# HREP Mussel Modeling

Michael Dougherty
Davi Michl
Dan Kelner
2019-03-02

# Contents

4 CONTENTS

# Preface

## 0.1 Abstract

6 CONTENTS

## Introduction

Species Distribution Modeling (SDM) has become a common method for mapping suitable areas for many species or taxa groups (?, ?). The dismo R package has been created to support SDM in R (see the dismo package vignette sdm, ?). Another good overview of SDM in R (?).

#### 1.1 Literature Review

- UMESC Pool 8 mussel model
- UMESC Pool 18 mussel model

### 1.2 Identify Available Datasets

The USACE Mussel Database HREP pre-project survey

- 2018 Steamboat survey
- 2014 Cordova survey
- 2010 Cordova survey

# Mussel Data Processing

#### 2.1 Prepare Mussel Sample Data

Working geodatabase: //mvrdfs/egis/Work/EMP/HREP\_Projects/SteamboatSlough/Mussels/SteamboatMusselModel/Database

- Import sample location point feature class for each survey.
- Import "individual" data tables for each survey.
- Add double fields 'Easting' and 'Northing' and calculate geometry for each sample location point feature class.
- Add double fields 'Easting' and 'Northing' to each mussel tabular data.
- Join the sample location point feature class to each mussel tabular data. Directionality of join is important: right-click mussel tabular data and select join, using SAMPLEID as join field.
- Field calculate 'Easting' and 'Northing' from the sample location point feature class. Then remove the join.
- Use the "Display XY" tool to create events from the mussel tabular data.
- Export the events to the geodatabase as: Cordova\_YEAR
- Create 3 new fields (type: long integer) titled 'NumberLive', 'NUM5YEAR', and 'NUM30MM'
- Use field calculator to populate each field with '0' indicating absence and '1' for presence in each new field (number live, number less than 5 years, and number less than 30mm in length).
- Create a definition query to select only those records where: NUMBER\_LIV > 0
- Dissolve the Cordova\_YEAR feature class by SAMPLEID to collapse mussel tabular data to the unique sample location points. The dissolve tool can also be used to collapse mussel tabular data to unique sample location using multiple fields; e.g., if a count by species is required for analysis, select SAMPLEID and ScientificName. In the statistics box, add the newly created fields 'Number\_Live', 'NUM5YEAR', and 'NUM30MM' and select 'SUM' as statistic to be calculated for each.
- Append all three datasets into one feature class in the .gdb called Cordova. Right-click this feature class and select 'load'.
- Export to .csv.

## Calculate Mussel Metrics

#### 3.1 Prepare Mussel Sample Metrics

The purpose of this section is to calculate various metrics of habitat quality for each sampled site. Each of these metrics will be used as presence data as input to create a set of Maxent habitat suitability models. The metrics calculated in this section were taken from the Upper Mississippi River Mussel Community Assessment Tool (MCAT) (?).

#### 3.2 MCAT Metrics

The following MCAT metrics were selected for inclusion to the model:

- percent listed species: percent of listed threatened or endangered species (federal or bordering states) is a measure of sensitive species
- percent tolerant: percent of tolerant species (Amblema plicata, Quadrula quadrula, and Obliquaria reflexa) is a measure of a highly disturbed mussel assemblage; i.e., dominated by species tolerant of unstable substrates, silt accumulation, low current velocities, and fluctuating flow conditions.
- percent tribe Lampsilini: percent of assemblage that falls within tribe Lampsilini is a measure of species composition, life history, and behavioral characteristics
- percent juveniles: percent of mussels <= 5 years-old is a measure of recruitment into an assemblage over the past five years
- percent >= 15 yrs: percent of mussels >= 15 years-old is a measure of age distribution in an assemblage
- Q75 abundance: a measure of abundance at the 75th quartile calculated as the density (no./m^2)
- species evenness: species evenness represents the dominance of an assemblage by a few species using Pielou's evenness index (range 0 to 1)
- tribe evenness: tribe evenness represents the dominance of a particular taxonomic group using Pielou's evenness index (range 0 to 1)
- ES\_100: the expected number of species with a sample size of 100 estimated by rarefaction is a measure of a healthy mussel assemblage

#### 3.3 Create the Individuals table

The purpose of this step is to assemble the "individual" data for each of the surveys within the study area that will be used in the study.

will be about in the study.

Working geodatabase: //mvrdfs/EGIS/Work/EMP/HREP\_Projects/SteamboatSlough/Mussels/SteamboatMusselMode

- Create a feature class respresenting all individual mussel samples in the study area, using a definition query to include only live individuals and name it Cordova\_all\_individuals. Note: the Q75 abundance metric will be based on live mussel counts for each quadrat.
- Add the following fields (Type=long integer) to the attribute table: Listed, Tolerant, Lampsilini, Juveniles, Over\_15yrs, tribe

#### 3.4 Create Percent Listed

The purpose of this step is to calculate the percent listed MCAT metric for each sampled site.

- Dissolve this feature class by SAMPLE\_ID and ENAME (Scientific name). Add 'SUM\_NumberLive' and 'SUM\_Listed' fields in the Statistics section of the tool. Uncheck create multipart features. Set Output Feature Class to the working .gdb and name it PercentListed.
- Open the PercentListed attribute table. Add a field titled 'Perc\_listed' Type=double and enter the expression ([SUM\_Listed] / [SUM\_NumberLive]) \* 100 into the field calculator to calculate % of individual listed species per quadrat sample.
- Add 3 fields titled 'species' Type=string and Length=20; 'longitude' Type=double; and 'latitude' Type=double; in that order.
- Use field calculator to populate the field with species="perc\_listed". Use Calculate Geometry to populate the 'longitude' and 'latitude' fields, ensuring that the coordinate system is set to 'NAD 1983 2011 US Feet' and 'StatePlane Illinois West FIPS 1202' and the units are set to US feet.
- Use the Symbology tab to select a suitability threshold based on quartiles (e.g., only selecting "fair" or "good" locations with >=50% listed species). In the Symbology tab, Show: Quantities; field value 'perc listed'; classes=5; and method=quantile. Export selected to .csv for Maxent model run.

#### 3.5 Create Percent Tolerant

The purpose of this step is to calculate the percent tolerant MCAT metric for each sampled site.

### 3.6 Create Percent Tribe Lampsilini

The purpose of this step is to calculate the percent tribe Lampsilini MCAT metric for each sampled site.

#### 3.7 Create Percent Juveniles

The purpose of this step is to calculate the percent juveniles MCAT metric for each sampled site.

#### 3.8 Create Percent Over 15 Years

The purpose of this step is to calculate the percent over 15 years old MCAT metric for each sampled site.

#### 3.9 Create Abundance

The purpose of this step is to calculate the Abundance MCAT metric for each sampled site.

• For the abundance metric, dissolve Cordova\_all\_individuals by SAMPLE\_ID and ENAME (Scientific name). Add 'SUM\_NumberLive' field in the Statistics section of the tool. Uncheck create multipart features. Set Output Feature Class to the working .gdb and name it Abundance.

# Import Abundance from geodatabase
abundance\_sp <- arc2sp("//mvrdfs/EGIS/Work/EMP/HREP\_Projects/SteamboatSlough/Mussels/SteamboatMusselMod</pre>

```
# Convert sp object to a data frame
abundance <- abundance_sp@data

# Calculate the quantiles
q <- quantile(abundance$SUM_NumberLive)
q

## 0% 25% 50% 75% 100%
## 0 1 1 2 33

# Create a new df num live >= 2nd quartile (50%)
q50abund <- abundance[abundance$SUM_NumberLive >= q[3], ]

# Create a new df num live >= 3rd quartile (75%)
q75abund <- abundance[abundance$SUM_NumberLive >= q[4], ]
```

#### 3.10 Create Species Evenness

The purpose of this step is to calculate the Pielou's evenness index (range 0-1), estimated at the species level for each sampled site.

- Create a table of SampleID by species by abundance (number live). Use the Dissolve tool:
  - Input Features: Cordova\_all\_individuals
  - Output feature class: Cordova\_sampleid\_species
  - Dissolve Fields: SampleID, EName
  - Statistics Fields: NumberLive, SUM
  - Create Multipart: unchecked

Ensure that the Cordova\_sampleid\_species feature class is of type Point Features and not Multipoint Features. If multipoint, then use the Multipart to Singlepart tool to convert from multipoint to point feature type.

```
# Import ArcGIS feature class into R
cordova_sampleid_species <- arc2sp("//mvrdfs/EGIS/Work/EMP/HREP_Projects/SteamboatSlough/Mussels/Steamb</pre>
# Convert sp object to a data frame
sample_species <- cordova_sampleid_species@data</pre>
# Remove OBJECTID field
sample_species <- sample_species[,-1]</pre>
# Convert to vegan community data matrix-like format using labdsv::matrify
sample_species_matrix <- labdsv::matrify(sample_species)</pre>
# Calculate the Shannon-Weaver diversity index
shannon_diversity <- vegan::diversity(sample_species_matrix, index = "shannon")</pre>
# Calculate the number species per site (see ?veqan::diversity)
species_number <- vegan::specnumber(sample_species_matrix)</pre>
# Use the following equation to calculate Pielou's evenness
# See the vegan Diversity Vignette for details
pielou_evenness <- shannon_diversity/log(species_number)</pre>
# Set NaN pielou_evenness values to zero
```

• Create a table of SampleID by species Pielou evenness score. Use the Dissolve tool:

pielou\_evenness <- ifelse(is.nan(pielou\_evenness), 0, pielou\_evenness)</pre>

- Input Features: sample\_species\_pielou
- Output feature class: species\_pielou
- Dissolve Fields: SampleID
- Statistics Fields: shannon\_diversity, MEAN species\_number, MEAN pielou\_evenness, MEAN
- Create Multipart: unchecked
- Add "SWD" fields.

#### 3.11 Create Tribe Eveness

The purpose of this step is to calculate the Pielou's evenness index (range 0-1), estimated at the tribe level. (will probably need to add a new column for tribe; dissolve on this with sample ID)

- In the feature class Cordova all individuals, create a new text variable named tribe.
- Use the Field Calculator tool on the tribe field to calculate its value using the following Python expression: !Ename!.split()[0]
- Create a table of SampleID by tribe by abundance (number live). Use the Dissolve tool:
  - Input Features: Cordova\_all\_individuals
  - Output feature class: Cordova\_sampleid\_tribe
  - Dissolve Fields: SampleID, tribe
  - Statistics Fields: NumberLive, SUM
  - Create Multipart: unchecked

```
# Import ArcGIS feature class into R
cordova_sampleid_tribe <- arc2sp("//mvrdfs/EGIS/Work/EMP/HREP_Projects/SteamboatSlough/Mussels/Steamboa
# Convert sp object to a data frame
sample_tribe <- cordova_sampleid_tribe@data
# Remove OBJECTID field
sample_tribe <- sample_tribe[,-1]
# Convert to vegan community data matrix-like format using labdsv::matrify
sample_tribe_matrix <- labdsv::matrify(sample_tribe)
# Calculate the Shannon-Weaver diversity index
tribe_shannon_diversity <- vegan::diversity(sample_tribe_matrix, index = "shannon")
# Calculate the number species per site (see ?vegan::diversity)</pre>
```

3.12. CREATE ES 100

```
tribe_number <- vegan::specnumber(sample_tribe_matrix)</pre>
# Use the following equation to calculate Pielou's evenness
# See the vegan Diversity Vignette for details
tribe_pielou_evenness <- tribe_shannon_diversity/log(tribe_number)</pre>
# Set NaN pielou_evenness values to zero
tribe_pielou_evenness <- ifelse(is.nan(tribe_pielou_evenness), 0,</pre>
                                 tribe_pielou_evenness)
# Create a data frame of results
tribe_pielou <- data.frame(sampleid = names(tribe_pielou_evenness),</pre>
                            tribe_shannon_diversity,
                            tribe number,
                            tribe_pielou_evenness)
# Merge the Pielou scores back onto the sp object
sample tribe pielou <- sp::merge(x = cordova sampleid tribe,
                                  y = tribe_pielou,
                                  by.x = "SampleID", by.y = "sampleid")
# Export to geodatabase
sp2arc(sample_tribe_pielou, fc_path = "//mvrdfs/EGIS/Work/EMP/HREP_Projects/SteamboatSlough/Mussels/Ste
```

- Create a table of SampleID by tribe Pielou evenness score. Use the Dissolve tool:
  - Input Features: sample\_tribe\_pielou
  - Output feature class: tribe\_pielou
  - Dissolve Fields: SampleID
  - Statistics Fields: tribe\_shannon\_diversity, MEAN tribe\_number, MEAN tribe\_pielou\_evenness, MEAN
  - Create Multipart: unchecked
- Add "SWD" fields.

#### 3.12 Create ES 100

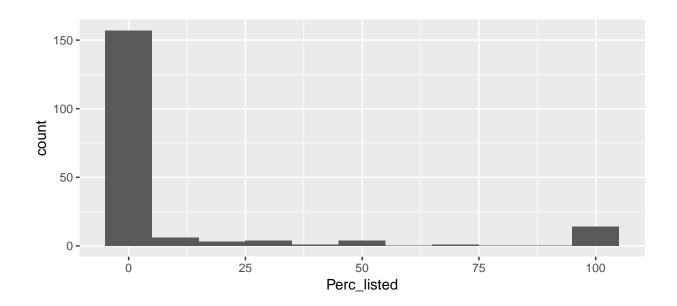
The purpose of this step is to calculate the ES 100 MCAT metric for each sampled site.

ES\_100-species richness estimated by rarefaction; caveat: sites need to be compared based on an equal sample size because # of species and # of individuals sampled are large correlated. Not sure how best to approach this one; I defer to you!

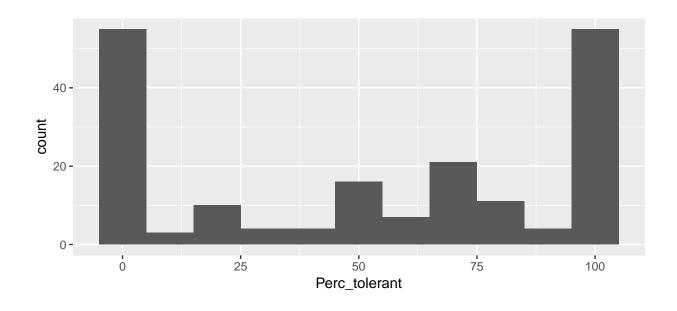
### 3.13 Determine MCAT Metric per site threshold

The purpose of this section is to display the range of calculated values for each of the MCAT metrics for each of the sites in the Steamboat Island study area. The MCAT report is clear about metric thresholds to use for evaluating quality mussel beds. However, this study is using the individual site unit of analysis. Therefore, decisions must be made on what is the appropriate metric threshold to use for the site level of analysis. The descriptive statistics in this section will help to inform that decision.

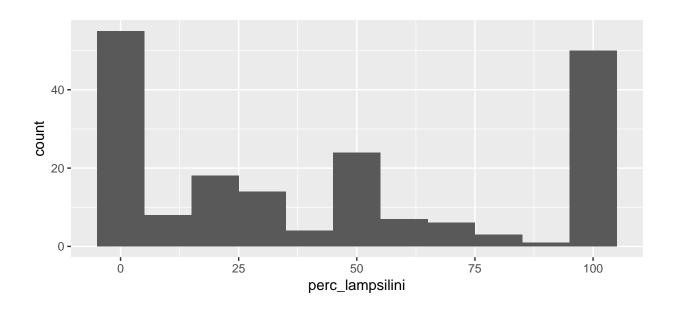
### 3.13.1 Percent Listed



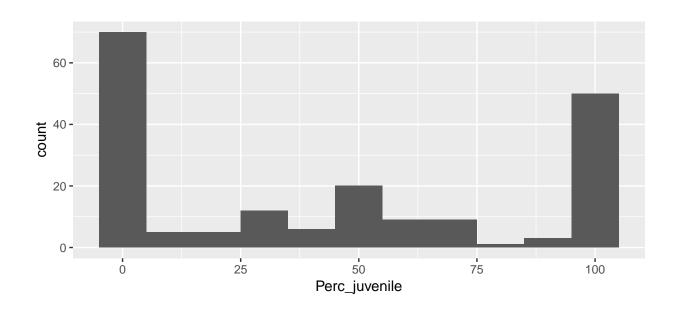
### 3.13.2 Percent Tolerant



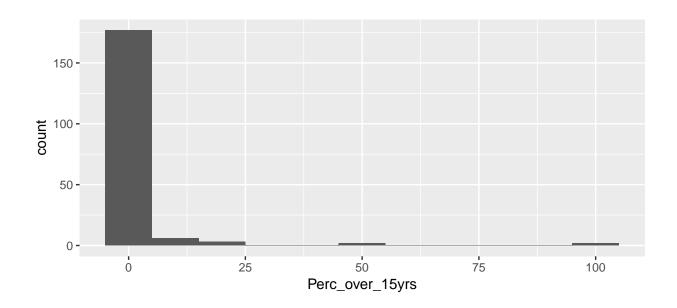
## 3.13.3 Percent Tribe Lampsilini



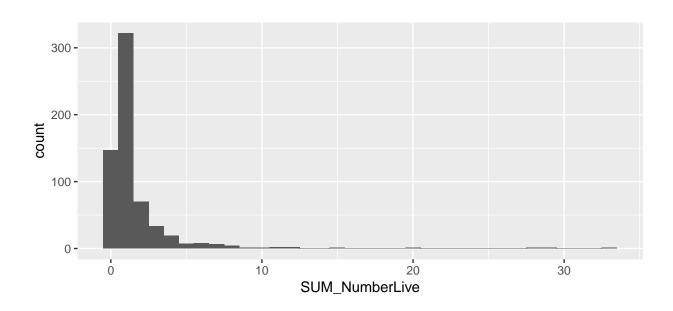
## 3.13.4 Percent Juveniles



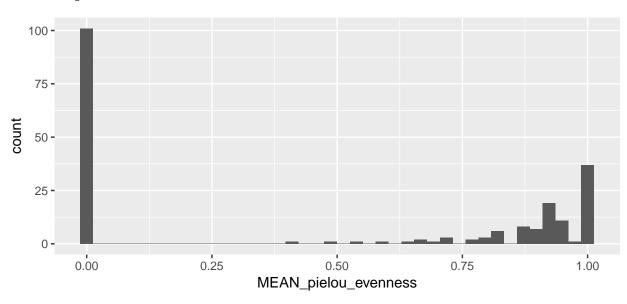
## 3.13.5 Percent over 15 years



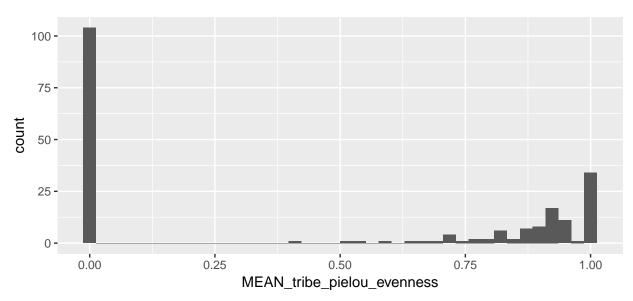
## 3.13.6 Abundance



### 3.13.7 Species Evenness



#### 3.13.8 Tribe Evenness



#### 3.13.9 ES\_100

### 3.14 Choice of Site Level MCAT Metric Thresholds

The discriptive statistics presented in the previous section were used to select MCAT metrics thresolds. Table 1.

## 3.15 Export to Maxent "SWD" format

## Create Backgound Data

### 4.1 Create Background Data

Create a background to constrain points to aquatic areas (more data resolution) and reduce Maxent run times.

- Use the Create Random Points (Data Management) tool to create a background restricted to aquatic
  areas:
  - Output location: \\mvrdfs\EGIS\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMusselMode
  - Output feature class: background
  - Constraining feature class:aquatic\_areas\_5m
  - Number of points: 10000
  - Minimum allowed distance: 25m
  - Leave all others as defaults
  - Click Ok to run the tool
- Open the background attribute table. Add 3 fields titled 'species' Type=string and Length=20; 'longitude' Type=double; and 'latitude' Type=double; in that order. Use field calculator to populate the field with species="background". Use Calculate Geometry to populate the 'longitude' and 'latitude' fields, ensuring that the coordinate system is set to 'NAD 1983 2011 US Feet' and 'StatePlane Illinois West FIPS 1202' and the units are set to US feet. Do the same for unionidae\_swd.
- Use Extract Multi Values to Points (spatial analyst) tool to write new values (velocity, depth, slope, and ss) onto feature classes unionidae\_swd and background.
  - a ss) onto leature classes unionidae\_swd and background.

     Input point features: \\mvrdfs\EGIS\Work\EMP\HREP Projects\SteamboatSlough\Mussels\SteamboatMusselN
  - Input rasters: velocity, depth, slope, and ss (use rasters in the Adh.gdb, not .bil) copy union-idae\_swd and delete 3 variable fields. Ensure all layers have same spatial reference before running the tool. Copy .gdb on local drive prior to running Maxent.
- Ensure all layers have same spatial reference before running the tool.
- May have to run Project (Data Mgmt) tool on the unionidae\_swd points feature class prior to running Multi Values to Points tool.
- Export background and each mussel metric as .csv files. Ensure ObjectID field is deleted in Excel prior to upload in Maxent.

# Prepare Environmental Predictor Data

#### 5.1 Create Mask

The goal of this step is to define no data areas outside of the study.

- Create a new polygon feature class named: \\mvrdfs\EGIS\\Work\EMP\\HREP\_Projects\\SteamboatSlough\\Mussels\\St
- Edit the 'study\_area' feature class by tracing one of the rasters created in the previous step (e.g., 'velocity'). The resulting polygon will become the clip feature for the next tool run.
- Use the Aquatic\_Areas\_2010 layer to prepare data for the Clip (Analysis) tool (\\mvrdfs\egis\Data\Layers\Biology
  - Input Features: Aquatic Areas 2010\Classifications\Aquatic Areas: Level 2
  - Clip Features: study\_area
  - Output Feature Class: \\mvrdfs\EGIS\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMusse
- Delete "Isolated Floodplain Lake" records from the aquatic\_areas attribute table.
- Use the Dissolve (Data Management tool) aquatic\_areas using 'LAND\_WATER' as the dissolve field.
  - iu.

- Output feature class: \mvrdfs\EGIS\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMussell

- Create multipart features: uncheck box
- Open an edit session to edit the aquatic\_areas\_dissolve attribute table. Select the row where 'LAND WATER' field is set to Land, and delete.
- Run the Multipart to Singlepart Tool (Data Management) on aquatic\_areas\_dissolve. Name the feature class aquatic\_areas\_dissolve\_singlepart.
- Turn on topobathy and use it as a guide to edit out areas where where it is too shallow in the aquatic\_areas\_dissolve\_singlepart feature class. Be sure to merge the polygons before saving edits and delete intermediates aquatic\_areas\_dissolve and aquatic\_areas\_dissolve\_singlepart.
- Use the Buffer (Analysis) tool to expand the study area beyond the extent of valid values. This expanded area will be used to define no data portions of the modeled Adh values.
  - Input features: aquatic\_areas
  - Output Feature Class: \mvrdfs\egis\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMusse
  - Distance: 20 meters

- Use the Polygon to Raster (Conversion) tool to convert the aquatic\_areas\_20m feature class to a
  raster.
  - Input Features: aquatic\_areas\_20m
  - Value field: OBJECTID
  - Output Raster Dataset: \mvrdfs\egis\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMuss
- Set Environments
  - Ouput Coordinates: NAD\_1983\_2011\_StatePlane\_Illinois\_West\_FIPS\_1202\_Ft\_US
  - Processing Extent: study area
  - Cell Size: 3 ft

#### 5.2 Set Environments

The purpose of this step is to set the Environments to mask so that all the subsequent Adh raster outputs will have the same coordinate system, extent, cell alignment, cell size, and no data area. \*Set Environments

```
* Ouput Coordinates: `mask`

* Processing Extent: `mask`

* Snap raster: `mask`

* Cell Size: `mask`
```

\* Mask: `mask`

#### 5.3 Process ADH Model Outputs

The goal of this step is to convert the raw Adaptive Hydraulics Model (Adh) data (depth, slope velocity, sheer stress, froude\_number, reynolds\_number) into raster format that can be input to the Maxent model. Adh models will be produced for the Q5 (high flow) and Q95 (low flow) flow conditions. This wil result in the creation of 12 raster (one for each of 6 Adh parameter) by 2 flow conditions.

- Create a new folder and geodatabase to represent the alternative that this Adh model represents. In this example we will be modeling the existing condition alternative: \\mvrdfs\egis\\Work\EMP\\HREP\_Projects\SteamboatS
- Get the Adh output from the hydraulic engineer. Copy it to the alternative folder you created in the last step. EC-HH Working directory: \mvr-netapp2\MVRDATA\ED\ec-hh\Mississippi\_River\Steamboat Island\AdH\Model Runs\
- Use Excel to convert the space-delimited .txt file to .csv. Open a blank Excel workbook and open the .txt file. Use the import wizard to convert the space-delimited file into an Excel workbook. Ensure the import works correctly, delete any header information above the row of field names, and then save as a .csv file.
- Import the .csv file into the existingcondition geodatabase. Name the table by the flow condition it represents (Q5, Q95).
- Add the Q5 table into a blank ArcMap document.
- Right-click the Q5 table and click "Display XY Data".
- Set appropriate 'X', 'Y', and 'Z' fields.
- Identify the datum and coordinate system recorded in the Adh table; all future output feature classes and rasters will need to be in the same projected coordinate system. In this example, Datum: 'NAD 1983 2011 US Feet' and Coordinate System: 'StatePlane Illinois West FIPS 1202'.
- Export Q5\_Events as new point feature class named Q5\_pts.
- Use the IDW (Spatial Analyst) tool to interpolate a raster surface.

- Input point features: Q5\_pts.
- Z value field: velocity
- Output raster: \\mvrdfs\EGIS\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMusselModel
- Complete these steps for the other variables, 'depth', 'slope', and 'sheer stress', 'froude\_number', 'reynolds number'.

### 5.4 Convert to .bil format

The goal of this step is to convert the Adh output into a format that can be used by Maxent. The .bil format is one of the formats common to both Maxent and Esri. This step will repeated for all Adh (Q5, Q95) predictors.

- Use the Copy Raster (Data Management) tool to covert the Adh predictor rasters into .bil format.
  - Input Raster: q5\_velocity
  - Output Raster Dataset: \mvrdfs\egis\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMuss (Note: cannot save .bil format within the .gdb)
  - Format: Esri BIL

### 5.5 Calculate Slope

The goal of the step is to calculate slope from the Adh depth variable created in the previous step.

- Use the Slope (Spatial Analyst) tool to calculate slope.
  - Input raster: q5\_depth
  - Output raster: \mvrdfs\egis\Work\EMP\HREP\_Projects\SteamboatSlough\Mussels\SteamboatMusselModel
  - Output measurement: degree
  - Z factor = 1 (linear units same)
- Use Copy Raster (Data Management) tool to convert the q5\_slope raster into .bil format, as outlined in previous step.

# Prepare Wind and Wave Data

- 6.1 Identify Appropriate Weather Stations
- 6.2 Obtain Wind Data
- 6.3 Determine Period of Record

# Develop Maxent Mussel Model

#### 7.1 Convert from .bil to .mxe format

To speed model creation (from 10's of hours to 10's of minutes), convert the predictors to the Maxent .mxe format.

- Open command prompt and change directory by typing c: and then cd C:\Users\b6pdpdem\Documents\SteamboatMus
- Create predictor\_mxe folder

 $\label{thm:convert} \verb| C:\Users\b6pdpdem\Documents\SteamboatMusselModel\predictors bil C:\Users\b6pdpdem\Documents\SteamboatMusselModel\predictors\_mxe mxe | mxe$ 

#### 7.2 Run Maxent

### 7.3 Convert model output from

To convert the .mxe format back to a format that can be read by ESRI.

java -Xmx5g -cp maxent.jar density.Convert C:\Users\b6pdpdem\Documents\SteamboatMusselModel\sb\_existing
mxe C:\Users\b6pdpdem\Documents\SteamboatMusselModel\sb\_existing\_20181203\_2 bil

• Use the Define Projection tool to set the projection of the .bil rasters.

### 7.4 Maxent speed Issues

Follow these simple performance guidelines to reduce Maxent run time from a week to half an hour.

- Run the model on a physical computer, not on a virtual machine.
- Select a computer with the fastest processor and most memory available.
- Increase the memory available to the java virtual machine (JVM). Edit the first line of the maxent.bat file to something like the following: java -Xmx5g -jar maxent.jar, where the -Xmx parameter sets the maximum heap size for the JVM. Set it to something less than the total amount of physical memory on your computer.
- Do not run the model across the LAN (inputs and outputs stored on a network file system). Copy all inputs and write all outputs to the local file system.
- Use a solid state drive for storing inputs and outputs.
- Use the "SWD" format for specifying both samples (species occurances) and environmental layers (background points).
- Convert all predictor grids to the Maxent .mxe format.

- $\bullet~$  Write all outputs to the Maxent .mxe format.
- Set the threads parameter to at least the number of cores on your computer (or a large percentage of logical processors) to speed processing of replicates.

# Model Interpretation

8.1 ?