Northeastern University

June 01, 2007

Identification of the Main Biotic Integrity Stressors and their Relationships using Cluster and Neuron Analysis with Self-Organizing Maps in Ohio, Maryland and Minnesota

David Bedoya
Northeastern University

Identification of the main biotic integrity stressors and their relationships using cluster and neuron analysis with Self-Organizing Maps in Ohio, Maryland and Minnesota

TECHNICAL REPORT NO. 13

David Bedoya, PhD Candidate

Center for Urban and Environmental Studies
Northeastern University
Boston,MA

Vladimir Novotny, Ph.D., P.E.
Primary Investigator
Center for Urban Environmental Studies
Northeastern University, Boston, MA

Bernice Smith
EPA Project Officer

Iris Goodman
EPA Program Director

June 2007

Abstract

Big environmental databases from different public agencies were obtained in the states of Minnesota, Ohio and Maryland. Biotic indices along with physical and chemical environmental variables and habitat metrics were some of the data available. We used Self-Organizing Maps (SOM) to group the data into physically and chemically homogeneous stressor groups. These were either groups of similar SOM neurons (clusters) or the SOM neurons themselves. When working with clusters of neurons, the biotic index's values statistical differences among clusters were identified using multiple range tests. Subsequently, the same procedure was applied to all the available environmental variables. Variables with similar homogeneous groups distributions to the biotic integrity indices was interpreted as a variable with an important effect on biotic integrity. The neuron-based analysis focused on regressing the neuron environmental variables values versus the neuron-based biotic index. The parameters with highest correlations were considered as most important. Both methodologies seemed to work well, especially in the case of Ohio and in the cluster-based analysis in Minnesota. Maryland also showed promising results and the separation of the sites in different strata clearly showed how the stressors are different in coastal sites than in the rest. The neuron-based analysis usually identified the same stressors in biotic integrity as the cluster-based analysis. Moreover, some of the relationships among off-stream and in-stream environmental variables as well as some of the in-stream physical variables and chemical elements could be explained. The SOM is a very powerful tool in identifying highly dimensional, with high natural variability, non-linear problems by means of data organization and pattern recognition

Acknowledgements

TABLE OF CONTENTS

1. INTRODUCTION

Biotic Integrity is defined as the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition and functional organization comparable to that of natural habitat in the region (Frey, 1975, Karr and Dudley, 1981). Karr and Kerans (1991) identified five elements which are the main constituents of biotic integrity: water quality, habitat structure and quality, flow regime, energy sources and biotic interactions. Any alteration of the natural state of any of these five components will ultimately affect biotic integrity. One of the main advantages of measuring biotic integrity as a parameter of stream health is because it reflects any impairment taking place within the stream, physical or chemical, and it has memory, being able to reflect for a period of time past polluting events (Novotny, 2003).

Traditionally, water quality has been mainly assessed by measuring only the chemical composition of surface waters. However, water quality is defined in the Clean Water Act as the chemical, physical, biological or radiological integrity of the waters. Chemical quality is just a part of the puzzle that leads to a final biotic integrity. Many studies show how the chemistry approach alone fails many times in identifying impaired biotic integrity. Some cases are in Ohio, in which water quality alone failed to identify 50% of the impaired water bodies (Rankin et al., 1990). Identification or prediction of a stream's biotic integrity is a complicated task and predicting the outcome when some of the five main components of biotic integrity is modified is not an easy challenge. All components are intertwined and the modification of one of them will inevitably affect all or some of the other components and, ultimately, have an effect on the integrity of that stream.

Biotic integrity is measured in the United States with a multi-metric approach. The so called indices of biotic integrity involve extensive sampling of fish or benthic organisms. The results of the sampling are then compared to reference sites, which represent the values that should be expected in the case of no human impairment. In the U.S., the biotic indices are usually based on the work by Karr et al. (1986). This index is comprised of twelve different metrics grouped in three categories: species richness and composition, trophic composition and fish abundance and condition. It consists of fish sampling and scoring for each metric. The scoring is based on a scale in which the highest scores correspond to sites that resemble reference sites and viceversa. Many states have developed their own fish IBI (Ohio EPA, 1987, Niemela and Feist,2000, Niemela and Feist, 2002, Roth et al., 2000). Also a myriad of benthic community indices exist. Some examples are the Hilsenhoff index (Hilsenhoff, 1987), the ICI or Ohio's Invertebrate Community Index (Ohio EPA, 1987), the Benthic Index of Biotic Integrity (BIBI) in Maryland (Stribling et al., 1998), or the Macroinvertebrate Index of biological Integrity (MIBI) in Minnesota (Chirart, 2003, Genet and Chirart, 2004). The advantages of measuring macroinvertebrates instead of fish are that they are relatively immobile, easy to collect at low cost, they occupy all stream habitats and are quick to react to environmental change (Ohio EPA, 1987, Mason, 1991).

Usually, fish and/or macroinvertebrate sampling goes along with water quality sampling and habitat and physical features assessment. Habitat is also usually measured with multimetric indices, which often times are state-based. Examples are the Qualitative Habitat Evaluation

Indices (QHEI) in Ohio or the Physical Habitat Indices (PHI) in Maryland (Rankin, 1989, Hall Jr. et al., 1999, Paul et al., 2003). Even though a great variety of stream habitat indices and sampling methodologies exist (Kauffman et al., 1999, Lazorchak et al., 2000, Barbour et al.,1999), efforts have been made in unifying criteria and simplifying habitat quality evaluation with methodologies such as the *Rapid Bioassessment Protocols for Use in Wadeable Streams and Rivers* by Barbour et al. (1999).

Evaluation of the endpoint (biotic integrity) and the main stressors (habitat modification, impaired water quality etc.) should allow watershed managers identify priorities in order to make strategic decisions towards a better integrity of U.S. streams. Many times this decision goes through the improvement of not only water quality but the understanding that the fresh water systems are highly dimensional. Identification of those parameters whose improvement will yield a larger increment in integrity is paramount.

2. METHODOLOGY

The main objective was to identify the principal stressors affecting biotic integrity in each of the three states evaluated in the present report. Grouping the sites with similar types of stressors and compare these with the biotic integrity in each homogeneous group was important in order to identify those that showed a major influence. The stressor and biological indices data were available in different databases that are explained later in the present report. SOM were the tool used to identify the clusters or homogenous groups of stressors and they are explained in detail in the present report. Basically, SOM are a tool to organize highly dimensional data in homogeneous groups or clusters in which the data belonging to these groups are as similar as possible. The SOM were used in two different ways:

1.   **SOM followed by multiple range tests within clusters:** the SOM were run using all the chemical and physical environmental variables and habitat metrics. A number of optimum clusters was then found. Subsequently, the distributions among clusters of the available indices of biotic integrity (fish for Minnesota and Ohio and benthic for Maryland) were plotted and a multiple range test among clusters was performed to determine if the differences within the clusters were statistically significant. A 95% confidence interval was used. The different statistically significant homogeneous groups distribution was obtained. The same process was then repeated for each one of the variables used in the clustering process and the distribution of the homogeneous groups was then compared to the distribution of the biotic indices. Those metrics that showed equal or similar distributions were considered to be the most important for biotic integrity.

2.   **SOM neuron-analysis:** in this case we considered the neurons as the minimal, most homogeneous group of environmental values. In a SOM, one neuron groups a few sites with very similar characteristics. The values of each environmental variable and the biotic index in each neuron were averaged. The neuron-based environmental variables were then regressed against the neuron-based biotic index. Those variables with highest correlation were considered the most important for biotic integrity. Subsequently, we analyzed the relationships among different environmental variables, especially the relationships between

off-stream and in-stream habitat parameters as well as the relationships between physical variables and chemical quality values. This was done by a simple neuron-based regression among the different variables.

3. SOM AS CLUSTERING TOOL

SOM were an interesting tool for us because they are able to represent highly dimensional environmental vectors in a 2D plot with a meaningful order. SOM are composed of multiple units called cells or neurons in which each environmental vector corresponding to each sampled different site is placed after a weighting algorithm. SOM were first developed by Kohonen in 1984. They are considered a type of unsupervised Artificial Neural Network (ANN). The SOM consist of a topologically ordered mapping of the input space (in our case multiple environmental variables) onto a two-dimensional space according to a meaningful order (Kohonen, 2001). All the input parameters are located following a weighting algorithm onto different sites (called neurons or cells) on the map depending on the similarities (euclidean distances) with the neighboring cells. Therefore, similar groups of data or clusters are easily identified. SOM have been widely used in different fields such as speech recognition or economics, and are now being discovered as a great tool for environmental purposes (Brosse et al., 2001, Virani et al., 2005).

In a SOM, each vector in the input layer has a weighted connection with the neurons in the SOM. The euclidean distance between the SOM neurons with their initially assigned weights and the environmental vectors is calculated to find the most suitable or closest cell called the *Best Matching Unit (BMU)* using equation 1.

$$d_{ij} = \sum_{i=1}^{n} (x_i(t) - w_{ij}(t))^2$$

Equation 1. Euclidean distance calculation in the SOM

where $X_i(t)$ represents the environmental vectors and $W_{ij}(t)$ the neuron weights.

Once this initialization layout is obtained, the algorithm constantly updates the weights by comparing the values among neighboring cells to further reduce the distances among neurons until convergence is reached (Kohonen, 2001). These weights are usually known as codebook vectors. The training is usually performed in two phases: relatively large initial learning rates and neighborhood radius are used in the first phase to initiate the SOM. In the second phase, both learning rates and neighborhood radius are then initially small to achieve further fine-tuning of the SOM. In our case, the first tuning had 100 epochs and the fine tuning 20.

After the ordination process, the different clusters are obtained by observing the distances within the different neurons in the SOM. Small distances between two neurons or a group of neurons mean that they belong to a same cluster, while large distances may indicate a cluster separation. The *k-means* method in combination with the Davies-Bouldin Index was used to determine the number of clusters in the SOM. A detailed description of the *k-means* method and the Davies-

Bouldin Index can be found in Legendre and Legendre (1998) and Davies and Bouldin (1979) respectively.

Another issue in the SOM is determining the number of neurons. The optimal number of neurons was set by choosing the number that offered the minimum topographic and quantization errors (Kohonen,2001, Kiviluoto,1996).

4. DESCRIPTION OF THE DATABASES

1.   State of Ohio: the database used for our analysis from Ohio was obtained from the Ohio Environmental Protection Agency (Ohio EPA). It consisted of 1,848 sites with observations of in-stream habitat quality scores, water quality chemical parameters, invertebrate community indices (ICI), qualitative habitat evaluation indices (QHEI), fish indices of biotic integrity (IBI) and its respective metrics, and fish counts at each of these sites for more than 150 species. Other parameters such as stream immediate land use type, drainage area, longitude and latitude, as well as sampling dates and hydrologic unit codes (HUC) were also included. The observations in the Ohio database ranged from 1995 to 2000. Drainage area information was available at 1,328 sites and only 429 sites had all records with no blank values for any field. Table 4-1 describes each one of the fields in Ohio's database.

2.   State of Maryland: the database for Maryland was obtained from the Maryland Biological Stream Survey (MBSS). It consisted of 955 observations of some in-stream chemical parameters, in-stream habitat quality scores, indices of biotic integrity (fish, benthic and Hilsenhoff indices), some stream morphology parameters as well as land use percentages in the drainage area, drainage area, and fish and fish species' counts were also available. Other parameters also present were latitude and longitude, type of strata (piedmont, coastal or highland), dates of sampling, ecorregion, HUC, and basin name. The observations ranged from 1995 to 1997. A total of 905 sites had all the records for every field. Table 4-3 shows the description of the fields available for Maryland.

3.   State of Minnesota: the database was obtained from Minnesota's Pollution Control Agency (MPCA) and consisted of 1,134 observations of some in-stream chemical parameters, in-stream habitat quality scores, detailed information about percentages of substrate types, some stream morphology parameters, percentage of disturbed and undisturbed land uses in the buffer area (30 meters from the stream) and beyond the buffer area (from 30 to 100 meters from the stream). Fish IBI and QHEI were also present as well as latitude, longitude, drainage area, HUC, and dates of sampling which ranged from 1990 to 2006. The problem with the Minnesota database was that not many observations had values for all the fields. A total of 404 had in-stream habitat scores, 272 sites had observations for in-stream habitat scores and morphology/substrate quality, 167 had habitat scores and IBI observations, and only 91 had all the records for every field. Table 4-2 shows the description of the fields available for Minnesota.

TYPE OF DATA	NAME	DESCRIPTION	UNITS
CHEMICAL PARAMETERS	TEMPERATURE	Water temperature	Degrees centigrade
	CONDUCTIVITY	Water conductivity	
	DO	Dissolved Oxygen	mg/L
	BOD	Biological Oxygen Demand	mg/L
	PH	Water pH	Standard units
	TSS	Total Suspended Solids	mg/L
	AMMONIA	Ammonia in water	mg/L
	NITRITE	Nitrite in water	mg/L
	TKN	Total Kjeldahl Nitgrogen	mg/L
	NITRATE	Nitrate in water	mg/L
	PHOSPHORUS	Total Phosphorus in water	mg/L
	HARDNESS	Hardness in water	ppm
	CALCIUM	Dssolved calcium	mg/L
	MAGNESIUM	Dissolved Magnesium	mg/L
	CHLORIDE	Dissolved Chloride	mg/L
	SULFATE	Dissolved Sulfate	mg/L
	ARSENIC	Dissolved Arsenic	mg/L
	CD	Dissolved Cadmium	mg/L
	CU	Dissolved Copper	mg/L
	IRON	Dissolved Iron	mg/L
	PB	Dissoved Lead	mg/L
	ZN	Dissolved Zinc	mg/L
PHYSICAL/HABITAT PARAMETERS	SUBSTRATE	Substrate quality and type	Score from 0 to 20
	EMBEDDED	Degree to which the parent material is covered by fine sediment	Scale from 0 to 4
	COVER	Amount and type of stream vegetal cover	Score from 0 to 20
	CHANNEL	Quality of the stream with regard to creation and stability of macrohabitat	Score from 0 to 20
	RIPARIAN	Riparian zone width and type of vegetation and bank erosion	Score from 0 to 10
	POOL	Maximum depth of pool and type and morphology	Score form 0 to 12
	RIFFLE	Riffle depth, stability and embeddedness	Score from 0 to 8
	GRADIENT_S	Elevation drop through the sampling area	Score from 0 to 10
	PER_AG	Percentage of agriculture in buffer area	0,25,50,75 or 100%
	PER_FORWET	Percentage of forest and/or wetlands in buffer area	0,25,50,75 or 100%
	PER_URBDEV	Percentage of urban/developed in buffer area	0,25,50,75 or 100%
	AREA	Drainage area of the site	Square miles

Table 4-1. Environmental variable description in the Ohio database

TYPE	NAME	DESCRIPTION	UNITS
Habitat metrics and physical parameters	**Score Riparian**	QHEI metric	0 to 15
	MbufferWidth	Buffer width	Meters
	MBankEros	Bank Erosion	Percentage
	Score Substrate	QHEI metric	0 to 27
	PctEmbed	Embeddedness	Percentage
	Fines depth	Mean depth of fines	Cm
	PctRock	% of coarse substrates in transect	Percent
	PctBoulder	% of cover made of boulders	Percent
	Pctfine	% of fine substrate in transect	Percent
	PctPoolRun	% of reach that's pool and run	Percent
	PctRiffle	% of reach that is riffle	Percent
	PctRun	% of reach that is run	Percent
	PctPool	% of reach that is pool	Percent
	Score Cover	QHEI metric	0 to 17
	PctEmerMac	% of cover that is emergent macrophytes	Percent
	PctSubMac	% of cover that is submerged macrophytes	Percent
	PctWoody	% of cover that are woody elements	Percent
	PctOverVeg	% of cover that is overhanging vegetation	Percent
	PctOtherCov	% of cover that is other cover	Percent
	PctUnderCut	% of cover that is undercut	Percent
	PctCover	% cover for fish	Percent
	Score Channel	QHEI metric	0 to 36
	MWidth	Mean width	Meters
	MthalDepth	Maximum thalweg depth	Cm
	MDepth	Mean water depth at transect points	Cm
	Sinuosity	Ratio between stream length and straight distance	Ratio
	WDRatio	Width-depth ratio	Ratio
	Score Land use	QHEI metric	0 to 5
	PctDistLU	% disturbed land use in DA	Percentage
	PctUnDistLU	% undisturbed land use in DA	Percentage
	PctDistLU30	% disturbed land use in 30-meter buffer	Percentage
	PctUnDistLU30	% undisturbed land use in 30-meter buffer	Percentage
	DA	Drainage area	Sq. miles
	Gradient	Site slope	m/Km
Chemical parameters	Cond	Specific Conductance	
	DO	Dissolved oxygen	mg/L
	NH4	Ammonia	mg/L
	Nitrogen	Total nitrogen	mg/L
	pH	pH	Standard Units
	Phosphorus	Total phosphorus	mg/L
	Temp	Temperature	Degrees Celsius
	TSS	Total Suspended Solids	mg/L
	Turbid	Turbidity	

Table 4-2. Physical and chemical environmental variables used for clustering in the Minnesota database

TYPE OF DATA	NAME	DESCRIPTION	UNITS*
PHYSICAL, HABITAT AND MORPHOLOGIC PARAMETERS	Remoteness (REMOTE)	Rate based on the absence of human activity and difficulty of access	Score from 0 to 20*
	Shading (SHADING)	Rate based on estimates of the degree and duration of shading during the summer	Percentage *
	Epifaunal Substrate (EPI_SUB)	Amount of variety of hard, stable substrates usable by benthic macroinvertebrates	Score from 0 to 20*
	Instream habitat (INSTRHAB)	Perceived value of habitat to the fish community	Score from 0 to 20*
	Woody elements (WOOD)	Number of woody debris and rootwads in the control site	Number*
	Bank Stability (BANKSTAB)	Presence/absence of riparian vegetation and other stabilizing bank materials.	Score from 0 to 20*
	Velocity-depth diversity (VEL_DPTH)	Variety of velocity-depth regimes present at the site	Score from 0 to 20*
	Pool quality (POOLQUAL)	Variety and spatial complexity of slow or still water habitat	Score from 0 to 20*
	Riffle Quality (RIFFQUAL)	Depth, complexity and functional importance of riffle/run habitat	Score from 0 to 20*
	Channel alteration (CHAN_ALT)	Measure of large scale changes in the shape of the stream channel	Score from 0 to 20*
	Embeddedness (EMBEDDED)	Fraction of surface area of larger particles surrounded by fine sediment	Percentage*
	Channel Flow Status (CH_FLOW)	Fraction of the area of the stream that is covered by water	Percentage
	Aesthetics (AESTHET)	Visual appeal of the site and presence/absence of human refuse	Score from 0 to 20*
	Max. depth (MAXDEPTH)	Maximum depth at the site	Centimeters
	Riparian buffer width (RIP_WID)	Width of the riparian strip along the stream	Meters*
	Gradient (ST_GRAD)	Stream gradient	Percentage
	Average width (AVGWID)	Average wetted width	Meters
	Average thalweg (AVGTHAL)	Average thalweg depth	Centimeters
	Average velocity (AVG_VEL)	Average velocity	Meters per second
	Urban land use (URBAN)	Fraction of urban land use in drainage area	Percentage
	Forest, wetland, water land uses (FORWETWAT)	Fraction of unimpacted land uses in drainage area	Percentage
	Agricultural and barren land uses (AGRIBARR)	Fraction of agriculture/bare soil in drainage area	Percentage
	Drainage area (ACREAGE)	Catchment area at the site	Acres

TYPE OF DATA	NAME	DESCRIPTION	UNITS
CHEMICAL PARAMETERS	Temperature_FLD (TEMP-FLD)	Water temperature	Degrees Celsius
	Dissolved Oxygen_FLD (DO_FLD)	Dissolved oxygen	ppm
	pH_FLD (PH_FLD)	pH in summer time	Standard units
	Conductance_FLD (COND_FLD)	Specific conductance in summer time	μmho/cm
	Dissolved Organic carbon (DOC_LAB)	Dissolved organic carbon concentration	mg/L
	Nitrate (NO3_LAB)	Nitrate-Nitrogen concentration	mg/L
	Sulfate (SO4_LAB)	Sulfate concentration	mg/L

Table 4-3. Description of the environmental variables included in the MBSS database.

*The scoring system shown in the table corresponds to the old PHI. The scores for the new metrics were calculated with the guidelines from Paul et al. (2003) using the metrics in the old PHI

5. RESULTS I: SOM AND MULTIPLE RANGE TESTS

The environmental vectors available in the databases were used to find sets with similar characteristics. The clustering procedure was performed using all chemical and physical environmental variables. Subsequently, the biotic integrity indices and the environmental variables distribution within the clusters were plotted. A comparison between the distributions of the metrics and the biotic indices was performed in order to distinguish the most important metrics affecting biotic integrity. Multiple range tests were used to identify statistically significant differences within the cluster means for the biotic and habitat indices and each one of the environmental variables and metrics. Those that followed the same or very similar distribution than the biotic indices were considered as the variables having the greatest impact in the biotic community.

5.1. OHIO

i. Clustering the database

The metrics used for clustering in this case are summarized in Table 4-1. As stated before in the database description, a total of 429 sites had values for each field. For this case the optimum number of clusters determined by the Davies-Bolduin index was three. Even though the absolute minimum was obtained for seven clusters, we decided to choose three because it was easier for the sake of data interpretation and understanding. The SOM used in this case had a total of sixty neurons or cells.

Figure 5-1. Optimum number of clusters. Ohio, all environmental variables

ii. Habitat and biotic indices cluster distribution and analysis

The distribution of the habitat and biological indices using all the variables are as follows (in box plots, top line means 75th percentile, red line is 50th percentile and bottom line is 25th percentile).
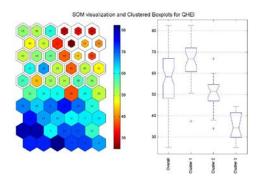
Figure 5-2. QHEI distribution among clusters
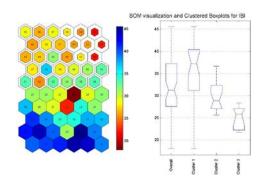
Figure 5-3. Fish IBI distribution among clusters
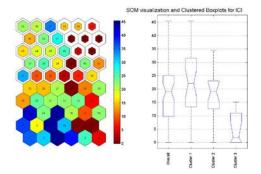
Figure 5-4. ICI distribution among clusters

The MRT tests were run to determine if the means' differences within the three clusters were statistically significant. Three homogeneous groups were found corresponding to each cluster. The MRT tests homogeneous groups are shown as follows:

Fish IBI:

```
-----------------------------------------------------------------------------
---
Method: 95.0 percent LSD
              Count     Mean              Homogeneous Groups
-----------------------------------------------------------------------------
---
IBI3          53        24.4528           X
IBI2          145       29.6966            X
IBI1          231       36.8658             X
-----------------------------------------------------------------------------
---
Contrast                                  Difference        +/-  Limits
-----------------------------------------------------------------------------
---
IBI1 - IBI2                               *7.16925          1.79754
IBI1 - IBI3                               *12.413           2.58398
IBI2 - IBI3                               *5.24372          2.72323
```
Table 5-1. Fish IBI homogeneous groups within Ohio's clusters

ICI:

```
-----------------------------------------------------------------------------
---
Method: 95.0 percent LSD
              Count     Mean              Homogeneous Groups
-----------------------------------------------------------------------------
---
ICI3          53        6.37736           X
ICI2          145       18.4552            X
ICI1          231       24.3896             X
-----------------------------------------------------------------------------
---
Contrast                                  Difference        +/-  Limits
-----------------------------------------------------------------------------
---
ICI1 - ICI2                               *5.93444          4.17147
ICI1 - ICI3                               *18.0123          5.99654
ICI2 - ICI3                               *12.0778          6.3197
-----------------------------------------------------------------------------
---
```
Table 5-2. ICI homogeneous groups within Ohio's clusters

QHEI:
```
------------------------------------------------------------------------------
---
Method: 95.0 percent LSD
             Count     Mean           Homogeneous Groups
------------------------------------------------------------------------------
---
QHEI3          53       29.0189        X
QHEI2         145       48.3155         X
QHEI1         231       71.2915          X
------------------------------------------------------------------------------
---
Contrast                                 Difference        +/-  Limits
------------------------------------------------------------------------------
---
QHEI1 - QHEI2                             *22.976           1.85745
QHEI1 - QHEI3                             *42.2726          2.67011
QHEI2 - QHEI3                             *19.2966          2.814
------------------------------------------------------------------------------
---
```
* denotes a statistically significant difference.

Table 5-3. QHEI homogeneous groups within Ohio's clusters

iii. Environmental variable cluster distribution and analysis

Land use and riparian area cluster distribution
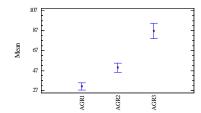
Figure 5-5. **Agriculture LU means distribution**

Figure 5-6. **Riparian score means distribution**

Figure 5-7. **Forest/wetlands LU means distrib.**

In-stream physical environmental variables with the same cluster distribution
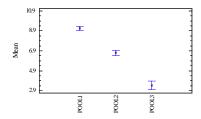
Figure 5-8. **Pool score means distribution**

Figure 5-9. **Channel score means distribution**

Figure 5-10. **Cover score means distribution**
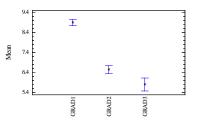
Figure 5-11. **Gradient score means distribution**

Figure 5-12. **Riffle score means distribution**
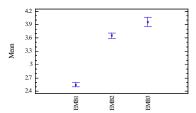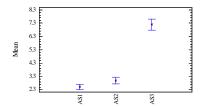
Figure 5-13. **Embeddedness score means distrib.**

Figure 5-14. Substrate score means distribution

Chemical parameters with the same cluster distribution
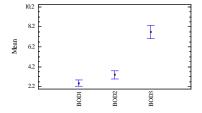
Figure 5-15. **Arsenic means distribution**

Figure 5-16. **BOD means distribution**

Other metrics showed statistically significant differences in only one of the clusters. This is the case of iron , TSS, and nitrate in cluster 1, which were lower (TSS and iron) or higher (nitrate). Also most of the chemical parameters had higher values in cluster 3: ammonia, calcium, chloride, hardness, magnesium, phosphorus, sulfate, TKN, and conductivity. The parameters different in cluster 2 were pH and temperature (lower values).

Other metrics didn't show any differences among clusters: drainage area, cadmium, copper, nitrite, lead, percent of urban/developed, and zinc.

5.2. MARYLAND

iv. Clustering the database

The SOM were run for each of the strata that MBSS has determined: coastal, piedmont and highland areas. In each of these strata the clustering was performed using all the environmental variables available in the database and described in Table 4-3. The habitat metrics that are used to calculate the PHI in each strata in the new PHI developed by Paul et al.(2003) were used. The rest of the physical variables and old habitat metrics were also used in the clustering but the original values were kept. In each of the three strata, five clusters were found.
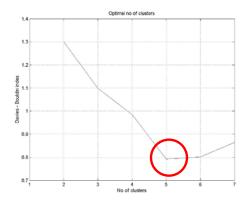
Figure 5-17. **Number of clusters used in coastal sites**
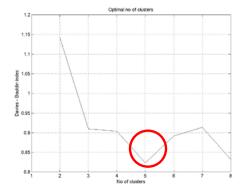
Figure 5-18. **Number of clusters in piedmont sites**
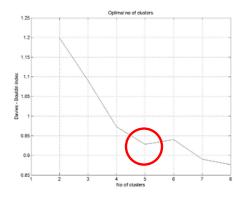
Figure 5-19. Number of clusters in highland sites

v. Habitat and biotic indices cluster distribution and analysis

The new PHI along with the fish and benthic IBI distributions were plotted and the differences among clusters were studied. Significantly different distributions between fish and benthic IBI were found. The differences are, most likely, due to a bias that exists in the fish IBI with watershed size as reported by Southerland et al. (2005). Therefore, the benthic IBI distribution was used as the reference in the environmental variable analysis.

Coastal sites

The following plots show the means distribution within clusters found with the SOM.
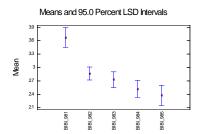
Figure 5-20. **Benthic IBI means distribution**

Figure 5-21. **Fish IBI means distribution**

Figure 5-22. **Hilsenhoff index means distribution**

Figure 5-23. **PHI means distribution**

The MRT are shown as follows:

		CL1	CL2	CL3	CL4	CL5
FIBI	**LEVEL1**					
	LEVEL2	Y				
	LEVEL3		Y	Y		
	LEVEL4				Y	Y
	LEVEL5					
BIBI	**LEVEL1**					
	LEVEL2	Y				
	LEVEL3		Y	YY		
	LEVEL4			YY	Y	Y
	LEVEL5					
HILSHOFF	**LEVEL1**					
	LEVEL2			Y		Y
	LEVEL3	Y	Y		Y	
	LEVEL4					
	LEVEL5					
PHI	**LEVEL1**					
	LEVEL2	Y				
	LEVEL3		Y		Y	YY
	LEVEL4			Y		YY
	LEVEL5					

Table 5-4. Multiple Range Test for the biotic indices and PHI. Non overlapping Ys mean significant difference and viceversa. Level means homogeneous groups.

Piedmont areas

The biotic and PHI means cluster distributions is shown as follows
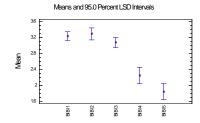
Figure 5-24. Benthic IBI means distribution

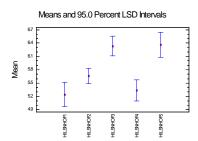**Figure 5-25. Fish IBI means distribution**
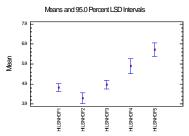
Figure 5-26. Hilsenhoff index means distribution

Figure 5-27. PHI means distribution

The MRT are as follows:

		CL1	CL2	CL3	CL4	CL5
FIBI	**LEVEL1**					
	LEVEL2	Y		Y		
	LEVEL3		Y			
	LEVEL4				Y	Y
	LEVEL5					
BIBI	**LEVEL1**					
	LEVEL2	Y	Y	Y		
	LEVEL3				Y	Y
	LEVEL4					
	LEVEL5					
HILSHOFF	**LEVEL1**					Y
	LEVEL2				Y	
	LEVEL3	Y		Y		
	LEVEL4		Y			
	LEVEL5					
PHI	**LEVEL1**					
	LEVEL2	Y	YY			
	LEVEL3		YY	Y	Y	
	LEVEL4					Y
	LEVEL5					

Table 5-5. Multiple Range Test for the biotic indices and PHI. Non overlapping Ys mean significant difference and viceversa. Level means homogeneous groups.

Highland sites

Figure 5-28. Benthic IBI means distribution
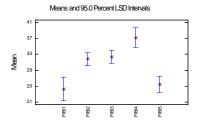
Figure 5-29. Fish IBI means distribution
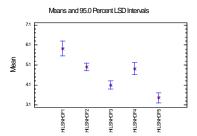
Figure 5-30. Hilsenhoff index means distribution
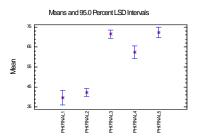
Figure 5-31. PHI means distribution
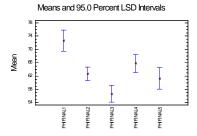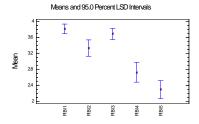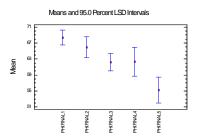
The MRT in highland sites are as follows:

		CL1	CL2	CL3	CL4	CL5
FIBI	LEVEL1					
	LEVEL2				Y	
	LEVEL3		Y	Y		
	LEVEL4	Y				Y
	LEVEL5					
BIBI	LEVEL1					
	LEVEL2			Y		Y
	LEVEL3	Y	Y		Y	
	LEVEL4					
	LEVEL5					
HILSHOFF	LEVEL1	Y				
	LEVEL2		Y		Y	
	LEVEL3			Y		
	LEVEL4					Y
	LEVEL5					
PHI	LEVEL1					
	LEVEL2			Y		Y
	LEVEL3				Y	
	LEVEL4	Y	Y			
	LEVEL5					

Table 5-6. Multiple Range Test for the biotic indices and PHI. Non overlapping Ys mean significant difference and viceversa. Level means homogeneous groups

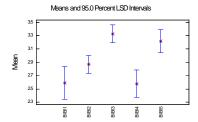### vi. Environmental variable cluster distribution and analysis

Coastal sites

Land use and riparian quality cluster distributions
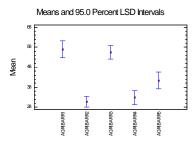
Figure 5-32. Agriculture LU means distribution

Figure 5-33. Forest/wetlands LU means distrib.

Figure 5-34. Urban LU means distribution

Figure 5-35. Riparian score means cluster distrib.

In-stream physical parameters with similar distributions as the BIBI or Hilsenhoff index

Figure 5-36. Average velocity means distrib.

Figure 5-37. Channel flow means distribution

Figure 5-38. Embeddedness means dist.
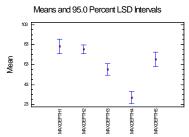
Figure 5-39.. Max. depth means distribution
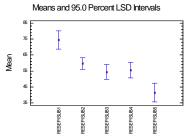
Figure 5-40. **Epif. Substrate means dist.**

Figure 5-41. **Pool quality means distribution**

Figure 5-42. **Riffle quality means dist.**
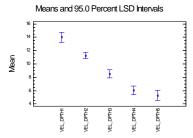
Figure 5-43. **Vel-depth diversity means distrib.**

Chemical parameters with similar distributions as the BIBI or Hilsenhoff index

DOC, and concentrations had similar patterns to the Hilsenhoff index.
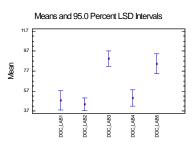
Figure 5-44. **DOC means distribution**
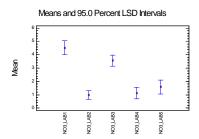
Figure 5-45. **Nitrate means distribution**

Other chemical variables such as DO or pH were significantly smaller in cluster 5, which might indicate that this cluster is heavily impacted by impaired water quality

Piedmont sites

Land use and riparian quality cluster distributions
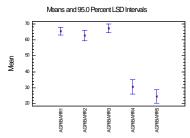
Figure 5-46. **Agriculture LU means dist.**
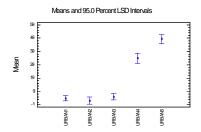
Figure 5-47. **Urban LU means dist.**
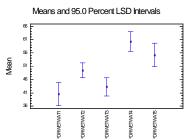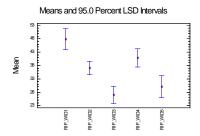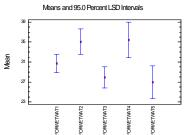
Figure 5-48. **Forest/wetlands LU means dist.**
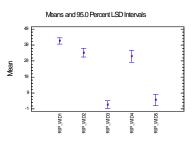
Figure 5-49. **Riparian quality means dist.**

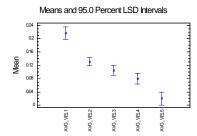## In-stream physical parameters with similar distributions as the BIBI or Hilsenhoff index

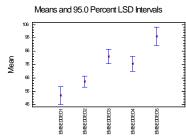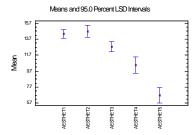**Figure** 5-50. **Aesthetic quality means distrib.**
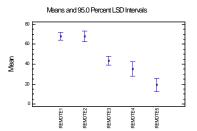
Figure 5-51. **Remoteness means distribution**

Chemical parameters with similar distributions as the BIBI or Hilsenhoff index
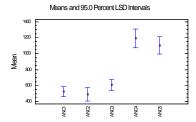
Figure 5-52. **ANC means cluster distribution**
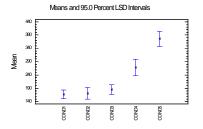
Figure 5-53. **Cond. means cluster distribution**
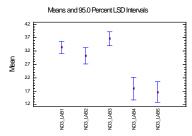
Figure 5-54. **Nitrate means cluster distribution**
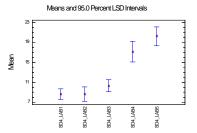
Figure 5-55. **Sulfate means cluster distribution**
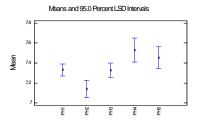
Figure 5-56. **pH means cluster distribution**

Other chemical parameters such as DO or DOC showed only significant higher levels in cluster 5

Highland sites

Land use and riparian quality cluster distributions
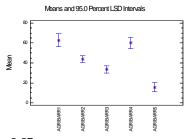
Figure 5-57. **Agriculture LU means distrib.**

Figure 5-58. **Forest/wetland LU means distrib.**

Figure 5-59. **Urban LU means distribution**
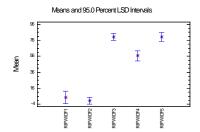
Figure 5-60. **Riparian score means distribution**

In-stream physical parameters with similar distributions as the BIBI or Hilsenhoff index
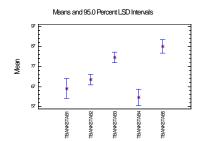
Figure 5-61. **Bank stability means cluster dist.**
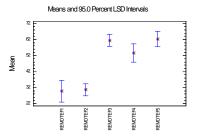
Figure 5-62. **Remoteness means cluster distrib.**
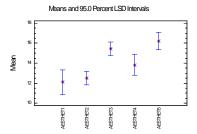
Figure 5-63. **Aesthetic quality cluster distrib.**

Chemical parameters with similar distributions as the BIBI or Hilsenhoff index
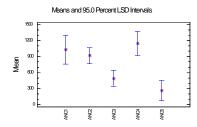
Figure 5-64. **ANC means cluster distribution**

Figure 5-65. **Conductivity cluster distribution**

Figure 5-66. **Nitrate means cluster distribution**
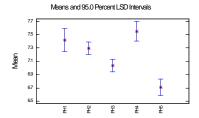
Figure 5-67. **pH means cluster distribution**

Figure 5-68. **Temperature means cluster distribution**

5.3. Minnesota

vii.   Clustering the database

When the clustering with the forty-three variables was performed three clusters were found as shown in the Davies-Boldwin index. Even though the index didn't converge to a minimum value after 5 clusters or more, a local minimum in three clusters was observed and chosen as optimum for data interpretation.
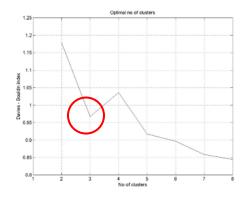
Figure 5-69. Number of clusters in Minnesota´s database

viii.   Habitat and biotic indices cluster distribution and analysis

The fish IBI as well as the QHEI used in Minnesota showed the following distributions among clusters
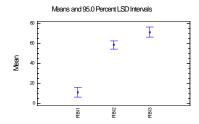
Figure 5-70**. Fish IBI means cluster distribution**
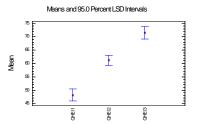
Figure 5-71**. QHEI means cluster distribution**

The homogeneous groups found with the MRT analysis are shown as follows

		CL1	CL2	CL3
FIBI	**LEVEL1**			
	LEVEL2			**Y**
	LEVEL3		**Y**	
	LEVEL4	**Y**		
	LEVEL5			
QHEI	**LEVEL1**			
	LEVEL2			**Y**
	LEVEL3		**Y**	
	LEVEL4	**Y**		
	LEVEL5			

Table 5-7**. Multiple Range Test for the biotic indices and QHEI. Non overlapping Ys mean significant difference and viceversa. Level means homogeneous groups.**

ix. Environmental variable cluster distribution and analysis

Land use and riparian quality cluster distributions
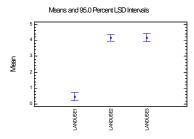
Figure 5-72. **QHEI´s land use score means cluster dist.**

Figure 5-73. **% disturbed LU in DA distrib.**

Figure 5-74. **% disturbed LU in 30-meter buffer dist.**
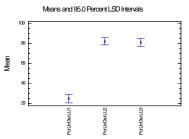
Figure 5-75. **% undisturbed LU in DA dist.**

Figure 5-76. **% undisturbed LU in 30-meter buffer dist.**
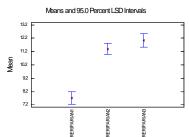
Figure 5-77. **QHEI´s riparian score distribution**

Figure 5-78. **Mean buffer width cluster distribution**

In-stream physical parameters with similar distributions as the Fish IBI

Three out of the five metrics that comprise Minnesota´s QHEI showed distributions in which clusters 2 and 3 were nor significantly different (land use, riparian and cover scores). The differences in the final QHEI score between clusters 2 and 3 are due to substrate and channel

quality. The scores' cluster distributions and the distributions of some of these metrics raw measurements are shown below.
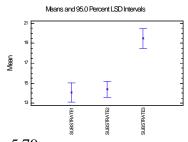
Figure 5-79. **QHEI´S substrate score distribution**
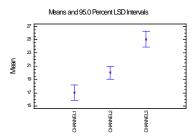
Figure 5-80. **QHEI's channel score distribution**

Morphologic parameters
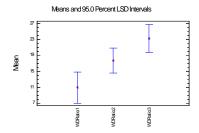
Figure 5-81. **Width-depth ratio cluster dist.**
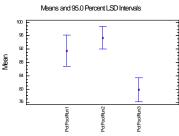
Figure 5-82. **% pool-run cluster distribution**
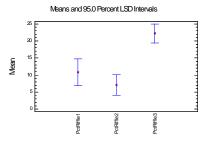
Figure 5-83. **% riffle cluster distribution**
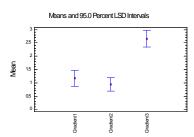
Figure 5-84. **Stream gradient cluster distrib.**

Substrate parameters
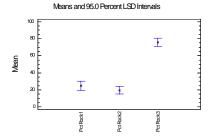
Figure 5-85. **% rock cluster distribution**
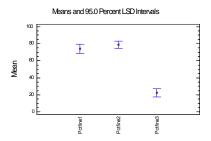
Figure 5-86. **% fines cluster distribution**

Chemical parameters with similar distributions as the Fish IBI

All the chemical showed significantly higher values in cluster 1 but not between clusters 2 and 3. Two examples of the pattern for the chemical variables are shown below.
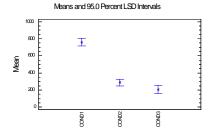
Figure 5-87. **Conductivity cluster dist.**
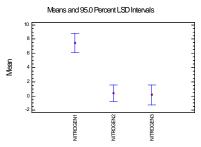
Figure 5-88. **Total nitrogen cluster distribution**

6. RESULTS II: NEURON-BASED ANALYSIS

6.1. Ohio

Effect of habitat on biotic integrity

The average value of each one of the SOM neurons was taken and each habitat parameter plotted versus the fish IBI. As identified by the SOM+MRT analysis, both substrate and morphologic parameters were the ones with better correlation with IBI.
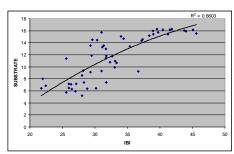
Figure 6-1. **IBI vs. substrate score**
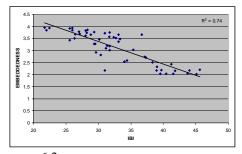
Figure 6-2. **IBI versus embeddedness**

Figure 6-3. **IBI vs channel**

Two landscape features seem to be responsible for the quality of the habitat in each site: riparian quality and gradient. While gradient has a deep impact on substrate quality (i.e. impounded areas), the riparian buffer has a linear impact on cover, channel quality and pool quality and a non linear impact in the amount of fine sediment reaching the receiving stream. Therefore, the substrate quality is a non-linear combination of riparian quality and gradient as shown in the following plots.
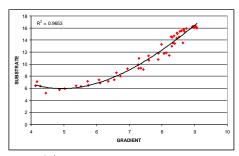
Figure 6-4. **Gradient vs. substrate score**

Figure 6-5. **Gradient vs. embeddedness**

Figure 6-6. **Gradient vs channel**
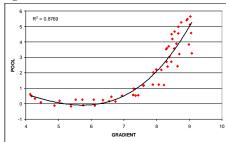
Figure 6-7. **Gradient vs pool score**

Figure 6-8. **Gradient vs. riffle score**
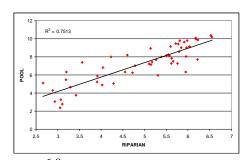
Figure 6-9. **Riparian quality vs. pool quality**

Figure 6-10. **Riparian quality vs. channel**
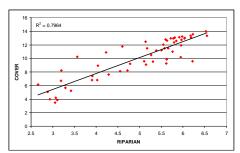
Figure 6-11. **Riparian quality vs. cover**
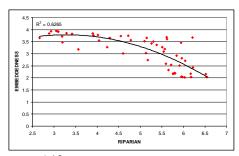
Figure 6-12. **Riparian quality vs. embeddedness**

Even though great variability exists in natural environments, the overall trends shows how general morphologic parameters (gradient) and immediate off-stream land use (riparian area) drive in-stream habitat quality. Substrate quality is a non-linear combination of both, gradient and riparian quality. Streams in impounded areas or agricultural lands are usually slow moving waters in which fine sediment can deposit. Habitat quality will decline steeply with gradient scores lower than seven, which seems to be the threshold for fine sediment deposition. Elimination of impoundments or structures that can increase the gradient will have a great effect on the stream substrate if a gradient score greater than 7 can be achieved. Also, any improvement in the riparian buffer will decrease importantly the amount of sediment entering the stream and improve linearly the pool, channel, cover quality.

Effect of water quality over biotic integrity

As successfully identified by the SOM+MRT, Arsenic and BOD showed the highest correlation with biotic integrity. Also ammonia (and therefore TKN) showed a good relationship with fish IBI in the neuron-based regressions.
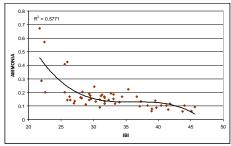
Figure 6-13. **IBI vs ammonia**

Figure 6-14. **IBI vs BOD**
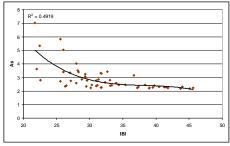
Figure 6-15. **IBI vs. arsenic**
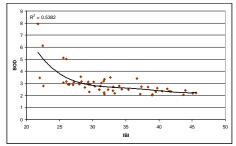
The relationships between riparian quality and water quality are not straightforward due to the potential presence of point sources such as Waste Water Treatment Plants (WWTP) or local discharges. The high correlation between water quality and riparian quality might mean that Ohio is mainly impaired by non-point sources, which are related to habitat degradation.

6.2. Maryland
i. Coastal areas

Effect of habitat on biotic integrity

The results in Maryland's coastal areas weren't as good as the ones in Ohio. However, it is remarkable that the metrics that showed better relationships with biotic integrity were almost the same. Gradient didn't seem to have a direct effect on biotic integrity but a very strong indirect effect by affecting the substrate and channel morphology.

The riparian width didn't have an important correlation with biotic integrity either. However, it showed important correlations with the parameters that are included in the habitat index such as in-stream habitat quality or remoteness as well as other variables not included in the PHI such as the degree of embeddedness.

The environmental variables and metrics that showed better correlations with Benthic IBI were again those identified by the SOM+MRT and are as follows:
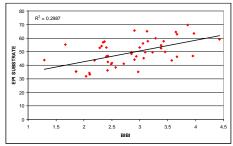
Figure 6-16. **BIBI vs. epifaunal substrate**

Figure 6-17. **BIBI vs riffle quality**

Figure 6-18. **BIBI vs avg. velocity**
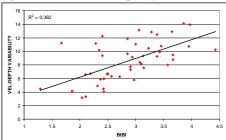
Figure 6-19.**BIBI vs vel-depth variability**
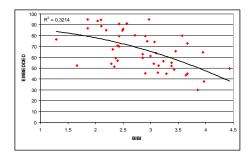
Figure 6-20. BIBI vs. embeddedness

It should be noticed that only one out of six metrics included in the coastal PHI showed some degree of correlation with biotic integrity. The correlations of gradient and riparian width over substrate, channel and other physical parameters are shown as follows. Other strong relationships were found but the plots were not included because they were straightforward such as riparian width versus remoteness.
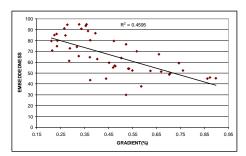
Figure 6-21. Gradient vs. embeddedness

Figure 6-22. Gradient vs. epifaunal substrate

Figure 6-23. Gradient vs.channel quality
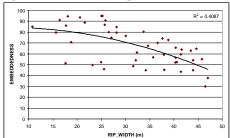
Figure 6-24. Riparian width vs embeddedness

Effect of water quality over biotic integrity
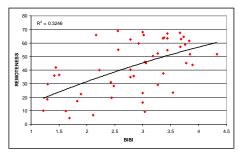
None of the chemical parameters measured showed a strong correlation with either the benthic or fish IBI or the Hilsenhoff index.

ii. Piedmont areas

Effect of habitat on biotic integrity

Maryland's piedmont areas are usually highly urbanized or developed lands. Even though habitat is not necessarily impaired (mean PHI Is 66 out of 100) water quality is the main responsible for

biotic integrity impairment. The correlation between habitat and benthic IBI is very small and only the two metrics identified in the SOM+MRT showed some degree of correlation. Remoteness and aesthetic quality were the selected physical metrics and enhance the idea that water quality is the main responsible for impaired integrity. More remote or "more beautiful" locations are less likely to have population nearby that could send pollution towards the receiving stream. The rest of the metrics as well as the PHI didn't seem to have a clear effect on benthic communities.
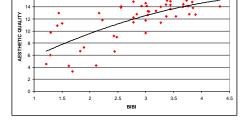
Figure 6-25. **BIBI vs. remoteness**

Figure 6-26. **BIBI vs. aesthetic quality**

In this case, stream gradient only had a negative effect over pool quality and velocity-depth variability. However, these two metrics didn't seem to affect biotic quality at all.

Interestingly, the presence of agricultural land in piedmont areas seems to have a positive effect over the benthic community as opposed to coastal areas. This might suggest that urban impairment is heavier or more acute than that caused by agriculture. The following plot shows this relationship:
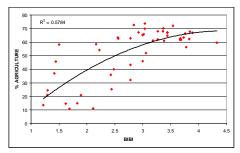
Figure 6-27. **BIBI vs. percent of agriculture**

Effect of water quality on biotic integrity

The SOM+MRT identified five water quality parameters as most disciminant. The SOM neuron analysis confirmed this and showed a strong correlation between water quality and benthic community, confirming that piedmont sites integrity is mainly impaired by water quality issues and not so much by habitat degradation.
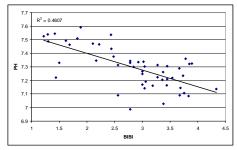
Figure 6-28. **BIBI vs pH**

Figure 6-29. **BIBI vs. sulfate**
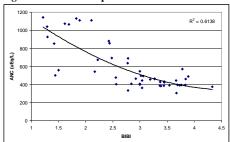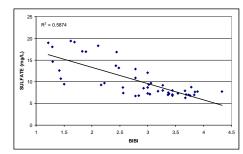
Figure 6-30. **BIBI vs. ANC**
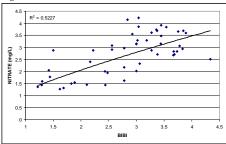
Figure 6-31. **BIBI vs nitrate**

Apparently, the two selected physical metrics, and especially aesthetic rating (which accounts for presence/absence of human refuse and state of the riparian area) are very good indicators of water quality in piedmont sites. The following plots are just two examples.
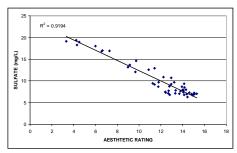
Figure 6-32. **Aesthetic rating vs. sulfate**
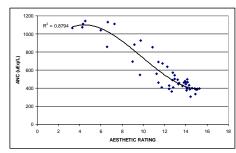
Figure 6-33. **Aesthetic rating vs. ANC**
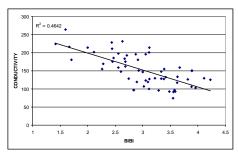
iii. Highland areas

Effect of habitat on biotic integrity

The effect of habitat in highland sites remains unclear. Even though the variables selected with the SOM+MRT showed promising distributions, their correlation with the benthic IBI was very poor (maximum $r^2 = 0.28$ for aesthetic rating). The correlation with land use was somewhat stronger in the case of urban land uses (negative correlation with $r^2 = 0.47$), which might suggest again that water quality problems are the main stressors in these areas.

Effect of water quality on biotic integrity

Even though there seems to be more correlation with chemical parameters, it wasn't as strong as expected. Conductivity and ANC were the parameters that showed higher correlations.

Figure 6-34. **BIBI vs. conductivity**
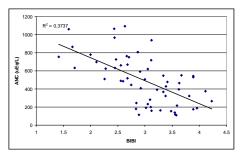
Figure 6-35. **BIBI vs. ANC**

In this case, aesthetic rating didn't have such a strong correlation with water quality parameters as it was observed in piedmont sites. Instead, riparian quality seemed to be important for water quality. Better riparian area meant less ANC and conductivity.

6.3.   Minnesota

Effect of habitat on biotic integrity

Even though the SOM+MRT analysis showed clearly the relationship between habitat and biotic integrity, the results of the neuron-based regressions weren't as good as it would be expected. As discussed previously in the present report, the habitat index successfully separates sites with high degree of impairment (cluster 1) and accounts for habitat quality differences due to substrate and morphologic parameters (clusters 2 and 3). The means distribution of the actual measurements (i.e percent fines or percent riffle) showed how in pristine or semi-pristine streams (clusters 2 and 3), substrate and morphologic differences accounted for the most part of the differences in biotic integrity. This was also clearly seen in Ohio and in Maryland's coastal areas. The relationships between substrate and morphology and gradient and riparian quality were also clear. In the case of Minnesota, the neuron-based analysis did not yield a very strong relationship between the habitat index and biotic integrity ($r^2 = 0.39$). The following plots show the metrics that had strongest correlations with the fish IBI.
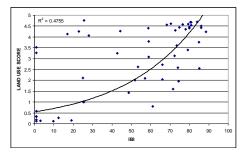
Figure 6-36. **IBI vs. land use quality**
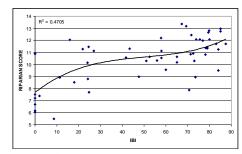
Figure 6-37. **IBI vs. riparian quality**
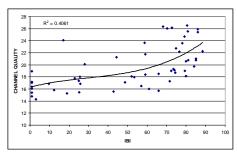
Figure 6-38. **IBI vs. channel quality**

Again, gradient and riparian quality seem to be the drivers of most of the metrics or parameters included in the metrics as shown in the following plots.
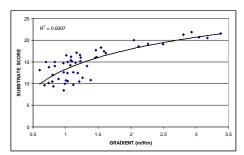
Figure 6-39. **Gradient vs substrate score**

Figure 6-40. **Gradient vs channel score**

Figure 6-41. **Gradient vs type of substrate**
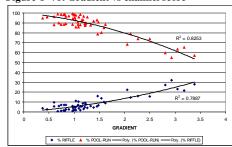
Figure 6-42. **Gradient vs percent of riffle/pools**

Figure 6-43. **Buffer width (up to 10 meters) vs cover**
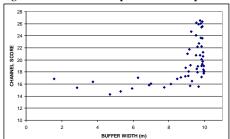
Figure 6-44. **Buffer width (up to 10m) vs channel**

Figure 6-45. **Buffer width (10m) vs percent pool/run**
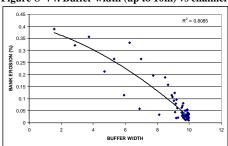
Figure 6-46. **Buffer width vs bank erosion**

Effect of water quality on biotic integrity

As shown in the chemical parameters distribution by the SOM, water quality has an important effect in Minnesota. The most affected sites are those located in cluster 1.
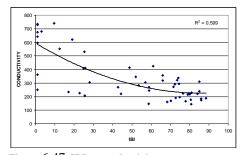
Figure 6-47. **IBI vs conductivity**

Figure 6-48. **IBI vs total nitrogen**

Figure 6-49. IBI vs phosphorus
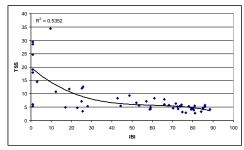
Figure 6-50. **IBI vs TSS**

Surprisingly, riparian width seemed to have an extremely important effect on water quality. Also, other morphologic parameters showed an important correlation towards some of the chemical parameters, which could indicate that habitat features and diversity are not only important because they affect stream's fauna but also have a role in water chemistry.
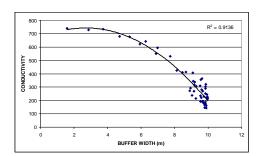
Figure 6-51. **Buffer width vs. conductivity**

Figure 6-52. **Buffer width vs. total nitrogen**

Figure 6-53. Buffer width vs phosphorus
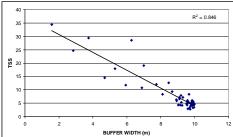
Figure 6-54. **Buffer width vs TSS**
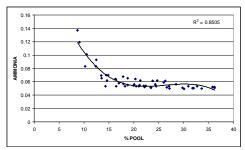
Figure 6-55. **% pool vs ammonia**
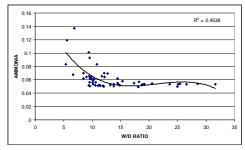
Figure 6-56. **Width-depth ratio vs ammonia**

7. CONCLUSIONS

SOM were an extremely useful tool in identifying sites with similar environmental stressors and were successful in revealing some of the very convoluted relationships among physical and chemical stressors and biotic integrity or among the physical and chemical stressors themselves. The clustering performed by the SOM followed by an analysis of the significant differences among clusters using Multiple Range Tests, and the subsequent comparison between biological and stressors' distributions, proved to be highly effective and successfully identified the variables that play a key role in biotic integrity, as proved in the SOM-neuron analysis.

In all three states, either with the SOM+MRT analysis, the SOM-neuron analysis, or both, it was found that substrate and channel morphologic features are the two in-stream habitat parameters that have a deeper impact on biotic integrity. This was particularly clear in Ohio and Maryland's coastal sites. In Minnesota, the SOM+MRT analysis also identified substrate and channel morphologic parameters as responsible for the differences in biotic integrity found between clusters 2 and 3, which had very similar water quality. Sometimes, the relationship between habitat parameters and biotic integrity is not straightforward because biotic integrity not only reflects the effects of human activity over habitat, but also over chemical quality of the stream (i.e. a point source is not associated with habitat impairment but decreases water quality and biotic integrity).

In Ohio, the SOM showed this is a habitat-driven state. This means that the greatest cause for biotic integrity impairment comes from habitat degradation more than water quality issues. This doesn't mean that Ohio's waters are not facing water quality problems, but these problems are usually originated from non-point sources which are usually associated with land use and management practices that ultimately affect in-stream habitat as well. In fact, Ohio's cluster 3 shows the poorest water quality values in the state and is clearly associated with the poorest habitat scores and, therefore, biotic integrity. It is important to highlight that Ohio's land use is highly dominated by agriculture. In this state, the QHEI seemed to be highly effective in identifying sites with different habitat quality and the association between habitat and biotic integrity was very clear.

In Maryland, two different types of environmental responses were observed. On the one hand, coastal sites seemed to have a similar response to Ohio, on the other hand piedmont and highland sites were clearly dominated by water quality impairment. In coastal sites, habitat degradation is mainly caused by substrate and morphology-related parameters. Water quality impairment

wasn't clear in this case, since none of the parameters measured showed a strong correlation with biotic integrity in the SOM-neuron analysis. However, DOC and nitrate were the most relevant as shown by the SOM+MRT analysis, which agrees with Ohio's most significant water quality parameters: ammonia and BOD.

Piedmont and highland sites seemed to be mainly driven by water quality issues. This was particularly clear in the case of piedmont areas in which the most significant habitat parameters were remoteness and aesthetic rating (which accounts for presence of human refuse). Several water quality parameters proved to be highly significant for biotic integrity and the relationship between aesthetic quality and the chemical measurements was extremely accurate (see Figure 6-32). In the case of Maryland, the PHI used for habitat assessment didn't prove to be as accurate as the QHEI in Ohio. In coastal sites, only one out of the six metrics included in the PHI (epifaunal substrate) seemed to be important. In piedmont only one out of eight (remoteness) and in highland only one out of five (riparian width) were identified as significant. A big variability in habitat types as well as big influence of water quality impairment by non-habitat related sources, could explain the difficulty in linking biotic integrity to habitat quality. Nevertheless, in our opinion, Maryland's PHI should be reviewed and particular attention should be paid in identifying reference sites with no or little impaired water quality, especially in piedmont and highland sites.

In Minnesota, the SOM+MRT analysis showed a very similar profile as that shown for Ohio. In terms of habitat, all the metrics included in the state's QHEI were significant, but this was particularly true in the case of substrate and channel morphology parameters, which accounted for the differences between clusters 2 and 3, impaired mainly by habitat degradation. The SOM-neuron based analysis only showed important correlations with morphologic and land use scores. The overall performance of the QHEI in Minnesota was mediocre if compared with Ohio. The QHEI used in this state is based either in Ohio's QHEI by Rankin (1989) or Wisconsin's QHEI by Simonson et al. (1994). Specific QHEI development implies calibration and determination of reference sites and criteria for an area with particular features. Minnesota's QHEI is based on either Ohio or Wisconsin streams' reference sites which might be different from those in Minnesota. Even though the QHEI in Minnesota does a good job in separating clusters with different habitat quality, its correlation with biotic integrity could be further improved by calibrating this index with reference criteria for this particular state.

The relationships between some of the in-stream habitat parameters and off-stream parameters (riparian quality or width and gradient) were unveiled. As shown in Ohio, Maryland's coastal sites and Minnesota, gradient is the key for both substrate quality and variability and channel morphology quality. In the case of Ohio and Minnesota, substrate and morphologic parameters show a non-linear threshold with gradient (a score of seven or 1.5m/km respectively). Greater values mean steep decrease in embeddedness and increase in substrate, channel, pool and riffle quality. In the case of Maryland's coastal sites, gradient seems to have a linear effect over embeddedness, channel and epifaunal substrate. The difference in response with respect to Ohio and Minnesota could be explained because SOM in Ohio and Minnesota were performed for the whole state, while Maryland was subdivided in strata. The effect of gradient over stream's integrity has been demonstrated and we think that respecting the natural slope variability during human development is paramount. Human-induced stream's gradient changes such as

impoundments, or channel modifications have a big impact on stream's health. Removal of dams and structures that modify the natural flow regime should be one of the priorities in stream restoration and biotic integrity improvement projects.

The second off-stream variable with a deep impact on in-stream's habitat and chemical quality is the riparian strip quality and width. Riparian quality and width has an important role in preventing fine sediment from reaching the stream and avoiding channel and bank erosion and degradation. In the case of Ohio, riparian width seems to have a linear response with substrate, cover and channel scores. In the case of Minnesota and Maryland's coastal sites, a non linear relationship exists. In Minnesota, a steep increase in habitat quality can be achieved wit buffers widths greater than 9 meters. Unfortunately, in Minnesota, buffers with widths greater than 10 meters were recorded as 10, which made it impossible to determine a full range correlation. In Maryland's coastal sites, buffer widths greater than 30 meters guarantee a degree of embeddedness lower than 70%, while widths greater than 45 meters guarantee 40% or less of embeddedness (see Figure 6-23). Buffer strips not only proved to be highly effective in terms of sediment delivery control, supply cover for fauna and bank and channel erosion protection, but they also act as very effective filters for some chemical elements. As shown in Minnesota, chemical water quality will improve rapidly with buffer width after a 5 to 6-meter width threshold is achieved (see Figure 6-51 to 6-54). Obviously, this statement only applies for non-point pollution sources. The minimal riparian width, however, should be determined depending on the type of land use beyond the riparian strip.

Finally, we considered relevant to mention some interesting relationships found between some morphologic features and some water quality parameters. In the case of Minnesota, pool and width-depth ratio were closely related to ammonia. Higher values in both variables meant less concentration of ammonia (see Figure 6-55 and 6-56). This is relevant because with the modification of off-stream parameters, other in-stream features will be affected and therefore biotic integrity will be impaired not only by habitat degradation but because of the alteration of naturally occurring chemical reactions that take place in specific environments with the right conditions. Severe modifications of gradient and/or buffer quality, will not only have an important effect on habitat, but also in the capacity of that stream to develop its normal natural chemical reactions.

8. REFERENCES

Barbour, M.T., J.Gerritsen, B.D. Snyder and J.B. Stribling (1999). Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington D.C.

Brosse, S., Giraudel, J.L., and Lek, S. (2001) Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages, Ecological Modelling, 146: 159-166, 2001.

Chirart,J. (2003). Development of a macroinvertebrate Index of Biological Integrity (MIBI) for rivers and streams of the St. Croix River basin in Minnesota. Minnesota Pollution Control Agency .   Biological Monitoring Program. St. Paul, MN.
Available at: http://www.pca.state.mn.us/publications/reports/biomonitoring-mibi-stcroix.pdf

Davies, J.L., and Bouldin, D.W. (1979) A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(4): 224-227

Frey, D.G. (1975). Biological integrity of water-an historical approach. Pp. 127-140 in R.K. Ballentine and L.J. Guarraia (eds.). The Integrity of Water. U.S. Environmental Protection Agency, Office of Water and Hazardous Materials, Washington, D.C.

Genet, J., Chirart, J.(2004).). Development of a macroinvertebrate Index of biological Integrity (MIBI) for rivers and streams of the Upper Mississipi river basin. Minnesota Pollution Control Agency.Biological Monitoring Program. St. Paul, MN.
Available at: http://www.pca.state.mn.us/publications/reports/biomonitoring-mibi-uppermiss.pdf

Hall Jr., L.W., Morgan II, R.P., Perry, E.S., Waltz, A. (1999). Development of a Provisional Physical Habitat Index for Maryland Freshwater Streams. Maryland Department of Natural Resources. Chesapeake Bay and Watershed Programs. Monitoring and Non-tidal Assessments

Hilsenhoff, W.L.(1987). An improved biotic index of organic stream pollution. Great lakes Entomol. 20:1-39

Karr, J.R. and Dudley, D.R. (1981). Ecological perspective on water quality goals. Environmental Management 5: 55-68

Karr, J.R., Fausch, K.D., Angermeier, P.L., Yant, P.R., Schlosser, I.J. (1986). Assessing Biological Integrity of Running Waters. A Method and its Rationale. Spec. Publ. 5, Illinois Natural History Survey, Champaign, IL.

Karr, J.R., Kerans, B.L. (1991). Components of Biological Integrity: Their Definition and Use in Development of an Invertebrate IBI. 1991 Midwest Pollution Control Biologists Meeting. Environmental Indicators: Measurement and Assessment Endpoints. Lincolnwood, Illinois

Kauffman, P.R., Levine,P., Robison, E.G., Seeliger,C., Peck, D.V. (1999). Quantifying Physical Habitat in Wadeable Streams. EPA/620/R-99/003. U.S. Environmental Protection Agency, Washington D.C.

Kiviluoto, K. (1996). Topology preservation in self-organizing maps, Proceedings of IEEE International Conference on Neural Networks, 1: 294-299

Kohonen, T.(2001) Self-Organizing Maps. Third Edition. Springer, Berlin.

Legendre, P.L., Legendre,L. (1998). Numerical Ecology, 2nd edition. Elsevier Science, B.V., Amsterdam

Lazorchak, J.M., Hill, B.H., Averill, D.K., D.V. Peck and D.J. Klemm (editors). (2000). Environmental Monitoring and Assessment Program-Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Non-Wadeable Rivers and Streams. U.S. Environmental Protection Agency, Cincinnati, Ohio

Mason,C.F.(1991). Biology of fresh water pollution, 2nd ed., Longman Science and Technical, J Wiley, NY

Niemela S.N., Feist, M.D. (2002). Index of Biotic Integrity Guidance for Coolwater Rivers and Streams of the Upper Mississipi River Basin. Minnesota Pollution Control Agency . Biological Monitoring Program. St. Paul, MN.
Available at: http://www.pca.state.mn.us/water/biomonitoring/bio-streams-fish.html

Niemela S.N., Feist, M.D. (2000). Index of Biotic Integrity Guidance for Coolwater Rivers and Streams of the St. Croix River Basin. Minnesota Pollution Control Agency Biological Monitoring Progra,. St. Paul, MN.
Available at: http://www.pca.state.mn.us/water/biomonitoring/bio-streams-fish.html

Novotny, V. (2003). Water Quality. Diffuse Pollution and Watershed Management. 2nd edition. John Wiley and Sons. NY

Ohio Environmental Protection Agency (1987). Biological Criteria for the Protection of Aquatic Life. Columbus, OH

Paul, M.J., Stribling, J.B., Klauda, R.J., Kazyak, P.F., Southerland, M.T., Roth, N.E. (2003). A Physical Habitat Index for Freshwater Wadeable Streams in Maryland. Final Report. Maryland Department of Natural Resources. Monitoring and non-tidal assessment. Available at: http://www.dnr.state.md.us/streams/mbss/mbss_pubs.html

Rankin, E.T. (1989). The Qualitative Habitat Evaluation Index (QHEI): Rationale, Methods and Application. Ohio EPA. Division of Surface Water. Ecological Assessment Section.

Rankin, E.T., Yoder, C.O., Mishne,D. (1990). 1990 Ohio water resource inventory. Ohio Environmental Protection Agency, Division of Water Quality Planning and Assessment, Ecological Assessment Section, Columbus, OH

Roth, N.E., Southerland,M.T., Chaillou,J.C., Kazyak,P.F., Stranko,S.A. (2000). Refinement and validation of a fish index of biotic integrity for Maryland streams. Maryland Department of Natural Resources. Chesapeake Bay and watershed programs monitoring and non-tidal assessment. Available at: http://www.dnr.state.md.us/streams/pubs/ea00-2_fibi.pdf

Southerland, M.T., Rogers,G.M., Kline,M.J., Morgan, R.P., Boward, D.M., Kazyak,P.F., Klauda, R.J., Stranko, S.A. (2005). New biological indicators to better assess the condition of Maryland streams. Maryland Department of Natural Resources. Chesapeake Bay and watershed programs monitoring and non-tidal assessment. Available at: http://www.dnr.state.md.us/streams/pubs/ea00-2_fibi.pdf

Stribling, J.B., Jessup,B.K., White, J.S. (1998). Development of a benthic Index of Biotic Integrity for Maryland streams. . Maryland Department of Natural Resources. Chesapeake Bay and watershed programs monitoring and non-tidal assessment. Available at: http://www.dnr.state.md.us/streams/pubs/1998_benthic_ibi.pdf

Virani, H, Manolakos, E., Novotny, V. (2005). Self Organizing Feature Maps Combined with Ecological Ordination Techniques for Effective Watershed Management. Technical Report No. 4. Center for Urban Environmental Studies. Northeastern University, Boston. Available at: http://www.coe.neu.edu/environment/publications.htm .