Received: 19 September 2023 | Accepted: 12 September 2024

DOI: 10.1111/2041-210X.14467

Methods in Ecology and Evolution

RESEARCH ARTICLE

Challenges and opportunities for data integration to improve estimation of migratory connectivity

Jeffrey A. Hostetler[1,2] | **Emily B. Cohen**[3] | **Christen M. Bossu**[4] | **Amy L. Scarpignato**[5] | **Kristen Ruegg**[4] | **Andrea Contina**[6] | **Clark S. Rushing**[7] | **Michael T. Hallworth**[8]

[1]U.S. Geological Survey, Eastern Ecological Science Center, Laurel, Maryland, USA; [2]Division of Migratory Bird Management, US Fish and Wildlife Service, Laurel, Maryland, USA; [3]University of Maryland Center for Environmental Science, Appalachian Laboratory, Frostburg, Maryland, USA; [4]Department of Biology, Colorado State University, Fort Collins, Colorado, USA; [5]Smithsonian's National Zoo and Conservation Biology Institute, Migratory Bird Center, National Zoological Park, Washington, District of Columbia, USA; [6]School of Integrative Biological and Chemical Sciences, The University of Texas Rio Grande Valley, Brownsville, Texas, USA; [7]Warnell School of Forestry and Natural Resources, University of Georgia, Athena, Georgia, USA and [8]Vermont Center for Ecostudies, White River Junction, Vermont, USA

Correspondence
Jeffrey A. Hostetler
Email: jhostetler@usgs.gov

Emily B. Cohen
Email: emily.cohen@umces.edu

Funding information
National Science Foundation, Grant/Award Number: 1942313

Handling Editor: Chris Sutherland

Abstract

1. Understanding migratory connectivity, or the linkage of populations between seasons, is critical for effective conservation and management of migratory wildlife. A growing number of tools are available for understanding where migratory individuals and populations occur throughout the annual cycle. Integration of the diverse measures of migratory movements can help elucidate migratory connectivity patterns with methodology that accounts for differences in sampling design, directionality, effort, precision and bias inherent to each data type.
2. The R package *MigConnectivity* was developed to estimate population-specific connectivity and the range-wide strength of those connections. New functions allow users to integrate intrinsic markers, tracking and long-distance reencounter data, collected from the same or different individuals, to estimate population-specific transition probabilities (estTransition) and the range-wide strength of those transition probabilities (estStrength). We used simulation and real-world case studies to explore the challenges and limitations of data integration based on data from three migratory bird species, Painted Bunting (*Passerina ciris*), Yellow Warbler (*Setophaga petechia*) and Bald Eagle (*Haliaeetus leucocephalus*), two of which had bidirectional data.
3. We found data integration is useful for quantifying migratory connectivity, as single data sources are less likely to be available across the species range. Furthermore, accurate strength estimates can be obtained from either breeding-to-nonbreeding or nonbreeding-to-breeding data. For bidirectional data,

Jeffrey A. Hostetler and Emily B. Cohen contributed equally to this manuscript.

integration can lead to more accurate estimates when data are available from all regions in at least one season.

4. The ability to conduct combined analyses that account for the unique limitations and biases of each data type is a promising possibility for overcoming the challenge of range-wide coverage that has been hard to achieve using single data types. The best-case scenario for data integration is to have data from all regions, especially if the question is range-wide or data are bidirectional. Multiple data types on animal movements are becoming increasingly available and integration of these growing datasets will lead to a better understanding of the full annual cycle of migratory animals.

KEYWORDS
animal migration, data integration, full annual cycle, genoscape, MC, migratory connectivity, population distribution, transition probability

1 | INTRODUCTION

Migratory connectivity describes the distribution of populations between seasons that results from animal migration. Tracking and other technologies are increasing our ability to describe migratory movements for more species and in greater detail. Understanding the distribution of populations between seasons is critical to elucidating the ecology and evolution of migratory species, information that has become more urgent with the declines of migratory animals (Bairlein, 2016; Rosenberg et al., 2019; Wilcove & Wikelski, 2008). However, the variability of migratory behaviours and differences in the technologies used to measure them has made it challenging to quantify their migratory connectivity.

Migratory connectivity is a growing field of study (Gregory et al., 2023; Marra et al., 2018), but recognition of the need to understand it is not new. Salomonsen (1955) originally defined the terms synhiemy and allohiemy to describe the convergence or segregation, respectively, of breeding populations on the nonbreeding grounds. Webster et al. (2002) coined the term migratory connectivity, to describe both the distribution of population-specific connections and the convergence or segregation of populations that occurs along a continuum from strong to weak. The field has further expanded to clarify the importance of the distribution of breeding populations during migratory phases, which inherently have a temporal component (Briedis & Bauer, 2018; Cohen et al., 2019; Knight et al., 2021). Thus, the term migratory connectivity is inclusive of both the **pattern**, the geographic linking of populations between seasons (i.e. transition probabilities or movement probabilities), and the **strength**, the extent or degree of population convergence between two seasons (e.g. MC metric; Cohen et al., 2018). For example, when migratory connectivity is strong (e.g. Willets [*Tringa semipalmata*]; Huysman et al., 2022), breeding populations largely remain segregated during the nonbreeding seasons, as is expected to be the case under chain or leapfrog migration (Newton, 2010). On the other hand, when migratory connectivity is weak (e.g. Prothonotary warbler [*Protonotaria citrea*]; Tonra et al., 2019), many breeding populations converge in the same nonbreeding areas as is expected to be the case under telescopic migration (Newton, 2010).

There are many different tools and technologies for studying migratory connectivity, but most studies remain limited by small sample sizes and incomplete sampling. In particular, sampling rarely covers the full extent of a species' range, and decisions about where to sample are often dictated by logistics rather than information about population structure (Huysman et al., 2022). Incomplete sampling can bias estimates; when migratory connectivity is measured from only part of the range, estimates of the strength tend to be lower than when they are measured from across the range of the species (Sharp et al., 2023). For example, Phipps et al. (2019) found strong migratory connectivity for Egyptian vultures (*Neophron percnopterus*) among subspecies but weak connectivity when sampling only some populations. While range-wide measures are likely to be the most biologically meaningful, they require more complete sampling of populations than has been possible from most single data types.

Data types for studying seasonal distributions of migratory animals include individual tracking (i.e. from geolocator or global positioning tags), assignments from intrinsic markers (i.e. using stable isotope values from inert tissues such as feathers, or genetic markers), and mark reencounters (i.e. from banding or ringing schemes; Marra et al., 2018). Each of these tools has limitations in precision, bias and/or reencounter probability. As data to measure the connections of migratory populations continue to become available, the integration of multiple data types is a potential solution to improve estimation of migratory connectivity in the face of sampling constraints. Beyond differences in bias and precision, data types often differ in both sample sizes and distribution across the range (Hagelin et al., 2021; Hill & Renfrew, 2019; Huysman et al., 2022; Tonra et al., 2019). Furthermore, sampling efforts have often not accounted for differences in abundance among the sampled populations (Cohen et al., 2018). Combined analyses that account for the

unique limitations and biases of each data type is a promising possibility for overcoming the challenge of range-wide coverage that has been hard to achieve using single data types (Contina et al., 2022), especially given the high costs and logistical challenges in collecting these data across the broad ranges of many migratory species (Korner-Nievergelt et al., 2017; Procházka et al., 2017).

The *MigConnectivity* R package, which quantifies the pattern and strength of migratory connectivity, was developed to facilitate quantitative comparisons across studies, data types and taxa while accounting for uneven sampling and the bias inherent to different data types for studying migratory movement (Cohen et al., 2018; Hostetler & Hallworth, 2024). We developed new functionality in the package to overcome key challenges prohibiting integration of animal movement data types and improve estimates of migratory connectivity. One key challenge to data integration is in the diversity of data types available for studying seasonal distributions of migratory animals. However, some prominent data types used to measure migratory connectivity were not included in previous versions (Cohen et al., 2014). Genoscapes are spatially explicit maps of genetic variation across the breeding range, to assign individuals sampled during nonbreeding and migratory periods to their most likely breeding population; it is a relatively new method that is increasingly used to measure migratory connectivity (Ruegg et al., 2014). Banding, on the contrary, may be the oldest method for measuring migratory movements, and considerable data have accumulated for some bird species, but sampling is typically incomplete with highly variable reencounter probabilities (Thorup et al., 2014). We address the data integration challenge of the diversity of available data types by adding functionality for genoscape assignment and capture–mark–reencounter (CMR; hereafter, banding) data.

Another key challenge to the integration of migration data is incorporating measures of movement in opposite directions (i.e. breeding to nonbreeding and vice versa) with assignment error (also known as location error) on different sides. For example, archival geolocators are often placed on animals in the breeding season and, if retrieved in the subsequent breeding season, provide information about where that animal spent the nonbreeding season with varying degrees of assignment error. Isotope or genetic samples are often obtained from tissues collected during the nonbreeding season, providing information about where that animal spent the previous breeding season, again with varying degrees of assignment error. We developed methods to account for assignment error when integrated data are collected in two directions, the 'to' and 'from' side of estimated transition probabilities.

Here we introduce new analytical approaches for estimating both the pattern and strength of migratory connectivity through data integration. We provide methods for using any combination of GPS/telemetry, light-level geolocator, isoscape (spatially explicit maps of environmental isotopic variation, such as hydrogen isotope ratios in precipitation, used to assign individuals sampled elsewhere to the location of origin of tissue growth; West et al., 2009), genoscape and band reencounter data collected from the same or different individual animals to estimate patterns and strength between any two phases of the annual cycle (Table 1). Using simulation, we assess when data integration can improve accuracy including the role of sampling relative to abundance, incomplete sampling and sample size. We also apply these methods to real-world case studies for three species: Painted Bunting (*Passerina ciris*), Yellow Warbler (*Setophaga petechia*) and Bald Eagle (*Haliaeetus leucocephalus*).

2 | METHODS

2.1 | Estimating migratory connectivity

In earlier versions of the *MigConnectivity* R package, transition probabilities were an intermediate step in estimating the strength of migratory connectivity (MC; Cohen et al., 2018, 2019). We adapted these methods to place greater emphasis on population-specific patterns of connectivity by separating estimation of MC into two new functions, estStrength and estTransition (Table 1). Here, we describe these new functions, two new data types added to the package, and data integration methods for estimating the pattern and strength of migratory connectivity, including bidirectional data.

Consider a migratory species where the breeding range is divided into B regions and its nonbreeding range is divided into NB regions. The **pattern** of migratory connectivity and abundance for this species can be given as transition probabilities ($\boldsymbol{\Psi}$), a matrix with each cell $\psi_{b,w}$ being the probability that an animal will migrate to nonbreeding region w given that it's from breeding region b and R, a vector of the relative (proportional) abundances of the B breeding regions. We can more generally call the first set of regions **origin regions** and the second set **target regions**, referring to the direction of migration conceptualized by these transition probabilities, not necessarily to where the data are collected. For example, if one wishes to estimate transition probabilities from nonbreeding regions to breeding regions ($\gamma_{w,b}$), along with nonbreeding relative abundances (W), then the nonbreeding regions become the origin regions and the breeding regions the target regions, regardless of where tags were deployed or intrinsic data collected. These patterns can also be reparametrized in at least one other way (Figure 1; Supporting Information 1). If one is referring to a single migration season or animals do not disperse between breeding or between nonbreeding regions (migrating animals return to the region whence they came), these parameterizations are equivalent, and it is simple to convert between them using the new *MigConnectivity* function reverseTransition (version 0.4, Table 1; also see Supporting Information 1 for the straightforward conversion equations).

The **strength** of migratory connectivity summarizes the pattern of migratory connectivity as the degree of population convergence expressed by MC (Cohen et al., 2018). We calculate MC from the distances between each pair of breeding regions (D), distances between each pair of nonbreeding regions (V), $\boldsymbol{\Psi}$ and R. We assume that the distances between regions can be calculated without relevant error or uncertainty, but that is generally not true of transition probabilities or relative abundances (among either breeding or nonbreeding regions).

TABLE 1 Names and descriptions of new functions added to the *MigConnectivity* R package (version 0.4; Hostetler & Hallworth, 2024) and used in these analyses.

Function	Description
calcTransition	Calculate maximum likelihood $\boldsymbol{\Psi}$ (transition probabilities between regions in two phases of the annual cycle) without estimating uncertainty or accounting for potential assignment error (e.g. from geolocators). If there are no banding data, $\boldsymbol{\Psi}$ is proportional to the number of observed animals in each combination of regions. If there are banding data, the function accounts for possible differences in reencounter probability
estTransition	Estimate $\boldsymbol{\Psi}$ (transition probabilities between regions in two phases of the annual cycle) with uncertainty. Data can be from any combination of geolocators (GL), telemetry/GPS, intrinsic markers such as isotopes and genetics, and banding data
estStrength[a]	Estimate **MC** (migratory connectivity strength) from estimates of $\boldsymbol{\Psi}$ and R (origin region relative abundance), D (distances between each pair of origin regions) and V (distances between each pair of target regions)
reverseTransition	Reverse $\boldsymbol{\Psi}$ and R estimates to calculate or estimate target site to origin site transition probabilities ($\boldsymbol{\Gamma}$), target region relative abundances (W) and origin/target site combination probabilities ($\boldsymbol{\Pi}$)
simCMRData	Simulate banding (capture–mark–reencounter; CMR) migratory movement data
simGLData	Simulate geolocator (GL) migratory movement data
simProbData	Simulate Dirichlet-based probability table data
simTelemetryData	Simulate telemetry/GPS migratory movement data

[a]Function estStrength was introduced in Roberts et al. (2023).

When one or both are quantified accounting for and expressing its uncertainty, we call that an estimate. The new function estStrength replaces and extends the function estMC, providing estimates of MC from estimates of transition probabilities and relative abundances, and propagating uncertainty from both of those estimates into the MC estimates (Table 1; Roberts et al., 2023). We focus here on estimating transition probabilities with data integration, as estimating abundance has been covered elsewhere (e.g. Johnston et al., 2015; Royle, 2004; Sauer & Link, 2011).

Another function added to the *MigConnectivity* package is estTransition (Table 1). It estimates $\boldsymbol{\Psi}$ from a variety of possible data types: light-level geolocators (point locations in two seasons with assignment error in one of them), telemetry (defined for our purposes as point locations in two seasons with no assignment error or detection heterogeneity, such as from satellite GPS), genetic or stable isotope assignments (point locations in one season with a raster layer with probabilities of assignment in the other or point locations in one season with probability tables of assignment in the other), and banding (point locations in two seasons with detection heterogeneity in one). The user can choose between estimation methods: a combined nonparametric/parametric bootstrap or Markov chain Monte Carlo (MCMC; the latter currently only handles telemetry and banding data, but generally runs much faster). Besides speed, the biggest practical effect of choosing MCMC over bootstrap is that when data are sparse, the priors included in MCMC (or any Bayesian analysis) will influence $\boldsymbol{\Psi}$ estimates. In some cases, the user may have a priori information that certain transition probabilities ($\psi_{b,w}$) are unlikely (Thorup et al., 2014; Thorup & Conn, 2009), and there are two options for using that information in the analysis: fixing them to zero (either estimator) or changing the default flat Dirichlet prior on $\psi_{b,*}$ from all ones to something biologically meaningful and informative (MCMC estimator only; Kéry & Royle, 2020). The user can also indicate when individual animals have more than one type of marking, which may affect the uncertainty of estimates.

2.1.1 | Banding data

To estimate transition probabilities from banding data, we used a simplified multistate model closely related to the division coefficient (Kania & Busse, 1987; Korner-Nievergelt et al., 2010). We developed methods to fit this model using both MCMC (von Rönn et al., 2020) and bootstrapping approaches. Both of these methods facilitate integrating banding data with other data types to estimate $\boldsymbol{\Psi}$; we allow for the integration of banding and telemetry data using MCMC and banding and geolocator, telemetry, genetic (see Section 2.1.2) and/or stable isotope data using the bootstrap method. For the bootstrapping approach, the estTransition function samples with replacement from all banded animals, including those never reencountered. It then calls the new calcTransition function (Table 1) for that set of animals, which, for banding data, finds the maximum likelihood solution to equation S18 (Supporting Information 1) using the optim function (Bolker, 2008). For the MCMC approach, the estTransition function calls the program JAGS using the R2jags interface

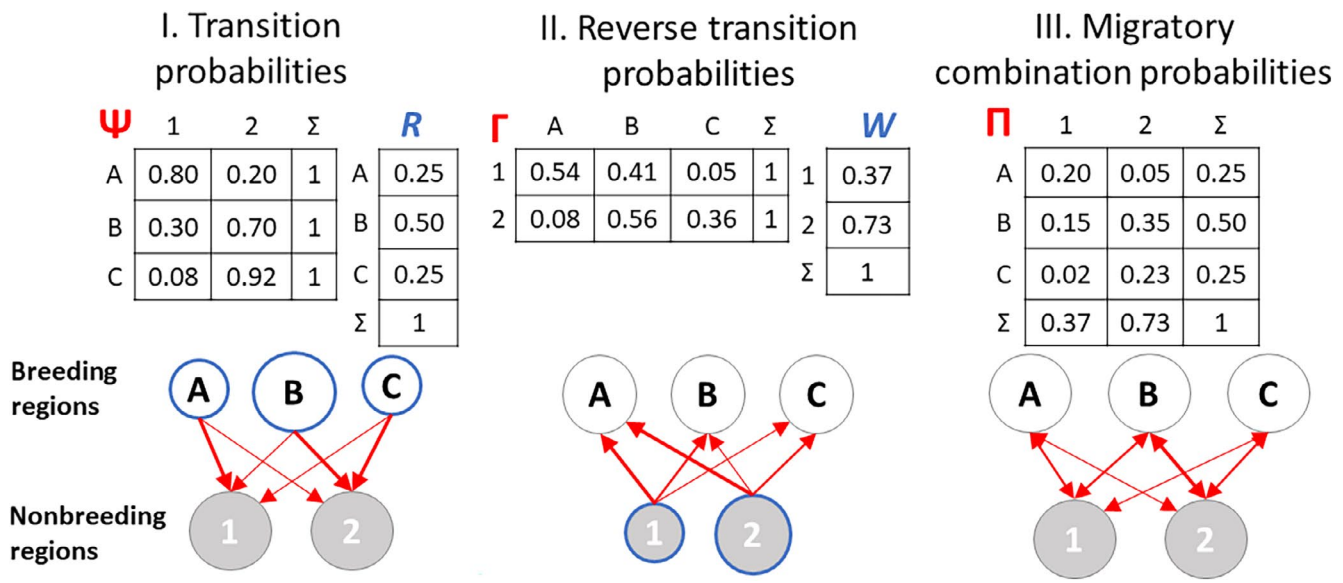


FIGURE 1 Examples of the three equivalent parameterizations of abundance and migratory connectivity patterns for a species with *B* = three breeding origin regions (depicted here as A, B and C) and NB = two target nonbreeding regions (depicted here as 1 and 2). In parameterization I (first column), which could correspond to geolocator data sampled on the breeding ground, $\boldsymbol{\Psi}$ is transition probabilities, a *B* by NB matrix where each row sums to 1 and *R* is relative abundances among the breeding regions, a *B* length vector that sums to 1. In parameterization II (second column), which could be an example of intrinsic markers like genetics or stable isotopes, $\boldsymbol{\Gamma}$ is reverse transition probabilities, a NB by *B* matrix where each row sums to 1, and *W* is a NB length vector of nonbreeding relative abundances that sums to 1. The values for $\boldsymbol{\Gamma}$ provided in the figure are rounded. In parameterization III (third column), $\boldsymbol{\Pi}$ is the migratory combination probabilities, a *B* by NB matrix where all entries together sum to 1. The directions of the arrows implicit in each parameterization differ: $\boldsymbol{\Psi}$ are breeding-to-nonbreeding transition probabilities, $\boldsymbol{\Gamma}$ are the nonbreeding-to-breeding transition probabilities, and $\boldsymbol{\Pi}$ are the joint probabilities of making each migratory transition. Note that *R* is identical to the row sums of $\boldsymbol{\Pi}$ and *W* is identical to its column sums.

(Plummer, 2003; Su & Yajima, 2021), with BUGS code adapted from von Rönn et al. (2020).

2.1.2 | Genetic data

A genoscape is a spatially explicit map of genetic variation across the geographic extent of the breeding range (Ruegg et al., 2014). The genoscape is constructed from genomic sequencing of hundreds of individuals across a species breeding range followed by population structure analyses using admixture (Alexander et al., 2009) or NgsAdmix (Skotte et al., 2013) to provide the baseline genetic structure of a species. We used two approaches to assign individuals sampled on the nonbreeding range to their breeding origin. For population-level questions, we assign individuals to genetically distinct breeding populations characterized by the genoscape using the R package *rubias* (Anderson & Moran, 2022). The output is a probability table of assignment to each defined breeding population. For questions that require a finer resolution estimate of breeding location, we estimate the geographic origin of individuals based on their genetic backgrounds using the R package *OriGen* (Rañola et al., 2014). The output of *OriGen* is a raster surface for each individual corresponding to the probability of assignment to each pixel within the breeding range of the species.

We can use these individual genoscape assignments to estimate species-level $\boldsymbol{\Psi}$ including propagation of assignment uncertainty, using a combined nonparametric and parametric bootstrap, in which we sample animals with replacement and then sample from the uncertainty associated with each sampled individual animal's location in the breeding season (Cohen et al., 2018). The estimation of $\boldsymbol{\Psi}$ (and MC) from isotope raster surfaces is well established (Cohen et al., 2019), and sampling from genoscape raster surfaces can be done in a similar way. To integrate intrinsic markers (genetic and stable isotope raster data), one can use an application of Bayes' Rule to create a combined raster (Ruegg et al., 2017) before calling estTransition. Sampling from probability of assignment tables is even simpler (and much faster computationally). After sampling with replacement from individual animals with probability table data, estTransition samples a breeding region for each animal, with probabilities of selecting regions taken from that animal's row of the probability table.

2.1.3 | Data integration

In previous efforts, $\boldsymbol{\Psi}$ was estimated using only animals first captured on the breeding range, where any assignment error or detection heterogeneity is on the nonbreeding range (Cohen et al., 2018). The reverse transition probabilities ($\boldsymbol{\Gamma}$) were only estimated where any assignment or detection error was on the breeding range. If one has estimates of relative abundance from the same range as animals were captured, that range can be designated the origin range, and $\boldsymbol{\Psi}$ can be converted to $\boldsymbol{\Gamma}$ or vice versa

(Figure 1). However, interest in integrating disparate data sources to estimate migratory connectivity has motivated a new question: under what circumstances can Ψ be estimated by integrating bidirectional data?

In Supporting Information 1, we demonstrate that for bidirectional or nonbreeding-alone data the combined nonparametric and parametric bootstrap with equal weighting of all data points introduced in Cohen et al. (2018) provides an unbiased estimate of Ψ only when sampling occurs at all nonbreeding regions in proportion to abundance (regardless of how the breeding regions were sampled). However, an unbiased estimate of Ψ is possible with a weighted bootstrap, sampling data points with replacement in proportion to $\frac{W}{n^{(NB)}}$, where $n^{(NB)}$ is the sample size collected in each nonbreeding region (see Supporting Information 1 for the weights with data going both directions; Hesterberg, 2011). However, when not all nonbreeding regions are sampled, this approach won't work, as there are no data points from those regions to weight. Therefore, for Ψ to be estimated with bidirectional data, sampling should occur across all regions on the nonbreeding side; and since sampling in proportion to abundance is unlikely, bootstrap samples of nonbreeding data should be weighted by estimates of relative abundance.

2.2 | Testing migratory connectivity estimates

We used simulation scenarios and three real data case studies to show the conditions under which it is possible to estimate migratory connectivity from data integration, including bidirectional data. The fundamental purpose of our simulations and case studies was to address a few important questions about the feasibility and helpfulness of (bidirectional) data integration given real-world sampling limitations. The full set is in Supporting Information 2, but key questions include: *Does data integration improve estimate accuracy under realistic sample sizes? Does it matter whether sampling was not in proportion to abundance? Does it matter whether sample sizes differ for the data types? Can we obtain accurate estimates by integrating data when some regions are not sampled?*

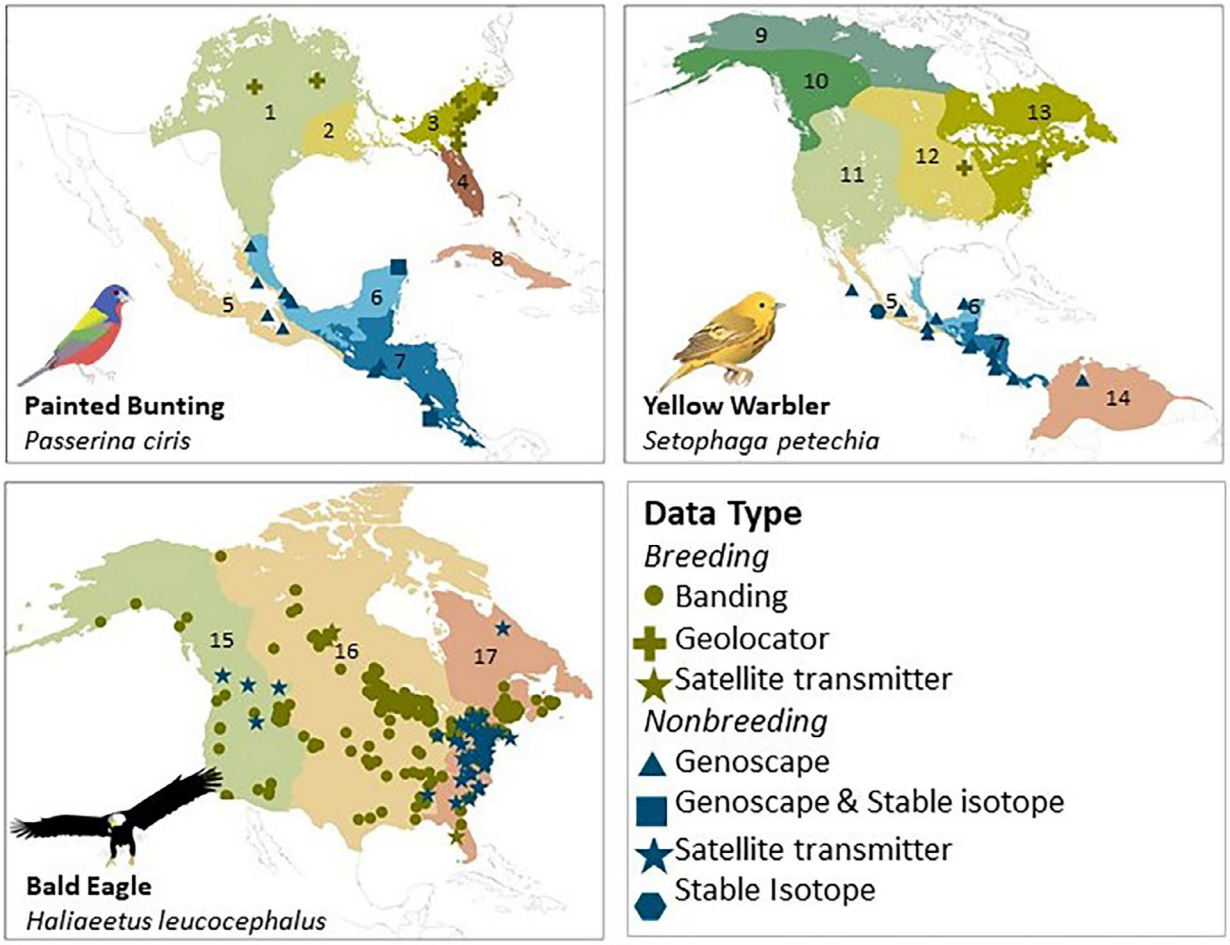
2.2.1 | Analyses with simulated data

We simulated a range of data sampling scenarios to evaluate whether estimates of Ψ and MC improve with data integration methods and the role of sampling design in the value of data integration (Supporting Information 2; Table S1). All sampling scenarios were tested on a biological example loosely based on the real-world case study for Yellow Warbler (see *Case studies with real data*), including plausible Ψ and R values and the number of breeding populations and nonbreeding regions, as well as the distances between them. For comparison with the biologically informed example, a subset of sampling scenarios was tested on a simpler, more hypothetical example with more uniform values of Ψ and W and more variable R values (Supporting Information 2). We simulated 1000 datasets for each scenario using functions simGLData and simProbData, and from each dataset estimated Ψ using the function estTransition (Table 1). For simplicity, we assumed relative abundances among breeding populations and among nonbreeding regions were known (without uncertainty). We estimated MC from Ψ and R using the function estStrength (Table 1). We evaluated the performance of Ψ and MC estimates for different scenarios using metrics for bias (mean of estimates minus the true known value), precision (spread of estimates) and accuracy (mean absolute error). A fourth metric, coverage, measures the accuracy of the error estimates (proportion of 95% CI that overlap the true value). The results for each metric were categorized as good, fair or poor (Table S2; see Supporting Information 2 for more details about the simulations).

2.2.2 | Case studies with real data

We demonstrate the methods and explore the usefulness of data integration for three bird species with multiple available datasets that differ in type, sample size and distribution within the range (Figure 2). For two species, the Painted Bunting and Yellow Warbler, we implement these methods using published genetic and stable isotope data on migratory movements from birds captured and sampled in the nonbreeding season and geolocator data from birds tagged and re-captured in the breeding season. For a third species, the Bald Eagle, we integrate published satellite tracking (Robinson et al., 2010) and long-distance band reencounter data between seasons. Estimation of range-wide migratory connectivity requires the measurement of movement between seasonally defined regions. For the first two species, we used genetically defined breeding population extents (Bay et al., 2021; Rueda-Hernández et al., 2023) and pre-defined nonbreeding ecoregions (Commission for Environmental Cooperation, 1997; Griffith et al., 1998) and for Bald Eagle, we used the three North American flyways boundaries (Pacific, Midcontinent and Atlantic). We estimated MC using the function estStrength from estimated Ψ and R.

The three case studies differed in the data types used to estimate Ψ, but all three used community science database eBird average **relative abundance** estimates during the breeding season at 9 km resolution (Fink et al., 2022), summed over the defined breeding regions and divided by the total to calculate R. Although the eBird relative abundance estimates provide measures of uncertainty (standard deviation and 95% confidence interval [CI]), they do not yet include covariance between cells, making it impossible to properly estimate uncertainty in R. Therefore, we assumed no uncertainty in relative abundances for the purpose of this demonstration. Depending on the needs of the study, one can use posterior samples of R estimates from one's own abundance analysis in estStrength (e.g. from Breeding Bird Survey data; simplified version of analysis available in modelCountDataJAGS function; Hostetler & Hallworth, 2024). For the Yellow Warbler, we



Map label	Region name	Species and season
1	Central	Painted Bunting Breeding
2	Gulf Coast	Painted Bunting Breeding
3	East	Painted Bunting Breeding
4	Southeastern US	Painted Bunting Nonbreeding
5	Pacific and Central Mexico	Painted Bunting & Yellow Warbler Nonbreeding
6	Atlantic Lowland Mexico	Painted Bunting & Yellow Warbler Nonbreeding
7	Central America	Painted Bunting & Yellow Warbler Nonbreeding
8	Caribbean	Painted Bunting Nonbreeding
9	Arctic	Yellow Warbler Breeding
10	Pacific Northwest	Yellow Warbler Breeding
11	Southwest	Yellow Warbler Breeding
12	Central	Yellow Warbler Breeding
13	East	Yellow Warbler Breeding
14	South America	Yellow Warbler Nonbreeding
15	Pacific	Bald Eagle Breeding and Nonbreeding
16	Midcontinent	Bald Eagle Breeding and Nonbreeding
17	Atlantic	Bald Eagle Breeding and Nonbreeding

FIGURE 2 Sampling locations for data types integrated to measure transition probabilities and strength of migratory connectivity between breeding and nonbreeding regions for the Painted Bunting, Yellow Warbler and Bald Eagle. For the Painted Bunting and Yellow Warbler, the breeding regions are defined by genoscape-defined populations and the nonbreeding regions were identified by ecoregions across the ranges. For the Bald Eagle, the breeding and nonbreeding regions are the three North American administrative flyways. See text for data source citations.

had data from all nonbreeding regions that we weighted by nonbreeding relative abundance *W*, and we used eBird data and the methods described above for breeding regions to calculate this. See Supporting Information 3 for more details about the data and methods used in the case studies.

3 | RESULTS

3.1 | Analyses with simulated data

We used simulation to explore the influence of sampling on estimates of **Ψ** and MC with data integration and found two general results that were consistent across scenarios. One broad finding was that where the true ψ value was on or near a boundary (0 or 1, meaning no movement or complete movement between regions), the coverage was never good, and particularly, it was always poor when the true value was exactly 0 (Table S6). However, the point estimate bias, precision, accuracy for boundary ψ were still good (or fair), even where most other ψ in the scenario were not (Tables S3–S5). For the hypothetical example, where there were no ψ values at boundaries (values were all at least 0.1 and less than 0.9), the coverages were still mostly fair or poor, suggesting poor coverage may also be influenced by small sample sizes (Table S11). Another general result was that all MC estimates from the biologically informed example had minimal bias and good accuracy and precision across sampling scenarios (Table S7). Most of the biologically informed example's sampling scenarios also had good MC coverage; only one related to incomplete sampling (see below) had poor coverage. This was not true for the hypothetical example scenarios, demonstrating that MC is robust to inaccuracy in **Ψ** estimates in some cases but not in all. Below, we summarize the results related to our key questions about the role of sampling for data integration (Table S1), and additional results can be found in Supporting Information 2.

3.1.1 | Does data integration improve estimate accuracy with realistic sample sizes?

When breeding and nonbreeding-collected data were both proportional to regional abundances, integrating the data sources greatly improved **Ψ** precision over breeding data alone but had little benefit over nonbreeding data alone, likely due to the differences in sample sizes (60 geolocators vs. 300 genetic samples; Table S4; Figure S4). None of the scenarios showed much bias (Table S3). Accuracy results showed some improvement over breeding data alone but little over nonbreeding data alone (Table S5). The coverage from the data integration estimates was most similar to the coverage from the estimates using only nonbreeding data (mostly fair; Table S6), even though coverage of estimates based only on breeding data was better, again suggesting the strong influence of sample sizes. The MC estimates from data integration had slightly higher precision than the estimates from nonbreeding data alone and much higher precision than estimates from breeding data alone (Table S7). When breeding and nonbreeding data were not proportional to abundance, but nonbreeding data were weighted, the same patterns held (both examples; Figures S5 and S11; Tables S3–S12). In summary, data integration with real-world sample sizes improves precision of **Ψ** estimates, and it can improve accuracy and coverage; however, estimates are strongly influenced by the data type with the larger sample size, and all **Ψ** estimates are generally unbiased, with or without data integration.

3.1.2 | Does it matter whether sampling was not in proportion to abundance?

In the biologically informed example, integration of data that were either sampled proportional to abundance or nonproportional but weighted by abundance generally both resulted in transition probability estimates that were unbiased, precise and accurate, although coverage was generally only fair (Figure S6; Tables S3–S6). When sampling was nonproportional and not weighted by abundance, estimates were generally biased (Figure S6; Table S3). The largest negative biases were away from the nonbreeding region with the highest abundance, and the largest positive biases were toward the nonbreeding regions with the lowest abundance. This also impacted the accuracy and coverage of this scenario (Tables S4 and S6). In the hypothetical example, the unweighted ψ estimates were also biased, but less so than in the biologically informed example, likely due to the nonbreeding abundances being closer to equal (Table S8; Figure S12). However, in this example the biased transition probabilities also resulted in a biased estimate of MC (Figure S12; Table S12). These results demonstrate it is important to either sample in proportion to abundance or weight nonproportional data by abundance, the latter of which is generally possible to do after samples are collected.

3.1.3 | Does it matter whether sample sizes differ for the data types?

Genetic and stable isotope sampling from nonbreeding to breeding typically results in far greater sample sizes than does tracking from breeding to nonbreeding. However, if nonbreeding sample sizes decrease to approximate typical breeding sample sizes, the benefits of bidirectional data integration became clearer (Figures S5 and S7; Figure 3): precision and accuracy of **Ψ** estimates improve with more data, including integrated data (Tables S4 and S5). Coverage showed a different pattern; breeding estimates had the best coverage, so less nonbreeding data improved the coverage of integrated estimates (Table S6). None of these scenarios had much bias in **Ψ** estimates (Table S3). MC estimates were also more precise with data integration and larger total sample sizes (Table S7; Figure 3).

Methods in Ecology and Evolution

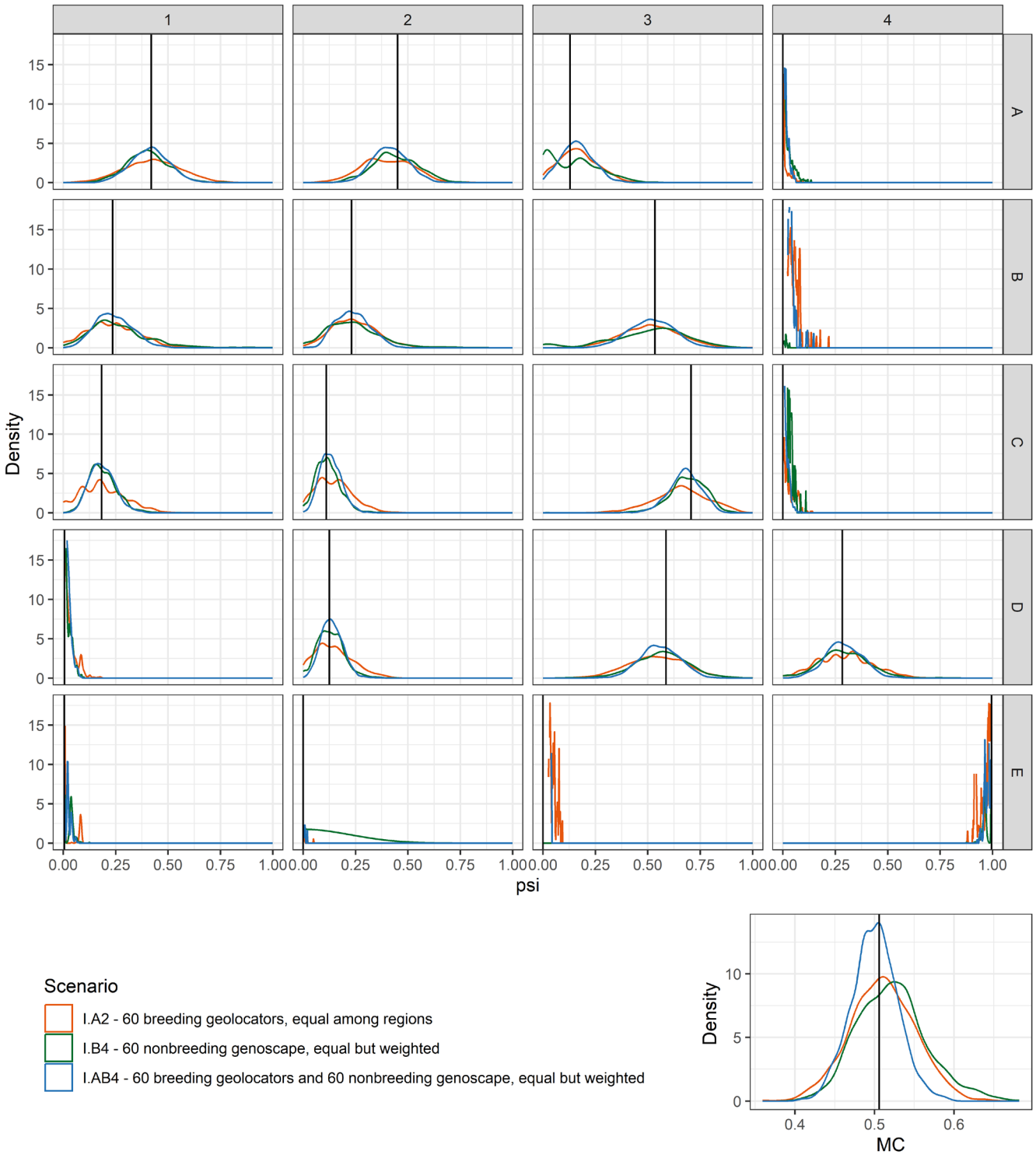


FIGURE 3 Does sampling at breeding sites need to be in proportion to abundance? How well does integration perform if the sample sizes are the same for the data types? Density plots of transition probability (Ψ) and migratory connectivity strength (MC) estimates from simulated data. The row indicates the breeding region (A–E), and the column indicates the nonbreeding region (1–4). The bottom panel is for MC. The black vertical line indicates the simulating values, from example I (biologically informed). The curves indicate the smoothed density plots of 1000 simulations of three sampling scenarios: 60 breeding animals (light-level geolocators); 60 nonbreeding animals (genoscape); and the combination of 60 breeding animals and 60 nonbreeding animals.

3.1.4 | Can we obtain accurate estimates by integrating data types when not all regions are sampled?

When geolocator data were not collected from all breeding regions but genetic data were collected from all nonbreeding regions, estimates were very similar to when both types of data were collected from all regions (Figure S8; Tables S3–S7). However, when data were also only collected from some nonbreeding sites, most ψ estimates were biased (Table S3). The direction of this bias was toward the nonbreeding regions with data and away from those without. Although precision remained good for this scenario (Table S4), Ψ accuracy was also poor (Tables S5–S6). All scenarios had good bias, precision and accuracy levels for their MC estimates, but MC had poor coverage when nonbreeding sampling was incomplete (Figure S8; Table S7). In summary, bidirectional data integration that includes incomplete target sampling (i.e. missing data from some nonbreeding regions) often leads to estimates that are biased and have poor accuracy.

3.2 | Case studies with real data

3.2.1 | Painted bunting

We estimated Ψ for Painted Buntings with geolocator sampling from parts of the breeding range and isoscape and genoscape sampling from across the western nonbreeding range (Figure 4). The estimate of MC was 0.293 (95% CI: 0.228–0.373). However, sampling by each data type was incomplete across the range, and these estimates may have been biased by the lack of data from eastern nonbreeding regions (Southeastern United States and the Caribbean).

3.2.2 | Yellow Warbler

We estimated Ψ for Yellow Warblers with geolocator data from two breeding populations, stable isotope data from one nonbreeding region and genetic data from across their full nonbreeding range (Figure 4).

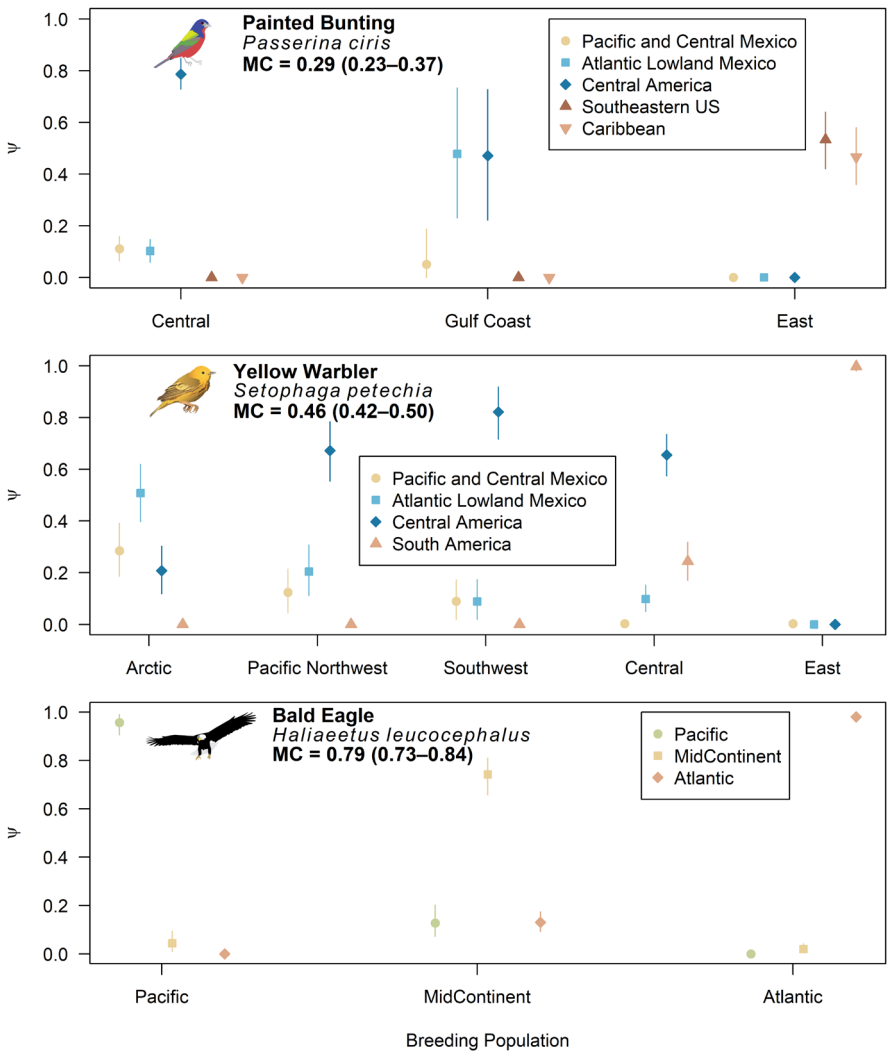


FIGURE 4 Transition probabilities (Ψ) and migratory connectivity strength (MC) results of data integration case studies for three species. Symbols indicate means and lines 95% confidence intervals. On each panel, the *x*-axis indicates the breeding population and the symbols and colours indicate the nonbreeding region of the migratory transition.

The estimate of MC was 0.458 (0.420–0.499). Comparing estimates from the full integrated dataset with estimates missing one or more data types, we found the estimates of Ψ without the geolocator data were virtually identical to those from the full dataset (Figure S13). However, excluding the stable isotope data or both stable isotope and geolocator data reduced the precision and strongly influenced some ψ estimates (lower from the Arctic and higher from the Southwest to Pacific and Central Mexico) (Figure S13). Estimates of Ψ where the nonbreeding data were not weighted by nonbreeding region abundances were quite different, with much higher ψ to Pacific and Central Mexico, among other effects (Figure S13). MC estimates, however, were generally similar for data subsets and without weighting (Figure S13).

3.2.3 | Bald Eagle

We estimated Ψ for Bald Eagles with satellite tracking and banding data. The estimate of MC was 0.79 (0.73–0.84), with most remaining in the same flyway between seasons (Figure 4). The Ψ estimates from banding data alone were similar to those from the integrated data but with even stronger MC (Figure S14). Precision was poor for estimates from the telemetry data alone, with substantial differences between estimators. This difference is likely due to the default flat Dirichlet (1,1,1) prior on ψ_b, in the MCMC analysis. Otherwise, the estimates from the MCMC estimator were similar to those from the bootstrap estimator (Figure S14).

4 | DISCUSSION

We developed methods to integrate migratory movement data types to estimate the pattern and strength of migratory connectivity. Available data types to measure individual movement now include genetic (Ruegg et al., 2014) and banding (Korner-Nievergelt et al., 2010; Roberts et al., 2023) data, which can be integrated with previously developed methods for geolocation, telemetry and stable isotope data (Cohen et al., 2018, 2019). We apply these methods in simulations and real-world case studies to explore key questions about the usefulness of data integration. We show that one can obtain reasonably accurate estimates with either breeding-to-nonbreeding data (e.g. telemetry or geolocation) or nonbreeding-to-breeding data (e.g. intrinsic markers), and that integration of these data can in some cases lead to more accurate estimates. Notably, the range-wide estimate of MC was in some cases more robust to sampling variation than the transition probabilities themselves.

We found data integration for migratory connectivity requires some key considerations. Perhaps not surprisingly, we found adding a small quantity of data to a larger one is likely to have little benefit, although this will also depend on the relative precisions of the two data types. We also found it is possible to use nonbreeding-to-breeding data to estimate breeding-to-nonbreeding transition probabilities, but this requires data from all nonbreeding regions and robust estimates of nonbreeding relative abundance. In terms of sampling across regions, this requirement may become easier to meet as data accumulate for many species and with increased tissue sampling that can be used for isotopic and genetic analyses. The relative abundance data (e.g. from eBird) can be applied to weighting the resampling so that it is not necessary to sample a priori in proportion to abundance. In the absence of reliable relative abundance estimates, there are a couple of options: (1) use only data going in one direction and estimate transition probabilities going that same direction and (2) estimate transition probabilities using bidirectional data without relative abundance weighting and live with potentially biased estimates. Given the large biases from following the latter and the often-limited benefits of data integration seen in simulation, we suspect that the former will generally lead to more accurate estimates. This highlights the value of not only collecting nonbreeding data for migratory species but also following that through with a nonbreeding focus to modelling efforts (in this case, estimating nonbreeding-to-breeding transition probabilities; Marra et al., 2015).

Another key consideration is the spatial extent of the data in each season. For example, considerable data were available for the Painted Bunting, but because the sampling was incomplete across the nonbreeding range (the west was well covered, but the east was not sampled) and sparse across the breeding range we were not able to obtain unbiased estimates. A solution to this would be to add genoscape samples from the Southeastern United States and the Caribbean. For example, the Yellow Warbler case study met these conditions with genetic data from across the full nonbreeding range along with geolocator data from two breeding populations and stable isotope data from one nonbreeding region. In this case, the addition of even the limited stable isotope and geolocator data improved the precision of estimates.

Range-wide analyses of migratory connectivity avoid the issue of the scale dependence (Cresswell & Patchett, 2024; Sharp et al., 2023; Vickers et al., 2021) and, as we demonstrate here, integration of bidirectional data can improve range-wide measures. In terms of sampling, assignment from tissues collected in all regions of the nonbreeding range is likely to be possible at a much lower cost (they can include historical collections) than tagging and tracking (often requiring recapture to recover archival tags) a comparable number of animals from all breeding regions, although it is important to remember the need for breeding range reference samples when calculating cost (Vincent et al., 2022). Individual tracking data are generally higher-resolution and inarguably key to understanding many aspects of migratory behaviour (Scarpignato et al., 2023), but given the strong influence of sample sizes, it may often be worthwhile to complement tracking data with intrinsic markers from across the nonbreeding range for robust range-wide estimates of migratory connectivity. As we demonstrated for the Bald Eagle, some species may have considerable banding reencounter data and integrating these with tracking data can improve estimates from tracking alone which are rarely comprehensive in coverage across the range.

Our analysis does have some limitations and considerations. We did not attempt to fully explore the sampling or species biology space, so all conclusions are limited to what we did test. Our simulations did

not include mark–reencounter data, which differs from the simulated geolocator and genetic data in having detection heterogeneity error instead of location or assignment error. Preliminary simulations and our Bald Eagle case study suggest that transition probabilities estimated from breeding-to-nonbreeding banding data are accurate, but our conclusions about how nonbreeding-to-breeding data can be used may not apply to mark reencounter data. None of our examples included relative abundance uncertainty, which should be included when possible. For example, if one estimates relative abundances of regions oneself using MCMC, one can feed in the full posterior samples to our functions rather than just the point estimates. Cohen et al. (2018) found that abundance variation can be less important for MC than transition probability variation, but there is value to including all relevant sources of uncertainty. In some simulation sampling scenarios, MC estimates were accurate even when some transition probabilities were not. However, this was not always the case, and we caution against assuming that migratory connectivity strength estimates are always accurate. We suggest that those interested in estimating migratory connectivity simulate data using our simulation functions (Table 1) to determine whether the data that the user has available may provide accurate estimates. Finally, the coverage of our estimates was often poor, especially for transition probabilities near the boundary (close to 0 or 1). We suggest being cautious in the interpretation of confidence intervals and standard errors of estimates less than 0.1 or greater than 0.9. Improving coverages of estimates on the boundaries would be a helpful advance for these estimates.

There are ways that the estimation of migratory connectivity and the *MigConnectivity* R package could be extended further. Several of the assumptions of our analyses, such as constant survival across nonbreeding regions and constant tag recapture probabilities across breeding regions, could be relaxed (Rushing et al., 2021). The MCMC estimators have some advantages over bootstrap ones (such as speed), and we could further extend the data types that the former can handle. In addition, recent studies have shown that relative abundances in the two seasons, along with some untested assumptions, could be sufficient for estimating transition probabilities (at least for species that meet those assumptions; Fuentes et al., 2023; Somveille et al., 2021; Vincent et al., 2022). It is probable that whether these relative abundance data were carefully integrated with individual-based migration data that some or all of these assumptions could be relaxed. Future developments for the package could include expanding user-friendly data importation options from sources such as Movebank and the Motus Wildlife Tracking System, a network of automated radio-telemetry stations (Kays et al., 2022; Taylor et al., 2017), and the addition of tracking with barometric geolocators tags (Nussbaumer et al., 2023).

Information about the full annual cycle distributions of populations is necessary to understand and conserve migratory animals (Marra et al., 2015). Our ability to address range-wide questions about the seasonal distributions of migratory species across studies, data types and taxonomic groups has often been limited by sparse and incomplete sampling (Cresswell & Patchett, 2024), but this may be changing as datasets from many diverse data types are increasingly available. Furthermore, the estimation of migratory connectivity requires the identification of meaningful regions and accounting for sampling effort relative to abundance among those regions. The designation of populations or regions is project-specific and can be based on management (Roberts et al., 2023), population trends (Rushing et al., 2016) and genetics (i.e. genoscape, Ruegg et al., 2014), depending on the question. Datasets to account for relative abundance among regions are now available for most avian species (e.g. eBird, Breeding Bird Survey, waterfowl population counts; Fink et al., 2022; Sauer & Link, 2011; U.S. Fish and Wildlife Service, 2022). Given all this, the future of migratory connectivity research clearly includes data integration (Gregory et al., 2023; Korner-Nievergelt et al., 2017; Meehan et al., 2022)—these methods and the *MigConnectivity* R package will facilitate this work.

AUTHOR CONTRIBUTIONS

Jeffrey A. Hostetler, Emily B. Cohen and Michael T. Hallworth conceived the ideas with input from all authors. Jeffrey A. Hostetler designed the methodology. Christen M. Bossu, Amy L. Scarpignato, Kristen Ruegg, Andrea Contina, Clark S. Rushing and Michael T. Hallworth compiled the data for the case studies. Jeffrey A. Hostetler, Michael T. Hallworth, Amy L. Scarpignato and Christen M. Bossu analysed the case study data. Jeffrey A. Hostetler, Michael T. Hallworth and Andrea Contina ran simulations. Emily B. Cohen and Jeffrey A. Hostetler led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank many volunteers, the Burke Museum and Institute for Bird Populations Monitoring Avian Productivity and Survivorship volunteers for the Yellow Warbler and Painted Bunting sample collection, the submitters of Bald Eagle band recoveries to the Bird Banding Lab, Simon Valdez-Juarez for the published Yellow Warbler isotope data, and Martin Witynski and Andrew Sharp for the geolocator data. We thank Jim Hines for his help with high-performance computing, Jaime Ashander for improving the clarity of the proof in Supporting Information 1, and Kim Kraeer, Lucy Van Essen-Fishman, Jane Thomas, Jane Hawkey and the University of Maryland Center for Environmental Science Integration and Application Network (ian.umces.edu/media-library) for the bird symbols (Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) licence). We also thank Nathan Cooper and at least two anonymous reviewers for helpful comments. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

FUNDING INFORMATION

This study received funding from the National Science Foundation 1942313 to K.R. and E.B.C.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14467.

DATA AVAILABILITY STATEMENT

Painted Bunting population definition and genoscape data were published in Rueda-Hernández et al. (2023). Painted Bunting isotope data were published in Contina et al. (2023). Painted Bunting geolocator data have also been published (Contina et al., 2013; Sharp et al., 2023). Yellow Warbler population definition and genoscape data were published in Bay et al. (2021). Yellow Warbler isotope data were published in Valdez-Juárez et al. (2018). Yellow Warbler geolocator data were published in Witynski and Bonter (2018). Bald Eagle telemetry data were published in several different sources (see Supporting Information 3). Bald Eagle capture–mark–reencounter data for the years 1954–2015 were obtained from the North American Bird Banding Program, with data available from Nakash et al. (2023). eBird-based estimates of relative abundance were obtained from the Cornell Laboratory of Ornithology using the ebirdst R package. As a courtesy to users of our R package, most of the Yellow Warbler data is also available for download from our GitHub repository (https://github.com/SMBC-NZP/MigConnectivity/tree/master/data-raw/YEWA). Code for the *MigConnectivity* R package (Hostetler & Hallworth, 2024) is available at https://doi.org/10.5281/zenodo.11075382. Other codes associated with this paper are available at https://doi.org/10.5066/P13YLKFJ.

ORCID

Jeffrey A. Hostetler https://orcid.org/0000-0003-3669-1758
Emily B. Cohen https://orcid.org/0000-0002-4978-4278
Christen M. Bossu https://orcid.org/0000-0002-0458-9305
Amy L. Scarpignato https://orcid.org/0000-0003-0192-2296
Kristen Ruegg https://orcid.org/0000-0001-5579-941X
Andrea Contina https://orcid.org/0000-0002-0484-6711
Clark S. Rushing https://orcid.org/0000-0002-9283-6563
Michael T. Hallworth https://orcid.org/0000-0002-6385-3815

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Anderson, E. C., & Moran, B. (2022). *Package 'rubias'* [computer software].

Bairlein, F. (2016). Migratory birds under threat. *Science*, *354*(6312), 547–548. https://doi.org/10.1126/science.aah6647

Bay, R. A., Karp, D. S., Saracco, J. F., Anderegg, W. R., Frishkoff, L. O., Wiedenfeld, D., Smith, T. B., & Ruegg, K. (2021). Genetic variation reveals individual-level climate tracking across the annual cycle of a migratory bird. *Ecology Letters*, *24*(4), 819–828.

Bolker, B. M. (2008). *Ecological models and data in R*. Princeton University Press. http://books.google.com/books?hl=en&lr=&id=7ZdtfhGHoz8C&oi=fnd&pg=PP2&dq=Bolker&ots=a_sA2pgojE&sig=I5O79IZmlkXtxVNWhBTeEVSGQUI

Briedis, M., & Bauer, S. (2018). Migratory connectivity in the context of differential migration. *Biology Letters*, *14*(12), 20180679.

Cohen, E. B., Hostetler, J. A., Hallworth, M. T., Rushing, C. S., Sillett, T. S., & Marra, P. P. (2018). Quantifying the strength of migratory connectivity. *Methods in Ecology and Evolution*, *9*, 513–524.

Cohen, E. B., Hostetler, J. A., Royle, J. A., & Marra, P. P. (2014). Estimating migratory connectivity of birds when re-encounter probabilities are heterogeneous. *Ecology and Evolution*, *4*(9), 1659–1670.

Cohen, E. B., Rushing, C. R., Moore, F. R., Hallworth, M. T., Hostetler, J. A., Ramirez, M. G., & Marra, P. P. (2019). The strength of migratory connectivity for birds en route to breeding through the Gulf of Mexico. *Ecography*, *42*(4), 658–669.

Commission for Environmental Cooperation. (1997). *Ecological regions of North America: Toward a common perspective* (Secretariat Ed.). The Commission.

Contina, A., Bossu, C. M., Allen, D., Wunder, M. B., & Ruegg, K. C. (2023). Genetic and ecological drivers of molt in a migratory bird. *Scientific Reports*, *13*(1), 814.

Contina, A., Bridge, E. S., Seavy, N. E., Duckles, J. M., & Kelly, J. F. (2013). Using geologgers to investigate bimodal isotope patterns in painted buntings (*Passerina ciris*). *The Auk*, *130*(2), 265–272.

Contina, A., Magozzi, S., Vander Zanden, H. B., Bowen, G. J., & Wunder, M. B. (2022). Optimizing stable isotope sampling design in terrestrial movement ecology research. *Methods in Ecology and Evolution*, *13*(6), 1237–1249.

Cresswell, W., & Patchett, R. (2024). Comparing migratory connectivity across species: The importance of considering the pattern of sampling and the processes that lead to connectivity. *Ibis*, *166*(2), 666–681. https://doi.org/10.1111/ibi.13261

Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M., Ligocki, S., Robinson, O., Hochachka, W. M., Jaromczyk, L., Rodewald, A. D., Wood, C., Davies, I., & Spencer, A. (2022). *eBird status and trends, data version: 2021; released: 2022* (p. 10). Cornell Lab of Ornithology. https://doi.org/10.2173/ebirdst.2021

Fuentes, M., Van Doren, B. M., Fink, D., & Sheldon, D. (2023). BirdFlow: Learning seasonal bird movements from eBird data. *Methods in Ecology and Evolution*, *14*(3), 923–938.

Gregory, K. A., Francesiaz, C., Jiguet, F., & Besnard, A. (2023). A synthesis of recent tools and perspectives in migratory connectivity studies. *Movement Ecology*, *11*(1), 69. https://doi.org/10.1186/s40462-023-00388-z

Griffith, G. E., Omernik, J. M., & Azevedo, S. H. (1998). *Ecological classification of the Western Hemisphere*. US Environmental Protection Agency.

Hagelin, J. C., Hallworth, M. T., Barger, C. P., Johnson, J. A., DuBour, K. A., Pendelton, G. W., DeCicco, L. H., McDuffie, L. A., Matsuoka, S. M., & Snively, M. A. (2021). Revealing migratory path, important stopovers and non-breeding areas of a boreal songbird in steep decline. *Animal Migration*, *8*(1), 168–191.

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(6), 497–526.

Hill, J. M., & Renfrew, R. B. (2019). Migratory patterns and connectivity of two North American grassland bird species. *Ecology and Evolution*, *9*(1), 680–692.

Hostetler, J. A., & Hallworth, M. T. (2024). *MigConnectivity: Estimate migratory connectivity for migratory animals* (version R package version 0.4.7) [computer software]. https://doi.org/10.5281/zenodo.11075382

Huysman, A. E., Cooper, N. W., Smith, J. A., Haig, S. M., Heath, S. A., Johnson, L., Olson, E., Regan, K., Wilson, J. K., & Marra, P. P. (2022).

Strong migratory connectivity indicates willets need subspecies-specific conservation strategies. *Ornithological Applications*, *124*(3), duac015.

Johnston, A., Fink, D., Reynolds, M. D., Hochachka, W. M., Sullivan, B. L., Bruns, N. E., Hallstein, E., Merrifield, M. S., Matsumoto, S., & Kelling, S. (2015). Abundance models improve spatial and temporal prioritization of conservation resources. *Ecological Applications*, *25*(7), 1749–1756.

Kania, W. M., & Busse, P. (1987). An analysis of the recovery distribution based on finding probabilities. *Acta Ornithologica*, *23*(1), 121–128.

Kays, R., Davidson, S. C., Berger, M., Bohrer, G., Fiedler, W., Flack, A., Hirt, J., Hahn, C., Gauggel, D., & Russell, B. (2022). The Movebank system for studying global animal movement and demography. *Methods in Ecology and Evolution*, *13*(2), 419–431.

Kéry, M., & Royle, J. A. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.

Knight, E. C., Harrison, A.-L., Scarpignato, A. L., Van Wilgenburg, S. L., Bayne, E. M., Ng, J. W., Angell, E., Bowman, R., Brigham, R. M., & Drolet, B. (2021). Comprehensive estimation of spatial and temporal migratory connectivity across the annual cycle to direct conservation efforts. *Ecography*, *44*(5), 665–679.

Korner-Nievergelt, F., Prévot, C., Hahn, S., Jenni, L., & Liechti, F. (2017). The integration of mark re-encounter and tracking data to quantify migratory connectivity. *Ecological Modelling*, *344*, 87–94.

Korner-Nievergelt, F., Schaub, M., Thorup, K., Vock, M., & Kania, W. (2010). Estimation of bird distribution based on ring re-encounters: Precision and bias of the division coefficient and its relation to multi-state models. *Bird Study*, *57*(1), 56–68.

Marra, P. P., Cohen, E. B., Harrison, A.-L., Studds, C. E., & Webster, M. (2018). Migratory connectivity. In *Encyclopedia of animal behavior* (2nd ed., pp. 1–12). Elsevier.

Marra, P. P., Cohen, E. B., Loss, S. R., Rutter, J. E., & Tonra, C. M. (2015). A call for full annual cycle research in animal ecology. *Biology Letters*, *11*(8), 20150552.

Meehan, T. D., Saunders, S. P., DeLuca, W. V., Michel, N. L., Grand, J., Deppe, J. L., Jimenez, M. F., Knight, E. J., Seavy, N. E., & Smith, M. A. (2022). Integrating data types to estimate spatial patterns of avian migration across the Western Hemisphere. *Ecological Applications*, *32*(7), e2679.

Nakash, E., Malorodova, M., Howes, L.-A., & Celis-Murillo, A. (2023). *North American bird banding program dataset 1960–2023 retrieved 2023-07-12* [dataset]. U.S. Geological Survey data release. https://doi.org/10.5066/P97LQNHY

Newton, I. (2010). *The migration ecology of birds*. Elsevier.

Nussbaumer, R., Gravey, M., Briedis, M., & Liechti, F. (2023). Global positioning with animal-borne pressure sensors. *Methods in Ecology and Evolution*, *14*(4), 1104–1117. https://doi.org/10.1111/2041-210X.14043

Phipps, W. L., López-López, P., Buechley, E. R., Oppel, S., Álvarez, E., Arkumarev, V., Bekmansurov, R., Berger-Tal, O., Bermejo, A., & Bounas, A. (2019). Spatial and temporal variability in migration of a soaring raptor across three continents. *Frontiers in Ecology and Evolution*, *7*, 323.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March, 20–22.* http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/Plummer.pdf

Procházka, P., Hahn, S., Rolland, S., van der Jeugd, H., Csörgő, T., Jiguet, F., Mokwa, T., Liechti, F., Vangeluwe, D., & Korner-Nievergelt, F. (2017). Delineating large-scale migratory connectivity of reed warblers using integrated multistate models. *Diversity and Distributions*, *23*(1), 27–40.

Rañola, J. M., Novembre, J., & Lange, K. (2014). Fast spatial ancestry via flexible allele frequency surfaces. *Bioinformatics*, *30*(20), 2915–2922.

Roberts, A., Scarpignato, A. L., Huysman, A., Hostetler, J. A., & Cohen, E. B. (2023). Migratory connectivity of north American waterfowl across administrative flyways. *Ecological Applications*, *33*(3), e2788.

Robinson, W. D., Bowlin, M. S., Bisson, I., Shamoun-Baranes, J., Thorup, K., Diehl, R. H., Kunz, T. H., Mabey, S., & Winkler, D. W. (2010). Integrating concepts and technologies to advance the study of bird migration. *Frontiers in Ecology and the Environment*, *8*(7), 354–361.

Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., & Marra, P. P. (2019). Decline of the north American avifauna. *Science*, *366*(6461), 120–124. https://doi.org/10.1126/science.aaw1313

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, *60*(1), 108–115.

Rueda-Hernández, R., Bossu, C. M., Smith, T. B., Contina, A., Canales del Castillo, R., Ruegg, K., & Hernández-Baños, B. E. (2023). Winter connectivity and leapfrog migration in a migratory passerine. *Ecology and Evolution*, *13*(2), e9769.

Ruegg, K. C., Anderson, E. C., Harrigan, R. J., Paxton, K. L., Kelly, J. F., Moore, F., & Smith, T. B. (2017). Genetic assignment with isotopes and habitat suitability (GAIAH), a migratory bird case study. *Methods in Ecology and Evolution*, *8*(10), 1241–1252.

Ruegg, K. C., Anderson, E. C., Paxton, K. L., Apkenas, V., Lao, S., Siegel, R. B., DeSante, D. F., Moore, F., & Smith, T. B. (2014). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology*, *23*(23), 5726–5739.

Rushing, C. S., Ryder, T. B., Scarpignato, A. L., Saracco, J. F., & Marra, P. P. (2016). Using demographic attributes from long-term monitoring data to delineate natural population structure. *Journal of Applied Ecology*, *53*(2), 491–500.

Rushing, C. S., Van Tatenhove, A. M., Sharp, A., Ruiz-Gutierrez, V., Freeman, M. C., Sykes, P. W., Jr., Given, A. M., & Sillett, T. S. (2021). Integrating tracking and resight data enables unbiased inferences about migratory connectivity and winter range survival from archival tags. *The Condor*, *123*(2), duab010.

Salomonsen, F. (1955). *The evolutionary significance of bird-migration*. I komission hos Munksgaard.

Sauer, J. R., & Link, W. A. (2011). Analysis of the north American breeding bird survey using hierarchical models. *The Auk*, *128*(1), 87–98.

Scarpignato, A. L., Huysman, A. E., Jimenez, M. F., Witko, C. J., Harrison, A.-L., Seavy, N. E., Smith, M. A., Deppe, J. L., Wilsey, C. B., & Marra, P. P. (2023). Shortfalls in tracking data available to inform north American migratory bird conservation. *Biological Conservation*, *286*, 110224.

Sharp, A. J., Contina, A., Ruiz-Gutiérrez, V., Sillett, T. S., Bridge, E. S., Besozzi, E. M., Muller, J. A., Kelly, J., Given, A. M., & Rushing, C. S. (2023). The strength of migratory connectivity in painted buntings is spatial scale dependent and shaped by molting behavior. *Journal of Field Ornithology*, *94*(1), 7.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702.

Somveille, M., Bay, R. A., Smith, T. B., Marra, P. P., & Ruegg, K. C. (2021). A general theory of avian migratory connectivity. *Ecology Letters*, *24*(9), 1848–1858.

Su, Y.-S., & Yajima, M. (2021). *R2jags: Using R to Run "JAGS"*. https://CRAN.R-project.org/package=R2jags

Taylor, P., Crewe, T., Mackenzie, S., Lepage, D., Aubry, Y., Crysler, Z., Finney, G., Francis, C., Guglielmo, C., Hamilton, D., Holberton, R., Loring, P., Mitchell, G., Norris, D. R., Paquet, J., Ronconi, R., Smetzer, J., Smith, P., Welch, L., & Woodworth, B. (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

Thorup, K., & Conn, P. B. (2009). *Estimating the seasonal distribution of migrant bird species: Can standard ringing data be used?* (pp. 1107–1117). Modeling demographic processes in marked populations.

Thorup, K., Korner-Nievergelt, F., Cohen, E. B., & Baillie, S. R. (2014). Large-scale spatial analysis of ringing and re-encounter data to infer movement patterns: A review including methodological perspectives. *Methods in Ecology and Evolution*, *5*(12), 1337–1350.

Tonra, C. M., Hallworth, M. T., Boves, T. J., Reese, J., Bulluck, L. P., Johnson, M., Viverette, C., Percy, K., Ames, E. M., & Matthews, A. (2019). Concentration of a widespread breeding population in a few critically important nonbreeding areas: Migratory connectivity in the Prothonotary warbler. *The Condor*, *121*(2), duz019.

U.S. Fish and Wildlife Service. (2022). *Waterfowl population status, 2022*. U.S. Department of Interior. https://www.fws.gov/media/waterfowl-population-status-2022

Valdez-Juárez, S. O., Drake, A., Kardynal, K. J., Hobson, K. A., Krebs, E. A., & Green, D. J. (2018). Use of natural and anthropogenic land cover by wintering yellow warblers: The influence of sex and breeding origin. *The Condor: Ornithological Applications*, *120*(2), 427–438.

Vickers, S. H., Franco, A. M., & Gilroy, J. J. (2021). Sensitivity of migratory connectivity estimates to spatial sampling design. *Movement Ecology*, *9*, 1–12.

Vincent, J. G., Schuster, R., Wilson, S., Fink, D., & Bennett, J. R. (2022). Clustering community science data to infer songbird migratory connectivity in the Western Hemisphere. *Ecosphere*, *13*(4), e4011.

von Rönn, J. A., Grüebler, M. U., Fransson, T., Köppen, U., & Korner-Nievergelt, F. (2020). Integrating stable isotopes, parasite, and ring-reencounter data to quantify migratory connectivity—A case study with barn swallows breeding in Switzerland, Germany, Sweden, and Finland. *Ecology and Evolution*, *10*(4), 2225–2237.

Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S., & Holmes, R. T. (2002). Links between worlds: Unraveling migratory connectivity. *Trends in Ecology & Evolution*, *17*(2), 76–83.

West, J. B., Bowen, G. J., Dawson, T. E., & Tu, K. P. (Eds.). (2009). *Isoscapes: Understanding movement, pattern, and process on earth through isotope mapping*. Springer Netherlands. https://doi.org/10.1007/978-90-481-3354-3

Wilcove, D. S., & Wikelski, M. (2008). Going, going, gone: Is animal migration disappearing. *PLoS Biology*, *6*(7), e188.

Witynski, M. L., & Bonter, D. N. (2018). Crosswise migration by yellow warblers, nearctic-neotropical passerine migrants. *Journal of Field Ornithology*, *89*(1), 37–46.

SUPPORTING INFORMATION

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

Supporting Information 1. Estimating migratory connectivity.

Supporting Information 2. Simulations details.

Supporting Information 3. Additional case study details.

How to cite this article: Hostetler, J. A., Cohen, E. B., Bossu, C. M., Scarpignato, A. L., Ruegg, K., Contina, A., Rushing, C. S., & Hallworth, M. T. (2025). Challenges and opportunities for data integration to improve estimation of migratory connectivity. *Methods in Ecology and Evolution*, *00*, 1–15. https://doi.org/10.1111/2041-210X.14467