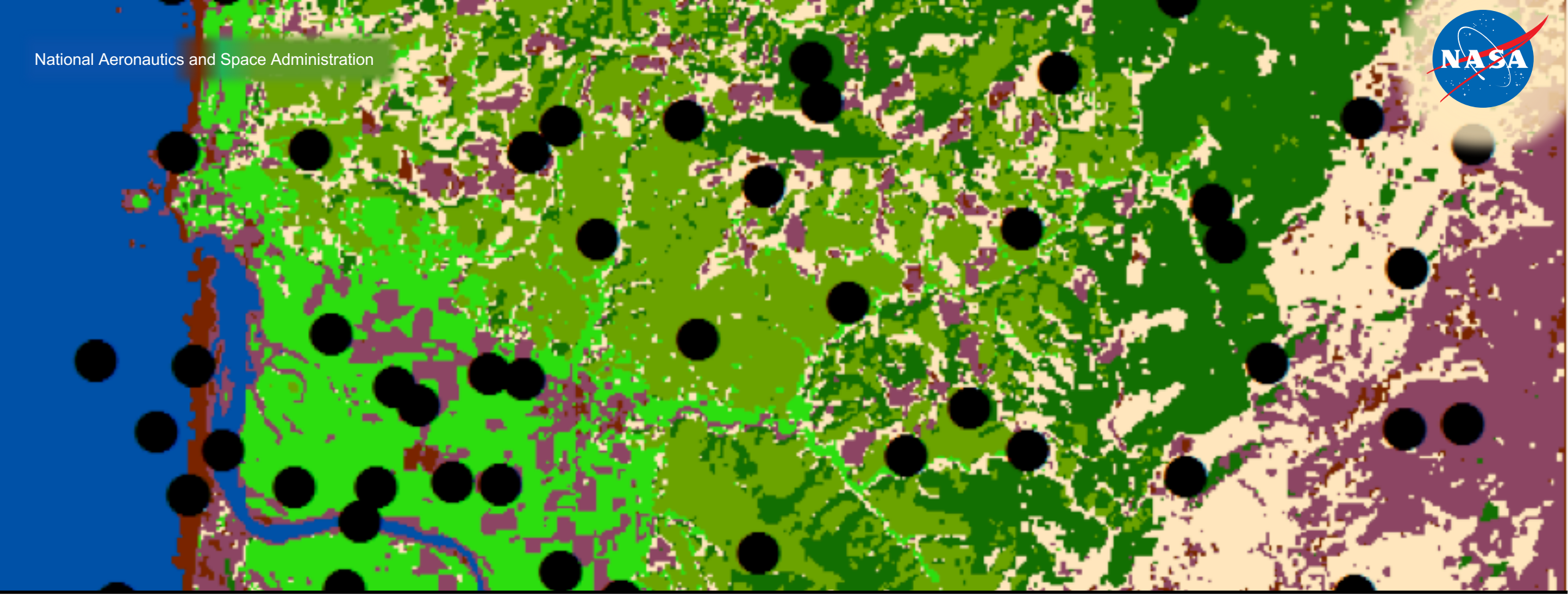
Welcome to Assessing the Accuracy of Land Cover Classifications

- **We will begin promptly at 11:00 EDT (UTC-5)**

Course Format:

- Two, two hour sessions
- Sessions will be held Tuesdays, February 13 and February 20, 2018
- We will have two sessions to reach our international audience. Please only sign up for and attend one session each week.
  - Session times: 11:00 - 13:00 EST and 23:00 - 01:00 EST (UTC-5)
- All attendees will be muted automatically upon entry
- This session will be recorded and made available to you within two days

Please be sure you have reviewed the content within Session 1 of *Fundamentals of Remote Sensing*: http://arset.gsfc.nasa.gov/webinars/fundamentals-remote-sensing



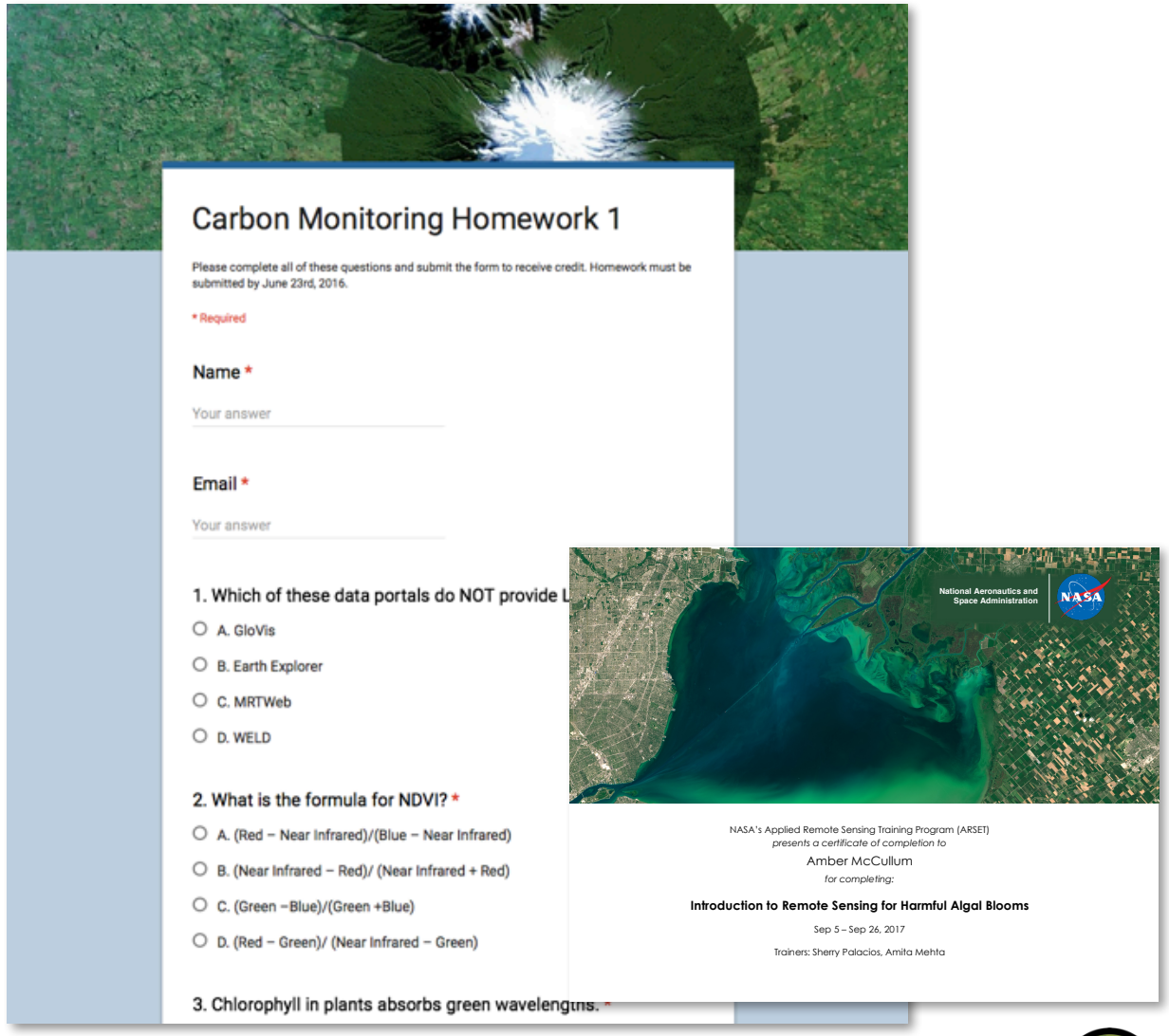
Assessing the Accuracy of Land Cover Classifications
Week 2: Unbiased Area Estimation

Cindy Schmidt and Amber McCullum

Week 2: February 20, 2018

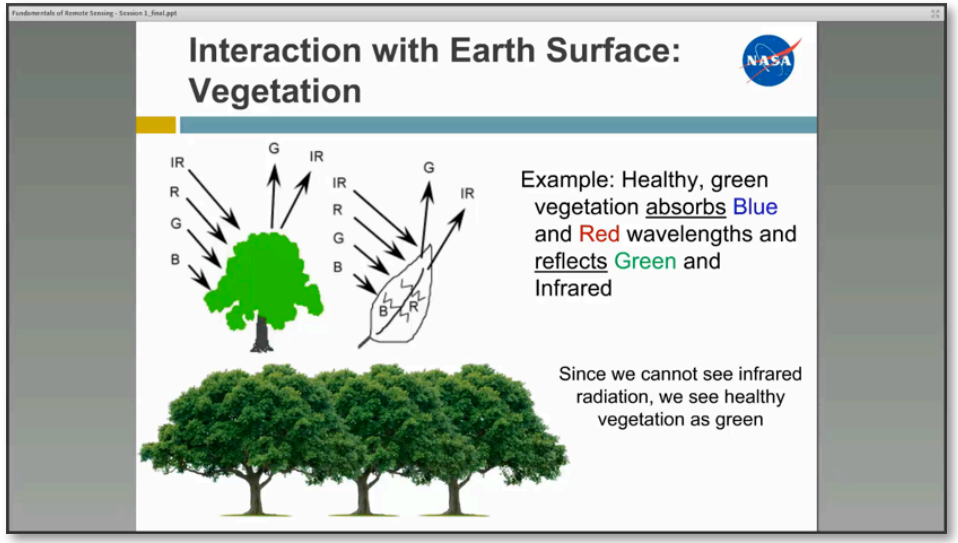
Homework and Certificates

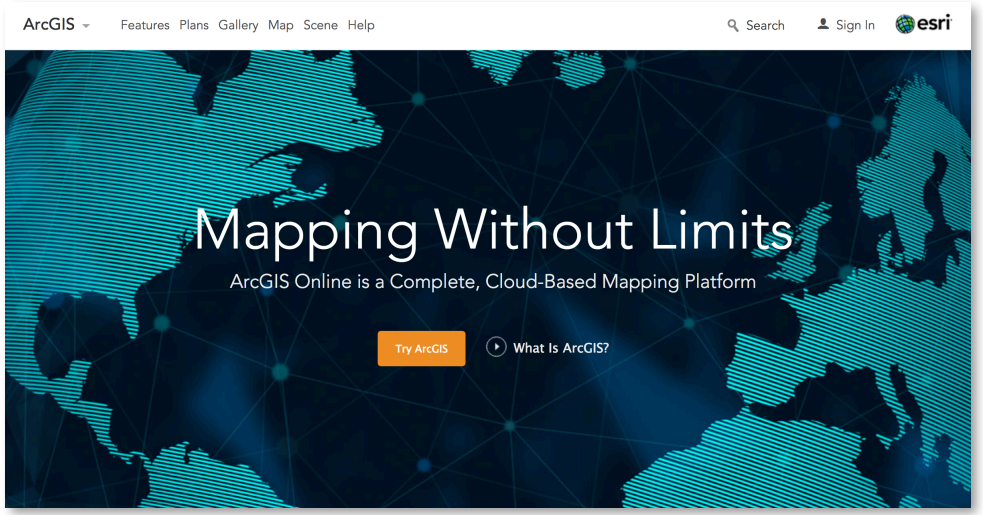
- Homework
  - One homework assignment
  - Answers must be submitted via Google Forms
- Certificate of Completion:
  - Attend both live webinars
  - Complete the homework assignment by the deadline (access from ARSET website)
    - **HW Deadline: March 6th**
  - You will receive certificates approx. two months after the completion of the course from: marines.martins@ssaihq.com



Prerequisites

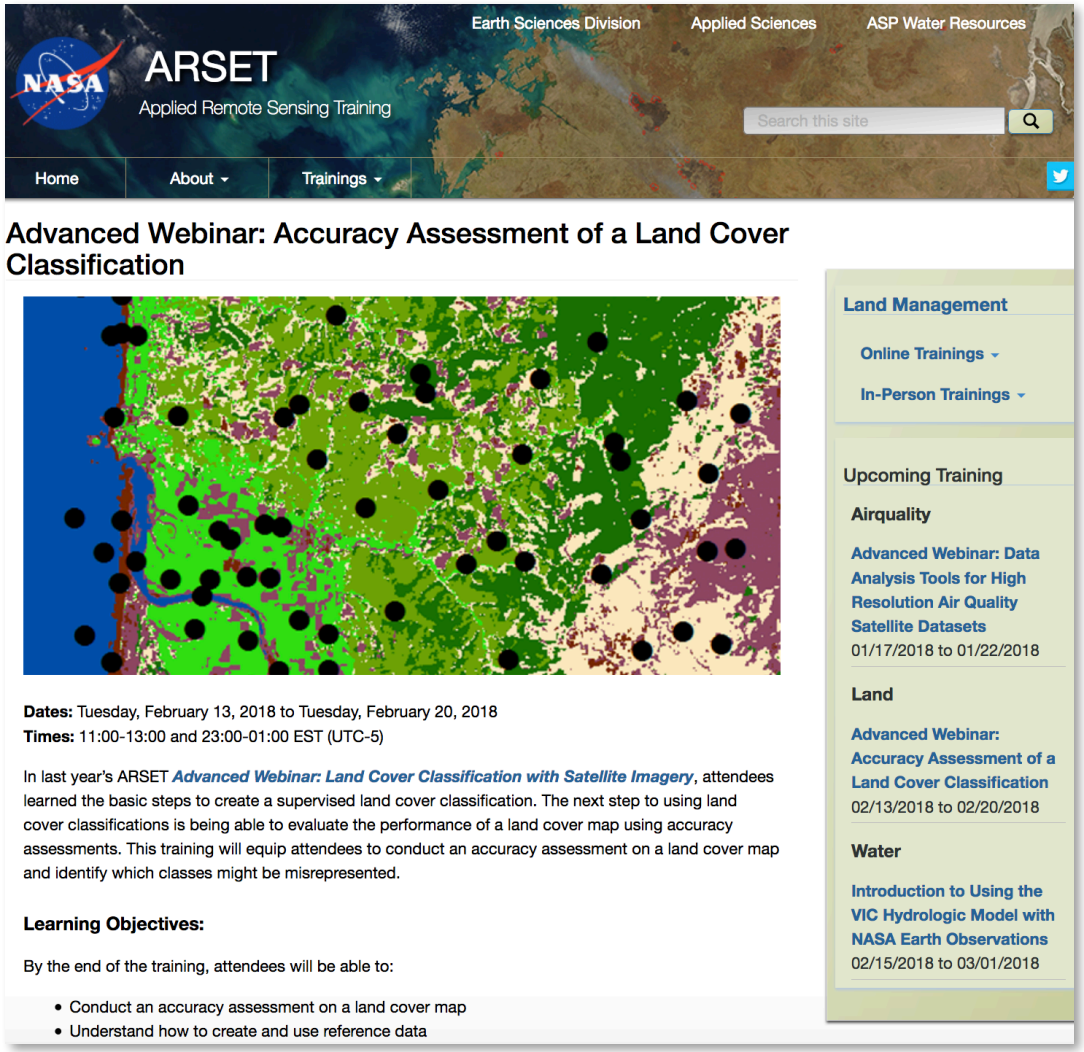
- Complete Sessions 1 & 2A of Fundamentals of Remote Sensing, or equivalent experience
  - https://arset.gsfc.nasa.gov/webinars/fundamentals-remote-sensing
- Complete the Advanced Webinar: Land Cover Classification with Satellite Imagery
  - https://arset.gsfc.nasa.gov/land/webinars/advanced-land-classification
- Download and install ArcGIS 10.4 or higher and all accompanying software
- Download and install Excel 2011 or higher

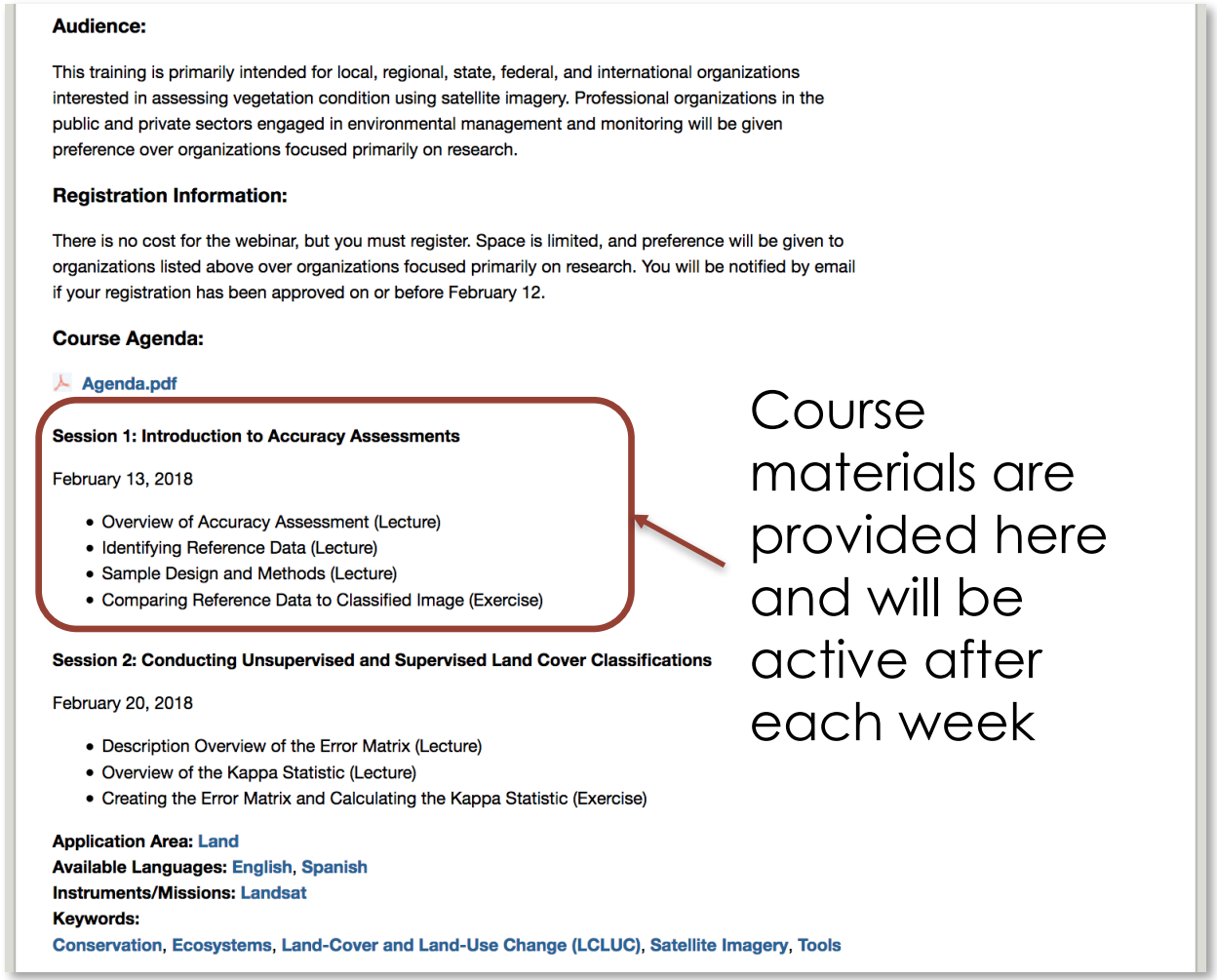




Accessing Course Materials

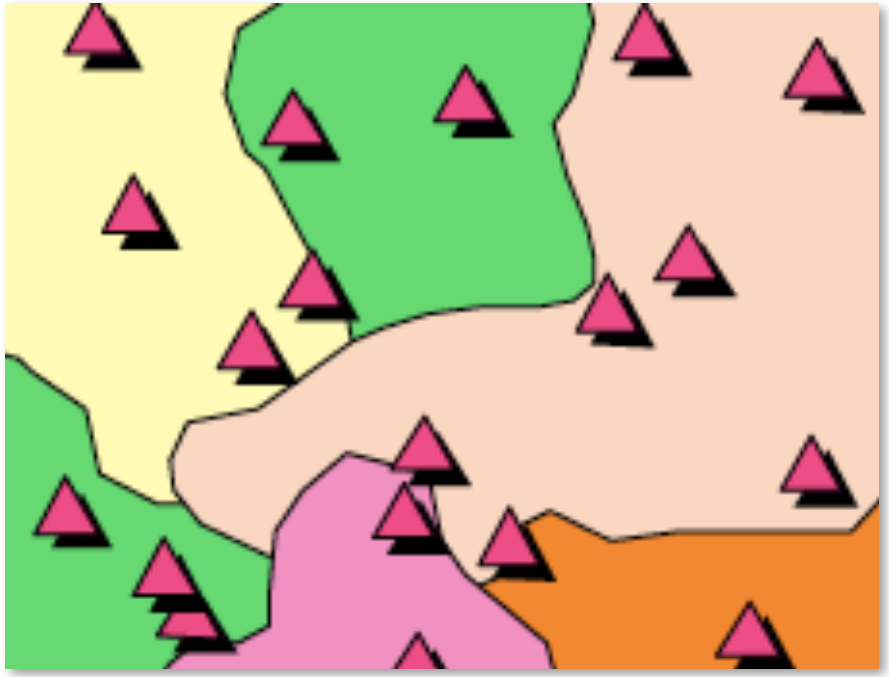
https://arset.gsfc.nasa.gov/land/webinars/18adv-land-classification



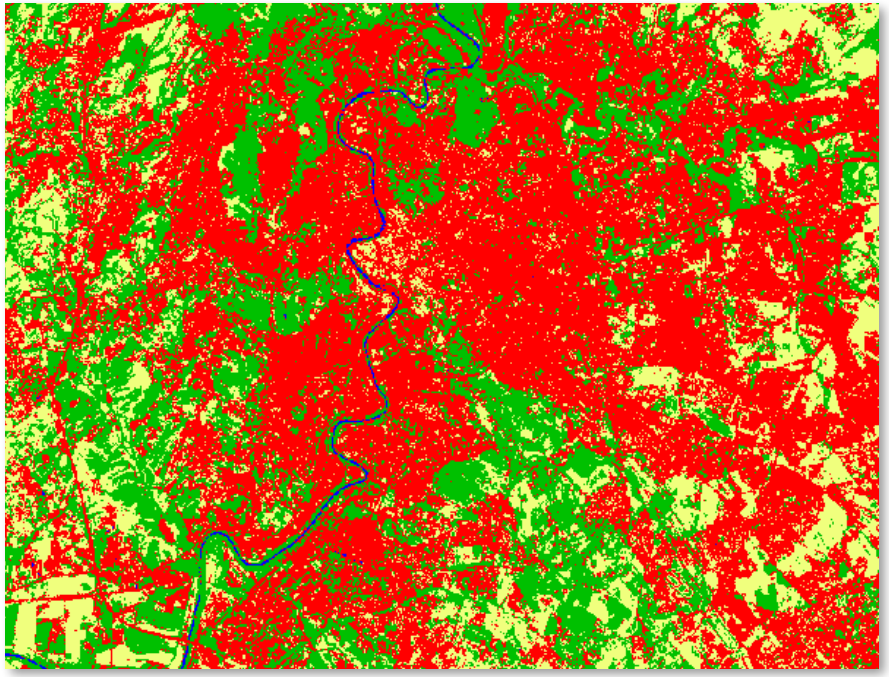


Course materials are provided here and will be active after each week

Course Outline



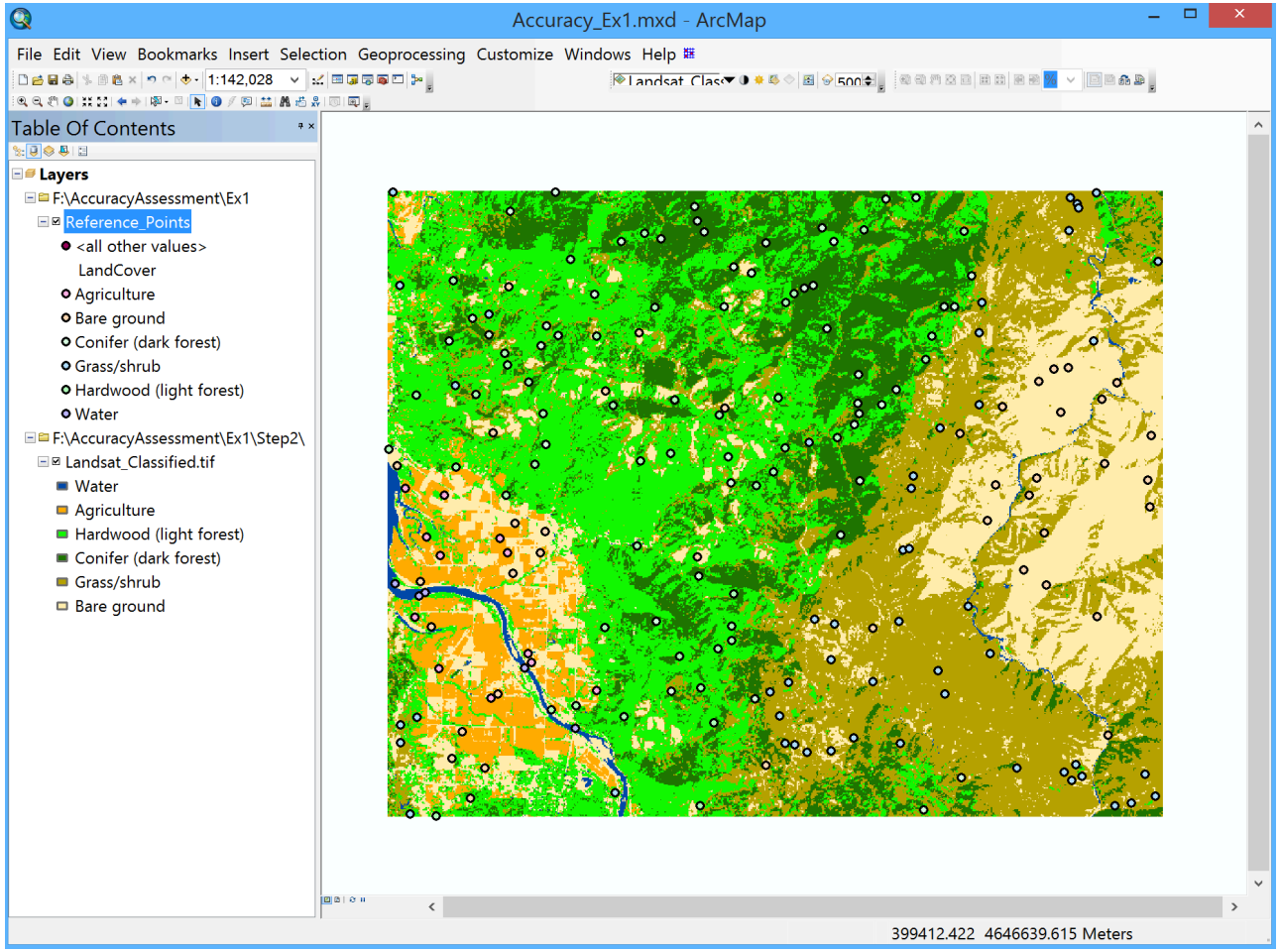
Session 1: Accuracy Assessment Basics



Session 2: Area Estimation

Session 2 Agenda

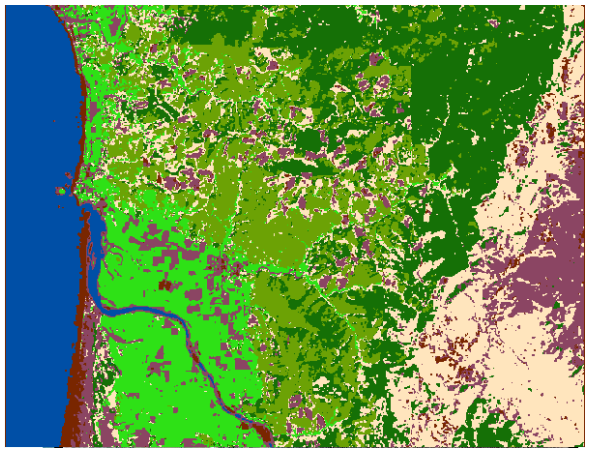
- Overview of Unbiased Area Estimation
- Error Matrix Review
- Estimated Area Calculations
  - Specific to class designations
- Standard Error Calculations
- Unbiased User and Producer Accuracy
- Error Matrix Calculations
- Exercise 2
- Q&A Session



Classified image and reference points in ArcMap

Unbiased Area Estimation

- Error matrices provide information about the accuracy of a classified map based on reference data
- Error matrices can also provide information about area estimates of each class (strata) adjusted for the map bias and characterize uncertainty
  - If the reference data were collected using a statistical approach (i.e. stratified random sample), then those data are considered unbiased
  - The classified map is considered biased
  - The adjusted area estimates require both map data and reference data



Land cover map:

Hardwood	45,000 ha
Conifer	36,000 ha
Other	18,000 ha

These estimates are biased

Review: Creating an Error Matrix

Landcover Classes

1 – Hardwood
2 – Conifer
3 - Other

Landsat Land Cover	Reference Point	Frequency
1	1	24
1	2	3
2	2	30
2	3	1
3	3	23
3	4	5

Review: Creating an Error Matrix

Landcover Classes
1 – Hardwood
2 – Conifer
3 - Other

Landsat Land Cover	Reference Point	Frequency
1	1	24
1	2	10
2	2	30
2	3	2
3	3	23
3	2	1

Correctly classified

Review: Error Matrix

- Compares the reference data to the classified map in terms of pixel counts
- Here is an example of an Error Matrix using three classes:
  - hardwoods, conifers, and other

Classification Data	Reference Data: Hardwood	Reference Data: Conifer	Reference Data: Other	Reference Data: Total
Hardwood	24	10	4	38
Conifer	5	30	2	37
Other	1	1	23	25
Total	30	41	29	100

New Error Matrix

Additions to the error matrix:

	Hardwood	Conifer	Other	Total	Total Pixels	Total Area (ha)
Hardwood	24	10	4	38	500,000	45,000
Conifer	5	30	2	37	400,000	36,000
Other	1	1	23	25	200,000	18,000
Total	**30**	**41**	**29**	**100**	**1,100,000**	**99,000**

Total Pixels: total number of pixels in each land cover class

Total Area (ha): $\frac{\text{(total pixels x (30 x 30))}}{\text{(100 x 100)}}$

Step 1: Calculate Area Proportion (W_i)

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	24	10	4	38	500,000	45,000	0.455
Conifer	5	30	2	37	400,000	36,000	0.364
Other	1	1	23	25	200,000	18,000	0.182
Total	**30**	**41**	**29**	**100**	**1,100,000**	**99,000**	**1.000**

$$W_i = \frac{\text{Total Area (ha) for each class}}{\text{Total Area (ha)}}$$

$$\frac{45{,}000}{99{,}000} = 0.455$$

Step 2: Calculate Estimated Area Proportions

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	24	10	4	38	500,000	45,000	0.455
Conifer	5	30	2	37	400,000	36,000	0.364
Other	1	1	23	25	200,000	18,000	0.182
Total	**30**	**41**	**29**	**100**	**1,100,000**	**99,000**	**1.000**

$$\text{Area proportion estimates for each class} = W_i * \frac{\textit{pixels in each class}}{\textit{total pixels for each class}}$$

$$= 0.455 * \frac{24}{38} = 0.2871$$

Step 2: Calculate Estimated Area Proportions

Old Error Matrix

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	24	10	4	38	500,000	45,000	0.455
Conifer	5	30	2	37	400,000	36,000	0.364
Other	1	1	23	25	200,000	18,000	0.182
Total	**30**	**41**	**29**	**100**	**1,100,000**	**99,000**	**1.000**

New Error Matrix

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871				500,000	45,000	0.455
Conifer					400,000	36,000	0.364
Other					200,000	18,000	0.182
Total					**1,100,000**	**99,000**	**1.000**

New Error Matrix

We get a new error matrix with estimated area proportions for each class

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	**0.3435**	**0.4217**	**0.2348**	**1.000**	**1,100,000**	**99,000**	**1.000**

Step 3: Calculate Class Area Estimates (A^)

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
A^ (ha)	**34,007**						

Area estimate (A^) = 0.3435 * 99,000 = 34,007

Step 3: Calculate Class Area Estimates (A^)

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
A^ (ha)	**34,007**	**41,751**	**23,243**				

Step 4: Calculate Standard Error (S(A^))

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
A^ (ha)	34,007	41,751	23,243				
S (A^)							

$$\sqrt{\sum_{i=1}^{3} \frac{W_i \text{ x } \hat{p}_{ij} - \hat{p}_{ij}^{\,2}}{n_{i+} - 1}}$$

For Hardwood:

W_i = 0.455 0.364 0.182

P_i = 0.2871 0.0491 0.0073

Step 4: Calculate Standard Error (S(A^))

For n_i we need to go back to our first error matrix

	Hardwood	Conifer	Other	Total	Total Pixels	Total Area (ha)
Hardwood	24	10	4	38	500,000	45,000
Conifer	5	30	2	37	400,000	36,000
Other	1	1	23	25	200,000	18,000
Total	**30**	**41**	**29**	**100**	**1,100,000**	**99,000**

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

n_i = the total number of classified pixels in each class

n_i = 38 pixels
37 pixels
25 pixels

Step 4: Calculate Standard Error (S(A^))

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
A^ (ha)	34,007	41,751	23,243				
S (A^)	**0.0422**	**0.0412**	**0.0286**				

Step 5: Convert Standard Error to Hectares (S(A^)(ha))

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
A^ (ha)	34,007	41,751	23,243				
S (A^)	0.0422	0.0412	0.0286				
S (A^) (ha)	**4,179**						

$$\frac{(0.0422 * 1{,}100{,}000) * 30^2}{100^2} = 4{,}179$$

Step 5: Convert Standard Error to Hectares (S(A^)(ha))

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
A^ (ha)	34,007	41,751	23,243				
S (A^)	0.0422	0.0412	0.0286				
S (A^) (ha)	**4,179**	**4,081**	**2,326**				

Step 6: Calculate Confidence Intervals

A 95% CI is 1.96 * S(A^)(ha)

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
A^ (ha)	34,007	41,751	23,243				
S (A^)	0.0422	0.0412	0.0286				
S (A^) (ha)	4,179	4,081	2,326				
95% CI (ha)	**8,190**	**7,998**	**5,540**				

Step 7: Calculate User's, Producer's, and Overall Accuracy

User's Accuracy

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
UA	0.632						
PA							
Overall							

$$\frac{0.2871}{0.455} = 0.632$$

Step 7: Calculate User's, Producer's, and Overall Accuracy

Producer's Accuracy

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
UA	0.632						
PA	0.836						
Overall							

$$\frac{0.2871}{0.3435} = 0.836$$

Step 7: Calculate User's, Producer's, and Overall Accuracy

User's and Producer's Accuracy

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
UA	0.632	0.811	0.920				
PA	0.836	0.699	0.712				
Overall							

Step 7: Calculate User's, Producer's, and Overall Accuracy

Overall Accuracy

	Hardwood	Conifer	Other	Total	Total pixels	Total Area (ha)	W_i
Hardwood	0.2871	0.1196	0.0478	0.455	500,000	45,000	0.455
Conifer	0.0491	0.2948	0.0197	0.364	400,000	36,000	0.364
Other	0.0073	0.0073	0.1673	0.182	200,000	18,000	0.182
Total	0.3435	0.4217	0.2348	1.000	1,100,000	99,000	1.000
UA	0.632	0.811	0.920				
PA	0.836	0.699	0.712				
Overall				**0.749**			

$$0.2871 + 0.2948 + 0.1673 = 0.749$$

Results

These are the unbiased area estimates for each class, including a 95% confidence interval based on both the reference data and the land cover classification

Class	Area (ha)	+/- 95% CI	User's	Producers	Overall
Hardwood	34,007	8,190	0.63	0.84	0.75
Conifer	41,751	7,998	0.81	0.70	
Other	23,243	5,540	0.92	0.71	

Error Matrix Comparisons

Unbiased Area Estimators

Class	Area (ha)	+/- 95% CI	User's	Producers	Overall
Hardwood	34,007	8,190	0.63	0.84	0.75
Conifer	41,751	7,998	0.81	0.70	
Other	23,243	5,540	0.92	0.71	

Error Matrix Using Sample Counts

Class	Area (ha)	User's	Producers	Overall
Hardwood	45,000	0.63	0.80	0.77
Conifer	36,000	0.81	0.73	
Other	18,000	0.92	0.79	

Error Matrix Comparisons

Unbiased Area Estimators

Class	Area (ha)	+/- 95% CI	User's	Producers	Overall
Hardwood	34,007	[illegible]	[illegible]	[illegible]	0.75
Conifer	41,751	[illegible]	[illegible]	[illegible]	
Other	23,243	5,540	0.92	0.71	

The land cover map overestimated hardwood

Error Matrix Using Sample Counts

Class	Area (ha)	User's	Producers	Overall
Hardwood	45,000	0.63	0.80	0.77
Conifer	36,000	0.81	0.73	
Other	18,000	0.92	0.79	

Error Matrix Comparisons

Unbiased Area Estimators

Class	Area (ha)	+/- 95% CI	User's	Producers	Overall
Hardwood	34,007				
Conifer	41,751				0.75
Other	23,243			0.71	

The land cover map overestimated hardwood and underestimated conifer

Error Matrix Using Sample Counts

Class	Area (ha)	User's	Producers	Overall
Hardwood	45,000	0.63	0.80	
Conifer	36,000	0.81	0.73	0.77
Other	18,000	0.92	0.79	

Summary

- Reference points must be derived using an unbiased method: stratified random sampling is preferred for categorical maps
- Error matrix is important for understanding land cover map error
- Using only sample counts (week 1) gives some indication of map error, but…
- Incorporating area proportions results in more informative error matrix
  - Results in an unbiased area estimator
  - Also includes measures of uncertainty (confidence intervals)