Biological Conservation 208 (2017) 121–126

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/bioc

Lessons from citizen science: Assessing volunteer-collected plant phenology data with Mountain Watch

Caitlin McDonough MacKenzie [a,*], Georgia Murray [b], Richard Primack [a], Doug Weihrauch [b]

[a] Boston University, Department of Biology, 5 Cummington Street, Boston, MA 02215, USA
[b] Research Department, Appalachian Mountain Club, 361 Route 16, Gorham, NH 03581, USA

ARTICLE INFO

Article history:
Received 9 December 2015
Received in revised form 11 July 2016
Accepted 22 July 2016
Available online 31 July 2016

Keywords:
Citizen science
Phenology
Alpine
White Mountain National Forest

ABSTRACT

Citizen science has the potential to expand the scope of data collection, engage the public in research, and answer big scientific questions. But, the quality of volunteer-collected data is often called into question, and citizen science programs must find ways to assess the validity of this concern. Here, we review five years of volunteer-collected data from an alpine flower monitoring citizen science project and present our efforts to investigate the quality of the volunteer-collected data. We found disparity between citizen scientists' self-assessed and actual plant species identification skills, indicating error in either true plant identification or reported location, consequently limiting the use of this dataset. Citizen science programs, including this project, must assess their data, and then make adjustments — in training, data collection methods, or goals — in order to produce quality data consistent with their scientific intentions. Indeed, this project now relies only on observations from seasonal trained staff and a handful of skilled volunteers in light of these findings.

1. Introduction

Citizen science projects with research-oriented goals must develop methods for assessing and improving the quality of their volunteer-collected data. Validating the quality of this volunteer-collected data to uphold the scientific integrity of a project is a common theme among citizen science literature, however no universal rules of data quality have emerged, perhaps because projects vary so much in their scope, scale, and study systems (Bonney et al., 2009; Miller-Rushing et al., 2012). Further, many citizen science programs have education in addition to research goals or have pre-existing audiences with varying skill sets. The high volumes of data from the dispersed data collection model of citizen science can reduce the inherent error in volunteer-collected data (Dickinson et al., 2012), however programs currently engaging in citizen science must still employ a range of Quality Assurance and Quality Control (QA/QC) approaches to fit both the types of data gathered and the audiences that participate. Research-oriented citizen science programs in ecology, climate change biology, or conservation must assess the species identification skills, the field measurements, the qualitative classifications, and the quantitative counts recorded by their citizen scientists. It is important for the citizen science community to share lessons from the fields, the workshops, the classrooms, or the websites where they work to assess and control the quality of volunteer observations. There is a special need for examples of programs that have experienced problems, rather than reporting only on projects that were successful.

Assessing and controlling the quality of volunteer-collected data is often heralded, but practical examples of implementing these measures are missing or folded into larger papers without thorough examination (Cooper et al., 2014; Sullivan et al., 2009). In addition, the largest and most well-known citizen programs have access to resources including infrastructure, experts, and software programming that allow for streamlined QA/QC and adjustments within programs; smaller, local programs often cannot afford these luxuries (Bonter and Cooper, 2012; Wiggins, 2013). A 2010 survey of 128 citizen science programs with a focus on monitoring invasive species — most of which fit this smaller, local category — found that only 39% incorporated quality checks on volunteer-collected data (Crall et al., 2010). Forty percent of the programs in this survey reported that they obtained a majority of their funding from grants; across all types of citizen science programs, short-term funding like this is a common obstacle to efforts to assess volunteer-collected data (Crall et al., 2010).

A recent review of the peer-reviewed literature on the quality of volunteer-collected data in biological monitoring found that most studies assessing citizen science focused on the act of data collection; the most common method reported was comparing volunteers with experts or professionals (Lewandowski and Specht, 2015). In this vein, vegetation surveys have re-sampled permanent transects with professional botanists (Brandon et al., 2003; Galloway et al., 2006), while monitoring programs for pollinators (Kremen et al., 2011), aquatic invertebrates (Delaney et al., 2007), terrestrial invertebrates (Lovell et

* Corresponding author.
E-mail address: mackenz@bu.edu (C. McDonough MacKenzie).

http://dx.doi.org/10.1016/j.biocon.2016.07.027

al., 2009), and benthic macroinvertebrates (Engel and Voshell, 2002) have compared volunteer observations to data collected by researchers in the same sites. Across these case studies, data comparisons with experts validated the data collection models and improved the associated programs; the volunteer-collected data was rated as high quality, or indistinguishable from the experts, reflecting a "good" citizen science program. However, honest accounts of programs identifying unreliable data and evaluating faults in an underlying data collection model are missing from the literature, and would provide valuable information, especially to smaller, more local citizen science programs with limited resources.

Here, we present a case study of one citizen science program, a project to assess its volunteer-collected data, and the lessons from this QA/QC effort. The Appalachian Mountain Club (AMC), a nonprofit organization dedicated to conservation, education, and recreation in the northeastern United States, launched the Mountain Watch Alpine Flower Watch (Mountain Watch) citizen science program in 2005 to collect long-term alpine plant phenology data in the White Mountain National Forest, New Hampshire. Alpine ecosystems are generally sensitive to changes in climate (Pauli et al., 2014) and phenological timing has implications for the success and long-term persistence of the plants within those systems (Inouye, 2008), but the remote location of alpine habitats makes it a challenging place to obtain observational data with good spatial and temporal resolution.

Mountain Watch solicits hikers to become citizen scientists, and asks volunteers to record flowering phenology observations along the trails in New Hampshire's White Mountains. With this data, the AMC planned to track the local ecological effects of climate change on plant communities in the small and fragmented alpine habitats of New Hampshire. The citizen science program also has core educational goals to engage the hiking community in the issue of climate change through hands-on monitoring. The available audiences were the large number of hikers (~500,000 per year) visiting AMC facilities in the White Mountains and a self-selected group of already-active volunteers. In addition to the volunteer-collected data from Mountain Watch, the AMC has utilized research staff, as well as seasonal naturalists and interns, to record phenology data at permanent plots in the White Mountains since 2005. Both the citizen science project and the research staff observations follow the same monitoring protocol, however only the research staff observations have resulted in a scientific publication to date.

In 2014, the AMC's research department used long-term weather records from the Mount Washington Observatory and alpine plant phenology data gathered by research staff to hindcast flowering phenology and assess late-spring/early-summer frost risks for three of the Mountain Watch plant species (Kimball et al., 2014). The volunteer-collected data from the Mountain Watch program could broaden the geographic scope of this research (from the twelve plots proximate to Mount Washington's meteorological station included in this analysis to alpine habitats across the northeastern United States) and provide long-term phenology data (expanding on the four years of data included in this analysis with on-going citizen science efforts). However, the potential of the Mountain Watch dataset is dependent on its quality. To this end, we looked at the first five years of volunteer-collected Mountain Watch data from the perspective of quality assurance and quality control.

We reviewed the volunteer-collected Mountain Watch data from 2005 to 2009, conducted vegetation surveys at locations recorded by volunteers, and assessed the Mountain Watch data collection model. We used chi-square tests to describe the relationships between species identification rates and characteristics including relative abundance, phenophase, and the volunteers' self-assessed certainty of identification. In the process, we identified two main challenges in QA/QC for the Mountain Watch data: 1) our ability to review the data hinged on the precision of the geographic location descriptions provided by volunteers, and 2) for the majority of volunteers, we did not know their plant identification skills or prior knowledge of the alpine habitat aside from their self-assessed certainty of identification on the datasheet. From our review of five years of volunteer-collected data, we were able to identify potential shortcomings in the original Mountain Watch data collection model, adjust the citizen science program, and share lessons in QA/QC methods for a small, local program with limited resources.

2. Study area

Alpine habitat in the northeastern United States, is limited to ~34 km^2 of fragmented ridges and summits above treeline. The largest of these alpine areas comprises ~11.3 km^2 in the Presidential Range of the White Mountain National Forest, New Hampshire (Kimball and Weihrauch, 2000). The Presidential Range includes New England's highest peak, Mt. Washington (1917 m a.s.l.), three AMC backcountry huts catering to backpackers, and some of the most popular hiking trails in the White Mountain National Forest. This case study focuses on data collected here.

Six common and charismatic alpine plant species were chosen as Mountain Watch target species: ericaceous shrubs *Rhododendron groenlandicum, Vaccinium uliginosum* and *Vaccinium vitis-idaea;* herbaceous *Geum peckii* which is endemic to the White Mountains and Nova Scotia; alpine sedge *Carex bigelowii;* and the circumpolar pin-cushion plant *Diapensia lapponica.* Criteria considered in target species choice included ease of identification, limited look-alike species, ease in phenophase observation, and a variety of life histories and phenological timing. All six are slow growing, long-lived perennials; the plant communities and species composition in the Presidential Range has not changed over the duration of this study.

3. Methods — Mountain Watch program

The AMC Mountain Watch program builds on the popularity of the White Mountain National Forest trail system and recruits hikers to become citizen scientists. The only prerequisites for participating in Mountain Watch are interest, a species identification field guide, and a blank datasheet, which are available online or at any AMC lodge or backcountry hut. Mountain Watch training was provided at backcountry huts as an evening nature program, but the frequency of these programs varied and AMC did not track which volunteers had attended a training program over the years examined. The datasheet asks volunteers to identify the six target alpine plant species, and record the current phenophase (i.e.: before flowering, flowering, or after flowering) for each observation (Fig 1). Volunteers record an observation by checking each phenophase present and circling the dominant phenophase.

Volunteers also rank their certainty of identification (CID) for each species on a scale from 1 (uncertain) to 3 (very certain) on the datasheet. The target species and locations are unmarked to protect the integrity of the National Forest and to encourage data collection across the alpine habitats of the White Mountains, and volunteers are asked to record the geographic location of their observations in an open-ended space on the datasheet (Fig 1). A map of the Presidential Range was printed on the reverse of the datasheet to provide guidance for the observation location. Occasionally, volunteers provided GPS coordinates in this space, but most often they simply wrote a description of their location. During the years examined in this study, cellphone service in the Presidential Range was spotty to nonexistent, and GPS-enabled smartphones had not yet become ubiquitous accessories for hikers (Wiggins, 2013).

4. Methods — Mountain Watch QA/QC

In 2009 we surveyed the vegetation at geographic locations in the Presidential Range recorded by volunteers in an effort to assess the

ALPINE FLOWER BASIC DATA SHEET

Date _________________ **Name** _____________________________

# of people on hike _____ **To receive updates provide email address** ___________________________

Alpine area or nearest major town/highway ________________________________ **State** _____

Describe your hike route ____

Instructions:

Stop at a place you can find on a map, like a trail junction, summit or large stream crossing, and **look** to see if there is one or more of the target species growing near that spot. You are unlikely to find all 6 target species at one site, but many occur together.

Describe your observation location in the first box below (more room on back). Enter elevation & lat/long (if known). Mark your location and label it with the location # on a sketch map on reverse side.

Record flowering status of the target plant(s). The plant(s) at your location may exhibit more than one flowering stage. Mark ALL stages that apply with an "**X**", and CIRCLE the stage that appears to be dominant; use the field guide for guidance. Indicate your certainty of identification (ID) on a scale of 1-3 (1=uncertain, 2=somewhat certain, 3=most certain).

Return this sheet to any NH AMC destination frontdesk, mail to: *AMC, Mtn. Watch P.O. 298, Gorham NH 03581* or enter online at www.outdoors.org/mountainwatch/volunteers

MOUNTAIN WATCH

PLEASE MAKE ALL OBSERVATIONS FROM THE TRAIL

Observation location description	**Species**	**Before Flowering**	**Flowering**	**After Flowering**	**Certainty of ID (1-3)**
Observation location #1 Elevation ________ (feet) Lat/Long ________ (decimal degrees)	Alpine bilberry				
	Bigelow's sedge				
	Mountain avens				
	Diapensia				
	Mountain cranberry				
	Labrador tea				
Observation location #2 Elevation ________ (feet) Lat/Long ________ (decimal degrees)	Alpine bilberry				
	Bigelow's sedge				
	Mountain avens				
	Diapensia				
	Mountain cranberry				
	Labrador tea				
Observation location #3 Elevation ________ (feet) Lat/Long ________ (decimal degrees)	Alpine bilberry				
	Bigelow's sedge				
	Mountain avens				
	Diapensia				
	Mountain cranberry				
	Labrador tea				
Observation location #4 Elevation ________ (feet) Lat/Long ________ (decimal degrees)	Alpine bilberry				
	Bigelow's sedge				
	Mountain avens				
	Diapensia				
	Mountain cranberry				
	Labrador tea				
Observation location #5 Elevation ________ (feet) Lat/Long ________ (decimal degrees)	Alpine bilberry				
	Bigelow's sedge				
	Mountain avens				
	Diapensia				
	Mountain cranberry				
	Labrador tea				

Thanks for your help! Find your data, forms for your next hike and more information at www.outdoors.org/mountainwatch

Fig. 1. Mountain Watch datasheet for volunteer-collected data during the period of this case study (2005–2009).

quality of the volunteer-collected data. First, we coded all geographic location descriptions provided by volunteers into "best guess", "general", and "precise" categories, based on the specificity of their description. "Best guess" descriptions were expansive (i.e. "on the Crawford Path"); "precise" descriptions reported specific, particular locations ("Intersection of the West Side Trail and Gulfside Trail"); "general" descriptions fell between ("on the Crawford Path about halfway between junctions with Davis Path and Camel Trail").

We surveyed the nineteen most popular precise locations recorded by volunteers in the Presidential Range. At each location, we recorded the presence/absence of each target species found within 10 m. The "best guess" or "general" locations could not be surveyed because those locations could not be pinpointed to within a 10 m radius. Volunteer observations of absent target species were categorized as "misidentified". Non-Mountain Watch species that could be misidentified as a target species (look-alikes) were also recorded across at these observation locations (Table 1, see also Appendix S1: Fig S1).

To determine identification rates, the volunteer-collected data was compared to our survey results. For example, volunteers recorded *D. lapponica* in 19 locations, but the survey validated the presence of *D. lapponica* in only 11 of those locations. The volunteer observations from the 11 locations verified by the survey were classified as "correct identifications"; volunteer observations of *D. lapponica* from the other locations were classified as "misidentified".

A series of chi-square analyses explored the relationships between these misidentifications and a volunteer's self-assessed CID, a plant's phenophase, and a species' abundance. To estimate the relative abundance of each target species, we used the number of survey locations where a target species was present in the Presidential Range (i.e. *D. lapponica* was present at 11 of the 19 locations; its relative abundance = 0.58). For these analyses, "blank" CID responses were disregarded and Fisher's exact test was used to interpret small cell counts.

5. Results — geography & plant identification

Mountain Watch received 1775 volunteer-collected observations in the Presidential Range during its first five years as a citizen science program. Of these, 1223 observations (69%) were recorded at "precise" locations, while 149 observations were at "best guess" locations, 197 were

Table 1
Mountain Watch target species and their "look-alike" species in the Presidential Range alpine habitats, New Hampshire.

Target species	Look-alike species
Vaccinium uliginosum	*Alnus viridis*
	Betula cordifolia
	Betula glandulosa
	Salix spp.
	Vaccinium angustifolium
	Vaccinium boreale
	Vaccinium vitis-idaea
Rhododendron groenlandicum	*Chamaedaphne calyculata*
	Kalmia angustifolia
	Kalmia polifolia
	Rhododendron canadense
Diapensia lapponica	*Empetrum nigrum*
	Kalmia procumbens
	Rhododendron lapponicum
Vaccinium vitis-idaea	*Empetrum nigrum*
	Gaultheria hispidula
	Kalmia procumbens
	Vaccinium uliginosum
Geum peckii	*Geum macrophyllum*
	Ribes glandulosum
	Rubus chamaemorus
Carex bigelowii	*Deschampsia flexuosa*
	Juncus trifidus
	Luzula spicata
	Trichophorum caespitosum

Table 2
Summary of Mountain Watch volunteer-collected data included in this QA/QC study. For each target species, we present the number of observations and locations recorded by volunteers at 19 precise locations in the Presidential Range (2005–2009) and the results of our 2009 QA/QC survey. The rate of species identification is defined as the percent of volunteer observations that agree with our survey. Volunteer observations of target species noted as absent at a location during the survey were categorized as "misidentified". Relative abundance of each target species is calculated as the number of survey locations where a target species was present in the Presidential Range divided by 19 (the number of precise locations surveyed).

Alpine plant species	Common name	Number of observations recorded by volunteers	Number of locations recorded by volunteers	Number of locations recorded by survey	Rate of species identification	Relative abundance
Vaccinium uliginosum	Alpine bilberry	113	18	17	96.5	0.95
Rhododendron groenlandicum	Labrador tea	128	17	5	27.3	0.26
Diapensia lapponica	Diapensia	216	19	11	73.6	0.58
Vaccinium vitis-idaea	Mountain cranberry	158	19	10	60.8	0.53
Geum peckii	Mountain aven	93	15	3	28.0	0.16
Carex bigelowii	Bigelow's sedge	157	19	18	98.1	0.95

at "general" locations, and 206 were entered into the database as "unknown" locations. Imprecise location descriptions rendered over five hundred volunteer observations (nearly a third of all observations) from the Presidential Range ineligible for our QA/QC analysis.

Our survey was conducted at the 19 most popular precise locations, comprising a subset of 865 observations in the Presidential Range. At these 19 locations, our survey found a 33.6% plant misidentification rate among volunteers (Table 2).

The rate of identification varied with species, ranging from 98.1% (*C. bigelowii*) to 27.3% (*R. groenlandicum*). It appears that the large variation in misidentification rates is influenced, in part, by the variation in the relative abundance of the target species; while *C. bigelowii* was found in all but one location and has a very low misidentification rate, the species with the lowest relative abundances (*G. peckii* and *R. greonlandicum*) carry the highest misidentification rates (Table 2). For *G. peckii* and *R. groenlandicum*, the volunteers' identification rates are no better than random — the species reports occur independently of where the species actually grow.

5.1. Phenophase effects:

A significant relationship existed between flowering phenophase and CID (Certainty of Identification) for all observations and all species (chi-square test, $p = 0.03$, Table 3). Over 60% of observations with the highest CID were associated with plants in flower. The presence of flowers boosts the volunteers' confidence in their ability to identify a plant.

Similarly, a significant relationship existed between flowering and correct species identification (chi-square test, $p = 0.002$). However, while 72.8% of "no flower" observations were correctly identified, only 64% of "flowering" observations and 59% of "dominant flowering" observations were correctly identified.

When the observations were analyzed by species, the relationship between accuracy and flowering phenology was driven by two species: *R. groenlandicum* and *D. lapponica* (Table 3). Though the identification rates for *R. groenlandicum* were much lower than identification rates for *D. lapponica* across the board (Table 2), for each species, "no flower" observations were correctly identified at a higher rate (36% and 83% respectively) than "flowering" (19% and 64%) and "dominant flowering" observations (11 and 73%).

5.2. Self-assessed certainty of identification:

A review of the CID rankings reveals a bimodal distribution of very certain (456), and volunteers not reporting a level of certainty at all (for 338 observations, the CID was blank). Very few observations ($n = 27$ and 44 respectively) were recorded with "1" (low) or "2" (medium) CID. The QA/QC survey found no significant relationship between CID and correct species identification. For each of the Mountain Watch species, accuracy is independent of CID (Table 3). Therefore, self-reported CID rankings appear insufficient to assess a volunteer's actual ability to identify a given plant.

6. Discussion

Implementing effective QA/QC methods in volunteer-collected data is a challenging endeavor, but lessons from these efforts can provide valuable information to the community of programs practicing citizen science. In the case of Mountain Watch, the need to accumulate enough citizen science observations for evaluation, and the limited resources of a small research department restricted the AMC's ability to conduct a more thorough QA/QC prior to 2009. Although minor changes to the program were made in response to anecdotal feedback during this development time, our assessment indicates that those adjustments were not sufficient and demonstrates the need to incorporate quantitative assessments early in citizen science project development.

We identify two limitations to our QA/QC study design: location errors and look-alike errors. In the case of location errors, volunteers may have been correctly identifying target species while misreporting their geographic location. For example, we have anecdotal evidence that volunteer observations recorded during evening naturalist programs at one of the huts included a misleading location description provided by

Table 3
Summary of the chi-square test results for the QA/QC analyses of Mountain Watch volunteer-collected data. A series of chi-square analyses explored the relationship between the rate of species identification (accuracy), a volunteer's self-assessed Certainty of Identification, and a plant's phenophase. *P* values are included for significant results.

Alpine plant species	Certainty of ID & accuracy	Flowering & certainty of ID	Flowering & accuracy
Vaccinium uliginosum	No relationship	Significant ($p = 0.005$)	No relationship
Rhododendron groenlandicum	No relationship	Significant ($p = 0.043$)	Significant ($p = 0.023$)
Diapensia lapponica	No relationship	Significant ($p = 0.045$)	Significant ($p = 0.014$)
Vaccinium vitis-idaea	No relationship	No relationship	No relationship
Geum peckii	No relationship	No relationship	No relationship
Carex bigelowii	No relationship	No relationship	No relationship
Across all species	No relationship	Significant ($p = 0.03$)	Significant ($p = 0.002$)

the naturalist. However, since our assessment was restricted to volunteer-collected data associated with precise geographic location descriptions, which were often detailed and specific, we were comfortable assuming that the naturalist program described above was an outlier.

During our assessment we compiled a list of look-alikes for each target species and noted that the target and look-alike species often grew together at our survey sites (Table 1, see also Appendix S1: Fig S1). We realize that look-alike errors could potentially bias our assessment: if target species and look-alike species grew together, the volunteer's observation would be in agreement with the survey when they may have actually been observing a misidentified look-alike species. Because of this, the target species with higher relative abundances in the survey are more likely to have inflated correct identification rates. The advent of smartphones and citizen science apps might alleviate both location errors and lookalike errors as observations can be associated with GPS coordinates and attached photographs of species *in situ* allow for validation of uncertain identifications (Crall et al., 2010; Newman et al., 2012). However, these were nascent technologies in 2005, and are still inaccessible to programs that occur in remote areas of the backcountry and/or have limited resources (Dickinson et al., 2012; Wiggins, 2013).

Despite the possible limitations associated with our survey methods, we are confident in the quality of presence/absence data captured. During the development of our assessment, we informally shadowed volunteers and naturalists as they recorded Mountain Watch data. While this allowed us to compile a look-alike species list, it was not feasible to assess individual volunteers at a larger scale and the survey provided a comprehensive, quantitative method of assessment with a small investment in time and resources. In addition, the ubiquity of professional re-surveys in the scientific literature (Lewandowski and Specht, 2015) reinforces our survey as a useful method of assessment.

The results of our QA/QC analysis revealed shortcomings in the original Mountain Watch data collection model. We found that open-ended data sheets return precise location descriptions at a rate under 70%, leaving nearly a third of the Mountain Watch data collected in the Presidential Range ineligible for our QA/QC analysis. The imprecise location descriptions could impact the future utility of Mountain Watch data, for example, resulting in a more coarse GIS analysis. The majority of hikers that we solicit to volunteer for Mountain Watch likely have limited identification skills to accurately monitor alpine plants, and flowering phenophases did not seem to improve their abilities.

While flowers increased the volunteers' confidence in their certainty of identification, in fact alpine plants in bloom were more likely to be misidentified in the case of *R. groenlandicum* and *D. lapponica*. We expected that flowering phenophases would aid identification skills. This is especially puzzling as the white inflorescences of *R. groenlandicum* and *D. lapponica* offer a contrast from their purple- and pink- flowered look-alikes (see Appendix S1: Fig S1 for photographs, *R. groenlandicum:* rhodora, bog laurel, sheep's laurel; *D. lapponica:* alpine azalea, lapland rosebay). It is possible that location errors are contributing to this relationship between flowering and reduced identification rates, but we assume the general hiking population is more likely to be adept at navigation (i.e. geographical location) and inept at botany (i.e. species identification), rather than the reverse. However, since our location surveys consisted of a 10 m radius, it is possible that volunteers correctly identifying a species located outside of this survey area would be considered incorrect in our results.

In the course of this QA/QC analysis we uncovered an apparent bias toward *D. lapponica*. The alpine zone is an unfamiliar habitat for most citizen scientists, but *D. lapponica* seems to be well-known among visitors to the AMC's backcountry huts: its photos decorate hut walls and brochures. At 16 of the 19 sites, *D. lapponica* was the most recorded Mountain Watch species even though at five of those locations, *D. lapponica* was not found in *any* abundance during our survey. In these five sites, volunteers were reporting the phenophases for some other plant while under the impression that they were monitoring *D. lapponica.* Among all observations in the Presidential Range (including non-precise locations), *D. lapponica* accounted for almost one-fourth (441) of the 1775 observations. Perhaps this *D. lapponica* fervor is equivalent to the phenomenon other programs have reported of volunteer biases toward charismatic bird species (Lepczyk, 2005), rare species (Lewandowski and Specht, 2015), and unique or large trees (Galloway et al., 2006).

After this QA/QC analysis, the AMC amended its data collection model. The process of assessing the Mountain Watch data led to discussions within the organization about the goals, priorities, and utility of citizen science in climate change research. Many citizen science programs without extensive training programs rely on volunteers who already have experience with identification as life-long birders (Sullivan et al., 2009) or amateur botanists (Beaubien and Hamann, 2011; Lawrence, 2009; Mayer, 2010), while the majority of the Mountain Watch audience is composed of hikers staying at an AMC hut who may not have a strong interest and knowledge in the flowering plants they are observing, and were not available for more extensive training. Mountain Watch assumed that adding a self-assessed “certainty of ID” metric would sort the volunteer-collected observations into “good” and “bad” plant identifications. However, our QA/QC survey revealed that this was not the case; other citizen science programs must similarly test their assumptions of their volunteers' abilities. The citizen science datasheets were modified to direct volunteers to specific, permanent locations where a list of target species known to present was provided (Wiggins, 2013). However, due to lack of direct funding necessary to maintain a robust QA/QC of the general hiking audience data, the program has been scaled back over time, and is now limited to trained seasonal employees, research staff, and a select group of well-trained volunteers.

The current, smaller cohort of Mountain Watch observers has had more success with the more constrained and less subjective tasks of recording data for lists of target species at specific, permanent locations (Kimball et al., 2014). Similarly, other citizen science programs have found that volunteers were more accurate with concrete tasks, for example, reporting measurements instead of classifications (Brandon et al., 2003; Galloway et al., 2006; Lovell et al., 2009). While Mountain Watch is specific to a place and unique pool of hiker-volunteers, it exhibits qualities common to many small, local citizen science programs. It is much smaller than the national phenology networks (Mayer, 2010), the alpine zone is much less familiar than “backyard monitoring” schemes like eBird (Sullivan et al., 2009), and the AMC is not able to provide hours- or day-long training sessions for its volunteers (Bois et al., 2011; Jordan et al., 2009; Kremen et al., 2011; Lovell et al., 2009). Explorations of data quality in big-name, well-funded citizen science programs that draw on large pools of volunteers are important, but most programs must attempt to assess their volunteer-collected data at a smaller scale and with fewer available resources (Engel and Voshell, 2002). This case study underscores the repeated call for well-structured data collection models and training that matches the requested volunteer skillset with a clear assessment of an imperfect model (Bonney et al., 2009; Dickinson et al., 2010; Lovell et al., 2009). Other citizen science programs must follow the lead of the AMC, take the time to study their own data, and then make adjustments — in training, data collection methods, or goals — in order to produce data of a quality consistent with their scientific intentions.

Acknowledgements

Financial support for Mountain Watch was provided by the Office of Oceanic and Atmospheric Research at NOAA under grants NA06OAR460029, NA06OAR4600180 and NA09OAR4590208, and grants from EnTrust and The Waterman Fund. We gratefully acknowledge the many Mountain Watch volunteers and the AMC naturalists and research interns who contributed to the field data collected, especially our dedicated phenology interns Katherine Siner, Ben Brush, and Matt Labonville. CMM thanks her M.S. committee at

University of Vermont Jeff Hughes, Deane Wang, Jennifer Pontius, and Dave Barrington for their mentoring on this project. Additional thanks to Ken Kimball for his support and suggestions on this manuscript, and Margaret Kosmala and two anonymous reviewers whose suggestions improved the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.biocon.2016.07.027.

References

Beaubien, E.G., Hamann, A., 2011. Plant phenology networks of citizen scientists: recommendations from two decades of experience in Canada. Int. J. Biometeorol. 55, 833–841. http://dx.doi.org/10.1007/s00484-011-0457-y.

Bois, S.T., Silander, J.A., Mehrhoff, L.J., 2011. Invasive plant atlas of New England: the role of citizens in the science of invasive alien species detection. Bioscience 61, 763–770. http://dx.doi.org/10.1525/bio.2011.61.10.6.

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, 977–984. http://dx.doi.org/10.1525/bio.2009.59.11.9.

Bonter, D.N., Cooper, C.B., 2012. Data validation in citizen science: a case study from project FeederWatch. Front. Ecol. Environ. 10, 305–307. http://dx.doi.org/10.1890/110273.

Brandon, A., Spyreas, G., Molano-Flores, B., 2003. Can volunteers provide reliable data for forest vegetation surveys? Nat. Areas.

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, 5, e106508. http://dx.doi.org/10.1371/journal.pone.0106508.

Crall, A., Newman, G., Jarnevich, C., Stohlgren, T., Waller, D., Graham, J., 2010. Improving and integrating data on invasive species collected by citizen scientists. Biol. Invasions 1–10 http://dx.doi.org/10.1007/s10530–010–9740-9.

Delaney, D.G., Sperling, C.D., Adams, C.S., Leung, B., 2007. Marine invasive species: validation of citizen science and implications for national monitoring networks. Biol. Invasions 10, 117–128. http://dx.doi.org/10.1007/s10530-007-9114-0.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T., Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. Front. Ecol. Environ. 10, 291–297. http://dx.doi.org/10.1890/110236.

Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: challenges and benefits. Annu. Rev. Ecol. Evol. Syst. 41, 149–172. http://dx.doi.org/10.1146/annurev-ecolsys-102209-144636.

Engel, S.R., Voshell, J.R., 2002. Volunteer biological monitoring: can it accurately assess the ecological condition of streams? Am. Entomol. 48, 164–177. http://dx.doi.org/10.1093/ae/48.3.164.

Galloway, A.W.E., Tudor, M.T., Haegen, W.M.V., 2006. The reliability of citizen science: a case study of Oregon white oak stand surveys. Wildl. Soc. Bull. 34, 1425–1429. http://dx.doi.org/10.2193/0091-7648(2006)34[1425:TROCSA]2.0.CO;2.

Inouye, D.W., 2008. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. Ecology 89, 353–362. http://dx.doi.org/10.1890/06-2128.1.

Jordan, R., Singer, F., Vaughan, J., Berkowitz, A., 2009. What should every citizen know about ecology? Front. Ecol. Environ. 7, 495–500. http://dx.doi.org/10.1890/070113.

Kimball, K.D., Davis, M.L., Weihrauch, D.M., Murray, G.L.D., Rancourt, K., 2014. Limited alpine climatic warming and modeled phenology advancement for three alpine species in the Northeast United States. Am. J. Bot. 101, 1437–1446. http://dx.doi.org/10.3732/ajb.1400214.

Kimball, K.D., Weihrauch, D.M., 2000. Alpine vegetation communities and the alpine-treeline ecotone boundary in New England as biomonitors for climate change. USDA Forest Service Proceedings.

Kremen, C., Ullman, K.S., Thorp, R.W., 2011. Evaluating the quality of citizen-scientist data on pollinator communities. Conserv. Biol. 25, 607–617. http://dx.doi.org/10.1111/j.1523-1739.2011.01657.x.

Lawrence, A., 2009. The first cuckoo in winter: phenology, recording, credibility and meaning in Britain. Glob. Environ. Chang. 19, 173–179. http://dx.doi.org/10.1016/j.gloenvcha.2009.01.006.

Lepczyk, C.A., 2005. Integrating published data and citizen science to describe bird diversity across a landscape. J. Appl. Ecol. 42, 672–677. http://dx.doi.org/10.1111/j.1365-2664.2005.01059.x.

Lewandowski, E., Specht, H., 2015. Influence of volunteer and project characteristics on data quality of biological surveys. Conserv. Biol. 29, 713–723. http://dx.doi.org/10.1111/cobi.12481.

Lovell, S., Hamer, M., Slotow, R., Herbert, D., 2009. An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. Biodivers. Conserv. 18, 3295–3307. http://dx.doi.org/10.1007/s10531-009-9642-2.

Mayer, A., 2010. Phenology and citizen science volunteers have documented seasonal events for more than a century, and scientific studies are benefiting from the data. Bioscience 60, 172–175. http://dx.doi.org/10.1525/bio.2010.60.3.3.

Miller-Rushing, A., Primack, R., Bonney, R., 2012. The history of public participation in ecological research. Front. Ecol. Environ. 10, 285–290. http://dx.doi.org/10.1890/110278.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., Crowston, K., 2012. The future of citizen science: emerging technologies and shifting paradigms. Front. Ecol. Environ. 10, 298–304. http://dx.doi.org/10.1890/110294.

Pauli, H., Gottfried, M., Grabherr, G., 2014. Effects of climate change on the alpine and nival vegetation of the Alps. J. Mountain Ecol. 7.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142, 2282–2292. http://dx.doi.org/10.1016/j.biocon.2009.05.006.

Wiggins, A., 2013. Free as in Puppies: Compensating for Ict Constraints in Citizen Science, the 2013 Conference. ACM, New York, New York, USA http://dx.doi.org/10.1145/2441776.2441942.