Performance Report Summary

Period Covered by the Report:	May 1, 2005 – May 31, 2006
Date of Report:	July 20, 2006
EPA Agreement Number:	R83-0885-010
Title:	**Developing of Risk Propagation Model for Estimating Ecological Responses of Streams to Anthropogenic Watershed Stresses and Stream Modifications**
Investigators:	Vladimir Novotny, Timothy Ehlinger, Elias Manolakos, Alena Bartošová
Institutions:	Northeastern University, Boston, MA (lead institution) University of Wisconsin, Milwaukee, WI, Illinois State Water Survey (University of Illinois), Champaign, IL
EPA Project Officer:	Iris Goodman**,** Bernice Smith
Research Category:	**Developing Regional-Scale Stressor-Response Models for Use in Environmental Decision-Making, Water and Watersheds**
Project Period:	May 1, 2003 – May 31, 2007

Objectives of the Research Project

The goal of this research is the development of a regionalized watershed-scale model to determine aquatic ecosystem vulnerability to anthropogenic watershed changes, pollutant loads and stream modifications (such as impoundments and riverine navigation). The model will assist watershed managers in their decisions on methods to mitigate stream degradation and biological impairment, assess potential watershed impacts, and identify watershed restoration opportunities. The layered hierarchical model system, developed by Artificial Neural Net modeling and analysis, will be based on probabilistic risk propagation and linking the stresses with ecologic endpoints, from physical attributes of the watershed and water body and pollutant loadings at the lowest level to measures of biotic integrity, such as the Index of Biotic Integrity (IBI), at the highest level.

The main objectives and outcomes of the research are: (1) Developing a model that would consider pollutant effects of impoundments for navigation and other purposes, channelization, watershed modification, and riparian corridor and land use changes as the key root stressors, using primarily data obtained from the Midwest streams; (2) Developing layered hierarchical progression of risks from the basic root stressors to the biotic endpoints (fish and macroinvertebrate IBIs); (3) Using the model to study the possibility of mitigating the stressors in a way that would have the most beneficial impact on the biotic endpoints; (4) Developing a manual for watershed managers and other users; and (5) Investigating adaptability and transferability of the model to a stressed New England stream.

Brief Progress Summary/Accomplishments:

In the first phase of our research (first two years) we established and demonstrated that a particular artificial neural network (ANN) structure, the Self Organizing Maps (SOM) can be

used to pattern and profile the distribution of stressors in large stream ecosystems, and discriminate sampling sites according to multi-stressor impact. SOM were used to analyze the biological integrity of the streams in the states of Ohio and Maryland. This type of ANN analysis is called unsupervised learning. Using Canonical Correspondence Analysis (CCA) and Principal Component Analysis (PCA) the research teams are identifying ranking of stressors as to their impact on IBI and Cluster Dominating Parameters (CDP).

In the second phase of the research (third year) we capitalized on the very promising results of first phase. We added supervised ANN based prediction capabilities as a step following the hierarchical unsupervised nonlinear clustering of sampling sites according to fish IBI metrics vectors distribution. Our objective is to be able to built simple yet powerful models that could be used to predict the IBI and its metrics (both fish and macroinvertebrate).

The team at the University of Wisconsin (Milwaukee) has assembled and is using extensive fish, habitat and land cover database for the State of Wisconsin and has developed a GIS based system to be used for analyzing impacts of stream habitat and fragmentation, hydrological and hydraulic parameters, and watershed land use of the stream biological integrity, using also SOMs and correlating it to the various “stressor” metrics calculated from GIS Database

Because habitat parameters have been identified as the stressors that have the greatest impact, significant effort is now devoted by both teams to synthesize and analyze the habitat metrics. Most of the current metrics identified, for example, in the Rapid Biotic Assessment Protocols are observational, i.e., they cannot be predicted. Our teams are striving towards developing predictive habitat indices and measures.

We have also completed our data base for storing and querying vast amounts of data from several states (Ohio, Maryland, Massachusetts, Wisconsin, Minnesota).

An extensive effort was devoted in the first 6 months of 2007 to the development methodology and execution of the supervised ANN modeling that was finding the best relationship between the IBI, its metrics and environmental variables (habitat metrics, land use, and water quality). Supervised Artificial Neural Network (ANN) modeling is a powerful tool to predict fish IBI or fish metrics from the watershed stressors. In this extensive effort datasets from Ohio EPA (OEPA) were used.

OEPA data set had 1848 records of 34 watershed stressor parameters, IBI and 12 fish metrics that were used to construct IBI for 1193 stations. This set included missing values for the watershed stressor parameters at few stations. For the optimal use of the available data, missing values were substituted either with the average values (if the number of missing values is small) or with the kriged GIS values (if missing values are large in number).

Maryland data was extracted from Maryland Biological Stream Survey (MBSS) data at 955 sampling stations for the period of 1995-1997. The sampling stations were spread throughout west (Appalachian Plateau), central (Piedomont) and east (Costal Plains) regions. The dataset contained quantitative and qualitative data of chemical and habitat and land use parameters. In addition, Benthic Index of Biological Integrity (BIBI) and Physical Habitat Index (PHI) were also available.

Three approaches were used to predict Fish IBI; single supervised ANN for the whole preprocessed Ohio data, three supervised ANNs for the clustered data, and multivariate regression models. The best set of input watershed stressors was selected from CCA and Principal Component based analyses. A three layer ANN (input layer-hidden layer-output layer)

was employed for prediction. 60% of the data was used for training, and the rest was used in validation and testing.

For developing prediction models from data four approaches were tested.

1. Traditional Multivariate Regression model,
2. Supervised ANN models with whole dataset,
3. Supervised ANN models for clusters, and SOM based Prediction model.

Over 80 IBI models were developed and analyzed from the Ohio data by supervised learning. The best models can account for more than 50% of the IBI variability.

The UW-M team demonstrated in a similar analysis that the SOM Clusters are detecting and patterning differently than the traditional IBI developed and calibrated for Wisconsin. What the SOM's are showing is that each of the clusters is defining a unique class of fish community that may each have their own combination of natural and anthropogenic stressors. This may help rethink the standard concept of degraded versus non-degraded sites on a large geographic level.

Publications/Presentations

The following publication has been accepted and will be published:

D.N. Beach and V. Novotny, Modeling variability of in-stream nitrogen concentrations using principal components analysis, submitted to *Journal of AWRA, accepted for publication, 2006*

The following publications are under review

E. Manolakos, H. Virani, V. Novotny, and T. Ehlinger, Visualization of biological integrity and analysis of stressors-response relationships in stream ecosystems using Self-organizing feature maps, *Water Research,* submitted and under review, 2006

Two publications/presentations were prepared and presented in 2006.

V. Novotny (2006) Agricultural diffuse pollution: Are we on the right track to successful abatement? Invited Key Note Presentation, Proc. SEPA/SAC Biennial Conference – Agriculture and the Environment – Managing Rural Diffuse Pollution, April 5-6 2006, Edinburgh, Scotland

V. Novotny and E. Manolakos (2006) Ecological clustering of integrity and nonlinear impact of environmental variables, Invited Keynote presentation, Proc. RESLIM 2006 International Conference, August 27 – September 1, Brno, Czech Republic

Future (Concluding) Activities

The time between now and May 2007 will be devoted to completion of the entire projects. It is expected that subcontractors will finish their work and submit technical reports by the end of 2006. This work will include the results of the biological studies in Wisconsin and the risk model by the University of Illinois.

The Northeastern University team will compete the following studies:

Analysis of habitat clustering for Ohio, Maryland, Minnesota and Wisconsin

Macroinvetebrate clustering and modeling, Ohio, Maryland and Massachusetts (test state)

Impact of impoundments

Manual for SOM analyses and modeling

The final report will be compiled and written in 2007.

A workshop proposal will be sent to US EPA in July-August 2006.

Relevant Websites:

A web site has been created by the team members where all reports and other publications or their abstracts are or will be available **www.coe.neu.edu/environment.**