# STUDY PERFORMANCE REPORT

**State:** Michigan **Project No.:** F-80-R-6

**Study No.:** 230738 **Title:** Improve and validate river segment identification and classification models for assessing fishery potential and environmental impairment in Michigan.

**Period Covered:** October 1, 2004 to September 30, 2005

**Study Objective:** The overall objective is to develop tools for assessing river fishery potential and environmental impairment for Michigan. The specific goals are to: (1) improve and validate models for predicting thermal and hydrological regimes, expected fish distribution, and expected physical habitat characteristics; (2) classify river segments; (3) conduct environmental impairment assessment; (4) provide site specific information required for establishing management goals; and (5) better understand how the management goals can be achieved.

**Findings**: Jobs 1, 2, 3, 4, and 6 were active for 2004-05. This project is co-funded by the Federal Aid in Sport Fish Restoration Program, Project F-80-R-6 and by the National Center for Environmental Research (NCER) STAR program of the U.S. Environmental Protection Agency, grants # R-83059601-0. The co-PIs funded by the Sport Fish Restoration Program are responsible for coordinating overall job planning, supervising research associates (funded by the NCER STAR program), providing technical guidance, and partially involving in the model development and classification processes.

**Job 1. Title: Assemble and cross walk GIS database layers; delineate watershed boundaries and calculate reach characteristics; and clip out all GIS layers at catchment, riparian, and reach scales.**

(1) GIS database layers of land use/cover, surficial geology, bedrock geology, bedrock depth, groundwater delivery potential, annual precipitation, annual growing-degree-days, and July mean air temperature have been assembled and cross walked for the entire State of Michigan.

(2) All river reaches defined by confluence to confluence that appear on 1:100,000 National Hydrographic Data have been identified. Catchment boundaries associated with each reach have been delineated. River reach sinuosity, gradient, and network connectivity were calculated.

(3) Landscape databases including surficial geology, soil, bedrock depth, bedrock geology, catchment slope, groundwater delivery potential, and climate factors associated with each river reach have been clipped for entire upstream catchment, network riparian buffer, local catchment (areas where surface water drains directly to the reach), and local riparian buffer for each river reach.

**Job 2. Title: Develop models for predicting water temperature, flow discharge, and fish assemblages.**

**Temperature Model:** Preliminary modeling work was conducted for Michigan's mean July stream temperature. Data collected by the MDNR were available from 260 sites located throughout Michigan. The locations of temperature sampling sites were geo-referenced and landscape data were summarized at multiple spatial scales for each site. A number of approaches have been used to explore the best way of developing temperature-predicting models. These methods included: (1) stepwise multiple linear regression models; (2) semi-parametric modeling of temperature through generalized additive models; (3) incorporating spatial trend surfaces in models; (4) temporal detrending of the temperature combined with regression prediction; (5) spatial regression, which included an autocorrelative error structure based on distances between sampling points; and (6) combining regression (both linear and nonlinear) predictions with kriging and co-kriging of residuals. The last approach developed by geo-statistical and generalized additive modeling has been proven to be satisfactory. In this approach, kriging was used to interpolate broad spatial patterns in mean July stream temperature from 260 sites where temperature loggers were deployed. Spatial autocorrelation in the kriged estimates was determined through a spherical semivariogram, which was estimated via nonlinear least squares in S-PLUS (Insightful Corp., Seattle, Washington). Deviations of observed mean July stream temperatures from kriged temperature estimates were then computed for the temperature logger sites. Generalized additive modeling was then used to develop a regression model for predicting deviations in observed temperatures from the kriged estimates based on landscape-level environmental variables. Variables used for predicting the deviations were determined using forward selection. Selected variables were natural log transformed network catchment area, mean July air temperature in the network catchment, potential groundwater loading in the network catchment, mean soil permeability in the reach catchment, percent barren land type in the reach catchment, percent forested wetland in the reach catchment, percent ignaceous/metamorphic bedrock in the reach buffer, and percent agricultural land use in the reach buffer. Predicted mean July stream temperatures were calculated by summing the kriged temperature estimates with the predicted deviations from the fitted generalized additive model. This combined geo-statistical and generalized additive modeling approach to predicting stream temperature yielded temperature estimates that were within two degrees Celsius of observed mean July stream temperature for approximately 90% of a hold-out dataset. The correlation between predicted and observed stream temperatures for the hold-out dataset was approximately 80%.The temperature models were recently used to predict July mean water temperature for all river reaches in the State of Michigan, which completes the temperature modeling task of this study.

**Flow model:** Statewide stream flow models were developed based on daily flow measurement from 68 U.S. Geological Survey (USGS) gauging stations that meet the criteria for flow model development. The criteria for selecting gauging stations were (1) having 20 years or more continuous records that included the 1995 water year; (2) being not close to or downstream of a dam; (3) being not listed in the USGS water book as having flow regulated by dams (including power plants, lake outflow, and mill dams); (4) having no effluent additions, diversions, mine pumpage; (5) having no diurnal fluctuations at low flow. These models were developed based on multiple regression between flow discharge data from USGS gauging stations and their corresponding catchment area, slope, annual precipitation, surficial geology, and land cover. Multiple linear regression was used to develop predictive models of annual and August exceedence flows (5%, 10%, 25%, 50%, 75%, 90%, and 95%). Model development began with a simple hydraulic geometry equation that included the catchment area, precipitation, and valley slope. Additional predictors were added to this base model in a stepwise fashion beginning with surficial geology summaries. When the base model with surficial geology had been developed, agricultural and urban land covers were added to all models because we were particularly interested in the influence of these predictors. These models were initially developed using the median flow and then checked against high and low flows for fit using the adjusted $R^2$ after each predictor was added. These models explained between 72% and 98% of the variation in flows. The flow models were recently used to predict exceedence flows for all river reaches in the State of Michigan, which completes the flow modeling task of this study.

**Fish modeling:** A fish assemblage database was assembled from various sources, including Michigan DNR and University of Michigan. This database contains 920 stream sites with fish assemblage collected using comparable sampling methods (backpack or tow-barge) and sampling seasons (May to September). Fish assemblages were summarized into 32 metrics. Different scale landscape data were linked to each of the fish sampling site. A preliminary modeling work has been conducted using a combination of regression (both linear and nonlinear) predictions with kriging and co-kriging of residuals.

**Jobs 3 and 4. Title: Identify and classify stream/river segments.**–A visual Basic program has been developed to merge adjacent river reaches that have similar physical, chemical, morphological, and biological characteristics using Cluster Affinity Search Technique (CAST). The CAST clustering routine is a non-hierarchical procedure that does not require pre-specification as to the total number of clusters (i.e., river segments), thus it represents a significant advantage over other clustering procedures. Some modification of the CAST clustering routine was required for it to deal effectively with the spatial structure of river networks (i.e., not all river reaches can be considered spatial neighbors), but such modifications would have been necessary for any clustering procedure used with digital river networks.

River segment identification by CAST begins by specification of an affinity threshold parameter, which determines the minimum level of similarity for neighboring river reaches to be clustered into a segment. The affinity threshold influences the final number of river segments that will be formed. Once the affinity threshold value has been specified, CAST begins by selecting an individual reach and comparing the level of similarity between the selected reach and its neighboring reaches. Only linearly-related river reaches are considered neighbors. If the selected reach has at least one neighboring reach that is sufficiently similar, then the two reaches are merged into a segment. If none of the reaches are sufficiently similar, the selected reach is considered to be its own segment.

**Job 6. Title: Write progress report.**–This progress report has been prepared as scheduled.

**Prepared by:** Lizhu Wang
**Date:** September 30, 2005