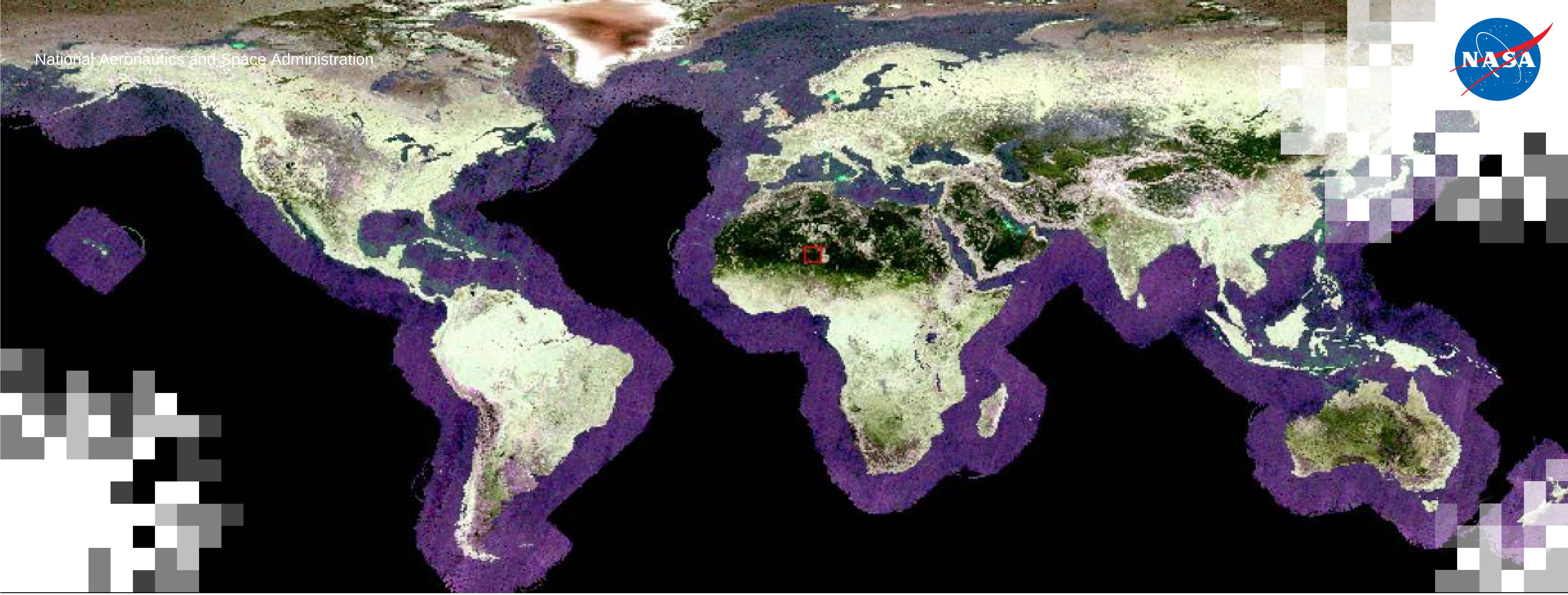
National Aeronautics and Space Administration



SAR para el Monitoreo Agrícola

Heather McNairn, Xianfeng Jiao, Sarah Banks y Amir Behnamian

4 de septiembre de 2019

Objetivos de Aprendizaje

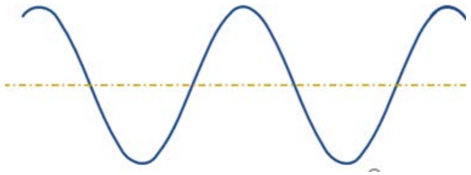
Al finalizar esta presentación podrán entender:

- cómo las diferentes configuraciones de SAR afectan la señal en relación a los suelos y cultivos
- el contenido informático en las imágenes SAR relevante a las condiciones del suelo y los cultivos
- los parámetros óptimos de los sensores SAR para aplicaciones agrícolas
- cómo asimilar, pre-procesar y procesar datos SAR para la clasificación de cultivos y la estimación de la humedad del suelo

SAR- Aspectos por Considerar

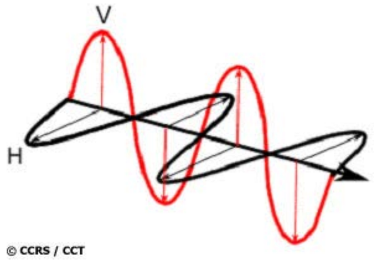
- Al planificar la recolección de datos SAR y al interpretar la señal de SAR, siempre hay que considerar tres características del sistema

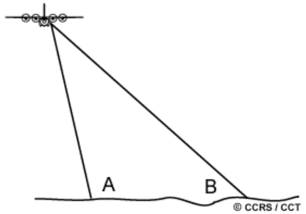
- La interpretación de la señal de SAR siempre se hace en relación a estas características



frecuencia o longitud de onda

polarización





geometría (ángulo de incidencia y dirección de mirada o look direction)

Fuente de las Imágenes: Polarization & Geometry CCRS/CCT

Frecuencia o Longitud de Onda

Selección de la frecuencia más efectiva

- considere el tamaño de los elementos en la superficie en relación a la frecuencia del SAR. Para maximizar la dispersión, elija ondas con longitudes de tamaño comparable o menor a estos elementos
- ¿es importante penetrar a través del medio (vegetación, suelo), o es la meta maximizar la dispersión en la superficie? Las frecuencias bajas (ondas de mayor longitud) ofrecen mayor penetración
- ¿es la meta maximizar o minimizar la sensibilidad a la rugosidad de la superficie? Una onda de frecuencia baja verá una superficie como lisa mientras que una onda de frecuencia alta verá la misma superficie como rugosa

¿Cuál es la mejor frecuencia para el monitoreo agrícola? ¡Depende!

- humedad del suelo: las ondas de mayor longitud (como Banda-L) son mejores porque penetran la vegetación mejor e interactúan con el suelo
- clasificación de cultivos y modelación biofísica: depende de la vegetación
- es necesario penetrar lo suficiente a través de la vegetación (por ejemplo, utilizando Bandas L- o C para maíz,) pero no demasiado ya que puede haber interferencia del suelo (Bandas C o X para cultivos de menor biomasa como soya)

El Poder de Utilizar Múltiples Frecuencias

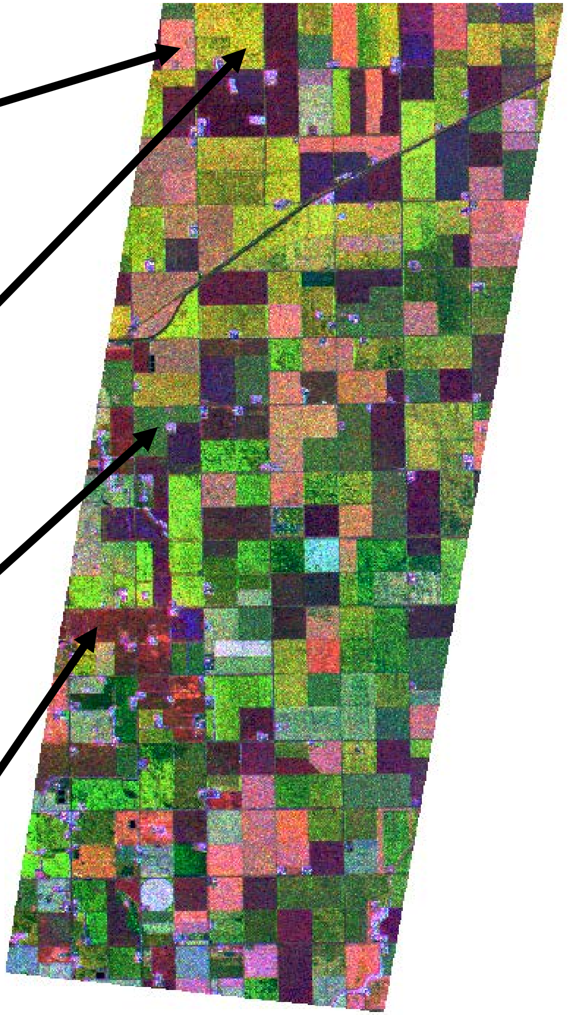
Integración de datos de RADARSAT-2, ALOS y TerraSAR-X, Manitoba (Canadá)

canola

soya

maíz

trigo



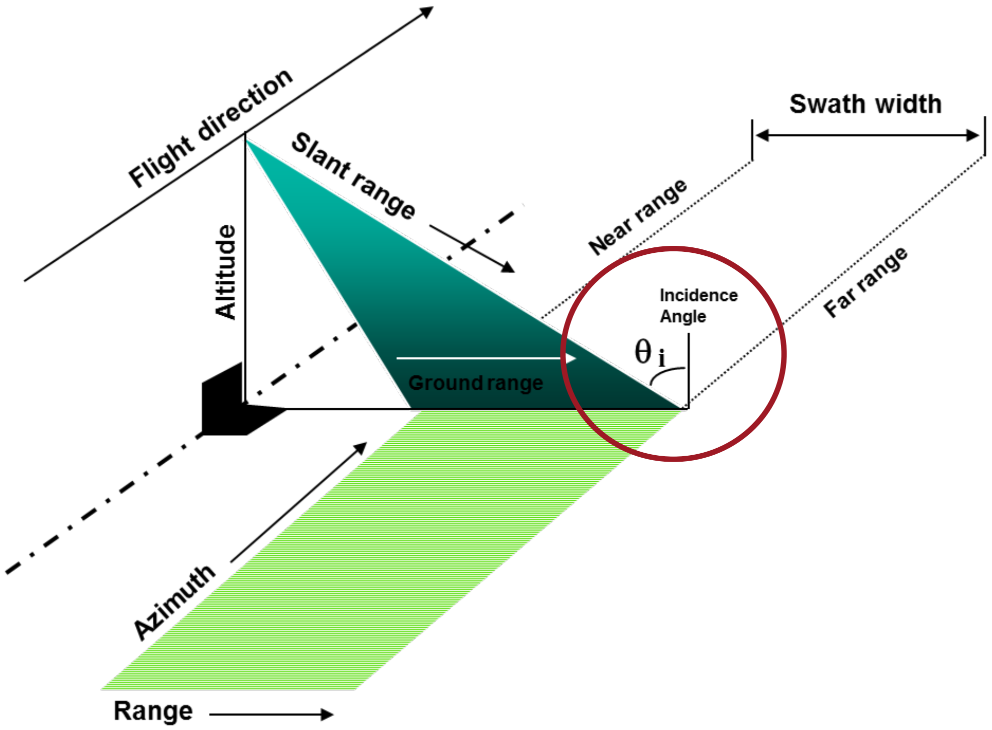
X (3,1 cm)
C (5,5 cm)
L (22,9 cm)

RADARSAT-2 VH 10 de julio de 2016

TerraSAR-X VV 10 de julio de 2016

ALOS VH 9 de julio de 2016

Ángulo de Incidencia

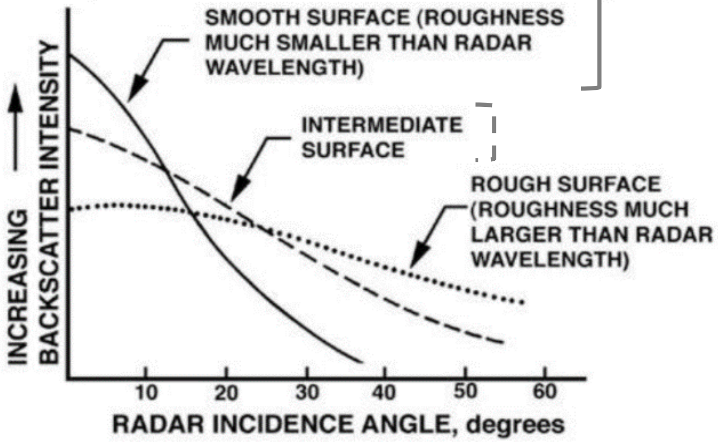


El ángulo de incidencia determina

- la contribución de diferentes elementos de la superficie a la retrodispersión. Los ángulos menores interactúan más con el follaje de la vegetación; los ángulos mayores permiten que la señal pase de largo hasta el suelo sin interactuar con el follaje
- cuán áspero o rugoso el SAR percibe los elementos en la superficie. Las superficies se ven "más lisas" desde un ángulo mayor. Los efectos más significativos del ángulo de incidencia se observan en las superficies más lisas
- la profundidad de la penetración de la señal a través del medio

Ángulo de Incidencia

- la retrodispersión disminuye al incrementar el ángulo de incidencia
- la tasa y función de la disminución dependen de las características de la superficie
- por lo tanto, cuando un radar observa la misma superficie en diferentes ángulos, la retrodispersión será diferente
- **PRECAUCIÓN:** para detectar cambios a través del tiempo, no mezcle ángulos (utilice repeticiones exactas)
- para estimaciones biofísicas, está bien mezclar ángulos si es que el modelo toma en cuenta el ángulo de incidencia



más áspero	más liso
menor tasa de cambio con el ángulo	mayor tasa de cambio con el ángulo

Fuentes de Imágenes: (izq.) Ulaby et al. (1981a); (der) SOEST University of Hawai'i

Polarización

La polarización determina

- cómo las microondas transmitidas interactúan con la superficie
  - si la superficie (por ejemplo, vegetación) tiene una estructura vertical dominante, las ondas en polarización vertical (V) se alinean con esta estructura y crean más retrodispersión. Con las ondas de polarización horizontal (H), menos de la energía interactúa con la estructura vertical de la vegetación y más ondas llegan al suelo a través del follaje
- al considerar las señales de transmisión y recepción, la cantidad de energía que es repolarizada (se transmite H pero se recibe V: se transmite V pero se recibe H) para crear polarizaciones cruzadas (HV o VH), depende de la estructura de la superficie

¿Cuál es la mejor polarización para el monitoreo agrícola?

- HV o VH es la mejor polarización para la identificación de cultivos o para la estimación biofísica de cultivos
- la segunda mejor polarización normalmente es VV

Características de la Superficie y su Influencia sobre la Dispersión

Los SARs responden a dos características fundamentales de la superficie: (1) estructura o rugosidad, (2) contenido de agua

- **Rugosidad (o aspereza) :** es caracterizada por dos parámetros, la varianza de la media cuadrática (root mean square o RMS) y la longitud de correlación (l) de la superficie
- **Raíz de la media cuadrática (RMS o root mean square):** *la variación estadística del componente aleatorio de la altura de la superficie relativo a una superficie de referencia (en mm o cm)*
- **Longitud de correlación (l):** *función de autocorrelación que mide la independencia estadística de la altura de la superficie en dos puntos, separados por una distancia x'. La longitud de correlación es igual a la distancia de desplazamiento x' para la que p(x') es igual a 1/e. Si los dos puntos están separados por una distancia mayor a l, sus alturas se consideran estadísticamente independientes.*

Para los suelos esto significa

- rugosidad aleatoria causada por la labranza (y otras operaciones agrícolas) modificada por la erosión y fenómenos meteorológicos
- estructuras periódicas lineales son causadas al arar o sembrar

Efectos de la Rugosidad sobre la Retrodispersión

- la retrodispersión incrementa con la rugosidad del suelo
- los suelos más rugosos se ven más brillantes en las imágenes SAR

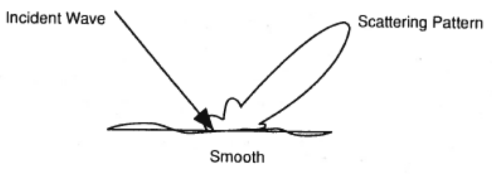
El impacto de la rugosidad en la retrodispersión depende de la frecuencia y el ángulo de incidencia. La rugosidad es un concepto relativo.

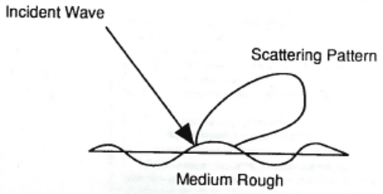
Según el criterio de Rayleigh, un suelo es liso si es que

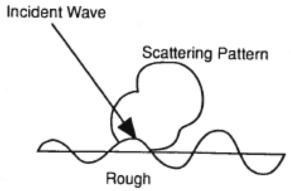
$$h < \frac{\lambda}{8\cos\theta}$$

donde h representa la variación en la altura de la superficie en cm, λ es la longitud de onda en cm y θ es el ángulo de incidencia en grados

La señal de radar anticipada







Todo es Relativo

Una rugosidad menor que “h” sería “vista” como lisa por el SAR

Ángulo de incidencia de 30º	
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
Ángulo de incidencia de 50º	
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

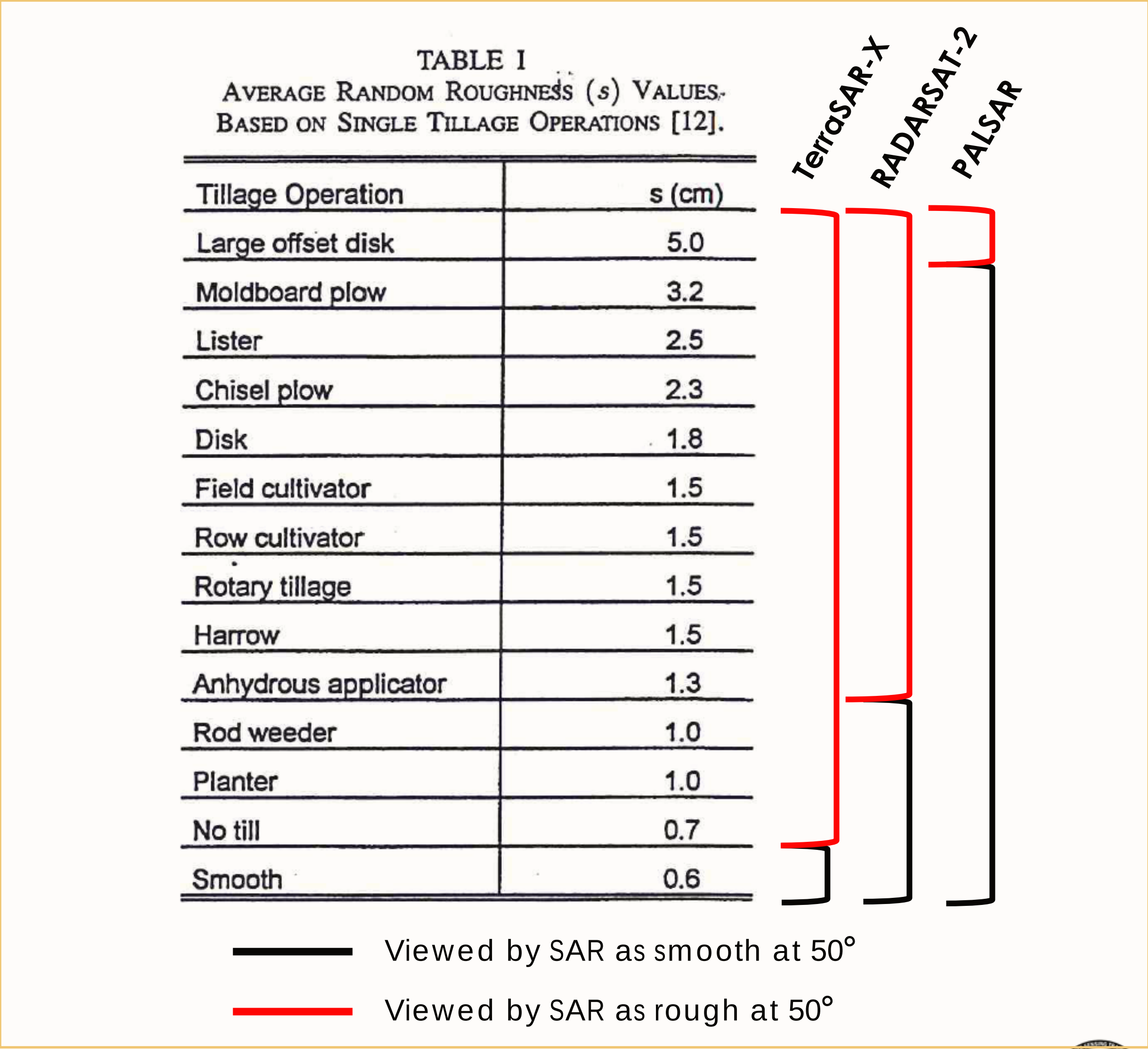


TABLE I
AVERAGE RANDOM ROUGHNESS (*s*) VALUES BASED ON SINGLE TILLAGE OPERATIONS [12].

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

Fuente: 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.

Efectos de la Vegetación

La escala es muy diferente a la que se usa para imágenes ópticas

La dispersión de microondas de mayor longitud es afectada por

a) estructuras de cierta escala (tamaño, forma y orientación de las hojas, tallos y frutos)

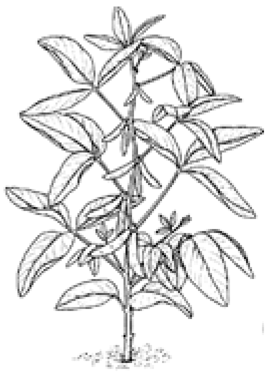
b) el volumen de agua en la vegetación (a nivel molecular)

¿Por qué SAR es tan sensitivo al tipo de cultivo y el desarrollo de estos?

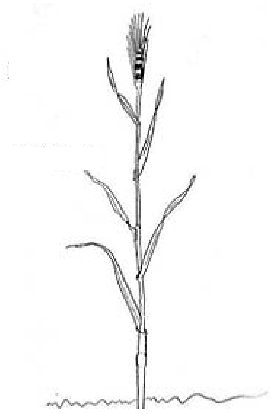
- La estructura de los cultivos varía de manera significativa de un tipo de cultivo a otro y durante sus diferentes etapas de crecimiento

La estructura de los cultivos de soya, trigo y maíz varía de manera significativa

La estructura también cambia según el crecimiento del cultivo



Soya



Trigo

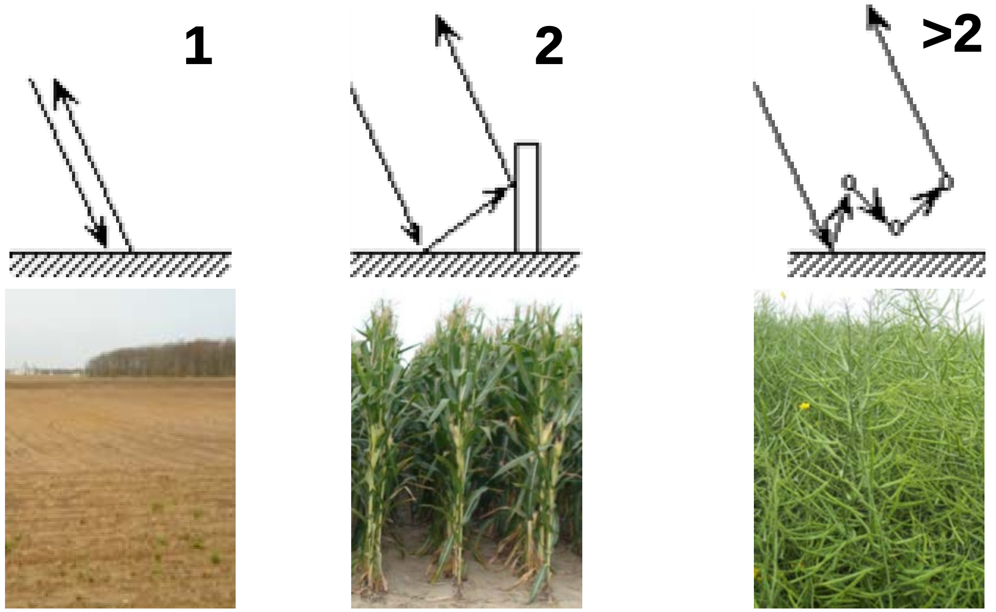
Maíz

Tipos de Dispersión

- cuando la señal de microondas llega a una superficie, la onda experimentará uno, dos, o más eventos de dispersión
- un cambio de polarización (ej. de H a V) se conoce como **repolarización**.
- el número de eventos determina el tipo de dispersión, la intensidad de esta y los cambios en fase
- los eventos de dispersión dependen de la estructura y geometría de la superficie
- normalmente hay un tipo de dispersión dominante
- en superficies distribuidos a menudo ocurre dispersión secundaria o terciaria y una mezcla de tipos de dispersión a menudo caracteriza estas superficies

tipo de dispersión + mezcla de eventos de dispersión + características de la fase + intensidad
= indicaciones sobre tipos y condiciones de cultivos

Número de Eventos de Dispersión



suelo expuesto

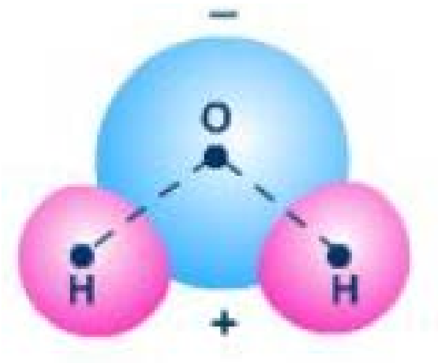
maizal con estructura vertical definida y un espaciado amplio entre las filas

cultivo de canola en etapa de formación

¿Cuál es la Influencia del Agua?

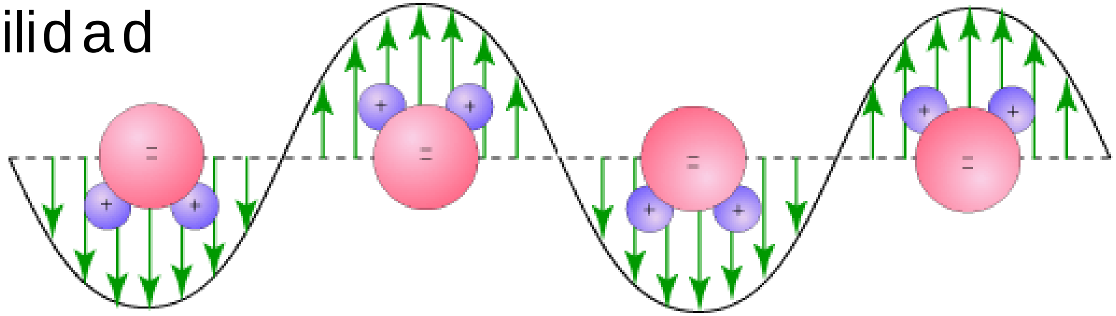
Se sabe que SAR es sensitivo a la humedad, pero **¿por qué?**

- el agua (H_2O) es un dipolo: el lado de la molécula donde está el oxígeno lleva una carga neta negativa mientras que el lado con los dos átomos de hidrógeno tiene una carga neta positiva
- por lo tanto, cuando se aplica un campo eléctrico (por ejemplo, microondas), la molécula rotará y se alineará con este campo



¿ Cuál es la Influencia del Agua?

- el agua (H_2O) es un dipolo: el lado de la molécula donde está el oxígeno lleva una carga neta negativa mientras que el lado con los dos átomos de hidrógeno tiene una carga neta positiva
- por lo tanto, cuando se aplica un campo eléctrico (por ejemplo, microondas), la molécula rotará y se alineará con este campo aplicado
- constante dieléctrica: una medida de la facilidad con la que las moléculas dipolares rotan en respuesta a la aplicación de un campo



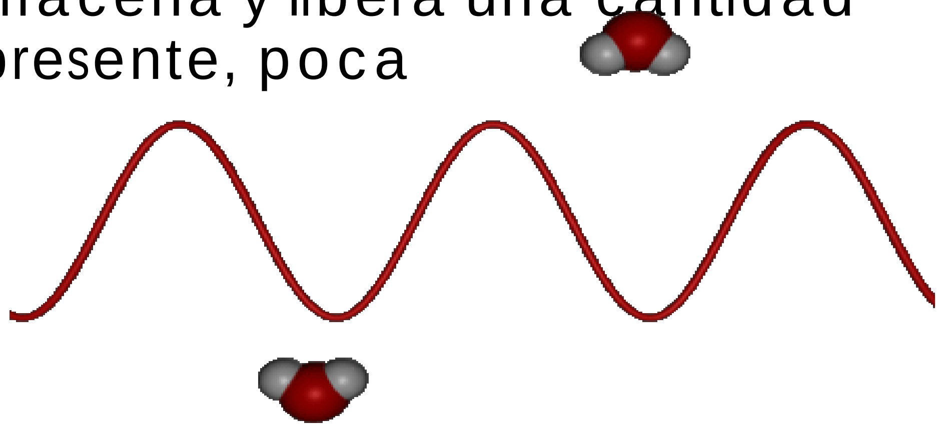
- constante dieléctrica (ε): un valor complejo que caracteriza tanto la permitividad (ε') (real) como la conductividad (ε'') (imaginaria) de un material

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

- la constante dieléctrica real varía entre ~3 (muy seco) y 80 (agua)

¿Qué Significa Esto para SAR?

- una microonda continuará a propagarse hasta encontrar una discontinuidad dieléctrica, como sucede cuando hay agua en la superficie
- cuando se aplica un campo eléctrico, las moléculas de agua que están libres (**no fuertemente entrelazadas entre sí**) rotan fácilmente para alinearse con el campo eléctrico (positivo a negativo)
- hay poca resistencia y poca de la energía almacenada en la rotación se pierde cuando la onda pasa y la molécula se relaja. La mayor parte de la energía almacenada se libera.
- si hay muchas moléculas de agua presentes, se almacena y libera una cantidad significativa de energía. Cuando hay poca agua presente, poca energía es almacenada.
- cuando esta energía almacenada se libera (y dependiendo de la estructura) esta energía será dispersada de vuelta hacia la antena del radar



Fuente de la Imagen: Anton Paar

¿Qué Significa Esto para SAR?

- una fuerte relación positiva entre la constante dieléctrica real y la retrodispersión
- una fuerte relación positiva entre la constante dieléctrica real y la humedad del suelo
- más agua en la superficie = mayor retrodispersión = retornos más brillantes
- aplica a CUALQUIER superficie (e.g. suelo, vegetación)
- La profundidad de penetración (δ_p) en el suelo y/o cultivos es determinada por la dieléctrica (ε), longitud de onda (λ) y ángulo de incidencia
- la penetración aumenta con la longitud de onda y es mayor cuando la superficie (suelo o cultivo) está más seca

Vista Compuesta Multi-fecha de RADARSAT-1, Saskatchewan (Canadá)

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

Una Complicación: El Medio Ambiente

Antes de utilizar datos SAR siempre, siempre, siempre revise las condiciones meteorológicas en el momento de adquisición de la imagen

Regla No. 1: Nunca utilice SAR si estuvo lloviendo en el momento de adquisición

- **¿Por qué?** Aunque SAR supuestamente funciona bajo cualquier condición meteorológica, no incluye la recolección de datos durante precipitaciones porque el agua en la atmósfera causa que la señal de SAR se disperse. En algunas regiones del mundo este riesgo es diurno.

Regla No. 2: Nunca utilice SAR para estimar la humedad del suelo si el suelo está congelado

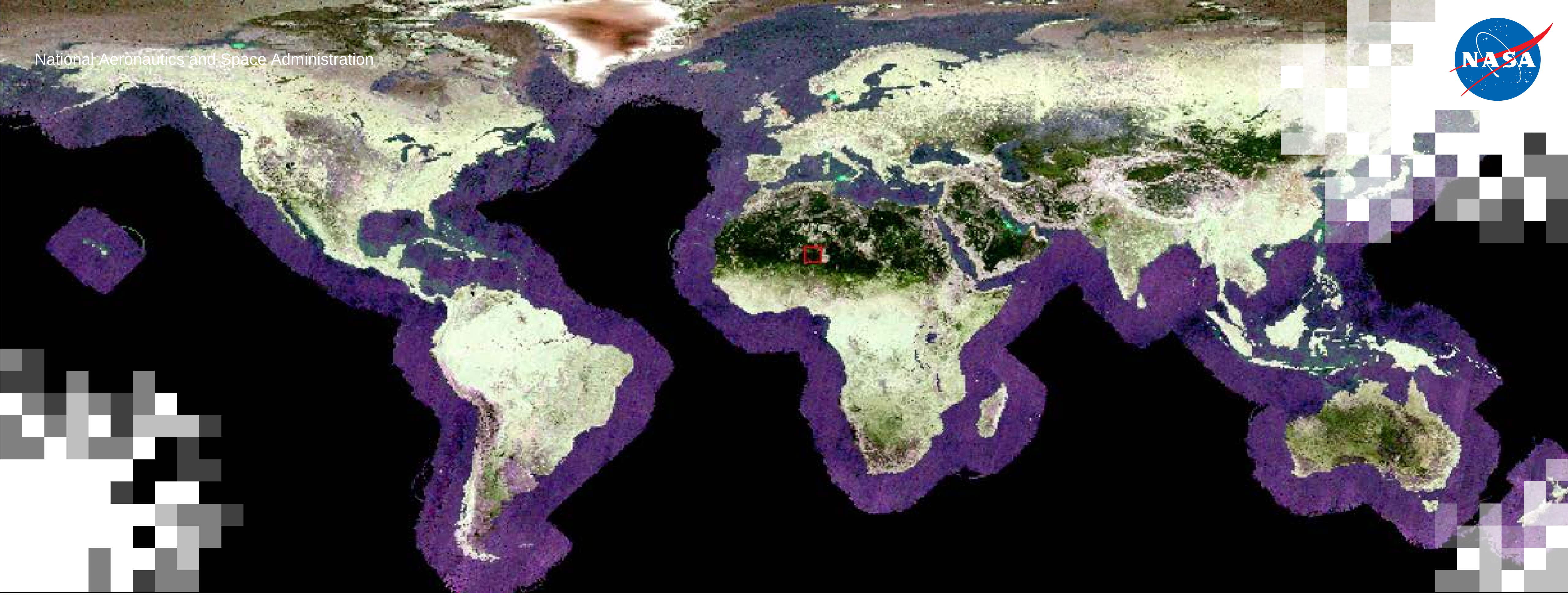
- **¿Por qué?** La constante dieléctrica se reduce a casi cero cuando el agua se congela. Por lo tanto, aunque haya agua en el suelo, el SAR verá el suelo como seco. SAR puede detectar eventos de congelamiento/descongelamiento. El congelamiento a menudo ocurre de noche.

Una Complicación: El Medio Ambiente

Regla No. 3: Considere la posibilidad de la presencia de rocío en las adquisiciones matutinas

- **¿Por qué?** La presencia de agua sobre las hojas incrementará la retrodispersión (un gran problema para la modelación biofísica). Si hay una cantidad significativa de agua sobre la vegetación (inmediatamente después de una lluvia), el contraste entre diferentes superficies puede ser reducido. El rocío es más prominente en regiones templadas en las primeras horas de la mañana.
- Elija órbitas (ascendente – vespertino; descendente – matutino) con precaución
- Siempre averigüe las condiciones meteorológicas

National Aeronautics and Space Administration



Aplicaciones de SAR para el Monitoreo Agrícola

Heather McNairn, Xianfeng Jiao, Sarah Banks y Amir Behnamian

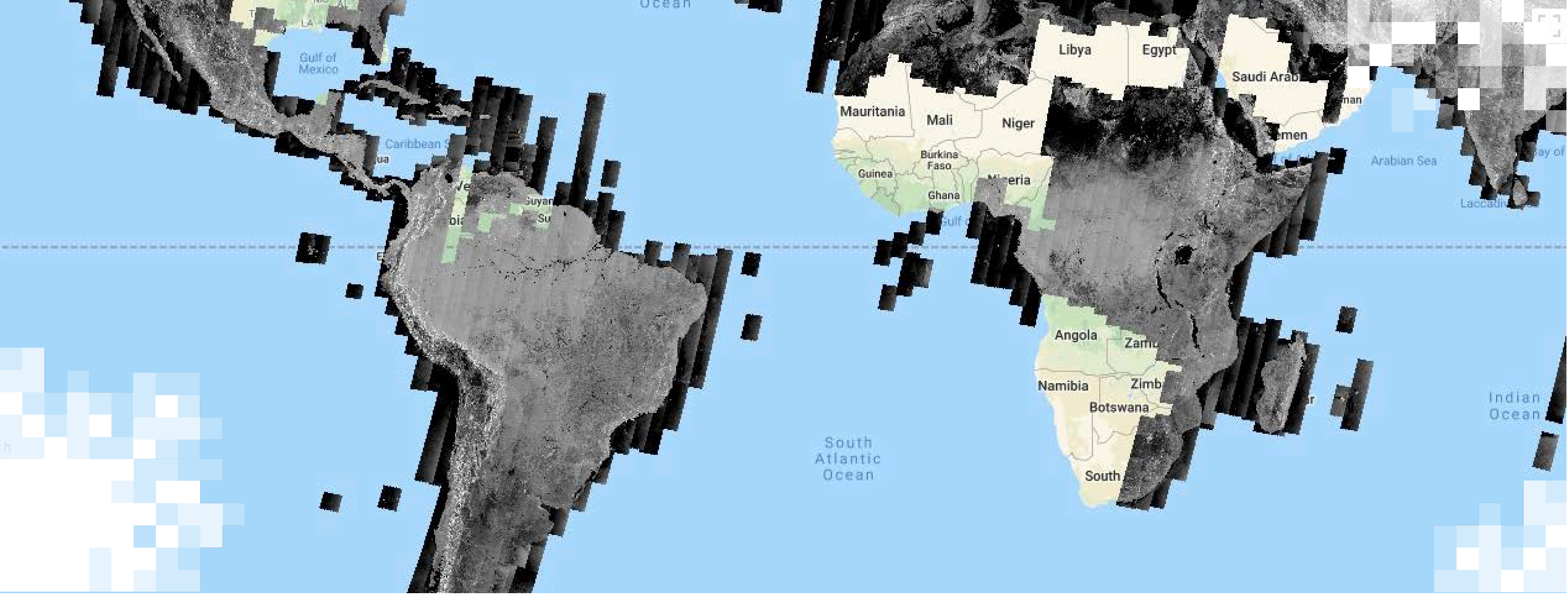
4 de septiembre de 2019

Objetivos de Aprendizaje

- Al finalizar esta presentación, usted podrá entender:
  – cómo estimar la humedad del suelo a partir de datos RADARSAT-2
  – cómo procesar datos multifrecuencia para la clasificación de cultivos

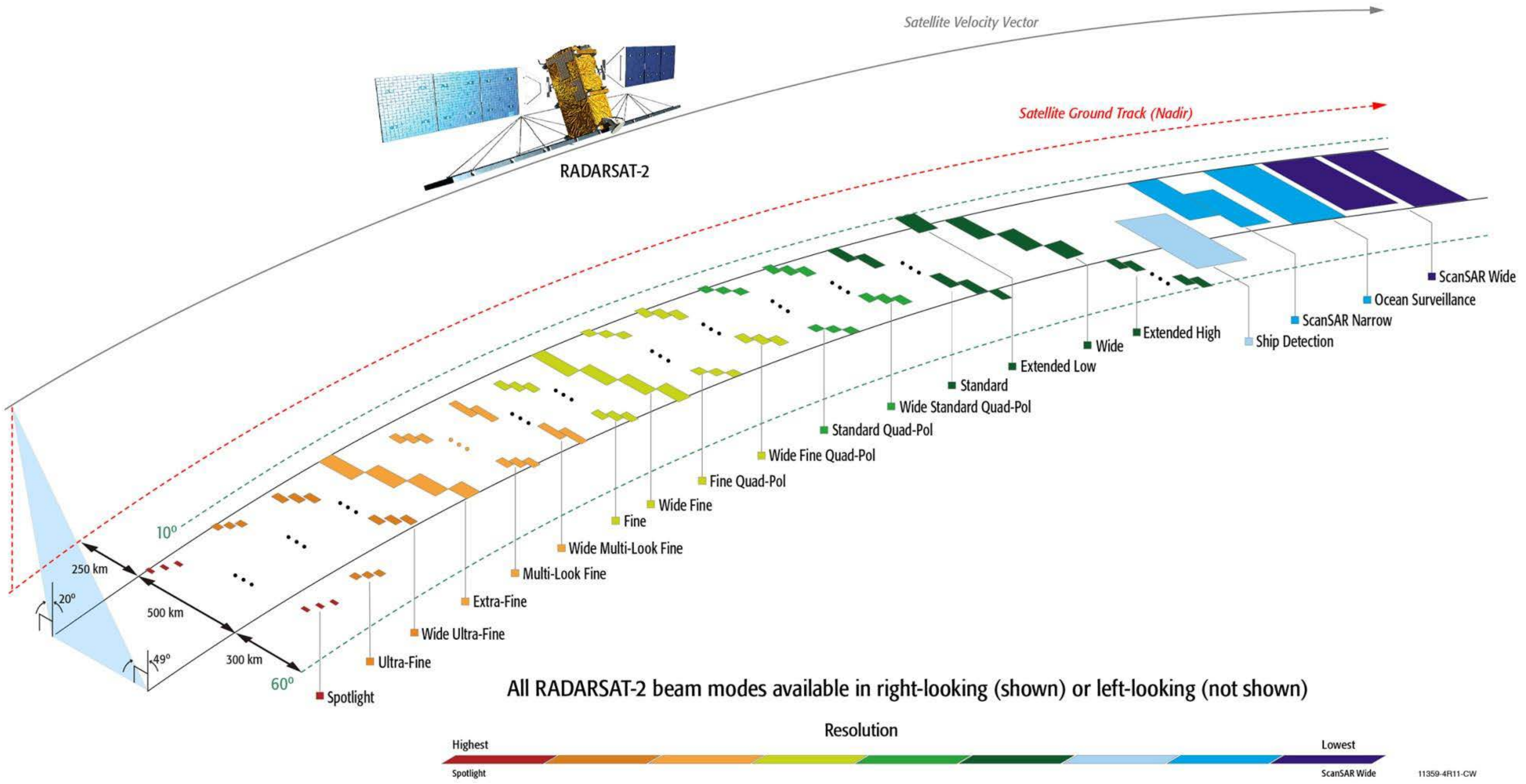
SNAP: Sentinel's Application Platform

- ESA SNAP es el Toolbox gratuito y de fuente abierta para procesar y analizar datos de imágenes obtenidos por satélites de observación de la tierra de la ESA y de terceros
- Pueda descargar los últimos instaladores de SNAP en esta página:
  – http://step.esa.int/main/download/snap-download/



Estimando la Humedad del Suelo con Datos RADARSAT-2

RADARSAT-2



RADARSAT-2 SAR- Modos de Haz – Tiempo de Revisita: 24 días

Fuente de la Imagen: [MDA RADARSAT-2 Product Description]

Preprocesamiento de Datos RADARSAT-2 con SNAP

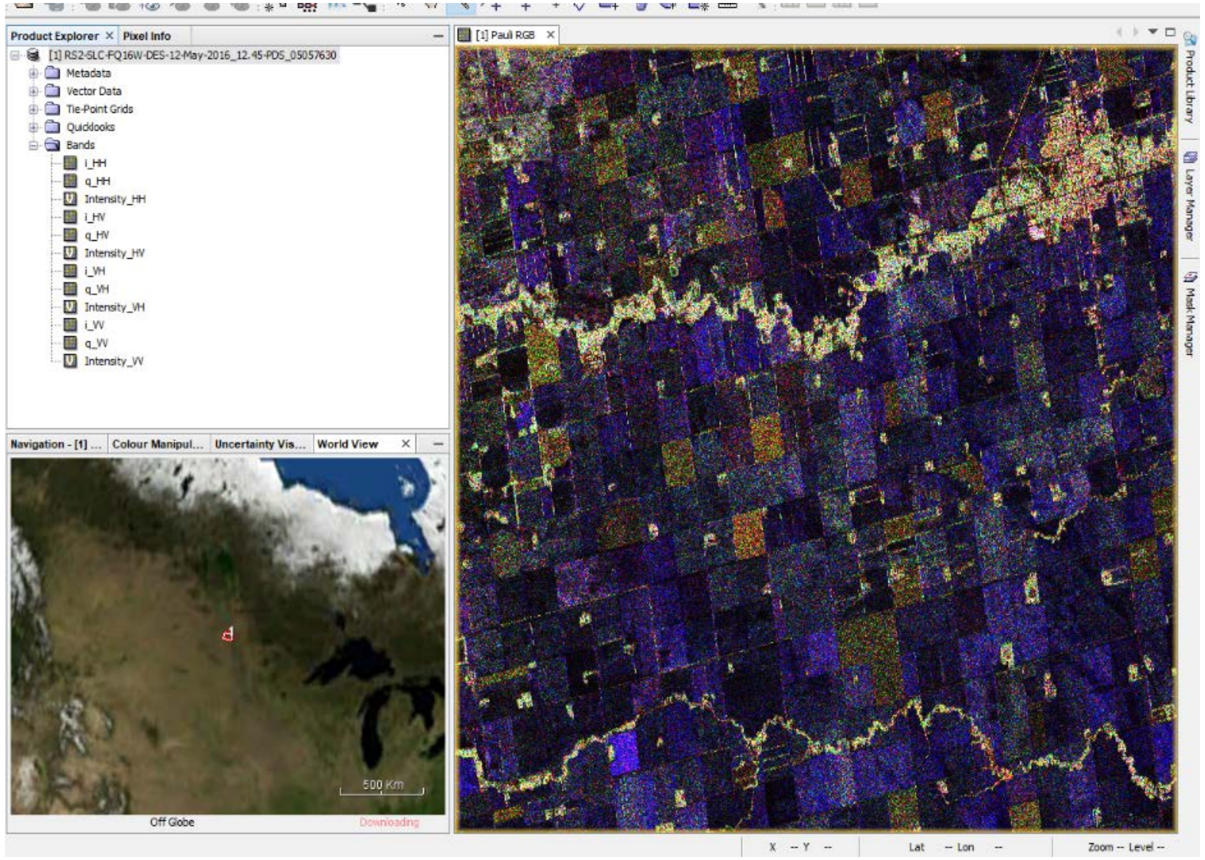
Extraer Retrodispersión

Leer → Calibración → Filtro para Speckle → Corrección Topográfica → Escribir

Preprocesamiento de Datos RADARSAT-2 con SNAP

Leer Imagen

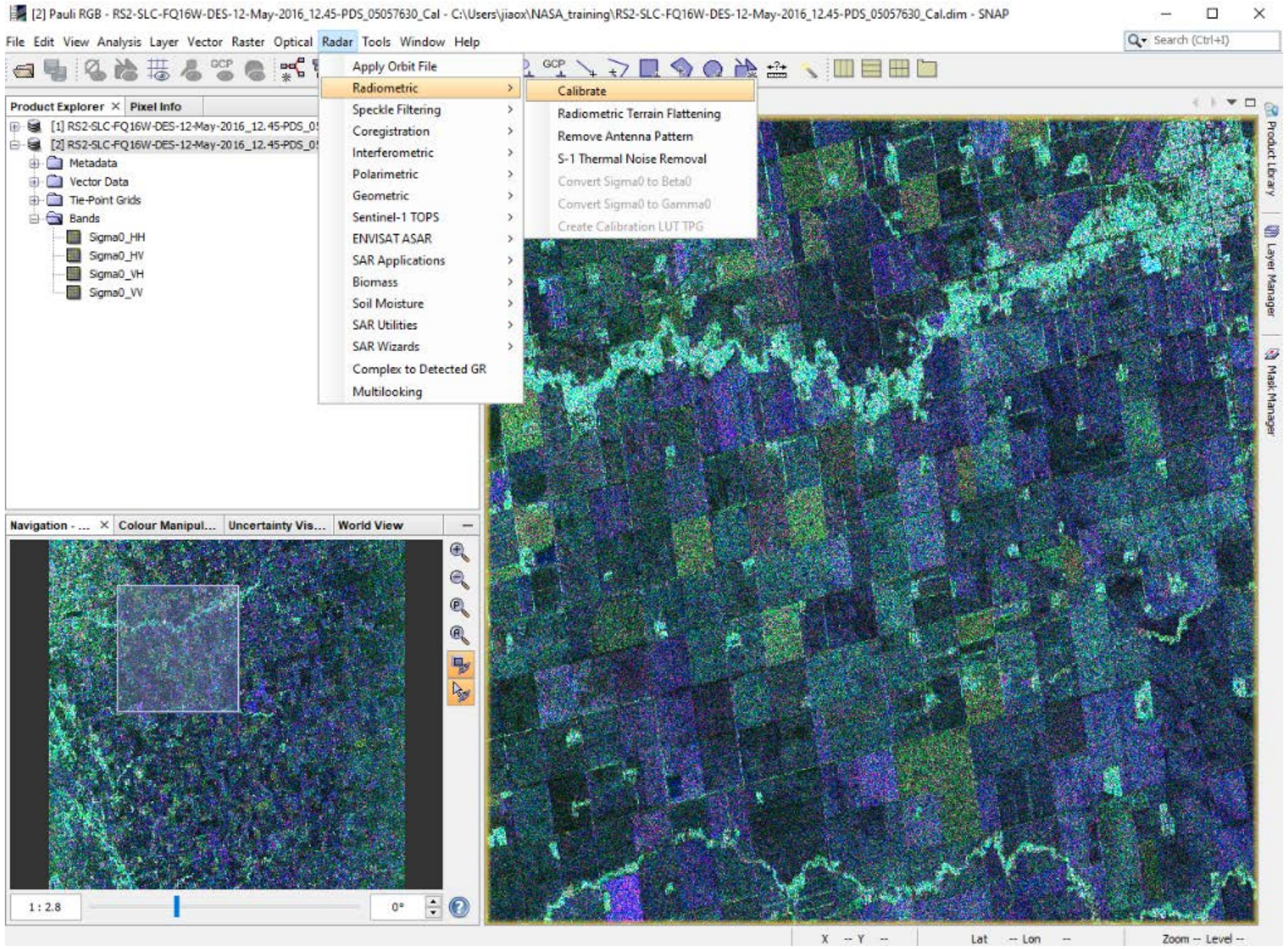
- Producto RADARSAT-2 Wide Fine Quad-Pol single look complex (SLC):
  - Resolución Nominal: 5.2m (Rango) * 7.6 m (Azimut)
  - Tamaño Nominal de Escena : 50 Km (Rango) * 25 Km (Azimut)
  - Polarización cuádruple (HH, HV,VH y VV) + fase
  - Producto complejo de mirada singular de distancia oblicua, contiene información sobre la amplitud y la fase



Datos RADARSAT-2 Wide Fine Quad-Pol FQ16W SLC descendente adquiridos el 12 de mayo de 2016 sobre Carman, Manitoba, Canadá

Preprocesamiento de Datos RADARSAT-2 con SNAP

Calibración: Convertir Valores de Pixeles en Retrodispersión de la Señal de Radar



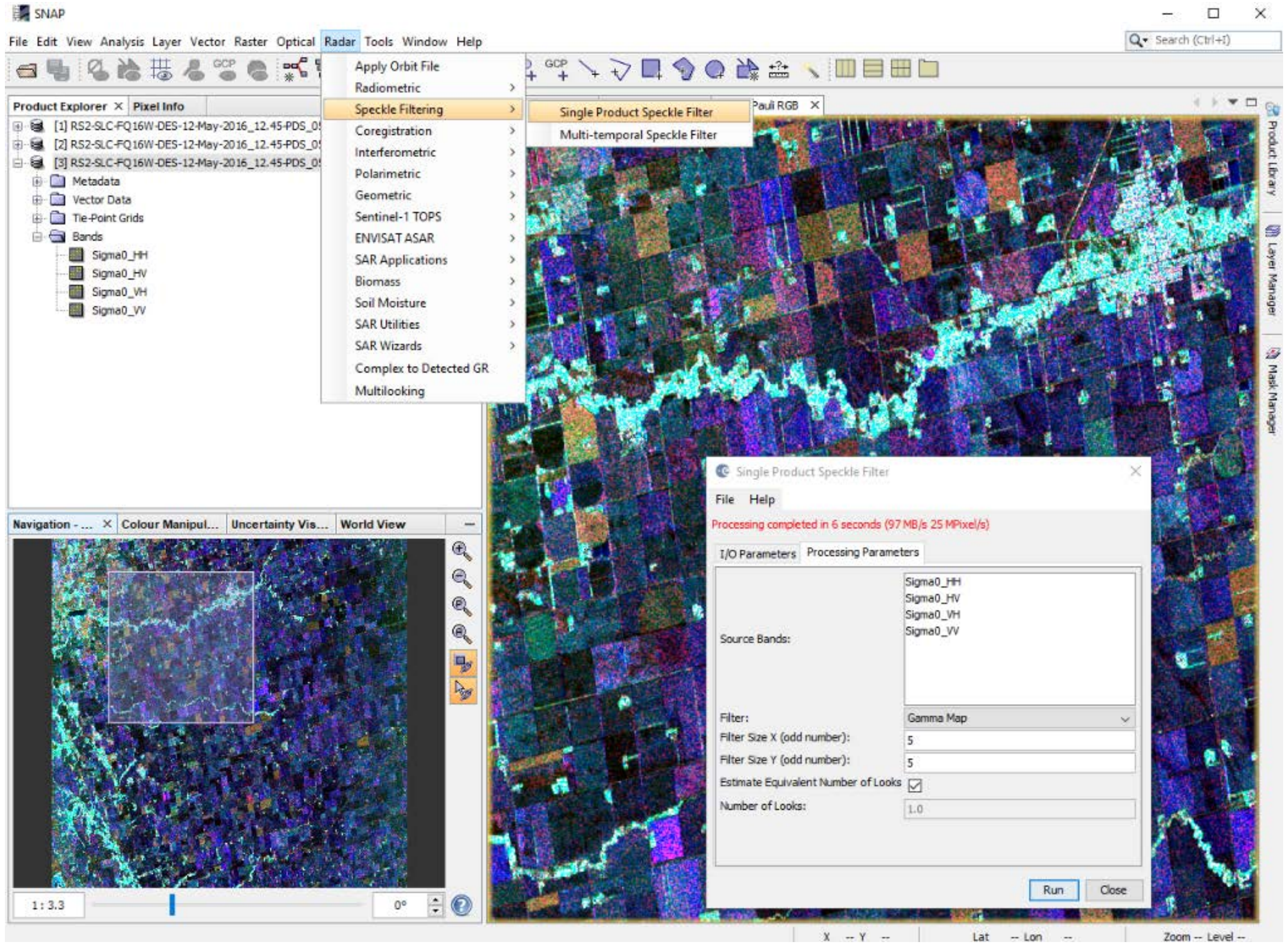
Filtrado de Speckle

- El filtrado de speckle no es una ciencia exacta → depende de la imagen y del objetivo
- Un filtro para speckle ideal :
  - reduce el speckle
  - preserva la nitidez en los bordes
  - preserva el contraste entre líneas y puntos en el objetivo
  - retiene los valores medios en regiones homogéneas
  - retiene información sobre la textura

- Filtrado de Speckle
  - Filtrado Adaptivo
    - Lee
    - Frost
    - Gamma-Map
    - Multi-Temporal
    - Intensity Drive Adaptive Neighborhood (IDAN)
    - Lee Refinado
    - Lee Sigma
  - Filtrado No Adaptivo
    - Media
    - Mediana
    - Moda

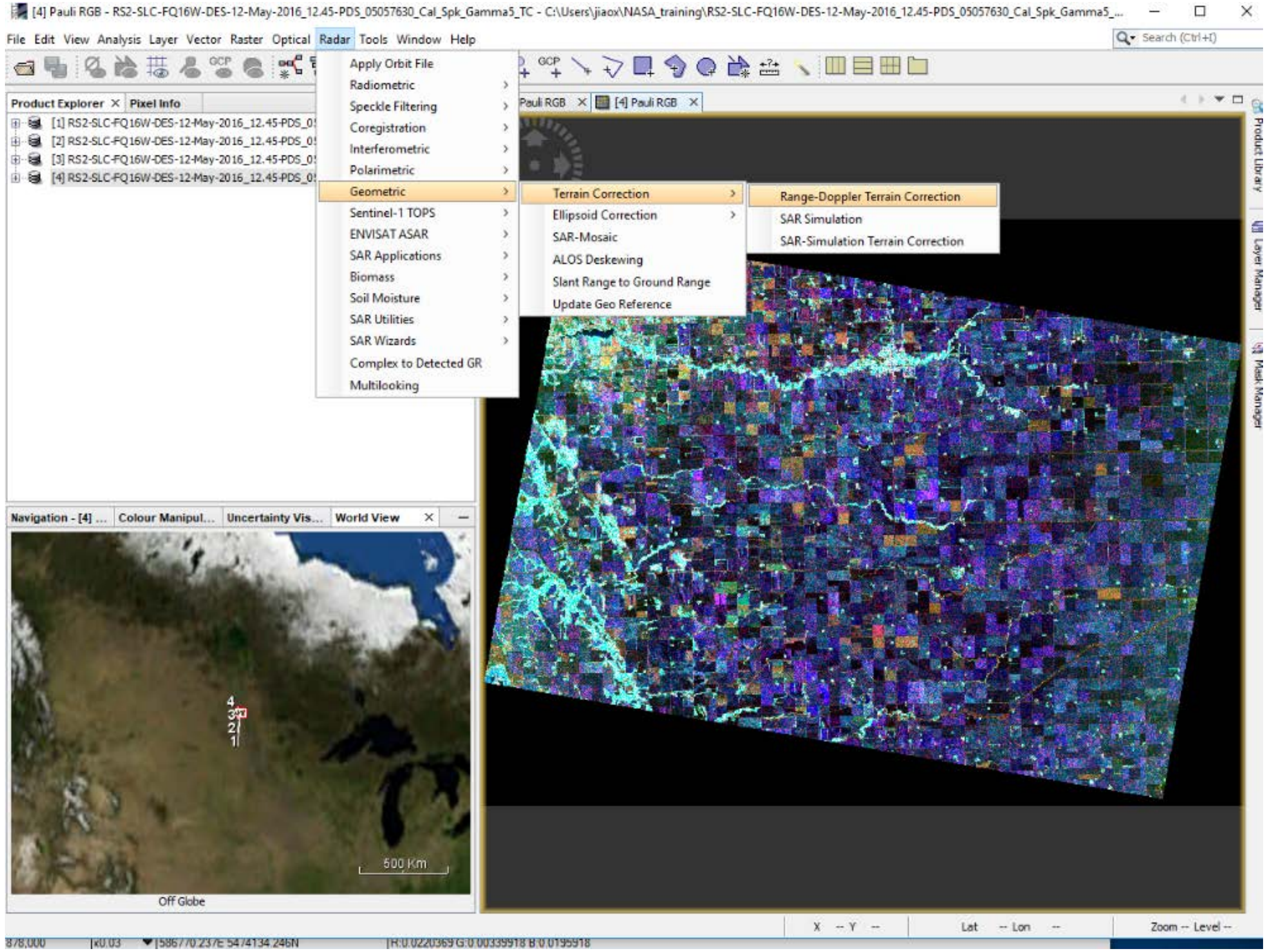
Preprocesamiento de Datos RADARSAT-2 con SNAP

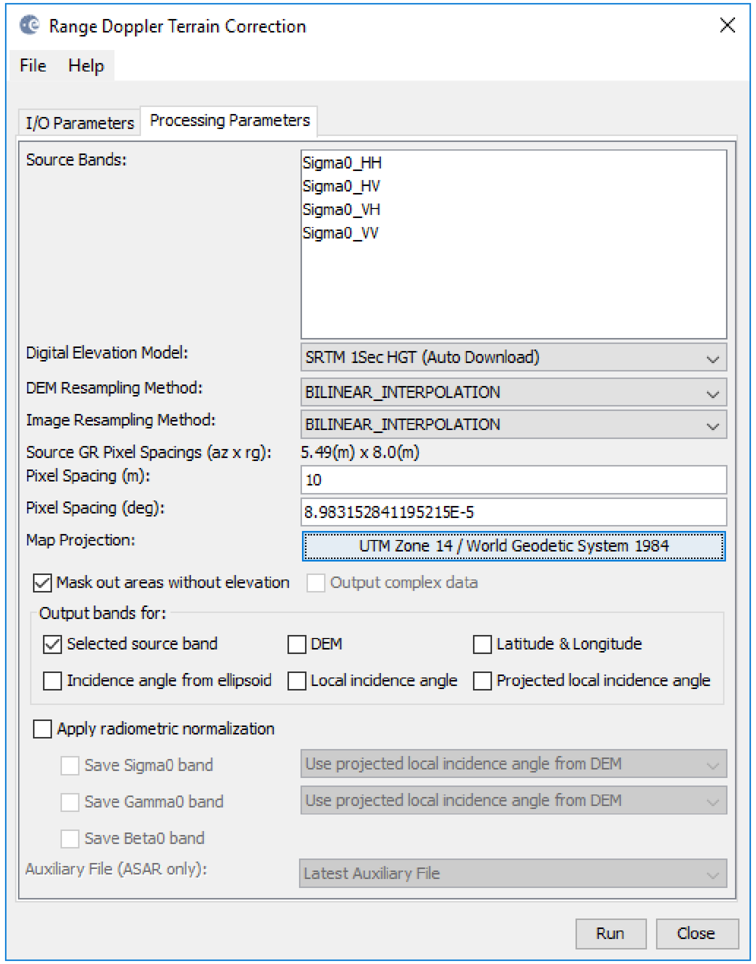
Filtro para Speckle– Gamma Map 5 x 5



Preprocesamiento de Datos RADARSAT-2 con SNAP

Corrección Topográfica





Preprocesamiento de Datos RADARSAT-2 con SNAP

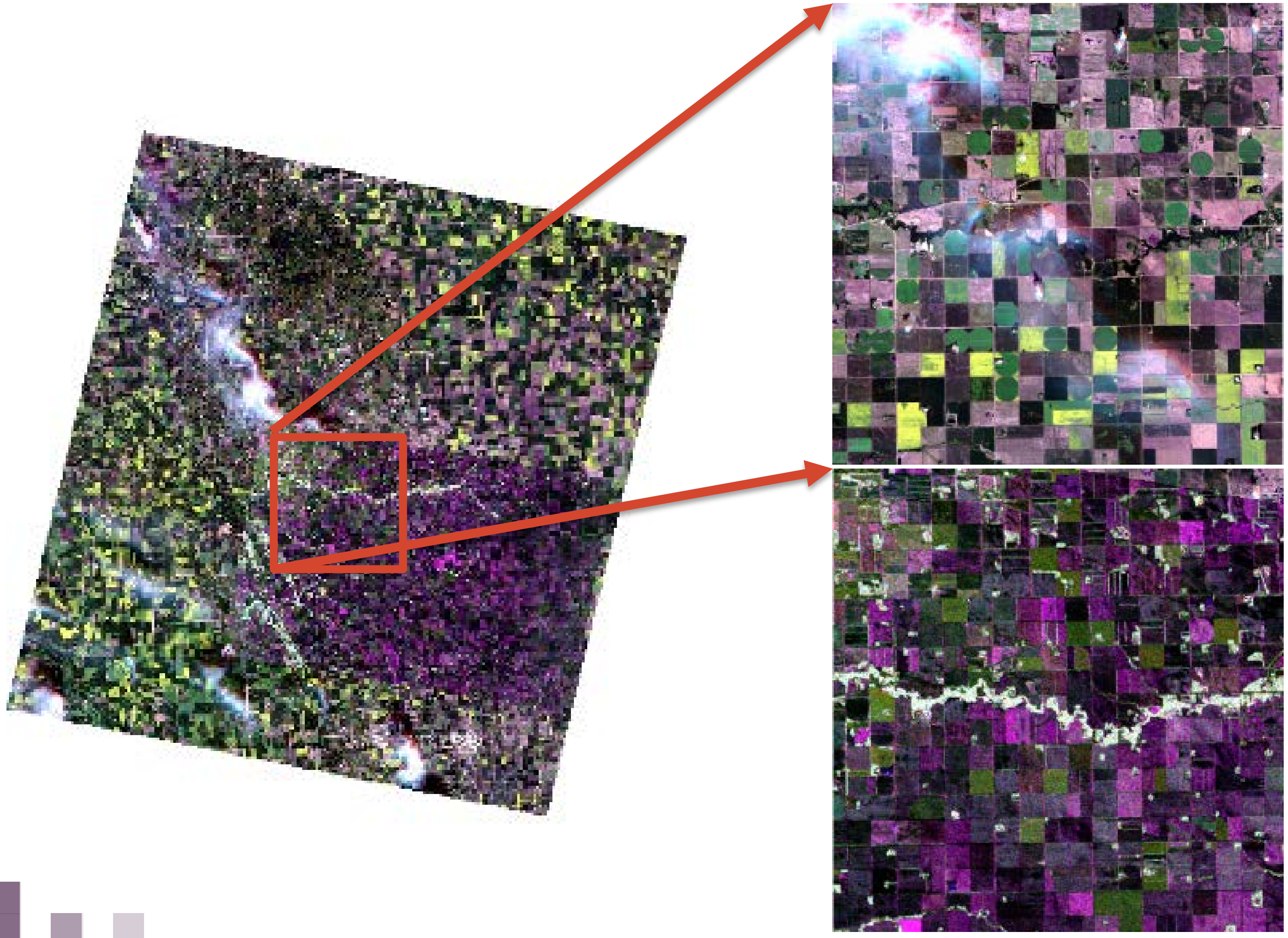


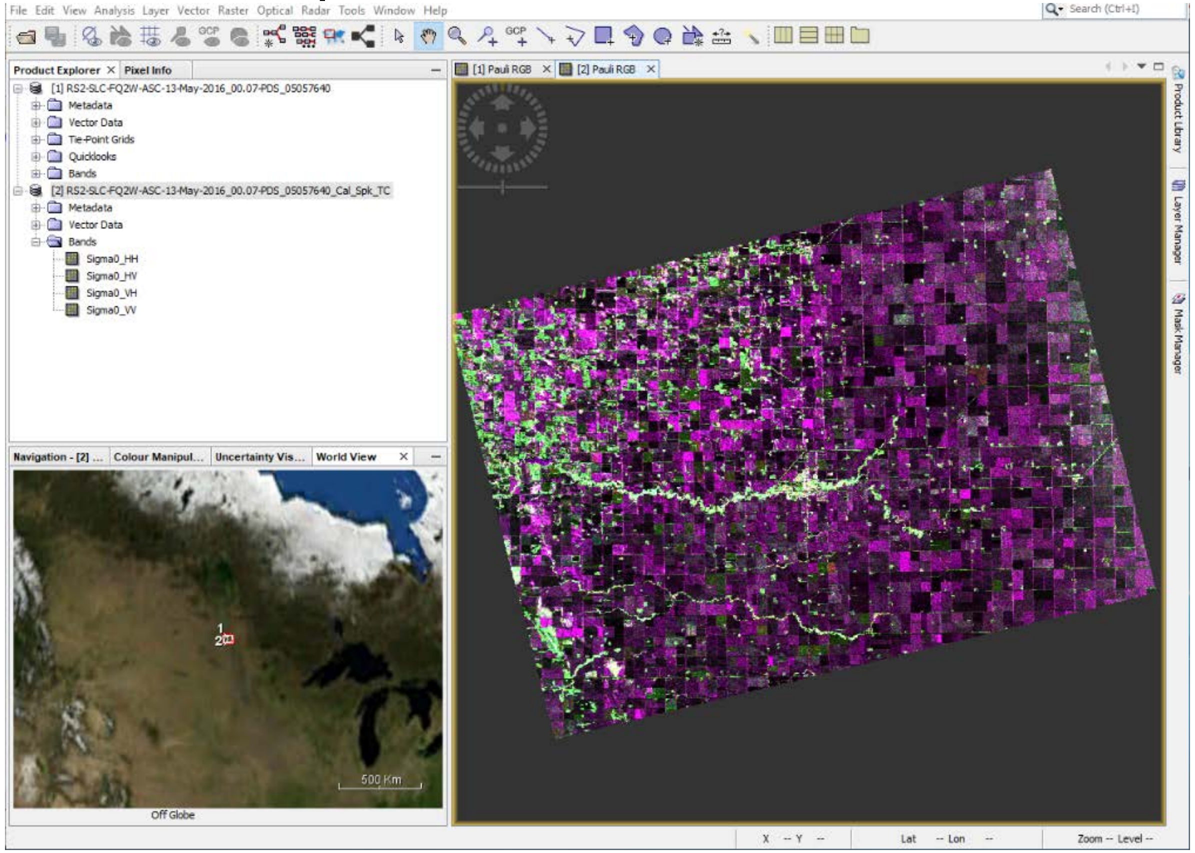
Imagen RapidEye de color natural adquirida el 4 de julio de 2016

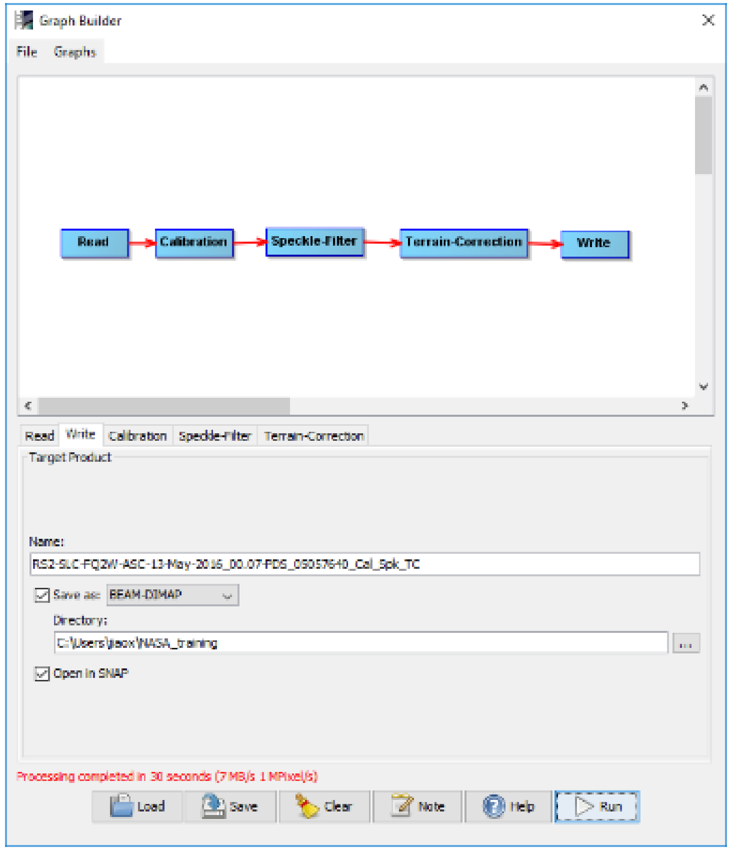
Imagen RGB compuesta de RADARSAT-2 adquirida el 12 de mayo de 2016

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

Preprocesamiento de Datos RADARSAT-2 con SNAP

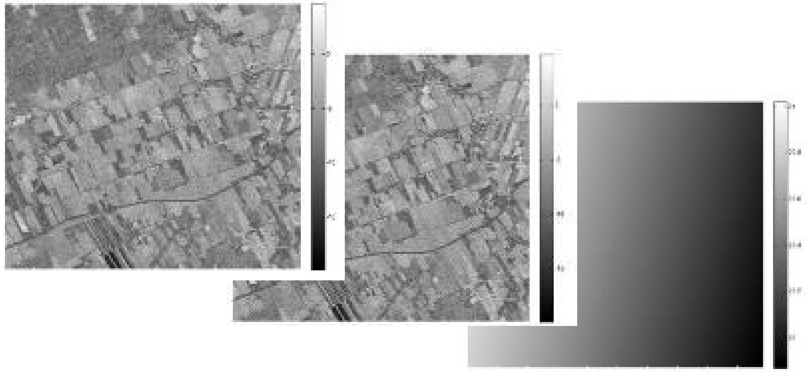
- Graph builder: se utiliza para el procesamiento por lotes y para cadenas de procesamiento personalizadas





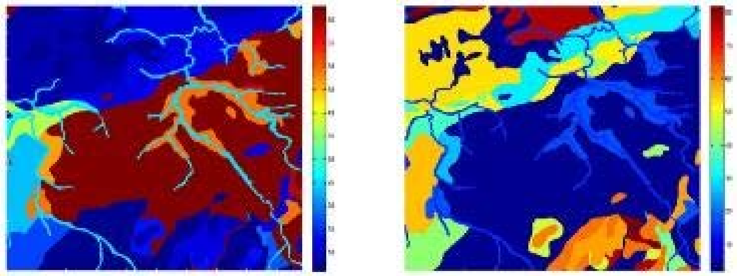
Técnica de Modelación Física para Estimar la Humedad del Suelo

Datos RADARSAT-2



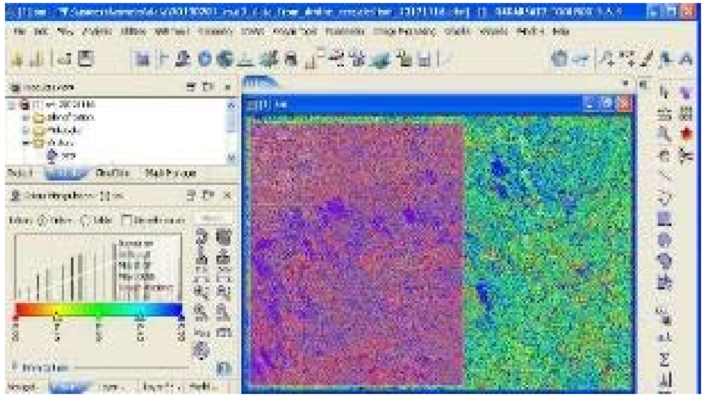
Datos Radar
(Retrodispersión HH y VV)
(Ángulos de Radar)

Datos de Suelos
(Fracciones de Arcilla y Arena)



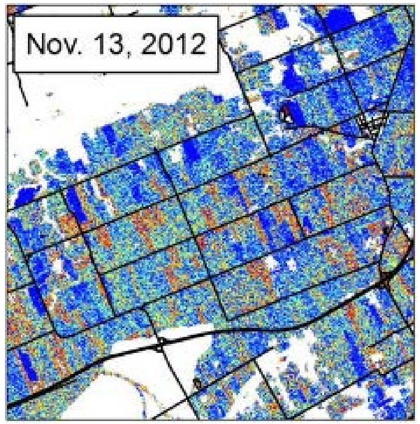
Procesamiento de Datos

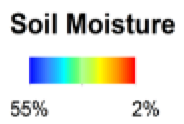
SNAP Soil Moisture Tool Box

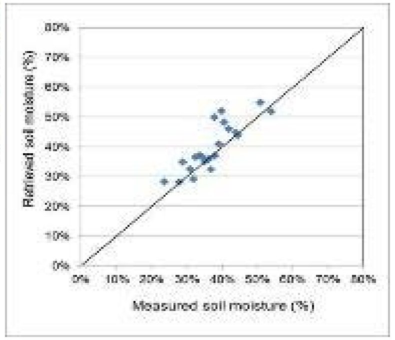


- Utiliza un modelo de ecuación integral (Integral Equation Model o IEM)
- No necesita ninguna información a priori
- Se recupera la humedad utilizando solo datos SAR (retrodispersión y ángulo de incidencia)

Salidas y Validación







Fuente de la Imagen:
Merzouki y McNairn, 2015

Esquemas de Inversión para Modelos de Retrodispersión

Dos Imágenes
Adquiridas a diferentes ángulos de incidencia (θ1 y θ2) con polarizaciones HH **y** VV

→ IEM →

Inversión híbrida
$HH_{\theta1} – VV_{\theta1} – HH_{\theta2}$
$HH_{\theta1} – VV_{\theta1} – VV_{\theta2}$
$HH_{\theta2} – VV_{\theta2} – HH_{\theta1}$
$HH_{\theta2} – VV_{\theta2} – VV_{\theta1}$
$HH_{\theta1} – VV_{\theta1} – HH_{\theta2} – VV_{\theta2}$

Una Imagen
Adquirida al ángulo de incidencia θ1 con polarizaciones HH **y** VV
($HH_{\theta1}$, $VV_{\theta1}$)

→ IEM Calibrado →

Inversión de Polarización Múltiple
$HH_{\theta1}$ y $VV_{\theta1}$

Dos Imágenes
Adquiridas a diferentes ángulos de incidencia (θ1 y θ2) con polarizaciones HH **o** VV

→ IEM Calibrado →

Inversión multiángulo
$HH_{\theta1} – HH_{\theta2}$
$VV_{\theta1} – VV_{\theta2}$

Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

- Mapa de la humedad del suelo derivado de una adquisición matutina de RADARSAT-2 y una adquisición vespertina de RADARSAT-2 12 de unas horas después.

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

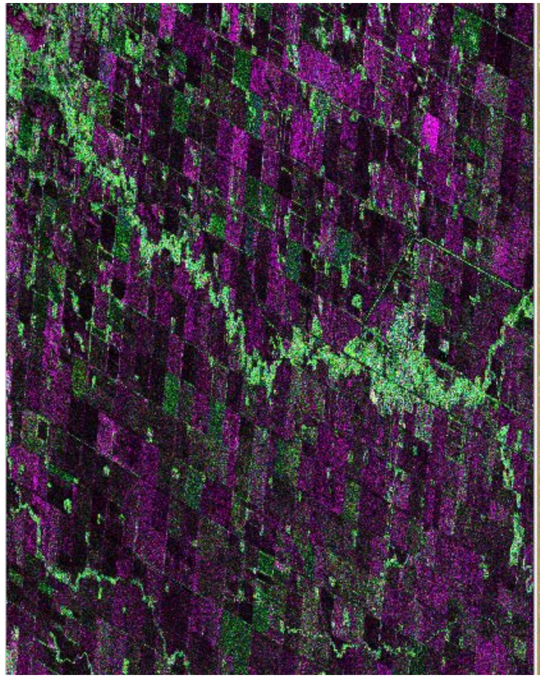


Imagen adquirida el 12 de mayo de 2016, FQ16W Paso descendente

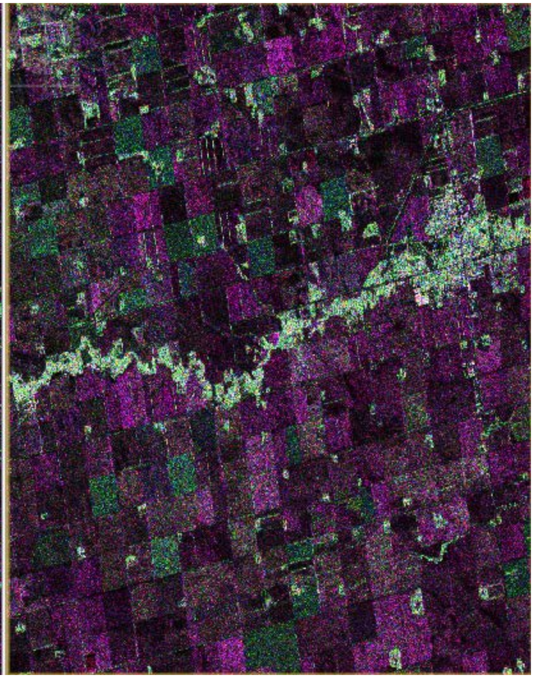
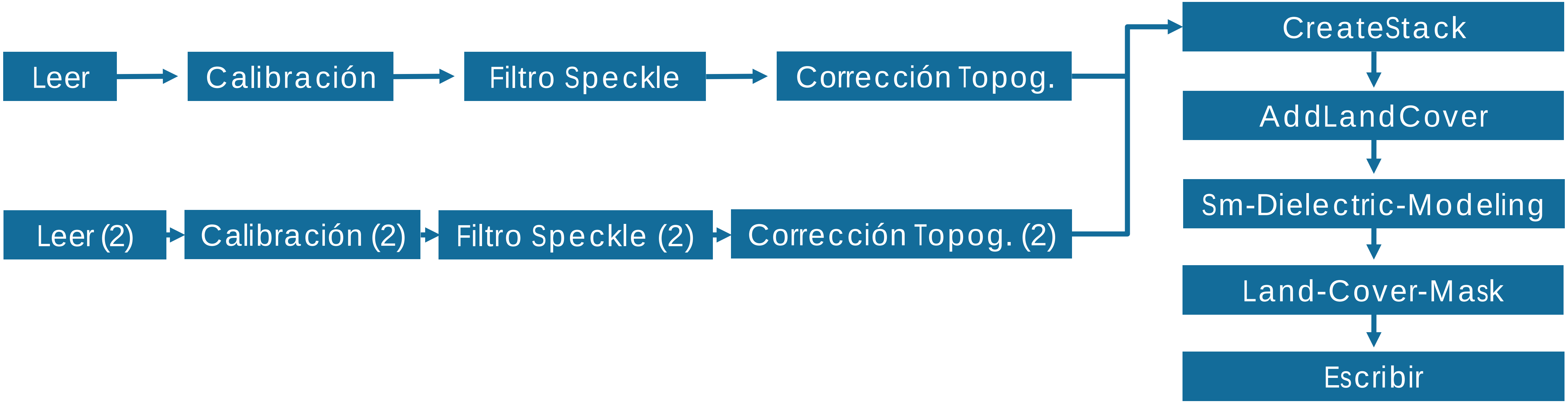


Imagen adquirida el 13 de mayo de 2016, FQ2W Paso ascendente

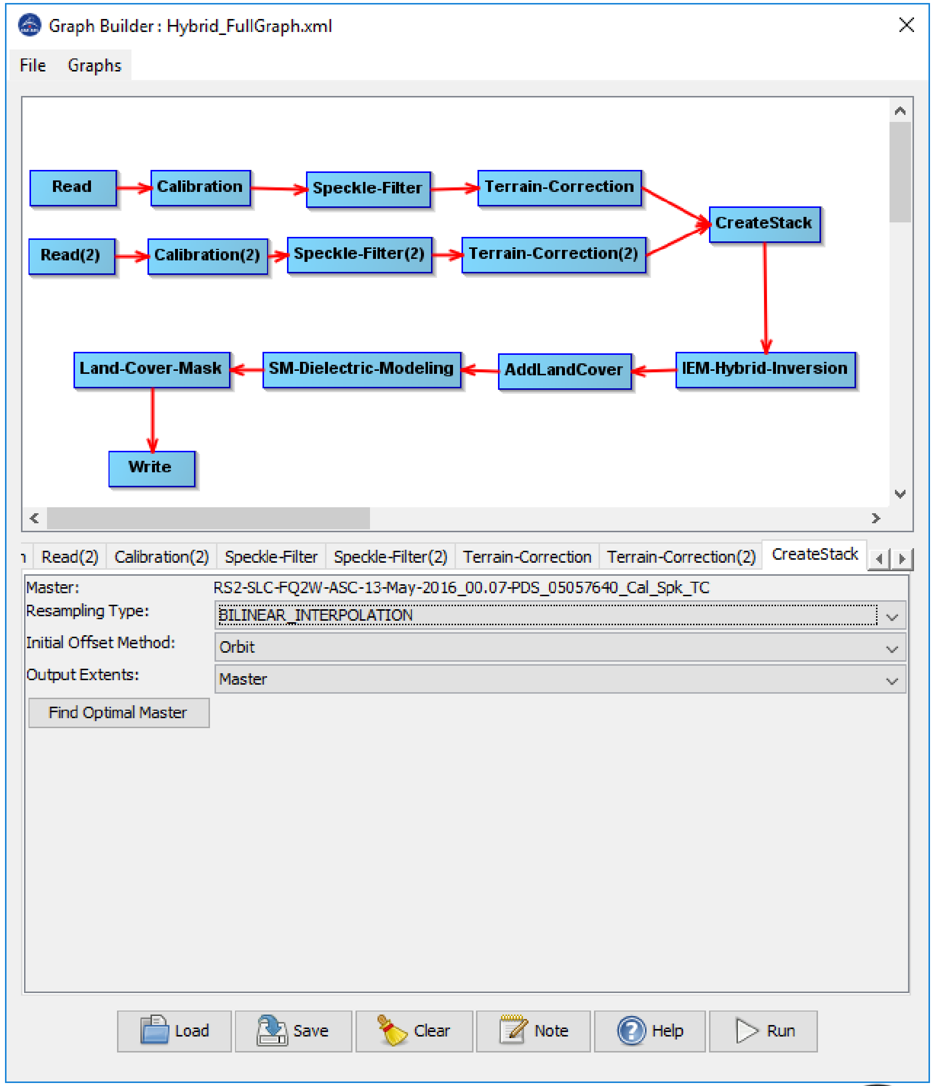
Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

- mediante una técnica de inversión híbrida, el método de inversión se realiza en la zona geográfica de solapamiento



Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

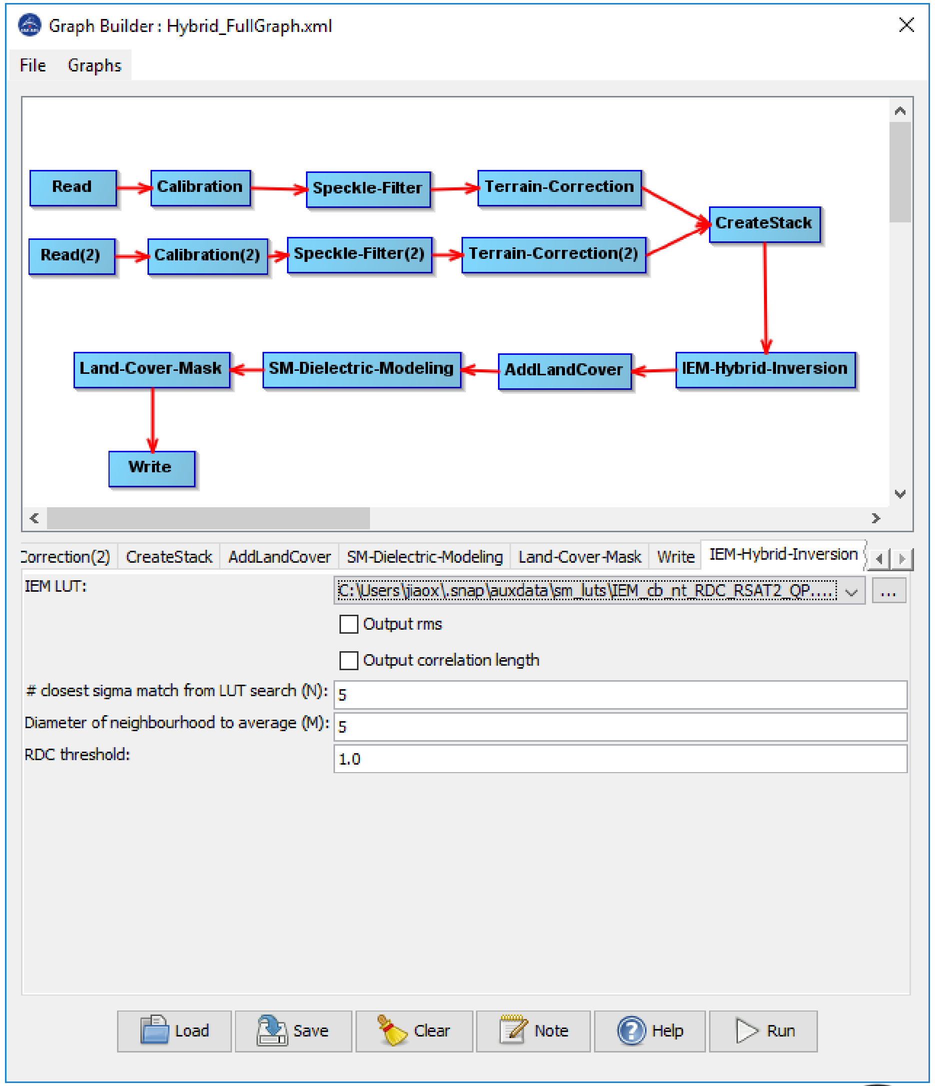
- Alineación espacial de imágenes adquiridas en la mañana y en la tarde
- Create Stack (Crear Pila):
  1. Tipo de Remuestreo: Interpolación Bilineal
  2. Haga clic en **Find Optimal Master**



Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

IEM Hybrid Inversion (Inversión Híbrida con IEM)

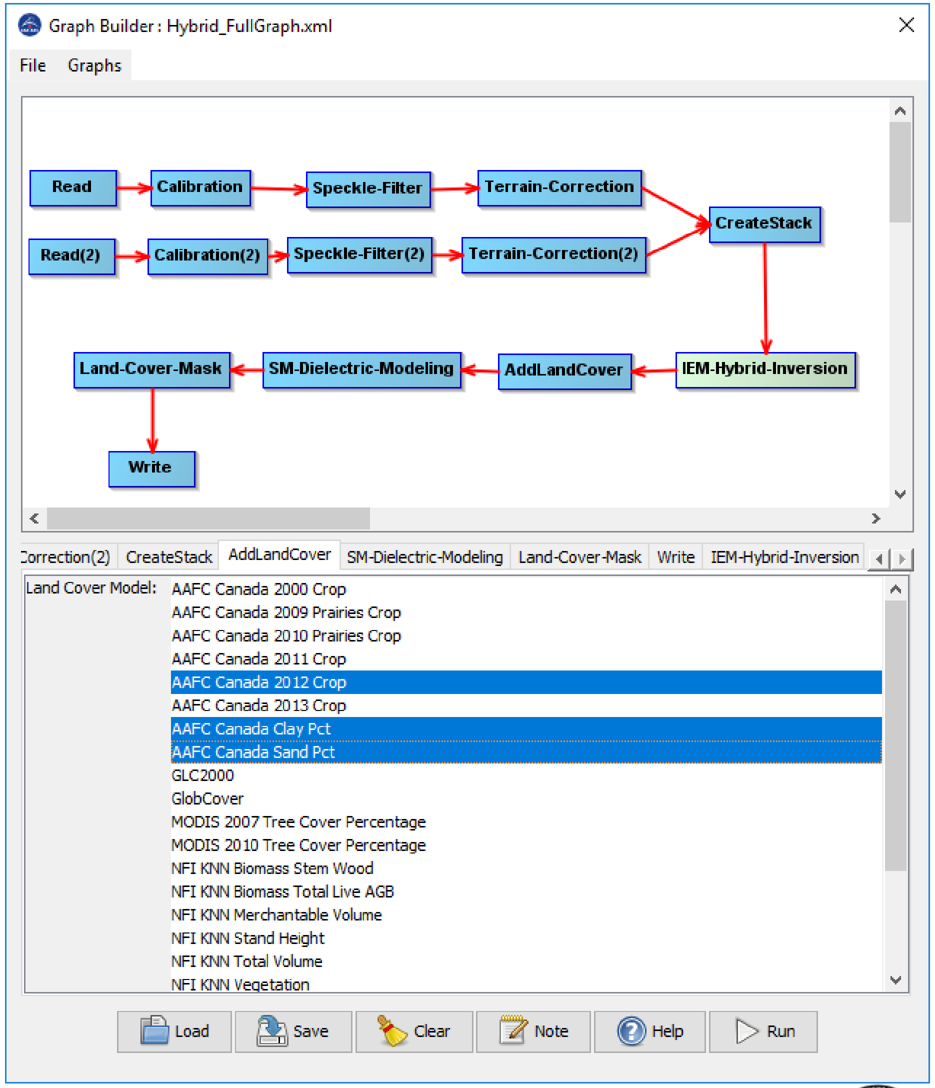
1. Seleccione **calibrated IEM LUT**
2. Use la configuración preprogramada para los otros parámetros



Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

AddLandCover (Agregar Cobertura Terrestre)

1. Seleccione el archivo de la cobertura terrestre
2. Seleccione los mapas con las fracciones de arena y arcilla (“Clay” y “Sand”)

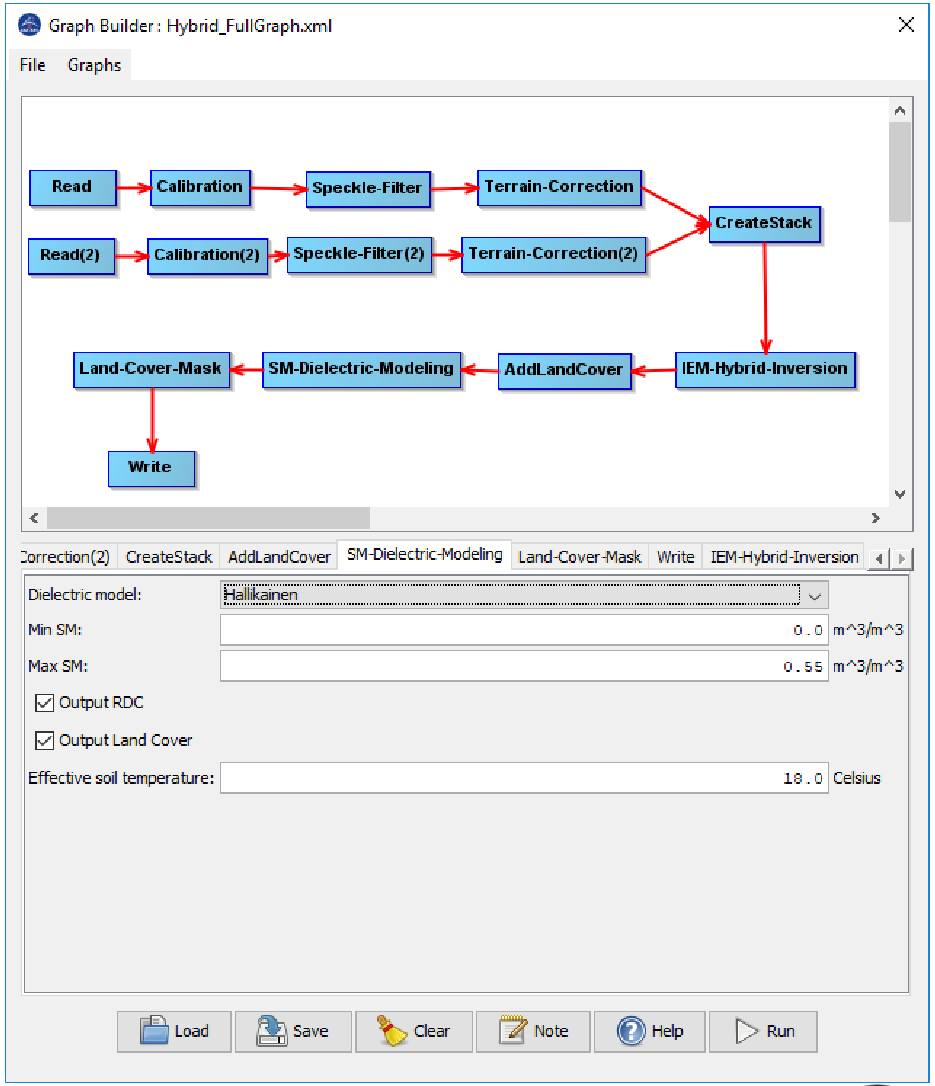


Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

SM Dielectric Modeling (Modelación Dieléctrica de la Humedad del Suelo):

1. Seleccione el modelo de mezclas **Hallikainen**
2. Use la configuración preprogramada para los otros parámetros

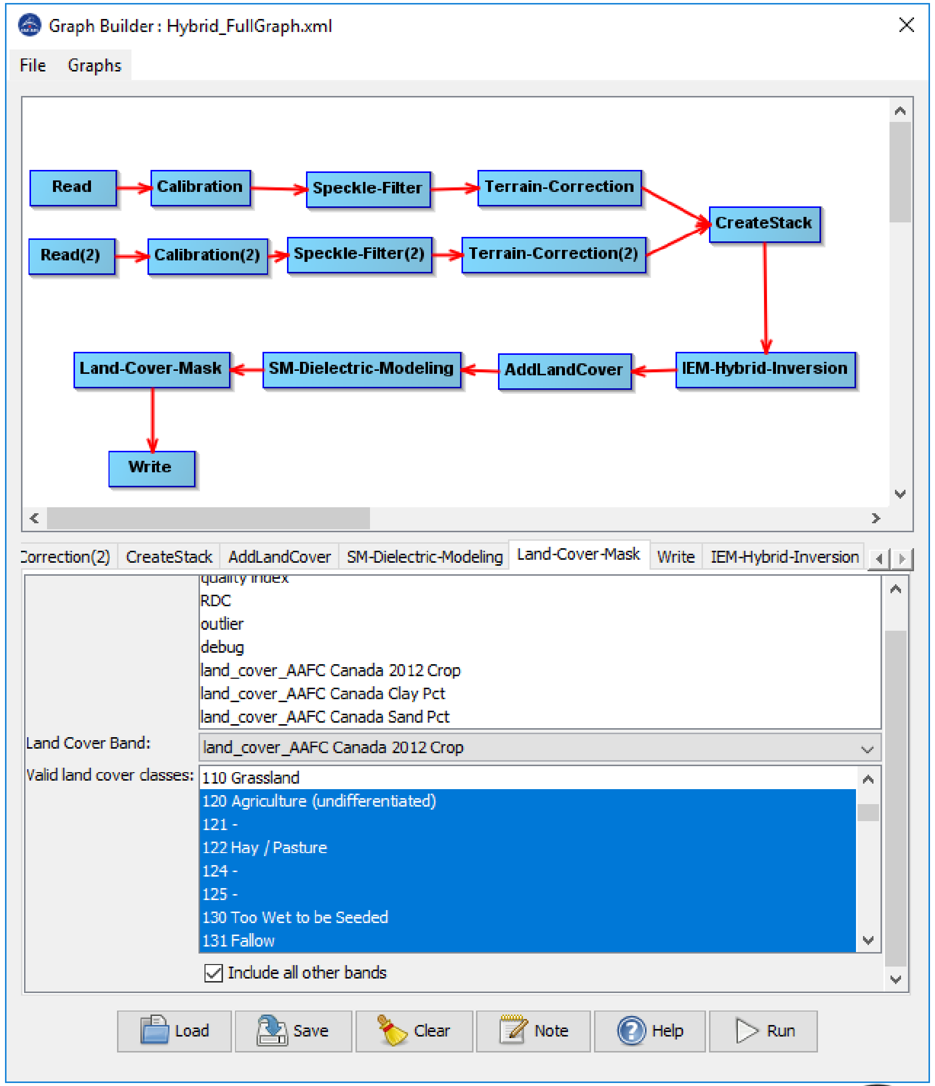
Se utilizó el modelo Haillkainen para estimar la humedad del suelo volumétrica, la cual está basada en los valores dieléctricos recuperados. Este modelo requiere información sobre la textura del suelo (fracciones de arcilla y arena)



Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

Land Cover Mask (Máscara de Cobertura Terrestre)

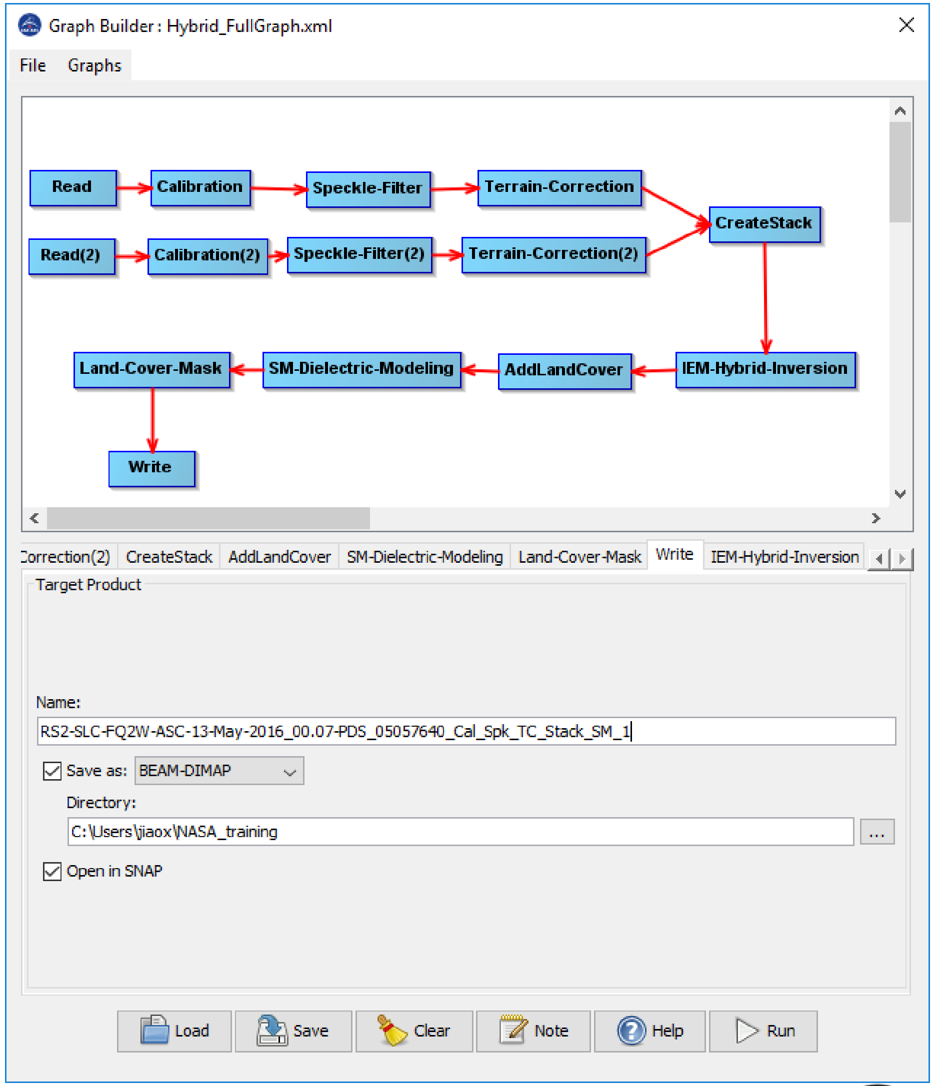
1. Seleccione las clases de cobertura terrestre agrícolas válidas
2. Señale **Exclude all other bands**



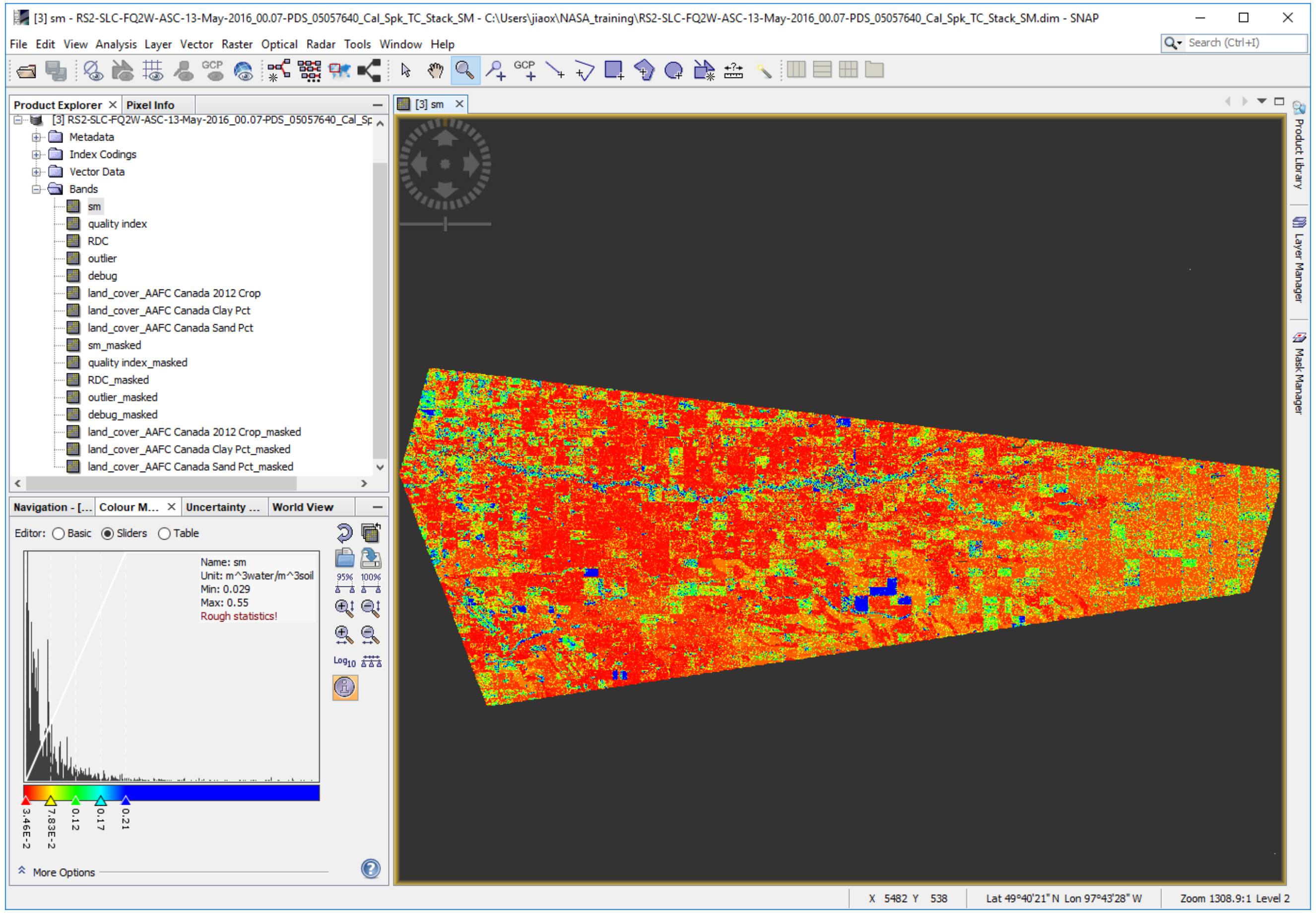
Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido

Write Output (Escribir Salida)

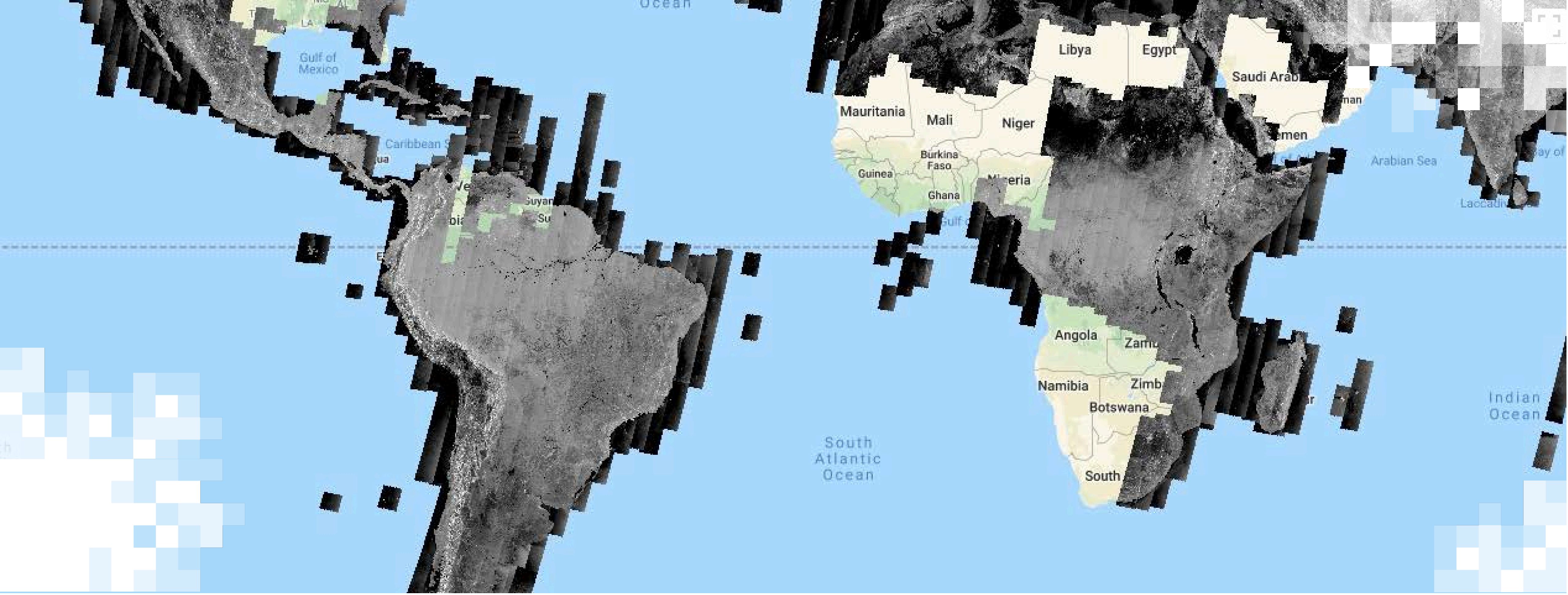
1. Explore y póngale nombre al archivo de salida
2. Seleccione el formato apropiado para el producto de la humedad del suelo recuperada
3. Ejecute el módulo



Procesando la Humedad del Suelo con el Soil Moisture Toolbox en SNAP – Esquema Híbrido



Mapa de la humedad del suelo adquirido mediante la inversión con el IEM y utilizando un par de imágenes RADARSAT-2 adquiridas el 12 y el 13 de mayo de 2016 en el sur de Manitoba

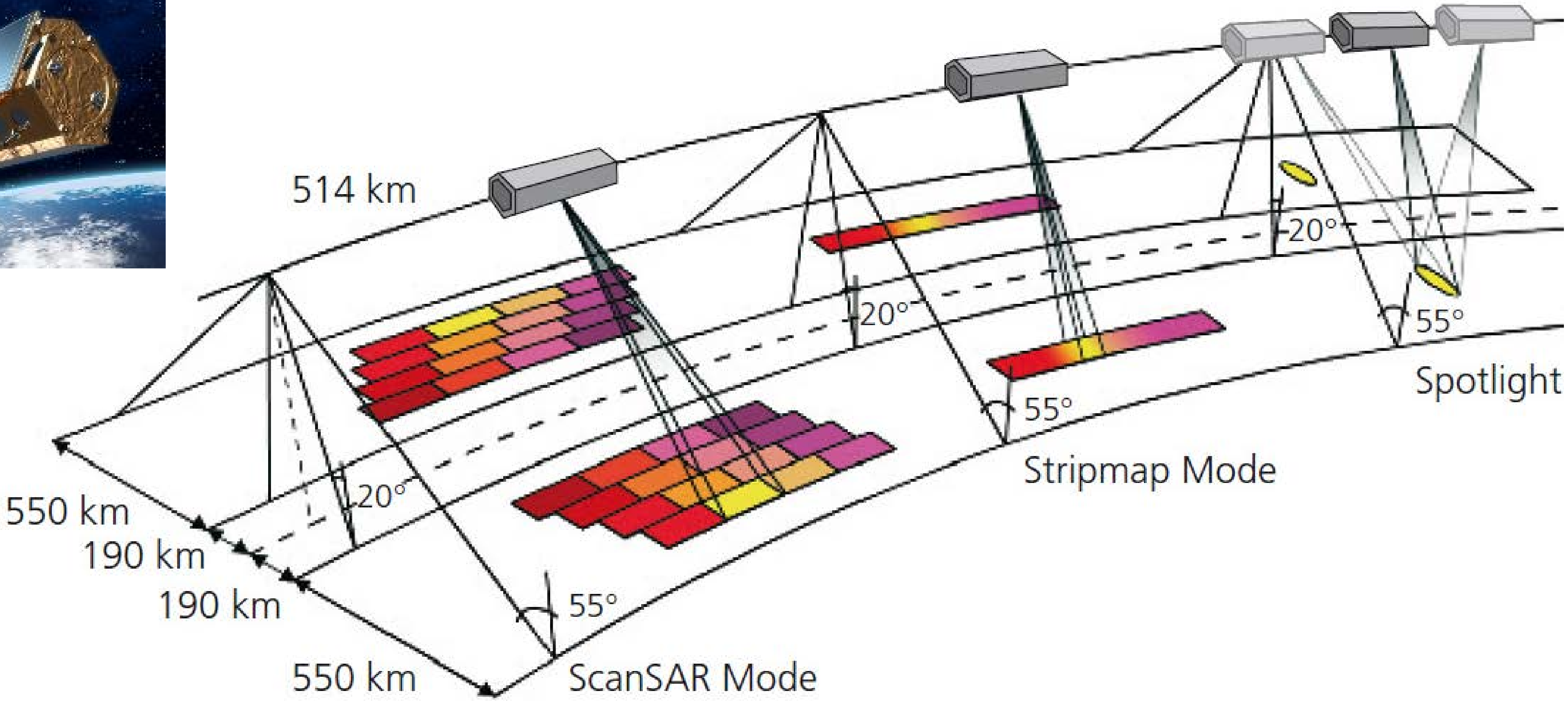


Procesando Datos Multifrecuencia para la Clasificación de Cultivos

Las Imágenes son Pasadas al Clasificador

- Imágenes RADARSAT-2 Wide Fine Quad-Pol adquiridas el 3 de julio, el 27 de julio y el 20 de agosto de 2016
- Imágenes TerraSAR-X StripMap dual pol MGD adquiridas el 26 de agosto y el 17 de agosto de 2016
- Imágenes Sentinel-1 IW mode GRDH adquiridas el 7, 13 y 31 de julio de 2016

TerraSAR-X



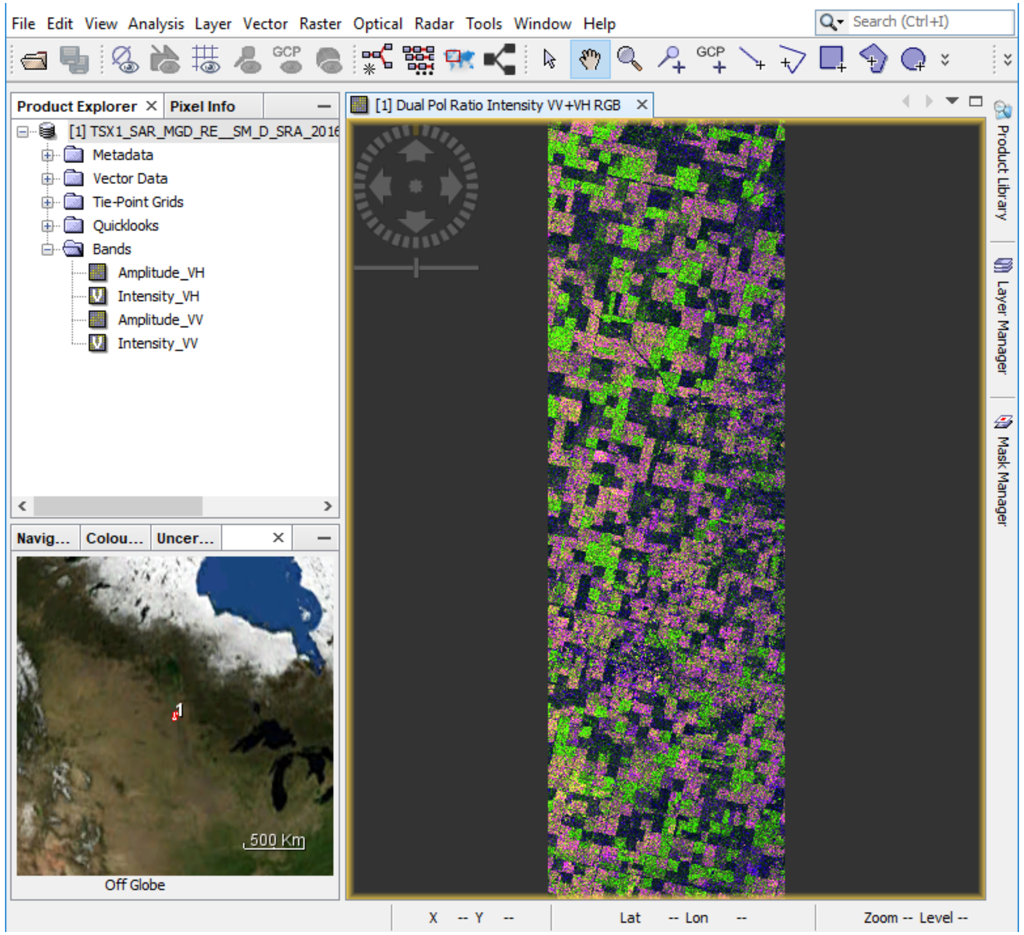
Resumen de los Modos de Escaneo de TerraSAR-X – Tiempo de revisita: 11 días

Fuente de la Imagen: [DLR]

Datos TerraSAR-X StripMap de Polarización Dual

Producto TerraSAR-X StripMap de polarización dual Multi-Look Ground Range Detected (MGD)

- Resolución Nominal: 1.2 m (rango) * 6.6 m (azimut)
- Tamaño Nominal de Escena: 15 km (rango) * 50 km (azimut)
- Polarización Dual: HHyVV, HHyHV, o VV y HH



Datos TerraSAR-X StripMap dual pol MGD adquiridos el 27 de julio de 2016 sobre Carman, Manitoba, Canadá

Pre-Procesamiento de Datos TerraSAR-X Data con SNAP

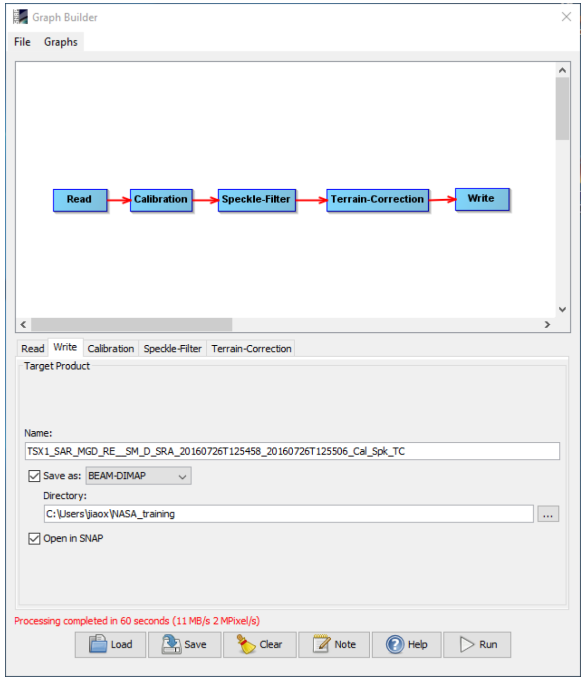
Extraer Retrodispersión

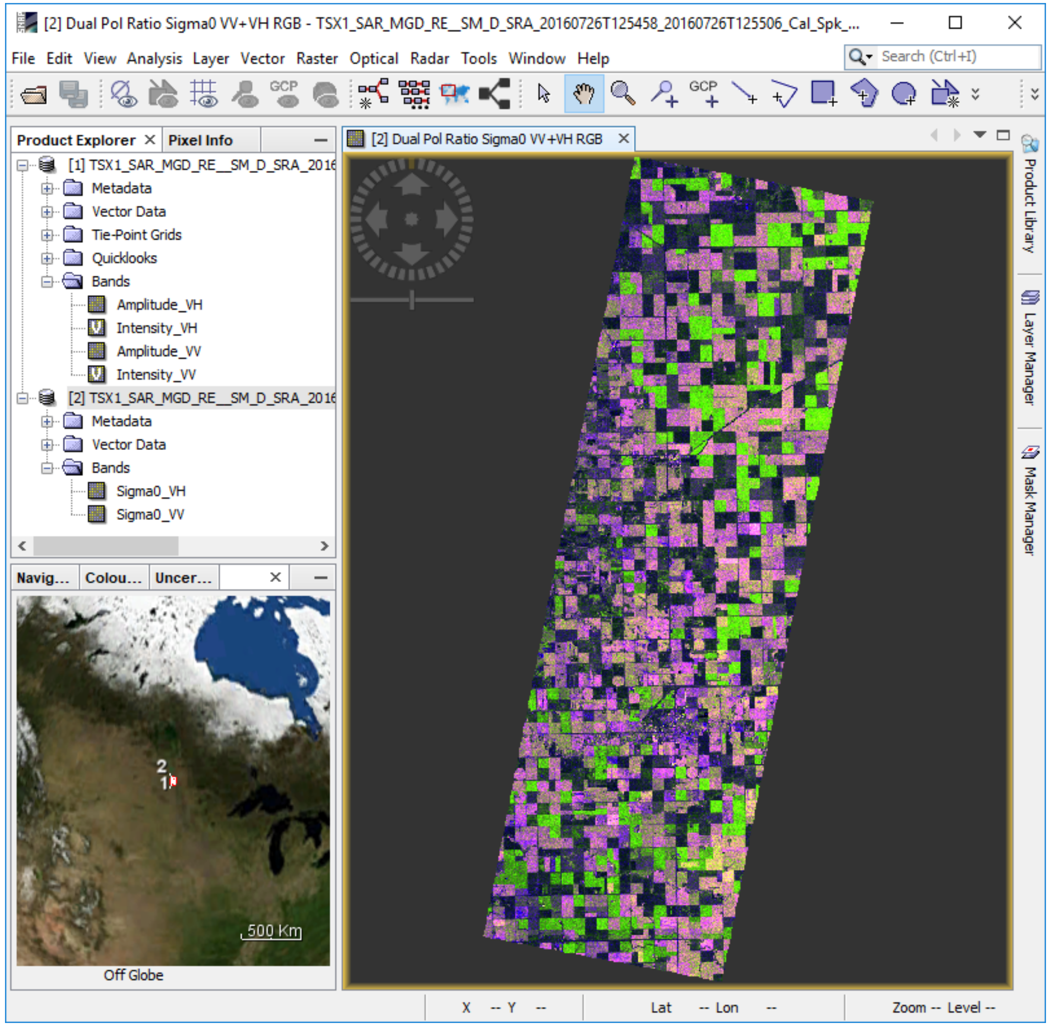
Leer → Calibración → Filtro para Speckle → Corrección Topográfica → Escribir

Pre-Procesamiento de Datos TerraSAR-X Data con SNAP

Graph Builder (Constructor de Gráficos)

- Importe los datos
- Calibre el valor de los pixeles para representar la retrodispersión del radar
- Filtrado “Gamma” en el mapa, ventanilla 7 x 7
- Corrección Topográfica:
  - remuestreo de interpolación bilineal
  - espaciado 5 m para pixeles
  - Proyección UTM zona 14





Sentinel-1

- Cobertura
  - Sentinel-1 comprende dos satélites: A (2014) y B (2016)
  - Cada satélite de Sentinel-1 tiene un ciclo de repetición de12 días
  - Los dos satélites ofrecen un ciclo de repetición exacta de 6 días en el ecuador en el modo de adquisición de barrido interferométrico ancho

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.

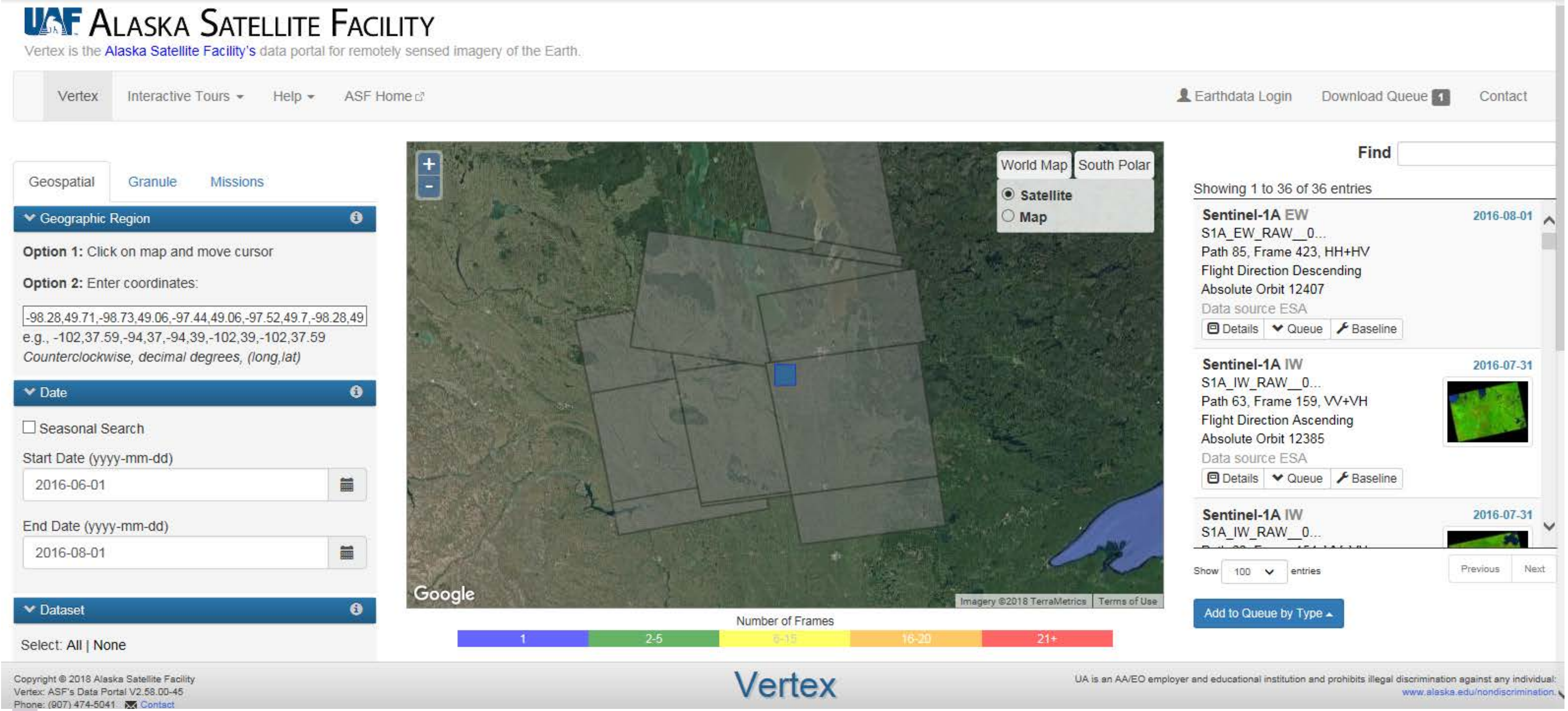
Tipo de Producto para Modo IW:

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

Fuente de Imagen Satelital: ESA/ATG medialab

Acceso a Datos de Sentinel-1 SAR Data desde Vertex

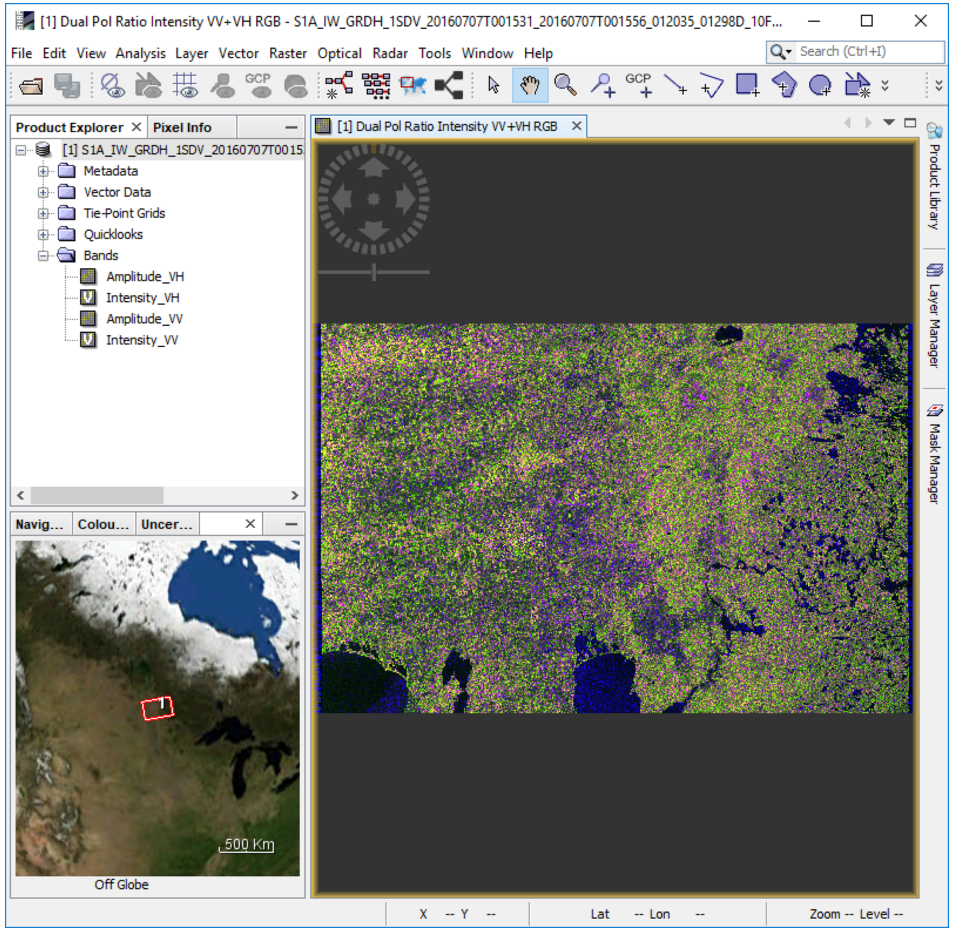
https://vertex.daac.asf.alaska.edu/



Datos del Rango Terrestre Detectados por Sentinel-1

Producto Sentinel-1 IW del alcance terrestre detectado:

- Resolución Nominal: 20 m (rango) * 22 m (azimut)
- Barrido de 250 km
- Polarización Dual : HH /HV, o VV/VH
- GRDH:
  - ground range detected (rango terrestre detectado)
  - high resolution (alta resolución)
  - Multi-Looked: 5 (rango)*1 (azimut)
  - Número de miradas y proyecta a lo largo del rango terrestre
- Se pierde la información sobre la fase



Datos Sentinel-1 GRDH dual pol adquiridos el 7 de julio de 2016, sobre Carman, MB, Canadá

Pre-Procesamiento de Datos Sentinel-1 SAR GRDH con SNAP

Extraer Retrodispersión

1. Leer
2. Aplicar Archivo de Órbita
3. Calibración (Convertir DN en Retrodispersión)
4. Filtro Para Speckle
5. Corrección Topográfica
6. Escribir

Aplique el Archivo de Órbita Preciso

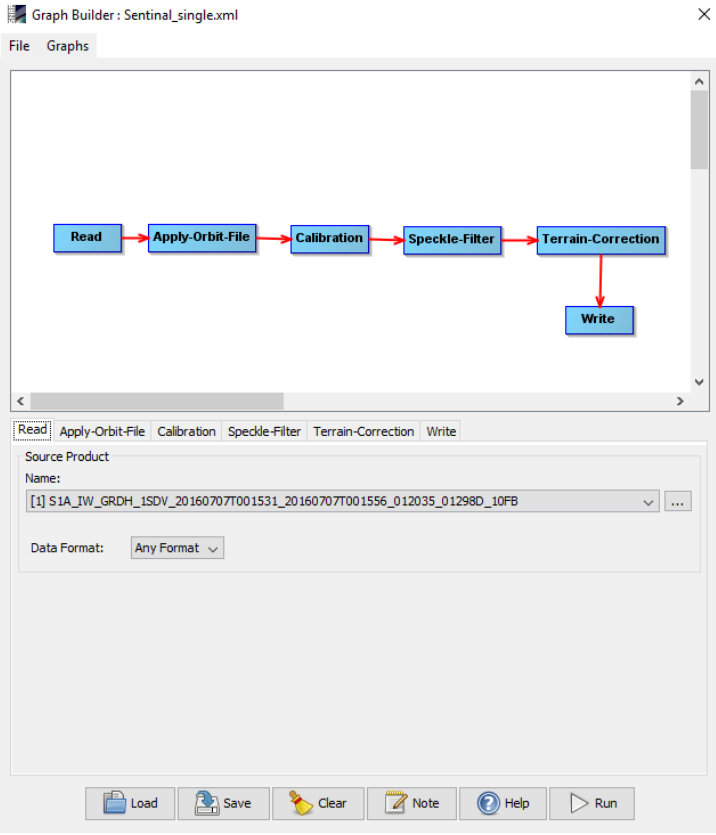
- Para Sentinel-1:
  - Durante la adquisición, la posición del satélite es registrada por un sistema global de navegación por satélite (Global Navigation Satellite System o GNSS)
  - Para garantizar la provisión rápida de productos de Sentinel-1 la información orbital generada por un sistema de navegación a bordo se almacena dentro de los productos de Sentinel-1 Nivel-1
  - Las posiciones orbitales son refinadas después por el servicio de Determinación Precisa de Órbita (Precise Orbit Determination o POD) de Copernicus
  - Los archivos de órbita precisa tienen una exactitud de menos de 5 cm y se producen dentro de 20 días después de la adquisición de datos
  - La exactitud de los archivos de órbita restituidos es de menos de 10 cm. Estos archivos están disponibles 3 horas después de la adquisición de datos.
  - Puede descargar información sobre las órbitas de Sentinel 1 de la página web de la ESA (https://qc.sentinel1.eo.esa.int/)

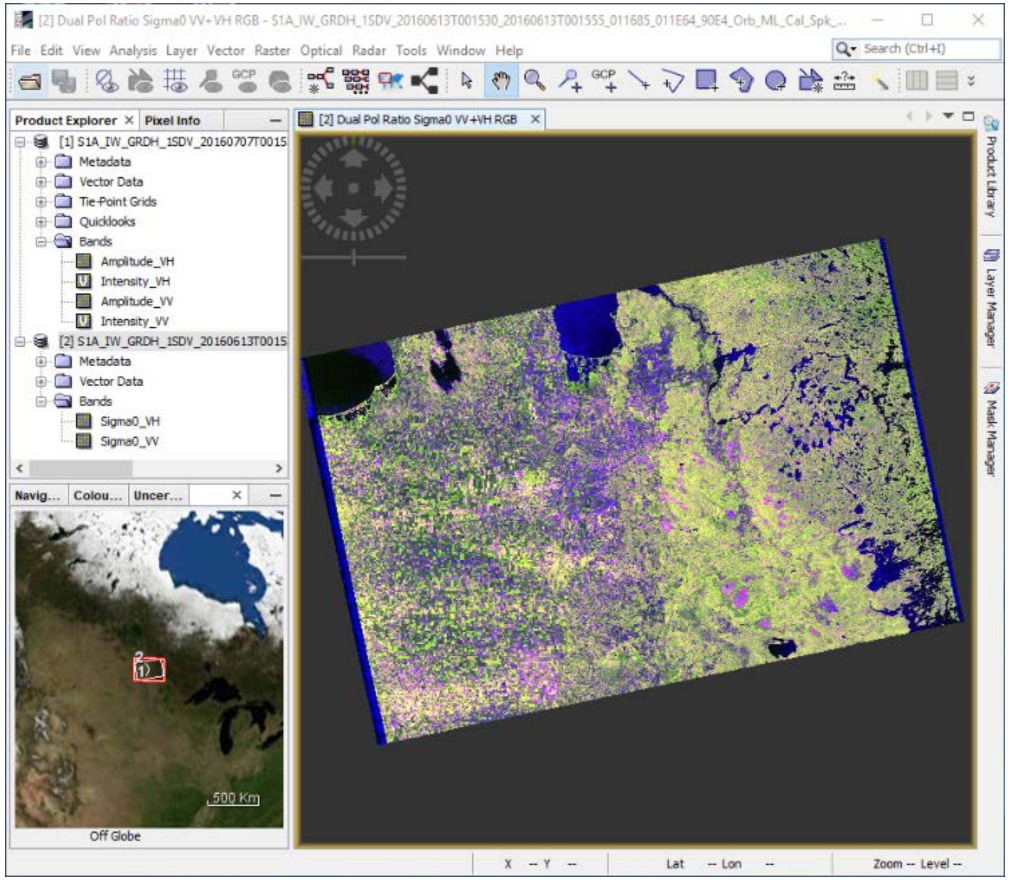
SNAP descarga los archivos de órbita y los almacena en la carpeta …/auxdata/Orbits/Sentinel-1/ https://www.asf.alaska.edu/sentinel/data/

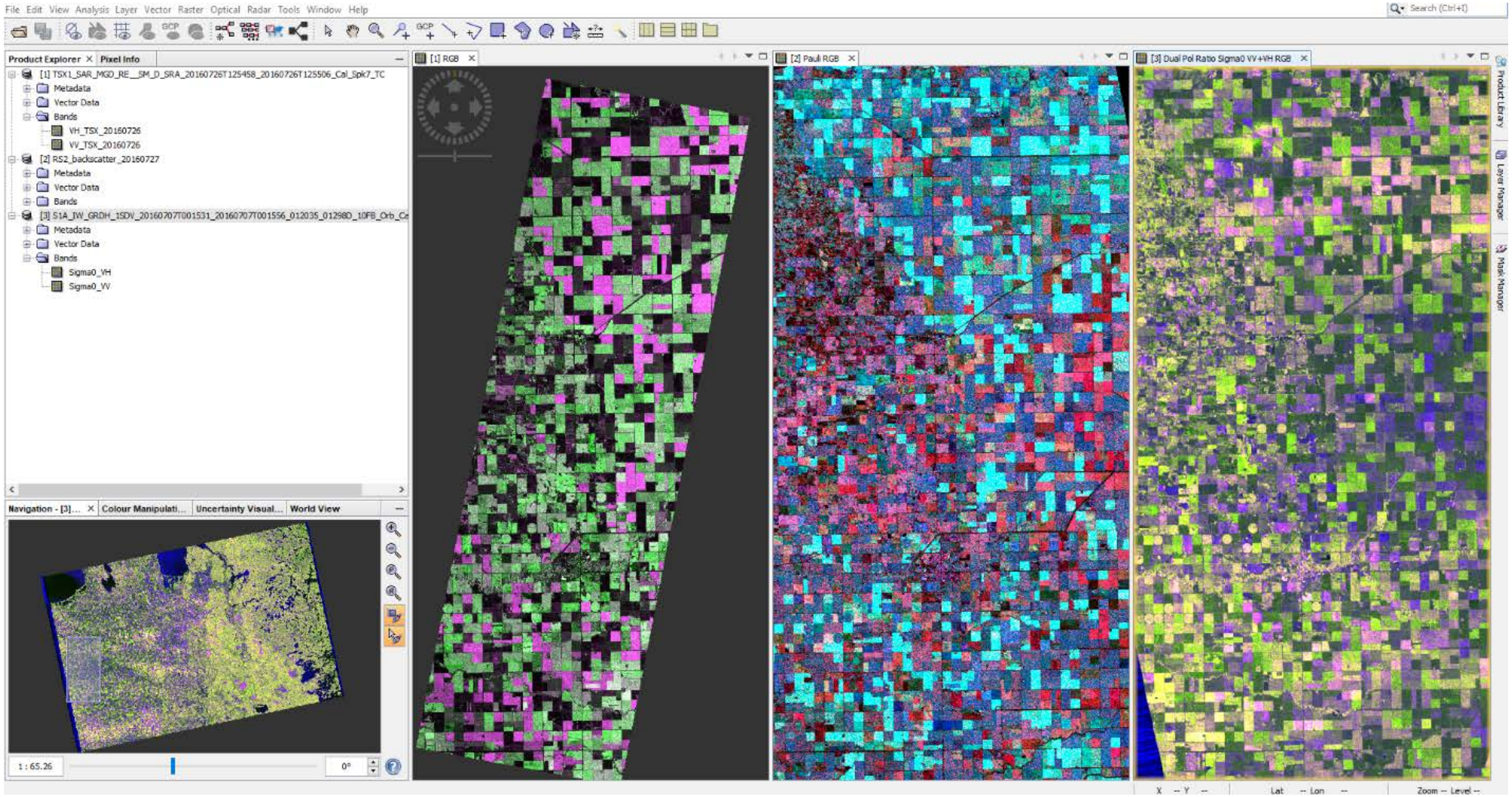
Pre-Procesamiento de Datos Sentinel-1 SAR GRDH con SNAP

Graph Builder (Constructor de Gráficos)

- Importe los datos
- Calibre el valor de los pixeles para representar la retrodispersión del radar
- Filtrado “Gamma” en el mapa, ventanilla 3 x 3
- Corrección Topográfica:
  - remuestreo de interpolación bilineal
  - espaciado de 30 m para pixeles
  - Proyección UTM Zona 14







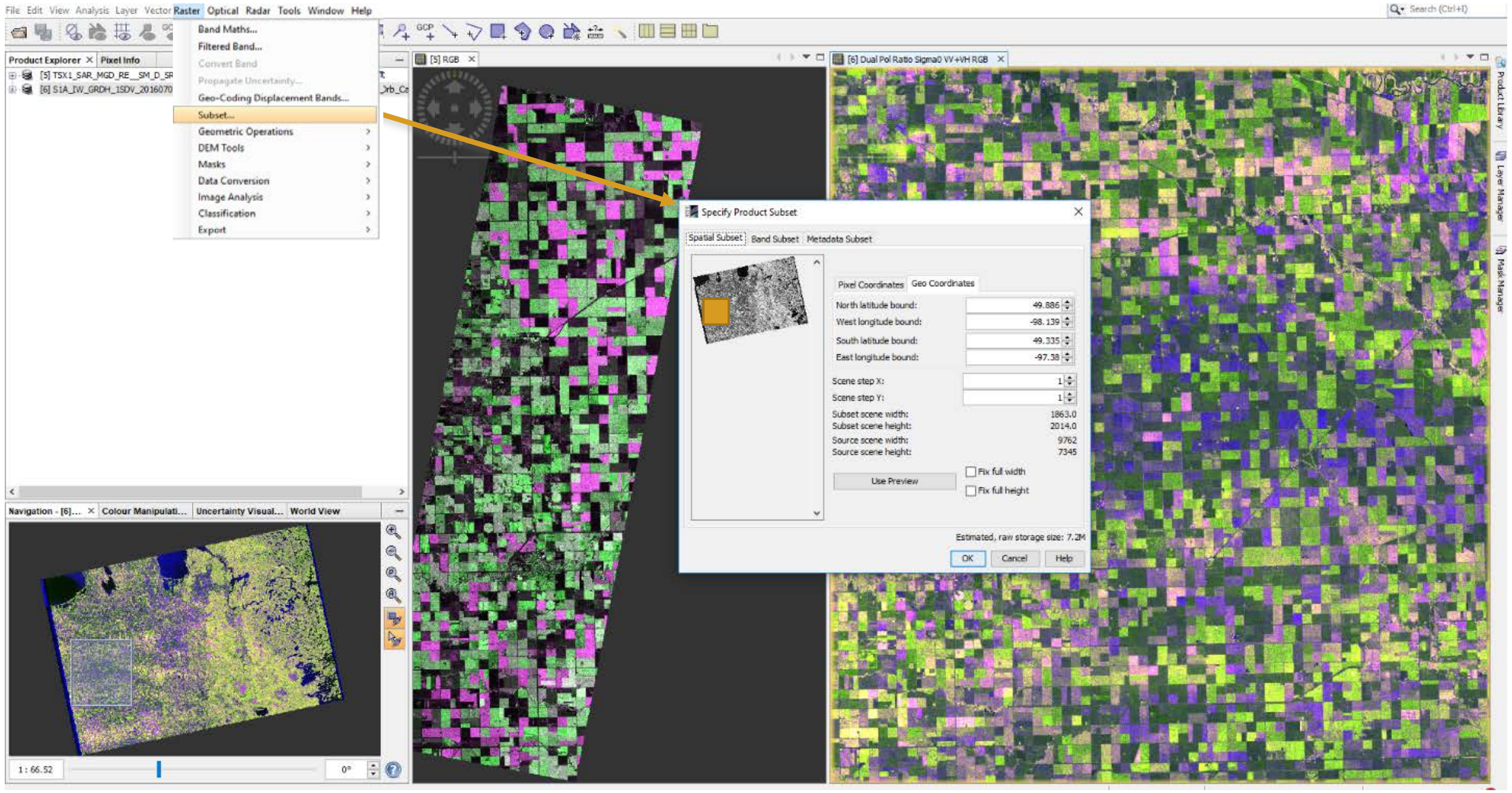
Imágenes TerraSAR-X StripMap dual pol MGD adquiridas el 26 de julio y el 7 de agosto de 2016

Imágenes RADARSAT-2 Wide Fine Quad-Pol adquiridas el 3 de julio, 27 de julio y el 20 de agosto de 2016

Imágenes Sentinel-1 IW modo GRDH adquiridas el 13 de julio, 7 de julio y 31 de julio de 2016

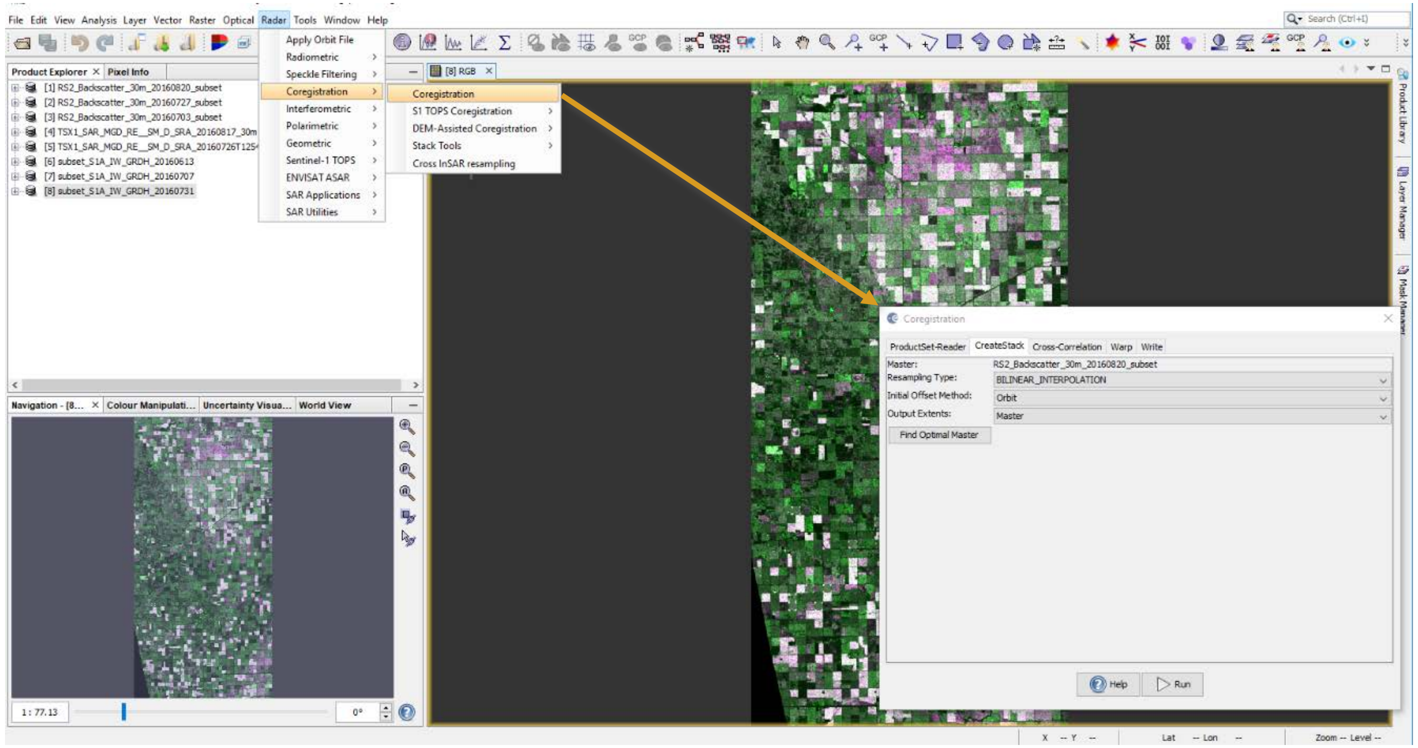
Formar Sub-set con el Ráster para AOI

- Vaya al menu de Raster >> Subset:
  a) Pestaña Spatial Subset→ ingrese las coordenadas para el extremo superior izquierdo y el extremo inferior derecho bajo “geo coordinates”
  b) Band Subset → seleccione las bandas con las que quiere crear sub-sets
  c) Metadata Subset: dejar con el valor preprogramado



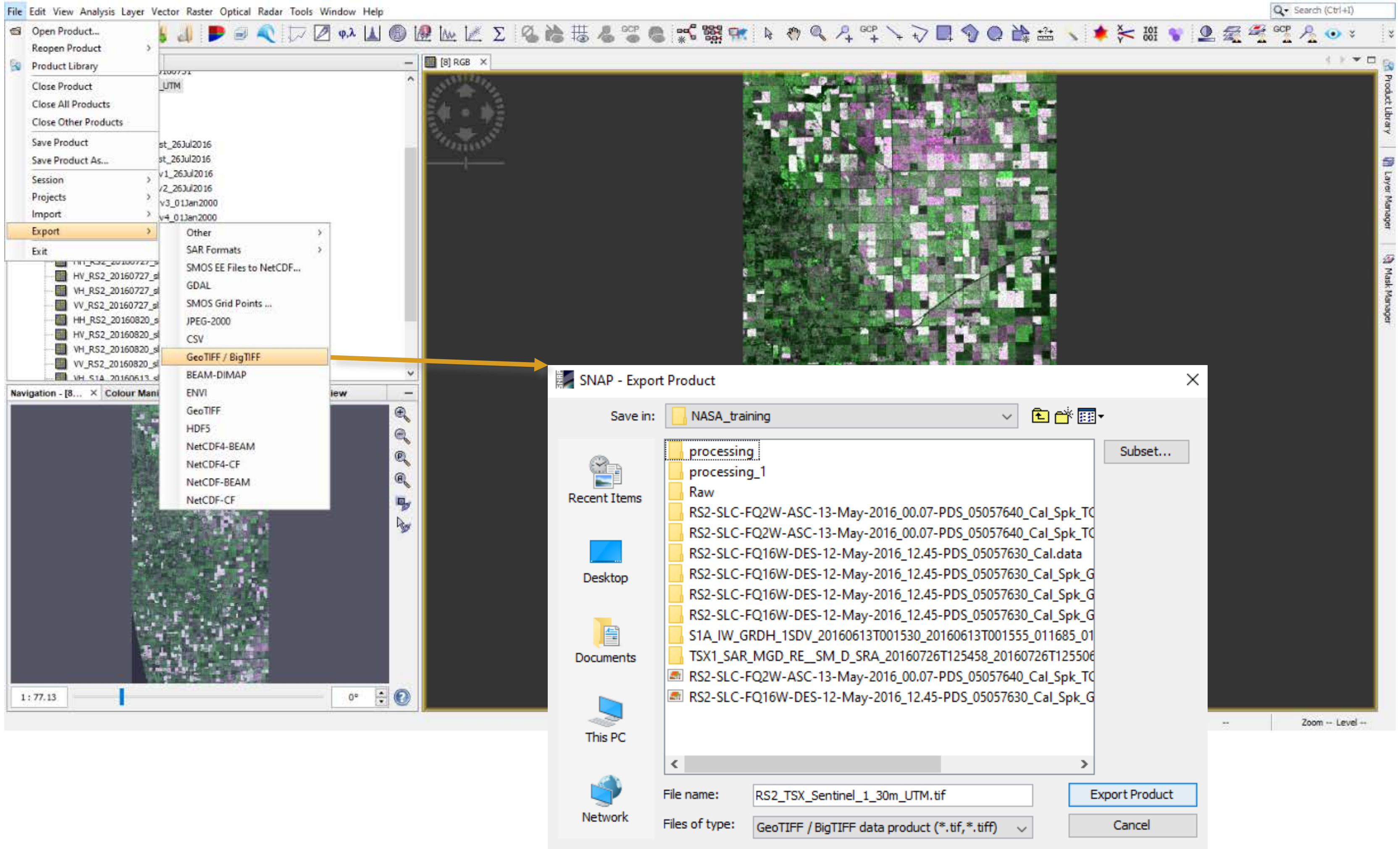
Co-Registración de Imágenes para el Subconjunto

- Alineación espacial de imágenes
- Vaya al menú de Radar >> Coregistration >> Coregistration:
  a) ProductSet-Reader: Haga clic en el signo más (+) con una línea encima para agregar todas las imágenes abiertas → Hacer clic en la flecha giratoria refresca los metadatos
  b) Create Stack: Resampling Type → Bilineal_Interpolation → Haga clic en Find Optimal Master
  c) Otras pestañas: dejar con sus valores preprogramados; verifique que la carpeta Write no esté sobrescribiendo archivos anteriores
  d) Hada clic en Run y cierre la ventanilla cuando haya terminado

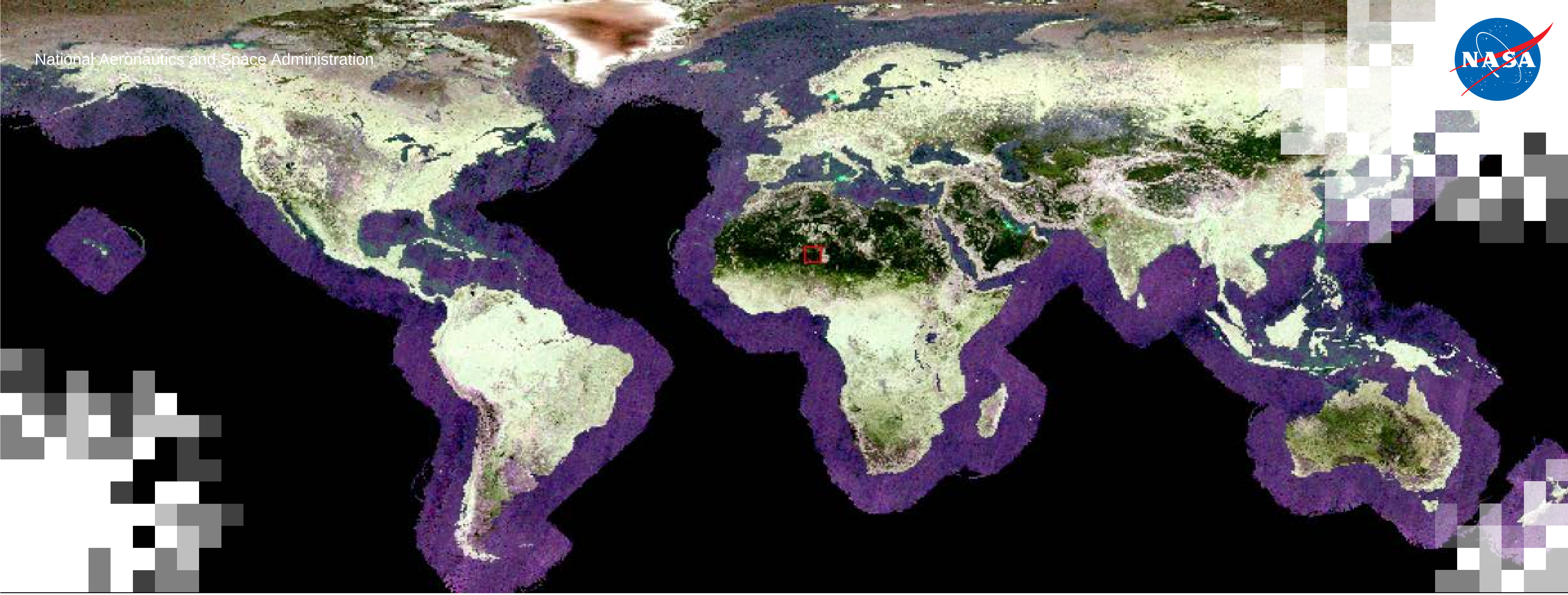


Exportar Datos Apilados Fuera de SNAP

- Alineación espacial de imágenes
  - Vaya al menú de Radar >> File >> export >> seleccione el formato .tif
- Se puede usar.tif en R, Python etc.



National Aeronautics and Space Administration



Aplicaciones de SAR para el Monitoreo Agrícola

Heather McNairn, Xianfeng Jiao, Sarah Banks y Amir Behnamian

4 de septiembre de 2019

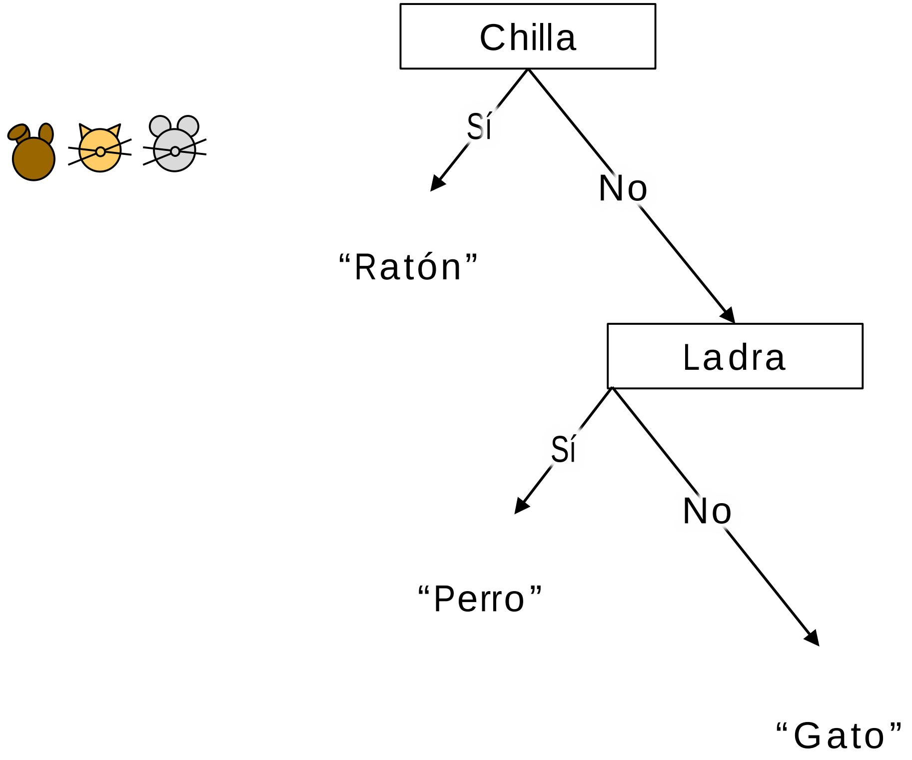
Objetivos de Aprendizaje

Al finalizar esta presentación, Usted podrá entender:

- Los Árboles de Clasificación y Regresión (CART- Classification and Regression Trees)
- Las Ventajas de Random Forests
- Random Forests:
  - Fundamentos
  - Archivos de entrada y parámetros
- Cómo implementar Random Forests en R

Árboles de Clasificación y Regresión (CART): Fundamentos

- Dividen los datos basado en un conjunto de reglas binarias
- Producen un “árbol” de decisiones fácil de interpretar
- Intentan hacer grupos homogéneos hasta donde sea posible



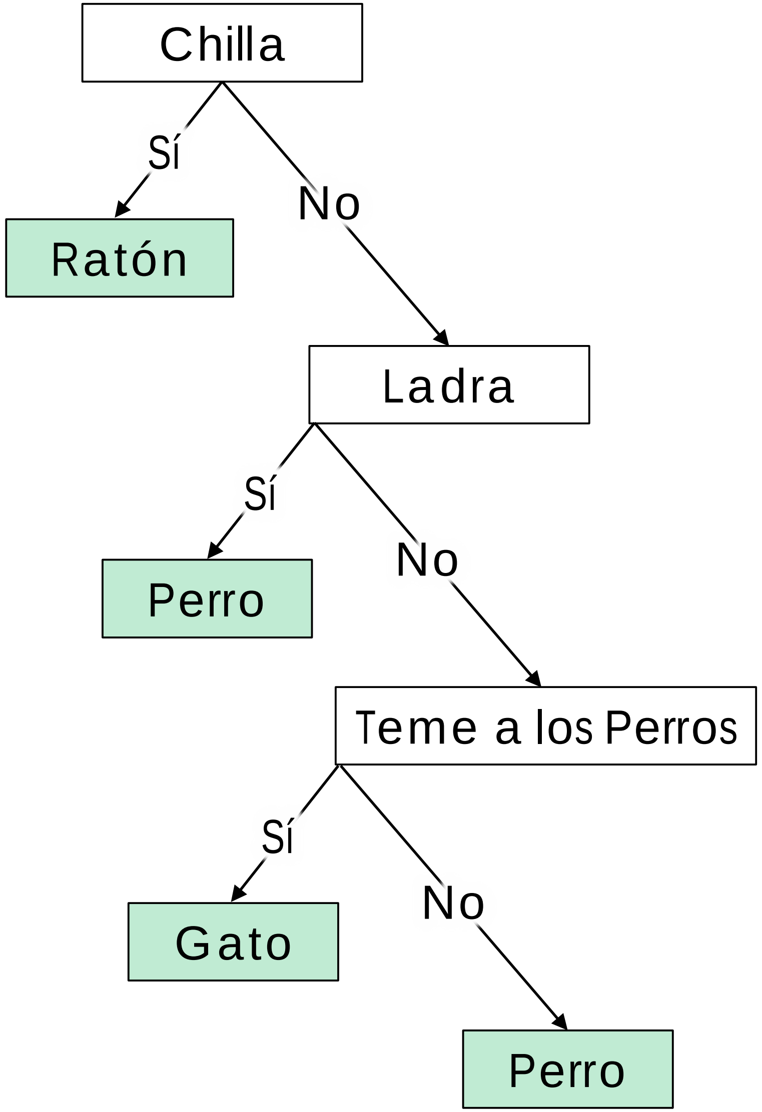
Árboles de Clasificación y Regresión (ARC): Fundamentos

- Términos Importantes:

	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	[illegible]	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

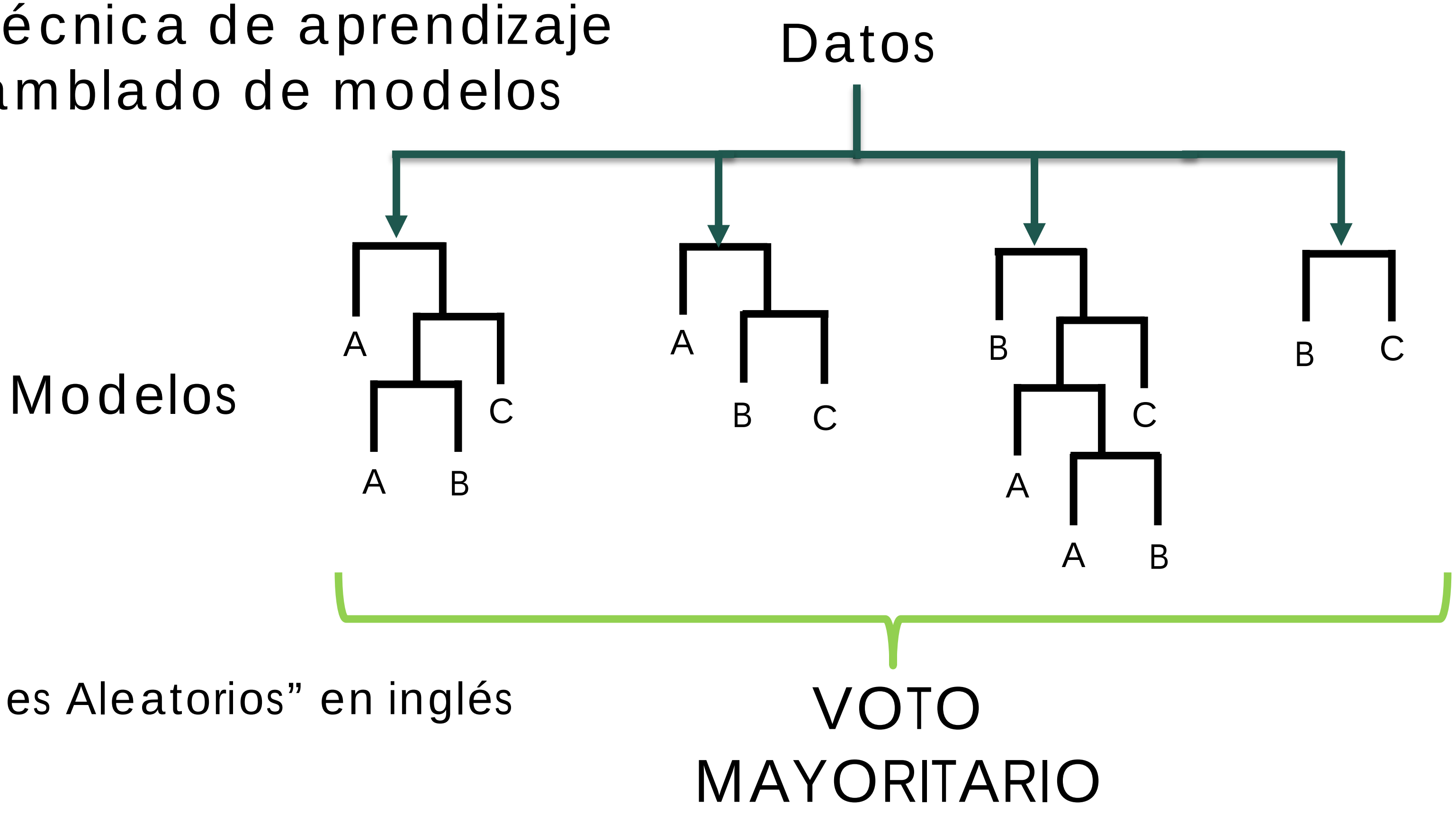
RAMA
ÁRBOL
IZQUIERDA
DERECHA

-



Random Forests*: Fundamentos

- Un “bosque” de árboles de decisión binarios (Breiman, 2001)
- Funciona bien con conjuntos de datos de grandes dimensiones (continuos o categóricos)
- Es una técnica de aprendizaje de ensamblado de modelos



*”Bosques Aleatorios” en inglés

Random Forests: Fundamentos

- Random Forests vs. CART

	CART	Random Forests
Número de árboles	1	n>>1
Podar	Aplica	Todos completamente crecidos
Variables probadas para divisiones	Todas	m<<M (todas las variables)
Conjunto de Entrenamiento	Todos los datos	≈ 2/3
Exactitud	Requiere que sea independiente	Internamente estimada (OOBE)

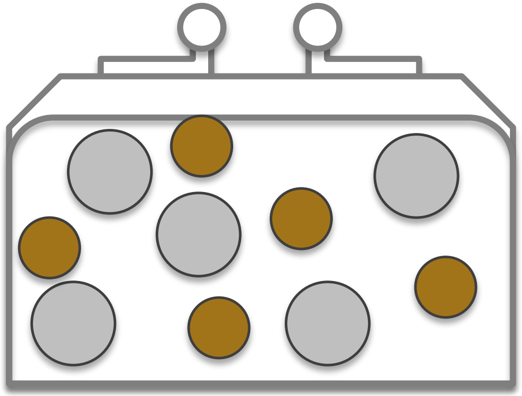
Random Forests: Parámetros y Términos Importantes

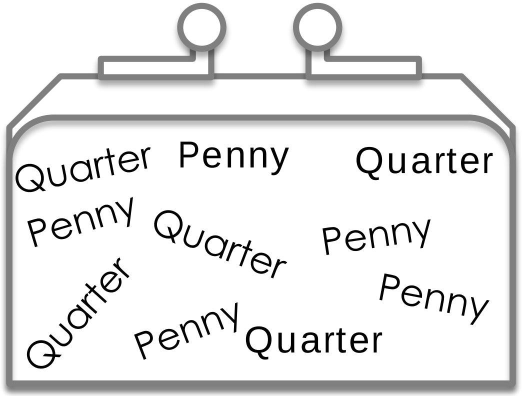
- ntree – el número de árboles por ser generados por bosque.
- Probabilidad – el número de árboles que votaron con la mayoría dividido por el número total de árboles.
- mtry – el número de variables probadas para determinar la división óptima en cada nodo.
- Exactitud "Out of Bag" (fuera de la bolsa) – validación interna; basada en ≈ $^1/_3$ del conjunto de datos que no se utilizó durante la construcción de un árbol particular.
- Mean Decrease in Accuracy (MDA) – Disminución media de la exactitud: cuantifica la "importancia" de las variables midiendo los cambios en la exactitud cuando los valores de la variable son aleatoriamente permutados.

Random Forests: Fundamentos

- Impureza de Gini: la probabilidad de que un elemento elegido aleatoriamente sea clasificado incorrectamente
- Elija un dato de un conjunto aleatoriamente y clasifíquelo de manera aleatoria según la distribución de clases
- 25% + 25% = 50%; Impureza de Gini = 0.5

Evento	Probabilidad
Elegir quarter (50%), clasificar quarter (50%)	50% x 50% = 25%
Elegir quarter (50%), clasificar penny (50%)	50% x 50% = 25%
Elegir penny (50%), clasificar penny (50%)	50% x 50% = 25%
Elegir penny (50%), clasificar quarter (50%)	50% x 50% = 25%

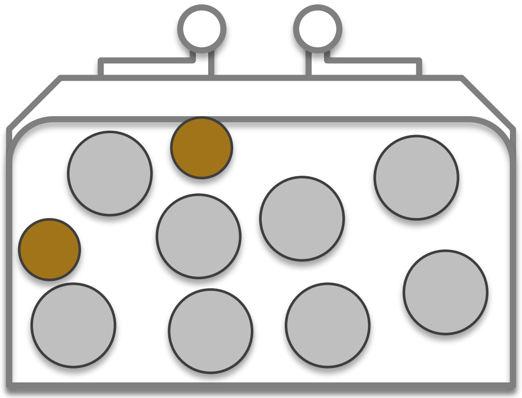


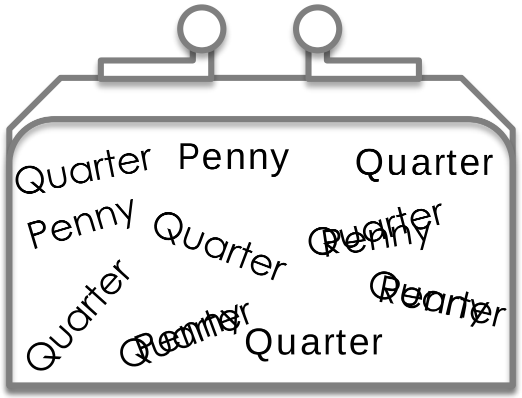


Random Forests: Fundamentos

- Impureza de Gini: la probabilidad de que un elemento elegido aleatoriamente sea clasificado incorrectamente
- Elija un dato de un conjunto aleatoriamente y clasifíquelo de manera aleatoria según la distribución de clases
- 16% + 16% = 32%; Impureza de Gini = 0.32

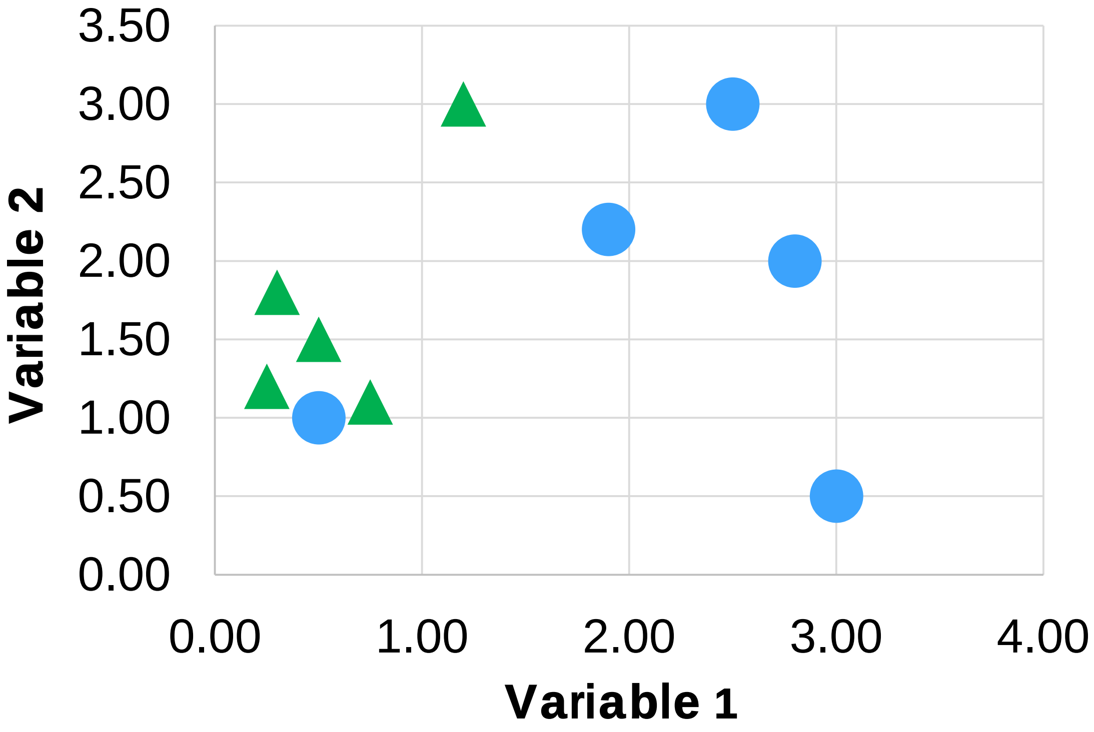
Evento	Probabilidad
Elegir quarter (80%), clasificar quarter (80%)	80% x 80% = 64%
Elegir quarter (80%), clasificar penny (20%)	80% x 20% = 16%
Elegir penny (20%), clasificar penny (20%)	20% x 20% = 4%
Elegir penny (20%), clasificar quarter (80%)	20% x 80% = 16%





Random Forests: Fundamentos

- Cómo se utiliza la Impureza de Gini para dividir:
  - No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  - Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene



Random Forests: Fundamentos

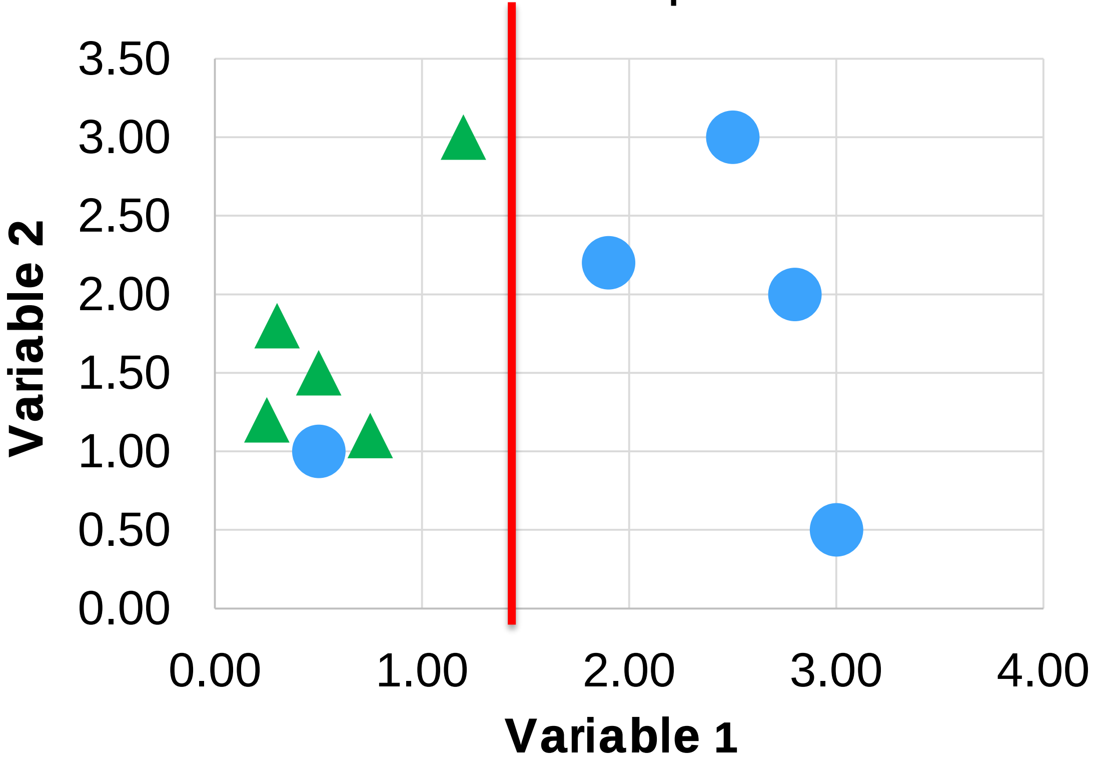
- Cómo se utiliza la Impureza de Gini para dividir:
  - No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  - Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene

Impurezas de Gini

Antes de la división= 0.50

Nodo Derecho = 0.00

Nodo Izquierdo = 0.28



Random Forests: Fundamentos

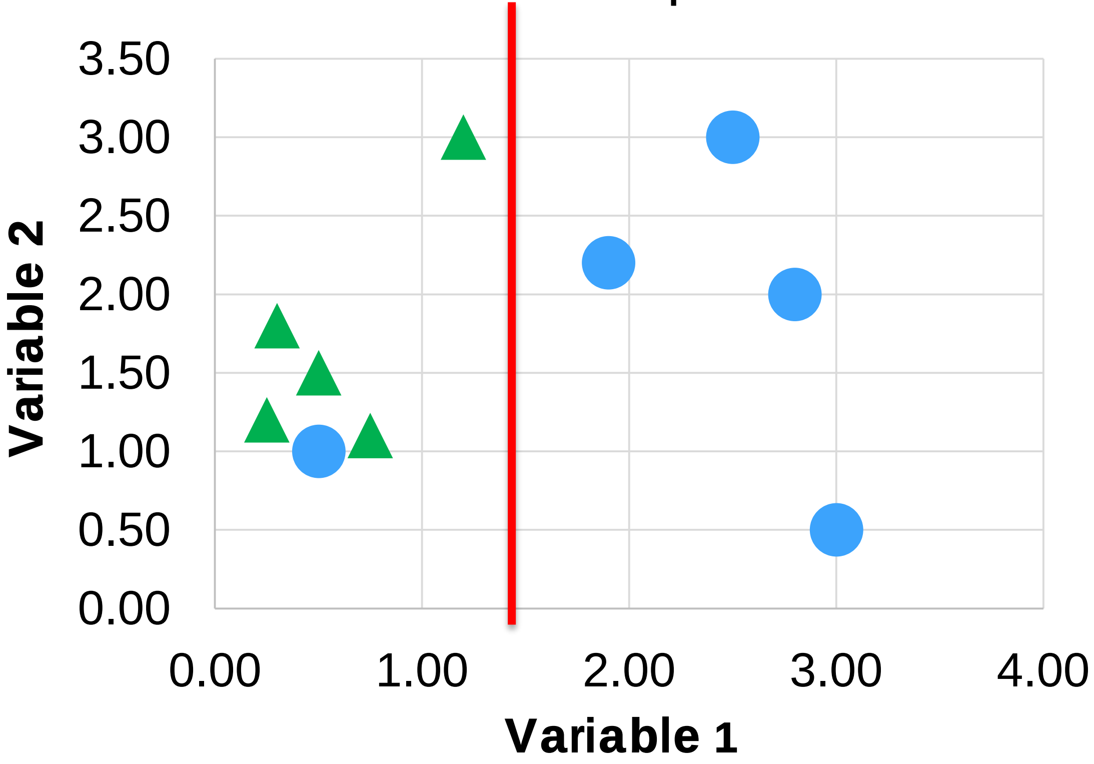
- Cómo se utiliza la Impureza de Gini para dividir:
  – No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  – Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene

Impurezas de Gini

Antes de la división= 0.50

Nodo Derecho = 0.00

Nodo Izquierdo = 0.28



El nodo tiene el 40% de los datos, el nodo izquierdo tiene el 60%

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

Con esta división la cantidad de impureza removida es:

0.5 – 0.17 = 0.33

Random Forests: Fundamentos

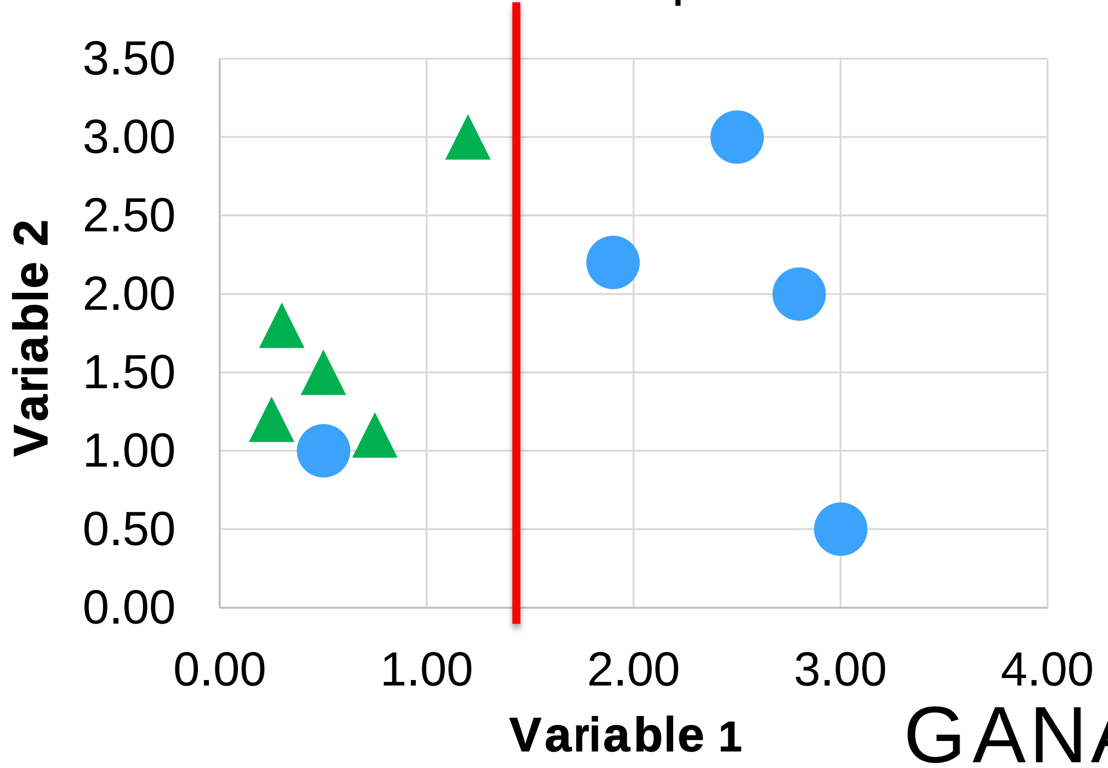
- Cómo se utiliza la Impureza de Gini para dividir:
  - No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  - Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene

Impurezas de Gini

Antes de la división= 0.50

Nodo Derecho = 0.00

Nodo Izquierdo = 0.28



El nodo tiene el 40% de los datos, el nodo izquierdo tiene el 60%

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

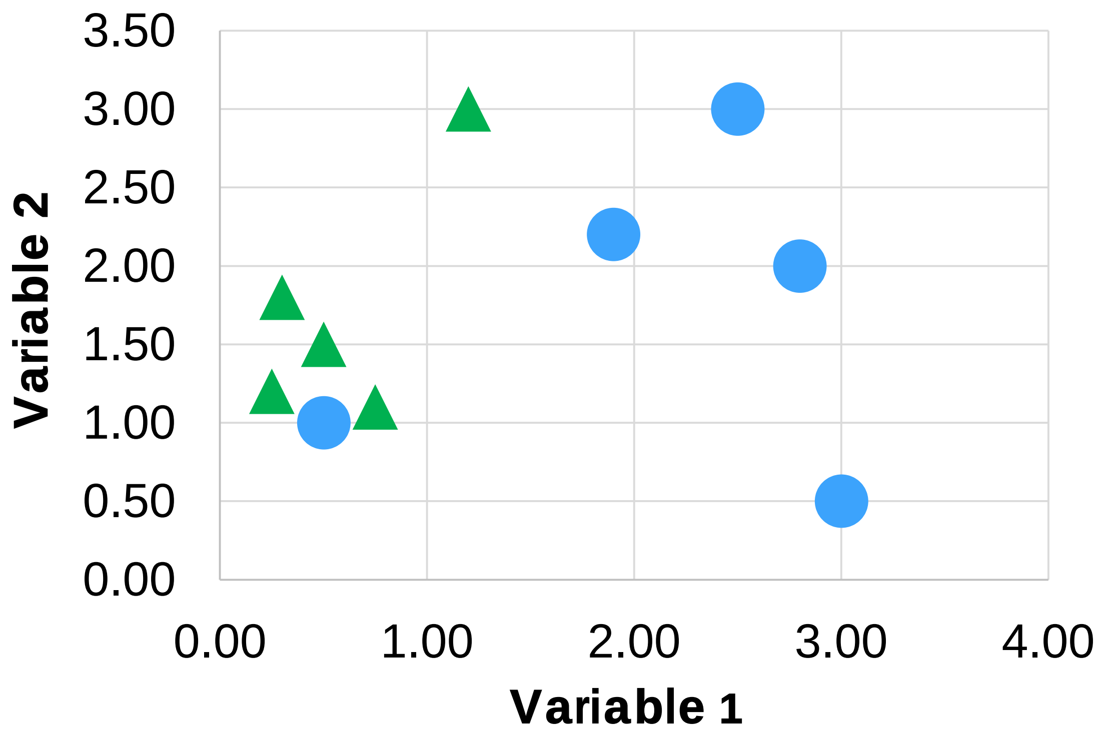
Con esta división la cantidad de impureza removida es:

0.5 – 0.17 = 0.33

GANANCIA DE INFORMACIÓN

Random Forests: Fundamentos

- Cómo se utiliza la Impureza de Gini para dividir:
  - No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  - Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene



Random Forests: Fundamentos

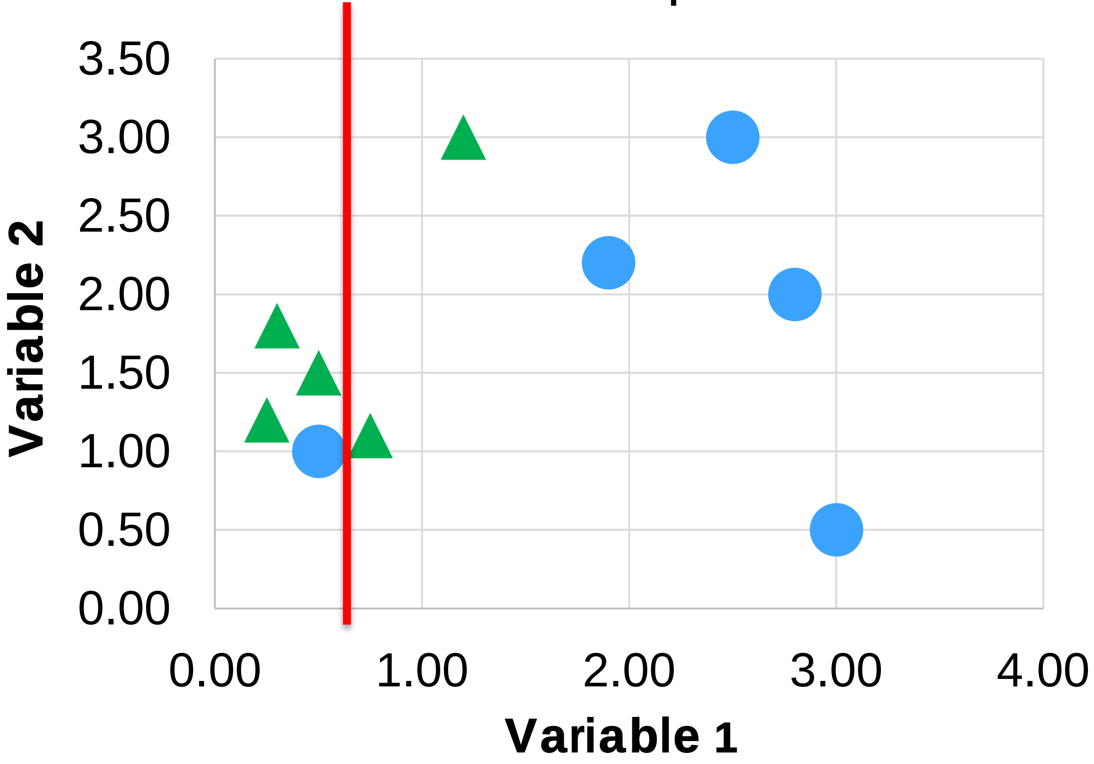
- Cómo se utiliza la Impureza de Gini para dividir:
  - No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  - Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene

Impurezas de Gini

Antes de la División = 0.50

Nodo Derecho = 0.44

Nodo Izquierdo = 0.38



Random Forests: Fundamentos

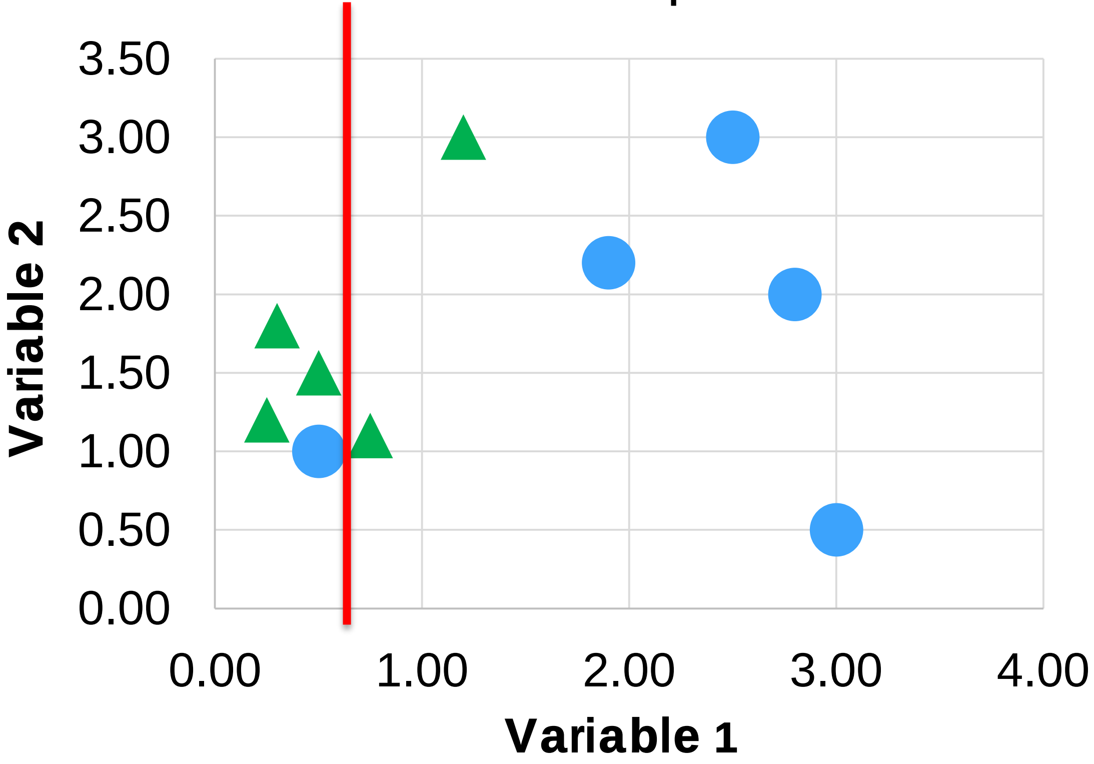
- Cómo se utiliza la Impureza de Gini para dividir:
  - No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  - Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene

Impurezas de Gini

Antes de la División = 0.50

Nodo Derecho = 0.44

Nodo Izquierdo = 0.38



El nodo derecho tiene el 60% de los datos, el nodo izquierdo tiene el 40%

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

Con esta división la cantidad de impureza removida es:

0.5 – 0.42 = 0.08

Random Forests: Fundamentos

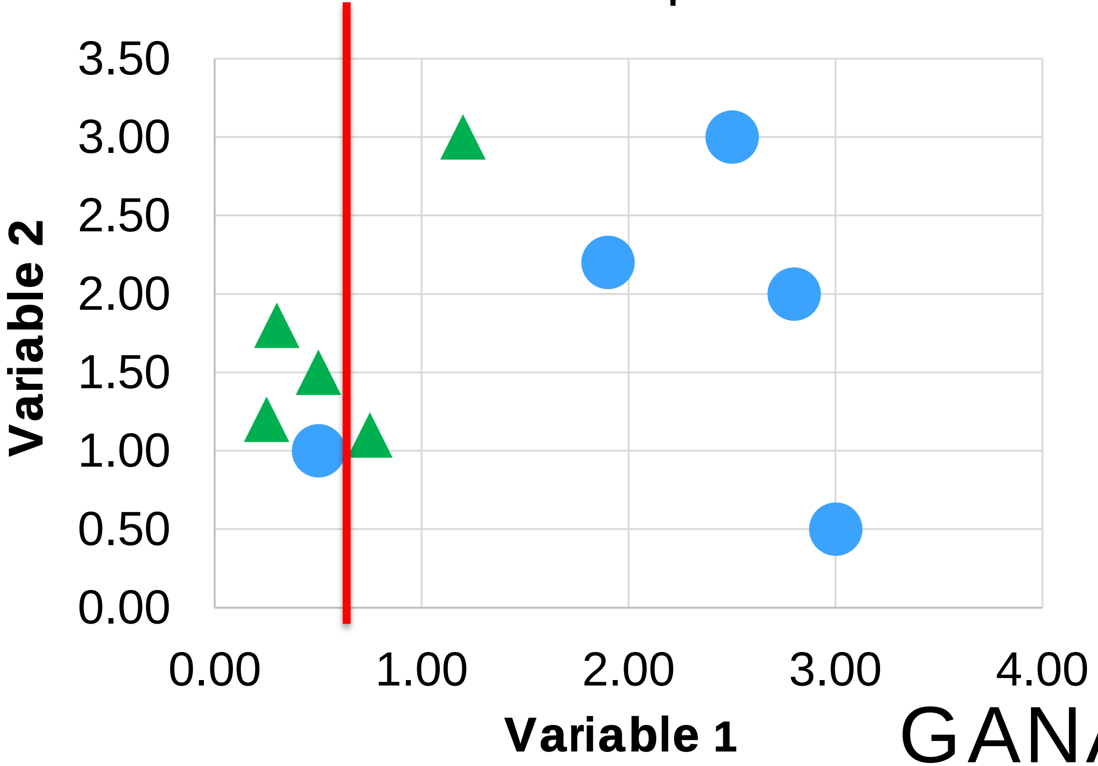
- Cómo se utiliza la Impureza de Gini para dividir:
  - No sabemos dónde sería la mejor división, pero podemos probar todas las divisiones posibles
  - Determina la calidad de la división midiendo la impureza de los nodos subsiguientes según el número de datos que contiene

Impurezas de Gini

Antes de la División = 0.50

Nodo Derecho = 0.44

Nodo Izquierdo = 0.38



El nodo derecho tiene el 60% de los datos, el nodo izquierdo tiene el 40%

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

Con esta división la cantidad de impureza removida es:

0.5 – 0.42 = 0.08

GANANCIA DE INFORMACIÓN

Random Forests: Fundamentos

- Entradas para Random Forests:
- Clases indicadas con variables predictoras asociadas
  - Datos categóricos o continuos

	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: Ejemplo

1) Crear los datos de entrenamiento para crecer el árbol
   – Para un número N de datos, aleatoriamente muestree N casos (con remplazo)

Random Forests: Ejemplo

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

Datos de Entrenamiento

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

Random Forests: Ejemplo

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

Datos de Entrenamiento

	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: Ejemplo

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

Datos de Entrenamiento

	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: Ejemplo

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

Datos de Entrenamiento

	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: Ejemplo

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

Datos de Entrenamiento

	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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Yes	Yes	0.36	0.17	Yes
16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Yes	Yes	0.3	0.16	Yes
16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Yes	Yes	0.3	0.16	Yes
16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Yes	Yes	0.3	0.16	Yes
18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Yes	Yes	0.33	0.22	Yes
18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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

Restantes para Evaluación de Precisión Interna $\approx {}^{1}/_{3}$

Random Forests: Ejemplo

2) Para un número M de variables, seleccione un subconjunto (m<<M) aleatoriamente para determinar cómo cada nodo se ramifica (**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: Ejemplo

Datos de Entrenamiento

	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

Resultados de Clasificación Basados en Puntos de División

	Mouse	Cat	Dog
Yes	8	0	0
No	0	8	11

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32

Random Forests: Ejemplo

Datos de Entrenamiento

	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

Resultados de Clasificación Basados en Puntos de División

	Mouse	Cat	Dog
Yes	0	7	0
No	8	1	11

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32
Meows	Yes o No	0.28

Random Forests: Ejemplo

Datos de Entrenamiento

	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

Resultados de Clasificación Basados en Puntos de División

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

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32
Meows	Yes o No	0.28
Height	<=0.05	0.03

Random Forests: Ejemplo

Datos de Entrenamiento

	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

Resultados de Clasificación Basados en Puntos de División

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

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32
Meows	Yes o No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09

Random Forests: Ejemplo

Datos de Entrenamiento

	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

Resultados de Clasificación Basados en Puntos de División

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

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32
Meows	Yes o No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09
Height	<=0.07	0.17

...

Random Forests: Ejemplo

Datos de Entrenamiento

	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

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32
Meows	Yes o 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: Ejemplo

Datos de Entrenamiento

	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

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32
Meows	Yes o 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

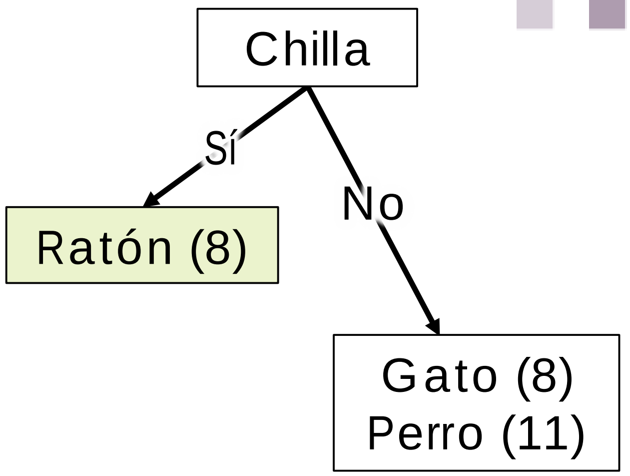
Random Forests: Ejemplo

Datos de Entrenamiento

	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

Ganancia de Gini para identificar la división óptima

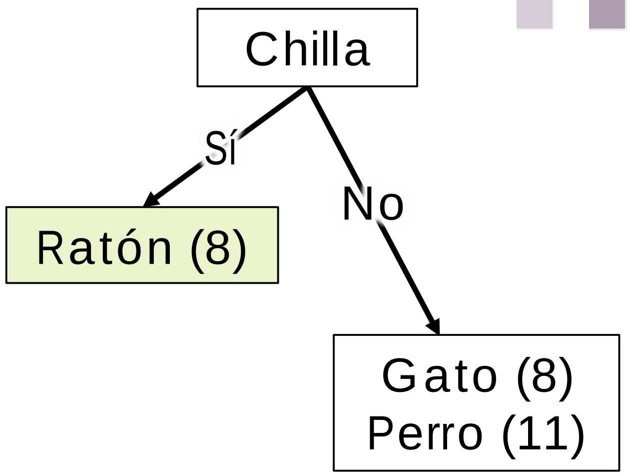
Variable	Puntos de División Evaluados	Gini
Squeaks	Yes o No	0.32
Meows	Yes o No	0.28
Height	<=0.05	0.03
Height	<=0.06	0.09
Height	<=0.07	0.17
Height	<=0.08	0.26
Height	<=0.09	0.28
Height	<=0.15	0.25
Height	<=0.16	0.26
Height	<=0.17	0.28
Height	<=0.22	0.27
Height	<=0.26	0.31
Height	<=0.27	0.22
Height	<=0.29	0.18
Height	<=0.32	0.09
Height	<=0.33	0.04
Height	<=0.35	0.00



Random Forests: Ejemplo

Datos de Entrenamiento

	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: Ejemplo

Datos de Entrenamiento

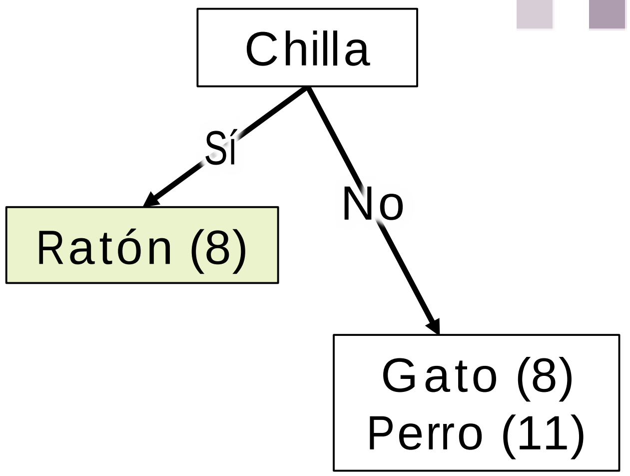
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	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

Resultados de Clasificación Basados en Puntos de División

	Cat	Dog
Yes	1	0
No	7	11

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Cheese	Yes o No	0.04



Random Forests: Ejemplo

Datos de Entrenamiento

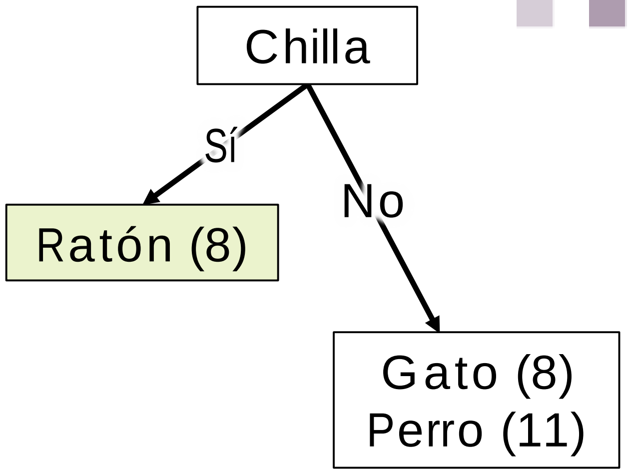
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	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

Resultados de Clasificación Basados en Puntos de División

	Cat	Dog
Yes	0	8
No	8	3

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Cheese	Yes o No	0.04
Barks	Yes o No	0.26



Random Forests: Ejemplo

Datos de Entrenamiento

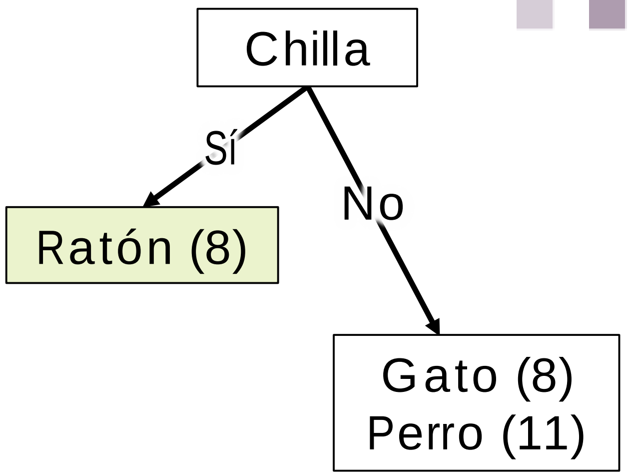
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	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

Resultados de Clasificación Basados en Puntos de División

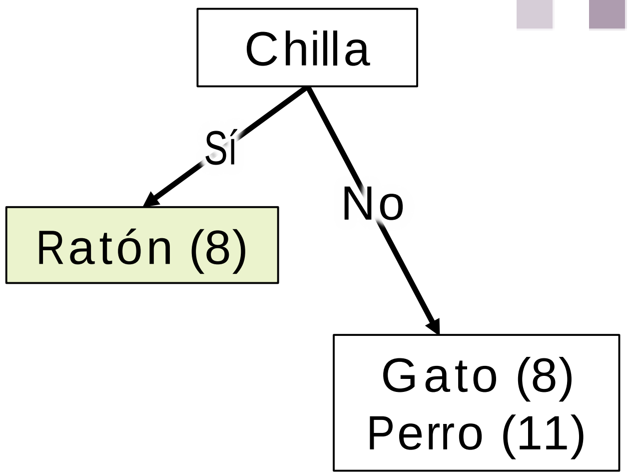
	Cat	Dog
Yes	8	10
No	0	1

Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Cheese	Yes o No	0.04
Barks	Yes o No	0.26
Collar	Yes o No	0.02



Random Forests: Ejemplo



Datos de Entrenamiento

	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

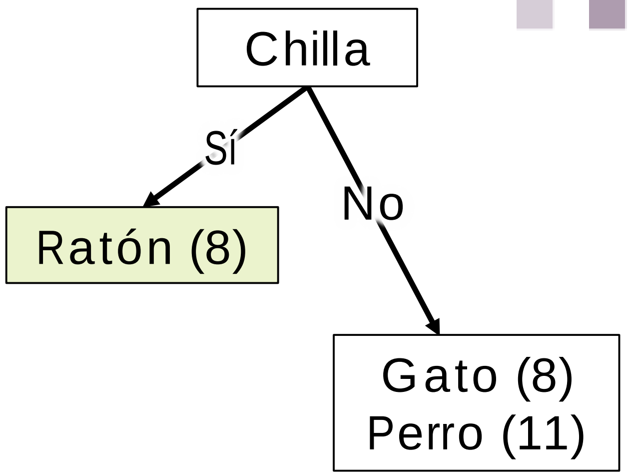
Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Cheese	Yes o No	0.04
Barks	Yes o No	0.26
Collar	Yes o No	0.02

Random Forests: Ejemplo

Datos de Entrenamiento

	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



Ganancia de Gini para identificar la división óptima

Variable	Puntos de División Evaluados	Gini
Cheese	Yes o No	0.04
Barks	Yes o No	0.26
Collar	Yes o No	0.02

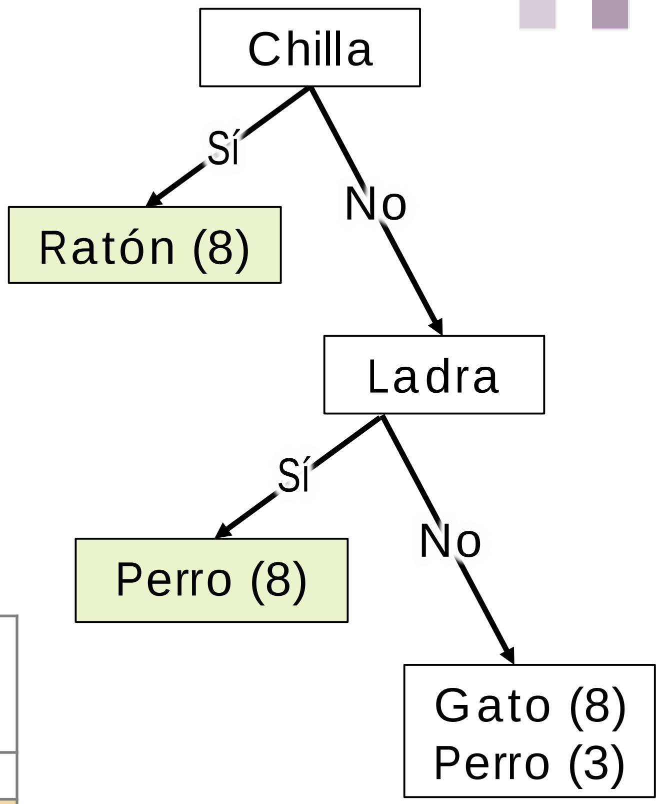
Random Forests: Ejemplo

Datos de Entrenamiento

	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

Ganancia de Gini para identificar la división óptima

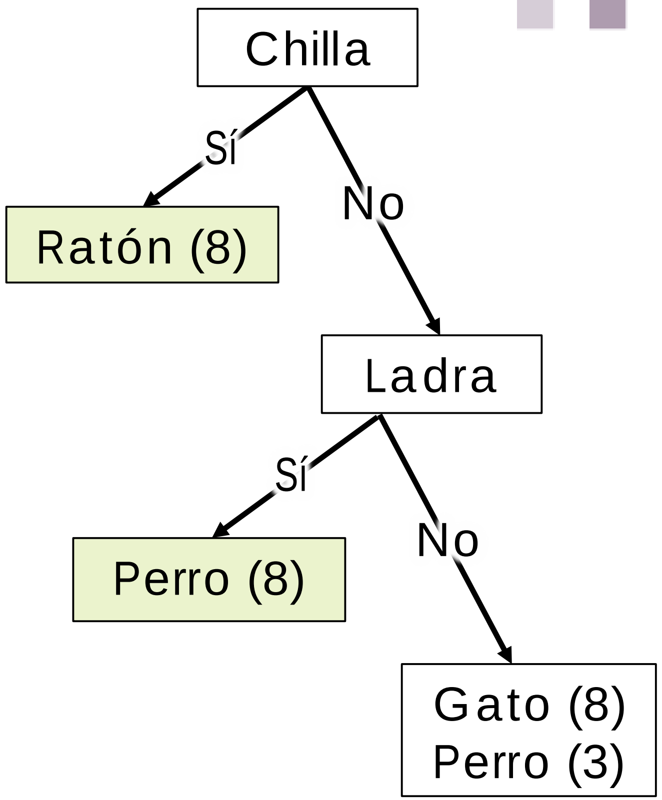
Variable	Puntos de División Evaluados	Gini
Cheese	Yes o No	0.04
Barks	Yes o No	0.26
Collar	Yes o No	0.02



Random Forests: Ejemplo

Datos de Entrenamiento

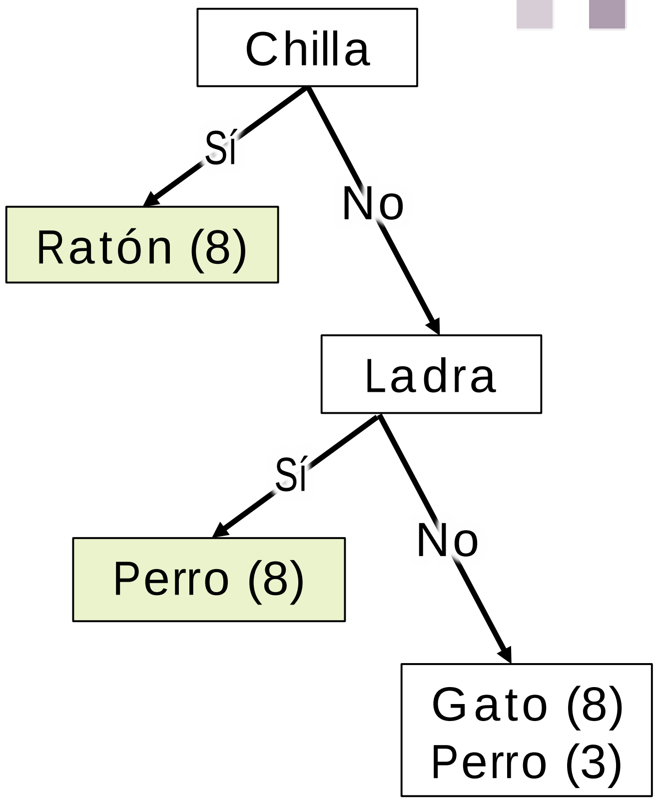
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	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: Ejemplo

Datos de Entrenamiento

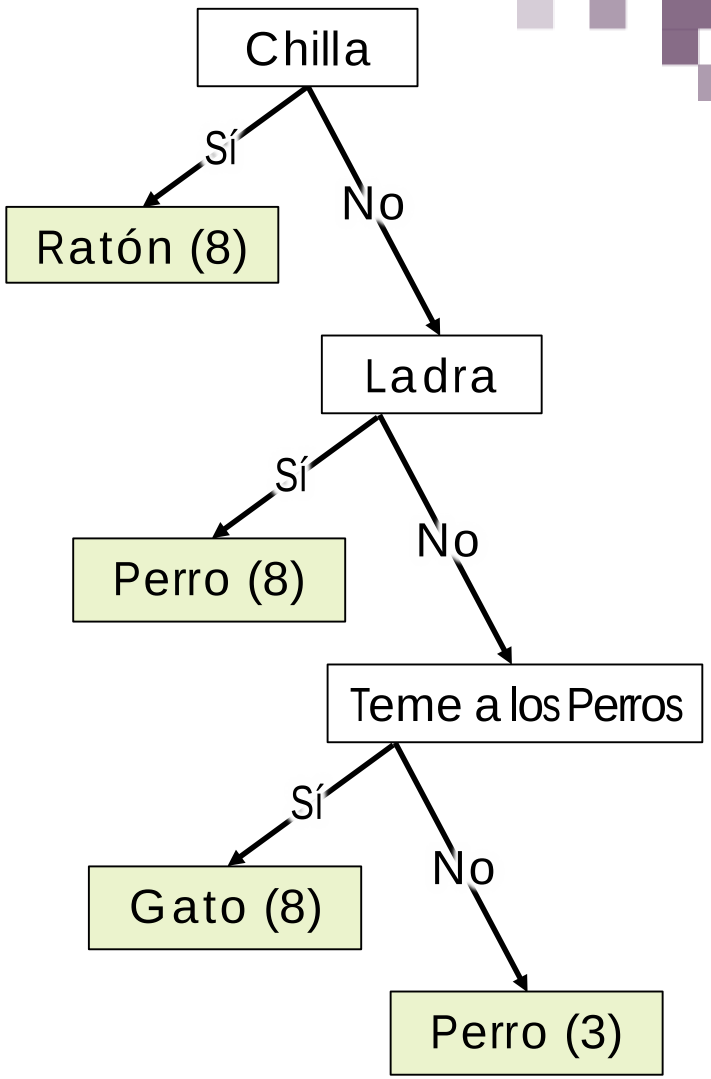
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
12	Cat	No	Yes	No	No	Yes	Yes	0.4	0.15	No
13	Cat	No	Yes	No	Yes	Yes	Yes	0.37	0.09	Yes
14	Cat	No	Yes	No	Yes	Yes	Yes	0.36	0.17	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	Yes	No	Yes	Yes	Yes	0.3	0.16	Yes
16	Cat	No	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: Ejemplo

Datos de Entrenamiento

	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

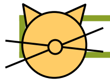


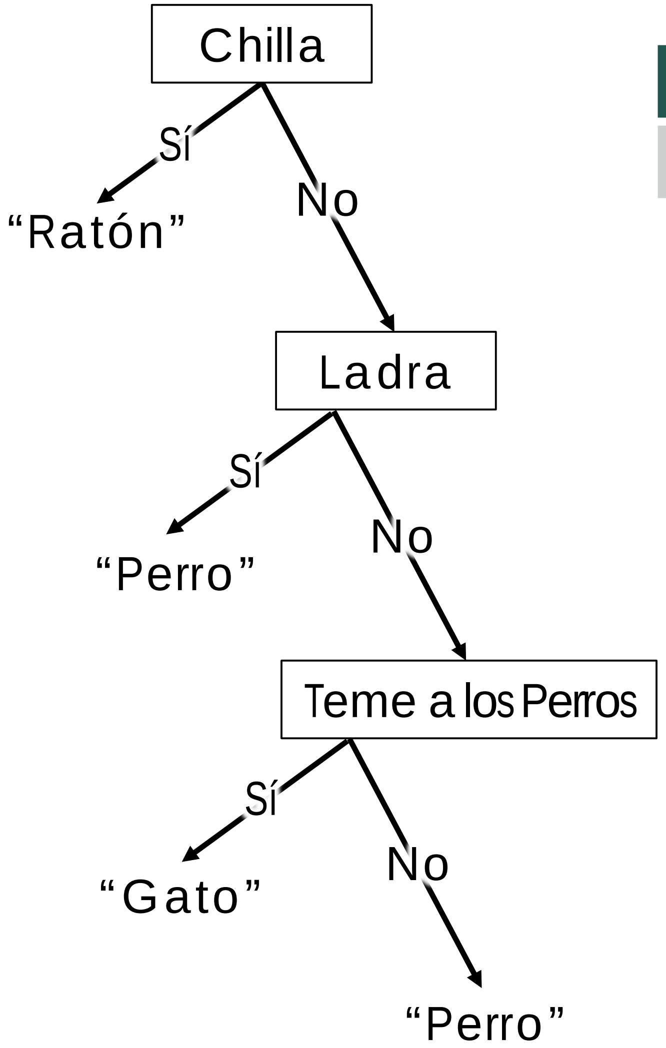
Random Forests: Ejemplo

3. Tome la predicción promediada de todos los árboles (**ntree**) para determinar la clasificación final

Random Forests: Ejemplo

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

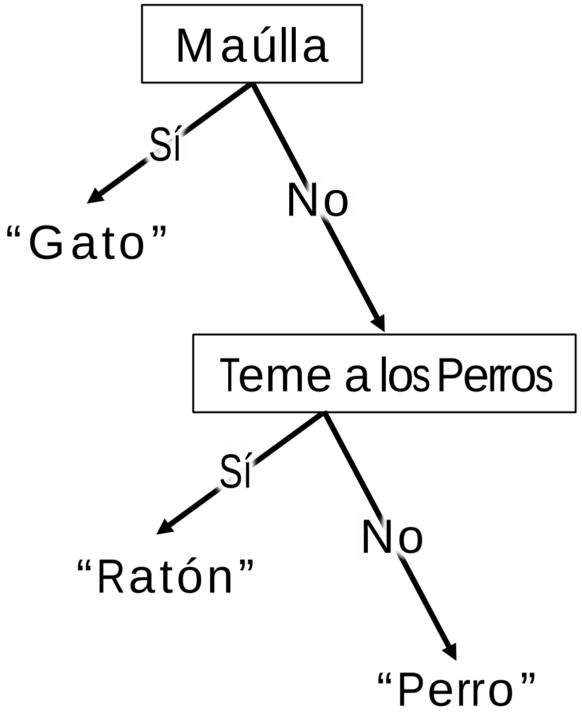


Árbol	Voto
1	Perro

Random Forests: Ejemplo

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

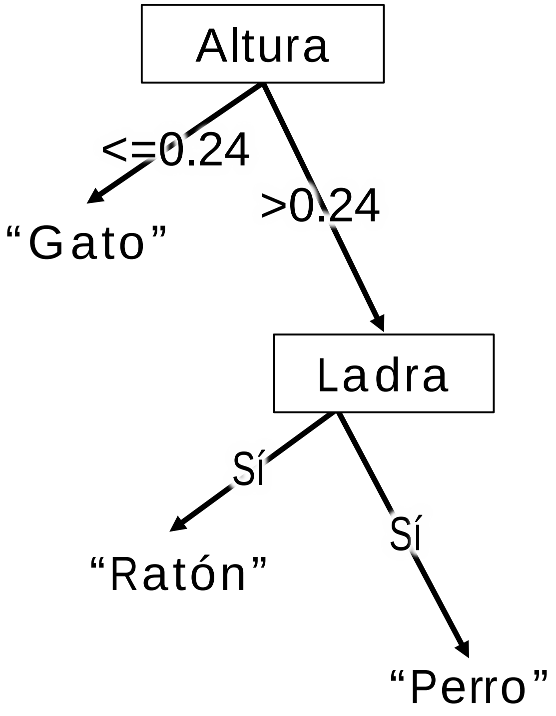


Árbol	Voto
1	Perro
2	Gato

Random Forests: Ejemplo

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



Árbol	Voto
1	Perro
2	Gato
3	Gato

Random Forests: Ejemplo

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

Árbol	Voto
1	Perro
2	Gato
3	Gato
4	Ratón
5	Gato
6	Gato
7	Gato
8	Gato
9	Perro

Mayoría = Gato

Clasificación Final = Gato

Random Forests: Ejemplo

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

Árbol	Voto
1	Perro
2	Gato
3	Gato
4	Ratón
5	Gato
6	Gato
7	Gato
8	Gato
9	Gato

Probabilidad:
$= {}^{2}/_{3}$
=67%

Random Forests: Ejemplo

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

Árbol	Voto
1	Perro
2	Gato
3	Gato
4	Ratón
5	Gato
6	Gato
7	Gato
8	Gato
9	Gato

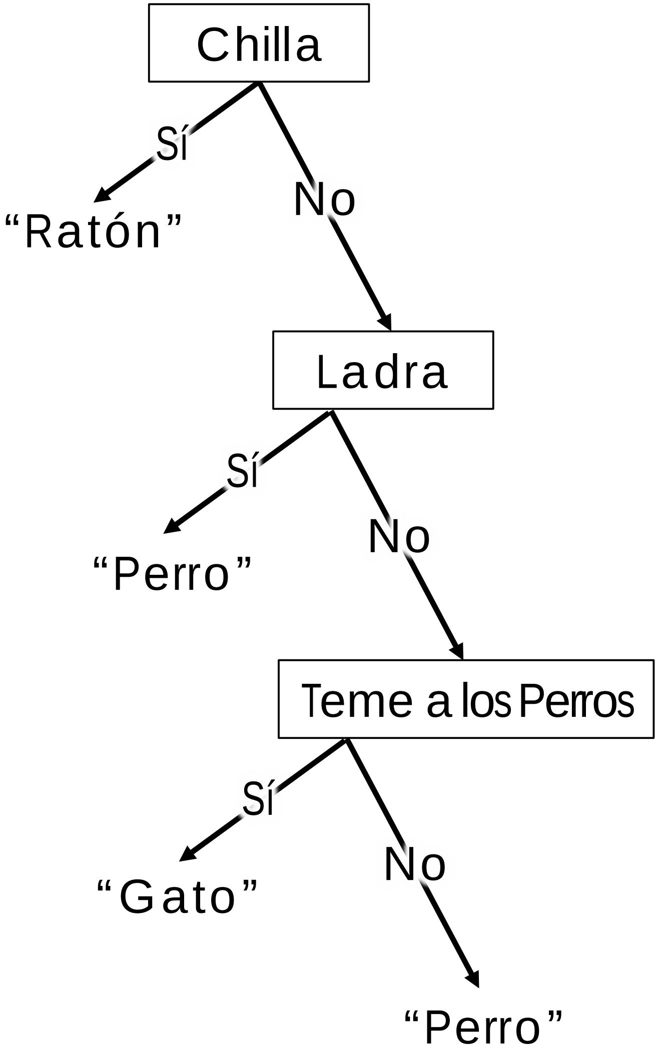
Random Forests: Ejemplo

- Pase las muestras “Out of Bag” por el árbol y calcule el error Out of Bag (para un árbol particular)

Random Forests: Ejemplo

	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
[illegible]	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
[illegible]	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
[illegible]	Mouse	No	No	Yes	No	No	Yes	0.16	0.08	Yes
[illegible]	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

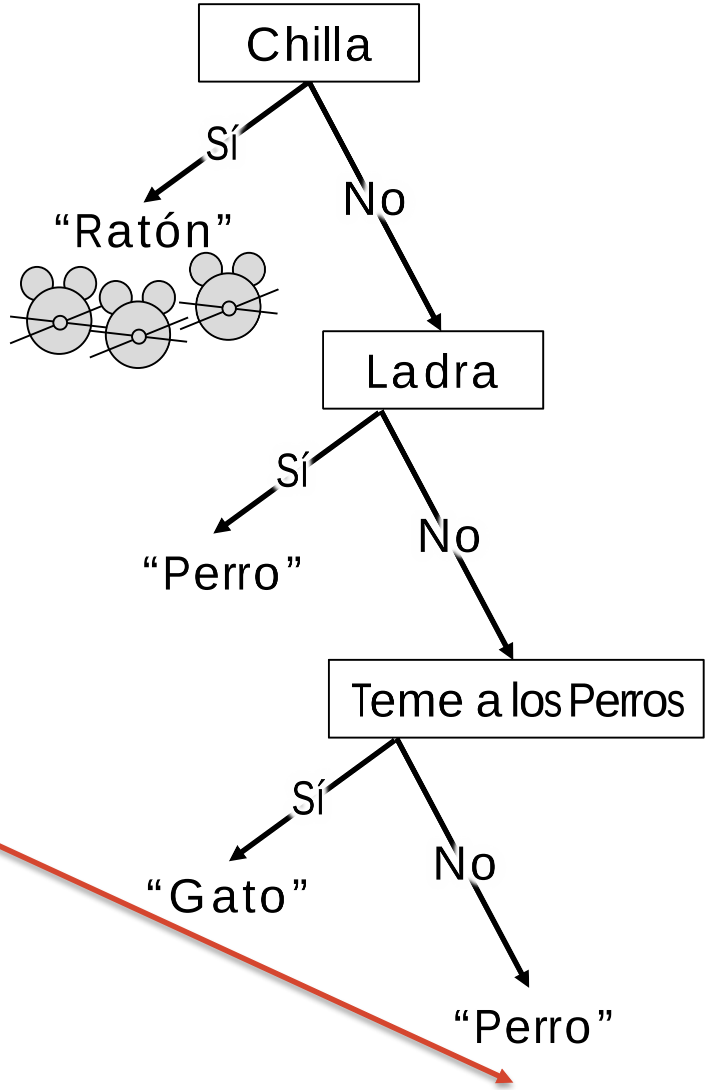
Restantes para Evaluación de Precisión Interna ≈$^1/_3$



Random Forests: Ejemplo

Restantes para Evaluación de Precisión Interna $\approx {}^{1}/_{3}$

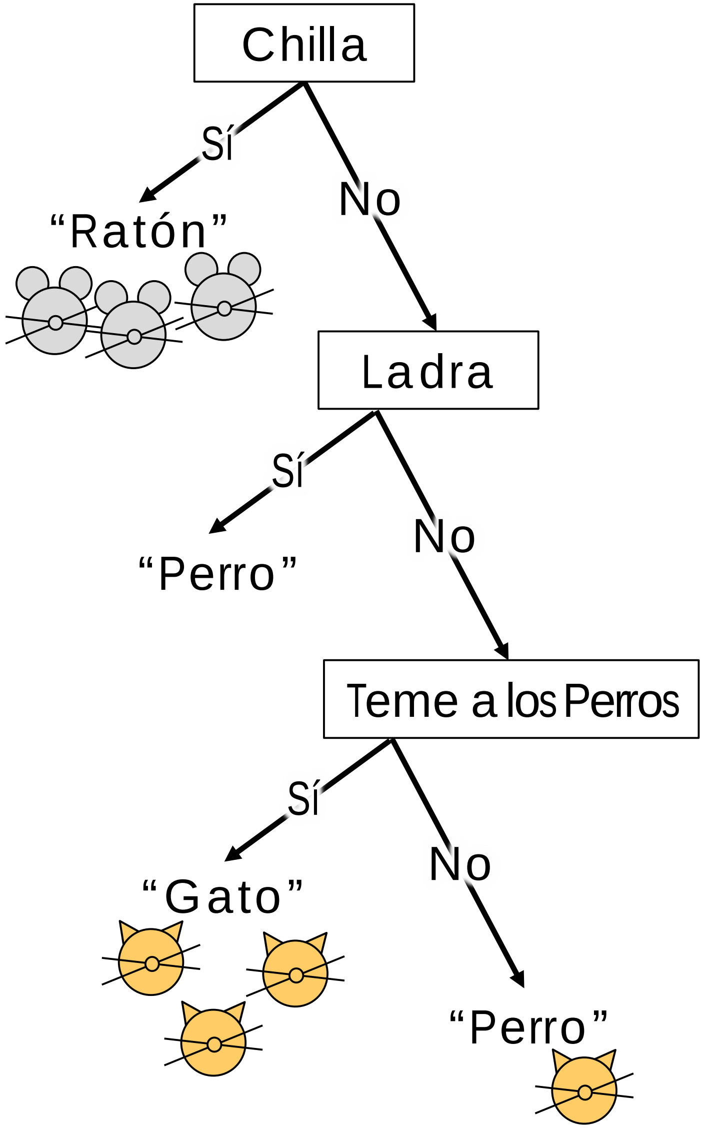
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	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
[illegible]	Cat	No	Yes	No	Yes	Yes	Yes	0.31	0.19	Yes
[illegible]	Cat	No	Yes	No	Yes	No	Yes	0.38	0.20	Yes
[illegible]	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
[illegible]	Cat	No	No	No	Yes	No	No	0.32	0.22	Yes
[illegible]	Cat	No	Yes	No	Yes	Yes	Yes	0.30	0.16	Yes
[illegible]	Cat	No	Yes	No	Yes	Yes	Yes	0.35	0.24	Yes
[illegible]	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: Ejemplo

Restantes para Evaluación de Precisión Interna $\approx {}^{1}/_{3}$

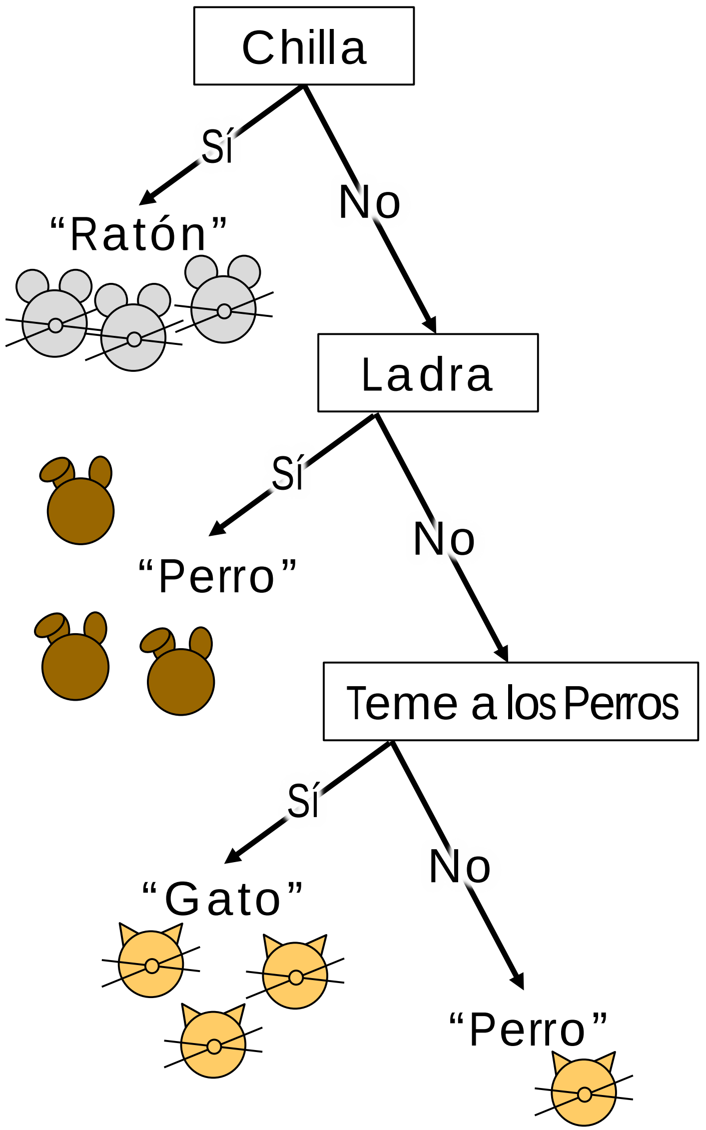
	Species	Barks	Pet	Squeaks	Meows	Collar	Afraid of Dogs	Length	Height	Likes Cheese
1	Mouse	No	Yes	Yes	No	No	Yes	0.14	0.07	Yes
2	Mouse	No	No	Yes	No	No	Yes	0.18	0.09	Yes
3	Mouse	No	No	Yes	No	No	Yes	0.13	0.06	Yes
4	Mouse	No	Yes	Yes	No	No	Yes	0.13	0.06	Yes
5	Mouse	No	No	Yes	No	No	Yes	0.13	0.07	Yes
6	Mouse	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
[illegible]	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
[illegible]	Dog	Yes	Yes	No	No	Yes	No	0.53	0.27	Yes
[illegible]	Dog	Yes	Yes	No	No	Yes	No	0.37	0.16	Yes
[illegible]	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: Ejemplo

Restantes para Evaluación de Precisión Interna ≈$^{1}/_{3}$

	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



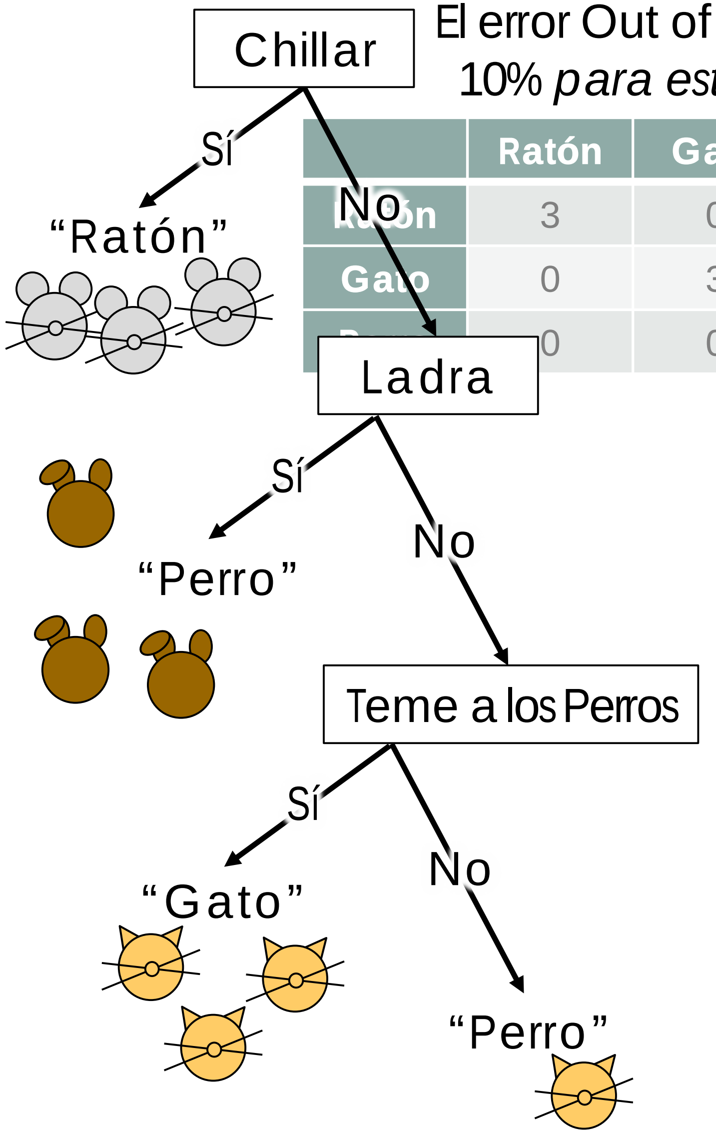
Random Forests: Ejemplo

Restantes para Evaluación de Precisión Interna $\approx {}^{1}/_{3}$

	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

El error Out of Bag es del 10% *para este árbol…*

	Ratón	Gato	Perro
Ratón	3	0	0
Gato	0	3	1
Perro	0	0	3

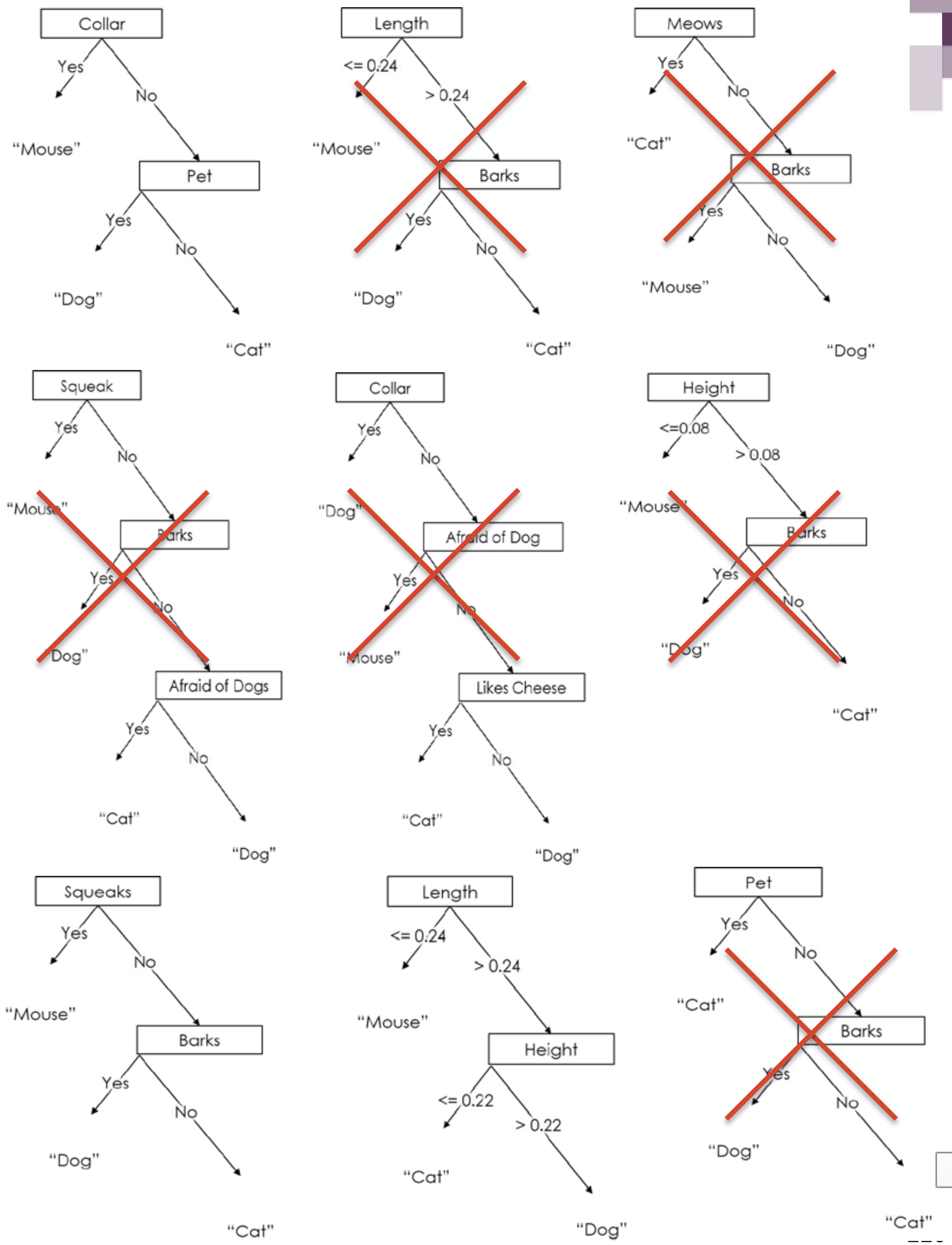


Random Forests: Ejemplo

- Calcule el error Out of Bag general

	nTree (=9)									Majority Vote		
Species	1	2	3	4	5	6	7	8	9	Mouse	Cat	Dog
1 Mouse												

	Ratón	Gato	Perro	Error
Mouse	9	1	0	10%
Gato				
Perro				



Random Forests: Ejemplo

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

5. Calculando el Mean Decrease in Accuracy (**MDA** o Disminución Media de la Exactitud)
   - Los valores son permutados en todos los árboles
   - MDA: disminución media de la exactitud; normalice mediante la desviación del estándar

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

Datos de Longitud
Aleatoriamente Permutados

- Largo
  - <= 0.24 → "Ratón"
  - > 0.24 → Ladra
    - Sí → "Perro"
    - No → "Gato"

Exactitud Out of Bag Original- 90%

	Ratón	Gato	Perro
Ratón	3	0	0
Gato	0	3	1
Perro	0	0	3

Exactitud Out of Bag Después de Permutar- 40%

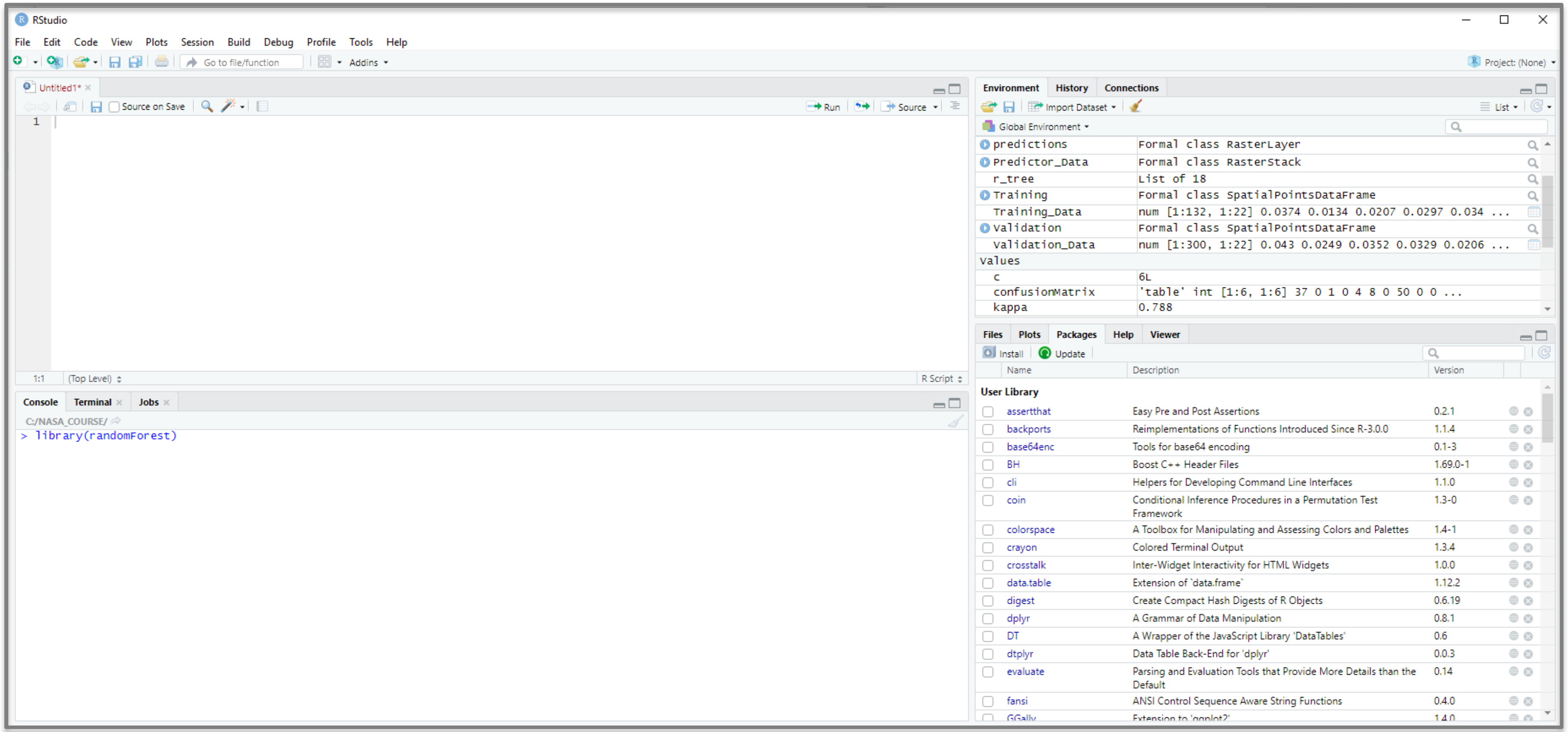
	Ratón	Gato	Perro
Ratón	0	3	0
Gato	1	3	0
Perro	2	0	1

Random Forests: Demostración Práctica de Clasificación de Cultivos

1. Instalar la última versión de R y RStudio
   – https://cran.r-project.org/bin/windows/base/
   – https://www.rstudio.com/products/rstudio/download/
2. Abrir Rstudio

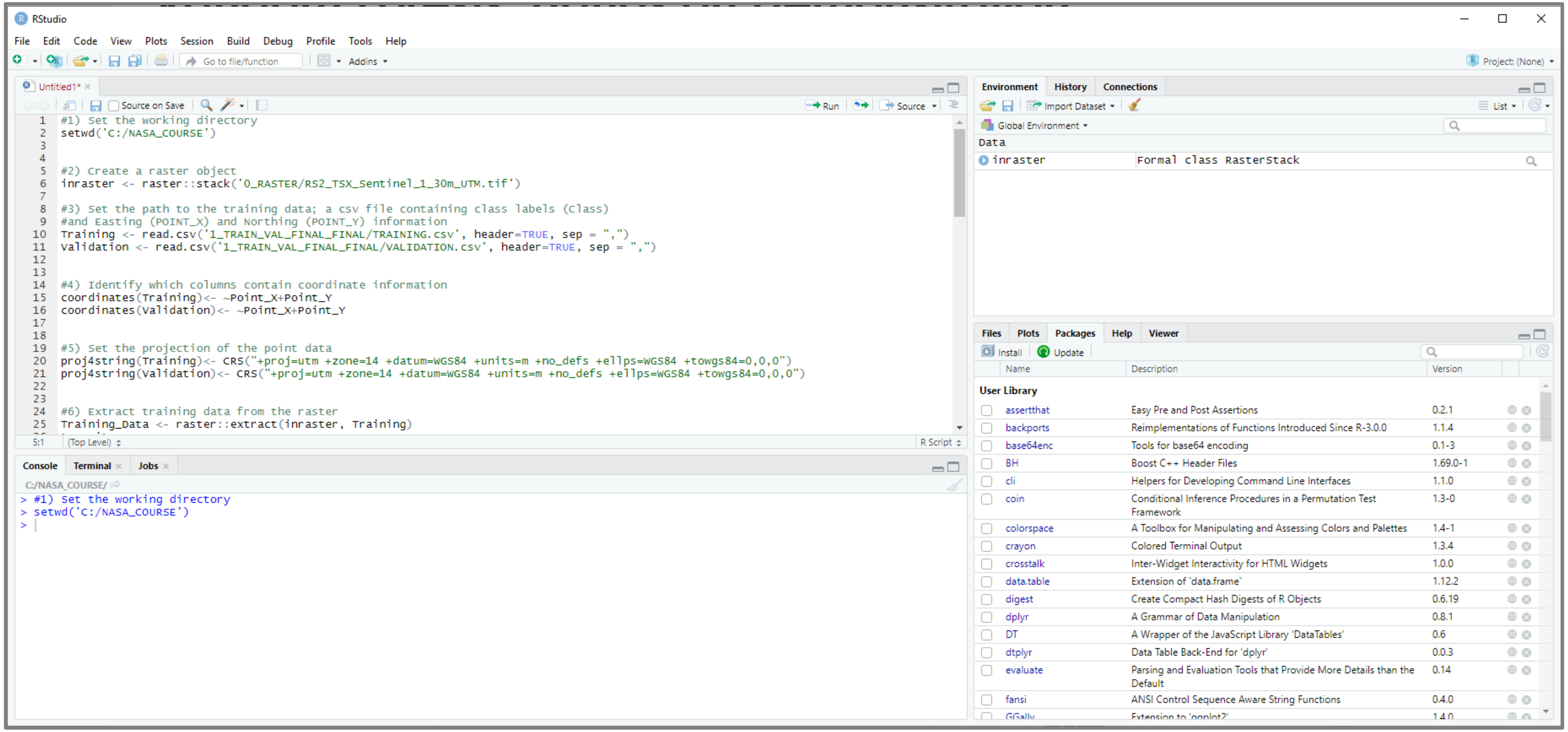
Random Forests: Demostración Práctica

3. Instalar (raster, randomForest, sp, rgdal)
4. Cargar (raster, randomForest, sp, rgdal)



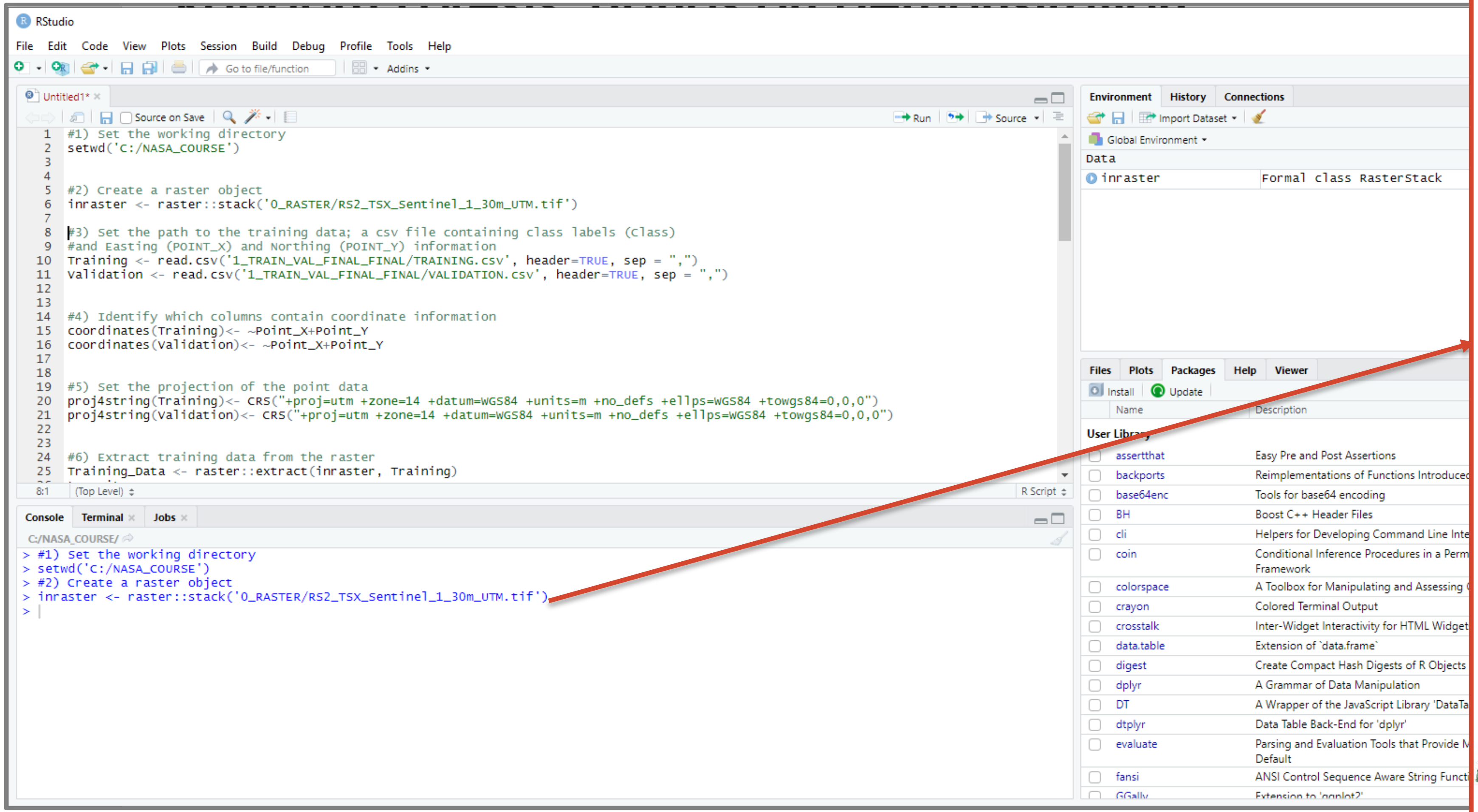
Random Forests: Demostración Práctica

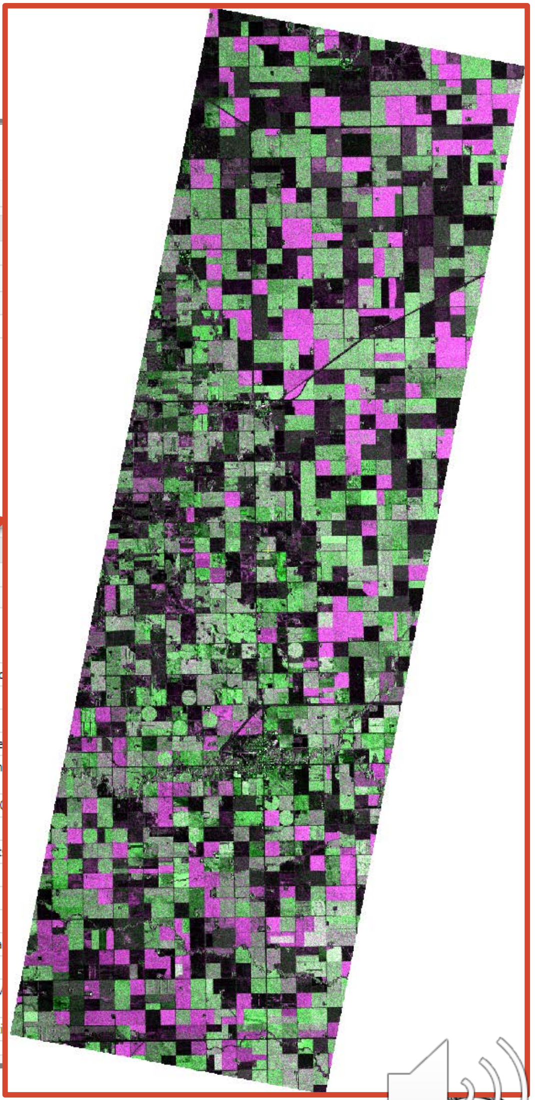
5. Configurar el Directorio Operativo



Random Forests: Demostración Práctica

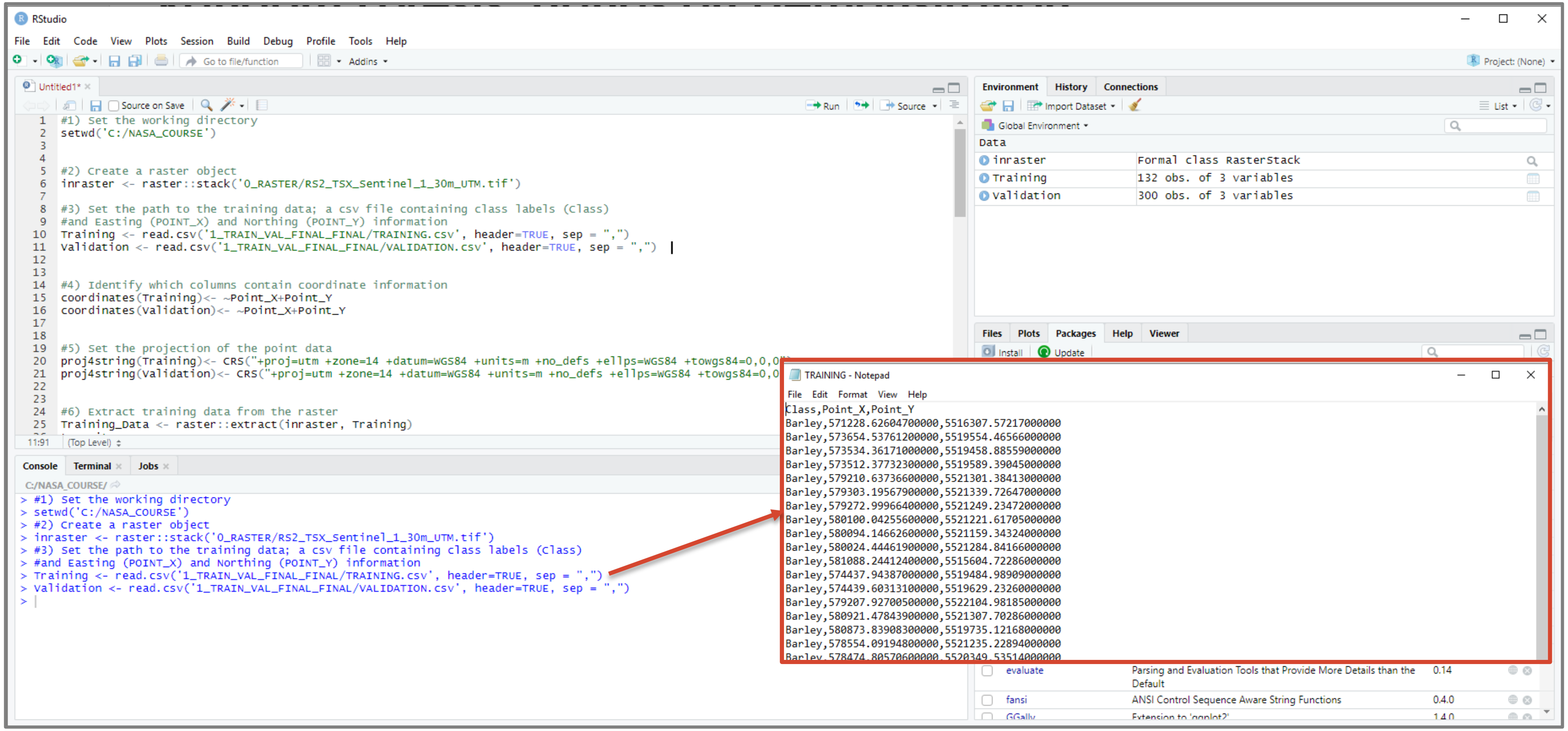
6. Crear un objeto ráster





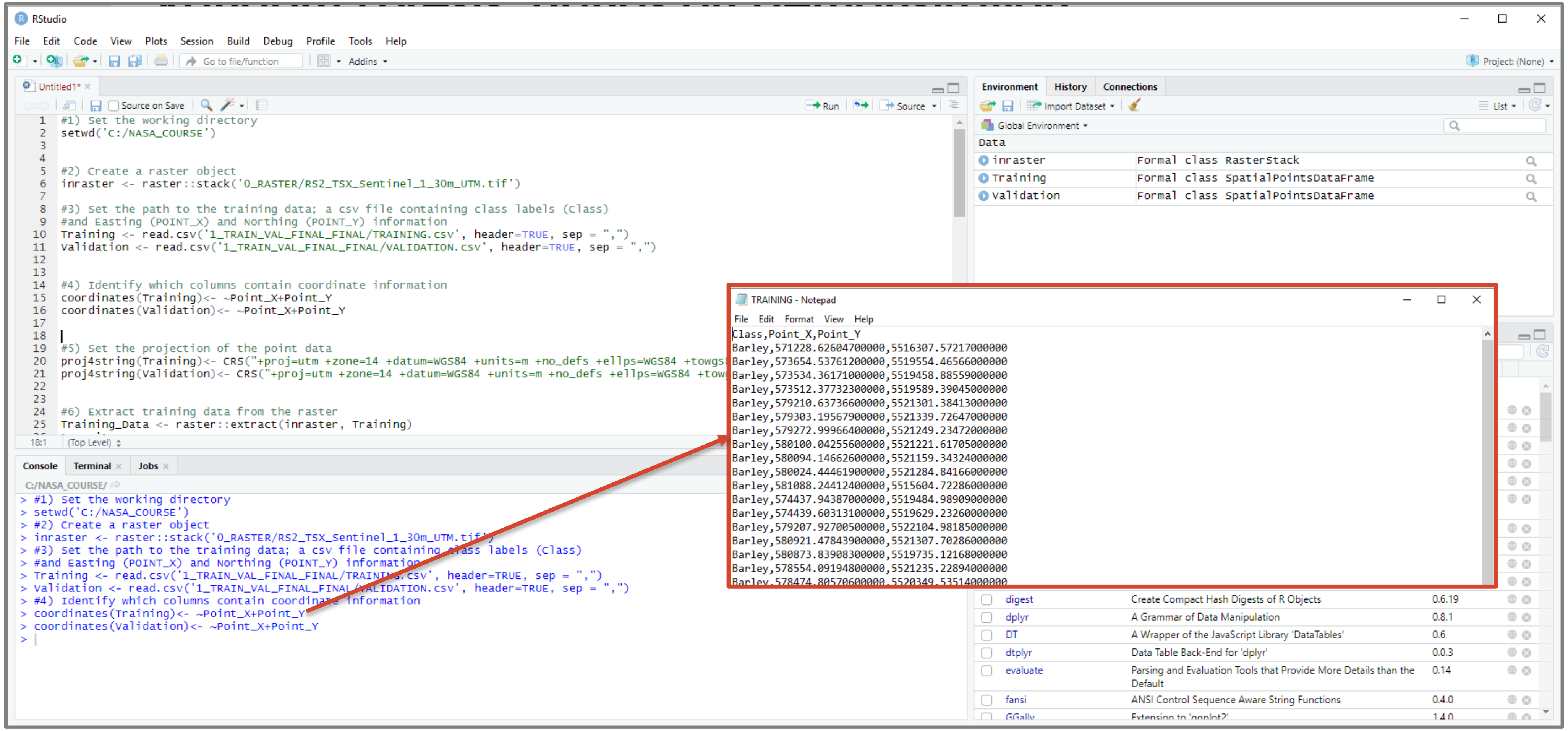
Random Forests: Demostración Práctica

7. Leer datos de entrenamiento y validación (archivos csv)



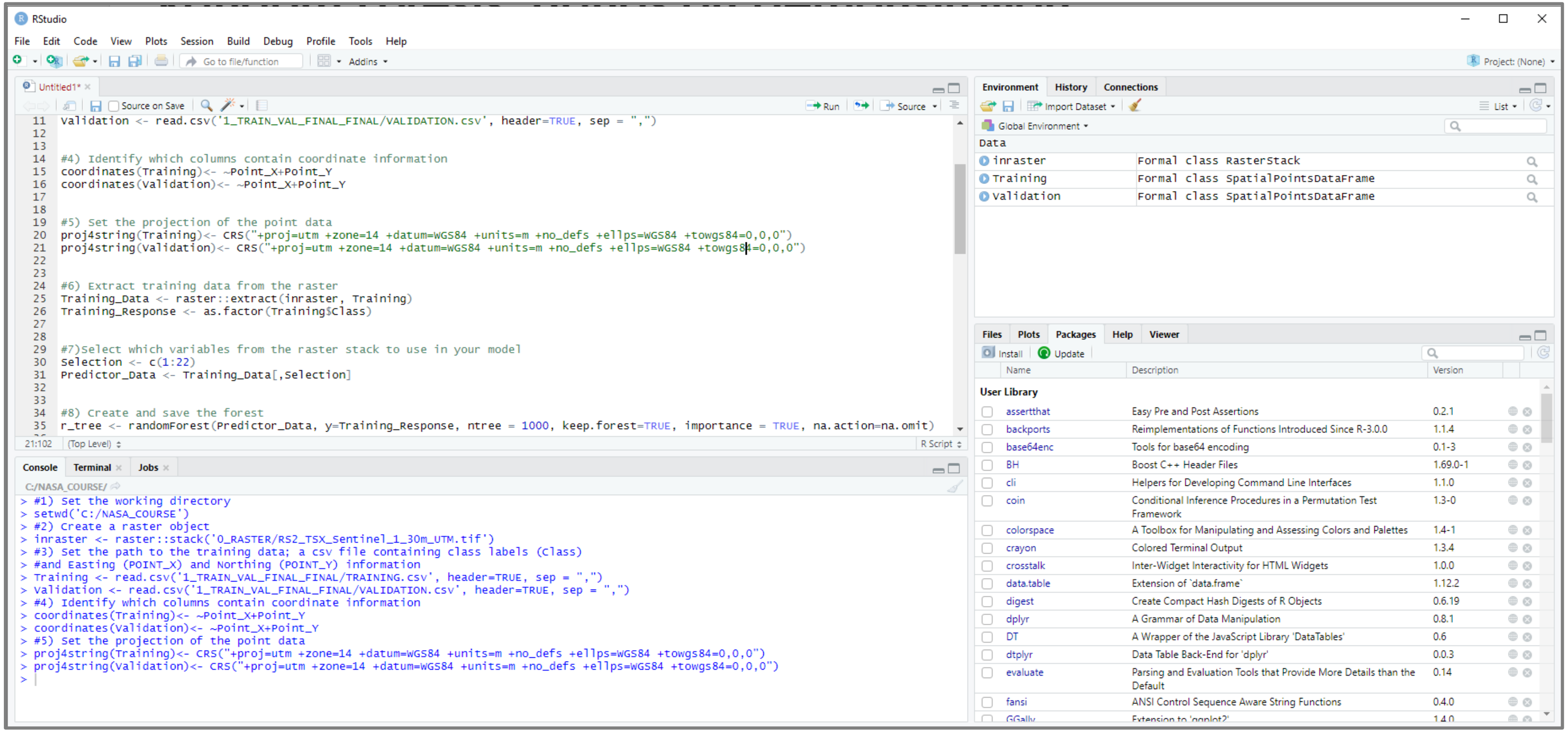
Random Forests: Demostración Práctica

8. Identificar cuáles columnas de los archivos csv contienen datos sobre coordenadas



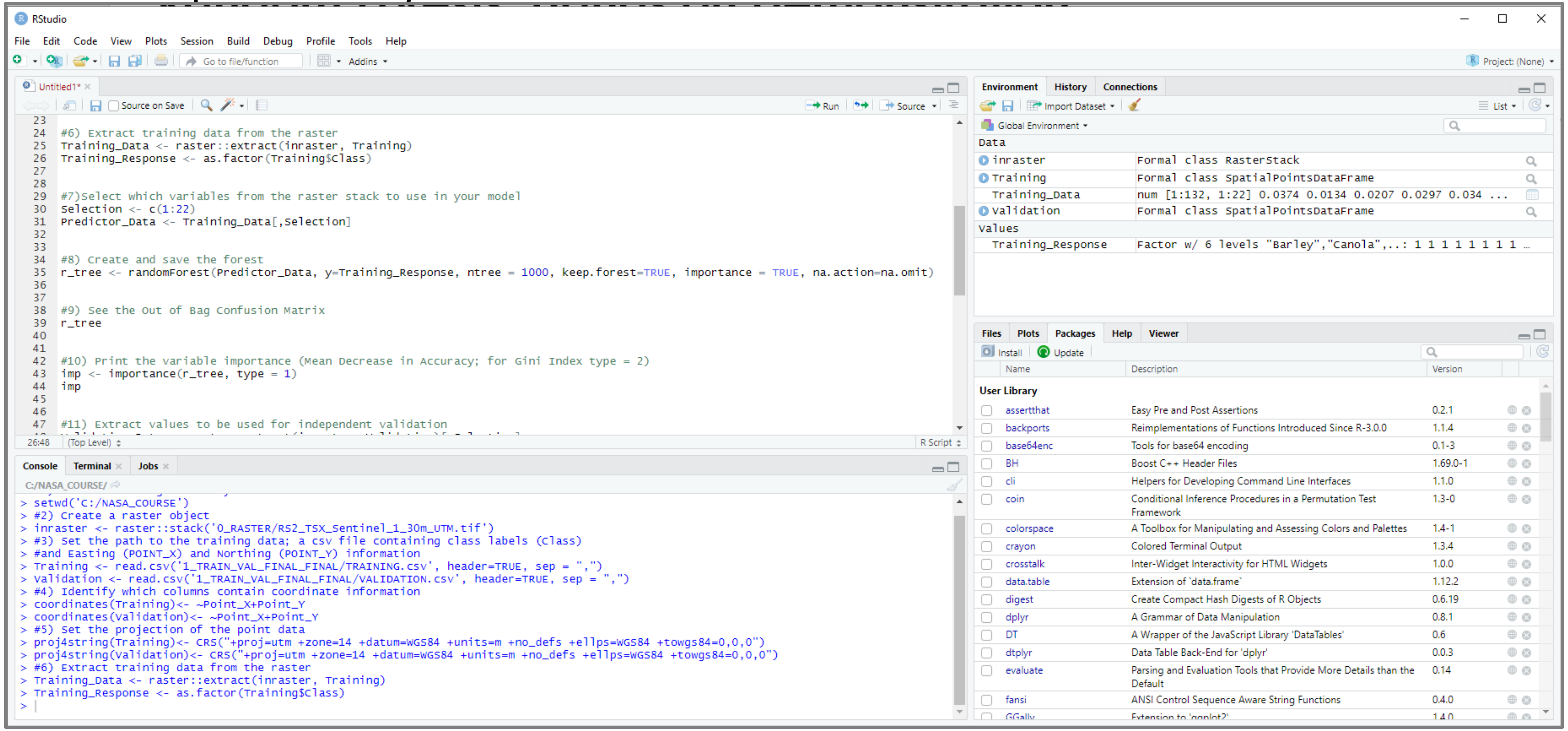
Random Forests: Demostración Práctica

9. Definir la proyección de los datos puntuales en los archivos csv



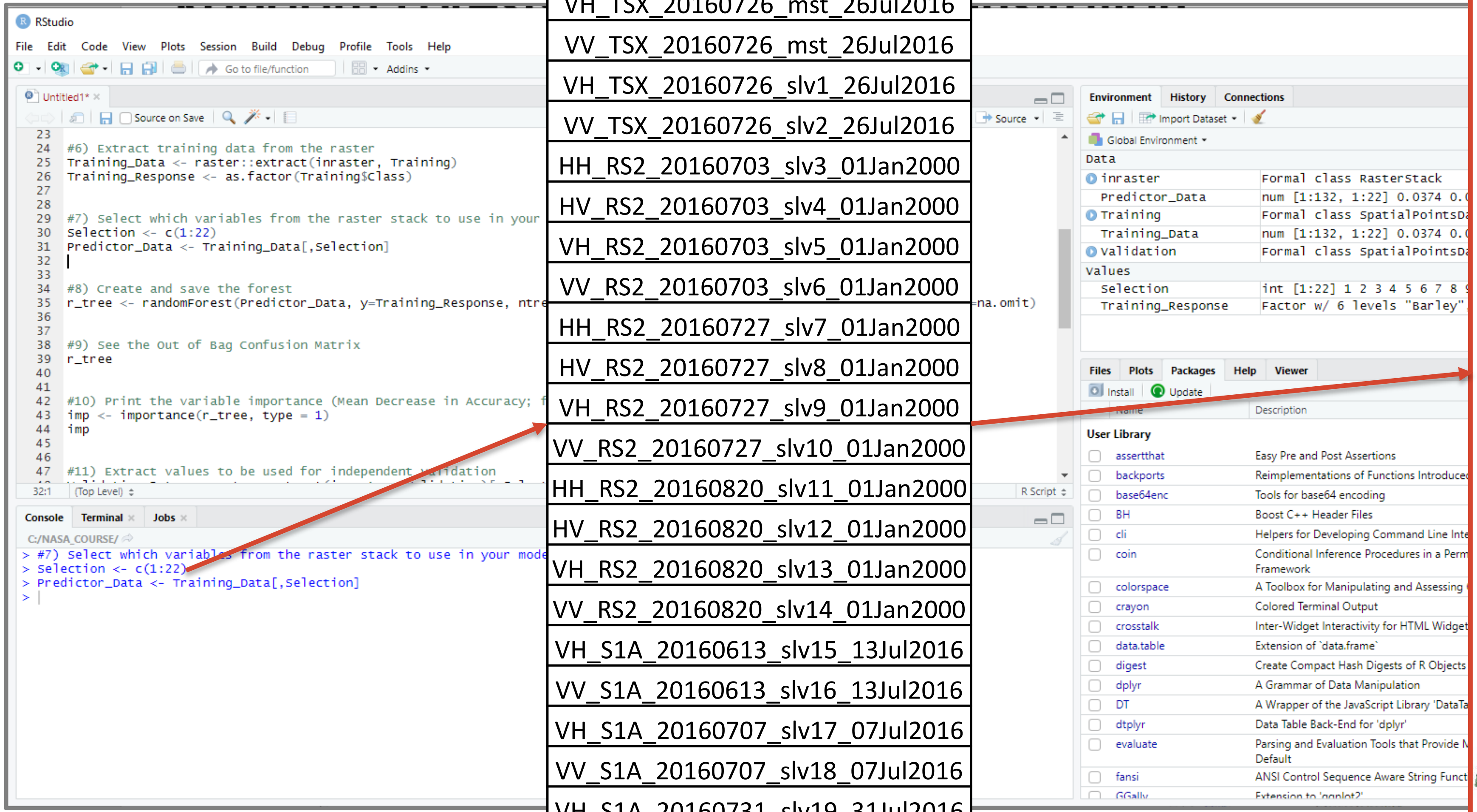
Random Forests: Demostración Práctica

10. Extraer datos de entrenamiento (valores de cada banda de ráster coincidentes con datos puntuales)

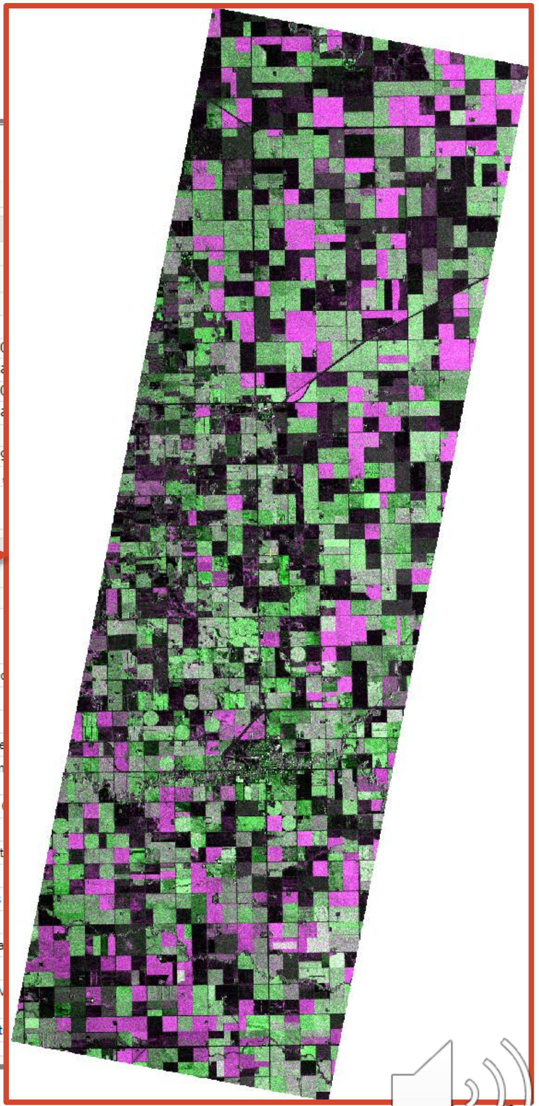


Random Forests: Demostración Práctica

11. Seleccionar variables para utilizar en el modelo

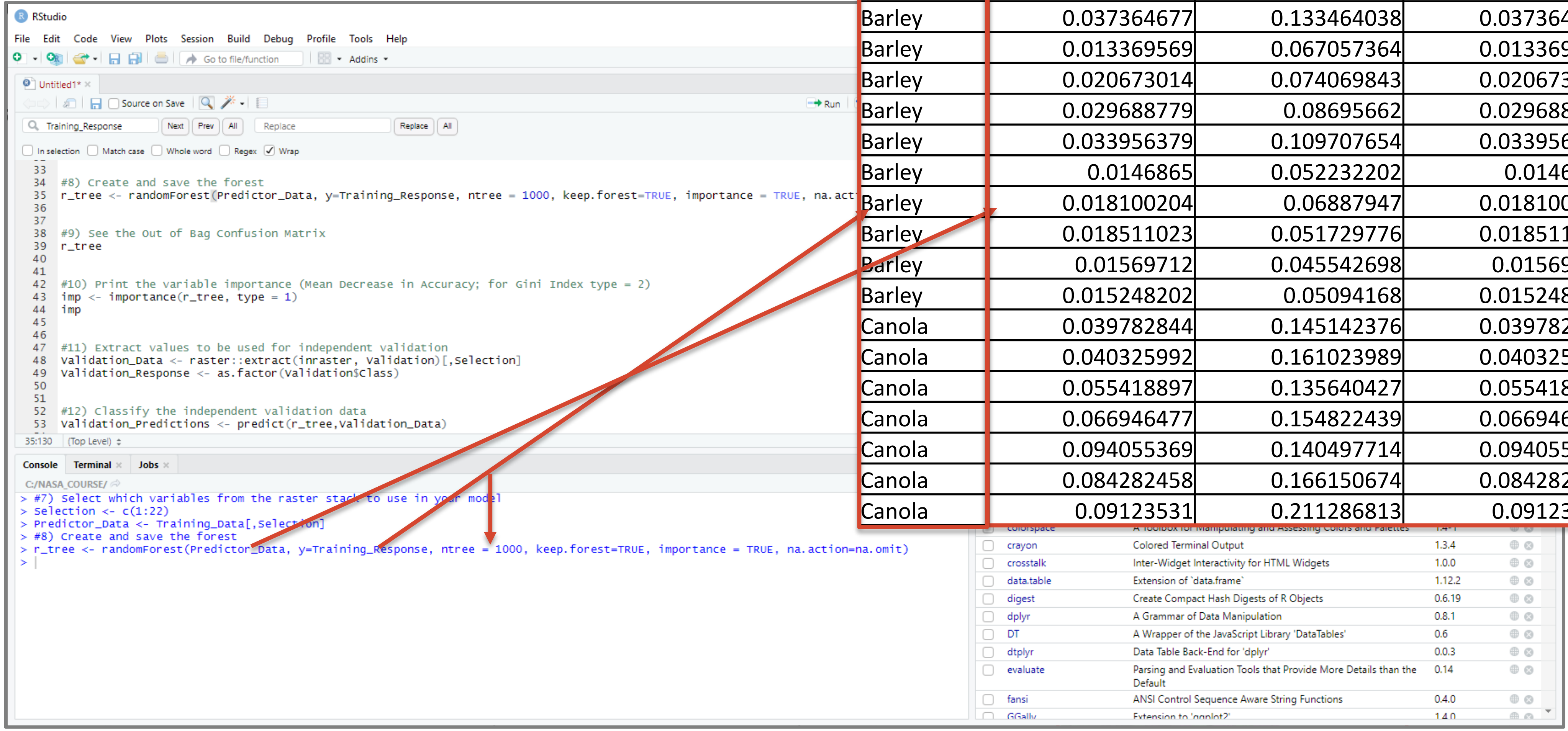


VH_TSX_20160726_mst_26Jul2016
VV_TSX_20160726_mst_26Jul2016
VH_TSX_20160726_slv1_26Jul2016
VV_TSX_20160726_slv2_26Jul2016
HH_RS2_20160703_slv3_01Jan2000
HV_RS2_20160703_slv4_01Jan2000
VH_RS2_20160703_slv5_01Jan2000
VV_RS2_20160703_slv6_01Jan2000
HH_RS2_20160727_slv7_01Jan2000
HV_RS2_20160727_slv8_01Jan2000
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: Demostración Práctica

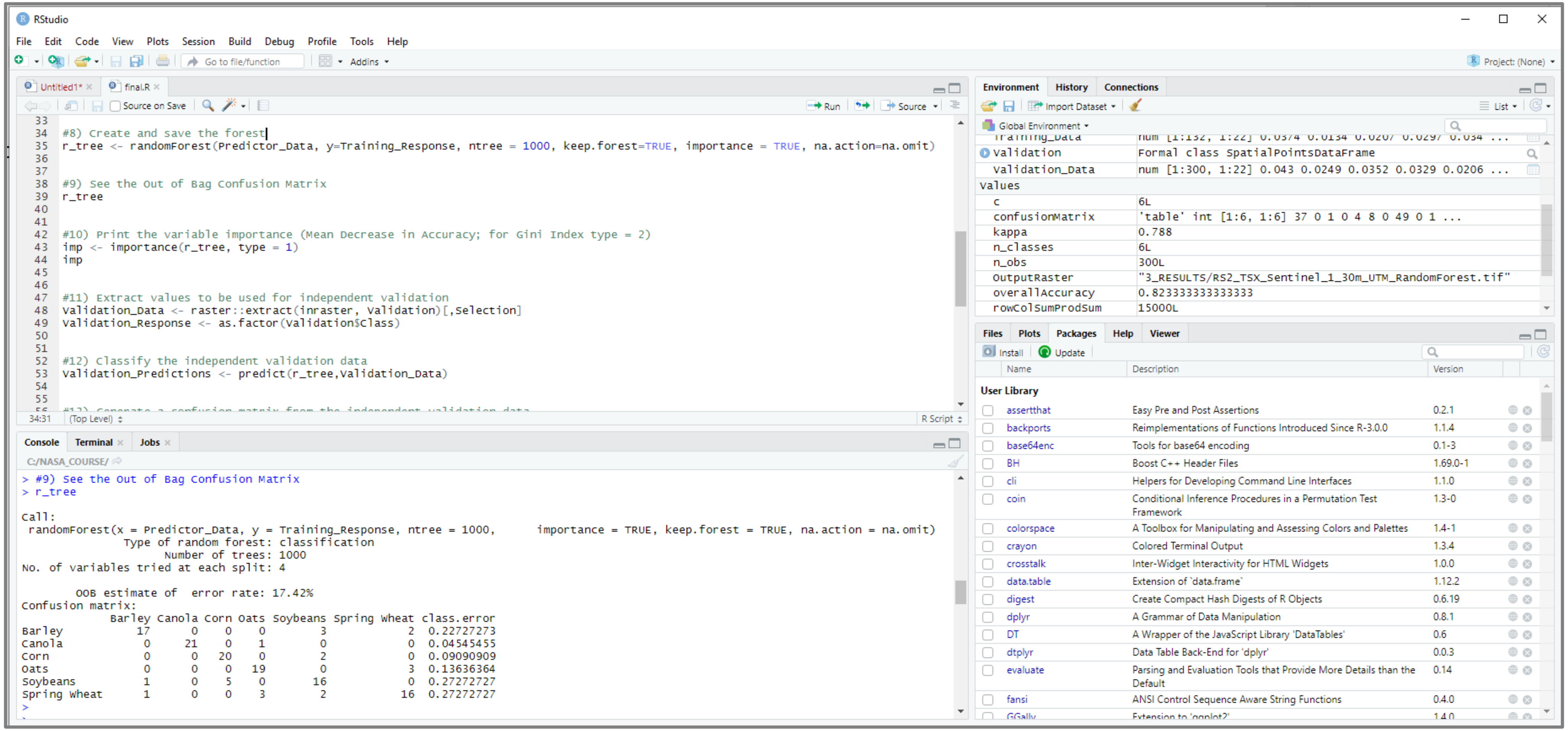
12. Crear su propio modelo Random Forest



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

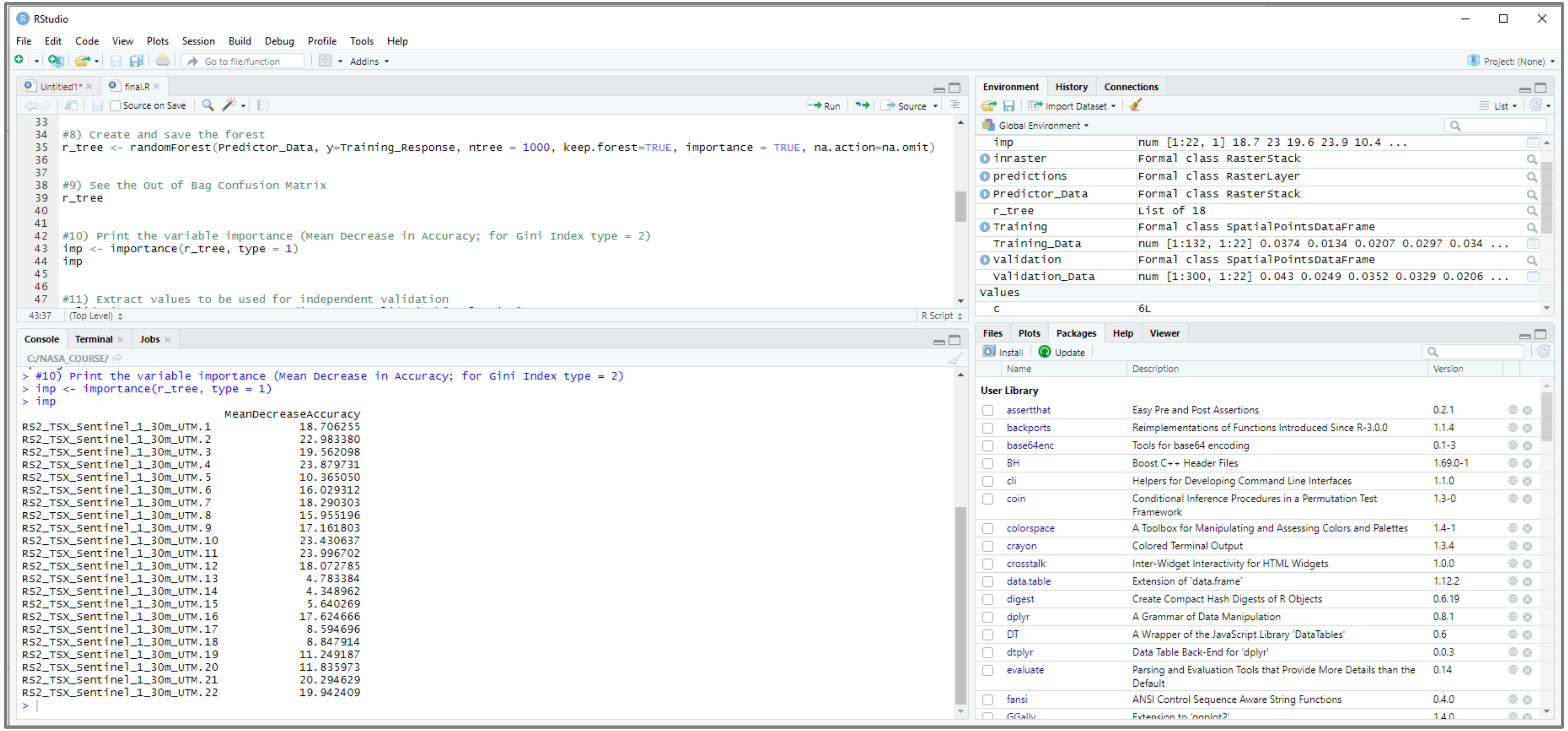
Random Forests: Demostración Práctica

13. Imprimir la matriz de confusión Out of Bag



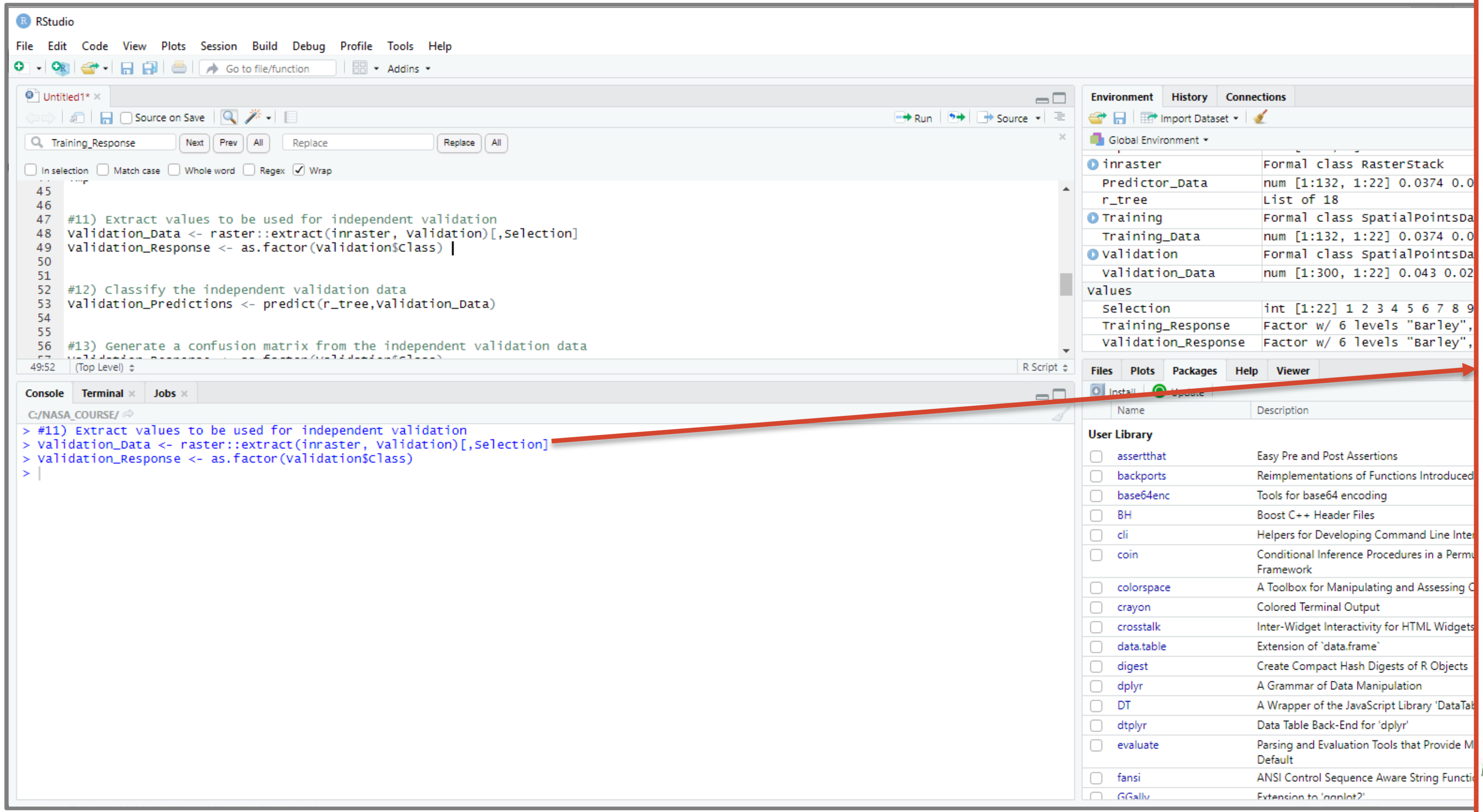
Random Forests: Demostración Práctica

14. Imprimir los valores de importancia variable



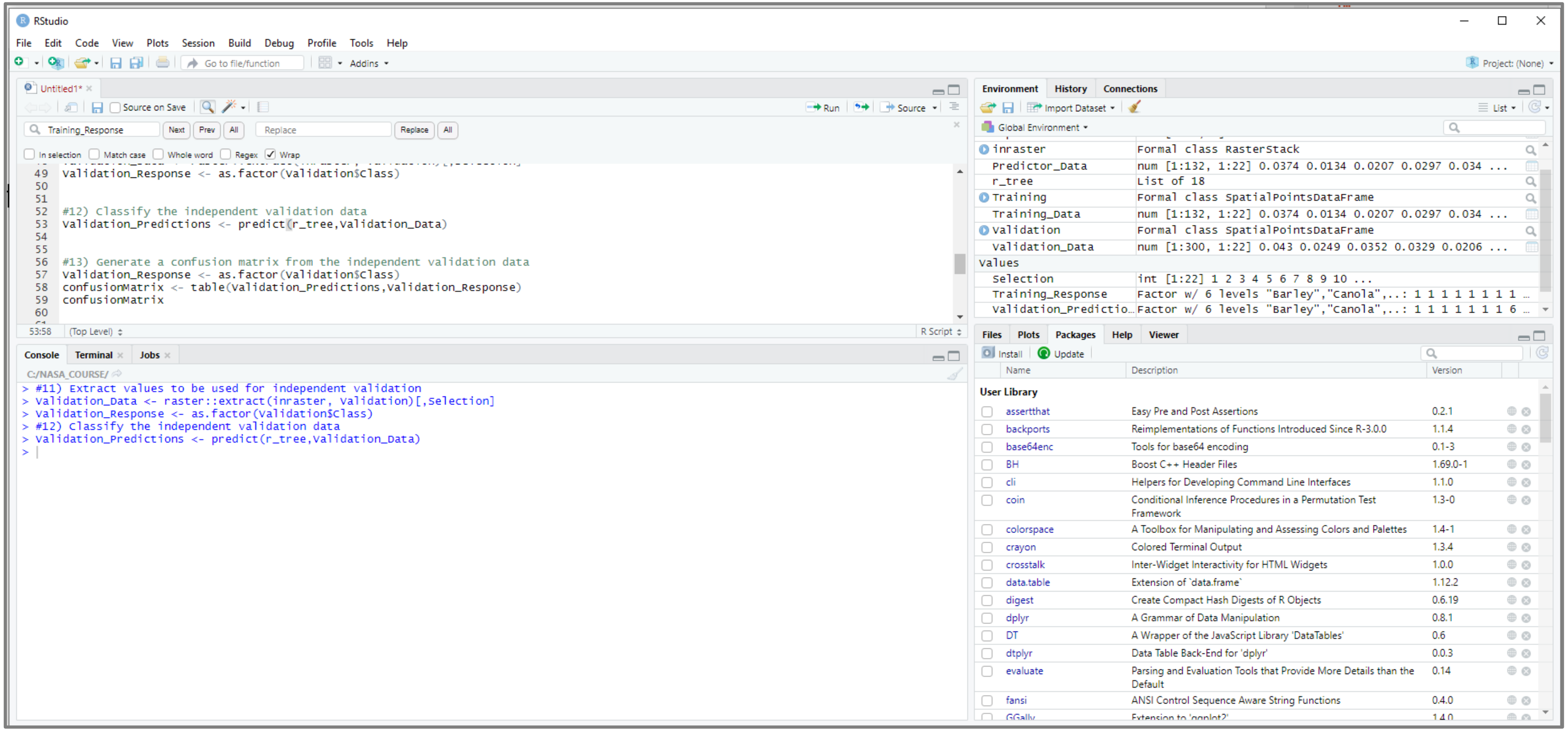
Random Forests: Demostración Práctica

15. Extraer los datos de validación



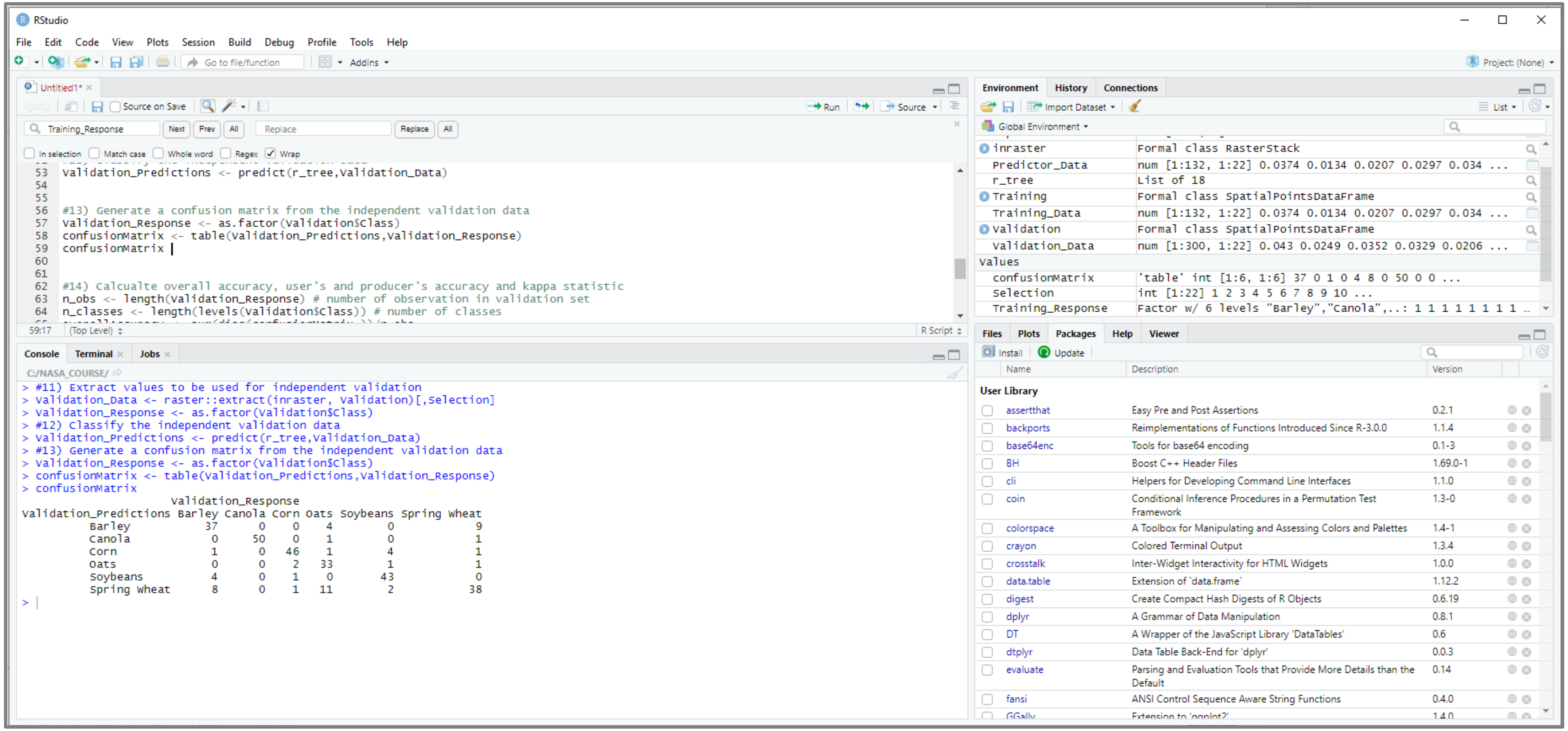
Random Forests: Demostración Práctica

16. Clasificar los datos de validación independientes



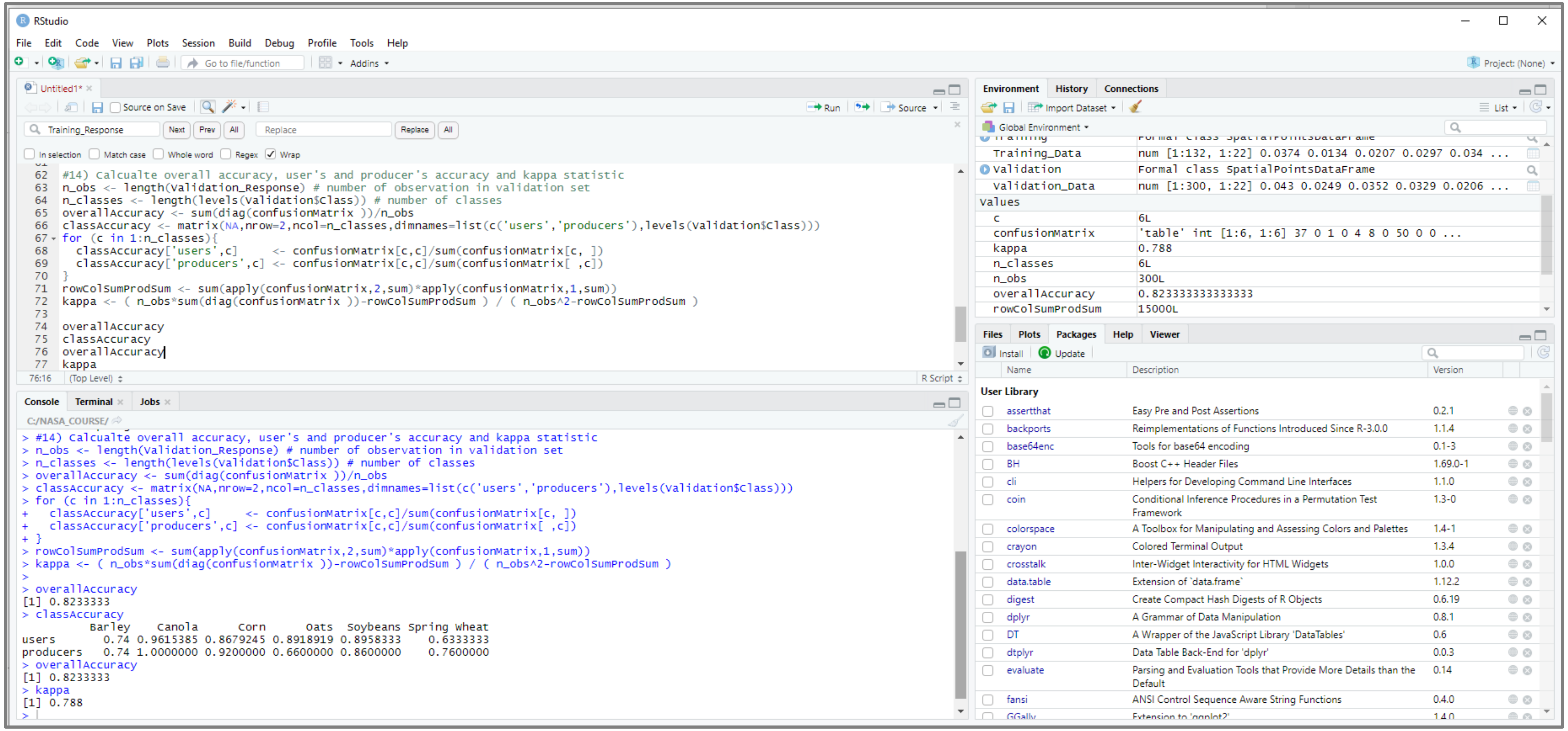
Random Forests: Demostración Práctica

17. Generar una matriz de confusión



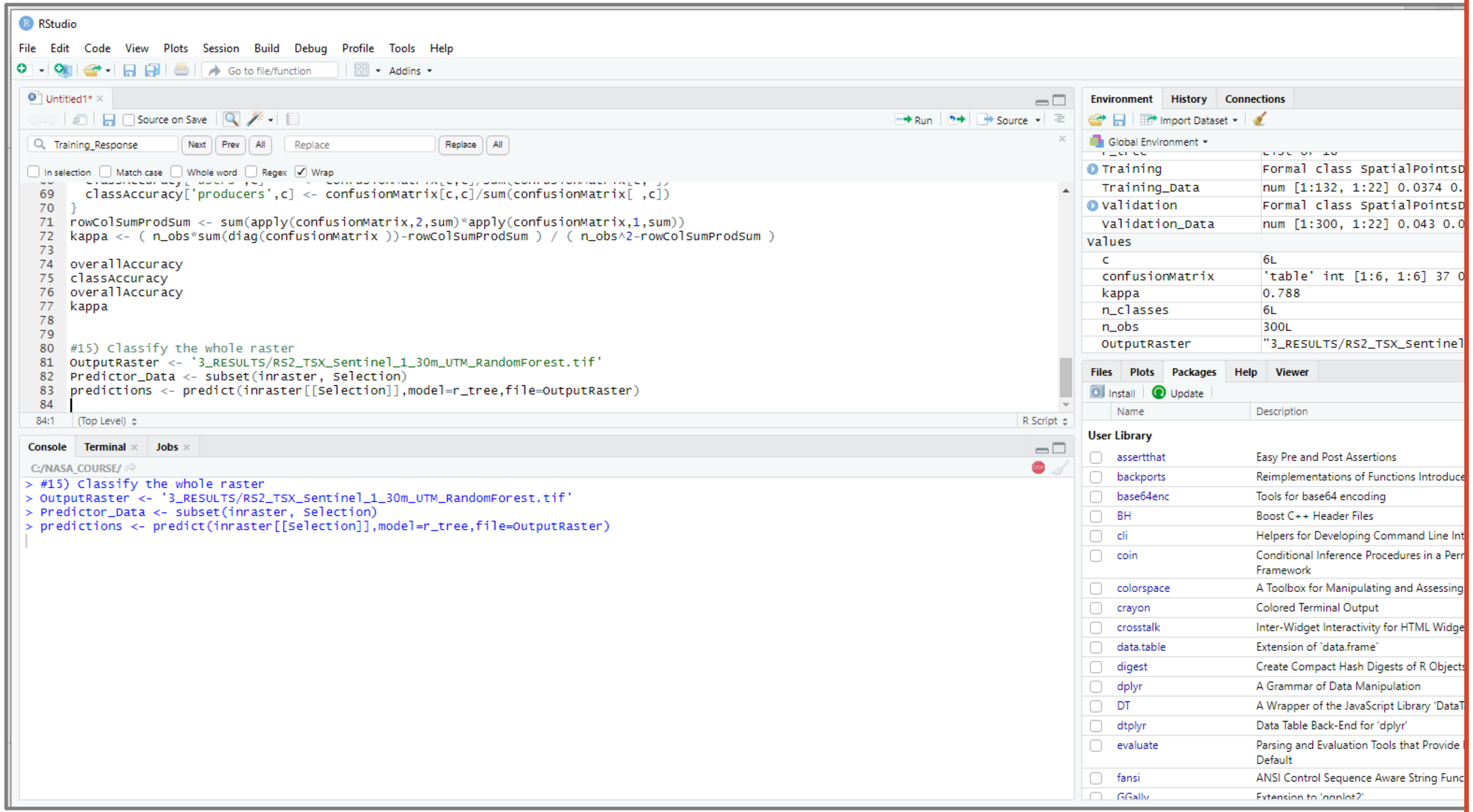
Random Forests: Demostración Práctica

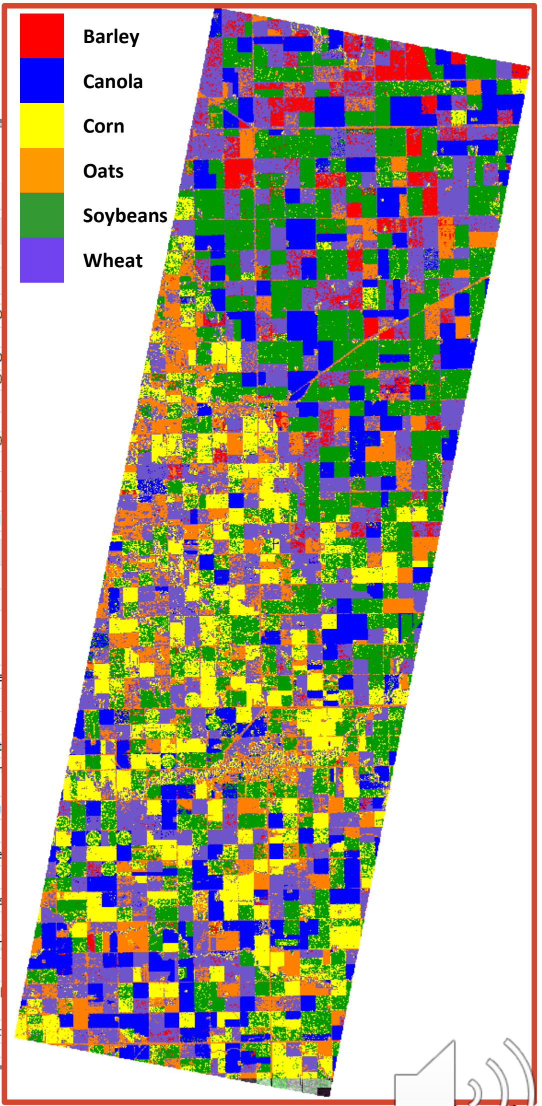
18. Calcular las exactitudes independientes y la estadística kappa



Random Forests: Demostración Práctica

19. Clasificar el ráster completo





Referencias

1. Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.
2. Liaw, Andy, y 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, y 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, y 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, y 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.

Contribuidores

- 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