

National Aeronautics and Space Administration  
Desarrollado en colaboración con  
Agricultura y Agroalimentación de  
Canadá.



 **UCLouvain**

 **esa**



**RUS**

 Copernicus

# Clasificación de cultivos agrícolas con radar de apertura sintética y teledetección óptica

Tereza Roth | RUS Copernicus  
(presentado por Amalia Castro Gómez (RSAC c/o ESA))





# Estructura del curso

5 Octubre, 2021

Recordatorio de Radar de Apertura Sintética (SAR)

7 Octubre, 2021

Recordatorio de Teledetección con Datos Ópticos & Introducción a SNAP

12 Octubre, 2021

Guía para la Clasificación Operativa de Cultivos con Datos Ópticos y de SAR (Parte 1)

**14 Octubre, 2021**

**Guía para la Clasificación Operativa de Cultivos con Datos Ópticos y de SAR (Parte 2)**

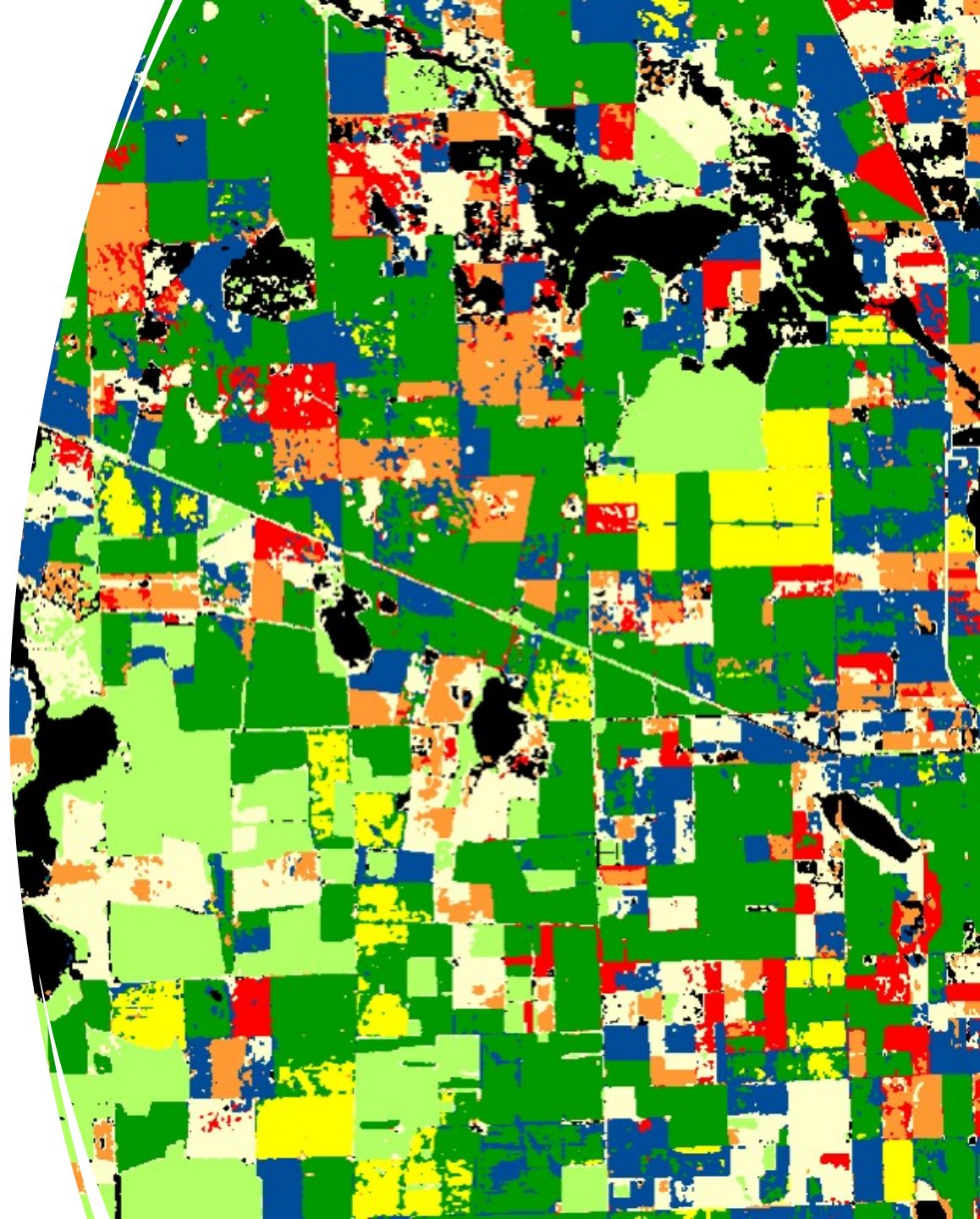
19 Octubre, 2021

Extracción de Variables Biofísicas con Datos Ópticos, Como Apoyo al Monitoreo de Prácticas Agrícolas



# PARTE 4 Descripción General

- Aprendizaje Automático (Machine Learning) para la clasificación de cultivos
- Introducción al Bosque Aleatorio (Random Forest)
- Introducción a la Máquina de Vectores de Soporte (Support Vector Machine)
- Introducción a K-Means
- Demostración: Python para la clasificación de cultivos
  - Bibliotecas de Python seleccionadas
  - Datos satelitales y de capacitación / validación
  - Jupyter lab y clasificación de cultivos
- Sesión de preguntas y respuestas







Aprendizaje Automático (Machine Learning)  
para la clasificación de cultivos



# Aprendizaje Automático (Machine Learning)

Rama de la inteligencia artificial y la informática centrada en el uso de datos y algoritmos para imitar el aprendizaje humano.

## Métodos de Aprendizaje Automático Supervisados

- Usan conjuntos de datos etiquetados para entrenar algoritmos en la clasificación de datos o la predicción de resultados con precisión.
- Se utiliza para la clasificación (e.g. clasificadores lineales, árboles de decisión, **Bosque aleatorio (Random Forest)**, **Máquinas de Vectores de Soporte (Support Vector Machine)**) y Regresión (regresión lineal, regresión logística, regresión polinomial).

## Métodos de Aprendizaje Automático No Supervisados

- Usan algoritmos de Aprendizaje Automático para analizar y agrupar conjuntos de datos sin etiquetar.
- tres tareas principales: agrupamiento (**K-means**), asociación y reducción de dimensionalidad



# Aprendizaje Automático para clasificación de cultivos

- El Aprendizaje Automático es usado sobre todo para la clasificación operativa de cultivos
- **Bosque Aleatorio (Random Forest) es el algoritmo más popular seguido por los algoritmos de Máquinas de Vectores de Soporte (Support Vector Machine)**
- Cualquier error en el conjunto de datos de referencia degradará la precisión de nuestro modelo. → **control de calidad crítico** (mín. 3000 a 10000+ muestras)
- Los modelos de Aprendizaje Automático generalmente no son transferibles a otras áreas o a otras estaciones
- Bosque Aleatorio: mejor para grandes conjuntos de datos con muchas clases
- Máquina de Vectores de Soporte: mejor para cuando el número de clases es pequeño, sensible a la proporción equilibrada de muestras entre clases



# Evaluación de la precisión

Precisión general engañosa → **Métricas F-Score**

$$F - Score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

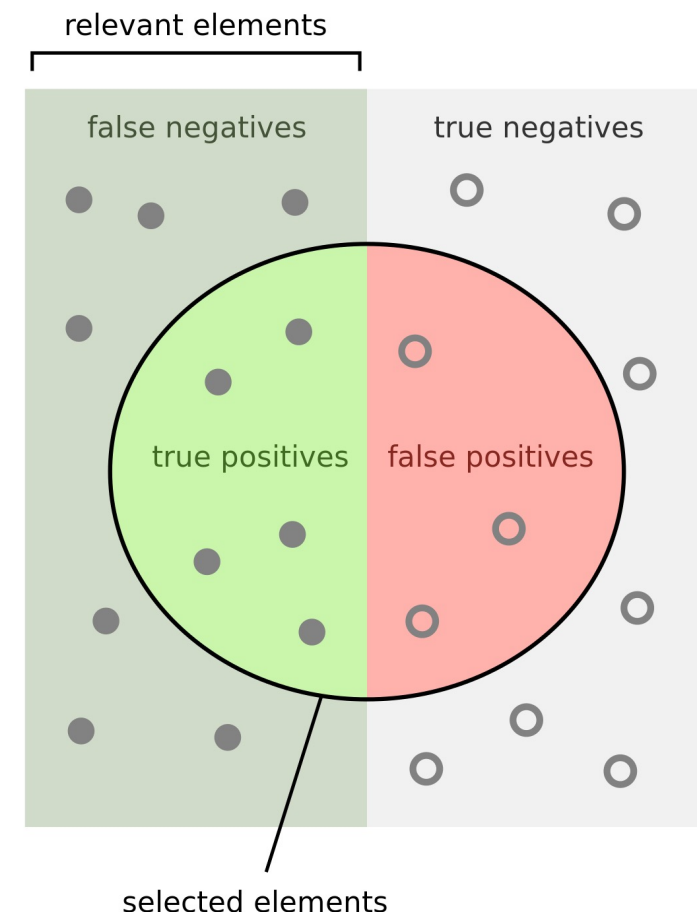
Criterios para una evaluación de precisión válida:

- Estadísticamente riguroso, calidad asegurada, confiable, transparente, reproducible

## Fuentes:

P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, 'Good practices for estimating area and assessing accuracy of land change', *Remote Sensing of Environment*, vol. 148, pp. 42–57, May 2014, doi: [10.1016/j.rse.2014.02.015](https://doi.org/10.1016/j.rse.2014.02.015).

S. V. Stehman and G. M. Foody, 'Key issues in rigorous accuracy assessment of land cover products', *Remote Sensing of Environment*, vol. 231, p. 111199, Sep. 2019, doi: [10.1016/j.rse.2019.05.018](https://doi.org/10.1016/j.rse.2019.05.018).



How many selected items are relevant?

$$Precision = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$Recall = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$





# Introducción al Bosque Aleatorio (Random Forest)



# Bosque Aleatorio (Random Forest)

- El Bosque Aleatorio es un **Algoritmo de Aprendizaje Automático Supervisado** que se utiliza ampliamente en problemas de clasificación y regresión.
- **Clasificador de conjunto** - consta de una gran cantidad de árboles de decisión individuales. Cada árbol proporciona una predicción; la clase con más votos es la predicción del modelo.
- **Embolsado (Agregación Bootstrap)** - cada árbol individual toma muestras al azar del conjunto de datos con reemplazo.
- **Puntuación OOB (Out-Of-Bag)** – da una idea sobre el rendimiento del modelo ya durante la fase de entrenamiento

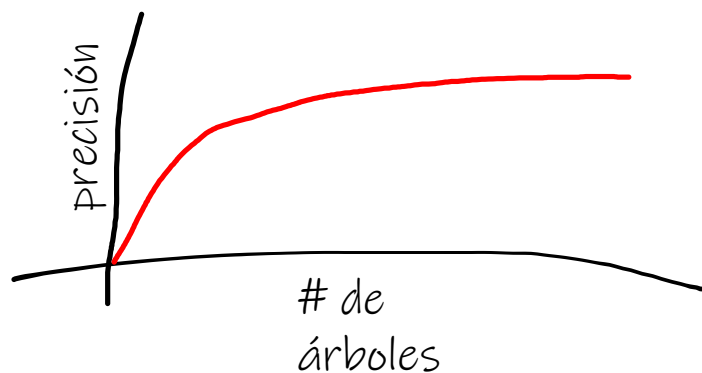




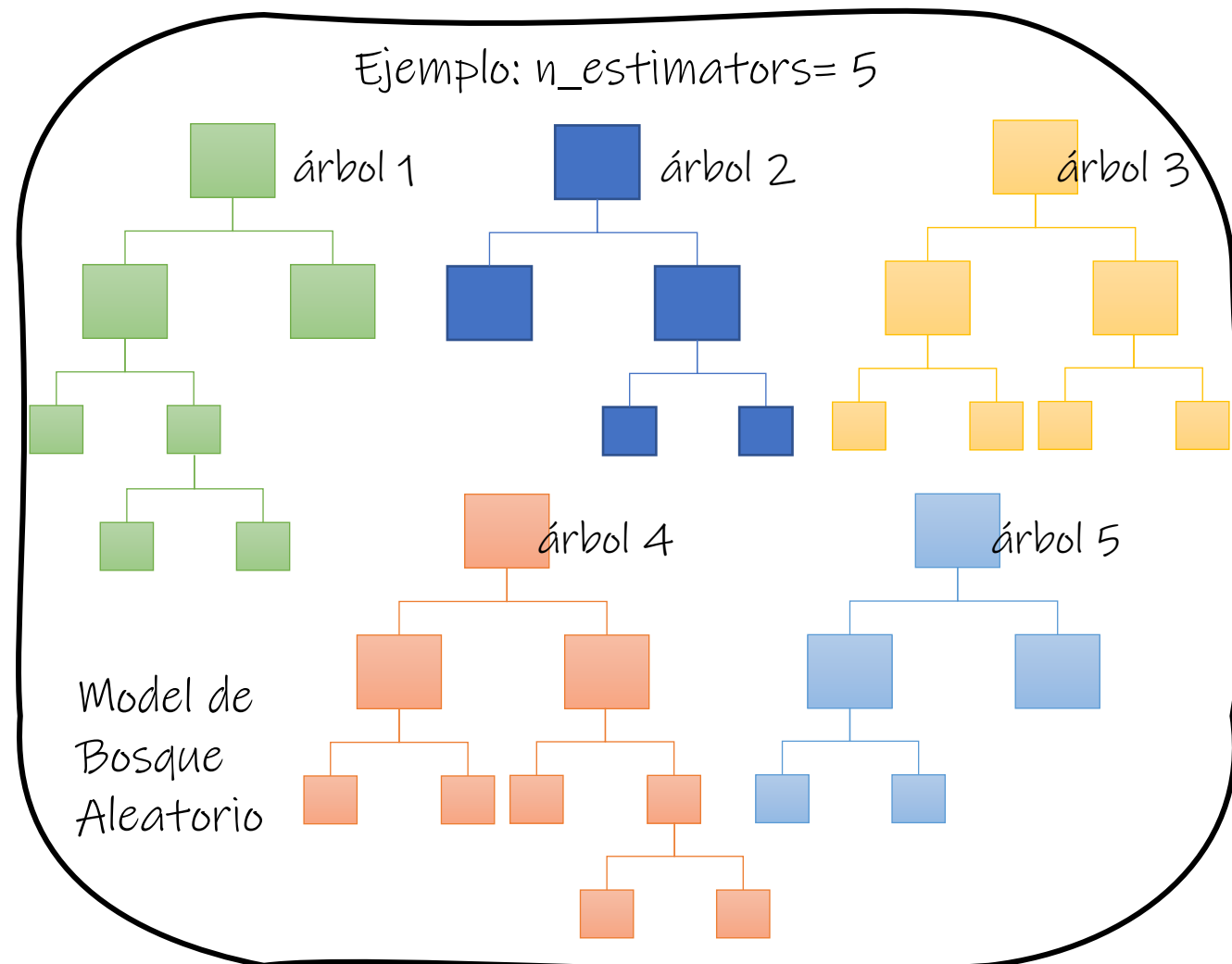
# Bosque Aleatorio: parámetros

**Numero de arboles** (número de estimadores,  $n\_estimators = 100$ )

- $\uparrow$  número de árboles ==  $\uparrow$  precisión



- $\uparrow$  número de árboles ==  $\uparrow$  costo computacional
- Para clasificación en Observación de la Tierra, generalmente **500** árboles
- El número óptimo se puede determinar utilizando la tasa de error OOB



# Bosque Aleatorio: parámetros

## Criterio (criterio = gini)

- Función para medir la calidad de una división
- *Gini* -  $[0,5]$  menos exigente de computar
- *Entropía* -  $[0,1]$

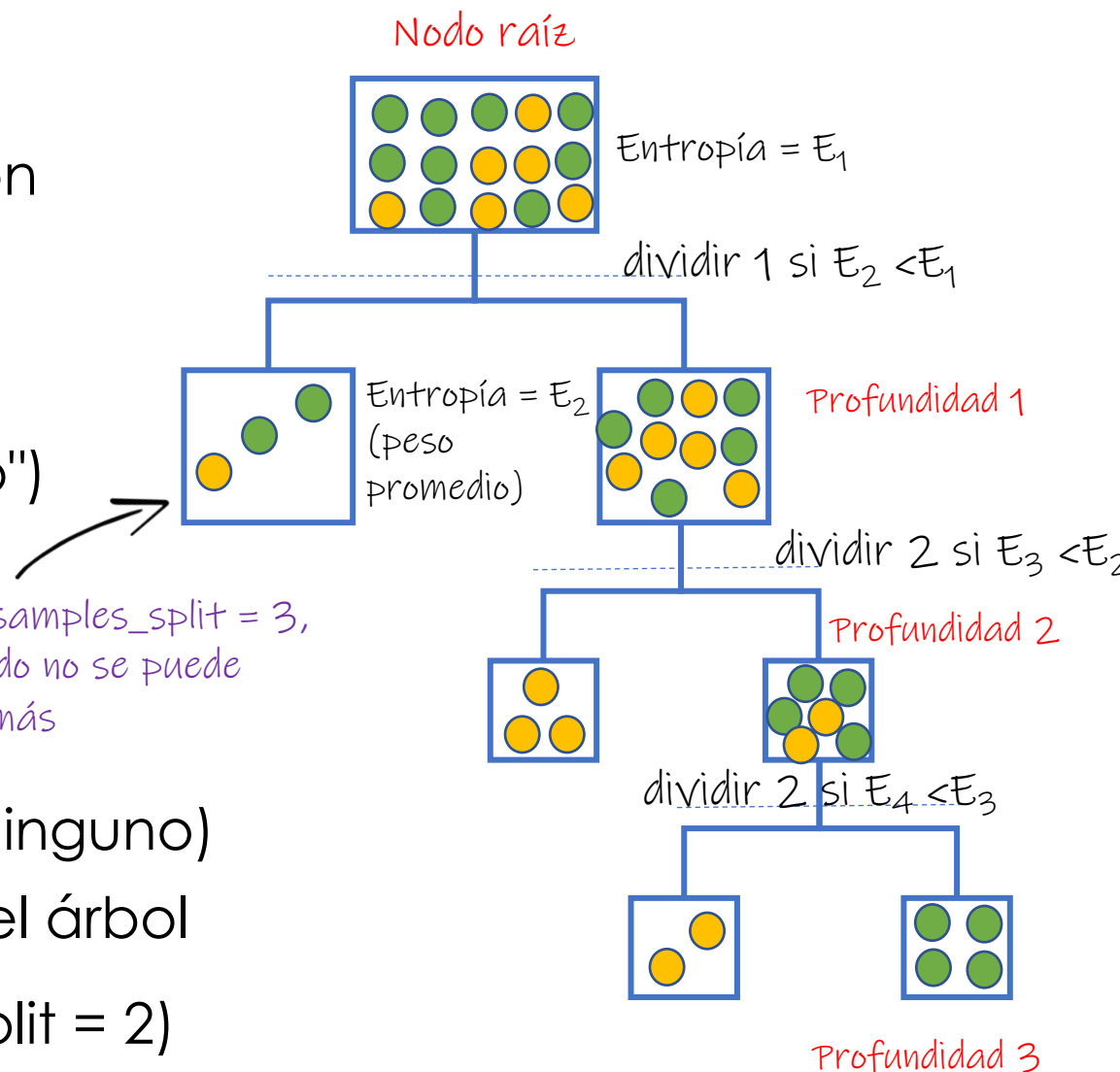
## Características máximas (max\_features = "Auto")

- Características máximas probadas en cada división

## Profundidad máxima del árbol (max\_depth = Ninguno)

- medida de cuánto más se puede expandir el árbol

## Número mínimo de elementos (min\_samples\_split = 2)





# Bosque Aleatorio: parámetros

## **Bootstrap** (bootstrap = Verdadero)

- Si es Falso, se usa todo el conjunto de datos para construir cada árbol.

## **Tamaño de la muestra** (max\_samples = Ninguno)

- Ninguno → mismo tamaño que el conjunto de datos de entrenamiento original
- Int → número específico de muestras
- Flotante [0, 1] → porcentaje del conjunto de datos de entrenamiento

Tamaño de muestra más grande → mayor precisión → mayor costo computacional

**Muchos otros parámetros ...**

**Datos de  
entrenamiento de  
tamaño N = 6**

[1, 2, 3, 4, 5, 6]

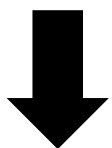
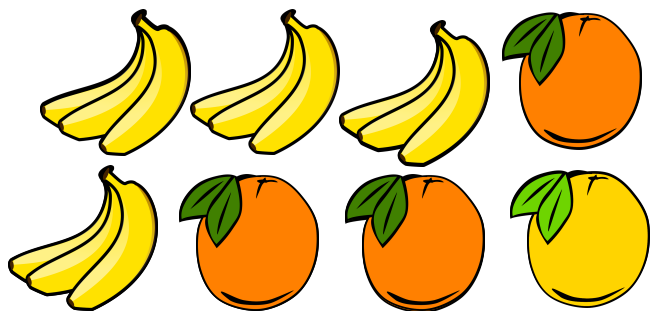


**Muestra de Bootstrap  
de talla N = 6**

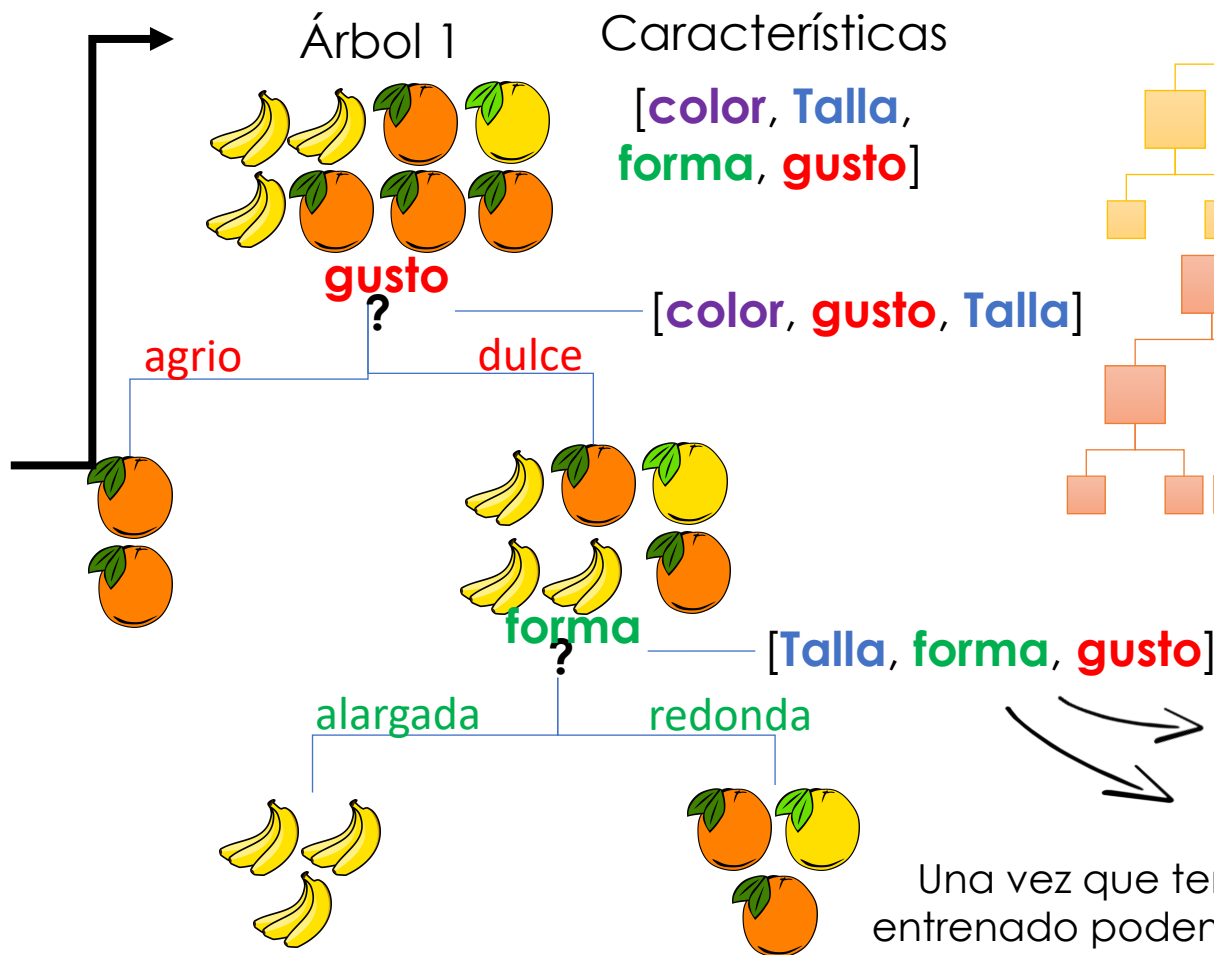
[1, 2, 2, 3, 6, 6]

# Bosque aleatorio - Ejemplo

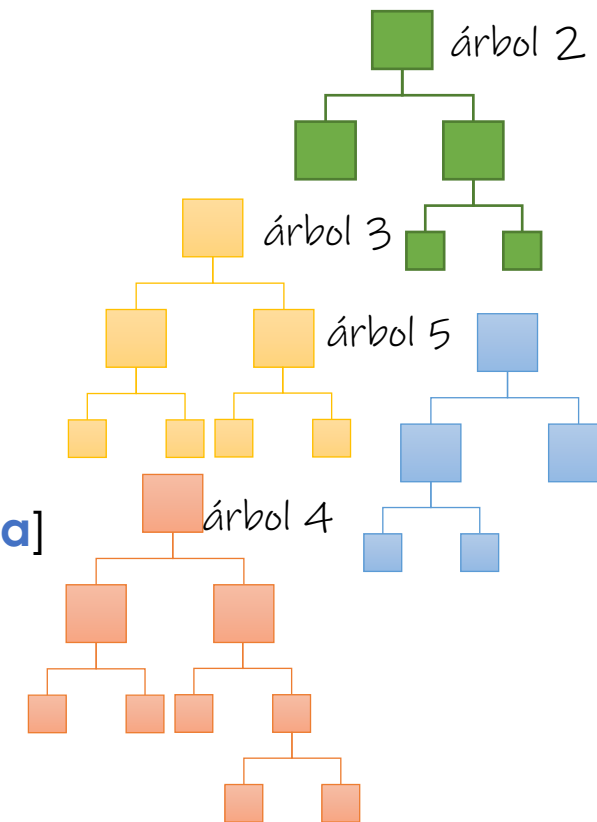
Datos de entrenamiento  
(tamaño = 8)



Muestra aleatoria 1 con reemplazo  
(tamaño de muestra = 8)



n\_estimators = 5  
bootstrap = Verdadero  
max\_features = 3  
min\_samples\_split = 2  
max\_depth = Ninguno  
max\_samples = Ninguno



Estimación de la puntuación OOB

Una vez que tenemos el modelo entrenado podemos alimentarlo con una nueva observación sin saber la clase



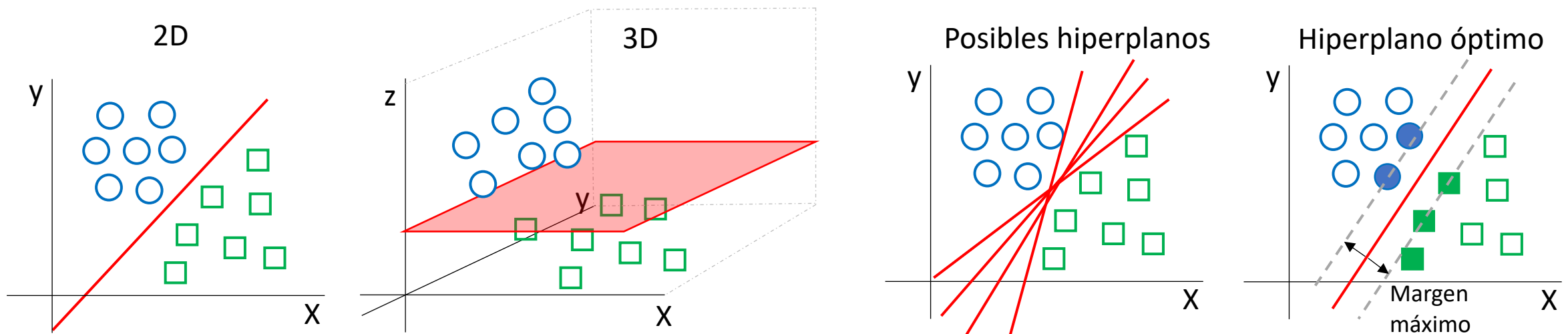


# Máquinas de Vectores Soporte (Support Vector Machines): Introducción



# Máquina de Vectores de Soporte (Support Vector Machine, SVM)

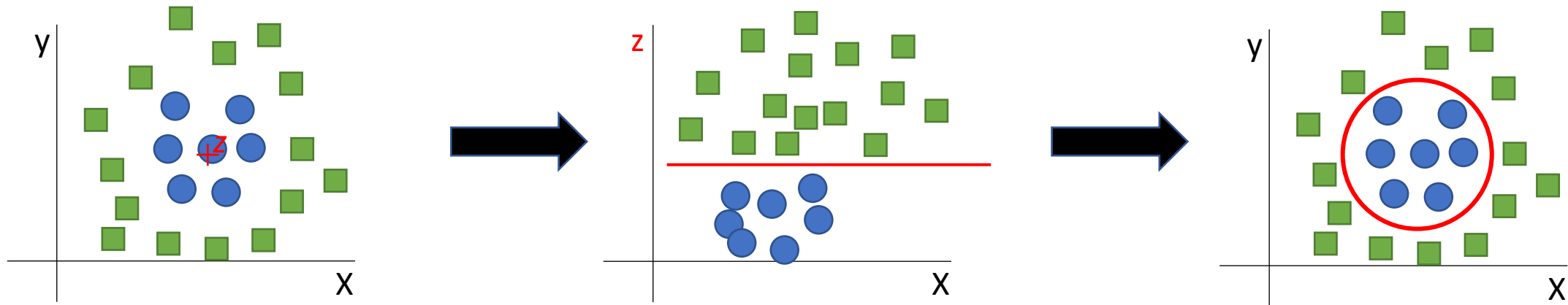
- La máquina de vectores de soporte (SVM) es un **Algoritmo de aprendizaje automático supervisado** utilizado para clasificación y regresión.
- El objetivo del algoritmo SVM es encontrar un **hiperplano** que clasifica claramente los puntos de datos.
- La dimensión del hiperplano depende del número de características.
- **Vectores de apoyo** - puntos de datos que están más cerca del hiperplano e influyen en la posición y orientación del hiperplano.



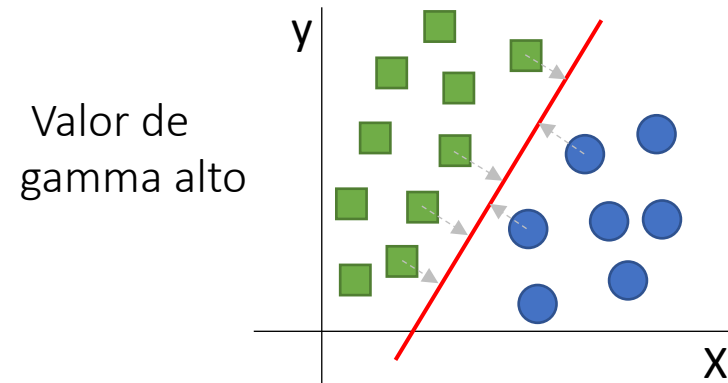
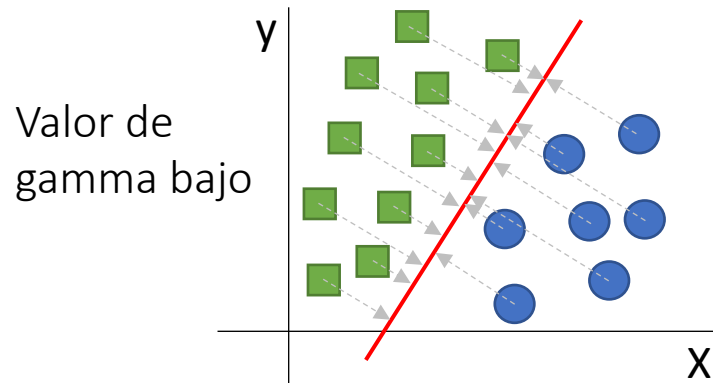


# SVM - Parámetros

**Kernel** (kernel = 'rbf') - función matemática que transforma los datos de entrada a la forma requerida, para poder ajustar un hiperplano.

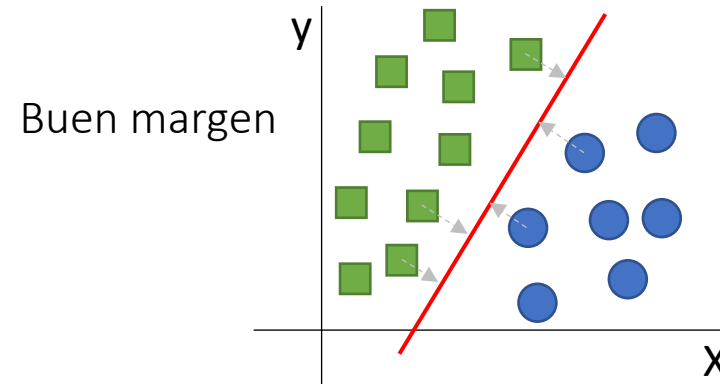
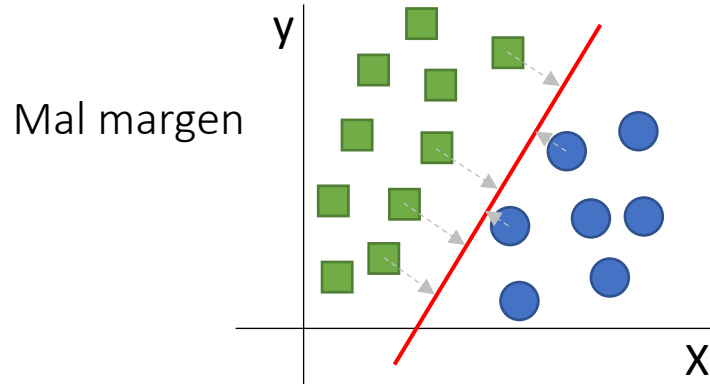


**Gamma** (gamma = 'escala', solo se usa si el kernel es polinomial, rbf o sigmoideo) - controla la distancia de influencia de un solo punto de entrenamiento.

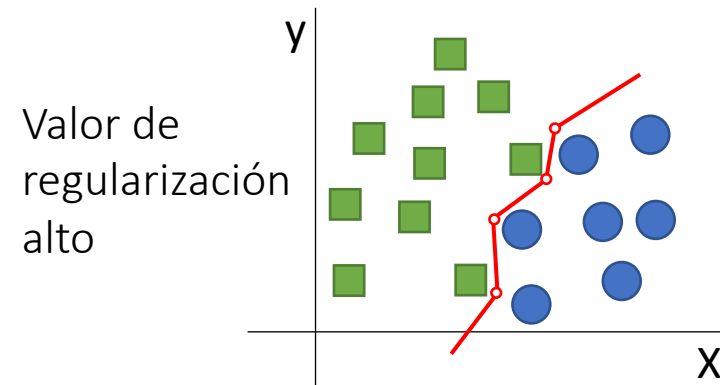
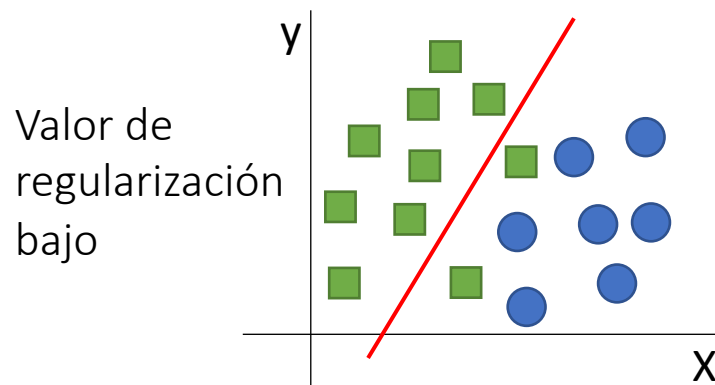


# SVM - Parámetros

**Margen** - una separación (distancia) de la línea / hiperplano a los puntos de clase más cercanos.



**Regularización (C)** - le dice a la optimización de SVM cuánto deseamos evitar clasificar erróneamente cada ejemplo de entrenamiento.





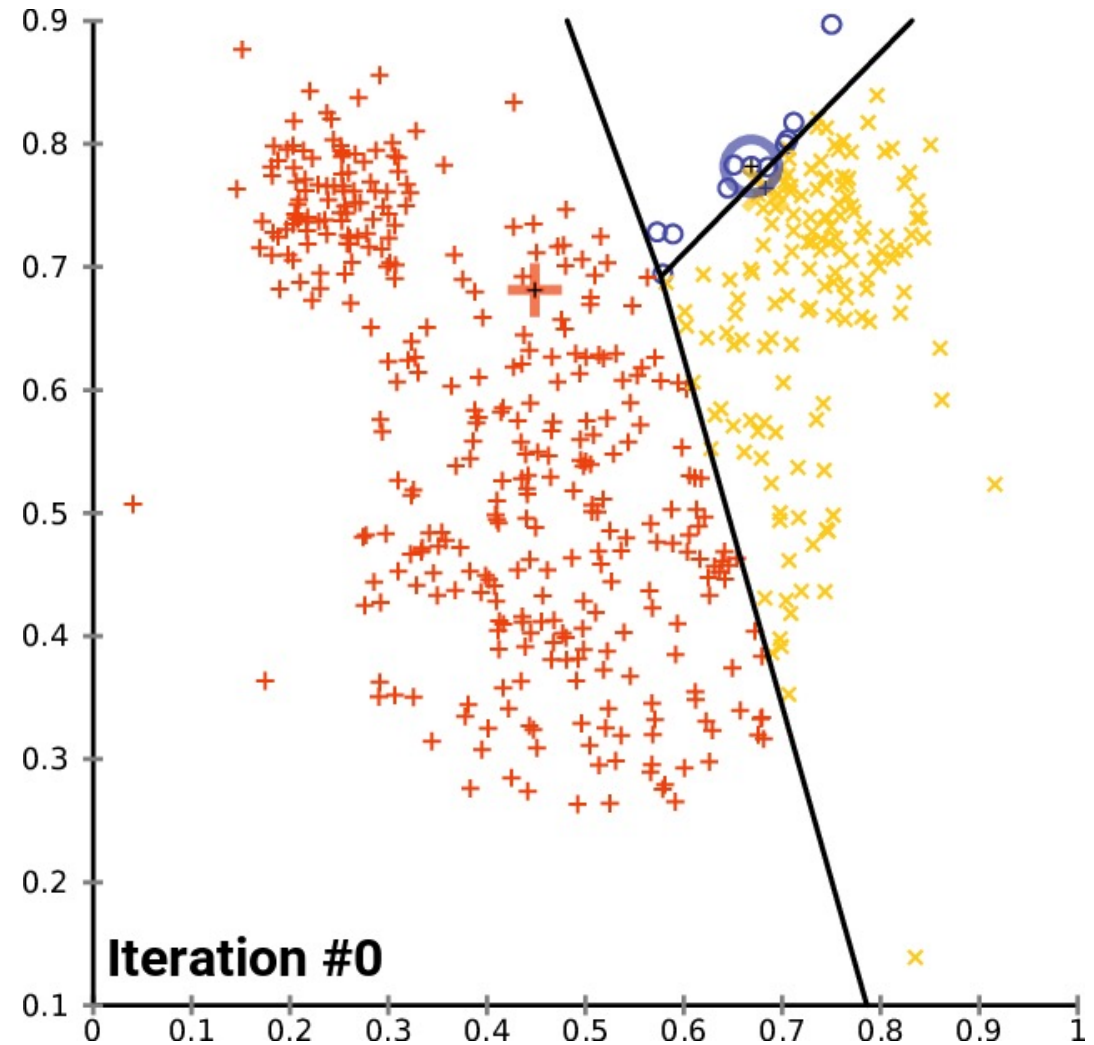


## Introducción a K-Means



# K-means

- Algoritmo de agrupación en clústeres no supervisado
- No necesita datos de entrenamiento
- Algoritmo iterativo que divide el conjunto de datos en  $K$  subgrupos (clústeres) distintos predefinidos que no se superponen.
- Minimiza la suma de la distancia al cuadrado entre los puntos de datos y el centroide.
- Se recomienda la normalización de datos.





# K-means - Parámetros

## **Numero de clusters** (n\_clusters= 8)

- Número de clusters que esperamos, o número óptimo de grupos derivados, por ejemplo, por el método del codo y otros.

## **Centroides de clúster inicial** (init = default='k-means++')

- Si se conocen, podemos proporcionar ubicaciones aproximadas de los centroides de los conglomerados.

## **Número de selecciones de centroide iniciales** (n\_init = 10)

- El algoritmo inicializará los centroides x veces y seleccionará el valor más convergente como el mejor ajuste.

## **Estado aleatorio** (random\_state)

- Establece una semilla aleatoria. Es útil si queremos reproducir clústeres exactos una y otra vez.



Demostración: Python para la clasificación de cultivos



# Paquetes de Python Seleccionados para Aprendizaje Automático



## Scikit learn

- Se integra fácilmente con diferentes bibliotecas de programación de Aprendizaje Automático como NumPy y Pandas.



## Tensor Flow

- Plataforma de código abierto para el Aprendizaje Automático. Funcionalidad avanzada para modelos de Aprendizaje Profundo y Redes Neuronales.



## Keras

- Una de las bibliotecas de Redes Neuronales de código abierto más populares para Python. Amplía la usabilidad de TensorFlow con funciones adicionales para la programación con Aprendizaje Automático y Aprendizaje Profundo.

# Conjunto de datos de partida ópticos y de radar

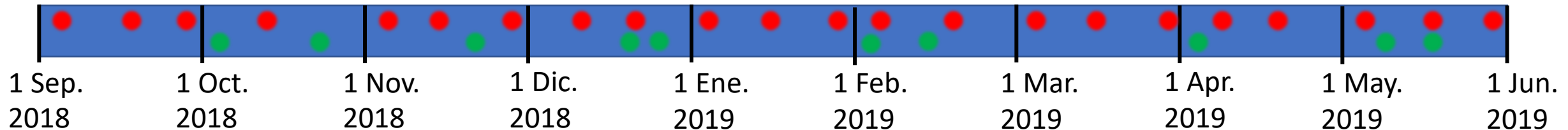
## Sentinel-1 IW GRDH - 22 imágenes ●

- Preprocesado como se muestra en la Parte 3 de esta serie de seminarios web



## Sentinel-2 L1C - 10 imágenes ●

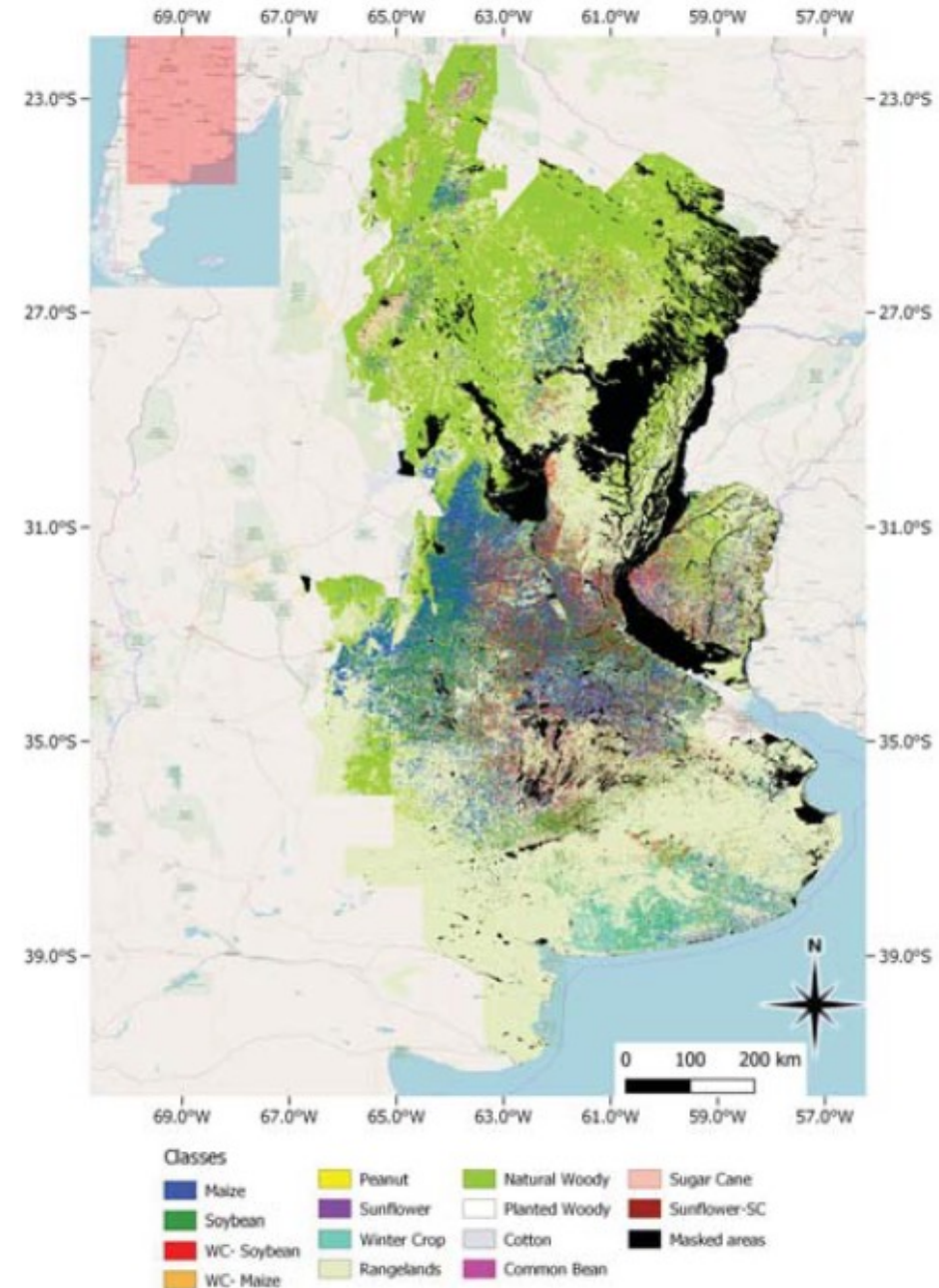
- Bandas usadas: B2 (azul), B3 (verde), B4 (rojo), B8 (NIR), B11 (SWIR 1), B12 (SWIR 2)
- Remuestreado a 10m y creación de subconjunto al área de estudio



# Datos de entrenamiento

- Mapa basado en clasificación de Bosque Aleatorio con Landsat-8 y muestras de campo en 14 zonas agrícolas; precisión general del 90% en zona X (utilizada aquí)
- Los datos de nuestra área de estudio se vectorizaron, se eliminaron los polígonos pequeños y se redujo el número de polígonos para las clases principales.
- 8 clases presentes
- División del 70/30 en datos de entrenamiento y validación
- Se extrajeron al azar 1000 (200) puntos por clase de los datos de entrenamiento (validación)

D. de Abelleira *et al.*, 'First Large Extent and High Resolution Cropland and Crop Type Map of Argentina', in *2020 IEEE Latin American GRSS ISPRS Remote Sensing Conference (LAGIRS)*, Mar. 2020, pp. 392–396. doi: [10.1109/LAGIRS48042.2020.9165610](https://doi.org/10.1109/LAGIRS48042.2020.9165610).



Créditos: de Abelleira *et al.* (2020)



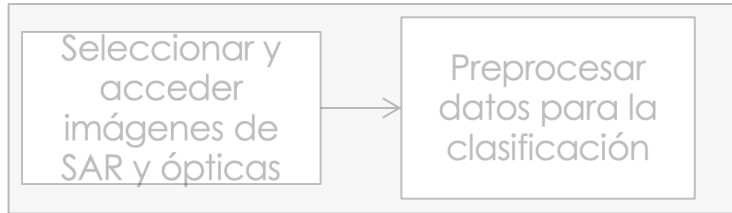


Clasificación con Python

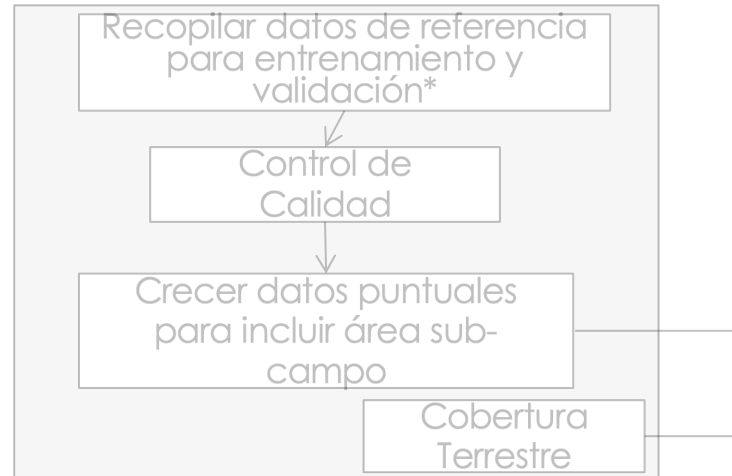


# Metodología Operativa para un Inventario de Cultivos

## 1. Adquirir y Procesar Imágenes Satelitales

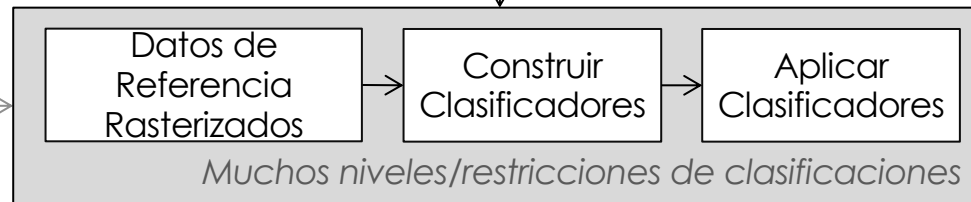


## 2. Datos de Entrenamiento/Validación



## 3. Creación de Región

## 4. Clasificación



## 5. Producto Final





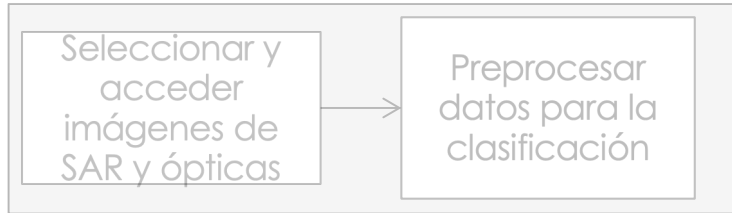


Para Crear el Producto Final

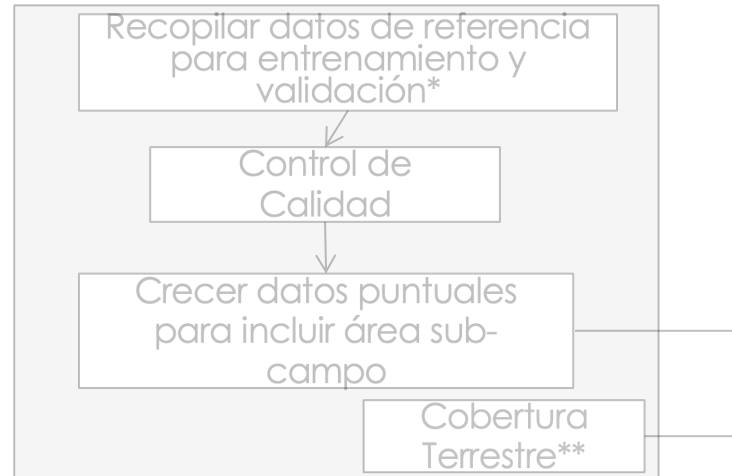


# Metodología Operativa para un Inventario de Cultivos

## 1. Adquirir y Procesar Imágenes Satelitales



## 2. Datos de Entrenamiento/Validación

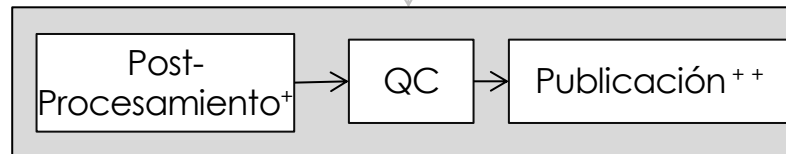


## 3. Creación de Región

## 4. Clasificación



## 5. Producto Final



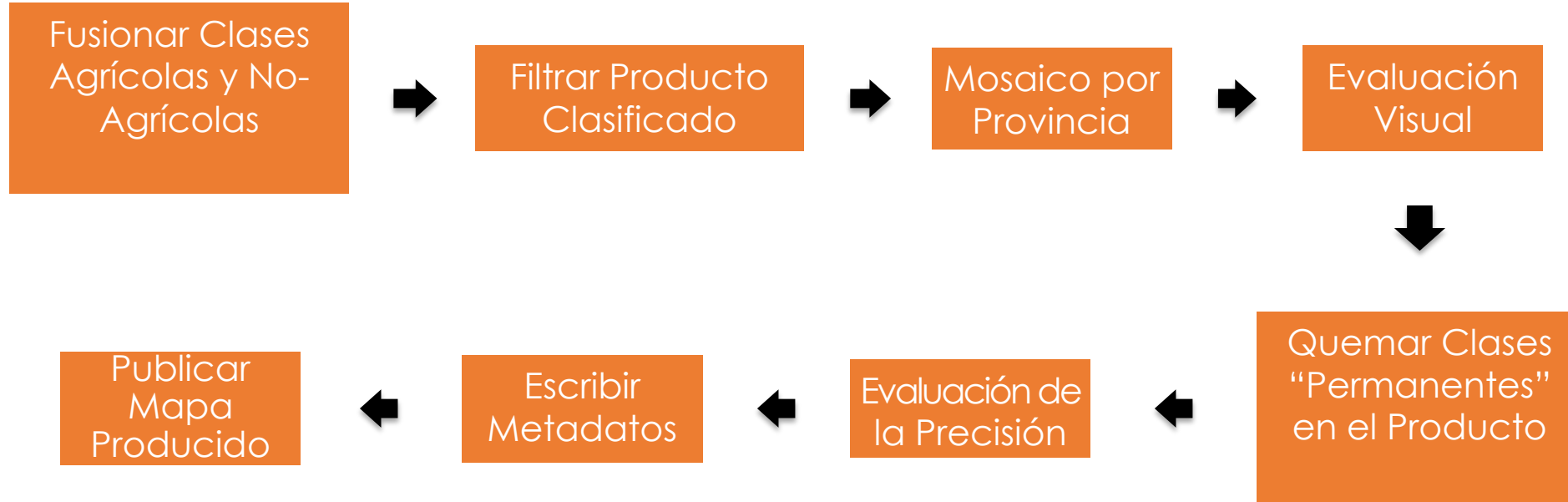
+Mosaico de filtro temático

++GoC Open Data Portal  
AAFC Geospatial Viewer

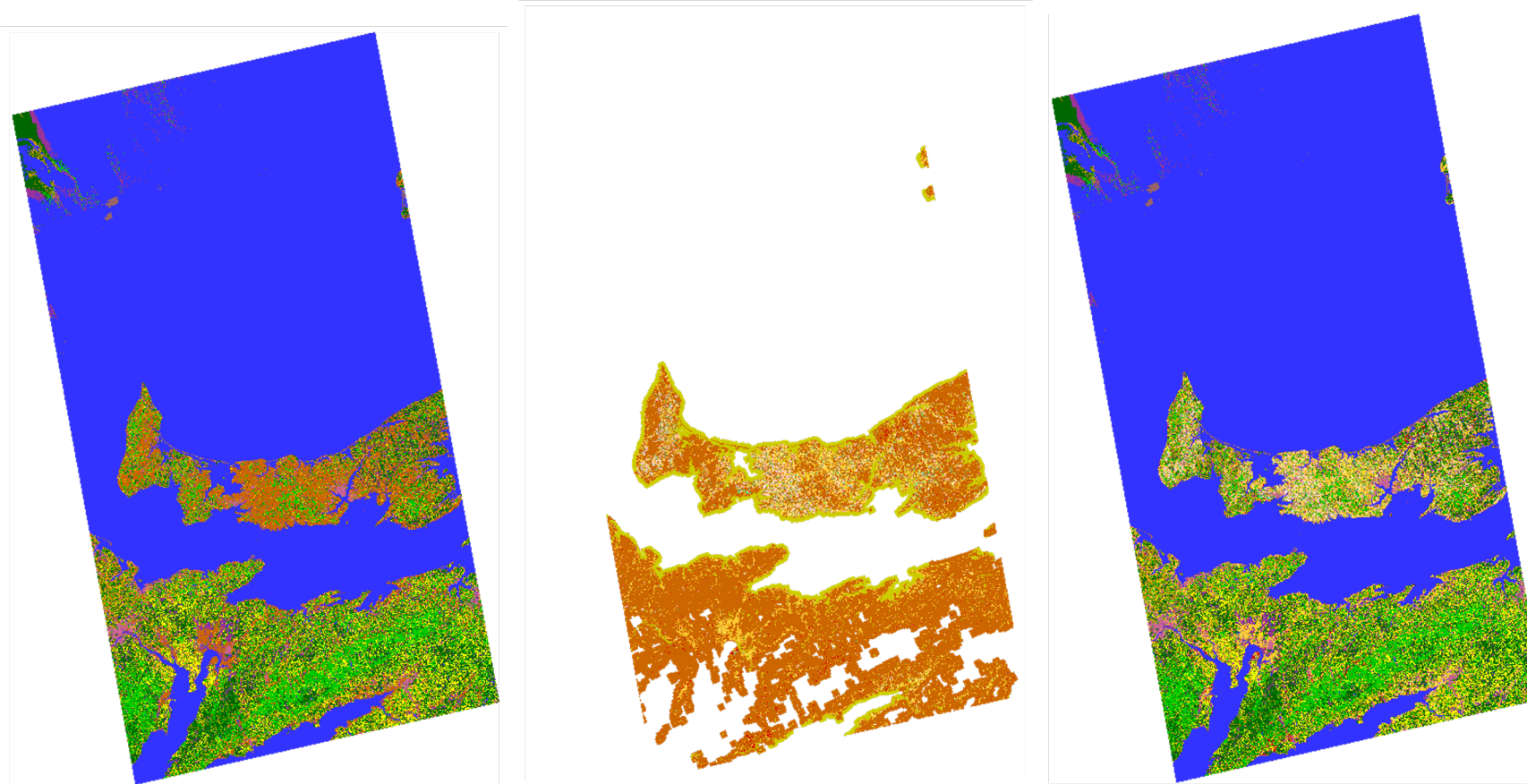




# Pasos para el Post-Procesamiento



# Fusionar Clases Agrícolas y No Agrícolas



Cortesía: Agricultura y Agroalimentación de Canadá

Las clasificaciones de cobertura terrestre (izq.) y cultivos (medio) se fusionan para formar la clasificación final (der.).



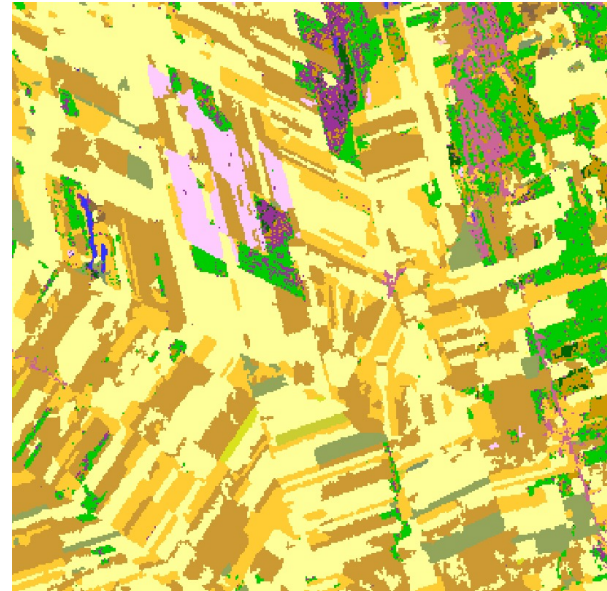
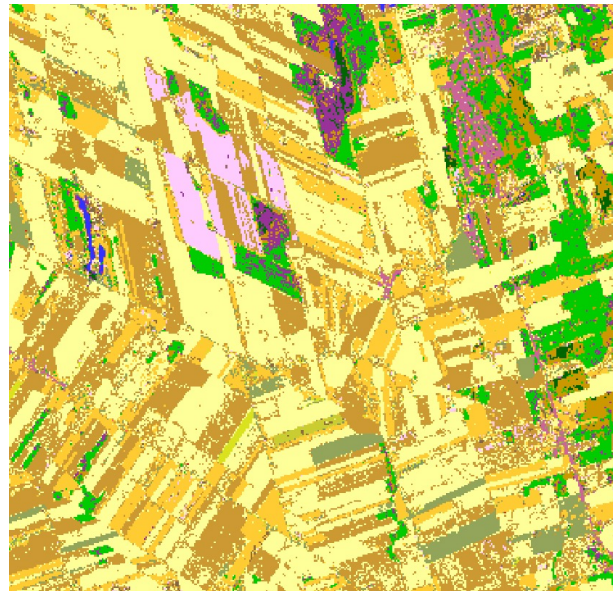


# Filtrar Mapa Producto

AAFC usa una clasificación basada en píxeles que puede ocasionar que haya algunos píxeles “huérfanos” dentro de un campo clasificado.

Se aplica un filtro de moda para reasignar estos píxeles huérfanos a la clase mayoritaria.

- Un filtro de moda calcula la moda de los valores de los píxeles del centro (el valor que ocurre con mayor frecuencia) dentro de una ventana de un filtro particular.
- Recuerde considerar el tamaño de los campos cuando elija el tamaño de kernel para el filtrado.
- Piense en las instancias en las que los píxeles huérfanos tienen sentido (intercultiivos; árboles dentro de campos; agua en los campos etc.).



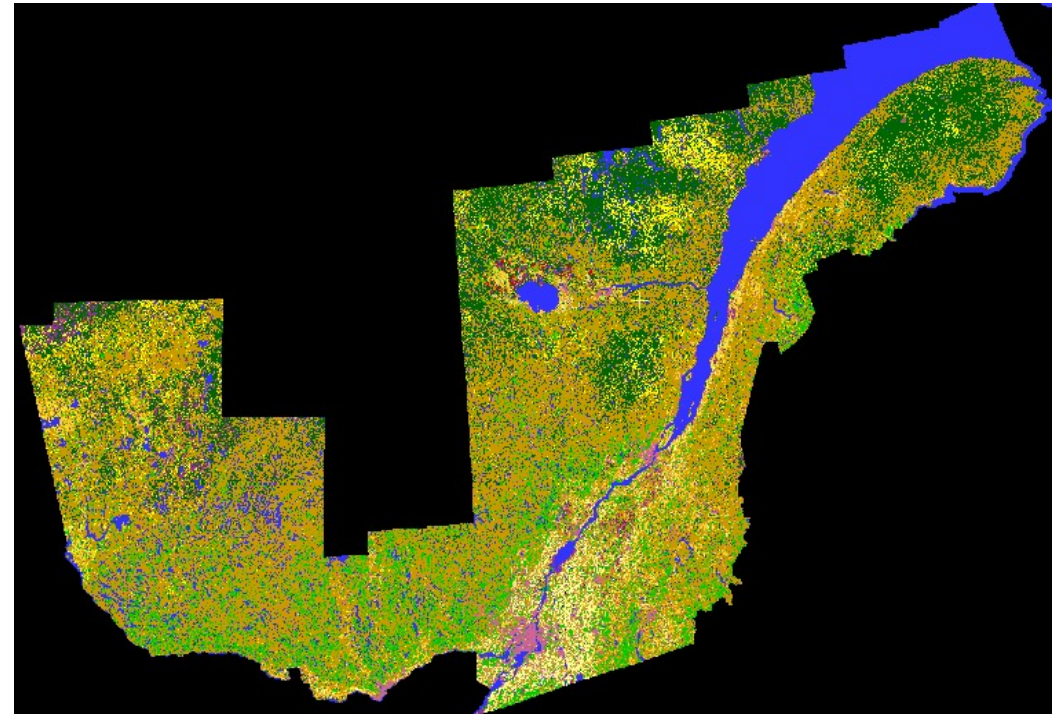
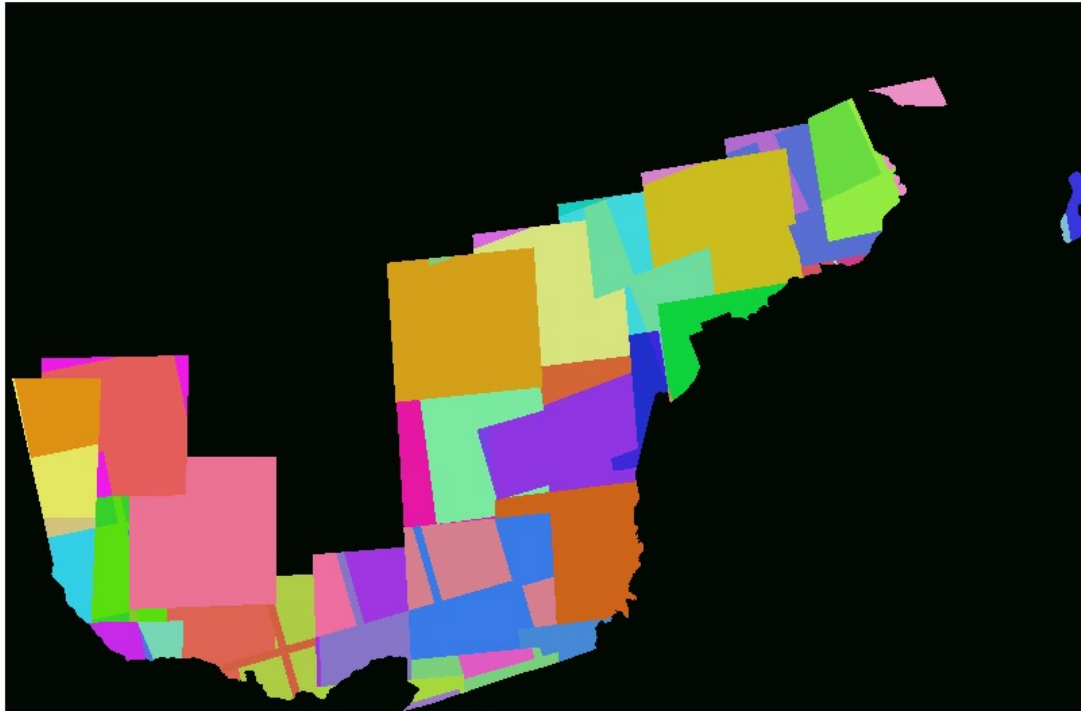
Clasificación cruda (izq.) vs. Clasificación filtrada (der.)



# Creación del Mosaico

AAFC crea mosaicos dentro de una provincia y después crea mosaicos entre provincias.

Es importante revisar la calidad en los bordes y los traslapos.



*Áreas de traslapo individuales en la pila del proceso de crear el mosaico para Quebec.*

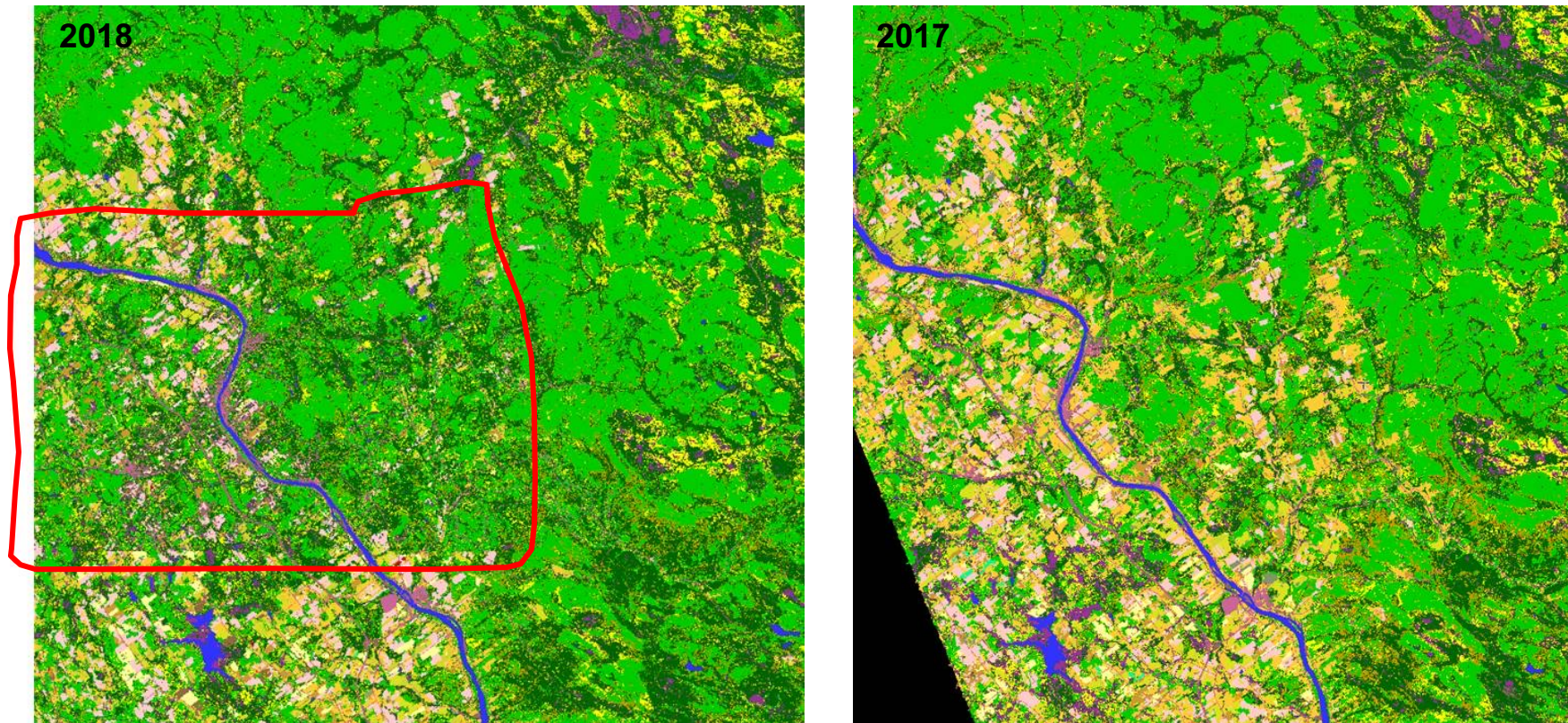




# Evaluación Visual

Las operaciones de la AAFC realizan una inspección y revisión visual manual, pero solo para errores mayores.

Considerando la naturaleza operativa de la ACI, es importante encontrar un equilibrio entre la calidad del mapa y los recursos (el tiempo de las/los analistas).



El área dentro del polígono rojo subestima el número de campos agrícolas en comparación con el año anterior. Este error fue detectado por inspección visual y luego corregido por un analista.

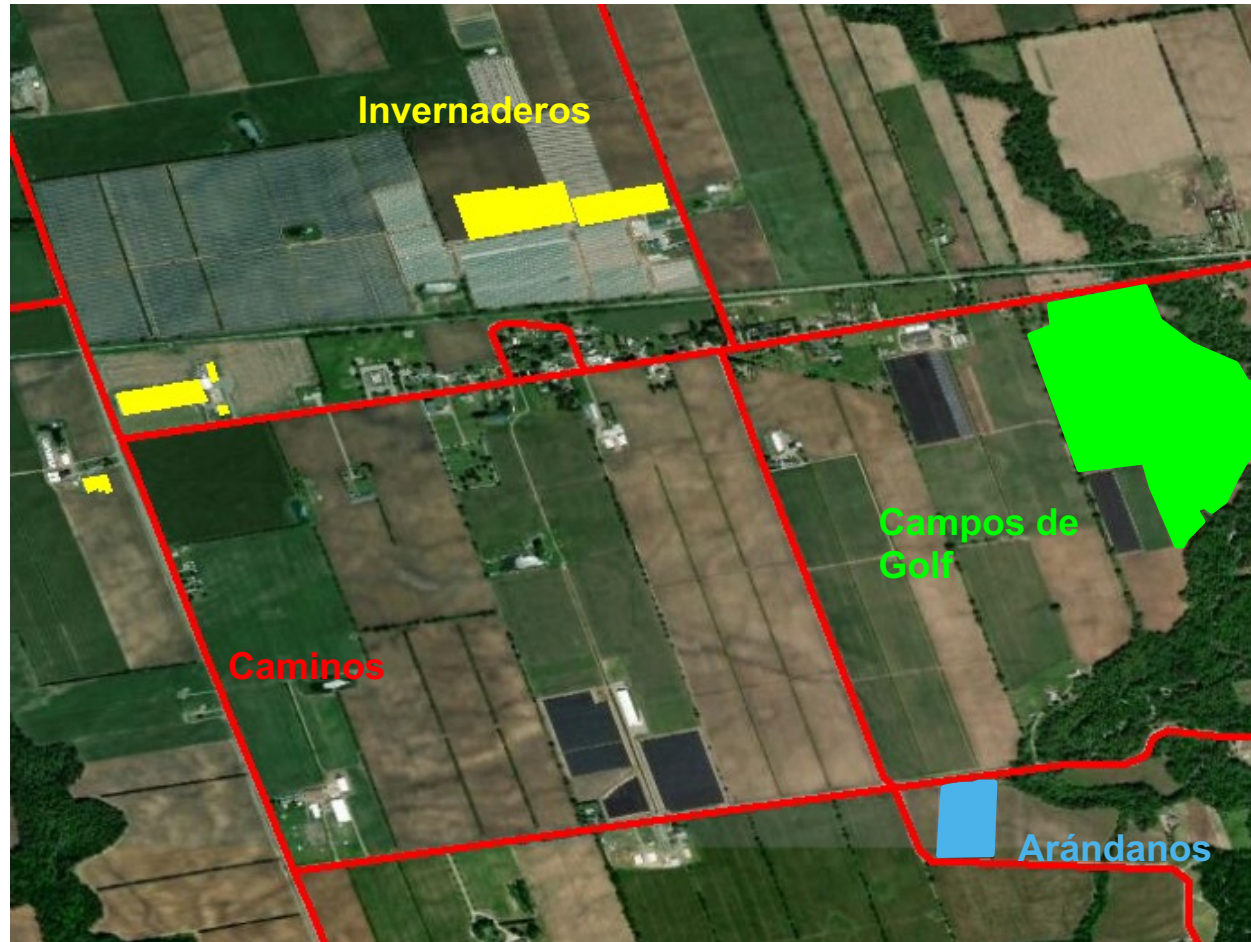




# Quemar Clases “Permanentes”

Las clases “permanentes” se queman en el producto clasificado.

Estas clase incluyen, por ejemplo, invernaderos, caminos, huertos solares, campos de golf etc.

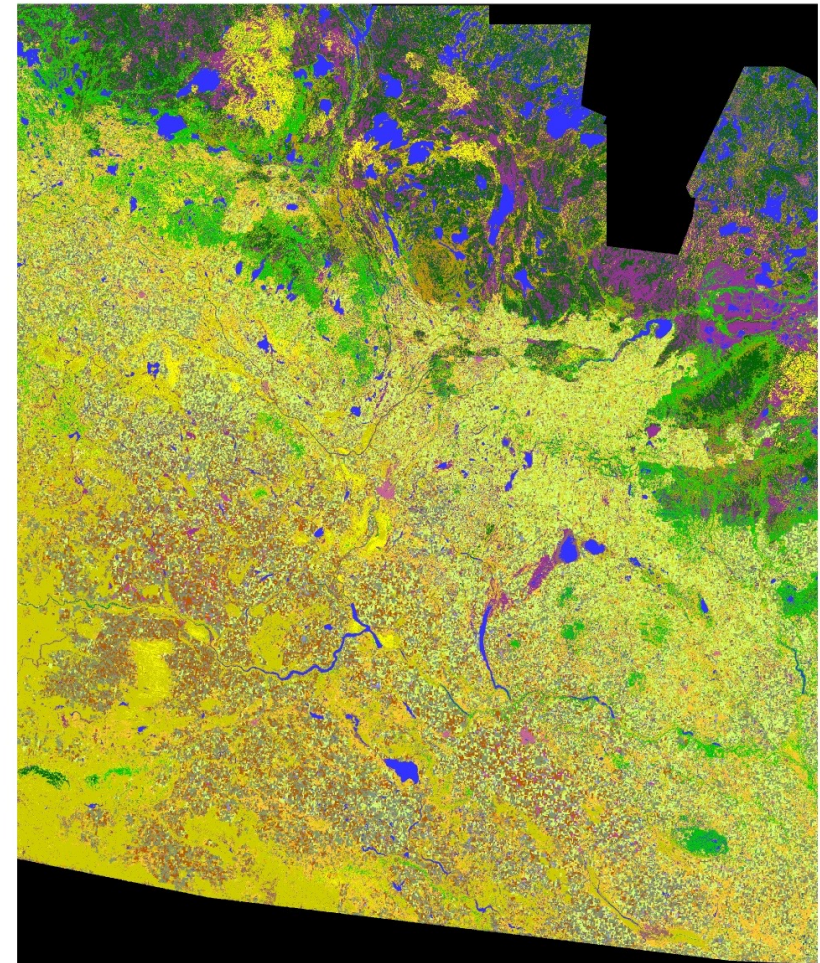




# Evaluación de la Precisión

Matriz de confusión, precisión del usuario y del productor (errores de omisión y comisión);  
Precisión general; coeficiente kappa

Evaluación de la Precisión - Saskatchewan 2020		
	Precisión del Usuario	Precisión del Productor
Pastoreo/Forraje	85.91	90.70
Muy húmedo para sembrar	65.46	33.48
Barbecho	83.84	76.91
Cebada	91.13	87.56
Mijo	0.00	0.00
Avena	87.13	78.86
Centeno	92.60	61.36
Farro	98.02	75.51
Triticale	74.38	24.52
Quínoa	94.47	37.20
Trigo Invernal	87.53	59.52
Trigo Primavera	91.91	96.49
Maíz	76.88	69.65
Camelina	91.36	37.77
Canola	96.60	98.39
Linaza	93.19	81.45
Mostaza	96.77	73.73
Girasol	92.19	94.41
Soya	95.08	88.92
Arveja	96.21	92.28
Garbanzo	96.48	84.94
Frijol	79.21	37.46
Haba	94.55	83.77
Lenteja	94.44	96.26
Papa	39.41	88.33
Hierbas Aromáticas	98.69	48.56
Alforfón	93.25	45.41
Alpiste	96.10	74.00
Cáñamo	85.43	40.86
Precisión General	93.87	
Kappa:	0.92	



0 60 120 240 Kilometers

Cortesía: Agricultura y Agroalimentación de Canadá



# Escribir Metadatos

- [ISO 19131-Inventario Annual de Cultivos – Especificaciones de Productos de Datos](#)
- Resumen de la metodología
- Precisiones generales por provincia; Se pueden obtener precisiones de cultivos individuales a pedido.
- Traducción de productos geospaciales y metadatos al francés

## 3.2.12. Annual Crop Inventory, 2020

Title	Annual Crop Inventory, 2020
Alternate Title	AAFC Crop Type Mapping, 2020
Abstract	In 2020, the Earth Observation Team of the Science and Technology Branch (STB) at Agriculture and Agri-Food Canada (AAFC) repeated the process of generating annual crop inventory digital maps using satellite imagery to for all of Canada, in support of a national crop inventory. A Decision Tree (DT) based methodology was applied using optical (Landsat-8, Sentinel-2) based satellite images, and having a final spatial resolution of 30m. In conjunction with satellite acquisitions, ground-truth information was provided by: provincial crop insurance companies in Alberta, Manitoba, & Quebec; point observations from the PEI Department of Environment, Water and Climate Change; the Ontario Ministry of Agriculture, Food and Rural Affairs; and data collection supported by our regional AAFC Research and Development Centres in St. John's, Charlottetown, Fredericton, and Guelph. <b>Due to COVID-19 travel restrictions, complete sampling coverages in NL, NS, NB and BC were not possible, as a result the general agriculture class (120) is found in these provinces in areas where there was no ground data collected.</b>
Purpose	An annual national crop type map
Topic Category	Farming; Environment; GeoscientificInformation; imagery; BaseMaps; EarthCover;
Spatial Representation Type	grid
Spatial Resolution	30 m
Geographic Description	Canada
Supplemental Information	<p>Overall provincial accuracies for crop classes are:</p> <p>Newfoundland: 95.08%            Prince Edward Island: 85.85%            Nova Scotia: N/A            New Brunswick: 95.74%            Quebec: 91.20%            Ontario: 88.26%            Manitoba: 93.58%            Saskatchewan: 93.87%            Alberta: 90.99%            British Columbia: 85.16%</p> <p>Overall provincial accuracies for non-agriculture land cover are:</p> <p>Newfoundland: 75.09%            Prince Edward Island: 75.74%            Nova Scotia: 70.52%            New Brunswick: 68.57%            Quebec: 72.85%            Ontario: 75.84%            Manitoba: 68.97%            Saskatchewan: 69.93%            Alberta: 65.59%            British Columbia: 77.09%</p> <p>Citation: Agriculture and Agri-Food Canada, 2020, "Annual Space-Based Crop Inventory for Canada, 2020", Agroclimate, Geomatics and Earth Observation Division, Science and Technology Branch. <a href="https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9">https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9</a></p>
Constraints	Data are subject to the Government of Canada Open Data Licence : <a href="http://open.canada.ca">http://open.canada.ca</a>
Keywords	Government of Canada Core Subject Thesaurus (2000-02-01) - Remote Sensing, Satellites, Agriculture, Crops, Crop insurance, Farmlands, Forage crops, Land cover, Geomatics, Geographic Information Systems, Geographic data, maps, Geography
Scope Identification	dataset
Feature Attribute Names	Class(Value)





# Publicar en “Open Canada”

- [Demo](#)

## Open Government Portal

Found 33 records

**Order By** Best match ▾

► Organization

► Portal Type

► Collection Type

[Annual Crop Inventory](#) Federal

Starting in 2009, the Earth Observation Team of the Science and Technology Branch (STB) at Agriculture and Agri-Food Canada (AAFC) began the process of generating **annual crop** type digital maps.

**Last Updated:** Dec. 9, 2020      **Date Published:** Nov. 4, 2013

**Organization:** Agriculture and Agri-Food Canada

**Formats:**

Agriculture, Satellites, Crops



# ¿Preguntas?

- Introduzca sus preguntas en la sección Q&A. Responderemos en el orden en el que sean recibidos.
- Publicaremos las preguntas y respuestas en el sitio web del curso tras la conclusión del webinar.



# Contactos

- Instructores:
  - Tereza Roth: [eotraining@serco.com](mailto:eotraining@serco.com)
  - Georgia Karadimou: [eotraining@serco.com](mailto:eotraining@serco.com)
  - Amalia Castro Gómez: [Amalia.castro.Gomez@esa.int](mailto:Amalia.castro.Gomez@esa.int) (para consultas en español)
- Página web de formación:  
<https://appliedsciences.nasa.gov/join-mission/training/spanish/arset-clasificacion-de-cultivos-agricolas-con-radar-de-apertura>
- Sitio web EO4Society de la ESA:  
<https://eo4society.esa.int/training-education/>
- Twitter: [@EOOpenScience](https://twitter.com/EOOpenScience)



**¡Gracias!**

