Presentation at the US EPA Global Change Workshop

November 3, 2005 Arlington, VA

Linking Pollution to Water Body Integrity - Second Year of Research

Vladimir Novotny

Elias Manolakos

Northeastern University

STAR WATERSHED PROJECT FUNDED BY USEPA 2003-2007
Development of Risk Propagation Model for Estimating Ecological Response of Streams to Anthropogenic Stresses and Stream Modification

- Project Team
  - PI - Vladimir Novotny, NEU Center for Urban Environmental Studies
  - Co-PI's NEU CUER
    - Elias Manolakos
    - Ramanitharan Kandiah
  - Co-PI Univ. of Wisconsin Milwaukee
    - Timothy Ehlinger
  - Co-PI Illinois State Water Survey
    - Alena Bartosova
  - 5 graduate students

Project objectives

- A model that will link stresses to biotic end points
    - Pollutant inputs
    - Watershed and water body modification
        - Land use changes
        - Chanelization and impoundments
        - Riparian corridor modifications
        - Development of a quantitative layered risk propagation from basic landscape and watershed stressors to the biotic IBI endpoints
    - Study the possibility of mitigating the stresses that would have the most beneficial impact on the biotic endpoints
- Apply the model to another geographic region

NUMERIC INDICES OF BIOTIC INTEGRITY (ENDPOINTS)

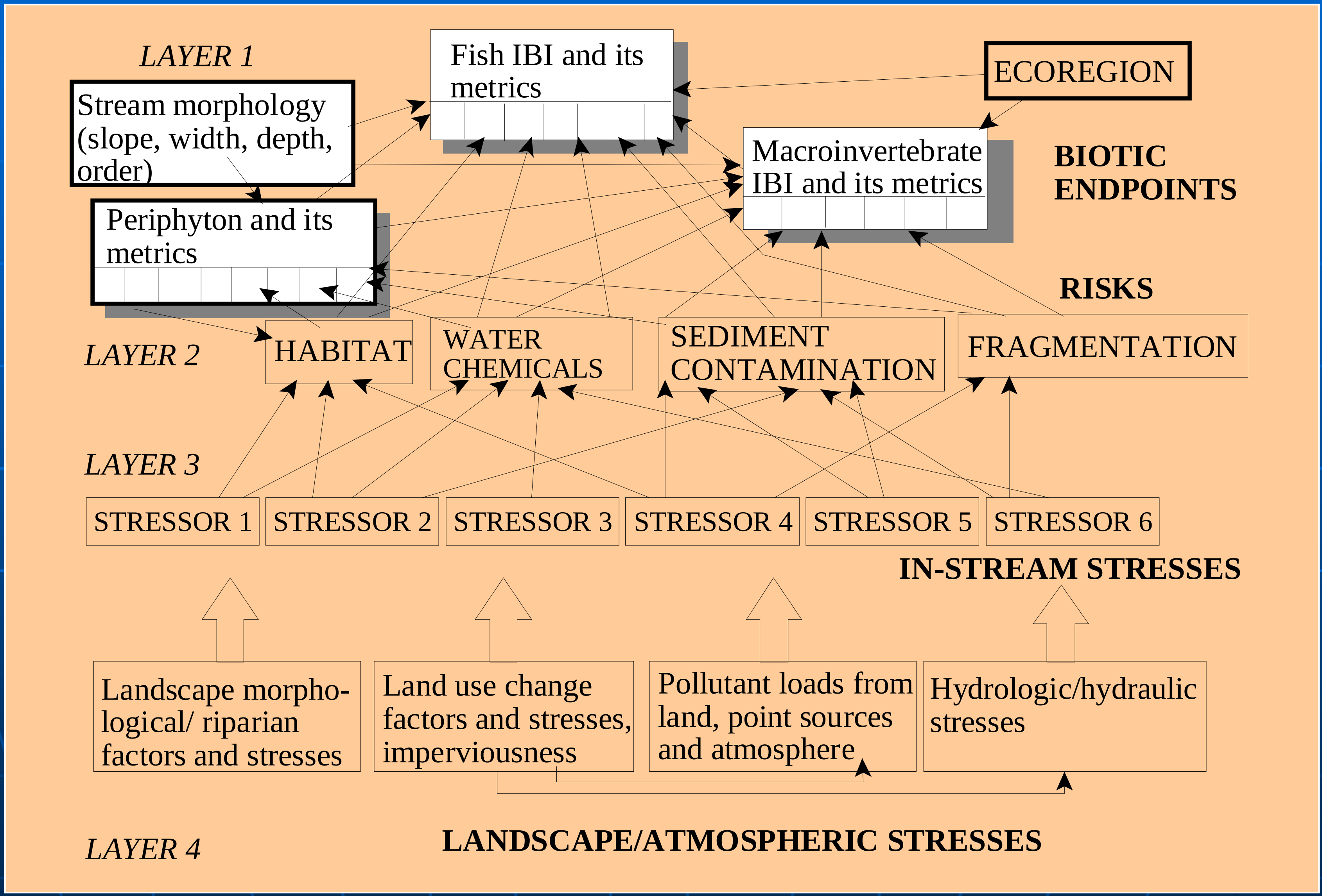
- Fish
- Benthic macroinvertebrates
- Physical - Habitat

INPUTS INTO THE MODELS

About 50 habitat, chemical, land use and riparian quality parameters

(Ohio and Maryland)

LINKING IBIs AND STRESSORS - STRUCTURAL AND FUNCTIONAL MAZE



RESEARCH ACCOMPLISHMENTS First and Second Year

- First year
  - Teams were formed
  - Data base shell development (USGS vs. our own)
  - Technical review reports
  - Several publications
- Second Year
  - Data base assembled (Maryland, Ohio, Minnesota)
  - Five Technical Reports completed
  - ANN Unsupervised Model combined with Canonical Correspondence Analysis model completed and ready for (β – testing)
  - Three publications submitted and four presentations were made (IWA Kyoto, Great Lakes Research Association -Ann Arbor, AWRA Conference Seattle)

DATA BASE SHELL

STAR Environmental Data Base STARED

Alena Bartosova, Ramanitharan Kandiah and Dweight Osmond

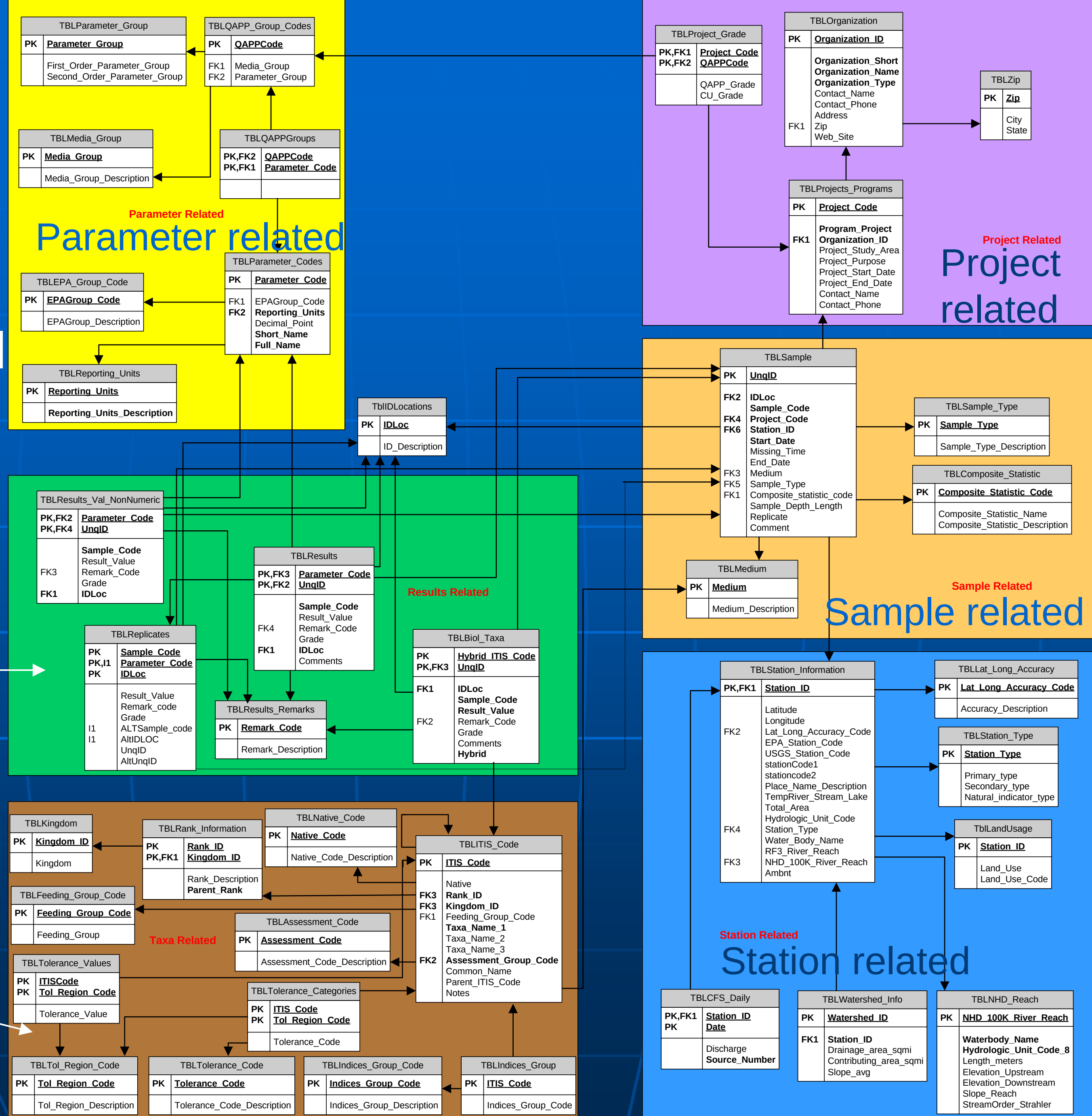
- Developed from FoxDB
- Significant modifications were made. The data base includes
  - environmental data, including water quality,
  - sediment chemistry,
  - biological indices,
  - stream hydrology,
  - and habitat
- Master database is stored in Microsoft SQL Server 2000 format on a server connected to a network through the Northeastern University system
- Data base currently includes complete data from Ohio, Maryland and parrtial data from Minnesota, Wisconsin, Illinois and Massachusetts

STAR

Environmental

Database

Structure



Results related →

Technical Report # 5

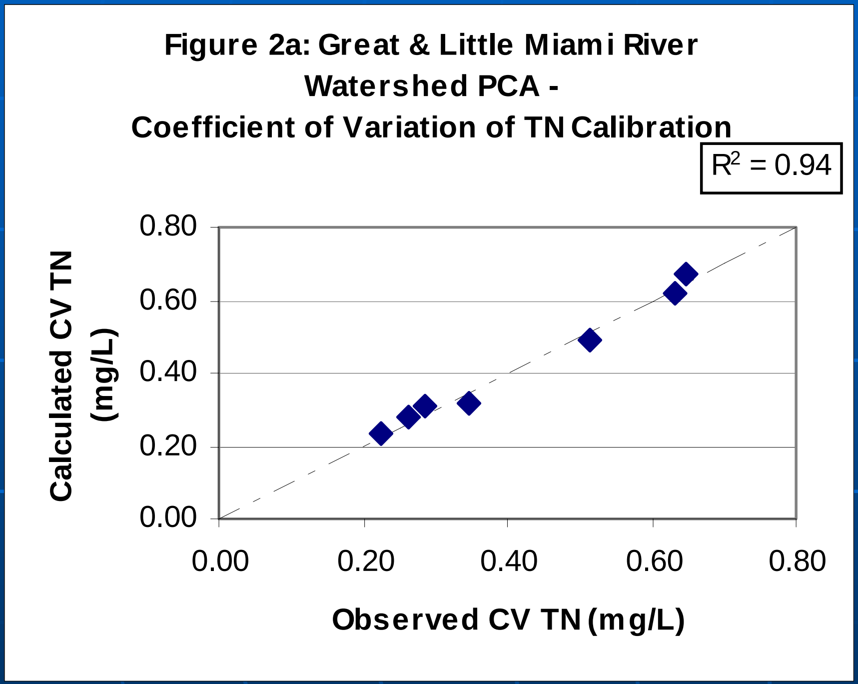
Taxa Related →

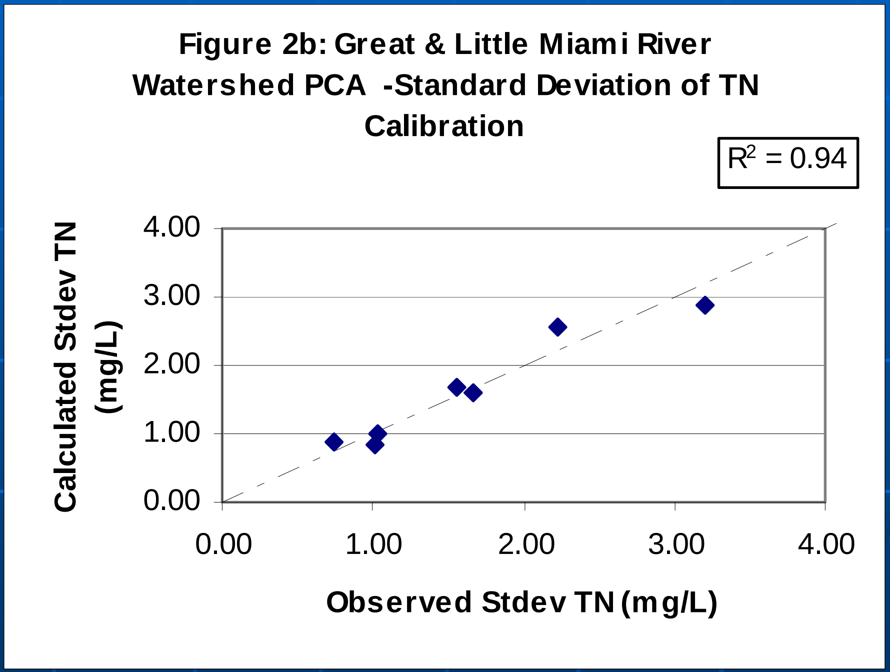
GIS – based Water Quality Model for Nitrogen by PCA

David Beach and V. Novotny

- SPARROW and similar models provide mean annual or seasonal concentrations
- The model developed by the team predicts both means and logarithmic standard deviation. Extreme (e.g., 99 percentile) concentrations can be estimated

The model was developed for the Fox River (Illinois) and the Miami River (Ohio)



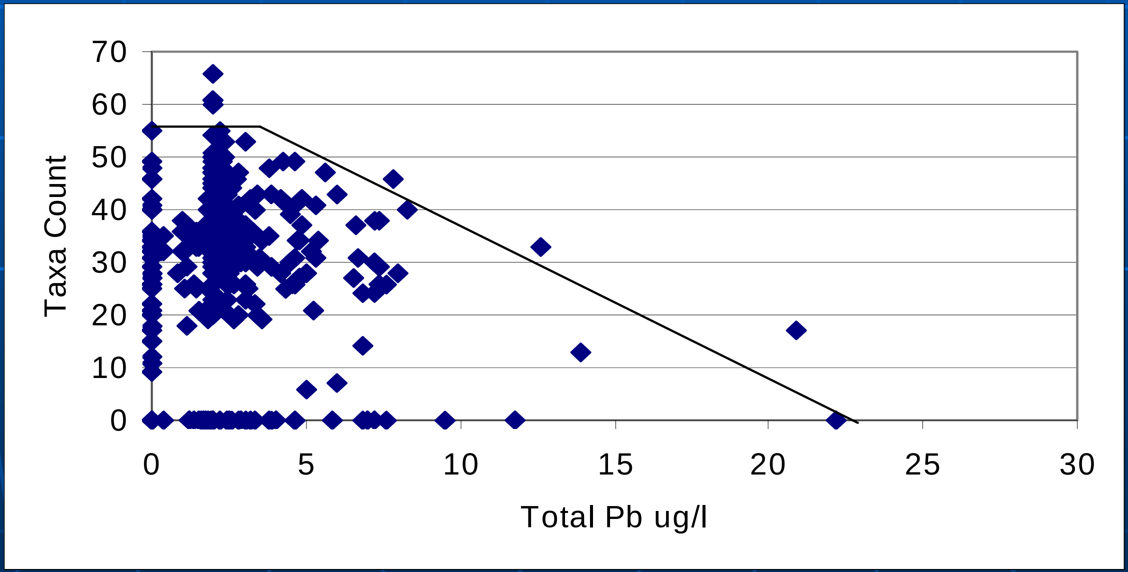


Univariate Functions Linking Stressors to Impairment of Metrics (Species Abundance)

Jessica Brooks and V. Novotny

Maximum Specie Richness (MSR) functions were developed from field data (Ohio) and US EPA toxicity bioassay data.

MSR can be related to risk of species disappearance



Self Organizing Feature Maps combined with Ecological Ordination Techniques for Effective Watershed Management

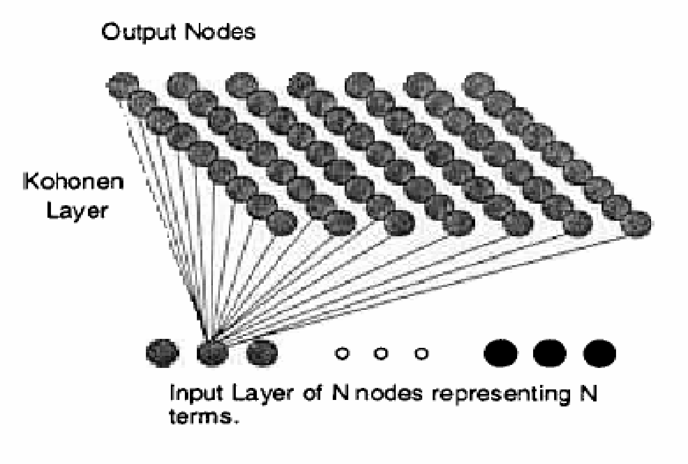
Hardik Virani Elias S. Manolakos

Data base for Ohio provided by Ed Rankin
Midwest Biodiversity Institute, Columbus, OH

Modeling Approaches

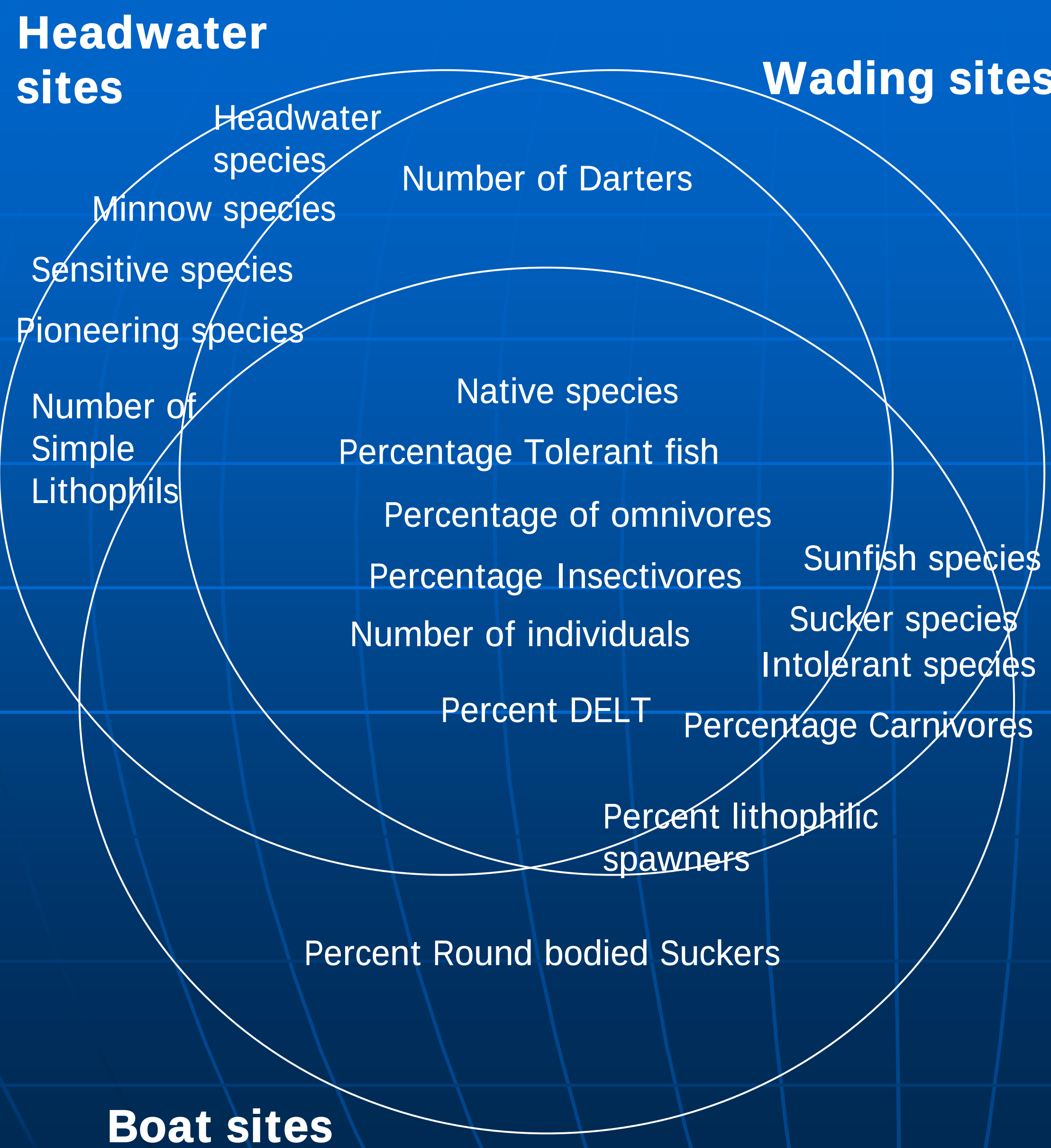
- Artificial Neural Networks
  - Learning
    - *Supervised: Output tracking*
    - *Unsupervised: Pattern detection*
  - Neural network applications to ecological modeling or, more general, to ecology are quite recent (early 90s)
  - Initial applications revolved around back propagation algorithm to predict the outcome from a set of environmental variables, with emphasis on comparison with regression techniques.
  - References: predicting brown trout density (Lek et al., 1996), predicting rainfall runoff (Anctil and Tape, 2004), modeling trihalomethane residuals in water treatment (Lewin et al., 2004)

Self Organizing Feature Map (SOM)



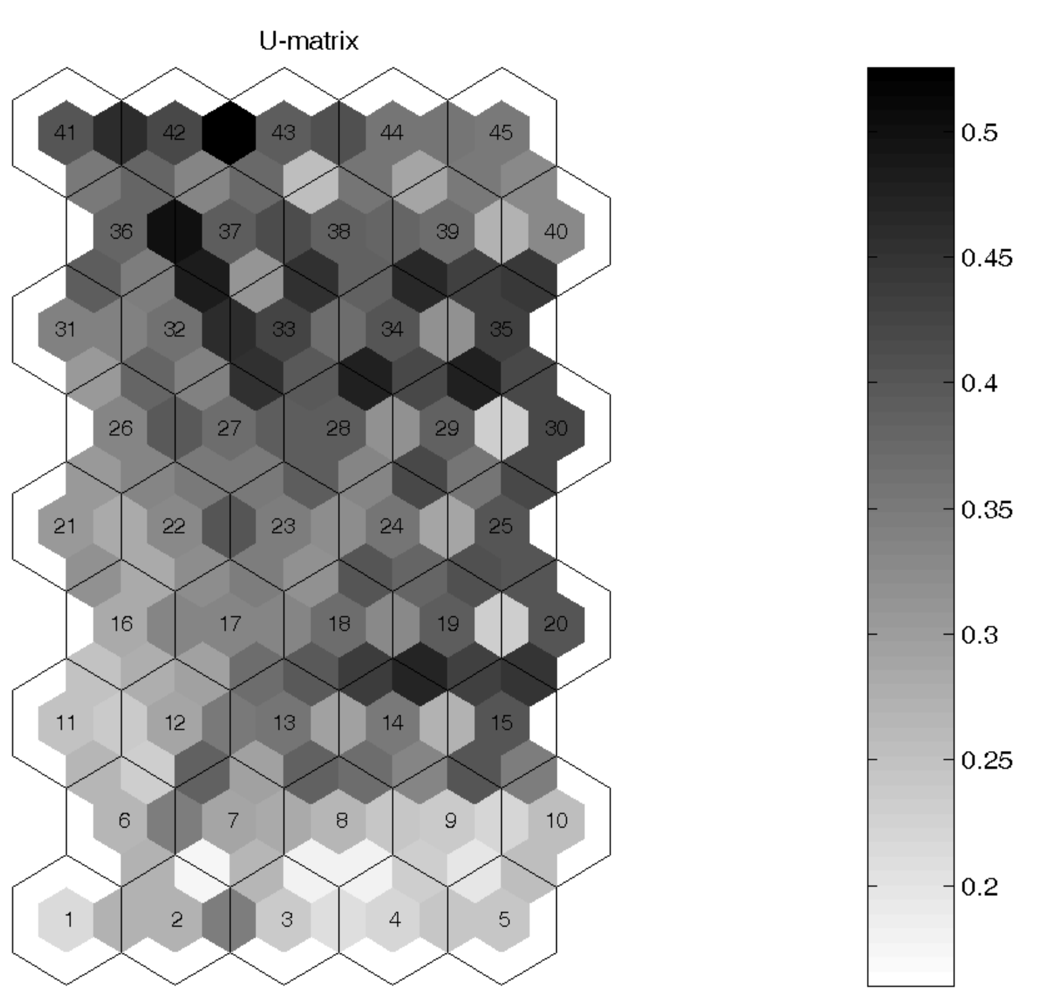
- Unsupervised artificial neural network (ANN) based on competitive and cooperative learning.
- Mapping high-dimensional data into a two-dimensional representation space.
- SOM is an approximation to the probability density function of the input data
- Used for data visualization, clustering (or classification), estimation etc.

Ohio EPA



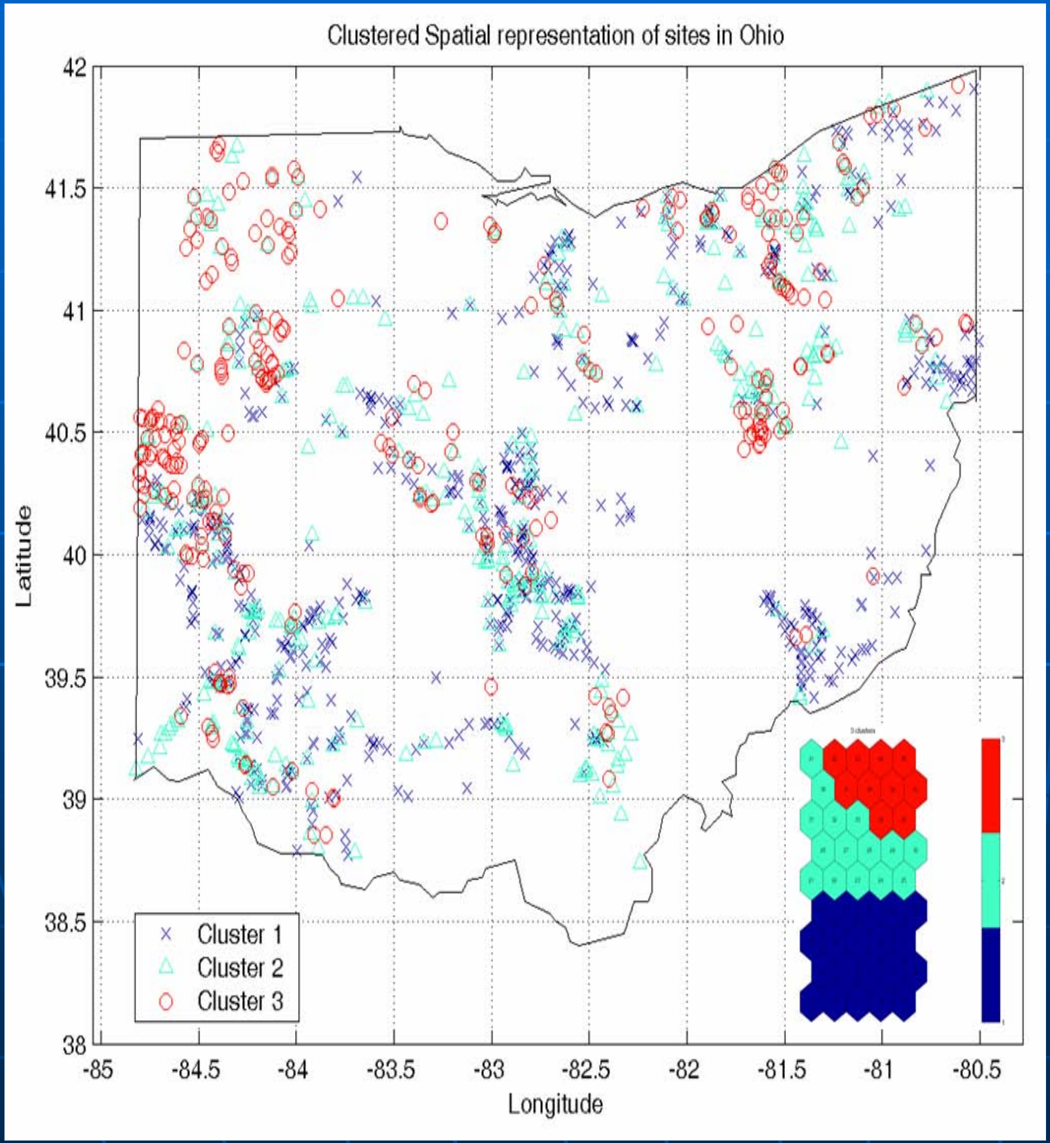
- 1848 sampling sites between 1995 and 2000.
- Sampling period:
  - *Summer (July-September)* Fish, Physical Habitat, and in-situ Water Chemistry.
- Metrics based on the type of sites.

SOM U Matrix - OH



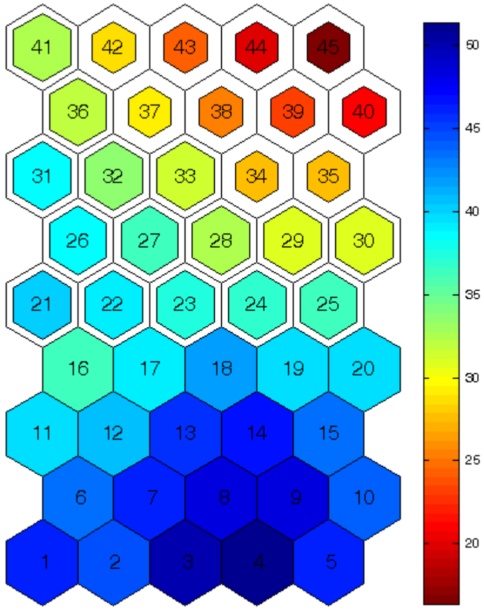
- U-matrix gives the distances between neighboring map units.
- Element for each neuron is either the min, max, median (default) or mean of the surrounding values.
- Dark color = large distance and thus a gap between the codebook values in the input space.
- Good for visualization.
- Light color = codebook vectors are close to each other in the input space.
- Light areas can be thought as clusters and dark areas as cluster separators.

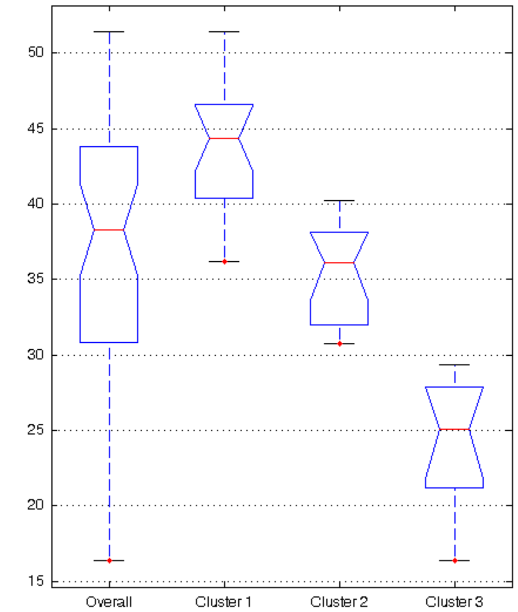
Spatial representation of the clustered data in Ohio



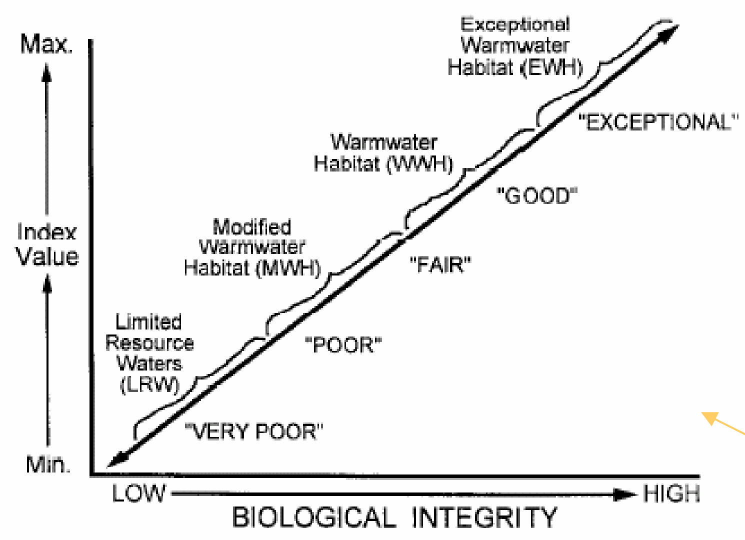
- 3 clusters (good in Cluster 1, intermediate in Cluster 2, and marginal in Cluster 3)
- Clusters scattered all over the Ohio state.
- The northwestern and Cleveland /Akron regions are dominated by Cluster 3.

Fish IBI - OH



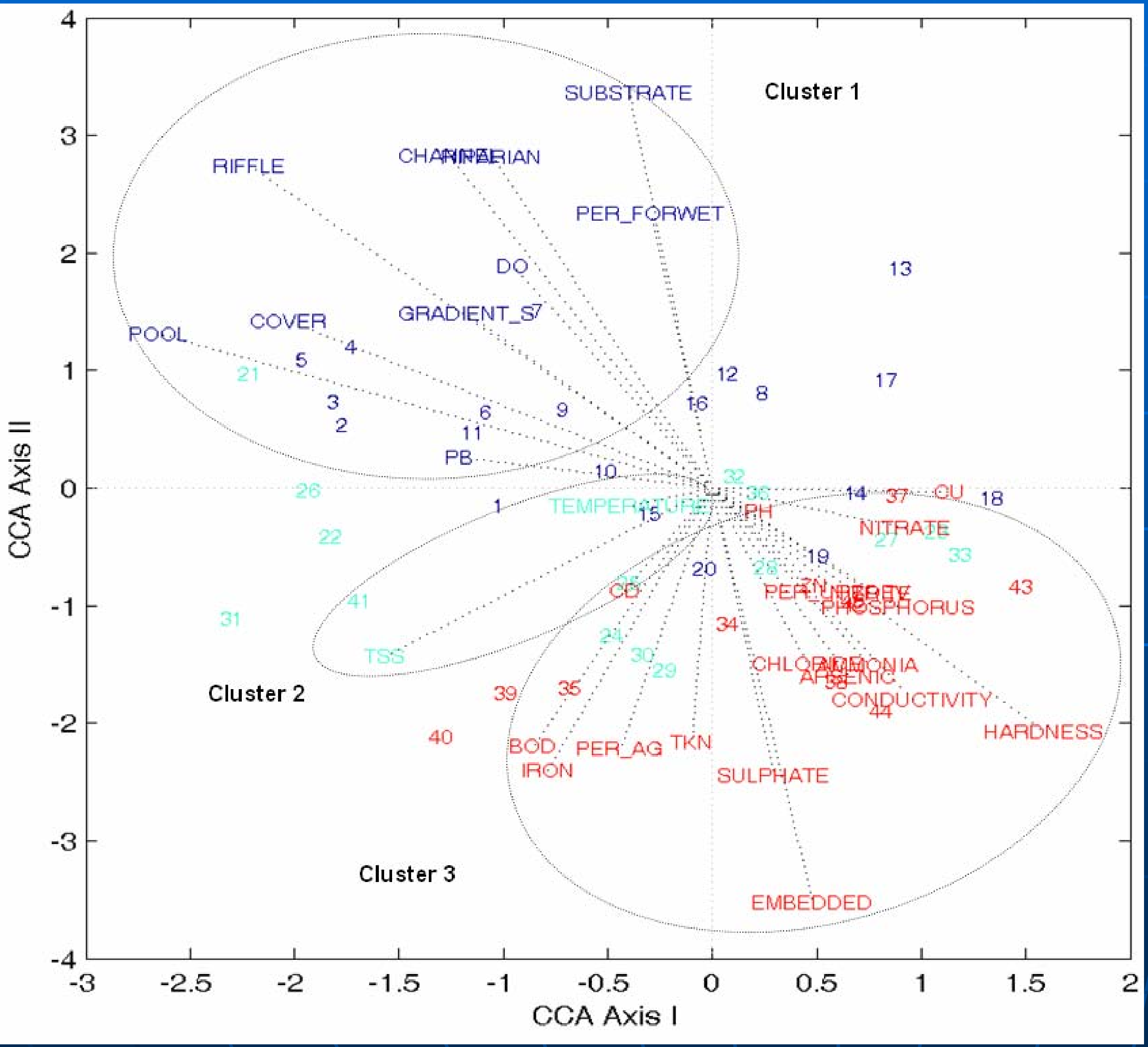


- IBI is patterned on the SOM
- 3 distinct clusters in the fish IBI
- Distinct Habitat use ranges by Ohio EPA
  - EWH(> 48): Cluster 1
  - WWH (32-44): Clusters 1 & 2
  - MWH(22-30): Cluster 3



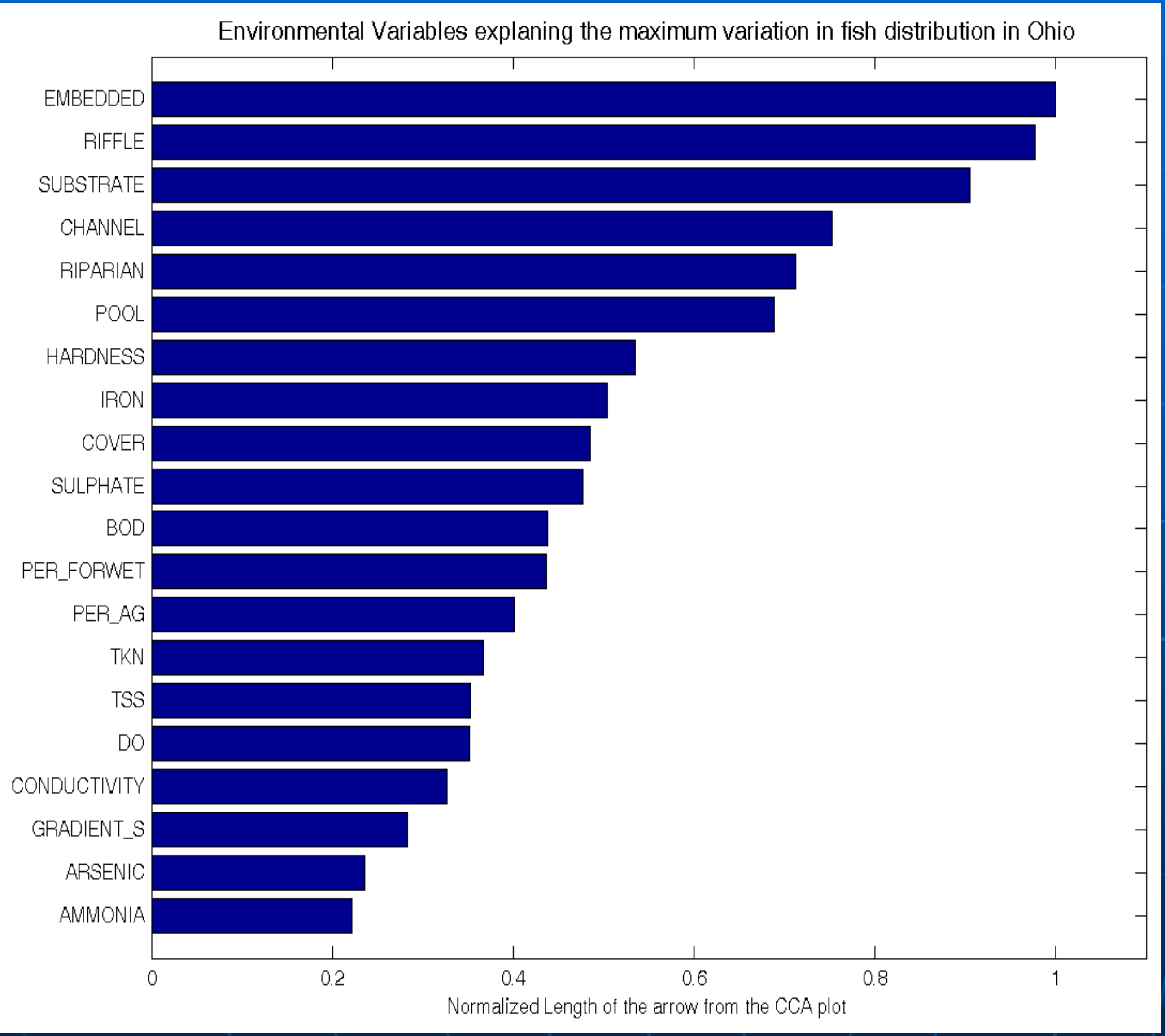
Ohio ranking of wadeable waters

Canonical Correspondence Analysis



- First 2 axes explain 28% and 19%
- Direction of arrow indicates inter-correlations between variables.
- Neuron Projections on arrow indicates relative associations.
- Per-cluster median projection to assign cluster labels to environmental variables.
- Effective segregation of the environmental gradients in Cluster 1 and 3.

Relative Rankings of environmental variables



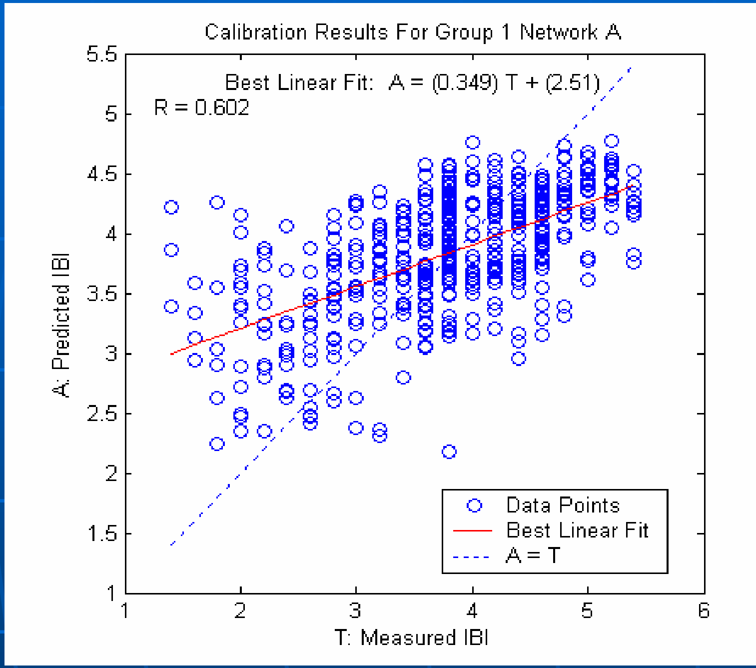
- Length of the environmental variables in CCA indicates their importance in explaining the variation in species distribution
- Habitat parameters dominate the top 10.

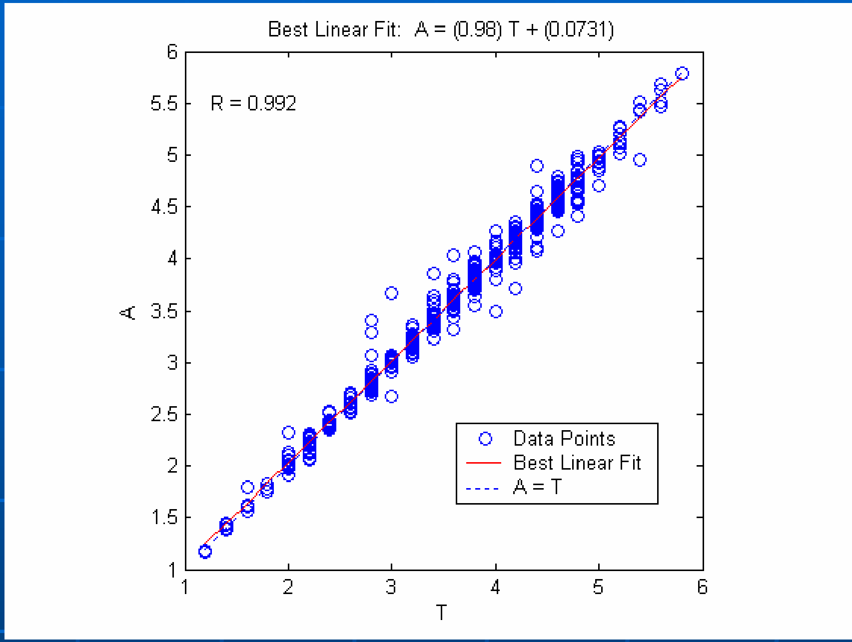
Supervised learning by ANN Multiple stressors relationship to IBI

Jessica Brooks, V. Novotny, H. Virani, E. Manolakos

- ANN were used in supervised learning.
- Data from Ohio data base were used
- First attempts using intuitively selected chemical parameters, habitat index, and macroinvertebrates index as inputs and IBIs as output were not successful
- SOM and Canonical Correspondence Analysis identified the most important variables. Using top 12 stressor variables the ANN performance improved dramatically

Training Results (Calibration 75% of data)





TEMPERATURE DISSOLVED OXYGEN
PH TOTAL SUSPENDED SOLIDS
TOTAL AMMONIA BOD, 5-DAY
HARDNESS CHLORIDE SULFATE
TOTAL CADMIUM TOTAL LEAD TOTAL ZINC
INVERTEBRATE COMMUNITY INDEX
QUALITATIVE HABITAT EVALUATION INDEX

INTUITION INPUTS

RIFFLE POOL EMBEDEDDNESS
CHANNEL SUBSTRATE RIPARIAN
PERCENT FORRESTED HARDNESS
TOTAL IRON BOD, 5-DAY,
DISSOLVED OXYGEN
INVERTEBRATE COMMUNITY INDEX

TOP 12 INPUTS AFTER SOM AND CCA

Discussion

- The results indicate the efficiency of the SOM in visualizing the state of streams in Maryland and Ohio and aid the watershed manager in making and implementing decisions which will ultimately provide knowledge necessary for restoration of the degraded watersheds.
- Habitat parameters surpassed chemical parameters as important variables related to species compositions, which in turn decides the fish IBI.
- Embeddedness was found to be an important variable in both the datasets; hence work on improving the watershed should revolve around analyzing the effects of embeddedness on the species composition.
- River mouths were the most degraded regions in both the states, which supports the general hypothesis about regions with poor fish IBI. In the case of Maryland, we found that the Coastal areas around the Chesapeake Bay are the most degraded regions. The Lake Erie region and the northwestern region in Ohio were found to have poor fish IBI compared to the rest of the state.

Summary

- SOM analysis provides an ordered set of sites across the state with similar metric characteristics, ultimately reflecting in the fish IBI.
- Clustering analysis allowed us to divide the examination of the entire state into 3 clusters, which helps us to compare the clusters and define their key ecological characteristics.
- CCA help in drawing conclusions about specific fish species and the role of the environmental variables in maintaining the perfect habitat and water quality for fishes.
- The results provide a comprehensive state-based study for future monitoring that can address short-term and long-term trends.
- Different visualization approaches, based on a detailed analysis give the watershed manager ample scope for working out new strategies, focused on improving the overall heath of the streams today.

Future Work (This and Future Projects)

- Prediction networks (supervised ANNs) to predict the fish (an macroinvertebrate) IBI.
- Development of hydrologically and morphologically based multidimensional habitat index
- Finalize effects of impoundments and other fragmentation
- Possible to integrate ANNs with other modeling approaches (genetic algorithms), or tools (GIS), or mechanistic models (Hybrid models)
- Specialized networks dealing with individual clusters to obtain relationships associated with similar metric traits.
- Apply the models to Northeast (the Charles River)
- Integrated tool with guidance manual to cater to a wide section of the audience snf s workshop for users.
- Consider sediment contamination as a stressor.
- Work still to be done to realize the final hierarchical risk based model

Thank you!!