

Metodología para la generación de áreas construidas con Sentinel 2



Delimitación de áreas construidas a partir de datos geoespaciales en el Salvador

Ana Villalobos/DOA-MARN - 2024

1. OBJETIVO

Identificar las zonas construidas a partir de imágenes Sentinel-2 en El Salvador

2. ALCANCE

En este trabajo, las áreas construidas se definen como un conjunto de estructuras urbanas, rurales, industriales, mineras o asfaltadas. Estas áreas serán mapeadas a partir de las imágenes satelitales de Sentinel-2, considerando que la mínima unidad mapeada será el tamaño del píxel de la imagen (10x10 m). Se definió el área mínima a mapear de 5 veces el tamaño del píxel, siendo un área de 500 m².



3. DATOS

Para este análisis, se utilizaron imágenes ópticas del Instrumento Multiespectral (MSI) de Sentinel-2 de la Agencia Espacial Europea, con nivel de procesamiento L2A, que incluye correcciones atmosféricas. El periodo analizado corresponde al año 2024. Las imágenes contienen 13 bandas espectrales con una resolución espacial de entre 10 y 60 m, como se muestra en la tabla siguiente:

Nº de Banda	Nombre	Longitud de Onda (nm)	Resolución Espacial (m)
1	Azul	490	10
2	Verde	560	10
3	Rojo	665	10
4	Borde Rojo 1	705	20
5	Borde Rojo 2	740	20
6	Borde Rojo 3	783	20
7	NIR	842	10
8	NIR (cercano)	865	10
8A	NIR (cercano)	865	20
9	Vapor de Agua	945	60
10	SWIR 1	1375	60
11	SWIR 2	1610	20
12	SWIR 3	2190	20

Las imágenes utilizadas, libres de nubes, corresponden a diferentes meses de 2024, principalmente en julio, aunque también se incluyeron algunas de enero y octubre. El análisis se realizó por separado según la fecha y el Tile de la imagen. Para mejorar los resultados, también se incorporaron imágenes del 2020, lo que permitió reducir el efecto de las fronteras en la metodología aplicada.



4. MÉTODOS

Los métodos utilizados para la identificación de áreas construidas en todo el país a partir de las imágenes Sentinel-2 considera dos procesos importantes. Estos son los siguientes:

A. Análisis de Componentes Principales

El Análisis de Componentes Principales (PCA) es una técnica estadística multivariante de simplificación que permite transformar un conjunto de variables originales correlacionadas en un conjunto sintético de variables no correlacionadas. Las nuevas variables son combinaciones lineales de las variables originales, ordenadas en función de su dispersión en el conjunto de datos.

En este estudio, se adoptó el método de PCA para cada fecha seleccionada, siendo las variables las 12 bandas de Sentinel-2, debido a que la mayor parte de la información de cobertura urbana se concentran principalmente en algunas componentes se eligieron dichas componentes para generar las muestras de entrenamiento para la posterior clasificación.



B. Clasificación de imágenes: Random Forest

Random Forest (Bosque Aleatorio) es un algoritmo de clasificación que se usa para analizar imágenes de satélite y clasificar los diferentes elementos que aparecen en ellas (como bosques, agua, ciudades, etc.). Esta técnica utiliza árboles de decisión para tomar decisiones sobre cómo clasificar los píxeles de la imagen. Donde, cada árbol de decisión es un modelo que, al analizar una imagen, hace preguntas sobre las características de cada píxel (por ejemplo, "¿El valor del píxel en el canal rojo es mayor que X?") y, según las respuestas, decide a qué clase pertenece ese píxel (bosque, agua, etc.).

Este algoritmo usa muchos árboles de decisión entrenados con una muestra aleatoria de los datos de la imagen. Además, en cada árbol se eligen de manera aleatoria solo algunas de las características (como los valores de los diferentes canales espectrales de la imagen).



Después de que todos los árboles hacen su predicción, el modelo final asigna a cada píxel la clase que más árboles hayan votado. Lo que permite obtener mejor la precisión de la clasificación, ya que combina las decisiones de muchos árboles y reduce los errores que podrían cometer los árboles individuales [1].

Siguiendo este enfoque, se implementó el algoritmo *Random Forest* junto con el PCA para identificar las zonas construidas. Además, se aplicó un filtro espacial con una ventana de 8x8 píxeles para eliminar el ruido. Los resultados obtenidos se compararon con las imágenes de color verdadero de Sentinel-2. Posteriormente, se realizó la conversión de la capa ráster a vector y la validación de los resultados de la clasificación. Los pasos de este proceso están representados en el flujograma de la Figura 1.



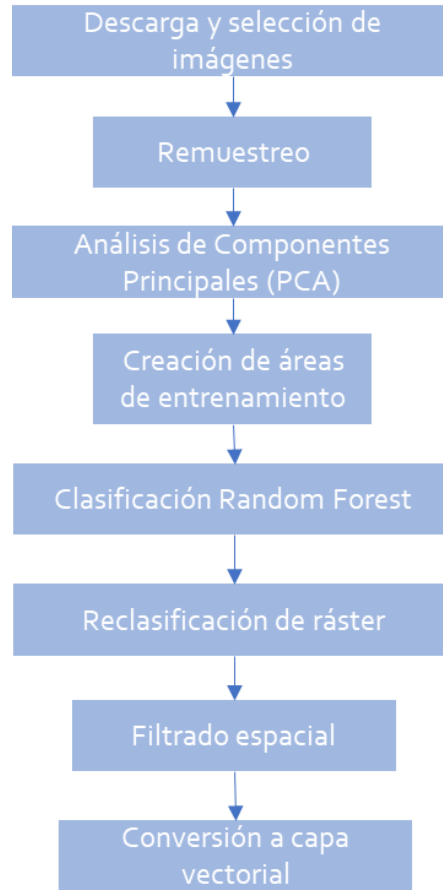


Figura 1: Flujograma de procesos para la delimitación de áreas construidas.



4. VALIDACIÓN DE LOS RESULTADOS

A. CÁLCULO DEL TAMAÑO DE LA MUESTRA

La fórmula utilizada para calcular el tamaño de la muestra para una población finita es la siguiente:

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1 - p)}{E^2 \cdot (N - 1) + Z^2 \cdot p \cdot (1 - p)}$$

Donde:

- n = tamaño de la muestra
- N = tamaño de la población total
- Z = valor de Z según el nivel de confianza (por lo general, se usa 1.96 para un nivel de confianza del 95%)
- p = proporción estimada de la población que tiene la característica de interés (0.5)
- E = margen de error aceptable (por ejemplo, 0.05 para un 5% de error)

Con una población de 59, 251 polígonos resultantes de la identificación de áreas construidas, un nivel de confianza del 95% y un margen de error del 5%, el tamaño de la muestra necesario sería aproximadamente 385 puntos.



B. PROCESO DE VALIDACIÓN

Se generaron puntos aleatorios sobre la capa de áreas construidas.

Se extrajeron aleatoriamente 385 puntos.

Para cada punto, se clasificó como "construida" o "no construida" utilizando como referencia las imágenes de Google.

Los resultados obtenidos para todo tipo de áreas construidas fueron los siguientes:

- Verdaderos Positivos (VP): 318 (áreas construidas correctamente identificadas).
- Falsos Positivos (FP): 67 (áreas construidas mal clasificadas).
- Precisión: Aproximadamente 82.6%.
- Tasa de Error: Aproximadamente 17.4%.



De estos resultados, la clasificación en las zonas urbanas fue:

- **Verdaderos Positivos (VP):** 21 (zonas urbanas correctamente identificadas).
- **Falsos Positivos (FP):** 2 (zonas no urbanas mal clasificadas como urbanas).
- **Precisión:** Aproximadamente 91.3 %.
- **Tasa de Error:** Aproximadamente 8.7%.

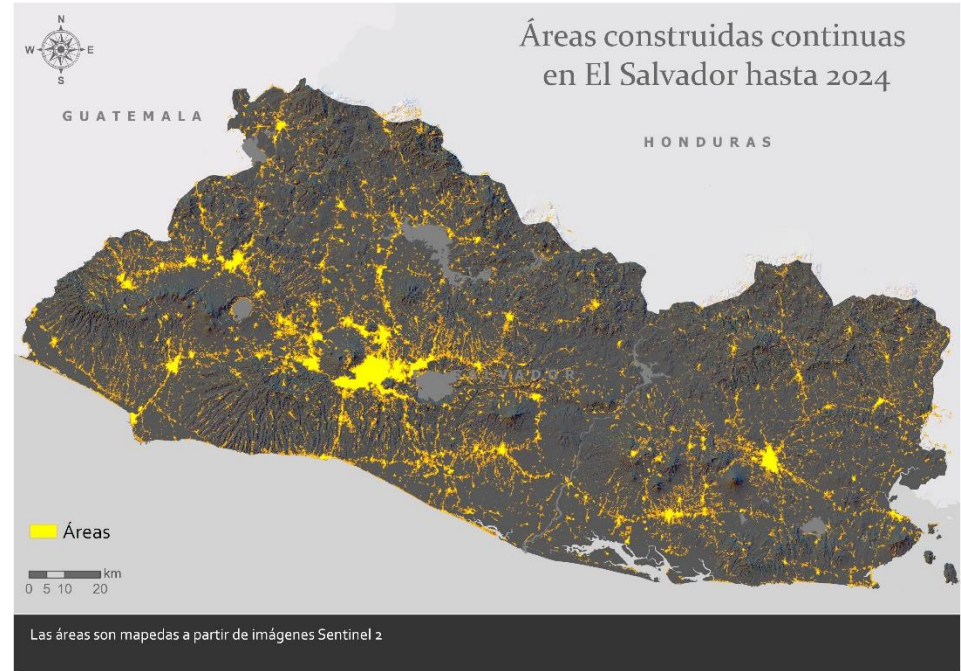


Figura 2: Áreas construidas identificadas con imágenes Sentinel-2 para el año 2024.



CONCLUSIÓN

El análisis de áreas construidas en El Salvador utilizando imágenes satelitales de Sentinel-2, combinado con técnicas avanzadas como el Análisis de Componentes Principales (*PCA*) y la clasificación mediante el algoritmo *Random Forest*, ha permitido obtener una delimitación precisa de las zonas urbanas y construidas en el país. A pesar de algunos errores de clasificación, particularmente en áreas rurales, la metodología aplicada mostró una precisión razonable de aproximadamente 82.63%. Mientras que en las áreas urbanas la tasa de error es menos del 9%. Esto demuestra la viabilidad de utilizar datos satelitales para el monitoreo de la expansión urbana y la identificación de infraestructuras construidas, lo que puede ser de gran utilidad para la planificación territorial, la gestión urbana y la toma de decisiones a nivel gubernamental. Sin embargo, se recomienda seguir refinando el modelo y explorar otras fuentes de datos y técnicas de clasificación para mejorar la precisión, especialmente en zonas rurales o menos urbanizadas.

REFERENCIAS

[1] https://pages.cms.hu-berlin.de/EOL/geo_rs/S08_Image_classification2.html

