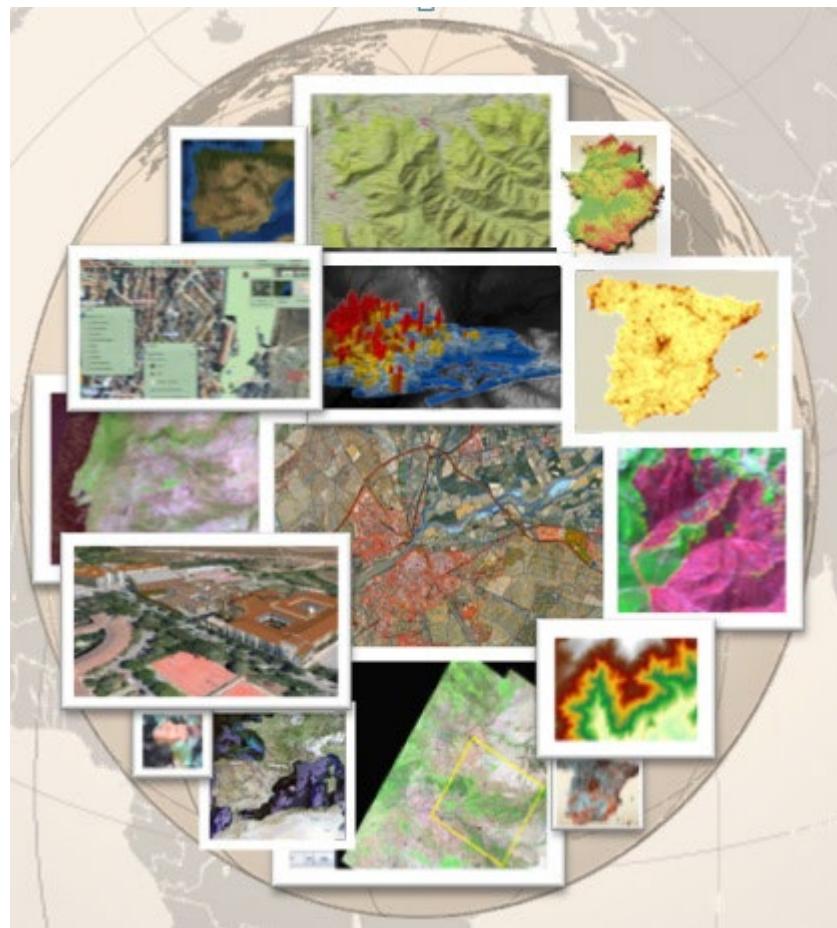


Máster en Tecnologías de la Información Geográfica: SIG y Teledetección



SIG y Teledetección

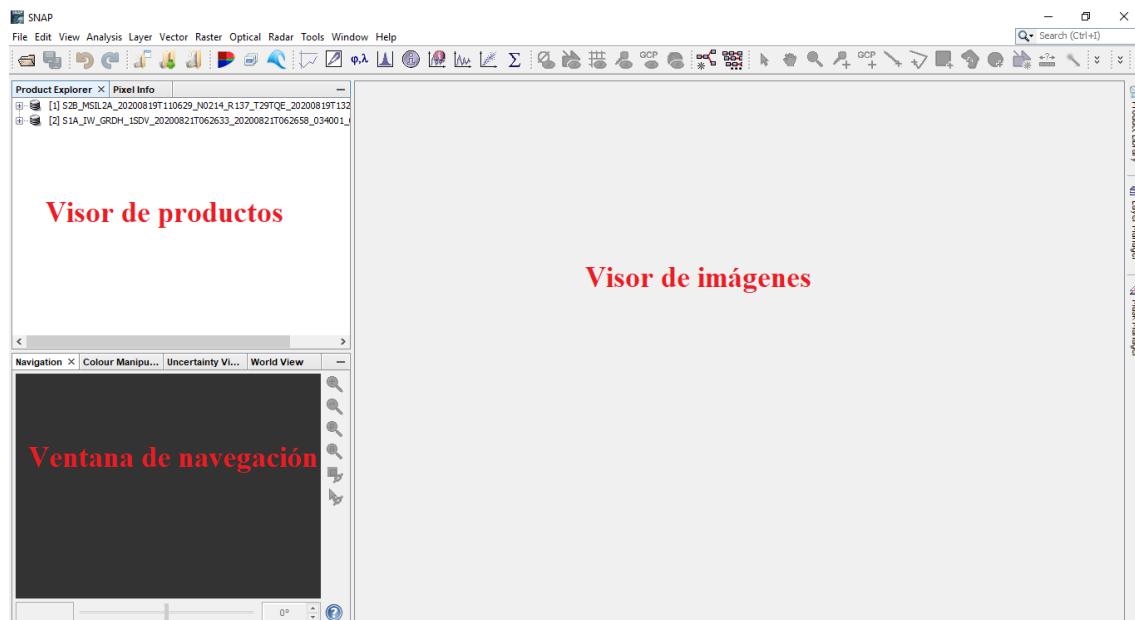
Tema 3. Imágenes Sentinel y su clasificación

Contenido

1. Preprocesado de las imágenes.....	2
1.1 Exploración de la imagen S2	3
1.2 Preprocesado de la imagen S2	13
1.3 Preprocesado de la imagen S1	20
1.4. Índices de Vegetación y de Suelos con S2	37
1.5. Fusión de imágenes S2 y S1	44
2. Clasificación de imágenes.....	46
2.1. Clasificación no supervisada.....	46
2.1.1. Análisis de conglomerados K-means	48
2.1.2. Análisis de clúster de maximización de expectativas (EM).....	50
2.2. Clasificación supervisada.....	50
2.2.1 Definición de áreas de entrenamiento	51
2.2.2 Importación de las áreas de entrenamiento en SNAP.....	66
2.2.3 Algoritmo de clasificación de máxima verosimilitud	70
2.2.4 Algoritmo clasificación Random Forest	74
2.2.5 Clasificación Random Forest con SNAP.....	78
3. Bibliografía.....	84

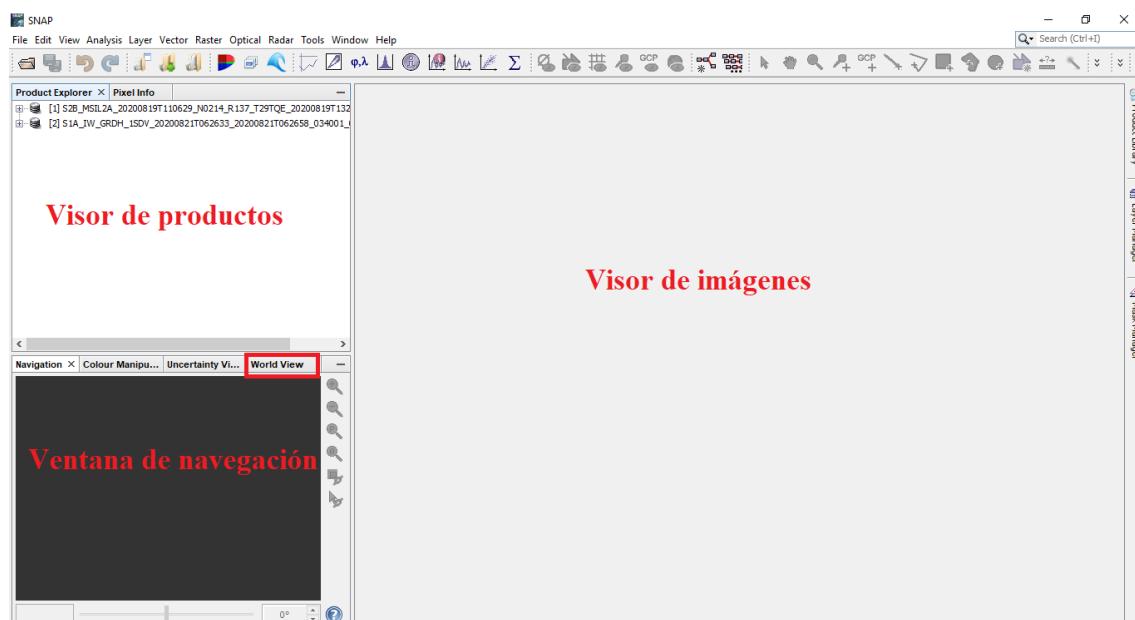
1. Preprocesado de las imágenes

SNAP está dividido en tres áreas de trabajo claramente diferenciadas:



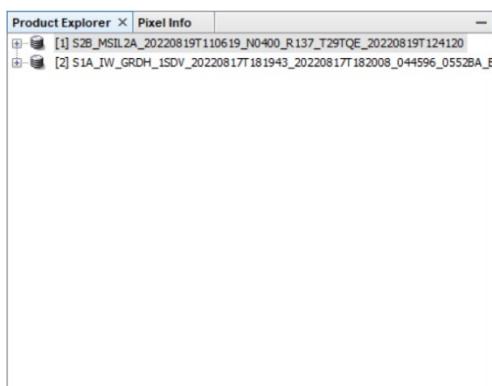
A medida que vayamos trabajando con las imágenes, podremos ir viendo lo que nos ofrece cada área de trabajo.

Así, por ejemplo, si pinchamos en la vista del mundo, podremos observar en qué zona de la superficie terrestre estamos trabajando.





En este caso, se puede ver lo que ocupa la imagen 1, que es la Sentinel 2 (S2) y la imagen 2, mucho más amplia, que es la Sentinel 1 (S1).



1.1 Exploración de la imagen S2

La misión Sentinel-2 proporciona imágenes multiespectrales de alta resolución y gran alcance. Se trata de una constelación de dos satélites de órbita polar colocados en la misma órbita sincrónica del sol, desfasados a 180° uno del otro. La captura de datos se realiza en una gran anchura de banda (290 km) y tiene un período de revisita de 10 días en el ecuador con un satélite, y 5 días con 2 satélites.

La misión está diseñada para observar la superficie terrestre y pretende proporcionar continuidad de datos de imágenes de tipo SPOT y LANDSAT, dando servicio al programa de observación de la tierra Copernicus en distintas áreas temáticas: cobertura terrestre, la agricultura y la silvicultura, gestión de desastres, la cartografía de riesgos y también en temas de seguridad nacional.

Las imágenes S2 están disponibles en dos niveles de procesamiento, en una cuadrícula de 100x100 km² en proyección UTM/WGS84.

Name	High-Level Description	Production & Distribution	Data Volume
Level-1C	Top-Of-Atmosphere reflectances in cartographic geometry	Systematic generation and online distribution	~600 MB (each 100km x 100km ²)
Level-2A	Bottom-Of-Atmosphere reflectances in cartographic geometry	Systematic and on-User side (using Sentinel-2 Toolbox)	~800 MB (each 100km x 100km ²)

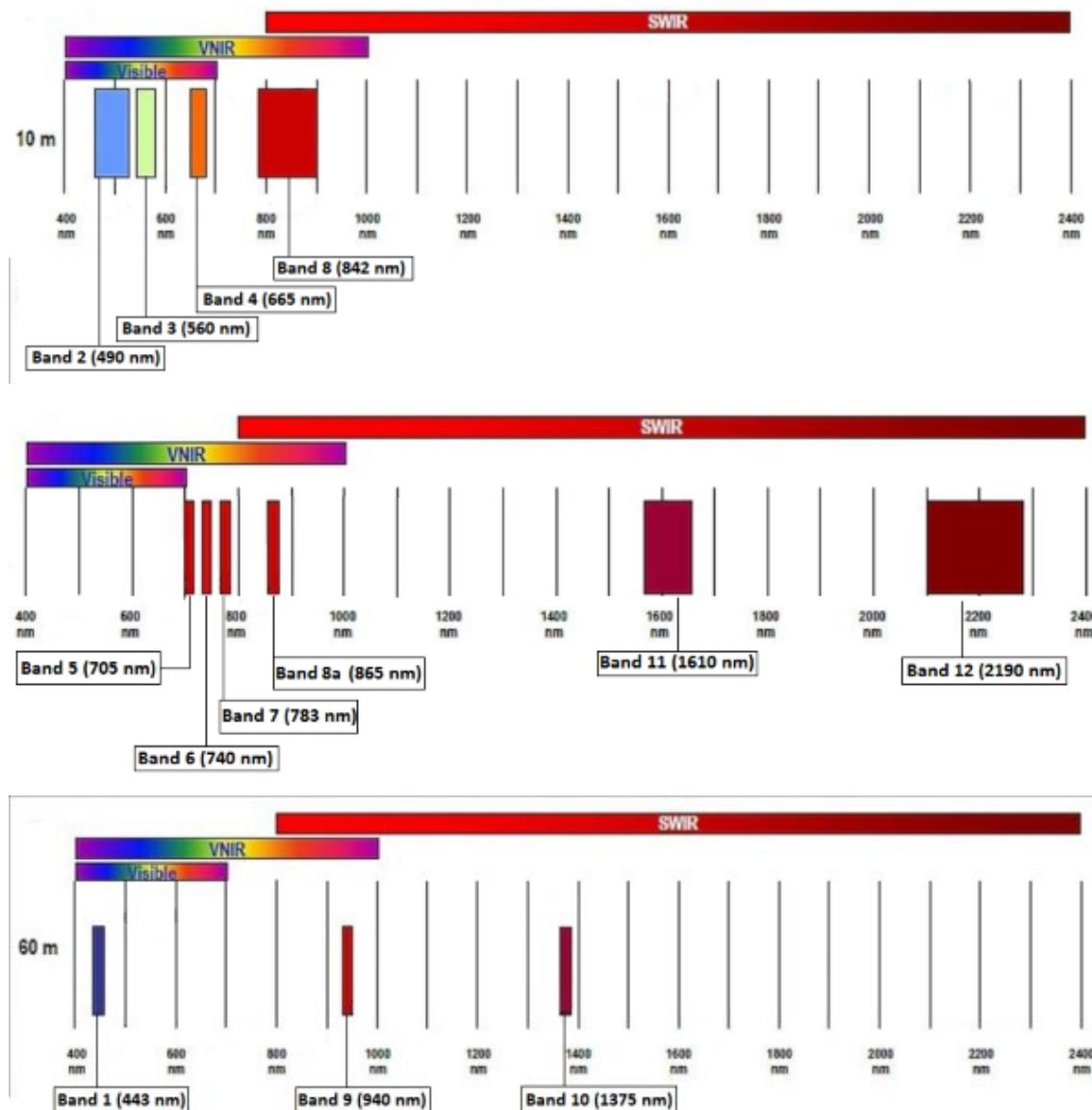
Fuente: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>

Las resoluciones de la misión Sentinel-2 son tres: resolución temporal, resolución radiométrica y resolución espacial:

La resolución temporal es la frecuencia de revisita de cada uno de los satélites Sentinel-2. Esta resolución temporal es de 10 días y la revisita combinada de la constelación es de 5 días.

La resolución radiométrica es el rango de reflectancia que puede ser medido por el instrumento del satélite. Cuanto mayor sea la resolución radiométrica, mayor será la capacidad del dispositivo para detectar diferencias de intensidad o reflectancia. En el caso del Sentinel-2 consta de 13 bandas espectrales, que van desde el espectro visible del azul (443 nm) al infrarrojo de onda corta (SWIR) (2190 nm).

Por último, la resolución espacial, que se refiere al tamaño del píxel en el terreno y que en el caso del Sentinel-2 depende de la bandapectral que se considere: cuatro bandas a 10 m, seis bandas a 20 m y tres bandas a 60 m de resolución espacial.

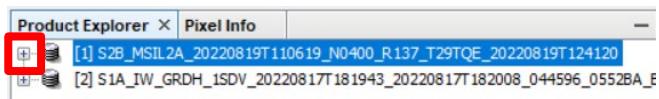


Sentinel-2

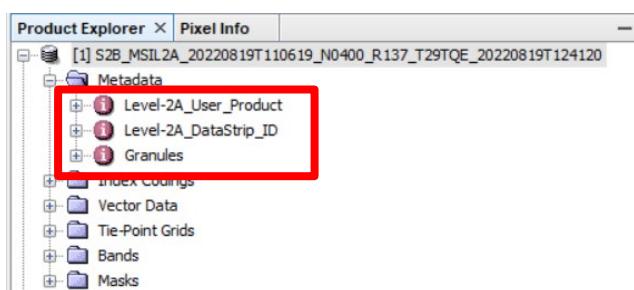
Espectro electromagnético	Longitud de onda central (nm)	Ancho de banda (nm)	Banda	Resolución espacial
VIS	443	21	B1	60 m
	490	65	B2	10 m
	560	35	B3	10 m
	665	30	B4	10 m
NIR	705	15	B5	20 m
	740	15	B6	20 m
	783	20	B7	20 m
	842	115	B8	10 m
	865	20	B8A	20 m
	945	20	B9	60 m
SWIR	1375	30	B10	60 m
	1610	90	B11	20 m
	2190	180	B12	20 m

Una vez descritas las características de las imágenes S2, vamos a explorar la imagen y ver la información que nos ofrecen sus metadatos.

Para ello, desplegamos la información de la imagen en el explorador de productos:



Comenzando con la primera carpeta de información, podremos observar los metadatos de la imagen.



Dentro de los metadatos de la imagen, existe infinidad de información. Destacaremos entonces el porcentaje de nubes:

El número de filas y columnas en función de la resolución de cada banda:

Product Explorer | Pixel Info

[1] S2B_MSIL1C_20220819T110619_N0400_R137_T29TQE_20220819T124120

- Metadata
 - Level-2A_User_Product
 - Level-2A_DataSrc_ID
 - Granules
 - Level-2A_Tile_29TQE
- Index Codings
- Vector Data
- Tie-Point Grids
- Bands
- Masks

[2] S1A_IW_GRDH_1SDV_20220817T181943_20220817T182008_044596_05528AE

Navigation | World View | Colour Manipulation | Uncertainty Visualis...

200 km

Sentinel 2 MSI Natural Colors RGB | Level-2A_User_Product | Granules | Level-2A_Tile_29TQE

Name	Value	Type	Unit	Description
General_Info				
Geometric_Info				
Tile_Geocoding				
Size				
resolution	10	asci		
NROWS	10980	asci		
NCOLS	10980	asci		
Size				
resolution	20	asci		
NROWS	5490	asci		
NCOLS	5490	asci		
Size				
resolution	60	asci		
NROWS	1830	asci		
NCOLS	1830	asci		
Geoposition				
Geoposition				
Geoposition				
metadataLevel	Brief	asci		
HORIZONTAL_CS_NAME	WGS84 / UTM zone 29N	asci		
HORIZONTAL_CS_CODE	EPSG:32629	asci		
Tile_Angles				
Quality_Indicators_Info				

La cantidad de píxeles de cada clase, según la clasificación previa:

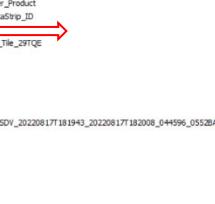
Product Explorer × Pixel Info

[1] S2B_MSIL2A_20220819T110619_N0400_R137_T29TQE_20220819T124120

- Metadata
 - Level-2A_User_Product
 - Level-2A_Datastrip_ID
 - Granules
 - Level-2A_Tile_29TQE
- Index Codings
- Vector Data
- Tie-Point Grids
- Bands
- Masks

[2] S1A_IW_GRDH_1SDV_20220817T181943_20220817T182008_044596_0552BA_E

Navigation World View × Colour Manipulation Uncertainty Visual... >



SENTINEL-2 MSI Natural Colors RGB

Level-2A_User_Product

[1] Granules

[1] Level-2A_Tile_29TQE

[1] Granules (2)

Name Value Type Unit Description

LEVEL_2A_TILE_29TQE			
GENERAL_INFO			
GEOMETRIC_INFO			
QUALITY_INDICATORS_INFO			
IMAGE_CONTENT_QI			
CLOUDY_PIXEL_PERCENTAGE	0.004097	asci	
CLOUDY_PIXEL_OVER_LAND_PERCENTAGE	0.000411	asci	
DEGRADED_MSI_DATA_PERCENTAGE	0.010600	asci	
NODATA_PIXEL_PERCENTAGE	0.000003	asci	
SATURATED_DEFECTIVE_PIXEL_PERCENTAGE	0.000000	asci	
DARK_FEATURES_PERCENTAGE	0.135365	asci	
CLOUD_SHADOW_PERCENTAGE	0.000000	asci	
VEGETATION_PERCENTAGE	31.771761	asci	
NOT_VEGETATED_PERCENTAGE	66.837865	asci	
WATER_PERCENTAGE	1.097276	asci	
UNCLASSIFIED_PERCENTAGE	0.105454	asci	
MEDIUM_PROBA_CLOUDS_PERCENTAGE	0.003344	asci	
HIGH_PROBA_CLOUDS_PERCENTAGE	0.000152	asci	
THIN_CIRRUS_PERCENTAGE	0.0000362	asci	
SNOW_ICE_PERCENTAGE	0.000000	asci	
RADIATIVE_TRANSFER_ACCURACY	0.0	asci	
WATER_VAPOUR_RETRIEVAL_ACCURACY	0.0	asci	
AOT_RETRIEVAL_ACCURACY	0.0	asci	
AOT_RETRIEVAL_METHOD	SEN2COR_DDV	asci	
GRANULE_MEAN_AOT	0.183382	asci	
GRANULE_MEAN_WV	1.277513	asci	
OZONE_SOURCE	AUX_ECMWF	asci	
OZONE_VALUE	300.937531	asci	

Pixel_Level_QI

metadatalevel Standard ascii

PVI_FILERNAME GRANULE/2A_T29TQE_A028477_20220819T1108*

O el sistema de coordenadas:

Name	Value	Type	Unit	Description
HORIZONTAL_CS_NAME	WGS84 / UTM zone 29N	ascii		
HORIZONTAL_CS_CODE	EPSG:32629	ascii		

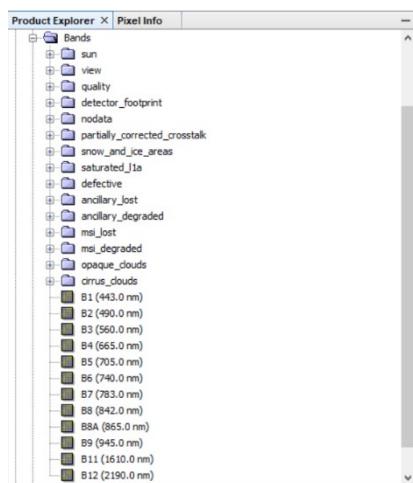
A continuación, tenemos los códigos de indexación, que corresponden con la calidad de la clasificación de la escena.

Name	Value	Type	Unit	Description
NODATA	0	uint32		No data
SATURATED_DEFECTIVE	1	uint32		Saturated or defective
TOPOGRAPHIC_AND_CASTED_SHADOWS	2	uint32		topographic and casted shadows
CLOUD_SHADOW	3	uint32		Cloud shadow
VEGETATION	4	uint32		Vegetation
NOT_VEGETATED	5	uint32		Not vegetated
WATER	6	uint32		Water
UNCLASSIFIED	7	uint32		Unclassified
CLOUD_MEDIUM_PROBABILITY	8	uint32		Cloud (medium probability)
CLOUD_HIGH_PROBABILITY	9	uint32		Cloud (high probability)
THIN_CIRRUS	10	uint32		Thin cirrus
SNOW_ICE	11	uint32		Snow or ice

Dichos códigos corresponden con las clases de una preclasificación general que viene por defecto en la imagen S2.

La tercera carpeta, contiene los vectores asociados a la imagen, y en nuestro caso, aún no hay elementos aún.

La quinta, corresponde con las bandas de la imagen:



A primera vista, podemos ver las 12 bandas y sus correspondientes longitudes de onda. Cada banda tiene una resolución espacial diferente:

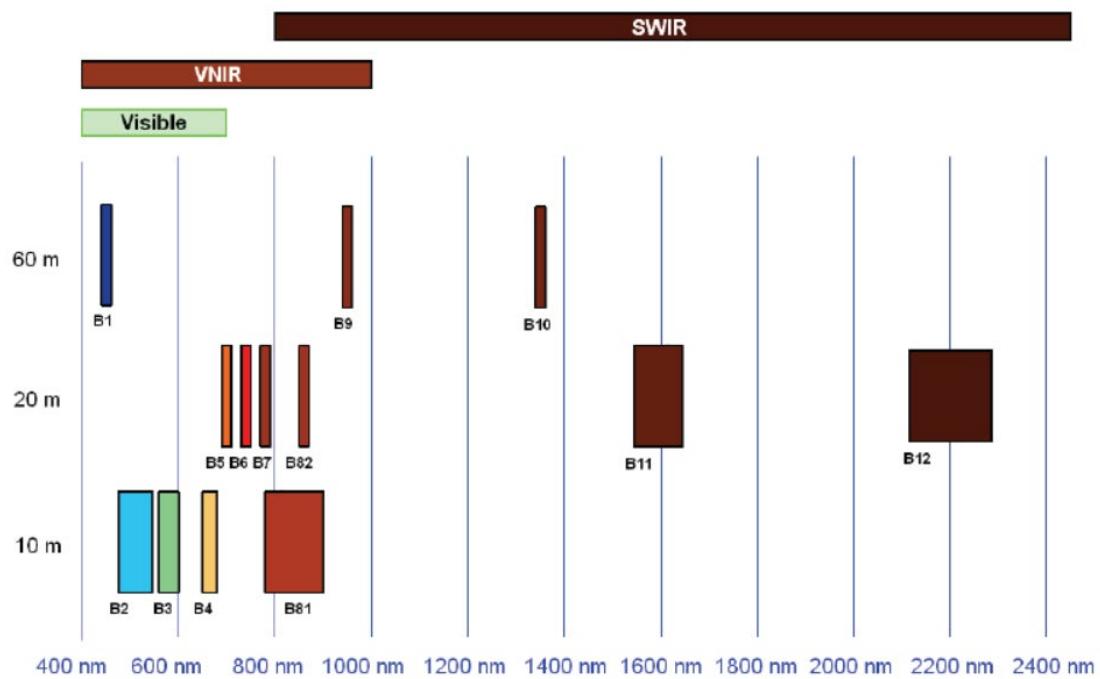
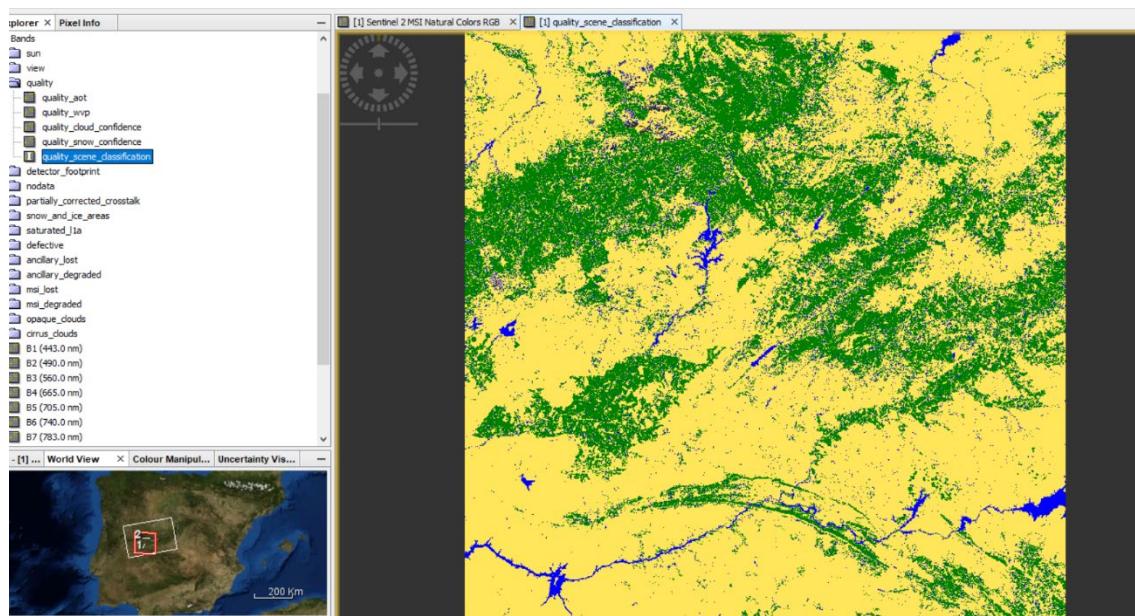


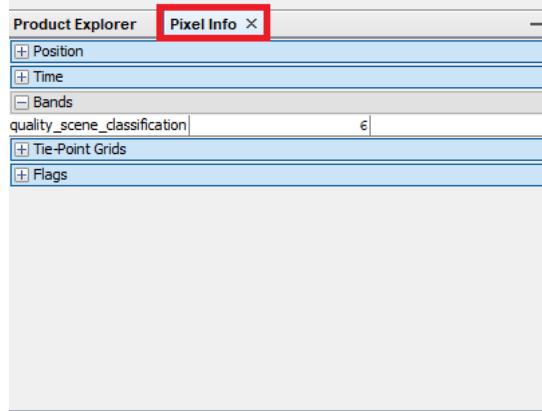
Figure 3-1 – Sentinel-2 Spectral Bands and Resolutions

Fuente: S2-PDGS-MPC-L2A-SUM-V2.8

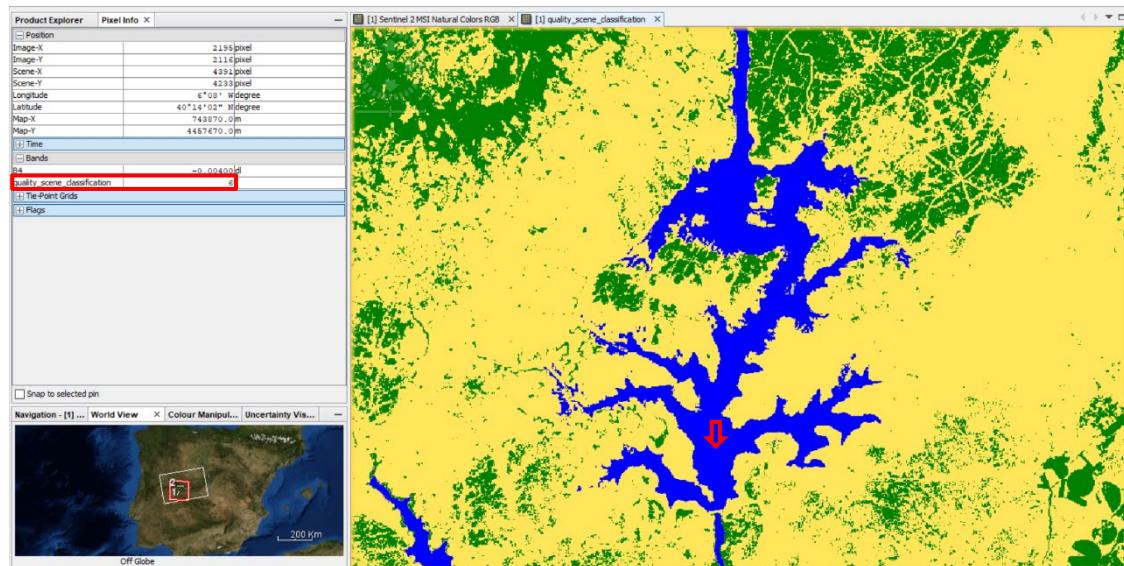
Además de esas bandas existen otras, como las referentes al estado del sol en el momento de la toma, la de las vistas para cada banda y las de calidad. De las tres, la que más nos interesa es la carpeta de calidad y en especial, la de clasificación de la imagen. Si la abrimos haciendo doble click, nos aparecerá la clasificación:



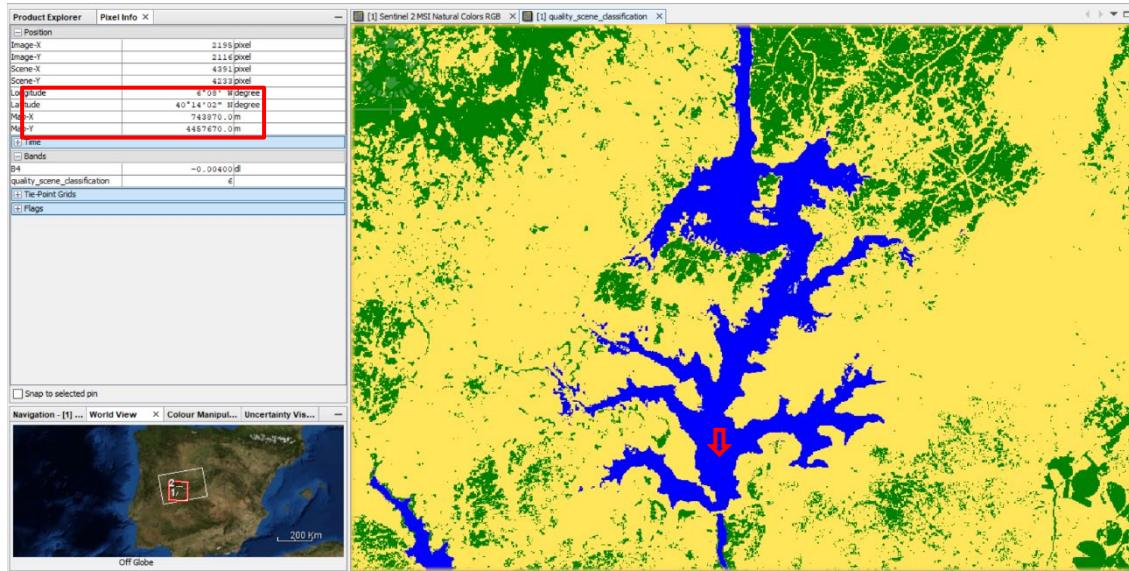
Para poder consultar la clasificación, nos iremos a la pestaña de información del píxel y podremos ver el código de cada clase, que corresponderá con la clase de los códigos de indexación explicados anteriormente:



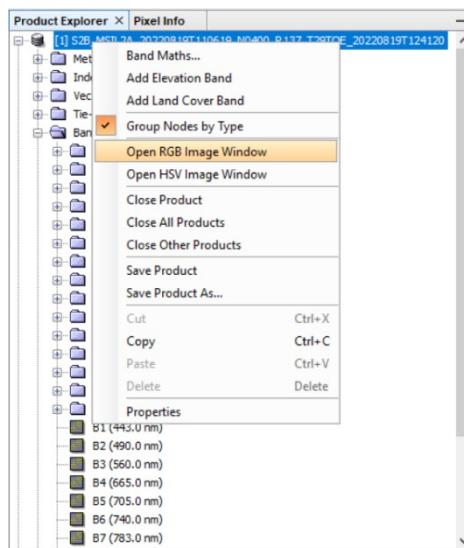
Si queremos ver una zona con más detalle, moveremos la rueda del ratón hacia nosotros y podremos ir comprobando los valores de los píxeles. Por ejemplo, el agua con valor 6.



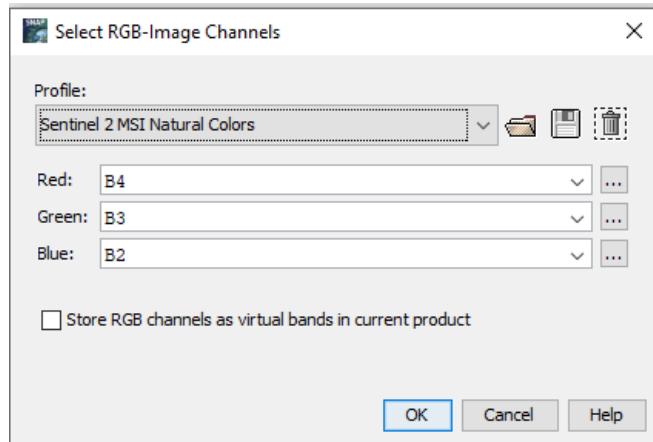
También podría ser interesante explorar las coordenadas del pixel en cuestión



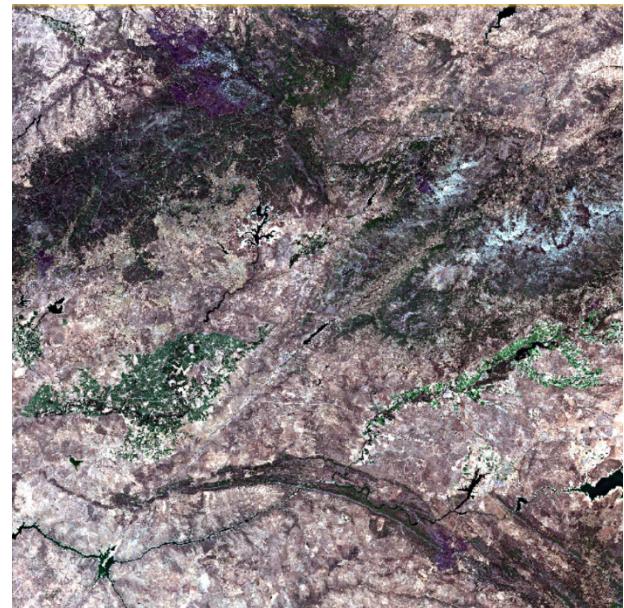
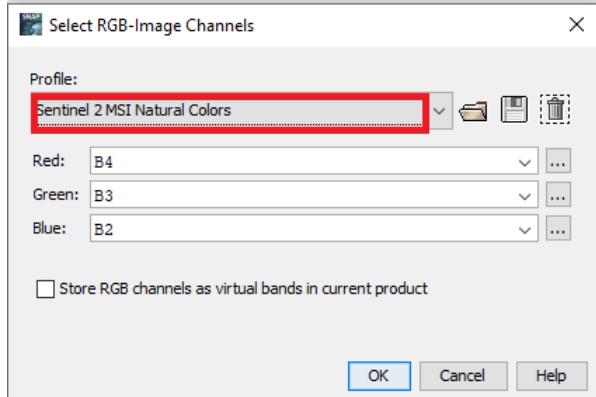
Para ver las bandas de la imagen, podríamos ir abriendo una por una, pero hay una forma de visualización predeterminada para combinarlas y verlas a color. De este modo, si pinchamos encima de la imagen y pulsamos el botón derecho del ratón, nos aparecerá el siguiente menú:



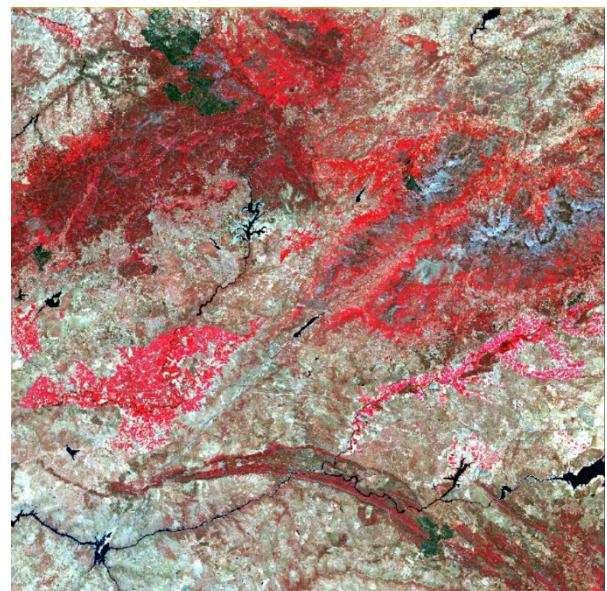
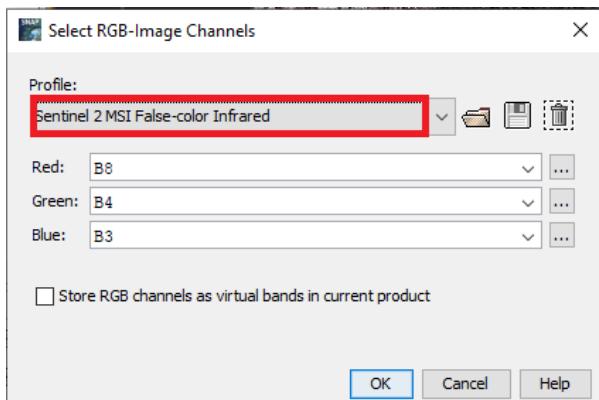
En este caso, podríamos indicarle que nos abra la imagen en RGB. Nos solicitará qué combinación de bandas queremos realizar. Existen una serie de combinaciones predefinidas en el desplegable.



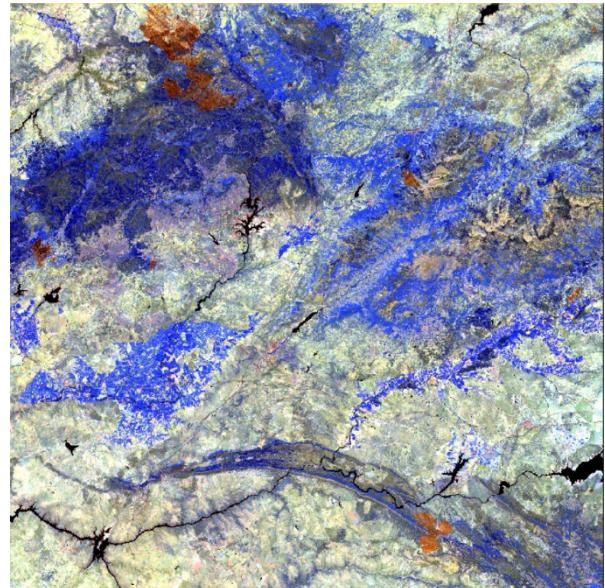
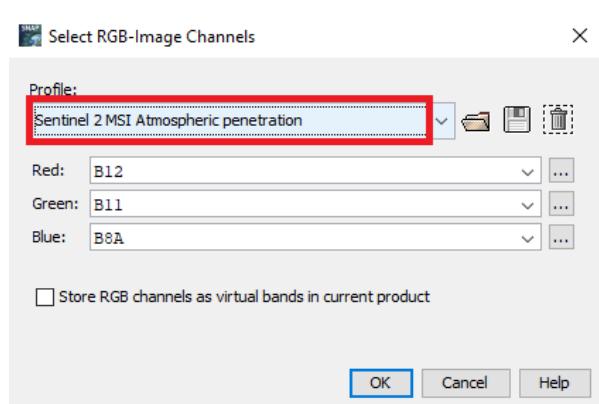
Podremos por tanto ver la imagen en color natural:



En falso color infrarrojo:

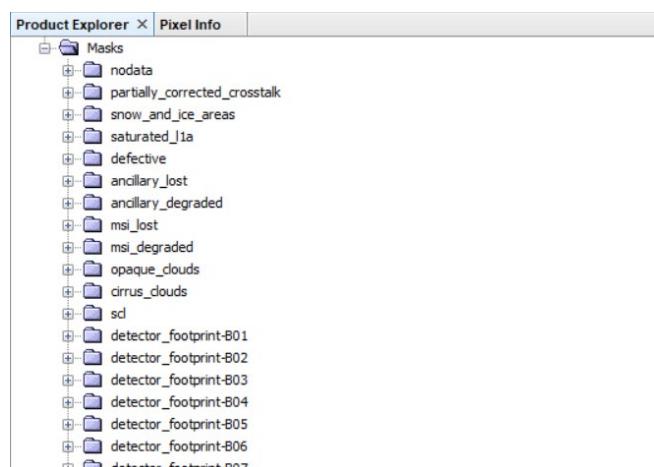


O en modo penetración atmosférica:

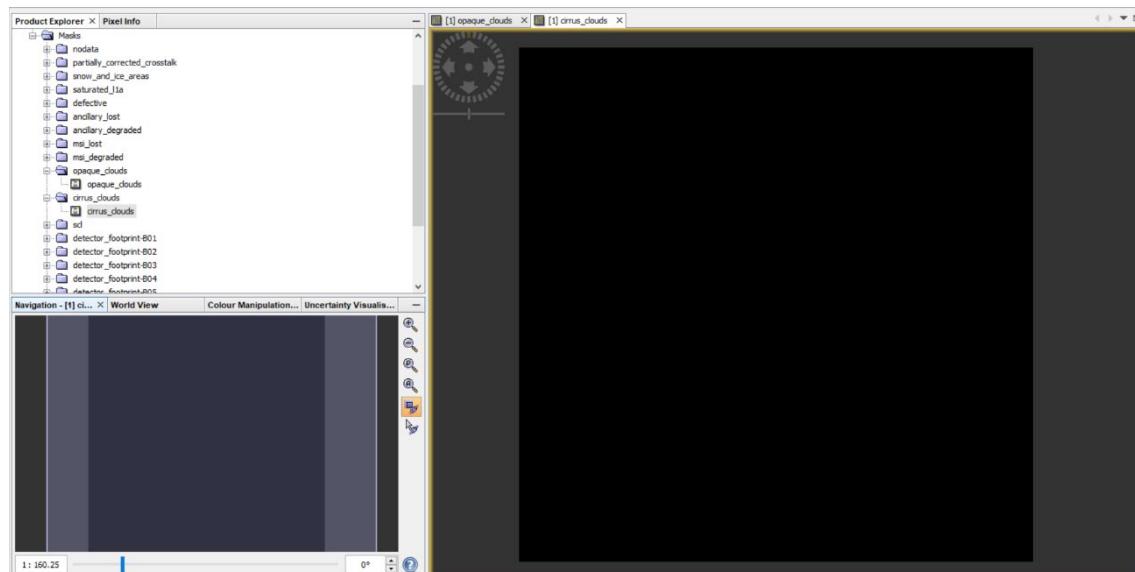


Aunque también existe la posibilidad de asignar los canales del RGB.

Finalmente, podremos ver también que existe una carpeta con máscaras:



De las que destacaremos las de nubes, que en nuestra imagen no hay:

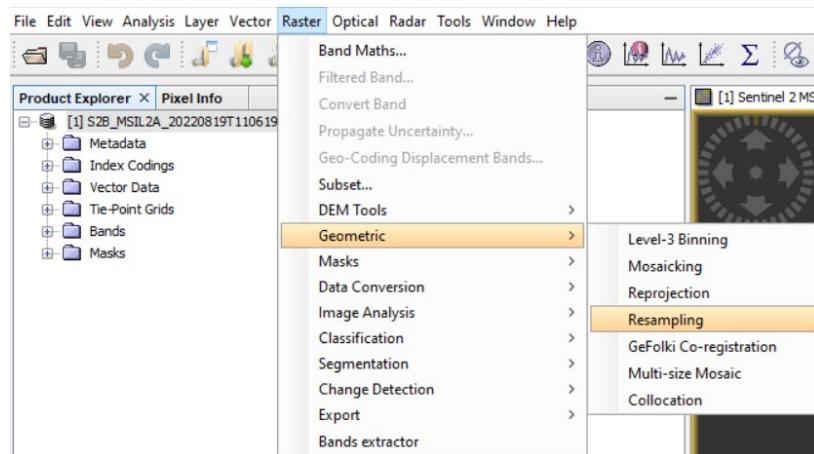


Si la máscara de nubes indicara presencia en nuestra área de estudio, tendríamos la información suficiente como para borrar esa parte de la imagen.

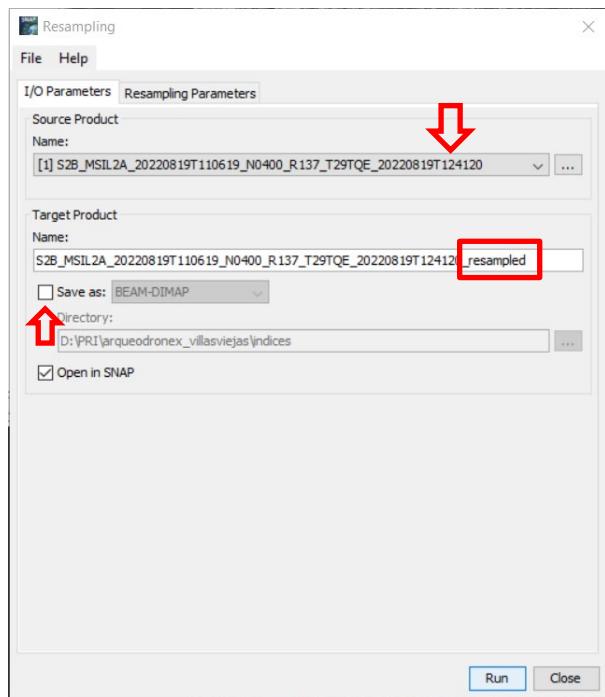
1.2 Preprocesado de la imagen S2

Como ya sabemos, la imagen S2 tiene distintas resoluciones en función de cada banda. Esta característica nos puede dificultar ciertas operaciones, por lo que, lo primero que vamos a hacer es remuestrear todas las bandas a la mínima resolución de 10 m.

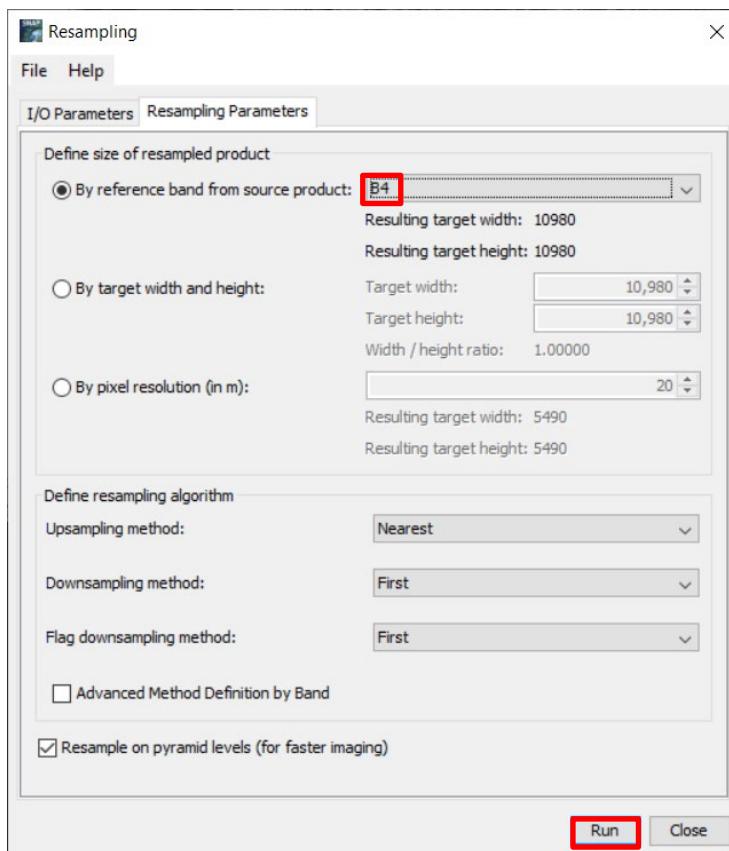
Para ello, una vez marcada la imagen S2, nos vamos a la opción remuestrear:



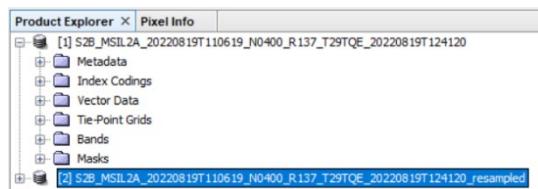
Nos aparecerá el menú de remuestreado, en el que podremos distinguir varias opciones. En la solapa de los parámetros I/O (*In/Out*), podremos indicar el nombre de la imagen de entrada, que, si hemos marcado con anterioridad la imagen S2, vendrá ya seleccionada. Si no lo hubiéramos hecho, podremos seleccionarla del desplegable. En segundo lugar, hay que indicar el archivo de salida. SNAP por defecto añade siempre sufijos a los nombres de las imágenes de entrada, en cada operación que realiza. En este caso, le pone el sufijo de *_resampled*. Otro aspecto a destacar de SNAP, es que puede trabajar con resultados virtuales, de tal manera que no hay que guardar los archivos de operaciones intermedias si no se desea, facilitando mucho la liberación de espacio en el equipo de trabajo. Si quisieramos guardar el resultado hay que indicarlo en la casilla correspondiente y darle un nombre, que no es nuestro caso.



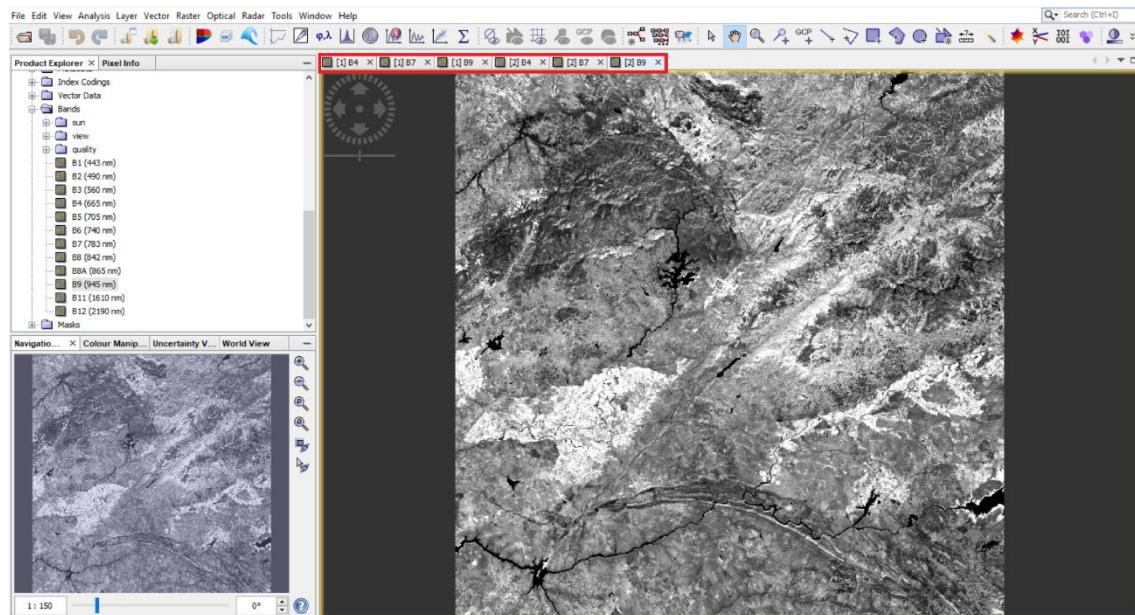
En la solapa de parámetros de remuestreado (*Resampling Parameters*), se configurará el modo de cálculo. Así, podremos remuestrear copiando la distribución de píxeles de otras bandas de la propia imagen, por un número de píxeles de salida en ancho y alto, o seleccionando el tamaño de píxel requerido. Por otro lado, podremos también configurar el algoritmo de interpolación. En nuestro caso, vamos a indicar que queremos que nos remuestree siguiendo la distribución de píxeles de una de las bandas de 10 m de resolución (B4) y dejaremos el resto de los parámetros por defecto.



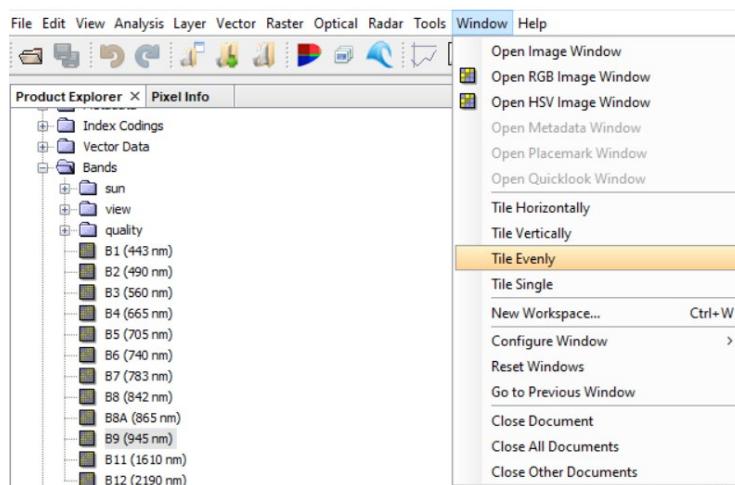
El resultado aparecerá en el explorador de productos y entonces, cerraremos la ventana de remuestreado y podremos comprobar el resultado.

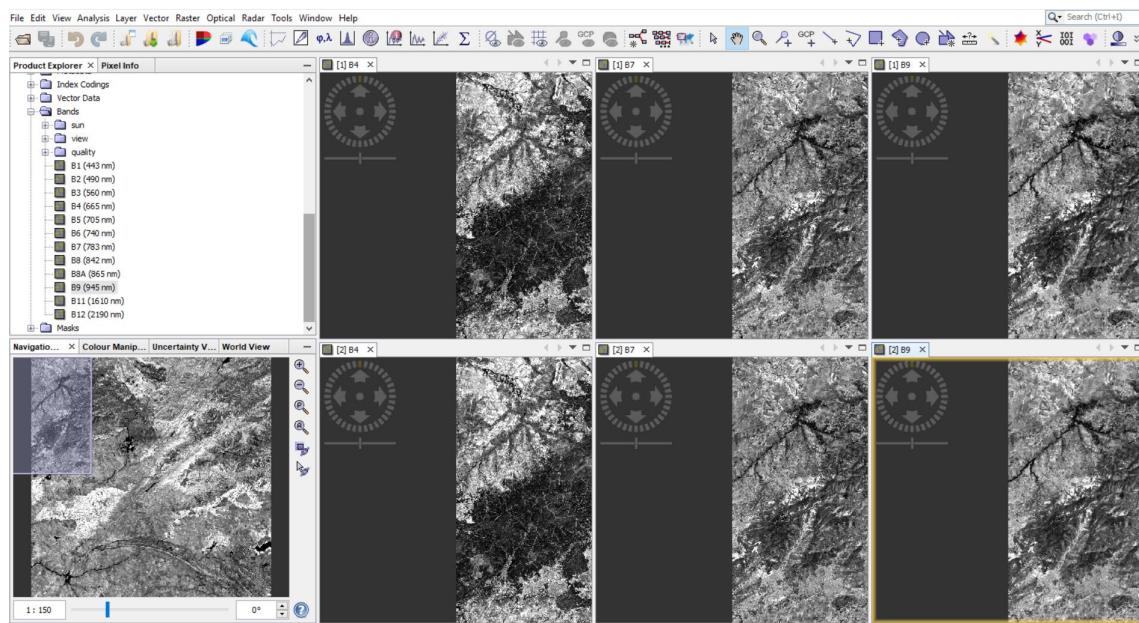


Para comprobarlo, vamos a abrir varias ventanas a la vez para ver las distintas bandas y vamos a vincularlas. Abrimos entonces las bandas 4, 7 y 9 de la imagen original y de la imagen remuestreada.

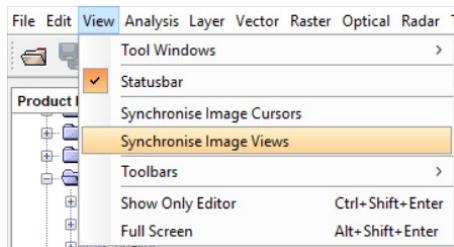


Y ahora las mostramos en modo de mosaico uniforme

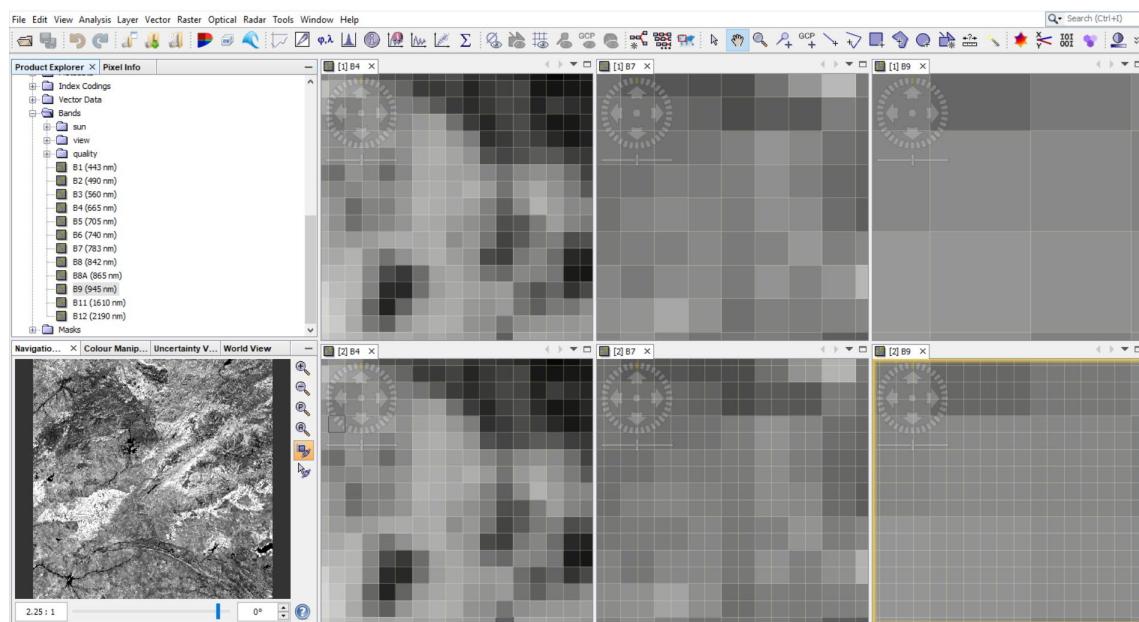




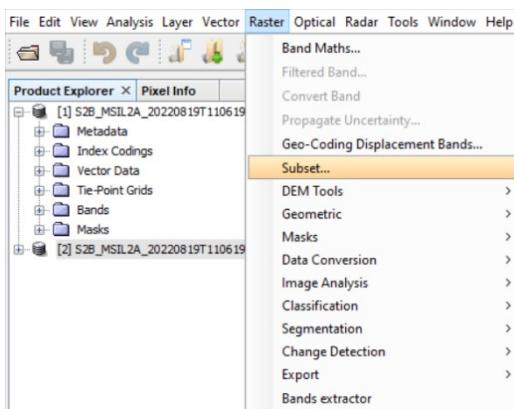
Si queremos vincular las ventanas de las vistas entre ellas, tendremos que configurarlo en el menú de vista o en la ventana de navegación.



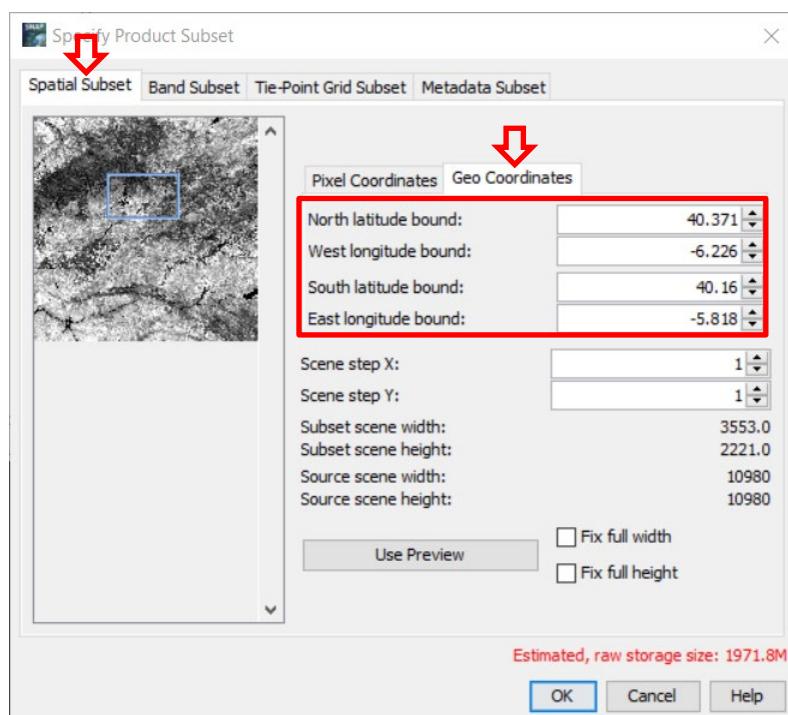
En ese momento, podremos hacer un zoom en la ventana “directora” de tal forma que lleguemos a apreciar los distintos tamaños de píxeles entre bandas en la imagen sin remuestrear y la igualdad en el caso de la imagen remuestreada.



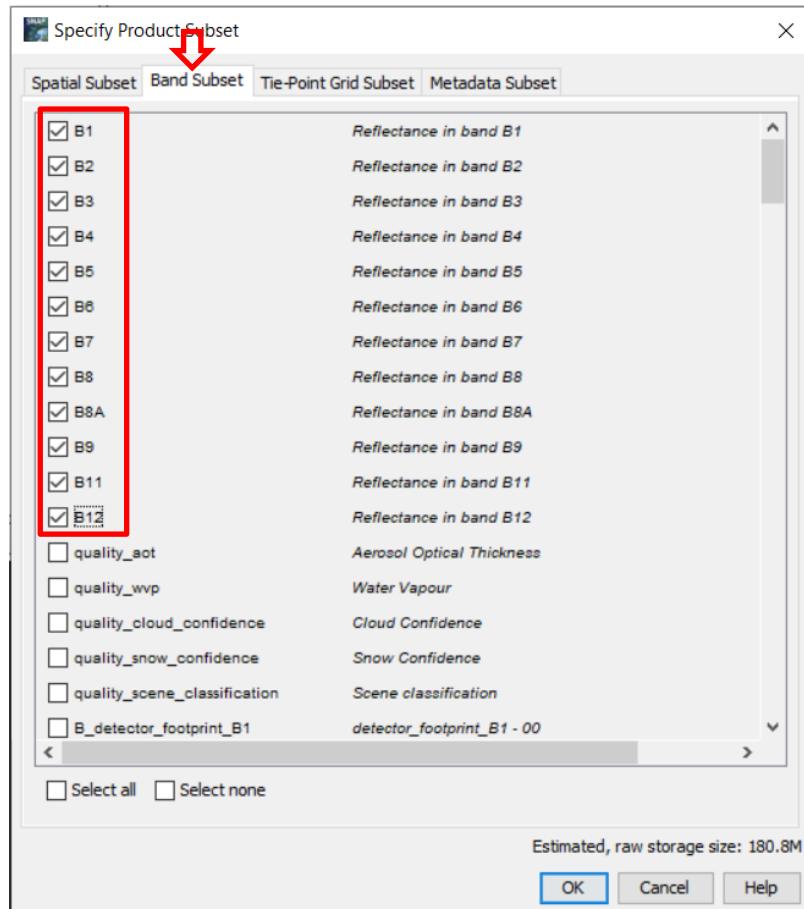
Con el fin de no consumir muchos recursos de memoria y hacer las siguientes operaciones con mayor celeridad, vamos a recortar una zona pequeña de la imagen. Para ello utilizaremos el comando *subset*.



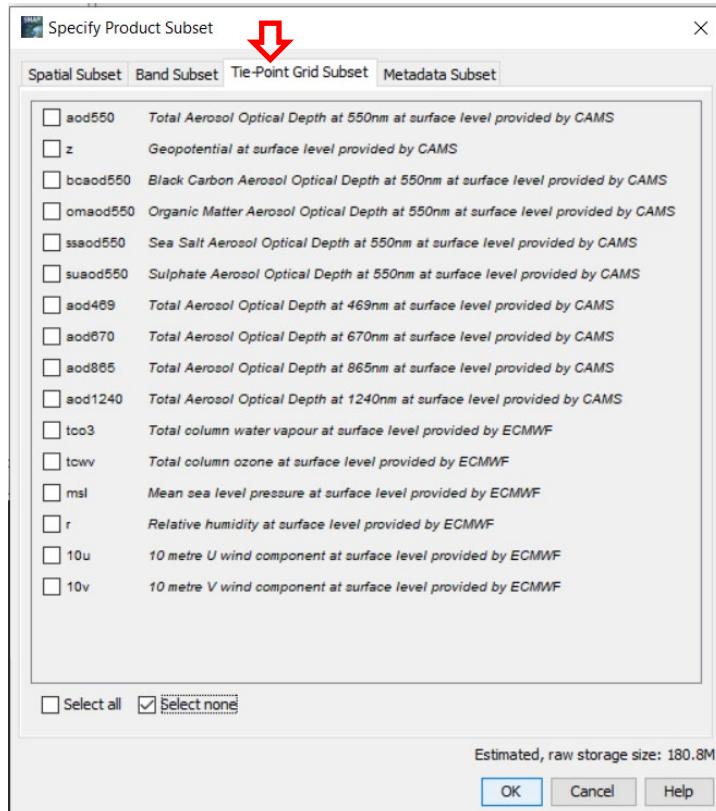
Configuraremos la ventana de recorte con las siguientes coordenadas:



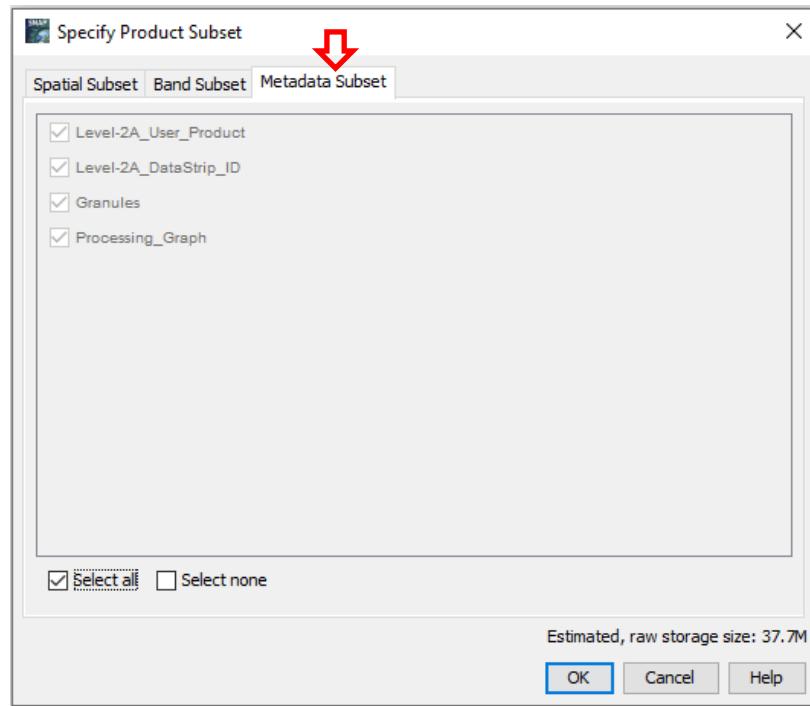
En la solapa de bandas, también vamos a indicarle que sólo nos mantenga las bandas espectrales:



Y no seleccionamos ningún dato de la grid de puntos

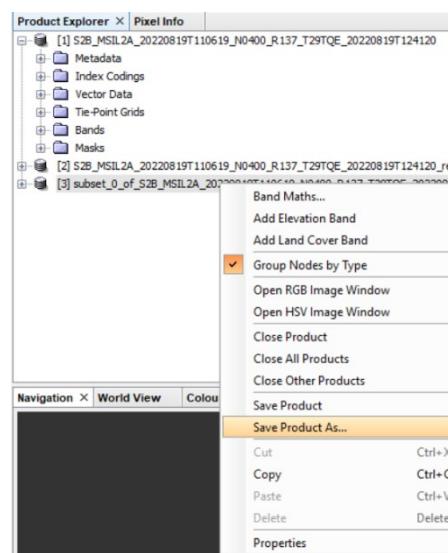


Y, finalmente, dejamos la configuración de la solapa de metadatos como está, para que nos los mantenga también.

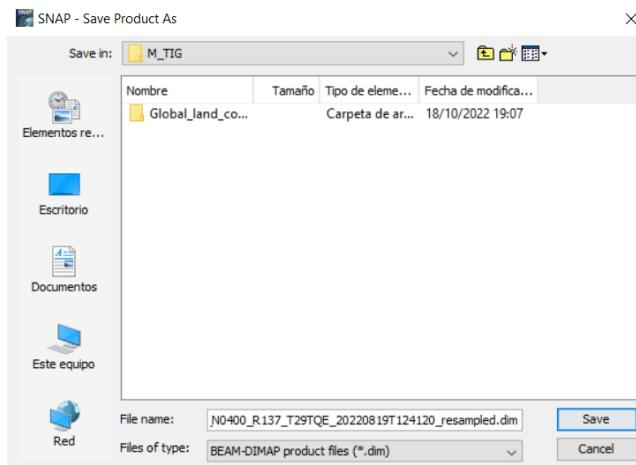


Le damos a calcular y aparecerá una nueva imagen en el explorador de productos con un prefijo *subset_0_of_*. Guardaremos esa imagen, que será con la que vamos a trabajar a partir de ahora.

Para guardarla, tendremos que marcarla y darle botón derecho del ratón e indicar que la queremos guardar como imagen.

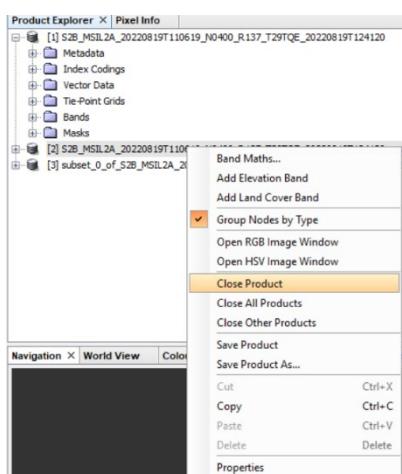


Nos saldrá un aviso de que va a tardar varios minutos y aceptamos. Nos pedirá entonces el nombre y ubicación.



Se guarda en formato *.dim*, que es el nativo de SNAP.

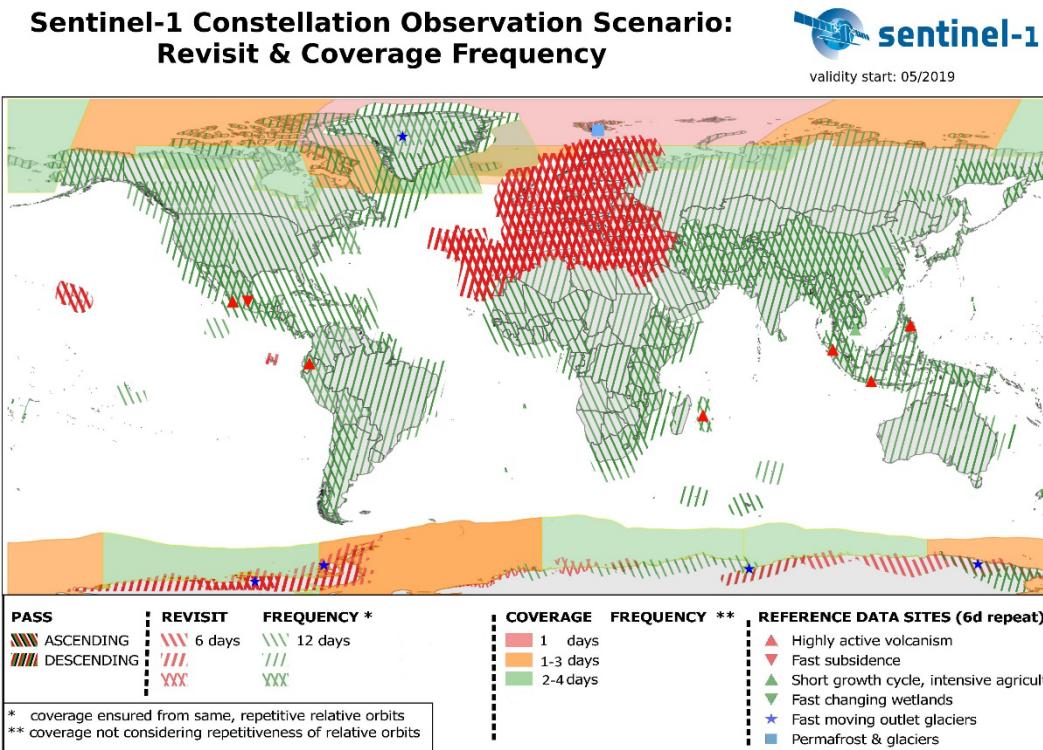
Podríamos ya cerrar la imagen “virtual” intermedia que nos ha llevado hasta aquí. Para ello, la marcaremos y con el botón derecho del ratón seleccionaremos la opción de cerrar producto.



Podemos observar cómo se nos ofrece la posibilidad de cerrar a la vez todos los productos o, incluso, seleccionar sólo la imagen que queremos dejar y cerrar el resto.

1.3 Preprocesado de la imagen S1

La misión Sentinel-1 es una constelación formada por dos satélites de órbita polar (hasta diciembre de 2021), que operaban día y noche realizando imágenes de radar de apertura sintética (SAR) en la banda C. Actualmente sólo existe uno que proporciona datos con polarización dual y cuenta con tiempos de revisita muy cortos.

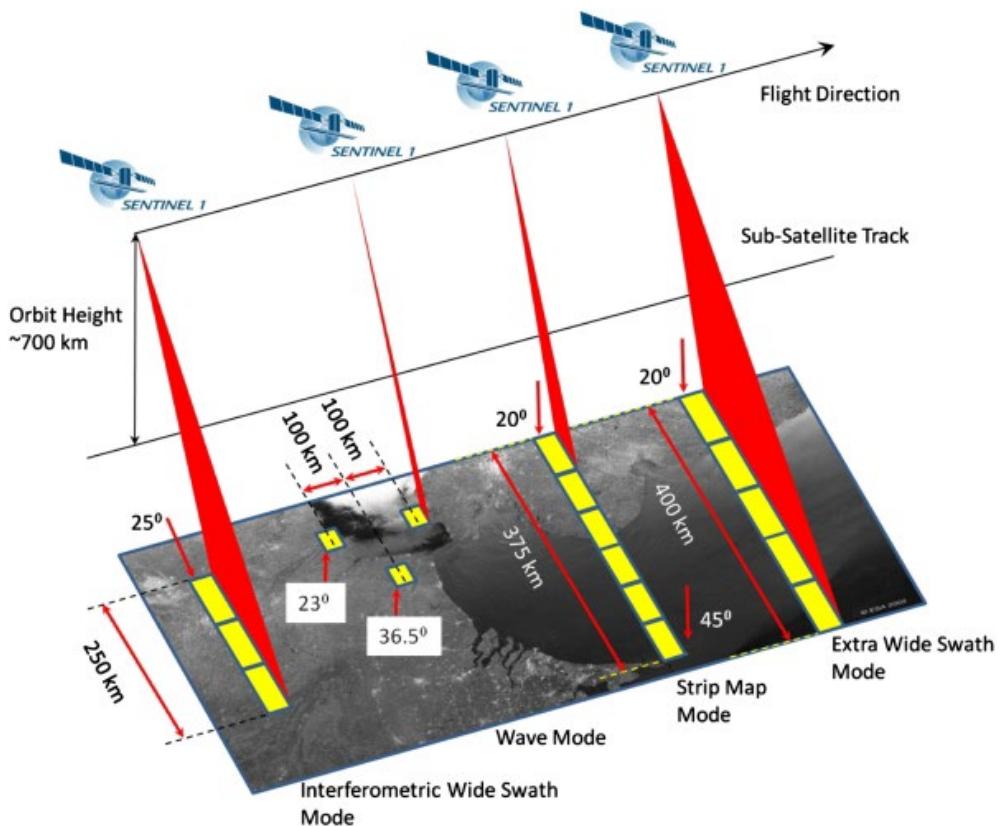


Fuente: <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>

El Radar de Apertura Sintética (SAR) tiene la ventaja de no estar limitado por la cobertura de nubes o la falta de iluminación y puede adquirir datos sobre un sitio durante el día o la noche en todas las condiciones meteorológicas. Dentro del plan de observación, la constelación Sentinel-1, garantiza observaciones para los siguientes servicios: vigilancia marítima, vigilancia terrestre y gestión de emergencias.

Sentinel-1 opera en cuatro modos de adquisición de datos:

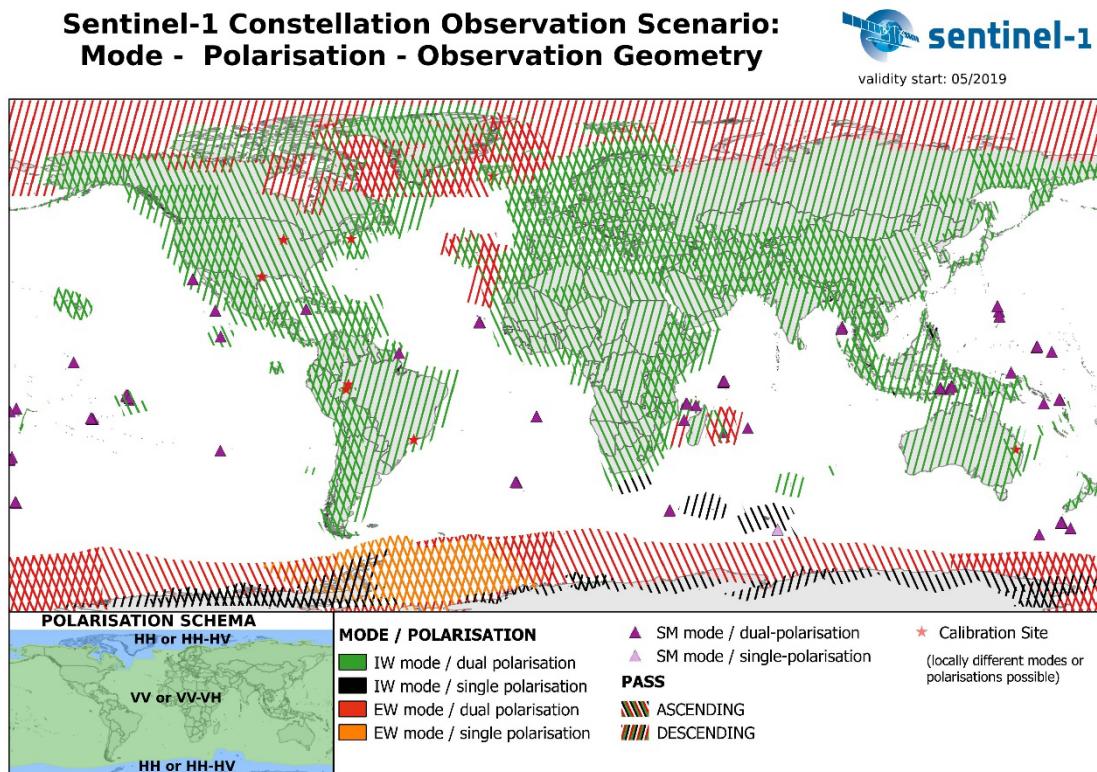
- Franja o Stripmap (SM): Proporciona datos con una resolución espacial de 5x5 m y una escena de 80 km.
- Interferometría de gran amplitud o Interferometric Wide swath (IW): Proporciona datos con una resolución espacial de 5x20 m y una escena de 250 km.
- Franja extra ancha o Extra-Wide swath (EW): Se emplea sobre zonas marítimas y polares, donde se necesita una gran cobertura y tiempos de revisita cortos. Proporciona datos con una resolución espacial de 20x40 m y una escena de 400 km.
- Onda o Wave (WV). Se utiliza para medir la dirección y altura de olas en el océano. Las imágenes tienen una resolución de 20x20km con dos ángulos de incidencia cada 100 km.



Fuente: <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>

Como ya se ha mencionado, el modo de Onda (WV) diseñado para capturar información sobre los océanos. Los modos de ondas interferométricas de gran amplitud (IW) y de ondas extra anchas (EW) operan sobre zonas las siguientes zonas geográficas predefinidas:

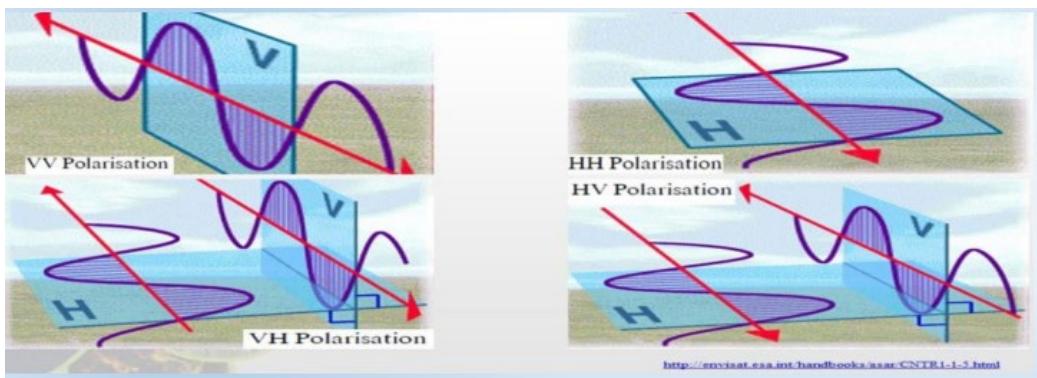
- sobre las zonas terrestres y costeras: el modo predefinido es IW.
- sobre los mares y las zonas polares, y las zonas relevantes para el océano, el modo predefinido es IW o EW.



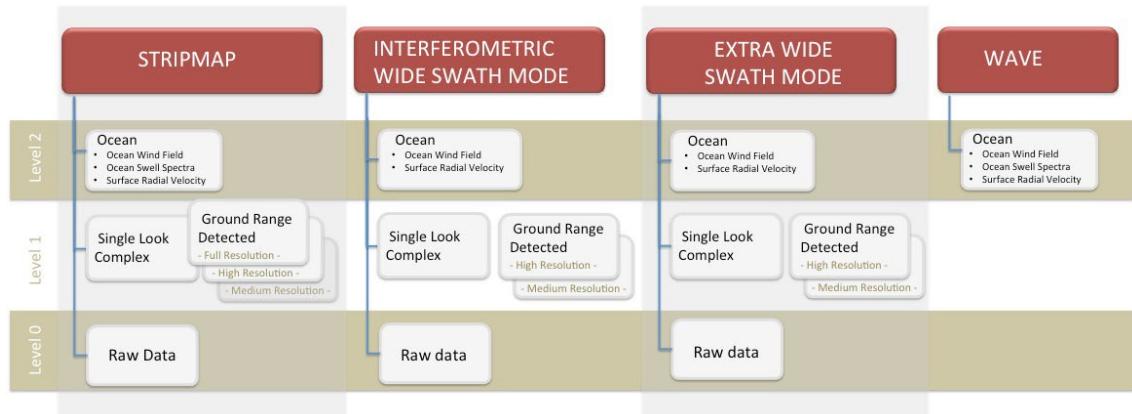
Fuente: <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>

En tierra, se utiliza sistemáticamente la misma polarización SAR en una zona determinada, para garantizar series cronológicas de datos en las mismas condiciones. Dependiendo de la zona, la selección es vertical u horizontal, y el tipo se selecciona en función de la aplicación principal del producto. Como principio general, el esquema de polarización utilizada es el siguiente:

- Polarización HH-HV o HH para la vigilancia de los polos y zonas de hielo marino.
- Polarización VV-VH o VV para todas las demás zonas de observación (con la excepción del Mar Báltico que también es observado parcialmente en HH-HV con el Sentinel-1B durante el invierno boreal).



Todos los Sentinel-1 adquiridos en los modos IW, EW y SM se procesan para convertirlos en productos GRD de nivel 0 y nivel 1.



Fuente: <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>

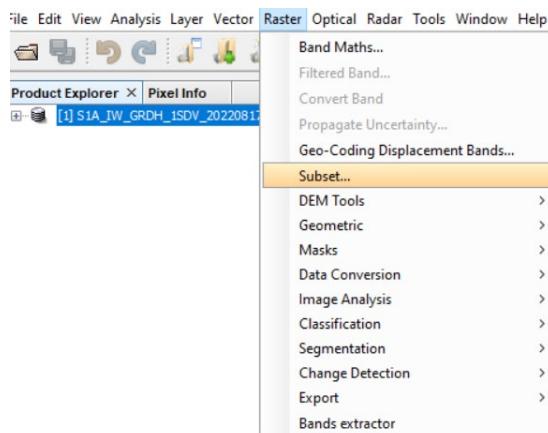
Especificamente para el nivel 1, los productos derivados son el Single Look Complex (SLC) y Ground Range Detected (GRD).

Los productos SLC consisten en datos de SAR georeferenciados utilizando datos de órbita y actitud del satélite y proporcionados en geometría de rango de inclinación de cero Doppler. Los productos incluyen un aspecto único en cada dimensión utilizando todo el ancho de banda de la señal de transmisión y consisten en muestras complejas que preservan la información de fase.

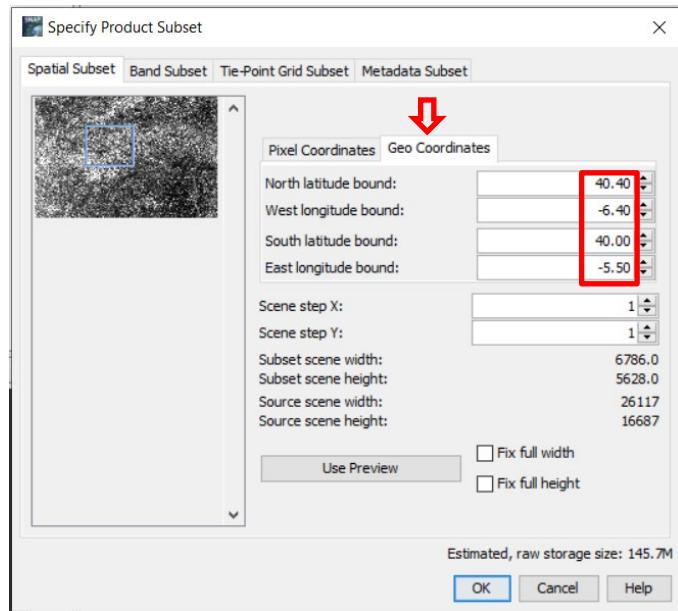
Los productos GRD consisten en datos de SAR proyectados a nivel del suelo usando un modelo de elipsode terrestre. Los productos GRD pueden estar en una de tres resoluciones: Full Resolution (FR), High Resolution (HR) y Medium Resolution (MR). Los productos GRD de nivel 1 están disponibles en MR y HR para los modos IW y EW, MR para el modo WV y MR, HR y FR para el modo SM.

Las imágenes S1 necesitan un procesado diferente del de las S2, ya que no se encuentran corregidas desde un principio.

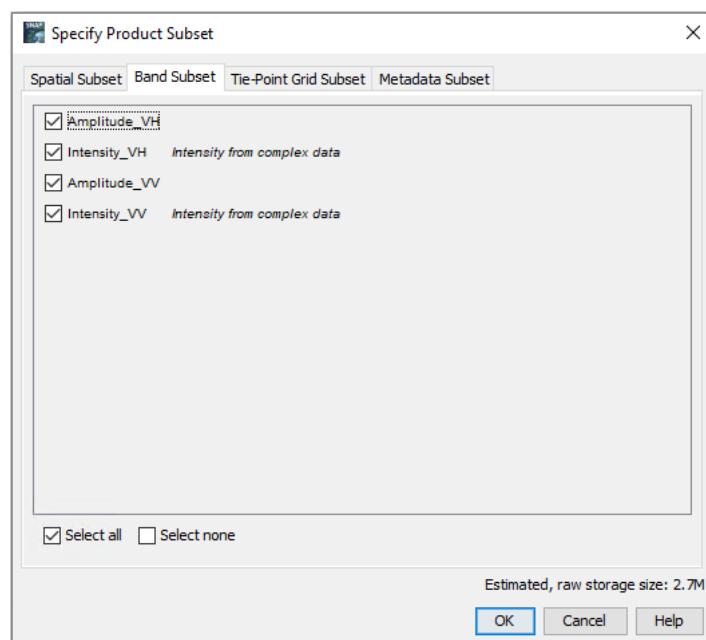
Lo primero que vamos a realizar en este caso es el recorte de la imagen, puesto que las imágenes S1 son más grandes y el resto de procesos nos llevarían mucho tiempo.

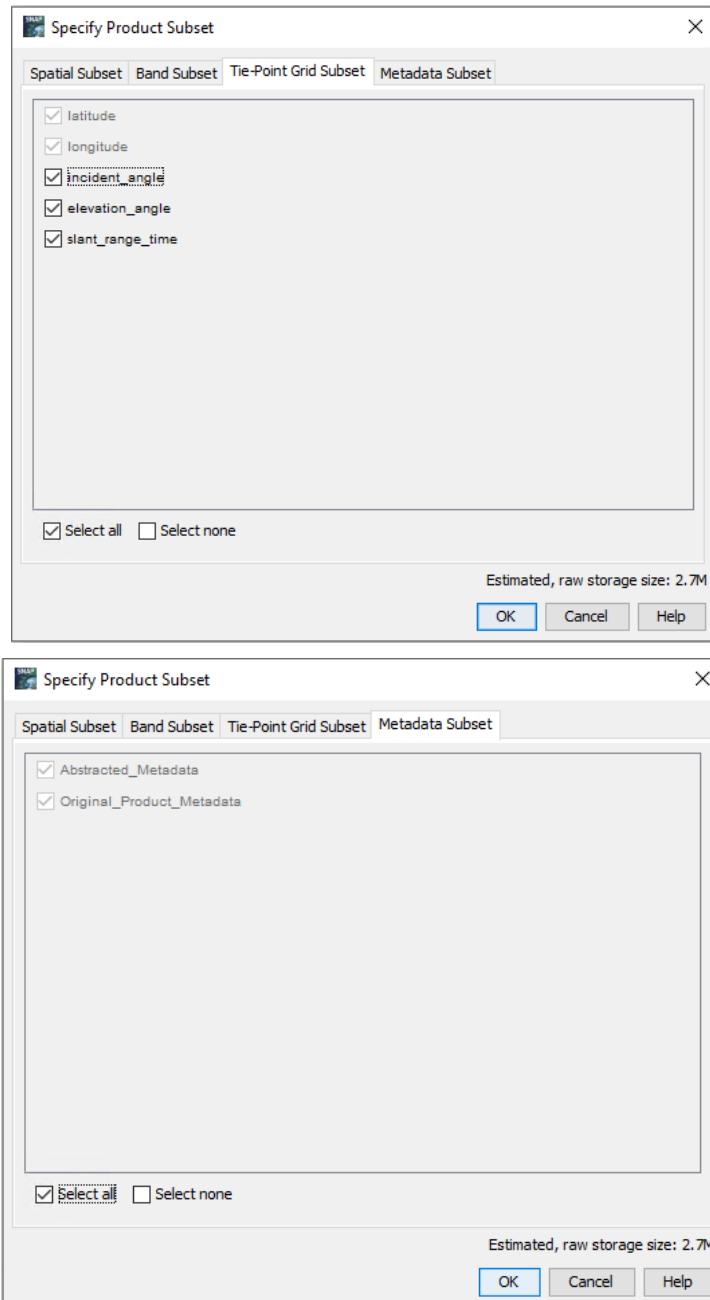


Recortaremos entonces por una zona más grande de la que le dimos a la imagen S2, ya que luego al proyectarla, se gira según la proyección UTM y no cuadraría con el tamaño exacto de la S2:



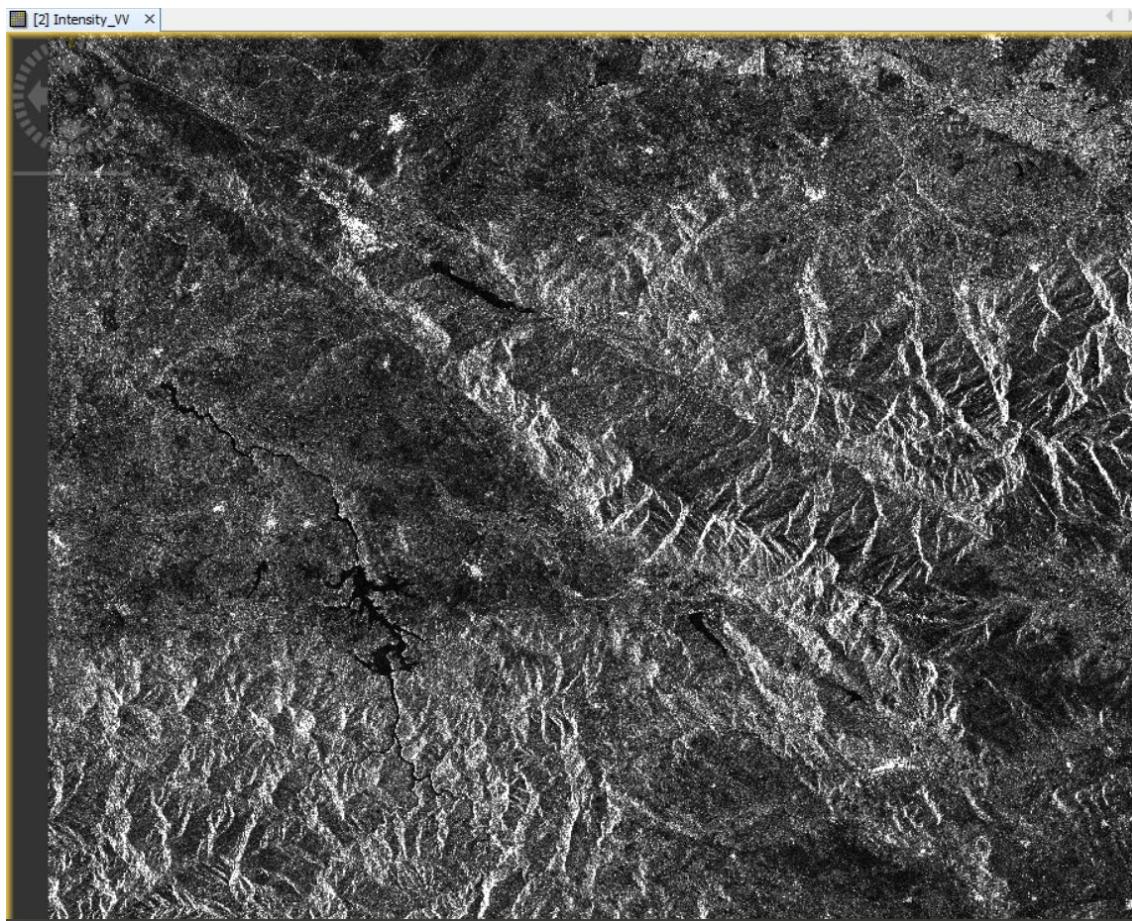
En el resto de solapas no cambiaremos nada.



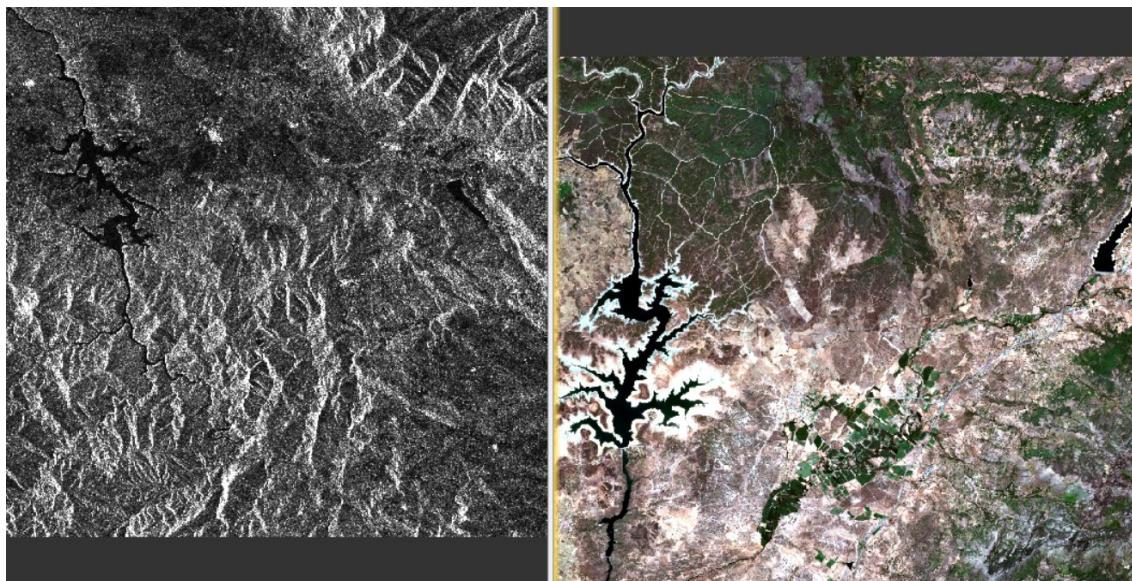


Habrá creado entonces una imagen con el prefijo *subset_0_of_*, de la que abriremos una banda para visualizarla (si no cerramos con anterioridad el proyecto y seguimos trabajando sobre el mismo, el *subset* será el 1, porque guarda en memoria las operaciones anteriores).

Si abrimos, por ejemplo, la banda de intensidad de la polarización VV, podremos observar la imagen en su estado actual.



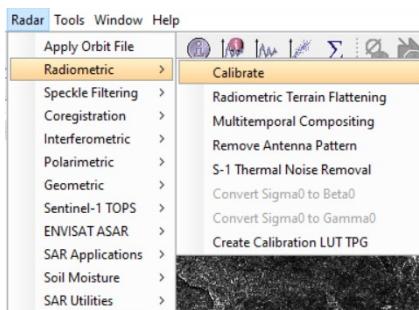
Si la comparamos con la imagen del S2 (a la derecha de la siguiente figura), podremos apreciar que tiene distinta disposición, por el efecto del relieve que explicaremos más adelante, además de ser de mayor tamaño:



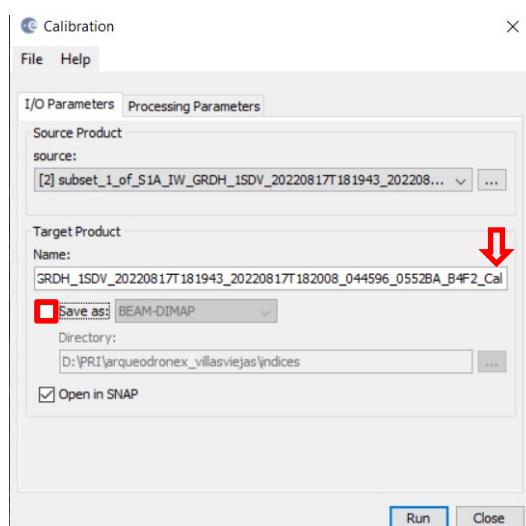
El procesamiento típico de imágenes SAR de nivel 1, no incluye correcciones radiométricas. Por lo tanto, es necesario aplicar la corrección radiométrica, para que los valores de los píxeles de las imágenes representen realmente la retrodispersión de radar de la superficie reflectante. La corrección radiométrica también es necesaria para la comparación de las imágenes de la SAR adquiridas con diferentes sensores, o adquiridas

del mismo sensor, pero en diferentes momentos, en diferentes modos, o procesadas por diferentes procesadores.

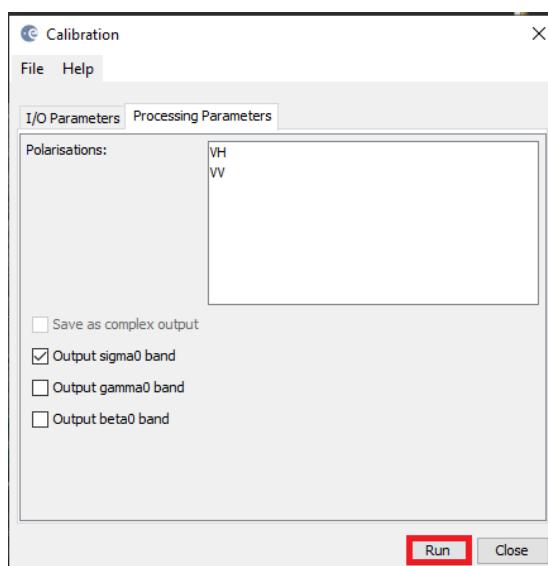
Continuamos entonces con la calibración de la imagen:



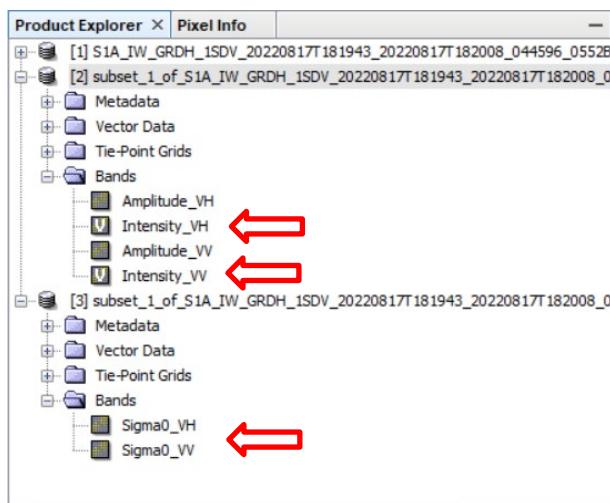
Tendremos la precaución de trabajar en virtual, para no consumir mucho espacio de almacenamiento. Nos generará entonces una imagen virtual calibrada y con el sufijo *_cal*.



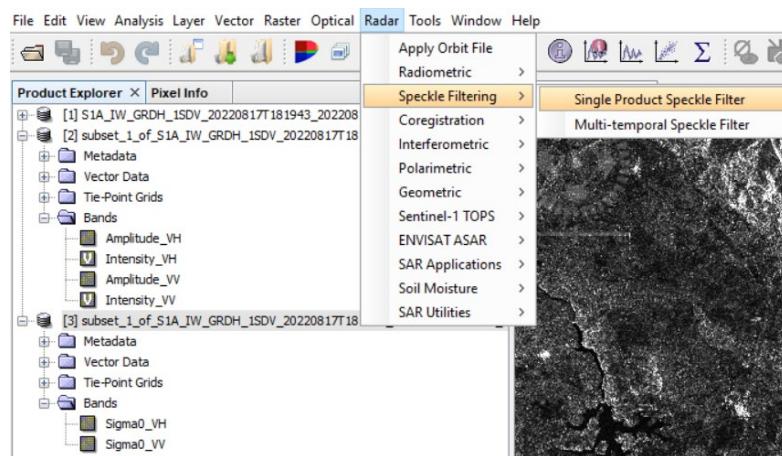
No cambiamos nada de los parámetros de procesado y le damos a calcular:



Vemos como ahora, ya se han reducido el número de bandas, ya que las bandas de intensidad de la imagen recortada eran virtuales .

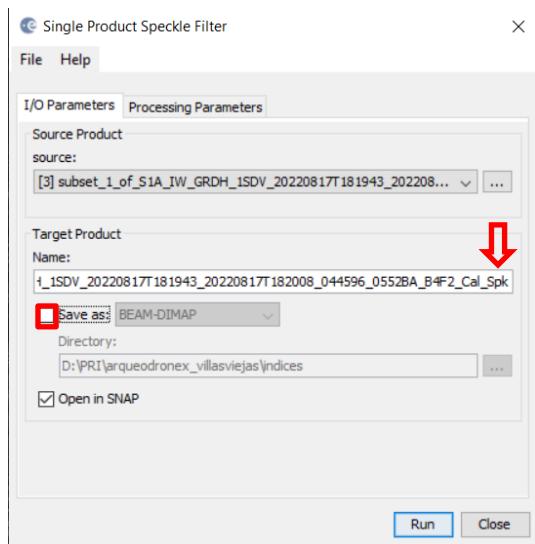


Las imágenes S1 tienen un efecto granular, también denominado “sal y pimienta”, *speckle* o ruido que requieren de tratamientos (filtros) para su correcta interpretación. El ruido se debe a oscilaciones aleatorias de la señal devuelta al sensor, debidas a la interacción de la onda emitida con las superficies rugosas del terreno. Por ese motivo, procedemos entonces a pasarle a la imagen un filtro especular:

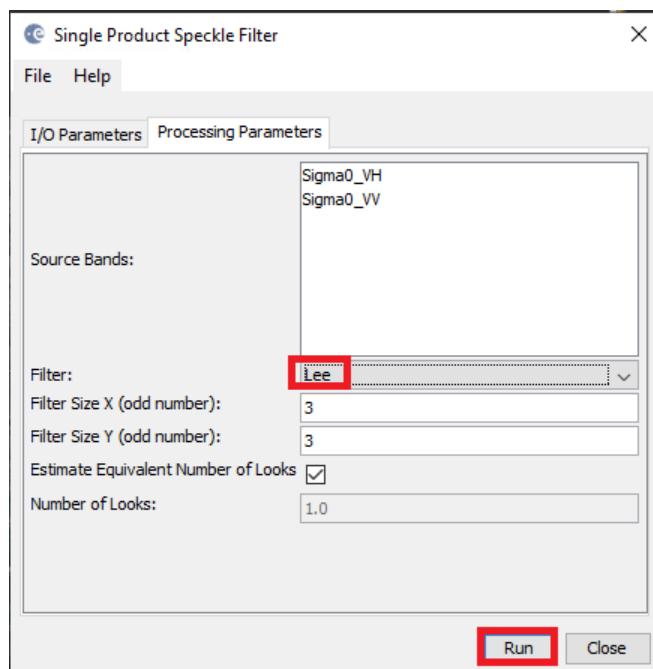


En este caso, como no estamos trabajando con una serie temporal de imágenes, le aplicamos el de producto único.

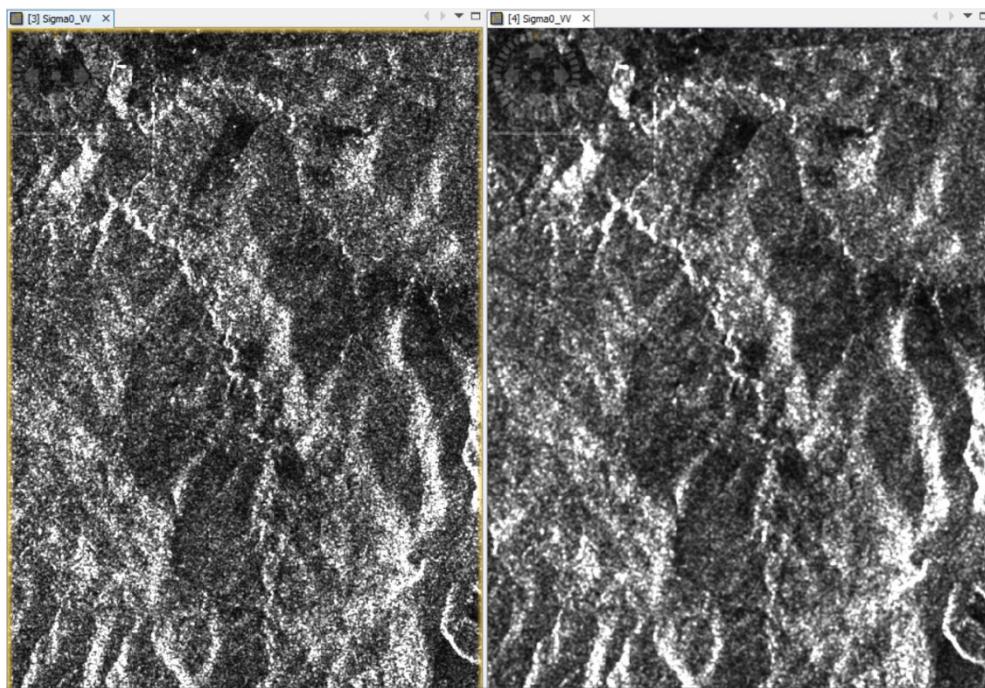
A la hora de indicar los parámetros de entrada/salida, vamos a trabajar de nuevo de forma virtual y vemos cómo le va a añadir el sufijo *_spk*.



Y en la solapa de parámetros de procesado le indicaremos que vamos a aplicar el filtro Lee:

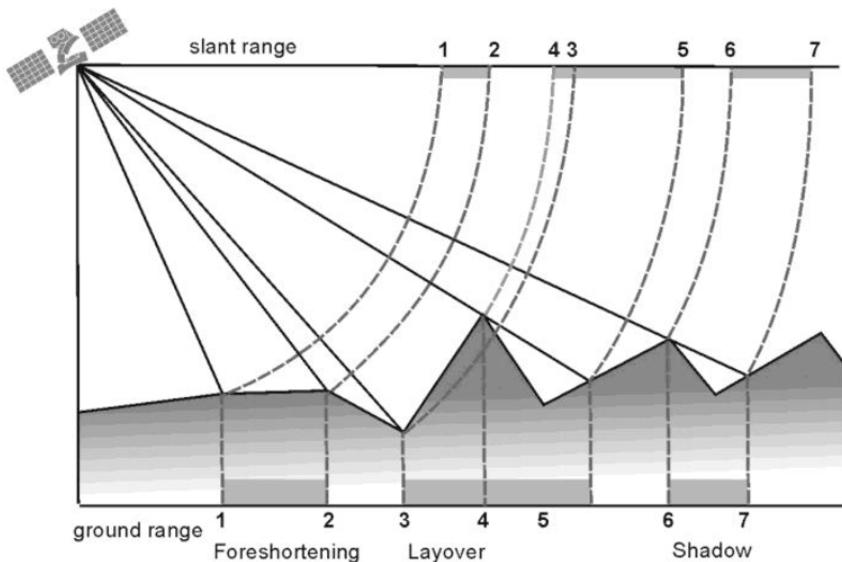


Si queremos ver el efecto de la aplicación del filtro, procederemos a abrir la misma banda:

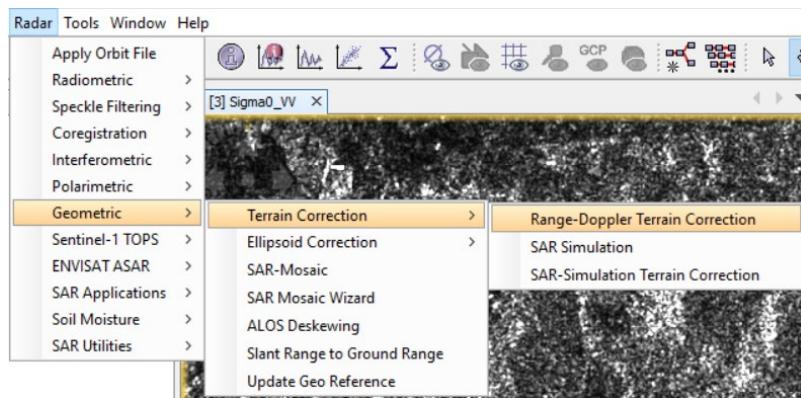


Finalmente, learemos la corrección del terreno, para eliminar las distorsiones en el relieve.

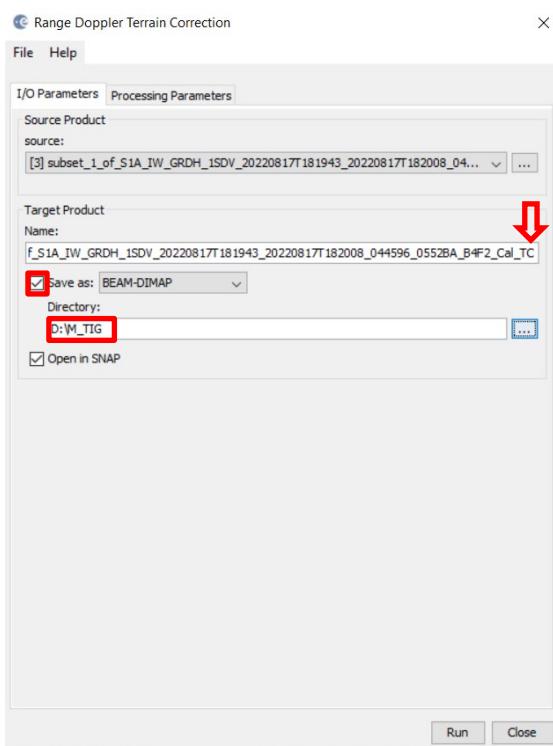
Debido a la emisión de ondas de captación por parte del sensor en una dirección lateral, se producen una serie de distorsiones geométricas que es necesario conocer y corregir. Una de ellas es la distorsión por distancia oblicua (*slant range distortion*), que genera zonas más comprimidas en la imagen en las áreas de captación más cercanas al sensor y más amplias en las más alejadas. Ello nos lleva a resultados no reales de distancia en la imagen. Otras distorsiones son: inversión por relieve (*layover*), desplazamiento de estructuras (*foreshortening*) y la generación de sombras; dándose estos casos, principalmente, en zonas con pendientes abruptas:



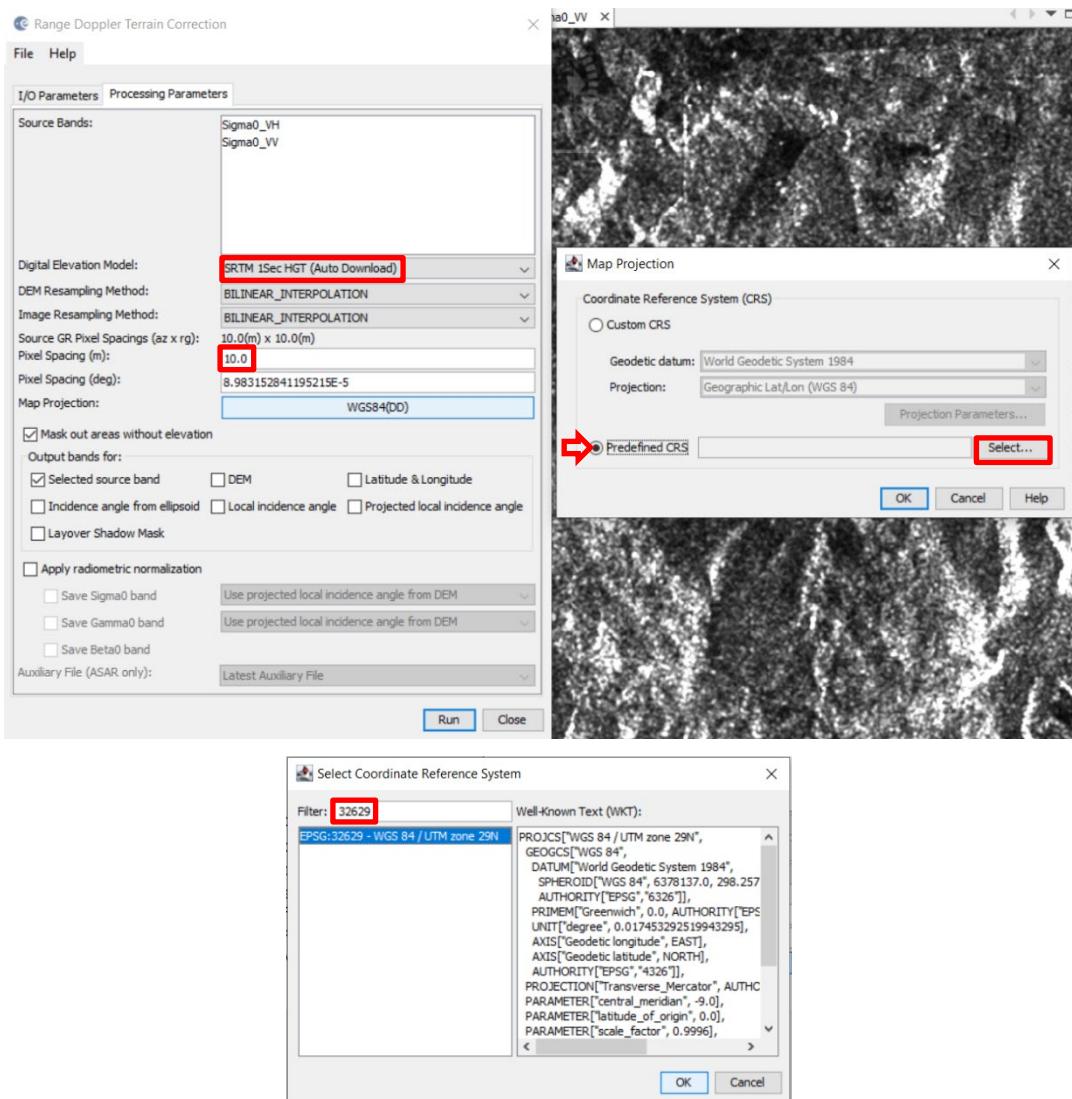
Fuente: Lenzano, MG.

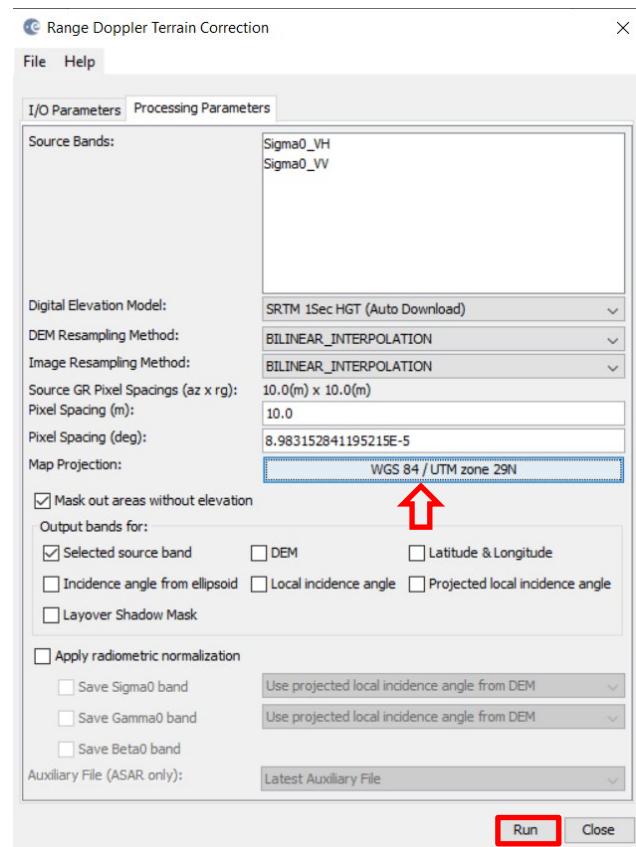


En esta ocasión, sí vamos a guardar el archivo resultante, puesto que será ya la imagen final con la que vamos a trabajar. Como siempre, SNAP aplicará al nombre un sufijo que en esta operación será _TC.

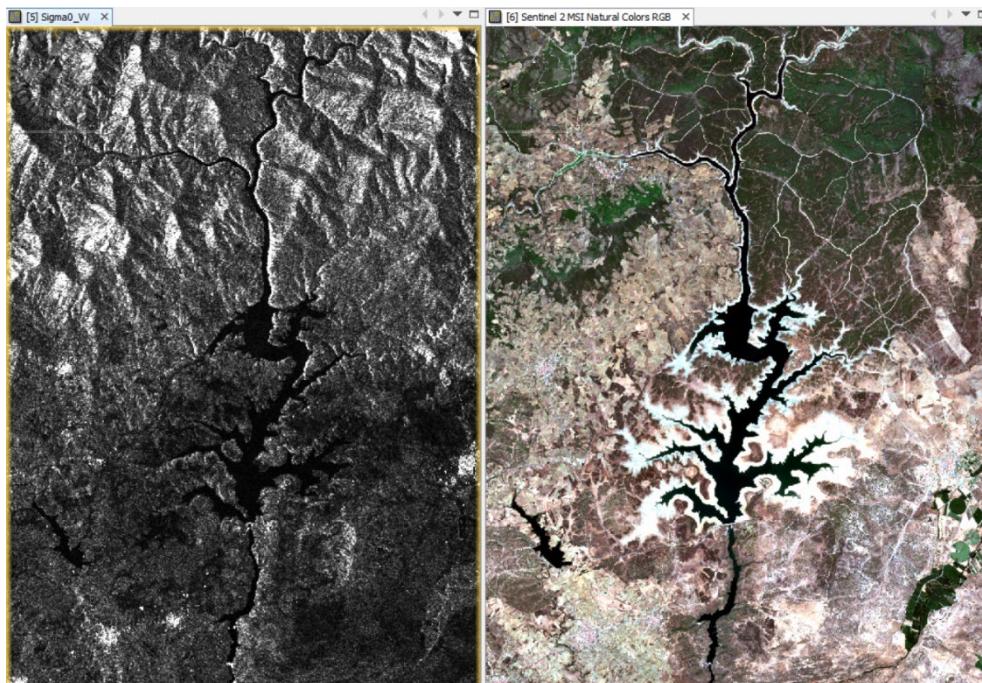


En el caso de los parámetros de procesado, le indicaremos que utilice como Modelo Digital de Elevaciones, para realizar la corrección, el SRTM de 1 segundo de resolución, que se descargará automáticamente al hacer la corrección. En cuanto a la resolución de la imagen de salida le indicaremos que sea de 10 m para que luego podamos fusionarla con la S2 y, por tanto, el sistema de referencia de salida también que ser el mismo de la S2.

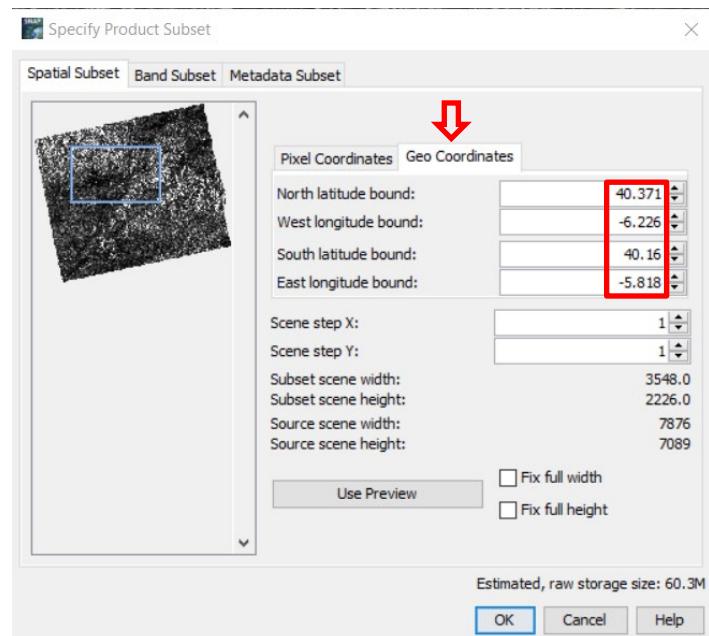




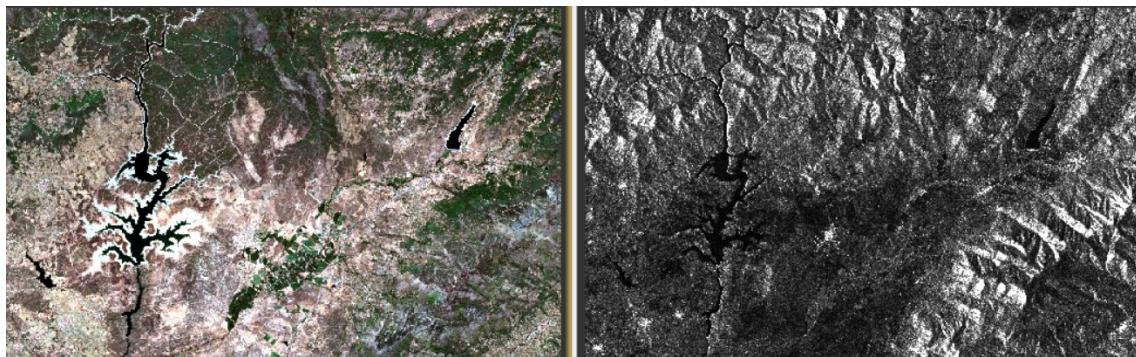
El proceso durará unos minutos. Si visualizamos el resultado final, podremos observar como ya sí la geometría de la imagen es la correcta y cuadra con la del S2.



Procederemos de nuevo a recortarla para que se adapte a las coordenadas de la S2 y así luego poder fusionarlas.



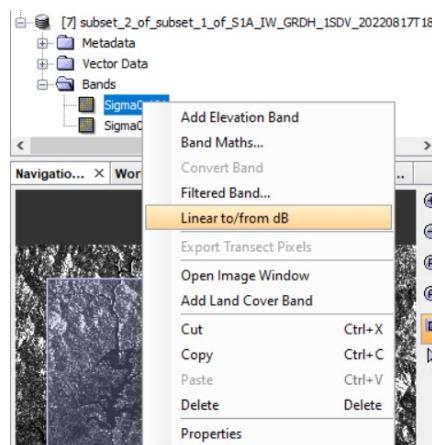
Comprobaremos entonces que son idénticas de tamaño.



Como los valores Sigma0 muestran la intensidad de la retrodispersión en escala lineal, la mayoría es oscura, mientras que sólo una pequeña proporción es brillante. Esto no es ideal en un sentido estadístico y puede dificultar la interpretación de la imagen, porque los valores inferiores a 1 tienen valores de gris similares.

Para conseguir una distribución normal de los valores, se aplica la función logarítmica a la imagen del radar. Esta función traduce los valores de los píxeles en una escala logarítmica y da lugar a mayores contrastes, ya que los valores brillantes se desplazados hacia la media, mientras que los valores oscuros se extienden sobre una gama de colores más amplia.

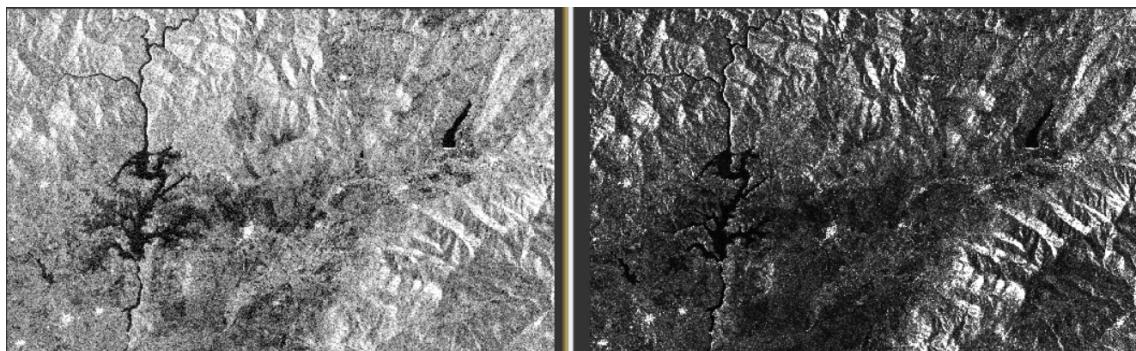
Para ello actuaremos sobre cada banda (abriendo el correspondiente menú con el botón derecho del ratón) para cambiar la distribución:



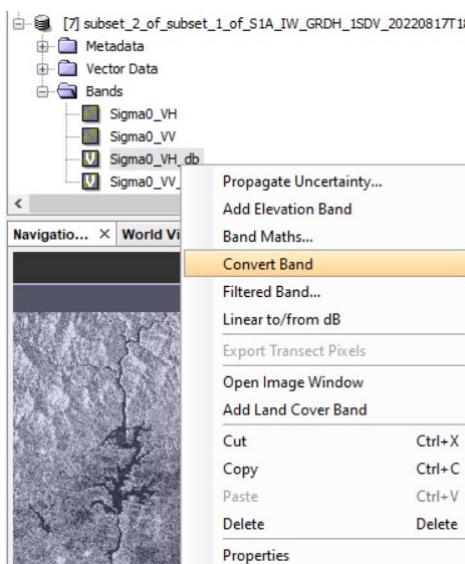
Nos preguntará si queremos crear una nueva banda virtual, a lo cual responderemos que sí.



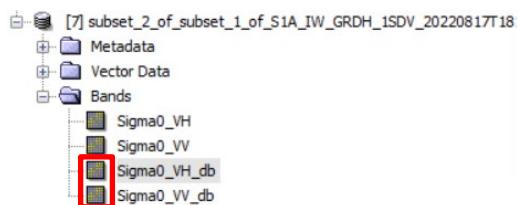
Podremos ahora apreciar el efecto producido:



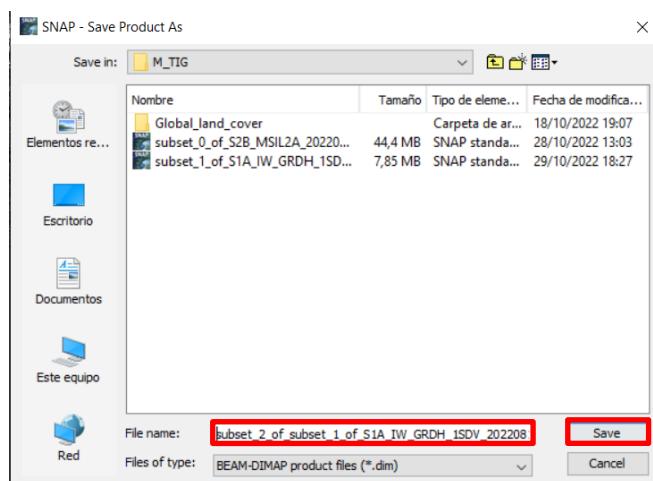
Haremos lo mismo con la otra banda y posteriormente convertiremos las bandas virtuales en bandas de la imagen:



Se podrá observar desde ese momento que ha desaparecido la marca de banda virtual de las dos bandas:



Ahora ya, guardamos el resultado final como archivo *.dim* para poder fusionarla posteriormente con la S2.



1.4. Índices de Vegetación y de Suelos con S2

Los índices de vegetación se consiguen con la combinación de las bandas espectrales, para producir un valor que indique la cantidad o vigor de vegetación dentro de un píxel. Nos permiten, de ese modo, estimar y evaluar el estado de salud de la vegetación.

Existen múltiples variedades de índices y con la aparición de las imágenes S2, el número ha sido ampliado.

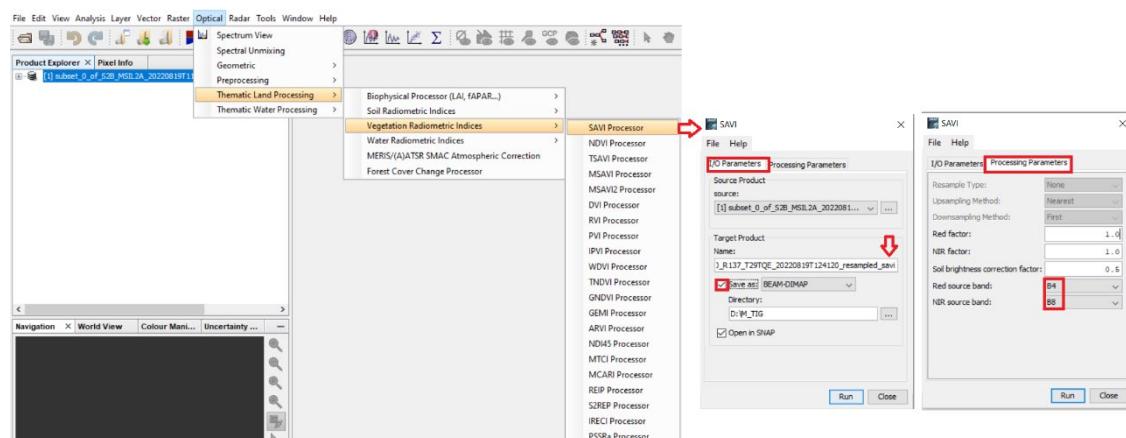
El más común de ellos es el NDVI, que es función de la energía absorbida o reflejada por las plantas. Mientras que en el visible los pigmentos de la hoja absorben la mayor parte de la energía que reciben, en el NIR, las paredes de las células de las hojas, que se encuentran llenas de agua, reflejan la mayor cantidad de energía. Sin embargo, cuando la vegetación no está sana, la cantidad de agua disminuye por lo que la reflectividad disminuye en el NIR y aumenta proporcionalmente en el rango del rojo al tener menor absorción clorofílica. También se puede calcular el GNDVI es una variante del NDVI, pero contrasta el estado de la vegetación entre el NIR y la banda del verde (GreenNDVI). El S2REP y el NDI45 a su vez, se emplea para estimar variables biofísicas como el índice de área foliar (LAI)

Por otro lado, y con respecto a los índices de suelo, encontramos el SAVI y MSAVI2 que pretenden minimizar el efecto del suelo sobre los índices de vegetación, especialmente manifiesto en superficies parcialmente recubiertas. También el BI que es índice de brillo y usa las bandas del rojo e infrarrojo cercano y que tiene en cuenta el albedo de la superficie terrestre para diferenciar entre las cubiertas vegetales y los suelos. Este índice es sensible a la luminosidad de los elementos, así cuando el suelo es muy húmedo puede aparecer en las imágenes con un color muy oscuro. Finalmente, el CI o índice de color que se relaciona con la concentración de carbonatos o sulfatos en los suelos.

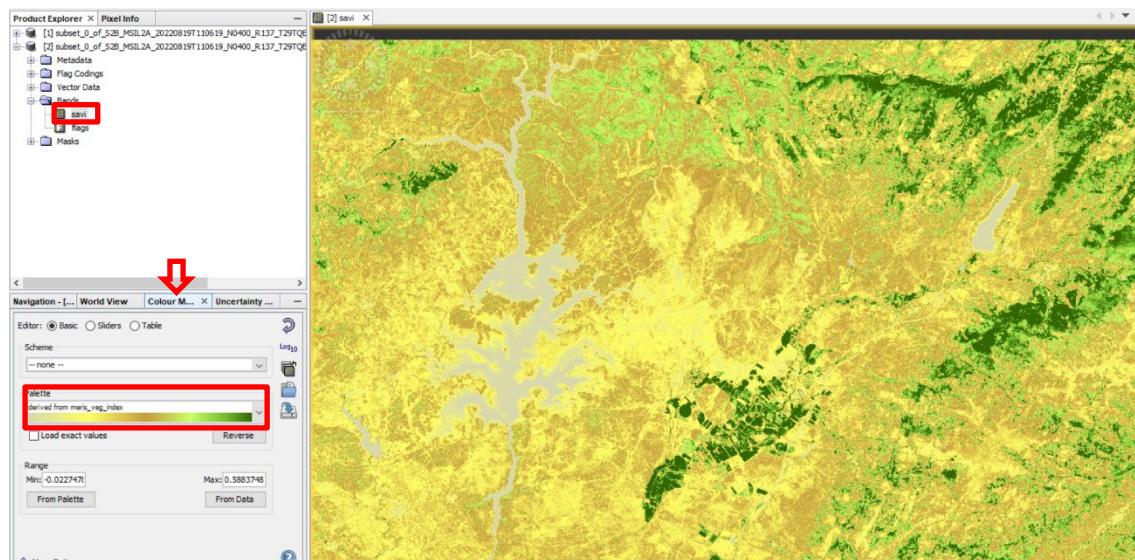
Índice	Ecuación
GNDVI	$GNDVI = \frac{B7 - B3}{B7 + B3}$
NDVI	$NDVI = \frac{B8 - B4}{B8 + B4}$
NDI45	$NDI45 = \frac{B5 - B4}{B5 + B4}$
S2REP	$S2REP = 705 + 35 \cdot \frac{\frac{B4 + B7}{2} - B5}{B6 - B5}$
SAVI	$SAVI = (1 + L) \cdot \frac{B8 - B4}{B8 + B4 + L}$
MSAVI2	$MSAVI2 = \frac{1}{2} \cdot [(2 \cdot (B8 + 1) - \sqrt{(2B8 + 1)^2 - 8(B8 - B4)})]$
BI	$BI = \sqrt{\frac{B4^2 + B3^2}{2}}$
CI	$CI = \frac{B4 - B3}{B4 + B3}$

Procederemos, por tanto, a calcular los índices de vegetación descritos anteriormente (aunque para cada trabajo existirán unos índices más o menos discriminatorios de las clases que queramos clasificar).

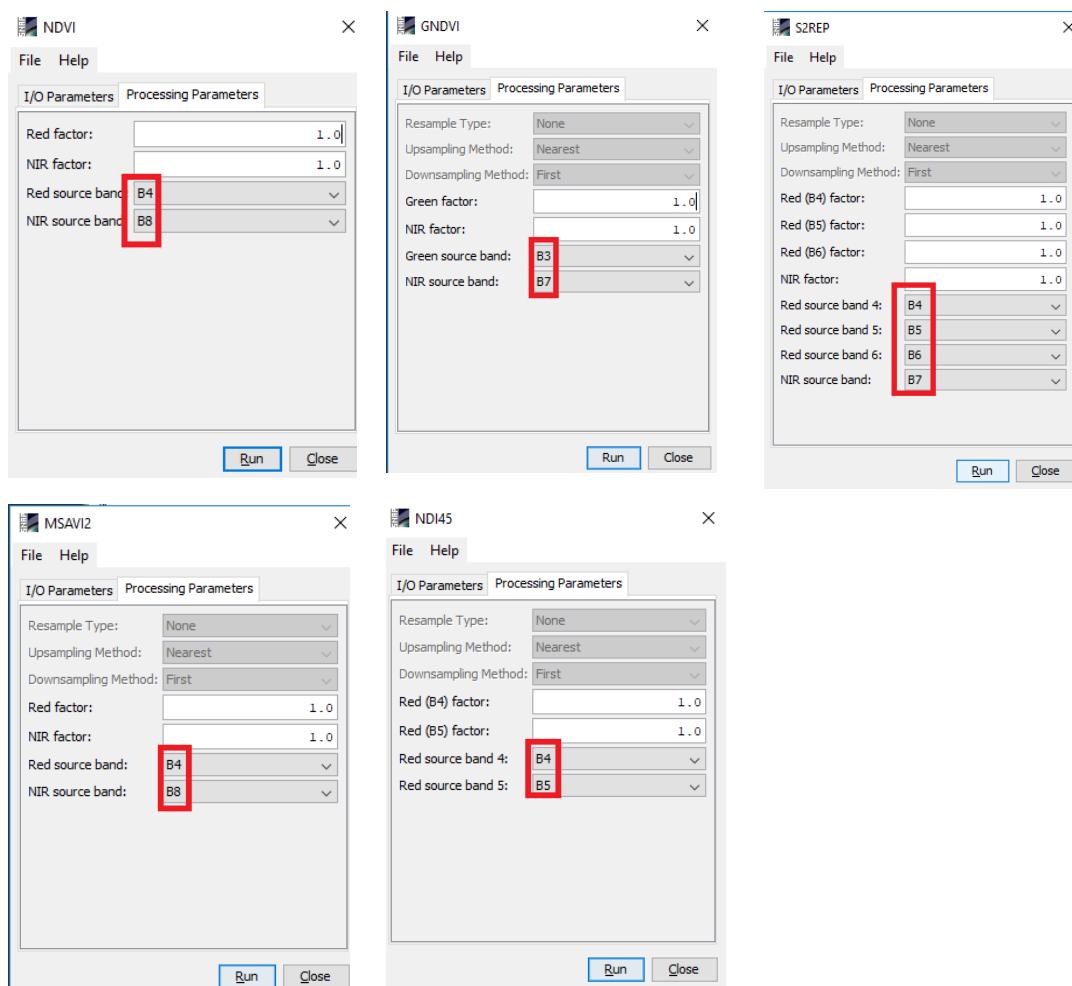
Cada índice se calculará como una imagen independiente, para que no ocupen mucho espacio en memoria, podemos calcularlas de forma virtual y luego ya añadirlas a la S2 como bandas auxiliares para la clasificación.



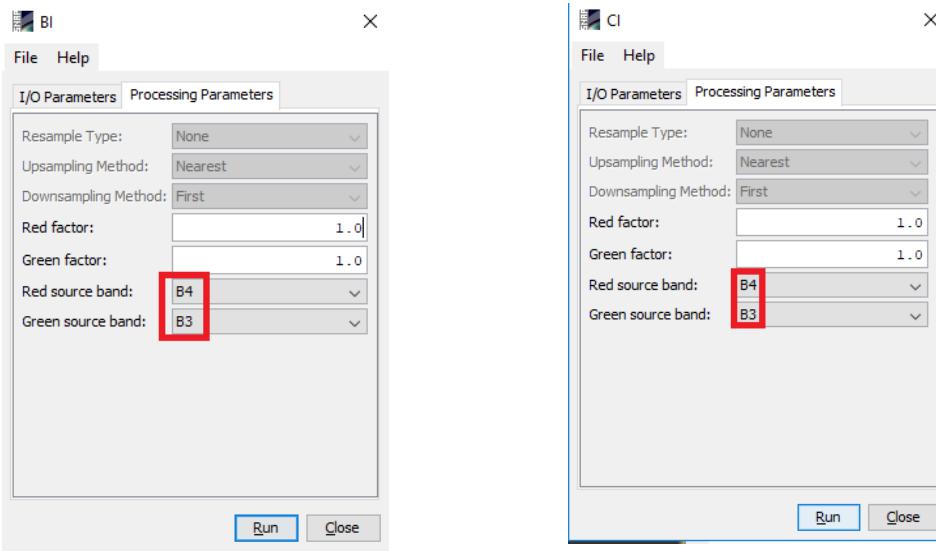
El resultado del SAVI, en este caso, será el siguiente:



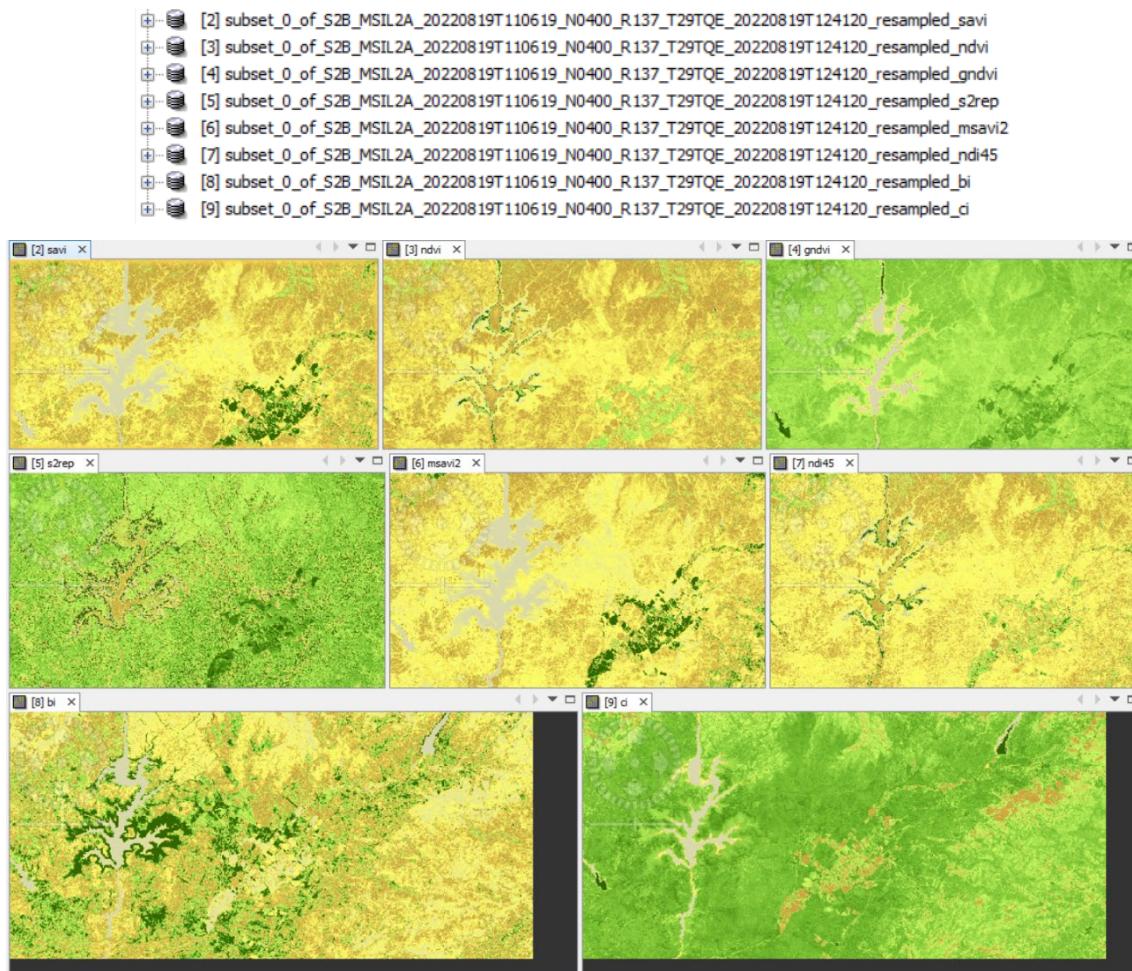
Y así procederemos con todos los índices mostrados en la tabla:



Y lo mismo para los índices de suelos

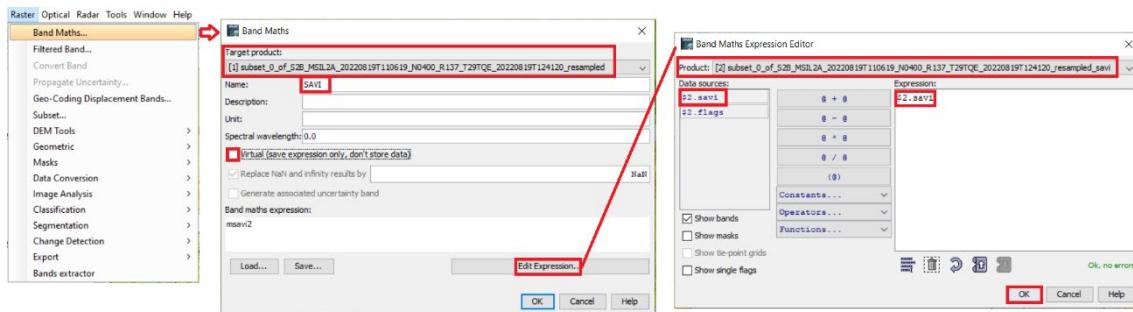


De este modo, tendremos todas las imágenes creadas con los índices de vegetación y suelo calculados.

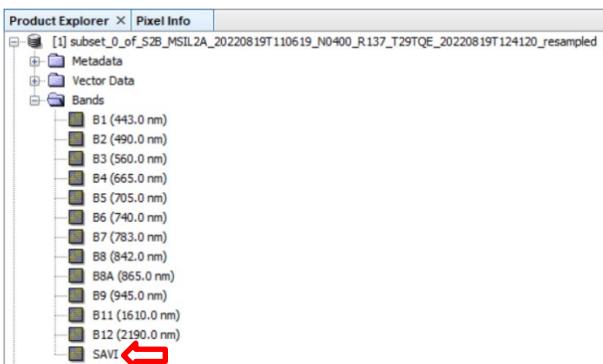


Procederemos ahora a añadirlas todas, como bandas, en la imagen S2 original.

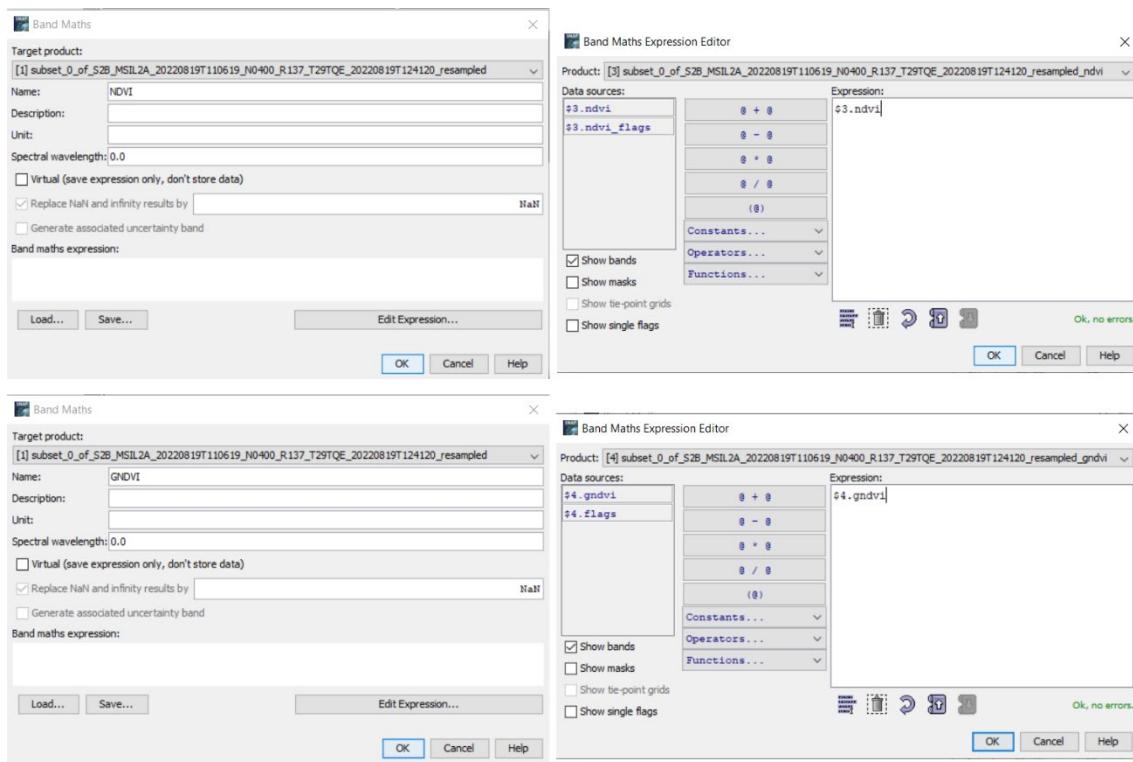
Como proceden todas del mismo sensor, podríamos hacerlo tanto con una simple colocación o con la calculadora de bandas. En este caso, para aprender a manejar las dos herramientas, lo realizaremos desde la calculadora de bandas.



A partir de ese momento, tendremos una nueva banda en la imagen original, que corresponde con el índice SAVI.



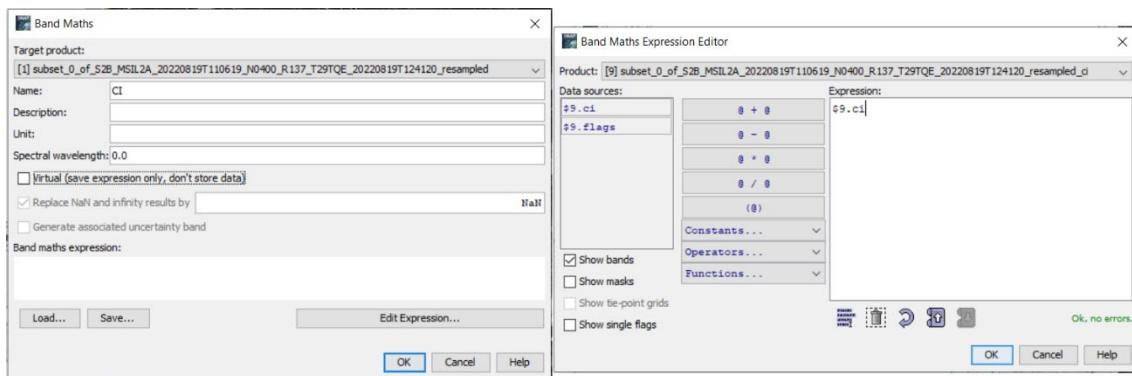
Y así sucesivamente con el resto de índices que hemos calculado:



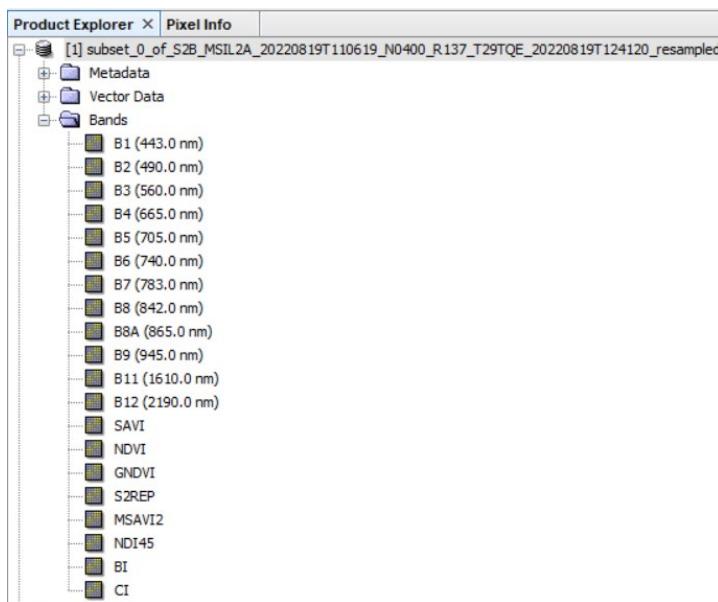
The figure consists of five rows, each containing a 'Band Maths' dialog box on the left and a 'Band Maths Expression Editor' dialog box on the right. Each row illustrates the creation of a specific band product:

- Row 1:** Target product is [1] subset_0_of_S2B_MSIL2A_20220819T110619_N0400_R137_T29TQE_20220819T124120_resampled. The expression in the editor is `$5.s2rep`.
- Row 2:** Target product is [6] subset_0_of_S2B_MSIL2A_20220819T110619_N0400_R137_T29TQE_20220819T124120_resampled_msavi2. The expression in the editor is `$6.msavi2`.
- Row 3:** Target product is [7] subset_0_of_S2B_MSIL2A_20220819T110619_N0400_R137_T29TQE_20220819T124120_resampled_ndi45. The expression in the editor is `$7.ndi45`.
- Row 4:** Target product is [8] subset_0_of_S2B_MSIL2A_20220819T110619_N0400_R137_T29TQE_20220819T124120_resampled_b1. The expression in the editor is `$8.b1`.
- Row 5:** This row shows the 'Band maths expression' field empty in the left dialog, while the right editor shows the expression `$8.b1`.

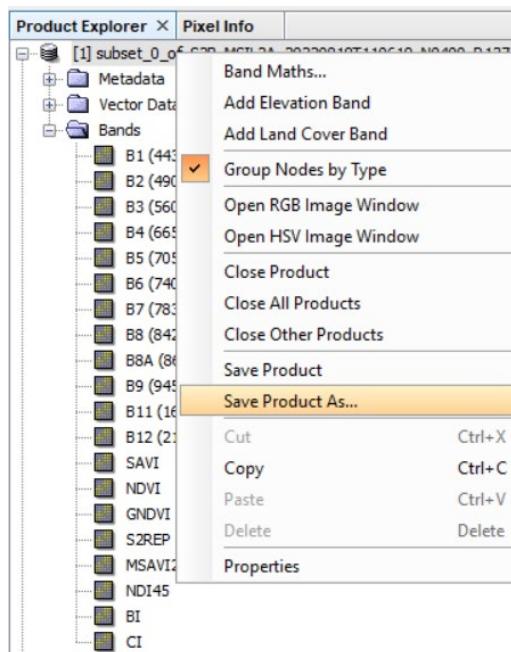
In all dialogs, the 'Target product' dropdown lists the same subset product. The 'Name' fields are populated with 'S2REP', 'MSAVI2', 'NDI45', 'B1', and an empty string respectively. The 'Edit Expression...' button is present in the bottom right of each dialog.



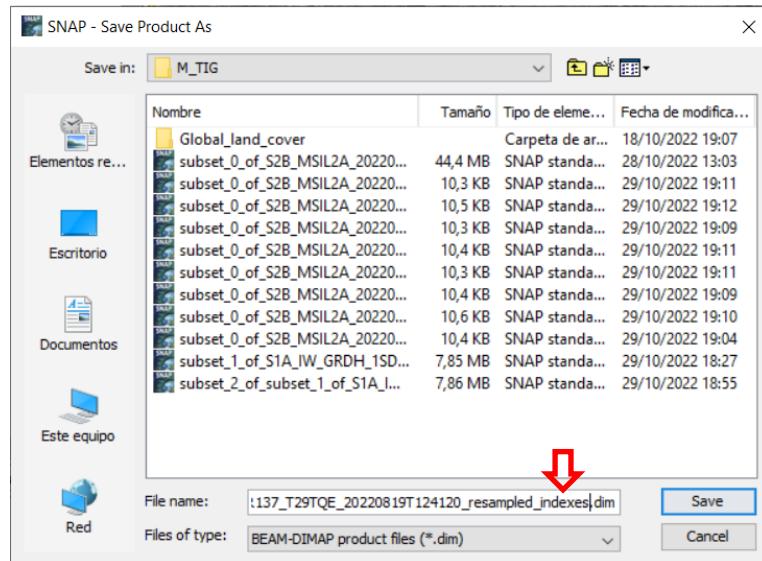
De ese modo, ya tendríamos todas las bandas añadidas a nuestra imagen S2.



En este momento guardaremos la imagen con los índices de vegetación y suelo para tener una copia con la que trabajar posteriormente.



Le pondremos el sufijo de _indexes, para saber que es la imagen con las bandas de los índices añadidos.

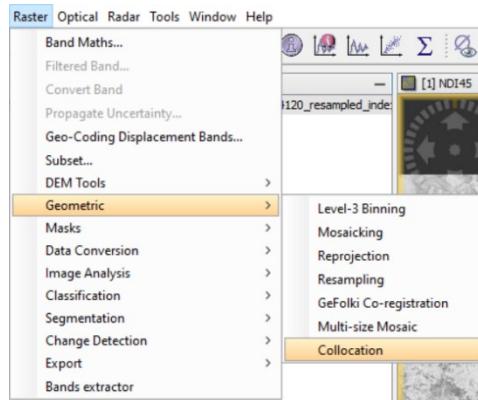


1.5. Fusión de imágenes S2 y S1

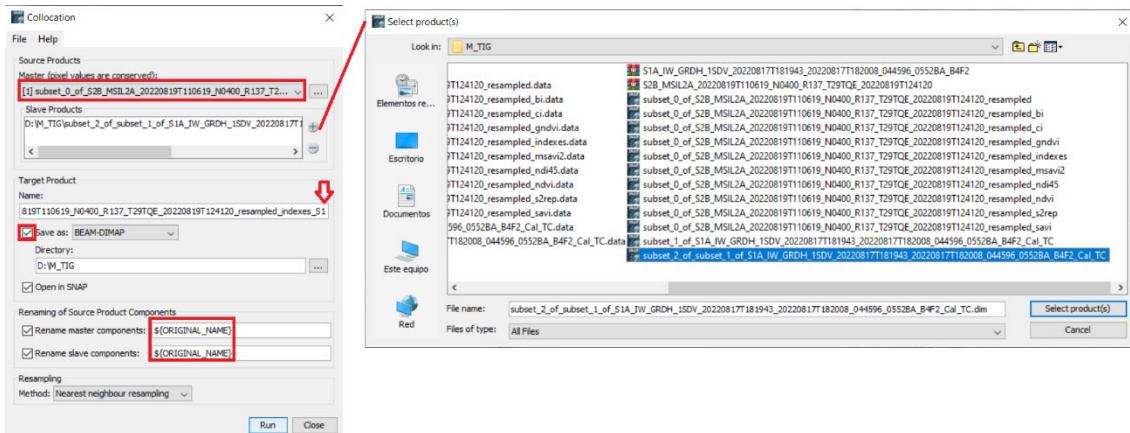
Por un lado, las imágenes S2 se utilizan para monitorizar, suelo y agua, vías de navegación interior y zonas costeras. Las imágenes S2 también pueden proporcionar información valiosa para servicios de emergencia. Por otro lado, se ha comprobado la utilidad de imágenes SAR del S1 en la detección y monitorización de inundaciones, vulcanología, etc. Sin embargo, cada vez son más los estudios que utilizan la fusión de imágenes de ambas constelaciones para realizar estudios variados.

En nuestro caso, vamos a fusionar dichas imágenes para posteriormente, en el método de clasificación Random Forest, clasificar las coberturas de nuestra zona de trabajo con ambas imágenes.

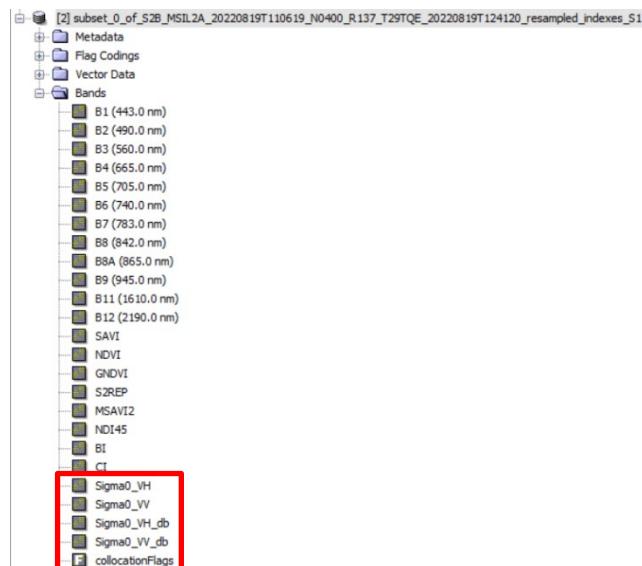
En esta ocasión, utilizaremos el comando *collocation* para unirlas.



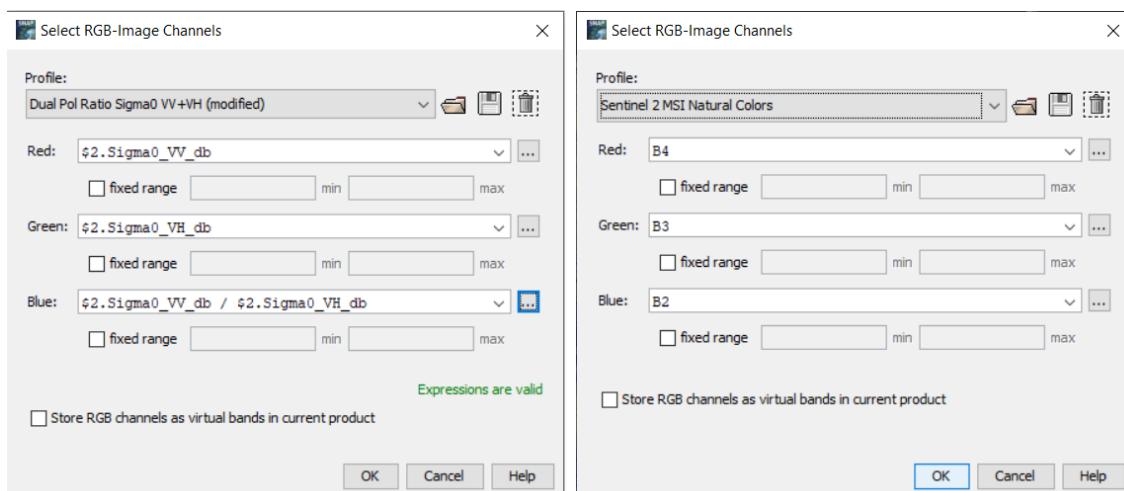
Y configuramos del siguiente modo los parámetros de la fusión:

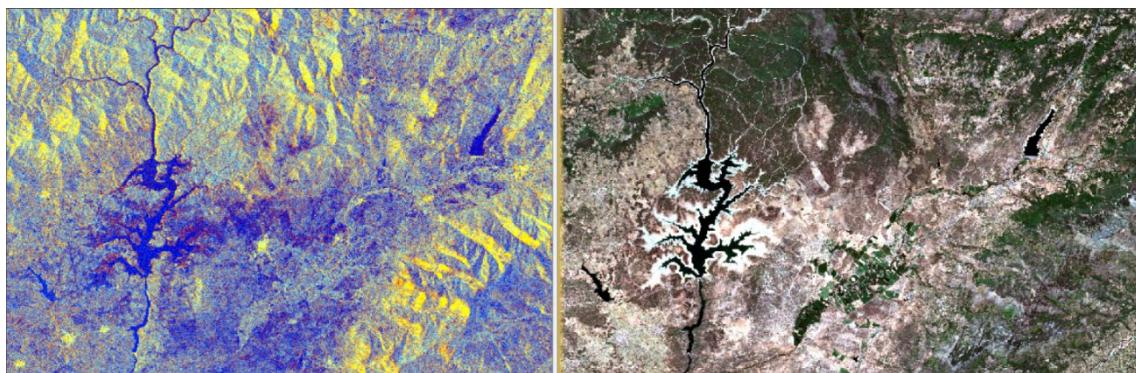


Obtendremos entonces una imagen con todas las bandas anteriores más la información del S1 fusionada.



Y tendremos por tanto la zona capturada desde el espectro visible, o desde la polarización SAR todo en un mismo fichero de trabajo.





2. Clasificación de imágenes

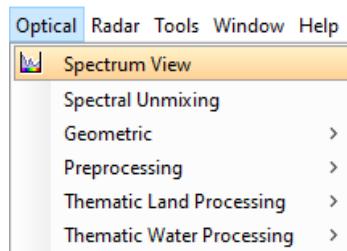
Uno de los objetivos principales de la teledetección es la de obtener cartografía temática de la superficie de la tierra. Dicha cartografía es la representación de los distintos tipos de cobertura que se pueden detectar desde el sensor. Para diferenciar las distintas coberturas, es necesario realizar una clasificación de las imágenes.

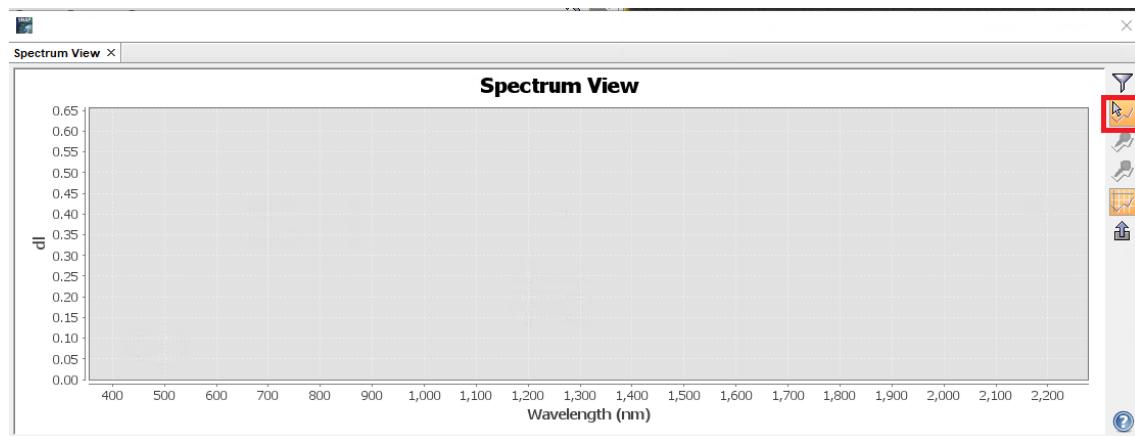
La clasificación, en el contexto de la teledetección, es un proceso que asigna a cada píxel de la imagen original una etiqueta correspondiente a una de las clases que, o bien predefinimos con anterioridad al proceso de clasificación o bien, por semejanza de características de los píxeles de la imagen a clasificar, se agrupan automáticamente en el proceso. En el primero de los casos, estaríamos ante un método de clasificación supervisada y en el segundo ante un método de clasificación no supervisada.

2.1. Clasificación no supervisada

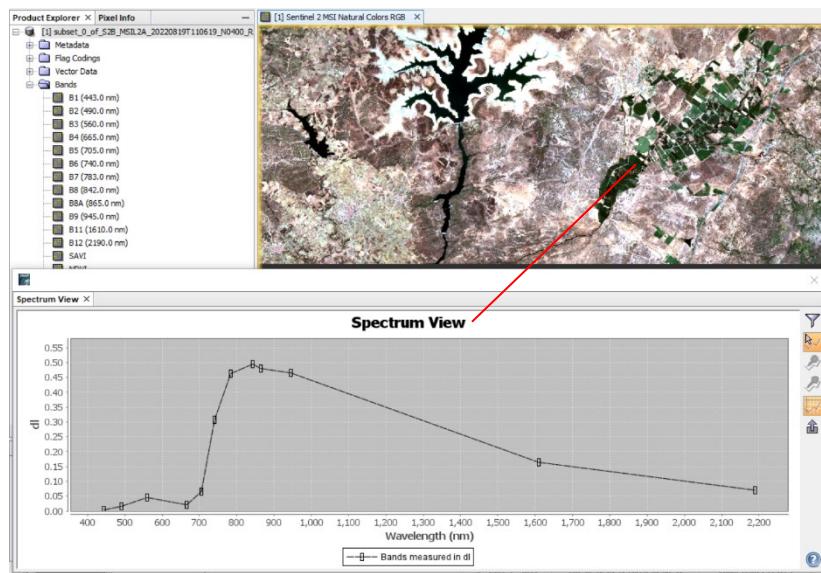
Es aquella que no precisa de información adicional, más allá de la información espectral de los píxeles de la imagen. Así, en función de los grupos con las mismas firmas espectrales, se establecerán las clases de la imagen, sin saber bien a qué cobertura corresponden. Por ejemplo, en el caso de nuestra imagen S2, todos los píxeles que tengan valores parecidos en las 12 bandas serán agrupados en una misma clase.

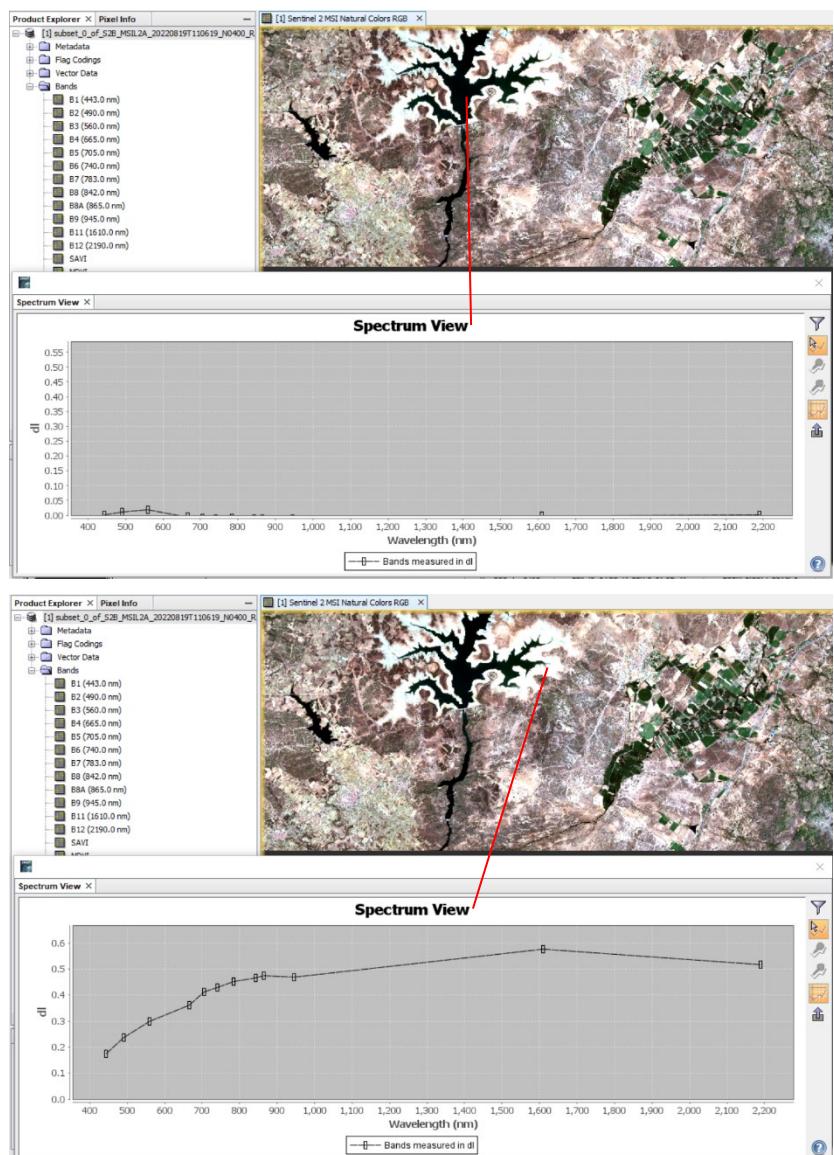
Para entenderlo mejor, abriremos la vista del espectro y compararemos dichos valores en distintas zonas de la imagen:





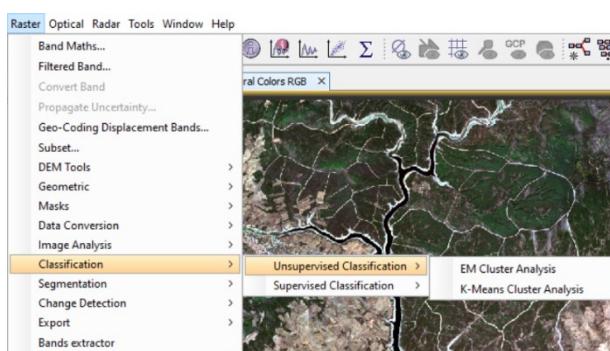
Si tenemos activada la casilla de mostrar el espectro en la posición del cursor y nos vamos moviendo en la imagen por las distintas coberturas que se aprecian a simple vista, podremos ir viendo los distintos valores que tienen en las distintas bandas y cómo estas son diferentes entre sí.





En el caso de la clasificación no supervisada, se agruparán los píxeles con espectros similares.

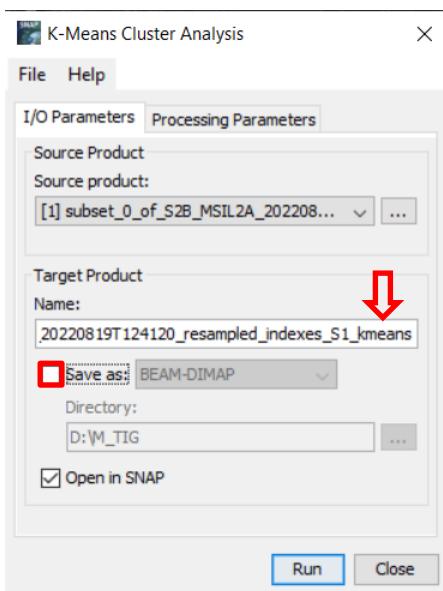
Existen varios métodos de clasificación no supervisada. En el caso de SNAP tiene implementados los algoritmos de EM y K means.



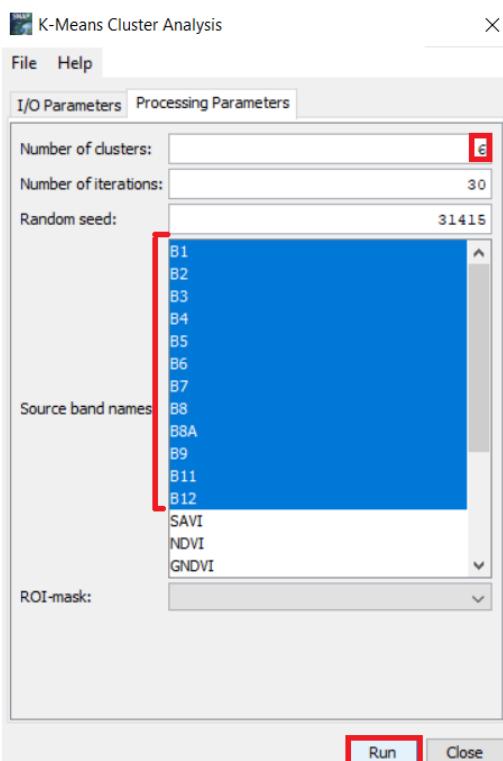
2.1.1. Análisis de conglomerados K-means

K-Means Cluster Analysis (análisis de conglomerados) es la partición de un conjunto de datos en subconjuntos (conglomerados o clases), de modo que los datos en cada subconjunto (idealmente) comparten algún rasgo común, a menudo proximidad según alguna medida de distancia definida.

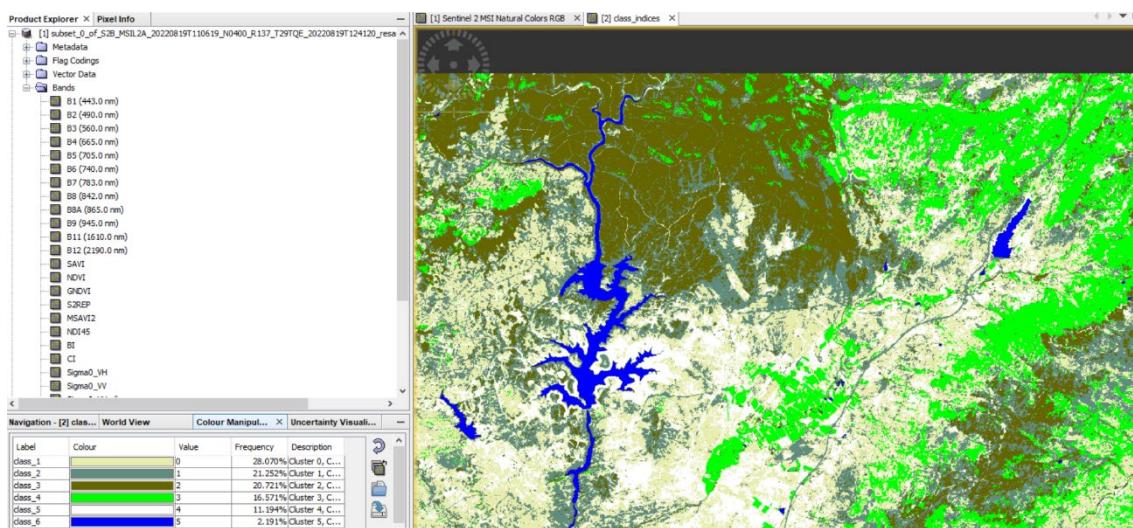
Configuraremos entonces el cálculo, para que lo haga de forma virtual, ya que no nos interesa almacenar el resultado.



Es muy importante que le marquemos, como bandas de trabajo, sólo las 12 bandas espectrales. En este caso pondremos 6 clusters y dejaremos el número de iteraciones por defecto.



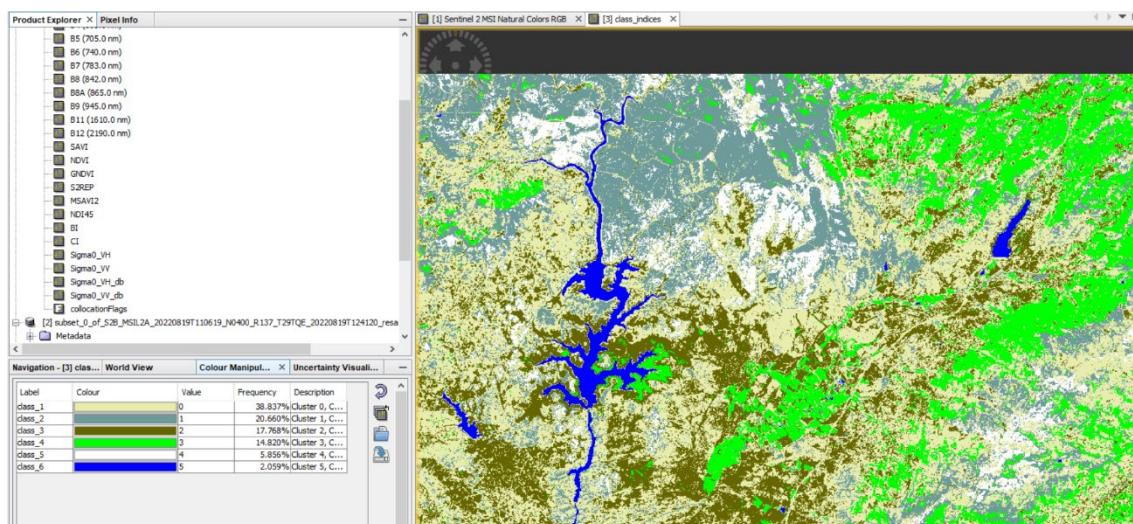
Visualizaremos el resultado, pudiendo variar la paleta de colores de la clasificación:



2.1.2. Análisis de clúster de maximización de expectativas (EM)

Análisis de clúster de maximización de expectativas (EM) es una generalización del algoritmo k-means donde los grupos son elipsoides definidos por un centro y una matriz de covarianza. Este algoritmo es que no se ve afectado por diferentes escalas de las dimensiones de los datos y las correlaciones entre ellos. Su desventaja es una velocidad considerablemente menor, que prácticamente limita la aplicabilidad solo a conjuntos de datos más pequeños.

Procederemos configurando los parámetros de cálculo de igual forma que en el caso anterior y obteniendo el siguiente resultado:



Estos tipos de clasificación no supervisada, nos sirven para hacernos una idea de las clases que a priori pueden distinguirse en la imagen.

2.2. Clasificación supervisada

Como ya se ha mencionado, la clasificación supervisada es aquella que compara las características espectrales de cada píxel de la imagen con las de un conjunto de píxeles indicados con anterioridad y que pertenecen a clases de coberturas conocidas. Ese conjunto de píxeles son las llamadas “Regiones de Interés” o “Áreas de Entrenamiento”. De la exactitud de dichas muestras depende en gran medida la exactitud del resultado final de la clasificación.

Las áreas de entrenamiento utilizadas para entrenar a los clasificadores supervisados en general deben cumplir una serie de requisitos:

- las muestras de entrenamiento deben tener un equilibrio de clases.
- las muestras de entrenamiento deben ser representativas de las clases objetivo.
- las muestras de entrenamiento deben ser lo suficientemente grande para adaptarse al número creciente bandas.
- los datos de entrenamiento y validación deben ser estadísticamente independientes.

Existen muchas teorías para la definición de las muestras de entrenamiento, pero la más extendida es la proporcional al tamaño de la clase, así las clases con mayor superficie requieren más muestras de entrenamiento que las que ocupan áreas más pequeñas.

En cuanto a los algoritmos desarrollados para realizar clasificaciones supervisadas, hay una variedad mucho más amplia que la de no supervisadas. Existe una taxonomía que divide a los clasificadores supervisados en paramétricos y no paramétricos. En el caso de los paramétricos, suponen una distribución de probabilidad de las clases conocida (normalmente distribución normal). Si la dimensionalidad de los datos es alta se necesitan muchas áreas de entrenamiento. Sin embargo, los no paramétricos no presuponen la forma de la distribución de las probabilidades. Estiman la densidad de probabilidad de cada observación usando las áreas de entrenamiento.

Entre los métodos paramétricos se encuentran los más clasificadores más extendidos como son el de máxima verosimilitud, el de mínima distancia euclídea o el de la distancia de Mahalanobis.

También son conocidos algunos de los no paramétricos tales como el del vecino más próximo, el del paralelepípedo, el del ángulopectral, las redes neuronales o los árboles de decisión.

Veremos en detalle uno del primer grupo, el de máxima verosimilitud y otro del grupo de los no paramétricos que será el de Random Forest.

2.2.1 Definición de áreas de entrenamiento

Para ejecutar cualquier algoritmo de clasificación supervisada, lo primero que necesitaremos es indicar las áreas de entrenamiento de las clases que vamos a clasificar. Existen varios métodos de definición de las áreas de entrenamiento. El más certero es el de la visita a campo con toma de datos mediante coordenadas para la definición de las clases de cobertura contenidas en nuestra imagen. A pesar de ser el método más exacto, esta forma de definición de las áreas es muy costosa y, en ocasiones, difícil de llevar a cabo. Por ese motivo, otra forma de definición de áreas de entrenamiento es la de definir las clases con las que vamos a trabajar, tomando como base otras cartografías temáticas u otras imágenes con mayor resolución y, por tanto, con mejor definición para la delimitación espacial de las clases.

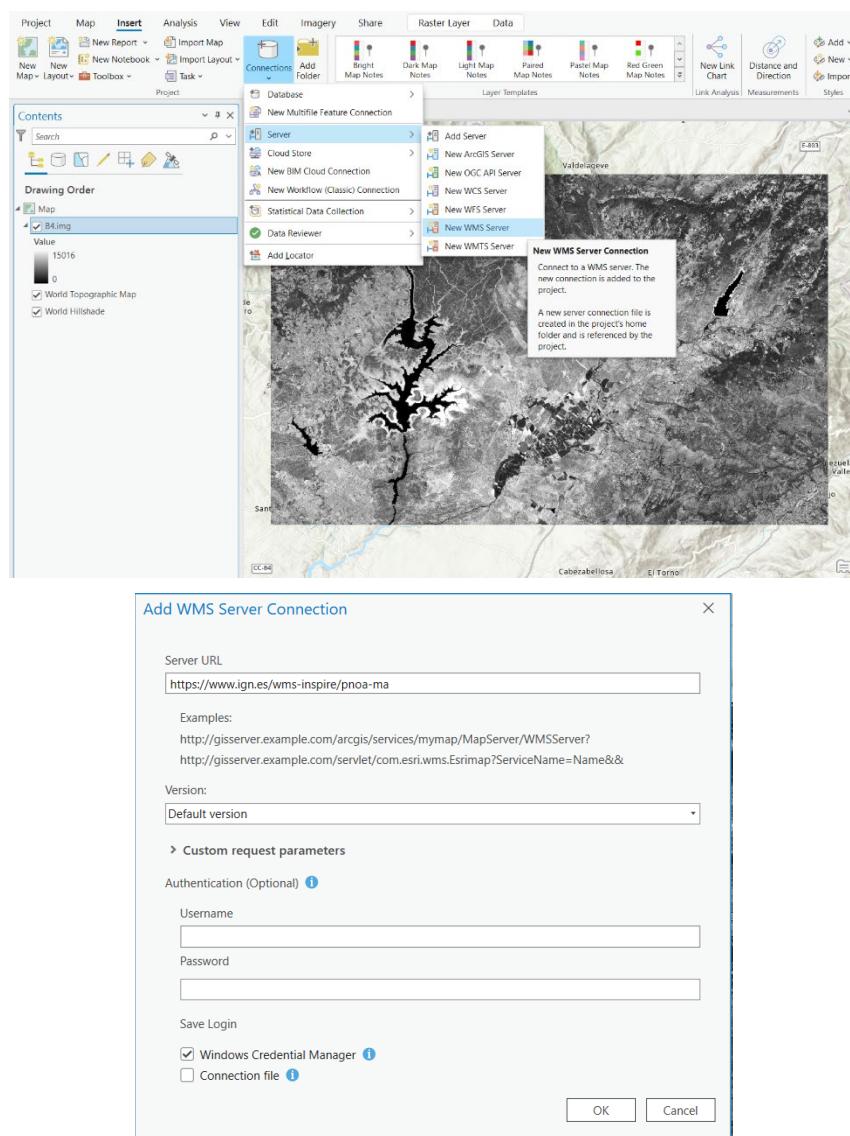
En nuestro caso, optaremos por este segundo método. Vamos a definir nuestras clases mediante el mapa del WorldCover map, recientemente publicado, y sobre la ortofoto del PNOA de máxima actualidad.

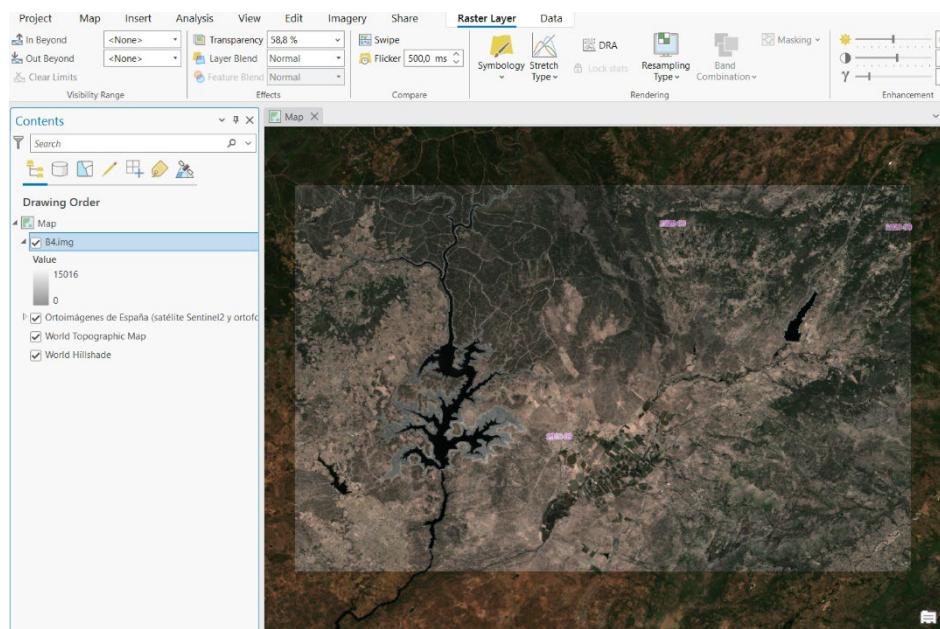
Para ello, cargaremos una de las bandas de la imagen del S2 en Arcgis, donde delinearemos los polígonos que posteriormente utilizaremos como áreas de interés.

Sobre un proyecto nuevo de Arcgis, insertaremos una de las bandas de la S2. Aunque la imagen tiene formato .dim, en el mismo directorio hay una carpeta con el mismo nombre del archivo y que contiene todas las bandas en formato .img y que están georreferenciadas. Cargaremos una de esas imágenes para saber la delimitación espacial con la que estamos trabajando.

Nombre	Fecha de modificación	Tipo	Tamaño
vector_data	30/10/2022 17:09	Carpeta de archivos	
B1.hdr	29/10/2022 20:00	Archivo HDR	2 KB
B1	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B2.hdr	29/10/2022 20:00	Archivo HDR	2 KB
B2	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B3.hdr	29/10/2022 20:00	Archivo HDR	2 KB
B3	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B4.hdr	29/10/2022 20:00	Archivo HDR	2 KB
B4	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B4.aux	30/10/2022 18:07	Documento XML	4 KB
B4.mng.ovr	30/10/2022 18:07	Archivo OVR	3.983 KB
B5.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B5	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B6.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B6	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B7.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B7	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B8.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B8	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B8A.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B8A	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B9.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B9	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B11.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B11	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B12.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B12	29/10/2022 20:01	Archivo de imagen de d...	15.630 KB
B13.hdr	29/10/2022 20:01	Archivo HDR	2 KB
B13	29/10/2022 20:01	Archivo de imagen de d...	30.800 KB
C1.hdr	29/10/2022 20:01	Archivo HDR	2 KB
C1	29/10/2022 20:01	Archivo de imagen de d...	30.800 KB
collocationFlag.hdr	29/10/2022 20:01	Archivo HDR	2 KB
collocationFlag	29/10/2022 20:01	Archivo de imagen de d...	7.715 KB
GNDVI.hdr	29/10/2022 20:01	Archivo HDR	2 KB
GNDVI	29/10/2022 20:01	Archivo de imagen de d...	30.690 KB
MSAV12.hdr	29/10/2022 20:01	Archivo HDR	2 KB
MSAV12	29/10/2022 20:01	Archivo de imagen de d...	30.690 KB
ND45.hdr	29/10/2022 20:01	Archivo HDR	2 KB
ND45	29/10/2022 20:01	Archivo de imagen de d...	30.690 KB
NDVI.hdr	29/10/2022 20:01	Archivo HDR	2 KB
NDVI	29/10/2022 20:01	Archivo de imagen de d...	30.690 KB
SRREP.hdr	29/10/2022 20:01	Archivo HDR	2 KB
SRREP	29/10/2022 20:01	Archivo de imagen de d...	30.690 KB
S4M.hdr	29/10/2022 20:01	Archivo HDR	2 KB
S4M	29/10/2022 20:01	Archivo de imagen de d...	30.690 KB

Añadiremos después el servidor WMS con la ortofoto del PNOA de máxima actualidad, aplicándole a la B4 de la S2 una transparencia del 50%:





Con el fin de definir los píxeles de la imagen Sentinel con la máxima pureza para hacer el entrenamiento, vamos a delimitarlos mediante polígonos y así, posteriormente, definir las áreas según estas localizaciones.

Por tanto, vamos a poligonizar la imagen:

Geoprocessing Raster to Polygon

Parameters

Input raster: B4.img

Field: Value

Output polygon features: pixeles_b4.shp

Simplify polygons

Create multipart features

Maximum vertices per polygon feature: []

Contents

Drawing Order

Map

B4.img

Ortoimágenes de España (satélite Sentinel2 y ortofoto)

World Topographic Map

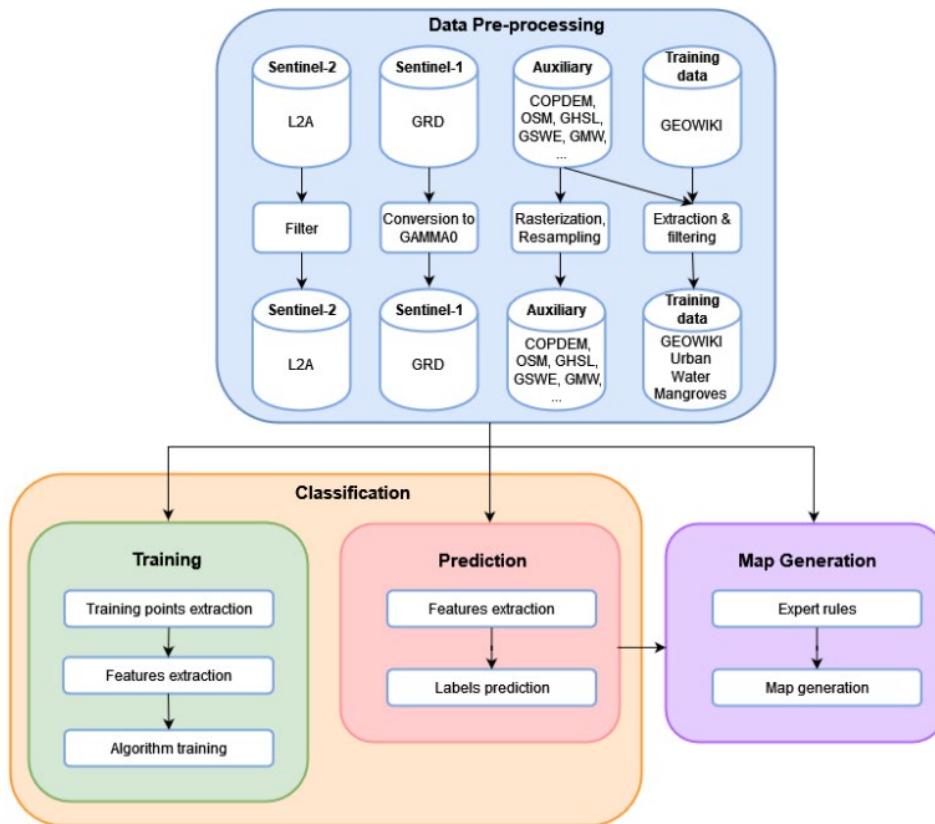
World Hillshade

Map

The map view shows a satellite image of a landscape with a dense grid overlay representing the polygonized areas from the 'Raster to Polygon' process. The grid covers most of the image, indicating where pixels have been converted into polygon features.

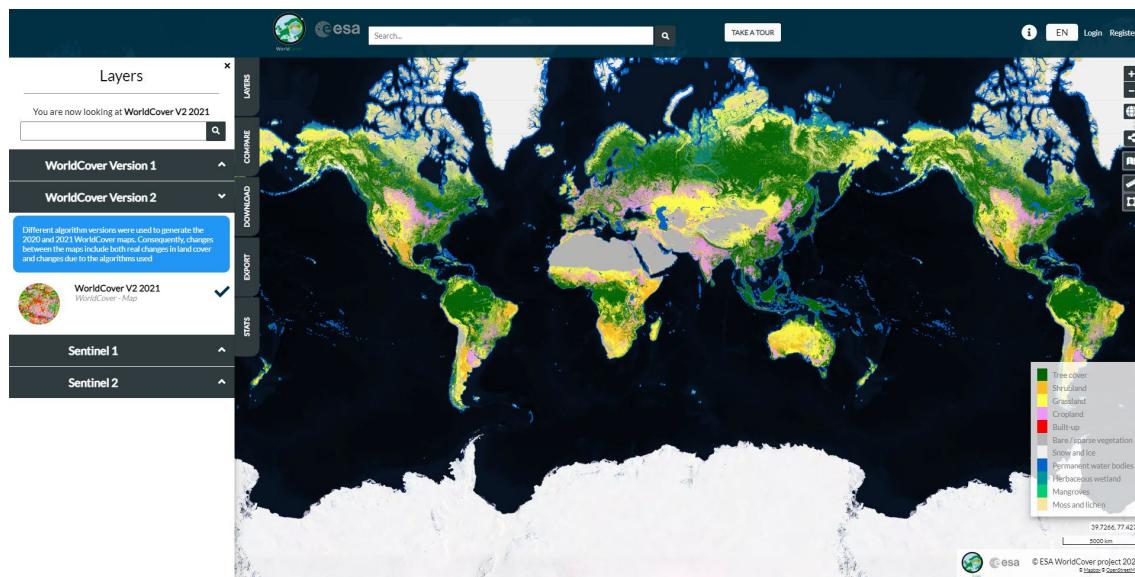
Una vez realizados todos estos pasos intermedios, estaremos en disposición de definir los polígonos de entrenamiento.

Como hemos dicho anteriormente para definir las clases sin visitar el terreno nos ayudaremos del WorldCover map, que se ha publicado el 24 de octubre de 2022. Se trata de un mapa de coberturas con resolución de 10 m que se ha realizado con una metodología similar a la que vamos a emplear nosotros en este tema.

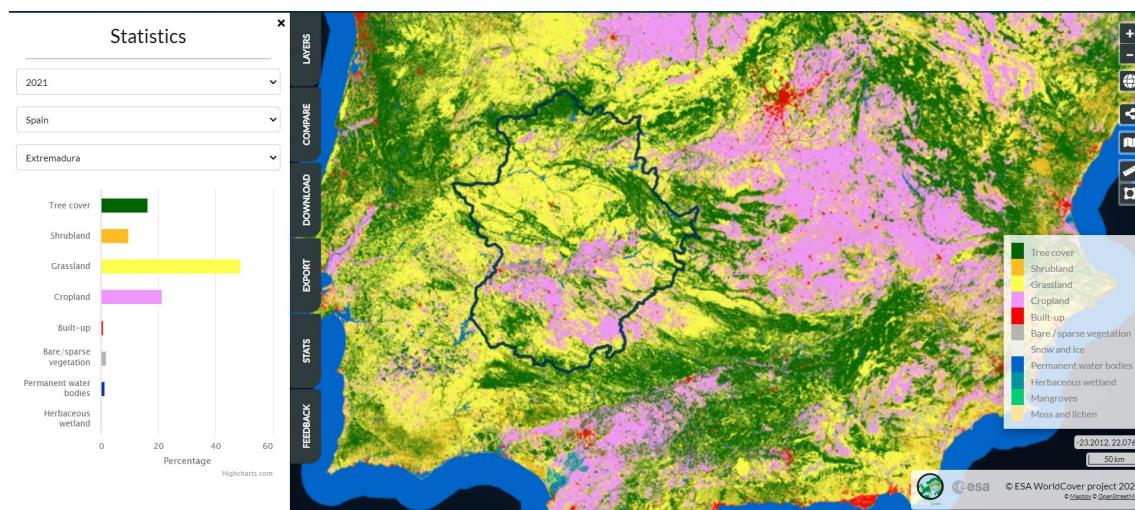
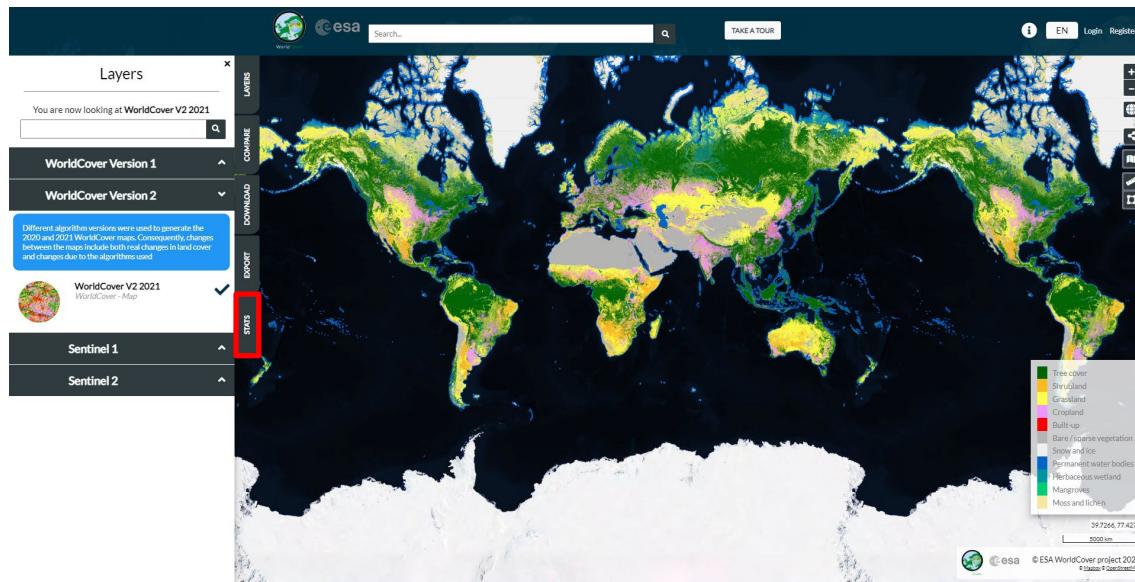


Fuente: WorldCover, Manual del usuario del producto

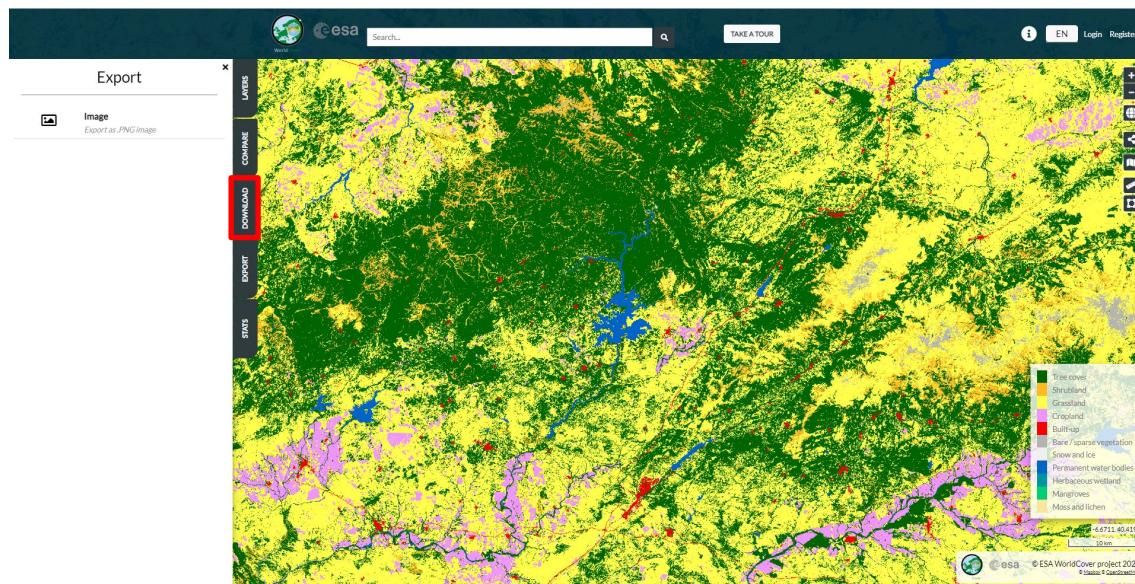
El mapa se puede consultar en la siguiente dirección: <https://esa-worldcover.org/en>



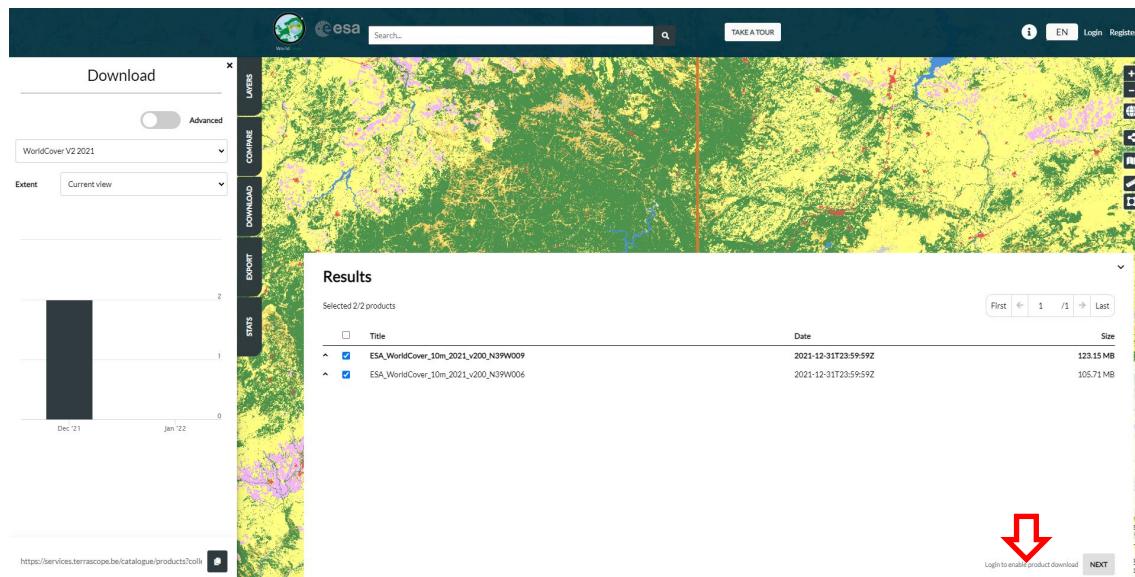
Si activamos la solapa de estadísticas, podremos mirar los porcentajes y tipos de cobertura de las distintas regiones



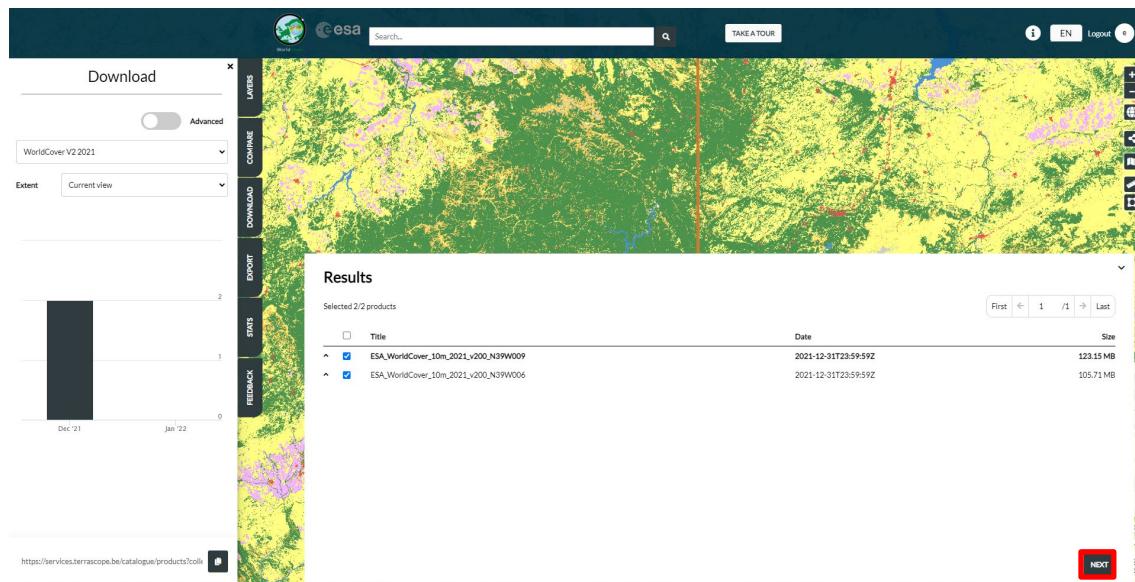
Para descargarnos el mapa en una zona concreta de trabajo habría que hacer zoom en el área en concreto e ir a la solapa de descargar.



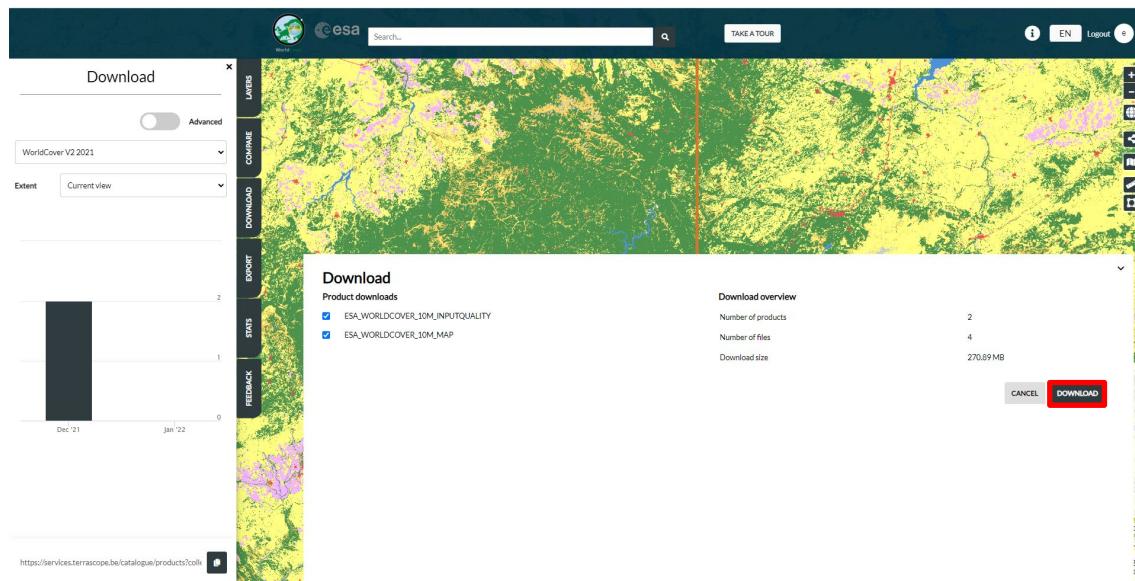
Veremos como nuestra zona de estudio está localizada entre dos hojas distintas del mapa, por lo que procederemos a descargar las dos.



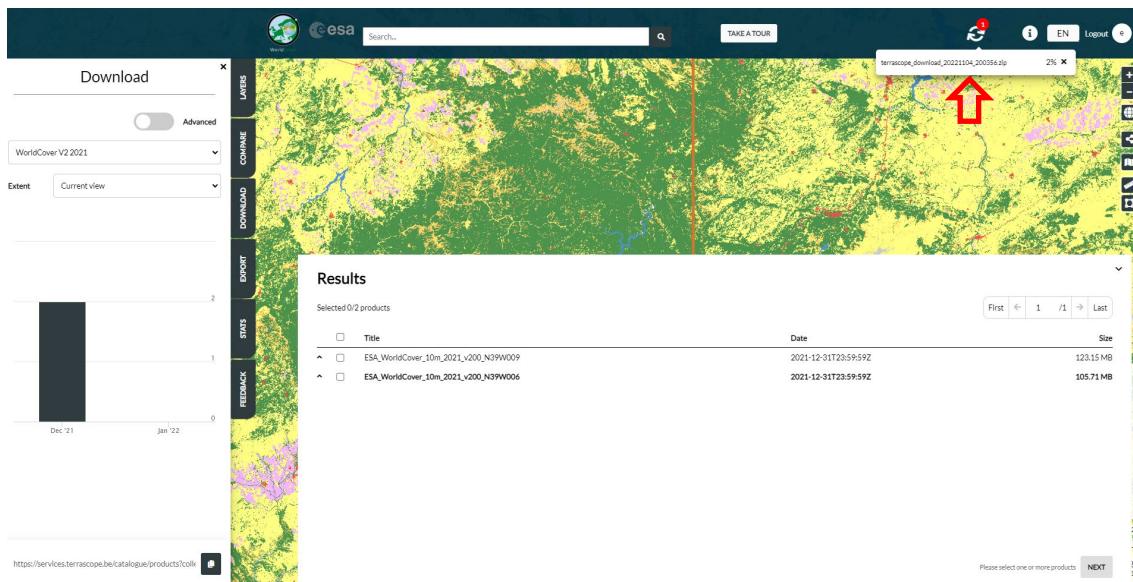
Para descargar, tendremos que darnos de alta en la aplicación y una vez logeados, se nos activará el botón para descargar las hojas:



Descargaremos tanto el mapa como la información de la calidad del mismo.



En la parte superior se indicará el estado del proceso de descarga hasta la finalización de la misma:

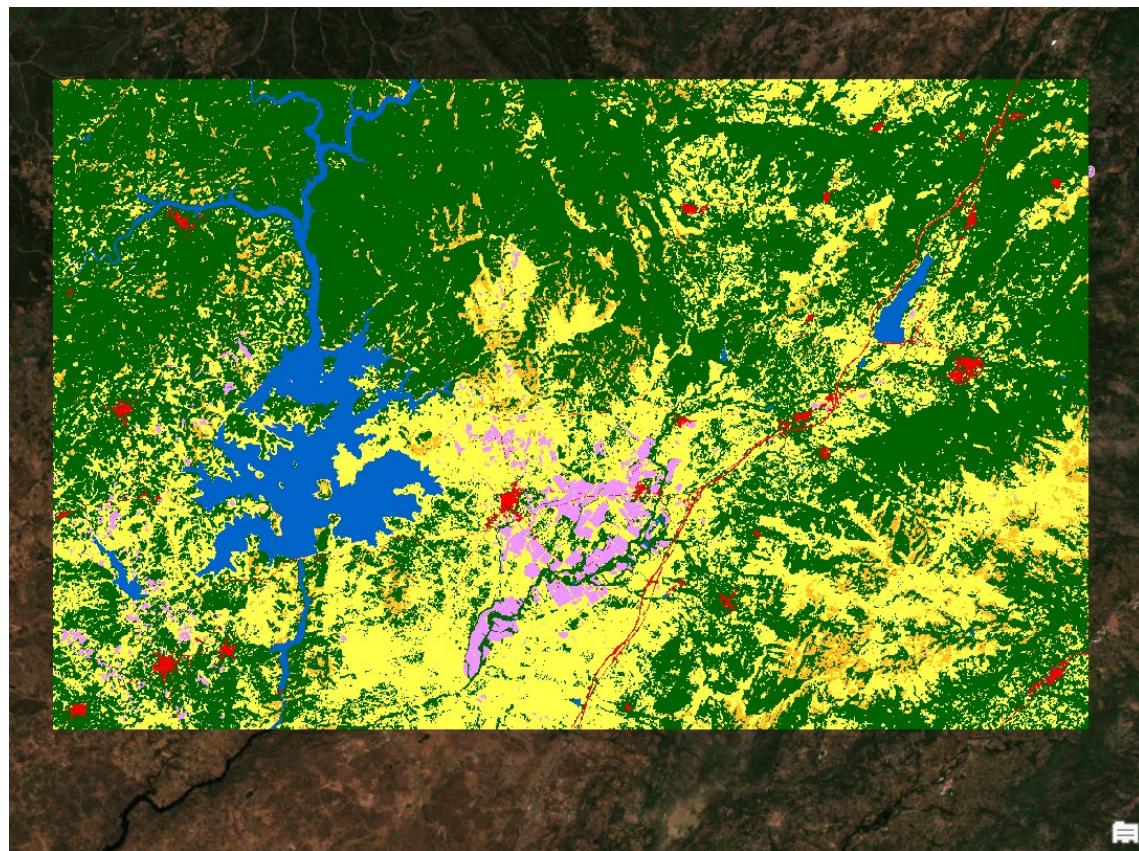
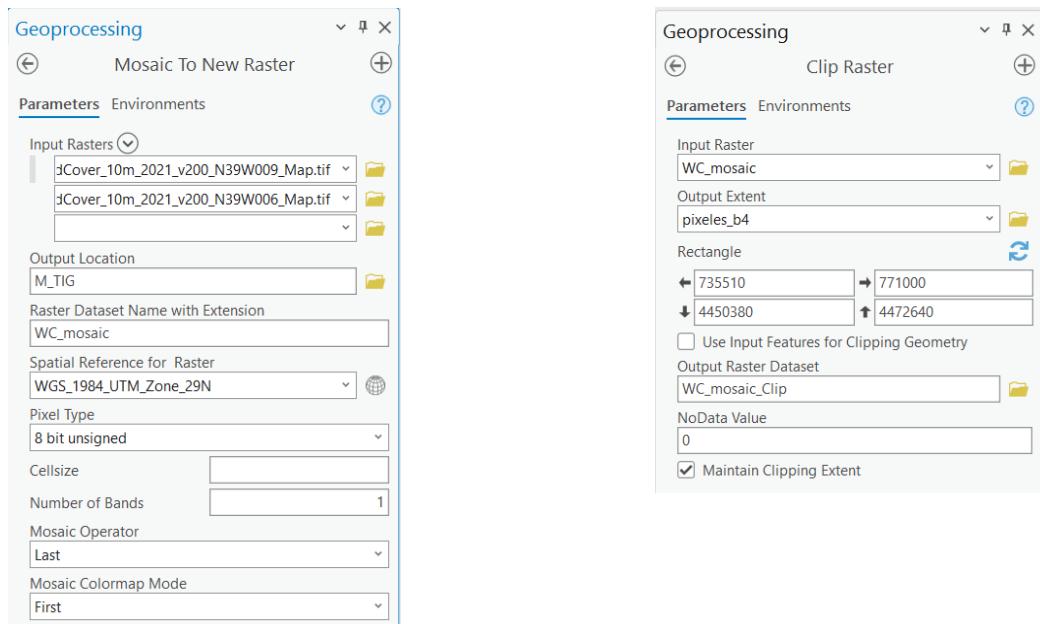


Las clases o categorías de clasificación del mapa son las que se muestran en la siguiente tabla. Estas clases son las generales, no teniendo por qué estar todas presentes en nuestra zona de estudio.

Código del mapa	Clase de cobertura del suelo	Código LCCS	Definición	Código de color (RGB)
10	Cubierta de árboles	A12A3 // A11A1 A24A3 C1(C2) - R1(R2)	Esta clase incluye cualquier área geográfica dominada por árboles con una cobertura del 10% o más. Otras clases de cobertura del suelo (arbustos y/o hierbas en el sotobosque, edificación, masas de agua permanentes, ...) pueden estar presentes bajo el dosel, incluso con una densidad superior a la de los árboles. Se incluyen en esta clase las áreas plantadas con árboles con fines de forestación y las plantaciones (por ejemplo, de palma aceitera, olivos). Esta clase también incluye las áreas cubiertas de árboles inundadas estacional o permanentemente con agua dulce, excepto los manglares.	0,100,0
20	Matorrales	A12A4 // A11A2	Esta clase incluye cualquier área geográfica dominada por arbustos naturales con una cobertura del 10% o más. Los arbustos se definen como plantas leñosas perennes con tallos persistentes y leñosos y sin ningún tallo principal definido de menos de 5 m de altura. Los árboles pueden estar presentes de forma dispersa si su cobertura es inferior al 10%. Las plantas herbáceas también pueden estar presentes en cualquier densidad. El follaje de los arbustos puede ser de hoja perenne o caduca.	255, 187, 34

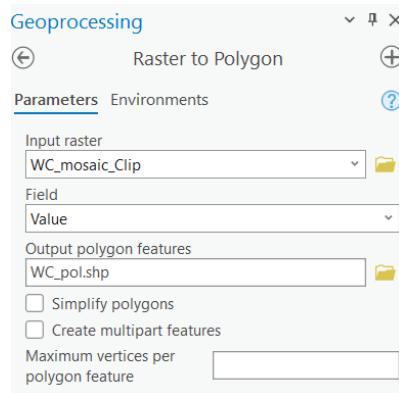
30	Pastizales	A12A2	Esta clase incluye cualquier área geográfica dominada por plantas herbáceas naturales (Plantas sin tallo o brotes persistentes sobre el suelo y que carecen de una estructura firme definida): (pastizales, praderas, estepas, sabanas, pastos) con una cobertura del 10% o más, independientemente de las diferentes actividades humanas y/o animales, como: pastoreo, gestión selectiva de incendios, etc. Las plantas leñosas (árboles y/o arbustos) pueden estar presentes siempre que su cobertura sea inferior al 10%. También puede contener zonas de cultivo no cultivadas (sin cosecha/suelo desnudo período) en el año de referencia	255, 255, 76
40	Tierras de cultivo	A11A3(A4)(A5) // A23	Tierra cubierta de cultivos anuales que se siembra/planta y se puede cosechar al menos una vez en los 12 meses siguientes a la fecha de siembra/plantación. La tierra de cultivo anual produce una cubierta herbácea y a veces se combina con alguna vegetación arbórea o leñosa. Obsérvese que los cultivos leñosos perennes se clasificarán como el tipo de cubierta vegetal arbórea o arbustiva correspondiente. Los invernaderos son se considera edificado.	240, 150, 255
50	Construido	B15A1	Terreno cubierto por edificios, carreteras y otras estructuras hechas por el hombre, como las vías férreas. Los edificios incluyen tanto los residenciales como los industriales. El verde urbano (parques, instalaciones deportivas) no se incluye en esta clase. Los depósitos de residuos y los lugares de extracción se consideran desnudo.	250, 0, 0
60	Vegetación desnuda / escasa	B16 A1(A2) // B15 A2	Terrenos con suelo expuesto, arena o rocas y nunca tiene más de 10 % de cobertura vegetal durante cualquier época del año	180, 180, 180
70	Nieve y hielo	B28A2(A3)	Esta clase incluye cualquier área geográfica cubierta por nieve o glaciares de forma persistente	240, 240, 240
80	Masas de agua permanente	B28 A1(B1) // B27 A1(B1)	Esta clase incluye cualquier área geográfica cubierta durante la mayor parte del año (más de 9 meses) por masas de agua: lagos, embalses y ríos. Pueden ser masas de agua dulce o salada. En algunos casos el agua puede congelarse durante una parte del año (menos de 9 meses).	0, 100, 200
90	Humedal herbáceo	A24A2	Terreno dominado por vegetación herbácea natural (cobertura del 10% o más) que está permanente o regularmente inundado por agua dulce, salobre o salada. Excluye los sedimentos sin vegetación (véase 60), los pantanos bosques (clasificados como manglares ver 95)	0, 150, 160

Una vez abiertas las dos hojas en ArcGIS Pro, lo primero que haremos será unir y recortar el mapa por nuestra zona de estudio.

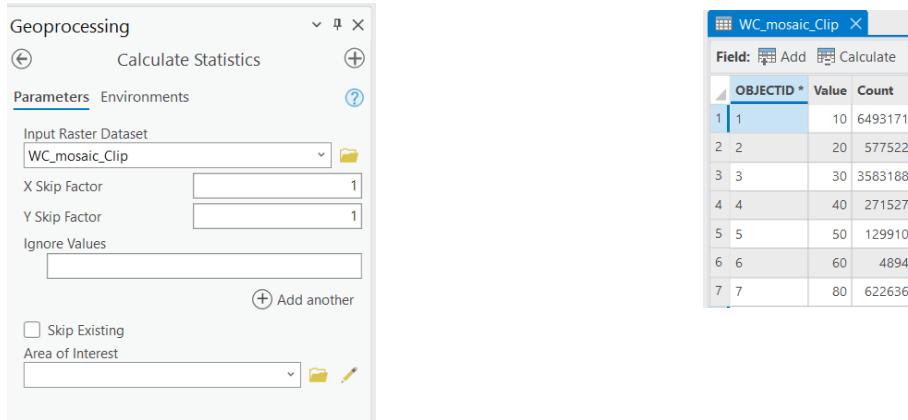


Si hacemos un zoom hasta llegar al tamaño del pixel de nuestra imagen, podremos ver que los píxeles del mapa no coinciden con los de la imagen de trabajo. No obstante, en este momento no faremos ninguna operación para subsanarlo, puesto que, para el entrenamiento, extraeremos sólo los píxeles que tengan el 100% del valor de su superficie única.

Para ello en primer lugar, vamos a poligonizar el raster del WorldCover.

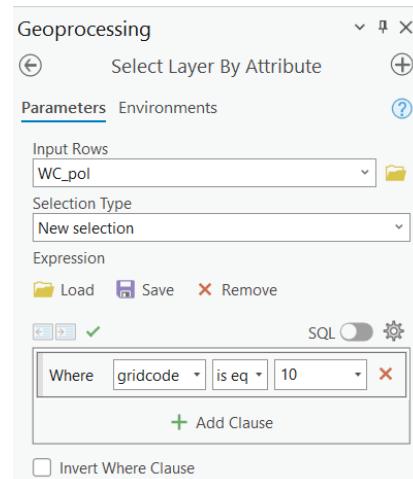


Una vez poligonizado, procederemos a seleccionar un porcentaje pequeño y aleatorio de polígonos para que nos sirvan de muestras de entrenamiento. Por tanto, necesitaremos tener una idea de qué clases, según el WorldCover, son las que cubren nuestra área de estudio y qué cantidad de polígonos tiene cada una:

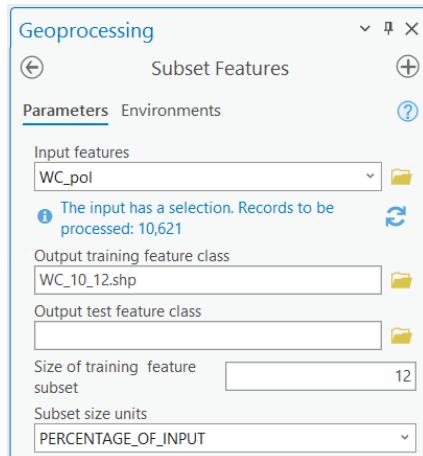


Analizando los datos, vemos como la clase 60 es la que menos ocupa en nuestra imagen y la 10 la que más.

Entonces ya estaremos en disposición de ir seleccionando clase por clase y posteriormente hacer subconjuntos aleatorios proporcionales. Por ejemplo, para la clase 10 procederemos de la siguiente forma:



Una vez seleccionado sólo esa clase, procedemos a hacer el subconjunto del 12% (cada estudiante lo puede hacer del porcentaje que estime oportuno) y guardaremos esos polígonos como un archivo independiente:



Haremos lo mismo para cada clase, tomando la precaución de inspeccionar visualmente la selección y, si nos elige algún polígono demasiado grande, habremos de eliminar el subconjunto y volver a seleccionar aleatoriamente, para que las áreas de entrenamiento no queden demasiado extensas, y, al contrario:

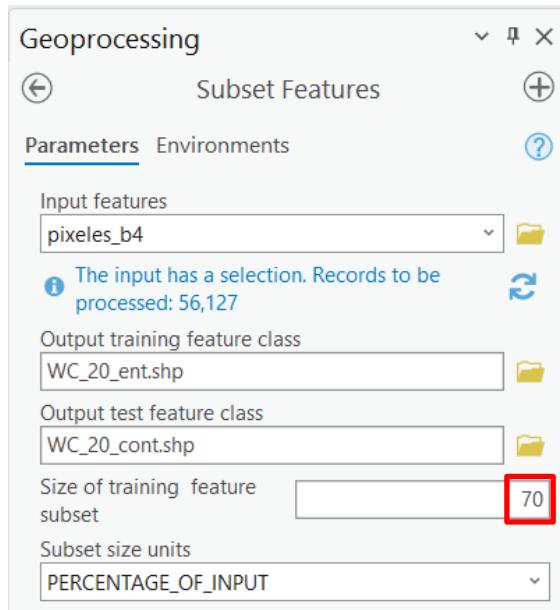
Clase	Porcentaje
10	12
20	20
30	20
40	20
50	30
60	100
80	10

De ese modo, tendremos ya un fichero *shp* con ciertos polígonos para cada clase. Ahora tendremos que buscar, dentro de cada uno de ellos, los píxeles que vamos a emplear para entrenar el clasificador y los que vamos a emplear para hacer un control del error. Como con anterioridad observamos que los píxeles del WordCover no correspondían espacialmente con los de nuestra imagen de trabajo, vamos a seleccionar, para asegurarnos pureza de clases, sólo los píxeles de la imagen que están por completo dentro de los polígonos de cada clase.

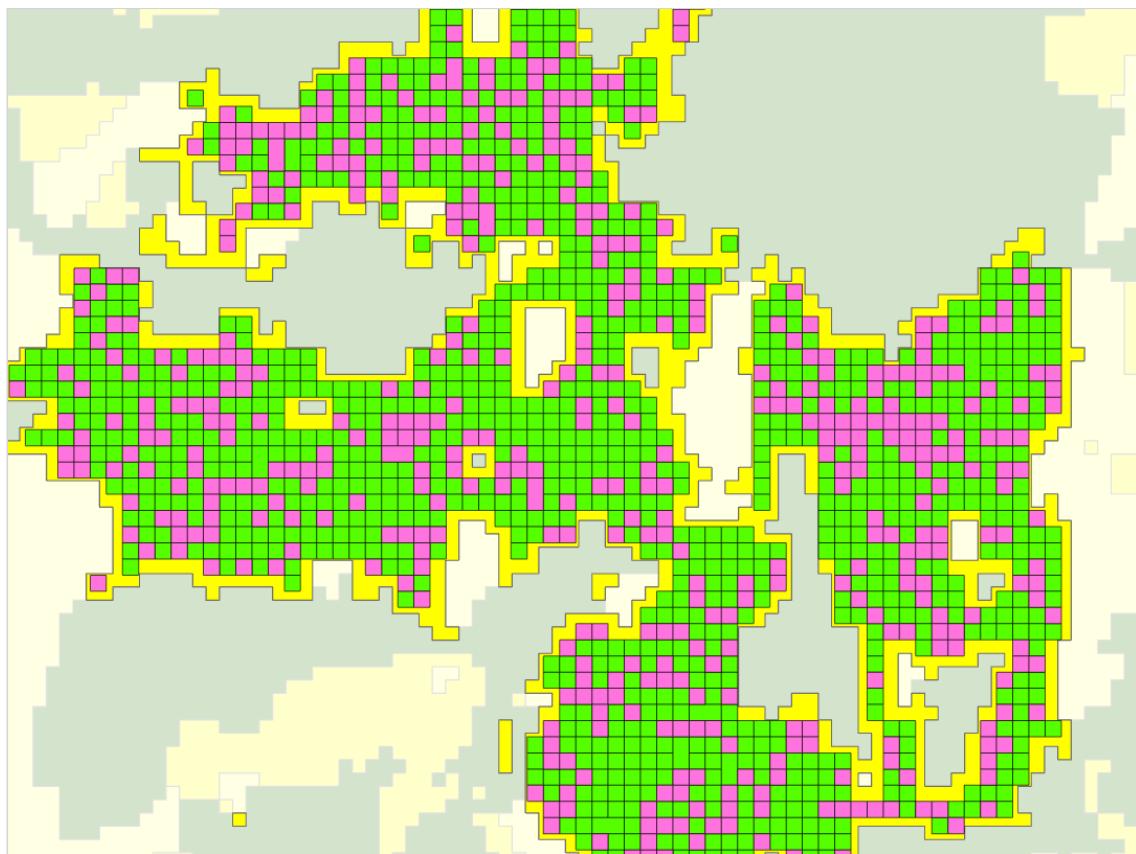
Para ello, haremos la misma operación para cada una de las clases. Por ejemplo, para la 20 (de la que hemos seleccionado el 20% de los polígonos) procederemos de la siguiente forma:



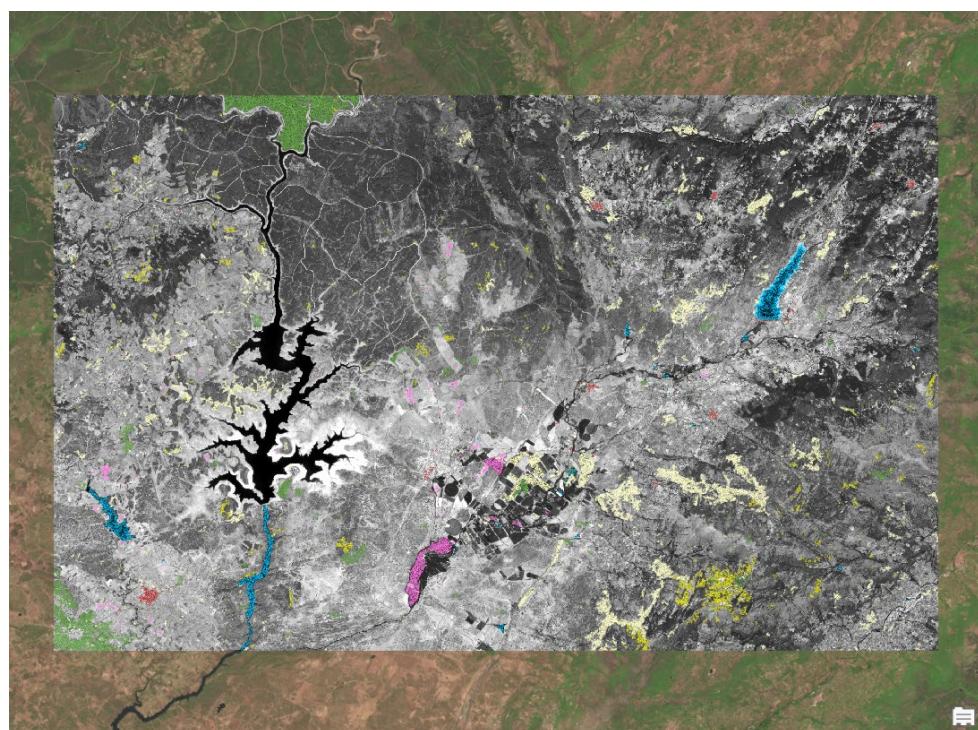
Una vez seleccionados los contornos de los píxeles que están completamente dentro, haremos dos subconjuntos: El 70% lo emplearemos para el entrenamiento de la clasificación supervisada y el 30% para el control del error del tema 4.



De ese modo, podremos observar cómo nuestros píxeles, tanto de entrenamiento (color verde), como de control (color magenta), son todos “puros” según el WorldCover y están aleatoriamente distribuidos.



Procederemos de la misma forma con todas las clases para finalmente tener todas las áreas de entrenamiento definidas (y las de control que emplearemos en el tema 4).

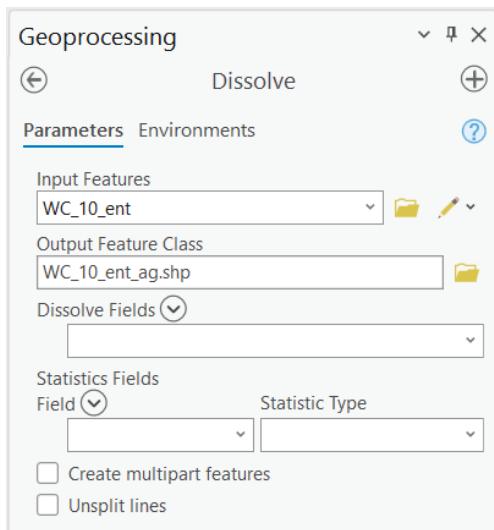


Para hacernos una idea del número aproximado de píxeles que emplearemos para clasificar, consultaremos uno por uno los ficheros de entrenamiento generados.

Clase	Nº de píxeles (aprox)
10	74.318
20	39.289
30	136.075
40	18.346
50	7.945
60	676
80	25.205

Aunque SNAP lee archivos *shp*, las geometrías de dichos *shp* no deben ser muy complejas, porque en caso contrario, pueden ocurrir errores de importación. Además, si intentamos importar un número muy elevado de polígonos, como es nuestro caso, el tiempo de procesado puede ser demasiado extenso. Por tanto, vamos a realizar las dos últimas operaciones antes de importar nuestras áreas de entrenamiento en SNAP.

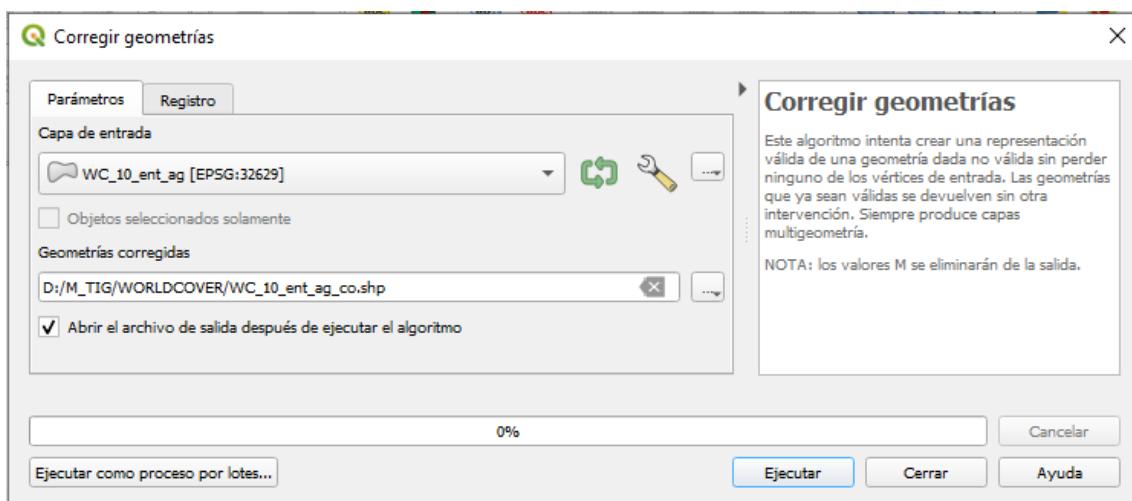
En primer lugar, vamos a agrupar o disolver los polígonos para disminuir su número.



Al hacer esta operación podremos observar cómo se han aligerado mucho el número de polígonos con el que trabajaremos:



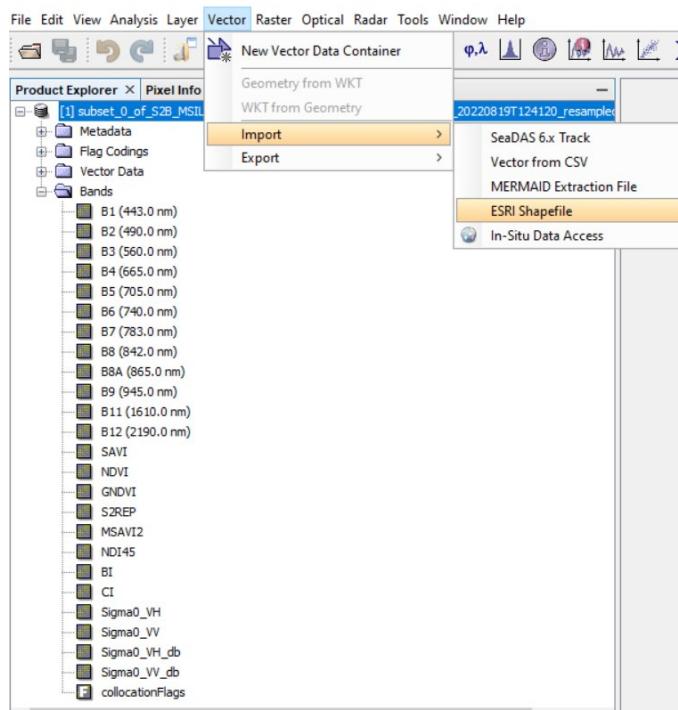
Para que SNAP pueda leer estas geometrías tan complejas, tendremos que utilizar Qgis y reparar su geometría.



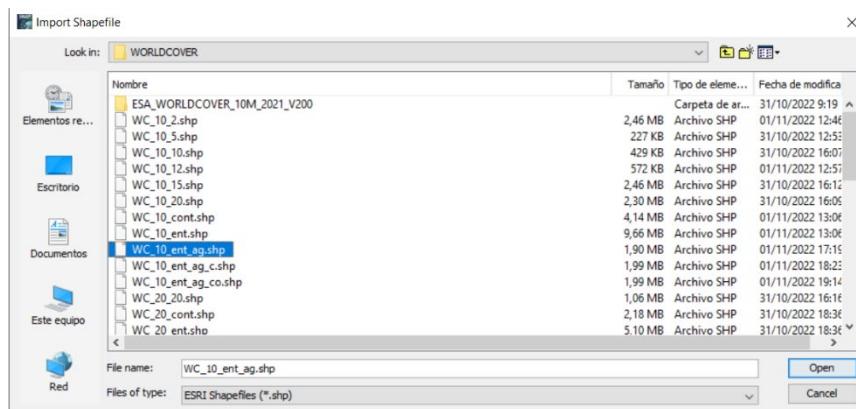
Una vez finalizadas todas estas operaciones en todas las categorías del mapa, comenzaremos con la clasificación.

2.2.2 Importación de las áreas de entrenamiento en SNAP

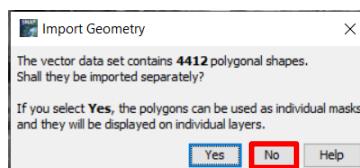
En primer lugar, cargaremos en SNAP los vectores de las áreas de entrenamiento de cada una de las clases.



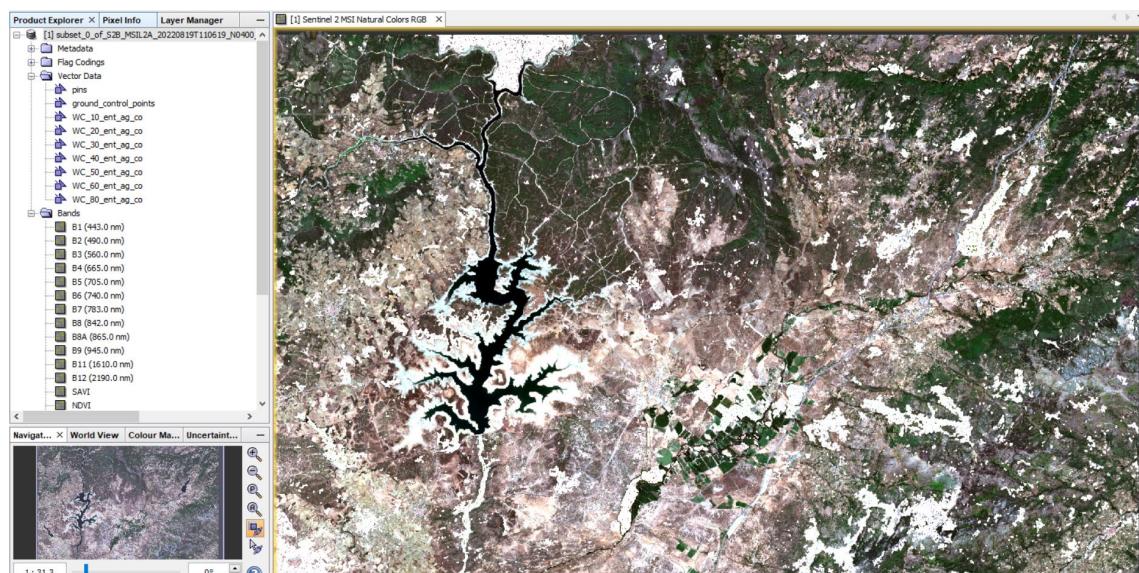
Iremos archivo por archivo cargándolo de forma independiente. Por ejemplo, cargando la clase 10:



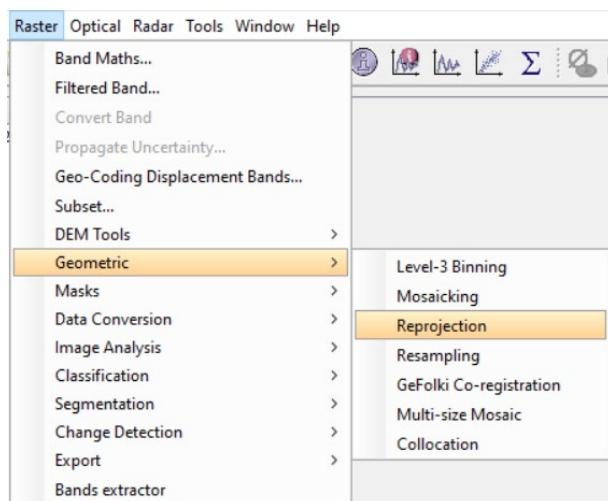
Le diremos que no queremos cargar cada polígono de agua como archivo independiente, sino que todos los polígonos estén dentro de la misma capa:

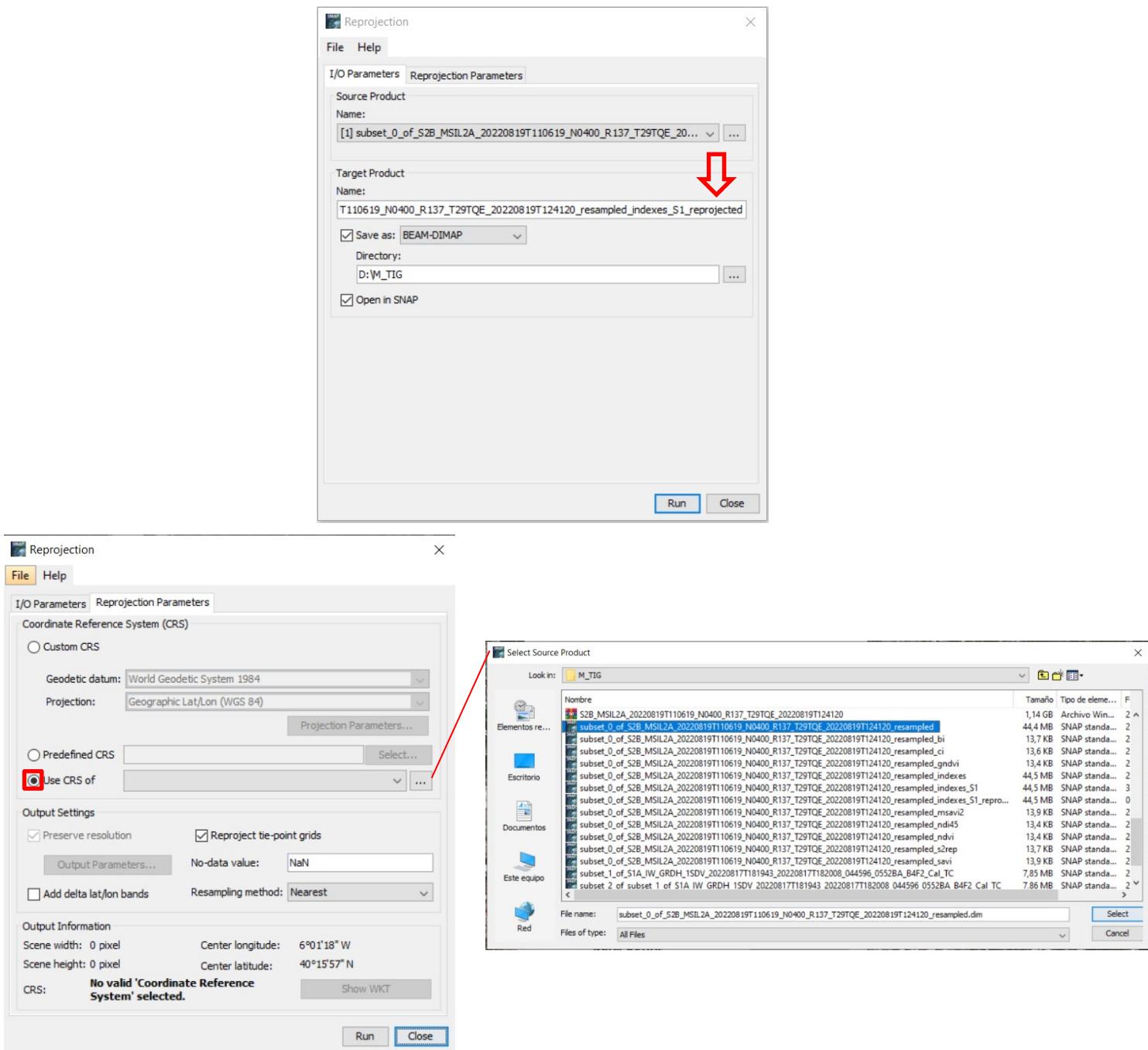


Procederemos de la misma forma para todas las clases que tenemos exportadas (el proceso, si tiene muchos vectores puede tardar varios minutos).

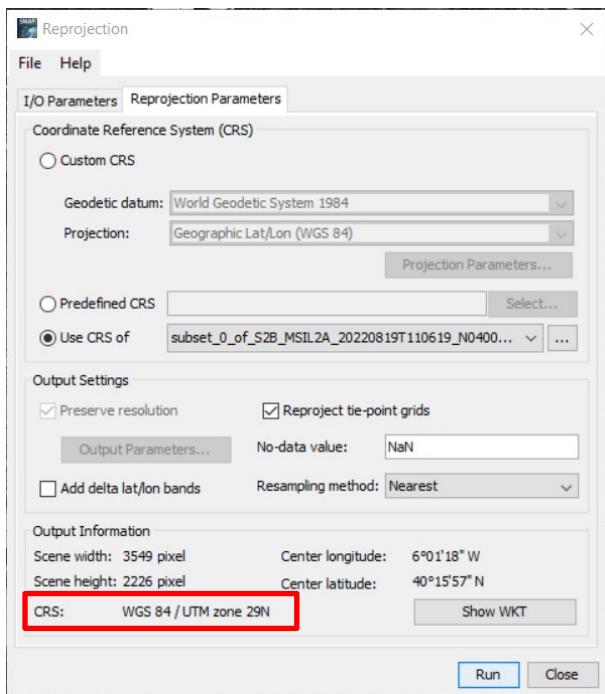


Esta versión del programa SNAP, pierde la proyección al introducir vectores en las imágenes y, al realizar operaciones posteriores, puede dar problemas de cálculo. Por ello, una vez importados los vectores, procederemos a reproyectar la imagen a su sistema de coordenadas:





Entonces comprobaremos que ha cogido correctamente el CRS:

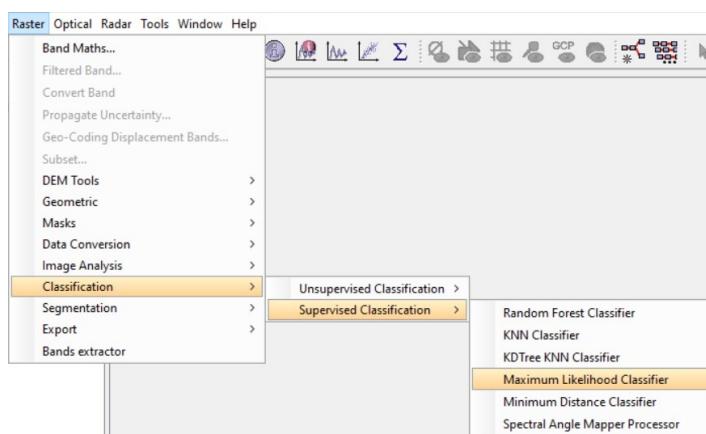


De este modo, ya tendríamos finalizada la imagen, con los vectores de entrenamiento cargados para proceder a su clasificación.

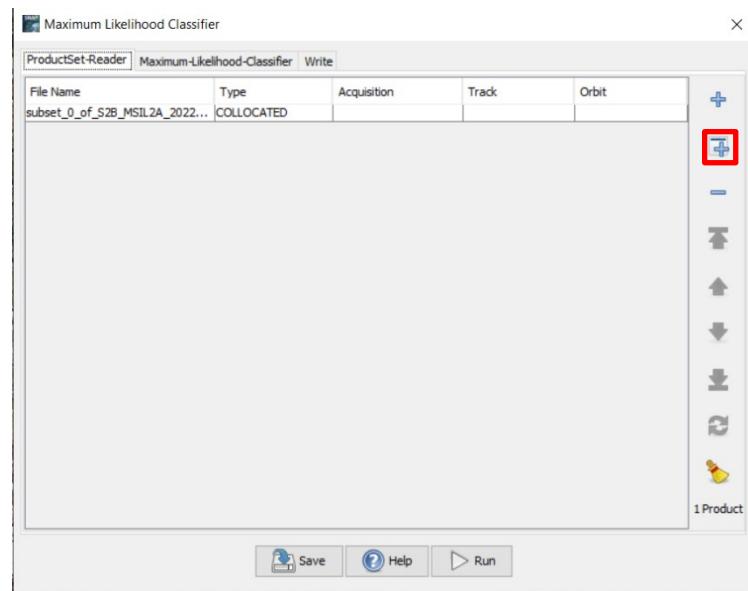
2.2.3 Algoritmo de clasificación de máxima verosimilitud

Este método se engloba dentro de los denominados clasificadores cuadráticos y considera que los valores espectrales de cada clase se ajustan a una distribución normal. Esto nos permite describir esa categoría por una función de probabilidad, a partir de su vector de medias y matriz de varianza-covarianza. Para determinar a qué clase pertenece un determinado píxel, se calcula la probabilidad de pertenencia a todas las clases y se asignará a la clase a la cual hayamos obtenido la máxima probabilidad.

Para ejecutar el algoritmo, lo buscaremos dentro del grupo de los clasificadores supervisados.

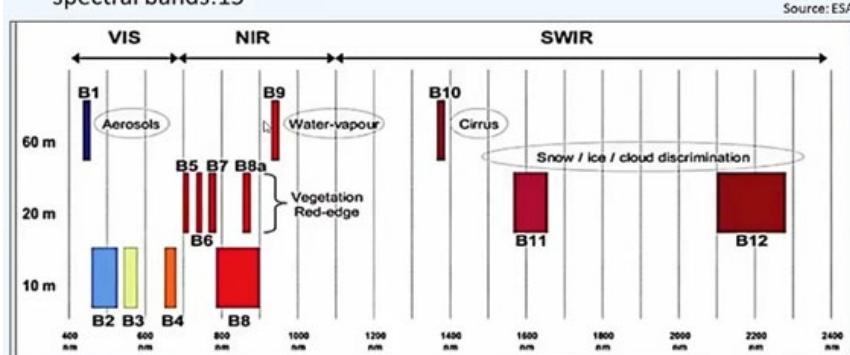


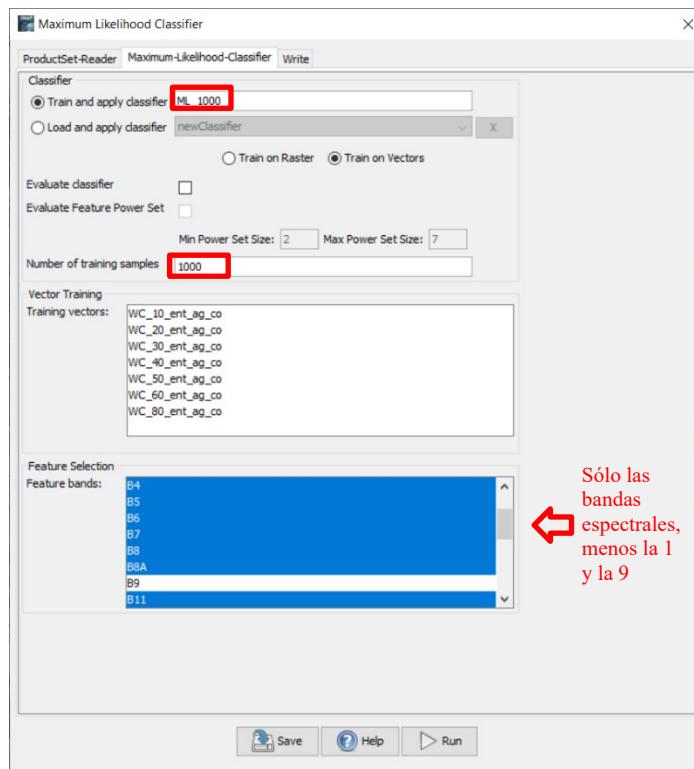
En ese momento nos pedirá, en la primera solapa, la imagen que queremos clasificar. Para ello, cargamos todas las imágenes que haya en la sesión. En este caso sólo está abierta la S2 con los índices, y será por tanto la única que nos ofrezca.



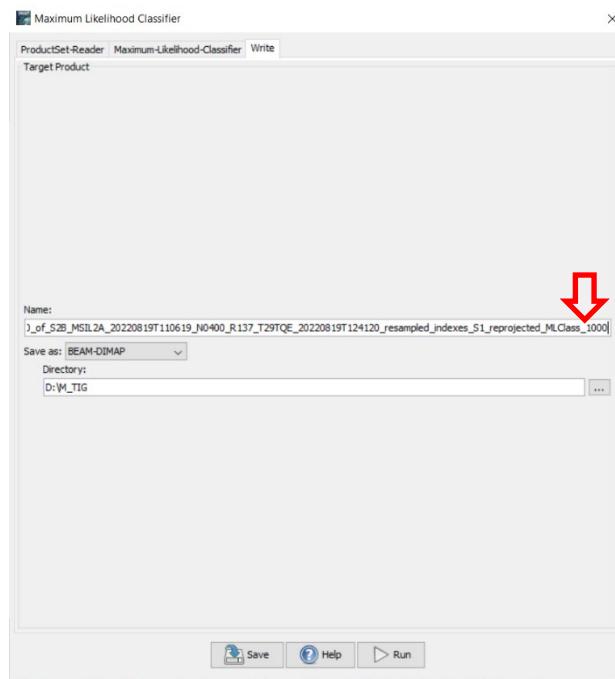
La siguiente solapa, nos solicita los parámetros propios con los que vamos a configurar la clasificación. SNAP tiene la posibilidad de guardar la configuración, por si en otras ocasiones la queremos volver a utilizar. Por ello, guardaremos esa configuración con el nombre “ML_1000”. También nos ofrece la posibilidad de entrenar con otro ráster o con vectores de entrenamiento. En este caso, utilizaremos esta segunda opción, puesto que ya los tenemos cargados. Seguidamente le indicaremos el número de píxeles con los que vamos a entrenar. En nuestro caso, a la vista del número de píxeles aproximado de cada una de las clases, pondremos 1000. A continuación, solicita las clases que vamos a clasificar, si no marcáramos ninguna, por defecto SNAP utilizará todas. Finalmente hay que indicar qué bandas se emplean para la clasificación. Por tanto, le marcaremos todas las bandas espectrales propias del S2 **menos la 1 y la 9** que están destinadas al uso atmosférico.

- Sentinel-2A and 2B: polar orbit, phased at 180° to each other
- Multispectral instrument with spectral bands:13
- Spatial resolution: 10m / 20m / 60m (based on spectral band)
- Revisit: 5 days at equator

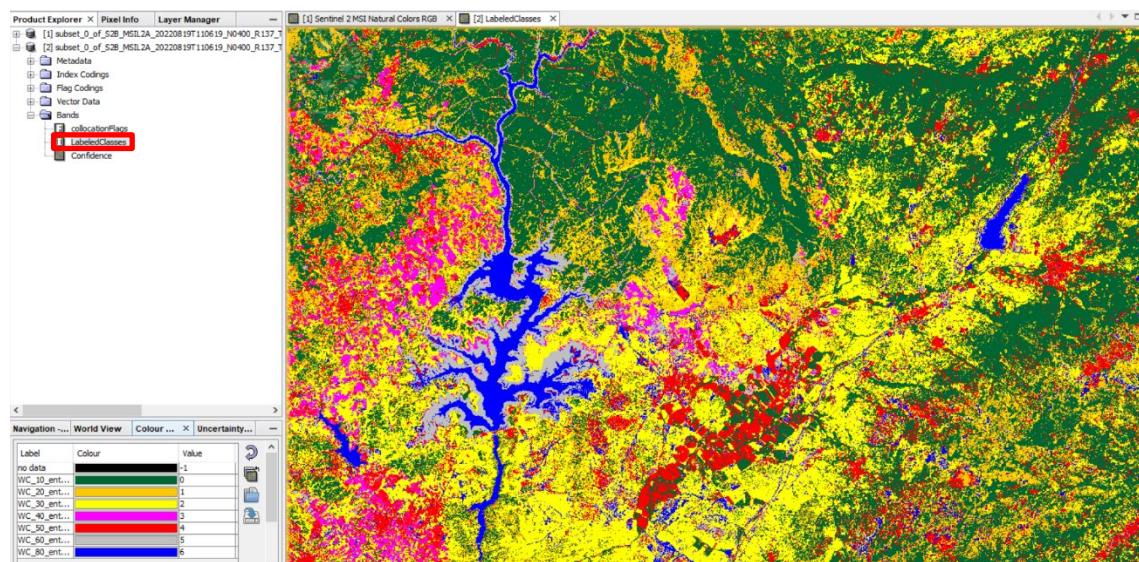




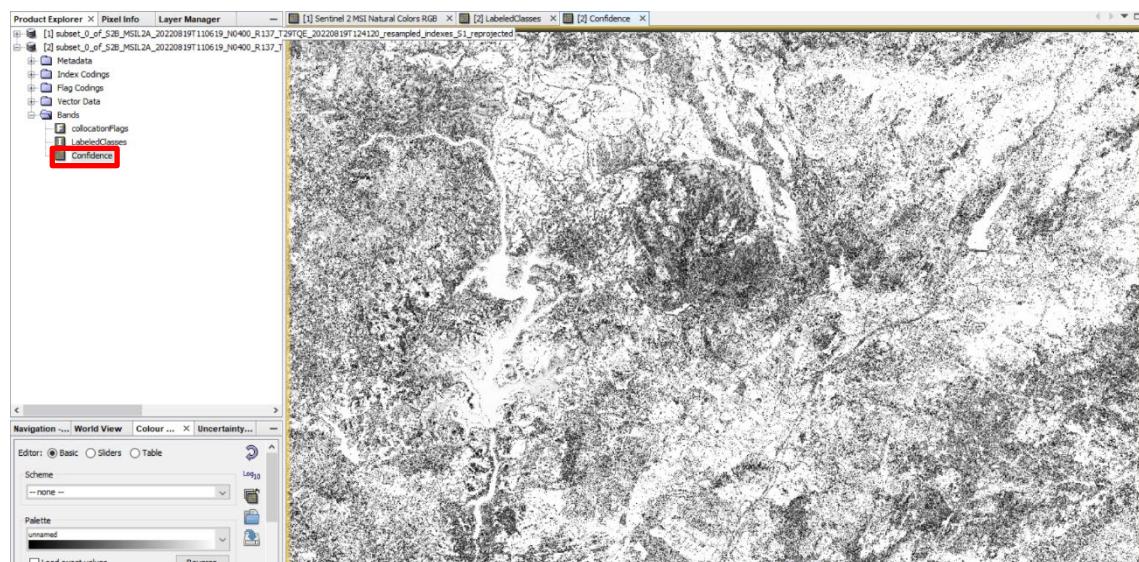
Finalmente, en la última solapa configuraremos el archivo de salida, en el que nos ofrece, como siempre el mismo nombre con un sufijo, en este caso terminado en “MLClass”.



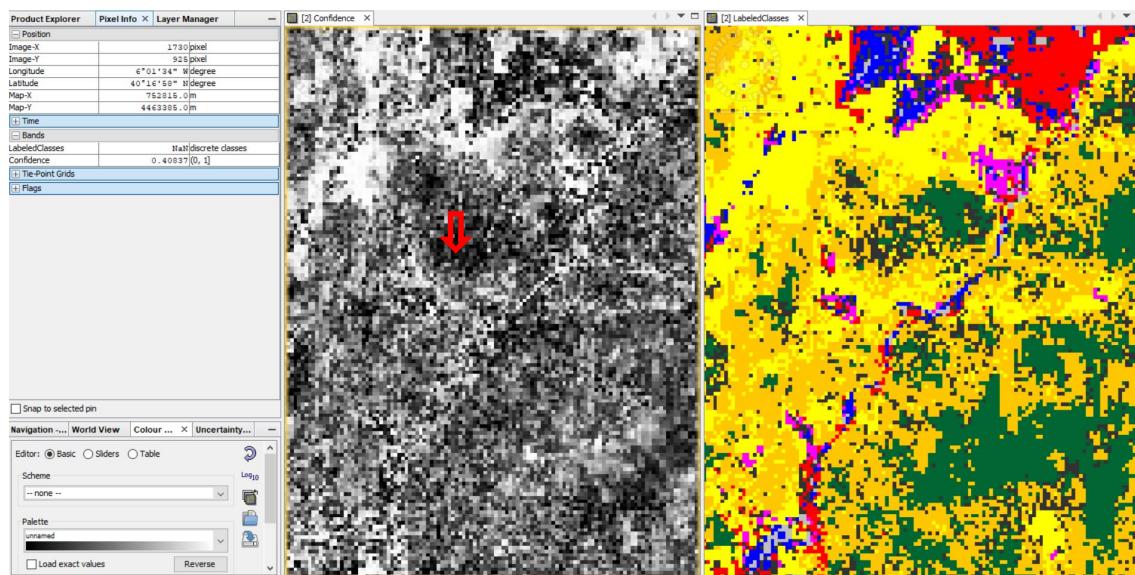
Una vez ejecutado el clasificador, a los pocos segundos, en el explorador de productos tendremos el resultado.



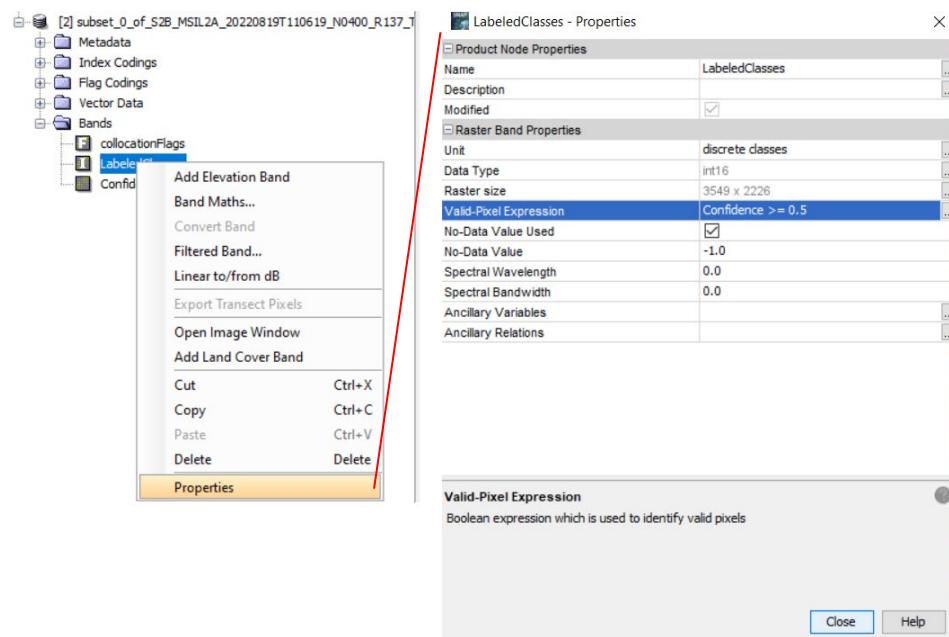
Pudiendo comprobar que también se ha creado una banda llamada “confidence” en la que se muestra el nivel de confianza de la clasificación.



Se puede observar cómo los píxeles clasificados con una menor confianza, no se representan en la ventana de la clasificación, aunque si que tienen asignada una clase.



El límite de confianza de la clasificación para que sea representado se puede cambiar en las propiedades de la banda:



Ejercicio propuesto 1:

Variar los parámetros/clases de la clasificación para lograr obtener una clasificación mejor.

2.2.4 Algoritmo clasificación Random Forest

Los clasificadores paramétricos supervisados, como la clasificación de Máxima Verosimilitud (MLC), ofrecen excelentes resultados cuando se trata de datos unimodales (de un solo sensor). Sin embargo, tienen limitaciones al tratar con conjuntos de datos de entrada multimodales, porque estos clasificadores suponen una distribución de datos normal.

Random Forest (RF) es una técnica clasificación creada por Leo Breiman en 2001 que mejora la precisión en la clasificación mediante la incorporación de aleatoriedad.

Esta aleatorización puede introducirse tanto por la partición del espacio (construcción del árbol), así como en la muestra de entrenamiento.

En las últimas dos décadas, el uso del clasificador Random Forest ha recibido una creciente atención debido a los excelentes resultados de clasificación obtenidos y la velocidad de procesamiento.

El clasificador de RF produce clasificaciones confiables usando predicciones derivadas de un conjunto de árboles de decisión. Además, este clasificador se puede utilizar con éxito para seleccionar y clasificar aquellas variables con mayor capacidad de discriminación entre las clases objetivo.

Para entender cómo funciona el algoritmo imaginaremos que queremos clasificar la población entre hombres y mujeres.



Para ello necesitaremos características o atributos de los elementos que queremos clasificar. Por ejemplo:

- Altura
- Peso
- Color de ojos
- Color de pelo
- Longitud de pelo
- ...

Dichas características pueden ser o no las que nos ayuden a discriminar entre hombres y mujeres.

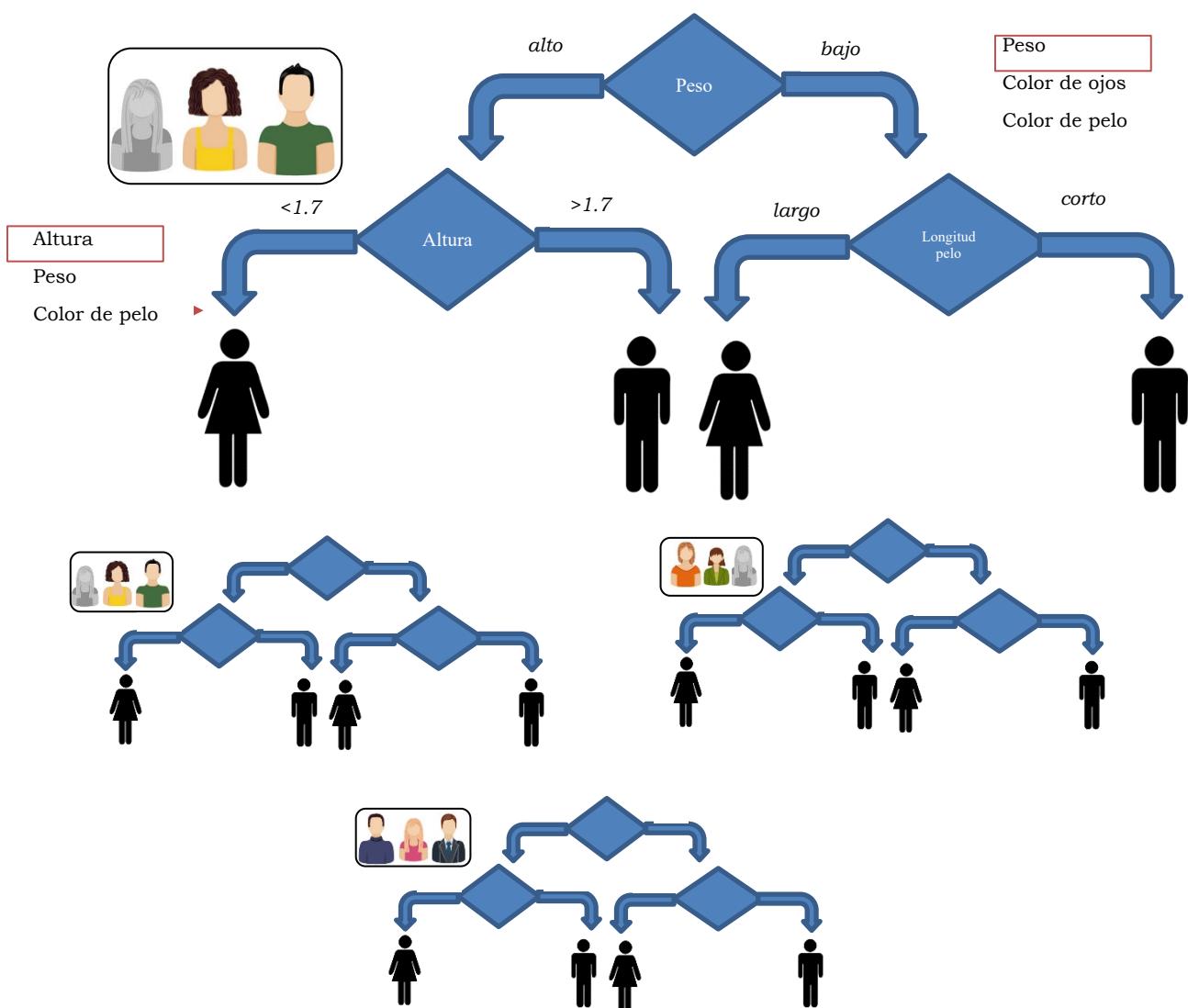
Como realizaremos una clasificación supervisada, dispondremos de unas muestras de entrenamiento (sujetos) de las que podremos extraer sus características:



El algoritmo escogería un conjunto de las muestras de entrenamiento aleatoriamente y las devuelve al total de la muestra, de tal modo que, al coger otro nuevo conjunto, puede repetir sujetos:

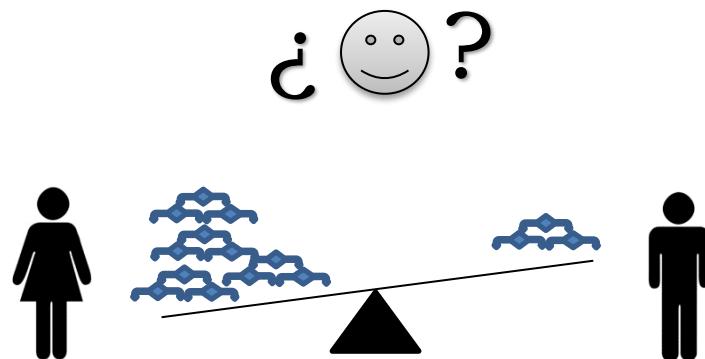


El siguiente paso será el de construir los árboles de decisión binarios haciendo preguntas, en cada subconjunto, sobre las características o atributos que vayamos a estudiar, que también serán seleccionados aleatoriamente:



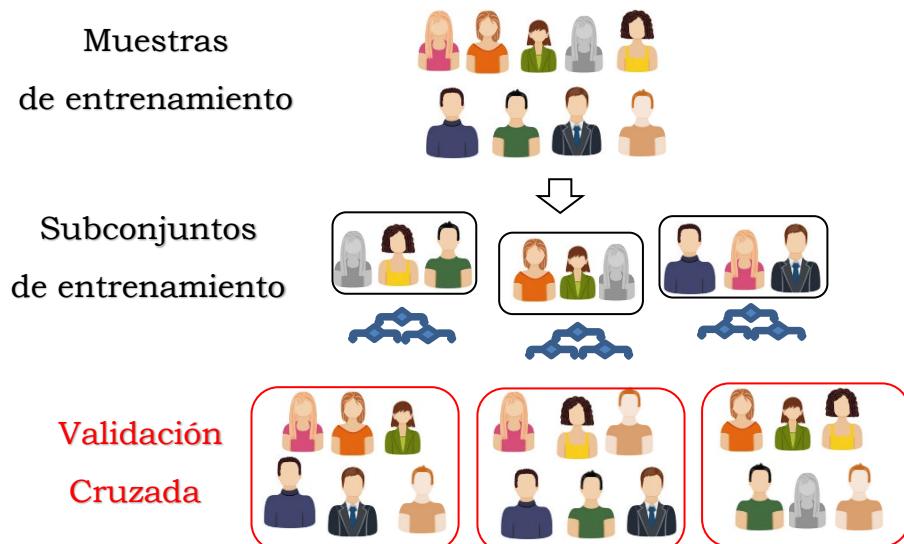
Con cada subconjunto de la muestra se construirá un árbol que tendrá sus propias reglas de clasificación entre hombres y mujeres. Por tanto, tendremos un grupo de árboles de decisión que formarán el “bosque” al que hace alusión el nombre del clasificador.

Una vez terminada la fase de entrenamiento, habrá que clasificar cualquier sujeto del globo terráqueo mediante estas reglas. De tal manera que cada individuo será sometido a las reglas formadas en la fase de entrenamiento y, por tanto, según cada árbol construido en dicha fase, el individuo será clasificado como hombre o como mujer. De ese modo, el individuo será clasificado en la clase que tenga más árboles a su favor.

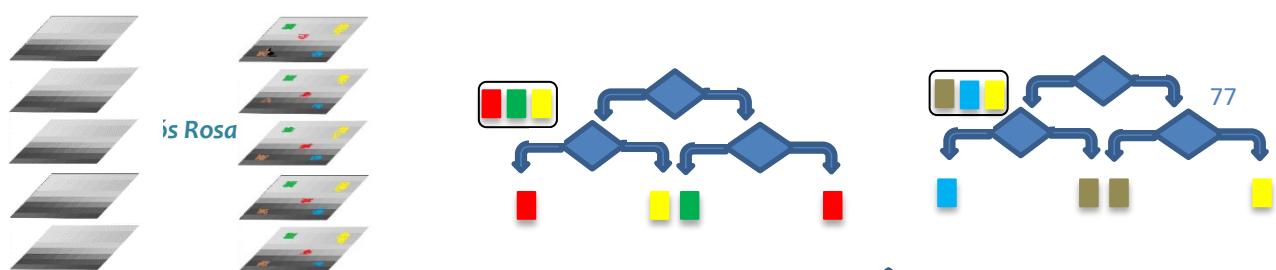


Además, el algoritmo de clasificación RF, al trabajar con subconjuntos de las muestras de entrenamiento, tiene una forma de estimar el error del entrenamiento (que no es el error de la clasificación final) y que nos da una idea de si las muestras y las características o atributos son los correctos.

Así, con las muestras de entrenamiento con las que no construye el árbol de decisión, comprueba el error, ya que al ser muestras de entrenamiento, se conoce si son hombres o mujeres.



En el caso de la teledetección, tenemos que clasificar una imagen con una serie de bandas compuestas a su vez por un número de píxeles. Los atributos o características serán las bandas de la imagen y las muestras de entrenamiento serán ciertos píxeles de clases conocidas.



Alrededor de dos tercios de las muestras de entrenamiento son utilizadas para construir los árboles y el tercio restante para la validación cruzada.

El usuario define el **número de características** que se seleccionan aleatoriamente y el **número de árboles** que se crearán. Las investigaciones actuales resaltan que la exactitud de la clasificación es menos sensible al número de árboles que al atributos o características (bandas de la imagen). Como el clasificador RF es computacionalmente muy eficiente, el número de árboles puede ser muy elevado. La mayoría de los estudios de investigación utilizan 500 árboles porque los errores se estabilizan antes de que se logre este número de árboles de clasificación, pero también existen trabajos de investigación que utilizan menos árboles y obtienen resultados con buena exactitud.

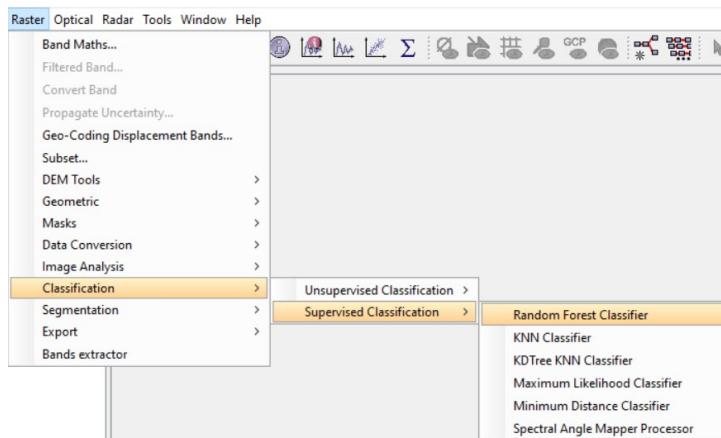
Por otro lado, el número de características del entrenamiento (número de bandas del entrenamiento) generalmente se establece en la raíz cuadrada del número de variables de entrada (número de bandas total de la imagen). Normalmente, ese valor es la raíz cuadrada del número de bandas de la imagen a clasificar.

Si nos centramos en el clasificador RF, existen estudios recientes que indican que las muestras de entrenamiento, además de cumplir con los requisitos generales para cualquier clasificación supervisada, deberían representar el 0.25% del área de estudio.

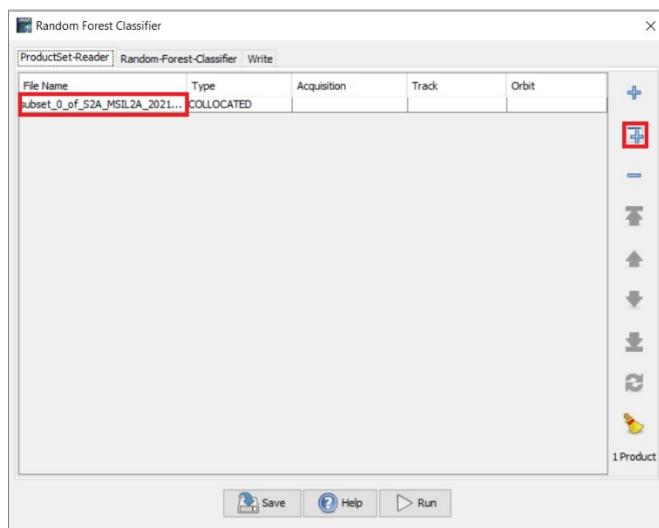
2.2.5 Clasificación Random Forest con SNAP

La mayor ventaja del clasificador Random Forest, como se ha dicho anteriormente, la capacidad de trabajar con datos multidimensionales. De tal modo que, aprovechando dicha disponibilidad, podremos combinar distintos tipos de sensores en la clasificación y extraer sacar las mejores ventajas de cada uno de ellos. Por ese motivo, con la combinación de varias fuentes de información (S2+IV+IS+S1) podrían mejorarse los resultados la clasificación.

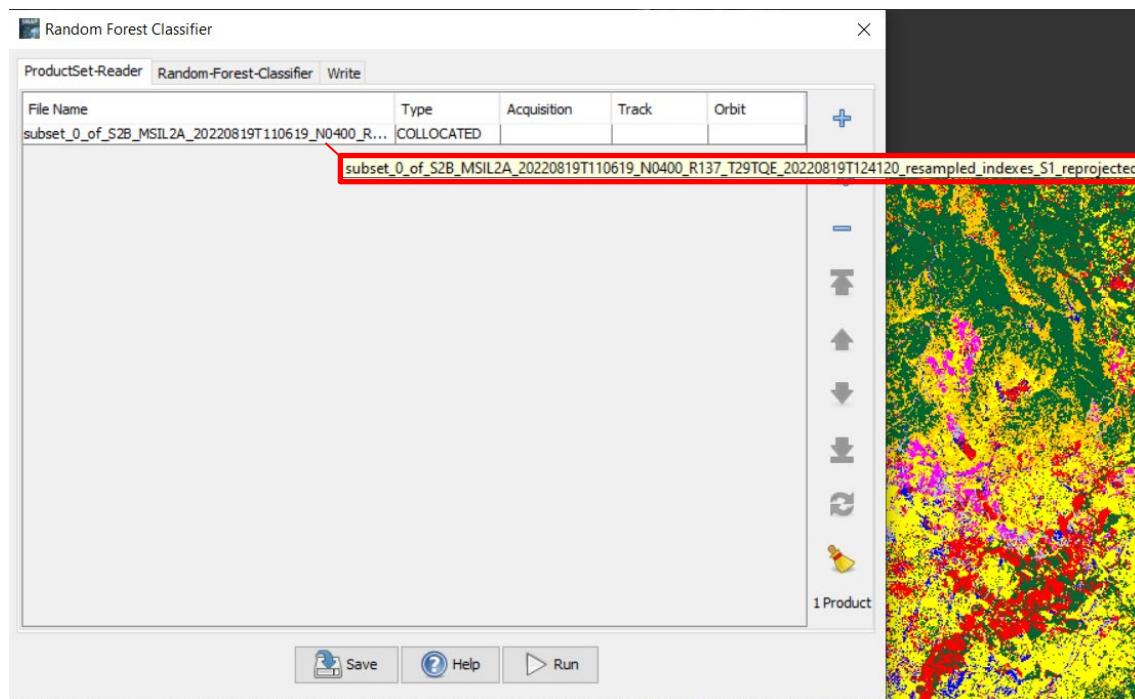
Para realizar la clasificación, buscamos el algoritmo Random Forest entre los clasificadores supervisados.



Añadiremos todas las imágenes que tengamos abiertas en la sesión.

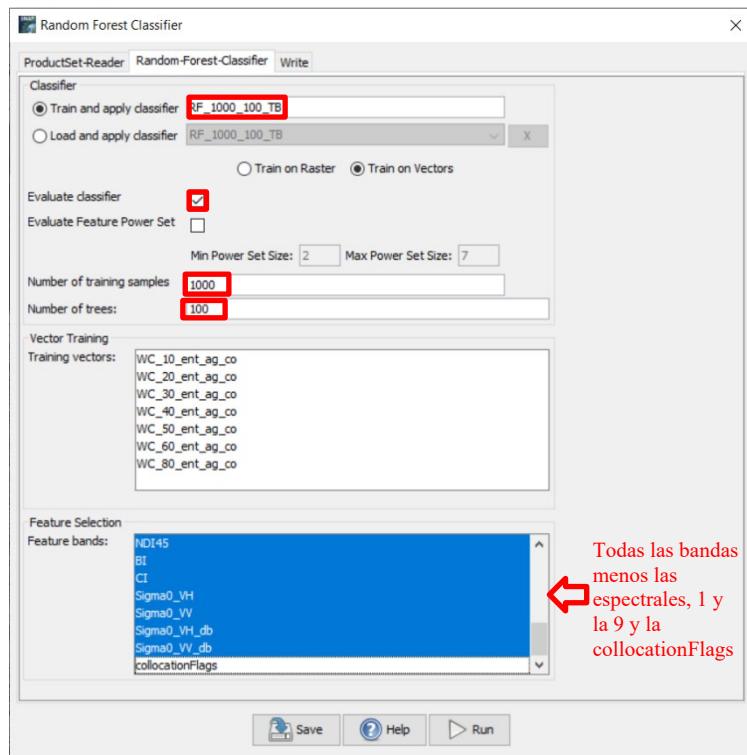


Si hubiera más de una imagen, las seleccionaríamos y le daríamos al botón para quedarnos sólo con la imagen a clasificar.

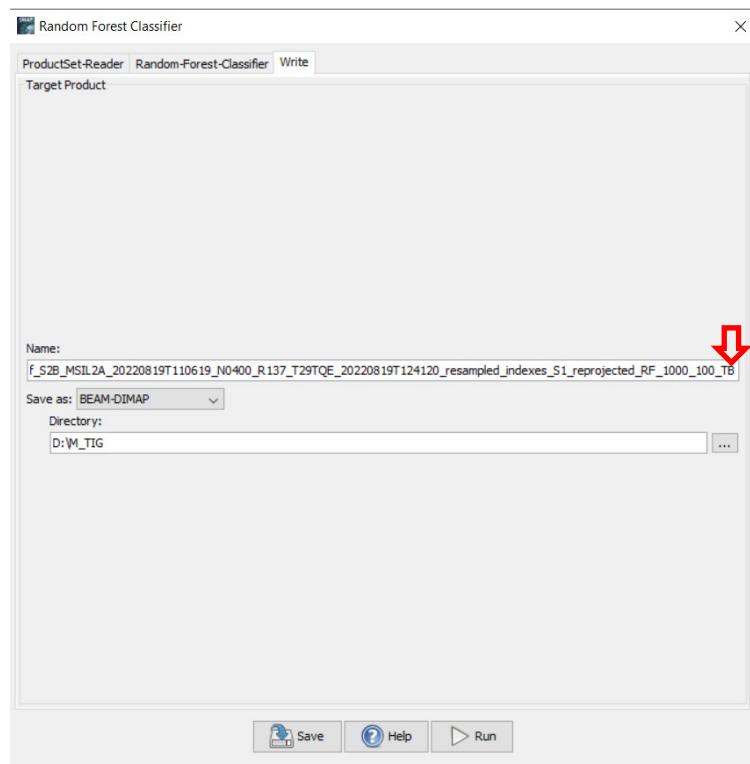


En la siguiente solapa, configuraremos los parámetros de clasificación.

Podemos guardar los parámetros de la configuración de la clasificación con un nombre por si posteriormente quisiéramos utilizarlos de nuevo. Además, es conveniente guardarlos porque el nombre que la configuración tenga, será el que adquiera el archivo de texto con la validación cruzada que posteriormente analizaremos. En nuestro caso vamos a llamar a la clasificación RF_1000_100_TB (Clasificación RF, con 1000 píxeles de entrenamiento, 100 árboles y todas las bandas de la imagen, menos la 1 y la 9 y *collocation flags*)



Y en la última solapa configuraremos el archivo de salida:



Una vez presionado el botón run, comenzará con el entrenamiento, y a mitad de proceso se nos abrirá una ventana con un fichero de texto en el que se muestra la validación cruzada, lo cerraremos y evaluaremos posteriormente.

```

Computing raster data...
Extracting data... 100% Cancel

Save Help Run

RF_1000_100_TB Bloc de notas
Archivo Edición Formato Ver Aplicación Randomforest classifier RF_1000_100_TB

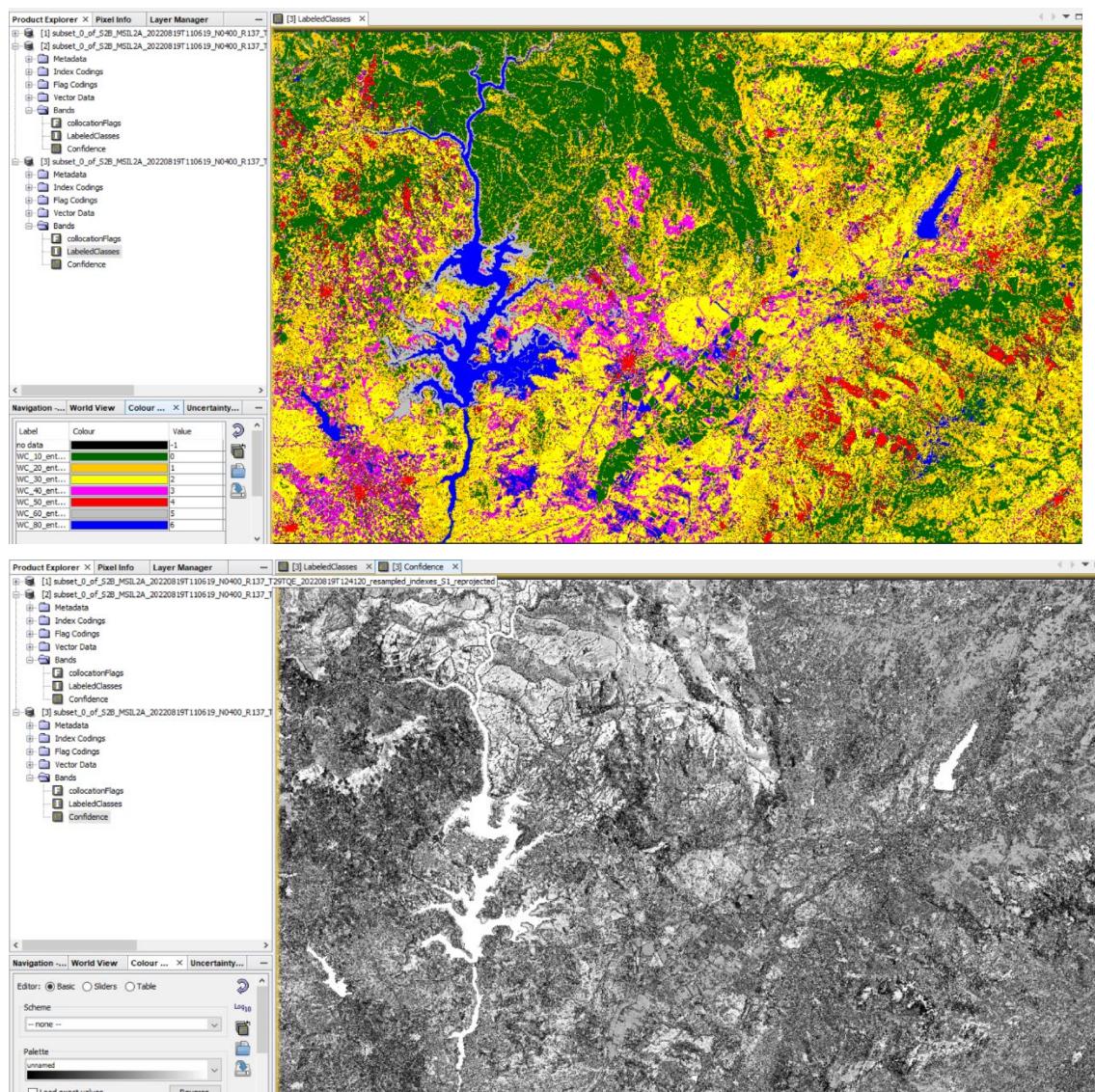
Cross Validation
Number of classes = 7
class 0.0: WC_10_ent_ag_co
accuracy = 0.9659 precision = 0.8911 correlation = 0.8595 errorRate = 0.0350
TruePositives = 122.0000 FalsePositives = 14.0000 TrueNegatives = 843.0000 FalseNegatives = 21.0000
class 1.0: WC_20_ent_ag_co
accuracy = 0.9454 precision = 0.7654 correlation = 0.7293 errorRate = 0.0570
TruePositives = 124.0000 FalsePositives = 38.0000 TrueNegatives = 819.0000 FalseNegatives = 19.0000
class 2.0: WC_30_ent_ag_co
accuracy = 0.9503 precision = 0.8630 correlation = 0.8559 errorRate = 0.0370
TruePositives = 126.0000 FalsePositives = 26.0000 TrueNegatives = 837.0000 FalseNegatives = 17.0000
class 3.0: WC_40_ent_ag_co
accuracy = 0.9780 precision = 0.9116 correlation = 0.9133 errorRate = 0.0220
TruePositives = 134.0000 FalsePositives = 13.0000 TrueNegatives = 844.0000 FalseNegatives = 0.0000
class 4.0: WC_50_ent_ag_co
accuracy = 0.9488 precision = 0.9488 correlation = 0.9187 errorRate = 0.0200
TruePositives = 130.0000 FalsePositives = 7.0000 TrueNegatives = 850.0000 FalseNegatives = 13.0000
class 5.0: WC_60_ent_ag_co
accuracy = 0.9728 precision = 0.8954 correlation = 0.8890 errorRate = 0.0280
TruePositives = 132.0000 FalsePositives = 15.0000 TrueNegatives = 843.0000 FalseNegatives = 11.0000
class 6.0: WC_80_ent_ag_co
accuracy = 0.9810 precision = 0.9844 correlation = 0.9214 errorRate = 0.0190
TruePositives = 126.0000 FalsePositives = 2.0000 TrueNegatives = 855.0000 FalseNegatives = 17.0000

Using Testing dataset, % correct predictions = 89.1000
Total samples = 2000
RMSE = 0.7395944834804239
Bias = -0.067000000000000017

Distribution:
class 0.0: WC_10_ent_ag_co      286    (14.30000)
class 1.0: WC_20_ent_ag_co      286    (14.30000)
class 2.0: WC_30_ent_ag_co      286    (14.30000)
class 3.0: WC_40_ent_ag_co      286    (14.30000)
class 4.0: WC_50_ent_ag_co      286    (14.30000)
class 5.0: WC_60_ent_ag_co      284    (14.20000)
class 6.0: WC_80_ent_ag_co      284    (14.20000)

```

Cuando el proceso termine, podremos cerrar la ventana y visualizar el resultado:

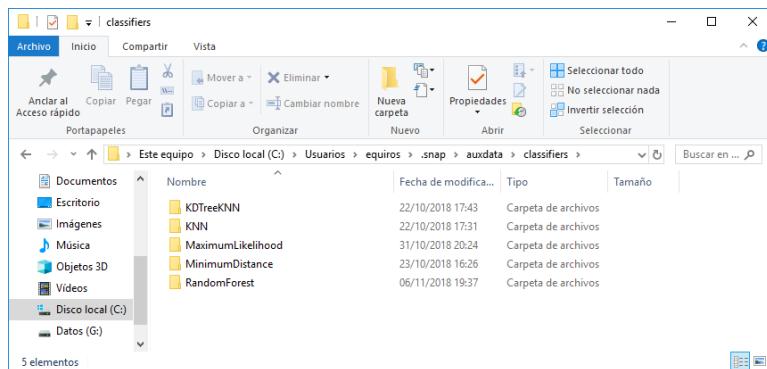


Así observaremos que lo que está por debajo del 0.5, no lo representa como clasificado (aunque sí lo está), ya que estima que no es suficiente como para asegurar que pertenece a una clase en concreto (no hay al menos un 50% de los árboles que hayan coincidido en la clasificación del píxel). Por tanto, esto podría ser una nueva pista para variar las áreas de entrenamiento o clases de nuestra clasificación.

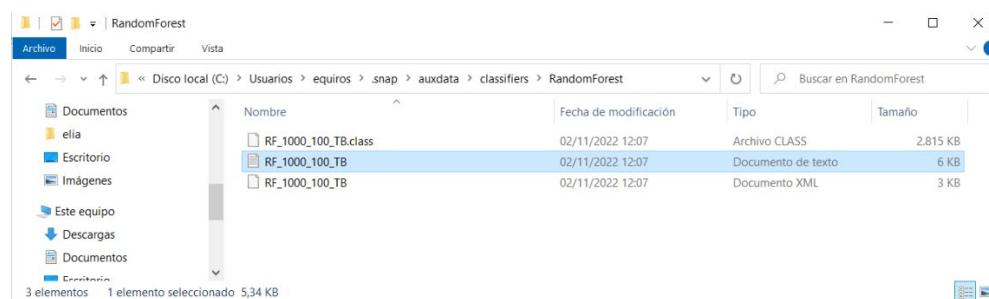
2.2.5.1 Validación cruzada

Finalmente, procederemos a revisar la validación cruzada que ha realizado el clasificador y que nos dará una idea de cómo ha funcionado el clasificador con respecto a las muestras de entrenamiento.

En la carpeta “.snap” que se crea en el usuario del equipo hay una carpeta en la que se guardan todas las evaluaciones por validación cruzada de las clasificaciones.



En nuestro caso, dentro de la carpeta de random forest podremos observar que se han guardado tres archivos con el nombre de la clasificación. Dos son de la configuración y el tercero (archivo de texto) es el de la validación cruzada. Este archivo es el que nos apareció en una ventana emergente mientras hacíamos la clasificación.



Al abrirlo podremos observar toda la información de la validación:

```

RF_1000_100_TB: Bloc de notas
Archivo Edición Formato Ver Ayuda
RandomForest classifier RF_1000_100_TB

Cross Validation
Number of classes = 7
class 0:0: WC_10_ent_ag_co
accuracy = 0.9550 precision = 0.8971 correlation = 0.8595 errorRate = 0.0350
TruePositives = 122.0000 FalsePositives = 14.0000 TrueNegatives = 843.0000 FalseNegatives = 21.0000
class 1:0: WC_20_ent_ag_co
accuracy = 0.9438 precision = 0.7654 correlation = 0.7929 errorRate = 0.0570
TruePositives = 124.0000 FalsePositives = 38.0000 TrueNegatives = 819.0000 FalseNegatives = 19.0000
class 2:0: WC_30_ent_ag_co
accuracy = 0.9520 precision = 0.8530 correlation = 0.8599 errorRate = 0.0370
TruePositives = 126.0000 FalsePositives = 28.0000 TrueNegatives = 837.0000 FalseNegatives = 17.0000
class 3:0: WC_40_ent_ag_co
accuracy = 0.9798 precision = 0.9116 correlation = 0.9133 errorRate = 0.0220
TruePositives = 134.0000 FalsePositives = 13.0000 TrueNegatives = 844.0000 FalseNegatives = 9.0000
class 4:0: WC_50_ent_ag_co
accuracy = 0.9600 precision = 0.9480 correlation = 0.9187 errorRate = 0.0100
TruePositives = 138.0000 FalsePositives = 7.0000 TrueNegatives = 850.0000 FalseNegatives = 13.0000
class 5:0: WC_60_ent_ag_co
accuracy = 0.9720 precision = 0.8958 correlation = 0.8899 errorRate = 0.0280
TruePositives = 129.0000 FalsePositives = 15.0000 TrueNegatives = 843.0000 FalseNegatives = 13.0000
class 6:0: WC_88_ent_ag_co
accuracy = 0.9810 precision = 0.9844 correlation = 0.9214 errorRate = 0.0190
TruePositives = 125.0000 FalsePositives = 2.0000 TrueNegatives = 855.0000 FalseNegatives = 17.0000

Using Testing dataset, % correct predictions = 89.1000
Total samples = 2000
RMSE = 0.735944434846239
Bias = -0.00700000000000017

Distribution:
class 0:0: WC_10_ent_ag_co      286 (14.3000%)
class 1:0: WC_20_ent_ag_co      286 (14.3000%)
class 2:0: WC_30_ent_ag_co      286 (14.3000%)
class 3:0: WC_40_ent_ag_co      286 (14.3000%)
class 4:0: WC_50_ent_ag_co      286 (14.3000%)
class 5:0: WC_60_ent_ag_co      284 (14.2000%)
class 6:0: WC_88_ent_ag_co      286 (14.3000%)
} Exactitud general

Testing feature importance:
Each feature is present in 3 times and the % correct predictions are averaged
The Importance score is the original % correct prediction - average
rank 1 feature 21 : Sigma0_VH_db      score: tp=0.0310 accuracy=0.0089 precision=0.0298 correlation=0.0351 errorRate=0.0089 cost=0.0331 GainRatio = 0.2008
rank 2 feature 9 : B1L      score: tp=0.0287 accuracy=0.0059 precision=0.0192 correlation=0.0232 errorRate=0.0059 cost=0.0217 GainRatio = 0.3795
rank 3 feature 9 : SAVI      score: tp=0.0198 accuracy=0.0054 precision=0.0154 correlation=0.0176 errorRate=0.0054 cost=0.0199 GainRatio = 0.3545
rank 4 feature 22 : Sigma0_VV_db      score: tp=0.0190 accuracy=0.0051 precision=0.0151 correlation=0.0176 errorRate=0.0051 cost=0.0196 GainRatio = 0.3498
rank 5 feature 12 : B2      score: tp=0.0160 accuracy=0.0046 precision=0.0156 correlation=0.0182 errorRate=0.0046 cost=0.0160 GainRatio = 0.4313
rank 6 feature 15 : MSAV12      score: tp=0.0093 accuracy=0.0027 precision=0.0088 correlation=0.0185 errorRate=0.0027 cost=0.0093 GainRatio = 0.3444
rank 7 feature 8 : BBA      score: tp=0.0065 accuracy=0.0017 precision=0.0057 correlation=0.0068 errorRate=0.0017 cost=0.0062 GainRatio = 0.3655
rank 8 feature 7 : BB      score: tp=0.0047 accuracy=0.0013 precision=0.0045 correlation=0.0053 errorRate=0.0013 cost=0.0049 GainRatio = 0.3024
rank 9 feature 4 : BS      score: tp=0.0043 accuracy=0.0012 precision=0.0042 correlation=0.0051 errorRate=0.0012 cost=0.0046 GainRatio = 0.4615
rank 10 feature 19 : Sigma0_VH      score: tp=0.0040 accuracy=0.0009 precision=0.0038 correlation=0.0050 errorRate=0.0009 cost=0.0040 GainRatio = 0.3244
rank 11 feature 19 : Sigma0_VH      score: tp=0.0017 accuracy=0.0005 precision=0.0015 correlation=0.0019 errorRate=0.0005 cost=0.0017 GainRatio = 0.3335
rank 12 feature 2 : B3      score: tp=0.0010 accuracy=0.0003 precision=0.0009 correlation=0.0011 errorRate=0.0003 cost=0.0019 GainRatio = 0.5083
rank 13 feature 14 : SZREP      score: tp=0.0000 accuracy=0.0000 precision=0.0000 correlation=0.0000 errorRate=0.0000 cost=0.0000 GainRatio = 0.1786
rank 14 feature 13 : GNDVI      score: tp=0.0003 accuracy=0.0001 precision=0.0003 correlation=0.0004 errorRate=0.0001 cost=0.0004 GainRatio = 0.3112
} Porcentaje de cada clase utilizado en la clasificación

} Ranking de importancia de cada banda en la clasificación

```

Ejercicio propuesto 2:

Variar los parámetros/bandas/clases de la clasificación para lograr obtener una clasificación más completa y con una buena exactitud en la validación cruzada.

3. Bibliografía

- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- European Spatial Agency. SNAP software help (2018). <http://step.esa.int/main/toolboxes/snap/>.
- Fragoso-Campón, L., Quirós, E., Mora, J., Gutiérrez, J. A., & Durán-Barroso, P. (2018). Accuracy Enhancement for Land Cover Classification Using LiDAR and Multitemporal Sentinel 2 Images in a Forested Watershed. *Multidisciplinary Digital Publishing Institute Proceedings*, 2(20), 1280.
- Quirós Rosado, E. (2009). Clasificación de imágenes multiespectrales ASTER mediante funciones adaptativas.
- RUS Copernicus Training (2018). Crop mapping with Sentinel-2. <https://rus-training.eu/training/crop-mapping-with-sentinel-2>.
- WorldCover 2021 v200: Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoíno, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. <https://doi.org/10.5281/zenodo.7254221>