CITIZEN SCIENCE AND BENTHIC MACROINVERTEBRATE POPULATION TRENDS IN THE MOUNTAINS OF WESTERN NORTH CAROLINA

by
Virginia Louise Hamilton

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science
in
Land Resources and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 2021

ACKNOWLEDGEMENTS

I would like to thank the many people who were patient with me while I worked to finish this degree. My advisor on this project, Dr. Adam Sigler for his invaluable guidance in writing R scripts, scientific writing coaching, and pushing me hard to produce something I was proud of. My committee for their valuable comments and review; Scott Powel, Bob Peterson, and Tracey Sterling. The good folks at EQI for providing me access to their data and for getting the public so excited about water quality, especially Robert Hastings, Anne-Marie Traylor, Dr. David Penrose, and Dr. Robert Zinna. My partner Armin Weise, for cooking dinners and keeping me on track. My dear friend and former boss Asher Wright, for letting me pursue a degree while we were managing a farm and molding minds. My former students, for keeping my work days lighthearted so my nights could be spent reading and working. My colleagues at Warren Wilson College for telling me that I was making the right choice. My beautiful friends for understanding why I was so busy and for being my support system. My parents and grandparents for paving the way, especially my grandfather Joseph LeConte Smith. I might not have gone into thermodynamics at MIT like you, but I'm sure you would be proud regardless. And a big thank you to the tiny creatures of the world, administering the myriad processes that allow me to look into a crystal clear pool in a trout stream in Appalachia and feel a true sense of wonder.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Citizen science and the efforts of volunteer data collectors provide an enormous amount of data to non-profits and government agencies at all levels across many scientific fields. Volunteer freshwater biomonitoring programs around the world have trained and used volunteers to collect, count, and identify benthic macroinvertebrates. Benthic macroinvertebrates have a variety of sensitivities to pollutants and certain groups' presence, absence, and population trends can be used to infer water quality. The Environmental Quality Institute (EQI) in western North Carolina has been collecting benthic macroinvertebrate data using volunteers for over a decade. Data from EQI were imported into a single data set using the statistical program R. These data were used to calculate volunteer identification accuracy and plot changes in benthic community composition across physiographic variables over time. Of the 71% of quality control samples that were of high quality, volunteers had 86% agreement in identifications when compared to a professional biologist. Volunteer data was used to assess differences in mean richness, abundance, and EPT richness at high and low elevation sites and at agricultural and non-agricultural sites. High elevation sites have significantly higher richness, abundance, and EPT richness, while high agriculture sites have significantly lower EPT richness. This suggests that the dataset is relatively robust and can support meaningful evaluation of trends in ecosystem health. As funding for government sponsored water quality monitoring continues to be decreased, well-trained volunteer monitors will need to continue providing supplemental data to decision makers. This project highlights the usefulness of organizations like EQI and the potential for the data that their volunteers collect.

CITIZEN SCIENCE AND BENTHIC MACROINVERTEBRATE POPULATION TRENDS IN THE MOUNTAINS OF WESTERN NORTH CAROLINA

Introduction

The Appalachian Mountains, stretching from Georgia to Maine, are one of the older mountain ranges on the planet. The oldest rocks in the southern Appalachians are estimated to be at least a billion years old (Faill, 1997). The Appalachians were formed by several episodes of continental collision and separation. The shallow sea that formed during separation, volcanic activity early in our planet's history, and the immense pressure created by the collision of tectonic plates created a patchwork of granite, limestone, sandstone, and metasedimentary rocks across the region. Subsequent erosion exposed layers buried millions of years ago and caused high variation in soil types and conditions (Clark, 2001).

The southern end of this mountain chain was never glaciated during the advance and retreat of the Laurentide ice sheet 18,000 years BP. The cold and dry climate from the leading edge of the glacier facilitated the southward movement of boreal communities from higher latitudes, the remnants of which can still be seen today on the crest of the Blue Ridge (Carroll et al., 2002). The preservation of the plethora of Pleistocene species in addition to the rapidly changing environment of glacial advance and retreat resulted in highly heterogenous habitat currently supporting the highest levels of biodiversity in the country (Delcourt & Delcourt, 1998). Precipitation in western North Carolina is driven by both marine and subtropical airmasses, leading to precipitation rates up to twice the contiguous national average of 76 cm (between 109 cm and 180 cm in the western counties) (*NOAA*, 2021). These mountain landscapes are shaped by dense networks of springs, streams, and rivers and are heavily influenced by high mean annual precipitation rates (Kozar et al., 2001).

More than 6,400 square kilometers (1.6 million acres) of land are protected in western North Carolina, meaning that the headwaters of the majority of the area's waterways are protected. The region has one of the highest concentrations of protected public land on the east coast (*USFS*, 2013). Once water leaves these pristine areas, a suite of other factors control its quality. Modern land use in western North Carolina is highly varied and supports industries including row crop farming, beef cattle production, paper product manufacturing, outdoor recreation, and craft beverage production. These industries rely heavily on the region's water resources, which continue to be a major draw for economic development in the region. In addition to supporting local economies, these waterbodies are home to a diverse suite of freshwater organisms.

Benthic macroinvertebrates represent a large fraction of the taxonomic families representing overall freshwater aquatic diversity in North Carolina streams. Benthic macroinvertebrates are small aquatic invertebrates that are visible without the aid of a microscope. They spend all or part of (in the case of some insect larvae) their lives under water (US EPA, 2013). These species inhabit a variety of stream and river types and make use of the heterogeneous nature of stream beds and stream banks. They occupy a broad spectrum of ecological niches within the stream channel, from predator to decomposer (Poff et al., 2006). These species also display a variety of tolerance to various physical conditions, chemical and biological pollutants, and physical or chemical water qualities such as dissolved oxygen concentration and pH. Some species may have a very wide tolerance while other species are only found under a set of specific conditions (J. G. Smith et al., 2011).

It is for this reason that these creatures have become important in the field of biomonitoring. Biomonitoring is defined as "the systematic use of living organisms or their responses to determine the condition or changes of the environment" (Oertel & Salánki, 2003). While instantaneous chemical monitoring can provide information about a specific set of parameters, it does not effectively capture cumulative effects of stream conditions caused by a wide variety of influences (Li et al., 2010). The utility of biomonitoring for providing a time integrated sample, as well as being a more direct quantification of aquatic ecosystem health has led to government agencies adopting these methods in addition to chemical monitoring.

Water quality in western North Carolina is monitored at the state level by the North Carolina Department of Environmental Quality's Division of Water Resources (NC DEQ). The NC DEQ is responsible for implementing federal policies that oversee safe drinking water requirements. The biological assessment branch of the division is responsible for evaluating water quality by using the biological communities of those waterbodies as indicators (*NC DEQ: Water Sciences Section*, n.d.). NC DEQ employees use kick net, sweep net, leaf-pack, log wash, and visual surveys to collect insects and identify them to species level as detailed in the NC DEQ's Standard Operating Procedures for the Collection and Analysis of Benthic Macroinvertebrates document (*NC DEQ*, 2016). The resulting data are presented in a publicly accessible map that reports the Ephemeroptera-Plecoptera-Trichoptera (EPT) taxa richness, Biotic Index, and Bio-classification score for the waterbody in question (*NC DENR*, 2021). Due to limited staff and an ever-shrinking budget, NC DEQ sampling occurs with less frequency and at fewer locations each year (Ross, 2020). As in many places across the United States and around the world, citizen science programs have been implemented to supplement the data that the NC DEQ and other regulatory agencies are able to provide.

The Environmental Quality Institute (EQI) began as a state funded branch of the University of North Carolina system in 1988. The Environmental Quality Institute was formed to offer another source of high-quality analytical data to fill in the gaps created by chronic underfunding in the state regulatory bodies. The University of North Carolina ended the program due to budget cuts in 2009, and EQI was reformed as a standalone non-profit in 2010 (Postelle, 2009). The EQI uses a network of trained volunteers to sample more than 150 sites monthly and tests them for pH, alkalinity, turbidity, total suspended solids, conductivity, ammonia-nitrogen, nitrate/nitrite-nitrogen, orthophosphate, total phosphorus, and fecal coliform analysis in a lab certified to NC Wastewater laboratory standards (*NC DEQ: Laboratory Certification Branch*, 2021).

In addition to laboratory analyses under their Volunteer Water Information Network (VWIN) program, EQI began monitoring benthic macroinvertebrates under their Stream Monitoring Information Exchange (SMIE) project. Volunteers in the SMIE program attend at least six hours of training in identifying macroinvertebrates and assist teams of more experienced data collectors in monitoring more than 70 sites twice per year (Figure 1). Data are reported to EQI as the abundances of 46 family and genera groups found using three collection methods (kick net, leaf-pack, and visual survey). Data from volunteers are used by EQI employees to calculate multiple indices describing stream health. This data is publicly available upon request and is regularly used by municipalities and non-profits to assess water quality and inform decision making processes. EQI's SMIE data is considered "Tier 2" data by the Environmental Protection Agency, meaning that the data are not considered robust enough for citation in regulatory documents or court proceedings (Penrose & Call, 1995).

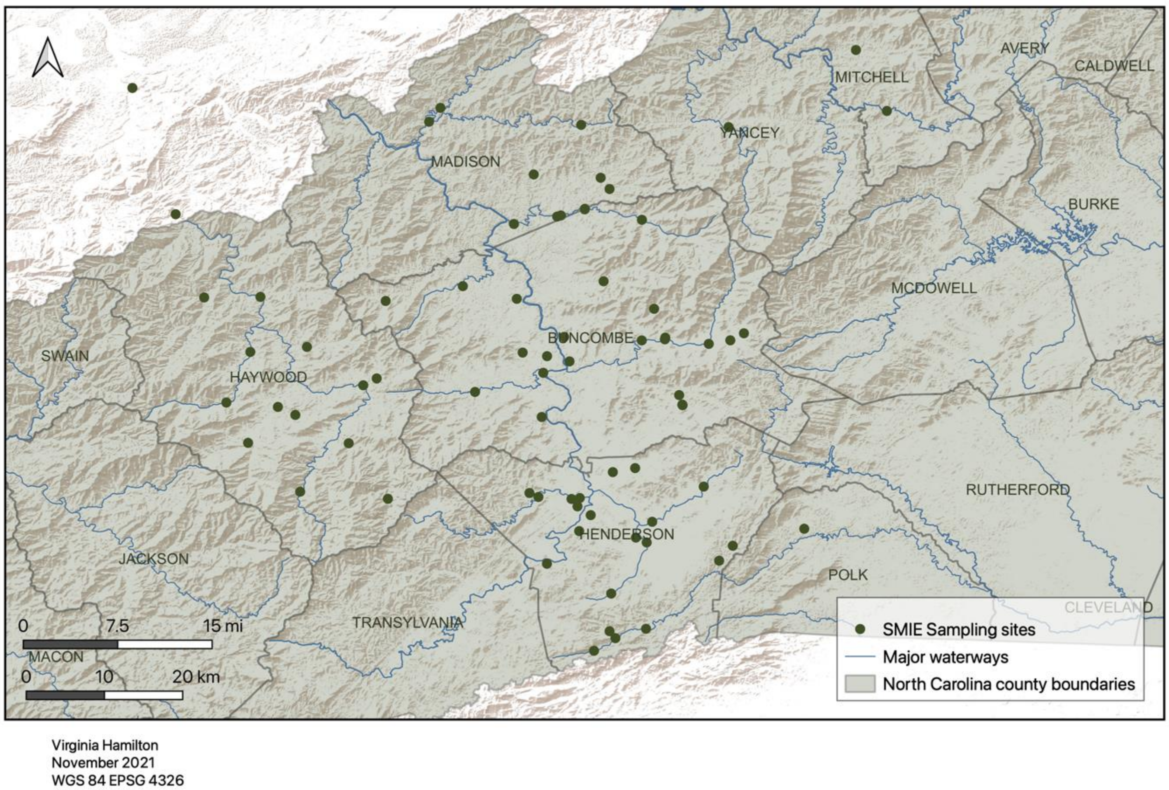


Figure 1: The Environmental Quality Institute's Stream Monitoring Information Exchange monitoring sites. The shaded area represents the boundary of the state of North Carolina and its counties. Counties are labeled. Major waterways appear in blue. SMIE sampling sites appear in dark green.

In 2020, two major citizen science organizations in the region, including EQI, reported an alarming drop in numbers of benthic macroinvertebrates in the lower end of the Mills River Watershed in Henderson County, North Carolina. The Mills River Watershed drains a portion of the Pisgah National Forest. With the Blue Ridge Parkway and Mount Pisgah to the north, water drains from high elevations into the community of Mills River via the North and South Forks of the Mills River. The Mills River supports a large agricultural community in the valley and supplies drinking water for Henderson County, one of the most populous counties in western North Carolina. Maintenance of clean water in the Mills River watershed is a high priority for the community for a number of reasons, including human drinking water, industry, and recreation. One survey team reported only eight individual macroinvertebrates in a kick net collection, causing several partner organizations to consider the possibility that an external stressor such as agricultural pollution was affecting the watershed (personal communication – Hamilton & EQI, 2021). The low macroinvertebrate count prompted an assessment of the VWIN program's data, which revealed no water quality results for the Mills River that violated water quality standards for drinking water (*North Carolina Division of Water Resources Surface Water Quality Standards, Criteria & In-Stream Target Values*, 2019).

Was this low number of benthic macroinvertebrates an indicator of an unknown and unmonitored stressor in the watershed, or simply the result of volunteer sampling error? This incident inspired the following questions that are addressed in my work; Are data collected by volunteers accurate when compared to a professional biologist? Are there differences in richness, abundance, and EPT richness across different physiographic variables in western North Carolina?

Methods

Environmental Quality Institute Methods

In the Field

The raw data used in my study was obtained from the EQI's biomonitoring efforts. Surveys are conducted twice a year, once in the spring and once in the fall. Two classes of volunteer are assigned to each survey site. Group Leaders are volunteers with more volunteer hours who have passed a test requiring them to correctly identify invertebrate specimens with high accuracy. At every site, a group leader is paired with other regular volunteers. These regular volunteers have attended more than six hours of volunteer training. Regular volunteers receive approximately four hours of identification and sampling procedure training either via a video meeting platform or in person, and then head to a demo site to perform a stream sample. At each site, the stream is surveyed using three collection methods: the kick net, leaf-pack, and visual survey. These methods are consistent with methods used by the NCDEQ and other citizen science efforts across the United States (*Standard Operating Procedures for the Collection and Analysis of Benthic Macroinvertebrates*, 2016). They were chosen for their concurrency with regulatory sampling methods that produce Type 1 data, their ease of use by non-professionals, and the way that each method samples a different area of the stream strata.

The kick net sample is taken by selecting a riffle in the stream. This is defined as a rocky area with turbulent water. The water should be shallow enough for the volunteers to move safely. The kick net is a 500 μm mesh net attached to two poles. The downstream volunteer installs the kick net by placing the two poles in the stream bed perpendicular to the current and ensuring that the bottom of the net is in contact with the bottom of the stream. The bottom of the net is weighted down with several stones. The upstream volunteer holds the net in such a way that the net curves with the current and no water escapes over the top of the net. The upstream volunteer uses their feet to disturb rocks in the streambed and uses their hands to “wash” larger rocks to dislodge small organisms. This is done in approximately a one and a half meter by one meter rectangle upstream of the kick net for sixty seconds, after which the downstream volunteer carefully removes the kick net from the stream in an upward sweeping motion to avoid washing away any organisms now caught in the net. The net is taken to a card table or white plastic sheet to be picked for 20 minutes. Specimens are placed in ice cube trays filled with stream water to await identification.

Leaf-pack sampling is conducted next. A volunteer takes a white pan and walks the stream sampling location for three minutes, removing masses of organic debris caught in the streambed behind rocks or logs and placing them in the pan. Stream water is added to the pan and agitated to dislodge organisms from the organic material. Individual pieces of organic detritus are washed over the white pan with stream water to ensure collection of all organisms present in the sample. The remaining water in the pan (containing the organisms) can be strained through a 500 μm bucket sieve. Organisms remaining in the sieve are separated into ice cube trays by type or similar appearance for further identification.

The final sample collected is a visual survey. Two volunteers search the stream area for habitats like seeps, logs, or large boulders. At each suitable habitat the volunteers search for organisms for five minutes. Any visible organisms are removed from the substrate and placed in a white pan filled with stream water. During picking, these organisms are removed and placed in white ice cube trays for identification. Great care is taken to keep specimens from the different sampling methods separate. The group leader is responsible for recording the team's findings on the "Field Data Sheet". These data will be referred to in this paper as "abundance data".

Organisms are identified to one of 43 groups that EQI has defined. These groups are generally a family or group of families that are physically similar and occupy similar niches in the stream ecosystem. Although professional organizations may identify organisms to species, this requires a more specialized skillset, and, often, high powered magnification. The EQI groups are named using descriptive titles that are easy for non-scientists to understand (Table 1).

Table 1: The Environmental Quality Institute's 43 taxa groups with their corresponding common names, orders, and families.

Group Name	*Common Name*	*Order*	*Families*
giant shredder	Stonefly	Plecoptera	Pteronarcyidae
roach shredder	Stonefly	Plecoptera	Peltoperlodae
quick crawling predator	Stonefly	Plecoptera	Perlidae, Chloroperlidae, Perlodidae
fragile detritivore	Stonefly	Plecoptera	Capniidae, Leuctridae, Nemouridae, Taenopterygidae
flattened scrapers	Mayflies	Ephemeroptera	Heptageniidae
spiny crawler	Mayflies	Ephemeroptera	Ephemerellidae, Caenidae, Tricorythidae, Neophemeridae
round headed swimmer	Mayflies	Ephemeroptera	Ameletidae, Baetidae, Leptophlebiidae, Siphlonuridae
burrowing mayflies	Mayflies	Ephemeroptera	Ephemeridae
spiny turtle mayfly	Mayflies	Ephemeroptera	Baetiscidae
filter mayfly	Mayflies	Ephemeroptera	Oligoneuridae, Isonychiidae
net spinner	Caddisflies	Trichoptera	Hydropsychidae, Psychomyiidae
small head caddis	Caddisflies	Trichoptera	Rhyacophilidae
stick bait caddis	Caddisflies	Trichoptera	Limnephilidae, Genus Pycnopsyche
square log cabin caddis	Caddisflies	Trichoptera	Brachycentridae
sand and stick case caddis	Caddisflies	Trichoptera	Leptoceridae
vegetative case caddis	Caddisflies	Trichoptera	All other organic cased caddisflies.
gravel coffin case caddis	Caddisflies	Trichoptera	Thremmatidae, Glossosomatidae
sand snail case	Caddisflies	Trichoptera	Helicopsychidae
sand or mineral case caddis	Caddisflies	Trichoptera	All other inorganic cased caddisflies
water penny	Water Penny	Coleoptera	Psephenidae
predator beetle	Diving beetle, water tiger	Coleoptera	Dytiscidae, Hydrophilidae, Gyrinidae

Table 1: continued

Group Name	*Common Name*	*Order*	*Families*
adult riffle beetle	Riffle beetle	Coleoptera	Elmidae
larval riffle beetle	Riffle beetle	Coleoptera	Elmidae
hellgrammite	Dobsonflies	Megaloptera	Corydalidae
fishfly	Fishflies	Megaloptera	Corydalidae, subfamily Chauliodinae
alderfly	Alderflies	Megaloptera	Sialidae
oligochaete	Worms	Oligochaeta	Lumbrichulidae, Tubificidae, Nadidae
leech	Leeches	Euhirudinea	Hirudinidae
watersnipe	Water snipe flies	Diptera	Athericidae
water worm	Worms	Diptera	Tipulidae
fat head cranefly	Crane fly, mosquito hawk	Diptera	Tipulidae
chironomid midge	Midge	Diptera	Chironomidae
red midge	Red midge, bloodworm	Diptera	Chironomidae
blackfly	Blackfly	Diptera	Simuliidae
crayfish	Crayfish, crawdad	Decapoda	Cambaridae
sowbug isopod	Sowbug, Pillbug	Crustacea	Asellidae
scud amphipod	Scud	Crustacea	Gammaridae
coiled left face snail	Snail	Basommatophora	Physidae
coiled right face snail	Snail	var.	Pleuroceridae, Hydrobiidae, Lymnaeidae, Viviparidae
rounded right face snail	Snail	Basommatophora	Planorbidae
clams and mussels	Clam, Mussel	var.	Corbiculidae, Sphaeriidae, Unionidae
damselfly	Damselfly	Zygoptera	Coenagrionidae, Calopterygidae, Lestidae
dragonfly	Dragonfly	Odonata	Aeshnidae, Cordulesgastridae, Cordulidae, Gomphidae, Libellulidae, Macromiidae

Quality Control Procedures

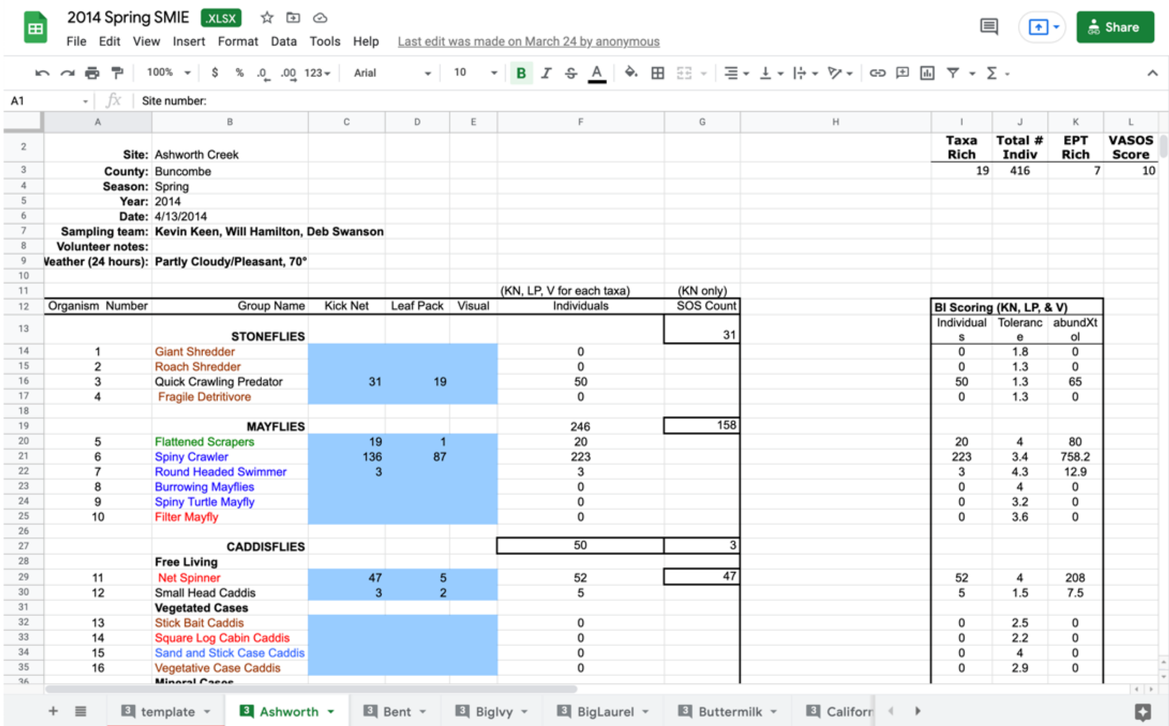
EQI aims to perform quality control on about 10 sites per season, or on average about 20% of total sites sampled. These sites tend to be sites that have been in the program the longest, and many of them repeat year to year. At these sites, the group leader is instructed to preserve every specimen that they encounter in vials filled with alcohol and labeled with the collection method. If any specimens are not preserved (too rare, too large), they must be noted on the data sheet (Robinson, 2014). Specimens are sent to Dr. David Penrose for professional identification (*Dave's Background* | Penrose Environmental Consulting Website). The biologist's identifications are recorded alongside the original volunteer observations in a separate excel worksheet. Any major discrepancies are used to prompt retraining or recertification of group leaders. This dataset will be referred to in this paper as “quality control data”.

In the Office

Field data sheets are transcribed into Excel spreadsheets by an EQI employee. As of Fall 2021, EQI’s practice is to store the data in multiple Microsoft Excel files. Each sampling season (ex: Fall 2011) is recorded in a separate workbook, and each sample location (ex: Ashworth Creek) is recorded in its own sheet. The total number of abundance data workbooks initially shared with me was 30. The Quality Control data are stored in a single excel workbook with multiple sheets representing sampling seasons. The data may be compiled into an official annual report depending on employee availability and labor budgeting. All data are available to the public, advocacy groups, or government groups upon request in the form of a copy of the Excel document.

Processing EQI’s Data

My study examines data from 2005 to 2016. In Spring 2017, the way that taxa groups are described by the organization was changed by the organization at the recommendation of consulting biologists. Some groups were split into two, resulting in a higher number of taxa groups. I chose the dataset from 2005 to 2016 to avoid mixing organism naming practices. The original raw data for this study is in the form of Excel spreadsheets with a structure that closely mirrors the data collection sheets that volunteers bring into the field (Figure 2).



2014 Spring SMIE .XLSX

File Edit View Insert Format Data Tools Help Last edit was made on March 24 by anonymous

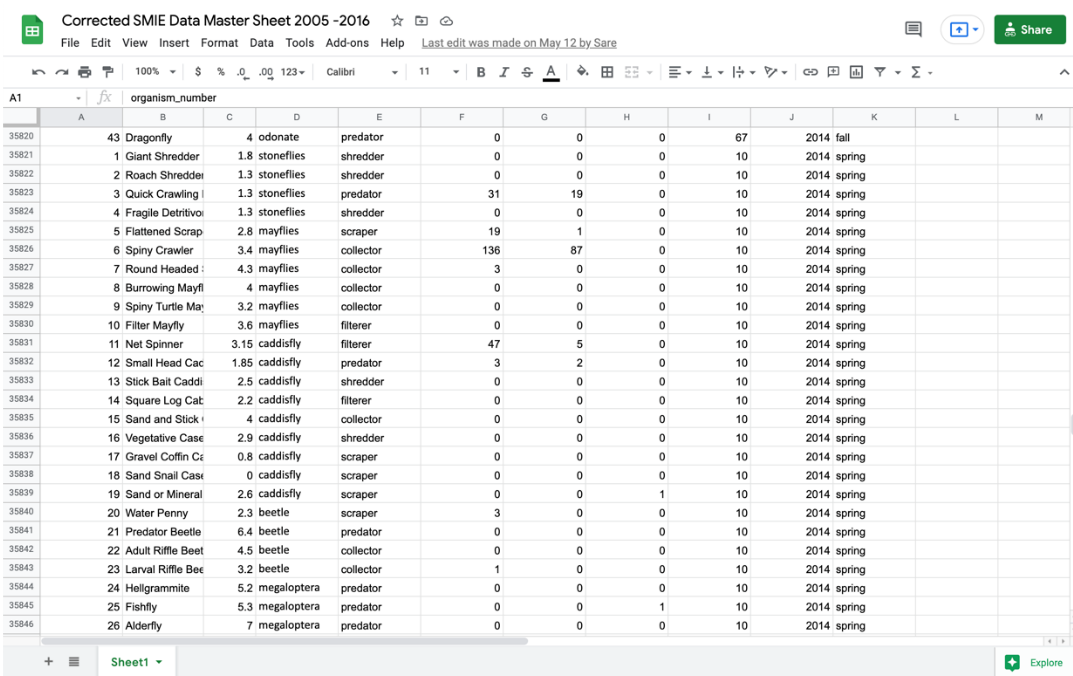
A1 Site number:

	A	B	C	D	E	F	G	H	I	J	K	L
2		Site: Ashworth Creek							Taxa Rich	Total # Indiv	EPT Rich	VASOS Score
3		County: Buncombe							19	416	7	10
4		Season: Spring										
5		Year: 2014										
6		Date: 4/13/2014										
7		Sampling team: Kevin Keen, Will Hamilton, Deb Swanson										
8		Volunteer notes:										
9		Veather (24 hours): Partly Cloudy/Pleasant, 70°										
10												
11						(KN, LP, V for each taxa)	(KN only)					
12	Organism Number	Group Name	Kick Net	Leaf Pack	Visual	Individuals	SOS Count		BI Scoring (KN, LP, & V)			
13		STONEFLIES					31		Individuals	Tolerance	abundXtol	
14	1	Giant Shredder				0			0	1.8	0	
15	2	Roach Shredder				0			0	1.3	0	
16	3	Quick Crawling Predator	31	19		50			50	1.3	65	
17	4	Fragile Detritivore				0			0	1.3	0	
18												
19		MAYFLIES				246	158					
20	5	Flattened Scrapers	19	1		20			20	4	80	
21	6	Spiny Crawler	136	87		223			223	3.4	758.2	
22	7	Round Headed Swimmer	3			3			3	4.3	12.9	
23	8	Burrowing Mayflies				0			0	4	0	
24	9	Spiny Turtle Mayfly				0			0	3.2	0	
25	10	Filter Mayfly				0			0	3.6	0	
26												
27		CADDISFLIES				50	3					
28		Free Living										
29	11	Net Spinner	47	5		52	47		52	4	208	
30	12	Small Head Caddis	3	2		5			5	1.5	7.5	
31		Vegetated Cases										
32	13	Stick Bait Caddis				0			0	2.5	0	
33	14	Square Log Cabin Caddis				0			0	2.2	0	
34	15	Sand and Stick Case Caddis				0			0	4	0	
35	16	Vegetative Case Caddis				0			0	2.9	0	

template | Ashworth | Bent | BigIvy | BigLaurel | Buttermilk | Califorr

Figure 2: Example of EQI's Excel sheet structure for Ashworth Creek, Spring 2014.

Although this structure makes data entry and index calculations easy and potentially reduces transcription errors, it causes difficulty for analyzing and importing the data into R. To facilitate ease of analysis, all data would need to be in a standardized format. The first step of data curation was completed by an EQI board member, Dr Robert Zinna of Mars Hill University with the help of an intern. During this first step, individual Excel sheets in the abundance dataset were compiled into a single master data sheet representing all observations from 2005 to 2016 (Figure 3).



Corrected SMIE Data Master Sheet 2005 -2016

File Edit View Insert Format Data Tools Add-ons Help Last edit was made on May 12 by Sare

A1 organism_number

	A	B	C	D	E	F	G	H	I	J	K	L	M
35820	43	Dragonfly	4	odonate	predator	0	0	0	67	2014	fall		
35821	1	Giant Shredder	1.8	stoneflies	shredder	0	0	0	10	2014	spring		
35822	2	Roach Shredder	1.3	stoneflies	shredder	0	0	0	10	2014	spring		
35823	3	Quick Crawling I	1.3	stoneflies	predator	31	19	0	10	2014	spring		
35824	4	Fragile Detritivo	1.3	stoneflies	shredder	0	0	0	10	2014	spring		
35825	5	Flattened Scrap	2.8	mayflies	scraper	19	1	0	10	2014	spring		
35826	6	Spiny Crawler	3.4	mayflies	collector	136	87	0	10	2014	spring		
35827	7	Round Headed	4.3	mayflies	collector	3	0	0	10	2014	spring		
35828	8	Burrowing Mayfl	4	mayflies	collector	0	0	0	10	2014	spring		
35829	9	Spiny Turtle May	3.2	mayflies	collector	0	0	0	10	2014	spring		
35830	10	Filter Mayfly	3.6	mayflies	filterer	0	0	0	10	2014	spring		
35831	11	Net Spinner	3.15	caddisfly	filterer	47	5	0	10	2014	spring		
35832	12	Small Head Cac	1.85	caddisfly	predator	3	2	0	10	2014	spring		
35833	13	Stick Bait Caddi	2.5	caddisfly	shredder	0	0	0	10	2014	spring		
35834	14	Square Log Cab	2.2	caddisfly	filterer	0	0	0	10	2014	spring		
35835	15	Sand and Stick	4	caddisfly	collector	0	0	0	10	2014	spring		
35836	16	Vegetative Case	2.9	caddisfly	shredder	0	0	0	10	2014	spring		
35837	17	Gravel Coffin Ca	0.8	caddisfly	scraper	0	0	0	10	2014	spring		
35838	18	Sand Snail Case	0	caddisfly	scraper	0	0	0	10	2014	spring		
35839	19	Sand or Mineral	2.6	caddisfly	scraper	0	0	1	10	2014	spring		
35840	20	Water Penny	2.3	beetle	scraper	3	0	0	10	2014	spring		
35841	21	Predator Beetle	6.4	beetle	predator	0	0	0	10	2014	spring		
35842	22	Adult Riffle Beet	4.5	beetle	collector	0	0	0	10	2014	spring		
35843	23	Larval Riffle Bee	3.2	beetle	collector	1	0	0	10	2014	spring		
35844	24	Hellgrammite	5.2	megaloptera	predator	0	0	0	10	2014	spring		
35845	25	Fishfly	5.3	megaloptera	predator	0	0	1	10	2014	spring		
35846	26	Alderfly	7	megaloptera	predator	0	0	0	10	2014	spring		

Sheet1 Explore

Figure 3: Example of compiled data from Mars Hill team, showing Ashworth Creek for Spring, 2014.

My project used that master data sheet as a starting point for further processing. Several common errors prevented a simple import into R. The most common error was a discrepancy in character counts resulting from inadvertent inclusion of spaces that are not readily visible in Excel. For example, R will read in “2005” (four characters) and “2005 ” (five characters) as two different factor levels. Other errors included blank cells, cells with value “NA” instead of 0, and string class characters appearing in numeric columns. Once the issues in this dataset were manually corrected, I used R to structure the data according to Tidy Data principles (described below). EQI provided an Excel sheet with site coordinates and another sheet with site qualities such as elevation, land use, watershed area, county, and site sponsor. These sheets were merged into a single data frame to provide more variables for analysis (Figure 4).

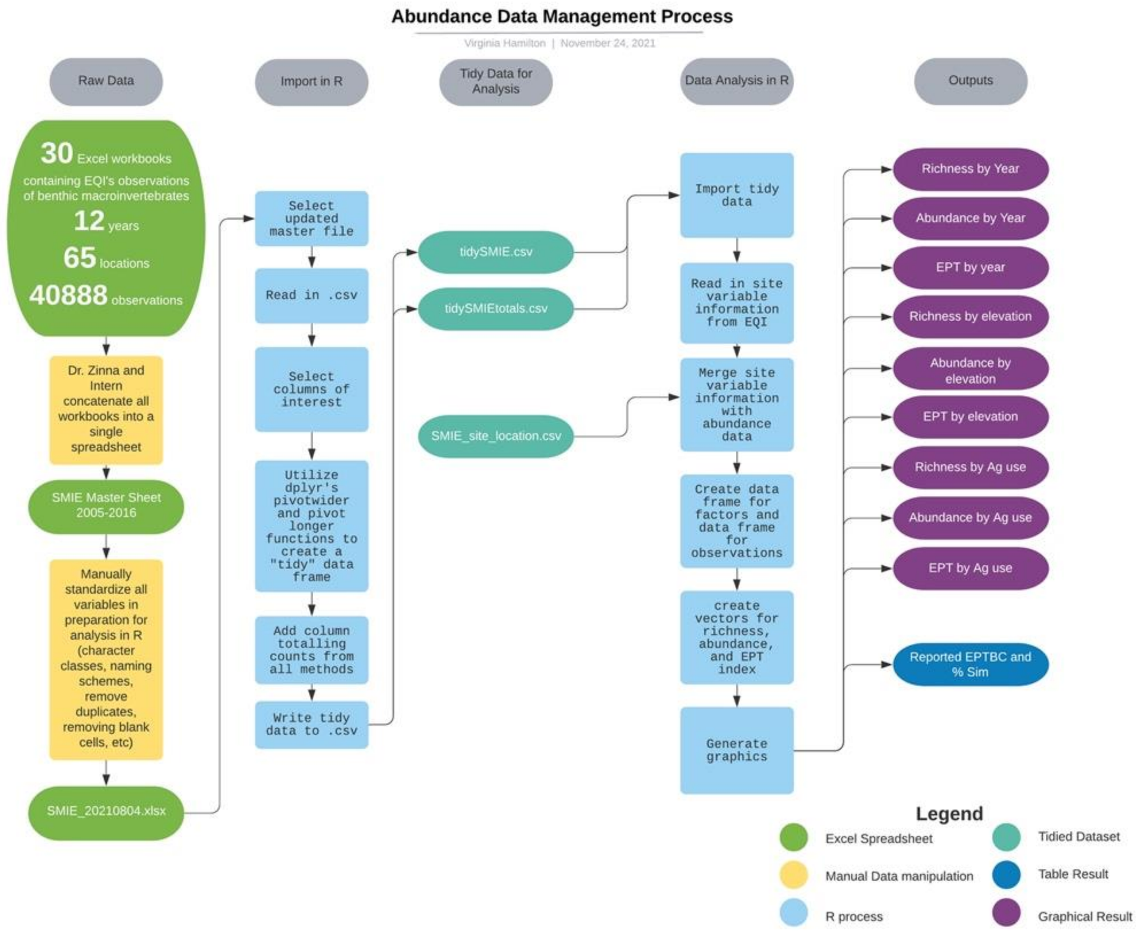


Figure 4: Abundance Data management process.

The quality control dataset was also restructured into a single Excel worksheet. To expedite the process and decrease the likelihood of transcription errors, Dr Sigler and I wrote a script in the program R to extract the necessary information from each Excel workbook and compile it into a single data frame. The original Excel file was configured with multiple tables next to one another in a single worksheet, which would be problematic for an automatic import to R. First, I separated the large quality control dataset into individual workbooks within Excel by season with a new worksheet for each location. New worksheets were reviewed to ensure that moving data did not affect formulas or other characteristics of the original dataset. The necessary cells were extracted from all files in the specified folder into a single data frame using a "for loop". I reorganized this data frame into a tidy data format using pivot functions from the dplyr library (Figure 5).

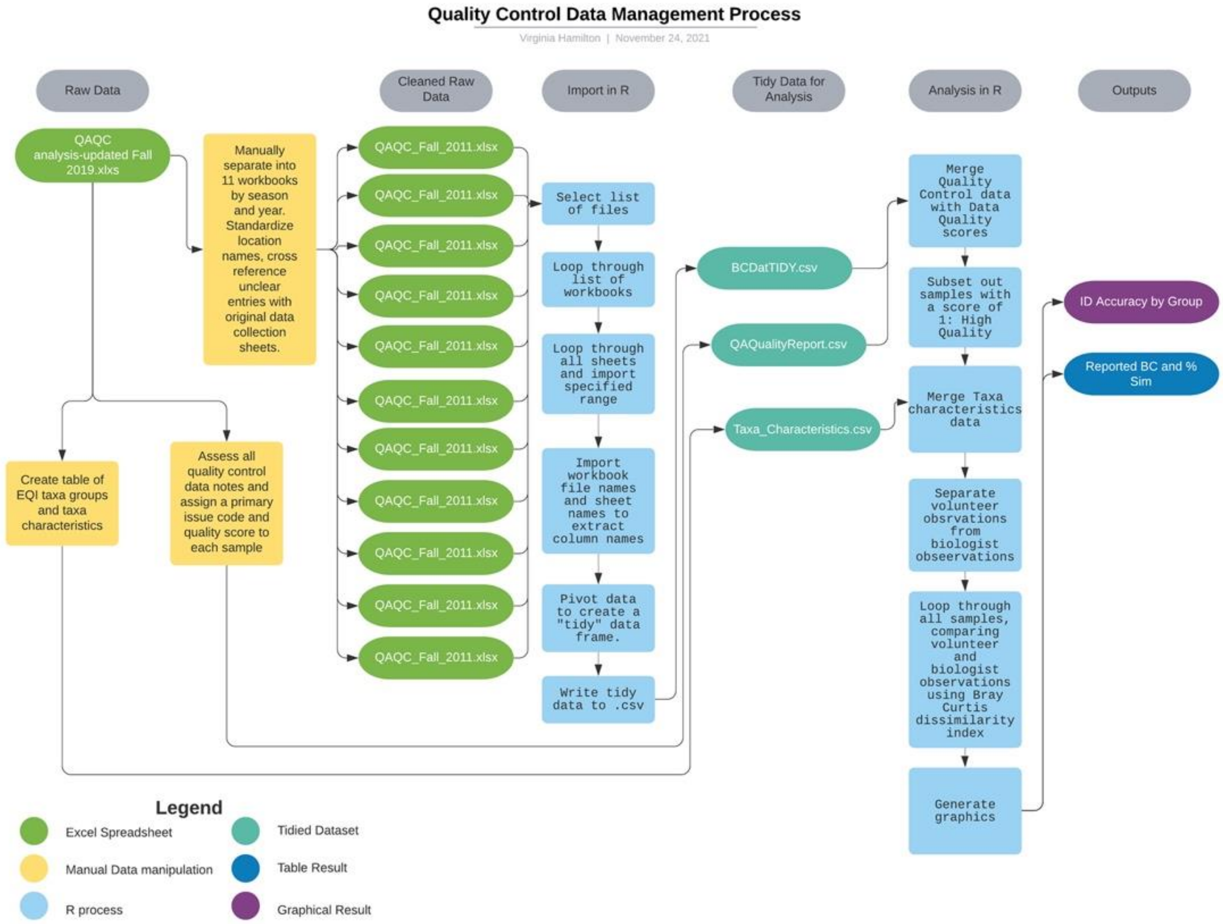


Figure 5: Quality Control Data Management Process.

Tidy Data Principles

The term "tidy data" gained popularity in the data science community after the publication of the paper "Tidy Data" by Hadley Wickham (Wickham 2014). Wickham states that

"80% of data analysis is spent on the process of cleaning and preparing the data" and we found that to be true in this work. Wickham defines tidying as "structuring datasets to facilitate analysis" (Wickham, 2014). The code written to organize the abundance and quality control datasets was written with the goal of making the data "tidy", meaning each variable is its own column, and each observation is its own row. Subsetting, filtering, and selecting data by factors is much easier when all the observations are in this format. In turn, this makes applying necessary functions much more straight forward (Figure 6).

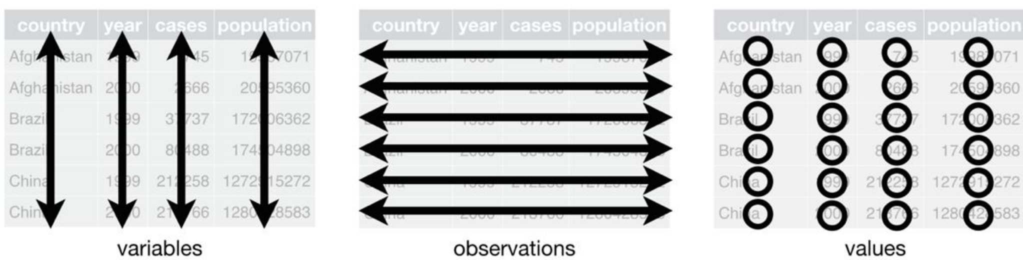


Figure 6: Tidy Data structure places all variables in unique columns. Each observation then occupies its own row. This makes sorting or subsetting by variable simple (Wickham 2014).

<u>Data Analysis</u>

<u>Quality Control Data</u>

My goal was to distill EQI's quality control data into a concise picture of how good volunteers are at identifying benthic macroinvertebrates. First, the raw quality control data was assessed for quality (Table 2). This was to ensure that other variables like ability to follow labeling directions or illegible handwriting were not interfering with an accurate assessment of volunteers' identification skills. In the raw data sheet, EQI employees and the professional biologist left comments describing the volunteer sample. We defined "sample" as each unique combination of season, year, location, and sample method. Using this information, each sample was assigned a number indicating the primary issue with the data. A quality score system was implemented where issues corresponded to a quality score. Seven issues were found in the original quality control dataset. Samples with no issues noted were given a quality score of 1, or “good”. Samples with small samples sizes (less than 3 taxa) were given a quality score of 2, or “use with caution”. Small sample sizes were disregarded because they prone to producing exceedingly high or low similarity values based on limited information that can skew the results of analysis. Samples where the group leader failed to follow collection or recording protocol produced a quality score of 3, or “poor”. In this way, samples of poor quality could be excluded and common errors in protocol could be addressed.

Table 2: Issue codes and quality codes used for Quality Control data.

Issue Code	*Issue Type*	*Quality*	*Descriptor*
0	No Issue	1	Good
1	Failed to preserve all specimens - data incomplete	3	Poor
2	No sample preserved	3	Poor
3	Improper preservation - specimens damaged	3	Poor
4	Vial not labeled properly	3	Poor
5	Taxa count less than or equal to 3	2	Use with Caution
6	Data sheet illegible	3	Poor

The Bray-Curtis dissimilarity index was chosen as a measure of identification agreement (Kosmala et al., 2016). Bray-Curtis is used to describe how similar species abundances are between two different sites (Bray & Curtis, 1957). When assessing observer agreement, the two observers act as the two different “sites”. Dissimilarity between the two observers (*BCvb*) is calculated using Equation 1.

$$BC_{vb} = 1 - \frac{2C_{vb}}{S_v + S_b}$$

Equation 1: Bray-Curtis Dissimilarity index

Where S_v and S_b are the total number of specimens in a sample across all species counted by the volunteer and the biologist respectively. C_{vb} is the sum of the lesser of the two abundance values for each species counted at each site. For example, if the volunteer counted 3 spiny crawlers and 2 roach shredders and the biologist counted 4 spiny crawlers and 1 roach shredder, C_{vb} would be equal to 4 (3 spiny crawlers plus 1 roach shredder). Here, I used an R script to produce a Bray-Curtis dissimilarity index on the volunteer and biologist observations for each sample. Bray-Curtis produces a dissimilarity score expressed as a decimal. The casual reader is more likely to understand the number when it is presented as percent similarity so a simple transformation was applied (Equation 2).

$$Percent\ similarity = (1 - BC_{vb}) \times 100$$

Equation 2: Percent similarity index derived from Bray Curtis Index (Equation 1)

<u>Abundance Data</u>

I calculated three community composition metrics. Richness (Equation 3) is the measure of the number of distinct taxa present in a sample, or

$$Richness = n_{taxa\ groups}$$

Equation 3: Species Richness

Abundance (Equation 4) is the sum of all individuals present in a sample, or

$$Abundance = \ n_{species\ A} + \ n_{species\ b} + n_{species\ c} \dots$$

Equation 4: Abundance

EPT richness is the number of distinct Ephemeroptera, Plecoptera, and Trichoptera taxa present (Lenat & Penrose, 1996).

Results

Assessment of Volunteer Data Accuracy

The data quality assessment process resulted in 71.4% (255/357) of the paired volunteer/professional samples being rated in the highest quality category of “good” (Table 3). Twenty-nine samples (8.1%) had sample sizes less than three and were rated as “use with caution”. Seventy-three samples (20.5%) were ranked as having “poor” quality.

Table 3: Quality control data issue types and counts. Issues were assigned a number. Issue numbers were assigned a quality code number.

Quality	*Issues*	*Descriptor*	*Number of Samples*	*Percentage*
1	0	Good	255	71.4
2	5	Use with caution	29	8.1
3	1, 2, 3, 4, 6	Poor	73	20.4

The most common data quality issue was that no sample was preserved (34 samples, 9.52%) (Table 4). This was usually the result of failing to preserve the specimens from one of the collection methods used. In one case, a group leader forgot that they were supposed to preserve specimens at all that day. The next most common issue was that the sample had a taxa count of less than or equal to 3 (29 samples, 8.12%). In 14 samples (3.92%), group leaders failed to preserve all specimens found as dictated in the quality control procedure. In 9 samples (2.52%) improper preservation of specimens rendered them too poorly preserved to be identified. In 12 samples (3.36%) the sample vials were improperly labeled or unlabeled, meaning that the biologist guessed which method was used or did not identify the specimens. Finally, 4 samples (1.12%) were illegible due to damaged collection sheets, unclear tallying methods, or poor handwriting.

Table 4: Quality control data issues. N=357.

Issue Code	*Issue Type*	*Number of Samples*	*Percentage*
0	No Issue	255	71.4
1	Failed to preserve all specimens - data incomplete	14	3.9
2	No sample preserved	34	9.5
3	Improper preservation - specimens damaged	9	2.5
4	Vial not labeled properly	12	3.4
5	Taxa count less than or equal to 3	29	8.1
6	Data sheet illegible	4	1.1

The average Bray-Curtis dissimilarity score of all volunteer and biologist observations, regardless of sample quality, was 0.223 or 77.8% similar. The average Bray-Curtis dissimilarity score of all samples with a quality score of 1 (no issues reported) was 0.1339 or 86.61% similar. When volunteer accuracy was compared across taxa groups using a regression model and an ANOVA model, there was no evidence that volunteer identification accuracy differed significantly among the taxa groups (p = 0.41) (Figure 7). There were no taxa groups that were correctly identified 100% of the time.

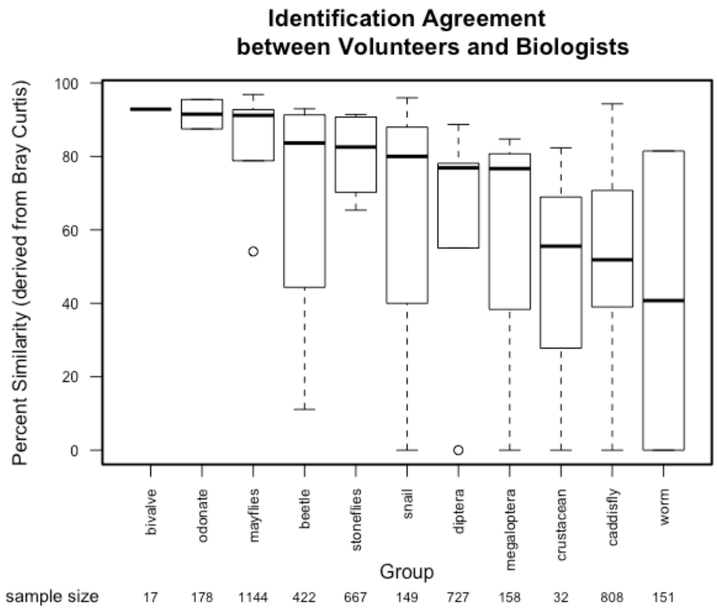


Figure 7: Bray Curtis based percent similarity of identification by volunteers versus biologists using samples rated “good”. Taxonomic groups are ordered with decreasing median similarity from left to right. Sample sizes are indicated below the X axis labels.

Spatial and Temporal Trends

The abundance dataset represented 64 sites sampled over 24 sampling seasons (unique combinations of year and season). Of those 64 sites, 49 sites were sampled on average once per year or more. Sites may not be sampled every season for a variety of reasons including dangerous stream conditions or changes in access permission. Excluding sites that were sampled at low frequency reduces the likelihood that unusual samples might change our assessments of trends over time. High frequency sites were used to plot and assess community composition trends over time (Figure 8). There are no significant trends across all sites for richness, abundance, and EPT taxa richness between 2005 and 2016.

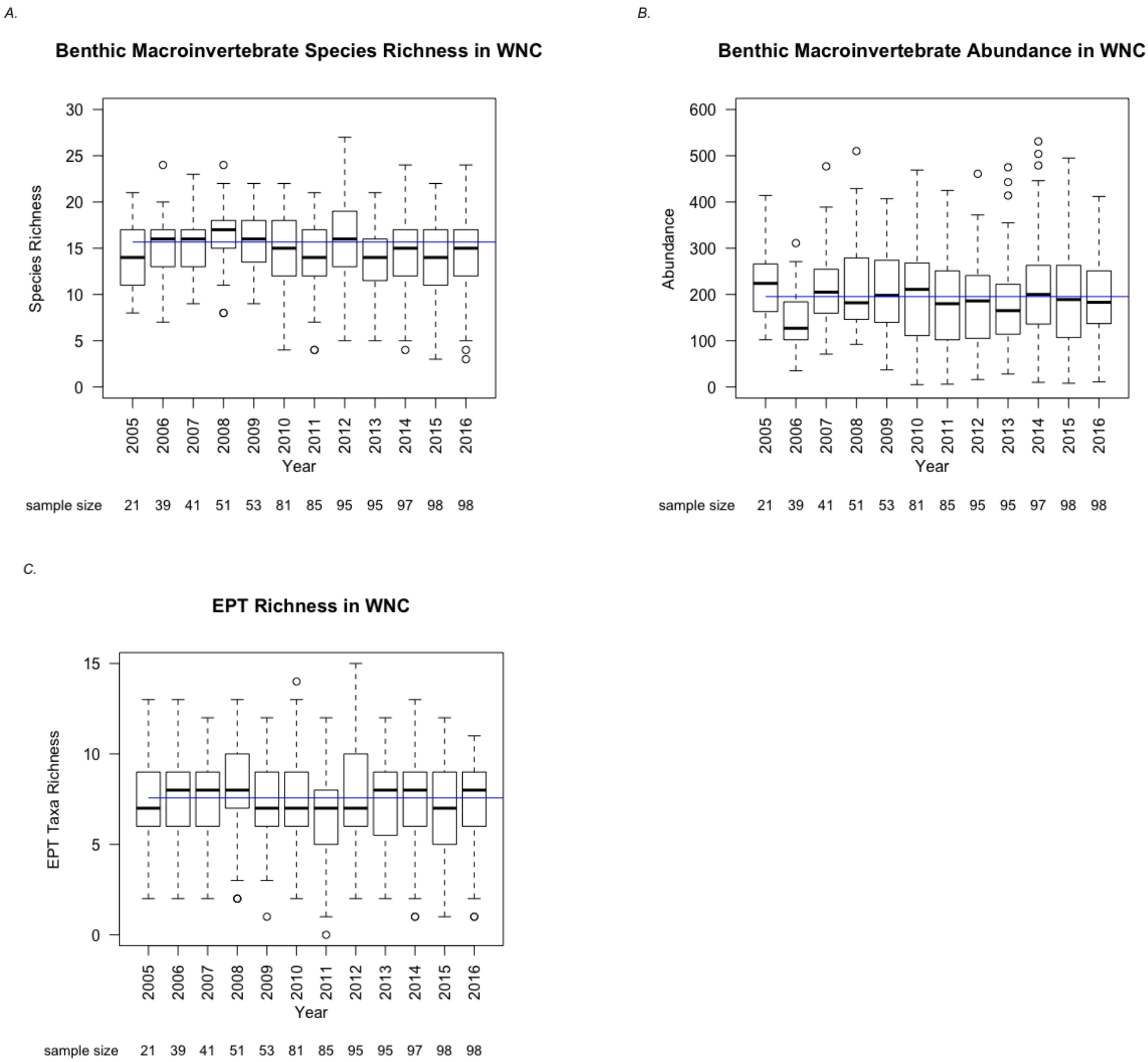


Figure 8: Boxplots of abundance (panel A), richness (panel B), and EPT richness (panel C) in western North Carolina between 2005 and 2016. Sample size (number of sampling occurrences that year) is presented underneath the x-axis. Trends of the mean values are represented in all plots by a blue line. Trends were calculated using a linear regression. Slopes of the trend lines are as follows; Plot A: -.004165 , Plot B: -0.1428, Plot C: -0.0327. No relations were found to be significant.

Average richness at high elevation sites (15) was significantly greater ($p < 0.0001$) than the average richness at low elevation sites (14) (Figure 9). High elevation sites showed a slightly positive trend in richness over the eleven-year period, while low elevation sites showed a slightly negative trend in richness over the same period. Average abundance at high elevation sites (219) was significantly greater ($p < 0.0001$) than average abundance at low elevation sites (187). High elevation sites showed a positive trend in abundance, while low elevation sites showed a slightly negative trend in abundance. Average EPT taxa richness at high elevation sites (8) was significantly greater ($p < 0.0001$) than average EPT richness at low elevation sites (7). High elevation sites showed a slight positive trend in EPT taxa richness, while low elevation sites showed a slightly negative trend in EPT taxa richness.

Average richness of agricultural sites (14) and non-agricultural sites (15) was not significantly different ($p = 0.14$). Both agricultural and non-agricultural sites showed a slightly negative trend in richness over time (Figure 10). Average abundance of agricultural sites (207) and non-agricultural sites (191) was not significantly different ($p = 0.06$). Agricultural sites showed a positive trend in abundance, while non-agricultural sites showed a slightly positive trend in abundance. Average EPT taxa richness of agricultural sites (7) was significantly lower ($p < 0.0001$) than average EPT taxa richness of non-agricultural sites (8). Agricultural sites showed a slightly positive trend in EPT taxa richness, while non-agricultural sites showed almost no trend in abundance over time.

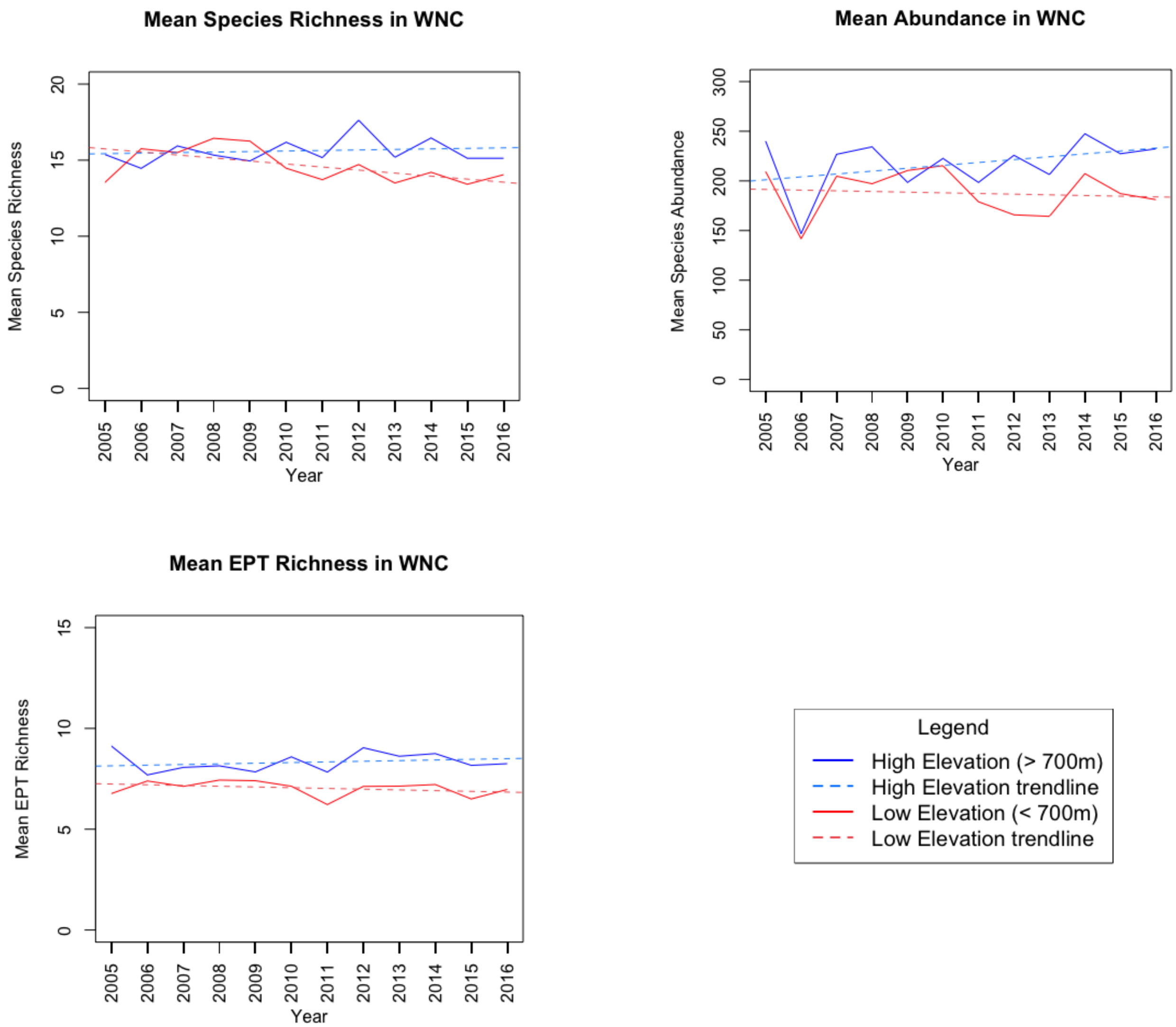


Figure 9: Charts comparing mean abundance, richness, and EPT richness at high and low elevation sites in WNC from 2005 to 2016. Blue lines represent sites higher than 700 m, while red lines represent sites lower than 700m. Solid lines represent mean values. Dashed lines represent linear models created using mean values.

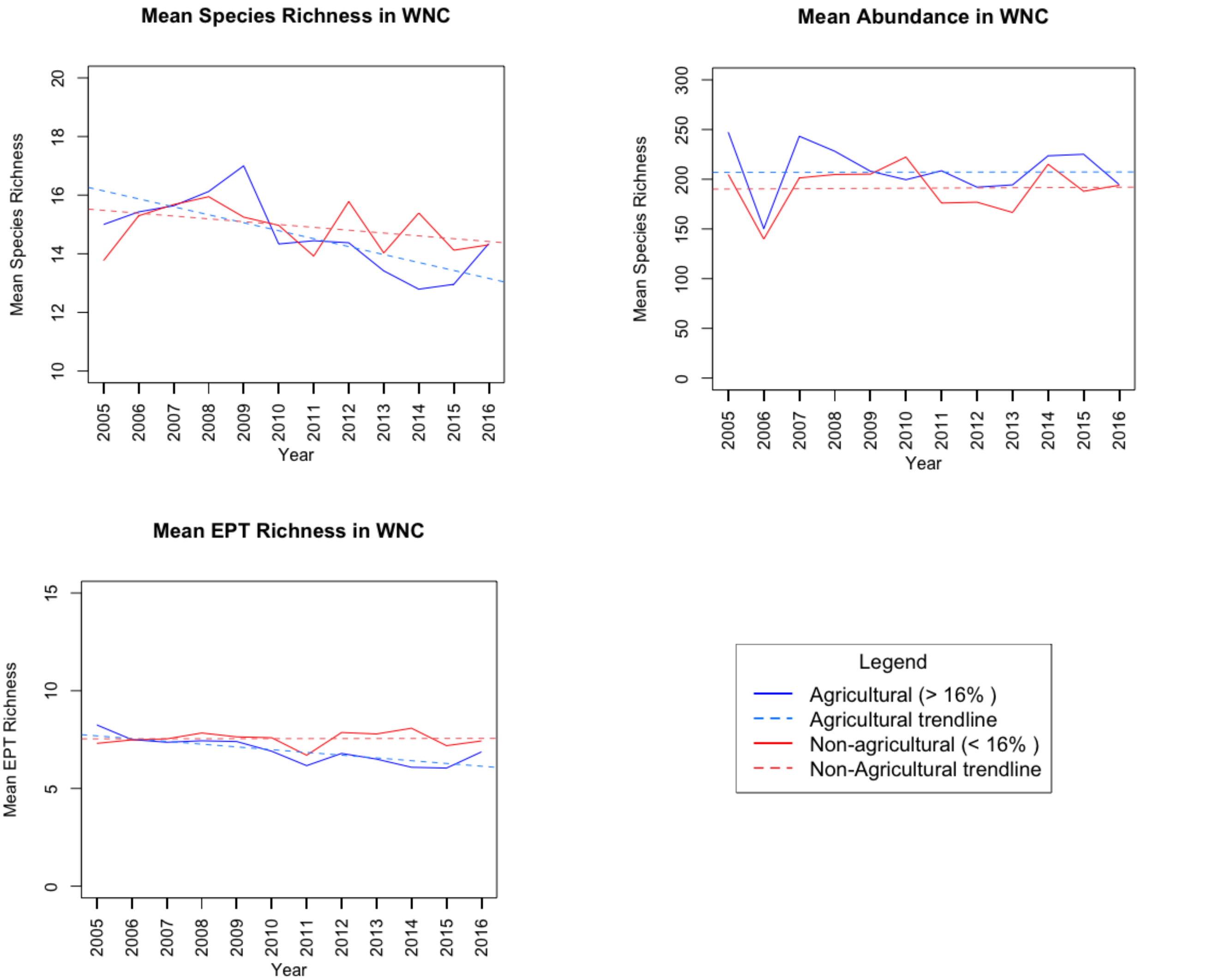


Figure 10: Charts comparing mean abundance, richness, and EPT richness at high and low agricultural use sites in WNC from 2005 to 2016. Blue lines represent agricultural sites where >16% of the watershed is in agriculture while red lines represent non-agricultural sites where <16% of the watershed is in agriculture. Solid lines represent mean values. Dashed lines represent linear models created using mean values.

Discussion

Volunteer Accuracy

EQI Volunteer observations made between 2005 and 2016 provide adequate accuracy to perform their function of informing local agencies and regulatory bodies alongside professional data. Of the "good" quality samples (71%), volunteers had 86% agreement with the professional biologist. The combination of lecture style and hands-on training seems to prepare volunteers well for identifying these animals and performing the collection methods effectively. When compared to other analyses of citizen science data of similar scope, EQI volunteer accuracy is comparable to the accuracy of other groups. Kosmala et al. (2016) found that "rates of 70–95% accuracy are typical for volunteer species identification across a diverse array of systems and taxa." Fore et. al. (2001) found that there was no statistically significant difference between volunteer and professional assessments of stream health using benthic macroinvertebrate data. The NCDEQ sets its threshold for acceptable data at 90% agreement between professionals during their quality control procedure. EQI's quality control data set exhibits a similar percent accuracy rating to that of a professional organization.

Although we did not see a significant difference in volunteers' ability to identify the different groups of taxa, some general patterns suggest possible explanations for differences in agreement across taxa groups. First, there were no large taxa groups (beetles, dipterans, caddisflies, ect) that were correctly identified 100% of the time. This suggests that there may be a limit on the level of identification accuracy that volunteers can gain in a single day of training. Volunteers more accurately identified animals that tend to be more charismatic or those with recognizable patterning. This kind of bias has been observed in citizen science volunteers that more accurately identify well known or highly distinct species (Barbato et al., 2021). Volunteers were more likely to correctly identify "iconic" species (of many diverse taxa) than lesser known species or species that appear very similar to other species (Swanson et al., 2016). For example, stoneflies were accurately identified much of the time. These animals move rapidly, have striking patterns, and are a little more visually stunning than a group like the megalopterans (Figure 11). While the animals in the megaloptera group have distinguishing features, they might be less compelling for volunteers to observe and identify. This is a form of sampling bias that may need to be addressed in EQI's sampling training or protocol.

Figure 11: Photos of a stonefly larva (left) and a meglopteran or Dobsonfly larva (right) (U.S. National Park Service, n.d.) (Hellgrammite- dobsonfly larva, *Corydalus cornutus*, n.d.)

Spatial and Temporal Trends

Our broad scope analysis of community composition metrics supports our assumptions that high elevation sites would have more "pristine" characteristics, such as higher EPT taxa richness. As mentioned, the headwaters of many of these sites fall within protected land. Protection from development, and in some cases recreation, protects high elevation waters from excessive pollution and sedimentation. Sites within agricultural watersheds had higher richness and abundance, but lower EPT taxa richness. This finding is consistent with phenomena observed in other aquatic systems where increased nutrients from runoff increases stream biota abundance, but decreases occurrence of sensitive taxa and overall richness (Lenat, 1984). Western North Carolina has seen large scale anthropogenic change in the last decade. This once less populous mountain region has become one of the United States' top tourism hubs, has experienced increased property values and development, and has an emerging role as a destination for wealthy climate refugees from Florida and California (Johnson, 2021). Further development of the region is unavoidable and understanding freshwater ecology trends as they relate to water quality will help the region plan for the future.

Sedimentation may be one of the major factors behind the trends we observed in this data set. Sediment alters the streambed in many ways that affect small organisms. Increased transport of small particle sediments fills in pore spaces in stream substrate, decreasing living areas for small animals and decreasing surface area for periphyton to grow. Periphyton is the name for the assemblage of algae, fungi, and bacteria that provide food and shelter for many scraping and filtering organisms, such as caddisflies (Trichoptera) and some mayflies (Ephemeroptera) (Harrison et al., 2007). Sedimentation also leaves periphyton communities more vulnerable to die-off during high flow events. Sediment becomes suspended in turbid water and stresses periphyton communities by obscuring sunlight and decreasing photosynthesis (Smith & Duncan, 1999). Scraping and filtering organisms that live in the periphyton are often prey for predatory swimming invertebrates, meaning that their loss can cause a ripple effect on other trophic levels. Waterways at lower elevations (less than 700 m) had significantly lower richness, abundance, and EPT richness than high elevation sites (Figure 9). Exposed and disturbed soil is an unavoidable aspect of development, but without proper erosion control measures sediment can easily enter waterbodies via runoff (Figure 12b). Increased development pressure is a result of increased interest in western North Carolina, largely due to its recreation opportunities. Increased heavy use of popular trails (such as the Black Balsam Trail along the Blue Ridge Parkway, Figure 12a) results in sediment loss from hiking and mountain biking. These trails are often at high altitude, but the effects of their degradation are felt at lower altitudes where sediment accumulates. Although there are strict regulations in place regarding construction site sediment management, enforcement is underfunded.

In addition to sedimentation, changes in stormwater runoff amount and intensity may also be influencing trends in benthic macroinvertebrate communities. In Asheville city limits (central Buncombe County, see Figure 1), the city lost 6.4% of its tree canopy cover in a ten year period (2008-2018) (McDaniel, 2019). Trees have greater protection at higher altitudes under the county's steep slope ordinances which protect trees on steep slopes at altitudes greater than 716 m, but below that there are few restrictions on tree removal. The cutoff for "high elevation" sites in our study was 700 m. The Urban Tree Canopy Study for Asheville, North Carolina conducted by the Davey Resource Group estimated that "The loss of tree canopy resulted in an average increase of [82,000 m3] of stormwater control, or equivalent to 27 Olympic swimming pools.

The 10-year increase in runoff was over 1,300 gallons per acre or approximately $1,600,000 [in increased stormwater management costs] overall" (Davey Resource Group, 2019). This effect goes hand in hand with the pressures of development. Increases in impermeable surfaces and decreases in canopy cover increases the severity of heavy rain events, contributing to the differences we observed in streambed communities.

A. *B.*

Figure 12: A. Intense erosion at a popular hiking trail in the Balsams. The trail bed sits up to 5 feet below ground level (see hiking pole at left) in some places. An influx of visitors during the pandemic and a decrease in trail maintenance funds has contributed to major soil loss (Weemhoff, 2014). B. While laws exist to prevent sediment loss at new developments, they are not always enforced. The road below this development is less than 100 feet from the bank of the French Broad River (Chávez, 2020).

In agricultural communities, a suite of factors may have contributed to our finding that agricultural areas have lower EPT richness. Our finding of highly variable macroinvertebrate abundance and richness associated with agricultural land use (Figure 10) is consistent with findings from studies that attribute the pattern to nutrient inputs from agricultural operations (Lenat, 1984). The region is home to farms with a variety of crops and management practices, all of which would cause different stressors on their watersheds (Lenat, 1984). It is interesting to note that the United States has seen the average acreage of farms increase while the number of farm businesses has decreased (USDA, 2020). This points to farm consolidation. If we assume that western North Carolina has followed national trends of farm land loss and consolidaiton, we can assume that production has moved from small scale or subsistance agriculture to large scale commercial agriculture (MacDonald, 2018). Consolidated farms may have access to more capital (being corporations or wealthy landowners with more equity to borrow against). This may allow for increased purchase and use of large tillage equipment as well as large scale use of synthetic fertilizers and agrochemicals, the costs of which have risen astronomically in the last 30 years, making them unavailable to small growers (Schnitky et al., 2021).

Increased sedimentation from more powerful tillage and herbicide use could be negatively affecting overall richness and EPT taxa richness in agricultural areas by decreasing habitat for more sensitive organisms. Depending on the crop and application (rate, timing, form) of the pesticide, pesticide residues may make their way into stream systems. The pesticide immidicloprid, for example, has been shown to accumulate in benthic macroinvertebrates and is a common pesticide in commercial tomato operations which becoming more common in western North Carolina (Crayton et al., 2020). It is important to note that “agricultural” encompasses a wide variety of land management practices and crops. Having more data on agricultural management practices would help assess the specific variables driving differences in population trends.

Opportunities for This Dataset

Questions regarding the region’s aquatic life can now be compared by elevation, land use, slope, watershed, size, watershed location, and date sampled by using the data frame created during this project. Having data specific to site and time makes it possible to compare the abundance data set, for example, to other datasets capturing change over time or ecological characteristics in the region. The EQI also administers a program called the Volunteer Water Information Network (VWIN) that collects monthly water samples from the region’s streams and lakes. These samples are tested for pH, alkalinity, turbidity, total suspended solids, conductivity, ammonia-nitrogen, nitrate/nitrite-nitrogen, orthophosphate, total phosphorus, and fecal coliform. Cleaning and organizing this data set is a clear next step that would allow comparison between water quality characteristics and benthic macroinvertebrate population trends. For example, is there a relationship between benthic macroinvertebrate population dynamics and total suspended solids or turbidity? Do higher fecal coliform counts appear to correspond to agricultural areas and how is that effecting EPT taxa richness? Assessing correlation between these factors would allow us to explore the possible trend explanations presented earlier.

One of the biggest opportunities for this data set is its integration with a geographic information systems program (GIS). The current table format for the abundance dataset could be readily included in a shapefile attribute table. This presents numerous possibilities for study of the region and its benthic communities. GIS data sets are readily available from county data clearinghouses and may include information such as building trends, impervious surfaces, point-source pollution sites, and slope data. In the future, it would be beneficial to have this dataset hosted within a publicly accessible data repository for easier access. The R code from this project is not proprietary and could easily be modified and shared with other citizen science projects or community stakeholders via a platform like GitHub.

Opportunities and Suggestions for Citizen Science Projects

This citizen science data set and this project offer many lessons for other citizen science data projects. Starting with the end goal of an analysis-ready dataset, the first step is training volunteers in the organization's preferred data recording process. Instructing volunteers in the importance of consistent data recording should be emphasized. The statistical power of the data is decreased when volunteers fail to follow preservation protocol, record an incomplete site name, or use an unknown tallying system for counting specimens. These errors all occurred within these datasets and caused 20% of data to be unusable. Explicit instruction for things like tallying, date formats, and site could reduce errors that directly contributed to data loss (Table 4).

While Excel and other spreadsheet programs are widely available and user-friendly, it is worth noting that use of data processing programs like R are on the rise. As of 2017, 58% of peer-reviewed articles in 30 ecology journals used R's statistical software, with a linear pattern of growth (Lai et al., 2019). R provides several unique advantages including processing speed, replication of analysis, packages for discipline-specific analysis, easy handling of large data sets, and improved visualization of data. Although it may not make sense to switch all data recording to a format immediately compatible with R analysis (like our tidied data), it is worth structuring data in Excel in a way that makes import into R simpler. R scripts can be written to extract data from uniform workbook and folder structures. The most critical characteristic of an easy-import dataset is consistency. To use a function to extract information from multiple workbooks or multiple sheets the same types of data need to be in the same place (row, column location) in every spreadsheet. Workbooks and folder structures also need to be organized consistently. In the workbooks themselves, column names, variable levels, and numbers need to have consistent formatting. Best practices include naming for files and for column headings that are concise, descriptive, and without spaces. To ensure consistent string variables (like locations or conditions), descriptors should always be typed the same way to avoid creating too many levels in a variable. Better yet, these variables could be selected from a list during data entry, so they are identical every time. For example, inconsistencies in site names within the abundance data set ("Asheworth Creek", "Ashworth creek", and "Asheworth Crk") resulted in extra work and a lot of manual correction. The ideal situation would involve a data input structure that is easy for the organization to use paired with an R script designed to pull data into a tidy format for analysis.

Conclusions

The current staffing levels of both government and non-profit water quality monitoring efforts simply cannot provide the level of frequency needed to capture a complete picture of waterway health in western North Carolina. Volunteers have and will need to supplement monitoring efforts. The sampling and training methodology that the Environmental Quality Institute uses has led to macroinvertebrate identification by volunteers that are 86% similar to that for professionals when data is collected according to protocols (20% of data did not follow protocol). This places their accuracy within the ranges of other citizen science efforts and is appropriate for the end use of EQI's data as an informative tool.

Volunteer-collected macroinvertebrate data reveal distinct patterns in richness, abundance, and EPT richness that are dependent on elevation and agricultural land use. This dataset also contains information about other land uses and watershed characteristic to be explored. The land surrounding these aquatic ecosystems very much affects the organisms living in them. Data collected by EQI volunteers supports known population dynamic phenomenon associated with elevation and proximity to agricultural land and has the potential to help explore many other questions about the drivers that influence aquatic systems.

As we move forward with the reality that volunteers will be critical in our understanding of the health of aquatic systems, citizen science projects need to ensure that end-use of their data is considered from the beginning by structuring their data storage with analysis in mind. Ensuring a clean path between data collection and data utilization in the modern world of statistical analysis will ensure that volunteers' hard work is useable. With effective training volunteer benthic macroinvertebrate monitors can provide accurate and critically important data that can serve as an early warning system for trained professionals.

REFERENCES CITED

Barbato, D., Benocci, A., Guasconi, M., & Manganelli, G. (2021). Light and shade of citizen science for less charismatic invertebrate groups: Quality assessment of iNaturalist nonmarine mollusc observations in central Italy. *Journal of Molluscan Studies*, *87*(4). https://doi.org/10.1093/mollus/eyab033

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. https://doi.org/10.2307/1942268

Carroll, W. D., Kapeluck, P. R., Harper, R. A., & Lear, D. H. V. (2002). Paper: Historical Overview of the Southern Forest Landscape and Associated Resources. *Background Paper*, 24.

Chávez, K. (2020, 03). *Fallen tree on Asheville RAD hillside above Amboy Road leads to "mess" but no imminent danger*. The Asheville Citizen Times. https://www.citizen-times.com/story/news/local/2020/08/03/asheville-developer-and-city-clash-over-steep-slope/5519879002/

Clark, S. H. B. (2001). *Birth of the Mountains; The Geologic Story of the Southern Appalachian Mountains*. U.S. Department of the Interior.

Crayton, S. M., Wood, P. B., Brown, D. J., Millikin, A. R., McManus, T. J., Simpson, T. J., Ku, K.-M., & Park, Y.-L. (2020). Bioaccumulation of the pesticide imidacloprid in stream organisms and sublethal effects on salamanders. *Global Ecology and Conservation, 24,* e01292. https://doi.org/10.1016/j.gecco.2020.e01292

Penrose Environmental Consulting Website: *Dave's Background | Penrose Environmental Consulting*. (n.d.). Retrieved November 2, 2021, from https://penrose.consulting/daves-background/

Davey Resource Group. (2019). *Urban Tree Canopy Study for Asheville NC*. https://drive.google.com/file/d/0B5y5xB5sGeofNlIzQ1ZiT09PVGZnZWJLMFI5dUZwcTY5am1v/view?resourcekey=0-rpAjXutJC6Cd1Lbv_qMSWw&usp=embed_facebook

Delcourt, P. A., & Delcourt, H. R. (1998). Paleoecological Insights on Conservation of Biodiversity:a Focus on Species, Ecosystems, and Landscapes. *Ecological Applications*, *8*(4), 921–934. https://doi.org/10.1890/1051-0761(1998)008[0921:PIOCOB]2.0.CO;2

Faill, R. T. (1997). A geologic history of the north-central Appalachians; Part 1, Orogenesis from the Mesoproterozoic through the Taconic Orogeny. *American Journal of Science*, *297*(6), 551–619. https://doi.org/10.2475/ajs.297.6.551

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

Harrison, E. T., Norris, R. H., & Wilkinson, S. N. (2007). *The impact of fine sediment accumulation on benthic macroinvertebrates: Implications for river management*.

Hellgrammite- Dobsonfly Larva—Corydalus cornutus. (n.d.). Retrieved November 12, 2021, from https://bugguide.net/node/view/185765

Johnson, E. (2021, February 3). *Western North Carolina, Climate Refuge?* Pulitzer Center. https://pulitzercenter.org/stories/western-north-carolina-climate-refuge

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*(10), 551–560. https://doi.org/10.1002/fee.1436

Kozar, M. D., Sheets, C. J., & Hughes, C. A. (2001). *Ground-water quality and geohydrology of the Blue Ridge physiographic province, New River basin, Virginia and North Carolina*. National Water Quality Assessment Program. https://doi.org/10.3133/wri004270

Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., & Ma, K. (2019). Evaluating the popularity of R in ecology. *Ecosphere*, *10*(1), e02567. https://doi.org/10.1002/ecs2.2567

Lenat, D. & Penrose. (1996). History of the EPT Taxa Richness Metric. *Bulletin of the North American Benthological Society*, *13*, 305–306.

Lenat, D. R. (1984). Agriculture and stream water quality: A biological evaluation of erosion control practices. *Environmental Management*, *8*(4), 333–343. https://doi.org/10.1007/BF01868032

Li, L., Zheng, B., & Liu, L. (2010). Biomonitoring and Bioindicators Used for River Ecosystems: Definitions, Approaches and Trends. *Procedia Environmental Sciences*, *2*, 1510–1524. https://doi.org/10.1016/j.proenv.2010.10.164

MacDonald, J. M. (2018). *Three Decades of Consolidation in U.S. Agriculture*. 61.

McDaniel, P. (2019, October 29). *Asheville Urban Tree Canopy Study looks at canopy loss and opportunities to preserve it*. The City of Asheville. https://www.ashevillenc.gov/news/asheville-urban-tree-canopy-study-looks-at-canopy-loss-and-opportunities-to-preserve-it/

National Forests in North Carolina—About the Forest. (2013). https://www.fs.usda.gov/main/nfsnc/about-forest

NC DENR. (2021). *Benthic Monitoring Data*. https://ncdenr.maps.arcgis.com/apps/webappviewer/index.html?id=7d4f2137e9154742bae16916e35a87a0

NC DEQ: Laboratory Certification Branch. (2021). Laboratory Certification Branch. https://deq.nc.gov/about/divisions/water-resources/water-resources-data/water-sciences-home-page/laboratory-certification-branch

NC DEQ: Water Sciences Section. (n.d.). Retrieved April 25, 2021, from https://deq.nc.gov/about/divisions/water-resources/water-resources-data/water-sciences-home-page

North Carolina Division of Water Resources Surface Water Quality Standards, Criteria & In-Stream Target Values. (2019). NC DEQ. https://deq.nc.gov/documents/nc-stdstable-06102019

Oertel, N., & Salánki, J. (2003). *Biomonitoring and Bioindicators in Aquatic Ecosystems* (pp. 219–246). https://doi.org/10.1007/978-1-4615-0221-0_10

Penrose, D., & Call, S. M. (1995). Volunteer Monitoring of Benthic Macroinvertebrates: Regulatory Biologists' Perspectives. *Journal of the North American Benthological Society*, *14*(1), 203–209. https://doi.org/10.2307/1467735

Poff, N. L., Olden, J. D., Vieira, N. K. M., Finn, D. S., Simmons, M. P., & Kondratieff, B. C. (2006). Functional trait niches of North American lotic insects: Traits-based ecological applications in light of phylogenetic relationships. *Journal of the North American*

Benthological Society, *25*(4), 730–755. https://doi.org/10.1899/0887-3593(2006)025[0730:FTNONA]2.0.CO;2

Postelle, B. (2009, August 26). Director seeks new home for water-testing center cut by UNCA. *Mountain Xpress*. http://mountainx.com/news/082609director_seeks_new_home/

Public Lands. (2021). *MountainTrue*. https://mountaintrue.org/forests/public-lands/

Ross, K. (2020, August 3). *Bracing for budget cuts: Slashing expected at NC Department of Environmental Quality*. Carolina Public Press. https://carolinapublicpress.org/36865/nc-budget-stalemate-cuts-department-environmental-quality-deq/

Schnitky, G., Paulson, N., Zulauf, C., & Swanson, K. (2021, August 3). 2021 Fertilizer Price Increases in Perspective, with Implications for 2022 Costs. *Farmdoc Daily*. https://farmdocdaily.illinois.edu/2021/08/2021-fertilizer-price-increases-in-perspective-with-implications-for-2022-costs.html

Smith, J. G., Brandt, C. C., & Christensen, S. W. (2011). Long-Term Benthic Macroinvertebrate Community Monitoring to Assess Pollution Abatement Effectiveness. *Environmental Management*, *47*(6), 1077–1095. https://doi.org/10.1007/s00267-010-9610-3

Smith, R. A., & Duncan, M. J. (1999). Velocity and Sediment Disturbance of Periphyton in Headwater Streams: Biomass and Metabolism. *Journal of the North American Benthological Society*, *18*(2), 222–241. https://doi.org/10.2307/1468462

Standard Operating Procedures for the Collection and Analysis of Benthic Macroinvertebrates. (2016). North Carolina Department of Environmental Quality. https://files.nc.gov/ncdeq/Water%20Quality/Environmental%20Sciences/BAU/NCDWRMacroinvertebrate-SOP-February%202016_final.pdf

State of the Climate: National Climate Report for Annual 2020. (2021). NOAA National Centers for Environmental Information. https://www.ncdc.noaa.gov/sotc/national/202013

Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, *30*(3), 520–531. https://doi.org/10.1111/cobi.12695

US EPA, O. (2013, November 21). *Indicators: Benthic Macroinvertebrates* [Overviews and Factsheets]. US EPA. https://www.epa.gov/national-aquatic-resource-surveys/indicators-benthic-macroinvertebrates

U.S. National Park Service. (n.d.). *Natural Resource Monitoring at Great Smoky Mountains National Park—Great Smoky Mountains National Park*. Retrieved November 12, 2021, from https://www.nps.gov/grsm/learn/nature/monitoring.htm

USDA. (2020). *Farms and Land in Farms 2019 Summary 02/20/2020* (p. 17).

Weemhoff, D. (2014, May 27). *Black Balsam-Graveyard Fields Loop (10.0 miles; d=13.70)—Dan Weemhoff (dwhike)*. https://www.dwhike.com/Hikes-in-the-South/North-Carolina-Hikes/Pisgah-NF-Pisgah-District/Black-Balsam-Graveyard-Fields