

ISSUES : DATA SET

Investigating Human Impacts on Stream Ecology: Locally and Nationally

Amelia Nuding¹ and Stephanie Hampton^{1,2}

1 - National Center for Ecological Analysis & Synthesis, Barbara, Santa Barbara, CA 93101-3351

2 - Corresponding Author: Stephanie Hampton (hampton@nceas.ucsb.edu)

THE ECOLOGICAL QUESTION:

How does nutrient pollution impact stream ecosystems locally and nationally?

ECOLOGICAL CONTENT:

Organic nutrients, land cover, biological indicators, macroinvertebrate index, stream ecosystem structure and function

WHAT STUDENTS DO:

- Explore maps of streams and land use in the US
- Analyze national water quality data sets from the US Environmental Protection Agency at regional and national scales
- Perform basic calculations from data set and generate graphs
- Relate water quality with land use
- Relate water quality with stream ecology

STUDENT-ACTIVE APPROACHES:

The questions posed throughout the activity are based on the six levels of [Bloom's taxonomy](#) and critical thinking. The students are asked to generate hypotheses, and this can be done in pairs, for example as a "[turn to your neighbor](#)" exercise.

Students' connection to this exercise will be enhanced if combined with hands-on activities such as fieldwork to collect macroinvertebrates and water samples in a stream, and/or laboratory work to measure water quality parameters.

SKILLS:

Navigating and manipulating large data sets, reading metadata (descriptions of the data), creating graphs, connecting the analysis of data to scientific concepts

- **What students should learn from this lesson:** Students should understand the connections between land use and water quality, the effects of nutrient pollution, how bioindicators are useful to ecologists, and how to plot and interpret stream water quality and ecosystem integrity indicators.



Photo 1: Waterfall in Quinalt near Olympic National Park. Photo by Robert Kraft.

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

- **What skills students should acquire from this lesson:** map interpretation, manipulation of datasets, construction and interpretation of box plots, simple linear regressions, log plots

ASSESSABLE OUTCOMES:

The activity is designed as a series of investigations and questions which require students to summarize raw data, create charts and plots, and interpret maps and figures. The deliverables are: graphs generated by students, short answers to provided questions, and a short essay which synthesizes the concepts presented in this exercise. Prediction, analysis and synthesis are the core elements of the deliverables. In addition, pre- and post-test questions can be administered to uncover what lessons were really learned by students in this activity.

TRANSFERABILITY:

This exercise is designed for college biology majors and non-majors, as well as students in a basic statistics class. Navigating, interpreting, and analyzing large data sets require basic skills that are applicable across all professional fields, from marketing to the hard sciences, and working with spreadsheets is necessary in many jobs and in studies that may be encountered from high school through post-graduate positions. This exercise may also be appropriate for pre-college students, though perhaps with some alterations to the lesson. It is designed to be appropriate for students with basic math and computer skills, and the analysis of the data can be made more or less complex at the instructor's discretion. Because this data set covers virtually all regions across the contiguous US, the lesson is also designed to examine trends in the state/region of the academic institution, allowing students and instructors to create a stronger connection between the national scope of this work and the local environment experienced by students.

The URLs included were live at the time of publication; in most cases, URLs will change over time and so we have uploaded static data and maps that may be readily used in this exercise and supplemented with new internet resources at the instructor's discretion.

SOURCE:

U.S. Environmental Protection Agency. 2010. Wadeable streams assessment. (http://www.epa.gov/owow/streamsurvey/pdf/WSA_Assessment_May2007.pdf) Accessed March 10, 2012. (Note this is a larger document that takes extra time to load.)

U.S. Environmental Protection Agency. 2010. Probabilistic sampling databrowser. (<http://www.epa.gov/bioiweb1/statprimer/ProbabilisticSampling/index.html>) Accessed March 10, 2012.

U.S. Environmental Protection Agency. 2010. Teacher 'use case' for the national WSA browser. (<http://www.epa.gov/bioindicators/statprimer/WSAUseCaseTeacher.html>) Accessed March 10, 2012.

U.S. Geological Survey. 2010. MRLC consortium viewer. (<http://gisdata.usgs.net/website/MRLC/viewer.htm>) Accessed March 10, 2012.
Note that you will need to find the key for this map separately – e.g., [Supplement 1 for Land Cover](#).

U.S. Geological Survey. 2012. The National Land Cover Database. (<http://pubs.usgs.gov/fs/2012/3020>) Accessed March 10, 2012.

National Atlas. 2010. National atlas map maker. (<http://nationalatlas.gov/mapmaker>) Accessed August 17, 2010.

National Atlas. 2010. Map layer info: streams and water bodies. (<http://www.nationalatlas.gov/mld/hydrogm.html>). Accessed March 10, 2012.

U.S. Geological Survey 2010. The Gulf of Mexico hypoxic zone. (http://toxics.usgs.gov/hypoxia/hypoxic_zone.html). Accessed March 10, 2012.

OVERVIEW OF THE ECOLOGICAL BACKGROUND

This exercise was developed to give students the opportunity to work with large, spatially diverse data sets on water quality and other ecological parameters of streams in the United States. The data set used in this lesson comes from stream samples collected by the U.S. Environmental Protection Agency (EPA) across the 48 contiguous states, in 2004 and 2005. The EPA is required by the Clean Water Act to report to Congress on the health of the nation's waters, and in 2002 the EPA began to develop a stream sampling regime to characterize the condition of all the waterways throughout the continental US. This survey of "wadeable" streams is the EPA's largest effort to make a scientifically and statistically defensible claim about the state of the nation's waters (EPA 2010). Wadeable streams provide a strong link between land use and water quality, and they contribute to larger rivers systems, so they are a good indicator of the status of waters throughout the entire US (EPA 2010). Even though wadeable streams are relatively small and shallow, they comprise about 90% of the length of all perennial waterways.

The sampling regime was designed to be representative and random, delivering statistically robust results that characterize the water quality and ecosystem integrity of the entire nation's wadeable streams. Researchers determined the length and location of all wadeable streams in the US using the National Hydrography dataset and the River Reach File. A total of 1,392 sites across the 48 states were selected using a probability-based sampling design, while ensuring that the selections were controlled for even spatial distribution and varied stream-order size. Each site was sampled once during summer sometime between 2000 and 2004, using a standardized survey method, with multiple quality control measures in place. More detailed information about the sampling design and data collection can be found in Chapter 1 (http://www.epa.gov/owow/streamsurvey/pdf/WSA_Assessment_May2007.pdf). The design of the sampling regime could be a lesson in itself, though it is not the focus of this activity. The entire report contains very useful, clear information, and therefore it is highly recommended as supplemental reading to the instructor.

The stream parameters investigated in this exercise are total nitrogen, total phosphorus and a macroinvertebrate index for the 10 EPA regions in the contiguous US. These nutrient parameters were selected because they are widespread pollutants and can have cascading effects on stream ecosystems. The macroinvertebrate index was selected because it is among the most widely used indicators of ecological integrity, and can be correlated with total nitrogen and total phosphorus.

REFERENCES

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

U.S. Environmental Protection Agency. 2010. Wadeable streams assessment. (http://www.epa.gov/owow/streamsurvey/pdf/WSA_Assessment_May2007.pdf) Accessed March 10, 2012.

DATA SETS

The following are links to the publicly available data sets used in this lesson.

- For this exercise we have merged three files from the Wadeable Stream Assessment data - “water chemistry,” “site info” and “benmet300” – the resulting single data set is used in this exercise and is available to the public via the Knowledge Network for Biocomplexity:

<http://knb.ecoinformatics.org/knb/metacat/nuding.9.4/knb>. The metadata for this dataset is available from this link as well. A full description of the multi metric index for benthic macroinvertebrates (MMI WSABEST) can be found in the student activity, [Section III](#).

- If you wish to look at the raw data produced by the EPA, please visit Wadeable Stream Assessment data and metadata: http://www.epa.gov/owow/streamsurvey/web_data.html. These datasets are downloadable separately, and were not easily combined at the time of this publication.

STUDENT INSTRUCTIONS

Introduction

Small streams are vital to the United States' major rivers because they deliver water and provide pathways for the movement of fish and other aquatic organisms. Humans rely on both large and small waterways for drinking, irrigation, industrial uses and transportation. Thus, if the water quality of any particular stream is impacted, it affects not only *local* fish, aquatic organisms and plants, but also non-local organisms, in other parts of the state or country, and ultimately the livelihood of humans. While the term “environmental health” is a scientifically controversial term (Simberloff 1998), we may find it useful to think about stream “health” as describing the conditions that humans find desirable such as water with low levels of pollutants and pathogens, high levels of oxygen, and providing otherwise good habitat for fish and wildlife (Meyer 1997; Gordon et al. 2004). A variety of factors affect which aquatic organisms are present, such as the stream's size and morphology, geographic location, stream flow (volume and speed of water), available light, temperature, and water quality (Vannote et al. 1980; Poff 1997).

In this exercise we will examine two water quality parameters of great importance in streams: nitrogen and phosphorus. These two nutrients are among the most common

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

pollutants in streams, lakes and coastal waters, resulting in degraded water quality (EPA 2007). Nitrogen and phosphorus generally come from fertilizer applied to farm lands (and from lawns, to a lesser extent), as well as sewage, and they are carried by ground water, rain water, or irrigation water from the land into streams (Vitousek et al. 1997; Smith 2003; Dodds 2006; EPA 2007). In addition, these nutrients may be deposited on land or water through the air after the combustion of fossil fuels or other industrial operations (Driscoll et al. 2006; Pepper et al. 2006). Specifically in this activity you will be looking at values of “total nitrogen” and “total phosphorus,” two very commonly measured water quality parameters. Total nitrogen includes all organic and inorganic nitrogen-containing compounds in the water. Inorganic forms are nitrate (NO_3^-), nitrite (NO_2^-), ammonia (NH_3), and ammonium (NH_4^+). Organic forms include proteins, peptides, nucleic acids, urea and synthetic organic materials (Pepper et al. 2006). Total phosphorus, similarly, includes all phosphorus-containing compounds in the water, which includes orthophosphate (PO_4) and organically bound phosphate.

When these nutrients become very high, *algae* (photosynthetic organisms) can grow extremely quickly and the waters can become cloudy, reducing the light availability in the stream. Importantly, when these algal blooms die, this large amount of dead algae fuels bacterial growth in the water. The bacteria decompose the algae and in doing so they consume much of the oxygen available in the water (Mallin et al. 2006). In many cases the waters become uninhabitable by aquatic organisms because of the lack of dissolved oxygen in the water. An extreme example of this is the “dead zone” in the Gulf of Mexico, which is an area of the ocean about as large as Connecticut, in which few aquatic organisms can survive, primarily due to the input of nutrients from the Mississippi River (USGS 2010).

Aquatic *benthic macroinvertebrates* are insects and other small invertebrates (like crustaceans, mollusks and aquatic worms) that live in streams and other aquatic habitats. “Benthic” refers to the lowest level of a water body, which in this case is the stream bed, and this is where these animals reside. “Macro” means that these animals are large enough to be seen with the naked eye, and “invertebrates” means that they have no spine. Many common flying insects have larval stages in streams and lakes, such as mayflies, stoneflies and dragonflies. These stream macroinvertebrates play very important roles in the stream ecosystem. For example, by shredding leaves and other detritus that falls into streams, they convert terrestrial carbon and other nutrients into forms available to other stream organisms (Vannote et al. 1980; Wallace and Webster 1996). Some macroinvertebrates eat algae, and others are predators on small invertebrates. Most macroinvertebrates eventually become an important food resource to fish and birds (Vannote et al. 1980; Wallace and Webster 1996).

Because benthic macroinvertebrates are such important members of the food web and they respond relatively quickly to ecological changes, they can be very useful to

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

humans as indicators of the health of a stream. Accordingly, they are called “bioindicators.” Some macroinvertebrates require high levels of oxygen and low pollutant levels, and so they are useful as indicators of good water quality, while other groups of macroinvertebrates can tolerate low oxygen and high pollutant levels, thereby indicating lower water quality (EPA 2007). They are not highly mobile (especially compared with fish) and so they are susceptible to the water quality conditions around them and whatever pollutants may have accumulated in the sediment of the stream bed.

Data Set Description

In 2002, the US Environmental Protection Agency (EPA) set out to characterize the health of all the waterways throughout the continental US. The EPA is required by law as set forth in the Clean Water Act to report to Congress on the health of the nation’s waters. This survey of “wadeable” streams – streams shallow enough to sample without a boat – is the EPA’s largest effort to make a scientifically and statistically defensible claim about how healthy, or unhealthy, the nation’s waters are (EPA 2007).

A statistically sound sampling design was necessary for the EPA to be able to detect major trends in stream quality across the nation. Ideally the EPA would take samples from every waterway in the US, but that is totally infeasible; it would be a huge, expensive, and very time-consuming endeavor. Therefore, the EPA devised a sampling regime of wadeable streams. Wadeable streams provide a strong link between land use and water quality, and they contribute to larger rivers systems, so they are a good indicator of the health of waters throughout the entire US (EPA 2007). Even though wadeable streams are relatively small and shallow, they comprise about 90% of the length of all perennial waterways.

A total of 1,392 sites were sampled in the 48 states. The type of sampling design selected for any ecological study or experiment is key to making more general assertions about the status of waterways throughout the nation. The sampling was designed to ensure that the site selection was representative and random. More information about the sampling design can be found in Chapter 1 in the section of the WSA Report entitled “How Were Sampling Sites Chosen?” (pgs. 15 – 17, http://www.epa.gov/owow/streamsurvey/pdf/WSA_Assessment_May2007.pdf).

Literature Cited

Dodds, W.K. 2006. Eutrophication and trophic state in rivers and streams. *Limnology and Oceanography* 51:671-680.

Driscoll, C., D. Whitall, J. Aber, E. Boyer, M. Castro, C. Cronan, C. Goodale, et al. 2003. Nitrogen pollution in the northeastern United States: sources, effects, and management options. *BioScience* 53:357-374.

Gordon, D.N., T.A. McMahon, B.L. Finlayson, C.J. Gippel, R.J. Nathan. 2004. Stream hydrology: an introduction for ecologists. Wiley, New York, New York, USA.

Mallin, M., V. Johnson, S. Ensign, and T. MacPherson. 2006. Factors contributing to hypoxia in rivers, lakes, and streams. *Limnology and Oceanography* 51:690-701.

Meyer, J.L. 1997. Stream health: Incorporating the human dimension to advance stream ecology. *Journal of the North American Benthological Society* 16:439-447.

Pepper, I.A., C.P. Gerba, M.L. Brusseau. 2006. Environmental and Pollution Science. Second edition. Academic Press, Elsevier, Burlington, MA, USA.

Poff, N. 1997. Landscape filters and species traits: Towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society* 16:391-409.

Simberloff, D. 1998. Flagships, umbrellas, and keystones: is single-species management passé in the landscape era? *Biological Conservation* 83: 247-257.

Smith, V.H. 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environ Science and Pollution Research* 10:126-139.

U.S. Environmental Protection Agency. 2006. Wadeable Streams Assessment: a collaborative survey of the nation's streams. (<http://www.epa.gov/owow/streamsurvey/>) Accessed August 17, 2010.

U.S. Geological Survey 2010. The Gulf of Mexico hypoxic zone. (http://toxics.usgs.gov/hypoxia/hypoxic_zone.html). Accessed August 17, 2010.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, C.E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 11:130-137.

Wallace, J. and J. Webster. 1996. The role of macroinvertebrates in stream ecosystem function. *Annual Review of Entomology* 41:115-139.

Exploring Data Sets

I. Land Cover

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

Many agencies and other organizations that have a lot of ecological data, like the U.S. Geological Survey, provide those data online with some tools to explore the data. Look at the map of NLCD land cover ([Supplement 1](#)), or on a website provided by your instructor, to answer the following questions. You may find it helpful to print out a color copy of [Supplement 1](#), or view it on your screen at a larger size (e.g., 200%) to see more detail.

1. What are the dominant types of land cover in your region? Based on your knowledge of sources of nutrients, which EPA Region(s) do you predict will have the highest and lowest nutrient values? If you are using an online tool to determine land cover of vegetation types you can refer to [Supplement 2](#) for a map of the EPA Regions.

2. Now look at the table of mean (average) values of total nitrogen (NTL) and total phosphorus (PTL) in each EPA region, located in [Supplement 3](#). Which regions have the lowest and highest mean nutrients, and how does this compare with your prediction? Are the mean values useful at characterizing streams in a region? Why or why not? Offer some suggestions about what additional information would be useful to more fully characterize nutrient levels within an EPA region.

II. Nutrients

A common first step in analyzing the distribution of data points is through the use of box plots. Box plots show you where the median of your data is – the point at which half (50%) of the data points are below that value, and half of them are above. The lower quartile, Q1, is the point below which 25% of the data is contained. The second quartile (Q2) is the same as your median. The upper quartile, Q3, is the point below which 75% of the data is contained. To find Q1, you take the median of the data points that lie between your lowest value and Q2. To find Q3, you take the median of the data points that lie between your highest value and Q2. The upper extreme is the largest value that is less than $Q3 + 1.5*(Q3-Q1)$. The lower extreme is the smallest value that is greater than $Q1 - 1.5*(Q3-Q1)$. “Outliers” are points far outside the range of the majority of the data - specifically, if their value is larger than the upper extreme, $Q3 + 1.5*(Q3-Q1)$, or less than the lower extreme, $Q1 - 1.5*(Q3-Q1)$.

3. Open the WSA data table (WSA_data_for_students.csv, also accessible from: <http://knb.ecoinformatics.org/knb/metacat/nuding.9.4/knb>) by choosing Download Data

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

for the file named WSA_data_for_students.txt. After opening the file find the median, quartile, extreme and outlier values of total nitrogen for Regions 7 and 10 separately, as indicated in Table 1. The median and quartiles can be found using the following Excel functions, or your instructor may suggest using alternative methods.

The parentheses should contain the column or row of data you wish to analyze. Here we use cells E1:E70, just as an example.

To calculate the median, type (without the quotation marks) “=MEDIAN(E1:E70)”

To calculate the Q1, type “=QUARTILE(E1:E70,1)”

To calculate Q3, type “=QUARTILE(E1:E70,3)”

The remainder of the boxplot data points can be obtained from the equations provided, and from inspection of the data.

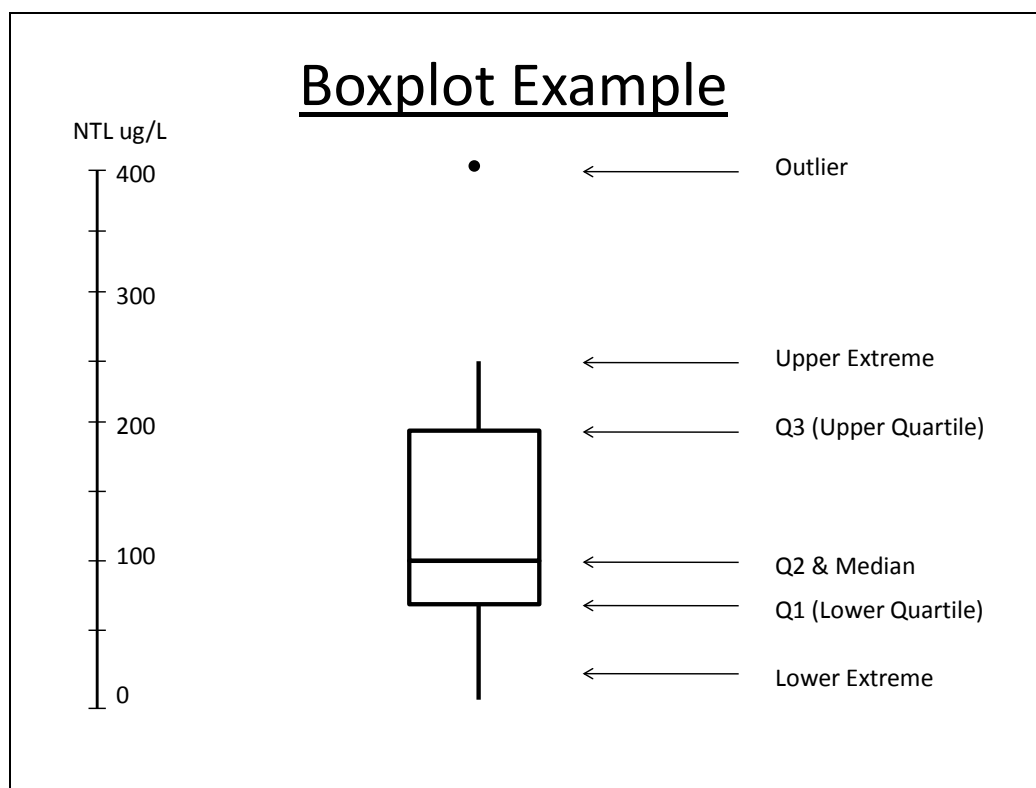


Figure 1. An example of a box plot table for total nitrogen (NTL).

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

4. Using these data, sketch box plots for each region on the same chart, using paper or software as indicated by your instructor. What percent of the data lies within the box? Describe the distribution of the data outside the box, e.g. are the points relatively close or spread widely, where are outliers, etc. Given what you have learned so far, would you expect an “outlier below Q1” to have more or less algae than other streams?

Table 1. Levels of total nitrogen (NTL) for EPA Regions 7 and 10

	Median (µg/L)	Q1 (µg/L)	Q3 (µg/L)	Upper Extreme (µg/L)	Lower Extreme (µg/L)	# outliers above Q3 (upper extreme)	# outliers below Q1 (lower extreme)
EPA Region							
EPA Region 10							

In a separate nutrient study by the EPA (EPA 2001), the lower quartile (Q1) value from water samples was recommended as the level below which nutrients in streams should be maintained in each *ecoregion* - note this is not the same as an *EPA Region*, as you can see in the map provided in [Supplement 4](#). Below are the **recommended total nitrogen values** for the ecoregions within EPA Regions 7 and 10. The full data set and a map of the ecoregions are available in [Supplement 4](#).

EPA REGION 7:

Ecoregion IV (Great Plains Grass and Shrublands) – 560 ug/L
 Ecoregion V (South Central Cultivated Great Plains) – 880 ug/L
 Ecoregion VI (Corn Belt and Northern Great Plains) – 2180 ug/L

EPA REGION 10:

Ecoregion II (Western Forested Mountain) – 120 ug/L
 Ecoregion III (Xeric West) – 380 ug/L

Compare the information you generated in Table 1 with the EPA recommended nutrient criteria.

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

5. How do the values you calculated for Region 7 and Region 10 compare to EPA's recommendations for various ecosystem types, or ecoregions, within these two regions?
6. Based on the data and maps you've looked at, can you assert whether the waters in EPA Regions 7 and 10 are "clean" or "polluted"? Explain your answer using data from parts I and II. If you feel more information is needed to make an assertion, state what information you would need to have. Note: there is not one "right" answer here, so what is important is how well you explain your answer.
7. How does the region in which you live compare to other regions? Briefly hypothesize factors that may influence nutrient levels in your region.

III. Macroinvertebrates

Streams often contain a large number and a wide variety of benthic macroinvertebrates. Stream ecologists often look for the pollution-sensitive insects, because they serve as bioindicators of stream health. The following taxonomic orders of pollutant-intolerant insects are collectively referred to as "EPT" because they are such a useful grouping of insects that react strongly to pollution: Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies). These insects begin their lives in the water, and later emerge as adults to live on the land.

Macroinvertebrates can be sampled by shuffling a net along the bottom of the stream bed and counting the number and types of insects brought up in the net. Ecologists have developed sophisticated methods of characterizing these benthic macroinvertebrate communities – rather than just counting the number of insects, we consider things like the diversity of the species present, and how many of those species are known to be pollution intolerant (e.g., EPT).

The EPA developed a method of characterization that worked best for the Wadeable Stream Assessment and called it "MMI" for "Multi Metric Index" (EPA 2007). It includes the following six categories:

1. Taxonomic richness – the number of distinct taxa (e.g., species)
2. Taxonomic composition – a measure of the abundance of the ecologically important taxa in the sample
3. Taxonomic diversity – the distribution of the numbers of organisms in different taxa
4. Feeding groups – the distribution of macroinvertebrates that have different feeding habits (e.g. leaf shredders vs. algal feeders)
5. Habits – the distribution of macroinvertebrates that burrow, cling, crawl and/or swim

TIEE

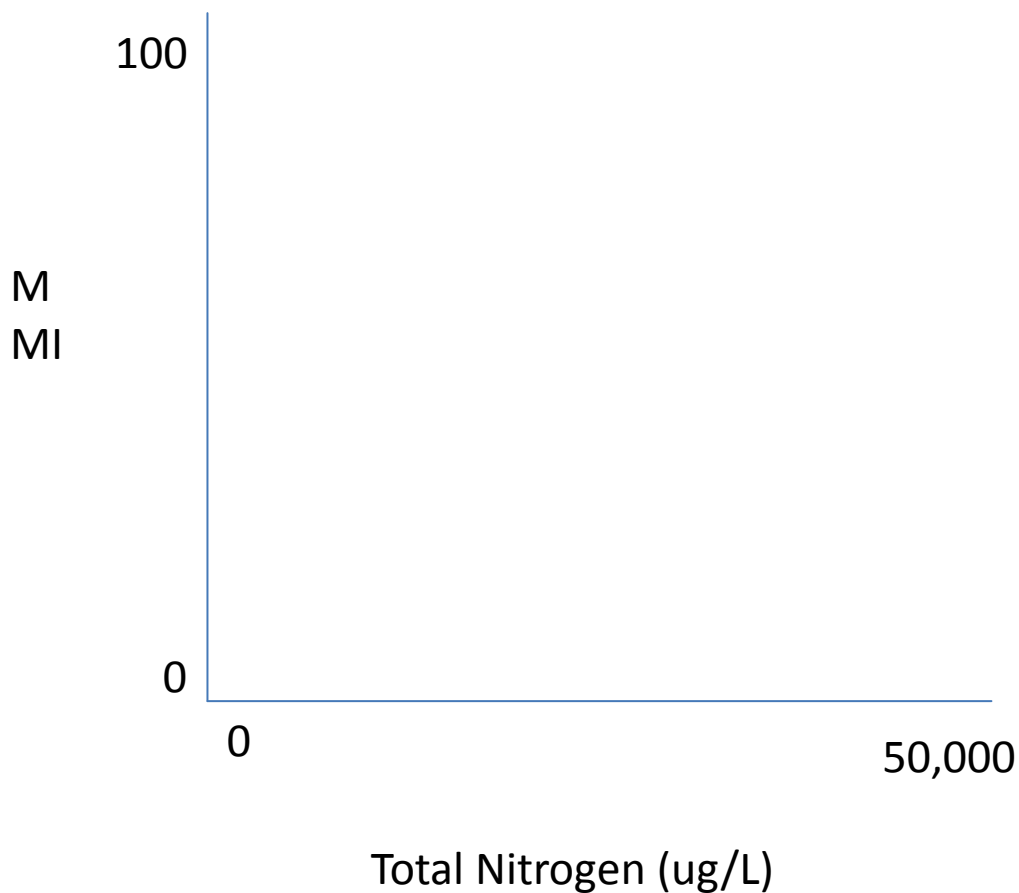
Teaching Issues and Experiments in Ecology - Volume 8, March 2012

6. Pollution tolerance – a measure of how many taxa present are pollution tolerant and intolerant

More information about these groups can be found in the WSA report (Ch 2, pgs. 27 – 29). In all cases, a higher value indicates a greater diversity of organisms and is considered to be indicative of a healthier stream. A high MMI score (max = 100) tends to indicate a healthy stream, and low score (min = 0) tends to indicate an impaired stream.

8. Make a Prediction. If we plotted the total nitrogen values (for the entire nation) on the x-axis, and the benthic macroinvertebrate index values on the y-axis, what do you think the graph would look like? Draw a line that expresses your prediction on the graph below.

MMI vs. Total Nitrogen



TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

9. Using actual WSA data, create a scatter plot of MMI vs. NTL and MMI vs. PTL, with a trend line for each. When you do a linear regression of MMI with NTL, and MMI with PTL, the R^2 can tell you how well the trend line fits the data (how strong is the relationship between the x and y variables). Display the R^2 on each chart.
- How does it appear that the macroinvertebrate community changes at low and high levels of total nitrogen and phosphorus?
 - What are the similarities and differences between the two graphs? Is this what you expected?

10. Because the nitrogen data have such a wide range of values (over three orders of magnitude) and many data points are clustered at one end, we may want to use the logarithm (log) of the nitrogen data to plot on the x-axis instead. Create this plot with MMI on the y-axis and Log NTL on the x-axis, and add a trend line.

Comment on the results. Explain why the numbers on the x-axis changed as they did - please give a sample calculation. How did the log transformation change the distribution of the points and the fit of the trend line (the R^2 value)? Would you also perform this operation on the total phosphorus data? Why or why not?

11. On the Log chart you just plotted, draw circles over the areas where you would expect data from Region 7 and Region 10 to be located.

12. Construct a regional comparison of data. Create one plot of MMI vs. Log NTL for Region 7 and one for Region 10. Make sure your axes on both graphs have identical value ranges to allow for easier comparison. In a paragraph, discuss differences in the slopes of the lines, and the distribution of points between the two plots. Do these graphs demonstrate what you expected to see? Why or why not?

IV. Synthesis Activities

One of the primary purposes of the Wadeable Stream Assessment was to be able to make a statement about the condition of the nation's waters. In this exercise you used the same data to look at distributions of nutrient data, to create charts of ecological trends both nationally and regionally, and connected these concepts to land use patterns. Write a 2-page response paper based on the following questions. Tables and diagrams are expected, and they do not count toward the page limit. Also, cite all your sources you use in your investigation (e.g. peer-reviewed papers, published reports, or datasets).

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

A. Give the nation a grade (A+ - F) based on its level of nutrient pollution in the waters. Your explanation for the grade should be based in part on the results and analyses done in this exercise, in part from external resources and in part from your original ideas. At the end, make suggestions about how the nation can improve its grade.

B. Compare macroinvertebrate populations and stream conditions from two EPA Regions, or from two states, using EPA's on-line WSA data plotter: <http://www.epa.gov/bioiweb1/statprimer/ProbabilisticSampling/index.html>. (If the link isn't working, do a quick search of EPA's website.) Water chemistry, habitat and watershed characteristics can affect macroinvertebrate populations. The U.S. Environmental Protection Agency has published data on a number of these pollutants, and the information can generally be found at <http://www.epa.gov>. The graphs are easy to generate, but the options are nearly endless, so you will need to do some research first to inform your investigation. Some of the resources in this exercise may give a helpful start as you develop your questions, e.g. for understanding nutrient relationships with low oxygen (http://toxics.usgs.gov/hypoxia/hypoxic_zone.html) or [why macroinvertebrates are good indicators of pollution](http://www.epa.gov/bioindicators/statprimer/WSAUseCaseTeacher.html) (<http://www.epa.gov/bioindicators/statprimer/WSAUseCaseTeacher.html>).

As a starting point you may want to think about the grade you gave the nation, and the EPA Regions that seem to be pushing that grade higher or lower, according to your explanation.

Be sure to include graphs as appropriate. You can generate box plots to compare a single metric between two EPA Regions ([Supplement 2](#)), and scatter plots to investigate how one metric affects another. Suggest possible causes for any differences you observe between the two places, and suggest ways to determine the actual causes of those differences (e.g. specific scientific investigations).

ADDITIONAL RESOURCES

Benthic Macroinvertebrate Indices:

Stoddard, J.L., A.T. Herlihy, D.V. Peck, R.M. Hughes, T.R. Whittier, E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society* 27: 878-891.

Klemm, D.J., K.A. Blocksom, F.A. Fulk, A.T. Herlihy, R.M. Hughes, P.R. Kaufmann, D.V. Peck, J.L. Stoddard, W.T. Theony, M.B. Griffith. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic highland stress. *Environmental Management*. 31:656-669.

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

U.S. Environmental Protection Agency. 2001. Ambient water quality criteria recommendations: information supporting the development of state and tribal nutrient criteria.

(http://water.epa.gov/scitech/swguidance/waterquality/standards/criteria/aqlife/pollutants/nutrient/rivers_index.cfm). Accessed April 17, 2010.

U.S. Environmental Protection Agency. 2009. Invertebrates as indicators.

(<http://www.epa.gov/bioiweb1/html/invertebrate.html>) Accessed June 2010.

NOTES TO FACULTY

Students will get the most out of this lesson if they have had some exposure to stream ecology, especially to nutrient pollution sources and effects, water quality metrics, and aquatic macroinvertebrates. One way to introduce some of these basic concepts in stream ecology in advance of this lesson may be to use the jigsaw method, in which teams of students learn different aspects of stream ecology – such as stream water quality parameters, some macroinvertebrate or algal ecology, nutrient pollution sources and impacts - and then reassemble in new groups to teach each other. For a more detailed description of this technique, visit:

http://tiee.esa.org/teach/teach_glossary.html#studentactive. Alternatively, these topics could be addressed in lecture and/or in readings. Additionally, instructors may use this lesson as an opportunity to discuss the concept of “ecosystem health”, a controversial topic in science but with intuitive resonance among the general public (Simberloff 1998).

Some of the work in this exercise involves simple data exploration, and some of it includes more complex techniques such as box plots, logarithmic transformations, and simple linear regressions. The tools used to construct these plots are left to the instructor’s discretion – a spreadsheet like Microsoft Excel works well for the linear regressions and logarithmic transformations, though the box plots are cumbersome in Excel. In this exercise, we recommend using Excel functions to determine the quartiles, median and outliers, and have the students sketch out the box plots on paper. However, many resources are available on-line to help construct box plots in Excel, using min, max, quartile and plot functions

(<http://support.microsoft.com/kb/155130>). Alternatively, statistical software may be utilized.

This activity in its entirety incorporates all thinking levels identified in Bloom’s Taxonomy: knowledge, comprehension, application, analysis, synthesis, and evaluation. One option to help students complete the activity is to have students work in pairs or in small groups. Techniques such as “Think-Pair-Share,” or “Turn-to-Your-Neighbor” could be effectively utilized on questions 1, 7, 8 and 11 (see above TIEE link for descriptions of these student active approaches). Using these techniques may

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

require informing students in advance that they may/should work in pairs on some or all questions, since in a computer lab setting it is rare that all students are addressing the same question simultaneously.

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

Exploring Data Sets

I. Land Cover

At the time of publication, the U.S. Geological Survey provides an online application that allows students to explore spatial data:

<http://gisdata.usgs.net/website/MRLC/viewer.htm>. The following instructions can be provided to your students if this link is live at the time you teach, or you may be able to find this or a similar application elsewhere on the USGS website.

Many agencies and other organizations that have a lot of ecological data, like the U.S. Geological Survey, provide those data online with some tools to explore the data. Look at the map of NLCD land cover ([Supplement 1](#)), or on a website provided by your instructor, to answer the following questions. You may find it helpful to print out a color copy of [Supplement 1](#), or view it on your screen at a larger size (e.g., 200%) to see more detail.

1. What are the dominant types of land cover in your region? Based on your knowledge of sources of nutrients, which EPA Region(s) do you predict will have the highest and lowest nutrient values? If you are using an online tool to determine land cover of vegetation types you can refer to [Supplement 2](#) for a map of the EPA Regions.
2. Now look at the table of mean (average) values of total nitrogen (NTL) and total phosphorus (PTL) in each EPA region, located in [Supplement 3](#). Which regions have the lowest and highest mean nutrients, and how does this compare with your prediction? Are the mean values useful at characterizing streams in a region? Why or why not? Offer some suggestions about what additional information would be useful to more fully characterize nutrient levels within an EPA region.

The region with the highest mean NTL is Region 7, and the region with the lowest mean NTL is Region 10. The region with the highest PTL is Region 7 (though 6 is close behind) and the regions with the lowest PTL are Regions 10 and 21.

A mean is an average value – therefore with this single number it compresses all the data points into one. Therefore it does not tell you any information about extremely high or low values. Also, the mean may not be a common value in a region – the region could have a lot of very low values, a lot of very high values, and the mean would be somewhere in the middle.

Additional information could include an understanding of the distribution of the data: the extreme nutrient values in a region, the median, the mode. It would also be useful to know the spatial distribution of nutrient levels along smaller streams and larger rivers, lakes, etc.

II. Nutrients

A common first step in analyzing the distribution of data points is through the use of box plots. Box plots show you where the median of your data is – the point at which half (50%) of the data points are below that value, and half of them are above. The lower quartile, Q1, is the point below which 25% of the data is contained. The second quartile (Q2) is the same as your median. The upper quartile, Q3, is the point below which 75% of the data is contained. To find Q1, you take the median of the data points that lie between your lowest value and Q2. To find Q3, you take the median of the data points that lie between your highest value and Q2. The upper extreme is the largest value that is less than $Q3 + 1.5*(Q3 - Q1)$. The lower extreme is the smallest value that is greater than $Q1 - 1.5*(Q3 - Q1)$. “Outliers” are points far outside the range of the majority of the data - specifically, if their value is larger than the upper extreme, $Q3 + 1.5*(Q3 - Q1)$, or less than the lower extreme, $Q1 - 1.5*(Q3 - Q1)$.

3. Open the WSA data table (WSA_data_for_students.csv, also accessible from: <http://knb.ecoinformatics.org/knb/metacat/nuding.9.4/knb>) by choosing Download Data for the file named WSA_data_for_students.txt. After opening the file find the median, quartile, extreme and outlier values of total nitrogen for Regions 7 and 10 separately, as indicated in Table 1. The median and quartiles can be found using the following Excel functions, or your instructor may suggest using alternative methods.

The parentheses should contain the column or row of data you wish to analyze. Here we use cells E1:E70, just as an example.

To calculate the median, type (without the quotation marks)
“=MEDIAN(E1:E70)”

To calculate the Q1, type “=QUARTILE(E1:E70,1)”

To calculate Q3, type “=QUARTILE(E1:E70,3)”

The remainder of the boxplot data points can be obtained from the equations provided, and from inspection of the data.

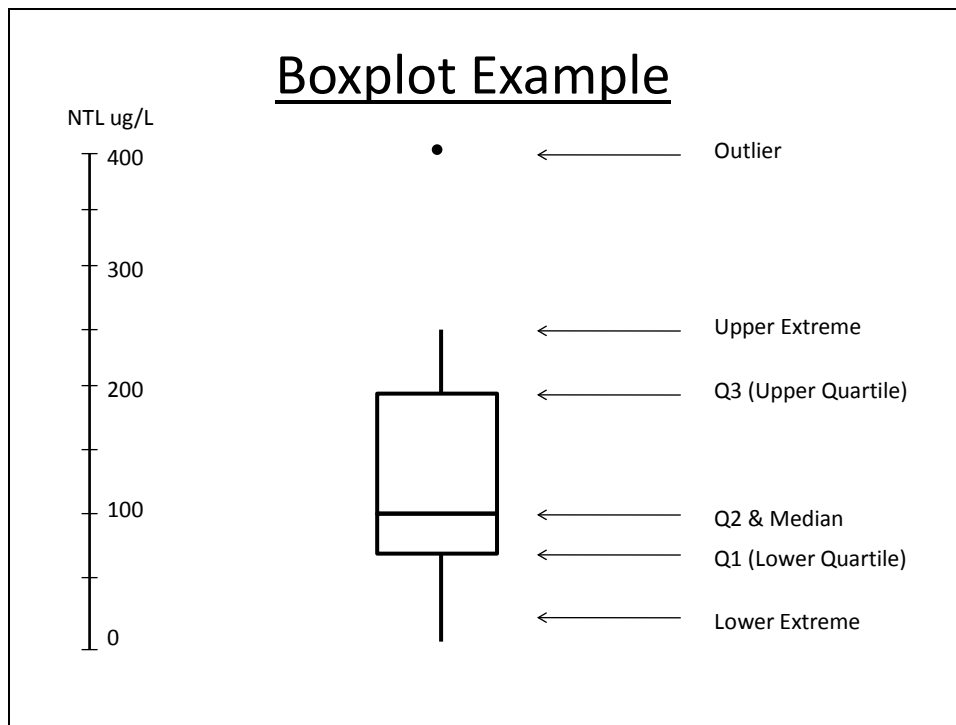


Figure 1. An example of a box plot table for total nitrogen (NTL).

4. Using these data, sketch box plots for each region on the same chart, using paper or software as indicated by your instructor. What percent of the data lies within the box? Describe the distribution of the data outside the box, e.g. are the points relatively close or spread widely, where are outliers, etc. Given what you have learned so far, would you expect an “outlier below Q1” to have more or less algae than other streams?

Table 1. Levels of total nitrogen (NTL) for EPA Regions 7 and 10

	Median (µg/L)	Q1 (µg/L)	Q3 (µg/L)	Upper Extreme (µg/L)	Lower Extreme (µg/L)	# outliers above Q3 (upper extreme)	# outliers below Q1 (lower extreme)
EPA Region 7	3450	759	9225	21924	154	2	0
EPA Region 10	99	55	177.5	361.25	11	17	0

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

This is a question designed to make the students stop and think about what the statistical terms really mean. An outlier below Q1 is likely to have lower algal biomass, since the stream has lower nutrients.

In a separate nutrient study by the EPA (EPA 2001), the lower quartile (Q1) value from water samples was recommended as the level below which nutrients in streams should be maintained in each *ecoregion* - note this is not the same as an *EPA Region*, as you can see in the map provided in [Supplement 4](#). Below are the **recommended total nitrogen values** for the ecoregions within EPA Regions 7 and 10. The full data set and a map of the ecoregions are available in [Supplement 4](#).

EPA REGION 7:

Ecoregion IV (Great Plains Grass and Shrublands) – 560 ug/L

Ecoregion V (South Central Cultivated Great Plains) – 880 ug/L

Ecoregion VI (Corn Belt and Northern Great Plains) – 2180 ug/L

EPA REGION 10:

Ecoregion II (Western Forested Mountain) – 120 ug/L

Ecoregion III (Xeric West) – 380 ug/L

Compare the information you generated in Table 1 with the EPA recommended nutrient criteria.

5. How do the values you calculated for Region 7 and Region 10 compare to EPA's recommendations for various ecosystem types, or ecoregions, within these two regions?

6. Based on the data and maps you've looked at, can you assert whether the waters in EPA Regions 7 and 10 are "clean" or "polluted"? Explain your answer using data from parts I and II. If you feel more information is needed to make an assertion, state what information you would need to have. Note: there is not one "right" answer here, so what is important is how well you explain your answer.

The responses would ideally include mention of the following:

- Region 7's Q1 falls between Ecoregion 4 and 5 levels, but Ecoregion 6 is quite high, comparatively.
- Region 10's Q1 is quite a bit below the ecoregion levels, but differences are similar to the differences between Region 7's Q1 and ecoregions 4 and 5 data.
- The overall trend of nitrogen levels in Regions 7 compared with Region 10 is consistent across both data sets, and also in keeping with land use trends (i.e. agricultural regions tend to have higher nutrient values).

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

- The number of sampled points in Regions 10 is more than 3 times that of Region 7, which could yield less reliable results
- Additional information might include other important pollutants not considered here, measures of aquatic habitat and ecosystem integrity, box plots of second study, the number of sites sampled in second study

7. How does the region in which you live compare to other regions? Briefly hypothesize factors that may influence nutrient levels in your region.

III. Macroinvertebrates

Streams often contain a large number and a wide variety of benthic macroinvertebrates. Stream ecologists often look for the pollution-sensitive insects, because they serve as bioindicators of stream health. The following taxonomic orders of pollutant-intolerant insects are collectively referred to as “EPT” because they are such a useful grouping of insects that react strongly to pollution: Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddisflies). These insects begin their lives in the water, and later emerge as adults to live on the land.

Macroinvertebrates can be sampled by shuffling a net along the bottom of the stream bed and counting the number and types of insects brought up in the net. Ecologists have developed sophisticated methods of characterizing these benthic macroinvertebrate communities – rather than just counting the number of insects, we consider things like the diversity of the species present, and how many of those species are known to be pollution intolerant (e.g., EPT).

The EPA developed a method of characterization that worked best for the Wadeable Stream Assessment and called it “MMI” for “Multi Metric Index” (EPA 2007). It includes the following six categories:

1. Taxonomic richness – the number of distinct taxa (e.g., species)
2. Taxonomic composition – a measure of the abundance of the ecologically important taxa in the sample
3. Taxonomic diversity – the distribution of the numbers of organisms in different taxa
4. Feeding groups – the distribution of macroinvertebrates that have different feeding habits (e.g. leaf shredders vs. algal feeders)
5. Habits – the distribution of macroinvertebrates that burrow, cling, crawl and/or swim
6. Pollution tolerance – a measure of how many taxa present are pollution tolerant and intolerant

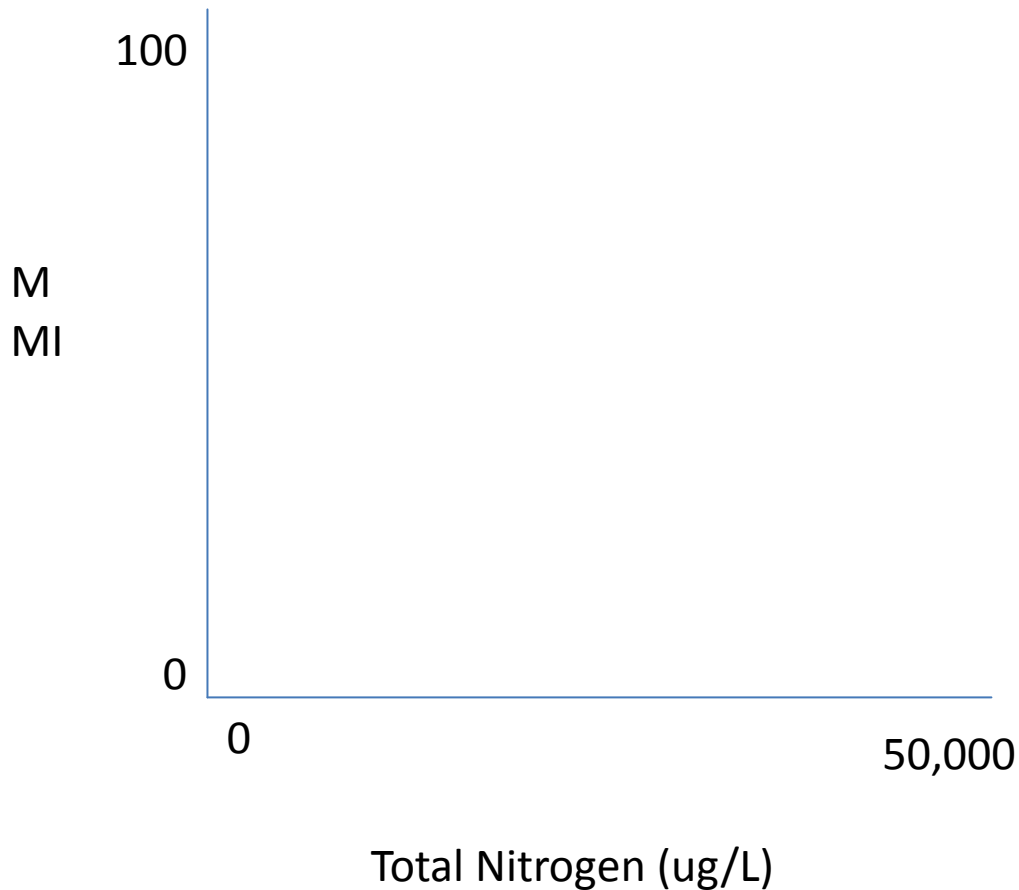
More information about these groups can be found in the WSA report (Ch 2, pgs. 27 – 29). In all cases, a higher value indicates a greater diversity of organisms and is considered to be indicative of a healthier stream. A high MMI score (max = 100) tends to indicate a healthy stream, and low score (min = 0) tends to indicate an impaired stream.

In our opinion, presenting the calculation for MMI may lengthen this exercise beyond normal limits, but we strongly urge instructors to read about the MMI in the WSA report before class in case students are confused.

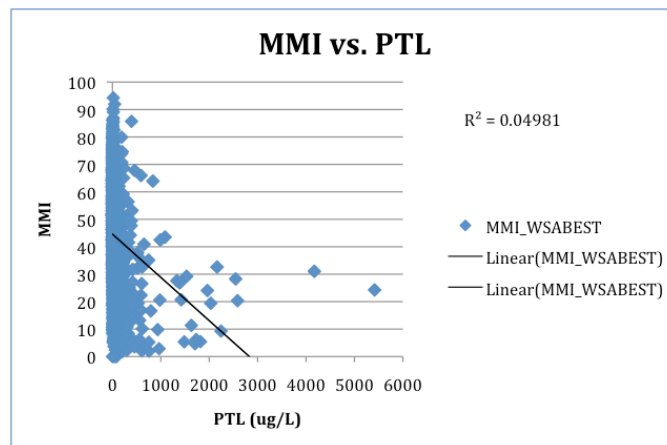
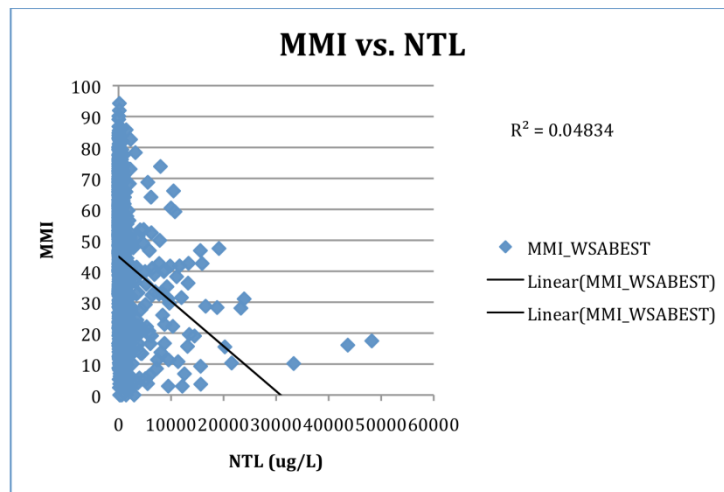
8. Make a Prediction. If we plotted the total nitrogen values (for the entire nation) on the x-axis, and the benthic macroinvertebrate index values on the y-axis, what do you think the graph would look like? Draw a line that expresses your prediction on the graph below.

Again this is designed to make students stop and think about the relationships that have been discussed. More nutrients generally means an “unhealthy” stream, and low values usually indicate a healthy stream, so we might expect negative relationships overall. You should be aware that there are certainly complications with stating a simple relationship like that – e.g., naturally nutrient-poor, healthy systems may be relatively low in diversity, you could imagine an alpine stream with high UV light exposure presenting an environment where relatively few taxa can persist. See Clarke et al. (2008) for examples of some variability in headwater systems.

MMI vs. Total Nitrogen



9. Using actual WSA data, create a scatter plot of MMI vs. NTL and MMI vs. PTL, with a trend line for each. When you do a linear regression of MMI with NTL, and MMI with PTL, the R^2 can tell you how well the trend line fits the data (how strong is the relationship between the x and y variables). Display the R^2 on each chart.
- How does it appear that the macroinvertebrate community changes at low and high levels of total nitrogen and phosphorus?
 - What are the similarities and differences between the two graphs? Is this what you expected?



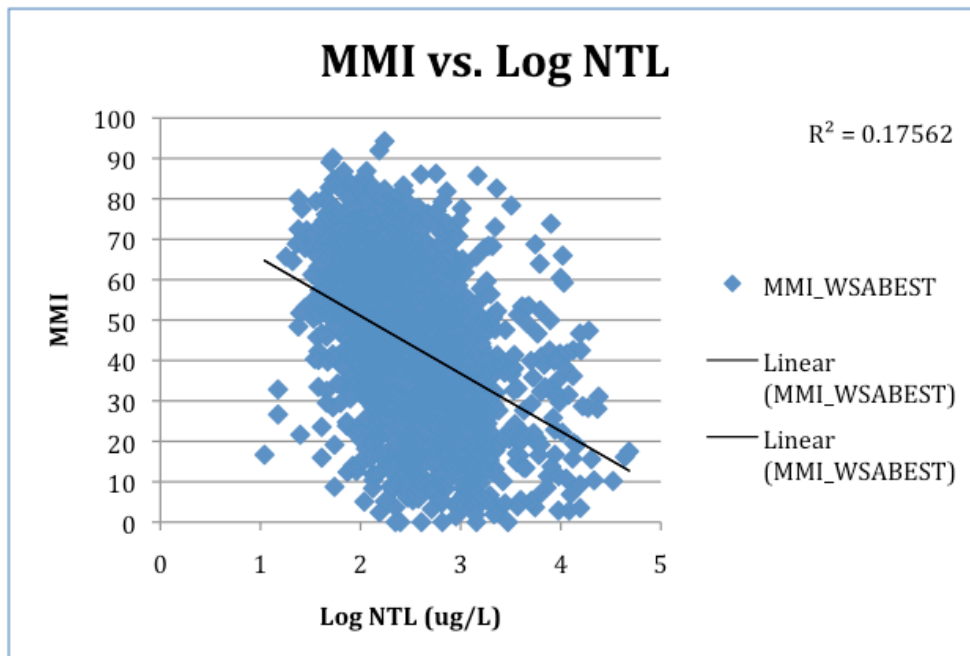
The graphs illustrate that a (nearly) full range of MMI values can exist at low levels of total nitrogen and phosphorus, however at high levels of the nutrient only lower MMI scores are present. This is likely an indication that many macroinvertebrates cannot persist at high levels of nutrient pollution, but the absence of nutrients in the water does not guarantee their presence as there may be other environmental factors controlling their presence.

The graphs have a very similar distribution of data points, indicating that many streams have relatively low levels of nutrients present and only a few streams have very high levels of nutrients present. Also, as stated earlier, they have a similar effect on macroinvertebrate species. The difference in scale shows that nitrogen is present

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

in higher quantities by an order of magnitude on average, which may be attributable to the composition of fertilizer (assuming fertilizer is the primary source of nutrient pollution).



10. Because the nitrogen data have such a wide range of values (over three orders of magnitude) and many data points are clustered at one end, we may want to use the logarithm (log) of the nitrogen data to plot on the x-axis instead. Create this plot with MMI on the y-axis and Log NTL on the x-axis, and add a trend line.

Comment on the results. Explain why the numbers on the x-axis changed as they did - please give a sample calculation. How did the log transformation change the distribution of the points and the fit of the trend line (the R^2 value)? Would you also perform this operation on the total phosphorus data? Why or why not?

The x-axis now reflects the log of NTL, which is the exponent to which 10 must be raised to give the NTL value (in ug/L):

$$\text{Log}_{10} 100 = 2$$

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

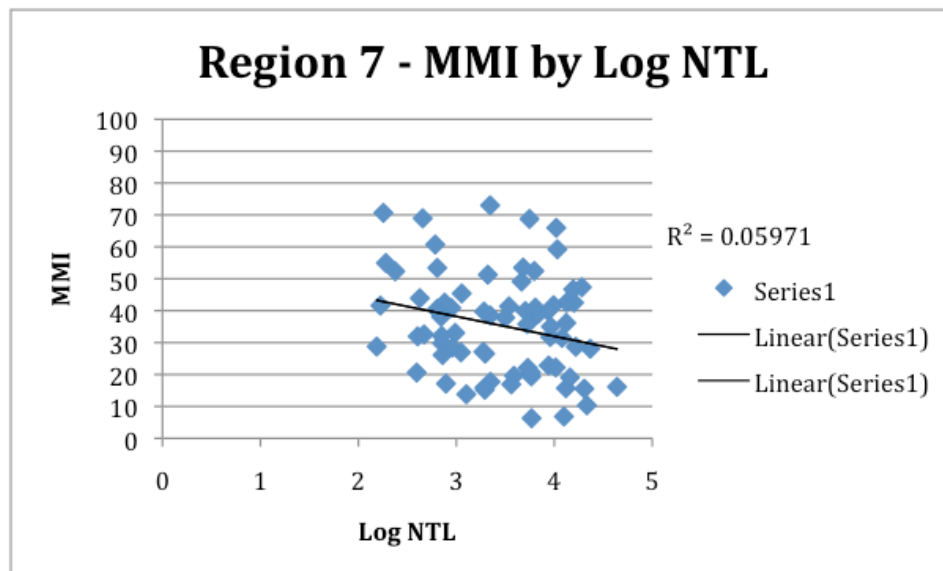
The “100” here is a sample value from the total nitrogen values we plotted along the original x-axis. These numbers ranged from 0 – 50,000 ug/L. We performed a log transformation of those nitrogen values, which subsequently yielded a range of 0 - 5 on the new x-axis.

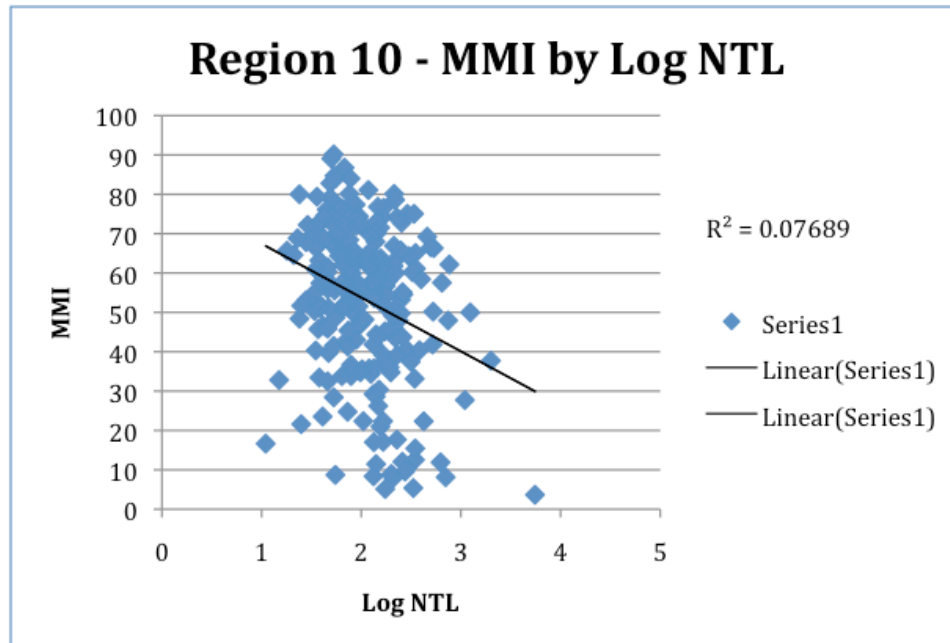
The shape of the distribution has changed significantly since we reduced the wide spread of the data to a smaller scale. Whereas in the previous graph a lot of the data were clustered at the left at values less than 5,000 ug/L, the second graph shows that the majority of the data is located between 100 and 1000 ug/L levels.

Yes the log transform process would be appropriate for the phosphorus data because the data spans over a few orders of magnitude (0 – 5000 ug/L), the majority of which is clustered at the lower levels.

11. On the Log chart you just plotted, draw circles over the areas where you would expect data from Region 7 and Region 10 to be located.

12. Construct a regional comparison of data. Create one plot of MMI vs. Log NTL for Region 7 and one for Region 10. Make sure your axes on both graphs have identical value ranges to allow for easier comparison. In a paragraph, discuss differences in the slopes of the lines, and the distribution of points between the two plots. Do these graphs demonstrate what you expected to see? Why or why not?





Both graphs have negative slopes, indicating that as nitrogen increases, MMI (and likely stream “health”) goes down. The log nutrient values are lower in Region 10, as expected and the range of MMI reaches higher, which is consistent with the ecology of macroinvertebrates.

IV. Synthesis Activities

One of the primary purposes of the Wadeable Stream Assessment was to be able to make a statement about the condition of the nation’s waters. In this exercise you used the same data to look at distributions of nutrient data, to create charts of ecological trends both nationally and regionally, and connected these concepts to land use patterns. Choose one of the following, and write a 1 – 2 page response paper. Tables and diagrams are expected, and they do not count toward the page limit. Also, cite all your sources you use in your investigation (e.g. peer-reviewed papers, published reports, or datasets).

- A. Give the nation a grade (A+ - F) based on its level of nutrient pollution in the waters. Your explanation for the grade should be based in part on the results and analyses done in this exercise, in part from external resources and in part from your original ideas. At the end, make suggestions about how the nation can improve its grade.

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

- B. Compare macroinvertebrate populations and stream conditions from two EPA Regions, or from two states, using EPA's on-line WSA data plotter: <http://www.epa.gov/bioiweb1/statprimer/ProbabilisticSampling/index.html>. (If the link isn't working, do a quick search of EPA's website.) Water chemistry, habitat and watershed characteristics can affect macroinvertebrate populations. The U.S. Environmental Protection Agency has published data on a number of these pollutants, and the information can generally be found at <http://www.epa.gov>. The graphs are easy to generate, but the options are nearly endless, so you will need to do some research first to inform your investigation. Some of the resources in this exercise may give a helpful start as you develop your questions, e.g. for understanding nutrient relationships with low oxygen (http://toxics.usgs.gov/hypoxia/hypoxic_zone.html) or why macroinvertebrates are good indicators of pollution (<http://www.epa.gov/bioindicators/statprimer/WSAUseCaseTeacher.html>).

As a starting point you may want to think about the grade you gave the nation, and the EPA Regions that seem to be pushing that grade higher or lower, according to your explanation.

Be sure to include graphs as appropriate. You can generate box plots to compare a single metric between two EPA Regions ([Supplement 2](#)), and scatter plots to investigate how one metric affects another. Suggest possible causes for any differences you observe between the two places, and suggest ways to determine the actual causes of those differences (e.g. specific scientific investigations).

The possibilities are potentially overwhelming in this WSA data plotter. If you ask students to do this, you will certainly want to spend some time looking through the possibilities on your own before class. Students may most fruitfully focus in on 2 areas that they know well, or on 2 areas where they expect to see big differences in land use. Depending on their comfort with statistics and graphing, you may ask them to produce their graphs simply from the "PDF" button on the data plotter, which would allow them to cut and paste the resulting map(s) into a report, or they can right-click on "CSV data" to save the underlying data to their desktops and open the data sheet in Excel or another application (e.g., R). If students use this option, they would be able to download several data sets separately (e.g., a nutrient value, and EPT) and then merge them into a single data sheet to run a regression similar to what they did in the preceding exercise. If you are comfortable with geospatial functions yourself, you may also ask them to download data in "KML" which opens up broader possibilities for geospatial visualizations.

The WSA report provides more information on the parameters that you see in the dropdown menus for the WSA plotter. However, some of the words may seem a little

TIEE

Teaching Issues and Experiments in Ecology - Volume 8, March 2012

cryptic if you are not a stream ecologist! The Vannote et al. 1980 paper will provide further guidance for what one might expect among different categories of macroinvertebrates, and we also strongly suggest having Merritt and Cummins (1996) on hand so that students can get much more detailed information about the macroinvertebrate categories, if they wish.

ADDITIONAL RESOURCES

Clarke, A., R. MacNally, N. Bond, and P.S. Lake. 2008. Macroinvertebrate diversity in headwater streams: a review. *Freshwater Biology* 53:1707-1721.

Merritt, R.W. and K.W. Cummins. 1996. An introduction to the aquatic insects of North America. Dubuque, Iowa: Kendall/Hunt Publishing Company.

Simberloff, D. 1998. Flagships, umbrellas, and keystones: is single-species management passé in the landscape era? *Biological Conservation* 83:247-257.

COPYRIGHT STATEMENT

The Ecological Society of America (ESA) holds the copyright for TIEE Volume 8, and the authors retain the copyright for the content of individual contributions (although some text, figures, and data sets may bear further copyright notice). No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner. Use solely at one's own institution with **no intent for profit** is excluded from the preceding copyright restriction, unless otherwise noted. Proper credit to this publication must be included in your lecture or laboratory course materials (print, electronic, or other means of reproduction) for each use.

To reiterate, you are welcome to download some or all of the material posted at this site for your use in your course(s), which does not include commercial uses for profit. Also, please be aware of the legal restrictions on copyright use for published materials posted at this site. We have obtained permission to use all copyrighted materials, data, figures, tables, images, etc. posted at this site solely for the uses described at TIEE site.

GENERIC DISCLAIMER

Adult supervision is recommended when performing this lab activity. We also recommend that common sense and proper safety precautions be followed by all participants. No responsibility is implied or taken by the contributing author, the editors of this Volume, nor anyone associated with maintaining the TIEE web site, nor by their academic employers, nor by the Ecological Society of America for anyone who sustains injuries as a result of using the materials or ideas, or performing the procedures put forth at the TIEE web site, or in any printed materials that derive therefrom.