

Assessing the Accuracy of Land Cover Classifications

Cindy Schmidt and Amber McCullum

Week 1: February 13, 2018



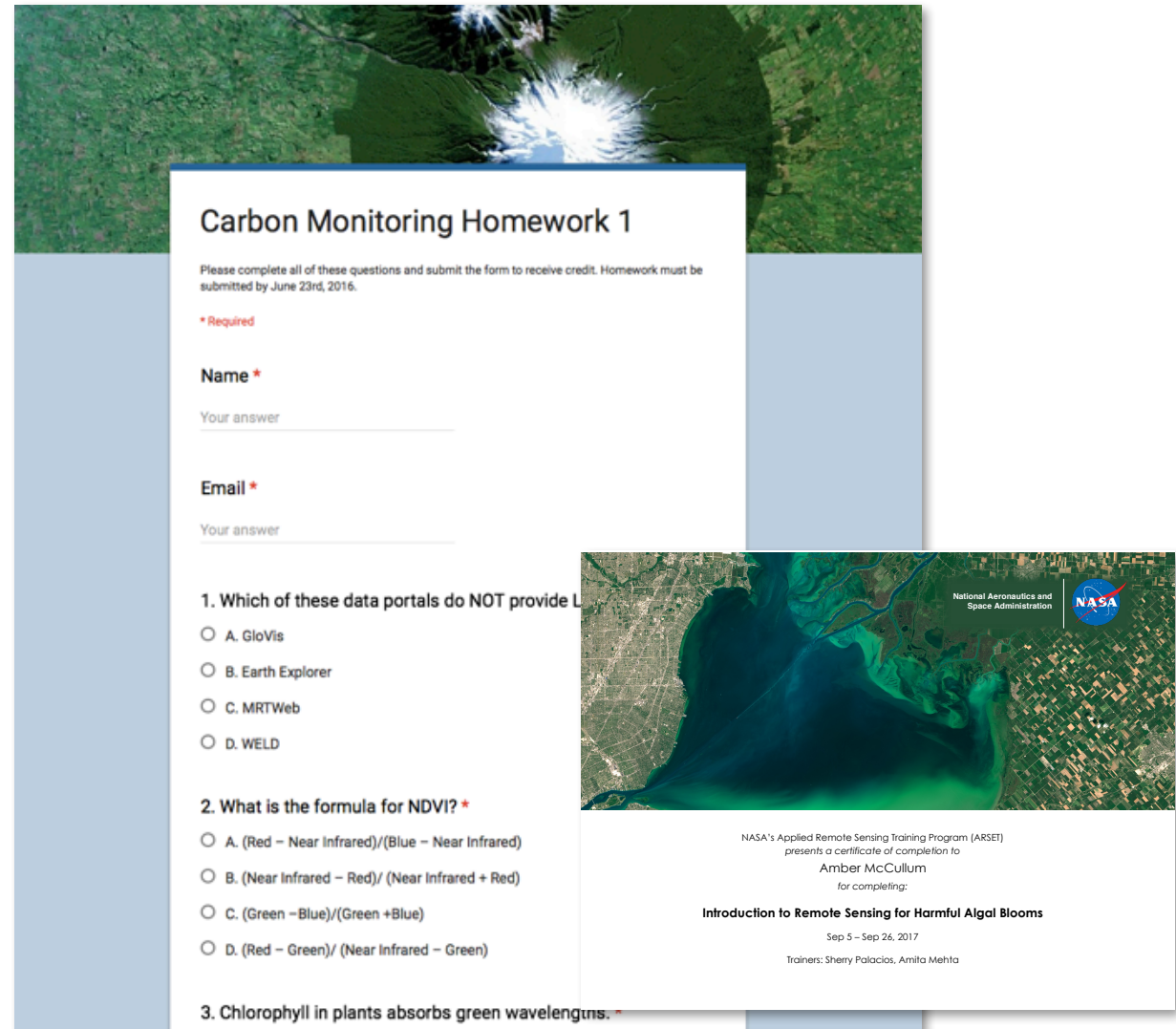
Course Structure

- 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)
- Each session will consist of a lecture followed by a hands-on exercise
- Webinar recordings, PowerPoint presentations, exercises, and the homework assignment can be found after each session at:
 - <https://arset.gsfc.nasa.gov/land/webinars/18adv-land-classification>
 - Q&A: Following each lecture and/or by email
 - cynthia.l.schmidt@nasa.gov, or
 - amberjean.mccullum@nasa.gov



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



Carbon Monitoring Homework 1

Please complete all of these questions and submit the form to receive credit. Homework must be submitted by June 23rd, 2016.

* Required

Name *

Your answer _____

Email *

Your answer _____

1. Which of these data portals do NOT provide L

A. GloVis

B. Earth Explorer

C. MRTWeb

D. WELD

2. What is the formula for NDVI? *

A. $(\text{Red} - \text{Near Infrared}) / (\text{Blue} - \text{Near Infrared})$

B. $(\text{Near Infrared} - \text{Red}) / (\text{Near Infrared} + \text{Red})$

C. $(\text{Green} - \text{Blue}) / (\text{Green} + \text{Blue})$

D. $(\text{Red} - \text{Green}) / (\text{Near Infrared} - \text{Green})$

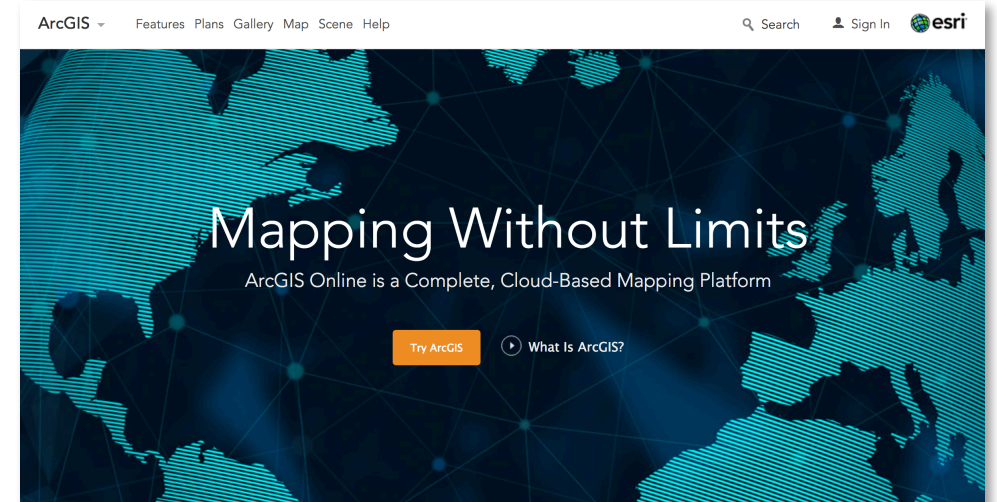
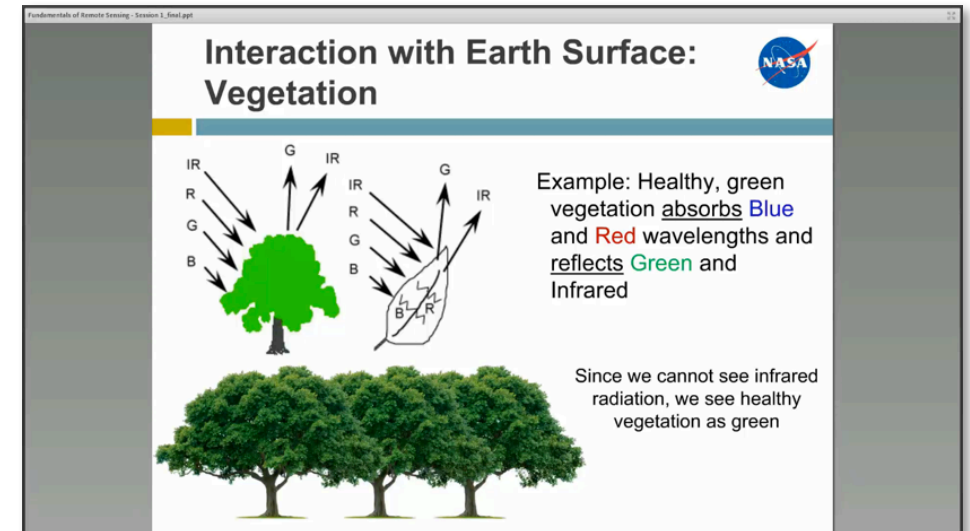
3. Chlorophyll in plants absorbs green wavelengths. *

NASA's Applied Remote Sensing Training Program (ARSET)
presents a certificate of completion to
Amber McCullum
for completing:
Introduction to Remote Sensing for Harmful Algal Blooms
Sep 5 - Sep 26, 2017
Trainers: Sherry Palacios, Amita Mehta



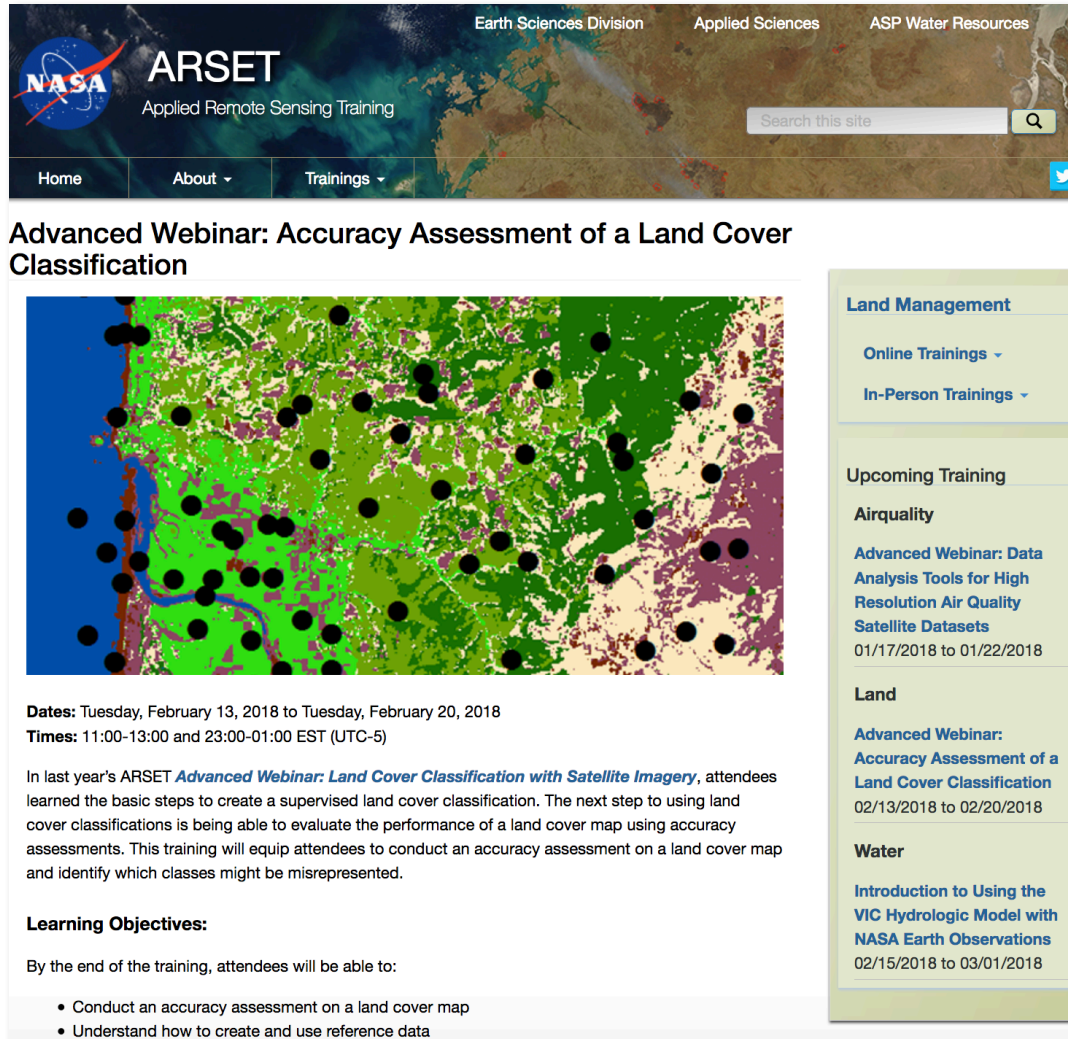
Prerequisites

- 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>



The screenshot shows the ARSET website header with the NASA logo and 'ARSET Applied Remote Sensing Training'. Navigation links include Home, About, and Trainings. A search bar is present. The main content area features a satellite image of a land cover classification with black dots indicating reference data points. The sidebar lists various training categories: Land Management (Online and In-Person Trainings), Upcoming Training (Airquality, Land, Water), and a specific webinar entry: 'Advanced Webinar: Accuracy Assessment of a Land Cover Classification' from 02/13/2018 to 02/20/2018.

Advanced Webinar: Accuracy Assessment of a Land Cover Classification

Dates: Tuesday, February 13, 2018 to Tuesday, February 20, 2018
Times: 11:00-13:00 and 23:00-01:00 EST (UTC-5)

In last year's ARSET [Advanced Webinar: Land Cover Classification with Satellite Imagery](#), attendees learned the basic steps to create a supervised land cover classification. The next step to using land cover classifications is being able to evaluate the performance of a land cover map using accuracy assessments. This training will equip attendees to conduct an accuracy assessment on a land cover map and identify which classes might be misrepresented.

Learning Objectives:

By the end of the training, attendees will be able to:

- Conduct an accuracy assessment on a land cover map
- Understand how to create and use reference data

Audience:

This training is primarily intended for local, regional, state, federal, and international organizations interested in assessing vegetation condition using satellite imagery. Professional organizations in the public and private sectors engaged in environmental management and monitoring will be given preference over organizations focused primarily on research.

Registration Information:

There is no cost for the webinar, but you must register. Space is limited, and preference will be given to organizations listed above over organizations focused primarily on research. You will be notified by email if your registration has been approved on or before February 12.

Course Agenda:

[Agenda.pdf](#)

Session 1: Introduction to Accuracy Assessments

February 13, 2018

- Overview of Accuracy Assessment (Lecture)
- Identifying Reference Data (Lecture)
- Sample Design and Methods (Lecture)
- Comparing Reference Data to Classified Image (Exercise)

Session 2: Conducting Unsupervised and Supervised Land Cover Classifications

February 20, 2018

- Description Overview of the Error Matrix (Lecture)
- Overview of the Kappa Statistic (Lecture)
- Creating the Error Matrix and Calculating the Kappa Statistic (Exercise)

Application Area: Land

Available Languages: English, Spanish

Instruments/Missions: Landsat

Keywords:

Conservation, Ecosystems, Land-Cover and Land-Use Change (LCLUC), Satellite Imagery, Tools

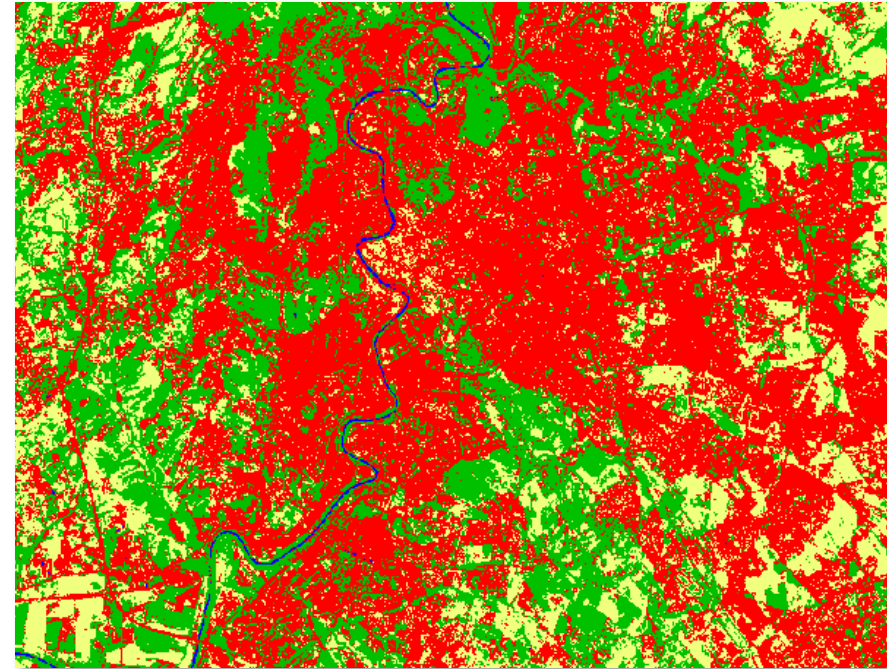
Course materials are provided here and will be active after each week



Course Outline



Session 1: Accuracy
Assessment Basics

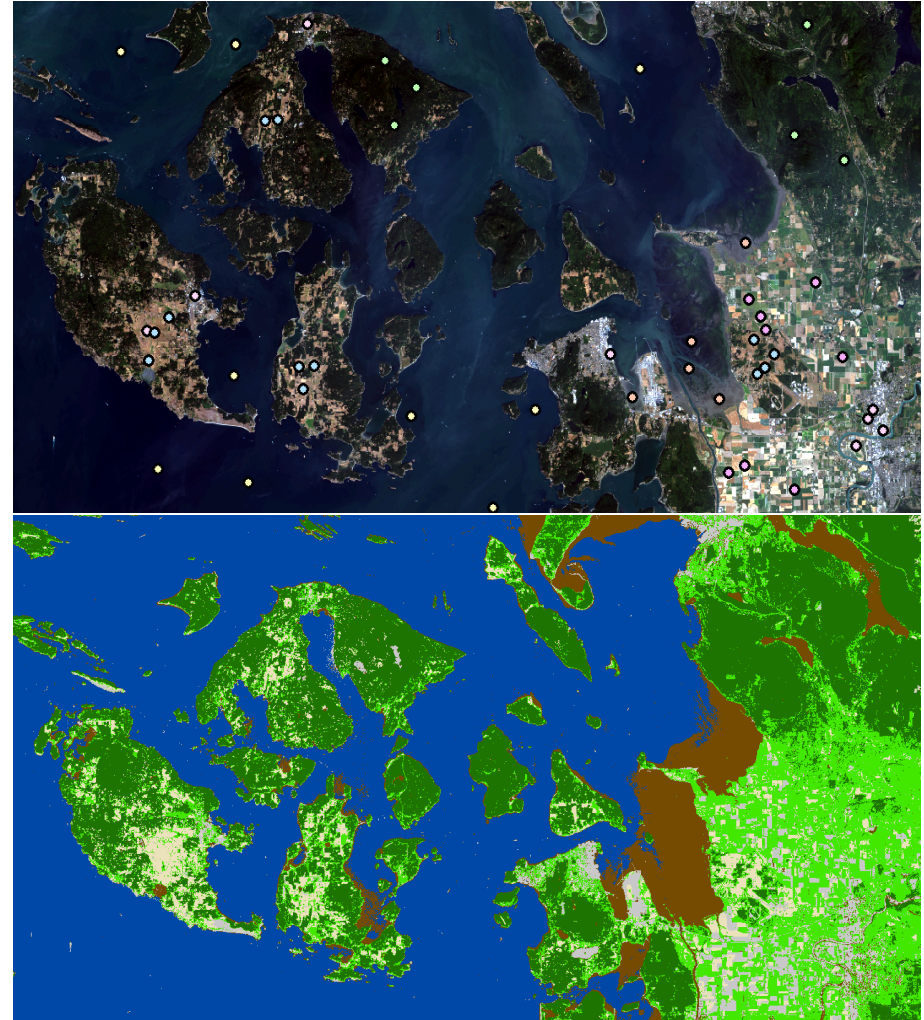


Session 2: Area Estimation



Session 1 Agenda

- What is accuracy assessment?
- Sample Design
 - Sampling Methodologies
- Reference Data
- Creating an Error Matrix
- Class Accuracy
- Exercise 1
- Q&A Session

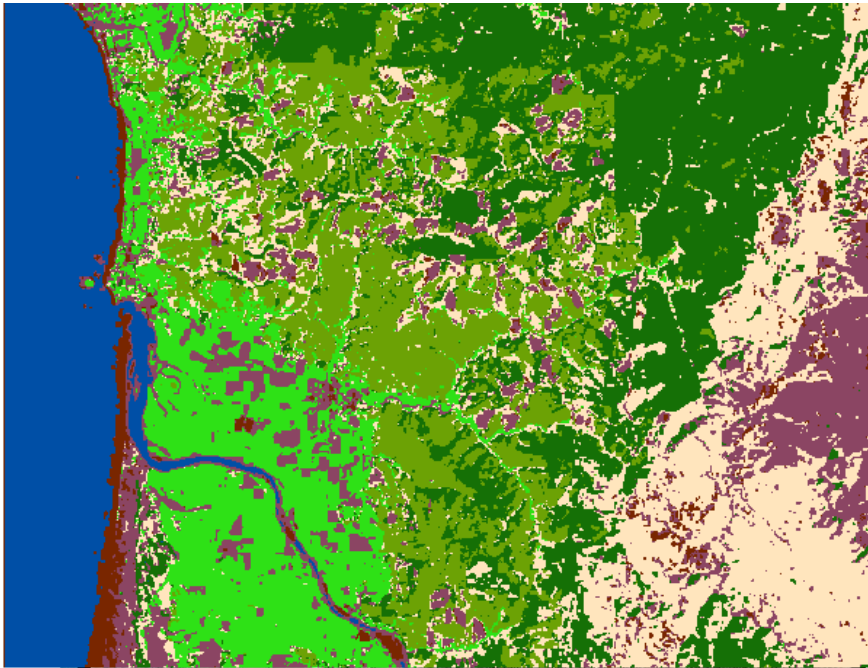


Example of sampling locations (top) and a classified image (bottom)



What is Accuracy Assessment?

- The process by which the accuracy or correctness of an image classification is evaluated
- Involves the comparison of the image classification to reference data that are assumed to be true



Reference (text): Maus, P., & Golden, M. L. (1996). *Guidelines for the use of digital imagery for vegetation mapping*. Washington, DC: U.S. Dept. of Agriculture, Forest Service



Sample Design

- Objective: obtain information that is representative of the population (i.e. the number of classes and the size of the classes)
- Basic Rules:
 - Data used for accuracy assessment must be independent of training data (for supervised classification)
 - Avoiding bias is key
 - All classes must be sampled
- Since you cannot identify every pixel, a reference point sample design is needed to determine point location:
 - Sample size
 - Sampling method

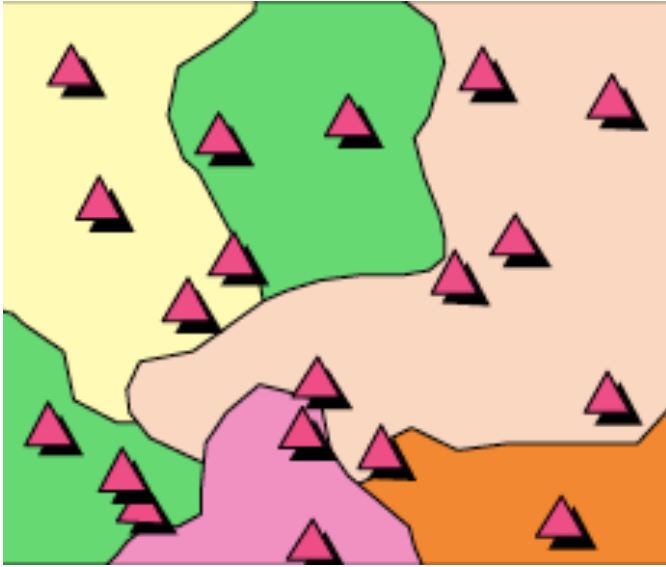


Sample Size

- The number of samples must be sufficiently large to statistically represent all classes, with enough in each class
- The minimum number of samples can be determined for each class using statistical methods:
 - Binomial Probability Theory – uses the expected percent accuracy and allowable error
 - Other methods use the proportion of area of each class
- Generally, at least 50 points per class is recommended
- For more information:
 - ARSET online webinar: Remote Sensing of Forest Cover and Change Assessment for Carbon Monitoring Session 4: Accuracy Assessment
 - BEEODA (Boston Education in Earth Observation Data Analysis):
<https://beeoda.org>

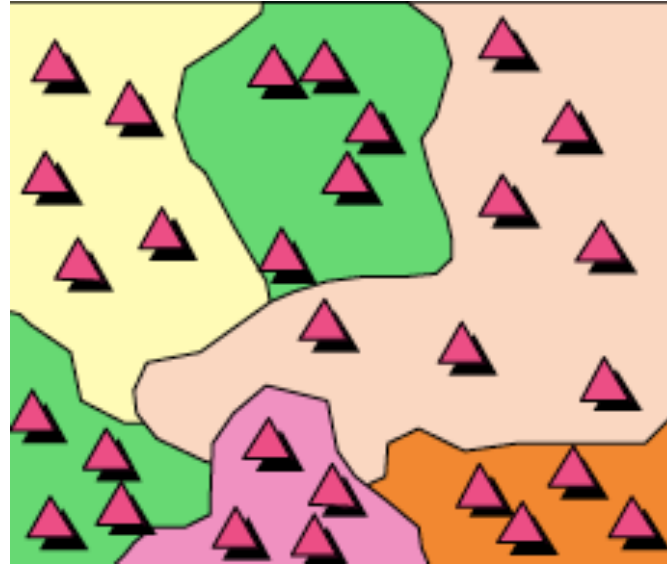


Sampling Methods



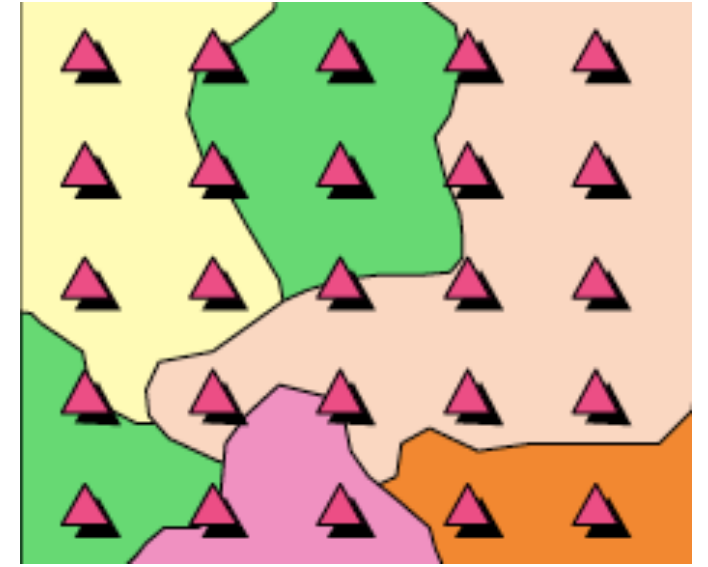
Simple Random Sampling:

Observations are randomly placed



Stratified Random Sampling:

A minimum number of observations are randomly placed in each category



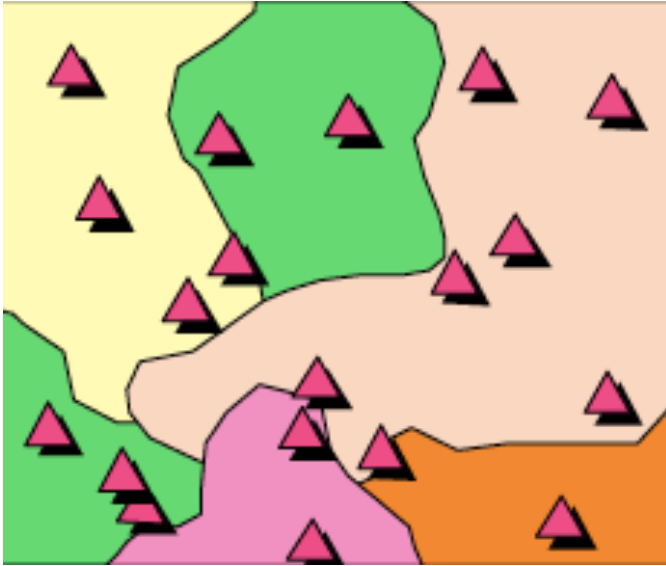
Systematic Sampling:

Observations are placed at equal intervals according to a strategy



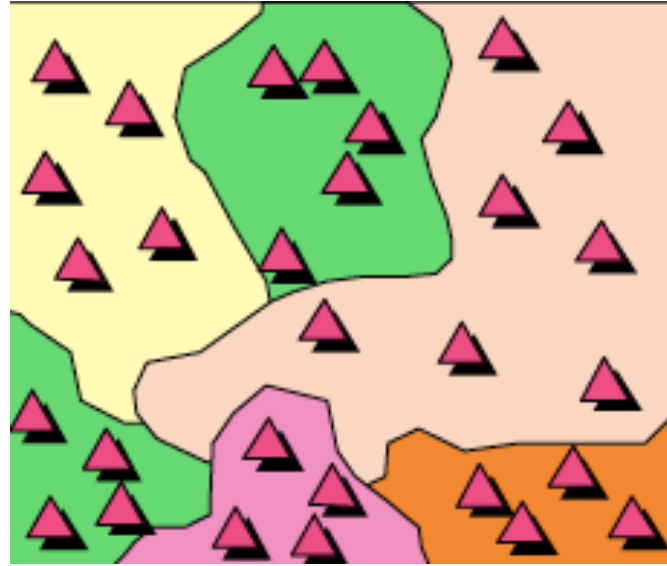
Sampling Methods

Preferred



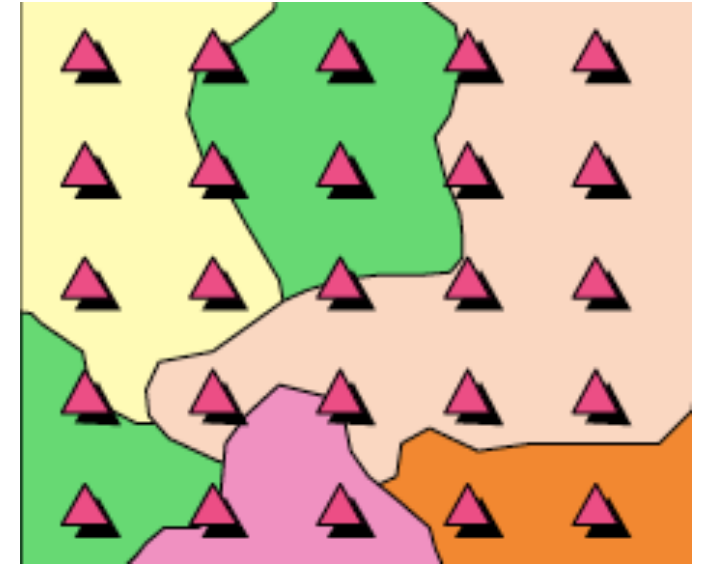
Simple Random Sampling:

Observations are randomly placed



Stratified Random Sampling:

A minimum number of observations are randomly placed in each category



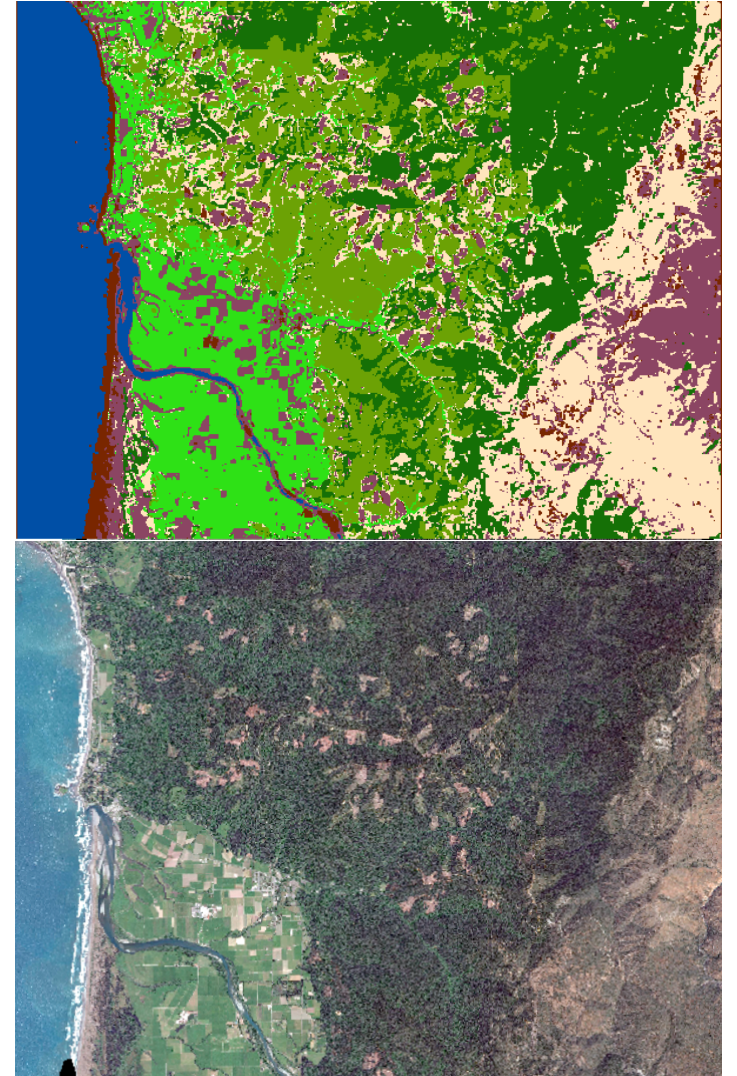
Systematic Sampling:

Observations are placed at equal intervals according to a strategy



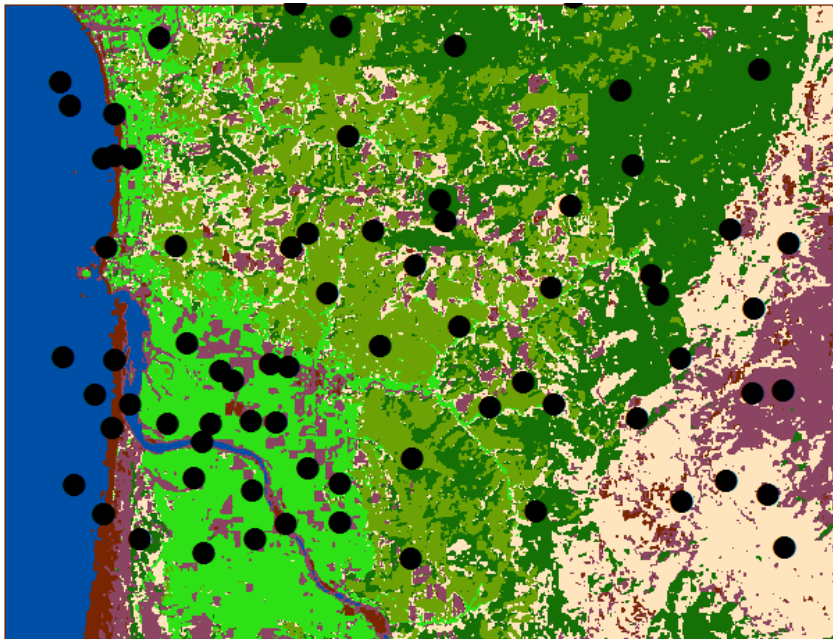
Reference Data

- An existing map
 - Existing inventory data
 - Aerial photo interpretation
 - High resolution satellite imagery
 - Ground collected (usually preferred but most expensive)
-
- Ideally, the reference data should be acquired at or near the same time as the classified image



Using Reference Data to Select Your Pixels

- Place reference points on your classified image and select those pixels
 - May be created with data from GPS
 - May be interpreted from maps or photos



- The result is a table that gives you the land cover class, the associated reference point, and the number of points (frequency) for each combination

Landsat Land Cover	Reference Point	Frequency
1	1	24
1	2	10
2	2	30
2	3	5
3	3	23
3	1	1



Creating an Error Matrix

Land Cover Classes

1 – Hardwood

2 – Conifer

3 - Other

Landsat Land Cover	Reference Point	Frequency
1	1	24
1	2	10
2	2	30
2	3	5
3	3	23
3	1	1



Creating an Error Matrix

Land Cover Classes

- 1 – Hardwood
- 2 – Conifer
- 3 - Other

Landsat Land Cover	Reference Point	Frequency
1	1	24
1	2	10
2	2	30
2	3	5
3	3	23
3	1	1

Correctly classified



Error Matrix

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

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



Error Matrix

Reference Data

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

Classification Data

Diagonals are pixels classified correctly according to the reference data



Error Matrix

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

Column totals are the total number of reference pixels in each class



Error Matrix

Reference Data

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

Row totals are the total number of classified pixels in each class



Error Matrix

Reference Data

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

Off diagonal numbers represent errors in the classification



Error Matrix

Reference Data

Classification Data

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

Overall accuracy is the sum of the diagonals divided by the total:
 $(24 + 30 + 23) / 100 = 77/100 = 77\%$



Error Matrix

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

Individual class accuracy is the diagonal value divided by the row **or** column sum:

Conifer: $30/37 = 0.81 = \mathbf{81\%}$ OR $30/41 = 0.73 = 73\%$



Individual Class Accuracy

- Classification error occurs when a pixel belonging to one category is assigned to another category
- There are two types of errors: Omission and Commission
 - Errors of omission occur when a pixel is left out of the category being evaluated (of all the pixels that are actually conifers, those that have been classified as hardwood)
 - Errors of commission occur when a pixel is incorrectly included in the category being evaluated (of all the classified hardwood pixels, those that are actually conifers)
- Measures of accuracy (% correct) are called User's and Producer's Accuracy
 - User's Accuracy measures errors of commission
 - Producer's Accuracy measures errors of omission



Error Matrix

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

User's accuracy (errors of commission): Diagonal value divided by the row total:

$$30 / 37 = 81\%$$

Of the 37 pixels classified as conifers, 30 were identified as conifers in the reference data. 5 hardwood and 2 other pixels were included in the conifer classification.



Error Matrix

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

Producer's accuracy (errors of omission): diagonal value divided by the column total

$$30 / 41 = 73\%$$

Of the 41 pixels that were referenced as conifer, only 30 were correctly classified as conifer. 10 were classified as hardwood and 1 was classified as other



Error Matrix

Remember that the overall accuracy is 77%

Reference Data

Classification
Data

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

User's Accuracies

- Hardwood: $24/38 = 63\%$
- Conifer: $30/37 = 81\%$
- Other: $23/25 = 92\%$

Producer's Accuracies

- Hardwood: $24/30 = 80\%$
- Conifer: $30/41 = 73\%$
- Other: $23/29 = 79\%$



Accuracy Assessment Conclusions

- Determining sample size is not trivial!
- Avoiding bias is important
- Generally, 80% overall accuracy is considered good
- But...identifying error is more complicated than simply reporting overall accuracy
 - It is important to know which classes have the greatest error
- Next week: Using the error matrix to calculate unbiased area estimates

