



Apéndice - Minería

Tema transversal en Colección 1

Versión 1

Equipo responsable

Andrés Medina

Karen Andrea Huertas
Emanuel Valero

Octubre 2023

Índice

1. Introducción	3
2. Construcción y definición de mosaicos Landsat	3
3. Resumen metodológico	5
4. Variables de clasificación	5
5. Clasificación	6
5.1. Búsqueda y actualización de datos	6
5.2. Preparación de los datos	8
5.3. Muestras de entrenamiento (minería/no minería)	9
5.4. Clasificación con algoritmo Random Forests	11
6. Post-clasificación	13
5.1 Gap fill	13
5.2 Filtro espacial	14
5.3 Filtro Temporal	14
5.3 Integración	16
7. Referencias	18

Lista de figuras

<i>Figura 1: Distribución de la cuadrícula Landsat para el transversal de minería</i>	4
<i>Figura 2: Pasos para la clasificación de minería</i>	5
<i>Figura 3: Presencia de minería identificada por cada fuente de referencia</i>	7
<i>Figura 4: Máscara de clasificación minera para Colombia colección 1</i>	8
<i>Figura 5: Fuentes de información adicionales</i>	9
<i>Figura 6: Localización de las regiones de clasificación para Colombia colección 1</i>	10
<i>Figura 7: Polígonos de zonas mineras - Cerrejón</i>	10
<i>Figura 8: Polígono de Mina y No-Mina Minería a cielo abierto - Cerrejón</i>	11
<i>Figura 9: Polígono de Mina y No-Mina - Minería Aluvial - Taparal - Chocó</i>	11
<i>Figura 10: Clasificación de minería</i>	12
<i>Figura 11: Esquema funcionalidad Filtro Gap Fill</i>	13
<i>Figura 12: Esquema de la funcionalidad del filtro espacial</i>	14
<i>Figura 13: Efecto del filtro temporal</i>	15
<i>Figura 14: Filtro temporal regla de primer y último año</i>	15
<i>Figura 15: Integración clase minera.</i>	19

Lista de Tablas

<i>Tabla 1: Variables para la clasificación de zonas mineras de la Colección 1, MapBiomias Colombia</i>	6
<i>Tabla 2: Jerarquía integración en la Orinoquía</i>	16
<i>Tabla 3: Jerarquía integración en el Caribe</i>	17
<i>Tabla 4: Jerarquía integración en los Andes</i>	17
<i>Tabla 5: Jerarquía integración en el Pacífico</i>	18

1. Introducción

La cobertura asociada a la minera se clasifica como un tema transversal dentro de la metodología de MapBiomias, debido a que genera confusión en las clasificaciones que incluyen varias clases de cobertura como los centros urbanos, los bancos de arena y las áreas sin vegetación; al no poder diferenciarse de manera adecuada de otras clases como herbazales, pastos o agricultura. La metodología ha sido ejecutada previamente en varios países como Bolivia, Ecuador, Perú y Venezuela dando resultados positivos en la identificación de la clase; en Colombia se ha implementado esta metodología de mapeo en la colección 4 de la amazonía presentada en el 2022.

Este documento describe la metodología utilizada en la colección 1 para la clasificación de las zonas mineras en el territorio colombiano; definidas como todas las áreas de extracción de materiales pétreos o minerales, entendida como la sustancia cristalina generalmente inorgánica con características físicas y químicas asociadas a una agrupación atómica específica; materiales de construcción, materiales de arrastre como arenas, gravas, piedras en cercanías a los cauces y orillas de los ríos, vegas de inundación y demás terrenos aluviales, producción industrial con clara exposición del suelo, sin hacer diferenciación del tipo de actividad (industrial, artesanal, legal o ilegal) (Congreso de Colombia, 2001) (IDEAM, 2010). En este proceso metodológico de clasificación de minería no se identifican las áreas asociadas a salinas, ya que estas se están identificando y clasificando dentro de la clase de salares (clase 32) desde el proceso de clasificación general de la colección 1 de Colombia.

Es importante señalar que la metodología original del trabajo se basó en la experiencia del equipo de MapBiomias Brasil, y en especial de Cesar Diniz (Solutions in Geoinformation, Solved, Brasil), quien ofreció una presentación al equipo MapBiomias para la identificación de la minería industrial de ese país. Esta fue implementada en la herramienta Google Earth Engine con ajustes menores para mejorar el rendimiento y la calidad.

2. Construcción y definición de mosaicos Landsat

Para fines de procesamiento, la metodología de MapBiomias, divide el área del proyecto en una cuadrícula regular definida con base a la grilla de imágenes de Landsat para las regiones de Andes, Caribe, Caribe Insular, Pacífico, Pacífico Insular y Orinoquía, cada cuadrícula cubre un área de aproximadamente 2.7 Mha. En la figura 1 se muestra el área de estudio de Colombia conformada por 43 cuadrículas (figura 1)

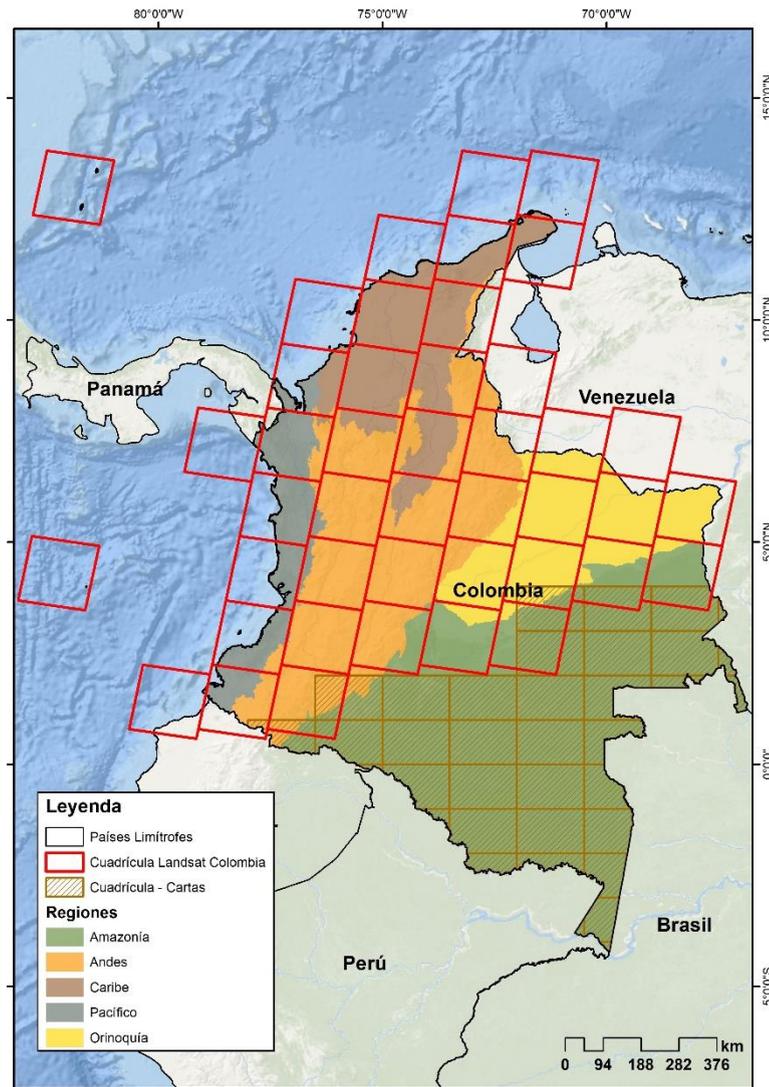


Figura 1: Distribución de la cuadrícula Landsat para el transversal de minería

En cada una de estas o celdas de la cuadrícula, se construyen los mosaicos anuales de imágenes Landsat. Esto con el objetivo de generar imágenes periódicas sin nubes que permitan su clasificación e identificación de cambios. Para mayor detalle revisar la sección 4 del documento: Apéndice - Colombia Algorithm Theoretical Basis Document - ATBD Colección 1.0. En esta colección para la construcción de los mosaicos se utilizó Landsat Collection 2, la cual presenta la versión más actualizada del reprocesamiento de los archivos de Landsat, con varias mejoras en la calidad de las imágenes.

La construcción de los mosaicos se inicia con la selección de las imágenes Landsat. Para ello se construye una tabla de parametrización con los siguientes parámetros, almacenados como metadatos en cada asset construido:

- ID: Identificador o código único de la unidad carta/región
- Año: Año de la serie (comprendiendo el período 1985 a 2022) al que corresponde el mosaico.
- Fecha inicial/ Fecha final: Periodo del año (fecha de inicio y final) para la selección de imágenes del catálogo de datos de imágenes Landsat de Google Earth Engine.

- Sensor: El satélite y su respectivo sensor: Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI, Landsat 9 o una combinación de Landsat 5 y Landsat 7.
- Porcentaje de nubes: Porcentaje máximo de cobertura de nubes aceptado de cada imagen Landsat que será usado para construir el mosaico de imágenes. Este dato proviene de la metadata de las imágenes Landsat.
- Blacklist: Imágenes que por su calidad son excluidas de la construcción del mosaico.

Con esta información se construyen los mosaicos anuales de imágenes, mediante la plataforma de Google Earth Engine. Todos los mosaicos generados se almacenan como una colección de imágenes de Google Earth Engine y están disponibles en la ruta: projects/mapbiomas-raisg/MOSAICOS/mosaics-pathrow-2; los cuales presentan inicialmente los valores de mediana de cada una de las bandas de los satélites Landsat. Con base en estos datos, se agregaron a cada mosaico bandas adicionales de índices espectrales, fracciones espectrales, variables de textura e índices topográfico-geomorfológicos derivados del modelo de elevación digital entre otros, con la finalidad de mejorar la capacidad de detección de la minería en el país mediante una diferenciación de presencia y ausencia de la clase de interés. Se aplica un cálculo estadístico a cada una de las bandas de los mosaicos, para generar información específica sobre: mediana, mediana estación seca (corresponde al percentil 25), mediana estación lluviosa (correspondiente al percentil 75), amplitud (rango de valores), la desviación estándar y el valor mínimo. Todos estos datos componen nuestras variables, que se utilizan como fuente para el clasificador de random forest.

3. Resumen metodológico

Los 6 pasos de clasificación del tema transversal de minería se representan esquemáticamente en el siguiente gráfico (figura 2).

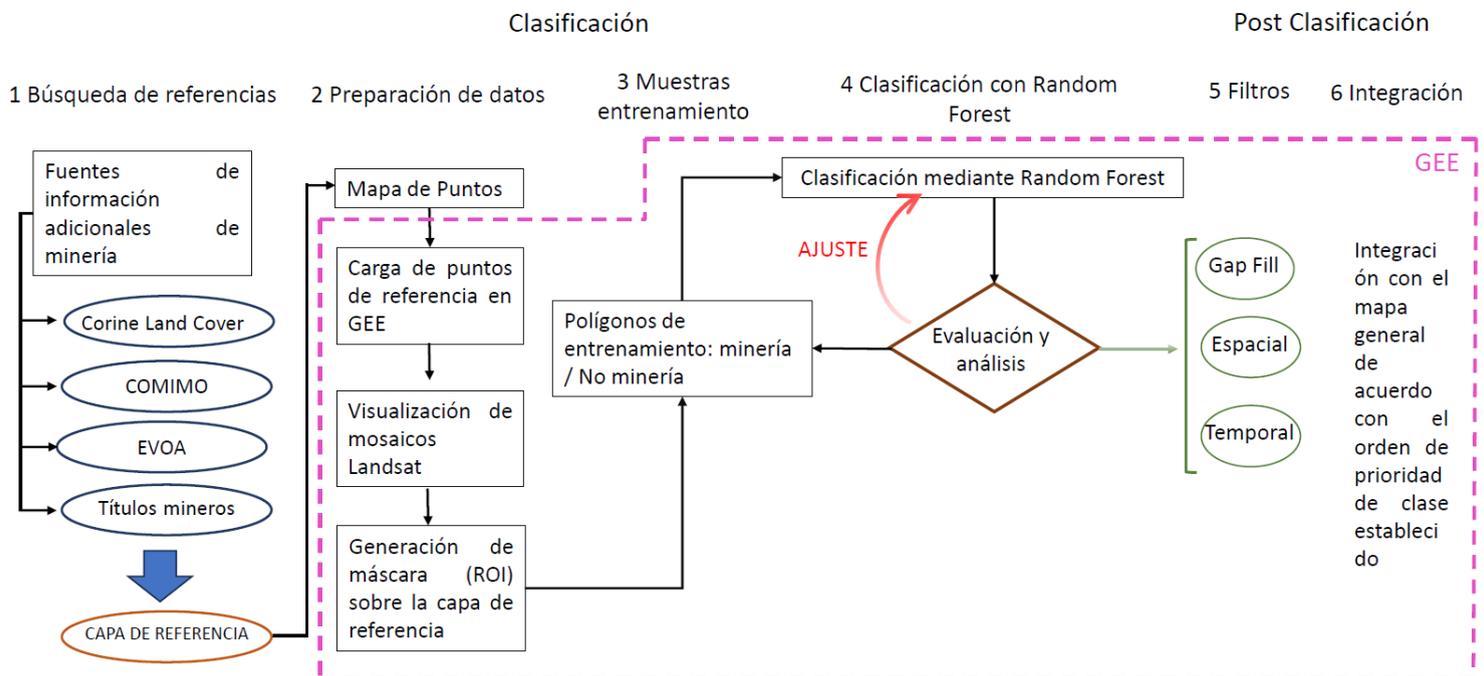


Figura 2: Pasos para la clasificación de minería

4. Variables de clasificación

La selección de variables (de los mosaicos) para la identificación de las zonas mineras en Colombia es el paso inicial para la clasificación. Se realizaron pruebas que permitieron identificar un conjunto de variables específicas para la detección de la minería. En total se usaron 12 variables, las cuales se presentan en la tabla 1.

Banda	Expresión / Descripción	Reductor
blue	Valor de la banda azul	Mediana
green	Valor de la banda verde	Mediana
red	Valor de la banda roja	Mediana
nir	Valor de la banda infrarrojo cercano (NIR)	Mediana
swir1	Valor de la banda SWIR1	Mediana
swir2	Valor de la banda SWIR2	Mediana
ndfi		Mediana
ndfib		Mediana
npv		Mediana
soil		Mediana, Amplitud
evi2	$2.5 * ((\text{NIR} - \text{Red}) / (\text{NIR} + 2.4 * \text{Red} + 1))$	Mediana
ndvi	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	Mediana
ndwi_gao	$(\text{Green} - \text{SWIR1}) / (\text{Green} + \text{SWIR1})$	Mediana
ndwi_mcfeters	$(\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR})$	Mediana
mndwi	$(\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR})$	Mediana
gcvi	$(\text{nir} / \text{green}) - 1$	Mediana
nuaci	$UNTL * (1 - \sqrt{(\text{NDWI} - a\text{NDWI})^2 + (\text{NDVI} - a\text{NDVI})^2} + (\text{NDBI} - a\text{NDBI})^2)$	Mediana
pri	$(\text{blue} - \text{green}) / (\text{blue} + \text{green})$	Mediana

Tabla 1 - Variables para la clasificación de zonas mineras de la Colección 1, MapBiomias Colombia

5. Clasificación

Para generar la clasificación se desarrolló un esquema metodológico dividido en seis etapas, cuatro de las cuales corresponden al proceso de clasificación y las restantes 2 a la etapa de post-clasificación; los resultados generados se almacenaron en Google Earth Engine. El proceso involucró la identificación de las áreas mineras por fuentes externas oficiales, toma de muestras y clasificación en cada una de las regiones identificadas, para luego integrar todas las minas identificadas en una capa única y, finalmente, con el mapa general (colección 1) de MapBiomias Colombia. A continuación, se describe cada paso asociado a la clasificación.

5.1. Búsqueda y actualización de datos

Se realizó una búsqueda de fuentes de datos auxiliares oficiales que reportan presencia de actividad minera en diferentes rangos de tiempo, entre las cuales podemos encontrar COMIMO (colombian mining monitoring), EVOA (explotación de oro de aluvión), Corine Land Cover y los títulos mineros (Figura 3), con el objetivo de generar la máscara de referencia para la identificación de minería, la cual es usada como el

punto de partida para el proceso de identificación y clasificación de la minería en el país (ver figura 4).

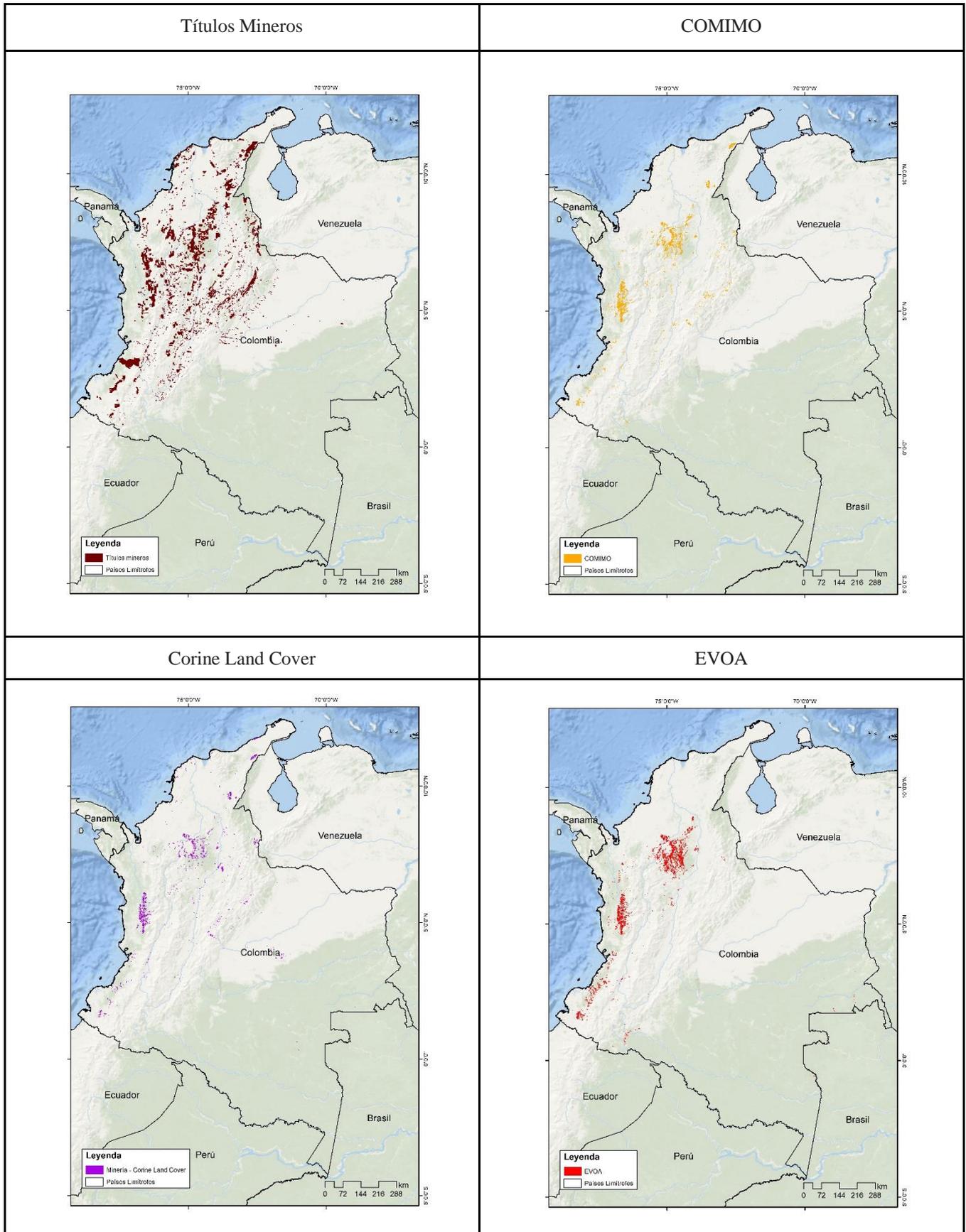


Figura 3: Presencia de minería identificada por cada fuente de referencia

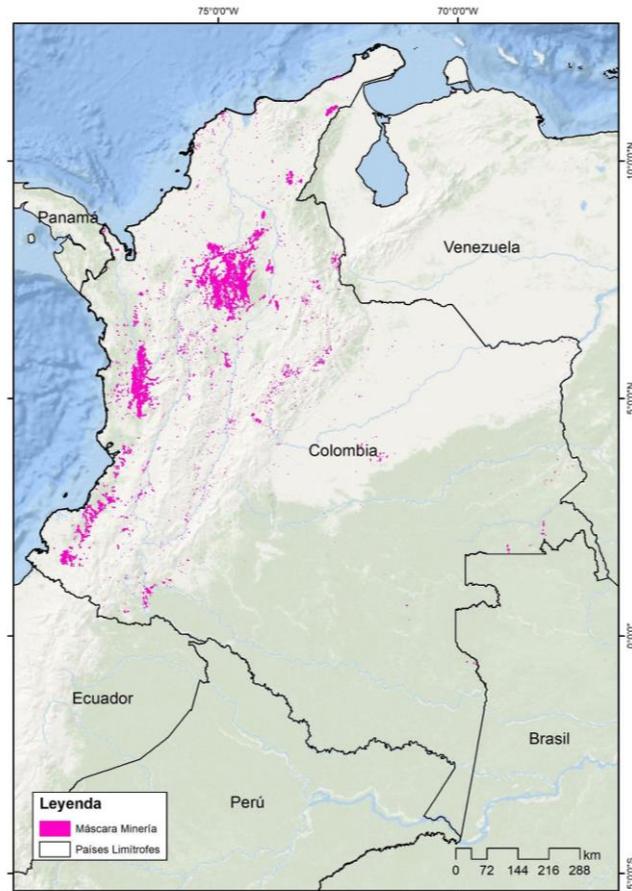


Figura 4: Máscara de clasificación minera para Colombia colección 1

De igual manera se utilizaron otras fuentes de información espacial de apoyo con el fin de identificar la presencia de minería en una región, el tipo de minería y algunos actores que pueden estar explotando dichos recursos en la zona. Las fuentes de información adicional son: zonas potenciales de minerales para la industria, zonas potenciales para minerales energéticos, metales potenciales para la industria del acero, metales especiales, metales básicos, materiales de construcción y zonas minero-éticas. (figura 5)

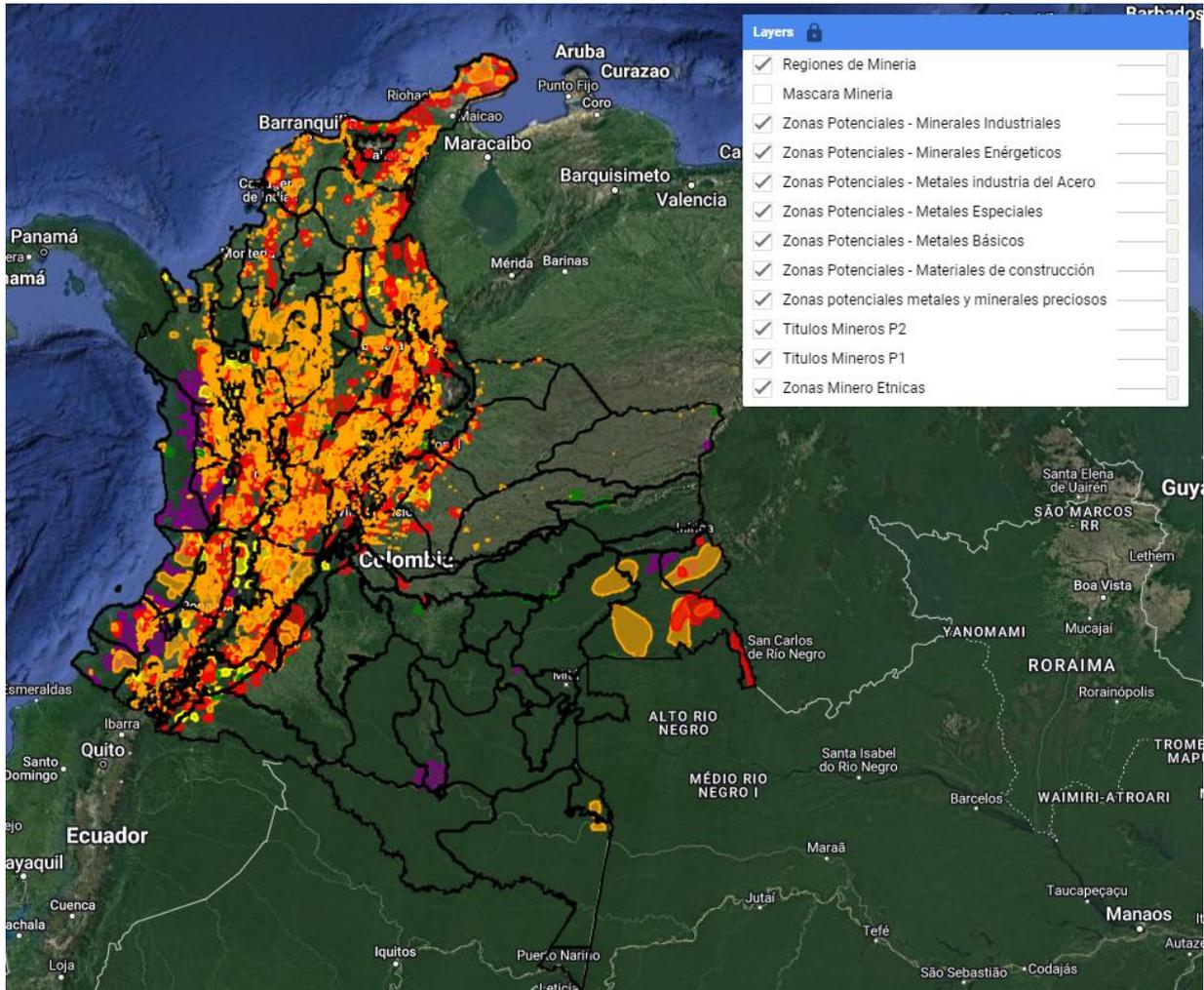


Figura 5: Fuentes de información adicionales

5.2. Preparación de los datos

Teniendo como insumo principal la capa de la máscara de regiones de clasificación (ver figura 4) y junto con imágenes de alta resolución del base map de Google Maps como apoyo, se procede a la delimitación de las regiones de clasificación en las cuales se identifica la presencia de actividad minera desde 1985 hasta el 2022 obteniendo como resultado 48 regiones de clasificación en las cuales se va a ejecutar la metodología de interpretación de la clase de minería (ver figura 6). Dicha capa se carga en Google Earth Engine (GEE) para iniciar el proceso de clasificación con la toma de muestras de entrenamiento.

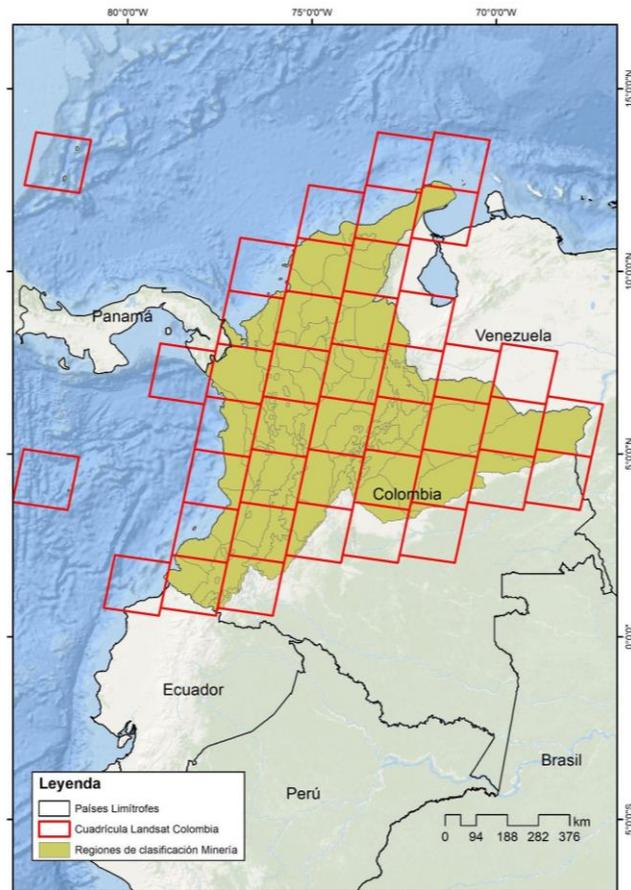


Figura 6: Localización de las regiones de clasificación para Colombia colección 1

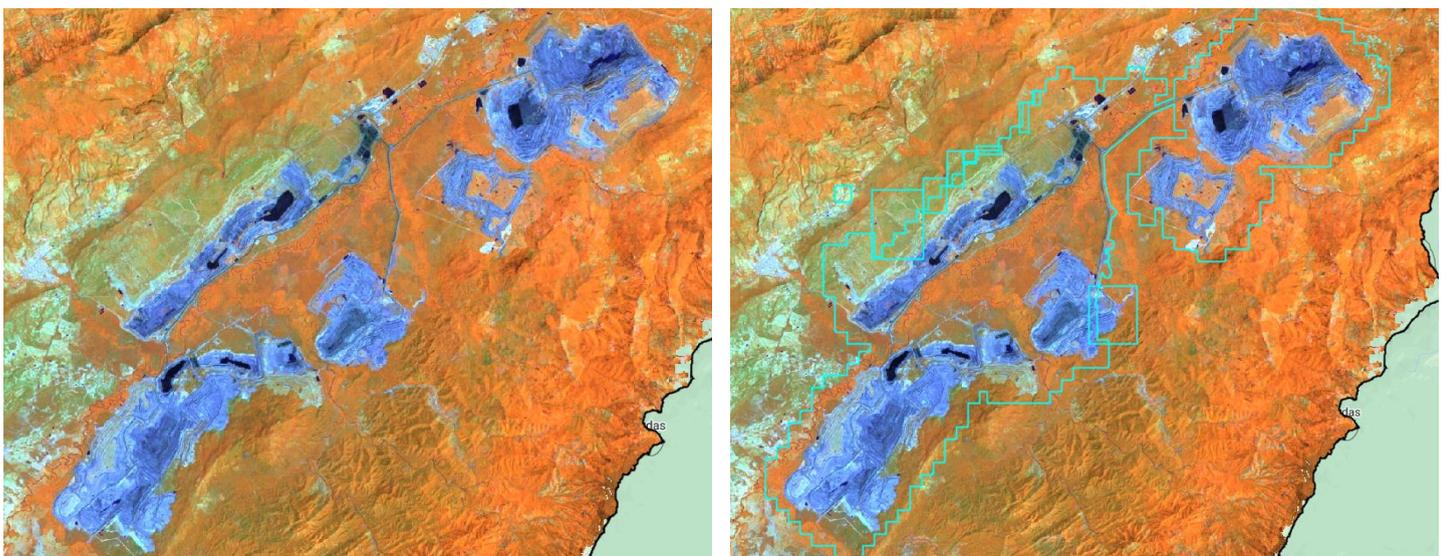
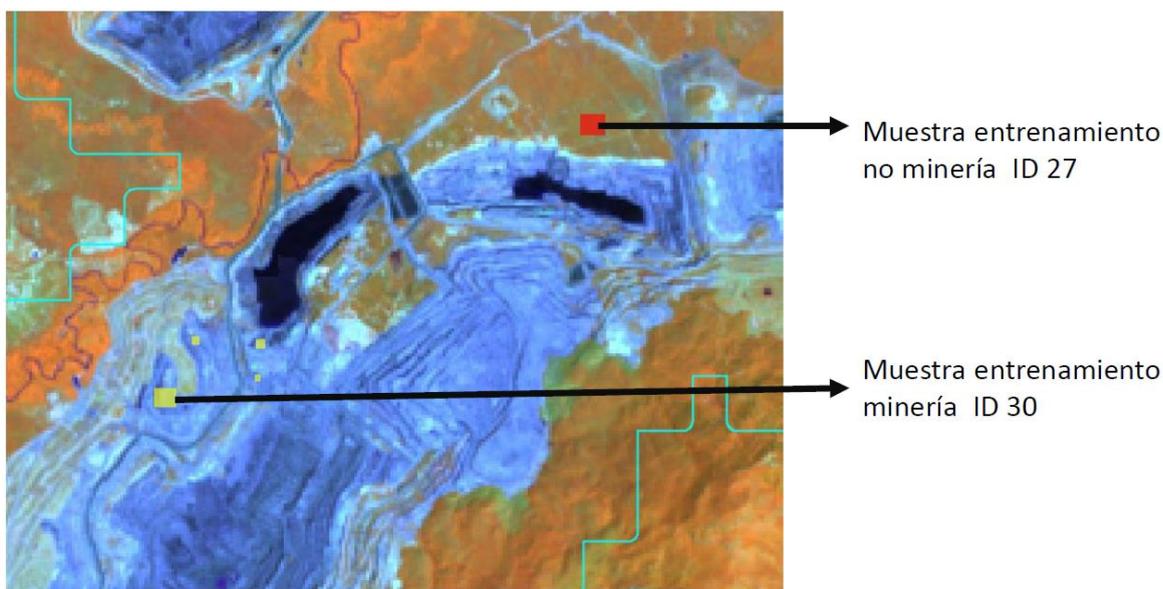


Figura 7: Polígonos de zonas mineras - Cerrejón

5.3. Muestras de entrenamiento (minería/no minería)

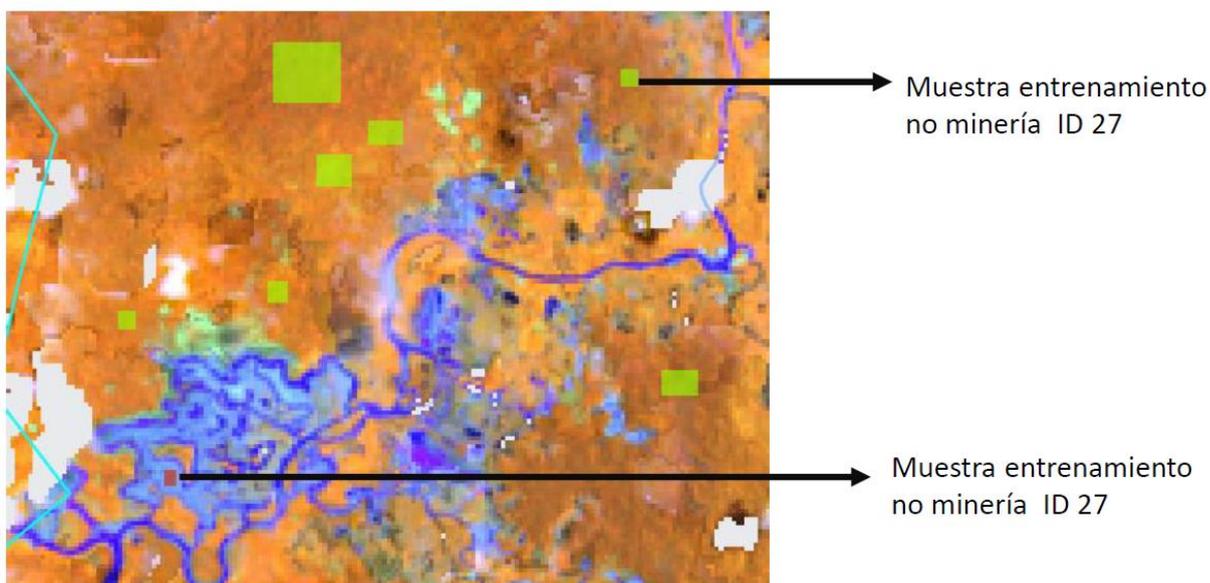
Dentro de los polígonos que delimitan las áreas mineras de interés (ver figura 4 y 7) se definieron e identificaron manualmente dos tipos de polígonos de entrenamiento: uno definido como 'Mina' (ID 30), que contiene estrictamente las zonas que hacen parte del área de explotación minera o de la mina y el segundo polígono abarca la clase 'no observado' o aquellas áreas que no corresponde a la actividad minera (ID 27) (figuras 8-9). Las muestras de entrenamiento tomadas dentro de la máscara son las que se utilizan para generar la clasificación de la zona de minería. Todo lo que está fuera de la zona delimitada no es clasificado ya que se considera como zona no minera y/o asignada a otra clase dentro de la leyenda MapBiomás.



Muestra entrenamiento no minería ID 27

Muestra entrenamiento minería ID 30

Figura 8: Polígono de Mina y No-Mina Minería a cielo abierto - Cerrejón



Muestra entrenamiento no minería ID 27

Muestra entrenamiento no minería ID 27

Figura 9: Polígono de Mina y No-Mina - Minería Aluvial - Taparal - Chocó

5.4. Clasificación con algoritmo Random Forests

El proceso de clasificación se basa en el uso del clasificador Random Forest, el cual mediante el uso de algoritmos de inteligencia artificial (IA) y procesos de aprendizaje automático (Machine Learning) permite la diferenciación de las zonas con presencia de minería en las regiones determinadas. A partir de la construcción de una colección de árboles estructurados decisiones independientes, se asigna una clase a cada uno de los píxeles (Breiman, 2001) realizando una serie de particiones binarias, permitiendo así realizar predicciones futuras con base en esta clasificación, estos árboles de decisión se establecen de forma aleatoria a fin de reducir la correlación entre los mismos. La clasificación de las zonas mineras se llevó a cabo en la plataforma de Google Earth Engine con base en la metodología implementada por el equipo de MapBiomias Brasil (Solutions in Geoinformation, Solved, Brasil) para la identificación de la minería industrial de ese país.

La clasificación se ejecutó dentro de las zonas delimitadas en los polígonos dibujados (ver figura 4) principalmente y dentro de cada una de las regiones (ver figura 6) con el fin de identificar nuevas zonas mineras que no se habían referenciado por fuentes externas. El resultado fue una imagen formada por píxeles con valor binario indicando la presencia o ausencia de zonas mineras (área definida en color rojo en la figura 10), dentro de cada área definida previamente en la máscara de clasificación.

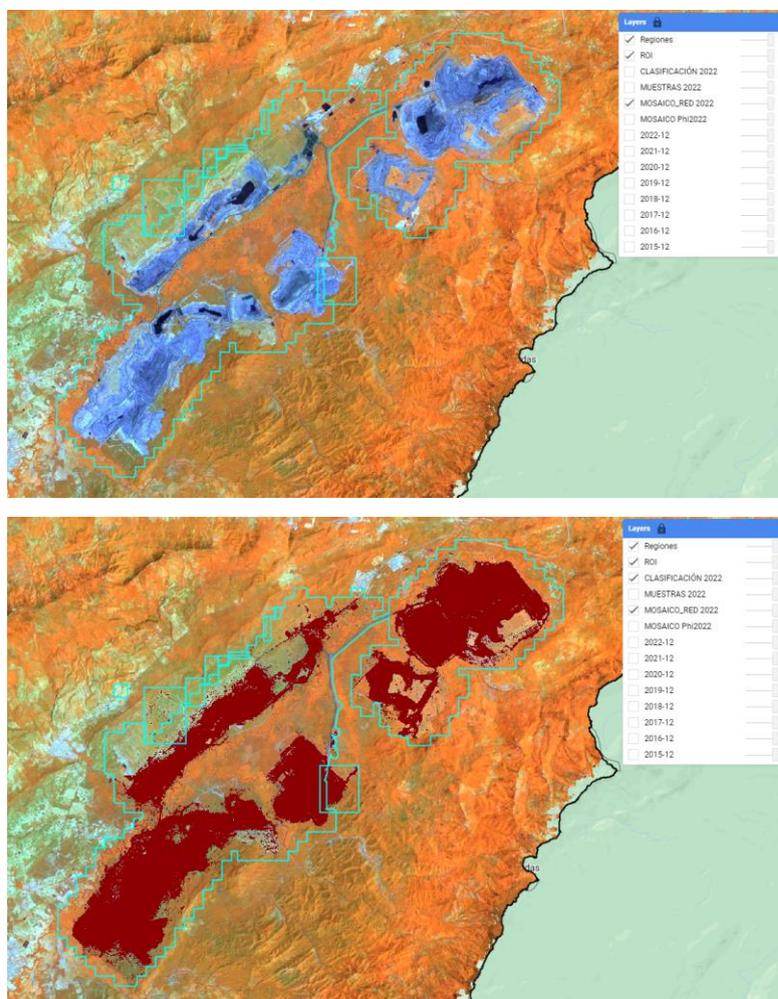


Figura 10: Clasificación de minería. Arriba. Mina a cielo abierto - El Cerrejón. Abajo. Clasificación de zona minera

Dentro del proceso de clasificación, con el apoyo de las capas de referencia se identificaron un grupo de características asociados a las zonas mineras, entre estas podemos encontrar el tipo de minería: de cantera, de cielo abierto, subterránea extractiva, aluvial, arrastre aluvial, de pequeña y mediana escala. Así mismo permite identificar si la actividad minera está asociada a alguna comunidad étnica, dentro de las cuales 33 no están vinculadas a alguna comunidad y 15 si son ejercidas por alguna comunidad. Dentro de los minerales explotados se pueden encontrar: arenas, gravas, oro, piedra caliza, mármol, calcita, travertino, cobre, anticita, anhidrita, azufre, yeso, piedras preciosas, platino níquel entre otros.

6. Post-clasificación

Los resultados de la clasificación se sometieron a la aplicación de una secuencia de filtros con el propósito de reducir las inconsistencias temporales, vacíos de información, ruidos de sobrestimación y otros problemas de clasificación. Para ello se utilizó gap fill, un filtro espacial y uno de temporal.

5.1 Gap fill

El territorio colombiano debido a su posición geográfica y a su topografía presenta una alta presencia de nubes a lo largo del año, adicional a esto debido a los problemas de bandeamiento presentados en el sensor Landsat 7 los mosaicos presentan pérdida de información o pixeles sin información (No Data). Debido a esto, es necesario rellenar estos vacíos de información o información nula mediante el uso del Gap fill, el cual nos permite reducir estos vacíos con el valor de píxel temporalmente más cercano.

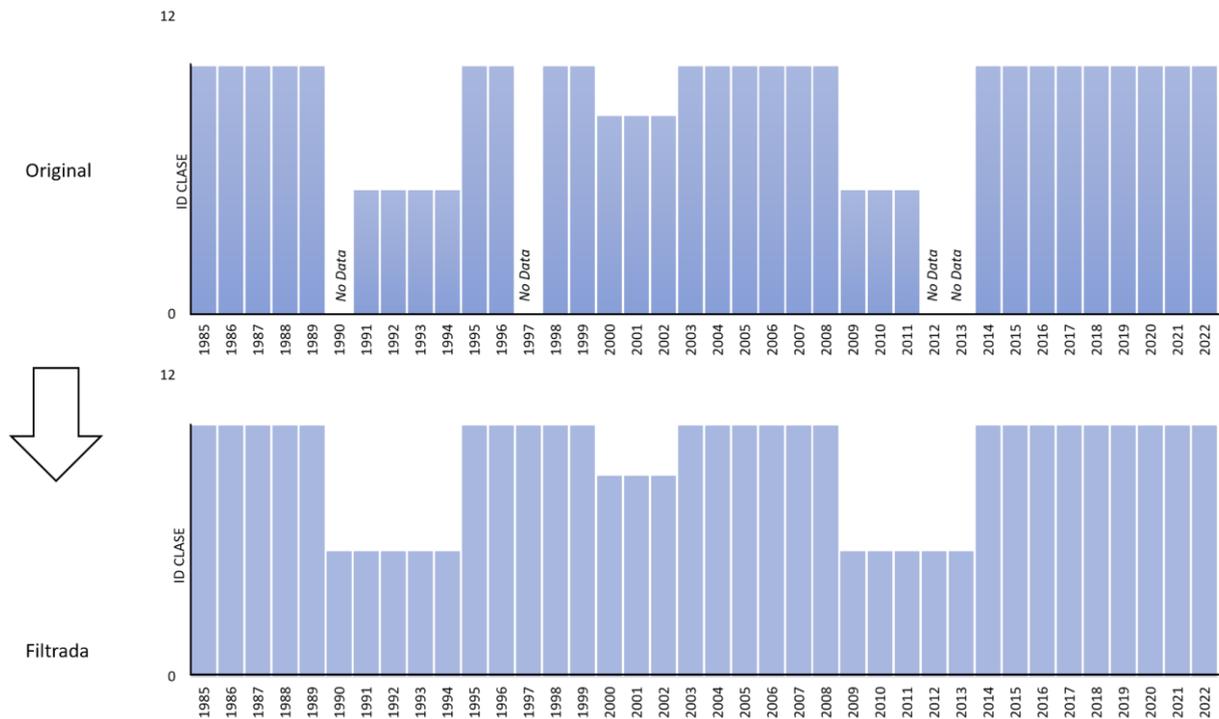


Figura 11: Esquema funcionalidad Filtro Gap Fill

5.2 Filtro espacial

El filtro espacial está basado en la función "connectedPixelCount" de Google Earth Engine, que identifica al conjunto de píxeles vecinos que comparten un mismo valor. Utilizando esta función y considerando una cantidad mínima preestablecida de píxeles idénticos, aquellos que no presentan conexión (valores diferentes) se consideran píxeles aislados y fueron reclasificados (figura 12). Para ambas clases (ID: 27, ID: 30), la conexión mínima establecida fue de 5 píxeles (aprox. 0.5 ha).

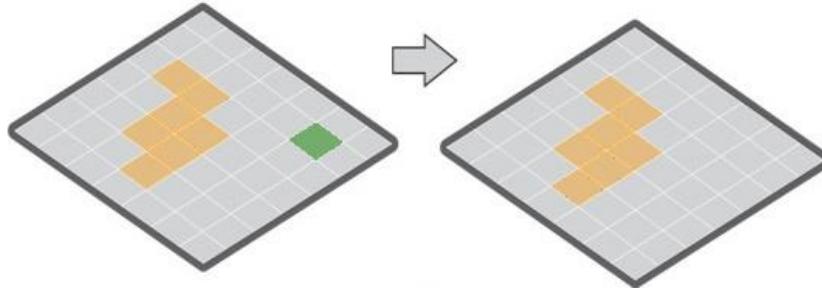


Figura 12: Esquema de la funcionalidad del filtro espacial

5.3 Filtro Temporal

Este filtro consiste en el reemplazo de los píxeles denominados como ruido o efecto sal y pimienta, los cuales representan inconsistencias que se generan en la clasificación de las coberturas respecto a su cambio paulatino año a año y la falta o distorsión de información en un mosaico. En este sentido, se utilizan 3 de filtros dentro del filtro temporal, entre ellos los de año intermedio, primer y último año, explicados a continuación.

Año intermedio: El filtro se ejecuta teniendo en cuenta el primer y último píxel de la serie temporal, cambiando la clasificación de los píxeles ubicados en los años intermedios a los clasificados en los extremos de la serie temporal (1986-2021). En la parte inferior de la figura 13, se puede observar que para la regla de 3 años se tiene un píxel clasificado como presencia de la clase de interés en color verde (bosque, agricultura, minería, etc.) en los años t_1 y t_3 , y para t_2 se presenta la ausencia de la clase de interés en color beige en este sentido, después de aplicar el filtro, éste estabiliza los cambios abruptos en la clasificación, cambiando el píxel intermedio a los clasificados por sus pares iniciales y finales. Del mismo modo, este se aplica para series temporales de 4 y 5 años, y de forma inversa como se explica en la parte superior de la figura 13. Es importante aclarar que para los años de inicio y fin del periodo analizado en este caso 1985 y 2022, no se aplica la regla de año intermedio debido a que no existen años para comparar.

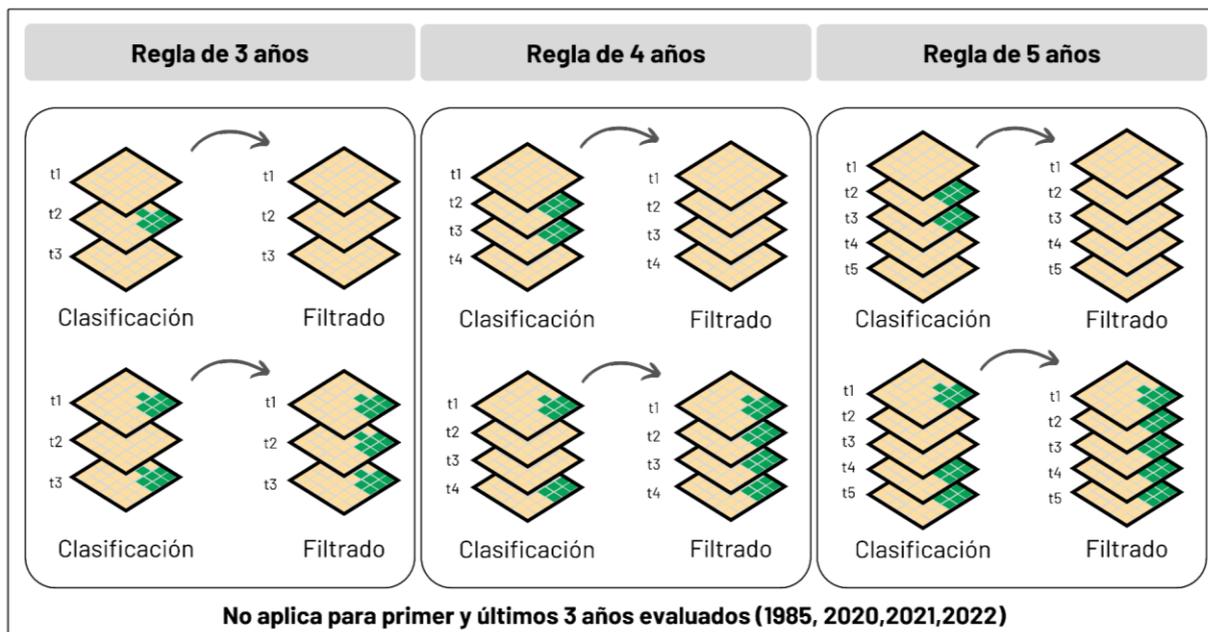


Figura 13: Efecto del filtro temporal

Año intermedio del primer y último año: La regla de primer año corrige inconsistencias de la serie temporal debido a falta de información para el año de 1985. El ejemplo enseguida muestra una transición de ausencia de áreas de la clase de interés (minería) para el primer año a presencia de áreas mineras en el primer año.

La regla de último año corrige el píxel mal clasificado al final de la serie temporal (2022) por cambios abruptos como el mencionado anteriormente, en la figura 14, se observa como el filtro actúa con el píxel clasificado al final de la serie temporal como presencia de minería (en verde) y después de ejecutado el filtro, este es reclasificado ausencia de minería homogeneizando la serie.

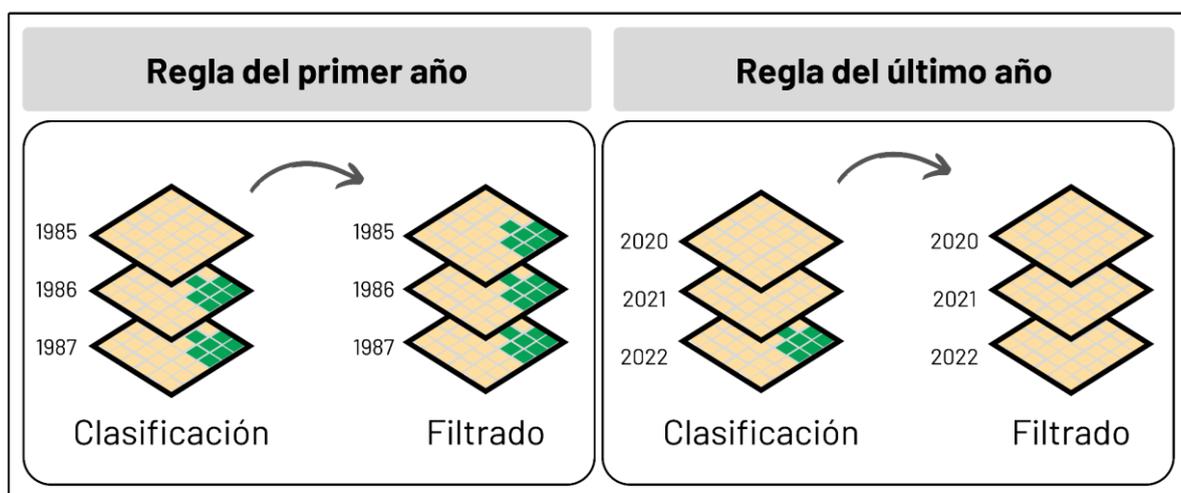


Figura 14: Filtro temporal regla de primer y último año

Debido a la naturaleza de la clase de minería, la cual puede aparecer y desaparecer en determinadas zonas a causa de que esta puede ser ilegal y no contar con los permisos para su funcionamiento constante a lo largo del tiempo, el intérprete evalúa en qué

situaciones ejecutar este filtro y su respectiva configuración para de esta forma evitar la sobrestimación o subestimación de la clase a lo largo de la serie temporal. Para concluir, el objetivo principal de este set de filtros es tener una serie temporal estable, sin saltos inconsistentes que puedan afectar los reportes estadísticos y la clasificación en sí misma, de igual manera se pueden excluir años de la serie temporal para ajustar más el comportamiento a lo observado en los mosaicos, esto ocurre principalmente en años intermedios, cuando una serie de años se encuentra estable, pero en otros periodos específicos no.

5.3 Integración

Al obtener la clasificación de las zonas mineras se hizo una primera integración de todas las zonas identificadas con minas para cada país y este resultado se integró al mapa general de clasificación nacional junto con las demás clases identificadas (figura 14). Esta integración se generó con base a una serie de reglas de prevalencia u orden de integración para cada una de las regiones (Andes, Caribe, Orinoquía y Pacífico) debido a que, en cada uno de estos, las coberturas presentan unas condiciones particulares (tablas 2 - 3 - 4 - 5).

Orinoquía		
Clase	CLASE ID	Orden integración
Minería	30	1
Infraestructura Urbana	24	2
Silvicultura	9	3
Playas dunas y arenas naturales	23	4
Agua	33	5
Palma de aceite	35	6
Acuicultura	31	7
Mosaico de agricultura o pasto	21	8
Otra área sin vegetación	25	9
FNNF inundable	11	10
FNNF inundable	11	11
Formación herbácea	12	12
Otra FNNF	13	13
Bosque inundable	6	14
Bosque inundable	6	15
Bosque	3	16

Tabla 2: Jerarquía integración en la Orinoquía

Caribe		
Clase	CLASE ID	Orden integración
Glaciar	34	1
Afloramientos rocosos	29	2
Planicie de mar hipersalina	32	3
Minería	30	4
Infraestructura Urbana	24	5
Silvicultura	9	6
Playas dunas y arenas naturales	23	7
Agua	33	8
Palma de aceite	35	9
Acuicultura	31	10
Mosaico de agricultura o pasto	21	11
Otra área sin vegetación	25	12
FNNF inundable	11	13
FNNF inundable	11	14
Otra FNNF	13	15
Manglar	5	16
Bosque inundable	6	17
Bosque inundable	6	18
Bosque	3	19
Vegetación leñosa sobre arena	49	20
Vegetación herbácea sobre arena	50	21

Tabla 3: Jerarquía integración en el Caribe

Andes		
Clase	CLASE ID	Orden integración
Glaciar	34	1
Afloramientos rocosos	29	2
Minería	30	3
Infraestructura Urbana	24	4
Silvicultura	9	5
Playas dunas y arenas naturales	23	6
Agua	33	7
Palma de aceite	35	8
Acuicultura	31	9
Mosaico de agricultura o pasto	21	10
Otra área sin vegetación	25	11
FNNF inundable	11	12
FNNF inundable	11	13
Otra formación natural no forestal	13	14
Bosque	3	15

Tabla 4: Jerarquía integración en los Andes

Pacífico		
Clase	CLASE ID	Orden integración
Minería	30	1
Infraestructura Urbana	24	2
Silvicultura	9	3
Playas dunas o arenas naturales	23	4
Agua	33	5
Palma de aceite	35	6
Acuicultura	31	7
Mosaico de agricultura y pasto	21	8
Afloramientos rocosos	29	9
Otra área sin vegetación	25	10
FNNF inundable	11	11
FNNF inundable	11	12
Otra FNNF	13	13
Manglar	5	14
Bosque inundable	6	15
Bosque inundable	6	16
Bosque	3	17

Tabla 5: Jerarquía integración en el Pacífico

Teniendo en cuenta el orden de integración presentado anteriormente, se obtiene como resultado el mapa final con la presencia de minería en las diferentes zonas del país:

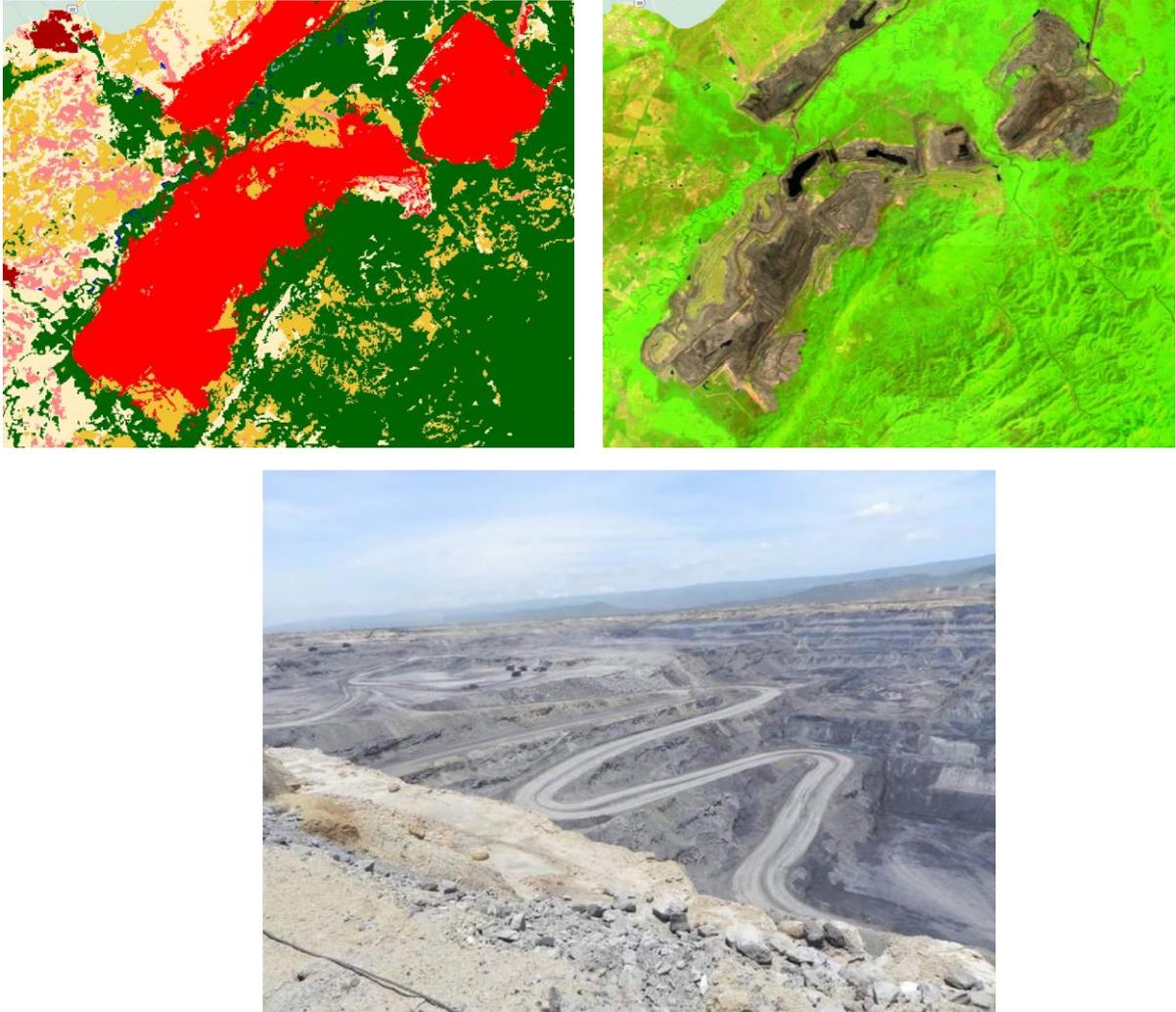


Figura 15: Integración clase minera. Izquierda, clasificación integrada de las clases identificadas en la Colección 1 MapBiomás Colombia. Derecha Imagen satelital de la zona. Abajo Imagen de paisaje de la zona minera identificada.

7. Referencias

- Breiman, L. (2001). *Random Forests. Machine Learning*, 45(1), 5-32.
<https://doi.org/10.1023/A:1010933404324>
- Congreso de Colombia. (2001). *LEY 685 DE 2001 Código de Minas*.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=9202>
- IDEAM. (2010). *Leyenda Nacional de Coberturas de la Tierra. Metodología CORINE Land Cover adaptada para Colombia Escala 1:100.000*.