

Standard Operating Procedure 16: Data analysis, version 1.0

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

Conclusions of ecological studies based on fish 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 summaries should be produced during every year that monitoring is conducted. These summaries provide visual displays of select metrics, largely in the form of control charts (see below). Annual summaries should not include formal analyses of trend or in-depth data reviews unless control chart thresholds are crossed. In-depth data analyses should occur every ten years or when thresholds indicate further data investigations are prudent. Ten year analyses should include formal assessment of trend and more rigorous statistical review (see Jastram et al. (2013) for an example).

This primary audience for most data analyses (and reports – see SOP 17) 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 reviews) generally will benefit from the use of multiple data analysis methods.

Primary approaches to analyzing fish data relies on metric estimation. Potential metrics and estimated variables for analysis of stream fish communities include:

1. Species richness (number of species). Species richness will be calculated for the sample site.

2. Simpson's Diversity Index (D). Simpson's Diversity Index will be calculated for the sample reach. Simpson's Diversity Index is preferable to the Shannon diversity index and will be used for data analysis because this index is independent of sample size. Simpson's Diversity Index is calculated with the formula shown below.

$$D = \sum ((n^2 - n) / (N^2 - N))$$

n = total number of individuals of a fish species, and N = total number of fish of all species.

3. Catch per area (#individuals/100m²). Catch per area (Ci) will be calculated for each captured species at the reach scale. For game fish, catch per area will also be calculated for each identifiable age class (generally age-0 and age-1+, as identified by length-frequency analysis). A length frequency analysis can be produced by creating a histogram of fish lengths from a site.

Generally, the age 0 age class can be determined through a visual assessment of the histogram and the determination of the maximum age-0 length. All individuals less than that length are age-0 and all above are age-1+. Based on 3-pass depletion, catch will be estimated using maximum likelihood methods as described in Van Deventer and Platts (1985).

$$C_i / A$$

C_i = catch of individuals of species i (either as a whole or by age)

A = area of sample reach (m^2 ; average wetted width times reach length)

4. Biomass. Biomass per area (BPA_i ; kg/ha) will be calculated for each captured species at the reach scale. For game fish, biomass per area will also be calculated for differing age classes (generally age-0 and age 1+, as identified by length-frequency analysis). To calculate biomass for game fish, simply sum the biomass of all weighed individuals by species and divide by the sample reach area. Calculate biomass per area for non-game species (B_i) via the following:

$$BPA_i = W_i N_i / A$$

W_i = average weight of fish species i (either as a whole or by age)

N_i = number of individuals of species i (either as a whole or by age)

A = area of the sample reach

5. Size Structure. Size structures of fish populations within the community can be indicative of a disturbance or resource problem. A community with primarily larger fish indicates that there is little recruitment to keep the community self-sustaining. A community with primarily smaller fish could indicate inadequate resources (e.g., food resources) or environmental stresses (e.g., acidification) that limit growth. Average length and weight (and ranges) for each gamefish species at the reach will be calculated. Averages will also be calculated for differing age classes (generally age-0 and age 1+, as identified by length-frequency analysis). For non-game fish, only average weight will be calculated (length of individual non-game fish is not collected).

6. Mean Relative Weight. Mean relative weight for each gamefish species in the reach. Mean relative weight (Wege and Andersen 1978) will be calculated for each gamefish species using established length-specific weight predicted from a published weight-length regression for the species (Blackwell et al. 2000; Hyatt and Hubert 2001).

7. Percent Composition. Percent composition by biomass will be calculated for each species in the reach. To calculate this:

First, calculate individual biomass for each species (B_i):

$$B_i = W_i N_i$$

W_i = average weight of fish species i

N_i = number of individuals of species i

Second, calculate biomass per area for each species (BPA_i) and total biomass per area (BPA_t):

$$BPA_i = B_i / A$$

$$BPA_t = \sum BPA_i$$

A = area of the sample reach

Lastly, calculate percent composition for each species (C_i)

$$C_i = BPA_i / BPA_t * 100$$

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 canopy cover) the median value of the cover class will be used in calculating means and standard errors. For those parameters where presence/absence data are collected, a percentage will be calculated for the reach.

Annual Data Summaries – Control Charts

Using several of these estimated variables listed above, control charts can be employed to visualize trends and changes in fish communities (Morrison 2008). The construction and interpretation of control charts is covered in many texts focusing on quality control in industry (Beauregar et al. 1992; Gyra 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 texts by McBean and Rovers (1998) and Manly (2001). 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 plus or minus two standard deviations around the ten year mean (i.e. baseline) of the statistic (see Figure 1 for example). Depending on the expected fish response to perturbation/stress, thresholds may occur above, below, or both above and below baseline values. During years where thresholds are crossed, more targeted data

analyses may be conducted to attempt 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 mean plus or minus 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, although decisions will always have to be made and an analyst familiar with control charts should ideally be consulted.

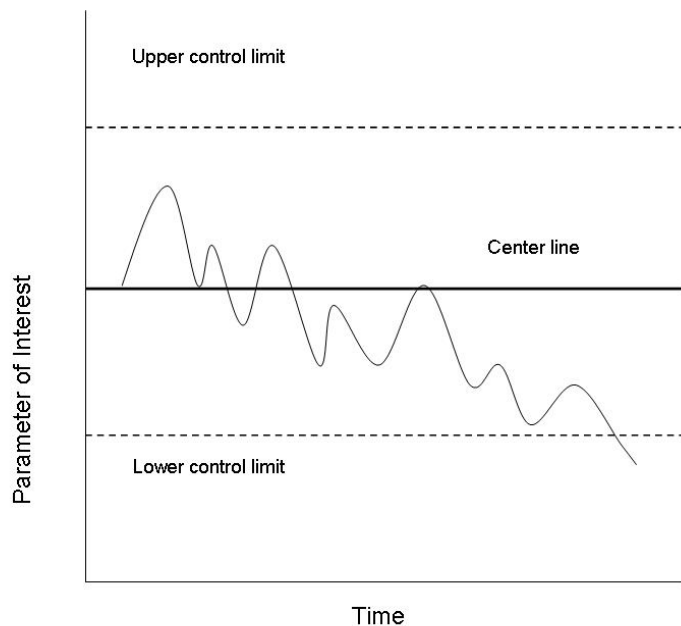


Figure 1. A univariate control chart.

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

1. Determine the parameter of interest. This may be any of the metrics presented above.
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. Determined a “center-line” value for this parameter, which could represent a mean of the observations, a target value, or some other value.
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.
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.

Control charts should be constructed after several data points are available, and updated regularly. Additional control charts can be constructed from other variables of interest as described above.

Comprehensive analyses

Comprehensive analyses provide a more formal analysis of trend and should 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.

In general, comprehensive analyses should 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).

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.

- 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.
- 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.
- Timm, N. H. 2002. Applied multivariate analysis. University of Pittsburgh Press, Pittsburgh, PA.
- Van Deventer, J.S. and W.S. Platts. 1985. A computer software system for entering, managing, and analyzing fish capture data from streams. Research Note INT-352, Intermountain For. & Range Research Station. U.S.D.A. Forest Service. Ogden, UT.

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 #