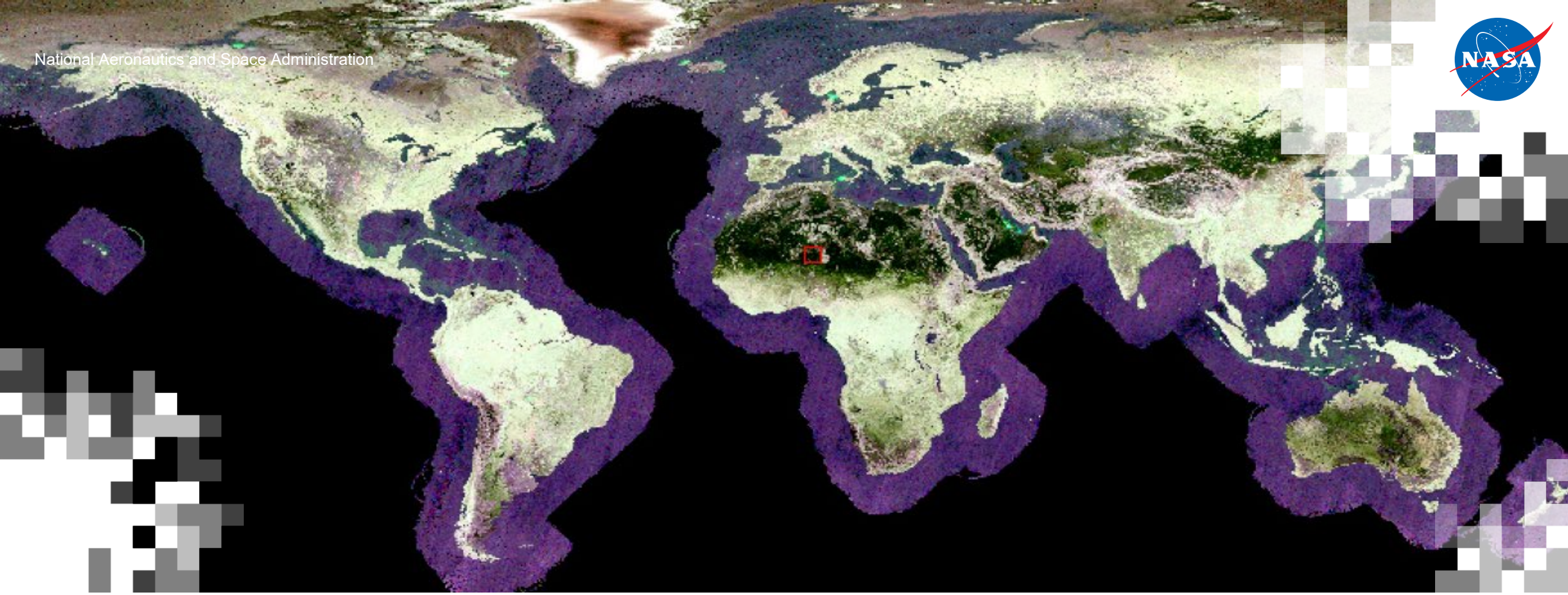


Exploiting SAR to Monitor Agriculture

Heather McNairn, Xianfeng Jiao, Sarah Banks and Amir Behnamian

4 September 2019

Learning Objectives

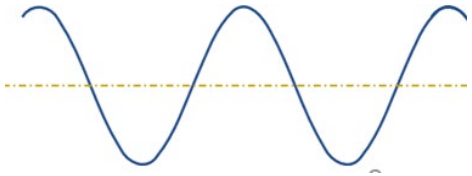
By the end of this presentation, you will be able to understand:

- how SAR configurations affect response from soils and crops
- the information content in SAR images relevant to soil and crop conditions
- the optimal sensor parameters for agriculture applications
- how to ingest, pre-process, and process SAR data for use in crop classification and soil moisture estimation

SAR System Considerations

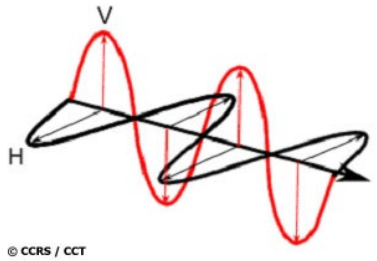
- When planning SAR data collections and when interpreting SAR response, the three fundamental system characteristics must always be considered

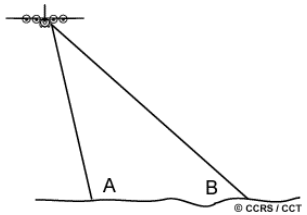
- Interpreting SAR response is always done, relative to these characteristics



frequency or wavelength

polarization





geometry (incidence angle and look direction)

Image Credits: Polarization & Geometry CCRS/CCT

Frequency or Wavelength

The best frequency should be selected

- consider the size of the target elements relative to SAR frequency. To maximize scattering, select wavelengths that are comparable in size or smaller than these elements
- is it important to penetrate into the target or is the goal to maximize surface scattering? Lower frequency (longer wavelengths) provide greater penetration
- is the goal to maximize or minimize sensitivity to surface roughness? A low frequency wave will see a surface as smooth while a high frequency wave will see this same surface as rough

What is the best frequency for agricultural monitoring? It depends!

- soil moisture: longer wavelengths (like L-band) are better as they penetrate deeper into the canopy and interact with soil
- crop classification and biophysical modeling: depends on canopy
- need enough penetration into canopy (L- or C-band for corn, for example) but not too deep so that we have soil interference (C- or X-band for lower biomass crops like soybeans)

The Power of Multiple Frequencies

Integration of data from RADARSAT-2, ALOS and TerraSAR-X, Manitoba (Canada)

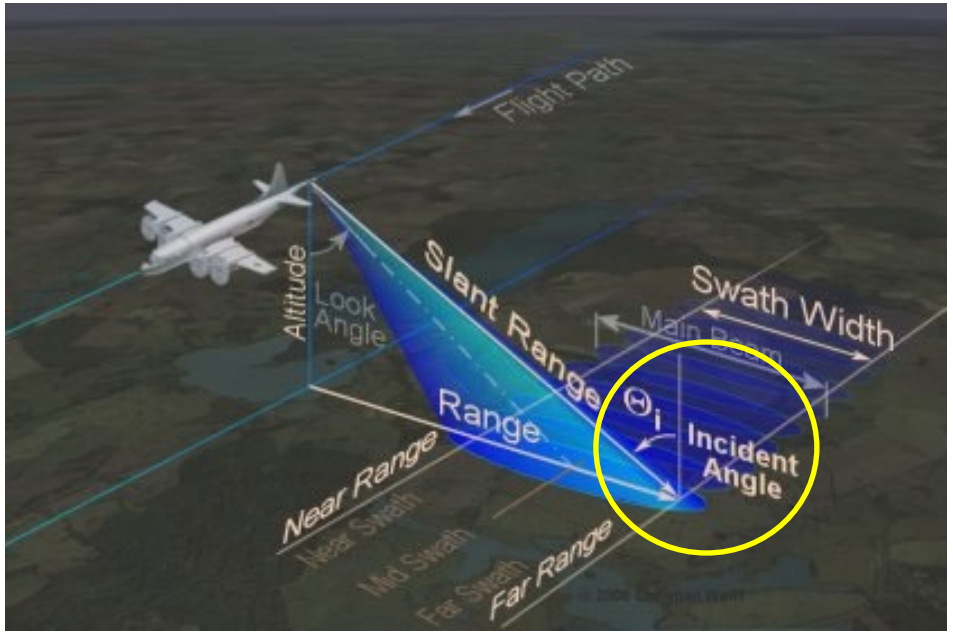
X (3.1 cm)
C (5.5 cm)
L (22.9 cm)

RADARSAT-2 VH July 10 2016

TerraSAR-X VV July 10 2016

ALOS VH July 9 2016

Incidence Angle



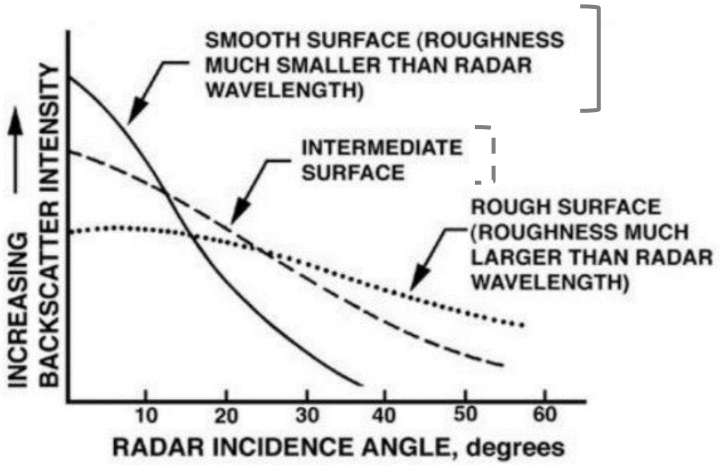
SAR incidence angle determines

- the contribution of different target elements to backscatter. Shallower angles interact more with the vegetation canopy; at steeper angles more of the signal can pass through to the ground without interacting with the canopy
- how rough the target appears to the SAR. Surfaces appear "smoother" at larger angles. The most significant incidence angle effects are observed on smoother surfaces
- how deep the microwave signal penetrates into the target

Image Credit: www.radartutorial.eu/20.airborne/ab06.en.html

Incidence Angle

- backscatter decreases with increasing incidence angle
- rate and function of decrease is target specific
- as a result, when a radar is viewing the same target at different angles, the backscatter will be different
- **CAUTION:** for temporal change detection, do not mix angles (use exact repeats)
- for biophysical estimation, ok to mix angles if the model accounts for the incident angle



rougher lower rate of change with angle

smoother higher rate of change with angle

Image Credits: (left) Ulaby et al. (1981a); (right) extension.umn.edu/soil-management-and-health/tillage-implements

Polarization

SAR polarization determines

- how transmitted microwaves interact with the target
  - if the target (such as vegetation) has a dominant vertical structure, V-polarized waves align with this structure and create greater scattering. With H-polarized waves, less of the energy interacts with the vertically structured target and more often, waves make their way through the canopy to the ground
- when considering transmit and receive signals, the amount of energy that is **re-polarized** (from H-transmit to V-receive; from V-transmit to H-receive) to create a cross-polarized response (HV or VH), depends on the structure of the target

What is the best frequency for agricultural monitoring?

- HV or VH is the single best polarization for either crop identification or crop biophysical estimation
- next best polarization is usually VV

What Target Characteristics Drive SAR Scattering

SARs respond to two fundamental characteristics of a target: (1) structure or roughness, (2) water content

- **Roughness:** characterized by two parameters, the root mean square variance (RMS) and the surface correlation length (l)
- **RMS (root mean square):** *the statistical variation of the random component of the surface height relative to a reference surface (in mm or cm)*
- **Correlation length (l):** *autocorrelation function measuring the statistical independence of surface heights at two points, spatially separated by a distance x'. The correlation length is equal to the displacement distance x' for which p(x') is equal to 1/e. If two points are separated by a distance greater than l, their surface heights are considered statistically independent*

For soils this means

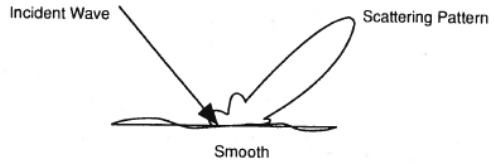
- random roughness caused by tillage (and other farm operations) modified by soil erosion and weathering effects
- periodic row structures caused by tillage and planting

How Does Roughness Affect Backscatter?

- backscatter will increase as soil roughness increases
- rougher soils appear brighter in SAR images

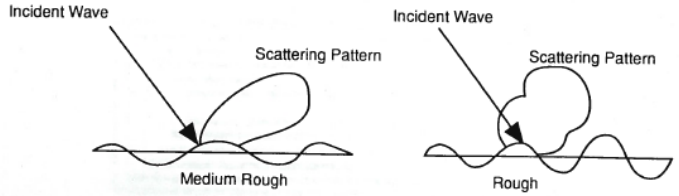
expected relative SAR response

The impact of roughness on backscatter depends on the frequency and incident angle of the SAR. Roughness is a relative concept.



According to the Rayleigh Criterion, a soil is smooth if

$$h < \frac{\lambda}{8\cos\theta}$$



where h is surface height variation in cm, λ is the wavelength in cm and θ is the incident angle in degrees

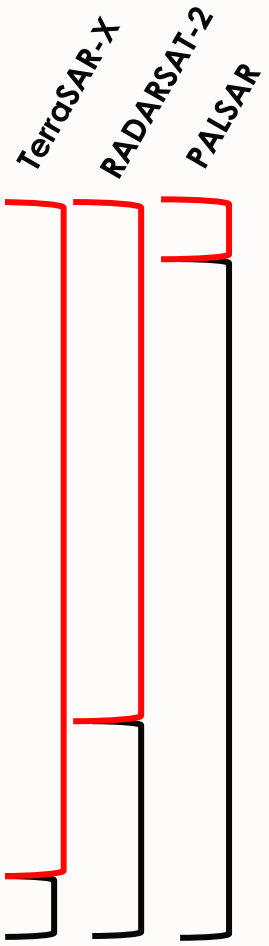
It's All Relative

Roughness less than "h" would be viewed as smooth by SAR

Incident Angle of 30°	
TerraSAR-X (3.1 cm)	h < 0.45 cm
RADARSAT-2 (5.6 cm)	h < 0.81 cm
PALSAR (23.6 cm)	h < 3.42 cm
Incident Angle of 50°	
TerraSAR-X (3.1 cm)	h < 0.60 cm
RADARSAT-2 (5.6 cm)	h < 1.09 cm
PALSAR (23.6 cm)	h < 4.59 cm

TABLE I
AVERAGE RANDOM ROUGHNESS (s) VALUES BASED ON SINGLE TILLAGE OPERATIONS [12].

Tillage Operation	s (cm)
Large offset disk	5.0
Moldboard plow	3.2
Lister	2.5
Chisel plow	2.3
Disk	1.8
Field cultivator	1.5
Row cultivator	1.5
Rotary tillage	1.5
Harrow	1.5
Anhydrous applicator	1.3
Rod weeder	1.0
Planter	1.0
No till	0.7
Smooth	0.6



Viewed by SAR as smooth at 50°

Viewed by SAR as rough at 50°

Source: Jackson, T.J., McNairn, H., Weltz, M.A., Brisco, B. and Brown, R.J. (1997). First order surface roughness correction of active microwave observations for estimating soil moisture. IEEE Transactions on Geoscience and Remote Sensing 35:1065-1069.

Vegetation Effects

The scale is very different from optical

Scattering of longer-wavelength microwaves is driven by

a) larger scale structures (size, shape and orientation of leaves, stems and fruit)

b) the volume of water in the vegetation canopy (at the molecular level)

So why is SAR sensitive to crop type and crop development?

- Crop structure changes significantly from one crop to the next, and as crops move through their growth stages

Crop structure varies significantly among soybeans, wheat, and corn

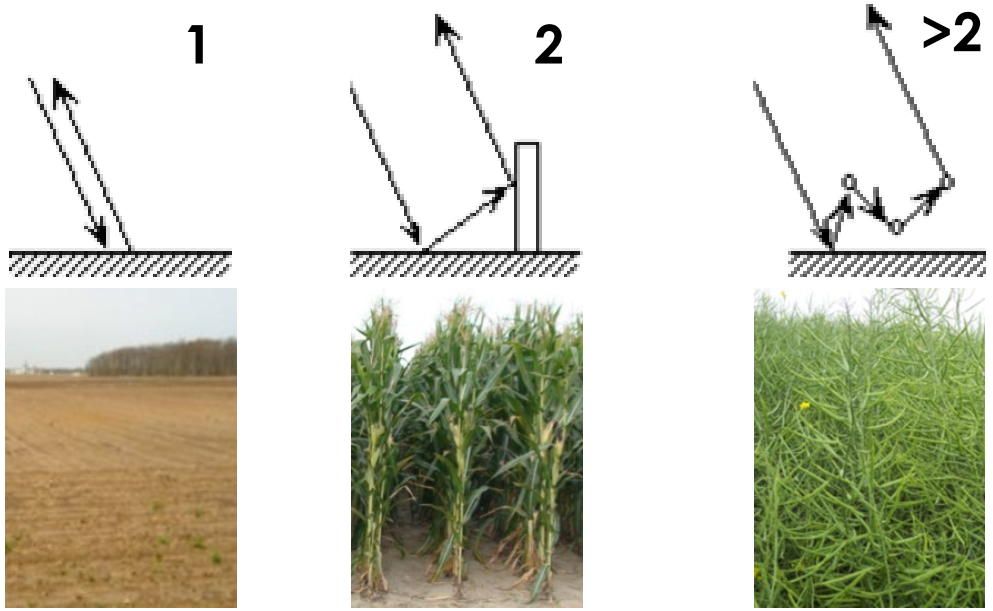
Structure also changes as crops grow

Types of Scattering

- as the microwave signal hits a target, the wave will undergo one, two or more scattering events
- a change from one polarization (i.e. H) to another polarization (i.e. V) is termed **re-polarization**.
- the number of events determines the type of scattering, intensity of scattering, and changes in the phase
- scattering events are dependent upon the structure and geometry of the target
- typically one scattering type dominates
- for distributed targets often secondary or tertiary scattering events occur and thus a mixture of scattering types often characterizes these targets

type of scattering + mixture of scattering events + phase characteristics + intensity = clues about type and condition of crops

Number of Scattering Events



bare soil

corn field with distinct vertical structure and wide row spacing

canola field in podding stage with an entanglement of scatterers

What About Water in the Target?

SAR is known to be sensitive to moisture, but **why?**

- water (H_2O) is a dipole: oxygen side of the molecule carries a net negative charge, while the side with the two hydrogen atoms has a net positive electrical charge
- as such, when an electric field (such as a microwave) is applied, the water molecule will rotate and align itself to this applied field

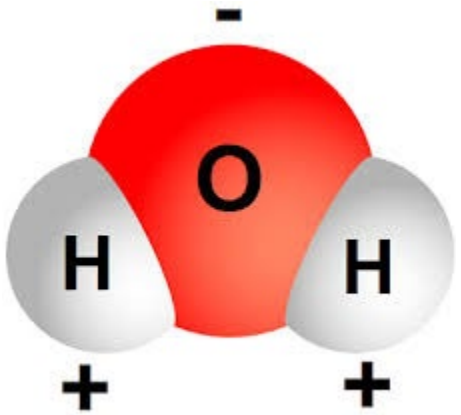


Image Credit: socratic.org/questions/to-which-part-of-a-water-molecule-would-li-be-attracted

What About Water in the Target?

- water (H_2O) is a dipole: oxygen side of the molecule carries a net negative charge, while the side with the two hydrogen atoms has a net positive electrical charge
- as such, when an electric field (such as a microwave) is applied, the water molecule will rotate and align itself to this applied field
- dielectric constant: a measure of the ease with which dipolar molecules rotate in response to an applied field
- dielectric constant (ε): a complex value which characterizes both the permittivity (ε') (real) and conductivity (ε'') (imaginary) of a material

$$\varepsilon = \varepsilon' - j\varepsilon''$$

- real dielectric ranges from ~3 (very dry) to 80 (water)

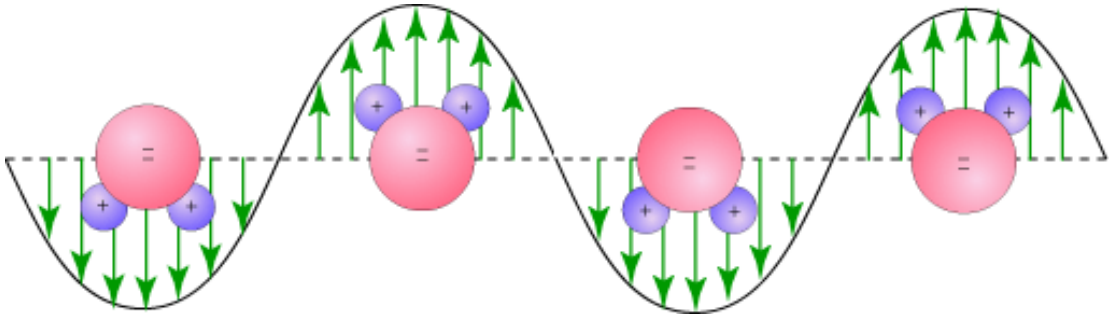


Image Credit: http://macrotomicro.blogspot.com/2011/02/how-microwaveoven-works.html

What Does This Mean for SAR?

- a microwave will continue to propagate until a dielectric discontinuity is encountered, as happens when water is present in the target
- when an electric field is applied, free water molecules (**not tightly bound**) easily rotate to align with the field (positive to negative)
- frictional resistance is low and little energy stored in the rotation is lost when the wave passes and the molecule relaxes. Most of the stored energy is released.
- if many water molecules are present, a significant amount of energy is stored and released. When little water is present, little energy is stored.
- when this stored energy is released, and depending on structure of the target, this energy will be scattered back towards the radar antenna

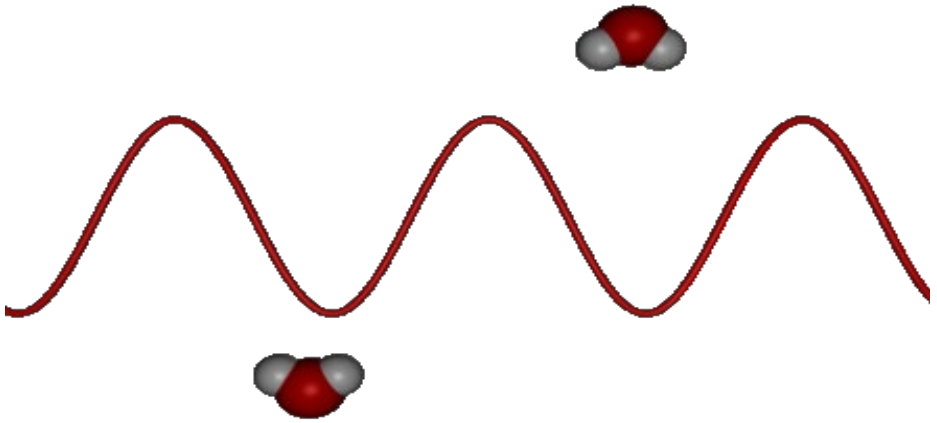


Image Credit: Anton Paar

What Does This Mean for SAR?

- strong positive relationship between real dielectric constant and backscatter
- strong positive relationship between real dielectric constant and soil moisture
- more water in the target = higher backscatter = brighter returns
- applies to ANY target (soil, vegetation etc.)
- Penetration depth (δ_ρ) into soil and/or crops is defined by the dielectric (ε), wavelength (λ), and incident angle
- penetration increases with wavelength and is greater when the target (soil or crops) is drier

Multi-Date RADARSAT-1 Composite Outlook, Saskatchewan (Canada)

$$\delta_p = \frac{\lambda\sqrt{\varepsilon'}}{2\pi\varepsilon''}$$

A Complication: The Environment

Always, always, always check the environmental conditions at the time of image acquisition before using SAR data

Rule 1: Never use SAR if it was raining at the time of the acquisition

- **Why?** Although SAR is considered "all weather" this does not include imaging during rain events as water in the atmosphere will cause SAR scattering. In some regions of world, risks are diurnally dependent.

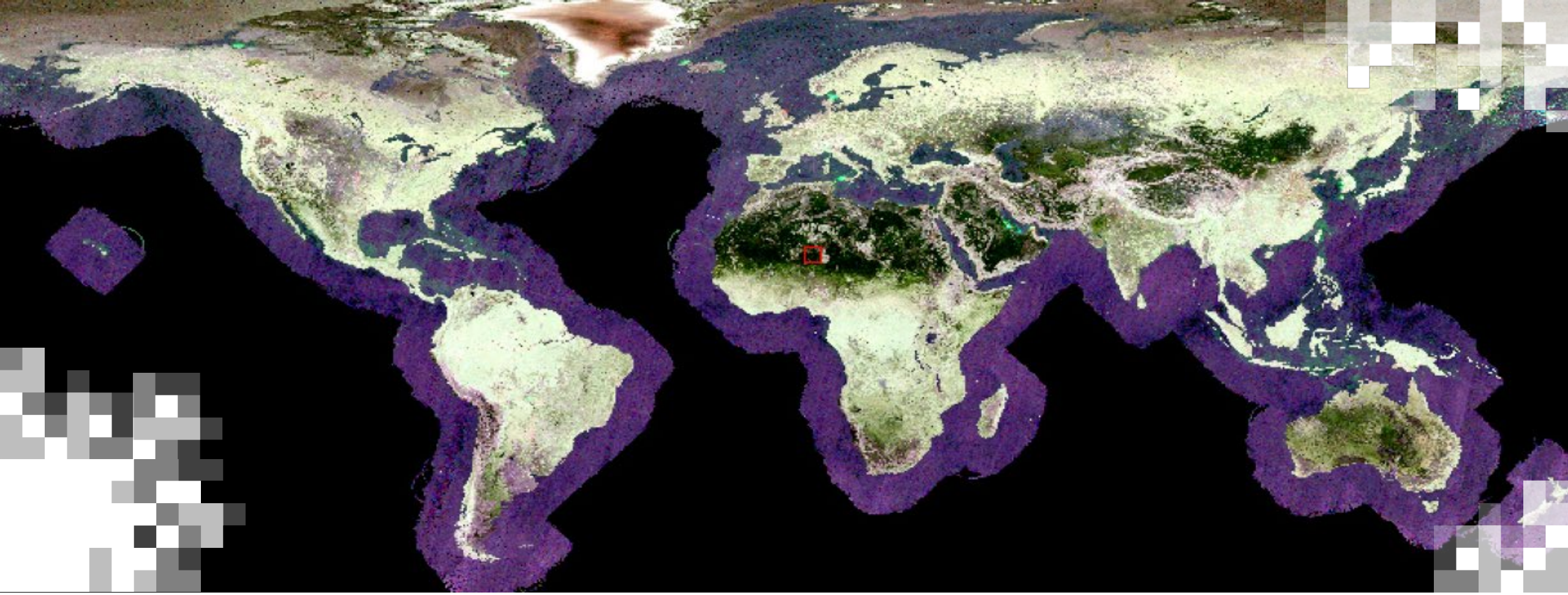
Rule 2: Never use SAR to estimate soil moisture if the ground is frozen

- **Why?** The dielectric constant drops close to zero when water changes to a frozen state. Thus even if there is water in the soil, the SAR will view the soil as dry. SAR can detect freeze/thaw events. Freezing often occurs overnight.

A Complication: The Environment

Rule 3: Consider if dew might be present during early morning acquisitions

- **Why?** Presence of water on leaves will increase backscatter (big problem for biophysical modelling). If water on canopy is significant (just after rain), contrast between targets can be reduced. Dew is most prominent in temperate regions in early morning hours.
- Select orbits (ascending – evening; descending – morning) carefully
- Always check in with meteorological stations



Exploiting SAR to Monitor Agriculture

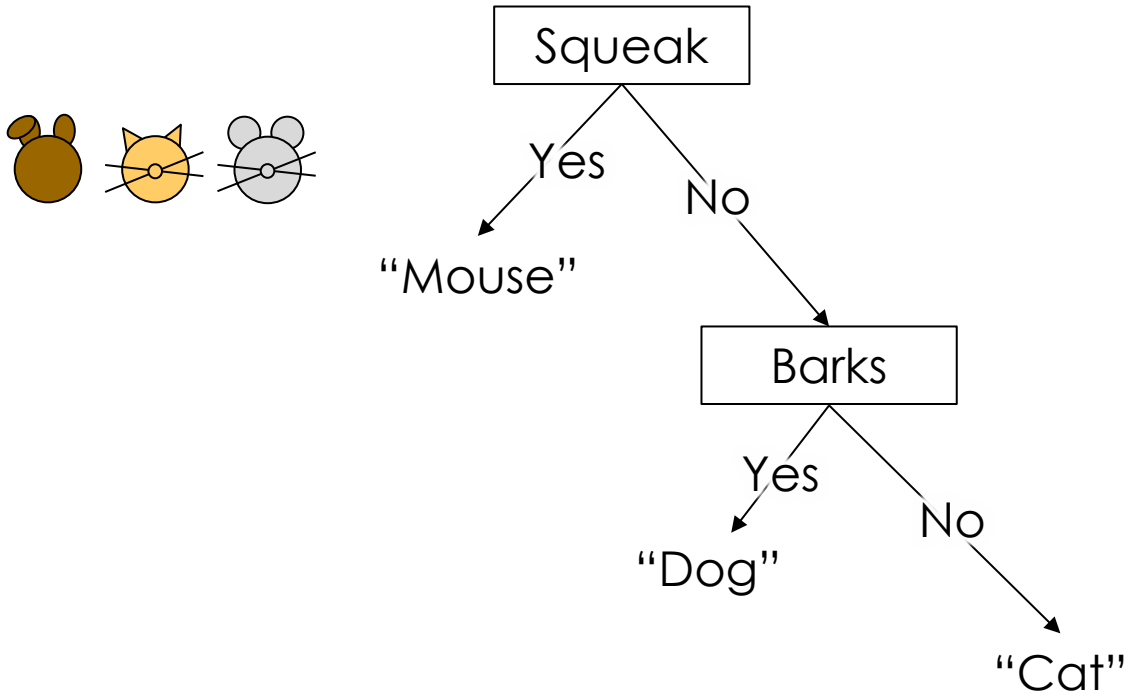
Learning Objectives

By the end of this presentation, you will be able to understand:

- Classification and Regression Trees (CART)
- Random Forests Advantages
- Random Forests:
  - Basics
  - Inputs and parameters
- How to implement Random Forests in R

Classification and Regression Trees (CART): Basics

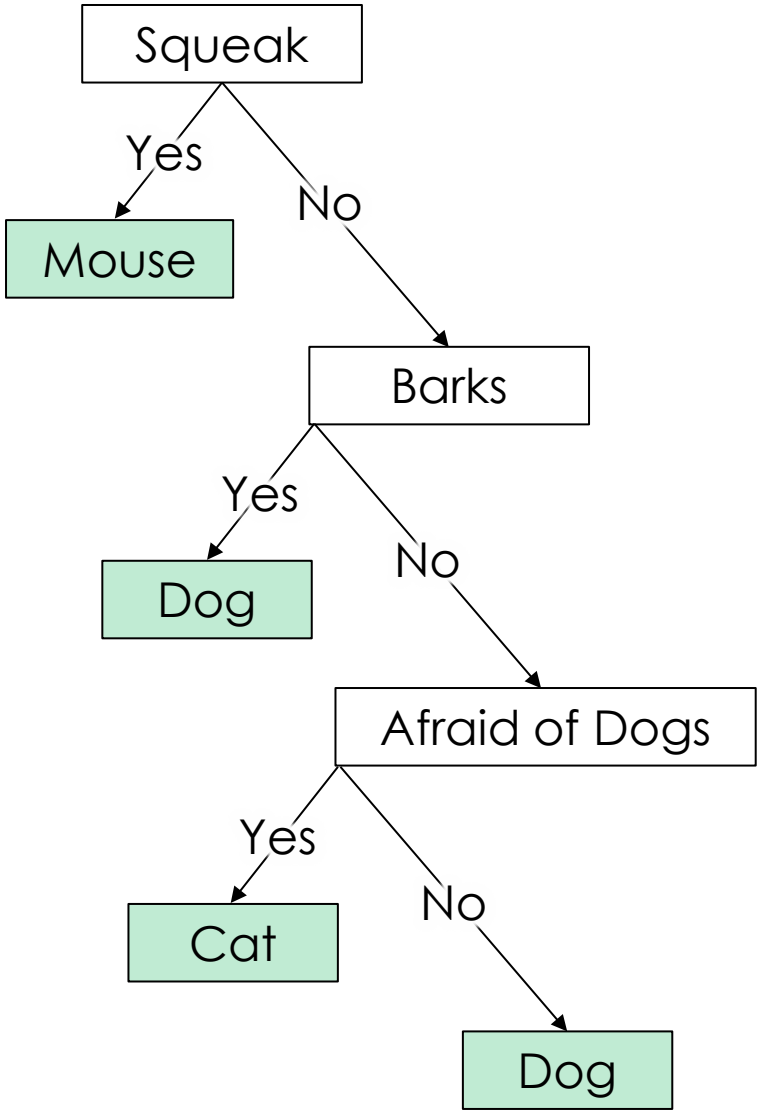
- Partitions or "splits" the data based on a set of binary tests
- Produces an easily-interpretable "tree" of tests and outcomes
- Tries to make groups that are as homogeneous as possible



Classification and Regression Trees (CART): Basics

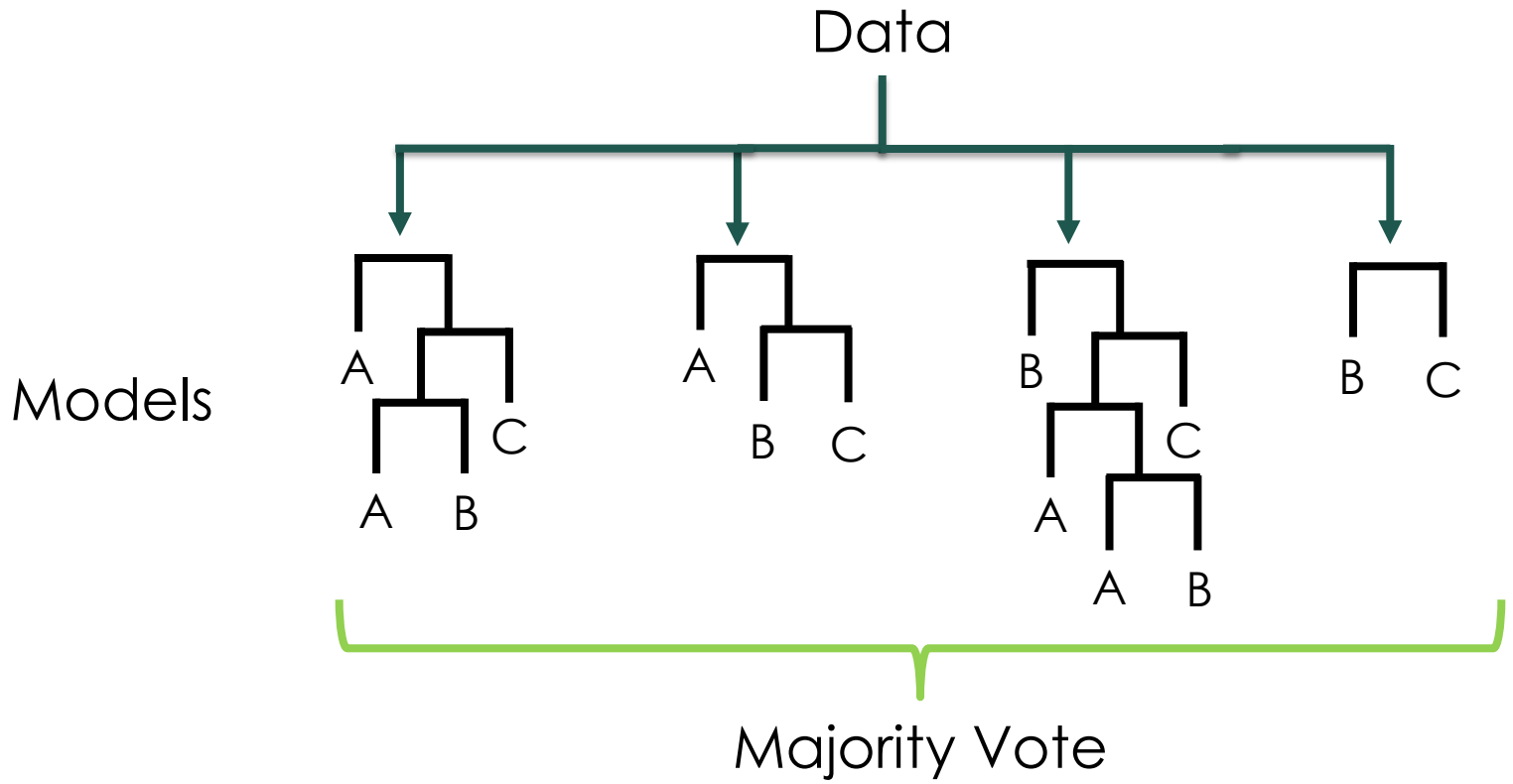
- Important Terms:

	Species	Barks	Pet	Squeeks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	Yes	No	Yes	Yes	No	0.30	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes



Random Forests: Basics

- A "forest" of binary decision trees (Breiman, 2001)
- Works well with high-dimensional datasets (continuous or categorical)
- Ensemble learning technique



Random Forests: Basics

- Random Forests vs. CART

	CART	Random Forests
Number of Trees	1	n>>1
Pruning	Applied	All fully grown
Variables tested for splitting	All	m<<M (all variables)
Training Set	All	≈ 2/3
Accuracy	Independent required	Internally estimated (OOBE)

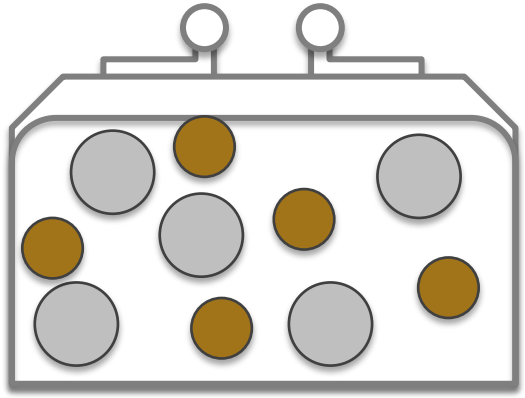
Random Forests: Parameters and Important Terms

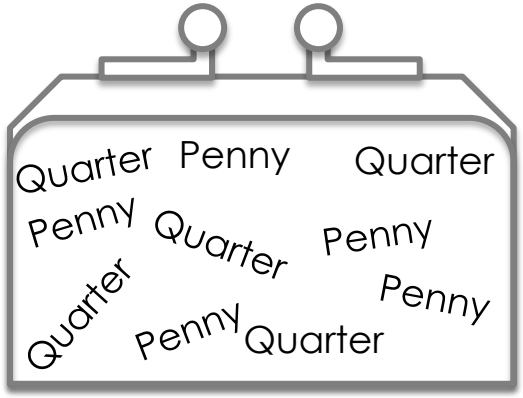
- ntree – number of trees per forest.
- Probability – number of trees that voted with the majority divided by the total number of trees.
- mtry – number of variables tested to determine optimal split at each node.
- Out of Bag Accuracy – internal validation; based on ≈ 1/3 of the dataset not used during the construction of a given tree.
- Mean Decrease in Accuracy (MDA) – quantifies variable "importance" by measuring the change in accuracy after the values of the variable are randomly permuted.

Random Forests: Basics

- Gini Impurity: probability of classifying a data point incorrectly
- Randomly pick data point & randomly classify it according to class distribution
- 25% + 25% = 50%; Gini Impurity = 0.5

Event	Probability
Choose quarter (50%), classify penny (50%)	50% x 50% = 25%
Choose penny (50%), classify quarter (50%)	50% x 50% = 25%

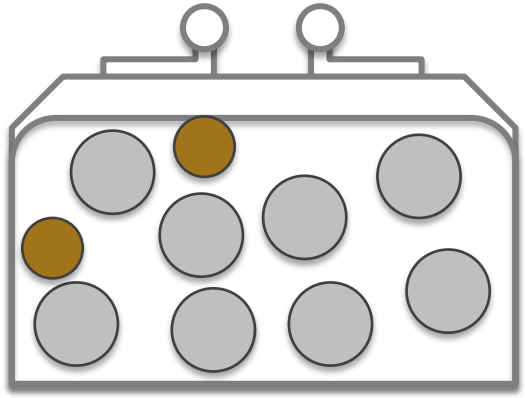


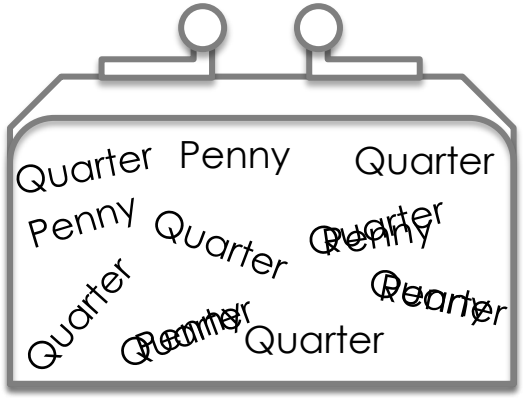


Random Forests: Basics

- Gini Impurity: probability of classifying a data point incorrectly
- Randomly pick data point & randomly classify it according to class distribution
- 16% + 16% = 32%; Gini Impurity = 0.32

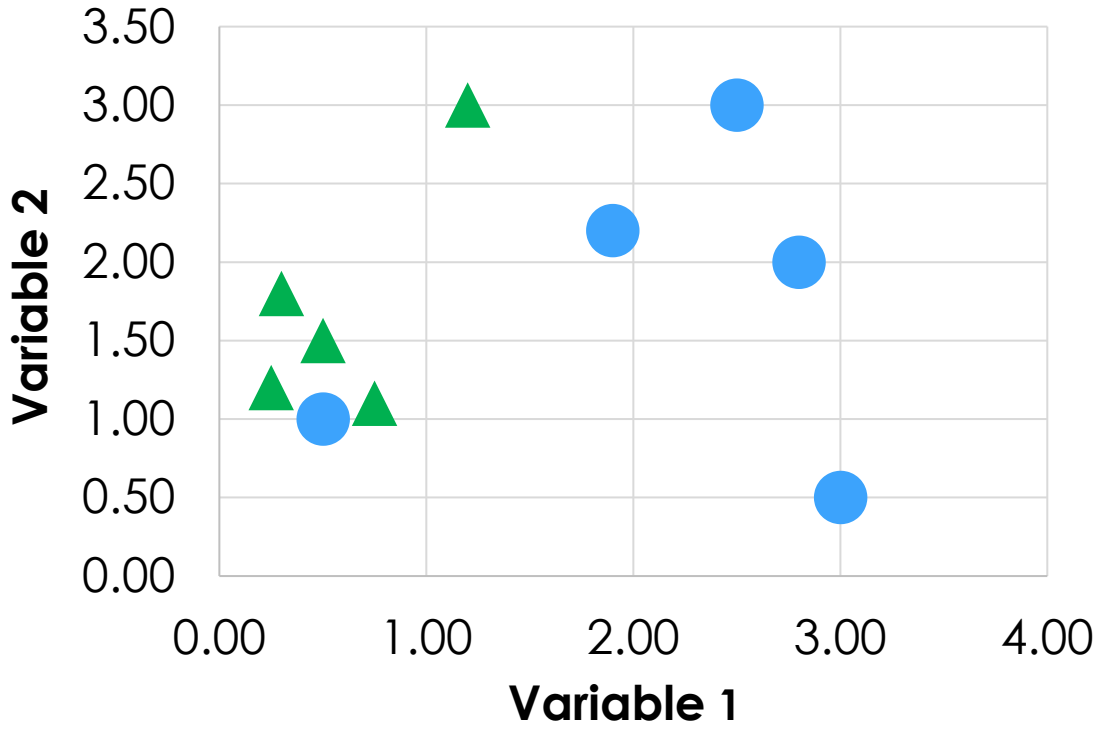
Event	Probability
Choose quarter (80%), classify penny (20%)	80% x 20% = 16%
Choose penny (20%), classify quarter (80%)	20% x 80% = 16%





Random Forests: Basics

- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has



Random Forests: Basics

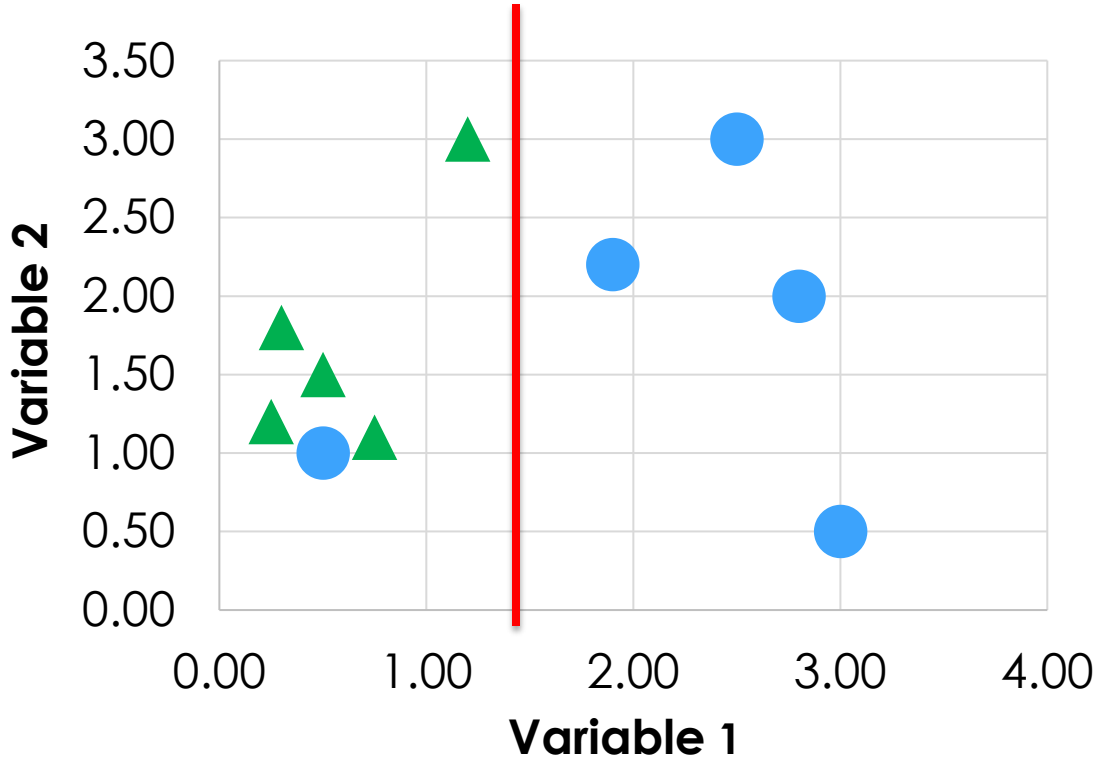
- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has

Gini Impurities

Before the Split= 0.50

Right Note = 0.00

Left Node = 0.28



Random Forests: Basics

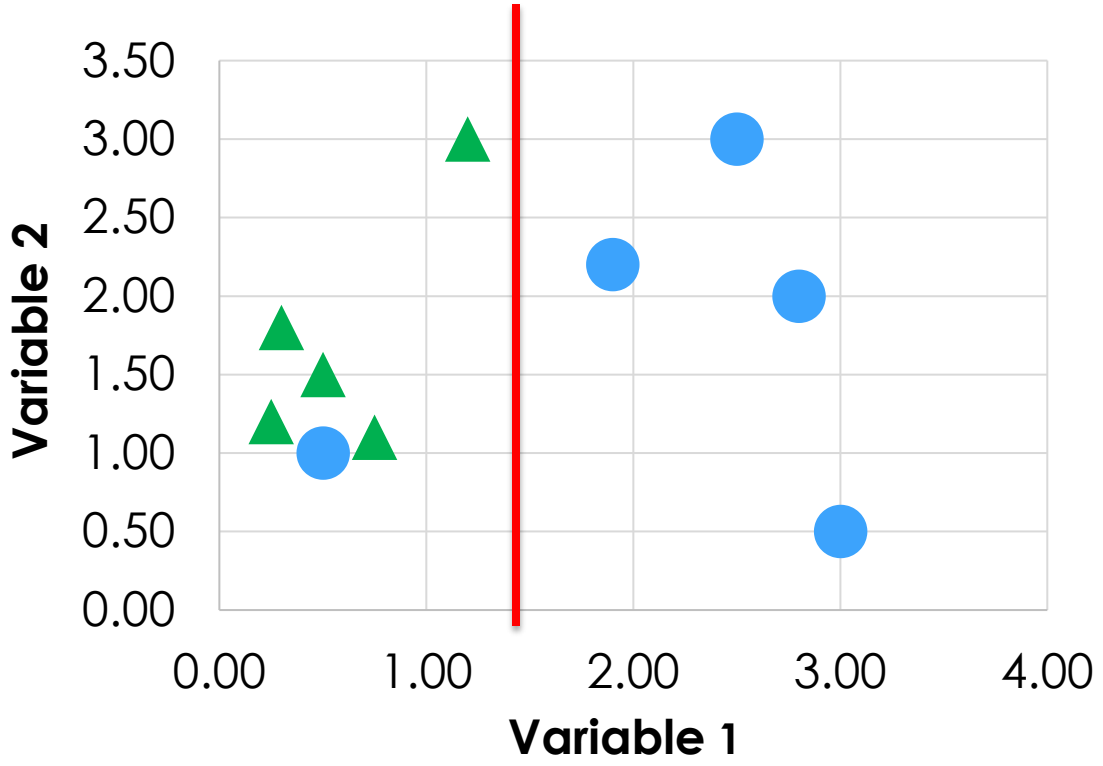
- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has

Gini Impurities

Before the Split= 0.50

Right Note = 0.00

Left Node = 0.28



Right node has 40% data points, left node has 60%

(0.40*0.00) + (0.60*0.28) = 0.17

With this split, the amount of impurity removed is:

0.5 – 0.17 = 0.33

Random Forests: Basics

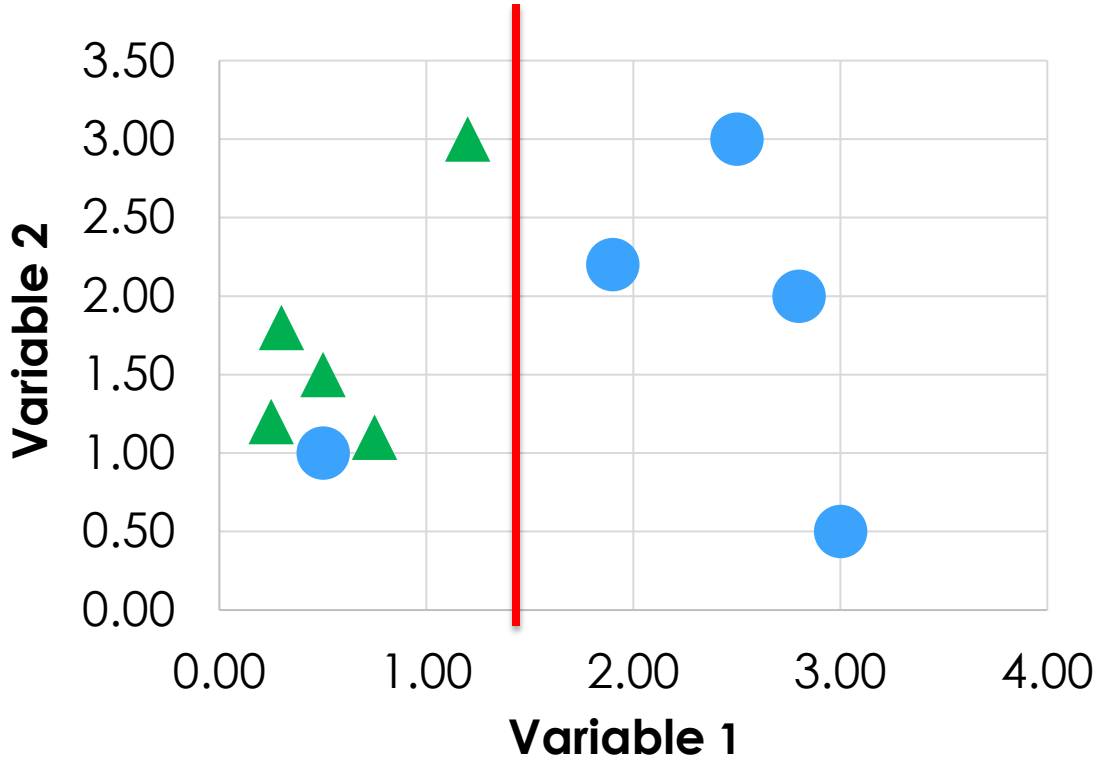
- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has

Gini Impurities

Before the Split= 0.50

Right Note = 0.00

Left Node = 0.28



Right node has 40% data points, left node has 60%

(0.40*0.00) + (0.60*0.28) = 0.17

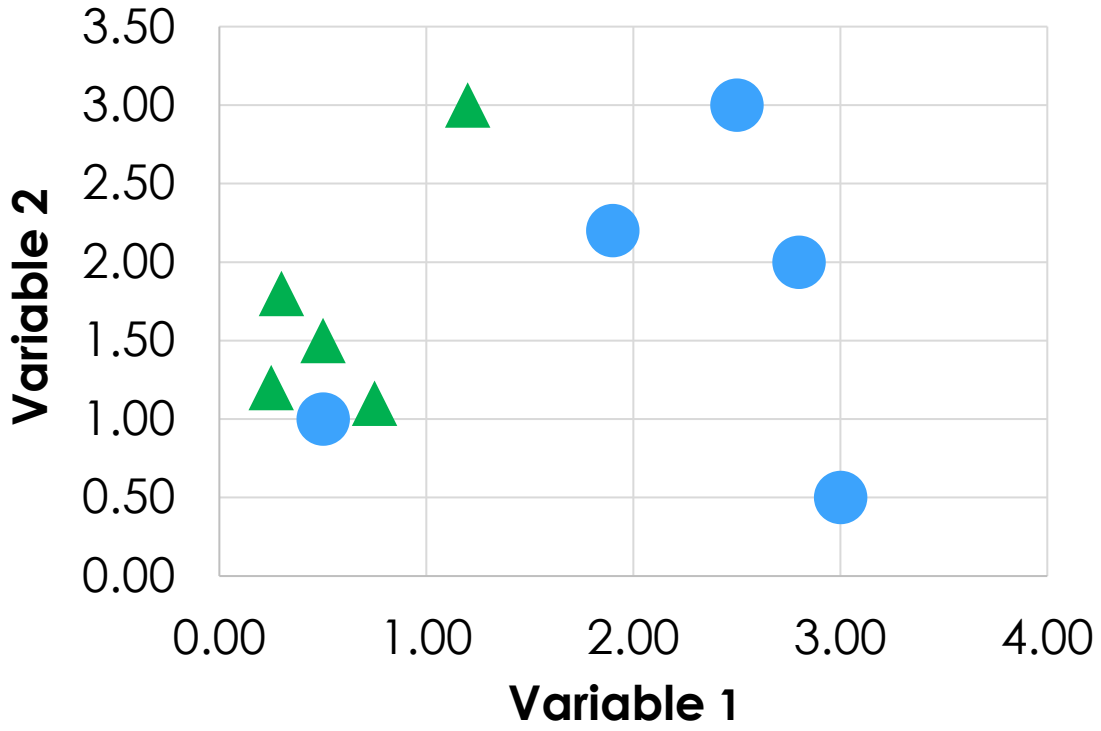
With this split, the amount of impurity removed is:

0.5 – 0.17 = 0.33

GINI GAIN

Random Forests: Basics

- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has



Random Forests: Basics

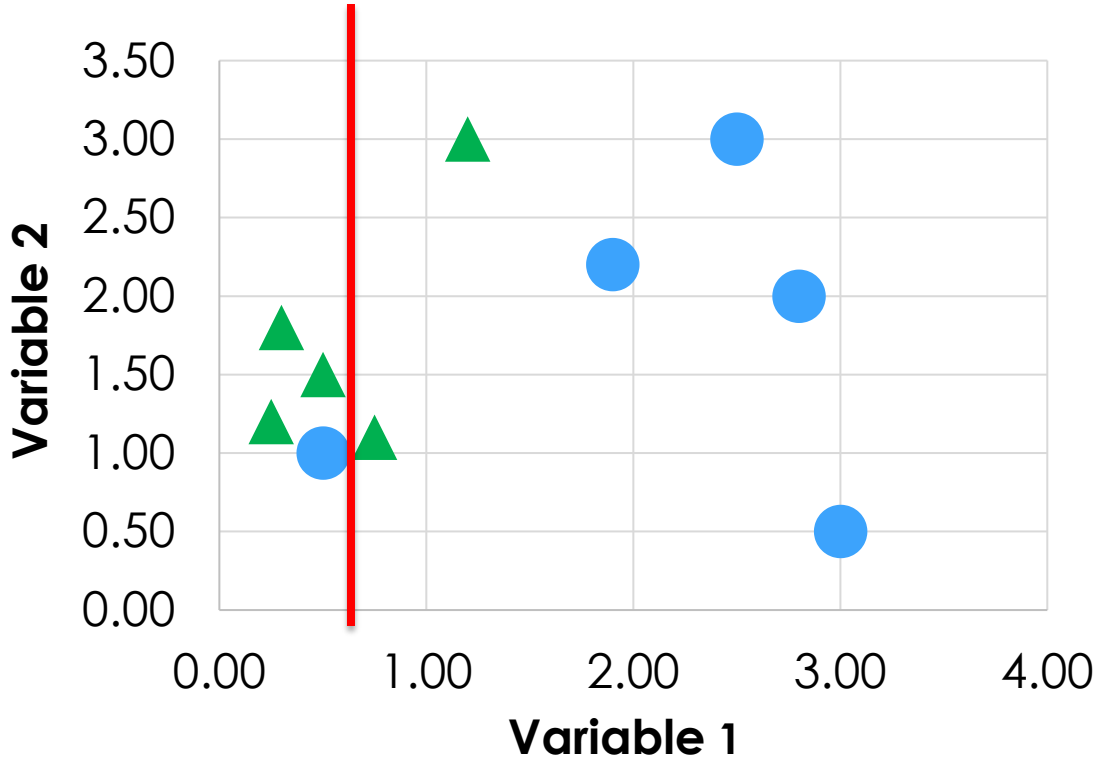
- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has

Gini Impurities

Before the Split= 0.50

Right Note = 0.44

Left Node = 0.38



Random Forests: Basics

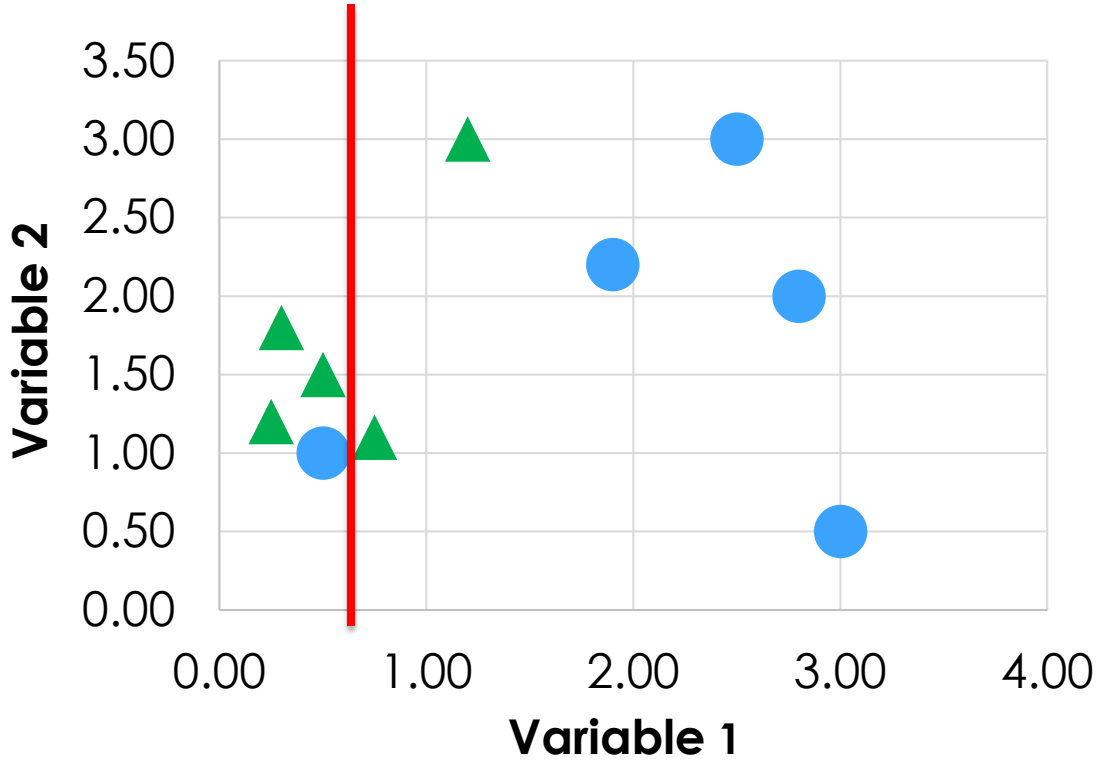
- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has

Gini Impurities

Before the Split= 0.50

Right Note = 0.44

Left Node = 0.38



Right node has 60% data points, left node has 40%

(0.60*0.38) + (0.40*0.44)=0.42

With this split the amount of impurity removed is:

0.5 – 0.42 = 0.08

Random Forests: Basics

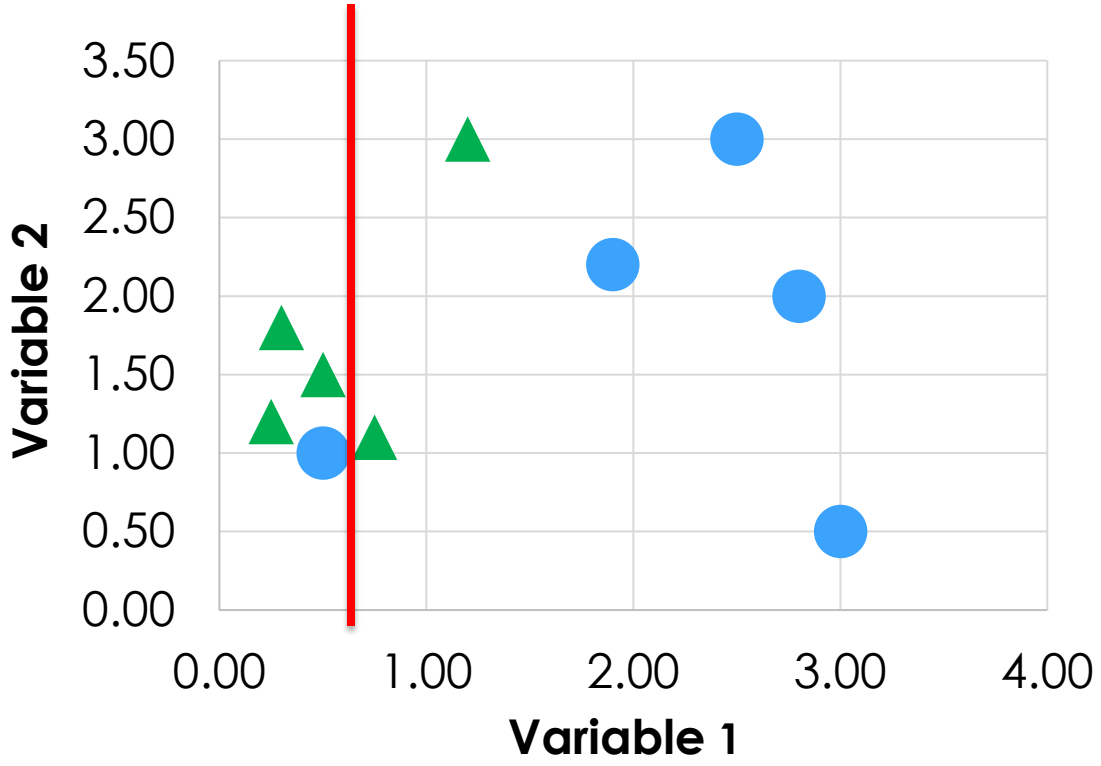
- How Gini Impurity is used for splitting:
  - Don't know where the perfect split is, but can test all possible splits
  - Determine split quality by weighting impurity of subsequent nodes by number of data points it has

Gini Impurities

Before the Split= 0.50

Right Note = 0.44

Left Node = 0.38



Right node has 60% data points, left node has 40%

(0.60*0.38) + (0.40*0.44)=0.42

With this split the amount of impurity removed is:

0.5 – 0.42 = 0.08

GINI GAIN

Random Forests: Basics

- Random Forests Inputs:
- Labelled classes with associated predictor variables
  - Categorical or continuous

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Random Forests: Example

1. Create the training data to grow the tree
   - For N number of data points, randomly sample N cases (with replacement)

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
23	Dog	Yes	No	No	No	No	No	0.52	0.26	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.21	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.21	Yes
14						Yes	Yes	0.36	0.17	Yes
16						Yes	Yes	0.3	0.16	Yes
16						Yes	Yes	0.3	0.16	Yes
16						Yes	Yes	0.3	0.16	Yes
18						Yes	Yes	0.33	0.22	Yes
18						Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	Yes	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Remaining for internal accuracy assessment ≈ 1/3

Random Forests: Example

2. For M number of variables, randomly select a subset (m<<M) to determine how each node is split (**mtry**)

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Mouse	Cat	Dog
Yes	8	0	0
No	0	8	11

Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Mouse	Cat	Dog
Yes	0	7	0
No	8	1	11

Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32
Meows	Yes or No	0.28

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Mouse	Cat	Dog
<=0.05	1	0	0
>0.05	7	8	11

Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32
Meows	Yes or No	0.28
Height	<=0.05	0.03

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Mouse	Cat	Dog
<=0.06	3	0	0
>0.06	5	8	11

Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32
Meows	Yes or No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Mouse	Cat	Dog
<=0.07	5	0	0
>0.07	3	8	11

Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32
Meows	Yes or No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09
Height	<=0.07	0.17

...

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32
Meows	Yes or No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09
Height	<=0.07	0.17
Height	<=0.08	0.26
Height	<=0.09	0.28
Height	<=0.15	0.25
Height	<=0.16	0.26
Height	<=0.17	0.28
Height	<=0.22	0.27
Height	<=0.26	0.31
Height	<=0.27	0.22
Height	<=0.29	0.18
Height	<=0.32	0.09
Height	<=0.33	0.04
Height	<=0.35	0.00

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32
Meows	Yes or No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09
Height	<=0.07	0.17
Height	<=0.08	0.26
Height	<=0.09	0.28
Height	<=0.15	0.25
Height	<=0.16	0.26
Height	<=0.17	0.28
Height	<=0.22	0.27
Height	<=0.26	0.31
Height	<=0.27	0.22
Height	<=0.29	0.18
Height	<=0.32	0.09
Height	<=0.33	0.04
Height	<=0.35	0.00

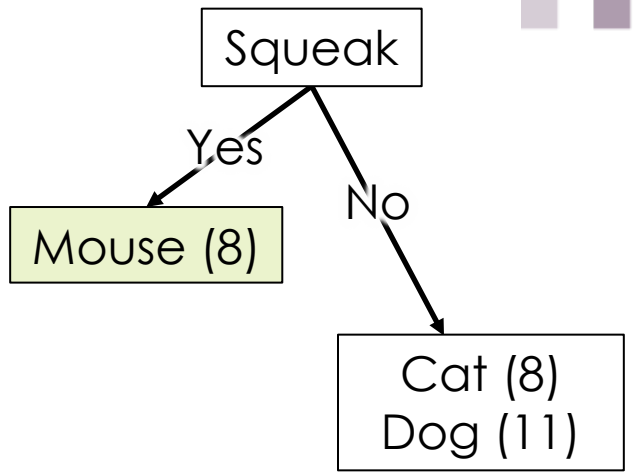
Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Gini Gain to Identify Optimal Split

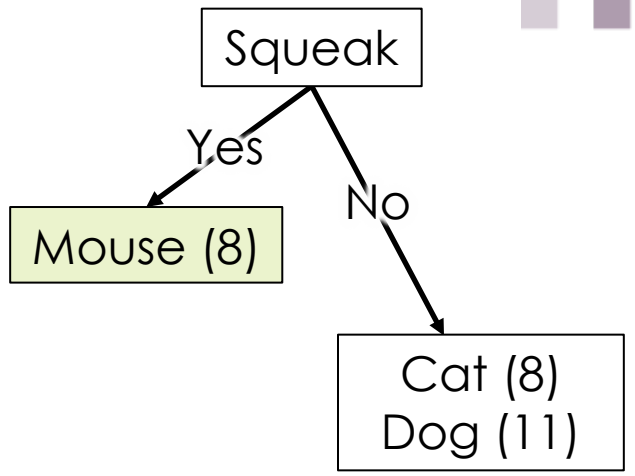
Variable	Evaluated Split Points	Gini
Squeaks	Yes or No	0.32
Meows	Yes or No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09
Height	<=0.07	0.17
Height	<=0.08	0.26
Height	<=0.09	0.28
Height	<=0.15	0.25
Height	<=0.16	0.26
Height	<=0.17	0.28
Height	<=0.22	0.27
Height	<=0.26	0.31
Height	<=0.27	0.22
Height	<=0.29	0.18
Height	<=0.32	0.09
Height	<=0.33	0.04
Height	<=0.35	0.00



Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes



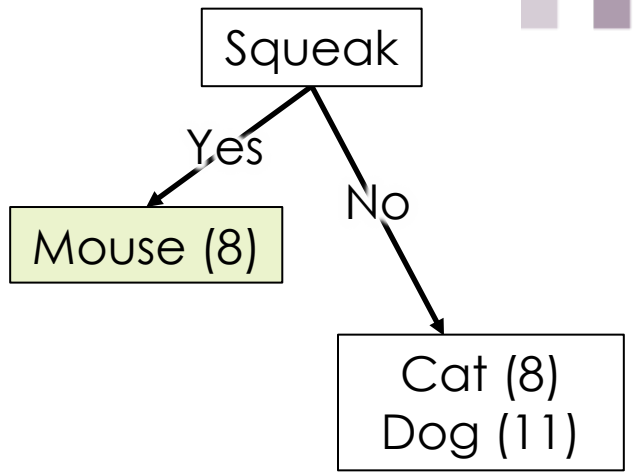
Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Cat	Dog
Yes	1	0
No	7	11



Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Cheese	Yes or No	0.04

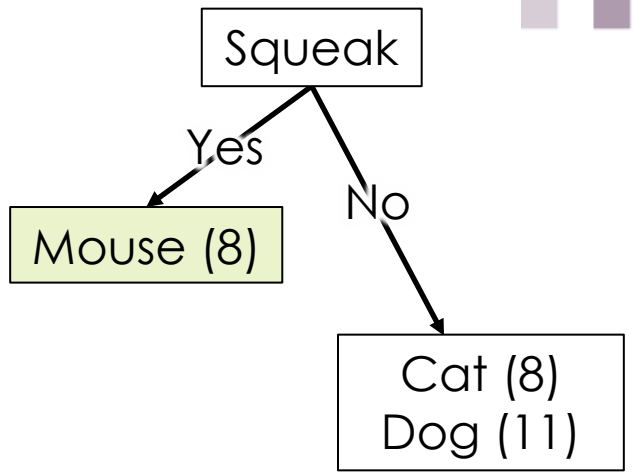
Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Cat	Dog
Yes	0	8
No	8	3



Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Cheese	Yes or No	0.04
Barks	Yes or No	0.26

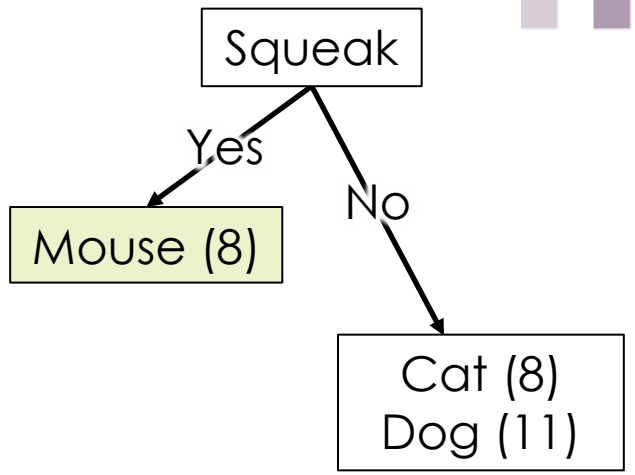
Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Classification Results Based on Split Points

	Cat	Dog
Yes	8	10
No	0	1



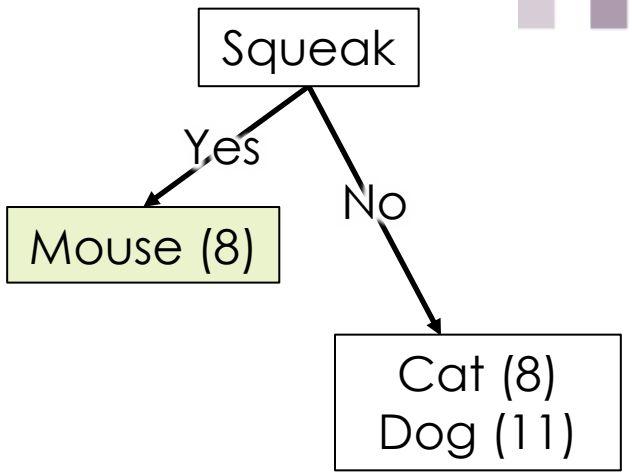
Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Cheese	Yes or No	0.04
Barks	Yes or No	0.26
Collar	Yes or No	0.02

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes



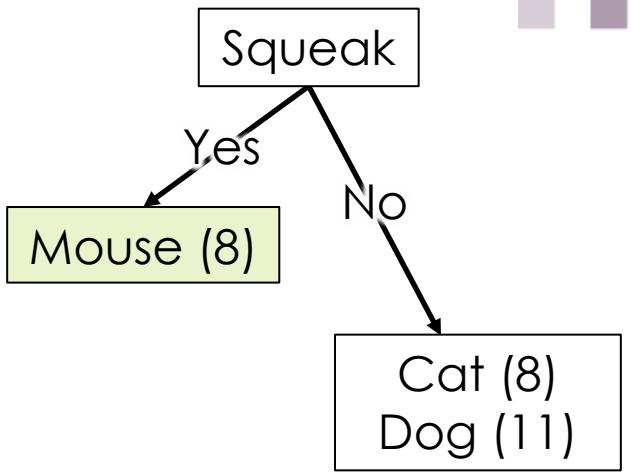
Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Cheese	Yes or No	0.04
Barks	Yes or No	0.26
Collar	Yes or No	0.02

Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes



Gini Gain to Identify Optimal Split

Variable	Evaluated Split Points	Gini
Cheese	Yes or No	0.04
Barks	Yes or No	0.26
Collar	Yes or No	0.02

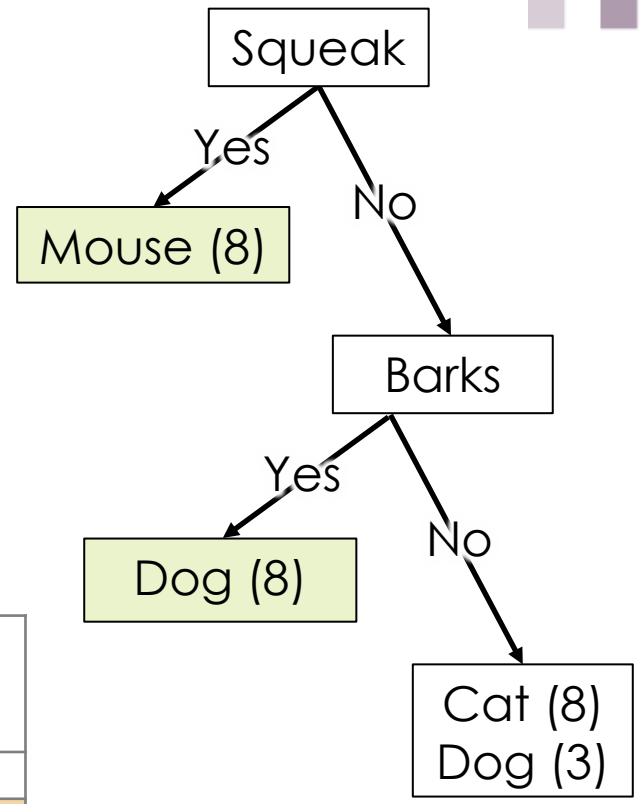
Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes

Gini Gain to Identify Optimal Split

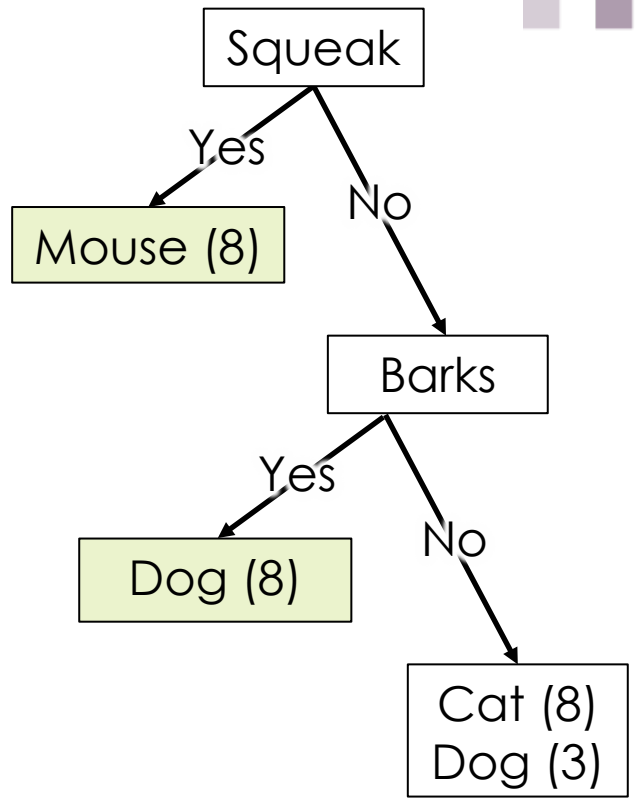
Variable	Evaluated Split Points	Gini
Cheese	Yes or No	0.04
Barks	Yes or No	0.26
Collar	Yes or No	0.02



Random Forests: Example

Training Set

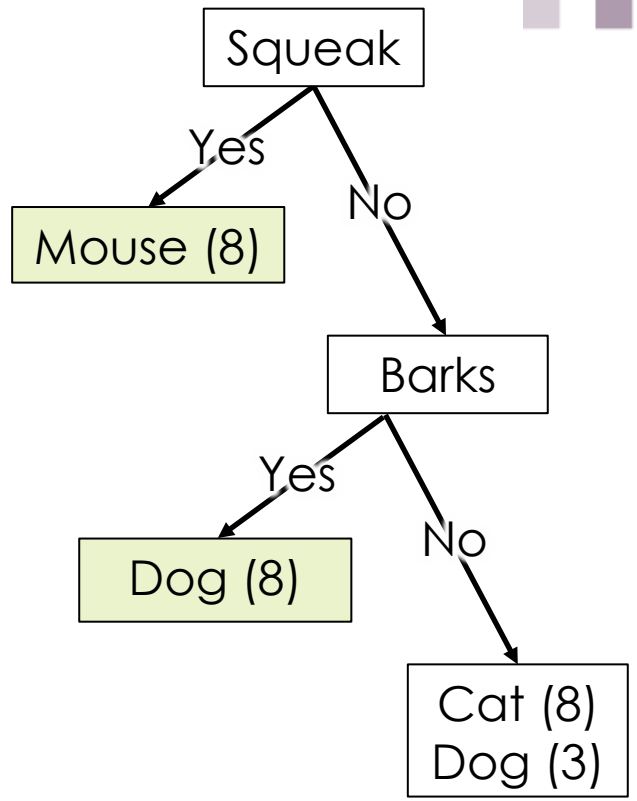
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes



Random Forests: Example

Training Set

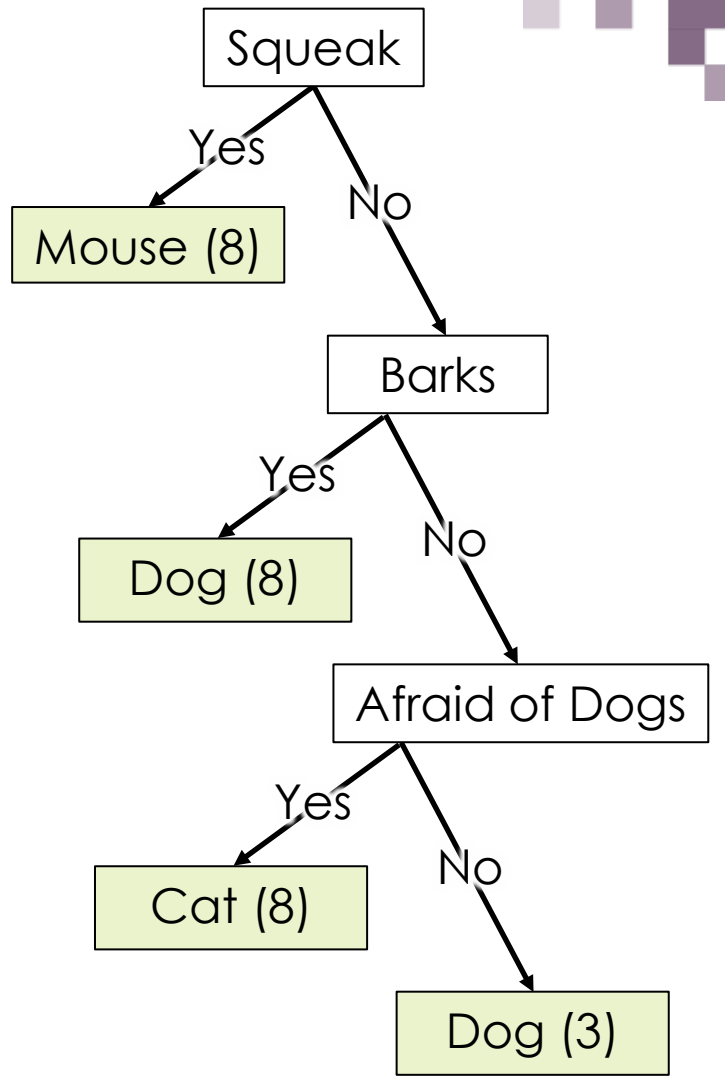
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes



Random Forests: Example

Training Set

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes
27	Dog	No	Yes	No	No	Yes	No	0.58	0.29	Yes



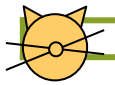
Random Forests: Example

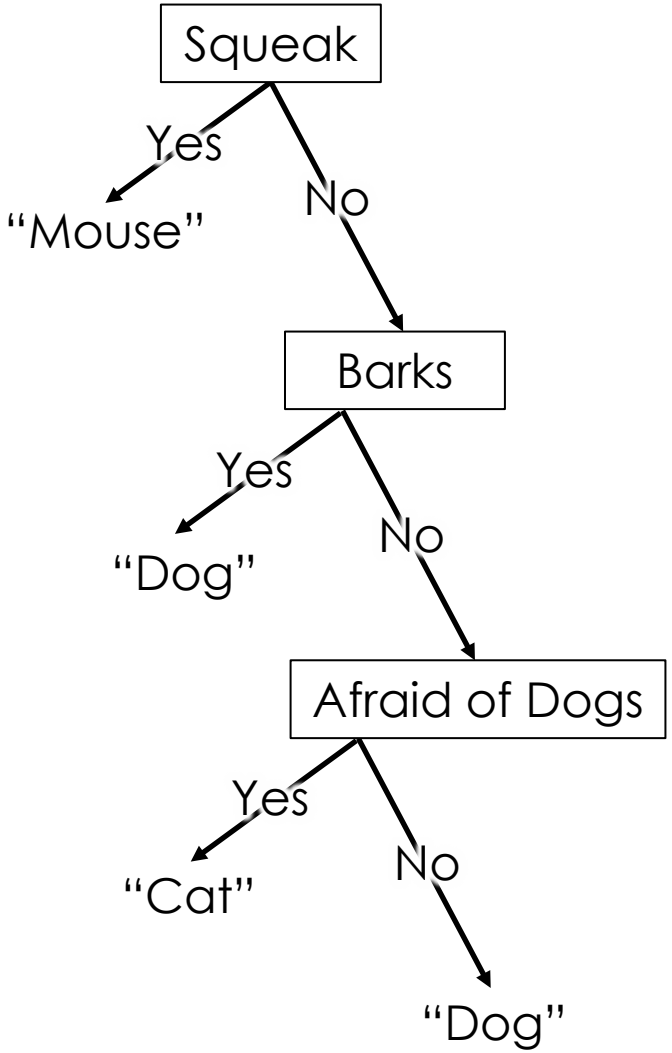
3. Take the mode prediction of all trees (**ntree**) to determine the final classification

Random Forests: Example

Original

Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
Cat	No	Yes	No	Yes	Yes	No	0.35	0.24	Yes



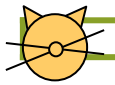


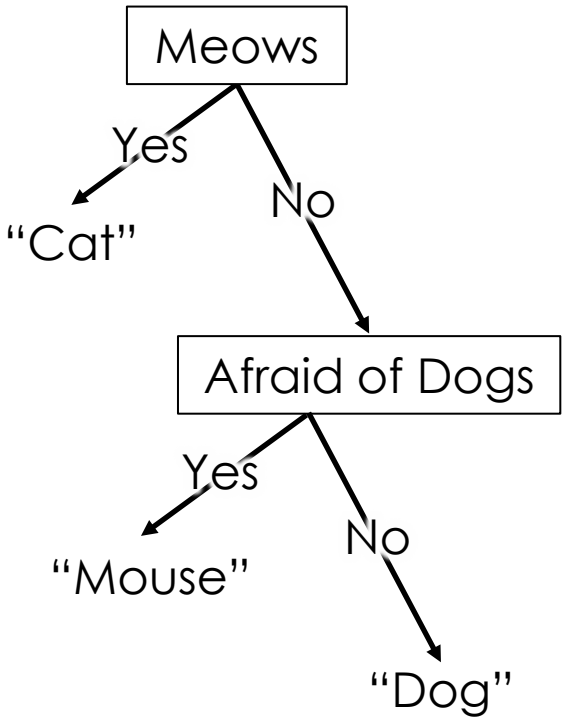
Tree	Vote
1	Dog

Random Forests: Example

Original

Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
Cat	No	Yes	No	Yes	Yes	No	0.35	0.24	Yes



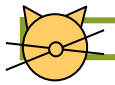


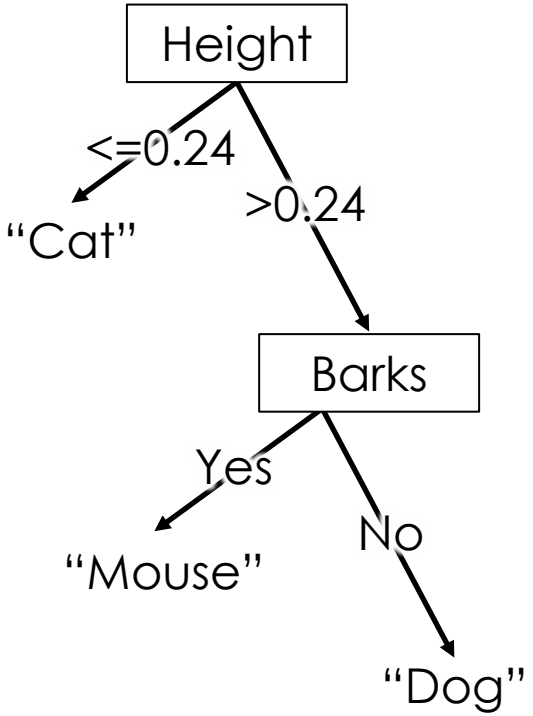
Tree	Vote
1	Dog
2	Cat

Random Forests: Example

Original

Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
Cat	No	Yes	No	Yes	Yes	No	0.35	0.24	Yes





Tree	Vote
1	Dog
2	Cat
3	Cat

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
17	Cat	No	Yes	No	Yes	Yes	No	0.35	0.24	Yes

Tree	Vote
1	Dog
2	Cat
3	Cat
4	Mouse
5	Cat
6	Cat
7	Cat
8	Cat
9	Dog

Majority = Cat

Final Classification= Cat

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
17	Cat	No	Yes	No	Yes	Yes	No	0.35	0.24	Yes

Tree	Vote
1	Dog
2	Cat
3	Cat
4	Mouse
5	Cat
6	Cat
7	Cat
8	Cat
9	Cat

Probability:
=2/3
=67%

Random Forests: Example

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
17	Cat	No	Yes	No	Yes	Yes	No	0.35	0.24	Yes

ntree = 9

Tree	Vote
1	Dog
2	Cat
3	Cat
4	Mouse
5	Cat
6	Cat
7	Cat
8	Cat
9	Cat

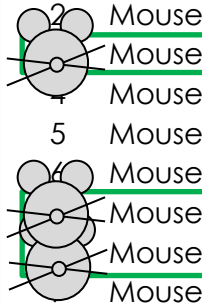
Random Forests: Example

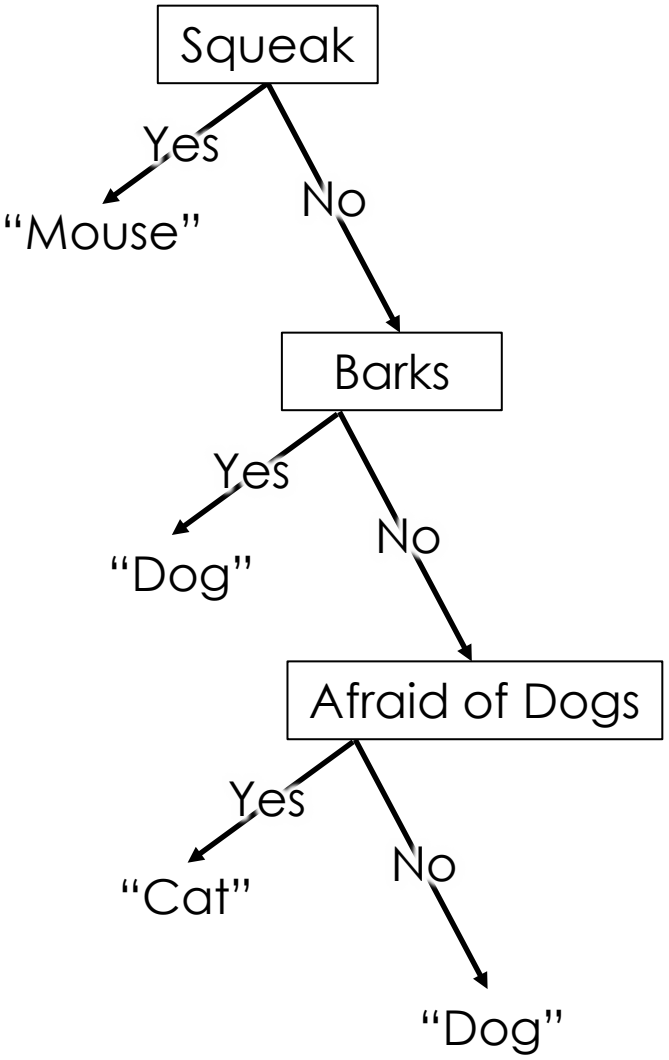
- Run the Out of Bag Samples down the tree and calculate the Out of Bag Error (for a given tree)

Random Forests: Example

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Remaining for internal accuracy assessment ≈ 1/3

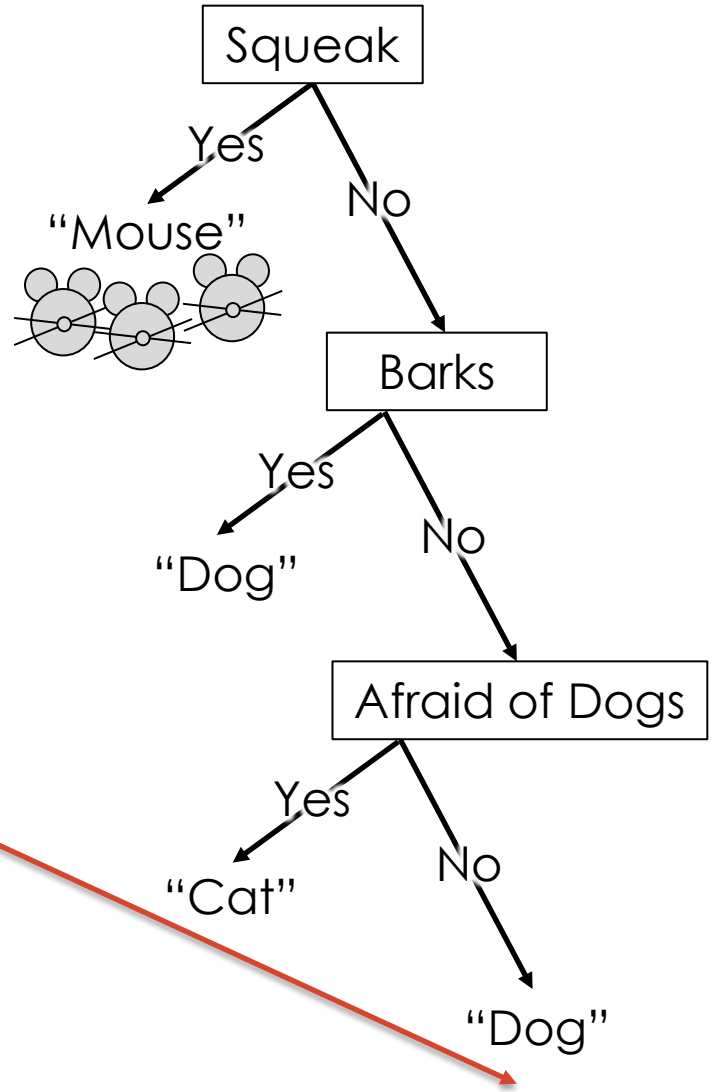




Random Forests: Example

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

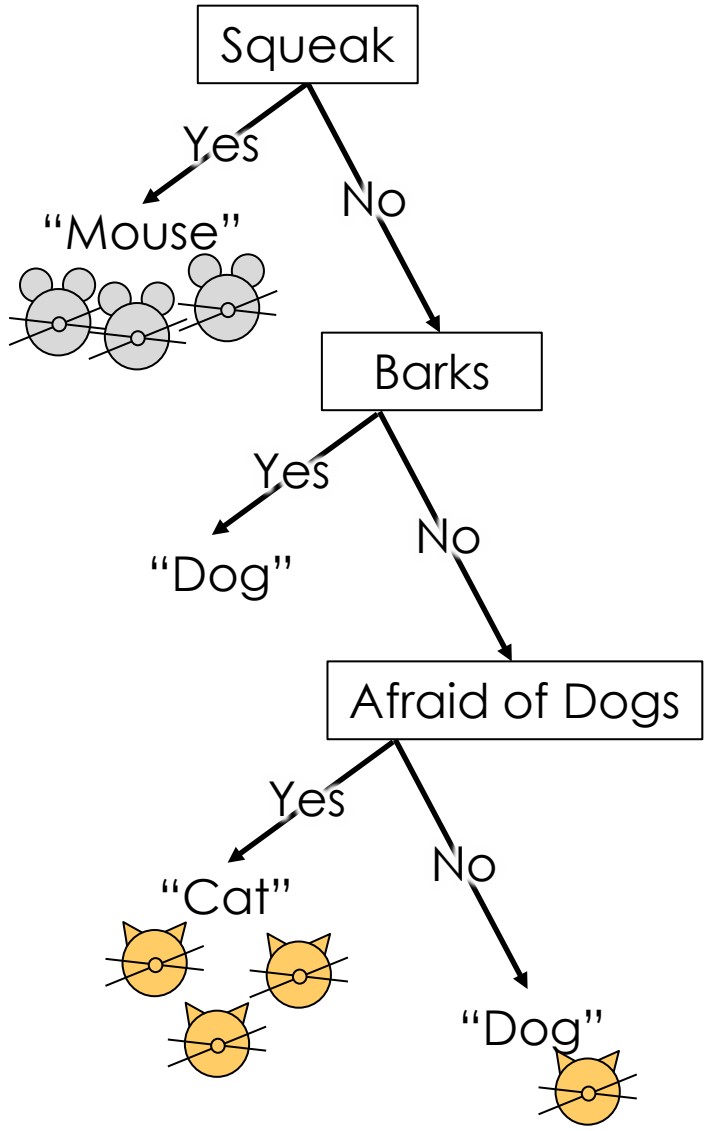
Remaining for internal accuracy assessment ≈ 1/3



Random Forests: Example

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

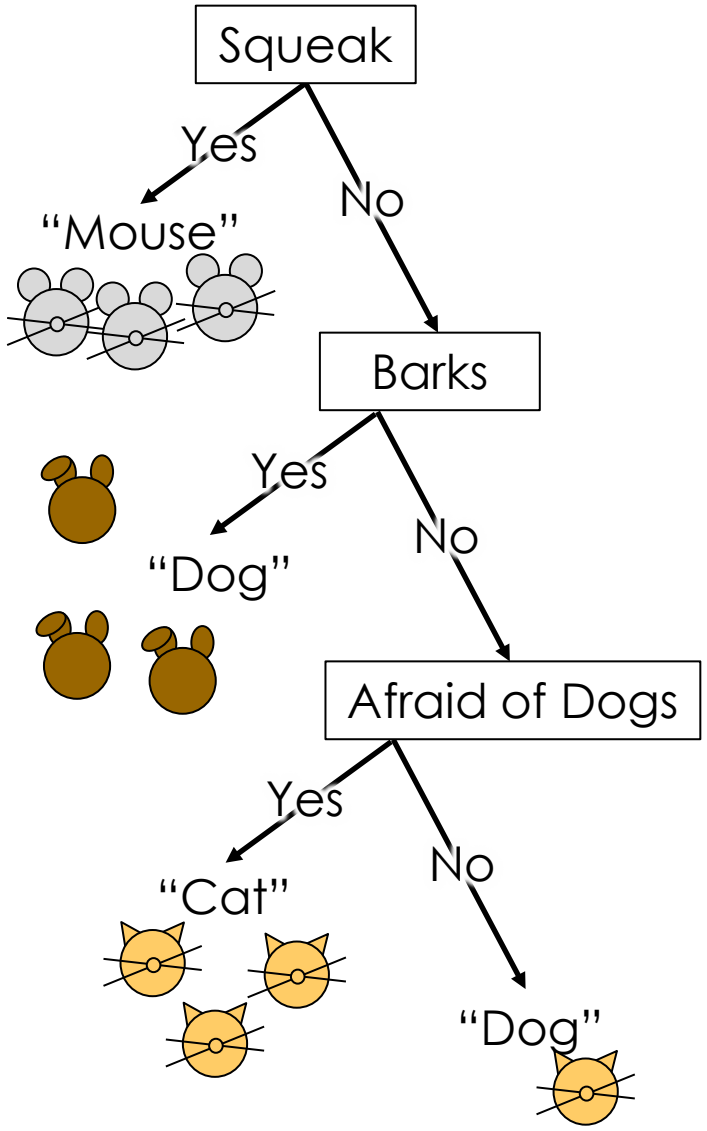
Remaining for internal accuracy assessment ≈ 1/3



Random Forests: Example

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

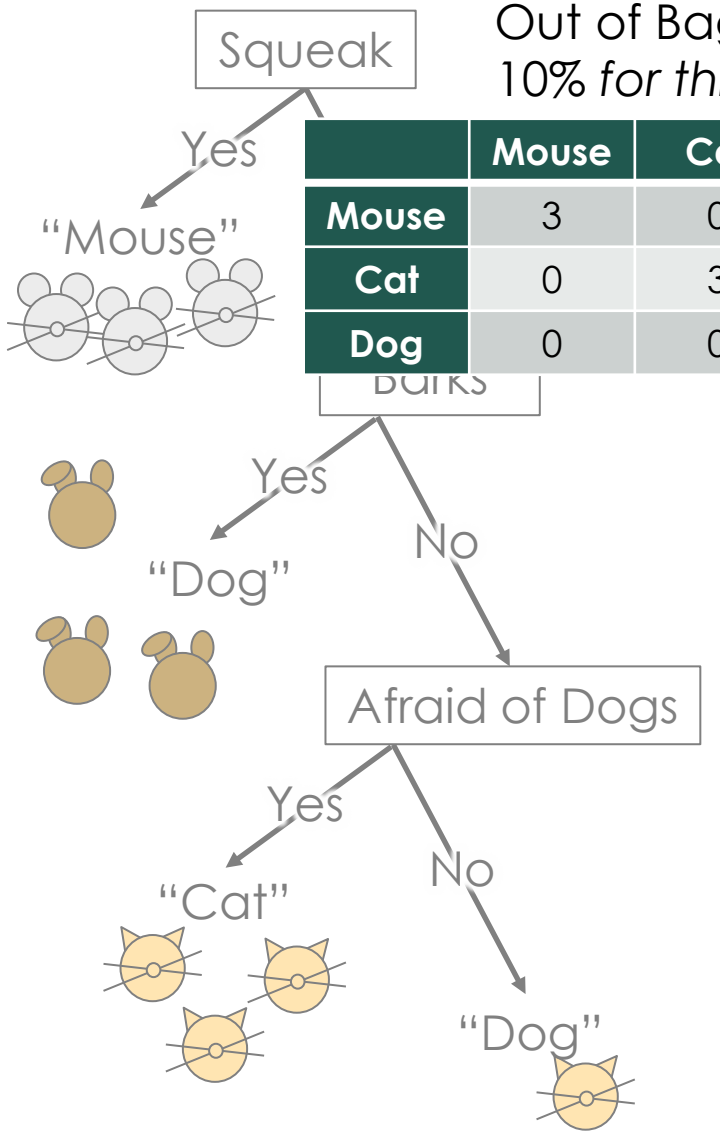
Remaining for internal accuracy assessment ≈ 1/3



Random Forests: Example

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	No	No	Yes	No	No	Yes	0.11	0.05	Yes
7	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
8	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
9	Mouse	No	No	Yes	No	No	Yes	0.15	0.08	Yes
10	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
11	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
12	Cat	No	Yes	No	Yes	Yes	Yes	0.40	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
15	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
17	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
18	Cat	No	Yes	No	Yes	Yes	Yes	0.33	0.22	Yes
19	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
20	Dog	Yes	Yes	No	No	Yes	No	0.53	0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No	0.51	0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No	0.16	0.32	Yes
23	Dog	No	No	No	No	No	No	0.52	0.26	Yes
24	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
25	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
26	Dog	Yes	Yes	No	No	Yes	No	0.53	0.29	Yes
27	Dog	Yes	Yes	No	No	Yes	No	0.58	0.29	Yes

Remaining for internal accuracy assessment ≈ 1/3

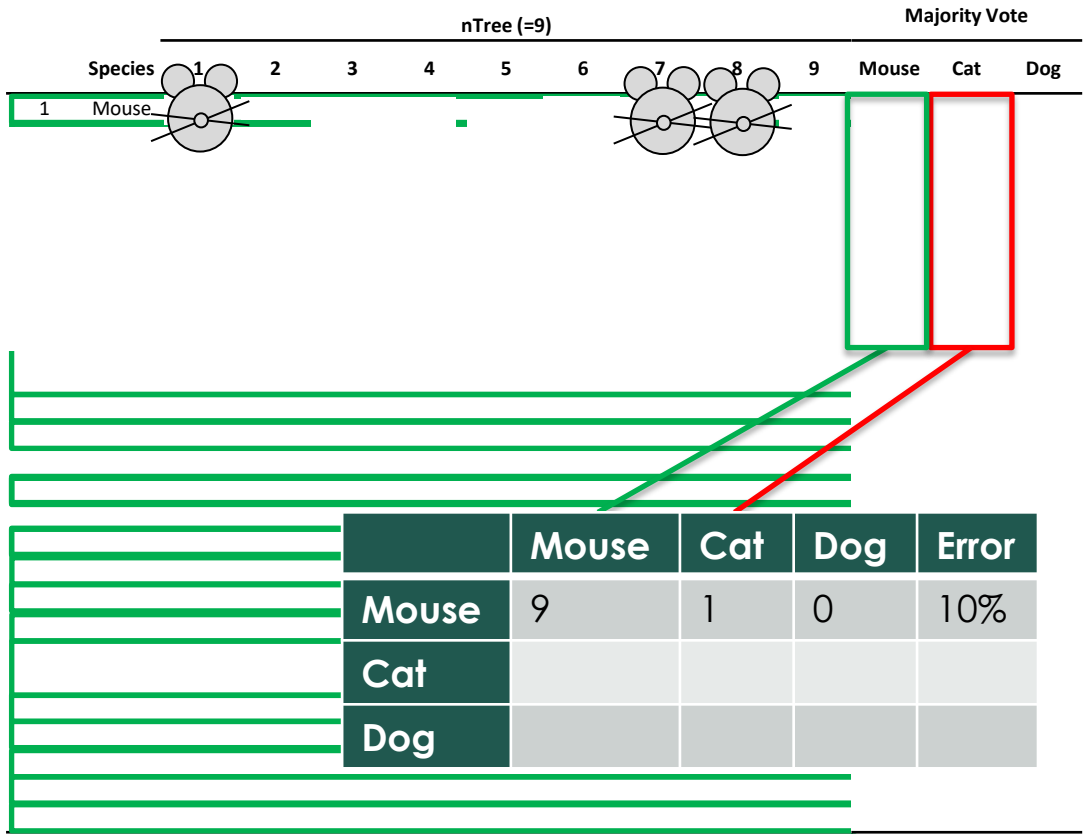


Out of Bag Error is 10% *for this tree…*

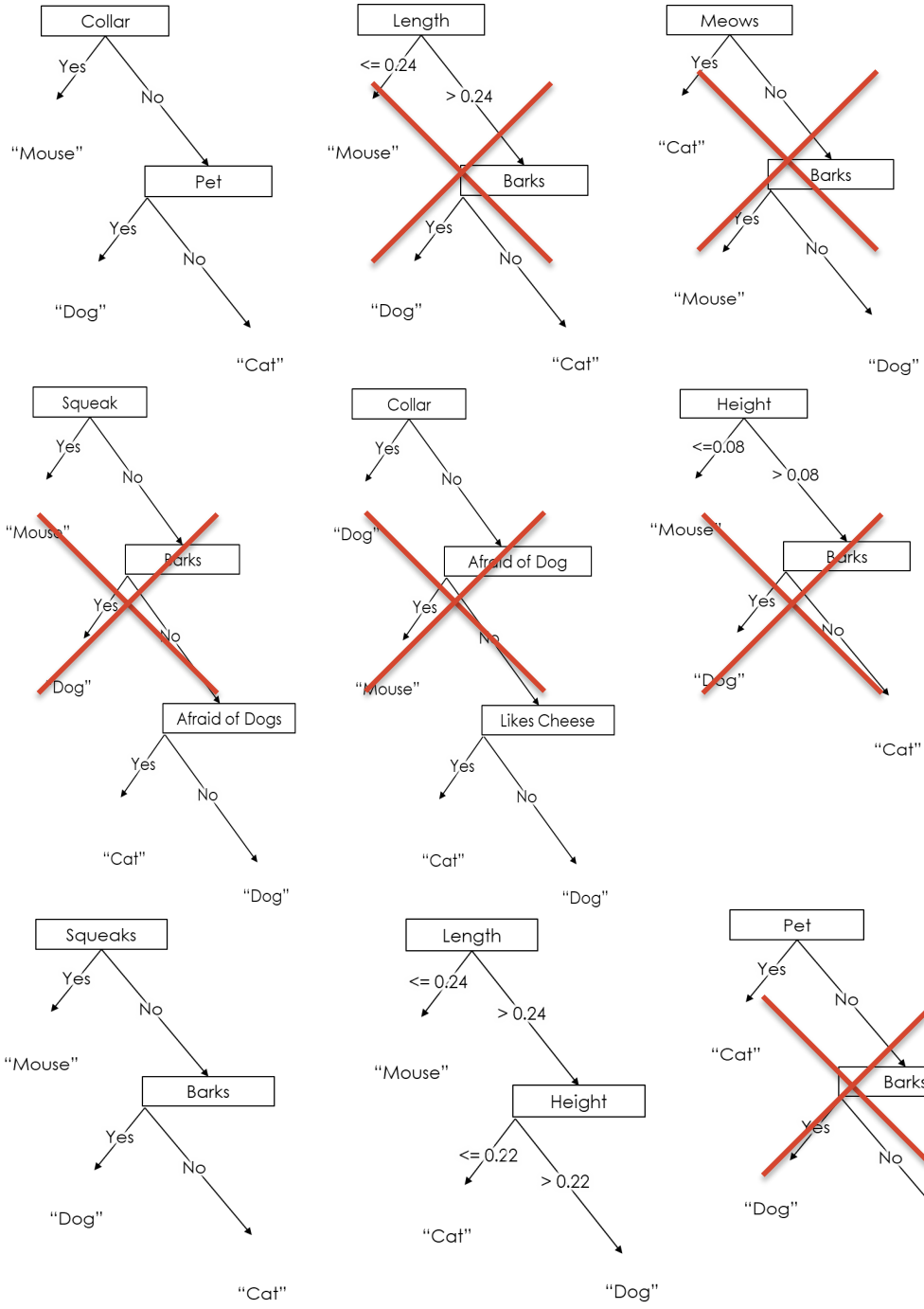
	Mouse	Cat	Dog
Mouse	3	0	0
Cat	0	3	1
Dog	0	0	3

Random Forests: Example

- Calculate the overall Out of Bag Error



	Mouse	Cat	Dog	Error
Mouse	9	1	0	10%
Cat				
Dog				



Random Forests: Example

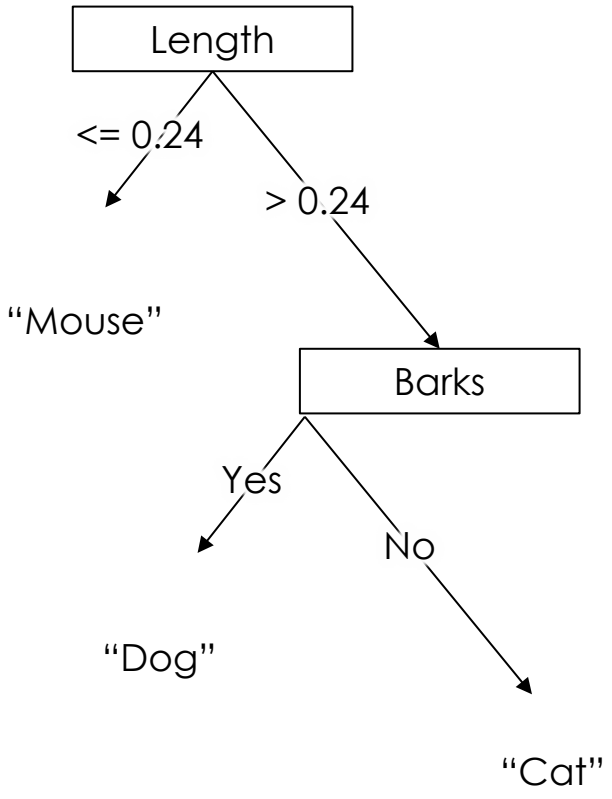
5. Calculating the Mean Decrease in Accuracy (**MDA**)
   - Values permuted across all trees
   - Average decrease in accuracy; normalize by standard deviation

Original

	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes		0.07	Yes
	Mouse	No	No	Yes	No	No	Yes		0.09	Yes
	Mouse	No	No	Yes	No	No	Yes	0.31	0.06	Yes
	Mouse	No	Yes	Yes	No	No	Yes		0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes		0.07	Yes
	Mouse	No	No	Yes	No	No	Yes		0.05	Yes
	Mouse	No	No	Yes	No	No	Yes	0.32	0.06	Yes
	Mouse	No	No	Yes	No	No	Yes	0.38	0.08	Yes
	Mouse	No	No	Yes	No	No	Yes		0.08	Yes
	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.19	Yes
	Cat	No	Yes	No	Yes	No	Yes	0.53	0.20	Yes
	Cat	No	Yes	No	Yes	Yes	Yes		0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes		0.09	Yes
	Cat	No	Yes	No	Yes	Yes	Yes		0.17	Yes
	Cat	No	No	No	Yes	No	No	0.35	0.22	Yes
	Cat	No	Yes	No	Yes	Yes	Yes		0.16	Yes
	Cat	No	Yes	No	Yes	Yes	Yes	0.13	0.24	Yes
	Cat	No	Yes	No	Yes	Yes	Yes		0.22	Yes
	Dog	Yes	No	No	No	No	No	0.58	0.33	Yes
	Dog	Yes	Yes	No	No	Yes	No		0.35	Yes
21	Dog	Yes	Yes	No	No	Yes	No		0.33	Yes
22	Dog	Yes	Yes	No	No	Yes	No		0.32	Yes
23	Dog	No	No	No	No	No	No		0.26	Yes
	Dog	Yes	Yes	No	No	Yes	No		0.27	Yes
	Dog	Yes	Yes	No	No	Yes	No	0.13	0.16	Yes
	Dog	Yes	Yes	No	No	Yes	No	0.16	0.29	Yes
	Dog	Yes	Yes	No	No	Yes	No		0.29	Yes

Original Length Data

Randomly Permuted Length Data



Decrease in Accuracy
= 90% – 40%
= 50%

Original Out of Bag Accuracy 90%

	Mouse	Cat	Dog
Mouse	3	0	0
Cat	0	3	1
Dog	0	0	3

Out of Bay Accuracy After Permutation 40%

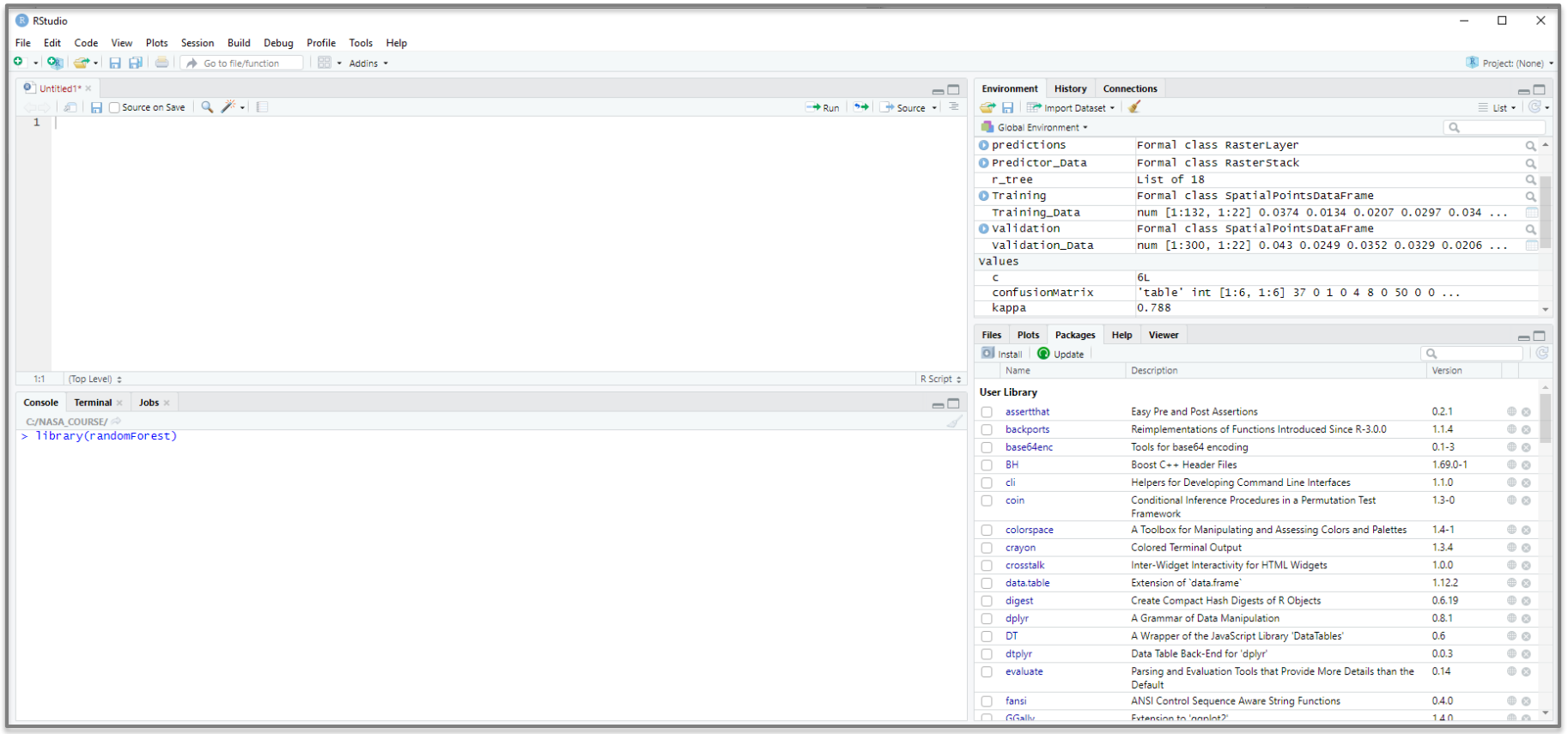
	Mouse	Cat	Dog
Mouse	0	3	0
Cat	1	3	0
Dog	2	0	1

Random Forests: Hands on Demonstration of Crop Classification

1. Install the latest version of R and RStudio
   - https://cran.r-project.org/bin/windows/base/
   - https://www.rstudio.com/products/rstudio/download/
2. Open Rstudio

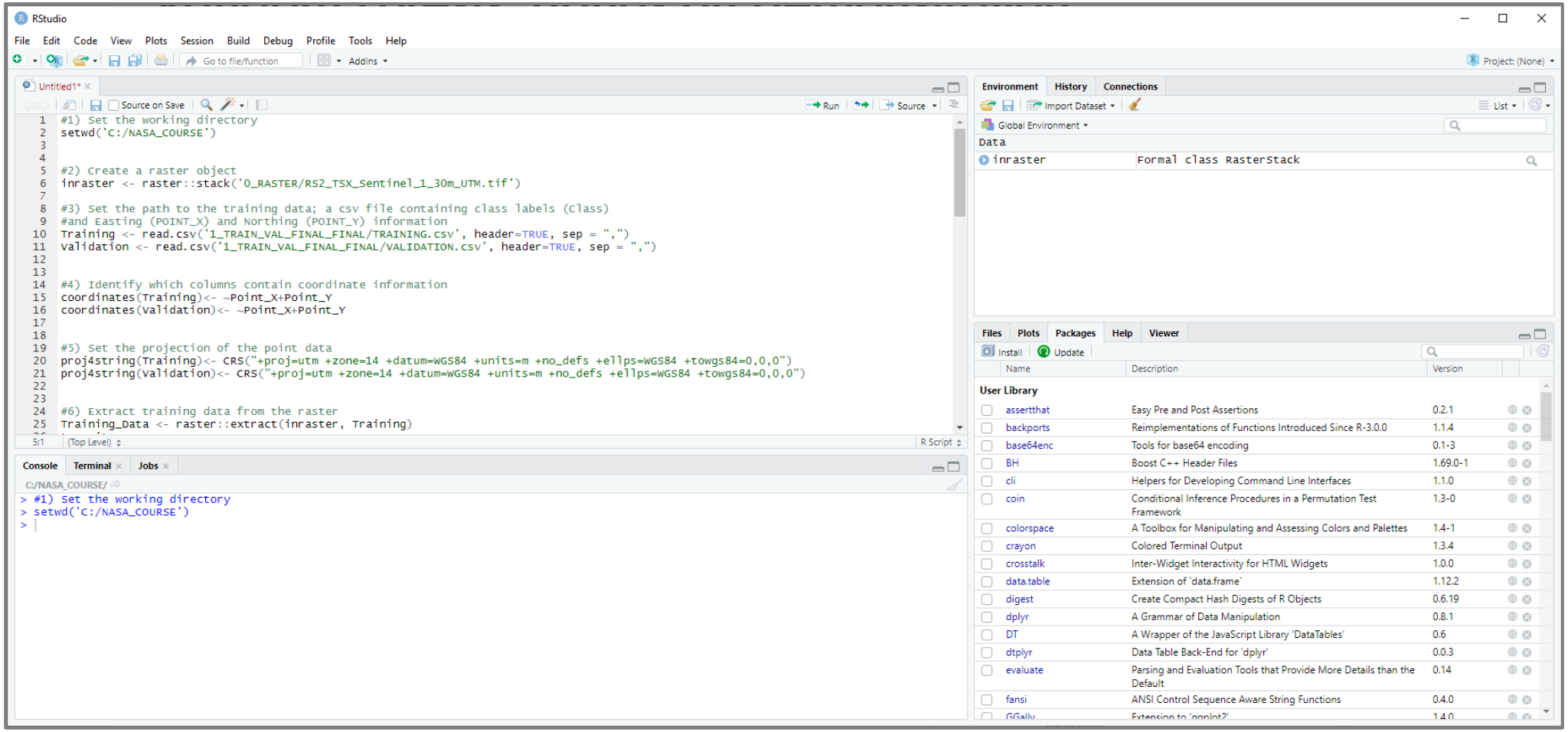
Random Forests: Hands on Demonstration

3. Install (raster, randomForest, sp, rgdal)
4. Load (raster, randomForest, sp, rgdal)



Random Forests: Hands on Demonstration

5. Set the Working Directory

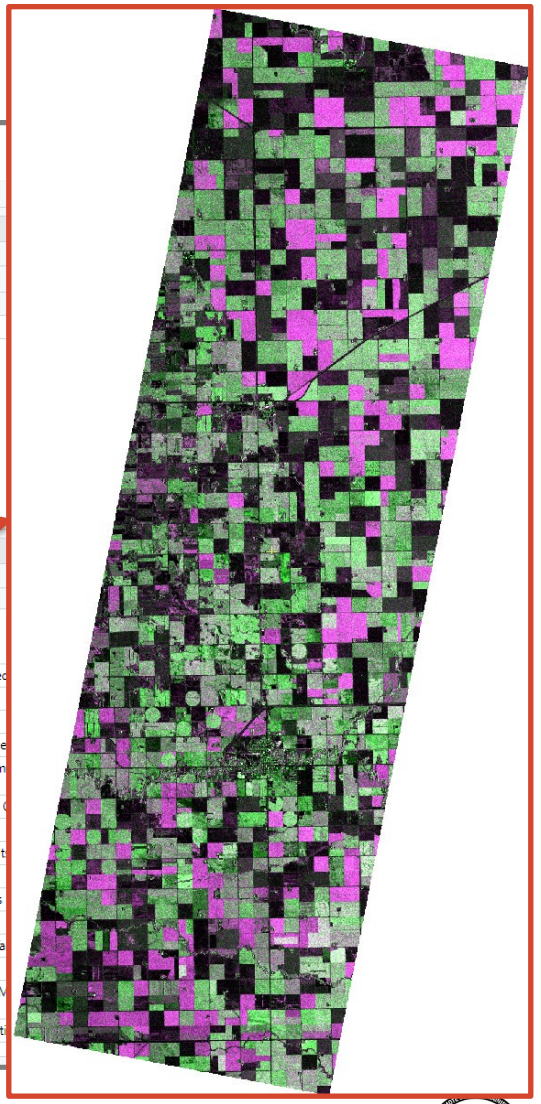


Random Forests: Hands on Demonstration

6. Create a raster object

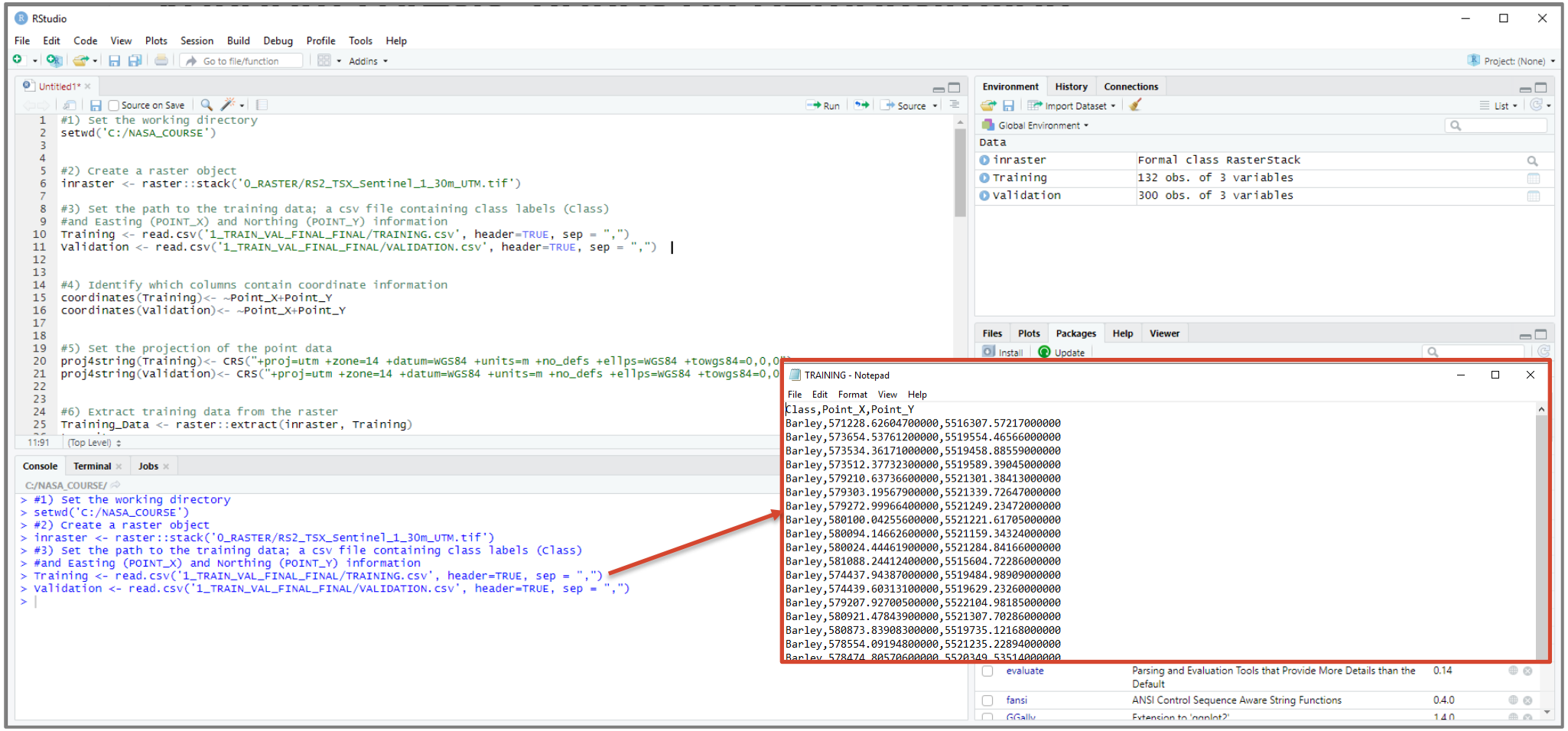
```
#1) Set the working directory
setwd('C:/NASA_COURSE')


#2) Create a raster object
inraster <- raster::stack('0_RASTER/RS2_TSX_Sentinel_1_30m_UTM.tif')

#3) Set the path to the training data; a csv file containing class labels (Class)
#and Easting (POINT_X) and Northing (POINT_Y) information
Training <- read.csv('1_TRAIN_VAL_FINAL_FINAL/TRAINING.csv', header=TRUE, sep = ",")
Validation <- read.csv('1_TRAIN_VAL_FINAL_FINAL/VALIDATION.csv', header=TRUE, sep = ",")


#4) Identify which columns contain coordinate information
coordinates(Training)<- ~Point_X+Point_Y
coordinates(Validation)<- ~Point_X+Point_Y


#5) Set the projection of the point data
proj4string(Training)<- CRS("+proj=utm +zone=14 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0")
proj4string(Validation)<- CRS("+proj=utm +zone=14 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0")


#6) Extract training data from the raster
Training_Data <- raster::extract(inraster, Training)
```

```
> #1) Set the working directory
> setwd('C:/NASA_COURSE')
> #2) Create a raster object
> inraster <- raster::stack('0_RASTER/RS2_TSX_Sentinel_1_30m_UTM.tif')
>
```



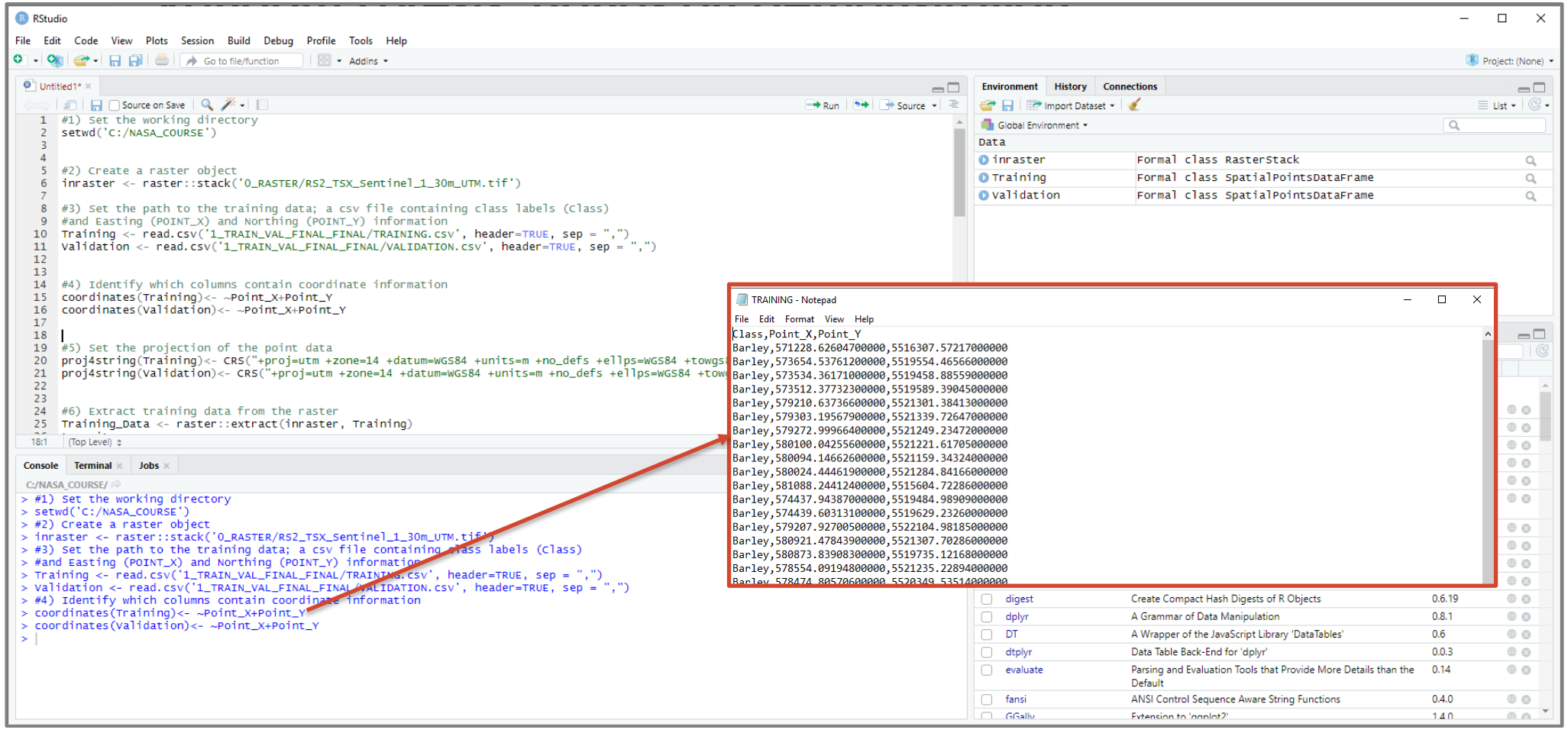
Random Forests: Hands on Demonstration

7. Read training and validation data (csv files)



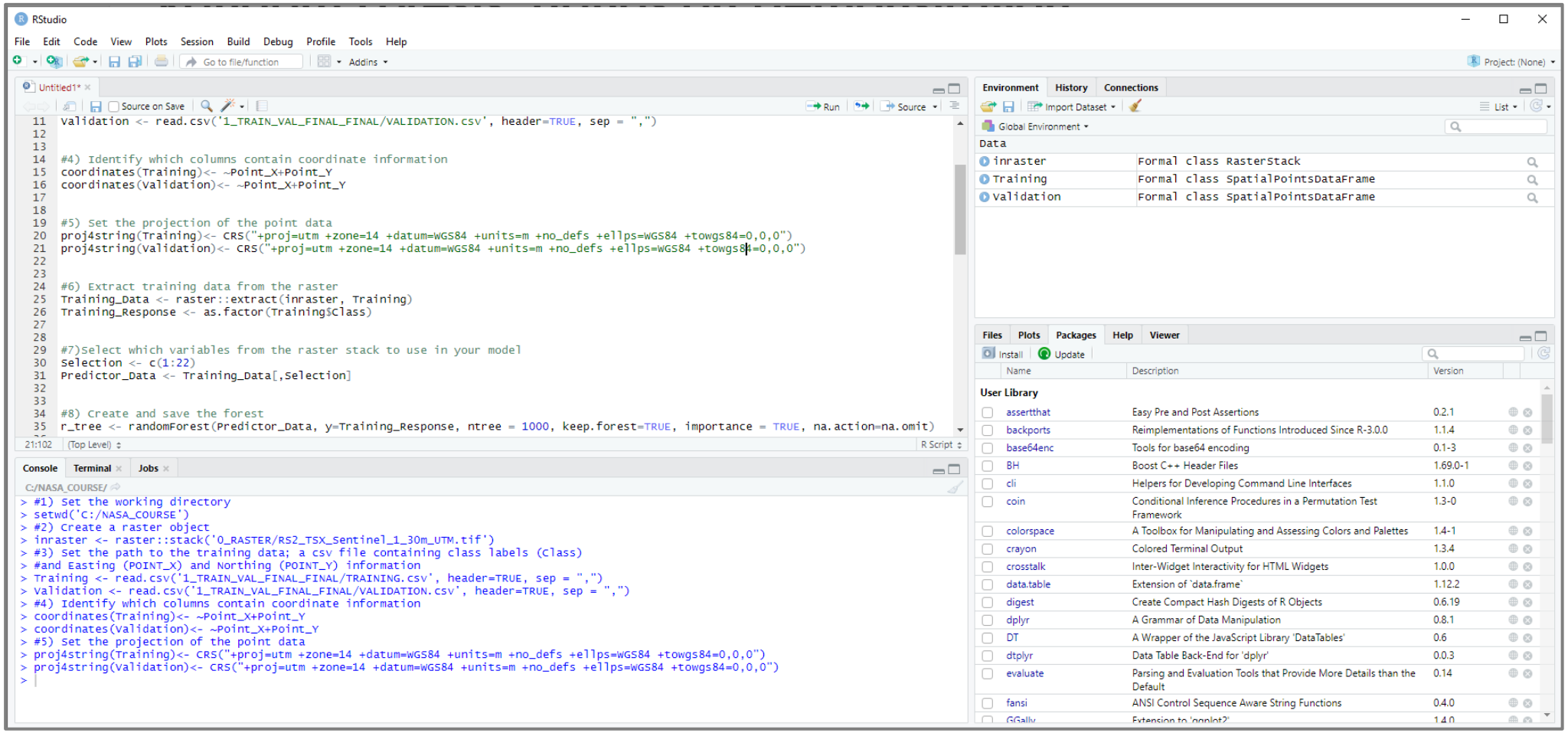
Random Forests: Hands on Demonstration

8. Identify which columns from the csv files contain coordinate information



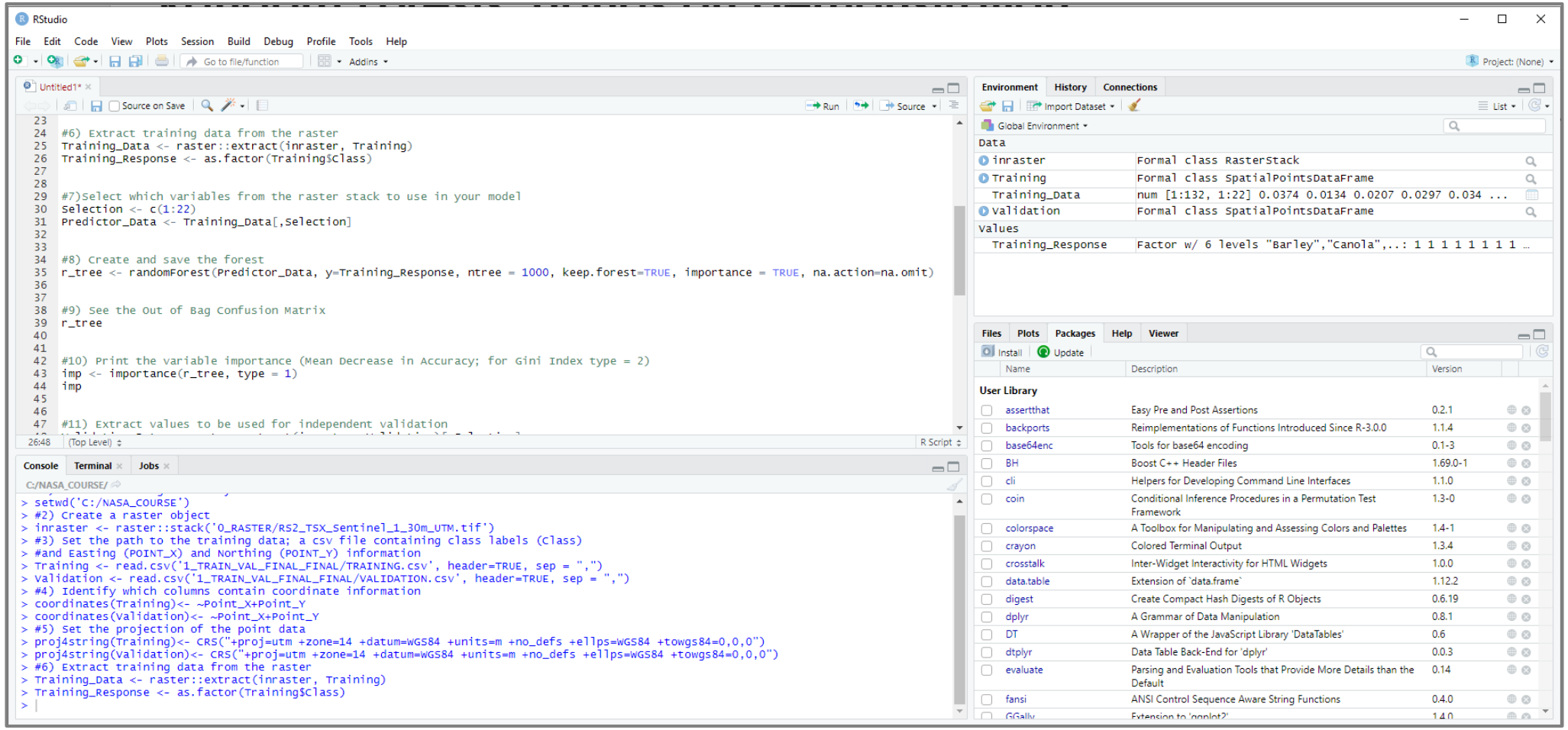
Random Forests: Hands on Demonstration

9. Define the projection of the point data in the csv files



Random Forests: Hands on Demonstration

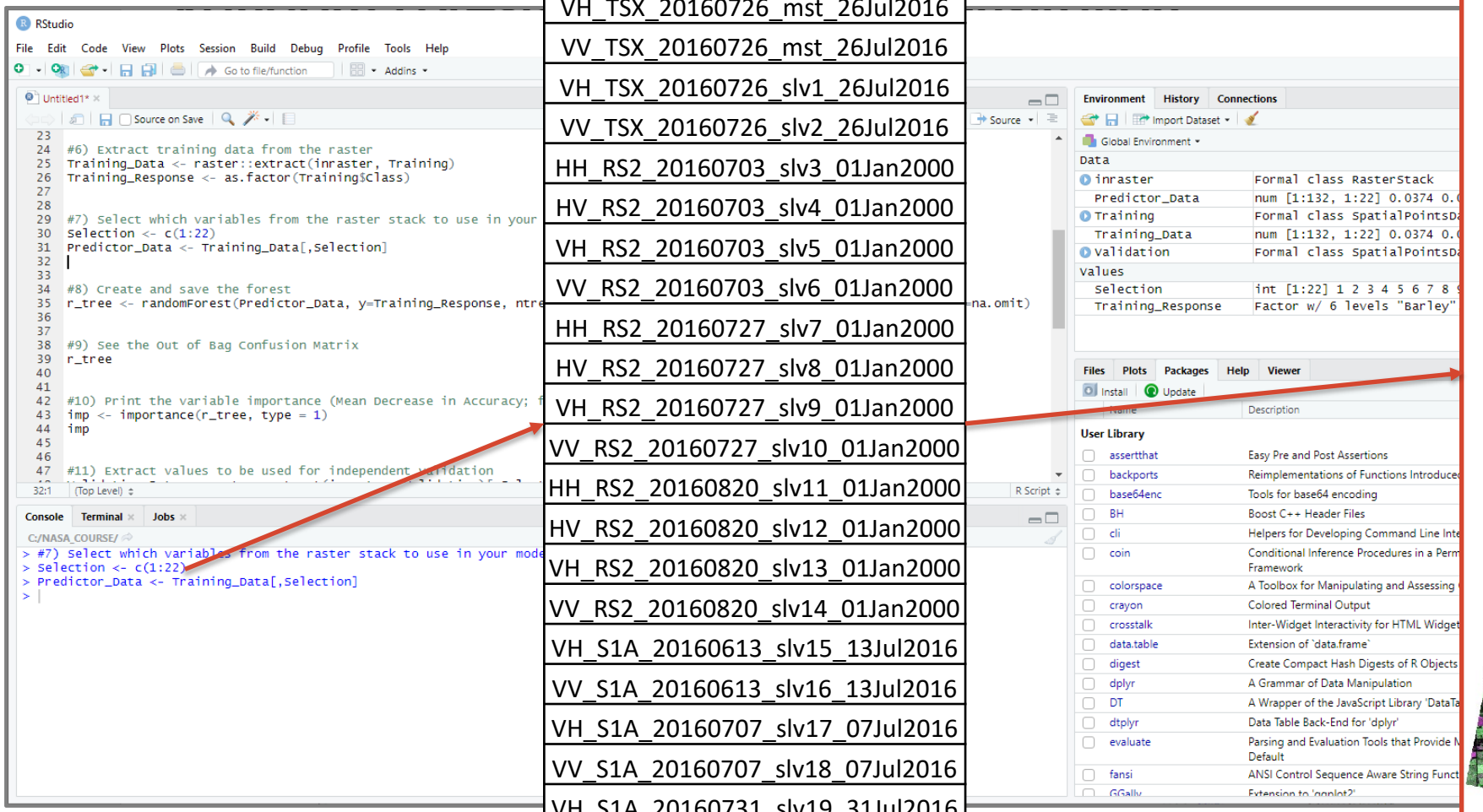
10. Extract the training data (values of each raster band coincident with point data)

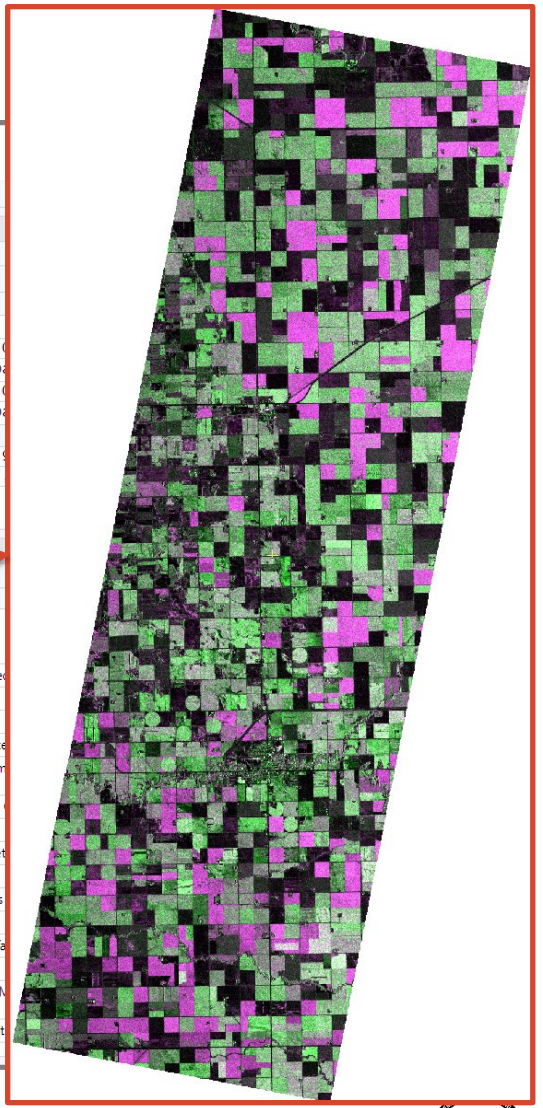


Random Forests: Hands on Demonstration

11. Select which variables to use in the model

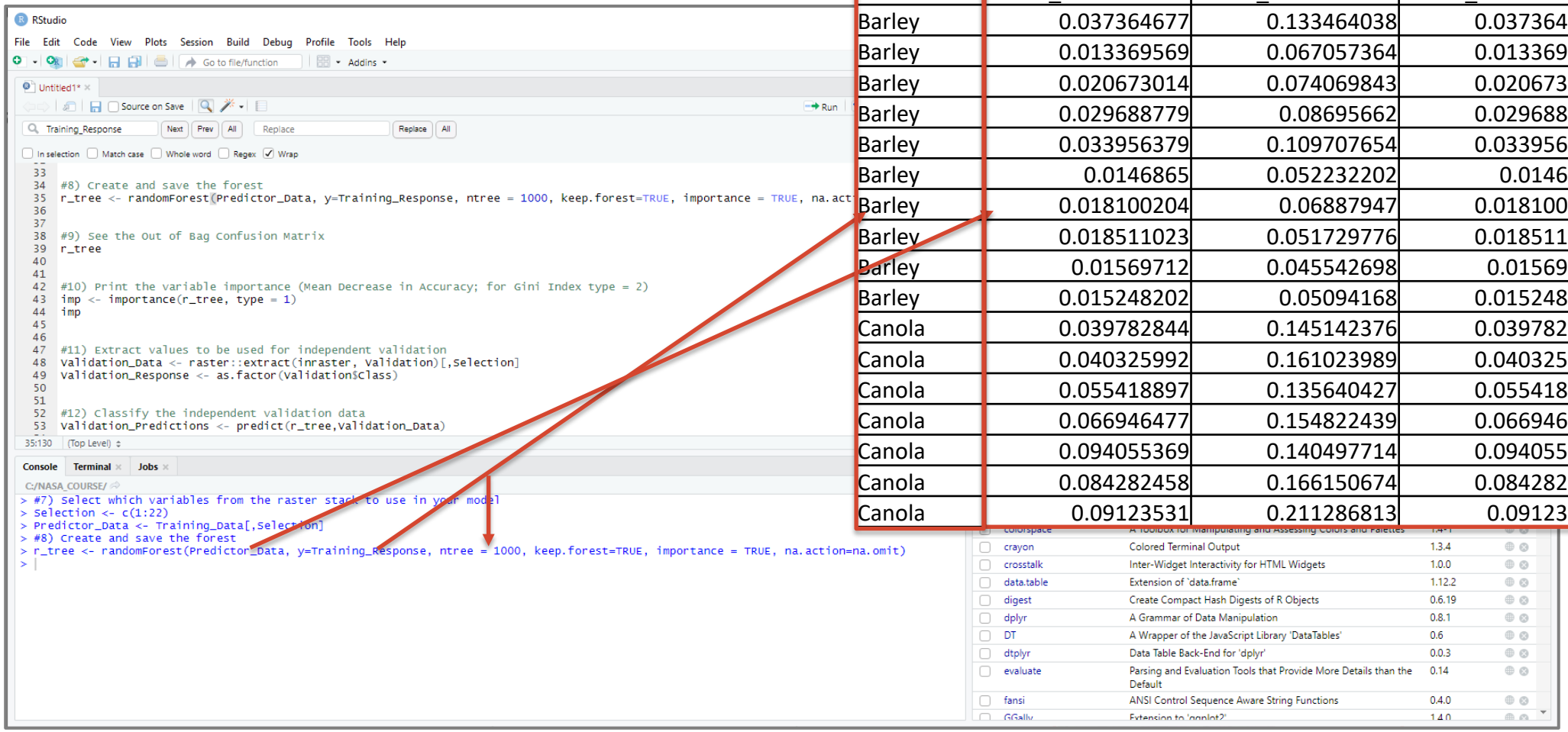
VH_TSX_20160726_mst_26Jul2016
VV_TSX_20160726_mst_26Jul2016
VH_TSX_20160726_slv1_26Jul2016
VV_TSX_20160726_slv2_26Jul2016
HH_RS2_20160703_slv3_01Jan2000
HV_RS2_20160703_slv4_01Jan2000
VH_RS2_20160703_slv5_01Jan2000
VV_RS2_20160703_slv6_01Jan2000
HH_RS2_20160727_slv7_01Jan2000
HV_RS2_20160727_slv8_01Jan2000
VH_RS2_20160727_slv9_01Jan2000
VV_RS2_20160727_slv10_01Jan2000
HH_RS2_20160820_slv11_01Jan2000
HV_RS2_20160820_slv12_01Jan2000
VH_RS2_20160820_slv13_01Jan2000
VV_RS2_20160820_slv14_01Jan2000
VH_S1A_20160613_slv15_13Jul2016
VV_S1A_20160613_slv16_13Jul2016
VH_S1A_20160707_slv17_07Jul2016
VV_S1A_20160707_slv18_07Jul2016
VH_S1A_20160731_slv19_31Jul2016
VV_S1A_20160731_slv20_31Jul2016





Random Forests: Hands on Demonstration

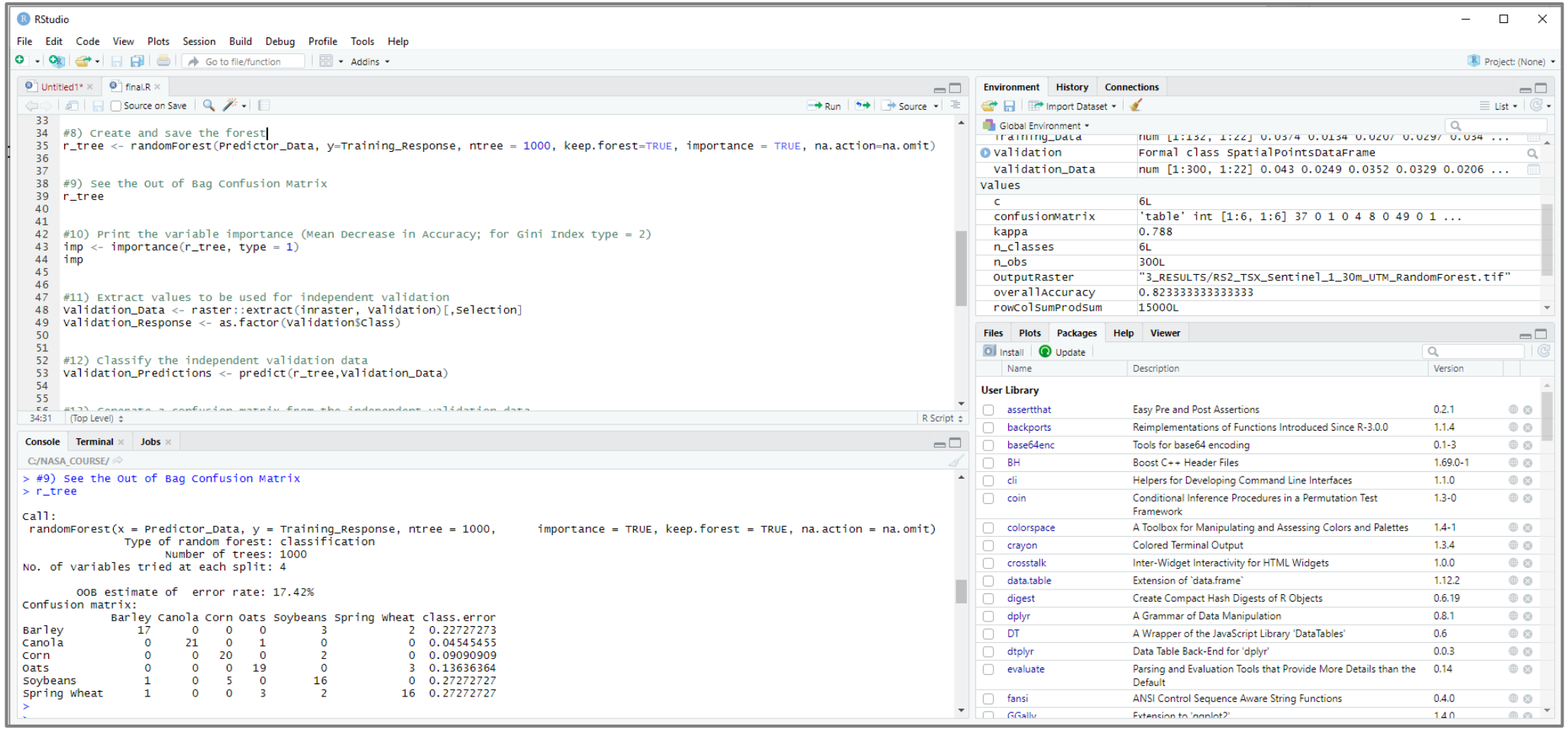
12. Create your Random Forest Model



Class	VH_TSX_20160726_mst_26Jul2016	VV_TSX_20160726_mst_26Jul2016	VH_TSX_20160726_slv1_26Jul2016
Barley	0.037364677	0.133464038	0.037364677
Barley	0.013369569	0.067057364	0.013369569
Barley	0.020673014	0.074069843	0.020673014
Barley	0.029688779	0.08695662	0.029688779
Barley	0.033956379	0.109707654	0.033956379
Barley	0.0146865	0.052232202	0.0146865
Barley	0.018100204	0.06887947	0.018100204
Barley	0.018511023	0.051729776	0.018511023
Barley	0.01569712	0.045542698	0.01569712
Barley	0.015248202	0.05094168	0.015248202
Canola	0.039782844	0.145142376	0.039782844
Canola	0.040325992	0.161023989	0.040325992
Canola	0.055418897	0.135640427	0.055418897
Canola	0.066946477	0.154822439	0.066946477
Canola	0.094055369	0.140497714	0.094055369
Canola	0.084282458	0.166150674	0.084282458
Canola	0.09123531	0.211286813	0.09123531

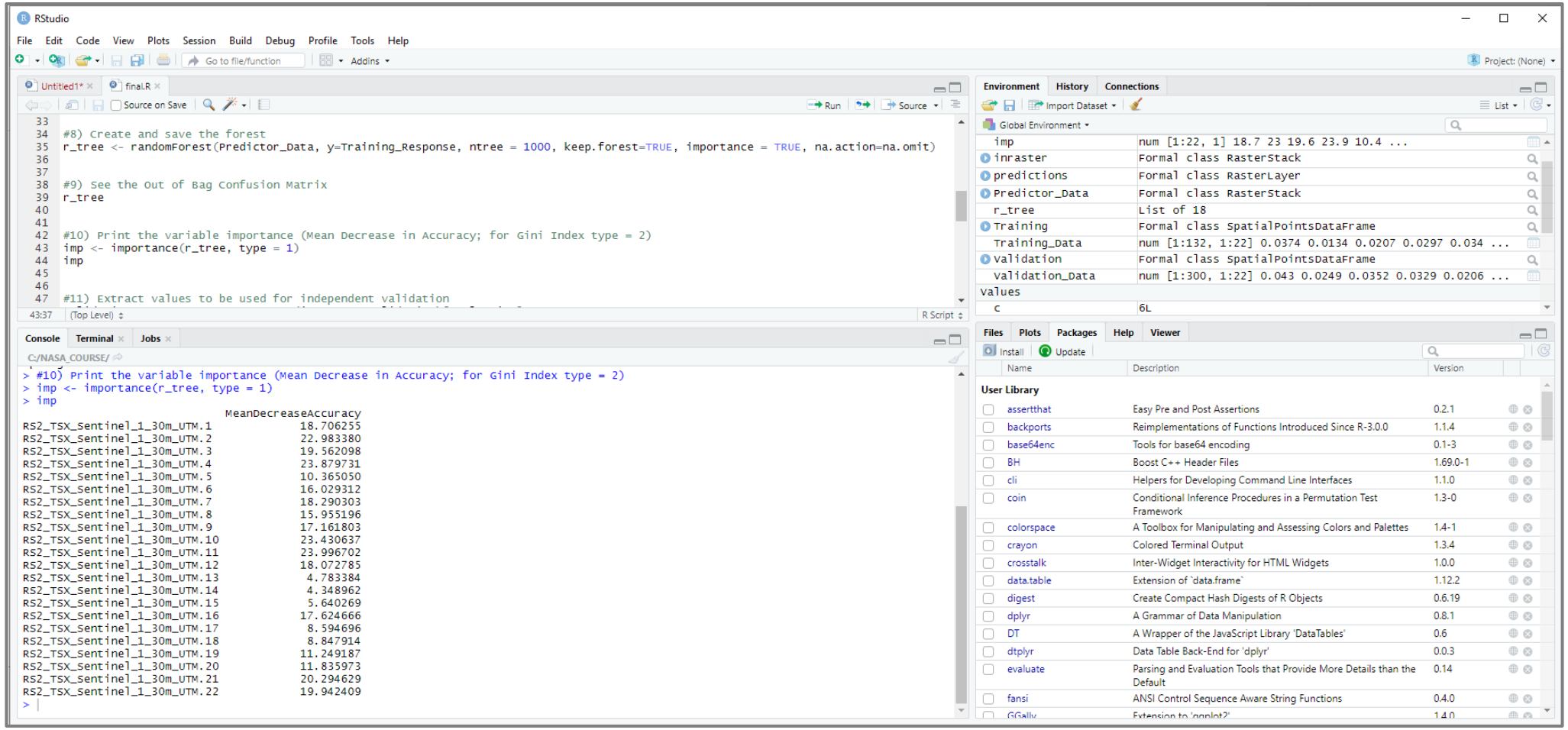
Random Forests: Hands on Demonstration

13. Print the Out of Bag confusion matrix



Random Forests: Hands on Demonstration

14. Print the variable importance values

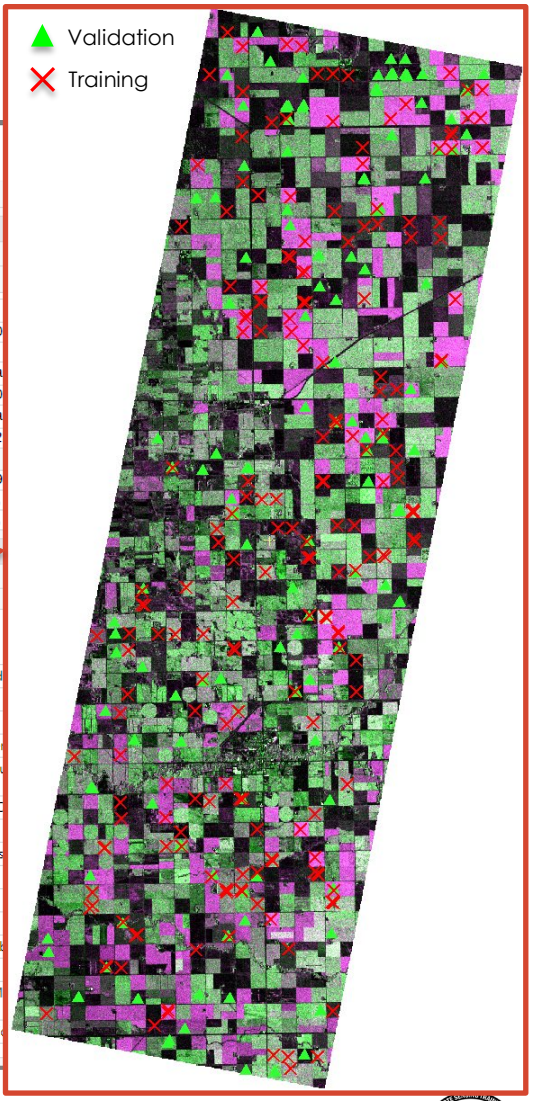


Random Forests: Hands on Demonstration

15. Extract the validation data

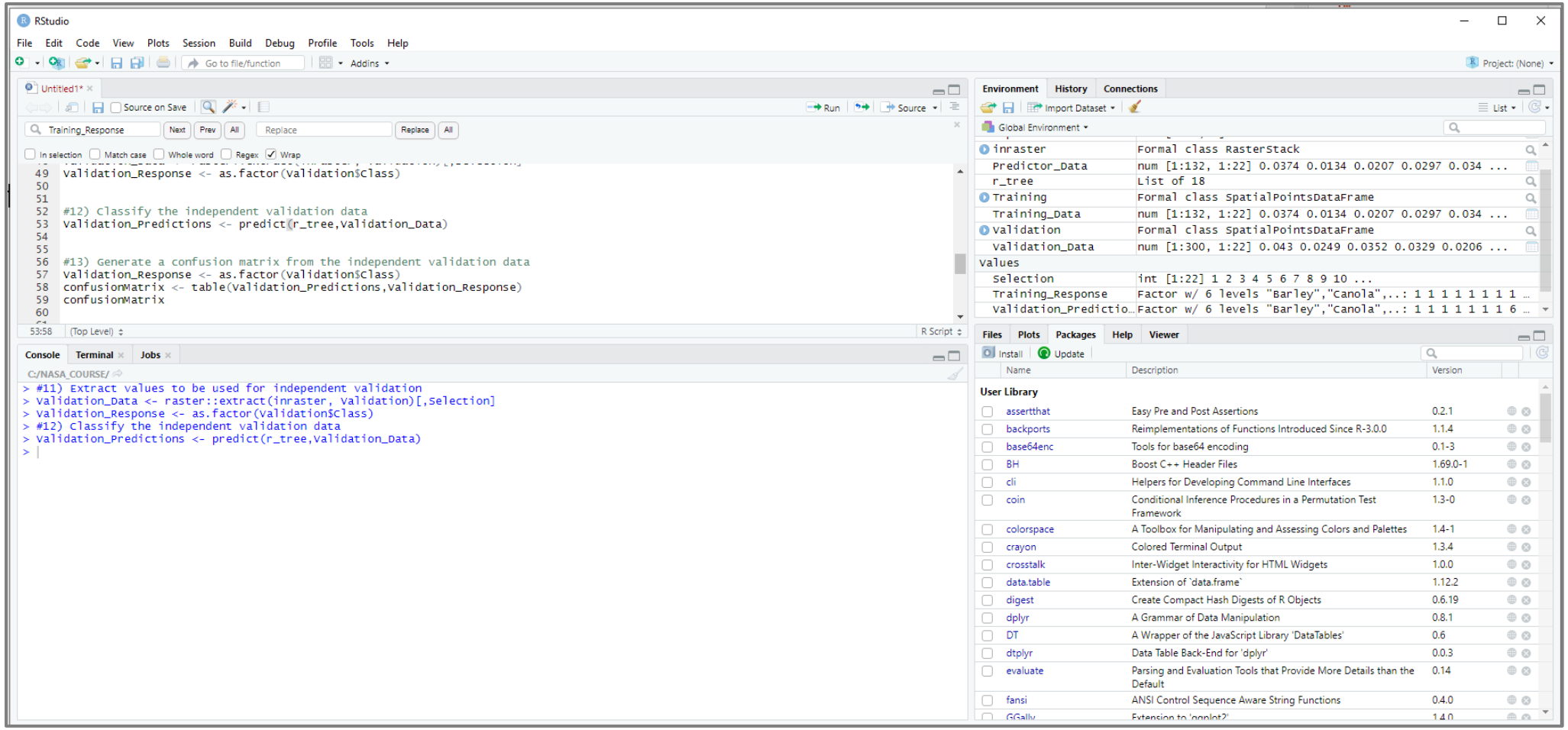
```
#11) Extract values to be used for independent validation
Validation_Data <- raster::extract(inraster, Validation)[,Selection]
Validation_Response <- as.factor(Validation$Class)

#12) Classify the independent validation data
Validation_Predictions <- predict(r_tree,Validation_Data)

#13) Generate a confusion matrix from the independent validation data
```

```
> #11) Extract values to be used for independent validation
> Validation_Data <- raster::extract(inraster, Validation)[,Selection]
> Validation_Response <- as.factor(Validation$Class)
>
```



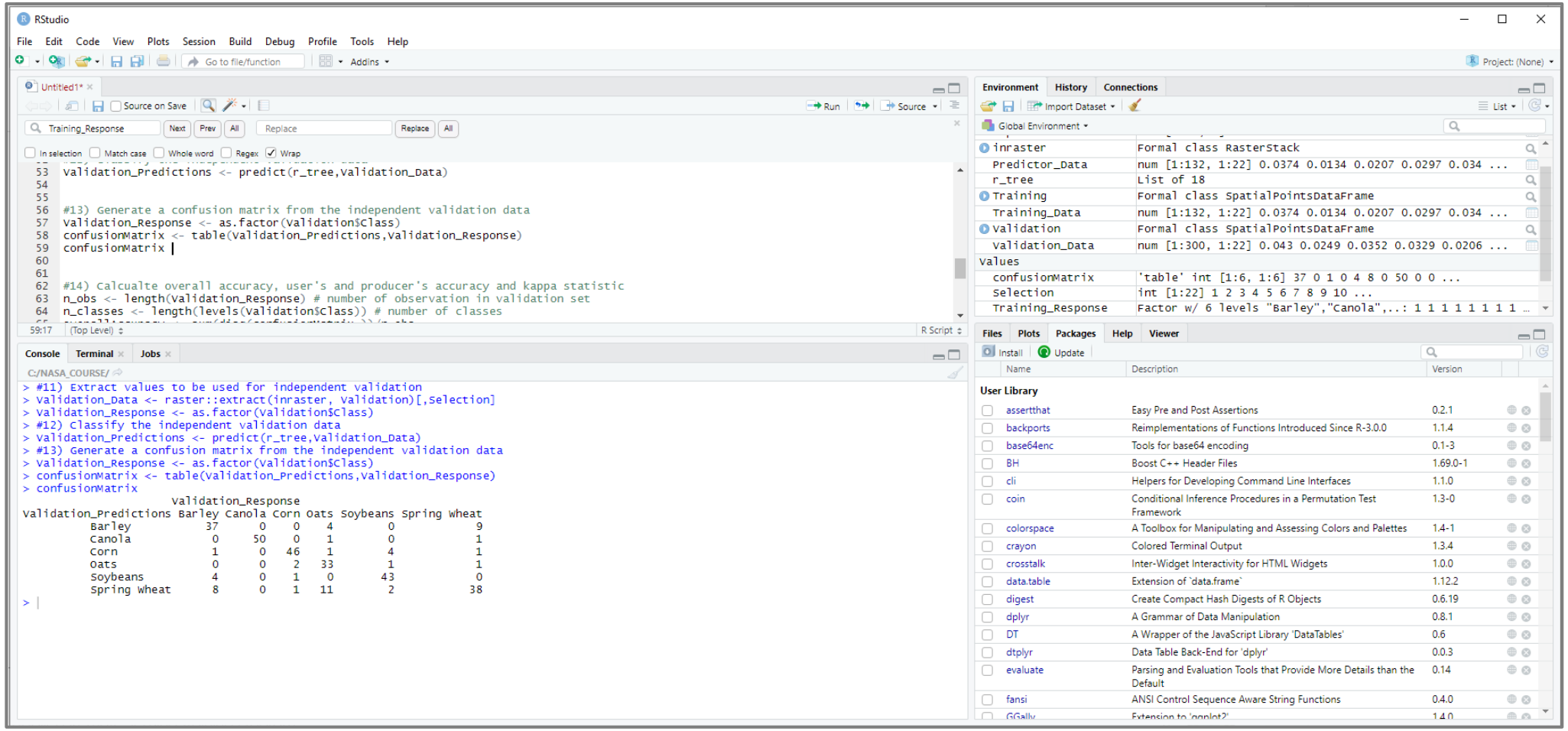
Random Forests: Hands on Demonstration

16. Classify the independent validation data



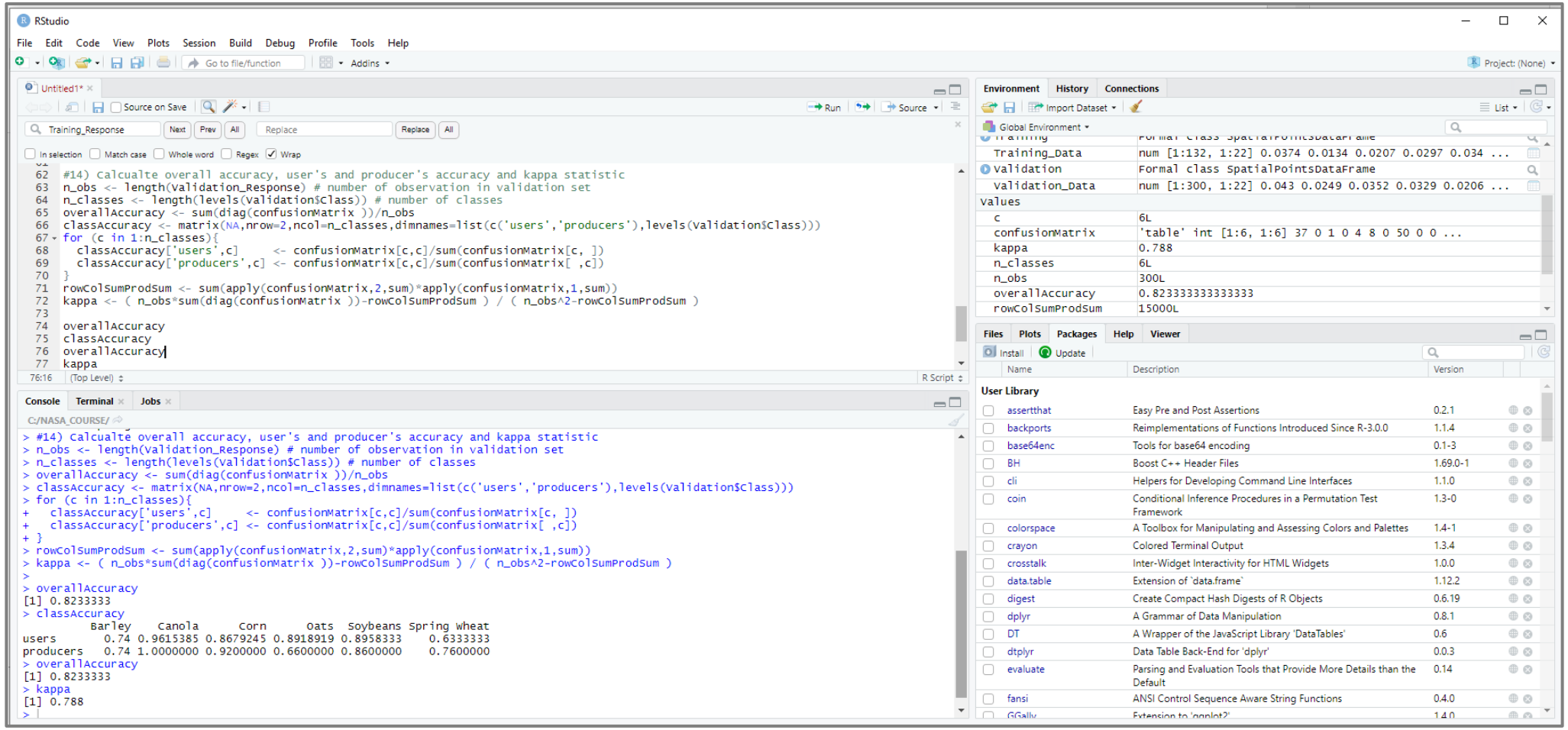
Random Forests: Hands on Demonstration

17. Generate a confusion matrix



Random Forests: Hands on Demonstration

18. Calculate the independent accuracies and kappa statistic

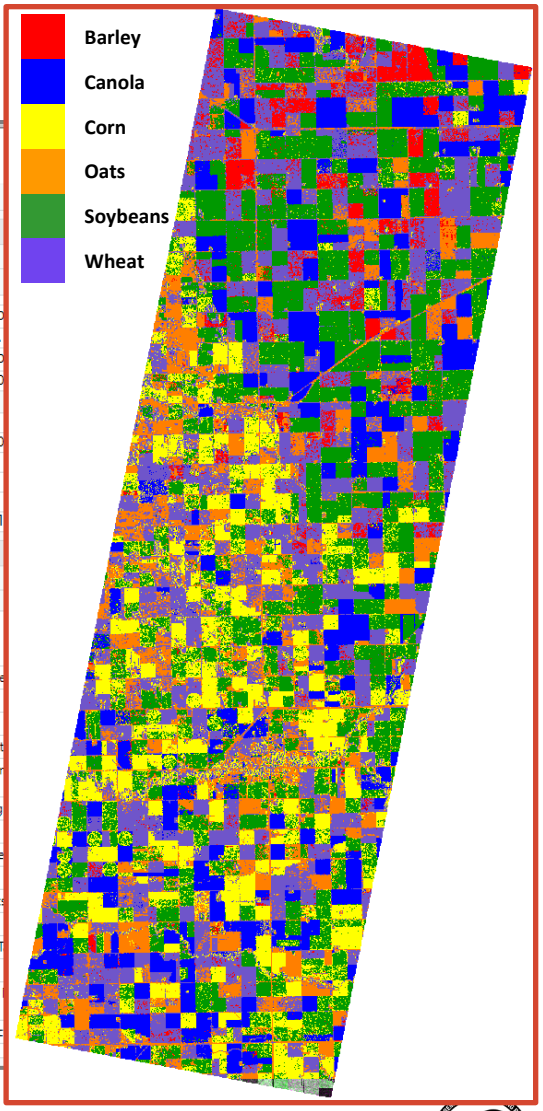


Random Forests: Hands on Demonstration

19. Classify the whole raster

```
69    classAccuracy['producers',c] <- confusionMatrix[c,c]/sum(confusionMatrix[ ,c])
70  }
71  rowColSumProdSum <- sum(apply(confusionMatrix,2,sum)*apply(confusionMatrix,1,sum))
72  kappa <- ( n_obs*sum(diag(confusionMatrix ))-rowColSumProdSum ) / ( n_obs^2-rowColSumProdSum )
73
74  overallAccuracy
75  classAccuracy
76  overallAccuracy
77  kappa
78
79
80  #15) Classify the whole raster
81  OutputRaster <- '3_RESULTS/RS2_TSX_Sentinel_1_30m_UTM_RandomForest.tif'
82  Predictor_Data <- subset(inraster, Selection)
83  predictions <- predict(inraster[[Selection]],model=r_tree,file=OutputRaster)
84
```

```
> #15) Classify the whole raster
> OutputRaster <- '3_RESULTS/RS2_TSX_Sentinel_1_30m_UTM_RandomForest.tif'
> Predictor_Data <- subset(inraster, Selection)
> predictions <- predict(inraster[[Selection]],model=r_tree,file=OutputRaster)
```

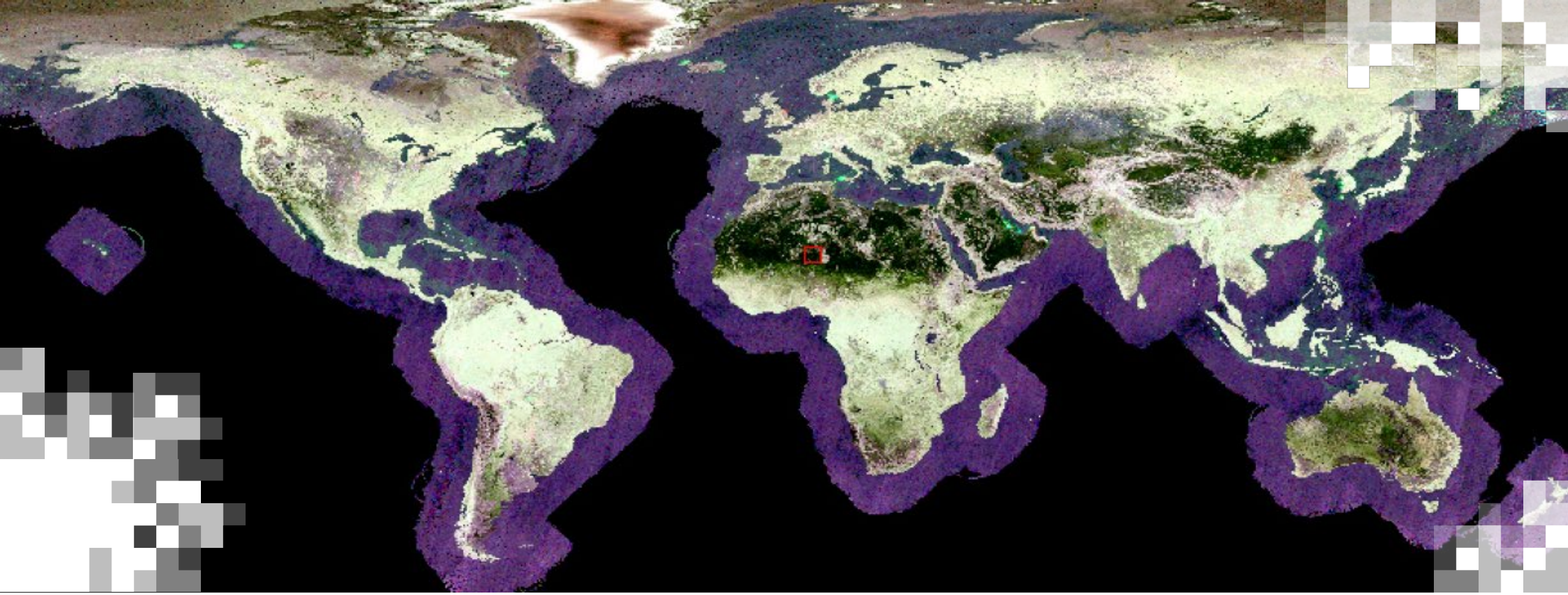


References

1. Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.
2. Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest." R news 2.3 (2002): 18-22.
3. https://victorzhou.com/blog/gini-impurity/
4. Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. "Variable selection using random forests." Pattern Recognition Letters 31.14 (2010): 2225-2236.
5. Behnamian, Amir, et al. "A systematic approach for variable selection with random forests: achieving stable variable importance values." IEEE Geoscience and Remote Sensing Letters 14.11 (2017): 1988-1992.
6. Banks, Sarah, et al. "Assessing the potential to operationalize shoreline sensitivity mapping: Classifying multiple Wide Fine Quadrature Polarized RADARSAT-2 and Landsat 5 scenes with a single Random Forest model." Remote Sensing 7.10 (2015): 13528-13563.
7. Banks, Sarah, et al. "Contributions of Actual and Simulated Satellite SAR Data for Substrate Type Differentiation and Shoreline Mapping in the Canadian Arctic." Remote Sensing 9.12 (2017): 1206.
8. Banks, Sarah, et al. "Wetland Classification with Multi-Angle/Temporal SAR Using Random Forests." Remote Sensing 11.6 (2019): 670.
9. White, Lori, et al. "Moving to the RADARSAT constellation mission: Comparing synthesized compact polarimetry and dual polarimetry data with fully polarimetric RADARSAT-2 data for image classification of peatlands." Remote Sensing 9.6 (2017): 573.
10. Millard, Koreen, and Murray Richardson. "On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping." Remote sensing 7.7 (2015): 8489-8515.
11. Millard, Koreen, and Murray Richardson. "Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier." Canadian Journal of Remote Sensing 39.4 (2013): 290-307.
12. Planet Team (2017). Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com
13. RADARSAT-2 Data and Products © Maxar Technologies Ltd. (2018) – All Rights Reserved. RADARSAT is an official mark of the Canadian Space Agency.

Contributors

- Sarah Banks, Environment and Climate Change Canada
- Dr. Amir Behnamian, Environment and Climate Change Canada
- Dr. Koreen Millard, Carleton University, Department of Geography and Environmental Studies
- Dr. Cai, Carleton University, Department of Mathematics and Statistics
- Dr. Liaw, Merck

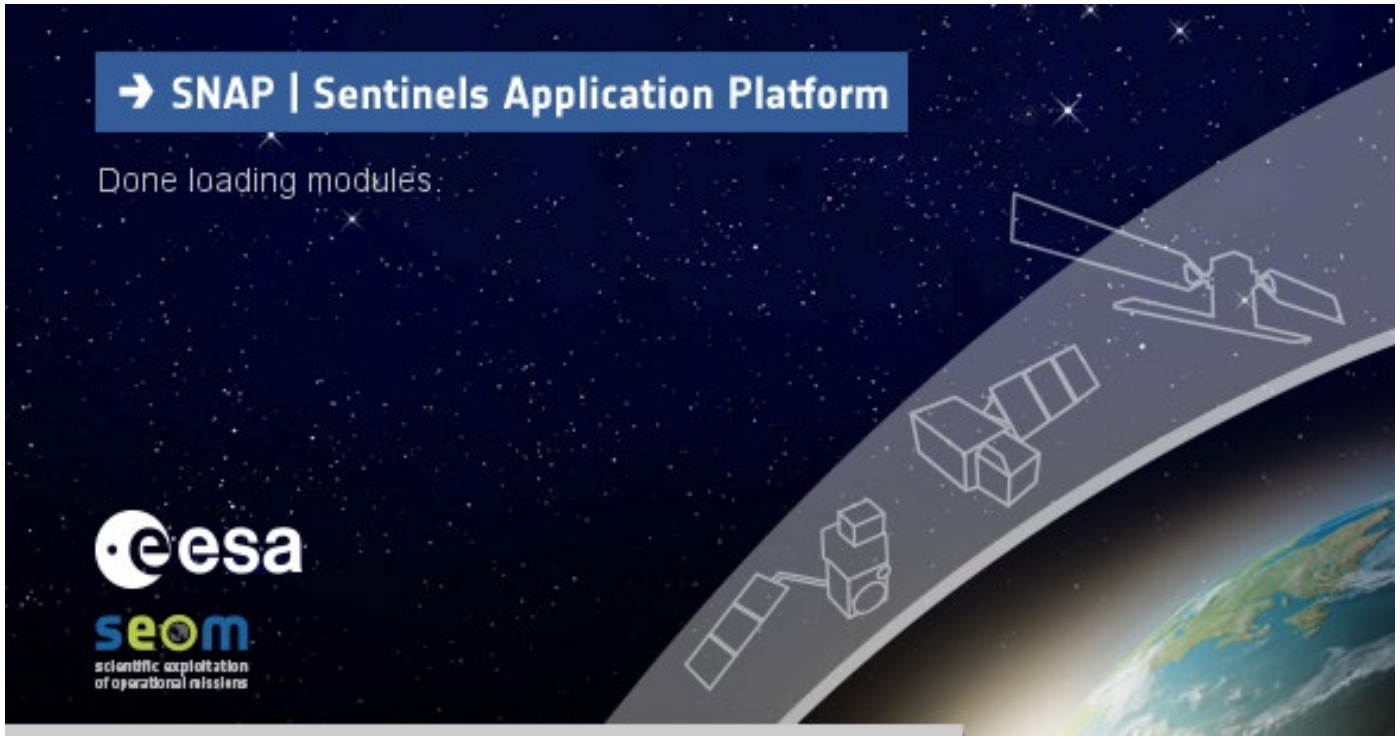


Exploiting SAR to Monitor Agriculture

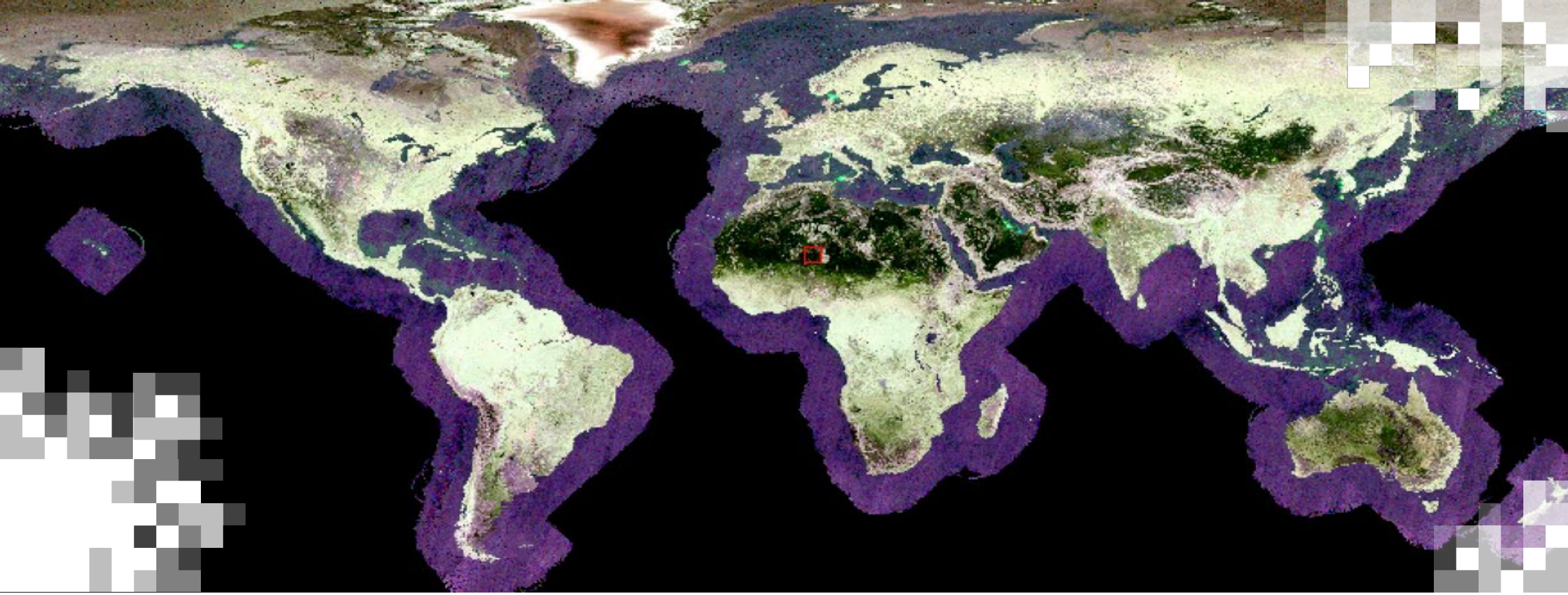
Learning Objectives

- By the end of this presentation, you will be able to understand:
  - how to estimate soil moisture from RADARSAT-2 data
  - how to process multi-frequency data for use in crop classification

SNAP: Sentinel's Application Platform

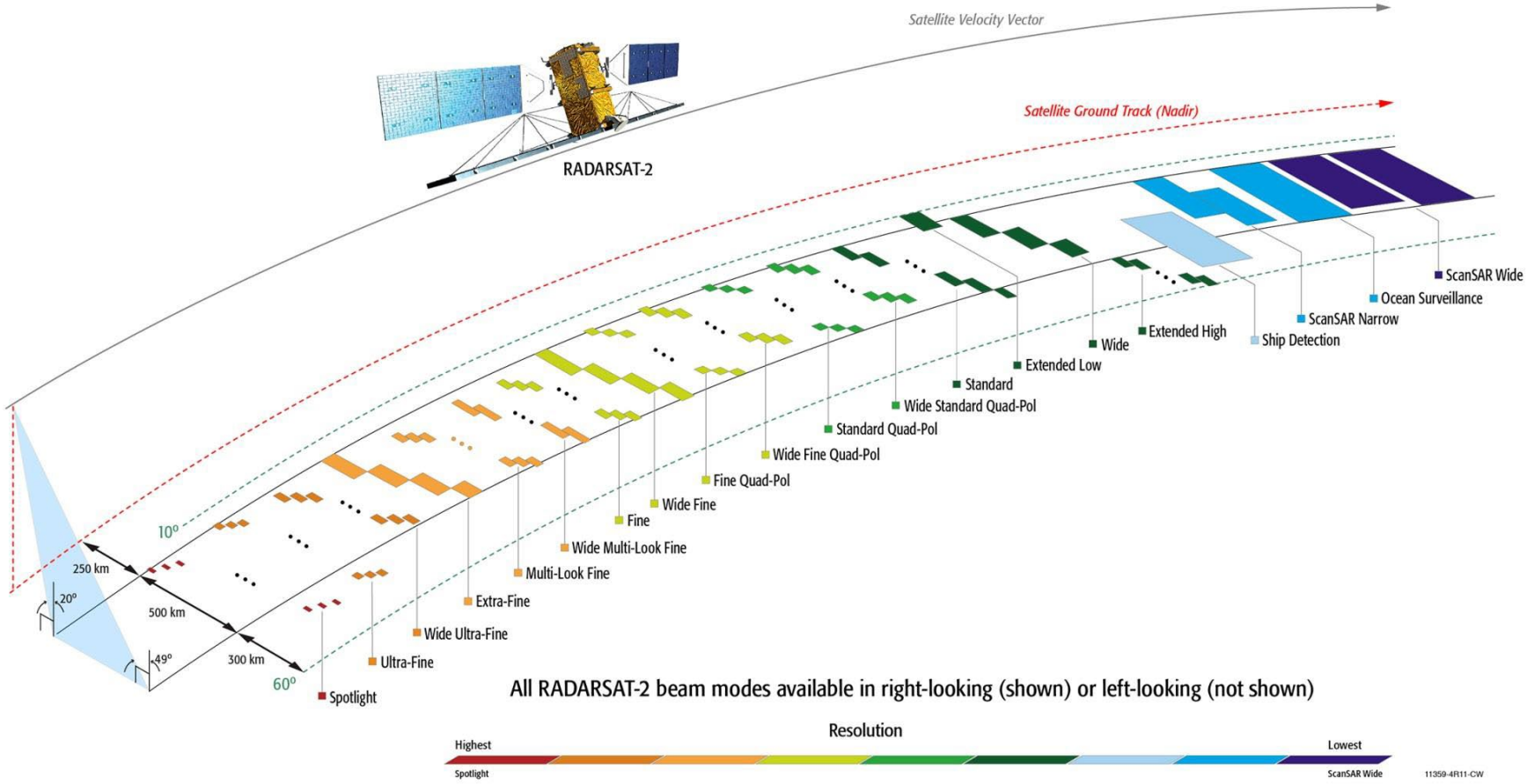


- ESA SNAP is the free and open-source toolbox for processing and analyzing ESA and 3rd part EO satellite image data
- You can download the latest installers for SNAP from:
  - http://step.esa.int/main/download/snap-download/



Estimating Soil Moisture from RADARSAT-2 Data

RADARSAT-2



RADARSAT-2 SAR Beam Modes – Revisit time: 24 days

Image Credits: MDA RADARSAT-2 Product Description

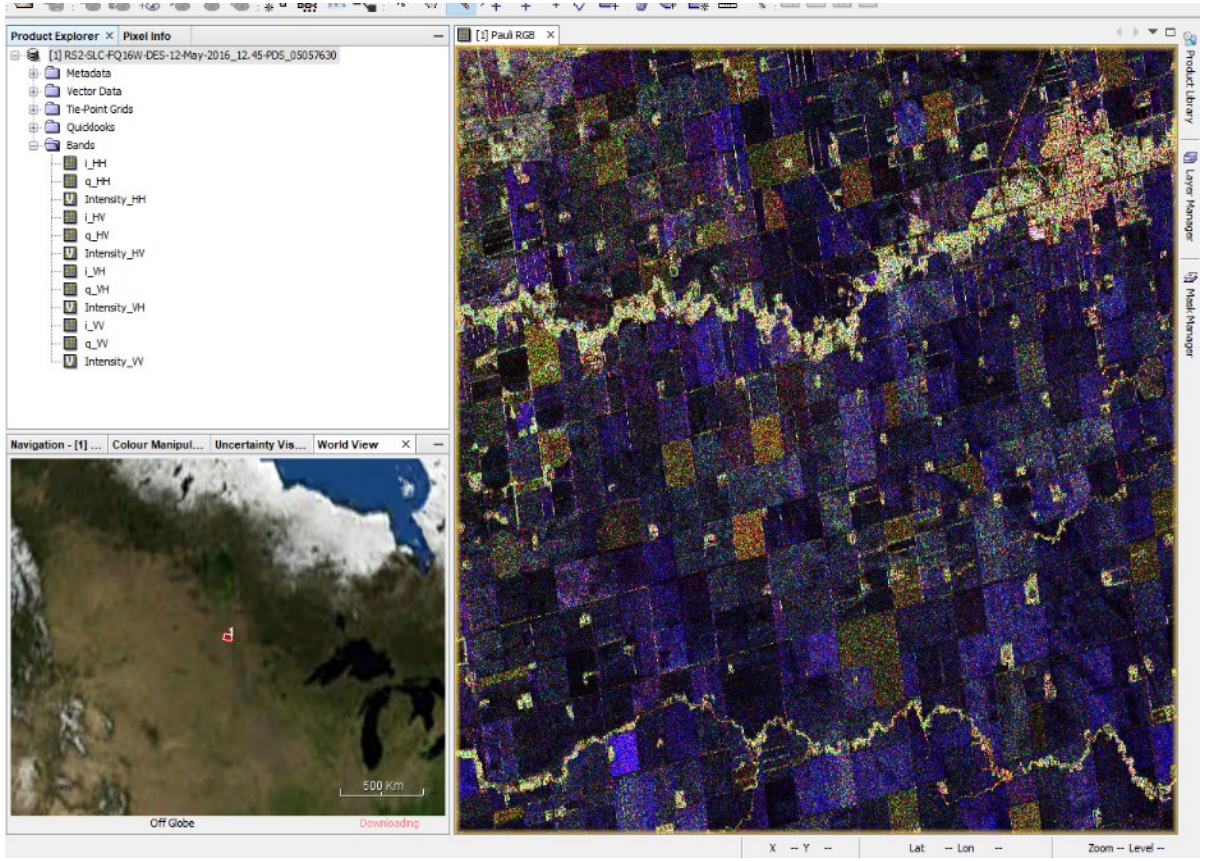
Pre-Processing RADARSAT-2 Data with SNAP

Extract Backscatter

Pre-Processing RADARSAT-2 Data with SNAP

Read Image

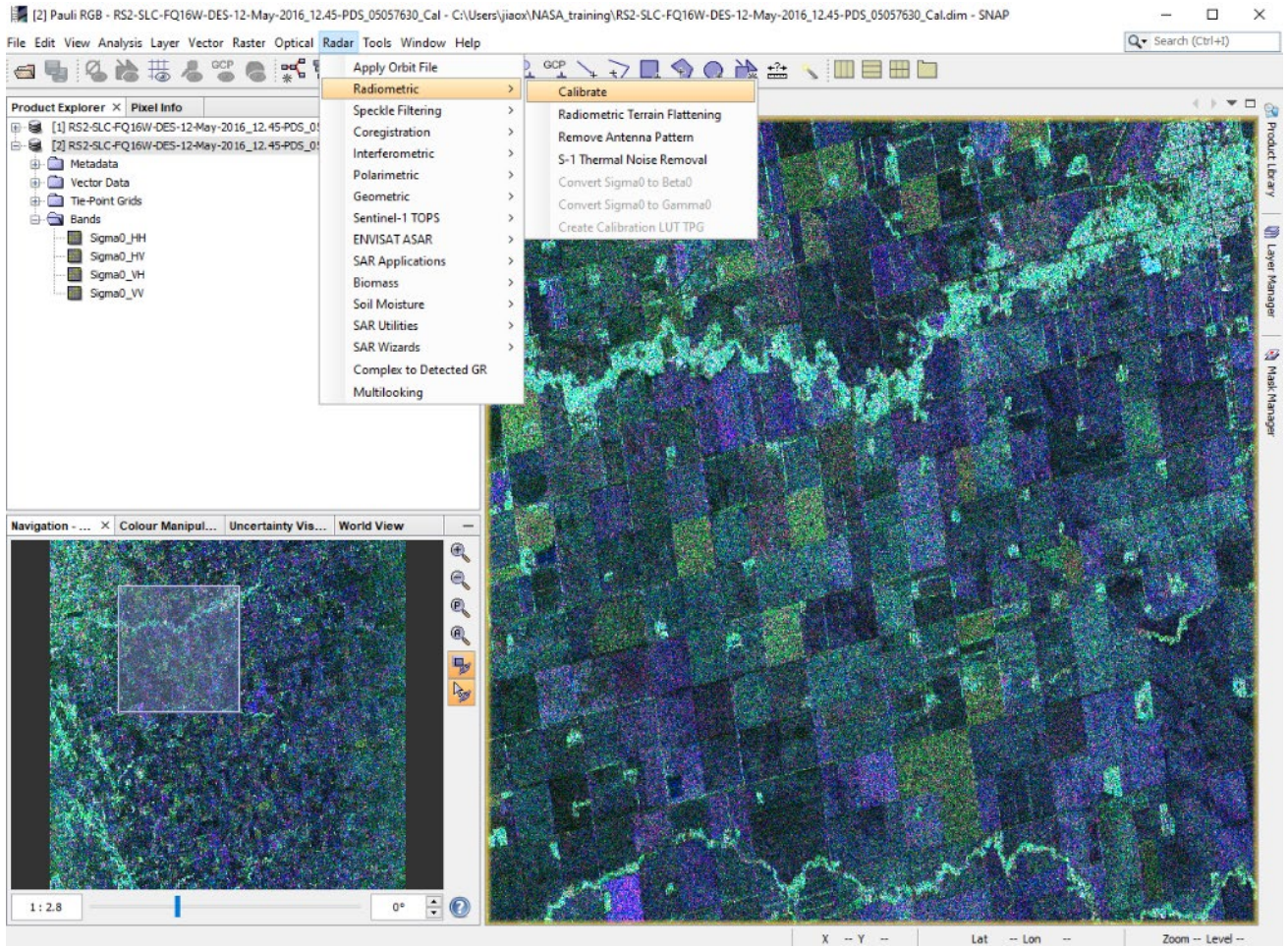
- RADARSAT-2 Wide Fine Quad-Pol single look complex(SLC) product:
  - Nominal Resolution: 5.2m (Range) * 7.6 m (Azimuth)
  - Nominal Scene Size: 50 Km (Range) * 25 Km (Azimuth)
  - Quad (HH, HV,VH and VV) polarizations + phase
  - Slant range single look complex, contains both amplitude and phase information



RADARSAT-2 Wide Fine Quad-Pol FQ16W descending SLC data acquired on May 12, 2016, over Carman, Manitoba, Canada

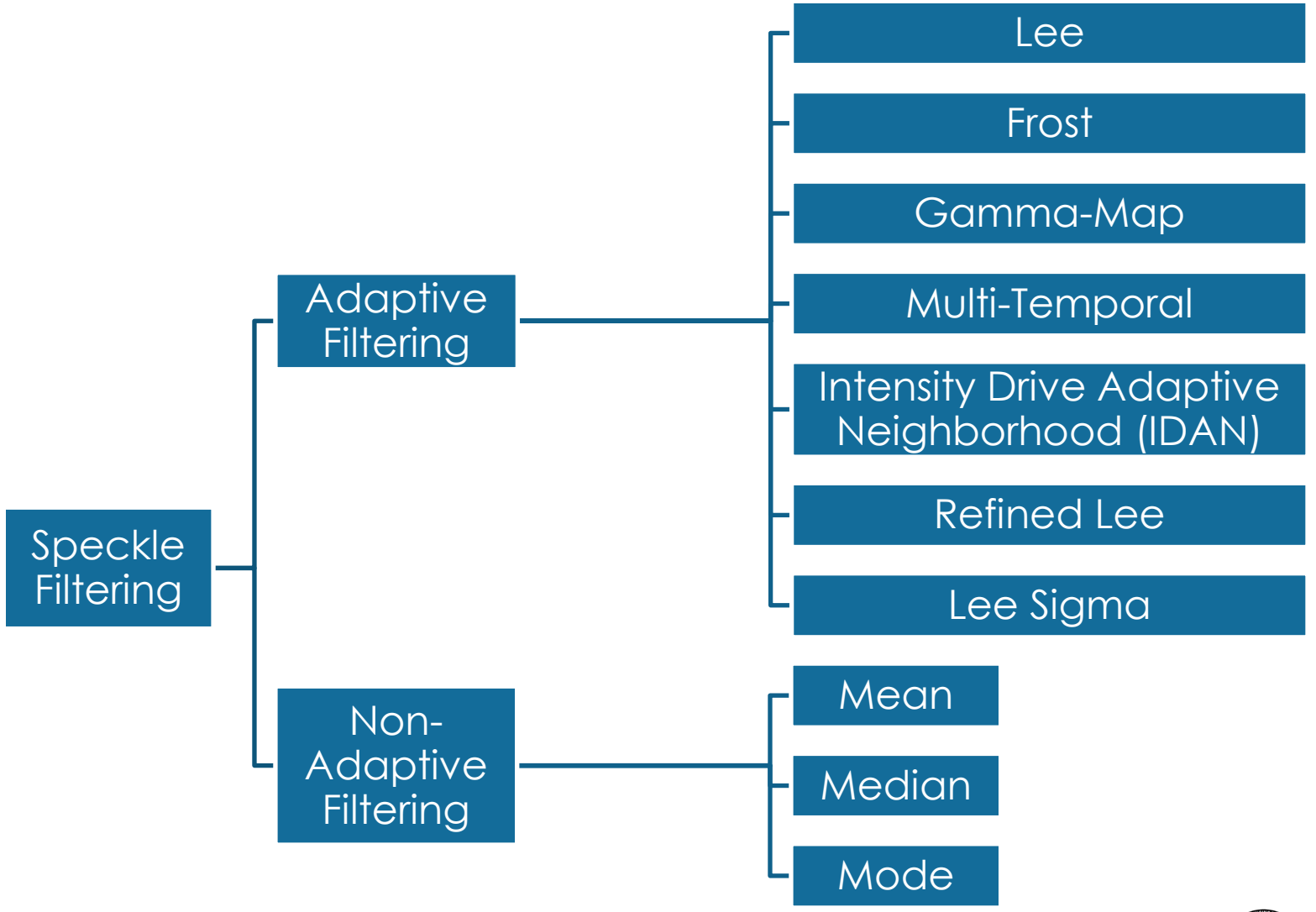
Pre-Processing RADARSAT-2 Data with SNAP

Calibration: Convert Pixel Values to Radar Backscatter



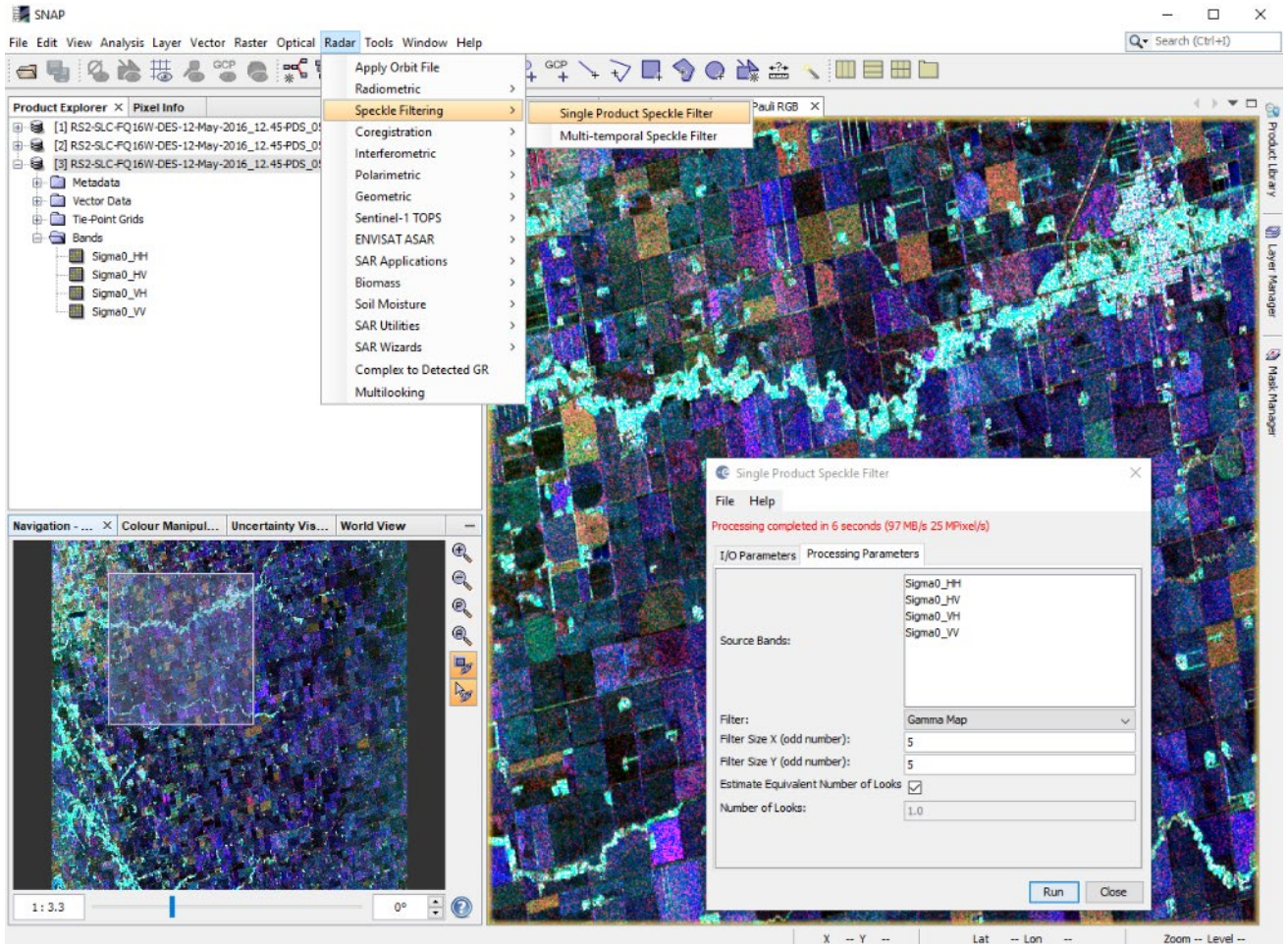
Speckle Filtering

- Speckle filtering is not an exact science → image and target dependent
- An ideal speckle filter will:
  - reduce speckle
  - preserve edge sharpness
  - preserve line and point target contrast
  - retain mean values in homogeneous regions
  - retain texture information



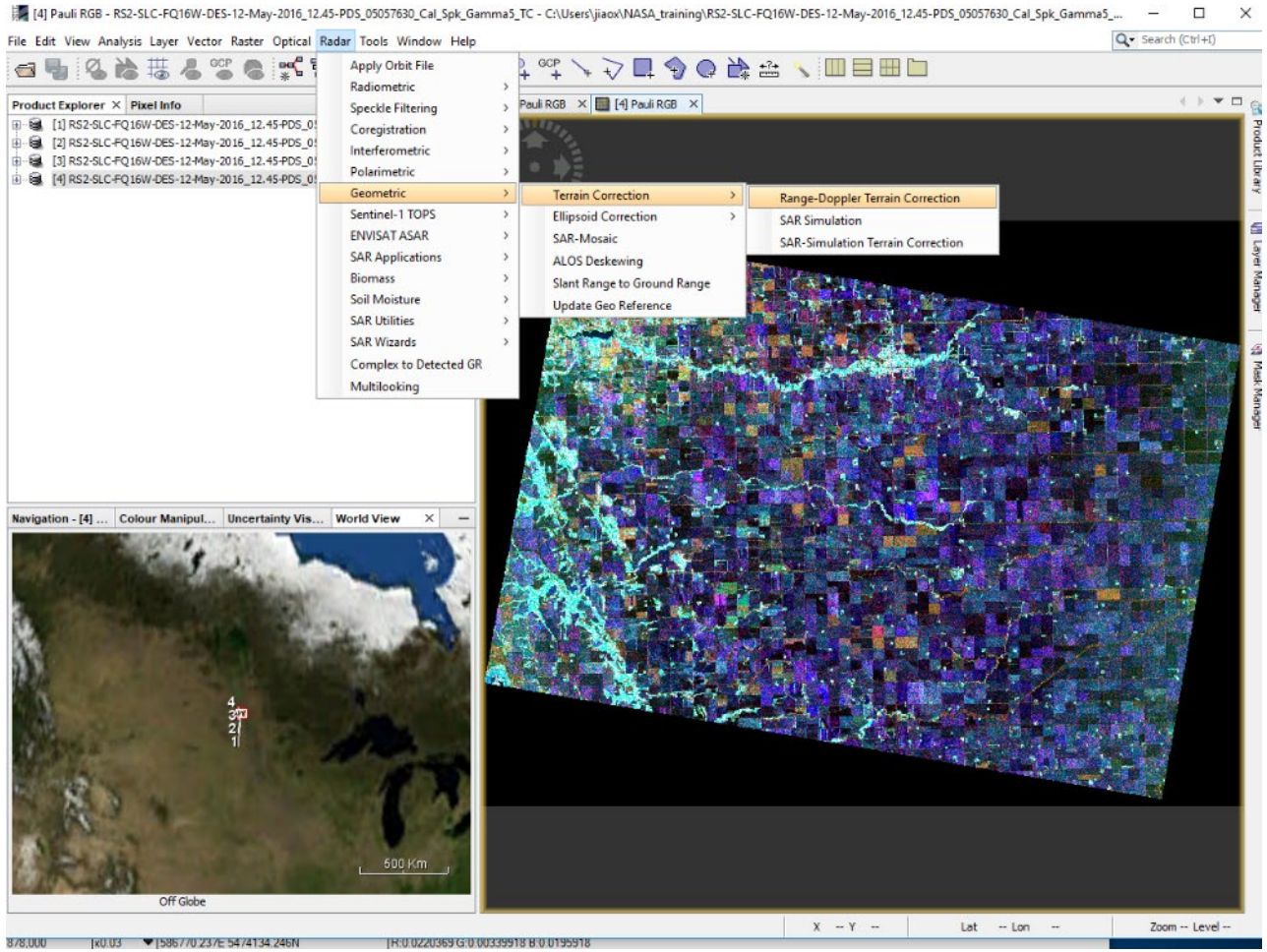
Pre-Processing RADARSAT-2 Data with SNAP

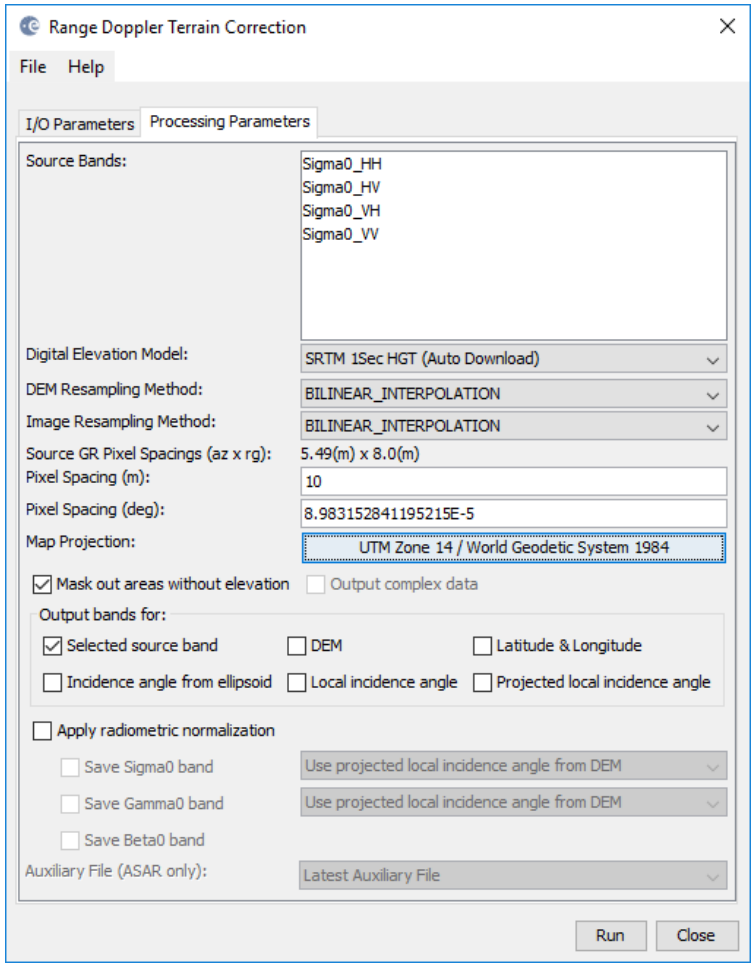
Speckle Filter – 5 by 5 Gamma Map



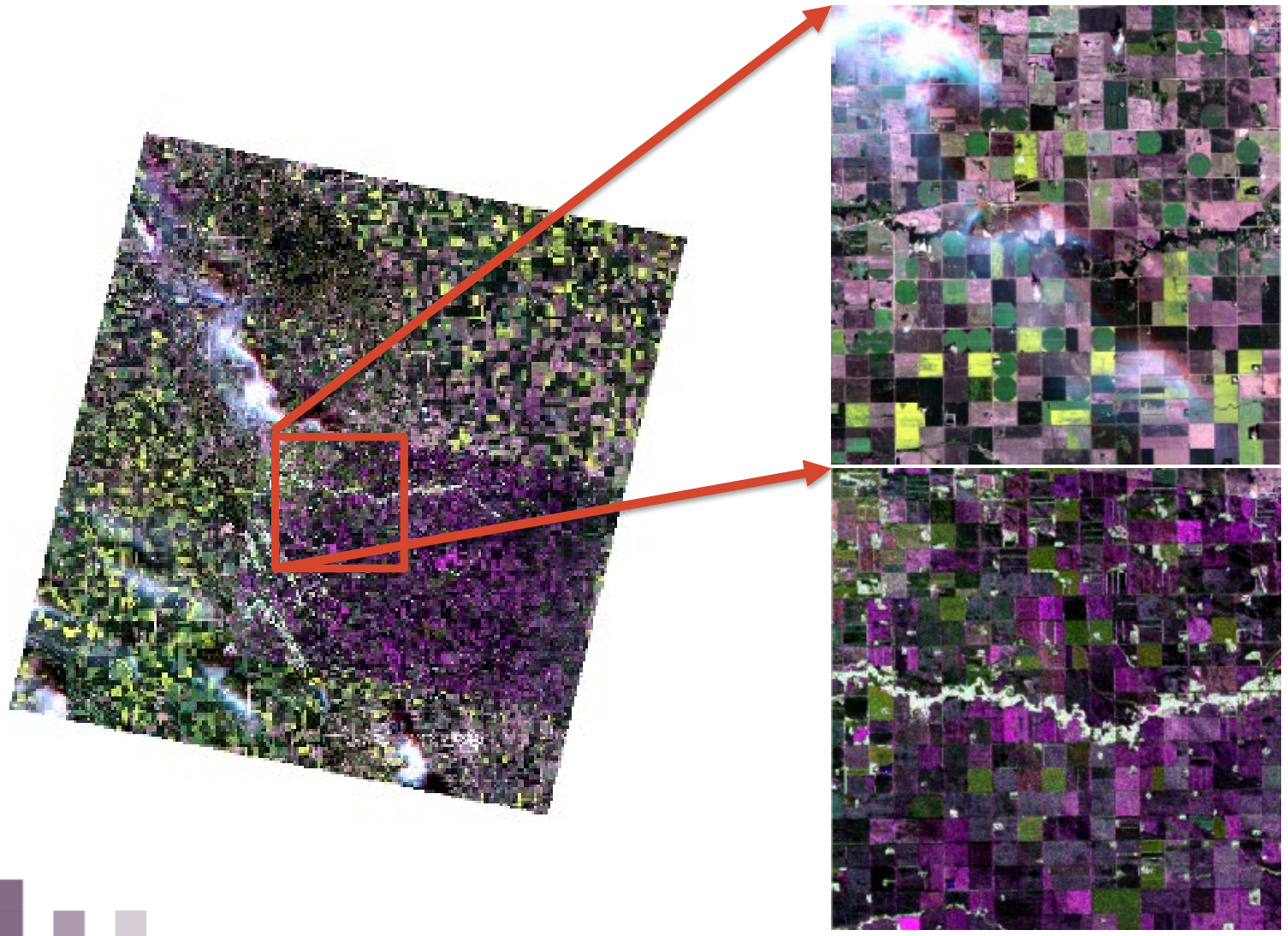
Pre-Processing RADARSAT-2 Data with SNAP

Terrain Correction





Pre-Processing RADARSAT-2 Data with SNAP



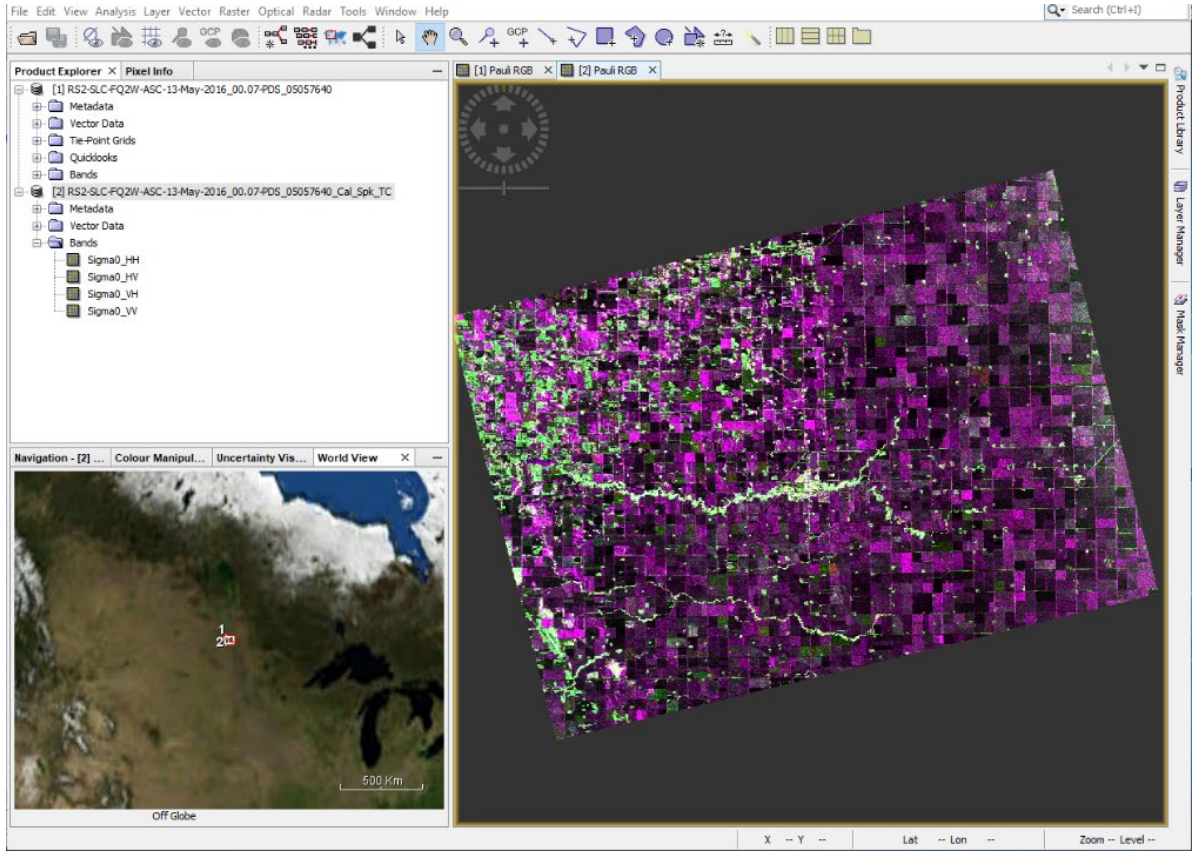
RapidEye natural color image acquired on July 4, 2016

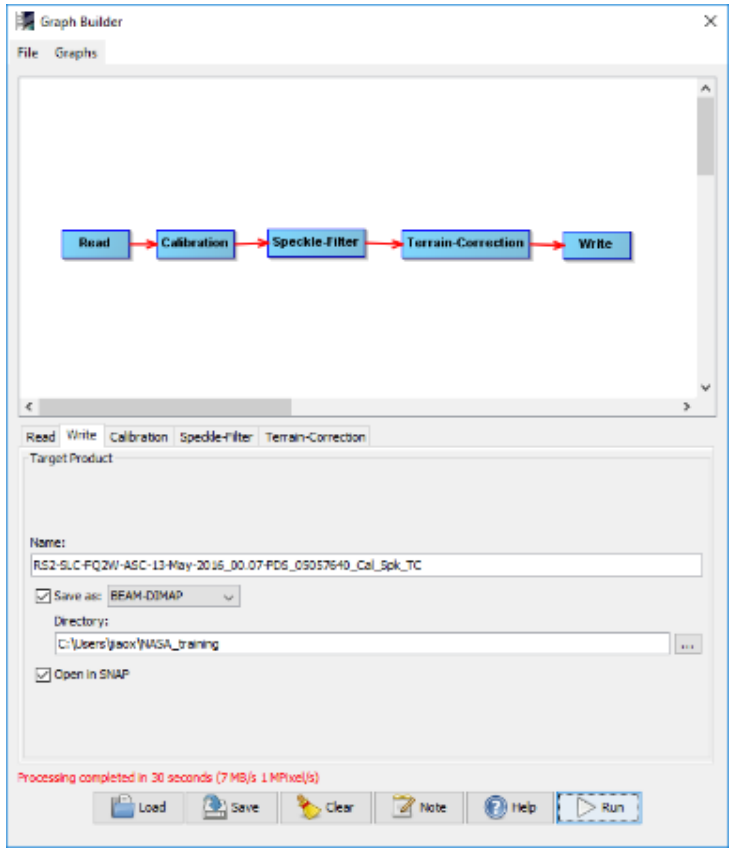
RGB color composite of RADARSAT-2 image acquired May 12, 2016

(**R=HH**, **G=HV**, **B=VV**)

Pre-Processing RADARSAT-2 Data with SNAP

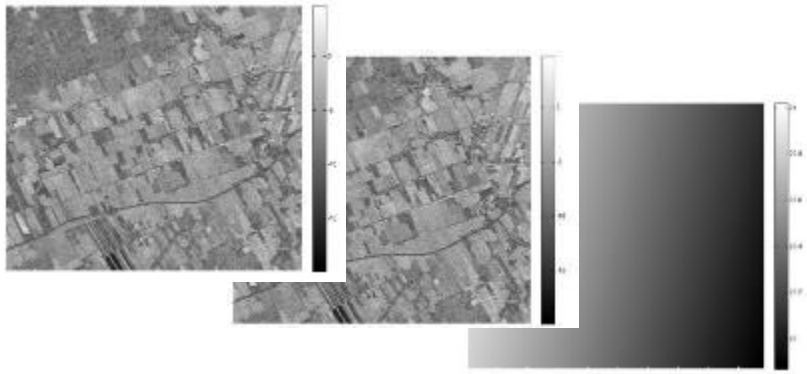
- Graph builder: used for batch processing and customized processing chains





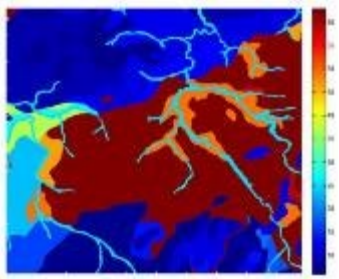
Physical Modeling Approach for Soil Moisture Estimation

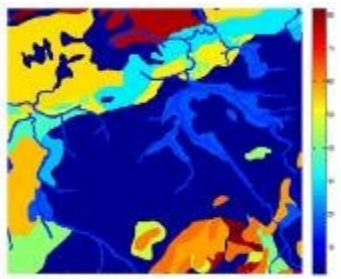
RADARSAT-2 Data



Radar Data
(HH and VV Backscatter)
(Radar Angle)

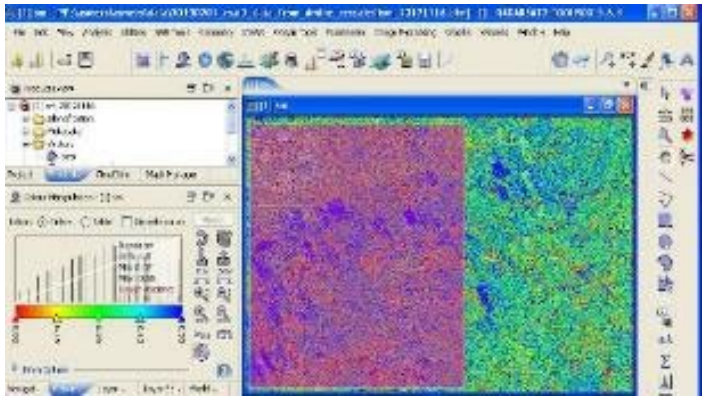
Soils Data
(Clay and Sand Fractions)





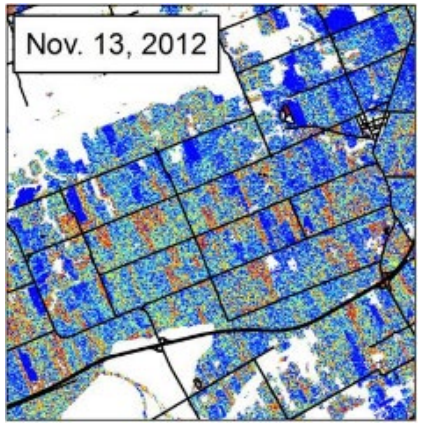
Data Processing

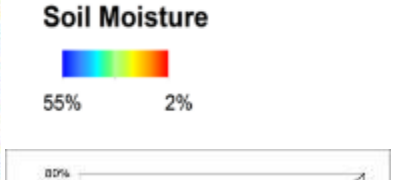
SNAP Soil Moisture Tool Box



- Uses Integral Equation Model (IDEM)
- No a priori information is needed
- Soil moisture is retrieved using only SAR data (backscatter and incidence angle)

Output and Validation





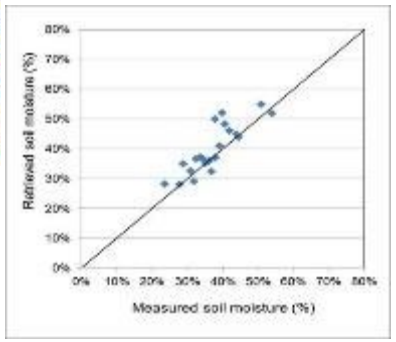
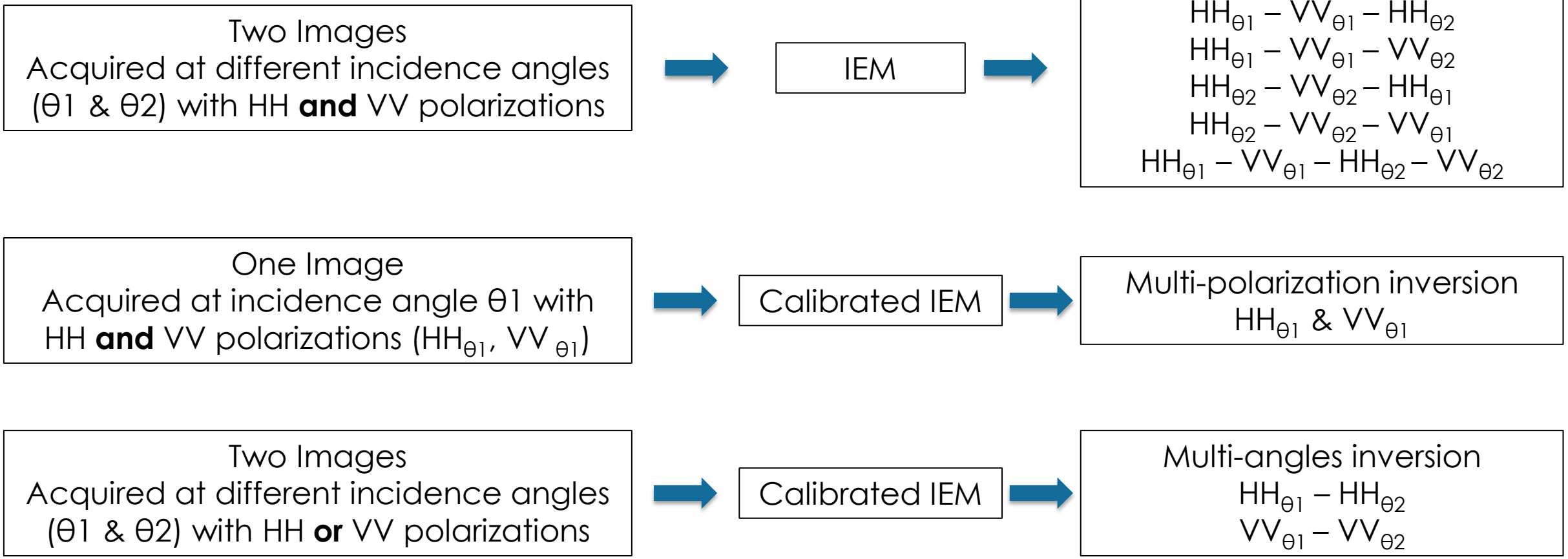


Image Source: Merzouki and McNairn, 2015

Backscatter Model Inversion Schemes



Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

- Soil moisture map derived from one a.m. RADARSAT-2 acquisition and one p.m. RADARSAT-2 acquisition, 12 hours apart

2016-05-12 RADARSAT-2 Acquisition Pair (FQ16-FQ2)

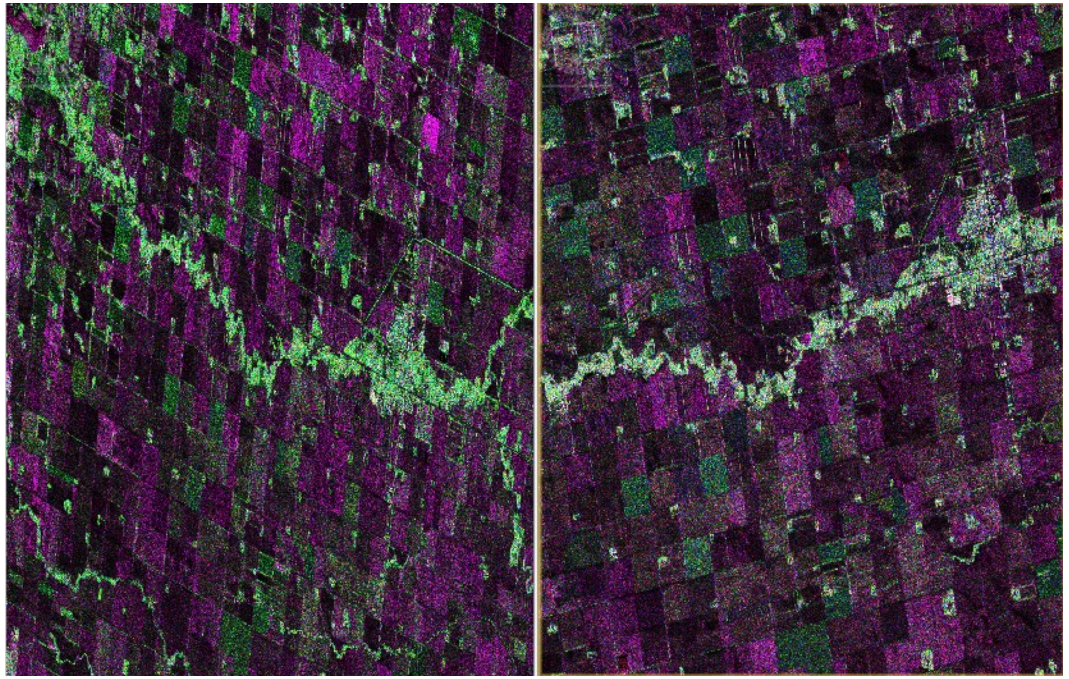
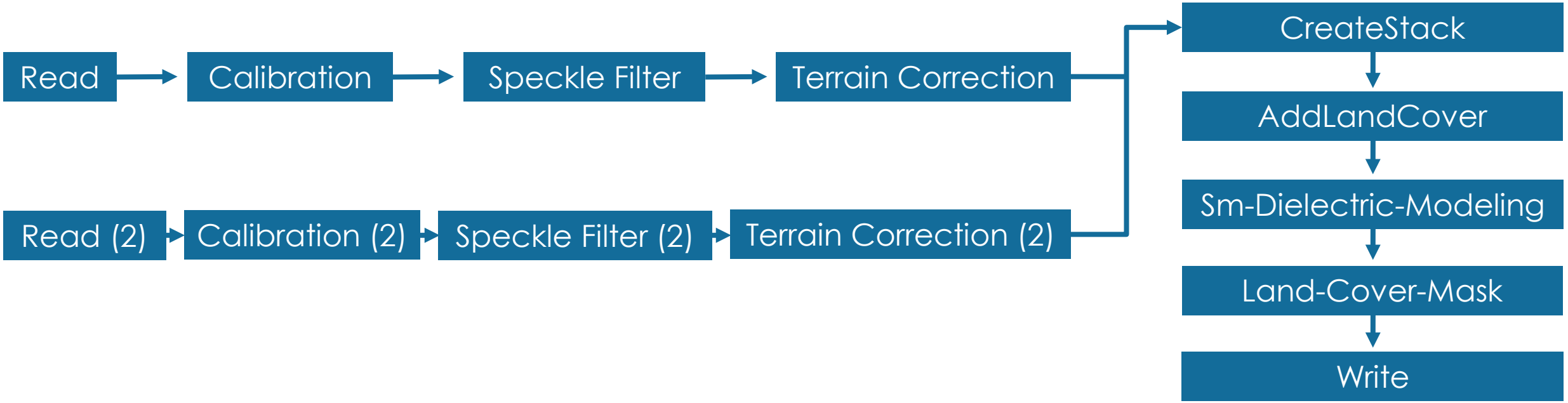


Image acquired on May 12, 2016, FQ16W Descending pass

Image acquired on May 13, 2016, FQ2W Ascending pass

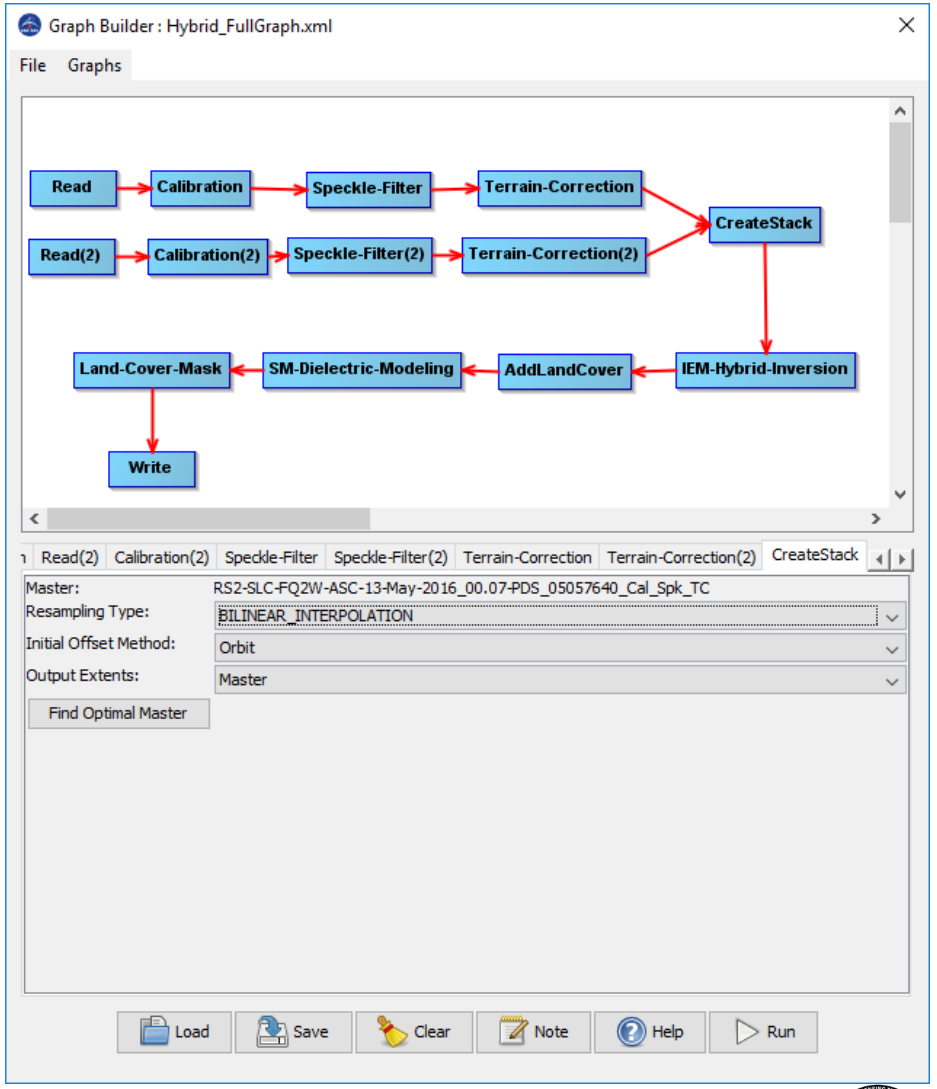
Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

- using a hybrid inversion approach, the inversion method is performed on the overlapping geographic area



Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

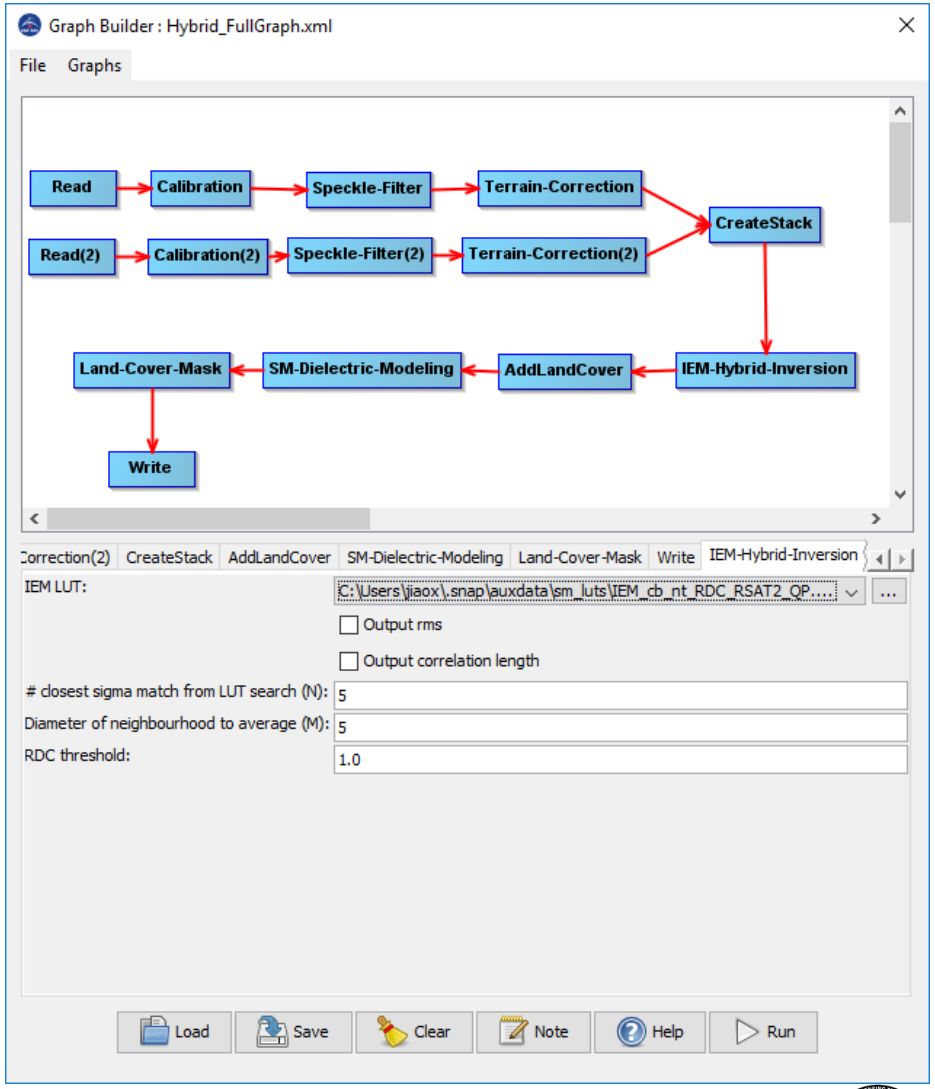
- Spatial alignment of images acquired in the morning and evening
- Create Stack:
  1. Resampling Type: Bilinear Interpolation
  2. Click **Find Optimal Master**



Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

IEM Hybrid Inversion

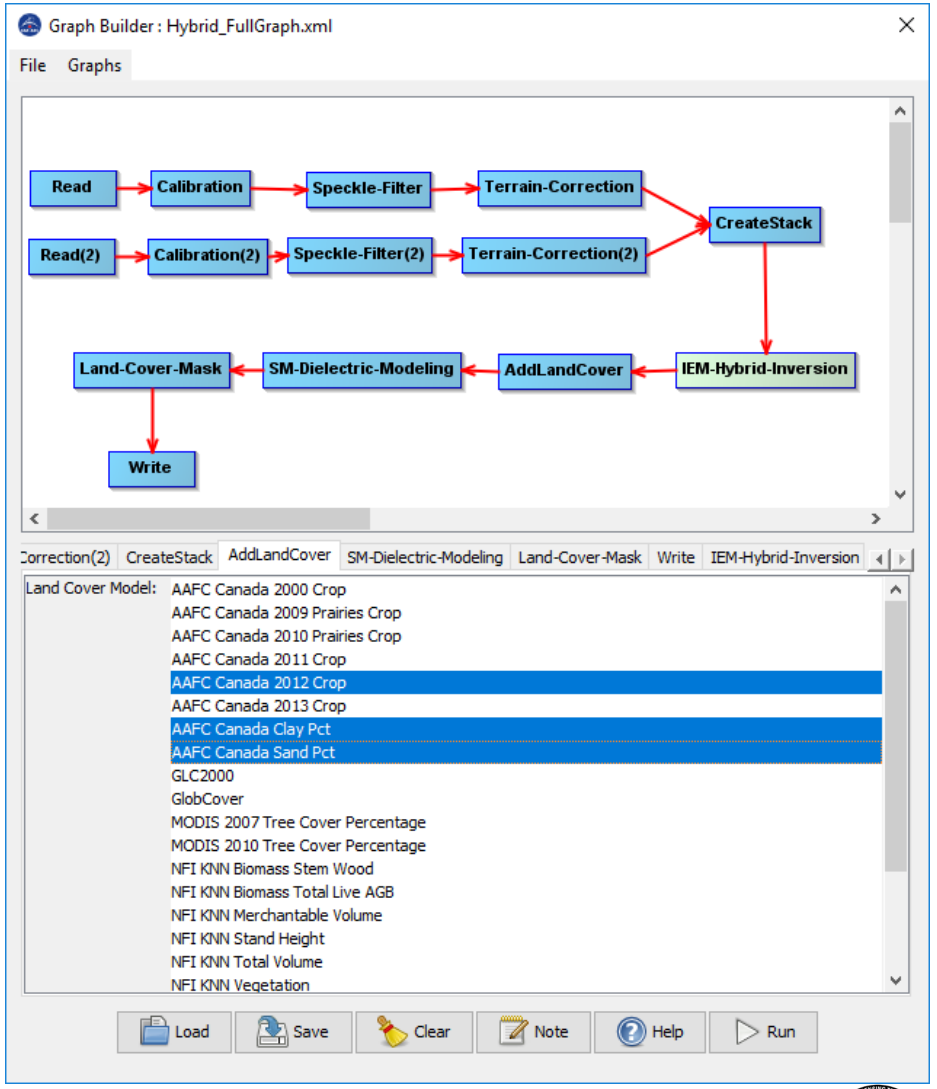
1. Select calibrated IEM LUT
2. Use the default setting for the other parameters



Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

Add Land Cover

1. Select the land cover file
2. Select sand and clay fractions maps

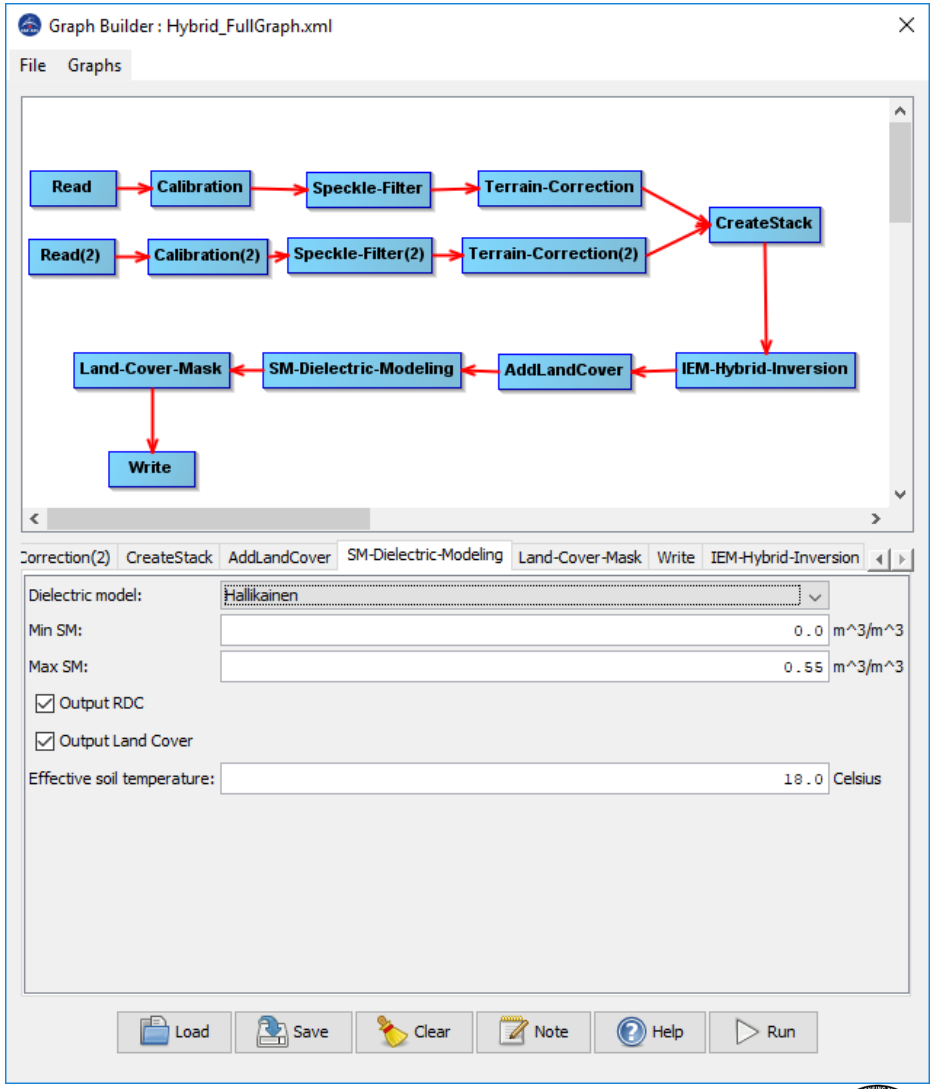


Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

SM Dielectric Modeling:

1. Select **Hallikainen** mixture model
2. Use the default values for the other parameters

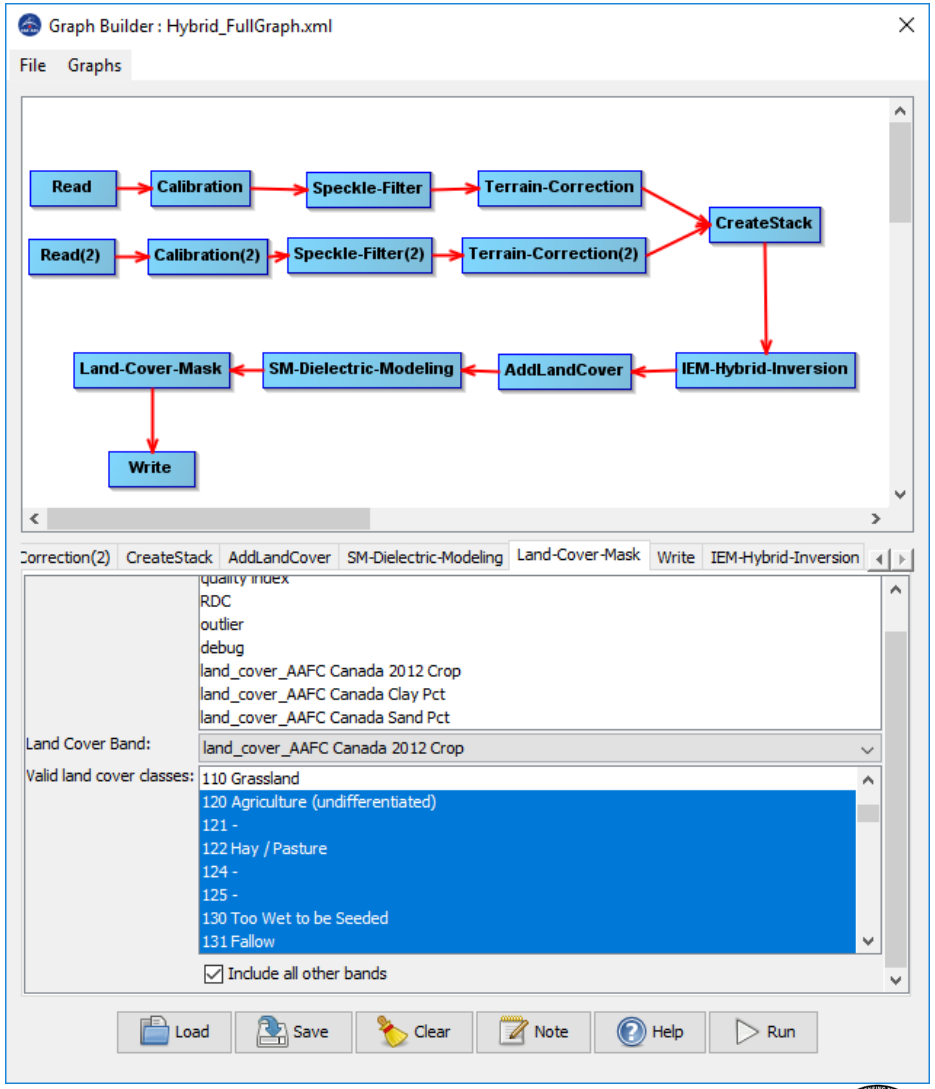
The Haillkainen model was used to estimate the volumetric soil moisture, which is based on retrieved dielectric values. This model requires the soil texture information (clay and sand fractions)



Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

Land Cover Mask

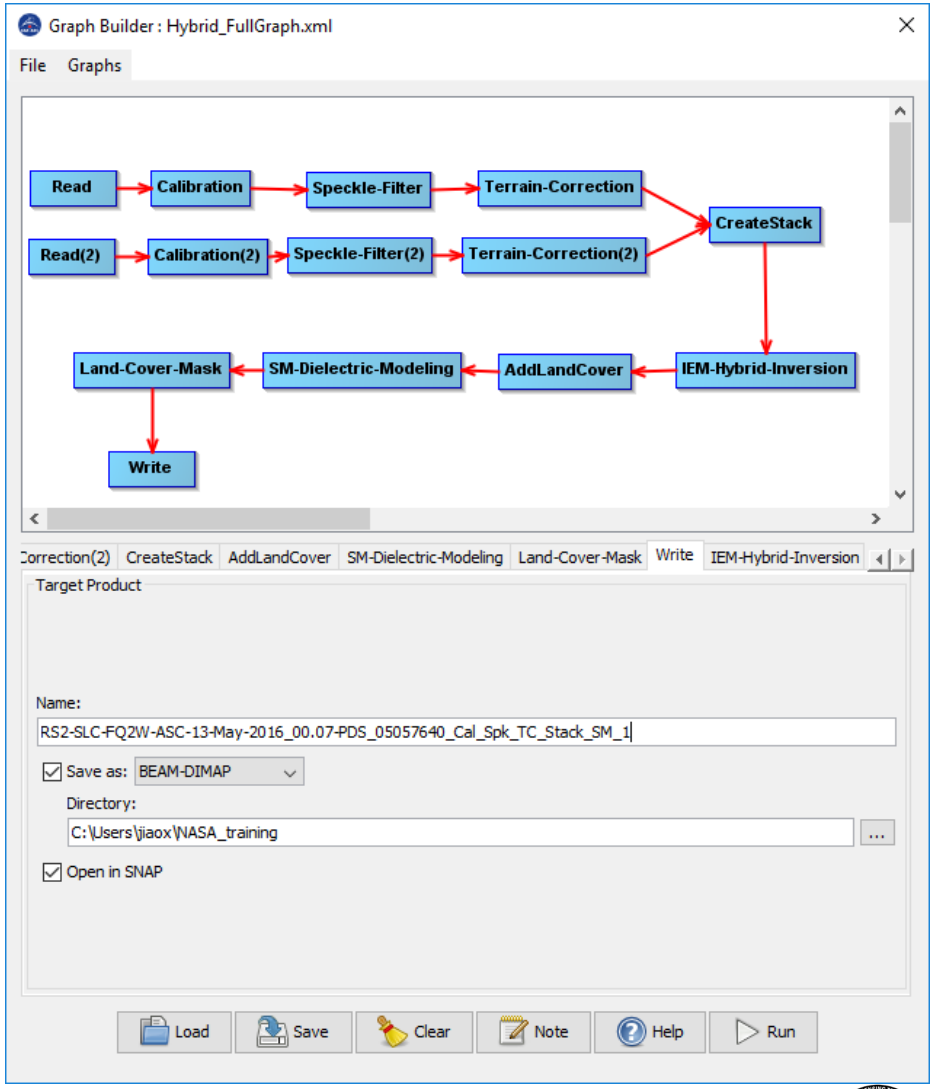
1. Select land cover valid agriculture classes
2. Check **Exclude all other bands**



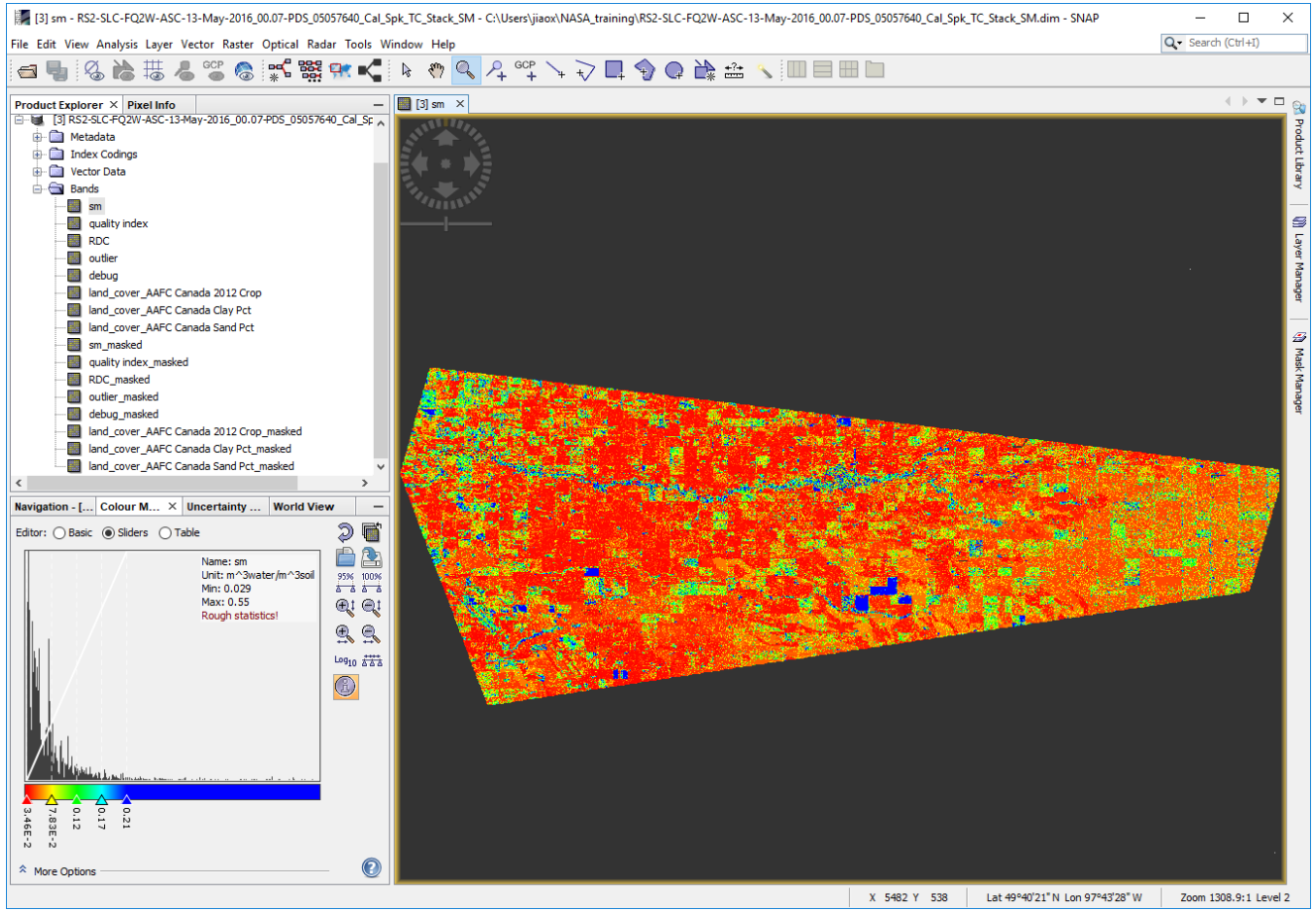
Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

Write Output

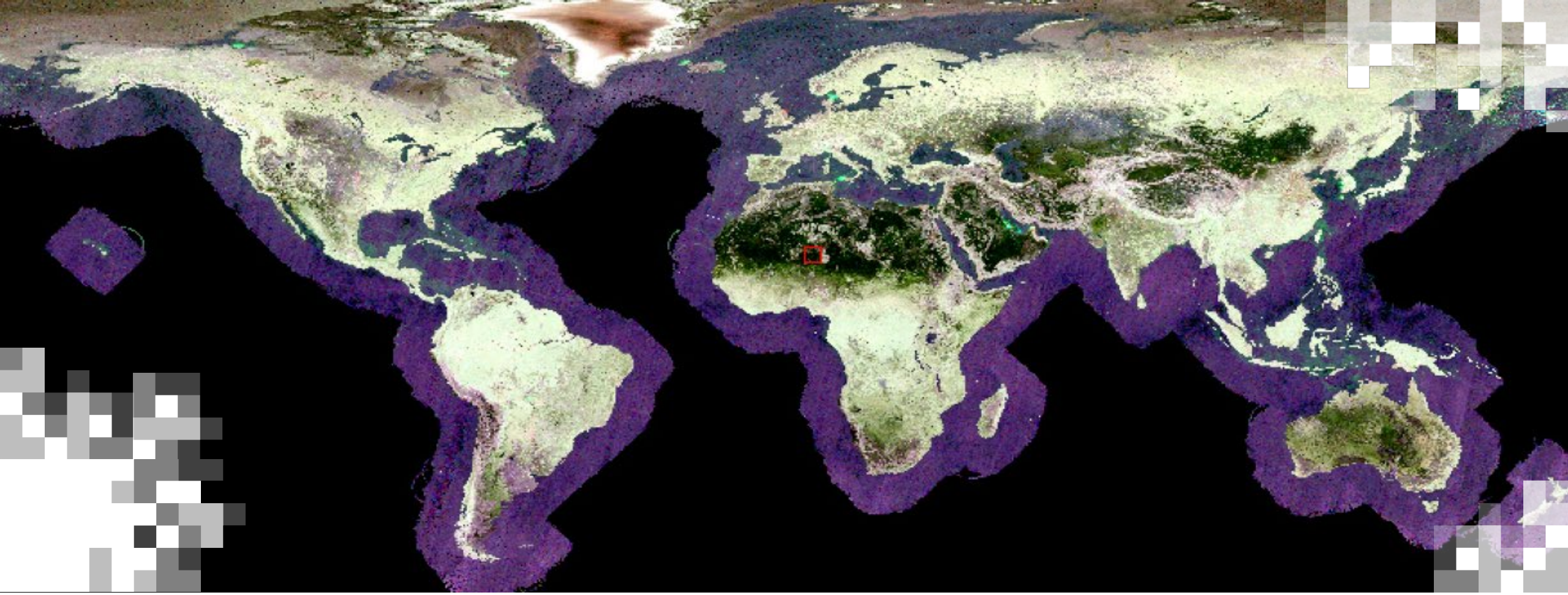
1. Browse and name the output file
2. Select the appropriate format for the retrieved soil moisture product
3. Run the module



Soil Moisture Processing with the Soil Moisture Toolbox in SNAP – Hybrid Scheme

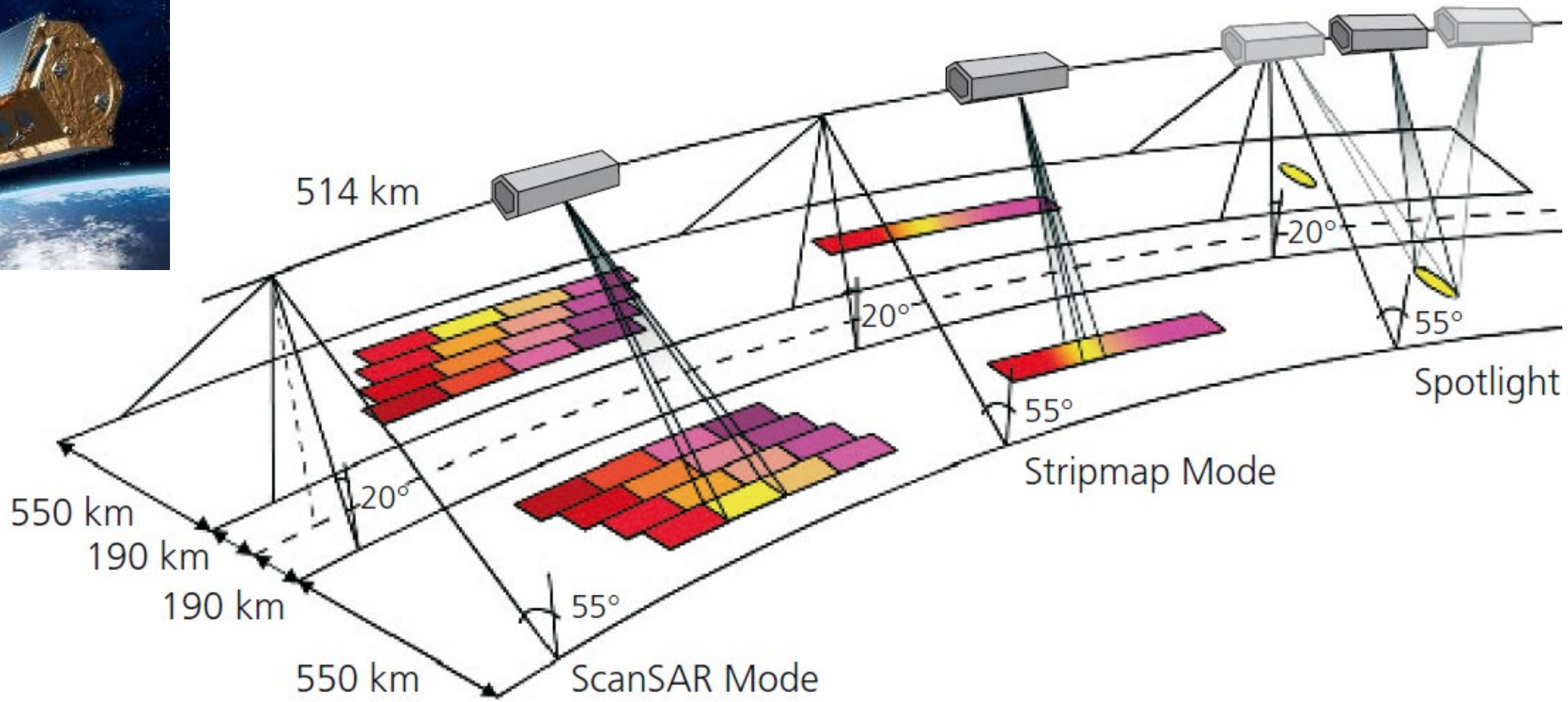


Soil moisture map obtained by inverting the IEM and using a pair of RADARSAT-2 images acquired May 12/13, 2016, in Southern Manitoba



Processing Multi-Frequency Data for Use in Crop Classification

TerraSAR-X



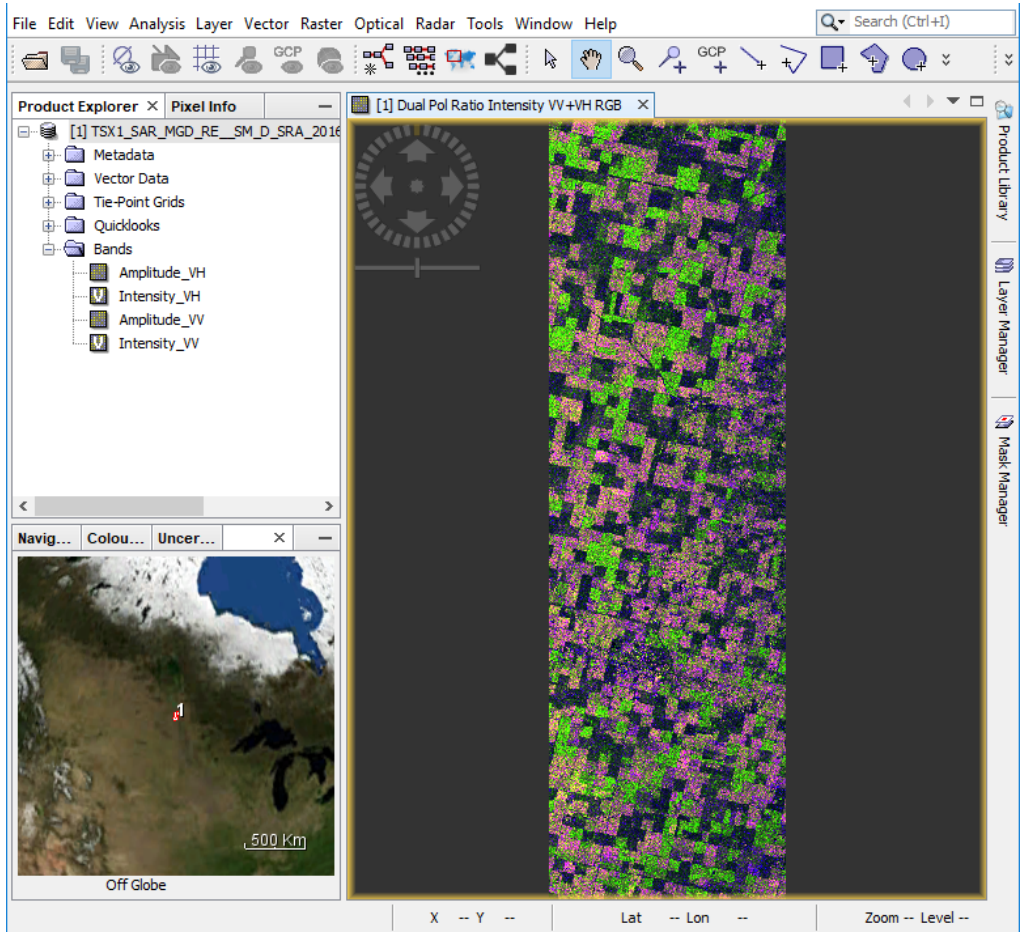
Overview of the TerraSAR-X Scanning Modes - Revisit time: 11 days

Image Credits: DLR

TerraSAR-X StripMap Dual Pol Data

TerraSAR-X StripMap dual polarization Multi-Look Ground Range Detected (MGD) product

- Nominal Resolution: 1.2 m (range) * 6.6 m (azimuth)
- Nominal Scene Size: 15 km (range) * 50 km (azimuth)
- Dual Polarization: HH&VV, HH&HV, or VV and HH



TerraSAR-X StripMap dual pol MGD data acquired on July 27, 2016, over Carman, Manitoba, Canada

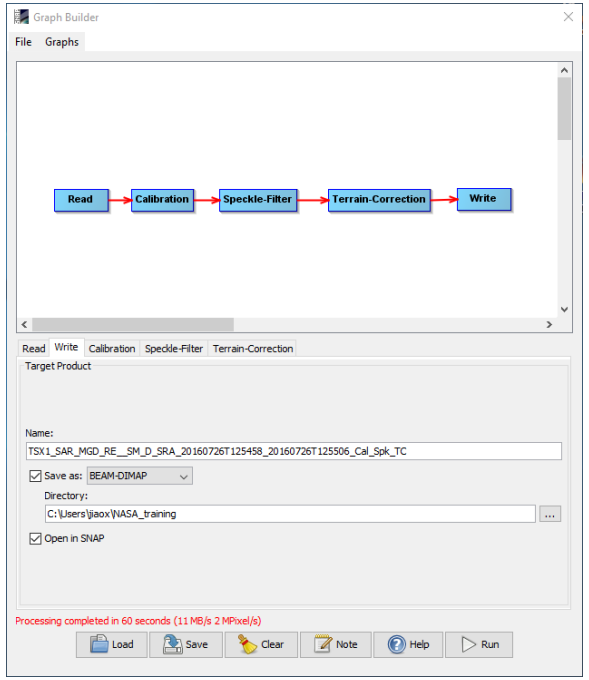
Pre-Processing RADARSAT-2 Data with SNAP

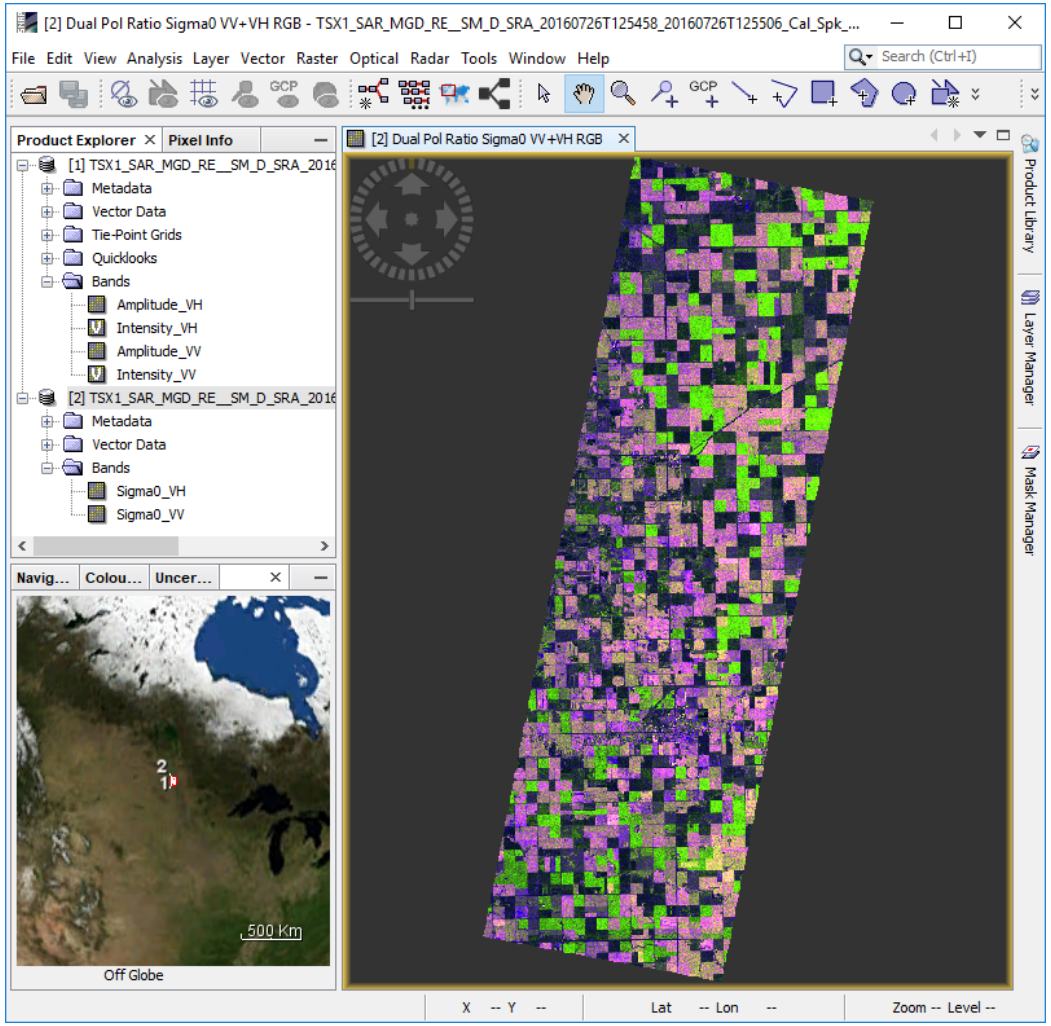
Extract Backscatter

Pre-Processing TerraSAR-X Data with SNAP

Graph Builder

- Import the data
- Calibrate pixel value to radar backscatter
- Gamma map filtering, 7 by 7 window size
- Terrain Correction:
  - bilinear interpolation resampling
  - 5 m pixel spacing
  - UTM zone 14 projection





Sentinel-1

- Coverage
  - Sentinel-1consists of two satellites: A (2014) and B (2016)
  - Each Sentinel-1 satellite has a 12 day repeat cycle
  - The two satellites offer a 6-day exact repeat cycle at the equator in the interferometric wide swath mode of acquisition

Extra Wide Swath (EW)	Interferometric Wide Swath (IW)	Stripmap (SM)	Wave (WV)
Acquired with TOPSAR using 5 sub-swaths instead of 3, resulting in lower resolution (20m-x-40m). Intended for maritime, ice, and polar zone services requiring wide coverage and short revisit times.	Acquired with TOPSAR. Default mode over land; 250km swath width; 5m-x-20m ground resolution.	Used in rare circumstances to support emergency-management services, 5m-x-5m resolution over an 80km swath width.	Default mode over oceans; VV polarization. Data acquired in 20km-x-20km vignettes, 5m-x-20m resolution, every 100km along the orbit.

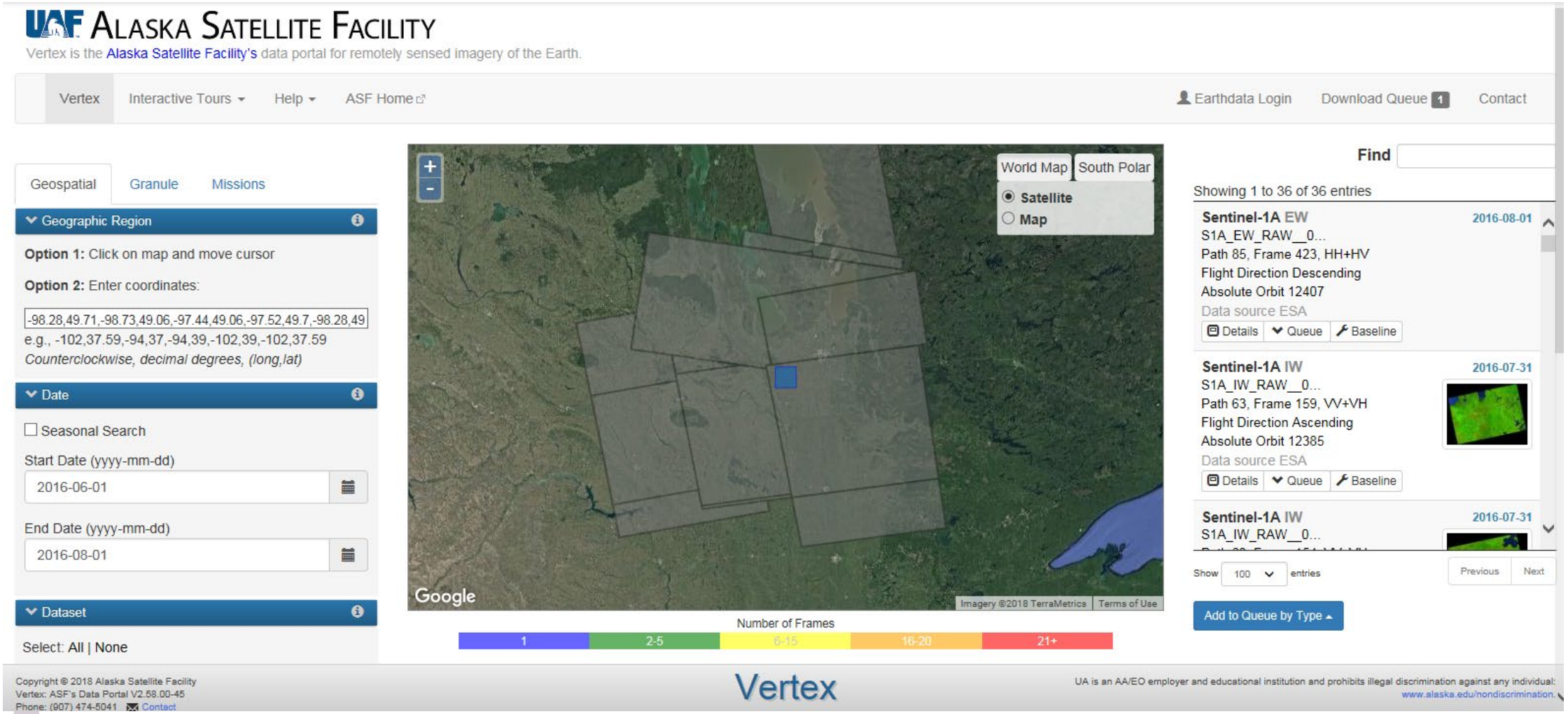
Product Type for IW Mode:

Acq. Mode	Product Type	Resolution Class	Resolution [1,2] [Rng x Azi] [3] [m]	No. Looks [Rng x Azi]
IW	SLC		2.7 x 22 to 3.5 x 22	1
	GRD	HR	20 x 22	5 x 1
		MR	88 x 87	22 x 5

Satellite Image Credit: ESA/ATG medialab

Access Sentinel-1 SAR Data from Vertex

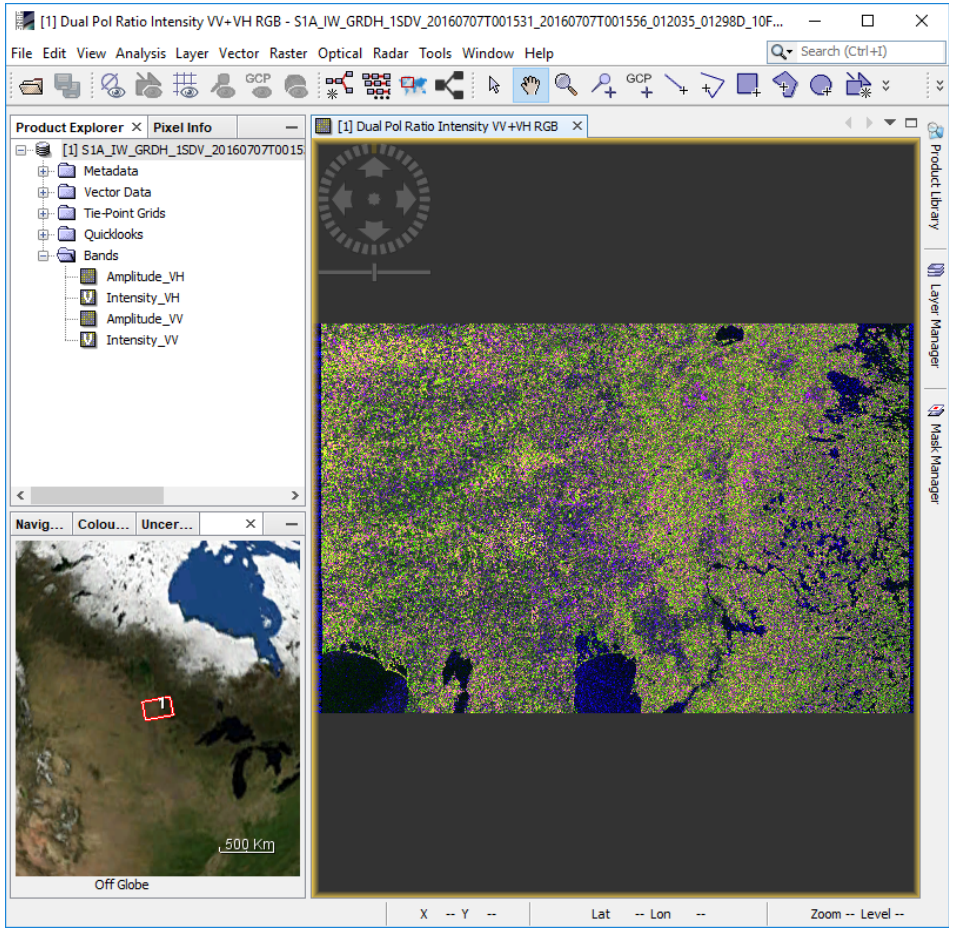
https://vertex.daac.asf.alaska.edu/



Sentinel-1 Ground Range Detected Data

Sentinel-1 IW mode ground range detected product:

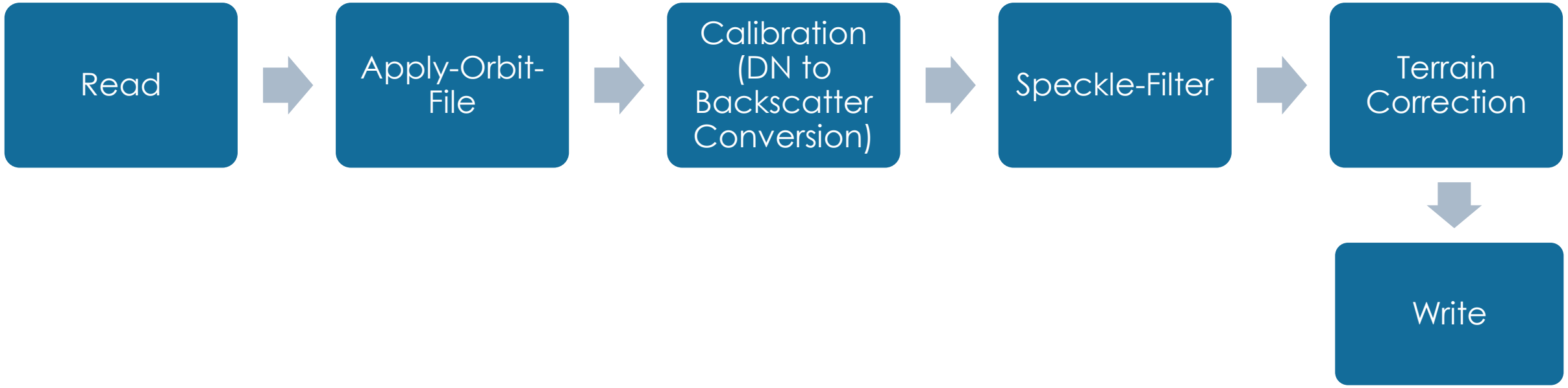
- Nominal Resolution: 20 m (range) * 22 m (azimuth)
- 250 km swatch
- Dual Polarization: HH /HV, or VV/VH
- GRDH:
  - ground range detected
  - high resolution
  - Multi-Looked: 5 (range)*1 (azimuth)
  - Number of looks and project to ground range
- Phase information is lost



Sentinel-1 GRDH dual pol data acquired on July 7, 2016, over Carman, MB, Canada

Pre-Processing Sentinel-1 SAR GRDH Data with SNAP

Extract Backscatter



Apply the Precise Orbit File

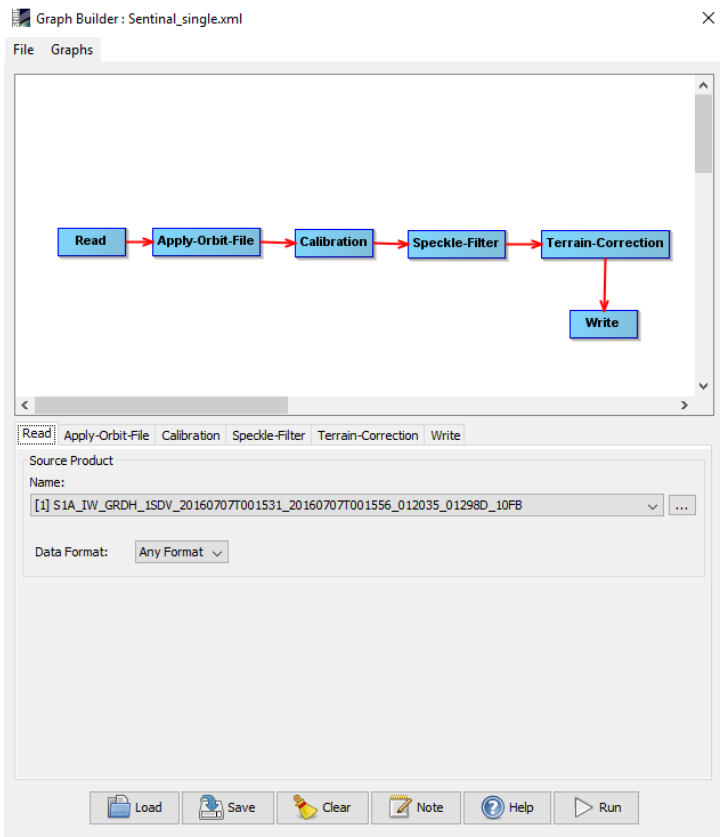
- For Sentinel-1:
  - During the acquisition the satellite position is recorded by a Global Navigation Satellite System (GNSS)
  - To assure fast delivery of Sentinel-1 products orbit information generated by an on-board navigation solution are stored within the Sentinel-1 Level-1 products
  - The orbit positions are later refined by the Copernicus Precise Orbit Determination (POD) Service
  - Precise orbit files have less than 5 cm accuracy and are delivered within 20 days after data acquisitions
  - The accuracy of restituted orbit files is less than 10 cm. The files are available 3 hours after data acquisitions
  - The orbit information of Sentinel 1 can be downloaded from the ESA website (https://qc.sentinel1.eo.esa.int/)

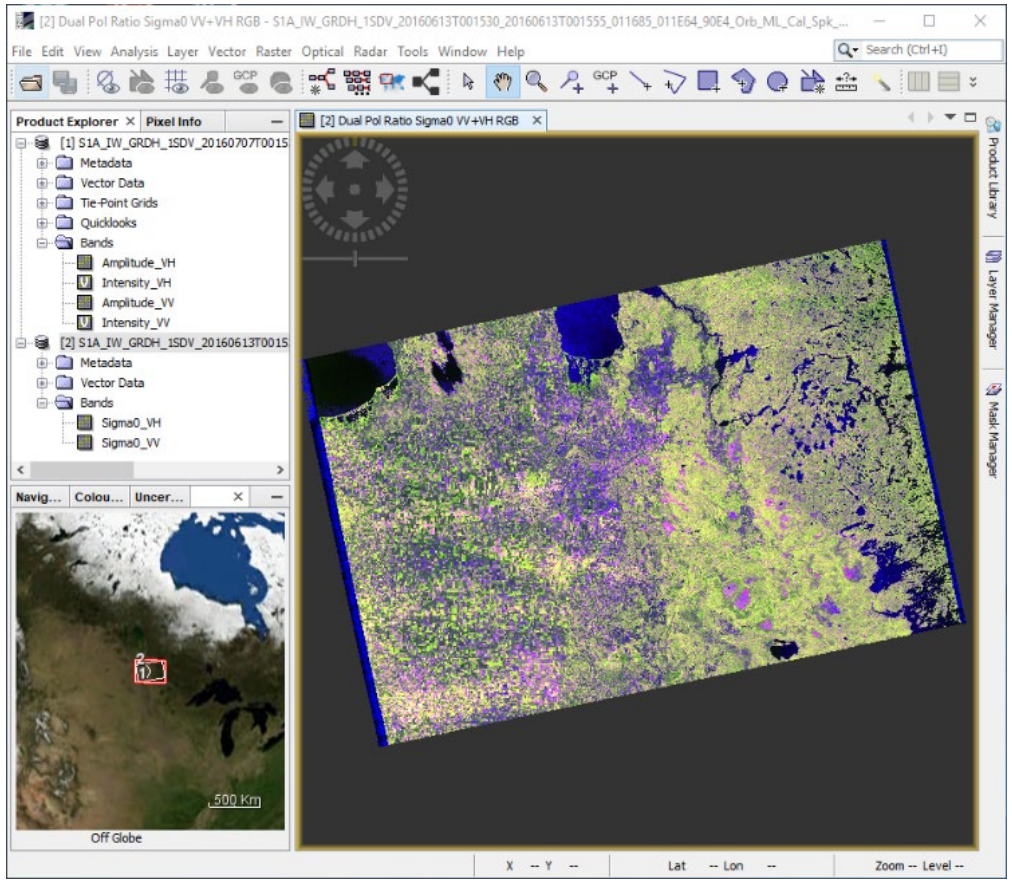
SNAP downloads orbit files and stores these into the folder …/auxdata/Orbits/Sentinel-1/
https://www.asf.alaska.edu/sentinel/data/

Pre-Processing Sentinel-1 SAR GRDH Data with SNAP

Graph Builder

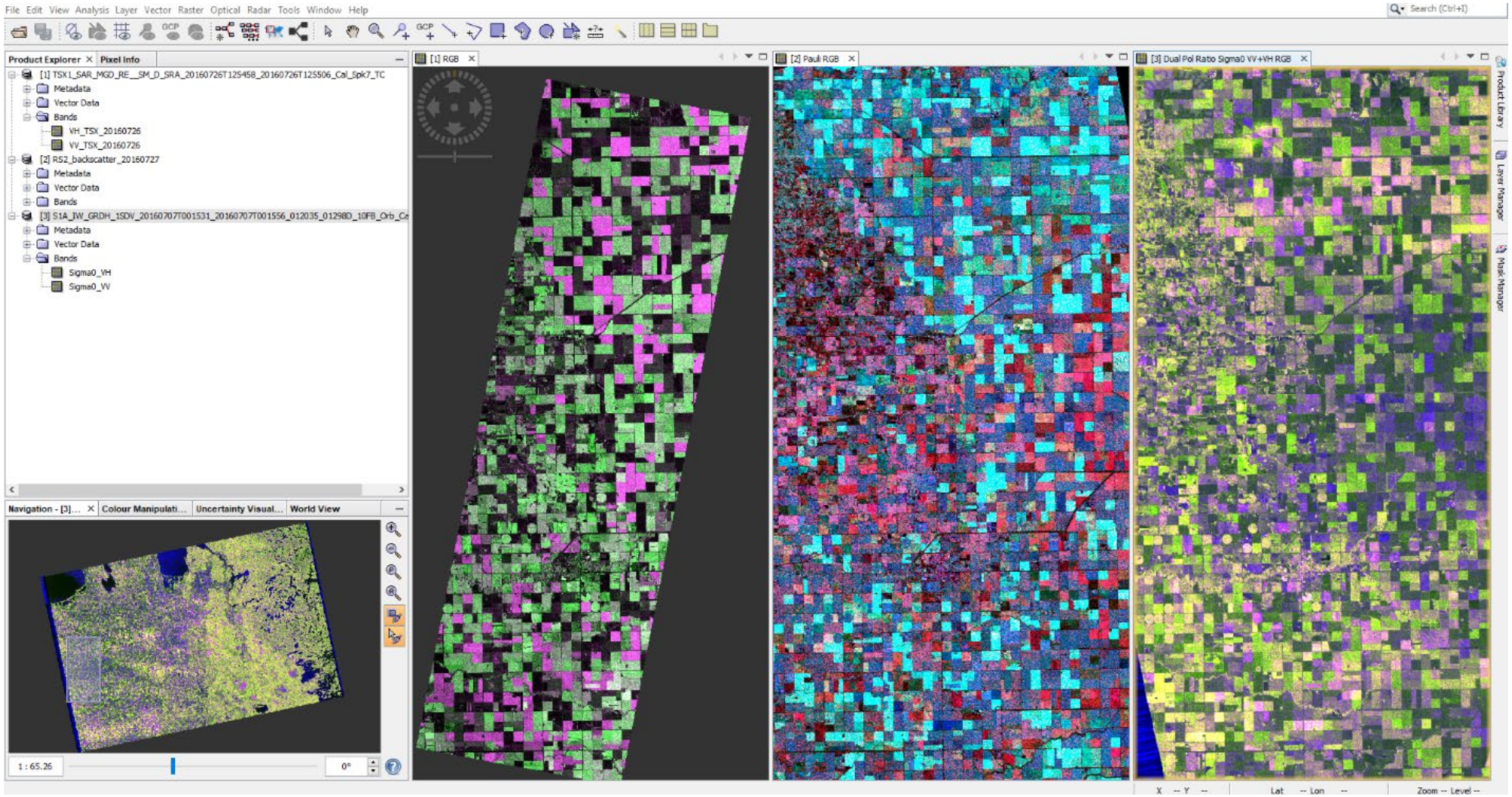
- Import the data
- Apply orbit file
- Calibrate pixel value to radar backscatter
- Gamma Map filtering, 3 by 3 window size
- Terrain Correction:
  - Bilinear interpolation resampling
  - 30 m pixel spacing
  - UTM Zone 14 Projection





Images Feed Into the Classifier

- RADARSAT-2 Wide Fine Quad-Pol images acquired on July 3rd 2016, July 27th and August 20th 2016
- TerraSAR-X StripMap dual pol MGD images acquired on July 26th and August 17th, 2016
- Sentinel-1 IW mode GRDH images acquired on July 13th, July 7th and July 31th, 2016



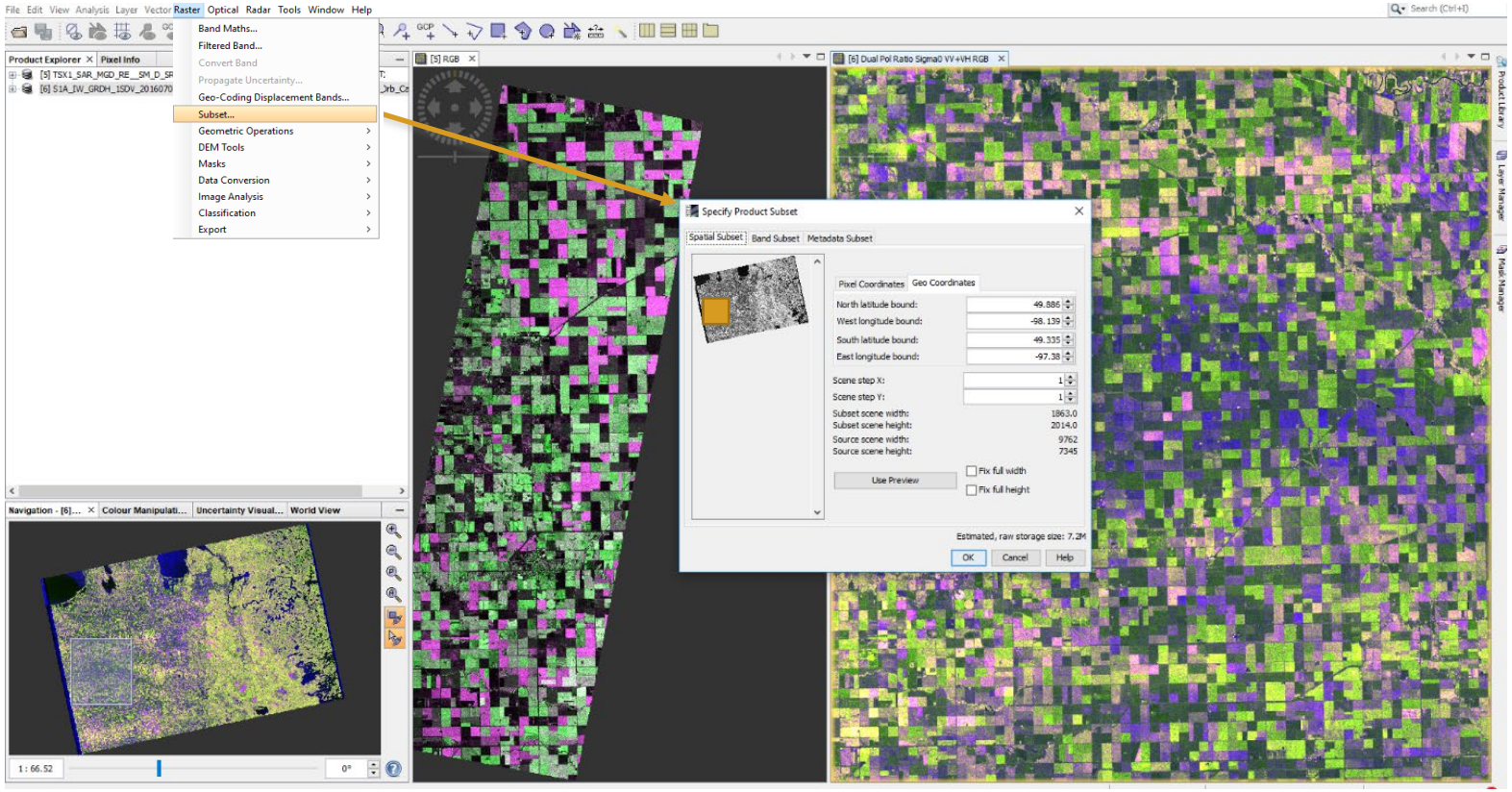
TerraSAR-X StripMap dual pol MGD images acquired on Jul 26 and Aug 7, 2016

RADARSAT-2 Wide Fine Quad-Pol images acquired on Jul 3, Jul 27, and Aug 20, 2016

Sentinel-1 IW mode GRDH images acquired on Jul 13, Jul 7, and Jul 31, 2016

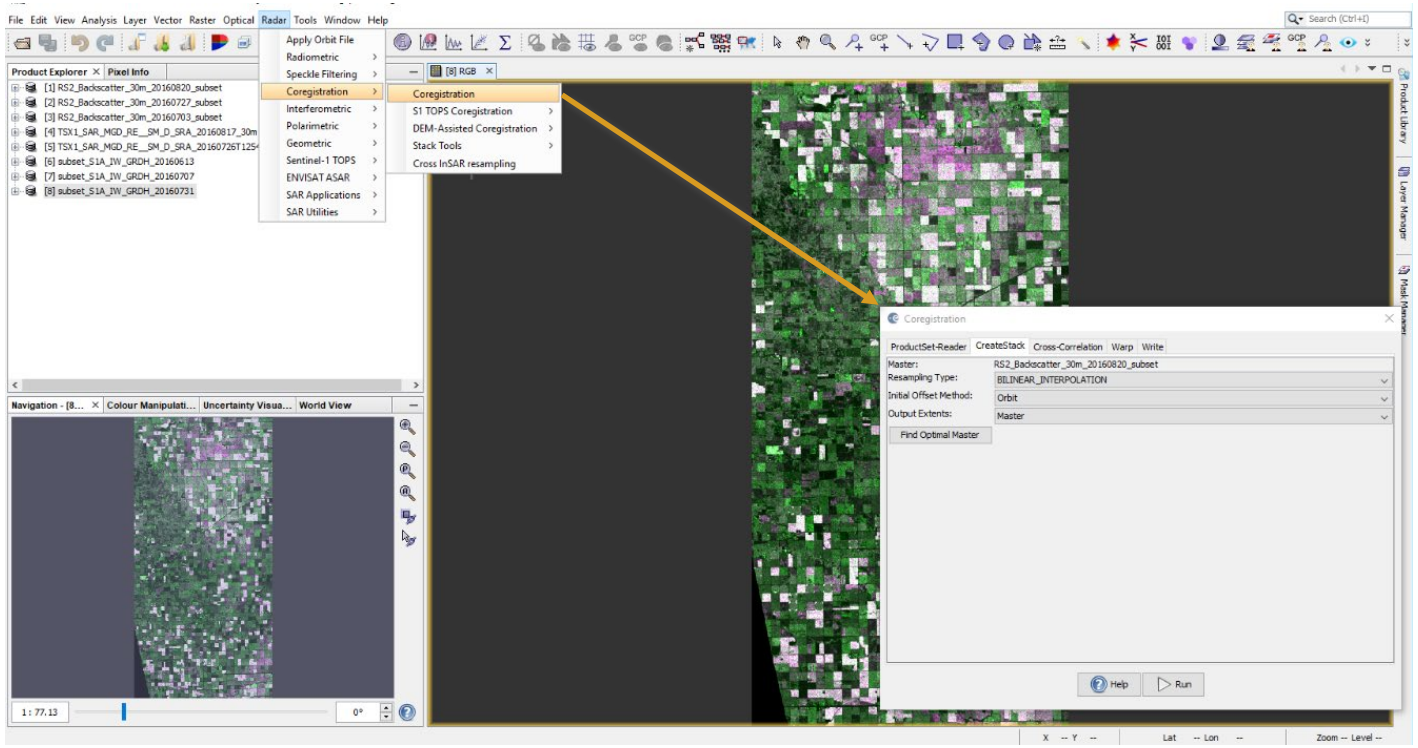
Subset Raster to AOI

- Go to Raster Menu >> Subset:
  a) Spatial Subset tab → enter the upper left and lower right coordinate under geo coordinates
  b) Band Subset → select bands you wish to subset
  c) Metadata Subset: leave as default



Co-Registration of Subset Images

- Spatial alignment of images
- Go to Radar Menu >> Coregistration >> Coregistration:
  a) ProductSet-Reader: Click Plus sign with line over top adds all open imagery → Click Revolving Arrows refreshes metadata
  b) Create Stack: Resampling Type → Bilinear_Interpolation → Click Find Optimal Master
  c) Other tabs: leave as default; ensure Write folder is not over-writing previous files
  d) Click Run and Close window when completed



Exporting Data Stacks Out of SNAP

- Spatial alignment of images
  - Go to Radar Menu >> File >> export >> select file format .tif
- Can use .tif in R, Python etc.