Manuscript submitted to: Volume 1, Issue 1, 38-44.

AIMS Environmental Science

DOI: 10.3934/environsci.2014.1.38

Received date 19 December 2013, Accepted date 13 February 2014, Published date 21 February 2014

Research Article

Understanding plant invasions: An example of working with citizen scientists to collect environmental data

Rebecca Jordan*, Wesley R. Brooks, Amanda Sorensen, Joan Ehrenfeld (deceased)

Program in Science Learning, Departments of Human Ecology and Ecology, Evolution, and Natural Resources, Rutgers University, New Jersey, USA

*** Correspondence:** E-mail: Rebecca.jordan@rutgers.edu; Tel: +1-848/932-9164.

Abstract: Citizen science programs are useful tools for collecting important environmental science data. To ensure data quality, however, it must be shown that data collected by volunteers can produce reliable results. We engaged 143 volunteers over four years to map and estimate abundance of invasive plants in New York and New Jersey parklands. We found that off trail abundance of only a few of our targeted invasive species were positively correlated with on trail abundance. Our results support that citizen science programs can be a useful and sometimes a much needed addition to environmental science protocols.

Key words: citizen science; environmental monitoring; invasive plants

1. Introduction

The ability of environmental scientists to collect and analyze data across large spatial scales is clearly limited, especially given rapid environmental change. A trained workforce of volunteers can be a means to further the reach of environmental scientists. It is clear that there is potential for engaging the public as a means for scientists and resource managers to expand their capacity to collect data [1,2].

Citizen science programs are defined as projects that enable public participation in authentic research. Bonney et al. describe three models of citizen science [3]. In the contributory model, scientists often initiate the project and capitalize on an expanded volunteer workforce to generate data that are often published in the primary, peer-reviewed literature. In the collaborative model, participants have the opportunity to participate in data analysis and interpretation, and in the co-created model, participants have the opportunity to engage as scientists would and develop ownership over the project [3], and expanded in [4]. Further, while scientists and resource professionals may engage volunteers out of a desire to grow a dataset beyond that in which they are capable of collecting themselves, they may also be interested in broader impacts. Participants have been shown to increase their understanding on the underlying science behind environmental issues [5,6,7]. Additionally when considering environmental issues, citizen science initiatives can successfully promote civic engagement [8]. Citizen groups have also played a role in shaping environmental policy [9,10] and may translate into increased socio-ecological resilience [11].

Given these potential benefits of citizen science programs, it is worthwhile for environmental scientists to ask whether the data are reliable and accurate. Certainly concern over data quality exists and not all tasks are necessarily suitable for volunteers [12]. Many studies, however, are able to show accuracy through data validation measures (e.g., plant related citizen science projects [11,13]). Perhaps a more important question is to consider the extent to which citizen science can improve environmental scientific research.

In this paper, we compare two data sets: one collected by experts and one where expert data are combined with volunteers. First, we describe how we were able to collect data to test the hypothesis: Weedy plant invasion on trails is associated with invasion off trails. Next, we use this context to highlight critical issues in the use of non-expert volunteers in environmental monitoring citizen science.

2. Project Context

In native ecosystems, invasive plants can pose a serious threat to biodiversity. The presence of many invasive plant species in the United States is known, however, knowledge of their distribution in the northeast in relatively scant [14]. Natural resource managers and policy-makers are limited in their ability to respond to issues concerning invasive plants, as well as establish policies and regulations intended to limit further spread of invasives because of this deficiency. This project set out to learn more about invasive plants in forested parkland in the Highlands physiographic province along the border between New Jersey and New York. The areas in this study include Harriman and Near Mountain State Parks in New York, a series of parks in the Ramapo Mountains in New Jersey, and other nearby protected lands. This study area is within the Highland geomorphic province (based on Precambrian granitic gneisses and schists), and are included within the Hudson Highlands ecozone, a largely hardwood region recognized for its high biodiversity, rare species habitats, and conservation value.

Plant distribution on-trail and off-trail were compared to test the hypothesis that trails were positively associated with invasive plant species distribution. Trained graduate students completed off-trail transects. For logistical purposes, using volunteers for on-trail (versus off-trail) sampling was preferred by park managers and allowed for us to test hypotheses about volunteer participation in this project.

3. Methods

3.1. Volunteers recruitment and training

In February of 2006, 2007, 2008, and 2009, volunteers from the New York-New Jersey Trail Conference (NYNJTC) were recruited (58, 35, 26, and 24 individuals respectively). We chose to work with the NYNJTC because these individuals were already hiking in the region of interest. The volunteers were recruited via email sent to the entire membership of about 10,000 individuals and 100 clubs, and no material incentives were offered. Volunteers were accepted if they could undertake the hiking and attend the training sessions. Those volunteers that wished to participate in subsequent years from 2006 were allowed to, but their data was excluded because of the likelihood their abilities increased.

The authors led one all-day training session that volunteers attended in early June, followed by a 'debriefing' session after volunteers collected their data in early July. Participants were given background information about the ecology and impacts of invasive species at the training session. Volunteers also received hands-on species identification training for a target list of invasive plants (22 species in 2006, but reduced to 12 in 2007, and 13 in 2008 and 2009, therefore we restrict our analysis to 13 species. Table 1). Volunteers were instructed to scan in a stratified manner from canopy to ground for the categorized target plants in four groups: trees, shrubs, vines, and herbs. Volunteers were provided with a field ID guide created for the project and were trained in field-based semi-quantitative data collection protocol.

Table 1. Targeted species name and abundance for the citizen science program. More information about these data can be found in Jordan et al. 2012.

Common Name	Species	No. of validated sightings
Japanese barberry	*Berberis thunbergii*	343
Japanese stiltgrass	*Microstegium vimineum*	218
Multiflora rose	*Rosa multiflora*	85
Garlic mustard	*Alliaria petiolata*	81
Oriental bittersweet	*Celastrus orbiculatus*	79
Wineberry	*Rubus phoenicolasius*	48
Tree of heaven	*Ailanthus altissima*	47
Japanese angelic tree	*Aralia elata*	37
Burning bush	*Euonymus alata*	30
Norway maple	*Acer platanoides*	24
Mile a minute	*Persicaria perfoliata*	24
Swallow wort	*Cynachum louiseae*	14
Japanese honeysuckle	*Lonicera japonica*	6

3.2. Invasive plant collection

The following protocol was created by JE for previous studies and was replicated here for comparison to previous work. Volunteers collected data in pairs and each pair was assigned a 2 miles (~ 3.2 km) length of trail to survey. The volunteer pairs hiked an assigned stretch of the trail, stopping every 0.1 mile (0.16 km) to record presence and abundance of target species as well as collect samples when they first encountered a target species. Volunteers surveyed approximated 150 miles (~ 241 km) of trails. Volunteers were instructed not to leave the trail and scan for target species in two zones on the right and left sides of the trail at each stop. This trailside zone was 6 feet (1.83 m) by 20 feet (6.10 m) estimated by volunteers in 2006, but delineated with rope in the following three years. Relative abundances were estimated as 'few', 'some', or 'many' by volunteers. Volunteers were provided with bags and labeling stickers for collecting samples on trail, and a plant press for sample preservation. Volunteers were trained in sample preservation and samples were returned to the authors in the press.

3.3. *Reliability of volunteer data*

Multiple methods were used to assess the accuracy of volunteer-generated data: 1) Pressed samples to test species recognition, 2) trail-point validation to measure volunteer accuracy on trails, 3) common trail data collection to measure the repeatability of volunteer-generated data. Volunteer accuracy is discussed at length in [11]; below we summarize our methods of doing so.

1) Pressed Samples. Authors assessed accuracy of specimen reporting by volunteers.

2) Trail Point Validation. Data collected by volunteers at a subset of points were compared with data collected by three pairs of specially trained staff, henceforth referred to as 'validators', and used to assess volunteer ability to detect target plants and estimate abundance along trails. Validators assisted with training and have well-developed and specialized skill for identifying the target plants. High repeatability in validator-generated data, coupled with much higher variability in volunteer data gives us confidence that we can use validator-generated data as an acceptable standard by which to measure accuracy. In 2006, 30% of volunteer collected data were validated, and roughly 50% were validated in 2007, 2008, and 2009.

3) Common Validation Trail. To examine repeatability of volunteer-produced data we had all volunteers collect data along 1 mile (1.6 km) 'validation' trails that were explicitly marked at 11 points each. Separate trails were used for NJ and NY volunteers in 2006 and 2008, except a single trail in 2007 and 2009, where the most experienced validator team surveyed the trails, and we used that team's data as the standard. An accuracy score was calculated for each sampling volunteer pair on both the experimental trail and validation trail.

Our volunteer data produced accurate results when inspecting plant pressed specimens, N = 70, only two species were occasionally misidentified [11]. Given the high level of sites without plants, volunteers were 97.3% accurate on the trail point validation, and volunteers were only about 15% less accurate than the professionals on the validation trail [11]. Once we were confident that we had a reliable dataset, we generated a level of invasion metric by which on trail and off trail data could be compared.

This invasion metric was computed by using the mean abundance for each species along a particular trail section that had at least 11 observations of that plant species (i.e., at each point where the volunteers stopped to collect data). Off trail data were collected by a single highly trained expert along a transect perpendicular to the trail originating 25 m beyond trail locations. For this, there were four locations along that transect were averaged in terms of species abundance. These data, when in sufficient numbers, were compared using a Pearson's correlation statistic (Minitab 16, Minitab, Inc.) by plant species. This analysis was conducted on the combined expert (N = 30) and citizen collected (N = 41) data set and the expert data set alone. That is to say, the on trail expert only and the on trail combined data were each separately compared with the expert off trail data in the correlation analysis.

4. Results

Of the plant species surveyed, only four were present in high enough frequency to compare data on and off trail. Some species simply lacked both positive on and off trail observations and others lacked positive off trail observations or positive on trail observations, but not both. All other correlations were not able to be computed because of insufficient sample size. In Table 2, we present data for each data set (total data and expert only). Note that these were among the most common species found in the study site (Table 1). While the correlations are near significance or are significant at the (a = 0.05) level, they are low (i.e., r ranging from 0.219–0.318) in the total dataset with roughly 50–60% of the variation in the data being explained by the relationship between on trail and off trail data. Interesting, however, is the variation in results in the expert only data set. Recall the expert data set is smaller (N = 30). Further inspection into both data sets indicate that there are still a great number of no sightings on and off trail, which could lend itself to results that are due to chance alone.

Table 2. Below are the four species for which a sufficient number of observations were made to calculate Pearson's r. For the two on trail data sets calculated (total data, which equals high performing volunteers and experts, and expert only), r is given as a correlation with off trail data. Associated p-values are in parentheses. In bold are significant correlations. For the four species listed below, there is some evidence to suggest off trail presence/absence is associated with on trail presence/absence.

Plant	*Total Data*	*Expert Only*
Japanese Barberry	**0.318 (p = 0.009)**	0.189 (p = 0.344)
Oriental Bittersweet	0.219 (p = 0.095)	**0.476 (p = 0.014)**
Japanese Angelic Tree	0.247 (p = 0.066)	**
Wineberry	**0.272 (p = 0.042)**	**0.466 (p = 0.019)**

** Not enough observations made

5. Discussion

In this paper, we tested the hypothesis that trails correlated with an increase in plant invasion in off trail locations using two data sets, one with expert data only and the other including citizen volunteers. Only with the inclusion of volunteer data, were we confident that we had a sufficient sample size to test our ideas. It appears that off trail abundance of only a few targeted invasive species are positively correlated with on trail abundance.

We found that some species lacked only sufficient positive off trail observations. This could be caused through a few mechanisms: 1) something about trails makes them a suitable habitat for being invaded by this species but not off trail, or 2) trails may be allowing invasives into what would otherwise be uninvaded forest (from that individual species' perspective) and we may be catching the invasion relatively early (thus we would expect spread to off trail sites later). Others as mentioned above had enough positive data on and off trail and of these, all had a significant or almost significant relationships between on and off trail level of invasion. Several hypotheses remain to be tested and different species may not be affected by similar mechanisms. Hypotheses include 1) invaders may not discriminate between on and off trail (e.g., Japanese Barberry because of bird dispersal) or 2) trails may be allowing invasives into what would otherwise be uninvaded forest (from that individual species' perspective) and we may be catching the invasion relatively late (thus already see spread to off trail sites).

More importantly, however is that our data support the notion that volunteer participation can enhance data generated by scientists alone. The combined trail length surveyed was clearly beyond that which project personnel could have completed alone. When looking at our data, one may be inclined to disregard the volunteer included data because of the different results. However, when the smaller data set was inspected, it was clear that insufficient observations could result in spurious conclusions. When the greater data set is inspected, there are more data points, which are evenly distributed. Plus, we were able to use biological information regarding habitat requirements to help us consider which data may be more accurate. The additional data were necessary for us to test our hypothesis regarding on and off trail species abundance.

The volunteer monitoring efforts allowed us to determine that certain species, though abundant, are simply not present in the understory off trails (e.g., Japanese Stilt Grass, Multiflora Rose, Garlic Mustard). Volunteer effort has also helped generate new hypothesis regarding past land use, trail placement, and trailside disturbance regimes that may be associated with invasive plant species. Data collected from our project has enabled us to create better profiles of likely understory invaders in Northeastern US forests. Here are three examples: Japanese barberry, which was one of our most common invasive plants and had a significant correlation with off trail abundance is likely spread by bird consumption which does not follow trails. Wineberry, is often spread by birds and mammals, the latter of which, use trails and could be source of spread. The latter of the two prefer at least low light which can help to reduce likelihood in dense cover. Barberry, however, can thrive in a number of habitats.

An additional 60% of trails were able to be surveyed by volunteers. The increase of 41 trails would be equivalent to 41 days of labor. At 6 hours per day and the going rate of $22 per hour for ecology field assistance here in NJ, $5412 dollars would be needed to cover this expense. Whereas a single day of scientist time would cost only $375 and this would be devoted to volunteer training. Finding volunteers was relatively easy, because we partnered with an organization in which the volunteers were already engaged in the target activity. In addition to gathering the necessary data, volunteers were able to enjoy benefits in learning and becoming more environmentally aware [11].

This study can provide insight for future monitoring programs. A number of tradeoffs exist when engaging volunteers in projects like ours. A considerable amount of thought and effort is necessary to ensure data validation measures. The three measures described above required additional thought and personnel hours. Certain tasks, such as hiking off trail were simply not amenable to volunteers with only minimal training, again requiring experts to also hike into some of the regions. Finally, volunteers were better at identifying certain plants than others, and are likely to be less useful when considering rarer plants (see [11] for data and further discussion).

Other projects with more specific and detailed data collection protocols may create greater data uncertainty using volunteers. This uncertainty may lead to considerably more effort on the part of the scientists and volunteer trainers. It may also be difficult to gather volunteers without the help of an organization or using tasks that are commonly conducted by public audiences (e.g., hiking). When conditions are right, however, environmental scientists may find that volunteers can provide additional and necessary data with minimal costs. And in our case, a relatively easy training protocol that has been validated can be used as part of greater monitoring efforts. Such an outcome is especially desirable in current times of rapid global change.

Acknowledgements

Funding was made possible through the USDA CSREES # 05-2221 and all work was conducted in accordance to Institutional Review Board policy. We thank Ed Goodell and the people of the NY-NJ Trail Conference. Additionally, we thank Kristen Ross, David Mellor, and Edwin McGowan. We give a special thanks to our numerous volunteers. Finally, we dedicate the manuscript to Joan Ehrenfeld, a fine scholar, mentor, and most importantly, friend.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References:

1. Evans C, Abrams E, Reitsma R, et al. (2005) The neighborhood nestwatch program: participant outcomes of a citizen-science ecological research project. *Conserv Biol* 19: 589-594.
2. Brossard D, Lewenstein, B, Bonney R (2005) Scientific knowledge and attitude change: The impact of a citizen science project. *Inter J Sci* 27: 1099-1121.
3. Bonney R, Ballard H, Jordan R, et al. (2009) Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report Center for Advancement of Informal Science Education, Washington, DC, USA. Accessed online March 15, 2010: http://caise.insci.org/uploads/docs/PPSR%20report%20FINAL.pdf.
4. Shirk JL, Ballard HL, Wilderman CC, et al. (2012) Public participation in scientificresearch: a framework for deliberate design. *Ecol Soc* 17: 29.
5. Krasny M, Bonney R (2005) Environmental education through citizen science and participatory action research. In: Johnson EA, Mappin MJ editors, *Environmental Education or Advocacy: Perspectives of Ecology and Education in Environmental Education.* New York: Cambridge University Press, pp 292-319.
6. McCormick S, Brown P, Zavestoski S (2003) The personal is scientific, the scientific is political: the public paradigm of the environmental breast cancer movement. *Sociol Forum* 18: 545-576.
7. Jordan RC, Gray ST, Howe DV, et al. (2011) Conservation education and citizen science: using an invasive plant case study to highlight issues of practice. *Conserv Biol* doi:10.1111/j.1523-1739.
8. Weber EP (2000) A new vanguard for the environment: grass-roots ecosystem management as a new environmental movement. *Soc Nat Resour* 13: 237-259.
9. Dunlap R (1992). Trends in public opinion toward environmental issues: 1965-1990, In: Dunlap R, Mertig A editors, *American Environmentalism: The U.S.Environmental Movement, 1970-1990*. Philadelphia, Pennsylvania: Taylor and Francis, 89-116.
10. Beierle TC, Cayford J (2002) Democracy in practice: public participation in environmental decisions. Washington, DC: RFF.
11. Jordan RC, Brooks WR, Howe DV, et al. (2012) Evaluating the performance of volunteers in mapping plants in public conservation lands. *Environ Manage* 49: 425-434.
12. Darwall W, Dulvy N (1996) An evaluation of the suitability of nonspecialist volunteer researchers for coral reef fish surveys, Mafia Island, Tanzania—a case study. *Biol Conserv* 78: 223-31.
13. Crall AW, Newman GJ, Stohlgre TJ, et al. (2011) Assessing citizen science data quality: an invasive species case study. *Conserv Lett* 4: 433-442.
14. Luken JO (2003) Invasions of forests in the Eastern United States. In: Gilliam FS, Roberts MR editors, *The Herbaceous Layer in Forests of Eastern North America*. New York: Oxford University Press, pp 283-301.