Cuad. Soc. Esp. Cienc. For. 49(2): 161-184 (2023)
Doi: https://doi.org/10.31167/csecfv5i49.19935

Cuadernos de la Sociedad Española de Ciencias Forestales

Acceso abierto disponible en http://secforestales.org/publicaciones/index.php/cuadernossecf/index

Accésit SECF 2022 en el apartado de Trabajo Fin de Máster Oficial

Clasificación automática de coberturas forestales a partir de imágenes Sentinel-2 y datos auxiliares mediante Google Earth Engine

Automatic classification of forest cover using Sentinel-2 imagery and auxiliary data through Google Earth Engine

Teijido-Murias, I.[1*]; Barrio-Anta, M.[1]; López-Sánchez, C.A.[1]

[1] *Grupo de Investigación SmartForest, Departamento de Biología de Organismos y Sistemas, Escuela Politécnica de Mieres, Universidad de Oviedo. 33600 Mieres, España.*

[*] Autor de correspondencia: teijidoiyan@uniovi.es

Resumen

El objetivo de este trabajo fue realizar una clasificación automática de coberturas forestales con una resolución espacial de 10 m/pixel^{-1} en Asturias (Norte de España). Esta región tiene una vocación eminentemente forestal con una gran superficie de plantaciones forestales de *Pinus pinaster*, *Pinus radiata* y *Eucalyptus globulus*, cuyas cortas representan el 91% de la madera aprovechada anualmente en la región. La clasificación se basó en una serie de datos multitemporales del sensor Sentinel^{-2} y en datos auxiliares climáticos y topográficos. Como fuentes de datos se utilizaron: parcelas de campo (Inventario Forestal Nacional), clasificación de coberturas terrestres (Corine Land Cover), imágenes de satélite (constelación Sentinel^{-2}) y varias fuentes de variables auxiliares. El código fue implementado en Google Earth Engine® (GEE) y para la clasificación se usó el algoritmo Random Forest. Se realizó una división aleatoria de la base de datos, destinando un 70% para el entrenamiento de los modelos y un 30% para su validación. Finalmente se clasificaron las coberturas forestales del territorio en 30 clases, obteniéndose un valor del estadístico Kappa de 0,5723. Al analizar la matriz de confusión, se observó que los fallos de clasificación eran principalmente consecuencia de la similitud de las firmas espectrales de las coberturas analizadas. En los próximos años, con la mayor diversidad y mejora de los datos capturados por sensores remotos, el perfeccionamiento de plataformas como GEE, y el desarrollo de nuevos algoritmos, permitirán obtener clasificaciones de cobertura automáticas más precisas y probablemente se convertirán en herramientas indispensables.

Palabras clave: *Sensores remotos, Inventario Forestal Nacional, Google Earth Engine, Sentinel-2, Random Forest.*

Abstract

The aim of this work was to carry out an automatic classification of forest covers with a spatial resolution of 10 m pixel^{-1}, in the region of Asturias (North Spain). This region has an eminently forest vocation and present an important area constituted by forest plantations of *Pinus pinaster*, *Pinus radiata* and *Eucalyptus globulus*, which represent 91% of the annual timber harvested in the region; so, the classification was mainly emphasized on these species. The classification was based on a multitemporal series data from the Sentinel-2 sensor and auxiliary data (climate and topography). We use the following data sources: field plots (Spanish National Forest Inventory); land cover classification of the Corine Land Cover project; satellite images (Sentinel^{-2} constellation); and different sources of auxiliary variables. The code was implemented in Google Earth Engine®(GEE). We use the Random Forest algorithm, to obtain the training-validation datasets was made a random division (70-30%) of the global database. The forest cover level classifies the territory into 30 large use classes. The Kappa value obtained for this level was 0.5723. Analyzing the confusion matrix, it was observed that confusion between classes was a consequence of the similarity of the analyzed coverage spectral signatures. In the coming years, with the greater diversity and improvement of data captured by remote sensors, the development of platforms such as GEE, and the acquisition of new algorithms will allow obtaining coverage classifications that will significantly improve their accuracy and will likely become an indispensable tool.

Keywords: *Remote Sensing, National Forest Inventory, Google Earth Engine, Sentinel-2, Random Forest.*

1. Introducción

La planificación del territorio se basa en un buen conocimiento de la extensión y distribución los distintos tipos de coberturas o usos del mismo. En el sector forestal, el conocimiento de las superficies ocupadas por las distintas formaciones es esencial para determinar los recursos disponibles y para su gestión.

Para analizar estas superficies se puede recurrir a dos enfoques diferentes:

1. El tradicional basado en trabajo de campo.

2. El basado en nuevas tecnologías.

Este segundo enfoque, en comparación con el primero, reduce el tiempo de dedicación y permite analizar áreas extensas y/o inaccesibles y ha recibido un gran impulso desde 2017 con la disponibilidad de dos satélites de la constelación Sentinel-2, el 2A (puesto en órbita en 2015) y el 2B (en 2017) con trece bandas espectrales de alta resolución espacial (10 m px^{-1}) que, combinados, proporcionan una resolución temporal de 5 días.

A partir de datos satelitales y del trabajo de campo, se pueden llevar a cabo clasificaciones de coberturas orientadas a objetos o a píxeles, pudiendo ser estas supervisadas o no supervisadas. Estos análisis son fácilmente integrables dentro de herramientas de computación en la nube como Google Earth Engine (GEE), permitiendo la generación a gran escala de productos espacio-temporales de gran interés práctico.

Actualmente, el único método de actualización de superficies de masas forestales de la región es el Mapa Forestal de España (MFE), actualizado decenalmente. Sin embargo, existen importantes razones que justifican el desarrollo de un método automático de clasificación:

1. La alta tasa de cortas en algunas especies.

2. La necesaria actualización anual de los recursos disponibles.

3. La existencia de fuentes de datos públicas robustas y fiables que lo permiten.

El objetivo de este trabajo fue obtener una clasificación supervisada a nivel píxel de las coberturas forestales del Principado de Asturias con una resolución espacial de 10 m $pixel^{-1}$, utilizando imágenes de las cuatro estaciones del año de la constelación de satélites Sentinel-2.

2. Material y Métodos

2.1. Área de estudio y especies de interés

La Comunidad Autónoma del Principado de Asturias se encuentra en el norte de

España y tiene una extensión de 10 603,57 km^2. El 72% del territorio está clasificada como terreno forestal, siendo un 60% de este arbolado. Aunque dominan las formaciones forestales autóctonas, 70%, el 30% restante está constituido por plantaciones forestales, principalmente de *Pinus pinaster*, *Pinus radiata* y *Eucalyptus globulus* (Ministerio de Agricultura, 2012), cuyas cortas supusieron el 91% de la madera cortada anualmente (916.000 m^3 año^{-1} de promedio en los últimos 15 años) (Ministerio de Agricultura, 2019).

Se realizó una clasificación de coberturas forestales completa, atendiendo a las principales especies arbóreas (hasta 30 especies o grupos de especies) (*tablas 1*, *2*, *3* y *4*).

2.2. Fuentes de datos

Se han utilizado cuatro fuentes de datos diferentes:

1. Parcelas de campo del Inventario Forestal Nacional (IFN).

2. Clasificación de coberturas del proyecto Corine Land Cover de la Agencia Espacial Europea (ESA).

3. Imágenes de satélite (Sentinel-2 de la ESA).

4. Diferentes clases de variables auxiliares.

2.2.1 Parcelas de campo del Inventario Forestal Nacional

Se usaron los datos de las parcelas de campo del Cuarto Inventario Forestal Nacional (IFN4) y su actualización de 2018 (IFN4.5) para las tres especies más productivas y sus correspondientes Mapas Forestales de España (MFE4 y MFE4.5). Estas parcelas se localizan en nodos de una cuadrícula de 1km x 1km que cubre todo el país. En la zona de estudio se localizaron 1109 parcelas sobre terreno catalogado como forestal *(Tabla 1)*.

2.2.2. Clasificación jerárquica de coberturas terrestres

Para la jerarquización de las parcelas en clases de uso del suelo se usaron las clases del proyecto Corine Land Cover de 2018 (CLC, 2018) para Asturias. La clasificación CLC2018 define tres niveles jerárquicos de coberturas (L1, L2 y L3) a la que se le añadió un nivel más (L4) para discriminar las especies forestales. Se usaron al menos 30 parcelas para una de las clases del nivel L4, que fueron fotointerpretadas mediante el uso de ortofotografías del Plan Nacional de Ortofotografía Aérea (PNOA). El total de parcelas utilizadas ha sido de 1648, cuya distribución se observa en la *tabla 4*.

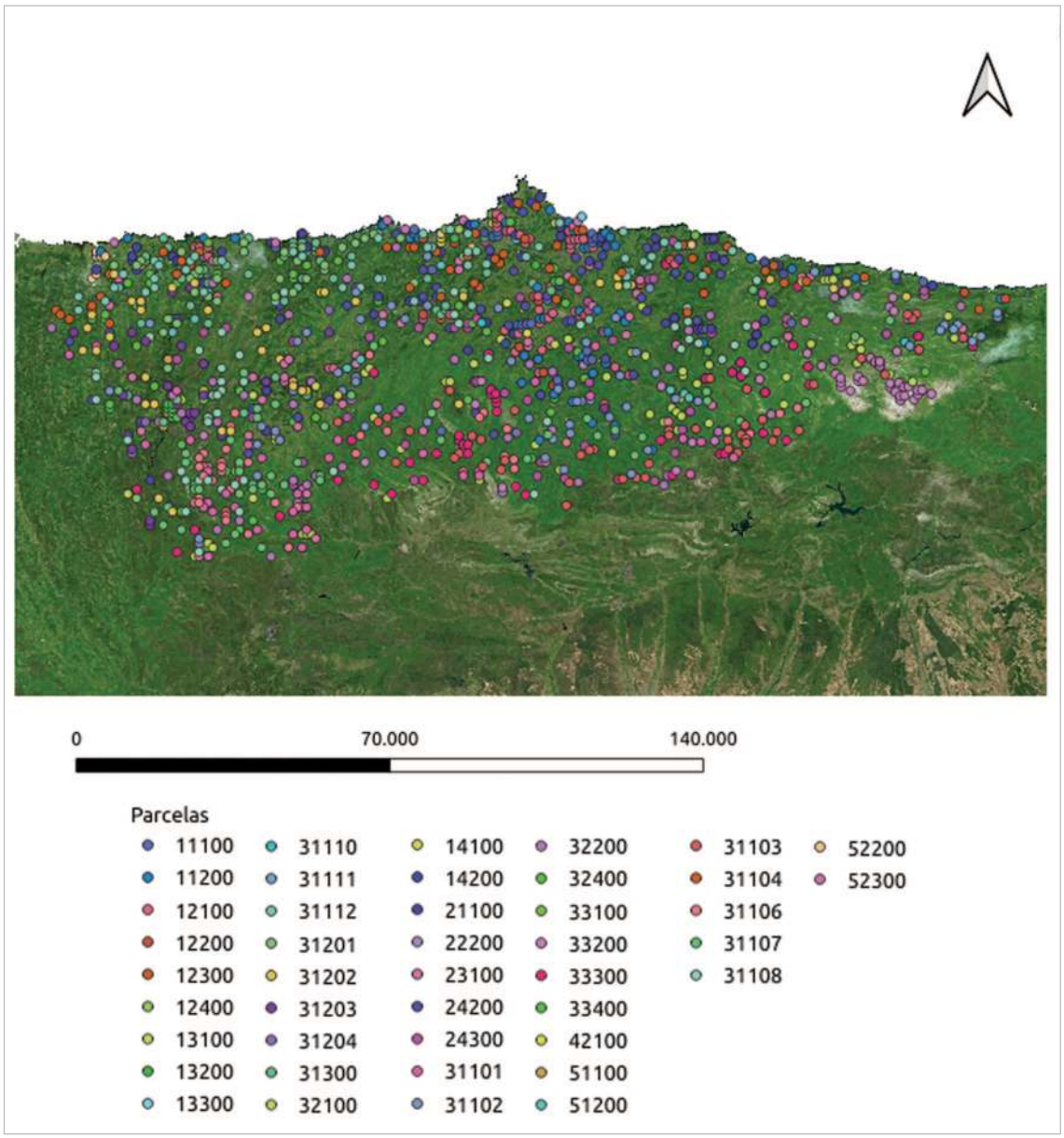

Figura 1. Parcelas de campo utilizadas para el ajuste del modelo con las clases L4 (SRC: 32629).

Tabla 1. Clasificación nivel L1, incluye el número de parcelas utilizadas y el nombre de la clase usado en Google Earth Engine (GEE).

L1	Clase	Nº parcelas	Denominación clase
1	Superficies Artificiales	227	0
2	Zonas agrícolas	269	1
3	Zonas forestales	1109	2
4	Zonas húmedas	3	3
5	Superficies de agua	40	4

Tabla 2. Clasificación nivel L2, incluye el número de parcelas utilizadas y el nombre de la clase usado en GEE.

L2	Clase	Nº parcelas	Denominación clase
11	Zonas Urbanas	93	0
12	Zonas industriales, comerciales y de transportes	82	1
13	Zonas de extracción minera, vertederos y de construcción	37	2
14	Zonas verdes artificiales, no agrícolas	15	3
21	Tierras de labor	63	4
22	Cultivos permanentes	2	5
23	Prados y praderas	75	6
24	Zonas agrícolas heterogéneas	129	7
31	Bosques	732	8
32	Espacios de vegetación arbustiva y/o herbácea	214	9
33	Espacios abiertos con poca o sin vegetación	163	10
41	Zonas húmedas continentales	0	–
42	Zonas húmedas litorales	3	11
51	Aguas continentales	20	12
52	Aguas marinas	20	13

Tabla 3. Clasificación nivel L3, incluye el número de parcelas utilizadas y el nombre de la clase usado en GEE.

L3	Clase	Nº parcelas	Denominación clase
110	Zonas Urbanas	93	0
120	Zonas industriales, comerciales y de transportes	82	1
130	Zonas de extracción minera, vertederos y de construcción	37	2
140	Zonas verdes artificiales, no agrícolas	15	3
210	Tierras de labor	63	4
220	Cultivos permanentes	2	5
230	Prados y praderas	75	6
240	Zonas agrícolas heterogéneas	129	7
311	Bosques de frondosas	514	12
312	Bosques de coníferas	157	13
313	Bosque mixto	60	14
321	Pastizales naturales	83	15
322	Landas y matorrales mesófilos	73	16
324	Matorral boscoso de transición	58	17
330	Espacios abiertos con poca o sin vegetación	163	8
410	Zonas húmedas continentales	0	–
420	Zonas húmedas litorales	3	9
510	Aguas continentales	10	10
520	Aguas marinas	20	11

Tabla 4. Clasificación nivel L4, incluye el número de parcelas utilizadas y el nombre de la clase usado en GEE.

L4	Clase	Nº parcelas	Denominación clase
11000	Zonas Urbanas	93	0
12000	Zonas industriales, comerciales y de transportes	82	1
13000	Zonas de extracción minera, vertederos y de construcción	37	2
14000	Zonas verdes artificiales, no agrícolas	15	3
21000	Tierras de labor	63	4
22000	Cultivos permanentes	2	5
23000	Prados y praderas	75	6
24000	Zonas agrícolas heterogéneas	129	7
31101	*Betula alba*	48	8
31102	*Castanea sativa*	67	9
31103	*Fagus sylvatica*	68	10
31104	*Eucalyptus globulus*	58	11
31105	*Eucalyptus nitens*	–	
31106	*Quercus petraea*	44	12
31107	*Quercus pyrenaica*	66	13
31108	*Quercus robur*	80	14
31109	*Quercus suber*	–	–
31110	*Populus spp*	3	15
31111	*Quercus ilex*	22	16
31112	Repoblación de *Eucalyptus spp.*	58	17
31201	*Pinus pinaster*	52	18
31202	*Pinus radiata*	68	19
31203	*Pinus sylvestris*	30	20
31204	*Pseudotsuga menziesii*	7	21
31301	Mezcla de especies	60	22
32100	Pastizales naturales	83	23
32200	Landas y matorrales mesófilos	73	24
32400	Matorral boscoso de transición	58	25
33400	Espacios abiertos con poca o sin vegetación	163	26
41000	Zonas húmedas continentales	0	–
42000	Zonas húmedas litorales	3	27
51000	Aguas continentales	20	28
52000	Aguas marinas	20	29
	TOTAL	1647	

3.2.3. Imágenes de satélite

Las imágenes Sentinel-2 usadas se obtuvieron del repositorio de la plataforma GEE. El nivel de procesamiento utilizado fue el 2A (valores de reflectancia en superficie, SR) (Baret, 2010), estando calibradas radiométricamente, corregidas geométricamente y representando su valor la reflectancia sobre la superficie terrestre (Rumora *et al.*, 2020). De acuerdo con Caparros-Santiago et al. (2021), para capturar el estado fenológico de las diferentes coberturas vegetales, se emplearon 4 conjuntos de imágenes multiespectrales (4 periodos de 3 meses cada uno) de la misión Sentinel-2 del año 2018 *(Tabla 5)*.

Tabla 5. Datos colección de imágenes SENTINEL

COPERNICUS/S2_SR			
Nombre_script	Fecha	% Nubes	Estadistico aplicado
sentinel_efm	2018/01/01 - 2018/03/31	5	Mediana
sentinel_amj	2018/04/01 - 2018/06/30	5	Mediana
sentinel_jas	2018/07/01 - 2018/09/30	5	Mediana
sentinel_ond	2018/10/01 - 2018/12/31	5	Mediana

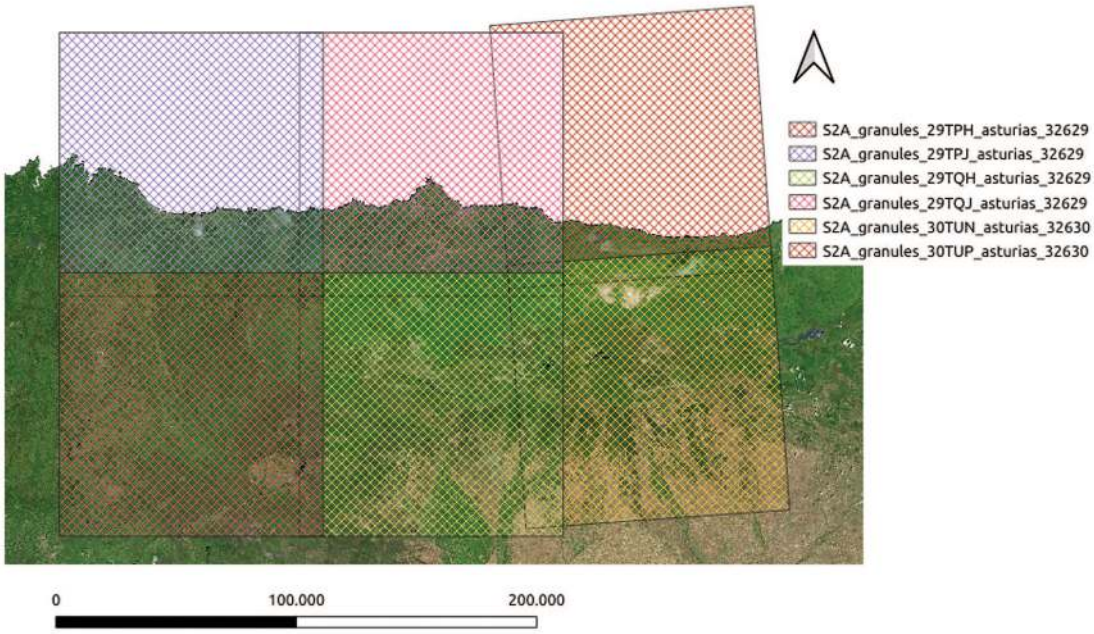

Figura 2. Superposición de los seis *granules* utilizados (SRC: 32629).

3.2.4. Variables predictoras usadas

Además de los valores de reflectancia, para el ajuste de los modelos también se utilizaron variables biofísicas, índices espectrales y variables auxiliares sobre climatología, topografía e hidrología a nivel píxel *(Tabla 6)*.

Tabla 6. Variables auxiliares utilizadas.

Topografía	Climatología	Hidrología	Variables de textura	Variables Biofísicas
Altitude	mnri2ge	dvca	ASM	LAI
Aspect	mtri2ge		Contrast	FAPAR
Curvature	mxri2ge		Dissimilarit	FCOVER
Heatload	plrs1ge		y Energy	
Plancurvature	radge1		Entropy	
Profilecurvature			GLCMCorrelation	
Slope			GLCMMean	
Slopeposition			GLCMVariance	
Terrshpind2pix			Homogeneity	
Wetness			Max	

Siendo Altitude, altitud; Aspect, orientación; Curvature, curvatura; Heatload, índice de carga de calor; Plancurvature, curvatura del plano; Profilecurvature, curvatura del perfil; Slope, pendiente; Slopeposition, posición de la pendiente; Terrshpind2pix, índice de forma del terreno; Wetness, índice de humedad; plrs1ge, precipitación mensual acumulada; mtri2ge, temperatura media diaria al mes; mxri2ge, temperaturas máximas medias diarias al mes; mnri2ge, temperaturas mínimas medias diarias al mes; radge1, radiación mensual acumulada por mes; dvca, Distancia vertical a los cuerpos de agua; ASM, segundo momento angular; Contrast, contraste; Dissimilarity, disimilitud; Energy, Entropy, entropía; GLCMCorrelation, correlación de la matriz de coocurrencia de nivel de gris (GLCM); GLCMMean, media de GLCM; GLCMVariance, varianza de la GLCM; Homogeneity, homogeneidad; Max, máximo; LAI, índice de área foliar; FAPAR: fracción de radiación activa fotosintéticamente absorbida; FCOVER, fracción de la cubierta vegetal.

3.3. Análisis de datos

En la *Figura 3* se muestre el flujo de procesos seguido en este estudio para la toma y el análisis de datos.

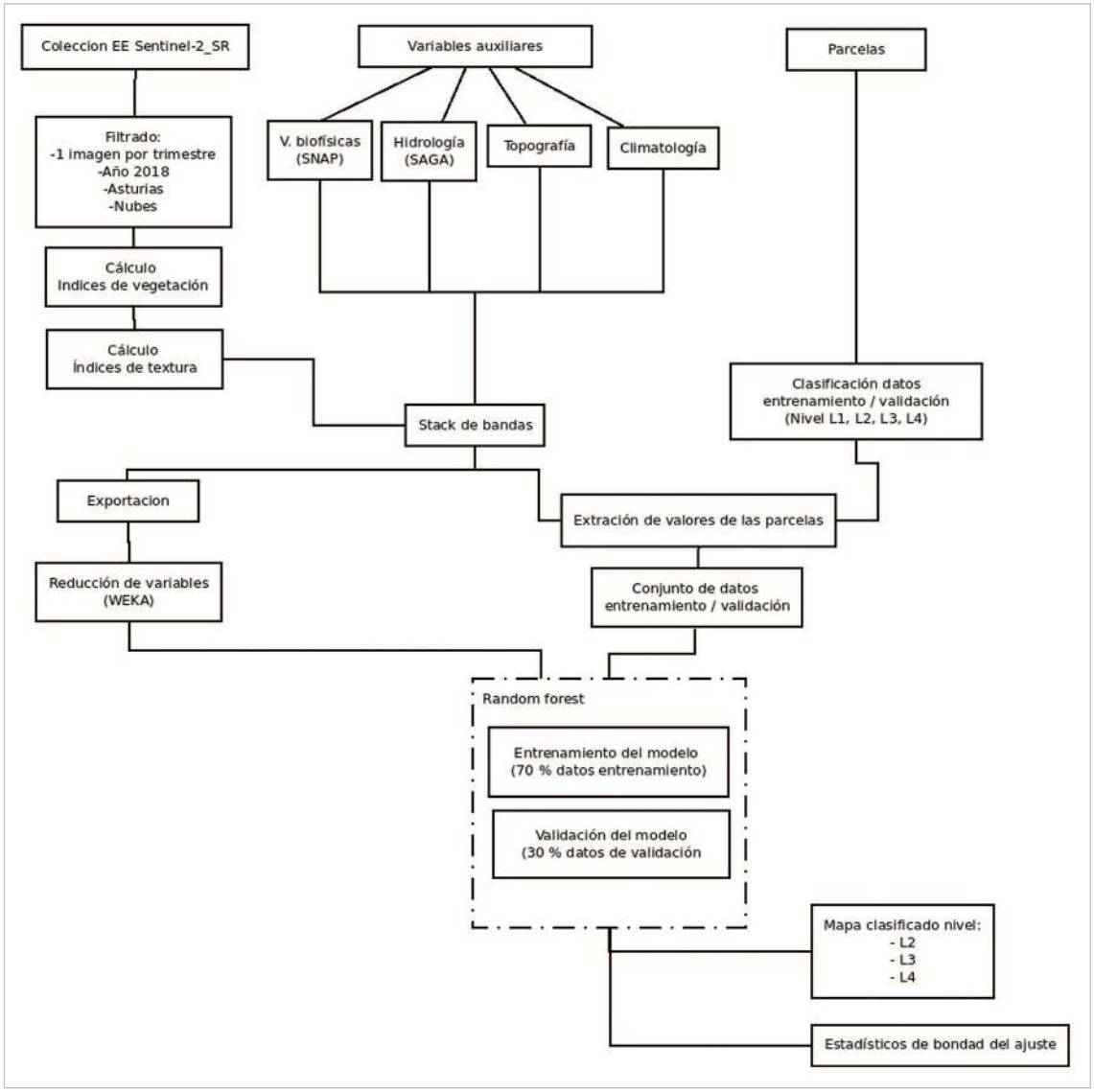

Figura 3. Flujo de procesos para la clasificación de coberturas forestales.

Previamente al ajuste de los modelos de clasificación, se realizó una reducción de variables mediante el método 'feature selection with wrapper' (*data dimensionality*) y el algoritmo *Random Forest* en el software WEKA®, para seleccionar los predictores del modelo final, se utilizó la medida de la importancia de la variable (VIM) normalizada. Para la clasificación se utilizó el algoritmo Random Forest (Breiman, 2001) con una división aleatoria de la base de datos (70 % entrenamiento y 30 % validación). Para evaluar la precisión de las predicciones se obtuvieron los siguientes estadísticos a partir de la matriz de confusión (Fielding y Bell, 1997): precisión global (OA), error de comisión, error emisión y el estadístico Kappa.

Tabla 7. Clasificación del estadístico Kappa (obtenida de Landis, J.R. y Koch, G.G., 1977).

Estadístico Kappa	Concordancia
< 0,00	Sin acuerdo
0,00 - 0,20	Insignificante
0,21 - 0,40	Mediano
0,41 - 0,60	Moderado
0,61 - 0,80	Sustancial
0,81 - 1,00	Casi perfecto

3. Resultados y Discusión

Teniendo en cuenta los valores del estadístico Kappa obtenidos tras los ajuste, estos se pueden considerar moderados según Landis (1977). El nivel L2 de la clasificación obtuvo un Kappa de 0,7; valor bastante aceptable, aunque se encuentra por debajo de los valores obtenidos por Qu *et al.* (2021) para una clasificación similar. Los estadísticos de bondad del ajuste se muestran en las *tablas 8* y *9* para el nivel L2 y en *tablas 10* y *11* para el nivel L3. Las *figuras 4* y *6* muestran la importancia de variables predictoras de cada clase. En las *figuras 5* y *7* se puede ver un mapa para cada nivel de clasificación.

Tabla 8. Ajustes del modelo RF para la Clasificación nivel L2.

Variable	Entrenamiento	Validación
Precisión	0,9991	0,7821
Kappa	0,9988	0,7005

Tabla 9. Estadísticos del modelo RF (validación) para la Clasificación nivel L2

Clase / Estadístico	P. usuario	P. productor	Error de Comisión	Error de Omisión
0	0,6800	0,6071	0,3200	0,3929
1	0,7272	0,8000	0,2728	0,2000
2	0,8571	0,6666	0,1429	0,3333
3	1,0000	0,2500	0,0000	0,7500
4	0,7500	0,5294	0,2500	0,4706
5	0,0000	0,0000	1,0000	1,0000
6	0,8333	0,5263	0,1667	0,4737
7	0,6944	0,5555	0,3056	0,4445
8	0,7538	0,9800	0,2462	0,0200
9	0,9714	0,5230	0,0286	0,4770
10	0,9302	0,8510	0,0698	0,1490
11	0,0000	0,0000	1,0000	1,0000
12	1,0000	0,8000	0,0000	0,2000
13	0,7272	1,0000	0,2728	0,0000

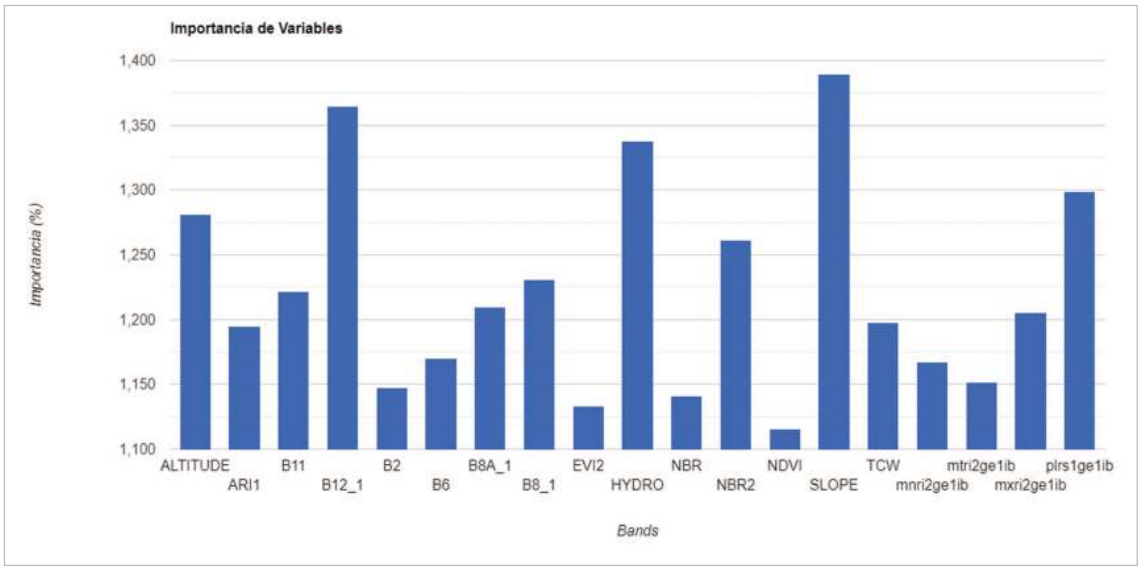

Figura 4. Importancia de las variables utilizadas para el modelo RF a nivel L2.

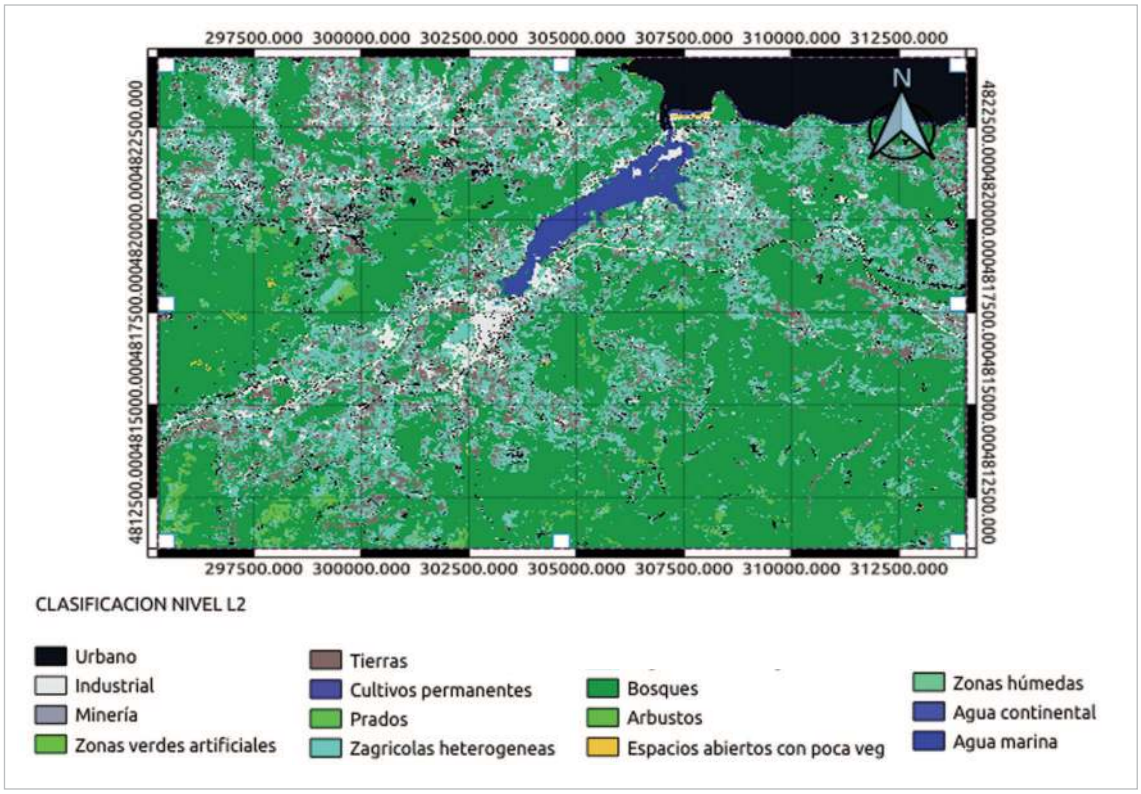

Figura 5. Vista del mapa a nivel L2 de coberturas para el concejo de Villaviciosa. EPSG: 32630.

Tabla 10. Ajustes del modelo RF para la Clasificación nivel L3

Variable	Entrenamiento	Validación
Precisión	1	0,7175
Kappa	1	0,6593

Tabla 11. Estadísticos del modelo RF (validación) para la Clasificación nivel L3

Clase / Estadístico	P. usuario	P. productor	Error de Comisión	Error de Omisión
0	0,6667	0,7200	0,3333	0,2800
1	0,6667	0,7200	0,3333	0,2800
2	0,8889	0,5333	0,1111	0,4667
3	1,0000	0,4000	0,0000	0,6000
4	0,6364	0,5385	0,3636	0,4615
5	1,0000	1,0000	0,0000	0,0000
6	0,9000	0,7826	0,1000	0,2174
7	0,4737	0,5294	0,5263	0,4706
8	0,7358	0,8125	0,2642	0,1875
9	0,0000	0,0000	1,0000	1,0000
10	1,0000	1,0000	0,0000	0,0000
11	0,8750	0,8750	0,1250	0,1250
12	0,7277	0,8634	0,2723	0,1366
13	0,7037	0,7917	0,2963	0,2083
14	0,5000	0,3000	0,5000	0,7000
15	0,8333	0,5556	0,1667	0,4444
16	0,7778	0,4667	0,2222	0,5333
17	0,7500	0,1579	0,2500	0,8421

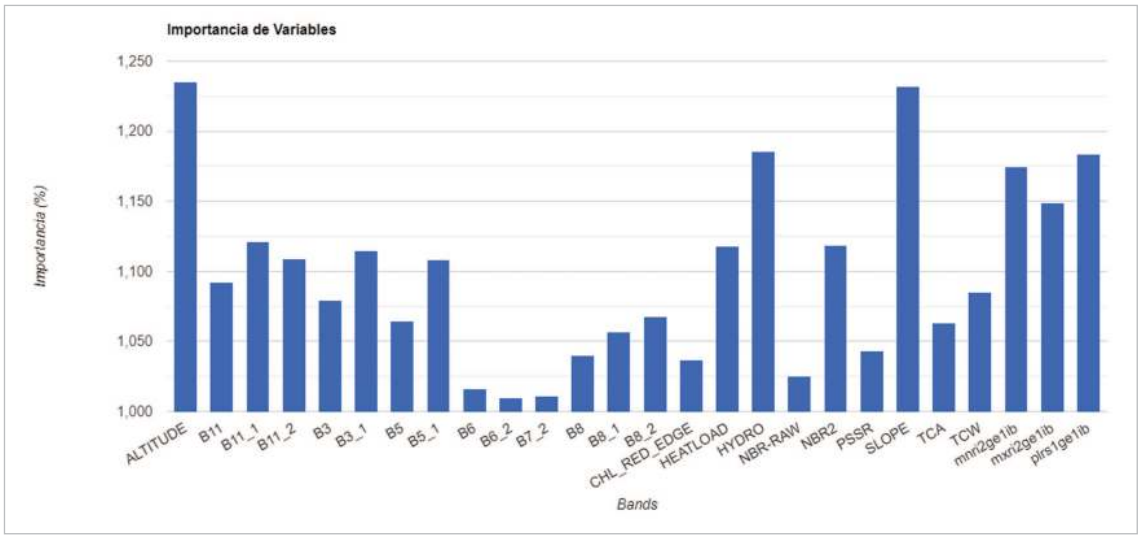

Figura 6. Importancia de las variables utilizadas para el modelo RF a nivel L3

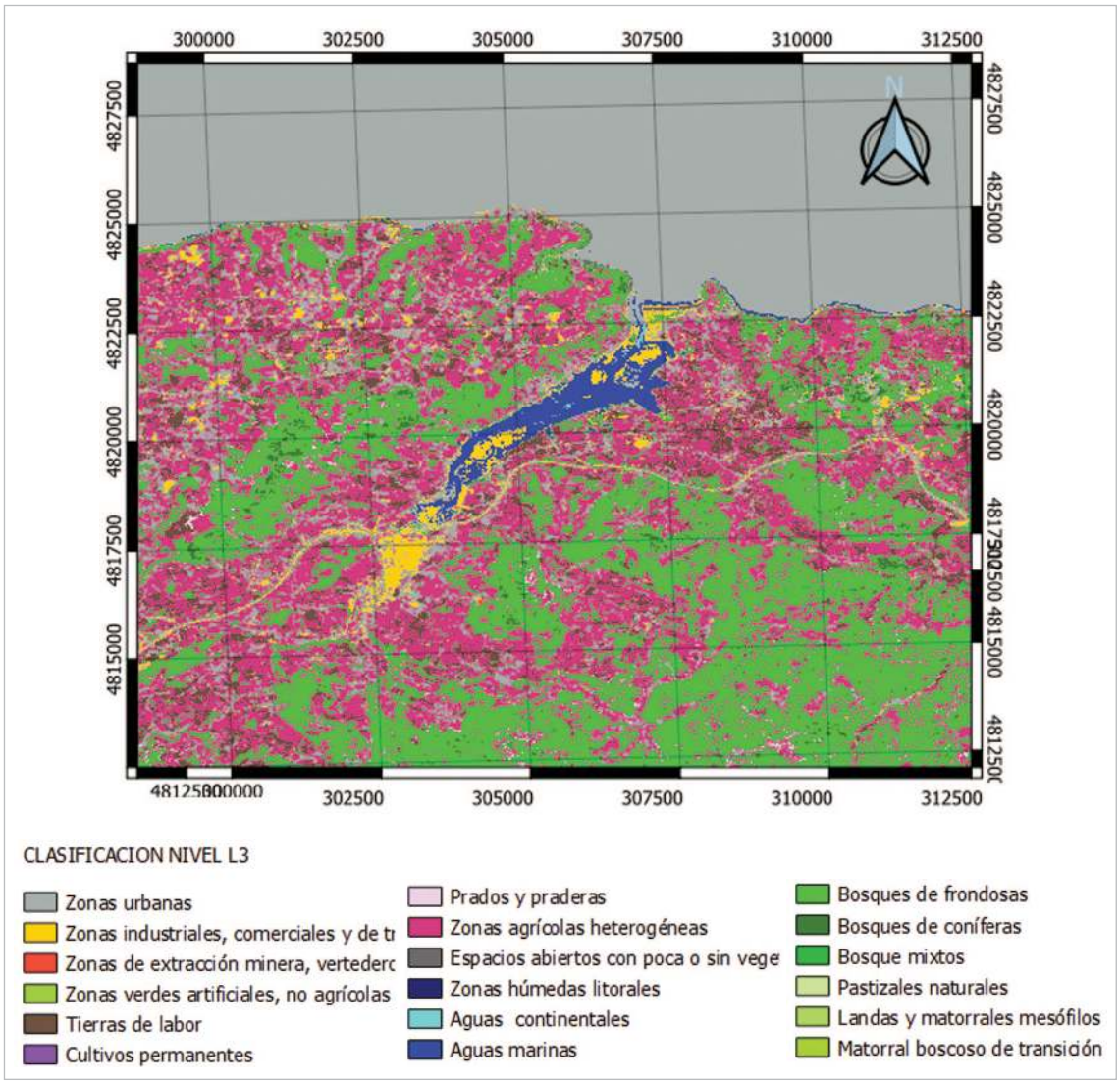

Figura 7. Vista del mapa a nivel L3 de coberturas para el concejo de Villaviciosa EPSG: 32630.

El nivel L4 clasifica el territorio en 30 grandes clases de uso *(tabla 4)*. El método *Wrapper* de reducción de la dimensionalidad seleccionó 17 variables predictoras de las 82 consideradas inicialmente. La relación de variables empleadas junto con su importancia en el modelo se presenta en la *figura 8*. Los estadísticos de bondad del ajuste pueden verse en la *tabla 12* y por clase en la *tabla 13*. La *figura 9* muestra un mapa espacialmente continuo con el resultado de la clasificación. Para el nivel L4, el valor de Kappa obtenido fue de 0,5723.

Tabla 12. Ajustes del modelo RF para la Clasificación nivel L4

Variable	Entrenamiento	Validación
Precisión	1	0,5938
Kappa	1	0,5723

Tabla 13. Estadísticos del modelo RF (validación) para la Clasificación nivel L4

Clase / Estadístico	P. usuario	P. productor	Error de Comisión	Error de Omisión
0	0,5714	0,8000	0,4286	0,2000
1	0,7727	0,6800	0,2273	0,3200
2	0,8000	0,5333	0,2000	0,4667
3	1,0000	0,4000	0,0000	0,6000
4	0,5714	0,6154	0,4286	0,3846
5	1,0000	1,0000	0,0000	0,0000
6	0,6538	0,7391	0,3462	0,2609
7	0,3818	0,6176	0,6182	0,3824
8	0,8000	0,6667	0,2000	0,3333
9	0,7500	0,5000	0,2500	0,5000
10	0,6818	0,7895	0,3182	0,2105
11	0,5000	0,4375	0,5000	0,5625
12	0,6667	0,4286	0,3333	0,5714
13	0,7222	0,5652	0,2778	0,4348
14	0,8095	0,5313	0,1905	0,4688
15	0,0000	0,0000	1,0000	1,0000
16	1,0000	0,7500	0,0000	0,2500
17	0,5385	0,3182	0,4615	0,6818
18	0,3478	0,4706	0,6522	0,5294
19	0,8000	0,6000	0,2000	0,4000
20	0,3333	0,5556	0,6667	0,4444
21	0,0000	0,0000	1,0000	1,0000
22	0,2500	0,5000	0,7500	0,5000
23	0,6296	0,6296	0,3704	0,3704
24	0,3750	0,4000	0,6250	0,6000
25	0,3333	0,1579	0,6667	0,8421
26	0,6780	0,8333	0,3220	0,1667
27	0,0000	0,0000	1,0000	1,0000
28	1,0000	0,8571	0,0000	0,1429
29	0,8750	0,8750	0,1250	0,1250

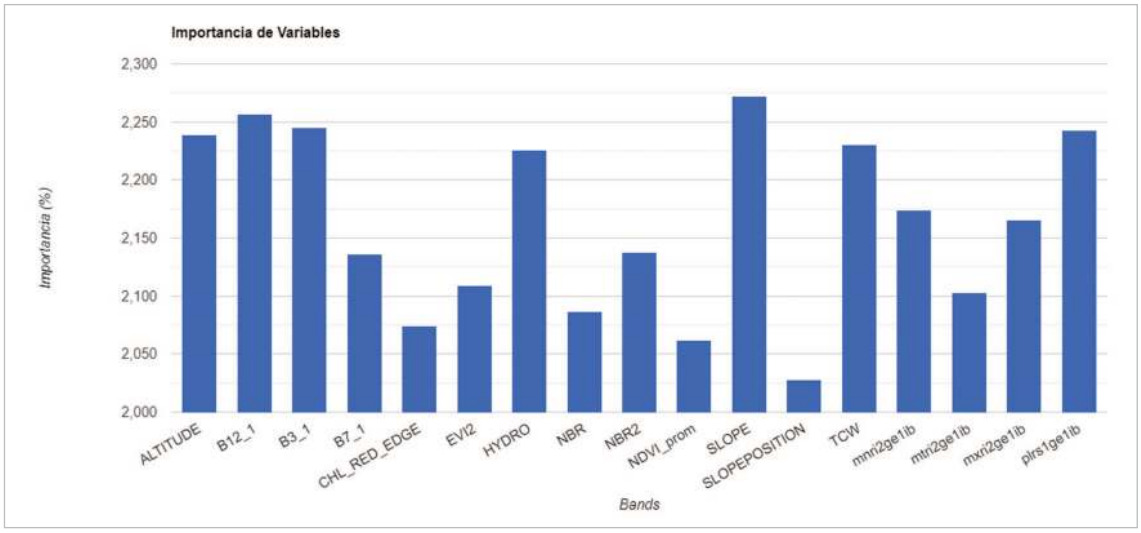

Figura 8. Importancia de las variables utilizadas para el modelo RF a nivel L4

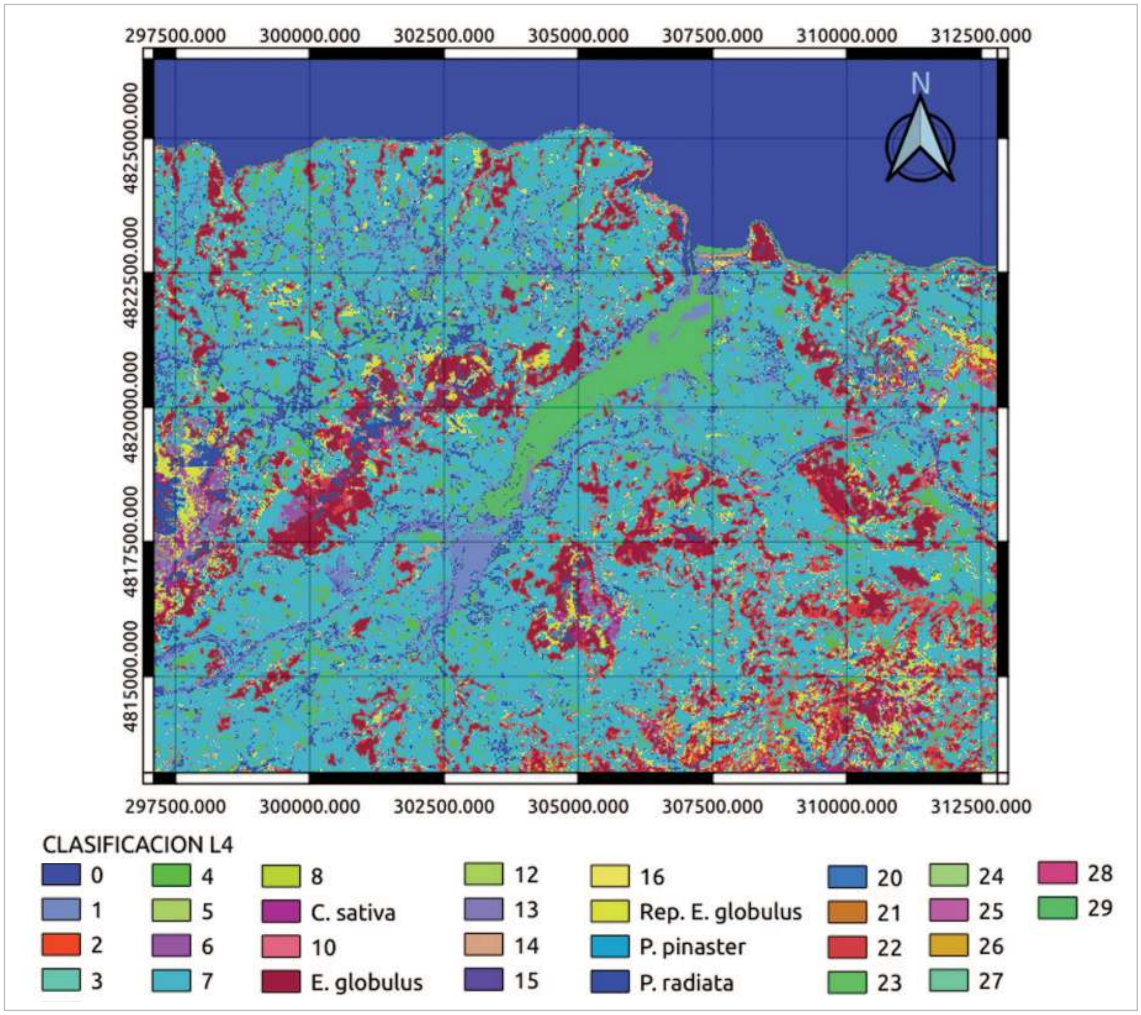

Figura 9. Vista del mapa a nivel L4 de coberturas para el concejo de Villaviciosa.

Tabla 14. Matriz de confusión de validación para el nivel L4

C/C	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
0	**20**	1	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	**17**	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	**8**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3	2	0	0	**2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	**8**	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	1	0	0	0	0	**17**	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	2	0	0	0	1	0	4	**21**	0	1	1	0	0	0	0	0	0	2	1	0	0	0	1	1	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	**8**	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0
9	0	0	0	0	0	0	0	4	0	**9**	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	**15**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	2	0	1	0	**7**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	0	0	2	0	1	**13**	1	0	0	0	0	0	0	0	0	3	0	0	1	0	0	0
14	1	0	0	0	2	0	1	2	1	0	1	0	2	0	**17**	0	0	0	0	0	0	0	3	1	0	0	2	0	0	0
15	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	1	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	2	4	0	0	0	1	0	0	0	0	0	**7**	2	0	0	0	4	1	0	1	0	0	0	0
18	0	0	0	0	2	0	0	0	0	0	0	3	0	0	0	0	0	0	**8**	1	1	0	2	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	**12**	2	0	0	0	0	0	0	0	0	0
20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	**5**	0	0	0	1	0	1	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0	1	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	**5**	0	1	1	1	0	0	0
23	0	0	0	0	0	0	0	3	0	0	1	0	0	0	0	0	0	0	2	0	0	0	1	**17**	0	0	3	0	0	0
24	0	0	0	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	**6**	0	1	0	0	0
25	0	1	0	0	0	0	2	4	0	0	0	0	0	0	0	0	0	1	3	0	1	0	0	1	0	**3**	2	0	0	0
26	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	1	1	2	0	**40**		0	0
27	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0
28	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	0
29	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**7**

Observando la matriz de confusión para nivel L4 *(tabla 14)*, el número de clases con errores elevados aumentan con respecto a los otros niveles, llegando a siete clases; 7, 18, 20, 21, 22, 24 y 25, correspondientes a: zonas agrícolas heterogéneas (2400), *Pinus pinaster* (31201), *Pinus sylvestris* (31203), *Pseudotsuga menziesii* (31204), mezcla de especies (31301), landas y matorrales mesófilos (32200) y matorral boscoso de transición (32400). Esta confusión entre estas clases se debe principalmente a la similitud de sus firmas espectrales.

Por ejemplo, la confusión entre rodales de *Castanea sativa* y *Betula alba* se debe principalmente a la presencia de pies de una especie en el rodal de la otra, introduciendo "ruido" a sus firmas espectrales a la hora de ajustar los modelos.

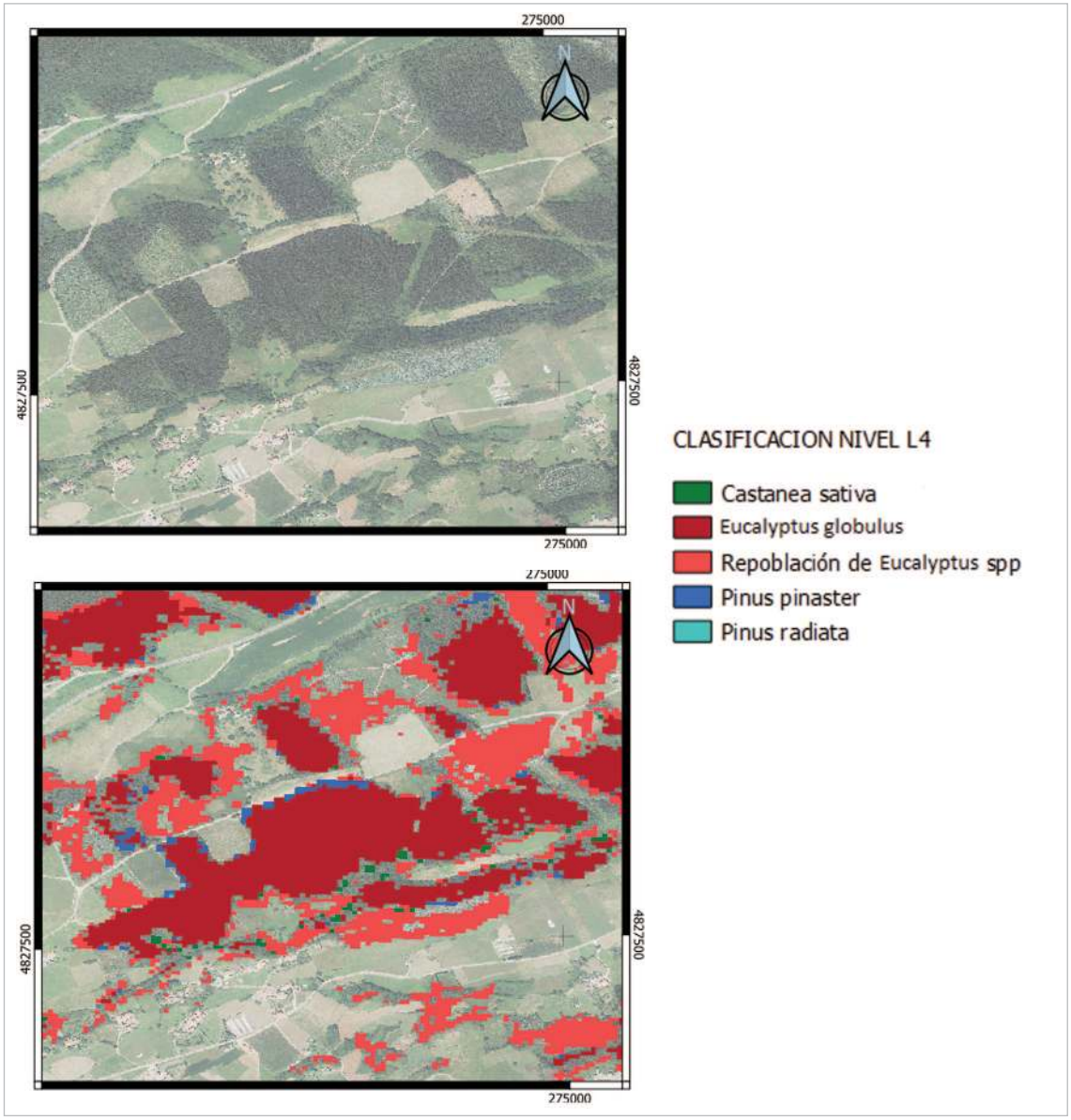

Figura 10. Comparación del nivel L4 de coberturas con la ortofoto del PNOA para el concejo de Carreño. EPSG: 32630

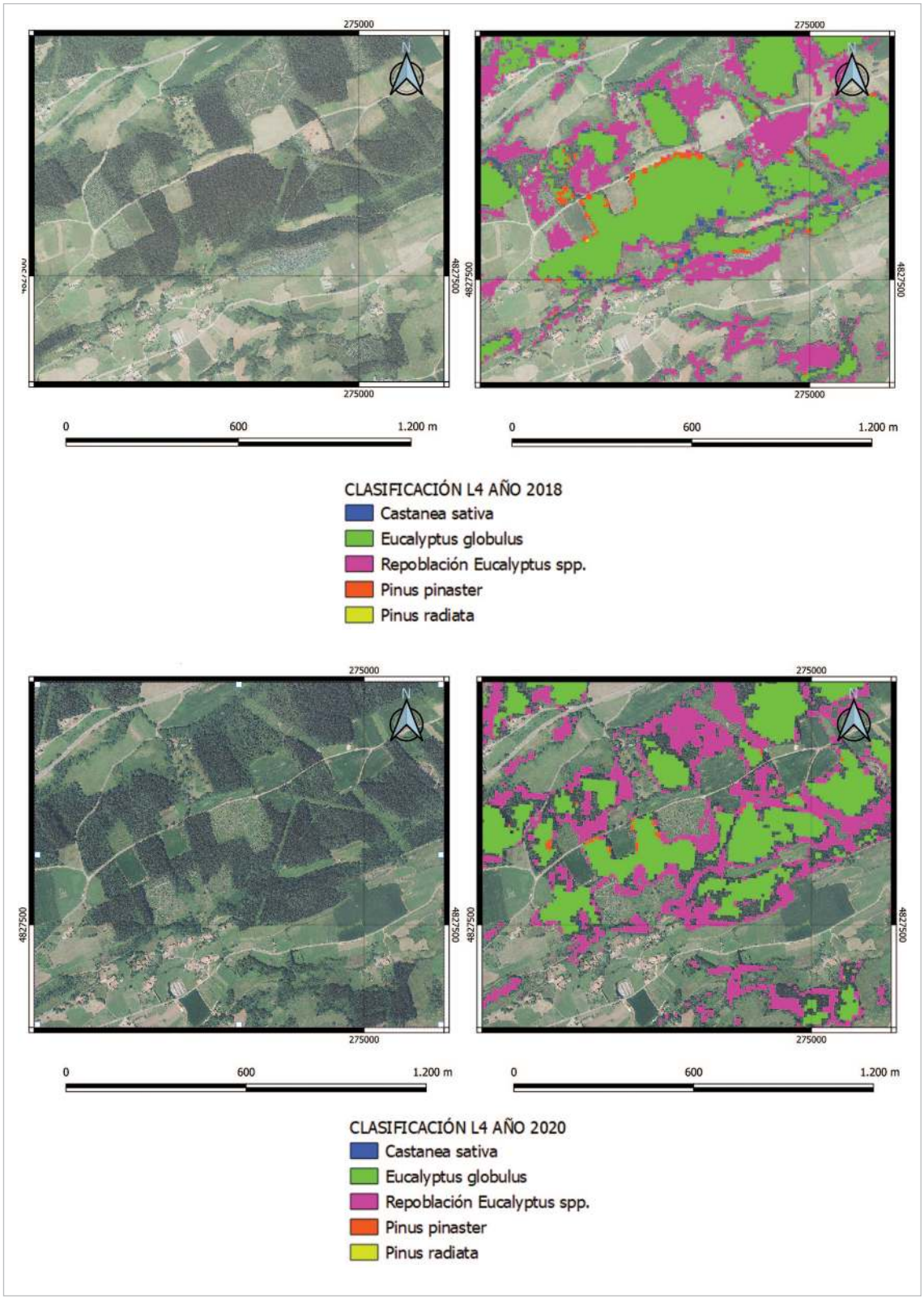

Figura 11. Comparación del modelo aplicado en la colección de imágenes de 2018 (superior) y de 2020 (inferior); a la izquierda de cada una se observa la ortofotografía del PNOA para ese año. EPSG: 32630.

En ocasiones, para *Eucalyptus globulus*, el modelo, clasifica masas correspondientes a coníferas como eucalipto y viceversa *(figuras 12* y *13)*. Las masas de *Pinus pinaster* y *Pinus radiata* además de lo anterior, presentan dificultades ya que el modelo también clasifica masas de estas especies como de matorral o mezcla de especies, pudiendo deberse a su baja densidad cuando se encuentran cercanas al turno. En la *figura 11* se muestra el modelo aplicado a dos colecciones de imágenes diferentes.

Al obtener los datos de clasificación a partir de la red de parcelas del IFN (nodos de cuadrícula UTM de 1 x 1 km), existen dificultades para obtener un número aceptable de parcelas por clase en algunas coberturas. Autores como Lean Qu *et al.*, (2021) han utilizado 10 000 parcelas por clase; capturando en mucha mayor medida la transición en la variabilidad de las firmas espectrales de las diferentes coberturas analizadas. A la vista de esto, un aumento del número de parcelas por clase mejoraría seguramente la precisión obtenida.

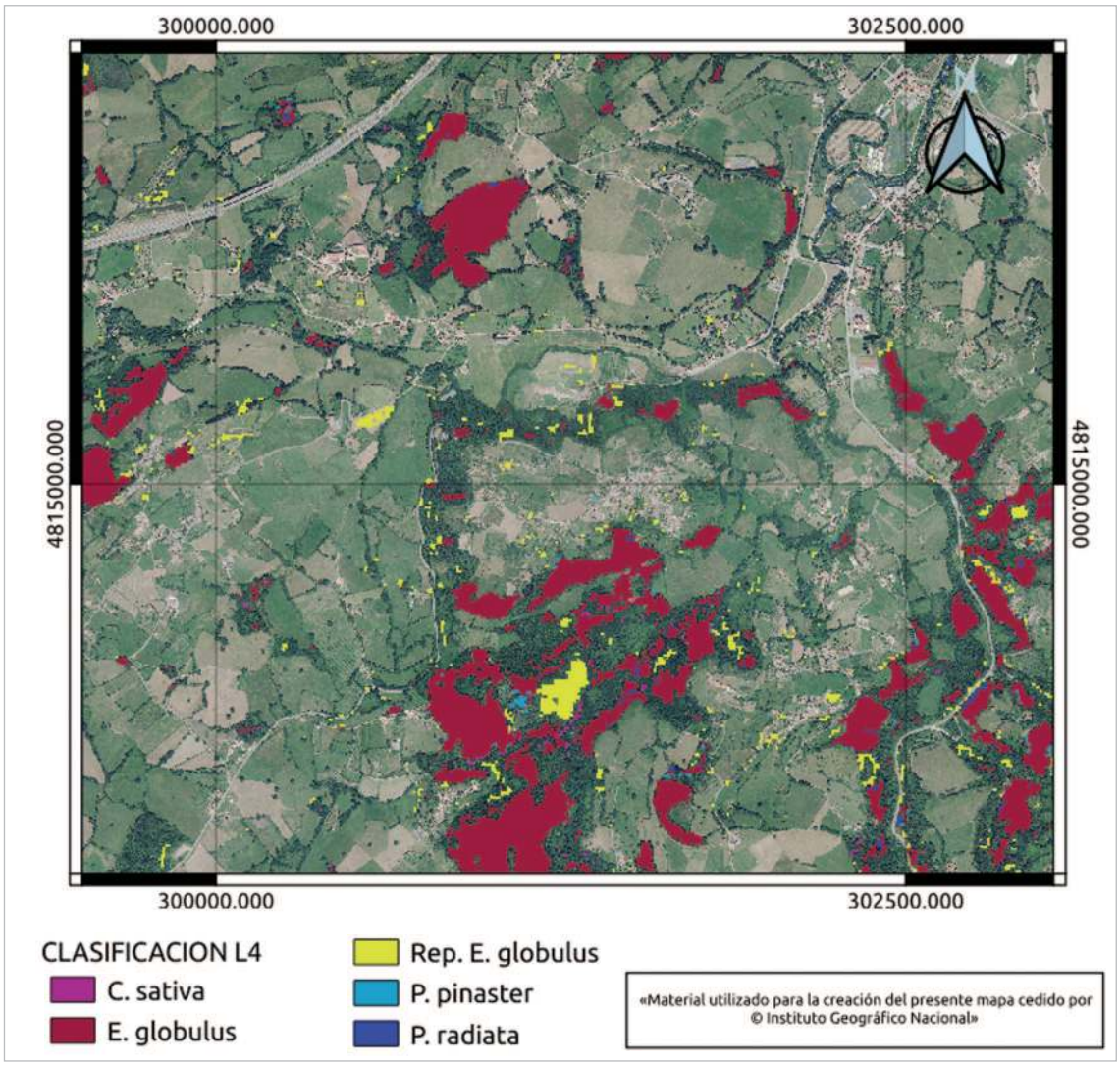

Figura 12. Efecto "sal y pimienta" en el mapa a nivel L4 de coberturas EPSG: 32630

En la *figura 12*, se muestran unidades homogéneas salpicadas por pixeles dispersos de otras categorías (efecto denominado como "sal y pimienta"), que se podría subsanar mediante el empleo de un enfoque de clasificación supervisada orientada a objetos (Weih *et al.*, 2010) en vez de a píxeles. Así, en paisajes heterogéneos, como el asturiano, el empleo de parámetros de segmentación adecuados junto con la consideración de variables auxiliares, permitiría obtener mapas de clasificación con mayor precisión (Ersoy *et al.*, 2021).

Autores como He *et al.*, (2022) recomiendan realizar un procesamiento topográfico de las imágenes satelitales puesto que las diferentes iluminaciones entre estaciones influyen en la clasificación. La corrección del sombreado topográfico homogenizaría posibles variaciones, generando mapas de coberturas vegetales más precisos. Debido a que los servidores de Google Earth Engine® no disponen de un

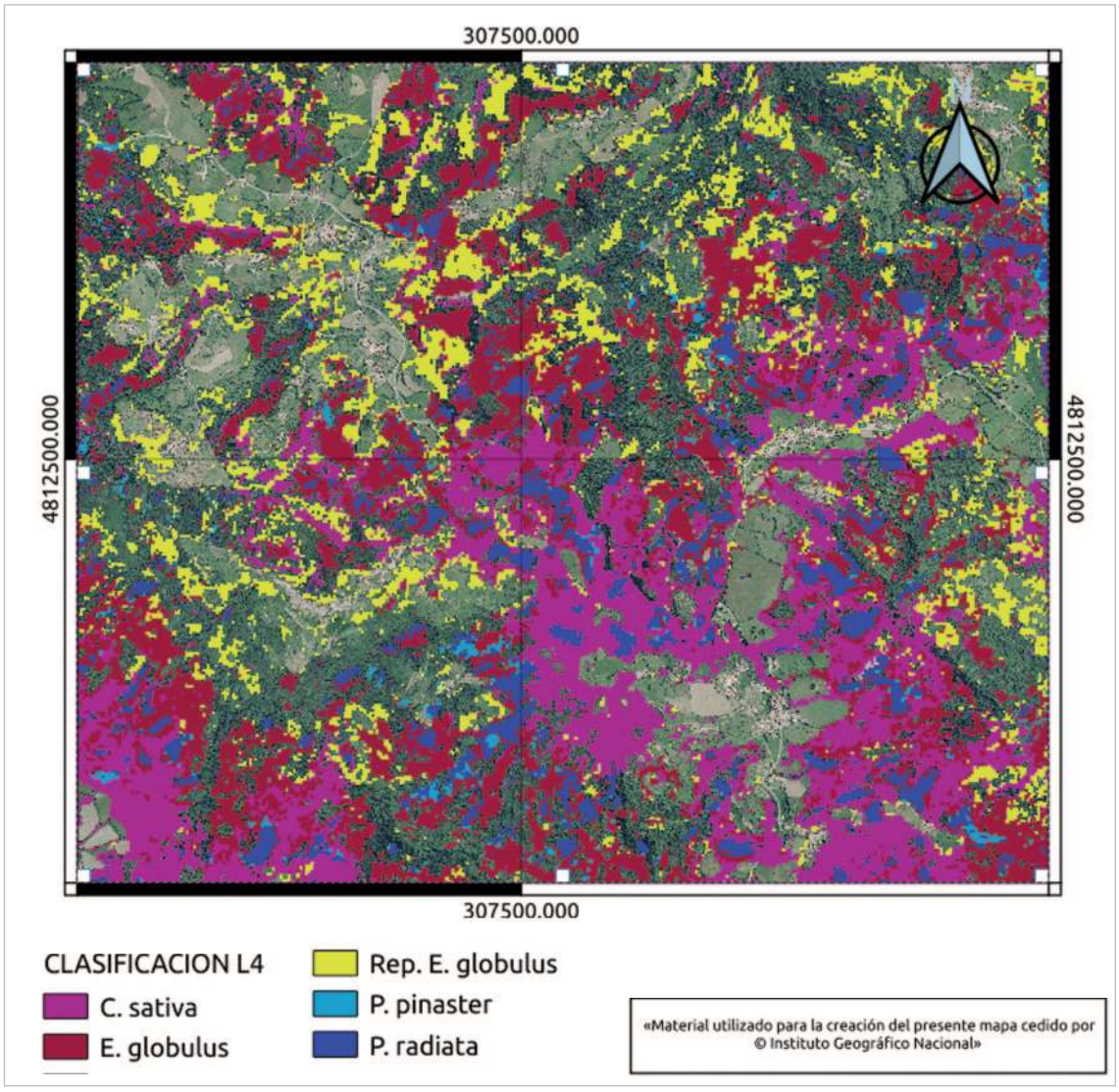

Figura 13. Clasificación de coberturas con menor precisión en zonas del interior de Villaviciosa, valles cerrados con elevada pendiente.

repositorio de imágenes con ese nivel de procesamiento, las correcciones topográficas no se han tenido en cuenta, comprometiendo los resultados en las zonas con orografía más compleja.

En la *figura 13* se pueden observar píxeles clasificados como *Pinus pinaster* o *Pinus radiata* en masas de *Eucalyptus globulus*, atribuible también a la similitud entre las firmas espectrales de las especies (Forstmaier *et al.*, 2020). En zonas de borde, el eucalipto parece tener una firma espectral similar a la del pino radiata, esto puede ser debido al tamaño de pixel (10 m), que no captura la zona perteneciente a la masa de eucalipto si no que toma información de las parcelas colindantes (matorrales o cultivos) siendo los valores de reflectancia del pixel más similares a los del pino que a la masa pura de eucalipto.

Se aprecia una aparente clasificación correcta de las superficies con repoblaciones de eucalipto en la zona costera. No obstante, en las zonas más montañosas esta clasificación no es precisa, confundiendo matorral con repoblaciones de eucalipto probablemente consecuencia de la similitud de las firmas espectrales, lo cual puede verse favorecido por la ausencia de correcciones topográficas.

4. Conclusiones

El trabajo utiliza una metodología automatizada para la clasificación supervisada de coberturas a nivel píxel replicable en las sucesivas imágenes satelitales (Sentinel-2). La precisión que ofrece el sensor es aceptable, aunque tiene ciertas limitaciones debido a la complejidad del medio estudiado. Por otro lado, la herramienta de Google Earth Engine (GEE) aporta gran capacidad de cómputo en la nube por lo que es muy recomendable para este tipo de aplicaciones. La mejora de la clasificación pasaría por:

1. Una segmentación previa del territorio.

2. La implementación de correcciones topográficas a las imágenes.

3. Incrementar el número de parcelas de campo. En los próximos años, la mayor diversidad y mejora de los datos capturados por los sensores remotos, el desarrollo de plataformas como GEE, y la obtención de nuevos algoritmos (Deep Learning) permitirán obtener mejores clasificaciones de coberturas y probablemente se convertirán en herramientas indispensables en la gestión de los recursos naturales.

Agradecimientos/Financiación

El trabajo no recibió financiación pública o privada alguna.

5. Bibliografía

Baret, F.; Hagolle, O.; Geiger, B.; Bicheron, P.; Miras, B.; Huc, M.; . . . Leroy, M.; (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. *Remote Sensing of Environment*, 110(3), 275-286. https://doi.org/10.1016/j.rse.2007.02.018

Baret, F.; Weiss, M.; Bicheron, P.; Berthelot, B.; (2010). S2 PGDS - *Sentinel-2-Level-2A-Algorithm-Theoretical-Basis-Document-ATBD*. European Spatial Agency.

Breiman, L. (2001). *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/a:1010933404324

Caparros-Santiago, J.A.; Rodriguez-Galiano, V.; Dash, J.; (2021). Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171, 330-347. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.11.019

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. https://doi.org/10.1177/001316446002000104

Ersoy Tonyaloglu, E.; Erdogan, N.; Çavdar, B.; Kurtsan, K.; Nurlu, E.; (2021). Comparison of pixel and object based classification methods on rapideye satellite image. *Turkish Journal of Forest Science*, 5(1), 1-11. https://doi.org/10.32328/turkjforsci.741030

Fielding, A.H.; Bell, J.F.; (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models | *Environmental Conservation* | Cambridge Core. Environmental Conservation, 24(1), 28-49. https://doi.org/doi:10.1017/S0376892997000088

Forstmaier, Andreas; Shekhar, Ankit; Chen, Jia.; (2020). Mapping of Eucalyptus in Natura 2000 Areas Using Sentinel 2 Imagery and Artificial Neural Networks, Remote Sensing. 12. 2176. 10.3390/rs12142176.

He Yin, Bin Tan, David Frantz, Volker C. Radeloff, Integrated topographic corrections improve forest mapping using Landsat imagery, International Journal of Applied Earth Observation and Geoinformation, Volume 108. 2022, 102716 ISSN 1569-8432. https://doi.org/10.1016/j.jag.2022.102716. /2529310 https://www.sciencedirect.com/science/article/ pil/S0303243422000423

JR, L.; GG, K.; (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1). https://doi.org/10.2307/2529310

Ministerio de Agricultura, A.; M.A.; (2012). Anuario de Estadística Forestal 2012. Retrieved 10/04/2022 from https://www.miteco.gob.es/es/biodiversidad/estadisticas/forestal_anuario_2012.aspx

Ministerio de Agricultura, P.; A.; (2019). Anuario de Estadística Forestal 2019. Retrieved 05/04/2022 from https://www.miteco.gob.es/es/biodiversidad/estadisticas/forestal_anuario_2019.aspx

Qu, L.A.; Chen, Z.; Li, M.; Zhi, J.; Wang, H.; (2021). Accuracy Improvements to Pixel-Based and Object-Based LULC Classification with Auxiliary Datasets from Google Earth Engine. *Remote Sensing*, 13(3), 453. https://doi.org/10.3390/rs13030453

Rumora, L.; Miler, M.; Medak, D.; (2020). Impact of Various Atmospheric Corrections on Sen-

tinel-2 Land Cover Classification Accuracy Using Machine Learning Classifiers [Article]. ISPRS *International Journal of Geo-Information*, 9(4), 277. https://doi.org/10.3390/ijgi9040277

Weih; Robert Riggan, Norman; (2010). *Object-bassed classification vs. Pixel-based classification: Comparative importance of multi-resolution imagery.* International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 38.