

Lopez-Nieta, D.; Rodríguez-Galiano, V.; Guisado-Pintado, E. Caracterización de los sistemas duranes costeros del Atlántico SW de España mediante un método semi-supervisado a partir de compuestos de Sentinel-2

Caracterización de los sistemas duranes costeros del Atlántico SW de España mediante un método semi-supervisado a partir de compuestos de Sentinel-2

Lopez-Nieta, Diego¹ **Rodríguez-Galiano, Victor**¹ **Guisado-Pintado, Emilia**¹

¹ Departamento de Geografía Física y Análisis Geográfico Regional, Facultad de Geografía e Historia, Universidad de Sevilla

ORCID: Lopez-Nieta 0000-0002-6408-2637 Rodríguez-Galiano 0000-0002-5422-8305 Guisado-Pintado 0000-0003-1115-4059

Correspondencia: dlopezdelanieta@us.es vrgaliano@us.es eguisado@us.es

RESUMEN

Se presenta una metodología orientada al estudio y la monitorización de dunas costeras mediante el uso de imágenes de alta resolución de Sentinel-2 y algoritmos de machine learning. Se ha desarrollado una metodología de clasificación orientada a objetos semi-supervisada (en lugar del enfoque tradicional basado en píxeles) que integra el contexto espacial de la vegetación y sus relaciones al agrupar píxeles homogéneos. El uso de un algoritmo optimizado de Multiresolution Segmentation (MRS), junto con variables de textura e índices de vegetación estacionales (NDVI), permitió una representación más fiel de la estructura del ecosistema. El flujo de trabajo combina el etiquetado automático mediante K-means junto con la optimización del entrenamiento del algoritmo Random Forest a partir de diferentes subconjuntos de datos, mejorando la representatividad de las muestras y la replicabilidad del proceso. Aplicado a las dunas costeras de la costa atlántica andaluza, el modelo mostró una alta capacidad para discriminar las distintas clases de vegetación, consolidándose como una herramienta eficaz para el monitoreo ambiental y la toma de decisiones en entornos costeros dinámicos.



Palabras clave: *Sentinel-2, NDVI, segmentación, Random Forest*


Fecha de recepción: 18 febrero 2026 · Fecha de aceptación: 18 febrero 2026

Caracterización de los sistemas duranes costeros del Atlántico SW de España mediante un método semi-supervisado a partir de compuestos de Sentinel-2

Lopez-Nieta, Diego ⁽¹⁾, Rodriguez-Galiano, Victor ⁽¹⁾, Guisado-Pintado, Emilia ⁽¹⁾

⁽¹⁾ Departamento de Geografía Física y Análisis Geográfico Regional, Facultad de Geografía e Historia, Universidad de Sevilla.

 0000-0002-6408-2637, dlopezdelanieta@us.es ;  0000-0002-5422-8305, vrgaliano@us.es

 0000-0003-1115-4059, eguisado@us.es

Resumen: Se presenta una metodología orientada al estudio y la monitorización de dunas costeras mediante el uso de imágenes de alta resolución de Sentinel-2 y algoritmos de *machine learning*. Se ha desarrollado una metodología de clasificación orientada a objetos semi-supervisada (en lugar del enfoque tradicional basado en píxeles) que integra el contexto espacial de la vegetación y sus relaciones al agrupar píxeles homogéneos. El uso de un algoritmo optimizado de *Multiresolution Segmentation* (MRS), junto con variables de textura e índices de vegetación estacionales (NDVI), permitió una representación más fiel de la estructura del ecosistema. El flujo de trabajo combina el etiquetado automático mediante *K-means* junto con la optimización del entrenamiento del algoritmo *Random Forest* a partir de diferentes subconjuntos de datos, mejorando la representatividad de las muestras y la replicabilidad del proceso. Aplicado a las dunas costeras de la costa atlántica andaluza, el modelo mostró una alta capacidad para discriminar las distintas clases de vegetación, consolidándose como una herramienta eficaz para el monitoreo ambiental y la toma de decisiones en entornos costeros dinámicos.

Palabras clave: Sentinel-2, NDVI, segmentación, Random Forest

Semi-supervised approach for coastal dune vegetation classification

Abstract: *We introduce a methodology oriented to study and monitoring coastal dunes using high-resolution Sentinel-2 imagery and machine learning algorithms. Unlike traditional pixel-based approaches, the study applies an object-oriented classification that incorporates spatial context by grouping homogeneous pixels. To this end, an optimised Multiresolution Segmentation (MRS) algorithm is employed, together with texture variables and seasonal vegetation indices (NDVI), enabling a more faithful representation of ecosystem structure. The workflow combines automatic labelling using K-means with the optimisation of training data through Random Forest, thereby improving sample representativeness and process reproducibility. When applied to the coastal dune systems of the Andalusian Atlantic coast, the model demonstrated high capacity to discriminate among different vegetation classes, establishing itself as an effective tool for environmental monitoring and decision-making in dynamic coastal environments.*

Keywords: *Sentinel-2, NDVI, segmentation, Random Forest*

1. INTRODUCCIÓN

Las dunas costeras son ecosistemas dinámicos y complejos, formados por la interacción entre el viento, el oleaje y la vegetación (Hesp, 2002). Su importancia no radica únicamente en la elevada biodiversidad que albergan, sino también en su papel fundamental como barreras naturales frente al cambio climático, el aumento del nivel del mar y la erosión asociada a eventos extremos (Jackson *et al.*, 2019). Además, estos sistemas se encuentran seriamente amenazados por la presión antrópica (cambios en el uso del suelo, el turismo de masas, etc.), lo que hace imprescindible el desarrollo de herramientas más precisas para su gestión sostenible (García-Romero *et al.*, 2019).

En las últimas décadas, la teledetección, combinada con algoritmos de *machine learning*, ha revolucionado el mapeo y análisis de estos entornos (Agrillo *et al.*, 2023; Gupta *et al.*, 2024). El uso de imágenes de muy alta resolución espacial, obtenidas mediante drones o satélites (por ejemplo, WorldView), es costoso y de acceso limitado, dificultando el seguimiento continuo que requieren estos ecosistemas tan dinámicos (Gupta *et al.*, 2024). En este contexto, la misión Sentinel-2 se presenta como una alternativa potencialmente valiosa, gracias a su acceso gratuito y su alta frecuencia de revisita, aproximadamente cinco días. Sin embargo, su resolución espacial de 10 m plantea un reto metodológico debido a la marcada heterogeneidad del paisaje costero dunar. Para abordar esta limitación, se propone un enfoque basado en objetos (*Object-Based*

Image Analysis, OBIA), que permite agrupar píxeles en unidades con significado ecológico, integrando información espectral, espacial y de textura, y ofreciendo así una representación más fiel de la realidad del terreno (Guisan & Zimmermann, 2000).

Este estudio presenta una metodología innovadora basada en etiquetado híbrido y aprendizaje semi-supervisado para clasificar la vegetación en sistemas dunares. El objetivo principal es reducir la dependencia del trabajo de campo intensivo y de la fotointerpretación experta, procesos que suelen ser costosos, lentos y subjetivos.

2. MATERIAL Y MÉTODOS

2.1. Área de estudio

El estudio se centra en los sistemas dunares de la costa atlántica de Andalucía (Fig. 1), donde la suave pendiente de la plataforma continental y la influencia de grandes ríos como el Guadalquivir y el Guadalquivir (Guisado-Pintado & Málvarez-García, 2015) favorecen la formación de extensos sistemas dunares. Se seleccionaron los sistemas que presentan continuidad espacial, condiciones naturales o seminaturales, disponibilidad de espacio de acomodación y presencia de vegetación semipermanente.

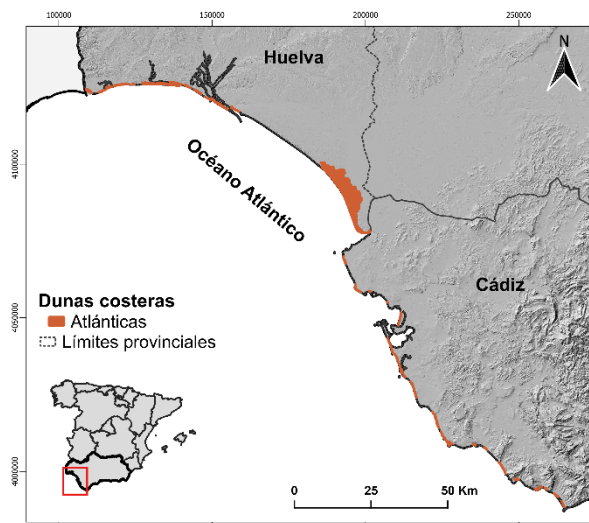


Figura 1. Área de estudio. En naranja se muestran los sistemas dunares costeros estudiados.

Desde el punto de vista ecológico, la vegetación de estos sistemas dunares atlánticos presenta una marcada zonificación. Esta se inicia en las zonas más expuestas, dominadas por comunidades pioneras de *Ammophila arenaria*, y evoluciona hacia áreas interiores progresivamente más estables. En estas zonas se desarrollan formaciones arbustivas compuestas por especies como el enebro o la retama, culminando en las áreas más protegidas del sistema con masas forestales de *Pinus pinea* (Muñoz Vallés *et al.*, 2015).

2.2. Adquisición y preprocesamiento de los datos

El área de estudio se delimitó digitalizando manualmente

los sistemas dunares a partir de ortofotos de muy alta resolución (0,25 m). El procesamiento de los datos satelitales se realizó en *Google Earth Engine* (GEE), utilizando la colección armonizada de imágenes Sentinel-2 Nivel 2. Se aplicó una máscara de nubes, cirros y sombras. El conjunto de datos incluyó bandas multispectrales del espectro visible (VIS), infrarrojo cercano (NIR) e infrarrojo de onda corta (SWIR1 y SWIR2), así como el NDVI (*Normalized Difference Vegetation Index*) y las variables de textura GLCM de Haralick (Haralick *et al.*, 1973). Con el fin de capturar la variabilidad temporal del ecosistema, se generaron compuestos anuales y estacionales (trimestrales) correspondientes al año 2017, utilizando la mediana para reducir la influencia de valores atípicos. Finalmente, las bandas SWIR1 y SWIR2, originalmente a 20 m, se remuestrearon a 10 m.

2.3. Clasificación orientada a objetos

La generación de los objetos se llevó a cabo mediante el algoritmo de Segmentación Multirresolución (*Multiresolution Segmentation*, MRS; Baatz, 2000) en el software eCognition 10.3, otorgando mayor peso a la información espectral frente a la forma con el fin de representar adecuadamente la vegetación. El tamaño óptimo de los objetos se determinó automáticamente utilizando la herramienta ESP2 (*Estimation Scale Parameter 2*; Drăguț *et al.*, 2014). Una vez segmentada la imagen, los objetos se dividieron en dos subconjuntos: el 80 % se destinó al entrenamiento del modelo y fue clasificado automáticamente mediante el algoritmo *K-means* (Hartigan & Wong, 1979); el 20 % restante se usó como validación independiente y fue etiquetado manualmente siguiendo la clasificación ecológica EUNIS (Chytrý *et al.*, 2020).

Para definir el número óptimo de clases, se aplicó el método *Elbow* (Thorndike 1973), obteniendo cuatro clases. Con el objetivo de reducir errores de clasificación y eliminar ruido en los datos de entrenamiento, se calculó la distancia espectral de cada objeto al centro de su clase (Chen & Xia, 2009). Posteriormente se agruparon los objetos en diferentes subconjuntos de entrenamiento en función del percentil de distancia (P10, P20, etc.). Estos se evaluaron para cada clase utilizando el algoritmo de clasificación supervisada *Random Forest* (RF; Breiman, 2001). La mejor combinación de subconjuntos sirvió como entrenamiento para la clasificación final. Este enfoque semisupervisado permitió depurar de forma automatizada las muestras de entrenamiento antes de la fase final de clasificación. La clasificación final se realizó también mediante el algoritmo RF, incorporando variables como el NDVI estacional, NIR, SWIR1 y SWIR2, y texturas GLCM de Haralick. Para determinar los valores óptimos de los hiperparámetros de RF (ntree: 500, 1000, 5000; mtry: 1–13; número de variables aleatorias) y seleccionar las características más importantes, se utilizó una validación cruzada de 10 pliegues (Rodríguez-Galiano *et al.*, 2012). Para evaluar la exactitud cartográfica se consideraron las siguientes métricas: exactitud global, errores de omisión y comisión y coeficiente Kappa.

3. RESULTADOS

Los resultados de la segmentación dieron lugar a un total de 21.789 objetos. Su amplia extensión y la complejidad interna de los sistemas dunares costeros atlánticos se reflejan en los valores del parámetro de escala identificados por el algoritmo ESP2, que oscilaron entre 25 y 38. Esta variabilidad pone de manifiesto un litoral altamente heterogéneo, con una marcada diversidad en el tamaño, la morfología y la composición de las dunas.

Desde el punto de vista de la cobertura vegetal, se distinguieron cuatro categorías principales que siguen el gradiente natural de sucesión ecológica. La vegetación escasa (VE), que se asocia principalmente a dunas embrionarias y móviles, con una presencia destacada en áreas como Doñana y Valdevaqueros (Fig. 2). La vegetación herbácea (VH), que se encuentra fundamentalmente asociada a las dunas móviles. La vegetación mixta (VM), la cual representa una etapa de transición hacia condiciones de mayor estabilidad. Finalmente, la vegetación leñosa (VA) corresponde a la fase más madura del sistema, formaciones forestales como los pinares de *P. pinea*.

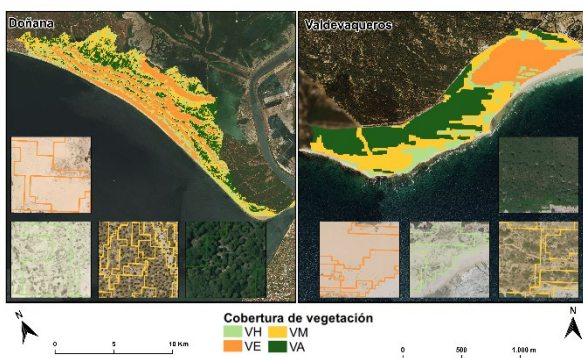


Figura 2. Mapa final de la clasificación RF de las dunas costeras de la región atlántica. La izq. se corresponde al P.N. de Doñana (Huelva) y la der. a las dunas de Valdevaqueros (Cádiz). Los recuadros de detalle muestran la ortofotografía original utilizada para la validación visual y el resultado de la segmentación de objetos.

Tabla 1. Resumen de la exactitud de la clasificación de RF.

Clase	Err. Omisión	Err. Comisión
VE	0,08	0,18
VH	0,29	0,17
VM	0,18	0,16
VA	0,04	0,25
Exact. global		0,82
Kappa		0,75

En cuanto a la optimización del entrenamiento mediante la selección de subconjuntos, esta región mostró un comportamiento donde la mayor capacidad de generalización se alcanzó en los percentiles bajos (20 y 30) para casi todas las categorías. La única excepción fue la clase de VM, que necesitó alcanzar el percentil 70 para mostrar su mejor desempeño. Bajo esta configuración óptima, las clases de VE y VA demostraron una exactitud excepcional al presentar errores de omisión inferiores al 10%, lo que indica que apenas son subestimadas por el modelo. Sin embargo, las clases más complejas, como VH y VM, lograron errores de

omisión del 25% y 30%, respectivamente. Estos valores reflejan una mayor dificultad de detección, lo que resulta en una subestimación moderada de estas categorías por parte del modelo. El mapa de clasificación basado en RF (Fig. 2) mostró un rendimiento robusto en ambos sistemas, alcanzando una exactitud global de 0,82 y un índice Kappa de 0,75 (Tabla 1). Las clases situadas en los extremos del gradiente sucesional presentaron los mejores resultados, con errores de omisión reducidos (0,08 para VE y 0,04 para VA) y una alta concordancia entre los datos observados y los predichos. En contraste, en las zonas de transición se obtuvieron errores de omisión más elevados, especialmente en VH (29%) y VM (18%), lo que refleja una mayor dificultad del modelo para delimitar estas etapas intermedias.

4. DISCUSIÓN

La región atlántica destaca por albergar los sistemas dunares costeros más extensos y desarrollados del litoral andaluz. Su complejidad geomorfológica y climática se traduce en una gran diversidad de especies vegetales (Guisado-Pintado & Málvarez-García, 2015). A diferencia de otras zonas más fragmentadas, el litoral atlántico permite observar una sucesión ecológica completa y bien estructurada (Muñoz Vallés *et al.*, 2015). El uso de datos de Sentinel-2 permitió la identificación de las distintas etapas de presencia de vegetación en estos sistemas dunares costeros, desde las dunas embrionarias hasta las dunas fijas.

El análisis subraya que la vegetación de los extremos de la duna es la más fácil de catalogar debido a su homogeneidad. Por un lado, las zonas de vegetación escasa en el frente costero tienen una firma visual muy clara sobre la arena blanca; por otro, los bosques de las dunas fijas interiores presentan una alta actividad fotosintética que el sensor detecta sin dificultad. Para optimizar la clasificación de la vegetación de esta región, se comprobó que basta con seleccionar una muestra más pequeña de observaciones más fiables (percentiles bajos), lo que demuestra la robustez de las categorías vegetales en estos ecosistemas de gran escala. Sin embargo, el principal desafío se encontró en las denominadas dunas grises o de transición. En lugares con una dinámica muy activa, como el Parque Nacional de Doñana, los límites entre la vegetación herbácea y los arbustos son difusos, creando un mosaico donde las especies se mezclan (Muñoz Vallés *et al.*, 2015). Esta heterogeneidad estructural genera ambigüedad, ya que el modelo a veces confunde áreas donde conviven pastizales con individuos leñosos dispersos (Hesp, 2002). A pesar de estas dificultades en las zonas intermedias, el método permitió una alta identificación, lo cual podría ser fundamental para estudios continuos que apoyen la gestión de estos entornos frente a la presión humana y el cambio climático.

5. CONCLUSIONES

El estudio propone una metodología avanzada que combina teledetección y aprendizaje automático para la clasificación de la vegetación en sistemas dunares costeros, caracterizados por su alta complejidad y dinamismo. Mediante imágenes Sentinel-2, se emplea

una clasificación orientada a objetos que permite el monitoreo de estos ecosistemas con una frecuencia de cinco días. El modelo muestra una buena exactitud en la identificación de áreas con vegetación escasa y formaciones forestales densas, gracias a firmas espectrales claras y niveles de biomasa estables, aunque presenta limitaciones en zonas de vegetación herbácea o mixta debido a su heterogeneidad. En un futuro, la integración de datos multitemporales y modelos digitales del terreno supondría una mejora. No obstante, se plantea como una herramienta eficiente, replicable y de bajo coste, con potencial aplicación en ecosistemas similares a escala global.

6. AGRADECIMIENTOS

El primer autor es beneficiario de una ayuda FPU financiada por el Ministerio de Universidades (Referencia FPU21/01033). Esta actividad ha sido financiada por la Unión Europea-NextGeneration EU y la Consejería de Universidad, Investigación e Innovación de la Junta de Andalucía, a través del Plan de Recuperación, Transformación y Resiliencia (PRTR) y del Plan Complementario de «Ciencias Marinas», subproyecto C17.I01.P01.S22, Proyecto (PCM_00009), mediante subvención pública. Componente 17 «Reforma institucional y fortalecimiento de las capacidades del sistema nacional de ciencia, tecnología e innovación», Inversión I1 «Planes Complementarios con las Comunidades Autónomas». Este trabajo constituye una contribución al IGCP 725 «Forecasting Coastal Change».

7. REFERENCIAS

- Agrillo, E. *et al.* (2023). Modeling approach for coastal dune habitat detection on coastal ecosystems combining very high-resolution UAV imagery and field survey. *Remote Sens. Ecol. Conserv.*, 9(2), 251-267. <https://doi.org/10.1002/rse2.308>
- Baatz, M., & Schäpe, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J. *et al.* (eds.): *Angew. Geogr. Inf. XII*. Wichmann, Heidelberg, pp. 12-23.
- Breiman, L. (2001). Random Forests. *Mach. Learn.*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Chytrý, M. *et al.* (2020). EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. *Appl. Veg. Sci.*, 23(4), 648-675. <https://doi.org/10.1111/avsc.12519>
- Chen, Z. & Xia, S. (2009). K-means Clustering Algorithm with Improved Initial Center. *2nd Int. Workshop Knowl. Discov. Data Min.*
- Drăguț, L. *et al.* (2014). Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.*, 88, 119-127. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2013.11.018>
- Gracia-Romero, L. *et al.* (2019). Biogeomorphological processes in an arid transgressive dunefield as indicators of human impact by urbanization. *Sci. Total Environ.*, 650, 73-86. <https://doi.org/10.1016/j.scitotenv.2018.08.429>
- Guisado-Pintado, E., & Malvárez García, G. C. (2015). El estado morfodinámico de las playas a través de modelización numérica de propagación y asomeramiento del oleaje: el frente litoral de Doñana. *GeoFocus*, 0(15), 163-180. <https://geofocus.org/index.php/geofocus/article/view/411>
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecol. Modell.*, 135(2), 147-186. [https://doi.org/https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/https://doi.org/10.1016/S0304-3800(00)00354-9)
- Gupta, S. *et al.* (2024). Unveiling the Invasion: Advancing Ecological Mapping of *Heterotheca Subaxillaris* Through Integrated Remote Sensing Techniques with Drones and Satellites. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 17, 7193-7211. <https://doi.org/10.1109/JSTARS.2024.3374232>
- Haralick, R. M. *et al.* (1973). Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.*, SMC-3(6), 610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *J. R. Stat. Soc. Ser. C (Appl. Stat.)*, 28(1), 100-108. <https://doi.org/10.2307/2346830>
- Hesp, P. (2002). Foredunes and blowouts: initiation, geomorphology and dynamics. *Geomorphology*, 48(1), 245-268. [https://doi.org/https://doi.org/10.1016/S0169-555X\(02\)00184-8](https://doi.org/https://doi.org/10.1016/S0169-555X(02)00184-8)
- Jackson, D. W. T. *et al.* (2019). A global 'greening' of coastal dunes: An integrated consequence of climate change? *Glob. Planet. Change*, 182. <https://doi.org/10.1016/j.gloplacha.2019.103026>
- Muñoz Vallés, S. *et al.* (2015). Effect of soil characteristics on plant distribution in coastal ecosystems of SW Iberian Peninsula sand spits. *Plant Ecol*, 216(11), 1551-1570. <https://doi.org/10.1007/s11258-015-0537-x>
- Rodríguez-Galiano, V. F. *et al.* (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.*, 67, 93-104. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4), 267-276. <https://doi.org/10.1007/BF02289263>