

MAPBIOMAS
[COLOMBIA]

Apéndice - Infraestructura Urbana

Tema transversal

Colección 2

Versión 1

Noviembre 2024

Equipo responsable

Karen Andrea Huertas
Efrain Yury Turpo Cayo
Luis Medina
Natalia García

Índice

1 Introducción	2
2 Mosaicos de imágenes satelitales.	3
2.1 Composición de mosaicos	3
2.2 Selección de variables de clasificación	4
3 Clasificación	6
3.1 Definición de la clase Infraestructura Urbana (ID: 24)	7
3.2 Algoritmos de clasificación	9
3.3 Información de referencia	9
4 Post-clasificación	11
4.1 Llenado de vacíos de información - Filtro Gap fill	11
4.2 Filtro temporal	12
4.3 Filtro de frecuencia ajustado	13
4.4 Filtro espacial	14
6 Referencias	16

Índice de tablas

Tabla 1. Colecciones de datos Landsat utilizadas en la Colección 1.0.	3
Tabla 2. Feature space: Variables para la detección de superficies de Infraestructura Urbana con el reductor asociado.	5
Tabla 3. Reductores estadísticos empleados en la composición de imágenes Infraestructura Urbana Colección 4.0.	6
Tabla 4. Definición detallada de Infraestructura Urbana Colección 4.0.	8
Tabla 5. Información por país empleada como referencia global en el proceso de clasificación de infraestructura urbana.	10
Tabla 6. Información por país empleada como referencia nacional en el proceso de clasificación de infraestructura urbana.	11

Índice de figuras

Figura 1. Área de estudio para la detección de la cobertura de Infraestructura Urbana en la Colección 4.0 de MapBiomias Amazonía.	2
Figura 2. Método de generación de mosaicos anuales Colección 4.0 MapBiomias Amazonía.	4
Figura 3. Flujo metodológico de la clasificación de Infraestructura Urbana Colección 4.0	7
Figura 4. Detección Infraestructura Urbana Colección 4.0. (A) - Mosaico Landsat, (B) - Clasificación superficie urbana ID:24, (C) - Imágen de alta resolución Google Earth. (Florencia - Colombia)	8
Figura 5. Fuentes de referencia disponibles para el área de estudio de MapBiomias Amazonía.	10
Figura 6. Filtro gapfill.	12
Figura 7. Filtro temporal.	13
Figura 8. Filtro de frecuencia.	13
Figura 9. Filtro espacial.	14

1 Introducción

La colección 2 de mapas de MapBiomias Colombia incorpora la segunda versión del mapeo de infraestructura urbana como capa binaria de tipo transversal en la superficie que cubre todo el territorio de Colombia en el periodo comprendido entre los años de 1985 y 2023 (Figura 1). Esta clase de cobertura está constituida por superficies de infraestructura asociada a grandes y pequeños centros urbanos (pueblos) y aquellas áreas periféricas que están siendo incorporadas mediante un proceso gradual de urbanización o de cambio del uso del suelo en transición hacia fines residenciales, zonas industriales, redes viales, bases militares, redes de comunicación, aeropuertos y zonas verdes urbanas artificiales, no agrícolas y zonas de infraestructura asociadas a la explotación de hidrocarburos, así como pistas de aterrizaje convencionales y no convencionales.

Los paisajes urbanos presentes en la región están configurados espacialmente en numerosos focos de concentración de infraestructura que yacen sobre todo el territorio. Algunas zonas urbanas han surgido principalmente de la concentración demográfica rural que con el tiempo se han integrado en el conjunto de ciudades en diferentes jerarquías según su condición territorial. En este mapeo se incluyen ciudades intermedias, ciudades gemelas, ciudades fronterizas y centros rurales en paulatina expansión. Las conocidas ciudades intermedias predominan en la región y están definidas como centros en donde su economía está basada en el territorio circundante. Estos nodos tienen características transnacionales que generan proximidad con su entorno para diversificar sus mecanismos socioeconómicos (UCLG, 2018).

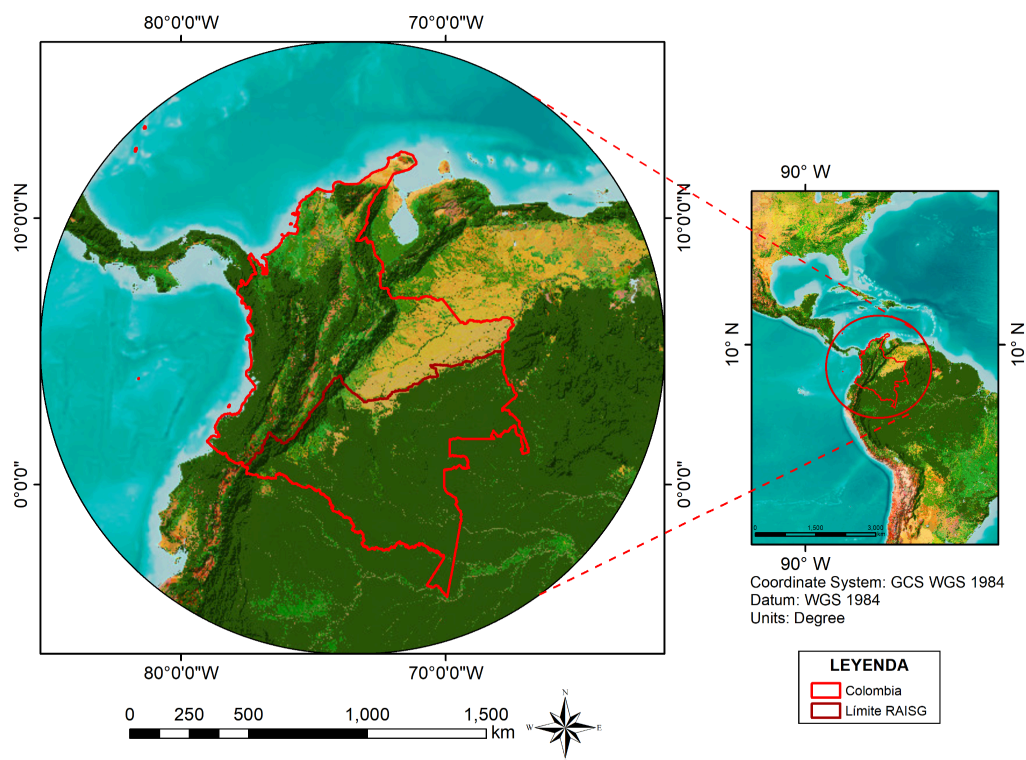


Figura 1. Área de estudio para la detección de la cobertura de Infraestructura Urbana en la Colección 2.0 de MapBiomias Colombia.

Otras iniciativas mundiales han desarrollado el mapeo del suelo urbano utilizando diferentes técnicas de teledetección y fuentes de datos (Schneider *et al.*, 2010), se encuentran disponibles productos globales como FROM-GLC de Gong *et al.* (2013) y GlobeLand30 de Chen *et al.* (2015) no obstante, la mayoría de esta información está disponible para un rango de años restringido; esto ha dificultado el monitoreo del comportamiento urbano durante un periodo histórico prolongado.

La serie de datos Landsat ha permitido perfeccionar la disponibilidad de información anual para MapBiomias Colombia a nivel regional como punto de partida en la presente colección a través del procesamiento automatizado en la nube desde Google Earth Engine. Los fundamentos metodológicos utilizados para el mapeo multitemporal de Infraestructura Urbana (ID:24) durante los últimos 39 años se describen a continuación teniendo en cuenta la siguiente estructura:

- A. Composición de mosaicos de imágenes satelitales.
- B. Clasificación multitemporal binaria de tipo transversal.
- C. Post-clasificación.

2 Mosaicos de imágenes satelitales.

2.1 Composición de mosaicos

La composición de mosaicos para la clasificación transversal de “Infraestructura Urbana” utilizó los mosaicos de imágenes generados previamente para la clasificación general de coberturas de la tierra de la Colección 2.0 de MapBiomias Colombia. Se realizó la composición a través de imágenes satelitales de reflectancia de superficie con corrección atmosférica del sensor Landsat 4, 5, 7, 8 y 9 Colección 2 que cubren la superficie asociada al límite de Colombia, generados anualmente en una serie temporal de 39 años comprendidos desde 1985 hasta 2023¹. La Tabla 1 indica las colecciones de imágenes alojadas en los repositorios de Google Earth Engine.

Tabla 1. Colecciones de datos Landsat utilizadas en la Colección 2.0.

Sensor	ID Colección en GEE
Landsat 4 TM	LANDSAT/LT04/CO2/T1_SR
Landsat 5 TM	LANDSAT/LT05/CO2/T1_SR
Landsat 7 ETM+	LANDSAT/LE07/CO2/T1_SR
Landsat 8 OLI /TIRS	LANDSAT/LC08/CO2/T1_SR
Landsat 9 OLI 2 /TIRS 2	LANDSAT/LC09/CO2/T1_SR

En la Figura 2 se describe la metodología general para la producción de mosaicos, primero se realizó la composición de imágenes según la disponibilidad de las colecciones de imágenes Landsat año a año desde 1985 hasta 2023 desde los repositorios de Google Earth Engine. La composición de imágenes utilizó la grilla de cartas millonésimas mundiales en donde se desarrolló la parametrización idónea para producir cada mosaico según criterios asociados a la temporalidad

¹ La composición de mosaicos se generó a partir del reductor de pixel de mediana con una resolución espacial de 30 metros.

(fecha inicial y fecha final), sensor, cobertura de nubes y escenas de exclusión que interfieren con la calidad de la imagen de mediana.

Adicionalmente se utilizó la capa de referencia disponible en Google Earth Engine - DEM STRM (Shuttle Radar Topography Mission - V03).

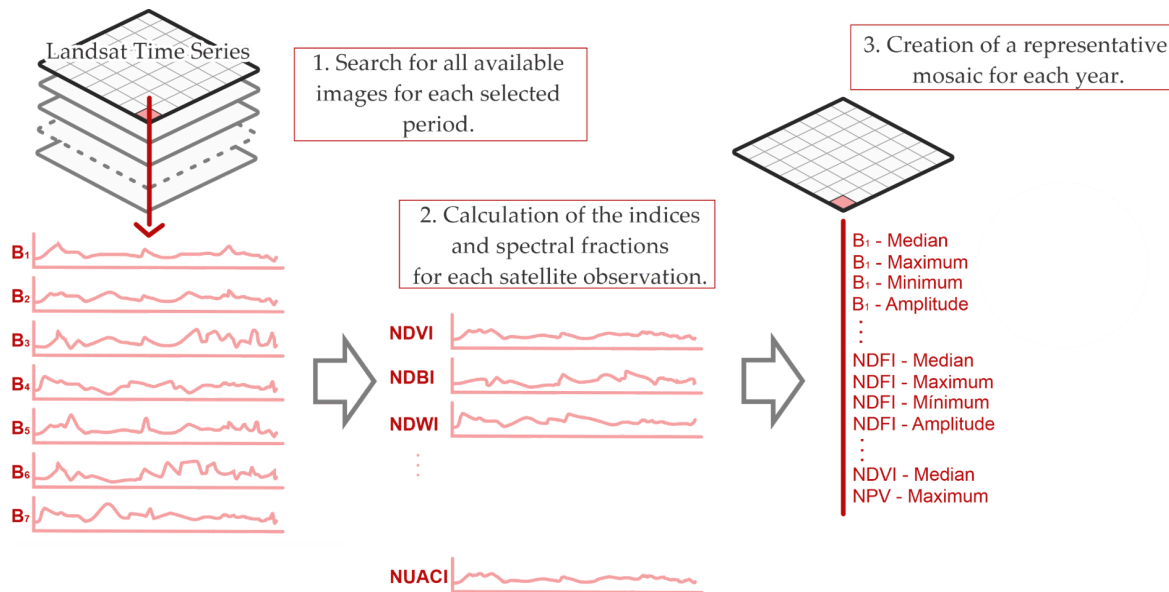


Figura 2. Método de generación de mosaicos anuales Colección 2.0 MapBiomias Colombia.

A partir de la parametrización, el preprocesamiento incluyó la aplicación de los algoritmos Cfmask y CloudScore que para Landsat 4, 5 y 7 (mosaicos de 1985 a 2012) empleó dos bandas, **pixel_qa** - Atributos de calidad de píxel generados a partir del algoritmo CFMASK y **sr_atmos_opacity** - Opacidad atmosférica; <0,1 = claro; 0,1 - 0,3 = promedio; > 0,3 = brumoso. Lo anterior, si se establece que el bit de nube (5) y la confianza de la nube (7) es alta, o si se establece el bit de sombra de nube (3), se considera un píxel incorrecto y por lo tanto se eliminan los píxeles de borde que no ocurren en todas las bandas. Para Landsat 8 y 9 (mosaicos de 2013 a 2023), se utilizó la banda **pixel_qa**, considerando que los bits 3 y 5 son nubes y sombras. Después de la eliminación, se realizó el cálculo de la mediana de la colección de imágenes sin nubes que genera el mosaico anual.

2.2 Selección de variables de clasificación

El siguiente paso consistió en calcular los índices y fracciones espectrales para cada observación satelital (feature space) que incluyen las bandas espectrales de Landsat, índices espectrales, información fraccional y de textura derivados de los índices y fracciones anteriores. También se emplearon variables estáticas: *HAND*, *shademask2*, *slppost*, *altitud*, *pendiente*, *latitud* y *longitud*; como apoyo a la clasificación de clases que espectralmente son muy similares pero se pueden lograr discriminar por las condiciones topográficas.

Tabla 2. Feature space: Variables para la detección de superficies de Infraestructura Urbana con el reductor asociado.

Variable	Description	Reductor
blue	Landsat BLUE band	Median
green	Landsat GREEN band	Median
red	Landsat RED band	Median
nir	Landsat NIR band	Median
swir1	Landsat SWIR1 band	Median
swir2	Landsat SWIR2 band	Median
soil	Soil Fraction	Median
snow	Snow Fraction	Median
cloud	Cloud Fraction	Median
slope	Slope	Static topographic variable
ndvi	Normalized Difference Vegetation Index	Median
ndwi_mcfeters	Normalized Difference Water Index (Mcfeters)	Median
ndbi	Normalized Difference Built-up	Median
nuaci	Normalized Urban Areas Composite Index	Median

Finalmente, se creó un mosaico representativo para cada año constituido por 156 bandas en total² fundamentados en el cálculo de reductores estadísticos para generar los valores de cada píxel. Estos reductores corresponden a:

Tabla 3. Reductores estadísticos empleados en la composición de imágenes Infraestructura Urbana Colección 2.0.

Reductor	Descripción
Mediana	Mediana de todos los valores disponibles en el mosaico anual para esa ubicación (píxel).
Mediana época seca	Cálculo de la mediana estadística aplicada a los píxeles del cuartil 25 (con menores valores) de NDVI (proxy de época seca).
Mediana época húmeda	Cálculo de mediana estadística aplicada a los píxeles del cuartil 75 (con los mayores valores) de NDVI (proxy de época lluviosa).
Amplitud	Extensión de la variación entre todos los píxeles disponibles en el mosaico anual.
Desviación estándar	Desviación estándar de los valores de todos los píxeles disponibles en el mosaico anual para una ubicación determinada.

² Ver detalle completo en la descripción de bandas y variables en la ATBD General de MapBiomás Colombia Colección 2.0.

Mínimo	Menor valor de todos los píxeles disponibles en el mosaico anual en una ubicación determinada.
Máximo	Mayor valor de todos los píxeles disponibles en el mosaico anual en una ubicación determinada.
Mínimo del periodo seco	Cálculo del menor valor de todos los píxeles disponibles de las imágenes del cuartil con los menores valores de NDVI (proxy de época seca).
Mínimo del periodo húmedo	Cálculo del menor valor de todos los píxeles disponibles de las imágenes del cuartil con los mayores valores de NDVI (proxy de época lluviosa).
Máximo del periodo seco	Cálculo del mayor valor de todos los píxeles disponibles de las imágenes del cuartil con los menores valores de NDVI (proxy de época seca).
Máximo del periodo húmedo	Cálculo del mayor valor de todos los píxeles disponibles de las imágenes del cuartil con los mayores valores de NDVI (proxy de época lluviosa).
QMO del periodo seco	El valor más alto que tiene la banda en el índice evi2 en la estación seca.
QMO del periodo lluvioso	El valor más alto que tiene la banda en el índice evi2 en la estación húmeda.

3 Clasificación

El objetivo de la clasificación consistió en producir una capa binaria entre infraestructura urbana y áreas no urbanas; considerando únicamente dos posibilidades para cada píxel (1: Presencia de Infraestructura Urbana, 2: Ausencia de Infraestructura Urbana). La clasificación a partir de muestras de entrenamiento sobre los mosaicos Landsat se realizó por medio del algoritmo de clasificación Random Forest (RF) (Breiman, 2001).

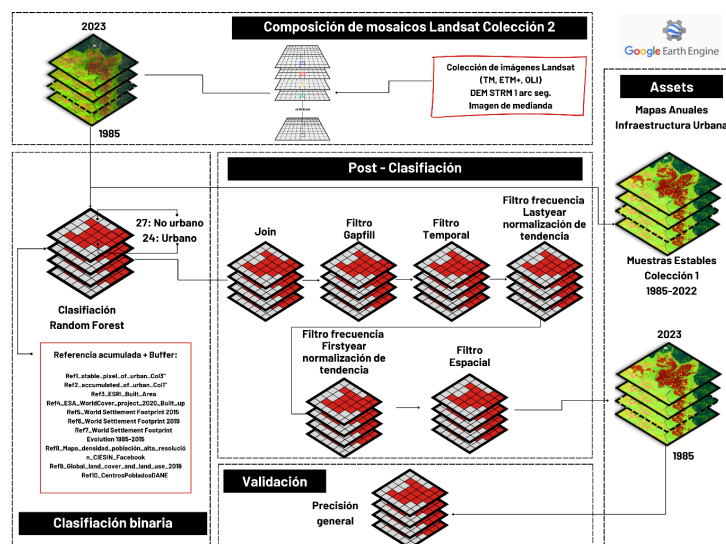


Figura 3. Flujo metodológico de la clasificación de Infraestructura Urbana Colección 2.0

La clasificación se realizó sobre las 82 regiones de clasificación para el transversal de la Colección 2.0 con presencia de la clase de infraestructura urbana según fuentes de referencia conocidas e información de imágenes de alta resolución.

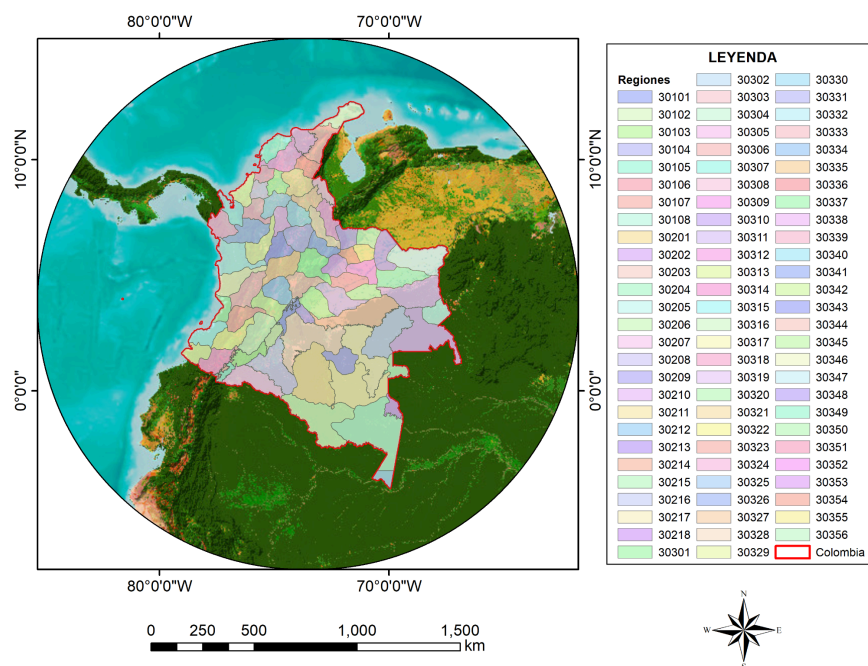


Figura 4. Regiones de clasificación de Infraestructura Urbana Colección 2.0

3.1 Definición de la clase Infraestructura Urbana (ID: 24)

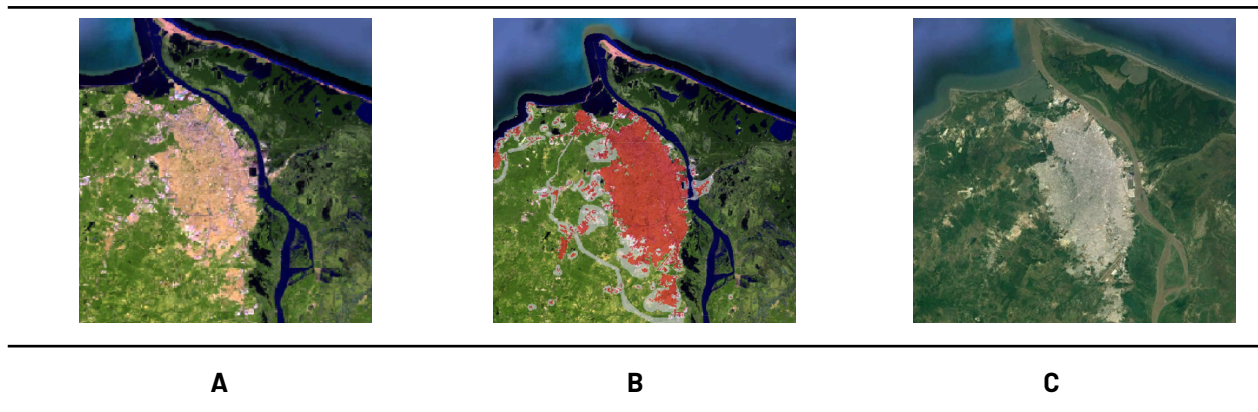
Esta clase de cobertura está definida como superficies de infraestructura asociada a grandes y pequeños centros urbanos (pueblos) y aquellas áreas periféricas que están siendo incorporadas mediante un proceso gradual de urbanización o de cambio del uso del suelo³ en transición hacia fines residenciales, zonas industriales, redes viales, bases militares, redes de comunicación, aeropuertos y zonas verdes urbanas artificiales, no agrícolas (Tabla 4).

Tabla 4. Definición detallada de Infraestructura Urbana Colección 2.0.

Incluye	Excluye
1. Zonas urbanizadas	Comunidades Indígenas
1.1. Tejido urbano continuo (Grandes centros poblados)	Carreteras y caminos sin pavimentar en zonas rurales.
1.2. Tejido urbano discontinuo (Pueblos)	
2. Redes viales	
2.1. Red vial, ferroviaria y terrenos asociados	
3. Otras zonas artificializadas	
3.1. Obras de explotación de hidrocarburos (Bloques petroleros)	
3.2. Hidroeléctricas	
3.3. Bases militares	
3.4. Aeropuertos (Pistas de aterrizaje en centros urbanos)	
3.5. Zonas portuarias	
3.6. Otro tipo de obra de escala industrial	
4. Zonas verdes (no agrícolas)	
4.1. Instalaciones recreativas en centros urbanos	
4.2. Prados urbanos y separadores viales	
Pistas de aterrizaje no convencionales en zonas rurales	

³ Área mínima de mapeo: mayor o igual a media hectárea (≥ 0.5 ha): aproximadamente 5 píxeles Landsat de 30 m^2 .

Tabla 5. Detección Infraestructura Urbana Colección 2.0 (A) - Mosaico Landsat, (B) - Clasificación superficie urbana ID:24, (C) - Imágen de alta resolución Google Earth. (Barranquilla - Colombia).



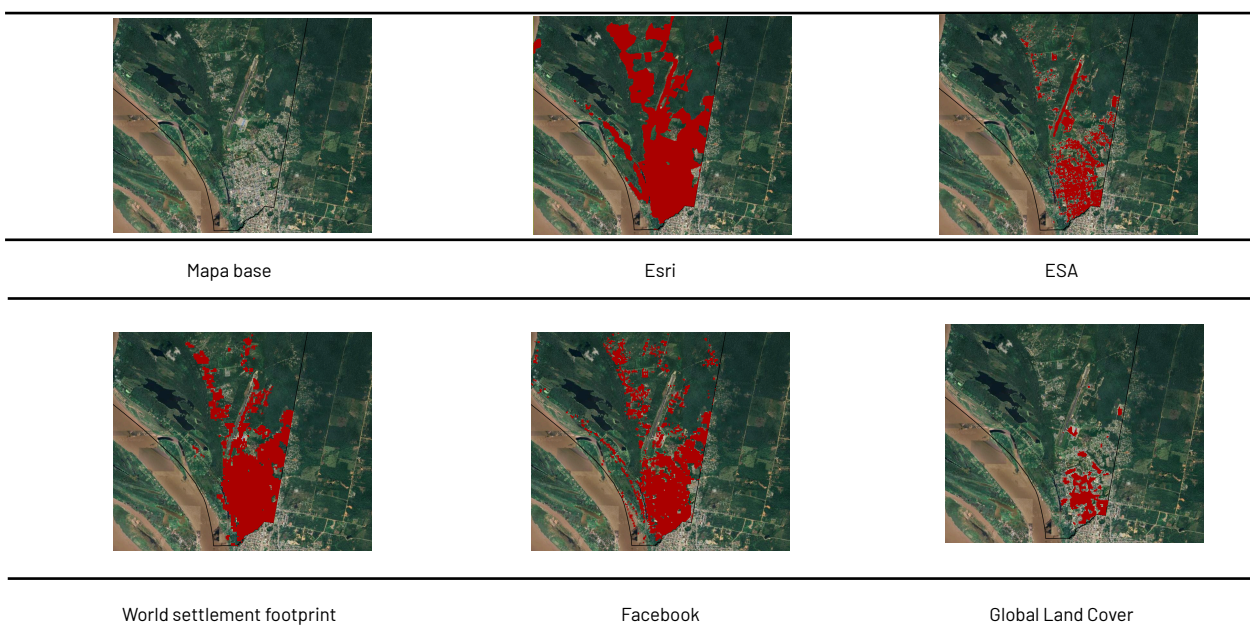
3.2 Algoritmos de clasificación

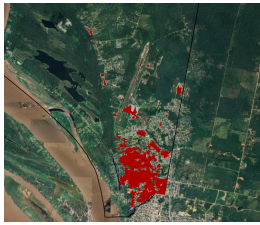
Se implementó el algoritmo de aprendizaje automático bosques aleatorios (Random Forest) (Breiman, 2001), en la plataforma de Google Earth Engine. Se generaron muestras de entrenamiento y se clasificó posteriormente en función al comportamiento de los índices empleados y el apoyo sustentado en mapas de referencia disponibles a nivel regional como OpenStreetMap e información de referencia nacional y mundial.

3.3 Información de referencia

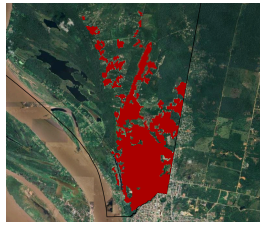
Para Colombia, en el algoritmo de Random Forest se utilizaron nueve fuentes de referencia, recopiladas desde proyectos con información geográfica que dispone de datos mundiales en diferentes temporalidades incluyendo además, el acumulado y muestras estables de la Colección 1.0 de MapBiomás Colombia a partir de la información anual desde 1985 a 2022. La información fue filtrada de acuerdo a la presencia de superficies de infraestructura urbana o conglomerados urbanos con atributos de carácter residencial, comercial, servicios, industrial, parques entre otros.

Tabla 6. Fuentes de referencia disponibles para el área de estudio de MapBiomás Colombia

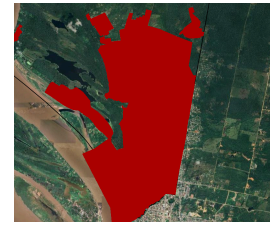




Muestras estables Col 1



Acomulado Col 1



DANE

La Tabla 7 y 8 describe la información secundaria empleada en Colombia como fuente de referencia en el proceso de clasificación de infraestructura urbana.

Tabla 7. Información de Colombia empleada como referencia global en el proceso de clasificación de infraestructura urbana.

Referencias Globales	
Referencias Globales	Colombia
Ref1_stable_pixel_of_urban_Col3	●
Ref2_accumulated_of_urban_Col3	●
Ref3_ESRI_Built_Area	
Ref4_ESA_WorldCover_project_2020_Built_up	●
Ref5_World Settlement Footprint 2015	
Ref6_World Settlement Footprint 2019	●
Ref7_World Settlement Footprint Evolution 1985-2015	
Ref8_Map_a_densidad_población_alta_resolución_CIESIN_Facebook	●
Ref9_Global_land_cover_and_land_use_2019	●

Tabla 8. Información de Colombia empleada como referencia nacional en el proceso de clasificación de infraestructura urbana.

Referencias nacionales	Fuente
Zonas urbanizadas, zonas industriales o comerciales y redes de comunicación, zonas verdes artificializadas (no agrícolas), vialidad, zonas residenciales, construcciones. Zona Urbana integrado	Coberturas de la Tierra para Colombia metodología CORINE Land Cover - IDEAM. (2012). Mapa de ecosistemas continentales, costeros y marinos para Colombia - IDEAM (2017). OpenStreetMap (2021) Capa de centros poblados para Colombia -DANE (2018).

4 Post-clasificación

Debido a la naturaleza basada en píxeles del método de clasificación y el trabajo sobre una serie temporal extensa, se ejecutó un proceso de post clasificación que incluye la aplicación de filtros (Gapfill, temporal, espacial y de frecuencia) sobre la discriminación binaria de

Infraestructura Urbana con el propósito de reducir inconsistencias temporales, reducir ruidos de clasificación menores a la unidad mínima de mapeo y llenar vacíos de información. Adicionalmente se incluyó un “Join” para unir los años clasificados o reclasificados en la colección dos con los años ya procesados de la colección 1.

4.1 Unión de años colección 1 y colección 2

Para la colección dos se implementó un paso previo a los filtros post clasificación. Este paso consiste en unir los años que se clasificaron para la colección dos (e.j 2023), con los ya clasificados y procesados en la colección 1 (e.j 1875-2022). El resultado de este paso es un solo stack de imágenes que posteriormente son procesadas con los filtros tradicionales. Adicionalmente, este paso también permitió modificar la clasificación de áreas específicas en los años procesados anteriores al 2023; esto con el fin de mejorar la calidad de clasificación en toda la serie temporal.

4.2 Llenado de vacíos de información - Filtro Gap fill

La secuencia de filtros para el tema transversal de infraestructura urbana comienza con el relleno de vacíos de información - *Gap fill* considerando que la extensión que compromete MapBiomás Colombia y las superficies urbanas a su vez experimentan condiciones atmosféricas y climáticas que resultan en mosaicos anuales con algunos píxeles sin información o ausencia de datos observados por el satélite (“gaps”). Este filtro tiene la capacidad de reducir estos vacíos residuales reemplazando el valor temporalmente más cercano. Cuando un píxel “futuro” carece de valor (sin datos), el filtro de Gap fill le asigna el valor del año más próximo al mismo. Este filtro hace una revisión de la serie donde primero llena vacíos haciendo un barrido de “atrás hacia adelante”, donde los vacíos son llenados con datos de los años próximos que anteceden al año sin datos. De haber vacíos remanentes, estos son llenados con datos del año más próximo precedente. Para cada píxel cuyo valor ha sido completado empleando este filtro, el cambio se registra en un archivo de metadata el cual registra la historia del píxel. Se pueden utilizar varios años para cubrir vacíos. Por lo tanto, las brechas de información solo persisten si un píxel dado se ha clasificado permanentemente como sin información en toda la serie temporal.

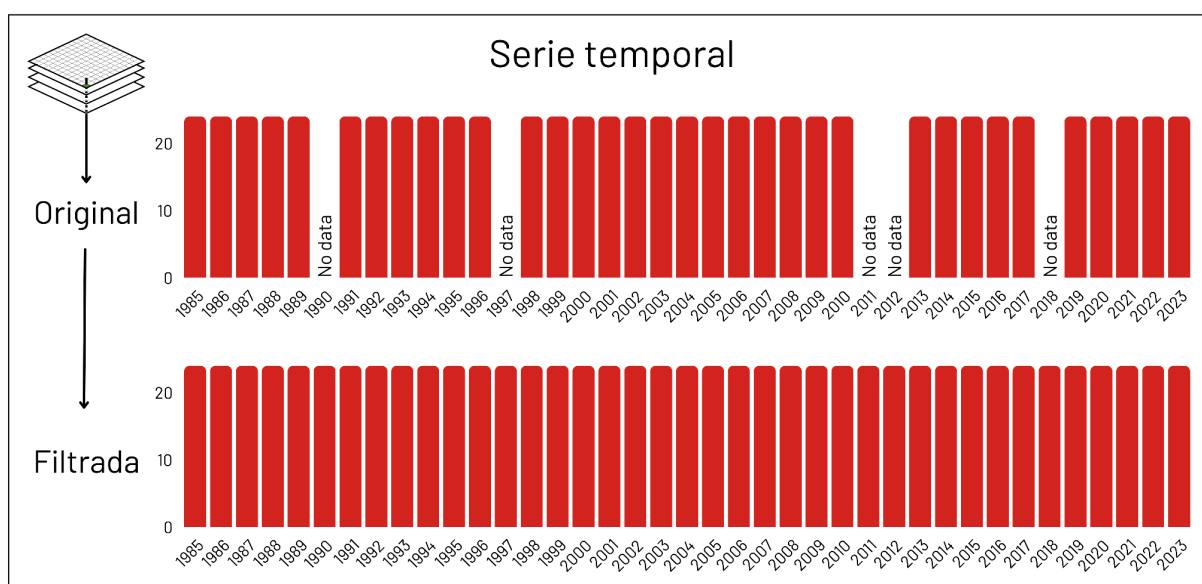


Figura 6. Filtro gapfill.

4.3 Filtro temporal

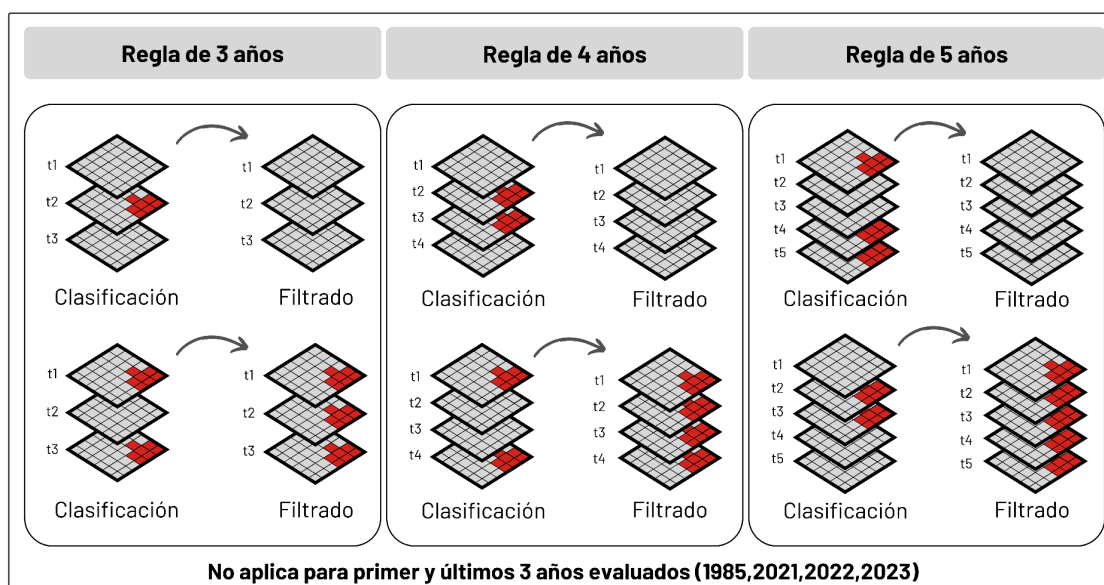
Se ejecutó un filtro temporal utilizando ventanas de 3, 4 y 5 años en toda la serie de tiempo. El filtro inspecciona el valor de cada píxel clasificado en relación al valor de ese píxel en clasificaciones temporalmente consecutivas. Para ello emplea una ventana móvil unidireccional que toma en consideración secuencias de clasificaciones de 3 a 5 años e identifica transiciones temporales no permitidas. El filtro temporal se aplica a cada píxel de todos los años de la colección. Dependiendo del año que la regla modificará, existen tres tipos de reglas:

Reglas generales (RG). Aplicadas a píxeles de años en posiciones intermedias en secuencias de 3 a 5 años. Esta regla se aplica únicamente en casos donde hay una inconsistencia temporal; por ejemplo, cuando secuencias de años consecutivos tienen valores idénticos a excepción del píxel en posición central. En estos casos, el filtro modificará el valor del píxel central para que guarde consistencia con los píxeles que le anteceden y suceden. En el caso de secuencias de 3 años, solo existe una opción de posición central o año intermedio. En el caso de secuencias de 4 o 5 años, existen dos o tres alternativas de posiciones centrales. Esta regla modifica los valores de las clasificaciones de los años 1986 al 2022.

Reglas de primer año (RP). Aplicadas únicamente al primer año de la serie temporal. Esta regla modifica los valores de la clasificación del año 1985.

Reglas de último año (RU). Aplicada al último año de la clasificación. Esta regla modifica los valores de la clasificación del año 2023.

De este modo, los filtros temporales reducen vacíos de información e inconsistencias temporales o cambios que no son posibles o no están permitidos.



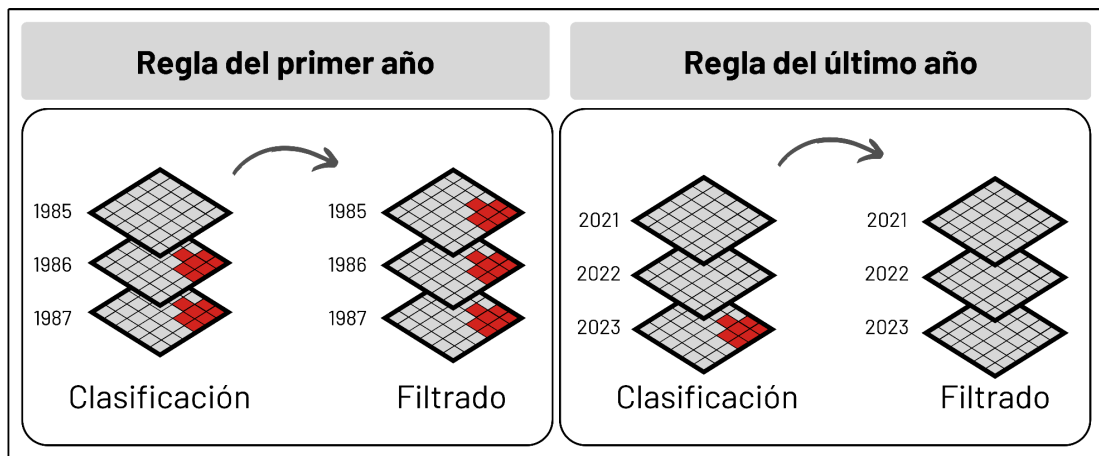


Figura 7. Filtro temporal.

4.4 Filtro de frecuencia ajustado

Este filtro considera la ocurrencia de la clase a lo largo de la serie temporal para normalizar la trayectoria histórica de la clase. Teniendo en cuenta que el comportamiento de infraestructura urbana manifiesta una tendencia incremental y de expansión a lo largo de la serie se aplicaron dos filtros (primer y último año) con el fin de regular el aumento consistente de los píxeles urbanos y evitar fluctuaciones anómalas de un año a otro.

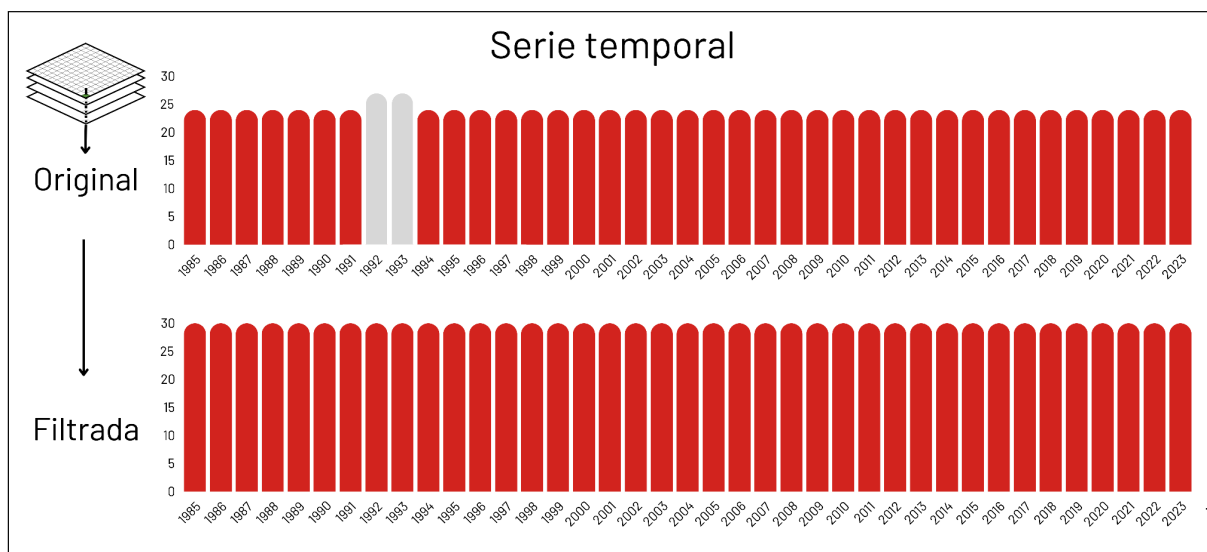


Figura 8. Filtro de frecuencia.

4.5 Filtro espacial

El último filtro de la secuencia aplicada en post clasificación fue el filtro espacial basado en la función "connectedPixelCount" de Google Earth Engine. Esta función localiza píxeles conectados (vecinos) que comparten el mismo valor empleando una ventana móvil. Únicamente los píxeles que no comparten una conexión con un número predefinido de vecinos idénticos son considerados como píxeles aislados. En el caso de MapBiomás Amazonía, la unidad mínima de mapeo fue

definida como 0.5 ha (aproximadamente 5 píxeles). Consecuentemente, se requirió que por lo menos cinco píxeles estén conectados para cumplir con el criterio de conexión mínima. De este modo, el filtro espacial suaviza diferencias locales al eliminar píxeles aislados o de borde menores a 0.5 ha, incrementando la consistencia espacial de la clase.

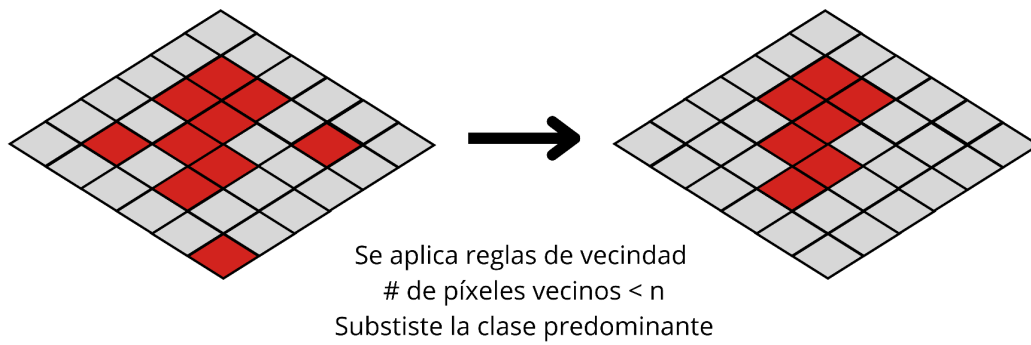


Figura 9. Filtro espacial.

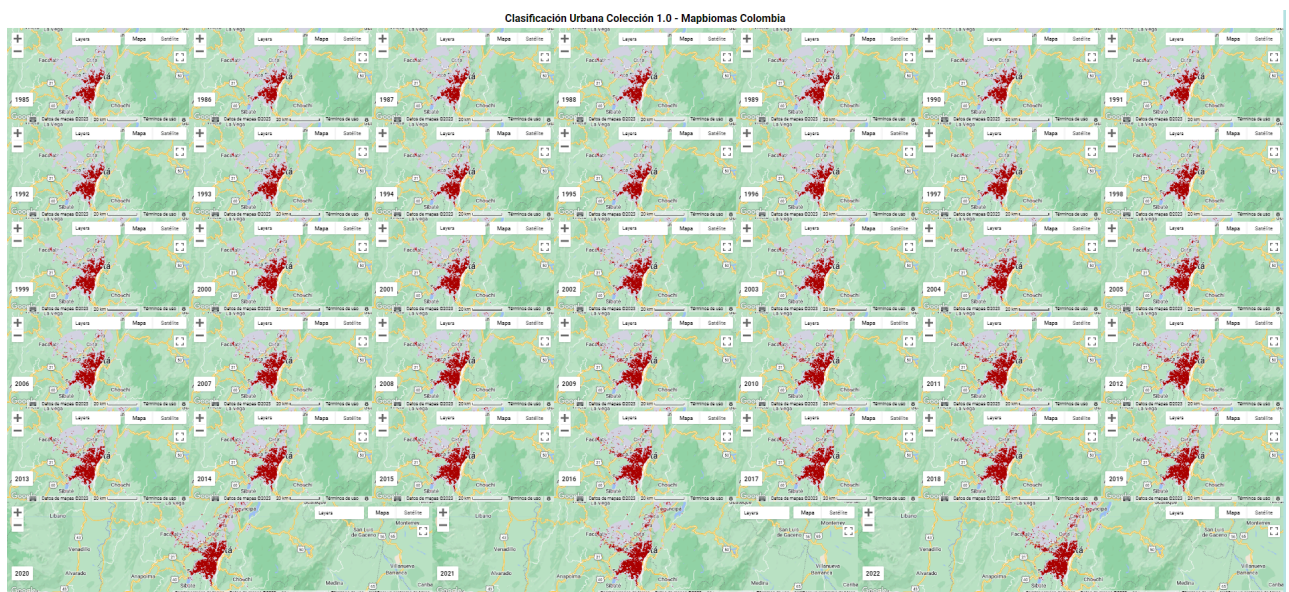


Figura 10. Clasificación multitemporal Infraestructura Urbana Colección 2.0 - 1985 a 2022 (caso Bogotá - Colombia).

6 Referencias

Breiman, L. (2001). Random Forest. Statistics Department University of California Berkeley. 1-30.

Chen, J., Chen J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S. & Lu, M. (2015). Global land cover mapping at 30 m resolution: a POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* 103, 7-27.

Gong, P., Wang, J., Yu, L., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H. & Liu, S. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat™ and ETM+ data. *Int. J. Remote Sens.* 34, 2607-2654.

Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. (2014) Good Practices for Estimating Area and Assessing Accuracy of Land Change. *Remote Sens. Environ.* 148, 42–57, doi:10.1016/j.rse.2014.02.015.

Schneider, A., Friedl, M.A., Potere, D. (2010). Mapping global urban areas using MODIS 500-m data: new methods and datasets based on 'urban ecoregions'. *Remote Sens. Environ.* 114, 1773-1746.

Souza, C., Monteiro, D., Shimbo, J. & Azevedo, T. (2019). Apéndice Infraestrutura Urbana. Versión 1. Recuperado el 2 de Julio de 2021:https://mapbiomas-br-site.s3.amazonaws.com/Infraestrutura_Urbana__Appendix_-_ATBD_Cole%C3%A7%C3%A3o_4_dez_2019.pdf

UCLG (2018). Amazonian Cities. Peer Learning about the sustainable use of terrestrial ecosystems #23. United Cities and Local Governments. Bolivia. 36 p.