Ecological Clustering of Integrity and Nonlinear Impact of Environmental Variables on Cyanobacteria

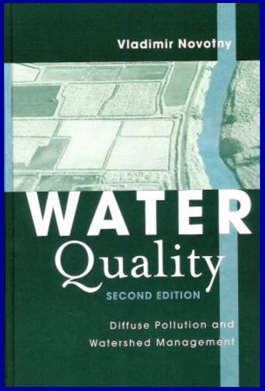
Vladimir Novotny, Elias Manolakos and Ferdinand Hellweger

Northeastern University

Expanded Keynote Presentation at the 5th International Conference on Reservoir Limnology and Water Quality

Brno, August 2006

Ch. 12

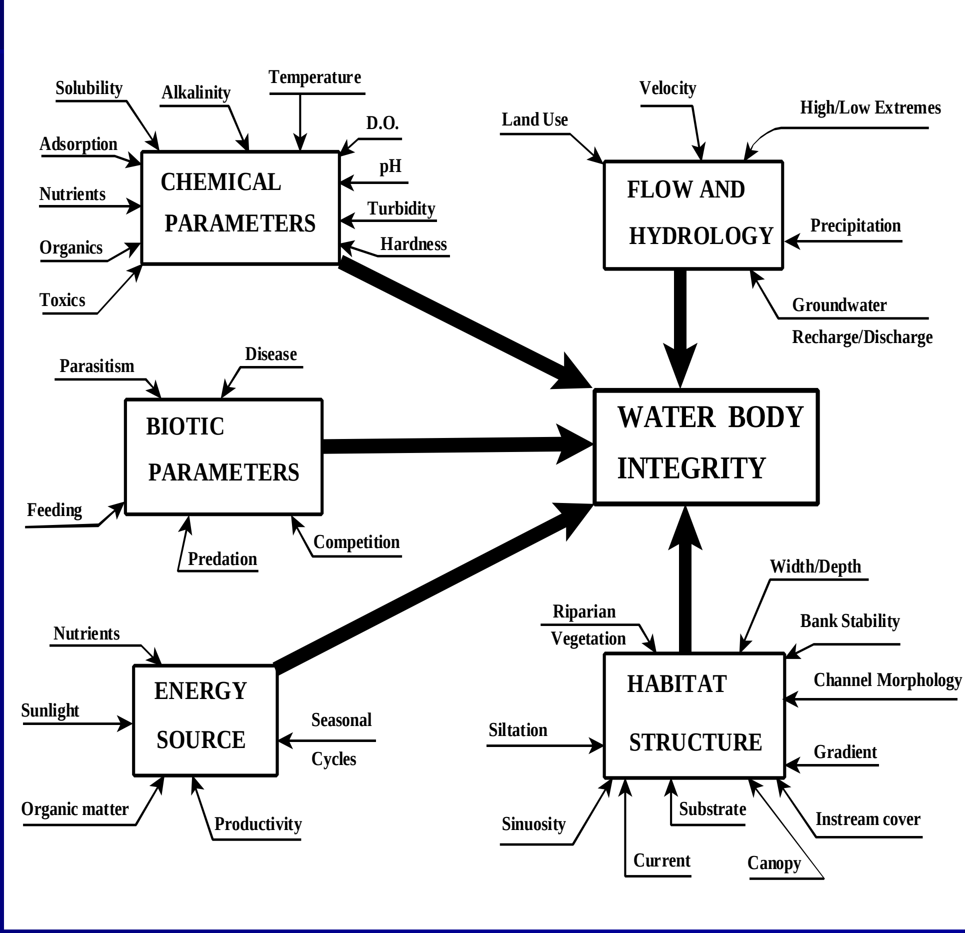


Outline

i. Introduction of measures of biotic integrity
ii. Application to a highly stressed river
iii. Regional analyses of clustering of biotic integrity – Self organizing maps (SOM)
iv. Predictive models of biotic integrity
v. Application to reservoirs and blue greens
vi. New models for blue green algae (cyanobacteria)

ECOLOGIC POTENTIAL WATERBODY INTEGRITY

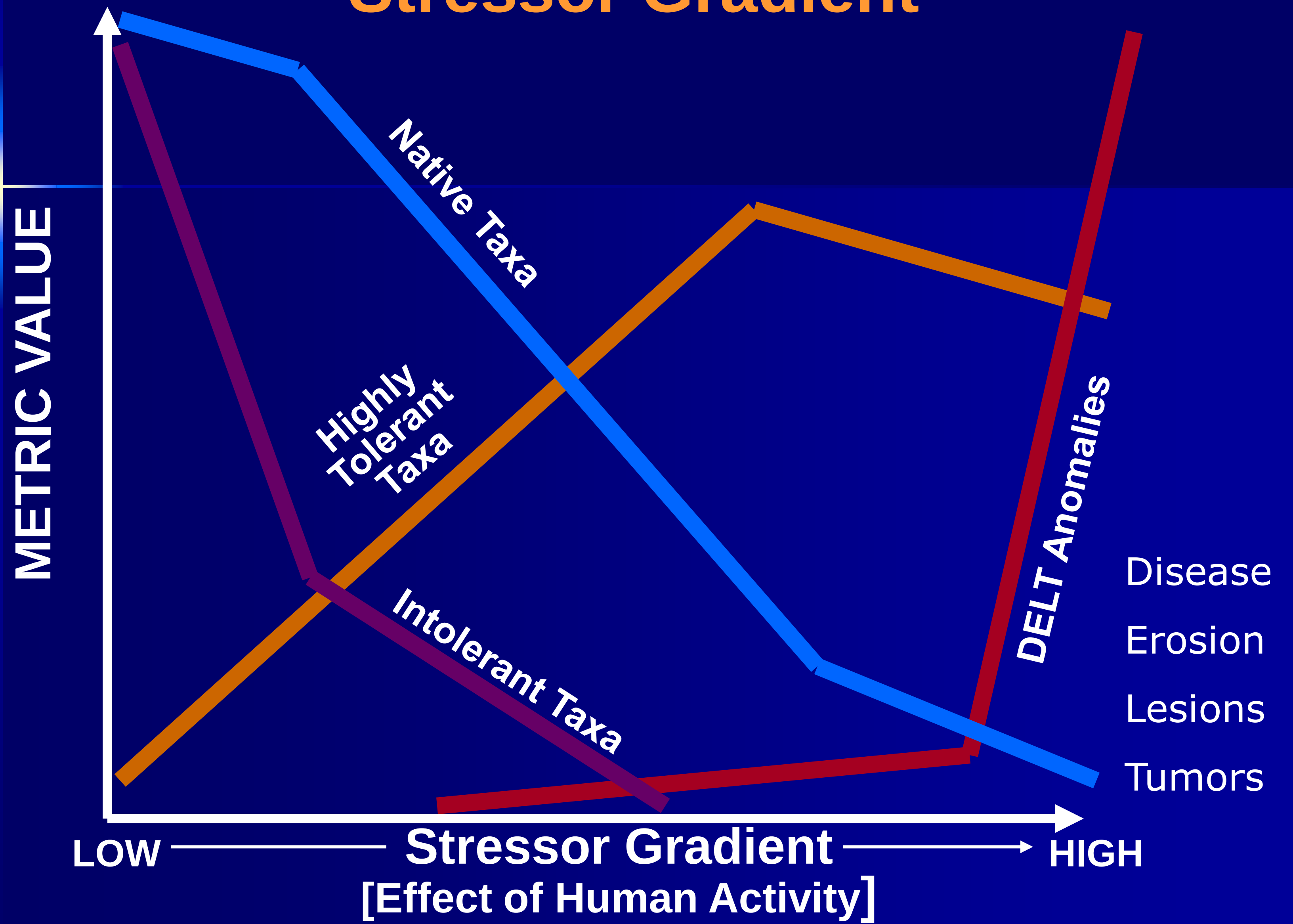
- Maintaining and improving water body integrity is the goal of the Clean Water Act
  - Habitat (physical)
  - Chemical
  - Biological
- Integrity: “A balanced, adaptive community of organisms having a species composition and diversity comparable to or approaching that of natural biota of the region” (Karr et al.)



Measures of Water body Integrity

- Physical — Habitat index
- Chemical — Water Quality Standards
- Biological — Indices of Biotic Integrity
  - Fish
  - Macroinvertebrates
    - ICI
    - Hilsenhoff/, Saprobien
- Periphyton (?)

IBI Metric Behavior Along the Stressor Gradient



Index of Biotic Integrity (Karr 1981)

12 Metrics

- Species richness
- #Darter species
- #Sunfish species
- #Sucker species
- %Intolerant species
- %Green sunfish
- %Omnivores
- %Insectivores
- %Top Carnivores
- %Hybrids
- %Diseased individuals
- Number of Fish

Community Composition

Environmental Tolerance

Community Function

Community Condition

- 5,3,1 metric scoring categories.
- 12 to 60 scoring range.
- Calibrated on a regional basis.
- Scoring adjust-ments needed for very low numbers.

Invertebrate Community Index (Ohio EPA 1987; DeShon 1995)

- Taxa Richness
- #Mayfly taxa
- #Caddisfly taxa
- #Dipteran taxa
- %Mayflies
- %Caddisflies
- %Tanytarsini Midges
- %Other Diptera/Non-Insects
- %Tolerant taxa
- Qualitative EPT taxa

- 6,4,2,0 metric scoring categories.
- 0 to 60 scoring range.
- Calibrated on regional basis.
- Scoring adjustments needed for very low numbers of specific taxa.

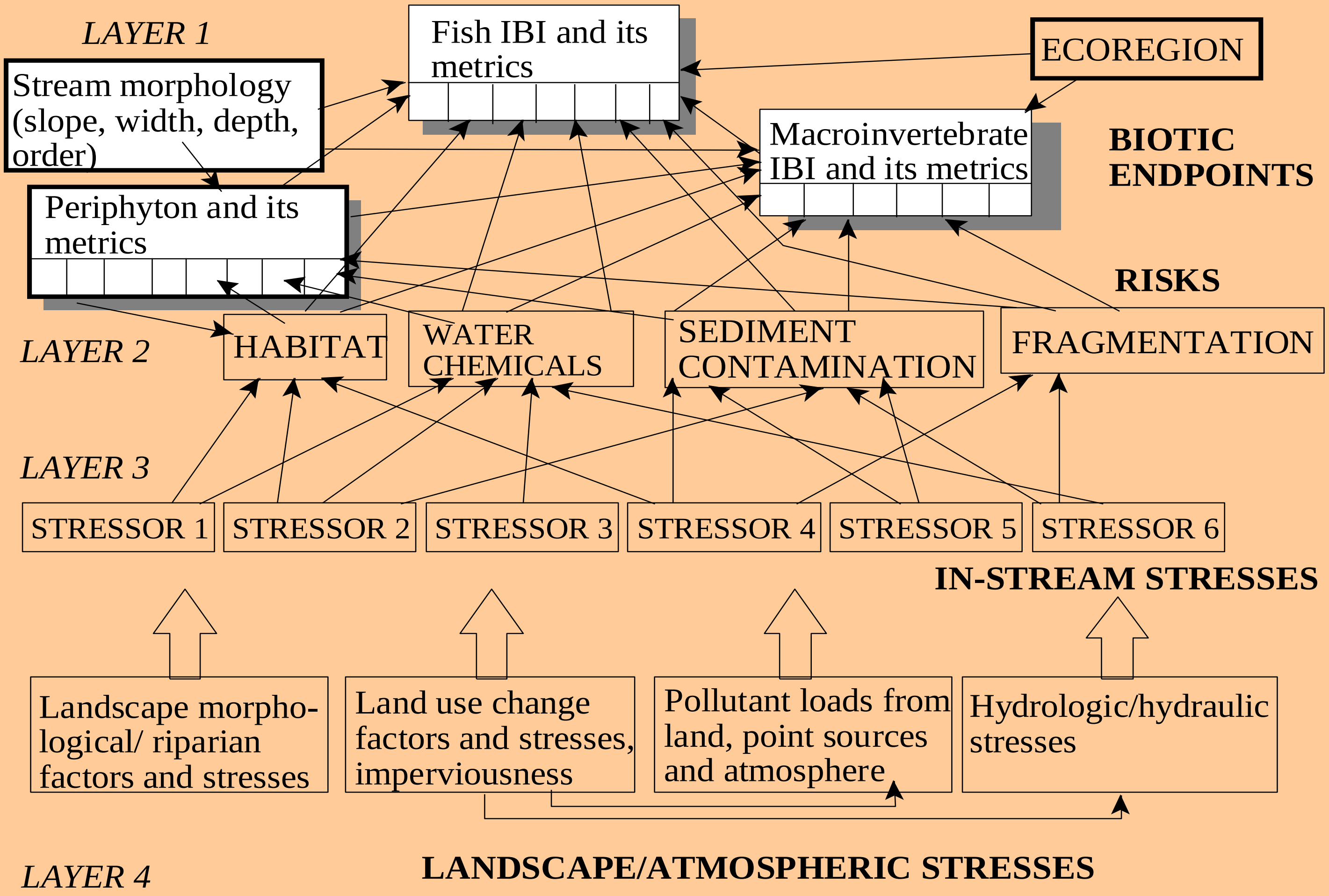
The Qualitative Habitat Evaluation Index (QHEI)

QHEI Includes Six Major Categories of Macrohabitat

- Substrate - types, origin, quality, embeddedness
- Instream Cover - types and amounts
- Channel Quality - sinuosity, development, stability
- Riparian/Bank Stability - width, quality, bank erosion
- Pool/Riffle/Run - max. depth, current types, morphology, substrate embeddedness
- Gradient - local gradient (varies by drainage area) Channelization

Source: The Qualitative Habitat Evaluation Index (Rankin 1989)

HIERARCHICAL MODEL CONCEPT OF RISK PROPAGATION FROM STRESSORS TO BIOTIC ENDPOINTS

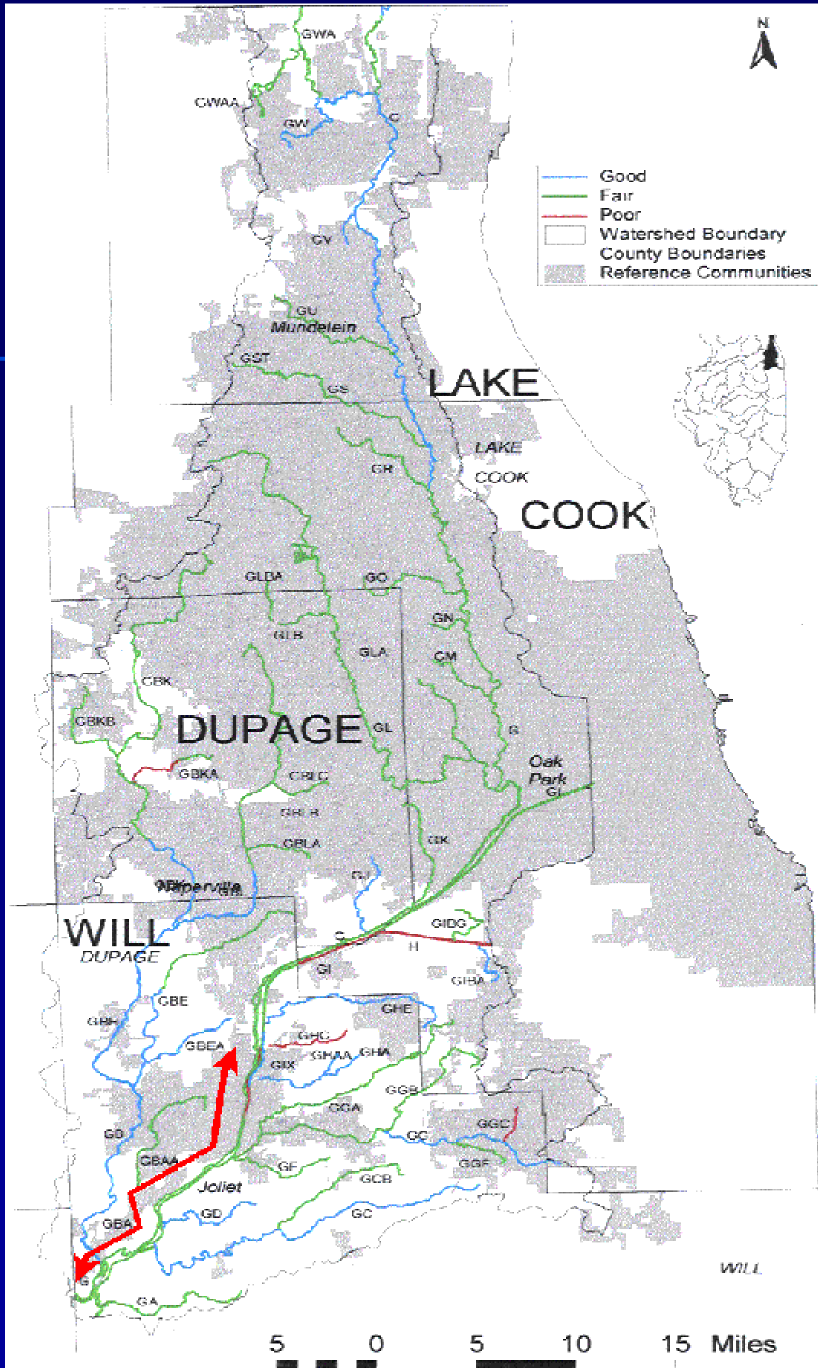


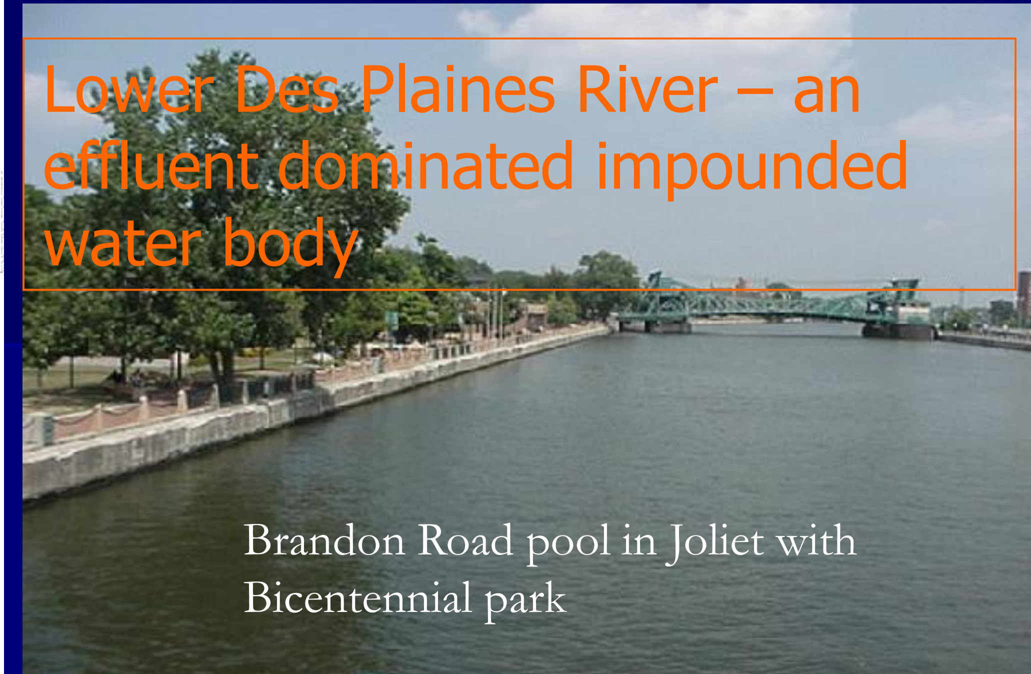
Novotny et al, Water Research, 2005

STREAMS AND RESERVOIRS

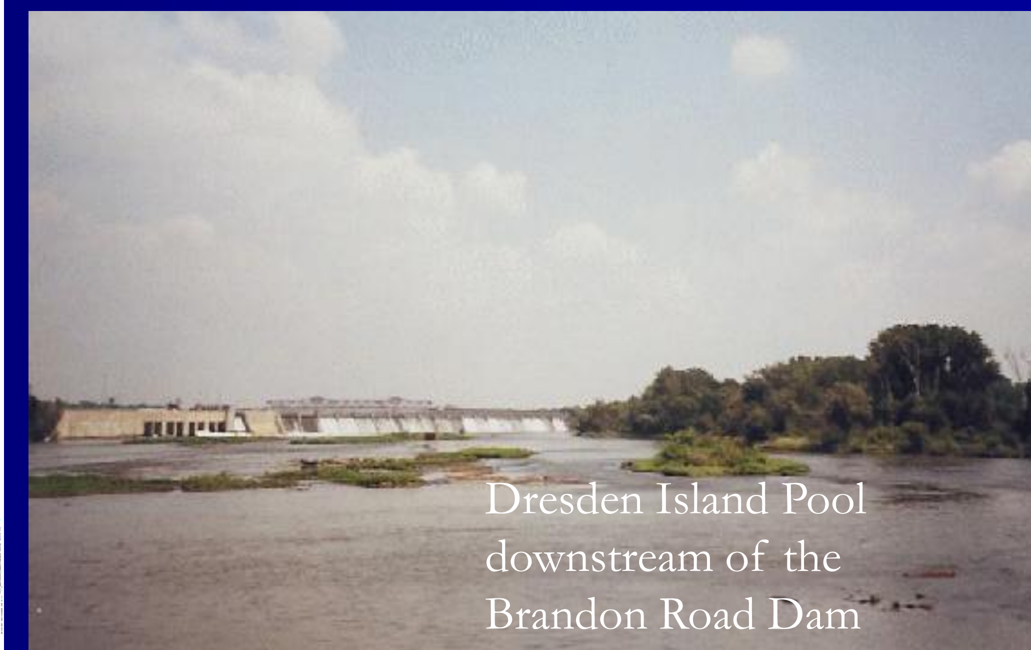
- Free flowing streams
- Deep and stratified reservoirs
- Shallow unstratified reservoirs
  - Navigation
  - Water supply
  - Power generation
  - Irrigation
  - Powered numerous mills before 1900's but have no purpose now
  - Some flood control

Lower Des Plaines River – an effluent dominated impounded water body





Brandon Road pool in Joliet with Bicentennial park



Dresden Island Pool downstream of the Brandon Road Dam

Dresden Island Pool

Near Empress Casino

I-55 bridge

Better Habitat

Thermal pollution by two large power plants

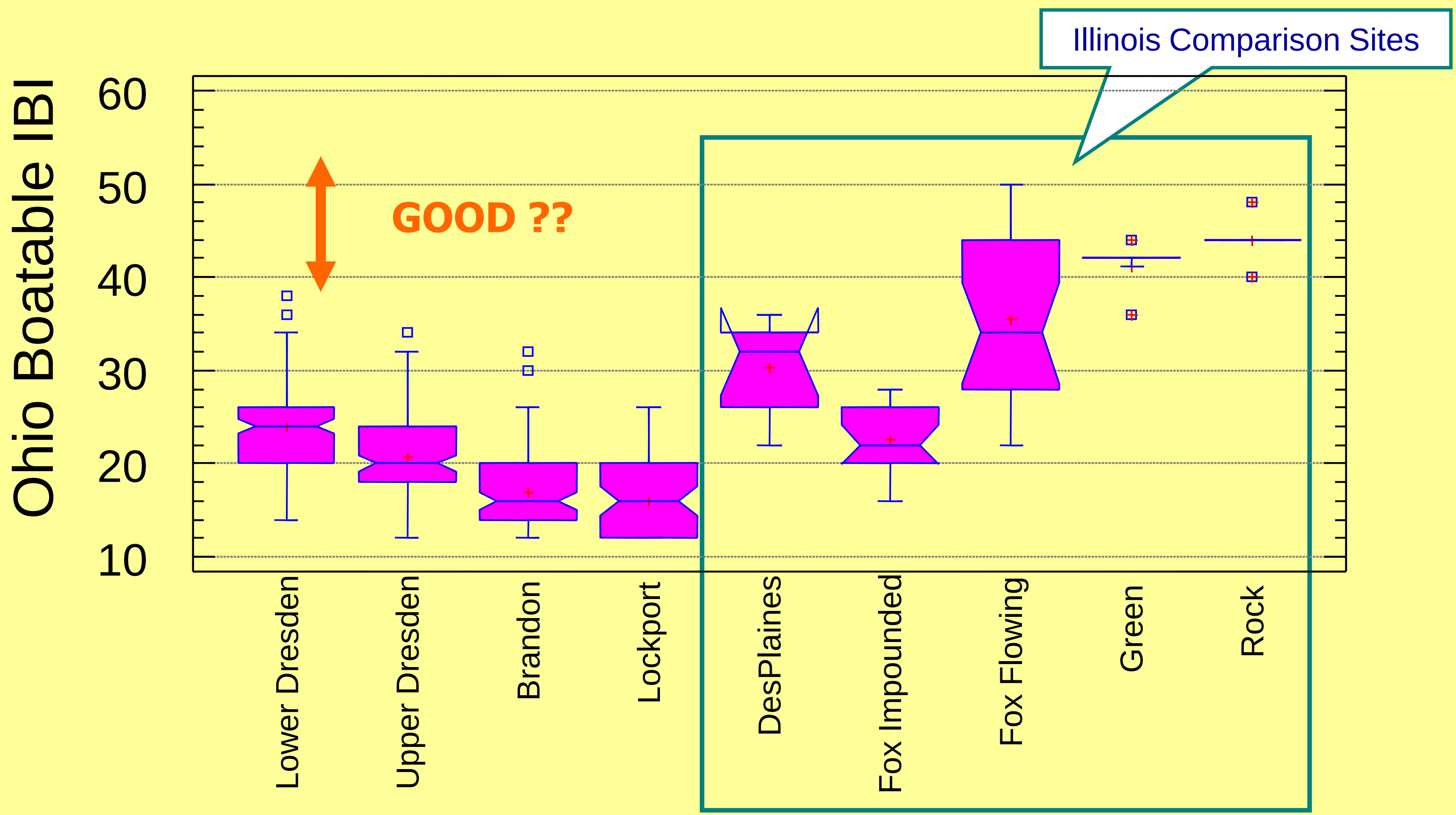
Enriched by nutrients

Navigation – minimal water level fluctuations

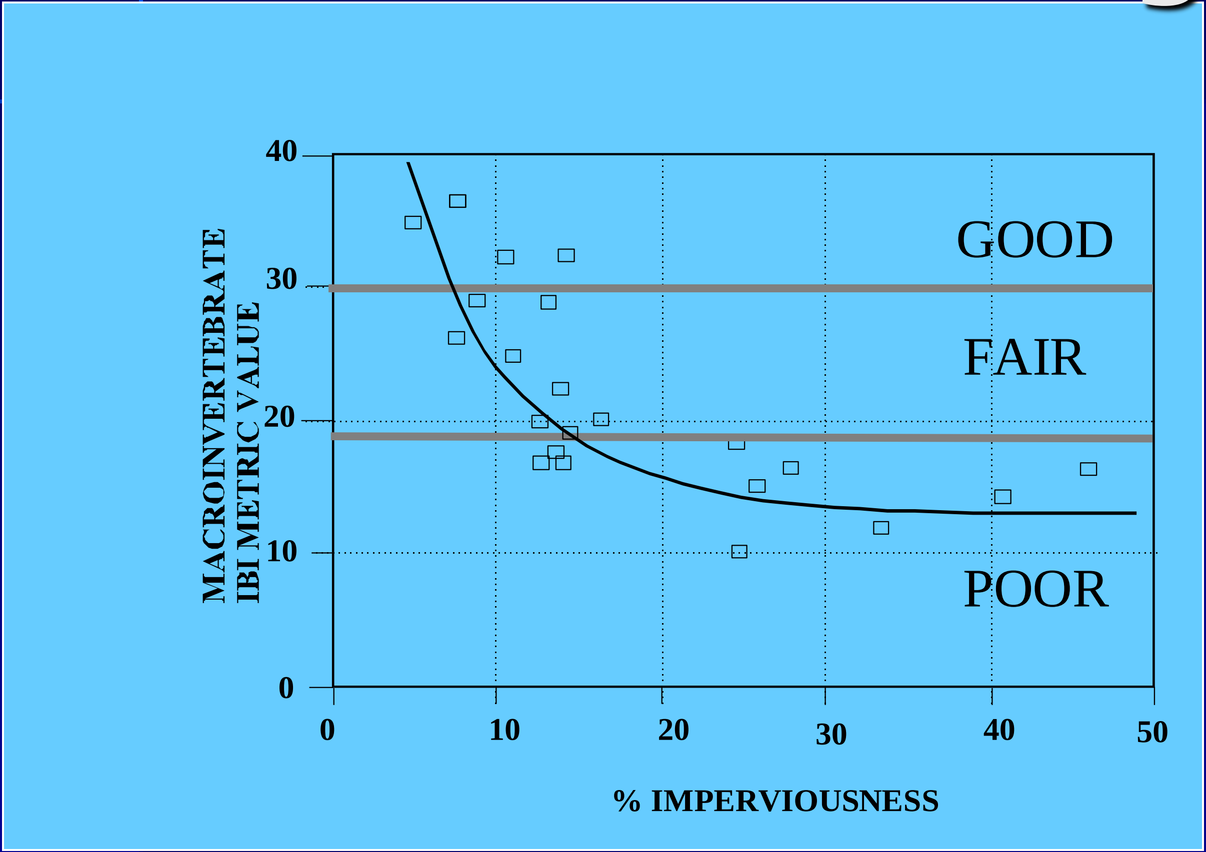
Limited recreation

Industrial zone

Ohio IBI Scores Calculated for Selected Impounded Illinois Waterways



Simplistic Linking of Stresses to Integrity



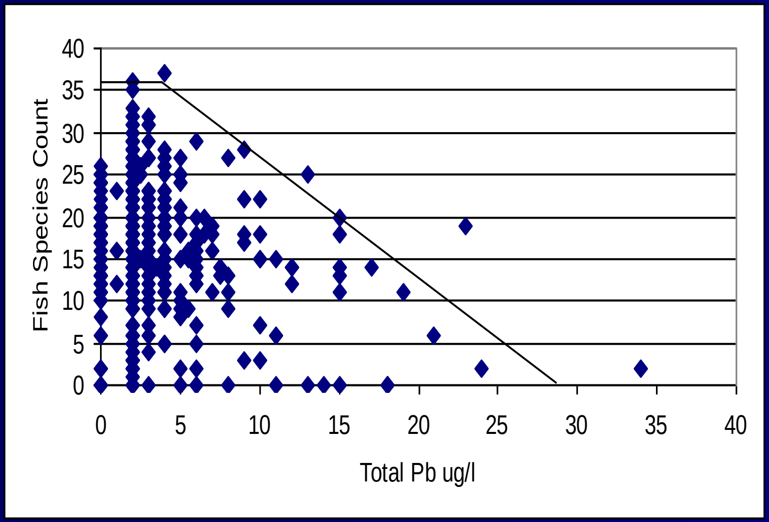
Schueller (1994)

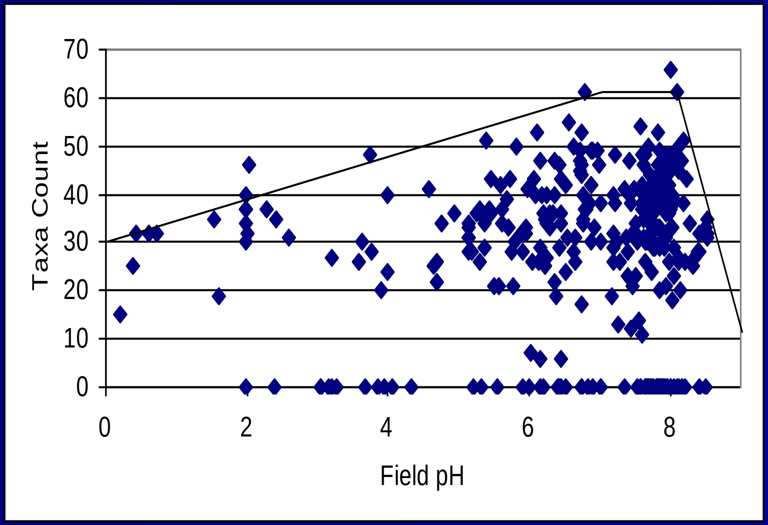
- A watershed is impaired if % imperviousness is more than X percent
- Reversible vs. irreversible stresses
- Imperviousness is a surrogate for many “bad” things that lead to impairment

Systems Being Modeled

- The Northeastern University team has retrieved from large data bases relationships between the fish and macro-invertebrate IBIs and their metrics (Level 1 of the risk propagation pyramid) and level 3 and 4 stressors for rivers
  - Benthic Macroinvetebrate IBI serves two purposes
    - It is a biotic endpoint
    - It is a surrogate for sediment contamination
- The models are statewide. The state date bases have 1000 to 2000 sites with often multiple observations
  - Ohio
  - Maryland
  - Wisconsin
  - Minnesota
  - Massachusetts

Sample MSRLs - Ohio





MSR = Maximum Species Richness

- defined by 95 percentile

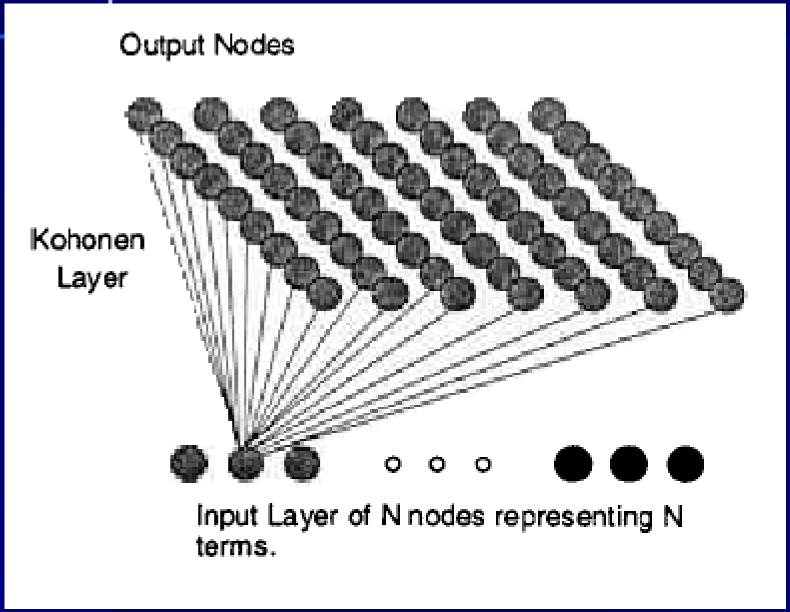
Dealing with Multiple Stressors and multiple endpoints

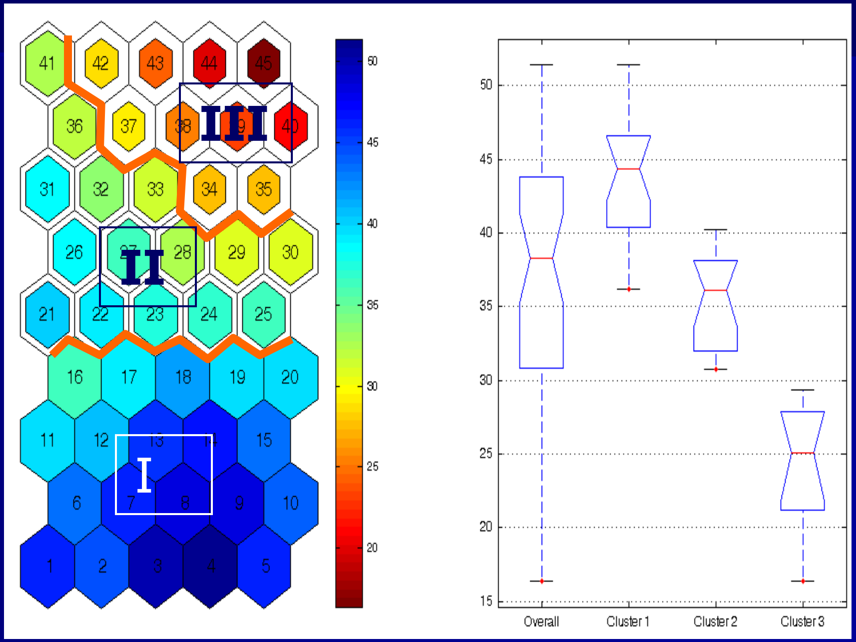
- Many current approaches are incapable of dealing with multiple stressors directly.
- Most single stressor risk assessments assume stressors are additive.
- Artificial neural networks (ANN) are capable of considering multiple inputs and outputs and evaluating their relative impact.
- Many stressors are cross-correlated

MODELING APPROACHES

- Artificial Neural Networks (ANN)
  - Learning
    - *Unsupervised: Pattern detection*
      - *Self-organized Mapping*
    - *Supervised: Output tracking and predictions*
      - *Feed forward ANN*
      - *Back propagation ANN*
- *Advanced Multi-regression Analysis*
  - *Canonical Correspondence Analysis*
    - *Used for post analysis of impact and determining of cluster dominating parameters*
  - *Principal Component Analysis*

Clustering of IBIs and their metrics



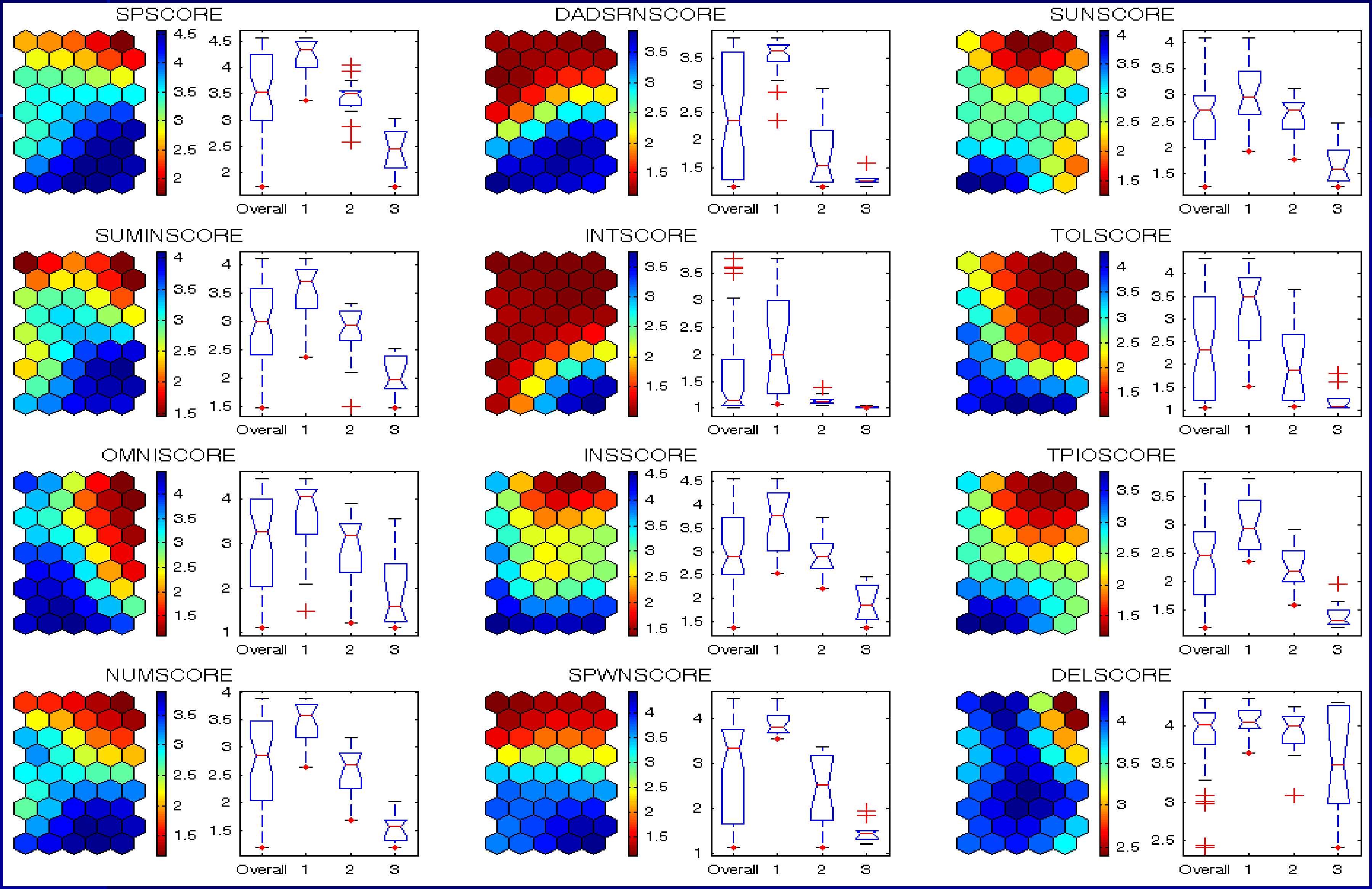


Self Organizing Map

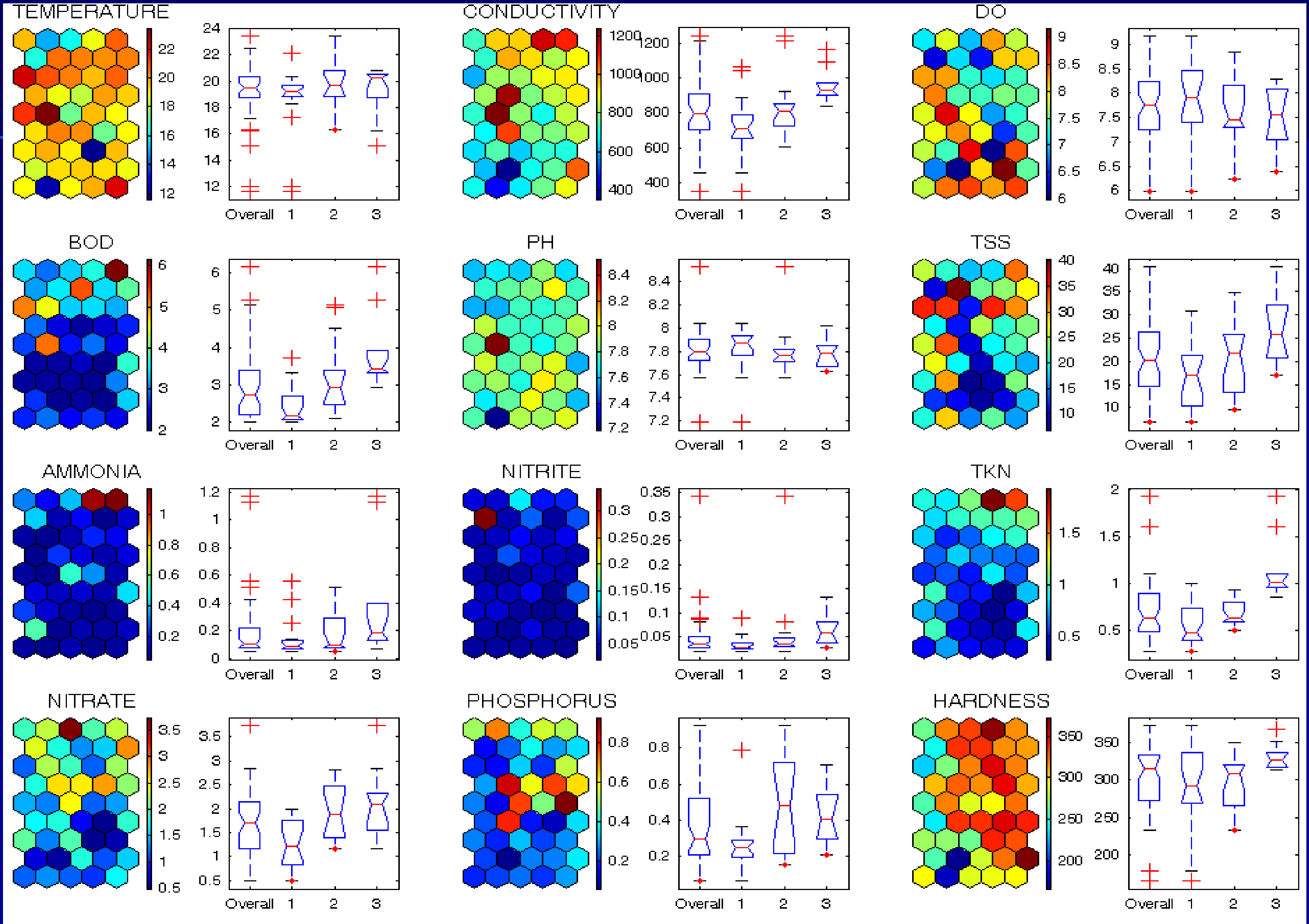
knowledge data mining by unsupervised learning with Artificial Neural Nets

SOM clustering of Fish IBI metrics for the State of Ohio

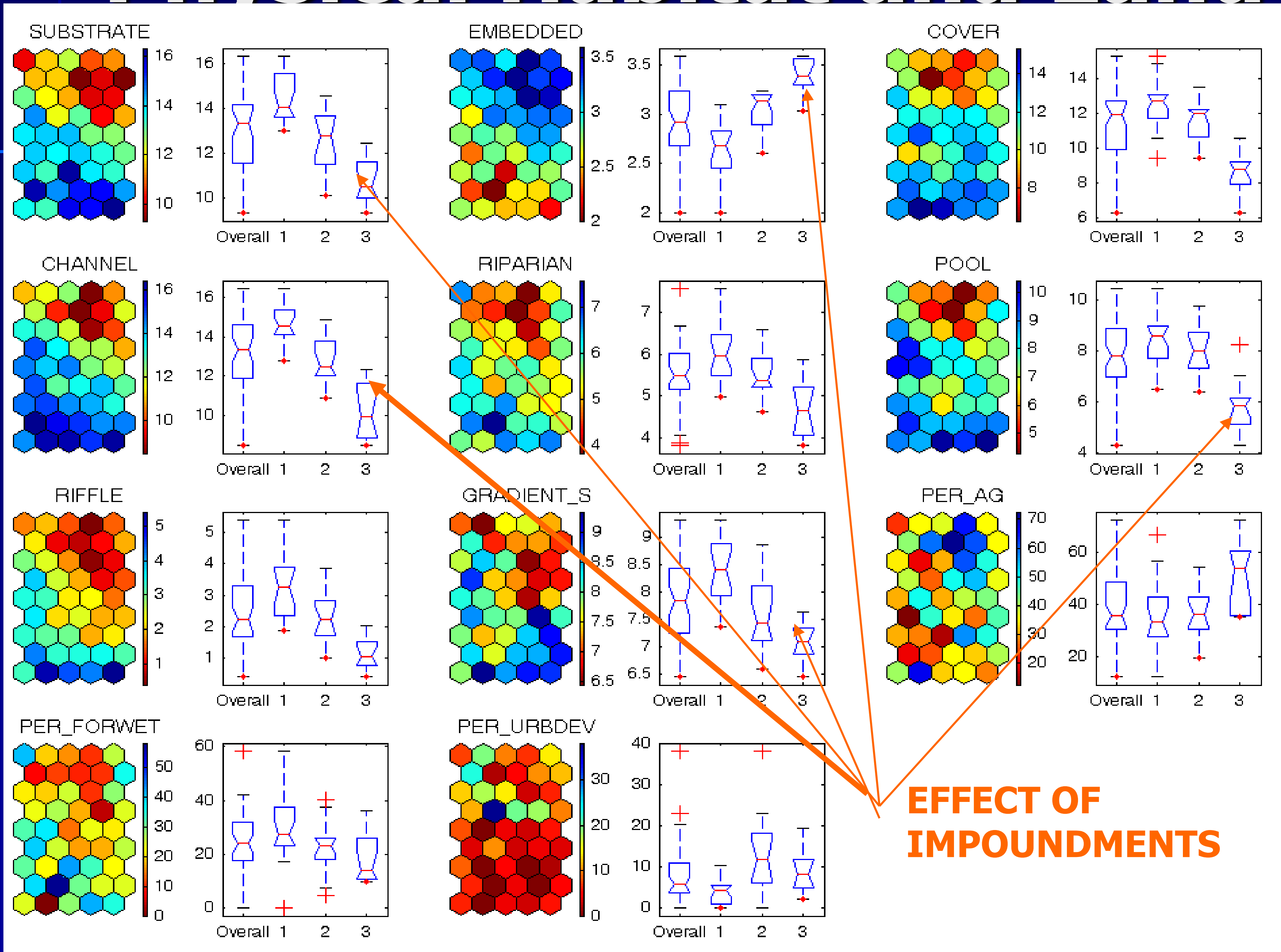
Distribution of Metrics - OH



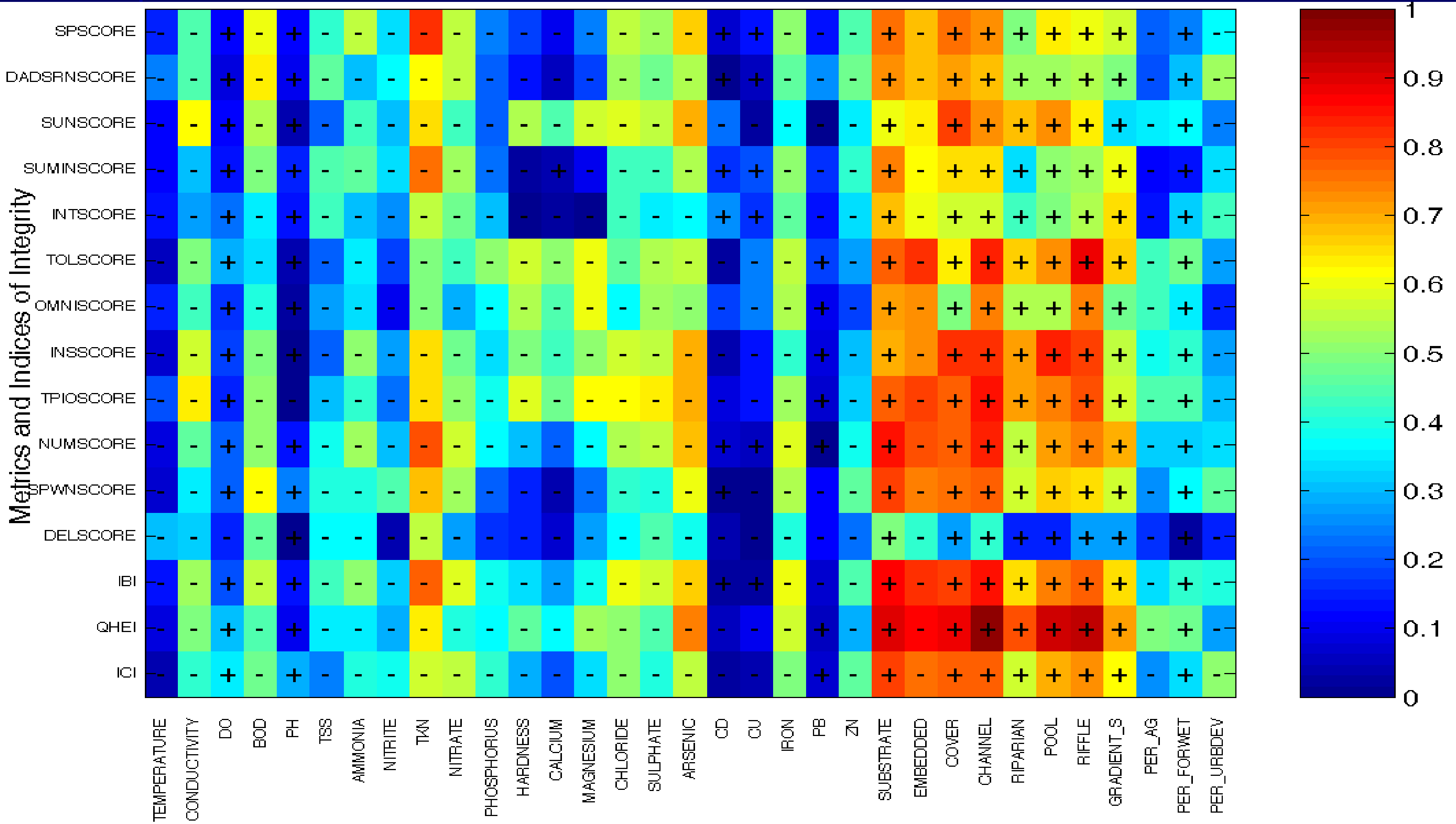
Water Chemistry I - OH



Physical Habitat and Land



Correlation matrix over the SOM

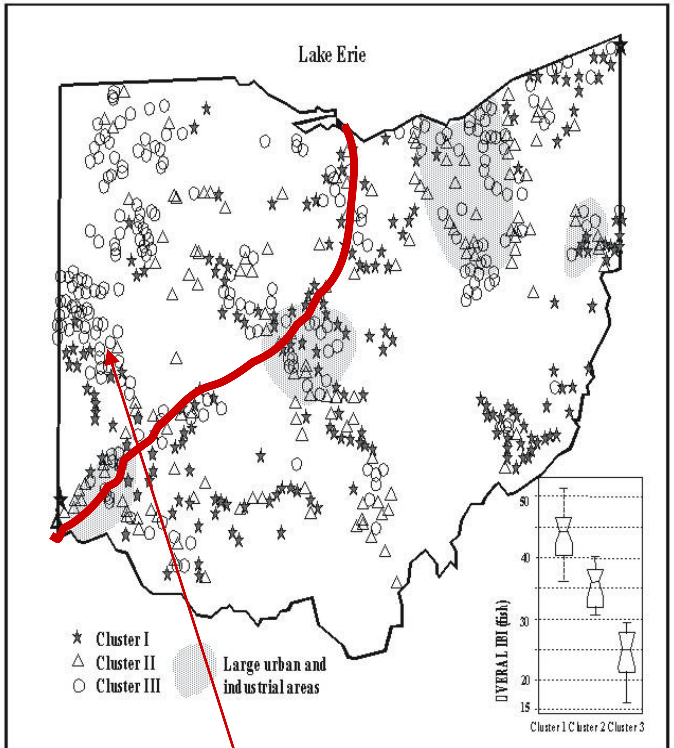


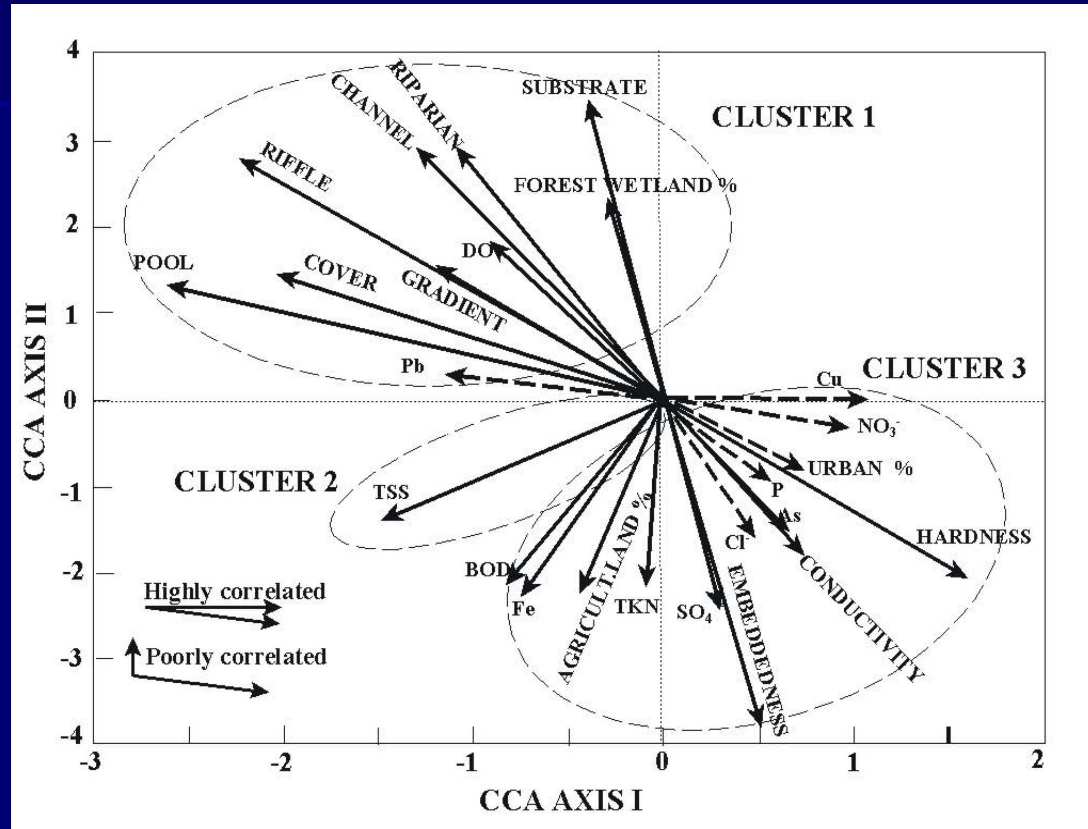
Water Quality

Habitat Quality

Land Cover

Distribution of clusters and cluster dominating parameters (CDP)



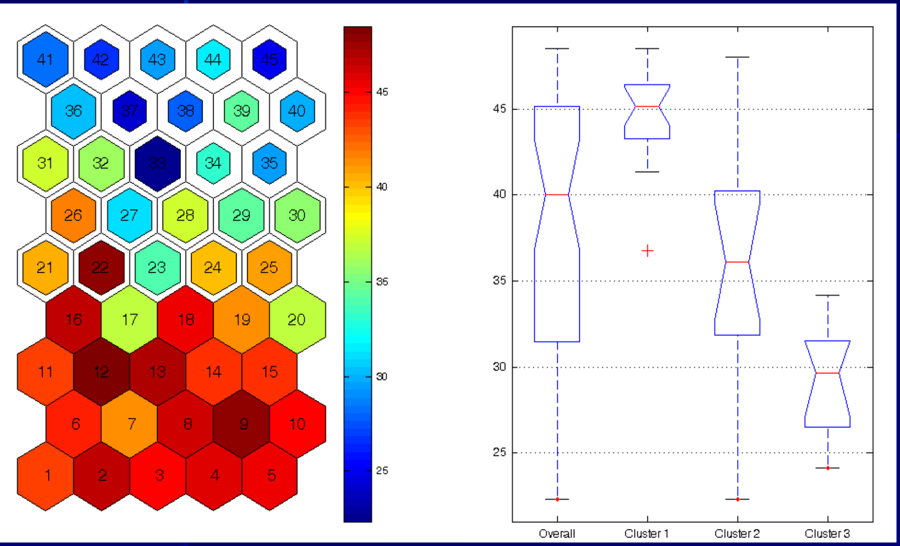


Agriculture in northwest Ohio

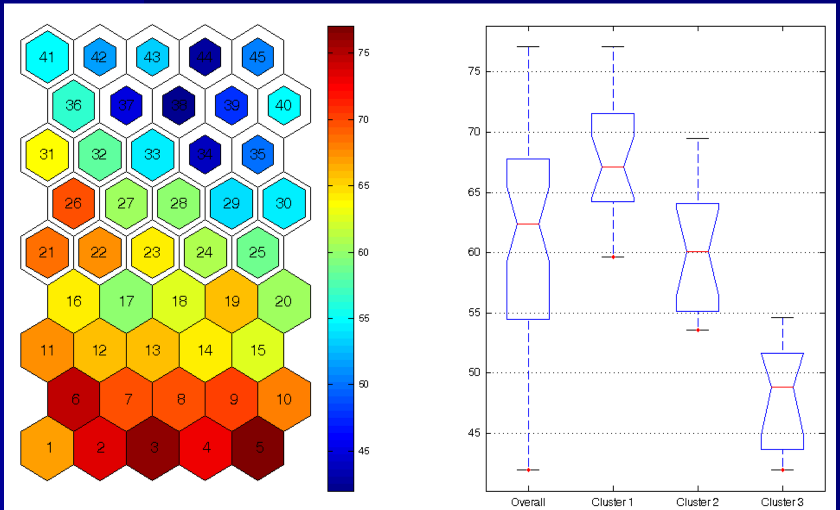
Canonical Correspondence Analysis of the impact of environmental variables within clusters determined Cluster Dominating Parameters

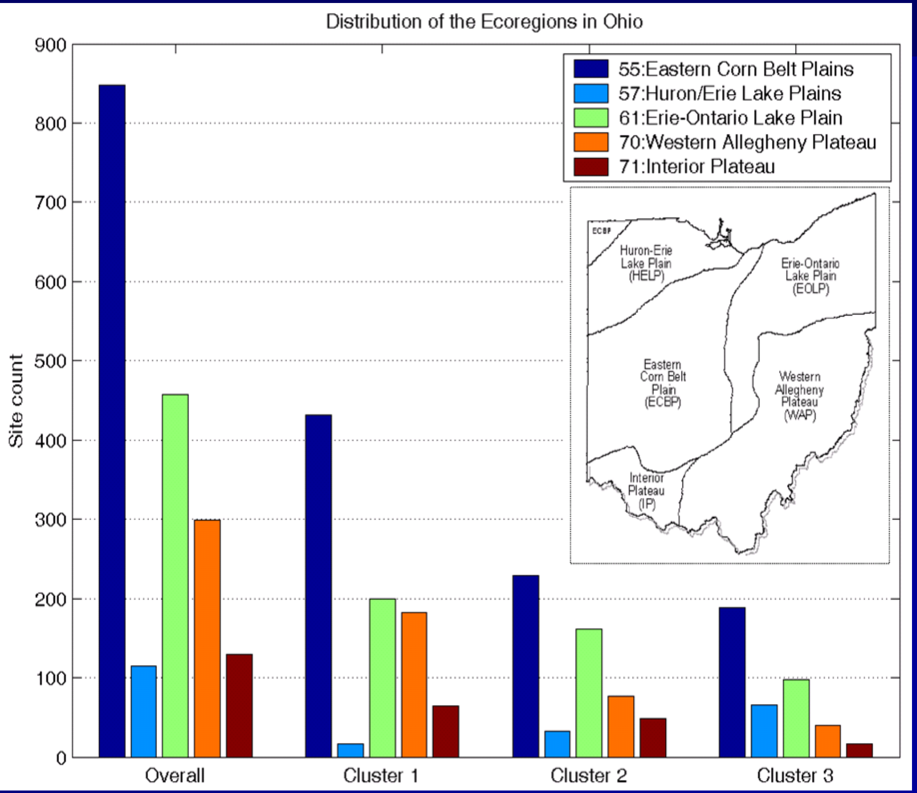
Visualizations of ICI and QHIE

SOM visualization and Clustered Boxplots for ICI



SOM visualization and Clustered Boxplots for QHEI

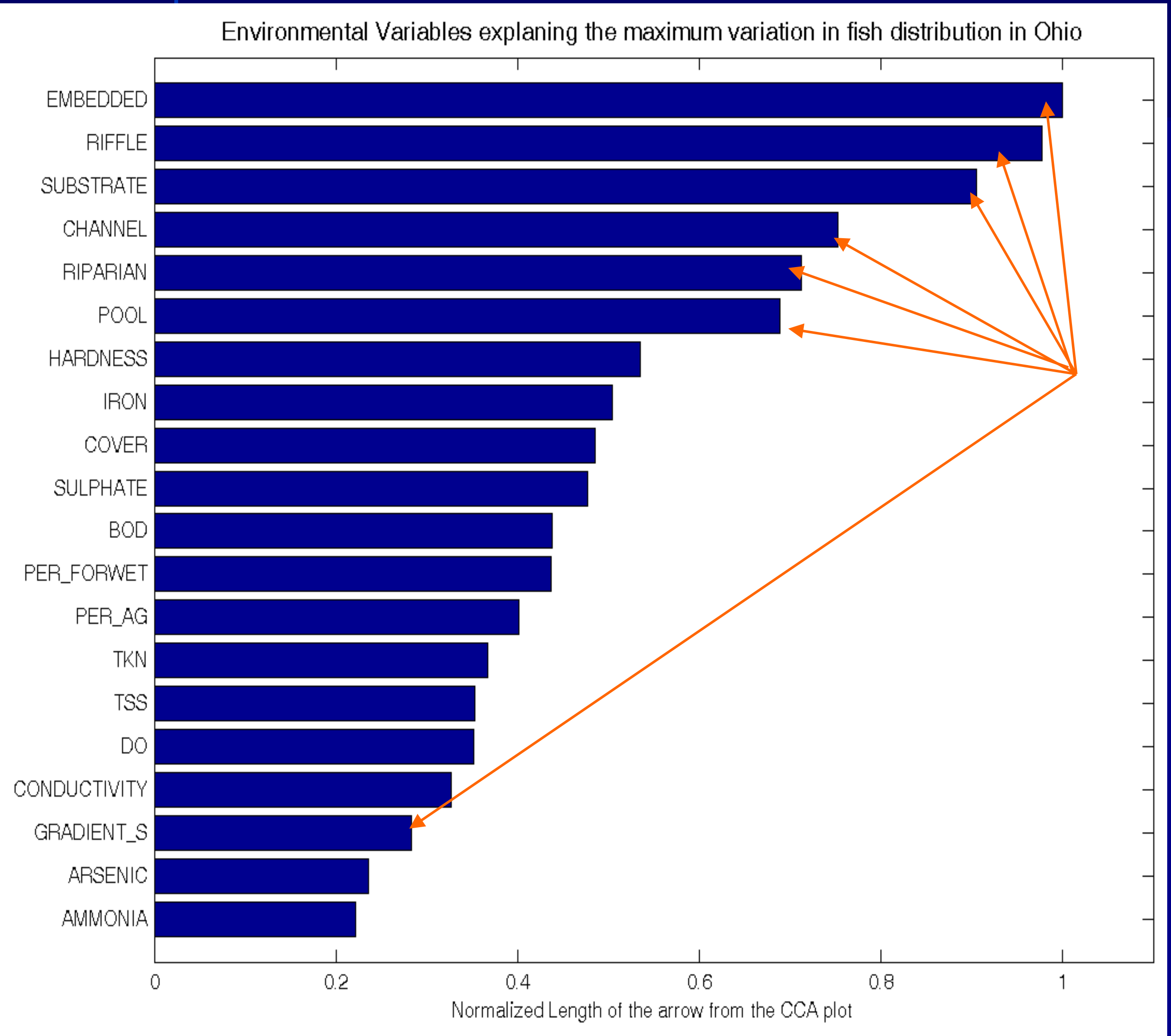




ICI macroinvertebrate index

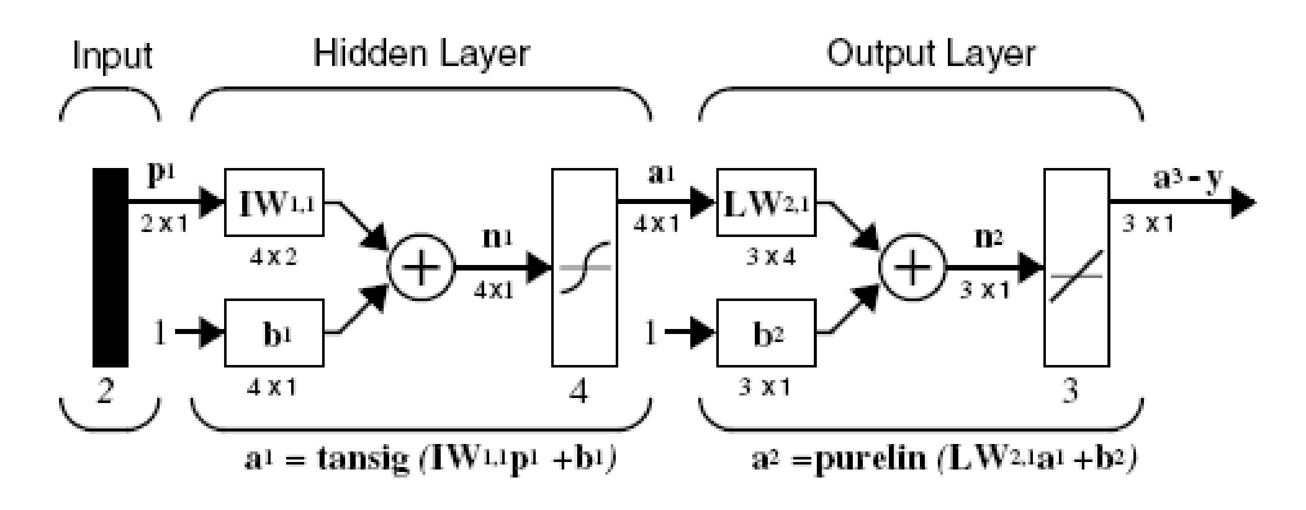
QHIE habitat quality index

Relative Rankings of environmental variables



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

Quick Introduction to Supervised ANN Models



A model of a layered network (Demuth and Beale 1992)

Supervised Learning ANN

- Supervised learning techniques used to develop prediction of Fish IBIs.
- Several feed-forward backpropagation networks developed and tested to assess what kind of structures could work with the given dataset
- Input included chemical and habitat quality measurements (and also ICI in some models).
- Targets for prediction: Fish IBI & metrics.
- Modeled: Ohio, Maryland, Wisconsin, Minnesota

Fish IBI Prediction Models (Ohio)

Model Identification	Model structure	r-train	r-valid	r-test
Clipped_10*	33v, 35 in, 50hn, 1on	0.703	0.635	0.615
Clipped Multiple Regression_1*	33v	0.617	0.568	0.389
C2_2	33v, 35, 35, 1	0.756	0.715	0.691
C2_6**	10v, 10, 20, 1	0.658	0.65	0.643
C2_14	33v, 35, 50, 1	0.857	0.707	0.662

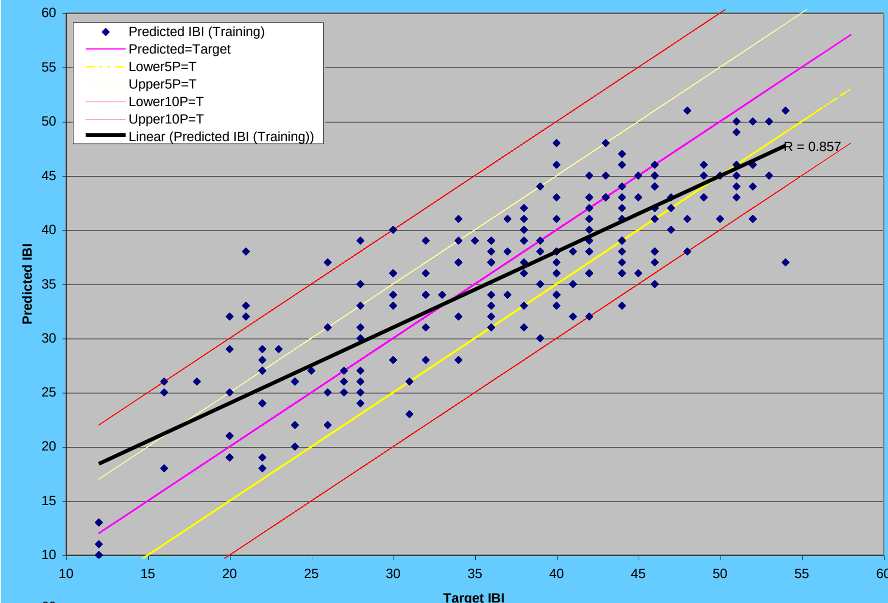
V = number of parameters ; in= # of input neurons, hn= # hidden layer neurons, on = # of output neurons

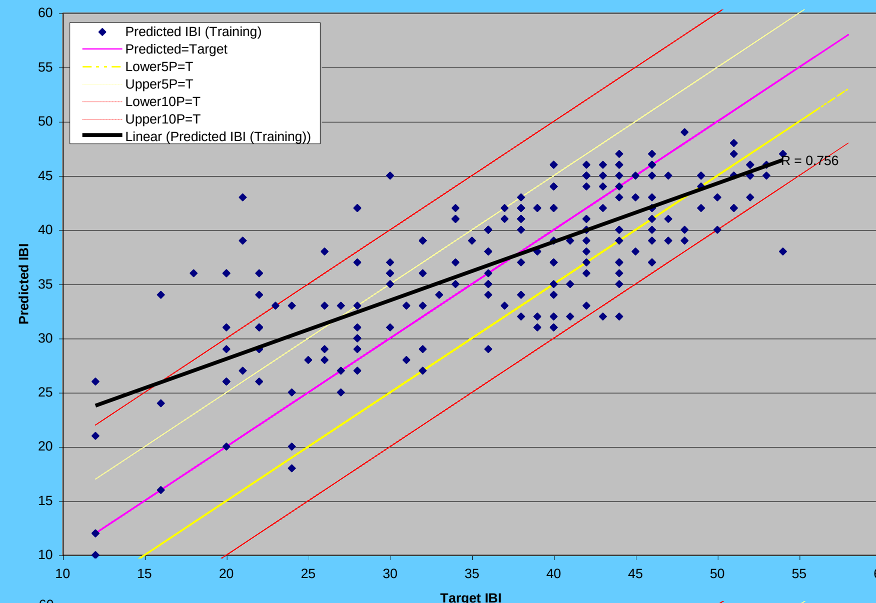
Out of 1149 data set 60% were used for training, 20% for validation and 20% for testing

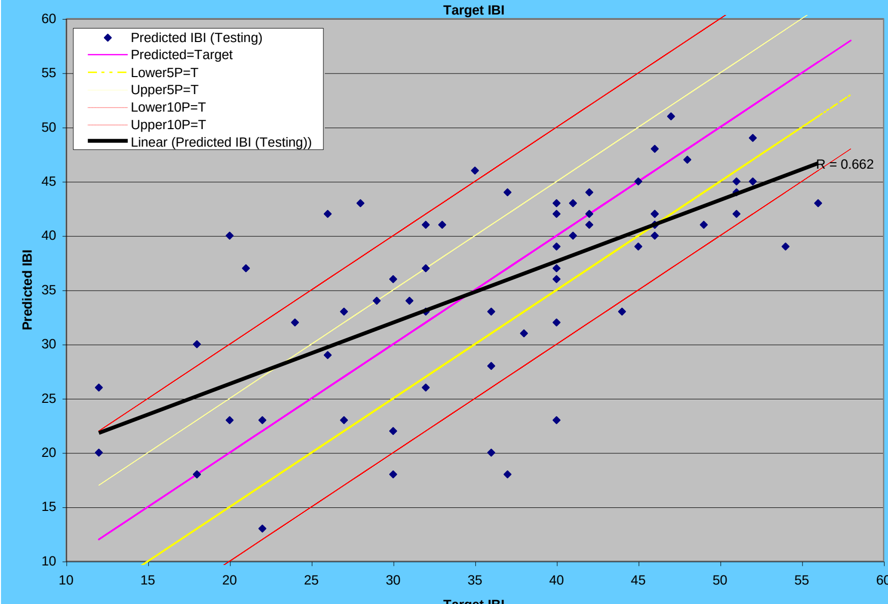
* Values with IBI<15 or > 57 not included (Full range of IBI is 12 to 60)

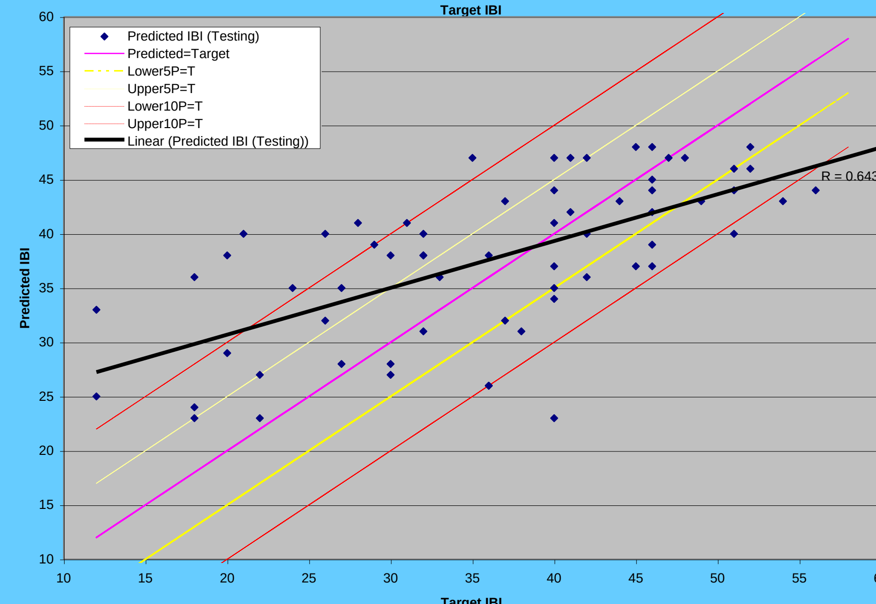
**Top 10 variables based on CCA ranking (Embeddedness, riffle, substrate, channel, riparian, pool, cover, iron, hardness, sulphate)

Models C2_14 (top) Vs. C2_6 (bottom): Training (left) & Testing (right) - OHIO









APPLICATIONS

- Site specific biotic criteria and ecologic potential can be ascertained
- Riverine impoundments affect adversely key parameters
  - Embeddedness and gradient (cross-correlated), siltation, cover, pool/riffle, spawning
  - DO (reaeration)
- Impact of remedial action (e.g., decommissioning of the impoundments) can be ascertained

Application to Blue - Greens

Orlík

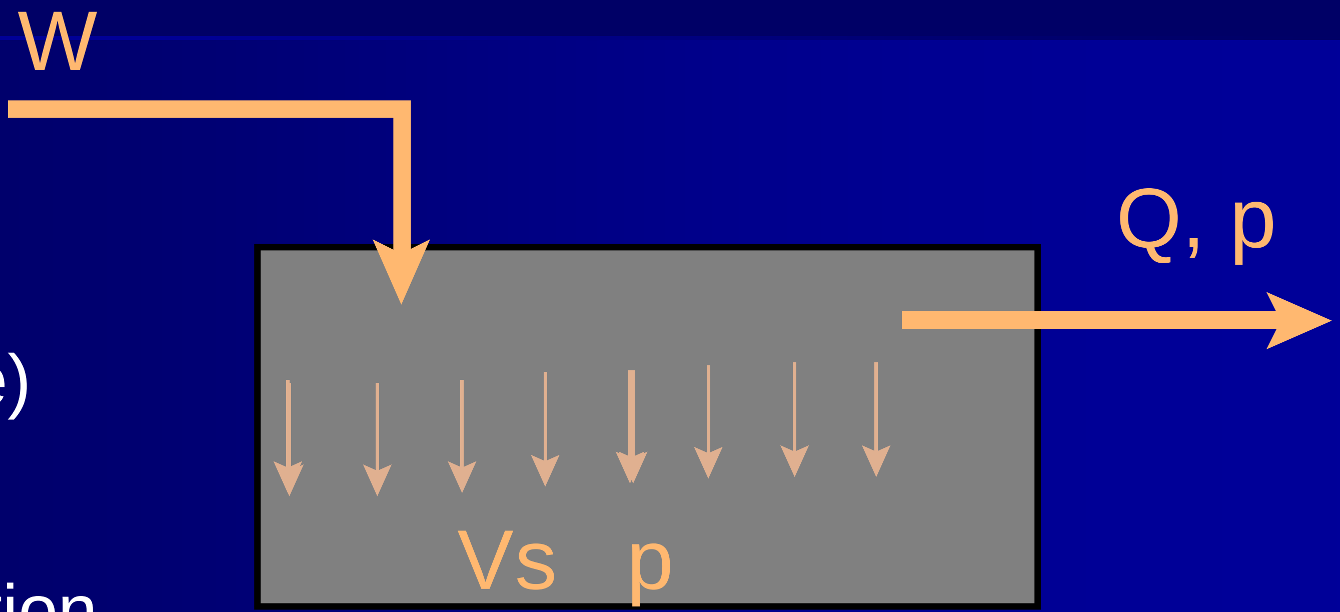
Sedlice

Czech Republic has about 24 000 reservoirs and ponds, 70% suffer from cyanobacteria blooms

Cyanobacteria (Blue Green Algae) in Lake Mendota, Madison, WI

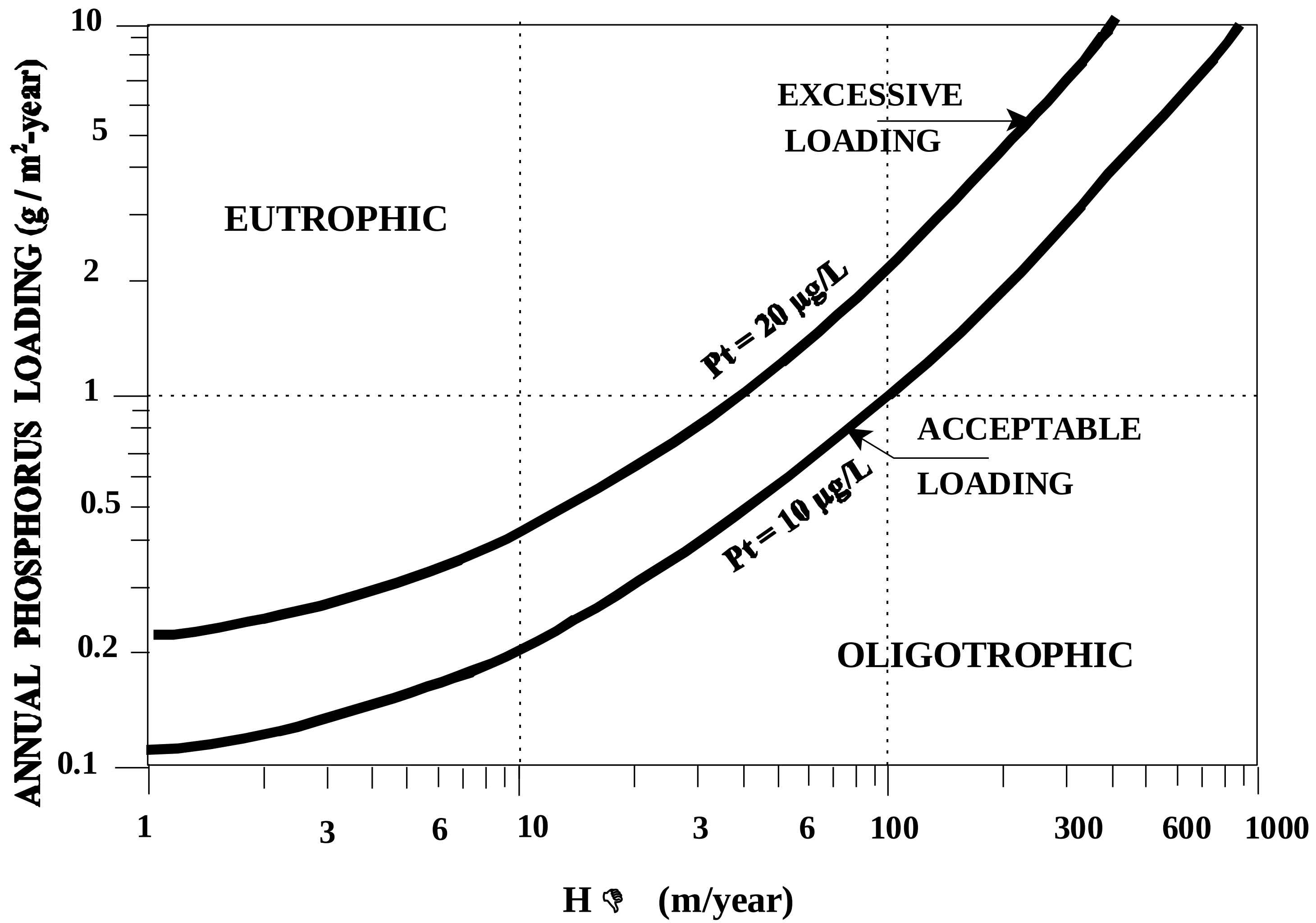
Cyanobacteria produce toxins, cause rash to swimmers, bad taste and odor

Vollenweider's Completely Mixed Lake Schematics



- W= loading (Mass/time)
- V=volume
- Q=flow
- p=concentration
- V_s = settling velocity
- A_s= surface area

$$V\frac{dp}{dt} = W - V_s A_s p - Qp$$



ANNUAL PHOSPHORUS LOADING (g / m²-year)
10
5
2
1
0.5
0.2
0.1
EXCESSIVE
LOADING
EUTROPHIC
Pt = 20 µg/L
Pt = 10 µg/L
ACCEPTABLE
LOADING
OLIGOTROPHIC
1
3
6
10
3
6
100
300
600
1000
H (m/year)

Current wide spread models

Simulate algal growth and nutrient uptake and impact on DO regime

QUAL2- quasi dynamic model

WASP - EUTRO for streams and stratified impoundments - a dynamic time variable model with some restrictions

HSP-F - receiving water quality model

Problem!! Most of the previous concepts do not work for cyanobacteria that are becoming a widespread worldwide nuisance

Blue – Greens in Czech Republic

Orlík

Sedlice

75% or reservoirs have been infected

PEA SOUP IN CHINA

Pea soup. Hans Paerl samples cyanobacteria in ailing Taihu Lake.

ECOLOGY

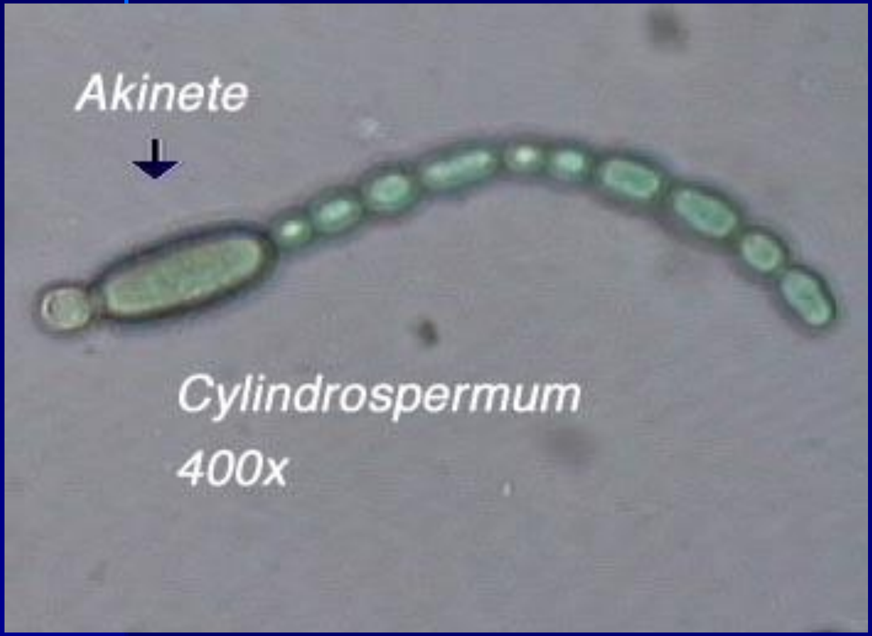
Doing Battle With the Green Monster of Taihu Lak

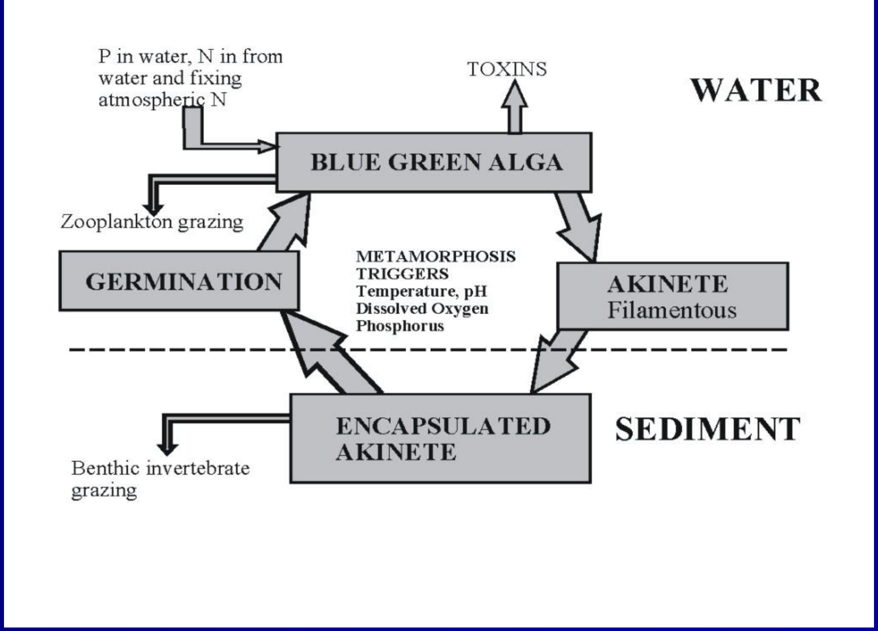
In attempting to subdue a vicious algal bloom, scientists aim to

Cyanobacteria (Blue Green Algae)

- Cyanobacteria have been around for more than 3 billion years
- These microorganisms are responsible for atmospheric dissolved oxygen
- Some species can assimilate atmospheric nitrogen
- They are ubiquitous to nature
- They prefer warmer water bodies
- They produce toxins and impair taste and odor

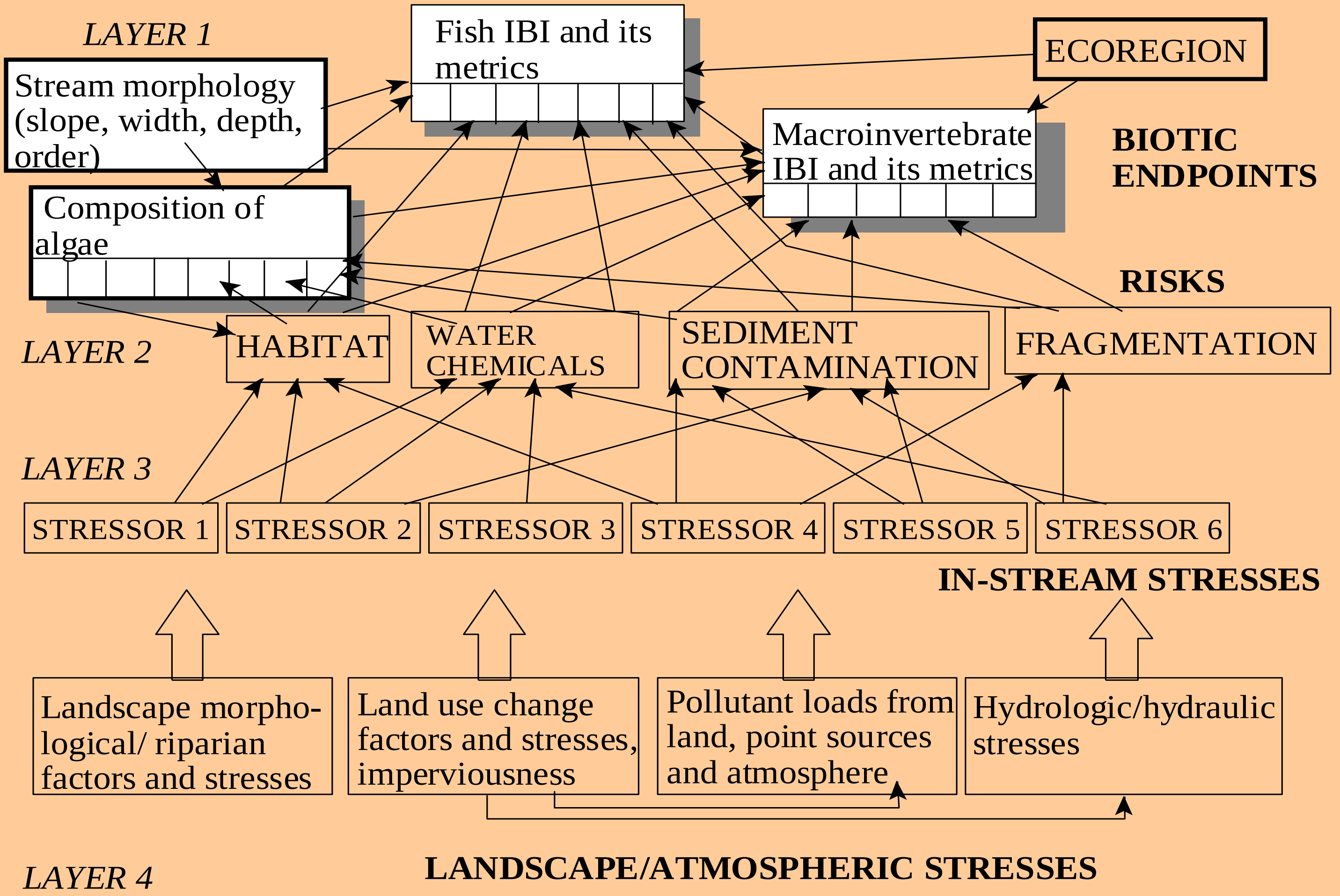
Some cyanobacteria fix atmospheric nitrogen and encapsulate into akinetes





Akinetes settle into sediments where they overwinter and can take up phosphorus. They can stay in the sediment for several years and rise into water when conditions are favorable. They have a preference for higher temperatures (impact of global warming)

HIERARCHICAL MODEL CONCEPT OF RISK PROPAGATION FROM STRESSORS TO BIOTIC ENDPOINTS



Novotny et al, Water Research, 2005

Hierarchical Organization – SOM Application

- Endpoints (Layer 1)
  - Diatoms, Greens, Blue-greens, fish IBI and its metrics, chlorophyll a, DO fluctuations
- Risks and water body stresses (2 and 3)
  - Nutrients, BOD/COD, silica, temperature, pH, alkalinity, hardness, light input
  - Retention time, stratification potential, depth
  - Physical: Elongation, shading, buffers, elevation, water level variability
- Allochthonous (Layer 4)
  - Pollutant inputs point and nonpoint, atmospheric
  - Landscape/land use, land use changes, elevation
  - Hydrologic – flow, flow variability

Minimum number of sites ~ 60s

Modeling

- ~~One site only – current mechanistic models (e.g., AquaTox, it does not account for dormant akinete conditions and settling into sediment) – PROTECH. Such models treat algae like chemicals, i.e. describe them as µg Chl/l~~
- Models developed by data mining –Many sites
  - SOM - unsupervised learning, clustering
  - Multiple input and multiple output by Supervised ANN learning with the key cluster dominating parameters
- Hybrid models - One site
  - Agent Based Modeling
  - Water and sediment quality by traditional mass balance dynamic models

Agent Based Modeling

- Describes the life cycle of cyanobacteria
- Describes the microorganisms as "supper individuals", i.e. thousands of groups of hundreds of individuals (out of millions)
- The organisms represented by a super individual respond to the environmental stresses in a similar fashion, e.g., together they die off, convert to akinete , settle to sediment, feed on phosphorus or germinate

Conditions affecting cyanobacteria (Kravchuk, 2006)

- Temperature
- Phosphorus concentration
- Hydraulic conditions
- Light
- Grazing by zooplankton

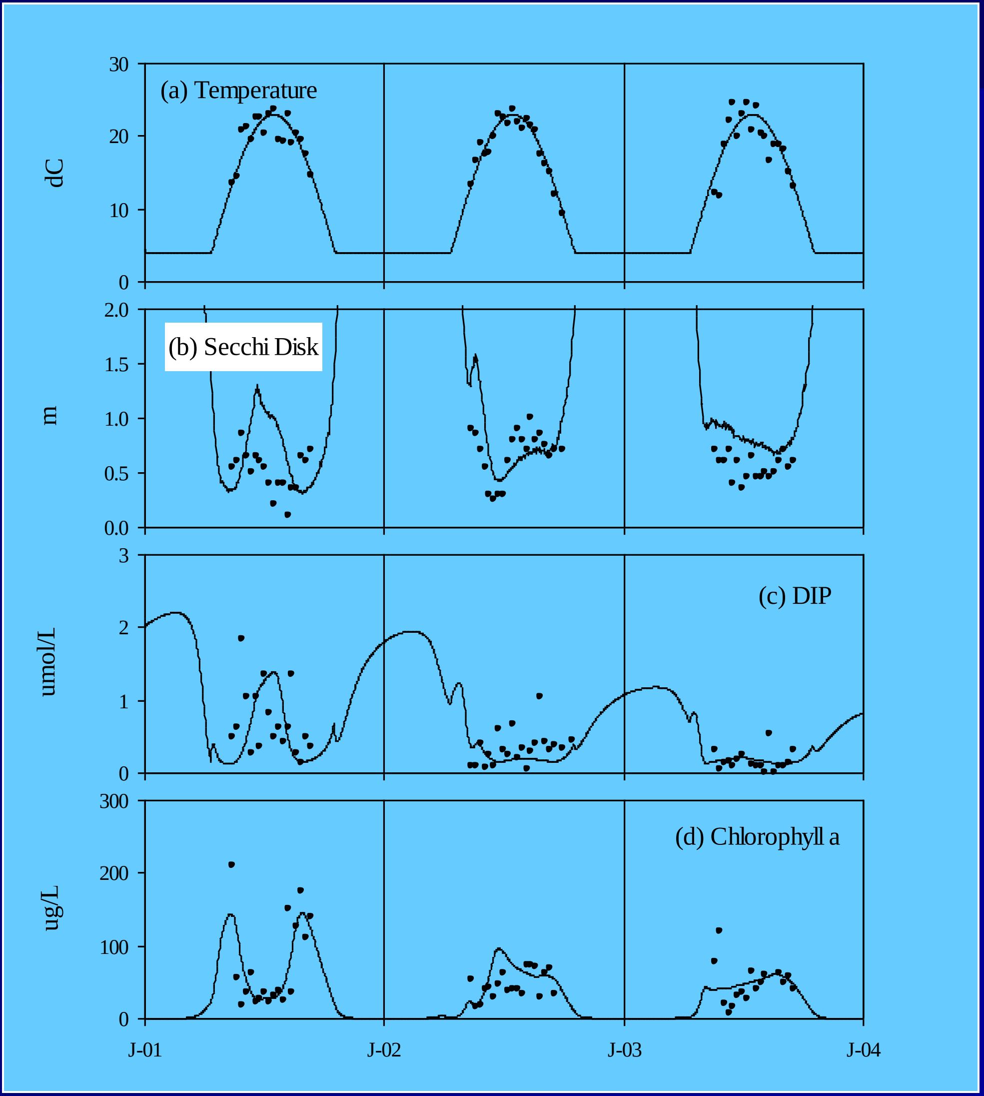
Algal blooms occur suddenly even when nutrients controls have been implemented (Lake Delavan, Charles River)

Agent Based Modeling

- See Hellweger, Kravchuk, Novotny and Gladyshev (2008) Agent based modeling of the complex lifecycle of Cyanobacterium (Anabena) in a shallow lake, *Limnologia and Oceanographia* ***53***(4), 2008
- Agent base models follow a history of of thousands of "superagents" and behavior, life cycle and resting stages in water and sediment

Preliminary results

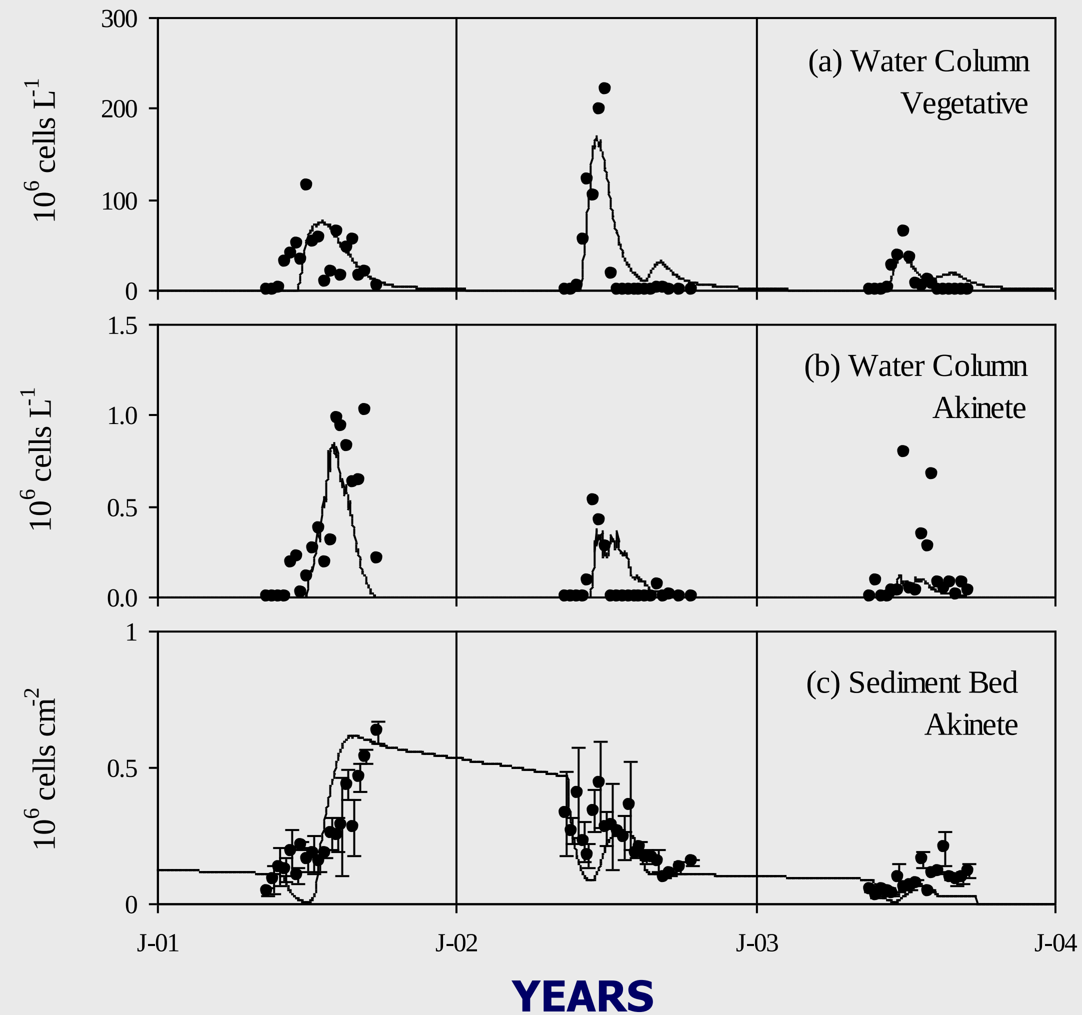
Conventional dynamic completely mixed lake model calculated environmental v variables affecting cyanobacteria. Both water and sediment were simulated.



Preliminary results

Model calibration to *Anabaena* (a) vegetative cells in the water column, (b) akinete cells in the water column, (c) akinete cells in the sediment bed, +/- 1 standard error (top 1 cm). Lines are model predictions and symbols are data from Kravchuk et al. (2006).

(Hellweger et al., 2008)



Conclusion on cyanobacteria

- The preliminary research documents the utility of agent based modeling for cyanobacteria
- Most of the phosphorus input occurs in the sediment in the akinete stage
- During the bloom most of the algae cells originate from germinated akinetes deposited into the sediment in the preceding years
- Formation of the resting stages (akinetes) is critical to the survival of the cyanobacteria

THANK YOU

ACKNOWLEDGEMENT

Illinois Environmental Protection Agency (The Des Plaines River)

US Environmental Protection Agency STAR Watershed Research Grant to Northeastern University

Also check how we are doing at

www.coe.neu.edu/environment