REVIEW AND SYNTHESIS PAPER

Usability of Existing Volunteer Water Monitoring Data: What Can the Literature Tell Us?

Kelly Albus*, Ruthanne Thompson* and Forrest Mitchell†

For decades citizen science has been used in environmental monitoring, and perhaps most commonly in water quality monitoring, as a tool to supplement professional data. Hundreds of volunteer monitoring efforts have generated datasets that cover large geographic areas over multiple years, and these large-scale datasets have been shown to be especially valuable for monitoring changes over time. Although volunteer water monitoring programs continue to grow worldwide, research shows that many of the existing datasets are still underutilized due to concerns about the accuracy of volunteer-collected data. An increasing number of "comparison studies" have attempted to address quality concerns by comparing volunteer data to professional data to assess relative accuracy, and the majority have reported that volunteer data are of a quality comparable to professional data. Nearly all of these studies, however, focused on a small subset of volunteer program data or data collected under experimental controls, and as such the results may not be applicable to existing, large-scale datasets with unknown controls and high levels of variation. Through a comprehensive look at water quality comparison studies to date, this review reveals a need for additional studies that specifically address the quality of highly variable, large-scale volunteer datasets and ultimately serve as a framework by which decades of volunteer efforts already in existence across the country can be better utilized.

Introduction and Purpose

Citizen science projects have involved volunteers in monitoring the environment for decades (Silvertown 2009; Devictor et al. 2010; Kobori et al. 2016). In the US, water monitoring citizen science programs have been increasing in number for many years, with numerous examples of local, state, and federal agencies utilizing volunteers to meet extensive water quality monitoring needs (USEPA "Volunteer Water Monitoring" website; Canfield et al. 2002; Savan et al. 2003; McKinley et al. 2016; Safford and Peters 2017). For example, the National Water Quality Monitoring Council (NWQMC) website includes more than 350 volunteer monitoring groups registered across the country as of 2018. Many of these groups have been collecting water quality sampling data such as dissolved oxygen (DO), pH, and conductivity for decades, and have generated continuous, long-term databases that often cover large regions, watersheds, or states. Long-term water monitoring datasets such as these can be useful in providing baseline information for streams and lakes for flagging further sampling needs, and/or providing evidence for impacts of land-use change or climate change

(Nicholson et al. 2002; Loperfido et al. 2010; Hoyer et al. 2012; Storey et al. 2016). However, although many potentially valuable datasets already exist, they are rarely used. This lack of use often results from concerns about the accuracy of volunteer-collected data (Canfield et al. 2002; Nicholson et al. 2002; Hoyer et al. 2012; Stepenuck 2013; Barrows et al. 2016; Safford and Peters 2017).

For existing datasets to be used, evidence suggests that professionals and scientists must be more confident in the accuracy of the volunteer-generated data (Burgess et al. 2016; Kosmala et al. 2016). To this end, a growing number of studies have attempted to determine the accuracy of volunteer-collected data by comparing them to similar data collected by professionals. By holding the professional data as the standard for accuracy, volunteer data are evaluated by relative variation. These studies have sought to provide a quantitative way to address data-quality concerns and have created a baseline body of knowledge that could allow potentially valuable datasets to be better utilized.

The purpose of this review is to summarize research from the burgeoning field of comparison studies related to volunteer water-quality monitoring efforts. Our goal is to provide a comprehensive look at what the current literature says about the accuracy of volunteer water quality data, along with data-quality issues that have not yet been addressed. Numerous volunteer monitoring

* University of North Texas, US

† Texas A&M AgriLife Research, US

Corresponding author: Ruthanne Thompson (rudi@unt.edu)

programs with diverse goals and methods are in existence, and no review could fully capture what accuracy, or more importantly "success," means for each individual program or for citizen science as a whole. However, by focusing on literature regarding volunteer stream water-quality sampling for common parameters such as DO and pH, and especially on data from long-running, large-scale programs, this review offers conclusions and areas for future development for this widely used and rapidly expanding segment of the citizen science population.

We first examine why citizen science has become more prevalent in environmental monitoring and in water monitoring specifically. We explicitly focus on the types of data collected by volunteers, if and how these data are used by scientists or professionals, and the possible reasons behind potential patterns. We then summarize water quality comparison studies from the literature, ultimately revealing those studies that address accuracy concerns which are more directly related to the use of existing data that many long-running volunteer monitor programs already have collected. Based on these findings, we draw conclusions about the existing literature and make recommendations for future research to address gaps in the field.

Citizen Science in Environmental Monitoring

Citizen science has become widely used in environmental monitoring, in part because of evidence that increased public involvement is beneficial for the participants, for scientists, and for the scientific field as a whole (Bonney et al. 2009a; Bonney et al. 2009b; Raddick et al. 2010; Jordan et al. 2011). Volunteer involvement in environmental monitoring is considered especially valuable as a method for collecting large datasets over a significant geographic area and for long periods of time. In large part, recent technological advancements have made projects involving the public more widely accessible and more efficient (Silvertown 2009; Newman et al. 2012; Valdes et al. 2012). Previous limitations inherent to using untrained volunteers in research included data access, standardization, and reliability. These have been countered by advancements in technology that are able to increase the access and interoperability of data, as well as the accuracy and verifiability of volunteer-collected data (Dickinson et al. 2010; Raddick et al. 2010; Gonsamo and D'Odorico 2014; Kobori et al. 2015; Paul et al. 2018).

In many cases, volunteer monitoring may now be considered more efficient for covering large areas quickly or continuously, and at less cost, then using professional scientists alone (Devictor et al. 2010; Hochachka et al. 2012). Citizen science projects have been used to monitor time-sensitive, large-scale biodiversity shifts, such as locating and removing invasive species, documenting evidence of climate change, or setting up continuous monitoring in threatened areas such as coral reefs and intertidal pools (Galloway et al. 2006; Delaney et al. 2008; Crabbe 2011; Cox et al. 2012; Kress et al. 2018). Many have argued that global conservation science would benefit from increased use of citizen science due to its unique ability to provide data that are broad-scale in scope, but with a fine grain resolution that only a lot of eyes on the ground can provide.

Beyond data collection, the rapid growth of environmental citizen science stems from the many benefits that can result from involving communities in research and management of natural resources. McKinley et al. (2016) conducted a review that involved citizen scientist experts and practitioners from multiple state and federal agencies and academic institutions across the US to better understand how citizen science was being used. This review determined that citizen science is already a substantial contributor to environmental science and natural resource management, providing both information and public engagement. The authors also argue that citizen science participants who are impacted by local natural resources in their daily lives can help refine research questions by making them more relevant to local needs and more useful to managers and local communities (McKinley et al. 2016). This ultimately helps researchers develop a more holistic perspective, which takes into account the connections between humans and their environment.

Citizen Science Water Monitoring Programs

Some of the most successful examples of long-term environmental monitoring are the volunteer water quality monitoring programs that have been collecting water quality data on streams and lakes for decades. Many of these programs have provided continuous environmental monitoring datasets that cover large areas or watersheds, and can inform ongoing management efforts by providing data for trend analyses or serving as early warning systems. In the US, there are numerous examples of state-supported volunteer monitoring programs such as the Colorado River Watch Network, Missouri Stream Team, Georgia Adopt-a-Stream, Alabama Water Watch, Texas Stream Team, Florida LAKEWATCH, and IOWATER (Canfield et al. 2002; Loperfido et al. 2010; Deutsch and Ruiz-Cordova 2015; Safford and Peters 2017). Programs of this type have been training volunteers to monitor alongside professional agencies and usually involve some form of volunteer certification process closely aligned with quality assurance that meets national standards (Stepenuck and Genskow 2018).

The continued growth of volunteer water monitoring programs is also tied to the actual or perceived impacts on the citizen scientists themselves. The global and shared nature of water management necessarily involves numerous stakeholders and members of the public in policy and decision making, and many studies have noted the benefits of involving local citizens in the monitoring of their own water resources (Penrose and Call 1995; Canfield et al. 2002; Nicholson et al. 2002; Loperfido et al. 2010; Hoyer et al. 2012). In the US, the White House office of Science and Technology Policy (OSTP) issued in 2015 a memorandum to the Heads of Executive Departments and Agencies titled "Addressing Societal and Scientific Challenges through Citizen Science and Crowdsourcing" (CS Inventory website). The idea behind this memorandum was to:

> "ensure future use of citizen science and crowdsourcing, and direct agencies to catalogue agency-specific citizen science and crowdsourcing

projects on a government-wide online database and website – to be developed by the General Services Administration" (CS Inventory website, last accessed 5/23/2018).

Although citizen scientists have been monitoring their water for many years, collecting data on chemical and biological parameters, flow, trash accumulation, and species diversity, the current uses of these data are limited. In a review of studies that assessed data quality in citizen science, Kosmala et al. (2016) reported that, despite the abundance of information that volunteers have generated and the resulting scientific discoveries, citizen science data are still met with skepticism. In their analysis, they reported that diverse types of datasets produced by volunteers can have reliably high quality that is on par with data produced by professionals, but that individual accuracy varies depending on program structure. Kosmala et al. (2016: 551) concluded that "each citizen science dataset should therefore be judged individually, according to project design and application, and not assumed to be substandard simply because volunteers generated it."

Stepenuck (2013) conducted a literature review on volunteer water monitoring programs that have appeared in peer-reviewed literature using the terms "volunteer monitoring," "citizen science," "participatory monitoring," "community-based monitoring," "locally-based monitoring," "public participation in scientific research," "community-based collaborative monitoring," and "environmental collaborative monitoring." From this review, Stepenuck concluded "there seemed to be a tendency for researchers to make conclusive statements about outcomes of volunteer water monitoring without providing data to support their remarks" and argues that these unsupported statements "reinforce doubts about the credibility of volunteer monitoring." And when it comes to including volunteer monitoring data in peer-reviewed literature, a survey involving 423 scientists conducted by Burgess et al. (2016) revealed that one of the main blockades to citizen science projects ending up in peer-reviewed journals are the perceptions of the scientists themselves. Although the proportion of scientists who reported that they would allow trained non-experts to collect their data was 79%, the proportion that has actually published a peer-reviewed paper using citizen science data was 35%.

These findings point to a professional bias against volunteer-collected data, a bias that several programs have certainly overcome, but one that still persists in citizen science. Many volunteer monitoring datasets are not fully utilized, especially in their entirety, by professionals and scientists (Penrose and Call 1995; Fore et al. 2001; Hoyer et al. 2012; Safford and Peters 2017). Perceptions about the accuracy of volunteer data, as well as applicability of the data, may be two of the main reasons for this lack of use (Dyer et al. 2014; Muenich et al. 2016). In the survey conducted by Burgess et al. (2016), scientists were asked what methods they required of a citizen science program to make the resulting data useful to them, and the top-ranked choices all related to data quality and included documentation of sample location, verifiability, and in-person training by an expert. Also according to the survey,

the perceptions that the scientists had about the quality of the data was the most important predictor of whether the data ended up in a publication. Specifically, Burgess et al. (2016) show that scientists' perceptions of data quality may be tied to the design of the program itself, including how the volunteers are trained to collect the data.

Note that this speaks only to the perceptions of the scientists, not to the actual quality of the data collected by volunteers. It is understandable for researchers to be skeptical of data that they did not collect. This may be especially true for volunteer monitoring data that are already in existence, and which are long-term, continuous, and cover a large geographic area. Although these dataset elements can be some of the most useful in terms of establishing a baseline for streams and waterbodies to inform management decisions (Nerbonne and Nelson 2004; MicKinley et al. 2016), they also contain large amounts of variability. Even with quality controls in place, large-scale datasets contain perhaps millions of samples from hundreds or thousands of sampling sites across varied ecoregions, collected by programs with personnel turnover and varied equipment maintenance schedules. These samples also reflect decades of natural and human-caused changes from diverse waterbodies with a range of baseline parameters, seasonal variations, and human impacts. Although these inherent variations are present in both volunteer and professional datasets of large size and scale, they nonetheless complicate any post-hoc analyses to assess the quality of such datasets. Therefore, existing volunteer datasets from long-running programs may be perceived as even less usable by scientists and professionals than smaller-scale, more controlled datasets. With numerous examples of long-term volunteer monitoring programs in existence, however, more information is needed on how the datasets that they have generated are being evaluated and utilized. This review seeks to provide a picture of how volunteer water quality datasets are evaluated for accuracy, and whether these methods also have been applied to large-scale volunteer datasets that are already in existence, especially in their entirety.

Water Monitoring Data Accuracy: Comparison Studies

A number of studies have attempted to address perceptions concerning volunteer accuracy by assessing the water monitoring data collected by volunteers in comparison to similar data collected by professionals (**Table 1**). This review first identifies the comparison studies from the

Table 1: Water Quality Data Comparison Studies. Breakdown of type and number of studies found (n = 15).

Description of Topic Addressed in Articles	Number of Related Articles Found
Water Quality Data Comparison Studies	15
1. Utilized Existing Volunteer Data	*6*
2. Large-scale Datasets (spatial, temporal)	*2*

literature which are directly related to water quality (15). Within that category, we then identify those that have utilized volunteer data already in existence prior to the study (6), and from those, which ones analyzed large-scale datasets that encompassed the program's full scope, either spatially or temporally (2).

To conduct this review, we first used online search sources such as Google Scholar™, EBSCO*host*®, *ERIC*, and Web of Science™, using the keywords "volunteer monitoring," "citizen science," "environmental monitoring," and "community monitoring." These searches returned articles from the 1970s to the present involving volunteer monitoring programs and came from multiple fields including environmental science, water resource management, biology, and geography. Only those studies in which the volunteers were credited and termed as such were considered. The results were further refined to include only studies in which evaluation of the monitoring data was the focus in regard to "accuracy," "variability," or "quality," by comparing volunteer-collected data to data collected by professionals or scientists. The filtered results were then reviewed and refined to include only studies that specifically dealt with water monitoring data of some form. Additionally, further resources were obtained through an email request via the USEPA volmonitor listserv to locate additional studies that were not generally accessible, or in which the above terms may not have been used in the description. This email list is supported by the USEPA and contains a large number of citizen science professionals and researchers.

Overall, this search located a total of 26 studies in which the accuracy or quality of volunteer water monitoring data were assessed and compared to similar data collected by a scientist, researcher, or professional, terms that we refer to interchangeably for the purposes of this study. Out of these studies, 11 volunteer and professional comparisons focused on indicator species such as birds, frogs, or macroinvertebrates and are not included in this review (Reynoldson et al. 1986; Penrose and Call 1995; Fore et al. 2001; Hoyer et al. 2001; Engel and Voshell 2002; Nerbonne and Vondracek 2003; Boudreau and Yan 2004; O'Leary et al. 2004; Gowan et al. 2007; Oscarson and Calhoun 2007; Moffett and Neale 2015).

A total of 15 studies focused specifically on water quality physical and chemical parameters (e.g., bacteria, DO, pH, temperature, conductivity,) and are included in this review (Obrecht et al. 1998; Au et al. 2000; Canfield et al. 2002; Nicholson et al. 2002; Loperfido et al. 2010; Sarnelle et al. 2010; Stepenuck et al. 2011; Hoyer et al. 2012; Coates 2013; Shelton 2013; Stepenuck 2013; Dyer et al. 2014; Muenich et al. 2016; Storey et al. 2016; Stafford and Peters 2017). Unlike indicator species data, these parameters are often monitored by professionals and volunteers using similar equipment and sampling protocols. They can be considered more straightforward than macroinvertebrate identification because there is no need for a biological index to relate the findings to stream health. In addition, these are the most commonly assessed parameters in volunteer monitoring as evidenced by their relative prevalence in existing datasets. In all of these studies, the

professional data were considered to be the standard for accuracy and the volunteer data were evaluated by their relative differences.

Water Quality Parameter Comparison Studies

Of the 15 water quality data comparison studies in this review, three studies focused on volunteer bacteria monitoring that involved a laboratory component, and the resulting comparisons also reflected differences in the volunteer and professional sampling methods (Au et al. 2000; Sarnelle et al. 2010; Stepenuck et al. 2011). For example, Au et al. (2000) compared the results of high school students using a simplified protocol to monitor local urban waterways for bacterial coliforms to those collected by professionals through traditional protocols, and found 83–96% agreement between volunteers and professionals at each of the monitored stations. All three studies concluded that volunteer bacterial data, although collected through a variety of methods, were comparable to those collected by professionals and could be used as effective monitoring tools. These results, however, do not directly apply to the data already collected by volunteer programs, especially regarding volunteer laboratory protocols designed for these comparisons alone. The 12 remaining studies used only field-based water quality sampling methods that were more comparable between volunteers and professionals.

From these 12 studies we next identified those that did not utilize datasets already in existence from an active volunteer water quality monitoring program, a total of six (**Table 1**). Two of these studies, Shelton (2013) and Muenich et al. (2016), did not use an existing program but instead recruited and trained volunteers specifically for the comparison study and, as with the bacteria studies, also compared different sampling methods between volunteers and professionals. The other four studies (Canfield et al. 2002; Hoyer et al. 2012; Stepenuck 2013; Storey et al. 2016) did use already active volunteer water quality monitoring programs to collect data, but only for sampling events that were part of the experimental design of the study. These studies did not include, compare, or analyze any data or datasets already in existence.

For example, Stepenuck (2013) compared data collected by the existing Wisconsin-based Water Action Volunteers (WAV) to data collected by professionals. For this study, volunteers sampled stream flow along with a professional for selected sampling events approximately four times a year for the two years of the study. Although the program had been existence longer and had additional datasets, only the experimental flow measurements collected for the study were included in the comparison analyses. Similarly, Storey et al. (2016) compared 12 water quality parameters collected by an existing volunteer monitoring group to those collected by professionals, but through a study design in which professionals and volunteers sampled side-by-side. Although they found that nine of the 12 parameters showed good agreement between the two groups, this comparison study analyzed data only from those experimentally designed samples and did not utilize any of the volunteer programs' previous data.

Canfield et al. (2002) compared data collected through the Florida LAKEWATCH program in which volunteers and researchers sampled side-by-side at 125 different lakes once in the summer and once in the winter over the course of one year. There were multiple parameters sampled, some in which volunteers were using modified protocols and some for which they were not, but overall the authors concluded that the mean values obtained by the volunteers were "strongly correlated" to those obtained by the professionals. Although they found significant differences between volunteers and professionals at some lakes, analysis of variance showed that the type of sampler (volunteer vs. professional) accounted for <1% of variance in the year surveyed (Canfield et al. 2002). In a follow-up study, Hoyer et al. (2012) analyzed data collected through the Florida LAKEWATCH program in which volunteers and professionals collected data on the same day. This study also was designed to test the comparability of the non-certified LAKEWATCH laboratory with the certified laboratory of the Florida DEP alongside professional- and volunteer-collected field data using similar protocols for the water quality parameters measured. Hoyer et al. (2012) also found no statistical difference between the data collected by volunteers and professionals.

However, neither of these studies included comparisons to the existing LAKEWATCH dataset. In 2012, Hoyer et al. (281) wrote that their study "demonstrates that FDEP can use the 25+ years of LAKEWATCH data on several hundred lakes and multiple estuaries to assess trends in nutrient and chlorophyll concentrations." Yet, without comparing their results to the existing 25+ year volunteer data, the results from their experiment may not necessarily apply to the existing data.

This is not to say that the existing datasets from this volunteer program or any other in these studies would be any less accurate, simply that they may be more open to professional bias. Although field collection methods for LAKEWATCH and other programs may not have changed significantly over time, and likely have been under quality assurance protocols throughout, most long-term programs will nonetheless experience many changes over their years of operation. Variations such as turnover in staff and volunteer personnel, as well as changes in equipment, geo-sensing, and database technology are common to professional and volunteer water quality monitoring operations alike. The natural accumulation of variation over time could make these long-term datasets more susceptible to professional bias, and ultimately less likely to be used. We feel that studies evaluating the consistency of volunteer-collected data over a long-term could be excellent support of ALL long-term citizen science efforts that have operated under quality assurance protocols throughout their monitoring history.

Comparison studies: Existing volunteer data

Thus far, this review has revealed that a majority of comparison studies used experimental protocols consisting of modified procedures and/or prescribed sampling events to collect volunteer data and demonstrated positive results regarding the ability of volunteers to collect data comparable to those collected by professionals. However, these results do not necessarily apply to the hundreds of existing, long-term, volunteer-collected datasets for which a myriad of events and/or variables could neither be prescribed nor controlled. The question then became: Does the current literature provide enough information for scientists and other professionals to better utilize existing, long-term, large-scale volunteer datasets?

The remaining six studies that compared volunteer and professional water quality data did utilize some portion of the existing data from a volunteer monitoring program for which the researchers had no control over the sampling times or sampling stations (Obrecht et al. 1998; Nicholson et al. 2002; Loperfido et al. 2010; Coates 2013; Dyer et al. 2014; Safford and Peters 2017). Although these did utilize existing volunteer datasets, most of them only used a portion of what was available, either focusing on a limited area or time frame within the larger-scale dataset.

For example, Obrecht et al. (1998) compared existing volunteer data collected by the Lakes of Missouri Volunteer Program with similar data that had been collected by professionals on the trophic classifications of lakes. By limiting the analysis to only those samples obtained from stations at which both volunteers and researchers had sampled and to the three volunteer samples that were closest in date to those collected by the researchers, they were able to compare the two existing datasets and to determine that the trophic classifications from the volunteers agreed with those of the professionals. Although the Lakes of Missouri Volunteer Program had been collecting data for longer, the dataset analyzed in this study included only volunteer observations from one year (Obrecht et al. 1998).

Additionally, Nicholson and Hodgkins (2002) used data from an existing volunteer program, Waterwatch Victoria, and compared them to professional data for turbidity, electrical conductivity (EC), pH, and TP. The authors found that the data were mostly comparable but that the level of agreement between volunteers and professionals varied by time and place, while also acknowledging that this could have been a factor of assessing data from only five stations with no more than one to five years of data at each station. Further, in 2013, Coates conducted a follow-up study using Waterwatch Victoria data, looking at trends over time as well as spatial differentiation of samples between stations. This study revealed evidence of seasonal cycling of parameters that varied with place. Coates' (2013) study spanned only one year, and the author acknowledged that this shorter time frame resulted in a low chance of detecting any significant differences.

One study that did use data from a larger area was Loperfido et al. (2010), in which data from the IOWATER program were compared to data collected concurrently by professionals at 971 stream stations across Iowa. This study used field samples of total nitrogen and total reactive phosphorus from the professionals and volunteers but also had lab samples to stand as the true value with which to compare and ultimately assess error or bias. The authors determined that the volunteer measurements, although significantly different from professional

measurements when looking at the specific values alone, were nevertheless successful in identifying and classifying most of the waters that violated USEPA standards, and the data accuracy improved when accounting for error and bias in the dataset.

These results from a statewide scale may reveal that larger volunteer datasets, although likely more useful for identifying trends and for management purposes, can contain higher levels of error and bias for professionals and volunteers alike. Volunteer (and professional) data accuracy can vary by place and time and additional years and/or site locations, and the additional variation they bring may impact the comparison results. For example, the previously mentioned Canfield et al. (2002) study that examined variability within their comparison data determined that type of sampler (volunteer or professional) accounted for <1% of variation, while lake-to-lake differences accounted for anywhere from 62–82% of the variance (Canfield et al. 2002). Thus, it may be hypothesized that long-term data that cover a larger sampling area would more likely generate even more variability within both volunteer and professional datasets, possibly affecting the resulting comparison analyses.

Comparison studies: Large scale volunteer data
With hundreds of volunteer water monitoring programs already in operation, many housing databases containing decades of long-term sampling, it can be argued that there is a need for more studies that specifically analyze the accuracy of large-scale, long-term existing volunteer datasets. As Safford and Peters (2017: 2) state,

> "previous studies of volunteer-collected water quality data have focused on highly controlled small-scale comparisons of data collected by volunteers alongside professionals or of data collected by volunteers and professionals sampling at the same or similar stations under similar conditions. This method, while rigorous, is resource-intensive and limited to relatively small sample sizes."

In fact, our review found only two studies that analyzed existing, long-term volunteer datasets that spanned five or more years. In one study, Safford and Peters (2017) collected citizen science data from online databases of two of the largest and longest-running statewide programs, Georgia Adopt-a-Stream and Rhode Island Watershed Watch. They compared the data from these programs with USGS field observations and USGS gauges in terms of dissolved oxygen (DO), while also using associated water temperature measurements to plot an expected range for DO based on an equation relating it to temperature. They found that volunteer-collected data fell within the expected relationship between temperature and DO, and that volunteer and professional data were in roughly the same range. This indicated that "data collected by large numbers of volunteers are as reliable as data collected under strict oversight of an agency such as the USGS" and "can provide reliable information about freshwater DO lev-

els" (Safford and Peters 2017: 1–2). One limitation of this study, however, was that all values for all stations within each state program were analyzed as a group, and the volunteer samples were not matched with professional samples by either location or date, so the results speak only to the overall values and cannot provide any information concerning variation over time or place.

In the second study, Dyer et al. (2014) compared data collected by Waterwatch volunteers in the Australian Capital Territory region to data collected by professional agencies between 2003–2012 in terms of conductivity, pH, turbidity, and DO. The volunteer samples were collected within a 10-day period and matched to a professional sample from the nearest station. Only stations with more than 20 matching samples were included, leaving 14 stations for which volunteer and professional samples were analyzed by comparing medians. The results showed excellent agreement for pH and conductivity, and good agreement for turbidity and DO. Dyer et al. (2014: 360) reported that this agreement was in spite of the fact that the researchers "retrospectively investigated the agreement between data collected by volunteers and as such was constrained to comparing data collected on different days and with different methods." The authors also noted that although a more detailed study could account for differences of method and sampling dates, given the observed agreement, such a study isn't necessary. The authors also mentioned that 1) professional data are assumed to be free of error and bias, which is not always the case; and 2) with volunteer data being used primarily to augment professional data, the quality of the data determined by this study is "fit for purpose." These assumptions allowed them to conclude that their results provided the necessary confidence for Waterwatch programs to be incorporated into monitoring strategies and allowed to augment existing monitoring efforts.

Conclusion and Directions for Future Research
The hundreds of citizen science water quality monitoring programs that already exist, as well as the growing number of programs being formed in recent years, has led to questions about volunteer data quality and its usefulness in water management (Canfield et al. 2002; Safford and Peters 2017). Many of the volunteer water monitoring programs discussed in this review, including the Rhode Island Watershed Watch and Georgia Adopt-a-Stream (Safford and Peters 2017), use a program structure that involves extensive volunteer training, quality assurance protocols, and professional oversight. There are many programs in other states such as the Missouri Stream Team, the Texas Stream Team, the Alabama Water Watch, and IOWATER that also have trained volunteers who use equipment that meets protocol standards and whose programs are designed in conjunction with the state's environmental agencies (Canfield et al. 2002; Loperfido et al. 2010; Deutsch and Ruiz-Cordova 2015; Safford and Peters 2017). There are likely many additional programs with a similar structure across the country, all of which may have been collecting high-quality data for many years. If this existing

data, moreover any volunteer data collected under these standards, is consistently shown to be of comparable accuracy to those of professionals, increasing amounts of volunteer data could be used to fill gaps in official data for management and restoration efforts.

A growing number of studies examining the reliability and accuracy of volunteer-collected water quality data through comparison to professional data have discovered that data accuracy between volunteers and professionals can be comparable across a variety of conditions. All of the studies outlined in this review, although necessarily based on assumptions of professional accuracy and relative comparability, have nonetheless paved the way for easing any professional bias against volunteer data that may exist.

Comparisons between professional and volunteer data are providing a valuable framework by which scientists and professionals can reframe their perceptions about non-expert data through its relationship to more familiar data for which the quality is known. Beyond the perceptions of professionals, and perhaps more importantly, the results from these comparison studies, also can provide important formative feedback for citizen science practitioners and volunteers to assure data quality, enhance reporting, justify funding, and continue to inform and improve program design.

While this review included 15 water quality data comparison studies, only two assessed the accuracy of large-scale, existing volunteer water quality monitoring datasets. To fully address any professional bias, especially toward large-scale datasets already in existence, more studies utilizing these kinds of datasets would be beneficial. These studies could provide much-needed information on relative consistency over time and space and ultimately increase confidence for expanded use of volunteer-collected data.

We recommend that more studies be conducted which utilize existing volunteer monitoring datasets that are long-term and cover a large geographic area for a comparison analysis to similar professional data. Both professional and volunteer datasets accumulate more variability as more years and locations are added to the dataset, as each new point has varying baseline conditions making overall accuracy more difficult to assess. The variability inherent to these types of datasets may make them more susceptible to professional bias, if for no other reason than they are more difficult to work with.

However, a well-planned comparison study should be able to account for this increased variation and still reveal whether the levels of agreement found between volunteers and professionals in the smaller-scale, more controlled studies is evidenced on the larger scale as well. Natural variation inherent in large-scale datasets should not mask similar patterns, which is what can be expected from both citizen science programs that have QA/QC measures in place and professional agencies that function under similar protocols. If both professionals and volunteers are getting the "right" answer, then studies in which long-term professional and volunteer data are compared should reveal similar patterns. These patterns

should manifest clearly in spite of the constraints of a post-hoc analysis, which would be necessary if using existing data.

The two studies in this review that did compare large-scale volunteer and professional sampling data, Dyer et al. (2014) and Safford and Peters (2017), also provide insight for future studies. Dyer et al. (2014) refined large volunteer and professional sample datasets to include only those samples nearest each other on the same stream and within a 10-day period. They concluded that, although there were some differences and biases, there was excellent agreement for the parameters assessed, especially considering they were comparing samples taken on different days using different sampling methods. Safford and Peters (2017) also concluded that the volunteer and professional measurements for DO lay within the same range and although there were slight differences, volunteer data could be used to provide reliable information. These studies revealed specific differences between volunteer and professional data for these large datasets which could inform future uses, and overall concluded that consistent patterns were evidenced for both volunteers and professionals.

Based on the findings from these studies and others in this review, we hypothesize that a comparison study utilizing long-term, existing data would not need to control for the variations in the dataset, such as individual sampler or sampling entity, exact location of sampling site, sampling season or time of day, or even exact sampling equipment used, as long as both volunteer and professional data being compared were collected under quality assurance protocols that are developed under the same overarching protocol (in these cases, the USEPA is the national standard) to achieve the best accuracy possible. Hopefully, this approach will make such a study more appealing to researchers. Although there will likely still be a need for extensive data refining before two disparate datasets can be comparable in a statistical analysis, we argue that the work may be well worth the reward. A few studies like this, if they demonstrate the pattern of consistency shown by smaller-scale, controlled studies, could reveal that quality and consistency over time need not be the purview of professionals alone.

Competing Interests
The authors have no competing interests to declare.

References
Au, J, Bagchi, P, Chen, B, Martinez, R, Dudley, SA and **Sorger, GJ.** 2000. Methodology for public monitoring of total coliforms, Escherichia coli and toxicity in waterways by Canadian high school students. *Journal of Environmental Management*, 58: 213–230. DOI: https://doi.org/10.1006/jema.2000.0323

Barrows, CW, Hoines, J, Vamstad, MS, Murphy-Mariscal, M, Lalumiere, K and **Heintz, J.** 2016. Using citizen scientists to assess climate change shifts in desert reptile communities. *Biological Conservation*, 195: 82–88. DOI: https://doi.org/10.1016/j.biocon.2015.12.027

Bonney, R, Ballard, H, Jordan, R, McCallie, E, Phillips, T, Shirk, J and **Wilderman, CC.** 2009a. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. CAISE Inquiry Group Report. Washington, DC, USA.

Bonney, R, Cooper, CB, Dickinson, J, Kelling, S, Phillips, T, Rosenberg, KV and **Shirk, J.** 2009b. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, 59: 977–984. DOI: https://doi.org/10.1525/bio.2009.59.11.9

Boudreau, SA and **Yan, ND.** 2004. Auditing the Accuracy of a Volunteer-Based Surveillance Program for an Aquatic Invader *Bythotrephes*. *Environ Monit Assess*, 91: 17–26. DOI: https://doi.org/10.1023/B:EMAS.0000009228.09204.b7

Burgess, HK, DeBey, LB, Froehlich, HE, Schmidt, N, Theobald, EJ, Ettinger, AK, HilleRisLambers, J, Tewksbury, J and **Parrish, JK.** 2016. The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation, The role of citizen science in biological conservation*, 208: 113–120. DOI: https://doi.org/10.1016/j.biocon.2016.05.014

Canfield, DE, Brown, CD, Bachmann, RW and **Hoyer, MV.** 2002. Volunteer Lake Monitoring: Testing the Reliability of Data Collected by the Florida LAKEWATCH Program. *Lake and Reservoir Management*, 18: 1–9. DOI: https://doi.org/10.1080/07438140209353924

Coates, A. 2013. Volunteer Monitoring of Water Quality in New Zealand: Where does the Value Lie? (Thesis). Canterbury, New Zealand: University of Canterbury.

Cox, TE, Philippoff, J, Baumgartner, E and **Smith, CM.** 2012. Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecological Applications*, 22: 1201–1212. DOI: https://doi.org/10.1890/11-1614.1

Delaney, DG, Sperling, CD, Adams, CS and **Leung, B.** 2008. Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biol Invasions*, 10: 117–128. DOI: https://doi.org/10.1007/s10530-007-9114-0

Deutsch, WG and **Ruiz-Córdova, SS.** 2015. Trends, challenges, and responses of a 20-year, volunteer water monitoring program in Alabama. *Ecology & Society*, 20: 154–163. DOI: https://doi.org/10.5751/ES-07578-200314

Devictor, V, Whittaker, RJ and **Beltrame, C.** 2010. Beyond scarcity: citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions*, 16: 354–362. DOI: https://doi.org/10.1111/j.1472-4642.2009.00615.x

Dickinson, JL, Zuckerberg, B and **Bonter, DN.** 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41: 149–172. DOI: https://doi.org/10.1146/annurev-ecolsys-102209-144636

Dyer, F, Harrison, E, Gruber, B, Nichols, S, Tschierschke, A and **O'Reilly, W.** 2014. Waterwatch data quality: an opportunity to augment professionally collected data sets. *Proceedings of the 7th Australian Stream Management Conference*, 27–30 Jul: 357–362.

Engel, SR and **Voshell, JR.** 2002. Volunteer Biological Monitoring: Can It Accurately Assess the Ecological Condition of Streams? *American Entomologist*, 48: 164–177. DOI: https://doi.org/10.1093/ae/48.3.164

Fore, LS, Paulsen, K and **O'Laughlin, K.** 2001. Assessing the performance of volunteers in monitoring streams. *Freshwater Biology*, 46: 109–123. DOI: https://doi.org/10.1111/j.1365-2427.2001.00640.x

Galloway, AWE, Tudor, MT and **Haegen, WMV.** 2006. The Reliability of Citizen Science: A Case Study of Oregon White Oak Stand Surveys. *Wildlife Society Bulletin*, 34: 1425–1429. DOI: https://doi.org/10.2193/0091-7648(2006)34[1425:TROCSA]2.0.CO;2

Gonsamo, A and **D'Odorico, P.** 2014. Citizen science: best practices to remove observer bias in trend analysis. *Int J Biometeorol*, 58: 2159–2163. DOI: https://doi.org/10.1007/s00484-014-0806-8

Gowan, C, Ruby, M, Knisley, R and **Grimmer, L.** 2007. Stream Monitoring Methods Suitable for Citizen Volunteers Working in the Coastal Plain and Lower Piedmont Regions of Virginia. *Am Entomol*, 53: 48–57. DOI: https://doi.org/10.1093/ae/53.1.48

Hochachka, WM, Fink, D, Hutchinson, RA, Sheldon, D, Wong, W-K and **Kelling, S.** 2012. Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution*, 27: 130–137. DOI: https://doi.org/10.1016/j.tree.2011.11.006

Hoyer, MV, Wellendorf, N, Frydenborg, R, Bartlett, D and **Canfield, DE, Jr.** 2012. A comparison between professionally (Florida Department of Environmental Protection) and volunteer (Florida LAKEWATCH) collected trophic state chemistry data in Florida. *Lake and Reservoir Management*, 28: 277–281. DOI: https://doi.org/10.1080/07438141.2012.736016

Hoyer, MV, Winn, J, Canfield, DE, Jr. 2001. Citizen Monitoring of Aquatic Bird Populations Using a Florida Lake. *Lake and Reservoir Management*, 17: 82–89. DOI: https://doi.org/10.1080/07438140109353976

Jordan, RC, Gray, SA, Howe, DV, Brooks, WR and **Ehrenfeld, JG.** 2011. Knowledge Gain and Behavioral Change in Citizen-Science Programs. *Conservation Biology*, 25: 1148–1154. DOI: https://doi.org/10.1111/j.1523-1739.2011.01745.x

Kobori, H, Dickinson, JL, Washitani, I, Sakurai, R, Amano, T, Komatsu, N, Kitamura, W, Takagawa, S, Koyama, K, Ogawara, T and **Miller-Rushing, AJ.** 2015. Citizen science: a new approach to advance ecology, education, and conservation. *Ecol Res*, 31: 1–19. DOI: https://doi.org/10.1007/s11284-015-1314-y

Kosmala, M, Wiggins, A, Swanson, A and **Simmons, B.** 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14: 551–560. DOI: https://doi.org/10.1002/fee.1436

Kress, WJ, Garcia-Robledo, C, Soares, JVB, Jacobs, D, Wilson, K, Lopez, IC and **Belhumeur, PN.** 2018. Citizen Science and Climate Change: Mapping the Range Expansions of Native and Exotic Plants with the Mobile App Leafsnap. *BioScience*, 68: 348–358. DOI: https://doi.org/10.1093/biosci/biy019

Loperfido, JV, Beyer, P, Just, CL and **Schnoor, JL** 2010. Uses and Biases of Volunteer Water Quality Data. *Environ. Sci. Technol*, 44: 7193–7199. DOI: https://doi.org/10.1021/es100164c

McKinley, DC, Miller-Rushing, AJ, Ballard, HL, Bonney, R, Brown, H, Cook-Patton, SC, Evans, DM, French, RA, Parrish, JK, Phillips, TB, Ryan, SF, Shanley, LA, Shirk, JL, Stepenuck, KF, Weltzin, JF, Wiggins, A, Boyle, OD, Briggs, RD, Chapin, SF, Hewitt, DA, Preuss, PW and **Soukup, MA.** 2016. Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation, The role of citizen science in biological conservation*, 208: 15–28. DOI: https://doi.org/10.1016/j.biocon.2016.05.015

Moffett, ER and **Neale, MW.** 2015. Volunteer and professional macroinvertebrate monitoring provide concordant assessments of stream health. *New Zealand Journal of Marine and Freshwater Research*, 49: 366–375. DOI: https://doi.org/10.1080/00288330.2015.1018913

Muenich, R, Peel, S, Bowling, L, Haas, M, Turco, R, Frankenberger, J and **Chaubey, I.** 2016. The Wabash Sampling Blitz: A Study on the Effectiveness of Citizen Science. *Citizen Science: Theory and Practice*, 1. DOI: https://doi.org/10.5334/cstp.1

Nerbonne, JF and **Nelson, KC.** 2004. Volunteer Macroinvertebrate Monitoring in the United States: Resource Mobilization and Comparative State Structures. *Society & Natural Resources*, 17: 817–839. DOI: https://doi.org/10.1080/08941920490493837

Nerbonne, JF and **Vondracek, B.** 2003. Volunteer Macroinvertebrate Monitoring: Assessing Training Needs through Examining Error and Bias in Untrained Volunteers. *Journal of the North American Benthological Society*, 22: 152–163. DOI: https://doi.org/10.2307/1467984

Newman, G, Wiggins, A, Crall, A, Graham, E, Newman, S and **Crowston, K.** 2012. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10: 298–304. DOI: https://doi.org/10.1890/110294

Nicholson, E, Ryan, J and **Hodgkins, D.** 2002. Community data – where does the value lie? Assessing confidence limits of community collected water quality data. *Water Science and Technology*, 45: 193–200. DOI: https://doi.org/10.2166/wst.2002.0395

Obrecht, DV, Milanick, M, Perkins, BD, Ready, D and **Jones, JR.** 1998. Evaluation of Data Generated from Lake Samples Collected by Volunteers. *Lake and Reservoir Management*, 14: 21–27. DOI: https://doi.org/10.1080/07438149809354106

O'Leary, N, Vawter, AT, Wagenet, LP and **Pfeffer, M.** 2004. Assessing Water Quality Using Two Taxonomic Levels of Benthic Macroinvertebrate Analysis: Implications for Volunteer Monitors. *Journal of Freshwater Ecology*, 19: 581–586. DOI: https://doi.org/10.1080/02705060.2004.9664738

Oscarson, DB and **Calhoun, AJK.** 2007. Developing vernal pool conservation plans at the local level using citizen-scientists. *Wetlands*, 27: 80–95. DOI: https://doi.org/10.1672/0277-5212(2007)27[80:DVPCPA]2.0.CO;2

Paul, JD, Buytaert, W, Allen, S, Ballesteros-Cánovas, JA, Bhusal, J, Cieslik, K, Clark, J, Dugar, S, Hannah, DM, Stoffel, M, Dewulf, A, Dhital, MR, Liu, W, Nayaval, JL, Neupane, B, Schiller, A, Smith, PJ and **Supper, R.** 2018. Citizen science for hydrological risk reduction and resilience building. *Wiley Interdisciplinary Reviews: Water*, 5: e1262. DOI: https://doi.org/10.1002/wat2.1262

Penrose, D and **Call, SM.** 1995. Volunteer Monitoring of Benthic Macroinvertebrates: Regulatory Biologists' Perspectives. *Journal of the North American Benthological Society*, 14: 203–209. DOI: https://doi.org/10.2307/1467735

Raddick, MJ, Bracey, G, Gay, PL, Lintott, CJ, Murray, P, Schawinski, K, Szalay, AS and **Vandenberg, J.** 2010. Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, 9. DOI: https://doi.org/10.3847/AER2009036

Reynoldson, T, Hampel, L and **Martin, J.** 1986. Biomonitoring networks operated by schoolchildren. *Environmental Pollution Series A, Ecological and Biological*, 41: 363–380. DOI: https://doi.org/10.1016/0143-1471(86)90027-9

Safford, H and **Peters, CA.** 2017. Citizen Science for Dissolved Oxygen Monitoring: Case Studies from Georgia and Rhode Island. *Environmental Engineering Science*, 35: 362–372. DOI: https://doi.org/10.1089/ees.2017.0218

Sarnelle, O, Morrison, J, Kaul, R, Horst, G, Wandell, H and **Bednarz, R.** 2010. Citizen monitoring: Testing hypotheses about the interactive influences of eutrophication and mussel invasion on a cyanobacterial toxin in lakes. *Water Research*, 44: 141–150. DOI: https://doi.org/10.1016/j.watres.2009.09.014

Savan, B, Morgan, AJ and **Gore, C.** 2003. Volunteer Environmental Monitoring and the Role of the Universities: The Case of Citizens' Environment Watch. *Environmental Management*, 31, 0561–0568. DOI: https://doi.org/10.1007/s00267-002-2897-y

Shelton, AM. 2013. The accuracy of water quality monitoring data: a comparison between citizen scientists and professionals.

Silvertown, J. 2009. A new dawn for citizen science. *Trends in Ecology & Evolution*, 24: 467–471. DOI: https://doi.org/10.1016/j.tree.2009.03.017

Stepenuck, KF. 2013. Improving Understanding of Outcomes and Credibility of Volunteer Environmental Monitoring Programs – ProQuest (Dissertation)., Madison, Wisconsin, USA: The University of Wisconsin – Madison.

Stepenuck, KF and **Genskow, KD.** 2018. Characterizing the Breadth and Depth of Volunteer Water Monitoring Programs in the United States. *Environmental Management*, 61: 46–57. DOI: https://doi.org/10.1007/s00267-017-0956-7

Stepenuck, KF, Wolfson, LG, Liukkonen, BW, Iles, JM and **Grant, TS.** 2011. Volunteer monitoring of E. coli in streams of the upper Midwestern United States: a comparison of methods. *Environ Monit Assess*, 174: 625–633. DOI: https://doi.org/10.1007/s10661-010-1483-7

Storey, R, Wright-Stow, A, Kin, E, Davies-Colley, R and **Stott, R.** 2016. Volunteer stream monitoring: Do the data quality and monitoring experience support increased community involvement in freshwater decision making? *Ecology and Society*, 21. DOI: https://doi.org/10.5751/ES-08934-210432

Valdes, C, Ferreirae, M, Feng, T, Wang, H, Tempel, K, Liu, S and **Shaer, O.** 2012. A collaborative environment for engaging novices in scientific inquiry. In: *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, 109–118. New York, NY, USA: ITS '12. ACM. DOI: https://doi.org/10.1145/2396636.2396654