

## GEO/EVS 425/525 Unit 12

### Classification Error Analysis

This exercise will enable you to evaluate the classifications you have carried out during the previous two weeks. In doing so, you will not use Imagine except to create printouts that you can use to carry out your evaluation. The primary authority for classification error analysis is *Assessing the Accuracy of Remote Sensed Data: Principles and Practice*, by Russell G. Congalton and Kass Green.

There are several ways in which to evaluate classifications, but all involve comparing the classification being evaluated with a reference. When one is dealing with vector files, for example, one can compare objects in the classified image with objects in the same place, either on the ground or in a reference image. In a raster image, one can either compare *zones*, where a number of pixels have the same attribute values, with zones on the reference, or one can compare the classified image to the reference on a pixel-by-pixel basis. Again, the reference can be a reference image, or it can be the actual situation on the ground.

For this exercise, you will compare your classified images – both the unsupervised and supervised classifications – with the situation on the ground. Because you did not classify your image relative to a standard set of land-cover classes found in a reference image, you will do your evaluation on a pixel-by-pixel basis.

The evaluation is carried out in five steps. The first two steps are done before you go into the field; the third is done in the field; the last two are done when you return to the lab.

Step one is to choose a representative set of pixels to evaluate. The total number should be at least 500 from your image. This may sound like a lot, but it really isn't. For a typical quadrangle, this amounts to about 1/3 of 1% of the total area of the image. On the other hand, each pixel in your image is 30 m x 30 m, or approximately 100 ft x 100 ft, or 10,000 square feet, so that 500 pixels constitutes about 5,000,000 square feet. Again, this sounds like a lot, but it is equivalent to less than 1/5 square mile. There are many ways to choose a field-check area. If you were doing this to support a research effort, you might use a random-number generator to choose the points you will evaluate. More informally, you might choose about 20 random points to which you could easily drive and then evaluate 25 pixels at random in the vicinity of each of those points. The choice is yours. But you should be prepared to discuss the methods by which you chose your points.

Step 2 is to prepare base maps and work sheets that you can take with you. To prepare a base map, you simply blow up the areas you have chosen so that you can tell what value you have ascribed to your classified image for each pixel you will evaluate when you get to the field. This is fairly easy if you have a relatively small number of classes (i.e. less than 10 or so). You can print out appropriate pieces of the image with the DLGs for roads and rivers superimposed over them, with the classes shown in different – and recognizable – colors. You should be able to get the appropriate portions of each image into 4-6 tiles each for each of your two classifications. To make a work sheet, you prepare a piece of paper on which you can record the *classified* land-cover type and the *actual* land-cover type for each pixel. This can be as simple or as complicated as you wish. Most probably, you will have a sheet of paper with a series of rows and three columns: one for the actual land-cover type and one column each for your classification in the supervised and unsupervised classifications.

Step 3 is to go to the field and carry out the field-checking. For each pixel you evaluate, you record the land-cover type you classified it as *and* the actual land-cover type you found in the field. It is important that you use exactly the same classification scheme for your evaluation as you did for your classification. When you are done in the field, your field notes will have three data fields for each data point: the actual land-cover type, your unsupervised classification for that data point, and your supervised classification for that data point.

In step 4, you will create an error matrix from these data for each of your classifications, such as the

matrix shown in Table 1. The process of making an error matrix is very simple. An error matrix is nothing more than a series of rows and columns, where the rows correspond to your classifications of each pixel, and the columns correspond to the actual situation for each data point. To make your error matrices, you record, for each data point for each of your two classifications, the class you *determined* from each classification against the class you *found* for that point in the field. With luck, most of your classes will agree. That is, if you determined that a particular pixel belonged in a “high-density residential” class, you will find that the actual land use for that pixel was high-density residential. These classes will go along the diagonal, as shown in Table 1. Sometimes, however, you will find that a pixel you classified as “high-density residential” will, in fact, be water – or at least something other than high-density residential. This pixel would be recorded in the *row* for high-density residential and the *column* for water. When you have recorded all of your field data, you add up all of the rows and columns. Each column total for a land-cover type will be the total number of pixels of a given type *found in the field*; each row total for a land-cover type will be the total number of pixels *determined for the classification*; the grand total will be the lower-right-hand total, which is the row total of all of the columns (or the column total of all of the rows).

Step 5 is the actual analysis, which is an evaluation of the error matrix to calculate the accuracy of the classification. Note that the form of the matrix allows us to ask two questions: “How accurate is the classification, when compared with the field data?”, and “To what degree is it possible to represent reality in a classified image?” The first question is one that the *producer* of the data would ask; the second is one that a *user* of the data would ask. To answer the first question, we look at the correctly-classified pixels in each land-cover class and compare them with the total number of pixels allocated by the classification process to that class. That is, we compare the correctly-classified pixels in each class with the column totals for that class. The ratio of the correctly-classified pixels to the column total represents the proportion of the pixels found in a particular class that were actually allocated to that class. This is called the *producer accuracy*, and it assesses the degree to which the producer of the image succeeded in allocating pixels in each class to their *correct* class.

In the same way, we can estimate the *user accuracy*, or the degree to which a user can rely on an image to represent reality. The ratio of the pixels found in a particular class to the row total in that class represents the proportion of the pixels classified to a particular class that were actually found in that class.

Finally, we can estimate the *overall accuracy*, which is an aggregate measure of both producer and user accuracy and assesses accuracy of *all* classes, rather than accuracy on a class-by-class basis, as is the case with producer and user accuracy. The overall accuracy can be defined as the ratio of the sum of all correctly-classified pixels to the total number of pixels evaluated.

Your results in this exercise will be two error matrices, one for your best unsupervised classification and the other for your supervised classification. **These two error matrices should be turned in as your portfolio contributions from this unit. If you think it might be helpful, you might also include a brief discussion of your results.**

#### Questions to Consider

1. Are the producer accuracy and the user accuracy equivalent to each other? Equivalent to the overall accuracy? Why or why not?
2. Are the accuracies of each row and column the same for each class, or do some classes typically have higher accuracy than others? Why?
3. In what way (if at all) does field-checking of classified imagery assist you in improving your classification? That is, if this were a research project, would the error analysis help you define a “next step”?

#### Portfolio

1. Error matrix from your best unsupervised classification of your quadrangle

2. Error matrix from your best supervised classification of your quadrangle.
3. (Optional) Discussion of your error matrices, including the differences between them and any follow-up you might be inclined to do.

**Table 1: Error Matrix for unsupervised classification based on Land-use/land cover classes**

Land Use/Land Cover Class	Park	Grass	Low-Dens Res.	Mid-Dens Res.	Hi-Dens Res	Commerce	Industrial	Pavement	Sum	User Accuracy
Urban Park	88	6	7	23	6	0	1	0	131	0.67
Grass	2	25	0	0	2	0	0	0	29	0.86
Low-Density Residential	4	0	43	2	1	0	0	0	50	0.86
Medium-Density Residential	2	4	0	61	4	2	4	0	77	0.79
High-Density Residential	4	0	1	7	67	8	13	2	102	0.66
Commercial	0	1	0	1	5	45	0	0	52	0.87
Industrial	2	0	0	6	4	2	50	2	66	0.76
Pavement	1	0	0	1	2	2	0	37	43	0.86
	103	36	51	101	91	59	68	41	550	
Producer Accuracy	0.85	0.69	0.84	0.60	0.74	0.76	0.74	0.90		0.76