Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity

STEVE KELLING, ALISON JOHNSTON, ALETTA BONN, DANIEL FINK, VIVIANA RUIZ-GUTIERREZ, RICK BONNEY, MIGUEL FERNANDEZ, WESLEY M. HOCHACHKA, ROMAIN JULLIARD, ROLAND KRAEMER, AND ROBERT GURALNICK

Biodiversity is being lost at an unprecedented rate, and monitoring is crucial for understanding the causal drivers and assessing solutions. Most biodiversity monitoring data are collected by volunteers through citizen science projects, and often crucial information is lacking to account for the inevitable biases that observers introduce during data collection. We contend that citizen science projects intended to support biodiversity monitoring must gather information about the observation process as well as species occurrence. We illustrate this using eBird, a global citizen science project that collects information on bird occurrences as well as vital contextual information on the observation process while maintaining broad participation. Our fundamental argument is that regardless of what species are being monitored, when citizen science projects collect a small set of basic information about how participants make their observations, the scientific value of the data collected will be dramatically improved.

Keywords: citizen science, biodiversity monitoring, species distributions, citizen, science survey design

Biodiversity monitoring provides essential information and evidence to develop species conservation strategies and inform the sustainable use of natural resources. Traditionally, monitoring programs rely on humans to collect field observations (Kelling et al. 2013); however, recent advances in machine learning are improving the ability of automated systems to detect and classify organisms (Schneider et al. 2018, Zhang et al. 2018). Because governments and scientific agencies often lack resources to support long-term biodiversity assessments by professional scientists (Balmford and Gaston 1999, Bland et al. 2015), many organizations recruit volunteers—both beginners and highly skilled ones—to meet these assessment goals (Danielsen et al. 2014, Pimm et al. 2015). Worldwide, up to 85% of the species-level information required by governments is collected by volunteers (Roy et al. 2012).

Thousands of citizen science projects enlist the public in gathering or processing scientific data (see http://scistarter.com), with hundreds of these projects collecting species observations (Theobald et al. 2015). Although biodiversity science has a long history in working with volunteered data—e.g., through museums—citizen science, as a formal activity, began to coalesce between 2006 and 2010, and the number of peer-reviewed publications using citizen science data began to grow exponentially (McKinley et al. 2015). Technical advances such as the Internet, social media, and mobile/handheld computers helped to engage many more participants, locally and globally, and these projects are now gathering or processing hundreds of millions of observations annually (Chandler et al. 2016). However, these projects vary greatly in the types of information that they collect, with important consequences for the ability of each to meet its intended outcomes for science and society.

In this article, we argue that citizen science projects aimed at robustly monitoring species distributions should follow several basic data-collection principles to provide a solid foundation for data analysis. We build on existing recommendations for biological monitoring programs that collect sufficient information on observation processes such that resulting data can be used for statistical analyses (Yoccoz et al. 2001). Although, in this article, we focus on birds and use eBird as an exemplar, our intent is to emphasize that anyone designing, or redesigning, a citizen science project that is gathering observational data on any organism or any taxa should follow these principles.

doi:10.1093/biosci/biz010

Table 1. Characteristics of unstructured, semistructured and structured citizen science projects as defined in this article.

	Project elements	Unstructured	Semistructured	Structured
Why What	**Objectives** • Clear objectives • Clear planned data analysis	•	●	⬤
How When Where	**Survey design** • Target sample size • Locations selected to fit design • Trained or expert observers			⬤
How When Where	**Rigorous protocol** • Preselected locations • Ability to estimate detectability • Specifications for date, time of day, etc. • Specialized equipment required		•	⬤
Who	**Open and flexible recruitment** • Open • Accessible • Appealing to wide range of observers	⬤	⬤	
How When Where Who	**Observation process recorded** • Time of day, date, location • Number of observers • Distance traveled • Style/type of survey		⬤	
	Variation accounted for by analytical techniques		⬤	
	Variation controlled by restricted protocol			⬤

Note: The size of the dots relates approximately to the extent or frequency with which projects of different types meet the various criteria.

We begin by describing the spectrum of data collection survey methodologies employed by citizen science projects. Next, we compare and contrast the ecological and observational processes involved in collecting data and how they affect the analyses of citizen science data. Third, we recommend several core types of information that all citizen science projects should collect that allow analysts to estimate species distribution. Finally, we describe how eBird data provide meaningful information on the distribution and abundance of bird populations. If citizen science projects collect a small set of basic information about how the participants make their observations, in addition to the species observations, the scientific value of the data collected will be dramatically improved.

The spectrum of citizen science biodiversity surveys

Citizen science projects that gather species observations of organisms can be considered on a continuum reflecting how data-collection events occur, from preplanned and structured to opportunistic and unstructured. Structured surveys are composed of rigorous protocols designed to meet well-defined objectives (table 1). For example, the United Kingdom Butterfly Monitoring Scheme, which is designed to monitor the population status of butterflies, has surveyed 1200 transects since 1976. Each transect is chosen to sample evenly the habitat types and management activity on varied sites. Transects are typically 2–4 kilometers long and are walked weekly between 1 April and 29 September, on days when weather conditions are favorable for encountering butterflies. Counts of all butterflies are recorded in a 5 meters wide band along each transect, and if a species is not reported, it is considered undetected or absent. The participants are highly motivated and trained in the survey methodology. Although the method is labor intensive, the project has allowed the calculation of robust site-level trend estimates for more than 50 species of butterflies (Fox et al. 2011) and has contributed to more than 100 peer-reviewed publications.

At the opposite end of the continuum unstructured projects have few requirements for data collection, and the participants report only when they opportunistically encounter a species (table 1). For example, iNaturalist is successful in recruiting volunteers to submit observations and images of species they detect. However, the observations have limited application for monitoring populations of organisms because the project collects little information about how the participant made their observations, which are crucial for subsequent scientific analyses (Guillera-Arroita et al. 2015). The majority of citizen science projects that gather ecological data can be characterized as unstructured (Pocock et al. 2017) and do not gather information on the observation process, meaning that there is no fully statistically defensible way of accounting for the biases inherent in the data collection.

To enlist a large number of participants while gathering interpretable data, a citizen science project must balance the trade-offs between strict data-collection protocols that provide high scientific value and flexible data-collection protocols that appeal to a wide audience (Bonney et al. 2009). To accomplish this balance we suggest a semistructured data collection process, which can yield enjoyable projects that attract large numbers of participants while collecting sufficient information to account for variation and bias in the data-collection process (table 1). Whereas a highly structured project attempts to eliminate variation in the data collection with strict protocols, a semistructured project gathers just enough data to control variation and bias during analysis (Fink et al. 2010).

One example of a semistructured survey project that allows open participation and observer-selected sites is eBird (Sullivan et al. 2014). What characterizes eBird as a semistructured survey is that it provides the option to collect information on how the participant made their observations; for example, the duration of data collection, the distance an observer traveled while collecting observations, inferring the nondetection of a species (e.g., from a complete checklist), and other facets of the data-collection event that can affect the probability that a species will be detected, identified, and recorded. Semistructured projects such as eBird can be very popular, and the volume of data they gather provides the largest and fastest-growing information on species occurrence (Amano et al. 2016).

A process for creating a citizen science project

In developing a citizen science project, it is critical for designers to understand that the data to be collected represent a combination of two processes: an ecological process that determines which species exist in a given location and an observation process that a participant uses to make a sighting, which has inherent biases. When the two processes are confounded in the collected data, critical interpretations of the data may be limited or misleading (figure 1). For example, a pattern observed about species occurrence depends not only on the actual distribution of the species but on the distribution of locations that were sampled. If data are lacking about the observation processes, then occurrence maps will be impossible to interpret accurately because the apparent distribution is determined to an unknown degree by where the participants are likely to sample.

Citizen science projects often overlook the importance of accounting for variation in the observation process. Bias and noise in data can be thought of as having two general sources: uneven sampling effort over space and time (Geldmann et al. 2016) and variation in rates at which species are detected, identified, and reported (Kelling et al. 2015, Kery and Schmid 2004). Structured projects control these sources of bias by implementing a sampling design and formal protocol that constrain variation in the data-collection process. The result is that structured projects focus on implementing a sampling design that feeds into a preselected analysis. An example is making multiple visits to a survey site to collect detection/nondetection observations that can be used to estimate species occupancy rates (e.g., MacKenzie et al. 2006). Locations and survey times are preselected, and the participants are well-trained in following the protocols. Structured surveys gather the most information-rich data but often require high participant dedication, such as the weekly data collection visits required by the United Kingdom Butterfy Monitoring Project, and thus significant effort in project coordination. In addition, maintaining interest by trained observers can be difficult, and extending structured surveys over broad spatial and temporal extents is often impossible, because of the small number of expert, trained, and motivated observers.

In contrast, semistructured projects have minimal survey design, use flexible and straightforward data-collection protocols, allow for broad participation, and do not limit where and when observers make observations. However, they do gather information on both the ecological and the observation processes including detection/nondetection data and information on features that vary among data-collection events, such as the location, time of day, duration of collection event, and distance traveled (table 1). Although analyses of semistructured data are complex, data collection biases can be addressed because the observation processes can be modeled. This provides a basis for estimating variation in abundance and distribution over space and time (Johnston et al. 2018).

Identifying the information required to describe the observation process

We emphasize that when designing a project, or currently managing an ongoing project, project managers should not

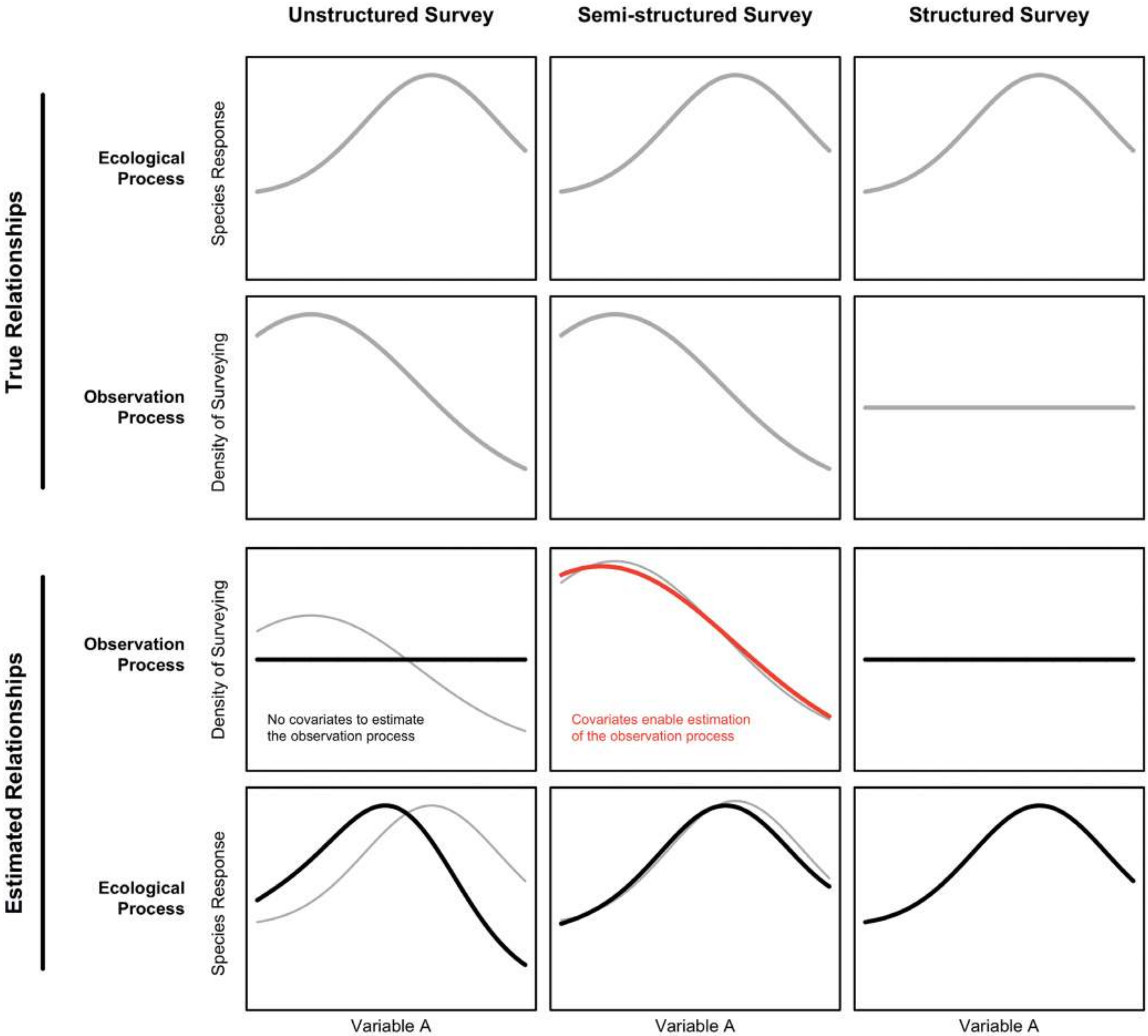

Figure 1. Schematic visualization of the observation process and ecological process for unstructured, semistructured and structured surveys, in relation to an imaginary variable A. The top two rows show the true ecological and observations processes in grey. Both unstructured and semistructured surveys show uneven sampling in relation to variable A. The lower two rows show the estimated ecological and observation processes in black and the true processes in grey underneath. The data on the observation process from the the semistructured survey are used to estimate the biased observation process (in red), which enables the estimated ecological process to be closer to the truth, when compared with the unstructured survey. When a biased observation process cannot be estimated analytically, the ecological and observation processes are confounded.

only determine what is to be observed, and also provide the means for the participants to describe the process by which each observation will be made. In the present article, we provide six key questions that, when answered, define the facets of the observation process important for controlling bias.

Why is the project being conducted? Every citizen science project should have a clearly articulated purpose to motivate the project design. This should be based on either a research question or monitoring agenda that will determine what should be observed and how data need to be collected (Bonney et al. 2009). We recommend clearly describing the project objectives and design, even if broad, so that an appropriate data-collection methodology can be implemented. Clearly articulated project objectives will be more engaging for the participants, and will lead to more homogeneous observation behavior, helping further analysis.

What is observed? Most citizen science monitoring projects identify a specific taxonomic or ecological scope, because developing protocols that effectively gather observations of one taxon (e.g., mammals) would be different from another (e.g., stream invertebrates). We recommend clearly defining the taxonomic scope to ensure that the observation process is well described and as straightforward as possible.

Where are the observations collected? Most semi- or unstructured projects allow the observer to select their preferred observation location. Regardless of how a location is selected, we recommend accurately recording where data are collected. Spatial aggregation to lower resolution is always possible at a later stage, e.g., for species of conservation concern, if needed.

When are the observations collected? Species detectability can change within a day or seasonally as animal behaviors change or plants alter their appearance. However, limiting sampling events to a short time frame in a given year may limit the participation of volunteers and miss key phenological events or changes. We recommend always recording the date and time when observations were made, including the duration of the observation event.

Who is making the observations? Most citizen science programs do not restrict participation and have a range of participants from very dedicated recorders to others who submit data only occasionally. Although the participants can develop enormous expertise in gathering information through participation, a large variation can exist between individuals in their behavior and their detection and classification skills (Fitzpatrick et al. 2009). Regardless of the monitoring project, knowing who collected each observation is important. In this way bias related to variability in observer behavior and expertise can be estimated and taken into account. We recommend retaining clear information about who collected the observations in the data management framework, by using coded identifiers that are unique to individual observer.

How are the observations collected? A high level of data quality can be maintained in the absence of a rigid sampling design and protocols if the observation event is well described. Not only should this description include information on what, where, when, and who made the observations, but also how observations of species are made. Recording all species generates a "complete checklist," which directly provides information on species detected and inferred information on species not detected. Knowing whether a checklist of species is "complete" is critical because the nondetections are required to estimate detection probabilities, used to infer true species presence and absences (Guillera-Arroita et al. 2015). Other key aspects of how the observations were collected may include the distance travelled or any variables that describe variation in the area surveyed or the method of surveying, which affect the detectability of individuals.

eBird, a semistructured citizen science project

eBird is a semistructured citizen science project with an open project design and minimal protocol requirements (Sullivan et al. 2014). To date, eBird's more than 400,000 observers have volunteered 40 million hours of effort in reporting bird observations. No restrictions are placed on who can participate, where and when they participate, or how they participate. However, observers are encouraged to submit "complete checklists." eBird has maintained 20% or more annual growth in data collection for more than a decade, and as of October 2018, eBird data include 31 million complete checklists containing 567 million observations of more than 10,000 bird species from every country in the world.

An important part of these complete checklists is the description of the observation process. eBird answers the six questions above as follows:

Why is eBird being conducted? eBird collects data that will be used to estimate the distribution and abundance of bird populations by taking advantage of the global network of bird enthusiasts who submit their observations to a central data repository.

What is observed? eBird gathers observations and counts of all wild bird species.

Where are the observations collected? eBird participants can select where they make their observations from any global location. All locations are georeferenced either through mapping tools provided on the eBird website or through the GPS system available on mobile phones and used within the freely available eBird App. The mobile App also records the track and distance the observer travelled while making observations.

When are the observations collected? eBird allows observers to record observations at any time of day or year. The start time of an observation event is recorded along with the total time spent making observations.

Who is making the observations? Individuals must register for eBird and then log in whenever they submit a checklist. In this way all submissions are linked to the individual who submitted them. Checklists can be shared to link multiple observers to the same checklist.

How are the observations collected? eBird collects lists of all species identified along with counts of the number of individuals of each species observed during the collection event. By asking observers if they are reporting a list of *all* the species they identified, analysts can better infer detectability and therefore absence of species. By collecting counts of each species, analysts can better estimate population abundance (Johnston et al. 2015). eBird also collects information on whether the observer was stationary or traveling, the distance they traveled, the duration of the checklist, and the number of observers.

eBird complete checklists are translated into information describing patterns of species occurrence and abundance in space and time using Species Distribution or Niche models (SDMs). These statistical models estimate the distribution or abundance of a species by estimating relationships between the observed patterns of species occurrence, or counts, and data describing the processes that give rise to these observations (Franklin 2009). To understand ecological processes, each location where eBird data are collected is linked to remotely sensed habitat and elevation data, which are included as explanatory variables in the SDM.

Comparable to modeling ecological processes, similar strategies can be used to account for sampling biases inherent in the observation process in the SDM by using the information that eBird collects about the observation process. First, to limit differences in sampling effort, the data used to train the SDM can be limited to complete checklists with similar levels of sampling effort. In addition, to account for the effects of variable search effort within the selected data, the duration and length of each search are included as explanatory variables in the SDM. To account for variation in detection rates associated with changes in species' behavior throughout each day, the time of day that each search is made is included as an explanatory variable. Finally, to account for the strong differences between observer's ability to detect and identify species, we have calculated the Checklist Calibration Index (Kelling et al. 2015), which is derived from the rates at which observers accumulate additional species with increasing effort. We also include the index as an explanatory variable in the SDM. In our experience, these explanatory variables that describe the observation process are some of the most important predictors of variation in the data fit by the SDMs. Including these variables leads to improved predictive performance and provides an inferential basis for separating the ecological and observation processes. This allows researchers to model the spatiotemporal variation in bird distributions throughout a species' annual cycle at continental scales (Fink et al. 2014, Fink et al. 2010).

Above, we list the principles of good practice for analyzing semistructured citizen science data, such as eBird. In a forthcoming article, we outline the practical steps for analyzing eBird data that follow these principles (Johnston et al. In prep). We provide researchers with code and practical guidance for applying these principles in their own analyses of species distributions. Appropriate filters on the observation process variables reduce the variation in the analysed data. Including the observation process variables as covariates in the analysis accounts for the remaining variation in the analysed data. Following these steps leads to better estimates of species distributions and more robust ecological conclusions.

For each individual species, SDMs generate a series of data products including range-wide, seasonal relative abundance estimates (figure 2) and weekly relative abundance estimates (figure 3). The range-wide, seasonal relative abundance estimates are meant to show the population of each species across its entire distribution. The inclusion of relative abundance identifies the core range of the species. The weekly relative abundance estimates show where a species occurs and its relative abundance for every week of the year. The estimates also show regions in which the species does not occur and locations where eBird does not have sufficient data.

When combined and analyzed appropriately, eBird data enable next generation species distribution models that provide full life cycle information about birds at relatively fine scales across broad spatial and temporal extents. Recently, estimates of the status and trends of more that 100 species of birds was released on the eBird website (https://ebird.org/science/status-and-trends). The eBird status and trends provide an unparalleled window into the full annual cycle of bird populations in North America. Maps, charts, and other products explore the range, abundance, habitat, and population trends for each species.

Together these metrics and species summaries can be used to contrast regions, seasons, and species, or to inform management decisions. For example, modeling how bird populations change throughout the year has uncovered complex and seasonally varying species–environment relationships (Zuckerberg et al. 2016), identified novel aspects of habitat associations that can affect bird populations during migration (La Sorte et al. 2017), and identified seasonal resources needed for supporting bird populations during critical stages of their life history (Johnston et al. 2015, Reynolds et al. 2017). Overall, eBird collects sufficient data to account for a large proportion of the variation in the observation process, which provides greater confidence in trends, maps, and other ecological metrics that are produced using eBird data.

Ensuring that citizen science projects effectively monitor biodiversity

Our focus in this article has been to provide a framework to improve the quality of citizen projects to best provide data that meet scientific objectives. Although proponents of citizen science monitoring projects may emphasize that the ability to gather large quantities of observational data can compensate for poor data quality (Munson et al. 2010), for many applications the quantity of data does not necessarily compensate for its quality. Our recommendations focus on improving the information collected regarding the observation process in order to account for the biases in the data caused by variation in the observation process. By doing this, the scope and quality of inferences that determine the distribution and abundance of species in space and time is enhanced. We argue that by using a semistructured approach it is possible to engage volunteers in citizen

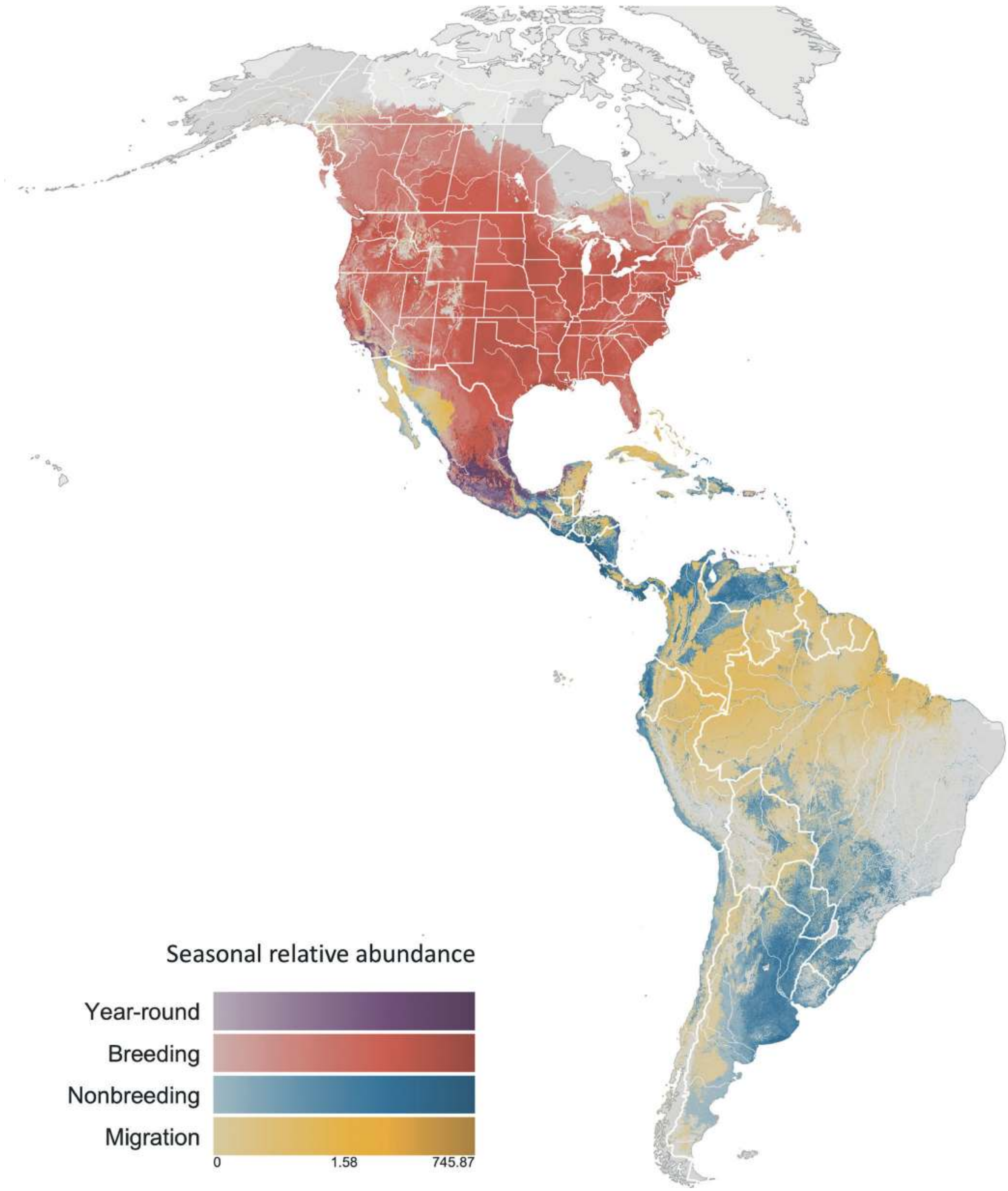

*Figure 2. Seasonal relative abundance of barn swallow (***Hirundo rustica***). This map shows the average relative abundance during each of the stationary seasons: breeding (June 11–July 23) and nonbreeding (December 18–February 11). The average relative abundance is also shown during the nonstationary migration seasons, and locations in which barn swallows occur year around. The stationary breeding and nonbreeding seasons are plotted on top of the other year-round and migration seasons, obscuring some aspects of the species' movements through the annual cycle. The areas denoted in palest grey currently have insufficient data with which to model relative abundance.*

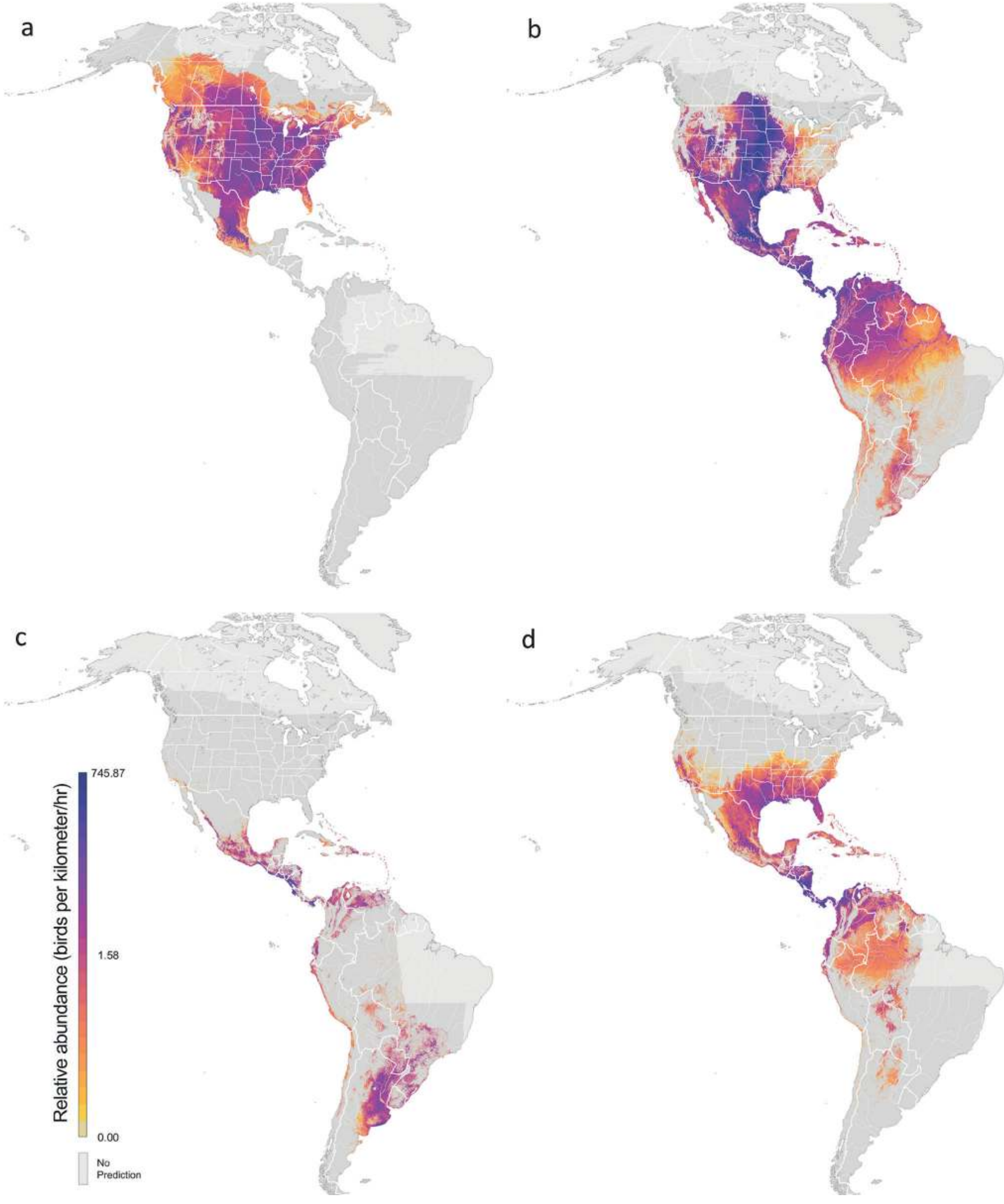

Figure 3. Barn swallow estimates of weekly relative abundance at 2.8 kilometers (km) × 2.8 km resolution representing the seasons: (a) breeding (June 18–24), (b) autumn migration (October 2–8), (c) nonbreeding (January 1–7), and (d) spring migration (March 26–April 1). The darker colors (pink and purple) indicate areas with higher abundance. Relative abundance was measured as the expected count of the species on a standardized 1-km survey conducted by a highly experienced participant.

science monitoring through broad participation, while still gathering sufficiently robust data for analysts to pursue these objectives. Although the additional data collection "costs" could still limit participation, recent advances in automated processes (e.g., GIS location and timing recording via smart phone apps) reduce these costs. For example, already more than 60% of all eBird data are being submitted via mobile apps, with this percentage growing. The eBird app automatically gathers most of the observation process data that eBird requires, with little additional effort required from the observer. Nevertheless, increasing rigor and standardization in data collection will need additional coordination and communication from project personnel. In summary, it is the responsibility of the project coordinators to carefully assess the project design and data collection necessary to ensure they are collecting semistructured rather than unstructured data to improve the scientific outcomes of the data collected.

Developments in the field of Artificial Intelligence, combined with increasingly efficient and powerful mobile technologies, will vastly improve the quality and quantity data of collected by citizen scientists. For example, projects such as the Cornell Lab of Ornithology's Merlin Bird ID App and iNaturalist already use powerful Deep Learning algorithms and computer vision techniques to identify images of thousands of organisms to species, helping observers get a species-level identification in the field (Van Horn et al. 2017). In addition, to encourage a more even distribution of surveys across a region or across habitat gradients, the eBird-based project called Avicaching used behavioral economics to incentivize site selection among the participants. Incentives were used to gather data from locations that were historically undersampled, using discrete choice models and machine learning to account for variable patterns of human behavior (Xue et al. 2016). The incentives resulted in significant shifts in participant effort to under sampled areas, demonstrating how incentives can be used to collect less-biased data in citizen science programs. However, as outlined above, any incentives should be carefully recorded as part of the observation process.

Equally important are continuing advances in analytical methodology for extracting signals from noisy citizen science data (Isaac et al. 2014). For example, recent methods have described a means to jointly analyze data collected in both structured and unstructured projects (Fithian et al. 2015, Giraud et al. 2016, Tenan et al. 2016). Other models leverage the strengths of both information-rich but relatively sparse structured data with semistructured data that has broad spatial coverage (Robinson et al. 2018).

We also note open-access data as another key component to consider in project development. Contributing data to data aggregators like the Global Biodiversity Information Facility (GBIF) or providing means to download data directly from a citizen science project's website, can significantly enhance the overall impact of collected data. For example, open accessibility of data to interested researchers can lead to analyses that would not have been possible by the project team alone, and feedback from external users of the data can feed into refinements in the data-collection process.

In conclusion, the enormous growth of digital networks, rapid advances in the field of artificial intelligence, the appearance of powerful computing devices that can fit in a pocket, new statistical analysis methods, and open data initiatives will improve the quantity and quality of citizen science project data. The potential benefits of technological advances, however, will be realized only if coupled with standardized reporting of basic ecological information, and related information on the observation process in order to best use these data in downstream analyses. We envision a global network of motivated observers rapidly collecting species lists and observation-process information usable for near real-time trend assessment of species and community health.

Acknowledgments

This work has its origin in a Future Earth supported meeting 6–10 March 2016 at Monte Verita, Switzerland entitled "Global Biodiversity Assessment and Monitoring—Science, Data and Infrastructure Needs for IPBES and Beyond." A follow-up meeting was supported by the Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118). Additional funding for this work came from the Wolf Creek Foundation, The Leon Levy Foundation, The Packard Foundation, and the National Science Foundation (grants no. ITR-0427914, no. DBI-0542868, no. IIS-0612031, and no. ABI-1356308 to S. Kelling; no. IIS-0920869 to D. Fink; no. CISE-0832782, no. ABI-1262600, and no. ABI-1458527 to R. Guralnick). We thank Matt Strimas-Mackey and Will Morris for assistance with the figures. We thank Brian Sullivan and three anonymous reviewers whose comments improved the manuscript.

References cited

Amano T, Lamming J, Sutherland W. 2016. Spatial gaps in global biodiversity information and the role of citizen science. BioScience 66: 393–400.

Balmford A, Gaston KJ. 1999. Why biodiversity surveys are good value. Nature 398: 204–205.

Bland L, Collen B, Orme C, Bielby J. 2015. Predicting the conservation status of data-deficient species. Conservation Biology 29: 250–259.

Bonney R, Cooper C, Dickinson J, Kelling S, Phillips T, Rosenberg K, Shirk J. 2009. Citizen science: A developing tool for expanding science knowledge and scientific literacy. BioScience 59: 977–984.

Chandler M, See L, Copas K, Bonde A, López B, Danielsen F, Legind J, Masinde S, Miller-Rushing A, Newman G. 2016. Contribution of citizen science towards international biodiversity monitoring. Biological Conservation 213: 280–294.

Danielsen F, Pirhofer-Walzl K, Adrian T, Kapijimpanga D, Burgess N, Jensen P, Bonney R, Funder M, Landa A, Levermann N. 2014. Linking public participation in scientific research to the indicators and needs of international environmental agreements. Conservation Letters 7: 12–24.

Fink D, Damoulas T, Bruns N, La Sorte F, Hochachka W, Gomes C, Kelling S. 2014. Crowdsourcing meets ecology: Hemisphere-wide spatiotemporal species distribution models. AI Magazine 35: 19–30.

Fink D, Hochachka W, Winkler D, Shaby B, Hooker G, Zuckerberg B, Munson M, Sheldon D, Riedewald M, Kelling S. 2010. Spatiotemporal exploratory models for large-scale survey data. Ecological Applications 20: 2131–2147.

Fithian W, Elith J, Hastie T, Keith DA. 2015. Bias correction in species distribution models: Pooling survey and collection data for multiple species. Methods In Ecology and Evolution 6: 424–438.

Fitzpatrick M, Preisser E, Ellison A, Elkinton J. 2009. Observer bias and the detection of low-density populations. Ecological Applications 19: 1673–1679.

Fox R, Warren M, Brereton T, Roy D, Robinson A. 2011. A new red list of British butterflies. Insect Conservation and Diversity 4: 13.

Franklin J. 2009. Mapping species distributions: Spatial inference and prediction. Cambridge University Press.

Geldmann J, Heilmann-Clausen J, Holm T, Levinsky I, Markussen B, Olsen K, Rahbek C, Tøttrup A. 2016. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. Diversity and Distributions: A Journal of Biological Invasions and Biodiversity 22: 1139–1149.

Giraud C, Calenge C, Coron C, Julliard R. 2016. Capitalizing on opportunistic data for monitoring relative abundances of species. Biometrics 72: 649–658.

Guillera-Arroita G, Lahoz-Monfort J, Elith J, Gordon A, Kujala H, Lentini P, McCarthy M, Tingley R, Wintle B. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography 24: 276–292.

Isaac N, Strien AJ, August TA, Zeeuw MP, Roy D. 2014. Statistics for citizen science: Extracting signals of change from noisy ecological data. Methods In Ecology and Evolution 5: 1052–1060.

Johnston A, Fink D, Hochachka W, Keling S. 2018. Estimates of observer expertise improve ecological inference from citizen science data. Methods In Ecology and Evolution 9: 88–97.

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

Kelling S, et al. 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? PLOS ONE 10 (art. e0139600).

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

Kery M, Schmid H. 2004. Monitoring programs need to take into account imperfect species detectability. Basic and Applied Ecology 5: 65–73.

La Sorte F, Fink D, Buler J, Farnsworth A, Cabrera-Cruz S. 2017. Seasonal associations with urban light pollution for nocturnally migrating bird populations. Global Change Biology 23: 4609–4619.

MacKenzie D, Nichols J, Royle A, Pollock K, Bailey L, HInes J. 2006. Occupancy Estimation and Modeling- Inferring Patterns and Dynamics of Species Occurrence. Elsevier.

McKinley D, et al. 2015. Investing in citizen science can improve natural resource management and environmental protection. Issues in Ecology 19: 27.

Munson MA, Caruana R, Fink DF, Hochachka WM, Iliff MI, Rosenberg KV, Sheldon DR, Sullivan BL, Wood CL, Kelling S. 2010. A method for measuring the relative information content of data from different monitoring protocols. Methods In Ecology and Evolution 1: 263–273.

Pimm SL, Alibhai S, Bergl R, Dehgan A, Giri C, Jewell Z, Joppa L, Kays R, Loarie S. 2015. Emerging Technologies to Conserve Biodiversity. Trends in Ecology and Evolution 30: 685–696.

Pocock M, Tweddle J, Savage J, Robinson L, Roy H. 2017. The diversity and evolution of ecological and environmental citizen science. PLOS ONE 12 (art. e0172579).

Reynolds M, et al. 2017. Dynamic conservation for migratory species. Science Advances 3: e1700707.

Robinson OJ, Ruiz-Gutierrez V, Fink D, Meese RJ, Holyoak M, Cooch E. 2018. Using citizen science data in integrated population models to inform conservation. Biological Conservation 227: 361–368.

Roy H, Pocock M, Preston C, Roy D, Savage J, Tweddle J, Robinson L. 2012. Understanding Citizen Science and Environmental Monitoring: Final Report on Behalf of UK Environmental Observation Framework. NERC Centre for Ecology and Hydrology and Natural History Museum http://nora.nerc.ac.uk/20679/1/N020679CR.pdf.

Schneider S, Taylor GW, Kremer SC. 2018. Deep learning object detection methods for ecological camera trap data. arXiv:1803.10842, vol. 1: 1–8.

Sullivan B, et al. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation 169: 31–40.

Tenan S, Pedrini P, Bragalanti N, Groff C, Sutherland C. 2016. Investigating the potential of opportunistic sighting data to inform wildlife conservation strategies. bioRxiv:075945.

Theobald E, et al. 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation 181: 236–244.

Van Horn G, Aodha OM, Song Y, Cui Y, Sun C, Shepard A, Adam H, Perona P, Belongie S. 2017. The iNaturalist species classification and detection data set. arXiv:1707.06642v2, (v2).

Xue Y, Davies I, Fink D, Wood C, Gomes C. 2016. Avicaching: A two stage game for bias reduction in citizen science. Pages 776–785 in Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems: International Foundation for Autonomous Agents and Multiagent Systems. AAMAS.

Yoccoz N, Nichols J, Boulinier T. 2001. Monitoring of biological diversity in space and time. Trends in Ecology and Evolution 16: 446–453.

Zhang S, Zhao Z, Xu Z, Bellisario K, Pijanowski BC. 2018. Automatic Bird Vocalization Identification Based on Fusion of Spectral Pattern and Texture Features. Paper presented at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB.

Zuckerberg B, Fink D, La Sorte FA, Hochachka WM, Kelling S. 2016. Novel seasonal land cover associations for eastern North American forest birds identified through dynamic species distribution modelling. Diversity and Distributions 22: 717–730.

Steve Kelling (stk2@cornell.edu), Daniel Fink, Viviana Ruiz-Gutierrez, Rick Bonney, and Wesley M. Hochachka are affiliated with the Cornell Lab of Ornithology, at Cornell University, in Ithaca New York. Alison Johnston is affiliated with the Cornell Lab of Ornithology and with the Department of Zoology at the University of Cambridge, in Cambridge, England. Aletta Bonn is affiliated with the Helmholtz Centre for Environmental Research—UFZ, Department of Ecosystem Services, in Leipzig, Germany; with the Institute of Biodiversity at Friedrich Schiller University Jena, in Jena, Germany; and with the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, in Leipzig. Miguel Fernandez is affiliated with NatureServe, in Arlington, Virginia; with the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig; and with the Environmental Science and Policy Department at George Mason University, in Fairfax, Virginia. Romain Julliard is affiliated with the Center for Ecology and Conservation Sciences (UMR CESCO), at the Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, in Paris, France. Roland Kraemer is affiliated with the Helmholtz Centre for Environmental Research—UFZ; with the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig; and with Humboldt-Universität zu Berlin's Institute of Geography, in Berlin, Germany. Robert Guralnick is affiliated with the Department of Natural History at the Florida Museum of Natural History and with the University of Florida's Biodiversity and Genetic Institutes, at the University of Florida, in Gainsville.