Appendix E

Site Classification Analysis

and

Additional maps –

* Baseflow
* Mean annual air temperature
* Mean summer stream temperature
* Elevation

# Site Classification Analysis

Site classification addresses the recognition that even with the least disturbance to streams, there might be different expectations of the sampled benthic assemblage due to natural effects and influences. Natural variation in stream slope, stream size, dominant substrates, temperature, and other factors that are components of ecoregional characteristics that might cause a sample to contain more or less of certain taxa groups, sensitive taxa, or functionally specialized taxa. These types of taxa and some of the metrics derived from their traits are expected to exhibit variation not only with natural variation but also with human disturbance and unnatural stressors. When we use the benthic assemblage to indicate biological conditions relative to disturbance, we attempt to account for different expectations due to the background natural setting.

Accounting for different biological expectations was explored by an investigation of natural variation in samples from the least-disturbed reference sites. If the variation in taxa or metrics can be associated with natural categories or gradients, then those categories or gradients can be used to characterize different reference conditions. Comparisons of metrics between reference sites and those with high disturbance will be more sensitive to stressors if the natural variation is filtered out through classification.

The classification investigation proceeded through the ordination of taxa and metrics in reference sites so that samples could be organized by similar biological characteristics. To increase the sample size for the classification analyses, one ‘borderline reference’ site was included in the reference dataset (n=27; these sites are marked in Attachment C). Non-metric multidimensional scaling (NMS) ordination was used to find sites with similar taxa. Principle components analysis (PCA) was used to organize sites by similar metric values, using 45 selected metrics. In each of these ordinations, the biological gradients were mapped in two dimensions, with each axis describing orthogonal composite aspects of the community. The axes were then associated with continuous natural variables through correlation. Categorical variables (e.g., level 4 ecoregions) were superimposed on the ordination diagrams to visually discern separation of categories. Any strong associations of environmental factors with the axes prompted further investigation of the factors as possible classification variables.

Due to the region and small data set, only a few discrete site classes could possibly be recognized before the separate classes would become too small to robustly represent the reference condition in each class or to allow comparisons between reference and disturbed data within each class. For adjustment of expectations along a continuous gradient, the optimal metric values were defined relative to the strongly correlated environmental variables. Metric scoring was thereby specific to the natural factor in each site.

## Non-metric multidimensional scaling (NMS)

The NMS ordinations were run on presence/absence data from the reference sites, with the dataset limited to 105 common taxa. Taxa that occurred in less than three sites were removed to prevent a bias in the sample similarities. The ordination resulted in a 2-dimensional solution with a final stress of 16.6 (< 20 is acceptable). The first two axes explained 84% of the variance in the data and these axes were explored with correlation analysis and visual inspection.

Level 4 ecoregions were fairly distinct for reference SNEP sites using presence/absence ordinations (Figure E-1). The first NMS axis (horizontal) was related to stream sinuosity, land slope, temperature, watershed size, and stream substrate (Table E-1). Larger, steeper streams with larger substrates in the western SNECPAH ecoregion were on the right of the diagram. These samples had higher numbers of clinger taxa and sensitive and Coleoptera, Odonata, Ephemeroptera, and Trichoptera (COET) taxa. On the left of the diagram were samples soft-bottom, sinuous streams from the eastern NBL ecoregion. These had greater percentages of non-insect individuals and more individuals were dominant in the five most common taxa. The second (vertical) axis was related to wide streams with diverse habitats at the top of the diagram. The opposite end of the axis had sites with higher temperatures in the east. The sinuosity measures traced in the GIS exercises did not confirm the relationship suggested by the NHD sinuosity. The top of the second axis had higher non-insect and clinger taxa and the bottom had higher percent Trichoptera individuals.

On the first axis, sinuosity, longitude, land slope, and substrate characteristics, and percent water and wetland cover in the watershed are the major correlated natural variables that might be useful for site classification (Table E-1). Drainage area was also correlated but might not be appropriate for classification. The correlation with drainage area was driven by three large sites (> 30 km2). In more disturbed non-reference sites, watersheds were up to 700 sq km. If drainage area was used in site classification, the reference condition derived mostly from small sites might represent a natural condition that would not be applicable to large non-reference sites.

Clinger taxa and percent Plecoptera, Odonata, Ephemeroptera, and Trichoptera (POET) taxa were associated with higher land slopes, larger drainage areas, and gravel and pebble substrates. More non-insect individuals and greater dominance of the five most common taxa per sample were associated with the sinuous, warmer, small sites with fine sediments.

Longitude is related to ecoregion and could be used as a continuous variable for classification whereas ecoregion would be categorical. Eastern sites in the lower left of the diagram are mostly in the Narragansett-Bristol Lowlands (NBL, L4 ecoregion 59e), though there is some overlap with sites of the Southern New England Coastal Plains and Hills (SNECPAH, L4 ecoregion 59c) (Figure E-1). The single Gulf of Maine Coastal Lowland (L4 ecoregion 59f) reference site is intermediate to the other ecoregions. Because of the overlap in longitudes without a distinctive break-point or threshold, the categorical ecoregions would be better classification variables than longitude.

Sinuosity was on the same axis as land slope and drainage area (Figure E-2). These three variables are often related, as large catchments are generally in flatter valleys with low slopes and meandering streams. In this data set, the three variables suggest that the flatter streams were sometimes more sinuous than steeper streams, but that these flatter, sinuous streams were in small catchments. The sinuosity measures traced in the GIS exercises did not confirm the relationship suggested by the NHD sinuosity.

Land slopes were steeper in the upper right of the diagram. Only two NBL sites had slopes similar to those in the SNECPAH sites (Figure E-3). Correlated substrate characteristics included the percent muck-mud in the reach and the percent gravel and cobble (estimated substrate areal percent). Though these had a higher correlation with the axis, there was no apparent threshold that could be used to classify biological types.

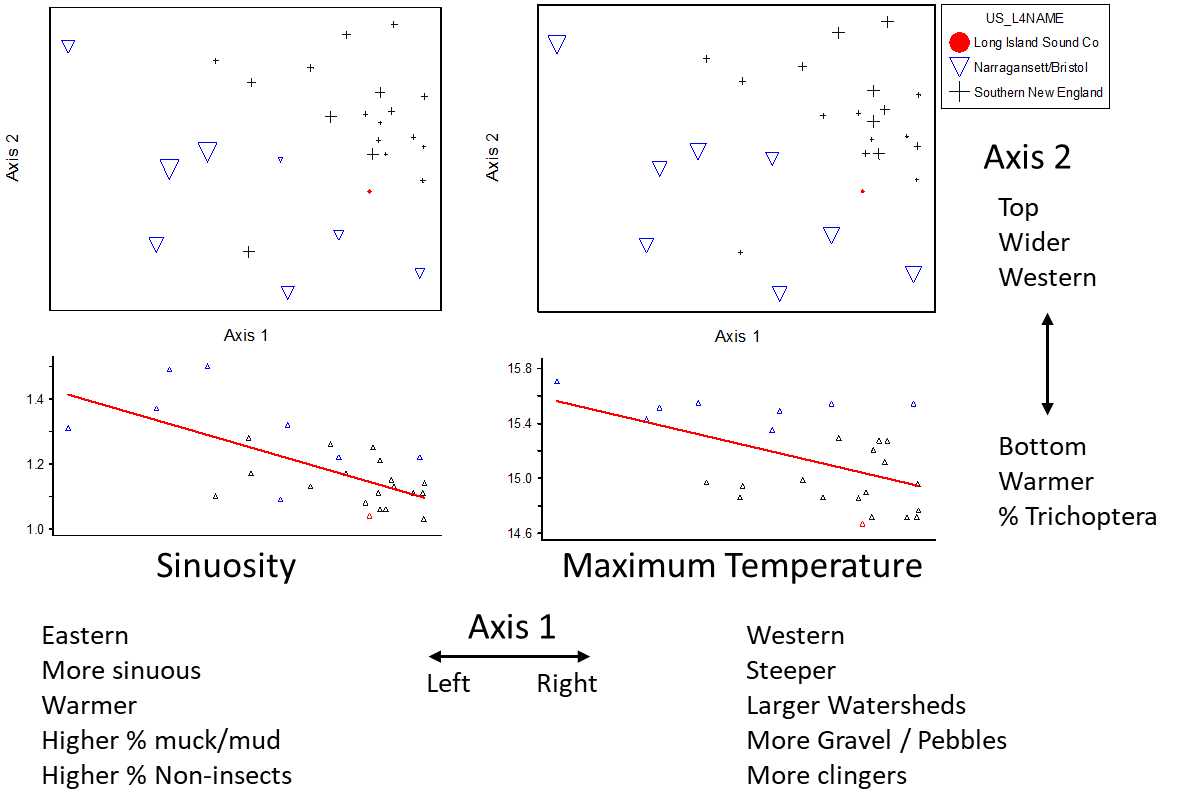
The second (vertical) axis was related to temperature, stream width, and longitude. The top of the diagram had warmer, wider, western sites, which were also related to the SNECPAH ecoregion. Temperature could be related to urbanization and clearing of riparian canopy and that could lead to misclassification of new sites. The top of the diagram also had greater richness of non-insect taxa, more clinger taxa, and fewer Trichoptera individuals.



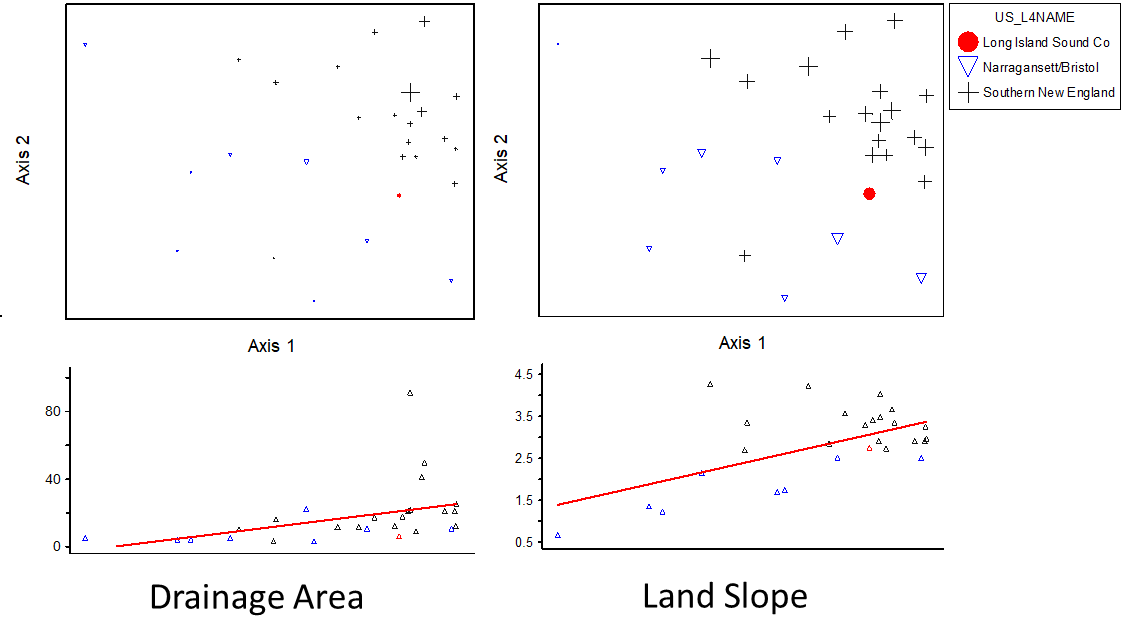
Figure E-1. Non-metric multidimensional scaling (NMS) ordination of taxa presence/absence in SNEP reference sites, with samples coded by Level 4 ecoregion. Samples with similar taxonomic composition are plotted in close proximity.

Table E-1. Correlation coefficients of the major environmental variables and biological metrics related to the non-metric multidimensional scaling (NMS) ordination axes.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Axis 1  (60% variance) | **Variables** | **r** | Axis 2  (23% variance) | **Variables** | **r** |
| Environmental | | Environmental | |
| Sinuosity | -0.66 | # Habitat Types | 0.66 |
| Land Slope | 0.59 | Minimum Temperature | -0.63 |
| Longitude | -0.54 | Stream Width | 0.61 |
| Maximum Temperature | -0.51 | Longitude | -0.54 |
| % Wetland & Water | -0.49 | Biological | |
| % Muck and Mud | -0.44 | % Trichoptera Taxa | -0.68 |
| Drainage Area | 0.41 | Non-insect Taxa | 0.66 |
| % Gravel & Pebble | 0.40 | Clinger Taxa | 0.62 |
| Biological | |  |  |  |
| Clinger Taxa | 0.81 |  |  |  |
| % COET Taxa | 0.81 |  |  |  |
| % Dominant 5 Taxa | -0.76 |  |  |  |
| % Non-insect Individuals | -0.75 |  |  |  |



*Figure E-2. NMS ordination of taxa presence/absence data in SNEP reference sites with samples coded by Level 4 ecoregion. The lower plots show how sinuosity and maximum water temperature relate to Axis 1.*



*Figure E-3. NMS ordination diagram of taxa presence/absence data in SNEP reference sites with samples coded by Level 4 ecoregion. The lower plots show how drainage area and land slope relate to Axis 1.*

## Principle components analysis (PCA)

To explore the effects of environmental variables on metric distributions, a PCA was performed with 45 metrics that represented a variety of metric formulations and taxa characteristics. The PCA identified the same variables on the first axis as were identified in the NMS of taxa presence absence, though in a slightly different order of importance. These included sinuosity, land slope, percent water and wetland cover in the watershed, longitude, and drainage area. Substrate characteristics were also correlated, though not as strongly. The first axis explained 39% of variance in the ordination. On the second axis, less variance (13%) was explained by the variables stream width, longitude, and temperature.

Richness metrics were associated strongly with the first axis, including clinger taxa, Coleoptera, Odonata, Ephemeroptera, and Trichoptera (COET) taxa, and insect taxa. These were on the end of the axis with steeper slopes. Patterns of the PCA ordination were similar to those seen in the NMS ordination, though the separation of ecoregions was less distinct (Figure E-4).



Figure E-4. Principle Components Analysis (PCA) ordination of 45 metrics in reference sites of the SNEP region, with larger marker size indicating eastern longitude of the sites

## Classification Summary

Classification schemes related to Level 4 ecoregions and drainage area were considered but ruled out based on results from the NMS and PCA analyses. Level 4 ecoregion did not cluster distinctly in the PCA ordination of metrics. Moreover, defining site classes based on Level 4 ecoregions might be untenable because it would result in small sample sizes for index calibration. All the reference sites in the NBL were <15 km2, which is smaller than the bulk of stressed sites, suggesting that a classification scheme based on drainage area or ecoregion would result in insufficient comparable samples for index calibration.

Continuous variables that showed potential for classification included: annual air temperature (PRISM 1981-2010), sinuosity, longitude, land slope, substrate types, and drainage area. Because there are no clear break-points to distinguish classes based on the continuous variables, scores for individual metrics that showed strong correlations with these natural variables were adjusted during index development (see Section 5.1).

# Additional Maps

Baseflow index - https://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml

Map

Description automatically generated

Mean annual air temperature (PRISM 1981-2010) - <https://prism.oregonstate.edu/normals/>

Map

Description automatically generated

Mean summer stream temperature (July-August) –

Hill, R.A., C.P. Hawkins, and D.M. Carlisle. 2013. Predicting thermal reference conditions for USA streams and rivers. Freshwater Science 32(1):39-55. doi:10.1899/12-009.1.

Map

Description automatically generated

Elevation

Map

Description automatically generated