USDA-CSREES 2007 National Water Quality Conference

Analyzing Stream Habitat Integrity by Self-Organizing Mapping

Self Organizing-Mapping (SOM) are used to find similarity clusters and linear and non/linear correlations in data by comparing Euclidean Distances existing among the different environmental input vectors. Habitat metrics used to calculate habitat quality indices in Ohio and Maryland were clustered using SOM. Subsequently, with the clusters obtained, the distributions of the final physical habitat indices as well as the fish and macroinvertebrate biological quality indices were represented. Three very prominent clusters were identified using the habitat metrics. The representation of the distribution of the habitat indices showed that the clusters could be identified as good, fair and poor. Moreover, the biological quality indices distribution matched almost perfectly the habitat index distribution and the impact of habitat impairment on stream fauna was very distinct. Stream's immediate land-use distribution showed that the worst habitat scores were associated primarily with agricultural land uses and exhibited poor biological integrity. In contrast, urban land-uses did not have low habitat scores but had the worst biological scores, suggesting that water quality stresses are more important than habitat. Maryland uses different habitat metrics expressed in the Provisional Physical habitat Index (PHI) and the Final PHI. The SOM were run using both types of metrics and for each of the different strata considered in each type of index. In coastal and piedmont sites, the clusters showed a poor separation of habitat index scores and much overlapping among the clusters. However, the biological quality indices distribution among the same clusters showed a very good separation in all biological communities. This suggested that the habitat quality indices used in Maryland are not sensitive enough to changes in habitat that seriously affect biological integrity. The clustering was significantly improved by selecting the four most discriminating metrics and ignoring those that apparently represented noise. The SOM are a promising tool for data analysis and selection for predictive models. Regression equations for biological integrity prediction are being developed, using the results from the SOM. Supervised Artificial Neural Nets and advanced statistical models such as the Principal Component and the Canonical Correspondence Analyses are being used.

Author: David Bedoya
University Affliation: Northeastern University
Co-Author(s): Vladimir Novotny and Elias Manolakos