

RESEARCH ARTICLE

Long-term analysis of persistence and size of swallow and martin roosts in the US Great Lakes

Maria Carolina T. D. Belotti¹ , Yuting Deng¹ , Wenlong Zhao² , Victoria F. Simons¹ , Zezhou Cheng² , Gustavo Perez² , Elske Tielens³ , Subhransu Maji² , Daniel Sheldon² , Jeffrey F. Kelly³  & Kyle G. Horton¹ 

¹Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA

²College of Information and Computer Sciences, University of Massachusetts-Amherst, Amherst, Massachusetts, USA

³Department of Biology, University of Oklahoma, Norman, Oklahoma, USA

Keywords

Aeroecology, animal aggregations, bird migration, machine learning, radar ornithology

Correspondence

Maria Carolina T. D. Belotti, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, USA. Tel: +1 (970) 491-6675; Fax: +1 (970) 491-0279; E-mail: maria.belotti@colostate.edu

Funding Information

This work was supported by the following National Science Foundation awards: Colorado State University (#2017554), University of Massachusetts-Amherst: (#2017756), Oklahoma University: (#2017582).

Editor: Vincent Lecours

Associate Editor: Matthew Van Den Broeke

Received: 8 August 2022; Revised: 28 November 2022; Accepted: 5 December 2022

doi: 10.1002/rse2.323

Remote Sensing in Ecology and Conservation 2023; **9** (4):469–482

Introduction

Swallows and martins (family: *Hirundinidae*) are highly specialized aerial insectivores that, like other species that rely on insects as their main energy source, have been facing regional and continental-wide declines in their breeding populations in the US and Canada (Rosenberg et al., 2019; Sauer et al., 2017; Smith et al., 2015). Some of the potential drivers of these declines include habitat

Abstract

In this study, we combined a machine learning pipeline and human supervision to identify and label swallow and martin roost locations on data captured from 2000 to 2020 by 12 Weather Surveillance Radars in the Great Lakes region of the US. We employed radar theory to extract the number of birds in each roost detected by our technique. With these data, we set out to investigate whether roosts formed consistently in the same geographic area over two decades and whether consistency was also predictive of roost size. We used a clustering algorithm to group individual roost locations into 104 high-density regions and extracted the number of years when each of these regions was used by birds to roost. In addition, we calculated the overall population size and analyzed the daily roost size distributions. Our results support the hypothesis that more persistent roosts are also gathering more birds, but we found that on average, most individuals congregate in roosts of smaller size. Given the concentrations and consistency of roosting of swallows and martins in specific areas throughout the Great Lakes, future changes in these patterns should be monitored because they may have important ecosystem and conservation implications.

loss, decreases in prey abundance, impacts from environmental contaminants and phenological mismatches due to climate change (Cox et al., 2019; Hallmann et al., 2014; Kardynal et al., 2020; Spiller & Dettmers, 2019). Particularly, in the case of the migratory species in this family, it is still a challenge to understand how these drivers of decline affect different stages of their annual cycles and how these effects interact between seasons to produce the overall pattern observed in their breeding populations

(Holmes, 2007; Marra et al., 2015; Webster et al., 2002). Gathering metrics of population size and the variation therein across broad spatial and temporal scales would have important conservation implications, while also providing insights into the fundamental ecology of these species.

Recent advances in tracking technologies have allowed us to address some of these questions, but they are limited to individual-level data over relatively short periods (Supp et al., 2021; Wikelski et al., 2007). Due to these limitations, tools such as community science data (iNaturalist, 2022; Sullivan et al., 2009; Wikiaves, 2022), stable isotope analysis (Imlay et al., 2018) and weather radars (Dokter et al., 2010, 2018; Horton et al., 2019) have recently been ascending as alternatives to observe population-wide patterns at large spatial and temporal scales. Particularly, in the case of swallows and martins, weather radars are an invaluable tool to collect data about their communal roosting behavior, though until recently we lacked the tools to leverage the potential of this data source.

After their breeding season, from mid-July until the end of September, species of the *Hirundinidae* family congregate in large communal premigratory roosts, which can tally more than 100 000 birds at the end of each day (Allen & Nice, 1952; Bent, 1942). The early departure of thousands of birds in synchrony from these roosting places is so conspicuous that they can easily be identified in weather radars, where they appear as donut-shaped rings that disperse outwards in the hours near sunrise (Russell et al., 1998). While the primary ecological importance of these roosts is largely unknown, especially prior to a long distance migration, these behaviors offer a unique and important snapshot to monitor population changes. The main challenge of using weather surveillance radars to obtain data about these aggregations is manually finding the roosts in the collection of millions of radar scans generated across decades. Additionally, establishing systematic criteria for classifying the multiple sources of noise and weather contamination inside the roosting signature are complicated by the vast diversity of patterns captured by the radars, especially at large spatial scales.

However, machine learning approaches have the potential to alleviate this barrier by disentangling the roost signatures and automatically discerning their locations and the boundaries of their dispersal (Cheng et al., 2020). With such tools, we can now analyze a dataset that spans two decades and can provide high spatial and temporal sampling frequency. We decided to test this approach in the Great Lakes region, where multiple species of *Hirundinidae* have been shown to roost. Large congregations of Purple Martins *Progne subis*, Tree Swallows *Tachycineta bicolor*, Barn Swallows *Hirundo rustica*, Bank Swallows

Riparia riparia and Cliff Swallows *Petrochelidon pyrrhonota* are regularly seen in this region between July and September (Bridge et al., 2015; Burney, 2002; Kelly et al., 2012). Additionally, we have found historical accounts tracing back to 1910 of roosts gathering Tree Swallows and Bank Swallows at the Black River Bay, in Lake Ontario, near Toledo, Ohio, at Cedar Point and also in Ohio (Bent, 1942, pp. 396 and 420). With more than two decades of radar remote sensing data available throughout this region, we set out to use these newly available tools to identify the locations of *Hirundinidae* roosts and provide some of the first estimates of macro-scale abundance.

With these data, we set out to quantify how consistently the population in our study region returned to the same regions to roost over the years (which we refer to as 'population-level site fidelity'), and whether site usage consistency is related to roost size. Investigations of the spatial and temporal patterns of site fidelity can help us understand the role played by competition and social dynamics as drivers of nonbreeding habitat selection in migratory bird species (e.g., see Marra & Holmes, 2001; Sherry & Holmes, 1996). Additionally, if site fidelity is driven by adults who return to good habitats based on experiences from previous seasons, then examining site fidelity patterns can provide indirect information about adult survival and overall population dynamics (Cresswell, 2014). Lastly, populations that strongly rely on a specific region during their migratory journey might be particularly vulnerable to anthropogenic impacts on that site, making them a species of conservation concern (Warkentin & Hernández, 1996).

Finally, an important step in understanding the ecological implications of such postbreeding aggregations is characterizing within-season variations in their size distribution. Until now, the techniques available to quantify the sizes of these large roosts suffered from observer variability, and thus were hard to reproduce from year to year, or even on a daily basis (Erwin, 1982; Frederick et al., 2003). With the use of weather radars, we can not only estimate the number of birds within each roost, but also examine within- and between-season changes. We describe the wide range of roost sizes found in the Great Lakes region and how this distribution changes as the overall population increases throughout the season.

Materials and Methods

Data source

Within the contiguous US, the National Weather Service, the Federal Aviation Administration and the United States Air Force jointly operate a network of 143S-band weather

radars that conduct volume scans of the atmosphere at regular intervals (5–10 min). Each scan is composed of multiple 360° horizontal sweeps of the radar antenna taken at different elevation angles. Scans are organized in a 3D array following the CfRadial convention in which cells, indexed by the horizontal angle of the antenna (each angle is called an azimuth), by distance from the radar (called range) and by elevation angle corresponding to measurements of the returned signal from a voxel of the volume covered by the radar beam (see Dixon & Lee, 2021 for a detailed specification of the CfRadial standard).

For the purposes of this paper, we downloaded all radar scans from the Great Lakes region (12 radar stations) available between June 1st and October 31st across 21 years (2000–2020, see Fig. 1A and Appendix S2, Fig. S4, for data completeness information). Each day, we processed the data in the period starting 30 min before local sunrise and ending 90 min afterwards (an average of 16.2 scans per day). Reflectivity factor and radial velocity from the first sweep of each scan were rendered in Cartesian grids with 600×600 pixels (1 pixel = 0.5 km).

These matrices were then used as input to the pretrained model described below. Each station's local sunrise time was calculated using PyEphem and pytz Python packages.

Automated detection and screening

The machine learning system developed by Cheng et al. (2020) operates in two steps—detection and tracking—following the paradigm proposed by Ren (2008) (see Fig. 1B). In the detection step, given some candidate proposal regions defined by positions in a radar scan and reference roost sizes, we trained a Faster R-CNN (acronym for Region-based Convolutional Neural Network) detector to predict the probability that a roost of each candidate size exists near each reference position (Ren et al., 2015). A high predicted probability indicates a roost detection according to the model. The detector also predicts bounding boxes for the detected roosts. To ease implementation, we leveraged the 'R101-FPN 3x' Faster R-CNN model from Detectron2, an open-source library in PyTorch (Wu et al., 2019). We picked 45 anchors at each position and otherwise used the default model

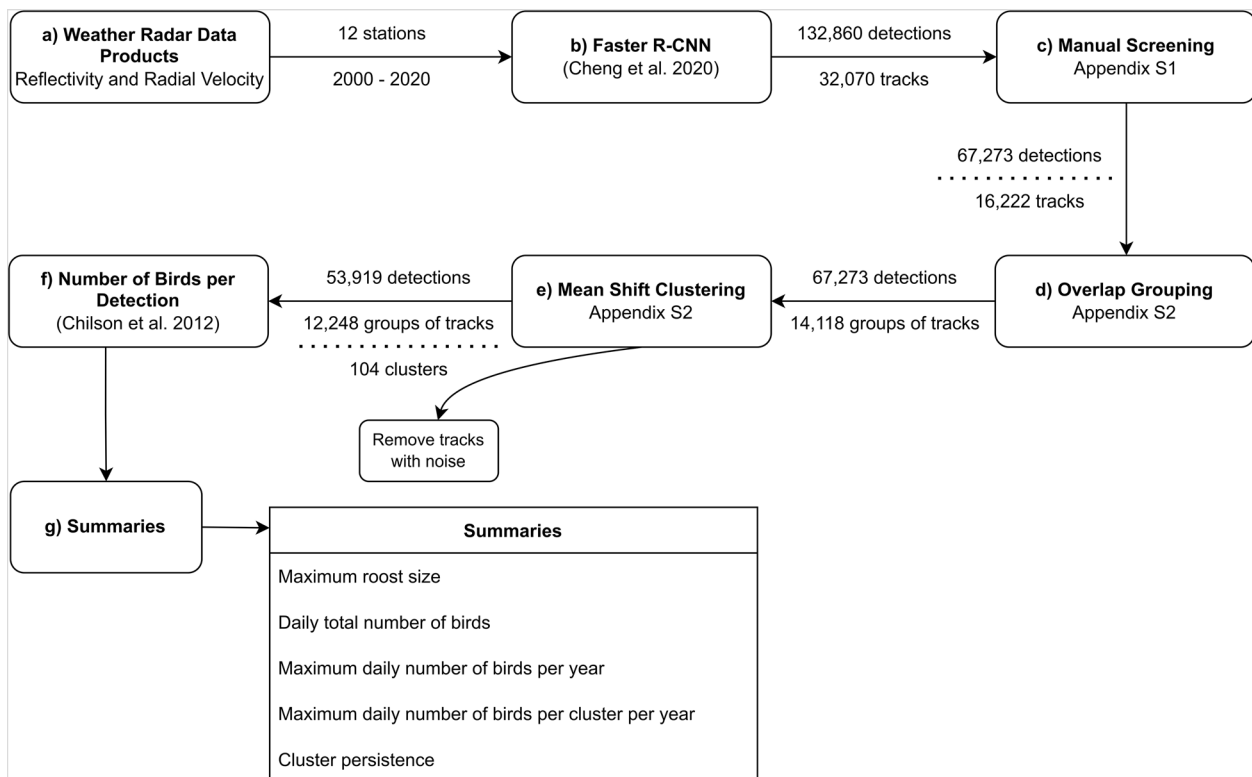


Figure 1. Diagram representing the data processing steps, beginning with the radar data downloaded from Amazon Web Services and ending at the extraction of biological summaries from the bird counts on each roost. At each step, we report the number of detections (single frames found in one scan), tracks (sequences of detections in subsequent scans) and groups of tracks (groups of overlapping tracks that correspond to a single roost). Each group of tracks corresponds to one roost's morning dispersal event.

configurations from the library. We loaded the model checkpoint trained for the COCO detection task from Detectron2 and fine-tuned it using the annotated training dataset developed for the paper by Laughlin et al. (2016). Since different annotators of the training data differ in their annotation styles, we adopted the factors from Cheng et al. (2020) to scale bounding box annotations, thus alleviating annotator biases. We used an Adam optimizer with a learning rate of 0.001 to train the detector (Kingma & Ba, 2014). For each scan, we kept 100 top-scoring roost detections with predicted probability scores of at least 0.05.

In the tracking stage, since the same roost is expected to show up in temporally consecutive scans, we reproduce the strategy described by Cheng et al. (2020) to predict the association among roost detections from individual scans. Finally, when possible, we removed rain false positives using dual-polarization data (specifically correlation coefficient) and wind farms according to the US Wind Turbine Database (Hoen et al., 2018). While polarimetric products provide a robust filter for the detection of weather contamination, they are not available throughout the entire temporal extent of our study (only from 2013 onward), and thus additional manual screening was necessary for removing potential nonbiological contaminants.

We manually picked a low score threshold (0.05) to ensure high recall regardless of the low precision of the predictions since we wanted to make sure that the pipeline would capture all roosts. To remove the false positives and detections with weather or noise contamination from our dataset, we developed a screening protocol (Fig. 1C) used by three different screeners (MB, YD and VS) to evaluate all the predicted positive detections (screening 1 year of data for a given station takes on average c. 45 min, see Appendix S1 for details of the screening protocol). We classified each true positive track, composed of a series of consecutive detections, as either clear, weather contaminated, having more than 20% anomalous propagation, containing other sources of unknown noise or tracking failures, in which the tracking algorithm failed to follow the roosts (e.g., see Fig. 2). Days that had more than 50% of the reflectivity renderings filled with weather or anomalous propagation were discarded from the analysis.

Overlap correction and mean shift clustering

Some of the weather radars selected for this study have clear overlapping ranges (see Fig. 3A). If a roost occurred inside this overlapping region, it could be detected by more than one station, which would result in duplicated tracks and inflated estimates of the number of birds. To

remove this bias, for each track, we computed the union of all its bounding boxes. We then grouped our dataset by day and calculated, for each pair of tracks, the overlap ratio, which we defined as the ratio between the area of the intersection of each pair of tracks and the area of their union (Fig. 1D). If the overlap ratio was greater than 0.2, the tracks were considered to be in the same group (see Appendix S2 for details about the selection of ratio threshold).

To establish regions of repeated use throughout the years, we captured the first detection of each group described above and used it as input to the mean shift clustering algorithm (implementation by Pedregosa et al., 2011, Fig. 1E), which finds density-based clusters while making minimal assumptions about cluster shape (see Appendix S2 for detailed methods of parameter choice). With this approach, we assume that regions consistently used by birds to roost would have a higher density of points and thus be included in the same cluster. We chose to use the first detection of each group because we observed that these were the most precise representatives of the actual roost's location. The cluster labels were then extrapolated to other points within each group, which is why the final clusters overlap.

Up until this postprocessing stage, we included the first detection of all the tracks classified by the screeners as true positives (bounding boxes that contained a verified roost dispersal), even the ones that were noise-contaminated or that the algorithm failed to track. These so-called 'contaminated' tracks provide useful information about roost location, but they cannot be used to calculate estimates of number of birds, since it cannot be assumed that all the scatterers in them are exclusively birds. For this reason, from this step onwards they were removed from the analysis.

Reflectivity extraction and postprocessing

Following manual screening to remove contaminants, we used coordinates of the center of each roost and its radius, both estimated by the detector, to obtain the bounding boxes of each departure event. We used these data to extract the raw reflectivity factor values from Level II radar data in polar coordinates (each voxel's location is defined by its range and azimuth) across all elevation angles of each full scan. To estimate the number of birds per roost, we followed the approach established by Chilson et al. (2012): we converted reflectivity factor (Z_e) measured at each range interval sample, originally in dBZ, to linear scale (mm^6/m^3), and then transformed it into reflectivity, which can be interpreted as the density of scatterers in the atmosphere (units are cm^2/km^3). We can then multiply this result by the

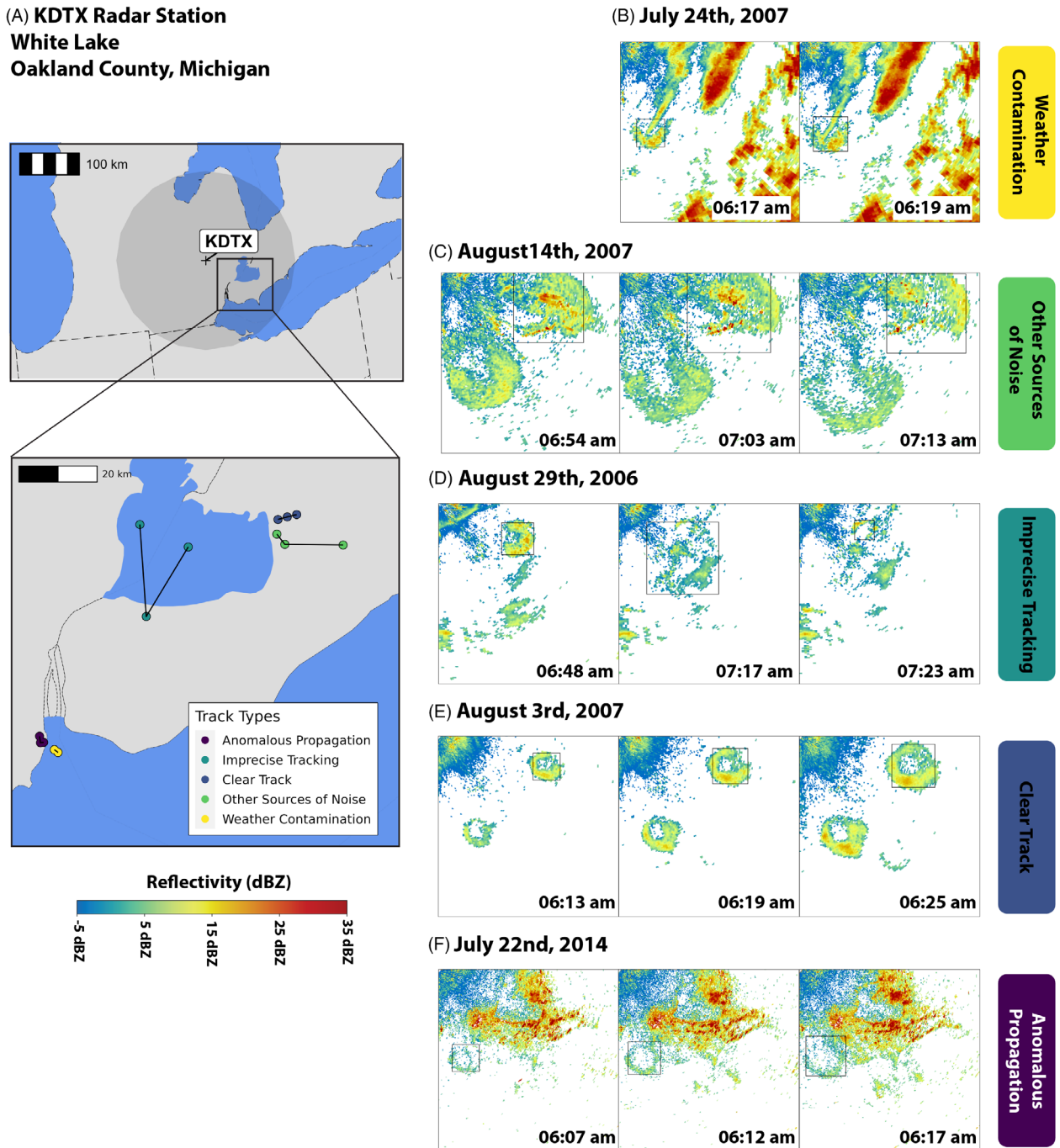


Figure 2. (A) A map of the 150 km buffer (shaded gray) around the KDTX station, located in Oakland County, Michigan, and an inset representing the locations of each detection shown on the right side of the panel; (B) A weather-contaminated track; (C) A track containing sources of noise other than weather and anomalous propagation (in this case, ground clutter); (D) A subset of detections from a track where the machine learning pipeline did not build the bounding boxes correctly; (E) A clear track that captures the roost without noise; (F) a track contaminated with noise from anomalous propagation in the upper right corner.

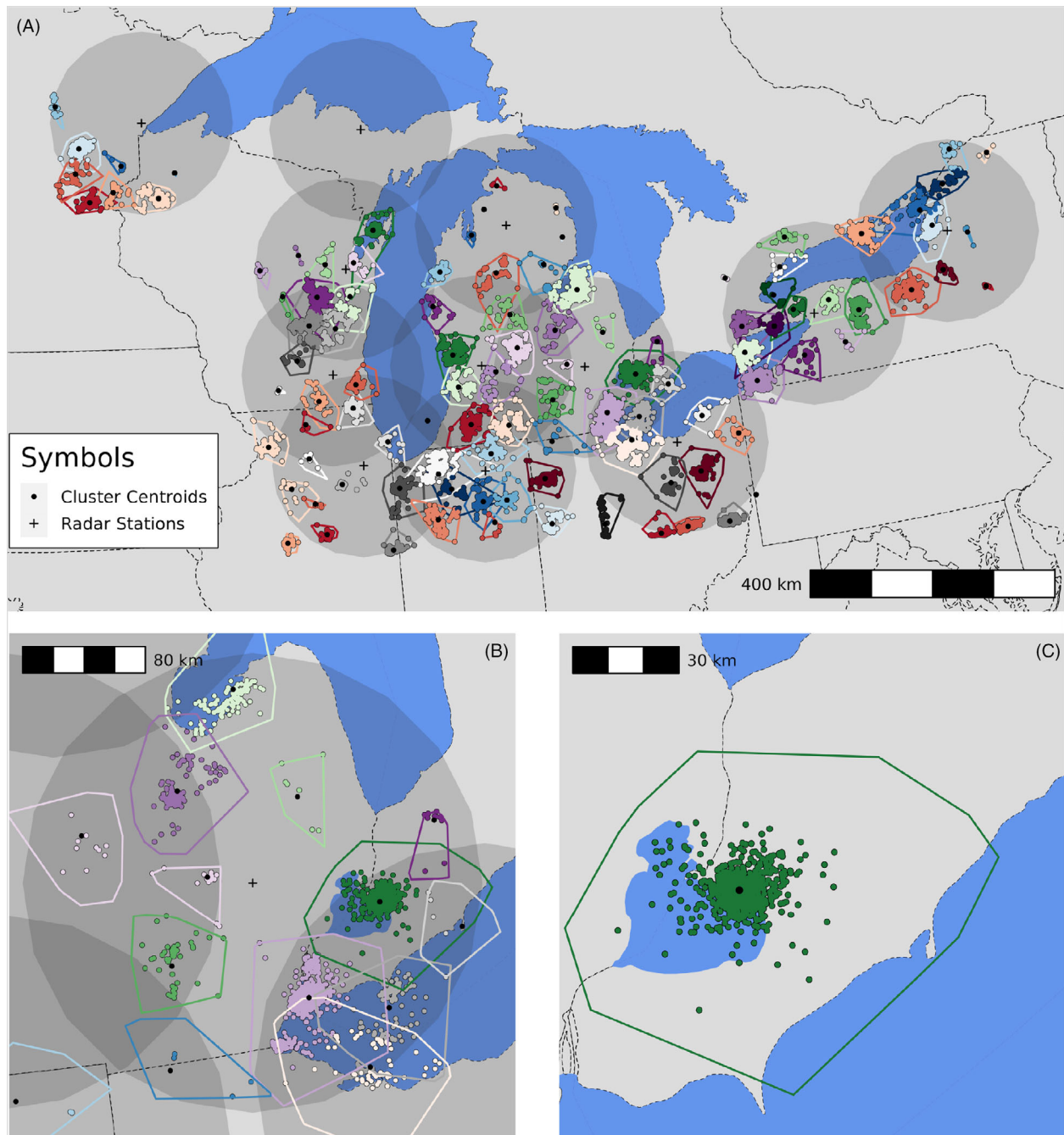


Figure 3. (A) Map of the study region, with circular radar ranges in shaded gray (150 km radius). Clusters and respective convex hulls are represented in the same color. (B) Inset showing clusters in the extent of a single station (KDTX) located in White Lake, Michigan. (C) Inset with a single cluster located on the north-east of Lake Saint Clair. This region saw martins and swallows across all 21 years of our series.

theoretical volume sampled by the radar beam at each range interval, and divide it by the radar cross section (RCS) of the scatterers assumed to be found in the roost (Fig. 1F). In summing the results of these operations in the region inside each bounding box, we obtain a count estimate for that sweep.

We opted to obtain conservative estimates of the number of birds by choosing Purple Martins as our benchmark radar cross section, since they are the largest species of *Hirundinidae* found in North America. Though previous studies suggest that the majority of roosts in the region are Purple Martins (Bridge et al., 2015; Kelly

et al., 2012), our ability to ascertain the species identification is still somewhat limited using radar data alone. The radar cross section of Purple Martins can be obtained from their mass—51 g, see Dunning (2008)—adopting the relationship between mass and RCS proposed by Horton et al. (2019): $\log(\text{RCS}) = 0.699 \times \log(\text{mass})$. Bank Swallows, the smallest species to participate in such aggregations, would yield counts c. 2.6 times higher, since their mass is 13 g according to Dunning (2008).

The volume sampled by the radar is assumed to be shaped like a truncated cone with an axis aligned with the antenna's peak power axis, cut by two parallel planes at each range gate. For data before the Super Resolution upgrade, the cone's apex angle was assumed to be 1° in width. After the 2007–2008 upgrade, we assumed an elliptical cone of 1° vertical beam width and 0.5° horizontal beam width (Torres, 2007). In order to further clean the data from spurious contamination, we removed voxels inside the detection's bounding box if the reflectivity factor in them exceeded 30 dBZ (Horton et al., 2019).

To prevent nonbiological scatterers from the upper levels of the atmosphere from artificially increasing our estimate of the number of birds, we only analyzed sweeps that had heights below 5000 m at the 150 km range. We counted birds across the entire vertical structure of the roost departure by grouping the multiple sweeps of each detection in bins of 1° intervals according to their elevation angle. Afterward, we took the mean bird count of each bin and summed the counts from all available bins. With this procedure, we hoped to avoid summing regions in the airspace sampled twice by consecutive sweeps that had elevation angles with $<1^\circ$ interval between them.

Analysis

To describe the peak activity in each group of tracks obtained as described above, we extracted the maximum number of birds in each series of detections. We then aggregated groups of tracks by day and summed the maximum bird counts obtained from roosts occurring on the same day. We did this assuming that we can sum the maximum bird count for each group without double counting since we have corrected most overlapping tracks. In doing so, we have a measurement of the maximum number of birds per day in the study region. We further aggregate this by cluster and year and compute the maximum number of birds per cluster per year, which we interpret as the peak number of birds throughout the region each year. To obtain a cluster's persistence, we sum the number of years when each cluster was detected (Fig. 1G).

Because radar stations had incomplete archives for some years, we have restricted the dataset to the 225 pairs of station-years, out of a total of 246 (91.4%), in which we had at least 130 days of sampling (see Appendix S2, Fig. S4 for data completeness information). All the analyses conducted in the rest of the paper will describe results from this filtered dataset. In Figure 6, however, we show clusters detected every year when their nearest stations were operational, regardless of the number of days, to highlight the full temporal scope of the dataset.

We used the peak activity in each group of tracks to measure changes in the distribution of roost sizes throughout the season. We tested the null hypothesis of no change in mean roost size between the months of July, August, September and October with the Kruskal–Wallis test (implemented in base R version 4.1.3). Since the Kruskal–Wallis test suggested that the means were in fact different (p -value <0.001), we performed pairwise comparisons with the Wilcoxon rank sum test using the Holm method to adjust the P -values for multiple comparisons (implemented in the base R version 4.1.3).

Results

Detections, tracks and groups of tracks

Across a total of 940 641 radar scans from 2000 to 2020, we obtained 67 273 detections, or snapshots of departure events, which the machine learning pipeline grouped into 16 222 tracks representing sequences of snapshots of the same dispersal. Of those tracks, 85.32% (13 841) were considered clear, 0.62% (100) had at least 20% of anomalous propagation inside the bounding boxes, 10.27% (1667) were contaminated by weather or other unknown sources of noise and 3.78% (614) were bad-tracks. After correcting for radar range overlaps, we had a total of 14 118 groups of tracks, each one corresponding to a unique dispersal event. Each track group had an average of 4.76 detections (SD = 3.45). The first detection of each group of tracks was then used as input to the mean shift clustering algorithm, which yielded 104 clusters with silhouette coefficient of 0.703 (represented in Fig. 3). After removal of tracks contaminated by noise or poorly captured by the machine learning pipeline, we had 12 248 groups of clear tracks, with an average of 4.40 detections (SD = 3.12).

Every year in which we had more than 130 sample days per station, we extracted the day when the maximum number of martins and swallows was detected in the area covered by the NEXRAD network in the Great Lakes region. This number was highest in 2007 (on August 17th), when 536 836 birds across 19 clusters were detected by the radars. In contrast, the peak number of birds was lowest in 2013, with a total 143 625 birds

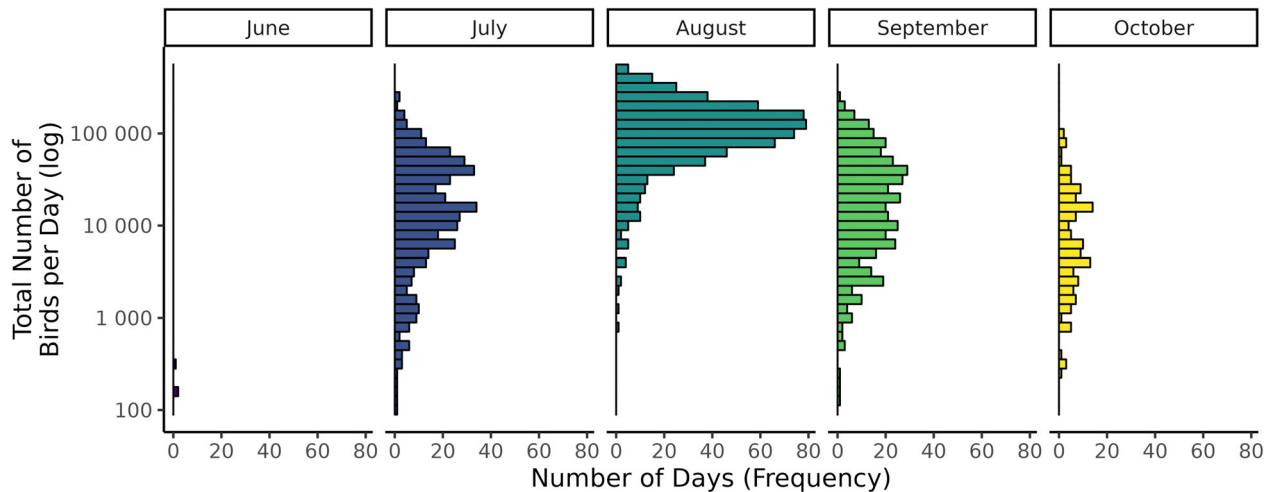


Figure 4. Histograms of the daily sum of the number of birds (on a logarithmic scale) throughout the Great Lakes study region, grouped by month, for pairs of station-year with data available on more than 130 days. Note how the total count increases in August.

spread across 13 clusters on 13 August 13. In Figure 4, we show, in each month, the frequency in days of each bird count bin (across all 21 years). The largest single roost in our series occurred on September 14th, 2005, in Fond du Lac, Wisconsin, when we estimated a departure of 187 736 birds.

Roost size distribution

The data suggest that small roost sizes are more common than larger ones since the daily size distribution throughout the years is heavy tailed (see Fig. 5A). Our results indicate that, in August, when the overall population is higher, there is an increase in the frequency of smaller sized roosts paired with the formation of roosts of size larger than 100 000 birds (Fig. 5C). Pairwise comparisons showed statistically significant (p -value < 0.001) differences in the mean group size of all pairs of months except for September and October (p -value = 0.11). These tests reveal an interesting asymmetry between the beginning and the end of the season, with the average group size in July (mean 5388, SD = 7998) being smaller than the average observed in September (mean 9011, SD = 14 123).

During July, September and October, more than 60% of the population can be found in smaller roosts, ranging from a few individuals up to 25 000 birds (Fig. 5B). This pattern changes in August, when roosts of larger size start to gather larger percentages of the population. Even though roosts of more than 100 000 individuals occur in every year of our 21-year series, they do not gather the majority of the *Hirundinidae* population in the Great Lakes region, since the proportion of the population aggregated in smaller roosts is larger.

Persistence and roost size analysis

Of 104 clusters, which we interpret as regions of frequent use by swallows and martins, 98 of them were detected in more than 1 year (94.2%), 11 of which occurred every year from 2000 to 2021 (10.6%, see Fig. 6). We found three clusters that were active on all years when the radar stations closer to them were active: two of them occurred in east Lake Ontario for 17 years (near KTYX), and one in northern Minnesota (close to KDLH) for 20 years (see Fig. 6). We used a simple linear regression to test if persistence was a significant predictor of cluster size, measured as the peak number of birds per cluster per year. To account for years when the entire season was unavailable on the archive, we used the number of persistent years over the total number of fully sampled (more than 130 days) years as a predictor variable. The fitted model was $\log(y) = 1.00x + 3.33$, and it was statistically significant with p -value < 0.001 , and R -squared = 0.441 (see Fig. 7), indicating that areas that were consistently used throughout the years hosted larger roosts.

Discussion

Weather surveillance radars have proven to be an invaluable source of data about bird migration ecology. Particularly in the case of some species of the *Hirundinidae* family, due to their specific postbreeding roost formation behavior and timing, weather radars offer a rare opportunity to monitor and inventory their populations with macroscale coverage (Kelly & Horton, 2016). The main challenge currently faced in the effort to reach

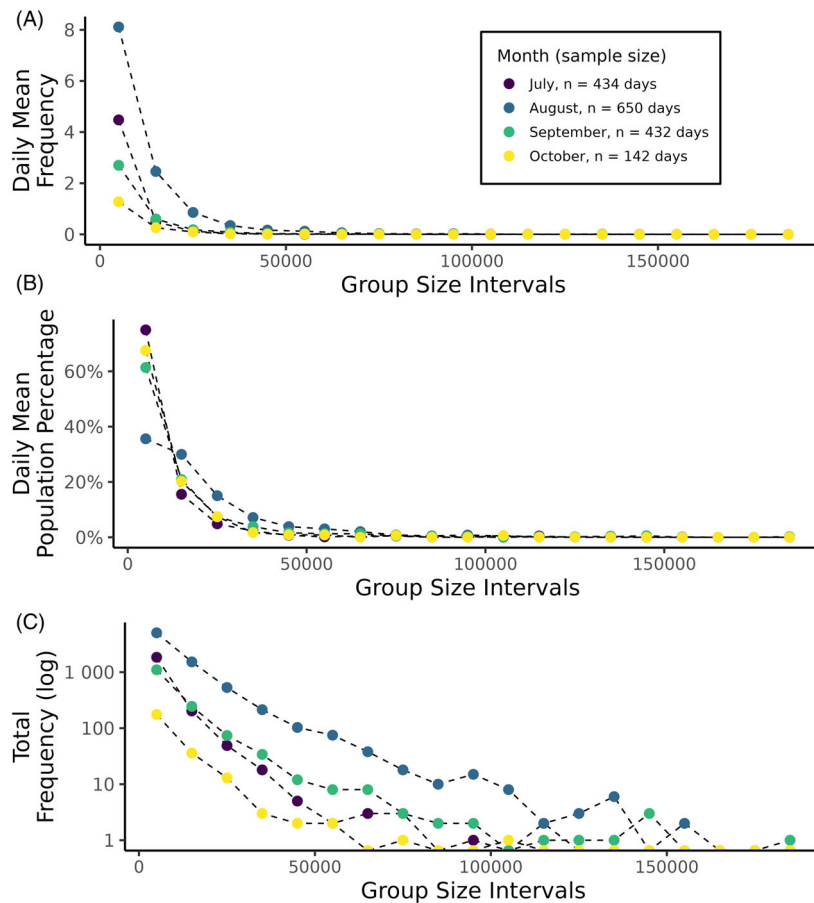


Figure 5. (A) Daily mean frequency of each roost size bin of 10 000 individuals, colored by month; (B) daily average percentage of the population found in each group size bin of 10 000 individuals, colored by month; (C) total frequency of each roost size, plotted on a logarithmic scale, in all station-year pairs with more than 130 sampled days.

such broad spatial and temporal scales is the detection of roosts, which is further complicated by the diverse sources of biological and nonbiological scattering that are typically captured by the radars. By integrating machine learning and human supervision, we have addressed both challenges, extending the temporal range from previous studies like Bridge et al. (2015) and Kelly et al. (2012), but more importantly, extracting critical indices of the magnitude of each communal roost throughout the years.

An important caveat of working with weather radar data is their lack of ability to sample lower heights above ground level, especially at greater distances to the radar station. When we can assume that the observed biomass density does not change with radar range, the systematic underestimation of density at greater distances to the radar station can be corrected using vertical profiles of reflectivity (Buler & Diehl, 2009). In the case of communal roosts, however, we can not assume that biomass density is constant at varying ranges, and thus, techniques to adjust

density estimates in this scenario based on radar data alone have not yet been developed. We remark, however, that in this study, we found little correlation between distance to the radar station and roost size or number of birds (see Appendix S2, Figs. S5 and S6), indicating that this issue likely has not influenced our conclusions.

We decided to use Purple Martins as our benchmark *Hirundinidae* because they are the largest species in this family, and thus produce conservative estimates of the number of birds per roost. Conservative abundance estimates can often be enough to diagnose population declines, monitor year-to-year abundance variation and inform conservation decision-making (e.g., see Hunter et al., 2010), yet they limit our ability to understand the full complexity of the *Hirundinidae* communal roosting behavior. Further studies pairing radar roost detections with on-the-ground species identification would allow us to characterize species-specific radar signatures in terms of their dispersal time, flight speed and polarimetric measurements (Stepanian et al., 2016). Moreover, verified

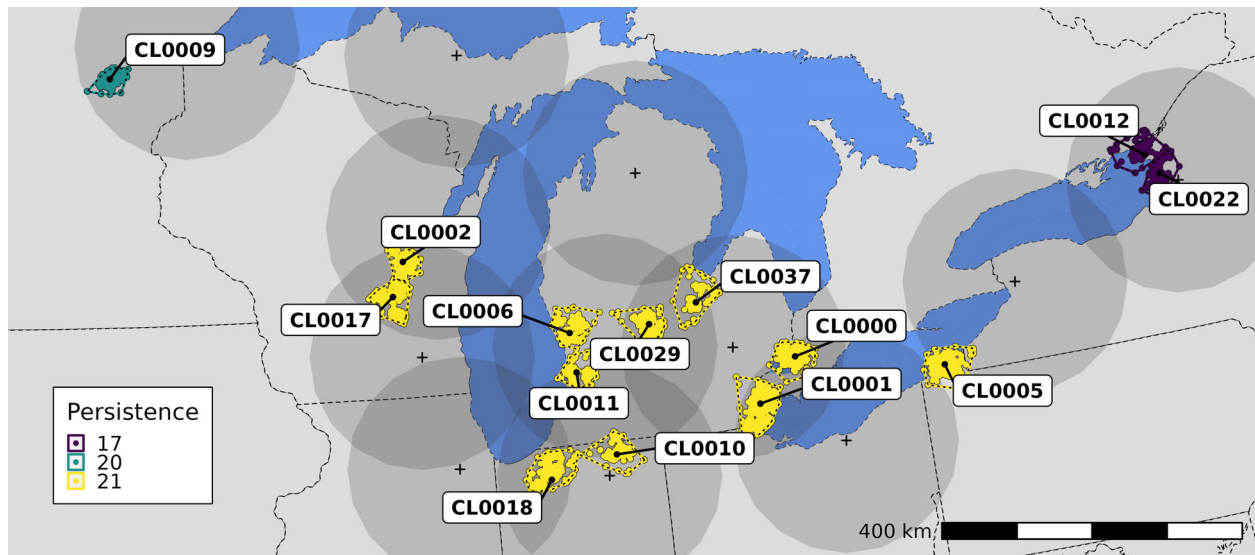


Figure 6. Throughout the Great Lakes region, the 17 clusters represented in this figure were detected on 100% of the years when radar stations near them were collecting data. The station with code KDLH, where CL0009 was detected, had inconsistent data in 2001. Data from KTYX, where clusters labeled CL0012 and CL0022 were located, were not made available from 2001 to 2004. Eleven roost clusters were detected every year from 2000 through 2020.

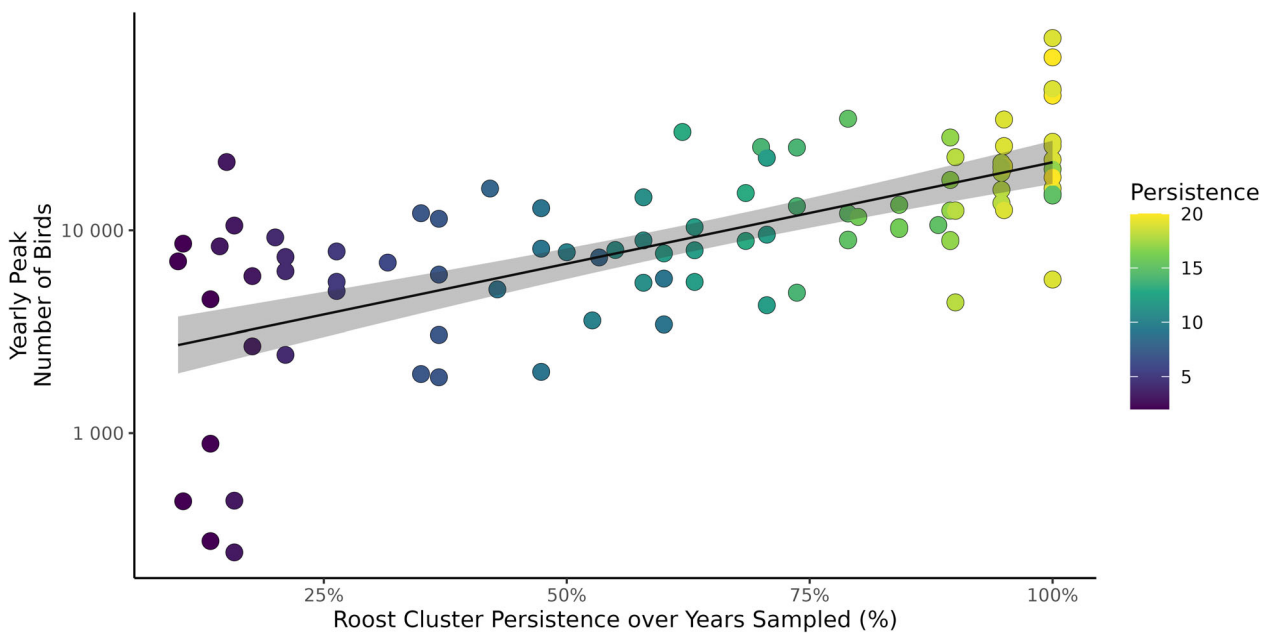


Figure 7. Relationship between yearly averaged peak number of birds and roost persistence corrected by the number of years each cluster was fully sampled. The fitted model was: $\log(y) = 1.00x + 3.33$, and it was statistically significant with p -value < 0.0001 , $F(1, 96) = 77.45$, and $R^2 = 0.4408$. The colors represent persistence of each cluster on the years when its nearest stations were fully sampled.

data could eventually be used as training data for a new model layer that assigns a species to each dispersal event.

Our results show 11 subregions (clusters) in the Great Lakes which were used year-after-year for the past 21 years, demonstrating how consistent the selection of

pre migratory roosting habitat is for these species (Allen & Nice, 1952; Bent, 1942; Bridge et al., 2015; Russell et al., 1998). This finding, together with the discovery that more persistent clusters were also larger in size, suggests that these regions might successfully support many birds,

and prompts further questions about the role swallows and martins have in those ecosystems. A study conducted by Powell et al. (1991) with much smaller avian aggregations demonstrated, for example, how seabirds in Florida Bay have an important role in translocating and concentrating nutrients in the regions close to their roost/colony sites, ultimately leading to changes in the adjacent seagrass communities.

We further exemplify the potential impact of a Purple Martin roost on the Great Lakes trophic network using the approach proposed by Kelly et al. (2013) to calculate the number of insects consumed on the day of peak activity in 2007 (August 17th). Assuming an adult Purple Martin needs 137 kJ per day to maintain regular metabolic functions and that each gram of an insect's dry mass provides 23 kJ, we can derive the total dry mass needed to fulfill the energetic requirements of a single adult. With this, we can employ the relationship between insect length and dry mass proposed by Sage (1982), and assume an average prey will have 20 mm, to determine how many insects would be needed to satisfy that daily dry mass requirement. In doing so, we estimate that on August 17, 2007, Purple Martins may have consumed more than 33 million insects. Examining summed daily martin activity, an average season may see as many as 338 million insects consumed throughout the Great Lakes region during the roosting period. While these simple calculations make many assumptions, they might serve as a basis for deeper investigations into the trophic role played by aerial insectivores across broad regions, and what the impacts of this guild's recent declines could be.

A secondary finding of our approach is related to how birds of the *Hirundinidae* family occupy the airspace. Our data suggests that, during their dawn departure, they might reach up to 4.5 km above ground level, which is higher than the maximum heights found by Helms et al. (2016) but similar to results from Dreelin et al. (2018) for swallows and martins during their breeding season. Combining altitude loggers and radar studies to explore demographic and seasonal species-specific patterns of flight height might provide valuable insight into how these aerial insectivores use different levels of the aerosphere during their annual cycle. Additionally, because the machine learning pipeline detects and tracks each roost's dispersal, it could be possible to rebuild a three-dimensional model of the departure that improves our understanding of how the birds coordinate their collective take off, and might eventually be used to correct the range bias mentioned above.

Empirical studies of animal aggregations in fish, ungulates and house sparrows *Passer domesticus* have found group size distributions of all these species to be typically heavy-tailed, with smaller groups being significantly more

common than larger ones (e.g., see Bonabeau et al., 1999; Gerard et al., 2002; Griesser et al., 2011). Theoretical studies have demonstrated that these distributions typically arise from the combined effects of animal movement and group splitting and amalgamation (also called fission and fusion) processes (Niwa, 2004; Okubo, 1986). Though these models accurately describe animals in movement, their translation to the communal roosting scenario is not straightforward, since roosts are not a direct result of movement dynamics, but rather of central place foragers daily gathering in a central place from which they depart the next day to forage (Lalla, 2022). Nonetheless, we have shown that the same pattern of group size distribution occurs in our study system. This result prompts further investigation into models that can describe the mechanisms responsible for generating the specific dynamic of roost formation behavior.

As expected, we found that the daily distribution of roost sizes changes throughout the season, yet we have, for the first time, quantified the progression of these within-season changes. Further studies of this pattern relating it to environmental variables and investigating its change throughout the years may reveal important factors influencing roost formation behavior and driving the roost size distribution. Since most of the *Hirundinidae* population in the Great Lakes seem to gather in groups with smaller sizes, it would be interesting to investigate why larger roosts are more persistent throughout the years, and what is the relationship between group size, roost persistence and fitness.

Acknowledgments

We thank Larissa Bailey, Matthew Ross and David Koons for their comments on the first version of this paper. This work was supported by the following National Science Foundation awards: Colorado State University (#2017554), University of Massachusetts-Amherst (#2017756), Oklahoma University: (#2017582).

Conflict of Interest

The authors report no conflict of interest.

Author Contributions

Maria Carolina T. D. Belotti: Conceptualization, Investigation, Methodology, Software, Data Curation, Writing Yuting Deng: Methodology, Investigation Wenlong Zhao: Methodology, Software, Investigation, Data Curation Victoria F. Simons: Investigation Zezhou Cheng: Methodology, Software, Gustavo Perez: Conceptualization, Software Elske Tielens: Conceptualization, Supervision Subhransu

Maji: Conceptualization, Supervision Daniel Sheldon: Conceptualization, Supervision Jeffrey F. Kelly: Conceptualization, Supervision Kyle G. Horton: Conceptualization, Methodology, Supervision, Validation, Writing.

References

- Allen, R.W. & Nice, M.M. (1952) A study of the breeding biology of the purple Martin (*Progne subis*). *American Midland Naturalist*, **47**(3), 606. <https://doi.org/10.2307/2422034>
- Bent, A.C. (1942) *Life histories of the north American flycatchers, larks, swallows and their allies*. Bulletin of the United States National Museum. <https://doi.org/10.5479/si.03629236.179.i>
- Bonabeau, E., Dagorn, D. & Fréon, P. (1999) Scaling in animal group-size distributions. *Proceedings of the National Academy of Sciences*, **96**(8), 4472–4477. <https://doi.org/10.1073/pnas.96.8.4472>
- Bridge, E.S., Pletschet, S.M., Fagin, T., Chilson, P.B., Horton, K.G., Broadfoot, K.R. et al. (2015) Persistence and habitat associations of purple Martin roosts quantified via weather surveillance radar. *Landscape Ecology*, **31**(1), 43–53. <https://doi.org/10.1007/s10980-015-0279-0>
- Buler, J.J. & Diehl, R.H. (2009) Quantifying bird density during migratory stopover using weather surveillance radar. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(8), 2741–2751. <https://doi.org/10.1109/TGRS.2009.2014463>
- Burney, C.W. (2002) *A study of swallow roosts found in the eastern United States*. Ithaca, NY: Cornell University.
- Cheng, Z., Gabriel, S., Bhambhani, P., Sheldon, D., Maji, S., Laughlin, A. et al. (2020) Detecting and tracking communal bird roosts in weather radar data. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**(1), 378–385. <https://doi.org/10.1609/aaai.v34i01.5373>
- Chilson, P.B., Frick, W.F., Stepanian, P.M., Shipley, J.R., Kunz, T.H. & Kelly, J.F. (2012) Estimating animal densities in the atmosphere using weather radar: to Z or not to Z? *Ecosphere*, **3**(8), 72. <https://doi.org/10.1890/ES12-00027.1>
- Cox, A.R., Robertson, R.J., Lendvai, Á.Z., Everitt, K. & Bonier, F. (2019) Rainy springs linked to poor nestling growth in a declining avian aerial insectivore (*Tachycineta bicolor*). *Proceedings of the Royal Society B: Biological Sciences*, **286** (1898), 20190018. <https://doi.org/10.1098/rspb.2019.0018>
- Cresswell, W. (2014) Migratory connectivity of Palaearctic–African migratory birds and their responses to environmental change: the serial residency hypothesis. *Ibis*, **156**(3), 493–510. <https://doi.org/10.1111/ibi.12168>
- Dixon, M. & Lee, W.C. (2021) CfRadial Data File Format. EOL, NCAR. <https://github.com/NCAR/CfRadial/blob/master/docs/CfRadialDoc.v1.5.20211201.pdf>
- Dokter, A.M., Farnsworth, A., Fink, D., Ruiz-Gutierrez, V., Hochachka, W.M., La Sorte, F.A. et al. (2018) Seasonal abundance and survival of North America's migratory avifauna determined by weather radar. *Nature Ecology & Evolution*, **2**(10), 1603–1609. <https://doi.org/10.1038/s41559-018-0666-4>
- Dokter, A.M., Liechti, F., Stark, H., Delobbe, L., Tabary, P. & Holleman, I. (2010) Bird migration flight altitudes studied by a network of operational weather radars. *Journal of the Royal Society Interface*, **8**(54), 30–43. <https://doi.org/10.1098/rsif.2010.0116>
- Dreelin, R.A., Shipley, J.R. & Winkler, D.W. (2018) Flight behavior of individual aerial insectivores revealed by novel altitudinal Dataloggers. *Frontiers in Ecology and Evolution*, **6**. <https://doi.org/10.3389/fevo.2018.00182>
- Dunning, J. (2008) *CRC handbook of avian body masses*, Vol. **384**, 2nd edition. <https://doi.org/10.1201/9781420064452>
- Erwin, M. (1982) Observer variability in estimating numbers. *Journal of Field Ornithology*, **53**(2), 159–167.
- Frederick, P.C., Hylton, B., Heath, J.A. & Ruane, M. (2003) Accuracy and variation in estimates of large numbers of birds by individual observers using an aerial survey simulator. *Journal of Field Ornithology*, **74**(3), 281–287.
- Gerard, J.F., Bideau, E., Maublanc, M.L., Loisel, P. & Marchal, C. (2002) Herd size in large herbivores: encoded in the individual or emergent? *Biological Bulletin*, **202**(3), 275–282. <https://doi.org/10.2307/1543479>
- Griesser, M., Ma, Q., Webber, S., Bowgen, K. & Sumpter, D.J.T. (2011) Understanding animal group size distributions. *PLoS One*, **6**(8), e23438. <https://doi.org/10.1371/journal.pone.0023438>
- Hallmann, C.A., Foppen, R.P.B., Turnhout, C.A.M.v., Kroon, H.d. & Jongejans, E. (2014) Declines in insectivorous birds are associated with high neonicotinoid concentrations. *Nature*, **511**(7509), 341–343. <https://doi.org/10.1038/nature13531>
- Helms, J.A., Godfrey, A.P., Ames, T. & Bridge, E.S. (2016) Predator foraging altitudes reveal the structure of aerial insect communities. *Scientific Reports*, **6**(1). <https://doi.org/10.1038/srep28670>
- Hoen, B., Diffendorfer, J.E., Rand, J., Kramer, L.A., Garrity, C.P., Roper, A.D. et al. (2018) United States wind turbine database. *U.S. Geological Survey*. <https://doi.org/10.5066/F7TX3DNO>
- Holmes, R.T. (2007) Understanding population change in migratory songbirds: long-term and experimental studies of Neotropical migrants in breeding and wintering areas. *Ibis*, **149**(s2), 2–13. <https://doi.org/10.1111/j.1474-919X.2007.00685.x>
- Horton, K.G., Doren, B.M.V., Sorte, F.A.L., Cohen, E.B., Clipp, H.L., Buler, J.J. et al. (2019) Holding steady: little change in intensity or timing of bird migration over the Gulf of Mexico. *Global Change Biology*, **25**(3), 1106–1118. <https://doi.org/10.1111/gcb.14540>
- Hunter, C.M., Caswell, H., Runge, M.C., Regehr, E.V., Amstrup, S.C. & Stirling, I. (2010) Climate change threatens polar bear

- populations: a stochastic demographic analysis. *Ecology*, **91** (10), 2883–2897. <https://doi.org/10.1890/09-1641.1>
- Imlay, T.L., Hobson, K.A., Roberto-Charron, A. & Leonard, M.L. (2018) Wintering areas, migratory connectivity and habitat fidelity of three declining Nearctic-Neotropical migrant swallows. *Animal Migration*, **5**(1), 1–16. <https://doi.org/10.1515/ami-2018-0001>
- iNaturalist. (2022) Available from <https://www.inaturalist.org> [Accessed 13th April, 2022].
- Kardynal, K.J., Jardine, T.D., Génier, C.S.V., Bumelis, K.H., Mitchell, G.W., Evans, M. et al. (2020) Mercury exposure to swallows breeding in Canada inferred from feathers grown on breeding and non-breeding grounds. *Ecotoxicology*, **29**(7), 876–891. <https://doi.org/10.1007/s10646-020-02249-6>
- Kelly, J.F., Bridge, E.S., Frick, W.F. & Chilson, P.B. (2013) Ecological energetics of an abundant aerial insectivore, the purple Martin. *PLoS One*, **8**(9), e76616. <https://doi.org/10.1371/journal.pone.0076616>
- Kelly, J.F. & Horton, K.G. (2016) Toward a predictive macrosystems framework for migration ecology. *Global Ecology and Biogeography*, **25**(10), 1159–1165. <https://doi.org/10.1111/geb.12473>
- Kelly, J.F., Shipley, J.R., Chilson, P.B., Howard, K.W., Frick, W.F. & Kunz, T.H. (2012) Quantifying animal phenology in the atmosphere at a continental scale using NEXRAD weather radars. *Ecosphere*, **3**(2), art16. <https://doi.org/10.1890/es11-00257.1>
- Kingma, D.P. & Ba, J. (2014) Adam: a method for stochastic optimization. ArXiv Preprint ArXiv:1412.6980.
- Lalla, K. (2022) *Purple martin (Progne subis) movement ecology during three stages of the annual cycle* [Master's Thesis]. Montreal, QC, Canada: McGill University.
- Laughlin, A.J., Sheldon, D.R., Winkler, D.W. & Taylor, C.M. (2016) Quantifying non-breeding season occupancy patterns and the timing and drivers of autumn migration for a migratory songbird using Doppler radar. *Ecography*, **39**(10), 1017–1024. <https://doi.org/10.1111/ecog.01988>
- Marra, P.P., Cohen, E.B., Loss, S.R., Rutter, J.E. & Tonra, C.M. (2015) A call for full annual cycle research in animal ecology. *Biology Letters*, **11**(8), 20150552. <https://doi.org/10.1098/rsbl.2015.0552>
- Marra, P.P. & Holmes, R.T. (2001) Consequences of dominance-mediated habitat segregation in American redstarts during the nonbreeding season. *The Auk*, **118**(1), 92–104. <https://doi.org/10.1093/auk/118.1.92>
- Niwa, H.S. (2004) Space-irrelevant scaling law for fish school sizes. *Journal of Theoretical Biology*, **228**(3), 347–357. <https://doi.org/10.1016/j.jtbi.2004.01.011>
- Okubo, A. (1986) Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in Biophysics*, **22**, 1–94. [https://doi.org/10.1016/0065-227X\(86\)90003-1](https://doi.org/10.1016/0065-227X(86)90003-1)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011) Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, **12**, 2825–2830.
- Powell, G.V.N., Fourqurean, J.W., Kenworthy, W.J. & Zieman, J.C. (1991) Bird colonies cause seagrass enrichment in a subtropical estuary: observational and experimental evidence. *Estuarine, Coastal and Shelf Science*, **32**(6), 567–579. [https://doi.org/10.1016/0272-7714\(91\)90075-M](https://doi.org/10.1016/0272-7714(91)90075-M)
- Ren, S. (2008) Finding people in archive films through tracking. *IEEE Conference on Computer Vision and Pattern Recognition*, **2008**, 1–8. <https://doi.org/10.1109/CVPR.2008.4587533>
- Ren, S., He, K., Girshick, R. & Sun, J. (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. & Garnett, R. (Eds.) *Advances in neural information processing systems*, Vol. **28**. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf>
- Rosenberg, K.V., Dokter, A.M., Blancher, P.J., Sauer, J.R., Smith, A.C., Smith, P.A. et al. (2019) Decline of the north American avifauna. *Science*, **366**(6461), 120–124. <https://doi.org/10.1126/science.aaw1313>
- Russell, K.R., Mizrahi, D.S. & Gauthreaux, S.A. (1998) Large-scale mapping of purple Martin pre-migratory roosts using WSR-88D weather surveillance. *Journal of Field Ornithology*, **69**(2), 316–325.
- Sage, R.D. (1982) Wet and dry-weight estimates of insects and spiders based on length. *The American Midland Naturalist*, **108**(2), 407–411. <https://doi.org/10.2307/2425505>
- Sauer, J.R., Niven, D.K., Pardieck, K.L., Ziolkowski, D.J., Jr. & Link, W.A. (2017) Expanding the north American breeding bird survey analysis to include additional species and regions. *Journal of Fish and Wildlife Management*, **8**(1), 154–172. <https://doi.org/10.3996/102015-JFWM-109>
- Sherry, T.W. & Holmes, R.T. (1996) Winter habitat quality, population limitation, and conservation of Neotropical-Nearctic migrant birds. *Ecology*, **77**(1), 36–48.
- Smith, A.C., Hudson, M.A.R., Downes, C.M. & Francis, C.M. (2015) Change points in the population trends of aerial-insectivorous birds in North America: synchronized in time across species and regions. *PLoS One*, **10**(7), e0130768. <https://doi.org/10.1371/journal.pone.0130768>
- Spiller, K.J. & Dettmers, R. (2019) Evidence for multiple drivers of aerial insectivore declines in North America. *The Condor*, **121**(2), duz010. <https://doi.org/10.1093/condor/duz010>
- Stepanian, P.M., Horton, K.G., Melnikov, V.M., Zrnić, D.S. & Gauthreaux, S.A. (2016) Dual-polarization radar products for biological applications. *Ecosphere*, **7**(11). <https://doi.org/10.1002/ecs2.1539>
- wSullivan, 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. *Biological Conservation*, **142** (10), 2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>

- Supp, S.R., Bohrer, G., Fieberg, J. & La Sorte, F.A. (2021) Estimating the movements of terrestrial animal populations using broad-scale occurrence data. *Movement Ecology*, **9**(1), 60. <https://doi.org/10.1186/s40462-021-00294-2>
- Torres, S.M. (2007) Initial implementation of super-resolution data on the NEXRAD network. American Meteorological Society Annual Meeting, San Antonio, Texas. https://ams.confex.com/ams/87ANNUAL/techprogram/paper_116240.htm.
- Warkentin, I.G. & Hernández, D. (1996) The conservation implications of site fidelity: a case study involving Nearctic-Neotropical migrant songbirds wintering in a Costa Rican mangrove. *Biological Conservation*, **77**(2–3), 143–150. [https://doi.org/10.1016/0006-3207\(95\)00146-8](https://doi.org/10.1016/0006-3207(95)00146-8)
- Webster, M.S., Marra, P.P., Haig, S.M., Bensch, S. & Holmes, R.T. (2002) Links between worlds: unraveling migratory connectivity. *Trends in Ecology & Evolution*, **17**(2), 76–83. [https://doi.org/10.1016/S0169-5347\(01\)02380-1](https://doi.org/10.1016/S0169-5347(01)02380-1)
- Wikelski, M., Kays, R.W., Kasdin, N.J., Thorup, K., Smith, J.A. & Swenson, G.W., Jr. (2007) Going wild: what a global small-animal tracking system could do for experimental biologists. *Journal of Experimental Biology*, **210**(2), 181–186. <https://doi.org/10.1242/jeb.02629>
- Wikiaves. (2022). Available from: <http://www.wikiaves.com.br> [Accessed 13th April 2022].
- Wu, Y., Kirillov, A., Massa, F., Lo, W.Y. & Girshick, R. (2019) Detectron2. <https://github.com/facebookresearch/detectron2>.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Detailed flowcharts and descriptions of the data screening protocol.

Appendix S2. Ancillary data analyses justifying the grouping of overlapping tracks and describing the clustering performance. Daily data completeness plots. Analyses of possible detection biases.