

Flujos de trabajo en los scripts generados para la construcción de mosaico de imágenes Sentinel-2 y la clasificación de coberturas de la tierra

Liliana Castillo Villamor

Este documento describe los flujos de trabajo de los scripts diseñados para generar un mosaico a partir de imágenes Sentinel-2 y una superficie de coberturas de la tierra.

Los archivos, escritos en JavaScript para ser ejecutados en la plataforma de GEE se encuentran disponibles en el repositorio [sinchi](https://github.com/SERVIR-Amazonia/sinchi) en GitHub <https://github.com/SERVIR-Amazonia/sinchi>.

Los usuarios con acceso al repositorio pueden añadirlo al Editor de Código utilizando el link https://code.earthengine.google.com/?accept_repo=users/an-sig/sinchi.

1. Construcción de mosaico de imágenes Sentinel-2 para la Amazonia Colombiana.

Al ejecutar en GEE el script mosaico, éste genera un mosaico de imágenes satelitales de Sentinel-2 (Colección Harmonized Sentinel-2 MSI) dentro de un periodo definido por el usuario. Las actividades que se ejecutan dentro del archivo composite se esbozan en la Ilustración 1.

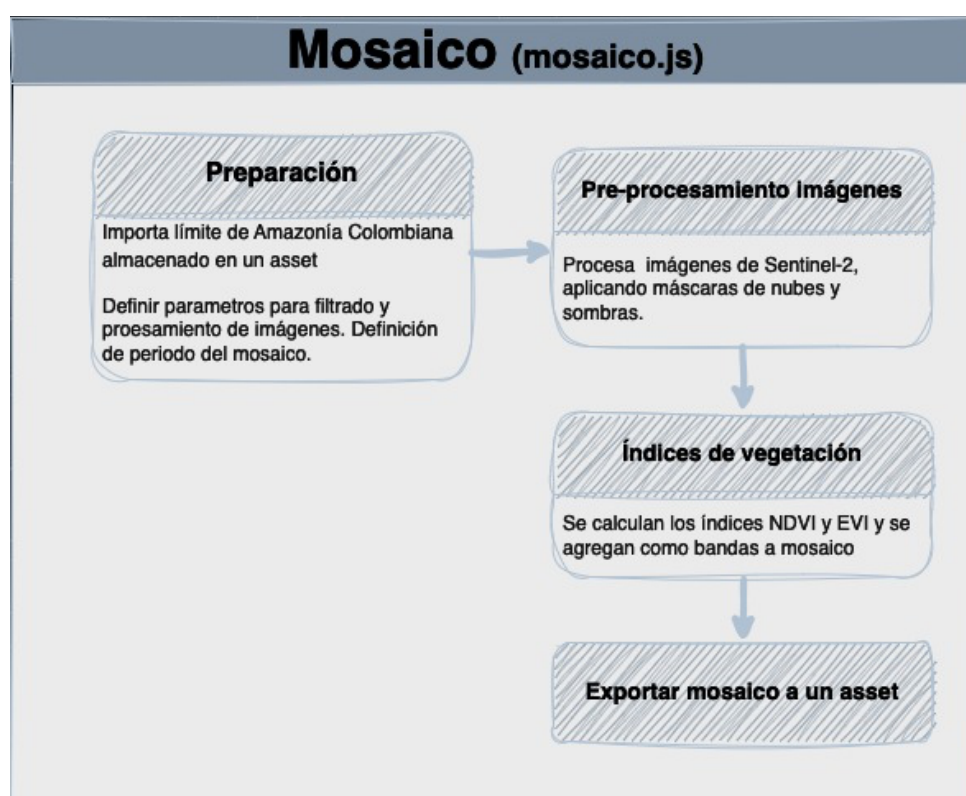


Ilustración 1. Flujo de trabajo del módulo mosaico

2. Algoritmo de clasificación de coberturas para zona piloto.

De manera general, los scripts desarrollados para realizar la clasificación de coberturas realizan los siguientes grupos de actividades:

1. Crear un mosaico de imágenes Sentinel-2 para un periodo específico
2. Generar una imagen de predictores que se incluirán para realizar la clasificación de coberturas, usando las bandas del mosaico Sentinel-2, índices de vegetación y variables geomorfológicas derivadas de un modelo de elevación de terreno (DTM).
3. Importar puntos de validación de coberturas en campo y asignar como atributos los valores de las covariables o predictores que se utilizarán en la clasificación.
4. Crear primitivos para las diferentes clases de cobertura (Probabilidad)
5. Construir mapa de coberturas a partir de la cobertura con la máxima probabilidad encontrada para un píxel.
6. Hacer una evaluación de la exactitud del modelo

Estas actividades se llevan a cabo usando las funciones que se encuentran en los módulos que se encuentran dentro de la carpeta [classification](#), disponible en el repositorio [sinchi](#) de GitHub. También se puede acceder al repositorio en GEE a través del link https://code.earthengine.google.com/?accept_repo=users/an-sig/sinchi, sin embargo, se requieren permisos de acceso.

2.1. Declaración de variables ([variables_definition](#))

En este módulo se definen las variables que se usarán en todo el proceso de tal manera que no se requiera ingresar datos en otros módulos.

Este módulo se diseñó para cambiar los valores de las variables requeridas antes de iniciar el proceso de clasificación. **NO** requiere ser ejecutado.

2.2. Preparación de los datos de campo e imágenes ([exec_prepare_data](#))

Al ejecutar el módulo [exe_prepare_data](#) se usan las funciones de los módulos [module_extract_samples](#), [composite](#) y [covariates](#) para pre-procesar imágenes de satélite Sentinel-2 y datos de puntos de muestreo con el fin de generar los insumos que se utilizarán en la clasificación de coberturas. Todo el proceso de clasificación se dividió en módulos y se generaron salidas intermedias con el fin de evitar cargar la plataforma con múltiples actividades de procesamiento simultáneas. Al ejecutar el módulo [exe_prepare_data](#) en GEE se ejecutan las actividades que se presentan en la Ilustración 2, las cuales se describen con más detalle a continuación. Es necesario haber importado con anterioridad los puntos de muestreo con la información real de la cobertura del suelo en un asset de GEE.

Este módulo se ejecuta antes de ejecutar el módulo **exec_classification**, y luego de ajustar los valores de las variables en el archivo **variables_definition**.

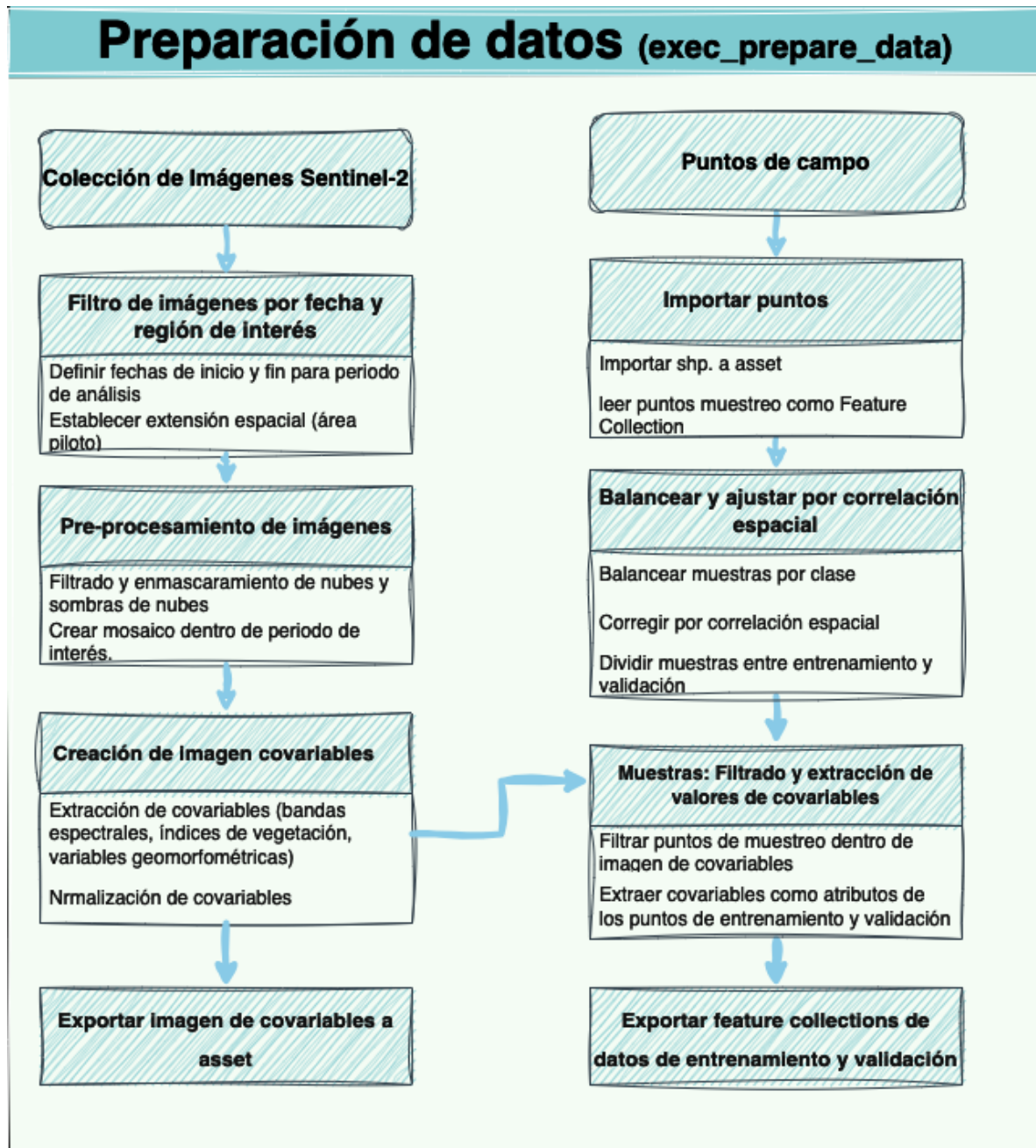


Ilustración 2. Flujo de trabajo de script `exe_prepare_data.js`

2.2.1. Inicio

- Al módulo comienza importando los módulos y define las fechas correspondientes al periodo para el cual se generará la clasificación de coberturas.

2.2.2. Importar puntos

- Se importan los datos de puntos de muestreo que contienen los valores reales de cobertura del suelo para posterior entrenamiento y validación del algoritmo. Estos puntos de muestreo se importan con feature collections para cada semestre (S1 y S2) desde una ubicación de almacenamiento (asset). Es decir que deben ser almacenados con anterioridad en el respectivo asset.

2.2.3. Balancear muestras y ajustar para evitar correlación espacial

- Las funciones almacenadas en el módulo **module_extract_samples** se usan para equilibrar las muestras por clase.
- Además, con el fin de evitar correlación espacial entre puntos de muestreo y validación, se define que las muestras estén separadas por más de una distancia específica (100 metros).
- Se generan dos Feature Collections: Una para puntos de entrenamiento y otra para almacenar los puntos de validación.

2.2.4. Filtro de imágenes de la colección Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A por fecha y área de interés.

- Se filtran las imágenes de la colección armonizada de Sentinel-2 considerando el periodo del análisis (Semestre 1 o 2) y la extensión del área piloto, la cual fue definida considerando las áreas con mayor cantidad de puntos con información de la cobertura real de la tierra.

2.2.5. Pre-procesamiento de imágenes.

El pre-procesamiento de las imágenes se realiza a través de las funciones disponibles en el módulo **composition**.

- Se realiza un filtrado de Nubes y Sombras de Nubes que pueden distorsionar el análisis
- Se crea un mosaico o composición para combinar las imágenes de la colección dentro del periodo de interés. Para cada píxel se selecciona el percentil 50 de los valores de reflectancia que toma el píxel en todas las imágenes disponibles para el rango de tiempo considerado.

2.2.6. Creación de imagen con covariables y normalización

Se genera una imagen multibanda con las covariables determinadas a partir de la imagen procesada de Sentinel-2 utilizando las funciones disponibles en el módulo **covariates**. Luego normaliza estas covariables para que se encuentren dentro del mismo rango y los algoritmos de aprendizaje de máquina tengan un mejor desempeño.

Las covariables incluyen las bandas originales, diferentes índices de vegetación y variables geomorfológicas. Se incluyó además el índice de vegetación SVVI reportado por Coulter et al. (2016), el cual mostró resultados prometedores al discriminar diferentes clases de vegetación y coberturas agrícolas en Ghana.

2.2.7. Muestras: Filtrado y extracción de valores de covariables

- Filtra los puntos de muestreo para conservar solo aquellos dentro del área de la imagen (Esto se realiza porque al trabajar con un área piloto algunos puntos quedan por fuera de la extensión de esta área).
- Prepara los datos de entrenamiento y prueba actualizando cada punto con las covariables de imagen correspondientes.

2.2.8. Salidas generadas

Las salidas del script **exe_prepare_data** son:

- **Datos de Entrenamiento** : Contiene puntos de muestreo geoespaciales que se usarán para entrenar el modelo de clasificación. Cada punto de muestreo tiene asociados como atributos las covariables derivadas de las imágenes de Sentinel-2. Además, cada punto de muestreo está etiquetado con la clase de que corresponde a la cobertura real de la tierra.
- **Datos de Prueba**: Contiene los puntos de validación en campo con sus respectivas covariables de imagen. Sin embargo, estos puntos se utilizan para evaluar la precisión del modelo de clasificación, no para entrenarlo.
- **Imagen de covariables**: Imagen multibanda con las covariables que se usarán para entrenar el modelo.

Los conjuntos de datos de entrenamiento y prueba con las covariables como atributos son exportados a una ubicación de almacenamiento definida, para facilitar el acceso y la referencia en análisis posteriores.

2.3. Clasificación de coberturas de la tierra (exec_classification)

El módulo **exec_classification** ejecuta un flujo de trabajo completo de clasificación de cobertura de tierra usando las funciones disponibles en el módulo **classify**. Utiliza un enfoque de máxima probabilidad basado en los primitivos, los cuales se construyen a partir de la imagen de covariables generada previamente con el módulo **exec_prepare_data**. Luego evalúa la exactitud de la clasificación con los datos de validación generados con módulo **exec_prepare_data**, los cuales fueron almacenados previamente en un asset. La Ilustración 3 presenta el flujo de trabajo del módulo **exec_classification**.

Este módulo se ejecuta luego de ejecutar el módulo **exec_prepare_data**, y de ajustar los valores de las variables en el archivo **variables_definition**.

2.3.1. Importar Módulos y Variables Iniciales

Importa los módulos **classify** y **exe_prepare_data** que contienen las funciones de clasificación y preparación de datos, respectivamente.

Establece variables que se utilizarán en todo el módulo, incluyendo la clase, el semestre, las rutas a los datos de entrenamiento y prueba, y la extensión del área piloto.

Carga los datos de entrenamiento y prueba actualizados y la imagen de covariables desde las rutas especificadas en el módulo **exe_prepare_data**.

2.3.2. Creación de Colección de primitivos

Utiliza la función `createLCPrimitiveCollection` del módulo **classify** para crear una colección de imágenes de primitivos de cobertura de tierra usando la imagen de covariables y los datos de entrenamiento.

2.3.3. Ensamble de máxima probabilidad

Aplica la función `assembleMaxProbability` del módulo **classify** para combinar las clasificaciones de probabilidad de la colección de primitivos en una única imagen que representa la clasificación de máxima probabilidad.

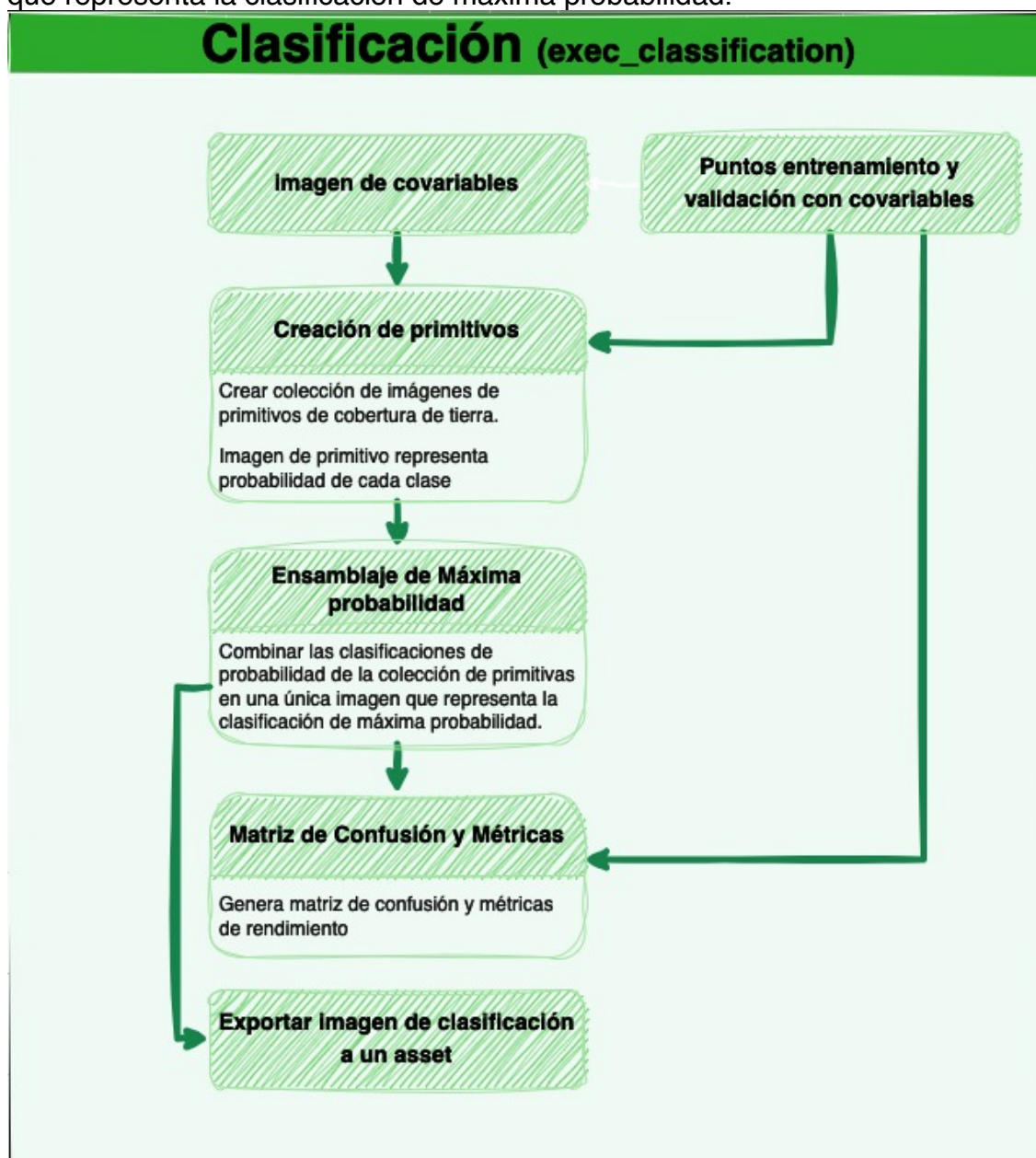


Ilustración 3. Flujo de trabajo de script `exec_classification.js`

2.3.4. Creación de Matriz de confusión y métricas

Calcula una matriz de confusión y métricas de rendimiento (i.e. la exactitud general, exactitud del productor, exactitud del consumidor y coeficiente kappa) para la imagen de clasificación ensamblada usando la función `calculateConfusionMatrixB` del módulo **classify**.

1.1.1. Exportar Imagen de coberturas

La imagen de coberturas para el periodo deseado se exporta a un asset.