Classification

N-STEPS Objectives

- Provide regions, states, and tribes with support related to nutrient criteria development
- Provide access to expert assistance with issues related to nutrient criteria development and implementation
- Improve communication nationwide.

What Is It?

Classification, in the nutrient criteria process, is the process of identifying groups of sites that have comparable characteristics and for which comparable nutrient behavior is expected. Classification reduces the variability associated with natural differences in nutrient conditions that will allow the development of criteria that focus on protecting against principally human-influenced nutrient enrichment. There are existing classification systems such as ecoregions, nutrient regions, and physiographic provinces. This sheet refers to techniques for classifying sites using data collected from sampling programs.

A number of statistical techniques exist for use in classification. This document briefly reviews only two very commonly used multivariate approaches: cluster analysis and ordination. Each of these, in and of themselves, could take entire fact sheets and they are advanced statistical techniques. For these reasons, we provide merely an overview.

Cluster Analysis refers to a suite of methods used to partition multivariate data into groups. Ordination refers to a suite of methods used to reduce multivariate data into one or a few axes that place sites in order along continuous underlying environmental gradients. Cluster results in discrete groupings whereas ordination generally results in representations of sites along a continuous gradient. Ordination techniques are often used to validate cluster analyses.

Example Question: Are there groups of streams/rivers within our state that make sense to treat separately in terms of developing nutrient targets?

How is it Applied to Nutrient Criteria Development?

Nutrient criteria development involves three main processes: identifying relationships between biological responses and nutrient stressors, examining these relationships, and establishing nutrient and/or biological thresholds or criteria.

Classification is an integral part of developing criteria because it identifies those groups for which the same set of criteria should be developed based on natural differences in nutrient expectations. Cluster analysis refers to a number of techniques used for identifying groups based on multivariate characteristics (traditionally biological composition) and ordination is a group of techniques used to place sites along synthetic axes, where the order of those sites reflects their relative position along the underlying environmental gradients.

How Does It Work?

Both cluster analysis and ordination are principally exploratory statistical techniques (not hypothesis testing techniques). They both start with similarity/dissimilarity matrices that measure the environmental similarity/dissimilarity between each pair of sites within a dataset based on all the variables considered simultaneously. There are a variety of similarity/dissimilarity measures to use – too many to review.

Agglomerative cluster analyses use these matrices as the basis for sequentially grouping sites and eventually linking groups based on their degree of similarity. There are also divisive cluster analyses, which start with all the sites in one group and sequentially split them into smaller groups. There are a variety of both agglomerative and divisive cluster analyses – too many to review here. All of them result in a grouping structure, most commonly represented with a dendrogram.

Ordination uses the matrices to generate a set of ordination axes, upon which sites are ordered based on their similarities with respect to all the variables simultaneously. There are also a variety of ordination techniques, some of the more common being principal components analysis, non-metric multidimensional scaling, detrended correspondence analysis, and canonical correspondence analysis.

Data Requirements

Cluster analyses are primarily performed on biological data, such as site by taxa matrices of presence/absence or abundance data. There is no reason physical/chemical data could not also be used. Ordination techniques are used for both biological and physical/chemical data, but certain techniques work better with biological data than others. For example, principal components analysis assumes a linear response to environmental gradients whereas organisms tend to respond in a non-linear fashion. As a result, PCA is less commonly used for biological analyses. The techniques vary in the requirements of the data and these should be researched before using any one technique.

What Should You Look For & Report?

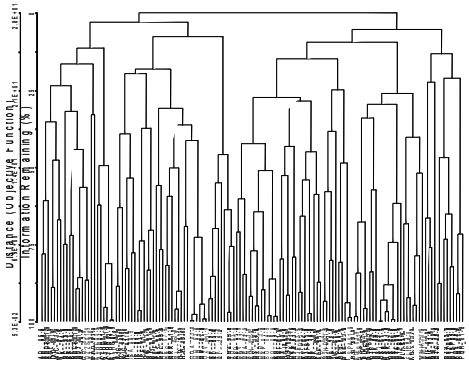
Because there are so many techniques, users should consult appropriate literature before using any one technique. Usually, the most valuable output are the plots: the dendrograms in the case of cluster analysis and the ordination plots. These show the relationship of sites to each other and allow one to visualize any patterns in the data, including groupings. However, there are also a variety of statistics generated by the different techniques that inform the user on how well the models perform and, therefore, how much faith can be put in the results.

Pros	Cons
• Effective ways to explore patterns in data	• Require substantial data
• Can be used to verify existing classifications (ecoregions) or suggest new ones	• Require advanced statistical knowledge
• Available in most standard software packages	• Many specific constraints, depending on method used

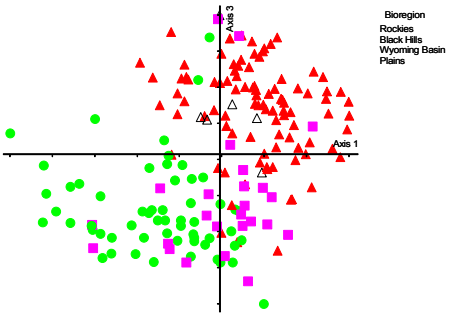
Citations

Legendre, P. and L. Legendre. 1998. Numerical Ecology. 2nd English Edition. Elsevier, New York.

Ordination Web Page. M. Palmer, Oklahoma State University. http://ordination.okstate.edu/



Cluster Analysis Dendogram



Ordination Plot (NMS)