Running title: Comprehensive Distribution Modeling

Modeling Avian Full Annual Cycle Distribution and Population Trends with Citizen Science Data

Daniel Fink, Tom Auer, Viviana Ruiz-Gutierrez, Wesley M. Hochachka, Alison Johnston, Frank A. La Sorte, and Steve Kelling*

Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA.

**Corresponding author: Email:* daniel.fink@cornell.edu

Authors' contributions. *DF, WMH, and STK conceived and designed this study. DF designed the statistical methodology. TA and DF designed the computational methodology, processed data, and distribution models. TA, VRG, WMH, AJ, and FAL designed the analysis of the model products. DF wrote the first draft of the manuscript, and all authors contributed substantially to revisions. All the authors have approved the final version of this manuscript and agree to be accountable for all aspects of the work.*

Data Accessibility: *Should the manuscript be accepted, the data supporting the results will be archived in an appropriate public repository such as Dryad or Figshare and the data DOI will be included at the end of the article.*

Abstract

Information on species' distributions and abundances, environmental associations, and how these change over time are central to the study and conservation of wildlife populations. This information is challenging to obtain at relevant scales across range-wide extents for two main reasons. First, local and regional processes that affect populations vary throughout the year and across species' ranges, requiring fine-scale, year-round information across broad — sometimes hemispheric — spatial extents. Second, while citizen science projects can collect data at these scales, using these data requires additional steps to address known sources of bias. Here we present an analytical framework to address these challenges and generate year-round, range-wide distributional information using citizen science data. To illustrate this approach, we apply the framework to Wood Thrush (*Hylocichla mustelina*), a long distance Neotropical migrant and species of conservation concern, using data from the citizen science project eBird. We estimate relative occupancy and abundance with enough spatiotemporal resolution to support inference across a range of spatial scales throughout the annual cycle. This includes intra-annual estimates of the range (quantified as the area of occupancy), intra-annual estimates of the associations between species and features of their local environment, and inter-annual season-specific trends in relative abundance. This is the first example of an analysis to capture intra- and inter-annual distributional dynamics across the entire range of a broadly distributed, highly mobile species.

Keywords: *conservation, biogeography, macroecology, eBird, bird migration, biodiversity monitoring, Wood Thrush*

MAIN TEXT:

(a) Introduction

Information about species' distribution and abundance are essential to the fields of applied ecology and conservation biology, for they are critical in the study of the processes that limit and regulate populations. Because the biotic and abiotic processes affecting species' populations vary seasonally and regionally, it is also important to generate this information both at the spatiotemporal *scales* at which these processes operate and across the full spatiotemporal *extents* over which these processes vary (Heffernan *et al.* 2014). Finally, for broadly distributed species, information that supports inference *across spatial and temporal scales* is necessary to understand how local, regional, and seasonal-scale processes interact to affect entire populations at continental extents.

For most well studies species groups, we still lack basic information on species distributions, especially at relevant spatiotemporal resolutions and extents. A limiting factor has been the ability to collect sufficient quantities of observational data both at fine scales and across broad spatiotemporal extents. Current information on species distributions often suffer from strong regional biases (Boakes *et al.* 2010) in coverage. Although range maps are available for a growing number of taxa, these often provide coarse spatial (Hurlbert & Jetz 2007) and temporal resolutions (Ridgeley *et al.* 2012). Further, most of this information depicts only the most basic information on species' range, often as expert drawn polygons, lacking more informative and ecologically valuable measures of occupancy rates, relative abundance, and habitat associations within

71 a species' range.

72

73 Most information on species distributions is also static and does not capture intra- or
74 inter-annual dynamics. The majority of studies on amphibians, reptiles, birds, and
75 mammals have been conducted only during the breeding season (Marra *et al.* 2015). Even
76 for birds, one of the best-surveyed classes of organisms, continental-scale trends in
77 abundance are routinely estimated only for the breeding season and only in North
78 America (Sauer & Link 2011) and Europe (European Bird Census Council 2016). The
79 inability to estimate changes in distribution and abundance is a serious deficiency in our
80 ecological knowledgebase, and underscores the need for scalable approaches to collect
81 biodiversity data and model intra- and inter-annual dynamics.

82

83 Citizen science projects that use crowdsourcing techniques to engage the public have
84 been very successful at collecting observational data across large areas throughout the
85 year (Dickinson *et al.* 2010). However, using these data to generate robust distributional
86 information is fraught with analytical challenges (Hochachka *et al.* 2012; Bird *et al.*
87 2014). These challenges have led to the development and application of a number of new
88 analytical tools. For example, bias related to heterogeneity in species' detection is a
89 significant challenge when analyzing citizen science data. Several approaches have been
90 used to deal with heterogeneous and imperfect observation processes, from the inclusion
91 of relevant fixed and random effects (Sauer & Link, 2011) to the development of
92 detection process models (Kéry & Royle, 2015). When citizen scientist participants
93 choose where and when to conduct their surveys, site selection biases lead to repeated

surveys in popular search locations and missing surveys in areas that are hard to access. Data-sampling methods have proven useful to mitigate the effects of these site-selection biases (Robinson *et al.* 2017). The ability to accurately capture the complex species-environment relationships underlying regional patterns of species' distributions and abundance is another challenge. Machine and statistical learning models have proven to be efficient tools for learning these species-environment relationships from large sets of environmental covariates (Elith & Leathwich 2009). Using citizen science data to estimate abundance often presents the statistical challenge of "zero inflation" where too many zero counts can degrade model performance. A wide variety of new abundance models have been proposed to deal with zero-inflation (Denes *et. al.* 2015).

Most of the methodological developments discussed above have been used to study regional-scale patterns of species' distribution and abundance during a single season of the year. To generate distributional information across larger spatiotemporal extents with citizen science data, requires tacking two additional challenges. First, most broad-scale citizen science projects exhibit strong variation in data density across large regions which can degrade model performance. Adaptive knot designs (Gelfand *et al.* 2012) and partitioning methods (Fink *et al.* 2013) have been proposed to deal with this challenge by adding multi-scale structure to broad extent analyses. Second, the spatial and temporal variation in a species' response to the same environmental conditions, statistical non-stationarity, can also degrade model performance. Non-stationary regression techniques function to add multi-scale structure to analyses and have proven to be a useful solution to this challenge (Fink *et al.* 2010; Finley 2011).

While previous studies have dealt with one or more of the analytical challenges outlined above (e.g. Johnston *et al.* 2015), to date, none have dealt with all of these challenges simultaneously at the relevant scales necessary to study broadly distributed species across the annual cycle. Here, we describe a simple analytical framework capable of estimating species' relative occupancy and abundance, year-round and range-wide with enough spatial and temporal resolution to support inference across a range scales. This includes intra-annual estimates of species ranges' (quantified as the area of occupancy), intra-annual estimates of the associations between species and features of their local environment, and inter-annual trends in relative abundance. For convenience, we will refer to this suite of parameter estimates as Cross-Scale Full-Annual Cycle (CS-FAC) distributional information.

As a case study, we analyzed data from the global citizen science project eBird (Sullivan *et al.* 2014) for the long-distance migratory songbird, Wood Thrush (*Hylocichla mustelina*) that breeds in eastern North America and winters largely in Mesoamerica. The Wood Thrush is a species of conservation concern, having suffered steep population declines over the past several decades. Despite numerous regional studies (e.g. Rushing *et al.* 2017), comprehensive information on abundance and distribution is still lacking. We used the CS-FAC analysis to fill important information gaps for this species, illustrating a method that is widely applicable to other species and taxonomical groups.

(b) Materials and methods

140

141 **Observational Data**

142 The bird observation data were obtained from the global bird monitoring project, eBird
143 (Sullivan *et al.* 2014) using the eBird Reference Dataset (ERD2016, Fink *et al.* 2017).
144 We used a subset of data in which the time, date, and location of the survey period were
145 reported and observers recorded the number of individuals of all bird species detected
146 and identified during the survey period, resulting in a complete 'checklist' of species on
147 the survey (Sullivan *et al.* 2009). The checklists used here were restricted to those using
148 the 'stationary', 'traveling', or 'areal search' protocols from January 1, 2004 to December
149 31, 2016 within the spatial extent between 180° to 30° W Longitude. Areal surveys were
150 restricted to those covering less than 56 sq. km. and traveling surveys were restricted to
151 those ≤ 15km. This resulted in a dataset consisting of 14 million checklists, of which 10%
152 were withheld for model validation.

153

154 **Predictor Data**

155 We incorporated three classes of predictors in the models: (1) Four observation-effort
156 predictors to account for variation in detection rates, (2) Three temporal predictors to
157 account trends at different scales, and (3) 79 site-specific predictors from remote sensing
158 data to capture associations of birds with elevation and a variety of habitats across the
159 hemisphere. The effort predictors included the duration spent searching for birds, whether
160 the observer was stationary or not, the distance traveled during the search, and the
161 number of people in the search party. The observation time of the day was used to model
162 variation in availability for detection, e.g. variation in behavior such as participation in

the dawn chorus (Diefenbach *et al.* 2007). The day of the year (1-366) on which the search was conducted was used to capture intra-annual variation and the year of the observation was included to account for inter-annual variation.

Elevation defines basic abiotic conditions that influence species distributions. To account for the effects of elevation and topography, each checklist location was associated with elevation, eastness, and northness at $1km^2$ resolution (Amatulli *et al.* 2017). To account for species' local habitat-selectivity each checklist was linked to a series of covariates derived from the NASA MODIS land cover data (Friedl *et al.* 2010). We selected this data product for its moderately high spatial resolution, annual temporal resolution, and global coverage. We used the University of Maryland (UMD) land cover classifications (Hansen *et al.* 2000) and derived water cover classes from the MODIS Land Cover Type QA Science Dataset resulting in a class label for each 500m pixel into one of 19 classes (Table 1).

Checklists were linked to the MODIS data by-year from 2004-2013 (checklist data after 2013 are matched to the 2013 data). All cover classes were summarized within a 2.8km × 2.8km (784 hectare) neighborhood centered on the checklist location. In each neighborhood, we computed the proportion of each class in the neighborhood (PLAND). To describe the spatial configuration of each class within each neighborhood we computed three statistics using FRAGSTATS (McGarigal *et al.* 2012; VanDerWal *et al.* 2014): LPI an index of the largest contiguous patch, PD an index of the patch density, and ED an index of the edge density.

186

187 **Analysis Overview**

188 To meet the analytical challenges of CS-FAC modeling with eBird data, we adopted an
189 ensemble modeling strategy combining pattern discovery and inference based on the
190 Spatio-Temporal Exploratory Model (STEM; Fink *et al.* 2010). STEM is an ensemble of
191 regional-seasonal regression models generated by repeatedly subsampling and
192 partitioning the study extent into 100 randomly located grids of overlapping
193 spatiotemporal blocks, called *stixels*, and then fitting an independent regression model,
194 called a *base model*, within each stixel. Together, the base models are used to form an
195 ensemble of local parameter estimates distributed uniformly across the study extent.
196 Using the fact that stixels overlap in space and time, parameters at a given location and
197 time are estimated by averaging across overlapping base models. Combining estimates
198 across the ensemble controls for inter-model variability (Efron 2014), providing a simple
199 way to control for overfitting while naturally adapting to non-stationary species-
200 environment relationships (Fink *et al.* 2010). Utilizing the fact that data are subsampled
201 for each base model, resampling techniques are employed to generate uncertainty
202 estimates for the ensemble parameter estimates. All analysis was conducted in R, version
203 3.4.2 (R Development Core Team 2017) and deployed using Apache Spark 2.1 (Zaharia,
204 *et. al.* 2016).

205

206 In the following sections, we describe the STEM base models, ensemble design, and the
207 spatial case-control sampling procedure. Then we describe how we used the ensemble to
208 estimate four population parameters: (1) landscape-scale relative occupancy and

abundance, (2) landscape-scale range boundaries quantified as the area of occupancy, (3) regional-scale habitat use and avoidance, and (4) regional-scale trends in occupancy and abundance.

STEM Base Models

Within each stixel, predictor-response (i.e. species-environment) relationships were assumed to be stationary. To estimate relative occupancy and abundance from the large predictor set while accounting for high numbers of zero counts, we used a two-step Zero-Inflated Boosted Regression Tree (ZI-BRT) model (Johnston *et. al.* 2015; Ridgeway *et al.* 2017). Effort and time covariates were included in both steps of the ZI-BRT to account for variation in detectability and variation in availability for detection. To generate estimates of the binary un/occupied state, the occupancy threshold value that maximized the Kappa statistic (Cohen 1960) was recorded for each ZI-BRT base model.

While the ZI-BRT base model can produce good estimates of the spatial patterns of occupancy and abundance from large sets of predictors, it is not well suited to making inference about inter-annual trends. Instead, we used the Zero-Inflated Generalized Additive Model (ZI-GAM) fit with the mcv package (Wood *et al.* 2016) to estimate trends, employing the additive structure to focus inference. The model equations defining the occupancy and abundance sub-models had the forms

$$logit(\psi) = \beta_0 + s(Year) + s(\psi spat) + s(lat, lon) + s(duration) + s(distance) + s(number) + stationary$$

$$log(N) = \alpha_0 + s(Year) + s(Nspat) + s(lat, lon),$$

where ψ is the occupancy and N is the abundance. Inter-annual variation in occupancy and abundance were both modeled using smooth functions, s, of the *Year*. To account for variation in the observation process, smooth functions of search *duration*, search *distance*, and the *number* of observers were included. The *stationary* indicator, was used to identify stationary counts, accounting for differences in search protocols. Spatial variation for both occupancy and abundance was modeled at two scales. Coarse-scale spatial patterns were captured by smooth two dimensional smooth functions of latitude and longitude with a limited number of knots. Fine-scale variation was modeled as smooth functions of ZI-BRT derived covariates describing landscape-scale variation in occupancy, ψ*spat*, and abundance, *Nspat*.

The ZI-BRT derived covariates were constructed to encapsulate landscape-scale spatial variation as a pair of covariates that could be used within the ZI-GAM. To do this we leveraged the strength of the ZI-BRT model as a high-dimensional regression model. The idea was to use the ZI-BRT model to adaptively select and estimate intra-seasonal landscape-scale effects from the large set of predictors describing land and water cover class and elevation. This was done by fitting the ZI-BRT model as described above, then predicting the expected relative occupancy and abundance for each checklist in the training data to generate the two covariates. Because we want to describe inter-annual variation in the ZI-GAM, an important part of generating these predictors was removing any inter-annual variation from the covariates. This was done simply by holding the checklist year predictor constant when predicting the covariate values. Similarly, to remove variation in detection associated with the effort predictors, all effort predictors

255 were also held constant.

256

257 By including the derived covariates in the ZI-GAM as smooth additive functions, the Zi-
258 GAM was able to adaptively calibrate the covariate information. Note, that by allowing
259 the day of the year predictor to vary when computing the derived covariates, the
260 covariates captured strong intra-seasonal changes in the spatial patterns of occupancy and
261 abundance. This is especially useful for modeling data from non-stationary periods.
262 Finally, we note that both ZI-BRT and ZI-GAM were fit independently within each stixel
263 in the ensemble. Thus, averaging across the ensemble naturally controls for the joint
264 variation of the ZI-BRT, the ZI-BRT derived covariates, and the ZI-GAM.

265

266 **Ensemble Design**

267 Stixel size controls a bias-variance tradeoff (Fink *et al.* 2013) and must strike a balance
268 between stixels that are large enough to achieve sufficient sample sizes, controlling
269 variance, and small enough to assume stationarity, controlling bias. For estimating
270 occupancy and abundance with ZI-BRT base model, 10° longitude × 10° latitude × 30-
271 contiguous day stixels were small enough to meet these requirements across much of the
272 northern study extent. To account for the relatively low data density south of 12° north
273 latitude we doubled the length of stixels to 20° longitude × 20° latitude × 30-days. For
274 estimating trends with ZI-GAM base model larger sample sizes were required to insure
275 representative sampling across years. To estimate breeding season trends, the same 10°
276 longitude × 10° latitude × 30-contiguous day stixels were used. To estimate trends during
277 the non-breeding season, when data density is lower, we increased stixel size to cover the

spatiotemporal region from southern Mexico to Panama from December 1 to February 28. See Appendix S1 in Supporting Information for additional information about the specification of the stixel ensemble design.

Spatial Case-Control Sampling

Within each stixel, a spatial case-control sampling strategy was used to address the challenges of highly imbalanced data and site selection bias. Imbalanced data arise when there are a very small number of species detections and a very large number of non-detections. This is a modeling concern because binary regression methods, like the first component of ZI-BRT model, become overwhelmed by the non-detections and perform poorly (Robinson *et al.* 2017). The low detection rates of many species, especially along seasonal range boundaries, can generate highly imbalanced training data and make data imbalance a defining challenge for broad-scale, year-round modeling. By sampling non/detection cases separately, case-control sampling (e.g. Fithian & Hastie 2014) improves data balance and model performance. Additionally, to alleviate spatial biases caused by the eBird site selection process, spatially balanced samples were drawn as part of the case-control sampling.

To generate spatially balanced samples for the case-control sampling, training data were drawn from a randomly located regular grid, with one checklist randomly selected per grid cell. North of 12° latitude, the grid cells were 10km × 10km and south of the cutoff they were 20km × 20km. As part of the case-control sampling, detection data were over-sampled, using the same spatially balanced procedure, when they represented less than

301 25% of the spatially balanced data. Because boosting, used in the ZI-BRT base models, is
302 driven more by the set of distinct data points than the number of tied data points
303 generated when oversampling, the spatial predictors of tied checklists were jittered to
304 break the ties (Mease et al. 2007). Finally, because case control sampling changes the
305 training data prevalence we back-transformed occupancy estimates to match the original
306 stixel prevalence rate.
307
308 For the ZI-GAM models, we also stratified the spatial case control samples by year and
309 restricted the total number of samples per year to control for inter-annual increases in
310 sample sizes resulting from increases in eBird participation rates. eBird participation rates
311 have been increasing at 20-30% per year since 2005. We set 2007 as our *reference year*,
312 fixing the sample size of spatially balanced training data for each year to equal that of
313 2007. We selected 2007 as the reference year as a balance between the amount of per-
314 year information available to estimate trends and the length of the trend.
315
316 **Local Relative Occupancy and Abundance**
317 We estimated relative occupancy and abundance once per week for all 52 weeks of the
318 calendar year. Estimates were made at 614,575 locations across the terrestrial Western
319 Hemisphere from a regular spatial grid with a density of one location per 8.4km × 8.4km
320 grid cell. Estimates at each location and date were made based on predictor values at that
321 location from all base models that contained the location and date. Then we averaged
322 across the model estimates using the upper 5% trimmed mean, a robust estimator
323 designed to guard against bias due to large outlying values. Uncertainty of the estimates

was estimated using the subsampling approach of Politis *et al.* (2009) following the computational strategy of Geyer (2013). See the Appendix S2 for more information about the subsampling procedures.

For all estimates we controlled for variation in detection rates associated with search effort by holding predictors for search duration, protocol, search length, and number of observers constant. Thus, the quantity we used to estimate relative occupancy was defined as the probability that an average eBird participant would detect the species on a search from 7–8AM while traveling 1 km on the given day at the given location. Relative abundance was estimated as the expected count of individuals of the species on the same standardized checklist. Although this approach accounts for variation in detection rates, it can not directly estimate the absolute detection probability. For this reason, we refer to the quantities estimated by the model as *relative* measures of occupancy and abundance.

Local Area of Occupancy

To estimate the Area of Occupancy (AOO) we predicted the binary un/occupied state for all weeks using the same 8.4km × 8.4km spatial grid of locations used for relative occupancy and abundance estimates. At each location and week, the AOO was estimated as the proportion of base model occupancy estimates that were larger than the thresholds for the corresponding base models. We call this the Proportion Above local Threshold (PAT) estimator. We say the site was occupied, and, therefore within the species' range, if the PAT estimator exceeded a specified value. One benefit of this ensemble estimator is that it naturally adapts to regional and seasonal variation in species prevalence and

347 detectability. By averaging across the ensemble, PAT controls for inter-model variation
348 in both occupancy and base model threshold estimates.
349
350 **Regional Habitat Associations**
351 For each base model, we quantified the strength and direction of association for each
352 cover class predictor. Predictor importance (PI) statistics measured the strength of the
353 over-all contribution of individual predictors as the change in predictive performance
354 between the model that includes all predictors and the same model with permuted values
355 of the given predictor (Breiman 2001). PI statistics capture both positive and negative
356 effects arising from both additive and interacting model components. Partial Dependence
357 (PD) statistics measured the functional form of the additive association for each
358 individual cover class predictor by averaging out the effects of all other predictors (Hastie
359 *et al.* 2009). To measure the direction of association, we estimated the slope of each PD
360 estimate using simple linear regression. Because the PI and PD statistics account for the
361 effects of all other predictors in the base models, they account for variation in
362 detectability associated with effort and the time of day.
363
364 To examine how species' habitat use and avoidance varied among regions and seasons,
365 we computed regional trajectories of the strength and direction of the cover class
366 associations. Given the region and the set of predictors to compare, the PI statistics were
367 standardized to sum to 1 across the predictor set for each base model within the specified
368 region. Then, a loess smoother (Cleveland *et al.* 1992) was used to estimate the
369 trajectories of relative predictor importance throughout the year for each predictor.

370 Similarly, a loess smoother was used to estimate the proportion of increasing PD
371 estimates throughout the year for each predictor. Predictors with proportions greater than
372 70% were considered to be increasing and predictors with proportions less than 30%
373 were considered to be decreasing. Predictors with inconsistent directions, those between
374 30 and 70%, were excluded from summaries.
375
376 **Regional Trends**
377 We estimated the inter-annual trends using ensembles of partial effect estimates for year
378 from the ZI-GAM base models. These partial effects are regional-scale estimates that
379 describe the systematic change in relative abundance averaged across the stixel, after
380 accounting for landscape-scale spatial variation associated with elevation and the cover
381 classes. The partial effects of abundance are quantified on the log-link scale where they
382 can be interpreted in units of percent-per-year change.
383
384 Three types of trend summaries were computed. First, to understand how seasonal trends
385 varied geographically, we used a GAM (Wood *et al.* 2016) to generate a spatially explicit
386 estimate of the average percent change per year based on the ensemble of base models
387 with stixel centroid dates falling within a specified season. Second, to identify regions
388 consistently estimated to be in decline, we use a binary response GAM to generate the
389 corresponding spatial estimate of the probability of decline. Third, we estimated the
390 expected trend within a specified region and season by averaging across the set of partial
391 effect estimates with stixel centroids lying within the specified extent. To generate
392 estimates of uncertainty for the expected trend, we computed 90% point-wise confidence

intervals from a Monte Carlo sample of expected trends. To insure that the uncertainty estimates were conservative, each expected trend was based on an average of 5 partial effect estimates, far fewer than the number available.

Model Validation

To assess model quality, we validated the model's ability to predict the observed patterns of occupancy and abundance using independent validation data. The statistics were evaluated using a Monte Carlo design of 25 spatially balanced samples to help control for the uneven spatial distribution of the validation data (Fink *et al.* 2010; Roberts *et al.* 2017). To quantify the predictive performance for the AOO we used the Area Under the Curve (AUC) and Kappa (Cohen 1960) statistics to describe the models' ability to classify un/occupied sites (Freeman & Moisen 2008). AUC measures a model's ability to discriminate between positive and negative observations (Fielding & Bell 1997) as the probability that the model will rank a randomly chosen positive observation higher than a randomly chosen negative one. Cohen's Kappa statistic (Cohen 1960) was designed to measure classification performance taking into account the background prevalence. To quantify the quality of the relative occupancy estimate as a rate within the AOO, we evaluated AUC and Kappa. To quantify the quality of the abundance estimates we computed Spearman's Rank Correlation (SRC) and the percent Poisson Deviance Explained (P-DE). SRC measures how well the abundance estimates rank the observed abundances and the P-DE measures the correspondence between the magnitude of the estimated counts and observed counts.

416 **(c) Results**

417

418 **Weekly AOO, Relative Occupancy and Abundance**

419 Using the Wood Thrush as exemplar analysis, we generated CS-FAC estimates of AOO,
420 relative occupancy and abundance at a spatiotemporal resolution of 8.4km × 8.4km ×
421 1week (Fig 1). A site was considered occupied if the PAT estimator was at least 0.05,
422 meaning that at least one individual of the species is expected to be detected on at least 1
423 out of every 20 independent, standardized eBird surveys of the site on the given day.
424 Unoccupied grid cells were considered to have zero occupancy and abundance, thus the
425 AOO was depicted as the boundary between pixels with and without color.

426

427 To assess the accuracy of estimates, we calculated range-wide validation estimates based
428 on spatially balanced samples of independent eBird observations. AOO weekly median
429 AUC scores were between 0.72 and 0.98 with mean 0.81 (Fig 2a) and AOO weekly
430 median Kappa scores were between 0.16 and 0.38 with mean 0.24 (Fig 2b). Relative
431 occupancy weekly median AUC scores were between 0.66 and 0.86 with mean 0.75 (Fig
432 2c) and relative occupancy weekly median Kappa scores were between 0.18 and 0.51
433 with mean 0.30 (Fig 2d). Relative abundance weekly median P-DE scores were between
434 0.01 and 0.71 with mean 0.20 (Fig 2e) and relative abundance weekly median SRC
435 scores were between 0.24 and 0.52 with mean 0.31 (Fig 2f). Weeks with insufficient
436 validation data were shown as zero. These weeks occurred during the migrations, when
437 detection rates and counts are lowest. Variation in predictive performance was highest
438 during the non-breeding season for all metrics, reflecting lower data densities in

Mesoamerica.

The AOO shows the seasonal changes in the population range size and shape while the abundance estimates capture regional and seasonal variation in population structure within the range. The breeding season range fills in the eastern deciduous forests east of the Great Plains with highest population concentrations in the Appalachian Mountains (Fig 1a). During autumn migration, the population concentrates in the southern part of the Appalachian Mountains (Fig 1b) before crossing the Gulf of Mexico into Central America. The winter distribution (Fig 1c) is concentrated in the Yucatán Peninsula, with lower concentrations extending north into Veracruz and south to Costa Rica and Panama. During the spring migration (Fig 1d), Wood Thrush crosses the Gulf, concentrating on the Gulf Coast and again in the southern part of the Appalachian Mountains.

Seasonal Habitat Use and Avoidance

To quantify changes in habitat use and avoidance throughout the annual cycle, we made weekly estimates of the association between Wood Thrush occupancy and the amount of each habitat class in the local landscape (Fig 3). For each week, the associations were summarized across the population core area, the 5° longitude × 5° latitude area located at the population center for that week. For each cover class, values were combined for both PLAND and LPI predictors to describe the relative strength and direction of the association. Larger absolute values indicate stronger associations and the sign of the value indicates class use or avoidance. Classes with inconsistent direction of association, were removed, resulting in total weekly relative importance that sums to less than 1.

The accuracy of the habitat associations follows from the strong validation results (Fig 2). The Wood Thrush breeding season is characterized by the strong positive association with deciduous broadleaf forest and the non-breeding season is characterized by the strong positive association tropical broadleaf forest. During the migrations, the population is associated with a wider variety of cover classes, and a more even distribution of associations, both positive and negative. This includes a notable positive association with the urban developed class.

Inter-annual Seasonal Trends

We estimated the 2004–16 inter-annual trends during the breeding (May 30–July 3) and non-breeding (Dec 1–Feb 28) seasons. The Wood Thrush population declined across most of its range during the breeding season (Fig 4a). The steepest declines reached 3 to 4% per year along the east coast, and the southeastern and northwest portions of the breeding range (Fig 4a). The green contour indicates where at least 95% of base model trend estimates were declining, a large region including areas in the northeast, the Mid-Atlantic coast, the Piedmont, and the Appalachian Mountains (Fig 4a). The breeding season trend in the Ohio/West Virginia area (Fig 4b), declined at an average of 1% per year, though this trend did show slight increases from 2009 to 2011. In the southern Appalachian Mountains area (Fig 4a) the breeding season trend declined at an average of 3% per year, with the strongest declines from 2004-10 (Fig 4b). The winter population (Fig 5) has declined at an average rate of 5.6% per year, with the largest declines from 2004-2010.

485

486 **(d) Discussion**

487 In this paper, we address challenges related to obtaining cross-scale, full annual cycle
488 information on patterns of abundance and distribution of bird species using citizen
489 science data. The resolution, extent, and comprehensiveness of the information generated
490 with this methodology is unprecedented, and has the potential to increase our knowledge
491 of information-poor species, regions and seasons (Runge *et al.* 2015).

492

493 The approach we present generated robust inferences about species' ranges, occupancy
494 and abundance (Fig. 2), habitat associations, and seasonal trends, confirming the
495 accuracy and utility of the approach. More broadly, the analysis presented here
496 demonstrates how citizen science data can be used to generate accurate species-level
497 information for broad-scale biodiversity monitoring like those outlined by the Group on
498 Earth Observations Biodiversity Observation Network (Kissling *et al.* 2017). It is worth
499 noting that without a single, comprehensive source of information, making population-
500 wide assessments requires the additional steps to acquire, analyze, and calibrate disparate
501 sources of information. Similarly, without critical ancillary information describing
502 participant search effort and information to infer the absence of species (e.g., complete
503 checklists), we would have been unable to account for the bias of imperfect detection. For
504 this reason, we advocate for other citizen science projects to collect ancillary information
505 sufficient to untangle the complexities of heterogeneous observation and ecological
506 processes.

507

508 To demonstrate how this methodology can be used to estimate complex patterns of
509 species' movement, phenology, and population concentration across the full annual cycle
510 we analyzed eBird data for Wood Thrush. This analysis captured important, known
511 patterns of movement, phenology, and concentration during the breeding (Fig 1a) and
512 non-breeding seasons (Fig 1c). Additionally, theses results filled important knowledge
513 gaps, providing novel population-level information during the less well-studied stages of
514 the annual cycle, such as migration and the overwintering period (Fig 1b and Fig 1d).
515 The estimated patterns of habitat use and avoidance (Fig 3) were consistent with
516 documented seasonal (Zuckerberg *et al.* 2016) and regional (Evans *et al.* 2011) patterns.
517 Notably, this is the first comprehensive population-level analysis of habitat associations
518 for not just Wood Thrush, but for any Neotropical migrant. In general, comprehensive
519 information on species' habitat associations will be useful for conservation planning and
520 prioritization, especially outside of the breeding season.
521
522 The population trend estimate for the non-breeding season is, to the best of our
523 knowledge, the first population-wide trend estimate for Wood Thrush outside the
524 breeding season. This trend estimate fills an important gap in understanding the role of
525 the autumn migration and non-breeding season on overall Wood Thrush population
526 health. The population-wide rate of decline during the non-breeding season (-5.6%) was
527 larger than the regional breeding season rates (1-3%), suggesting that limitation is
528 happening during the autumn migration and/or the non-breeding season. Although this is
529 consistent with the demographic models of Taylor & Stutchbury (2016), it contradicts the

530 results from Rushing *et al.* (2017), suggesting that there might be regional variation in
531 population declines in Central America.
532
533 Another novel aspect of the methodology is the ability to use citizen science data to
534 estimate trends in relative abundance - a task usually left to monitoring programs which
535 employ more stringent sampling protocols and are hard to deploy across broad extents.
536 The potential to use eBird data to accurately estimate population trends will grow with
537 the increasing volume and density of data. Increasing volumes and density of data will
538 further improve the precision and spatiotemporal resolution of trend estimates across a
539 wider geographic area than is currently possible. This can be seen when comparing the
540 Wood Thrush breeding and non-breeding season trends (Fig 4 & 5). The increased
541 volume and density of data in the breeding season, made it possible to estimate trends
542 with greater spatial resolution and higher precision than in the non-breeding season. The
543 increasing availability of population trends during the non-breeding season will help to
544 refine our understanding of where and when populations are limited or regulated,
545 complimenting migratory connectivity information derived from individual-level tracking
546 data.
547
548 Aside from filling knowledge gaps, the comprehensive nature of CS-FAC information
549 can be used for other novel and important applications. With CS-FAC distributional
550 information, it is straightforward to make population-wide comparisons and
551 prioritizations and to coordinate conservation activities across regions and seasons.
552 Moreover, once regions of interest have been identified, the resolution of CS-FAC can be

leveraged to seamlessly compare and prioritize landscapes within regions (e.g. Reynolds *et al*. 2017). In addition, the impact of regional and seasonal scale processes can be integrated across space throughout the year, making it possible to carry out accurate multi-scale population-wide impact assessments. This is important for studying a variety of broad-scale environmental and anthropogenic effects, many of which are themselves multi-scale processes, from land-use change to ecosystem services (e.g., La Sorte *et al.* 2017). The potential of our approach to to integrate effects also addresses an important multi-scale challenge in climate change studies (Ådahl *et al.* 2006; Small-Lorenz *et al.* 2013) where nearly all facets of climate (e.g. temperature and precipitation) exhibit strong regional-scale intra-annual variation.

(e) Acknowledgements

We thank the eBird participants for their contributions, the eBird team for their support, and reviewers for their constructive suggestions. This work was funded by The Leon Levy Foundation, The Wolf Creek Foundation, NASA (NNH12ZDA001N-ECOF), Microsoft Azure Research Award (CRM: 0518680), and the National Science Foundation (ABI sustaining: DBI-1356308; computing support from CNS-1059284 and CCF-1522054).

(f) References

Ådahl, E., Lundberg, P., Jonzen, N. (2006) From climate change to population change: the need to consider annual life cycles. *Glob. Change Biol.* 12, 1627– 1633. (doi:10.1111/j.1365-2486.2006.01196.x)

Amatulli, G., Domisch S., Tuanmu M.-N., Parmentier B., Ranipeta A., Malczyk J., and Jetz W. (2017). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, in press (accepted).

Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F. and Pecl, G.T., (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173, 144-154.

Breiman, L., (2001). Random forests. *Machine learning*, 45(1), 5-32.

Boakes, E.H., McGowan, P.J., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K. and Mace, G.M., (2010). Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS biology*, 8(6), p.e1000385.

Bohrer, G., Beck, P.S., Ngene, S.M., Skidmore, A.K. and Douglas-Hamilton, I., (2014). Elephant movement closely tracks precipitation-driven vegetation dynamics in a Kenyan forest-savanna landscape. *Movement Ecology*, 2(1), 2.

Cleveland, W.S., Grosse, E. and Shyu, W.M., (1992). Local regression models. *Statistical models in S*, 2, 309-376.

Cohen, J., (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.

Denes, F.V., Silveira, L.F. and Beissinger, S.R., (2015). Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution*, 6(5), 543-556.

Dickinson, J.L., Zuckerberg, B. and Bonter, D.N., (2010). Citizen science as an ecological research tool: challenges and benefits. *Annual review of ecology, evolution, and systematics*, 41, 149-172.

Diefenbach, D.R., Marshall, M.R., Mattice, J.A. and Brauning, D.W., (2007). Incorporating availability for detection in estimates of bird abundance. *The Auk*, 124(1), 96-106.

Elith, J. and Leathwick, J.R., (2009). Species distribution models: ecological explanation and prediction across space and time. Annual review of ecology, evolution, and systematics, 40, 677-697.

Efron, B., (2014). Estimation and accuracy after model selection. *Journal of the American Statistical Association*, 109(507), 991-1007.

European Bird Census Council, (2016) *Trends of common birds in Europe, 2016 update.* Available at: http://www.ebcc.info/index.php?ID=612. Last accessed 06 December 2017.

Evans, M., Gow, E., Roth, R. R. , Johnson, M. S. and Underwood, T. J. (2011). Wood Thrush (Hylocichla mustelina), version 2.0. In *The Birds of North America* (Rodewald, P. G. editor). Cornell Lab of Ornithology, Ithaca, New York, USA. https://doi.org/10.2173/bna.246

Faaborg, J., Holmes, R.T., Anders, A.D., Bildstein, K.L., Dugger, K.M., Gauthreaux, S.A., Heglund, P., Hobson, K.A., Jahn, A.E., Johnson, D.H. and Latta, S.C., (2010).

Recent advances in understanding migration systems of New World land birds. *Ecological monographs*, 80(1), 3-48.

Fielding, A.H. and Bell, J.F., (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation*, 24(1), 38-49.

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

Fink, D., Damoulas, T. and Dave, J., (2013), July. Adaptive Spatio-Temporal Exploratory Models: Hemisphere-wide species distributions from massively crowdsourced eBird data. In: Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13) July 14–18, 2013 in Bellevue, Washington, USA.

Fink, D., Damoulas, T., Bruns, N.E., La Sorte, F.A., Hochachka, W.M., Gomes, C.P. and Kelling, S., (2014). Crowdsourcing meets ecology: hemisphere-wide spatiotemporal species distribution models. *AI magazine*, 35(2), 19-30.

Fink, D., Auer,T., Obregon, F., Hochachka,W.M., Iliff,M., Sullivan,B., Wood,C., Davies,I., and Kelling, S. (2017). *The eBird Reference Dataset Version 2016 (ERD2016).* Available at: http://ebird.org/ebird/data/download/erd. Last accessed 13 December 2017.

Finley, A.O., (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2(2),143-154.

Fithian, W. and Hastie, T., (2014). Local case-control sampling: Efficient subsampling in imbalanced data sets. *Annals of statistics*, 42(5), 1693.

Freeman, E. A. and Moisen, G. (2008). PresenceAbsence: An R Package for Presence-Absence Model Analysis. *Journal of Statistical Software*, 23(11):1-31. http://www.jstatsoft.org/v23/i11

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment,* 114, 168–182.

Gelfand, A.E., Banerjee, S. and Finley, A.O., (2012). Spatial design for knot selection in knot-based dimension reduction models. In: *Spatio-Temporal Design: Advances in Efficient Data Acquisition*, eds. Mateu, J. and Müller, W.G., John Wiley & Sons., pp.142-169.

Geyer, C. J. (2013) *5601 Notes: The Subsampling Bootstrap*, Available at: http://www.stat.umn.edu/geyer/5601/notes/sub.pdf. Last accessed 06 December 2017.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at the 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21, 1331-1364.

Haste, T., Tibshirani, R., J. Friedman. (2009). *The elements of statistical learning (2nd ed)*. Springer series in statistics, New York.

Heffernan, J.B., Soranno, P.A., Angilletta, M.J., Buckley, L.B., Gruner, D.S., Keitt, T.H., Kellner, J.R., Kominoski, J.S., Rocha, A.V., Xiao, J. and Harms, T.K., (2014). Macrosystems ecology: understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment*, 12(1), 5-14.

Hochachka, W.M., Fink, D., Hutchinson, R.A., Sheldon, D., Wong, W.K. and Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in ecology & evolution*, 27(2), 130-137.

Hurlbert, A.H. and Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *PNAS*, 104(33), 13384-13389.

Johnston, A., Fink, D., Reynolds, M.D., Hochachka, W.M., Sullivan, B.L., Bruns, N.E., Hallstein, E., Merrifield, M.S., Matsumoto, S. and Kelling, S. (2015). Abundance models improve spatial and temporal prioritization of conservation resources. *Ecological Applications*, 25(7), 1749-1756.

Kissling, W.D., Ahumada, J.A., Bowser, A., Fernandez, M., Fernández, N., García, E.A., Guralnick, R.P., Isaac, N.J., Kelling, S., Los, W. and McRae, L. (2017). Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biological reviews*. doi: 10.1111/brv.12359

Kearney, M. & Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, 12, 334–350.

Kéry, M. and Royle, J.A., (2015). *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models.* Academic Press.

La Sorte, F.A., Fink, D., Blancher, P.J., Rodewald, A.D., Ruiz-Gutierrez, V., Rosenberg, K.V., Hochachka, W.M., Verburg, P.H. and Kelling, S., (2017). Global change and the distributional dynamics of migratory bird populations wintering in Central America. *Global Change Biology*, 23(12), 5284-5296.

Marra P.P., Cohen, E.B., Loss, S.R., Rutter, J.E., Tonra, C.M. (2015). A call for full annual cycle research in animal ecology. *Biol. Lett.* 11: 20150552. http://dx.doi.org/10.1098/rsbl.2015.0552

Mease, D., Wyner, A. and Buja, A., (2007). Cost-weighted boosting with jittering and over/under-sampling: JOUS-boost. *J. Machine Learning Research*, 8, 409-439.

McGarigal, K., Cushman, S.A., and Ene, E. (2012). *FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps.* Available at: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

Norris, D.R. and Marra, P.P. (2007). Seasonal interactions, habitat quality, and population dynamics in migratory birds. *The Condor*, 109(3), 535-547.

Pereira, J. A., and A. J. Novaro. (2014). Habitat-specific demography and conservation of Geoffroy's cats in a human-dominated landscape. *Journal of Mammalogy* 95, 1025–1035.

Politis, D. N., Romano, J.P., and Wolf, M. (1999). *Subsampling, Springer Series in Statistics*, Springer-Verlag New York, Inc.

R Development Core Team (2017). *R: A language and environment for statistical computing. R Foundation for Statistical Computing* http://www.R–project.org, Vienna, Austria.

Reynolds, M.D., Sullivan, B.L., Hallstein, E., Matsumoto, S., Kelling, S., Merrifield, M., Fink, D., Johnston, A., Hochachka, W.M., Bruns, N.E. and Reiter, M.E., 2017. Dynamic conservation for migratory species. *Science Advances*, 3(8), p.e1700707.

Ridgeley, R.S., Allnutt, T.F., Brooks, T., NcNicol, D.K., Mehlman, D.W., Young, B.E. and Zook, J.R., BirdLife International. 2012. Digital distribution maps of the birds of the Western Hemisphere, version, 5.

Ridgeway, G. with contributions from others. (2017). gbm: Generalized Boosted Regression Models. R package version 2.1.3. https://CRAN.R-project.org/package=gbm

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W. and Warton, D.I., (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.

Robinson, O.J., Ruiz-Gutierrez, V., Fink, D. (2017). Correcting for bias in distribution modeling for rare species using citizen science data. *Diversity and Distributions*, in press (accepted).

Runge, C.A., Watson, J.E., Butchart, S.H., Hanson, J.O., Possingham, H.P. and Fuller, R.A., (2015). Protected areas and global conservation of migratory birds. *Science*, 350(6265), pp.1255-1258.

Rushing, C.S., Hostetler, J.A., Sillett, T.S., Marra, P.P., Rotenberg, J.A. and Ryder, T.B., (2017). Spatial and temporal drivers of avian population dynamics across the annual cycle. *Ecology*, 98(11), 2837–2850.

Sauer, J.R. and Link, W.A., (2011). Analysis of the North American breeding bird survey using hierarchical models. *The Auk*, 128(1), 87-98.

Sherry, T.W. & Holmes, R.T. (1995) Summer versus winter limitation of populations: what are the issues and what is the evidence. In: *Ecology and Management of*

Neotropical Migratory Birds: a Synthesis and Review of Critical Issues (eds T.E. Martin & D.M. Finch), Oxford University Press, New York, pp. 85–120.

Small-Lorenz, S.L., Culp, L.A., Ryder, T.B., Will, T.C., Marra, P.P. (2013) A blind spot in climate change vulnerability assessments. *Nat. Clim. Change* 3, 91 – 93. (doi:10.1038/nclimate1810)

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

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dietterich, T., Farnsworth, A. and Fink, D., (2014). The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, 169, 31-40.

Wood, S.N., Pya, N. and Säfken, B., (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), pp.1548-1563.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L. and Storlie, C., (2014). SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. *R package version*, 1-1.

Zaharia, M., Xin, R.S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M.J. and Ghodsi, A., (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM*, 59(11), 56-65.

Zuckerberg, B., Fink, D., La Sorte, F.A., Hochachka, W.M. and Kelling, S., (2016). Novel seasonal land cover associations for eastern North American forest birds

identified through dynamic species distribution modelling. *Diversity and Distributions*, 22(6), 717-730.

Table 1: Land and water cover classes used for distribution modeling. All cover classes were summarized within a 2.8km × 2.8km (784 hectares) landscape centered on each checklist location. Within each landscape, we computed the proportion of each class, and three descriptions of the spatial configuration of the class within the landscape.

Figure 1: Wood Thrush estimates of area of occupancy and relative abundance at 8.4km × 8.4km resolution during (a) breeding (June 20), (b) autumn migration (October 3), (c) non-breeding (December 12), and (d) spring migration (March 28) seasons. Positive abundance is only shown in areas estimated to be occupied and the area of occupancy is depicted as the boundary between pixels with and without color. Brighter colors indicate areas occupied with higher abundance. Relative abundance was measured as the expected count of the species on a standardized 1km survey conducted from 7-8AM.

Figure 2: Boxplots of range-wide weekly predictive performance for area of occupancy, relative occupancy and abundance estimates across 25 Monte Carlo samples of spatially balanced validation data. (a) AUC and (b) Kappa scores for area of occupancy estimates. (c) AUC and (d) Kappa scores for relative occupancy estimates. (e) Spearman's Rank Correlation and (f) Percent Deviance Explained scores for relative abundance estimates.

Figure 3: The weekly relative importance for the amount of each land and water cover class for the core Wood Thrush population. Positive importance indicates class use and negative importance indicates class avoidance. The strength of the association with each class is proportional to the width of the class color. Classes with inconsistent direction of

association were removed, resulting in total weekly relative importance that sums to less than 1.

Figure 4: Wood Thrush breeding trend map and regional estimates. The map shows the average percent change per year in relative abundance from 2004–16 during the breeding season (May 30– July 3). The green contour indicates the region where the probability of decline is at least 95%. Breeding season regional trends and 90% point-wise confidence intervals are shown for the (A) Ohio-West Virginia and (B) Southern Appalachian regions as the deviation from the mean on the log scale.

Figure 5: Wood Thrush non-breeding trend map and regional estimate. The map shows the average percent change per year in relative abundance from 2004–16 during the non-breeding season (Dec 1–Feb 28). The regional trend and 90% point-wise confidence intervals are shown as the deviation from the mean on the log scale.

MODIS Class Codes	Land Cover	Water Cover
0		Shallow Ocean
1	Evergreen Needleleaf Forest	
2	Evergreen Broadleaf Forest	Ocean coastlines and lake shores
3	Deciduous Needleleaf Forest	Shallow inland water
4	Deciduous Broadleaf Forest	
5	Mixed Forest	Deep Inland Water
6	Closed Shrublands	Moderate or continental ocean
7	Open Shrublands	Deep Ocean
8	Woody Savannas	
9	Savannas	
10	Grasslands	
12	Croplands	
13	Urban and built-up	
16	Barren	

Table 1

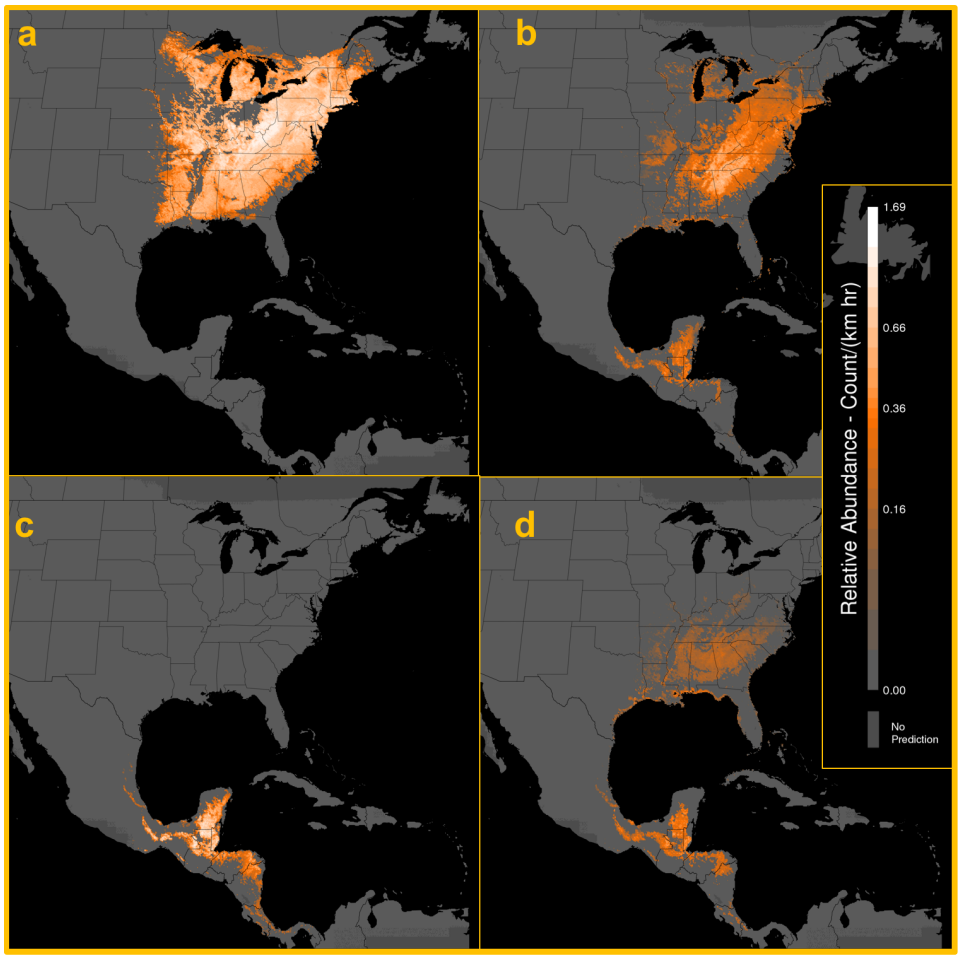
a
b
c
d
Relative Abundance - Count/(km hr)
1.69
0.66
0.36
0.16
0.00
No Prediction

Figure 1

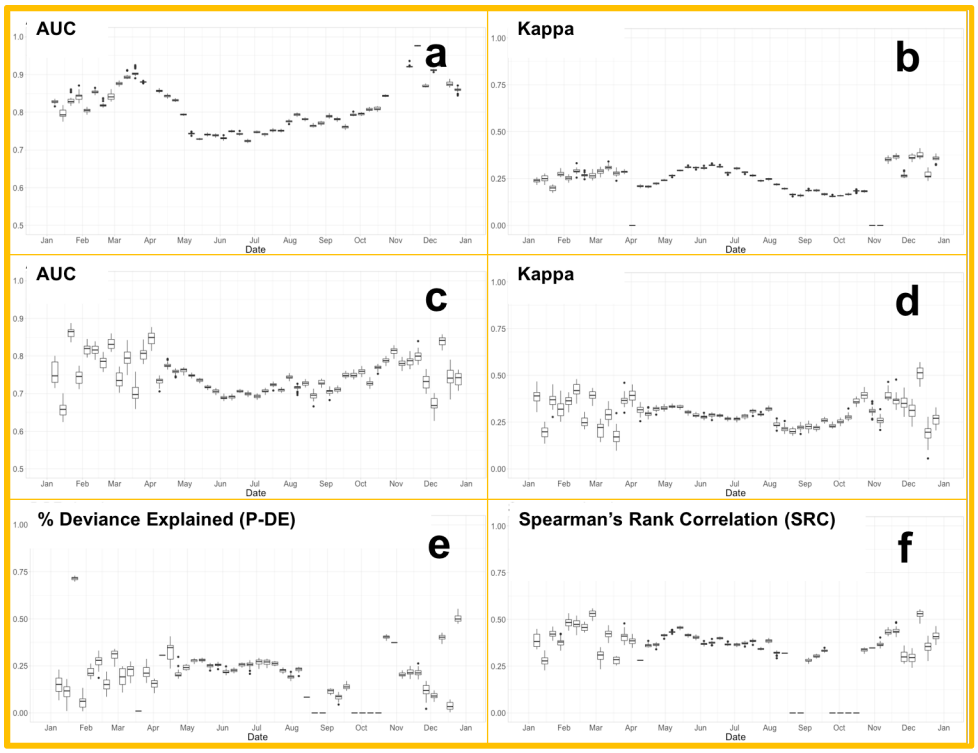
AUC
a
Kappa
b
AUC
c
Kappa
d
% Deviance Explained (P-DE)
e
Spearman's Rank Correlation (SRC)
f
Date

Figure 2

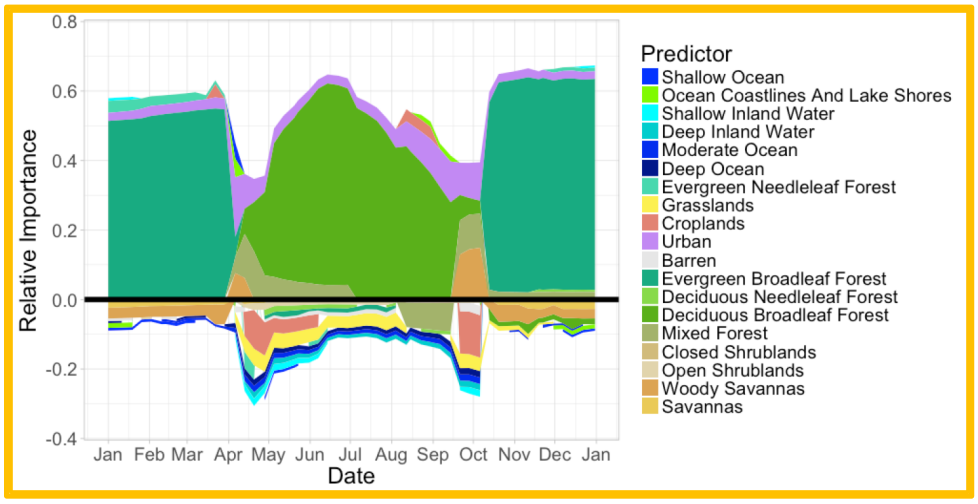
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
Relative Importance
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
Date
Predictor
Shallow Ocean
Ocean Coastlines And Lake Shores
Shallow Inland Water
Deep Inland Water
Moderate Ocean
Deep Ocean
Evergreen Needleleaf Forest
Grasslands
Croplands
Urban
Barren
Evergreen Broadleaf Forest
Deciduous Needleleaf Forest
Deciduous Broadleaf Forest
Mixed Forest
Closed Shrublands
Open Shrublands
Woody Savannas
Savannas

Figure 3

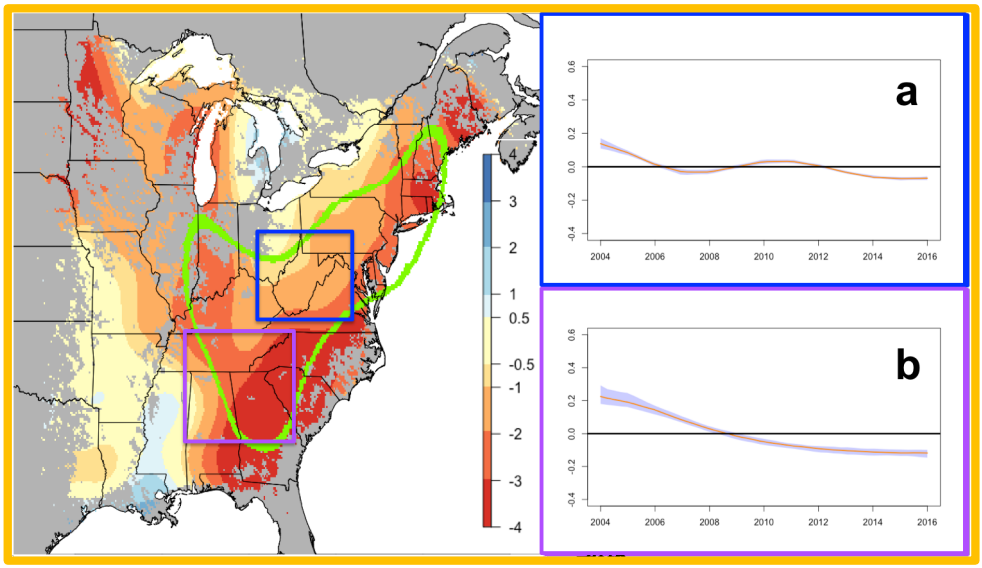

Figure 4

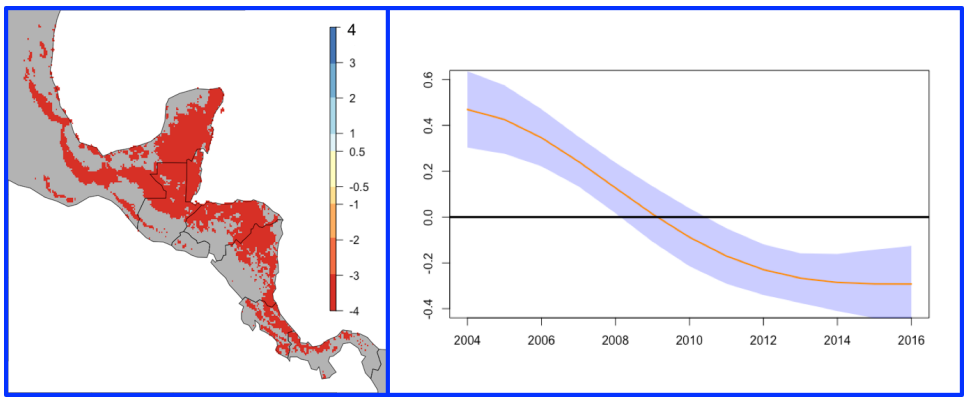
4
3
2
1
0.5
-0.5
-1
-2
-3
-4
0.6
0.4
0.2
0.0
-0.2
-0.4
2004
2006
2008
2010
2012
2014
2016

Figure 5

Supporting Information for

"Modeling Avian Full Annual Cycle Distribution and Population Trends with Citizen Science Data"

Daniel Fink, Tom Auer, Viviana Ruiz-Gutierrez, Wesley M. Hochachka, Alison Johnston, Frank A. La Sorte, and Steve Kelling

This document includes supplemental information on the

1. STEM ensemble design, and
2. Subsampling procedures to estimate uncertainty of the occupancy and abundance estimates.

These are described in the appendices below.

Appendix S1: Ensemble Design

The ensemble of stixels was designed as a Monte Carlo sample of 100 randomly located spatiotemporal partitions of the spatiotemporal study extent. This results in a sample of stixels uniformly distributed through out the spatiotemporal extent of study. Averaging across this sample helps control for biases associated with the arbitrary partitioning of data into stixels. We also use the Monte Carlo sample to generate estimates of uncertainty by incorporating subsampling into the selection of training data within stixels.

An important part of the STEM implementation was determining the spatial and temporal dimensions of the stixels. When averaging across the ensemble, stixel size controls an important bias-variance tradeoff (Fink *et al* 2010; Fink *et al.* 2013). Stixel size needs to

be chosen small enough to capture local predictor-response (i.e. species-environment) relationships, controlling the bias of base model estimates. Stixel size also needs to be chosen large enough to meet the minimum sample size requirements necessary for fitting the base models: this controls the variance when averaging across the ensemble.

As stixel size gets smaller, the training data sample size within a stixel also decreases. When training sample sizes are too small to fit base models, the number of base model estimates available for ensemble averaging also decreases, increasing the variance of the ensemble estimator. The number of stixels used to compute a local estimate across the ensemble is called the *ensemble support*. Ensemble support is important because it determines effectiveness of ensemble averaging to control inter-model variability. In general, ensemble support follows patterns of data density, filtered through a combination of stixel geometry and the base model minimum sample size requirements.

Because of the irregular and often sparse distribution of eBird data, selecting the spatial and temporal dimensions of stixels necessary to maintain ensemble support is a non-trivial process. Our goal was to select the stixel size parameters to maximize the spatiotemporal coverage of the analysis while maintaining sufficient ensemble support to guarantee good model performance. To operationalize this, we required an ensemble support of at least 50 stixels, throughout at least 75% the Western Hemisphere for each week of the year. Estimates of occupancy and abundance were only produced when ensemble support was above 50.

We began by specifying the temporal dimension of the partitions to be 30 contiguous days. A 30 day window is small enough to capture a wide variety of migration patterns across a diverse set of terrestrial species using eBird data (Johnston *et al.* 2015; La Sorte *et al.* 2017). For simplicity, we decided to divide space into longitude-latitude squares

using un-projected latitude × longitude space. To account for the relatively low data density (Fink *et al.* 2013) in the southern part of the study area we doubled the length of stixel squares south of 12° latitude. Given the 30-day temporal dimension and minimum sample size requirements (see below for details), we found that a stixel length of 10° in the north and 20° in the south met our requirements, resulting in a minimum weekly coverage of 79% of the terrestrial Western Hemisphere in December and up to 90% coverage of the Western Hemisphere in July. Figure S1 shows realizations of 3 randomly located spatial partitions used to define STEM stixels for the analysis. This image shows how stixels overlap across the randomly located partitions and it shows how stixel size varies between North and South America.

The base model minimum sample size requirements affect base model bias and variance as well as ensemble support. To fit a base model, we required that the training data meet the following three criteria. First, we required a minimum sample size of 50 checklists. Second, to insure minimum spatial coverage within each base model, we required that at least 50 cells from the spatially balanced case-control sampling procedure (see below for details) contained checklists. Finally, to insure a minimum signal to predict positive relative occupancy and abundance, we required at least 10 species detections, i.e. non-zero counts, among the checklists. To guard against the effects of replicate surveys at popular birding locations, only one detection per day is considered from each location.

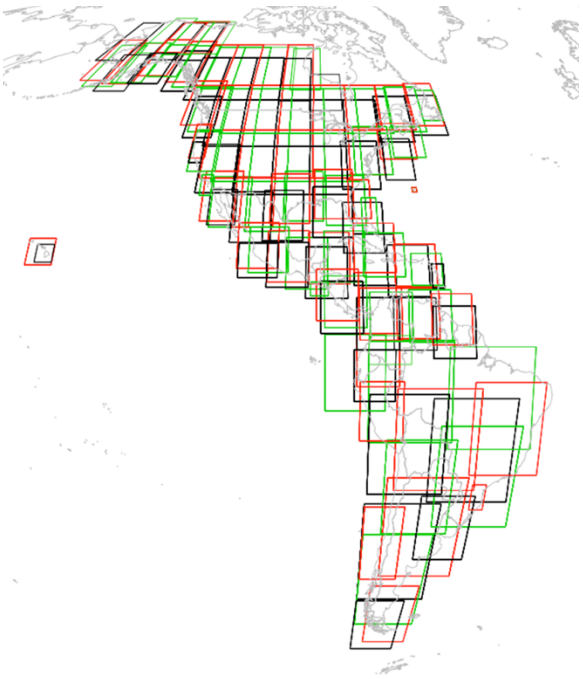

FIGURE S1: Realizations of three randomly located spatial partitions used to define STEM stixels. This image shows how stixels overlap across the randomly located partitions and it shows how stixel size varies between North and South America. One hundred randomized partitions were used for the analysis.

Appendix S2: Subsampling procedures to estimate uncertainty

The ensemble means are used to estimate relative occupancy and abundance. Consequently, the variation across the ensemble itself provides a conservative estimate of uncertainty for the ensemble mean (Efron 2014). A straightforward, brute force approach way to generate more accurate estimates of uncertainty for the ensemble mean can be computed by bootstrapping the ensemble mean, however, this is computationally prohibitive.

Instead, we employed a subsampling approach (Politis *et al.* 2009) following the computational strategy of Geyer (2013). We faced two challenges to implement this approach. First, the sample size, here, the ensemble support, was very small, 100 at most. Second, the computational efficiency of the approach was very important because we needed to compute uncertainty estimates for up to 86M quantities per species (= 614K locations / week * 52 weeks * 2 estimates per location [occupancy & abundance] + another 28M for the 14M training & testing checklists * 2 estimates per checklist). To deal these challenges we selected a set of parameter settings that balanced the quality of the interval estimates with the computational costs of generating them.

When the sample size was less than 10, subsampling was not performed and quantiles of the original sample were to estimate uncertainty. For sample sizes greater than or equal to ten, we computed the upper 90th confidence limit and lower 10th confidence limit. The subsampling was performed with two different sizes to facilitate estimation of the rate parameter used to correct the uncertainty estimates. Following Geyer (2013), we subsampled the square root of the ensemble support value and the -1.5 power of the ensemble support value.

To check these parameter settings, a small simulation test was run. We found that for sample sizes of 25 or less, the rate parameter estimates tended to be too small, resulting in intervals that were too small and had poor coverage. To mitigate this, we adjusted the rate parameter estimate upwards by 0.5 of the rate parameter's standard error, producing more

conservative uncertainty estimates. In the cases where the rate parameter estimate was negative, subsampling was not performed and quantiles of the entire sample were used producing conservative uncertainty estimates. Note that ensemble support requirements for the occupancy and abundance estimates excludes most of the estimates suffering from these small sample size complications.

References

Efron, B., (2014). Estimation and accuracy after model selection. Journal of the American Statistical Association, 109(507), 991-1007.

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

Fink, D., Damoulas, T. and Dave, J., (2013), July. Adaptive Spatio-Temporal Exploratory Models: Hemisphere-wide species distributions from massively crowdsourced eBird data. In: *Twenty-Seventh AAAI Conference on Artificial Intelligence* (AAAI-13) July 14–18, 2013 in Bellevue, Washington, USA.

Geyer, C. J. (2013) 5601 Notes: The Subsampling Bootstrap, Available at: http://www.stat.umn.edu/geyer/5601/notes/sub.pdf. Last accessed 06 December 2017.

Johnston, A., Fink, D., Reynolds, M.D., Hochachka, W.M., Sullivan, B.L., Bruns, N.E., Hallstein, E., Merrifield, M.S., Matsumoto, S. and Kelling, S. (2015). Abundance models improve spatial and temporal prioritization of conservation resources. *Ecological Applications*, 25(7), 1749-1756.

La Sorte, F.A., Fink, D., Blancher, P.J., Rodewald, A.D., Ruiz-Gutierrez, V., Rosenberg, K.V., Hochachka, W.M., Verburg, P.H. and Kelling, S., (2017). Global change and the distributional dynamics of migratory bird populations wintering in Central America. *Global Change Biology*, 23(12), 5284-5296.Mease *et al.* 2007

Politis, D. N., Romano, J.P., and Wolf, M. (1999). *Subsampling*, *Springer Series in Statistics*, Springer-Verlag New York, Inc.