Modeling Experts and Novices in Citizen Science Data for Species Distribution Modeling

Jun Yu
School of EECS
Oregon State University
Corvallis, OR 97331
yuju@eecs.oregonstate.edu

Weng-Keen Wong
School of EECS
Oregon State University
Corvallis, OR 97331
wong@eecs.oregonstate.edu

Rebecca A. Hutchinson
School of EECS
Oregon State University
Corvallis, OR 97331
rah@eecs.oregonstate.edu

Abstract—**Citizen scientists, who are volunteers from the community that participate as field assistants in scientific studies [3], enable research to be performed at much larger spatial and temporal scales than trained scientists can cover. Species distribution modeling [6], which involves understanding species-habitat relationships, is a research area that can benefit greatly from citizen science. The eBird project [16] is one of the largest citizen science programs in existence. By allowing birders to upload observations of bird species to an online database, eBird can provide useful data for species distribution modeling. However, since birders vary in their levels of expertise, the quality of data submitted to eBird is often questioned. In this paper, we develop a probabilistic model called the *Occupancy-Detection-Expertise* (ODE) model that incorporates the expertise of birders submitting data to eBird. We show that modeling the expertise of birders can improve the accuracy of predicting observations of a bird species at a site. In addition, we can use the ODE model for two other tasks: predicting birder expertise given their history of eBird checklists and identifying bird species that are difficult for novices to detect.**

Keywords-**Applications, Species Distribution Modeling, Citizen Science, Graphical Models, Contrast Mining**

I. INTRODUCTION

The term *Citizen Science* refers to scientific research in which volunteers from the community participate in scientific studies as field assistants [3]. Since data collection by *citizen scientists* can be done cheaply, citizen scientists allow research to be performed at much larger spatial and temporal scales than trained scientists can cover. For example, species distribution modeling (SDM) [6] with citizen scientists allows data to be collected from many geographic locations, thus achieving broad spatial coverage. Most citizen scientists, however, have little or no scientific training. Consequently, the quality of the data collected by citizen scientists is often questioned. Recent studies have shown that citizen scientists were able to provide accurate data for easily detected organisms [4]. However, for difficult-to-detect organisms, Fitzpatrick et al. [7] found differences between observations made by volunteers and by experienced scientists led to biases in their results.

The eBird project [16], launched in 2002 by the Cornell Lab of Ornithology and National Audubon Society, is one of the largest citizen science programs in existence. The eBird project maintains an online database that allows bird watchers (known as *birders*) to submit checklists that record the bird species they have seen or heard. eBird's goal is to maximize the utility and accessibility of the vast numbers of bird observations made each year by recreational and professional birders. As an example of the volume of data submitted, in January 2010, participants reported more than 1.5 million bird observations across North America.

SDM can, in theory, benefit greatly from data collected by eBird. The goal of SDM is to predict the presence/absence or abundance of a species at a geographic site. SDMs provide insight into species-habitat relationships, which in turn helps ecologists predict biodiversity, design reserves, predict species invasions, and identify areas at risk. A variety of methods have been used for SDM including envelope models [2], Genetic Algorithms [18], GLMs/GAMs [1], Hierarchical Bayesian models [12], Boosted Regression Trees [8], and Maximum Entropy models [17].

Since eBird data is contributed by citizen scientists, can accurate species distribution models be built from this data? Checklists submitted to eBird undergo a data verification process which consists of automated data filters which screen out obvious mistakes on checklists. Then, the checklists go through a review process by a network of experienced birders. Nevertheless, biases still exist due to differences in the expertise level of birders who submit the checklists. In our work, we show that modeling the expertise level of birders can be beneficial for SDM.

In order to incorporate birder expertise into a species distribution model, we need to distinguish between two processes that affect observations: *occupancy* and *detection*. Occupancy determines if a geographic site is viable habitat for a species. Factors influencing occupancy include environmental features of the site such as temperature, precipitation, elevation and land use. Detection describes the observer's ability to detect the species and depends on factors such as the difficulty of identifying the species, the effort put in by the birder, the current weather conditions, and the birder expertise. Neglecting to model the detection process

can result in misleading models [9]. For instance, a bird species might be wrongly declared as not occupying a site when in fact, this species is simply difficult to detect because of reclusive behavior during nesting. Although the focus of this paper is on species distribution modeling, the occupancy / detection problem is representative of a more general problem in domains such as object recognition and surveillance in which a detection process, conditioned on a set of features, corrupts a "true" value with noise to produce an observed value.

Mackenzie et al. [15] proposed a well-known site occupancy model that separates occupancy from detection. We refer to this model as the Occupancy-Detection (OD) model and describe it in detail in Section II-A. Recent work [10] has applied the OD model to citizen science checklist data similar to those from eBird. In our work, we introduce the Occupancy-Detection-Expertise (ODE) model which extends the OD model by incorporating the expertise of citizen scientists. We will show that the ODE model improves the prediction of observations of a bird species at a site, allows prediction of the expertise level of a birder given his or her submitted checklists, and identifies bird species that novices under/over-report as compared to experts.

II. METHODOLOGY

In this section, we first describe the OD model [15], [14] before extending it to incorporate birder expertise.

A. The Occupancy-Detection Model

Figure 1 illustrates the OD model for a single species as a graphical model [11], in which nodes represent random variables and directed edges can be interpreted as a direct influence from parent to child. Circles represent continuous random variables while squares represent discrete random variables. In addition, shaded nodes denote observed variables and unshaded ones denote latent variables. As shown in Figure 1, the true site occupancy at site i (Z_i) is latent while all other nodes are observed. The dotted boxes in Figure 1 represent plate notation used in graphical models in which the contents inside the dotted box are replicated as many times as indicated in the bottom right corner. The outer plate represents N sites and the inner plate represents the number of visits T_i to the ith site. In addition, o_i represents the occupancy probability of site i and d_{it} being the true detection probability at site i, visit t. The OD model is parameterized by occupancy parameters α and detection parameters β.

We model the relationship between the occupancy of the ith site (ie. the node Z_i) and the occupancy features X_i at that site using a logistic regression with parameters α. Occupancy features are environmental factors determining the suitability of the site as habitat.

The detection component captures the conditional probability of the observer detecting the species (ie. random
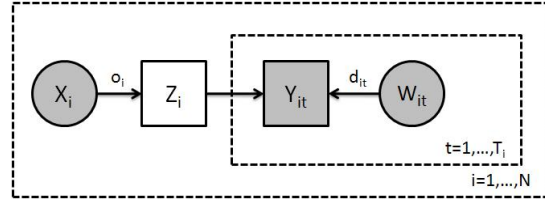

Figure 1. Graphical model representation of the Occupancy-Detection model for a single bird species.

variable Y_{it}), during a visit at site i and at time t conditioned on the site being occupied ie. $Z_i = 1$ and the detection features W_{it}. The detection features include factors affecting the observer's detection ability. We model the detection variable Y_{it} as a function of the detection features using logistic regression with parameters β.

Under the OD model, sites are visited multiple times and observations are made during each visit. The site detection history includes the observed presence or absence of the species on each visit at this site. The OD model makes two key assumptions. First, the population closure assumption [15] assumes that the species occupancy status at a site stays constant over the course of the visits. Second, the standard OD model does not allow for false detections. False detections occur when observers incorrectly declare a species to be present at a site when the site is in fact unoccupied by that species. Hence under the OD model, reporting the presence of a species at a site makes the site occupied by that species. Reporting the absence of a species at a site can be explained by either the site being truly unoccupied or the observer failing to detect the species.

B. The Occupancy-Detection-Expertise Model

The ODE model incorporates birder expertise by extending the OD model in two ways. First, we add to the OD graphical model an expertise component which influences the detection process. Birder expertise strongly influences the detectability of the species, such as when experts are more proficient at identifying certain bird species by sound rather than by sight. The occupancy component of the ODE model stays the same as in the OD model because the site occupancy is independent of the observer's expertise. The second extension we add to the OD model is to allow false detections by both novices and experts. A graphical model representation of the ODE model for a single bird species is shown in Figure 2.

In the expertise component, E_j is a binary random variable capturing the expertise (ie. 0 for novice, 1 for expert) of the jth birder and there are M birders in total. We use logistic regression, with parameters γ, to model E_j as a function of the expertise features U_j associated with the jth birder. Expertise features include features derived from the birder's personal information and history of checklists, such
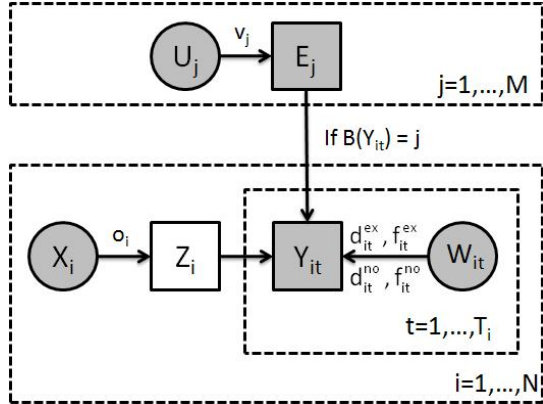
Figure 2. Graphical model representation of Occupancy-Detection-Expertise Model for a single bird species.

as the total number of checklists submitted and the total number of bird species identified on these checklists.

In order to incorporate birder expertise, we modify the detection process such that it consists of a mixture model in which one mixture component models the detection probability by experts and the other mixture component models the detection probability by novices. Each detection probability has a separate set of detection parameters for novices and for experts. These two separate feature sets are useful if the detection process is different for experts versus novices. For instance, experts can be very skilled at identifying birds by sound rather than by sight. Let $B(Y_{it})$ be the index of the birder who submits checklist Y_{it}. In Figure 2, the links from E_j to Y_{it} only exist if $B(Y_{it}) = j$ ie. the jth birder is the one submitting the checklist corresponding to Y_{it}.

In addition, we allow for false detections by both experts and novices. This step is necessary because allowing for false detections by experts and novices improves the predictive ability of the model. Experts are in fact often over-enthusiastic about reporting bird species that do not necessarily occupy a site but might occupy a neighboring site. For instance, experts are much more adept at identifying and reporting birds that fly over a site or are seen at a much farther distance from the current site. As a result, the detection probabilities for novices and experts in the ODE model are now separated into a total of 4 parts: true and false detection probabilities for experts (d_{it}^{ex} and f_{it}^{ex} respectively), and true and false detection probabilities for novices (d_{it}^{no} and f_{it}^{no} respectively). Each of these probabilities is modeled using logistic regression with an associated set of parameters. For more details, we refer the interested reader to the extended version of this paper [19].

C. Parameter Estimation and Regularization

The ODE model requires a labeled set of expert and novice birders to estimate the model parameters using Ex-

pectation Maximization [5]. In the E-step, EM computes the expected occupancies Z_i for each site i using Bayes rule. In the M-step, EM determines the values of parameters that maximize the expected joint log-likelihood in Equation 1; we use L-BFGS [13] to perform the optimization. A more complete description of the parameter estimation process for the ODE model can be found in [19].

$$Q = \mathbb{E}_{P(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{E})}[\log P(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{E}|\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{W})] \quad (1)$$

D. Inference

The ODE model can be used for three main inference tasks: prediction of site occupancy (Z_i), prediction of observations on a checklist (Y_{it}) and prediction of a birder's expertise (E_j). Although ecologists are extremely interested in the true species occupancy at a site, ground truth on site occupancy is typically unavailable. Consequently, we evaluate the ODE model on the latter two inference tasks, which we describe in detail below.

1) Predicting observations on a checklist: To predict Y_{it}, we compute the detection probability $P(Y_{it} = 1|\boldsymbol{X}_i, \boldsymbol{W}_{it}, \boldsymbol{U}_{B(Y_{it})})$. During prediction, we treat the expertise node E_j as a latent variable.

2) Predict birder's expertise: Prediction of birder expertise can alleviate the burden of manually classifying new birders as experts and novices. Let $\boldsymbol{Y}^j$ be the set of checklists that belong to birder j (with Y_{it}^j and $\boldsymbol{Y}_{i\cdot}^j$ extending our previous notation), let $\boldsymbol{W}_{it}^j$ be the detection features for Y_{it}^j and let $\boldsymbol{Z}^j$ be the set of sites at which birder j submitted checklists. We treat $\boldsymbol{Z}^j$ as latent variables during prediction and marginalize them out. To predict the expertise of birder j, we compute $P(E_j = 1|\boldsymbol{X}, \boldsymbol{Y}^j, \boldsymbol{W}, \boldsymbol{U}_j)$.

III. EVALUATION

In this section, we evaluate the ODE model over two prediction tasks: predicting observations on a birder's checklist and predicting the birder's expertise level based on the checklists submitted by the birder. In both evaluation tasks, we report the area under the ROC curve (AUC) as the evaluation metric. We also include results from a contrast mining task that illustrates the utility of the ODE model.

A. Data description

The eBird dataset consists of a database of checklists associated with a geographic site. Each checklist belongs to a specific birder and one checklist is submitted per visit to a site by a birder. In addition, each checklist stores the counts of all the bird species observed at that site by that birder. We convert the counts for each species into a Boolean presence/absence value. A number of other features are also associated with each site-checklist-birder combination: 1) the occupancy features associated with each site, 2) the detection features associated with each observation, and 3) the expertise features associated with each birder. The

observation history of each birder is used to construct two expertise features – the total number of checklists submitted and the total number of bird species identified. We use 19 occupancy features, 3 detection features and 2 expertise features in the experiment. For more details on these features, we refer the reader to [19] and [16].

In our experiments we use eBird data from New York state during the breeding season (May to June) in years 2006-2008. We choose the breeding season because many bird species are more easily detected during breeding and because the population closure assumption is reasonably valid during this time period. Furthermore, we group the checklists within a radius of 0.16 km of each other into one site and each checklist corresponds to one visit at that grouped site. The radius is set to be small so that the site occupancy is constant across all the checklists associated with that grouped site. Checklists associated with the same grouped site but from different years are considered to be from different sites. Ornithologists working with the eBird project at the Cornell Lab of Ornithology hand-labeled the expertise of birders in our training set using a variety of criterion including personal knowledge of birder reputation, number of checklists rejected during data verification, and manual inspection of eBird checklists. This training set consists of 32 expert and 88 novice birders with 2352 and 2107 total checklists respectively.

There are roughly 400 bird species that have been reported over the NY state area. Each bird species can be considered a different prediction problem. We evaluate our results over 3 groups with 4 bird species in each group. Group A consists of common bird species that are easily identified by novices and experts alike. Group B consists of bird species that are difficult for novices to detect; most of these birds are detected by sound rather than by sight. Finally, Group C consists of two pairs of birds – Hairy and Downy Woodpeckers and Purple and House Finches. Novices typically confuse members of a pair for each other.

B. Task 1: Prediction of observations on a checklist

Since the *occupancy* status of the site Z_i is not available, we can use the *observation* of a bird species as a substitute. We evaluate the accuracy of the ODE model when predicting detections against a Logistic Regression (LR) model and the classic OD model found in the ecology literature.

Evaluating predictions on spatial data is a challenging problem due to two key issues. First, a non-uniform spatial distribution of the data introduces a bias in which small regions with high sampling intensity have a very strong influence on the performance of the model. Secondly, spatial autocorrelation allows test data points that are close to training data points to be easily predicted by the model. To alleviate the effects of both of these problems, we superimpose a 9-by-16 checkerboard (each grid cell is roughly a 50 km x 33 km rectangle) over the data. The checkerboard grids

the NY state region into black and white cells. Data points falling into the black cells are grouped into one fold and those falling into the white cells are grouped into another fold. The black and white sets are used in a 2-fold cross validation. We also randomize the checkerboarding by randomly positioning the bottom left corner to create different datasets for the two folds. We run 20 such randomization iterations and within each iteration, we perform a 2-fold cross validation. We compute the average AUC across all 20 runs and show the results in Table I. Boldface indicates the best results. The $\star$ and $\dagger$ symbols indicate that the ODE model is a statistically significant improvement (paired t-test, $\alpha = 0.05$) over the LR and OD models respectively.

We use a validation set to tune the regularization terms of three different models. Data in one fold is divided into a training set and a validation set by using a 2-by-2 checkerboard on each cell. More specifically, each cell is further divided into a 2-by-2 subgrid, in which the top left and bottom right subgrid cells are used for training and the top right and bottom left subgrid cells are used for validation.

1. LR Model: A typical machine learning approach to this problem is to combine the occupancy and detection features into a single set of features. Since we are interested in the benefit of distinctly modeling occupancy and detection by having occupancy as a latent variable, we use this LR model as a baseline as it does not separate occupancy from detection. We use two LR models for our baseline. The first LR model predicts the birder's expertise using the birder's expertise features. The probability of the birder being an expert is then treated as a feature associated with each checklist from that birder. The second LR predicts the detection Y_{it} using the occupancy features, detection features and the expertise probability computed from the first LR.

2. OD Model: In order to incorporate birder expertise in the OD model, we also employ a LR to predict the birder expertise from the expertise features. We treat the probability of the birder being an expert as another detection feature associated with each checklist from that birder. Then, we use EM to train the OD model. To predict a detection, we first compute the expertise probability using coefficients from the first LR and then predict the detection using the occupancy features, detection features and the predicted expertise as an additional detection feature.

3. ODE Model: The ODE model is trained using EM and we predict Y_{it} as before. The birder expertise is observed during training but unobserved during testing.

C. Task 2: Prediction of birder's expertise

In this experiment, we compare the ODE model with LR to predict the birder's expertise.

1. LR Model: To train a LR to predict a birder's expertise, every checklist is treated as a single data instance. The set of features for each data instance include occupancy features, detection features, and expertise features. To predict

Table I
AVERAGE AUC FOR PREDICTING DETECTIONS ON TEST SET
CHECKLISTS FOR BIRD SPECIES

Group A Bird Species	LR	OD	ODE
Blue Jay	0.6726	0.6881	**0.7104***†
White-breasted Nuthatch	0.6283	0.6262	**0.6600***†
Northern Cardinal	0.6831	0.7073	**0.7085***
Great Blue Heron	0.6641	0.6691	**0.6959***†
Group B Bird Species	LR	OD	ODE
Brown Thrasher	0.6576	0.6920	**0.6954***
Blue-headed Vireo	0.7976	0.8055	**0.8325***†
Northern Rough-winged Swallow	0.6575	0.6609	**0.6872***†
Wood Thrush	0.6579	0.6643	**0.6903***†
Group C Bird Species	LR	OD	ODE
Hairy Woodpecker	0.6342	0.6283	**0.6759***†
Downy Woodpecker	0.5960	0.5622	**0.6183***†
Purple Finch	0.7249	0.7458	**0.7659***†
House Finch	0.5725	0.5809	**0.6036***†

Table II
AVERAGE AUC FOR PREDICTING BIRDER EXPERTISE ON A TEST SET OF
BIRDERS FOR BIRD SPECIES

Group A Bird Species	LR	ODE
Blue Jay	0.7265	**0.7417***
White-breasted Nuthatch	**0.7249**	0.7212
Northern Cardinal	0.7352	**0.7442**
Great Blue Heron	0.7472	**0.7661**
Group B Bird Species	LR	ODE
Brown Thrasher	0.7523	**0.7761***
Blue-headed Vireo	0.7869	**0.7981**
Northern Rough-winged Swallow	0.7792	**0.8052***
Wood Thrush	0.7675	**0.7937***
Group C Bird Species	LR	ODE
Hairy Woodpecker	0.7056	**0.7334***
Downy Woodpecker	0.7223	**0.7307**
Purple Finch	0.7481	**0.7739***
House Finch	0.7279	**0.7403***

Table III
AVERAGE Δ_{TD} FOR GROUP A AND B.

Group A Bird Species	Average Δ_{TD}
Blue Jay	0.0118
White-breasted Nuthatch	0.0077
Northern Cardinal	-0.0218
Great Blue Heron	0.0110
Group B Bird Species	Average Δ_{TD}
Brown Thrasher	0.1659
Blue-headed Vireo	0.1158
Northern Rough-winged Swallow	0.1618
Wood Thrush	0.0954

the expertise of a new birder, we first retrieve the checklists submitted by the birder, use LR to predict the birder's expertise on each checklist, and then average the predictions of expertise on each checklist to give the final probability.

2. ODE Model: The ODE model is trained using EM and we predict the birder's expertise using the model.

We evaluate on the same twelve bird species using a 2-fold cross validation across birders. We randomly divide the expert birders and novice birders into half so that we have an equal number of expert birders as well as novice birders in the two folds. Assigning birders to each fold will assign checklists associated with each birder to the corresponding fold. We use a validation set to tune the regularization terms of both the LR model and the ODE model. Of all birders in the training fold, half of the expert birders and the novice birders in that fold are randomly chosen as the actual training set and the other half serve as the validation set. Finally, we run 2-fold cross validation on the two folds and compute the AUC. For each bird species, we perform the 2-fold cross validation using 20 different random splits for the folds. In Table II we tabulate the mean AUC for each species, with boldface entries indicating the best results and $\star$ indicating that the ODE model is a statistically significant improvement (paired t-test, $\alpha = 0.05$) over LR.

D. Task 3: Contrast mining

In this contrast mining task, we identify bird species that are over/under reported by novices compared to experts. We compare the average Δ_{TD} values for Groups A and B, where Δ_{TD} is the difference of the true detection probabilities between expert and novice birders. We expect experts and novices to have similar true detection probabilities on species from Group A, which correspond to common, easily identified bird species. For Group B, which consists of species that are hard to detect, we expect widely different true detection probabilities. In order to carry out this case study, we first train the ODE model over all the

data described in Subsection III-A for a particular species. Then for each checklist, we compute the difference between the expert's true detection probability and the novice's true detection probability. We average this value over all the checklists. The results are shown in Table III.

IV. DISCUSSION

Since true site occupancies are typically not available for real-world species distribution data sets, predicting species observations at a site is a reasonable substitute for evaluating the performance of a SDM. Table I indicates that the top performing model over all 12 species is the ODE model. The ODE model offers a statistically significant improvement over LR in 12 species and over the OD model in 10 species. The two main advantages that the OD model has over LR are that it models occupancy separately from detection and it allows checklists from the same site i to share evidence through the latent variable Z_i. However, in 3 species, the OD model performs worse than the LR model. This decrease in AUC is largely due to the fact that the OD model does not allow for false detections. In contrast to the OD model, the ODE model allows for false detections by both novices and experts and it can incorporate the expertise of the observer into its predictions. Since the ODE model consistently outperforms the OD model, the improvement in accuracy is mainly due to these two advantages.

As shown in Table II, the ODE model outperforms LR on all species except for *White-breasted Nuthatch* when predicting expertise. The ODE model's results are statistically significant for almost all the Group B birds species, which are hard to detect, but not significant for Group A birds, which are much more obvious to detect. For Group C, the ODE model results are statistically significant for Hairy Woodpeckers, Purple Finches and House Finches. These results are consistent with behavior by birders. Both Purple Finch and Hairy Woodpeckers are rarer and experts are better at identifying then. In contrast, novices often confuse House Finches for Purple Finches and Downy Woodpeckers for Hairy Woodpeckers. Overall, the AUCs for most species are within the 0.70-0.80 range, which is an encouraging result for using the ODE model to predict birder expertise.

Finally, the results in Table III indicate that experts and novices appear to have very similar true detection probabilities for the common bird species in Group A. However, for the hard-to-detect bird species in Group B, the Δ_{TD} values are much larger. These results show that the ODE model is a promising approach for contrast mining, which can identify differences in how experts and novices report bird species.

V. CONCLUSION

We have presented the ODE model that has distinct components that capture occupancy, detection and observer expertise. We have shown that it produces more accurate predictions of species detections and birder's expertise than other models. More importantly, we can use this model to find differences between expert and novice observations of birds. This knowledge can be used to inform citizen scientists who are novice birders and thereby improve the reliability of their observations.

REFERENCES

[1] M. P. Austin. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Modell.*, 157:101–118, 2002.

[2] G. Carpenter, A. N. Gillison, and J. Winter. Domain: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, 2:667–680, 1993.

[3] J. P. Cohn. Citizen science: Can volunteers do real research? *BioScience*, 58(3):192–197, 2008.

[4] D. G. Delaney, C. D. Sperling, C. S. Adams, and B. Leung. Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions*, 10(1):117–128, 2008.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JRSS, Series B*, 39(1):1–38, 1977.

[6] J. Elith and J. Leathwick. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics*, 40:677–697, 2009.

[7] M. C. Fitzpatrick, E. L. Preisser, A. M. Ellison, and J. S. Elkinton. Observer bias and the detection of low-density populations. *Ecological Applications*, 19(7):1673–1679, 2009.

[8] J. H. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Ann. Stat.*, 28:337–407, 2000.

[9] M. Kéry, B. Gardner, and C. Monnerat. Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, 37:1851–1862, 2010.

[10] M. Kéry, J. A. Royle, H. Schmid, M. Schaub, B. Volet, G. Häfliger, and N. Zbinden. Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, 24(5):1388–1397, 2009.

[11] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge, MA, 2009.

[12] A. M. Latimer, S. Wu, A. E. Gelfand, and J. John A. Silander. Building statistical models to analyze species distributions. *Ecological Applications*, 16(1):33–50, 2006.

[13] D. C. Liu and J. Nocedal. On the limited memory method for large scale optimization. *Mathematical Programming B*, 45(3):503–528, 1989.

[14] D. I. Mackenzie, J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84(8):2200–2207, 2003.

[15] D. I. MacKenzie, J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255, 2002.

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

[17] S. J. Phillips, M. Dudik, and R. E. Schapire. A maximum entropy approach to species distribution modeling. In *Proceedings of the 21st ICML*, pages 83–91, 2004.

[18] D. Stockwell and D. Peters. The garp modelling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inform. Sci.*, 13:143–158, 1999.

[19] J. Yu, W.-K. Wong, and R. Hutchinson. Modeling experts and novices in citizen science data for species distribution modeling. Technical report, Oregon State University, 2010. http://hdl.handle.net/1957/18806.