

Standard Operating Procedure 19: Data Analysis, version 1.0

Adapted from Petersen et al (2008) and Dodd et al (2008).

Conclusions of ecological studies based on macroinvertebrate and other biological, chemical, and physical data are used by resource managers to better comprehend underlying system processes and develop environmental and management policies that best serve the resource.

Annual data analyses and summaries will be completed for every year that monitoring is conducted at Shenandoah National Park. Summaries can include short reports or resource briefs. These summaries will provide visual displays of select metrics, potentially in the form of control charts (Figure 1). Annual summaries will not include formal analyses of trend or in-depth data reviews unless data suggests that more analyses are needed (i.e. control chart thresholds are crossed). Comprehensive trend analyses will occur every ten years or when thresholds indicate further data investigations are prudent. Comprehensive analyses will include formal assessment of trend and more rigorous statistical review (see Jastram et al (2013) for an example).

The primary audience for most data analyses (and reports – see SOP 20) will consist of park resource managers, superintendents, and other staff who may not have an in depth background in statistical methods or may have limited time for evaluation of these analyses. Thus, to the extent possible, it is important that core data analysis and presentation methods are relatively straightforward to interpret, provide a standard format for evaluation of numerous variables, can be quickly updated whenever additional data become available, and work for many different types of indicators, whether univariate or multivariate. Control charts can fill this need. Additionally, the type and magnitude of variability or uncertainty associated with the results should be somewhat intuitive, and it may be necessary to indicate a threshold for potential management action. More detailed analysis of the data (comprehensive trend reports) generally will benefit from the use of multiple data analysis methods.

Ideally, analyses will occur across multiple spatial scales and should reflect trends at sites, streams (when multiple sites are sampled in a watershed), geologic groups or ANC groups, and elevations (i.e. $> 549\text{m}$). Numerous spatial scales of analyses support the stratification that occurred in the study design (larger scale analyses) as well assist in potentially identifying small scale threats that might occur at the site level.

Primary approaches to analyzing macroinvertebrate data will rely on metric estimation. Potential metrics for analysis of macroinvertebrate communities are outlined in the table below.

Table 1. Recommended Macroinvertebrate metrics for analyses at Shenandoah National Park.

Category	Metric	Definition	Expected response to perturbation
VA DEQ VSCI ¹	VA DEQ VSCI overall score	VA DEQ Stream condition index (Burton and Gerritsen 2003)	Decrease

MAIS	MAIS overall score	Multimetric index (Smith and Voshell 1997)	Decrease
Richness ¹	Number of taxa	The total number of taxa in the macroinvertebrate assemblage	Decrease
Richness	Number of EPT	Total number of taxa in the orders Ephemeroptera (E, mayflies), Plecoptera (P, stoneflies), and Trichoptera (T, caddisflies)	Decrease
Composition ¹	% EPT	Relative abundance of insects in the orders Ephemeroptera, Plecoptera, and Trichoptera	Decrease
Composition	% Ephemeroptera	Relative abundance of Ephemeroptera	Decrease
Composition	% Hydropsychidae/ Trichoptera	Percent abundance of insects in the family Hydropsychidae divided by the total number of caddisflies	Increase
Composition	% Leuctra/ Plecoptera	Percent abundance of the low pH tolerant stonefly Leuctra divided by the total number of stoneflies	Increase
Balance	% 5 dominant taxa	Relative abundance of the 5 most abundant taxa	Increase
Balance ¹	Simpson Diversity Index (1-D)	Accounts for species richness and abundance	Decrease
Tolerance	HBI (modified Hilsenhoff Biotic Index)	Weighted sum of total taxa by pollution tolerance values	Increase
Tolerance ¹	% Intolerant organisms	Percent abundance of macroinvertebrates with tolerance values of 0, 1, 2	Decrease
Trophic	% Scrapers	Relative abundance of the functional feeding group containing scrapers	Decrease
Trophic	% Shredders	Relative abundance of the functional feeding group containing shredders	Decrease
Habit	% Haptobenthos	Relative abundance of macroinvertebrates requiring clean, firm, coarse substrates (crawlers and clingers)	Decrease

¹Recommended for annual data summaries. All metrics will be analyzed in 10 year trend reports.

Annual Data Summaries

Using a subset of the metrics listed above, control charts can be employed to visualize trends and changes in macroinvertebrate communities (Morrison, 2008). The construction and interpretation of control charts is covered in many texts focusing on quality control in industry (Beauregard and others, 1992; Gyrna, 2001; Montgomery, 2001). The application of control charts for ecological purposes, however, is relatively straightforward. The use of control charts in environmental monitoring is discussed in the texts by McBean and Rovers (1998) and Manly (2001), although not as detailed as the texts referenced above focusing on industrial applications. Many different types of control charts could be constructed, depending upon the type of information desired. For example, control charts can be used to evaluate variables or attributes (for example, count data, richness, or biomass), to focus on measures of central tendency or dispersion, and in univariate or multivariate analyses.

Control charts allow for the visual display of trends and thresholds for management action or further data analyses. Initially, the first ten years of data will be serve as a baseline (see Morrison, 2008) and thresholds will be defined as $\pm$ two standard deviations around the ten year mean (i.e. baseline) of the statistic (see Figure 1 for example). During any year where thresholds are crossed, more targeted data analyses will be conducted to assess the driver behind the threshold transgression and to help determine the potential need for management action. Baselines and thresholds may be altered in the future to potentially better reflect meaningful management or biological thresholds that might not be captured in the initial ten year mean $\pm$ two standard deviations.

When appropriate, multivariate control charts may be used for community data (Anderson and Thompson, 2004). A software application entitled ControlChart.exe is available for constructing these types of multivariate control charts (see Anderson and Thompson, 2004). Multivariate temporal autocorrelation will violate the assumption of stochasticity upon which this method is based, however, and it is important to test for temporal autocorrelation using Mantel correlograms prior to using this method.

Although control charts have potentially wide applicability, each application may be different. A generic process for control chart construction is provided below.

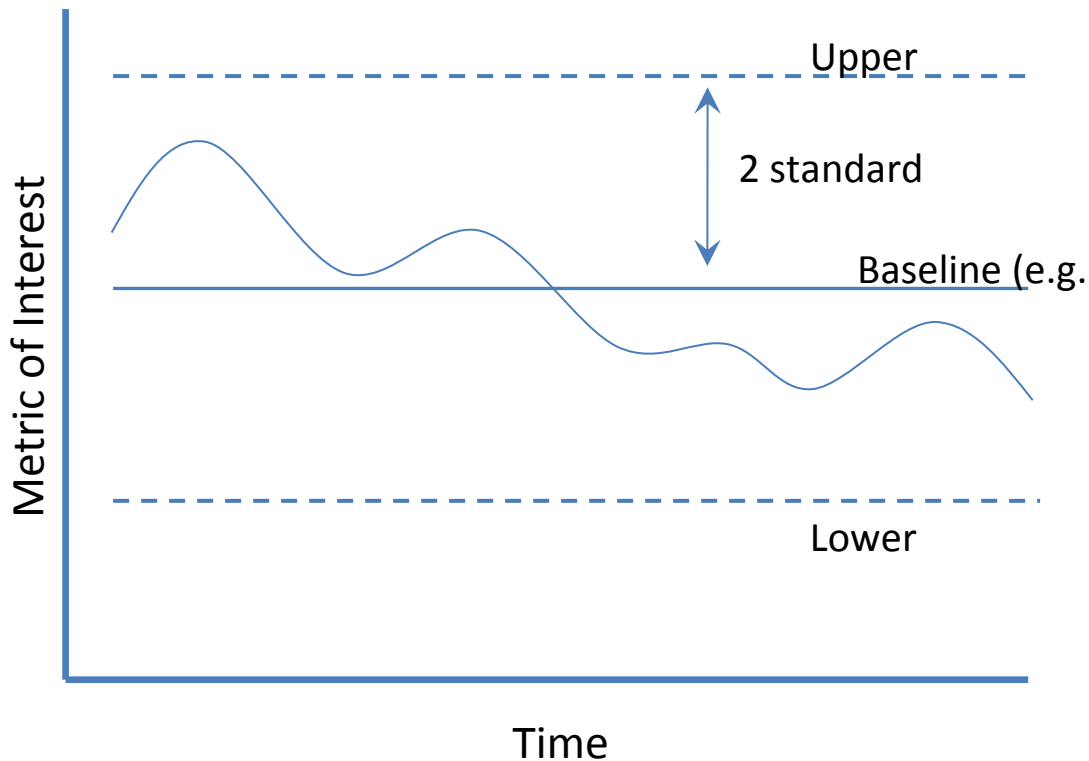


Figure 1. Control chart example.

Steps in constructing a univariate control chart (see Figure 1):

1. Determine the parameter of interest. This may be any of the metrics presented in Table 1.
2. After several data points are available, plot the values of the parameter of interest (on the y axis) against time (on the x-axis).
3. Determine a “center-line” value for this parameter, which could represent a mean of the observations, a target value, or some other value. As appropriate center-lines are identified, they will be incorporated into this SOP. For initial reporting purposes, the first ten years of monitoring data can be used to establish center-line values.
4. Establish control limits around the center-line. It is possible that only an upper control limit, or only a lower control limit, or both will be necessary, depending upon the parameter of interest and management concerns. Control limits may be based on a probability distribution and thus allow one to make statistical inferences, or they may be based on target levels set by management. Once again, determining appropriate control limits can be tricky, especially if statistical inferences are desired. For initial reporting purposes, two standard deviations around the center-line values from the first ten years of monitoring data can be used to establish control limits. Ideally, ecologically based thresholds will be identified.

5. Continue to plot values of the parameter of interest over time as new data become available. If an observation exceeds the control limit(s), this is indicative of the potential need for management action, or a more focused study.

Habitat and Water Quality Parameters

Physical and chemical habitat measurements will be estimated using summary statistics such as means, medians, standard errors and/or confidence intervals. In the simplest presentation of data, each parameter should be estimated in each year that data are available, and confidence intervals or standard errors calculated, where appropriate. For parameters where percentage cover class categories are used (such as riparian cover) the median value of the cover class will be used in calculating means and standard errors. Those parameters where presence/absence data are collected (such as fish cover and riparian cover types), a percentage will be calculated for the reach. Generally, habitat and water quality parameters will be used in comprehensive analyses to assess relationships between community data and environmental characteristics.

Comprehensive analyses

Comprehensive analyses will provide a more formal analysis of trend and will occur every ten years. Rather than requiring a detailed list of specific data analyses, the data analysis process should be flexible enough to allow the use of newly developed statistical and analytical techniques and tailoring of analyses for a variety of audiences with a variety of questions about the aquatic resources at Shenandoah (see Jastram et al (2013) for an example).

In general, comprehensive analyses will include multivariate and univariate assessments of community and environmental data. Univariate trend tests such as a seasonal Kendall test will be used to assess trends in univariate measures. Depending on the metric of interest, additional trend tests will be used with appropriate transformations. Multivariate analysis is another frequently used analysis technique and involves methods used to explain variability in community data and to identify the environmental variables that best explain, and have an assumed responsibility for, the variability measured (Gauch, 1982; Jongman and others, 1995; Everitt and Dunn, 2001; Timm, 2002). Multivariate techniques elicit a hypothesis from the biological data rather than disproving a null hypothesis. Two commonly used multivariate techniques include: ordination (such as principal components analysis, canonical correspondence analysis, and detrended correspondence analysis) and classification (such as two-way indicator species analysis). Detailed discussion of these methods can be found in several texts (Gauch, 1982; Jongman and others, 1995; Everitt and Dunn, 2001; Timm, 2002).

Comprehensive analyses, at a minimum, will include an assessment of the metrics outlined in Table 1.

References

Anderson, M. J., and A. A. Thompson. 2004. Multivariate control charts for ecological and environmental monitoring. *Ecological Applications* **14**:1921-1935.

- Beauregard, M. R., R. J. Mikulak, and B. A. Olson. 1992. A practical guide to statistical quality improvement: opening up the statistical toolbox. Van Nostrand Reinhold, New York, NY.
- Burton, J., and J. Gerritsen. 2003. A stream condition index for Virginia non-coastal streams. Owings Mills, MD.
- Dodd, H.R., D.G Peitz, G.A. Rowell, D.E. Bowles, and L.M. Morrison. 2008. Protocol for Monitoring Fish Communities in Small Streams in the Heartland Inventory and Monitoring Network. Natural Resource Report NPS/HTLN/NRR - 2008/052. National Park Service, Fort Collins, Colorado.
- Everitt, B. S., and G. Dunn. 2001. Applied Multivariate Data Analysis, 2nd edition. Hodder Arnold Publishers, London, England.
- Gauch, H. G., Jr. 1982. Multivariate analysis in community ecology. Cambridge University Press, London, England.
- Gyrna, F. M. 2001. Quality planning and analysis: from product development through use. McGraw-Hill Irwin, New York, NY.
- Jastram, J. D., C. D. Snyder, N. P. Hitt, and K. C. Rice. 2013. Synthesis and interpretation of surface-water quality and aquatic biota data collected in Shenandoah National Park, Virginia, 1979-2009. U.S. Geological Survey Scientific Investigations Report 2013-5157. Available online: <http://pubs.usgs.gov/sir/2013/5157/>.
- Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. 1995. Data analysis in community and landscape ecology. Cambridge University Press, London, England.
- Manly, B. F. J. 2001. Statistics for environmental science and management. Chapman & Hall/CRC, Boca Raton, FL.
- McBean, E. A., and F. A. Rovers. 1998. Statistical procedures for analysis of environmental monitoring data and risk assessment. Prentice Hall PTR, Upper Saddle River, NJ.
- Montgomery, D. C. 2001. Introduction to statistical quality control. John Wiley & Sons, Inc., New York, NY.
- Morrison, L.W. 2008. The use of control charts to interpret environmental monitoring data. Natural Areas Journal **28**:66-73.
- Petersen, J.C., Justus, B.J., Dodd, H.R., Bowles, D.E., Morrison, L.W., Williams, M.H., and Rowell, G.A., 2008. Methods for monitoring fish communities of Buffalo National River and Ozark National Scenic Riverways in the Ozark Plateaus of Arkansas and Missouri: Version 1: U.S. Geological Survey Open-File Report 2007-1302, 94 p. (PDF)
- Smith, E.P. and J.R. Voshell. 1997. Studies of benthic macroinvertebrates and fish in streams within EPA Region 3 for development of biological indicators of ecological condition; Park 1, Benthic

macroinvertebrates. U.S. Environmental Protection Agency, Cooperative Agreement
CF821462010.

Timm, N. H. 2002. Applied multivariate analysis. University of Pittsburgh Press, Pittsburg, PA.

SOP Revision History

All protocols, regardless of how sound they are originally, require modification as new and different information becomes available. SOP version numbers increase incrementally by tenths (e.g. version 1.1, version 1.2...etc.) for minor changes. Major revisions should be designated with the next whole number (e.g., version 2.0, 3.0, 4.0 ...). Small changes or additions to existing methods will be reviewed in-house by Shenandoah or Mid-Atlantic Network staff. However, if a complete change in methods is sought, an outside review is required.

Revision History Log:

Prev. Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #