

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/303876162>

# The science of citizen science: Exploring barriers to use as a primary research tool

Article in *Biological Conservation* · June 2016

DOI: 10.1016/j.biocon.2016.05.014

CITATIONS

9

READS

350

7 authors, including:



[Hillary K. Burgess](#)

University of Washington Seattle

8 PUBLICATIONS 99 CITATIONS

[SEE PROFILE](#)



[Lauren DeBey](#)

Tableau Software

12 PUBLICATIONS 116 CITATIONS

[SEE PROFILE](#)



[Halley E. Froehlich](#)

University of California, Santa Barbara

25 PUBLICATIONS 199 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Public perceptions of aquaculture: evaluating spatiotemporal patterns of sentiment around the world [View project](#)

All content following this page was uploaded by [Lauren DeBey](#) on 10 June 2016.

The user has requested enhancement of the downloaded file.



Contents lists available at ScienceDirect

## Biological Conservation

journal homepage: [www.elsevier.com/locate/bioc](http://www.elsevier.com/locate/bioc)

## The science of citizen science: Exploring barriers to use as a primary research tool

H.K. Burgess<sup>a,\*</sup>, L.B. DeBey<sup>b,1</sup>, H.E. Froehlich<sup>a,2</sup>, N. Schmidt<sup>c</sup>, E.J. Theobald<sup>b</sup>, A.K. Ettinger<sup>b,3</sup>, J. HilleRisLambers<sup>b</sup>, J. Tewksbury<sup>b</sup>, J.K. Parrish<sup>a</sup><sup>a</sup> School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington 98105, USA<sup>b</sup> Biology Department, University of Washington, Seattle, Washington 98195, USA<sup>c</sup> School of Environmental and Forest Sciences, University of Washington, Seattle, Washington 98195, USA

## ARTICLE INFO

## Article history:

Received 16 October 2015

Received in revised form 16 April 2016

Accepted 23 May 2016

Available online xxxx

## Keywords:

Biodiversity

Citizen science

Data quality

Outreach

Public participation in science

Research

## ABSTRACT

Biodiversity citizen science projects are growing in number, size, and scope, and are gaining recognition as valuable data sources that build public engagement. Yet publication rates indicate that citizen science is still infrequently used as a primary tool for conservation research and the causes of this apparent disconnect have not been quantitatively evaluated. To uncover the barriers to the use of citizen science as a research tool, we surveyed professional biodiversity scientists ( $n = 423$ ) and citizen science project managers ( $n = 125$ ). We conducted three analyses using non-parametric recursive modeling (random forest), using questions that addressed: scientists' perceptions and preferences regarding citizen science, scientists' requirements for their own data, and the actual practices of citizen science projects. For all three analyses we identified the most important factors that influence the probability of publication using citizen science data. Four general barriers emerged: a narrow awareness among scientists of citizen science projects that match their needs; the fact that not all biodiversity science is well-suited for citizen science; inconsistency in data quality across citizen science projects; and bias among scientists for certain data sources (institutions and ages/education levels of data collectors). Notably, we find limited evidence to suggest a relationship between citizen science projects that satisfy scientists' biases and data quality or probability of publication. These results illuminate the need for greater visibility of citizen science practices with respect to the requirements of biodiversity science and show that addressing bias among scientists could improve application of citizen science in conservation.

© 2016 Published by Elsevier Ltd.

## 1. Introduction

Public participation in scientific research, or citizen science, is a growing practice that could be a powerful addition to the conservation toolbox (Cooper et al., 2007; Danielsen et al., 2010; Cosquer et al., 2012; Theobald et al., 2015). From documenting climate change impacts (e.g., Boyle and Sigel, 2015) to informing land management (e.g., Martin, 2015), applications of citizen science data are broad across subjects and at scales relevant to today's conservation issues (Theobald et al., 2015). Many authors have argued that both ecology and conservation would benefit from greater use of citizen science due to its ability to provide data at the broad spatiotemporal scales and fine grain resolution needed to address global-scale conservation questions (Jiguet et al., 2005; Couvet et al., 2008; Schmeller et al., 2009; Devictor et al., 2010;

Magurran et al., 2010; Loss et al., 2015). This argument may be particularly strong for biodiversity science, because abundance and/or density of taxa is the focus of a large number of citizen science projects (Theobald et al., 2015).

Despite these arguments, citizen science has yet to be fully embraced by either the ecological or conservation communities (Silvertown, 2009; Riesch and Potter, 2013; Tulloch et al., 2013; Cooper et al., 2014; Bonney et al., 2014). Citizen science projects within these fields generally report only modest peer-reviewed publication rates (Theobald et al., 2015), and they have rarely generated well-known, highly cited data (Jiguet et al., 2005; Zuckerberg et al., 2009; Silvertown et al., 2011, but see Devictor et al., 2010 and Edgar et al., 2014 for exceptions).

One important obstacle to the scientific use of citizen science data may be the perceptions of scientists. For example, efforts to incorporate citizen-generated data into conservation are sometimes met with concerns regarding rigor of data collection and, ultimately, data quality. Specific critiques include lack of attention to study design (Newman et al., 2003; Krasny and Bonney, 2005), inconsistent or suboptimal training (Conrad and Hilchey, 2011), absent or problematic standardization and verification methods (Cohn, 2008; Dickinson et al., 2010; Bonter and

\* Corresponding author.

E-mail address: [hkb10@uw.edu](mailto:hkb10@uw.edu) (H.K. Burgess).<sup>1</sup> Burgess and DeBey contributed equally to the work.<sup>2</sup> National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, California 93106, USA.<sup>3</sup> Arnold Arboretum of Harvard University, Boston, Massachusetts 02131, USA.

Cooper, 2012), and observer or sampling biases (Galloway et al., 2006; Delaney et al., 2008; Dickinson et al., 2010). These criticisms are often countered by specific examples of citizen science projects producing data comparable in quality to that collected by professionals (e.g. Elbroch et al., 2011; Forrester et al., 2015; Lewandowski and Specht, 2015; Lin et al., 2015). This debate highlights the potential influence of perceptions in shaping the use of citizen science (Conrad and Hilchey, 2011; Henderson, 2012). For example, Riesch and Potter (2013) found that scientists' perceptions that citizen science data would not be well received by peers in the scientific community contributed to lack of use.

In this paper, we quantitatively and systematically examine factors related to scientific use of citizen science data, assessed as publication in the peer-reviewed literature. Our goals are to evaluate the influence of perception on use, and elucidate barriers to the use of citizen science data in conservation and ecology. We use multivariate analysis to explore survey results from 423 biodiversity scientists and 125 managers of biodiversity citizen science projects, with respect to three categories: scientists' perceptions of citizen science, scientists' requirements for their own data, and actualities of biodiversity citizen science projects. Our analyses focus on the influence of data source, quality, and visibility. Our results point to specific opportunities for expanded integration of citizen science into the fields of ecology and conservation science.

## 2. Methods

### 2.1. Surveys

#### 2.1.1. Scientists

To determine factors that influence whether biodiversity scientists use citizen science data, we created an online survey targeting biodiversity scientists (IRB approval 43,438) to assess: (1) the extent to which citizen science data are presently used in their research, (2) perceptions of citizen science and its resultant data, and (3) requirements for methods and data (e.g. standardization, procedures, measure of error) when conducting their own biodiversity research (see Appendix A Table A1 for full text of survey questions and response options). We defined biodiversity citizen science in the survey introduction as “programs collecting taxon-specific information at known locations and date/times.” The survey contained 25 multi-part questions with binomial (yes/no), multiple choice (inclusive and exclusive), and Likert scale (e.g., strongly agree to strongly disagree) answers, as well as free responses to select prompts. Survey respondents were free to skip any question, resulting in a variable sample size for each question.

We identified publishing biodiversity scientists using a Web of Science search, restricted to natural science, of corresponding authors of papers containing the word “biodiversity” in the title, abstract, or keywords; this yielded a pool of 3148 scientists as potential survey respondents. We contacted the corresponding author of each publication by email; 423 scientists completed the survey (13.4% response rate; see Table A1 for respondent demographics). It is possible that scientists who responded were more interested in, and/or aware of, citizen science compared to those who did not respond, and we consider this when interpreting our results.

#### 2.1.2. Citizen science

In order to compare scientists to citizen science programs, we surveyed citizen science project managers (IRB approval 43,438) regarding (1) project goals and details of project administration, (2) data collection protocols, and (3) participant demographics (see Table A2 for full text of survey questions and response options). We defined biodiversity citizen science in the survey introduction as “programs collecting taxon-specific information at known locations and date/times.” The survey contained 32 multi-part questions with similar types of questions as the scientist survey (i.e., binomial, multiple choice, Likert scale, and free response); where applicable, we asked identical questions of both

scientists and citizen science projects. Again, respondents were free to skip any question, resulting in a variable sample size for each question.

We identified potential respondents from a database of biodiversity citizen science projects that aggregates projects from seven publicly available databases (see Theobald et al., 2015 for database details). Of these 388 projects, 329 were extant and had contact information that enabled our communication with project managers via email at the time of survey administration. We received a total of 125 responses (38% response rate). Respondent projects were predominantly housed in North America (66.4%) followed by 9%, Europe, 2.5% Asia, 2.5% multiple, 1.6% Oceania with 18% Unknown.

### 2.2. Analysis

#### 2.2.1. Factors influencing publication of citizen science data

While peer-reviewed scientific publication is not the only outlet for citizen science research outcomes, patterns of publication of citizen science data in the peer-reviewed literature represent a measure of the current extent to which citizen science reaches professional scientists. We used a non-parametric modeling approach known as Random Forest analysis to derive explanatory patterns between the probability of publication using citizen science data and survey responses, for both professional scientists and citizen science project managers.

Random Forest analysis (RF) is a statistical technique that uses recursive and “out-of-bag” bootstrap sampling (i.e., predicting data not in the bootstrap sample) to construct binary partitions of predictor variables, fitting regression trees ( $n = 1000$ ) to the dataset, and ultimately combining the predictions from all trees (Breiman, 2001). Predictors are ranked by mean squared error (Breiman, 2001; Cutler et al., 2007); the order reflects the influence of each predictor on the response variable. We conducted three separate RF analyses: two using explanatory variables from scientists' survey responses (Table A1), and one using explanatory variables from citizen science project manager responses (Table A2). See Appendix A Supplemental Methods for additional details.

We distinguished two categories of variables a priori to explore via RF in association with scientists' potential engagement with citizen science projects (perceptions/preferences and requirements), and one set of variables for the citizen science managers. “Perceptions/preferences” captures opinions regarding the purpose of citizen science, the quality of citizen science data, and the degree of trust in various data sources. “Requirements” consists of awareness of citizen science projects that matched their research area and factors that are required to successfully conduct their particular research (e.g., specific methods, protocols, or data attributes). We assumed that for citizen science data to be used for research purposes by our respondent scientists these latter factors must be satisfied. We performed two separate RF analyses on these mutually exclusive sets of variables: scientists' perceptions/preferences (398 respondents, 29 predictor variables: 27 numeric, 3 binary), and scientists' requirements (388 respondents, 27 predictor variables: 23 binary, 2 factors, one 1 numeric), each with the binary response variables of either “have published using citizen science generated data” or “have not” (1 or 0, respectively; see Table A1, Question 21). We conducted a third RF analysis on citizen science survey responses (118 respondents, 49 predictor variables: 27 numeric, 22 binary) to predict the probability of whether a project reported having one or more peer-reviewed articles using project-generated data (1; see Table A2, Question 18), or no publications associated with that project's data (0).

The full RF models incorporated all possible respective variables, which we reduced to the best-fit model based on a subset of those predictors. The full RF models provided an initial ranked order of all predictor variables associated with publication for each dataset. In a stepwise elimination, starting from least to most influential, predictor variables were removed, without replacement, from the model and variance explained was determined at each step. We then compared all models, selecting the best-fit as the model that explained the greatest amount

of variance and contained the fewest number of parameters (principle of parsimony). Once the best-fit models were identified, we evaluated the individual predictors in the reduced models using partial dependence plots (PDP; Cutler et al., 2007; see Appendix A Supplemental Methods for additional details). RF results informed further use of descriptive statistics and data visualization. All analyses were conducted using the *randomForest* package in R (Cutler et al., 2007; R Core Team, 2013).

### 2.2.2. Comparing characteristics of citizen science to scientists' perceptions and preferences

Additional univariate analysis and descriptive approaches were used to make direct comparisons between citizen science and the preferences of scientists where possible (i.e., institution, data collector demographics, and data collector training), or to examine multi-part questions when one response option was part of the final model and the distribution of other responses provided context for interpretation. Specifically, we report the results of single tail two-sample *t*-tests assuming unequal variances (with associated *t* statistic, and degrees of freedom given in parentheses), and all associated means and standard deviations.

Finally, we assessed whether citizen science is currently meeting the needs of professional scientists and how scientists' perceptions reduce the use of citizen science data by scientists. To do this, we rank-ordered citizen science projects based on the degree to which they matched scientists' preferences and requirements, then successively eliminated those that met fewer and fewer preferences/requirements. We conducted two separate cumulative elimination rounds of citizen science projects. The first round included scientists' data requirements only; the second accommodated scientists' preferences first and data requirements second, to compare the extent that each category (preference

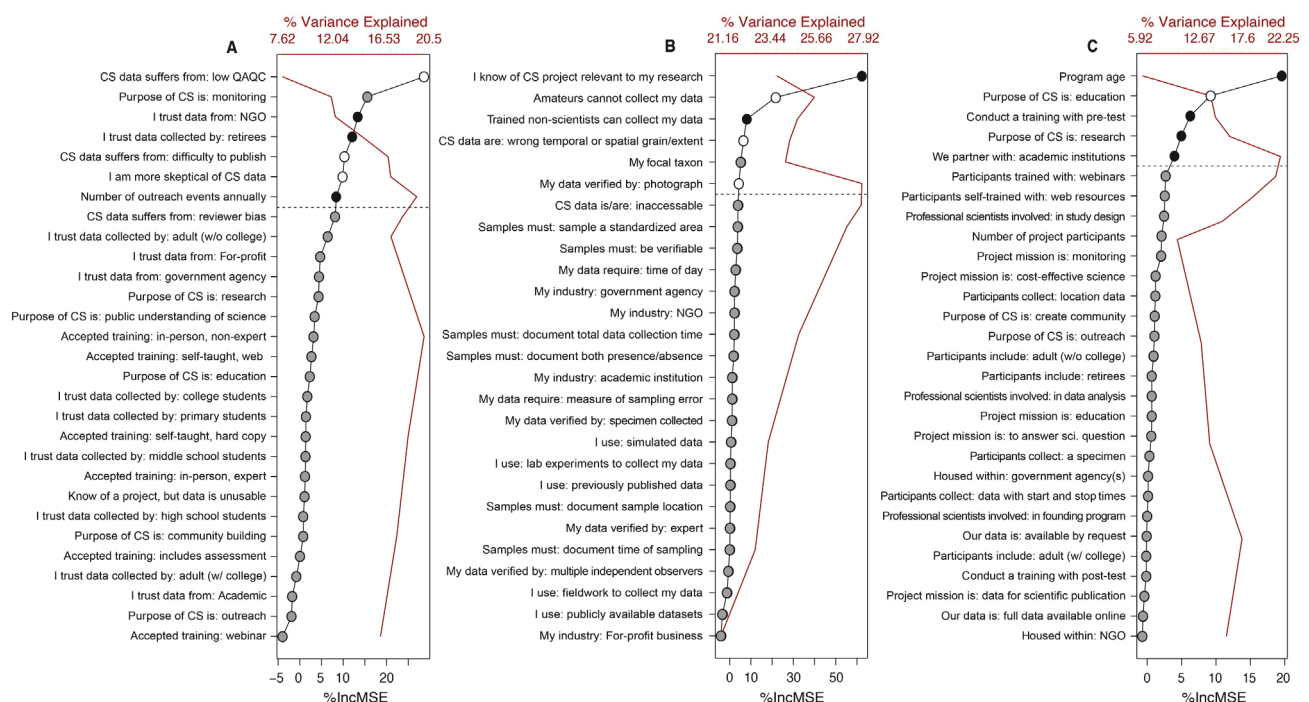
or requirement) limited project data that scientists would be willing to use. Ranking was measured by the proportion of “strongest” scientist opinions for Likert scale questions (perceptions/preferences), and the percent agreement with qualities required for data usability for which we collected data from citizen science projects (requirements). This allowed us to compare the match between projects and scientists' perceptions/preferences and requirements.

## 3. Results

### 3.1. Factors that influence publication of citizen science data

#### 3.1.1. Perceptions/preferences

We identified seven “perceptions/preferences” variables that best explain publication probability of citizen science data (16.53% and 19.55% variance explained, full and reduced models, respectively; Fig. 1A, Appendix B Tables B1, B2). The probability of publication was strongly affected by scientists' beliefs about the quality assurance/quality control present among citizen science data (i.e., scientists who agreed with the statement “citizen science data suffers from low QA/QC” had a lower probability of publication using citizen science data, Figs. 1A, B1, Table B1). Scientists who were more skeptical of citizen science data and believed that it was more difficult to publish these data were also less likely to have done so (Figs. 1A, B1). Additionally, scientists who indicated they would use data originating from NGOs (any data, not just citizen science-generated), would use data collected by retirees, and/or who engaged in more frequent outreach activities, all had a higher probability of publication using citizen science data than others (Figs. 1A, B1). Lastly, both scientists who strongly agreed, and those who strongly disagreed, that a purpose of citizen science is monitoring were



**Fig. 1.** Random Forest model results. Shown are all predictor variables (survey question responses) in rank order of influence on the publication of citizen science data according to (A) scientists' preferences for and perceptions of citizen science data, (B) scientists' requirements for their own data, and (C) responses of citizen science project managers. Questions above the dotted line are the most critical variables and maximize percent variance explained (% Variance Explained, full model) while minimizing the number of explanatory variables (see Table A2.2). Abbreviations: %IncMSE, percent Increase in Mean Standard Error. Point shading indicates direction of trend: black dots indicate a positive trend between the question and probability of publication using citizen science (CS) data; white dots indicate a negative trend; gray dots suggest either a non-directional trend (e.g., in A, “Purpose of CS is: monitoring”), a pattern not applicable to directionality (e.g., in B, “My focal taxon”), or non-significant questions and directional trends (below the dotted lines of significance). See text for explanation of questions and datasets; see Appendix 1 for survey questions, response options, and question type, and Appendix 2 for response distributions and partial dependence plots.



more likely to publish using citizen science-generated data than those with neutral opinions (Figs. 1A, B1).

### 3.1.2. Requirements

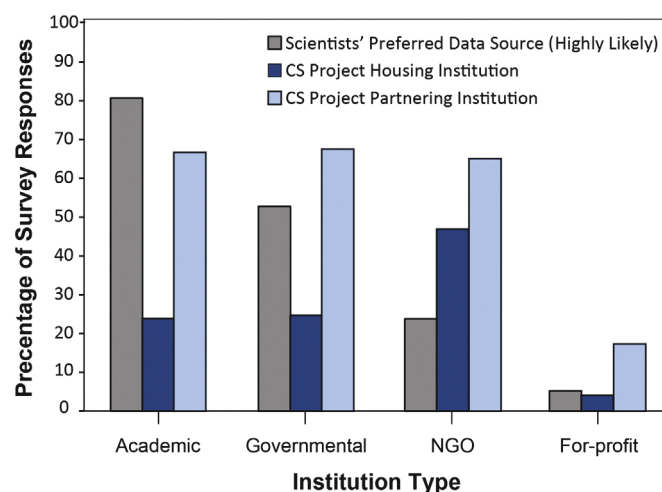
We identified six “requirement” variables that best predicted the probability of a scientist having published using citizen science-generated data (22.04% and 27.82% variance explained, full and reduced models, respectively; Fig. 1B, Tables B1, B2). Knowledge of a relevant program was the most influential factor (Figs. 1B, B2, Table B1); specifically, scientists who knew of a program relevant to their research had a higher probability of publication using citizen science data. Scientists who indicated that adequately trained citizens could collect their data also had a higher probability of publication using citizen science data compared to those who agreed with the statement “amateurs could not collect their data” (Figs. 1B, B2). Additionally, scientists who reported that citizen science sampling happens at the wrong scale, grain or extent for their research, and those who used photographs to verify their own data (Figs. 1B, B2) had a lower probability of publication. Scientists’ particular focal taxa was the last important predictor, with the highest probability of publication associated with scientists studying birds, reptiles, and algae, and the lowest probability for aquatic invertebrates, fungi, and mammals (Fig. B2).

### 3.1.3. Citizen Science Projects

We identified five citizen science project characteristics that best predicted probability of publication (9.55% and 21.76% variance explained, full and reduced models, respectively; Fig. 1C, Tables B1, B2). The strongest predictor was a positive relationship between publication and the age of the citizen science project (Figs. 1C, B3). Additionally, projects that conducted a pre-test during participant training, and those who partnered with academic institutions had higher probabilities of having peer-reviewed publications (Figs. 1C, B3). Lastly, projects with managers who considered the primary purpose of citizen science to be education had a lower probability of publication, whereas projects with managers who strongly agreed that the purpose of citizen science is research had a higher probability of having published (Figs. 1C, B3).

### 3.2. Comparing characteristics of citizen science to the perceptions and preferences of scientists

Survey results suggested biodiversity scientists and project managers had similar but slightly differing perspectives on the purpose of citizen science (Fig. 2). Citizen science project managers were more positive across all proposed purposes of citizen science compared to scientists (global project manager mean =  $1.64 \pm 0.60$ , global scientist mean =  $1.41 \pm 0.74$ ;  $t$ -test (1405) = 7.55,  $p < 0.001$ ). Perspectives on research as a purpose of citizen science were relatively low among both response groups, and scientists’ average responses ( $0.71 \pm 1.01$ )

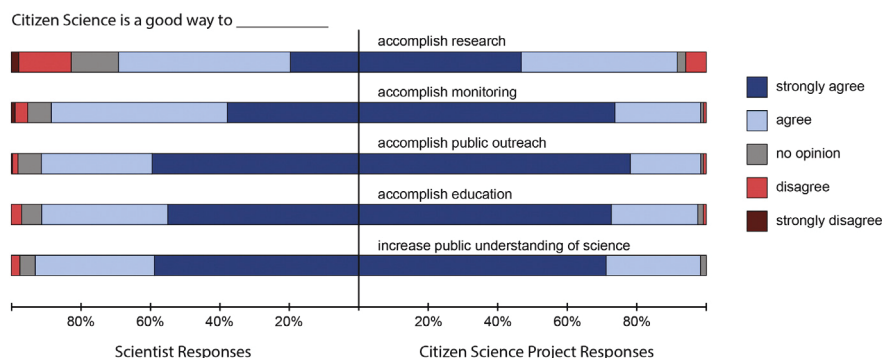


**Fig. 3.** Preferred data sources compared to citizen science. Comparison of respondent scientists' preferred data source (gray,  $n = 396$ ), citizen science project home institution (dark blue,  $n = 127$ ), and citizen science partnering institution (light blue,  $n = 122$ ; see Table B1). Scientists' preferred data source was calculated using only the “Highly Likely” to use responses to the Likert scale Question 12; citizen science housing and partnering institution used responses to Questions 30 and 31, respectively. Responses for citizen science partnering institution were not exclusive, and therefore the sum is greater than 100%. Abbreviations: CS, citizen science; NGO, non-governmental organization.

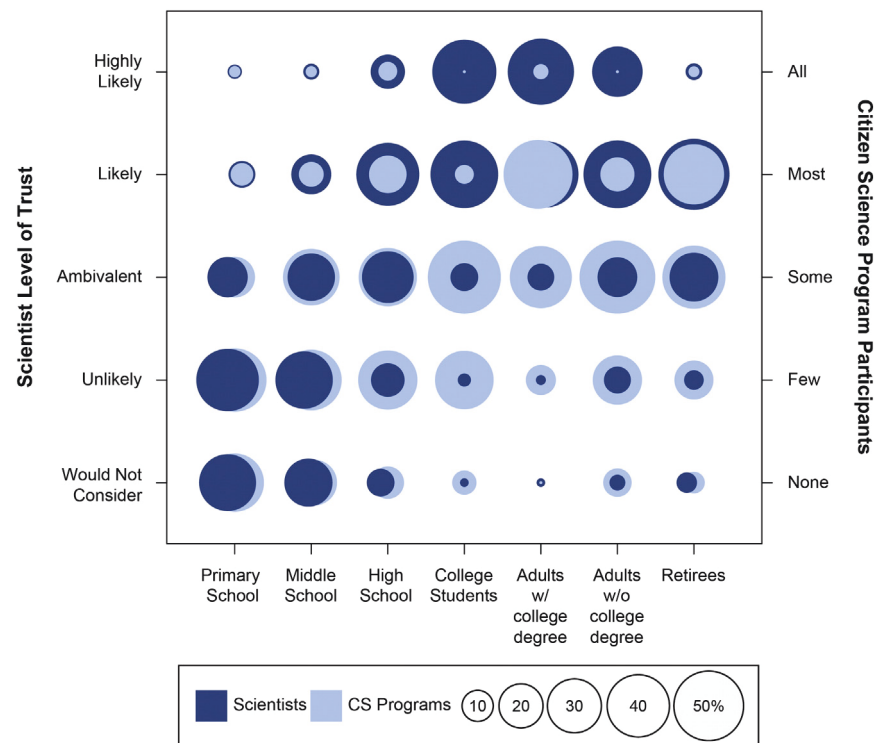
were significantly weaker than the responses of project managers ( $1.34 \pm 0.78$ ;  $t$ -test (268) = 7.20,  $p < 0.001$ ).

Housing institution of citizen science projects was an important factor for publication. Specifically, scientists had a higher probability of publication using citizen science data if they were amenable to using data from NGOs (Fig. 1A). Scientist respondents had a preference for data from academic institutions, followed by governmental agencies, NGOs and for-profit institutions, in descending order (Fig. 3, Appendix C Table C1); and scientists significantly preferred data from academic institutions to NGOs (academic mean =  $1.75 \pm 0.57$ , NGO mean =  $0.70 \pm 1.00$ ;  $t$ -test (609) = 17.90,  $p < 0.001$ ). Citizen science projects have a higher probability of publishing their data if they partner with an academic institution (Fig. 1C, B3), but we find citizen science projects are significantly more likely to be affiliated with (i.e., housed within, or partnered with) NGOs than academic institutions (academic mean =  $0.91 \pm 0.70$ , NGO mean =  $1.27 \pm 0.79$ ;  $t$ -test (256) = 3.97,  $p < 0.001$ ).

We found that publication of citizen science data by scientists was related to the demographics of data collectors. Specifically, scientists had a higher probability of publication using citizen science data if they were likely to use data collected by retirees (Fig. 1A, B1). However overall, scientists were significantly more likely to use data collected by



**Fig. 2.** Perspectives on the purpose of citizen science. Survey responses among scientists and citizen science projects to a multi-part question regarding the purposes of citizen science. In each survey, the question (number 14 for scientist survey; 5 for citizen science project managers) required responses to each statement (e.g., “Citizen Science is a good way to: accomplish research”). Responses are shown as proportions of total responses, with Likert scale response options coded from warm colors (strongly disagree and disagree) to cool colors (agree and strongly agree). See Appendix 1 for survey questions and response options; see Table A2.1 for response distributions and sample sizes.



**Fig. 4.** Preferred data collectors compared to citizen science. Comparison of citizen science participant demographics and scientists' preference for certain groups of data collectors (age and education level). Responses are for scientist survey Question 17 and citizen science project survey Question 11. Percent responses are indicated by circle diameter, according to the legend at the bottom. To facilitate readability, smaller circles are placed in front of corresponding larger circles, and dark and light blue circles are slightly offset one another if size is approximately identical. See Appendix A for survey questions and response options and Table B1 for results.

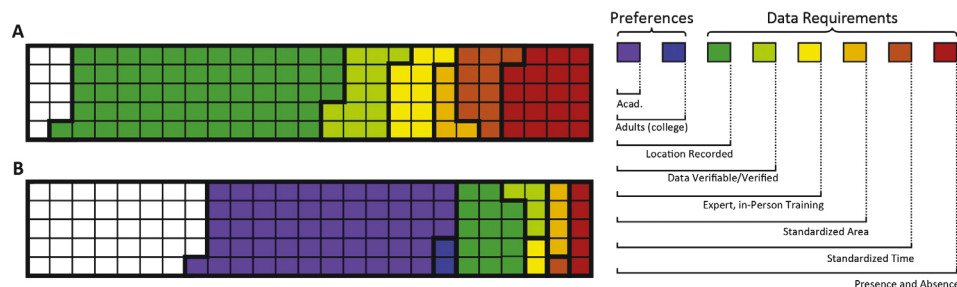
college students or adults with college degrees than from retirees (means =  $1.32 \pm 0.74$ , and  $0.98 \pm 0.93$ , respectively;  $t$ -test (634) = 6.12,  $p = <0.001$ ; Fig. 4, Table C1). They were similarly more likely to use data collected by college students or adults with college degrees than younger individuals (combined means =  $1.32 \pm 0.74$ , and  $-0.39 \pm 1.20$ , respectively;  $t$ -test (1893) = 38.45,  $p = <0.001$ . Citizen science projects that collected demographic data on their participants have some overlap with these preferences (Fig. 4, Table B2).

### 3.3. Is there a match between what scientists want and the data citizen science produces?

Because agreement with the prompt “citizen science data suffers from low QA/QC” was the most important predictor of publication with respect to scientists' perceptions (Fig. 1A), we explored data quality within citizen science by prioritizing methods that scientists required for data to be useful to them. In ranked order of preference

among scientist respondents they were: documentation of sample location, verifiability, sampling area recorded, in-person training by an expert, time spent or start and stop time recorded, and both presence and absence of focal species recorded (Table C1). Individually, each of these criteria were met by 49.6%–92.8% of projects (Table C2), but only 15.2% of citizen science projects surveyed met all criteria evaluated (Fig. 5A, Table C3). When scientists' preferences were prioritized above measures of data quality, even fewer (4.0%) citizen science projects satisfied scientists' criteria (Fig. 5B, Table C3). In particular, restricting data collection to adults with college degrees or college students severely limited the number of available citizen science projects (Fig. 5B).

Application of QA/QC measures were mixed among citizen science projects. With respect to quality assurance, scientists preferred in-person training by an expert above all other training methods (62.8% of scientists indicated they would use citizen science data if collectors were trained in-person by an expert); and 64.5% of project managers reported use of this method (Table C2). Pre-tests, a method of quality



**Fig. 5.** Scientists' criteria applied to citizen science projects. Cumulative elimination analyses via rank order of citizen science project characteristics. Each cell represents a single citizen science project ( $n = 125$ ). At the extremes of the rank order, white cells represent projects that fulfill none of the criteria outlined in the ranked legend at the right, and red cells fulfill all criteria. The top panel (A) shows citizen science projects that met all of the most important criteria scientists indicated they would need for citizen science data to be useful to them; the bottom panel (B) first satisfies scientists' preferences for data source (projects housed in, or partnering with Academic Institutions) and collectors (projects with mostly Adults with a college degree OR current college students as most or all of participants) before any requirements for data. Abbreviations: Acad., academic institutions. See text for methods, Appendix A for complete list of survey questions and response options, and Appendix C for ranking criteria and results.

assurance that was an important RF predictor for publications, were employed by only 15.6% of projects (Table B1). Post-tests, which provide a measure of confidence in data collectors' abilities and in the efficacy of the training program, were employed by 30.9% of projects (Table C2). With respect to quality control, 49.6% of projects reported requiring one or more methods allowing subsequent data verification (e.g., a photograph, collected specimen, an expert present, and/or other verifiable evidence collected).

#### 4. Discussion

We identified four emergent themes that characterize the use of citizen science in peer-reviewed publications: 1. Source preferences—biodiversity scientists place a higher value on data from other scientists, and especially academics, relative to that collected by non-scientists, amateurs and citizen scientists. 2. Data quality—scientists believe that citizen science data suffer from low data quality, and project managers report inconsistent data quality across citizen science projects. 3. Awareness—science professionals may not realize citizen science projects relevant to their research interests exist. 4. Suitability—not all types of science are good matches for citizen science.

Survey respondents who had used citizen science data in their published research appeared to be more “outward-facing,” characterized by a higher level of trust in nontraditional data sources (e.g., retirees, NGOs) and higher annual participation in public outreach events (Fig. 1A). These scientists tended to indicate that properly trained non-professionals could collect their data (Fig. 1B). However, one of the most important factors was simply whether they were aware of projects focused on relevant taxa and/or study sites (Fig. 1B). However, one of the most important factors was simply whether they were aware of projects focused on relevant taxa and/or study sites (Fig. 1B). There are several explanations for this result. A scientist could be unaware of a program because a) one doesn't exist, b) they haven't found out about it, or c) they have misinformation about it. All of these explanations point to latent potential in applications of citizen science for researchers who are otherwise open to the idea. For instance, we found that biodiversity scientists who study birds had the highest probability of publication using citizen science data. This is not surprising as birds are particularly well-known subjects of citizen science (e.g., Breeding Bird Surveys, Christmas Bird Count, eBird). Although citizen science oversamples birds relative to both biodiversity science and described species, despite the fact that the diversity of focal species across all of biodiversity citizen science is quite broad and generally matches that of mainstream biodiversity science (Theobald et al., 2015).

By contrast, biodiversity researchers who had not used citizen science data felt skeptical of project data, expressing the belief that these data are of lower quality, and are more difficult to publish (Fig. 1A), perceptions that echo the case study findings of Riesch and Potter (2013). In fact, the perception of low data quality was the most important predictor in our preferences/perceptions RF analysis (Fig. 1A), and is a frequently cited barrier preventing integration of citizen science as a recognized tool in conservation (Lewandowski and Specht, 2015; Cohn, 2008; Brandon et al., 2003). Researchers who had not used citizen science data also preferred data sources proximal to them: projects or data associated with an academic institution (Fig. 3), and data collected by college students or graduates who were not retired (Figs. 1 and 4). Given that respondents to our survey were more likely to be interested in and familiar with citizen science than the general population of biodiversity scientists, the true extent of these biases may be more widespread. Countering these perceptions is a growing body of literature showing that high quality data can be produced by non-expert data collectors over a wide range of age, education and experience, as long as the study design clearly matches the collectors' abilities (Couvett et al., 2008; Kremen et al., 2011; Cox et al., 2012; Nagy et al., 2012; Tregidgo et al., 2013). We are also unaware of studies that show citizen science

projects managed by NGOs produce less reliable data than projects managed by academic, or government, institutions.

At the same time, for the field to reach its full potential as a legitimate tool in the conservation toolbox, the “science” in citizen science must be rigorous in any project with science-oriented goals. Setting data source biases aside, the data requirements that biodiversity scientists identified as important to their research were far from uniformly adopted by citizen science programs (Fig. 5). We found that scientists would prefer participants to be trained in-person and by experts. Although many projects do train participants in-person, those practices that facilitate measurement of quality assurance, such as testing participant competence pre- and post-training, were implemented only sporadically by our respondents. Metrics for evaluating data quality post-hoc (i.e., quality control, error detection) were also likely to reduce scientists' skepticism of citizen science. Although we asked project managers to enumerate potential verification methods, we did not ask whether verification was actually conducted. Increased transparency of the extent to which project data are verified, allowing a measure of and potential to correct for error, would be useful (Gardiner et al., 2012; Bonter and Cooper, 2012; Fowler et al., 2013).

Some scientists also noted that the scale (i.e., extent, grain, sample size or time period) of their work did not match known citizen science programs (Fig. 1B); however, project age is clearly a factor in visibility within the scientific community, as older projects have a higher probability of associated publications (Fig. 1C). This finding is consistent with other works showing that both spatial and temporal aspects of scale (e.g., larger, longitudinal studies) were important predictors of publication (Theobald et al., 2015; Tulloch et al., 2013; Chandler et al., in this issue).

Related to scale is the tradeoff between opportunistic and standardized sample design, and the potential applications of resultant data (Sagarin and Pauchard, 2010; Sarda-Palomera et al., 2012). Most scientists identified “standardization” as a data requirement; however, project managers reported that effort control (i.e., documented and/or standardized time spent and/or area surveyed) were not common. Yet both standardized and opportunistic schemes can be useful when applied appropriately (Lewandowski and Specht, 2015), in part due to advancing statistical techniques that increase the inferential power of opportunistic data when sample sizes are large (e.g., distribution and population trend estimates; Szabo et al., 2010, 2012; Bird et al., 2013; Danielsen et al., 2013; van Strien et al., 2013).

Notably, there is a relationship between the purpose of citizen science, as reported by project managers, and publication outcomes (Fig. 1C). Some citizen science projects do not move their data into peer-reviewed publication venues because that is not their intent; citizen science includes many goals other than data collection explicitly for peer-reviewed publication (e.g., Fig. 2). In our study, projects with data in the peer-reviewed literature elevated research as a purpose of citizen science, while those without mainstream scientific publications elevated education (Fig. 1C). Study designs that match intended applications are most likely to successfully meet project and research goals (Cohn, 2008; Moyer-Horner et al. 2012; Tulloch et al., 2013), for example, see Newman et al. (in this issue) for a discussion of the benefits of a place-based approach for projects that wish to see their data utilized for conservation decision making. Best practices for scientific outcomes include attention to training, protocol, and materials that prepare participants to effectively collect high quality data (Bonney et al., 2009; Hunter et al., 2013; Riesch and Potter, 2013; Shirk et al., 2012). A rigorous approach to the science in citizen science will lend itself to any data dependent outcome, including publication, decision making, and resource management and conservation applications.

In addition, citizen science is not equally applicable to all biodiversity research or conservation issues. From deep-sea vents, to microbiomes explored with high-powered genomics—volunteers cannot go everywhere or be trained to do everything. But evolving technology allows members of the public to participate in monitoring and conservation-



oriented data collection without leaving home. Examples include classifying habitat types and species demographics (e.g., Seafloor Explorer, Cropland Capture, DigiVol), and monitoring the abundance and distribution of indicator, threatened and invasive species via camera traps and continuous video monitoring (e.g., Snapshot Serengeti, Plankton Portal, Citizen Sort; Sandbrook et al., 2015 and Eaton et al., this issue). Some survey respondents pointed out that amateurs would simply not be able to collect their data, a conclusion that could be true, based on experience, a perception. These respondents tended to have not published citizen science data, but it is possible therefore that their assessments are based on experience with citizen science that did not result in publication. Nonetheless, the discrepancy between the proportion of scientists who report that trained non-experts could collect their data (78.6%), and those who have published using citizen science data (34.6%), implies there is potential for a much greater role for citizen science.

Other factors may influence the application of citizen science and include demographic and cultural context. The practice and perception of citizen engagement in biodiversity and conservation science may be dependent on country, history and culture of volunteerism, NGO activity, and socio-economic landscapes. An exploration of what influences perceptions and practices was beyond the scope of this study, but we suggest that an important next step in developing context-specific recommendations.

## 5. Recommendations and conclusion

One step that would aid the use of citizen science by scientists is increased transparency and availability of methods and data attributes. Based on the surveys from scientists, we suggest that metadata, at a minimum, should include data availability, conservation issue/question, study taxon/system, scales of time and space measured, sampling intervals, standardization protocols, and training and verification methods and could draw on guidelines developed by DataONE (Wiggins et al., 2013). If this information was publicly available in a single resource, it would complement clearinghouses that target public participation in science, like Scistarter and Citizen Science Central, by facilitating scientist engagement with citizen science—making it easy to identify and evaluate matches for a data need or collaboration (e.g. Loss et al., 2015).

Citizen science is at a crossroads of demonstrated successes and unrealized potential. As the field of citizen science grows, its practitioners and potential collaborators need a common understanding of the goals, opportunities and shortcomings of citizen science (Conrad and Hilchey, 2011). An alignment across scientific objectives, rigorous and relevant methodologies, and accessible, high quality data are important steps if we want to meaningfully engage the public in scientific research that expands the ability to understand and address ecological problems of our time.

## Acknowledgements

We gratefully acknowledge the Dimensions of Biodiversity Distributed Graduate Seminar for which funding was provided by a National Science Foundation Grant to Sandy Andelman, and Julia Parrish (DEB1050680). Theobald, Ettinger, DeBey, and Froehlich were NSF pre-doctoral fellows during this project (DGE-1256082, DGE-0718124, DGE-0718124, and DGE-0718124/DGE-1256082, respectively). We also thank G. Leubhn, J. Newman, J. Weltzin, S. Reichard, and several anonymous reviewers for providing feedback that greatly improved this manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.biocon.2016.05.014>.

## References

- 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., Pecl, G.T., Barrett, N., Frusher, S., 2013. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* 173, 144–154.
- Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *Bioscience* 59 (11), 977–984.
- Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J., Parrish, J.K., 2014. Next steps for citizen science. *Science* 343 (6178), 1437.
- Bonter, D.N., Cooper, C.B., 2012. Data validation in citizen science: a case study from Project FeederWatch. *Front. Ecol. Environ.* 10 (6), 305–307.
- Boyle, A., Sigel, B., 2015. Ongoing changes in the avifauna of La Selva Biological Station, Costa Rica: twenty-three years of Christmas Bird Counts. *Biol. Conserv.* 188, 11–21.
- Brandon, A., Spyreas, G., Molano-Flores, B., Carroll, C., Ellis, J., 2003. Can volunteers provide reliable data for forest vegetation surveys? *Nat. Areas J.* 23, 254–261.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Chandler, M., Rullman, S., Cousins, J., Nafeesa, E., Begin, E., Venick, G., 2015. Ecological and social outcomes from 7 years of citizen science evaluation: an Earthwatch case study. *Biol. Conserv.* (in this issue).
- Cohn, J.P., 2008. Citizen science: can volunteers do real research? *Bioscience* 58 (3), 192–197.
- Conrad, C.C., Hilchey, K.G., 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environ. Monit. Assess.* 176, 273–291.
- Cooper, C.B., Dickinson, J., Phillips, T., Bonney, R., 2007. Citizen science as a tool for conservation in residential ecosystems. *Ecol. Soc.* 12 (2), 11.
- Cooper, C.B., Shirk, J., Zuckerberg, B., 2014. The invisible prevalence of citizen science in global research: migratory birds and climate change. *PLoS One* 9 (9).
- Cosquer, A., Raymond, R., Prevot-Julliard, A.C., 2012. Observations of everyday biodiversity: a new perspective for conservation? *Ecol. Soc.* 17 (4), 2.
- Couvet, D., Jiguet, F., Julliard, R., Levrel, H., Teyssedre, A., 2008. Enhancing citizen contributions to biodiversity science and public policy. *Interdiscip. Sci. Rev.* 33 (10), 95–103.
- Cox, T.E., Philippoff, J., Baumgartner, E., Smith, C.M., 2012. Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecol. Appl.* 22 (4), 1201–1212.
- Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792.
- Danielsen, F., Adrian, T., Brofeldt, S., van Noordwijk, M., Poulsen, M.K., Rahayu, S., Rutishauser, E., Theilade, I., Widayati, A., The An, N., Nguyen Bang, T., Budiman, A., Enghoff, M., Jensen, A.E., Kurniawan, Y., Li, Q., Mingxu, Z., Schmidt-Vogt, D., Priksa, S., Thouttone, V., Warta, Z., Burgess, N., 2013. Community monitoring for REDD+: international promises and field realities. *Ecol. Soc.* 18 (3), 41.
- Danielsen, F.T., Burgess, N.D., Jensen, P.M., Pirhofer-Waiz, K., 2010. Environmental monitoring: the scale and speed of implementation varies according to the degree of people's involvement. *J. Appl. Ecol.* 47 (6), 1166–1168.
- Delaney, D., Sperling, C., Adams, C., Leung, B., 2008. Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biol. Invasions* 10 (1), 117–128.
- Devictor, V., Whittaker, R.J., Beltrame, C., 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Divers. Distrib.* 16 (3), 354–362.
- Dickinson, J., Zuckerberg, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: challenges and benefits. *Annu. Rev. Ecol. Syst.* 41, 149–172.
- Edgar, G.J., Stuart-Smith, R.D., Willis, T.J., Kininmonth, S., Baker, S.C., Banks, S., Barrett, N.S., Becerro, M.A., Bernard, A.T.F., Berkhout, J., Buxton, C.D., Campbell, S.J., Cooper, A.T., Davey, M., Edgar, S.C., Forsterra, G., Galván, D.E., Irigoyen, A.J., Kushner, D.J., Moura, R., Parnell, P.E., Shears, N.T., Soler, G., Strain, E.M.A., Thomson, R.J., 2014. Global conservation outcomes depend on marine protected areas with five key features. *Nature* 506, 216–220.
- Elbroch, M., Mwampamba, T., Santos, M., Zylberberg, M., Liebenberg, L., Minye, J., Mosser, C., Reddy, E., 2011. The value, limitations, and challenges of employing local experts in conservation research. *Conserv. Biol.* 25 (6), 1195–1202.
- Forrester, G., Baily, P., Conetta, D., Forrester, L., Klintzing, E., Jarcki, L., 2015. Comparing monitoring data collected by volunteers and professionals shows that citizen scientists can detect long-term change on coral reefs. *J. Nat. Conserv.* 24, 1–9.
- Fowler, A., Whyatt, J.D., Davies, G., Ellis, R., 2013. How reliable are citizen-derived scientific data? Assessing the quality of contrail observations made by the general public. *Trans. GIS* 17 (4), 488–506.
- Galloway, A.W.E., Tudor, M.T., Vander Haegen, W.M., 2006. The reliability of citizen science: a case study of Oregon white oak stand surveys. *Wildl. Soc. Bull.* 34 (5), 1425–1429.
- Gardiner, M.M., Allee, L.L., Brown, P.M.J., Losey, J.E., Roy, H.E., Smyth, R.R., 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Front. Ecol. Environ.* 10 (9), 471–476.
- Henderson, S., 2012. Citizen science comes of age. *Front. Ecol. Environ.* 10 (6), 283.
- Hunter, J., Alabri, A., van Ingen, C., 2013. Assessing the quality and trustworthiness of citizen science data. *Concurr. Comput. Pract. Experience* 25 (4), 454–466.
- Jiguet, F., Julliard, R., Couvet, D., Petiau, A., 2005. Modeling spatial trends in estimated species richness using breeding bird survey data: a valuable tool in biodiversity assessment. *Biodivers. Conserv.* 14 (13), 3305–3324.
- Krasny, M., Bonney, R., 2005. Environmental education through citizen science and participatory action research. In: Johnson, E.A., Mappin, M.J. (Eds.), *Environmental Education or Advocacy: Perspectives of Ecology and Education in Environmental Education*. Cambridge University Press, New York, New York, USA, pp. 291–318.



- Kremen, C., Ullman, K.S., Thorp, R.W., 2011. Evaluating the quality of citizen-scientist data on pollinator communities. *Conserv. Biol.* 25 (3), 607–617.
- Lewandowski, E., Specht, S., 2015. Influence of volunteer and project characteristics on data quality of biological surveys. *Conserv. Biol.* 29 (3), 713–723.
- Lin, Y., Deng, D., Lin, W., Lemmens, R., Crossman, N.D., Henle, K., Schmeller, D.S., 2015. Uncertainty analysis of crowd-sourced and professionally collected field data use in species distribution models of Taiwanese moths. *Biol. Conserv.* 181, 102–110.
- Loss, S.R., Loss, S., Will, T., Marra, P.P., 2015. Linking place-based citizen science with large-scale conservation research: a case study of bird-building collisions and the role of professional scientists. *Biol. Conserv.* 184, 439–445.
- Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J., Watt, A.D., 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends Ecol. Evol.* 25 (10), 574–582.
- Martin, K., 2015. Citizen science on the beach: scientific uses and policy impacts of data from grunion greeters. Presented at the Citizen Science Association Meeting, San Jose, CA.
- Moyer-Horner, L., Smith, M.M., Belt, J., 2012. Citizen science and observer variability during American pika surveys. *J. Wildl. Manag.* 76 (1), 1472–1479.
- Nagy, C., Bardwell, K., Rockwell, R.F., Christie, R., Weckel, M., 2012. Validation of a citizen science-based model of site occupancy for Eastern Screech Owls with systematic data in suburban New York and Connecticut. *Northeast. Nat.* 19 (6), 143–158.
- Newman, C., Buesching, C.D., Macdonald, D.W., 2003. Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation—“*Sed quis custodiet ipsos custodiet?*”. *Biol. Conserv.* 113 (2), 189–197.
- Newman, G., Clyde, M., McGreavy, B., Chandler, M., Haklay, M., Ballard, H., Gray, S., Mellor, D., Gallo, J., 2015. Leveraging the power of place in citizen science for effective conservation decision making. *Biol. Conserv.* (in this issue).
- R Core Team, 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (ISBN 3-900051-07-0. [online] URL: <http://www.R-project.org/>).
- Riesch, H., Potter, C., 2013. Citizen science as seen by scientists: methodological, epistemological and ethical dimensions. *Public Underst. Sci.* 23 (1), 107–120.
- Sagarin, R., Pauchard, A., 2010. Observational approaches in ecology open new ground in a changing world. *Front. Ecol. Environ.* 8 (7), 379–386.
- Sandbrook, C., Adams, W., Monteferrri, B., 2015. Digital games and biodiversity conservation. *Conserv. Lett.* 8 (2), 118–124.
- Sarda-Palomera, F., Brotons, L., Villero, D., Sierdsema, H., Newson, S.E., Jiguet, F., 2012. Mapping from heterogeneous biodiversity monitoring data sources. *Biodivers. Conserv.* 21 (11), 2927–2948.
- Schmeller, D.S., Henry, P.Y., Julliard, R., Gruber, B., Clobert, J., Dziok, F., Lengyel, S., Nowicki, P., Déry, E., Budrys, E., Kull, T., Tali, K., Bauch, B., Settele, J., Van Swaay, C., Kobler, A., Babij, V., Papastergiadou, E., Henle, K., 2009. Advantages of volunteer-based biodiversity monitoring in Europe. *Conserv. Biol.* 23 (2), 307–316.
- Shirk, J.L., Ballard, H.L., Wilderman, C.C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B.V., Krasny, M.E., Bonney, R., 2012. Public participation in scientific research: a framework for deliberate design. *Ecol. Soc.* 17 (2), 29.
- Silvertown, J., 2009. A new dawn for citizen science. *Trends Ecol. Evol.* 24 (9), 467–471.
- Silvertown, J., Cook, L., Cameron, R., Dodd, M., McConway, K., Worthington, J., Skelton, P., Anton, C., Bossdorf, O., Baur, B., Schilthuizen, M., Fontaine, B., Sattmann, H., Bertorelle, G., Correia, M., Oliveira, C., Pokryszko, B., Özgo, M., Stalažs, A., Gill, E., Rammul, Ü., Sólmos, P., Féher, Z., Juan, X., 2011. Citizen science reveals unexpected continental-scale evolutionary change in a model organism. *PLoS One* 6 (4), e18927.
- van Strien, A.J., van Swaay, C.A.M., Termaat, T., 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analyzed with occupancy models. *J. Appl. Ecol.* 50 (6), 1450–1458.
- Szabo, J.K., Fuller, R.A., Possingham, H.P., 2012. A comparison of estimates of relative abundance from a weakly structured mass-participation bird atlas survey and a robustly designed monitoring scheme. *Ibis* 154 (3), 468–479.
- Szabo, J.K., Vesk, P.A., Baxter, P.W.J., Possingham, H.P., 2010. Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. *Ecol. Appl.* 20 (8), 2157–2169.
- Theobald, E.J., Ettinger, A.K., Burgess, H.K., DeBey, L.B., Schmidt, N.R., Froehlich, H.E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M.A., Parrish, J.K., 2015. Global change and local solutions: tapping the unrealized potential of citizen science for biodiversity research. *Biol. Conserv.* 181, 236–244.
- Tregidgo, D.J., West, S.E., Ashmore, M.R., 2013. Can citizen science produce good science? Testing the OPAL Air Survey methodology, using lichens as indicators of nitrogenous pollution. *Environ. Pollut.* 182, 448–451.
- Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J., Martin, T.G., 2013. Realizing the full potential of citizen science monitoring programs. *Biol. Conserv.* 165, 128–138.
- Wiggins, A., Bonney, R., Graham, E., Henderson, S., Kelling, S., LeBuhn, G., Littauer, R., Lotts, K., Michener, W., Newman, G., Russel, E., Stevenson, R., Welzen, J., 2013. Data Management Guide for Public Participation in Scientific Research. DataONE ([online] URL: <http://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf>).
- Zuckerberg, B., Woods, A.M., Porter, W.F., 2009. Poleward shifts in breeding bird distributions in New York State. *Glob. Chang. Biol.* 15 (8), 1866–1883.