ARTICLE

Macrosystems Ecology

Clustering community science data to infer songbird migratory connectivity in the Western Hemisphere

Jaimie G. Vincent[1] | **Richard Schuster**[1,2,3] | **Scott Wilson**[1,4] | **Daniel Fink**[5] | **Joseph R. Bennett**[1]

[1]Department of Biology, Carleton University, Ottawa, Ontario, Canada

[2]Ecosystem Science and Management Program, University of Northern British Columbia, Prince George, British Columbia, Canada

[3]The Nature Conservancy of Canada, Vancouver, BC, Canada

[4]Wildlife Research Division, Pacific Wildlife Research Centre, Environment and Climate Change Canada, Delta, British Columbia, Canada

[5]Cornell Lab of Ornithology, Ithaca, New York, USA

Correspondence
Jaimie G. Vincent
Email: jaimie.vincent@carleton.ca

Abstract

Migratory connectivity describes the spatial linkage among migrating individuals through time. Accounting for it is necessary for full annual cycle conservation planning, to avoid uneven protection leading to overall population declines. However, conventional methods used to study migratory connectivity usually demand substantial fiscal and human resources. We present a methodology that infers patterns of migratory connectivity for songbirds using relative abundance models created from eBird, a global community science program. We compare our inferences with previously described patterns of migratory connectivity for two species assumed to exhibit broadscale parallel migration strategies: wood thrush (*Hylocichla mustelina*) and Wilson's warbler (*Cardellina pusilla*). Initial findings suggest that our method has the potential to be a rapid and inexpensive way to infer broad patterns of connectivity for species that do not engage in leap-frog migration nor deviate much from parallel migration. Our flexible framework can be used to guide sampling designs for studies of migratory connectivity and to generate hypotheses for species in need of urgent conservation planning for which migratory connectivity has not yet been established.

KEYWORDS
Bayesian analysis, birds, community science, connectivity, conservation, migration

INTRODUCTION

Conservation plans must consider the full annual cycle to adequately conserve migratory birds (Marra et al., 2015; Runge et al., 2014; Webster et al., 2002). However, creating such plans comes with many challenges, not the least of which is the fact that migrants often traverse vast distances annually and spend the majority of the year in regions where monitoring has been historically sparse (Runge et al., 2014). Knowledge gaps in time and space that ensue throughout migration and the stationary nonbreeding period limit our ability to understand the full annual cycle ecology of migrants and hinder conservation planning.

Migratory connectivity represents an important foundation for research and conservation that address limiting factors across time and space (Rushing et al. 2016) because it describes the spatial linkage between individuals through time (Marra et al., 2019). For this reason, in cases where there is weak connectivity—defined as situations in which individuals scatter from one period to another (Cohen et al., 2018;

Webster et al., 2002), conservation action in a given area is predicted to have a range-wide effect on the entire population. In contrast, when individuals remain in geographic proximity and, thus, display strong connectivity, conservation action can be directed to specific areas used by a subpopulation of conservation concern. Additionally, migratory connectivity can allow for more strategic conservation actions that minimize cost and economic activities, such as dynamic habitat protection (Reynolds et al., 2017). Conservation plans that ignore migratory connectivity can lead to uneven protection, which can result in regional population declines if conservation effort is not targeted to maintain key habitats throughout the annual cycle (Martin et al., 2007).

Identifying migratory connectivity requires tracking representative samples of individuals from different populations of a species throughout the annual cycle. In the last few decades, increasingly precise methodologies and technologies have been developed for tracking migration (Faaborg et al., 2010; McKinnon & Love, 2018). Each tracking technology has its own benefits and limitations. Recaptures of banded individuals are generally too low to infer migratory connectivity for songbirds (Hobson, 2003; Plissner & Haig, 2011), and light-level geolocators, which also require the recapture of an individual to extract migration data, can also have low return rates for certain species and cannot provide information when mortality occurs during migration (McKinnon & Love, 2018). Further, satellite GPS tags can be expensive and too heavy for smaller songbirds (McKinnon & Love, 2018). Although light-level geolocators and satellite GPS tags provide detailed migration information per individual, extensive fieldwork is required to elucidate a migration network (Knight et al., 2018). On the contrary, inferring migratory connectivity using intrinsic markers such as genetic markers and stable isotope ratios does not require recapture of individuals and can readily incorporate historical samples from museum specimens. However, they can be limited by coarser spatial resolution (Hobson, 2011). Another recently developed tool is the Motus Wildlife Tracking System, which is a network of automated radio-telemetry arrays that can track the movements of any small flying animal over vast spatial scales; as more Motus stations are erected, this technology will be increasingly useful to pinpoint connections between populations (Taylor et al., 2017). Thus, there is not a single approach to spatially delineate migrating songbird populations in migratory connectivity studies, and boundaries among populations are not always identified based on movement data. For example, boundaries can be established with political borders or recognized conservation regions such as Bird Conservation Regions (Kramer et al., 2018), or by intuitively dividing regions onto a map. Delineations

between migrating populations have also been created by clustering demographic data (Rushing et al., 2016) and genetic data (Ruegg et al., 2014).

In this paper, we develop an approach to infer songbird migratory connectivity between the breeding and nonbreeding ranges using the spatially explicit estimates of species' relative abundance based on observations from the eBird community science program (Fink et al., 2018, 2020). Our goal is to present a methodology that capitalizes on readily available data products and showcase a model-based approach that should not replace (nor supersede) other methods of assessing migratory connectivity, but rather provide researchers with additional analytical options to explore migratory connectivity that is not dependent on individual tracking data. Our objective is to apply a partition-based clustering method that mirrors the methodology developed by Rushing et al. (2016) that delineates natural populations based on demographic data. In doing so, our approach provides a method for delineating migratory regions in both the breeding and nonbreeding ranges based on the Euclidean distance between simulated individuals. Next, we aim to infer migratory connectivity using Bayes' rule, incorporating the total relative abundance of a region as a prior and assuming parallel migration to calculate the likelihood of individuals migrating to a given breeding region. Parallel migration refers to a migratory system where individuals that breed in the western part of a species' range will also overwinter in the western part of the nonbreeding range and, similarly, where individuals that breed in the eastern part of the breeding range will overwinter in the eastern part of the nonbreeding range (Alerstam & Hedenström, 1998). Parallel migration is commonly observed in songbirds where migratory connectivity has been studied, and there is strong evidence in the literature that supports the persistence of broad East–West divides for many species throughout the annual cycle (e.g., Kelly & Hutto, 2005; Norris et al., 2006; Stanley et al., 2015). To evaluate the performance of our methods, we apply our methodology to two species for which there has been extensive migratory connectivity research: wood thrush (*Hylocichla mustelina*) (Stanley et al., 2015) and Wilson's warbler (*Cardellina pusilla*) (Ruegg et al., 2014). This novel, low-cost approach to inferring migratory connectivity with community science data and common migration patterns could be applied to understudied species as an initial approach that can inform further exploration (e.g., to species for which migratory connectivity estimates are required before undertaking field studies) and to refine migratory connectivity estimates for species for which data are already available, thereby helping managers to plan for full annual cycle conservation.

METHODS

Study species

The wood thrush is a forest songbird that breeds in the eastern United States and southeastern Canada and winters in Central America. The species is listed as near-threatened on the International Union for the Conservation of Nature (IUCN) Red List and threatened under Canada's *Species at Risk Act*. The strength of migratory connectivity during wood thrush migration is uncertain, although there is some evidence that spatial segregation diminishes en route (Cohen et al., 2018). Nevertheless, light-level geolocator tracking evidence suggests strong connectivity and parallel migration between the breeding and nonbreeding periods (Stanley et al., 2015). More specifically, breeders in the American Northeast are strongly associated with the eastern nonbreeding range (i.e., eastern Honduras to Costa Rica); breeders in the American Midwest are connected with the central and eastern nonbreeding ranges (i.e., from Guatemala to Costa Rica); breeders in the western and southern portions of the American breeding range are linked to the eastern and central nonbreeding ranges (i.e., Mexico to western Honduras); and breeders from the western breeding range disperse throughout the nonbreeding range, although individuals recaptured in the western nonbreeding range were exclusively from the western breeding range (Stanley et al., 2015).

Wilson's warbler is a shrub songbird that breeds mostly in northern forests of Canada and the northwestern United States and winters in Central America and along the Gulf of Mexico. A broad East–West divide in the breeding range is well documented for this species (Clegg et al., 2003; Irwin et al., 2011; Kelly et al., 2002), with further genetic differentiation within the western group more recently recognized (Ruegg et al., 2014). Continent-wide, there is evidence of moderate parallel migration, where eastern breeders from Quebec and the Maritimes migrate to the east of Veracruz, Mexico, and south to Panama, and individuals from the northwestern group tend to spread throughout the nonbreeding range (Irwin et al., 2011; Ruegg et al., 2014). Isotopic and genetic analyses have shown that individuals from the western group demonstrate leapfrog and parallel migration (Clegg et al., 2003; Kelly et al., 2002; Ruegg et al., 2014).

To judge the plausibility of our models, we compare our migratory connectivity inferences with migratory networks constructed with multiple intrinsic markers and light-level geolocator tracks for wood thrush (Stanley et al., 2015) and migratory networks built via genetic markers for Wilson's warbler (Clegg et al., 2003; Irwin et al., 2011; Kelly et al., 2002; Ruegg et al., 2014) in the "Discussion."

eBird relative abundance estimates

eBird is a global community science (aka citizen science) project where members of the public can submit checklists containing information on bird counts and survey effort for any region (Sullivan et al., 2009). We used the relative abundance estimates (Fink et al., 2018) from the eBird Status and Trends project to describe the relative abundance of wood thrush and Wilson's warbler during the 2016 breeding and 2016/2017 nonbreeding seasons. The relative abundance estimates from the eBird Status and Trends project have been successfully used for other broadscale studies that rely on full annual cycle connectivity information (Johnston et al., 2020; Schuster et al., 2019). These estimates are based on the Adaptive Spatio-Temporal Exploratory Models (AdaSTEM) (Fink et al., 2014; Fink et al., 2020) and data from the eBird community science program (Sullivan et al., 2014). Adaptive Spatio-Temporal Exploratory Model controls for variation in detectability associated with search effort by standardizing the relative abundance estimates as the expected number of individuals of a species an observer is likely to encounter between 7:00 AM and 8:00 AM while traveling 1 km at a pixel resolution of 8 km^2 for every week of the year.

To infer migratory connectivity between the breeding and nonbreeding seasons, we selected relative abundance for a single week to represent a "snapshot" of each static season. We selected the week of 4 July for the breeding season and the week of 18 January for the nonbreeding season for both species. For the purposes of this study, the breeding period is inclusive of breeding adults, adults with failed nesting attempts who have moved away from their original nesting territory, and older hatch-year birds, and the nonbreeding season does not include considerations for intra-seasonal movement. We included pixels located within the 95% home range estimate using a kernel utilization distribution function assuming a bivariate normal probability density function using the adehabitat R package (Calenge, 2006).

Delineating regions

Simulated counts

To simulate individual birds from relative abundance estimates, we transformed the modeled relative abundance values into pseudo-counts, hereafter referred to as "counts," by multiplying the relative abundance value of an 8.4-km^2 pixel by 10 and rounding to the nearest integer in order to capture relative abundance values <1. We considered these counts to be a proxy for the number of

individuals located within the pixel. To delineate the seasonal populations into regions based on proximity between individuals, we clustered simulated individual birds using the Clustering for Large Applications (CLARA) algorithm, a partition-based clustering method suitable for large datasets that separates objects into a user-defined number of clusters (Kaufman & Rousseeuw, 1990). Clustering for Large Applications therefore requires the distance between individual objects (Kaufman & Rousseeuw, 1990). In our case, an "object" refers to an individual simulated bird.

Determining the optimal number of clusters

We used two evaluative criteria to determine the optimal number of clusters (k): the average silhouette method and the gap statistic. The former measures an object's similarity (proximity) to other objects of its cluster compared with its similarity to data objects of other clusters (Kaufman & Rousseeuw, 1990). The latter compares the distribution of the data objects within a cluster to the expected distribution under a null reference set (Tibshirani et al., 2001).

We used the NbClust function from the NbClust R package to compute the average silhouette and gap statistic for solutions between two and eight clusters. We chose to test between two and eight clusters because of computing constraints and because we presumed that broadscale migratory connectivity inferences with fewer large regions would produce more conservative inferences than inferences created with a larger number of smaller regions.

We computed both evaluative criteria 100 times by randomly sampling 1000 counts each time. The mode of all the optimal k solutions from the 100 samples derived from the average silhouette index was compared with that of the Gap statistic. If there was disagreement between the two criteria, we considered the mode of the average silhouette index to be the optimal k (Long et al., 2010).

Clustering simulated counts

Clustering for Large Applications builds off the Partitioning Around Medoids (PAM) algorithm, which clusters objects in an iterative process that minimizes the dissimilarity (i.e., distance) of objects within clusters (Kaufman & Rousseeuw, 1990). Partitioning Around Medoids requires considerable computing power, which is why it is unsuitable for large datasets. Clustering for Large Applications randomly subsamples a large dataset multiple times and computes PAM on each sample. Clustering for Large Applications retains the set of clusters that minimizes the dissimilarity of all objects to their respective central object (Kaufman & Rousseeuw, 1990).

We applied the CLARA algorithm to the counts with the *cluster* R package (Maechler et al., 2018). Clustering for Large Applications requires the user to define the size of the samples (i.e., the number of individuals to input into the PAM algorithm) and the number of samples that it will draw from the dataset to find the optimal clustering solution (i.e., the number of times that it will compute the PAM algorithm). We defined the sample size to be 15% of the total number of counts for computational feasibility. We parameterized the algorithm to draw 100 samples to determine the optimal clustering solution. We used Euclidean distance to compute the distance between individuals. The resulting clusters delineated the regions for each seasonal range.

Migratory connectivity analysis using Bayes' rule

To infer connections between breeding and nonbreeding regions, we applied Bayes' rule to calculate the probabilities of connection for each simulated bird to every target region in the other stationary period. Considering that differences in total relative abundance between regions could influence the probability of belonging to one region (Gómez et al., 2019; Norris et al., 2006; Royle & Rubenstein, 2004; Wilgenburg & Hobson, 2011), we included a region's total relative abundance as a prior while applying Bayes' rule:

$$f(b|y) = \frac{f(y|b)f(b)}{f(y)} \quad (1)$$

The prior probability of connection with a location b is dependent on the total relative abundance in that region, noted as $f(b)$, calculated by summing of the counts of the region. We standardized the prior probabilities for each region to sum to 1.

The likelihood of an individual's assignment to a given target region b from a given pixel $y*$, denoted as $f(y*|b)$, is calculated with the underlying assumption that the study species demonstrates parallel migration. That is, individuals that breed in the most westerly area of the range will migrate to the most westerly area of the nonbreeding range, and birds that breed in the most easterly area of the breeding range will migrate to the most easterly area of the nonbreeding range. To represent this, we first standardized all of the individuals' longitudinal values across each range and, for ease of calculation, assumed that the distribution of standardized values within each cluster was normal (though in some cases, the distribution deviated from normal; see Appendix S1: Figures S3 and S6). We then calculated the likelihood of

all the simulated individuals to a target region from a given pixel, $f(y^*|b)$, with a normal density function:

$$f(y^*|b) = \frac{1}{\sqrt{2\pi}\sigma_b}\exp\left[-\frac{1}{2\sigma_b{}^2}(y^* - \mu_b)^2\right] \quad (2)$$

where y^* is an individual's standardized longitudinal value relative to its seasonal range, and μ_b and σ_b are the standardized longitudinal mean value and standard deviation of a target region, respectively. The marginal probability, noted as $f(y)$, is calculated using the following equation:

$$f(y) = \sum_{b_1}^{b_n} f(y|b)f(b) \quad (3)$$

For a given individual, the target region that resulted in the highest $f(b|y)$ value was considered the most likely assignment region. We calculated the $f(b|y)$ value per breeding region for all the individuals in the nonbreeding range and the $f(b|y)$ value per nonbreeding region for all the individuals in the breeding range.

RESULTS

Wood thrush

Based on the average silhouette index and the gap statistic, we determined the optimal number of clusters for wood thrush to be 3 in the breeding range and 2 in the nonbreeding range (Figure 1a, Appendix S2: Tables S1 and S2).

For wood thrush, when we connected breeding to nonbreeding regions (Figure 1b), all individuals from the western breeding cluster (Figure 1a, cluster 1) and the central breeding cluster (Figure 1a, cluster 2) were predicted to be connected to the western nonbreeding cluster. Most individuals (57%) from the eastern breeding cluster were assigned to the eastern nonbreeding cluster.

Almost all individuals located in the eastern nonbreeding cluster were predicted to be connected to the eastern breeding cluster (Figure 1c). Conversely, individuals located in the western nonbreeding cluster were estimated to be connected to all of the breeding clusters, with the majority being assigned to the central breeding cluster (Figure 1c).

Wilson's warbler

For Wilson's warbler, two clusters were optimal in the breeding range (Figure 2a). The nonbreeding range was divided into five clusters, although there was disagreement between the two evaluative criteria, where the average silhouette index suggested five clusters and the gap statistic suggested two clusters (Appendix S2: Table S4).

When assigning connections to simulated individuals in the breeding range (Figure 2b), most (75%) western breeders were predicted to migrate to central Mexico nonbreeding cluster (Figure 2, cluster 2), with 19% moving to northwestern Mexico (cluster 1), and only 6% to southern Mexico (cluster 3). Eastern breeders were most likely to migrate to the most eastern nonbreeding clusters 4 (36%) and 5 (58%), while 5% were predicted to migrate to southern Mexico (cluster 3).

When connecting simulated individuals in the nonbreeding range (Figure 2c), our results predict low mixing in central Mexico and in the southern United States along the Gulf of Mexico (cluster 2). More extensive mixing is predicted in southern Mexico and Guatemala (cluster 3).

DISCUSSION

Migratory connectivity is notoriously challenging to study for migratory birds, especially for songbirds because of their small size and low recapture rates (McKinnon & Love, 2018). Moreover, the high costs associated with conventional field-based or tracking methods tend not to resolve the issue of low sample sizes (although exceptions exist; see Fraser et al., 2012; Knight et al., 2018; Ruegg et al., 2014). Our research provides a method that can be widely adopted to infer migratory connectivity for species that are difficult to study with tracking technologies.

Our inferences are consistent with known migratory connectivity patterns for wood thrush, though less so for Wilson's warbler. Models suggested strong connectivity between southern/central breeding regions and western nonbreeding regions for wood thrush but only weak connectivity for northeastern breeders. For Wilson's warbler, our migratory connectivity model suggested that western and eastern breeders are most likely to mix during the nonbreeding period in central (cluster 2) and southern Mexico (cluster 3).

In accordance with published wood thrush migratory networks (Rushing et al., 2014; Stanley et al., 2015), our results suggest that broadscale connectivity exists for wood thrush along an East–West axis. Similar to the connectivity network described in Rushing et al. (2014), our models describe the high connectivity between the central and western breeding clusters with the western region of the nonbreeding range (Figures 1b). Likewise, we find that our inferences are also congruent with Stanley et al. (2015), where 102 wood thrush were tracked
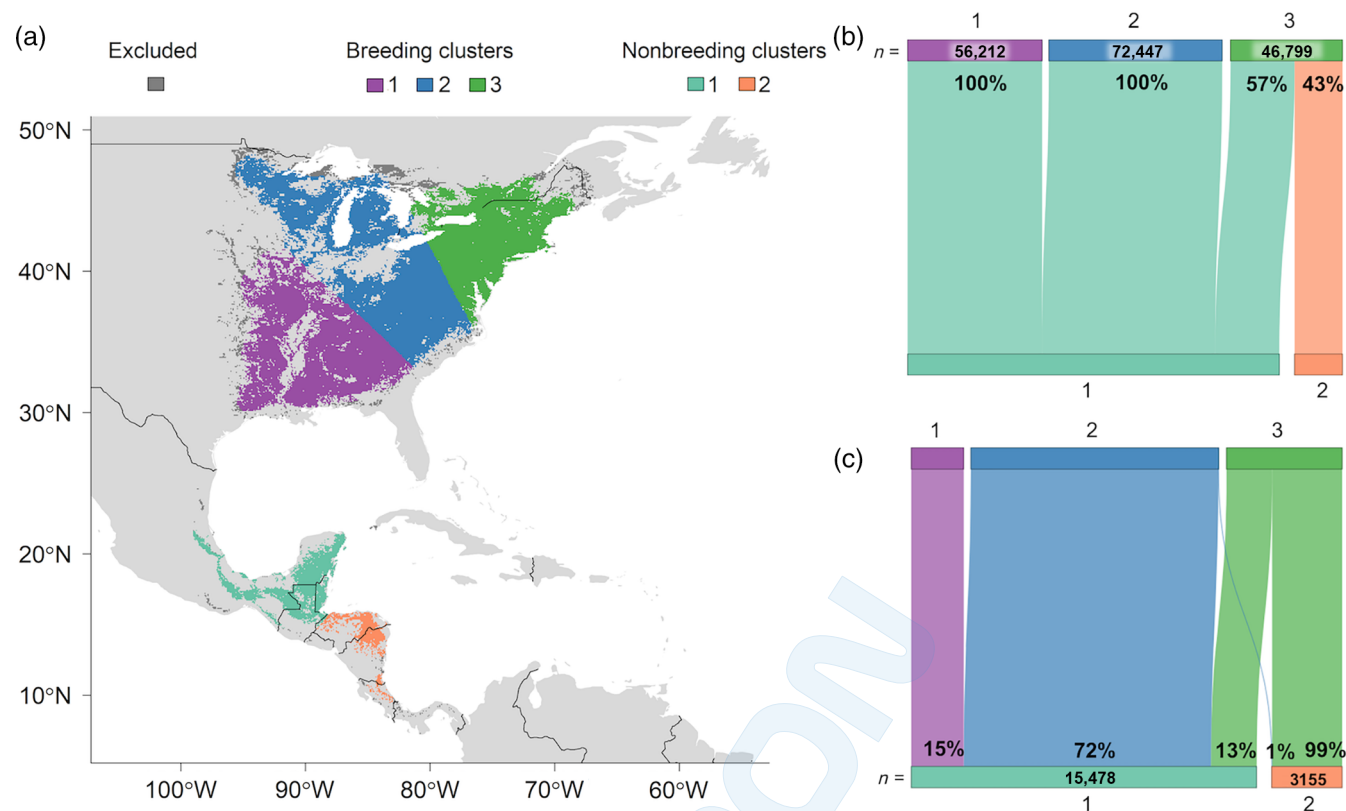
FIGURE 1 Migratory connectivity estimates for wood thrush. (a) Regions are defined by clustering eBird's wood thrush Adaptive Spatio-Temporal Exploratory Models (AdaSTEM) counts for the week of 4 July (breeding clusters) and 18 January (nonbreeding clusters). Dark gray pixels represent the excluded pixels from the home range (95%) analysis. The purple, blue, and green regions correspond to the breeding clusters 1, 2, and 3, respectively. The turquoise and orange regions correspond to the nonbreeding clusters 1 and 2, respectively. (b) The proportion of counts from the breeding range assigned to the nonbreeding clusters. (c) The proportion of counts from the nonbreeding range assigned to the breeding clusters

with light-level geolocators. However, we do note one dissimilarity; light-level geolocator tracking revealed higher connectivity between the central breeding cluster and the eastern nonbreeding cluster than our model suggested (Figure 1). This suggests that the parallel migration assumption can be insufficient to predict actual movement patterns by itself without field data to inform additional priors.

Although our analyses identified the broad East–West divide in the breeding range for Wilson's warbler (Clegg et al., 2003; Irwin et al., 2011; Kelly et al., 2002), they failed to reflect finer-scale genetic populations along the western coast of the United States (Ruegg et al., 2014). Wilson's warbler migratory connectivity estimates obtained from range-wide and high-resolution genetic markers (Ruegg et al., 2014) provide a rigorous backdrop against which our methodology's inferences can be compared and its strengths and limitations can be described.

There are two main areas of disagreement between our inferences and the genetic inferences for Wilson's warbler, which likely reflect the larger population size of the western than eastern breeders (see Appendix S1: Figure S4). First, while our estimates predict that few individuals from the central Mexico nonbreeding region (cluster 2) migrate to the eastern breeding cluster (Figure 2c), genetic markers did not predict a connection between eastern breeders and that region (Ruegg et al., 2014). For Wilson's warbler, eastern breeders are notoriously difficult to sample in the nonbreeding range even with an extensive sampling design (Irwin et al., 2011). Further, no genetic sampling occurred along the northwestern shore of the Gulf of Mexico (Ruegg et al., 2014), which is part of our central Mexico nonbreeding cluster. In this case, the differing migratory connectivity estimates show how our methodology can be helpful for filling in information gaps when it is not feasible to sample every part of a species' range.

Second, our estimates did not predict that western breeders overwinter along the entire longitudinal gradient of the nonbreeding range, as was shown via genetic analyses (Ruegg et al., 2014). This difference highlights a limitation of our methodology. Further work to improve
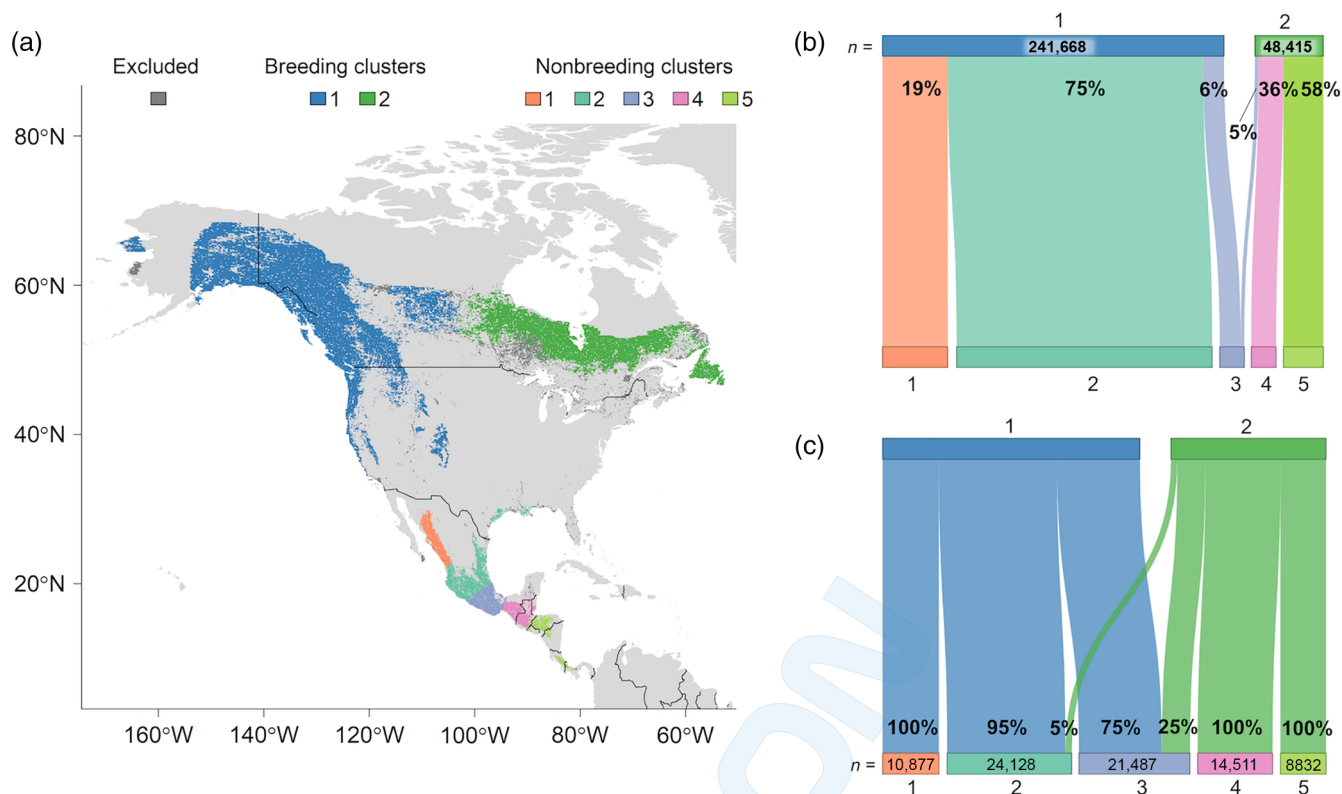
FIGURE 2 Migratory connectivity estimates for Wilson's warbler. (a) Regions are defined by clustering eBird's Wilson's warbler Adaptive Spatio-Temporal Exploratory Models (AdaSTEM) counts for the week of 4 July (breeding clusters) and 18 January (nonbreeding clusters). Dark gray pixels represent the excluded pixels from the home range (95%) analysis. The blue and green regions correspond to the breeding clusters 1 and 2, respectively. The orange, turquoise, lilac, pink, and light green regions correspond to the nonbreeding clusters 1, 2, 3, 4, and 5, respectively. (b) The proportion of counts from the breeding range assigned to the nonbreeding clusters. (c) The proportion of counts from the nonbreeding range assigned to the breeding clusters

our methodology could incorporate dispersal probabilities relative to the proportion of individuals from each breeding range and the total area occupied during the nonbreeding season.

Possible applications

Our methods provide a novel and feasible way to estimate plausible migratory connectivity patterns for species that are in need of conservation planning and have limited data to identify connectivity. Additionally, this model-based approach is not limited by survivorship bias, although it is not the only way to circumvent the issue (see Rushing et al., 2021). In some circumstances, our methods may not reflect actual migratory connectivity, which is why the results of connectivity networks constructed using this methodology can be thought of as a starting point; other tracking technologies could then be incorporated to provide a more integrated assessment of connectivity. We caution that connectivity inferences should be taken as coarse estimates and that the methods

presented in this study might not be appropriate for all songbird species. In any case, conservation plans should be robust enough to account for different connectivity scenarios in case migratory connectivity is not confirmed, poorly understood, or subject to change under different environmental conditions (Runge et al., 2014).

In cases where other data, such as band records, migratory tracks, stable isotopes, are available for a given species, the Bayesian framework can be used to refine migratory connectivity estimates according to new sources of information when parallel migration can be reasonably assumed. Indeed, methodologies that incorporate genetic markers and stable isotopes ratios into a Bayesian framework are already available in the literature (see, for e.g., Chabot et al., 2018; Rundel et al., 2013), and eBird data specifically have already been used to constrain migratory estimates derived from stable isotopes (Fournier et al., 2017). If the assumption of broadscale parallel migration can be made, our methodology can be easily used alongside other data types without adding significant costs.

For species that are subject to field studies, our methods can also be used to generate hypotheses of

coarse connectivity to improve the sampling design for future fieldwork in both stationary periods of the annual cycle that would then refine connectivity estimates. Migratory connectivity field studies should indeed aim to sample individuals across the entire extent of their range with an equal effort at each sampling site (Cohen et al., 2018; Knight et al., 2018). A sound sampling strategy would be to ensure sampling locations in each hypothesized cluster, with equal sampling effort among clusters in each period.

Caveats

We note several important caveats. First, our exploratory methodology infers migratory connectivity between the breeding and the nonbreeding ranges; it does not attempt to describe connectivity during migration. This is because parallel migration is not always maintained during migration, even if the final destinations of individuals often follow parallel patterns (e.g., Delmore et al., 2012). However, because eBird relative abundance models are available on a weekly basis, important common stopover areas could be identified in future studies if assumptions can be made about migratory routes. For example, geographically separate migration corridors such as waterfowl migration flyways could provide baseline assumptions for additional migratory connectivity analysis for certain species. Further, our methodology does not include assumptions of latitudinal migration strategies and eBird does not currently track individual movement; therefore, we do not take into account leapfrog or chain migration possibilities.

Second, we assumed that individuals' longitudinal values were normally distributed in each region to simplify calculations. An improved methodology would consider the unique distribution of longitudes of individuals in each region and apply different probability density functions accordingly. This would, however, still assume that "perfect" parallel migration occurs, which of course may never be the case. Relaxations of this assumption could perhaps be explored in the future, via additional simulations.

Third, the shape of the clustered regions is limited to the nature of the CLARA algorithm, which seeks to minimize the objective function based on the Euclidean distance of each cluster object to the centroid of each cluster. Therefore, CLARA clusters tend to be spherical, and oblong clusters are not recognized (Kaufman & Rousseeuw, 1990). Ultimately, when applying our method to a species in real-world conservation scenarios, the resulting clusters should be tempered with expert opinion and, if available, integrated with other data relevant to population delineation such as genetic markers, stable isotope ratios, or band returns. Some density-based clustering techniques such as DBSCAN can better recognize oblong clusters, but require careful consideration of user-defined parameters (Zerhari et al., 2015). To implement CLARA, the user must only define the number of clusters and sample sizes. For large datasets (such as ours), sample size is limited by computational constraints.

Finally, community science data in general have been critiqued for being biased toward areas with higher human activity (Chandler et al., 2017; Lloyd et al., 2020; Theobald et al., 2015). Nevertheless, when analyses account for species and spatial bias, community science is a valuable tool for conservation (McKinley et al., 2017). Indeed, models of distribution and relative abundance with eBird data account for variation in data density at regional–seasonal scales and for varying survey effort within region–seasons (Fink et al., 2010, 2013, 2014, 2020). Continent-wide weekly models such as the eBird relative abundance estimates present new opportunities for full annual cycle research (Schuster et al., 2019).

CONCLUSION

Conventional methods for tracking migratory connectivity can be challenging and expensive. To our knowledge, this is the first time that community science has been used as the only data source to explore migratory connectivity using a parallel migration assumption. Our work provides a low-cost opportunity to enhance our understanding of migration and can be applied to understudied species in need of conservation action. Our method also provides a flexible framework for producing hypotheses relevant to field studies. To adapt this methodology on a species-by-species basis, future work could incorporate all other available data for the species of interest (such as genetic markers and stable isotopes) to refine the clustering process and the inferred migratory connectivity and could explore stopover connectivity if appropriate individual movement data are available.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Jaimie G. Vincent https://orcid.org/0000-0002-1246-0587
Daniel Fink https://orcid.org/0000-0002-8368-1248

REFERENCES

Alerstam, T., and A. Hedenström. 1998. "The Development of Bird Migration Theory." *Journal of Avian Biology* 29(4): 343–69. https://doi.org/10.2307/3677155

Calenge, C. 2006. "The Package 'Adehabitat' for the R Software: A Tool for the Analysis of Space and Habitat Use by Animals." *Ecological Modelling* 197(3): 516–9. https://doi.org/10.1016/j.ecolmodel.2006.03.017

Chabot, A. A., K. A. Hobson, S. L. Van Wilgenburg, G. E. Pérez, and S. C. Lougheed. 2018. "Migratory Connectivity in the Loggerhead Shrike (*Lanius ludovicianus*)." *Ecology and Evolution* 8(22): 10662–72. https://doi.org/10.1002/ece3.4415

Chandler, M., L. See, K. Copas, A. M. Z. Bonde, B. C. López, F. Danielsen, J. K. Legind, et al. 2017. "Contribution of Citizen Science towards International Biodiversity Monitoring." *Biological Conservation* 213: 280–94. https://doi.org/10.1016/j.biocon.2016.09.004

Clegg, S. M., J. F. Kelly, M. Kimura, and T. B. Smith. 2003. "Combining Genetic Markers and Stable Isotopes to Reveal Population Connectivity and Migration Patterns in a Neotropical Migrant, Wilson's Warbler (*Wilsonia pusilla*)." *Molecular Ecology* 12(4): 819–30. https://doi.org/10.1046/j.1365-294X.2003.01757.x

Cohen, E. B., J. A. Hostetler, M. T. Hallworth, C. S. Rushing, T. Scott Sillett, and P. P. Marra. 2018. "Quantifying the Strength of Migratory Connectivity." *Methods in Ecology and Evolution* 9(3): 513–24. https://doi.org/10.1111/2041-210X.12916

Delmore, K. E., J. W. Fox, and D. E. Irwin. 2012. "Dramatic Intraspecific Differences in Migratory Routes, Stopover Sites and Wintering Areas, Revealed Using Light-Level Geolocators." *Proceedings of the Royal Society B: Biological Sciences* 279(1747): 4582–9. https://doi.org/10.1098/rspb.2012.1229

Faaborg, J., R. T. Holmes, A. D. Anders, K. L. Bildstein, K. M. Dugger, S. A. Gauthreaux, P. Heglund, et al. 2010. "Recent Advances in Understanding Migration Systems of New World Land Birds." *Ecological Monographs* 80(1): 3–48. https://doi.org/10.1890/09-0395.1

Fink, D., T. Damoulas, and J. Dave. 2013. "Adaptive Spatio-Temporal Exploratory Models: Hemisphere-Wide Species Distributions from Massively Crowdsourced eBird Data." In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Fink, D., T. Auer, A. Johnston, V. Ruiz-Gutierrez, W. M. Hochachka, and S. Kelling. 2020. "Modeling Avian Full Annual Cycle Distribution and Population Trends with Citizen Science Data." *Ecological Applications* 30(3): e02056. https://doi.org/10.1002/eap.2056

Fink, D., T. Auer, A. Johnston, M. Strimas-Mackey, M. J. Iliff, and S. T. Kelling. 2018. *eBird Status and Trends. Data Version: 2016 Released 2018*. Ithaca, NY: Cornell Lab of Ornithology. https://ebird.org/science/status-and-trends

Fink, D., T. Damoulas, N. E. Bruns, F. A. La Sorte, W. M. Hochachka, C. P. Gomes, and S. Kelling. 2014. "Crowdsourcing Meets Ecology: Hemisphere-Wide Spatiotemporal Species Distribution Models." *AI Magazine* 35(2): 19–30. https://doi.org/10.1609/aimag.v35i2.2533

Fink, D., W. M. Hochachka, B. Zuckerberg, D. W. Winkler, M. Ben Shaby, A. Munson, G. Hooker, M. Riedewald, D. Sheldon, and S. Kelling. 2010. "Spatiotemporal Exploratory Models for Broad-Scale Survey Data." *Ecological Applications* 20(8): 2131–47.

Fournier, A. M. V., K. L. Drake, and D. C. Tozer. 2017. "Using Citizen Science Monitoring Data in Species Distribution Models to Inform Isotopic Assignment of Migratory Connectivity in Wetland Birds." *Journal of Avian Biology* 48(12): 1556–62. https://doi.org/10.1111/jav.01273

Fraser, K. C., B. J. M. Stutchbury, P. M. Cassandra Silverio, J. B. Kramer, D. Newstead, N. Mickle, B. F. Cousens, et al. 2012. "Continent-Wide Tracking to Determine Migratory Connectivity and Tropical Habitat Associations of a Declining Aerial Insectivore." *Proceedings of the Royal Society B: Biological Sciences* 279(1749): 4901–6. https://doi.org/10.1098/rspb.2012.2207

Gómez, C., S. L. Guerrero, A. M. FitzGerald, N. J. Bayly, K. A. Hobson, and C. D. Cadena. 2019. "Range-Wide Populations of a Long-Distance Migratory Songbird Converge during Stopover in the Tropics." *Ecological Monographs* 89(2): e01349. https://doi.org/10.1002/ecm.1349

Hobson, K. A. 2003. "Making Migratory Connections with Stable Isotopes." In *Avian Migration*, edited by P. Berthold, E. Gwinner, and E. Sonnenschein, 379–91. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-05957-9_27

Hobson, K. A. 2011. "Isotopic Ornithology: A Perspective." *Journal of Ornithology* 152(S1): 49–66. https://doi.org/10.1007/s10336-011-0653-x

Irwin, D. E., J. H. Irwin, and T. B. Smith. 2011. "Genetic Variation and Seasonal Migratory Connectivity in Wilson's Warblers (Wilsonia Pusilla): Species-Level Differences in Nuclear DNA between Western and Eastern Populations." *Molecular Ecology* 20(15): 3102–15. https://doi.org/10.1111/j.1365-294X.2011.05159.x

Johnston, A., T. Auer, D. Fink, M. Strimas-Mackey, M. Iliff, K. V. Rosenberg, S. Brown, R. Lanctot, A. D. Rodewald, and S. Kelling. 2020. "Comparing Abundance Distributions and Range Maps in Spatial Conservation Planning for Migratory Species." *Ecological Applications* 30(3): e02058. https://doi.org/10.1002/eap.2058

Kaufman, L., and P. J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley. https://doi.org/10.1002/9780470316801

Kelly, J. F., V. Atudorei, Z. D. Sharp, and D. M. Finch. 2002. "Insights into Wilson's Warbler Migration from Analyses of Hydrogen Stable-Isotope Ratios." *Oecologia* 130(2): 216–21. https://doi.org/10.1007/s004420100789

Kelly, J. F., and R. L. Hutto. 2005. "An East–West Comparison of Migration in North American Wood Warblers." *The Condor* 107(2): 197–211. https://doi.org/10.1093/condor/107.2.197

Knight, S. M., D. W. Bradley, R. G. Clark, E. A. Gow, M. Bélisle, L. L. Berzins, T. Blake, et al. 2018. "Constructing and Evaluating a Continent-Wide Migratory Songbird Network across the Annual Cycle." *Ecological Monographs* 88(3): 445–60. https://doi.org/10.1002/ecm.1298

Kramer, G. R., D. E. Andersen, D. A. Buehler, P. B. Wood, S. M. Peterson, J. A. Lehman, K. R. Aldinger, et al. 2018. "Population Trends in Vermivora Warblers Are Linked to Strong Migratory Connectivity." *Proceedings of the National Academy of Sciences* 115(14): E3192–200. https://doi.org/10.1073/PNAS.1718985115

Lloyd, T. J., R. A. Fuller, J. L. Oliver, A. I. Tulloch, M. Barnes, and R. Steven. 2020. "Estimating the Spatial Coverage of Citizen Science for Monitoring Threatened Species." *Global Ecology and Conservation* 23: e01048. https://doi.org/10.1016/j.gecco.2020.e01048

Long, J., T. Nelson, and M. Wulder. 2010. "Regionalization of Landscape Pattern Indices Using Multivariate Cluster Analysis." *Environmental Management* 46(1): 134–42. https://doi.org/10.1007/s00267-010-9510-6

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2018. Cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0.

Marra, P. P., C. E. Studds, and M. Webster. 2010. "Migratory Connectivity." In *Encyclopedia of Animal Behavior*, edited by M. D. Breed and J. Moore, 455–61, Amsterdam: Elsevier.

Marra, P., E. B. Cohen, S. R. Loss, J. E. Rutter, and C. M. Tonra. 2015. "A Call for Full Annual Cycle Research in Animal Ecology." *Biology Letters* 11: 20150552. https://doi.org/10.1098/rsbl.2015.0552

Marra, P. P., E. Cohen, A.-L. Harrison, C. E. Studds, and M. Webster. 2019. "Migratory Connectivity." In *Encyclopedia of Animal Behavior*, 2nd ed., edited by J. C. Choe, 643–54. Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-809633-8.20761-9

Martin, T. G., I. Chadès, P. Arcese, P. P. Marra, H. P. Possingham, and D. Ryan Norris. 2007. "Optimal Conservation of Migratory Species." *PLoS One* 2: e751. https://doi.org/10.1371/journal.pone.0000751

McKinley, D. C., A. J. Miller-Rushing, H. L. Ballard, R. Bonney, H. Brown, S. C. Cook-Patton, D. M. Evans, et al. 2017. "Citizen Science Can Improve Conservation Science, Natural Resource Management, and Environmental Protection." *The Role of Citizen Science in Biological Conservation* 208: 15–28. https://doi.org/10.1016/j.biocon.2016.05.015

McKinnon, E. A., and O. P. Love. 2018. "Ten Years Tracking the Migrations of Small Landbirds: Lessons Learned in the Golden Age of Bio-Logging." *The Auk* 135(4): 834–56. https://doi.org/10.1642/AUK-17-202.1

Norris, D. R., P. P. Marra, G. J. Bowen, L. M. Ratcliffe, J. Andrew Royle, and T. Kurt Kyser. 2006. "Migratory Connectivity of a Widely Distributed Songbird, the American Redstart (*Setophaga ruticilla*)." *Ornithological Monographs* 61: 14–28. https://doi.org/10.2307/40166836

Plissner, J. H., and S. M. Haig. 2011. "Status of a Broadly Distributed Endangered Species: Results and Implications of the Second International Piping Plover Census." *Canadian Journal of Zoology* 78: 128. https://doi.org/10.1139/z99-179

Reynolds, M. D., B. L. Sullivan, E. Hallstein, S. Matsumoto, S. Kelling, M. Merrifield, D. Fink, et al. 2017. "Dynamic Conservation for Migratory Species." *Science Advances* 3(8): e1700707. https://doi.org/10.1126/sciadv.1700707

Royle, J. A., and D. R. Rubenstein. 2004. "The Role of Species Abundance in Determining Breeding Origins of Migratory Birds with Stable Isotopes." *Ecological Applications* 14(6): 1780–8. https://doi.org/10.1890/04-0175

Ruegg, K. C., E. C. Anderson, K. L. Paxton, V. Apkenas, S. Lao, R. B. Siegel, D. F. DeSante, F. Moore, and T. B. Smith. 2014. "Mapping Migration in a Songbird Using High-Resolution Genetic Markers." *Molecular Ecology* 23(23): 5726–39. https://doi.org/10.1111/mec.12977

Rundel, C. W., M. B. Wunder, A. H. Alvarado, K. C. Ruegg, R. Harrigan, A. Schuh, J. F. Kelly, et al. 2013. "Novel Statistical Methods for Integrating Genetic and Stable Isotope Data to Infer Individual-Level Migratory Connectivity." *Molecular Ecology* 22(16): 4163–76. https://doi.org/10.1111/mec.12393

Runge, C. A., T. G. Martin, H. P. Possingham, S. G. Willis, and R. A. Fuller. 2014. "Conserving Mobile Species." *Frontiers in Ecology and the Environment* 12: 395. https://doi.org/10.1890/130237

Rushing, R. S., T. B. Ryder, J. F. Saracco and P. P. Marra. 2014. "Assessing Migratory Connectivity for a Long-Distance Migratory Bird Using Multiple Intrinsic Markers." *Ecological Applications*, 24: 445–56. https://doi.org/10.1890/13-1091.1

Rushing, C. S., T. B. Ryder, and P. P. Marra. 2016. "Quantifying Drivers of Population Dynamics for a Migratory Bird throughout the Annual Cycle." *Proceedings of the Royal Society B: Biological Sciences* 283: 20152846. https://doi.org/10.1098/rspb.2015.2846

Rushing, C. S., T. B. Ryder, A. L. Scarpignato, J. F. Saracco, and P. P. Marra. 2016. "Using Demographic Attributes from Long-Term Monitoring Data to Delineate Natural Population Structure." *Journal of Applied Ecology* 53: 491. https://doi.org/10.1111/1365-2664.12579

Rushing, C. S., A. M. Van Tatenhove, A. Sharp, V. Ruiz-Gutierrez, M. C. Freeman, P. W. Sykes, Jr., A. M. Given, and T. S. Sillett. 2021. "Integrating Tracking and Resight Data Enables Unbiased Inferences about Migratory Connectivity and Winter Range Survival from Archival Tags." *Ornithological Applications* 123(2): duab010. https://doi.org/10.1093/ornithapp/duab010

Schuster, R., S. Wilson, A. D. Rodewald, P. Arcese, D. Fink, T. Auer, and J. R. Bennett. 2019. "Optimizing the Conservation of Migratory Species over Their Full Annual Cycle." *Nature Communications* 10(1): 1754. https://doi.org/10.1038/s41467-019-09723-8

Stanley, C. Q., E. A. McKinnon, K. C. Fraser, M. P. Macpherson, G. Casbourn, L. Friesen, P. P. Marra, et al. 2015. "Connectivity of Wood Thrush Breeding, Wintering, and Migration Sites Based on Range-Wide Tracking." *Conservation Biology* 29(1): 164–74. https://doi.org/10.1111/cobi.12352

Sullivan, B. L., J. L. Aycrigg, J. H. Barry, R. E. Bonney, N. Bruns, C. B. Cooper, T. Damoulas, et al. 2014. "The eBird Enterprise: An Integrated Approach to Development and Application of Citizen Science." *Biological Conservation* 169: 31–40. https://doi.org/10.1016/j.biocon.2013.11.003

Sullivan, B. L., C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling. 2009. "eBird: A Citizen-Based Bird Observation

Network in the Biological Sciences." *Biological Conservation* 142: 2282. https://doi.org/10.1016/j.biocon.2009.05.006

Taylor, P. D., T. L. Crewe, S. A. Mackenzie, D. Lepage, Y. Aubry, Z. Crysler, G. Finney, et al. 2017. "The Motus Wildlife Tracking System: A Collaborative Research Network to Enhance the Understanding of Wildlife Movement." *Avian Conservation and Ecology* 12(1): 8. https://doi.org/10.5751/ACE-00953-120108

Theobald, E. J., A. K. Ettinger, H. K. Burgess, L. B. DeBey, N. R. Schmidt, H. E. Froehlich, C. Wagner, et al. 2015. "Global Change and Local Solutions: Tapping the Unrealized Potential of Citizen Science for Biodiversity Research." *Biological Conservation* 181: 236–44. https://doi.org/10.1016/j.biocon.2014.10.021

Tibshirani, R., G. Walther, and T. Hastie. 2001. "Estimating the Number of Clusters in a Data Set via the Gap Statistic." *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 63: 411–23. https://doi.org/10.1016/j.scico.2012.08.004

Vincent, J. G. 2022. "Clustering Community Science Data to Infer Songbird Migratory Connectivity in the Western Hemisphere." OSF. https://doi.org/10.17605/OSF.IO/4WVTA.

Webster, M. S., P. P. Marra, S. M. Haig, S. Bensch, and R. T. Holmes. 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

Wilgenburg, S. L. V., and K. A. Hobson. 2011. "Combining Stable-Isotope (6D) and Band Recovery Data to Improve Probabilistic Assignment of Migratory Birds to Origin." *Ecological Applications* 21(4): 1340–51.

Zerhari, B., A. A. Lahcen, and S. Mouline. 2015. "Big Data Clustering: Algorithms and Challenges." In *International Conference on Big Data, Cloud and Applications*. Tetouan, Morocco.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Vincent, Jaimie G., Richard Schuster, Scott Wilson, Daniel Fink, and Joseph R. Bennett. 2022. "Clustering Community Science Data to Infer Songbird Migratory Connectivity in the Western Hemisphere." *Ecosphere* 13(4): e4011. https://doi.org/10.1002/ecs2.4011