Lab assignment #1: Supervised classification

In this lab you will perform a ***supervised classification*** to make a land cover map of one of five study areas: the greater Madison area, Horicon Marsh in southeastern Wisconsin, the vegetation of southern Florida, a mountainous area near Guatemala City, downtown New Orleans, or the forest mosaic of Romania. Because we would like a more synoptic view of a range of land cover types for this assignment, we will focus on using moderate resolution satellite data from the Landsat, SPOT and ASTER sensors. The goal of this lab is to generate a set of training sites, and then use these sites as input to the parallelepiped classification algorithm and the maximum likelihood classification algorithm.

1. Logistics

Date assigned:	January 29, 2016
Date due:	February 12, 2016
Points:	100 points

2. Introduction

In this lab, you will implement a supervised classification using your knowledge of the area, image interpretation skills, and understanding of the spectral classes needed to train the classifier. There are a number of supervised classification algorithms available, many of which have a statistical basis that requires specification of certain parameters in advance. We will test two methods: the *parallelepiped classifier* and a traditional *maximum likelihood classifier*. Note that the parallelepiped algorithm in ENVI is unique in that it chooses the upper and lower bounds of the training data based on *standard deviations from the mean*, rather than the raw min and max values. Remember that *standard deviation* is calculated by summing the distances of each pixel from the mean of the class, and then taking the square root. Statistically speaking, using standard deviation gives you more control over how you set the boundaries of your parallelepiped boxes.

3. Methods for classification

a. **Display** the image.

Open ENVI and display your image (note that you are required to use the study area you chose for lab 6 for this lab) in whatever color scheme you find useful for image interpretation, and which helps you discriminate land cover classes the best. Open Google Earth or other ancillary data sets.

b. **Digitize exemplars**.

Using your unsupervised classification results as a guide, revise your class scheme as needed (i.e. add classes or remove classes, or split classes). To start digitizing ROI polygons you must open the ***ROI Tool*** window; click on the following in the image window:

Overlay > Region of Interest

Inside the dialogue box, the name, color, and number of pixels of each polygon are displayed in spreadsheet format. To edit the name of a polygon, click on the ***ROI Name*** of the polygon (e.g. Region #1), change its name, and make sure to hit return. To draw a new polygon, click on ***New Region***, and repeat the above steps. If you make a mistake, you can ***Delete*** a polygon. When drawing training sites, it is important that you label each ROI *immediately* after you create it.

If you want to add another polygon to an existing region, highlight the region (an * will appear to the left of the name), and digitize additional areas anywhere in the image. Be careful that you have highlighted the region you want to add pixels to, or that you click on ***New Region*** before you begin digitizing new polygons. It is easy to merge polygons later, but quite a bit more difficult to split polygons apart within a given ROI. Save your ROIs to a file with a name such as "class_florida_original.roi".

c. **Evaluate** the training sites (or jump to steps d and e and return to this step later).

To help you understand why some classes might get confused, you should plot the *min*, *max* and *means* of your regions of interest (the merged file) to see which classes have similar patterns as a function of the different spectral bands. This will also help you understand the variance within each class (each ROI). Remember, some variance is okay, but a large amount of variance (i.e. a large number of pixels that are far from the class mean) can lead to over-classification.

On the main menu bar, go to

Basic Tools > Statistics > Compute Statistics

Select the image, and than click on ***Stats Subset***. A new window will appear, where you will see options for ***Calculate Stats*** *On.* Select ***ROI/EVF***.

Unfortunately, you will only be able to calculate stats for one region at a time. Select one ROI and hit ***OK***. Click ***OK*** in the original ***Compute Statistics*** window. Finally, check the boxes next to ***Basic Stats***, ***Histograms***, and ***Output to the Screen***, and hit ***OK***.

Repeat this process for each of your ROIs. In each ***Statistics Results*** window, click on

File > Save results to text file…

Save the file for each ROI. You will need to condense the information in these text files into one table, and include it in your final report.

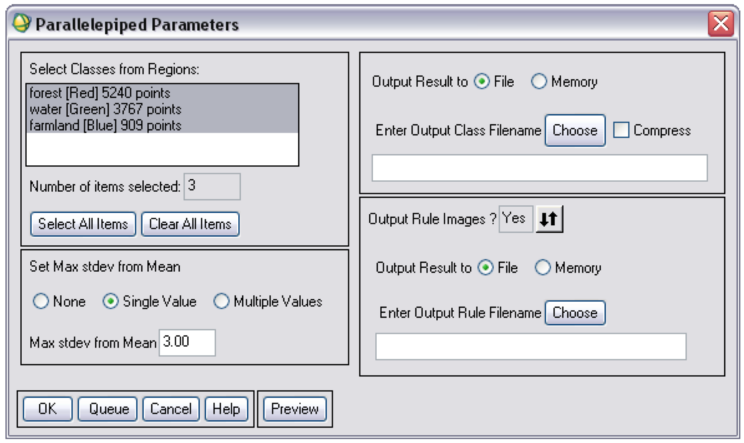
d. Run the **parallelepiped algorithm**.

The parallelepiped method is located on the ENVI main menu at

Classification > Supervised > Parallelepiped

First choose your image of interest. Note that if there are no available training regions already loaded, the parallelepiped method will not work. That is, if you come back to classify an image at any time other than when you generated and saved your ROIs, you will first have to reload the ROIs on the image.

Assuming that you've worked through the steps up until this point, the ***Parallelepiped Parameters*** dialogue will appear. Enter the following parameters:



Select Classes from Regions: Click on the desired ROI names in the list or click on ***Select All Items***.

Set Max stdev from Mean: This allows you to choose the number of standard deviations from the mean that you would like to use. If you use the default value of 3, this is nearly equivalent to using the class max and min values. If you choose a smaller number, you can, in effect, make the parallelepiped boundaries *smaller* or more restrictive. This will prevent too many image pixels from being put into that class. If you use a larger number, you are expanding the parallelepiped boundaries beyond the class max and min to grab ‘nearby’ pixels. Notice that you can also choose to set standard deviations *individually for each class* if you like, which could be helpful if you notice that two of your classes overlap in spectral feature space.

Output Rule Images: This is an optional parameter that allows you to show the intermediate classification product before final assignment of classes. The pixel values of the resulting rule images range from 0 to n (where n is the number of bands) and represent the number of bands that satisfied the parallelepiped criteria. In other words, for a given ROI class, a pixel’s value in the rule image represents the number of bands in which that pixel was within the boundaries of the parallelepiped box. Note: *If more than one match occurs, the first class to be evaluated (the first ROI from the selected list) carries over into the classified image*. If an image pixel did not fall into any parallelepipeds (a rule image value of 0), it will be labeled *unclassified*.

Output Filename: Select a name for the final classified image (such as “class_parallel_florida.bin”). Click ***OK*** to begin the classification.

There will be a status message box indicating that the procedure is running. Once completed, you will see the new map appear in the ***Available Bands List***.

Open a new display window for the classified map. What do you see? Are locations that you are most familiar with classified correctly? Make sure to look at any land cover features that you know might pose issues.

e. Run the **maximum likelihood classification algorithm**.

You will now compare the results of the parallelepiped algorithm to those from the maximum likelihood classifier. To do this, you will need to use exactly the same training sites (same ROI file) as input to both classifiers.

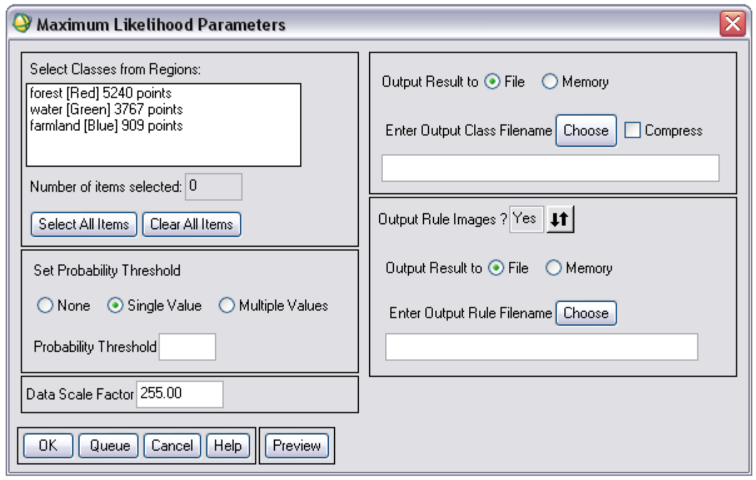
Go to

Main ENVI menu bar > Classification > Supervised > Maximum Likelihood

Select your input file and click ***OK***. The ***Maximum Likelihood Parameters*** dialog appears:

Input the following parameters:

Select Classes from Regions: Select the ROIs you wish to use for this classification.



Set Probability Threshold: For this setting, you will need to refer to your notes on the maximum likelihood classifier. Remember that the training sites are used to estimate a set of probabilities that give the relative likelihoods that a given pixel belongs to each land cover class (i.e. the classes you specify). It is implicit that every pixel will be classified into one of the available classes, no matter how low its class membership probabilities really are. To prevent this from happening (which could generate a poor classification), you may set the probability threshold so that only pixels above a certain probability (which is scaled 0 to 1) are labeled that class. For example, you may wish to set this value to 0.2; ENVI then will not classify pixels if their probability of class membership is below 0.2. To start, use 0 for this setting, but feel free to use a range of values (start low, such as 0.1, 0.2, etc.) if you feel it may improve your classification results. You may determine it is better to have unclassified pixels than mis-classified pixels in your final map

Data Scale Factor: The scale factor is a division factor used to convert integer scaled reflectance or radiance data into floating-point values. For uncalibrated integer data (such as the ASTER data we are using), set this to the maximum value the instrument can measure ($2^n - 1$, where n is the bit depth of the instrument). For 8-bit instruments, for example, set the scale factor to 255. So, make sure to put 255 for this value.

Output Result to File: Specify your directory and provide a name for your classified map.

Output Rule Images: This option will produce a multi-band image, where each band corresponds to one of your land cover classes. The values in each band indicate the probability of a given pixel belonging to that class. Once you get used to looking at these images (view these one band at a time in gray scale), this information can be very helpful for refining your training sites, determining where additional sites are needed, etc. However, this option will produce a very large image, so you will want to be mindful of zipping or removing the rule images if they are not needed.

f. **Analyze** the results.

There are several different ways to look at your results. First, open the results from both classifiers as well as the original multiband image, and compare the results visually. The second way is to look at the class statistics. This can be achieved by going to:

Classification > Post Classification > Class Statistics

In the first window, choose the input **classified image** and click OK. In the second window, choose the input image (the **original image**), which specifies the image from which the statistics will be generated. Click OK. In the ***Class Selection*** dialog, choose the classes of interest for which statistics will be extracted. If desired, ***Select All Items***. Click OK. At the ***Compute Statistics Parameter*** window, select ***Output to a Text Repor***t and make certain that the ***Basic Stats*** box is checked. Click OK. This will compute statistics on all of your individual classes and report a class statistic summary.

The third (and most important way) is to assess your classified map for accuracy. The steps to do this will be assigned next week.

g. **Edit** the training sites, re-run the classifier.

Once you have assessed the results of the two classification algorithms, you will most need to do two things: (1) decide which algorithm is working better, and (2) edit your training sites to correct mistakes in the map. Once you have added, deleted, or modified the training sites, run one of the two classification algorithms. Repeat this process until you achieve a satisfactory result.

4. Methods for accuracy assessment

For this portion of the lab, you will examine accuracy by generating a new, independent set of test sites, labeling these sites, and then comparing their labels to those in the classified map.

a. **Check the image header for the number of lines (y direction) and samples (x direction) in your image.**

Go to the ***Available Bands List*** window, and highlight the name of the file (right next to the little image icon). Right click on the file name, and select ***Edit Header***. A new window will appear – look for ***Samples*** and ***Lines***, and write down the number for each on a piece of paper.

b. **Select test sites using a random sample.**

You will now create an image of 'noise' (pixels with random values), which you will then sample to generate random test sites. First, create a random uniform image. Go to:

Main ENVI menu bar > File > Generate Test Data

In the ***Generate Image Parameters*** window, fill in the following:

Output image value: **Random (uniform)**

Min value: **0**
Max value: **10000**
Seed: (leave blank)

Samples: **lines and samples must be *exactly* the same as your image**
Lines: (from the ENVI header of the image you just opened)
Bands: **1**

Output data type: **Floating Point**
Enter output filename: (specify your directory and a file name)

Hit ***OK***. After the image has been created, view the image in a separate image window. Use the ***Cursor Location/Value*** tool to look at the pixel values.

Now, threshold the random value image, and turn a range of data values into an ROI layer:

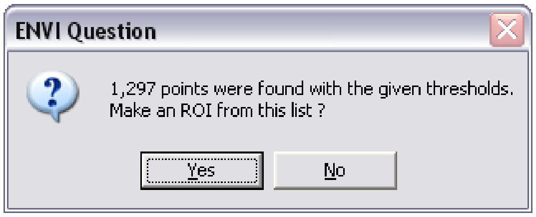
Main ENVI menu bar > Basic Tools > Region of Interest > Band Threshold to ROI

Select the random image for the input band. Now, plug in a few values for ***Min Thresh Value*** and ***Max Thresh Value*** and hit ***OK***.

ENVI will find all pixels in the random image that fall within the range specified. The software will prompt you with the number of pixels it will generate, and allow you to choose whether to proceed.

Here is an **example**:

Let's choose **0** and **1** as our min and max threshold values. After hitting ***OK***, the following window appears:



We are only interested in 100 random pixels, so 1297 'points' is far too many (ENVI uses points to refer to the pixels that would be selected). Choose no, and try again.

This time, let's take advantage of the floating point properties of the image and choose a much smaller slice. This time, we'll use **0** and **0.001**. This returns only 2 pixels. This is too small.

Finally, let's use **0** and **0.1**. This returns 144 pixels, and we are much closer to our goal of 100. With some additional refinement to our min and max values, we will generate a number very close to 100.

Once ENVI returns the desired number of pixels (e.g. please select at least 50-100 test pixels for this lab), choose ***Yes*** and then save the ROIs to a file. This is the set of test sites that you will now label using *only* Google Earth, the original Landsat data, and other ancillary data sources.

c. **Label the test sites.**

Open the original satellite image and restore the new ROIs (make sure to delete any other open ROIs to avoid confusion).

Open a *second viewer* with the same remotely sensed image, and link the viewers.

Your task is to now view each random pixel in Image Window #1, determine the land cover type in Image Window #2 (use the flicker tool here), and then <u>redigitize</u> a new ROI pixel and type in its land cover label (do this in Image Window #1) in the ROI tool box.

To get started, return to your classified maps, and look at your final classification scheme. In the ROI tool, add a series of new, empty ROIs for each of your classes. Make sure your class scheme here matches your maps exactly.

Now, highlight the random ROI category, and use the **Goto** command to view one of these pixels. Once you've located the pixel, decide on its class label. Highlight the corresponding class in your list of new, empty ROIs, and digitize a new region <u>*right on top of the red thresholded pixel*</u>. Make sure **Zoom** is selected for the digitizing window, and use the **Point** option from the ***ROI Type*** menu bar options.

Once complete, highlight the list of random ROIs again, and hit **Goto** to move on to the next pixel. Then repeat the labeling / digitizing process, go to the next pixel and repeat, etc.

Once you have finished labeling all your test pixels, save your ROIs to a file. Make sure to SAVE OFTEN.

d. **Estimate the confusion matrix.**

On the main menu bar, go to:

Classification > Post Classification > Confusion Matrix > Using Ground Truth ROIs

Select your final classified map from the image list, and then click ***OK***.

In the next window, you will see boxes where you can match the ground truth ROIs (what you just digitized and labeled) to the thematic classes in your classified map. To make a new match, click on a class (such as **forest)** in the ***Select Ground Truth ROI*** list in the upper left corner. Then click on the corresponding category in the ***Select Classification Image*** list in the upper right corner. Now hit ***Add Combination***. You will see your new choice in the list of ***Matched Classes*** at the bottom of the window.

When you have finished adding and checking your classes, click ***OK***. In the ***Confusion Matrix Parameters*** box, make sure ***Pixels***, ***Percent*** and ***Report Accuracy Assessment*** are all checked (by checking the ***Report*** button, overall accuracy, user's and producer's accuracy will all be calculated for you; otherwise, only the error matrix values will be reported). Click ***OK***. A new window will appear with your results in text format.

e. **Save the confusion matrix results.**

The results will be a little difficult to interpret at first, since they are not formatted very nicely. Unfortunately, you will not be able to print this information directly. To save this information, click on

File > Save Text to ASCII

in the ***Confusion Matrix*** window. Make sure you specify your directory, and give this file a name such as "test_site_accuracy.txt". Since you are saving many things for your labs, it is worth spending a few minutes to *intelligently* organize and name your files so that you can easily find what you are looking for later.

Once you save the file, you should import the file into Excel and edit/delete the accuracy results for your final report (please feel free to ask for instructions if you do not know how to do this step). You should take the time to format and edit the data into easy-to-read tables (i.e. type in the full names of classes, make sure all columns and rows have labels, eliminate any extra information that is not necessary in the table, and give the table a caption).

f. **Understanding the results.**

The most important part of any accuracy assessment is interpreting the results. You will need to understand and describe the results for four different measures: overall accuracy, the kappa statistic, user's accuracy for each class, and producer's accuracy for each class. Refer to chapter 13 of Jensen's *Digital Image Processing* textbook and your class notes as you look at the resulting values.

Since the goal of Lab 7 was to compare the parallelepiped and maximum likelihood classification algorithms, you will need to repeat the above steps for <u>both of your classified maps</u>, produce two confusion matrices, and import both matrices into Excel.

Report the results of the confusion matrices as outlined in the assignment below. Do not bias the estimates of class or map accuracy, since you will not be graded on *how high* your accuracy result is. Rather, you will be graded on your understanding, explanation, and discussion of the accuracy assessment results.

5. Assignment

You will need to submit *several* things for this lab: a lab report, your map results (with captions), and any tables/statistics (with captions).

<u>The final report</u>

In your own words, write a 2-3 page report that includes the following:

- Introduction: Explain your goal, and *briefly* describe what supervised classification is.

- Methods: Provide a detailed summary of the supervised classification process. Make sure you explain how each algorithm works, its advantages and disadvantages, and any issues that can arise from both the algorithm or from supervised classification in general, and ultimately how you can address these issues. Talk about your strategy and process for creating training sites. What worked and what did not? Again, reference any sources if you use them.

- Results: Explain the results of your classification with each classifier, provide and asses class statistics, etc. What class scheme did you choose? How well did each class do? Which classes were confused, and why do you think this confusion occurred? How did you remedy this? Which parameters did you try? Did you properly account for the variations in the land cover and the original image? Remember, you will need to save your maps as jpg files and insert them into your report with captions and legends, and you should also insert any class statistics (for the training sites or the final map) as condensed tables.

- Discussion: Describe what land cover classes were difficult to map or had confusion, and why this occurred. How did you solve the problem? What parameters did you manipulate during classification? Did these changes help or hurt? If you had problems in several classes, or different types of confusion, describe each and how you rectified the problem. Discuss whether you thought supervised or unsupervised classification worked better and why, and whether or not one was easier than the other.

Please do not answer these questions with bullet points or question numbers. Weave these questions into a short report including full sentences and complete ideas.

<u>Submitting the final report</u>:

For this lab, you must submit your final report electronically by emailing it to the class email account. Please put your name and the lab number in the subject heading of your email. For student Bucky Badger, the subject line would read:

Bucky Badger – Lab 1

The name of your *word document* must also have the lab number and your name in this format:

Bucky_Badger_lab1.doc