Obtaining Submerged Aquatic Vegetation Coverage from Satellite Imagery and Confusion Matrix Analysis

Brian Madore

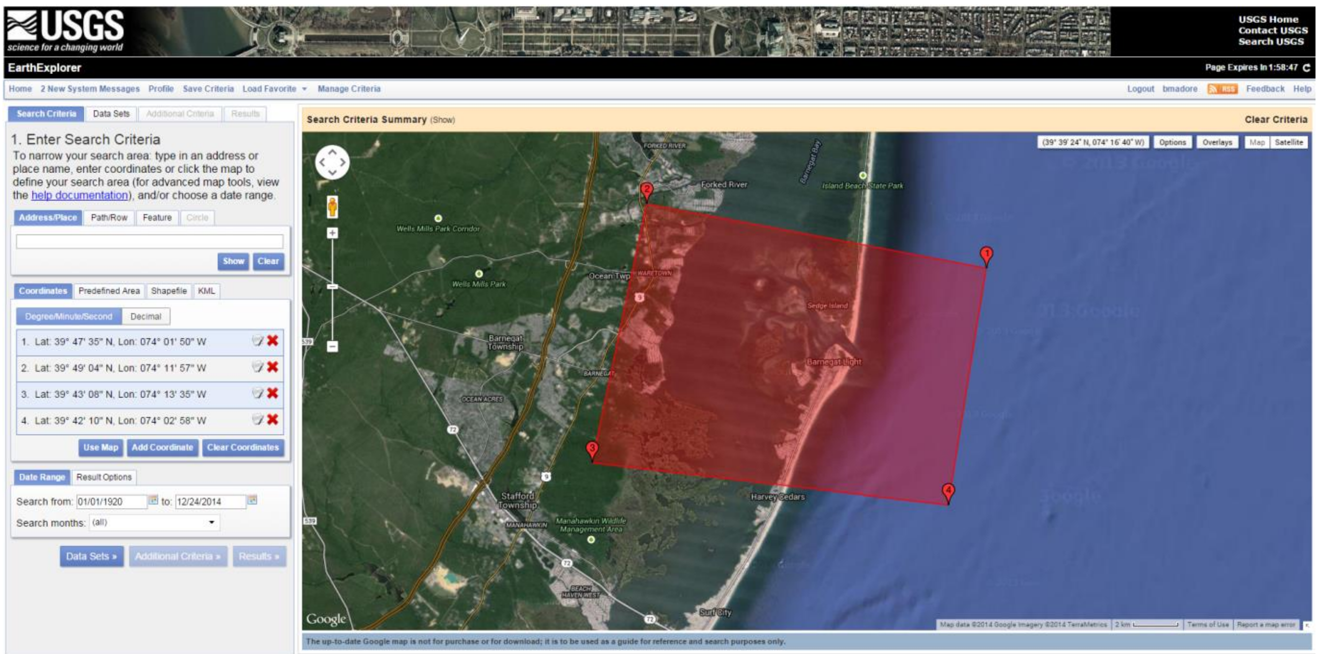
April 7, 2015

This document shows the procedure for obtaining a submerged aquatic vegetation (SAV) map using satellite imagery for a shallow coastal environment. The SAV detected in this process is not distinguished between different types or classifications of vegetation. The last part of the procedure documents how to perform a confusion matrix analysis of the SAV detection results. In order to conduct the analysis a reference dataset (or ground truth data) is required.

A. Obtaining Landsat 8 Imagery

A.1. Landsat 8 imagery can be downloaded from http://earthexplorer.usgs.gov/ . A login is required, however registration is free and easy to obtain.

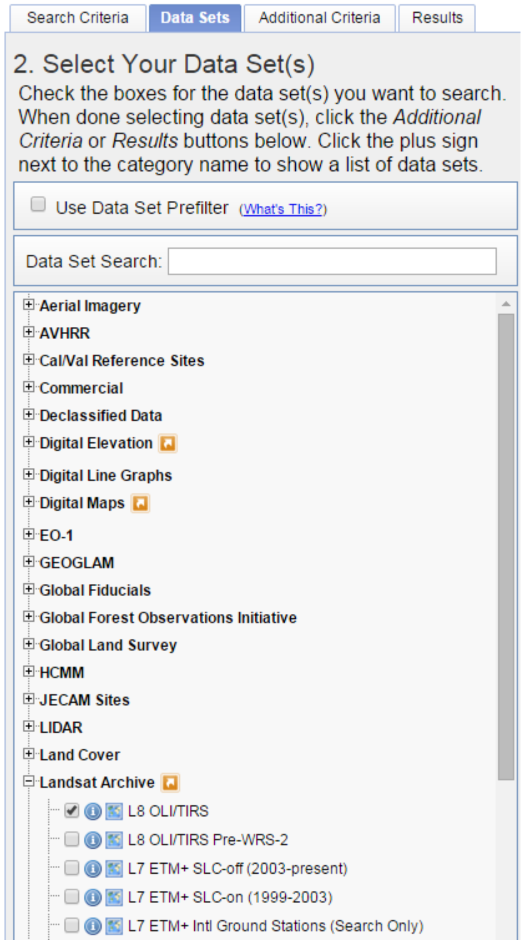
A.2. Once logged in zoom to the place of interest on the map. Create a polygon around the area of interest by clicking on the map to create the vertices of the polygon.



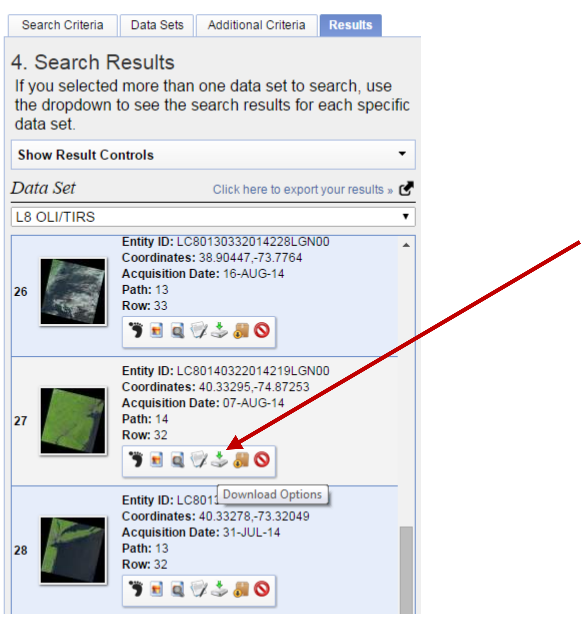
A.3. If there are any date constraints that need to be used input those under the Search Criteria tab.

A.4. Select **Data Sets.**

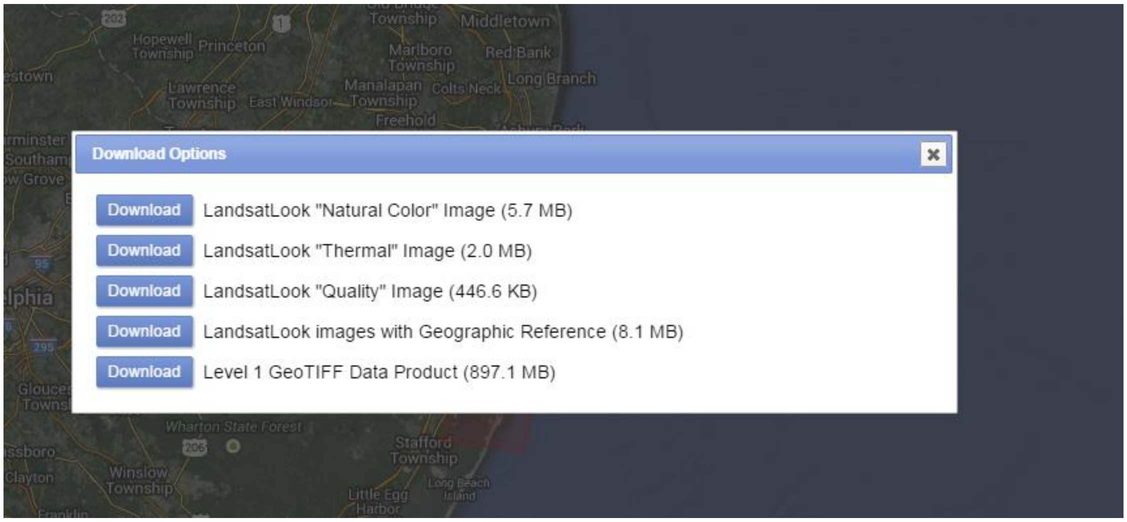
A.5. Open the **Landsat Archive** section and select **L8 OLI/TIRS**.



A.6. Open the **Results** tab and search for the image desired for download. Select **Download Options** on the desired image.



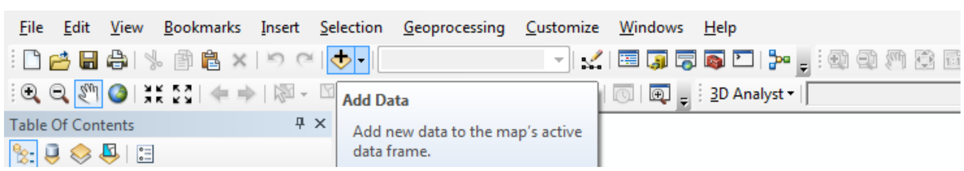
A.7. Select **Level 1 GeoTIFF Data Product** for download.



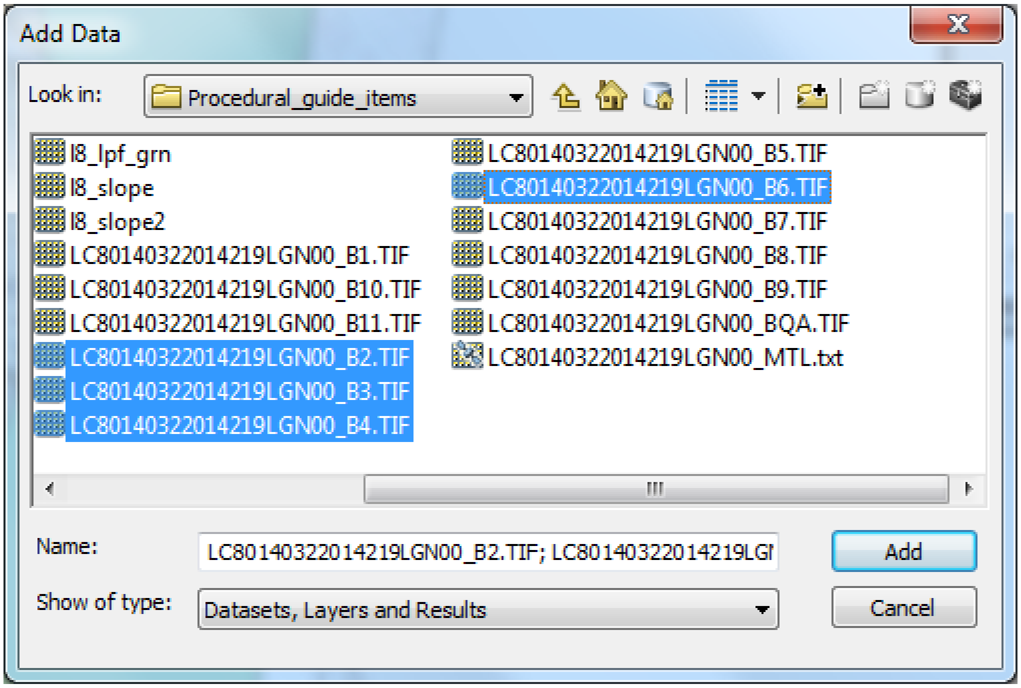
A.8. Once downloaded unzip the file and save the data to a desired folder. It is recommended that the folder the unzip files are saved in be created with the date of the satellite imagery's acquisition date.

B. Creating algorithms to mask deep channels

B.1. In ArcMap select **Add Data**.

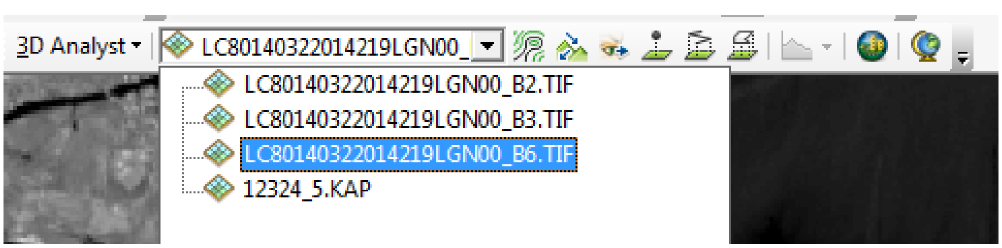


B.2. Add bands 2, 3, 4, and 6 to the data frame. If it asks to create pyramids, select yes.

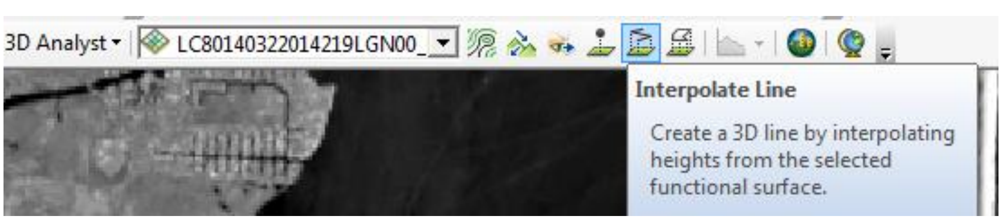


B.3. Using band 6, which is the IR band, the threshold between the land and water surfaces will need to be distinguished. Because water absorbs IR radiation this difference in the IR values for land and water surfaces will be used to mask out all of the land features. There are a couple of ways of determining the threshold.

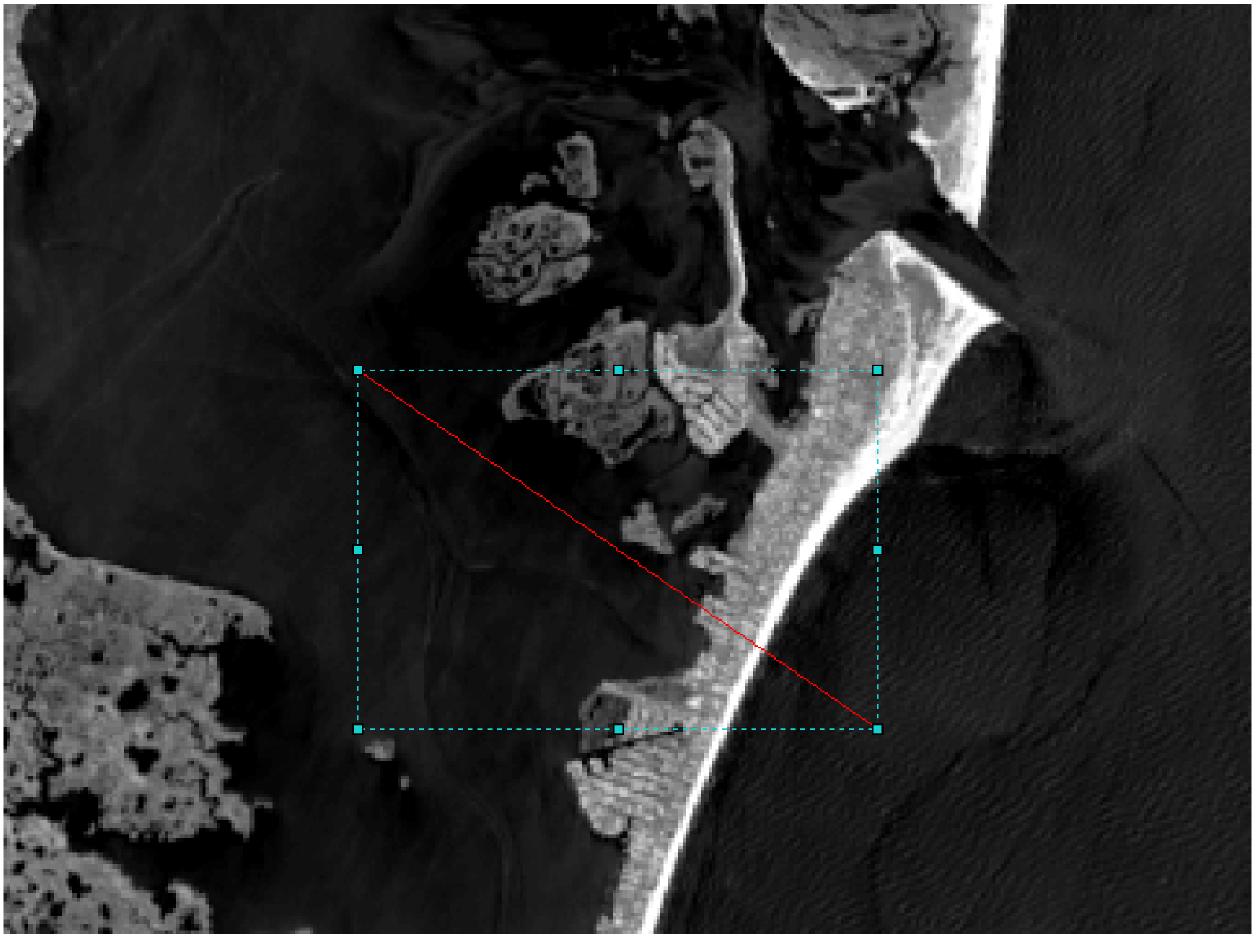
B.4. If possible, using the 3D analyst tool is the easiest way to determine the value between the land/water threshold. Go to **Customize>Toolbars** and select the 3D Analyst toolbar. In the toolbar there is a dropdown box which can be used to select different layers. Select the band 6 layer.



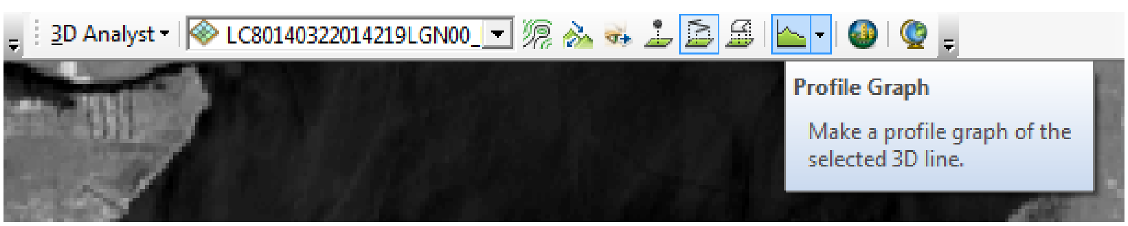
B.5. Select **Interpolate Line** on the toolbar.



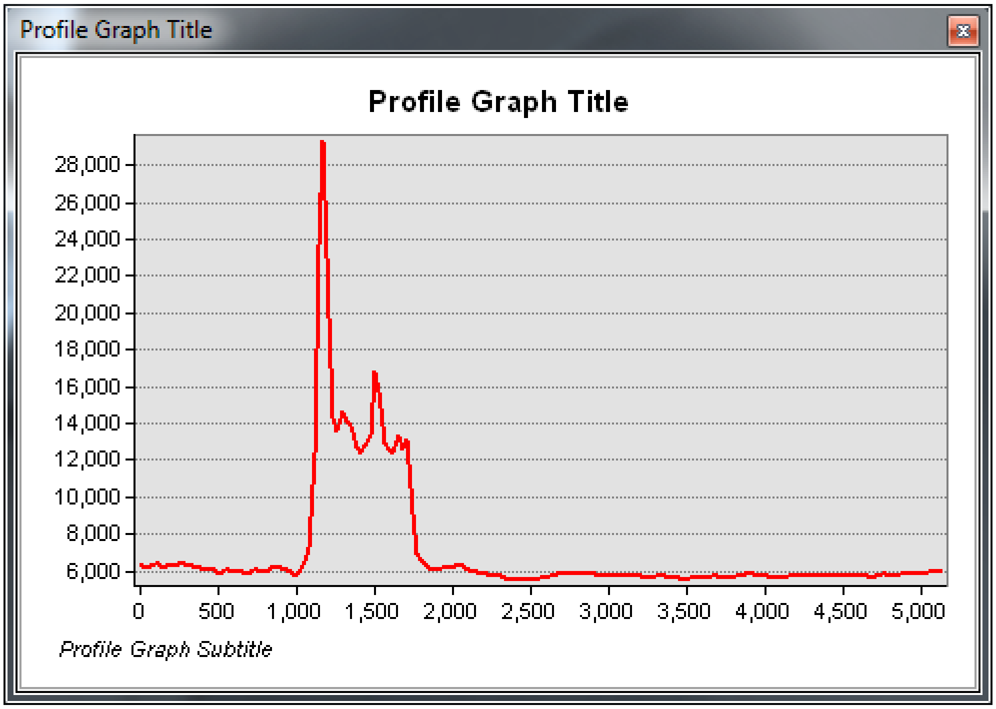
B.6. Draw a line that goes over both land and water.



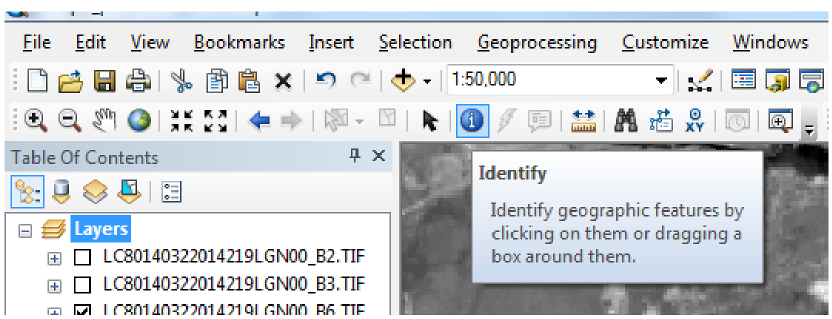
B.7. Select **Profile Graph** on the 3D Analyst toolbar.



B.8. The profile graph will show the values of the IR band over the red line extent. The land IR values will be higher values and the water values will be a consistent lower value. The graph below shows the IR values for this example. All of the water values are below the value of 7000, and the land values are all much higher. Therefore the value of 7000 will be the threshold for this specific image. **Important:** each satellite image, even if collected from the same satellite, will have a different threshold value.



B.9. If the 3D Analyst toolbar cannot be used, the other way to determine the threshold difference is to use the identify tool on the band 6 layer.

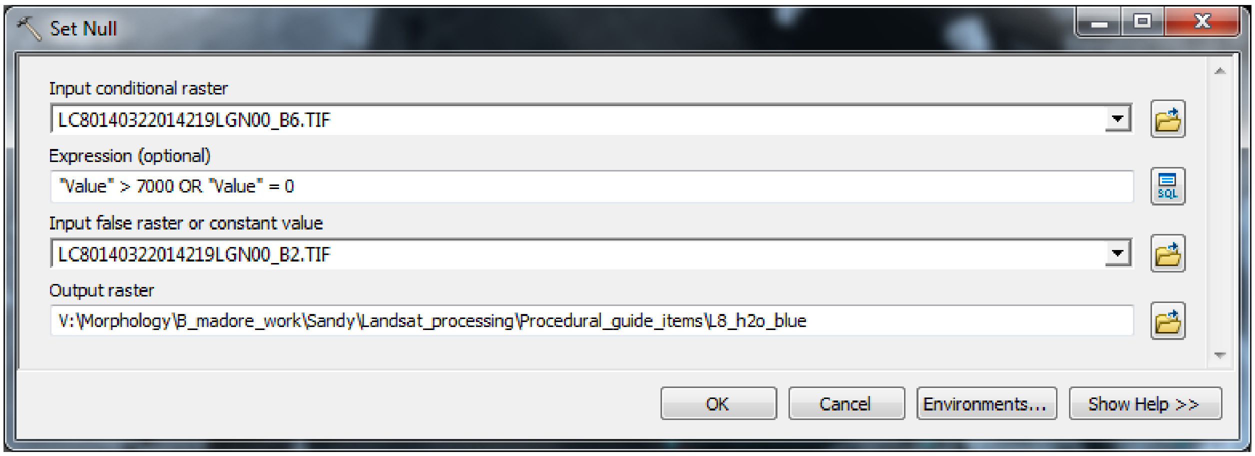


B.10. Select multiple points in the water, on land, and near the water/land boundary to determine what the typical values for the IR band are. A threshold value then should be selected where all values below the threshold value are water and all values above the threshold are land values.

B.11. In the **ArcToolbox** select **Spatial Analyst Tools>Conditional>Set Null**.

B.12. For the **input conditional raster** select the Landsat 8 band 6 (IR band). For the **Expression** use **"Value" >** (threshold value) **OR "Value" = 0**

IMPORTANT: Make sure the expression is used correctly, see the below expression as an example, but make sure to input the correct threshold value for the image being examined, instead of the 7000 which is the value for the image in this procedure. Select the band 2 (blue band) for the **false raster**. Select the folder icon next to **Output Raster** to designate a location to save the layer being created. Name the layer L8_h2o_blue (or similar). This layer being created is using the IR band to mask out all values from the IR band that are considered land by the threshold value on the blue band.

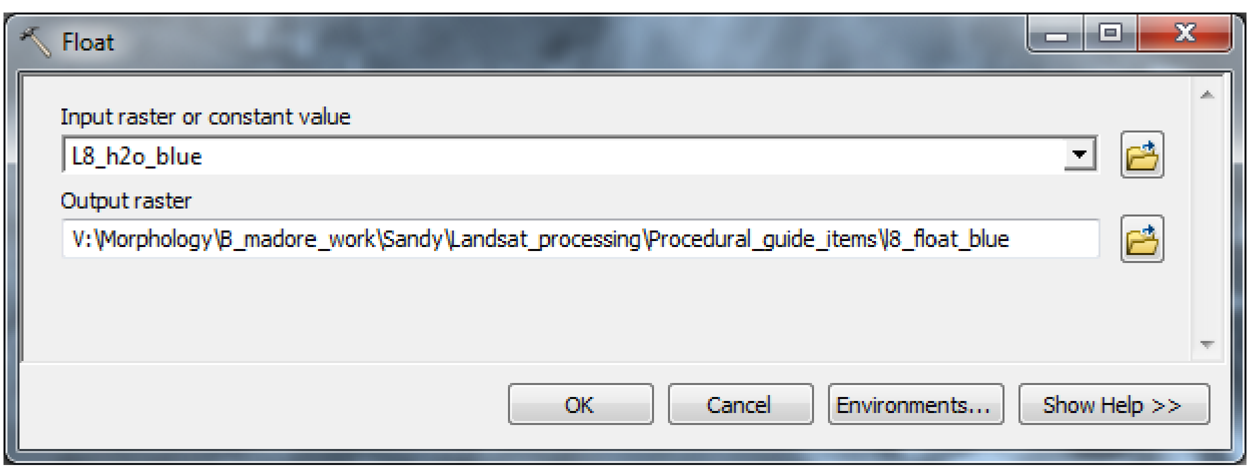


B.13. Repeat step B.11. and B.12. using band 3 as the **false raster** and saving the layer as L8_h2o_green (or similar).

B.14. Repeat step B.11. and B.12. using band 4 as the false raster and saving the layer as L8_h2o_red (or similar).

B.15. In the **ArcToolbox** select **Spatial Analyst Tools>Math>Float**.

B.16. For the **input raster** select L8_h2o_blue. Select a destination to save the **output raster** and name the layer L8_float_blue (or something similar).

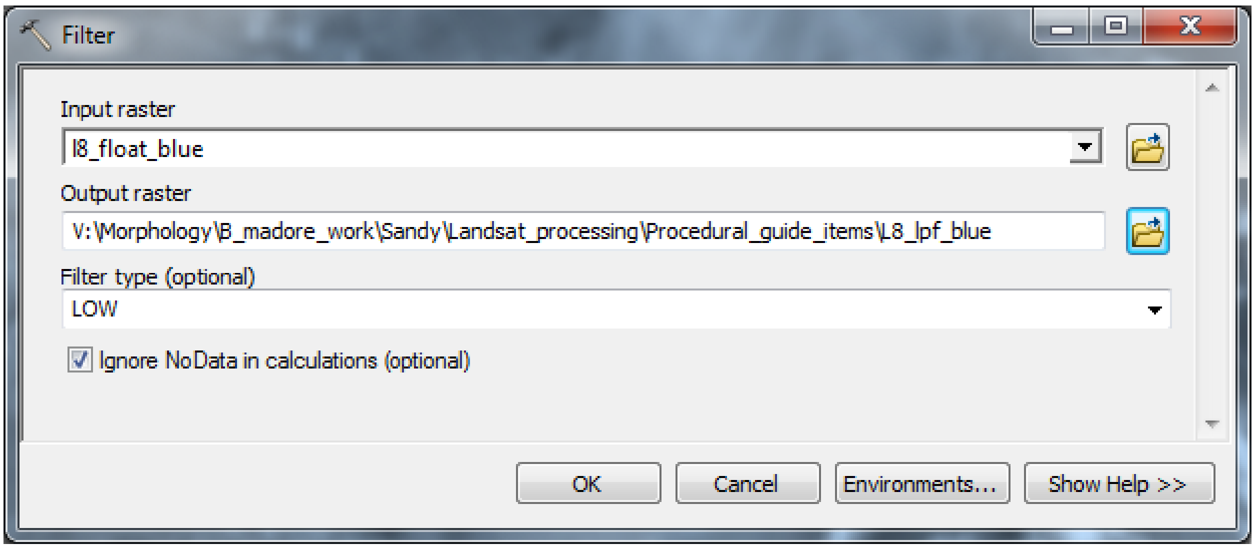


B.17. Repeat steps B.15. and B.16. using the L8_h2o_green for the **input raster** and saving the **output raster** as L8_float_green (or similar).

B.18. Repeat steps B.15. and B.16. using the L8_h2o_red for the **input raster** and saving the **output raster** as L8_float_red (or similar).

B.19. In the **ArcToolbox** select **Spatial Analyst Tools>Neighborhood>Filter**.

B.20. Select L8_float_blue for the **input raster**. Save the output raster as L8_lpf_blue (or similar). **Filter type** should be set to LOW.

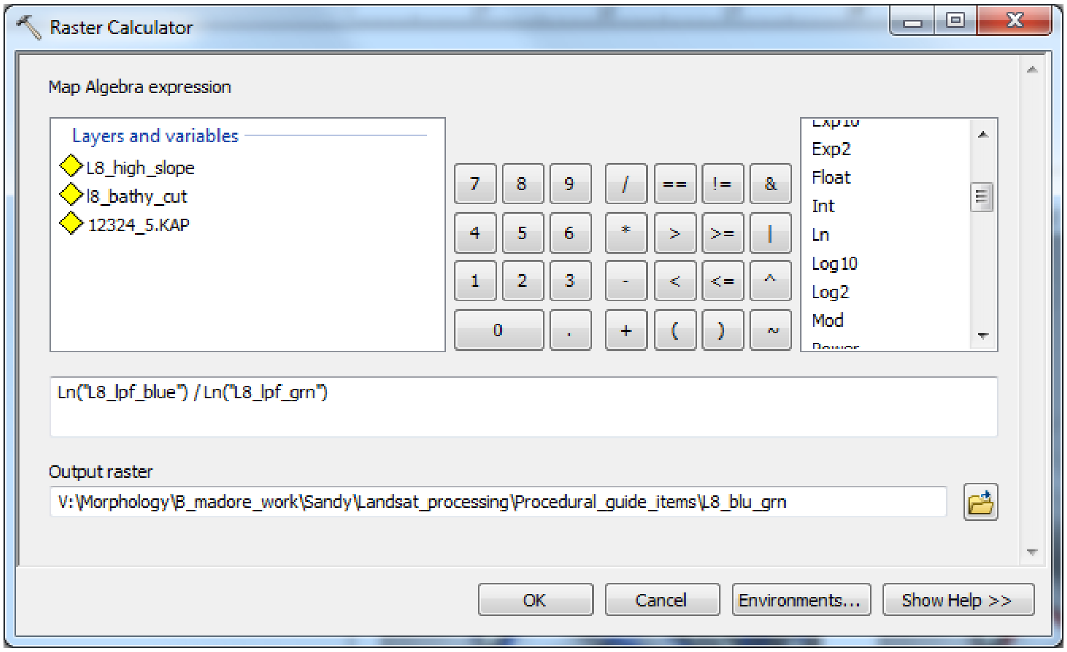


B.21. Repeat steps B.19. and B.20 using L8_float_green for the **input raster** and saving the **output raster** as L8_lpf_green (or similar).

B.22. Repeat steps B.19. and B.20 using L8_float_red for the **input raster** and saving the **output raster** as L8_lpf_red (or similar).

B.23. In the **ArcToolbox** select **Spatial Analyst Tools>Map Algebra>Raster Calculator**.

B.24. Input the equation **Ln("L8_lpf_blue")/Ln("L8_lpf_green")**. See the following figure for reference. Save the **output raster** as L8_blu_grn.



B.25. Repeat steps B.23. and B.24. using the equation

Ln("L8_lpf_green")/Ln("L8_lpf_red") Save the **output raster** as L8_grn_red.

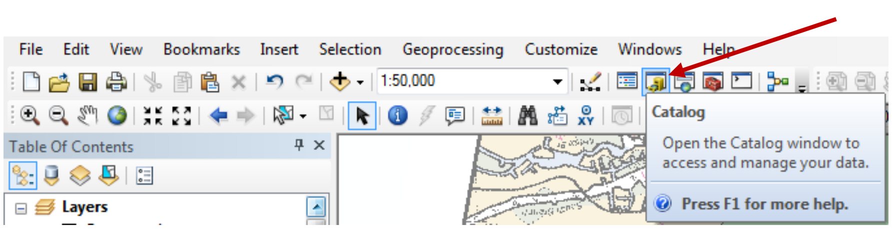
NOTE: The naming of the files should be monitored carefully. Since most likely multiple images from different dates will be analyzed and processed, saving the output rasters with the date in the layer name can be useful.

The naming convention for the layers used in this procedure are basic in order to provide structure to the workflow.

C. Extracting the study area

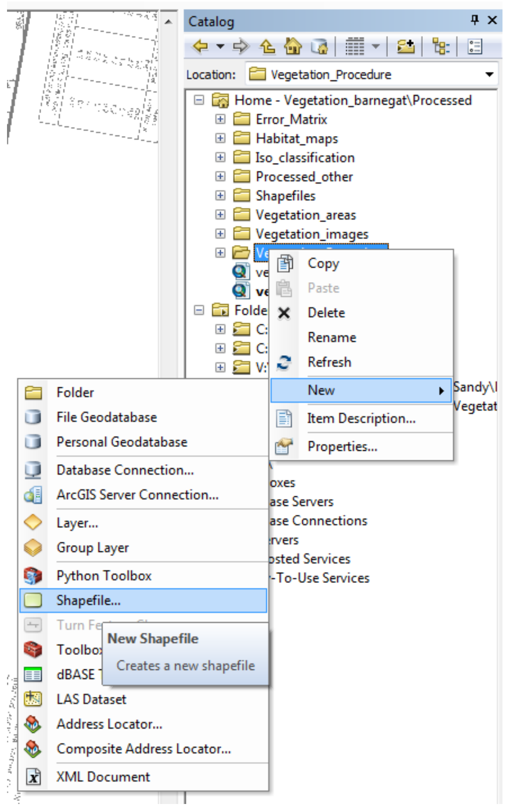
C.1. A shapefile of the study area will need to be created in order to extract the preceding layers. If a study area shapefile already exists, skip to C.10.

C.2. Open the **Catalog** by selecting the following icon.

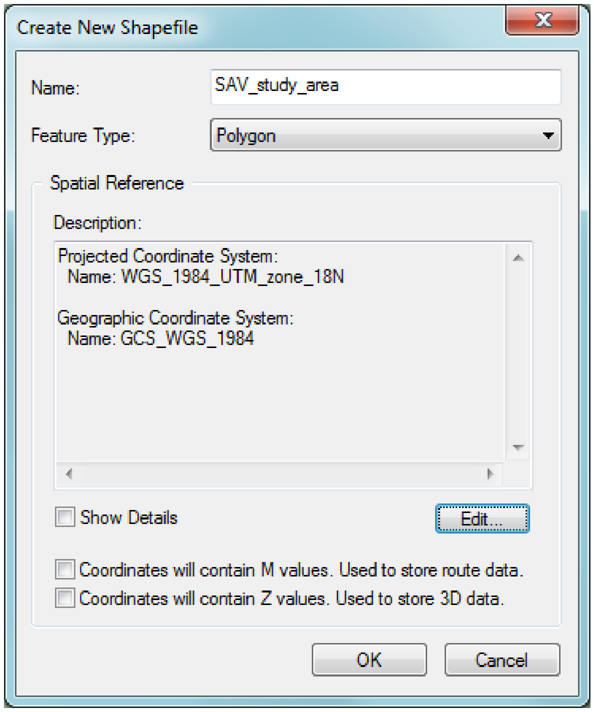


C.3. In the **Catalog** navigate to the folder where the shapefile being created will be saved.

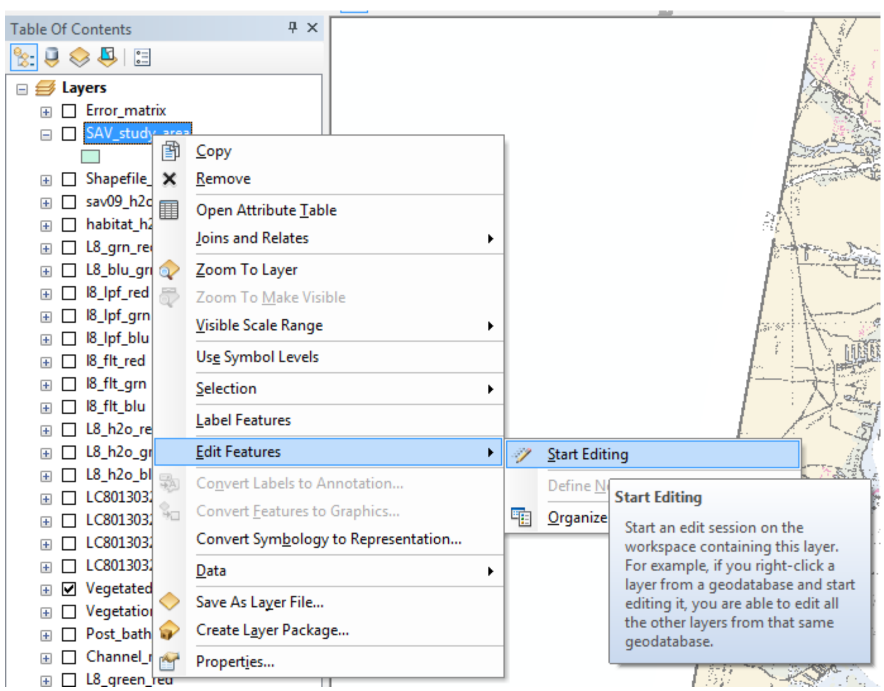
C.4. Right-click the folder and select **New > Shapefile.**



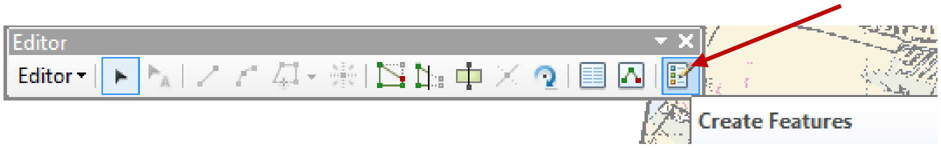
C.5. Name the shapefile SAV_study_area. Select **Polygon** for the feature type. Select **Edit** to choose the correct spatial reference for your shapefile (**NOTE:** This spatial reference is unique to your work and should not be copied from the one in this example, unless the work corresponds to that area). Select **OK.**



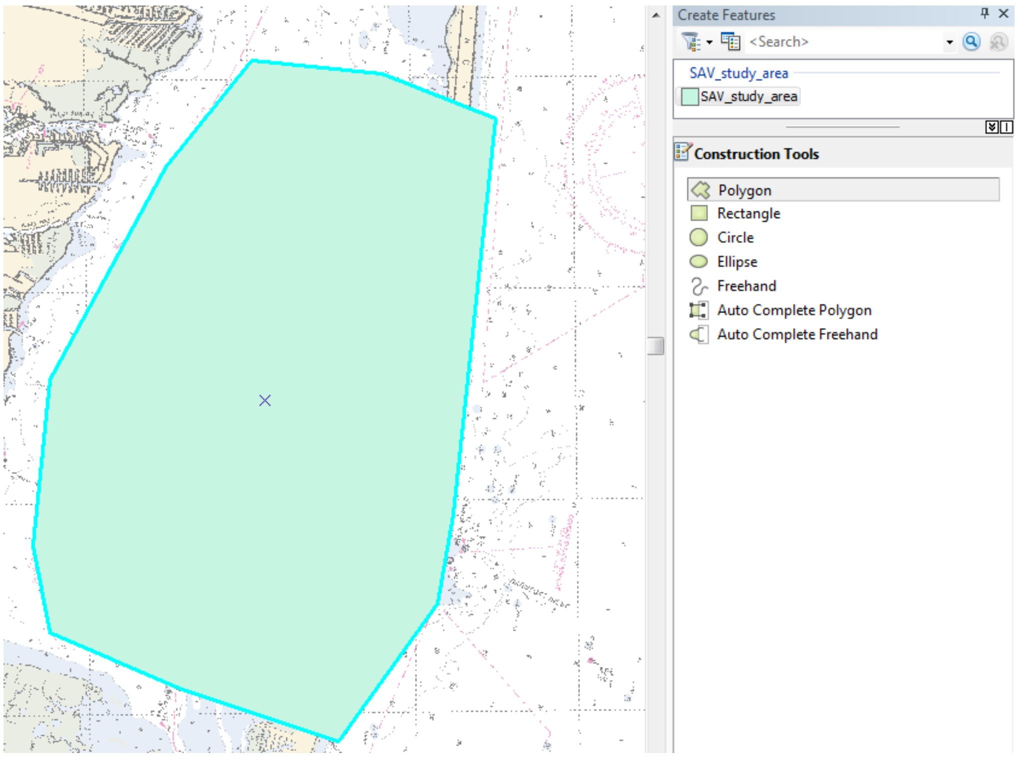
C.6. Right-click on the SAV_study_area shapefile in the table of contents. Select **Edit Features > Start Editing**.



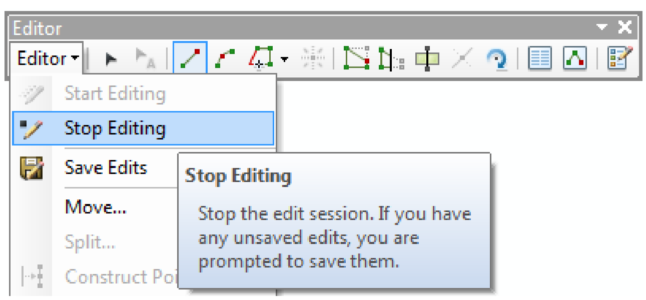
C.7. Open the **Create Features** tool on the **Editor** toolbar.



C.8. Select SAV_study_area in the **Create Features** window. Select **Polygon** to start constructing the shapefile. Create the polygon by clicking to create the vertices of the polygon. Once the polygon is finished double click the last vertex.

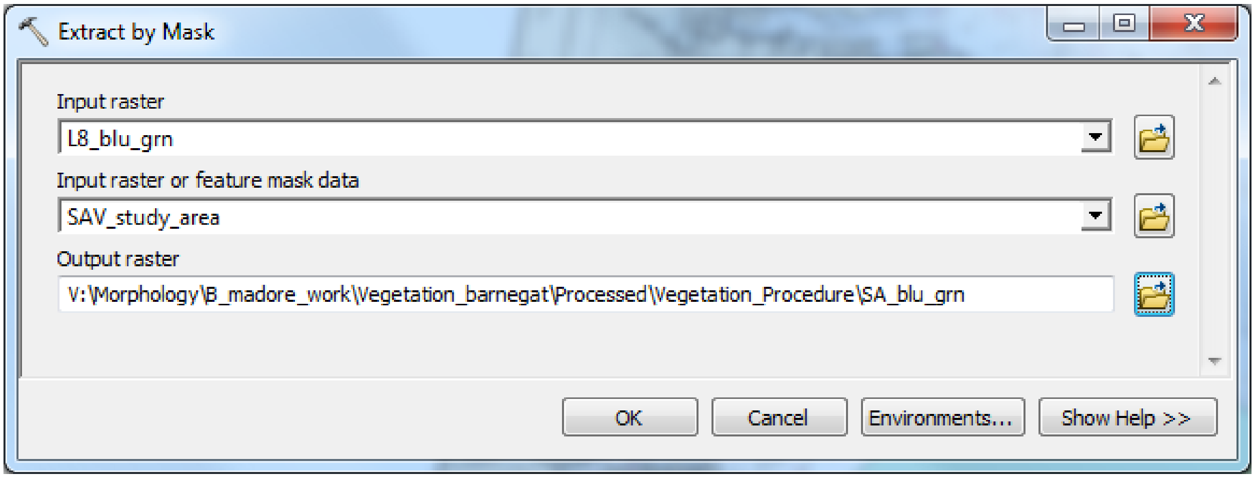


C.9. Select **Editor > Save Edits.** Then select **Editor > Stop Editing**.



C.10. In the **ArcToolbox** select **Spatial Analyst Tools>Extraction>Extract by Mask**.

C.11. Select L8_blu_grn for the **input raster**. Select the SAV_study_area (or other shapefile) for the **feature mask data**. Select a location to save the **output raster** and name the layer SA_blu_grn (or other name).



C.12. Repeat C.10 and C.11 using L8_grn_red for the **input raster.** Save the **output raster** as SA_grn_red (or other).

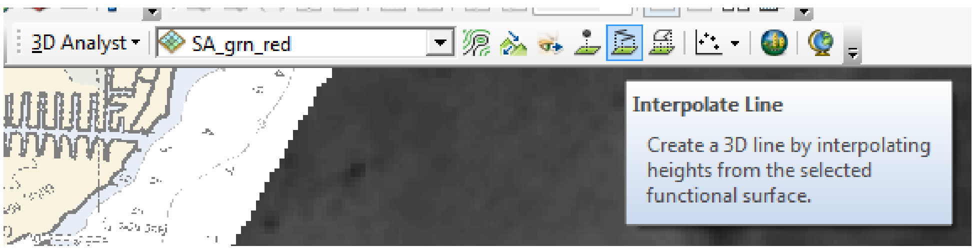
D. Removing the deep channels.

D.1. The algorithm layers created in step B are going to be used to mask out deep channels that are located throughout the study area. The reasoning for creating the algorithms are that the deep channels have the same optical signature as vegetation in the green wavelength and would be falsely identified as SAV in the proceeding steps. The algorithms are used to identify only the deep channels in order to remove them before proceeding to the steps that identify SAV. The blue/green algorithm and green/red algorithm were both created because certain satellite images will have a better detection of the deep channels of one or the other. Depending on the study area there may not be any deep channels or other optically similar areas to vegetation that need to be removed using this process.

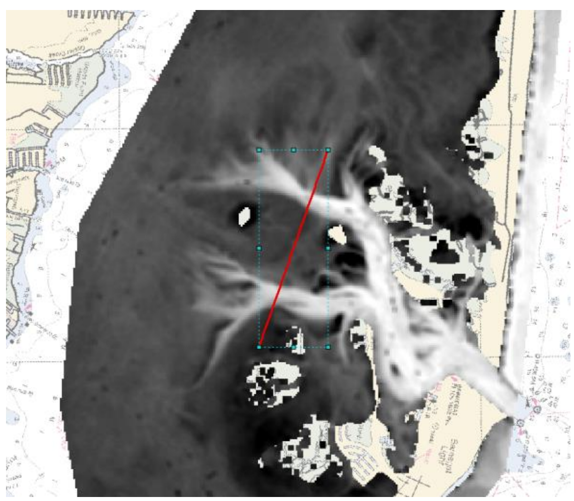
D.2. Determine which algorithm will be used to mask the channels by comparing the deep channel areas with the shallow areas for the SA_blu_grn and SA_grn_red. The deep channels should have a strong signal (either positive or negative) compared to the surrounding areas. It may be essential to examine RGB imagery of the area or to gain an understanding of the study area to in order to know the locations of the deep channels.

Based on these factors, this procedures SA_grn_red layer will be used to mask the deep channels. Note: determining which algorithm layer to use is based on visual inspection, it may be important to change the layers' symbology to get better contrast.

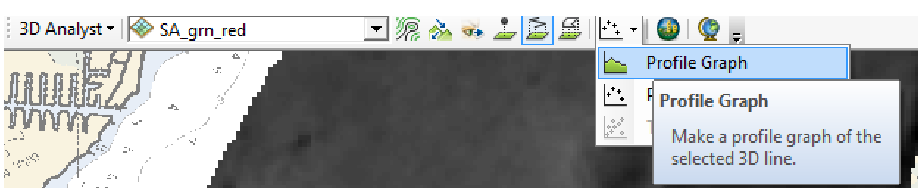
D.3. Once a layer has been chosen the threshold value which separates the deep channel from the non-deep areas must be determined. Using the 3D analyst toolbar is the easiest method. Select the SA_grn_red layer (or other depending on what was chosen) in the 3D analyst toolbar. Select **Interpolate Line** on the toolbar.



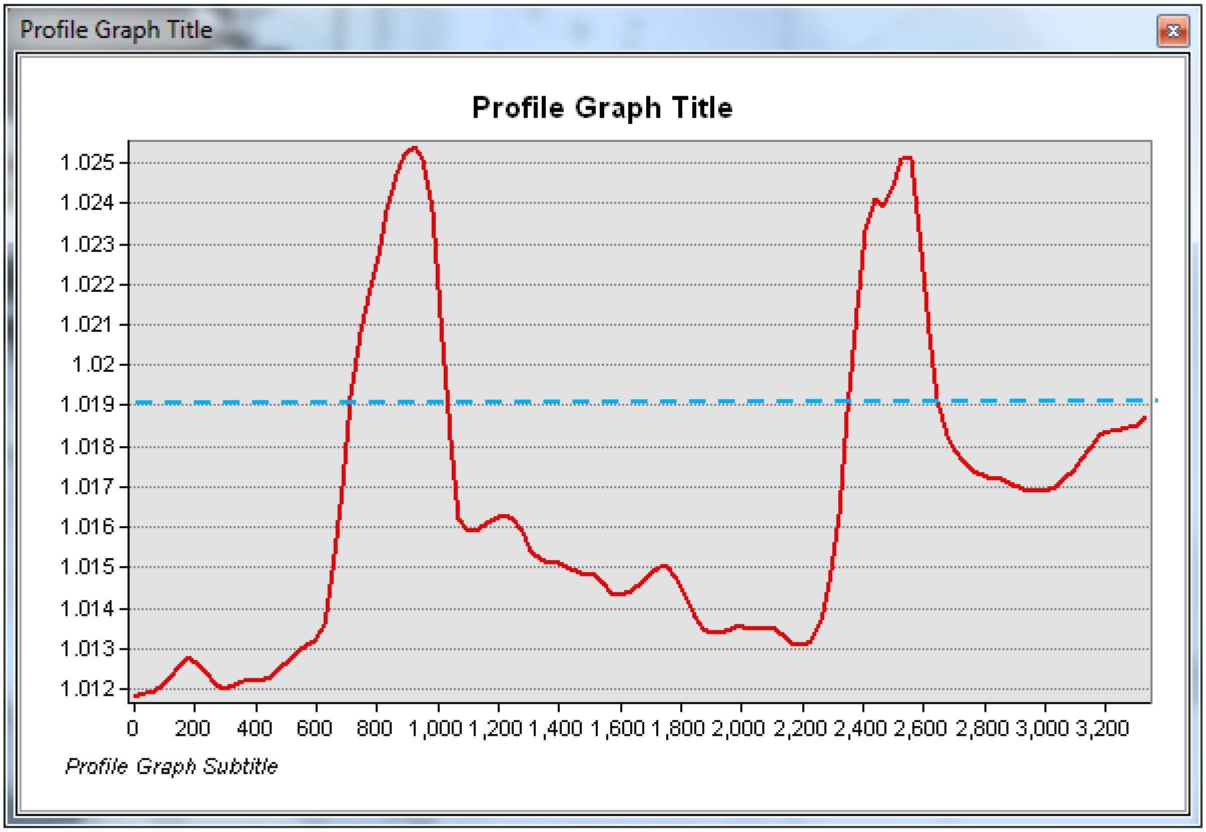
D.4. Create a line segment which crosses over at least one deep channel (make sure the line also goes over non-deep areas as well).



D.5. Select **Profile Graph** in the 3D analyst toolbar.

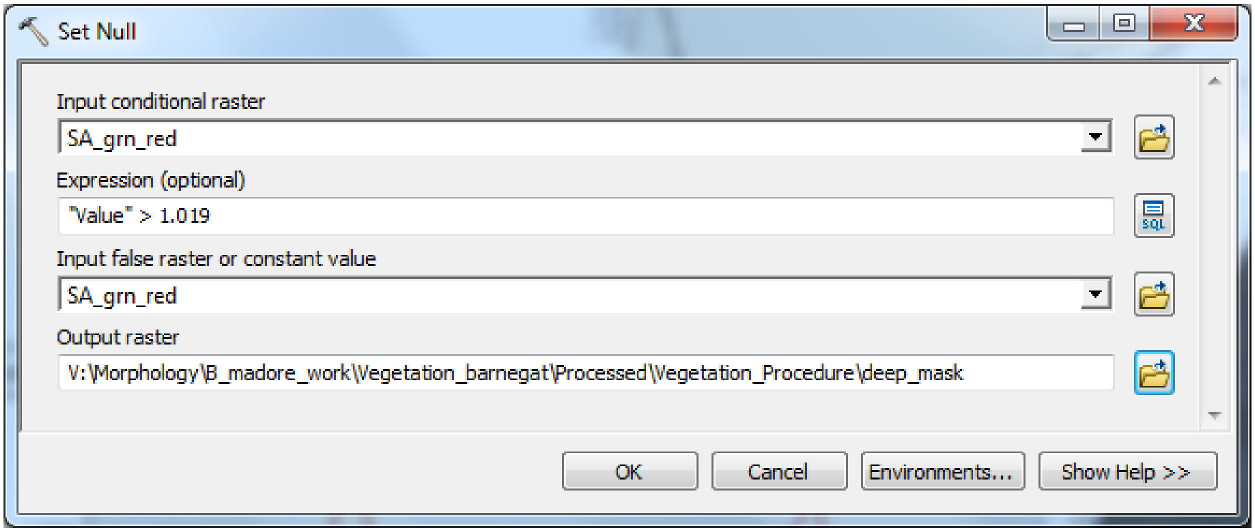


D.6. Determine where the threshold value between deep and non-deep water is by comparing the peaks of the graph for the deep channel areas. The following graph shows the corresponding deep channels at 800 – 1000 and at 2400 – 2600. In order to get rid of those peaks a threshold value of 1.019 will be used.



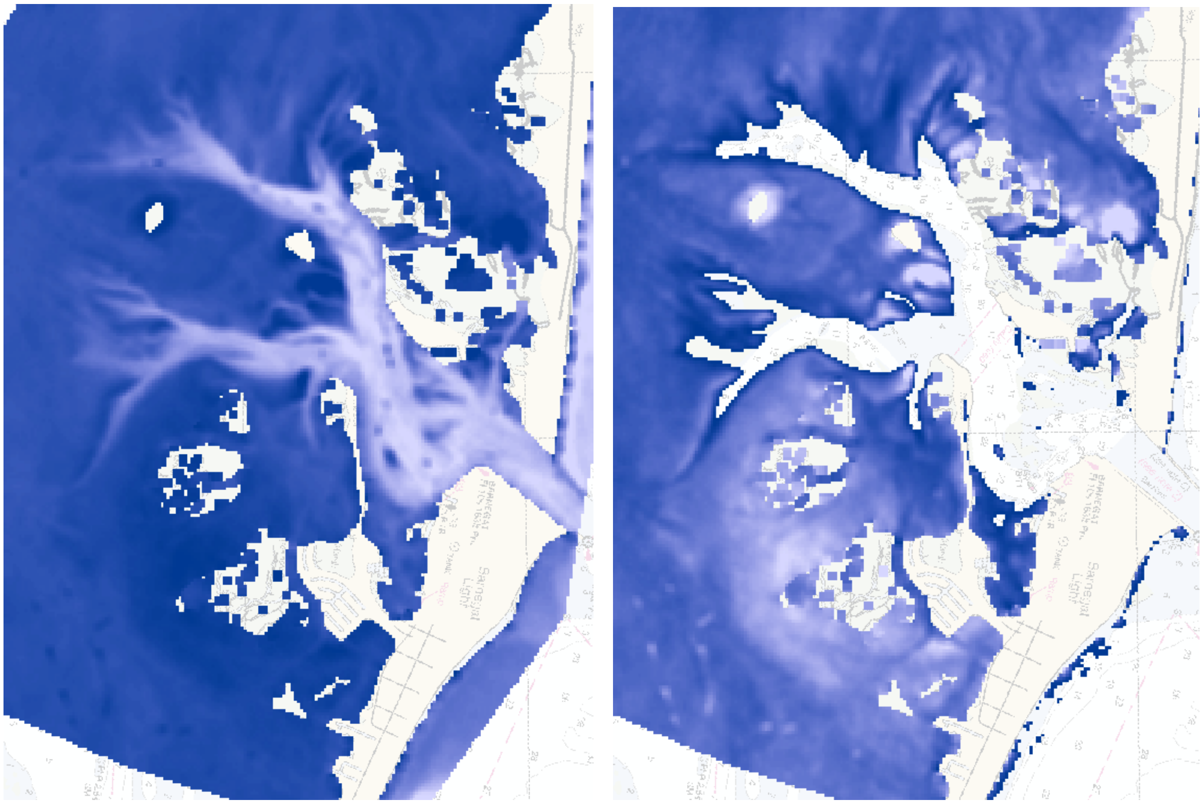
D.7. In the **ArcToolbox** select **Spatial Analyst Tools>Conditional>Set Null**.

D.8. Select SA_grn_red for the **input conditional raster** (or other). For the **Expression** use "Value" > (**Value determined from step D.6.**) Select SA_grn_red (or other) for the **false raster**. Save the **output raster** as deep_mask. See the following figure to help with the input for the **Expression.**



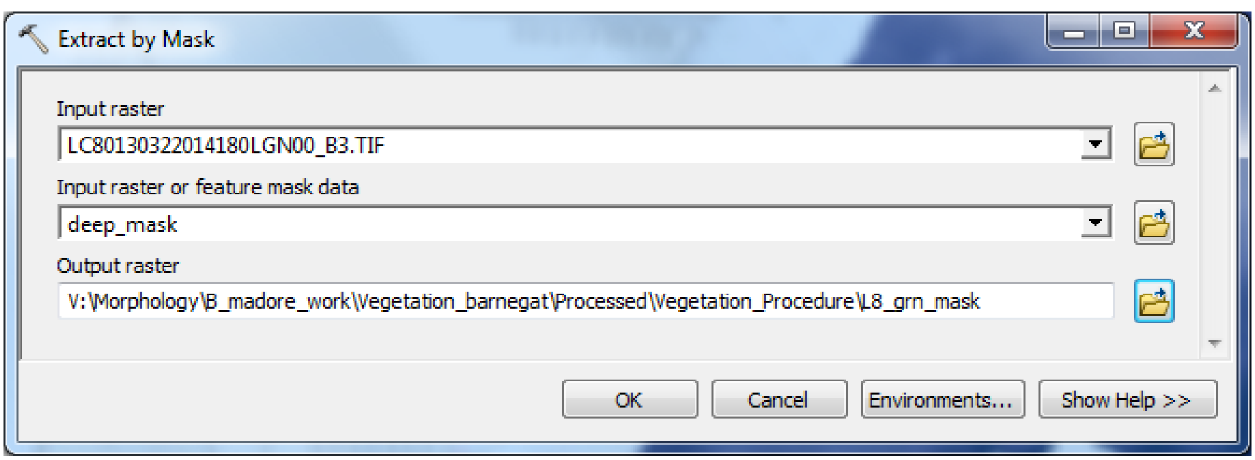
D.9. The deep_mask will cover the same area with the deep channels missing. It may be that the threshold value determined in D.6.was slightly off. If that is the case and a different threshold value should be tried then adjust the threshold value slightly in step D.8. until a coverage which eliminates the deep channels correctly is found. Note: Do not be overly strict with the channel removal. It is important to not remove shallow water areas.

D.10. The following figure shows the SA_grn_red on the left with the deep channels in lighter colors. The figure on the right is the deep_mask with the channels removed.



D.11. In the **ArcToolbox** select **Spatial Analyst Tools>Extraction>Extract by Mask**

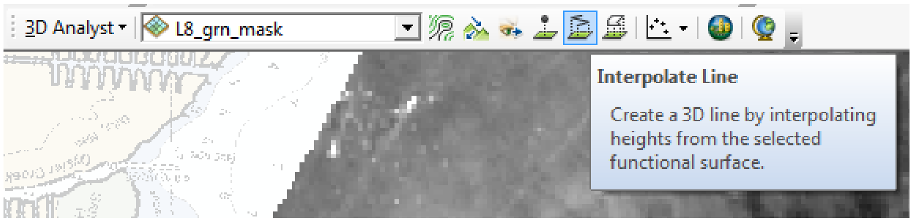
D.12. Select the green band from the satellite bands (this is band 3 for Landsat 8) for the **input raster**. Select the deep_mask for the **feature mask**. Save the **output raster** as L8_grn_mask (or other name, if desired).



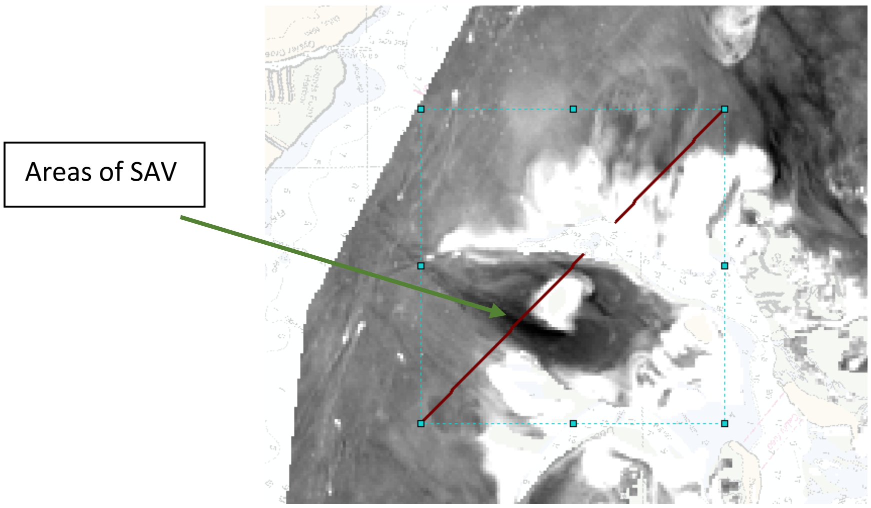
E. Determining the Extent of SAV

E.1. With the deep channels removed from the green band, the deep areas with the same optical signal as vegetation have been removed from the green band. Determining the extent of the SAV is similar to steps D.3. to D.6. where the threshold value between vegetated and non-vegetated needs to be determined.

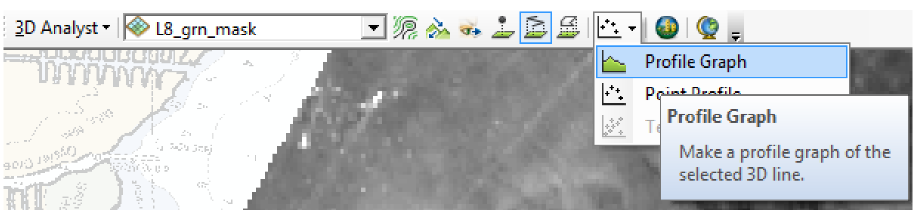
E.2. Select L8_grn_mask in the **3D analyst toolbar**. Select **interpolate line**.



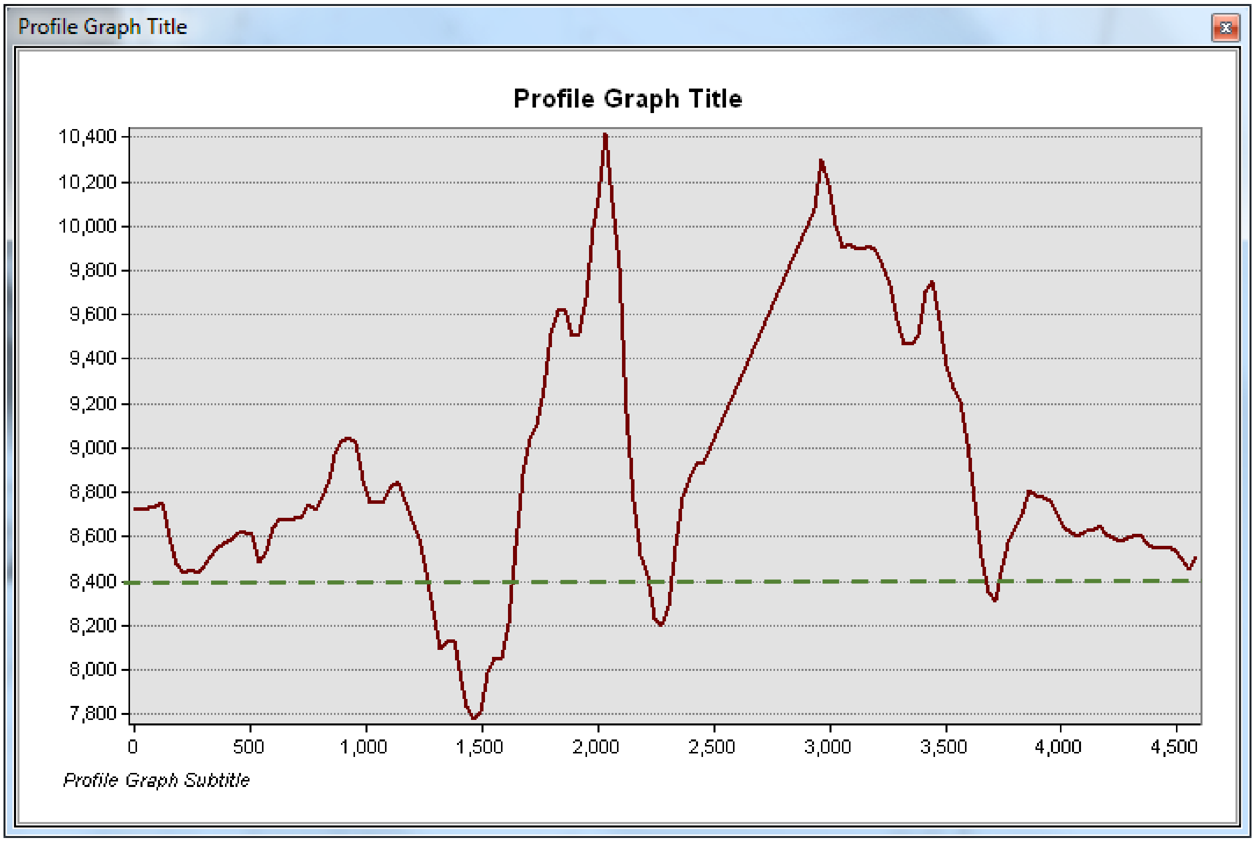
E.3. Draw a line segment that crosses over areas of vegetation (dark areas of L8_grn_mask) and non-vegetated areas as well. Note: for this particular satellite image the distinction between vegetation and non-vegetated values is very low. Using the symbology tab to adjust the visual properties is helpful in distinguishing visually the vegetated areas.



E.4. Select **Profile Graph** in the **3D analyst toolbar**.

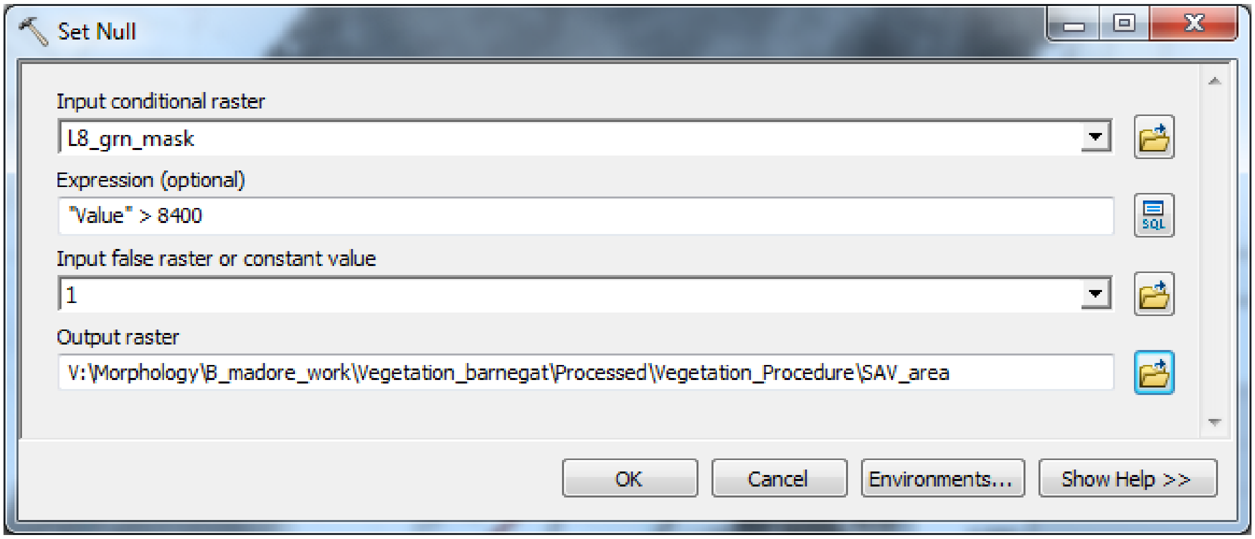


E.5. Examine the graph to find a threshold value that distinguishes vegetated regions from non-vegetated areas. For this graph the vegetated areas were determined to be at areas from 1300 – 1600 and 2200 – 2300. Based on these areas the vegetated areas are found at regions below a value of 8400.



E.6. In the **ArcToolbox** select **Spatial Analyst Tools>Conditional>Set Null**.

E.7. For the **input raster** select L8_grn_mask. For the **Expression** use "Value" > (**value determined in E.5.**) For the **false raster** input 1. Save the **output raster** as SAV_area. See the following figure for help inputting the **expression** correctly.



E.8. If the resulting SAV region needs to be adjusted try changing the threshold value in small amounts.

E.9. The SAV regions have now been determined. The figure on the left is L8_lpf_green and the figure on the right is L8_lpf_green with SAV_area (in green) covering it.

F. Confusion Matrix

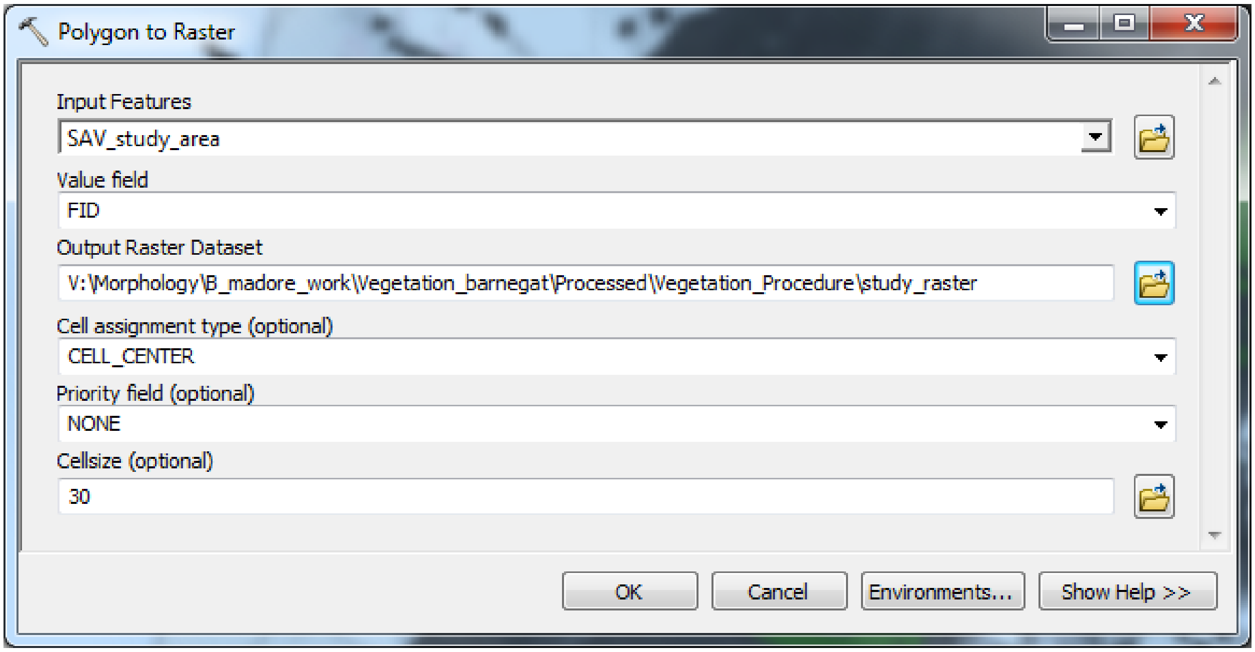
F.1. The confusion matrix is a technique used to statistically analyze the results of certain datasets. For the SAV datasets the confusion matrix is going to be used to compare the regions that were determined to be SAV (SAV_area layer created in step E.8.) with a dataset collected by Rutgers Universisty at the Grant F. Walton Center for Remote Sensing and Spatial Analysis. This dataset is from 2009 and gives SAV coverage over Barnegat Bay, NJ and will be the actual dataset (called reference dataset in this guide, as well), whereas the SAV_area will be the predicted dataset. The confusion matrix is going to compare vegetation and non-vegetated regions for the predicted and actual datasets. NOTE: Raster calculator is going to be used to determine the values for the confusion matrix, therefore some layers are going to be converted from shapefile/polygon to rasters.

The following table is how the confusion matrix is set up. The values representing when each case is correct will be explained as the procedure continues.

N=?	Predicted: Vegetated	Predicted: Non-vegetated
Actual/Reference: Vegetated	When value = 11	When value = 21
Reference: Non-vegetated	When value = 10	When value = 20

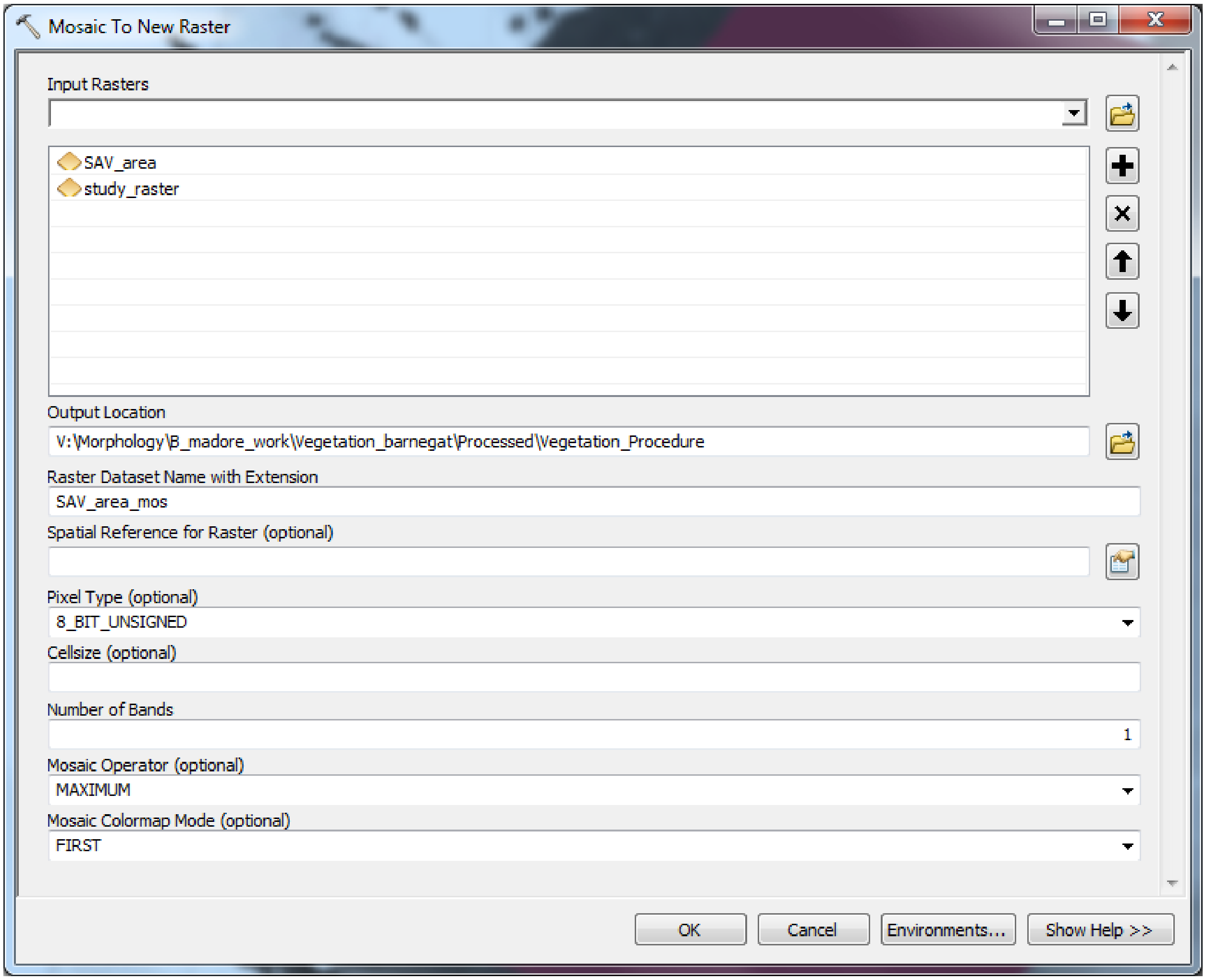
F.2. Since the SAV_area has only values for vegetated areas, a background needs to be added for all areas that are non-vegetated. In the **ArcToolbox** select **Conversion Tools>To Raster>Polygon to Raster.**

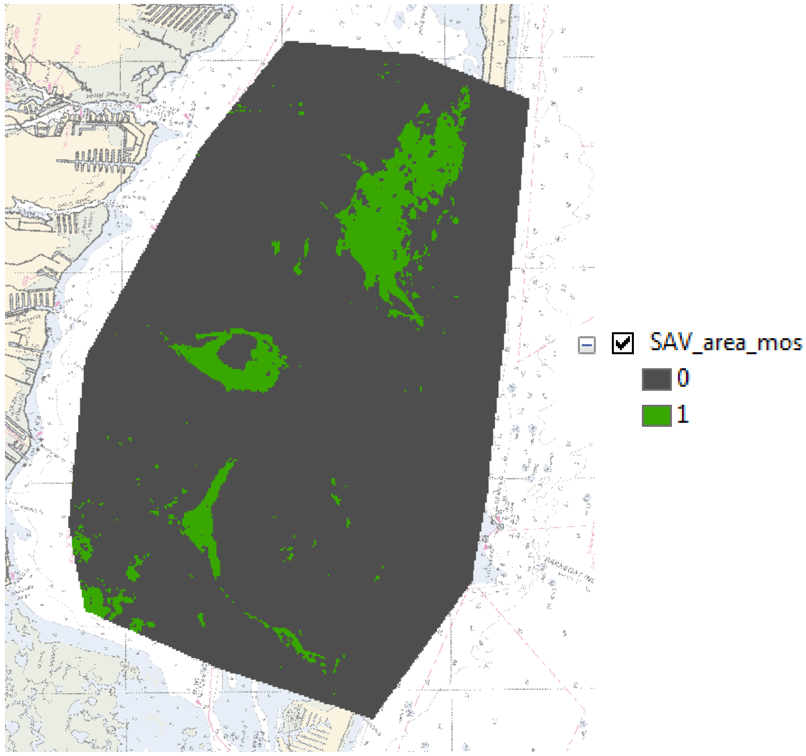
F.3. Select the SAV_study_area for the **input features**. Save the **output raster** as study_raster. Select 30 for the cellsize (this is because the Landsat 8 satellite imagery is at 30 meter resolution. If another imagery type was used to create the SAV_area use the corresponding cellsize and not 30). **IMPORTANT:** Make sure study_raster has only values of 0 when created.



F.4. In the **ArcToolbox** select **Data Management Tools>Raster>Raster Dataset>Mosaic To New Raster**.

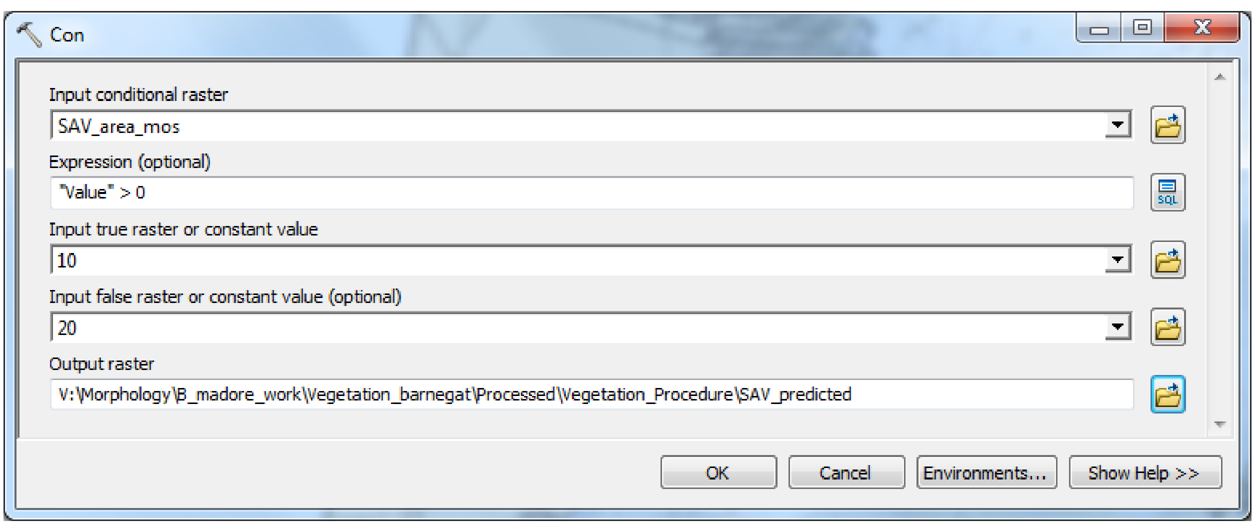
F.5. For the **input rasters** add SAV_area and study_raster. Set the **output location** to the folder where the raster will be saved. Use **Raster dataset name with extenstion** to name the raster as SAV_area_mos. Set **number of bands** to 1. Set **Mosaic Operator** to MAXIMUM. The resulting raster will contain values of 1 for SAV areas and a value of 0 for non-vegetated areas.


Mosaic To New Raster
Input Rasters
SAV_area
study_raster
Output Location
V:\Morphology\B_madore_work\Vegetation_barnegat\Processed\Vegetation_Procedure
Raster Dataset Name with Extension
SAV_area_mos
Spatial Reference for Raster (optional)
Pixel Type (optional)
8_BIT_UNSIGNED
Cellsize (optional)
Number of Bands
1
Mosaic Operator (optional)
MAXIMUM
Mosaic Colormap Mode (optional)
FIRST
OK
Cancel
Environments...
Show Help >>

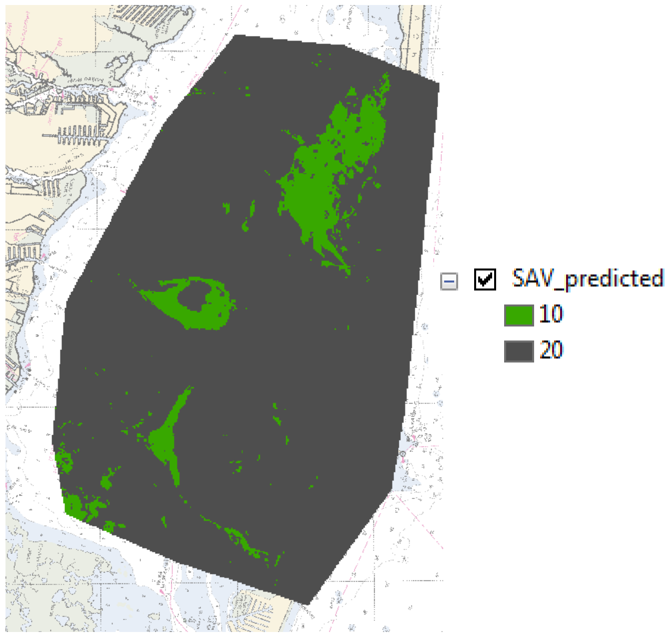


F.6. In the **ArcToolbox** select **Spatial Analyst Tools>Conditional>Con.**

F.7. For the **input raster** select SAV_area_mos. Set the **Expression** to: “Value” > 0 Set the **input true raster** to 10. Set the **false raster** to 20. Save the **output raster** as SAV_predicted. This step makes the SAV areas have a value of 10 and the non-vegetated areas have a value of 20.



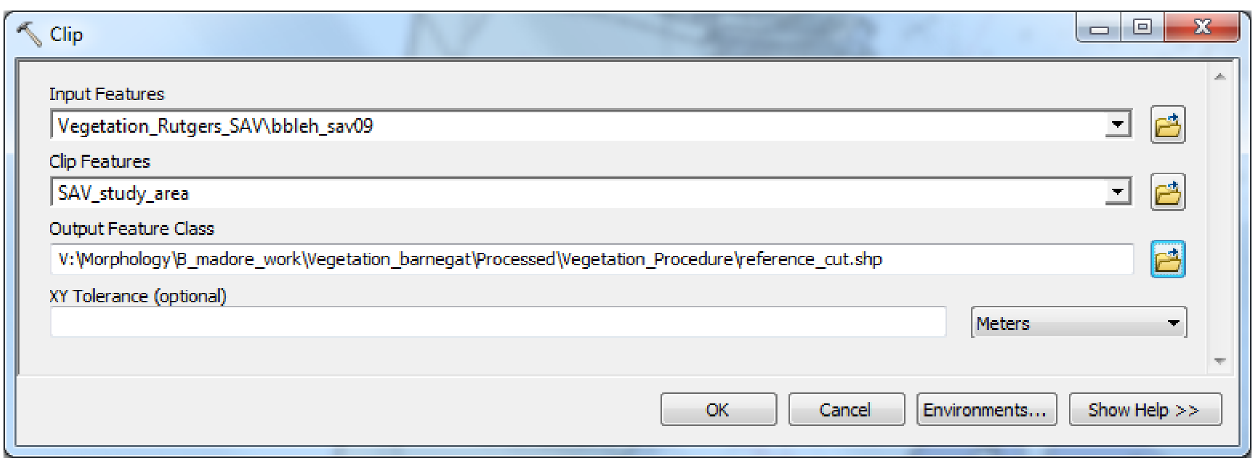
F.8. The resulting SAV_predicted will have two values. 10 for vegetated regions determined from the SAV_area from step E in this guide and value 20 for all other areas covering the study area.



F.9. The next steps are used to change the Rutgers dataset for SAV to a raster file. If the dataset being used to compare your SAV_area is not a raster file and is a shapefile, follow these steps to convert it into a raster layer.

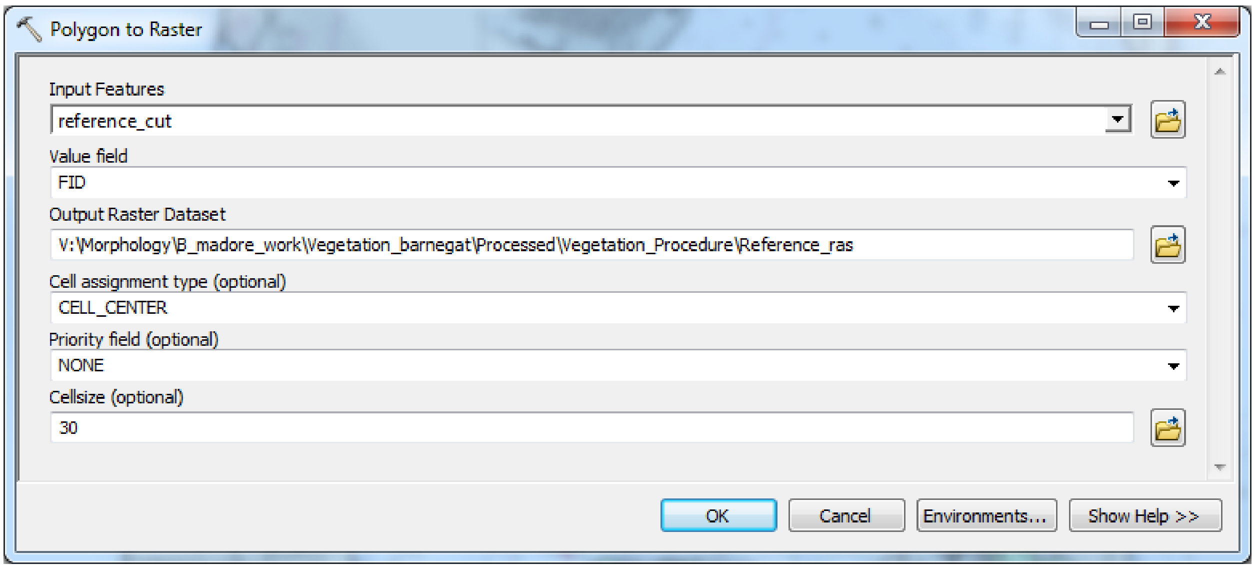
F.10. In the **ArcToolbox** select **Analysis Tools>Extract>Clip**.

F.11. For the **input feature** select the reference dataset you are utilizing. For the **clip features** select SAV_study_area. Save the **output feature class** as reference_cut.



F.12. In the **ArcToolbox** select **Conversion Tools>To Raster>Polygon to Raster**.

F.13. For the **input feature** select reference_cut. The value field is going to be dependent on how the shapefile is set up, but FID should be acceptable as long as the reference dataset is of only SAV regions, otherwise this may need to change. Select **output raster** and name it reference_ras. Set the **cellsize** to the same value as being used in step F.3. (Reminder: this depends on the satellite imagery used to create the predicted SAV_area).

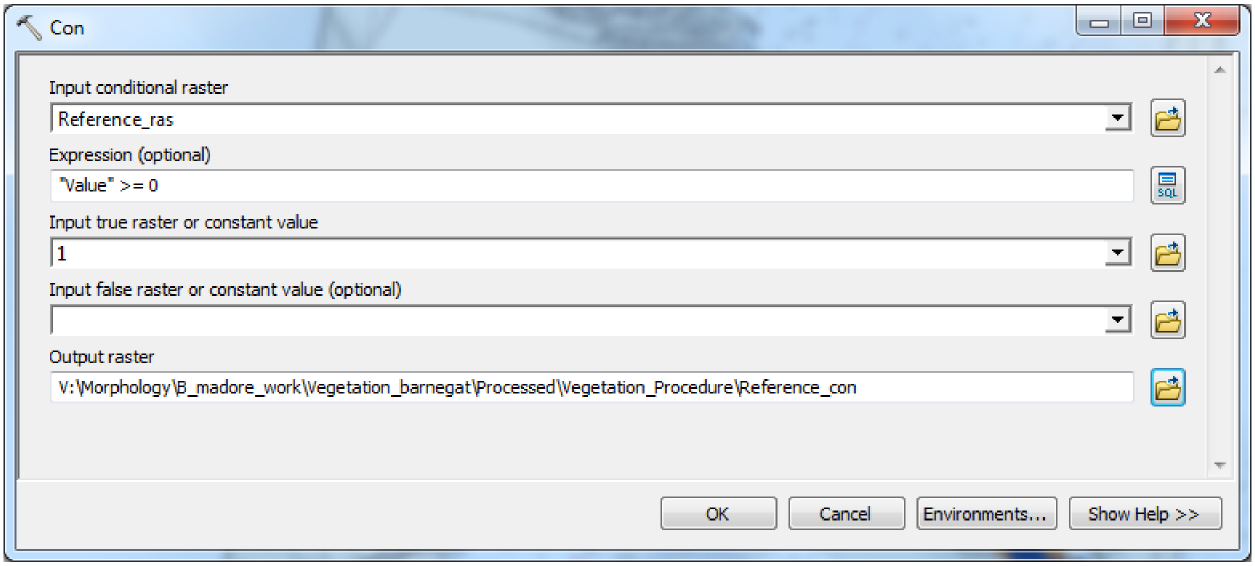


F.14. In the **ArcToolbox** select **Spatial Analyst Tools>Conditional>Con**.

F.15. For the **input raster** select Reference_ras. For the **Expression** use:

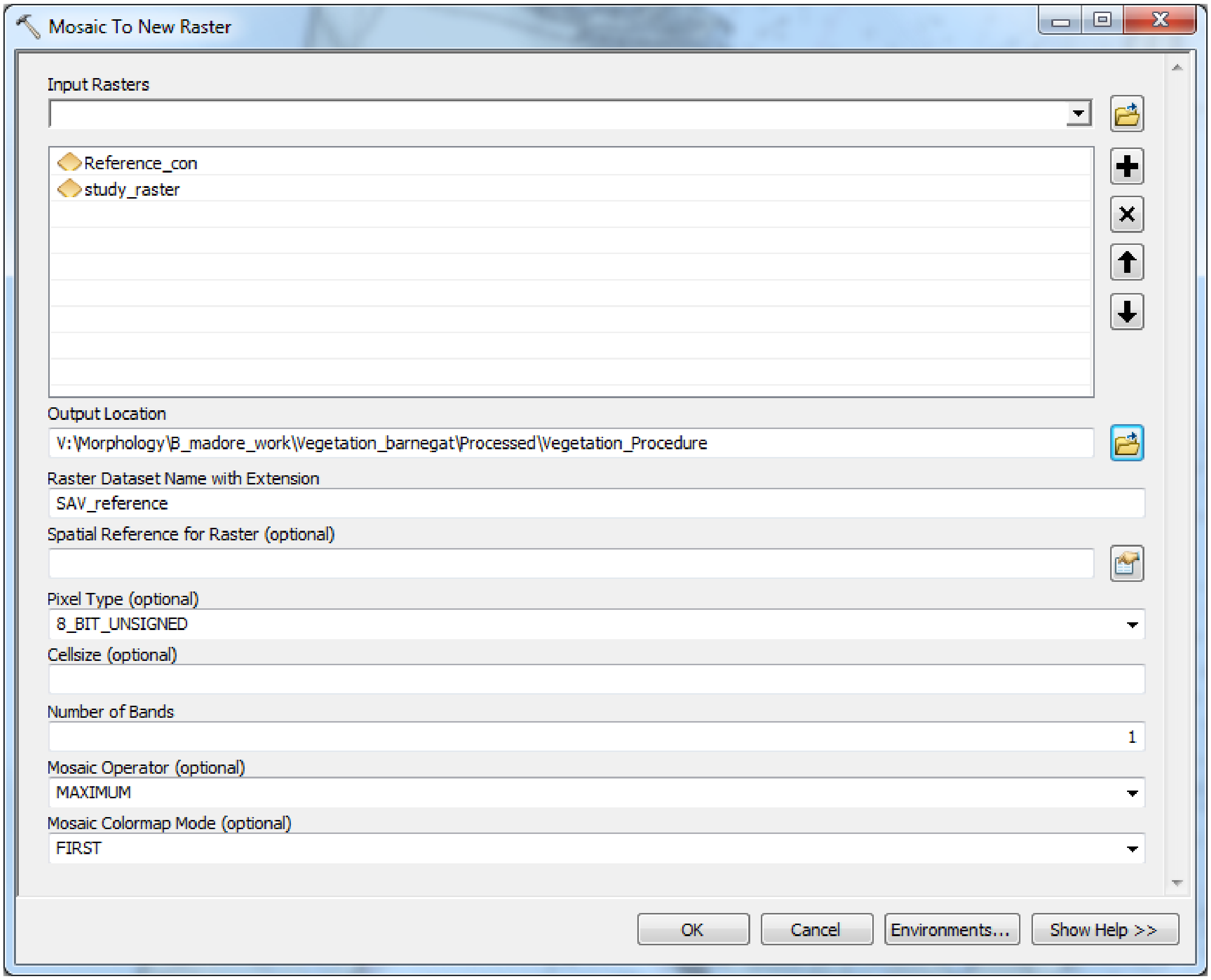
“Value” >= 0

Set the **true raster** value to 1 and leave the **false raster** blank. Save the raster as Reference_con.

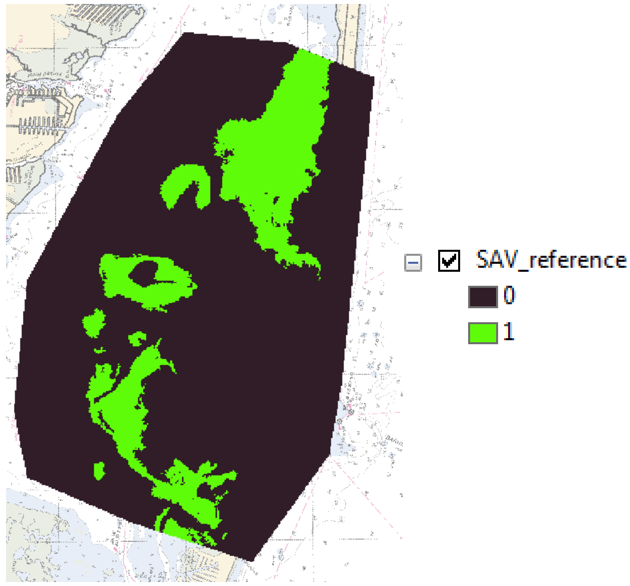


F.16. In the **ArcToolbox** select **Data Management Tools>Raster>Raster Dataset>Mosaic To New Raster**.

F.17. Add Reference_con and study_raster for the **input rasters.** Select the folder to save the raster in the **output location.** Name the raster as SAV_reference in **Raster dataset name with extension**. Set **number of bands** to 1. For the **Mosaic operator** choose MAXIMUM.



F.18. The resulting raster file for SAV_reference should have values of 1 for SAV regions and values of 0 for the non-vegetated areas.

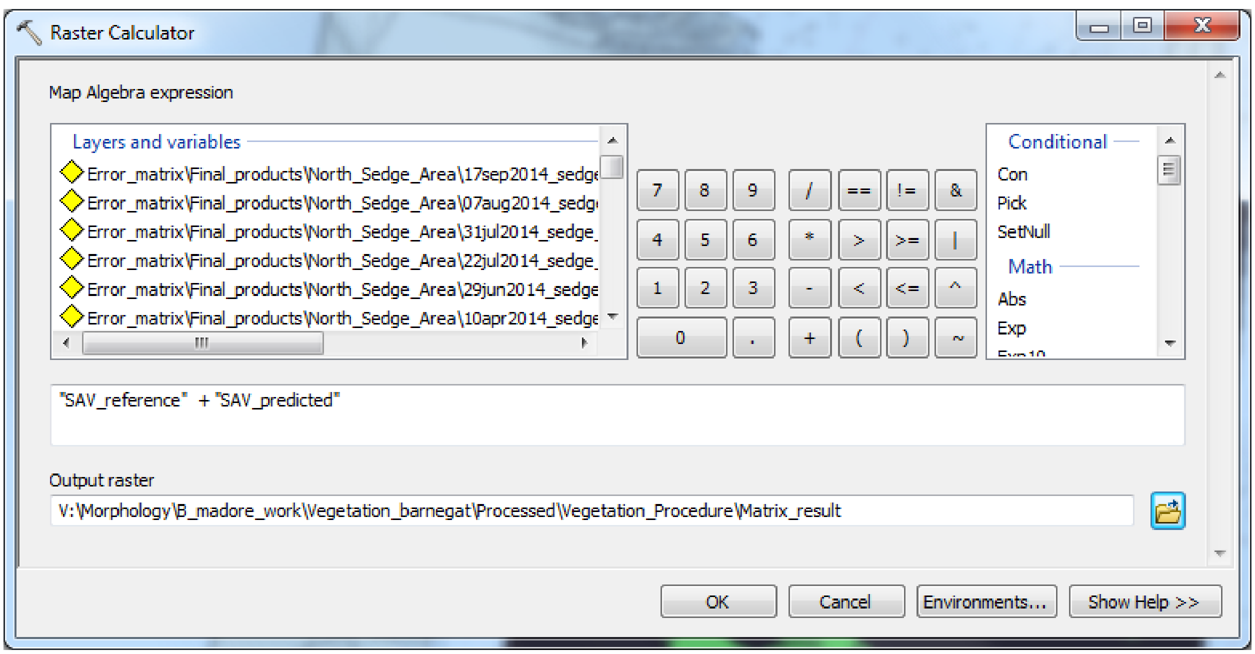


F.19. In the **ArcToolbox** select **Spatial Analyst Tools>Map Algebra>Raster Calculator**.

F.20. For the expression use:

"SAV_reference" + "SAV_predicted"

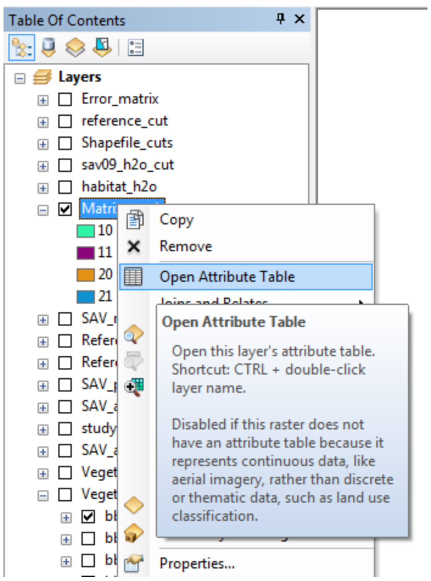
Save the **output raster** as Matrix_result.



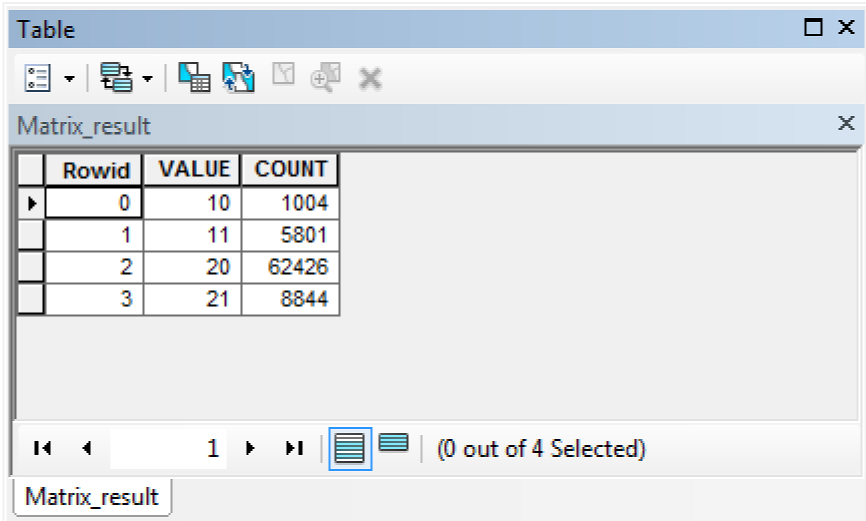
F.21. The resulting layer will have 4 unique values: 10, 11, 20, and 21. Each of these values corresponds to a different cell in the confusion matrix. The following table shows the corresponding results.

Value	Predicted Layer	Reference Layer
10	Vegetated Cell	Non-Vegetated Cell
11	Vegetated Cell	Vegetated Cell
20	Non-Vegetated Cell	Non-Vegetated Cell
21	Non-Vegetated Cell	Vegetated Cell

F.22. The amount of cells for each value, which can be translated to area, can be obtained by right-clicking on the Matrix_result layer in the table of contents and selecting **Open Attribute Table.**



F.23. The table shows the Value and the corresponding count which is the number of cells with that value. Using the cell size (which is 30 meter by 30 meter for this procedure) the area corresponding to each cell size can be obtained by multiplying COUNT by the cell size.



Table

Matrix_result

Rowid	VALUE	COUNT
0	10	1004
1	11	5801
2	20	62426
3	21	8844

(0 out of 4 Selected)

Matrix_result

F.24. Computing the accuracy of the confusion matrix can be done by adding the COUNT for the VALUE 11 with VALUE 20, and then dividing that by the total count.

(5801 + 62426)/(1004 + 5801 + 62426 + 8844) = .8739

G. Supplemental Confusion Matrix notes.

G.1. The confusion matrix example provided for this procedure in part F focused on comparing vegetation detection versus non-vegetated areas. With different imagery datasets the confusion matrix can be modified to include more fields. This include specific types of vegetation, density of vegetation, or various other detections/predictions that statistics want to be calculated for.

G.2. The changes that would be made to the datasets would involve creating rasters that have more values than the binary components created above. In order to do this slight changes to the procedure need to be made.

G.3. Adding fields to the prediction layer (beyond 2)

G.3.1. In part F the prediction layer was created by first mosaicing the two different datasets, then using the conditional tool to change the values to the confusion matrix values. For more than 2 different components in the prediction layer each raster will first be changed to its appropriate confusion matrix value then all the rasters will be mosaiced together.

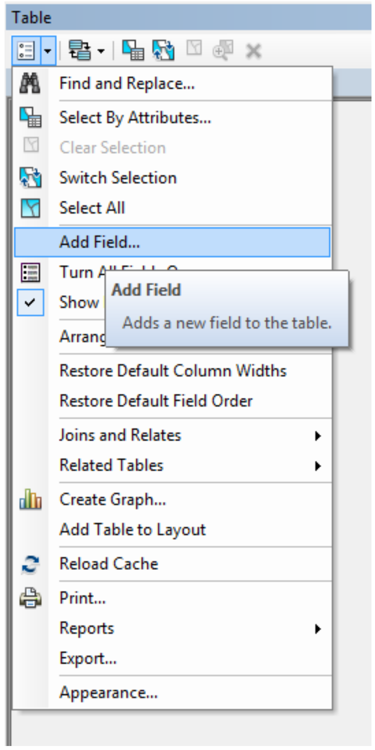
G.3.2. For the values to be used for the prediction layer of the confusion matrix it is important to use values of 10. Therefore this example with 3 different density type predictions of SAV the following values will be used. **NOTE:** If there are more components of the prediction layer being examined simply add more components. As long as the numeric value is a factor of 10 the confusion matrix process will work correctly.

Numeric Value	Prediction Layer Description of SAV
10	Dense Vegetation
20	Moderate Vegetation
30	Sparse Vegetation
40	Non-Vegetated Region

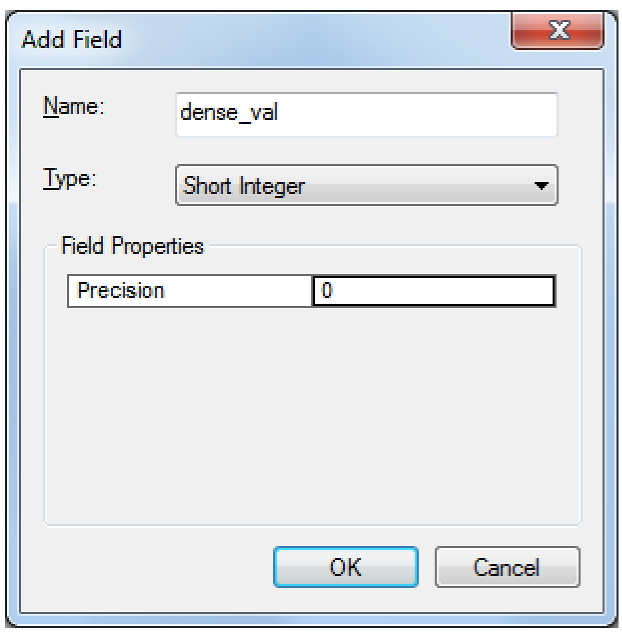
G.3.3. If the individual layers are still in shapefiles use the following steps. This will convert the values during the raster making process to the appropriate prediction layer value. This example will show the conversion of the dense vegetation prediction layer shapefile.

G.3.3.1. Right-click the shapefile layer (Dense_prediction for this example) in the table of contents and select **Open Attribute Table.**

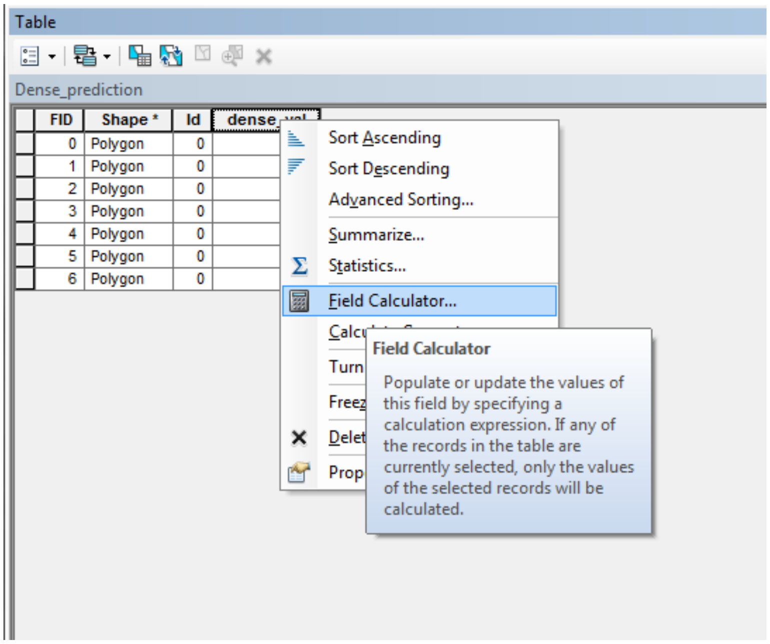
G.3.3.2. In the table options select **Add Field**.



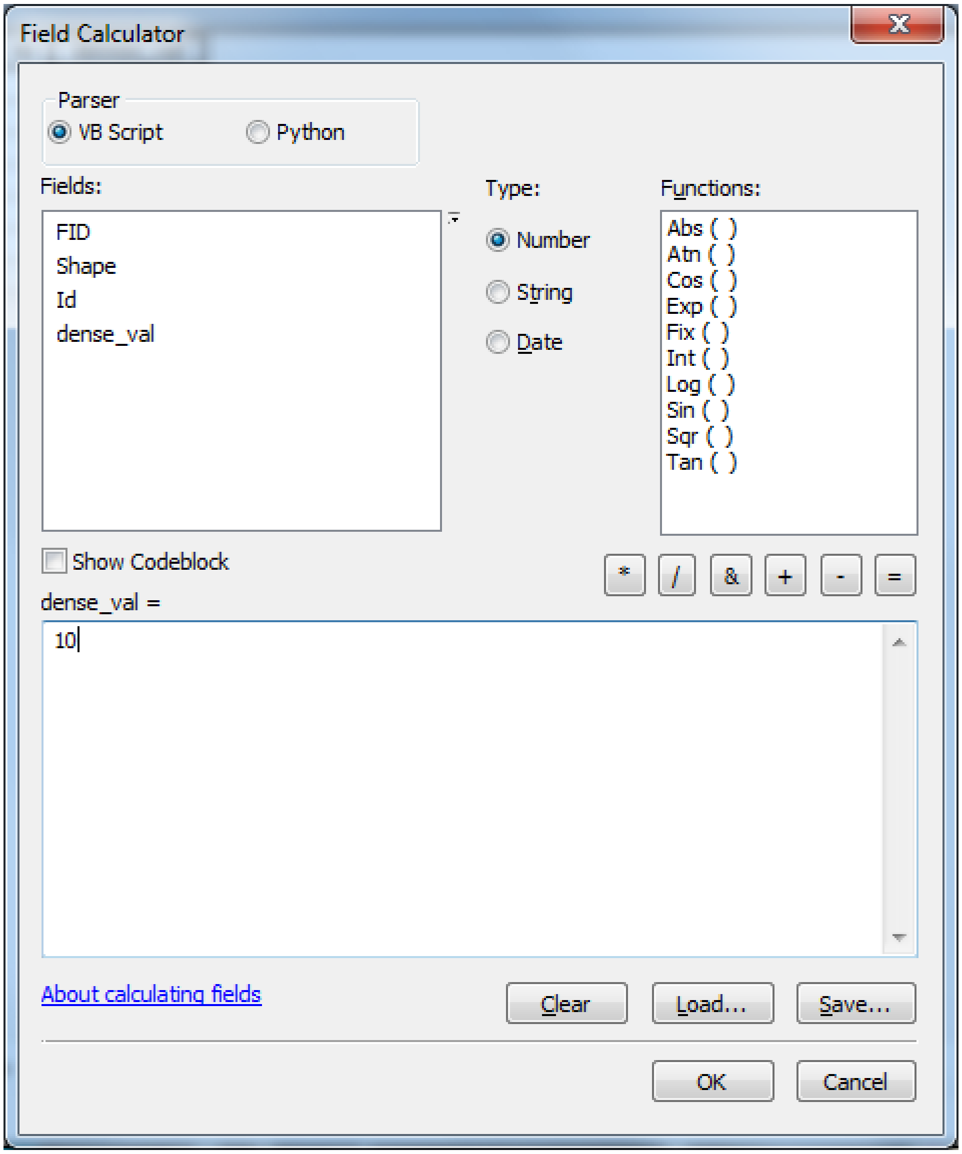
G.3.3.3. Name the field dense_val (or whatever is appropriate for the dataset being studied). Set the type as short integer, select OK.



G.3.3.4. Since this polygon covers only the dense predictions, all of the components to shapefile need to have the dense_val field changed to the appropriate value for the dense vegetation for the confusion matrix. This value is 10 as shown in the table at G.3.2. Right-click the dense_val header in the attribute table and select **Field Calculator**.



G.3.3.5. Input the value 10 into the dense_val expression window and select OK.

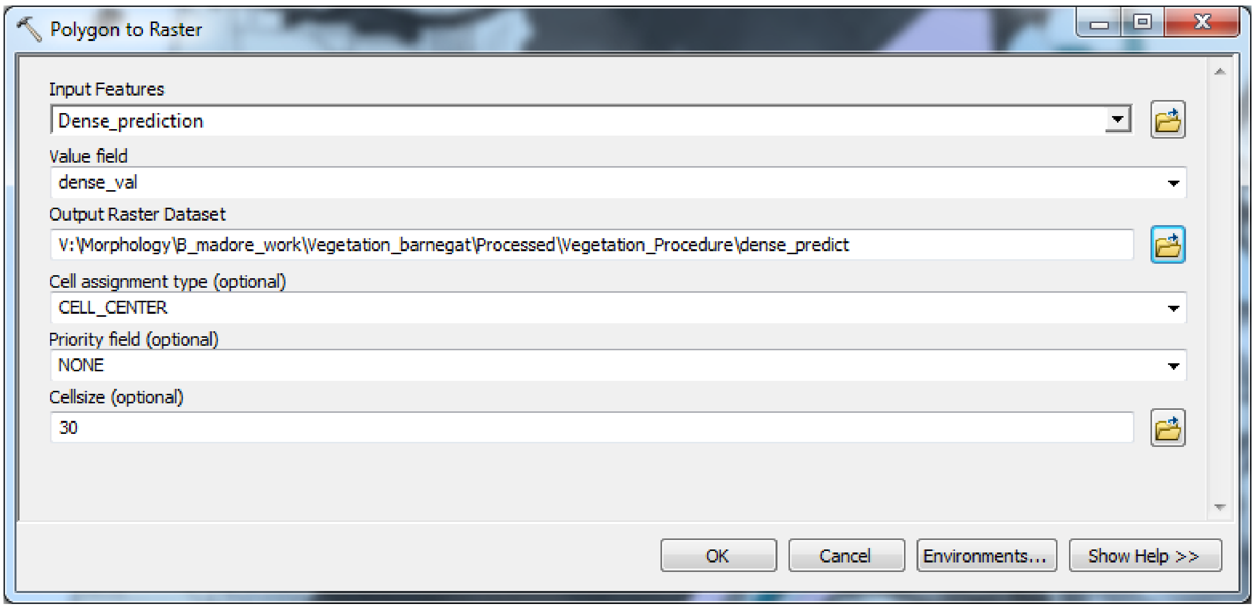


G.3.3.6. Repeat steps G.3.3.1 to G.3.3.5 for each different component of the prediction layer that is a shapefile, giving the appropriate value in step G.3.3.5 that corresponds to the confusion matrix in step G.3.2 (or the confusion matrix values chosen for the specific project at hand). For this example a moderate_prediction and sparse_prediction shapefile have also been processed.

G.3.3.7. To create the background layer (aka the non-vegetated layer) repeat steps G.3.3.1 to G.3.3.5 with the SAV_study_area shapefile created in part C (this shapefile will be used to cover the study site and represent the background for all non-vegetated areas) and give it a value of 40 (or the appropriate value based on your data/confusion matrix).

G.3.3.8. In the **ArcToolbox** select **ConversionTools>To Raster>Polygon to Raster** for converting the shapefiles into rasters.

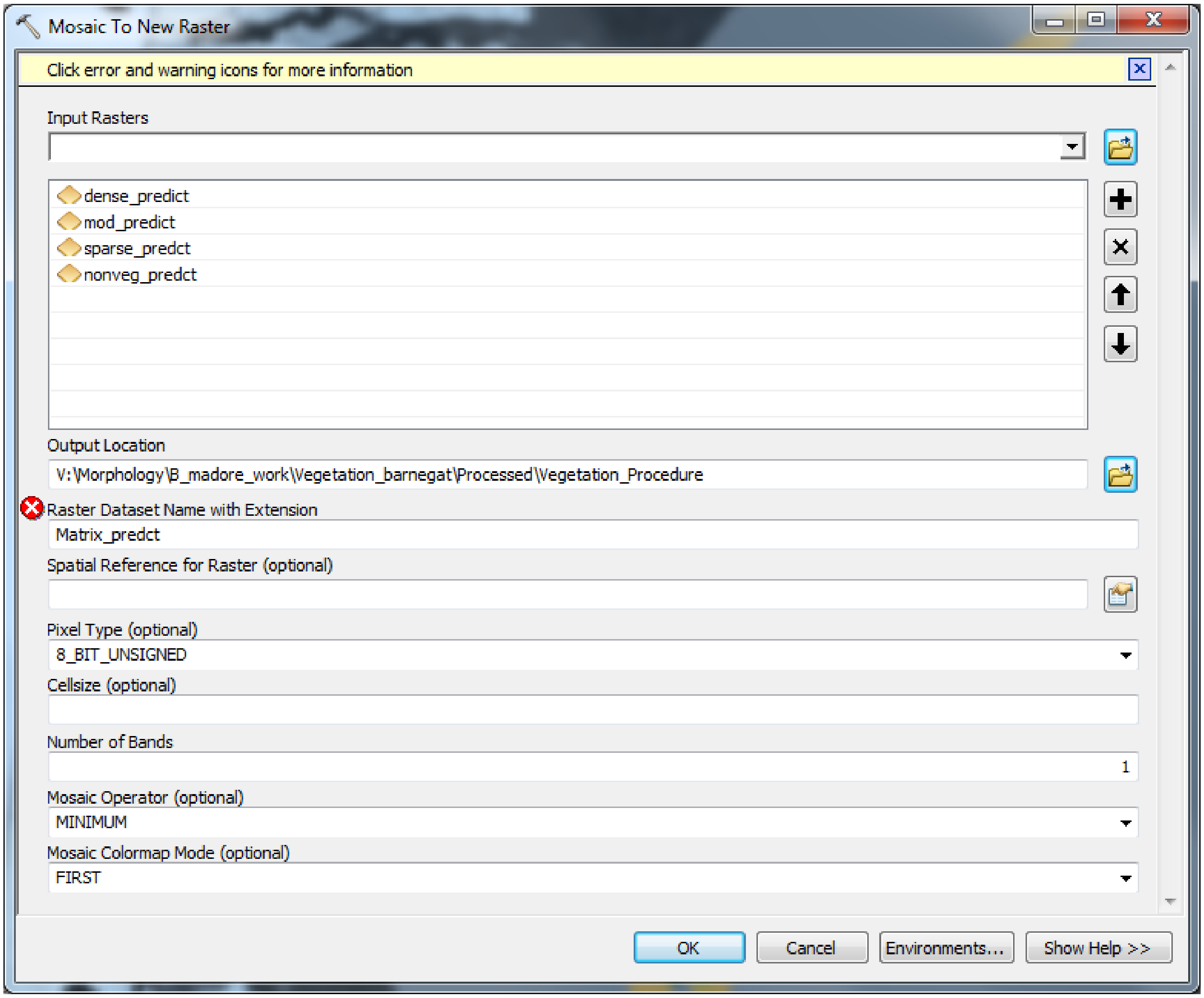
G.3.3.9. Input the dense_prediction shapefile (or other). Set the **value field** to dense_val (or the name of the field created which corresponds to the layers confusion matrix value). Save the dataset as dense_predict and set the **cellsize** to 30 (or to the appropriate value based on your datasets. This uses 30 meter cellsize as Landsat 8 is in 30 meter resolution).



G.3.3.10. Repeat step G.3.3.9. for each part of the prediction layer of the confusion matrix. For this example that includes the moderate prediction layer, spare layer, and the SAV_study_area (which is the non-vegetated layer).

G.3.3.11. In the ArcToolbox select **Data Management Tools>Raster>Raster Dataset>Mosaic To New Raster**.

G.3.3.12. Add all the rasters created in step G.3.3.9 and G.3.3.10 for the prediction part of the confusion matrix. Select the **output location** for which folder the file will be saved in. Set the name of the layer being created under **Raster Dataset name with extension**. Set the **number of bands** as 1. For the **Mosaic Operator** set the value based on how your datasets are set up. For overlapping components of this example a higher density layer will take priority (therefore if a cell contains both dense and moderate as a value the dense layer will take priority) (ALSO since the non_veg layer have the highest value in this set up, it should have least priority, feel free to use the LAST and FIRST decisions to customize which layers have priority when mosaicing the multiple rasters together). This example will use minimum for the operator.

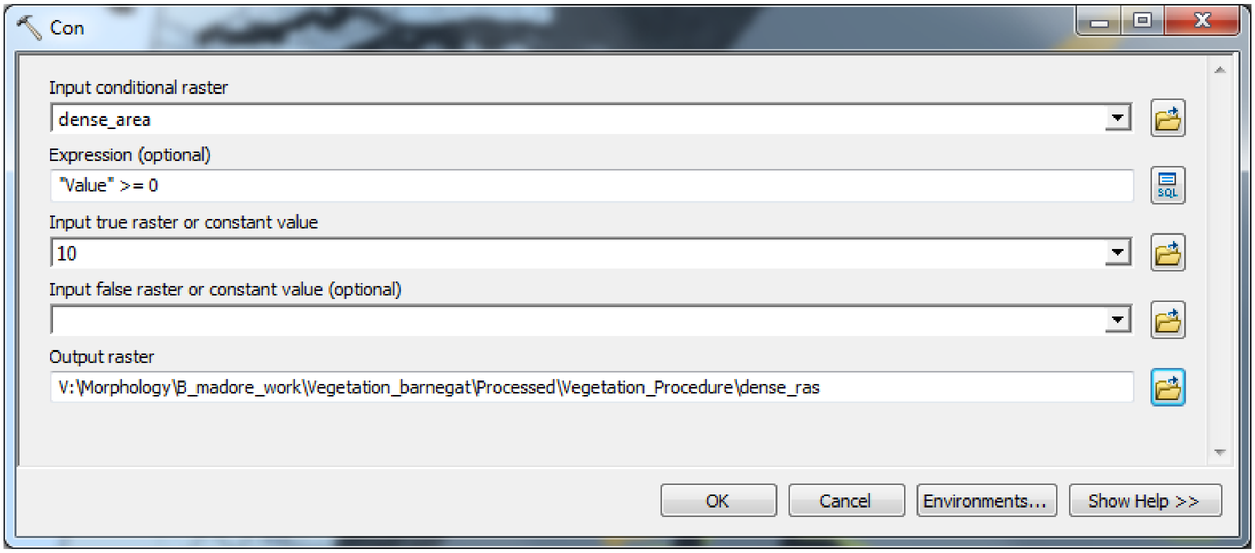


G.3.4. If the prediction layers are already individual rasters but with the incorrect values the Con tool will be used to change the values then the layers will be mosaiced together. It may be helpful to still read section G.3.3 entirely through to understand the process.

G.3.4.1. In the **ArcToolbox** select **Spatial Analyst Tools>Conditional>Con**.

G.3.4.2. For the **input** select the prediction layer that is having its value changed (for this example it is the dense vegetation area which is being changed to have the correct raster values associated with the table at G.3.2.). Set the **Expression** to encompass all values of the raster being changed, for this example the following is used: **"Value" >= 0**

Set the **true raster** value corresponding to the confusion matrix table for the association prediction layer. Save the **output raster** as a name fit for the layer being changed.

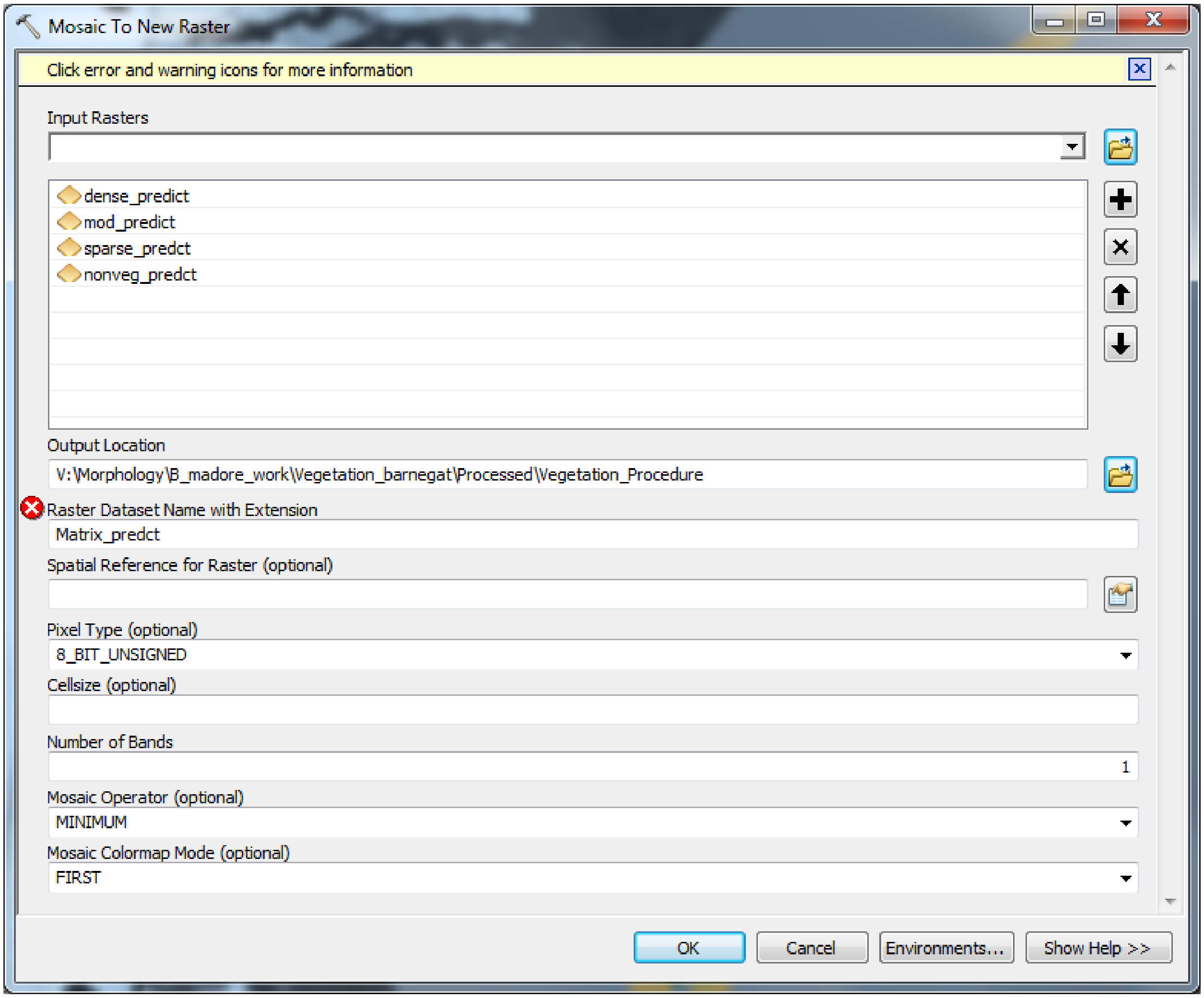


G.3.4.3. Repeat step G.3.4.1 to G.3.4.2 for each raster that needs to have its values changed to the appropriate confusion matrix table value.

G.3.4.4. In the ArcToolbox select **Data Management Tools>Raster>Raster Dataset>Mosaic To New Raster**.

G.3.4.5. Add all the rasters created in step G.3.4.3 and others (if applicable) for the prediction part of the confusion matrix. Select the **output location** for which folder the file will be saved in. Set the name of the layer being created under **Raster Dataset name with extension**. Set the **number of bands** as 1. For the **Mosaic Operator** set the value based on how your datasets are set up. For overlapping components of this example a higher density layer will take priority (therefore if a cell contains both dense and moderate as a value the dense layer will take priority) (ALSO since the non_veg layer have the highest value in this set up, it should have least priority, feel free to use the LAST and

FIRST decisions to customize which layers have priority when mosaicing the multiple rasters together). This example will use minimum for the operator.



G.4. Creating the actual (will be called reference in this procedure as well) layer for the confusion matrix may follow some steps described in G.3. but most likely the reference dataset will come from an outside data source and therefore will not be created by the user. Therefore only modifications to the layer will need to be made.

G.4.1. The most important thing about the reference layer is the value scheme that it will use for the confusion matrix. Whereas the prediction layer used factors of 10, the actual layer will use single digit values. Therefore for this example where dense, moderate, sparse, and non vegetated areas are being compared the following values will be used. **NOTE:** if your work requires or has more components simply add another digit value for each layer. However, this will only work for values up to 9.

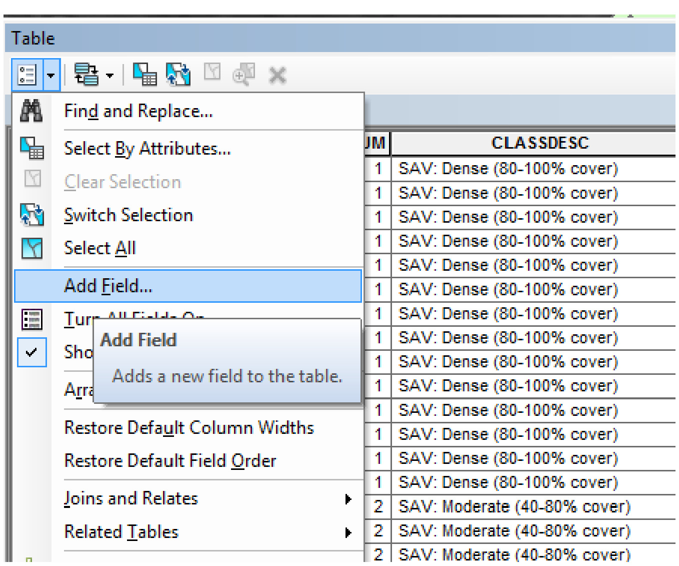
Numeric Value	Reference Layer Description of SAV
0	Non-Vegetated
1	Dense
2	Moderate
3	Sparse

G.4.2. This example for the procedure will assume a shapefile containing the appropriate SAV files was given to the user to compare against their own work. Therefore the changes that need to be changed are giving the shapefile the appropriate values for the correct description of SAV, convert the shapefile to a raster, and mosaic a non_vegetated background layer to the raster. **NOTE**: This assumes the shapefile has been clipped to the study area as described in F.10

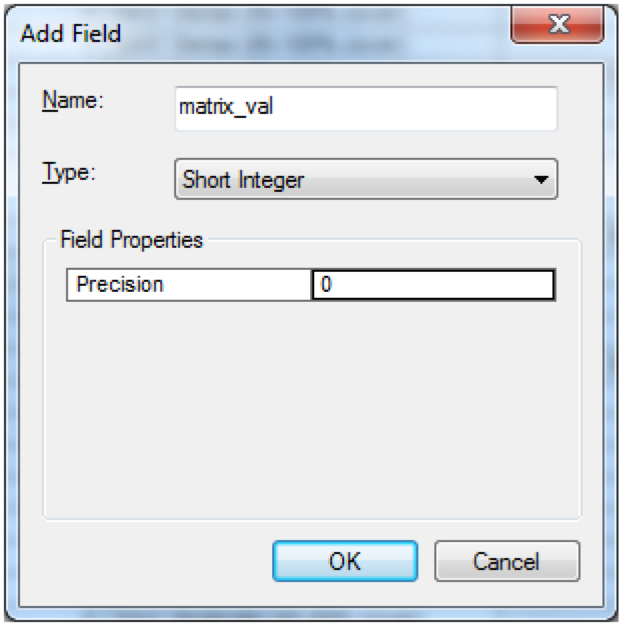
G.4.2.1. Right-click the reference shapefile containing the data in the table of contents and select **Open Attribute Table**.

G.4.2.2. Examine the shapefile and its description. For example this shapefile gives the descriptions of Dense, Moderate, and Sparse. Each of these components will need to be given the appropriate value as shown in the table of G.4.1.

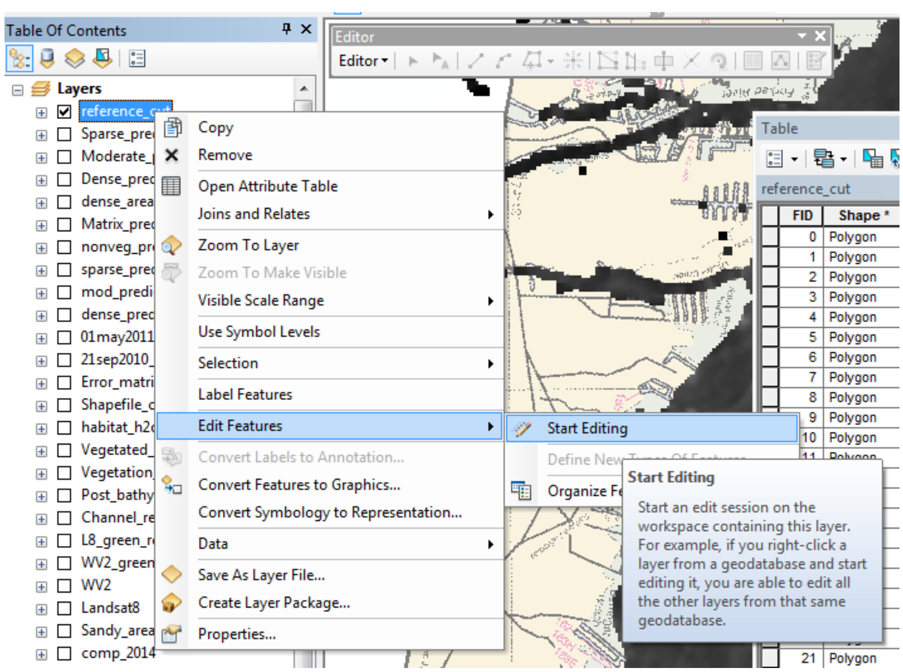
G.4.2.3. Select **table options** and select **Add Field**.



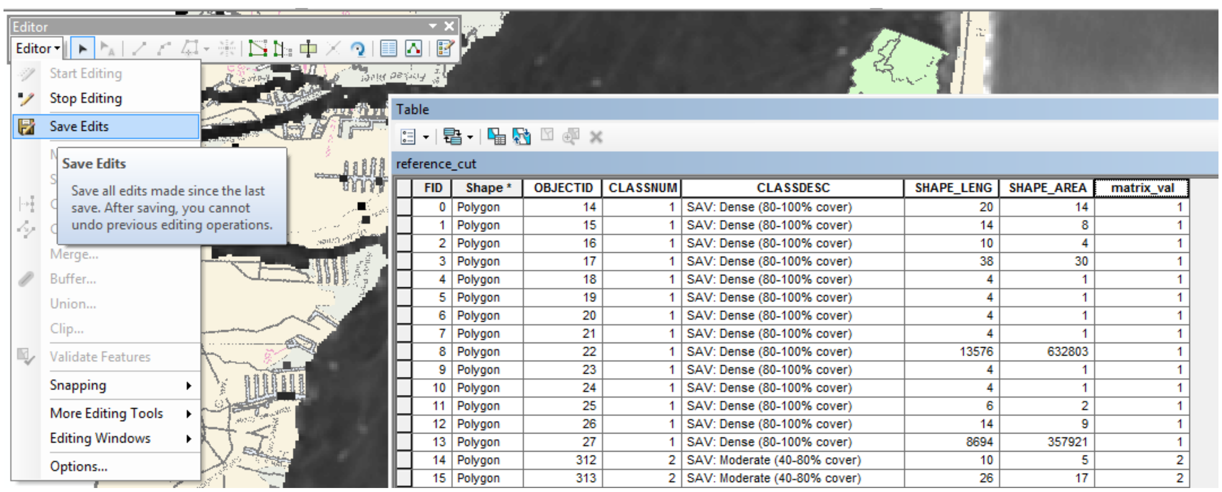
G.4.2.4. Name the field matrix_val. Set the type as short integer.



G.4.2.5. For this example, since the shapefile contains multiple components and these components have different values associated with the confusion matrix the values will need to be input manually. Keep the attribute table open, but go to the table of contents and right-click the reference shapefile and select **Edit Features** then **Start Editing**.

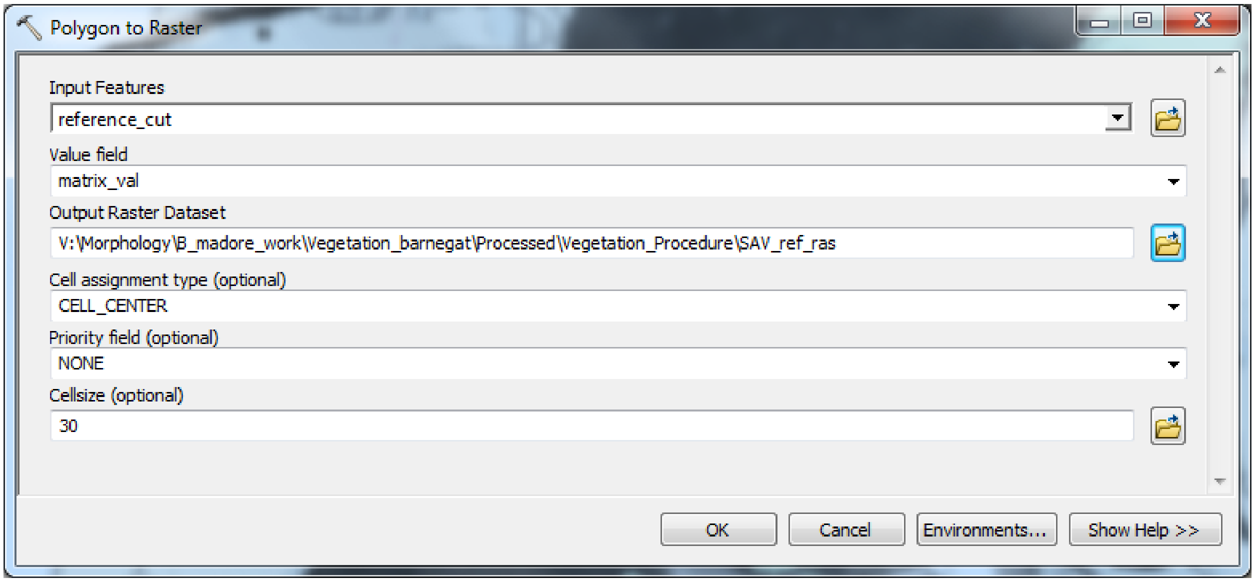


G.4.2.6. The values in the attribute table opened before can now be edited. For the matrix_val field begin to input the associated reference table value based on the information given in the table. Once all the values of have been input, select **editor** in the **editor toolbar** and choose **save edits.** Then select **editor** then **stop editing**. **NOTE:** There are other methods of adding values to the matrix_val component. If the reference shapefile contains so many components that manual input is unrealistic, using tools in the toolbox such as Split may be useful to separate the individual fields. These can then have their values changed separately before being mosaiced together again in then (this would be similar to the processes described in G.3.3.).



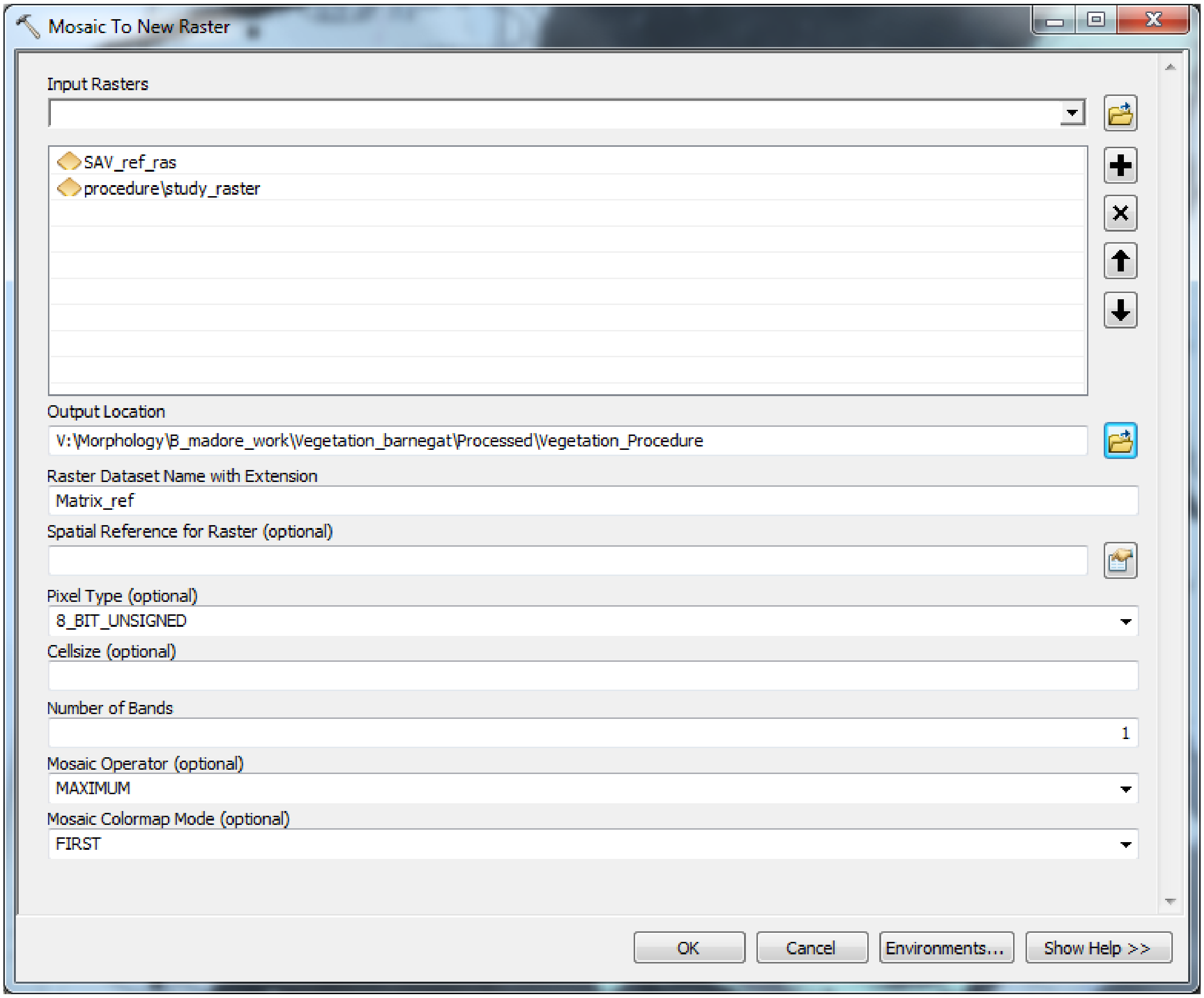
G.4.2.7. In the **ArcToolbox** select **ConversionTools>To Raster>Polygon to Raster** to convert the shapefile into a raster.

G.4.2.8. Input the shapefile from the previous steps. For the **value field** select the field created in G.4.2.4. Select a location and name to save the layer. For the cellsize use an appropriate value based on the datasets being examined. For this example 30 meters is used because that is the resolution of Landsat 8.



G.4.2.9. In the **ArcToolbox** select **Data Management Tools>Raster>Raster Dataset>Mosaic To New Raster** for adding the non_vegetated raster background to the data (see step F.3 for more information on the background raster).

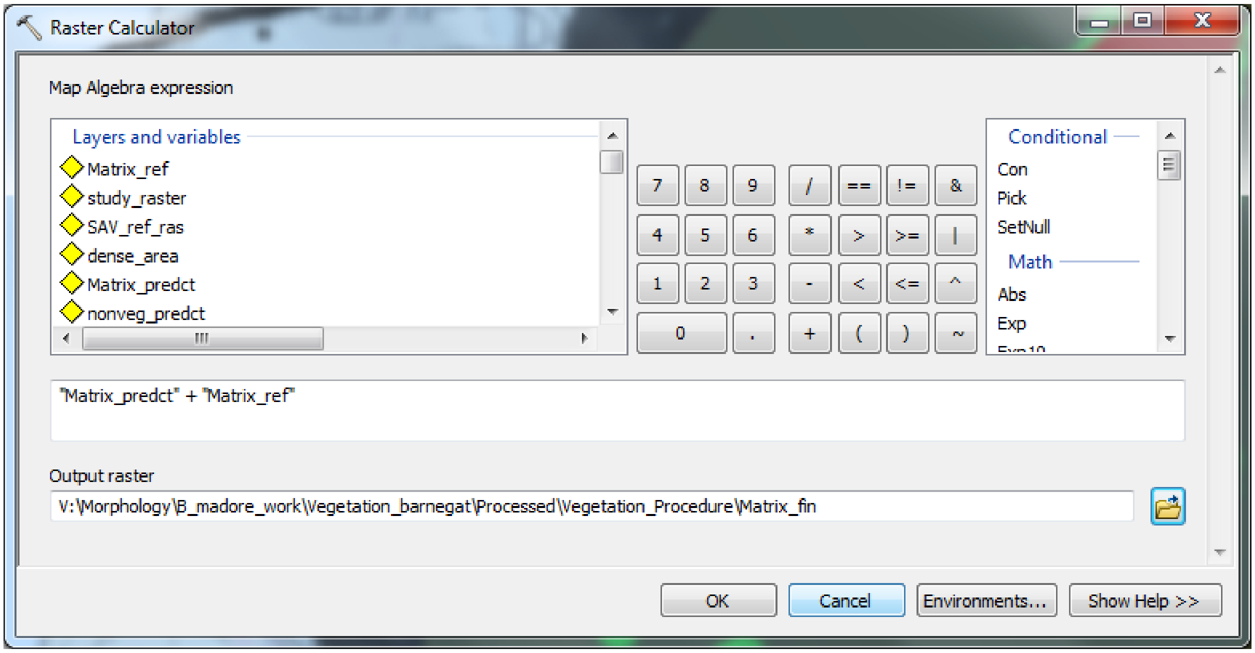
G.4.2.10. Add the reference raster from G.4.2.8. and the study_raster (from F.3). Select the **output location** for which folder the file will be saved in. Set the name of the layer being created under **Raster Dataset name with extension**. Set the **number of bands** as 1. For the **Mosaic Operator** set the value based on how your datasets are set up. For overlapping components of this example a higher density layer will take priority (therefore if a cell contains both dense and moderate as a value the dense layer will take priority)(For this example and process the background raster of non-vegetated areas (that is the study_raster layer) has values of 0 therefore the maximum value will be used for the operator). Other operators can be used, but make sure they go correctly with the input of rasters and how your data was entered.



G.5. Once the predicted layer and reference layers have been created with the multiple components the two datasets can be added together.

G.6. In the **ArcToolbox** select **Spatial Analyst Tools>Map Algebra>Raster Calculator**.

G.7. For the expression select the predicted layer created in G.3 and add to it the reference layer created in G.4.

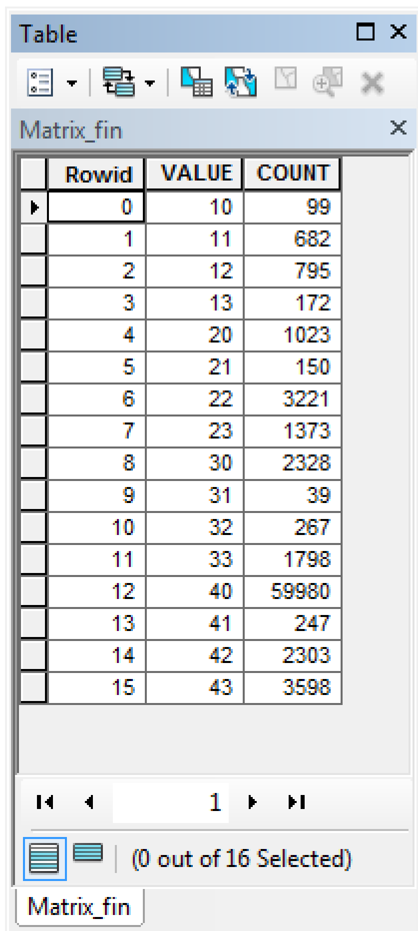


G.8. The resulting raster will now have up to 16 possible unique values based on the 4 possible prediction values and the 4 possible reference values. The table below shows the values and their explanations (for this example).

Numeric Value	Predicted Layer SAV description	Reference Layer SAV description
10	Dense Vegetation	No Vegetation
11	Dense Vegetation	Dense Vegetation
12	Dense Vegetation	Moderate Vegetation
13	Dense Vegetation	Sparse Vegetation
20	Moderate Vegetation	No Vegetation
21	Moderate Vegetation	Dense Vegetation
22	Moderate Vegetation	Moderate Vegetation
23	Moderate Vegetation	Sparse Vegetation
30	Sparse Vegetation	No Vegetation
31	Sparse Vegetation	Dense Vegetation
32	Sparse Vegetation	Moderate Vegetation
33	Sparse Vegetation	Sparse Vegetation
40	No Vegetation	No Vegetation
41	No Vegetation	Dense Vegetation

42	No Vegetation	Moderate Vegetation
43	No Vegetation	Sparse Vegetation

G.9. Opening the **Attribute Table** of the final raster will provide the number of cells (which can then be calculated into area) for each of the unique values. Calculating the accuracy can be accomplished by adding all of the matching cell values (that is, when predicted = reference, for example 11 is when predicted has dense vegetation and the reference layer has dense vegetation) and then dividing by the total number of cells. More information on confusion matrix and what can be calculated from the results can be found online.



Rowid	VALUE	COUNT
0	10	99
1	11	682
2	12	795
3	13	172
4	20	1023
5	21	150
6	22	3221
7	23	1373
8	30	2328
9	31	39
10	32	267
11	33	1798
12	40	59980
13	41	247
14	42	2303
15	43	3598

The table below uses values from the attribute table located above with the description of the values shown in step G.8.

Example of 4X4 vegetation confusion matrix	Predicted (Dense)	Predicted (Moderate)	Predicted (Sparse)	Predicted (No-SAV)	Total
Actual (Dense)	682	150	39	247	1118
Actual (Mod)	795	3221	267	2303	6586
Actual (Sparse)	172	1373	1798	3598	6941
Actual (No-SAV)	99	1023	2328	59980	63430
Total	1748	5767	4432	66128	78075

For the accuracy of the above table the gold shaded regions would be added together and then divided by the absolute total, represented in the green shaded cell. This value would be .841 (for the hypothetical example).

G.10.