Clustering Species Accumulation Curves to Identify Skill Levels of Citizen Scientists Participating in the eBird Project

Jun Yu
Department of EECS
Oregon State University
yuju@eecs.orst.edu

Weng-Keen Wong
Department of EECS
Oregon State University
wong@eecs.orst.edu

Steve Kelling
Cornell Lab of Ornithology
Cornell University
stk2@cornell.edu

Abstract

Although citizen science projects such as eBird can compile large volumes of data over broad spatial and temporal extents, the quality of this data is a concern due to differences in the skills of volunteers at identifying bird species. Species accumulation curves, which plot the number of unique species observed over time, are an effective way to quantify the skill level of an eBird participant. Intuitively, more skilled observers can identify a greater number of species per unit time than inexperienced birders, resulting in a steeper curve. We propose a mixture model for clustering species accumulation curves. These clusters enable the identification of distinct skill levels of eBird participants, which can then be used to build more accurate species distribution models and to develop automated data quality filters.

Introduction

Citizen science is a paradigm in which volunteers from the general public collect scientifically relevant data. This paradigm is especially useful when the scope of the data collection requires covering such a broad spatial and temporal extent that it cannot be performed by a limited number of trained scientists. Researchers in the ecological sciences are beginning to recognize the effectiveness of citizen science and as a result, citizen science projects for collecting observational data for biodiversity have proliferated in recent years. Examples of these projects include eBird for birds (Sullivan et al. 2009; Kelling et al. 2013), eButterfly for butterflies (Larrivee et al. 2014) and REEF for reef fish (Pattengill-Semmens and Semmens 2003). In addition, sites such as iNaturalist (http://www.inaturalist.org) and Project Noah (http://www.projectnoah.org) allow citizen scientists to record observations for a wide range of organisms.

Our work is in the context of the eBird project, which relies on a global network of citizen scientists to record checklists of bird observations, identified by species, through a protocol-driven process. These checklists are submitted via the web to the Cornell Lab of Ornithology, forming one of the largest biodiversity datasets in existence, with over 140 million observations reported by 150,000 birders worldwide. This data plays an important role in ecological research (Hochachka et al. 2012) and conservation (North American Bird Conservation Initiative, U.S. Committee 2013).

Data quality is an common concern for citizen science projects and it is especially important for eBird due to the high volume of data submitted. The strategies for improving data quality depend on whether citizen scientists play the role of *processors* or *sensors*. When citizen scientists act as *processors*, the same processing tasks are usually repeated and multiple volunteers can be assigned to work on the same task by design. For example, Zooniverse (Savage 2012) uses volunteers as processors by having them classify or extract information from images. Validation of this process can be done through a consensus of responses or directly from an expert (Sheppard and Terveen 2011). Consequently, the skill level of a citizen scientist can be assessed based on the accuracy of his or her finished tasks; participants acting as *processors* can be easily grouped into different skill levels.

When citizen scientists act as *sensors*, they report observations at their specific locations and times. Assessing their skill levels is difficult because ground truth is rarely available to validate their finished tasks and events can neither be repeated nor independently observed. In eBird, participants act as a global sensor network, actively collecting bird observational data over a broad spatial and temporal extent. Most sites are surveyed by a few participants and there is no ground truth of species occupancies at a site to validate their submissions. To ensure high quality data submissions, the current eBird system employs a regional filter based on expected occurrences of each species at specific times of the year. This filter flags anomalous observations and any flagged records are reviewed by a large network of volunteer reviewers. Observations are discarded if they do not pass the review stage; otherwise the data is accepted to the database.

A major factor influencing data quality in eBird is the variation in observer skill at detecting bird species. Past work has shown that accounting for observer skill can produce more accurate species distribution models (Sauer, Peterjohn, and Link 1994; Yu, Wong, and Hutchinson 2010) and can enable the development of automated data quality filters (Yu et al. 2012). Our goal is to employ a data-driven approach to identify distinct skill levels of observers. We measure an observer's skill by the number of species they submit on a checklist, operating under the assumption that more skilled observers report a greater number of bird species per unit of time. We propose to represent a participant's skill level using a Species Accumulation Curve (SAC) (Gotelli and Colwell 2001), which plots the cumulative number of unique species observed as a function of the cumulative effort expended (e.g. time). SACs are typically used in the ecological literature to quantify species richness (Chao and Jost 2012). However, we repurpose the use of SACs as a measure of an observer's skill at detecting species. We can compare the skills of participants using their SACs if we can control for the environment in which they make their observations. Intuitively, skilled birders rely on both sound and sight to identify bird species. Thus, they are able to identify more species per unit time than inexperienced birders in the same environment, resulting in a steeper SAC.

We can then cluster participant SACs to identify distinct skill levels of birders in the data. To accomplish this task, we develop a mixture model to cluster the SACs of eBird participants and derive a learning algorithm based on Expectation-Maximization. In our experiments, we apply our model to eBird data and show that the resulting clusters are meaningful and correspond to distinct skill levels.

The mixture of SACs model

In the mixture of SACs model, we assume that there is a fixed number K of distinct groups of observers and observers in the same group are at a similar skill level. As eBird is our application domain, we use *observer* and *birder* interchangeably. Figure 1 shows a plate diagram of the mixture of SACs model. The plate on the left represents K groups where group k is parameterized with $\boldsymbol{\beta}_k$. The outer plate on the right represents M birders. The variable $Z_i \in \{1, \cdots, K\}$ denotes the group membership of birder i. The inner plate represents N_i checklists submitted by birder i. The variable X_{ij} represents the amount of effort (e.g. duration) and Y_{ij} specifies the number of unique species reported on checklist j of birder i. Finally, let $\boldsymbol{X}_{ij}$ denote the vector consisting of the variable X_{ij} and the intercept term.

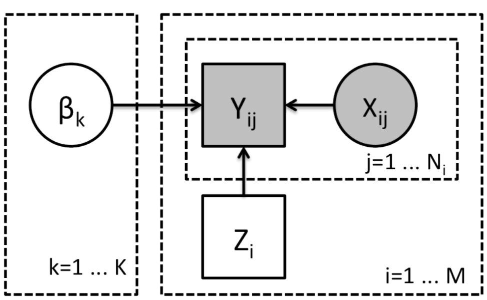


Figure 1: The mixture of Species Accumulation Curves model.

The observation variable Y_{ij} depends on the effort X_{ij} and the skill level of birder i, indicated by the group membership Z_i. To model their relationship in a SAC, we use a linear regression model with a square root transformation on X_{ij} (i.e. $Y_{ij} = \beta_0 + \beta_1\sqrt{X_{ij}}$). A number of alternative models could be used (eg. negative binomial regression) but we found that the square root model produced the best fit to the data, where the fit is measured in terms of mean squared error on a holdout set.

The structure of the mixture model corresponds to the following generative process. For each birder i, we first generate its group membership Z_i by drawing from a multinomial distribution with parameter $\boldsymbol{\pi}$. Next, birder i produces N_i checklists. On each checklist j, the expected number of species detected is $\boldsymbol{\beta}_{Z_i} \cdot \boldsymbol{X}_{ij}$ where $\boldsymbol{\beta}_{Z_i}$ are the parameters of group Z_i. Finally, the number of species actually reported (Y_{it}) is generated by drawing from a Gaussian distribution with mean $\boldsymbol{\beta}_{Z_i} \cdot \boldsymbol{X}_{ij}$ and variance σ^2. Here we assume SACs in different groups share the same variance σ^2. The log-likelihood for this mixture model is given in the following equation.

$$\sum_{i=1}^{M} \log \left(\sum_{k=1}^{K} P(Z_i = k; \boldsymbol{\pi}) \prod_{j=1}^{N_i} P(Y_{ij}|\boldsymbol{X}_{ij}, Z_i = k; \boldsymbol{\beta}, \sigma^2) \right)$$

Parameter estimation

During learning, we estimate the model parameters $\{\boldsymbol{\pi}, \boldsymbol{\beta}, \sigma^2\}$ and the latent group membership $\boldsymbol{Z}$ for each birder using Expectation Maximization (EM) (Dempster, Laird, and Rubin 1977). The EM algorithm iterates between performing the E-step and M-step until convergence. In the E-step, EM computes the expected group membership for every birder i. In the M-step, we re-estimate the model parameters $\{\boldsymbol{\pi}, \boldsymbol{\beta}, \sigma^2\}$ that maximize the expected complete log-likelihood in the equation below. In addition, let r_{ik} denote the expected group membership of birder i belonging to group k.

$$\begin{aligned} \mathcal{Q} &= E_{\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{X}}[\log(P(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{X}; \boldsymbol{\pi}, \boldsymbol{\beta}, \sigma^2))] \\ &= \sum_{i=1}^{M}\sum_{k=1}^{K} E_{\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{X}}[\mathbb{I}(Z_i = k)] \\ &\quad \log \left(P(Z_i = k; \boldsymbol{\pi}) \prod_{j=1}^{N_i} P(Y_{ij}|\boldsymbol{X}_{ij}, Z_i = k; \boldsymbol{\beta}, \sigma^2) \right) \end{aligned} \tag{1}$$

In the E-step, we keep the parameters $\{\boldsymbol{\pi}, \boldsymbol{\beta}, \sigma^2\}$ fixed and update the expected group membership r_{ik} for every birder i and group k. This expected membership can be computed as the posterior probability in the equation below.

$$\begin{aligned} & r_{ik} \\ &= P(Z_i = k|\boldsymbol{X}_{i\cdot}, \boldsymbol{Y}_{i\cdot}; \boldsymbol{\pi}, \boldsymbol{\beta}, \sigma^2) \\ &= \frac{P(Z_i = k; \boldsymbol{\pi}) \prod_{j=1}^{N_i} P(Y_{ij}|\boldsymbol{X}_{ij}, Z_i = k; \boldsymbol{\beta}, \sigma^2)}{\sum_{k'=1}^{K} P(Z_i = k'; \boldsymbol{\pi}) \prod_{j=1}^{N_i} P(Y_{ij}|\boldsymbol{X}_{ij}, Z_i = k'; \boldsymbol{\beta}, \sigma^2)} \end{aligned}$$

In the M-step, we re-estimate $\{\boldsymbol{\pi}, \boldsymbol{\beta}, \sigma^2\}$ using the expected membership computed in the E-step. To estimate π_k, we introduce a Lagrange multiplier λ to ensure that the constraint $\sum_{k=1}^{K} \pi_k = 1$ is satisfied. i.e. $\sum_{i=1}^{M} r_{ik} - \lambda\pi_k = 0$. Summing over all $k \in \{1, \cdots, K\}$, we find that $\lambda = \sum_i \sum_k r_{ik} = M$. Thus we plug λ into the equation above and get the updating equation for π_k.

$$\pi_k = \frac{1}{M}\sum_{i=1}^{M} r_{ik}$$

To estimate $\boldsymbol{\beta}_k$, we compute the gradient of $\boldsymbol{\beta}_k$ w.r.t. the expected complete log-likelihood $\mathcal{Q}$ in Equation 1. Notice that the gradient of $\boldsymbol{\beta}_k$ has the same form as that of a linear regression model, except each instance is associated with a weight of r_{ik}. Thus we can use the method of least squares to update $\boldsymbol{\beta}_k$ efficiently.

$$\frac{\partial \mathcal{Q}}{\partial \boldsymbol{\beta}_k} = \frac{1}{\sigma^2}\sum_{i=1}^{M} r_{ik}\sum_{j=1}^{N_i}(Y_{ij} - \boldsymbol{\beta}_k \boldsymbol{X}_{ij})\boldsymbol{X}_{ij}$$

Finally, we compute the gradient of σ^2 w.r.t. the expected complete log-likelihood $\mathcal{Q}$ and the updating equation for the parameter σ^2 has the closed-form solution.

$$\sigma^2 = \frac{\sum_{i=1}^{M}\sum_{k=1}^{K} r_{ik}\sum_{j=1}^{N_i}(Y_{ij} - \boldsymbol{\beta}_k \boldsymbol{X}_{ij})^2}{\sum_{i=1}^{M} N_i}$$

Since the EM algorithm may converge to a local maximum of the expected complete log-likelihood function, depending on initialization of the parameters, we use random restart by assigning each birder to a group randomly. The expected membership $r_{i\cdot}$ specifies a soft clustering of birder i. To partition birders in the training data, we assign each birder to the group with the largest expected membership.

Determining the number of groups

To determine the number of groups in the data, we start the mixture model with only one group (K=1) and gradually increase the value of K until it does not improve the average log-likelihood on a holdout set. The average log-likelihood is defined in the following equation. Unlike the log-likelihood function, we compute the data likelihood of a birder by averaging the observation probability $P(Y_{ij}|\boldsymbol{X}_{ij}, Z_i = k; \boldsymbol{\beta}, \sigma^2)$ over all the observations from that birder.

$$\sum_{i=1}^{M} \log\left(\sum_{k=1}^{K} P(Z_i = k; \boldsymbol{\pi})\frac{\sum_{j=1}^{N_i} P(Y_{ij}|\boldsymbol{X}_{ij}, Z_i = k; \boldsymbol{\beta}, \sigma^2)}{N_i}\right)$$

Evaluation and Discussion

We evaluate the mixture of SACs model using the eBird Reference Data (Munson et al. 2009) from 2012. We perform four separate studies using four species-rich states that have high levels of year-round eBird participation (New York, Florida, Texas, and California). We attempt to control for species richness in the environment by using state boundaries; in future work, we will look at using Bird Conservation Regions, which are more homogenous with regard to species richness than political boundaries. We evaluate the mixture model's ability to accurately cluster eBird participants into groups based on their skill levels. However, evaluating the clustering results is challenging due to the lack of ground truth on the participants' skill levels. Given the large number of birders, we can not validate the clusters by manually verifying each birder's submissions. Instead, we propose to validate the clusters based on an individual birder's ability to identify hard-to-detect species and use anecdotal information from the eBird project staff. We also run the same analyses on eBird hotspots where the number of observers is relatively small, allowing us to manually verify their skills and validate the clustering results.

Grouping eBird participants

First, we remove the birders who submitted fewer than 20 checklists because their data is too sparse to fit a SAC. Then, we limit our analysis to only include checklists with duration less than 2 hours. To find the number of distinct groups in the data, we split the data into training and validation sets. We train the mixture model on the training set with different values of $K \in \{1, \cdots, 5\}$, and then we calculate the average log-likelihood on the validation set. The best value of K is chosen when increasing K does not improve the average log-likelihood. In Table 1, we show the average log-likelihood on the holdout data in four states. The graphs clearly show that there are 3 distinct groups in all four states.

State	K=1	K=2	K=3	K=4	K=5
NY	-3.456	-3.407	**-3.396**	-3.400	-3.406
FL	-3.398	-3.389	**-3.387**	-3.393	-3.405
TX	-3.543	-3.496	**-3.491**	-3.495	-3.501
CA	-3.507	-3.483	**-3.481**	-3.489	-3.493

Table 1: The average log-likelihood of the holdout data in four states. The numbers in bold indicate the number of distinct groups found in that state.

State	The percent of birders			Avg CLs per birder		
	G1	G2	G3	G1	G2	G3
NY	7%	37%	56%	407	215	152
FL	19%	45%	36%	200	125	124
TX	18%	44%	38%	132	157	99
CA	12%	40%	48%	236	195	111

Table 2: The percent of birders and the average number of checklists submitted per birder in each group (G1-G3) in the four states.

Given the value of K chosen above, we re-estimate the mixture model using all the data in 2012 and show the SACs of different groups in four states in Figure 2. We sort the SACs by their slope coefficient β_1 in decreasing order so that the first group corresponds to the most skilled observers and the last group corresponds to the least skilled observers. The red curve corresponding to the top group has a consistently higher SAC than the other two groups across all four states. Birders in this top group are able to detect around 40 unique species during a 2-hour birding trip, while birders in group 2 and group 3 can only detect around 30 and 20 species. Though the number of distinct groups are the same in all four states, the proportions of groups are very

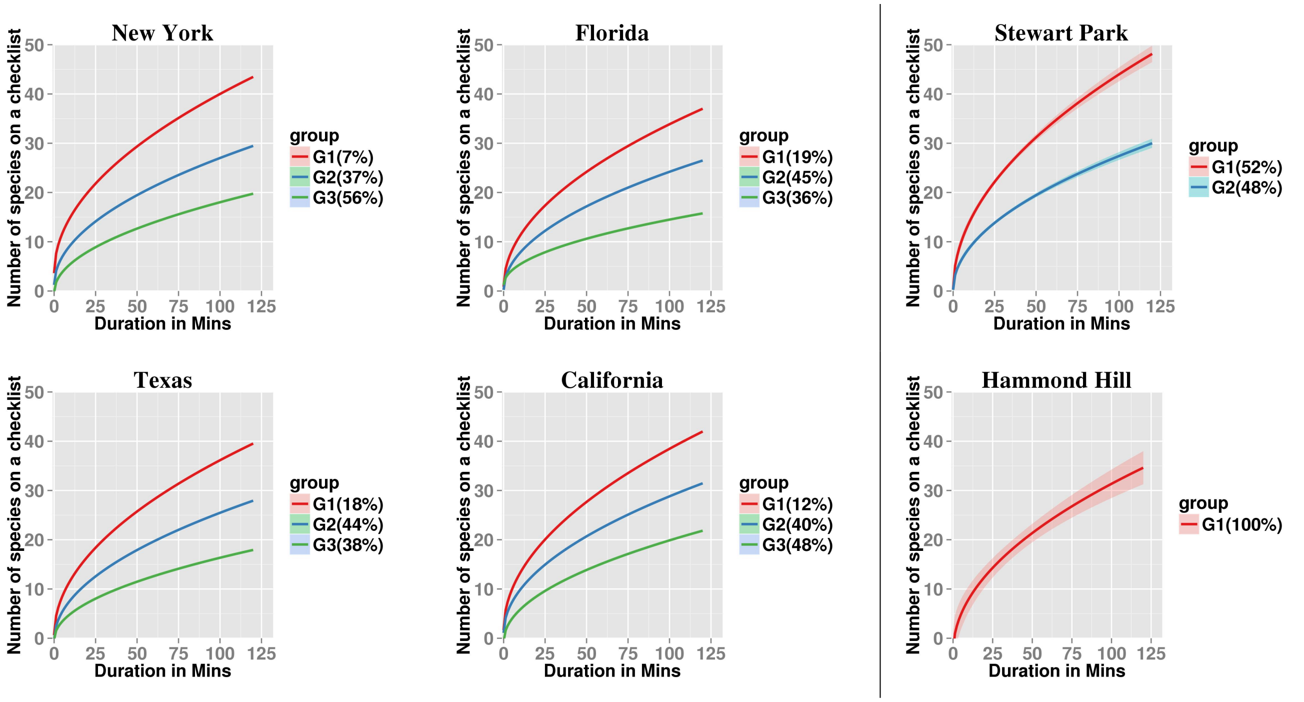


Figure 2: The SACs learned from the mixture of SACs model in four states. The proportion of birders in each group is indicated in parentheses. The shaded area of a curve shows the 95% confidence interval containing the mean of the curve.

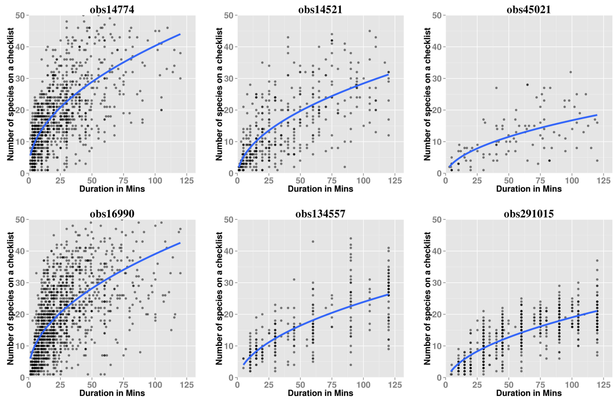
different. In New York and California, there are 7% and 12% participants falling into the top group as they are able to detect more species per unit of time. In Florida and Texas, the size of the top group is bigger, with 19% and 18% observers respectively. One explanation is that in New York, a small group of observers from the Cornell Lab of Ornithology are extremely skilled at identifying bird species and this elite skill level distinguishes them from the rest of the eBird participants in New York.

In Table 2, we report the proportion of observers and the average number of checklists submitted per birder in each group. The observers in the more skilled groups submit more checklists than observers in the less skilled groups. This matches our intuition that observers who are more active and involved in the eBird project tend to be more skilled at detecting bird species. To demonstrate the differences in birding skills of birders across groups, we randomly choose two birders from each group in New York and show their SACs in Figure 3. Birders in the top group are able to accumulate species much faster especially in the first 30-45 minutes than birders in groups 2 and 3.

Detection of hard-to-detect bird species

A good partition of birders leads to distinct differences in the skill levels of different groups. Since we do not have ground truth on birders' skills, we characterize their skill levels in terms of their ability to detect hard-to-detect bird species. Hard-to-detect species often require more experience and skills to be identified, e.g. some species can be detected by sound rather than by sight and some species can be detected only if observers know their habitats. In our experiment, we use 8 hard-to-detect species in each state suggested by experts at the Cornell Lab of Ornithology and calculate the average detection rate of observers within each group. An observer's detection rate of a species is defined to be the percent of one's checklists that report the detection of that species. In Figure 4, we show the average detection rate of the hard-to-detect species in each group. The top group has the highest detection rate across all species in all four states, showing that a steeper SAC does in fact correspond to a better skill level. As we go from group 1 to group 3, the detection rate of reporting these species keeps decreasing and shows statistically significant differences between two adjacent groups. These differences show that birders in different groups vary greatly in their skill levels and the mixture model is able to cluster birders of similar skills into the same group.

In addition, we sent a list of birder IDs in the top group for New York to the eBird project leaders and asked them to verify if these birders are reputable birders in the community. Out of 30 birders in the top group, 25 are experts from the Cornell Lab of Ornithology or known regional experts in New York while the other 5 observers are known to be reputable birders submitting high quality checklists to eBird. Thus, the mixture model is able to identify a group of top eBird contributors that are highly skilled birders and distinguish their behavior from the other groups of eBird participants.



(a) Group 1 (b) Group 2 (c) Group 3

Figure 3: The SACs of six eBird participants from New York, with two participants from each of the three groups. Each point represents a checklist submitted by the birder. The darkness of a point in the scatterplot specifies the number of checklists overlapped in the location. A darker color indicates more checklists overlapped at that point.

eBird Hotspots

Since validating the clusters of birders at each location in a state is not viable, we run the same analysis on data from two eBird hotspots (*Stewart Park* and *Hammond Hill* in New York), where the smaller number of observers allows human verification of the skill levels of individual birders. The eBird hotspots are public birding locations that are often heavily visited all year around. After training the mixture model using data submitted in these two hotspots, the model discovers 2 groups in Stewart Park and only 1 group in Hammond Hill. The SACs of these two eBird hotspots are shown in Figure 2. In Stewart Park, there are 25 birders submitting at least 10 checklists in 2012 and about half of the birders (13 birders) are classified into group 1. Members of the eBird project assessed the skill levels of the birders by the reputation of these birders in the local birding community and by manually inspecting their submitted checklists. All 13 birders in group 1 have been verified to be expert birders and 10 out of the other 12 birders have been verified to be novice birders. There are two skilled birders classified into group 2 because most of their submissions are short duration observations, making the curve fitting of their observations less accurate. In Hammond Hill, there are only 10 birders submitting at least 10 checklists in 2012 and all of them have been manually verified to be expert birders by the eBird project staff using the same process. Thus, the mixture model is able to find the correct number of groups and cluster birders with similar skill levels into the same group.

Related Work

The topic of user expertise has been explored in other crowdsourcing projects. In question answering communities, researchers are interested in identifying authoritative users by examining content ratings and user reputation (Bian et al. 2009), participation characteristics of experts (Pal et al. 2011) and the link structure of the community (Jurczyk and Agichtein 2007). In product recommendation, McAuley and Leskovec (2013) present a latent-factor recommender system that uses online product reviews to learn the progression of a user's level of expertise. Despite the overarching theme of modeling particpant expertise, there are fundamental differences between eBird and these domains in terms of the structure of the data as well as the underlying processes that generate the data. Furthermore, the indicators of expertise in eBird are more subtle and less overt than the explicit user ratings in question answering communities. Consequently, custom models need to be developed for discovering skill levels of eBird participants.

Yu et al. (2010) show that including the expertise of observers into a species distribution model can produce more accurate predictions of species occurrence. Their work assumes that training data labeled with the skill levels of eBird participants is available. Obtaining this training data is difficult and only possible on a limited scale. Our current work provides a more viable alternative to obtaining this training data by providing an unsupervised approach for identifying distinct observer skill levels and assigning birders to these skill levels.

Conclusion

We proposed a mixture model for SACs that was successful at identifying well-defined skill levels of eBird participants. In our experiments over four species-rich states, these clusters correspond to groups that distinctly vary in their ability to observe hard-to-detect bird species. Furthermore, when we applied the model to two birding hotspots in New York, the model produced accurate and meaningful clusters as verified by domain experts on the eBird project staff.

The mixture of SACs model can be readily applied to other citizen science projects that collect observational species data (eg. eButterfly) and we plan to evaluate its effectiveness on such data. The applicability of our mixture model to other citizen science or crowdsourcing projects depends on whether participant skill can be quantitatively described by a metric similar to a SAC. If such a metric exists and can be reasonably represented as a linear model, then the mixture model presented in our work can be easily modified for that data.

For future work, we plan to extend the model by accounting for other covariates that affect the SACs, such as location and time of year. In addition, we intend to integrate the mixture of SACs approach with the Occupancy-Detection-Expertise model in (Yu, Wong, and Hutchinson 2010). Finally, we would like to investigate the evolution of an observer's skill level over time.

Acknowledgements

The authors would like to thank Marshall Iliff, Brian Sullivan and Chris Wood for their bird expertise as well as Bistra Dilkina, Jeff Gerbracht, Carl Lagoze and Andrea Wiggins for their insight discussions. This material is based upon work supported by the National Science Foundation under Grants No. 0941748 and 1209714.

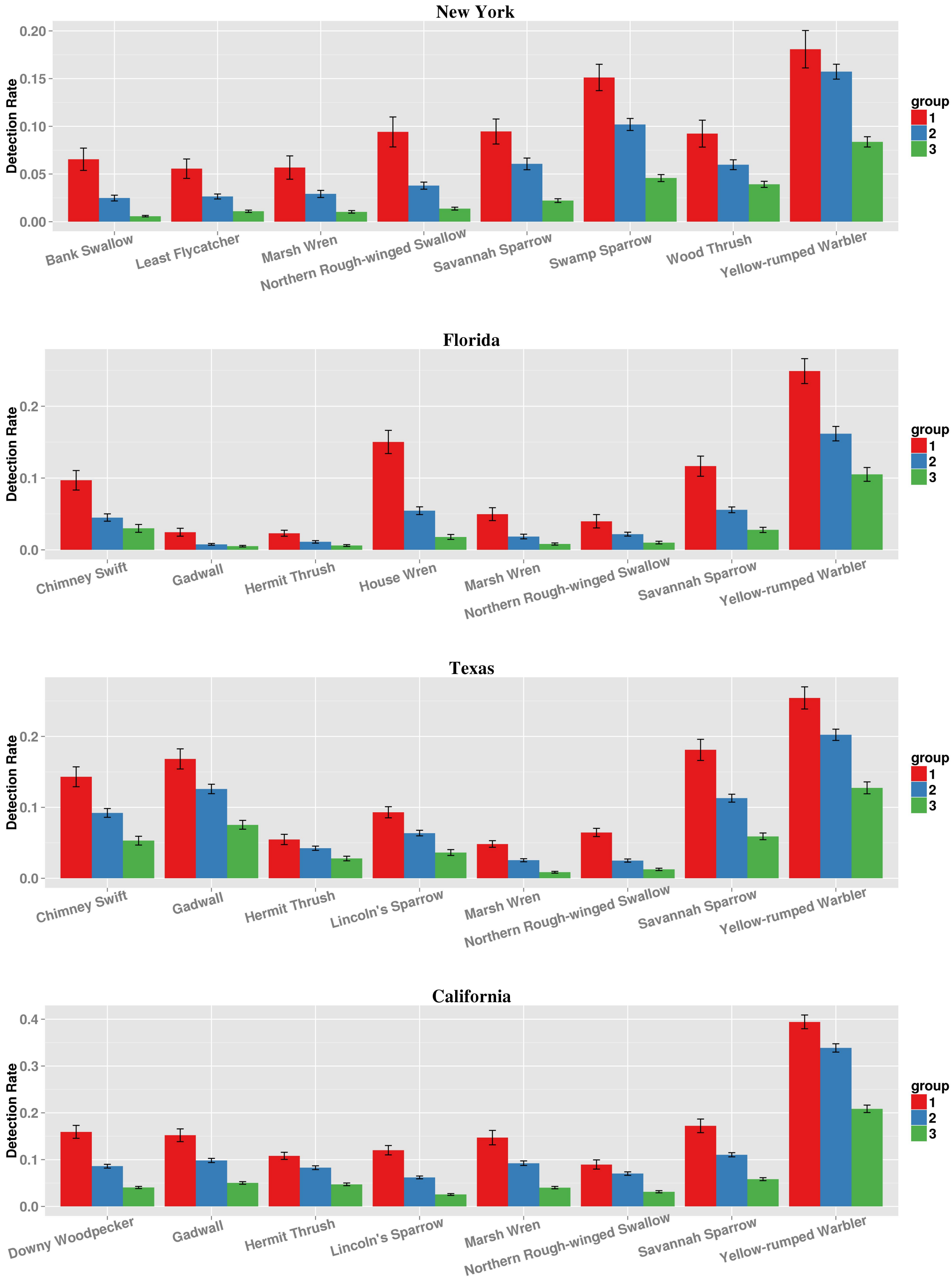


Figure 4: The average detection rate of three groups on 8 hard-to-detect species in NY, FL, TX and CA. The error bars represent the standard error of detection rate within a group.

References

Bian, J.; Liu, Y.; Zhou, D.; Agichtein, E.; and Zha, H. 2009. Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, 51–60. New York, NY, USA: ACM.

Chao, A., and Jost, L. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533–2547.

Dempster, A. P.; Laird, N. M.; and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of Royal Statistical Society* 39(1):1–38.

Gotelli, N., and Colwell, R. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 379–391.

Hochachka, W.; Fink, D.; Hutchinson, R.; Sheldon, D.; Wong, W.-K.; and Kelling, S. 2012. Data-intensive science applied to broad-scale citizen science. *Trends in Ecology and Evolution* 27(2):130–137.

Jurczyk, P., and Agichtein, E. 2007. Discovering authorities in question answer communities by using link analysis. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM '07, 919–922. New York, NY, USA: ACM.

Kelling, S.; Lagoze, C.; Wong, W.-K.; Yu, J.; Damoulas, T.; Gerbracht, J.; Fink, D.; and Gomes, C. P. 2013. ebird: A human/computer learning network to improve conservation and research. *AI Magazine* 34(1):10–20.

Larrivee, M.; Prudic, K. L.; McFarland, K.; and Kerr, J. 2014. ebutterfly: a citizen-based butterfly database in the biological sciences. http://www.e-butterfly.org.

McAuley, J. J., and Leskovec, J. 2013. From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, 897–908.

Munson, M. A.; Webb, K.; Sheldon, D.; Fink, D.; Hochachka, W. M.; Iliff, M.; Riedewald, M.; Sorokina, D.; Sullivan, B.; Wood, C.; and Kelling, S. 2009. The ebird reference dataset, version 1.0. Cornell Lab of Ornithology and National Audubon Society, Ithaca, NY.

North American Bird Conservation Initiative, U.S. Committee. 2013. The state of the birds 2013 report on private lands.

Pal, A.; Farzan, R.; Konstan, J. A.; and Kraut, R. E. 2011. Early detection of potential experts in question answering communities. In *Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization*, UMAP'11, 231–242. Berlin, Heidelberg: Springer-Verlag.

Pattengill-Semmens, C. V., and Semmens, B. X. 2003. Conservation and management applications of the reef volunteer fish monitoring program. *Env. Monitor. Assess. Journal.* 81:43–50.

Sauer, J. R.; Peterjohn, B. G.; and Link, W. A. 1994. Observer differences in the north american breeding bird survey. *The Auk* 111(1):50–62.

Savage, N. 2012. Gaining wisdom from crowds. *Commun. ACM* 55(3):13–15.

Sheppard, S. A., and Terveen, L. 2011. Quality is a verb: the operationalization of data quality in a citizen science community. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, 29–38.

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

Yu, J.; Kelling, S.; Gerbracht, J.; and Wong, W.-K. 2012. Automated data verification in a large-scale citizen science project: a case study. In *Proceedings of the 8th IEEE International Conference on E-Science*, 1–8.

Yu, J.; Wong, W.-K.; and Hutchinson, R. 2010. Modeling experts and novices in citizen science data for species distribution modeling. In *Proceedings of the 2010 IEEE International Conference on Data Mining*, 1157–1162.