Exercise 1: Accuracy Assessment

Objectives

- Perform an accuracy assessment of a classified Landsat image
- Construct an error matrix

Overview of Topics

- Review of sampling strategy and reference points
- Add reference points to a map and extract values
- Create an error matrix and assess the level of accuracy in the classification

Tools Needed

- Esri ArcGIS 10.x
- Microsoft Excel

Associated Data

Please ensure you have downloaded the necessary data and saved it to your computer prior to conducting this exercise. The data can be found on the ARSET website here: https://arset.gsfc.nasa.gov/land/webinars/18adv-land-classification. The necessary data:

- Clipped and classified Landsat 8 image (saved as Landsat_Classified.tif)
    - This image is from northern California on April 06, 2016. The original Landsat image filename is LC08_L1TP_028027_20160830. This is a rural, forested region and was classified using six distinct land cover types.
- Reference_Points.shp

Introduction

For this exercise we will use a pre-classified Landsat 8 image and we will assess the accuracy of that map using a set of reference points. This image was classified using the supervised classification method.

For information on how to conduct a land cover classification, please view the previous ARSET webinar: *Land Cover Classification with Satellite Imagery*, located here: https://arset.gsfc.nasa.gov/land/webinars/advanced-land-classification
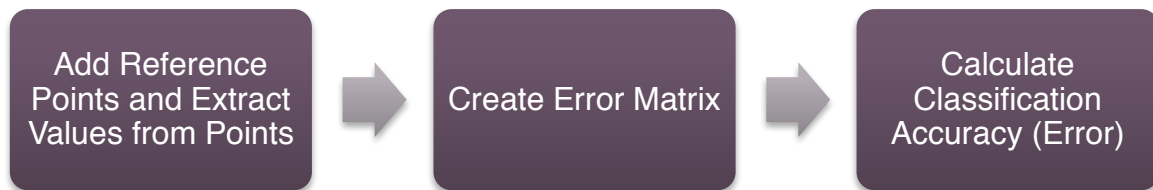
Now that we have a Landsat image classified using a supervised classification method, we want to quantitatively assess the error. Accuracy assessments allow us to evaluate how closely the classified image is to reference data from a different source of information (e.g. field data, aerial imagery, etc.). Specifically, "accuracy" is tested by comparing test pixels to the corresponding location in the classified image. One of the main reasons to check our accuracy or "error" is to determine where errors are occurring and whether the image needs to be resampled.

In this exercise, we have provided the reference data. These reference data were generated using the stratified random sampling approach that generates random points within subareas of a population, or strata. Here, we have improved upon a simple random approach because the number of reference points is proportional to the estimated area of each stratum. Thus, larger strata will have more points. To do this, we used the NOAA Biogeography Branch's Sampling Design Tool for ArcGIS.

More information on the Sampling Design Tool can be found here:
https://www.arcgis.com/home/item.html?id=ecbe1fc44f35465f9dea42ef9b63e785

Methods Outline

Add Reference Points and Extract Values from Points → Create Error Matrix → Calculate Classification Accuracy (Error)

*Outline of methodology for an accuracy assessment.*

Part 1: Set up Your Workplace and Load Associated Data

1. Launch **ArcMap** on your computer.
   a. Go to the **Microsoft Logo Start** icon at the lower left of screen, click it, and then choose **All Programs** then, **ArcGIS** then, **ArcMap**
2. Open a new **Blank Map**

First, make sure that your **Spatial Analyst** extension is activated. This is an important extension for much of the remote sensing processes we conduct in ArcMap.

3. At the top of the window, go to **Customize** > **Extensions**
4. Click the check mark for **Spatial Analyst**
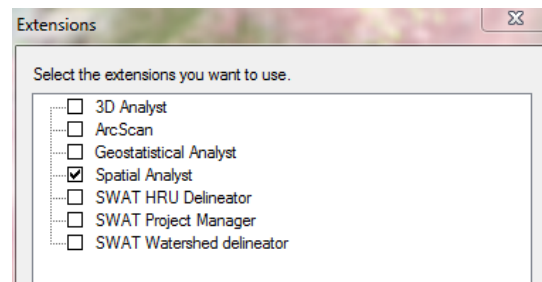
First, we will load the classified image:

5. On the toolbar, click the **Add Data** icon
6. Navigate to your Accuracy Assessment folder and add the **Landsat_Classified.tif**
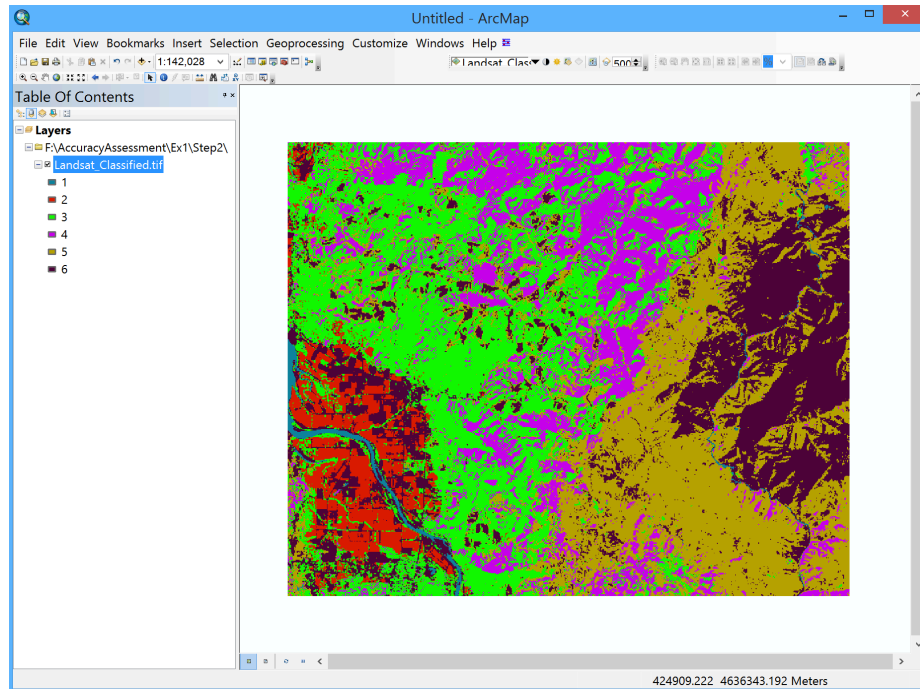
Now, we will rename our land cover categories and change the color scheme to be more intuitive.

7. Right click on the Landsat_Classified file in the **Table of Contents** and click on **Properties**

8. Click on the **Symbology** tab, and click on **Unique Values** on the left panel under **Show**

9. Click on each label and rename them to match our land cover categories of:
    a. Water
    b. Agriculture
    c. Hardwood (light forest)
    d. Conifer (dark forest)
    e. Grass/shrub
    f. Bare ground

10. Change the colors by clicking on each categories color icon. Pick colors that seem logical. Click **Apply** and **OK**.
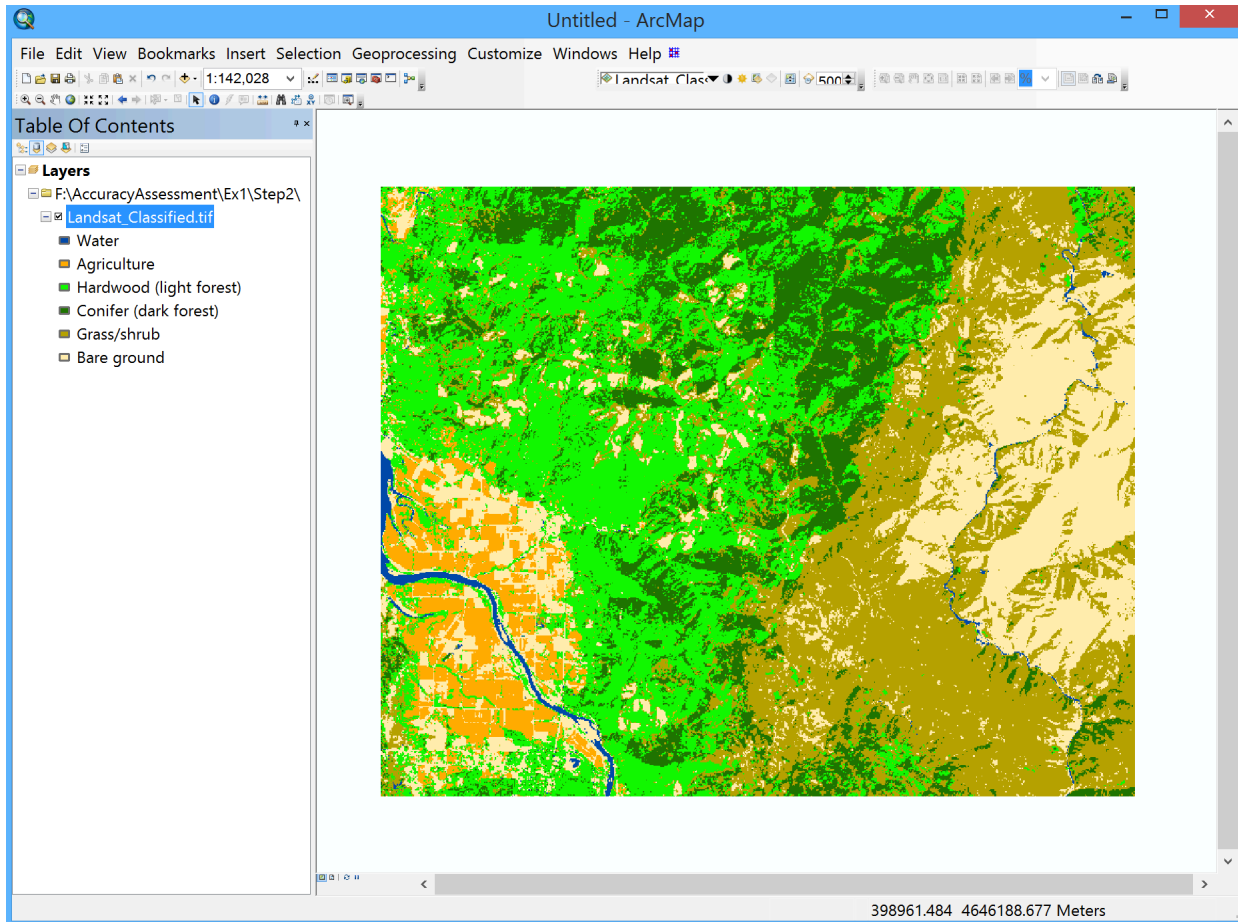
Saving Your ArcMap document

It is always good practice to save your map along the way.

11. Go to the top of the ArcMap window and click on **File** > **Save As**
12. Save your file as **Accuracy_Ex1.mxd** (the extension should add automatically) in your Accuracy Assessment folder.

Adding Your Reference Points

Ideally, when performing an accuracy assessment, you have reference data collected from the field. This provides "ground truth" for classification. For this exercise, we will use mock reference data as "ground truth."

13. On the toolbar, click the **Add Data** icon
14. Navigate to your Accuracy Assessment folder and add **Reference_Points.shp**
15. In the **Table of Contents**, right click on **Reference_Points** and click **Properties**. Click on the **Symbology** tab. On the left hand side, click **Categories** > **Unique values**.
16. Under the **Value Field**, select **LandCover**. Then click the **Add All Values** button. Click **Apply** then **OK.**
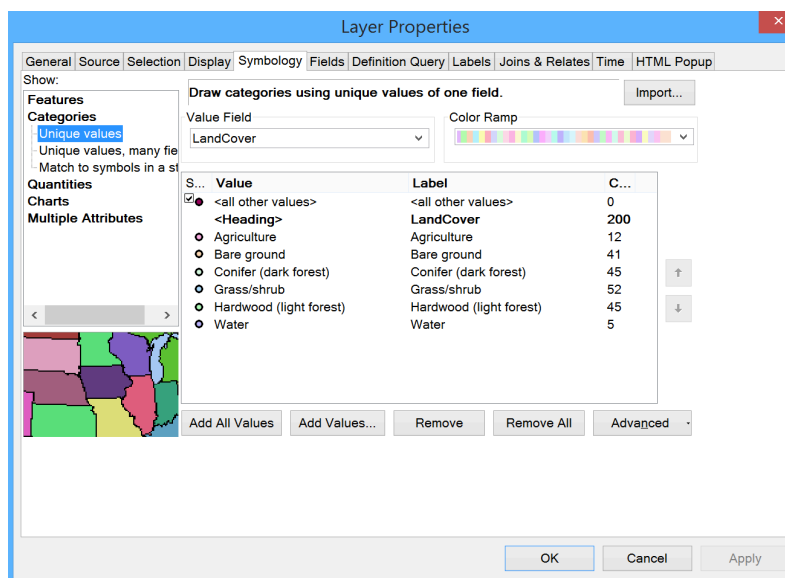
This will show the reference points for all the values. You should see six reference point classes, the same as your classified Landsat image (water, agriculture, etc.).

Part 2: Extract Pixel Values for Each Reference Point

Now that we have all the same points loaded, we want to get the pixel value for each point to compare them to our classified image.

1. Navigate to the **Extract Values to Points** tool in the ArcToolbox (**Spatial Analyst Tools** > **Extraction** > **Extract Values to Points**). This tool extracts the pixel value for each point.

2. For **Input point features**, select **Reference_Points**. For Input raster, navigate to your Accuracy Assessmnet folder and select **Landsat_Classified.tif**. Save the output as **Values.shp** in your Accuracy Assessment folder and click **OK**.
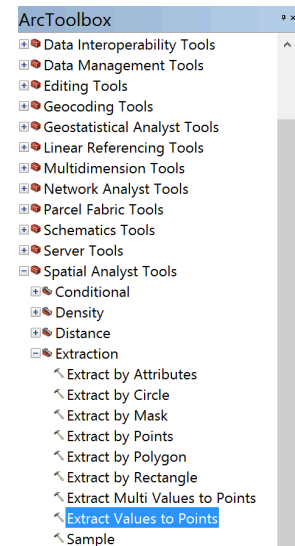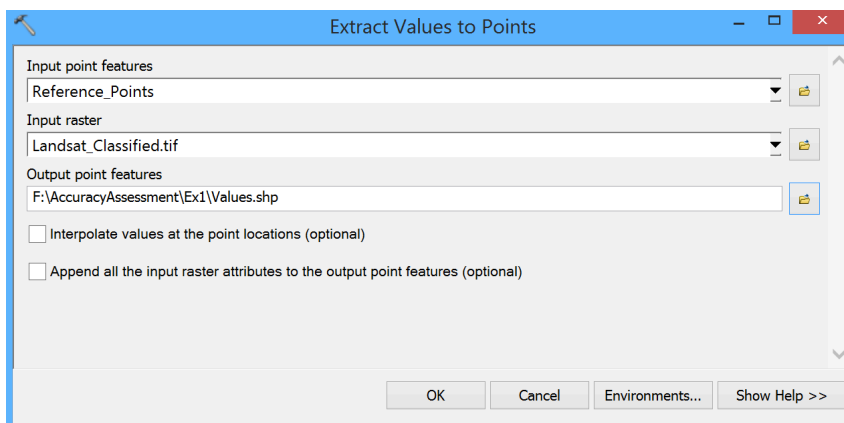
3. Once the processing is complete, the new file should be added to your map. In the **Table of Contents**, right click on the new **Values.shp** file and click on **Open Attribute Table.** You will see that your new classified values are in the field **RASTERVALU**.

FID	Shape *	Id	LandCover	Values	RASTERVALU
0	Point	0	Hardwood (light forest)	3	3
1	Point	0	Hardwood (light forest)	3	3
2	Point	0	Conifer (dark forest)	4	3
3	Point	0	Hardwood (light forest)	3	3
4	Point	0	Hardwood (light forest)	3	3
5	Point	0	Grass/shrub	5	3
6	Point	0	Hardwood (light forest)	3	3
7	Point	0	Hardwood (light forest)	3	3
8	Point	0	Hardwood (light forest)	3	3
9	Point	0	Conifer (dark forest)	4	3
10	Point	0	Hardwood (light forest)	3	3
11	Point	0	Hardwood (light forest)	3	3
12	Point	0	Hardwood (light forest)	3	3
13	Point	0	Hardwood (light forest)	3	3
14	Point	0	Hardwood (light forest)	3	3
15	Point	0	Hardwood (light forest)	3	3
16	Point	0	Hardwood (light forest)	3	3
17	Point	0	Hardwood (light forest)	3	3
18	Point	0	Hardwood (light forest)	3	3
19	Point	0	Hardwood (light forest)	3	3
20	Point	0	Hardwood (light forest)	3	3
21	Point	0	Conifer (dark forest)	4	3
22	Point	0	Hardwood (light forest)	3	3
23	Point	0	Hardwood (light forest)	3	3

(0 out of 200 Selected)

"RASTERVALU" and "Values" are not very explanatory names, so let's rename the columns.

4. Click on **Table Options** , and select **Add Field**. Name this field **Classified**. For type, select **Short Integer** and click **OK**.

5. Once the column appears, right click on the new **Classified** column and click **Field Calculator**. You may see a warning that says "You are about to do a calculation outside of an edit session…" Click **Yes.**

6. In the **Fields** section of the **Field Calculator**, double click on **RASTERVALU** and click **OK**. This will transfer the information from the **RASTERVALU** column to the **Classified** column

7. Do the same for **Values** to change the name to **Reference**

8. Select **Add Field**, rename the field **Reference**. For type, select **Short Integer**. Right click on the new **Reference** column, go to the **Field Calculator**, select **Values**, and click **OK.**
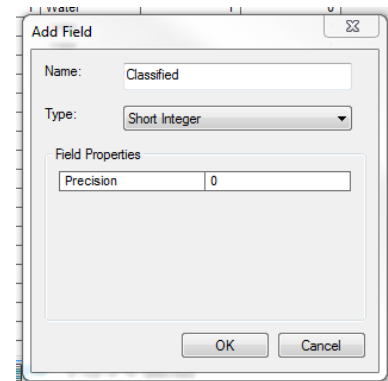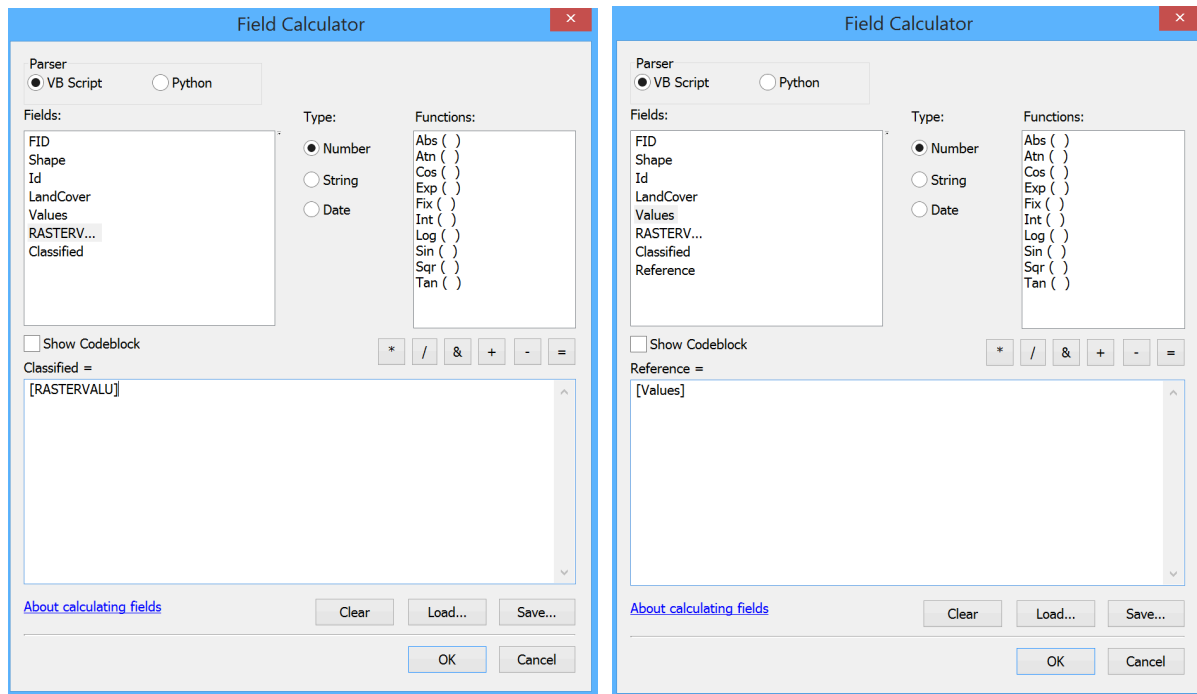
9. Delete the **RASTERVALU** and **Value** columns. To do this, right click on the **RASTERVALU** column and select **Delete Field.** When you see the **Confirm Delete Field** warning, click **Yes**.

10. For clarity, sort the table based on the land cover categories we specified in our classified image. Right click on the **Reference** column and click on **Sort Ascending**

The **Values.shp** table should look like this:

Part 3: Calculate Frequency

Now that we have the values of the reference points, we want to isolate the frequency values for the classified image and reference points.

1. Close the Attribute Table
2. Navigate to the **Frequency** tool in the ArcToolbox (**Analysis Tools** > **Statistics** > **Frequency**).
3. In the **Input Table**, select the **Values** file. This will automatically load the columns from the Attribute Table. In the **Output Table,** navigate to your Accuracy Assessment folder and save the file as **Frequency_Table.dbf**. For the Frequency fields, select **Classified** and **Reference**. Keep the Summary Field(s) unselected. Click **OK**.

FID	Shape *	Id	LandCover	Classified	Reference
195	Point	0	Water	1	1
196	Point	0	Water	1	1
197	Point	0	Water	1	1
198	Point	0	Water	1	1
199	Point	0	Water	1	1
34	Point	0	Agriculture	2	2
35	Point	0	Agriculture	2	2
36	Point	0	Agriculture	2	2
37	Point	0	Agriculture	2	2
38	Point	0	Agriculture	2	2
40	Point	0	Agriculture	2	2
41	Point	0	Agriculture	2	2
42	Point	0	Agriculture	2	2
43	Point	0	Agriculture	2	2
64	Point	0	Agriculture	6	2
67	Point	0	Agriculture	6	2
68	Point	0	Agriculture	6	2
0	Point	0	Hardwood (light forest)	3	3
1	Point	0	Hardwood (light forest)	3	3
3	Point	0	Hardwood (light forest)	3	3
4	Point	0	Hardwood (light forest)	3	3
6	Point	0	Hardwood (light forest)	3	3
7	Point	0	Hardwood (light forest)	3	3
8	Point	0	Hardwood (light forest)	3	3

(0 out of 200 Selected)

4. Once the process is complete you should see a **Frequency_table** under the **Values** file in the **Table of Contents**. Right click on **Frequency_table** and click **Open.**

a. This table shows us how many pixels were correctly predicted in the Landsat image for each reference point ("ground truth"). The frequency table gives us some information, but it's still hard to read. Let's put it in an easier form. Close the table.

Input the Frequency Into a Pivot Table

In these next steps, we'll create a pivot table to sort and display the frequencies.

5. Navigate to the **Pivot Table tool** (**Data Management Tools** > **Table** > **Pivot Table**).

6. For the **Input Table**, select **Frequency_Table**. For **Input Fields**, select **Classified**. For **Pivot Field**, select **Reference**. For value field, select **FREQUENCY**. this will create a new table with classified values as rows, reference points as columns, and the frequency of the values in each of the cells. Save as **Error_Matrix.dbf** and click **OK.**

7. After the process is complete, the Error Matrix table will be added to your map. In the **Table of Contents**, right click on the **Error_Matrix**, and select **Open.**

OID	Classified	Reference1	Reference2	Reference3	Reference4	Reference5	Reference6
0	1	5	0	0	0	0	1
1	2	0	9	1	0	0	4
2	3	0	0	39	10	6	1
3	4	0	0	3	33	1	0
4	5	0	0	2	1	40	5
5	6	0	3	1	0	5	30

8. You will see a table with six land cover class types and six reference point columns. Close the **Error_Matrix** table.

Export the Pivot Table to Excel

9. Navigate to the **Table to Excel tool** (**Conversion Tools** > **Excel** > **Table to Excel**).

10. For the Input Table, select **Error_Matrix**. For the Output Excel File, click on the yellow folder and save as **Error_Matrix.xls** in your Accuracy Assessment folder and click **OK.**

11. Once the process is complete, close the Table window. If you have not already, save your ArcMap document as **Accuracy_Ex1.mxd**

Part 4: Calculating Error in Excel

The accuracy of an image is most often reported as a percentage correct. As discussed in the lecture, there are different *types* of accuracy. Here we will calculate the overall accuracy, the user's accuracy, and the producer's accuracy.

1. Minimize your ArcMap window and open **Microsoft Excel**
2. Navigate to the Accuracy Assessment folder and open your **Error_Matrix** file

This is your error matrix (also called a "confusion matrix"). The error matrix will show your reference data in *columns* and your classified data in *rows*. It should look like this:

The classes are still labeled as numbers, so rename them in the rows under the Classified column and for each Reference label to match your class type (e.g., Reference1 as Water, Reference2 as Agriculture, etc.). You can refer to the beginning of this exercise for all the class types.

3. Delete the **OID** column

Your table should look like this:

4. Add a **Total Reference Points** column at the end and sum the values across. Use the **SUM** function to sum all the cells across. This will add up all the reference points for each land cover type.

Hold down the lower right hand corner of the cell and drag down to apply the same equation to each row.

5. Add a **Total Classified Points** row at the bottom of the **Classified** column and sum the values for each column (Water, Agriculture, etc.). Hold down the lower right hand corner of the cell and drag across to apply to the same equation. This column shows the number of pixels classified for that land cover class type.

6. After you sum across and down, your table should look like this:

7. To calculate the error, we want to sum the pixel values correctly matched and divide by the total reference points.
8. Underneath your table (skip a row to avoid confusion), make a new row and call it **Total Correct Reference Points**
   a. To calculate the total correct reference points, use the **SUM** equation and hold down **CTRL** to add the values diagonally (highlighted). If you look closely at the table these are the number of pixels that were classified as the same land cover type as the reference points (Water/Water, Agriculture/Agriculture)

9. Make another cell and call it **Total "True" Reference Points**. This is the total number of pixels provided in your reference points (blue box below). This is the value in the last row under **Total Reference Points**. So in the row next to **Total "True" Reference Points**, type in the equals sign and click on that cell (H8).

10. Add a row called **Percent Accuracy** (skip a row below the **Total "True" Reference Points** to avoid confusion). To calculate the percent accuracy, divide your total correct reference points by your total "true" reference points and multiply this by 100.

Percent Accuracy = [Total correct reference points / Total "true reference points]*100

11. For this example, the percent accuracy is approximately 78 percent. Percent accuracy explains how *accurate* your classification was; that is, the percent of correctly classified reference points out of total reference points. This is also called the "percent error."

12. Save your Excel workbook.

Errors of Omission and Commission

We can also quantify the accuracy of the individual classes through the calculation of errors of commission (user's accuracy) and omission (producer's accuracy). Errors of commission occur when a pixel is incorrectly included in a category being evaluated, while errors of omission occur when a pixel is left out of the category being evaluated.

First, let's start with the user's accuracy (errors of commission).

13. Add a row called **User's Accuracy** (skip a row below **Percent Accuracy** to avoid confusion).

14. Copy each of the land cover class and paste them below **User's Accuracy**

15. The user's accuracy is found by dividing the diagonal number by the row total and multiplying by 100. Input this calculation for each land cover class.

16. Change the cell format for each of the user's accuracies to show two digits after the decimal. Highlight all the cells, along the top click on **Format** > **Cells**. Under **Category**, click on **Number**, and next to **Decimal Places** enter 2. Click **OK**.

17. To keep the excel spreadsheet organized, merge and center cells 15A and 15B. Highlight both cells, and click on the **Merge and Center** function. ✓⊞ Merge & Center   Then make the background color for that cell the same as the headings along the top of the spreadsheet.

Let's do the same thing for the producer's error (errors of omission).

18. Copy the land cover classes in to the D column (starting at D16, see below). Create another heading called **Producer's Accuracy** in the D column. As you did in the previous step, **Merge and Center** the **Producer's Accuracy** cell (D15) with the cell to the right (E15) and make the background color the same as the other headings.

19. The producer's accuracy is found by dividing the diagonal number by the column total and multiplying by 100. Input this calculation for each land cover class.

20. As you did previously, change the cell format for each of the user's accuracies to show two digits after the decimal.

Conclusion

Accuracy assessment is an important step in analyzing your land classification. In this exercise, you gained an understanding of how to perform an accuracy assessment for a classified image. You extracted pixel values from reference points, created an error matrix, calculated the overall percent accuracy, and the user's and producer's accuracies. Assessing the accuracy is a good check point to determine whether you need to re-sample your classification, or are ready to move forth with your analyses. Next week we will focus on standard error in our classification and conduct some more complex statistical processes.

Additional Online Resources

Food and Agriculture Organization: Map Accuracy Assessment and Area Estimation:
http://www.fao.org/documents/card/en/c/e5ea45b8-3fd7-4692-ba29-fae7b140d07e/

Texas A&M Accuracy Assessment of an Image Classification in Arc Map:
https://www.youtube.com/watch?v=FaZGAUS_Nlo

Create Accuracy Assessment: http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/create-accuracy-assessment-points.htm