STATE WILDLIFE GRANT—INDIANA
Mussel Habitat Modeling

Purdue University graduate student Suse LaGory searches for freshwater mussels in Beaver Creek in Newton County.

CURRENT STATUS

First year of a two-year project

FUNDING SOURCES AND PARTNERS

State Wildlife Grant Program (T7R24)
Purdue University

PROJECT PERSONNEL

Dr. Tomas Höök, Principal Investigator, Purdue University
Carolyn Foley, Purdue University
Susanna LaGory, graduate student, Purdue University

BACKGROUND AND OBJECTIVES

The pearly mussels (superfamily *Unionoidea*) are a highly diverse and globally distributed group of freshwater bivalves. Comprising about 300 native species, the North American assemblage of freshwater mussels is the richest in the world. Although many aspects of their ecology and life history are not fully understood, freshwater mussels are known to support a number of critical ecosystem functions in the streams, rivers, and lakes they inhabit. As filter feeders, mussels improve water clarity, maintain connections between the upper and lower trophic levels of stream food webs, and help cycle nutrients within aquatic ecosystems. Additionally, the shell production and burrowing behavior of freshwater mussels can stabilize substrate and provide beneficial habitat structure that is used by fish, aquatic invertebrates, and other benthic organisms. They are also a food source for many bird and mammal species.

Despite their contributions to stream ecosystem health, freshwater mussels are one of the most globally imperiled groups of organisms. Indiana has historically supported a rich species assemblage of mussels; however, similar to mussel populations in other areas, the state's mussel populations have recently suffered rapid losses in abundance and biodiversity due to stressors such as pollution, loss of fish host species, habitat degradation and fragmentation, and overharvest.

While these general factors are thought to have

Purdue graduate student Suse LaGory holds a giant floater (*Pyganodon grandis*) sampled from Beaver Creek in Newton County.

A native freshwater mussel siphons in the sunshine in Beaver Creek near Willow Slough Fish & Wildlife Area. (Photo by Suse Lagory)

contributed to declines of mussel populations, more specific drivers of declines have not been identified for many species. Many of Indiana's remaining native freshwater mussel species are currently listed by state or federal agencies as endangered, threatened, or vulnerable to extinction; however, state conservation strategies have only recently been put in place to address these declines and prevent further loss of biodiversity. A crucial first step in these strategies has been a statewide survey to document the distribution of freshwater mussels in streams and rivers. These monitoring efforts have collected data on mussel species presence and relative abundance at thousands of locations around the state, creating a thorough documentation of current freshwater mussel distributions in Indiana.

We aim to improve state conservation strategies and fill knowledge gaps surrounding specific drivers of freshwater mussel declines and distributions in Indiana. Our specific objective is to use monitoring data collected by the DNR to develop species-specific models relating native mussel distributions to various environmental characteristics. The primary goal in developing these models is to identify specific environmental and habitat characteristics associated with the presence and relative abundance of freshwater mussels across the state. Identifying these characteristics and quantifying their relative importance will inform conservation strategies such as aiding in prioritizing sites for reintroducing locally extinct mussel populations or protecting critical mussel habitat.

METHODS

Field Data Collection

The dataset used for this study includes freshwater mussel presence data collected by the DNR. Most data were collected during timed surveys targeting mussels, and a subset of the data was collected via incidental observations of mussel presence during fish surveys. Monitoring data were collected annually from 1994 through 2017, generally between the months of May and November.

For timed surveys targeting mussels, sites were visually searched for a minimum of 30 minutes. The survey time for each site was calculated as the sum total of minutes searched by all individual surveyors (e.g., three surveyors at 10 minutes each = 30 survey minutes). Site transects were surveyed starting at the midpoint. Surveyors walked in opposite directions, meandering and visually scanning the stream bottom for live mussels or mussel shell material.

When a live mussel or piece of shell material was encountered, it was temporarily removed from the stream for species identification and immediately returned to its original location in the stream. The total number of live individuals was recorded for each detected species. If no live individuals were found but shell material of a species was present at the survey location, the condition of the shell material was recorded as one of the following categories: fresh dead, weathered dead, or sub-fossilized. For incidental observations where searches were not timed, presence and "best condition" (i.e., live, fresh dead, weathered dead, or sub-fossilized) of a species found at the survey site were recorded.

Model Development

We are using these survey data to develop statistical models that will help understand and predict where particular freshwater mussel species are present, as well as what might affect the abundance of a species present at a given location. These models are simplified representations of the natural world that are built with two main components: response and predictor variables. A response variable is the pattern or process

in the environment we are interested in explaining or learning more about through our model. In our study, the response variables are patterns of mussel species presence and abundance. Predictor variables are factors that may explain patterns in the response variable. In our models, the predictor variables are the different characteristics of land and streams across Indiana.

Response Variables

We broadly grouped response variables for our models into two categories: presence/absence and relative abundance. These two categories of response variables can be defined in many different ways from our data set. For our initial models, we will use two separate definitions of both presence/absence and relative abundance. Some survey sites were sampled multiple times, so to avoid issues with independence and unbalanced design in our models, we only included data collected during the most recent sampling event at each location.

The species we used in our preliminary analyses was the wavy-rayed lampmussel (*Lampsilis fasciola*). This species was recommended for our initial analyses by DNR mussel specialists because of its historic widespread distribution in the state and unknown causes of its local extinction at sites throughout its historic range. In the near-term, we plan to repeat these analyses for spike (*Elliptio dilatata*), kidneyshell (*Ptychobranchus fasciolaris*), and rainbow (*Villosa iris*) mussels, and will ultimately analyze additional species as the project progresses and our methods become further refined.

Potential Predictor Variables

To create models that can be broadly applied, we have limited our options for predictor variables to environmental and habitat data that are available for the entire state. We defined our predictor variables at multiple spatial scales, allowing us to examine effects of habitat and environmental characteristics at more localized watershed scales and broader, basin-level scales. Predictor variables included in our initial models are: major drainage basin, number of dams in the watershed, watershed land use composition, dominant bedrock type, and catchment area. We now plan to incorporate soil attribute data into our models and may also include climate variables such as mean annual precipitation, mean annual air temperature, and annual air temperature range.

Statistical Analyses

To improve our ability to draw reliable conclusions from these analyses, we will use an ensemble modeling approach. Using this method, we will draw conclusions from the combined results of multiple modeling methods rather than from the results from one type of model. To evaluate sites and their capacity to support the mussel species of interest, we will average and rank model residuals (the difference between model prediction and observed presence/absence or abundance). Smaller, more negative residuals will indicate sites that may not currently support the species of interest but would be good candidates for species reintroduction or habitat protection. Larger residuals will indicate sites that are less suitable for or may be at risk of losing the species of interest. Ultimately, this ranking and evaluation process will be used to identify and prioritize sites for conservation and management actions.

Our predictor variables are both continuous (i.e., numeric with infinite possible values) and categorical (i.e., fixed number of possible categories), so we will only use modeling methods that can accept both variable types. The first modeling approach we used is recursive partitioning with Classification and Regression Trees (CART). Classification trees are best suited for binary response data (i.e., two possible values) and were used to model presence/absence data for wavy-rayed lampmussel. Regression trees are best suited for continuous response data, and were used to model *L. fasciola* relative abundance data. We included all predictor variables in our initial CART models, and "pruned" trees to minimize the model's cross-validation

A plain pocketbook mussel (*Lampsilis cardium*) sampled from Beaver Creek near Willow Slough Fish & Wildlife Area during mussel surveys conducted by DNR staff. (Photo by Suse Lagory)

PREDICTOR VARIABLE	SCALE	SOURCE	YEAR	TYPE	LEVELS
Major Drainage Basin	basin	Indiana Geological Survey	1991	categorical	Lake Michigan, Lake Erie, Wabash River, Ohio River, Illinois River
Dam Count	HUC 10	Division of Water, IDNR	2018	continuous	not applicable
Percent Developed Land Use	cumulative upstream catchment	National Land Cover Database	2011	continuous	developed: high intensity, medium intensity, low intensity, open space
Percent Agricultural Land Use	HUC 12	National Land Cover Database	2011	continuous	cultivated crops, pasture/hay
Percent Natural Land Use	HUC 12	National Land Cover Database	2011	continuous	wetland: emergent herbaceous, woody; grassland/herbaceous; shrub/scrub; barren land; open water; forest: mixed, evergreen, deciduous
Dominant Bedrock Type	HUC 8	Indiana Geological Survey	2013	categorical	black shale, shale, limestone, dolostone, sandstone, siltstone, glacial drift, sand
Catchment Area	local catchment, cumulative upstream catchment	Tipping Point Planner	2018	continuous	not applicable

Description of predictor variables included in preliminary models. Scale specified indicates the spatial scale at which variables were summarized. All summary statistics for environmental predictor variables were calculated in ArcGIS Pro.

error. Pruning is a common method to improve CART model interpretability and to prevent overfitting of models. These analyses were performed using the "rpart" package in program R.

The second modeling approach we used is Generalized Linear Modeling (GLM). GLMs are a group of regression analyses that can accommodate non-linear relationships between predictor and response variables. The logistic family of GLMs is best suited for modeling our presence/absence data, and the Poisson family of GLMs is best for modeling our relative abundance data. To determine which predictor variables to include in each of our final GLMs, we used a backward stepwise regression process with Akaike's Information Criterion (AIC) as the selection criterion. During this process, we start with a model that includes all possible predictor variables. The model is then repeatedly re-run, each time removing one predictor variable and reassessing the model's performance with each "step." This stepwise process results in a final model that includes a combination of predictor variables that best explains patterns in the response variable. The regression coefficient of each of these predictor variables indicates its relative contributions to the distribution of our response variables—either presence/absence or relative abundance. All GLM analyses were performed using the 'MASS' package in program R.

PROGRESS TO DATE

We have developed and tested four initial models for wavy-rayed lampmussel using the previously described approaches. One significant challenge in the first year of this project has been assembling and managing large data sets such as our mussel survey data set and environmental characteristic data sets. Identifying the proper predictor variables and corresponding spatial scales for building our models has also presented a challenge. There is an enormous number of potential environmental variables that could be used as predictors in these models, and it is crucial to set

constraints to identify predictor variables that will enhance—rather than complicate—our understanding of patterns in freshwater mussel species distributions.

Because of these challenges and the nature of the modeling process, the initial model development phase has resulted in many different iterations of each model yielding slightly different outputs. There have, however, been consistencies among these preliminary models, indicating certain predictor variables as important drivers of wavy-rayed lampmussel presence/absence and relative abundance. Common predictor variables are catchment area, watershed land-use composition, number of dams in the watershed, and dominant bedrock type. These commonalities indicate that these variables are central underlying factors that may be contributing to the current distribution and local abundances of wavy-rayed lampmussel across Indiana. Preliminary model results suggest that presence and abundance of wavy-rayed lampmussels are both positively associated with catchment area and percentage of natural land use in the catchment, while presence and abundance are negatively associated with number of dams in the catchment.

Future Analyses

In addition to the previously mentioned predictor variables and additional species distribution data we will model in the future, we will build upon our initial analyses and further develop our ensemble of models by using additional modeling techniques. Our final ensemble will likely include Boosted Regression Trees or Random Forest models, both of which are more robust expansions of CART methods. Additionally, we will model our data using machine learning techniques such as Artificial Neural Networks (ANN) and Maximum Entropy (MaxEnt).

COST: $131,512 FOR THE COMPLETE TWO-YEAR PROJECT