Exercise 2: Unbiased Error Estimation

Objective:

- Understand how to create an error matrix
- Conceptualize biased and unbiased error
- Learn how to conduct area-based accuracies for different land cover classes

Overview of Topics:

- Review an error matrix from a classified Landsat image
- Conduct estimated area calculations for specific land cover classes
- Calculate standard error
- Calculate unbiased user and producer accuracies
- Compare biased and unbiased errors

Tools needed:

- Esri ArcGIS 10.x
- Microsoft Excel

Associated Data:

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

- Exercise2_ErrorMatrix_Template.xlsx
    - We will use this template for calculating errors
- Your saved ArcMap document from Exercise 1 (Accuracy_Ex1.mxd). This will include the clipped and classified Landsat 8 image (saved as Landsat_Classified.tif) you downloaded for Exercise 1.

Introduction

For this exercise we will build upon Exercise 1 and further assess the error of a classified Landsat 8 image. When an error matrix does not incorporate the standard error based on the total area of each land cover class, those estimates can be biased. Factors that can affect the accuracy of a specific land cover class include the total sample size, the number of strata (or classes), and the allocation of the total sample size to each strata.

It is important to include an area proportion or stratum weight (W_i) for each land cover class based on the pixel count for each class. Area estimates adjust the

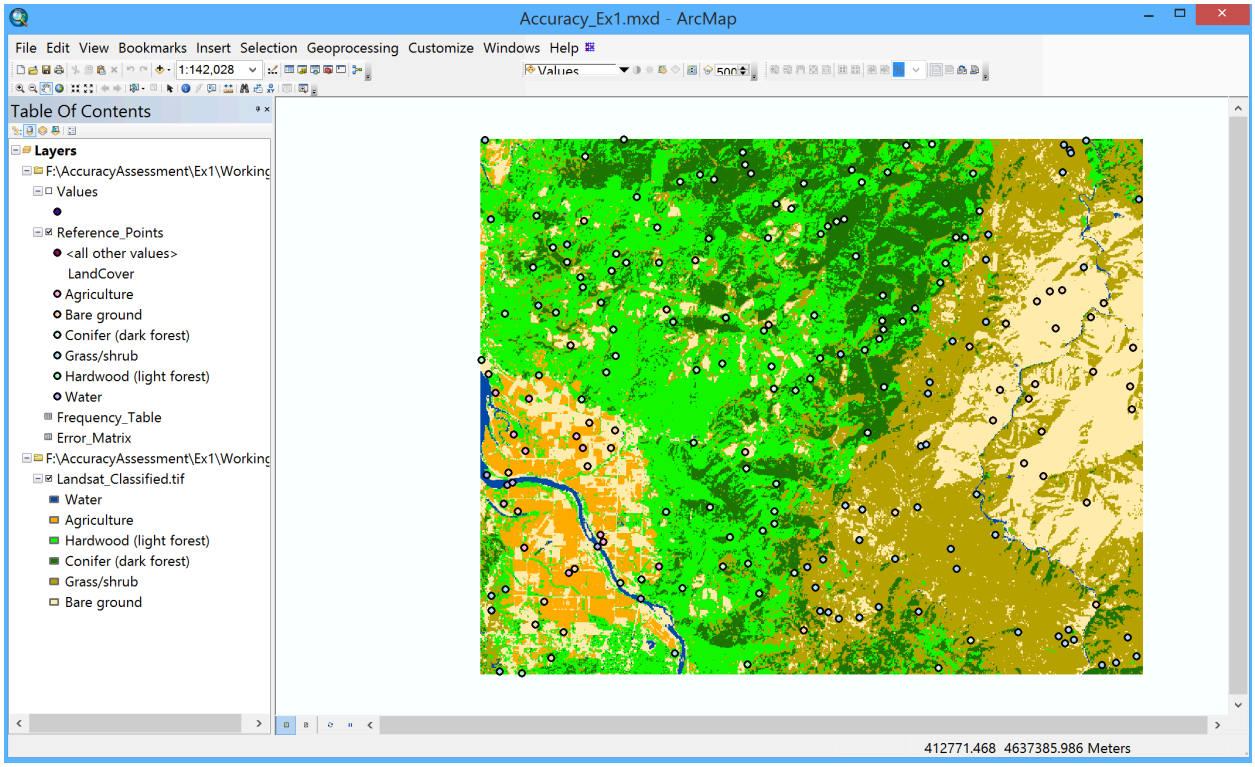
accuracy estimate and provide confidence intervals. This improves your ability to access the accuracy of the classified map. Estimated areas are included, then used to calculate the user's and producer's accuracy. The user's accuracy is assessed based on the map data, and can be calculated using the sample counts (as in Exercise 1) or with the area estimates (that we will conduct in Exercise 2). The producer's accuracy will differ depending on whether the calculation is based on sample counts or area estimates.

Further guidance can be found in a 2013 article in *Remote Sensing of Environment* (see full citation at the end of the exercise):
http://ftp-earth.bu.edu/public/olofsson/Estimation_Workshop_Lima_Aug2014/Articles/
and from the Boston Education in Earth Observation Analysis (BEEODA) here:
http://beeoda.org/

Part 1: Obtain Pixel Counts

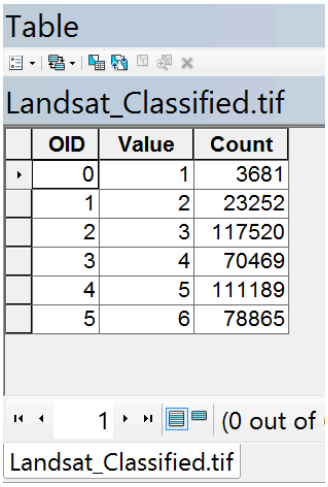
1. Launch **ArcMap** on your computer: Go to the **Microsoft Logo Start** icon at the lower left of screen, click it, and then choose **All Programs** then, **ArcGIS** then, **ArcMap**.

2. Navigate to your Accuracy Assessment folder and open your Exercise 1 map **Accuracy_Ex1.mxd**.



First, we need to obtain the pixel counts for each land cover class in order to calculate the area estimates.

3. Right click on the Landsat_Classified file in the **Table of Contents** and click on **Open Attribute Table**.



Table

Landsat_Classified.tif

	OID	Value	Count
▸	0	1	3681
	1	2	23252
	2	3	117520
	3	4	70469
	4	5	111189
	5	6	78865

1 (0 out of

Landsat_Classified.tif

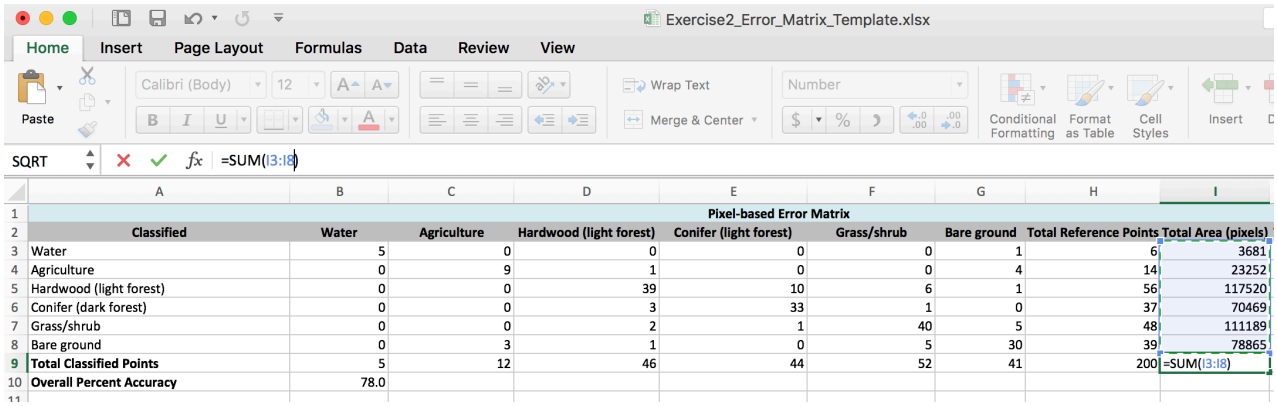
The **Count** category is the number of pixels in each class. Remember that the class **Value** corresponds to the land cover type. For reference:

1: Water
2: Agriculture
3: Hardwood (light forest)
4: Conifer (dark forest)
5: Grass/shrub
6: Bare ground

4. Keep the **Attribute Table** open in ArcMap. Navigate to your Accuracy Assessment folder and open the Excel file **Exercise2_ErrorMatrix_Template.xlsx**. This template includes the pixel counts and the overall classification that we completed in Exercise 1. Take a look at the **Total Area (pixels)** column (I). This is where we will add the pixel counts from our classified Landsat 8 image.

5. Copy the pixel counts from the **Attribute Table** into the **Total Area (pixels)** column of your **Exercise2_ErrorMatrix_Template.xlsx**.

6. In the cell I9, sum all of the **Total Area (pixels)** values (cells I3:I8) and click **Enter**.



	A	B	C	D	E	F	G	H	I
1					Pixel-based Error Matrix				
2	Classified	Water	Agriculture	Hardwood (light forest)	Conifer (light forest)	Grass/shrub	Bare ground	Total Reference Points	Total Area (pixels)
3	Water	5	0	0	0	0	1	6	3681
4	Agriculture	0	9	1	0	0	4	14	23252
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469
7	Grass/shrub	0	0	2	1	40	5	48	111189
8	Bare ground	0	3	1	0	5	30	39	78865
9	Total Classified Points	5	12	46	44	52	41	200	=SUM(I3:I8)
10	Overall Percent Accuracy	78.0							

7. Save your Excel file under a new name: **Exercise2_ErrorMatrix2.xlsx**. Also, notice that the user's and producer's accuracy are included in the template. This is the biased accuracy that you calculated in Exercise 1.

Accurray	User's Accuracy	Producer's Accuracy
Water	83.3	100.0
Agriculture	64.3	75.0
Hardwood (light forest)	69.6	84.8
Conifer (dark forest)	89.2	75.0
Grass/shrub	83.3	76.9
Bare ground	76.9	73.2

8. Close your ArcMap document. We will only be working from the Excel file for the remainder of this exercise.

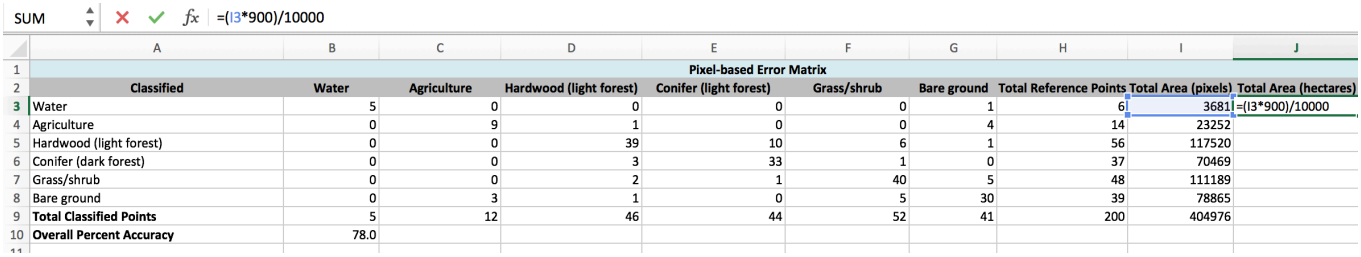
Part 2: Calculate Area and Stratum Weight (W_i)

The next step is to calculate the area for each class based on the pixel counts. Recall that each Landsat pixel is 30 x 30 meters (900 m^2). We would like our area estimates to be in hectares, a standard area unit of measurement. The conversion to hectares is 1 hectare = 10,000 square meters.

We need to do this conversion in Excel to convert our pixel count to hectares.

$$Total\ Area\ (ha) = total\ pixels \times \frac{900m^2}{1\ pixel} \times \frac{1ha}{10{,}000m^2}$$

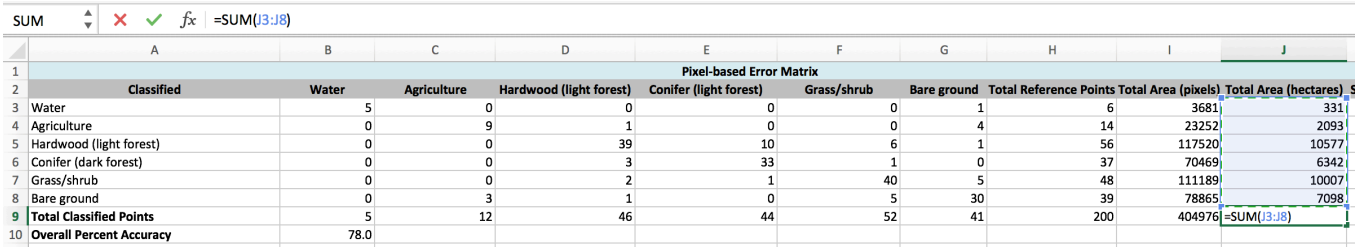
1. In the **Total Area (hectares)** column (J), multiply the **Total Area (pixels)** cell by 900, then divide by 10,000 and click **Enter**. For example, the equation for J3 will be: =(I3*900)/10000



	A	B	C	D	E	F	G	H	I	J
1					Pixel-based Error Matrix					
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Reference Points**	**Total Area (pixels)**	**Total Area (hectares)**
3	Water	5	0	0	0	0	1	6	3681	=(I3*900)/10000
4	Agriculture	0	9	1	0	0	4	14	23252	
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	
7	Grass/shrub	0	0	2	1	40	5	48	111189	
8	Bare ground	0	3	1	0	5	30	39	78865	
9	**Total Classified Points**	5	12	46	44	52	41	200	404976	
10	**Overall Percent Accuracy**	78.0								

2. Apply this equation for each land cover class (cells J3 to J8). To do this, you can hover over the bottom right of cell J3 until you see a black cross symbol, then click and drag it down. Alternatively, you can enter this same equation into each cell.

3. In cell J9, sum the area of the classified pixels and click **Enter**. The equation will be: =SUM(J3:J8)



	A	B	C	D	E	F	G	H	I	J
1					Pixel-based Error Matrix					
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Reference Points**	**Total Area (pixels)**	**Total Area (hectares)**
3	Water	5	0	0	0	0	1	6	3681	331
4	Agriculture	0	9	1	0	0	4	14	23252	2093
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342
7	Grass/shrub	0	0	2	1	40	5	48	111189	10007
8	Bare ground	0	3	1	0	5	30	39	78865	7098
9	**Total Classified Points**	5	12	46	44	52	41	200	404976	=SUM(J3:J8)
10	**Overall Percent Accuracy**	78.0								

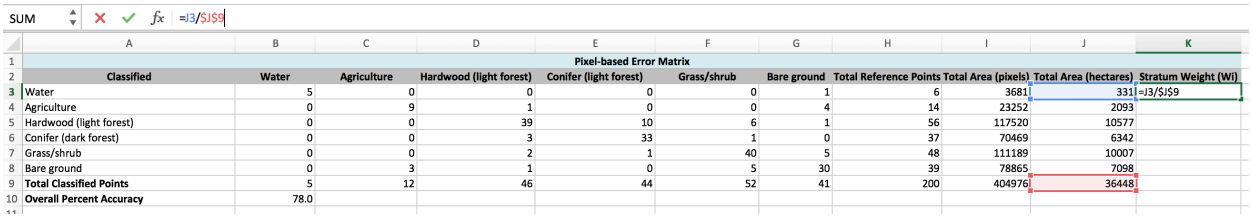
For each land cover class, we will calculate the area proportion of that class to the total area, also known as the **Stratum Weight (W_i)**.

The equation for this step is:

$$W_i = \frac{Total\ Individual\ Class\ Area\ (ha)}{Total\ Area\ (ha)}$$

4. In cell K3, divide the total area of the **Water** class by the total area of all classified pixels. This equation will be: = J3/J9 and click **Enter**. Repeat this for each land cover class.
    - Excel tip: If you would like to apply this equation to each of the land cover classes by dragging the bottom right of cell K3, you will need to anchor the

cells to all reference the area total (cell J9). To do this, you can you the dollar symbol ($J$9).



	A	B	C	D	E	F	G	H	I	J	K
1	Pixel-based Error Matrix										
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Reference Points**	**Total Area (pixels)**	**Total Area (hectares)**	**Stratum Weight (Wi)**
3	Water	5	0	0	0	0	1	6	3681	331	=I3/$J$9
4	Agriculture	0	9	1	0	0	4	14	23252	2093	
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	
7	Grass/shrub	0	0	2	1	40	5	48	111189	10007	
8	Bare ground	0	3	1	0	5	30	39	78865	7098	
9	**Total Classified Points**	5	12	46	44	52	41	200	404976	36448	
10	**Overall Percent Accuracy**	78.0									

5. Calculate the sum of all the **Stratum Weight (W_i)** in cell K9. The equation will be: =SUM(K3:K8). Click **Enter**.
    - This should equal 1, as each stratum weight is a percentage of the total area.

Pixel-based Error Matrix										
Classified	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Reference Points**	**Total Area (pixels)**	**Total Area (hectares)**	**Stratum Weight (Wi)**
Water	5	0	0	0	0	1	6	3681	331	0.00909
Agriculture	0	9	1	0	0	4	14	23252	2093	0.05742
Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	0.29019
Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	0.17401
Grass/shrub	0	0	2	1	40	5	48	111189	10007	0.27456
Bare ground	0	3	1	0	5	30	39	78865	7098	0.19474
Total Classified Points	5	12	46	44	52	41	200	404976	36448	1.00000
Overall Percent Accuracy	78.0									

Accurray	**User's Accuracy**	**Producer's Accuracy**
Water	83.3	100.0
Agriculture	64.3	75.0
Hardwood (light forest)	69.6	84.8
Conifer (dark forest)	89.2	75.0
Grass/shrub	83.3	76.9
Bare ground	76.9	73.2

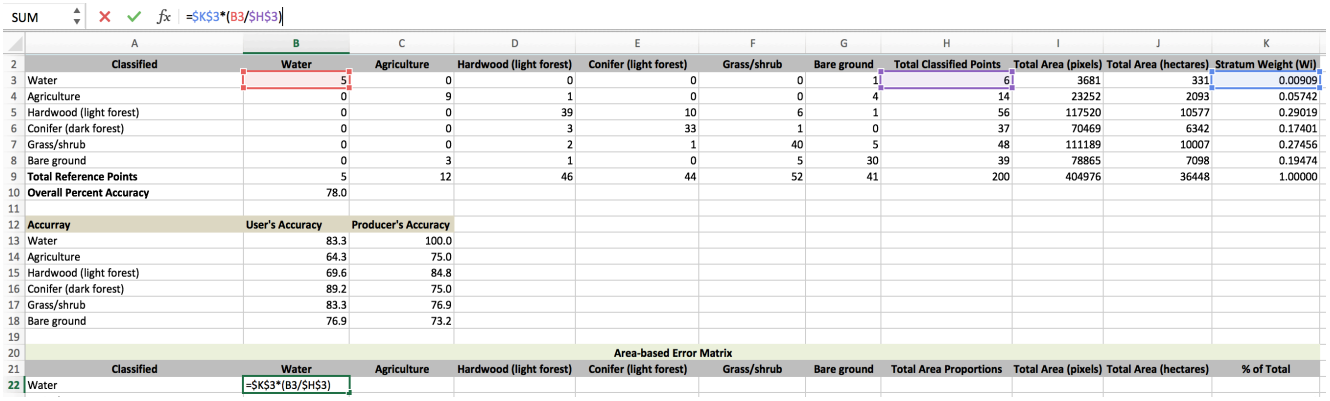
Now we have the pixel-based error matrix calculated and the area information needed to calculate our unbiased area-based error matrix.

Part 3: Create an Area-Based Error Matrix

Next, we will create an error matrix based on our area weights, which will be an unbiased error matrix. First, we will calculate the estimated area proportions for each class. The equation for the area proportion for each class is:

$$Area\ proportion = W_i \times \left(\frac{Pixels\ in\ each\ class}{Total\ pixels\ for\ each\ class}\right)$$

1. Calculate the area proportion for each class in the **Area-Based Error Matrix**. It will be very helpful for you to use the anchor ($) function when creating your equations here. For the first cell (B22) the equation will be: =$K$3*(B3/$H$3).



	A	B	C	D	E	F	G	H	I	J	K
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Classified Points**	**Total Area (pixels)**	**Total Area (hectares)**	**Stratum Weight (Wi)**
3	Water	5	0	0	0	0	1	6	3681	331	0.00909
4	Agriculture	0	9	1	0	0	4	14	23252	2093	0.05742
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	0.29019
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	0.17401
7	Grass/shrub	0	0	2	1	40	5	48	111189	10007	0.27456
8	Bare ground	0	3	1	0	5	30	39	78865	7098	0.19474
9	**Total Reference Points**	5	12	46	44	52	41	200	404976	36448	1.00000
10	**Overall Percent Accuracy**	78.0									
11											
12	**Accurray**	**User's Accuracy**	**Producer's Accuracy**								
13	Water	83.3	100.0								
14	Agriculture	64.3	75.0								
15	Hardwood (light forest)	69.6	84.8								
16	Conifer (dark forest)	89.2	75.0								
17	Grass/shrub	83.3	76.9								
18	Bare ground	76.9	73.2								
19											
20					**Area-based Error Matrix**						
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**	**Total Area (pixels)**	**Total Area (hectares)**	**% of Total**
22	Water	=$K$3*(B3/$H$3)									

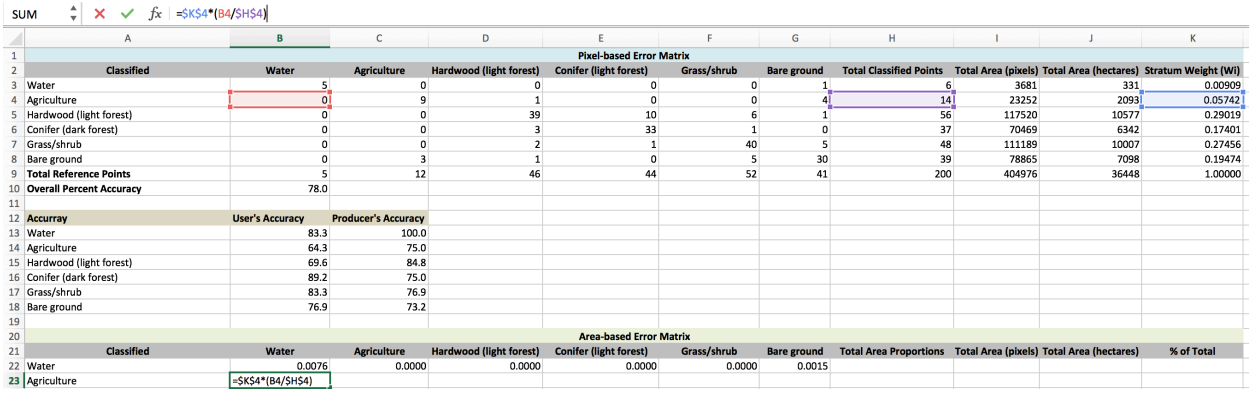
SUM =$K$3*(B3/$H$3)

2. Repeat the first step for all of the classified **Water** row (cells B22 to G22). If you used the $ sign in the equation, you can drag the bottom of cell B22 across to cell G22 and the equation should be correct. For each equation in this row, the only cell that will change is the "*pixels in each class*". For example, as you move from left to right across each row the equation will change like this:
    - Cell C22 =$K$3*(C3/$H$3)
    - Cell D22 = $K$3*(D3/$H$3)
    - Cell E22 = $K$3*(E3/$H$3)
    - Cell F22 = $K$3*(F3/$H$3)
    - Cell G22 = $K$3*(G3/$H$3)

	A	B	C	D	E	F	G	H	I	J	K
1					**Pixel-based Error Matrix**						
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Classified Points**	**Total Area (pixels)**	**Total Area (hectares)**	**Stratum Weight (Wi)**
3	Water	5	0	0	0	0	1	6	3681	331	0.00909
4	Agriculture	0	9	1	0	0	4	14	23252	2093	0.05742
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	0.29019
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	0.17401
7	Grass/shrub	0	0	2	1	40	5	48	111189	10007	0.27456
8	Bare ground	0	3	1	0	5	30	39	78865	7098	0.19474
9	**Total Reference Points**	5	12	46	44	52	41	200	404976	36448	1.00000
10	**Overall Percent Accuracy**	78.0									
11											
12	**Accurray**	**User's Accuracy**	**Producer's Accuracy**								
13	Water	83.3	100.0								
14	Agriculture	64.3	75.0								
15	Hardwood (light forest)	69.6	84.8								
16	Conifer (dark forest)	89.2	75.0								
17	Grass/shrub	83.3	76.9								
18	Bare ground	76.9	73.2								
19											
20					**Area-based Error Matrix**						
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**	**Total Area (pixels)**	**Total Area (hectares)**	**% of Total**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015				

3. Repeat steps 1 and 2 for the **Agriculture** row. For example, the equation for the first cell (B23) should look like this: =$K$4*(B4/$H$4). Again, remember to use the $ sign to save time creating each equation in each cell. For each equation in this row, the only cell that will change is the "*pixels in each class*". For example, as you move from left to right across each row the equation will change like this:

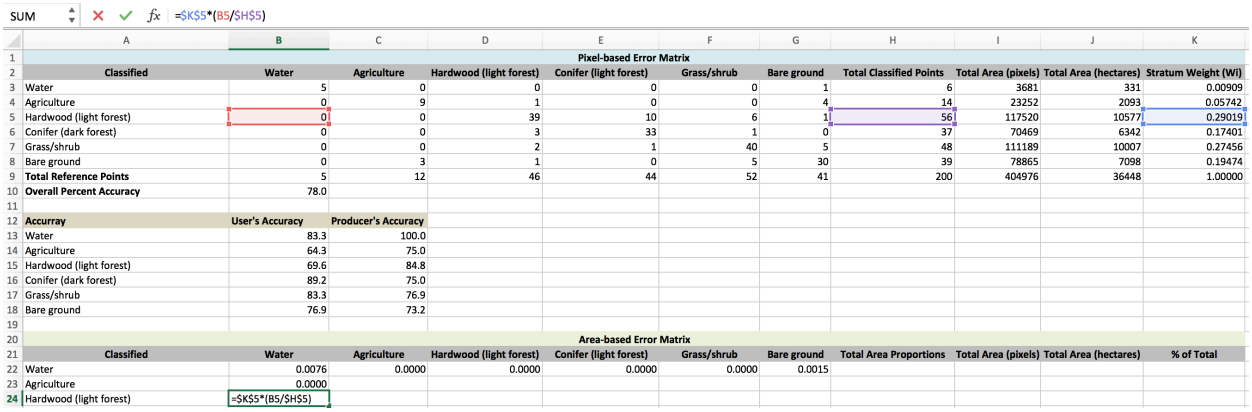
- Cell C23 =$K$4*(C4/$H$4)
- Cell D23 = $K$4*(D4/$H$4)
- Cell E23 = $K$4*(E4/$H$4)
- Cell F23 = $K$4*(F4/$H$4)
- Cell G23 = $K$4*(G4/$H$4)



SUM =$K$4*(B4/$H$4)

	A	B	C	D	E	F	G	H	I	J	K
1					Pixel-based Error Matrix						
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Classified Points**	**Total Area (pixels)**	**Total Area (hectares)**	**Stratum Weight (Wi)**
3	Water	5	0	0	0	0	1	6	3681	331	0.00909
4	Agriculture	0	9	1	0	0	4	14	23252	2093	0.05742
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	0.29019
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	0.17401
7	Grass/shrub	0	0	2	1	40	5	48	111189	10007	0.27456
8	Bare ground	0	3	1	0	5	30	39	78865	7098	0.19474
9	**Total Reference Points**	5	12	46	44	52	41	200	404976	36448	1.00000
10	**Overall Percent Accuracy**	78.0									
11											
12	**Accurray**	**User's Accuracy**	**Producer's Accuracy**								
13	Water	83.3	100.0								
14	Agriculture	64.3	75.0								
15	Hardwood (light forest)	69.6	84.8								
16	Conifer (dark forest)	89.2	75.0								
17	Grass/shrub	83.3	76.9								
18	Bare ground	76.9	73.2								
19											
20					Area-based Error Matrix						
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**	**Total Area (pixels)**	**Total Area (hectares)**	**% of Total**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015				
23	Agriculture	=$K$4*(B4/$H$4)									

4. Repeat steps 1 and 2 for the **Hardwood (light forest), Conifer (dark forest), Grass/shrub, and Bare ground** rows. For example, the equation for the first cell of the Hardwood (light forest) (B25) should look like this: =$K$5*(B5/$H$5).



SUM =$K$5*(B5/$H$5)

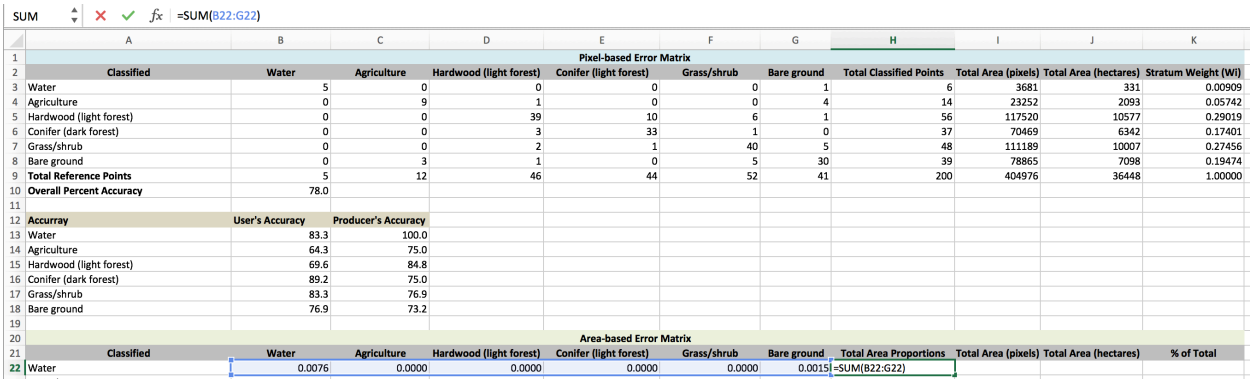
	A	B	C	D	E	F	G	H	I	J	K
1					Pixel-based Error Matrix						
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Classified Points**	**Total Area (pixels)**	**Total Area (hectares)**	**Stratum Weight (Wi)**
3	Water	5	0	0	0	0	1	6	3681	331	0.00909
4	Agriculture	0	9	1	0	0	4	14	23252	2093	0.05742
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	0.29019
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	0.17401
7	Grass/shrub	0	0	2	1	40	5	48	111189	10007	0.27456
8	Bare ground	0	3	1	0	5	30	39	78865	7098	0.19474
9	**Total Reference Points**	5	12	46	44	52	41	200	404976	36448	1.00000
10	**Overall Percent Accuracy**	78.0									
11											
12	**Accurray**	**User's Accuracy**	**Producer's Accuracy**								
13	Water	83.3	100.0								
14	Agriculture	64.3	75.0								
15	Hardwood (light forest)	69.6	84.8								
16	Conifer (dark forest)	89.2	75.0								
17	Grass/shrub	83.3	76.9								
18	Bare ground	76.9	73.2								
19											
20					Area-based Error Matrix						
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**	**Total Area (pixels)**	**Total Area (hectares)**	**% of Total**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015				
23	Agriculture	0.0000									
24	Hardwood (light forest)	=$K$5*(B5/$H$5)									

Your matrix should now have these values:

Area-based Error Matrix						
Classified	Water	Agriculture	Hardwood (light forest)	Conifer (light forest)	Grass/shrub	Bare ground
Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015
Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164
Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052
Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000
Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286
Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498

Next, we need to sum the area proportions for each class.

5. In the **Total Area Proportions** column of the **Area-Based Error Matrix**, sum the area proportions across each row and click **Enter**. For the **Water** row (cell H22), the equation will be: =SUM(B22:G22).



Pixel-based Error Matrix										
Classified	Water	Agriculture	Hardwood (light forest)	Conifer (light forest)	Grass/shrub	Bare ground	Total Classified Points	Total Area (pixels)	Total Area (hectares)	Stratum Weight (Wi)
Water	5	0	0	0	0	1	6	3681	331	0.00909
Agriculture	0	9	1	0	0	4	14	23252	2093	0.05742
Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	0.29019
Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	0.17401
Grass/shrub	0	0	2	1	40	5	48	111189	10007	0.27456
Bare ground	0	3	1	0	5	30	39	78865	7098	0.19474
Total Reference Points	5	12	46	44	52	41	200	404976	36448	1.00000
Overall Percent Accuracy	78.0									

Accurray	User's Accuracy	Producer's Accuracy
Water	83.3	100.0
Agriculture	64.3	75.0
Hardwood (light forest)	69.6	84.8
Conifer (dark forest)	89.2	75.0
Grass/shrub	83.3	76.9
Bare ground	76.9	73.2

Area-based Error Matrix										
Classified	Water	Agriculture	Hardwood (light forest)	Conifer (light forest)	Grass/shrub	Bare ground	Total Area Proportions	Total Area (pixels)	Total Area (hectares)	% of Total
Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015	=SUM(B22:G22)			

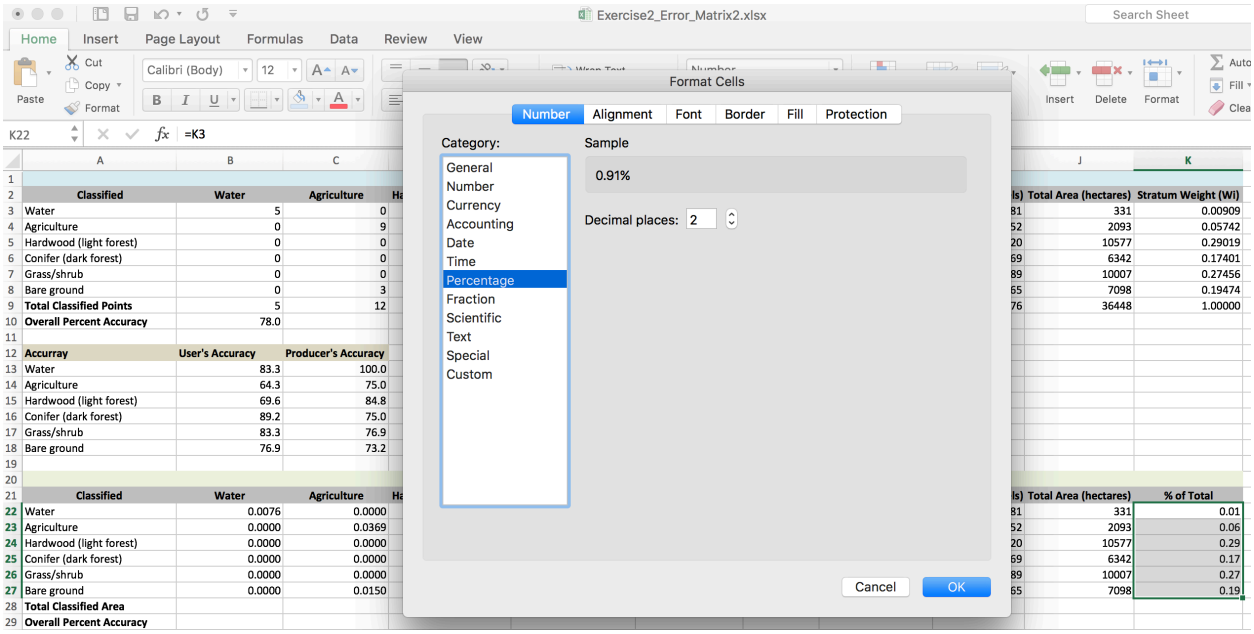
6. Repeat the previous step for each row in the **Total Area Proportions** column (to cell H27).

The next two columns for the **Area-Based Error Matrix** will be the same as the **Pixel-Based Error Matrix**.

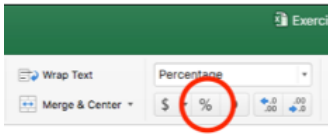
7. For the **Total Area (pixels)** column in the **Area-Based Error Matrix**, reference the same cells (values) as in the **Pixel-Based Error Matrix**. For example, the equation for cell I22 will be = I3.
8. Repeat this step for cells I23 to I29. Remember, you can drag the bottom right corner of the first cell to apply the equation to the cells below.
9. For the **Total Area (hectares)** column in the **Area-Based Error Matrix**, reference the same cells (values) as in the **Pixel-Based Error Matrix**. For example, the equation for cell J22 will be = J3.
10. Repeat this step for cells J23 to J29.

For the **% of Total** column in the **Area-Based Error Matrix**, this will be displayed as the stratum weight (W_i) in terms of percentage. However, we want the actual values to remain the same.

11. For the **% of Total** column in the **Area-Based Error Matrix**, reference the same **Stratum Weight (W_i)** cells from the **Pixel-Based Error Matrix**. The equation for cell K22 will be = K3.
12. Repeat this step for cells K23 to K29.
13. Now, display the values as a percentage by highlighting the entire column. Then, click on **Format** along the top panel of Excel, and select **Percentage**, and click **OK**.



- Alternatively, you can highlight the cells and click on the % button on the home ribbon.



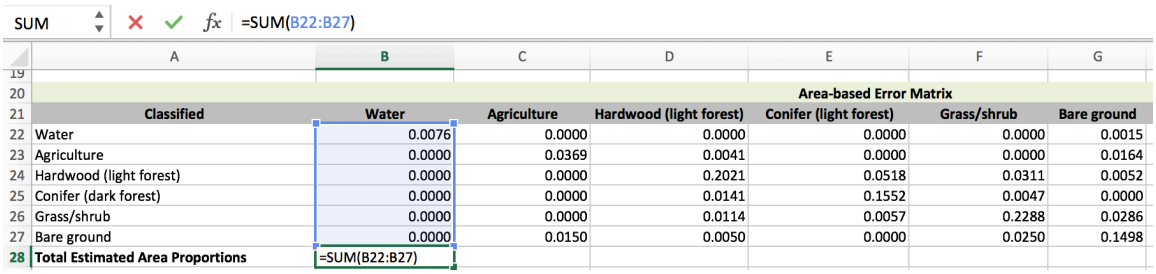
Your matrix should now have these values:

Area-based Error Matrix

Classified	Water	Agriculture	Hardwood (light forest)	Conifer (light forest)	Grass/shrub	Bare ground	Total Area Proportions	Total Area (pixels)	Total Area (hectares)	% of Total
Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015	0.0091	3681	331	0.91%
Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164	0.0574	23252	2093	5.74%
Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052	0.2902	117520	10577	29.02%
Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000	0.1740	70469	6342	17.40%
Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286	0.2746	111189	10007	27.46%
Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498	0.1947	78865	7098	19.47%

14. For the **Estimated Area Proportions** row (28), sum the cells of the **Area-Based Error Matrix** for each column. For example, the **Water** column equation will be: =SUM(B22:B27), the **Agriculture** column equation will be = SUM(C22:C27), and

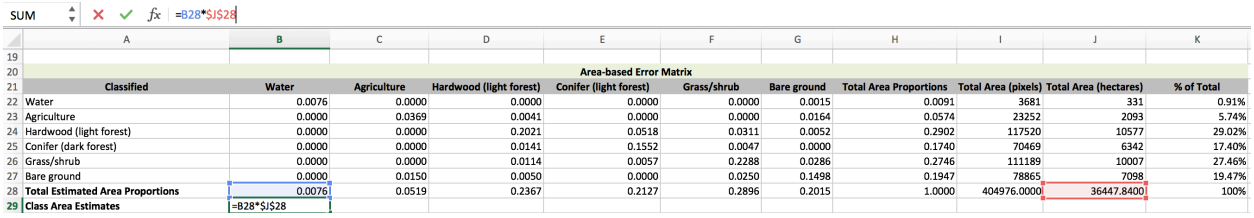
so on. Apply this SUM equation across all the columns in the **Area-based Error Matrix** table.



	A	B	C	D	E	F	G
20				Area-based Error Matrix			
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015
23	Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164
24	Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052
25	Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000
26	Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286
27	Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498
28	**Total Estimated Area Proportions**	=SUM(B22:B27)					

15. Double check to ensure that cell K28 is formatted as a percentage, as we did in step 13.

16. For the **Class Area Estimates** row (29), multiply the **Estimated Area Proportions** cell for each category by the **Total Area (hectares)**. Here is the Class Area Estimates equation:
Class Area Estimates = Estimated Area Proportion * Total Area (hectares)



	A	B	C	D	E	F	G	H	I	J	K
20					Area-based Error Matrix						
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**	**Total Area (pixels)**	**Total Area (hectares)**	**% of Total**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015	0.0091	3681	331	0.91%
23	Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164	0.0574	23252	2093	5.74%
24	Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052	0.2902	117520	10577	29.02%
25	Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000	0.1740	70469	6342	17.40%
26	Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286	0.2746	111189	10007	27.46%
27	Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498	0.1947	78865	7098	19.47%
28	**Total Estimated Area Proportions**	0.0076	0.0519	0.2367	0.2127	0.2896	0.2015	1.0000	404976.0000	36447.8400	100%
29	**Class Area Estimates**	=B28*$J$28									

For example, the **Water** column equation will be: =B28*$J$28. Notice that we used the $ anchors to keep referencing the same cell (J28) as we move across the **Area-Based Error Matrix**. Apply this equation to all the land cover classes up to column G (Bare ground).

Part 4: Calculate Standard Error

In this portion of the exercise, we will calculate the standard errors of the area estimates. These are generated for each land cover class and will allow us to calculate the error associated with each classified class and then the confidence intervals. Since we used a stratified random sampling technique, the equation is:

$$S(A\hat{}) = \sqrt{\sum_{i=1}^{3} \frac{W_i \times \hat{p}_{ij} - \hat{p}_{ij}^2}{n_{i+} - 1}}$$

This step will be tricky to get the equation correct in the spreadsheet.

1. Input the equation above into the **Standard Error of Area Estimate** row for each land cover class. For example, the **Water** column (cell B31) equation will be:

   =SQRT(($K$22*B22-B22^2)/($H$3-1)+($K$23*B23-B23^2)/($H$4-1)+($K$24*B24-B24^2)/($H$5-1)+($K$25*B25-B25^2)/($H$6-1)+($K$26*B26-B26^2)/($H$7-1)+($K$27*B27-B27^2)/($H$8-1))

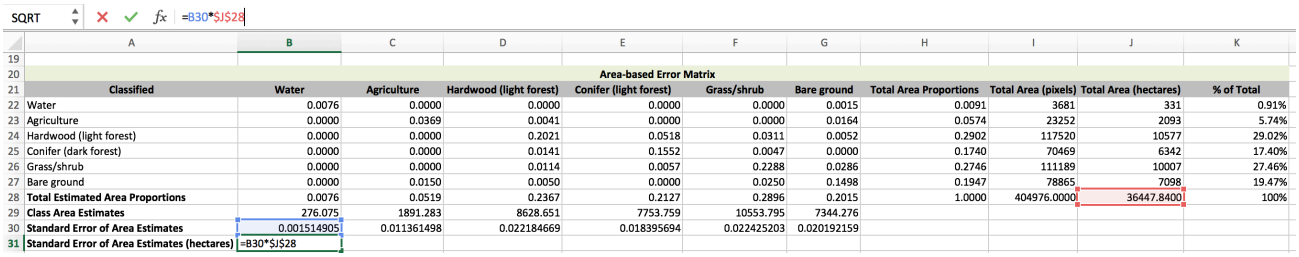
You will notice a pattern as you input this equation into Excel. Please double check it before clicking Enter.

SUM =SQRT(($K$22*B22-B22^2)/($H$3-1)+($K$23*B23-B23^2)/($H$4-1)+($K$24*B24-B24^2)/($H$5-1)+($K$25*B25-B25^2)/($H$6-1)+($K$26*B26-B26^2)/($H$7-1)+($K$27*B27-B27^2)/($H$8-1))

	A	B	C	D	E	F	G	H	I	J	K
1					Pixel-based Error Matrix						
2	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Classified Points**	**Total Area (pixels)**	**Total Area (hectares)**	**Stratum Weight (Wi)**
3	Water	5	0	0	0	0	1	6	3681	331	0.00909
4	Agriculture	0	9	1	0	0	4	14	23252	2093	0.05742
5	Hardwood (light forest)	0	0	39	10	6	1	56	117520	10577	0.29019
6	Conifer (dark forest)	0	0	3	33	1	0	37	70469	6342	0.17401
7	Grass/shrub	0	0	2	1	40	5	48	111189	10007	0.27456
8	Bare ground	0	3	1	0	5	30	39	78865	7098	0.19474
9	**Total Reference Points**	5	12	46	44	52	41	200	404976	36448	1.00000
10	**Overall Percent Accuracy**	78.0									
11											
12	**Accurray**	**User's Accuracy**	**Producer's Accuracy**								
13	Water	83.3	100.0								
14	Agriculture	64.3	75.0								
15	Hardwood (light forest)	69.6	84.8								
16	Conifer (dark forest)	89.2	75.0								
17	Grass/shrub	83.3	76.9								
18	Bare ground	76.9	73.2								
19											
20					Area-based Error Matrix						
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**	**Total Area (pixels)**	**Total Area (hectares)**	**% of Total**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015	0.0091	3681	331	0.91%
23	Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164	0.0574	23252	2093	5.74%
24	Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052	0.2902	117520	10577	29.02%
25	Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000	0.1740	70469	6342	17.40%
26	Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286	0.2746	111189	10007	27.46%
27	Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498	0.1947	78865	7098	19.47%
28	**Total Estimated Area Proportions**	0.0076	0.0519	0.2367	0.2127	0.2896	0.2015	1.0000	404976.0000	36447.8400	100%
29	**Class Area Estimates**	276.075	1891.283	8628.651	7753.759	10553.795	7344.276				
30	**Standard Error of Area Estimates**	$K$23*B23-B23^2)/(									

2. Repeat this step for each land cover class in the **Standard Error of Area Estimates** row. Notice that if you used the anchor ($) for all the correct cells when you input the equation initially, you can drag the bottom right corner of cell B30 across to cell G30. This will save time as you will not need to type in the long equation for each class.
3. Calculate the **Standard Error of Area Estimates (hectares)**. Multiply the **Standard Error of Area Estimates** by the **Total Area (hectares)** for the entire

classified area. For example, the equation for the **Water** column (B30) will be: =B30*$J$28.



	A	B	C	D	E	F	G	H	I	J	K
19											
20					Area-based Error Matrix						
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**	**Total Area (pixels)**	**Total Area (hectares)**	**% of Total**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015	0.0091	3681	331	0.91%
23	Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164	0.0574	23252	2093	5.74%
24	Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052	0.2902	117520	10577	29.02%
25	Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000	0.1740	70469	6342	17.40%
26	Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286	0.2746	111189	10007	27.46%
27	Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498	0.1947	78865	7098	19.47%
28	**Total Estimated Area Proportions**	0.0076	0.0519	0.2367	0.2127	0.2896	0.2015	1.0000	404976.0000	36447.8400	100%
29	**Class Area Estimates**	276.075	1891.283	8628.651	7753.759	10553.795	7344.276				
30	**Standard Error of Area Estimates**	0.001514905	0.011361498	0.022184669	0.018395694	0.022425203	0.020192159				
31	**Standard Error of Area Estimates (hectares)**	=B30*$J$28									

4. Repeat step 3 for all the land cover classes. Again, you can drag the bottom right corner of cell B30 across if you used the anchor ($) for the total area (hectares).

Next, we will calculate the confidence intervals (CI) based on 95% confidence. As noted in the lecture, the standard CI at 95% is 1.96. Therefore, we will multiply 1.96 by the **Standard Error of Area Estimates (hectares)**.

5. Calculate the 95% CI in hectares. Multiply the **Standard Error of Area Estimates (hectares)** for each land cover class by 1.96. For example, the equation for the Water column (B31) will be: = B31*1.96.

Classified	**Water**
Water	0.0076
Agriculture	0.0000
Hardwood (light forest)	0.0000
Conifer (dark forest)	0.0000
Grass/shrub	0.0000
Bare ground	0.0000
Total Estimated Area Proportions	0.0076
Class Area Estimates	276.075
Standard Error of Area Estimates	0.001514905
Standard Error of Area Estimates (hectares)	55.215
95% Confidence Interval in hectares	=B31*1.96

6. Repeat the previous step for all the land cover classes.

Your answers should look like this:

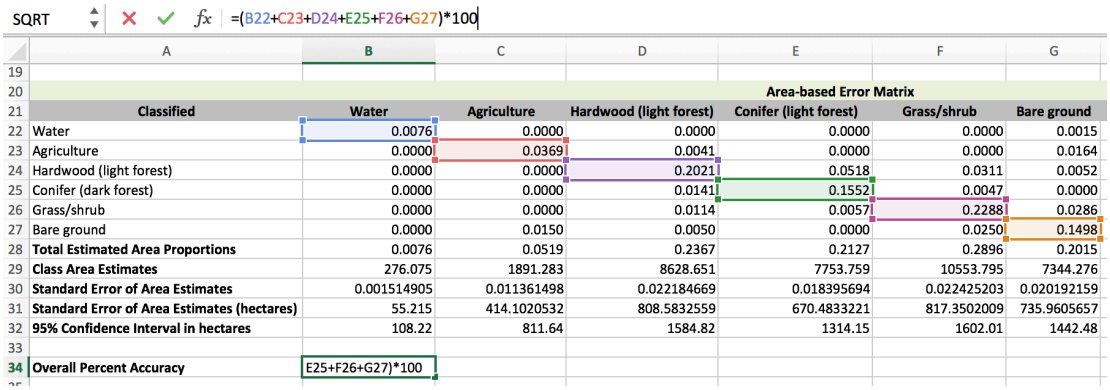
Total Estimated Area Proportions	0.0076	0.0519	0.2367	0.2127	0.2896	0.2015
Class Area Estimates	276.075	1891.283	8628.651	7753.759	10553.795	7344.276
Standard Error of Area Estimates	0.001514905	0.011361498	0.022184669	0.018395694	0.022425203	0.020192159
Standard Error of Area Estimates (hectares)	55.215	414.1020532	808.5832559	670.4833221	817.3502009	735.9605657
95% Confidence Interval in hectares	108.22	811.64	1584.82	1314.15	1602.01	1442.48

Next we will add the **Overall Percent Accuracy** using our **Area-Based error matrix**.

Next to the **Overall Percent Accuracy** cell (B34), sum the diagonal cells of the error matrix, divide by the total cells, and multiply by 100. This is similar to the steps we completed in Exercise 1. The equation will look like this: =(B22+C23+D24+E25+F26+G27) * 100.

- Note that you do not need to divide by the total here, since that will always be 1.

Percent Accuracy = [Total correct reference points] * 100



SQRT =(B22+C23+D24+E25+F26+G27)*100

	A	B	C	D	E	F	G
19							
20				Area-based Error Matrix			
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015
23	Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164
24	Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052
25	Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000
26	Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286
27	Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498
28	**Total Estimated Area Proportions**	0.0076	0.0519	0.2367	0.2127	0.2896	0.2015
29	**Class Area Estimates**	276.075	1891.283	8628.651	7753.759	10553.795	7344.276
30	**Standard Error of Area Estimates**	0.001514905	0.011361498	0.022184669	0.018395694	0.022425203	0.020192159
31	**Standard Error of Area Estimates (hectares)**	55.215	414.1020532	808.5832559	670.4833221	817.3502009	735.9605657
32	**95% Confidence Interval in hectares**	108.22	811.64	1584.82	1314.15	1602.01	1442.48
33							
34	**Overall Percent Accuracy**	E25+F26+G27)*100					

Part 5: Calculate Unbiased User's and Producer's Accuracy

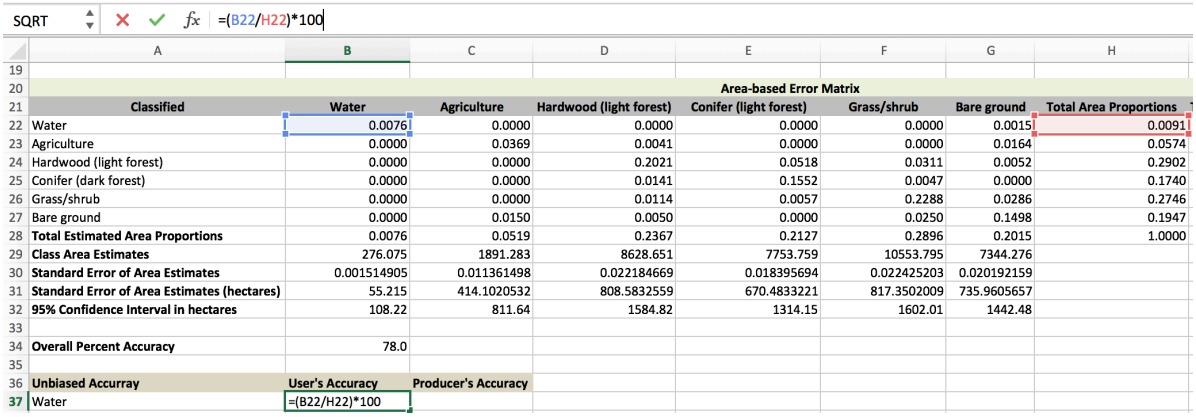
Remember, errors of commission (user's accuracy) occur when a pixel is incorrectly included in a category being evaluated, while errors of omission (producer's accuracy) occur when a pixel is left out of the category being evaluated. While the user's accuracy will not change, the producer's accuracy will be different based on the area estimates. However, it is best practice to estimate both types of accuracy based on the area-based calculations.

Let's start with the user's accuracy.

1. 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. For example, the equation for the **Water** row (B37) will be: =(B22/H22)*100

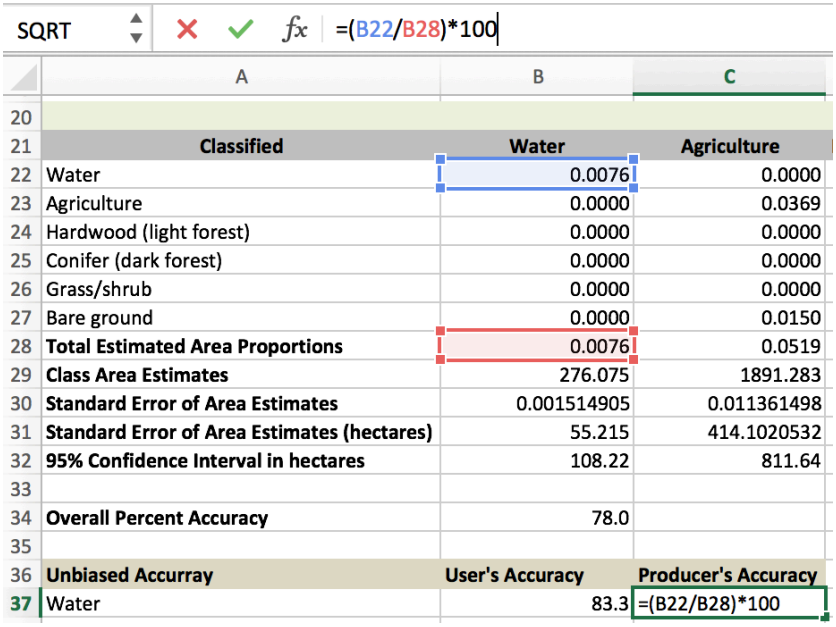
2. Repeat this step for each land cover class for the **User's Accuracy**.

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



	A	B	C	D	E	F	G	H
19								
20					Area-based Error Matrix			
21	**Classified**	**Water**	**Agriculture**	**Hardwood (light forest)**	**Conifer (light forest)**	**Grass/shrub**	**Bare ground**	**Total Area Proportions**
22	Water	0.0076	0.0000	0.0000	0.0000	0.0000	0.0015	0.0091
23	Agriculture	0.0000	0.0369	0.0041	0.0000	0.0000	0.0164	0.0574
24	Hardwood (light forest)	0.0000	0.0000	0.2021	0.0518	0.0311	0.0052	0.2902
25	Conifer (dark forest)	0.0000	0.0000	0.0141	0.1552	0.0047	0.0000	0.1740
26	Grass/shrub	0.0000	0.0000	0.0114	0.0057	0.2288	0.0286	0.2746
27	Bare ground	0.0000	0.0150	0.0050	0.0000	0.0250	0.1498	0.1947
28	**Total Estimated Area Proportions**	0.0076	0.0519	0.2367	0.2127	0.2896	0.2015	1.0000
29	**Class Area Estimates**	276.075	1891.283	8628.651	7753.759	10553.795	7344.276	
30	**Standard Error of Area Estimates**	0.001514905	0.011361498	0.022184669	0.018395694	0.022425203	0.020192159	
31	**Standard Error of Area Estimates (hectares)**	55.215	414.1020532	808.5832559	670.4833221	817.3502009	735.9605657	
32	**95% Confidence Interval in hectares**	108.22	811.64	1584.82	1314.15	1602.01	1442.48	
33								
34	**Overall Percent Accuracy**	78.0						
35								
36	**Unbiased Accurray**	**User's Accuracy**	**Producer's Accuracy**					
37	Water	=(B22/H22)*100						

3. 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. For example, the equation for the **Water** row (C37) will be: =(B22/B28)*100.



	A	B	C
20			
21	**Classified**	**Water**	**Agriculture**
22	Water	0.0076	0.0000
23	Agriculture	0.0000	0.0369
24	Hardwood (light forest)	0.0000	0.0000
25	Conifer (dark forest)	0.0000	0.0000
26	Grass/shrub	0.0000	0.0000
27	Bare ground	0.0000	0.0150
28	**Total Estimated Area Proportions**	0.0076	0.0519
29	**Class Area Estimates**	276.075	1891.283
30	**Standard Error of Area Estimates**	0.001514905	0.011361498
31	**Standard Error of Area Estimates (hectares)**	55.215	414.1020532
32	**95% Confidence Interval in hectares**	108.22	811.64
33			
34	**Overall Percent Accuracy**	78.0	
35			
36	**Unbiased Accurray**	**User's Accuracy**	**Producer's Accuracy**
37	Water	83.3	=(B22/B28)*100

4. Repeat this step for each land cover class for the **Producer's Accuracy**.

Now you will notice some slight differences in the producer's accuracy of the different land cover classes between the error matrix we did last week and the unbiased

assessment we just completed this week. This is an unbiased accuracy and is based on the area estimates of each land cover class.

Error Matrix from Session 1

Accurray	User's Accuracy	Producer's Accuracy
Water	83.3	100.0
Agriculture	64.3	75.0
Hardwood (light forest)	69.6	84.8
Conifer (dark forest)	89.2	75.0
Grass/shrub	83.3	76.9
Bare ground	76.9	73.2

Unbiased Error Matrix

Unbiased Accurray	User's Accuracy	Producer's Accuracy
Water	83.3	100.0
Agriculture	64.3	71.1
Hardwood (light forest)	69.6	85.4
Conifer (dark forest)	89.2	73.0
Grass/shrub	83.3	79.0
Bare ground	76.9	74.3

Summary Table of Unbiased Accuracy Assessment

You may now notice that the summary table at the bottom of your Excel has been compiled automatically. This table references the cells of the previous calculations for **Class Area Estimates, ± 95% CI**, the **User's Accuracy**, the **Producer's Accuracy**, and the **Overall Accuracy**. These are the standard metrics that are reported when you conduct a land cover classification.

Unbiased Accuracy Assessment Summary Table

Class	Area (hectares)	± 95% CI	User's Accuracy (%)	Producer's Accuracy (%)	Overall Accuracy (%)
Water	276.08	108.22	83.3	100.0	78.0
Agriculture	1,891.28	811.64	64.3	71.1	
Hardwood (light forest)	8,628.65	1,584.82	69.6	85.4	
Conifer (dark forest)	7,753.76	1,314.15	89.2	73.0	
Grass/shrub	10,553.80	1,602.01	83.3	79.0	
Bare ground	7,344.28	1,442.48	76.9	74.3	

5. As a final step, **Save** your Excel spreadsheet for future reference.

Conclusion

Unbiased accuracy assessment is key to reporting meaningful results for your land cover classification. In this exercise, you gained an understanding of how to calculate an area-based error matrix, and how to incorporate standard errors based on the area of each identified land cover class. You also included a typical confidence interval (95%) to your error estimates, which also aids in effective reporting. Now that you have completed these steps on a pre-identified land cover classification, you can apply these same techniques to your future classifications of your region of interest.

Additional Resources

Olofsson, P. et al., 2013, Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation, Remote Sensing of Environment, 129, 122-131, DOI: 10.1016/j.rse.2012.10.031.