



**HAL**  
open science

# Value and limits of GBIF botanical occurrence data in mapping mangrove cover via the KNN classification of Sentinel-2 images

Julien Andrieu, Oliver-James Crook

## ► To cite this version:

Julien Andrieu, Oliver-James Crook. Value and limits of GBIF botanical occurrence data in mapping mangrove cover via the KNN classification of Sentinel-2 images. *Boletín de Investigaciones Marinas y Costeras*, 2024, 54 (1), 10.25268/bimc.invenmar.2025.54.1.1313 . hal-04907486

HAL Id: hal-04907486

<https://hal.science/hal-04907486v1>

Submitted on 23 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Valor y límites de los datos de registro botánico de la GBIF en el mapeo de la extensión de manglar mediante la clasificación KNN de imágenes de Sentinel-2

## Value and limits of GBIF botanical occurrence data in mapping mangrove cover via the KNN classification of Sentinel-2 images

Julien Andrieu\*

0000-0002-0031-1672

julien.andrieu@ifpindia.org

Oliver-James Crook

0009-0001-1607-1482

Institut Français de Pondichéry (CNRS MEA)

\* Autor de correspondencia / Corresponding author

**Recibido / Received:** 01/12/2023

**Aceptado / Accepted:** 07/08/2024

**Citación / Citation:** Andrieu, J.; Crook, O-J. 2025. Valor y límites de los datos de registro botánico de la GBIF en el mapeo de la extensión de manglar mediante la clasificación KNN de imágenes de Sentinel-2. Bol. Invest. Mar. Cost., 54(1): 49-70

### RESUMEN

En el contexto actual, que enfatiza la importancia de los datos de acceso abierto, este artículo busca realizar la primera evaluación de la efectividad de la base de datos de la Global Biodiversity Information Facility (GBIF) para entrenar clasificadores de imágenes satelitales multiespectrales de acceso abierto con el fin cartografiar la extensión de manglares. Se eligieron seis áreas en Australia, Benín, Brasil, Colombia, Malasia y EEUU por su densidad de datos de presencia de manglar y su proximidad con otras áreas de vegetación boscosa húmeda. Se utilizaron tres conjuntos de datos de acceso abierto: imágenes Sentinel-2, el visualizador de alta resolución de Google Earth y datos de presencia descargados del sitio de la GBIF. En primer lugar, se procesó un algoritmo de k-vecinos más cercanos (KNN), entrenándolo con píxeles circundantes de los datos de la GBIF. Luego, se compararon los resultados con los de una clasificación no supervisada (K-means). La exactitud de los dos métodos se evaluó mediante matrices de error de un proceso de fotointerpretación doble-ciego de imágenes de muy alta resolución obtenidas de Google Earth. El algoritmo KNN obtuvo índices *kappa* de 0.85 a 0.94, muy similares a los del método no supervisado (de 0.95 a 0.96). La GBIF posee un conjunto de datos adecuado para lugares con información, por lo que se invita a la comunidad a llenar los principales vacíos geográficos en su base de datos.

**Palabras clave:** datos abiertos, manglar, cartografía, evaluación de precisión, GBIF.

### ABSTRACT

In the current context, which emphasizes the importance of open-access data, this article aims to conduct the first evaluation of the effectiveness of the Global Biodiversity Information Facility's (GBIF) database for training open-access multispectral satellite image classifiers to map mangrove extent. Six areas in Australia, Benin, Brazil, Colombia, Malaysia, and the USA were selected for their density of mangrove occurrence data and their proximity to other areas with wooded humid vegetation. Three open-access datasets were used: Sentinel-2 images, Google Earth's high-resolution viewer, and occurrence data downloaded from the GBIF website. First, a k-nearest neighbors (KNN) algorithm was processed and trained with pixels surrounding the GBIF data. Then, the results were compared to those of an unsupervised classification (K-means). The accuracy of the two methods was assessed using error matrices from a double-blind photointerpretation of very high-resolution images obtained from Google Earth. The KNN algorithm achieved *kappa* indices ranging from 0.85 to 0.94, very similar to those of the unsupervised method (between 0.95 and 0.96). The GBIF possesses a suitable dataset for areas with information, which is why the community is encouraged to fill the main geographical gaps in its database.

**Keywords:** open data, mangrove, mapping, accuracy assessment, GBIF.

## INTRODUCCIÓN

La teledetección se aplica con gran regularidad al mapeo de manglares. Para dar algunos ejemplos, se ha aplicado en el mapeo de LULCC (Nababa *et al.*, 2020), en el mapeo a escala fina de árboles y parches de árboles de manglar posterior (Biswas *et al.*, 2020), en la estimación de la biomasa aérea forestal (Hu *et al.* 2020; Zhu *et al.* 2020; Pham *et al.* 2020) o en la caracterización fenológica (Younes-Cardenas *et al.* 2020). Al menos 10 artículos han realizado revisiones de la literatura en teledetección aplicada a manglares. En 1998, Green listó un conjunto de enfoques (incluyendo la distinción entre clasificación supervisada y no supervisada) y probó diferentes imágenes en un pequeño conjunto de estudios de caso. Heumann *et al.* (2001) y Kuentzer *et al.* (2001) hicieron especial énfasis en los datos –para entonces– nuevos (SAR, VHR) y en los métodos asociados (análisis de textura, OBIA). Los primeros abogaron por más colaboración con científicos en campo para la calibración, la validación y la interpretación de productos de teledetección, mientras que el segundo recomendó “actividades de levantamiento de terreno mediante muestreo aleatorio para verificar y calibrar los resultados de análisis de imágenes”. Purnamasayangasukasih *et al.* (2016) se centraron en comparar las resoluciones ópticas media y muy alta y los datos de radar, listando sus beneficios y limitaciones. Estos autores no compararon distintos métodos y no mencionaron los datos de campo en relación con la clasificación supervisada. Por su parte, Giri (2016) insistió en la multiplicación de los datos, en las oportunidades que brinda una mayor cantidad de datos y en la automatización con referencia al Google Earth Engine. A su vez, Cárdenas *et al.* (2017) profundizaron en las series de tiempo complejas para comprender mareas y estaciones y realizar monitoreo a largo plazo, destacando la importancia de una perspectiva integrada del ecosistema con base en datos de campo y recolectados por ciudadanos. Thakur (2018) y Pham (2019) exploraron los sensores y los métodos actuales, haciendo un llamado a enfocarse en la resolución espacial, en el potencial de las aplicaciones de múltiples sensores y en las oportunidades que brinda el aprendizaje automático. Wang *et al.* (2019) rastrearón la evolución de la investigación, desde los asuntos “simples” relacionados con la extensión de manglar hasta las “nuevas” problemáticas como la captura de carbono. Su trabajo concluyó con un llamado a la comunidad de investigación en manglares para ampliar el banco de conjuntos de datos de referencia estandarizados disponibles al público. Finalmente, Maurya *et al.*

## INTRODUCTION

Remote sensing is very regularly applied to mangrove mapping. To give a few examples, it has been applied in LULCC mapping (Nababa *et al.*, 2020), in the fine-scale mapping of trees and tree patches in back mangroves (Biswas *et al.*, 2020), in aboveground biomass estimation in forests (Hu *et al.* 2020, Zhu *et al.* 2020, Pham *et al.* 2020), or in phenological characterization (Younes-Cardenas *et al.* 2020). At least 10 papers have conducted literature reviews regarding remote sensing applied to mangroves. In 1998, Green listed a set of approaches (including the distinction between supervised and unsupervised classification) and tested different images on a small set of case studies. Heumann *et al.* (2001) and Kuentzer *et al.* (2001) focused mainly on – by then – new data (SAR, VHR) and associated methods (texture analysis, OBIA). The former advocated for more collaboration with field scientists for the calibration, validation, and interpretation of remote sensing products, while the latter recommended “random-sampling ground-survey activities to verify and calibrate image-analyses results”. Purnamasayangasukasih *et al.* (2016) focused on comparing medium and very high optical resolutions as well as radar data, listing their benefits and limitations. These authors did not compare different methods and failed to mention field data in relation to supervised classification. On the other hand, Giri (2016) insisted on the multiplication of data, on the opportunities provided by larger amounts of data, and on automatization with reference to the Google Earth Engine. Cárdenas *et al.* (2017) delved into complex times series in order to understand tides and seasons and perform long-term monitoring, highlighting the importance of an integrated view of ecosystems based on field and citizen-collected data. Thakur (2018) and Pham (2019) explored the currently employed sensors and methods, calling for a focus on spatial resolution, on the potential of multi-sensor applications, and on the opportunities provided by machine learning. Wang *et al.* (2019) traced the evolution of research, from “simple” mangrove extension issues to “new” problematics such as carbon sequestration. Their work concluded with a call for the mangrove research community to enlarge the pool of publicly available, standardized ground-truth datasets. Finally, Maurya *et al.* (2020) also advocated for the use of multiresolution and multi-source image data and machine learning methods.

(2020) también abogaron por el uso de datos de imagen multi-resolución y multi-origen y métodos de aprendizaje automático.

Este resumen de 10 revisiones de la literatura revela dos cosas. La primera es que la teledetección aplicada a los manglares ha evolucionado rápidamente, adoptando una a una todas las nuevas tendencias, impulsada por artículos de revisión que insisten en la urgencia/necesidad de mantenerse al día.

Un problema de esta tendencia a recomendar la tecnología más reciente es que no ayuda a la comunidad de teledetección a adherirse a los principios FAIR y, en general, no favorece a la ciencia abierta. En efecto, repetir que la resolución más alta garantiza mejores resultados, en épocas en las que el campo de la VHR está liderado por actores privados, está cortando el acceso por parte de algunos académicos (especialmente en países de bajos ingresos y, en general, en países que asignan pocos fondos a la investigación) a algunos datos considerados por otros como los mejores. Este artículo, siguiendo los principios FAIR, tiene por objetivo probar únicamente datos de acceso abierto: datos satelitales abiertos, datos abiertos de observación en campo y algoritmos disponibles en *software* libre.

El segundo problema es que la mayoría de las revisiones se interesan únicamente en datos y algoritmos de teledetección. Ninguno de los trabajos anteriormente presentados analizó las fortalezas y debilidades de los datos de campo (ya sea como sitios de entrenamiento o de control); solo tres de ellos incluían un comentario general sobre su importancia. No obstante, los datos usados como sitios de control y/o entrenamiento son importantes, pues inciden en los resultados y en la capacidad de realizar evaluaciones de precisión. Esta debe ser una preocupación principal de los especialistas en teledetección (Yancho *et al.* 2020, Heumann *et al.* 2011).

Incluso sin analizar cuantitativamente el estado del arte sobre teledetección, se puede advertir fácilmente que las clasificaciones supervisadas son elegidas con más frecuencia por encima de las no supervisadas. En la mayoría de los estudios basados en métodos supervisados, los sitios de entrenamiento y/o control provienen de las propias imágenes. En el mejor de los casos, estos sitios son digitalizados con transparencia desde Google Earth (*i.e.*, describiendo los criterios de interpretación y/o presentándolos a manera de mapas o material suplementario). En el peor de los casos se utiliza la misma imagen para digitalizar el sitio de

This overview of 10 literature reviews reveals two things. The first one is that remote sensing applied to mangroves has quickly evolved, adopting, one by one, all the new trends, driven by review papers insisting on the urgency to stay up-to-date.

One issue with this trend towards recommending the newest technology is that it does not aid the remote sensing community in sticking to the FAIR principles, and, in general, it does not favor open science. Indeed, repeating that the highest resolution guarantees better results, in times when the field of VHR is led by private actors, is hindering access by some scholars (especially in low-income countries and, in general, in countries that allocate little funding to research) to some data regarded by others as the best. This paper, following the FAIR principles, aims to exclusively test open access data: open satellite data, open data from field observations, and algorithms available in free software.

The second issue is that most reviews only take interest in remote sensing data and algorithms. None of the above-presented works analyzed the strengths and weaknesses of field data (be it as training or control sites); only three of them included a general comment on their importance. Nonetheless, data used as control and/or training sites are important, as they influence the results and the ability to perform accuracy assessments. This should be a major concern of remote sensing specialists (Yancho *et al.* 2020; Heumann *et al.* 2011).

Even without a quantitative analysis of the state of the art on remote sensing, one can easily note that supervised classifications are more often selected over unsupervised ones. In most studies based on supervised methods, the training and/or control sites stem from the imagery itself. In the best of cases, these sites are transparently digitized from Google Earth (*i.e.*, describing the interpretation criteria and/or presenting them in the form of maps or supplementary material). In the worst cases, the same image is used to digitize the training site and to produce results and control sites. This last option might entail circular reasoning. Some rarer papers use field data, but these are, in general, very local studies (BihamtaToosi *et al.* 2020; Quang *et al.* 2020; Zanvoet *et al.* 2021).

entrenamiento y producir resultados y sitios de control. Esta última opción puede generar un razonamiento circular. Algunos artículos menos comunes utilizan datos de campo, pero estos son, en general, estudios muy locales (BihantaToosi *et al.*, 2020; Quang *et al.*, 2020; Zanvo *et al.* 2021).

En el universo de datos abiertos sobre biodiversidad, la Global Biodiversity Information Facility (GBIF, una red e infraestructura de datos internacional financiada por los gobiernos del mundo y orientada a proveer acceso abierto a datos sobre todos los tipos de vida sobre la tierra para cualquiera y en cualquier lugar) es un gran contribuyente, con 1 900 000 000 registros (*i.e.*, la observación de un individuo en un determinado momento y lugar) y 60 000 conjuntos de datos, así como 1700 instituciones editoras y 6000 artículos científicos utilizando estos datos en 2024.

Sobre esta base “un análisis metodológico poco común de la calidad de los datos utilizados para el entrenamiento, en el marco de un enfoque en satélites de acceso abierto entrenados con (diferentes) datos de la misma naturaleza” este artículo simplemente busca realizar un primer análisis de la eficiencia de la base de datos de la GBIF en el entrenamiento relacionado con el procesamiento de imágenes satelitales ópticas multi-espectro, con énfasis en el mapeo preciso de la extensión de manglar. Se aborda la siguiente pregunta: ¿pueden los datos de acceso abierto sobre registros botánicos de la GBIF mejorar la precisión del mapeo de manglares mediante teledetección con satélites de acceso abierto? Este artículo estimará la robustez de algunos de estos datos de campo como potencial conjunto de datos de entrenamiento, comparándolos con los de un método no supervisado. Si este artículo se mantiene al nivel de un conjunto pequeño y representativo de casos de estudio individualmente procesados (seis), dicha pregunta puede verse como la premisa de una cuestión mucho mayor: la posibilidad de que sistemas grandes de inteligencia artificial utilicen estos datos botánicos abiertos para entrenar aplicaciones de teledetección a escala global.

## CASOS DE ESTUDIO, DATOS, MÉTODO

Se seleccionaron seis áreas de manglar ubicadas en Australia, Benín, Brasil, Colombia, Malasia y los EEUU (ver 1. a. iii.), de acuerdo con su densidad de datos de registro de manglar y la proximidad de otros tipos de vegetación boscosa húmeda (Figura 1). Más precisamente, se computó la densidad de puntos de

In the universe of open biodiversity data, the Global Biodiversity Information Facility (GBIF, an international network and data infrastructure funded by the world’s governments and aimed at providing anyone, anywhere with open access to data about all types of life on Earth) is a major contributor, with 1 900 000 000 records (*i.e.*, the observation of one individual at a given time and place) and 60 000 datasets, as well as 1700 publishing institutions and 6000 scientific papers using these data as of 2024.

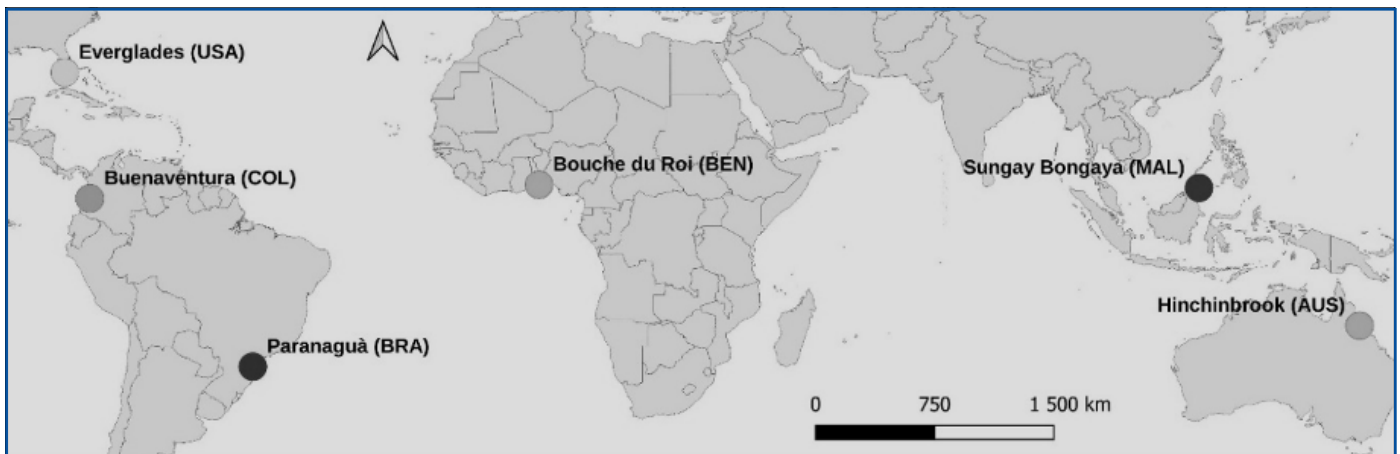
On such basis – a rare methodological analysis regarding the quality of the data used for training, within an approach focusing on open access satellites trained with (different) open-access data – this paper simply aims to perform a first assessment of the GBIF database’s efficiency in training activities related to the processing of multispectral optical satellite imagery, with a focus on accurate mangrove extension mapping. The following question is addressed: can the GBIF’s open-access botanical occurrence data enhance the accuracy of mangrove mapping via remote sensing with open-access satellites? This paper will estimate the robustness of some of these field data as a potential training dataset, comparing them to those from an unsupervised method. If this paper stays at the level of a small representative set of individually processed case studies (six), such question can be regarded as the premise of a much larger issue: the possibility for large artificial intelligence systems to use the aforementioned data to train remote sensing applications on global scale.

## STUDY CASES, DATA, METHOD

Six mangrove areas located in Australia, Benin, Brazil, Colombia, Malaysia, and the USA were selected (see 1. a. iii.) according to their mangrove occurrence data density and the proximity of other types of wooded humid vegetation (Figure 1). More precisely, each country’s point density was computed. Among the countries with high density, we discarded all arid climates (since mangroves are much easier to map when surrounded by environments with little or no vegetation). Then, we searched for mangrove areas with a very large number of points. The final selection aimed to

cada país. Entre los países con una alta densidad, descartamos todos los climas áridos (debido a que los manglares son mucho más fáciles de mapear cuando están rodeados por ambientes con poca o ninguna vegetación). Posteriormente, buscamos áreas de manglar con un gran número de puntos. La selección final buscó abarcar todos los continentes. Así, los manglares no fueron escogidos por sus puntos comunes; por el contrario, fueron escogidos para cubrir su diversidad biogeográfica (atlántico vs. indo-pacífico). Los puntos comunes fueron los climas húmedos, los bosques densos y los humedales alrededor de los manglares.

cover all continents. Thus, the mangroves were not selected for their common points but, on the contrary, while aiming to cover their biogeographical diversity (Atlantic vs. Indo-Pacific). The common points were the humid climates, dense forests, and wetlands surrounding the mangroves.



**Figura 1.** Mapa de seis casos de estudio con una alta densidad de datos de GBIF y una probable confusión espectral entre manglar y vegetación continental circundante.

**Figure 1.** Map of six study cases with a high density of GBIF data and a probable spectral confusion between mangroves and surrounding continental vegetation.

Los manglares de los Everglades pertenecen a la región atlántica (*Rhizophora mangle*, *Avicenniagerminans* y *Lagunculariaracemosa*). Se encuentran entre los manglares más septentrionales del mundo y se consideran la mayor extensión continua de bosque de manglar protegido en el hemisferio occidental (Sklar *et al.*, 2019). Los manglares de Buenaventura, en Colombia, se encuentran en la costa pacífica de Suramérica (*Rhizophora mangle*, *R. harrisonii*, *R. racemosa*, *Avicenniagerminans*, *Lagunculariaracemosa*, *Conocarpuserectus*, *Pelliceriarhizophorae* y *Mora oleifera*), y son conocidos por estar entre las copas de manglares más altas del mundo (Castellanos-Galindo *et al.*, 2014). Los manglares de Paranaguá están ubicados en la costa atlántica de Brasil (*Rhizophora mangle*, *Avicenniagerminans* y *Lagunculariaracemosa*), cerca del límite más meridional. Son un sistema estuarino que se compone de dos cuerpos de agua principales, donde el manglar está principalmente ubicado en el borde interior del sistema (Faraco *et al.*, 2010). Los manglares de Bouche du Roi (*Rhizophora mangle*

The mangroves of the Everglades belong to the Atlantic region (*Rhizophora mangle*, *Avicennia germinans*, and *Laguncularia racemosa*). They are among the northernmost mangroves worldwide and are considered to be the largest contiguous protected mangrove forest extension in the western hemisphere (Sklar *et al.*, 2019). In Colombia, the Buenaventura mangroves are located on the Pacific coast of South America (*R. mangle*, *R. harrisonii*, *R. racemosa*, *Avicennia germinans*, *L. racemosa*, *Conocarpus erectus*, *Pelliceria rhizophorae*, and *Mora oleifera*), and they are known to be among the highest mangrove canopies in the world (Castellanos-Galindo *et al.* 2014). The Paranaguá mangroves are located on the Atlantic coast of Brazil (*R. mangle*, *A. germinans*, and *L. racemosa*), close to its southernmost limit. They are an estuarine system comprising two main water bodies, with the mangrove mainly located at the inner edge of the system (Faraco *et al.* 2010). The Bouche du Roi (*R. mangle* and *A. germinans*) are situated in the southwestern part of Benin,

y *Avicenniagerminans*) están situados en la parte suroeste de Benín, en la zona costera, caracterizada por el clima guineo-congoleño, con dos temporadas lluviosas en un extenso sitio Ramsar (Zanvo *et al.*, 2021). Los manglares de Sungai Bongaya en el estado de Sabah, Malasia, están ubicados en la región indo-pacífica y albergan 28 especies verdaderas de manglares en un clima ecuatorial muy húmedo (Phua *et al.* 2008). Por último, los manglares de Hinchbrook –por tanto, en la región indo-pacífica, dominados por *Rhizophorastylosa*, *Rhizophoraapiculata* y *Ceriops australis*– son una de las áreas contiguas más extensas de bosque de manglar en Australia (Clough, 1998).

### Datos Sentinel-2

Sentinel-2, un componente del programa Copernicus de la Agencia Espacial Europea, suministra productos de teledetección con imágenes multiespectrales de alta resolución (Spoto *et al.*, 2010), *i.e.*, con una resolución espacial de 10 metros para sus bandas visible y de infrarrojo cercano y 20 metros para sus bandas de infrarrojo de onda corta. Además, el satélite tiene un tiempo de revisita de cinco días, único en su tipo, y una capacidad de cobertura global. Con 13 bandas que van desde el visible hasta el infrarrojo de onda corta, Sentinel-2 ofrece una buena resolución espectral, lo que lo hace ideal para una multitud de aplicaciones como la clasificación del uso del suelo, el monitoreo de la salud de los cultivos y la gestión forestal. Estos atributos, combinados con una política de datos abiertos y de libre acceso, han hecho de Sentinel-2 un recurso invaluable para la comunidad global de teledetección (Drusch *et al.*, 2012). Todas las imágenes escogidas (Tabla 1) son del nivel 2A, lo que significa que han sido preprocesadas adecuadamente (desde la proyección hasta la corrección atmosférica).

Tabla 1. Características de las imágenes de Sentinel-2 utilizadas.

País / Country	Caso de estudio / Case study	Fecha / Date	Referencia / Reference
Indonesia	Sungay Bongaya	04/06/2021	S2A_MSIL2A_20210604T022551_N0300_R046_T50NPM_20210604T055731
Benín	Bouche du Roi	29/12/2022	S2B_MSIL2A_20221229T101329_N0509_R022_T31NCH_20221229T125721
Colombia	Buenaventura	20/02/2022	S2A_MSIL2A_20220220T153621_N0400_R068_T18NTK_20220220T202405
EE. UU. / USA	Everglades	19/12/2022	S2A_MSIL2A_20221219T160701_N0509_R097_T17RMJ_20221219T220256
Brasil / Brazil	Paranaguá	15/11/2022	S2A_MSIL2A_20221115T132241_N0400_R038_T22JGS_20221115T183900

in the coastal zone, which is characterized by the Guineo–Congolian climate, with two rainy seasons in a large Ramsar site (Zanvoet *et al.* 2021). The mangroves of Sungai Bongaya, in the State of Sabah, Malaysia, are located in the Indo-Pacific region and harbor 28 true mangrove species in a very humid equatorial climate (Phua *et al.* 2008). Finally, the Hinchbrook mangroves – located in the Indo-Pacific region and dominated by *R.stylosa*, *R. apiculata*, and *Ceriops australis* – are one of the largest contiguous areas of mangrove forest in Australia (Clough, 1998).

### Data Sentinel-2

Sentinel-2, a component of the European Space Agency's Copernicus program, provides remote sensing products with high-resolution multispectral imagery (Spoto *et al.* 2010), *i.e.*, with a spatial resolution of 10 m for the visible and near-infrared bands and 20 m for its shortwave infrared bands. Additionally, the satellite has a unique revisit time of five days as well as a global coverage capability. With 13 bands ranging from visible to shortwave infrared, Sentinel-2 offers a good spectral resolution, making it ideal for a myriad of applications such as land cover classification, crop health monitoring, and forest management. These attributes, coupled with a freely accessible and open data policy, have made Sentinel-2 an invaluable resource for the global remote sensing community (Drusch *et al.*, 2012). All the selected images (Table 1) are in the 2A level, meaning that they have been adequately pre-processed (from projection to atmospheric correction).

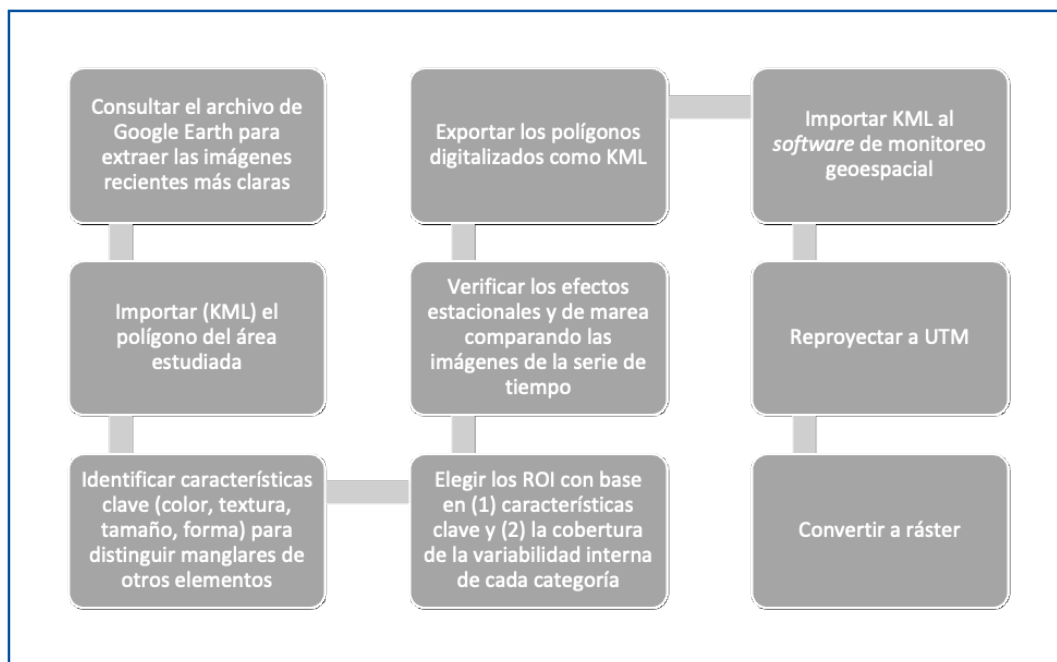
Table 1. Characteristics of the Sentinel-2 images used.

## VHRS de Google Earth para evaluar la precisión

En aras de obtener datos para la evaluación de precisión, se digitalizó un conjunto de sitios de control (polígonos en formato vectorial) con base en un proceso de interpretación visual doble-ciego de imágenes de Google Earth, llevado a cabo por dos especialistas en teledetección (que no estuvieron involucrados en el procesamiento de Sentinel-2). El enfoque elegido (Cha y Park, 2007; Dorais y Cardille, 2011) para extraer información robusta de las imágenes satelitales de Google Earth siguió los pasos descritos a continuación (Figura 2).

## Google Earth VHRS for accuracy assessment

To obtain data for the accuracy assessment, we digitized a set of control sites (polygons in vector format) based on a double-blind visual interpretation of Google Earth images carried out by two remote sensing specialists (who were not involved in Sentinel-2 processing). The approach selected (Cha and Park, 2007; Dorais and Cardille, 2011) for extracting robust information from the Google Earth satellite images followed the steps outlined below (Figure 2).



**Figura 2.** Diagrama de flujo de la interpretación visual de imágenes de Google Earth.

**Figure 2.** Flowchart of the Google Earth visual interpretation.

Google Earth brindó, para los seis sitios, varias imágenes VHR (separadas por unos cuantos meses) de buena calidad, permitiendo una clara distinción entre manglares (color verde oscuro, con textura oolítica densa y regular, usualmente rodeados por canales de agua) y otros elementos. Para cada uno de los seis sitios, una imagen reciente permitió cubrir adecuadamente el sitio, de modo que la calidad del material seleccionado para la foto-interpretación no fue debilitada por mosaicos de distintas imágenes.

Google Earth provided, for the six sites, several good-quality VHR images (which were a few months apart), enabling a clear distinction between mangroves (dark green, with dense and regular oolitic texture, usually surrounded by water channels) and other elements. For each of the six sites, a recent image allowed covering the site properly, so the quality of the material selected for photo-interpretation was not weakened by mosaics of different images.

## Datos de registro de la Global Biodiversity Information Facility

Todos los registros de especies de manglar en el mundo fueron descargados del sitio de la GBIF (alrededor de 99 000 registros

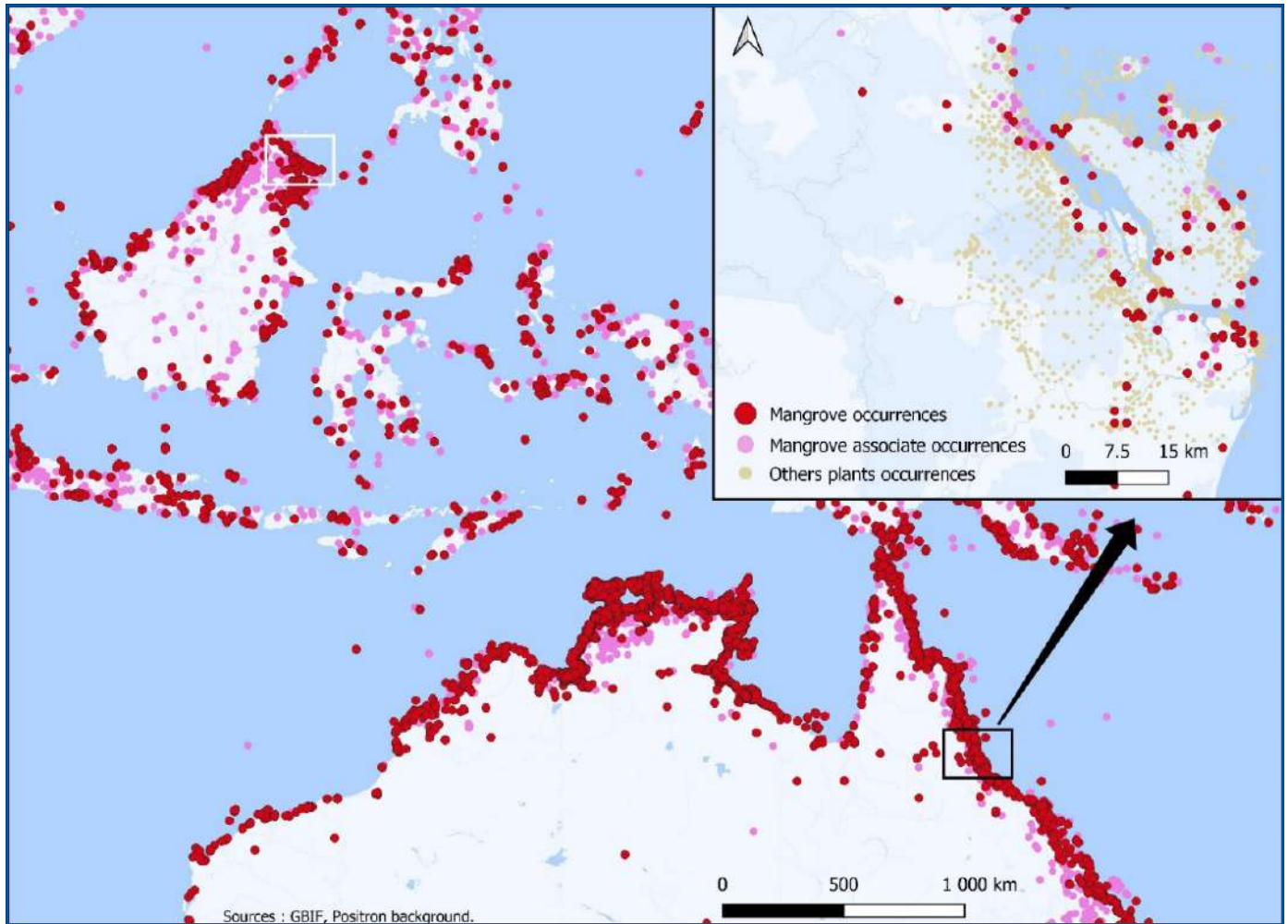
## Occurrence data from the Global Biodiversity Information Facility

All the occurrences of mangrove species in the world were downloaded from the Global Biodiversity Information Facility's



de especies de manglar verdaderas y aproximadamente 65 000 de manglar posterior o especies asociadas) (Figura 3). Este conjunto de datos tiene problemas significativos de precisión (*i.e.*, taxonomía, coordenadas). No se procedió con una curaduría completa de los datos globales. En cambio, solo se curaron los datos de los seis casos de estudio (ver sección 4.ii).

site (around 99 000 occurrences of true mangrove species and around 65 000 records of back mangroves or associated species) (Figure 3). This dataset exhibits significant accuracy issues (*i.e.*, taxonomy, coordinates). We did not proceed with a full curation of the global data. Instead, we only curated data on the six case studies (see section 4.ii).



**Figura 3.** Visión general de la distribución espacial de datos de manglar verdadero y posterior provenientes de la GBIF para un caso de estudio. Se incluyen los registros de áreas no ocupadas por manglares.

**Figure 3.** Overview of the spatial distribution of true and back mangrove data from the GBIF and for a case study. Non-mangrove occurrences are also included.

El primer paso de la curaduría fue eliminar 394 registros de estas categorías: espécimen fósil, observación automática, referencia de material y muestra de material. El segundo paso se centró en las alertas propuestas por la GBIF. Se eliminaron 749 registros con alguna de estas alertas del conjunto de datos: “precisión de coordenadas inválida”, “latitud presuntamente negada”, “longitud presuntamente negada”, “coordenada presuntamente intercambiada”, “coincidencia de taxón de rango superior”, “metros de incertidumbre de la coordenada inválidos”.

The first curation step was to remove 394 occurrences from these different categories: fossil specimen, machine observation, material citation, and material sample. The second step focused on the warnings proposed by the GBIF. 749 occurrences with one of these warnings were removed from the dataset: “coordinate precision invalid”, “presumed negated latitude”, “presumed negated longitude”, “presumed swapped coordinate”, “taxon match higherrank”, “coordinate uncertainty meters invalid”.

Seguidamente, también se descargaron, para los seis sitios, todos los datos de registros botánicos (no pertenecientes a manglares) alrededor de los manglares (5 500-26 500 registros) (Tabla 1). Se aplicó la misma curaduría, eliminando 1199 registros. Los registros de los casos de estudio se presentan en la Tabla 2.

Tabla 2. Datos de registro por caso de estudio.

País	Manglar verdadero / True mangrove	Manglar posterior / Back-Mangrove	No perteneciente a manglar / Non-mangrove
Brasil / Brazil	79	49	14 413
Malasia / Malaysia	176	119	9 455
Australia	303	199	5 546
Benin / Bénin	309	211	26 495
Colombia	2129	115	25 270
EE. UU. / USA	274	252	16 874

## Método

En efecto, tras décadas de aplicaciones de teledetección, aún se presentan imprecisiones cuando los manglares están rodeados por bosques húmedos densos u otros humedales con biomasa vegetal alta (Kuenzer *et al.*, 2011; Maurya *et al.*, 2021). Por ejemplo, Buenaventura solo fue mapeada con un índice  $\kappa$  de 0.7 por Perea-Ardila *et al.* (2019), Asbridge *et al.* (2018) cartografiaron la extensión de manglar en Hinchinbrook con un índice  $\kappa$  de 0.81 y Han *et al.* (2017) lo hicieron para los Everglades con un índice  $\kappa$  de 0.82.

En vista del objetivo de evaluar el interés y los límites de los datos de GBIF para el entrenamiento en aplicaciones de mapeo de manglares, a manera de muestra de entrenamiento, se aplicó una clasificación k-vecinos más próximos (KNN) a los píxeles que rodeaban los datos de la GBIF definidos mediante segmentación de imágenes (i). Esto se comparó con el método de K-means apilados (SKM), en representación de las técnicas basadas exclusivamente en datos radiométricos.

Se realizó un proceso doble-ciego de foto-interpretación de imágenes VHR de Google Earth (Figura 4), generando polígonos georreferenciados digitalizados que se usaron como sitios de control, *i.e.*, sitios que permiten comparar los mapas con la “realidad”. La matriz de error generada con estos sitios de control permite comparar las dos clasificaciones (Figura 4).

Afterwards, for the six sites, all the botanical (non-mangrove) occurrence data surrounding the mangroves (5500 to 26 500 occurrences) were also downloaded (Table 1). The same curation process was applied, removing 1199 occurrences. The occurrences for the case studies are presented in Table 2.

Table 2. Occurrence data by case study.

## Method

Indeed, after decades of remote sensing applications, inaccuracies are still observed when mangroves are surrounded by dense humid forests or other wetlands with high vegetation biomass (Kuenzer *et al.*, 2011; Maurya *et al.*, 2021). For example, Buenaventura was only mapped with a 0.7  $\kappa$  index by Perea-Ardila *et al.* (2019), Asbridge *et al.* (2018) mapped the mangrove extension in Hinchinbrook with a 0.81  $\kappa$  index, and Han *et al.* (2017) did the same for the Everglades with 0.82  $\kappa$  index.

As the objective was to evaluate the interest and limits of the GBIF data in training mangrove mapping applications, as a training sample, we applied k-nearest neighbor (KNN) classification to the pixels surrounding the GBIF data, as defined by image segmentation (i). This was compared against the use of a stacked k-means method (SKM) representing the techniques based only on radiometric data.

A double-blind photo-interpretation of Google Earth VHR images was carried out (Figure 4), generating digitized georeferenced polygons that were used as control sites, *i.e.*, sites that allow comparing maps to “reality”. The error matrix produced with these control sites allows comparing the two classifications (Figure 4).

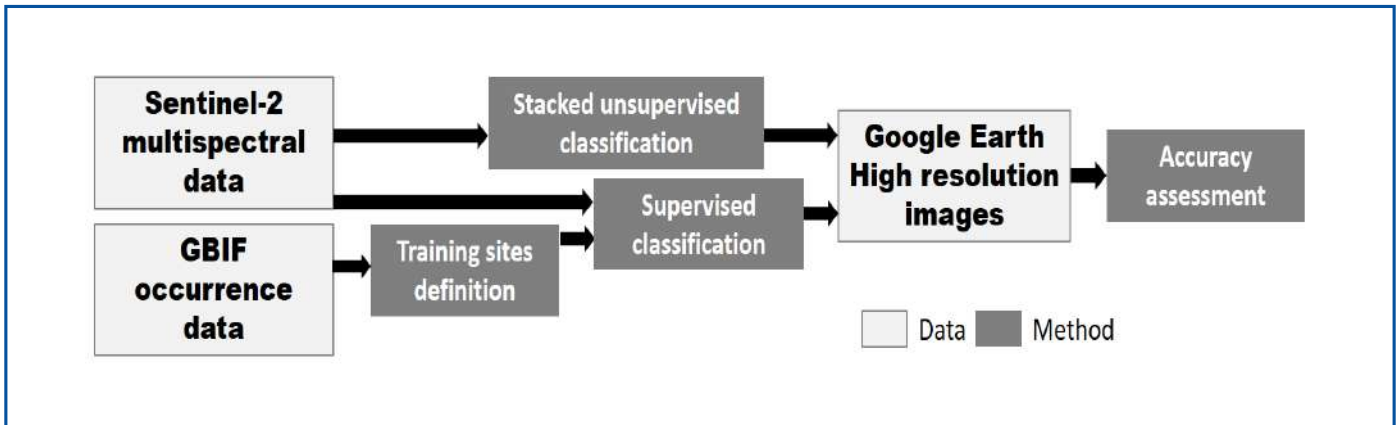


Figura 4. Diagrama de flujo del procesamiento de imágenes.

Figure 4. Image processing flowchart.

### i. Clasificación no supervisada basada en análisis espectral

El primer método se basó exclusivamente en datos espectrales del satélite seleccionado (Sentinel-2), pues los métodos no supervisados han demostrado efectividad en la teledetección (Cohen *et al.* 2002; Lunetta *et al.* 2006; Zhang *et al.* 2005; Steininger *et al.* 2001; Huang *et al.* 2009; Loveland *et al.* 2000). Las publicaciones citadas destacan colectivamente la efectividad de los métodos no supervisados en la teledetección, haciendo énfasis en su capacidad de manejar tareas complejas y ofrecer soluciones en ausencia de grandes conjuntos de datos etiquetados. El procedimiento descrito en Andrieu (2018) brinda instrucciones detalladas paso a paso con ilustraciones adjuntas. Este artículo brinda una visión general concisa del procedimiento.

Este método se basa en el análisis de curvas radiométricas e imposibilita enfoques completamente automatizados. El proceso involucra la clasificación de píxeles en 15 categorías distintas utilizando el algoritmo K-means, implementado en el *software* Terrset®. Este gran número de clases permite asignar, de manera consistente, al menos una clase a cada tipo de cobertura de suelo en la tipología deseada. Por consiguiente, se obtiene una imagen de 15 clases y se recolectan estadísticas para cada una de ellas, las cuales se sintetizan en curvas radiométricas. Mediante el escrutinio de estas curvas, se pueden condensar aún más las 15 clases, alineándolas con la nomenclatura (agua, manglar, bosque continental, uso artificial del suelo y, si aplica, llanura mareal). Este enfoque ha sido utilizado por Diedhiou *et al.* (2020), Marega *et al.* (2021), Solly *et al.* (2021) y Valdez Achucarro *et al.* (2023). Ha demostrado ser robusto (Andrieu *et al.*, 2019) en comparación con métodos más simplistas (Carney *et al.*, 2013).

### i. Stacked unsupervised classification based on spectral analysis

The first method was exclusively based on the use of spectral data from the selected satellite (Sentinel-2), as unsupervised methods have demonstrated their effectiveness in remote sensing (Cohen *et al.* 2002; Lunetta *et al.* 2006; Zhang *et al.* 2005; Steininger *et al.* 2001; Huang *et al.* 2009; Loveland *et al.* 2000). The cited publications collectively underscore the effectiveness of unsupervised methods in remote sensing, emphasizing their ability to handle complex tasks and offer solutions in the absence of large, labeled datasets. The procedure outlined in Andrieu (2018) provides detailed step-by-step instructions with accompanying illustrations. This paper provides a concise overview of the procedure.

This method hinges on the analysis of radiometric curves and precludes entirely automated approaches. The process involves the classification of pixels into 15 distinct categories, using the K-means algorithm as implemented in the Terrset® software. This large number of classes allows consistently assigning at least one class to each type of land cover in the desired typology. Consequently, a 15-class image is obtained, and statistics are collected for each of them, which are then synthesized in radiometric curves. By scrutinizing these curves, the 15 classes can be further condensed, aligning them with the nomenclature (water, mangrove, continental forest, artificial land use, and, when applicable, mudflats). This approach has been employed by Diedhiou *et al.* (2020), Marega *et al.* (2021), Solly *et al.* (2021), and Valdez Achucarro *et al.* (2023). It has proven to be robust (Andrieu *et al.* 2019) when compared to more simplistic methods (Carney *et al.* 2013).

El resultado de este paso inicial es la creación de un mapa preliminar de cobertura de suelo. Sin embargo, se debe admitir que este paso carece de la precisión deseada, lo que se debe principalmente a los retos inherentes de separar ciertos tipos de cobertura de suelo. En particular, hay distinciones espectrales limitadas entre algunos tipos, como es el caso de los manglares densos y algunas áreas boscosas en tierra seca, así como entre el suelo desnudo de color claro en tierra firme y algunas llanuras mareales secas. Para corregir esta limitación, procedimos a implementar un segundo paso que involucraba clasificaciones apiladas, con el fin de refinar progresivamente las definiciones de clase. En este proceso se generaron imágenes binarias, denominadas *máscaras* de aquí en adelante. Cada una de estas imágenes representaba una de las clases de cobertura del suelo. Para cada máscara, se realizó un análisis de componentes principales (PCA) en todos los canales espectrales, en aras de incrementar las diferencias numéricas entre píxeles de la misma clase. Acto seguido, se llevó a cabo una clasificación K-means de seis clases. Esto mejoró el contraste radiométrico, aunado a un conteo de píxeles más pequeños, facilitando la identificación de píxeles mal clasificados en el paso inicial. Se logró, entonces, reasignar estos píxeles a su respectiva clase de la nomenclatura, con base en su comportamiento radiométrico (Andrieu, 2018).

## ii. Clasificación supervisada utilizando datos de GBIF como muestra de entrenamiento

En estudios que aplican la teledetección en bosques, el algoritmo KNN es comúnmente empleado para integrar datos de campo de parcelas de muestra con datos de teledetección, con el fin de estimar variables forestales (Tomppo, 1991; Dasarathy, 1991; Franco-López *et al.*, 2001; Katila y Tomppo, 2001; Koukal, 2004). El método KNN, propuesto por Cover y Hart (1967), compara cada "píxel desconocido" que requiere estimación con "píxeles de parcela" de lotes terrestres conocidos. Se seleccionan los píxeles de la parcela k con las características espectrales más similares, y sus valores medios se utilizan como estimaciones para los píxeles desconocidos (Fazakas *et al.*, 1999; Kilki y Päivinen, 1987; McRoberts *et al.*, 2002). El método KNN es un enfoque no paramétrico que no asume ninguna distribución de píxeles en el espacio de características con respecto a variables forestales. Se necesitan números adecuados en los píxeles de parcela que corresponden a cada variable forestal para encontrar valores similares.

The outcome of this initial step is the creation of a preliminary land cover map. However, we acknowledge that this step lacks the desired precision, mainly due to the inherent challenges in separating certain land cover types. Notably, there are limited spectral distinctions between certain types, as is the case with dense mangroves and some wooded areas on dry land, as well as between light-colored bare soil on dry land and some dry mudflats. To address this limitation, we implemented a second step involving stacked classifications, with the purpose of progressively refining the class definitions. During this process, binary images were generated, hereafter referred to as *masks*. Each of these images represented one of the land cover classes. For each mask, a principal component analysis (PCA) was carried out on all spectral channels, in order to heighten the numeric differences between pixels of the same class. Subsequently, a six-class K-means classification was performed. This enhanced the radiometric contrast while also yielding a smaller pixel count, thereby facilitating the identification of misclassified pixels from the initial step. Thus, we managed to reassign these pixels to their corresponding nomenclature class based on their radiometric behavior (Andrieu, 2018).

## Supervised classification using GBIF data as a training sample

In studies applying remote sensing to forests, the KNN algorithm is commonly employed to integrate field data from sample plots with remote sensing data, in order to estimate forest variables (Tomppo, 1991; Dasarathy, 1991; Franco-Lopez *et al.*, 2001; Katila and Tomppo, 2001; Koukal, 2004). The KNN method, proposed by Cover and Hart (1967), compares each 'unknown pixel' requiring estimation against 'plot pixels' from known terrestrial plots. The k-plot pixels with the most similar spectral characteristics are selected, and their mean values are used as estimates for the unknown pixel (Fazakas *et al.*, 1999; Kilki and Päivinen, 1987; McRoberts *et al.*, 2002). The KNN method is a non-parametric approach that does not assume any distribution of pixels in the feature space with respect to forest variables. Adequate numbers in the plot pixels corresponding to each forest variable are necessary for finding similar values.

**Tabla 3.** Regiones de interés (ROI) utilizadas para entrenar el algoritmo KNN.

País / Country	ROI de manglar (en píxeles) / Mangrove ROI (in pixels)	ROI de manglar (conteo) / Mangrove ROI (count)
COL	1586	25
MAL	3956	13
USA	3763	51
BRA	824	13
BEN	1780	39
AUS	2214	30

Se utilizaron datos de la GBIF para entrenar el método KNN (**Tabla 3**). Para configurar un conjunto de datos de entrenamiento sin digitalizar manualmente las parcelas, procedimos a segmentar las imágenes (algoritmo de cuenca hidrográfica) en el *software* Terrset, donde

- Todas las bandas de Sentinel-2 tenían el mismo peso.
- La ventana móvil para calcular la capa de varianza se estableció en 3\*3.
- Se dio el mismo peso a los factores de media y de varianza.
- El factor de tolerancia de similitud se estableció en 10 para garantizar segmentos pequeños y homogéneos.

Para seleccionar segmentos para los sitios de entrenamiento con base en los datos de registro de GBIF, se añadieron cinco capas de información geográfica (**Figura 5**):

- Una capa de ráster (composición en color de Sentinel-2)
- Una capa vectorial (polígono, segmentación de imagen)
- Tres capas vectoriales (puntos; registros de manglares verdaderos y posteriores y de flora continental)

Cada registro de manglar ubicado en (o muy cerca a) un segmento cuya interpretación visual tuviera sentido se incluyó como sitio de entrenamiento. Esto puede considerarse como el último paso de la curaduría de datos. Con base este conjunto de puntos, se produjeron las ROI presentadas en la **Tabla 3**.

Los sitios de entrenamiento y todas las bandas de Sentinel-2 se movilizaron mediante un clasificador KNN.

**Table 3.** Regions of interest (ROI) used for training the KNN algorithm.

We used the GBIF data to train the KNN method (**Table 3**). To set up a training data set while avoiding the manual digitization of plots, we segmented the images (watershed algorithm) in the Terrset software, where

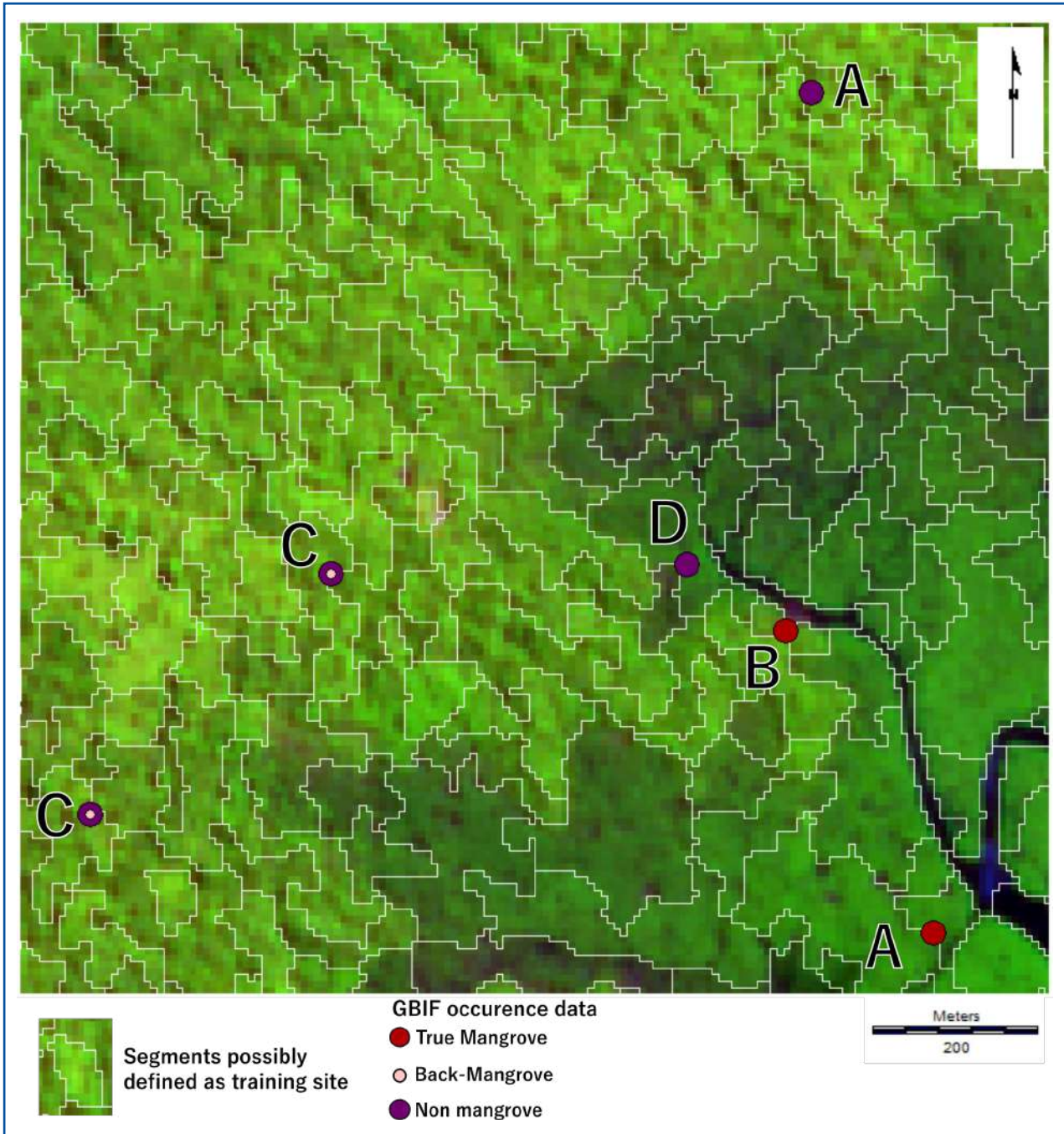
- All the bands from Sentinel-2 had the same weight.
- The moving window to compute the variance layer was set as 3\*3.
- The same weight was assigned to the mean and variance factors.
- The similarity tolerance factor was set at 10 to ensure small and homogeneous segments.

To select segments for the training sites based on the GBIF occurrence data, we added five layers of geographic information (**Figure 5**):

- One raster layer (color composite from Sentinel-2)
- One vector layer (polygon, image segmentation)
- Three vector layers (points; true mangrove, back mangrove, and continental flora occurrences)

Every mangrove occurrence located in (or very close to) a segment whose visual interpretation made sense was included as a training site. This can be regarded as the last step of the data curation process. Based on this set of points, the ROI presented in **Table 3** were produced.

The training sites and all the Sentinel-2 bands were mobilized via a KNN classifier.



**Figura 5.** Construcción del conjunto de datos de entrenamiento guiada por datos de la GBIF. A) Registro consistente con la interpretación visual; B) incertidumbre sobre la consistencia entre datos de registro y la interpretación visual; C) registro de dos tipos diferentes en la misma geolocalización; D) discrepancia entre la ocurrencia y la interpretación visual.

**Figure 5.** Training dataset construction guided by GBIF data. A) Occurrence consistent with the visual interpretation; B) uncertainty regarding the consistency between occurrence data and visual interpretation; C) occurrence of two different types in the same geolocation; D) discrepancy between occurrence and visual interpretation.

### iii. Evaluación de precisión

Se calcularon matrices de error (Darmawan *et al.* 2011) para comparar las clasificaciones y los sitios de control doble-ciego de alta resolución de Google Earth (cf. 4.a.ii). Estos sitios, para los manglares, oscilan entre los 12 641 y los 54 200 píxeles en una o más parcelas. Esto produce:

### iii. Accuracy assessment

Error matrices (Darmawan, *et al.*, 2011) were computed to compare the classifications and the Google Earth high-resolution, double-blind control sites (cf. 4.a.ii). These sites, for the mangroves, oscillate between 12641 and 54 200 pixels in one or several patches. This yields:

El índice *Kappa*, que mide el nivel de concordancia entre una clasificación observada y una esperada.

El error de omisión (falso negativo) en la clase de manglar, que representa los píxeles donde las clasificaciones no lograron reconocer alguna superficie de manglar. Esto también se conoce como *precisión de productor* o *KIA utilizando el mapa*.

El error de comisión (falso positivo) en la clase de manglar, que representa los píxeles incorrectamente atribuidos a manglares en una superficie perteneciente a otro tipo de cobertura de suelo, también denominado *precisión de usuario* o *KIA utilizando sitios de control*.

Por último, para ahondar en el análisis comparativo, se realizaron tabulaciones cruzadas de los dos resultados. Después se extrajeron los dos lotes más grandes de discrepancias (*i.e.*, manglares utilizando solamente el método supervisado y manglares utilizando solamente el método no supervisado), y se revisaron de nuevo con imágenes de alta resolución de Google Earth.

## RESULTADOS

### a. Evaluación de precisión a partir de matrices de error de interpretación visual de Google Earth

El contraste de ambos métodos con la interpretación visual de Google Earth generó algunos índices que describían los niveles de precisión: el índice *Kappa* general y, para los manglares, *KIA* utilizando el mapa y *KIA* utilizando sitios (**Tabla 4**).

**Tabla 4.** Índices principales de precisión en el mapeo de manglares, comparando el método no supervisado y el método supervisado entrenado mediante datos de la GBIF. Los números en negrita indican el mejor puntaje.

País / Country	No supervisado / Unsupervised			Supervisado, entrenado con datos de la gbif / Gbif-trained, supervised		
	Kappa	KIA Utilizando mapas / KIA using map	KIA Utilizando sitios / KIA using sites	Kappa	KIA Utilizando mapas / KIA using map	KIA Utilizando sitios / KIA using sites
EE. UU. / USA	0.85	0.96	0.99	0.85	0.98	0.99
Colombia	0.88	0.99	0.71	0.82	0.73	0.87
Brasil / Brazil	0.96	0.99	0.71	0.94	0.73	0.87
Benín / Benin	0.88	0.87	0.99	0.79	0.36	0.99
Malasia / Malaysia	0.95	0.62	0.89	0.92	0.46	0.99
Australia	0.95	0.99	0.91	0.92	0.84	0.92

The *kappa* index, which measures the level of agreement between an observed and an expected classification.

The omission error (false negative) for the mangrove class, which represents pixels where the classifications failed to recognize some mangrove surfaces. This is also known as *producer accuracy* or *KIA using map*.

The commission error (false positive) for the mangrove class, which represents pixels wrongly attributed to mangroves on a surface belonging to another land cover type. This is also called *user accuracy* or *KIA using control sites*.

Finally, to delve deeper into the comparative analysis, cross-tabulations of the two results were performed. Then, the two largest patches of discrepancies (*i.e.*, mangroves using only the supervised method and mangroves using only the unsupervised method) were extracted and once again checked using Google Earth high-resolution images.

## RESULTS

### a. Accuracy assessment based on error matrices from Google Earth visual interpretation

The comparison of both methods with the Google Earth visual interpretation generated some indices describing the levels of accuracy: the overall *kappa* index and, for the mangroves, *KIA* using map and *KIA* using sites (**Table 4**).

**Table 4.** Main accuracy indices in mangrove mapping, comparing the unsupervised method against the GBIF-trained, supervised method. Bold numbers indicate the best score.

Obviamente, ninguno de los dos métodos produjo mapas perfectos (Tabla 4). Sin embargo, a modo global, la mayoría de los resultados de ambos métodos serían aceptables para los estándares de teledetección (e.g., el USGS considera cualquier índice kappa mayor a 0.85 como aceptable).

Los índices kappa varían entre 0.85 y 0.96 para el método no supervisado y entre 0.85 y 0.94 para el supervisado y entrenado con datos GBIF. Con solo seis estudios de caso, una diferencia de 0.02 no se considera significativa. Mediante el método no supervisado, los seis sitios obtuvieron un índice kappa que no era significativamente mayor y superaba el kappa general por valores entre los 0.02 (Everglades, EEUU) y los 0.9 puntos (Bouche du Roi, Benín) (Figure 6).

Obviously, neither of the two methods produced perfect maps (Table 4). However, globally, most of the results from both methods would be acceptable according to remote sensing standards (e.g., the USGS deems any kappa index over 0.85 acceptable).

The kappa indices range from 0.85 to 0.96 for the unsupervised method and from 0.85 to 0.94 for the GBIF-trained, supervised one. With only six case studies, a 0.02 difference is not regarded as significant. By means of the unsupervised method, the six sites obtained a kappa index that was not significantly more accurate, exceeding the overall kappa by values between 0.02 points (Everglades, USA) and 0.9 points (Bouche du Roy, Benin) (Figure 6).

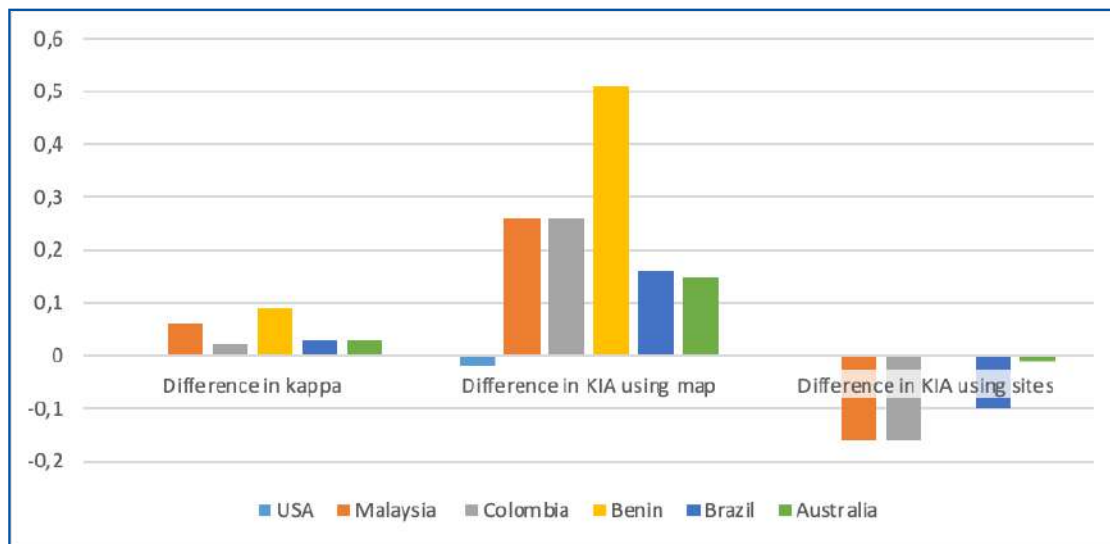


Figura 6. Diferencia de puntaje entre el método no supervisado y el supervisado y entrenado con datos GBIF. Los valores positivos representan mejores puntajes para el método no supervisado, y los valores negativos indican mejores puntajes para el supervisado y entrenado con datos GBIF.

Figure 6. Score difference between the unsupervised and the GBIF-trained, supervised methods. Positive values represent better scores for the unsupervised method, and negative values indicate better scores for the GBIF-trained, supervised one.

El sitio de los Everglades obtuvo un puntaje muy similar para ambos métodos. En todos los otros sitios, el SKM obtuvo puntajes particularmente buenos en la métrica de precisión de usuario, mientras que el KNN se desempeñó ligeramente mejor en cuanto a la precisión de productor de la clase de manglar para tres sitios (Malasia, Colombia y Brasil).

El KIA utilizando mapas del método no supervisado es bajo en el caso de Malasia, medio para Benín y excelente para los otros cuatro sitios, mientras que su KIA utilizando sitios es medio para Colombia y Brasil, bueno para Malasia y Australia, y excelente para Benín y los EEUU.

The Everglades site obtained a very similar score for both methods. In all other sites, the SKM scored particularly well in the user accuracy metric, while KNN scored slightly better regarding the producer accuracy of the mangrove class for three sites (Malaysia, Colombia, and Brazil).

The KIA using maps of the unsupervised method is low for Malaysia, medium for Benin, and excellent for the other four sites, whereas its KIA using sites is medium for Colombia and Brazil, good for Malaysia and Australia, and excellent for Benin and the USA.



El KIA utilizando mapas con el método no supervisado es muy bajo para Benín y Malasia, bajo para Colombia y Brasil, bueno para Australia y excelente para los EEUU, mientras que su KIA utilizando mapas es medio para Colombia y Brasil, bueno para Australia y excelente para los otros cuatro sitios.

**b. Evaluación de precisión a partir de mapas de tabulación cruzada**

La interpretación es consistente con la evaluación de precisión cuando se compara con la de las imágenes de alta resolución de Google Earth. La mayoría de las discrepancias entre los dos métodos se deben a limitaciones del método supervisado y entrenado con datos GBIF, exceptuando a Malasia, donde las dos parcelas (muy pequeñas) sí son de manglar. La mayoría de las imprecisiones del método KNN son errores de omisión de manglares que fueron mapeados adecuadamente por el método SKM. Más de la mitad de las parcelas mapeadas como cobertura de manglar por el método no supervisado solo son manglares según la interpretación de imágenes de Google Earth (Tabla 5). Entre las imprecisiones del método no supervisado, principalmente se encuentran parcelas muy pequeñas de errores de omisión, o lotes de errores de comisión encontrados de manera similar con ambos métodos (por ejemplo, en el caso del sitio en Brasil; ver Figura 7).

**Tabla 5.** Interpretación de Google Earth para las dos parcelas principales que fueron mapeadas como manglar por un solo método.

Parcela / Patch	Manglar en SKM únicamente / Mangrove in SKM only		Manglar en KNN únicamente / Mangrove in KNN only	
	Tamaño de las dos parcelas más grandes / Size of the two largest patches (ha)	Interpretación / Interpretation	Tamaño de las dos parcelas más grandes / Size of the two largest patches (ha)	Interpretación / Interpretation
Parcela 1 de Brasil / Brazil patch 1	115	Bosque continental / Continental Forest	513	Bosque continental / Continental forest
Parcela 1 de Brasil / Brazil patch 1		Manglar / Mangrove		Bosque continental / Continental forest
Parcela 1 de Colombia / Colombia patch 1	90	Manglar / Mangrove	411	Bosque continental / Continental forest
Parcela 2 de Colombia / Colombia patch 2		Manglar / Mangrove		Sin datos / No data
Parcela 1 de EE. UU. / USA patch 1	231	Manglar / Mangrove	29	Manglar / Mangrove
Parcela 2 de EE. UU. / USA patch 2		Manglar / Mangrove		Humedades continentales / Continental wetlands
Parcela 1 de Benín / Benin patch 1	49	Manglar / Mangrove	78	Mixta / Mixed
Parcela 2 de Benín / Benin patch 2		Humedales continentales / Continental wetlands		Mixta / Mixed

The KIA using maps of the unsupervised method is very low for Benin and Malaysia, low for Colombia and Brazil, good for Australia, and excellent for the USA, while its KIA using maps is medium for Colombia and Brazil, good for Australia, and excellent for the other four sites.

**b. Accuracy assessment from cross-tabulation maps**

The interpretation is consistent with the accuracy assessment when compared to that of the Google Earth high-resolution images. Most of the discrepancies between the two methods are due to shortcomings of the GBIF-trained, supervised method, except for Malaysia, where the two patches are indeed mangroves. Most of the inaccuracies of the KNN method are omission errors involving mangroves that were properly mapped by the SKM method. More than half of the patches mapped as mangrove cover by the unsupervised method are only mangroves according to the Google Earth interpretation (Table 5). Among the inaccuracies of the unsupervised method, very small patches of omission errors are mainly found, or patches of commission errors found similarly with both methods (for example, in the case of the Brazilian site; see Figure 7).

**Table 5.** Google Earth visual interpretation of the two main patches mapped as mangroves by only one method.

Parcela / Patch	Manglar en SKM únicamente / Mangrove in SKM only		Manglar en KNN únicamente / Mangrove in KNN only	
	Tamaño de las dos parcelas más grandes / Size of the two largest patches (ha)	Interpretación / Interpretation	Tamaño de las dos parcelas más grandes / Size of the two largest patches (ha)	Interpretación / Interpretation
Parcela 1 de Malasia / Malaysia Patch 1	16	Bosque continental / Continental forest	34	Manglar / Mangrove
Parcela 2 de Malasia / Malaysia Patch 2		Bosque continental / Continental forest		Manglar / Mangrove
Parcela 1 de Australia / Australia Patch 1	3	Balsas de tratamiento de agua / Water treatment basins	367	Bosque continental / Continental forest
Parcela 2 de Australia / Australia Patch 2		Manglar / Mangrove		Bosque continental / Continental forest

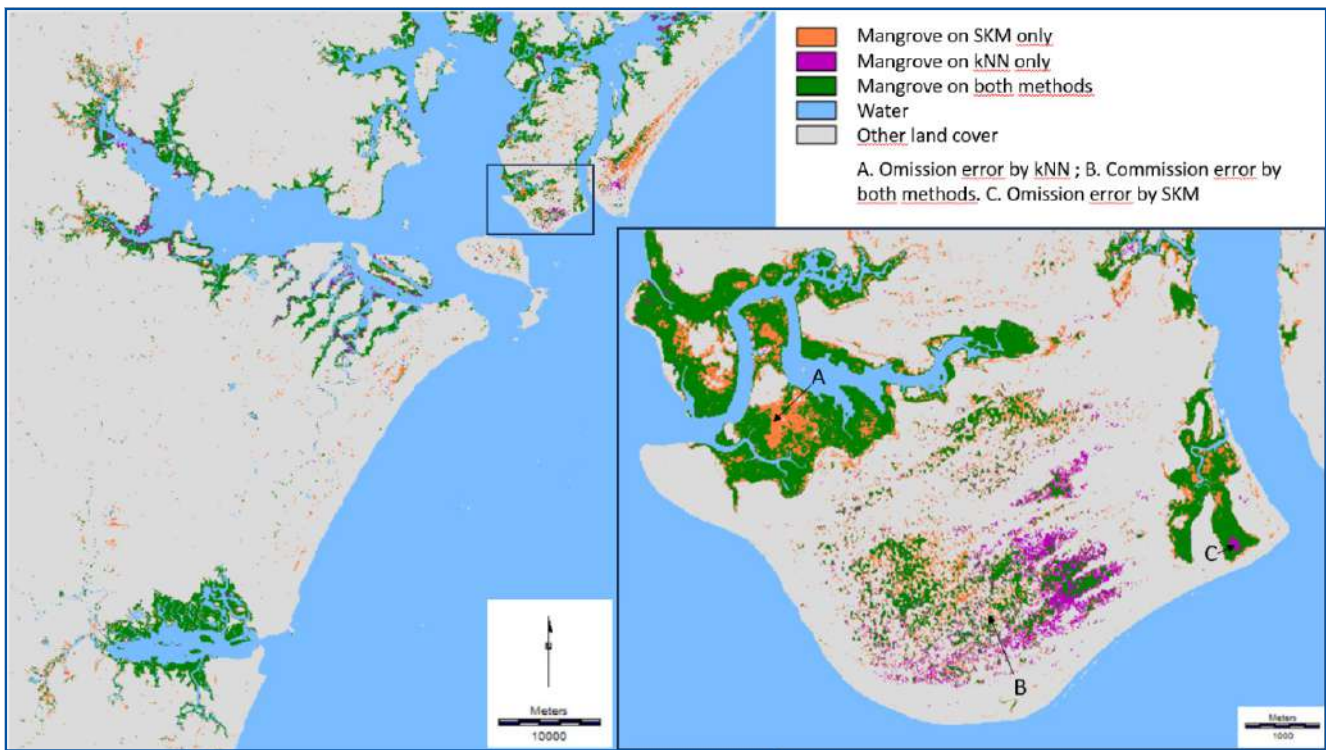


Figura 7. Mapa de la tabulación cruzada entre los dos métodos.

Figure 7. Map of the cross-tabulation between the two methods.

Los resultados de la comparación entre ambos métodos y las imágenes de alta resolución de Google Earth muestran que la GBIF, para los seis sitios seleccionados, dada su alta densidad de datos disponibles (con índices *kappa* entre 0.82 y 0.94), sí constituye un conjunto de entrenamiento satisfactorio que ayuda a elaborar mapas muy precisos. Para los investigadores que busquen reducir las imprecisiones tanto como sea posible, podríamos anotar que la clasificación simple, basada en radiometría, supervisada y apilada obtuvo un puntaje ligeramente mejor. La interpretación visual de las discrepancias entre ambos métodos también confirma que el método no supervisado es un poco más preciso.

The results of the comparison between both methods and the high-resolution Google Earth imagery show that the GBIF, for the six sites selected due to their very high density of available data (with *kappa* indices between 0.82 and 0.94), does constitute a satisfying training dataset that helps to elaborate very accurate maps. For researchers aiming to reduce inaccuracies as much as possible, we might note that the simple, radiometry-based, unsupervised, and stacked classification scored slightly better. The visual interpretation of the discrepancies between both methods also confirms that the unsupervised method is slightly more accurate.

## DISCUSIÓN

### a. Sobre la limitación metodológica de esta evaluación

El mundo es extenso y diverso, como los manglares. Los seis casos de estudio fueron cuidadosamente seleccionados para evaluar la hipótesis de que la GBIF constituye un buen conjunto de datos de entrenamiento. Sin embargo, los casos de estudio, en efecto, son solo seis y no presentan un conjunto homogéneo de resultados. Podría considerarse que este número de casos de estudio no es suficiente para evaluar la hipótesis a cabalidad. Estos casos presentan la densidad de datos más alta de especies de manglar verdaderas, por lo que este conjunto de datos es suficientemente bueno para elaborar mapas precisos, aun siendo un poco menos preciso que los métodos no supervisados.

Al comparar un método supervisado y uno no supervisado, en efecto comparamos dos algoritmos diferentes (Green, 1998). Parte de la diferencia en la precisión de los mapas evaluados puede provenir de la selección de algoritmos (Pham, 2019; Maurya, 2020). Podría uno preguntarse si los resultados serían los mismos si KNN hubiera sido reemplazado por un bosque aleatorio, o si el método K-means no estuviera apilado, o si se hubiera empleado el algoritmo Isodata en su lugar. Consideramos que K-means y KNN tenían características en común (son iterativos y se basan en la distancia euclidiana) y que habían demostrado su exactitud en su propia categoría. Obviamente, más pruebas siempre son bienvenidas en la investigación futura sobre el mapeo de manglares, pero este estudio no se trata de probar todos los algoritmos posibles, solo una hipótesis sobre los datos de la GBIF.

La interpretación visual de las imágenes de alta resolución de Google Earth era la mejor opción para construir un conjunto de datos de control para los seis casos de estudio. Esto se hizo con mucho cuidado, pero no se puede garantizar una precisión perfecta, por lo cual estos resultados deben tomarse con tanta cautela como aquella interpretación.

### b. Sobre el potencial de la base de datos de registro de GBIF como un conjunto de datos para la teledetección

Los mapas generados mediante la clasificación supervisada y entrenada con datos de GBIF obtuvieron índices *kappa* generales de entre 0.82 y 0.94. Estos puntajes de evaluación de exactitud son satisfactorios.

## DISCUSSION

### a. On the methodological limitation of this assessment

The world is wide and diverse, as are mangroves. The six case studies were carefully selected to test the hypothesis that the GBIF constitutes a good training dataset. However, there are indeed only six case studies that do not provide a homogeneous set of results. One might consider that this number of case studies is not enough to fully test the hypothesis. These cases exhibit the highest density of GBIF data on true mangrove species, which is why this dataset is good enough for accurate mapping, even though it is slightly less accurate than unsupervised methods.

By comparing a supervised and an unsupervised method, we indeed compare two different algorithms (Green, 1998). Part of the difference in the accuracy of the tested maps might stem from the choice of algorithms (Pham, 2019; Maurya, 2020). One might wonder if the results would be the same if KNN had been replaced by a random forest, or if K-means had not been stacked, or if the Isodata algorithm had been used instead of K-means. We considered that K-means and KNN had common characteristics (they are iterative and based on Euclidian distance), and that they had proven their accuracy in their own category. Obviously, more tests are most welcome in future research on mangrove mapping, but this study is not about testing all possible algorithms; it only tests a hypothesis on GBIF data.

The visual interpretation of high-resolution Google Earth imagery was the best option to build a control dataset for the six case studies. This was done with utmost care but does not ensure perfect accuracy, so these results must be taken with as much caution as those of the aforementioned interpretation.

### b. On the potential of the GBIF occurrences database as a training dataset for remote sensing

The maps produced by the GBIF-trained and supervised classification obtained overall *kappa* indices ranging from 0.82 to 0.94. These accuracy assessment scores are satisfactory.

A modo de comparación, Asbridge *et al.* (2018) mapearon la extensión de manglar en Hinchinbrook mediante el método de máxima verosimilitud, con un índice *kappa* de 0.81, mientras que el KNN entrenado con datos de GBIF obtuvo 0.92. En Buenaventura, para un área mapeada más pequeña, Perea-Ardila *et al.* (2019) obtuvieron un *kappa* de 0.7, mientras que nuestro método obtuvo 0.82. En Malasia, en una zona más extensa que la de nuestro estudio, Omar *et al.* (2018) realizaron una serie de clasificaciones locales de bosque aleatorio, obteniendo valores de *kappa* entre 0.81 y 0.91; nosotros obtuvimos 0.92. En Paranaguá, Lopes *et al.* (2023) emplearon datos multisensor en una clasificación orientada a objetos, generando un mapa de manglares con un índice *kappa* de 0.86. En cambio, nuestra clasificación alcanzó 0.92 con el KNN entrenado con datos de GBIF. En los Everglades, Han *et al.* (2017) utilizaron una clasificación basada en espectroscopía, multisensores y desmezcla de píxeles, logrando un índice *kappa* de 0.82, en comparación con nuestro 0.85. En Bouche du Roi, Zanzo *et al.* (2021) mapearon manglares mediante el método de máxima verosimilitud, con base en una intensa campaña de campo para seleccionar ROI. Estos autores obtuvieron 0.92, y este parece ser el único artículo cuya precisión es mejor que la nuestra, la cual fue obtenida con datos de la GBIF (0.79).

La GBIF es, por tanto, una base de datos satisfactoria, como lo demostró este análisis tras dos pasos de curaduría, uno basado en metadatos de la GBIF y otro en interpretación visual. El presente artículo, que constituye la primera evaluación del potencial de este conjunto de datos de GBIF para los propósitos mencionados, invita a aquellos dedicados a la teledetección que prefieren los enfoques supervisados a considerar GBIF para complementar sus datos cuando durante la construcción de sitios de entrenamiento.

## CONCLUSIÓN

La GBIF, con cerca de 100 000 registros de especies verdaderas de manglares en el mundo en 2023, necesita un importante proceso de curación para proporcionar a la comunidad de teledetección una serie de datos puntuales que no sea muy densa (y no esté muy dispersa en el espacio). En seis casos de estudio elegidos por la densidad de dichos datos y la probabilidad de confusión entre la cobertura de manglar y otra vegetación continental, se obtuvo una buena puntuación de precisión. Mientras tanto, la clasificación apilada no supervisada, basada únicamente en datos espectrales, puntuó igual de bien (más exactamente, no

For comparison, Asbridge *et al.* (2018) mapped the mangrove extension in Hinchinbrook via the maximum likelihood method, with a 0.81 *kappa* index, while the GBIF-trained KNN scored 0.92. In Buenaventura, for a smaller mapped area, Perea-Ardila *et al.* (2019) obtained a *kappa* of 0.7, whereas our method obtained 0.82. In Malaysia, in a larger zone than that of our study, Omar *et al.* (2018) performed a series of local random forest classifications, with *kappa* values between 0.81 and 0.91. In contrast, we obtained 0.92. In Paranaguá, Lopes *et al.* (2023) employed multi-sensor data within an object-oriented classification, generating a mangrove map with a 0.86 *kappa*, while our classification reached 0.92 with the GBIF-trained KNN. In the Everglades, Han *et al.* (2017) used a spectroscopy, multi-sensor, and pixel unmixing-based supervised classification, achieving a 0.82 *kappa* index, in comparison with our 0.85. In Bouche du Roi, Zanzo *et al.* (2021) mapped mangroves using the maximum likelihood method, based on an intense field campaign for selecting ROI. These authors obtained 0.92, and it appears that this is the only paper whose accuracy is better than ours, which was obtained with GBIF data (0.79).

Therefore, the GBIF is a satisfactory database, as proven by this analysis after two curation steps, one based on GBIF metadata and one based on visual interpretation. This paper, which constitutes the first assessment of the potential of this dataset for the aforementioned purposes, invites those dedicated to remote sensing who prefer supervised approaches to consider the GBIF as a complement to their other data when building training sites.

## CONCLUSION

The GBIF, with nearly 100 000 occurrences of true mangrove species worldwide as of 2023, necessitates a major curation process to provide the remote sensing community with series of punctual data that is not too dense (and not too sparsely distributed in space). In six case studies selected for the density of such data and the probability of confusion between mangrove cover and other types of continental vegetation, a good accuracy score was obtained. In the meantime, the unsupervised stacked classification, which was only based on spectral data, scored equally well (more precisely, not significantly better), so such approaches should not be discarded by the community dedicated to mapping mangroves with satellite imagery.

significativamente mejor), por lo que tales enfoques tampoco deberían ser descartados por la comunidad que cartografía la cobertura de manglar con imágenes satelitales.

Volviendo al valor y los límites de la GBIF como conjunto de datos de entrenamiento para mapear la extensión de manglares, las distintas pruebas realizadas demostraron que estos datos son apropiados en los sitios de estudio. La curaduría fue importante y, al final de ella, no eran tantos los registros útiles. Sin embargo, incluso reducida a un poco más de 10 ROI y a cubrir menos de 1000 píxeles, esta base de datos demostró ser eficiente, pero es preciso recordar que hay muchos manglares que no tienen registros en la base de la GBIF. Por lo tanto, invitamos a los académicos a publicar muchos más datos de manglares en ella para investigaciones futuras, que necesitarán más datos abiertos. Esto solo mejorará lo que se ha evaluado como un buen conjunto de datos.

As for the value and limits of the GBIF as a training dataset for mapping mangrove cover, the various tests conducted demonstrated that these data are appropriate for the studied sites. The curation process was important, and, at the end of it, there were not so many useful occurrences. Nevertheless, even when reduced to a little more than 10 ROI and covering less than 1000 pixels, this database proved to be efficient. Still, it should be noted that there are many mangroves with no occurrences in the GBIF database, so we call upon scholars to publish much more data on mangroves in it for future research, which will need more open access data. This will only improve that which has been shown to be a good dataset.

## BIBLIOGRAFÍA / LITERATURE CITED

- Andrieu, J. 2018. Land cover changes on the West-African coastline from the Saloum Delta (Senegal) to Rio Geba (Guinea-Bissau) between 1979 and 2015. *Eur. J. Remote Sens.*, 51(1): 314-325. <https://doi.org/10.1080/22797254.2018.1432295>
- Andrieu, J., M.C. Cormier-Salem, L. Descroix, T. Sané, E. Dièye and N. Ndour. 2014. Correctly assessing forest change in a priority West African mangrove ecosystem: 1986-2010 An answer to Carney *et al.*, paper "Assessing forest change in a priority West African mangrove ecosystem: 1986-2010". *Remote Sens. Appl.: Soc. Environ.*, 13. <https://doi.org/10.1016/j.rsase.2018.12.001>
- Asbridge, E., R. Lucas, K. Rogers and A. Accad. 2018. The extent of mangrove change and potential for recovery following severe Tropical Cyclone Yasi, Hinchinbrook Island, Queensland, Australia. *Ecol. Evol.* vol:8(21). P. 10416–10434. <https://doi.org/10.1002/ece3.4485>
- Bihanta Toosi, N., A.R. Soffianian, S. Fakheran, S. Pourmanafi, C. Ginzler and L. T. Waser. 2020. Land Cover Classification in Mangrove Ecosystems Based on VHR Satellite Data and Machine Learning—An Upscaling Approach. *Remote Sens.* vol:12(17). p. 2684. <https://doi.org/10.3390/rs12172684>
- Biswas, H., K. Zhang, M. S. Ross and D. Gann. 2020. Delineation of Tree Patches in a Mangrove-Marsh Transition Zone by Watershed Segmentation of Aerial Photographs. *Remote Sens.* vol:12(13). p. 2086. <https://doi.org/10.3390/rs12132086>
- Castellanos-Galindo, G., J. Cantera, U. Saint-Paul and D. Ferrol-Schulte. 2014. Threats to mangrove social-ecological systems in the most luxuriant coastal forests of the Neotropics. *Biodiversity and Conservation*. vol:24. p.4. <https://doi.org/10.1007/s10531-014-0827-y>
- Cha, S.Y. and C.H. Park. 2007. The Utilization of Google Earth Images as Reference Data for The Multitemporal Land Cover Classification with MODIS Data of North Korea. *Korean J. Remote Sens.* vol:23(5). p. 483–491. <https://doi.org/10.7780/kjrs.2007.23.5.483>
- Cohen, W.B., T. A. Spies, R. J. Alig, D. R. Oetter, T. K. Maier-Sperger and M. Fiorella. 2002. Characterizing 23 Years (1972–95) of Stand Replacement Disturbance in Western Oregon Forests with Landsat Imagery. *Ecosyst. vol:5(2)*. p. 122–137. <https://doi.org/10.1007/s10021-001-0060-X>
- Darmawan, S., D. K. Sari, K. Wikantika, A. Tridawati, R. Hernawati and M. K. Sedu. 2020. Identification before-after Forest Fire and Prediction of Mangrove Forest Based on Markov-Cellular Automata in Part of Sembilang National Park, Banyuasin, South Sumatra, Indonesia. *Remote Sens.* vol:12(22). p. 3700. <https://doi.org/10.3390/rs12223700>
- Diédhiou, I., C. Mering, O. Sy and T. Sané. 2020. Cartographie par télédétection l'occupation du sol et ses changements: Application à l'analyse de la dynamique des paysages forestiers sénégalais entre 1972 et 2016. *Echogeo*. vol:54. <https://doi.org/10.4000/echogeo.20510>
- Dorais, A. and J. Cardille. 2011. Strategies for Incorporating High-Resolution Google Earth Databases to Guide and Validate Classifications: Understanding Deforestation in Borneo. *Remote Sens.* vol:3(6). <https://doi.org/10.3390/rs3061157>
- Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Labertini, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese and P. Bargelini. 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* vol:120. pp. 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- Faraco, L., J. Andriuguetto-Filho and P. Lana. 2010. Methodology for Assessing the vulnerability of mangroves and fisherfolk to climate change. *Pan-Am. J. Aquat. Sci.* vol:5. pp. 205–223
- Franco-Lopez, H., A.R. Ek and M.E. Bauer. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens. Environ.* vol:77(3). pp. 251–274. [https://doi.org/10.1016/S0034-4257\(01\)00209-7](https://doi.org/10.1016/S0034-4257(01)00209-7)

- Giri, C. 2016. Observation and Monitoring of Mangrove Forests Using Remote Sensing: Opportunities and Challenges. *Remote Sens.* vol:8(9). p. 783. <https://doi.org/10.3390/rs8090783>
- Green, E.P., C.D. Clark, P.J. Mumby, A.J. Edwards and A.C. Ellis. 1988. Remote sensing techniques for mangrove mapping. *Int. J. Remote Sens.* vol:19(5). pp. 935–956. <https://doi.org/10.1080/014311698215801>
- Han, X., L. Feng, C. Hu and P. Kramer. 2018. Hurricane-Induced Changes in the Everglades National Park Mangrove Forest: Landsat Observations Between 1985 and 2017. *J. Geophys. Res.* vol:123(11). pp. 3470–3488. <https://doi.org/10.1029/2018JG004501>
- Heumann, B. W. 2011. Satellite remote sensing of mangrove forests: Recent advances and future opportunities. *Progress in Physical Geography: Earth and Environment.* vol:35(1). pp. 87–108. <https://doi.org/10.1177/0309133310385371>
- Hu, T., Y. Zhang, Y. Su, Y. Zheng, G. Lin and Q. Guo. 2020. Mapping the Global Mangrove Forest Aboveground Biomass Using Multisource Remote Sensing Data. *Remote Sens.* vol:12(10). p. 1690. <https://doi.org/10.3390/rs12101690>
- Huang C., S. Kimb, K. Song, J.R.G. Townshend, P. Davis, A. Altstatt, O. Rodas, A. Yanosky, R. Clay, C.J. Tucker, J. Musinsky. 2009. Assessment of Paraguay's forest cover change using Landsat observations. *Glob. Planet. Change.* vol:67(1). pp. 1–12. <https://doi.org/10.1016/j.gloplacha.2008.12.009>
- Katila, M. and E. Tomppo. 2001. Selecting estimation parameters for the finnish multisource national forest inventory. *Remote Sens. Environ.* vol:76(1). pp. 16–32. [https://doi.org/10.1016/S0034-4257\(00\)00188-7](https://doi.org/10.1016/S0034-4257(00)00188-7)
- Kuenzer, C., A. Bluemel, S. Gebhardt, T. V. Quoc and S. Dech. 2011. Remote Sensing of Mangrove Ecosystems: A Review. *Remote Sens.* vol:3(5). pp. 878–928. <https://doi.org/10.3390/rs3050878>
- Lombard, F. and J. Andrieu. 2021. Mapping Mangrove Zonation Changes in Senegal with Landsat Imagery Using an OBIA Approach Combined with Linear Spectral Unmixing. *Remote Sens.* vol:13(10). p. 1961. <https://doi.org/10.3390/rs13101961>
- Lombard, F., Soumaré, S., Andrieu, J. and D. Josselin. 2023. Mangrove zonation mapping in West Africa, at 10-m resolution, optimized for inter-annual monitoring. *Ecol. Inform. P.* 102027. Vol:75. <https://doi.org/10.1016/j.ecoinf.2023.102027>
- Lopes, J.P.N., W.R. Nascimento Jr, C.G. Diniz, and P.W.M. Souza-Filho. 2023. Mangrove changes over the past decade in South and Southeast Brazil using spaceborne optical and SAR imagery. *An. Acad. Bras. Ciênc.* vol:95. p. e20201533. <https://doi.org/10.1590/0001-3765202320201533>
- Loveland, T. R., B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L. Yang and J.W. Merchant. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* vol:21(6–7). pp. 1303–1330. <https://doi.org/10.1080/014311600210191>
- Lunetta, R.S., J. F. Knight, J. Ediriwickrema, J. G. Lyon and L. D. Worthy. 2006. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote Sens. Environ.* vol:105(2). pp. 142–154. <https://doi.org/10.1016/j.rse.2006.06.018>
- Marega, O., J.-L. San Emeterio, A. Fall and J. Andrieu. 2021. Cartographie par télédétection des variations spatio-temporelles de la couverture végétale spontanée face à la variabilité pluviométrique au Sahel: approche multiscale. *physio-geo.* Vol:16. pp. 1–28. <https://doi.org/10.4000/physio-geo.11977>
- Maurya, K., S. Mahajan and N. Chaube. 2021. Remote sensing techniques: mapping and monitoring of mangrove ecosystem—a review. *Complex Intell. Syst.* vol:7(6). pp. 2797–2818. <https://doi.org/10.1007/s40747-021-00457-z>
- McRoberts, R.E., M.D. Nelson and D.G. Wendt. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. *Remote Sens. Environ.* vol:82(2–3). pp. 457–468. [https://doi.org/10.1016/S0034-4257\(02\)00064-0](https://doi.org/10.1016/S0034-4257(02)00064-0)
- Nababa, I.I., E. Symeonakis, S. Koukoulas, T. P. Higginbottom, G. Cavan and S. Marsden. 2020. Land Cover Dynamics and Mangrove Degradation in the Niger Delta Region. *Remote Sens.* vol:12(21). p. 3619. <https://doi.org/10.3390/rs12213619>
- Omar, H., M. A. Misman and V. Linggok. 2018. Characterizing and monitoring of mangroves in Malaysia using Landsat-based spatial-spectral variability. *IOP Conf. Ser.: Earth Environ. Sci.* vol:169(1). p. 012037, <https://doi.org/10.1088/1755-1315/169/1/012037>
- Perea-Ardila, M.A., F. Oviedo-Barrero and J. Leal-Villamil. 2019. Mangrove forest mapping through remote sensing imagery: study case for Buenaventura, Colombia. *Revista de Teledetección.* vol: 53. pp. 73–86. <https://doi.org/10.4995/raet.2019.11684>
- Pham, T.D., N.N. Le, N.T. Ha, L.V. Nguyen, J. Xia, N. Yokoya, T.T. To, H.X. Trinh, L.Q. Kieu and W. Takeuchi. 2020. Estimating Mangrove Above-Ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve, Vietnam. *Remote Sens.* vol:12(5). p.777. <https://doi.org/10.3390/rs12050777>
- Pham, T.D., N. Yokoya, D.T. Bui, K. Yoshino and D.A. Friess. 2019. Remote Sensing Approaches for Monitoring Mangrove Species, Structure and Biomass: Opportunities and Challenges. *Remote Sens.* vol:11(3). p. 230. <https://doi.org/10.3390/rs11030230>
- Phua, M.H., O. Conrad, K.U. Kamlun, M. Ficher and J. Böhner. Multitemporal fragmentation analysis of peat swamp forest in the Klias Peninsula, Sabah, Malaysia using GIS and remote sensing techniques. 2007. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie.* p. 81–90.
- Purnamasayangasukasih, P.R., K. Norizah, A.A.M. Ismail and I. Shamsudin. 2016. A review of uses of satellite imagery in monitoring mangrove forests. *IOP Conf. Ser. Earth Environ. Sci.* vol:37(1). p. 012034. <https://doi.org/10.1088/1755-1315/37/1/012034>
- Quang, N.H., C.H. Quinn, L.C. Stringer, R. Carrie, C.R. Hackney, L.T. Van Hue, D. Van Tan and P.T.T. Nga. 2020. Multi-Decadal Changes in Mangrove Extent, Age and Species in the Red River Estuaries of Viet Nam. *Remote Sens.* vol:12(14). p. 2289. <https://doi.org/10.3390/rs12142289>
- Sklar, F.H., J.F. Meeder, T.G. Troxler, T. Dreschel, S. E. Davis and P.L. Ruiz. 2019. The Everglades: At the Forefront of Transition. in *Coasts and Estuaries.* Elsevier. pp. 277–292. <https://doi.org/10.1016/B978-0-12-814003-1.00016-2>
- Solly, B., E.H.B. Dieye, O. Sy, A.M. Jarju and S. Tidiane. 2021. Détection des zones de dégradation et de régénération de la couverture végétale dans le sud du Sénégal à travers l'analyse des tendances de séries temporelles MODIS NDVI et des changements d'occupation des sols à partir d'images LANDSAT. *Revue Française de Photogrammétrie et de Télédétection.* vol:223(1). <https://doi.org/10.52638/rfpt.2021.580>
- Spoto, F., O. Sy, P. Laberinti, P. Martimort, V. Fernandez, O. Colin, B. Hoersch, A. Meygret. 2012. Overview Of Sentinel-2. in 2012 *Int. Geosci. Remote Sens. Symp.* pp. 1707–1710. <https://doi.org/10.1109/IGARSS.2012.6351195>

- Steininger, M.K., C.J. Tucker, J.R.G. Townshend, T.J. Killeen, A. Desch, V. Bell and P. Ersts. 2001. Tropical deforestation in the Bolivian Amazon. *Environ. Conserv.* vol:28(2). pp. 127–134. <https://doi.org/10.1017/S0376892901000133>
- Thakur, S., I. Mondal, P.B. Ghosh, P. Das and T.K. De. 2020. A review of the application of multispectral remote sensing in the study of mangrove ecosystems with special emphasis on image processing techniques. *Spat. Inf. Res.* vol:28(1). pp. 39–51. <https://doi.org/10.1007/s41324-019-00268-y>
- Tomppo, E. and M. Katila. 1991. Satellite image-based national forest inventory of Finland for publication. in the igarss'91 digest," [Proceedings] IGARSS'91 Remote Sensing: Global Monitoring for Earth Management, Espoo, Finland, 1991, pp. 1141-1144, <https://doi.org/10.1109/IGARSS.1991.579272>
- Valdez-Achucarro, I.C., J. Andrieu and S. Bouissou. 2023. Land cover dynamics in the Paraguayan Pantanal in the timeframe 1987–2020. *Reg Environ Change*, vol:23. no.( 1). p. 11. <https://doi.org/10.1007/s10113-022-01990-4>
- Wang L., M. Jia M., D. Yin D. and J. Tian. J. 2019. A review of remote sensing for mangrove forests: 1956–2018. *Remote Sens. Environ.* vol. :231,. p. 111223. <https://doi.org/10.1016/j.rse.2019.111223>
- Yancho, J.M.M., T.G. Jones, S.R. Gandhi, C. Ferster, A. Lin and L. Glass. 2020. The Google Earth Engine Mangrove Mapping Methodology (GEEMMM). *Remote Sens.* vol. :12. no.( 22). p. 3758. <https://doi.org/10.3390/rs12223758>
- Younes-Cárdenas, N., K.E. Joyce and S.W. Maier. 2017. Monitoring mangrove forests: Are we taking full advantage of technology?. *Int. J. Appl. Earth Obs. Geoinf.* vol. :63. pp. 1-14. <https://doi.org/10.1016/j.jag.2017.07.004>
- Younes-Cárdenas, N., T.D. Northfield, K.E. Joyce, S.W. Maier, N.C. Duke and L. Lyburner. 2020. A Novel Approach to Modelling Mangrove Phenology from Satellite Images: A Case Study from Northern Australia. *Remote Sens.* vol. :12(24). p. 4008. <https://doi.org/10.3390/rs12244008>
- Zanvo, M.G.S., Y.S.S. Barima, V.K. Salako, K.A.N. Koua, M.A. Kolawole, A.E. Assogbadjo and R.L. Glele Kakaï. 2021. Mapping spatio-temporal changes in mangroves cover and projection in 2050 of their future state in Benin. *Bois & Forêts des Tropiques.* vol:350. pp. 29-42. <https://doi.org/10.19182/bft2021.350.a36828>
- Zhang, Q., D. Devers, A. Desch, C. O. Justice and J. Townshend. Mapping tropical deforestation in Central Africa. *Environ Monit Assess.* vol:101(1). pp. 69–83. <https://doi.org/10.1007/s10661-005-9132-2>
- Zhu Y., K. Liu, W.S. Myint, Z. Du, Y. Li, J. Cao, L. Liu, Z. Wu. 2020. Integration of GF2 Optical, GF3 SAR and UAV Data for Estimating Aboveground Biomass of China's Largest Artificially Planted Mangroves. *Remote Sens.* vol:12(12). p. 2039. doi: 10.3390/rs12122039

