Population-level migration modeling of North America's birds through data integration with BirdFlow

Yangkang Chen[1], David L. Slager[2], Ethan Plunkett[3], Miguel Fuentes[3], Yuting Deng[2], Stuart A. Mackenzie[4], Lucas E. Berrigan[4], Daniel Fink[2], Daniel Sheldon[3], Benjamin M. Van Doren[1]* & Adriaan M. Dokter[2]*

[1]Department of Natural Resources and Environmental Sciences, University of Illinois Urbana-Champaign, Champaign, Illinois, USA

[2]Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA

[3]College of Information and Computer Sciences, University of Massachusetts, Amherst, Massachusetts, USA

[4]Birds Canada, Port Rowan, Ontario, CANADA

*co-corresponding authors: amd427@cornell.edu (AD) & vandoren@illinois.edu (BVD)

Abstract

Background Accurate information on population-level movements of migratory animals is essential for understanding migration and for designing effective conservation strategies in a changing world. Yet such information remains scarce for most migratory species due to the effort and expense needed to collect data across their full distribution ranges. BirdFlow is a probabilistic modeling framework that infers population-level movements from weekly species distribution maps produced by the participatory science project eBird. However, BirdFlow

models have only been tuned for a handful of species using high-resolution individual tracking data, which is not available for most migratory species.

Methods Here, we introduce a general tuning and evaluation framework for BirdFlow that enables the first large-scale integration of distributional and individual-level data to infer animal movement across continents and hundreds of migratory species, eliminating reliance on any single individual-tracking data source. By generalizing the BirdFlow model parametrization, we enable tuning and validation using multiple complementary data sources, including GPS tracks, banding recoveries, and radio telemetry data from the Motus Wildlife Tracking System. We investigate the efficacy of this approach by (1) investigating predictive performance compared to null models; (2) validating the biological plausibility of BirdFlow models by comparing movement properties such as route straightness, number of stopovers, and migration speed between model-generated routes and real movement tracks; and (3) comparing the performance of models tuned on species-specific movement data to models tuned using hyperparameters transferred from other species.

Results Our results show that BirdFlow models produced by the new tuning framework achieve biologically realistic performance, even for prediction horizons of thousands of kilometers and several months. When species-specific data are unavailable, models can still be tuned using data from other phylogenetically adjacent species to achieve improved performance.

Conclusions By integrating eBird Status & Trends abundance surfaces with data from banding recaptures, radio telemetry, and GPS tracking, we scale BirdFlow model to 153 North American migratory species, representing the first collection of continental-scale population-level movement and forecasting models. Species-specific tuning improves population-level movement forecasts, while taxonomically informed hyperparameter transfer supports the modeling of data-limited species. Overall, our work offers a foundation for more accurate predictions across hundreds of species for research in ecology and conservation, disease surveillance, aviation, and public outreach.

Keywords: bird migration, probabilistic modeling, multi-source data, movement ecology, spatiotemporal modeling, big data, population ecology, macroecology

Background

Our understanding of continent-scale bird migration has been greatly fostered by modern tracking technologies and the rapid accumulation of participatory science data [1, 2]. Despite these advances, population-level variation in movement behavior remains poorly understood for most bird species, particularly across continental extents and throughout the annual cycle [3]. This gap poses a critical challenge because species' ranges often span diverse ecoregions and habitats, exposing different populations of the same species to distinct environmental barriers and threats (e.g., [4]). Spatial heterogeneity in threats, which include habitat degradation, artificial light at night (ALAN), climate change, and hunting risk, likely drives corresponding variation in bird population trends [5–10]. Therefore, understanding population-level movements within a species' range is essential for quantifying migratory connectivity and for targeting conservation actions in a changing world [11]. However, such information remains largely unavailable due to the difficulties of tracking multiple moving populations across species' range.

Several modeling frameworks have been proposed to infer large-scale avian movement by integrating multiple data types, including methods that combine individual tracking and species distribution information within integrated statistical frameworks [16, 37]. These approaches represent important advances and highlight the growing range of tools available for migration inference at broad spatial scales. However, these approaches remain constrained by the limited

availability of individual-level movement data, which limits their application in regions and for species for which tracking data remain sparse.

BirdFlow is a probabilistic modeling framework developed to address this need by inferring the population-level movements across a species' range from spatiotemporal abundance distributions [12]. The model builds on relative abundance estimates generated by the eBird Status & Trends project (hereafter, eBird S&T) [13, 14] and uses mechanistic optimization that accounts for energy costs and inter-population movement patterns. Performance of the BirdFlow model can be improved when hyperparameters are tuned using individual movement data from other sources, such as GPS and satellite tracking [12]. However, generalizing BirdFlow to hundreds of species has remained challenging because GPS tracking data are unavailable for most migratory birds [15].

To overcome the reliance on any single individual-tracking data source and produce BirdFlow models for many species, we introduce a unified model tuning framework to train BirdFlow models based on general mark-resight data. We also introduce a set of intuitive validation metrics for quantifying the quality and predictive performance of BirdFlow models. Our approach standardizes observations across multiple data sources, including GPS tracking data, banding recovery data, and Motus Wildlife Tracking System automated radio telemetry data (Motus; https://motus.rog; [17]). We address three primary questions: (1) Can integrating multiple data sources enable BirdFlow modeling across hundreds of species? (2) How consistent are optimal model parameterizations among species with similar ecological or morphological traits? and (3) Can such cross-species similarities be leveraged to model species lacking individual-level tracking data? Finally, we discussed how species-specific models tuned using this framework open new opportunities for research and outreach across ecology, evolution, and conservation biology.

Methods

We describe the tuning and validation framework in the sections below. A schematic summary is presented in Fig. 1.

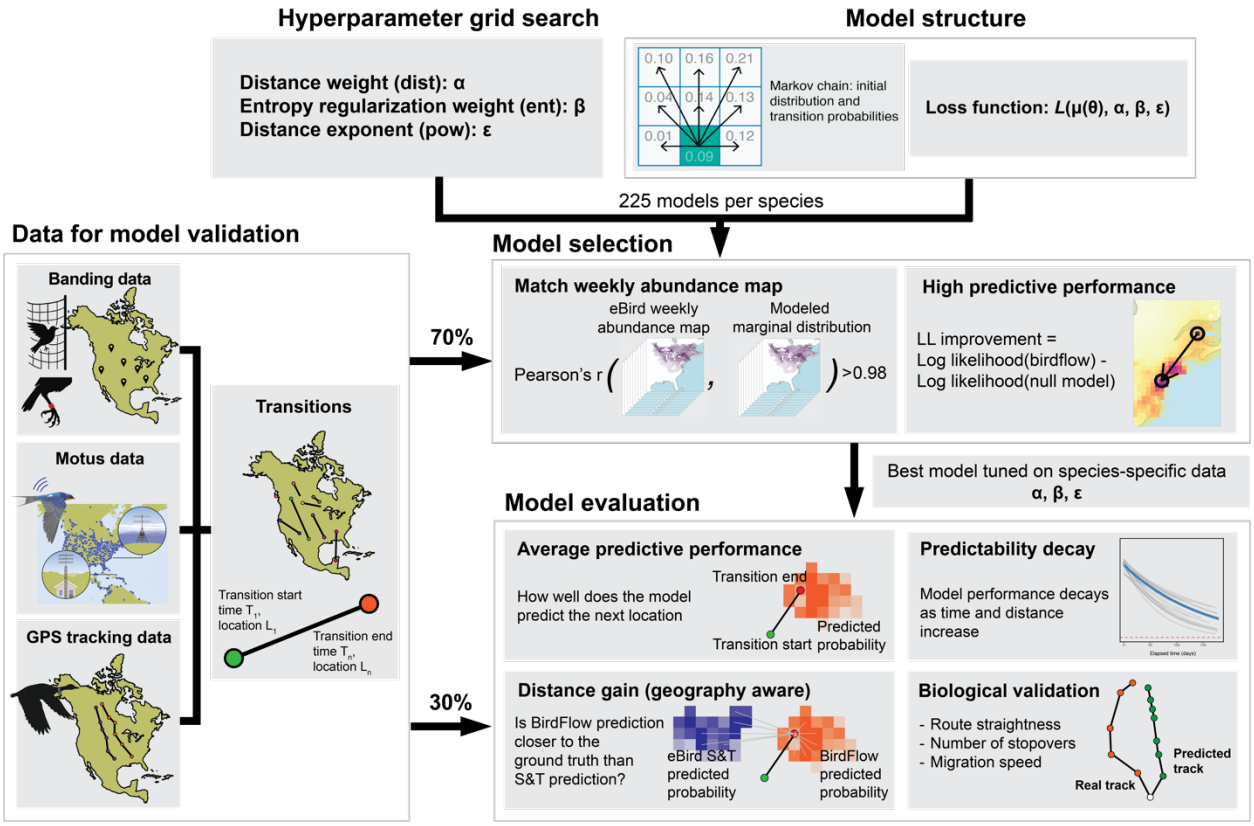

Figure 1. Schematic workflow for the BirdFlow species-specific model tuning and validation framework. The diagram illustrates the main steps of data preparation, model tuning, and validation methods used in this study.

The BirdFlow model

BirdFlow model structure and training methodology are described in detail in [12]. Briefly, the current version of BirdFlow is a Markovian probabilistic graphical model where each state (location and time) is projected to the next state with a certain transition probability. BirdFlow is trained by optimizing an objective function, with three terms controlled by three hyperparameters:

$$L_{obs}(\mu(\theta), \hat{\mu}) + \alpha L_{mov}(\mu(\theta), \epsilon) + \beta J(\mu(\theta)) \quad \text{(Eq. 1)}$$

The first term is the "observation" term, which is the same as the "location loss" term in [12]. The observation term minimizes the differences between the modeled probabilistic surface $\mu(\theta)$, given Markov chain parameter set θ and a species' spatiotemporal relative abundance surface $\hat{\mu}$.

The second term in Eq. 1 is the "distance" term, computed as a function of distance that penalizes longer movements. This requires simulated routes to minimize the movement cost for the population. This term is regulated by hyperparameter α (the distance weight) and hyperparameter ϵ (the distance exponent). The distance term for a single designated transition is formulated as

$$L_{mov}(\mu(\theta), \epsilon)_{x_t, x_{t+1}} = \mu(\theta)_{x_t, x_{t+1}} d(x_t, x_{t+1})^{\epsilon} \,/\, 100^{\epsilon} \quad \text{(Eq. 2)}$$

Where $\mu(\theta)_{x_t, x_{t+1}}$ is the pairwise marginal that specify the probability of moving from cell x_t at time t to cell x_{t+1} at time $t+1$. $d(x_t, x_{t+1})$ denotes the great circle distance between x_t and x_{t+1}, with ϵ regulating the trade-off of cost between a few long-distance jumps versus many

short distance hops. The exponentiated distance d is divided by 100^{ϵ} to make the magnitude of L_{mov} less dependent on ϵ, thereby reducing correlations in the optimized values of α and ϵ.

The third term in Eq. 1 is the "Shannon entropy" term and is regulated by hyperparameter β (the entropy weight), which controls the stochasticity of the routes (Fig. 1).

Species relative abundance surfaces

We acquired eBird Status & Trends weekly relative abundance distributions (data version 2023; hereafter, S&T abundance surfaces) using the ebirdst package version 3.2023.1 [18]. These surfaces are model-based products derived from eBird checklist data. In the present study, we used these S&T abundance surfaces as inputs to the BirdFlow framework but did not explicitly propagate their uncertainty through the tuning procedure, largely because doing so across many species would be computationally prohibitive. We therefore interpret them as model inputs rather than components of a fully joint uncertainty-propagating inferential pipeline. All species are assigned quality scores by experts during the S&T modeling process; we only included species assigned the highest score (3) across all seasons. We preprocessed S&T surfaces to restrict geographic scope to the Western Hemisphere. In addition, to avoid the influence of high-abundance outliers, we truncated abundance distributions by replacing all values above the 0.99 quantile with the 0.99 quantile value. The rationale behind this trimming was to mitigate the influence of imperfect S&T abundance surfaces on model predictions, where excessive counts in aggregation areas (e.g., due to roosting behavior) could bias abundance estimates. We justified the choice of quantile using a sensitivity analysis (Additional file 1: Fig. S1).

Individual mark-resight data (transitions)

We used empirical movement data to select the best hyperparameters from a set of 225 candidate models (see Hyperparameter tuning section) for each species. We acquired tracking data from Movebank (movebank.org) and published literature (Additional file 1: Table S1). We acquired bird banding data from the United States Geological Survey (USGS) Bird Banding Program (BBL) database [19]. Motus data were acquired following the Motus Collaboration Policy (https://motus.org/resources/policy) and preprocessed to reduce errors (Additional file 1: Supplementary methods).

We converted all data into species-specific BirdFlow spatiotemporal coordinates using the BirdFlowR package, an R-language interface that implements and operationalizes the BirdFlow model (version 0.1.0.9075) [20]. We refer to a set of detections for a single individual bird movement within a migratory season as a “track”. After compiling empirical movement data, we sampled mark-resight pairs (segments of movement) from the tracks, which we call “transitions”. We sampled up to 10000 transitions for each species, considering the tradeoff of computational cost and model performance, using an iterative sampling procedure (Additional file 1: Supplementary methods). For species to be included, we required availability of at least 20 transitions (justified using a sensitivity analysis; Additional file 1: Fig. S2) and high-quality abundance maps based on a sensitivity analysis, resulting in 153 migratory species in North America spanning 14 orders and 39 families (Additional file 1: Table S2). Across species, the number of training transitions varied from 29 to 7000 (mean = 1673). Banding data accounted for 69.2% of transitions per species on average, and Motus data made up 27.6%. GPS tracking data were available for only seven species, served as the primary data source (> 50% of transitions) for four, and comprised 11.8% of all transitions (Additional file 1: Fig. S3).

Hyperparameter tuning

For each species, we trained BirdFlow models with different combinations of hyperparameters (α, β, and ϵ introduced in the previous sections) in an exhaustive parameter grid (Additional file 1: Supplementary methods). In total, the training generated 225 models per species. We trained models with a spatial resolution of 150 km and temporal resolution of 1 week. We chose this spatial resolution to balance computation cost while maximizing resolution. The 1-week temporal resolution is inherited from the temporal resolution of the S&T abundance surfaces.

For each species, we used 70% of transitions as the training set for model selection in a hyperparameter grid search. We used two criteria when selecting the best model for a species:

(1) First, we selected models with a high correlation between the modeled probability surface and the eBird S&T relative abundance surface, indicating that the BirdFlow model reproduces an accurate abundance surface. We defined the mean correlation score as the average of the correlation score for pre-breeding migration and post-breeding migration seasons:

$$mean\ correlation\ =\ (\frac{1}{N}\sum_{n=1}^{N} cor_n + \frac{1}{M}\sum_{m=1}^{M} cor_m)/2 \qquad \text{(Eq. 3)}$$

Where cor_n and cor_m represent the Pearson correlation coefficient calculated across grid cells between BirdFlow marginal distributions and S&T abundance surfaces at week n of pre-breeding migration and week m of post-breeding migration, with N and M being the total number of weeks of pre-breeding migration and post-breeding migration. The mean correlation score thus reflects how well the model projection matches observations during migration, with value 1 representing a perfect match. We only considered models with mean correlation score >0.98, which is an empirical value we chose after investigating the models for tens of

species. This threshold ensures any candidate BirdFlow model accurately reproduces a species' abundance surface across areas of both high and low density.

(2) Second, we selected the best model from the remaining set of candidate models based on how well the model predicted the true resighting location of each transition, quantified by the weighted log-likelihood improvement (ΔLL). To calculate this metric, we first introduce the log-likelihood improvement (ΔLL_i) for a single transition i. The ΔLL_i tells us how much more accurate a BirdFlow model prediction is compared to a random sample from the S&T abundance surface for transition i.

$$\Delta LL_i = log\ p_{BF}(t'_i, x'_i \mid t_i, x_i) - log\ A(t'_i, x'_i) \qquad \text{(Eq. 4)}$$

Here, p_{BF} denotes the probability assigned by the BirdFlow model to a transition from a starting location and time to a destination location and time. Specifically, $log\ p_{BF}(t'_i, x'_i \mid t_i, x_i)$ represents the conditional log probability for the BirdFlow model to correctly predict the resighting location x' at time t', given that the bird is recorded at location x at time t, where $[(t_i, x_i), (t'_i, x'_i)]$ defines empirical transition i. In simple terms, it describes how probable the observed data are, given the model. $log\ A(t'_i, x'_i)$ is the normalized relative abundance at (t'_i, x'_i) calculated from the S&T abundance surface, representing the probability of randomly sampling the exact resighting location from the abundance surface.

The weighted average log-likelihood improvement (ΔLL) is calculated as

$$\Delta LL = \frac{\sum_{i=1}^{N} A_i\, \Delta LL_i}{\sum_{i=1}^{N} A_i} \qquad \text{(Eq. 5)}$$

Where N represents the total number of transitions and A_i is a weighting factor equal to the normalized weekly relative abundance of the S&T abundance surfaces at the starting location for transition i. The purpose of the weight is to mitigate potential spatiotemporal bias introduced by rare transitions that are hardly representative of the population.

For each species, we selected the best set of hyperparameters based on the two steps described above. We refer to the models using the selected set of hyperparameters as “species-specifically" tuned models.

Quantifying hyperparameter transferability

We conducted leave-one-out (LOO) analysis to determine the feasibility of transferring hyperparameters from other species in the scenario that tracking data are not available for a target species. In this process, we sought to determine whether taxonomic information could improve the accuracy of hyperparameter transfer. When applying LOO, we omitted transition data of the target species and selected hyperparameters based on the transition data of the remaining species. We only considered models with a mean correlation score >0.98, consistent with species-specific tuning. We then selected the hyperparameter configuration that had the highest average log-likelihood improvement score within the same taxonomic group as the target species. We compared four model tuning methods, including taxonomically informed (1) family-level LOO (Family-LOO), (2) order-level LOO (Order-LOO), and (3) LOO with all the other species except for the target species (All-LOO). We consider All-LOO as a baseline method since it incorporates no taxonomic information. We also included (4) species-specifically tuned models (no LOO) for comparison.

Model validation

After tuning, we evaluated tuned models using the remaining 30% of the transition data (test set). We again used log-likelihood improvement to evaluate tuned models as described in the hyperparameter tuning section. However, while log likelihood reflects the quality of prediction, it only validates the model at the exact resighting locations and therefore has limited biologically interpretability. Ideally, an evaluation of a spatiotemporal movement model should also consider how close the predictions are to the resighting locations. To address this need, we introduce a spatially explicit metric called "distance gain" (ΔD) and a variant, relative distance gain (ΔD^{rel}), as the primary validation metric. It measures how close the BirdFlow model prediction is to the empirical resighting location and compares this distance to a random sample of distances drawn from the abundance surface:

$$\Delta D = \frac{\sum_{i=1}^{N} A_i \, (D_{ST,i} - D_{BF,i})}{\sum_{i=1}^{N} A_i} \qquad \text{(Eq. 6)}$$

Where $D_{BF,i}$ is the weighted mean distance from the BirdFlow predicted locations to the empirical resighting location for transition i:

$$D_{BF,i} = \sum_{j=1}^{K} p(X_{t'} = j \mid X_t = x_i) \, d(j, x'_i) \qquad \text{(Eq. 7)}$$

Where j is a location index running over all spatial pixels K. $d(j, x'_i)$ represents the great circle distance between location j and the resighting location x'_i of the transition i. The weight for location j is the BirdFlow predicted probability (p) at location j time t' given the starting location x_i at time t.

Similarly, $D_{ST,i}$ is defined as the weighted mean distance from the normalized abundance surface (S&T) to the empirical resighting location for transition i, representing a "null distance" of drawing a random sample from all potential locations.

To facilitate multi-species comparison, we developed the relative distance gain (ΔD^{rel}) to account for the variation of distance gain among species with different migration distances:

$$\Delta D^{rel} = \frac{\sum_{i=1}^{N} A_i \left(D_{ST,i} - D_{BF,i}\right) / D_{ST,i}}{\sum_{i=1}^{N} A_i} \qquad \text{(Eq. 8)}$$

With value 1 representing a perfect prediction of the resighting location, value 0 representing predictive performance equivalent to a random sample from the S&T abundance surfaces, and negative values representing worse performance than a random sample.

Model decay analysis

Next, we quantified the pattern of decay in model performance across time and distance. We fitted exponential regression functions using the test set transitions and relative distance gain metric. We chose exponential functions because the visual patterns generally match well with exponential decay as transition interval increases, a pattern also seen in [12] (Additional file 1: Supplementary methods). The regression line is asymptotic to y=0, so we define the maximum functional distance horizon as the distance value where the regression line intersects a threshold of 0.05. Thus, the maximum functional distance represents the distance at which BirdFlow model predictions are <5% closer to true resighting locations than a random sample from the S&T distribution of the resighting week (Additional file 1: Supplementary methods, Fig. S4). We do not define a maximum functional time horizon since the models often performed well

305 even beyond the temporal threshold (see results). We also evaluated performance decay using
306 log-likelihood improvement metrics and validated prediction horizons using the same approach.
307

308 Empirical validation of biological metrics

309 To assess the biological plausibility of BirdFlow-generated trajectories, we compared the route
310 statistics of model-simulated and empirical routes. For all seven species for which GPS tracking
311 data were available at the time of analysis, we extracted track sections during pre-breeding and
312 post-breeding migration phases. Only tracks recorded for more than two weeks and with ≥1
313 observation per week were included to match the temporal resolution of BirdFlow models. For
314 each track, we sampled 100 trajectory predictions conditional on the starting time and location
315 and the ending time of the track. We then calculated three biological metrics frequently used in
316 migration research [21] for both simulated and observed routes: migration speed, route
317 straightness, and number of stopovers. We defined migration speed as the total distance
318 traveled by the simulated individual within the trajectory divided by the total length of the time.
319 Route straightness was the beeline distance of the routes divided by the total distance traveled.
320 We defined the number of stopovers as the number of 150-km resolution grid cells in which the
321 individual stayed >7 days; this temporal resolution is predefined and aligned with the eBird S&T
322 product. The route features were calculated using trajr package version 1.5.1 [22] in R version
323 4.5.

324 Modeling hyperparameters using species traits

325 We modeled the selected hyperparameters for each species using species-specific
326 morphological traits and distribution range statistics to understand hyperparameter variation
327 across species. We acquired species morphological traits from the AVONET dataset [23]. We

further calculated range-wide summaries as explanatory variables, including range size, spatial variation of abundance, and the minimum, maximum, and centroid of longitude and latitude for different seasons (Additional file 1: Supplementary methods). Some of these explanatory variables were highly correlated (Pearson's r >0.8); for a group of correlated variables, we only included one representative variable in the model (Additional file 1: Supplementary methods, Fig. S5). To account for potential overfitting, we tuned Random Forest regression models for each hyperparameter separately using 5-fold cross-validation on the species-level dataset and calculated the coefficient of determination (R^2) based on pooled predictions on out-of-fold datasets. We calculated impurity-based feature importance metrics for the best model and linearly regressed BirdFlow hyperparameters against the variable with the highest feature importance in the Random Forest model to understand how a single trait is associated with the hyperparameter values. The feature calculations are performed in R version 4.5 and the Random Forest model is fitted using the scikit-learn package version 1.6.1 [24] in Python version 3.11.11.

Results

Multi-source transition data improves model performance

BirdFlow models tuned using species-specific mark-resight data consistently predicted resight locations better than a random sample from the S&T abundance surface corresponding to the week of resighting (the null model) for all species (Fig. 2). This was reflected in both relative distance gain and log-likelihood (LL) improvement metrics. Averaged across species and transitions, tuned models were 27 times better in predicting the exact resighting location (average log-likelihood improvement 3.30, range: 0.27–7.13 across species; 95% percentile

interval 0.67–5.17) and 1058 km closer to the empirical resighting locations compared to the null model (range: 55–6793 km; 95% percentile interval 225–2352 km).

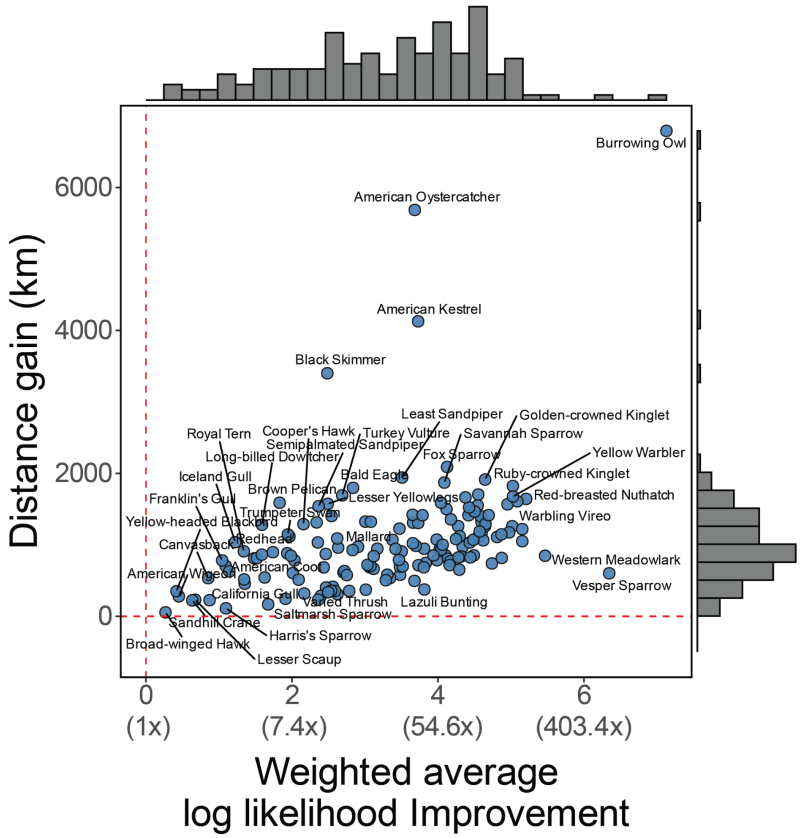

Figure 2. BirdFlow models for all 153 species perform better compared to a random sample from eBird Status & Trends (indicated by the dotted red line). Performance is evaluated by both weighted average log-likelihood improvement (x axis) and distance gain (y axis). Metrics are calculated across all hold-out transitions with different forecasting horizons. The log-likelihood improvement is additionally labeled using straightforward explanations (e.g., 7.4x means on average 7.4 times better in predicting the actual recovery location of the transitions compared to a random sample).

BirdFlow's predictive performance decreased with increasing forecasting distance and time, but even at long temporal and spatial horizons, performance was better than a random sample from the S&T abundance surface (Fig. 3, a-b). Across time, most BirdFlow models continued to achieve better performance compared to S&T after 270 days (Fig. 3a) based on the relative distance gain metric. Across space, the averaged maximum functional distance beyond which the BirdFlow model was no longer favored over the null S&T abundance surfaces was 2449 km (95% percentile interval: 242–6395 km; Fig. 3 b-c). Similar results for the log-likelihood improvement metric can be found in Additional file 1: Fig. S6.

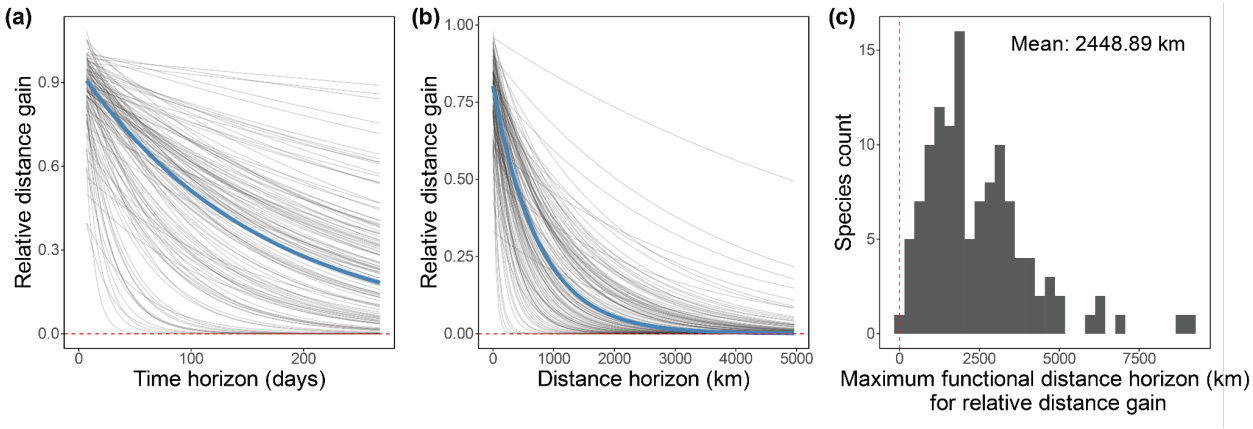

Figure 3. Model performance decays as time and distance horizon increase. Model performance decay is measured as the decrease of relative distance gain as the time horizon (a) and distance horizon (b) extends. Gray lines represent the fitted exponential regression function for each species, and blue lines represent the functions using cross-species-averaged parameters. (c) measures the maximum distance horizon that BirdFlow models still have advantages against a random sample from the abundance surface, defined as the horizon where the exponential decay curve (asymptotic to y=0) in (b) intersects with the y=0.05 line (5% closer to true resighting locations than S&T). Red dashed lines in (c) indicate zero distance horizon. See Additional file 1: Fig. S4 for schematic diagram of calculation in (c).

By comparing simulated trajectory statistics with empirical observations, we show that the biological characteristics of BirdFlow-generated routes are generally consistent with data from tracked individual birds (Fig. 4, Additional file 1: Fig. S7, Table S3). A biologically plausible model should expect around 95% of the empirically observed biological metrics to occur within the 95% simulated intervals and around 50% to occur within the 50% simulated intervals. For migration speed, on average 86% of the empirical observations fell within the 95% simulated intervals (range: 56%–100% across species), and 44% fell within the 50% simulated intervals (range: 22–58%) (Fig. 4). For route straightness, the averages are 90% (63%–100%) and 49% (11%–64%), respectively (Additional file 1: Fig. S7). For the number of stopovers, these values are 99.59% (97.14%–100%) and 98.44% (94.29%–100%), respectively (Additional file 1: Fig. S7). The results for the number of stopovers are less informative because the coarse weekly temporal resolution reduces the precision of stopover calculations.

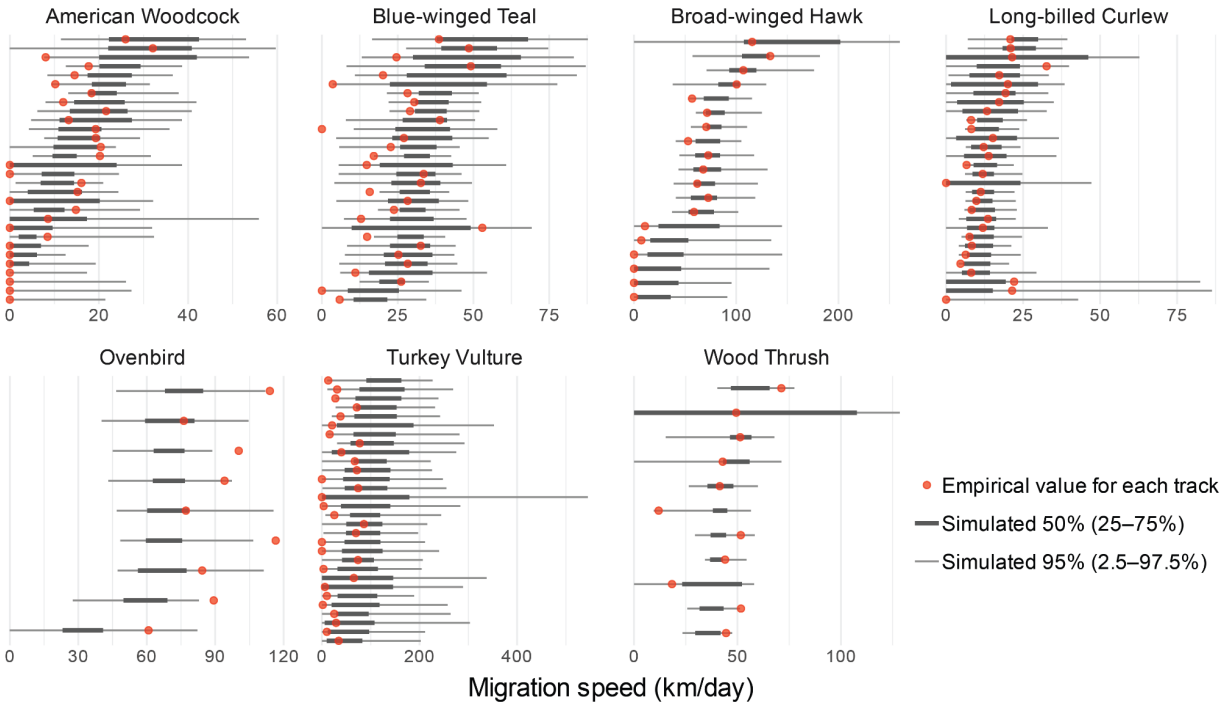

Figure 4. BirdFlow models simulate biologically plausible migration speeds. For each species, we randomly sampled at most 30 tracks from pre-breeding or post-breeding migration seasons. Red points indicate the empirical migration speeds calculated from GPS tracks and gray bars indicate 95% and 50% percentile intervals of the simulated migration speed. A quantitative summary of the figure can be found in Additional file 1: Table S3.

Hyperparameter transfer among related species enhances model performance

The taxonomically informed model tuning methods showed better performance than the All-LOO method, both measured by relative distance gain and log-likelihood (Table 1). Models tuned using information from more closely related species performed better relative to the baseline All-LOO: Species-specific tuning performed the best (89.5% species showed log-likelihood improvement; 82.9% species show positive distance gain improvement), followed by Family-LOO (37.3% species showed log-likelihood improvement; 45.5% species show positive distance gain improvement) and Order-LOO (20.4% species showed log-likelihood improvement; 18.4% species show positive distance gain improvement).

Table 1. Taxonomically informed model tuning shows improvements over the all-leave-one-out method (ALL-LOO)

Validation metric	Model tuning method	Metric improvement			
		Mean	95% percentile interval	Range	# of species performing better

Log likelihood	All-LOO (baseline)	0	–	–	–
	Order-LOO	0.017	-0.34–0.41	-1.25–1.23	30 (20.4%)
	Family-LOO	0.006	-0.64–0.44	-1.35–1.13	50 (37.3%)
	Species-specific tuning	0.248	-0.01–0.77	-0.08–5.28	136 (89.5%)
Relative distance gain	All-LOO (baseline)	0	–	–	–
	Order-LOO	0.007	-0.04–0.09	-0.14–0.27	27 (18.4%)
	Family-LOO	0.017	-0.05–0.13	-0.28–0.27	61 (45.5%)
	Species-specific tuning	0.071	-0.05–0.25	-0.23–0.32	126 (82.9%)

We further investigated the influence of transition data abundance on model tuning. The result shows that a small amount of training data during model tuning is sufficient for achieving a performance improvement compared to All-LOO baseline, and that more transition data for model tuning is not correlated with higher performance improvement across species (Additional file 1: Figure S8; Pearson's r: 0.0012, Spearman's r: 0.0942).

Species traits explain variation in hyperparameters

While we did not observe statistically meaningful clustering patterns of hyperparameters by taxonomic group (Fig. 5), the regression results show that species' morphological traits and distribution range summaries are predictive of their tuned hyperparameters (Additional file 1: Fig. S9a-c). A Random Forest regression model trained by 5-fold cross validation had moderate predictive power for the tuned values of hyperparameter β (R^2 = 0.27), ϵ (R^2 = 0.20), and α (R^2

= 0.27) across species. Body mass was the most important predictor of entropy weight and distance weight, with higher body mass correlated with higher entropy weights (Pearson's r = 55; Additional file 1: Fig. S9d) and lower distance weights (Pearson's r = -0.33; Additional file 1: Fig. S9f). Higher spatial variation of abundance and narrower range at breeding season were also correlated with higher distance weights (Pearson's r = 0.45 and -0.47, respectively; Additional file 1: Fig. S9e, Fig. S5).

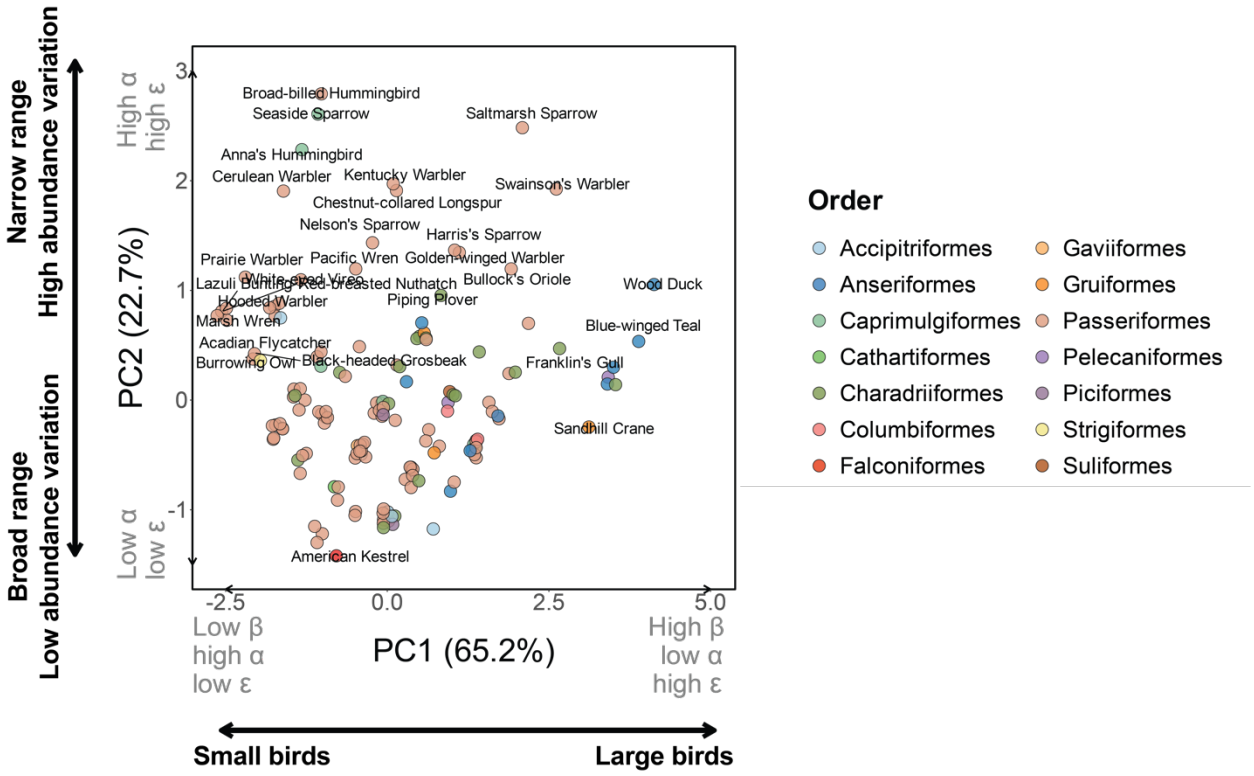

Figure 5. Associations between selected hyperparameters and species traits. Principal component analysis results for the hyperparameter patterns across species. Color denotes the order of the species. Gray text indicates the relative PC loadings of entropy weight β, distance weight α, and distance exponent ϵ, respectively. Higher values in PC1 are correlated with higher entropy weight, lower distance weight, and higher distance exponent. Higher values in PC2 are correlated with higher distance weight and higher distance exponent. Small jitters in x and y

axes were added for better visualization, since the best models for certain species share the same hyperparameter combination.

Discussion

Our results show that BirdFlow models tuned with individual-based movement data can simulate routes that closely match actual individual tracks. By integrating the movements of a small sample of tracked individuals with eBird distribution data, we inferred range-wide population-level movement patterns for 153 species. These models provide a first quantitative yardstick for delineating population-specific migration routes, even when individuals are tracked or banded in only part of that range. Specifically, BirdFlow predicted movement transitions from week *i* to week *j* outperformed a null baseline of randomly sampling from a species' abundance distributions corresponding to the resighting week. This result holds for >100 species (Fig. 2) across forecast horizons of a few thousand kilometers (on average 2449 km) and over 270 days (Fig. 3).

By integrating individual-based movement data from multiple sources into the hyperparameter tuning of BirdFlow, we greatly expand the number of species for which species-specific models can be tuned. Banding data alone enable tuning for 88.2% of the study species (defined as at least 20 transitions per species), but individual-level movement data remain scarce for many small-bodied passerines. Because of their low body mass, these species are difficult to track with current GPS technology [15, 25], and band-recovery data accumulate slowly due to low recovery rates [26]. In contrast, the Motus Wildlife Tracking System supports tuning for 49.7% of the study species, a smaller but rapidly growing fraction that is particularly valuable for small passerines [17]. As a result of their complementarity, combining banding and Motus data makes it possible to tune models for 98.7% of the studied species.

Our results demonstrate that improvements in model performance can be achieved with limited mark-resight data. Notably, simply increasing the number of individual transitions does not consistently lead to higher predictive performance across species (Additional file 1: Fig. S8). This pattern is encouraging because it indicates that a small amount of empirical data can meaningfully calibrate and improve BirdFlow predictions for a focal species. This result also suggests it may be possible to increase the flexibility (i.e. number of hyperparameters) of the current BirdFlow model structure to learn more detailed patterns from the individual data, which could lead to increasing performance with higher amounts of tuning data.

To our knowledge, this study provides the first large-scale integration of distribution and individual-level data to infer animal movement across continents and among hundreds of species. As part of this study, we release the batch of BirdFlow models for 153 North American migratory species (see data statement), representing a comprehensive collection of population-level movement models spanning the full spatial and temporal ranges of these species. Previous research on bird migration has developed quantitative and mechanistic models of movement and connectivity, but such approaches have typically been applied at the species level or to a limited number of intensively tracked populations. In contrast, BirdFlow resolves range-wide migration patterns at the population level across hundreds of species, addressing a key scaling gap given growing evidence that population-level variation in migratory strategies and population trends can be comparable to, or exceed, differences among species [10, 27, 28].

As expected, we see a decay of model performance along spatial and temporal forecasting horizons. A main explanation for this performance decay is the biological reality of populations mixing across space and time. For example, for species with low migratory connectivity, we expect the spatial connection patterns among populations to dissipate within a certain horizon

[29]. For these species, BirdFlow models are not expected to be able to distinguish the movement of individuals in the same spatiotemporal unit. Alternatively, the decayed model performance could stem from the accumulation of prediction errors. Disentangling these two different factors requires further investigation into the migration connectivity estimation of the modeled species.

While most species' models recovered the population-level characteristics of tracked individuals, some models exhibited consistent deviations. For example, the Ovenbird model consistently predicted lower migration speed and higher route straightness relative to empirical data, while the Turkey Vulture model consistently predicted higher migration speed and lower route straightness (Fig. 4, Additional file 1: Fig. S7). The biases indicate potential space to improve the parameterization and/or structure for those models.

Collectively, our results indicate that the three hyperparameters reflect species' functional traits and distributional properties. For example, small passerine birds generally showed low entropy weight, high distance weight, and low distance exponent (Fig. 5; lower PC1 values; less spread and higher movement cost), while large birds in Anseriformes (ducks and geese) showed the opposite (higher PC1 values; more spread and lower movement cost). This result aligns with our understanding that small songbirds often depend on frequent flights and refueling, whereas large waterfowl accumulate substantial reserves and undertake long-distance flights [30–32]. We also find that species with narrow ranges (low spatial coverage) and high variation in spatial abundance show higher distance weight and higher distance power weight (higher PC2 values), resulting in a migration pattern with higher movement cost and favoring long jumps over small hops. On the other hand, species with broad ranges and low abundance variation show the opposite. This may occur because narrow-ranged species, being restricted to small geographic areas or specialized habitats, may face stronger migration barriers and have fewer suitable

stopover habitats to choose from, which elevates the distance cost and forces long jumps rather than small hops (e.g., [33]).

The evidence of hyperparameter similarity among taxonomic groups opens the door for tuning species lacking any individual-level movement data. Hyperparameters could be transferred from close taxonomic relatives using a taxonomically informed leave-one-out approach, providing a practical fallback when species-specific movement data are unavailable. We show that BirdFlow models using hyperparameters transferred from phylogenetically adjacent species groups performed better than models using parameters transferred from distant or unrelated groups. In BirdFlow models, the three hyperparameters β, α and ϵ encode different aspects of the migration pattern and strategies: the diversity of the migration routes, the cost of energy expenditure, and the trade-off between long-distance jumps versus short-distance hops. These diverse migration patterns and strategies are driven by species-specific trade-offs that ideally maximize the survival and reproduction of the species [34, 35]. Species that are closer in taxonomic relationship or functional traits might be more similar in migration patterns and strategies due to ancestral traits inheritance or convergent evolution [36] and consequently may also be closer in BirdFlow hyperparameter space. Despite this evidence, we did not find statistically significant clustering patterns of hyperparameters across orders or families, which indicates that within-group hyperparameter variation is comparable to among-group variation, further emphasizing that species-specific tuning is still important when available.

Additionally, because the validation framework is insensitive to both model structure and data type, it opens opportunities for additional refinement of the BirdFlow modeling framework, from adjusting hyperparameters to exploring alternative model formulations. For example, we acknowledge that log-likelihood may not be the optimal metric for hyperparameter tuning in all cases. Future work could therefore explore additional tuning criteria, such as distance gain or

connectivity-based metrics. This tuning framework also provides a way to quantitatively evaluate different structures as more data become available, ensuring that results grow increasingly robust over time.

The probabilistic Markovian structure of BirdFlow models does have limitations in estimating the routes of individuals, in particular across long time scales and long distances. BirdFlow intrinsically models the synthetic movements of all individuals presented within a spatiotemporal unit, making it a model of space-time tracks along which the population's biomass redistribution occurs. The Markovian nature of the current BirdFlow model algorithm precludes "memory" for individual movement histories. Accordingly, BirdFlow-simulated tracks are not expected to resolve all individual-level variation in the observed tracks. Rather, they represent the average movement patterns across all individuals at a given location and time.

While the generalization of movement patterns to species full populations and ranges is one of the key contributions of BirdFlow model output, we should be aware that the individual transition data on which models are tuned still have spatiotemporal biases. Tracking data often concentrate in a few locations rather than spanning the full species distribution [15], and banding or Motus records are largely restricted to sites where stations are established. Moreover, relatively few transition data extend into Central and South America, which may bias the model to prioritize movement tuning for North America. While overfitting is unlikely with the current three-hyperparameter model, and models are heavily constrained by the fine-scale eBird distributional data, such biases warrant attention. If this tuning framework were applied to a more complex model structure, bias-control methods or balanced sampling strategies could be applied to quantify such potential bias.

Another limitation of this work, and of BirdFlow more generally, is that it does not fully propagate uncertainty in eBird S&T relative abundance surfaces and tracking observation inputs (e.g. cf. Buderman et al. [37]). Rather than jointly estimating movement processes and observation models from multiple data streams within a single hierarchical model, BirdFlow uses abundance surfaces as a ground truth, while additional data sources such as tracking and mark-resight data are used primarily to inform model tuning and evaluation. As a result, uncertainty is not explicitly propagated throughout the full inference process. In this study, we did not incorporate uncertainty in the eBird S&T abundance estimates because doing so at this scale would have been computationally prohibitive. Although the criterion requiring mean correlation score > 0.98 may provide some tolerance to uncertainty and help reduce overfitting during model tuning, it does not substitute for formal uncertainty propagation. Future extensions could more formally account not only for uncertainty in the S&T abundance surfaces, but also for observation-process biases, such as spatiotemporal bias in banding and Motus data. Such approaches may provide stronger uncertainty quantification and a clearer representation of heterogeneity among individuals or populations.

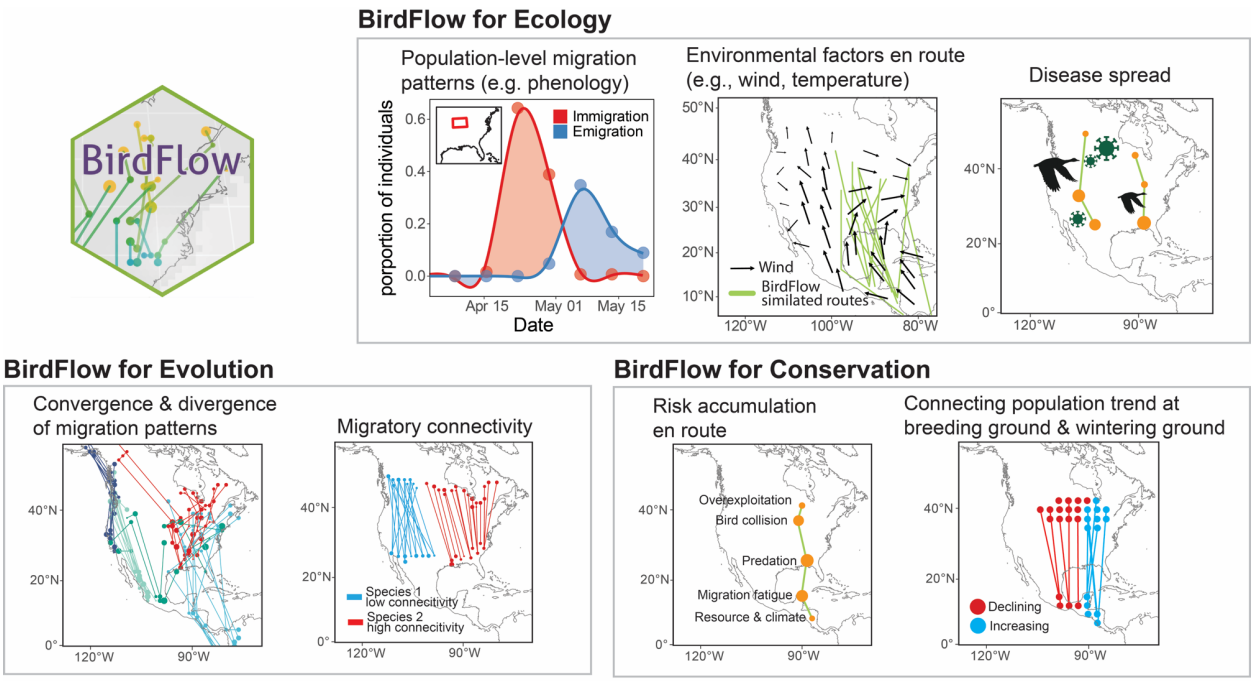

Figure 6. Applications of using BirdFlow for ecological, evolutionary, and conservation research. In ecological research, BirdFlow enables (i) quantification of population-level migration dynamics, such as regional influxes and effluxes, (ii) integration of migration patterns with environmental conditions to assess environmental drivers and barriers of movement, and (iii) linkage with surveillance data to study the spread, control, and prevention of wildlife-borne diseases. In evolutionary research, BirdFlow-simulated trajectories support analyses of (i) convergence and divergence in migration routes within and among species and (ii) variation in migratory connectivity, providing insights into the mechanisms shaping differences in migration strategies. In conservation applications, BirdFlow models facilitate (i) assessment of population-specific anthropogenic and environmental risks accumulated along migration routes and (ii) identification of the seasons and locations most influential for population trends by connecting breeding and wintering grounds.

As illustrated in Fig. 6, species-specifically tuned BirdFlow models provide an approach for addressing a broad range of ecological, evolutionary, and conservation questions. These models provide a population-resolved, continent-scale representation of migratory movement across the annual cycle, making it possible to link movement dynamics to environmental context and to quantify population differences without collapsing them into a single species-level pattern. This population-level resolution also allows direct, comparable tests of how migration strategies diverge or converge across populations and taxa. Finally, the same framework supports conservation-relevant attribution by connecting seasonal ranges, enabling researchers to evaluate which times of year and which regions are most strongly associated with among-population differences in trends and therefore most informative for explanation, prediction, and targeted intervention (Fig. 6). These questions often require range-wide coverage as provided by BirdFlow and are difficult to address with distributional data or sparse individual-tracking data alone. Together, these advances position BirdFlow as a general framework for integrating multi-source data to estimate range-wide bird movements at continental scales, opening new opportunities to link population ecology with applied management across hundreds of species.

Conclusions

By integrating eBird Status & Trends abundance surfaces with multi-source mark–resight observations, including banding recoveries, Motus telemetry, and GPS tracking, we parametrized BirdFlow models for 153 North American migratory bird species. This continent-wide inference of population movements over the full annual cycle provides a quantitative basis for delineating population-specific routes across species' full ranges, complementing existing individual-based tracking efforts and helping fill a key gap in ecological research and applied conservation. Tuned models consistently outperform abundance-based null expectations,

remain informative across forecasting horizons of thousands of kilometers and many months, and generate biologically realistic migration routes that align with individual movement data. When species-specific movement data are unavailable, taxonomically informed hyperparameter transfer enables improved model, demonstrating that migration strategies share partial predictability across related taxa. Together, these advances establish BirdFlow as a general, extensible framework for estimating the comprehensive migratory movements of entire populations for hundreds of species, offering a foundation for more accurate predictions for ecological understanding, conservation planning, disease surveillance, aviation risk assessment, and public outreach as new data continue to accumulate.

List of abbreviations

LL: log-likelihood

ΔLL: weighted log-likelihood

eBird S&T: eBird Status & Trends relative abundance surfaces

LOO: leave-one-out

BF: BirdFlow

p_{BF}: the probability assigned by the BirdFlow model to a possible destination in space and time, conditional on a specified starting location and time.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Code is available at https://github.com/birdflow-science/BirdFlowTuning. Data and the species-specifically tuned models are available at https://doi.org/10.5281/zenodo.17545007. Models can be loaded using the BirdFlowR package (https://github.com/birdflow-science/BirdFlowR). Tuned models that passed manual quality control will be released along with the BirdFlowR package in the future.

Competing interests

The authors declare that they have no competing interests

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 2210979 and 2210980, and Illinois Distinguished Fellowship (to YC). Computation was supported by the Unity computing platform at the Massachusetts Green High Performance Computing Center.

Authors' contributions

Conceptualization: DS, YC, DLS, BVD, AMD, EP, MF; Data curation: LEB, SAM, DLS, BVD, YC, EP, YD, AMD; Formal analysis and investigation: YC, DLS, EP, YD; Methodology: DS, YC, DLS, EP, BVD, AMD, MF, DF; Software: DLS, EP, YC; Validation: YC, DLS, EP, YD. Visualization: YC. Project administration and funding: DS, AMD, BVD; DF; Supervision: BVD, AMD, DS; Writing: All authors.

Acknowledgements

We thank the USGS bird banding lab for the bird banding data collection and Birds Canada for preparing and sharing the data of the Motus Wildlife Tracking System and the individual contributors to these projects. We are also grateful to the many thousands of participants, reviewers, and partner organizations around the world who support and contribute to eBird.

References

1. Sullivan BL, Aycrigg JL, Barry JH, Bonney R, Bruns N, Cooper CB, et al. The eBird enterprise: an integrated approach to development and application of citizen science. Biol Conserv. 2014;169:31–40.
2. Flack A, Aikens EO, Kölzsch A, Nourani E, Snell KRS, Fiedler W, et al. New frontiers in bird migration research. Curr Biol. 2022;32(20):R1187–R99.
3. Shaw AK. Causes and consequences of individual variation in animal movement. Mov Ecol. 2020;8(1):12.
4. Kramer GR, Andersen DE, Buehler DA, Wood PB, Peterson SM, Lehman JA, et al. Population trends in *Vermivora* warblers are linked to strong migratory connectivity. Proc Natl Acad Sci USA. 2018;115(14):E3192–E200.
5. Emmenegger T, Hahn S, Arlettaz R, Amrhein V, Zehtindjiev P, Bauer S. Shifts in vegetation phenology along flyways entail varying risks of mistiming in a migratory songbird. Ecosphere. 2016;7(6):e01385.
6. Zurell D, Graham CH, Gallien L, Thuiller W, Zimmermann NE. Long-distance migratory birds threatened by multiple independent risks from global change. Nat Clim Chang. 2018;8(11):992–6.
7. Rosenberg KV, Dokter AM, Blancher PJ, Sauer JR, Smith AC, Smith PA, et al. Decline of the North American avifauna. Science. 2019;366(6461):120–4.

8. Buchan C, Franco AMA, Catry I, Gamero A, Klvaňová A, Gilroy JJ, et al. Spatially explicit risk mapping reveals direct anthropogenic impacts on migratory birds. Glob Ecol Biogeogr. 2022;31(9):1707–25.
9. Horton KG, Buler JJ, Anderson SJ, Burt CS, Collins AC, Dokter AM, et al. Artificial light at night is a top predictor of bird migration stopover density. Nat Commun. 2023;14(1):7446.
10. Johnston A, Rodewald AD, Strimas-Mackey M, Auer T, Hochachka WM, Stillman AN, et al. North American bird declines are greatest where species are most abundant. Science. 2025;388(6746):532–37.
11. Rushing CS, Ryder TB, Scarpignato AL, Saracco JF, Marra PP. Using demographic attributes from long-term monitoring data to delineate natural population structure. J Appl Ecol. 2016;53(2):491–500.
12. Fuentes M, Van Doren BM, Fink D, Sheldon D. BirdFlow: learning seasonal bird movements from eBird data. Methods Ecol Evol. 2023;14(3):923–38.
13. Fink D, Auer T, Johnston A, Ruiz-Gutierrez V, Hochachka WM, Kelling S. Modeling avian full annual cycle distribution and population trends with citizen science data. Ecol Appl. 2020;30(3):e02056.
14. Fink D, Auer T, Johnston A, Strimas-Mackey M, Ligocki S, Robinson O, et al. eBird status and trends: data version 2023. Ithaca (NY): Cornell Lab of Ornithology; 2025.
15. Scarpignato AL, Huysman AE, Jimenez MF, Witko CJ, Harrison AL, Seavy NE, et al. Shortfalls in tracking data available to inform North American migratory bird conservation. Biol Conserv. 2023;286:110224.
16. Meehan TD, Saunders SP, DeLuca WV, Michel NL, Grand J, Deppe JL, et al. Integrating data types to estimate spatial patterns of avian migration across the Western Hemisphere. Ecol Appl. 2022;32(7):e2679.

17. Taylor PD, Crewe TL, Mackenzie SA, Lepage D, Aubry Y, Crysler Z, et al. The Motus Wildlife Tracking System: a collaborative research network to enhance the understanding of wildlife movement. Avian Conserv Ecol. 2017;12(1):art8.
18. Strimas-Mackey M, Ligocki S, Auer T, Fink D. ebirdst: access and analyze eBird status and trends data products. 2025. https://CRAN.R-project.org/package=ebirdst. Accessed 16 Dec 2025.
19. Celis-Murillo A, Malorodova M, Nakash E. North American Bird Banding Program dataset 1960–2022. U.S. Geological Survey data release. 2022. doi:10.5066/P9BSM38F.
20. Plunkett E. BirdFlowR: predict and visualize bird movement. 2025. https://github.com/birdflow-science/BirdFlowR. Accessed 16 Dec 2025.
21. Alerstam T. The geographical scale factor in orientation of migrating birds. J Exp Biol. 1996;199(1):9–19.
22. McLean DJ, Skowron Volponi MA. trajr: an R package for characterisation of animal trajectories. Ethology. 2018;124(6):440–8.
23. Tobias JA, Sheard C, Pigot AL, Devenish AJM, Schleuning M, et al. AVONET: morphological, ecological and geographical data for all birds. Ecol Lett. 2022;25(3):581–97.
24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
25. Bridge ES, Thorup K, Bowlin MS, Chilson PB, Diehl RH, Fléron RW, et al. Technology on the move: recent and forthcoming innovations for tracking migratory birds. BioScience. 2011;61(9):689–98.
26. Arnold TW, de Sobrino CN, Specht HM. Estimating annual survival rates from sparse band-recovery data: examples from migratory shore and upland gamebirds. Washington (DC): U.S. Fish and Wildlife Service; 2013.

27. Bonnet-Lebrun AS, Somveille M, Rodrigues AS, Manica A. Exploring intraspecific variation in migratory destinations to investigate the drivers of migration. Oikos. 2021;130(2):187–96.
28. Delmore KE, Fox JW, Irwin DE. Dramatic intraspecific differences in migratory routes, stopover sites and wintering areas revealed using light-level geolocators. Proc R Soc B. 2012;279(1747):4582–9.
29. Finch T, Butler SJ, Franco AM, Cresswell W. Low migratory connectivity is common in long-distance migrant birds. J Anim Ecol. 2017;86(3):662–73.
30. Catry P, Encarnação V, Araújo A, Fearon P, Fearon A, Armelin M, et al. Are long-distance migrant passerines faithful to their stopover sites? J Avian Biol. 2004;35(2):170–81.
31. O'Neal BJ, Stafford JD, Larkin RP, Michel ES. The effect of weather on the decision to migrate from stopover sites by autumn-migrating ducks. Mov Ecol. 2018;6(1):23.
32. Horton KG, Deng Y. Stopover strategies drive potential adaptability under changing environments. Proc Natl Acad Sci USA. 2024;121(20):e2406694121.
33. Studds CE, Kendall BE, Murray NJ, Wilson HB, Rogers DI, Clemens RS, et al. Rapid population decline in migratory shorebirds relying on Yellow Sea tidal mudflats as stopover sites. Nat Commun. 2017;8(1):14895.
34. Hedenström A. Adaptations to migration in birds: behavioural strategies, morphology and scaling effects. Philos Trans R Soc B Biol Sci. 2008;363(1490):287–99.
35. Linek N, Yanco SW, Volkmer T, Zuñiga D, Wikelski M, Partecke J. Migratory lifestyle carries no added overall energy cost in a partial migratory songbird. Nat Ecol Evol. 2024;8(12):2286–96.
36. Dufour P, Descamps S, Chantepie S, Renaud J, Guéguen M, Schiffers K, et al. Reconstructing the geographic and climatic origins of long-distance bird migrations. J Biogeogr. 2020;47(1):155–66.

37. Buderman FE, Hanks EM, Ruiz-Gutierrez V, Shull M, Murphy RK, Miller DA. Integrated movement models for individual tracking and species distribution data. Methods Ecol Evol. 2025;16(2):345–61.

Supplemental Information for:

Population-level migration modeling of North America's birds through data integration with BirdFlow

Table of Contents:

Supplementary Figure 1	Page 2
Supplementary Figure 2	Page 3
Supplementary Figure 3	Page 4
Supplementary Figure 4	Page 5
Supplementary Figure 5	Page 6
Supplementary Figure 6	Page 7
Supplementary Figure 7	Page 8
Supplementary Figure 8	Page 9
Supplementary Figure 9	Page 10
Supplementary Figure 10	Page 11
Supplementary Table 1	Page 12
Supplementary Table 2	Page 13 – 17
Supplementary Table 3	Page 18
Supplementary Methods	Page 19 – 22
References	Page 23

Supplementary Figures

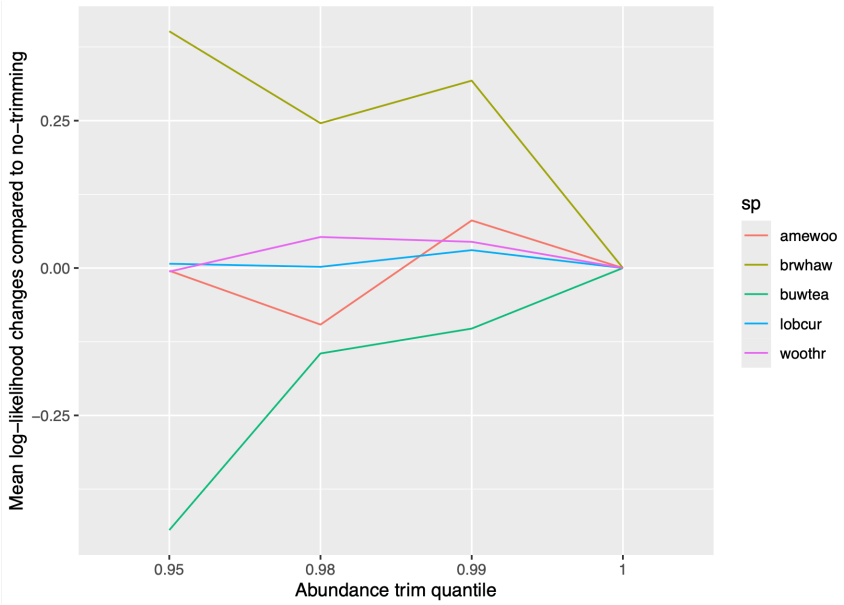

Figure S1. Sensitivity analysis of trimming quantile choices. To avoid the influence of high-abundance outliers, we truncated abundance distributions by replacing all values above the 0.99 quantile with the 0.99 quantile value. We showed that this outlier removal improved the average log-likelihood across the 225 models for 4 out of 5 species tested, for which individual tracking data were available at the time of analysis. The multi-species averaged log-likelihood improvement was the highest with 0.99 quantile trimming (ΔLL = 0.074), followed by 0.98 (ΔLL = 0.012), 1 (no trimming; ΔLL = 0), and 0.95 (ΔLL = -0.009).

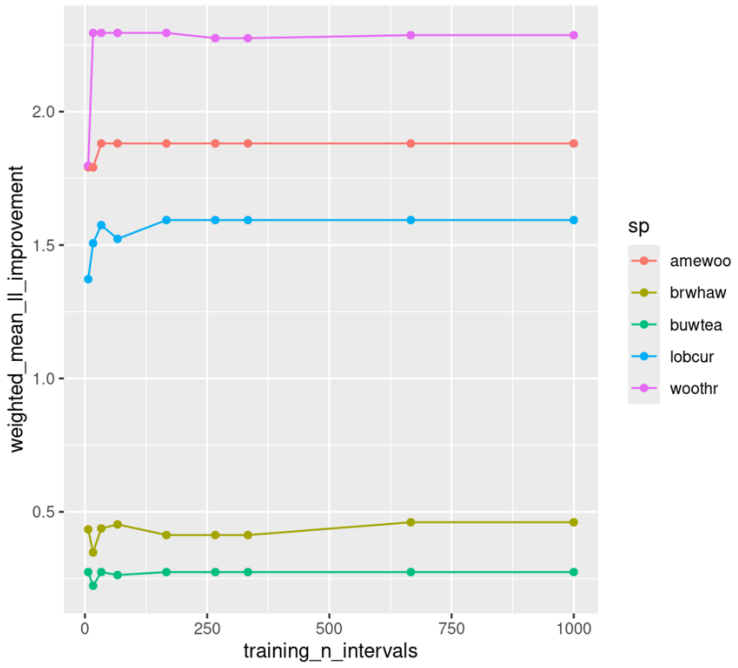

Figure S2. Sensitivity analysis of number of training transitions. To determine the minimum number of transitions required to train BirdFlow models, we applied a learning-curve analysis for five species for which individual tracking data were available at the time of analysis, each trained with increasing number of transitions. Model performance was evaluated on a hold-out test dataset as described in the Method section. We found that models trained with approximately 20–50 transitions achieved substantially improved performance (log-likelihood) and began to approach a performance plateau. We therefore chose a threshold of 20 transitions to train the models for maximum number of species.

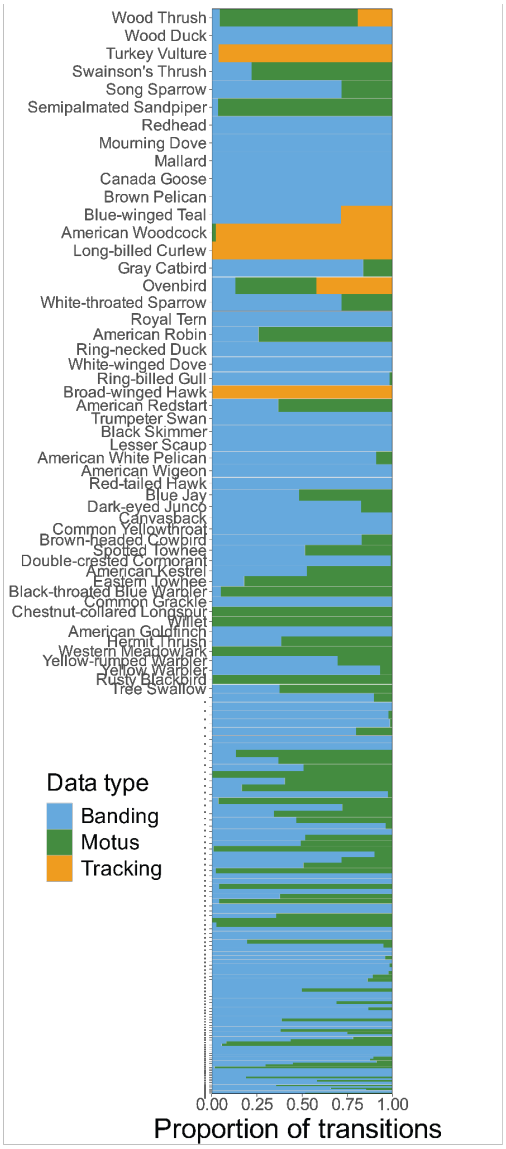

Figure S3. The transition data composition for each species, colored by data sources. The height of each block is proportional to the square root of total transition count available for the species. The width of the block represents the proportion of transitions from each data source for each species. The top 50 species with most abundant transition data are labeled with their eBird English common name.

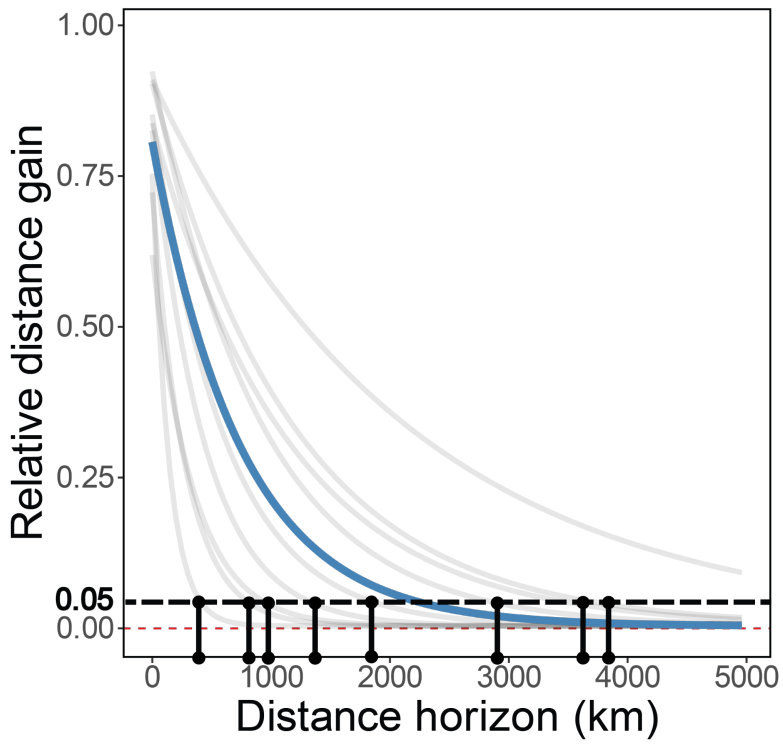

Figure S4. A schematic diagram of how the maximum functional forecasting horizon is calculated. We set a threshold of 0.05 for both relative distance gain and log-likelihood improvement, with values below the threshold being considered as BirdFlow models having no advantages over the baseline. We calculated the intersection of each species' model performance decay line with the threshold, and averaged the distance horizon at the intersection as a multispecies averaged maximum functional horizon.

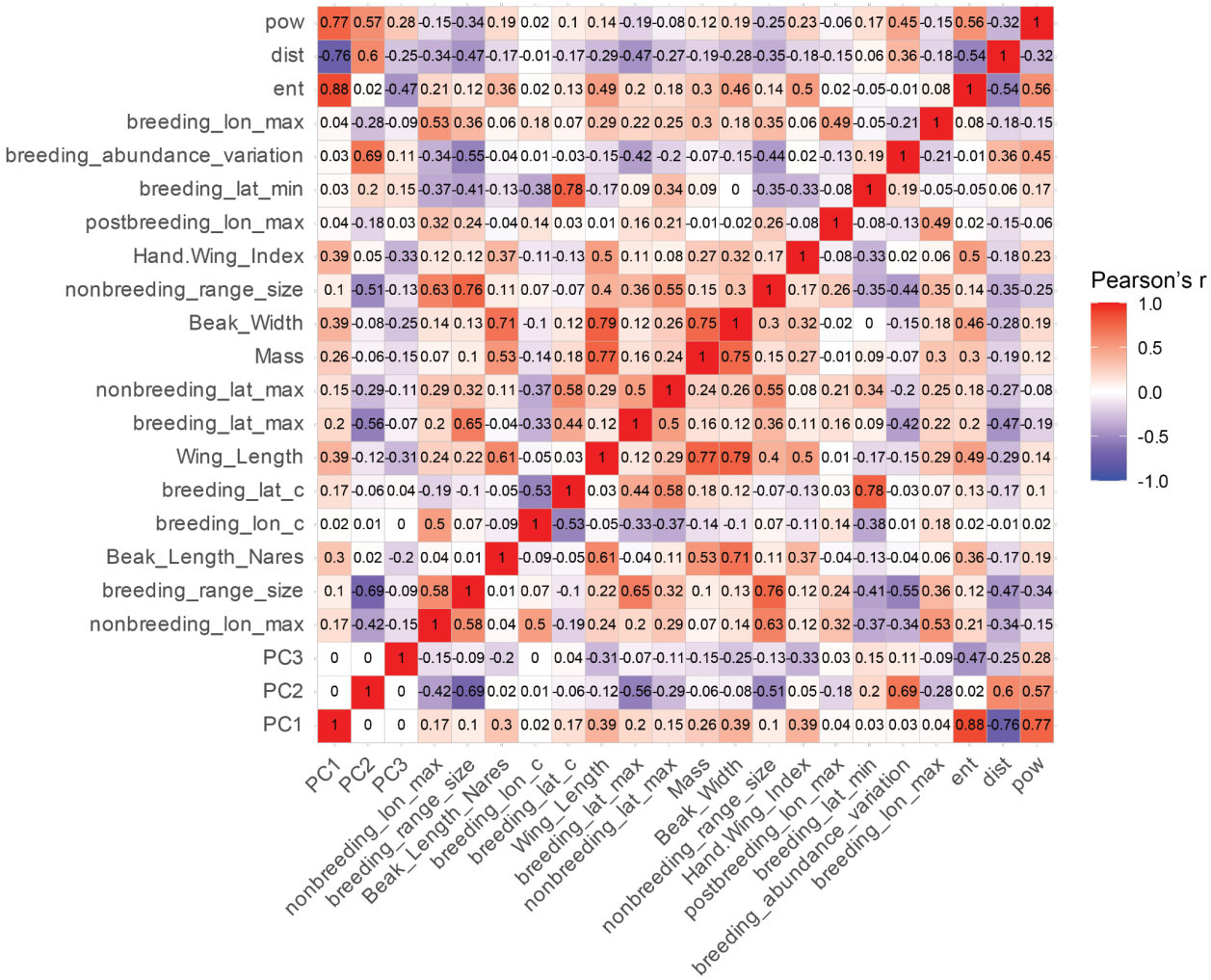

Figure S5. Pairwise Pearson's correlation coefficients among variables included in the Random Forest regression. The variables include the hyperparameters, principal components of the hyperparameters, species range summaries, and species morphological traits. Each datapoint is a species. Species range summaries and morphological traits with Pearson's correlation coefficients over 0.8 were considered highly collinear, and in such cases only one variable was retained while the other was removed before the analysis.

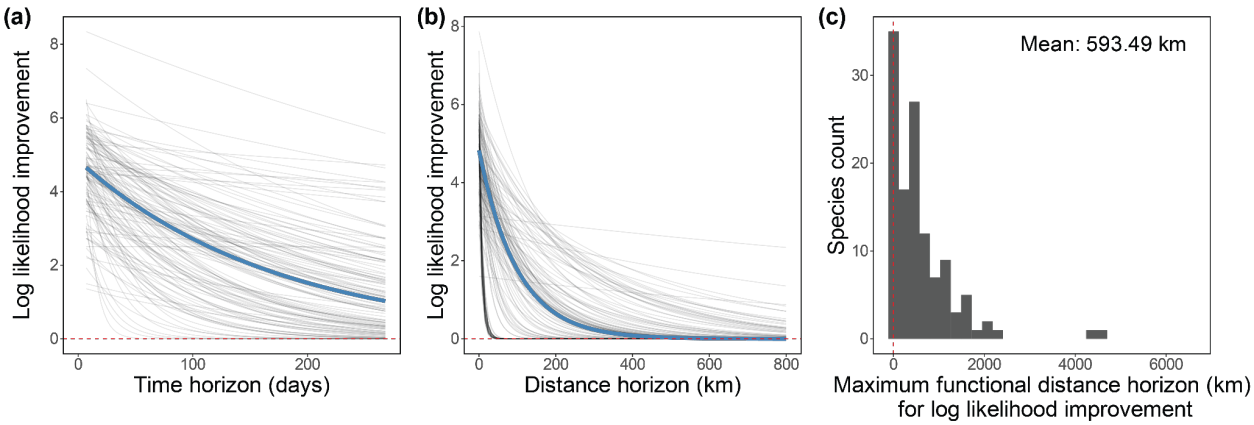

Figure S6. Model performance decays as time and distance horizon increase. Model performance decay is measured as the decrease of log-likelihood as the time horizon (a) and distance horizon (b) extends. Gray lines represent the fitted exponential regression function for each species, and blue lines represent the functions using cross-species-averaged parameters. (c) measures the maximum distance horizon that BirdFlow models still have advantages against a random sample from the abundance surface in terms of log-likelihood across species, defined as the horizon where the exponential decay curve in (b) intersects the y=0.05 horizontal line (1.05 times better than S&T). Red dashed lines in (c) indicate zero distance horizon. See Supplementary Fig. 4 for schematic diagram of calculation in (c). On average, the maximum functional distance horizon is 593.49 km for log-likelihood improvement (95% percentile interval: 31 km - 2145 km). The fact that the maximum functional distance horizon for log-likelihood improvement (593 km) is lower than that for relative distance gain (Fig. 3) demonstrates that log-likelihood is less tolerant to error accumulation along the prediction horizon, since it only measures the performance at the exact resighting location and does not consider geographic closeness.

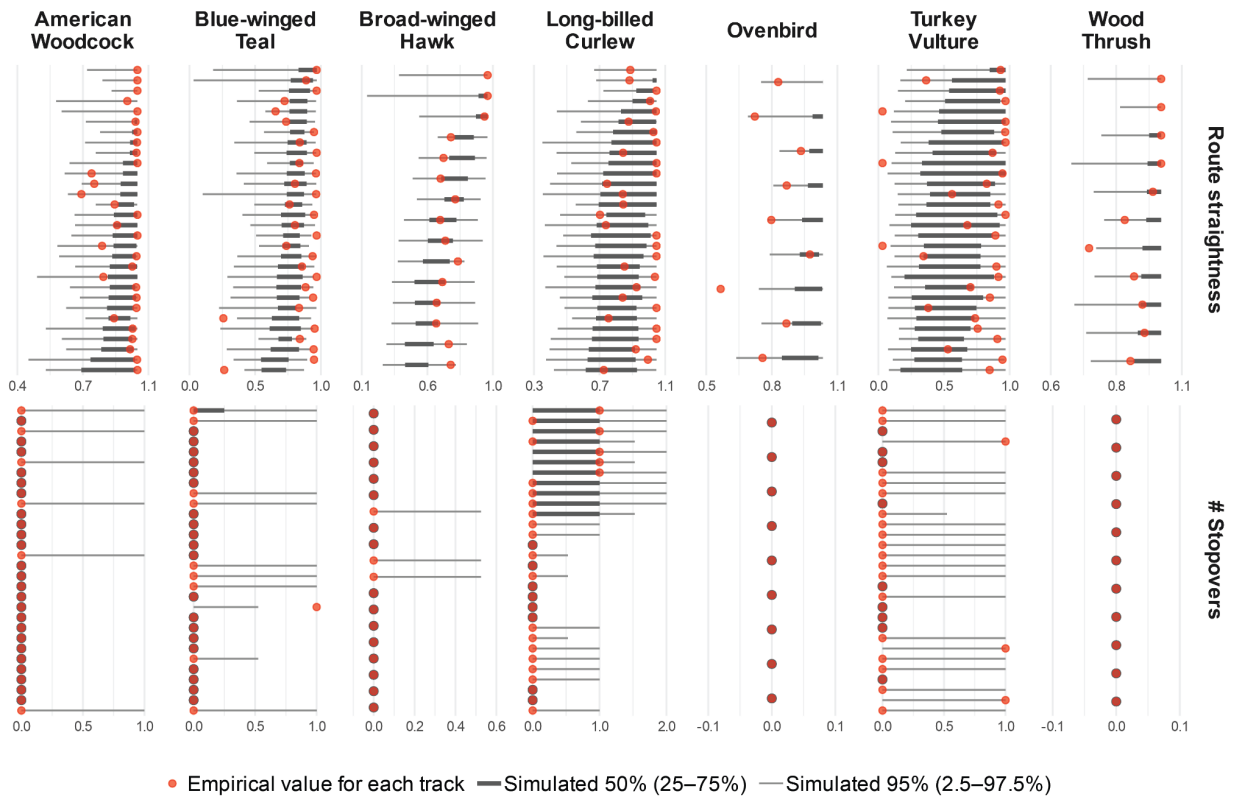

Figure S7. BirdFlow models simulate biologically plausible route straightness and number of stopovers. For each species, we randomly sampled at most 30 tracks from pre-breeding or post-breeding migration seasons. Red points indicate the empirical metrics calculated from the GPS tracks and the gray bars indicate 95% and 50% percentile intervals of the simulated metrics. A quantitative summary of the figure can be found in Additional file 1: Table S3.

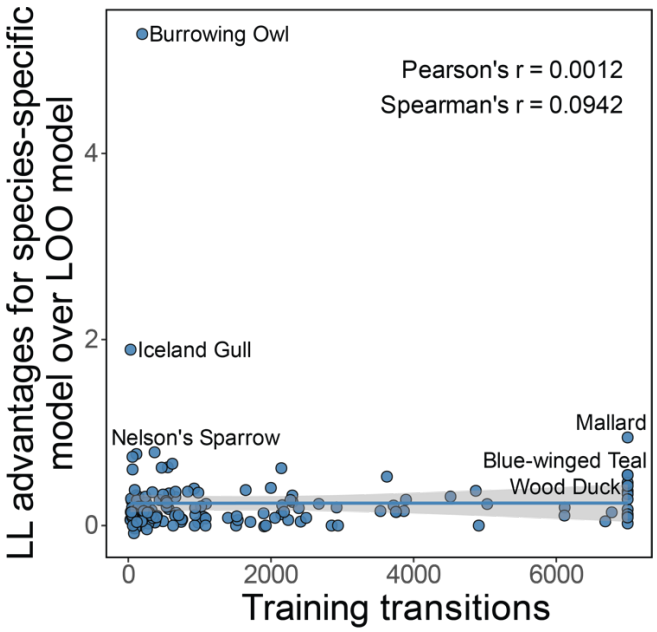

Figure S8. Lack of correlation between the improvements of species-specifically tuned models and the amount of training transitions. Although species-specific tuning improves model performance compared to All-LOO baselines, more training data does not necessarily lead to higher log-likelihood improvement across species. Each data point represents a best model for each species tuned on species-specific data.

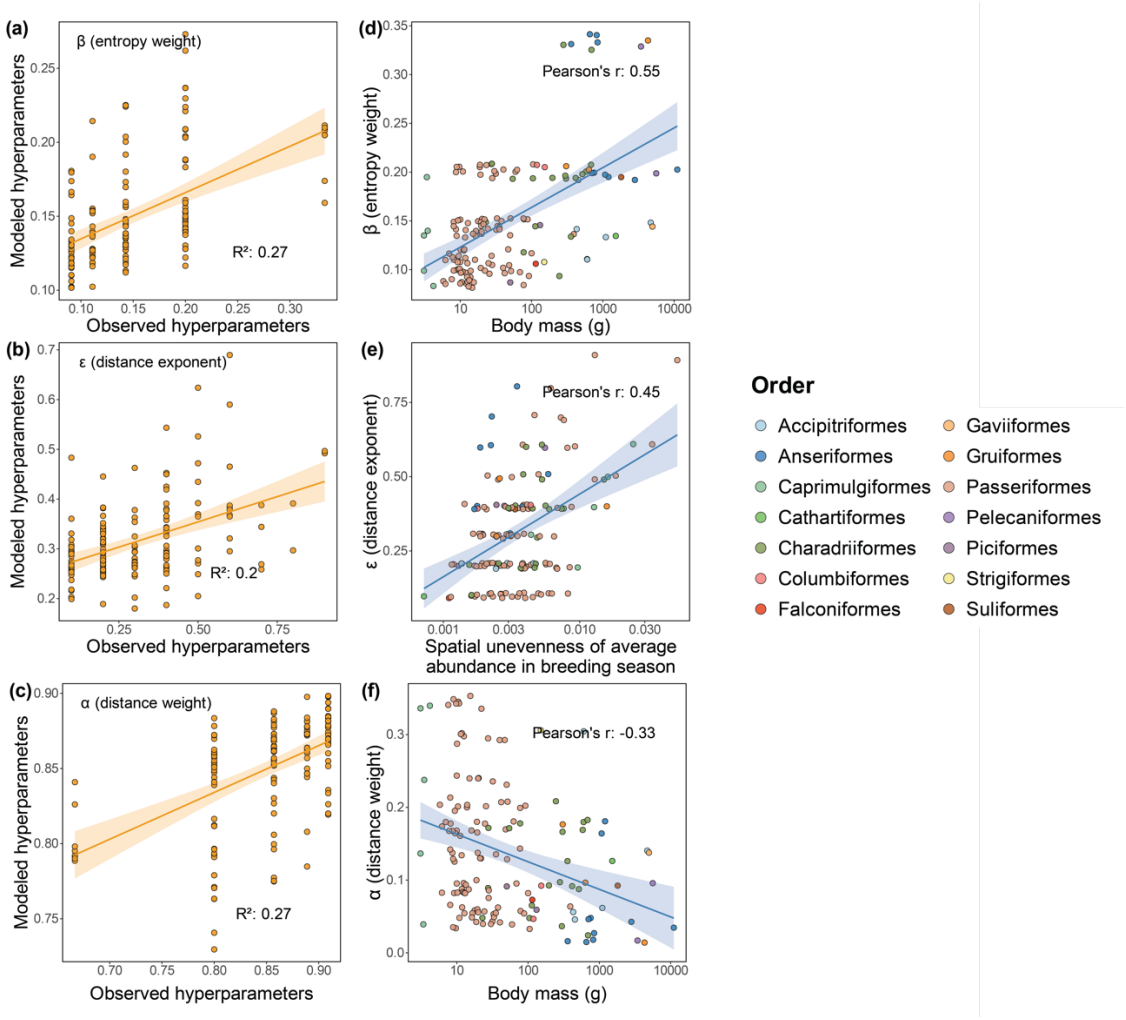

Figure 9. Explaining hyperparameters using species traits. We trained Random Forest models using species morphological traits and range summary statistics to explain the hyperparameter patterns among species. The Random Forest model was trained using 5-fold cross validation, and the out-of-fold prediction results for all species are shown for hyperparameter β (a), ϵ (b) and α (c), respectively. The hyperparameters are also linearly regressed against the explanatory variables with the highest feature importance score in the Random Forest models for β (d), ϵ (e) and α (f). The values on the horizontal axes are log-transformed to calculate the correlation coefficients in (d-f). Small jitters in x and y axes were added in (d-f) to disperse overlapping points.

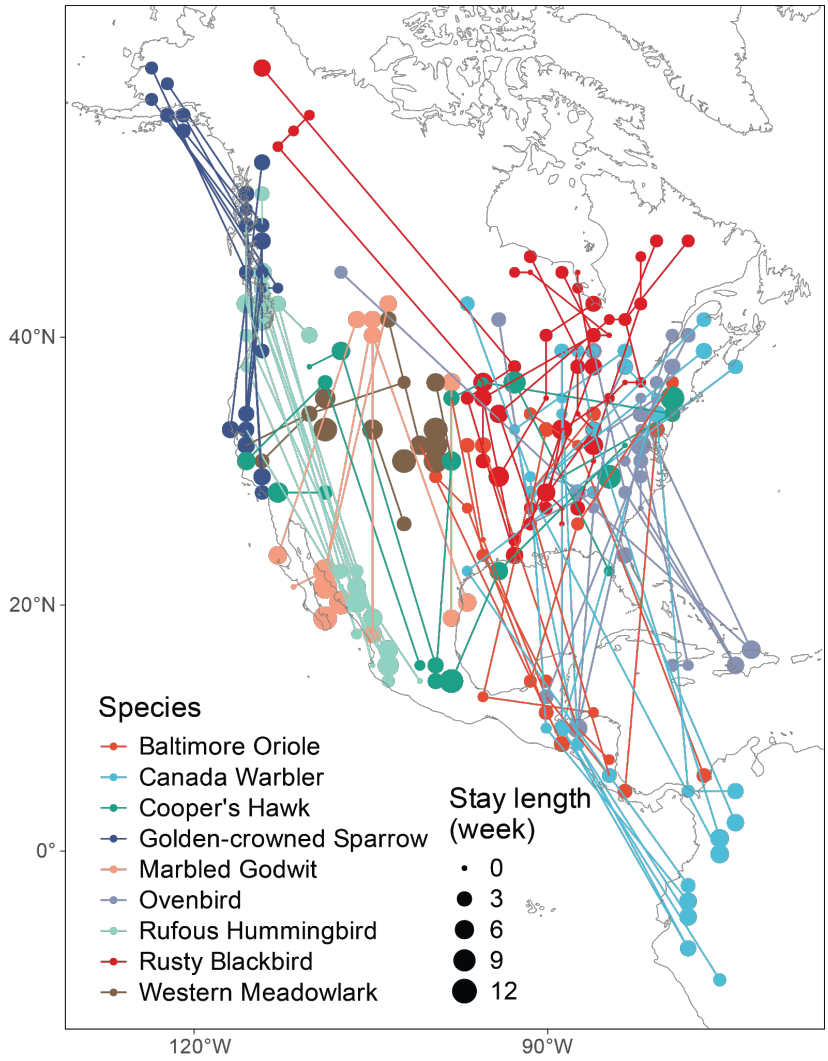

Figure S10. The BirdFlow-simulated spring migration routes for 9 migratory species. Stay length indicates the inferred total length of stay at the stopover sites in weekly temporal resolution, with zero stay length indicating non-stopping or stopover shorter than one week. 10 routes were simulated for each species.

Supplementary Tables

Table S1. The sources of tracking data for each species used in this research.

Common name	Scientific name	Data source
American Woodcock	*Scolopax minor*	Movebank ID: 12170798, 351564596
Turkey Vulture	*Cathartes aura*	Movebank ID: 16880941
Ovenbird	*Seiurus aurocapilla*	Movebank ID: 205765709
Long-billed Curlew	*Numenius americanus*	Movebank ID: 42451582
Blue-winged Teal	*Spatula discors*	Movebank ID: 446107870
Broad-winged Hawk	*Buteo platypterus*	Movebank ID: 28691134
Wood Thrush	*Hylocichla mustelina*	Supplementary material of Stanley et al. 2015

875 **Table S2.** Species list for the 153 modeled species.

Common name	eBird species code	Scientific name	Order	Family
Bald Eagle	baleag	*Haliaeetus leucocephalus*	Accipitriformes	Accipitridae
Broad-winged Hawk	brwhaw	*Buteo platypterus*	Accipitriformes	Accipitridae
Cooper's Hawk	coohaw	*Accipiter cooperii*	Accipitriformes	Accipitridae
Red-shouldered Hawk	reshaw	*Buteo lineatus*	Accipitriformes	Accipitridae
Red-tailed Hawk	rethaw	*Buteo jamaicensis*	Accipitriformes	Accipitridae
American Wigeon	amewig	*Mareca americana*	Anseriformes	Anatidae
Blue-winged Teal	buwtea	*Spatula discors*	Anseriformes	Anatidae
Canada Goose	cangoo	*Branta canadensis*	Anseriformes	Anatidae
Canvasback	canvas	*Aythya valisineria*	Anseriformes	Anatidae
Lesser Scaup	lessca	*Aythya affinis*	Anseriformes	Anatidae
Mallard	mallar3	*Anas platyrhynchos*	Anseriformes	Anatidae
Redhead	redhea	*Aythya americana*	Anseriformes	Anatidae
Ring-necked Duck	rinduc	*Aythya collaris*	Anseriformes	Anatidae
Trumpeter Swan	truswa	*Cygnus buccinator*	Anseriformes	Anatidae
Wood Duck	wooduc	*Aix sponsa*	Anseriformes	Anatidae
Anna's Hummingbird	annhum	*Calypte anna*	Caprimulgiformes	Trochilidae
Black-chinned Hummingbird	bkchum	*Archilochus alexandri*	Caprimulgiformes	Trochilidae
Broad-billed Hummingbird	brbhum	*Cynanthus latirostris*	Caprimulgiformes	Trochilidae
Ruby-throated Hummingbird	rthhum	*Archilochus colubris*	Caprimulgiformes	Trochilidae
Rufous Hummingbird	rufhum	*Selasphorus rufus*	Caprimulgiformes	Trochilidae
Turkey Vulture	turvul	*Cathartes aura*	Cathartiformes	Cathartidae
Piping Plover	pipplo	*Charadrius melodus*	Charadriiformes	Charadriidae
American Oystercatcher	ameoys	*Haematopus palliatus*	Charadriiformes	Haematopodidae
Black Skimmer	blkski	*Rynchops niger*	Charadriiformes	Laridae
California Gull	calgul	*Larus californicus*	Charadriiformes	Laridae
Franklin's Gull	fragul	*Leucophaeus pipixcan*	Charadriiformes	Laridae
Iceland Gull	y00478	*Larus glaucoides*	Charadriiformes	Laridae
Ring-billed Gull	ribgul	*Larus delawarensis*	Charadriiformes	Laridae
Royal Tern	royter1	*Thalasseus maximus*	Charadriiformes	Laridae
American Avocet	ameavo	*Recurvirostra americana*	Charadriiformes	Recurvirostridae
American Woodcock	amewoo	*Scolopax minor*	Charadriiformes	Scolopacidae

Least Sandpiper	leasan	*Calidris minutilla*	Charadriiformes	Scolopacidae
Lesser Yellowlegs	lesyel	*Tringa flavipes*	Charadriiformes	Scolopacidae
Long-billed Curlew	lobcur	*Numenius americanus*	Charadriiformes	Scolopacidae
Long-billed Dowitcher	lobdow	*Limnodromus scolopaceus*	Charadriiformes	Scolopacidae
Marbled Godwit	margod	*Limosa fedoa*	Charadriiformes	Scolopacidae
Semipalmated Sandpiper	semsan	*Calidris pusilla*	Charadriiformes	Scolopacidae
Western Sandpiper	wessan	*Calidris mauri*	Charadriiformes	Scolopacidae
Willet	willet1	*Tringa semipalmata*	Charadriiformes	Scolopacidae
Wilson's Snipe	wilsni1	*Gallinago delicata*	Charadriiformes	Scolopacidae
Mourning Dove	moudov	*Zenaida macroura*	Columbiformes	Columbidae
White-winged Dove	whwdov	*Zenaida asiatica*	Columbiformes	Columbidae
American Kestrel	amekes	*Falco sparverius*	Falconiformes	Falconidae
Common Loon	comloo	*Gavia immer*	Gaviiformes	Gaviidae
Sandhill Crane	sancra	*Antigone canadensis*	Gruiformes	Gruidae
American Coot	y00475	*Fulica americana*	Gruiformes	Rallidae
Clapper Rail	clarai11	*Rallus crepitans*	Gruiformes	Rallidae
Cedar Waxwing	cedwax	*Bombycilla cedrorum*	Passeriformes	Bombycillidae
Chestnut-collared Longspur	chclon	*Calcarius ornatus*	Passeriformes	Calcariidae
Black-headed Grosbeak	bkhgro	*Pheucticus melanocephalus*	Passeriformes	Cardinalidae
Indigo Bunting	indbun	*Passerina cyanea*	Passeriformes	Cardinalidae
Lazuli Bunting	lazbun	*Passerina amoena*	Passeriformes	Cardinalidae
Rose-breasted Grosbeak	robgro	*Pheucticus ludovicianus*	Passeriformes	Cardinalidae
Western Tanager	westan	*Piranga ludoviciana*	Passeriformes	Cardinalidae
Brown Creeper	brncre	*Certhia americana*	Passeriformes	Certhiidae
American Crow	amecro	*Corvus brachyrhynchos*	Passeriformes	Corvidae
Blue Jay	blujay	*Cyanocitta cristata*	Passeriformes	Corvidae
American Goldfinch	amegfi	*Spinus tristis*	Passeriformes	Fringillidae
Purple Finch	purfin	*Haemorhous purpureus*	Passeriformes	Fringillidae
Tree Swallow	treswa	*Tachycineta bicolor*	Passeriformes	Hirundinidae
Baltimore Oriole	balori	*Icterus galbula*	Passeriformes	Icteridae
Brown-headed Cowbird	bnhcow	*Molothrus ater*	Passeriformes	Icteridae
Bullock's Oriole	bulori	*Icterus bullockii*	Passeriformes	Icteridae
Common Grackle	comgra	*Quiscalus quiscula*	Passeriformes	Icteridae
Eastern Meadowlark	easmea	*Sturnella magna*	Passeriformes	Icteridae

Orchard Oriole	orcori	*Icterus spurius*	Passeriformes	Icteridae
Red-winged Blackbird	rewbla	*Agelaius phoeniceus*	Passeriformes	Icteridae
Rusty Blackbird	rusbla	*Euphagus carolinus*	Passeriformes	Icteridae
Western Meadowlark	wesmea	*Sturnella neglecta*	Passeriformes	Icteridae
Yellow-headed Blackbird	yehbla	*Xanthocephalus xanthocephalus*	Passeriformes	Icteridae
Yellow-breasted Chat	yebcha	*Icteria virens*	Passeriformes	Icteriidae
Loggerhead Shrike	logshr	*Lanius ludovicianus*	Passeriformes	Laniidae
Brown Thrasher	brnthr	*Toxostoma rufum*	Passeriformes	Mimidae
Gray Catbird	grycat	*Dumetella carolinensis*	Passeriformes	Mimidae
Northern Mockingbird	normoc	*Mimus polyglottos*	Passeriformes	Mimidae
American Redstart	amered	*Setophaga ruticilla*	Passeriformes	Parulidae
Bay-breasted Warbler	babwar	*Setophaga castanea*	Passeriformes	Parulidae
Black-and-white Warbler	bawwar	*Mniotilta varia*	Passeriformes	Parulidae
Black-throated Blue Warbler	btbwar	*Setophaga caerulescens*	Passeriformes	Parulidae
Black-throated Green Warbler	btnwar	*Setophaga virens*	Passeriformes	Parulidae
Blackburnian Warbler	bkbwar	*Setophaga fusca*	Passeriformes	Parulidae
Canada Warbler	canwar	*Cardellina canadensis*	Passeriformes	Parulidae
Cape May Warbler	camwar	*Setophaga tigrina*	Passeriformes	Parulidae
Cerulean Warbler	cerwar	*Setophaga cerulea*	Passeriformes	Parulidae
Chestnut-sided Warbler	chswar	*Setophaga pensylvanica*	Passeriformes	Parulidae
Common Yellowthroat	comyel	*Geothlypis trichas*	Passeriformes	Parulidae
Golden-winged Warbler	gowwar	*Vermivora chrysoptera*	Passeriformes	Parulidae
Hooded Warbler	hoowar	*Setophaga citrina*	Passeriformes	Parulidae
Kentucky Warbler	kenwar	*Geothlypis formosa*	Passeriformes	Parulidae
Louisiana Waterthrush	louwat	*Parkesia motacilla*	Passeriformes	Parulidae
MacGillivray's Warbler	macwar	*Geothlypis tolmiei*	Passeriformes	Parulidae
Magnolia Warbler	magwar	*Setophaga magnolia*	Passeriformes	Parulidae
Mourning Warbler	mouwar	*Geothlypis philadelphia*	Passeriformes	Parulidae
Nashville Warbler	naswar	*Leiothlypis ruficapilla*	Passeriformes	Parulidae
Northern Parula	norpar	*Setophaga americana*	Passeriformes	Parulidae
Northern Waterthrush	norwat	*Parkesia noveboracensis*	Passeriformes	Parulidae
Orange-crowned Warbler	orcwar	*Leiothlypis celata*	Passeriformes	Parulidae
Ovenbird	ovenbi1	*Seiurus aurocapilla*	Passeriformes	Parulidae

Pine Warbler	pinwar	*Setophaga pinus*	Passeriformes	Parulidae
Prairie Warbler	prawar	*Setophaga discolor*	Passeriformes	Parulidae
Swainson's Warbler	swawar	*Limnothlypis swainsonii*	Passeriformes	Parulidae
Tennessee Warbler	tenwar	*Leiothlypis peregrina*	Passeriformes	Parulidae
Wilson's Warbler	wlswar	*Cardellina pusilla*	Passeriformes	Parulidae
Yellow Warbler	yelwar	*Setophaga petechia*	Passeriformes	Parulidae
Yellow-rumped Warbler	yerwar	*Setophaga coronata*	Passeriformes	Parulidae
American Tree Sparrow	amtspa	*Spizelloides arborea*	Passeriformes	Passerellidae
Chipping Sparrow	chispa	*Spizella passerina*	Passeriformes	Passerellidae
Clay-colored Sparrow	clcspa	*Spizella pallida*	Passeriformes	Passerellidae
Dark-eyed Junco	daejun	*Junco hyemalis*	Passeriformes	Passerellidae
Eastern Towhee	eastow	*Pipilo erythrophthalmus*	Passeriformes	Passerellidae
Field Sparrow	fiespa	*Spizella pusilla*	Passeriformes	Passerellidae
Fox Sparrow	foxspa	*Passerella iliaca*	Passeriformes	Passerellidae
Golden-crowned Sparrow	gocspa	*Zonotrichia atricapilla*	Passeriformes	Passerellidae
Harris's Sparrow	harspa	*Zonotrichia querula*	Passeriformes	Passerellidae
Lincoln's Sparrow	linspa	*Melospiza lincolnii*	Passeriformes	Passerellidae
Nelson's Sparrow	nstspa	*Ammospiza nelsoni*	Passeriformes	Passerellidae
Saltmarsh Sparrow	sstspa	*Ammospiza caudacuta*	Passeriformes	Passerellidae
Savannah Sparrow	savspa	*Passerculus sandwichensis*	Passeriformes	Passerellidae
Seaside Sparrow	seaspa	*Ammospiza maritima*	Passeriformes	Passerellidae
Song Sparrow	sonspa	*Melospiza melodia*	Passeriformes	Passerellidae
Spotted Towhee	spotow	*Pipilo maculatus*	Passeriformes	Passerellidae
Swamp Sparrow	swaspa	*Melospiza georgiana*	Passeriformes	Passerellidae
Vesper Sparrow	vesspa	*Pooecetes gramineus*	Passeriformes	Passerellidae
White-crowned Sparrow	whcspa	*Zonotrichia leucophrys*	Passeriformes	Passerellidae
White-throated Sparrow	whtspa	*Zonotrichia albicollis*	Passeriformes	Passerellidae
Blue-gray Gnatcatcher	buggna	*Polioptila caerulea*	Passeriformes	Polioptilidae
Golden-crowned Kinglet	gockin	*Regulus satrapa*	Passeriformes	Regulidae
Ruby-crowned Kinglet	ruckin	*Corthylio calendula*	Passeriformes	Regulidae
Red-breasted Nuthatch	rebnut	*Sitta canadensis*	Passeriformes	Sittidae
European Starling	eursta	*Sturnus vulgaris*	Passeriformes	Sturnidae
Marsh Wren	marwre	*Cistothorus palustris*	Passeriformes	Troglodytidae
Pacific Wren	pacwre1	*Troglodytes pacificus*	Passeriformes	Troglodytidae
Winter Wren	winwre3	*Troglodytes hiemalis*	Passeriformes	Troglodytidae

American Robin	amerob	*Turdus migratorius*	Passeriformes	Turdidae
Eastern Bluebird	easblu	*Sialia sialis*	Passeriformes	Turdidae
Hermit Thrush	herthr	*Catharus guttatus*	Passeriformes	Turdidae
Swainson's Thrush	swathr	*Catharus ustulatus*	Passeriformes	Turdidae
Varied Thrush	varthr	*Ixoreus naevius*	Passeriformes	Turdidae
Wood Thrush	woothr	*Hylocichla mustelina*	Passeriformes	Turdidae
Acadian Flycatcher	acafly	*Empidonax virescens*	Passeriformes	Tyrannidae
Eastern Phoebe	easpho	*Sayornis phoebe*	Passeriformes	Tyrannidae
Eastern Wood-Pewee	eawpew	*Contopus virens*	Passeriformes	Tyrannidae
Least Flycatcher	leafly	*Empidonax minimus*	Passeriformes	Tyrannidae
Western Wood-Pewee	wewpew	*Contopus sordidulus*	Passeriformes	Tyrannidae
Willow Flycatcher	wilfly	*Empidonax traillii*	Passeriformes	Tyrannidae
Warbling Vireo	warvir	*Vireo gilvus*	Passeriformes	Vireonidae
White-eyed Vireo	whevir	*Vireo griseus*	Passeriformes	Vireonidae
American White Pelican	amwpel	*Pelecanus erythrorhynchos*	Pelecaniformes	Pelecanidae
Brown Pelican	brnpel	*Pelecanus occidentalis*	Pelecaniformes	Pelecanidae
Northern Flicker	norfli	*Colaptes auratus*	Piciformes	Picidae
Yellow-bellied Sapsucker	yebsap	*Sphyrapicus varius*	Piciformes	Picidae
Burrowing Owl	burowl	*Athene cunicularia*	Strigiformes	Strigidae
Double-crested Cormorant	doccor	*Nannopterum auritum*	Suliformes	Phalacrocoracidae

Table S3. Summarizing the biological plausibility of BirdFlow-simulated routes. A biologically plausible model should expect around 95% of the empirically observed biological metrics to occur within the 95% simulated intervals by BirdFlow models, and roughly 50% of the empirically observed biological metrics to occur within the 50% simulated intervals. This result indicates that BirdFlow-generated routes are biologically realistic and reflect population-level variation. More than 50% of observations occurred within the 50% simulated intervals for the number of stopovers, likely because the coarse weekly temporal resolution increases variability and reduces informativeness of stopover calculations. Stationary tracks were not included for straightness validation.

Metric	Species	# GPS tracks for validation	% of empirical observations fall into the 95% simulated intervals	% of empirical observations fall into the 50% simulated intervals
Straightness	American Woodcock	452	95.58%	61.73%
	Blue-winged Teal	63	63.49%	26.98%
	Broad-winged Hawk	15	100.00%	60.00%
	Long-billed Curlew	280	98.21%	63.93%
	Ovenbird	9	88.89%	11.11%
	Turkey Vulture	97	91.75%	55.67%
	Wood Thrush	11	90.91%	63.64%
	Averaged		**89.83%**	**49.01%**
# stopovers	American Woodcock	651	100.00%	99.85%
	Blue-winged Teal	70	97.14%	94.29%
	Broad-winged Hawk	19	100.00%	100.00%
	Long-billed Curlew	296	100.00%	100.00%
	Ovenbird	9	100.00%	100.00%
	Turkey Vulture	118	100.00%	94.92%
	Wood Thrush	11	100.00%	100.00%
	Averaged		**99.59%**	**98.44%**
Migration speed	American Woodcock	651	94.93%	56.22%
	Blue-winged Teal	70	74.29%	31.43%
	Broad-winged Hawk	19	100.00%	57.89%
	Long-billed Curlew	296	99.66%	73.65%
	Ovenbird	9	55.56%	22.22%
	Turkey Vulture	118	85.59%	36.44%
	Wood Thrush	11	90.91%	27.27%
	Averaged		**85.85%**	**43.59%**

Supplementary methods

Motus data preprocessing

Raw Motus data were summarized into “visits” for each time a bird visited a Motus station, with arrival and departure times for that event. These data were then validated using multiple criteria so that tracks generated from these summaries had reasonable estimated flight speeds, fell within expected species ranges, and there were at least 3 consecutive detections for any given visit. In addition, Motus stations which were deemed ‘too noisy’ were eliminated from the results to reduce the risk of introducing false positives.

Transition sampling:

We applied an iterative transition sampling procedure that considers the trade-off of computational cost and model performance improvement.

(1) First, we iteratively sampled a single one-week transition from each of the tracks without replacement, until the maximum transition count was met or the target tracks run out of transitions to sample (or without any one-week transition to sample, as often seen in banding data), e.g., (t_1, x_1) to (t_2, x_2), and (t_2, x_2) to (t_3, x_3), where the t is the weekly timestep and x is the location corresponding to the timestep.

(2) Second, if the size of the one-week transition set did not exceed the maximum transition count, we iteratively randomly sampled one n-timestep transition (n can be any positive integer from 2 to maximum week duration for the track) from each track without replacement, until the maximum transition count was met or the target tracks run out of transitions to sample, e.g., (t_1, x_1) to (t_5, x_5).

The rationale behind this sampling strategy is to prioritize the information-rich one-week transitions, while also integrating the information in longer transitions to tune and evaluate the prediction performance of BirdFlow models across seasons.

To constrain the transitions in a reasonable spatiotemporal range, only transitions that span less than 270 days with at least one end falling into migration seasons were sampled. We did not sample transitions with both ends falling into breeding or non-breeding seasons as we focused on the migratory periods, and since the species detectability can change vastly in stationary periods (especially during breeding and molting periods), which may cause spatiotemporal variation in S&T abundance surfaces that do not reflect movement (e.g., Socolar et al. 2025)

Hyperparameter grid search settings

We reparametrized the hyperparameters for easier interpretation, defining the relative proportion of the observation term $\rho = 1/(1 + \alpha + \beta)$ and the relative proportion of the distance and entropy terms $\phi = \beta/\alpha$, which allows us to rewrite Eq. 1 as

$$\rho L_{obs}(\mu(\theta), \hat{\mu}) + (1 - \rho)\left[\frac{1}{1+\phi} L_{mov}(\mu(\theta)) + \frac{\phi}{1+\phi} J(\mu(\theta))\right] \quad \text{(Supplementary Eq. 1)}$$

Since different combinations of parameters have different emphases on matching S&T abundance surfaces and regulating the route behavior patterns, this reparameterization allowed us to directly vary the overall emphasis of the observation weight with ρ and vary the degree of stochasticity with ϕ. In our grid search, we varied ρ with values (0.95, 0.975, 0.99, 0.999, 0.9999) to emphasize the high importance of matching S&T abundance surface, ϕ across values (2, 4, 6, 8, 10), and ϵ with values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9). We chose

these grid search values based on the empirical investigation in Fuentes et al. 2023 where none of the models with these parameter combinations perform the best with the extremal grid search values and the performance appears to vary smoothly as the hyperparameters change, which indicates a potentially optimal model enclosed in this hyperparameter surface. We configured parameters and passed them to the BirdFlowPy package (Fuentes 2022, Fuentes et al. 2023) for model fitting.

Model decay analysis

The exponential regression can be formulated as:

$$f(x) = ae^{-bx} + c \qquad \text{(Supplementary Eq. 2)}$$

Where a, b, and c are parameters and x is the independent variable (time or distance horizon). The multi-species median regression line is calculated by taking the median value of each parameter across species, respectively. The maximum functional distance horizon is calculated as the value on the x-axis where the regression line intersects with a threshold, which was set to 0.05 for both the relative distance gain and the log-likelihood improvement (supplementary Fig. 4). For relative distance gain, a 0.05 value represents a horizon where the BirdFlow prediction is only 5% closer to the empirical resighting locations than the S&T abundance surfaces, and for log-likelihood improvement, 0.05 means the BirdFlow prediction is 1.05 times more likely to recover the empirical observation than a random sample from the S&T abundance surfaces. While these thresholds are arbitrary, we believe they generally reflect a functional horizon, beyond which the BirdFlow model is no longer meaningfully better than the original S&T prediction.

Distributional summaries for each species

We calculated distributional summaries as explanatory variables for the best hyperparameters, including range size, spatial variation of abundance, the minimum, maximum, and centroid of longitude and latitude for different seasons. First, the eBird S&T abundance map was aggregated to 150km resolution and split into pre-breeding season, breeding season, post-breeding season, and non-breeding seasons. The season partitioning is queried from the eBird S&T dataset. Second, we averaged the abundance estimation across weeks for each spatial pixel by each season. Next, for each season, we calculated:

- Range size, as the maximum number of spatial pixels with abundance>0.
- Range abundance variation, as the standard deviation of the average-abundance variation across spatial pixels.
- Maximum and minimum range longitude and latitude: The maximum and minimum of the longitude and latitude across spatial pixels with abundance>0.
- Centroid of longitude and latitude: The abundance-weighted average longitude and latitude.

References:

Stanley CQ, McKinnon EA, Fraser KC, Macpherson MP, Casbourn G, Friesen L, et al. Connectivity of wood thrush breeding, wintering, and migration sites based on range-wide tracking. Conserv Biol. 2015;29(1):164–174.

Socolar, J., Galtbalt, B., Johnston, A., La Sorte, F. A., Robinson, O. J., Rosenberg, K. V. and Dokter, A. M. 2025. Seasonal macro-demography of North American bird populations revealed through participatory science. – Ecography 2025: e07349.

Fuentes, M. 2022. Miguel-Fuentes/birdflow: Birdflow (zenodo_release) – Zenodo <https://doi.org/10.5281/zenodo.7324572>.

Fuentes, M., Van Doren, B. M., Fink, D. and Sheldon, D. 2023. BirdFlow: Learning seasonal bird movements from eBird data. – Methods Ecol. Evol. 14: 923–938.