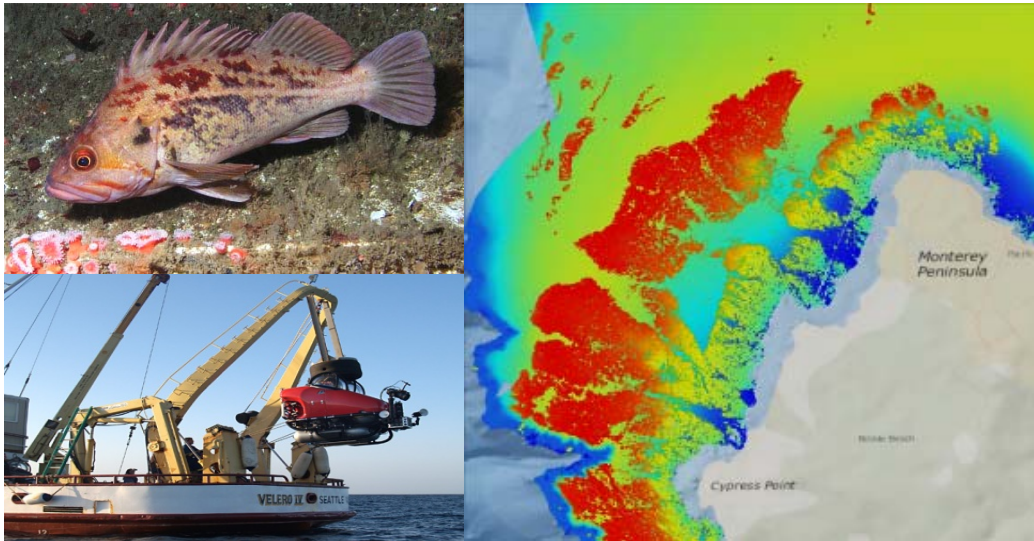


Species Distribution Modeling



ESM 296-4F Advanced Topics in GIS October 2014

Mary Young¹, Lisa Wedding², Jason Roberts³

¹UC Santa Cruz, Ecology and Evolutionary Biology Department

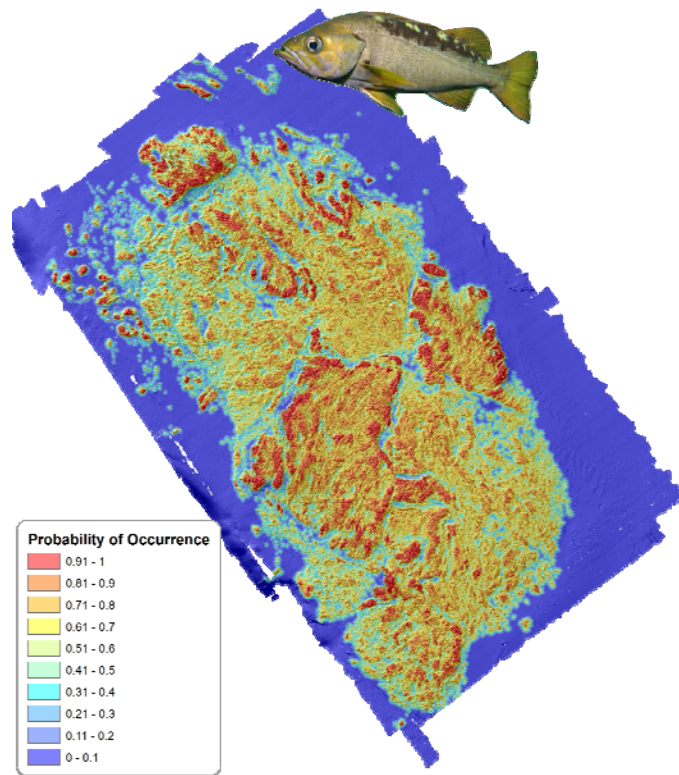
²Stanford University, Center for Ocean Solutions

³Duke University, Marine Geospatial Ecology Lab



**STANFORD
UNIVERSITY**

Species Distribution Modeling of Yellowtail Rockfish



GOALS:

1. Learn modeling workflow using the MGET (Marine Geospatial Ecology Tool).
2. Apply a GLM and GAM to Yellowtail rockfish presence/absence model.
3. Create a spatially predictive map.
4. Assess accuracy of predictive map products.

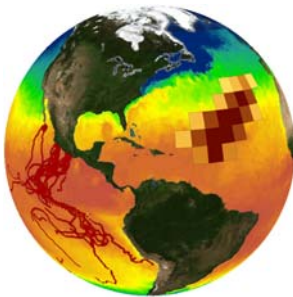
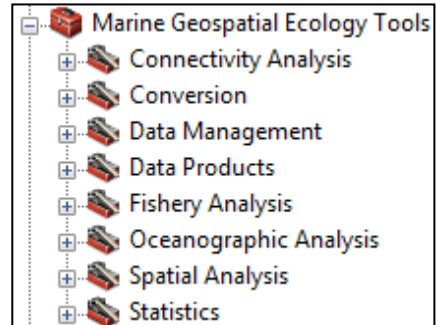
MATERIALS:

- Computers with access to internet
- ArcMap 10.1 with Spatial Analyst and MGET add ins installed
- Exercise datasets

MODULE 1: Species Distribution Modeling (GLM)

Overview:

This exercise will guide you through spatially predictive modeling using P/A data for yellowtail rockfish. We will be using the MGET tool, an ArcGIS toolbox created by the Marine Geospatial Ecology Lab at Duke University (<http://mgel.env.duke.edu/mget>). Marine Geospatial Ecology Tools (MGET) is a free, open-source geoprocessing toolbox that can help you solve a wide variety of marine research, conservation, and spatial planning problems. MGET can perform tasks such as:



- Downloading and converting oceanographic data into GIS-compatible formats
- Identifying ecologically-relevant features in remote sensing imagery
- Fitting, evaluating, and mapping statistical models of species habitat
- Modeling habitat connectivity by simulating the dispersal of larvae

For detailed information on the MGET tool please review the below links:

1. <http://swfsc.noaa.gov/textblock.aspx?Division=FED&ParentMenuId=54&id=18070> contains a PowerPoint that gives a quick tour of the toolbox.
2. The downloadable habitat modeling example (<http://code.nicholas.duke.edu/projects/mget/wiki/HabitatModeling>) illustrates what might be two of the most popular applications of MGET: accessing oceanographic data from ArcGIS and building a species distribution model from ArcGIS using MGET's integration with R.
3. The main MGET paper (<http://dx.doi.org/10.1016/j.envsoft.2010.03.029>) gives an overview of the software architecture of the toolbox and gives a habitat modeling example similar to #2 above. If you use MGET in your research cite the following paper: (Roberts, JJ, Best BD, Dunn DC, Trembl EA, Halpin PN (2010) Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software* 25: 1197-1207.)
4. Also, <http://mgel.env.duke.edu/mget/documentation/arcgis-tutorial/> contains a list of scenarios. This is not comprehensive but does cover many of the common applications.

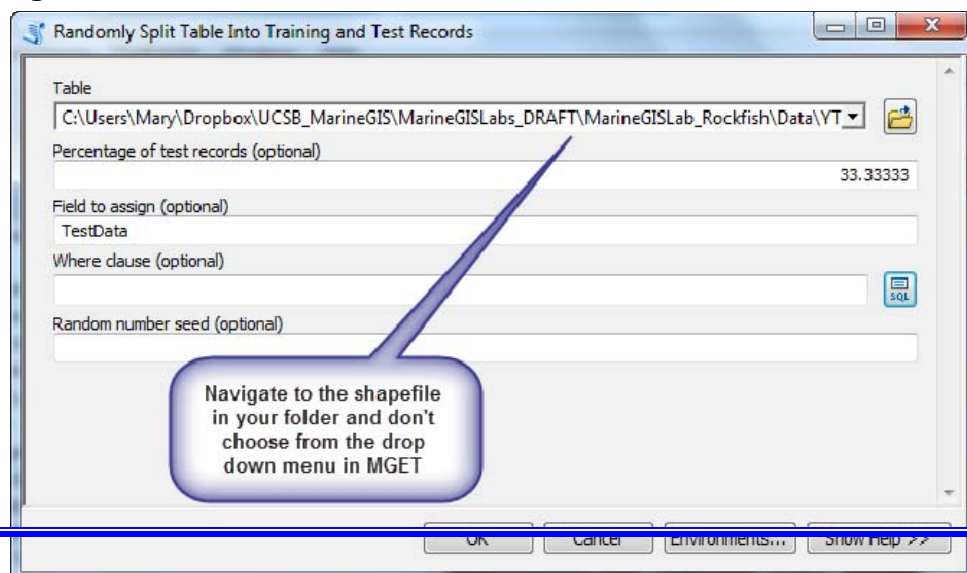
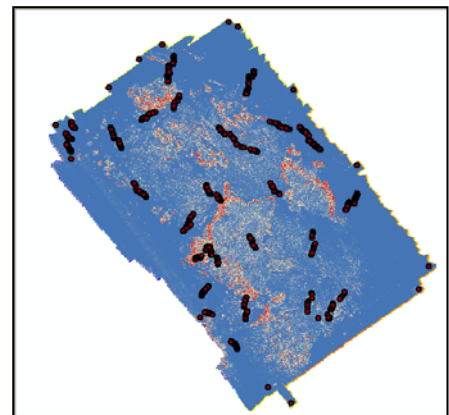
Background – Yellowtail Rockfish

Pacific coast rockfishes (*Sebastes* spp.) are a wide ranging and diverse genus that have extremely vulnerable life-history characteristics as they are generally long-lived, slow growing and slow to reach sexual maturity. Due to these reproductive strategies rockfishes cannot sustain high levels of fishing mortality. In order to best manage and assess stocks of rockfish species a better understanding is needed of the habitat variables that are driving distribution and abundance of these species. Species distribution modeling and spatially predictive mapping can support best management practices.

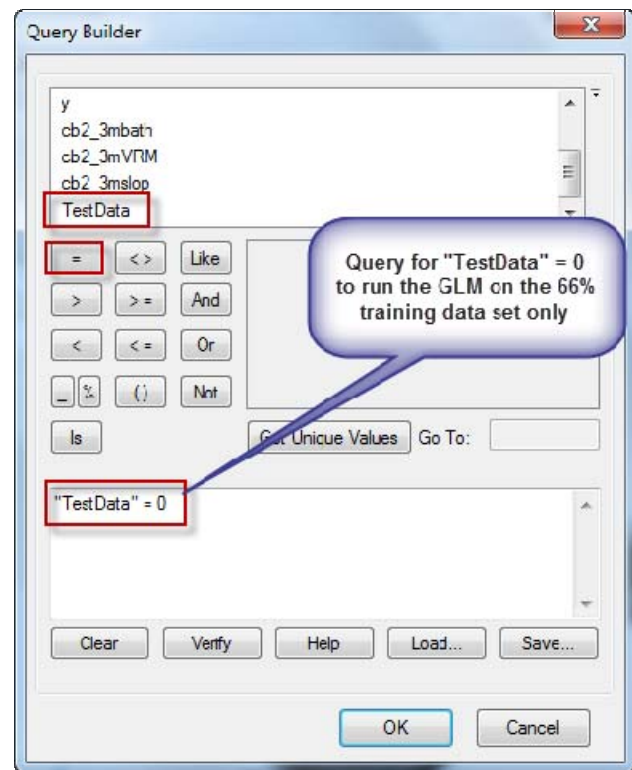


Yellowtail (*Sebastes flavidus*) rockfishes have common depth ranges in this study region from approximately 50 to 250 meters, frequently occurring in boulder fields and complex rock and cobble habitat. We will run through GLM (Generalized Linear Models) and GAM (Generalized Additive Models) examples with yellowtail rockfish data to understand the habitat variables driving the presence of this species in Cordell Bank.

1. Open the MarineGISLab_Rockfish.mxd file and view the ArcMap project with 3 raster baselayers (i.e. depth, slope, VRM) and rockfish observation data in a point shapefile. Right click on the YT_PresAbs_all.shp file and select 'Open Attribute Table' and review the data set.
2. Open the symbology in layer properties and change 'presence' of 1 to red dots and 'presence' of 0 to black dots to visually assess the presence/absence data for this species.
3. First we are going to split up the Rosy rockfish data into test and training data sets. This will designate 649 rows (2/3) as training records and 324 rows (1/3) as test records. Open the MGET tool/Statistics/Randomly Split Table Into Training and Test Records.



- In the query builder select 'TestData' = 0 and select ok. We will use 2/3 of the data to run our model and 1/3 of the data to test the accuracy of our predictive map output.



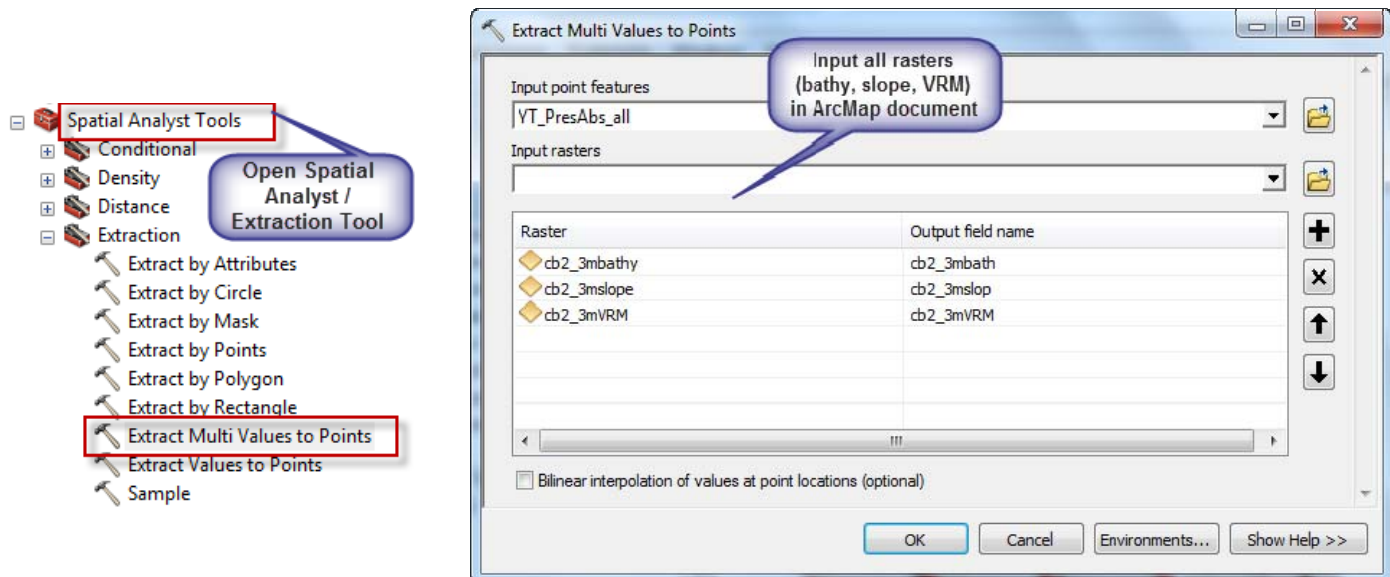
- Right click on the YT_PresAbs_all.shp file and open up the table to view the new TestData column that was created.

YT_PresAbs_all							
	presence	x	y	cb2_3mbath	cb2_3mVRM	cb2_3mslop	TestData
	0	454439.652923	4211844.96464	-180.38	0.000125	5.75555	0
	0	454448.302445	4211822.72479	-180.51	0.00005	6.30015	0
	0	454453.06466	4211811.38051	-180.85	0.000056	7.67843	0
	0	454454.659451	4211389.9546	-213.88	0.000679	16.2244	0
	0	454456.685711	4211799.93165	-180.67	0.000058	7.92774	0
	0	454463.371467	4211378.8102	-212.89	0.000084	16.7962	0
	0	454466.774408	4211784.00837	-180.62	0.000093	7.66381	0
	0	454472.083508	4211367.66581	-211.77	0.000491	15.4027	0
	0	454474.251299	4211756.0053	-180.55	0.000201	8.21259	0
	0	454474.313239	4211767.10072	-179.96	0.00002	6.74254	0
	0	454483.024868	4211755.95633	-179.57	0.000184	6.08432	1
	0	454489.445751	4211334.28164	-210.58	0.00022	15.1172	1
	0	454489.507661	4211345.37705	-209.46	0.000273	13.2576	0
	0	454495.945633	4211712.94362	-179.98	0.000146	8.71145	0
	0	454497.149251	4211771.41158	-177.04	0.000042	7.09174	1
	0	454503.933178	4211697.80882	-180.16	0.000309	7.88626	1
	0	454507.537328	4211714.76523	-178.45	0.000235	7.49189	0

1 (0 out of 973 Selected)

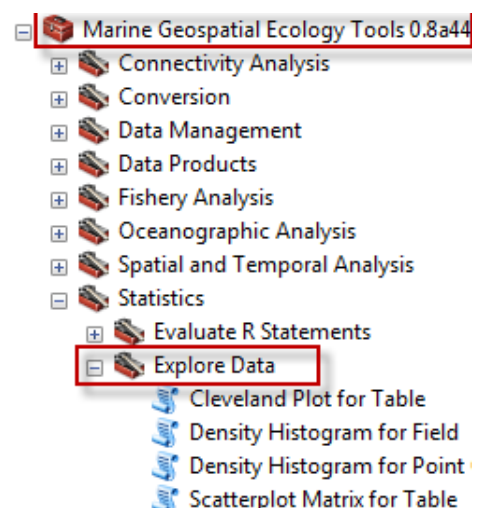
YT_PresAbs_all

- Next open the Spatial Analyst toolbox/Extraction tool/Extract Multi Values at Points. Select YT_PresAbs_all.shp for the input point feature and add all the raster data sets in the ArcMap document (i.e. bathy, slope and VRM). This tool will attribute the depth, slope and VRM value for each sampling location in the fish data set. These will serve as our independent variables in the subsequent model.



Data Exploration: The first step of any analysis should be a data exploration. For instance, it is important to look for outliers, collinearity in the explanatory variables, and relationships between response and explanatory variables. Why? Data exploration avoids type I and II errors and reduces the chance of incorrect ecological conclusions which is important for providing model results to support robust management and policy decisions. See the following paper for a detailed discussion of data exploration and supporting R code: (Zuur, A.F., Ineo, E.N., Elphick, C.S. (2009) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology & Evolution* 1: 3-14.)

- There is a subset of data exploration tools in MGET under the 'Explore Data' menu.
- Select 'Explore Data' and 'Cleveland Plot for Table'. Create a Cleveland dotplot for all three explanatory variables and review the dot plot and choose an appropriate data transformation if necessary. Cleveland dotplots are typically used to explore the distributions of values in a table and detect possible outliers. Each dotplot shows the row number of the record vs. the value of a field. Points that appear far to the left or right are extreme values of that field and may be statistical



outliers that should be investigated and possibly removed prior to further analysis of the data.

8. In reviewing the below Cleveland dotplot a transformation is necessary for slope and VRM. MGET is running R in the background and each entry is an R expression that operates on the vector *d* that represents the original values of that field. For example, the following expressions perform natural logarithm, base 10 logarithm, square root, and cube root transforms, respectively:

$\log(d)$
 $\log_{10}(d)$
 $d^{(1/2)}$
 $d^{(1/3)}$

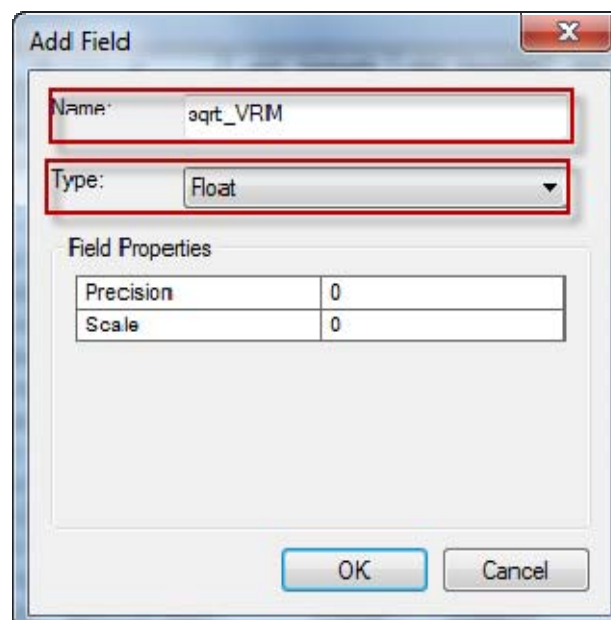
For example, if the input data represent water depth expressed as a negative number (e.g. -100 is 100 meters below the surface), and you want to transform the data using a base 10 logarithm, you must first take the absolute value of depth:

$\log_{10}(\text{abs}(d))$

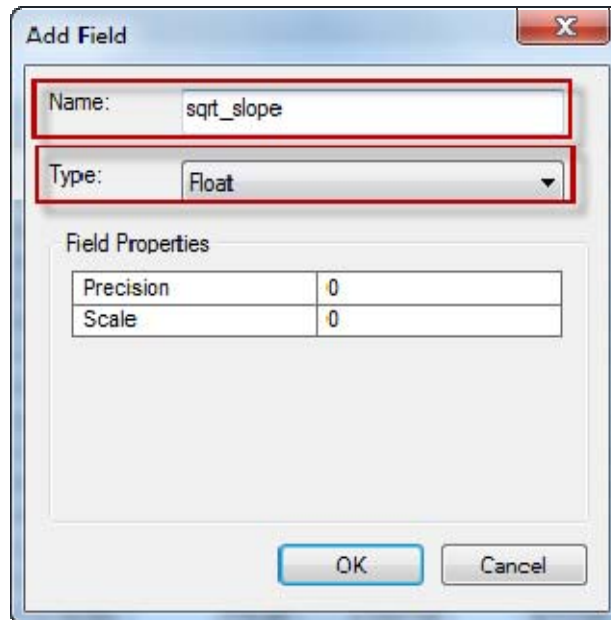
When you do not want to transform a field but are forced to provide a value, simply use *d* as the expression. For example, if you want to plot three fields--Depth, SST, and Chlorophyll--and want to \log_{10} transform just Depth and Chlorophyll, provide these expressions:

$\log_{10}(d)$
d
 $\log_{10}(d)$

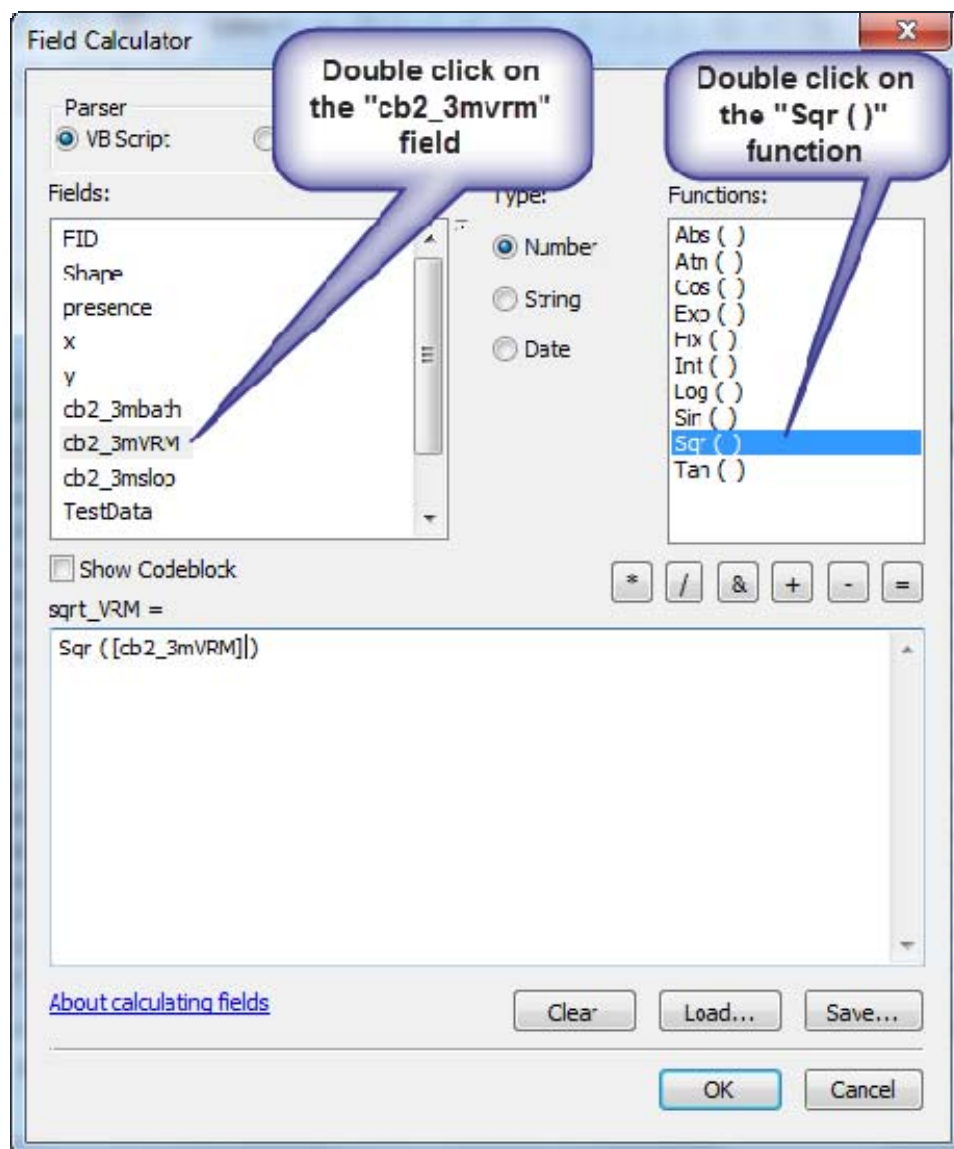
8. We are going to use field calculator to square root transform VRM and slope.
- Right click on the YT_PresAbs_all.shp point shapefile in the table of contents and choose "Open Attribute Table"
 - Add a new field (Table Options → Add Field...)
 - Name: Type in "sqrt_VRM"
 - Type: Select "Float" from the drop down menu



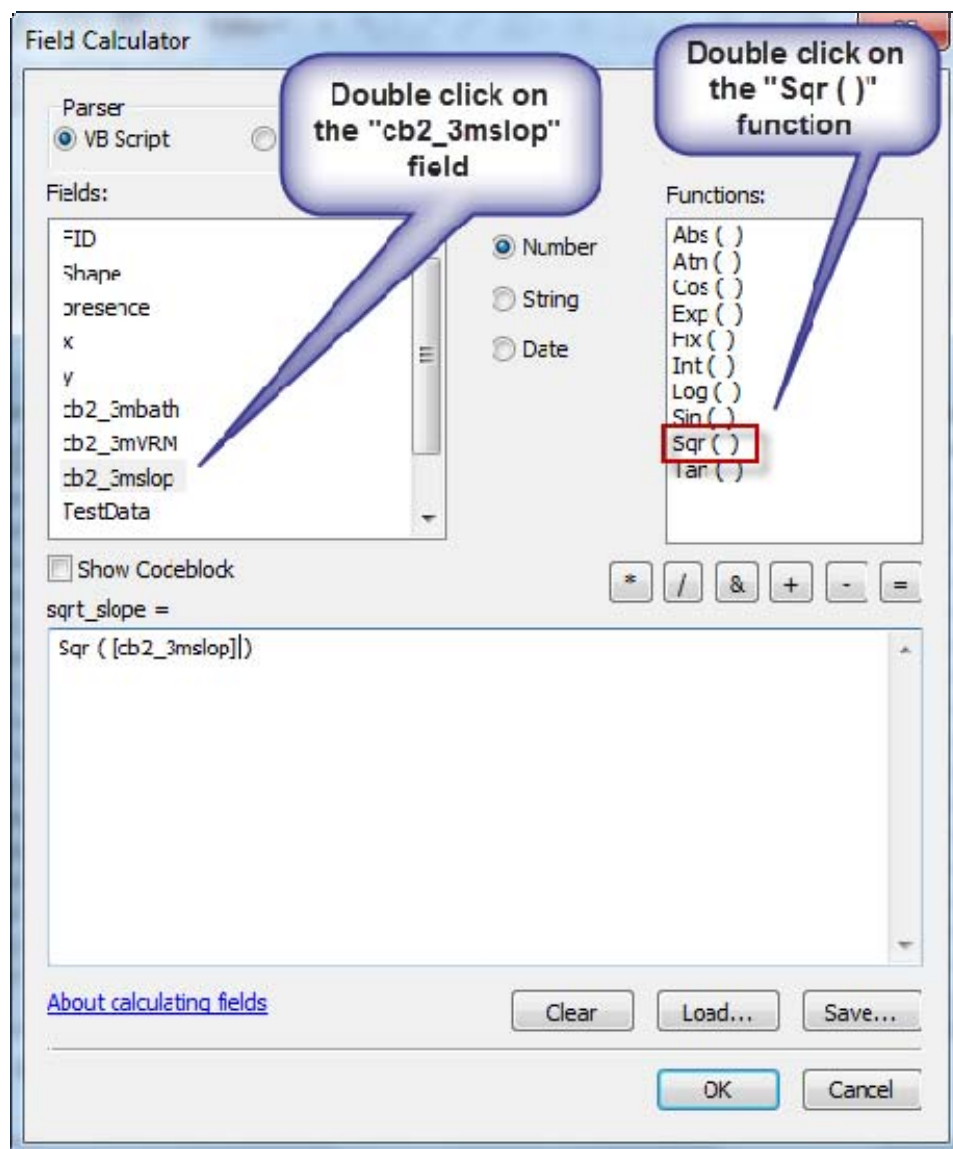
- ii. Add another new field (Table Options → Add Field...)
 1. Name: Type in “sqrt_slope”
 2. Type: Select “Float” from the drop down menu



- iii. Now we need to calculate the square root of each of slope and VRM so that we can use the transformed variables in the model.
 1. Right click on the “sqrt_VRM” field in the attribute table and choose “Field Calculator...”
 - a. Functions: Double click on the “Sqrt ()” function in the list to add it to the equation
 - b. With the cursor in the parentheses of the “Sqrt ()” function, double click on “cb2_3mvrn” in the Fields list.
 - c. Click OK and wait for the process to complete



2. Right click on the “sqrt_slope” field in the attribute table and choose “Field Calculator...”
 - a. Functions: Double click on the “Sqr ()” function in the list to add it to the equation
 - b. With the cursor in the parentheses of the “Sqr ()” function, double click on “cb2_3mslop” in the Fields list.
 - c. Click OK and wait for the process to complete



3. Both of those fields should now be populated with data.

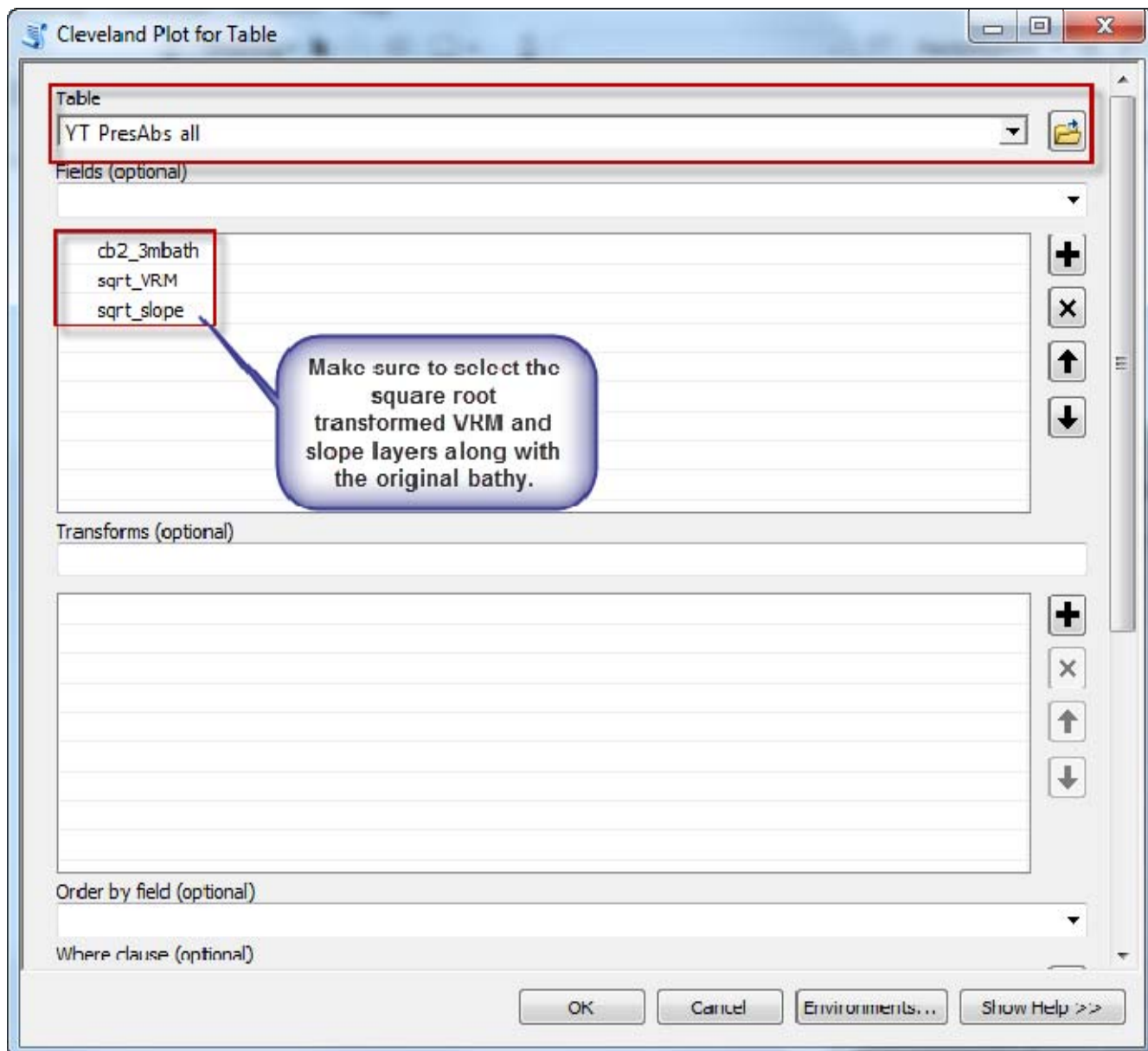
Shape *	presence	x	y	cb2_3mbath	cb2_3mVRM	cb2_3mslop	TestData	sqrt_VRM	sqrt_slope
Point	0	454439.652923	4211844.96464	-180.38	0.000125	5.75555	0	0.011172	2.39907
Point	0	454448.302445	4211822.72479	-180.51	0.00005	6.30015	0	0.007072	2.51001
Point	0	454453.06466	4211811.38051	-180.85	0.000056	7.67843	0	0.007481	2.771
Point	0	454454.659451	4211389.9546	-213.88	0.000679	16.2244	0	0.026064	4.02795
Point	0	454456.685711	4211799.93165	-180.67	0.000058	7.92774	0	0.007612	2.81562
Point	0	454463.371467	4211378.8102	-212.89	0.000084	16.7962	0	0.009184	4.09832
Point	0	454466.774408	4211784.00837	-180.62	0.000093	7.66381	0	0.009655	2.76836
Point	0	454472.083508	4211367.66581	-211.77	0.000491	15.4027	0	0.022155	3.92463
Point	0	454474.251299	4211756.0053	-180.55	0.000201	8.21259	0	0.014175	2.86576
Point	0	454474.313239	4211767.10072	-179.96	0.00002	6.74254	0	0.004508	2.59664
Point	0	454483.024868	4211755.95633	-179.57	0.000184	6.08432	1	0.013547	2.46664
Point	0	454489.445751	4211334.28164	-210.58	0.00022	15.1172	1	0.014847	3.88808
Point	0	454489.507661	4211345.37705	-209.46	0.000273	13.2576	0	0.016524	3.6411
Point	0	454495.945633	4211712.94362	-179.98	0.000146	8.71145	0	0.012084	2.95152
Point	0	454497.149251	4211771.41158	-177.04	0.000042	7.09174	1	0.006501	2.66303
Point	0	454503.933178	4211697.80882	-180.16	0.000309	7.88626	1	0.017578	2.80825
Point	0	454507.537328	4211714.76523	-178.45	0.000235	7.49189	0	0.015338	2.73713

9. We can now create the Cleveland dotplots again with the new, transformed values in the table.

a. Within ArcToolbox go to Marine Geospatial Ecology Tools 0.8a51 → Statistics → Explore Data and choose “Cleveland Plot for Table”

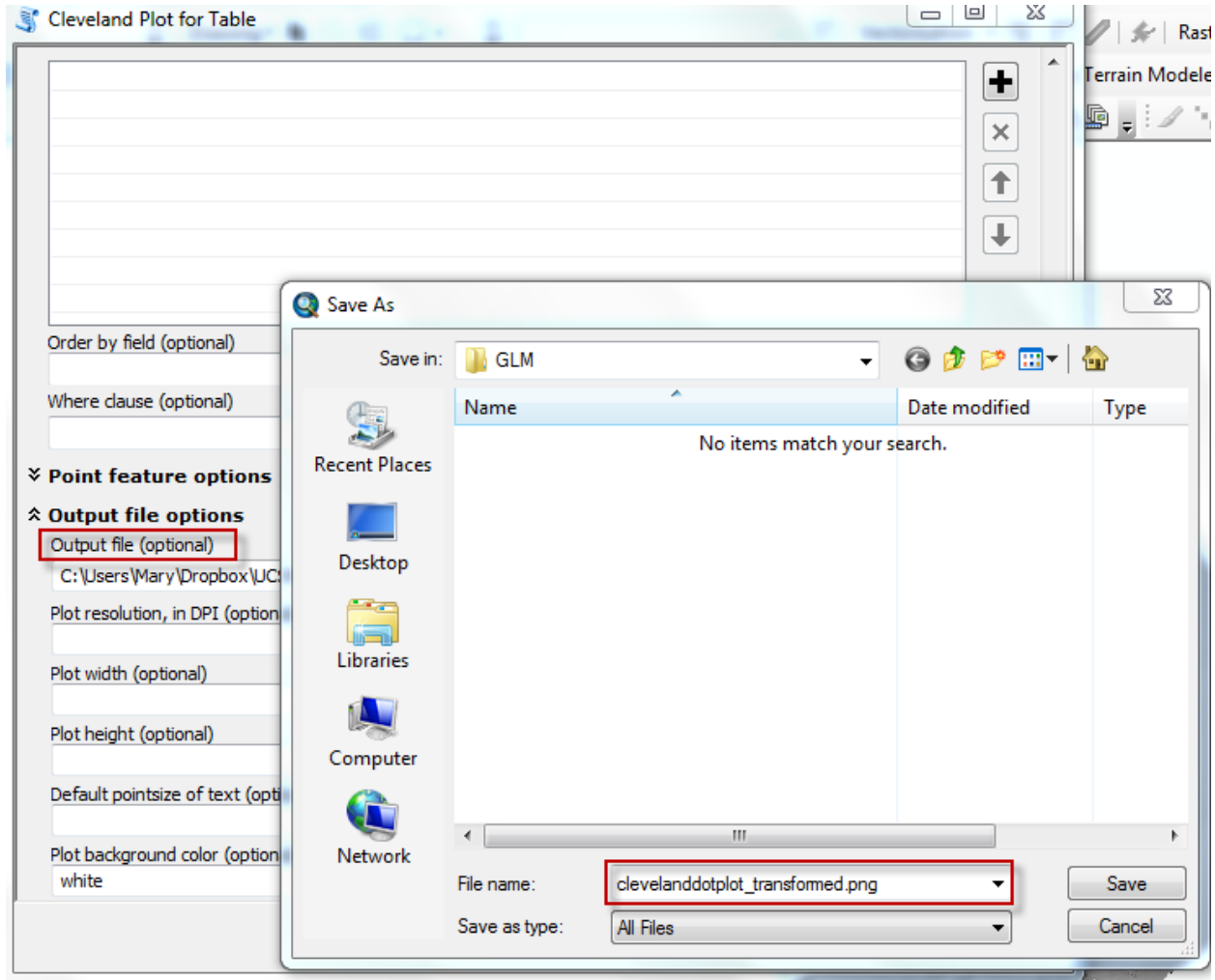
i. Table: Choose the “YT_PresAbs_all” shapefile from the drop down menu.

ii. Fields: Select the “cb2_3mbath” field, the “sqrt_VRM” field, and the “sqrt_slope” field and they will appear in the list.

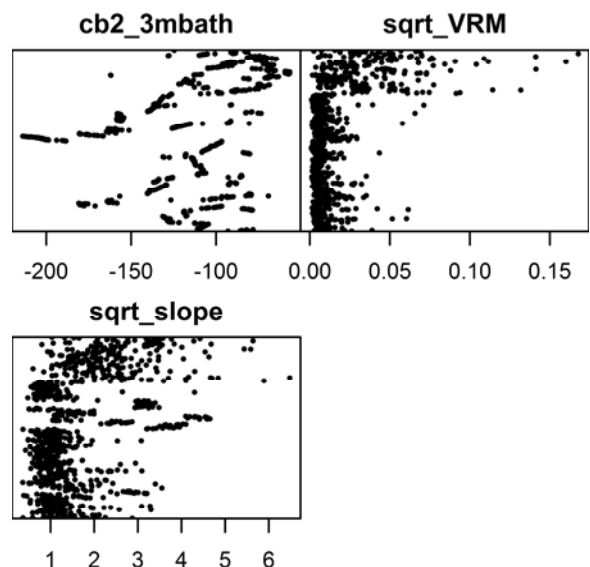


iii. You also have the option to save the resulting figure:

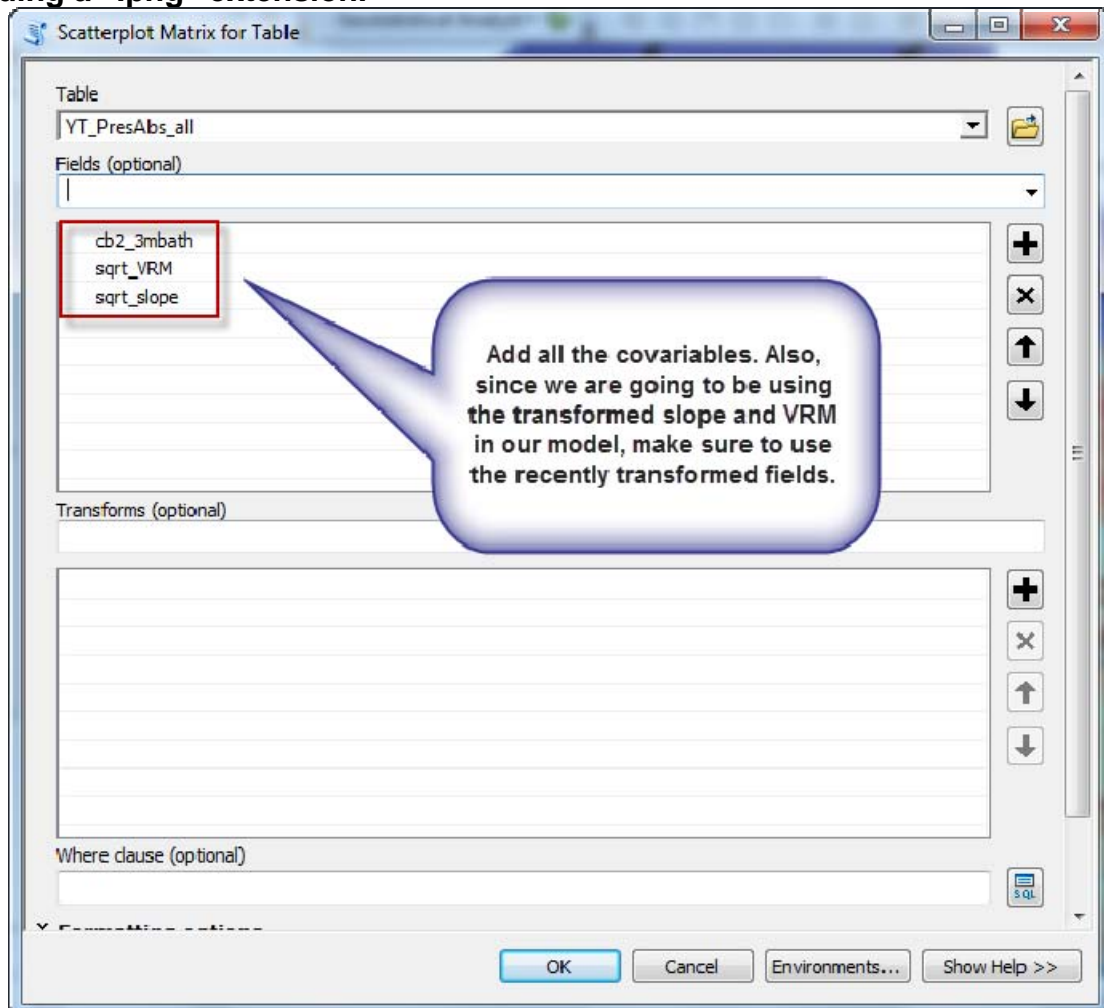
1. Expand the “Output file options” section of the dialog box and under “Output file” navigate to the “DataExploration” within the folder for this lab and save the Cleveland plot as “clevelanddotplot_transformed.png”



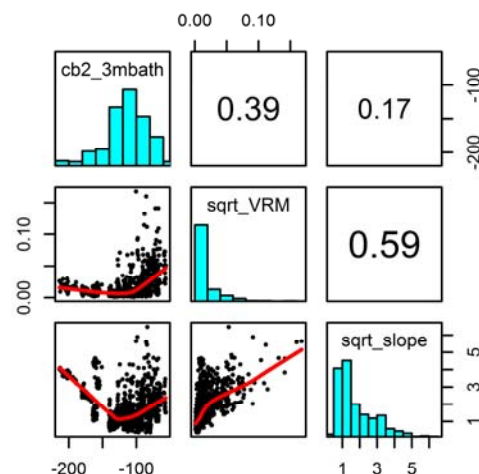
b. You can now navigate to the location where you saved the plots and view the results (shown to the right). You can see that the outliers are less pronounced. We are going to go ahead and move forward with the analysis of these transformed variables.



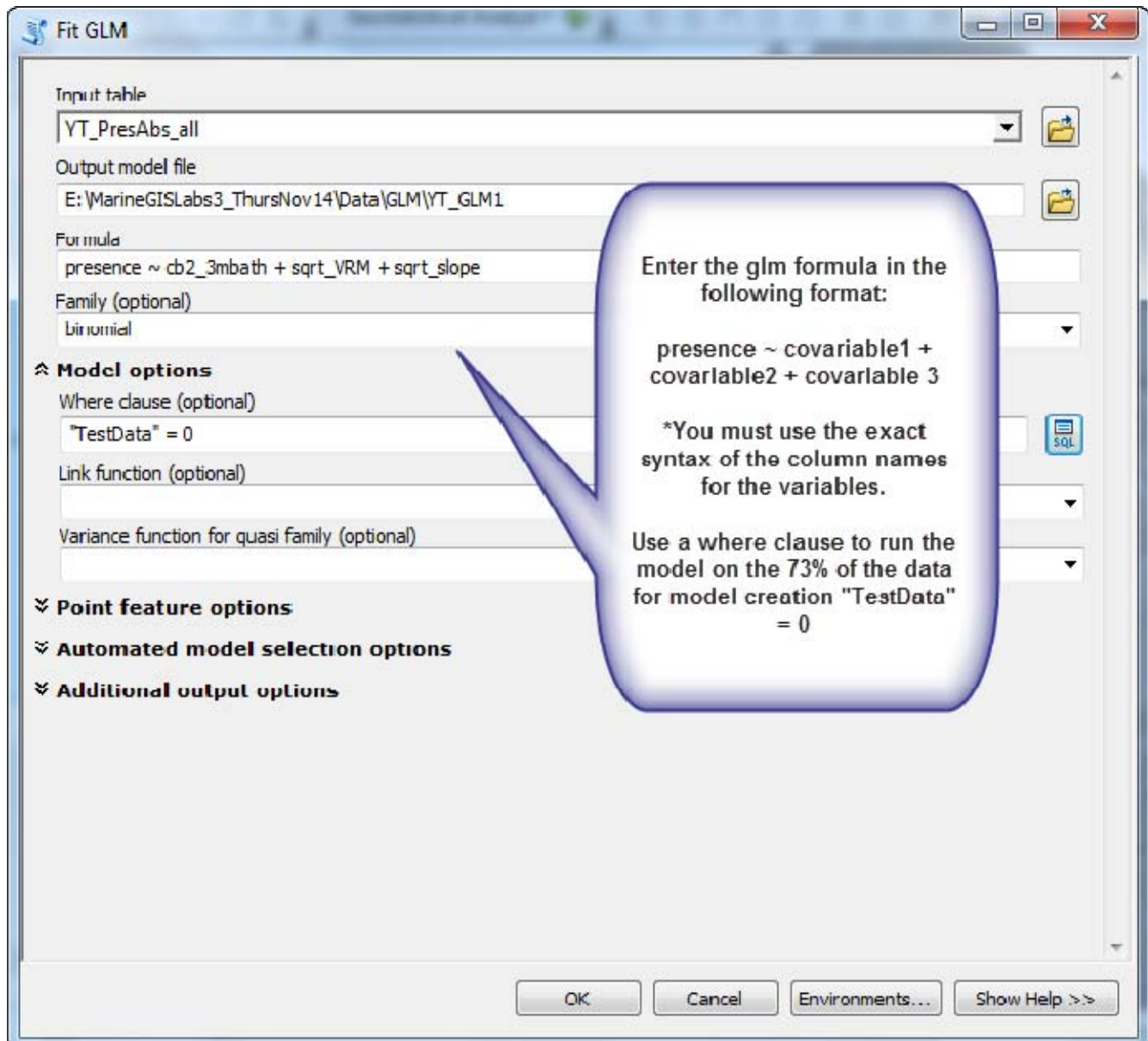
10. Within MGET run the Explore Data → Scatterplot Matrix for Table tool on the explanatory variables and review the output of the multi-panel scatter plot and Pearson correlation coefficients. If you want to save a copy of the scatterplot, you can expand “output file options” and save a copy of the scatterplot into the “DataExploration” folder. If you do save your scatterplot, which is recommended, you need to add an extension to the file. I would suggest adding a “.png” extension.



11. This data set demonstrates collinearity between slope and VRM (a measure of habitat complexity) at 0.59. Variance inflation factor (VIF) values can also compliment this evaluation of collinearity and can be calculated in R. We will move forward with this example using all explanatory covariables in the model based on prior evaluation of VIF values.



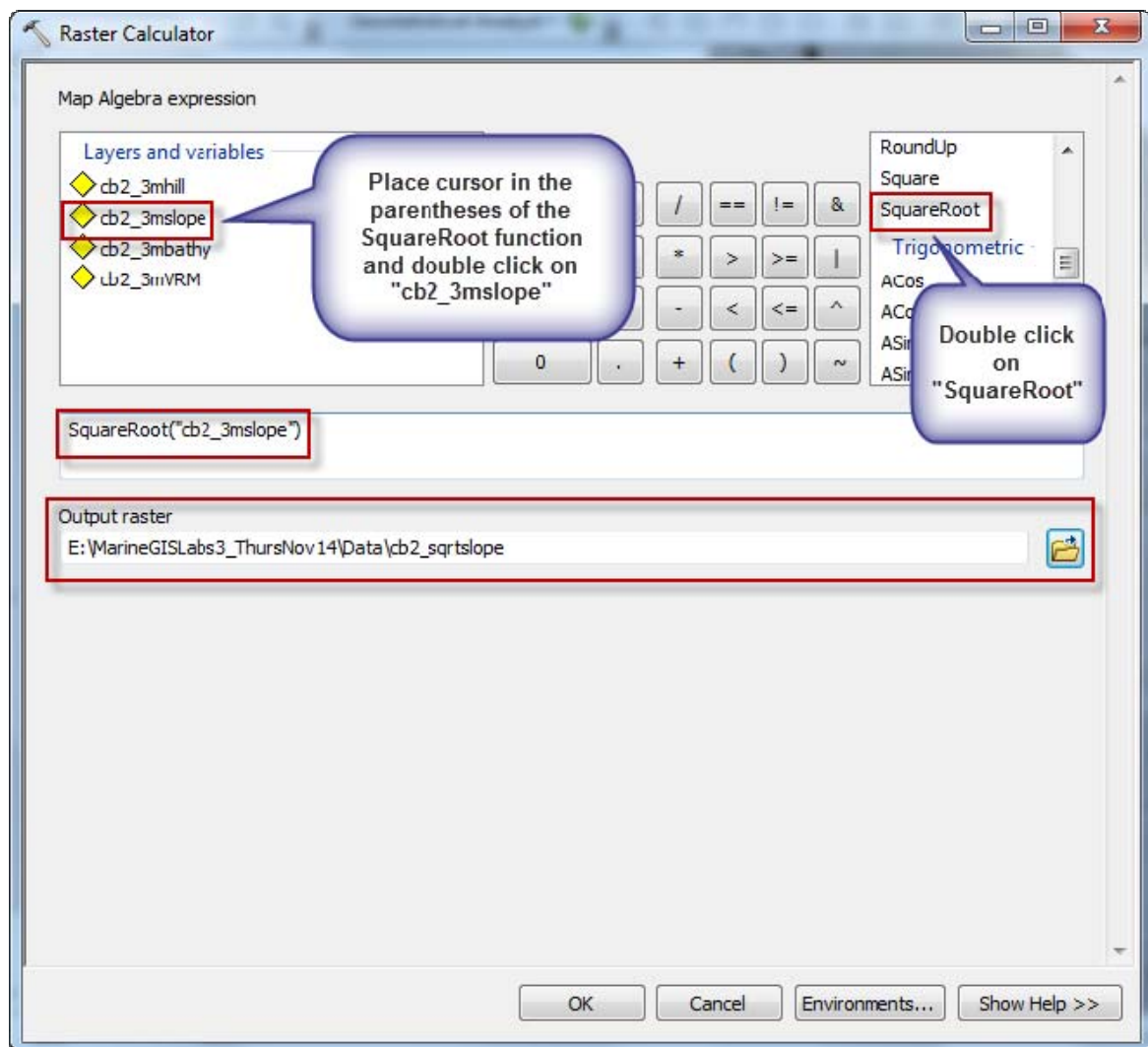
12. The next step is to fit a GLM to the data set using MGET/ Statistics / Model Data /Generalized Linear Models / Fit GLM. Select your input table for YT_PresAbs_all.shp and you will have to navigate to it using the folder icon and not using the drop down menu or an error will occur. Select your lab folder to create the output model and name it GLM_YT1. The formula will be presence ~ cb2_3mbath + sqrt_VRM + sqrt_slope using a binomial family. Also, query the test data under model options using 'TestData' = 0. This will allow you to run the model on the 2/3 training data set. Click OK and the tool will run.



- a. You can now view the summary file of the GLM that you just ran. To do this, navigate to the GLM folder you are saving your model outputs into and open the "YT_GLM1_summary.txt" file. Some highlights from the outputs include the coefficients of the model so you can see which variables are significant in the GLM and the coefficients associated with

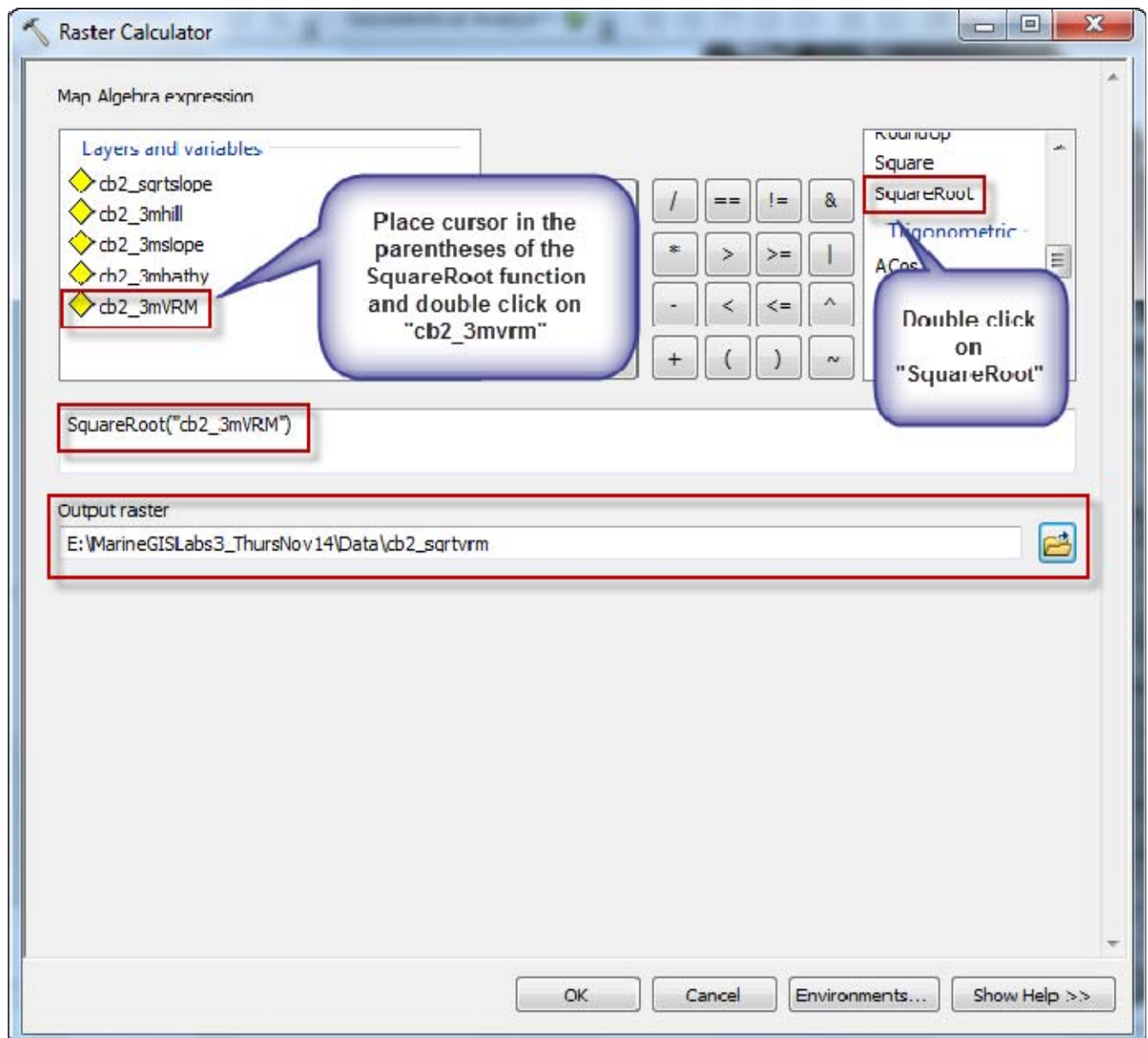
them. In this model, all variables are significant with bathy being the most significant followed by VRM and slope. You can also view the variance inflation factors (VIFs), which give information on the collinearity of the variables. If you are happy with the summary output from the GLM (FYI: You are extremely happy with the results for this lab and will be continuing forward).

13. Now that we have a model that we are happy with, we can create a predictive habitat suitability map for yellowtail rockfish so we can view where our GLM predicts the fish are likely to occur throughout Cordell Bank. Prior to doing this, however, we need to first create rasters to represent our transformed variables (sqrt_VRM and sqrt_slope). Because we transformed these variables to fit the model we will need to transform the rasters we will use to create the predictive output.
14. Within ArcToolbox select Spatial Analyst Tools → Map Algebra and select “Raster Calculator”
 - a. For the map algebra expression, scroll through the list on the right and select “SquareRoot” and double click on it. You will now see “SquareRoot()” in the Map Algebra expression area. Click in the parentheses of the expression and then double click on the “cb2_3mslope” layer. That layer will now appear in the SquareRoot function.
 - b. Output raster: navigate to the Data folder where all the other grids are located and name the new file “cb2_sqrtslope”. Click OK and wait for the map algebra to run. The new layer will appear in your ArcMap document.



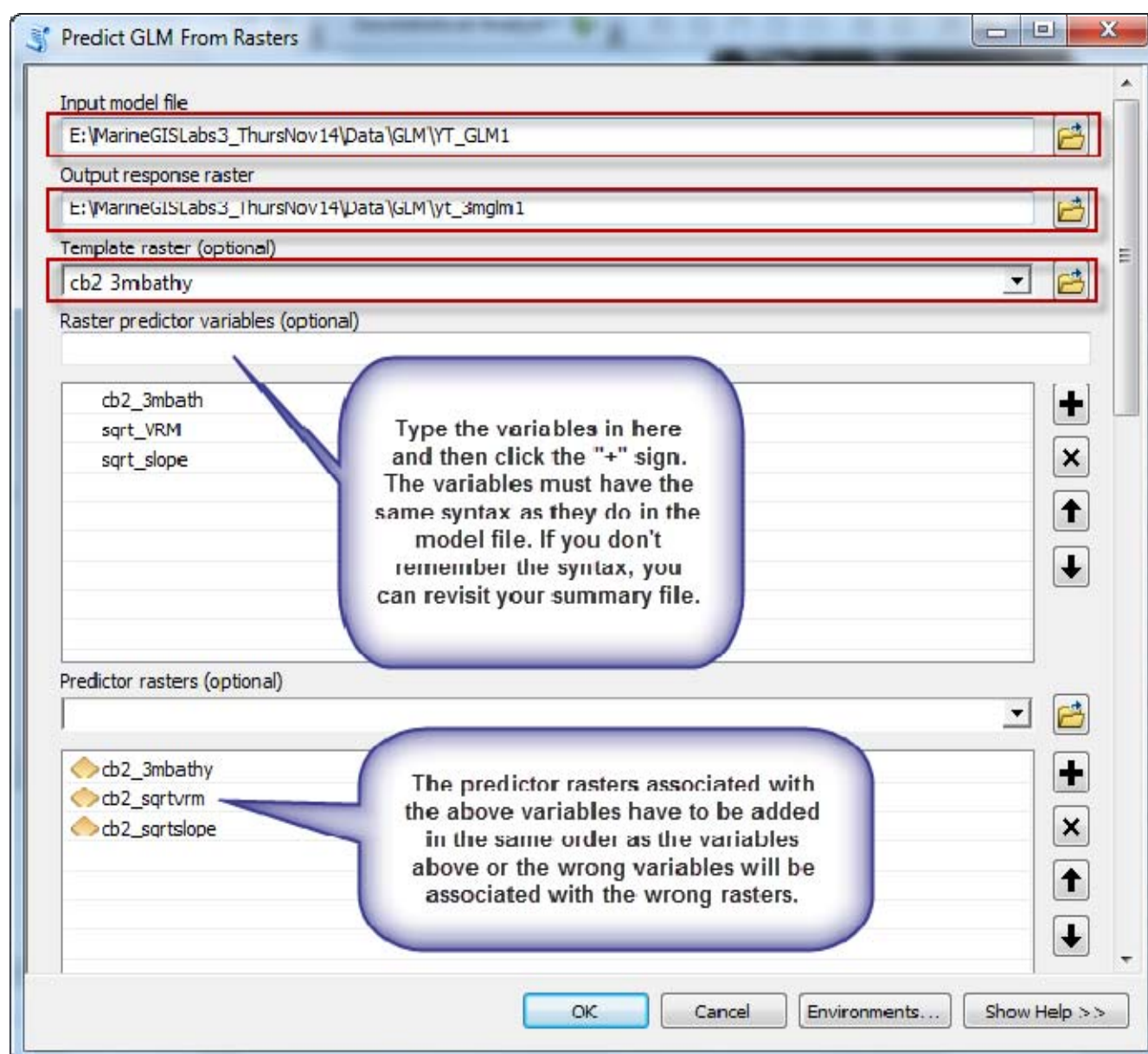
15. Select Raster Calculator again.

- a. For the map algebra expression, scroll through the list on the right and select “SquareRoot” and double click on it. You will now see “SquareRoot()” in the Map Algebra expression area. Click in the parentheses of the expression and then double click on the “cb2_3mVRM” layer. That layer will now appear in the SquareRoot function.
- b. Output raster: navigate to the Data folder where all the other grids are located and name the new file “cb2_sqrtvrn”. Click OK and wait for the map algebra to run. The new layer will appear in your ArcMap document.



- c. Within ArcToolbox go to Marine Geospatial Ecology Tools 0.8a51 → Statistics → Model Data → Generalized Linear Models and choose “Predict GLM from Rasters”
 - i. Input model file: navigate to the GLM folder within your data folder for this lab and select the “YT_GLM1” model file you created as an output when you were fitting your model.
 - ii. Output response raster: This raster will be your resulting habitat suitability map. Navigate to the GLM folder and name the resulting raster “yt_3mg1m1”
 - iii. Template raster: from the drop down menu, select the “cb2_3mbathy” raster. Any raster would do because they all have the same coordinate system and extent but this step just uses the information from the bathy raster.

- iv. **Raster predictor variables:** type in the variables one at a time in the order of the equation you did in the fit model steps: type in “cb_3mbath” and click the plus symbol, type in “sqrt_VRM” and click the plus symbol, and type in “sqrt_slope” and click the plus symbol. Remember that the syntax for the variables need to be the same as what was included in the original model file. Review the summary file if you don’t remember the proper syntax.
- v. **Predictor rasters:** Select the predictor rasters associated with the predictor variables above in the same order as the predictor variables using the drop down menu: “cb2_3mbathy”, “cb2_sqrtvrm”, and “cb2_sqrtslope”.

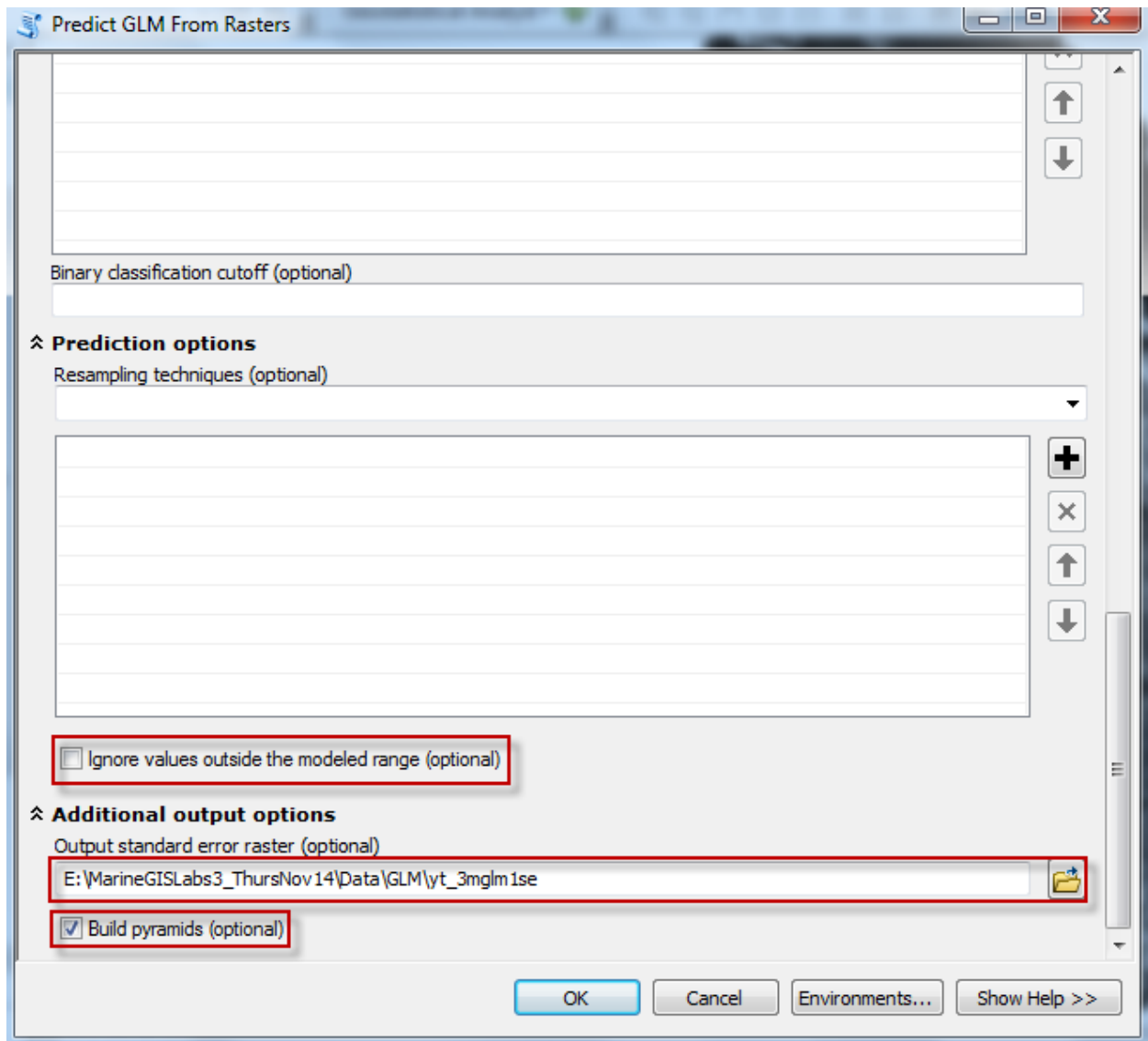


- vi. Expand the “Prediction Options” section and deactivate the “Ignore values outside the modeled range” check box. For this

lab we are going to create a full coverage habitat suitability map. Depending on what you are using the resulting map for you may or not want to create predictions outside the modeled range of your data.

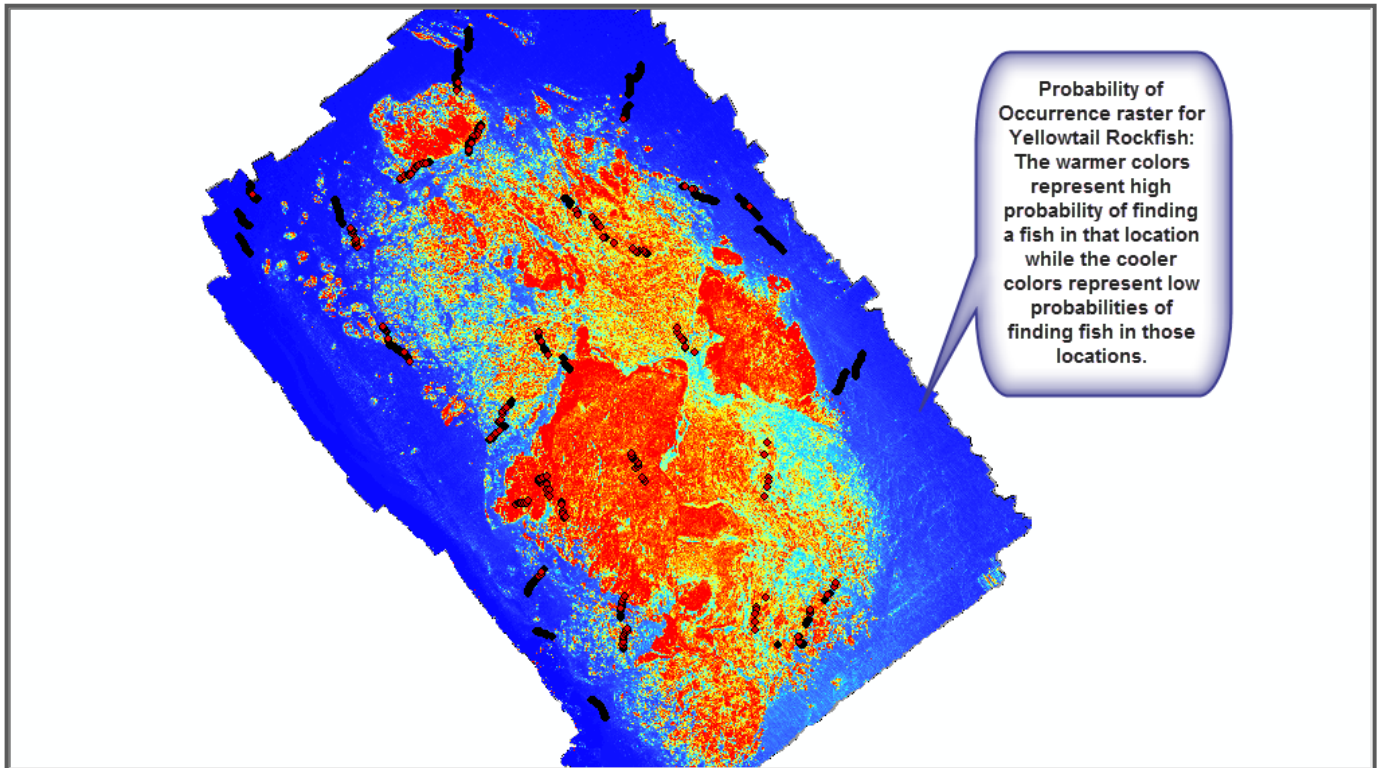
vii. Expand the “Additional output options” section:

1. Output standard error raster: navigate to the GLM folder inside your data folder for this lab and name the output raster “yt_3mgglm1se”
2. Activate build pyramids.

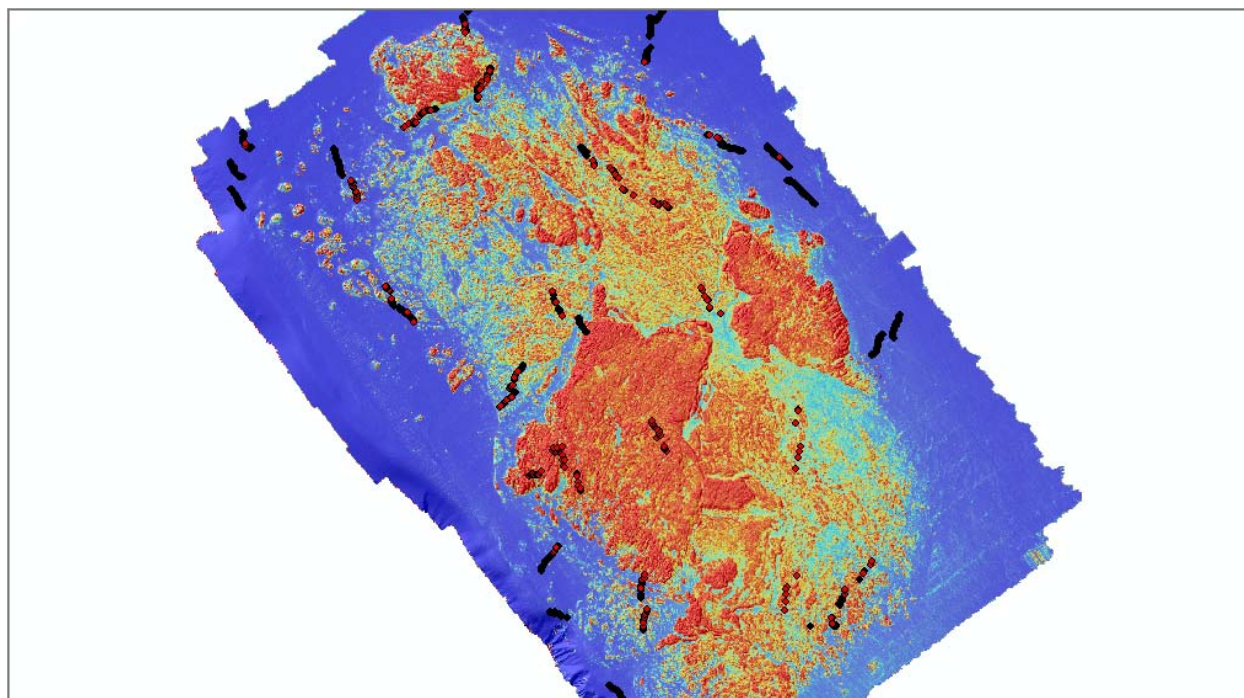
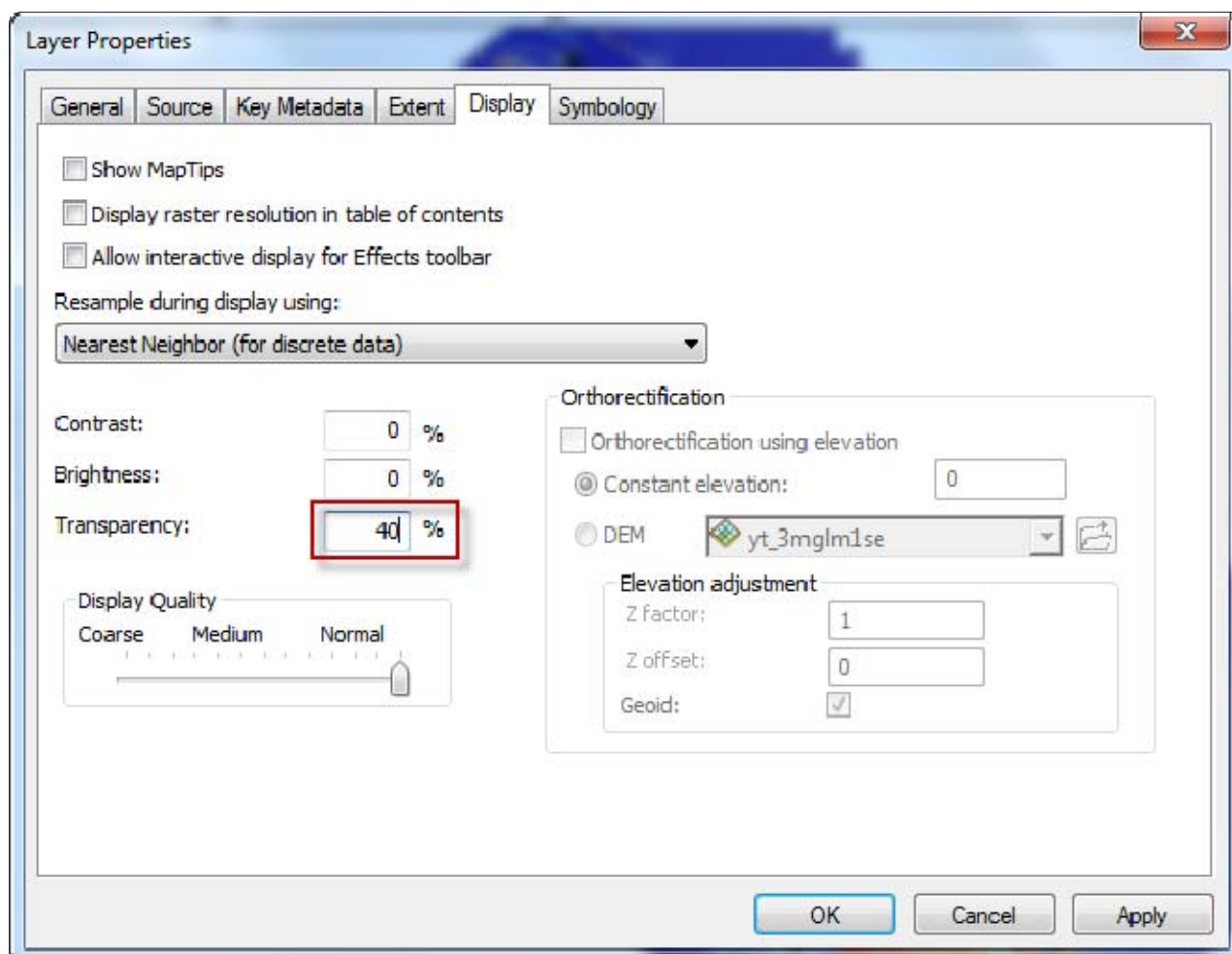


viii. Click OK and wait for the model to run. Once the process completes, a new raster will appear in your project with values ranging from 0 to 1. These values are associated with the

probability of finding a fish in each of the 3 m resolution cells. Below is an image of the probability raster with the symbology set with the warmer colors representing high probability of occurrence while the cooler colors represent low probability of occurrence.



- i. If you want to view the structure of the rocky habitat under the predictions, you can change the transparency of the glm grid and place it on top of the shaded relief imagery. Right click on the grid, select "Properties..." Click on the "Display" tab and type in 40 for the transparency.
- ii. You will want to make sure that there are no layers turned on between the predictive grid the shaded relief grid (cb2_3mhill) and you will be able to see the structure of the rock. This is just a more interesting way to view the prediction and you can get a better understanding for the three dimensional structure of the reef where the fish are most likely to occur.



16. Now we want to test the accuracy of the predictions from this model by using the “Predict GLM from Table” tool.

a. Before we do the model accuracy assessment, we need to create a field in the table to send the predicted values to.

i. Right click on the “YT_PresAbs_all” layer in the table of contents and select “Open Attribute Table”

1. Add a new field to the attribute table and call it “Prediction” with a type of “Float”

FID	Shape	presence	x	y	cb2_3mbath	cb2_3mVRM	cb2_3mslop	TestData	sqrt_VRM	sqrt_slope	Prediction
0	Point	0	457969.364875	4212756.38355	-92.96	0.000681	8.23046	0	0.0261	2.86888	0
1	Point	0	460480.67482	4212096.16939	-85.76	0.000479	3.37592	0	0.021878	1.83737	0
2	Point	0	460514.103356	4212078.47631	-85.75	0.000007	0.136021	0	0.002708	0.36881	0
3	Point	0	460568.590404	4211988.67072	-85.27	0.00001	0.346232	1	0.003202	0.588415	0
4	Point	0	460579.154329	4211959.77095	-84.77	0.000093	2.32666	0	0.009655	1.52534	0
5	Point	0	460702.815227	4211859.31324	-81.49	0.00151	5.07173	1	0.038853	2.25205	0
6	Point	0	460722.872207	4211833.91853	-82.15	0.000132	2.15412	0	0.011477	1.46769	0
7	Point	0	460725.024418	4211825.36451	-81.94	0.000113	0.74184	0	0.010622	0.861301	0
8	Point	0	460725.212811	4211809.8297	-82.13	0.000079	0.702711	0	0.00891	0.838279	0
9	Point	0	461101.406644	4211603.20111	-78.28	0.000072	1.3273	0	0.008471	1.15209	0
10	Point	0	463247.796963	4212040.95389	-110.54	0.000044	1.05939	0	0.006605	1.02927	0
11	Point	0	463217.549996	4212045.63947	-110.83	0.000047	1.80342	1	0.006858	1.34291	0
12	Point	0	463178.333328	4212065.12282	-110.93	0.000021	0.944814	0	0.004593	0.972015	0
13	Point	0	463149.017455	4212081.67676	-110.39	0.000101	2.40128	0	0.01006	1.54961	0
14	Point	0	463138.957433	4212088.15768	-110.47	0.000027	3.04821	0	0.005156	1.74591	0
15	Point	0	463109.792476	4212079.85691	-109.76	0.000131	1.6276	0	0.011425	1.27577	0
16	Point	0	463118.588203	4212084.81015	-110.37	0.000409	4.15759	1	0.020234	2.03902	0
17	Point	0	463113.322312	4212084.39014	-109.93	0.00016	3.8597	0	0.012662	1.96461	0

a. Open the “Predict GLM from Table” tool, which is located immediately below the “Predict GLM from Rasters” tool we just used.

a. Input Model File: Navigate to and select the YT_GLM1 model file we created when we fit the model (this is the same model file we used for the creation of the predictive probability raster).

b. Input Table: Select the “YT_PresAbs_all” shapefile from the drop down list.

c. Field containing actual values of the response variable: select the “presence” field from the drop down menu

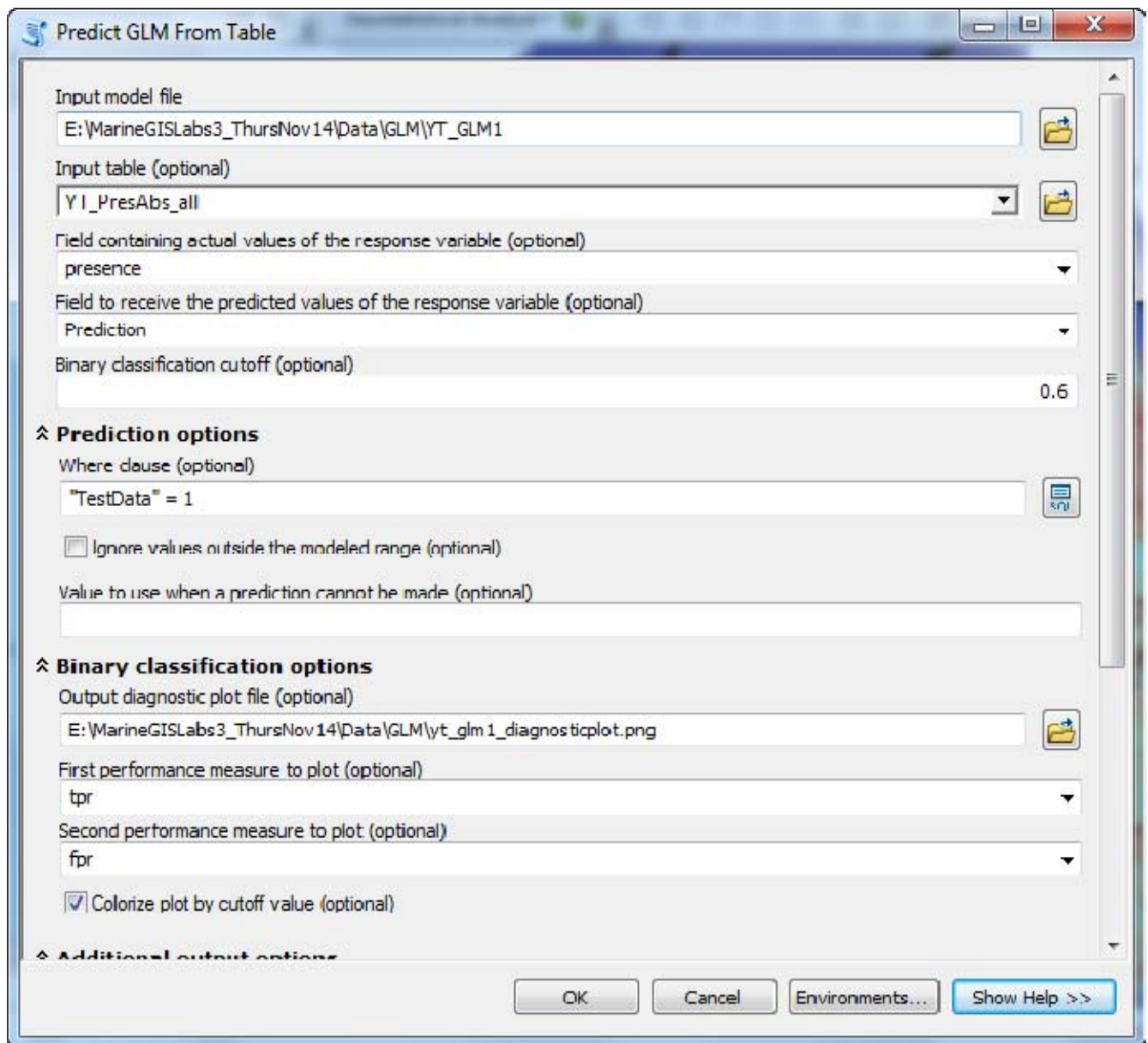
d. Field to receive the predicted values of the response variable: Select the “Prediction” field you just created.

e. Binary classification cutoff: type in “0.6”. This value will give us a cutoff of 0.6 for the probabilities. This value means that everything below 0.6 will be treated as “absent” and everything above will be treated as “present”

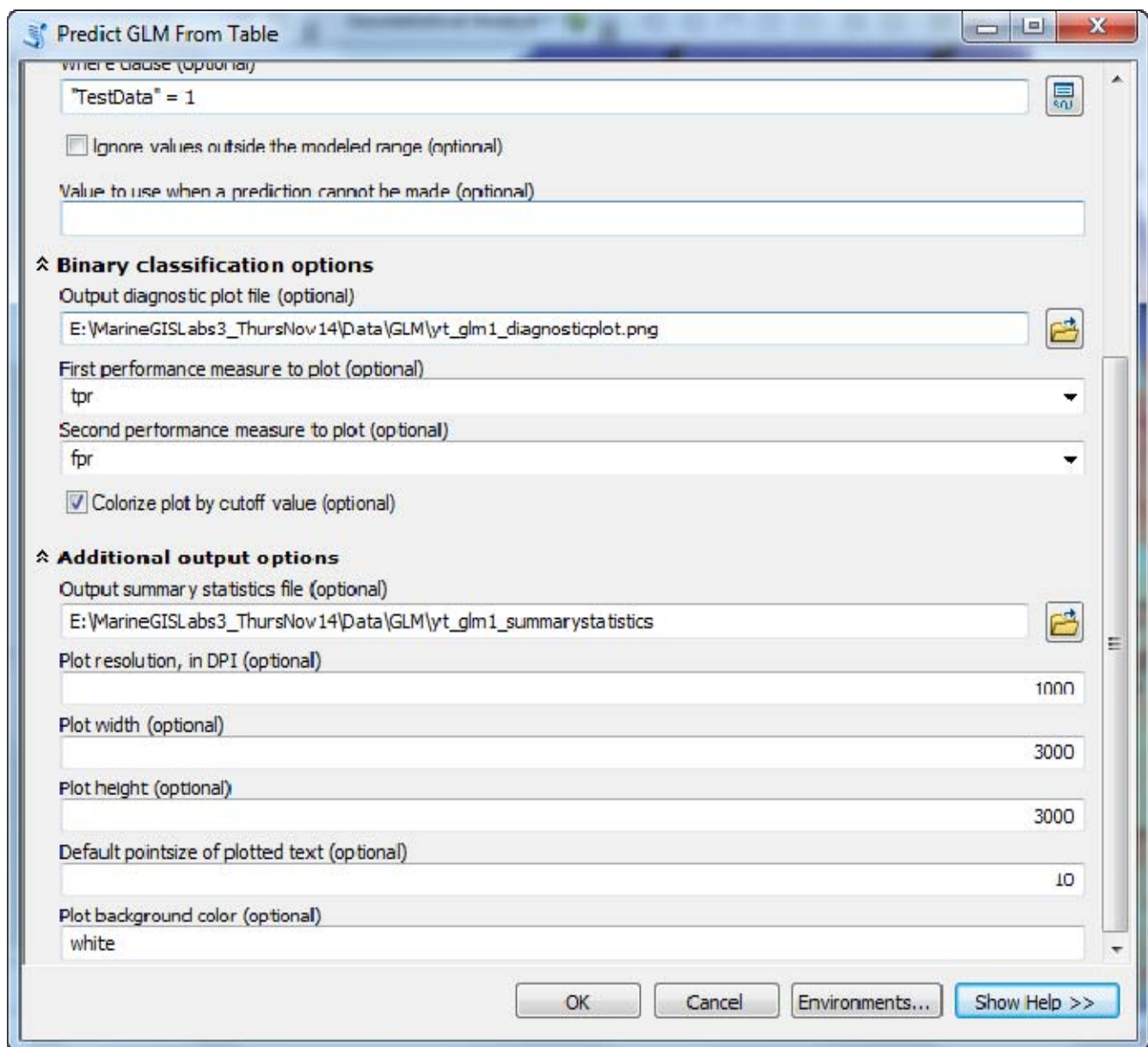
f. Expand the “Prediction Options”

i. Where clause “TestData” = 1 (We only want to compare the predicted values to the response values that were not used to create the model.

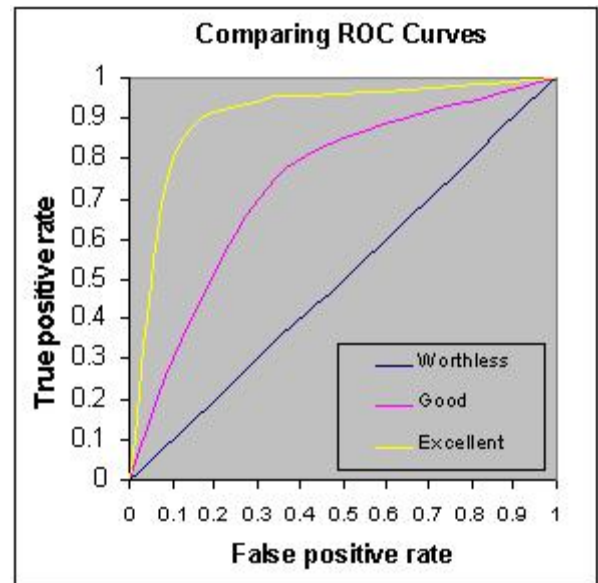
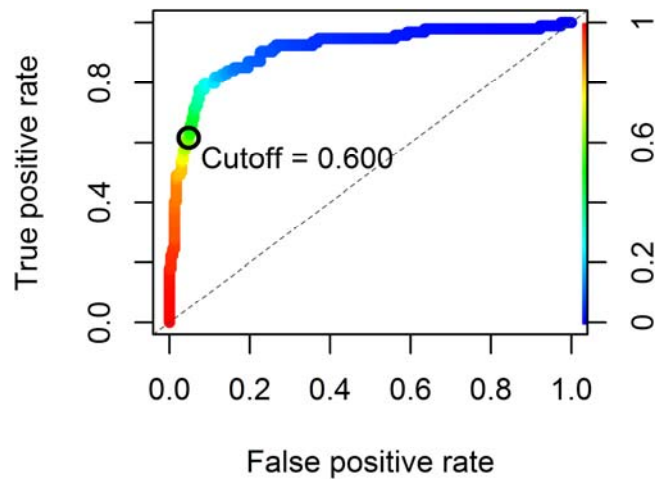
ii. Deactivate “Ignore values outside the modeled range”



- g. Expand the “Binary Classification Options”
 - i. Output diagnostic plot file: Navigate to your GLM folder and name the file “yt_glm1_diagnosticplot.png”
- h. Expand the “Additional Output Options”
 - i. Output summary statistics file: Navigate to your GLM folder and name the file “yt_glm1_summarystatistics”



- i. Click OK and allow the model to run. If you uncheck the “Close this dialog when completed successfully” option you can see the output in the geoprocessing window when it is completed.
- b. Navigate to the folder where you saved all your outputs from the model accuracy assessment.
 - a. Diagnostic Plot: This shows the False Positive and True Positive rates of the prediction and where the cutoff value lies within that.



- b. **Summary Statistics:** If you open this file in notepad or another text editor you can view the summary statistics of the predictions including the overall accuracy, the area under the roc curve (auc), and the Cohen's kappa value. The area under the roc curve is a measure of how well the model predicts presence or absence based on your chosen cutoff value with 1 being the best and 0 being the worst. In this case we have a pretty high auc of 0.909, which equates to "excellent" agreement. Cohen's kappa is a statistical measure of the agreement between the predicted and actual values. For this model we have a cohen's kappa of 0.618, which is okay. You would probably want to improve this value by adding other variables to the model to help to explain the distribution of the yellowtail rockfish and/or adjust the values you are using. Depending on the use of the resulting raster you will want to better understand the accuracy tests done and presented in this file to select the one the most useful to your project.

```
yt_glm1_summarystatistics - Notepad
File Edit Format View Help
MODEL PERFORMANCE SUMMARY:
=====
Area under the ROC curve (auc) = 0.909
Mean cross-entropy (mxe) = 0.347
Precision-recall break-even point (prbe) = 0.787
Root-mean square error (rmse) = 0.319

User-specified cutoff = 0.600

Confusion matrix for that cutoff:

      Actual 1   Actual 0   Total
Predicted 1     58       11     69
Predicted 0     36      219    255
      Total     94      230    324

Model performance statistics for that cutoff:

Accuracy (acc) = 0.855
Error rate (err) = 0.145
Rate of positive predictions (rpp) = 0.213
Rate of negative predictions (rnp) = 0.787

True positive rate (tpr, or sensitivity) = 0.617
False positive rate (fpr, or fallout) = 0.048
True negative rate (tnr, or specificity) = 0.952
False negative rate (fnr, or miss) = 0.383

Positive prediction value (ppv, or precision) = 0.841
Negative prediction value (npv) = 0.859
Prediction-conditioned fallout (pcfall) = 0.159
Prediction-conditioned miss (pcmiss) = 0.141

Matthews correlation coefficient (mcc) = 0.631
Odds ratio (odds) = 32.076
SAR = 0.694

Cohen's kappa (k) = 0.618
```

***Key Concept**

Although the model we have created in this case looks pretty good, there are a lot of statistical rules that we neglected. We didn't test for spatial autocorrelation in the residuals of the model output to determine if there is significant correlation. We also used variables that were collinear and we didn't assess the relationship between the response variable and the covariables. These are all problems that you can run into when doing these types of spatial models and they must be taken into consideration. This lab was meant to introduce you to the tools. If you would like to use this type of modeling in your own work, you need to become familiar with the statistics.