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

Abstract:

Accurate information on the population-level movements of migratory animals is highly valuable for migration research and critical for designing effective conservation strategies for a changing world. However, population-wide movement information is lacking for most migratory species due to the effort and expense needed to collect data across species' ranges. BirdFlow is a probabilistic modeling framework that infers population-level movements from weekly species distribution maps produced by the participatory science project eBird. Producing accurate species-specific BirdFlow models has required model tuning using high-resolution individual tracking data, which is not available for most migratory species. Here, we introduce a new model tuning framework that eliminates this reliance on tracking data and generalizes BirdFlow to hundreds of migratory species. This framework allows us to tune and validate BirdFlow models using a combination of data sources, including GPS tracks, banding recoveries, and radio telemetry data from the Motus Wildlife Tracking System. We investigate the generalizability 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. Our results show that the BirdFlow modeling framework achieves 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. Alongside this study, we release 153 tuned BirdFlow models, representing the first collection of large-scale population-level movement forecasting models and offering a foundation for more accurate predictions for applications in conservation, disease surveillance, aviation, and public outreach.

Keywords: population-level movement, probabilistic modeling, multi-source data, movement ecology, spatiotemporal, big data

Introduction:

Our understanding of continent-scale bird migration has been greatly fostered by modern tracking technologies and the rapid accumulation of participatory science data (Sullivan et al. 2014, Flack et al. 2022). 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 (Shaw 2020). 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., Kramer et al. 2018). 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 (Emmenegger et al. 2016, Zurell et al. 2018, Rosenberg et al. 2019, Buchan et al. 2022, Horton et al. 2023, Johnston et al. 2025). Therefore, understanding population-level movements within a species' range is essential for quantifying migratory connectivity and for targeting conservation actions in a changing world (Rushing et al. 2016). However, such information remains largely unavailable.

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 (Fuentes et al. 2023). The model builds on eBird Status & Trends relative abundance data (Fink et al. 2020, 2024; hereafter, eBird S&T) through a mechanistic optimization that accounts for energy costs and inter-population movement patterns. Performance of the model can be improved when hyperparameters are tuned using individual movement data from other sources, such as GPS and satellite tracking (Fuentes et al. 2023). However, generalizing BirdFlow to hundreds of species remains challenging because GPS tracking data are unavailable for most migratory birds (Scarpignato et al. 2023), a limitation shared by other continental-scale movement models (Meehan et al. 2022).

To overcome this limitation, we introduce a unified model tuning and validation framework to train BirdFlow models based on general mark-resight data. Our approach standardizes observations across multiple data sources, including GPS tracking data, banding recovery data, and Motus Wildlife Tracking System (Motus) automated radio telemetry data (Motus; https://motus.rog; Taylor et al. 2017). We address three primary questions: (1) Whether integrating multi-source data can enable BirdFlow modeling across hundreds of species; (2) Whether species with similar traits (e.g., morphology) share similar optimal model parameterizations; and (3) Whether species lacking individual-level data can still be modeled accurately by using tuned hyperparameters from other species.

Material and Methods:

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

The BirdFlow model

BirdFlow model structure and fitting methodology are described in detail in Fuentes et al. 2023. 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 fit by optimizing an objective function, with three terms controlled by three hyperparameters:

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

The first term is the "observation" term, which is the same as the "location loss" term in Fuentes et al. 2023. 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 "(negative Shannon) entropy" term and is regulated by hyperparameter β (the entropy weight), which controls the stochasticity of the routes (Fig. 1).

Species relative abundance data

We acquired eBird Status and Trends weekly relative abundance distributions (data version 2023; hereafter, S&T abundance surfaces) using the ebirdst package version 3.2023.1

(Strimas-Mackey et al. 2025). 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.

Individual mark-resight data (transitions)

We used empirical movement data to select the best hyperparameters from the 225 candidate models for each species. We acquired tracking data from Movebank (movebank.org) and published literature (see supplementary Table 1). We acquired bird banding data from the United States Geological Survey (USGS) Bird Banding Program (BBL) database (Celis-Murillo et al. 2022). Motus data was acquired following the Motus Collaboration Policy (Birds Canada 2023) and preprocessed to reduce errors (see supplementary information).

We converted all data into species-specific BirdFlow spatiotemporal coordinates using BirdFlowR package version 0.1.0.9075 (Plunkett 2025). We refer to a set of detections for a single individual bird movement 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 trade-off of computational cost and model performance, using an iterative sampling procedure (see supplementary information). For species to be included, we required availability of at least 20 transitions and high-quality abundance maps, resulting in 153 migratory species in North America spanning 14 orders and 39 families (Supplementary Table 2). Across species, the number of training transitions varied from 29 to 7000 (mean = 1673). Banding data accounted for 69.2% transitions per species on average, and Motus data made up 27.6%. GPS tracking data were available for only seven species, serving as the primary data source (> 50% of transitions) for four and comprising 11.8% of all transitions (Supplementary Fig. 2).

Model fitting and hyperparameter tuning

For each species, we fitted BirdFlow models with different combinations of hyperparameters (α, β, and ϵ introduced in the previous sections) in an exhaustive parameter grid (see supplementary information for details). In total, the fitting generated 225 models per species. We fitted models with a spatial resolution of 150 km and temporal resolution of 1 week. This spatial resolution balances computation cost while maximizing resolution. The 1-week temporal resolution matches the temporal resolution of the S&T data.

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 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:

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

$log\ p_{BF}(t'_i, x'_i \mid t_i, x_i)$ 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 is, given the model. $log\ A(t'_i, x'_i)$ is simply 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 ΔLL_i tells us how much more accurate the BirdFlow model prediction is compared to a random sample from the S&T abundance surface for transition i.

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 left the transition data of the target species out of the procedure and selected hyperparameters based on the transition data of the remaining species. We only considered models with a mean correlation score >0.98. 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 precludes 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 is less biologically interpretable. An evaluation of a spatiotemporal movement model should therefore 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 its 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:

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

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 further developed a 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 perfectly predicting the resighting location, value 0 representing equivalent predictive performance as 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 the performance decay as transition interval increases, the pattern of which is also seen in Fuentes et al. 2023. See supplementary information for the details of fitting the exponential curve. We define the maximum functional distance horizon as the distance value where the regression line intersects with 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 S&T (see supplementary methods and supplementary Fig. 4). We do not define maximum functional time horizon since the models often performed well even beyond the temporal threshold (see results). We also evaluated performance decay using log-likelihood improvement metrics and validated prediction horizons using the same approach.

Empirical validation of biological metrics

To assess the biological plausibility of BirdFlow-generated trajectories, we compared the route statistics of model-simulated and empirical routes. For seven species for which GPS tracking data were available, we extracted track sections during pre-breeding and post-breeding migration phases. Only tracks recorded for more than two weeks and with logging frequency over once a week were included. For each track, we sampled 100 trajectory predictions conditional on the starting time and location and the ending time of the track. We then

summarized the migration speed, route straightness, and number of stopovers of the simulated and observed routes. We defined migration speed as the total distance traveled by the simulated individual within the trajectory, divided by the total length of the time horizon. Route straightness was the beeline distance of the routes divided by the total distance traveled. We defined the number of stopovers as the number of 150-km resolution grid cells in which the individual stayed >7 days. This temporal resolution is predefined and aligned with the eBird S&T product. The route features were calculated using trajr package version 1.5.1 (McLean and Skowron Volponi 2018) in R version 4.5.

Modeling hyperparameters using species traits

We modeled the selected hyperparameters for each species using species-specific morphological traits and distribution range statistics to understand hyperparameter variation across species. We acquired species morphological traits from the AVONET dataset (Tobias et al. 2022). 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 (see supplementary information). 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 (see supplementary Fig. 7 for the remaining variables). 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 regression deterministic coefficients (R^2) based on pooled predictions on out-of-fold datasets. We calculated 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 (Pedregosa et al. 2011) in Python environment 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 surfaces (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).

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 2489 km (95% percentile interval: 242–6395 km; Fig. 3 b-c). Similar results for the log-likelihood improvement metric can be found in supplementary Fig. 5.

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, supplementary Fig. 8, supplementary table 3). 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 fall within the 95% simulated intervals (range: 56%–100% across species), and 44% fall within the 50% simulated intervals (range: 22–58%) (Fig. 4). For route straightness, the averages are 90% (63%–100%) and 49% (11%–64%), respectively (supplementary Fig. 8). For the number of stopovers, these values are 99.59% (97.14%–100%) and 98.44% (94.29%–100%), respectively (supplementary Fig. 8). The results for the number of stopovers are less informative because the coarse weekly temporal resolution reduces the precision of stopover calculations.

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).

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 (Supplementary Fig. 3; 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 (Supplementary Fig. 6a-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; Supplementary Fig. 6d) and lower distance weights (Pearson's r = -0.33; Supplementary Fig. 6f). Higher spatial variation of abundance and narrower range at breeding season were also correlated with higher distance weights (Pearson's r = 0.45, Supplementary Fig. 6e, Supplementary Fig. 7).

Discussion:

Strengths and limitations of BirdFlow predictions

Our validation results show that species-specifically tuned BirdFlow models provide superior information about population-level movement compared to a random sample from species' abundance distributions, a result that holds for >100 species (Fig. 2). BirdFlow models generally outperformed our baseline at forecast horizons of a few thousand kilometers (on average 2449 km). We also observed that most BirdFlow models continued to outperform the baseline for greater than 270 days, which was the maximum time horizon tested (Fig. 3). This robust performance at long time horizons may derive from stationary individual transitions, which are well represented in the transition data.

The decay of model performance along spatial and temporal forecasting horizons aligns with expectations. A possible explanation for the performance decay could be simply 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 short horizon (Finch et al. 2017). For these species, BirdFlow models are not expected to be able to distinguish the movement of individuals in the same spatiotemporal unit (see discussion below). 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.

Our results demonstrate that BirdFlow-simulated routes are generally biologically realistic, align with tracked individual trajectories, and reflect variation at the population level. 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, supplementary Fig. 8). The biases indicate potential space to improve parameterization for those models. An additional possibility is that these biases could reflect sampling bias if the tracked individuals are consistently fast/slow moving compared to the population. Regardless, it should be noted that BirdFlow is intrinsically modeling the synthetic movements of all the individuals that aggregate at the same 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 any "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 individuals.

The value of multi-source data integration

Our results show that integrating multi-source data greatly expands our ability to parameterize species-specific BirdFlow models. For example, banding data alone enabled species-specific tuning for 37.9% of our study species. Species-specific tuning yielded meaningful improvements in performance than hyperparameter transferring, enabling more accurate movement prediction. However, individual-level movement data remain scarce for many small-bodied passerine species. Because of their low body mass, these species are difficult to track with current GPS technology (Bridge et al. 2011, Scarpignato et al. 2023), and banding recovery data will continue to accumulate slowly due to low recovery rate (Arnold et al. 2013). In this context, Motus Wildlife Tracking System represents an especially promising resource for tuning models for small-bodied passerine birds, as its data are accumulating rapidly (Taylor et al. 2017).

Our results demonstrate that the meaningful improvements in model performance can be achieved with limited mark-resight data. Notably, simply increasing the number of transitions does not consistently lead to higher predictive performance across species (Supplementary Fig. 3). This pattern is encouraging because it indicates that a small amount of empirical data can meaningfully improve BirdFlow predictions for a focal species. At the same time, this result may also reflect the limited flexibility of the current BirdFlow model structure: with only three hyperparameters, desirable model performance can often be reached with a modest amount of tuning data.

Transition data also have spatiotemporal biases. Tracking data often concentrate in a few locations rather than spanning the full species distribution (Scarpignato et al. 2023), 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, such biases warrant attention. If this tuning framework were applied to a more complex model structure, bias-control methods or balanced sampling strategies would become important considerations.

Overall, the BirdFlow model's foundation on eBird S&T abundance data, combined with tuning from individual-based multi-source movement data, makes it a comprehensive realization of data integration. 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.

Species variation and the importance of species-specific tuning

We show that the 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 (Hedenström 2008, Linek et al. 2024). 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 (Dufour et al. 2020) and consequently may also be closer in BirdFlow hyperparameter space.

Hyperparameter transfer outcomes are consistent with species-specific clustering patterns (Fig. 5). For example, small passerine birds generally showed low entropy weight, high distance weight, and low distance exponent (lower PC1 values; less spread-out and high movement cost), while large birds in Anseriformes (ducks and geese) showed the opposite (higher PC1 values; more spread-out 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 (Catry et al. 2004, O'Neal et al. 2018, Horton and Deng 2024). We also find that spatially narrow-ranged species with high spatial abundance variation 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, while broad ranged species with low abundance variation show the opposite. This likely reflects that narrow-ranged species are facing stronger migration barriers and having fewer suitable stopover habitats, which elevates the distance cost and forces long jumps rather than small hops (e.g., Studds et al. 2017). Collectively, the results indicate that the three hyperparameters reflect species' functional traits and distributional properties. Despite this evidence, we did not find statistically significant clustering patterns across orders or families, which indicates that within-group hyperparameter variation is comparable to among-group variation, further emphasizing the importance of species-specific tuning.

The evidence of hyperparameter similarity among taxonomic groups opens the door for tuning species lacking any individual-level movement data. Hyperparameters can be transferred from close taxonomic relatives using a taxonomically informed leave-one-out approach, yielding acceptable prediction performance. This taxonomic grouping approach may hold even greater

potential when applied to tuning other biologically relevant metrics, such as migration connectivity, or to optimizing more hyperparameter-intensive model structures.

Continuous model improvement for applications

Because the validation framework is insensitive to both model structure and data type, it opens opportunities for continual refinement of BirdFlow, from adjusting hyperparameters to exploring alternative model formulations. It also provides a way to quantitatively evaluate different structures as more data become available, ensuring that results grow increasingly robust over time. Importantly, understanding which types of transitions are most informative (e.g., long-distance connections between breeding and non-breeding regions versus shorter, within-season transitions) can guide not only future model development but also strategic decisions about where and when to prioritize data collection.

As part of this study, we release the batch of BirdFlow models for 153 North American migratory species (see data statement), representing the first comprehensive set of population-level movement prediction models spanning broad spatial and temporal scales. The released models can be readily applied, extended, and updated as new data become available, offering a flexible foundation for advancing both ecological research and practical applications, including quantifying migratory connectivity, estimating collision risk, forecasting bird strike risk in aviation, modeling pathogen spread in epidemiology, inferring population-specific migration phenology, and guiding conservation strategies for migratory birds. Together, these advances position BirdFlow as a general framework for integrating multi-source data to predict bird movement at continental scales, opening new opportunities to link population ecology with applied management across hundreds of species.

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.

Funding statement: 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.

Conflict of interest statement:
The authors declare no conflict of interest.

Code and data availability:
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.16985511. 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.

Significance statement:
This study provides the first large-scale framework for modeling population-level migration across 153 bird species in North America by integrating GPS, banding, and Motus data into the BirdFlow modeling framework. Building on recent work by our team on BirdFlow (Fuentes et al. 2023) and eBird Status & Trends (Fink et al. 2025), we extend these approaches to produce biologically realistic migration routes and long-distance forecasts that capture spatiotemporal variation at continental scales. The resulting collection of models offers a novel resource for understanding migratory connectivity and advancing ecological research on spatial and temporal dynamics, making this journal an ideal venue for this contribution.

References:

Arnold, T. W., de Sobrino, C. N. and Specht, H. M. 2013. Estimating Annual Survival Rates from Sparse Band-recovery Data: Examples from Migratory Shore and Upland Gamebirds. – U.S. Fish and Wildlife Service.

Bridge, E. S., Thorup, K., Bowlin, M. S., Chilson, P. B., Diehl, R. H., Fléron, R. W., Hartl, P., Kays, R., Kelly, J. F., Robinson, W. D. and Wikelski, M. 2011. Technology on the Move: Recent and Forthcoming Innovations for Tracking Migratory Birds. – BioScience 61: 689–698.

Buchan, C., Franco, A. M. A., Catry, I., Gamero, A., Klvaňová, A., Gilroy, J. J. and Field, R. 2022. Spatially explicit risk mapping reveals direct anthropogenic impacts on migratory birds. – Glob. Ecol. Biogeogr. 31: 1707–1725.

Catry, P., Encarnação, V., Araújo, A., Fearon, P., Fearon, A., Armelin, M. and Delaloye, P. 2004. Are long-distance migrant passerines faithful to their stopover sites? – J. Avian Biol. 35: 170–181.

Dufour, P., Descamps, S., Chantepie, S., Renaud, J., Guéguen, M., Schiffers, K., Thuiller, W. and Lavergne, S. 2020. Reconstructing the geographic and climatic origins of long-distance bird migrations. – J. Biogeogr. 47: 155–166.

Emmenegger, T., Hahn, S., Arlettaz, R., Amrhein, V., Zehtindjiev, P. and Bauer, S. 2016. Shifts in vegetation phenology along flyways entail varying risks of mistiming in a migratory songbird. – Ecosphere 7: e01385.

Finch, T., Butler, S. J., Franco, A. M. A. and Cresswell, W. 2017. Low migratory connectivity is common in long-distance migrant birds. – J. Anim. Ecol. 86: 662–673.

Fink, D., Auer, T., Johnston, A., Ruiz-Gutierrez, V., Hochachka, W. M. and Kelling, S. 2020. Modeling avian full annual cycle distribution and population trends with citizen science data. – Ecol. Appl. 30: e02056.

Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M., Ligocki, S., Robinson, O., Hochachka, W., Jaromczyk, L., Crowley, C., Dunham, K., Stillman, A., Davis, C., Stokowski, M., Sharma, P., Pantoja, V., Burgin, D., Crowe, P., Bell, M., Ray, S., Davies, I., Ruiz-Gutierrez, V., Wood, C. and Rodewald, A. 2024. eBird Status and Trends, Data Version: 2023; Released: 2025.

Flack, A., Aikens, E. O., Kölzsch, A., Nourani, E., Snell, K. R. S., Fiedler, W., Linek, N., Bauer, H.-G., Thorup, K., Partecke, J., Wikelski, M. and Williams, H. J. 2022. New frontiers in bird migration research. – Curr. Biol. 32: R1187–R1199.

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.

Hedenström, A. 2008. Adaptations to migration in birds: behavioural strategies, morphology and scaling effects. – Phil. Trans. R. Soc. B 363: 287–299.

Horton, K. G. and Deng, Y. 2024. Stopover strategies drive potential adaptability under changing environments. – Proc. Natl. Acad. Sci. U.S.A. 121: e2406694121.

Horton, K. G., Buler, J. J., Anderson, S. J., Burt, C. S., Collins, A. C., Dokter, A. M., Guo, F., Sheldon, D., Tomaszewska, M. A. and Henebry, G. M. 2023. Artificial light at night is a top predictor of bird migration stopover density. – Nat. Commun. 14: 7446.

Johnston, A., Rodewald, A. D., Strimas, M., Auer, T., Hochachka, W. M., Stillman, A. N., Davis, C. L., Ruiz, V., Dokter, A. M., Miller, E. T., Robinson, O., Ligocki, S., Jaromczyk, L. O., Crowley, C., Wood, C. L. and Fink, D. 2025. North American bird declines are greatest where species are most abundant. – Science 388: 532–537.

Kramer, G. R., Andersen, D. E., Buehler, D. A., Wood, P. B., Peterson, S. M., Lehman, J. A., Aldinger K. R., Bulluck, L. P., Harding, S., Jones, J., Loegering, J. P., Smalling, C., Vallender, R. and Streby, H. M. 2018. Population trends in Vermivora warblers are linked

to strong migratory connectivity. – Proc. Natl. Acad. Sci. U.S.A. 115: E3192-E3200.
Linek, N., Yanco, S. W., Volkmer, T., Zuñiga, D., Wikelski, M. and Partecke, J. 2024. Migratory lifestyle carries no added overall energy cost in a partial migratory songbird. – Nat. Ecol. Evol. 8: 2286–2296.
McLean, D. J. and Skowron Volponi, M. A. 2018. trajr: An R package for characterisation of animal trajectories. – Ethology 124: 440–448.
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., Taylor, L., Witko, C., Akresh, M. E., Barber, D. R., Bayne, E. M., Beasley, J. C., Belant, J. L., Bierregaard, R. O., Bildstein, K. L., Boves, T. J., Brzorad, J. N., Campbell, S. P., Celis-Murillo, A., Cooke, H. A., Domenech, R., Goodrich, L., Gow, E. A., Haines, A., Hallworth, M. T., Hill, J. M., Holland, A. E., Jennings, S., Kays, R., King, D. T., Mackenzie, S. A., Marra, P. P., McCabe, R. A., McFarland, K. P., McGrady, M. J., Melcer, R., Norris, D. R., Norvell, R. E., Rhodes, O. E., Rimmer, C. C., Scarpignato, A. L., Shreading, A., Watson, J. L. and Wilsey, C. B. 2022. Integrating data types to estimate spatial patterns of avian migration across the Western Hemisphere. – Ecol. Appl. 32: e2679.
O'Neal, B. J., Stafford, J. D., Larkin, R. P. and Michel, E. S. 2018. The effect of weather on the decision to migrate from stopover sites by autumn-migrating ducks. – Mov. Ecol. 6: 23.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. – J. Mach. Learn. Res. 12: 2825–2830.
Plunkett, E. 2025. BirdFlowR: Predict and Visualize Bird Movement. – <https://github.com/birdflow-science/BirdFlowR>
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. and Marra, P. P. 2019. Decline of the North American avifauna. – Science 366: 120–124.
Rushing, C. S., Ryder, T. B., Scarpignato, A. L., Saracco, J. F. and Marra, P. P. 2016. Using demographic attributes from long-term monitoring data to delineate natural population structure. – J. Appl. Ecol. 53: 491–500.
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. and Marra, P. P. 2023. Shortfalls in tracking data available to inform North American migratory bird conservation. – Biol. Conserv. 286: 110224.
Stanley, C. Q., McKinnon, E. A., Fraser, K. C., Macpherson, M. P., Casbourn, G., Friesen, L., Marra, P. P., Studds, C., Ryder, T. B., Diggs, N. E. and Stutchbury, B. J. M. 2015. Connectivity of wood thrush breeding, wintering, and migration sites based on range-wide tracking. – Conserv. Biol. 29: 164–174.
Strimas-Mackey, M., Ligocki, S., Auer, T. and Fink, D. 2025. ebirdst: Access and Analyze eBird Status and Trends Data Products. – <https://CRAN.R-project.org/package=ebirdst>
Studds, C. E., Kendall, B. E., Murray, N. J., Wilson, H. B., Rogers, D. I., Clemens, R. S., Gosbell, K., Hassell, C. J., Jessop, R., Melville, D. S., Milton, D. A., Minton, C. D. T., Possingham, H. P., Riegen, A. C., Straw, P., Woehler, E. J. and Fuller, R. A. 2017. Rapid population decline in migratory shorebirds relying on Yellow Sea tidal mudflats as stopover sites. – Nat. Commun. 8: 14895.
Shaw, A. K. 2020. Causes and consequences of individual variation in animal movement – Mov. Ecol. 8: 12.
Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., Damoulas, T., Dhondt, A. A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J. W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W. M., Iliff, M. J., Lagoze, C., La Sorte, F. A., Merrifield, M., Morris, W., Phillips, T. B., Reynolds, M., Rodewald, A. D., Rosenberg, K.

V., Trautmann, N. M., Wiggins, A., Winkler, D. W., Wong, W.-K., Wood, C. L., Yu, J. and Kelling, S. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. – Biol. Conserv. 169: 31–40.

Taylor, P. D., Crewe, T. L., Mackenzie, S. A., Lepage, D., Aubry, Y., Crysler, Z., Finney, G., Francis, C. M., Guglielmo, C. G., Hamilton, D. J., Holberton, R. L., Loring, P. H., Mitchell, G. W., Norris, D. R., Paquet, J., Ronconi, R. A., Smetzer, J. R., Smith, P. A., Welch, L. J. and Woodworth, B. K. 2017. The Motus Wildlife Tracking System: a collaborative research network to enhance the understanding of wildlife movement. – Avian Conserv. Ecol. 12: art8.

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Schleuning, M., et al. 2022. AVONET: morphological, ecological and geographical data for all birds. – Ecol. Lett. 25: 581–597.

Celis-Murillo, A., Malorodova, M., and Nakash, E. 2022. North American Bird Banding Program Dataset 1960-2022 retrieved 2022-07-14 – U.S. Geological Survey data release <https://doi.org/10.5066/P9BSM38F>.

Zurell, D., Graham, C. H., Gallien, L., Thuiller, W. and Zimmermann, N. E. 2018. Long-distance migratory birds threatened by multiple independent risks from global change. – Nat. Clim. Chang. 8: 992–996.

Figures and tables

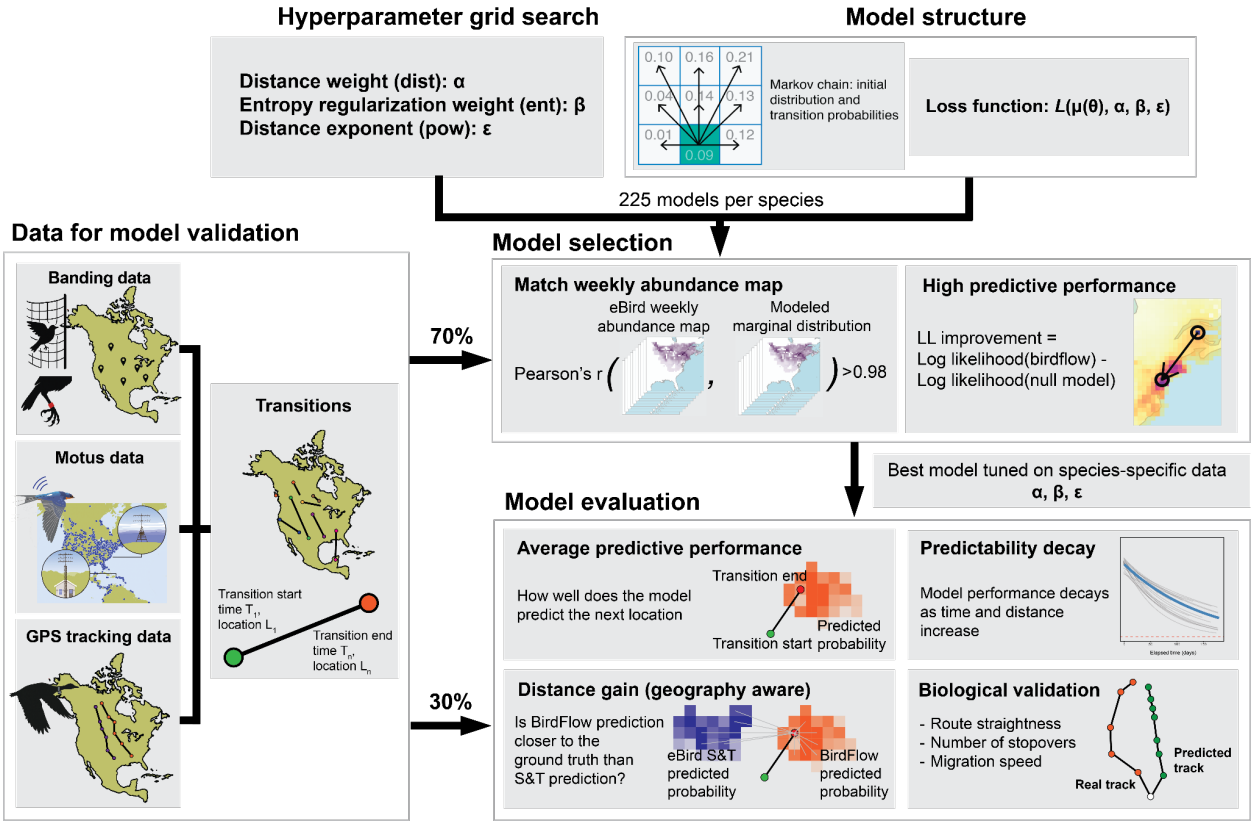


Figure 1. Schematic workflow for the BirdFlow species-specific model tuning and validation framework. The diagram illustrates the main steps of data preparation, model tuning, and validation methods used in this study.

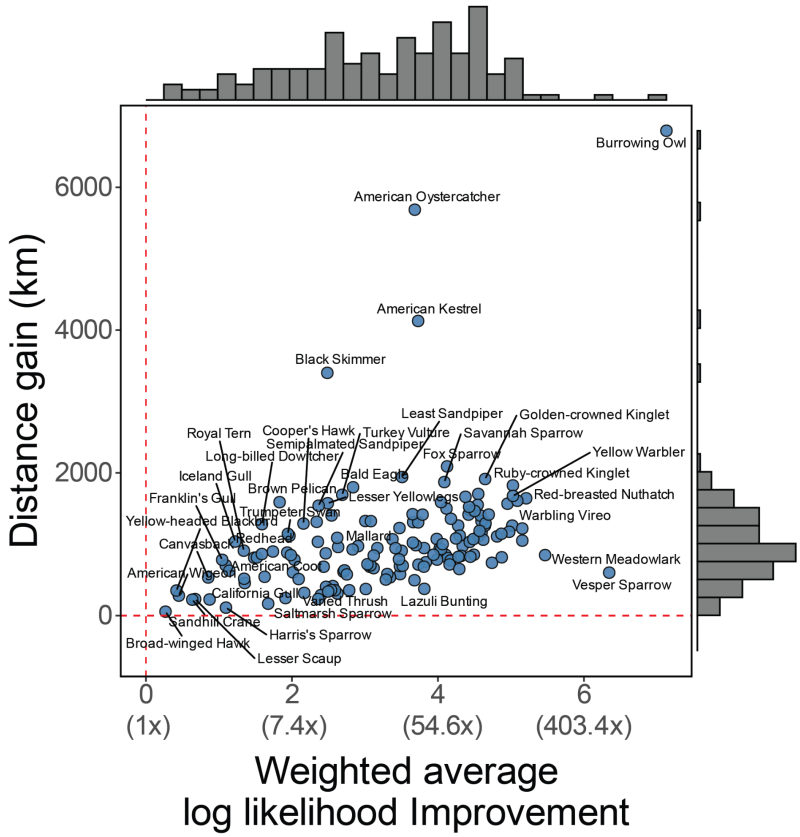


Figure 2. BirdFlow models for all 153 species perform better compared to a random sample from eBird Status and Trends (indicated by the dotted red line). Performance is evaluated by both weighted average log-likelihood improvement (x axis) and distance gain (y axis). Metrics are calculated across all 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).

688 **Table 1.** Taxonomically informed model tunings show improvements over the all-leave-one-out method
689 (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%)

690
691
692
693
694

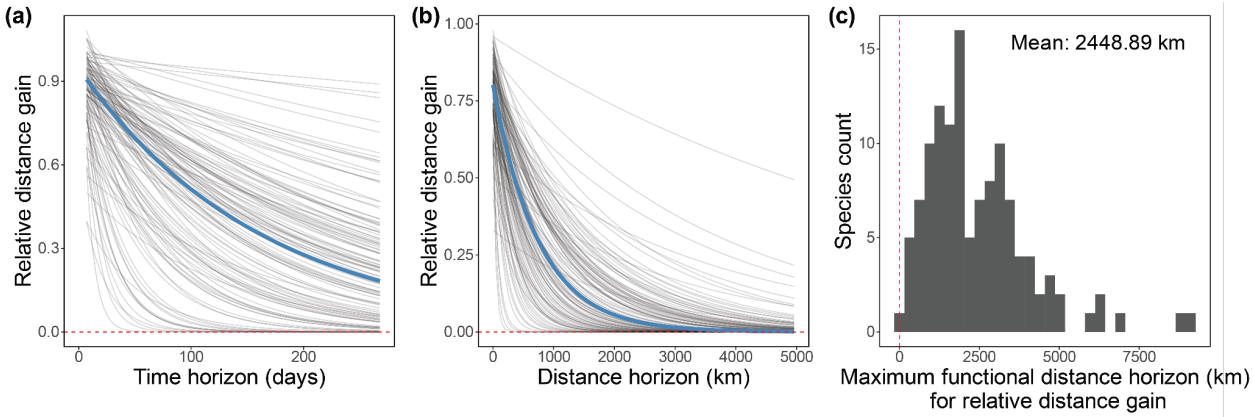


Figure 3. Model performance decays as time and distance horizon increase. Model performance decay is measured as the decrease of relative distance gain as the time horizon (a) and distance horizon (b) extends. Gray lines represent the fitted exponential regression function for each species, and blue 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 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 Supplementary Fig. 4 for schematic diagram of calculation in (c).

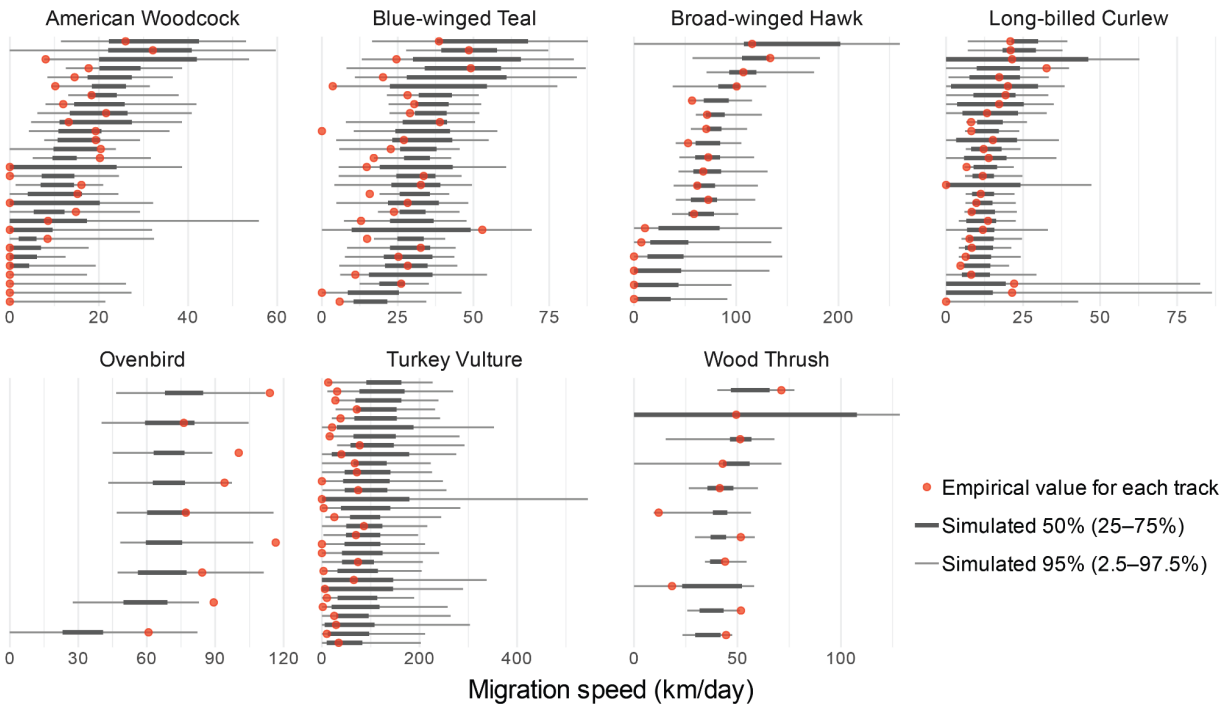


Figure 4. BirdFlow models simulate biologically plausible migration speeds. For each species, we randomly sampled at most 30 tracks from pre-breeding or post-breeding migration seasons. Red points indicate the empirical migration speeds calculated from GPS tracks and gray bars indicate 95% and 50% percentile intervals of the simulated migration speed. A quantitative summary of the figure can be found in supplementary table 3.

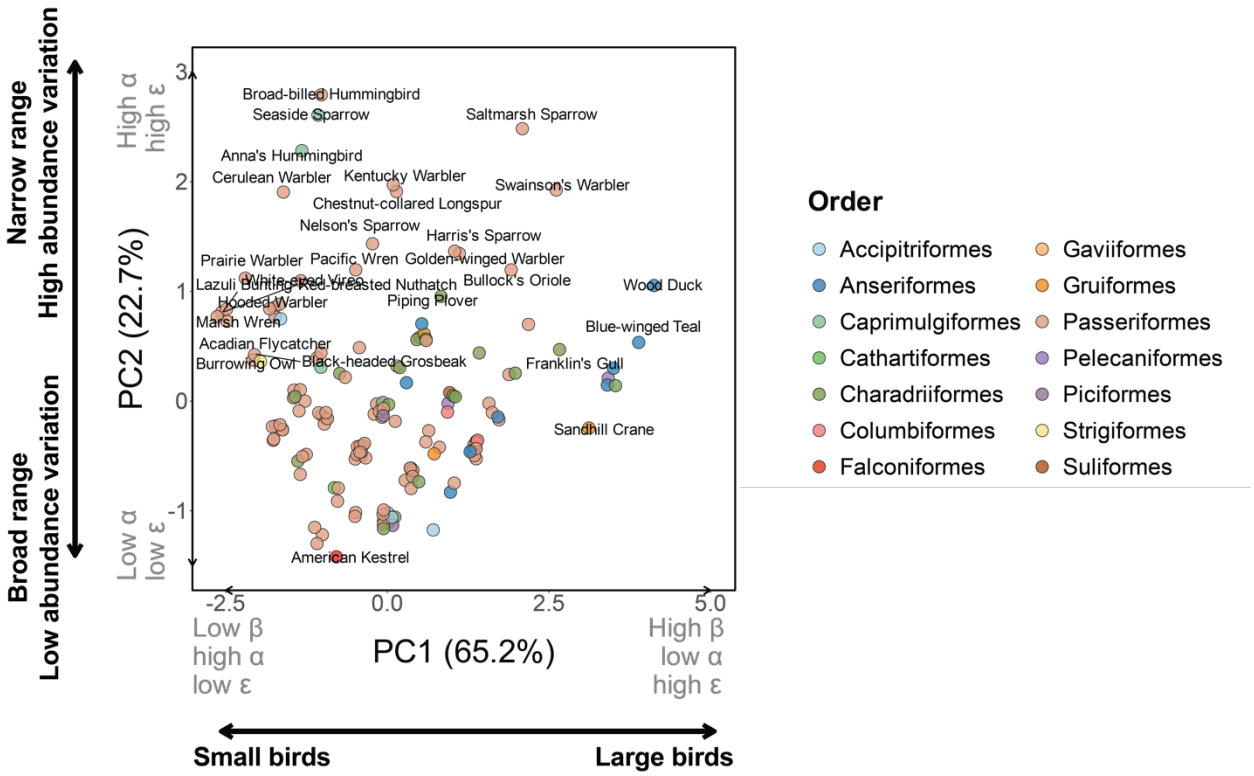


Figure 5. Associations between selected hyperparameters and species traits. Principal component analysis results for the hyperparameter patterns across species. Color denotes the order of the species. Gray text indicates the relative PC loadings of entropy weight β, distance weight α, and distance exponent 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.

Supplemental Information for:

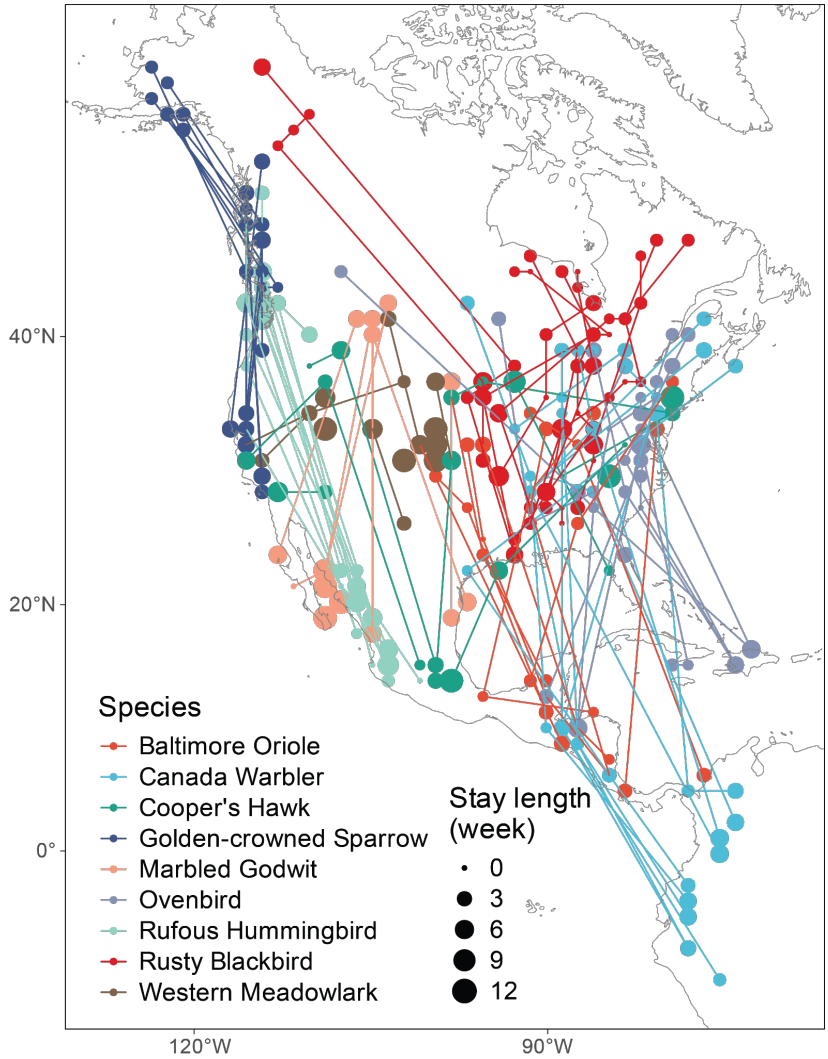
Population-level migration modeling of North America's birds through data integration with BirdFlow

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

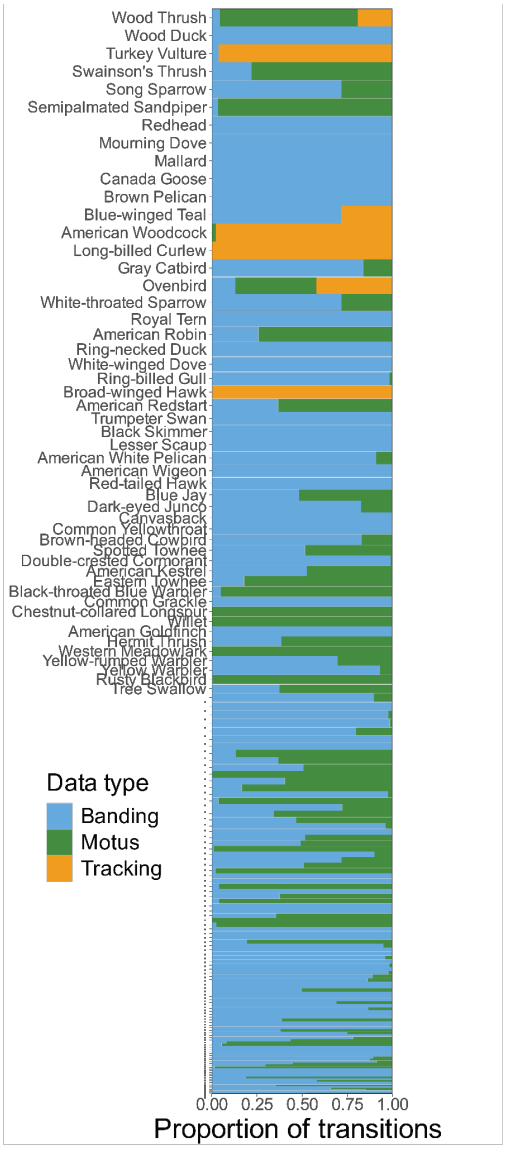
Table of Contents:

Supplementary Figure 1	Page 1
Supplementary Figure 2	Page 2
Supplementary Figure 3	Page 3
Supplementary Figure 4	Page 4
Supplementary Figure 5	Page 5
Supplementary Figure 6	Page 6
Supplementary Figure 7	Page 7
Supplementary Figure 8	Page 8
Supplementary Table 1	Page 9
Supplementary Table 2	Page 10 – 14
Supplementary Table 3	Page 15
Supplementary Methods	Page 16 – 18
References	Page 18

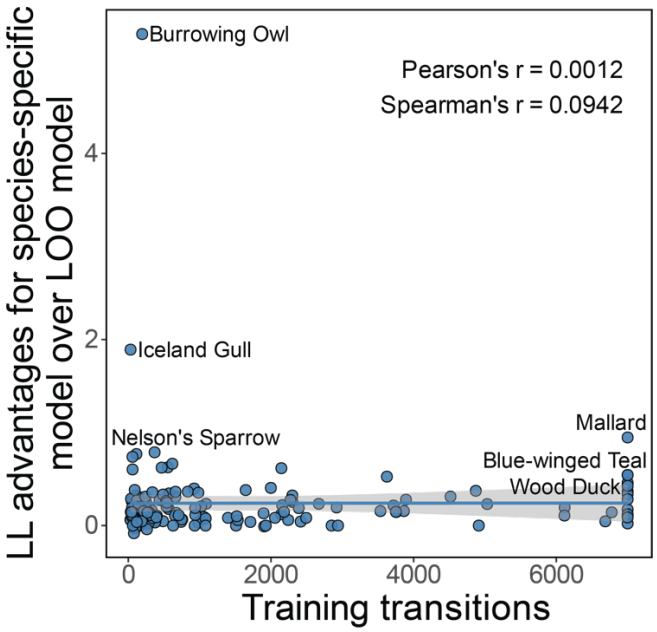
Supplementary Figures



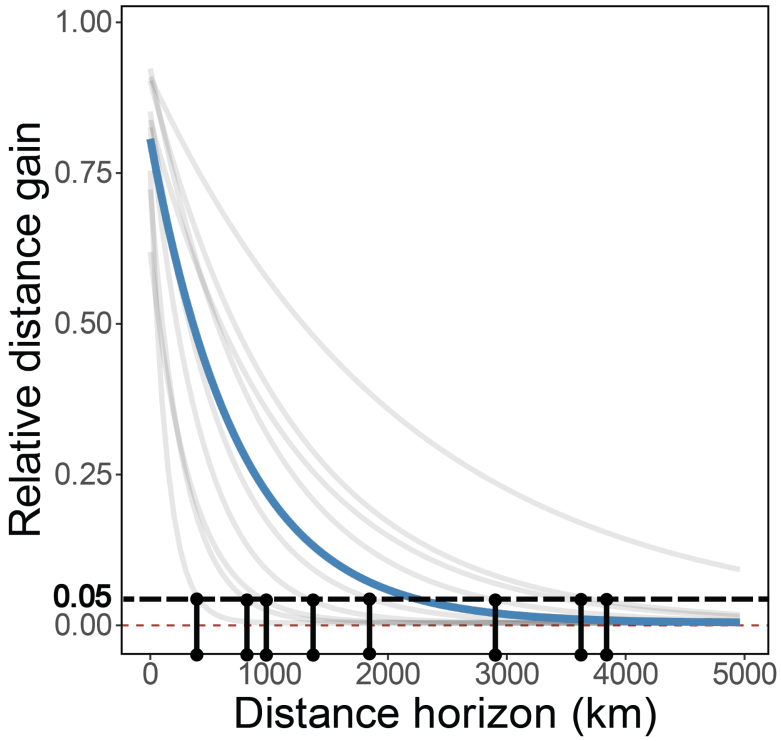
Supplementary Figure 1. The BirdFlow-simulated spring migration routes for 9 migratory species. Stay length indicates the inferred total length of stay at the stopover sites in weekly temporal resolution, with zero stay length indicating non-stopping or stopover shorter than one week. 10 routes were simulated for each species.



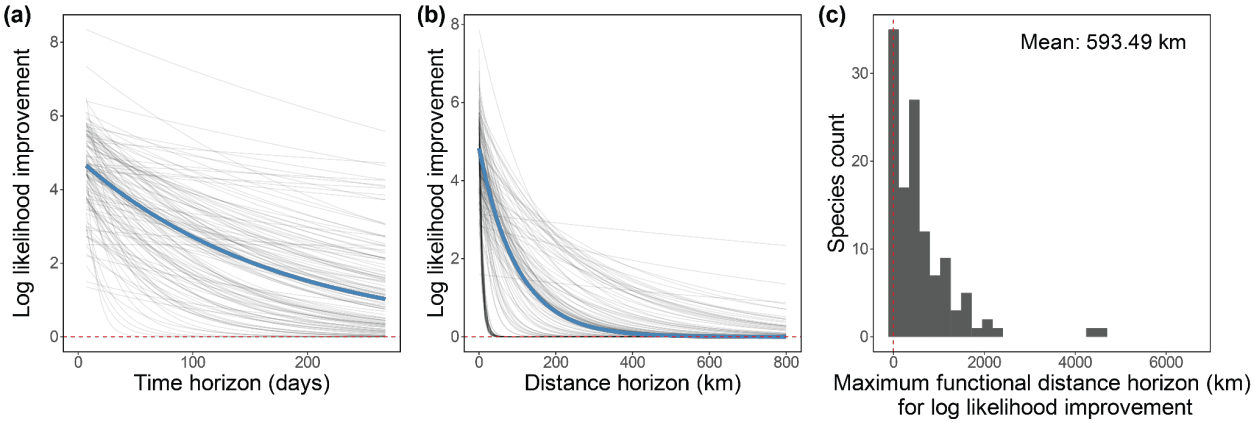
Supplementary Figure 2. The transition data composition for each species, colored by data sources. The height of each block is proportional to the square root of total transition count available for the species. The width of the block represents the proportion of transitions from each data source for each species. The top 50 species with most abundant transition data are labeled with their eBird English common name.



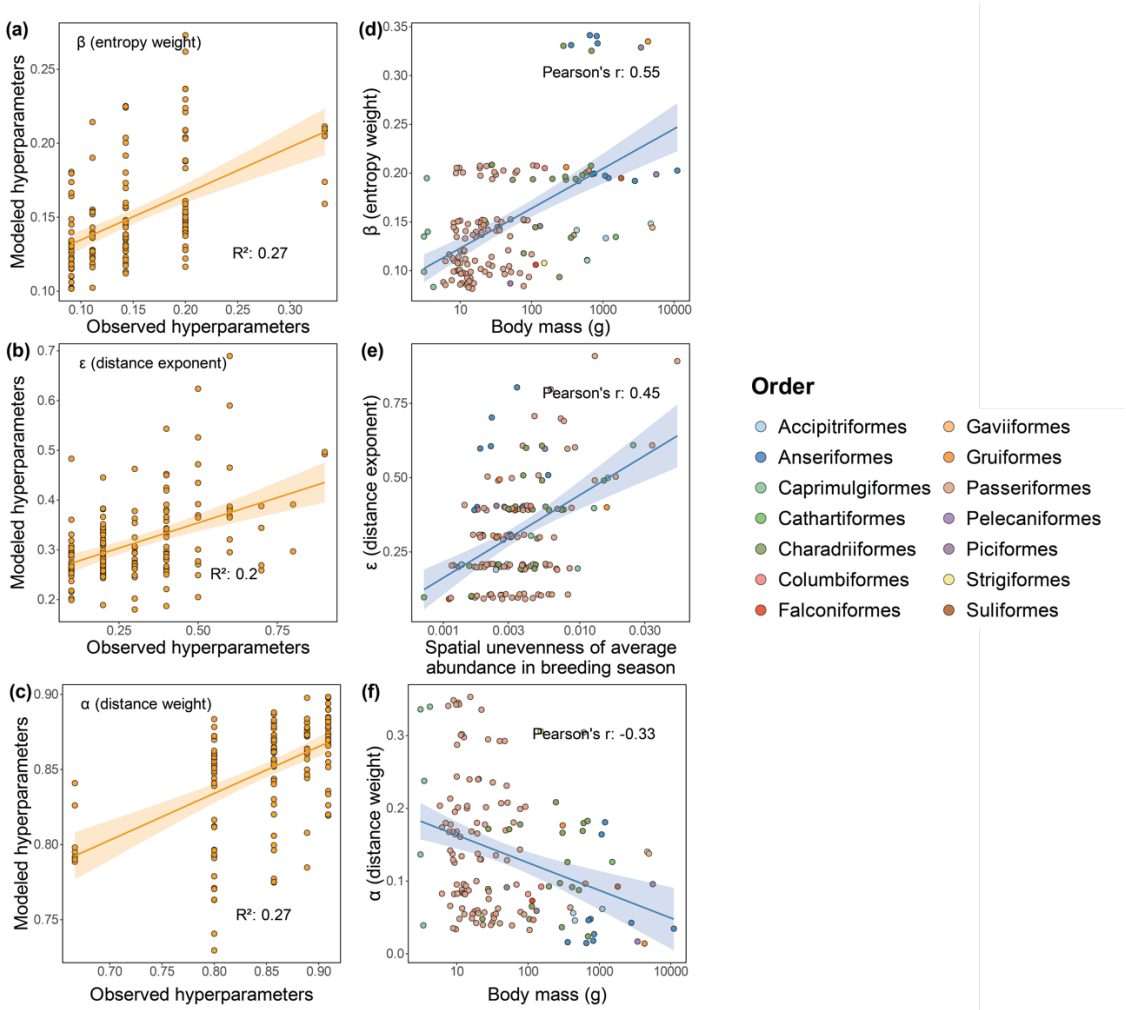
Supplementary Figure 3. Lack of correlation between the improvements of species-specifically tuned models and the amount of training transitions. Although species-specific tuning improves model performance compared to All-LOO baselines, more training data does not necessarily lead to higher log-likelihood improvement across species. Each data point represents a best model for each species tuned on species-specific data.



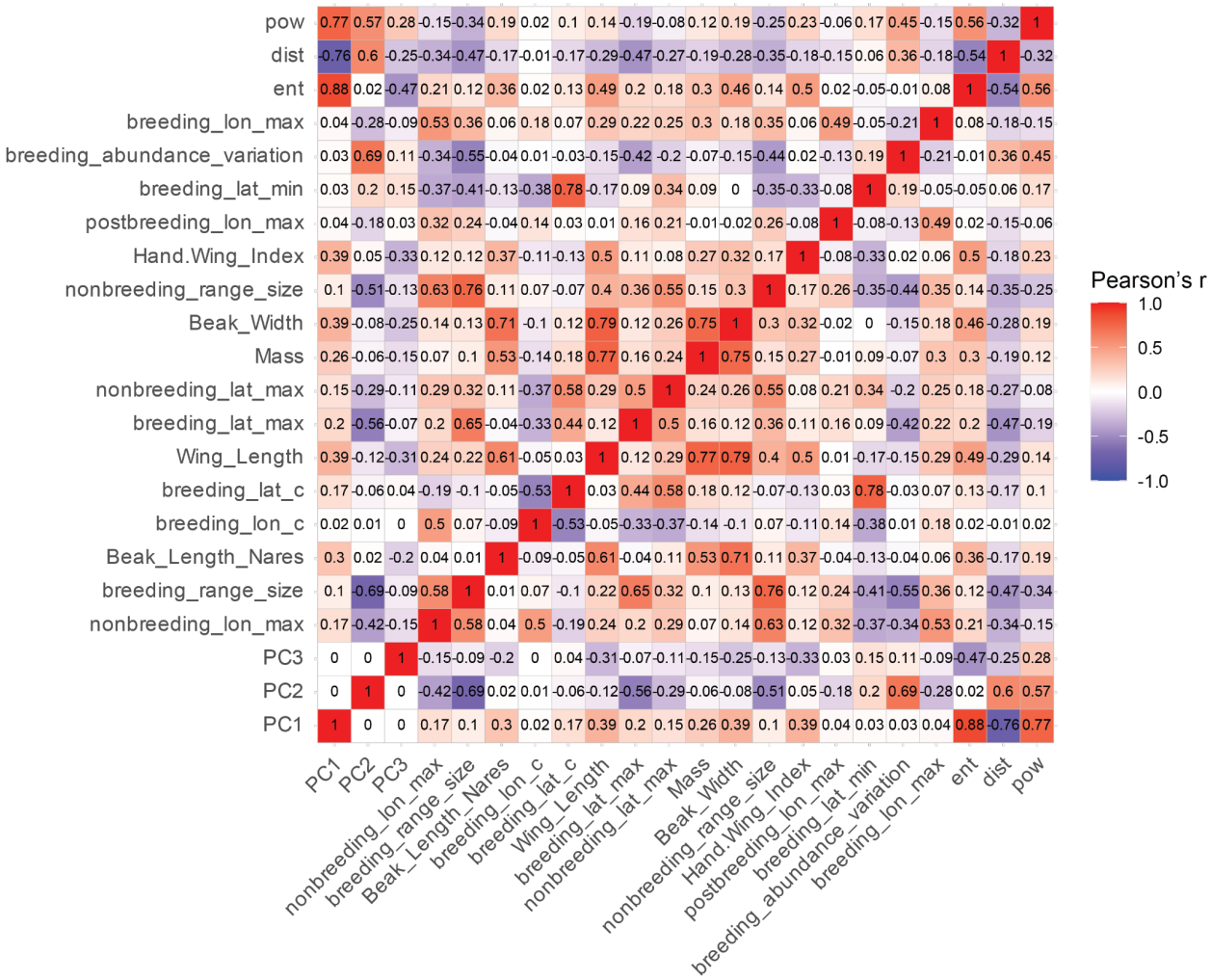
Supplementary Figure 4. A schematic diagram of how the maximum functional forecasting horizon is calculated. We set a threshold of 0.05 for both relative distance gain and log-likelihood improvement, with values below the threshold being considered as BirdFlow models having no advantages over the baseline. 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.



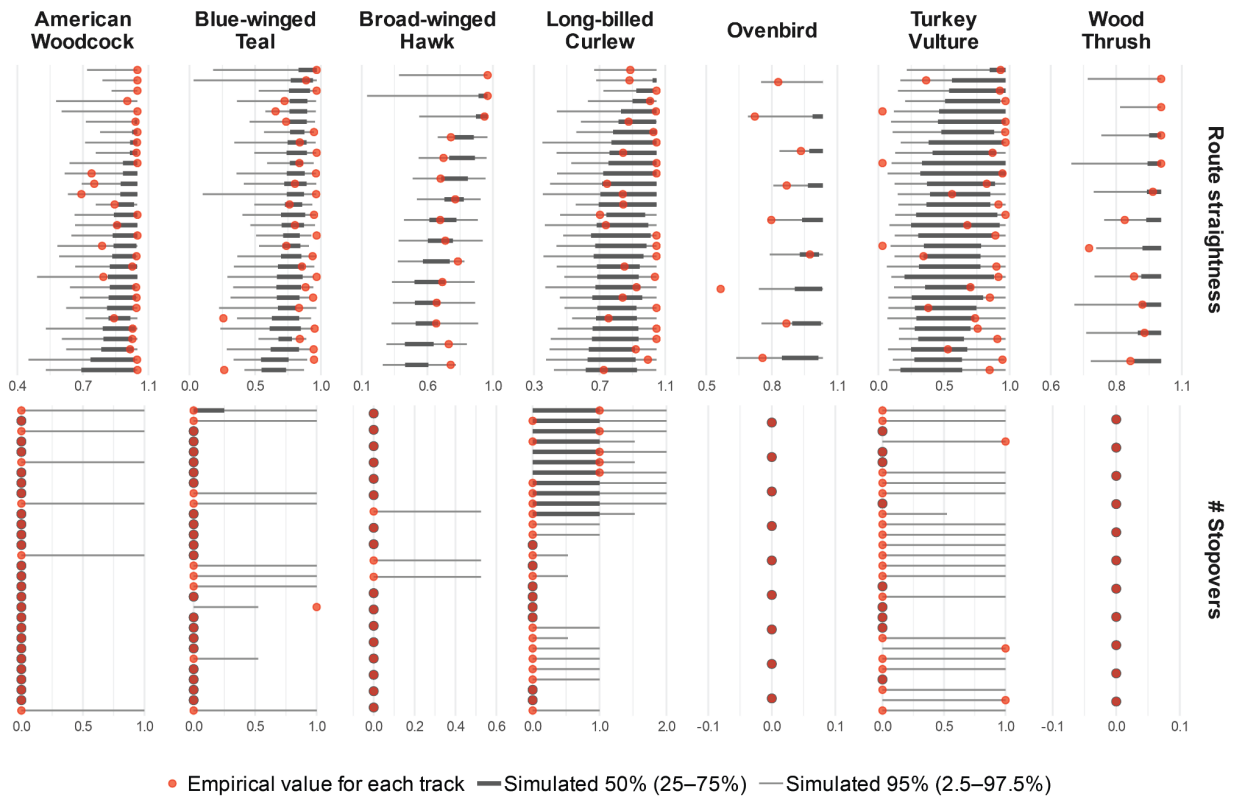
Supplementary Figure 5. Model performance decays as time and distance horizon increase. Model performance decay is measured as the decrease of log-likelihood as the time horizon (a) and distance horizon (b) extends. Gray lines represent the fitted exponential regression function for each species, and 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.



Supplementary Figure 6. 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.



Supplementary Figure 7. 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.



Supplementary Figure 8. 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 supplementary table 3.

Supplementary Tables

Supplementary Table 1. 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

824 **Supplementary Table 2.** 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

Supplementary Table 3. 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 occur 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:

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.