Integrating data types to estimate spatial patterns of avian migration across the Western Hemisphere

Timothy D. Meehan¹, Sarah P. Saunders¹, William V. DeLuca¹, Nicole L. Michel¹, Joanna Grand¹, Jill L. Deppe², Miguel F. Jimenez¹, Erika J. Knight¹, Nathaniel E. Seavy¹, Melanie A. Smith¹, Lotem Taylor¹, Chad Witko¹, Michael E. Akresh³, David R. Barber¹, Erin M. Bayne⁵, James C. Beasley⁶⁻⁷, Jerrold L. Belant⁸, Richard O. Bierregaard⁹, Keith L. Bildstein⁴, Than J. Boves¹⁰, John N. Brzorad¹¹⁻¹², Steven P. Campbell¹³, Antonio Celis-Murillo¹⁴, Hilary A. Cooke¹⁵, Robert Domenech¹⁶, Laurie Goodrich⁴, Elizabeth A. Gow¹⁷⁻¹⁸, Aaron Haines¹⁹, Michael T. Hallworth²⁰⁻²¹, Jason M. Hill¹¹, Amanda E. Holland⁶⁻⁷, Scott Jennings²², Roland Kays²³⁻²⁴, D. Tommy King²⁵, Stuart A. Mackenzie¹⁷, Peter P. Marra²⁶, Rebecca A. McCabe⁴, Kent P. McFarland²¹, Michael J. McGrady²⁷, Ron Melcer Jr.²⁸⁻²⁹, D. Ryan Norris¹⁸, Russell E. Norvell³⁰, Olin E. Rhodes Jr.⁶, Christopher C. Rimmer²⁰, Amy L. Scarpignato³¹, Adam Shreading¹⁶, Jesse L. Watson⁸⁻³², and Chad B. Wilsey¹

¹National Audubon Society, 225 Varick Street, New York, NY 10014, USA.
²National Audubon Society, 1200 18th Street NW, Suite 500, Washington DC, 20036, USA.
³Department of Environmental Studies, Antioch University New England, Keene, NH 03431, USA.
⁴Acopian Center for Conservation Learning, Hawk Mountain Sanctuary Association, 410 Summer Valley Road, Orwigsburg, PA 17961, USA.
⁵Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada.
⁶Savannah River Ecology Laboratory, Aiken, SC 29808, USA.
⁷University of Georgia, Warnell College of Forestry & Natural Resources, Athens, GA 30602, USA.
⁸Global Wildlife Conservation Center, State University of New York College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, USA.
⁹Department of Ornithology, Academy of Natural Sciences of Drexel University, 1900 Benjamin Franklin Parkway, Philadelphia, PA 19103, USA.
¹⁰Department of Biological Sciences, Arkansas State University, Jonesboro, AR 72401, USA.
¹¹1000 Herons. 7727 Wingmont Dr. Charlotte, NC 28269, USA.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/eap.2679

This article is protected by copyright. All rights reserved.
12 Lenoir-Rhyne University, 625 7th Ave. NE, Hickory, NC 28601, USA.
13 Albany Pine Bush Preserve Commission, 195 New Karner Road, Albany, NY 12205, USA.
14 U.S. Geological Survey, Eastern Ecological Center, Patuxent Research Refuge, 12100 Beech Forest Rd., Laurel, MD 20708
16 Raptor View Research Institute, PO Box 4323, Missoula, MT 59806, USA.
17 Birds Canada, Port Rowan, ON, Canada.
18 Department of Integrative Biology, University of Guelph, Guelph, ON, Canada.
19 Millersville University, Biology Department, PO Box 1002, Millersville, PA 17551, USA.
20 Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA.
21 Vermont Center for Ecostudies, PO Box 420, Norwich, VT 05055, USA.
22 Cypress Grove Research Center, Audubon Canyon Ranch, P.O. Box 808, Marshall, CA 94940, USA.
23 North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA.
24 Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27601, USA.
25 U.S. Department of Agriculture, Wildlife Services, National Wildlife Research Center, Mississippi Field Station, 2745 West Line Road, Mississippi State University, MS 39762, USA.
26 Department of Biology and McCourt School of Public Policy, Georgetown University, 37th and O Streets, NW, Washington, DC 20057, USA.
27 International Avian Research, Am Rosenhügel 59, A-3500 Krems, Austria.
28 California State Parks, 1416 9th Street, Sacramento, CA 95814, USA.
29 Geography Graduate Group, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA.
30 Utah Division of Wildlife Resources, 1594 W. North Temple, Suite 2110, Salt Lake City, UT 84114, USA.
31 Migratory Bird Center, Smithsonian Conservation Biology Institute, National Zoological Park, 3001 Connecticut Ave., NW Washington, D.C. 20008, USA.
32 HawkWatch International, Salt Lake City, UT 84106, USA.

Corresponding author: Sarah P. Saunders. E-mail: sarah.saunders@audubon.org. ORCID ID: 0000-0002-2688-9528.
Open Research Statement
Tracking datasets for the 12 bird species analyzed here are available on Movebank, band re-encounter data are available through the USGS Bird Banding Laboratory, and Cornell Lab of Ornithology eBird Status and Trends products are available at https://ebird.org/science/status-and-trends. Please see Appendix S4 for data attributions and contact person(s) for each species. Code to run analyses and species-specific migratory connectivity results are provided in Audubon’s GitHub repository and permanently archived in Zenodo at https://doi.org/10.5281/zenodo.5996798 (Meehan & Saunders 2022). Species-specific raster layers of the integrated index shown in Appendix S3 are available in the Zenodo at https://doi.org/10.5281/zenodo.6325814 (Meehan et al. 2022).
Abstract

For many avian species, spatial migration patterns remain largely undescribed, especially across hemispheric extents. Recent advancements in tracking technologies and high-resolution species distribution models (i.e., eBird Status and Trends products) provide new insights into migratory bird movements and offer a promising opportunity for integrating independent data sources to describe avian migration. Here, we present a three-stage modeling framework for estimating spatial patterns of avian migration. First, we integrate tracking and band re-encounter data to quantify migratory connectivity, defined as the relative proportions of individuals migrating between breeding and nonbreeding regions. Next, we use estimated connectivity proportions along with eBird occurrence probabilities to produce probabilistic least-cost path (LCP) indices. In a final step, we use generalized additive mixed models (GAMMs) both to evaluate the ability of LCP indices to accurately predict (i.e., as a covariate) observed locations derived from tracking and band re-encounter datasets versus pseudo-absence locations during migratory periods, and to create a fully integrated (i.e., eBird occurrence, LCP, and tracking/band re-encounter data) spatial prediction index for mapping species-specific seasonal migrations. To illustrate this approach, we apply this framework to describe seasonal migrations of 12 bird species across the Western Hemisphere during pre- and post-breeding migratory periods (i.e., spring and fall, respectively). We found that including LCP indices with eBird occurrence in GAMMs generally improved the ability to accurately predict observed migratory locations, when compared to models with eBird occurrence alone. Using three performance metrics, the eBird + LCP model demonstrated equivalent or superior fit relative to the eBird-only model for 22 of 24 species-season GAMMs. In particular, the integrated index filled in spatial gaps for species with over-water movements and those that migrated over land where there were few eBird sightings,
and thus, low predictive ability of eBird occurrence probabilities (e.g., Amazonian rainforest in South America). This methodology of combining individual-based seasonal movement data with temporally dynamic species distribution models provides a comprehensive approach for integrating multiple data types to describe broad-scale spatial patterns of animal movement. Further development and customization of this approach will continue to advance knowledge about the full annual cycle and conservation of migratory birds.

*Keywords:* annual cycle, band re-encounter, data integration, eBird, least-cost paths, migratory connectivity, tracking
Introduction

Billions of birds migrate across the globe each year, and their seasonal movements span a broad range of spatial and temporal scales. Approximately 73% of bird species breeding in North America are migratory, 54% of which are considered Nearctic-Neotropical migrants (Albert et al. 2020). Evidence suggests that migration costs can have significant impacts on individual survival and, ultimately, the viability of a population (Newton 2006). While consideration of animal movements is often necessary when designing management efforts for conservation of migratory birds (McGowan et al. 2017, Robertson et al. 2018), their extensive ranges and highly variable occurrence at fine spatial scales present considerable challenges (Kays et al. 2015, Nandintsetseg et al. 2019).

A comprehensive understanding of migratory connectivity, or the extent to which populations are linked throughout the annual cycle (Marra & Studds 2010), requires quantification of spatial variation in the distribution and abundance of individuals across stationary and migratory stages (Webster et al. 2002, Knight et al. 2021). For many species, spatial patterns during migration remain largely undescribed, especially across hemispheric extents (Nilsson et al. 2019). This knowledge gap contributes to uncertainty about the ecological factors limiting populations (e.g., climate vs. land cover; Howard et al. 2020); the timing, location, and potential delayed impact of those limiting factors (e.g., carryover effects; Akresh et al. 2019); the spatial distribution of threats to habitats (Wilson et al. 2019); and the optimal allocation of conservation efforts that will match species’ full-annual-cycle requirements (Runge et al. 2016). Deeper insights into species’ spatial migration patterns will not only shed light on these uncertainties but will further advance our understanding of migratory connectivity.
Advancements in tracking technologies, such as GPS, light-level geolocators (LLG), and satellite telemetry (e.g., platform transmitting terminal [PTT]), have improved our understanding of spatial migration patterns by allowing for increasingly accurate estimates of location and quantification of error in animal movements (Rakhimberdiev et al. 2016, McKinnon & Love 2018). For example, tracking information has been used to reveal individuals’ migration routes (e.g., DeLuca et al. 2019), migratory bottlenecks (e.g., Buechley et al. 2018), and conservation priorities for migratory birds (e.g., Knight et al. 2021). Individually banded and re-encountered birds have also contributed invaluable information to our understanding of migratory patterns (e.g., Kreakie & Keitt 2012, Deppe et al. 2015). However, individual-based data types are often limited by restricted spatial coverage, as well as relatively small sample sizes and low re-encounter rates. Moreover, non-random deployment of tracking units and banding efforts could misrepresent species-level patterns of movement and habitat requirements (Li et al. 2017, Buechley et al. 2018).

Recently developed high-resolution species distribution models based on widespread, volunteer-collected eBird data have proved to be an important advancement for capturing landscape-scale, intra-annual distributional dynamics for many avian species (e.g., Schuster et al. 2019, Johnston et al. 2020). The eBird project consists of a global network of volunteers (albeit with the majority located in North America) who submit bird sightings, which are then verified by regional reviewers (Sullivan et al. 2014). eBird Status and Trends products (ebird.org/science/status-and-trends) use observations submitted since 2005 in machine learning models to characterize species abundance and distribution at a 2.96-km resolution for each week of the year while accounting for the local environment and multiple sources of sampling bias (see Appendix S1 and Fink et al. 2020a for more details), effectively capturing localized spatial
patterns across broad species’ ranges. Yet the predictive ability of these models is limited in
geographic areas where eBird sightings are absent due to a lack of observers or because cryptic
behaviors make direct observation of bird movements challenging during transitory migration
periods.

The development of statistically robust methods to integrate individual-based seasonal
movement information (i.e., tracking and band re-encounter data) with these temporally dynamic
species distribution models (i.e., eBird Status and Trends products) will enable researchers to
leverage the unique strengths of these independent sources and advance our understanding of
bird migration at relevant spatial and temporal extents for full-annual-cycle conservation
planning. Data integration, or the merging of multiple data types within a unified analytical
framework, provides a powerful method for sharing information among available data sources
while also minimizing their inherent biases (Miller et al. 2019, Zipkin et al. 2021). Here, we
present a novel, three-stage approach (Fig. 1) for integrating multiple, distinct data types to
describe spatial patterns of bird migration across the Western Hemisphere during species-
specific pre- and post-breeding migratory periods, as defined by eBird Status and Trends
products (hereafter eBird; see Appendix S1 for species’ migration timings). In the first two
stages, we quantified migratory connectivity and then used probabilistic least-cost paths (LCPs;
Adrianensen et al. 2003, Storfer et al. 2007) to combine connectivity estimates with weekly
estimates of relative occurrence probabilities derived from eBird (Fink et al. 2020b).

Traditionally, LCPs have been used to model gene flow across landscapes (Storfer et al. 2007),
with resistance surfaces or costs representing geographic or physiological barriers to animal
movement (e.g., topographic, elevational, thermal stress; Pullinger & Johnson 2010). More
recently, LCPs have been adapted for estimating species-specific migration routes (e.g., Nourani
et al. 2018), and they offer a promising – yet uninvestigated – avenue for broader migration applications, such as filling spatial gaps where volunteer-collected data are sparse or unavailable.

In a final stage, we used generalized additive mixed models (GAMMs) both to evaluate the ability of estimated LCPs to accurately predict (i.e., as a covariate) observed animal locations derived from tracking and band re-encounter datasets compared to pseudo-absence locations during migratory periods, and to create fully integrated (i.e., eBird occurrence, LCP, and tracking/band re-encounter data) prediction surfaces describing spatial patterns of seasonal bird migrations. We apply this modeling framework to a suite of 12 bird species with a variety of migratory strategies and geographic ranges. Accounting for spatiotemporal dynamics strongly influences the success of conservation planning for wide-ranging migratory species (Runge et al. 2016, Johnston et al. 2020), and this methodology of integrating disparate migration data types provides a comprehensive approach for estimating broad-scale spatial patterns of animal movement to advance full-annual-cycle research and conservation efficacy.

Materials and Methods

Overview of modeling framework

To identify and evaluate spatial migration patterns using multiple data types, we developed a novel three-stage modeling framework (Fig. 1), which we used to describe seasonal migrations of 12 bird species across the Western Hemisphere. In the first stage, we integrated tracking and band re-encounter datasets to estimate migratory connectivity, defined as the relative proportions of individuals migrating between stationary breeding and nonbreeding regions for each species (also referred to as transition probabilities; Marra et al. 2006). In the second stage, we used migratory connectivity proportions developed in the first stage along with eBird occurrence
probabilities (i.e., occurrence rate estimates from eBird Status and Trends products; Fink et al. 2020b) as the conductance surface to produce probabilistic least-cost path passage indices (hereafter LCP indices) during pre- and post-breeding migratory periods (i.e., spring and fall, respectively). In the final stage, we used GAMMs to evaluate the strength of associations between presence/pseudo-absence data (derived from tracking and band re-encounter datasets) and both LCP indices and eBird occurrence probabilities. We used the resulting GAMM prediction surfaces, which represent a statistical integration of eBird occurrence probabilities and LCP indices (i.e., as predictor variables), as well as tracking and band re-encounter data (i.e., as part of the response variable), to produce the final maps of integrated migration patterns for each species during each migratory period. Below, we briefly describe each step of the three-stage modeling framework (Fig. 1); see Appendix S1 for additional methodological details.

Focal species

We applied our modeling framework to band re-encounter and tracking data from 12 focal species to estimate migration patterns during pre- and post-breeding migratory periods (see Appendix S1 for species-specific migration timings): American White Pelican (Pelecanus erythrorhynchos), Blackpoll Warbler (Setophaga striata; DeLuca et al. 2015, 2019), Broad-winged Hawk (Buteo platypus; Haines et al. 2003, McCabe et al. 2020), Grasshopper Sparrow (Ammodramus savannarum; Hill and Renfrew 2019), Great Egret (Ardea alba), Osprey (Pandion haliaetus; Horton et al. 2014, Martell et al. 2014), Ovenbird (Seiurus aurocapilla; Hallworth and Marra 2015, Haché et. Al. 2017), Prairie Warbler (Setophaga discolor), Prothonotary Warbler (Protonotaria citrea; Tonra et al. 2019), Swainson’s Hawk (Buteo swainsoni; Fuller et al. 1998, Kochert et al. 2011), Tree Swallow (Tachycineta bicolor; Knight et
al. 2018), and Turkey Vulture (*Cathartes aura*; Dodge et al. 2014, Holland et al. 2017, 2019). We selected species to represent a diversity of migratory strategies (e.g., overwater vs. overland, short- vs. long-distance), geographic range boundaries, avian guilds, tracking technologies, and available data quantities (see *Estimating migratory connectivity*). For the latter, we ensured that all selected species had tracking locations for ≥ 10 individuals (the approximate number to minimize spatial biases; Knight et al. 2021, O’Toole et al. 2021) because our evaluation procedure required use of this data type as part of the response variable (in addition to band re-encounters; see *Integrating data sources and evaluating relative model performance*).

*Estimating migratory connectivity*

To develop a generalizable approach for describing migratory connectivity between the stationary breeding and nonbreeding seasons across species, we divided the Western Hemisphere into distinct spatial units, hereafter ‘migratory connectivity regions’ or MCRs (Fig. 2a) that largely corresponded with either pre-defined Bird Conservation Regions (North America; Birds Canada and North American Bird Conservation Initiative 2014) or Level II ecoregions (Central and South America; Griffith et al. 1998, see Appendix S1 for more details). To ensure that we did not estimate migratory connectivity between marginal breeding and nonbreeding populations of each species, we only considered MCRs where summed seasonal eBird relative abundance estimates within a given MCR were ≥ 1% of the summed total seasonal relative abundance across a species’ range (see Appendix S1 for details on calculating seasonal abundances from eBird Status and Trends products). We estimated migratory connectivity as the relative proportions of individuals migrating between these selected breeding and nonbreeding MCRs for each species (e.g., Fig. 2b). To do this, we statistically integrated (via joint likelihood) band re-
encounter and tracking data during the stationary breeding and nonbreeding periods following
the method developed by Korner-Nievergelt et al. (2017), which controls for spatial
heterogeneity in re-encounter probabilities of marked birds. To maximize sample sizes for
inference, we used band re-encounter data obtained from the U.S. Geological Survey (USGS)
Bird Banding Laboratory (BBL) since 1930 (e.g., Ryder et al. 2011, Macdonald et al. 2012) and
tracking data (LLG, GPS, PTT) obtained by Audubon’s Migratory Bird Initiative in partnership
with individual researchers (see Appendix S1 for details on data cleaning and filtering
procedures). We fit species-specific migratory connectivity models within a Bayesian framework
via JAGS (Plummer 2003) called from R (version 4.0.1; R Core Team 2019) using the ‘jagsUI’
package (Kellner 2016). See Appendix S1 for further details on model implementation and
Appendix S2 for details on species-specific migratory connectivity results.

**Producing least-cost paths (LCPs)**

Within selected MCRs for each species, we ranked each 2.96-km cell (i.e., native resolution of
eBird Status and Trends products) by seasonal relative abundance and selected as the breeding
and nonbreeding cores (i.e., high-abundance clusters) the minimum number of cells that
represented 30% of the total sum of cells (*sensu* Schuster et al. 2019, Lin et al. 2020) within each
MCR (Fig. 2a). Relative abundance models for the breeding and nonbreeding seasons for each of
the focal species were expert-reviewed and met performance standards of eBird Status and
Trends products (Fink et al. 2020a). We used these breeding and nonbreeding cores for initiation
of LCPs (i.e., source of origin points) for post-breeding and pre-breeding migration, respectively,
because we were primarily interested in identifying major migratory pathways among population
clusters. We designed this approach to incorporate two important attributes of identifying natural
We randomly selected 50 grid cells within each breeding MCR core to serve as post-breeding migration origin points for species-specific LCPs. Next, we randomly paired each origin point with a destination point in a randomly selected grid cell ($n = 50$) within nonbreeding MCR cores using the relative proportions estimated from the migratory connectivity analysis described above (Fig. 2b). We then computed a probabilistic (i.e., randomized) LCP (Adrianensen et al. 2003, Storfer et al. 2007, Wang et al. 2009) between each breeding-nonbreeding core pixel pair that minimized the total cumulative cost, where the cost of moving between paired pixels was determined by the intervening distance weighted by a conductance surface representing average (i.e., arithmetic mean across weeks; see Appendix S1: Table S2) post-breeding occurrence probabilities obtained from eBird (Fig. 2c). Thus, higher occurrence values during migration coincided with higher conductance. We chose to use occurrence probabilities as the conductance surface, rather than relative abundances, because abundance values for several focal species were highly right-skewed (e.g., high-abundance aggregations of migrating Tree Swallows), which resulted in LCPs directed towards regions of unusually high abundance, masking the known movements of smaller, regional populations.

To create biologically reasonable LCPs, we modified conductance surfaces by adding minimum conductance values where eBird occurrence probabilities were zero or had missing values. These conductance values varied across species to reflect relevant migratory behaviors (e.g., likelihood of long-distance overwater movements). See Appendix S1 for more details on assigning values to each of the 12 focal species and the potential for customization in future analyses. Partial randomization of the deterministic LCPs was incorporated via a constrained
random walk using the passage() function from the ‘gdistance’ package in R (Van Etten 2017). The passage function simulates movements from a starting location to an ending location, with movements between intervening cells governed by a random process superimposed on a cost surface. Random movements are derived from a probability distribution constrained by a parameter that controls the degree of randomization (Saerens et al. 2009). During model development, we evaluated a wide variety of values and algorithms for choosing the degree of randomization, and ultimately opted to use half the minimum conductance value, which resulted in individual probabilistic LCPs that had a longitudinal range of approximately 100 km (see Appendix S1 for more details on the partial randomization process).

Each set of paired grid cells yielded one raster layer (rescaled to a 26.6-km resolution due to computational demands; see Appendix S1) with values between zero and one that described the probability of passage (i.e., LCP index) through a given grid cell during post-breeding migration. We repeated this process for the 50 sets of paired grid cells for each breeding MCR, and then repeated the entire procedure in reverse (i.e., nonbreeding season origin points paired with breeding season destination points) using average pre-breeding eBird occurrence probabilities as the conductance surface to produce LCP indices during pre-breeding migration. Finally, we averaged (i) across the 50 probabilistic LCP surfaces for each season-specific MCR, and (ii) across all nonbreeding or breeding MCRs to illustrate probable pathways during pre- and post-breeding migrations, respectively, for each species (Fig. 2d). For example, for a species with 10 breeding MCRs, 10 × 50 = 500 total LCPs were averaged to produce a post-breeding migration surface.
**Integrating data sources and evaluating relative model performance**

To integrate additional data sources with LCPs, we first assembled available band re-encounter and tracking data during the pre- and post-breeding migratory periods (i.e., seasonal observations that were not used in previous stages) and followed similar cleaning procedures as those used in migratory connectivity analyses (see Appendix S1 for more details). We also filtered redundant tracking observations by randomly selecting one observation per individual bird per 26.6-km grid cell for each migratory season. Observations were filtered in this way to reduce spatial and temporal autocorrelation (sensu Northrup et al. 2013). For LLG data, we also removed locations over a 20-day period centered on each equinox. At each of these remaining ‘known presence’ locations (e.g., Fig. 3a) during each migratory season, we extracted two underlying raster values using geographic coordinates: (i) eBird occurrence probability (Fig. 2c), and (ii) LCP index (Fig. 2d). We then extracted occurrence probabilities and LCP indices at 10,000 background (pseudo-absence) points drawn from a convex hull across each species’ full-annual-cycle range (i.e., the spatial extent of possible migration locations), which we defined by overlaying eBird Status and Trends seasonal ranges with BirdLife International seasonal ranges to encompass peripheral, unmodeled areas (BirdLife International 2017).

For each species and migratory season, we built three GAMMs assuming a binomial distribution with a logit-link function to evaluate associations between presence/pseudo-absence data and both eBird occurrence probabilities and LCP indices. The full model had a global intercept, a random intercept that represented either an individual bird (tracking data) or all banded birds (band re-encounter data), and a two-dimensional smooth effect of eBird occurrence probabilities and LCP indices, as we expected possible interacting and asymptotic relationships between the predictors and the response variable. The random effect was specified to control for
potential correlations in tracking locations within individuals, as well as possible error correlations associated with different data types, while minimizing the number of random-effect levels because of computational challenges (e.g., all band re-encounters were assigned the same level). Random intercept levels for pseudo-absence locations were randomly generated in proportion to the presence data (see Appendix S1 for details). In an eBird-only model, we included both intercept terms as well as a one-dimensional smooth effect of eBird occurrence. In a third model, we included only the two intercepts and considered this the null model. Despite filtering tracking data and specifying random effects for individual birds, there was residual autocorrelation in some species-season models (results not shown). However, we did not expect autocorrelation of tracking data to impact our inferences regarding the contributions of LCP indices relative to eBird occurrence probabilities because (i) tracking data were included as the response variable in all models (i.e., there were no comparisons of models with and without tracking data); and (ii) spatial or temporal autocorrelation is primarily of concern when hypothesis testing (e.g., resource selection modeling; Gillies et al. 2006, Koper & Manseau 2012), which we did not do here.

We assessed the added contribution of LCP-derived indices to describing eBird-derived migration patterns by comparing model fits of the full model (i.e., eBird + LCP) vs. eBird-only model using three evaluation metrics: Akaike’s Information Criterion (AIC), proportion deviance explained, and Area Under the Receiver Operator Curve (AUC). The same response data (i.e., tracking and band re-encounters during migration) were used for model comparisons within each species-season combination; see Appendix S1 for more details on GAMM fitting and results. We did not evaluate an LCP-only model because we were primarily interested in assessing whether LCP indices provided additional information, beyond that of eBird occurrence probabilities, for
describing spatial patterns of avian migration. Although eBird occurrence informed LCP indices, which may have resulted in correlated predictors for some locations, we note that our objective of determining orthogonal-added predictive value (rather than hypothesis testing) makes multicollinearity concerns less important.

The resulting GAMM prediction surface (e.g., Fig. 3b) for each species in each migratory season represented a statistical integration of eBird occurrence probabilities and LCP indices (i.e., predictor variables in GAMMs), as well as available tracking and band re-encounter data (i.e., as part of the response variable in GAMMs). Below we summarize results across the 12 focal species and illustrate specific results for Swainson’s Hawk during post-breeding migration as a fully worked case study; see Appendix S3 for all species-specific results and Appendix S4 for data attributions for each of the 12 species analyzed here.

Sensitivity analyses

We conducted post-hoc Spearman’s correlations using model fit results and quantities and qualities of data types to evaluate whether our inferences on LCP contributions relative to eBird occurrences were related to data availability and/or technology type. To further understand the relative effects of data quantity and quality, we ran a multiple linear regression with deviance explained by the full model as a function of the following four predictors that were centered scaled prior to analysis: number of tracked individuals, number of geolocator-tagged or banded individuals, number of individuals tracked with PTT or GPS tags (i.e., high-quality tracking data), and spatial comprehensiveness of tracking data (i.e., proportion of species-specific breeding and wintering MCRs with $\geq 1\%$ of the total number of tracking observations). We
included the latter metric to determine whether there was a relationship between spatial coverage of tracking data and model fit (see Appendix S1 for more details on sensitivity analyses).

Results

Migratory connectivity

For several focal species (e.g., American White Pelican, Osprey, Ovenbird, Prairie Warbler), tracking and band re-encounter data suggested relatively high migratory connectivity proportions (i.e., strong connectivity). For example, more than half (0.59) of marked Ovenbirds breeding in the northeastern U.S. (MCR 14) overwintered in Florida (MCR 31) and the Greater Antilles (MCR 162), while marked Ovenbirds breeding in the central/western Canadian Boreal (MCR 6) were generally equally distributed across wintering MCRs in Mexico, Central America, and the Caribbean Islands. For other species, tracking and band re-encounter data supported equivalent (i.e., weak) migratory connectivity between all breeding and nonbreeding MCRs. Given that the migratory connectivity model assumed equal proportions of individuals migrating between breeding and nonbreeding MCR pairs in the absence of available data (i.e., null hypothesis; see Appendix S1), this pattern of uniformly distributed connectivity proportions was a result of data limitations for certain species or for certain breeding-nonbreeding MCR connections within species (see Appendix S2: Table S1 for species-specific sample sizes).

For the Swainson’s Hawk case study (Fig. 2b), the average connectivity proportion between a given breeding MCR and each nonbreeding MCR was 0.13, which was close to the expected (null) proportion of 0.11 (i.e., 1 breeding MCR ÷ 9 nonbreeding MCRs; Fig. 2a). Yet our results revealed that proportions were greater than average (i.e., strong connectivity) between the Great Basin (MCR 9) and Northern Prairie (MCR 11) breeding regions of the U.S. and the
Pampas region of Argentina (MCR 232) during the nonbreeding season (0.42 and 0.29, respectively; MCR 11 connections shown in Fig. 2b). Further, greater-than-average proportions of individuals breeding in the Central Valley of California (MCR 15) overwintered in the desert regions of the southern U.S. (Sonoran and Mohave) and northern Mexico (Chihuahuan; MCR 33), and along the Transverse Volcanic Belt (MCR 134) in Mexico (0.23 and 0.22, respectively). Concurrently, Central Valley breeders had lower-than-average proportions of individuals overwintering in the Pampas and Gran Chaco (MCR 134) regions of Argentina (both proportions of 0.06). While we used these mean connectivity proportions (i.e., posterior means) as a method of incorporating migratory connectivity information into LCPs, we acknowledge that, given limited data, differences in mean connectivity proportions were seldom statistically significant (i.e., 95% credible intervals around mean estimates overlapped; see Appendix S2 for more information). All estimated connectivity proportions for each of the 12 focal species are available on Audubon’s GitHub repository: https://github.com/audubongit/migration_data_integration.

Least-cost paths
As designed, LCPs mapped probable migratory pathways and filled in gaps in eBird seasonal occurrence estimates, specifically in areas with a lack of eBird sightings. This pattern was especially apparent for Swainson’s Hawk during post-breeding migration, where LCP indices provided non-zero estimates over much of South America (e.g., Amazonian rainforest) in areas where eBird occurrence probabilities were zero (Fig. 2c&d). Similarly, LCP indices for overwater migrants were non-zero in locations over the Atlantic Ocean and Gulf of Mexico, indicating predicted migratory movements in marine areas where eBird observations were also unavailable (see Appendix S3 for all species-specific results).
In addition to filling in spatial gaps, LCPs mapped migratory bottlenecks, or locations where the geographic funneling of migrants could lead to relatively high probabilities of occurrence. For example, for Swainson’s Hawk during post-breeding migration, the LCP indices over Central America were particularly high relative to the seasonal average occurrence probabilities estimated from eBird (Fig. 2c&d). Likewise, several other species demonstrated high-use migratory areas along peninsulas via greater LCP indices (e.g., Blackpoll Warbler, Grasshopper Sparrow, Prairie Warbler, Tree Swallow in Florida, USA; Appendix S3), as well as stepping stone migratory movements via greater LCP indices in Cuba, where migrants are known to concentrate before or after passage over large expanses of open water (e.g., Blackpoll Warbler, Ovenbird, Prairie Warbler, Prothonotary Warbler; Appendix S3).

Relative model performance

When AIC and deviance explained were the evaluation metrics, GAMMs that included LCP indices (i.e., full models) had either equivalent or superior model fit than those that did not (i.e., eBird-only models; Fig. 4, Appendix S1: Table S4). However, according to AUC, two species-season models (Great Egret and Prairie Warbler during post-breeding migrations) demonstrated poorer fit when LCP indices were included, relative to the eBird-only model. Across the 24 combinations of 12 species and two migratory seasons, the inclusion of LCP indices as predictors in GAMMs increased the deviance explained by an average of 57% (range = 0% – 163%), and increased AUC by an average of 6% (range = -11% – 17%; Appendix S1: Table S4). For Swainson’s Hawk during post-breeding migration (Fig. 3), the addition of LCP indices to the GAMM increased the deviance explained by 67% (Appendix S1: Table S4).
**Integrated prediction surfaces**

The degree to which eBird occurrences and LCP indices were represented in each species’ GAMM-produced integrated prediction surface depended on the agreement between each predictor and observed locations obtained from tracking data and band re-encounters. For Swainson’s Hawk, high values of the integrated index in the northwestern plains of North America were largely due to the spatial agreement between band re-encounter locations (Fig. 3a) and high eBird occurrence probabilities (Fig. 2c). Conversely, high integrated index values across South America (Fig. 3b) were almost entirely due to a congruence between telemetry (GPS and PTT) locations and high LCP index values (Fig 2d). Thus, for this species during the post-breeding migration season, both eBird and LCP prediction surfaces appeared to make important contributions to the integrated surface. Across the remaining species, contributions of the two predictors varied considerably, and depended, in part, on the extent to which each species migrated over areas where eBird observations were sparse. See Appendix S3 for detailed results for each of the 12 focal species.

**Sensitivity analyses**

Correlations between data quality/quantity and model fit (Appendix S1: Table S5) and fit improvement (Appendix S1: Table S6) were all positive. When evaluated together via multiple linear regression, we found weak, positive effects of the number of tracked individuals ($\beta = 0.07$, $p = 0.02$) and number of high-quality tracked individuals ($\beta = 0.04$, $p = 0.06$); a weak, negative effect of number of geolocator-tagged individuals ($\beta = -0.05$, $p = 0.07$); and no detectable effect of spatial comprehensiveness ($\beta = -0.03$, $p = 0.21$).
Discussion

There are many approaches to tracking birds during migration (reviewed in Bridge et al. 2011, McKinnon & Love 2018, Bernard et al. 2021), yet there have been few attempts to unite these technologies with widespread eBird data within a single integrated modeling framework to share information across data sources and account for associated uncertainties and biases (but see Hallworth et al. 2015, Tonra et al. 2019). Our modeling framework leverages the strengths of multiple data types to describe spatial patterns of avian migration across the Western Hemisphere during pre- and post-breeding migratory seasons. We applied our approach to a suite of 12 species with varying life histories to illustrate its reproducibility across different migratory strategies, tracking technologies, range boundaries, and data quantities. Overall, we demonstrated that LCPs contributed additional information to eBird occurrence probabilities, broadening our understanding of migratory patterns at unprecedented spatial extents. Given the rapid declines in migratory species globally (Wilcove & Wikelski 2008, Wilson et al. 2018), knowledge of both the connectivity among populations and the high-use areas during migration can inform conservation strategies to effectively advance species recovery while avoiding expenditure of limited resources on locations that may have little ability to mitigate declines. Although the variability and complexity of biological systems, along with the nuances of data collection, make it difficult to develop a universal integrated model for describing bird migration, this method represents a powerful starting point for further development and customization based on researcher needs and data availabilities (see Appendix S1).

As compared to eBird occurrences alone, we found that inclusion of LCP indices as a covariate in GAMMs generally improved model fit for estimating observed animal locations derived from tracking and band re-encounter datasets. Out of 24 species-season models, all but two (Great Egret and Prairie Warbler during post-breeding migrations) demonstrated either
equivalent or superior fit of the full model (i.e., eBird + LCP) compared to the eBird-only model across three separate performance metrics (Fig. 4). For the two exceptions, fits of the full models were only slightly poorer than the eBird-only models according to AUC (Appendix S1: Table S4). From our post-hoc analysis evaluating correlations between data quantity/quality and model fit, we found that all correlations were positive, indicating that increased data availability and quality improves model fit (Appendix S1: Tables S5 & S6). When we evaluated data availability metrics together, we found that the number of tracked individuals was the most significant predictor of model fit. Specifically, the inclusion of high-quality tracking data (PTT and GPS tags) was positively associated with model fit, which likely accounts for the poorer fit of the full model for Prairie Warbler during post-breeding migration because only LLG data (i.e., tracking technology with the greatest location error) were available for this species (Appendix S3: Fig. S16). Although spatial comprehensiveness of the tracking data was not a significant predictor of model fit across species, we found that species with < 50% of seasonal MCRs containing tracking observations had more variable model performance, which may partly explain the poorer fit of the full model for Great Egret during post-breeding migration (32% of MCRs with tracking observations; Appendix S1: Table S3, Appendix S3: Fig. S10). Taken together, these results suggest that model performance is optimal when high-quality tracking data are well-distributed throughout a species’ range, thereby reducing spatial biases and uncertainties.

Species for which LCPs greatly improved model fit via AIC (i.e., positive outliers in Fig. 4a), compared to eBird occurrences alone, included Osprey and American White Pelican, both of which were rich in high-quality tracking data sources (GPS and PTT; Appendix S1: Table S3, Appendix S3: Figs. S1-2, S11-12). LCPs also offered a valuable contribution to describing avian migratory patterns for species with over-water movements (e.g., Blackpoll Warbler in both
seasons and Prairie Warbler during pre-breeding migration; Appendix S3: Figs. S3-4, S15), and those that migrate over land where there are presumably few eBird sightings (e.g., Broad-winged Hawk and Swainson’s Hawk in South America; Appendix S3: Figs. S5-6, S19-20) and thus low predictive ability of eBird Status and Trends models. Least-cost path analysis was designed to model the flow of individuals (or genes) between points on a heterogeneous landscape (Storfer et al. 2007), and our findings demonstrate that this same logic can be applied to birds migrating (i.e., species-specific occurrence estimates) between grid cells at weekly time steps. By supplementing this approach with prior knowledge on migratory connectivity derived from tracking and band re-encounters (i.e., akin to using informed prior distributions in Bayesian frameworks), we are integrating concepts and technologies, enabling LCPs to reflect the migratory behavior of birds at the species level more accurately.

The complexities of avian migration have contributed to a long-standing gap in our knowledge of the full annual cycle and migratory connectivity (Marra & Studds 2010, Marra et al. 2015). The LCP-derived integrated index of passage probability developed here represents bird space use during migratory seasons and thus captures average patterns at the species level. However, we acknowledge that the integrated index is only a proxy for spatially-explicit migratory abundances and suggest that it be interpreted within the context of a species’ life history to ensure consistency with known migratory patterns. In addition, exploring model modifications, such as spatially-varying coefficients for predictor surfaces or a spatial covariance structure, as well as alternative null hypotheses for connectivity analyses (e.g., abundance-weighted or informed by expert opinion or genetic structure), may help overcome potential species-specific challenges.
Although we selected focal species with the goal of developing a robust and widely applicable method, conducting more extensive sensitivity analyses to fully assess model fit across a range of species groups, sample sizes, and tracking technologies is also warranted (see Appendix S1 for further details). For example, we foresee difficulties in applying this modeling framework to species with unusual migrations (e.g., irruptive species like redpolls, grosbeaks, and nuthatches) or with spatially complex distributions of migratory and non-migratory populations (e.g., American Robin). Species that are primarily aquatic (e.g., terns, loons) or otherwise secretive (e.g., marshbirds), and thus have very low eBird occurrence probabilities terrestrially, will also likely face computational issues during LCP production. While integrating multiple data sources can alleviate some of the limitations of traditional independent analyses, such as filling in spatiotemporal data gaps and increasing precision of parameter estimates (Zipkin & Saunders 2018), the caliber of individual datasets ultimately determines the quality of inference and integrated modeling cannot completely remove biases inherent to each dataset. Nevertheless, the resulting integrated index represents a much-needed initial synthesis of several data types that are frequently collected on migratory birds, enabling new insights that may not have been obtained from separate analyses.

There are several potential avenues for further refining this framework (see Appendix S1: Table S7 for more details). Future developments could consider incorporating additional sources of information on migration, including genetic (Ruegg et al. 2020), isotope (Wassenaar 2019), and Motus (Taylor et al. 2017) data. Although all of these data types are not available for every species, we suggest that researchers could customize the method presented here based on data available for their species of interest, such that a complete, detailed understanding of migratory patterns and connectivity can be achieved. In the absence of tracking and band-re-encounter data,
integrated surfaces could also be produced using the maximum values across eBird occurrence and LCP index surfaces. Indeed, preliminary analyses suggested that using ‘max’ surfaces to generate the integrated index produced qualitatively similar results to those from GAMMs using presence/pseudo-absence locations (see Appendix S1 for more details). For species with long-term tracking data availability that are particularly sensitive to fluctuating environmental conditions (e.g., waterfowl in response to wetland availability), a temporally-varying version of this approach (e.g., annual or decadal) could be developed to understand potential shifts in migratory patterns with climate change. While evidence suggests that birds are generally shortening their migration distances as a result of warming temperatures globally (Visser et al. 2009), the rate and magnitude of these changes vary by species’ migratory behaviors and winter geographies (Rushing et al. 2020). Estimating and predicting human-induced alterations to animal migrations remains an active area of research (Kubelka et al. 2021).

The acquisition and synthesis of knowledge across technologies is needed to inform the ecology and conservation of migratory birds breeding in North America, which have declined by 2.5 billion since 1970 (Rosenberg et al. 2019). Designing multifunctional landscapes is key to supporting migratory species, and this modeling framework represents a critical step toward predicting where and when (e.g., pre- vs. post-breeding migratory periods) species will be moving through locations, which can help conservationists optimize provisioning of habitat (e.g., stopover) or abatement of threats (e.g., wind turbines, light pollution). Hence, we anticipate several creative uses of this work to advance migration and conservation sciences (Appendix S1: Table S7), including (i) filling knowledge gaps regarding the spatial distribution of species with cryptic life histories that hinder direct observation of movements, and (ii) prioritizing localized conservation actions that are responsive to migratory species’ spatiotemporal dynamics.
Uncovering species’ spatial migration patterns during the high-mortality periods of their annual cycles has important implications for migration ecology, migratory bird policies, and conservation actions for populations and species across their ranges. We encourage researchers to explore and advance the methods presented here for other migratory species or systems, which will further expand the scope of inference on spatial patterns of animal movement across hemispheric extents and the full annual cycle.

Acknowledgements
This work was made possible by generous gifts to the Audubon Migratory Bird Initiative (National Audubon Society) from J. Ellis, E. L. Doolin, and B. Doolin. The findings were also made possible thanks to data contributed by community science participants in eBird. This material uses data from the eBird Status and Trends Project at the Cornell Lab of Ornithology, eBird.org. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Cornell Lab of Ornithology. We thank the many field technicians who assisted on the species-specific tracking projects included here. Please see Appendix S4 for the full list of acknowledgements and funding sources related to the species-specific tracking projects. We are grateful to M. van Toor and an anonymous reviewer for insightful comments that improved the manuscript.

Conflict of Interest
The authors declare no conflict of interest.
Author Contributions

TDM conceived the idea in discussion with NLM, SPS, and WVD. JLD, MFJ, EK, NS, MS, LT, and CW obtained and processed tracking and band re-encounter data as part of Audubon’s Migratory Bird Initiative. TDM and SPS conducted analyses with help from WVD, NLM, and JG. SPS led the writing of the manuscript together with TDM and WVD. All authors contributed to the drafts and gave final approval for submission. Authors listed in Appendix S4 collected and analyzed the tracking data for the 12 focal species.

Literature Cited


Meehan et al.


Songbirds’ Departure Decisions, Routes, and Time It Takes to Cross the Gulf of Mexico.”


https://doi.org/10.1073/pnas.1503381112.

Dodge, Somayeh, Gil Bohrer, Keith Bildstein, Sarah C. Davidson, Rolf Weinzierl, Marc J. Bechard, David Barber, et al. 2014. “Environmental Drivers of Variability in the Movement Ecology of Turkey Vultures (Cathartes Aura) in North and South America.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1643): 20130195.


https://doi.org/10.1002/eap.2056.


Kreakie, Betty J., and Timothy H. Keitt. 2012. “Integration of Distance, Direction and Habitat into a Predictive Migratory Movement Model for Blue-Winged Teal (Anas Discors).” 


Nourani, Elham, Kamran Safi, Noriyuki M. Yamaguchi, and Hiroyoshi Higuchi. n.d. “Raptor Migration in an Oceanic Flyway: Wind and Geography Shape the Migratory Route of Grey-


Figure Legends

Figure 1. Conceptual schematic of the modeling framework, which integrates multiple data types to describe spatial patterns of avian migration across the Western Hemisphere. For each species and migratory season, we (a) statistically integrated tracking and band re-encounter data to (b) estimate migratory connectivity proportions. Migratory connectivity proportions were then used along with (c) eBird Status and Trends products (i.e., occurrence probabilities) to produce (d) least-cost paths. We used the least-cost path indices, as well as eBird occurrence probabilities, as covariates in generalized additive mixed models (GAMMs) to evaluate the strength of association between each predictor and presence/pseudo-absence data (i.e., response variable) obtained from tracking and band re-encounter datasets. The resulting (e) GAMM prediction surface represents the statistical integration (i.e., integrated index) of all available migration information shown.

Figure 2. (a) Migratory connectivity regions (MCRs) in North, Central, and South America, which we used as the spatial units of the modeling framework. We use Swainson’s Hawk as an example species to illustrate the breeding (green) and nonbreeding (purple) MCRs selected (i.e., contain ≥ 1% of total seasonal eBird relative abundance) for inclusion in analyses. Green and purple points represent high-abundance clusters (i.e., sum to 30% of abundance) within each MCR in stationary breeding and nonbreeding seasons, respectively. (b) Estimated proportions from the migratory connectivity analysis for Swainson’s Hawk that integrates tracking and band re-encounter data from stationary breeding and nonbreeding periods; for ease of illustration, the proportions shown refer to a single breeding MCR and thus sum to one. (c) Average post-breeding eBird Status and Trends (eBird ST) relative occurrence probabilities of Swainson’s
Hawk (i.e., arithmetic mean of weekly estimates during the post-breeding migration period; see Appendix S1: Table S2 for species-specific migration timings). (d) Least-cost path (LCP) index for Swainson’s Hawk illustrating the probable pathways during post-breeding migration. Icons in bottom left corner of each panel refer to a specific stage of the modeling framework and correspond with those shown in Fig. 1.

Figure 3. (a) Available band re-encounter (green), GPS (orange), and PTT (purple) data for Swainson’s Hawk during post-breeding migration. These data were considered ‘known presence’ locations and combined with 10,000 background points (i.e., pseudo-absence locations) and used as the response variable in generalized additive mixed models (GAMMs) to create (b) the integrated index illustrating bird migration patterns across the Western Hemisphere (Swainson’s Hawk shown here). See Appendix S1: Table S3 for sample sizes of tracking and band re-encounter datasets that were used in GAMMs for each species-season. Icons in bottom left corner of each panel refer to a specific stage of the modeling framework and correspond with those shown in Fig. 1.

Figure 4. Evaluation metrics, (a) Akaike’s Information Criterion (AIC), (b) Deviance explained, and (c) Area Under the Receiver Operator Curve (AUC), used to assess the difference (Δ) in fits between two generalized additive mixed models (GAMMs): (i) a full model including eBird occurrence probabilities + least-cost path (LCP) passage indices, and (ii) a model including only eBird occurrence probabilities as predictors of presence/pseudo-absence data (derived from tracking and band re-encounter datasets during pre- and post-breeding migrations). Positive values (above horizontal dashed line) indicate improvements in model fit with the addition of
LCP indices (e.g., $AIC_{eBird} - AIC_{full}$ in (a), $DevExp_{full} - DevExp_{eBird}$ in (b), $AUC_{full} - AUC_{eBird}$ in (c)); negative values (below horizontal dashed line) indicate poorer fit of the full model relative to the eBird-only model; values at zero (on horizontal dashed line) indicate equivalent model fits. Results from 24 species-season model comparisons are shown, with yellow boxplots showing the distribution of model fits during post-breeding migration (i.e., fall) and green boxplots showing the distribution of model fits during pre-breeding migration (i.e., spring). Horizontal lines in boxplots indicate median values, bounds of boxes indicate 25th and 75th percentiles, and labeled points represent outliers. See Appendix S1: Table S4 for additional model fit results.