Coarse Models for Bird Migrations Using Clustering and Non-Stationary Markov Chains

Nitin Jain and Bistra Dilkina

School of Computational Science and Engineering, College of Computing
Georgia Institute of Technology, Atlanta, GA
email:{nitin.jain1690, bdilkina6}@gatech.edu

Abstract

While great strides have been made in collecting presence data and developing accurate species distribution models, much less is known about the migratory process that guides the spatio-temporal changes in distributions for migrating species, especially birds. In this work, we address a challenging inference task, where given only aggregate and noisy data of the volume of birds for each spatial pixel and time window, we predict the likely transition links with their associated probabilities. We propose a framework to build such migration networks for different bird species and present a real world example of constructing a network using our approach.

Introduction

Understanding migration behavior and migratory patterns for bird species is critical to mitigating risks to bird populations caused by factors like climate change, urban expansion etc. and facilitating efficient conservation efforts. Obtaining data on bird migrations is challenging and data collected using ringed birds has been the most popular source of migration data. Recent work by (Ambrosini et al. 2014) uses ringing data to fit Conditional Autoregressive Models to assess changes in migration over different time periods.

Data based on observations at field locations is much easier to obtain and allows much larger amounts of data to be collected. But it is harder to infer migratory patterns from such aggregated data. The Spatio-Temporal Exploratory Model (STEM) (Fink et al. 2010) uses aggregated data on millions of bird observations from the *eBird* database (Sullivan et al. 2009) to provide fine-scale raster predictions across US for weekly presence of a bird species. These predictions provide some indirect insight into migration patterns for that species on a weekly scale.

In this paper, we propose a framework to infer actual likely movement segments for species using the STEM predictions. Similar work has been done by (Elmohamed, Kozen, and Sheldon 2007) and (Sheldon and Dietterich 2011) where variants of graphical models are used to infer migratory routes using synthetic and eBird datasets. We construct a migration network by identifying nodes as spatio-

temporal clusters and computing edges using variations of Markov chains. We present our results on data for the Tree Swallow (Tachycineta bicolor) bird species from the STEM species distribution model.

Methodology

Our framework goes through the following pipeline to get from STEM predictions to a migration network,

1. Define migration regions (K-means clustering)

2. Determine time-varying abundance levels in regions

3. Infer time-varying transitions model (Markov chains)

4. Measure accuracy for different models and interpret their results

Data

We use weekly presence scores (STEM predictions) for Tree Swallow on points sampled from a 3×3 km grid across US. We model the fall migration of tree swallows covering a period from early August to end of December for the year 2011. We restrict our study area to Eastern US extending from a latitude of $25°$ to $50°$ and longitude of $-97°$ to $-67°$. To discard points that are clearly not part of the Tree Swallow fall migratory range, we filter the raw STEM data by removing all points which lie below a weekly presence threshold (80 percentile in our case) for every week. Figure 1 shows the points that remain after filtering.
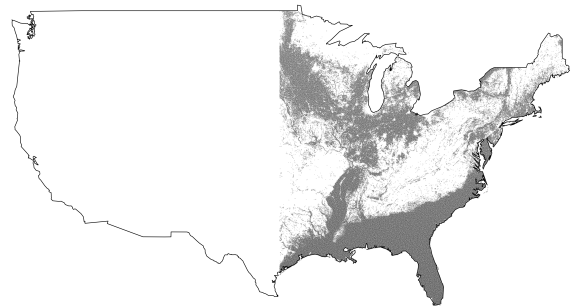

Figure 1: STEM data after threshold

Migration Regions

We use K-means clustering (MacQueen and others 1967) to partition the STEM locations into migration regions for tree swallows. Each STEM location i is associated with a feature vector $(x_i, y_i, \rho_i^1, ..., \rho_i^T)$ corresponding to its spatial coordinates x_i, y_i (longitude and latitude) and the weekly presence scores at i across the T weeks. A custom distance function is used for the clustering to capture temporal correlation as well as spatial coherence between STEM locations. For two locations i and j, the distance function is given by:

$$\mathcal{D}_{ij} = w\mathcal{D}_{ij}^{sp} + \mathcal{D}_{ij}^{tm}$$

$$\mathcal{D}_{ij}^{tm} = \left(1 - \frac{\sum_{t=1}^{T}(\rho_i^t - \bar{\rho}_i)(\rho_j^t - \bar{\rho}_j)}{\sqrt{\sum_{t=1}^{T}(\rho_i^t - \bar{\rho}_i)^2}\sqrt{\sum_{t=1}^{T}(\rho_j^t - \bar{\rho}_j)^2}}\right)$$

where $\mathcal{D}_{ij}^{sp}$ represents the Euclidean distance between locations i and j based on their spatial coordinates and $\mathcal{D}_{ij}^{tm}$ represents the Pearson distance for the temporal features. $\bar{\rho}_i$ is the mean of the weekly presence scores at location i. A weight w is used to adjust for the desired spatial coherence of the clusters.

We experiment with different number of clusters and use the gap statistic metric (Tibshirani, Walther, and Hastie 2001) and prior domain knowledge to choose appropriate K value for the number of clusters. We also adjust the weight w to assess spatial coherence. Figure 2 shows clusters obtained using $K = 11$ and $w = 4$.
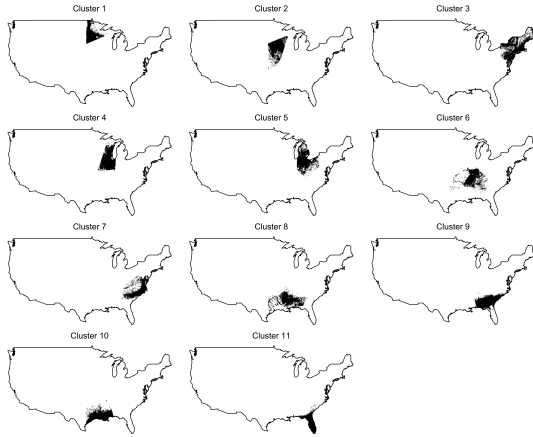

Figure 2: Spatio-temporally coherent clusters obtained from k-means. $K = 11$, $w = 4$.

Abundance Curves

We compute the abundance for a cluster k in week t as the sum of presence scores across all locations C_k assigned to the cluster:

$$q_t(k) = \sum_{i \in C_k} q_i^t,$$

where q_i^t is the STEM presence score for location i in week t. Every cluster is therefore associated with an abundance curve (time series), where we use the sum of presence scores as a surrogate for the expected number of birds across the cluster region in a particular week. In future work, we hope to replace this surrogate measure with actual abundance estimates for each location in each week.

Network Inference

The migration network is a graph $G(V, E)$, where each node is one of the K spatially-explicit migratory regions (clusters of STEM locations) and each edge connects a pair of regions that are close enough to be reachable within one week of migratory movement. For each pair of nodes i and j, $P_{ij}(t)$ is the transition probability from region i to region j in week t. It is zero for all $(i, j) \notin E$. Our goal is to infer transition probabilities on the edges of the network G that best explain the region abundance scores $q_t(k)$ across weeks. We use two different models to infer the transition probability matrix P: 1) a stationary Markov chain where the transition probabilities on the edges do not change across weeks, and 2) a non-stationary Markov chain where the transition probabilities change as a function of time-varying features of the migration regions.

Stationary Markov Chain

We estimate the stationary Markov chain model, by solving a *constrained linear least squares* problem. The mathematical formulation is,

$$\underset{P}{\text{minimize}} \quad \sum_{t=1}^{T} \|q_t P - q_{t+1}\|_2^2 \tag{1}$$

$$\text{subject to} \quad P_{ij} \geq 0, \ i,j \in \{1, ..., K\} \tag{2}$$

$$\sum_{j=1}^{K} P_{ij} \leq 1, \ i \in \{1, ..., K\} \tag{3}$$

$$P_{ij} = 0, \ (i,j) \notin E \text{ or } \text{lat}(i) \leq \text{lat}(j) \tag{4}$$

Constraint 3 indirectly models an outgoing edge to an additional sink node. A sink node is required since our system is not closed, but is essentially only a part of the full migration range. The Tree Swallow migratory range is broader than the US, and includes Canada to the North and Mexico to the South (Butler 1988).

Constraint 4 imposes the edge restrictions and the directionality of the migratory movement. Since the period we consider is the fall migration, it is known that the birds will move from North to South across the US. So the constraint forbids transitions from any node at a lower latitude to any other node at a higher latitude. The objective is to minimize the error in predicted regional abundance levels at each week based on the abundance levels of the previous week and the assigned transition probabilities.

Non-stationary Markov Chain

The non-stationary Markov chain is used to capture the dependence of transition probabilities on time-varying properties, such as temperature and wind, availability of food and other factors, which impact migration across regions. This

time-varying model is a more accurate representation of the dynamics of bird migration.

We use the *multinomial logit function* (MacRae 1977) to parameterize the transition probabilities in terms of time-varying environmental features of the source and destination migration region. We define $X_{ij}(t)$ as the features for edge (i, j) at week t based on features of the regions i and j. Under the multinomial logit representation,

$$P_{ij}(t) = \frac{e^{\beta^T X_{ij}(t)}}{1 + \sum_{j'=1}^{k} e^{\beta^T X_{ij'}(t)}}, \forall i, j \in \{1, ..., K\}$$

$$P_{is}(t) = \frac{1}{1 + \sum_{j=1}^{k} e^{\beta^T X_{ij}(t)}}, \forall i \in \{1, ..., K\}$$

$$ln\left(\frac{P_{ij}(t)}{P_{is}(t)}\right) = \beta^T X_{ij}(t), \forall i, j \in \{1, ..., K\}$$

We explicitly add a sink node s in this model. Under this form, the non-negativity constraint and the sum-to-one constraint for estimating a transition probability matrix are automatically modeled. This allows us to frame an *unconstrained nonlinear least squares* problem to solve for the transition probability matrix. The formulation is defined as,

$$\underset{\beta}{\text{minimize}} \quad \sum_{t=1}^{W} \|q_t P(\beta, t) - q_{t+1}\|_2^2$$

where

$$P_{ij}(\beta, t) = \frac{e^{\beta^T X_{ij}(t)}}{1 + \sum_{j'=1..k:\text{lat}(i)>\text{lat}(j')} e^{\beta^T X_{ij'}(t)}}$$

$$\forall i, j : \text{lat}(i) > \text{lat}(j)$$

$$P_{is}(\beta, t) = \frac{1}{1 + \sum_{j'=1..k:\text{lat}(i)>\text{lat}(j')} e^{\beta^T X_{ij'}(t)}}$$

$$P_{ij}(t) = 0, \forall i, j : \text{lat}(i) \leq \text{lat}(j)$$

The environmental features used for our model were the weekly mean temperature, precipitation obtained from the WorldClim database (Hijmans et al. 2005) and Normalized Difference Vegetation Index (NDVI) from the eMODIS database (Jenkerson, Maiersperger, and Schmidt 2010). We also used the inverse of the distance between the nodes as a feature where the distance is computed as the Euclidean distance among the coordinates of the nodes. To obtain the weekly D-dimensional environmental feature vector $X_k(t)$ for each cluster (node) k, we took the mean value of the point-wise weekly d feature $f_i^d(t)$ over all points i in the set of STEM locations C_k in cluster k:

$$X_k^d(t) = \frac{1}{|C_k|} \sum_{i \in C_k} f_i^d(t)$$

The edge features $X_{ij}(t)$ were computed as the gradient of the environment features between node i at week t and node j at week $t + 1$:

$$X_{ij}(t) = \left\{1, \frac{1}{1 + d_{ij}}, \{X_j^d(t+1) - X_i^d(t)\}_{d=1...D}\right\}$$

where d_{ij} is the distance between node i and node j and $X_i(t)$ is the feature vector of environment features for cluster i at week t.

Results

We use the Root Mean Squared Error (RMSE) metric to assess the fit of our models. The RMSE is calculated as,

$$\text{RMSE} = \sqrt{\sum_{i=1}^{k} \sum_{t=1}^{W} (q_{it} - qOpt_{it})^2}$$

where q_{it} is the presence score and $qOpt_{it}$ is the estimated presence score by the optimization solution at cluster i at week t.

Table 1 gives the overall RMSE values for the stationary and gradient Markov Chain models. A detailed summary for RMSE (overall, grouped by week and by cluster) is given in Table 3 in the Appendix. The coefficients for the features obtained from the non-stationary gradient model are given in Table 2.

A visual comparison among the weekly estimates for all clusters is provided in Figure 4 in the Appendix. Due to space constraints we show the trends in the weeks 40-50 from our data of weeks 32-52. It shows the expected fraction of the population (actual and estimated) which is the presence score for each week normalized by the total presence score in the initial week, which we assume corresponds to the maximum population in the time period considered.

Group	Stationary	Gradient
RMSE	0.0111	0.0120

Table 1: RMSE values for the Stationary and Gradient MC models

Features	Coefficients (β)
Intercept	2.6056
Distance	1.5560
Mean Temperature	0.4384
Precipitation	-0.0500
NDVI	0.7935

Table 2: Coefficients for the Gradient MC model

The migration networks obtained from the two models are shown in Figure 3. To display edges in the gradient model, for each possible edge between two clusters, we compute the maximum estimated transition probability across all weeks as $P_{ij}^{\max} = \max_t P_{ij}(t)$. We display only transition edges with $P_{ij}^{\max} > 0.01$.

Observations

We can make the following observations about the RMSE, the coefficients and the obtained migration networks,

- The Stationary model performs better than the Gradient model overall but both have particular clusters or weeks where one performs better or worse than the other.

- Both the models have bad estimates for the initial clusters in the later weeks where the abundance decreases a lot. The stationary model overestimates abundance more
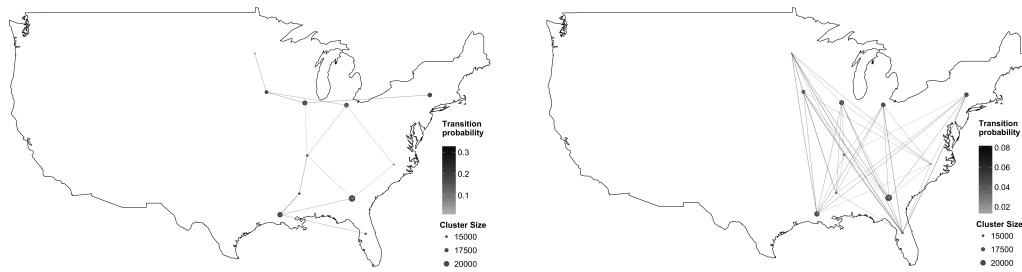
Figure 3: Migration Networks estimated by the Stationary and Gradient MC models, based on clusters shown in Fig. 2.

in such cases compared to gradient model. The gradient model underestimates abundance around weeks 40-44 but improves later on.

- All the features were normalized to have mean zero and standard deviation one for the optimization in the gradient model. This allows us to comment on the relative importance of features in determining transition. We see that the gradient model prefers shorter distances, higher NDVI, higher temperatures and lower precipitation for higher probabilities of transition.

- The stationary model network has an edge going from a node around the Connecticut region to a node near Lake Michigan. This is unexpected and might suggest that birds from Canada are coming into northern US around this time and the model is trying to compensate for this rise by sending birds from the Connecticut region.

- The network from the stationary model has much fewer edges than the gradient model network. The gradient model network spreads the probabilities across a lot of edges. We see that most of the edges going from nodes extreme north to south have higher transition probabilities in the gradient model. This might be related to higher differences in temperature and NDVI in such locations.

Conclusions

We present a framework to tackle the challenging problem of inferring migration structure from aggregate data about bird migration. We build a processing pipeline involving clustering data into migration regions, constructing a Markov chain to model transitions between such regions using both a stationary and non-stationary approach to capture dependence on time-varying environmental features. The framework is flexible enough to include domain knowledge at various stages and hence can be applied to study migrations for diverse species.

Future Work

There are some issues in our preliminary study that we wish to address in the future work.

- We currently use the STEM presence scores as a proxy to determine the abundance in a migration region.

- We do not consider *mortality* of the birds in our analysis. The migration out of the system into the sink node does not differentiate between population that was alive or

dead. In future work, we would like to cover the complete migration area for tree swallows and specifically deal with mortality issues.

- Coverage of the whole migration range would also likely improve the estimated networks by removing unexpected edges like the Connecticut to Lake Michigan one in the stationary model since all population will be accounted for.

- Currently the model does not include any *distance constraints*, i.e the distances over which birds can fly in a week. Including distance constraints in the model would potentially reduce the number of edges in the gradient model and increase local transition probabilities.

- The current estimates of the environmental features for temperature and precipitation are interpolations of observed data, representative of years 1950-2000. We would like to include conditions for current years and take yearly/monthly variations into account.

Acknowledgements

Appendix
References

Ambrosini, R.; Borgoni, R.; Rubolini, D.; Sicurella, B.; Fiedler, W.; Bairlein, F.; Baillie, S. R.; Robinson, R. A.; Clark, J. A.; Spina, F.; et al. 2014. Modelling the progression of bird migration with conditional autoregressive models applied to ringing data. *PloS one* 9(7):e102440.

Butler, R. W. 1988. Population dynamics and migration routes of tree swallows, tachycineta bicolor, in north america (dinámica de poblaciónes y rutas de migración de tachycineta bicolor en norte america). *Journal of Field Ornithology* 395–402.

Elmohamed, M.; Kozen, D.; and Sheldon, D. R. 2007. Collective inference on markov models for modeling bird migration. In *Advances in Neural Information Processing Systems*, 1321–1328.
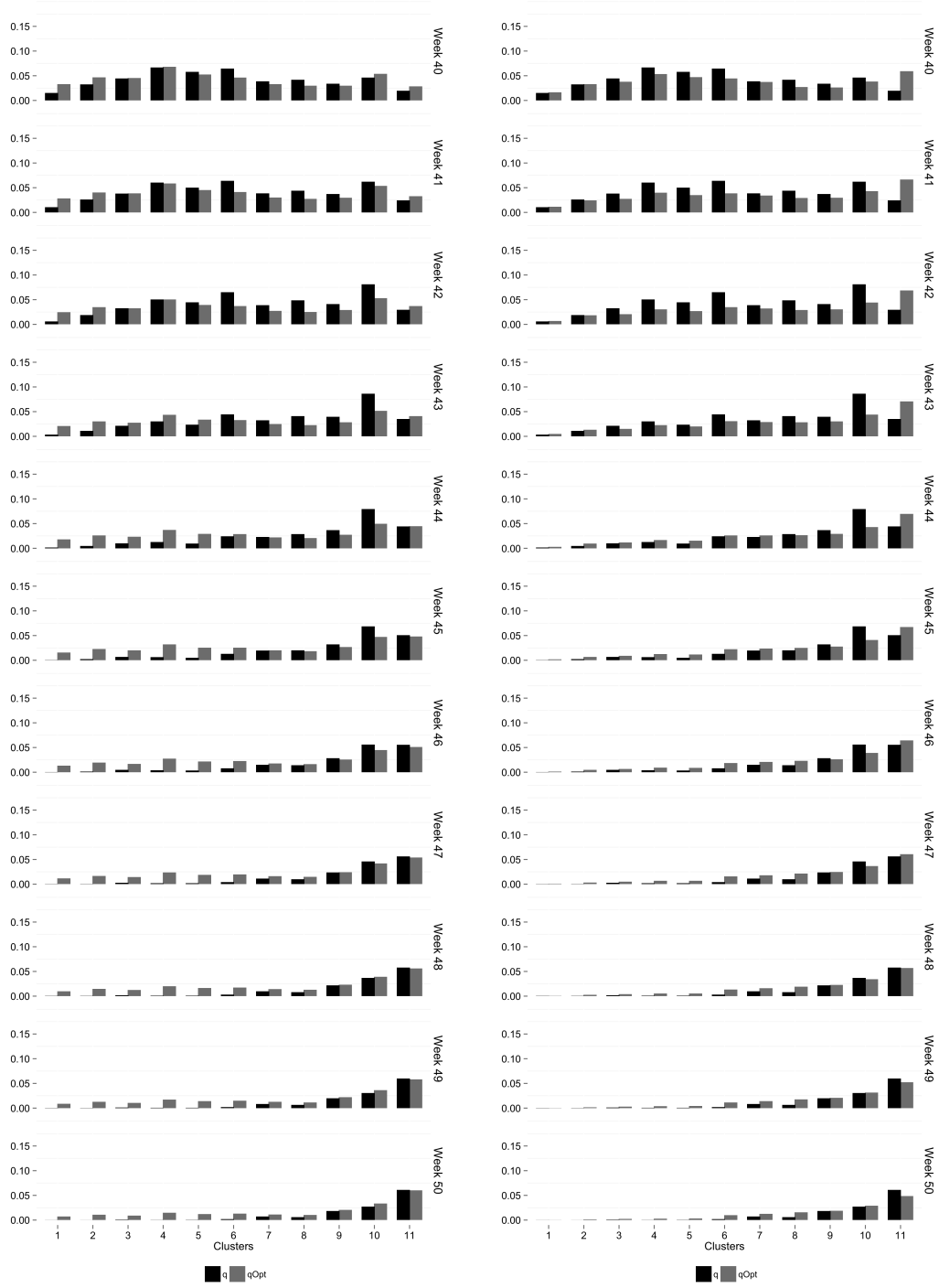
Figure 4: Comparing stationary and gradient models. The bars show the expected population fraction in red and the estimates from the models in blue.

Group RMSE	Stationary	Gradient
Total	0.0111	0.0120
Week 32	0.0000	0.0000
Week 33	0.0070	0.0048
Week 34	0.0104	0.0065
Week 35	0.0116	0.0101
Week 36	0.0106	0.0117
Week 37	0.0078	0.0146
Week 38	0.0083	0.0170
Week 39	0.0097	0.0132
Week 40	0.0104	0.0174
Week 41	0.0120	0.0191
Week 42	0.0167	0.0226
Week 43	0.0162	0.0179
Week 44	0.0164	0.0127
Week 45	0.0151	0.0096
Week 46	0.0132	0.0069
Week 47	0.0120	0.0062
Week 48	0.0107	0.0056
Week 49	0.0096	0.0057
Week 50	0.0084	0.0061
Week 51	0.0073	0.0064
Week 52	0.0062	0.0077
Cluster 1	0.0138	0.0082
Cluster 2	0.0144	0.0063
Cluster 3	0.0088	0.0119
Cluster 4	0.0136	0.0114
Cluster 5	0.0108	0.0124
Cluster 6	0.0128	0.0139
Cluster 7	0.0047	0.0044
Cluster 8	0.0095	0.0117
Cluster 9	0.0058	0.0056
Cluster 10	0.0160	0.0170
Cluster 11	0.0047	0.0195

Table 3: RMSE values for the two models averaged for each individual cluster and for each week.

Fink, D.; Hochachka, W. M.; Zuckerberg, B.; Winkler, D. W.; Shaby, B.; Munson, M. A.; Hooker, G.; Riedewald, M.; Sheldon, D.; and Kelling, S. 2010. Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications* 20(8):2131–2147.

Hijmans, R. J.; Cameron, S. E.; Parra, J. L.; Jones, P. G.; and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology* 25(15):1965–1978.

Jenkerson, C.; Maiersperger, T.; and Schmidt, G. 2010. emodis: a user-friendly data source. *US Geological Survey Open-File Report* 1055(10).

MacQueen, J., et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, 281–297. California, USA.

MacRae, E. C. 1977. Estimation of time-varying markov processes with aggregate data. *Econometrica: journal of the Econometric Society* 183–198.

Sheldon, D. R., and Dietterich, T. G. 2011. Collective graphical models. In *Advances in Neural Information Processing Systems*, 1161–1169.

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

Tibshirani, R.; Walther, G.; and Hastie, T. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2):411–423.