Ecological Informatics 63 (2021) 101319

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

A novel method for multi-trajectory reconstruction based on LoMcT for avian migration in population level

Shi Feng[a,b], Qinmin Yang[a,*], Alice C. Hughes[c], Jiming Chen[a], Huijie Qiao[b,*]

[a] State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310007, PR China
[b] Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, PR China
[c] Landscape Ecology Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunnan 666303, PR China

ARTICLE INFO

Keywords:
eBird
Mean-shift
LoMcT
Multi-trajectory reconstruction
GAM

ABSTRACT

Migration is an essential ecological process, and is usually used to describe the seasonal movements between the breeding grounds and wintering areas. During migration, bird populations often disperse in groups and aggregate together due to geographical barriers, topography, seasonal climatic changes, species specific physiology, or other factors. Recording and reconstructing these diverse migratory routes are important for identifying major stopover sites as well as migration bottlenecks which may include key foraging grounds and resting areas, and ensuring high-quality habitat to provide adequate resources. However, good data including individual tracking data are only available for some regions and species large enough to carry a transmitter. Better approaches using observational data are needed to enable better understanding in less-studied regions. To reconstruct and visualize the long-distance avian migration routes with observations from the citizen-science dataset eBird, we developed an interpretive avian multi-trajectory reconstruction framework based on Level-order-Minimum-cost-Traversal (LoMcT) algorithm. This approach uses linear interpolation for missing records, spatial outlier detection for abnormal values, unsupervised clustering by density-based Mean-Shift algorithm for sub-group centroids, LoMcT algorithm based on the distances among centroids, and multi-trajectory reconstruction based on generalized additive models. We have verified the feasibility of our reconstruction method using 15 bird species, and analyzed the trends of the distribution density of birds' population during the long-distance migration cycle. Our analysis could help obtain the important gathering time points and sites in the moving process based on the multiple routes we reconstructed. These can be used in comparisons of multi-trajectory migration strategies between the transoceanic migratory birds and non-transoceanic ones, and provide the ability to understand how species are moving in the absence of individual tracking data to help target conservation better. We have demonstrated that the proposed approach is capable of reconstructing trajectories based on observational citizen-science data.

1. Introduction

Birds play irreplaceable roles in ecosystems, acting as predators and preys across trophic groups, and as seed-dispersers as well as pollinators (Whelan et al., 2015). Currently, there are over 10,000 described species of bird, of which more than 20% are migratory birds (Roskov et al., 2020). Bird migration is a regular seasonal movement between breeding grounds and wintering grounds driven by environment, climate, atmospheric conditions, photoperiod, and other factors (Gwinner, 1990; Horton et al., 2016; Ramenofsky and Wingfield, 2007; Shamoun-Baranes et al., 2017; Somveille et al., 2020). The migration trajectory of birds can effectively reveal the behavior of birds and demonstrate the importance of certain areas during their migration. Assessing bird migration time, routes, population changes, and the conditions of breeding places and winter habitats can inform the protection of migratory birds.

A variety of techniques have been used to study bird migration including bird ringing, radio tracking, GPS tracking, satellite telemetry, isotopic analysis, and so on (Nilsson and Sjöberg, 2016; von Hünerbein et al., 2000; Klaassen et al., 2010; Fiedler, 2009; Seifert et al., 2016; Hobson, 1999). However, traditional individual tracking technology with devices or markers may not be suitable for most small-sized species

* Corresponding author.
E-mail addresses: qmyang@zju.edu.cn (Q. Yang), huijieqiao@gmail.com (H. Qiao).

https://doi.org/10.1016/j.ecoinf.2021.101319
Received 11 December 2020; Received in revised form 9 April 2021; Accepted 9 April 2021
Available online 14 May 2021

(Hobson, 1999), and isotopes could provide a limited resolution in analysis (Inger and Bearhop, 2008). So they may not be able to provide enough information for the migration routes in population level, and thus may not enable effective measures for the conservation of stopover sites which may be critical to species long-term viability. Therefore, more affordable and accessible approaches are needed for further insights into migration trajectories in less-studied regions, where little or no exact tracking data exists.

The recent development and popularization of citizen science provide a large volume of bird observation data shared via platforms such as eBird (Sullivan et al., 2009), iNaturalist (https://www.inaturalist.org/; iNaturalist, 2021), and so on. The growing spatial and temporal resolution of citizen science data makes it possible to study migration patterns (Supp et al., 2015; Weisshaupt et al., 2021; Schubert et al., 2019). It has already been used to effectively research questions about bird migration. Bounas, et al. described the migration trajectories of the red-footed falcon in the Mediterranean region from 2010 to 2017 with the help of the occurrence information from eBird to find the changes of its migration patterns and explore the stopover sites during the migration based on distribution modeling in 2020 (Bounas et al., 2020). Li J, Hughes A C, and Dudgeon D. constructed the first biodiversity hotspot map for the migratory waders along the East Asian-Australasian Flyway based on MaxEnt (Phillips et al., 2006) with the observations of 57 species from eBird (Li et al., 2019). Also, with the help of eBird, La Sorte, F. A. and Graham, C. H found out the phenological coupling between vegetation greenness and seasonal avian migration in North America (La Sorte and Graham, 2021). These models show the potential to study avian migration trajectories and patterns with discrete observation data from eBird. However, they are weak in providing detailed dynamic information about bird distributions during migration, which may be necessary for targeted management and ensuring the protection of key sites.

In recent years, further advances in data mining, machine learning, and other computer technologies have helped to reconstruct migratory routes. Avian trajectories have been widely studied by machine learning, statistical methods, and a variety of other methods. One such approach used the generalized additive models for location, scale, and shape (GAMLSS) model (Rigby and Stasinopoulos, 2005) to estimate the migration trajectories of 118 migratory species in the Western Hemisphere to study the convergence of migration strategies caused by geographical factors in 2016 (La Sorte et al., 2016). Somveille et al. have reconstructed bird migrations over the past 50,000 years with mechanical models and provided a baseline understanding of how species may respond to modern climate change (Somveille et al., 2020). Walker, J. and Taylor, P. used Bayesian regression models and range-wide data of 28 species in the U.S. and Canada from e-Bird to reconstruct and estimate the migratory trajectories of spring migration/breeding, fall migration, and wintering (Walker and Taylor, 2020).

The approaches mentioned above have greatly helped us improve our knowledge about bird migratory routes in the absence of individual specific data on species level, however, avian migration often follows multiple routes, and may not utilize the shortest route (Alerstam, 2001). So, simulations which can better account for and simulate these variable patterns in population level are needed. Existing models typically consider an average population movement or used one centroid over time, both of which may cause a bias towards better-known areas and fail to provide the details which may be needed for targeted conservation actions.

Here, we have developed a novel multi-trajectory reconstruction algorithm for avian migration in population level focusing on Level-order-Minimum-cost-Traversal (LoMcT) algorithm based on the observation dataset from eBird. This enables us to reconstruct multiple trajectories for locality-only observation records. The multi-trajectory reconstruction focuses on the dynamic changes of aggregation and separation in population level during the long migration circle. Realizing the grouping function and reconstructing the multiple trajectories could better simulate the birds' real migration process and contribute to the protection and cognition. Also, the method may help further research in areas where technologically advanced tracking approaches have rarely been applied and thus provide further insights into bird migratory processes in population level, including flyways which have received less attention and species which may not be easy for accurate individual tracking.

2. Material and methods

2.1. Data acquisition

We downloaded the observations from the eBird database within the Western Hemisphere between 170°E to 25°W longitude and 62°S to 85°N latitude in 2018. The avian occurrence data are organized into lists, including observed species with longitude, latitude, and observation date for each species. Here, we selected 15 species of birds whose observation date loss rates are less than 20%, and they are active in the North and South America with a total of 170,152 records in an annual circle. We used MATLAB (MATLAB, 2020) and R (R Core Team, 2020) for data mining and trajectory reconstruction.

2.2. Migration trajectory reconstruction

We proposed a bird trajectory reconstruction method based on the LoMcT algorithm to estimate the migration trajectories effectively, which aimed to find the paths of different groups belonging to one avian species population during migration. Our method includes four sections: data preprocessing, clustering, grouping, and trajectory reconstruction (Fig. 1). A flow chart with more details about the main components of the approach is shown in Fig. S1.

2.2.1. Preprocessing

In addition to species name, we used three data components from the eBird indices: latitude, longitude, and observation date. The repeated latitude and longitude observations on the same day were deleted to reduce redundancy. Missing values were estimated by linear interpolation and outlier records were cleaned according to the spatial local deviation factor (SLDF; Zhang and Wang, 2011). The following is a detailed description of the data preprocessing steps.

2.2.1.1. Missing record interpolation. On some occasions (e.g., Table S1), both time and location information were lost (the occurrence information on 2018/12/17). So, we need to use the time and location data which are k-nearest neighbors of the missing value and fill out the missing information with linear interpolation for the data continuity. We set k as 2 after testing higher numbers which have little effect. Then the processed data after interpolation were stored for further use.

2.2.1.2. Outlier detecting based on SLDF. When it comes to the outlier records, we calculated the local deviation degree of each observation data and detected the outliers with SLDF (Algorithm description S1), which is effective for handing large scale spatial datasets.

After ordering all the SLDF values of the observations of one day, we find and clean the m objects with higher SLDF values in the top 20% after testing several times for effective outlier detection.

2.2.2. Clustering based on mean-shift

The purpose of this section is to identify the main clusters for each day from the preprocessed dataset. Since we focused on the movement of the main populations of a given bird species, we chose unsupervised density-based cluster analysis to implement the approach. We used high-density clusters to identify the grouping process during the migration circle because the high-density sets mean more information and fewer outliers. Also, an unsupervised algorithm is necessary because we can't

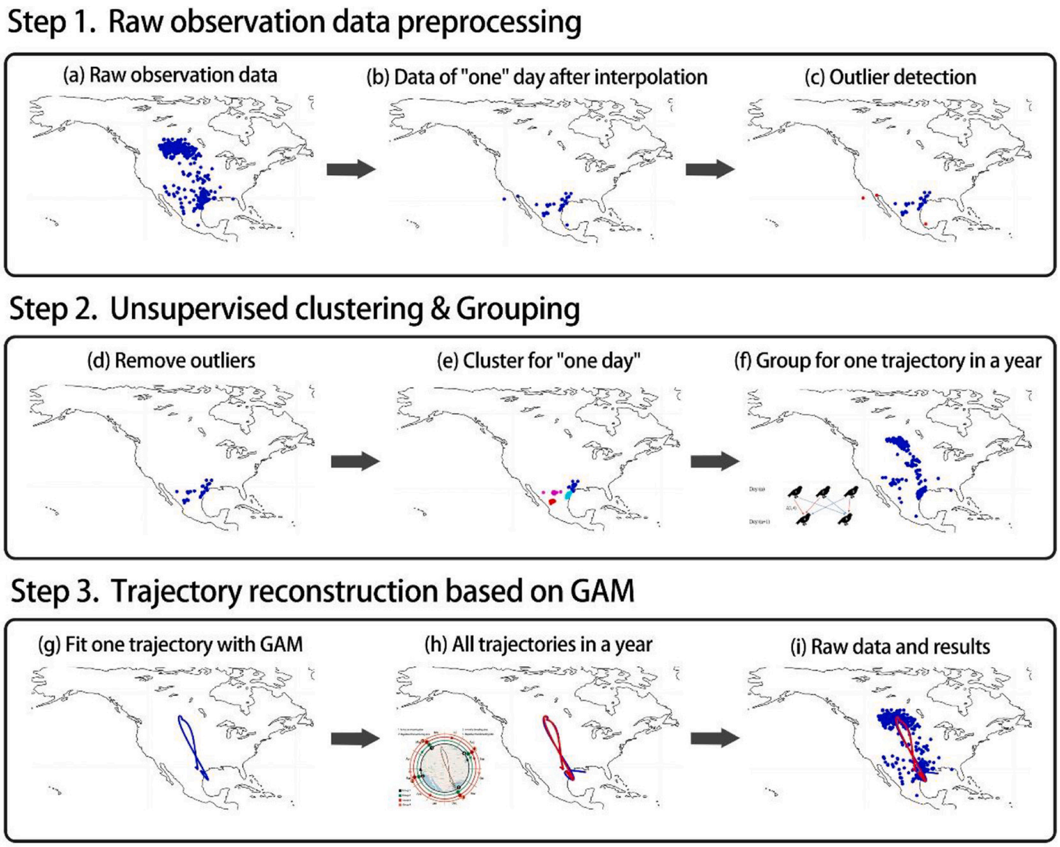


Fig. 1. Methodology for avian multi-trajectory reconstruction. (a) Raw observation data of *Anthus_spragueii* from eBird in 2018; (b) estimating out the missing observations via the *k*-nearest neighbors linear interpolation algorithm; (c, d) detecting and removing outliers with space local deviation factor (SLDF); (e, f) identifying the daily groupings with Mean-Shift cluster algorithm and grouping the birds with Level-order-Minimum-cost-Traversal (LoMcT) algorithm; (g-i) with the help of generalized additive model (GAM), we constructed trajectories of *Anthus_spragueii* in 2018.

confirm the number of bird populations in advance during migration, and how it varies across the cycle.

Based upon the requirements above and considering the common-used clustering algorithms, we chose Mean-Shift clustering algorithm (Derpanis, 2005), which is an unsupervised clustering method based on density with no need to define the number of clusters in advance, and has no restrictions on the shape of clusters. The effect of the algorithm of pushing the cluster center closer to the densest area is also very satisfactory. This process conforms to the needs of data-driven tasks, and performs well with large data volumes. All of these could meet the characteristics of our dataset well: large data volume, unsupervised, obvious density differences, and irregular shape. The density can be estimated by the number of data points in a region of a specified radius r. Its nuclear function helps us come closer to the highest-density clusters which may form during migration as the number of iterations increases. In each iteration, the algorithm continuously calculates the offset means from the center of the circle to the center of mass, and then approaches the center of mass gradually. The density within the drift circle is proportional to the number of data points in it. Once the center of mass is reached, the algorithm updates the center of mass and continues to move the circle closer to a higher density area. When the circle reaches the target, it finds itself unable to find more data points no matter which direction it drifts in. At this time, it is assumed to be in the densest area. The algorithm satisfies the final condition, and we get one intended clustering result. The flow chart is shown in Fig. S2.

However, it should be noted that the distribution of daily observation data of one species varied greatly as patterns of data collection vary across space and time. In order to solve this problem and ensure the stability of clustering results, we aggregated the observation data of 7 adjacent days to form the cluster dataset for one day based on the rolling window analysis (Zivot and Wang, 2007). By doing this, our method could stabilize the route construction and reduce sensitivity to uncertainty in areas with heterogeneous data collection.

2.2.3. Level-order-Minimum-cost-Traversal algorithm (LoMcT)

Birds usually migrate in different groups rather than one single population during their migration process, so we aim at reconstructing their trajectories for each subgroup by the LoMcT algorithm.

The LoMcT is the core and novel component of the multi-trajectory reconstruction method. Firstly, the centroid distances among adjacent days are calculated, and the clustering centroids of one day are considered to be on one level. Without accounting for climate, terrain, or other factors, we assume that the shortest path between the centroids has the minimum migration cost (Algorithm description S2), which forms the basis of the following calculations. Then, the flight path could be determined based on the distance D between the unsupervised cluster centroids of adjacent two days. The number of trajectories is consistent with the largest number of unsupervised cluster centroids in one day during the trajectory fitting cycle. The pseudo-code of the grouping algorithm is shown in Pseudo code S1.

After traversing the distance of two centroids of adjacent two days, we considered the two centroids with the minimum distance to be the same population group as shown in Fig. S3.

To be aware of the special circumstances, we need the help of the cluster centroids on Day (n-1) as shown in Fig. S4. When the number of cluster centroids on Day (n) is less than the one on Day (n-1) and Day ($n + 1$), we cannot judge the grouping conditions by relying solely on the centroids of adjacent two days, so we need to use the centroids' distance calculation between Day (n-1) and Day ($n + 1$) for further grouping determination.

To show our clustering algorithms more intuitively, it can be indicated by the schematic intuitively with a total of seven cases in Fig. S5.

2.2.4. Trajectory reconstruction based on the generalized additive model

The generalized additive model (GAM; Algorithm description S3; Hastie and Tibshirani, 1990) extends the generalized linear model by estimating the relationship between the dependent and independent variables by fitting nonparametric functions, and the concept is a nonparametric regression method in which potentially non-parametric functions can be processed by data smoothing techniques. We need to note that the additive terms are also allowed to be parametric function in GAMs, such as polynomials.

When generating the GAM models, we fit longitude and latitude with the date respectively because they are assumed to be independent from each other. In this way, we 1) matched the coordinates of longitude and latitude on the same day according to the chronological order, 2) obtained the trajectory coordinates of longitude and latitude under the time dimension, and 3) generated the migration trajectory results finally by linking the coordinates up.

3. Method verifications

3.1. Virtual data based verification

To verify the accuracy of the algorithm, we built a virtual species and artificially defined the group situation in the absence of bird calibration trajectories. The test dataset is the observation data of the virtual species across one year. The orange and blue lines are the calibration of the virtual species migration trajectory, and the green and red ones are the reconstructed trajectories through our grouping algorithm (Fig. S6).

In the process of grouping verification, considering that two or more paths represent a dichotomy, the validity of the two-groups can adequately prove the validity of the multi-group grouping algorithm. We assume that the virtual species located in the area named B as the breeding ground and the area named A as the wintering ground. The virtual birds migrate from area A to area B in January, and finally return to A through two trajectories colored orange and blue to complete the migration activities during one year. The test dataset is observations for virtual species within a year and is organized in a humanly defined calibration grouping with the grouping results being displayed in orange and blue curves on the map. The reconstructed results are displayed as green and red lines on the map when the test dataset is input our grouping algorithm. As can be seen from the figure, the grouping algorithm can effectively finish the grouping of the virtual migration observation data.

3.2. Empirical data based verification

The ideal reconstructed trajectories should across most of the high-density observation points of a given species. To verify it in empirical datasets, we invoked LoMcT algorithm to 15 species to reconstruct their population-level paths (Fig. S7–Fig. S11), then we generated a buffer around the trajectories via constructing a 95% confidence interval (Fig. 2), and used the one-sided independent two-sample t-test to test whether the density of the points inside of the buffer was significantly higher than the ones out of the buffer. Results are shown in Table S2.

4. Results

4.1. Grouping conditions

Detailed analysis of migratory trajectories was performed on 15 American species in 2018. The 15 species showed five different patterns of grouping across their migratory cycle. No species showed only one single migratory trajectory, and three species (*Helmitheros_vermivorum, Elaenia_parvirostris, Euphagus_carolinus*) showed two trajectories (Fig. S7). One species (*Empidonax_difficilis*) showed three migration trajectories (Fig. S8), four species had four trajectories (*Anthus_spragueii, Catharus_minimus, Icterus_spurius, Muscisaxicola_capistratus,* Fig. S9), and a further four species (*Calcarius_ornatus, Cardellina_canadensis, Contopus_cooperi, Vireo_philadelphicusi*) showed five migratory trajectories (Fig. S10). The highest number of migratory trajectories noted was six, which was found in two species (*Ammospiza_nelsoni, Passerina_amoena*) in Fig. S11.

4.2. Trajectories reliability assessment

Because of the scattered distribution of the observation data from eBird caused by less accessible areas and fewer engaged bird-watchers, it is difficult for us to make use of all the occurrence information. Therefore, our grouping algorithm focuses more on the areas where the density of the observation data points is sufficiently high for trajectory grouping to capture the movement information of the main populations.

As shown in Fig. 2, the red full line is one trajectory of *Anthus_spragueii*. The gray ribbon area is the buffer of the trajectory created by 95% confidence interval. We calculated the density of the points inside and outside of the buffer. The P-value of the one-sided independent two-sample t-test is $0.0064 < 0.01$, which meant the density of points inside of the buffer was significantly higher than the ones outside. In this way, we preliminarily proved that our trajectories could effectively characterize the major populations with higher densities in the observational

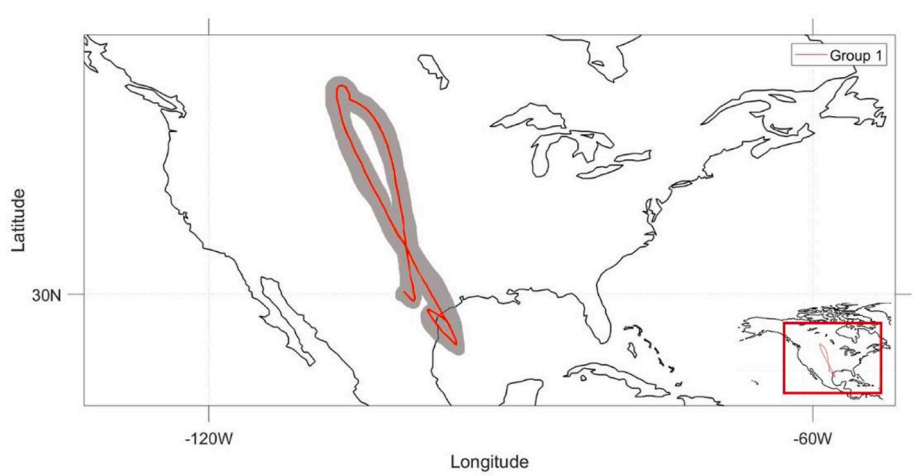


Fig. 2. A map for one trajectory of *Anthus_spragueii*. The red full line is the trajectory. The gray ribbon area is a buffered pathway with 95% confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data. 11 of the 15 species in our study showed the same pattern as *Anthus_spragueii*, while the rest of 4 species didn't give out a significant difference (Table S2).

5. Discussion

With the increasing global data availability and the greater pressure on natural habitats, there is a greater need for more accessible approaches to understanding species migratory routes. However, the type of data often used for such analysis is generally limited to a subset of species and regions. Based on that, we developed an avian multi-trajectory reconstruction method based on LoMcT, which could help reconstruct avian migration routes with observations from eBird. However, due to the limitations of eBird being an unsupervised citizen data, the dataset is limited by data availability, visibility, observation bias, and other factors, which can lead to some deviation, such as local closed loop in the trajectory reconstruction results, in areas where there is too much observation data or too little observation data. Through the effective analysis of avian occurrence information, the trajectory of bird migration can be simulated as the basis to analyze and visualize the stopover sites distribution, and the important time nodes of reaching wintering grounds, breeding grounds as well as stopover sites during avian migration. The simulated trajectories can help us evaluate the distribution of the movement sequences and the dynamic changes in the moving process, and was found to be credible by experimental trials and calibration.

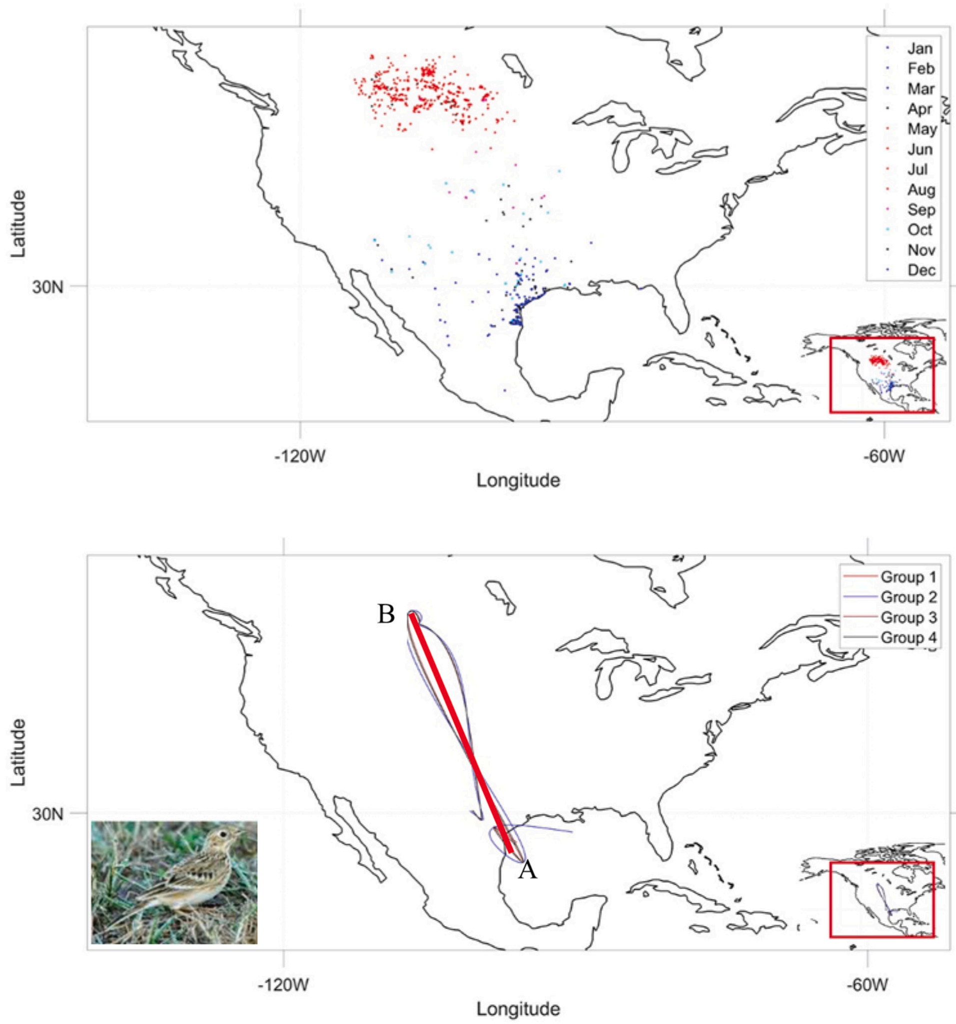
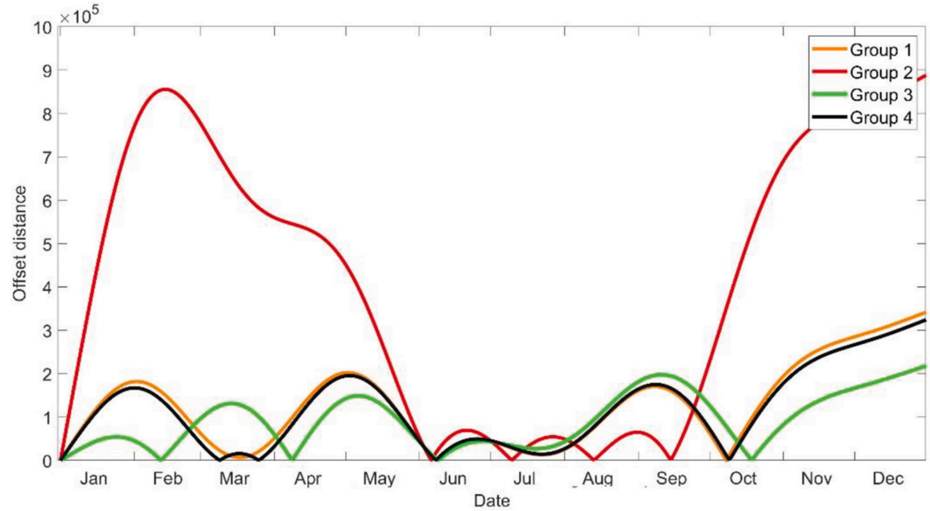
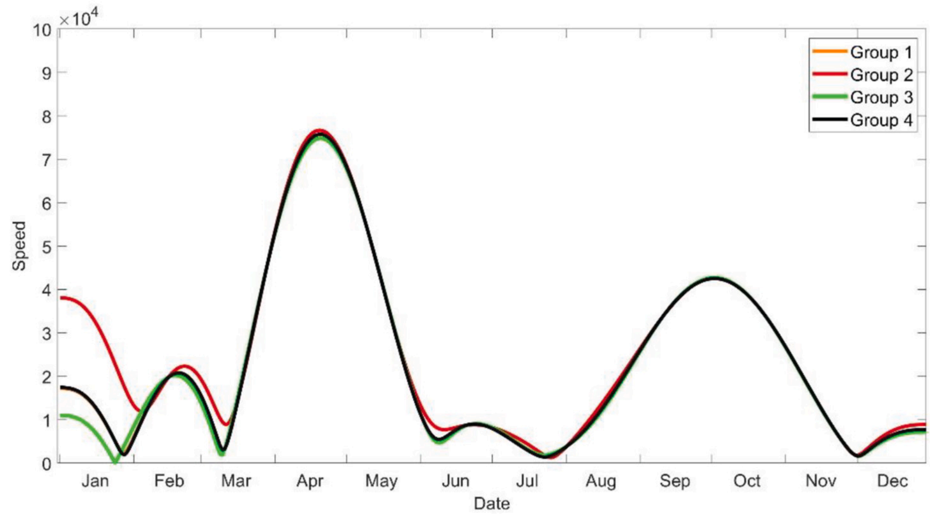


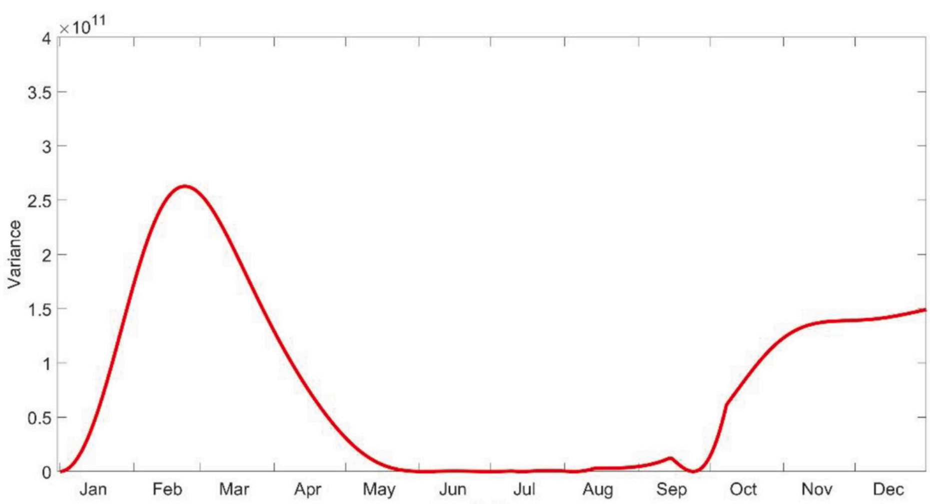
Fig. 3. (a) A migration axis and its migration trajectories schematic for *Anthus_ spragueii*. The line segment AB represents the migration axis between the winter range and the breeding range. We can get the changing trend of the offset degree graph between each route and the migration axis AB by working out its each-day Euclidean offset distances in the migration process. (b) Offset distances graph of *Anthus_spragueii* in 2018. The horizontal axis is time, and the vertical axis is the every-day offset Euclidean distance under the Mercator coordinate system between the position on each trajectory and the migration axis. The graph shows that the birds reach the breeding range in early June. The offset distances of each route change periodically. (c) Speed of *Anthus_spragueii* in 2018. By the speed graph above, we can see the birds arrive at the breeding ground in early June with decreasing speed and leave in late August with increasing speed. (d) Variance of each-day positions in the trajectories of *Anthus_ spragueii* in 2018. By the variance graph between the trajectories above, we can see the birds arrive at the breeding ground in early June and leave in late August with the minimum variance. Also, we find the birds arrive at a stopover site in late September. The variance graph shows the aggregate information of the population during the whole migration circle according to the distribution of different trajectories. They scatter with large variances and cluster with smaller ones.



(b) Offset distances graph of *Anthus_spragueii* in 2018.



(c) Speed of *Anthus_spragueii* in 2018.



(d) Variance of each-day positions in the trajectories of *Anthus_spragueii* in 2018.

Fig. 3. (*continued*).

5.1. Offset distance and their variance analysis

It seems necessary for birds to choose altitude and speed to maintain optimal migration direction and minimize energy consumption and risk. In order to analyze the strategy of flight behavior and migration, we calculated the daily Euclidean offset distances between the trajectories and the migration axis between the winter range and the breeding range during the migration cycle.

For example, *Anthus_spragueii* is a small songbird that breeds in the mixed-grass prairies of North America. It overwinters in the south-western United States and northern Mexico, and its summer habitat is primarily native grasslands in the north-central meadows of the United States and Canada. But the trajectory result shows that it does not utilize the route of the shortest distance.

In the static trajectory diagram map of *Anthus_spragueii* in Fig. 3a, the migration axis between the overwintering grounds and the breeding ground is reflected as line segment AB for the shortest distance. Daily Euclidean offset distances in the migration process enable calculation of the offset degree graph between each route and the shortest trajectory. In Fig. 3b, the horizontal axis is time, and the vertical axis is the every-day offset Euclidean distance under the Mercator coordinate system between the position on each trajectory and the migration axis. The graph shows that the birds reach the breeding range in early June with decreasing speed shown in Fig. 3c., and the offset distance of the migration trajectories from the migration axis changes periodically with roughly in the straight flight direction. Alerstam found that migratory birds often choose to make detours to avoid crossing barriers such as large areas of sea, sand, or ice (Alerstam, 2001). Considering the actual environment and their feeding habits, we demonstrate that *Anthus_spragueii* prefers to migrate across grassland for food-providing rather than using potential routes straightly through forests.

Furthermore, we figure out the daily population density of *Anthus_spragueii* during the yearly migration cycle (Fig. 3d). Combined with the variance graph between the trajectories in Fig. 3d, we can analyze the time of their arrival at the breeding ground (early June) and they leave in late August with the minimum variance. In the same way, the birds arrive at the stopover site in late September. By the variance graph, the aggregate information of the population during the whole migration cycle can be reflected clearly according to the distribution of different trajectories. The birds scatter with large variances and cluster with smaller ones. Fig. 4 shows the whole migration cycle along with the information we get above.

Variance graphs of birds' trajectories show the degree of scattering or concentration within each species population during the migration cycle. Birds disperse at high levels of variance and concentrate when the variance is low. Different groups of the same species converge in the same special geographical location likely denoting topographic drivers or high food availability, especially in breeding grounds or other stop-over sites. Birds come together in the spring for breeding, and they travel separately to the winter grounds after breeding.

By identifying areas with low variance, we can identify the areas of key importance for feeding, breeding, wintering, and stopover sites which may show important time nodes during the movement phases of migration.

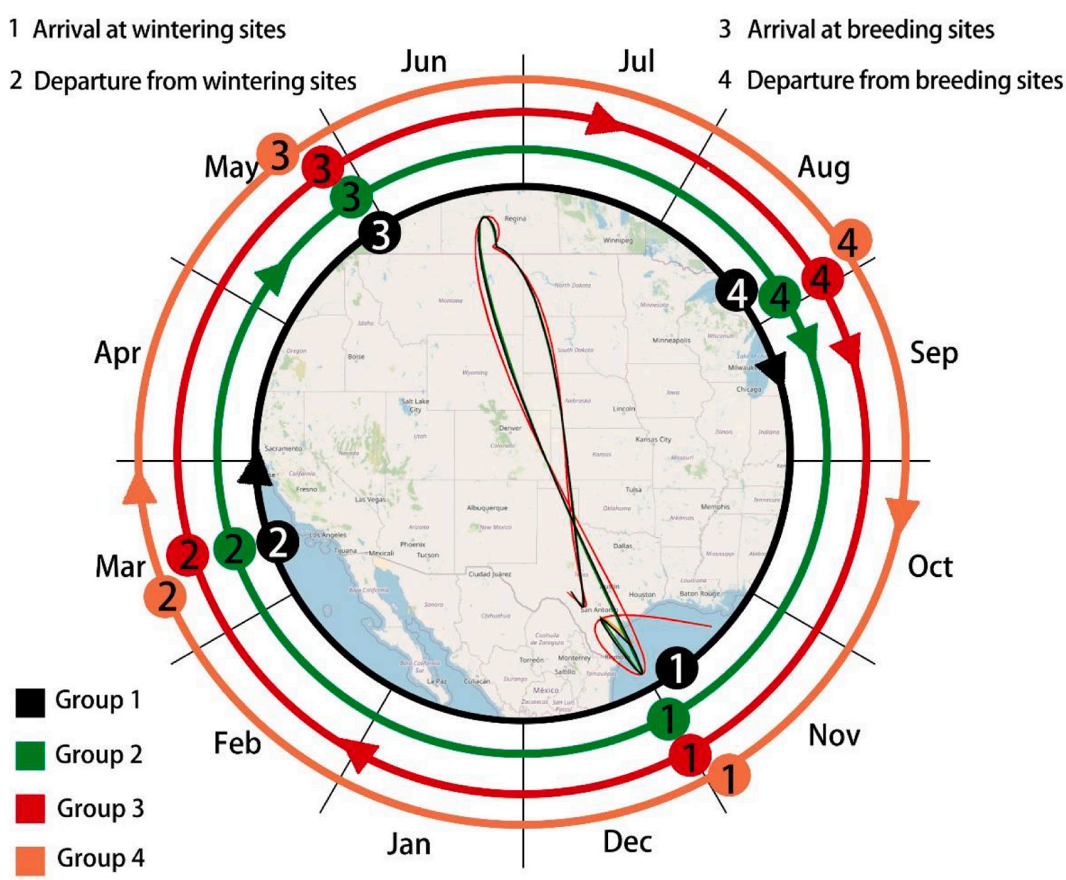


Fig. 4. A diagram of the whole migration cycle of *Anthus_spragueii* in 2018. Number 1 represents the time of their arrival at wintering sites. Number 3 represents the time of arrival at breeding sites. Number 2 represents the time of departure from wintering sites. Number 4 represents the time of departure from breeding sites. Four different colors represent four different groups during the migration cycle.

5.2. Analysis of migration strategies

The migration trajectories of birds can be categorized into broad or narrow front migration (Fig. S12) according to the breadth of the migration coverage area (Mead, 1983; Qin et al., 2008). Narrow-front migration refers to the movement of birds between breeding sites and wintering grounds with a narrow concentrated migratory route, which is typically narrower than the range used for breeding or overwintering. In broad-front migration, the paths of birds between the breeding area and the wintering area are more scattered, and cover a wider area than the narrow-front migration (Mead, 1983; Qin et al., 2008).

15 species are further divided into broad-front migration or narrow-front migration according to the offset degree. When the average offset distance away from the migration axis of the trajectories is greater than a Euclidean distance of 500,000 m, the bird migration belongs to the broad-front movement. When the offset Euclidean distance away from the migration axis is less than 500,000 m, the bird migration is narrow-front migration. These two types of migration trajectories are evenly distributed among the 15 species migration trajectories. The migration strategy results are shown in Table S3.

Seven species, including *Ammospiza nelsoni, Cardellin canadensis, Catharus minimus, Contopu cooperi, Euphagus carolinus, Icterus spurius,* and *Passerina amoena* are classified as broad-front migrants. The other eight species, including *Anthus_spragueii, Calcarius ornatus, Elaenia parvirostris, Empidonax difficilis, Geothlypis philadelphia, Helmitheros vermivorum, Muscisaxicola capistratus,* and *Vireo philadelphicus* are classified as narrow-front migrants. Birds with more groups are more likely to have broad-front migrations. The differences in the migration mode could help us obtain the important survival information of birds in migration, such as understanding the importance of stop-over sites during migration and ensuring that these key areas are maintained, especially in migratory bottlenecks.

Different groups of the same species show different offset distances from the migration axis. According to the variance of different offset distances, we can judge the distribution of the whole population in the entire migration cycle. When the variance is significant, then wide migratory pathways may be used, while low levels of variance indicate very narrow migratory routes. The distribution of different populations of the same species could be obtained across one year using the variance graph (Fig. 3d), which reflects the level of dispersal of each bird species in the face of geographical, climatic, and other environmental changes, and therefore may be used to pinpoint key stopover sites. This analysis also advances from previous methods (Bounas et al., 2020; La Sorte et al., 2016; Li et al., 2019; Somveille et al., 2020; Walker and Taylor, 2020), by better capturing the dynamic process of avian migration including population distribution bottlenecks across the year.

5.3. Moving forwards

Globally, there are five global flyways (UNEP/CMS, 2014), and huge efforts have gone into tracking and reconstructing the routes of the American flyways, but to a less extent gone for the African/west Eurasian flyway; data for species on other flyways is also much more limited. This represents a serious threat to the future survival of many species, as, for example, populations of some species on the East-Asian Australasian flyway have shown decrease rates at 5–9% per year (Li et al., 2019). Targeted conservation for many species has previously been impossible with limitation of both the capacity and resources in certain regions, and the low weights of some species which precludes the use of most tracking devices. Our approach provides many details of any analysis in terms of identifying the intensity of use across routes as well as their stopovers, and therefore enables targeted conservation efforts to ensure these regions are protected. Furthermore, in areas where data is particularly sparse, we also could help reconstruct routes, and additionally understand their dependencies over the course of migration (such as flying over grasslands) which may be essential for such species to be a viable weight for reproduction or over-wintering once they reach breeding or over-wintering grounds. By utilizing our approach within these less-known regions, we hope to enable new insights into the routes used in much less-known global flyways, and therefore enable better conservation, management, and further research.

6. Conclusion

Advancing global understanding of avian migratory routes calls for making better use of easy-to-acquire and access data. In this work, we detail a new method for multi-trajectory reconstruction method based on the LoMcT algorithm with observations from eBird, which combines with data mining and graph theory to enable sub-grouping trajectory reconstruction during bird migration. As bird databases grow, our method may be applicable to an increasing number of species all over the world, helping fill in gaps for small-bodied and little-known species.

What's more, our trajectory reconstruction work is carried out in an “ideal state” without taking the limits of accessibility due to terrain, prevailing wind, or other factors into account. Terrain, and climate information could be added to such analysis and would further enhance our understanding of bird migration. It's worth noted that our occurrence information is usually from areas with good observation conditions, and we need actively subsample such data to develop our trajectories. Inaccessible areas, such as mountain passes and oceans are challenging to produce exact migration trajectories in this study, especially for the species across the Atlantic, the Gulf of Mexico. Combining limited individual tracking data, such as from radio-tracked birds, as a complement to trajectory analysis would provide one way to assay accuracy across such regions. Also, as an important part of our multi-trajectory reconstruction algorithm, more available clustering methods would be tested and assessed in our framework for further improvement to make our approach more effective.

In synthesis, we showcase a helpful method for mapping multiple migration trajectories, and their stops based on observation data. The increased level of routes accuracy provides more accurate data to target conservation and management, and provides a tool that will greatly enhance our understanding of species on less known and studied flyways across the globe.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2021.101319.

Data and code available

https://github.com/shifengshierya/LoMcT

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The research was supported by Chinese National Natural Science Foundation (31772432) and Zhejiang University State Key Laboratory of Industrial Control Technology (ICT2021B47).

References

Alerstam, T., 2001. Detours in bird migration. J. Theor. Biol. 209 (3), 319–331. https://doi.org/10.1006/jtbi.2001.2266.

Bounas, A., Solanou, M., Panuccio, M., Barišić, S., Bino, T., Erciyas-Yavuz, K., Barboutis, C., 2020. Mining citizen science data to explore stopover sites and spatiotemporal variation in migration patterns of the red-footed falcon. Curr. Zool. 66 (5), 467–475. https://doi.org/10.1093/cz/zoaa008.

Derpanis, K.G., 2005. Mean shift clustering. Lecture notes. [Online]. 32, 1–4. http://www.cse.yorku.ca/%7Ekosta/CompVis_Notes/mean_shift.pdf (accessed 12 March 2019).

Fiedler, W., 2009. New technologies for monitoring bird migration and behaviour. Ringing Migr. 24 (3), 175–179. https://doi.org/10.1080/03078698.2009.9674389.
Gwinner, E., 1990. Circannual rhythms in bird migration: control of temporal patterns and interactions with photoperiod. In: Bird Migration. Springer, Berlin, Heidelberg, pp. 257–268.
Hastie, T.J., Tibshirani, R.J., 1990. Generalized additive models, Vol. 43. CRC Press, London.
Hobson, K.A., 1999. Tracing origins and migration of wildlife using stable isotopes: a review. Oecologia 120 (3), 314–326. https://doi.org/10.1007/s004420050865.
Horton, K.G., Van Doren, B.M., Stepanian, P.M., Farnsworth, A., Kelly, J.F., 2016. Seasonal differences in landbird migration strategies. Auk Ornithol. Adv. 133 (4), 761–769. https://doi.org/10.1642/AUK-16-105.1.
iNaturalist. Available from https://www.inaturalist.org (accessed 15 January 2021).
Inger, R., Bearhop, S., 2008. Applications of stable isotope analyses to avian ecology. Ibis 150 (3), 447–461. https://doi.org/10.1111/j.1474-919X.2008.00839.x.
Klaassen, R.H., Strandberg, R., Hake, M., Olofsson, P., Tøttrup, A.P., Alerstam, T., 2010. Loop migration in adult marsh harriers *Circus aeruginosus*, as revealed by satellite telemetry. J. Avian Biol. 41 (2), 200–207. https://doi.org/10.1111/j.1600-048X.2010.05058.x.
La Sorte, F.A., Graham, C.H., 2021. Phenological synchronization of seasonal bird migration with vegetation greenness across dietary guilds. J. Anim. Ecol. 90 (2), 343–355. https://doi.org/10.1111/1365-2656.13345.
La Sorte, F.A., Fink, D., Hochachka, W.M., Kelling, S., 2016. Convergence of broad-scale migration strategies in terrestrial birds. Proc. R. Soc. B 283 (1823), 20152588. https://doi.org/10.1098/rspb.2015.2588.
Li, J., Hughes, A.C., Dudgeon, D., 2019. Mapping wader biodiversity along the east Asian - Australasian flyway. PLoS One 14 (1), e0210552. https://doi.org/10.1371/journal.pone.0210552.
MATLAB, 2020. Version 9.9.0. The MathWorks Inc, Natick, Massachusetts.
Mead, C.T., 1983. Bird Migration. Newnes Books, Feltham.
Nilsson, C., Sjöberg, S., 2016. Causes and characteristics of reverse bird migration: an analysis based on radar, radio tracking and ringing at Falsterbo, Sweden. J. Avian Biol. 47 (3), 354–362. https://doi.org/10.1111/jav.00707.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.
Qin, J.Y., Chen, M.H., Xiang, Z.F., 2008. On seasonal migratory model and study method of aves. J. Anhui Agric. Sci. 36 (17), 7278–7281. https://doi.org/10.3969/j.issn.0517-6611.2008.17.097.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Ramenofsky, M., Wingfield, J.C., 2007. Regulation of migration. Bioscience 57 (2), 135–143. https://doi.org/10.1641/B570208.
Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. J. R. Stat. Soc.: Ser. C: Appl. Stat. 54 (3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x.
Roskov, Y., Ower, G., Orrell, T., Nicolson, D., Bailly, N., Kirk, P.M., Bourgoin, T., DeWalt, R.E., Decock, W., van Nieukerken, E.J., Penev, L., . Species 2000 & ITIS Catalogue of Life, 2020-12-01. Digital resource at. www.catalogueoflife.org. Species 2000: Naturalis, Leiden, the Netherlands. ISSN 2405-8858.
Schubert, S.C., Manica, L.T., Guaraldo, A.D.C., 2019. Revealing the potential of a huge citizen-science platform to study bird migration. Emu-Austral Ornithol. 119 (4), 364–373. https://doi.org/10.1080/01584197.2019.1609340.
Seifert, N., Haase, M., Van Wilgenburg, S.L., Voigt, C.C., Schmitz Ornés, A., 2016. Complex migration and breeding strategies in an elusive bird species illuminated by genetic and isotopic markers. J. Avian Biol. 47 (2), 275–287. https://doi.org/10.1111/jav.00751.
Shamoun-Baranes, J., Liechti, F., Vansteelant, W.M., 2017. Atmospheric conditions create freeways, detours and tailbacks for migrating birds. J. Comp. Physiol. A. 203 (6), 509–529. https://doi.org/10.1007/s00359-017-1181-9.
Somveille, M., Wikelski, M., Beyer, R.M., Rodrigues, A.S., Manica, A., Jetz, W., 2020. Simulation-based reconstruction of global bird migration over the past 50,000 years. Nat. Commun. 11 (1), 1–9. https://doi.org/10.1038/s41467-020-14589-2.
Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142 (10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006.
Supp, S.R., Sorte, F.A.L., Cormier, T.A., Lim, M.C., Powers, D.R., Wethington, S.M., Graham, C.H., 2015. Citizen-science data provides new insight into annual and seasonal variation in migration patterns. Ecosphere 6 (1), 1–19. https://doi.org/10.1890/ES14-00290.1.
UNEP/CMS, 2014. A Review of Migratory Bird Flyways and Priorities for Management. In: CMS Technical Series No. 27. UNEP / CMS Secretariat, Bonn, Germany.
von Hünerbein, K., Hamann, H.J., Rüter, E., Wiltschko, W., 2000. A GPS-based system for recording the flight paths of birds. Naturwissenschaften 87 (6), 278–279. https://doi.org/10.1007/s001140050721.
Walker, J., Taylor, P., 2020. Evaluating the efficacy of eBird data for modeling historical population trajectories of North American birds and for monitoring populations of boreal and Arctic breeding species. Avian Conserv. Ecol. 15 (2) https://doi.org/10.5751/ACE-01671-150210.
Weisshaupt, N., Lehikoinen, A., Mäkinen, T., Koistinen, J., 2021. Challenges and benefits of using unstructured citizen science data to estimate seasonal timing of bird migration across large scales. PLoS One 16 (2), e0246572. https://doi.org/10.1371/journal.pone.0246572.
Whelan, C.J., Şekercioğlu, Ç.H., Wenny, D.G., 2015. Why birds matter: from economic ornithology to ecosystem services. J. Ornithol. 156 (1), 227–238. https://doi.org/10.1007/s10336-015-1229-y.
Zhang, T.Y., Wang, X.L., 2011. Outlier detection algorithm based on space local deviation factor. Comput. Eng. 14. https://doi.org/10.3969/j.issn.1000-3428.2011.14.096.
Zivot, E., Wang, J., 2007. Modeling financial time series with S-Plus® (Vol. 191). Springer Science & Business Media.