Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research


A R T I C L E   I N F O

Article history:
Received 28 June 2014
Received in revised form 3 October 2014
Accepted 22 October 2014

Keywords:
Biodiversity
Citizen science
Crowd sourcing
Global change
Volunteer monitoring
Climate change

A B S T R A C T

The collective impact of humans on biodiversity rivals mass extinction events defining Earth’s history, but does our large population also present opportunities to document and contend with this crisis? We provide the first quantitative review of biodiversity-related citizen science to determine whether data collected by these projects can be, and are currently being, effectively used in biodiversity research. We find strong evidence of the potential of citizen science: within projects we sampled (n = 388), ~1.3 million volunteers participate, contributing up to $2.5 billion in-kind annually. These projects exceed most federally-funded studies in spatial and temporal extent, and collectively they sample a breadth of taxonomic diversity. However, only 12% of the 388 projects surveyed obviously provide data to peer-reviewed scientific articles, despite the fact that a third of these projects have verifiable, standardized data that are accessible online. Factors influencing publication included project spatial scale and longevity and having publically available data, as well as one measure of scientific rigor (taxonomic identification training). Because of the low rate at which citizen science data reach publication, the large and growing citizen science movement is likely only realizing a small portion of its potential impact on the scientific research community. Strengthening connections between professional and non-professional participants in the scientific process will enable this large data resource to be better harnessed to understand and address global change impacts on biodiversity.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

1. Introduction

Citizen science, also referred to as community science or public participation in scientific research, is a growing movement that enlists the public in scientific discovery, monitoring, and experimentation across a wide range of disciplines. From microbiology (Cooper et al., 2010) to astronomy (Lintott et al., 2008), millions of non-experts lend their time and problem-solving abilities to discoveries as fundamental as new galaxy types (Cardamone et al., 2009), and disciplinary intersections as societally relevant as genomics and wellness profiling (Dove et al., 2012).

Can this burgeoning public interest in the scientific process be fruitfully applied to the scales of inquiry needed to address global change impacts on biodiversity (Silvertown, 2009; Devictor et al., 2010; Danielsen et al., 2010; Hochachka et al., 2012; Bird et al., 2014; Pimm et al., 2014)? Specifically, can citizen science be harnessed to attend to the “evil quintet” (Brook et al., 2008): climate change, overexploitation, invasive species, land use change, and pollution (Vitousek, 1992; Dirzo and Raven, 2003; Butchart et al., 2010; Danielsen et al., 2010; Hochachka et al., 2012; Bird et al., 2014; Pimm et al., 2014)?
2010)? Scientific interest in biodiversity and global changes affecting it has increased dramatically in recent decades. However, tracking, understanding, and ameliorating biodiversity losses requires collecting fine-grain data over regional to continental extents and decadal time scales (Magurran et al., 2010; Andelman, 2011; Bellard et al., 2012; Jetz et al., 2012) – a nearly impossible task for professional scientists and resource managers alone (Erwin and Johnson, 2000; Millennium Ecosystem Assessment, 2005; Hochachka et al., 2012; Pimm et al., 2014). Since public involvement and volunteerism is a tradition in many countries (Anheier and Salamon 1999; Corporation for National and Community Service, 2011), and can easily amount to millions, even billions, in in-kind economic worth (Independent Sector, 2011; Bureau of Labor Statistics, United States Department of Labor, 2012), citizen science offers a potential source of increasing support for basic and applied science. In other words, could volunteer efforts from the ballooning global human population be harnessed to effectively contribute to biodiversity research?

Citizen science has been increasingly advocated as a means for scientists to address large-scale data limitations (Devictor et al., 2010; Danielsen et al., 2010; Hochachka et al., 2012; Bird et al., 2014). Despite many published reviews and case studies on citizen science (e.g. Bonney et al., 2009; Dickinson et al., 2010; Dickinson et al., 2012), a quantitative analysis of biodiversity citizen science, critical to assessing whether these projects can indeed “fill the data gap”, is lacking. Furthermore, the extent to which scientists already use citizen science data is unknown. To address these needs, we performed the largest quantitative assessment of biodiversity-focused citizen science projects to date (to our knowledge). Noting that citizen science projects address multiple goals, from research and monitoring to participant experiential learning and education (Bonney et al., 2009; Dickinson et al., 2010; Dickinson et al., 2012), and acknowledging that there are other types of citizen-oriented science that engage non-scientists (e.g. Local Ecological Knowledge, Traditional Ecological Knowledge) that have the potential to be useful in filling data gaps (Huntington, 2000), we focused our analysis on the following overarching question: How is citizen science currently contributing to biodiversity research and what is its potential to contribute? To address this broad question, we had three specific queries:

(1) What is the current scope of biodiversity citizen science, in terms of its spatial and temporal scales, diversity coverage (including taxonomic, genetic, and functional diversity), and economic worth of the volunteerism engaged?
(2) To what extent is citizen science already integrated into peer-reviewed biodiversity research, and what factors influence the likelihood of publication?
(3) What is the potential of citizen science for global change research, as measured by the rate of project initiation, relative to professional interest in biodiversity science?

2. Materials and methods

2.1. Data collection

To assess the scope, and current and potential contributions of citizen science to biodiversity research, we created a database of biodiversity citizen science projects from around the world. We compiled and used information found on project websites and via Web of Science bibliographic searches. To corroborate our results and investigate follow-up questions, we separately surveyed both citizen science project managers and biodiversity scientists.

2.1.1. Web-based biodiversity citizen science database

To find project websites, we searched seven prominent English-language citizen science web-based clearingsouses, together listing more than 500 unique projects (Table A1). To be included in our database, projects had to match our operational definitions of citizen science, scientific research, and biodiversity. We defined “citizen science” as projects engaging volunteers (i.e., unpaid, and not receiving K-12, college, or internship credit), to collect and/or process data as part of “scientific research,” which we defined as collecting quantifiable information related to a specific issue or question (Miller-Rushing et al., 2012). Thus, projects such as those geared solely toward picture-sharing (e.g. AnimalsandEarth: http://www.animalsandeart.com/en/), park operations and maintenance (e.g. Volunteers in Parks: http://www.parks.ca.gov/?page_id=886), or education (e.g. Monarchs in the Classroom: http://www.monarchlab.org/mitc/) were excluded from our study.

We included only projects that collected biodiversity data, defined as the presence and/or abundance of identified taxonomic (e.g. species, genus, family), genetic, or functional groups as well as contextual information (e.g. collection date and location). After resolving duplicates and removing projects that did not match the above criteria, our database included 388 biodiversity citizen science projects.

For each project included in our database, we assessed four overarching features: (1) project characteristics; (2) information on the type(s), rigor, and availability of data collected, including measures of data quality control and assurance; (3) participant recruitment; and (4) number of peer-reviewed scientific publications resulting from each project’s data. Data obtained from project websites were collected independently by two co-authors, and discrepancies were resolved through group discussion and consensus. In total, our database included 45 fields (for metadata, see Table A2 and Appendix A); in this paper we focus on a subset (15 fields) of the total dataset, which we enumerate below.

Project characteristics included in the present analysis: headquarter type (governmental agency, academic institution, nongovernmental organization, a partnership between the above mentioned institution types, or “other” institution, including individual people, or private institutions like zoos or museums); first year of activity; last year of activity (if the project was no longer active); project goal as stated on the project website, which we subsequently coded as education/outreach, data/research, or both; project mission as stated on the project website, which we subsequently coded as including one or more of the following global change drivers: invasion, climate change, and land-use change; and number of peer-reviewed articles in the scientific journal literature containing project data (attained by citations of published work on project web sites, and through Web of Science searches of topic fields using the project name).

Data information included the: spatial scale over which sampling was conducted (linear distance between the farthest sampling points in the dataset, binned into four log-scale categories); system within which data were collected (one or more of: terrestrial, marine, freshwater); biodiversity dimension (one or more of: taxonomic, genetic, functional; where the latter was included only if traits were measured for focal organisms); taxonomic group(s) considered; amount of data available online in some form, either raw or summarized (all, some, or none); training method for taxon identification (none, or one or more of the following: in-person, electronic/web including downloadable pdfs, other, or some combination); training method for sampling protocol (same list as for taxon identification training); degree that sampling was standardized for time- and/or distance-based sampling effort (all, some, or none); and amount of verifiable data as defined by one or more
specific verification techniques: specimen collected, photograph taken, expert present (all, some, or none).

Participant information included: cumulative number of participants in the project, number of participants involved in the most recent year, and actual or estimated hours contributed per volunteer annually.

The database we assembled is, to our knowledge, the largest publicly available database to comprehensively catalog aspects of effort, scale, and rigor of citizen science projects. Nonetheless, it should be acknowledged that our dataset is not a random sampling of global citizen science, as projects were found using English-language web-based clearinghouses. For instance, while project head-quarters in our database span six continents, 89% of projects were located in North America (Fig. A1). Our data collection methods may also have excluded many smaller scale and/or ad hoc projects without a web presence.

2.1.2. Surveys
We sent follow-up surveys to managers of all extant projects in our database that listed contact information on project websites (329 projects, IRB approval number 43438). In total, we received 125 responses (38% response rate). The survey asked 32 questions, including binomial (yes/no), multiple choice (inclusive and exclusive), and open response questions, all pertaining to their citizen science project. In this study, we use information from six questions: year founded, current and cumulative participant numbers, data availability to scientists, and two questions regarding the use of project data for “published peer-reviewed articles in the scientific journal literature” (the first is binomial, and the second consists of exclusive multiple choice answers of the binned number of publications from zero upwards). Questions used in this analysis can be found in the Supplemental Materials.

We also surveyed biodiversity scientists to solicit information about their own biodiversity research, as well as their awareness and perceptions of biodiversity citizen science (IRB approval number 43438). We identified biodiversity scientists as corresponding authors of any study with “biodiversity” in the topic fields (title, keywords, or abstract), and with a working email address, as identified by Web of Science; \( n = 3148 \) (search performed on May 5, 2012). In total, we received 423 responses (13% response rate, less than a third of the response rate of project managers). The survey asked 25 questions, including binomial (yes/no), multiple choice (inclusive and exclusive), and open response questions. For this study, we used information from two questions: taxon group(s) studied, and the maximum linear distance between sampling sites within a single research project. Once again, questions used in this analysis can be found in the Supplemental Materials.

Surveys were designed to be relatively short, to be salient to the respondents (e.g. either pertaining to their own project or citizen science as a whole), and were sent from an email address affiliated with an academic institution (i.e. containing the suffix “.edu”). Additionally, we sent follow-up emails reminding respondents of the survey and requesting responses. Each of these characteristics are recommended to increase email survey response rate (Sheehan, 2001).

2.2. Analyses

2.2.1. Current scope of biodiversity citizen science
Spatial and temporal extent: We used the above database to assess the spatial and temporal extent of each project. To examine whether larger spatial and temporal extents were associated with higher participant numbers, we fit separate Poisson regressions with log-link. The response variables were cumulative or current participation (omitting projects that do not report these values), and the explanatory variables were log of project spatial extent, or project longevity (a composite from first and most recent year). These regressions were performed in R version 3.0.2.

Taxonomic and system breadth: To assess the taxonomic breadth of citizen science, we categorized each project as monitoring one or more of the following taxonomic groups: Protozoa, Bacteria, Fungi, Plants, Invertebrates (including Orthoptera, Lepidoptera, Hymenoptera, Diptera, Coleoptera, Myriapoda, Crustacea, Chelicerata, Echinodermata, Annelida, Mollusca, Cnidaria), Fishes, Amphibians, Reptiles, Birds, Mammals. Using chi-square tests, we compared the number of citizen science projects that collect data within each of these groups to estimates of: (1) globally identified species in each of these groups (Groombridge and Jenkins, 2002; Chapman, 2009), and (2) mainstream biodiversity science interest in these groups as proxied by research projects conducted among scientists answering our survey. Using chi-square tests, we also compared the distribution of ecosystem types studied by citizen science projects (marine, including estuarine, freshwater, and/or terrestrial) to the distribution of these ecosystem types on Earth (ChartsBin statistics collector team, 2010). All chi-square tests were performed in R version 3.0.2.

Economic worth: To estimate the economic worth of volunteerism in biodiversity citizen science, we first estimated the number of volunteers participating in citizen science annually from 1930 to 2012. Since less than half of projects in our database (191 out of 388) reported the number of participants online (either cumulatively for the most recent year, we augmented this dataset with data provided by project managers. If managers reported a range, we used the minimum value for conservative estimates. To estimate the total annual volunteers across all projects, we assumed the same average participation levels for the non-reporting projects, and created a range of participation using both our web and project manager databases. To estimate the median number of hours individual participants spend volunteering for citizen scientist projects, we used a subset of projects in the data base \( n = 106 \). Finally, we calculated an annual estimate for total in-kind value of biodiversity citizen science volunteerism as the product of these estimates \( \text{number of volunteers} \times \text{volunteer time estimates} \times \text{hours} \) multiplied by the U.S. national volunteer hourly in-kind rate (Independent Sector, 2011). We chose to use the U.S. national volunteer hourly in-kind rate as the majority of the projects listed in our database (89%) were housed in North America.

2.2.2. Publication rate
To understand the extent to which citizen science is integrated into peer-reviewed biodiversity research, and what factors (if any) constrain its integration, we used a logistic regression, fit in a Bayesian framework. For this analysis, we restricted our dependent variable (publication yes/no) to our web-based dataset, as the project manager survey was not comprehensive across our dataset and indicated high variability in managers’ interpretation of the term “peer-reviewed publication”. In our analysis, we included attributes that we thought were likely to affect publication, and that did not covary. We included project characteristics: headquarter type, project longevity, project spatial extent, and whether data were publicly available (reduced to binary), and data quality characteristics: method of training for specimen identification and data collection, degree that sampling was standardized, and whether data were verifiable.

We fit logistic regression models within a Bayesian framework to account for missing data as well as correct for linear separation in one variable (data availability). Specifying weakly informative priors for the linearly separable variable, data availability, constrained the model likelihood and improved parameter convergence (Gelman et al., 2008), but only affected the width of the credible interval, not the qualitative results (Fig. A2). We used non-informative priors for all other variables. For more details on
variable levels and coding, treatment of missing data, and linear separation, see the Supplemental Materials.

The model was fit using OpenBugs called from the BRugs library in R version 3.0.2 (Thomas et al., 2006; R Core Team, 2013). We ran three chains each with a burn-in of 15,000 iterations, which was sufficient to ensure convergence, as judged by visual inspection of the chain histories and the Gelman-Rubin statistic (Brooks and Gelman, 1997). We then sampled the posterior distributions from a further 10,000 iterations of each chain. The importance of explanatory variables was assessed using 95% Bayesian credible intervals on these posterior distributions.

2.2.3. Project initiation rate

To assess the increase in interest in biodiversity citizen science compared to biodiversity research by the professional scientific community, we examined the rate of project initiation and compared this rate to rates of professional scientific interest in biodiversity and conservation issues surrounding impacts to biodiversity. As a proxy for professional scientific interest, we used the number of peer-reviewed articles published per year on biodiversity, over exploitation, invasion, climate change, land-use change, and pollution, as proportions of total peer-reviewed articles published in scientific journals each year, found in Web of Science (for specific search terms used, see Table A3). We then fit logistic regressions, regressing the number of citizen science projects initiated annually (out of the total number of projects across all years) or the number of scientific publications in a given category (biodiversity and each of the global change drivers) annually (out of the total number of publications each year). Explanatory variables included year, category (i.e., citizen science, biodiversity, over exploitation, etc.), and their interactions. We constrained our analyses to the past 30 years (1982–2011), the period during which trends in citizen science increased dramatically. All regressions were performed in R version 3.0.2.

3. Results

3.1. Current scope of biodiversity citizen science

3.1.1. Spatial and temporal extent

Citizen science projects are executed across a range of spatial and temporal scales, with some attaining scales at or above mainstream biodiversity science (Fig. 1). Of the 326 projects for which the spatial extent of data collection could be assessed, 10% were local-scale projects (i.e., with 10 km or less separating the farthest two data collection sites), 22% operated at the 10–100 km scale, and 67% had a regional or larger extent (100–10,000 km or larger, Fig. 1A). According to our scientist survey respondents, this is comparable to the spatial scales at which professional scientists conduct biodiversity research: the majority reported sampling at 100 km or greater (55%; 354 out of 642 projects; Fig. 1B). Across the 328 projects for which lifespans data were available, mean citizen science project longevity was 10.9 years (median = 7, min = 0, max = 132).

Projects with large spatial and temporal extents of sampling, tend to enlist more volunteers (Table 1). We found that projects with the largest spatial extent have 7.5–14 times the number of participants compared with projects with small spatial extent (depending on whether cumulative or current participation is considered; Table 1). We also calculated participant number per meter for each project, in order to better understand if a disproportionate number of volunteers participate in projects of a specific scale. We found that small to medium projects (between 10 and 1000 linear km) had the greatest number of current annual participants per linear km. Furthermore, as project longevity increases, cumulative and current participation also increase: cumulative participation doubles every 6.4 years of project lifespan, and for every 41.2 years that a project is active, current participation levels double. The difference in these figures (i.e., that cumulative participation doubles faster than current participation) indicates that projects experience attrition throughout their lifespan.

3.1.2. Taxonomic and system breadth

Collectively, projects spanned a wide range of vertebrates, as well as major invertebrate phyla, plant families, bacteria, fungi, and even protozoa (Fig. 2). Volunteers monitor everything from thermophyllic bacteria in home water heaters and protozoan parasites (Ophryocystis elektroscirrha) of monarch butterflies throughout North America, to house sparrows (Passer domesticus) in India, and whale sharks (Rhincodon typus) around the world. Of the 388 projects in our database, 97% monitored taxonomic diversity, with the vast majority (87%) collecting data on multiple species. Compared to their presence on Earth, vertebrate groups and terrestrial ecosystems were significantly over-sampled, and invertebrates and marine ecosystems were under-sampled by citizen science (taxonomic groups: \( \chi^2 = 6614, df = 10, p < 0.001; \) ecosystems: \( \chi^2 = 5814, df = 2, p < 0.001 ; \) Fig. 2). Within invertebrates, citizen science sampling was biased relative to presence on Earth, with butterflies and shellfish oversampled and beetles and flies undersampled (\( \chi^2 = 98.27, df = 11, p < 0.001 \)). Focal taxa studied by citizen science projects also differed significantly from those studied by professional biodiversity scientists (\( \chi^2 = 232, df = 10, p < 0.001 \)), with a greater number of projects focused on reptiles, amphibians, birds, and algae and fewer focused on plants (Fig. 2).

One-fifth of the projects we surveyed monitored functional diversity. Common traits measured included reproductive or migratory phenology in birds and other organisms (e.g. MigrantWatch: Tracking Bird Migration Across India), body length (e.g. North Carolina Sea Turtle Project, Seabird Ecological Assessment Network), and chlorophyll a concentration for algae (e.g. Coalition for Buzzard’s Bay, Wisconsin Citizen Lake Monitoring). Only 2% of projects studied genetic diversity, for example by collecting feather samples of raptors to identify genetic lines (e.g. Seward Park Eagle and Raptor
DNA Fingerprinting), or by examining genotypic differences across plant clones (e.g. Cloned Plants Project).

### 3.1.3. Economic worth

We estimate that between 1.36 million and 2.28 million people volunteer annually in the 388 projects we surveyed, though variation is great (website-derived data: average of 3505 people per project per year, median = 50, standard error = 1914; project manager-derived data: mean from minimum reported participation = 5037, median = 200). Across projects for which we obtained time estimates (n = 106), volunteers spent an average of 21–24 h per person annually collecting biodiversity data (range: 0.5–107.1 h average per participant per year), which is comparable to rates reported in the literature (e.g. 34 h (Corporation for National and Community Service. 2011), and 51 h (Bureau of Labor Statistics, United States Department of Labor, 2012). We estimated the range of in-kind contribution of the volunteerism in our 388 citizen science projects as between $667 million to $2.5 billion annually. Note that this represents a minimum estimate for biodiversity citizen science worldwide, as our project sampling was restricted to only projects reporting in English and found in major online citizen science clearinghouses.

### 3.2. Publication rate

Advancing scientific understanding was an explicit primary goal for 97% of the citizen science projects we surveyed; however, only 12% of reviewed projects listed peer-reviewed scientific publications on their websites and/or returned results in Web of Science searches with project names in the topic fields (446 publications across 46 projects). Estimates of publication rates differed between our web-derived data and survey data: in follow-up surveys to project managers, 60% (n = 122) responded that data from their projects had been used in “scientific publication in the peer-reviewed science journal literature”. Later in the survey, though, 45% responded that their projects had one or more "peer-reviewed published journal articles". The inconsistent responses to survey questions suggest that project manager respondents may be confused about what constitutes peer-reviewed literature, potentially misinterpreting “peer” or “journal”. This interpretation conforms to other findings that gray-literature and technical reports are often included in citizen science self-reported publication totals (Shirk et al., 2012).

We found that the likelihood that citizen science project data will be published in a peer-reviewed scientific journal was related to project extent (spatial and temporal), data availability, and one aspect of data quality (Fig. 3). Specifically, citizen science project data were more likely to be published in peer-reviewed scientific literature if projects sampled at a large spatial extent and had been sampling for decades (Figs. 2 and 3). Data availability was also a positive predictor of publication; in fact, all published projects made data available in some form on their website. Of the 88% of projects (n = 343) for which we could determine data availability, 37% made all of their data available online, and 50% made some of it available; only 13% did not provide data in any form. In our follow-up surveys to citizen science project managers, only 3% of respondents said they would not share data if contacted by a scientist. Finally, projects that trained volunteers in species identification methods, using in-person or online training, were more likely to be published than projects that provided no identification training or trained with a combina-

### Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficient</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative participation–spatial scale</td>
<td>Intercept</td>
<td>4.255a</td>
</tr>
<tr>
<td></td>
<td>Spatial scale</td>
<td>0.939a</td>
</tr>
<tr>
<td>Current participation–spatial scale</td>
<td>Intercept</td>
<td>3.155a</td>
</tr>
<tr>
<td></td>
<td>Spatial scale</td>
<td>1.547a</td>
</tr>
<tr>
<td>Cumulative participation–project longevity</td>
<td>Intercept</td>
<td>6.450a</td>
</tr>
<tr>
<td></td>
<td>Longevity</td>
<td>0.109a</td>
</tr>
<tr>
<td>Current participation–project longevity</td>
<td>Intercept</td>
<td>7.997a</td>
</tr>
<tr>
<td></td>
<td>Longevity</td>
<td>0.0167a</td>
</tr>
</tbody>
</table>

a Indicates statistical significance to \( p < 0.001 \).
tion of methods. By contrast, training in data collection methods did not influence publication likelihood.

Other factors we tested had minor or no effect on publication likelihood (Fig. 3). Headquarter type influenced publication, in that projects housed in government agencies were slightly less likely to publish compared to those housed at academic institutions. Projects housed in NGOs, in partnerships between institution types, or in other institutions like zoos or museums, were just as likely to publish as those housed in academic institutions. We found no relationship between publication likelihood and data standardization, or whether or not data were verifiable.

3.3. Project initiation rate

We found that the number of biodiversity-oriented citizen science projects has increased dramatically in the last 75 years (Fig. 4, Table 2). Project initiation rose particularly steeply during the most recent 30 years, coincident with the rising scientific interest in biodiversity in general (Table 2). In addition, although the proportion of papers published on over-exploitation, invasion, climate change, land-use change, and pollution have all increased (Fig. 4), none exceeds the rate of creation of citizen science projects during this time period (Table 2). Within our citizen science sample set, over a quarter of projects purported to address factors impacting biodiversity, specifically referencing land use change (29%), invasive species (24%), and/or climate change (16%).

4. Discussion

4.1. Citizen science as a resource

Biodiversity citizen science currently provides a valuable, albeit underutilized, resource for global change research. Citizen scientists act locally, so collectively they gather fine-grain data that achieve datasets reaching regional and even global scales, and long temporal extents (Fig. 1). Indeed, our data suggest that citizen scientist projects span spatial scales that are comparable to professional scientists (Fig. 1) and are, on average, 7 years longer than the mean length of NSF grants (National Science Foundation, 2012). These types of datasets – that reach great spatial and temporal extents without compromising fine-grain resolution – are particularly critical for tracking, understanding, and ameliorating global change impacts to biodiversity (Magurran et al., 2010; Andelman, 2011; Bellard et al., 2012; Jetz et al., 2012).

Additionally, despite the fact that many of the most well-known citizen science projects focus on birds (e.g. Christmas Bird Count, Breeding Bird Surveys, eBird), citizen science includes a broad range of taxa (Fig. 2). Although there are biases in citizen science sampling efforts relative to abundance on Earth, these biases are consistent with biases found in professional science (Fig. 2, Erwin and Johnson, 2000; Millennium Ecosystem Assessment, 2005; Martin et al., 2012). Furthermore, the time donated by the large number of engaged volunteers (approximately $0.7–2.5 billion annually) is equivalent to 11–42% of the annual U.S. National Science Foundation budget (National Science Foundation, 2013). Finally, the growth in interest in citizen science (Fig. 4) represents unprecedented opportunity and potential to contend with global changes with local observers.

Citizen science datasets are also useful resources for global change research because they are readily available. We found that a majority of citizen science projects already make their data available online, and 72% of citizen science project manager survey respondents ($n = 118$) said they would share data if contacted by a scientist. In fact, biodiversity citizen science projects appear to make their data publically available at higher rates than professional scientists in the same field, as only 8% of Division of Environ-
For example, the United Kingdom Butterfly Monitoring Scheme provides an example of citizen science projects that have played a significant role in understanding global change impacts to biodiversity, especially pre-perturbation, as well as shifts following global changes. Large-scale environmental crises simply by having more scientists may be better able to quickly contend with emergent, urgent issues, working collaboratively with millions of informed citizens, thereby increasing the likelihood of their data being published in peer-reviewed journals (Erwin and Johnson, 2000; Millennium Ecosystem Assessment, 2005; Hochachka et al., 2012; Pimm et al., 2014). However, by working collaboratively with millions of informed citizens, scientists may be better able to quickly contend with emergent, large-scale environmental crises—simply by having more widespread access to information regarding biodiversity patterns, as well as shifts following global changes. Our research identified measures of scale, both spatial scale and longevity, as positive predictors of publication (Fig. 3A). There are several examples of citizen science projects that have played important roles in understanding global change impacts to biodiversity—most of which are long-lived or large-scale. For example, the United Kingdom Butterfly Monitoring Scheme (established: 1976, spatial extent: <10 000 km) contributed data to one of the first large-scale documentations of pole-ward range shifts due to climate warming (Parmesan et al., 1999), Breeding Bird Survey (established: 1994, spatial extent: <1000 km) data were used to determine the spread of West Nile virus across North America (LaDeau et al., 2007), and Reef Environmental Education Foundation (established: 1990, spatial extent: >10 000 km) data have been used to inform marine conservation efforts through the documentation of large-scale changes in world-wide shark species abundance (Ward-Paige and Lotze, 2011; Ward-Paige et al., 2011).

Large-scale studies may be more likely to reach the scientific literature because they are better able to measure change over space and time (Tulloch et al., 2013; Bird et al., 2014), and therefore can quantify impacts of management and policy (e.g. Danielsen et al., 2014). Another possible explanation could be that scientists may be more aware of older, more widespread citizen science projects, and thus more likely to use their data in publications. Regardless of the cause, at least one other study has also found that broad-extent citizen science monitoring projects have higher impact in scientific literature, in addition to being more cost-effective, compared to more short-term, cross-sectional studies (Tulloch et al., 2013).

We were surprised that probability of publication was largely unaffected by the data quality assurance measures we assessed (Fig. 3B), since concerns about quality, consistency, and reliability of citizen science data are widespread (Silvertown, 2009; Dickinson et al., 2010; Bonter and Cooper, 2012; Bird et al., 2014). In our view, this result does not suggest that data quality is unimportant. Rather, it suggests that perhaps most projects have adequate data quality measures in place, or that non-professional data can be comparable to data collected by professional scientists, as others have suggested (Kremen et al., 2011; Holt et al., 2013; Cooper et al., 2014). Alternatively, given the nature of complex ecological data, individual observer variation may not contribute substantially to additional noise in the data (Bird et al., 2014). Taken together with the positive effect of temporal and spatial scale on publication (Fig. 3A), our findings suggest that even “messy” citizen science datasets are valuable if sample sizes are large, as variation among observers can be reconciled statistically (Schmeller et al., 2009; Dickinson et al., 2010; Bird et al., 2014).

Despite our findings that large-scale projects have a higher likelihood of reaching the peer-reviewed literature, we wish to emphasize that smaller-scale projects can still provide utility to the scientific community and to society. Specifically, projects that monitor single species or that are constrained to narrow geographic regions could inform or improve conservation efforts. For example, passage of the 1972 Clean Water Act in the U.S. required states to assess the quality of their surface water, which spurred many grassroots volunteer water monitoring efforts by lake and stream conservation groups (Lee, 1994). The legacy of these efforts is apparent in our dataset (Fig. 4), which includes many projects monitoring biological indicators of water quality, including presence and/or abundance of invertebrates (Fig. 3). Indeed regional projects may be likely to bypass the peer-reviewed literature and affect management or policy directly, or reside in the gray literature, where they may inform natural resource management practices. We do not wish to discount the value of these regional projects nor the utility and relevance of their data; nonetheless, the data will be more useful to the scientific community at large if they reach the peer-reviewed literature.

5. Conclusions

There is even greater potential for citizen science to address global changes moving forward, given the rapid increase in citizen...
science projects and people engaged in them (Fig. 4). However, much of this potential in citizen science will not be fully realized if citizen science data do not reach the peer-reviewed scientific literature. The scientific impact of citizen science could be much greater if it were embraced by, and better integrated into, established modes of scientific research.

For citizen science data to become integrated into established modes of inquiry, it must first be explicitly acknowledged as a source of information. While it is possible that more citizen science data reach the peer-reviewed literature than our search identified, our data suggest that professional science may not embrace citizen science data as usable. For example, the higher publication rates claimed by project manager survey respondents (45% or 60%) as opposed to our web-based search results (12%) could indicate that some project data are published without direct or prominent advertisement of its origins (e.g. project name lacking from topic fields), an interpretation also mentioned in Tulloch et al. (2013) and Cooper et al. (2014).

Integration may be improved through increasing scientists' awareness of and accessibility to citizen science data, and organizations such as the National Phenology Network (http://www.usanpn.org) and the recently created Citizen Science Association (http://www.citizensciencesassociation.org) may be crucial vehicles. These and other umbrella efforts will be most successful if they facilitate and deepen connections between mainstream science and citizen science. For example, if they serve as a matching service, connecting professional scientists and citizen science projects that are well-suited to one another, they may be able to reduce taxonomic biases or increase genetic and functional diversity sampling. Our results suggest that citizen science projects should focus on data quantity, covering large spatial and temporal scales, if they wish to be used in peer-reviewed scientific publications. Thus, designed as a cooperative matching service, umbrella efforts could further the utility of citizen science datasets by targeting expansion efforts to broad spatial scales. Scientists may be able to initiate their own citizen science project or help ensure the continued existence of high quality citizen science datasets by working more closely with citizen science project managers to advocate for sustaining projects over the long-term, to collaborate on grant applications and other funding resources, and to publish their data in the peer-reviewed scientific literature.

For biodiversity science, the era of ivory tower science is over (Devictor et al., 2010; Könneker and Lugger, 2013). We need a paradigm shift, wherein scientists and nonscientists work collaboratively to contend with emergent, large-scale environmental issues. If biodiversity science does not engage nonscientists, as biodiversity and ecosystem services continue to erode, it runs the risk of becoming irrelevant in the eyes of a public that may offer local solutions to global problems.

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 T. Daniels, J. Hoekstra, T. Ricketts, S. Reichard, E. Neeley, S. Andelman, and several anonymous reviewers for providing feedback that greatly improved this manuscript. Data reported here can be found housed in PANGEA.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bioccon.2014.10.021.

References


E.J. Theobald et al. / Biological Conservation 181 (2015) 236–244 243


