

heavy applications in the domains of computer linguistics or bioinformatics with the purpose of computing distances between sequences of text or genes or proteins. The last group, the shape-based distance models aim to extract the geometric properties of the trajectories.

For the FATBIRD Tool we use Dynamic Time Warping (DTW) to compute the pairwise distances of bird trajectories. It is left to future work to apply and evaluate in how far other trajectory-based distance models are more suitable, or to even construct an ensemble of the models to achieve best possible distance computation. Regarding the research performed on DTW, a wealth of literature and variants exists. Among them works by Petitjean et al. [9] which propose a global averaging method for DTW. Although we are using hierarchical clustering in this paper, an application of centroid-based clustering methods is of high interest with regards to future work. For this purpose, an averaging technique in sequences such as in Petitjeans work is indispensable to get centroids which represent a cluster of similar trajectories.

Computing the distance of trajectories implies to detect semantic information regarding a series of data points. Spacapietra et al. [10] contributed to this topic by analyzing and highlighting the requirements in order to get a trajectory representation with semantic enrichment. Examples for adding semantic aspects are e.g. velocity on a trajectory, additional information on the altitude of a trajectory, or a partitioning of a trajectory into sub trajectories based on certain criteria.

Having a distance model, finding an appropriate clustering method is the next step. Among various clustering models such as e.g. partitioning with k-Means [11], density based like DBSCAN [12], hierarchical like full-linkage [13], or spectral we choose here a hierarchical clustering approach. The choice for a hierarchical model comes with the idea in mind to utilize the hierarchical property to show the user in future work the outcome of the clustering, giving the domain expert the opportunity to choose at which level a cluster of birds makes sense. Especially in biology, explainability is an important

aspect which we focused on creating the FATBIRD Tool.

Comparing our FATBIRD Tool with other works such as e.g. TraClass by Lee et. al. [14], TraClass first performs a region-based clustering to obtain non-homogeneous regions. Then it partitions trajectories into a set of line segments based on the criterion that a split is performed whenever a rapid change of the moving direction occurs. The FATBIRD Tool in contrast determines the partitions of a trajectory by detecting dense areas where the moving objects have resided over a longer period of time. This design decision adds the semantic that a new part of a bird route emerges whenever an individual bird has resided for a longer period of time within an area.

IV. THE FATBIRD TOOL

In this Section, we dive into the technical concepts which have been used for the FATBIRD Tool. We first start taking a closer look on how the data is structured in order to access the vital information for the bird trajectories. Having the big picture regarding the data structure, we continue with elaborating on the core ingredients regarding the computation of trajectory distances. The distances as such serve as the very basis for computing clusters and deriving hierarchies of similarities between the bird trajectories. The whole process of mining the information from the data can be seen in Fig. 3.

A. Data Structure

Before we deal with the computation of the trajectory distances and the clusters, we first take a closer look at the structure of our data. As it can be seen in Fig. 4, in the beginning we have a data set. This data set contains of several individual birds. The records of a bird are further split by the years the data records are from. So far, we have for each bird and for each year a set of data records with longitude, latitude, and a timestamp. This information describes a trajectory. However, we need to consider the fact that the birds may remain for a longer period of time at certain spots. These spots are the breeding, wintering, or stop-over areas. Thus, we need to further split the trajectory into subtrajectories. Below, we will also elaborate on the definition of subtrajectories in more detail.

B. Partitioning Trajectories into Subtrajectories

Continuing on the data structure, we provide the following formal definition D1 which defines a trajectory \mathcal{T} :

D1 A trajectory \mathcal{T} is a series of locations l , which are 2-tuples of the values of the latitude lat and the longitude lon at specific times t . In our definition the temporal dimension is encoded implicitly over the position of



Fig. 3. Overview on the data structure.

the (lat, lon) coordinates in the trajectory tuple. Each position of the tuple \mathcal{T} represents a specific time t_i . Since we do not encompass in our study several migratory cycles within one trajectory, we can rely on this definition, otherwise a more general definition would be required.

$$\mathcal{T} = (l_{t_0}, l_{t_1}, \dots, l_{t_n}), l = (lat, lon) \quad (1)$$

Our goal, as for now, is to split a given trajectory \mathcal{T} into subtrajectories. The trajectory is split according to the breeding, wintering, or stop-over areas. This means that we need to detect those areas in order to get the points among which we can split \mathcal{T} . These areas are detected by applying the mode seeking clustering algorithm Mean Shift [15]. The core assumption here is that dense areas with many points will collapse to a single centroid leading to the following definition D2:

D2 Given a trajectory \mathcal{T} , its clustered trajectory \mathcal{T}_γ is defined as:

$$\begin{aligned} MeanShift(\mathcal{T}) &\mapsto \mathcal{T}_\gamma, \\ \mathcal{T}_\gamma &= (l_{\gamma_1}, l_{t_{|l_{\gamma_1}|}}, \dots, l_{t_{|l_{\gamma_1}|+p}}, l_{\gamma_2}) \end{aligned} \quad (2)$$

Here l_{γ_i} represent the centroids of a set of data points and $t_{|l_{\gamma_1}|}, \dots, t_{|l_{\gamma_1}|+p}$ represent all p data points between the centroids while $|l_{\gamma_1}|$ is the number of points contributing to the centroid l_{γ_1} . The centroids l_{γ_i} have to fulfill the property that the number of points belonging to them $|l_{\gamma_i}| > \tau$ where τ is a threshold set by the domain expert.

More intuitively spoken, definition D1 states that the clustered trajectory \mathcal{T}_γ consists of a series of centroids which emerged from a high amount of data points τ which *encloses* a series of data points over time. Based on the formalism of a clustered trajectory we can now extend to the definition D3 of subtrajectories:

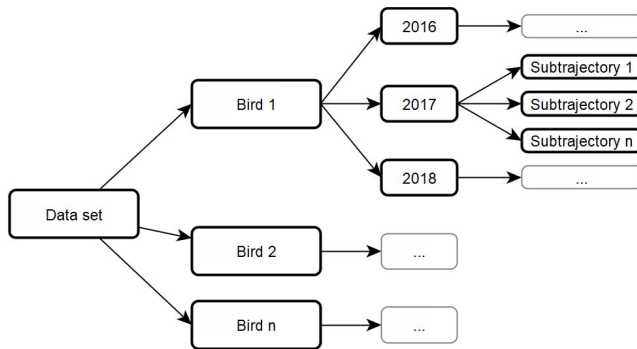


Fig. 4. Overview on the data structure.

D3 Given a clustered trajectory \mathcal{T}_γ consisting of a series of centroids $\mathcal{C}_\gamma = (l_{\gamma_0}, l_{\gamma_1}, \dots, l_{\gamma_n})$ a subtrajectory \mathcal{S}_γ is defined as a slice of \mathcal{T}_γ over the interval $[l_{\gamma_i} : l_{\gamma_{i+1}}]$.

Definition D2 tells us that a subtrajectory is a slice from the given clustered trajectory ranging from one cluster centroid to its immediate successor. It basically means that a clustered trajectory is split into its pieces at the centroids which have sufficient support. Mapping this to the biological context it means that the route of a bird is split into partial routes where the start and end point of these routes are locations where a bird resides over a longer duration.

C. Computing Distances of Trajectories

Having the clustered trajectories and the subtrajectories we proceed now with elaborating on the computation of the distances of trajectories. For this purpose, we use – as stated in Section III – the Dynamic Time Warping which is defined as definition D4:

D4 Given two trajectories $S = (s_0, \dots, s_m)$ and $T = (t_0, \dots, t_n)$, the matching cost is computed via a dynamic programming approach following Equation 3. Here $D(i, j)$ is the DTW distance between $s(1 : i)$ and $t(1 : j)$, i.e. between the first i points of trajectory S and the first j points of trajectory T . With that, $D(S, T)$ is defined as $D(m, n)$ using

$$\begin{aligned} D(i, j) &= 0, \text{ if } i = j = 0, \\ D(i, j) &= d(s_i, t_j) + \min \left\{ \begin{array}{l} D(i, j-1) \\ D(i-1, j) \\ D(i-1, j-1) \end{array} \right\}, \\ &\quad \text{if } i > 0, j > 0 \\ D(i, j) &= \infty, \text{ else.} \end{aligned} \quad (3)$$

We use Dynamic Time Warping $D(\mathcal{T}_i, \mathcal{T}_j)$ to compute a distance matrix \mathcal{D}_{dtw} which is characterized by definition D5:

D5 Given $n + 1$ trajectories $\mathcal{T}_0, \dots, \mathcal{T}_n$ the distance matrix \mathcal{D}_{dtw} is yielded by computing the pairwise Dynamic Time Warping distances $D(\mathcal{T}_i, \mathcal{T}_j)$.

$$\mathcal{D}_{dtw} = \begin{pmatrix} D(\mathcal{T}_0, \mathcal{T}_0) & D(\mathcal{T}_0, \mathcal{T}_1) & \dots & D(\mathcal{T}_0, \mathcal{T}_n) \\ D(\mathcal{T}_1, \mathcal{T}_0) & D(\mathcal{T}_1, \mathcal{T}_1) & \dots & D(\mathcal{T}_1, \mathcal{T}_n) \\ D(\mathcal{T}_2, \mathcal{T}_0) & D(\mathcal{T}_2, \mathcal{T}_1) & \dots & D(\mathcal{T}_2, \mathcal{T}_n) \\ \vdots & \vdots & \ddots & \vdots \\ D(\mathcal{T}_n, \mathcal{T}_0) & D(\mathcal{T}_n, \mathcal{T}_1) & \dots & D(\mathcal{T}_n, \mathcal{T}_n) \end{pmatrix} \quad (4)$$

The results of such a distance matrix can be seen in the use case in Section V where the distance matrix is represented in form of a heat map. The distance matrix serves as the very basis for the next step, the clustering of similar trajectories.

At this point we want to highlight that Dynamic Time Warping is not the definite solution for computing distances between bird trajectories. It is more of an entry point in

the field of distance computation of distances between bird flyways. Further, much more exhaustive research is required – which would be far beyond the scope of this paper – in order to determine potential properties of bird trajectories and to distil the criteria which are required for a good distance model in context of bird movements.

D. Computing Clusters of Trajectories

Based on the definition of the distance matrix in the previous Subsection we now utilize this distance matrix for a hierarchical clustering on bird trajectories. Here we present two types: The agglomerative and the divisive approach for hierarchical clustering. [16]

The agglomerative approach works in a bottom-up manner. Each object is initially considered as a single-element cluster. At each step of the algorithm, the two clusters that are the most similar are combined into a new bigger cluster. This procedure is iterated until all points are member of just one single big cluster. The result is a tree which can be plotted as a dendrogram.

The divisive approach works in a top-down manner. The algorithm is an inverse order of the agglomerative approach. It begins with the root, in which all objects are included in a single cluster. At each step of iteration, the most heterogeneous cluster is divided into two. The process is iterated until all objects are in their own cluster.

Hierarchical clustering methods vary regarding the choice of their linkage criteria, among which are listed complete-linkage, single-linkage, average linkage (UPGMA), centroid-linkage, or minimum energy clustering. Each of these linkage criteria have an impact on the respective cluster outcome.

As we use the full-linkage criterion as an example, it is defined in definition D6:

D6 Given two clusters of trajectories $\mathcal{C}_{i\tau}$, $\mathcal{C}_{j\tau}$, the distance of these two clusters is computed as

$$d_{\text{clink}}(\mathcal{C}_{i\tau}, \mathcal{C}_{j\tau}) = \max_{\tau \in \mathcal{C}_{i\tau}, \phi \in \mathcal{C}_{j\tau}} d(\tau, \phi) \quad (5)$$

The distance $d(\tau, \phi)$ in definition D6 poses another potential target where multiple choices are possible and their outcomes require further research in the context of our use case. We use the Euclidean distance, however, one can as well use Manhattan distance, Maximum distance, Mahalanobis, Bray–Curtis dissimilarity or even a self-engineered distance metric.

The core insight, which can and should be gained in our technical part of this paper, is to raise the awareness that there are many areas where further research is required such as the choice of the distance model for the trajectories, the choice of a clustering algorithm, the choice of the linkage type, and the choice of the distance metric. We hope at this point to have stimulated the reader to proceed with discovering the possibilities as well as limitations of the above-mentioned aspects in trajectory similarity of bird flights.

V. USE CASE: TRAJECTORIES OF EURASIAN CURLEWS

The *Bavarian Society for the Protection of Birds* (LBV)² is one of the largest species and biotope protection associations in Bavaria, Germany. It engages in several species conservation projects for birds, mammals, insects, plants, and biotopes.

One of the major scientific issues in species conservation is the population decline of animal species. To research the reasons for the decline of birds, the LBV started several telemetry projects in the recent years. The birds are equipped with modern satellite or GPS/GSM transmitters which enable the transmission of precise geo-data (longitude and latitude). These transmitters allow to manually set the time interval between two measurement points. The intervals range from 30 minutes to a maximum of 24 hours. In case of the GSM network is not available at the time of a measurement, the information is stored and sent later. The gathered data is also stored for collaborations in “Movebank” [17], a database for animal tracking data hosted by the Max Planck Institute for Ornithology³. This allows the researchers to track the position of the birds in order to be able to make reliable statements about the position of the wintering areas, the site-fidelity, the exchange of birds between different breeding areas, and the trajectories of individuals.

Recently, the LBV run telemetry projects for white storks (*Ciconia ciconia*), black storks (*Ciconia nigra*), common cuckoos (*Cuculus canorus*), and Eurasian curlews (*Numenius arquata*). The purpose of these studies is to answer species conservation questions and develop conservation measures having obtained knowledge about migration threats. [18]

For the GPS/GSM telemetry project of Eurasian curlews, the LBV equipped nine⁴ different curlews in Bavaria, Germany, with GPS/GSM transmitters, which are attached similar to a small backpack (on adult birds) and leg-loop (on juvenile

²Landesbund für Vogelschutz in Bayern e.V., <http://www.lbv.de>

³<http://www.orn.mpg.de/en>

⁴as of May 2018



Fig. 5. An Eurasian curlew (*Numenius arquata*) with an equipped GPS/GSM transmitter on the back. Image: Wolfgang Nerb.

birds) on the animals, as seen in Fig. 5. The project will be continued until 2023. Over these years a total of 30 curlews from five breeding areas in Bavaria, Germany, will be equipped with these transmitters.

In this use case, we take the recorded telemetry data of the curlews. The birds that were tagged until today generally move from the breeding areas in Bavaria, Germany, to the wintering areas in the Iberian Peninsula (Portugal and Spain) and Morocco, respectively back. As an example, the trajectories of six of the tracked curlews from 2016 until 2018 are visualized in the map of Fig. 1. Considering all available data of all nine curlews builds the basis for our use case.

Regarding the realization of the tool we used Python 3.6 as the programming language with auxiliary libraries such as e.g. numpy/scipy for numerical tasks, matplotlib for visualization scikit-learn for hierarchical clustering and FastDTW from the pip repositories for fast DTW computation which is based on the work by Salvador and Chan [19].

Step 1: Determination of Trajectories

After loading the raw data of the transmitters, the FATBIRD Tool first uses the bird ID (name and year) and the geo-data (longitude and latitude) to detect available areas in which a specific bird stays for a longer time, that means it is located in a breeding, wintering, or stop-over area. Therefore, the position points are clustered to identify these areas. The cluster recognition in our data, that means the amount of the nearby spots of an area, is indicated in the diagram in Fig. 6. The threshold level, that means the number of nearby spots that is required to form a cluster, can be set up by the researchers.

By replacing the points within these clusters with a corresponding centroid, this information is used to identify the individual trajectories of the bird. Each cluster, which actually represents a breeding, wintering, or stop-over area, is either the start or the end point of a trajectory. Regarding the start point and all points in chronological order until the end point, we collate all points that are corresponding to one trajectory. This is visualized in Fig. 7. After this step, we have split all points from the annual data of the bird into groups of points which all belongs to the same trajectories.

Finally, we gain a full representation of the trajectories of a bird by connecting the centroid of the start and end area with the position points in a chronological order, as can be seen in Fig. 8. At this point, we are also able to provide an overview and a visualization about the pathway of a single trajectory.

Step 2: Comparison of Trajectories

At this point, the FATBIRD Tool has recognized the available trajectories of all birds. Now it starts to set these trajectories in relation.

The FATBIRD Tool is supposed to support the analyses of the researchers to figure out similarities in the behavior during the flyover of the birds. Therefore, it provides two visualizations for the researchers: a distance matrix and a hierarchical clustering dendrogram.

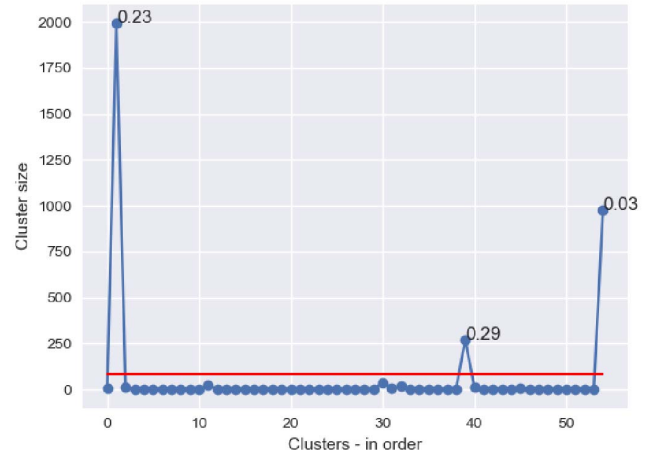


Fig. 6. Identification of the breeding, wintering, or stop-over area of the data of one curlew in one annual migration cycle. The points with a large cluster size are equivalent with these areas. The red line represents the threshold value.

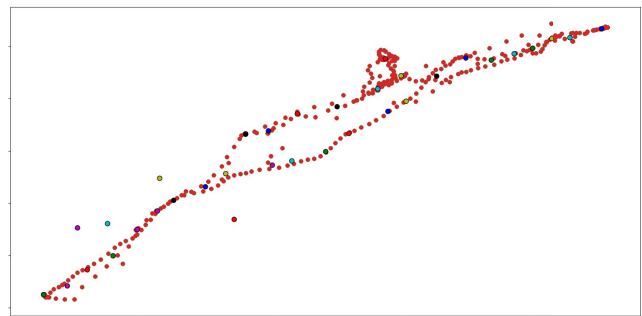


Fig. 7. The red colored points describe all available positions of one bird in 2.5 annual migration cycles. The varicolored points describe the centroid of the calculated clusters of the position of the bird which arise from a lower threshold. Units are omitted to protect the birds by keeping exact position of the breeding places unknown.

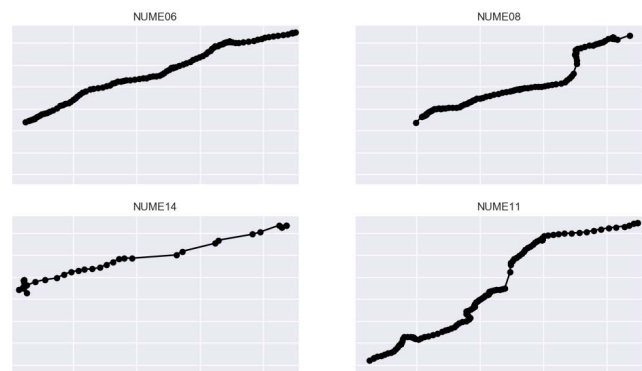


Fig. 8. The trajectory of the curlews of 2017, exemplary for the four curlews “NUME06”, “NUME08”, “NUME14”, and “NUME11” which migrated from the breeding areas in Bavaria, Germany, to the wintering areas in Portugal, Spain, and Morocco.

The distance matrix displays the differences of the trajectories. For our use case, we took the flight data from the birds in the year 2017 and considered only the route from Bavaria, Germany, to the wintering areas in Spain, Portugal, and Morocco. The visualized distance matrix, as seen in Fig. 9, sets the trajectories of the single birds in relation and displays the distance value in two ways: the textual value and the visualization as heat map colors. The lower the value and the darker the color, the more similar are the trajectories of the compared trajectories. In our example we can read out that the comparison of the trajectories of ‘NUME11’ and ‘NUME25’ have the lowest value, so they are the most similar trajectories. The trajectory of ‘NUME07’ is very comparable to the ones of all of the other birds. The trajectories with the lowest similarity are the ones of ‘NUME14’ and ‘NUME11’.

The same data is visualized in a different way in the hierarchical clustering dendrogram, which can be seen in Fig. 10. It sets tuples of trajectories, respectively groups of trajectories in relationship. Here it is visible that the tuple of the routes of ‘NUME11’ and ‘NUME25’ is the most similar one because of the lowest value (‘0.0’). ‘NUME07’ and ‘NUME08’ have the second-highest similarity to each other (value ‘1.0’), and ‘NUME06’ and ‘NUME14’ the third-highest one (‘2.0’). This information is similar to the comparison of the one of the distance matrix.

This information can be used by the biological researchers to identify similar, but also diverging routes of the birds. Starting with this automatically generated information, the researchers can now start identifying reasons why the birds choose similar trajectories, or possibly identify reasons for choosing different routes. Possible reasons for similar trajectories are geographic (e.g. river course, natural elevations, etc.) or climate reasons. The reason for differing trajectories could be local weather conditions, e.g. thunderstorms or wind conditions. While the FATBIRD Tool itself does not allow the identification of these reasons, it provides information about where to start to research.

Finally, it is possible for the researchers to select single trajectories of the birds which they want to compare visually. Because of demonstration purposes for this use case, we selected all six available trajectories. The visualization of these is visible in Fig. 11. There we can recognize that the trajectories of ‘NUME11’ and ‘NUME25’ are almost identical on long distances, besides the deviation in the center section. In contrast, the trajectories of ‘NUME11’ and ‘NUME14’ are completely different. While the first third of the two routes are quite identical, the last two thirds of the two trajectories and also the final destination is completely different. Because the first third of the routes of all birds is quite similar, the very different last two thirds of the routes of ‘NUME11’ and ‘NUME14’ are the reason for the large differences that were calculated in the distance matrix and also the hierarchical clustering dendrogram.

Finally, this information serves as a basis for further investigations by ornithological domain experts.

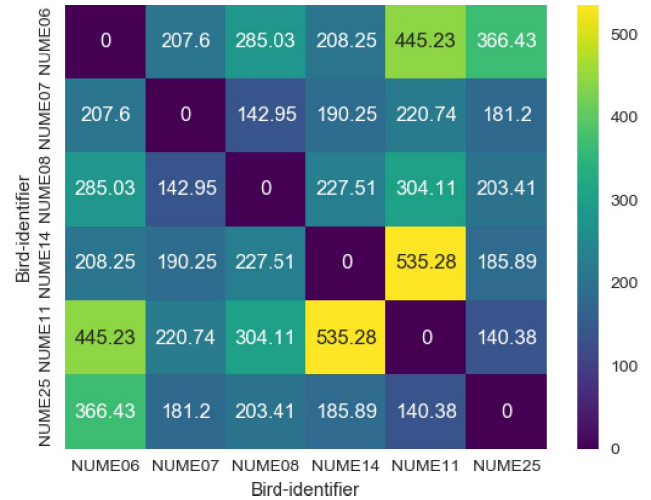


Fig. 9. The calculated distance matrix comparing the differences between the trajectories from the breeding areas to the wintering areas of the birds ‘NUME06’, ‘NUME07’, ‘NUME08’, ‘NUME14’, ‘NUME11’, and ‘NUME25’ of the year 2017. The darker the color, the more similar are the compared trajectories.

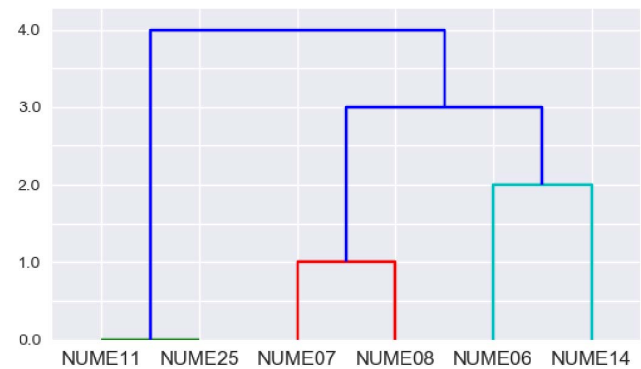


Fig. 10. The hierarchical clustering dendrogram to compare the same birds than in Fig. 9.

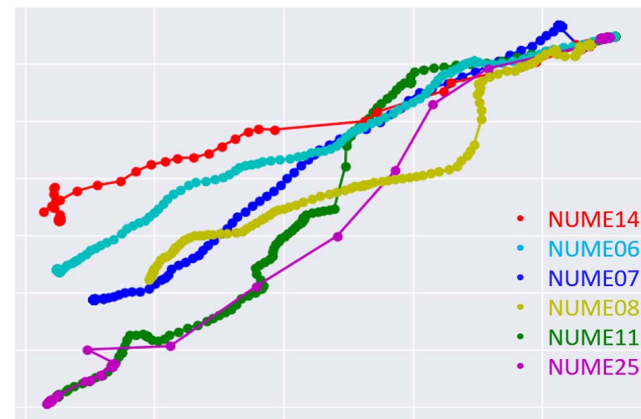


Fig. 11. A visual comparison of the trajectories from the breeding areas to the wintering areas all six curlews (cf. Fig. 9 and Fig. 10) in 2017.

VI. CONCLUSION AND FUTURE PROSPECTS

In this paper, we described the FATBIRD Tool, a new interactive tool which was developed to support ornithologists in their research to evaluate flyways of migrating birds.

After the discussion of the problem and the related work, we explained the concepts of the visualization and the computation of the flyway similarities. Finally, we applied real telemetry data of Eurasian curlews and proved the usefulness of our tool.

However, the FATBIRD Tool is already working, there is still a lot of potential to extend the functionality:

Depending on the hardware type, the transmitters are able to collect different enriched geo-data, e.g. temperature, altitude, etc. Currently we focus on the position of the birds at a certain time. However, the FATBIRD Tool could be extended to also consider this data in the automatic analysis process. Also, the consideration of several data sets, e.g. from different research institutes, could lead to interesting results. It could also be interesting not only to consider the data of a specific species. Different bird species, which have similar annual flyway behaviors could be analyzed together.

Of course, the automatic search for stop-over areas cannot replace the trained view of biologists, who know e.g. the parameters a species needs in its natural habitat, but an application like the FATBIRD Tool can support them by executing a first, very rapid search of stop-overs. This result can then be evaluated by the researchers in more detail.

In this paper, we presented one use case for the telemetry data of Eurasian curlews (*Numenius arquata*). It is planned that the FATBIRD Tool will be further developed with the intent of doing a more extensive analysis in an upcoming publication, considering the biological view. There we will integrate the analysis of further parameters, like ground relief, wind, and weather information.

Furthermore, the scope of application of the FATBIRD Tool is not limited to the research of ornithologists. Other biologists with telemetry studies, for example mammal researchers who work on mammalian telemetry studies, could also benefit from the functionality of our tool. The analysis can also be executed from data gathered by different transmitter technologies.

ACKNOWLEDGMENT

We like to thank the *Bavarian Society for the Protection of Birds* (LBV)⁵ for the provision of the GPS/GSM telemetry data of Eurasian curlews that was used in the use case of this work.

The *Bavarian Society for the Protection of Birds* (LBV) wants to thank the *Bavarian Nature Conservation Funds*⁶ for sponsoring the telemetry study, its cooperation partner *Bavarian State Office for Environment*⁷, the further nature conservation authorities in Bavaria, and all the fieldwork assistants.

REFERENCES

- [1] Landesbund für Vogelschutz in Bayern e.V., “Telemetrie Großer Brachvogel”, available from <https://www.lbv.de/naturschutz/arten-schuetzen/voegel/grosser-brachvogel/telemetrie-grosser-brachvogel/>, accessed on June 13, 2018.
- [2] S. Hahn, S. Bauer, and F. Liechti, “The natural link between Europe and Africa - 2.1 billion birds on migration”, *OIKOS* 118(4), pp. 624–626, 2009.
- [3] M. Willemoes, R. Strandberg, R. H. G. Klaassen, A. P. Tøttrup, Y. Vardanis, P. W. Howey, K. Thorup, M. Wikelski, and T. Alerstam, “Narrow-Front Loop Migration in a Population of the Common Cuckoo *Cuculus canorus*, as Revealed by Satellite Telemetry”, *PLOS ONE* 9(1): e83515, 2014.
- [4] M. L. Vega, M. Willemoes, R. L. Thomson, J. Tolvanen, J. Ruttila, P. Samaš, R. Strandberg, T. Grim, F. Fossøy, B. G. Stokke, and K. Thorup, “First-Time Migration in Juvenile Common Cuckoos Documented by Satellite Tracking”, *PLOS ONE* 11(12): e0168940, 2016.
- [5] Y. Vardanis, R. H. G. Klaassen, R. Strandberg, and T. Alerstam, “Individuality in bird migration: routes and timing”, *Biology Letters* 8(4), pp. 502–505, 2011.
- [6] C. M. Hewson, K. Thorup, J. W. Pearce-Higgins, and P. W. Atkinson, “Population decline is linked to migration route in the Common Cuckoo”, *Nature Communications* 7: 12296, 2016.
- [7] P. C. Besse, B. Guilloet, J. M. Loubes and F. Royer, “Review and Perspective for Distance-Based Clustering of Vehicle Trajectories”, in *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11, pp. 3306–3317, Nov. 2016.
- [8] H. Wang, H. Su, K. Zheng, S. Sadiq, and X. Zhou, “An effectiveness study on trajectory similarity measures”, In H. Wang and R. Zhang (Eds), “Proceedings of the Twenty-Fourth Australasian Database Conference – Volume 137 (ADC ’13)”, pp. 13–22, 2013.
- [9] F. Petitjean, A. Ketterlin, and P. Ganarski, “A global averaging method for dynamic time warping, with applications to clustering.” *Pattern Recogn.* 44, 3 (March 2011), 678–693.
- [10] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, C. Vangenot, “A conceptual view on trajectories, Data and Knowledge Engineering” 65(1), p. 126–146, 2008.
- [11] J. B. MacQueen, “Some Methods for classification and Analysis of Multivariate Observations”. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. Volume 1. University of California Press, pp. 281–297, 1967.
- [12] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: Evangelos Simoudis, Jiawei Han, Usama M. Fayyad (Eds): “Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)”. AAAI Press, pp. 226–231, 1996.
- [13] D. Defays, “An efficient algorithm for a complete link method”, *The Computer Journal* 20(4), pp. 364–366, 1977.
- [14] J.-G. Lee, J. Han, X. Li, and H. Gonzalez. “TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering.” *Proc. VLDB Endow.* 1, 1 (August 2008), 1081–1094.
- [15] Comaniciu, Dorin, and Peter Meer. “Mean shift: A robust approach toward feature space analysis.” *IEEE Transactions on pattern analysis and machine intelligence* 24.5 (2002): 603–619.
- [16] J. Abfal, C. Böhm, K. Borgwardt, M. Ester, E. Januzaj, K. Kailing, P. Kröger, J. Sander, M. Schubert, and A. Zimek, “Knowledge Discovery in Databases. Kapitel 5: Clustering”, 2003. Slides available from <http://www.dbs.ifi.lmu.de/Lehre/KDD/WS1011/skript/kdd-5-clustering1-2.pdf>
- [17] M. Wikelski and R. Kays. “Movebank: archive, analysis and sharing of animal movement data”, available from <http://www.movebank.org>, accessed on September 16, 2018.
- [18] Landesbund für Vogelschutz in Bayern e.V., “Modernste Satelliten-Technik für den Artenschutz”, available from <http://www.lbv.de/naturschutz/satellitentelemetrie>, accessed on June 13, 2018.
- [19] S. Salvador and P. Chan, “FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space”, *Journal of Intelligent Data Analysis*, 11.5 (2007): 561–580.

⁵Landesbund für Vogelschutz in Bayern e.V., <http://www.lbv.de>

⁶Bayerischer Naturschutzfonds, <http://www.naturschutzfonds.bayern.de>

⁷Bayerisches Landesamt für Umwelt, <http://www.lfu.bayern.de/>