# Lesson 6: Accuracy Assessment

## INTRODUCTION

This lesson will provide a general overview of some of the basic concepts of accuracy assessment and how it can be used specifically for qualifying a land cover data set derived from remotely sensed imagery using image classification methods. The last lesson focused on different land cover classification methods. The classification process involved a number of steps and processes to actually generate the result, the classified land cover map; however, up to this point we do not know how good the result is.

## LESSON OBJECTIVES

By the end of this lesson, you will be able to:

1. Explain the importance of accuracy assessments and why they are used for image classification project(s).
2. Describe the computed measures of a typical accuracy assessment.

3. Perform an accuracy assessment on the products of remote sensing workflows.

4. Incorporate accuracy assessment results into interpretation and analysis of workflow outputs.

## LEARNING SEQUENCE

|  |  |
| --- | --- |
|  | |
| Required Reading | Read the following:  Accuracy Assessment   * Image Classification Assessment Accuracy |
| Assignments | Complete the following:   * Quiz * Lab |

## INSTRUCTION

## Importance of Accuracy Assessments

### Accuracy Assessment: Classification Process

During the classification process, the analyst fills in the blanks with inputs and outputs and a few other parameters and clicks the OK button to generate a result. The analyst may have color coded the results or converted spectral classes into information classes, but we still don’t know how good the answer is.

For a real-world product, the data must be qualified with statistical information as to how good the result is and how well the methods used to produce the end product worked.

Besides the quantitative assessment of the product, a written document often accompanies the results that provide the explanation and interpretation of the results.

Some software packages provide statistical analysis tools which can be used, but having some background such as education and/or training in statistics is often helpful, since the quantitative analysis, interpretation, and written documentation and explanation of the results are left to the analyst who performed the image analysis work.

### Assessment Measures

The final classification an analyst obtains from a set of digital image processing methods is not finished until there is a quantitative assessment conducted.

The assessment provides two basic measures:

1. It provides a quantitative measure of how “good” the final classification is
2. It provides some inference as to how well the methods such as the training sets, the spectral signature analysis, and image processes methods were to generate the product.

If the image classification accuracy assessment turned out well, then one can conclude that the methods used to generate the result were appropriately conducted and executed.

### Sampling Options

When talking about statistics one often refers to sampling methods and design to obtain samples that can be used to generate the statistics. Two common sampling methods that are used with image classification projects are random sampling and stratified random sampling.

Often times the stratified random sampling approach is used to collect training and accuracy assessment samples, since there are categories of land cover types that are being evaluated and so samples are often chosen within the geographic extents of these land cover types.

In any case, the samples taken must be independent. Samples should not be exactly adjacent to one another, since pixels in the adjacent area could be in more than one sample. The samples should also represent the spectral breadth of each land cover type, so samples should be taken throughout the extent of each category. Samples used for training the image classifier cannot be used for performing the accuracy assessment.

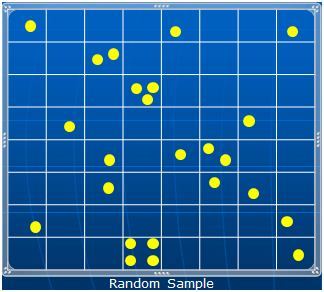
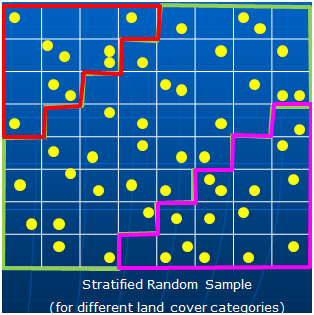
Ideally, an analyst would take enough samples for each land cover type that would result in “statistically valid” results. In many practical situations, this may not be possible. Regardless, taking independent samples to perform any quantitative analysis is better than no quantitative analysis.

### Sampling Methods

The images provided illustrate two sampling methods: random sampling and stratified random sampling. The samples are taken at random locations throughout the image, irrespective of the land cover type.

Each of the colors represents specific land cover types. Within each category, random samples are chosen throughout the extent of the respective category. Normally, spectral samples are taken early in the land cover classification process before an actual image classification algorithms is executed. Once the land cover type for each spectral sample is known (often through the use of ground truth or photo interpretation), the samples are divided into training samples and accuracy assessment samples.

**Random Sample Stratified Random Sample**

## Image Classification Assessment Accuracy

### Performing an Accuracy Assessment

This section focuses on the methods used to perform an accuracy assessment for an image classification. Typically, the stratified random sampling method is used to generate samples for the accuracy assessment. Depending on the software package an accuracy assessment can be performed using either individual pixels or groups of pixels. In most current image classification work, a group of pixels (or polygons) is the base unit for conducting the accuracy assessment. As mentioned earlier, the polygons used as training sites for the image classifier should not be used for assessing the accuracy of the image.

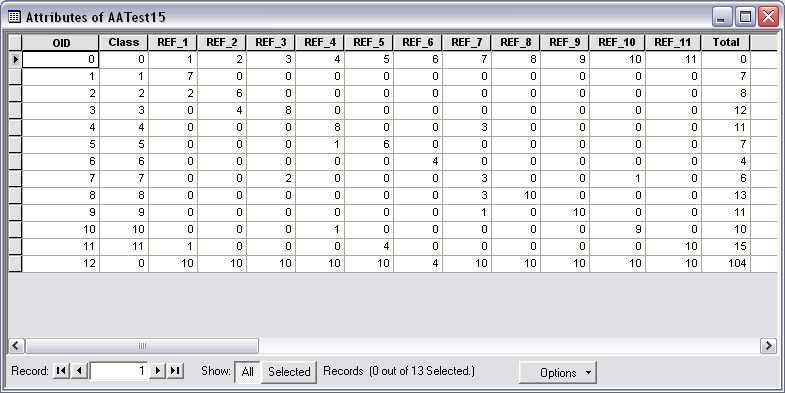
An analyst must have some kind of program to perform the accuracy assessment. Since the accuracy assessment computes statistical values, a statistics software package is needed such as Statistical Analysis System (SAS) or Statistical Package for the Social Sciences (SPSS) if the image processing software does not have an accuracy assessment routine or module. In the lab accompanying this unit, a custom-built accuracy assessment program has been created by the author that students can use for this course as well as their own image classification work. This program functions within the ArcGIS software. See the lab for more details.

A general rule of thumb for parsing out training versus accuracy assessment sites is to use approximately two thirds of the sites for training the classifier and about one third are set aside for performing the accuracy assessment. Since all of the spectral samples are delineated at the same time and the land cover type identified through ground truth or air photo interpretation, the sites can be divided into training and accuracy assessment sites for each land cover type before executing the image classifier.

### Error Matrix

This is an example of an Error Matrix, a table that is created and filled in with values when an accuracy assessment program is executed. The matrix is essential a square table that has the same number of rows and columns which represent the different land cover types in the classified image.

When the accuracy assessment routine is run, the error matrix is filled in with values. The values along the diagonal, which are shown as the “red” boxes, represent the total number of accuracy assessment polygons that were categorized or classified correctly. The values that are off of the diagonal are accuracy assessment polygons that are misclassified.



### Measures Computed

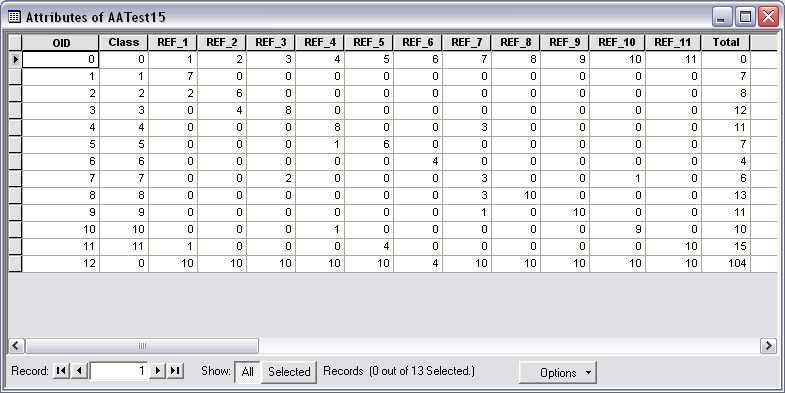
Several measures are computed as part of the accuracy assessment.

1. The error matrix provides user’s and producer’s categorical accuracies for each land cover type
   1. A user’s accuracy describes how well each land cover type is categorized in the data set
   2. A producer’s accuracy describes how well the methods were implemented to achieve the land cover data set.
2. The overall accuracy
3. The Khat or Kappa statistic compares the overall accuracy with the possibility of randomly assigning land cover types to the image pixels.
4. In some accuracy programs a set of confidence intervals are provided for both the user’s and producer’s accuracy. The accuracy assessment program used in this lab computes confidence intervals.

### Error Matrix – Commission Error (Included)

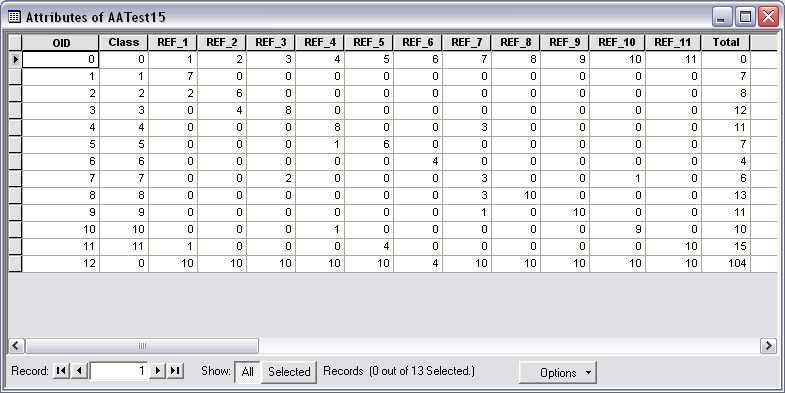
Each land cover category can be reviewed and evaluated in an error matrix. By looking at each row the analyst can look at all of the non-diagonal values to see what kind of confusion exists between the different land cover types. If the off-diagonal values are high or there are a large number of off-diagonal values within the error matrix, this can infer there is more confusion between the category that was assessed and other land cover types found in the image classification. These errors are grouped into two categories, commission error and omission error.

Commission errors are those non-diagonal values that occur along each row. In the example shown here 1 site of category 1 and 4 sites of category 5 were classified (or included) as category 11.



### Error Matrix – Omission Error (Excluded)

In a similar manner the omission error can be evaluated by looking at the non-diagonal values in each column. In this case, seven sites that should have been categorized as class 7 were categorized as another class. In this example, three of these sites were categorized as class 4, three were categorized as class 8, and one site was classified as class 9. As can be seen, class 7, class 4, and class 8 may have significant confused between the three classes and should each be re-evaluated before completing a final accuracy assessment to see if one or more of these mapped land cover types can be improved.



7 Sites of Category 7 were omitted from this category

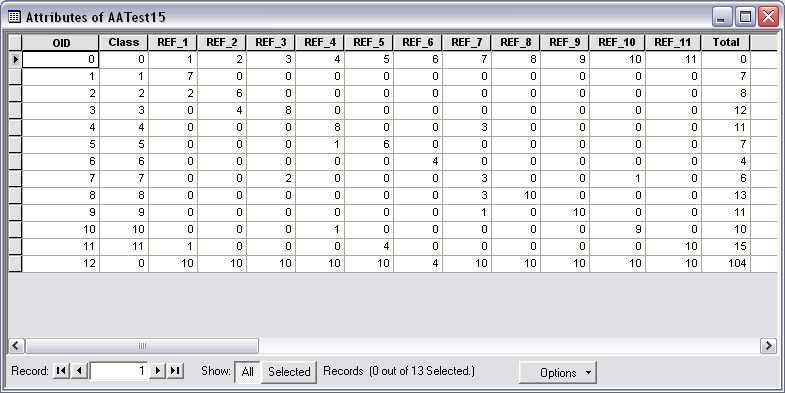
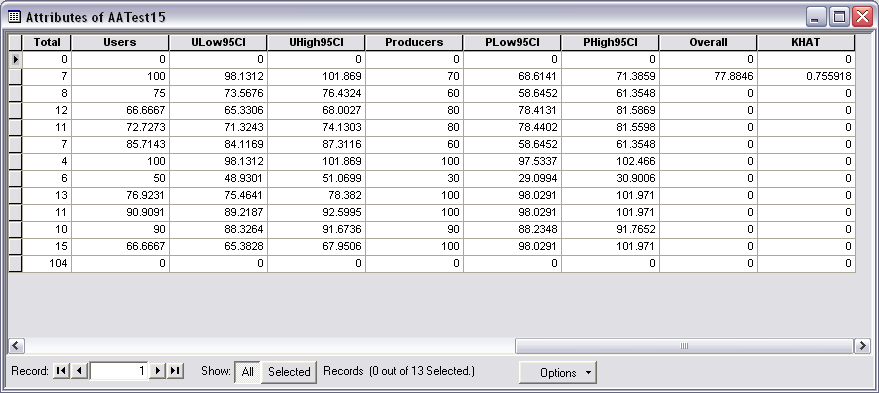
(i.e. classified as some other cover type instead of Category 7)

3 sites classified as category 4; 3 sites classified as category 8 and

1 site classified as category 9

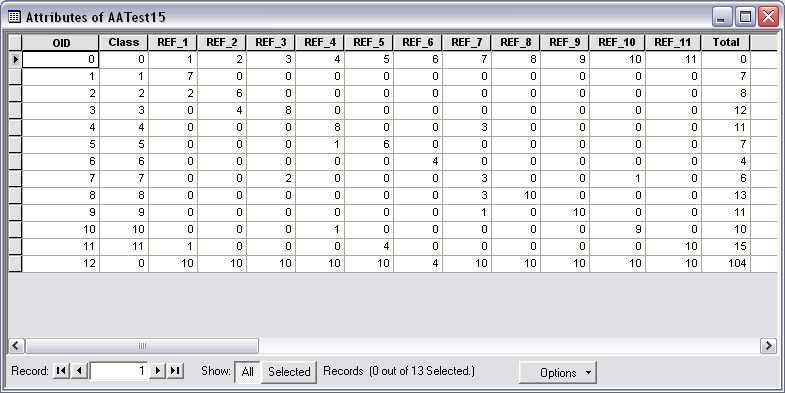
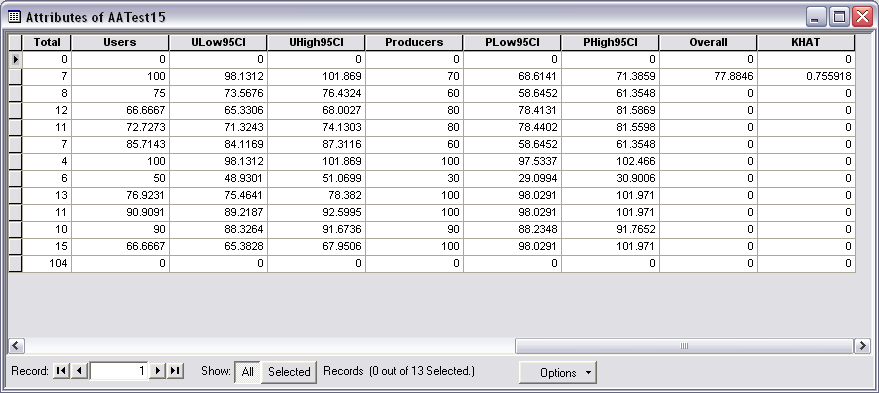
### Error Matrix – User’s Accuracy

The user’s accuracy can be computed by taking the number of polygons that were correctly classified and divided by the row total. These are shown in the yellow boxes. In addition, the table on the right shows another part of the error matrix table that contains the actual user’s accuracy for this land cover class. Also, the respective confidence intervals are computed as well.

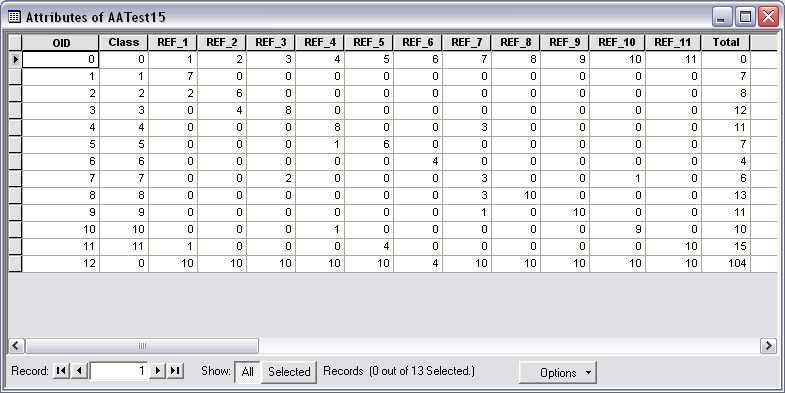
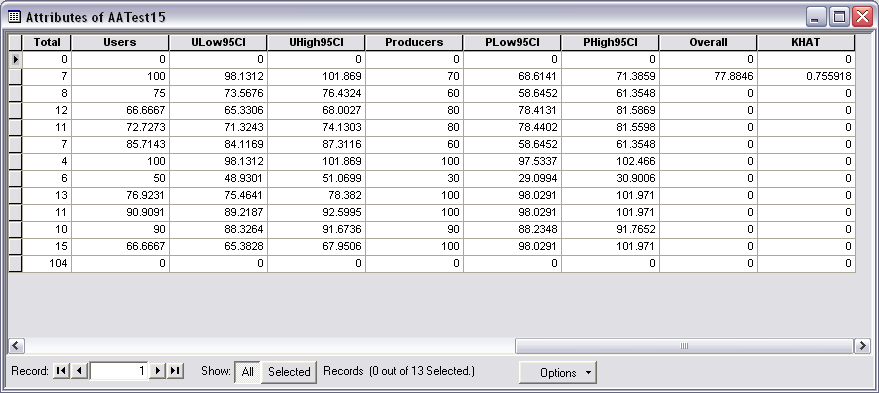
### Error Matrix – Producer’s Accuracy

Likewise, the producer’s accuracy can be computed by taking the number of correctly classified polygons and dividing by the column total as shown in the yellow boxes. The right table shows the respective producer’s accuracy and the accompanying confidence intervals.

** **

### Error Matrix – Overall Accuracy

The overall accuracy can be computed by taking summing the individual values along the diagonal and dividing by the total number of accuracy assessment sites. The values along the diagonal are accuracy assessment polygons that have been correctly classified. The overall accuracy assessment is shown in the table on the right.

### Kappa “Khat” Statistics

The Kappa or “Khat” statistic is also commonly computed and shown in addition to the overall accuracy value. This value provides a measure that compares the actual accuracy assessment to that of randomly assigning pixels to land cover categories or chance agreement. This slide shows the basic computation that is involved. The actual computation can be found in the *Remote Sensing and Image Interpretation* book by Lillesand, Kiefer, and Chipman, (2007). The same computation logic is used in the accuracy assessment program provided in the Unit 6 lab.

As an example, if a Khat value of 0.70 was computed, this means that the image classification is 70% better than an image classification resulting from chance (or random assignment of pixels to land cover categories).

Khat = Observed Accuracy – Chance Agreement

1 – Chance Agreement

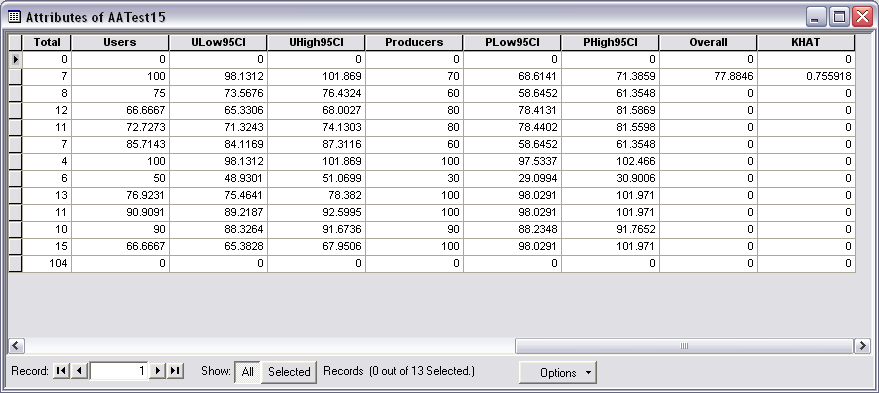
Khat = 0.70 indicates an observed classification is 70% better than one resulting from chance.

**Key Elements of Khat Statistics**

* Khat tends to be stated as a lower and more conservative value than the overall accuracy
* A low Khat value indicates that an observed classification accuracy is no better than a random assignment of pixels
* The Khat computation incorporates diagonal and non-diagonal components
* Both the overall and the Khat statistics are stated as part of an accuracy assessment.

### Confidence Intervals

The confidence intervals for each category are also useful since they can provide a low and high range for the user’s and producer’s accuracy to fall within and can provide a measure of how variable the categorical accuracy can be throughout the resulting image classification.



### Accuracy

To conclude, an image analyst is more concerned with the quality of each class being classified correctly, that is, the producer’s accuracy. The user is often concerned about how well each class has been identified, that is the user’s accuracy. Also, the categorical accuracies are often more important that the overall accuracy because the end-user really does want to be able to have confidence in the mapped land cover types, since the specific categories may be used in subsequent processes and analyses.

## Summary

In this lesson you learned about accuracy assessment and image classification assessment accuracy. The classification process and assessment measures were explained and sampling examples were provided for your understanding. This lesson also provided examples and an explanation of an error matrix used to determine the accuracy of images.

## Assignments

1. Quiz
2. Lab