

---

# Modeling Bird Migration by Disaggregating Population Level Observations

---

Miguel Fuentes\*<sup>1</sup> Benjamin Van Doren\*<sup>2</sup> Daniel Sheldon<sup>1</sup>

## Abstract

Birds are shifting migratory routes and timing in response to climate change, but modeling migration to better understand these changes is difficult. Some recent work leverages fluid dynamics models, but this requires individual flight speed and directional data which may not be readily available. We developed an alternate modeling method which only requires population level positional data and use it to model migration routes of the American Woodcock (*Scolopax minor*). We use our model to sample simulated bird trajectories and compare them to real trajectories in order to evaluate the model.

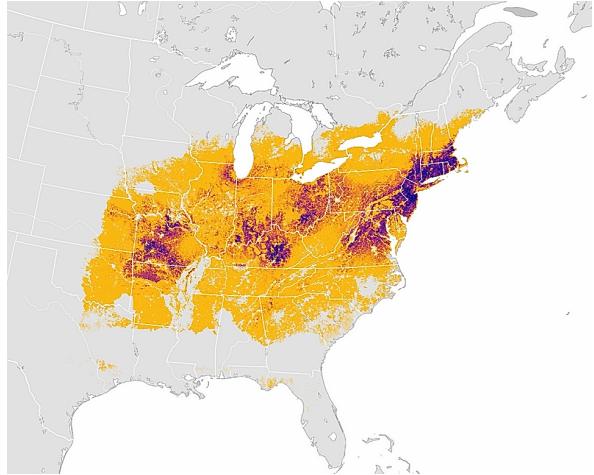


Figure 1. Abundance map from eBird Project for the American Woodcock

## 1. Introduction

Bird migration is an important ecological phenomenon. Migrating birds act as long distance dispersal agents for many other species of plants, invertebrates, and microbes and serve many other ecological functions (Viana et al., 2016). In recent decades, the migratory timing of migratory birds has shifted on a continental scale and climate change is an important factor in this change (Horton et al., 2020; Youngflesh et al., 2021). In order to fully understand these changes, we need effective methods for modeling bird migration.

There are various sources of data that can be used to model migration. Individual bird tracks are a useful source of data but they are not generally available in large quantities and they are expensive to collect. Weather radar can be processed in order to infer bird density, flight speed, and flight direction, and this kind of data has been used recently in a fluid dynamics model (Nussbaumer et al., 2021). However, processing the radar data can be difficult and it does not identify the species of bird. One of the largest sources of data is eBird, which is a citizen science project that aggregates the observations of recreational birders (Sullivan et al.,

2009). This data is readily available on a species level but it does not include flight speed or direction, so the dynamics approaches are not applicable to this data. Our method uses the abundance maps generated by the eBird project to infer a distribution of bird trajectories.

## 2. Data

We use weekly modeled species abundance estimated derived from eBird data (Fink et al., 2020) to infer population movement over time. The abundance model takes into account sources of bias in the citizen science data to produce a weekly estimate of abundance for many bird species. Abundance refers to the number of birds of a species that an observer will see at a location within a given amount of time. An example of one of these weekly abundance maps for the American Woodcock (*Scolopax minor*) can be seen in Figure 1.

We treat a bird’s position during these weekly snapshots as discrete random variables  $X_1, \dots, X_T$  over the grid cells in the map. In order to use the abundance model as a “ground truth” distribution, we aggregate the abundance to a coarser grid and normalize it so that it sums to one. We will refer to these “ground truth” distributions as  $p^*(X_t)$  for the weeks  $t \in [1, \dots, T]$ .

\*Equal contribution <sup>1</sup>University of Massachusetts, Amherst  
<sup>2</sup>Cornell University. Correspondence to: Miguel Fuentes <mmfuentes@umass.edu>.

### 3. Method

We present a fully differentiable parameterization of the single time-step and transition marginals of an arbitrary Markov chain, and develop a loss function of those marginals to infer the flow of birds through different spatial locations.

#### 3.1. Model

In order to simplify the migration model, we will assume that a bird's trajectory is Markovian. This means that the bird's position one week only depends on the position the week before. While this is not a perfect assumption, parameterizing the full distribution  $p(X_1, \dots, X_T)$  without this assumption would require exponentially many parameters in  $T$ . So, we parameterize the model as follows:

Initial Parameters:  $Z^{(1)} \in \mathbb{R}^n$ ,

Transition Parameters:  $Z^{(t,t+1)} \in \mathbb{R}^{n \times n}$ .

Here,  $n$  is the number of cells in the map. In order to go from unconstrained parameters to a distribution we use the softmax function.

$$\sigma(\mathbf{z})_i = \frac{\exp(\mathbf{z}_i)}{\sum_{j=1}^n \exp(\mathbf{z}_j)}$$

This ensures the entries are non-negative and sum to one. Applying it to our parameters gives us the following distributions:

$$p_Z(X_1 = i) = \sigma(Z^{(1)})_i,$$

$$p_Z(X_{t+1} = j | X_t = i) = \sigma(Z_i^{(t,t+1)})_j.$$

Note that  $(Z_i^{(t,t+1)})$  refers to the  $i$ 'th row of the matrix  $Z^{(t,t+1)}$ . It will be convenient to represent these distributions as vectors and matrices:

$$\boldsymbol{\mu}_t(i) = p_Z(X_t = i),$$

$$\mathbf{T}_{t,t+1}(i, j) = p_Z(X_{t+1} = j | X_t = i).$$

From these distributions, we can infer the weekly marginal distributions and the joint distributions over subsequent weeks from the model.

$$\boldsymbol{\mu}_t^T = \boldsymbol{\mu}_1^T \prod_{k=1}^{t-1} \mathbf{T}_{k,k+1},$$

$$\begin{aligned} \mathbf{M}_{t,t+1}(i, j) &= p_Z(X_t = i, X_{t+1} = j) \\ &= p_Z(X_t = i) p_Z(X_{t+1} = j | X_t = i) \\ &= \boldsymbol{\mu}_t(i) \mathbf{T}_{t,t+1}(i, j). \end{aligned}$$

#### 3.2. Optimization

Simply matching the ground truth distribution is an under-specified problem. There are many transition parameters

which could match the ground truth distribution and some of them do not correspond to realistic bird behavior. So, to learn this model we will minimize an objective with three components: the mean squared error  $\mathcal{S}$  between the inferred marginal distributions and the ground truth marginal distributions, a distance penalty  $\mathcal{B}$ , and the entropy  $\mathcal{H}$ .

Minimizing the mean squared error between the ground truth distributions  $\boldsymbol{\mu}_t^*$  and the model predictions  $\boldsymbol{\mu}_t$  ensures that the model reflects the abundance data from eBird.

$$\mathcal{S} = \frac{1}{nT} \sum_{t=1}^T \|\boldsymbol{\mu}_t^* - \boldsymbol{\mu}_t\|_2^2.$$

We need a term to reflect the fact that birds avoid making unnecessary large-scale movements. So, we encode the distance between two cells on the map in the distance matrix  $D$  and penalize birds for the distance they travel. We refer to this as the distance penalty.

$$\mathcal{B} = \frac{1}{n^2(T-1)} \sum_{t=1}^{T-1} \sum_{i=1}^n \sum_{j=1}^n p_Z(X_t = i, X_{t+1} = j) D(i, j),$$

We also encourage the model to avoid solutions which are overly restrictive. We do this by taking into account the entropy of the predicted joint distribution  $p_Z(X_1, \dots, X_T)$ .

$$\begin{aligned} \mathcal{H} &= \sum_{t=1}^{T-2} (\mathcal{H}(X_t, X_{t+1}) - \mathcal{H}(X_{t+1})) \\ &\quad + \mathcal{H}(X_{T-1}, X_T) \\ \mathcal{H}(X_t) &= \sum_{i=1}^n -p_Z(X_t = i) \log p_Z(X_t = i) \\ \mathcal{H}(X_t, X_{t+1}) &= \sum_{i=1}^n \sum_{j=1}^n -p_Z(X_t = i, X_{t+1} = j) \\ &\quad \cdot \log p_Z(X_t = i, X_{t+1} = j) \end{aligned}$$

Finally, we combine these components into one loss term  $\mathcal{L}$

$$\mathcal{L}(Z) = \theta_s \mathcal{S} + \theta_b \mathcal{B} - \theta_h \mathcal{H}$$

Here the weights  $\theta$  are hyper-parameters which we will choose. Note that the entropy terms is subtracted because we minimize the loss and we want to maximize entropy. We minimize this objective by using gradient descent, specifically the Adam optimizer (Kingma & Ba, 2014). The gradients are calculated using automatic differentiation.

#### 3.3. Connection to Collective Graphical Models

Our model is inspired by collective graphical models (CGMs) (Sheldon & Dietterich, 2011) and solves a similar

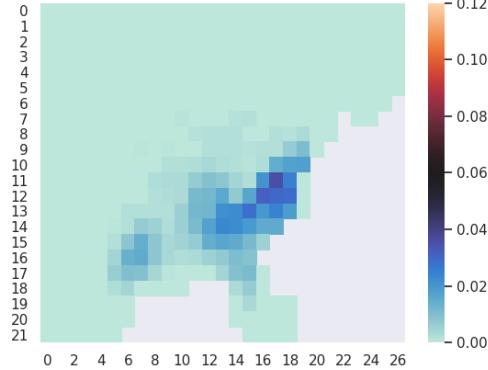


Figure 2. Learned Initial Distribution

optimization problem but is simplified to not attempt a fully generative model. It is instead oriented toward: (1) practical gradient descent optimization on modern deep learning frameworks, (2) effective modeling of the domain application. Our ability to model bird trajectories is substantially more mature than prior applications of CGMs.

## 4. Experiments

We applied our model to the abundance data for the American Woodcock. We chose this species because [Moore et al. \(2021\)](#) recently released individual American Woodcock tracks via the Movebank data repository ([Kranstauber et al., 2011](#)). The availability of individual tracks simplifies the process of evaluating the model. The model was implemented using the JAX package for automatic differentiation and the Optax package for optimization ([Bradbury et al., 2018](#); [Hessel et al., 2020](#)).

After training until convergence, the model had a final loss value of about 1.611. We can see the learned distribution for the first week in Figure 2. However, looking at the weekly distributions cannot indicate how well the model has inferred movement. We can evaluate the quality of the inferred movement by sampling trajectories.

### 4.1. Sampling

Once the parameters are learned, we can sample a trajectory  $x_1, \dots, x_T$  from the model by sampling from the initial distribution and each of the conditional distributions.

$$\begin{aligned} x_1 &\sim p_Z(X_1), \\ x_2 &\sim p_Z(X_2|X_1 = x_1), \\ &\vdots \\ x_T &\sim p_Z(X_T|X_{T-1} = x_{T-1}). \end{aligned}$$

An example sample drawn from the model can be seen in Figure 3. Note that some weeks the bird will not move,

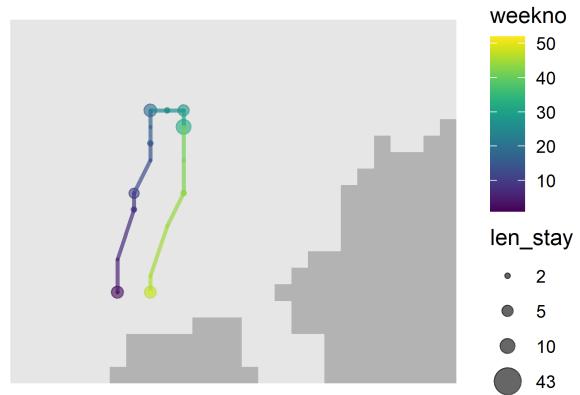


Figure 3. American Woodcock Trajectory Generated by the Model

the length of the bird’s stay in one particular location is indicated by the size of the bubble.

The large scale trends in the sampled trajectories can be seen in Figure 4. Here, we aggregated 400 sampled trajectories and split them across the spring and fall seasons. This shows that the model has learned trajectories which follow the proper seasonal patterns and proper positioning for this species.

### 4.2. Comparing to Real Tracks

The fact that the predicted distribution has the proper position according to the ground truth distribution and the proper seasonal direction is encouraging, but it does not show that the sampled trajectories will be similar to real trajectories. In order to show that, we need to look at true trajectories.

The positions from the tracks from [Moore et al. \(2021\)](#) were not sampled once a week, so we had to line up the timescale with our model. Any birds which were tracked for more than one year were split into several trajectories, one for every year. For each trajectory we recorded the start week  $t_1$  as the week of the year when the first position was recorded and the position  $l_{t_1}$  as the map cell corresponding to the coordinates of the bird at that time. Then, for each subsequent week  $t \in [t_2, \dots, t_F]$  we set  $l_t$  to be the grid cell corresponding to the latest coordinates which were observed that week. Some of the observed trajectories only lasted a few weeks and some of them lasted an entire year.

In order to compare the true trajectories to the sampled trajectories we sampled trajectories starting at the same time and place as a real trajectory. To do this we set the sampled trajectories initial position  $x_{t_1} = l_{t_1}$  and then sequentially sampled  $x_t \sim p_Z(X_t|X_{t-1} = x_{t-1})$  for  $t \in [t_2, \dots, t_F]$ . We plotted some of the real trajectories along with 50 sampled tracks, this can be seen in Figure 5.

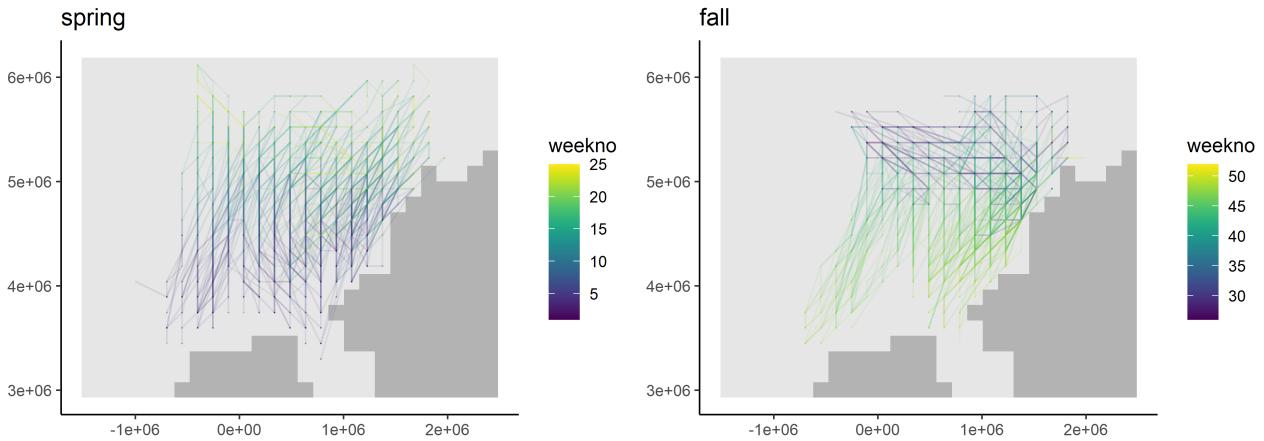


Figure 4. Aggregated American Woodcock Trajectories Generated by the Model

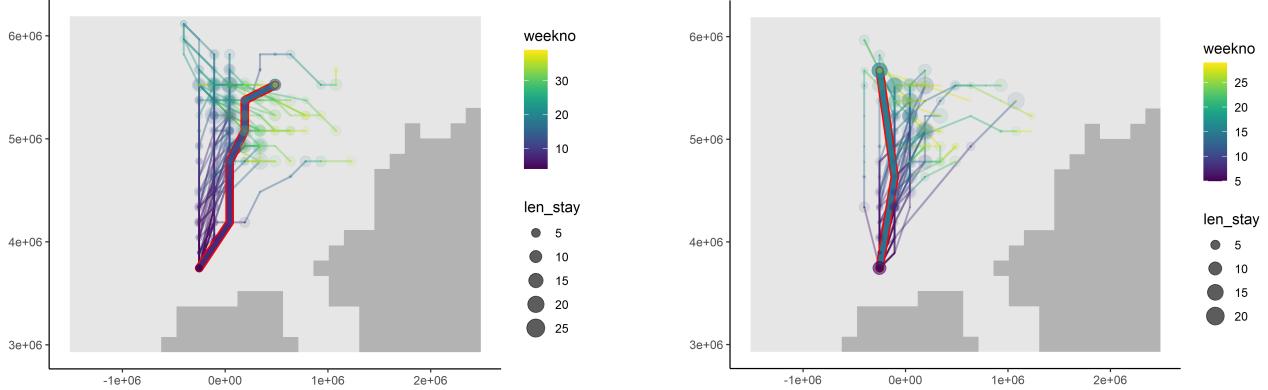


Figure 5. Real American Woodcocks Compared to Model Samples - True Track Outlined in Red

## 5. Discussion

Based on the comparison of the sampled trajectories to the real trajectories, it seems that the model does a good job of positional forecasting. The sampled trajectories and the real trajectories cover a similar distance and travel in a similar direction. Another positive indicator is that the real trajectories seem to lie within the “spread” of the samples. This shows that the model is not learning trajectories which are too restrictive.

However, there are some ways in which the model’s sampled trajectories differ from the real ones. The model’s trajectories tend to move more often than true trajectories which tend to consist of just a few large movements. Also, there is a tendency, because of the Markovian assumption, for some of the year long sampled trajectories to start and end in different locations whereas the true trajectories almost

always end very close to where they start.

Our method can be a useful new tool for scientists trying to investigate changes in the migration patterns of individual species. By using the learned distribution from our model it may be possible to infer behavioral changes with the use of fewer real tracks.

## References

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. JAX: composable transformations of Python+NumPy programs, 2018. URL <http://github.com/google/jax>.

Fink, D., Auer, T., Johnston, A., Ruiz-Gutierrez, V., Hochachka, W. M., and Kelling, S. Modeling avian full annual cycle distribution and population trends with citi-

---

zen science data. *Ecological Applications*, 30(3):e02056, 2020.

Hessel, M., Budden, D., Viola, F., Rosca, M., Sezener, E., and Hennigan, T. Optax: composable gradient transformation and optimisation, in *jax!*, 2020. URL <http://github.com/deepmind/optax>.

Horton, K. G., La Sorte, F. A., Sheldon, D., Lin, T.-Y., Winner, K., Bernstein, G., Maji, S., Hochachka, W. M., and Farnsworth, A. Phenology of nocturnal avian migration has shifted at the continental scale. *Nature Climate Change*, 10(1):63–68, 2020.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M., and Kays, R. The movebank data model for animal tracking. *Environmental Modelling & Software*, 26(6):834–835, 2011.

Moore, J. D., Andersen, D. E., Cooper, T., Duguay, J. P., Oldenburger, S. L., Al Stewart, C., and Krementz, D. G. Migration phenology and patterns of american woodcock in central north america derived using satellite telemetry. *Wildlife Biology*, 2021(1):wlb–00816, 2021.

Nussbaumer, R., Bauer, S., Benoit, L., Mariethoz, G., Liechti, F., and Schmid, B. Quantifying year-round nocturnal bird migration with a fluid dynamics model. *bioRxiv*, pp. 2020–10, 2021.

Sheldon, D. R. and Dietterich, T. G. Collective graphical models. In *NIPS*, pp. 1161–1169. Citeseer, 2011.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., and Kelling, S. ebird: A citizen-based bird observation network in the biological sciences. *Biological conservation*, 142(10):2282–2292, 2009.

Viana, D. S., Santamaría, L., and Figuerola, J. Migratory birds as global dispersal vectors. *Trends in Ecology & Evolution*, 31(10):763–775, 2016. ISSN 0169-5347. doi: <https://doi.org/10.1016/j.tree.2016.07.005>. URL <https://www.sciencedirect.com/science/article/pii/S0169534716301136>.

Youngflesh, C., Socolar, J., Amaral, B. R., Arab, A., Guralnick, R. P., Hurlbert, A. H., LaFrance, R., Mayor, S. J., Miller, D. A., and Tingley, M. W. Migratory strategy drives species-level variation in bird sensitivity to vegetation green-up. *Nature Ecology & Evolution*, pp. 1–8, 2021.