Image Processing Lab Exercise #6:

Dr Bagher Bayat, RS/GIS Group, Faculty of Earth Sciences, SBU

Classification I: Unsupervised Classification

Objectives:

- **Perform an unsupervised image classification for land cover mapping via ISODATA and K-Mean algorithms**
- **Perform post-classification statistical analysis and accuracy assessments via class statistics**
- **Begin to determine the ideal number of classes for a supervised classification**

About Images Used in the Lab

*******TO FIND OUT ADDITIONAL INFORMATION ABOUT AN IMAGE, RIGHT CLICK ON IT IN THE LAYER MANAGER AND SELECT **VIEW METADATA**. ********

1. *tm-siestakey_17dec1997.img*: this is a Landsat TM image of Siesta Key, located in Florida from December 17, 1997.

More images that will not be used in this lab but can be used for further challenges.

1. *gnv-subset-tm-23_oct_2011*: this is a Landsat TM image of Gainesville, Florida from October 23, 2011.
2. *L8_Oct_24_2013*: This is a Landsat 8 image of Gainesville, Florida from October 24, 2013…look at differences between this and the 2011 image (MORE UPDATED AND NEW SENSOR SO EXPLORE THIS FILE!!! THE BANDS DO NOT CORRESPOND TO TM OR ETM+ BANDS)
3. *L8_Mar_14_2014*: This is a Landsat 8 image of Gainesville, Florida from March 30, 2014

*Also, explore seasonal differences around Gainesville on the two Landsat 8 images between March (looks drier) and October (looks wetter).

Note: To support post-processing of unsupervised classification results in this lab we need Google Earth Pro software program being installed on your laptop. Please install it!

Background:

Unsupervised classification is used to cluster pixels in a dataset based on statistics only, without any user-defined training classes. In this case, the categorization is done initially by the computer, followed by labeling the identified classes by the user.

The unsupervised classification technique requires specification of the number of spectral classes to be defined. This number is obtained in a few different ways. There obviously has to be at least as many spectral classes as land cover classes in the map. Generally it is good to have a lot more spectral classes than land cover classes, as each land cover may require several spectral classes. So think about how many classes you think you will have (urban, forest, suburban, wetlands, grass, water, etc.) and then try multiplying that number by about 2-3 to determine the number of spectral classes. It will be hard to decide on a final legend for the map until after you have seen the results of the unsupervised classification. You may hope to separate classes that prove difficult using supervised classification, and you may end up revising your legend accordingly.

One of the strengths of unsupervised classification is that it will help you understand which classes can be separated in your map, and which cannot.

The user determines the minimum and maximum number of the classes the computer will use to classify the image(s). After the initial classification map is produced by the computer, the interpreter labels the classes, giving them names corresponding to the classification system. Note that each information class can have more than one spectral appearance, therefore more than one spectral class. If this is the case we can lump those classes together, but think about it before doing this. There is obviously some spectral difference, maybe deep water vs. shallow. You must decide whether it will benefit your research to lump these together.

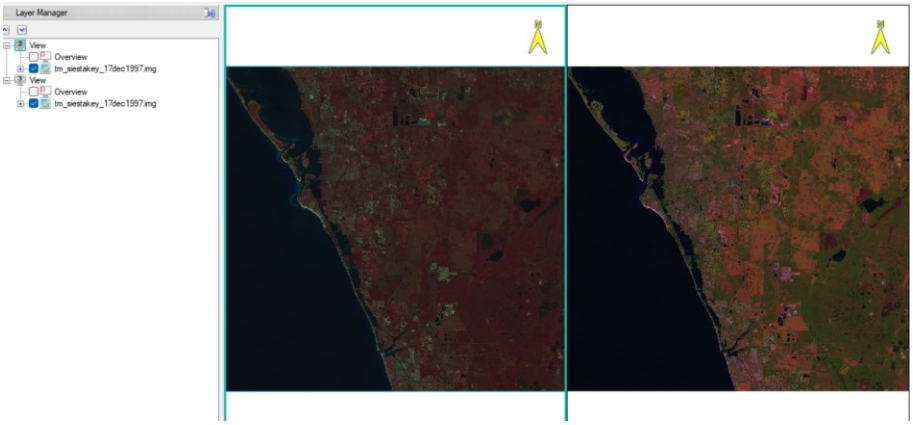
ENVI's simplest classification methods are the Unsupervised ISODATA (Iterative Self-Organizing Data Analysis Technique Algorithm) and K-Means classifiers, often referred to as *clustering routines*.

ISODATA unsupervised classification calculates class means evenly distributed in the data (feature) space then iteratively clusters the remaining pixels using minimum distance techniques as described in the lecture. Each iteration recalculates means and reclassifies pixels with respect to the new means. Iterative class splitting, merging, and deleting are done based on input threshold parameters. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria. This process continues until the number of pixels in each class changes by less than the selected pixel change threshold or the maximum number of iterations is reached.

K-means clustering unsupervised classification asks you to specify a number of output classes. Generally, ISODATA and K-Means classify all pixels in the image, but you can specify a threshold to leave some pixels unclassified. The K-means classifier takes its name from the fact that it will divide the data into a specified number of classes (denoted by K) and take the "correct" spectra by the mean (actually a multivariate centroid) of each class of points. However finding the means will be an iterative process because initially we can't know which pixels really belongs to which classes. To begin, the program starts with a random guess for the "center" of each of the K clusters, remember K is user defined, then it iterates (or corrects) the classification as follows. It assigns each pixel to whichever cluster is closest then once it has assigned all the pixels it recomputes the center of each clusters from the mean of all of the pixels currently in each cluster. Once it has finished it begins a new iteration, reassigning pixels and recomputing means. It keeps track of how many pixels are reassigned to a new cluster during an iteration, and also how much the cluster centers move. If the number reassigned is above some specified threshold or a cluster center moves by more than a specified amount then the iteration repeats. If not, the classification is assumed to have converged. In ENVI only the number-of-pixels test is used to check convergence. ENVI does also offer one final option, related to the center of the clusters.

Part 1: Unsupervised Classification via ISODATA

1. Open the tm-siestakey_17dec1997.img image included in the lab 6 folder and create a 4,3,2 and 5,4,3 composite. Remove any composites other than these and adjust the brightness so that you can see the images clearly.



2. From the ENVI toolbox, select ***Classification → Unsupervised Classification → IsoData Classification.*** The Input File dialog appears.

3. Select the input file, be sure that the Spatial Subset is Full Scene and the Spectral Subset is 7/7 Bands, and that no Mask is selected. Click ***OK***

4. Enter the minimum and maximum ***Number of Classes.*** In this image we are looking at a Landsat TM image of Siesta Key, Fl with covers such as water, forest, residential, pasture, and commercial. For this first ISODATA classification use **5 and 30** for your minimum and maximum values for number of classes.

5. Enter the maximum number of iterations as **20** in the ***Maximum Iterations*** field and a change threshold of **5%** in the ***Change Threshold % (0-100)*** field.

The classification ends when either this threshold is met or the maximum number of iterations is reached.

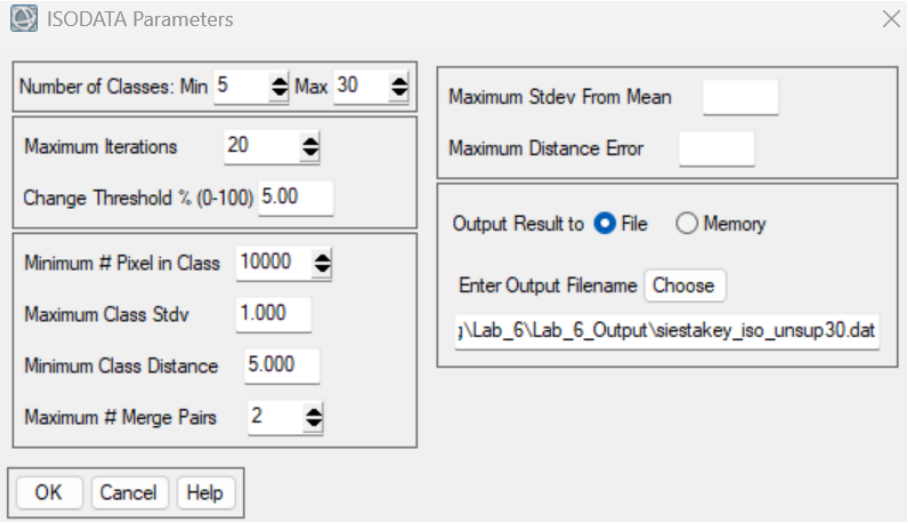
6. Enter the minimum number of pixels needed to form a class as 10,000 in the ***Minimum # Pixels in Class*** field. If there are fewer than the minimum number of pixels in a class than ENVI deletes that class and the pixels are placed in the class(es) nearest to them.

7. Enter the maximum class standard deviation (in DN) as 1 in the ***Maximum Class Stdv*** field. If the standard deviation of a class is larger than this threshold then the class is split into two classes.

8. Enter the minimum class distance (in DN) as 5. If the distance between class means is less than the minimum value entered, then ENVI merges the classes. The maximum number of class pairs to merge is set by the **Maximum # Merge Pairs** parameter. For now leave this as the default 2.

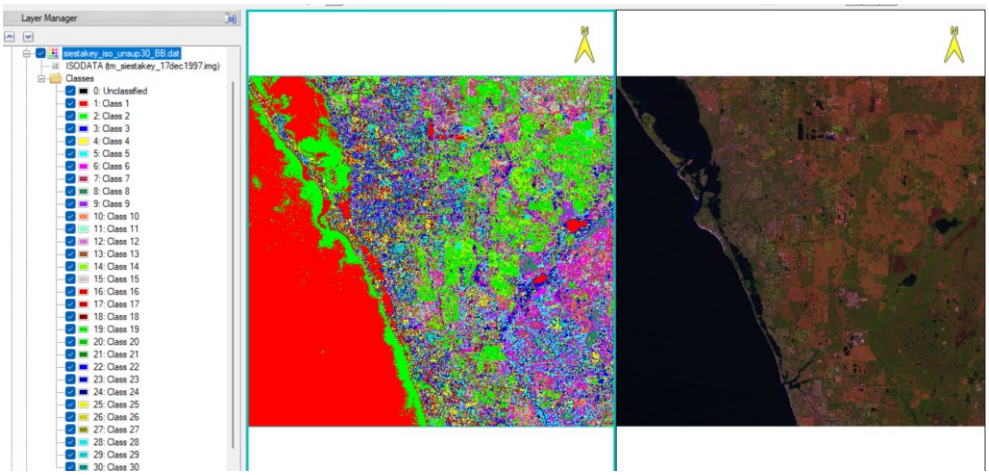
Leave the "Maximum Stdev From Mean" and "Maximum Distance Error" options blank. This way: (i) Every pixel will be assigned to the closest class, and (ii) You can inspect the result, then later test these thresholds to refine your results if needed. So, if you provide values for these options, the classification uses the smaller of the two to determine which pixels to classify. If you do not enter a value for either parameter, then all pixels are classified.

9. Select output to ***File*** and select a file name. Use the **Choose** button to make sure that the classification is being saved to your lab 6 OUTPUT folder. Name the output: siestakey_iso_unsup30.dat



10. Click ***OK.*** If each iteration takes more than a few seconds, it may take a while so please be patient. If it seems to be moving quickly, watch to see how many iterations it takes to stop.

11. Examine the output – these are spectral classes with arbitrary colors assigned to each one. Now you must figure out the information class for each spectral class.



NEXT STEPS IN UNSUPERVISED CLASSIFICATION: NAMING CLASS AND CLUMPING

Changing Class Colors and Names:

Now we need to organize our classes and figure out what each means. In the end we will lump all of our classes into one of five final classes: water, forest, pasture/grassland, residential, or commercial. These classes are described in the table below.

Table 1. Desired classification classes and their appearances in a 5-4-3/R-G-B composite Landsat TM image.

Class number	Name	Color in color composite Image (RGB=543)	Texture
1	Water	Black; Dark blue	Smooth
2	Forest	Green	Moderate
3	Pasture/Grassland	Pink to light green/yellow	Moderate
4	Residential	Bright blue	Rough
5	Commercial/industrial	Blue/Grey	Rough

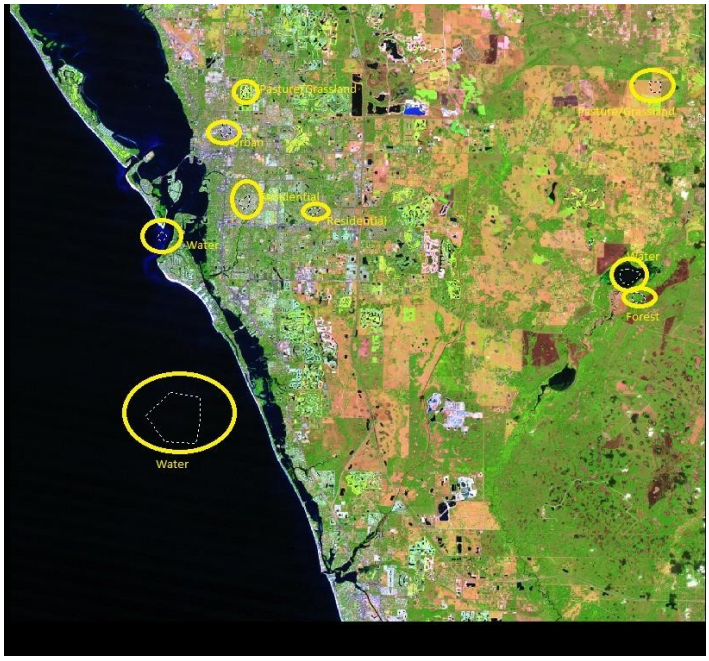


Figure 1. False color composite of Siesta Key, FL (RGB=543). Major land cover types are found within the ellipses.

To do this most simply open a blank Excel spreadsheet and enter the number of classes (1-30) that the classification derived in one column and in the next column you will describe what land cover is being detected. The figure below also lists the classes in the table that we want to in the

final land-cover image. BEWARE that these may not be correct in your judgement, so be sure to check the first several spectral classes ("Class") before you input them to your spreadsheet.

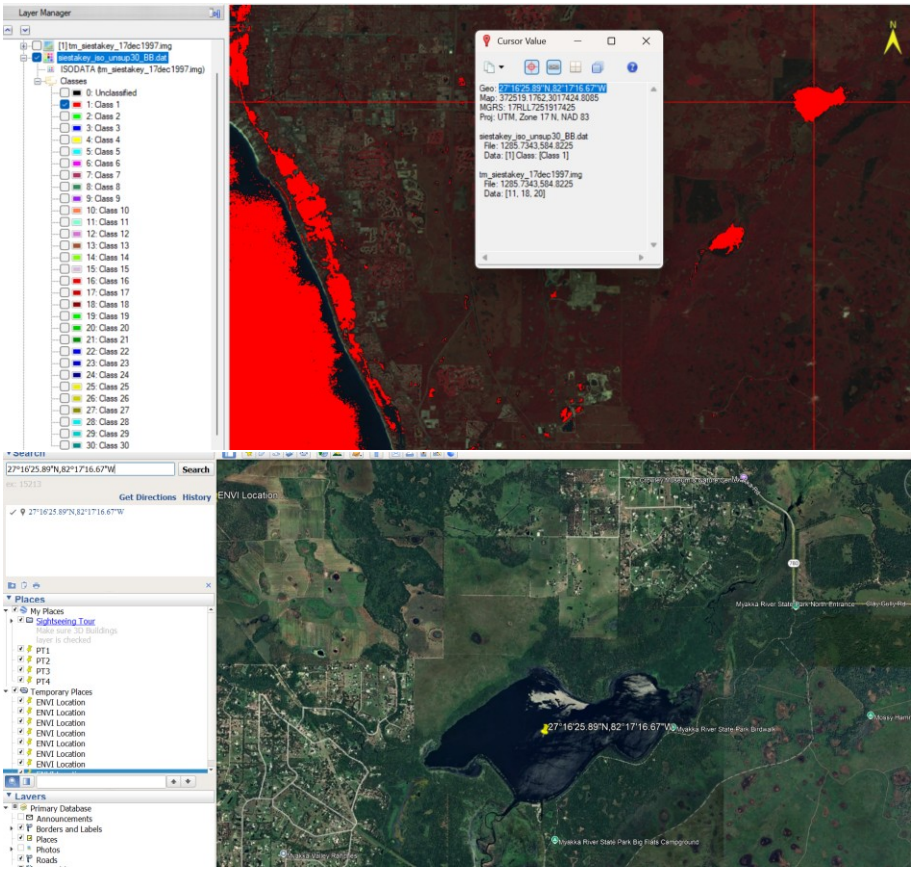
1. Examine each class by using the cursor value function in multiple instances of each class, and describe the land cover that is being detected. This is where the View Flicker or View Swipe functions come in handy. The first few classes are shown below in a spreadsheet. Note that I gave the land cover a detailed name first, then aggregated them into the 5 classes that we want for the final land-cover image. You don't have to do this but it is very useful for re-analysis or even devising a different land cover than was thought appropriate at the beginning.

	A	B	C
1	Class	Detailed Land Cover	Land Cover Information Class
2	1	Open Water - Gulf	Water
3	2	Shallower Water - Gulf	Water
4	3	Shallower Water	Water
5	4	Shallowest Water	Water
6	5	Treed Residential	Residential
7	6	Wetland Forest and Reside	Residential
8	7	Wetland Forest and Reside	Residential
9	8	Wetland Forest	Forest
10	9	Residential	Residential
11	10	Wetland Forest	Forest

Class number	Name	Color in color composite Image (RGB=543)	Texture
1	Water	Black; Dark blue	Smooth
2	Forest	Green	Moderate
3	Pasture/Grassland	Pink to light green/yellow	Moderate
4	Residential	Bright blue	Rough
5	Commercial/industrial	Blue/Grey	Rough

2. You will notice that some classes may be represented by several different covers or vice-versa (e.g. Wetland Forest and Residential – 2 different land covers - in my detailed description became Residential), multiple classes are assigned to the same land cover. For each class decide whether the dominant land cover being represented is: water, forest, pasture/grassland, residential, or commercial. In some cases this will be difficult as commercial and residential may seem to overlap, or pasture and residential may seem to overlap because of large yards. Use your judgement and assign each category to one of the five classes. You will start to understand the "convenient fiction" of land-cover classes as you go through this exercise and you may feel frustrated by the difficulty of assigning a spectral class to a named land-cover class. In your own work, you can change the number of land-cover classes to meet the needs of your study, and this exercise is a good way to decide on the LC classes that you will use.

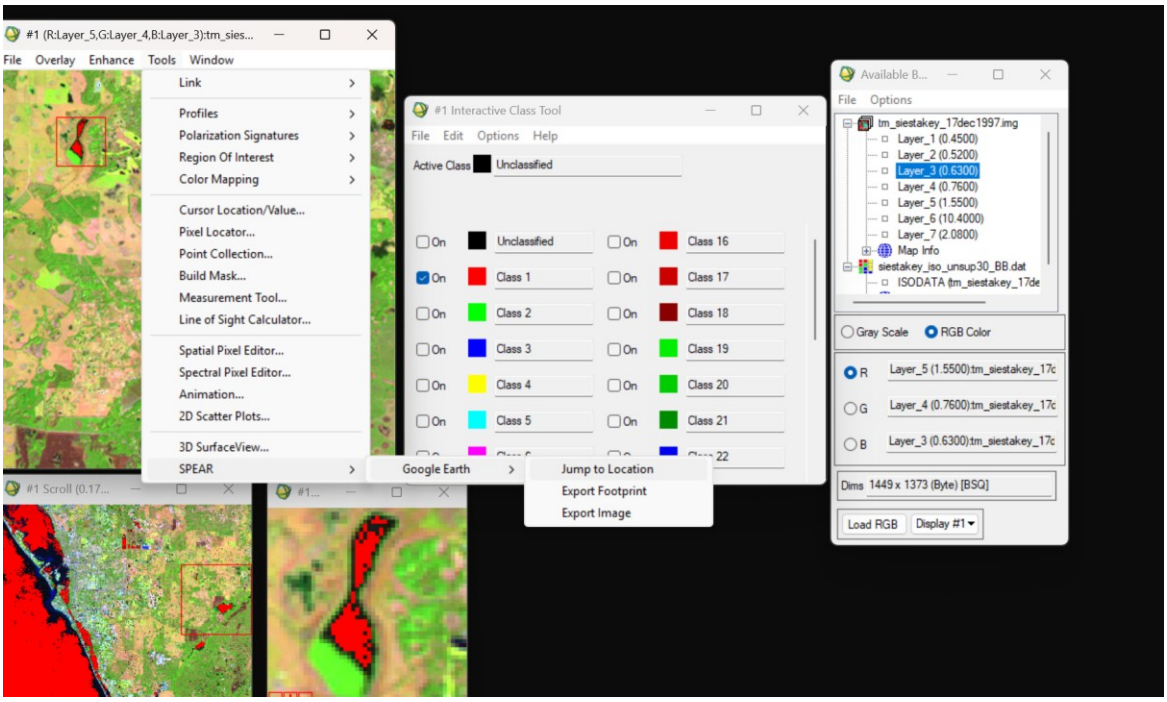
3. As you go through the classes, fill in the excel spread sheet class by class. You may have multiple classes of the same land cover. You will lump these together in ENVI later. Use the image and table above to help you with this, also use the composites already displayed 5,4,3 and 4,3,2. You may also wish to use Google Earth to compare against a finer-scale set of images. Location determined in your ENVI window (cursor value) can be typed into the search window of Google Earth (If you are working with decimal degrees, be sure to make the longitude negative, which denotes West longitude in Google Earth.). See example below:



4.Lastly, note that you can turn the classes on and off (uncheck the boxes) this will help with determine the meaning of each class. For example, in the above example I turned off all the classes except class 1, and overlaid the classification results with the tm image. It helps, right?

5.One useful and interesting feature in ***ENVI 5.6 <u>Classic</u>*** is the ability to quickly jump to a location in Google Earth directly from the image display window. You can access this through:
First keep your Google Earth Pro software program open.
Then:

a. Open ENVI Classic.
b. Load the TM image (tm_siestakey-17dec1997.img) in 543 RGB color composite via File>open.
c. Then, overlay your classified image (siestakey_iso_unsup30.dat) on top of it by Overlay>classification option available in the image window.
   Once the layers are loaded, you can toggle individual classes on or off and use the Jump to Location tool (Tools > SPEAR > Google Earth > Jump to Location) to directly view that area in Google Earth. Note that you need to have already Google Earth Pro software installed on your laptop and be connected to internet for utilizing this feature.
   This is especially helpful for visually comparing your classification results with high-resolution imagery in Google Earth.



After utilizing this feature, you can close the ENVI Classic version and continue again with ENVI 64-bit version.

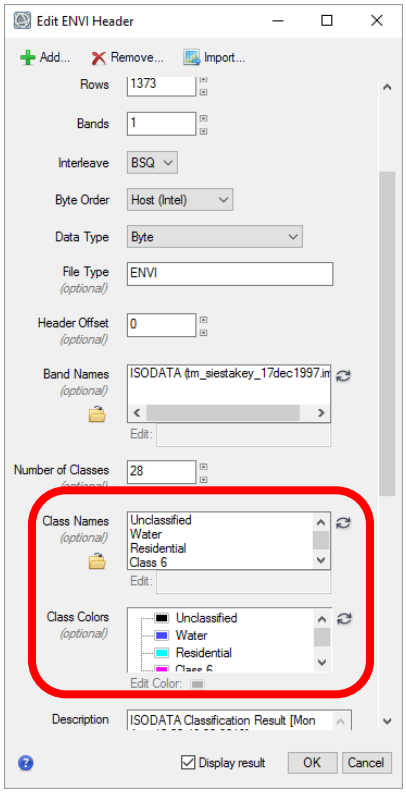
6. Once you have determined the land-cover representation of each of the spectral classes, you will change the labels and colors of the classification raster.

7. In the ENVI toolbox search bar enter **Edit ENVI Header**.

8. Double-click on **Edit ENVI Header.**

9. Choose your classification file as the input.

10. Go to the Class Names window:



11. Go through the class names and rename the FIRST instance of a class to the name of that class.

12. Select the class you want to rename, in the class name edit box enter the new name. If you want to change the colors you can do that at this point as well. It is a good idea to use colors that are evocative of the land cover, e.g. dark green for forest, blue for water, etc.

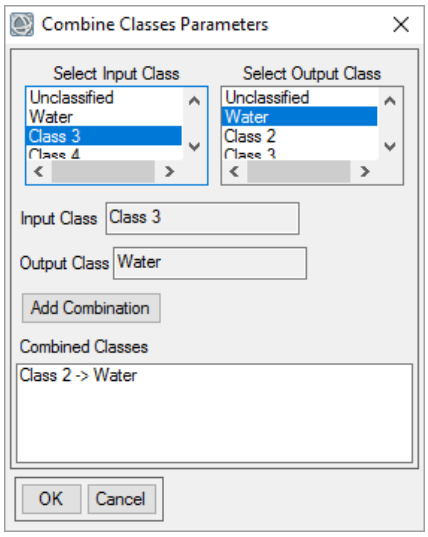
13. When finished click **OK.**

Combining Classes:

Now use the toolbox command ***Classification:Post Classification:Combine Classes*** to combine classes in classified images so that instead of the approximately 30 classes you have, only the five. You may have an additional class that is unclassified. It is fine to leave this for now.

1. Run the ***Classification:Post Classification:Combine Classes*** command in the Toolbox.

2. The Combine Classes Parameters Dialog box appears:

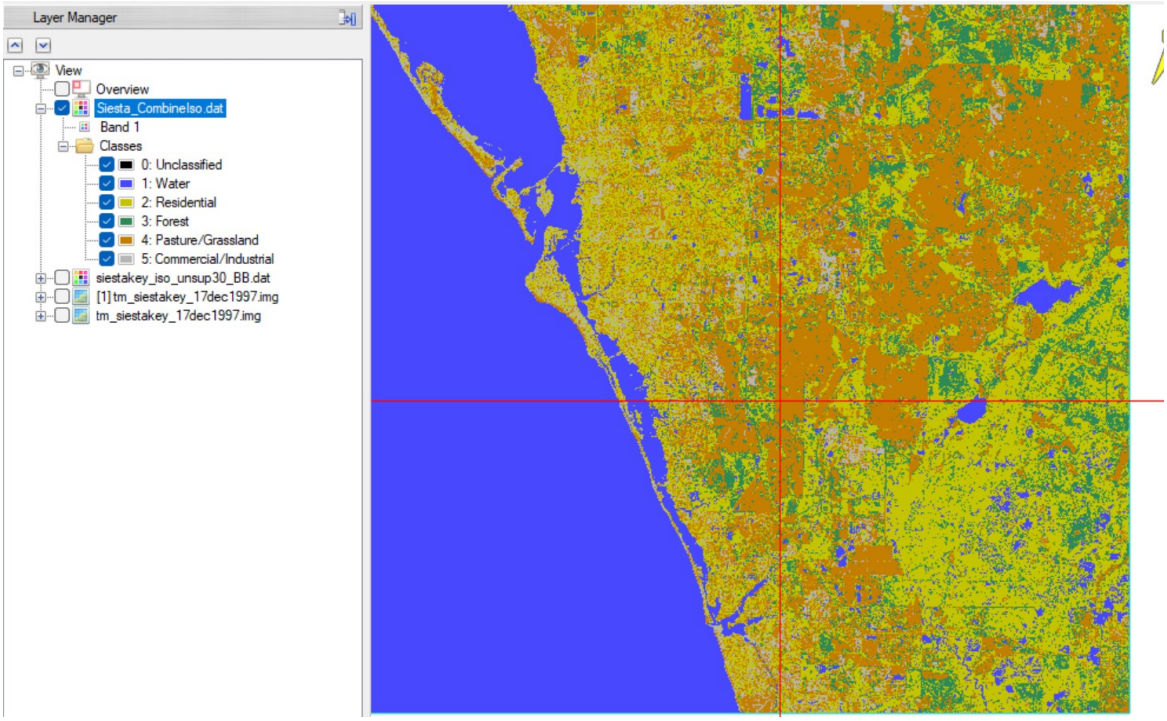


3. You already named some of the classes by using **Edit ENVI Header**.

4. Select the unnamed class by clicking on a class name in the ***Select Input Class*** list, and then the class into this class will be combined in the Select Output Class, then click on **Add Combination**. The example shows that Class 3 is to be combined into the Water class.

Class 2 has already been assigned to Water. As you Add Combinations the list in the Combined Classes window will grow.

5. Continue to Combine all unnamed classes into the named classes until you have the five land covers completely merged. Click OK.
6. Set Remove Empty Classes to Yes. This is very important. Do NOT forget it!
7. Save the merged file with File:Save as…
8. Name the output **Siesta_CombineIso.dat**.



Calculating Class Statistics

Use ***Class Statistics*** to calculate statistics based upon classification results for any input file. Basic statistics include the number of points in a class and minimum, maximum, mean, and standard deviation for each band for that class. You can plot minimum, maximum, mean, and standard deviation plots for each class. Histograms for each class can also be reported and plotted and the covariance matrix, correlation matrix, and eigenvalues and eigenvectors calculated. A summary report for all classes will also be displayed.

1. From the ENVI toolbox, select ***Classification → Post Classification → Class Statistics***. The Input File dialog appears.
2. From the **classification input file** dialog select your **Siesta_CombineIso.dat** file. Click ***OK***.

3. In the **Statistics Input File** dialog select the original landsat tm image of Siesta Key. Do a spectral subset (click on Spectral Subset and select only bands 1,2,3,4,5,and 7. Basically we are getting rid of the thermal band (6).

4. In the Class Selection dialog, select all classes except Unclassified. Click **OK.**

5. In the Compute Selection Dialog, keep Basic Stats checked, also check **Histogram**. Click OK.

6. In the Class Statistics Results dialog that appears, the mean spectra for all classes are shown in the plot window. Below that the Class Distribution Summary lists the number of pixels and percentage of the image assigned to each class.

7. Below the Class Distribution Summary is a statistics table for one of your classes. You can choose which class to show stats for by selecting from the **Stats for** pull-down menu at the top of the dialog. Choose one of the classes, say water or residential.

8. From the **Select Plot** pull-down menu select **All Histograms: All Bands**. These histograms show the values for a particular class in all bands. Change to another class by selecting a different class from the **Select Plot** pull-down menu.

9. To see the histogram for all classes in a specific band click on **Select Plot→All Histograms: Band 1**.

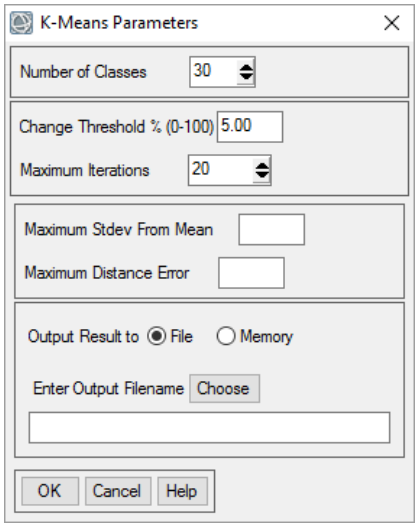
10. Click the small pointer on the right margin of the histogram plot to show the list of histograms by class. Uncheck all but one, then add another to compare the differences among the two. Continue to do this on a pairwise basis.

11. When you are finished examining the statistics, close the Class Statistics Results dialog.

Part II: K-Means Unsupervised Classification

1. From the ENVI toolbar, select **Classification > Unsupervised > K-Means**

2. Select the same Siesta Key TM images file as in part 1 and click **OK**. The K-Means Parameters dialog appears.



3. Enter **30** for the number of classes, and change the maximum iterations to **20**. Select the **File** radio button, and name the file **siestakey_kmeans_unsup30.dat,** click **OK**. This will not take as long as the previous classification, but still will take long enough for you to take a break.

4. After the classification has been calculated then follow a similar procedure as you did above with the ISODATA classification by creating an Excel spreadsheet and determining the dominate land cover from the same five used above (water, forest, pasture/grassland, residential, or commercial), the class colors and names, clumping like classes, and performing class statistics.

Well done on completing the exercise. Congratulations!