Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models

Orin J. Robinson1 | Viviana. Ruiz-Gutierrez1 | Mark D. Reynolds2 | Gregory H. Golet3 | Matthew Strimas-Mackey1 | Daniel Fink1

Abstract

Aim: Information on species’ habitat associations and distributions, across a wide range of spatial and temporal scales, is a fundamental source of ecological knowledge. However, collecting information at relevant scales is often cost prohibitive, although it is essential for framing the broader context of more focused research and conservation efforts. Citizen science has been signalled as an increasingly important source to fill in data gaps where information is needed to make comprehensive and robust inferences on species distributions. However, there are perceived trade-offs of combining highly structured, scientific survey data with largely un-structured, citizen science data.

Methods: We explore these trade-offs by applying a simplified approach of filtering citizen science data to resemble structured survey data and analyse both sources of data under a common framework. To accomplish this, we integrated high-resolution survey data on shorebirds in the northern Central Valley of California with observations in eBird for the entire region that were filtered to improve their quality.

Results: The integration of survey data with the filtered citizen science data resulted in improved inference and increased the extent and accuracy of distribution models on shorebirds for the Central Valley. The structured surveys improved the overall accuracy of ecological inference over models using citizen science data only by increasing the representation of data collected from high-quality habitats for shorebirds.

Main conclusions: The practical approach we have shown for data integration can also be used to improve the efficiency of designing biological surveys in the context of larger, citizen science monitoring efforts, ultimately reducing the financial and time expenditures typically required of monitoring programs and focused research. The simple method we present can be used to integrate other types of data with more localized efforts, ultimately improving our ecological knowledge on the distribution and habitat associations of species of conservation concern worldwide.

Keywords
citizen science, data filtering, data integration, data quality, ecological niche model, species distribution model, structured survey
1 | INTRODUCTION

Information on species’ habitat associations and distributions is a fundamental source of ecological knowledge (Sofaer et al., 2019). This information is often of interest across a broad range of spatial and temporal scales, from high-resolution information that is more relevant for research on habitat selection (Matthiopoulos, Hebblewhite, Aarts, & Fieberg, 2011) or needed to inform management objectives (Zipkin, Andrew Royle, Dawson, & Bates, 2010) to larger-scale inferences that are useful to address broader questions (e.g. potential range shifts with changing climatic conditions; Lyon, Debinski, & Rangwala, 2019). However, the process of collecting biological observations across large spatial scales is often cost prohibitive for most research and monitoring efforts. At best, researchers and practitioners are able to monitor plant and animal communities just within their study regions, and during specific times of year. However, the need to make inferences beyond the sampled range of environmental conditions and seasons often limits our understanding of the broader context of our results and can limit the use of applied research to inform future monitoring efforts and effective conservation actions.

Citizen science data have been signalled as a promising source of information to fill in information gaps needed to model species distributions (Bradter et al., 2018; Gouraguine et al., 2019). The collection of citizen science data is growing rapidly, and for a number of taxa, large databases of observations exist. For example, the National Moth Recording Scheme in Great Britain has collected over 11 million observations (Fox et al., 2014). There are many more taxa for which rich citizen science data sets exist, for example birds (Sauer et al., 2017; Sullivan et al., 2014), invertebrates (Howard, Aschen, & Davis, 2010), bats (Newson, Evans, & Gillings, 2015), whales (Tonachella, Natasi, Kaufman, Maldini, & Rankin, 2012) and frogs (Westgate et al., 2015) among others. However, there are few examples of the potential trade-offs of combining observational data collected at smaller spatial extents with citizen science data collected across vast extents. This might be due in part to inherent differences that often exist between the two data types. On the one hand, you may have data that is collected at a high spatial resolution using skilled observers, sampling effort is often standardized, and sampling occurs across a habitat gradient that is representative of the region of interest. On the other hand, is citizen science data, which can be collected across a wide range of sampling conditions by observers that vary widely in their level of expertise, collected across a wide range of spatial resolutions, and sampling effort is not standardized. This has created a perceived trade-off between data quality and quantity among data collected from structured, scientific surveys and data collected from larger-scale, volunteer-based monitoring efforts (Figure 1). The assumption being made is that an increase in quantity of citizen science data comes at a significant cost to quality. Therefore, the logical framework to integrate these two sources of information would be one that would treat them as independent sources of information used to inform a common underlying distribution for a given species (Pacifici et al., 2017).

The integration of different data sources is a growing area of methodological development in ecological statistics, and recent advances have been made to develop ways of integrating survey data (e.g. structured) with citizen science data (e.g. un-structured data) (Miller, Pacifici, Sanderlin, & Reich, 2019). For observational data collected at discrete locations, these methods include specifying a joint likelihood for the two data sources to estimate the underlying species distribution (Miller et al., 2019). In cases where this is not possible, the data source that is deemed as of lower quality (e.g. citizen science data, museum observations) can be used in two ways: (a) modelled as a covariate of the underlying distribution or (b) used to estimate a separate species distribution, where a correlation structure is specified to share information across data sources. Pacifici et al. (2017) tested these different approaches to integrate observational data from the citizen science project eBird (Sullivan et al., 2014) with more structured data from the North American Breeding Bird Survey (BBS; Sauer et al., 2017). Their results showed that the joint-likelihood approach of combining eBird and BBS data outperformed all other approaches, including using BBS data alone.

The approach used by Pacifici et al. (2017) and Miller et al. (2019) summarized observational data at a coarse, grid-level in order to account for differences in effort, sampling approach, and other variables that are known to influence detectability (GuillerArroita, 2017). In addition, Pacifici et al. (2017) wanted to reduce potential bias related to the degree of uncertainty about the spatial

![Figure 1](image-url)
scale that observations were collected for each independent eBird checklist. This mismatch in scales between the two data sources is what often makes data integration between high-resolution survey data (e.g. point count observations) with lower-resolution citizen science data non-trivial. However, many citizen science programs collect high-scale resolution information (e.g. camera traps) in ways that we can infer absences, and collect additional information on effort (e.g. distance travelled, number of hours sampled) that is highly valuable for improving the accuracy of SDMs (Kelling et al., 2019).

Here, we explore a practical approach for data integration between high-quality, citizen science data with structured survey data that builds upon existing methods for “data pooling” (e.g. Fithian, Elith, Hastie, & Keith, 2015). We explore the trade-offs in inference of using citizen science data alone, more localized and structured data alone, and pooling together both data sets combined. In addition, we examine specific trade-offs when combining structured and un-structured data sources by exploring the performance of increasing the quantity of citizen science data through simulations, versus. the addition of data from more targeted survey efforts. To accomplish this, we use The Nature Conservancy’s (TNC) BirdReturns project as a case study (Reynolds et al., 2017). We explore ways of combining high-resolution bird survey data collected for shorebirds on rice fields in the northern region of the Central Valley in California, with observations in eBird for the entire Central Valley that are filtered to improve their quality. The specific aim of the case study is to provide a framework for leveraging survey data with citizen science data to build more accurate distribution models for shorebird species across the extent of the Central Valley.

2 | METHODS

2.1 | Data

We used point counts carried out during spring surveys (February 1–May 31; n = 8,192) as part of the TNC BirdReturns project conducted in 2014–2017. This project used predicted shorebird occurrence and abundance (Johnston et al., 2015) along with predicted surface water in the Sacramento River Valley to identify times and locations that were likely to be important for migrating shorebirds. TNC used a reverse auction approach to select and incentivize rice farmers in the identified locations to flood their fields during the spring and fall, making temporary wetlands available to the migrating shorebirds. Observers made point counts at fields enrolled in the program and at unenrolled control sites, surveying a semi-circle with a 200 m fixed radius. Each site was surveyed for at least two minutes and lasted as long as necessary to count all birds present (for more detail on count methods see Golet et al., 2018). Effort for each point count consisted of date, time started and ended, and name of observer.

We combined the point counts with data from the citizen science project eBird (Sullivan et al., 2014) collected during the same time period as the point counts. The eBird data were restricted to the Central Valley of California, USA, and to complete checklists so that non-detection could be inferred (Johnston et al., 2019). We also restricted the eBird data to stationary checklists and travelling checklists limited to 300 m. After filtering these data, we were left with 12,891 checklists. Effort variables for the eBird data set were date, time observations started, duration of observation in minutes, survey protocol (stationary or travelling), distance travelled in metres and number of observers.

We calculated the effort variables from the point counts to match those of the eBird data set. Observer name in the point count data set was converted into number of observers (however, it was always one for this study), time started and ended for the point counts was used to calculate duration in minutes, and each point count was treated as a stationary count (distance travelled = 0). By doing this, the two data sets contained the same effort information and were identical in structure which allowed us to simply join them into one combined data set. We added a variable to the combined data set to note whether an observation was a TNC point count or an eBird checklist. We also calculated a checklist calibration index (CCI) for each checklist in the combined data set to account for variation in expertise among observers (e.g. expertise score; Johnston, Fink, Hochachka, & Kelling, 2018). As environmental variables, we attached the remotely sensed Cropland Data Layer (CDL; Boryan, Yang, Mueller, & Craig, 2011; Han, Yang, Di, & Mueller, 2012) to the combined data set by computing the per cent of each land cover or crop type in the CDL that was present within a 300 m radius centred on each point count or eBird checklist. We also similarly attached cloud-filled data from Water Tracker (Reiter, Elliott, Barbaree, & Moody, 2018), a high spatial and temporal resolution surface water tracking system for the Central Valley of California.

A comprehensive data set was created using all of the above eBird checklists (12,891) plus simulated eBird checklists equal to the number of TNC point counts that were included in the combined data set (8,192). This resulted in a data set that was the same magnitude (21,083) as the combined data set and allowed us to determine whether the improvement in accuracy from combining the data sets was simply a function of increasing the sample size. The simulated eBird data were created by adding a small amount of noise (via the jitter function in base R; R Core Team, 2019) to the spatial covariates of 8,192 randomly chosen checklists, similar to the oversampling procedure in Fink et al. (2019); however, we used all checklists rather than only positive observations here. This ensured that we were maintaining roughly the same prevalence rate in the data set and also were not simply making exact copies of the randomly chosen checklists.

2.2 | Spatial filtering and class imbalance

As spatial bias is always a concern when using citizen science data (Geldmann et al., 2016), we spatially subsampled the combined data set. The data were sparse for many species (proportion of detections for a species <0.05 for all checklists in the data set), so class imbalance was also a concern (He & Garcia, 2009). We spatially undersampled the data (e.g. Robinson, Ruiz-Gutierrez, & Fink, 2017) by...
first creating a hexagonal grid of 3.5 km (~10 times the radius of each survey) cells over the region from which our observations came via the dggridR package (Barnes et al., 2018) in R (R Core Team, 2019). This was done to reduce the chance that we selected overlapping observations at each run of each model. We then split the data for a single species into checklists on which the species was detected (positive observations) and those on which the species was not detected (negative observations). We selected one checklist from the negative observations from within each grid cell and recombined the filtered negative observations with the positive observations. As almost all of the spatial bias was from the negative observations (i.e. only a small percentage of the total number of observations were positive observations), this procedure relieves much of the spatial bias, and because only negative observations were filtered out of the data set, class imbalance is also addressed here (King & Zeng, 2001; Robinson et al., 2017).

To alleviate the effects of class imbalance, after spatially sampling eBird checklists for training distribution and population trend models, Fink et al. (2019) oversampled eBird checklists for species that had a prevalence rate of less than 25%. After spatially undersampling our data for the current study, class imbalance was still a concern, as our undersampling improved the imbalance, but it did not improve class balance to 25% for species other than Yellowlegs. Following the recommendation of Fink et al. (2019), we oversampled the positive observations for each of our species and data sets if prevalence was below 25% before training the distribution models. We used the synthetic minority oversampling technique (SMOTE; Chawla, Bowyer, Hall, & Kegelmeyer, 2002) to create one new example of the positive class in the training data set for every positive observation and a nearest neighbour. We did not randomly oversample our data as is recommended when using SMOTE because our data had already been spatially undersampled as described above. The spatial undersampling and SMOTE procedure were done to each of the data sets. As the spatial undersampling randomly chooses from negative observations within a grid cell, many negative observations may not be included in training the model. Therefore, we sampled each of the four data sets in the study using this procedure 100 times, creating 100 unique data sets unless it met the 25% prevalence criteria described above. In that case, the data were only spatially undersampled to create the 100 unique data sets.

2.3 | Analysis

We selected and spatially subsampled (but did not oversample) 15% of the combined data set to be the testing data for evaluation of each model. This test set was selected because our goal is to make accurate predictions across the entire Central Valley and give equal importance to each location. Therefore, the test set must include observations across the spatial extent of evaluation and be spatially balanced. For the training data sets, we removed any checklist or point count that was in the test set. We repeated this process 100 times creating 100 unique data sets against which our models were tested.

We used the R package “ranger” (Wright, Wager, Probst, & Maintainer, 2019) to train a random forest model for each of the seven species (or combined species; Table 1) and for each of four data sets: (a) TNC point counts alone, (b) eBird checklists alone, (c) an oversampled eBird data set, and (d) the combined TNC and eBird data set. For each species, 1,000 trees were grown in the ensemble. The number of variables from which the model could select at each split for each tree was initially set to the square root of the number of variables included in the model \( n = 12 \approx \sqrt{142} \); however, we allowed this to vary by half, double and triple this common rule of thumb. We selected the output from the model that maximized accuracy (Data S1). We evaluated the accuracy of the models using multiple predictive performance metrics (PPMs). We did not

| TABLE 1 | List of species for which we modelled spring distribution in the Central Valley of California with each of the four data sets |
| --- | --- | --- |
| Common Name | Latin Name | Abbreviation in Tables & Figures |
| American Avocet | Recurvirostra americana | AMAV |
| Dunlin | Calidris alpina | DUNL |
| Greater Yellowlegs* | Tringa melanoleuca | YLEG* |
| Least Sandpiper | Calidris minutilla | LESA |
| Lesser Yellowlegs* | Tringa flavipes | YLEG* |
| Long-billed Curlew | Numenius americanus | LBCU |
| Long-Billed Dowitcher* | Limnodromus scolopaceus | DOWI* |
| Short-billed Dowitcher* | Limnodromus griseus | DOWI* |
| Western Sandpiper | Calidris mauri | WESA |

*Combined in analysis and referred to collectively as “Yellowlegs” in this study (e.g. Golet et al., 2018)

*Combined in analysis and referred to collectively as “Dowitcher” in this study (e.g. Golet et al., 2018)
evaluate the models using the out of bag (OOB) metric calculated by random forest internally as it favours total accuracy over correctly predicting the minority class and is highly sensitive to class imbalance. We used the test data set to evaluate mean squared error (MSE) between the model predictions of presence or absence and the true presence or absence in the test set. We also evaluated error using Brier score (Brier, 1950), the mean squared error of the probabilistic model predictions and the true presence or absence in the test set. We evaluated each model’s ability to rank positive observations higher than negative ones using the area under the curve (AUC; Fielding & Bell, 1997). We evaluated each model’s ability to predict presence or absence using Cohen’s Kappa (Kappa; Cohen, 1960) and its components, sensitivity (true positive rate), and specificity (true negative rate). We produced distribution maps for each species and data set, and we recorded the predictor importance metrics from models trained on each data set. All variables included in the model may be found in Data S2.

3 | RESULTS

The fields that are stored as part of the eBird project allowed us to filter the data to be of a similar protocol as the TNC point counts. This filtering reduced uncertainty in the location of eBird checklists and eliminated the need for coarse level summaries of the data for integration (Miller et al., 2019; Pacifi ci et al., 2017). For all species, the combined data set had higher predictive accuracy than either the TNC point counts or the eBird checklist data sets on their own; however, error (MSE, Brier score) was relatively low and AUC was relatively high for all species and all data sets, particularly for the three data sets where eBird checklists were included (Figure 2; Figure S3). Improvement in accuracy varied among species, however; for the Kappa statistic (predicting presence and absence against the test set), the combined data set was an improvement over all three of the other data sets evaluated (with exception of western sandpiper; −3%–25.5% improvement; Figure 3). The improvement or loss in the error statistics was usually negligible (with the exception of LBCU; Figure 4; Figure S4) for the combined data set versus the next best data set; however, error metrics were already very low for most species, so great improvement here was not likely. Likewise, AUC was relatively high for most species and for the three data sets containing eBird checklists; therefore, the gain was negligible for many species (improvement of up to 6%; Figure S4). For the few species/metrics combinations where the combined data set was not the best performing model, it was the second best, with the best being the eBird checklists with simulated eBird checklists added. Neither TNC

---

**FIGURE 2** Box plots for MSE, AUC, Brier score, Cohen’s Kappa, Sensitivity and Specificity for models run for Least Sandpiper (LESA) with each data set: TNC point counts (TNC), eBird checklists (eBird), eBird checklists with additional simulated eBird data (eBird + sim), and the data set of TNC point counts and eBird checklists combined (Combined).
point counts, nor eBird checklists alone performed better for any metric than the data set where the two were combined.

We produced distribution maps for each species (Figure 5; Figure S5) to determine whether there was a difference in the overall pattern of distribution estimated by models trained on the different data sets. For most species, there was greater contrast between the presence estimates and absence estimates for the combined data set when compared to the other three. Visually, this means more obvious differentiation between high predicted probability of presence and absence (e.g. hotter “hotspots” and darker regions where absence is predicted; Figure 5; Figure S5). We collected the important variables identified by the model (via Gini index) when run with each data set. For all species, the importance of rice and water was apparent as it was among the most important variables for each of the data sets; however, it was not until the data sets were combined that each rice and the Water Tracker layer had a high importance score (Figure 6; Figure S6). The combination of the data allowed the models to home in on rice and surface water as highly important, where they had only moderate importance comparatively when using the other data sets. This is likely because only 1.5% of eBird checklists in our study come from locations where the per cent landcover is >50% rice. Conversely, almost 60% of the TNC point counts come from locations where the per cent landcover is >50% rice.
Discussion

Our results lend further support to efforts looking to combine data from multiple sources to improve the inference and/or predictive ability of distribution models (Miller et al., 2019; Pacifici et al., 2017). We have shown that citizen science data can be filtered to generate a high-quality data set that can closely match the resolution and sampling approach of structured surveys, supporting the call for current and future citizen science projects to collect essential information related to location and effort, as well as complete surveys (Kelling et al., 2019). The integration of survey data with the filtered citizen science data in eBird resulted in improved inference, predictive ability, and ultimately increased the extent of inference of the structured surveys. In turn, the structured surveys were able to improve the ecological inference of the citizen science data, by improving the representation of sampled habitats that are key for shorebird species. Most importantly, the practical approach we have shown for data integration is an improvement on simpler “data pooling” approaches for data integration and can be used to improve the efficiency of designing biological surveys to collect distribution information in the context of larger, citizen science monitoring efforts, ultimately reducing the financial and time expenditures typically required of monitoring programs and focused research (Reich, Pacifici, & Stallings, 2018).

The combined data set resulted in improved accuracy across all metrics relative to the TNC survey data or eBird data alone, for all of the species considered in this study. Our combined approach also predicted presence/absence via agreement with a test data set more accurately than the different permutations of data sets considered (e.g. TNC alone, eBird alone and eBird plus simulated eBird). The observed improvement in accuracy is not likely a function of simply increasing sample size, as augmenting eBird data with simulated data did not show the same levels of improvement. The addition of the TNC survey data to eBird data improved the coverage of the data both spatially and in targeted habitats. Previous work has shown that migrating shorebirds heavily use flooded rice fields in the Central Valley, and will also use unflooded rice fields (Elphick & Oring, 1998; Golet et al., 2018). Rice fields are the main focal habitat of the TNC BirdReturns program, ~60% of the survey data from this project was carried out at sites where at least 50% of the landcover is rice. In contrast, the majority of the rice fields in the Central Valley are not accessible to regular eBird participants, as they are often on private lands. This is likely why we see so few (~1.5%) eBird checklists from the Central Valley in locations where at least 50% of the landcover is rice. The addition of more data from high-quality habitat is what provided the improvement in accuracy of the combined eBird and TNC data set. This highlights the importance and value of more targeted research and survey efforts within the context of large-scale citizen science monitoring efforts. Given that private lands make up more than half of the land in the United States, supporting wildlife monitoring efforts on privately held lands that are linked into large-scale efforts such as eBird can greatly improve inferences on species distributions and habitat associations across scales of interest for all stakeholders (Hilty & Merenlender, 2003).

Interestingly, models based only on the TNC point counts struggled to learn where each species was likely to be absent across the entire Central Valley because the data came largely from “good” shorebird habitat in the northern portion of the Central Valley. The original intent of this project was not focused on species distribution modelling, so this does not come as a surprise. However, given that absence information allows for more accurate distribution models (Brotos, Thuiller, Araújo, & Hirzel, 2004), the addition of eBird data was able to provide information on where species are likely to be absent, and improved inferences on what habitat types most benefit shorebirds, but were not surveyed as part of the TNC monitoring efforts. On the other hand, the models using the eBird data alone were able to predict absences well and had relatively high accuracy when predicting presence/absence, but overall ecological inference was improved when combined with the point count data set. The TNC point counts acted as targeted surveys in under-surveyed habitat, which previous work has shown can improve the accuracy of distribution models using eBird checklists (e.g. Xue, Davies, Fink, Wood, & Gomes, 2016). The complementary nature of the data sets is also shown when examining the important predictors. For Dunlin (Figure 6), the rice landcover and the Water Tracker layer are important predictors for each data set, however, their importance values more than double when the combined data set is used to train the model. While other species do not show such a drastic shift in the importance value for these two habitat variables, the pattern is similar for most in that these variables become more important to the model’s predictions when the data sets are combined.

The approach we present for combining data is applicable for other conservation monitoring programs and ongoing research efforts, given that a significant amount of research is conducted over small spatial and temporal scales (Heidorn, 2008), and often difficult to scale-up without similarly structured data (Poisot, Bruneau, Gonzalez, Gravel, & Peres-Neto, 2019). The data fields that exist in eBird data facilitated further processing and filtering to match the structure of the individual data set of interest and allowed us to leverage the strengths of both data sources. However, even citizen science data sets that do collect effort information are often lacking information on sampling locations, although incentivizing participants to collect data in these data-poor locations has been shown to improve the accuracy of distribution models (e.g. avicaching; Xue et al., 2016). Similarly, data from smaller-scale, individual research projects can also help fill in these gaps in citizen science data. Given that inferences from small-scale studies cannot be extrapolated to larger spatial extents (Sandel & Smith, 2009), the approach we present here for combining data from small-scale studies with citizen science data filtered to match the existing data structure will increase the overall extent of inference, and improve our ability to conceptualize conservation actions within the larger context of the target population(s) of interest.

Information on species distributions across large scales is one of the most fundamental information needs for basic and applied research fields in ecology. However, this level of information often requires large-scale, coordinated surveys that can be time consuming...
FIGURE 5  Predicted probability of an expert surveyor detecting a Dunlin in the spring of 2016 on a one-hour long checklist where the distance travelled was < 300m., as estimated by the model using TNC point counts alone (A), eBird checklists alone (B), eBird checklists with added simulated eBird checklists (C), and the combined data set of TNC point counts and eBird checklists.
and costly to manage. In addition, models used to estimate species distributions are often data hungry, and are often unable to generate information at the spatial and temporal scales that are most relevant for research and conservation efforts. Citizen science data is a growing source of additional, minimal cost surveys for a number of taxa including invertebrates, bats, marine mammals, lichens, amphibians and more (Deutsch, Bilenca, & Agostini, 2017; Devictor, Whittaker, & Beltrame, 2010; Howard et al., 2010; Newson et al., 2015; Tonachella et al., 2012; Westgate et al., 2015). We have shown the utility of combining survey data with semi-structured citizen science data (Kelling et al., 2019) for improving accuracy in species distribution models, which can result in more efficient and cost-effective surveys (Miller et al., 2019; Pacifici et al., 2017; Reich et al., 2018).

The simple method we present for citizen science data allows for the integration of a small-scale point count data set with eBird checklists, can be used to integrate similar types of data being collected by citizen scientists (e.g. camera traps) with more localized efforts (e.g. patrolling by park rangers), ultimately improving our ecological knowledge on the distribution and habitat associations of species of conservation concern worldwide.

ACKNOWLEDGEMENTS
We would like to thank Point Blue for supplying Central Valley water data. We would also like to thank Tom Auer, Ali Johnston, Steve Kelling and Matt Reiter for comments that improved this project and the manuscript.

DATA AVAILABILITY STATEMENT
The Nature Conservancy point count data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.724v

The eBird data used for analyses in this manuscript may be downloaded from https://ebird.org/data/download

Cropland Data Layer may be downloaded from: https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php

Water Tracker Data may be downloaded from Point Blue: https://data.pointblue.org/apps/awater/

ORCID
Orin J. Robinson https://orcid.org/0000-0001-8935-1242

REFERENCES
Barnes, R., Sahr, K., Evenden, G., Johnson, A., Warmerdam, F., & Maintainer, J. (2018). Package “dggridR” Type Package Title Discrete Global Grids. Retrieved from https://github.com/r-barnes/dggridR/

Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: The US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, 26(5), 341–358. https://doi.org/10.1080/10106049.2011.562309

Bradter, U., Mair, L., Jönsson, M., Knappe, J., Singer, A., & Snäll, T. (2018). Can opportunistically collected Citizen Science data fill a data gap for habitat suitability models of less common species? *Methods in Ecology and Evolution*, 9(7), 1667–1678. https://doi.org/10.1111/2041-210X.13012

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Brotons, L., Thuiller, W., Araújo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4), 437–448. https://doi.org/10.1111/j.0906-7590.2004.03764.x

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Keegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953
Sandel, B., & Smith, A. B. (2009). Scale as a lurking factor: Incorporating scale-dependence in experimental ecology. Oikos, 118(9), 1284-1291. https://doi.org/10.1111/j.1600-0706.2009.17421.x
Sauer, J. R., Pardieck, K. L., Ziolkowski, D. J., Smith, A. C., Hudson, M.-A.-R., Rodriguez, V., & Link, W. A. (2017). The first 50 years of the North American Breeding Bird Survey. The Condor, 119(3), 576–593. https://doi.org/10.1650/condor-17-83.1
Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., … Hamilton, H. (2019). Development and Delivery of Species Distribution Models to Inform Decision-Making. BioScience, 69(7), 544–557. https://doi.org/10.1093/biosci/biz045
Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., … Kelling, S. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation, 169, 31–40. https://doi.org/10.1016/j.biocon.2013.11.003
Tonachella, N., Natasi, A., Kaufman, G., Maldini, D., & Rankin, R. W. (2012). Predicting trends in humpback whale (Megaptera novaeangliae) abundance using citizen science. Pacific Conservation Biology, 18, 297-309. https://doi.org/10.1071/PC120297
Westgate, M. J., Scheele, B. C., Ikin, K., Hoefer, A. M., Beaty, R. M., Evans, M., … Driscoll, D. A. (2015). Citizen Science Program Shows Urban Areas Have Lower Occurrence of Frog Species, but Not Accelerated Declines. PLoS ONE, 10(11), e0140973. https://doi.org/10.1371/journal.pone.0140973
Wright, M. N., Wager, S., & Probst, P. (2019). Package “ranger” A Fast Implementation of Random Forests. https://doi.org/10.1080/10618 600.2014.983641
Xue, Y., Davies, I., Fink, D., Wood, C., & Gomes, C. P. (2016). Behavior identification in two-stage games for incentivizing citizen science exploration. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9892 LNCS, 701–717. https://doi.org/10.1007/978-3-319-44953-1_44
Zipkin, E. F., Royle, J. A., Dawson, D. K., & Bates, S. (2010). Multi-species occurrence models to evaluate the effects of conservation and management actions. Biological Conservation, 143(2), 479-484. https://doi.org/10.1016/j.biocon.2009.11.016
Biosketch
Orin Robinson is an ecologist at the Cornell Lab of Ornithology interested in using and developing quantitative tools to learn about vertebrate population and community ecology, and using lessons learned to inform conservation.

Supporting Information
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Robinson OJ, Ruiz-Gutierrez V, Reynolds MD, Golet GH, Strimas-Mackey M, Fink D. Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. Divers Distrib. 2020;26:976–986. https://doi.org/10.1111/ ddi.13068