RESEARCH PAPER

The Wabash Sampling Blitz: A Study on the Effectiveness of Citizen Science

Rebecca Logsdon Muenich[*], Sara Peel[†], Laura C. Bowling[‡], Megan Heller Haas[‡], Ronald F. Turco[‡], Jane R. Frankenberger[§] and Indrajeet Chaubey[‖]

The increasing number of citizen science projects around the world brings the need to evaluate the effectiveness of these projects and to show the applicability of the data they collect. This research describes the Wabash River Sampling Blitz, a volunteer water-quality monitoring program in Central Indiana developed by the Wabash River Enhancement Corporation (WREC). Results indicate that field test strips for nitrate+nitrite-N read by volunteers generally agree with lab-determined values. Orthophosphate results are less transferable owing to low observed concentrations, although the field test strip values from unfiltered samples consistently over-predicted the lab values. Hierarchical cluster analysis (HCA) applied to volunteer-collected data groups sampling sites into meaningful management clusters that can help to identify water-quality priorities across the watershed as a proof of concept for watershed managers. Results of the HCA provide an opportunity for WREC to target future programs, education, and activities by analyzing the data collected by citizen scientists. Overall this study demonstrates how citizen science water quality data can be validated and applied in subsequent watershed management strategies.

Keywords: Citizen science; water quality monitoring; watershed groups; cluster analysis; Wabash River

Introduction

Freshwater ecosystems provide many benefits to society, including food, water, flood control, aesthetics, and recreation (Finlayson et al. 2005). Despite state and federal regulations aimed at protecting these resources, 52% of all assessed streams in the United States are impaired (USEPA 2012). Recognizing the degraded state of their waters, many public stakeholders across the world have formed watershed associations in efforts to improve the health of their local watersheds (Cline and Collins 2003). Watershed groups have been shown to enhance their community's chance of receiving funds to improve their watershed and to develop programs to protect and enhance water quality (Cline and Collins 2003). However, monitoring water quality is not only complex logistically, it is also expensive (Maas et al. 1991). Many watershed groups turn to citizen science to both engage the public and collect large amounts of data that they need to address their concerns. For example, a 9th grade class in New Jersey worked with their local watershed partnership to determine their community's willingness to pay for the restoration of ecosystem services in the watershed (Nicosia et al. 2014).

Citizen science involves engaging and collaborating with members of the public to gather data to address scientific problems (Cohn 2008; Dickinson et al. 2012; Miller-Rushing et al. 2012). Recently, citizen science projects have grown from a few examples to thousands (Conrad and Hilchey 2009; Shirk et al. 2012). Many benefits have been identified as reasons for including citizens in scientific work, including increased public knowledge of science, ability to capture large amounts of data across space and time, advancement of scientific knowledge, lowered cost of collection and processing, increased social capital, and government and ecosystem benefits (Bonney et al. 2009; Cohn 2008; Conrad and Hilchey 2011; Silvertown et al. 2009). Kolok and Schoenfuss (2011) specifically describe citizen science as a meaningful approach for monitoring waterways. Despite all of the benefits that citizen science projects provide, there are still continued concerns about the validity and subsequent application of the data collected (Bonney et al. 2014; Conrad and Hilchey 2011; Dickinson et al. 2010; Kim et al. 2011; Kolok et al. 2011). Specifically, Conrad and Hilchey (2009) note that citizen science data often are not used in the decision-making process, either because of concerns with collection methods or the inability to get data to decision-makers.

[*] Department of Agricultural & Biological Engineering, Purdue University, West Lafayette, Indiana, USA; Graham Sustainability Institute, University of Michigan, Ann Arbor, Michigan, USA

[†] Wabash River Enhancement Corporation, Lafayette, Indiana, USA

[‡] Department of Agronomy, Purdue University, West Lafayette, Indiana, USA

[§] Department of Agricultural & Biological Engineering, Purdue University, West Lafayette, Indiana, USA

[‖] Department of Earth, Atmospheric, & Planetary Sciences and Department of Agricultural & Biological Engineering, Purdue University, West Lafayette, Indiana, USA

Corresponding author: Rebecca Logsdon Muenich (rlogsdon@umich.edu)

Many examples of citizen science for monitoring water quality exist in the peer-reviewed scientific literature (Au et al. 2000; Buytaert et al. 2014; Canfield et al. 2002; Fore at al. 2001; Lottig et al. 2014; Maas et al. 1991; Nicholson et al. 2002; Overdevest et al. 2004; Peckenham et al. 2012; Savan et al. 2003) and more certainly exist in practice. In assessing the validity of volunteer-collected and volunteer-analyzed water chemistry data, Nicholson et al. (2002) found mixed results depending on the variable, though their assessment was between yearly means of two different datasets, not direct sample comparison. Savan et al. (2003) found that 40% of their citizen science program's water chemistry variables failed quality control checks, leading them to use biological measures of water quality over chemical measures. Fore et al. (2001) and Canfield et al. (2002) both found no significant difference between volunteer collected biological and chemical water quality data, while Maas et al. (1991) chose to run their water chemistry samples through a university lab to avoid volunteer error. Au et al. (2000) found that local high school students were able to evaluate toxicity of *Escherichia coli* (*E. coli*) similarly to experts after they were trained, and Peckenham et al. (2012) determined that middle to high-school aged students were able to accurately analyze pH and conductivity, but that additional quality assurance was needed for hardness, chloride, and nitrate testing. There is still a need to assess current water monitoring programs and provide examples of applications of citizen science to collect and analyze water quality data for improved watershed management.

The overall goal of this research is to address two of the main issues surrounding citizen science: data validity and data application. The first objective is to compare volunteer-collected and volunteer-analyzed water quality data to volunteer-collected and laboratory-analyzed water quality data to assess the validity of the volunteer-analyzed data. The second objective is to provide a proof of concept of how the data collected by volunteers can be used by watershed groups to target management strategies and priorities.

Methods
Program description
The Wabash River Enhancement Corporation (WREC) is a 501c3 nonprofit agency established in 2004 and based in Lafayette, Tippecanoe County, Indiana (www.wabashriver. net). The goal of WREC is to lead efforts within the community to improve and enhance the local Wabash River corridor as well as to engage and educate the community in the implementation of projects, programs, and activities that enhance the Wabash River ecosystem. WREC has established many programs to achieve their goals, including cost-share programs for urban and agricultural best management practices, green business certification, and riverfront development.

In 2009, WREC—in partnership with researchers at Purdue University—established a citizen science water quality monitoring program called the Wabash Sampling Blitz (Blitz). The two main goals of the Blitz are first, to provide a hands-on opportunity for the community to experience the Wabash River and its tributaries through citizen science, and second, to obtain uniform, simultaneous water quality data throughout the area that WREC serves. This large-scale and simultaneous collection of data can help by providing "hot spot" identification for future watershed priorities (Kolok et al. 2011). The Blitz occurs twice a year in the spring (April) and fall (September), when approximately 250 volunteers sample 206 sites within the Region of the Great Bend of the Wabash River Watershed (**Figure 1**). Land use in the watershed is mostly row-crop agriculture (corn and soybean), but the watershed is also host to the urban areas of West Lafayette and Lafayette (~100,000 population). Since its establishment in 2009 and until the fall of 2013, the Blitz has benefited from the contribution of 889 unique community volunteers (192 repeat volunteers) giving more than 3,000 hours of their time (**Figure 2**).

The Blitz is held for approximately four hours on one afternoon. Volunteers may arrive at any time during the sampling window. Volunteers are pre-assigned to one of three staging locations where they either meet up with or are matched with at least one other sampling partner. Staging location organizers detail the sampling methods and objectives with volunteer groups. Volunteers then travel in their own vehicles to 3–4 sites where they collect water samples in stream, measure water transparency with a transparency tube, and measure in-stream water temperature. Volunteers then return to their staging location where additional volunteers help them filter a portion of their samples to use in subsequent lab analyses. The remainders of their water samples are used to test for nutrients and contaminants on-site using field test strips. Volunteers then color in selected constituent (nitrate+nitrite-N and water temperature) levels on a map of the watershed so they can easily compare their results to other portions of the watershed, as well as to data from the previous year. The constituents tested in lab and by participants have varied from year to year depending on funding availability, but many have been consistently analyzed (**Table 1**). The Purdue University Soil Science Laboratory used an AQ2 Discrete Analyzer to measure concentrations of ammonia (mg/L; AQ2 method EPA-103-A Rev. 10), nitrate+nitrite-N (mg/L; AQ2 method EPA-114-A Rev. 9), and orthophosphate-P (mg/L; AQ2 method EPA-118-A Rev. 5-subsequently converted to orthophosphate). Dissolved organic carbon concentration (mg/L) was measured with a Shimadzu TOC-V CSH. Field test strips were used by volunteers to determine concentrations for nitrate+nitrite-N (mg/L; Hach Aquacheck Cat. 27454-25) and orthophosphate (mg/L; Hach Aquacheck Cat. 27571-50), and pH levels (Sigma P-4411). Spring 2010 samples only were analyzed using WaterWorks Nine-Way Test Kits that included pH, nitrate+nitrite-N, and other tests. Volunteers also used a transparency tube, with a secchi disc, and marked in units of cm to record in-stream transparency and took water temperature readings using alcohol thermometers (°C).
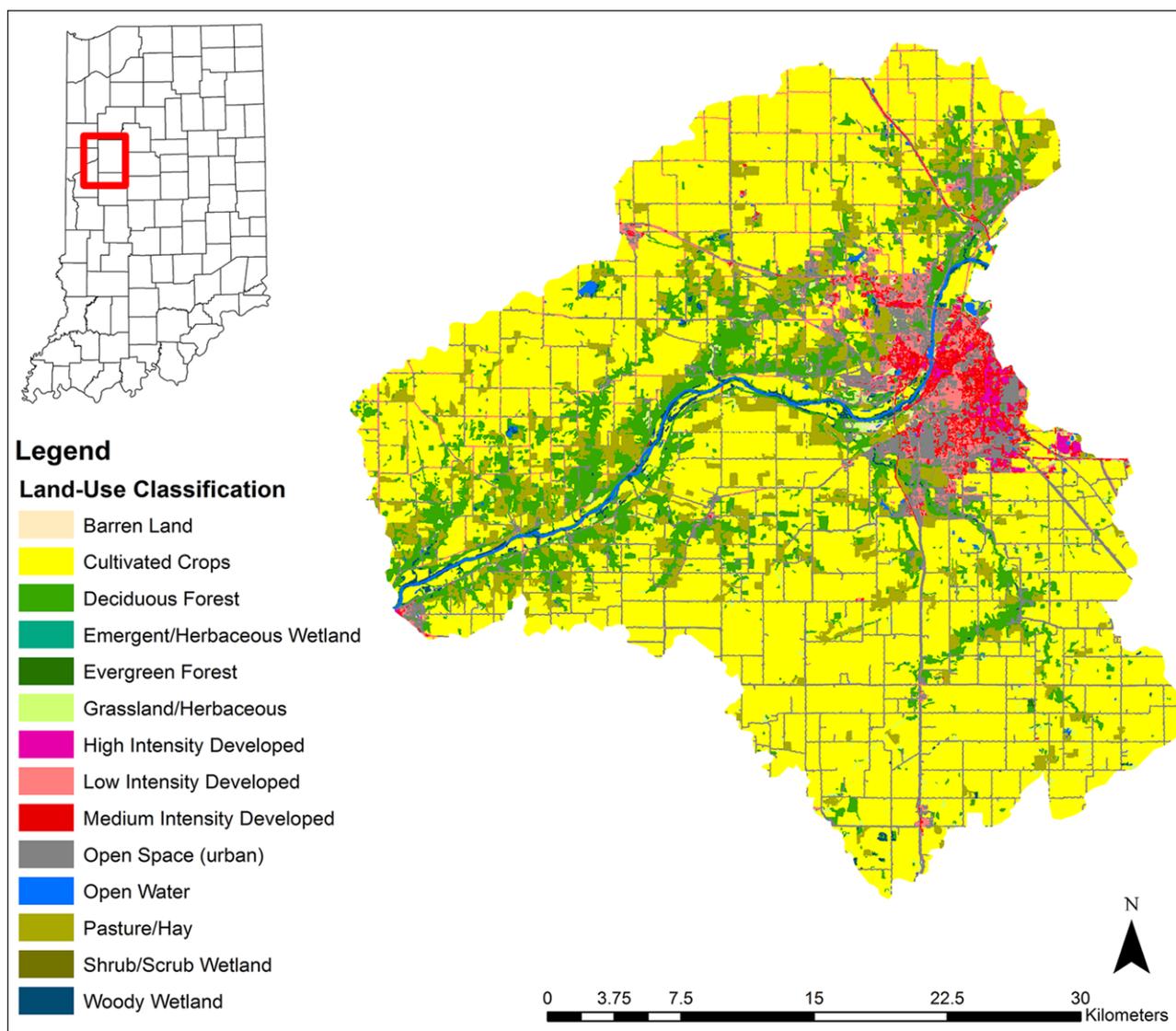
Figure 1: Location and land-use of the Great Bend of the Wabash River Watershed from the 2001 United States Geological Survey National Land Cover Dataset.

Comparison of volunteer and lab-determined data
For this study, the volunteer-collected nitrate+nitrite-N and orthophosphate sample concentrations were compared to the lab-determined sample concentrations because these two variables had the most field and lab data available. The field test strips are used on unfiltered samples while the lab analyses are performed on samples filtered through a 0.45 micron glass fiber filter. While this should not impact the nitrate+nitrite-N comparisons, it could lead to volunteer overestimation of orthophosphate due to the affinity of phosphorus to sorb to sediments (Zhou et al. 2005). However, given the high transparency of the water samples on average (**Table 3**), this may not have a large influence on the readings. Another issue with comparing these datasets is that the field test strips used by volunteers have a binned, colored scale comparison for volunteers to read the level of the constituent. These bins essentially make the data provided by volunteers categorical. Thus, in order to compare the two datasets, the lab-determined data were binned to match the test strips in

order to make them categorical as well (**Table 2**). The test strip scales provide a single concentration value associated with each color, which was assumed to represent the mid-point of the represented concentration range. Bins were therefore centered around the test strip concentration values. For example, the first nitrate+nitrite-N bin ranges from zero to halfway between the first and second concentration value (0.5). Because nitrate and nitrite are combined from the lab analysis, the field strip nitrate and nitrite values were also added together. Because nitrate concentrations are larger and because most volunteer-read nitrite values were close to zero, nitrate values were used to make the bins.

Once the datasets were categorized as described in **Table 2**, three measures of agreement were used to determine how well the volunteer-read data compared with actual lab concentrations: The percent agreement, the unweighted and weighted Cohen's Kappa Statistic (hereafter referred to as Kappa), and the unweighted and weighted Bangdiwala *B* statistic (hereafter referred to as *B*).
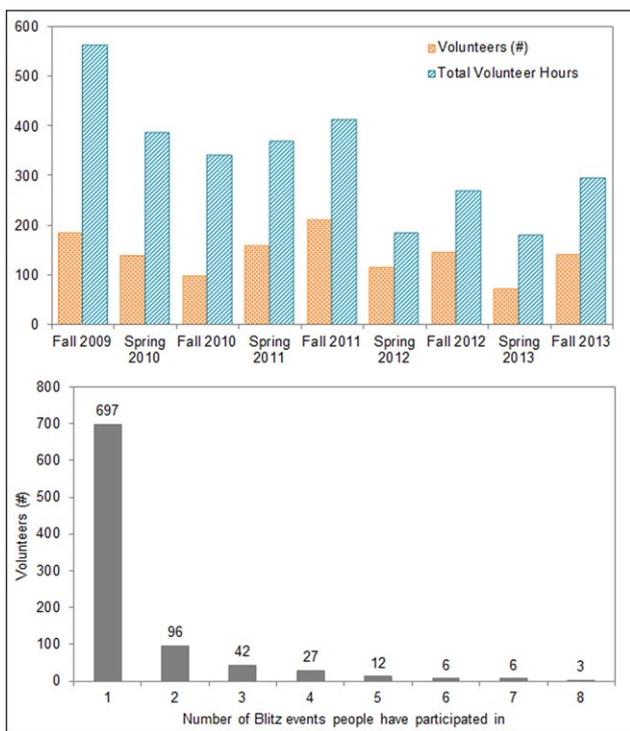
Figure 2: Total number of volunteers and hours worked by volunteers for each Blitz event (top) and the number of Blitz events in which unique volunteers have participated (bottom); the sum of all bars on the bottom figure indicates the total number of unique volunteers who participated in the Blitz events from the fall of 2009 to the fall of 2013.

Three measurements were chosen to provide more certainty of the conclusions and to demonstrate multiple methods that can be used to assess agreement. As a first order evaluation, the exact percent agreement and the percent agreement within one category were calculated to provide a straightforward assessment of the percent of volunteer-read observations that fell exactly into the same bin as lab-analyzed values (exact percent agreement) or the percent of observations that fell into the same bin or one bin higher or lower of lab-analyzed values (percent agreement within one category). The Kappa statistic and the B statistic are two different ways to evaluate the agreement between two independently classified observations, provided the datasets have the same categories (Munoz and Bangdiwala 1997). The B and Kappa statistics both go beyond percent agreement by taking into consideration that some agreement could occur by chance (Banerjee et al. 1999; Munoz and Bangdiwala 1997). The B statistic is calculated based on a graphical "area of agreement" whereas the Kappa statistic is based on the observed proportion of agreement (Munoz and Bangdiwala 1997). The higher the Kappa and B statistics, the better the agreement between the two datasets. The Kappa statistic is also known to be more conservative in measuring agreement when most values fall into one category, known as the prevalence problem, which is important for interpretation (Hallgren 2012; Viera and Garrett 2005). The unweighted statistics only compare how well the two observation sets match up for each category bin. The weighted statistics consider how far the observations are from exact agreement (i.e., within one or two categories). More weight is thus given to agreement in categories closer to exact agreement. The weighted and unweighted Kappa and B statistics were determined for each Blitz, as well as for all Blitz events combined using the 'vcd' package of R (Meyer et al. 2014). The interpretation guidelines developed by Munoz and Bangdiwala (1997) were then used to determine the qualitative level of agreement. Lastly, to further evaluate the levels of agreement between the field and lab concentrations, bubble plots were created

Lab analysis (completed by professionals)	Fall 2009	Spring 2010	Fall 2010	Spring 2011	Fall 2011	Spring 2012	Fall 2012	Spring 2013	Fall 2013
Dissolved organic carbon (DOC; mg/L)	x	x	x	x	—	x	x	—	x
Nitrate+nitrate-N (mg/L)	x	x	x	x	x	x	x	x	x
Orthophosphate (mg/L)	x	x	x	x	x	x	x	x	x
pH	—	—	—	—	—	x*	—	—	—
On-site analysis (completed by volunteers)	**Fall 2009**	**Spring 2010**	**Fall 2010**	**Spring 2011**	**Fall 2011**	**Spring 2012**	**Fall 2012**	**Spring 2013**	**Fall 2013**
Nitrate+nitrite-N strip (mg/L)	x	x	x	x	x	x	x	x	x
Orthophosphate strip (mg/L)	x	—	—	—	—	—	x	x	x
pH strip	x	x	x	x	x	—	x	x	x
Temperature (°C)	x	x	x	x	x	x	x	x	x
Transparency tube (cm of visibility)	—	—	x	x	x	x	x	x	x

Table 1: List of water quality data measured in the lab and field for each of the Blitz events. An "x" indicates the constituent was measured, a "—" indicates that it was not measured.

*In Spring 2012, pH was measured by a Purdue lab technician using the test strips due to contaminated test strips at staging locations.

Nitrate-N test strip scale value (mg/L)	Assigned nitrate+nitrite bins (mg/L)	Orthophosphate test strip scale value (mg/L)	Assigned orthophosphate bins (mg/L)
0.0	<0.5	0.0	<2.5
1.0	0.5–1.5	5.0	2.5–10.0
2.0	1.6–3.5	15	10.1–22.5
5.0	3.6–7.5	30	22.6–40.0
10.0	7.6–15.0	50	>40.0
20.0	15.1–35.0		
50.0	>35.0		

Table 2: Bins used to compare volunteer-determined and lab-determined concentrations for nitrate+nitrite-N and orthophosphate.

in *R*. These plots provide a visual interpretation of agreement between two observed datasets. Perfect agreement is shown along the upward diagonal of the plot, with the number of data points that fall within a category provided in each bubble and proportional to bubble size.

Cluster analyses

The second objective of this research was to utilize data collected from the Blitz events (both volunteer-analyzed and lab-analyzed) to help target outreach and education within the watershed. Because multiple variables were available over the period of nine Blitz events over five years at 206 sites, multivariate cluster analyses techniques were employed to examine the large dataset. Cluster analysis is a multivariate statistical technique that can aid in interpreting very large datasets by grouping objects (e.g., sampling sites) with similar characteristics together, and is a common tool used in riverine systems (Bierman et al. 2011). While many studies have used cluster analysis techniques to interpret water chemistry data (Alberto et al. 2001; Daughney et al. 2012; Güler et al. 2002; Kim et al. 2005; Mavukkandy et al. 2014; Najar et al. 2012; Pati et al. 2014; Shrestha and Kazama 2007; Simeonov et al. 2003; Singh et al. 2004; Singh et al. 2005; Templ et al. 2008; Wang et al. 2013; Wang et al. 2014), none of these studies were conducted as part of a citizen science effort or used data collected by citizen science volunteers.

Six variables that had the most data available were used in the cluster analyses. Variables were first pre-tested to see if the fall and spring Blitz samples had different means, as temporal variation has been shown to be important in cluster analysis (Gamble and Babbar-Sebens 2012). All observations for a given variable in the spring were tested versus all observations for a given variable in the fall using two sample t-tests (**Table 3**). All six variables showed significant differences in fall and spring values, so these were separated into different variables (i.e., fall pH and spring pH were treated as two separate variables). The individual sampling event observations were averaged into fall and spring variables for each location, because some sites were not sampled in a given year owing to low/high water levels, site inaccessibility, or a lack of volunteers. Thus, the final cluster analysis was completed to categorize each sampling location based on its average spring and fall water quality (12 variables, 206 sites).

There are many types of clustering techniques; however, for this project hierarchical clustering was employed as it has been previously applied to the classification of water quality data and is the most common approach (Shrestha and Kazama 2007). Hierarchical clustering connects similar data points based on a chosen distance measure and seeks to minimize within-cluster variation and to maximize between cluster variations. There are two main types of hierarchical clustering: Agglomerative, starting with individual data and grouping like observations, and divisive, starting with all data in one group and then dividing into groups. An agglomerative technique was used because these methods are very efficient (Alberto et al. 2001) and often have been used for water chemistry clustering. Before the cluster analysis was completed, the variables were transformed to achieve normal distribution using either a log10 transformation or a three-parameter lognormal or log10 transformation (**Table 4**) and then standardized to meet the normality and equal variance assumptions of cluster analysis (Güler et al. 2002). The cluster analysis was completed using *R* statistical software employing the Ward's Method using a Euclidean distance measure (Alberto et al. 2001; Güler et al. 2002; Kim et al. 2005; Shrestha and Kazama 2007; Simeonov et al. 2003; Singh et al. 2004; Singh et al. 2005). Cluster numbers were determined using $D_{max}*0.66$ as the cutoff criteria where D_{max} is the maximum distance between clusters (Singh et al. 2005). A subsequent principal components analysis (PCA) was used to identify important variables in the cluster analysis (see the Supplementary Materials for details). Principal components analysis is most often employed to reduce a large dimension dataset into smaller dimensions by creating combinations of variables called principal components (Güler et al. 2002). Boxplots summarizing the distribution of the variables which contributed most to principal component loadings were constructed to examine the results of the cluster analysis.

Results

Comparison of volunteer and lab-determined data

There was a good agreement for nitrate+nitrite-N between the volunteer-analyzed and lab-analyzed data (**Figure 3**). The exact (same bin) percentage agreement between the two datasets for nitrate+nitrite-N was 55% and went up to 84% if considering agreement within one category, i.e.,

Water Quality Variable	Fall Mean	Spring Mean	*p*-value
Nitrate+nitrite-N (mg/L)	1.15	**4.39**	<2.2e-16
Orthophosphate (mg/L)	**0.30**	0.04	2.03e-07
Temperature (°C)	**16.5**	12.5	<2.2e-16
Dissolved organic carbon (mg/L)	**3.66**	2.88	1.54e-05
pH	7.31	**7.97**	1.72e-03
Transparency (cm)	81.5	**95.6**	1.83e-07

Table 3: Fall and spring mean comparison using a two sample t-test. **Bold** indicates the significantly higher value.

Normalization Technique	Water Quality Variable Transformed
no transformation (already normally distributed)	fall pH, spring pH, fall temperature, spring temperature
log10 transformation	fall dissolved organic carbon, spring dissolved organic carbon
three parameter lognormal transformation; shift parameter determined using a quantile lower bound estimator	fall nitrate+nitrite-N, spring nitrate+nitrite-N, fall orthophosphate, spring orthophosphate
three parameter log10 transformation; shift parameter estimated as 1 plus data maximum	fall turbidity, spring turbidity

Table 4: Transformations used to normalize variables prior to cluster analysis.
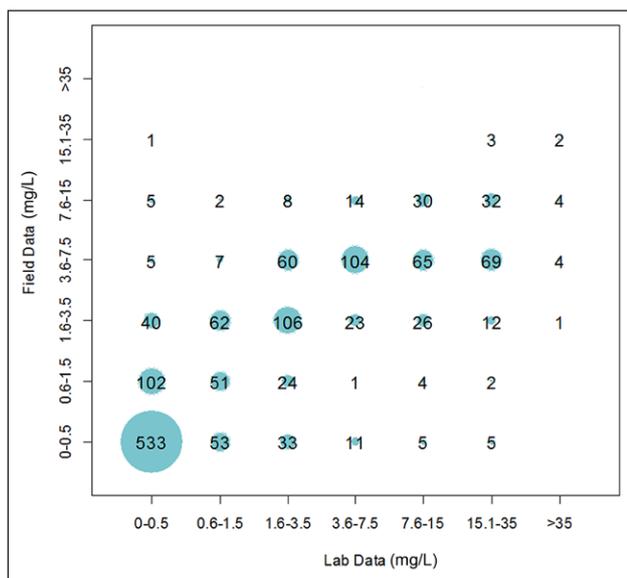
Figure 3: Bubble plot showing agreement between field (volunteer readings) and lab (professional analyses) estimated nitrate+nitrite-N for all Blitz events. The number inside the bubble indicates how many observations fell into that category.

the volunteer-read concentrations fell within one bin of the lab-determined values (**Table 5**). The Kappa and *B* statistics show that field strip volunteer-read nitrate+nitrite-N concentration data agree moderately to substantially well (Munoz and Bangdiwala 1997) with lab-determined values most of the time (**Figures 3** and **4**). As seen in **Figure 3**, most observations fell into the lowest bin. This was especially true for the fall of 2013, for which the *B* statistics are high while the Kappa statistics are low. This is likely because the Kappa statistic does not do well when there are very few categories (Viera and Garrett 2005). The low Kappa and *B* values in the spring of 2010 are likely

due to the fact that different test strips were used in this Blitz than in all other Blitz events. This change in strips may have led to incorrect readings by volunteers, or these strips could have been faulty. Because of this, overall statistics were calculated both with ("All") and without ("All-S10") those values.

The bubble plot illustrating the agreement between volunteer- and lab-analyzed orthophosphate values shows that the range of the data was even lower than that of the nitrate+nitrite-N data (**Figure 5**). The overall percentage agreement was only 33%, but went up to 99% if considering agreement +/- one bin (**Table 5**). The Kappa and unweighted *B* statistics for the orthophosphate comparison are fair to moderate overall (**Figure 6**). This is because most of the data were below 5 mg/L, thus putting them into one of the lowest two bins (**Figure 5**). The weighted *B* statistic is very good because all except a few of the data points were within the lower two categories. Because the actual orthophosphate levels fell primarily into the lower category, the results of this comparison may not be broadly transferable to other studies. Although the majority of samples were overestimated, the majority of samples were also within one bin +/- bin, indicating that the volunteers were not estimating values completely incorrectly. This consistent overprediction bias by the volunteers could be due to the fact that most samples were actually in the lowest category or perhaps the volunteers were reading unfiltered samples and the lab data were for filtered samples.

Cluster analyses

Ward's Method using Euclidean distance measures was applied to the water quality data in order to group the sampling sites into similar clusters. Applying the D_{max} *0.66 criterion, three distinct clusters emerged. To evaluate cluster membership a PCA was performed on the variables

Event	Nitrate+nitrite-N		Orthophosphate	
	Exact Agreement	*Within One Category*	*Exact Agreement*	*Within One Category*
All	**55%**	**84%**	**33%**	**99%**
Fall 2009	73%	91%	37%	99%
Spring 2010	6%	37%	–	–
Fall 2010	57%	80%	–	–
Spring 2011	36%	93%	–	–
Fall 2011	73%	93%	–	–
Spring 2012	53%	94%	–	–
Fall 2012	72%	96%	12%	99%
Spring 2013	62%	96%	42%	100%
Fall 2013	67%	83%	43%	98%

Table 5: Percent agreement between field and lab tested nitrate+nitrite-N and orthophosphate data.

(see Supplementary Materials). The variables contributing the greatest loading to the first three principal components were summarized using boxplots, grouped by cluster (**Figure 7**) and included: Spring and fall DOC, spring nitrate+nitrite-N, spring temperature, and fall and spring orthophosphate. Additionally, the cluster means are summarized in a spider plot to visualize cluster separation (**Figure 8**). Boxplots for all other variables are included in the Supplementary Materials.

The clusters were mapped within the watershed (**Figure 9**) and show a striking similarity to the land-use of the watershed (**Figure 1**) which can serve as a reasonable way to evaluate the quality of the clusters (Templ et al. 2008). In comparing the land-use percentages of each cluster, Cluster 1 had the greatest percentage of urban and suburban land use, Cluster 2 had the greatest percentage of agricultural land use, and Cluster 3 had a fairly even mix of all land-use types.

The results showed that Cluster 1 generally had the highest fall and spring DOC, highest spring temperature, and highest spring orthophosphate. Additionally, this cluster had low spring nitrate+nitrite-N, lower fall orthophosphate concentrations, and lower transparency than the other clusters. Cluster 2 was characterized by some of the highest fall and spring nitrate+nitrite-N, generally higher fall orthophosphate, and lower DOC and greater transparency values compared to cluster 1. Cluster 3 was the relatively "cleanest" cluster, having generally lower nutrients and DOC compared with the other two clusters while maintaining high transparency, average spring temperatures, and the lowest fall temperatures. The pH did not seem to vary greatly across the clusters.

Discussion and Conclusions
Citizen science data
The greatest challenge in comparing volunteer-determined and lab-analyzed water quality data was that the test strip methods of measuring nitrate+nitrite-N and orthophosphate as read by volunteers created categori-

cal datasets because volunteers picked values only on the scale provided within the strips. This can create a challenge in analyzing samples using common statistical methods. Similarly to Peckenham et al. (2012), we chose to address this issue by binning the continuous lab data into comparable categories; we then used multiple types of agreement analysis methods to compare the volunteer-read and lab-tested data. The results demonstrated that for nitrate+nitrite-N, volunteers were consistently able to estimate concentrations using field test strips with moderate to substantial agreement to lab values, although the potential biases of volunteer-read data were not evaluated. This is consistent with a study that demonstrated that nitrate test strips showed good precision when read by students (Peckenham and Peckenham 2014). Agreement between volunteer-read data and lab-analyzed data across the Blitz events and overall in this study supports the conclusion that citizen-collected data can be scientifically valid for water quality assessment, which is key to demonstrating the usefulness of citizen science (Bonter and Cooper 2012). In addition to providing meaningful data, the fact that participants were able to accurately evaluate on-site nitrate+nitrite-N concentrations enhances the educational outcomes for the Blitz participants (Jordan et al. 2012).

The validity of orthophosphate observations measured on-site by volunteers was more difficult to assess considering there was little variability of measurements within the test strip categories (most of the lab-measured orthophosphate concentrations were very low) and there was a consistent overestimation by the volunteers. Similarly to this finding, Peckenham and Peckenham (2014) found that overestimation occurred when using nitrate+nitrite test strips when actual concentrations were low. However, some of the overestimation in orthophosphate comparisons could also result from the fact that the test strips measured orthophosphate in unfiltered water (e.g., with more sediment-bound phosphate) and the lab analysis was performed on filtered samples (e.g., with less sediment-bound phosphate).
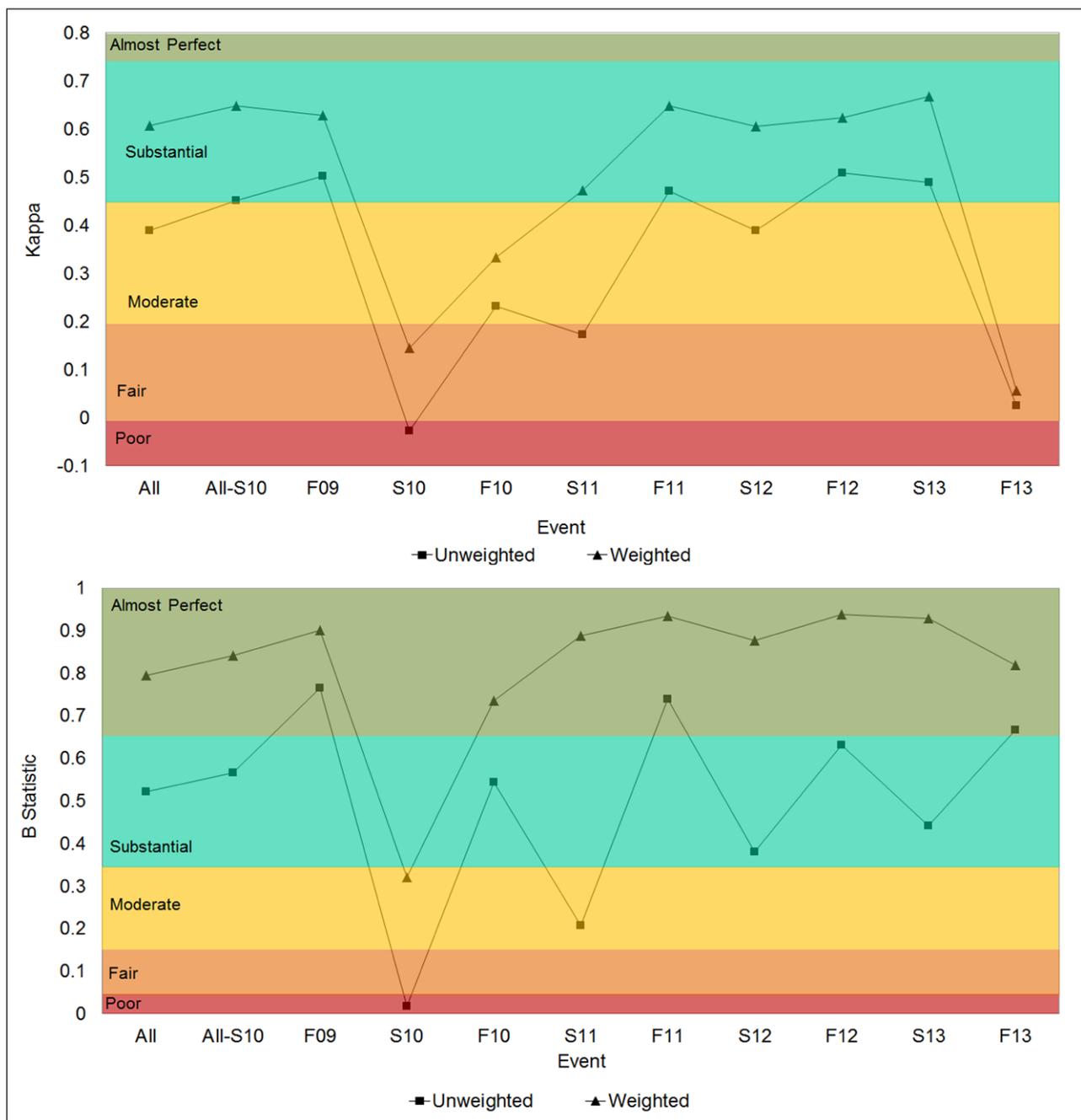
Figure 4: *B* statistics (top) and Cohen's Kappa statistics (bottom) determined for volunteer-read versus lab-determined nitrate+nitrite-N data agreement. Qualitative values are based on Munoz and Bangdiwala (1997).

This, combined with the fact that most of the data fell into the lowest test strip category, suggests that future orthophosphate testing—especially in this watershed—would benefit from test strips that have more categories between 0 and 5 mg/L and from testing samples that have been filtered for better comparison. Overall, the results support the idea that water quality data observed by volunteers can be acceptable for an educational experience and informative for watershed groups.

Cluster interpretation for watershed management
The cluster analysis and characterization was completed only for volunteer-collected water quality data. Similarly to other studies (Kim et al. 2005; Shrestha and Kazama 2006; Simeonov et al. 2003; Singh et al. 2004; Varol et al. 2012), the cluster analysis revealed unique primary management zones (**Figure 8**) that can be used to target education and conservation strategies in the future, as follows:

- Cluster 1: Urban/suburban management zone with high spring and fall DOC and generally lower nutrients and transparency.
- Cluster 2: Agricultural management zone with the highest nutrients and lower DOC.
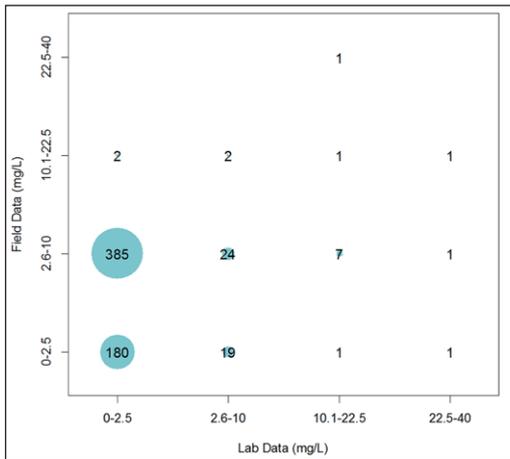- Cluster 3: Minimal management zone with the greatest transparency and lower nutrient and DOC concentrations.

Figure 5: Bubble plot showing agreement between field (volunteer readings) and lab (professional analyses) estimated orthophosphate for all Blitz events. The number inside the bubble indicates how many observations fell into that category.

These general relationships are useful for identifying the persistent water quality impacts associated with different land uses (Foley et al. 2005) and serve to confirm targeted conservation strategies within the watershed. For example, although nutrient management is certainly important for agricultural land (Vitousek et al. 2010), sediment pollution may be a greater problem in urban streams (Tayler and Owens 2009) to the extent that transparency reflects sediment load. In contrast, exceptions to the general land use pattern can help to identify areas which might have specific polluters that are unrelated to land use. For example, one sampling site that falls into Cluster 2 is primarily forested and urban land, not agriculture. This specific area is host to a golf course, and previous research has shown that nutrient loadings from golf courses can be similar to those of agriculture (King et al. 2007), which likely explains why the site would fall into a cluster with primarily agricultural land use. Another sampling site that is primarily forest and urban also was placed into Cluster 2. A small
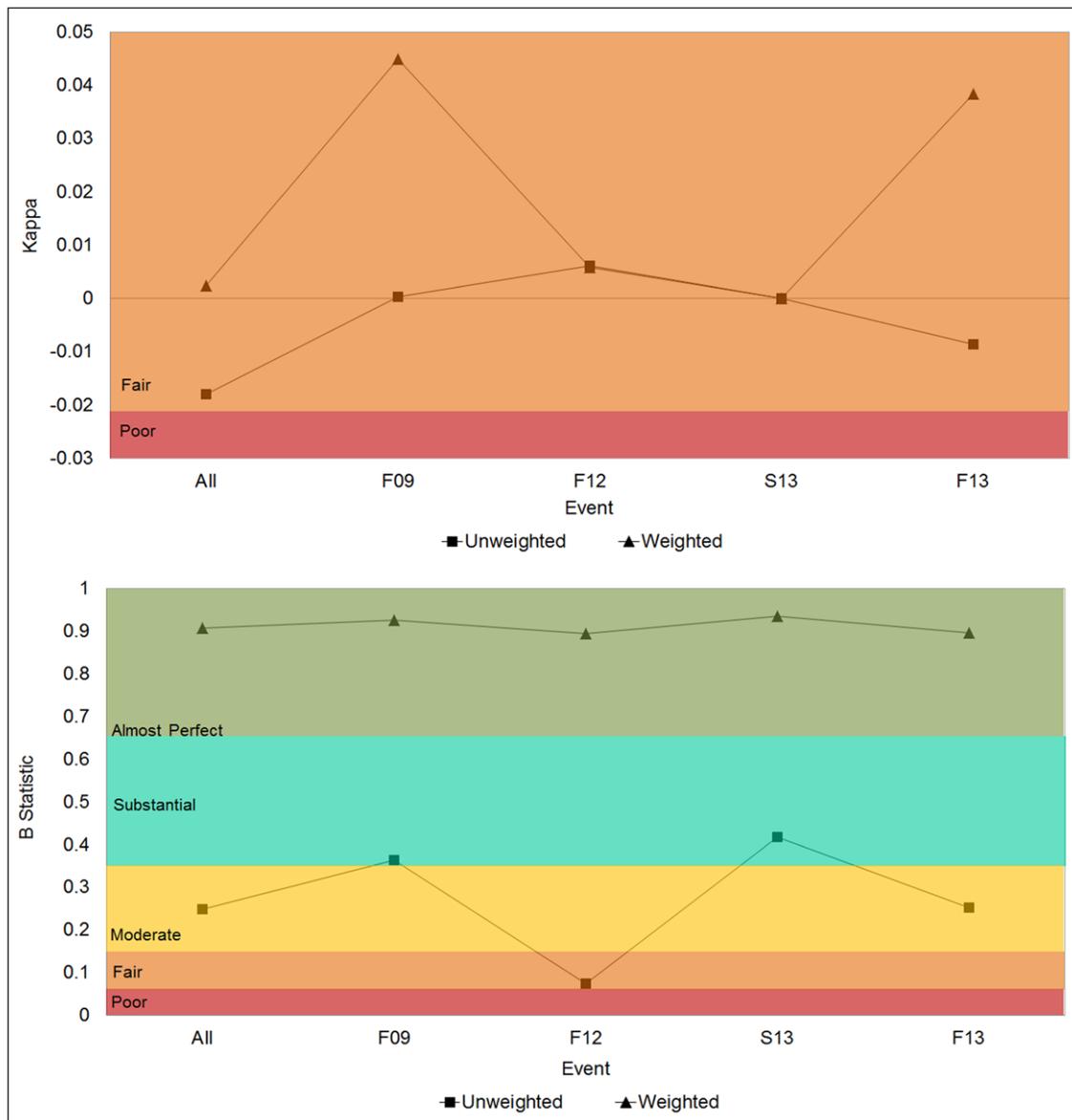
Figure 6: *B* statistics (top) and Cohen's Kappa statistics (bottom) determined for volunteer-read versus lab-determined orthophosphate data agreement. Qualitative values are based on Munoz and Bangdiwala (1997).
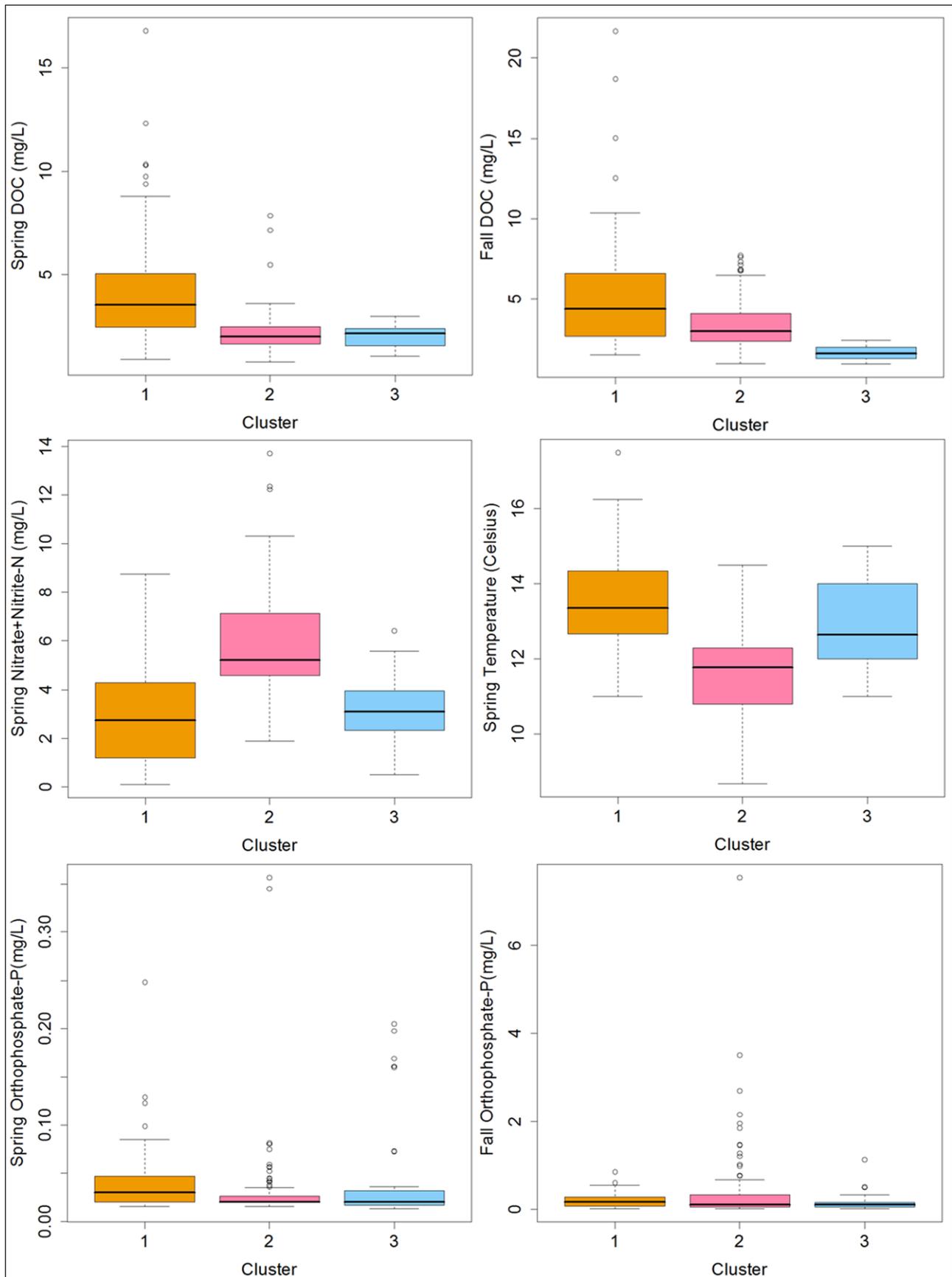
Figure 7: Boxplots of spring and fall DOC and orthophosphate, as well as spring temperature and nitrate+nitrite-N, grouped by cluster membership. These six variables had significant loading on the first three principal components. All other variable boxplots are provided in the Supplementary Materials.
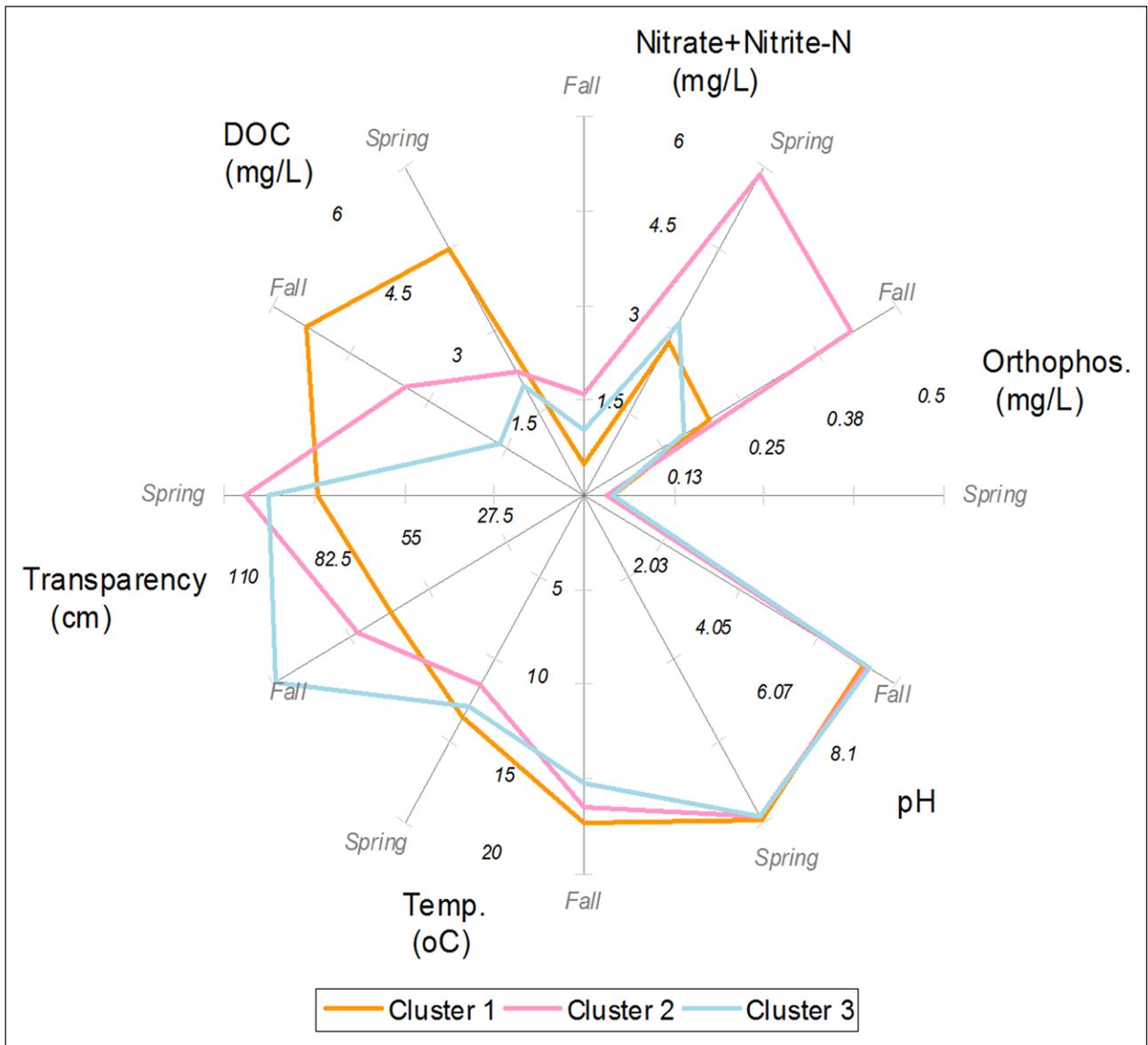
Figure 8: Spider plot displaying the mean of each cluster for all twelve variables.

wastewater treatment plant is located within this site, and the constant nitrate+nitrite-N signals likely explain its inclusion in this category. Overall, these clusters help provide WREC with insight into the specific water quality concerns seen throughout the watershed, so that management and education strategies can be improved (Brezonik et al. 1999).

The cluster results also can be used by WREC to determine which sites need no further testing given a shortage of volunteers or a reduction in budget (Wang et al. 2014). Lastly, this cluster analysis demonstrates how volunteer-collected and tested data (transparency, temperature, pH) can be used along with volunteer-collected and lab-tested data (nutrients, DOC) to perform more complex and informative analyses of water quality data.

Citizen science approach for water quality monitoring
Watershed groups exist in all parts of the US and the world, and many operate as nonprofit organizations (Lubell et al. 2002). By collaborating with local citizen scientists,

these groups can not only maximize their resources but also educate and involve the local community in water protection efforts (Cline and Collins 2003). Such groups, along with other citizen science-based organizations, are under increasing pressure to show the effectiveness of their programs (Conrad and Hilchey 2011). Overall, our research illustrates that citizen science-produced data can be highly valuable for use by watershed groups. Twice a year, hundreds of citizen scientists in Indiana help to sample 206 sites to provide a snapshot of water quality conditions in the Great Bend of the Wabash River Watershed that would otherwise not be achievable. By utilizing relatively inexpensive field test strips, volunteers are able to instantly evaluate the quality of the water they are sampling, which provides not only important data but a great educational opportunity. The test strips are inexpensive compared to lab analyses, and our analeyses show that they can be informative to water quality managers even when read by the members of the public. The cluster analysis provides a replicable example of how citi-
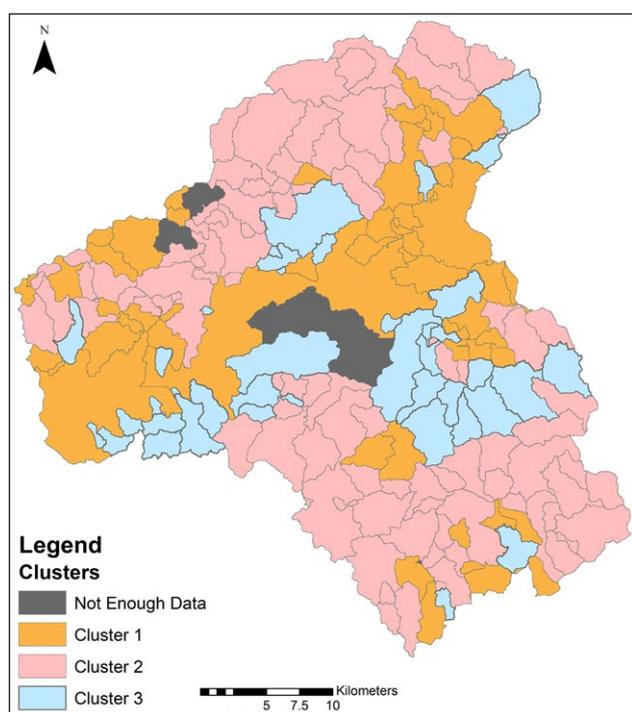
Figure 9: Watershed sub-basins mapped by cluster membership.

zen science-collected data can be used to further inform watershed management decisions. Overall, this work supports the increasing body of scientific knowledge demonstrating that citizen scientists can contribute worthwhile data which can easily be used in planning by watershed groups.

Supplementary Materials
Supplementary material relating to this article is available at http://dx.doi.org/10.5334/cstp.1.s1

Acknowledgements

Competing Interests
The authors declare that they have no competing interests.

References
Alberto, W.D., Diaz, M.D.P., Valeria, A.M., Fabiana, P.S., Cecilia, H.A. and De Los Angeles, B.M., 2001. Pattern recognition techniques for the evaluation of spatial and temporal variation in water quality. A case study: Suquia River Basin (Cordoba-Argentina). *Water Resources*, 35(12): 2881–2894. DOI: http://dx.doi.org/10.1016/S0043-1354(00)00592-3

Au, J., Bagchi, P., Chen, B., Martinez, R., Dudley, S.A. and Sorger, G.J., 2000. Methodology for public monitoring of total coliforms, *Escherichia coli* and toxicity in waterways by Canadian high school students. *Journal of Environmental Management*, 58(3): 213–230. DOI: http://dx.doi.org/10.1006/jema.2000.0323

Banerjee, M., Capozzoli, M., McSweeney, L. and Sinha, D., 1999. Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, 27(1): 3–23. DOI: http://dx.doi.org/10.2307/3315487

Bierman, P., Lewis, M., Ostendorf, B. and Tanner, J., 2011. A review of methods for analysing spatial and temporal patterns in coastal water quality. *Ecological Indicators*, 11(1): 103–144. DOI: http://dx.doi.org/10.1016/j.ecolind.2009.11.001

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T.B., Rosenberg, K.V. and Shirk, J., 2009. Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59: 977–984. DOI: http://dx.doi.org/10.1525/bio.2009.59.11.9

Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J. and Parrish, J.K., 2014. Next steps for citizen science. *Science*, 343(6178): 1436–1437. DOI: http://dx.doi.org/10.1126/science.1251554

Bonter, D.N. and Cooper, C.B., 2012. Data validation in citizen science: a case study from Project Feeder-Watch. *Frontiers in Ecology and the Environment*, 10(6): 305–307. DOI: http://dx.doi.org/10.1890/110273

Brezonik, P.L., Easter, K.W., Hatch, L., Mulla, D. and Perry, J., 1999. Management of diffuse pollution in agricultural watersheds: Lessons learned from the Minnesota River Basin. *Water Science and Technology*, 39(12): 323–330. DOI: http://dx.doi.org/10.1016/S0273-1223(99)00350-9

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T.C., Bastiaensen, J., De Bievre, B., Bhusal, J., Clark, J., Dewulf, A. and Foggin, M., 2014. Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Hydrosphere*, 2: 26. DOI: http://dx.doi.org/10.3389/feart.2014.00026

Canfield, D.E. Jr., Brown, C.D., Bachmann, R.W. and Hoyer, M.V., 2002. Volunteer lake monitoring: Testing the reliability of data collected by the Florida LAKEWATCH program. *Lake and Reservoir Management*, 18(1): 1–9. DOI: http://dx.doi.org/10.1080/07438140209353924

Cline, S.A. and Collins, A.R., 2003. Watershed associations in West Virginia: Their impact on environmental protection. *Journal of Environmental Management*, 67: 373–383. DOI: http://dx.doi.org/10.1016/S0301-4797(02)00222-0

Cohn, J.P., 2008. Citizen science: Can volunteers do real research? *BioScience*, 58(3): 192–197. DOI: http://dx.doi.org/10.1641/B580303

Conrad, C.C. and Hilchey, K.G., 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment*, 176: 273–291. DOI: http://dx.doi.org/10.1007/s10661-010-1582-5

Daughney, C.J., Raiber, M., Moreau-Fournier, M., Morgenstern, U. and van der Raaij, R., 2012. Use of hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality monitoring network: comparison of New Zealand's national and regional groundwater monitoring programs. *Hydrogeology Journal*, 20(1): 185–200. DOI: http://dx.doi.org/10.1007/s10040-011-0786-2

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T. and Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, 10(6): 291–297. DOI: http://dx.doi.org/10.1890/110236

Dickinson, J.L., Zuckerberg, B. and Bonter, D.N., 2010. Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41: 149–172. DOI: http://dx.doi.org/10.1146/annurev-ecolsys-102209-144636

Finlayson, C.M., D'Cruz, R., Aladin, N., Barker, D.R., Beltram, G., Brouwer, J., Davidson, N., Duker, L., Junk, W., Kaplowitz, M.D. and Ketelaars, H., 2005. Inland water systems. In: *Ecosystems and human well-being: Current state and trends*, Vol 1, pp. 553–583.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K. and Helkowski, J.H., 2005. Global consequences of land use. *Science*, 309(5734): 570–574. DOI: http://dx.doi.org/10.1126/science.1111772

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

Gamble, A. and Babbar-Sebens, M., 2012. On the use of multivariate statistical methods for combining in-stream monitoring data and spatial analysis to characterize water quality conditions in the White River Basin, Indiana, USA. *Environmental Monitoring and Assessment*, 184(2): 845–875. DOI: http://dx.doi.org/10.1007/s10661-011-2005-y

Güler, C., Thyne, G.D., McCray, J.E. and Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*, 10: 455–474. DOI: http://dx.doi.org/10.1007/s10040-002-0196-6

Hallgren, K.A., 2012. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1): 23–24.

Jordan, R.C., Ballard, H.L. and Phillips, T.B., 2012. Key issues and new approaches for evaluating citizen-science learning outcomes. *Frontiers in Ecology and the Environment*, 10(6): 307–309. DOI: http://dx.doi.org/10.1890/110280

Kim, J-H., Kim, R-H., Lee, J., Cheong, T-J., Yum, B-W. and Chang, H-W., 2005. Multivariate statistical analysis to identify the major factors governing groundwater quality in the coastal area of Kimje, South Korea. *Hydrological Processes*, 19: 1261–1276. DOI: http://dx.doi.org/10.1002/hyp.5565

Kim, S., Robson, C., Zimmerman, T., Pierce, J. and Haber, E.M., 2011. Creek Watch: Pairing usefulness and usability for successful citizen science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, British Columbia, May 7–12, 2011, pp. 2125–2134. DOI: http://dx.doi.org/10.1145/1978942.1979251

King, K.W., Balogh, J.C., Hughes, K.L. and Harmel, R.D., 2007. Nutrient load generated by storm event runoff from a golf course watershed. *Journal of Environmental Quality*, 36(4): 1021–1030. DOI: http://dx.doi.org/10.2134/jeq2006.0387

Kolok, A.S. and Schoenfuss, H.L., 2011. Environmental scientists, biologically active compounds, and sustainability: The vital role for small-scale science. *Environmental Science and Technology*, 45: 39–44. DOI: http://dx.doi.org/10.1021/es100455d

Kolok, A.S., Schoenfuss, H.L., Propper, C.R. and Vail, T.L., 2011. Empowering citizen scientists: The strength of many in monitoring biologically active environmental contaminants. *BioScience*, 61(8): 626–630. DOI: http://dx.doi.org/10.1525/bio.2011.61.8.9

Lottig, N.R., Wagner, T., Henry, E.N., Cheruvelil, K.S., Webster, K.E., Downing, J.A. and Stow, C.A., 2014. Long-term citizen-collected data reveal geographical patterns and temporal trends in lake water clarity. *PLoS One*, 9(4): e95769. DOI: http://dx.doi.org/10.1371/journal.pone.0095769

Lubell, M., Schneider, M., Scholz, J.T. and Mete, M., 2002. Watershed partnerships and the emergence of collective action institutions. *American Journal of Political Science*, 46(1): 148–163. DOI: http://dx.doi.org/10.2307/3088419

Maas, R.P., Kucken, D.J. and Gregutt, P.F., 1991. Developing a rigorous water quality database through a volunteer monitoring network. *Lake and Reservoir Management*, 7(1): 123–126. DOI: http://dx.doi.org/10.1080/07438149109354262

Mavukkandy, M.O., Karmakar, S. and Harikumar, P.S., 2014. Assessment and rationalization of water quality monitoring network: a multivariate statistical approach to the Kabbini River (India). *Environmental Science and Pollution Research*, 21(17): 10045–10066. DOI: http://dx.doi.org/10.1007/s11356-014-3000-y

Meyer, D., Zeileis, A., Hornik, K., Gerber, F. and Friendly, M., 2014. Package 'vcd', 22 December 2014. Available at http://cran.r-project.org/web/packages/vcd/vcd.pdf [Last accessed 19 January 2015].

Miller-Rushing, A., Primack, R. and Bonney, R., 2012. The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, 10(6): 285–290. DOI: http://dx.doi.org/10.1890/110278

Munoz, S.R. and Bangdiwala, S.I., 1997. Interpretation of Kappa and *B* statistics measures of agreement. *Journal of Applied Statistics*, 24(1): 105–111. DOI: http://dx.doi.org/10.1080/02664769723918

Najar, I.A. and Khan, A.B., 2012. Assessment of water quality and identification of pollution sources of three lakes in Kashmir, India, using multivariate analysis.

Environmental Earth Sciences, 66(8): 2367–2378. DOI: http://dx.doi.org/10.1007/s12665-011-1458-1

Nicholson, E.J., Ryan, J. and Hodgkins, D., 2002. Community data-where does the value lie? Assessing confidence limits of community collected water quality data. *Water Science & Technology*, 45(11): 193–200.

Nicosia, K., Daaram, S., Edelman, B., Gedrich, L., He, E., McNeilly, S., Shenoy, V., Velagapudi, A., Wu, W., Zhang, L. and Barvalia, A., 2014. Determining the willingness to pay for ecosystem service restoration in a degraded coastal watershed: A ninth grade investigation. *Ecological Economics*, 104: 145–151. DOI: http://dx.doi.org/10.1016/j.ecolecon.2014.02.010

Overdevest, C., Orr, C.H. and Stepenuck, K., 2004. Volunteer stream monitoring and local participation in natural resource issues. *Research in Human Ecology*, 11(2): 177–185.

Pati, S., Dash, M.K., Mukherjee, C.K., Dash, B. and Pokhrel, S., 2014. Assessment of water quality using multivariate statistical techniques in the coastal region of Visakhapatnam, India. *Environmental Monitoring and Assessment*, 186(10): 6385–6402. DOI: http://dx.doi.org/10.1007/s10661-014-3862-y

Peckenham, J.M. and Peckenham, S.K., 2014. Assessment of quality for middle level and high school student-generated water quality data. *Journal of the American Water Resources Association*, 50(6): 1477–1487. DOI: http://dx.doi.org/10.1111/jawr.12213

Peckenham, J.M., Thornton, T. and Peckenham, P., 2012. Validation of student generated data for assessment of ground water quality. *Journal of Science Education and Technology*, 21(2): 287–294. DOI: http://dx.doi.org/10.1007/s10956-011-9317-0

Savan, B., Morgan, A.J. and Gore, C., 2003. Volunteer environmental monitoring and the role of the universities: The case of citizens' environment watch. *Environmental Management*, 31(5): 561–568. DOI: http://dx.doi.org/10.1007/s00267-002-2897-y

Shirk, J.L., Ballard, H.L., Wilderman, C.C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B.V., Krasny, M.E. and Bonney, R., 2012. Public participation in scientific research: a framework for deliberate design. *Ecology and Society*, 17(2): 29. DOI: http://dx.doi.org/10.5751/ES-04705-170229

Shrestha, S. and Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, 22: 464–475. DOI: http://dx.doi.org/10.1016/j.envsoft.2006.02.001

Silvertown, J., 2009. A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9): 467–471. DOI: http://dx.doi.org/10.1016/j.tree.2009.03.017

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, T., 2003. Assessment of the surface water quality in Northern Greece. *Water Research*, 37: 4119–4124. DOI: http://dx.doi.org/10.1016/S0043-1354(03)00398-1

Singh, K.P., Malik, A., Mohan, D. and Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research*, 38: 3980–3992. DOI: http://dx.doi.org/10.1016/j.watres.2004.06.011

Singh, K.P., Malik, A. and Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. *Analytica Chimica Acta*, 538(1–2): 355–374. DOI: http://dx.doi.org/10.1016/j.aca.2005.02.006

Taylor, K.G. and Owens, P.N., 2009. Sediments in urban river basins: A review of sediment-contaminant dynamics in an environmental system conditioned by human activities. *Journal of Soils and Sediments*, 9(4): 281–303. DOI: http://dx.doi.org/10.1007/s11368-009-0103-z

Templ, M., Filzmoser, P. and Reimann, C., 2008. Cluster analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry*, 23: 2198–2213. DOI: http://dx.doi.org/10.1016/j.apgeochem.2008.03.004

United States Environmental Protection Agency, Office of Water, 2012. *Summary of Water Quality Assessments for Each Waterbody Type for Reporting Year 2012*. Available at http://ofmpub.epa.gov/waters10/attains_index.control (Last accessed 13 July 2015).

Varol, M., Gökot, B., Bekleyen, A. and Sen, B., 2012. Water quality assessment and apportionment of pollution sources of Tigris River (Turkey) using multivariate statistical techniques—A case study. *River Research and Applications*, 28(9): 1428–1438. DOI: http://dx.doi.org/10.1002/rra.1533

Viera, A.J. and Garrett, J.M., 2005. Understanding inter-observer agreement: The Kappa statistic. *Family Medicine*, 37(5): 360–3.

Vitousek, P.M., Naylor, R., Crews, M., David, M.B., Drinkwater, L.E., Holland, E., Johnes, P.J., Katzenberger, J., Martinelli, L.A., Matson, P.A. and Nziguheba, G., 2010. Nutrient imbalances in agricultural development. *Science*, 324(5934): 1519. DOI: http://dx.doi.org/10.1126/science.1170261

Wang, Y-B., Liu, C-W., Liao, P-Y. and Lee, J-J., 2014. Spatial pattern assessment of river water quality: implications of reducing the number of monitoring stations and chemical parameters. *Environmental Monitoring and Assessment*, 184(2): 1781–1792. DOI: http://dx.doi.org/10.1007/s10661-013-3492-9

Wang, Y., Wang, P., Bai, Y., Tian, Z., Li, J., Shao, X., Mustavich, L.C. and Li, B-L., 2013. Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. *Journal of Hydro-environment Research*, 7(1): 30–40. DOI: http://dx.doi.org/10.1016/j.jher.2012.10.003

Zhou, A., Tang, H. and Wang, D., 2005. Phosphorus adsorption on natural sediments: Modeling and effects of pH and sediment composition. *Water Research*, 39(7): 1245–1254. DOI: http://dx.doi.org/10.1016/j.watres.2005.01.026

]u[*Citizen Science: Theory and Practice* is a peer-reviewed open access journal published by Ubiquity Press.   OPEN ACCESS