

Supplemental structured surveys and pre-existing detection models improve fine-scale density and population estimation with opportunistic community science data

Hallman, Tyler A; Robinson, W Douglas

Scientific Reports

DOI:
[10.1038/s41598-024-61582-6](https://doi.org/10.1038/s41598-024-61582-6)

Published: 14/05/2024

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Hallman, T. A., & Robinson, W. D. (2024). Supplemental structured surveys and pre-existing detection models improve fine-scale density and population estimation with opportunistic community science data. *Scientific Reports*, 14(1), 11070. <https://doi.org/10.1038/s41598-024-61582-6>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **Title: Supplemental structured surveys and pre-existing detection models improve fine-**
2 **scale density and population estimation with opportunistic community science data**

3

4 **Authors:** Tyler A. Hallman^{1,2,3,4*} and W. Douglas Robinson¹

5

6 **Affiliations:**

7 ¹ Oak Creek Lab of Biology, Department of Fisheries, Wildlife, and Conservation Science,
8 Oregon State University, Corvallis, Oregon USA;

9 ² Swiss Ornithological Institute, Seerose 1, 6204 Sempach, Switzerland;

10 ³ Department of Biology and Chemistry, Queens University of Charlotte, Charlotte, North
11 Carolina, USA;

12 ⁴ School of Environmental and Natural Sciences, Bangor University, Bangor LL57 2DG, UK

13

14 ***Corresponding author:** t.hallman@bangor.ac.uk

15 **Abstract**

16 Density and population estimates aid in conservation and stakeholder communication.
17 While free and broadly available community science data can effectively inform species
18 distribution models, they often lack the information necessary to estimate imperfect detection
19 and area sampled, thus limiting their use in fine-scale density modeling. We used structured
20 distance-sampling surveys to model detection probability and calculate survey-specific detection
21 offsets in community science models. We estimated density and population for 16 songbird
22 species under three frameworks: 1) a fixed framework that assumes perfect detection within a
23 specified survey radius, 2) an independent framework that calculates offsets from an independent
24 source, and 3) a calibration framework that calculates offsets from supplemental surveys. Within
25 the calibration framework, we examined the effects of calibration dataset size and data pooling.
26 Estimates of density and population size were consistently biased low in the fixed framework.
27 The independent and calibration frameworks produced reliable estimates for some species, but
28 biased estimates for others, indicating discrepancies in detection probability between structured
29 and community science surveys. The calibration framework produced reliable population
30 estimates with as few as 10 calibration surveys with positive detections. Data pooling
31 dramatically decreased bias. This study provides conservationists and managers with a cost-
32 effective method of estimating density and population.

33 **Introduction**

34 Population estimates are exceptionally valuable for conservation practitioners. They
35 provide tangible and engaging numbers that aid in communicating with stakeholders, including
36 policy makers and the public ¹⁻³. Further, these estimates allow practitioners to set population-
37 based conservation goals, monitor the effects of management actions, and identify conservation
38 successes. Conservation organizations would benefit from the development of cost-effective
39 methods of estimating population size at local and regional scales ^{4,5}.

40 The growth of opportunistic community science projects (also known as citizen science
41 and participatory science) such as eBird, Birdtrack, and Ornitho, provide immense opportunities
42 to develop cost-effective methods of population estimation. Through these community science
43 projects, the spatial and temporal breadth of available biodiversity data has reached
44 unprecedented levels ^{6,7}. To increase participation, many community science projects such as
45 eBird, encourage contributions from observers of all skill levels and allow a large variety of
46 survey methods to be employed. While eBird does not control when or where surveys are
47 conducted, it is classified as a semi-structured community science database. Semi-structured
48 databases are separated from unstructured databases (e.g., iNaturalist) by asking observers to
49 specify the survey protocol used and additional information during data submission (e.g., time,
50 date, etc.). In contrast, more structured databases (e.g., North American Breeding Bird Survey)
51 generally use strict survey methods, predefined survey locations, and trained observers. In terms
52 of the overall information within either project type, the sheer quantity of semi-structured data
53 may compensate for the higher per-datum quality of data from structured community science
54 projects ⁶.

55 Increased participation through the use of less strict protocols, however, is not without
56 drawbacks. Persistent questions of data quality fuel ongoing research on statistical methods that
57 make better use of semi-structured community science data. To date, extensive methods have
58 been developed to improve the performance of community science based species distribution
59 models⁸⁻¹⁰. Conservation planning based on abundance, however, is generally more effective
60 than based on occurrence alone¹¹⁻¹³. Further, for commonly used population-based conservation
61 goals¹⁴, relative or observed abundance information is insufficient as density is required for
62 population estimation. The relative difficulty of modeling density and population size has led to
63 the frequent use of species occurrence as a proxy for density^{15,16}. Although occurrence and
64 density of a species are linked, their relationship is complex and nonlinear, making the direct
65 substitution of one for the other problematic^{11,14}. Estimating density and population from
66 community science data, however, presents a unique set of challenges. While abundance is
67 increasingly available in large community science databases, densities of organisms that allow
68 for population estimation are not.

69 While distribution models built on opportunistic community science data can produce
70 predictions comparable to those informed by professional surveys, abundance information in
71 community science data can be considerably biased and options for estimating density are
72 limited^{17,18}. Addressing these biases, while estimating density and population from observed
73 abundance, requires additional information. Distance sampling data, for example, can address
74 biases through the explicit estimation of individual detection probability. Perhaps more
75 importantly, estimates of area surveyed are essential to converting observed abundance to
76 density. Due to the complexity of implementation, however, both are generally absent from
77 community science databases. The use of structured surveys that include such information, to

78 address the biases in community science data, may allow for unbiased estimates of density and
79 population size.

80 We evaluated the use of highly structured, professional surveys to address the biases in
81 observed abundance in community science data while estimating density and population.
82 Specifically, we used models of detection probability built on structured survey data to estimate
83 survey and location specific detection offsets that were included in community science based
84 density models for 16 songbird species. Our primary objective was to assess whether models of
85 imperfect detection from independently gathered, structured data could be used to adjust
86 community science surveys to produce comparable detectability-adjusted estimates of density
87 and population. We approached this objective with three frameworks that emulate realistic
88 scenarios experienced by researchers and conservationists: 1) a post-hoc implementation of an
89 assumed fixed survey radius that ignores imperfect detection and requires no structured data, 2)
90 an independent, pre-existing source of modeled detection probabilities without access to
91 additional data, and 3) an additional, supplemental, calibration dataset collected specifically to
92 adjust available community science data. This final Calibration framework simulates the
93 collection of structured data, specifically intended to adjust existing community science data,
94 when large, independent, structured datasets are unavailable. Additionally, within the Calibration
95 framework, we investigated the effects of calibration dataset size and data pooling on the degree
96 of bias in estimates of density and population. Throughout, density models from the structured
97 dataset were used as benchmarks to compare community science derived estimates of density
98 and regional population size.

99

100 **Methods**

101 *Study area and species*

102 We compiled environmental and avian survey data from Benton and Polk counties,
103 Oregon, USA. These counties are located along the western edge of the Willamette Valley and
104 the eastern slope of the Oregon Coast Mountains. The Willamette Valley is dominated by a
105 patchwork of agricultural land whose primary crops include festucoid grasses (turf seed
106 production) and tree- and vine-borne fruits such as hazelnuts and grapes. Remnant fragments of
107 native oak woodlands are dispersed throughout lower elevations, with the largest patches within
108 two National Wildlife Refuges. The coastal mountains are dominated by moist Douglas-Fir
109 (*Pseudotsuga menziesii*) forest. An active timber industry diversifies the age structure of the
110 landscape. Elevation ranges from 150 m to 1248 m.

111 We selected 16 species of passerine that regularly breed in the study area. The selected
112 species represent a wide range of sample size (number of positive occurrences in the dataset) and
113 sample prevalence (proportion of surveys within the community science data in which the
114 species occurs; Table 1), factors that can influence species distribution model (SDM)
115 performance¹⁹.

116 *Survey Datasets & Data Processing*

117 We used two sources of wildlife survey data throughout our analyses: a highly-structured,
118 professionally-gathered dataset from the Oregon 2020 project²⁰, and an opportunistically
119 gathered, semi-structured, community science dataset from eBird.

120 ***Structured Dataset.*** From 2011 to 2013, the Oregon 2020 project conducted 2,912
121 structured bird surveys throughout the study area (Fig. 1)²⁰. Trained and experienced observers
122 recorded every bird detected by sight or sound during structured, 5-minute, stationary counts.

123 The counts were conducted every 0.8 km along all accessible roads and every 0.2 km off roads
124 within targeted natural habitats. Surveys were conducted during the breeding season (April 30-
125 July 9) from just before sunrise until song activity declined, sometimes up to 7 hours after
126 sunrise. To address issues of imperfect detection, time-of-detection ²¹ and distance sampling ²²
127 methods were implemented. For time-of-detection, observers tracked and recorded a detection
128 history for each individual bird through five sequential one-minute intervals. For distance
129 sampling, observers estimated the distance to each individual bird at its initial point of detection
130 and confirmed distances with laser rangefinders. We used Oregon 2020's highly structured avian
131 surveys in two ways described in depth below. First, these data informed density models using
132 current best practices to estimate densities and populations that serve as benchmarks, against
133 which results from the community science data could be compared. Second, these data were used
134 to model detection probability and calculate offsets to address imperfect detection in the
135 community science dataset. We refer to the Oregon 2020 data as structured data throughout this
136 paper.

137 For each species, we created benchmark datasets from this structured data. These
138 benchmark datasets were used to inform density models, as described below. Results of
139 community science based models were compared against these benchmark estimates, which were
140 intended to represent current best practices in density modeling. To create benchmark datasets,
141 for each species, we randomly sampled the complete structured dataset without replacement to
142 match the sample sizes of the community science datasets described below. This simultaneously
143 created benchmark datasets and independent test data (e.g. the remaining structured data that
144 were not included in the benchmark dataset) for the calculation of AUC. This process also

145 reduced effects of uneven sample sizes on the comparative performance between community
146 science and benchmark datasets as, generally, models with more data perform better.

147 ***Opportunistic semi-structured dataset.*** We downloaded complete eBird checklists from
148 the study area, date range, and years matching the Oregon 2020 surveys (version ebd_relNov-
149 2017). For each species, we created a separate dataset through stringent filtering. We limited our
150 focus to stationary counts so that environmental data could be directly related to eBird checklist
151 locations. We selected personal locations, as they correspond more closely to the exact locations
152 of stationary counts. We restricted counts to seven hours after nautical dawn and durations to 3
153 to 30 minutes. We removed any remaining eBird checklists that contained presence information
154 (e.g. “X”) instead of counts, which resulted in slightly variable numbers of checklists among
155 species (number of surveys ranged from 1,060 to 1,073 across species). Finally, we used
156 geographic sampling to reduce overrepresentation of birds on territories near popular birding
157 locations. To do this, we created a 200 by 200 m grid over the entire study area and randomly
158 selected one checklist from each grid cell, independent of whether the species was detected.
159 Ideally, there would be sufficient community science surveys within the years in which the
160 structured surveys were conducted, but due to the small numbers of eBird checklists remaining
161 after stringent filtering, we expanded our criteria to include eBird data from 2011-2017 (Fig. 1).
162 While expanding criteria temporally greatly augments the number of community science surveys
163 available, it assumes a constant distribution, density, and population size, within the timeframe.

164 ***Environmental Data***

165 We compiled data from 25 environmental variables previously used to characterize the
166 conditions in Benton and Polk counties for avian SDMs (Table S1)^{19,23}. These variables describe
167 topographic, land cover, and forest structure information acquired from freely available raster

168 datasets^{24,25}. We used focal statistics in ArcMap to calculate percent land cover at five spatial
169 scales shown to be relevant to birds: 75 m, 165 m, 315 m, 615 m, and 1215 m radii from cell
170 centers^{19,23,26,27}. We used focal statistics to calculate the mean values for all topographic and
171 forest structure variables at the same spatial extents.

172 ***Frameworks Implemented***

173 We implemented three frameworks to mimic the circumstances of real-world researchers
174 and conservationists attempting to model local and regional population sizes from community
175 science data (Table 2). The results of these three community science based models were then
176 compared against benchmark estimates.

177 ***Fixed framework.*** This framework represents a scenario in which no independent source
178 of distance sampling surveys or detection functions are available and the decision is made to
179 assume a fixed survey radius for all opportunistic community science surveys. The fixed
180 framework assumes perfect detection (i.e., does not account for imperfect detection) within the
181 defined survey area. It is important to note that this survey area is not a part of the field methods
182 employed during surveys, but is defined at the stage of modeling. In contrast to the frameworks
183 described below, where structured data are used to adjust observed abundance for imperfect
184 detection in community science surveys, no structured data are used in, or required for, the Fixed
185 framework. Since this framework uses a constant survey radius across species, resulting
186 “density” estimates are directly related to observed, or unadjusted abundance. For this
187 framework, we converted observed abundance from community science counts to densities using
188 a fixed 200 m survey radius. We chose 200 m because it is a common distance within which
189 most individuals of our set of species and many North American landbird species would be
190 detected easily by sound (Table S10).

191 ***Independent framework.*** This framework represents a scenario in which an independent
192 source of detection functions are available, but the distance sampling surveys used to inform
193 those detection functions are not available. In this framework, the decision is made to use
194 detection functions from another source to calculate offsets that account for imperfect detection
195 and area surveyed without the option for data pooling. This framework could be particularly
196 valuable as researchers and practitioners would not need to conduct structured surveys, but could
197 apply models of detection probability from independent sources to account for imperfect
198 detection in local semi-structured community science data. The use of this framework is now
199 possible, and will likely increase with the growing availability of such models ²⁸. For this
200 framework, we used the complete structured dataset (2,912 surveys) to model detection
201 probability and estimate survey-specific detection offsets. We included detection offsets in
202 density models built on community science surveys.

203 ***Calibration framework.*** This framework represents a scenario in which no independent
204 source of distance sampling surveys or detection functions are available and the decision is made
205 to collect supplemental distance sampling surveys with which to model detection probability. As
206 large, pre-existing, structured datasets are rare, this scenario is commonly encountered by
207 conservationists looking to estimate local density and population size with community science
208 data. In this framework, we used subsets of the benchmark datasets to model detection
209 probability and estimate survey-specific detection offsets that account for area surveyed. Within
210 this framework, we examined the effects of two pertinent factors for this scenario: sample size
211 within the calibration dataset, and pooling of calibration and community science data. We
212 created calibration datasets with a range of sample sizes to investigate the degree of survey effort
213 necessary to effectively address bias in community science data. We implemented the calibration

214 framework with and without the pooling of calibration and community science datasets, used in
215 density models. Data pooling of even small calibration datasets may decrease bias of results by
216 ensuring that some of the data included in density models experienced the exact detection
217 processes present in modeled detection probabilities. In this way we investigated the influence
218 sample size and data pooling on the efficacy of calibration datasets.

219 To create calibration datasets, we randomly sampled benchmark datasets without
220 replacement, until the desired number of surveys with at least one detection of the species
221 reached 10, 30, 100, and 250 occurrences (herein referred to as sample size). Sampling was
222 performed separately within each iteration of the analysis, so calibration datasets were not
223 identical. We used the number of surveys with at least one detection instead of the number of
224 surveys overall, as rarer species might not be detected in a random sample of all surveys, making
225 offset calculation impossible. We used the number of surveys with at least one detection instead
226 of the number of individuals detected to increase the potential environmental variability
227 incorporated in the calculation of offsets (i.e., if one site had 10 individuals and no other points
228 were selected there would be no variation in environmental variables). Due to the low prevalence
229 of some species, only eight species had sufficient detections for inclusion in the largest (N=250
230 detections) calibration dataset (Table 1).

231 ***Zero-inflated Density Models***

232 For each species and each framework we ran zero-inflated boosted regression tree (BRT)
233 density models (Fig.2). Generally, zero-inflated BRTs are a three-step process that includes
234 fitting an SDM (logistic BRT) to estimate probability of occurrence, converting probability of
235 occurrence to suitable and unsuitable habitat with a threshold, and fitting a Poisson BRT to
236 estimate abundance within suitable habitat ¹¹. We modified this method by adding an

237 intermediate step, in which offsets for detection probability that are calculated from structured
238 survey data are included in Poisson BRTs to convert resulting abundance estimates to density
239 (Fig. 2).

240 ***Zero-inflation.*** For each species, we fit SDMs with logistic regression BRTs²⁹. We set
241 tree complexity to 3, bag fraction to 0.75, and optimized the learning rate so that the optimal
242 number of trees fell between 1000 and 5000. We used a 10-fold cross-validation method to
243 construct boosted regression trees and used a multi-scale SDM framework in which we included
244 all environmental variables at all radii²³. To evaluate models we calculated AUC with the
245 independent test dataset. We then used the sample prevalence of a species within its dataset as
246 the threshold to transform continuous habitat suitability (or probability of occurrence) to
247 binomial suitable and unsuitable habitat^{23,30}. We restricted counts used in Poisson BRTs to those
248 occurring in suitable habitat. This first step of the zero-inflated BRT reduces excess zeroes and
249 the influence of counts in unsuitable habitat prior to modeling abundance¹¹.

250 ***Detection offset calculation.*** In the Independent and Calibration frameworks, detection
251 offsets for community science counts were calculated from detection models built on surveys
252 from the structured dataset, using the QPAD method³¹. No offsets were included in the Fixed
253 framework. Before building models of detection probability, we restricted either the full
254 structured dataset (Independent framework), or the calibration dataset (Calibration framework) to
255 habitat predicted to be suitable for each species, which allowed for the estimation of detection
256 probability within suitable habitat. Imperfect detection is comprised of two components:
257 availability and perceptibility³². We ran removal³³ and distance sampling²² model sets for each
258 species to estimate availability and perceptibility, respectively. For removal models, we reduced
259 our time-of-detection data, which included detection histories at each interval within the five-

260 minute count, to removal data by recording the first interval of detection for each individual (i.e.
261 the removal interval). In removal model sets we included combinations of Julian date, time of
262 day (minutes since dawn), and quadratic terms and compared models with AICc.

263 For distance sampling model sets, we first included distance to the nearest river and
264 distance to the nearest highway (sources of noise) as explanatory variables. We included
265 quadratic terms, log-transformed values, and combinations of distance to river and distance to
266 nearest highway in models. We compared these models to the null with AICc and perpetuated
267 the structure of the top AICc model. We included canopy cover, percent high and medium
268 density urban land cover, and percent total urban land cover, as well as combinations of canopy
269 cover and each of the two urban land cover variables in the subsequent model set. We
270 characterized all land cover covariates in distance models as the mean value within a 75-m radius
271 from cell centers. For all distance models we used 50 m distance bins for distances up to 200 m
272 and included a final bin of all observations over 200 m in distance. The unlimited distance
273 inherent in opportunistic community-science checklists (i.e., observers do not use truncation
274 distances) necessitates an unlimited distance framework³¹. As there is no finite truncation
275 distance, the area sampled is effectively infinite, and estimation of density over an infinite area is
276 impossible. We therefore estimated the effective detection radius (EDR), the radius where the
277 estimated number of individuals missed within the EDR (e.g. not detected) equals the number of
278 individuals detected outside of the EDR, to estimate the effective area sampled. We used the top
279 AICc removal and distance sampling models to calculate offsets (i.e. correction factors) at each
280 survey location³¹. Offsets were calculated as the product of the estimated perceptibility,
281 availability, and effective area sampled. By definition, the perceptibility within the effective area
282 sampled is set to 1.

283 **Density model.** In the Independent and Calibration frameworks, detection offsets were
284 included within Poisson BRTs to convert resulting abundances to densities. In the Fixed
285 framework, no offsets were included as area surveyed was assumed to be constant and detection
286 probability was assumed to be 1. We set tree complexity to 3, bag fraction to 0.75, and optimized
287 learning rates so that the optimal number of trees fell between 1000 and 5000. To avoid
288 overfitting, we included only pseudo-scale optimized environmental variables previously found
289 to be influential for each species in Poisson BRTs^{19,23}. To assess the predictive performance of
290 models, we calculated predictive correlation with the independent test dataset as the correlation
291 between the predicted count at a site derived from estimated densities and offsets, and the
292 observed count. Population estimates were derived from estimated densities. Due to stochasticity
293 involved in the BRT algorithm, and the random sampling of calibration datasets, we ran ten
294 iterations of the above process for each dataset (e.g. each species x dataset combination). These
295 ten iterations were used to assess variability in the results. Zero-inflated density models were run
296 with the *dismo*, *gbm*, and *QPAD* packages in R (version 4.0.3)^{31,34-36}.

297 ***Quantifying Comparative Performance***

298 While we highlight species-specific results below, we were most interested in
299 overarching patterns in comparative performance of each framework's density models. We
300 therefore converted estimates of each endpoint (AUC, area of suitable habitat, mean density, and
301 population) to a percent of the species-specific benchmark estimate. For each endpoint,
302 benchmarks were calculated as the species-specific median value of the benchmark's ten
303 iterations. We divided estimates from individual iterations within each framework by these
304 species-specific benchmark values. To reduce the influence of outliers, within our results, we
305 report medians for all summary statistics.

306

307 **Results**

308 Stringent filtering reduced 12,572 community science checklists to between 1,060 and
309 1,073 (depending on the species) once all criteria were applied (91% reduction; Fig. S1). In
310 benchmark datasets, sample prevalence ranged from 0.03 for Bushtit, Marsh Wren, and White-
311 breasted Nuthatch, to 0.49 for American Robin. In community science datasets, prevalence
312 followed a similar pattern and ranged from 0.01 for Marsh Wren to 0.42 for American Robin
313 (Table 1). Due to low prevalence in some species, only eight species had sufficient detections to
314 create the largest calibration dataset (Table 1).

315 SDMs built on community science data generated similar AUCs to benchmarks (Fig. 3).
316 The median AUC from community science SDMs within our zero-inflated BRTs was 0.77 across
317 species, averaging 97% of benchmark values. Median suitable area estimated from community
318 science data was biased marginally low across frameworks (91% of benchmarks). While
319 accuracy of estimates was high for most species, estimates of suitable area for Marsh Wren were
320 a median of ten times higher than benchmarks (Fig. S2). This bias in suitable area for Marsh
321 Wrens was dramatically reduced to 142% of the benchmark by data pooling within the 30-
322 occurrence calibration dataset. Within zero-inflated BRTs, AUC and estimated suitable area are
323 calculated before offsets of detection probability are incorporated. Therefore, across frameworks,
324 in the absence of data pooling, AUC and estimated suitable area remained constant. When data
325 were pooled in the Calibration framework, precision and accuracy of AUC and estimated
326 suitable area increased with calibration dataset size, a pattern especially evident in the eight most
327 common species, which had sufficient detections to create larger calibration datasets (Fig. 3). In
328 species with lower prevalence, pooling of even small calibration datasets had a large influence

329 on AUC. The variance of AUC and estimated suitable area was higher in species with lower
330 prevalence and lower numbers of detections.

331 Density estimated from community science data was relatively unbiased with a median of
332 95% of the benchmark values. Precision of density estimates increased with increasing size of
333 the calibration datasets, particularly in sample sizes of 100 and 250 occurrences (Fig. 4). While
334 unbiased across most frameworks, estimated density was biased extremely low in the Fixed
335 framework, with a median of 17% of the benchmark values. Also, while density estimates from
336 the other frameworks were unbiased for most species, for Marsh Wren densities were biased
337 extremely low, with a median of 11% of the benchmark values (Table S6, Fig. S2). Density
338 estimates of House Finch and Black-throated Gray Warbler were also biased low, with 52% and
339 68% of the benchmark estimate, respectively. Sample sizes implemented in the Calibration
340 frameworks were robust to random variation above $N=30$ but substantial variability was apparent
341 at the lowest sample size of $N=10$ (Table S7).

342 Overall, population sizes estimated from community science data were biased low, with
343 a median of 87% of the benchmark values (Fig. 4). Similar to density, population estimates from
344 the Fixed framework were biased extremely low, with a median of 21% of the benchmark
345 estimates. With the exception of rare species, the precision of population estimates across the
346 Calibration frameworks improved with calibration dataset size, with greater improvements in the
347 presence of data-pooling. With data pooling, increased calibration dataset size generally
348 decreased bias (i.e., estimates were closer to benchmarks). In contrast, without data pooling,
349 greater negative bias was introduced with increased calibration dataset size (Fig. 4). Population
350 estimates for House Finch and Black-throated Gray Warbler were biased low, and increases in

351 calibration dataset size, even in data pooling frameworks, did not always result in improved
352 estimates (Table S8, Fig. S2).

353

354 **Discussion**

355 We found that even small subsets of structured surveys can be used to address detection
356 bias in free and broadly available community science bird survey data, allowing for the reliable
357 estimation of density and population. The ability to reduce detection bias in community science
358 data, which typically lack the necessary information to account for imperfect detection, while
359 simultaneously estimating an effective survey area, greatly amplifies their conservation value.
360 The substantial bias in our Fixed framework, which lacks adjustments for imperfect detection,
361 emphasizes the risk of estimating populations while taking observed abundance at face value.
362 While this Fixed framework could be greatly improved by using species-specific values, such as
363 maximum detection distance, resulting estimates would remain biased low if detection
364 probability within these distances is ignored³⁷. While the bias of community science derived
365 density and population estimates were greatly reduced in both the Calibration and Independent
366 frameworks, we advise a degree of caution when using such methods as the accuracy of
367 estimates were species-specific.

368 The application of detection functions from the full structured dataset to calculate
369 detection offsets in community science based density models (e.g., the Independent framework),
370 resulted in reliable estimates of density and population for most species. As detection functions
371 with which to calculate these offsets are now available for over 300 landbird species across
372 North America, the use of this Independent framework will likely grow²⁸. The species-specific

373 bias of density and population estimates in our study, however, indicate that care must be taken
374 in the use of structured surveys to adjust community science data. Increasing the similarity
375 between structured and community science datasets through stringent filtering, increases the
376 performance of SDMs ¹⁰, and is likely an essential first step in reducing bias in density and
377 population estimates.

378 We increased alignment of important survey characteristics through stringent filtering,
379 based on count duration, time of day, and locational precision. As our models involve predicting
380 distributions based on habitat characteristics around count locations, community science data
381 must be limited to those surveys using stationary protocols with reliable location information. In
382 eBird, many checklists contributed by birders are traveling counts or stationary counts associated
383 with Hotspot locations. Use of either for our models adds noise and muddles the relationship
384 between observed counts and habitat information. Restricting eBird data to stationary counts at
385 “personal locations” is critical to fine-scale modeling as it reduces much of the locational noise
386 inherent in checklists using other types of protocol (e.g., traveling or incidental) and location
387 (e.g., Hotspot). The use of complete checklists is likewise essential as this allows us to infer
388 absences in checklists without abundances for the species recorded. Geographic sampling or
389 spatial subsampling reduces geographic bias by removing large numbers of counts from
390 popularly surveyed areas. While this may be an important step, it can greatly reduce sample
391 sizes. Here, for example, even with a relatively fine grain of 200 m, geographic sampling
392 reduced the number of opportunistic checklists in our study by around 50 percent (Fig. S1).
393 Geographic bias can have relatively minor impacts on distribution modeling, indicating that
394 geographic sampling may not be strictly necessary ⁸. As we did not rerun our models without
395 geographic sampling, we cannot speak to its impacts on our results. Skipping this step, however,

396 might greatly increase sample sizes for rarer species (discussed more below). Although excising
397 the remaining data from analyses greatly reduces sample sizes (Fig. S1), community science
398 datasets are often large enough that sufficient data remain to justify such filtering.

399 Even with stringent filtering, density and population estimates were biased high in our
400 Independent framework for some species, and low for others. Using models of detection
401 probability, built on structured data, to adjust community science counts, assumes that the
402 detection processes in structured and community science surveys do not differ. For the most part,
403 this seems a reasonable assumption, as important factors such as habitat, extraneous noise, and
404 time-of-day likely impact observers similarly and can be accounted for in stringent filtering and
405 survey-specific detection offsets^{31,38,39}. Differences in observer-specific detection probability,
406 however, are not included in these models. For some species, such as American Robin, Lazuli
407 Bunting, Common Yellowthroat, and White-crowned Sparrow, this assumption appeared to be
408 met, as estimates from the Independent and Calibration frameworks matched benchmarks well.
409 For abundant and conspicuous species such as these, models of detection probability from a
410 previously existing source or a supplementary calibration dataset can be used to effectively
411 estimate spatially explicit densities and populations.

412 For other species, however, including Pacific Wren, Orange-crowned Warbler, Song
413 Sparrow, Spotted Towhee, Swainson's Thrush, and Wrentit, population estimates from
414 Independent and Calibration frameworks were biased low. This bias can likely be attributed to a
415 violation of the assumption that detection probabilities between structured and community
416 science surveys do not differ. In these cases, higher detection probability in structured surveys
417 would lead to lower population estimates. Heterogeneity in the discrepancies of observed
418 abundance between professional and community science counts tend to be species-and-observer-

419 specific¹⁸. Whereas observed abundances in community science counts may be accurate for
420 some species, they tend to be biased low for others. In the case of the songbirds listed here, the
421 proportion of detections that are purely auditory could be quite high. These species tend to sing
422 from cover and visual detection can be difficult. On average, counts from community scientists
423 may be more accurate with species detected visually, than aurally. Masking of auditory cues, and
424 the additional effort required to differentiate multiple vocalizing individuals of the same species,
425 may depress count values from community scientists. While there may be little difference in the
426 number of detections between novice and experienced observers for conspicuous and easy-to-
427 identify species, observer expertise is strongly correlated with observed counts in stationary
428 surveys for species that are more difficult to identify⁴⁰. Data pooling of calibration datasets can
429 reduce the bias of estimates and is an important step when discrepancies in detection probability
430 exist. Truly integrated models that allow for the explicit estimation of observer-specific detection
431 probabilities would further address this assumption⁴¹.

432 Understanding the reasons for discrepancies in detection probability between structured
433 and community science datasets would greatly increase our confidence in these methods. While
434 eBird's checklist calibration index, which uses species accumulation curves to account for
435 observer differences in species detection, improves SDM performance^{42,43}, no index currently
436 exists to account for differences in the reliability of species counts. Such an index would differ
437 from general detection probabilities as it would need to address common observer-specific
438 behaviors, such as rounding of observed counts, recording numbers from memory well after
439 surveys have ended, and reductions in effort in the detection of subsequent individuals following
440 the initial detection of a species. The development of such an index may greatly reduce the bias
441 of community science based density and population estimates. The choice of when and where to

442 begin a survey also introduces bias in opportunistic community science data if the detection of
443 birds or specific species motivates observers to begin surveys. Databases such as eBird, for
444 example, likely include few surveys where no individuals are detected and many surveys where
445 charismatic or vagrant species are detected. Data pooling of calibration datasets may help to
446 address biases associated with choice of survey initiation.

447 Large benchmark datasets may not exist for all species in all locations, and conducting
448 large numbers of surveys to create one can be prohibitively expensive. We therefore evaluated
449 the efficacy of collecting smaller supplementary datasets with which to model detection
450 probability and adjust community science data in our Calibration framework. We found that
451 supplementary calibration datasets with as few as ten surveys in which the target species was
452 detected, could produce unbiased community science based estimates of density and population.
453 Combined with the large bias in the one framework where detection probability was ignored
454 (e.g., Fixed Framework), these results strongly suggest that any community science based
455 estimates of density and population should incorporate the explicit estimation of detection
456 probability, even if very few structured surveys can be conducted. If small calibration datasets
457 can be used effectively, financial and temporal limitations may pose much less of a barrier.
458 While bias was low in small calibration datasets, precision of estimates was greatly improved
459 with increasing calibration dataset size, whether or not calibration data were pooled with
460 community science data. As bias in this framework increased with calibration dataset size for
461 some species, researchers should default to data pooling calibration datasets with community
462 science data. The sample size required to produce precise and unbiased estimates is case-
463 specific, but precision of estimates were greatly improved with 30 or more checklists in which
464 target species were detected.

465 *Small Sample Size & Additional Considerations*

466 Small sample sizes in less common species, such as Bushtit, White-breasted Nuthatch,
467 and Marsh Wren, led to some additional challenges in density modeling. False positives, for
468 example, have a very strong influence when sample sizes are low. Marsh Wren is a habitat
469 specialist, only found in marshes, a rare habitat in the study area. Without data pooling, models
470 predicted extremely large areas of suitable habitat and very low densities throughout. The habitat
471 suitability without data pooling was unambiguously incorrect. Inaccurate species distribution
472 models may be due to two primary factors. First, there may be false positives in the community
473 science data. Given the small sample size in this species, any false positives outside of a marsh
474 could have large impacts on an algorithm's ability to differentiate between suitable and
475 unsuitable habitat. Second, there may be true positives in small marshes not accurately identified
476 by satellite imagery. From a modelling perspective, this would present the same issues as false
477 positives. Data pooling greatly increased the accuracy of habitat suitability models for this
478 species, especially when calibration datasets included at least 30 surveys with positive
479 detections.

480 While estimates of suitable habitat were improved with data pooling, densities within
481 areas of suitable habitat were biased low. There may be at least three contributing factors to
482 biased densities. First, observers may not be visiting marshes when this species is most vocal,
483 and reported abundances may be lower. Second, as this species is primarily identified by sound,
484 community scientists may have lower detection probabilities as some observers may not know
485 how to identify the vocalizations. Third, this species occurs at high densities. At high densities,
486 singing individuals may mask one another, leading to inaccurate counts in community science
487 surveys when effort isn't put into accurately deciphering the number vocalizing. One of the

488 strengths of community science data is that its large quantity can overwhelm a lower per datum
489 information of structured data ⁶. The eBird database continues to grow, and practitioners using
490 community science data can increase sample size by increasing the geographic or temporal scope
491 of the surveys incorporated. For example, we used seven years of data for a single population
492 estimate for each species. Using data from fewer years and larger areas may be more suited to
493 those interested in assessing changes in population size through time. Had we increased the
494 geographic breadth of our study, distribution and density estimates for Marsh Wren may have
495 been much improved. Alternatively, had we chosen to forgo geographic sampling in our study
496 area, rare and geographically restricted marsh habitats would have been far more highly
497 represented as these sites tend to be popular with birders and therefore contain far more
498 opportunistic community science surveys.

499 On lands without public access, allowing community scientists access can increase data
500 in desired locations without the costs associated with wildlife monitoring. Actively inviting
501 community scientists to conduct surveys, year-round or on restricted dates, could further
502 augment desired data while simultaneously engaging the public. Active participation in local
503 conservation can improve conservation actions and help address current biodiversity issues ^{44,45}.
504 When conducting supplemental calibration surveys, encompassing the environmental variability
505 of the geographic area of interest and the variability in survey level characteristics (e.g. time of
506 day or day of year), is important to minimize extrapolation. Although geographic overlap is the
507 clear ideal for a structured dataset, it may not be essential if overlap in environmental space is
508 sufficient.

509 *Suitable Area & Threshold Selection*

510 For many species, the estimated area of suitable habitat was biased low compared to
511 benchmarks. While discrepancies in detection probabilities likely play a role, differences in the
512 breadth of geographic and environmental sampling may also be a contributing factor. Machine
513 learning algorithms such as BRTs can include erroneous relationships between environmental
514 variables and occurrence or abundance when there is insufficient variation to inform models
515 across environmental space. Greater geographic clustering in community science data can lead to
516 less variation in environmental variables, which can lead to lower estimated area of suitable
517 habitat (Fig. 1). In the presence of clustering in community science surveys, collection and
518 pooling of additional data from structured surveys collected in habitats or locations unsurveyed
519 by the community science data may address this issue.

520 It is also important to note that we used species prevalence within opportunistic
521 community science checklists as thresholds when converting continuous habitat suitability to
522 binary suitable habitat. We chose prevalence because it is species-specific and easily calculated
523 from community science data. Many options for thresholding exist, and we did not specifically
524 examine the sensitivity of our modeling to threshold selection. Area of suitable habitat based on
525 binary suitability is sensitive to threshold selection and thresholding results from species
526 distribution models can impact conservation prioritization ^{46, 47}. This may therefore be an
527 important area of continued research and threshold selection should be considered when using
528 these methods. Alternative methods of thresholding can easily be incorporated into the modeling
529 methods used here. As the modeling methods employed here only use data from within suitable
530 habitat to model density, population estimates in this study are somewhat robust to threshold
531 selection. For a given species, lower thresholds result in larger areas of predicted suitable habitat.
532 Generally, these larger areas incorporate a greater number and proportion of low abundance

533 counts, which reduce the model predicted densities. Higher thresholds result in smaller areas of
534 predicted suitable habitat. Generally, these smaller areas include a greater proportion of high
535 abundance counts, resulting in higher predicted densities.

536 *Conclusion*

537 Although opportunistic community science data can be used to produce high performing
538 species distribution models^{10,17}, moving beyond predicted distributions to densities greatly
539 benefits conservation and management⁴⁸. Density and derived population estimates allow
540 conservationists to assess the system's current state, set conservation goals, and evaluate the
541 success of management actions. Biodiversity monitoring is expensive, currently making up a
542 significant portion of total conservation costs⁴⁹. Many local conservation organizations report
543 fiscal barriers to the monitoring necessary to assess the success of their conservation actions⁴.
544 Furthermore, Combined with freely available, remotely sensed environmental data, community
545 science data provide a cost-effective method of monitoring wildlife populations⁵⁰. These
546 methods will be used moving forward²⁸, so understanding their strengths and limitations is
547 essential.

548 In this study, we used independent structured survey data to model detection functions
549 and calculate offsets for community science surveys. Reliable estimation of density and
550 population size with community science data would greatly increase their conservation value.
551 We found that although independent detection functions could be used to produce accurate
552 estimates for some species, there was relatively high bias in others. The collection of
553 supplemental calibration survey data with which to model detection probability was similarly
554 accurate for some species and biased for others. Data pooling of calibration datasets greatly

555 decreased bias, and should be implemented in conjunction with stringent filtering and geographic
556 sampling, where sample sizes are sufficient.

557 **References**

- 558 1. Bickford, D., Posa, M. R. C., Qie, L., Campos-Arceiz, A. & Kudavidanage, E. P. Science
559 communication for biodiversity conservation. *Biological Conservation* **151**, 74–76 (2012).
- 560 2. Briggs, S. V. Integrating policy and science in natural resources: Why so difficult?
561 *Ecological Management & Restoration* **7**, 37–39 (2006).
- 562 3. Thogmartin, W. E. *et al.* A Review of the Population Estimation Approach of the North
563 American Landbird Conservation Plan. *The Auk* **123**, 892–904 (2006).
- 564 4. Anderson, M., Rodewald, A. D. & Dayer, A. A. Regional Variation in US Land Trust
565 Capacities and Activities Related to Bird Conservation. *naar* **41**, 39–46 (2021).
- 566 5. Merenlender, A. M., Huntsinger, L., Guthey, G. & Fairfax, S. K. Land Trusts and
567 Conservation Easements: Who Is Conserving What for Whom? *Conservation Biology* **18**,
568 65–76 (2004).
- 569 6. Munson, M. A. *et al.* A method for measuring the relative information content of data from
570 different monitoring protocols. *Methods Ecol. Evol.* **1**, 263–273 (2010).
- 571 7. Sorte, F. A. L. *et al.* Opportunities and challenges for big data ornithology. *cond* **120**, 414–
572 426 (2018).
- 573 8. Johnston, A. *et al.* Analytical guidelines to increase the value of community science data: An
574 example using eBird data to estimate species distributions. *Diversity and Distributions* **27**,
575 1265-1277 (2021).
- 576 9. Kelling, S. Improving Data Quality in eBird-the Expert Reviewer Network. *Biodiversity*
577 *Information Science and Standards* (2018).

- 578 10. Steen, V. A., Elphick, C. S. & Tingley, M. W. An evaluation of stringent filtering to improve
579 species distribution models from citizen science data. *Diversity and Distributions* **25**, 1857–
580 1869 (2019).
- 581 11. Johnston, A. *et al.* Abundance models improve spatial and temporal prioritization of
582 conservation resources. *Ecological Applications* **25**, 1749–1756 (2015).
- 583 12. Johnston, A. *et al.* Comparing abundance distributions and range maps in spatial
584 conservation planning for migratory species. *Ecological Applications* **30**, e02058 (2020).
- 585 13. Veloz, S. *et al.* Improving effectiveness of systematic conservation planning with density
586 data. *Conservation Biology* **29**, 1217–1227 (2015).
- 587 14. Hallman, T. A., Robinson, W. D., Curtis, J. R. & Alverson, E. R. Building a better baseline
588 to estimate 160 years of avian population change and create historically informed
589 conservation targets. *Conservation Biology* **35**, 1256–1267 (2021).
- 590 15. MacKenzie, D. I. & Nichols, J. D. Occupancy as a surrogate for abundance estimation.
591 *Animal biodiversity and conservation* **27**, 461–467 (2004).
- 592 16. Pearce, J. & Ferrier, S. The practical value of modelling relative abundance of species for
593 regional conservation planning: a case study. *Biological Conservation* **98**, 33–43 (2001).
- 594 17. Henckel, L., Bradter, U., Jönsson, M., Isaac, N. J. B. & Snäll, T. Assessing the usefulness of
595 citizen science data for habitat suitability modelling: Opportunistic reporting versus sampling
596 based on a systematic protocol. *Diversity and Distributions* **26**, 1276–1290 (2020).
- 597 18. Robinson, W. D., Hallman, T. A. & Hutchinson, R. A. Benchmark Bird Surveys Help
598 Quantify Counting Accuracy in a Citizen-Science Database. *Front. Ecol. Evol.* **9**, (2021).
- 599 19. Hallman, T. A. & Robinson, W. D. Deciphering ecology from statistical artefacts:
600 Competing influence of sample size, prevalence and habitat specialization on species

- 601 distribution models and how small evaluation datasets can inflate metrics of performance.
602 *Diversity and Distributions* **26**, 315–328 (2020).
- 603 20. Robinson, W. D., Hallman, T. A. & Curtis, J. R. Benchmarking the avian diversity of
604 Oregon in an era of rapid change. *nwnt* **101**, 180–193 (2020).
- 605 21. Alldredge, M. W., Pollock, K. H., Simons, T. R., Collazo, J. A. & Shriener, S. A. Time-of-
606 detection method for estimating abundance from point-count surveys. *The Auk* **124**, 653
607 (2007).
- 608 22. Buckland, S. T. *et al.* *Introduction to distance sampling estimating abundance of biological*
609 *populations*. (Oxford University Press, 2001).
- 610 23. Hallman, T. A. & Robinson, W. D. Comparing multi- and single-scale species distribution
611 and abundance models built with the boosted regression tree algorithm. *Landscape Ecol* **35**,
612 1161–1174 (2020).
- 613 24. United States Geological Survey. National Gap Analysis Project.
614 <https://gapanalysis.usgs.gov/gaplandcover/data/download/> (2011).
- 615 25. Landscape Ecology, Modeling, Mapping, and Analysis. GNN Structure Maps.
616 <https://lemma.forestry.oregonstate.edu/data/structure-maps> (2014).
- 617 26. Boscolo, D. & Metzger, J. P. Is bird incidence in Atlantic forest fragments influenced by
618 landscape patterns at multiple scales? *Landscape Ecology* **24**, 907–918 (2009).
- 619 27. Shirley, S. M. *et al.* Species distribution modelling for the people: unclassified landsat TM
620 imagery predicts bird occurrence at fine resolutions. *Diversity and Distributions* **19**, 855–866
621 (2013).
- 622 28. Edwards, B. P. M. *et al.* Point count offsets for estimating population sizes of north
623 American landbirds. *Ibis* **165**, 482–503 (2023).

- 624 29. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *Journal*
625 *of Animal Ecology* **77**, 802–813 (2008).
- 626 30. Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. Selecting thresholds of occurrence in
627 the prediction of species distributions. *Ecography* **28**, 385–393 (2005).
- 628 31. Sóllymos, P. *et al.* Calibrating indices of avian density from non-standardized survey data:
629 making the most of a messy situation. *Methods in Ecology and Evolution* **4**, 1047–1058
630 (2013).
- 631 32. Nichols, J. D., Thomas, L. & Conn, P. B. Inferences about landbird abundance from count
632 data: recent advances and future directions. in *Modeling demographic processes in marked*
633 *populations* 201–235 (Springer, 2009).
- 634 33. Farnsworth, G. L. *et al.* A removal model for estimating detection probabilities from point-
635 count surveys. *AUK* **119**, 414–425 (2002).
- 636 34. Greenwell, B., Boehmke, B., Cunningham, J. & Developers, G. `_gbm: Generalized Boosted`
637 `Regression Models_`. R package version 2.1.8. (2020).
- 638 35. Hijmans, R., Phillips, S., Leathwick, J. & Elith, J. `_dismo: Species Distribution Modeling_`.
639 R package version 1.3-5. (2021).
- 640 36. R Core Team. R: A language and environment for statistical computing. (2022).
- 641 37. Matsuoka, S. M. *et al.* Using binomial distance-sampling models to estimate the effective
642 detection radius of point-count surveys across boreal Canada - Utilisation de modèles
643 binomiaux d'échantillonnages basés sur la distance pour estimer le rayon de détection
644 effectif lors d'inventaires par points d'écoute en région boréale canadienne. *The Auk* **129**,
645 268–282 (2012).

- 646 38. Simons, T. R., Alldredge, M. W., Pollock, K. H. & Wettroth, J. M. Experimental Analysis of
647 The Auditory Detection Process on Avian Point Counts. *The Auk* **124**, 986–999 (2007).
- 648 39. Alldredge, M. W., Simons, T. R. & Pollock, K. H. Factors Affecting Aural Detections of
649 Songbirds. *Ecological Applications* **17**, 948–955 (2007).
- 650 40. Bergen, N., Ruyck, C. C. D. & Koper, N. Effects of observer skill and survey method on
651 forest bird abundance data: recommendations for citizen science conservation monitoring in
652 the Caribbean. *JCO* **36**, 45–61 (2023).
- 653 41. Kery, M. *et al.* Integrated distance sampling models for simple point counts. Preprint at
654 <https://doi.org/10.48550/arXiv.2211.17229> (2022).
- 655 42. Johnston, A., Fink, D., Hochachka, W. M. & Kelling, S. Estimates of observer expertise
656 improve species distributions from citizen science data. *Methods in Ecology and Evolution* **9**,
657 88–97 (2018).
- 658 43. Kelling, S. *et al.* Can Observation Skills of Citizen Scientists Be Estimated Using Species
659 Accumulation Curves? *PLOS ONE* **10**, e0139600 (2015).
- 660 44. McKinley, D. C. *et al.* Citizen science can improve conservation science, natural resource
661 management, and environmental protection. *Biological Conservation* **208**, 15–28 (2017).
- 662 45. Novacek, M. J. Engaging the public in biodiversity issues. *PNAS* **105**, 11571–11578 (2008).
- 663 46. Muscatello, A., Elith, J. & Kujala, H. How decisions about fitting species distribution models
664 affect conservation outcomes. *Conservation Biology* **35**, 1309–1320 (2021).
- 665 47. Scherrer, D., D’Amen, M., Fernandes, R. F., Mateo, R. G. & Guisan, A. How to best
666 threshold and validate stacked species assemblages? Community optimisation might hold the
667 answer. *Methods in Ecology and Evolution* **9**, 2155–2166 (2018).
- 668

- 669 48. Nichols, J. D. The Role of Abundance Estimates in Conservation Decision-Making. in
670 *Applied Ecology and Human Dimensions in Biological Conservation* (eds. Verdade, L. M.,
671 Lyra-Jorge, M. C. & Piña, C. I.) 117–131 (Springer, 2014). doi:10.1007/978-3-642-54751-
672 5_8.
- 673 49. Buxton, R. T. et al. Half of resources in threatened species conservation plans are allocated
674 to research and monitoring. *Nat Commun* **11**, 4668 (2020).
- 675 50. Chandler, M. et al. Contribution of citizen science towards international biodiversity
676 monitoring. *Biological Conservation* **213**, 280–294 (2017).
- 677 51. Chesser, R. T. et al. Fifty-eighth supplement to the American Ornithological Society’s
678 Check-list of North American Birds. *tauK* **134**, 751–773 (2017).
- 679

680 **Acknowledgements**

681 We appreciate the contribution of community scientists to the eBird database, assistance with
682 surveys from R. Moore, database work and maintenance from R. DeMoss, and data entry from
683 the Robinson Lab. Thank you to A. T. Peterson, J. Jones, and S. Gregory for providing guidance
684 and insight. The work was supported by the generous endowment of the Bob and Phyllis Mace
685 Watchable Wildlife Professorship.

686 **Author Contributions**

687 T.A.H. designed the study, analyzed the data, and prepared the tables and figures. W.D.R.
688 conducted the professional surveys used within and supervised the research. T.A.H and W.D.R.
689 both contributed to ideas and writing.

690 **Competing Interests**

691 The authors declare no competing interests.

692 **Data availability statement**

693 All data used in this study are available at:

694 [https://figshare.com/articles/dataset/R_Scripts_and_data_used_for_peer-](https://figshare.com/articles/dataset/R_Scripts_and_data_used_for_peer-reviewed_paper_/24723756)
695 [reviewed_paper_/24723756](https://figshare.com/articles/dataset/R_Scripts_and_data_used_for_peer-reviewed_paper_/24723756)

696 **Additional Information**

697 **Supplementary Information**

698 **Figure Legends**

699 **Figure 1.** Stringently filtered survey locations in the community science (green) and structured
700 (orange) datasets for American Robin. After stringent filtering and geographic sampling, 1,060
701 community science surveys remained for this species. The structured survey dataset was sampled
702 without replacement to match survey number.

703

704 **Figure 2.** Workflow for analyses including a) the frameworks and datasets, b) the zero-inflated
705 density modeling method, and c) the calculation of detection probability offsets used within
706 density models. The fixed framework incorporates no offsets and assumes a constant area
707 surveyed of 200m and perfect detection.

708

709 **Figure 3.** Results from the zero-inflated portion of two-step density models for each framework,
710 including AUC (A, B, and C) and estimated area of suitable habitat (D, E, and F), compared
711 against a best-practices reference (benchmark). To allow for summarization across species, for
712 each species, the results of each of the ten iterations within a framework were adjusted to the
713 percentage of the median species-specific reference value. Results are divided into species that
714 are common (A and D; 8 species), uncommon (B and E; 4 species), and rare (C and F; 4 species)
715 within our study area as rarer species had insufficient data for the use of larger calibration
716 datasets (Table 1).

717

718 **Figure 4.** Results from the density portion of two-step density models for each framework,
719 including mean density (A, B, and C) and estimated population (D, E, and F), compared against
720 a best-practices reference (benchmark). To allow for summarization across species, for each
721 species, the results of each of the ten iterations within a framework were adjusted to the
722 percentage of the median species-specific reference value. Results are divided into species that
723 are common (A and D; 8 species), uncommon (B and E; 4 species), and rare (C and F; 4 species)
724 within our study area as rarer species had insufficient data for the use of larger calibration
725 datasets (Table 1).

726 **Tables and Figures**

727 **Table 1.** Descriptive statistics for the 16 study species, including species 4-letter codes, in the structured professional and
 728 opportunistic community science (eBird) datasets. For each species, total number of surveys ranged between 1,060 and 1,073 and was
 729 equal between the two datasets. Species names and sequences follow American Ornithological Society ⁴⁹. Prev. = Prevalence, Obs.
 730 Occ. = Number of sites observed occupied, and Ind. Det. = Number of individuals detected. Local rarity within the study area was
 731 assigned based on number of occurrences and the largest calibration dataset used within the study: C = Common, U = Uncommon, and
 732 R = Rare.

Species	Scientific Name	Structured Professional Dataset			Opportunistic Community Science Dataset			Largest Calibration Dataset	Local Rarity
		Prev.	Obs. Occ.	Ind. Det.	Prev.	Obs. Occ.	Ind. Det.		
Bushtit	<i>Psaltriparus minimus</i>	0.03	35	59	0.02	19	61	30	R
Wrentit	<i>Chamaea fasciata</i>	0.04	46	56	0.03	35	43	30	R
White-breasted Nuthatch	<i>Sitta carolinensis</i>	0.03	29	32	0.03	27	31	30	R
House Wren	<i>Troglodytes aedon</i>	0.11	119	149	0.16	169	263	100	U
Pacific Wren	<i>Troglodytes pacificus</i>	0.22	238	373	0.15	157	212	250	C
Marsh Wren	<i>Cistothorus palustris</i>	0.03	34	64	0.01	15	34	30	R
Swainson's Thrush	<i>Catharus ustulatus</i>	0.46	497	736	0.39	416	755	250	C
American Robin	<i>Turdus migratorius</i>	0.49	524	754	0.42	440	819	250	C
House Finch	<i>Haemorhous mexicanus</i>	0.12	128	164	0.05	54	100	100	U
White-crowned Sparrow	<i>Zonotrichia leucophrys</i>	0.23	244	389	0.22	232	438	250	C
Song Sparrow	<i>Melospiza melodia</i>	0.47	506	750	0.38	406	582	250	C
Spotted Towhee	<i>Pipilo maculatus</i>	0.33	357	478	0.37	399	581	250	C

Orange-crowned Warbler	<i>Leiothlypis celata</i>	0.26	281	360	0.28	301	408	250	C
Common Yellowthroat	<i>Geothlypis trichas</i>	0.25	269	412	0.25	269	460	250	C
Black-throated Gray Warbler	<i>Setophaga nigrescens</i>	0.13	144	182	0.12	124	157	100	U
Lazuli Bunting	<i>Passerina amoena</i>	0.09	95	116	0.09	98	133	100	U

733

734 **Table 2.** Brief descriptions of the three frameworks implemented in this study. Each framework adjusts community science bird
 735 survey data to allow for density estimation. Only the Fixed framework adjusts surveys without the explicit estimation of detection
 736 probability and survey area. The Calibration framework was run with and without data pooling in density models to investigate the
 737 influence of data pooling and sample sizes on density estimates. See methods for more in-depth descriptions of frameworks.

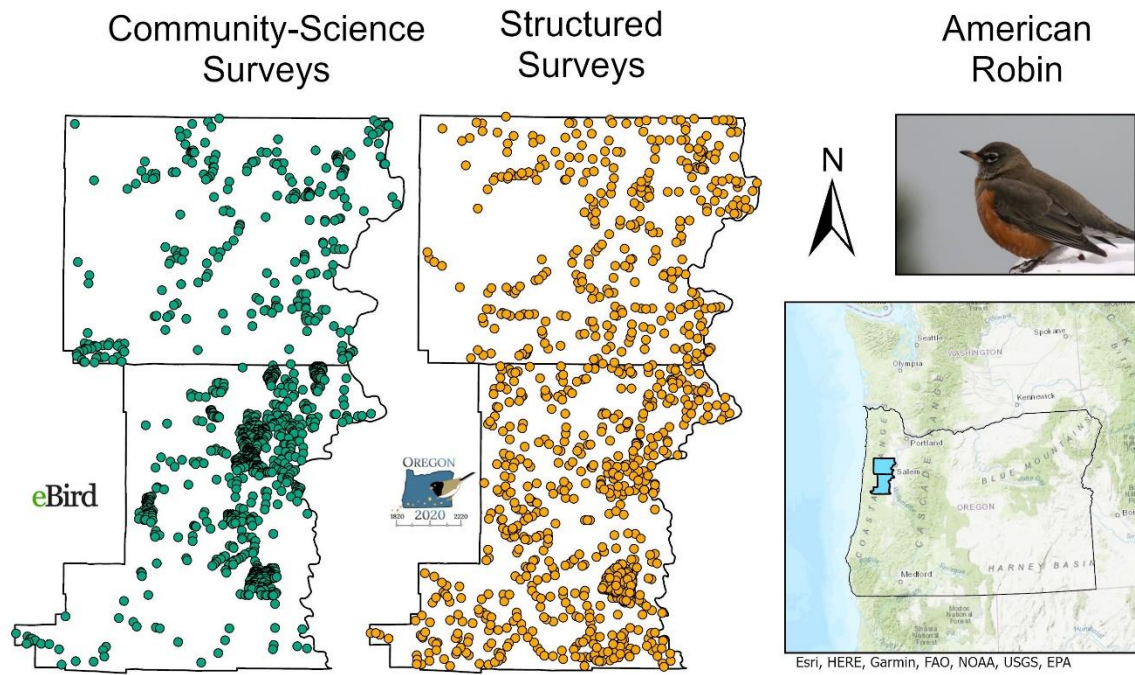
Framework	Converts Abundance to Density	Estimates Variable Survey Area	Adjusts for Imperfect Detection	Includes Data Pooling	Calibration Sample Size
Fixed	<i>Yes</i>	<i>No</i>	<i>No</i>	<i>No</i>	NA
Independent	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	NA
Calibration	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	10
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	30
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	100
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	250
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	10
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	30
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	100
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	250

738

Framework	Converts Abundance to Density	Estimates Variable Survey Area	Adjusts for Imperfect Detection	Includes Data Pooling	Calibration Sample Size
Fixed	<i>Yes</i>	<i>No</i>	<i>No</i>	<i>No</i>	NA
Independent	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	NA
Calibration	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	10
	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	30

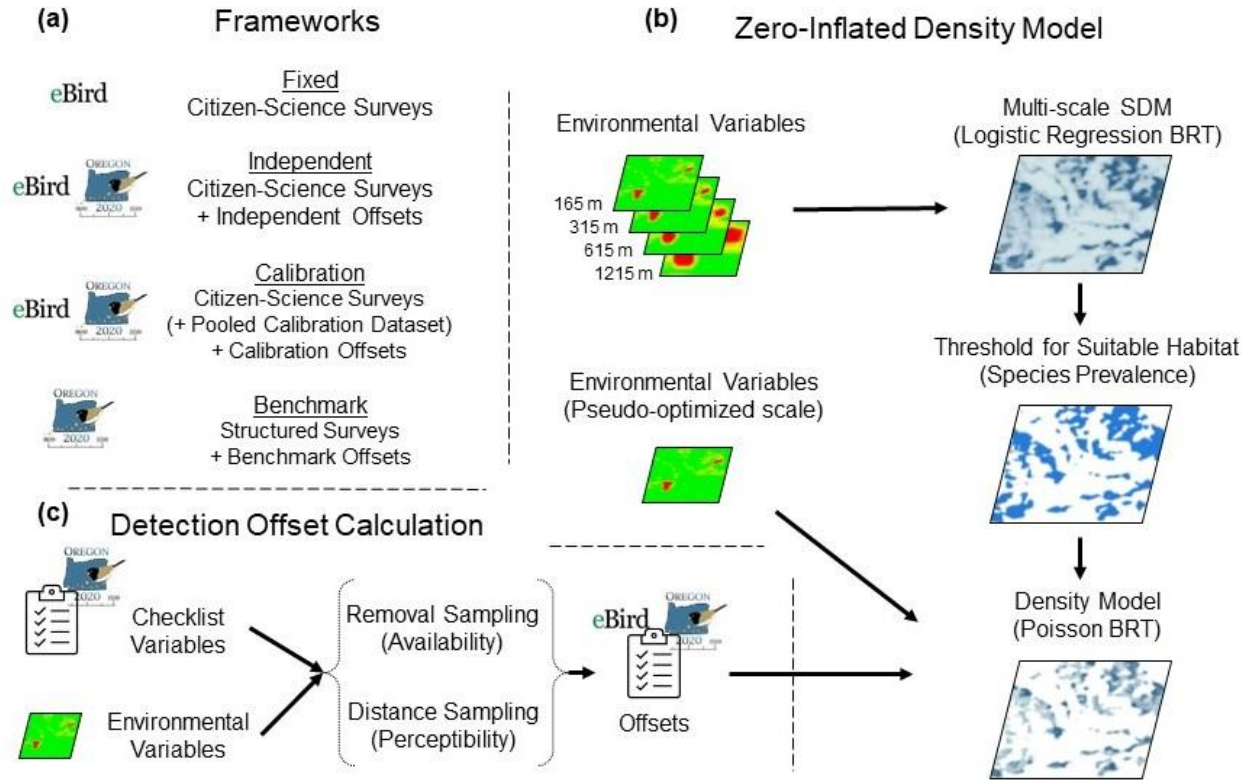
<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	100
<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>No</i>	250
<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	10
<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	30
<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	100
<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	250

739



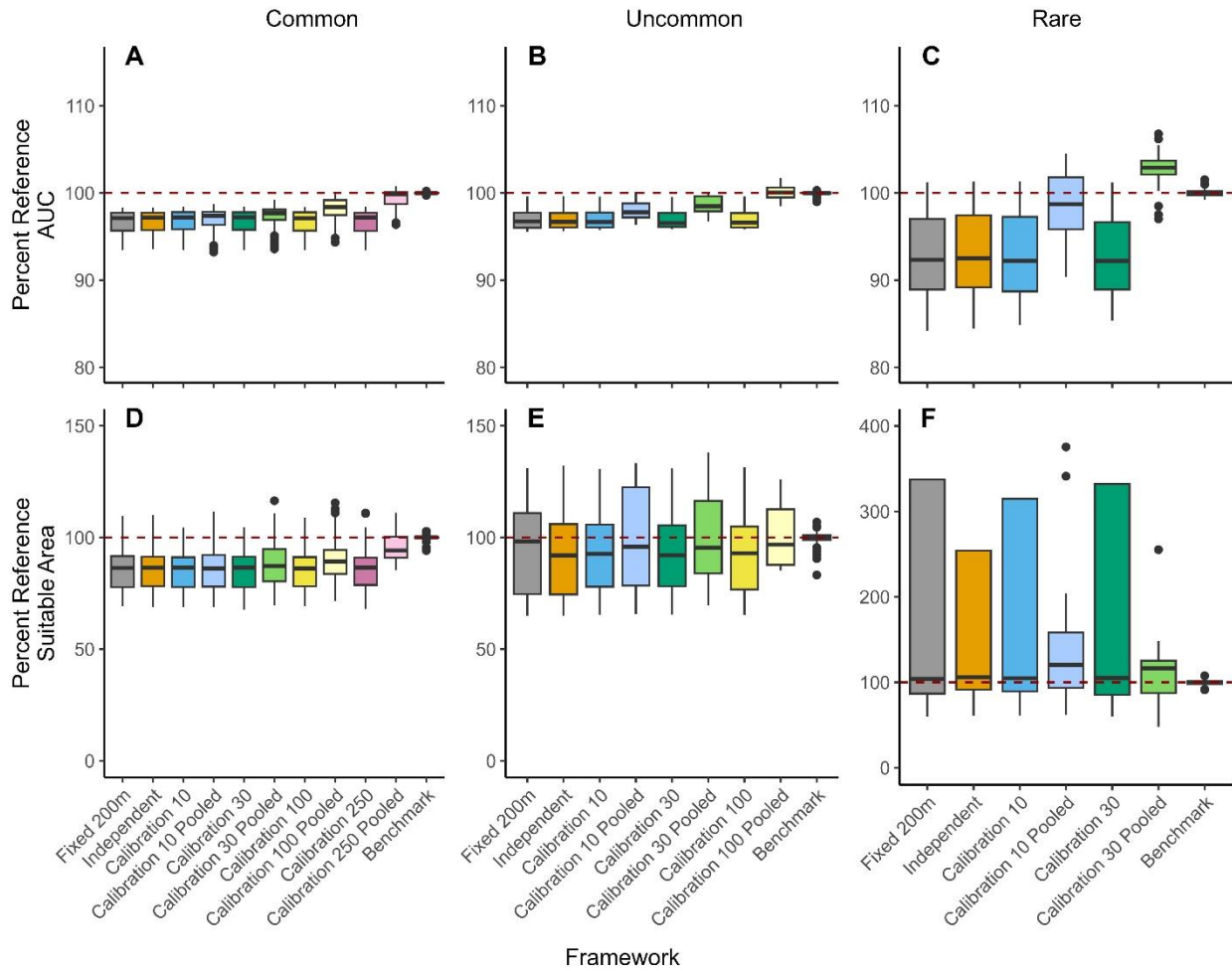
740

741 **Figure 1.** Stringently filtered survey locations in the community science (green) and structured
 742 (orange) datasets for American Robin. After stringent filtering and geographic sampling, 1,060
 743 community science surveys remained for this species. The structured survey dataset was sampled
 744 without replacement to match survey number.



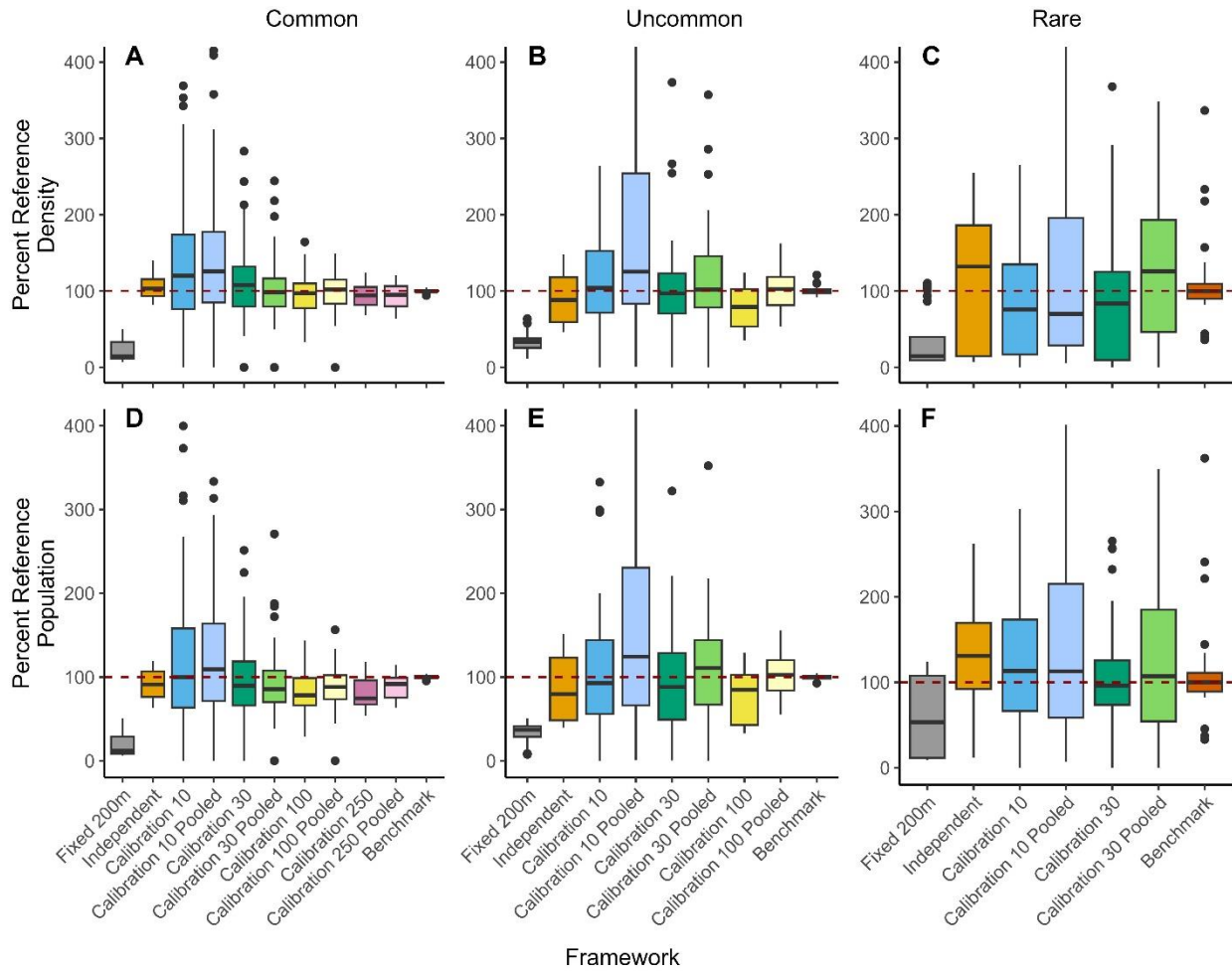
745

746 **Figure 2.** Workflow for analyses including a) the frameworks and datasets, b) the zero-inflated
 747 density modeling method, and c) the calculation of detection probability offsets used within
 748 density models. The fixed framework incorporates no offsets and assumes a constant area
 749 surveyed of 200m and perfect detection.



750

751 **Figure 3.** Results from the zero-inflated portion of two-step density models for each framework,
 752 including AUC (A, B, and C) and estimated area of suitable habitat (D, E, and F), compared
 753 against a best-practices reference (benchmark). To allow for summarization across species, for
 754 each species, the results of each of the ten iterations within a framework were adjusted to the
 755 percentage of the median species-specific reference value. Results are divided into species that
 756 are common (A and D; 8 species), uncommon (B and E; 4 species), and rare (C and F; 4 species)
 757 within our study area as rarer species had insufficient data for the use of larger calibration
 758 datasets (Table 1).



759

760 **Figure 4.** Results from the density portion of two-step density models for each framework,
 761 including mean density (A, B, and C) and estimated population (D, E, and F), compared against
 762 a best-practices reference (benchmark). To allow for summarization across species, for each
 763 species, the results of each of the ten iterations within a framework were adjusted to the
 764 percentage of the median species-specific reference value. Results are divided into species that
 765 are common (A and D; 8 species), uncommon (B and E; 4 species), and rare (C and F; 4 species)
 766 within our study area as rarer species had insufficient data for the use of larger calibration
 767 datasets (Table 1).