GEO/EVS 425/525 Unit 15
Quantitative Supervised Classification

In the last laboratory, we carried out a quantitative approach to unsupervised classification of satellite imagery. In this laboratory, we will use a new classification-oriented program, eCognition, to do a quantitative supervised classification of the same quadrangle. Quantitative supervised classification differs from standard supervised classification in that it is much more objective. Specifically, it does not require you to establish *a priori* training sites or even to specify the classes you will use. eCognition will do this for you. But it does so in a way that establishes specific signatures much as would be done if you set up your own training areas, and the logic by which it proceeds is very much that of a supervised classification.

If you were to read the description of eCognition, you would find it described as "image segmentation" software. What that means is that it segments an image into what are, in essence, training areas based on similarity of the sets of digital numbers within various regions. You can alter some of the parameters that determine the shape of the regions considered or the degree of similarity required within an image segment, but image segmentation is a much more objective process than standard training-area identification nevertheless.

To begin eCognition, click on the eCognition icon on your desktop. It will take a while for the program to open, but when it does, open your subset TM image. Be sure that the file type is "ERAS Imagine Image." You will see the 6 or 7 layers of your image listed as having been loaded.

Recall that you had to define a set of land-cover classes, and that you had to take a relatively large number of statistical classes and assign each to the most appropriate land-cover class. Many of you objected that some of the land-cover classes to which you were assigning the statistical classes did not seem as appropriate as you would have liked, especially considering that you had a reasonable basis for defining "appropriate" based on the supervised classification you did in exercise 10. In this laboratory, we will do a much less common, but more meaningful, approach to classification based on data-fusion and identification of a series of quantitative continua along which statistical classes can be associated. Assignment of statistical classes to land-cover classes, if desired, can be done by dynamic definition of land-cover classes in terms of the quantitative continua.

You should already have done the basic calculation needed to carry out quantitative unsupervised classification. That is, you have carried out an unsupervised ISODATA classification in which you derived 50 statistical classes. You can use these classes. If you didn't derive this many, simply do an ISODATA classification in which you will get 50 statistical classes from your quadrangle.

It is probably easiest to introduce you to quantitative unsupervised classification by comparing it to the qualitative unsupervised classification you have already done. Each type has 4 steps, but they are quire different:

Step	Qualitative Unsupervised Classification	Quantitative Unsupervised Classification
1	Carry out an ISODATA classification on the image of interest	
2	Develop a set of qualitative classes which are meaningful in the light of the focus of the study	Develop a set of continua which are meaningful in the light of the focus of the study
3	Place each statistical class derived in step 1 in one (and only one) of the classes identified in step 2	Place each statistical class derived in step 1 at a position along each and every continuum identified in step 2

4	Field-check the results of the classification

Obviously the first and last step are the same. The essence of qualitative classification is that its goal is nominal classes to which statistical class will *belong*. No statistical class can belong to more than one land-cover class. The land-cover classes are mutually exclusive, and much of the significance of a classification exercise is determined by how well the list of land-cover classes chosen covers the universe of land covers found in the study area. The essence of quantitative classification is that land cover can be described in terms of a series of factors which vary continuously from a minimum to a maximum, typically 0% to 100%, although other continua are possible in principle. It is not necessary to insure that these continua include all aspects of land cover, depending on the needs of the study being carried out, but it is necessary that the analyst define each continuum in a way that insures that all statistical classes can have a place within it.

In qualitative unsupervised classification, the classes chosen in step 2 will be things like “residential,” “forest,” “water,” etc. In quantitative unsupervised classification, the continua chosen in step 2 will be things like “% impervious surface,” “% water,” “% canopy cover,” etc. As long as percentages are used for the continuum, the stricture that the analyst must insure that all statistical classes derived in step 1 have a place along each continuum is met by definition.

It is obvious that any land-cover type can be described in terms of a set of criteria. For example, typical high-density development such as might be found in the City of Cleveland might be described as having about 70% impervious surface, 25% grass, 10% canopy cover, and 0% open water. The 4 variables need not sum to 100%, because some land is covered by more than one variable (e.g. the canopy of trees shades grass, sidewalks, or buildings) and there may be issues that are not considered in the analysis that are locally important (e.g. specialized land cover such as landfill). What is important is that the analyst state the *type* of information he or she wishes to derive from the imagery by classification rather than the topology of *results*.

In carrying out your quantitative classification, you should identify the continua you wish to consider. The most commonly appropriate are likely to be % impervious surface, % grass, % canopy cover, and % water, but specialized uses will have reason to use others. Your choice should be made on the basis of what is most meaningful to you.

The next step is to fuse data at as high resolution as possible with your image. In this exercise, you should use the DOQQs as your high-resolution data set. Basically, what you do is to take the DOQQs for your quadrangle and put them into a viewer and then put the classified image you are analyzing on top of it. Change the opacity of each statistical class to about 0.4, so that you can see the details of the underlying DOQQ but still tell colors apart. Keep the raster attribute editor open! Go to the most common class (i.e. the one with the largest number in the “histogram” column), and change its opacity to 0. You are now looking through the classified image to the DOQQ underlying it. Every clear pixel (i.e. not covered by a classified image pixel with an opacity of 0.4) belongs to the statistical class currently under consideration. Look at a large number of pixels of this class. How do you describe it, both qualitatively and quantitatively? Specifically, where does the average pixel in this class fall along *each* of the continua you are considering? That is, on the average, what is the percentage of impervious surface, water, tree canopy, and grass for the class? You should be able to estimate an average value – as well as a judgement as to the variability and/or consistency of the class – for each continuum you have defined.

When you have finished with your assessment of the most common type of statistical class, go on to the next most common type, and so on until you have considered every statistical class derived in the ISODATA calculations, and each statistical class has been associated with a specific level on *each* of the continua you are considering.

You can now do one (or both) of two things. First, what you have done is to give quantitative

significance to the statistical classes you derived via ISODATA. You can recode your classified image to reflect that, or you can carry out a dynamic land-cover class definition. To do the first, all you have to do is to recode your classified image (i.e. Interpreter -> GIS Analysis -> Recode) so that each class in the input image is changed to the quantitative level of one of the continua in the output image. For example, if the estimate of % impervious surface for class 6 is 50%, then you would recode class 6 to be equal to 50 in the impervious-surface image. The result of these transformations is a set of images, one for each continuum considered in the analysis, in which each pixel is represented by a ratio number, typically a percentage, describing its status with regard to that continuum. The overall image describes the state of the landscape under consideration with regard to that continuum. For example, continuing the example given earlier in this paragraph, the result of recoding all statistical classes as the estimate of their % impervious surface would be a map of impervious surface. When you have completed your recodes, you should compare the maps of your quantitative classifications with the qualitative classifications you did in units 10 and 11. Which contains more information?

You should then convert your continua to land-cover classes. To do this, you identify a series of land-cover classes (as you did previously), and then you prepare definitions of each class in terms of the continua you derived from your analysis. For example, high-density residential land might have 50 - 75% impervious surface, 0 - 25% canopy cover, 10 - 30% grass, and 0 - 10% open water. You can easily develop a conditional model to do this dynamic definition of land-cover classes. You might ask why one should define land-cover classes this way rather than doing it directly with a qualitative approach as in the standard approach. The answer is simple. When you do a standard classification, you can't do much with it. You are limited to a single definition of land-cover types, and once you have associated each statistical class with a land-cover type, you can't easily change that. When you develop a series of continua within which you can define land-cover classes, you can easily adjust the numbers in your dynamic classification to adapt the analysis to different interpretations or different studies. Try making some of these changes.

When you are satisfied with the results of your analyses, you should print the maps of each continuum you have considered, and you should also print a land-cover map reflecting one or more dynamic classifications you feel show some validity. **These images should be included in your portfolio for this unit.**

Questions to Consider

1. Which is more meaningful, quantitative or qualitative classification?

2. Which of the two approaches to unsupervised classification provides the best results? That is, what do you mean by "best results"?

3. Is the extra time required by quantitative classification worth it, based on the overall level of retrieval of information from an image and the significance of the information retrieved?

Portfolio

1. A set of maps based on your classified TM image, showing the levels of each continuum you have considered in the classification process.

2. A map showing land-cover classes based on the continua you have considered in this exercise, with a brief summary of the definition of each class in your land-cover assessment.