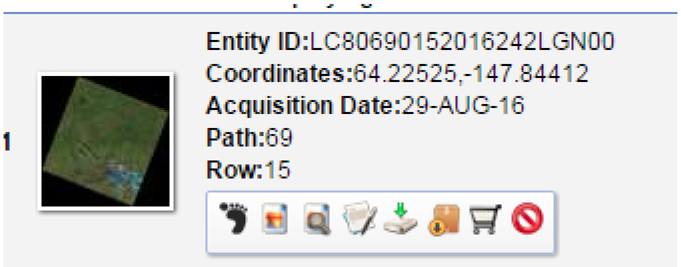
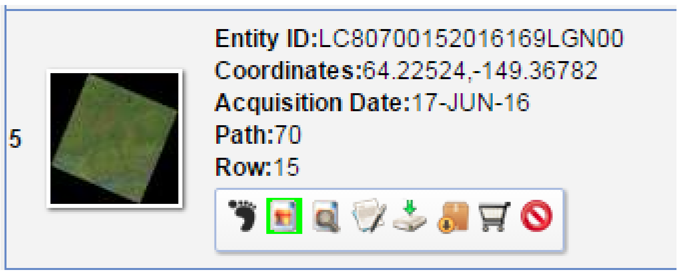
Lab#10: Image Classification

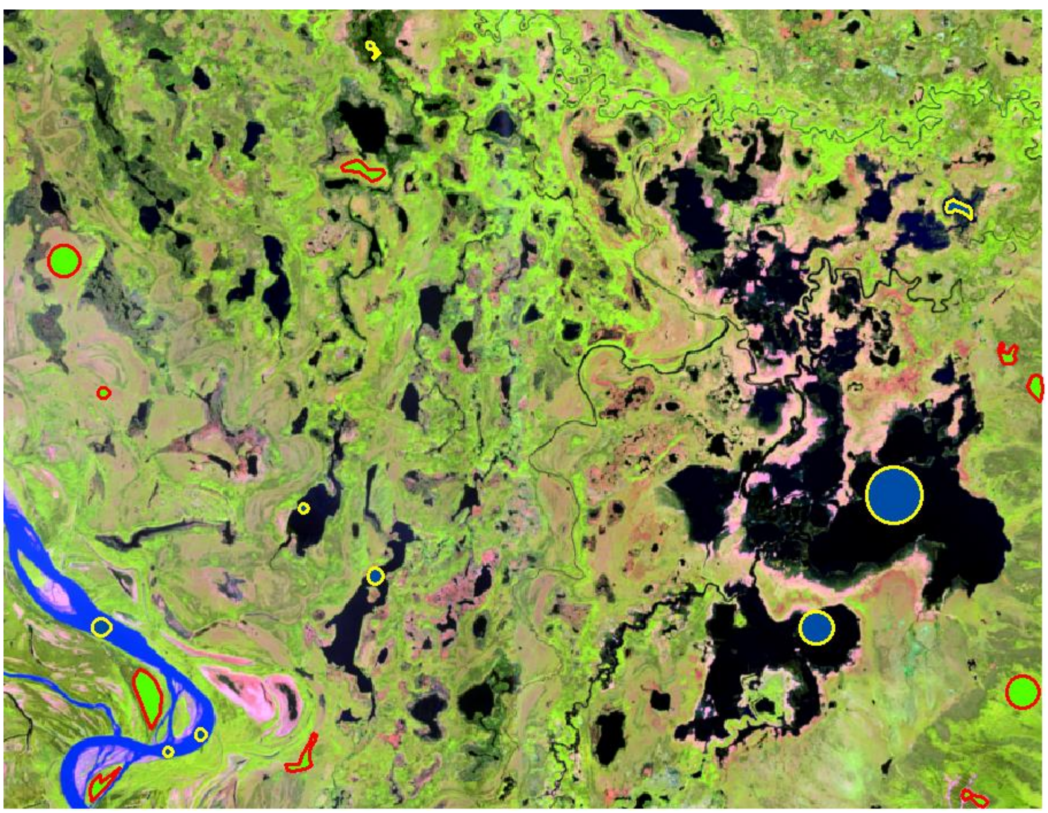
In this lab, you will develop a map of flooded area You will use a part of a Landsat-8 scene from 29-August-2016 and 17-June-2016





You will use ground-truth location points to estimate the classification accuracy of your map. You will do the following:

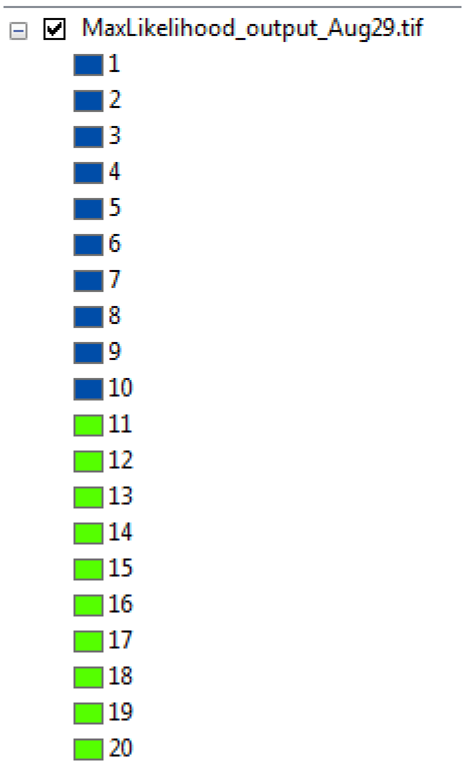
- Edit a polygon theme and assign coded domain values of water/land



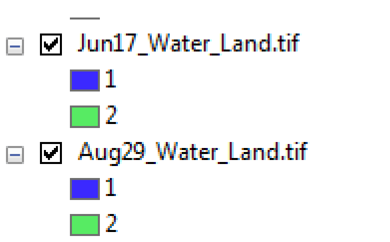
- Extract spectral statistics from image pixels inside each polygon

```
#  Class ID      Number of Cells     Class Name
        10                134         10
# Layers              1              2              3
# Means
            6.632209e+003 5.776955e+003 5.195858e+003
#  Class ID      Number of Cells     Class Name
        11                 50         11
# Layers              1              2              3
# Means
            7.466820e+003 1.303706e+004 1.165364e+004
```

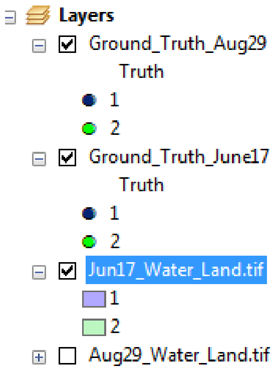
- For each pixel, determine which training polygon to pixel is most similar to

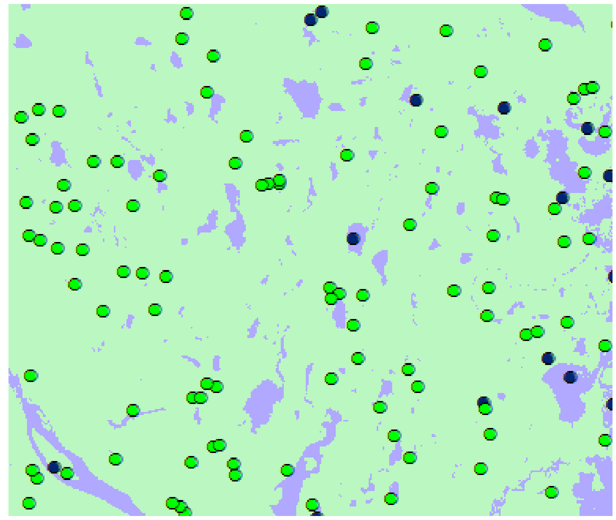


- Reclassify raster to 1=water 0=land



- Compare your image classification with ground truth points to estimate your classification accuracy





Error_Matrix_June17

	Predict	Truth1	Truth2
	1	13	0
	2	5	82

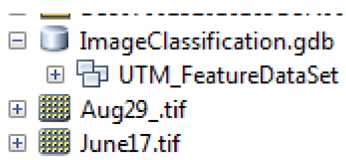
Error_Matrix_Aug29

	Predict	Truth1	Truth2
	1	15	0
	2	4	81

Step 1) Download and unzip data file. You can download the data file ***Lab10_Image_Classification.zip*** from the data folder of the nrm338 website:

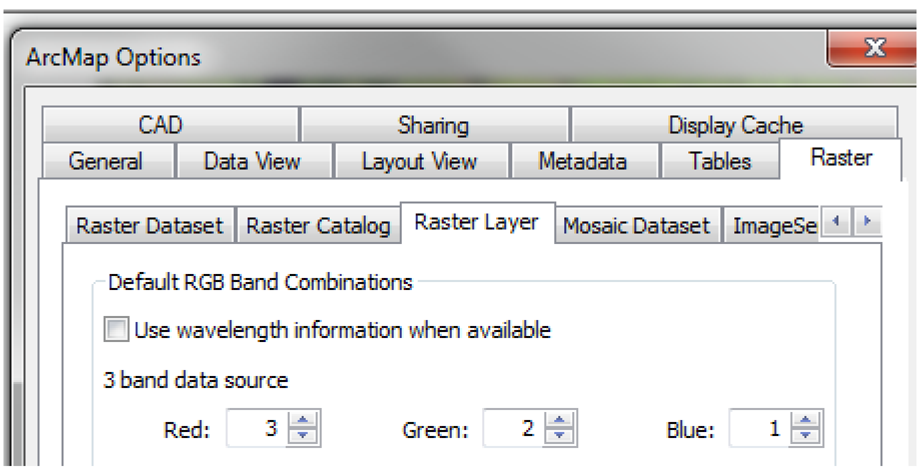
http://dverbyla.net/nrm338/data

Download and unzip your data to your own folder…there will be a file geodatabase, and 2 tif raster images.

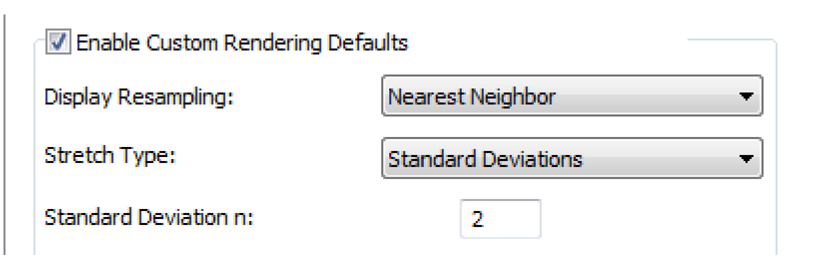


The 2 tif images have been clipped to a study area in the Minto Flats State Game Refuge, northwest of Fairbanks.

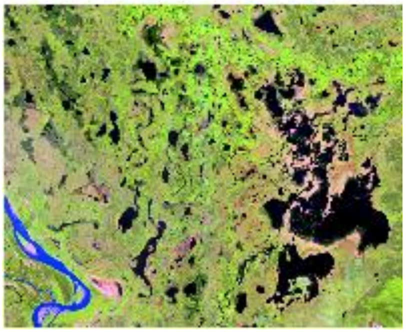
Each tif image has 3 spectral bands. Before adding these rasters to your data frame, set up a default display with the first value controlling the red video intensity, the second pixel value controlling the green video intensity, and the third pixel value controlling the blue video intensity. (Customize Menu-→ArcMap Options)

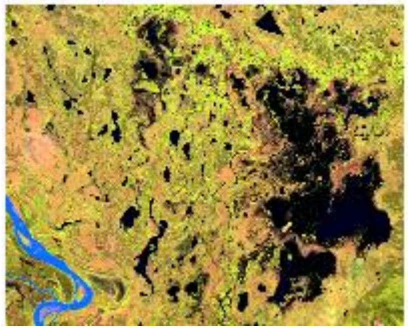


And set your contrast stretch to the mean pixel value +/- 2 standard deviations:

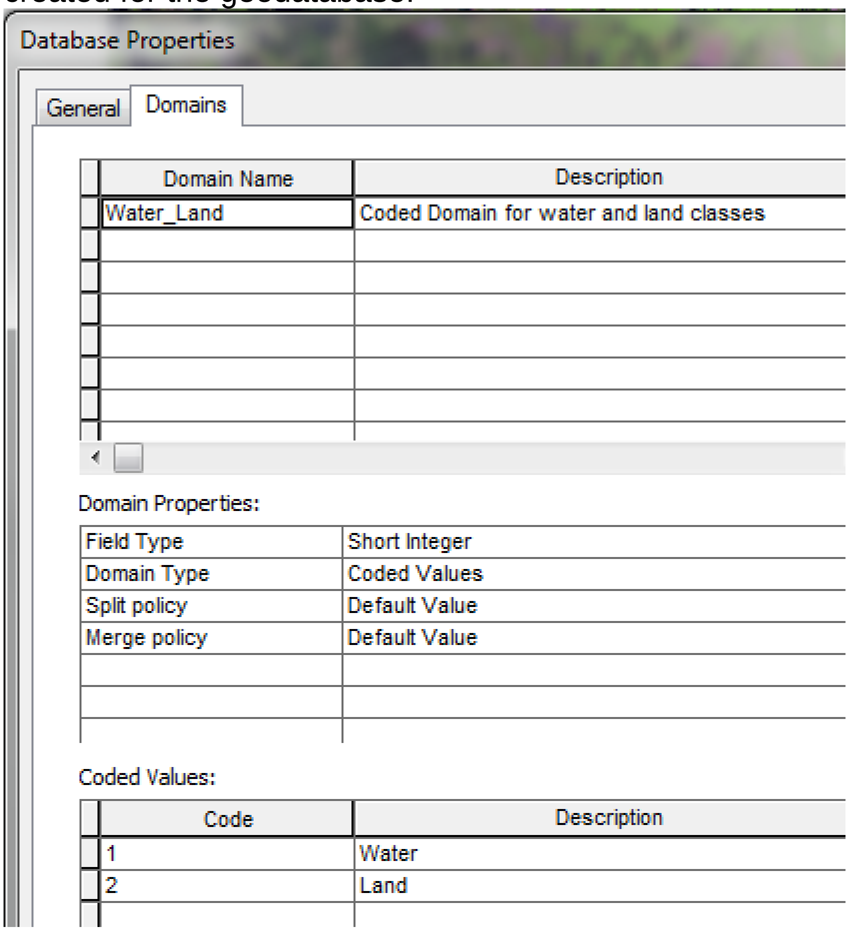


Then add your 2 tif images to your data frame:

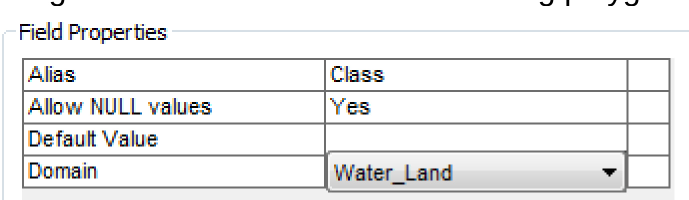




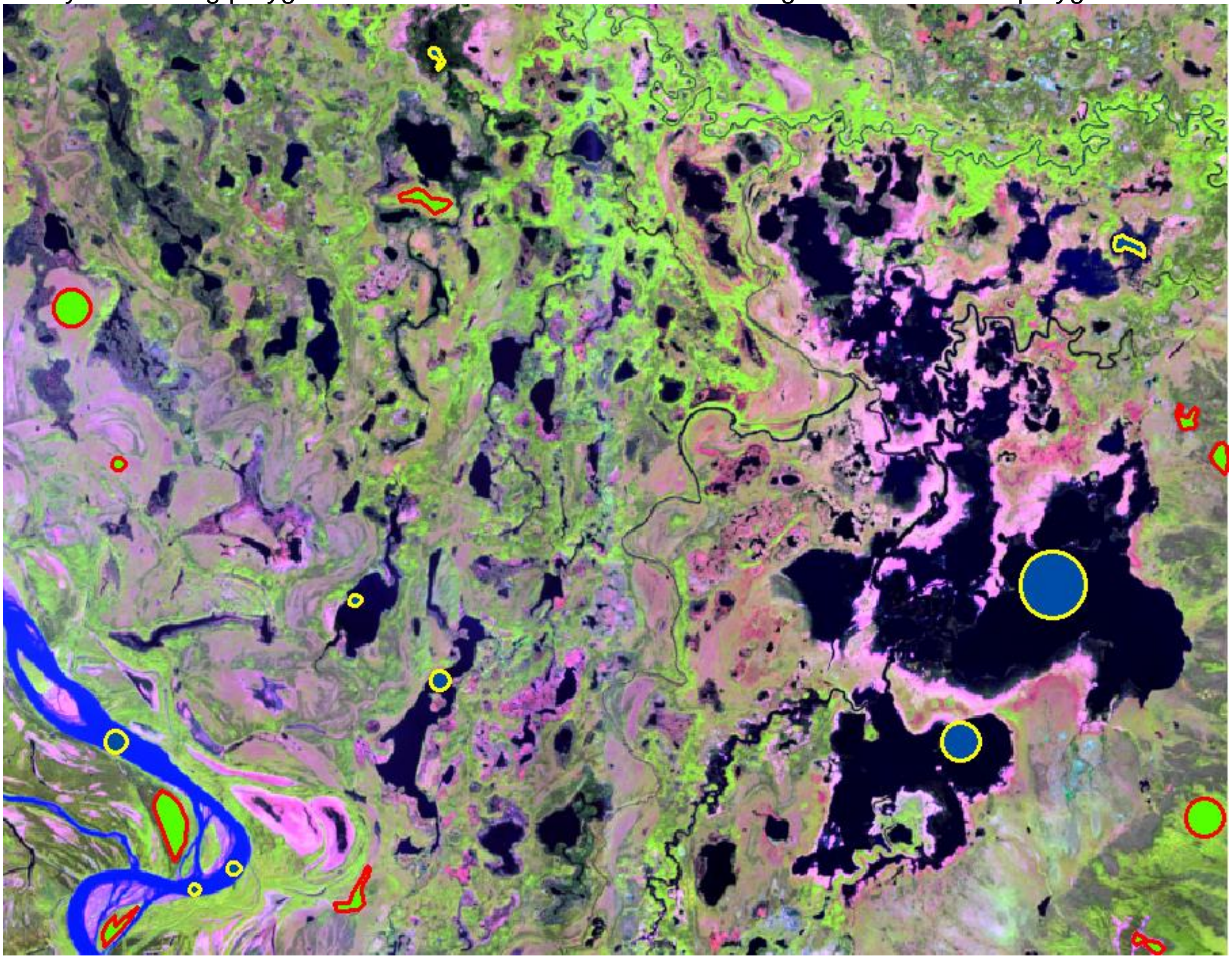
Step 2) Assign water/land classes to your polygons. A short integer coded domain was created for the geodatabase:



Assign this coded domain in the training polygons class field…

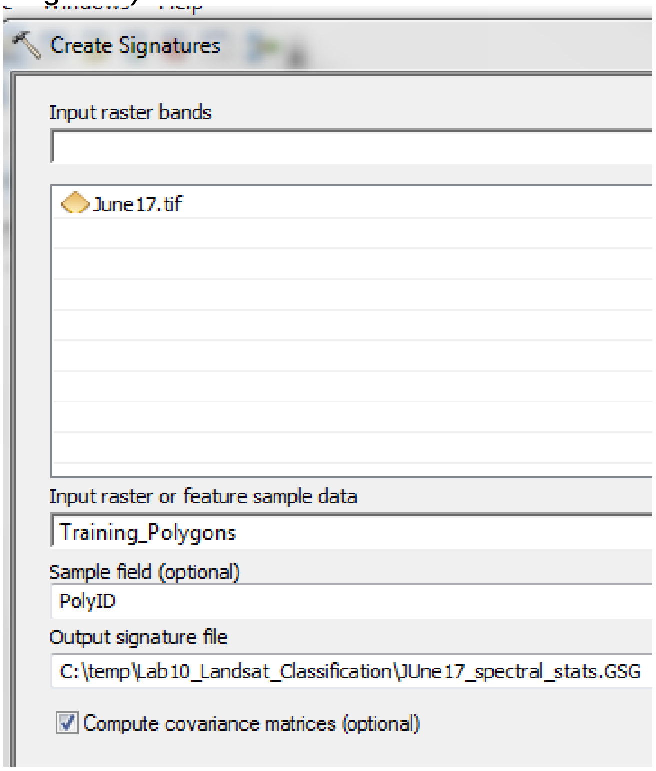


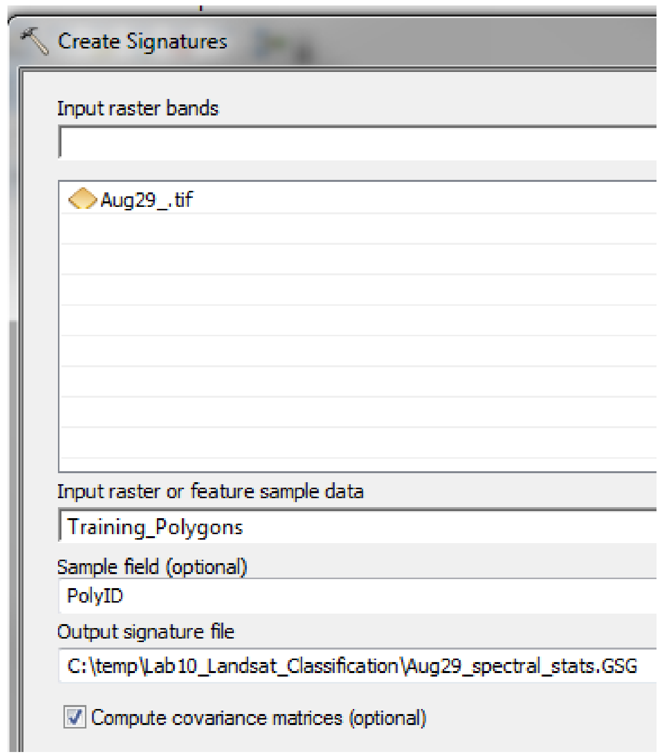
Add your training polygons to your data frame.
Edit your training polygons and use this coded domain to assign a class to each polygon.



You should have ten water and ten land polygons.

Step 3) Extract spectral values from inside each training polygon. Use the ***Create Signatures*** geoprocessing tool to extract the image values from inside each polygon by ***using the PolyID field*** as the field to sample by….we want the mean and covariance for the pixels inside each polygon. Create spectral statistics (signatures) for each raster (June17.tif and Aug29.tif)





The tool requires the Spatial Analyst extension…

Your output .gsg file is a text file formatted the maximum likelihood tool can understand it. You can open this text file using either notepad or worpad:

spectralstats_june17.gsg - Notepad

File Edit Format View Help

```
# Signatures Produced by ClassSig from
#    Class-Grid __1000001
#    and Stack __1000000

#    Number of selected grids
/*             3
#    Layer-Number     Band-name
/*             1      June17.tif\Band_1
/*             2      June17.tif\Band_2
/*             3      June17.tif\Band_3

#  Type   Number of Classes   Number of Layers    Number of Parametric Layers
    1             20                 3                         3
# ===================================================================

#  Class ID     Number of Cells    Class Name
       1             1415           1
# Layers             1             2             3
# Means
           6.204606e+003 5.622343e+003 5.144924e+003
# Covariance
    1      2.616492e+002 8.486609e+001 4.432023e+001
    2      8.486609e+001 1.479046e+002 4.211651e+001
    3      4.432023e+001 4.211651e+001 6.704590e+001
# -------------------------------------------------------------------

#  Class ID     Number of Cells    Class Name
       2              480           2
# Layers             1             2             3
# Means
           6.202923e+003 5.676242e+003 5.182779e+003
# Covariance
    1      1.180347e+003 7.196071e+003 5.096025e+003
    2      7.196071e+003 5.516918e+004 3.785958e+004
    3      5.096025e+003 3.785958e+004 2.655340e+004
# -------------------------------------------------------------------

#  Class ID     Number of Cells    Class Name
       3               42           3
```

Use a text editor like wordpad to examine your spectral statistics file…there should be 20 classes from your 20 polygons and the statistics are from your 3 spectral bands.

Notice that the classes from water polygons typically had different means compared to land polygons. For example:

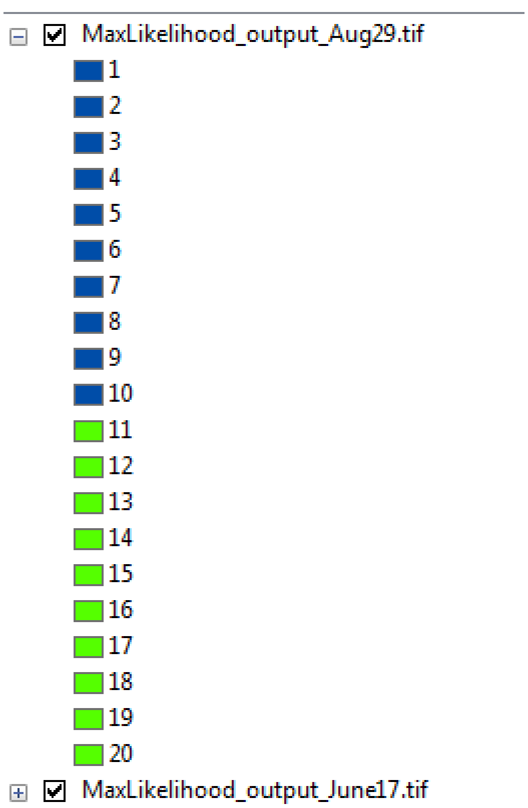
	Polygon	9	Water
	Polygon	10	Water
	Polygon	11	Land

Water polygonID 10:

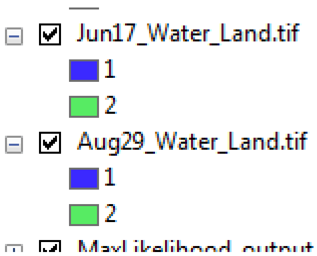
```
#  Class ID     Number of Cells    Class Name
       10                 134        10
# Layers              1             2              3
# Means
            6.632209e+003 5.776955e+003 5.195858e+003
#  Class ID     Number of Cells    Class Name
       11                  50        11
# Layers              1             2              3
# Means
            7.466820e+003 1.303706e+004 1.165364e+004
```

Land had higher mean spectral values in bands 1,2,3

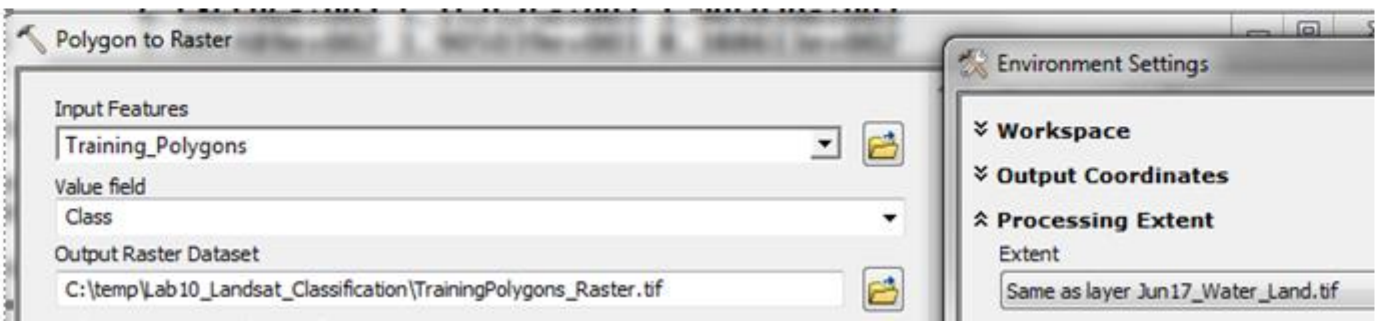
Use the ***Maximum Likelihood*** geoprocessing tool to predict every pixel in the image to belong to one of the twenty training polygons…

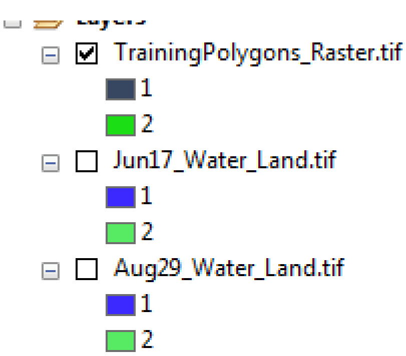


Step 5) Reclassify your 20 class raster into water/land. Your output from the maximum likelihood classification is a prediction of what training polygon each pixel is most likely to be a member of. The training polygons with ids 1 through 10 were from water areas, while the polygons with IDs of 11 through 20 were from land areas. So use the ***reclassify tool*** to create a new raster where **1** represents an **water**pixel, **2** represents a **land** pixel.



Did you correctly classify all pixels in training polygons Convert your training polygons to a tif raster, setting your **output extent,** ***cell size*** and ***snap raster*** to your classified raster. Click on the Environments button when you start the Polygon to Raster geoprocessing tool.





And use the Combine geoprocessing tool to see how well the pixels were classified within the training polygons.

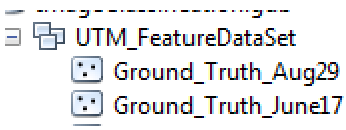
June17_combine_out.tif

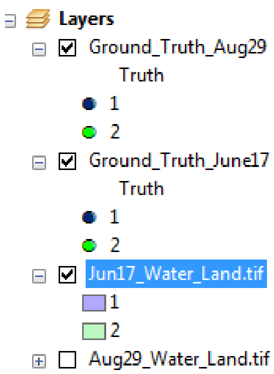
TrainingPo	Jun17_Wate	Count
1	1	2579
2	2	2432
1	2	7

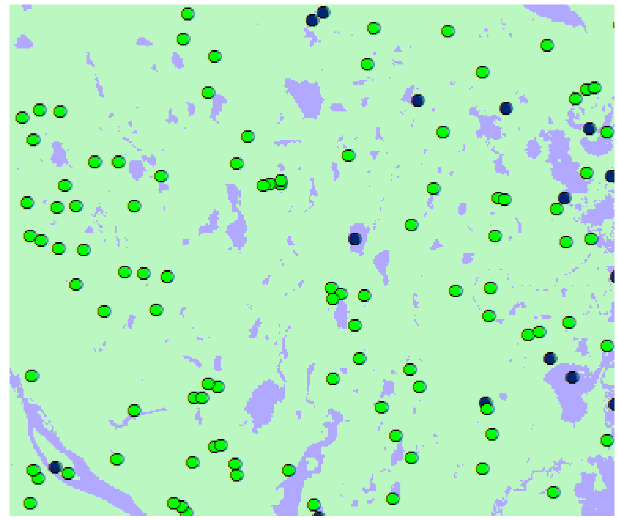
TrainingPo	Aug29_Wate	Count
1	1	2582
2	2	2432
1	2	4

So only 7 of 5018 pixels in the training polygons was incorrectly classified for June 17.
And only 4 of 5018 pixels in the training polygons was incorrectly classified for Aug 29.

Step 6) Compare classification with ground truth points. Add the Ground Truth Points feature class to your Arcmap frame.







For each validation point, we need to know the classified prediction, use the ***Extract Values to Points*** geoprocessing tool.

Make a new field name Predict and replace the RASTERVALU field with your Predict field.

SHAPE *	Truth	Predict
Point	1	1
Point	1	1
Point	1	1
Point	1	1
Point	1	1
Point	1	1

Use the ***Frequency*** geoprocessing tool to summarize by Truth and Prediction fields:

Frequency_Aug29

Truth	Predict	FREQUENCY
1	1	15
1	2	4
2	2	81

Predict_Truth_June17 | Predict_Truth_Au

Frequency_June17

Truth	Predict	FREQUENCY
1	1	13
1	2	5
2	2	82

Use the ***Pivot Table*** geoprocessing tool to create an error matrix with groun truth as columns, predictions (pixel class value) as rows.

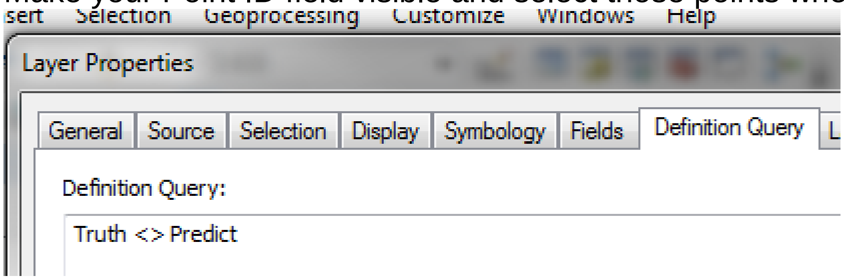
Error_Matrix_June17

Predict	Truth1	Truth2
1	13	0
2	5	82

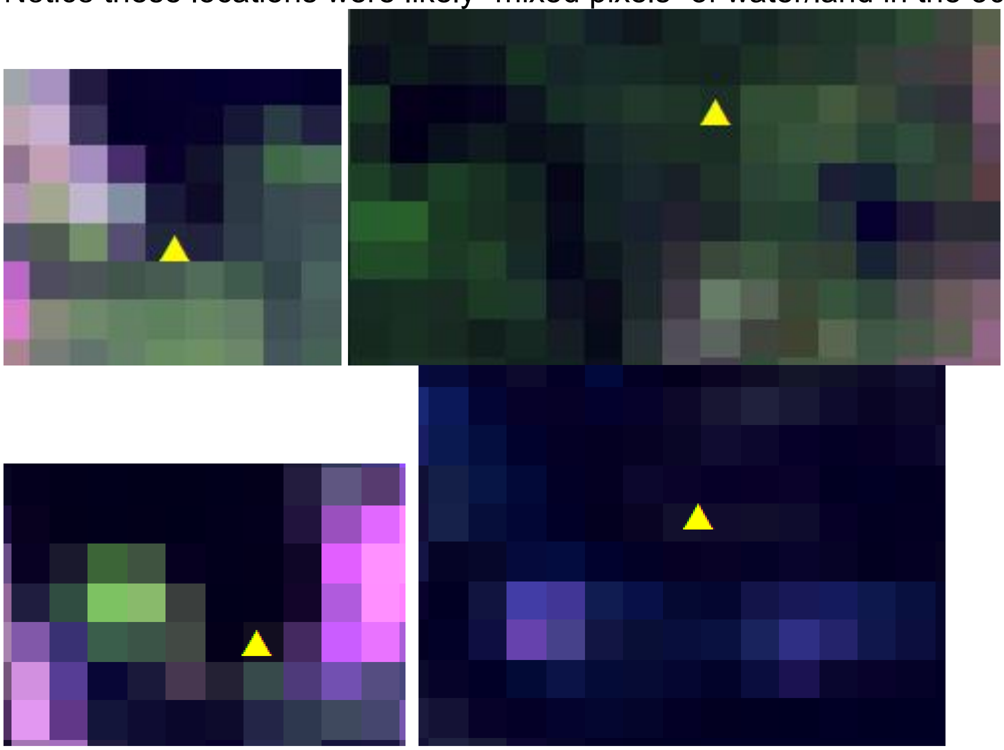
Error_Matrix_Aug29

Predict	Truth1	Truth2
1	15	0
2	4	81

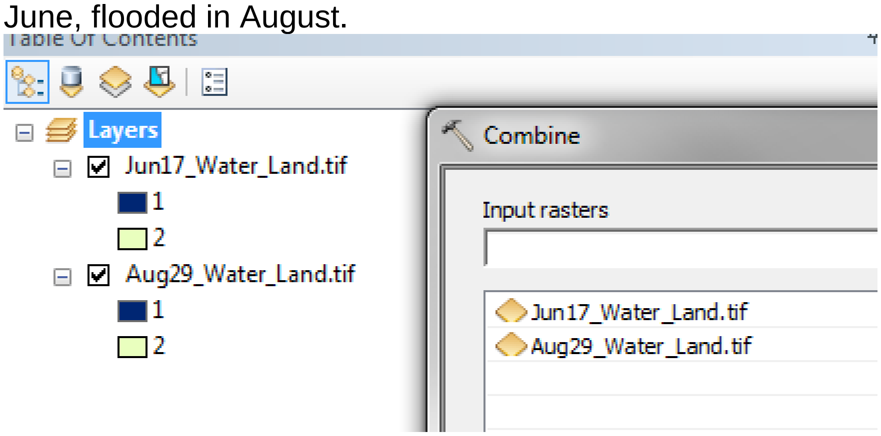
What are the point IDs of the validation points that were not correctly predicted?
Make your Point ID field visible and select those points where Truth <> Prediction



Notice these locations were likely "mixed pixels" of water/land in the 30mX30m pixel area:



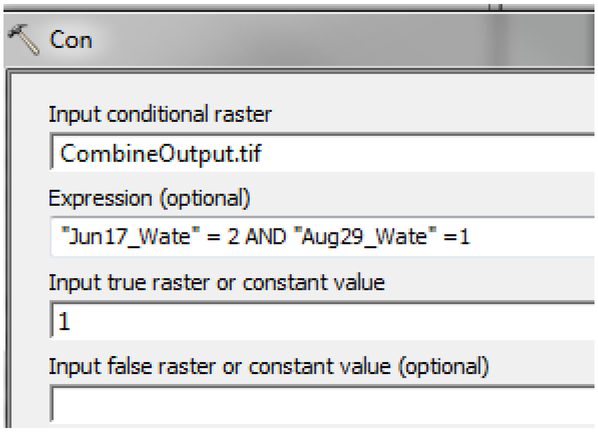
Assuming your classification accuracy is adequate, create a polygon of areas that were land in June, flooded in August.



CombineOutput.tif

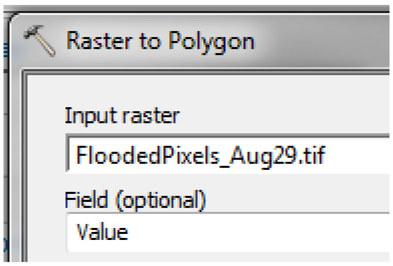
Value	Count	Jun17_Wate	Aug29_Wate
1	435440	2	2
2	4095	1	2
3	54610	1	1
4	21210	2	1

So there were 21,210 pixels that were classified as land on June17, water on Aug 29



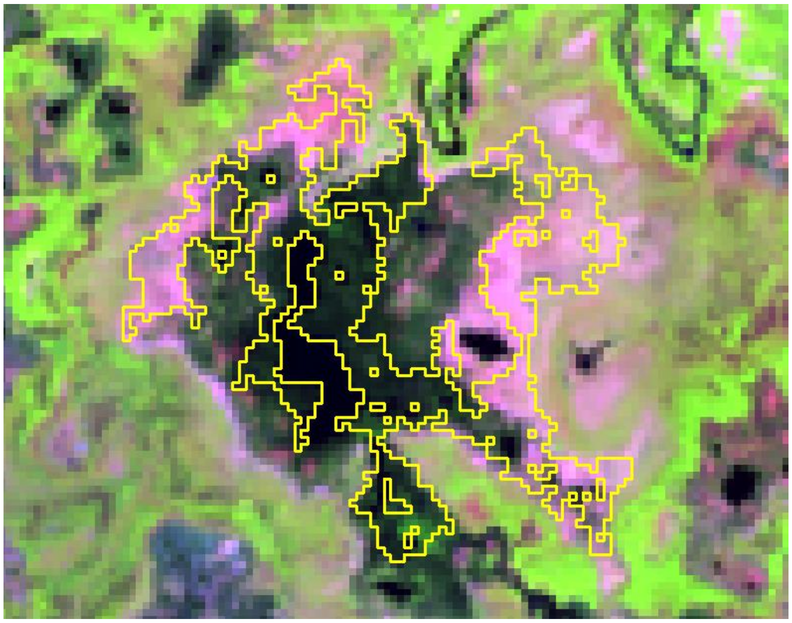
FloodedPixels_Aug29.tif

Value	Count
1	21210



Display all flood polygon that had an area above 10 hectares

June 17



Aug 29