Remote Sensing 1: GEOG 4/585

Lecture 4.2.

Image classification 2

Johnny Ryan (he/him/his)
jryan4@uoregon.edu
Office hours: Monday 15:00-17:00
in 165 Condon Hall

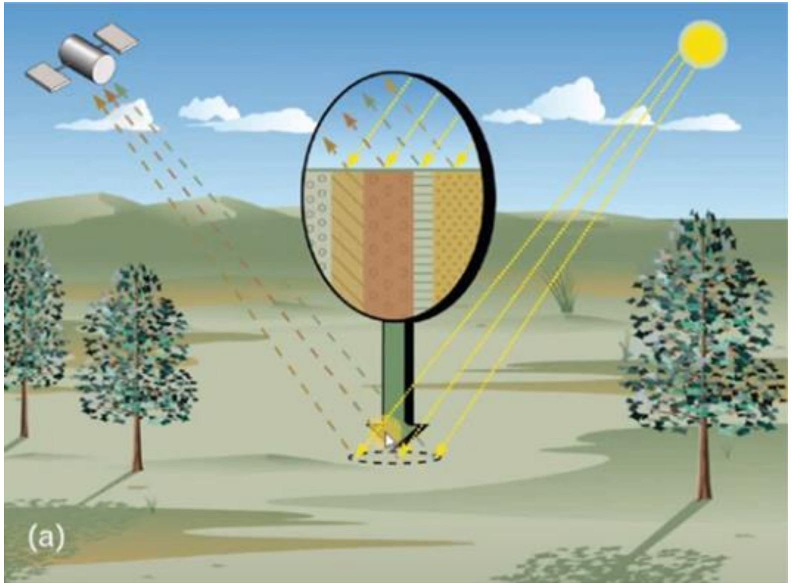
<u>Required reading:</u>
Principles of Remote Sensing pp 280-306

Overview

- Soft classification approaches
    - Spectral unmixing
    - Fuzzy classification
- Machine learning
- Post-classification label refinement
- Accuracy assessment

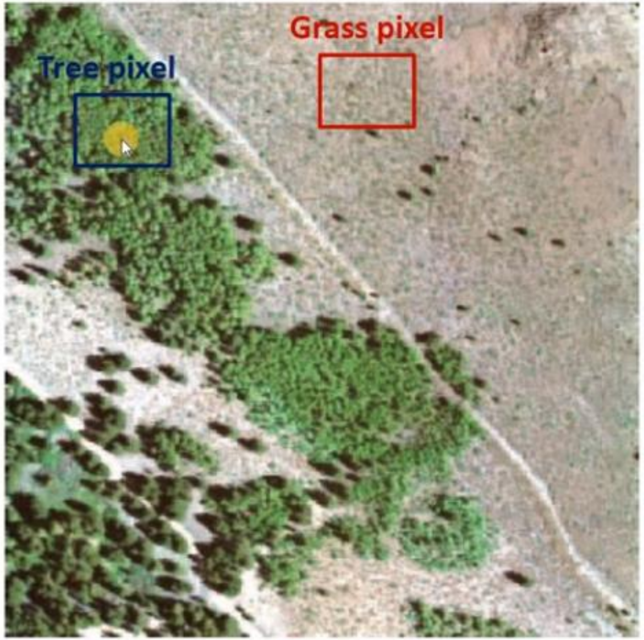
Spectral unmixing (or sub-pixel classification)

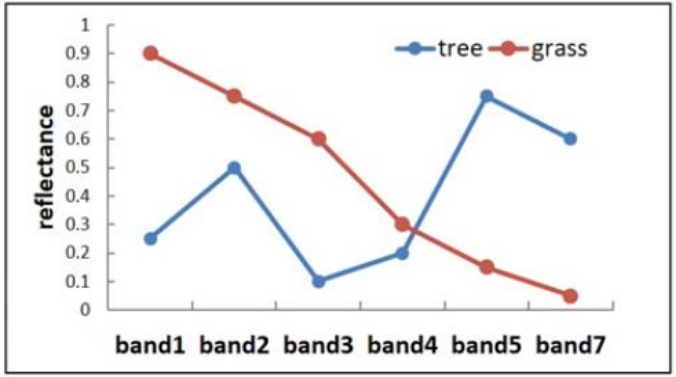
- Often the observed reflectance of a pixel is a combination (spectral mixture) of a number of ground classes present at the surface
- Spectral unmixing strives to find the relative fractions of surface types that together contribute to the observed reflectance of a pixel on the ground
- Post-classification procedure



Spectral unmixing

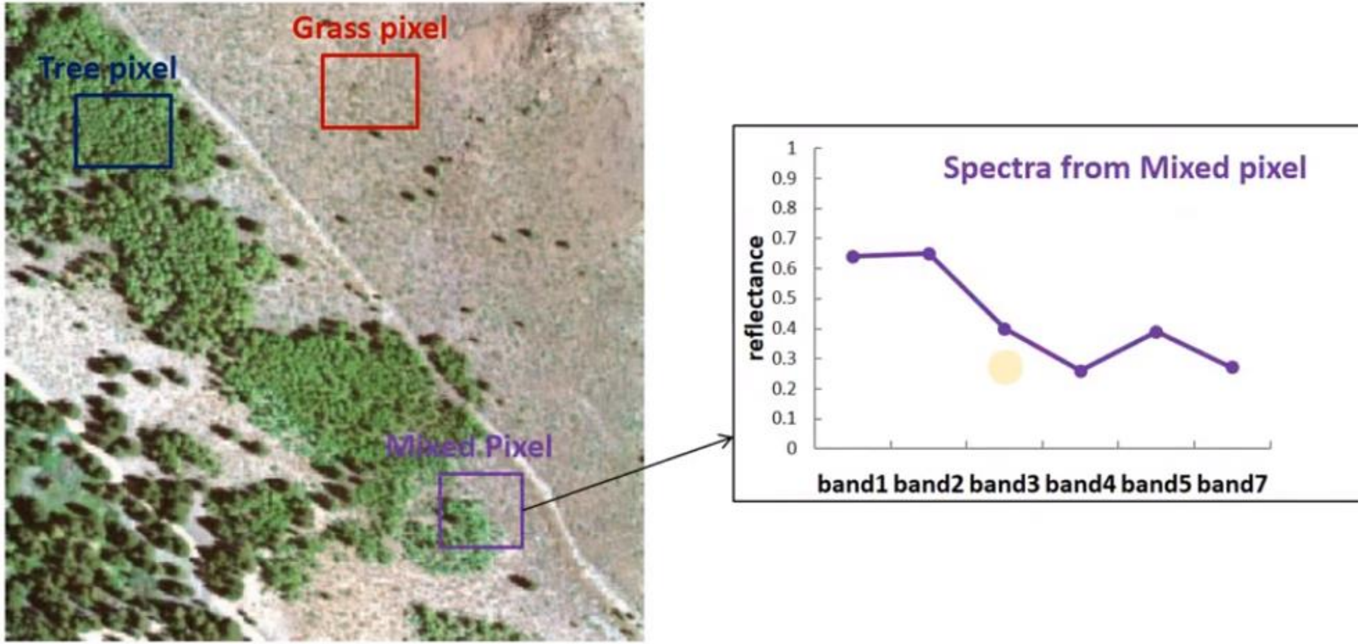
- Step 1: identify end-member spectral reflectance curves



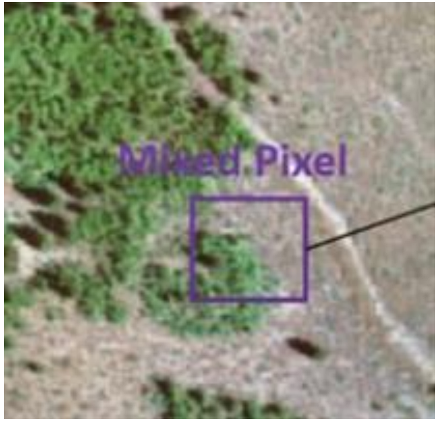


Spectral unmixing

- Step 2: identify a mixed-pixel (or mixel)



Spectral unmixing for one band



- Step 3: compute mixing fractions
  - Tree reflectance (R_{Tree}) = 10
  - Grass reflectance (R_{Grass}) = 50
  - Mixed pixel reflectance (R_{Mix}) = 30

$$R_{Mix} = (Fraction_{Tree} * R_{Tree}) + (Fraction_{Grass} * R_{Grass})$$

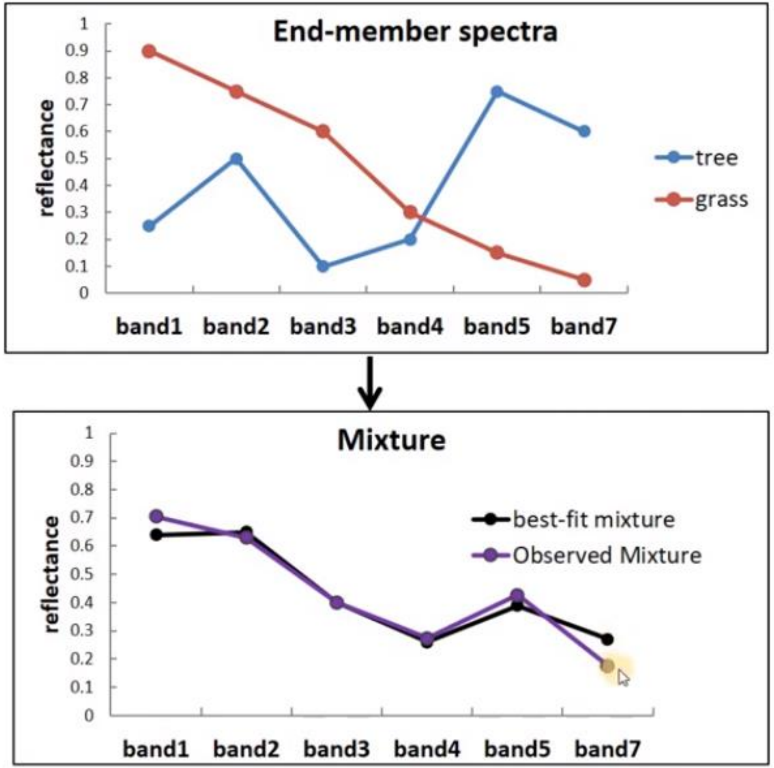
$$30 = (Fraction_{Tree} * 10) + (Fraction_{Grass} * 50)$$

Recognize “unit sum constraint”:

$$Fraction_{Tree} + Fraction_{Grass} = 1$$

$$30 = (Fraction_{Tree} * 10) + ((1 - Fraction_{Tree}) * 50)$$

Spectral unmixing



- A single set of mixing fractions rarely produces a perfect match for all bands due to:
  - Uncertainty in end-member spectra
  - Presence of unidentified end-member components

- We can use an iterative approach that computes mixing fractions for all combinations of end-members

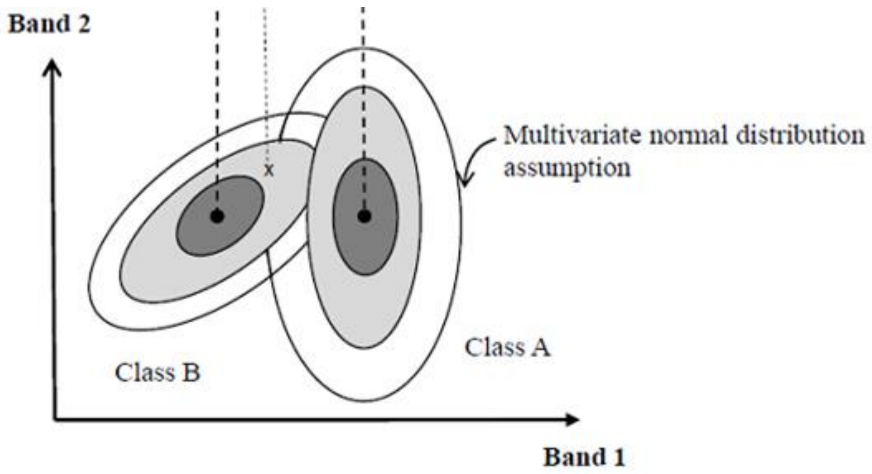
- Goal is to minimize the summed square mismatch between best-fit and observed mixtures

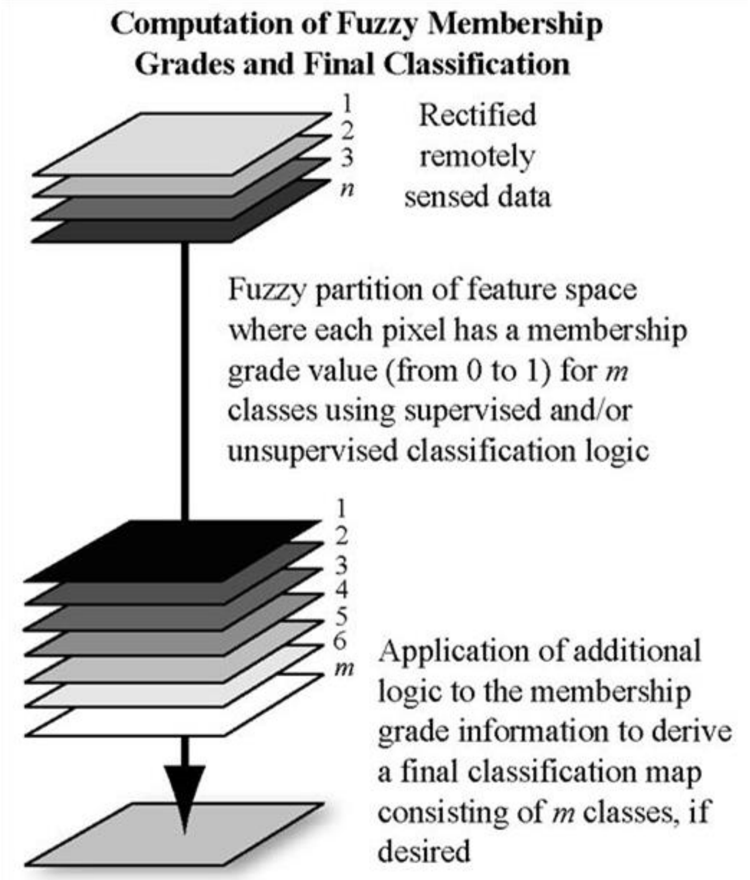
Spectral unmixing assumptions

- Identified all end-members
- All end-members contribute to signal (i.e. no shielding or interfering)
- No interaction between materials occurs (i.e. each photon sees only one material)
- The scale of mixing is very large as opposed to the grain size of the materials

Fuzzy classification

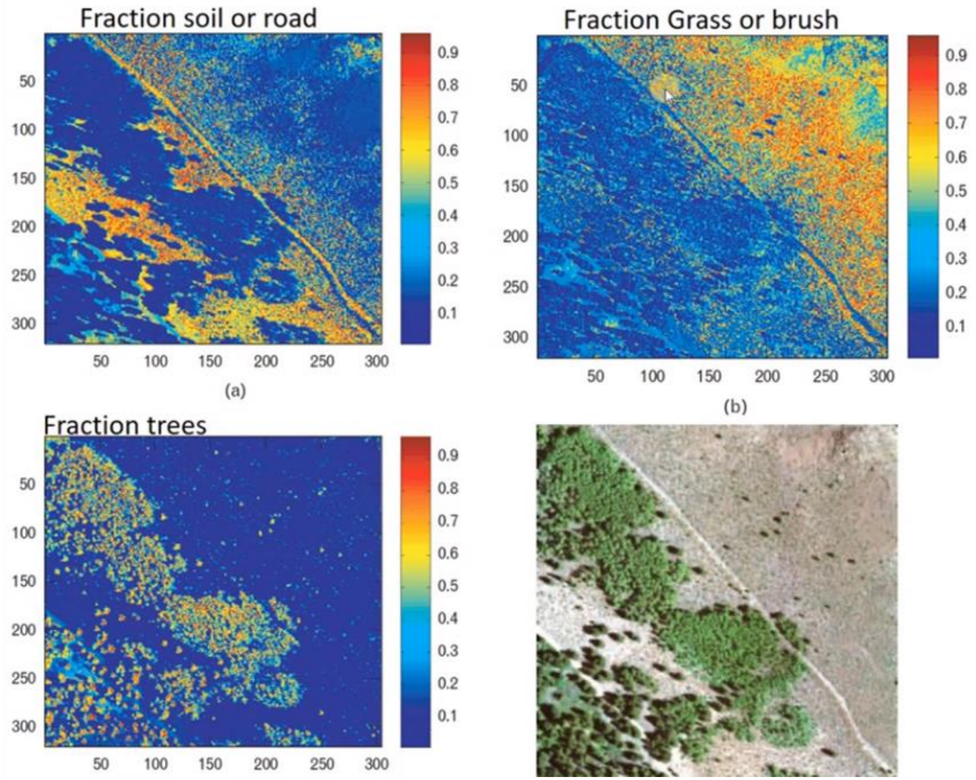
- It is also possible to use fuzzy set logic where each pixel can be a member of multiple classes (usually between 0 and 1)
- For example, retain probabilities from Maximum Likelihood classification



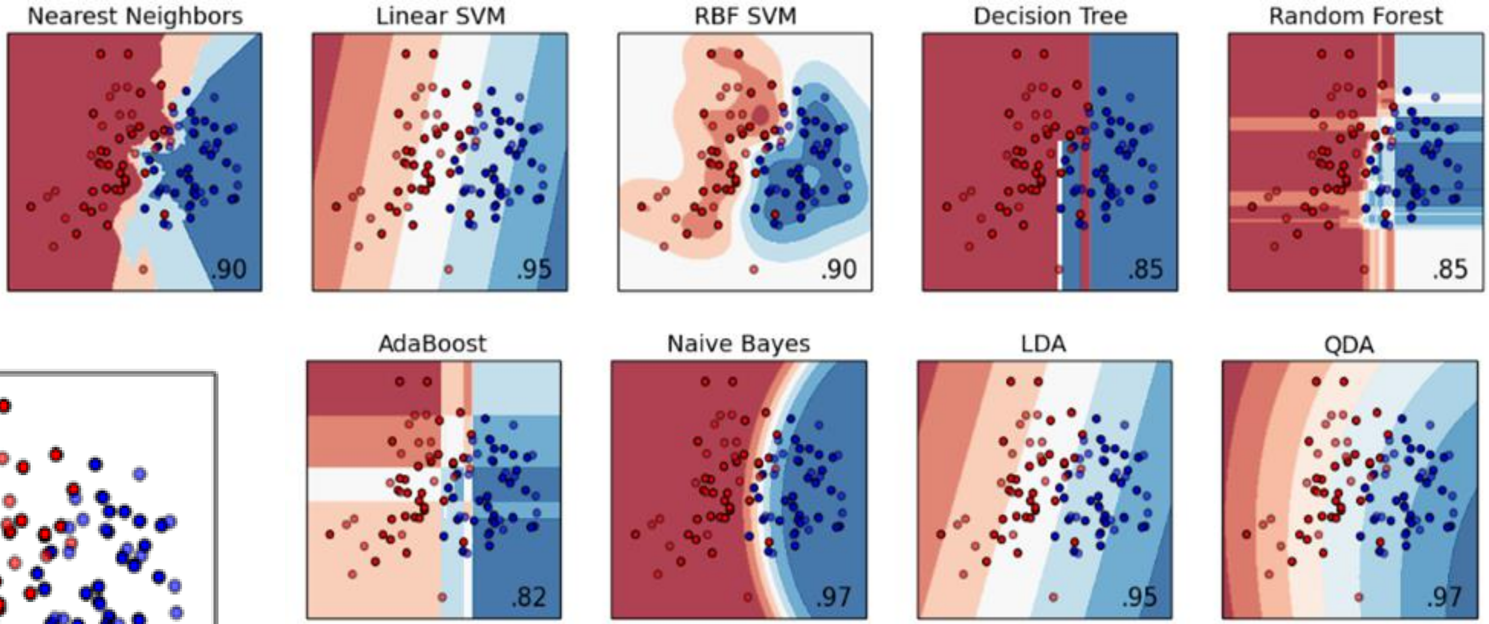


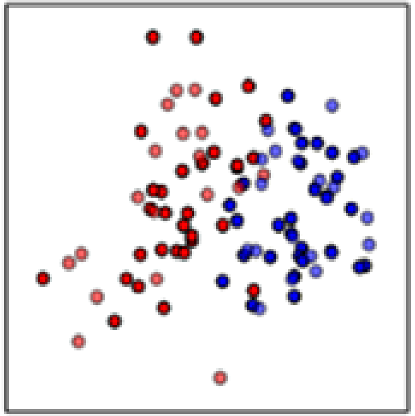
Advantages of “soft” classification approaches

- Often more accurate than “hard” classification approaches
- Encourages us to acknowledge complexity in surface types
- Allows simple computation of uncertainties
- Can update and modify thresholds for membership (e.g. optimize a classification against ground truth data)



Machine learning





Post-classification label refinements

Processing required to refine output thematic or quantitative images to suitable format for map presentation

- Filtering
  - Reclassifies isolated pixels or noise based on neighboring pixels (e.g. median filter)

- Smoothing
  - Smooths the ragged class boundaries and clumps the classes.

Original Image

with Median Filter

Accuracy assessment

- Accuracy assessment consists of validating or evaluating your classification
- Usually done by a sampling approach in which a number of raster values are selected and both the classification result and the true world class are compared
- Main steps:

  a) Select sampling scheme

  b) Obtain high-quality testing data at sampling sites

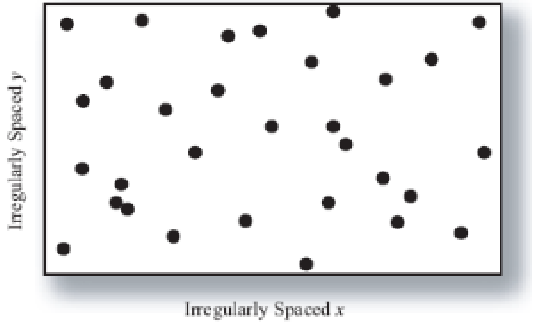
  c) Conduct error or (confusion) matrix analysis

Sampling scheme

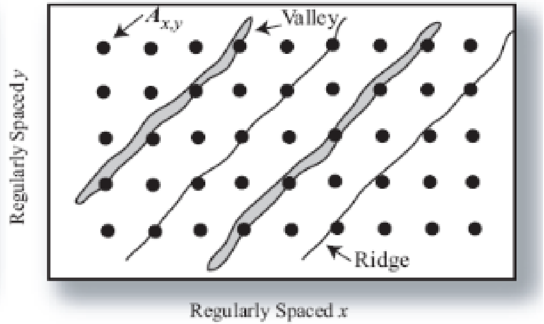
The location of the sample locations must be selected randomly without bias!

Five common sampling designs used to collect ground reference test data for assessing the accuracy of a remote sensing–derived thematic map:

- Random sampling
  - Random number generator is used to generate random x, y coordinates
  - Disadvantage: random sampling may undersample small but (possibly) very important classes

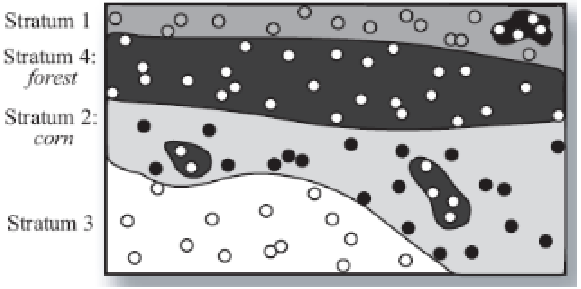


- Systematic sampling
  - Sample along a line of coordinates (starting point chosen randomly)
  - Disadvantage: Not useful if periodicities exist in the data

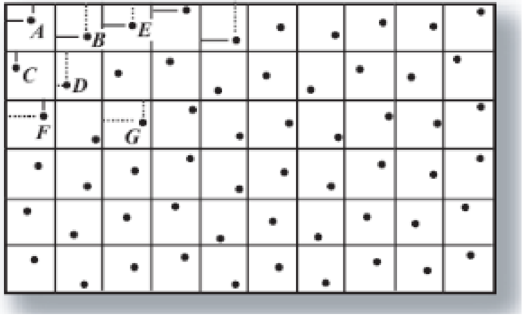


Sampling scheme

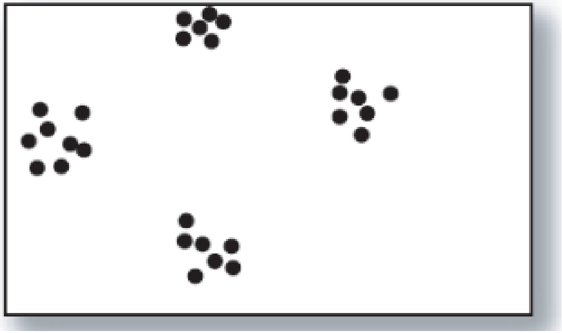
- Stratified random sampling
  - Minimum number of samples are selected from each strata (i.e., land cover category)
  - Assures even samples per class but could be biased



- Stratified systematic unaligned sampling
  - Combines randomness and stratification with a systematic interval
  - Point(s) selected at random within a regularly spaced grid



- Cluster sampling
  - Sometimes it is difficult to go into the field to randomly selected sites; so, several samples can be collected at a single random location
  - Each pixel in the cluster is not independent of the others



Obtain validation data

- Remember, we use training data to fit the model, and validation (test) data to test the model

- Similar to obtaining training data, there are a number of ways to obtain testing data:
  - Collection of in situ information
  - Use of external imagery, preferably with superior resolution
  - On-screen selection of polygonal training data…Regions of Interest (ROIs)

- There exist a number of statistical methods for determining minimum number of pixels per class based on:
  - Number of classes
  - Desired level of confidence
  - Anticipated areal cover by a certain class

- Rule of thumb, however, > 50 pixels per class in general
  - Fewer if class is homogeneous and more if class if heterogeneous

Error/confusion matrix

- To correctly perform a classification accuracy (or error) assessment, you will need to compare two sources of information:
  - Ground reference test information (validation/input) – columns
  - Pixels in a remote sensing-derived classification map (classification/output) – rows

		Ground Reference Test Information				
Remote Sensing Classification	Class	1	2	3	*k*	Row total
	1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,k}$	x_{1+}
	2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,k}$	x_{2+}
	3	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,k}$	x_{3+}
	k	$x_{k,1}$	$x_{k,2}$	$x_{k,3}$	$x_{k,k}$	x_{k+}
	Column total	x_{+1}	x_{+2}	x_{+3}	x_{+k}	N

Descriptive statistics

- Overall accuracy
  - How many of the test pixels were correctly classified (regardless of class)?

- Producer’s accuracy ~ Omission error
  - “How many of my test pixels did I correctly classify for each class?”
  - (How well can a certain area be classified?)

- User’s accuracy ~ Commission error
  - “How many of the pixels in the classified image are correctly classified for each class?”

- For each of the above, the accuracy is essentially the opposite of the error, such that:
  - Producer’s Acc. + Omission Err. = 100%
  - User’s Acc. + Commission Err. = 100%

Overall accuracy

- The overall accuracy of the classification map is by taking the total correct pixels (sum of the major diagonal) and dividing by the total number of pixels in the error matrix (N).

Ground truth

Remote sensing classification

	Residential	Commercial	Wetland	Forest	Water	Total	UA
Residential	**70**	5	0	13	0	88	*80%*
Commercial	3	**55**	0	0	0	58	*95%*
Wetland	0	0	**99**	0	0	99	*100%*
Forest	0	0	4	**37**	0	41	*90%*
Water	0	0	0	0	**121**	121	*100%*
Total	73	60	103	50	121	**407**	
PA (%)	*96%*	*92%*	*96%*	*74%*	*100%*		*94%*

Producer's accuracy (or omission error)

- Probability of a pixel being correctly classified
- Defined as the number of reference pixels classified accurately divided by the actual total number of pixels for that class.

Ground truth

Remote sensing classification

	Residential	Commercial	Wetland	Forest	Water	Total	UA
Residential	**70**	5	0	13	0	88	*80%*
Commercial	**3**	55	0	0	0	58	*95%*
Wetland	**0**	0	99	0	0	99	*100%*
Forest	**0**	0	4	37	0	41	*90%*
Water	**0**	0	0	0	121	121	*100%*
Total	**73**	60	103	50	121	407	
PA	***96%***	*92%*	*96%*	*74%*	*100%*		*94%*

User's accuracy (or commission error)

- Probability that a pixel classified on the map actually represents that category on the ground
- How often the class on the map will actually be present on the ground

Ground truth

Remote sensing classification

	Residential	Commercial	Wetland	Forest	Water	Total	UA
Residential	**70**	**5**	**0**	**13**	**0**	**88**	***80%***
Commercial	3	55	0	0	0	58	*95%*
Wetland	0	0	99	0	0	99	*100%*
Forest	0	0	4	37	0	41	*90%*
Water	0	0	0	0	121	121	*100%*
Total	73	60	103	50	121	407	
PA	*96%*	*92%*	*96%*	*74%*	*100%*		*94%*

Kappa coefficient

- The Kappa Coefficient of Agreement is the measure of agreement or accuracy between the remote sensing–derived classification map and the reference data as indicated by:
  - Observed accuracy (the major diagonal), and
  - Chance agreement (row and column totals referred to as marginals)

- Kappa essentially evaluates how well the classification performed as compared to just randomly assigning values.
  - It ranges from -1 to 1
  - Value of 0: classification is no better than a random classification
  - Negative value: classification is worse than random
  - Positive value: classification is better than random

$$\hat{K} = \frac{\text{observed accuracy - chance agreement}}{\text{100\% accuracy - chance agreement}}$$

Kappa coefficient

- 0.92 for our confusion matrix

Ground truth

Remote sensing classification

	Residential	Commercial	Wetland	Forest	Water	Total	UA
Residential	70	5	0	13	0	88	*80%*
Commercial	3	55	0	0	0	58	*95%*
Wetland	0	0	99	0	0	99	*100%*
Forest	0	0	4	37	0	41	*90%*
Water	0	0	0	0	121	121	*100%*
Total	73	60	103	50	121	407	
PA	*96%*	*92%*	*96%*	*74%*	*100%*		*94%*

Class activity

- Calculate overall accuracy and errors of omission and commission for this error matrix
- Which class has highest error of omission? Commission?

Ground truth

Remote sensing classification

Class	1	2	3	4	Total
1	70	10	15	5	100
2	8	67	20	5	100
3	0	11	88	1	100
4	4	10	14	72	100

Course survey

- The Midway Student Experience Survey for your courses is currently open and will close at 6PM on Fri, Oct 22, 2021 PDT
- “OU Course Survey” from left-hand-side menu on Canvas

Today’s lab

Lab Assignment #4: Supervised and unsupervised classification

Objectives:

- We will produce a land cover map of the McKenzie River Valley in July 2021 following the Holiday Farm Fire using both a supervised and unsupervised classification in QGIS.

Deadline: October 26 Tuesday 11:59 pm