

MAPA DE COBERTURAS DE LA TIERRA PARA EL AÑO 2018, SOBRE EL DEPARTAMENTO DE AHUACHAPÁN, EL SALVADOR



heig-vd

Haute Ecole d'Ingénierie et de Gestion
du Canton de Vaud

KING'S
College
LONDON



Citación sugerida:

PEREZ-ESCOBAR, J., TELLO J.J., PAZ-GARCIA, P. & LEVERON S. (2019). Mapa de coberturas de la tierra para el año 2018 sobre el departamento de Ahuahcpán, El Salvador.

CONTENIDO

SIGLAS Y ACRÓNIMOS.....	3
1. INTRODUCCIÓN	4
2. COBERTURAS DE LA TIERRA	5
3. DATOS DE ENTRADA	8
3.1. Imágenes de satélite	8
3.1.1. Landsat8.....	8
3.1.2. Sentinel-2	9
3.1.3. Puntos de entrenamiento y evaluación	10
4. METODOLOGÍA	10
4.1. Pre-proceso de imagen de satélite Sentinel-2	10
4.1.1. Corrección de imágenes	11
4.1.2. Índice de vegetación NDVI.....	11
4.1.3. Reducción de imágenes.....	11
4.1.4. Estandarización de la imagen	11
4.2. Selección de datos de entrenamiento.....	12
4.3. Proceso de clasificación.....	12
4.3.1. Clasificación no supervisada.....	12
4.3.2. Clasificación supervisada.....	13
4.3.3. Comparación de modelos de clasificación.....	15
5. RESULTADOS.....	16
5.1. Clasificación no supervisada.....	16
5.2. Clasificación supervisada.....	16
6. POST-PROCESO DE CLASIFICACIÓN	18
6.1. Filtro mayoritario	18
6.2. Refinado de límites	19
6.3. Regiones de grupos	19
6.4. Establecer nulos y relleno de datos	19
7. VALIDACIÓN	20
8. CONCLUSIÓN	23
9. BIBLIOGRAFÍA	24

SIGLAS Y ACRÓNIMOS

<i>Acrónimo</i>	<i>Definición</i>
K-means	Método K-medias
SVM	Máquina de soporte vectorial Support Vector Machine por sus siglas en inglés
RF	Árboles aleatorios Random Forest por sus siglas en inglés
MLP	Perceptrón multicapa Multi Layer Perceptrón por sus siglas en ingles
µm	Micras
nm	Nanómetros
GEE	Google Earth Engine
S2	Sentinel-2
SCS-c	Sensor de Capota Solar - c Sun Canopy Sensor – c
NDVI	Índice de vegetación de la diferencia normalizada Normalized Difference Vegetation Index por sus siglas en inglés

1. INTRODUCCIÓN

El presente documento tiene como propósito mostrar la metodología para la elaboración de mapas de coberturas de la tierra para el año 2018 en el departamento de Ahuachapán, El Salvador. A continuación, se muestra la utilidad de las imágenes satelitales y las técnicas de aprendizaje automático para la generación de mapas de coberturas.

El paso a paso de la metodología, inicia desde el pre-procesamiento de la colección de imágenes de satélite Sentinel-2, la selección de datos geográficos de muy alta resolución en Google Earth Pro y mapas de coberturas previos, con la intención de alimentar los algoritmos de aprendizaje automático, clasificar las coberturas, realizar un post-proceso de clasificación para eliminar ruido, y por último validar el producto generado.

Diferentes clasificadores supervisados y no supervisados se evaluaron con la intención de seleccionar el más adecuado para los datos de entrenamiento y así clasificar, probando desde el K-means, como método no supervisado, y los supervisados como Máquina de Soporte Vectorial, Naive Bayes, Random Forest y Perceptrón Multicapa.

Las categorías a clasificar se establecieron de acuerdo a los mapas previos generados en el departamento, como el Corine, de la Universidad del Salvador para el año 2010, el mapa de coberturas de uso del suelo del MARN para el año 2012, y el Mapa de cobertura forestal y uso de la tierra del CCAD-GIZ para el año 2015. Esto se hizo para llevar una continuidad en las cubiertas de la tierra y hacerlos comparables en el tiempo. El mapa está clasificado en dos niveles, donde el primero está compuesto por las categorías de Bosque y no bosque, mientras que el segundo nivel es más detallado y está compuesto por 15 categorías, desglosando las que se encuentran en el primer nivel.

2. COBERTURAS DE LA TIERRA

Las coberturas de la tierra hace referencia a los elementos que componen la superficie de la terrestre, ya sean naturales o creados por la mano del hombre (IDEAM, 2014), identificando los patrones de la naturaleza y las actividades sociales, que segmentan las cubiertas de acuerdo al uso que le dan al suelo. (Sharma, Liu, & Xiaojun, 2018)



Las coberturas mapeadas sobre la región, se encuentran en la Tabla 1, que se presenta a continuación:






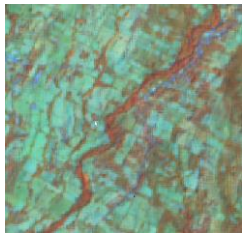




Tabla 1. Coberturas de la tierra


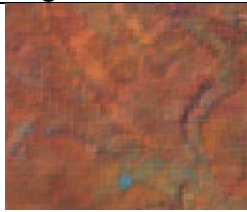









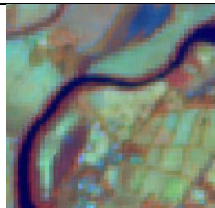
Coberturas nivel 1	Coberturas nivel 2	Etiquetas
<i>Bosque</i>	Bosque caducifolio	BC
	Bosque de mangle	BM
	Bosque siempre verde	BSV
	Bosque de galería	BG
<i>No bosque</i>	Plantaciones de bosque monoespecífico	PBM
	Palmeras oleíferas	PO
	Café	CF
	Vegetación secundaria	VS
	Frutales	FR
	Caña de azúcar	CAZ
	Mosaico de pastos y cultivos	MPYC
	Cuerpos de agua	CA
	Suelo desnudo	SD
	Urbano	U
	Playas	PL







Estas coberturas tienen una definición establecida y una manera de identificarse en campo, además de apreciarse fácilmente sobre imágenes satelitales de acuerdo a la combinación de bandas. Esto se presenta a continuación en la tabla 2.

Tabla 2. Definición de cobertura de la tierra, visualización en campo y sobre imágenes de satélite sentinel-2

Cobertura	Definición	Imagen de cobertura	Imagen en sentinel-2
<i>Bosque caducifolio</i>	Corresponde a los bosques que botan las hojas en por temporadas. se encuentran en zonas con precipitaciones anuales de 1000 a 2000 mm. (WWF, 2019)	 <p>(Bissanti, 2017)</p>	 <p>Bandas 453</p>

Cobertura	Definición	Imagen de cobertura	Imagen en sentinel-2
Bosque de mangle	El bosque de mangle está formado por arboles esclerófilos siempre verdes de raíces zancos. Su altura varía entre 5 y 30 metros.	 (ECOicias.com-PNUMA, 2010)	 Bandas 453
Bosque siempre verde	Corresponde a la especie de bosque del cual el 75% de su cobertura no pierde las hojas. Sus árboles alcanzan alturas entre los 40 y 50 metros. (REDD-CCAD-GIZ, 2011)	 (REDD-CCAD-GIZ, 2011)	 Bandas 453
Bosque de galería	Es aquella vegetación arbórea que se encuentra a las orillas de los ríos y normalmente están ubicadas en las áreas de sabanas. (Instituto Sinchi, 2009)	 (Instituto Sinchi, 2009)	 Bandas 453
Plantaciones de bosque monoespecifico	Las plantaciones de bosque monoespecifico corresponden a un tipo de cultivo arbóreo que contiene una misma especie. (FAO, 2010)	 (Espino Ortega, 2017)	 Bandas 453
Palmeras oleíferas	Las palmas son un tipo de vegetación que pertenece a clima cálido, plantadas hasta los 500 metros sobre el nivel del mar.	 (Sánchez, 2018)	 Bandas 453
Café	Los cafetales hacen parte de los cultivos perennes que se siembran en condiciones climáticas y topográficas específicas, en donde la temperatura debe oscilar entre los 17 y 23 grados centígrados y los 1100 y 1800 metros		

Cobertura	Definición	Imagen de cobertura	Imagen en sentinel-2
	sobre el nivel medio del mar. (Federación Nacional de cafeteros de Colombia, 2010)	 (Urquillo, 2018)	 Bandas 453
Vegetación secundaria	Corresponde por lo general a la sucesión vegetal de los bosque que fueron talados, en donde algunos de sus árboles presentan menor tamaño mezclándose con matorrales.	 (Instituto Sinchi, 2009)	 Bandas 453
Frutales	Los frutales corresponden a cierto tipo de plantaciones de árboles que producen frutas.	 (García Abadía, 2014)	 Bandas 453
Caña de azúcar	Cultivo que pertenece a la familia de las gramíneas, que tiene como característica un tallo leñoso, lleno de tejido esponjoso y dulce del que se extrae azúcar. (Procaña, 2019)	 (Diario Electrónico Cuarto Poder, 2018)	 Bandas 453
Mosaico de pastos y cultivos	Corresponde a las zonas utilizadas para la actividad de agrícola y ganadera, comúnmente con algunas áreas de vegetación secundaria.	 (Instituto Sichi, 2009)	 Bandas 453
Cuerpos de agua	Son todos los cuerpos de agua, entre los que se encuentran los ríos, las lagunas naturales, lagunas artificiales y los mares.	 (Álvarez G., 2016)	 Bandas 453

Cobertura	Definición	Imagen de cobertura	Imagen en sentinel-2
Suelo desnudo	Zonas sin evidencia de vegetación, ya sea por causas naturales, o provocadas por el hombre.	 (La estrella de Panamá, 2017)	 Bandas 321
Urbano	Zona conformada por edificaciones con una baja proporción de vegetación en comparación a las áreas construidas	 (Diario el Mundo, 2017)	 Bandas 321
Playas	Acumulaciones de arenas en las zonas de costa, adyacentes al mar	 (elsalvador.com, 2018)	 Bandas 321

3. DATOS DE ENTRADA

3.1. Imágenes de satélite

3.1.1. Landsat8

Landsat8 es la colección más reciente de la Landsat, con datos ópticos y térmicos, lo que permite calcular una gran cantidad de índices espectrales que son útiles en áreas como la agricultura, deforestación, mapeo de coberturas, entre otras. Estas imágenes poseen resolución temporal de 16 días, mientras que la resolución espacial es de 30 metros.

Las características de la imagen Landsat8 se encuentran a continuación:

Tabla 3. Características de la colección Landsat8. Fuente: (Google Earth Engine, 2017)

Banda	Nombre	Longitud de onda	Resolución espacial
B1	Aerosol costero	0.43 – 0.45 (µm)	30 (metros)
B2	Azul	0.45 – 0.51 (µm)	30 (metros)

Banda	Nombre	Longitud de onda	Resolución espacial
B3	Verde	0.53 – 0.59 (μm)	30 (metros)
B4	Rojo	0.64 – 0.67 (μm)	30 (metros)
B5	Infrarrojo cercano	0.85 – 0.88 (μm)	30 (metros)
B6	Infrarrojo de onda corta 1	1.57 – 1.65 (μm)	30 (metros)
B7	Infrarrojo de onda corta 2	2.11 – 2.29 (μm)	30 (metros)
B8	Pancromática	0.52 – 0.90 (μm)	15 (metros)
B9	Cirro	1.36 – 1.38 (μm)	15 (metros)
B10	Infrarrojo térmico 1	10.60 – 11.19 (μm)	30 (metros)
B11	Infrarrojo térmico 2	11.50 – 12.51 (μm)	30 (metros)
BQA	Banda de calidad		

De esta lista de bandas, únicamente se usó la Banda 10 que corresponde a una de las bandas térmicas, debido a que acordando con la USGS, la B11 contiene grandes incertidumbres en su calibración (Zhao, et al., 2019). Este dato se requiere para alimentar y reforzar el modelo de entrenamiento.

3.1.2. Sentinel-2

Sentinel-2 está compuesta por satélites que capturan imágenes de alta resolución espacial y multiespectral al servicio de la ciencia. Esta información contiene 13 bandas espectrales que se capturan cada 5 días, además de la banda de calidad que contiene datos que enmascaran las nubes.

Las características de las imágenes son las siguientes:

Tabla 4. Características de la colección Sentinel-2. Fuente: (Google Earth Engine, 2017)

Banda	Nombre	Longitud de onda	Resolución espacial
B1	Aerosoles	443.9nm	60 metros
B2	Azul	496.6nm	10 metros
B3	Verde	560nm	10 metros
B4	Rojo	664.5nm	10 metros
B5	Borde rojo 1	703.9nm	20 metros
B6	Borde rojo 2	740.2nm	20 metros
B7	Borde rojo 3	782.5nm	20 metros
B8	Infrarrojo cercano	835.1nm	10 metros
B8a	Borde rojo 4	864.8nm	20 metros
B9	Vapor de agua	945nm	60 metros
B10	Cirro	1373.5nm	60 metros
B11	Infrarrojo de onda corta 1	1613.7nm	20 metros
B12	Infrarrojo de onda corto 2	2202.4nm	20 metros
QA10	Siempre vacío		10 metros
QA20	Siempre vacío		20 metros
QA60	Máscara de nube		60 metros

Las bandas usadas para la clasificación son la B2, B3, B4, B8, B11 y B12, que corresponden al espectro azul, verde, rojo, infrarrojo cercano, infrarrojo de onda corta 1, e infrarrojo de onda corta 2. Además, la banda QA60, máscara de nube, que se usa para realizar la limpieza de nubes y cirrus.

3.1.3. Puntos de entrenamiento y evaluación

Los puntos de evaluación se crean a partir de referencias como mapas de coberturas de la tierra de años previos, además de imágenes de alta resolución disponibles sobre Google Earth Pro. Estas combinaciones de insumos son complementarias, y permiten entender cómo se ven las coberturas sobre el terreno real sobre las imágenes de Google Earth Engine. Es necesario identificar de manera correcta las coberturas, debido a que estas en el tiempo cambian debido a las actividades antrópicas.

4. METODOLOGÍA

La elaboración del mapa de cobertura consiste en ejecutar una serie de pasos que inicia desde la colección y elaboración de datos de entrada, pre-proceso de imágenes de satélite, clasificación de coberturas y validación. (Figura 1)

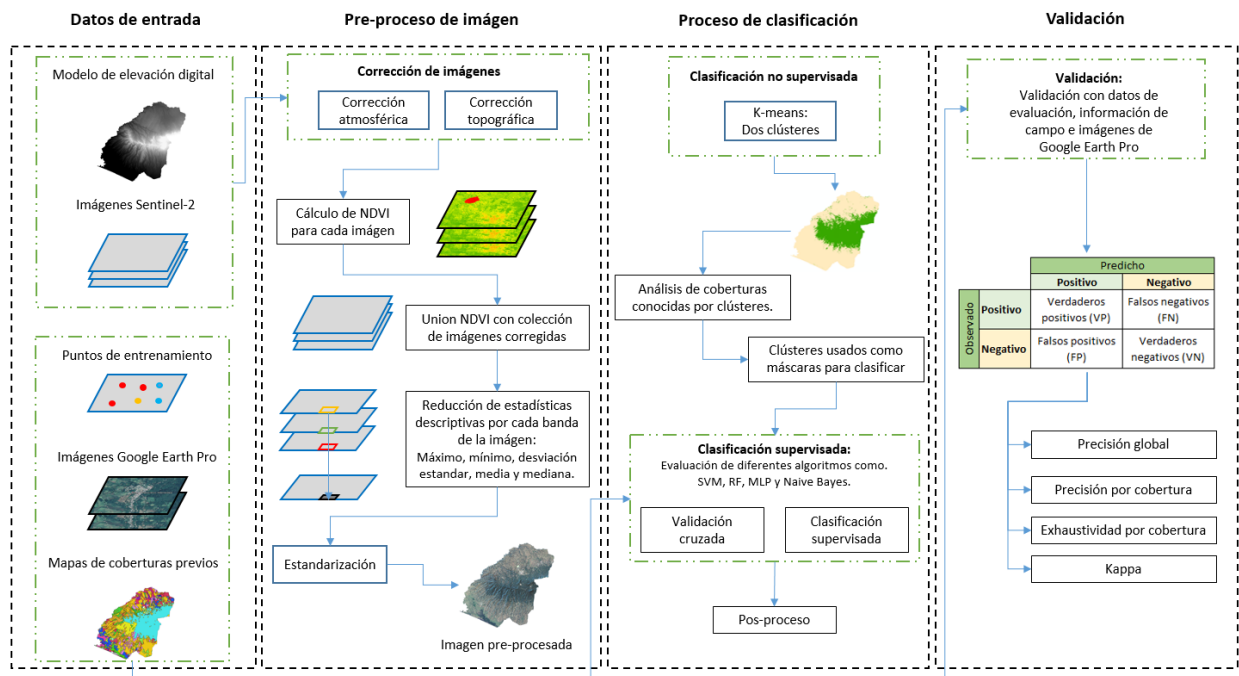


Figura 1. Metodología para elaboración de mapa de coberturas de la tierra.

4.1. Pre-proceso de imagen de satélite Sentinel-2

Para obtener una imagen satelital limpia y corregida de errores atmosféricos y topográficos sobre el territorio de Ahuachapán, se procesaron 142 imágenes Sentinel-2 del año 2018 con resolución espacial de 10 metros, sobre la plataforma GEE.

Dentro de cada imagen S2 disponible en la colección de imágenes, se encuentran diferentes bandas espectrales, y de esta se seleccionaron 8 de cada una, que corresponden a la banda azul, banda verde, banda roja, banda infrarrojo cercano, banda infrarroja de onda corta 1, banda infrarroja de onda corta 2 y la banda de calidad.

4.1.1. Corrección de imágenes

A cada una de las imágenes se les aplicó las siguientes correcciones:

- Corrección atmosférica: Corrección realizada para la eliminación de nubes y de sombras, procesada a través de la plataforma de GEE, haciendo uso de la banda de calidad o máscara de nubes llamada QA60.
- Corrección Topográfica: Esta corrección se realizó a partir del método SCS-c, el cual facilita la eliminación de los efectos de sombras, permitiendo que las coberturas similares no presenten grandes diferencias por efecto del terreno. (Telliet, Guindon, & Goodenough, 1982). Su fórmula se explica de la siguiente manera:

$$\rho_{\lambda,h,i} = \rho_{\lambda,i} * \left(\frac{\cos \theta_s + c_\lambda}{\cos \gamma_i + c_\lambda} \right) \quad (1)$$

Donde:

$\rho_{\lambda,h,i}$ = Reflectancia superficial corregida topográficamente

$\rho_{\lambda,i}$ = Reflectancia superficial de la banda espectral (radianes)

γ_i = Ángulo de iluminación (radianes)

c_λ = constante empírica para cada banda, generada a partir de la regresión lineal entre el ángulo de iluminación y cada banda espectral. (Marina-Ramona, Bogdan, Bogdan-Andrei, Constantin, & Lonut, 2017)

4.1.2. Índice de vegetación NDVI

El índice de la diferencia normalizada de la vegetación, se genera para cada una de las imágenes de la colección S2 con la intención de separar de mejor manera las diferentes coberturas, de acuerdo al nivel de verdor. Este se calcula a través de la siguiente ecuación.

$$NDVI = \frac{\text{infrarojo cercano} - \text{rojo}}{\text{infrarojo cercano} + \text{rojo}} \quad (2)$$

4.1.3. Reducción de imágenes

El proceso consiste en reducir a una imagen las 142 correspondientes de la colección. Esto quiere decir que cada banda espectral es llevada a cinco métricas estadísticas como lo son la media, mediana, desviación estándar, mínimo y máximo, para obtener una sola imagen con 41 bandas espectrales. Esto se calcula para identificar de mejor manera los diferentes comportamientos que tiene la tierra para cada una de sus coberturas, ya que este resultado es una composición que representa todo el año 2018.

4.1.4. Estandarización de la imagen

Este paso se realiza como pre-proceso de los datos en el momento de entrenar un clasificador (Kuhn, 2018), para eliminar la información que generen ruido en cada una de las bandas espectrales.

$$\frac{((img - m) / (3 * sd)) + 1}{2} \quad (3)$$

Donde:

- img: imagen Sentinel-2 con cada una de sus bandas
- m: media
- sd: desviación estándar

4.2. Selección de datos de entrenamiento

La selección de datos de entrenamiento se realizó a través de información secundaria tales como:

- Corine, Universidad del Salvador para el año 2010.
- Cobertura uso del suelo, MARN, para el año 2012.
- Mapa de cobertura forestal y uso de la tierra para el año 2015, CCAD-GIZ.

De las anteriores fuentes de información se seleccionó como dato base, la clasificación de coberturas del MARN, 2012, pero se tiene en cuenta las otras fuentes para corroborar la escogida.

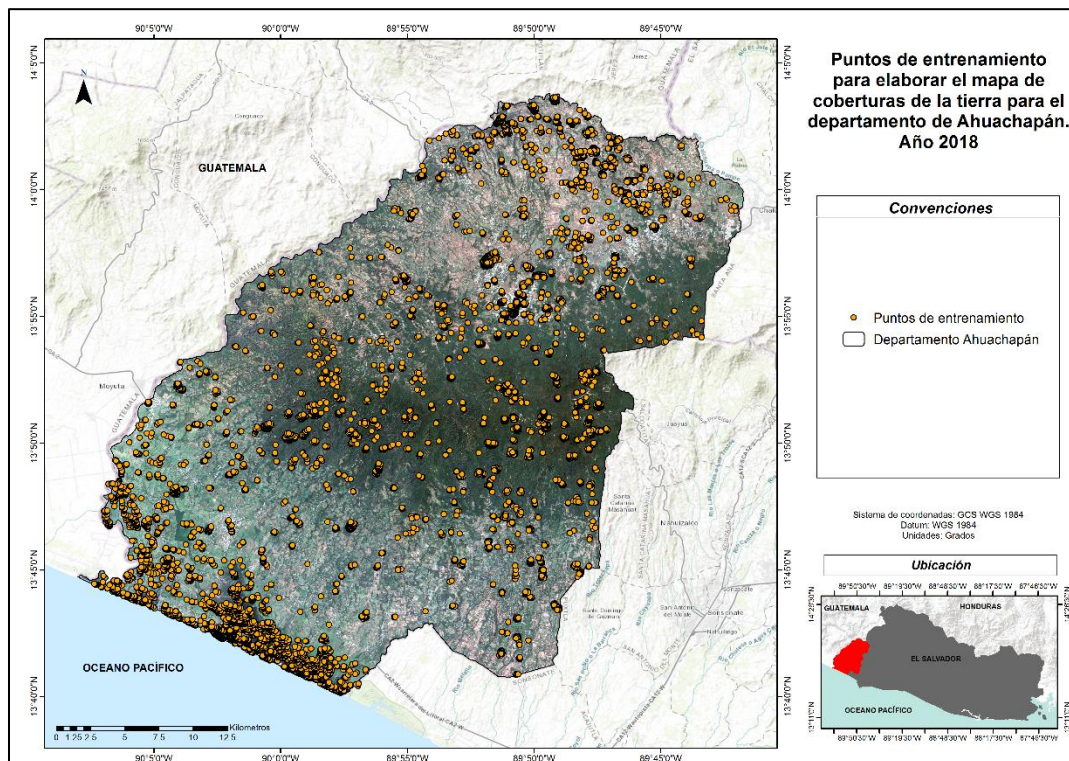


Figura 2. Puntos de entrenamiento para elaborar el mapa de coberturas de la tierra para el departamento de Ahuachapán.

4.3. Proceso de clasificación

4.3.1. Clasificación no supervisada

La clasificación no supervisada es utilizada para asociar un set de datos en N cantidades de grupos de acuerdo a sus valores. Esta cantidad de grupos se establece según el tipo de información que se tenga, el método deseado y con diferentes índices estadísticos.

El método de agrupación con el que se realiza la clasificación no supervisada es el K-means, debido a que ha sido uno de los más utilizados por diferentes autores, obteniendo resultados satisfactorios. Inicialmente, se estableció la cantidad de clústeres mediante el programa R-Studio a través de la herramienta NbCluster del paquete NbClust, sobre el cual se evaluaron 30 diferentes métodos y se

determinaron la cantidad optima de agrupaciones. El resultado determinó que la cantidad optima de clúster para la imagen es de dos (2) agrupaciones, que son introducidas al modelo de clasificación no supervisado.

Existen diferentes tipos de métodos para implementar el K-means así como se evidencia y se explica en la figura 3, pero para este caso, se seleccionó el agrupamiento plano sin generalizar el modelo, que es el que viene por defecto.

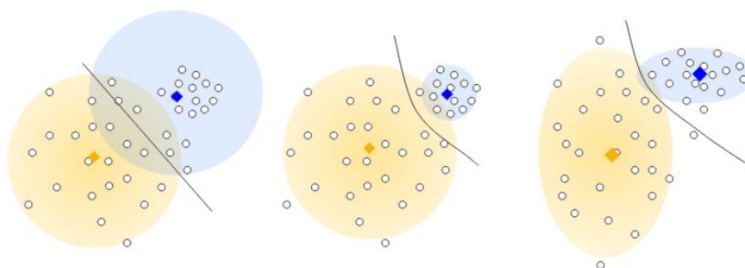


Figura 3. Diferentes tipos de K-means. En el gráfico izquierdo, se puede evidenciar un cluster no intuitivo. En el gráfico central, se puede establecer diferentes anchos de cluster, haciéndolo más intuitivo. En el gráfico derecho se muestra diferentes tamaños de grupos y anchos por dimensión. Fuente: (Google Developers, 2019)

4.3.2. Clasificación supervisada

Diferentes clasificadores se deben probar para seleccionar cual es el que mejor se comporta de acuerdo al set de datos que se quiere clasificar. Por esta razón, se probaron clasificadores como el SVM, Perceptrón Multicapa, Random Forest y NB, usando una validación cruzada con cada uno, dividiendo los datos de entrenamiento en cinco grupos y clasificando la misma cantidad de veces, con el objetivo de generar 5 precisiones por clasificador, promediarlas y determinar cuál es el mejor promedio.

4.3.2.1. SVM

SVM es un método que mapea la información con la que se entrena el modelo en hiperplanos, que separan las clases a clasificar, minimizando los errores. Teóricamente, este se entrena con los datos que definen el borde del hiperplano máximo de margen, descartando el resto de información y logrando una alta precisión con pocos datos de entrada. (Zhen et al., 2015)

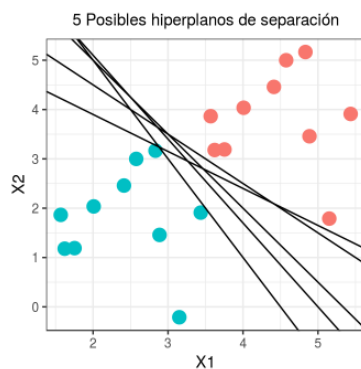


Figura 4. Hiperplanos posibles de separación entre clases. Fuente: (Amat, 2017)

4.3.2.2. Naive Bayes

Naive Bayes es un clasificador supervisado que basado en el teorema de bayes, y en la hipotesis de que sus predictores son independientes. Esto quiere decir que la presencia de una característica no está asociada o relacionada con la presencia de otra (Pereira-Toledo et al, 2017). La ecuación en la que se basa es la siguiente:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (4)$$

Donde $P(x|c)$ es la probabilidad posterior de la clase objetivo c , dado el predictor x , $P(c)$ es la probabilidad previa de clase y $P(x)$ es la probabilidad previa del predictor.

4.3.2.3. Perceptrón multicapa

Perceptrón multicapa es un tipo de red neuronal poderosa y popular. Este tipo de arquitecturas están formadas por capas de entrada y capas ocultas que se establecen arbitrariamente, más una capa de salida. La capa de entrada contiene la misma cantidad de neuronas que las etiquetas que se representarán. Como clasificador supervisado, usa las etiquetas de entrada como variable de respuesta para representarla en la capa de salida como variable categórica. (Mejías et al., 2013)

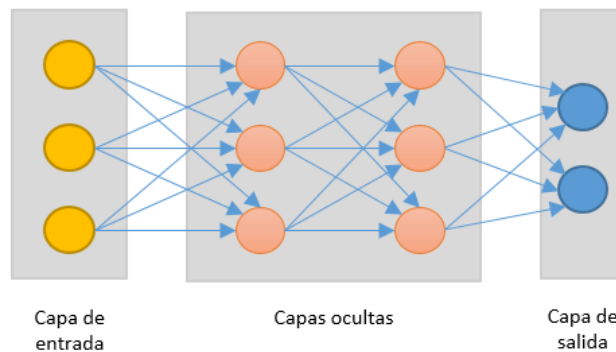


Figura 5. Perceptrón multicapa.

4.3.2.4. Random Forest

Random Forest es un método de aprendizaje automatizado, el cual construye árboles de decisión que cuando se combinan encuentran una variable de respuesta confiable. Cada árbol de decisión usa un subconjunto de datos de la muestra original, con el cual genera diversos entrenamientos y evaluación para calcular la precisión del algoritmo (Uhorchak, 2017).

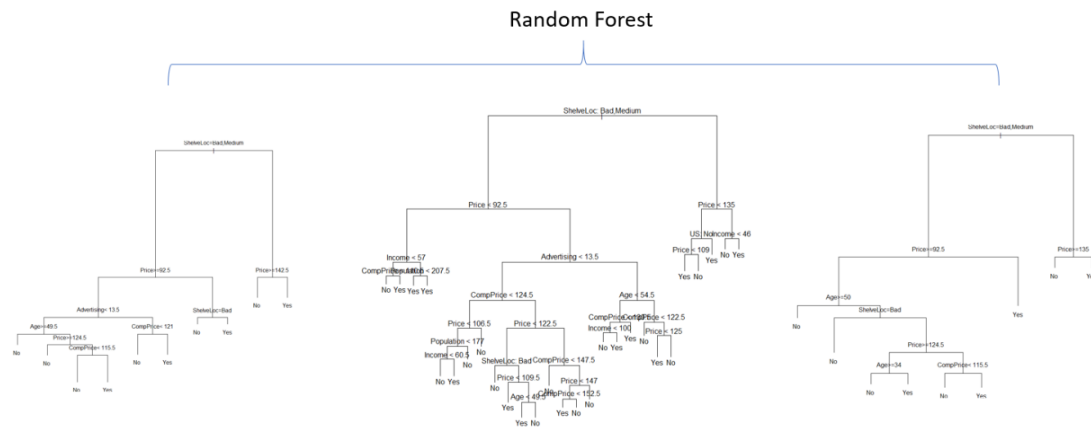


Figura 6. Ejemplo de Random Forest. Fuente: (Uhorchak, 2017)

4.3.3. Comparación de modelos de clasificación

La comparación entre los modelos de clasificación, se realizó a través de dos plataformas de procesamiento como Google Earth Engine y R-Studio, utilizando métricas como la precisión global y el índice Kappa debido a que son las comúnmente usadas para seleccionar el modelo adecuado a los datos de entrenamiento.

Tabla 5. Métricas estadísticas para comparación de modelos

	PRECISIÓN				KAPPA			
	Mín.	Mediana	Media	Máx.	Mín.	Mediana	Media	Máx.
SVM	0.8157	0.8178	0.8178	0.8198	0.8015	0.8037	0.8037	0.8060
RF	0.8550	0.8611	0.8611	0.8673	0.8438	0.8504	0.8504	0.8569
MLP	0.7150	0.7205	0.7205	0.7259	0.6927	0.6986	0.6986	0.7045
NB	0.6635	0.6846	0.6847	0.7143	0.6465	0.6631	0.6632	0.6844

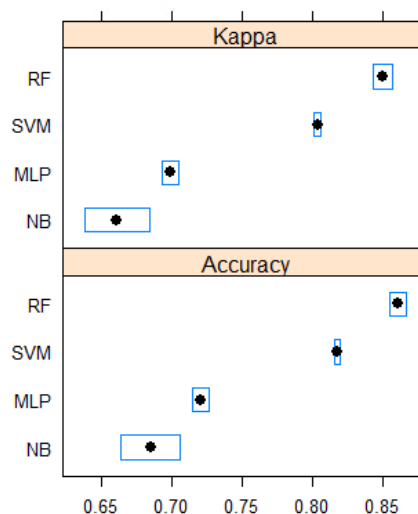


Figura 7. Métricas estadísticas para comparación de modelos

El resultado que se evidencia en la tabla 5, muestra que el algoritmo que mejor se adapta a los datos de entrenamiento es el Random Forest, y tomando como punto de comparación la mediana, este modelo alcanzó un 86%, superando por el 5% al SVM, 18% NB y 14% al MLP. Por esta razón, se aplica el RF para generar el mapa de coberturas de la tierra.

5. RESULTADOS

5.1. Clasificación no supervisada

Como primer resultado, la clasificación no supervisada se realizó con dos (2) clústeres, en los que se parecía que las variables de temperatura y elevación tuvieron gran influencia en los resultados, ya que las zonas con mayor área y menor temperatura se encuentran en un mismo grupo. (ver figura 8)

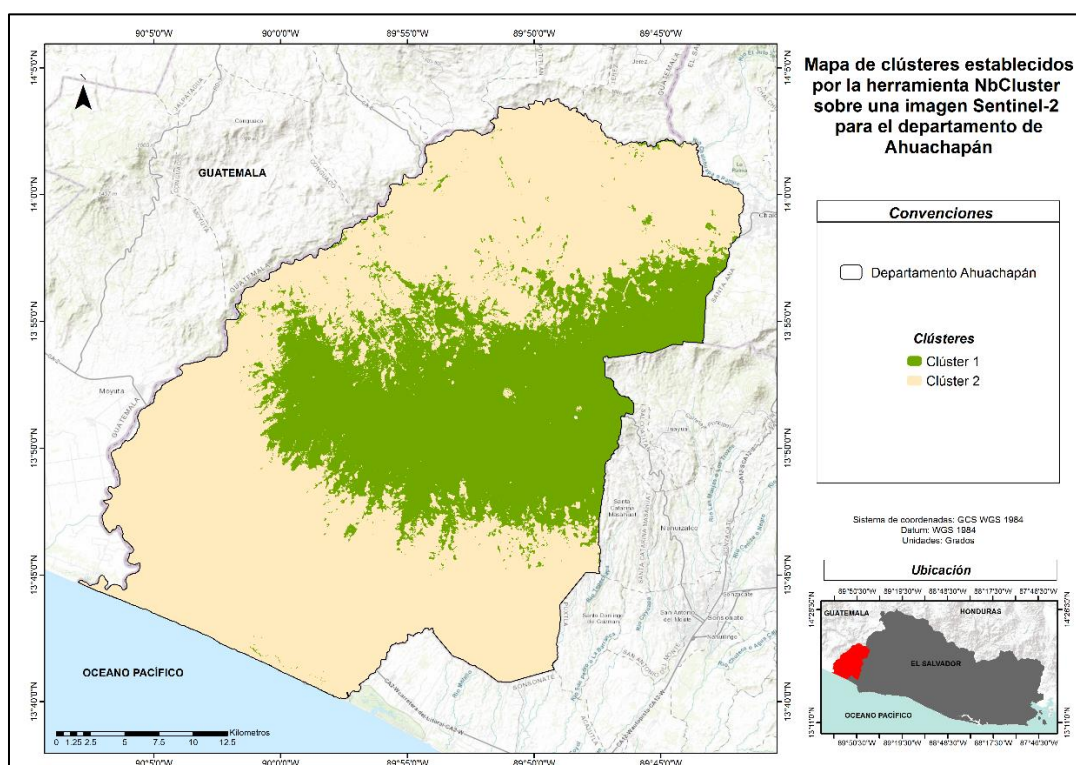


Figura 8. Mapa de clústeres establecidos por la herramienta NbCluster sobre una imagen Sentinel-2 para el año 2018 en el departamento de Ahuachapán, El Salvador.

5.2. Clasificación supervisada

Posteriormente, los clústeres generados son usados como máscaras para clasificar solamente las coberturas que se encuentren en esa región, obteniendo los resultados que se evidencian en la figura 9 y figura 10.

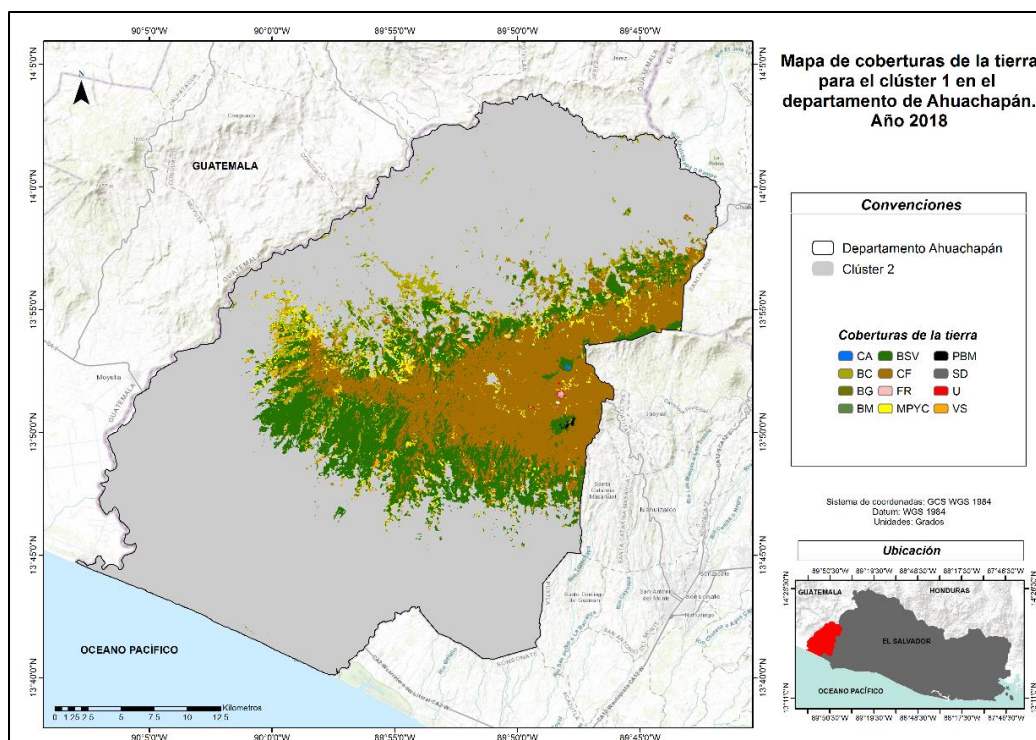


Figura 9. Mapa de coberturas de la tierra para el clúster 1 en el departamento de Ahuachapán. Año 2018

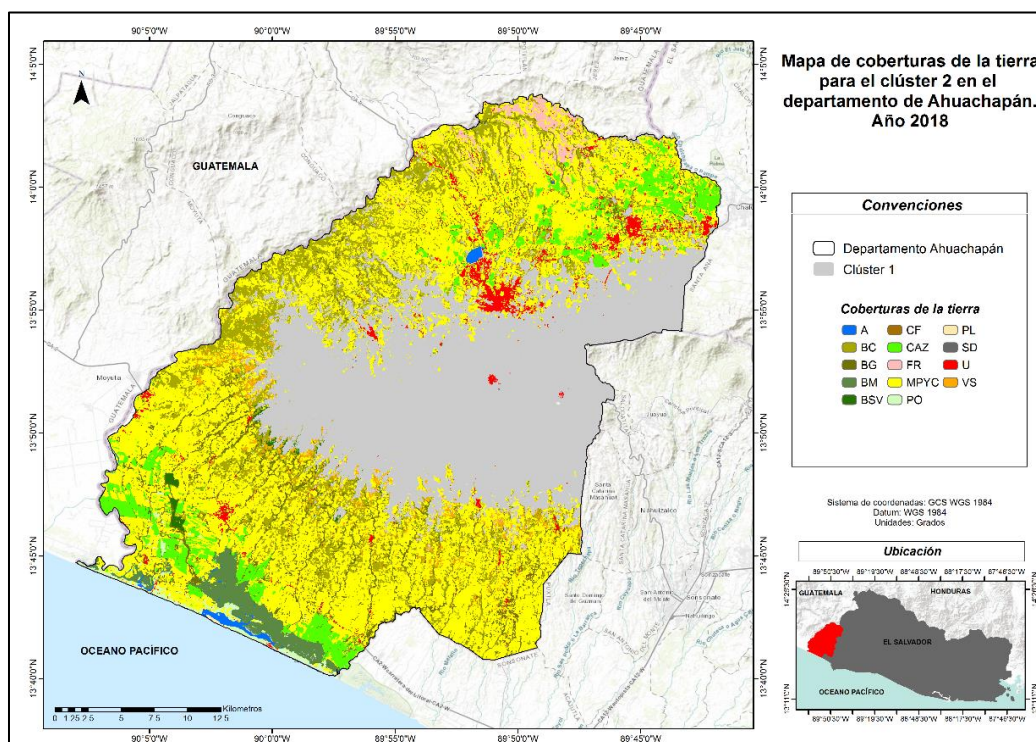


Figura 10. Mapa de coberturas de la tierra para el clúster 2 en el departamento de Ahuachapán. Año 2018

Con estas coberturas por clúster, se realiza un mosaico para obtener un solo mapa que represente el departamento, que será post-procesada con la intención de eliminar el efecto de “sal y pimienta” y substraer las pequeñas regiones mal clasificadas.

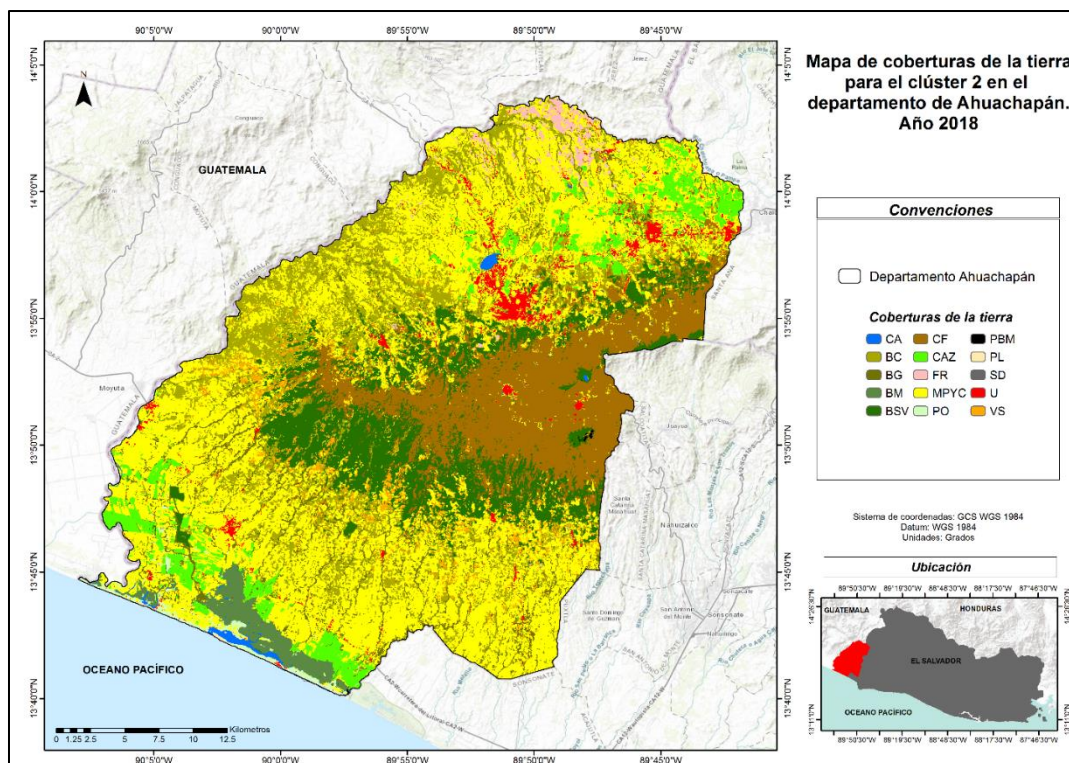


Figura 11. Mapa de coberturas de la tierra en el departamento de Ahuachapán. Año 2018

6. POST-PROCESO DE CLASIFICACIÓN

El proceso de post-clasificación, es una manera de eliminar los píxeles mal clasificados, como regiones aisladas o no conectadas. Esto se realiza a través de pasos como el filtro mayoritario, el refinado de límites y eliminación de pequeñas regiones de grupos.

6.1. Filtro mayoritario

Este proceso se realiza para eliminar píxeles que se encuentran aislados, reemplazando su valor de acuerdo a los 8 valores vecinos más próximos. (ESRI, 2016).

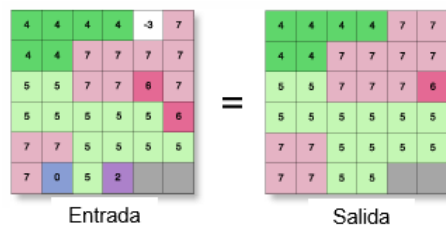


Figura 12. Filtro mayoritario. Fuente: (ESRI, 2016)

6.2. Refinado de límites

Este proceso se usa para refinar los bordes a través de encoger o comprimir algunas regiones. Las zonas de mayor prioridad invaden las zonas vecinas de menor prioridad en una ventana de 3x3.

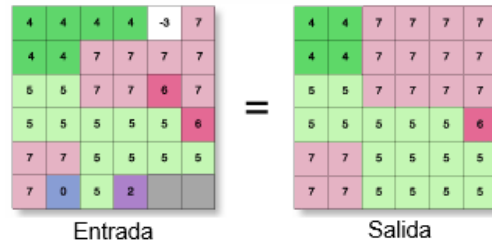


Figura 13. Refinado de límites. Fuente: (ESRI, 2016)

6.3. Regiones de grupos

Este proceso se realiza para agrupar regiones de celdas con un mismo valor, creando para cada región un ID y un conteo de las celdas por identificador. Se realiza con la finalidad de identificar cuáles son las regiones más pequeñas, reemplazarlas y establecerlas como nulos para posteriormente cambiar su valor. (ESRI, 2016)

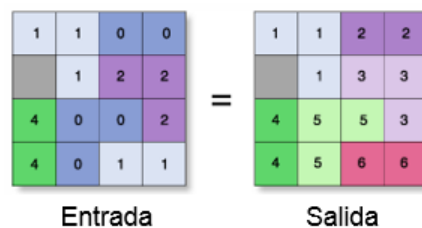


Figura 14. Regiones de grupos. Fuente: (ESRI, 2016)

6.4. Establecer nulos y relleno de datos

Las regiones que presentan un menor tamaño se establecen como valores nulos, suponiendo que son regiones aisladas (ESRI, 2016), a las que posteriormente se les reasigna un valor de pixel de acuerdo a los vecinos más próximos. (ESRI, 2016)

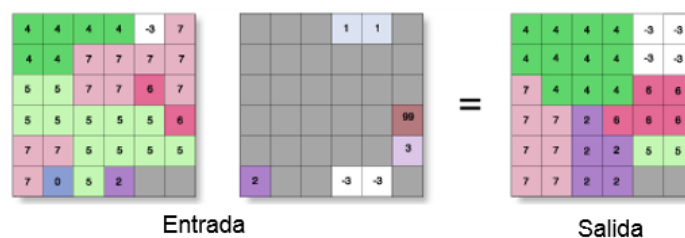


Figura 15. Establecer nulos y relleno de datos. Fuente: (ESRI, 2016) & (ESRI, 2016)

7. VALIDACIÓN

En campo, 464 puntos se validaron mediante el uso de vehículos aéreos no tripulados, capturando imágenes de muy alta resolución, además de 500 puntos extra, validados con imágenes de alta resolución de GEP. Los puntos evaluados se seleccionaron de acuerdo a las metodologías de Oloffson, (2014), la cual recomienda utilizar un muestreo estratificado, de acuerdo al tamaño de cada una de las coberturas de la tierra a validar.

La fórmula para seleccionar las muestras aleatorias en cada una de las coberturas, identificar el peso que tiene cada cubierta en la validación y determinar la cantidad de puntos a validar, es la siguiente

$$\frac{\frac{Z^2 * p(1 - p)}{e^2}}{1 + \left(\frac{Z^2 * p(1 - p)}{e^2 N} \right)} \quad (5)$$

Donde Z corresponde al z-score, p corresponde al 0.5, e es el error permitido que varía entre 0 y 10% y N es la población total, que contiene todas las coberturas.

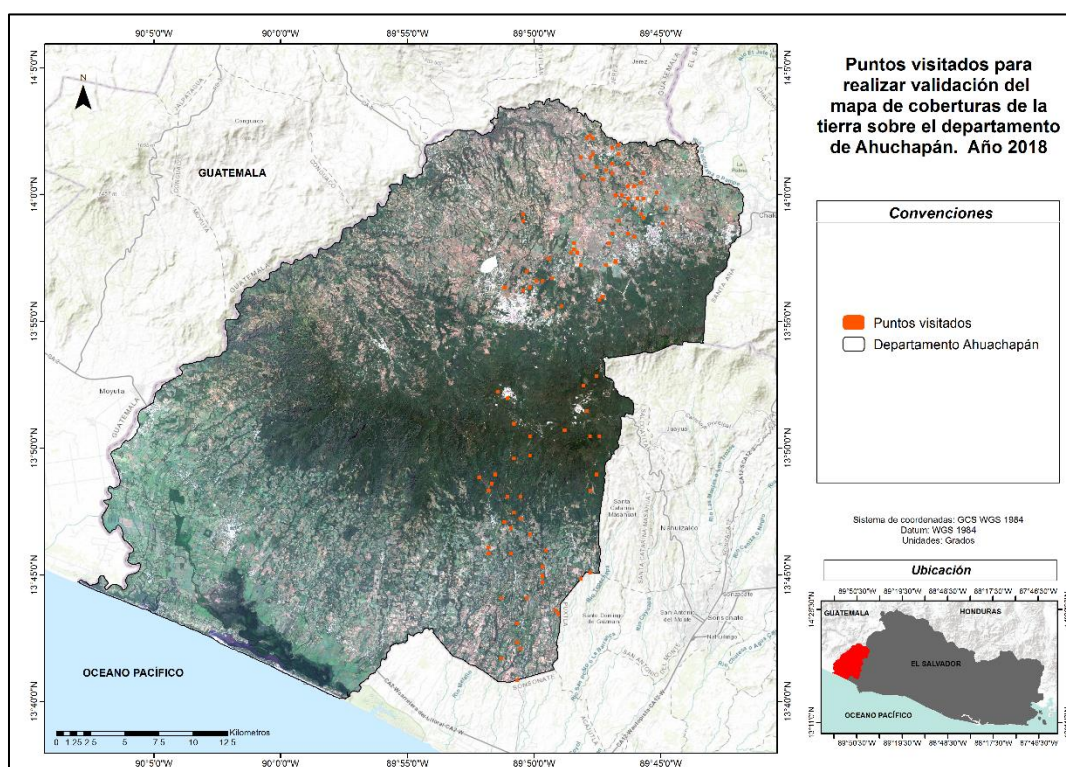


Figura 16. Mapa de puntos visitados para validación de campo.

Los puntos en campo se validaron de acuerdo a la metodología de IPCC, (2003) en donde se selecciona un pixel de gran tamaño (pixel de 250 metros) y dentro de este, se visitaron 8 muestras alrededor del punto

central (PC) (ver figura 17) De las nueve muestras validadas, al menos 1 de ellas fue visitada en el lugar exacto y a través de la herramienta Kobocollect se recolectaron los datos observados, como las coberturas que adyacentes en los cuatro puntos cardinales del sitio, fotos, fecha de la visita, entre otra información. Las otras muestras fueron observadas a través de imágenes de muy alta resolución capturadas con tres vehículos aéreos no tripulados como el phantom 3, phantom 4 y Mavic.



Figura 17. Pixel con submuestras validados en campo.

Los otros puntos no validados en campo, se presentan en la figura 18, y estos se revisaron con imágenes de alta resolución de Google Earth Pro.

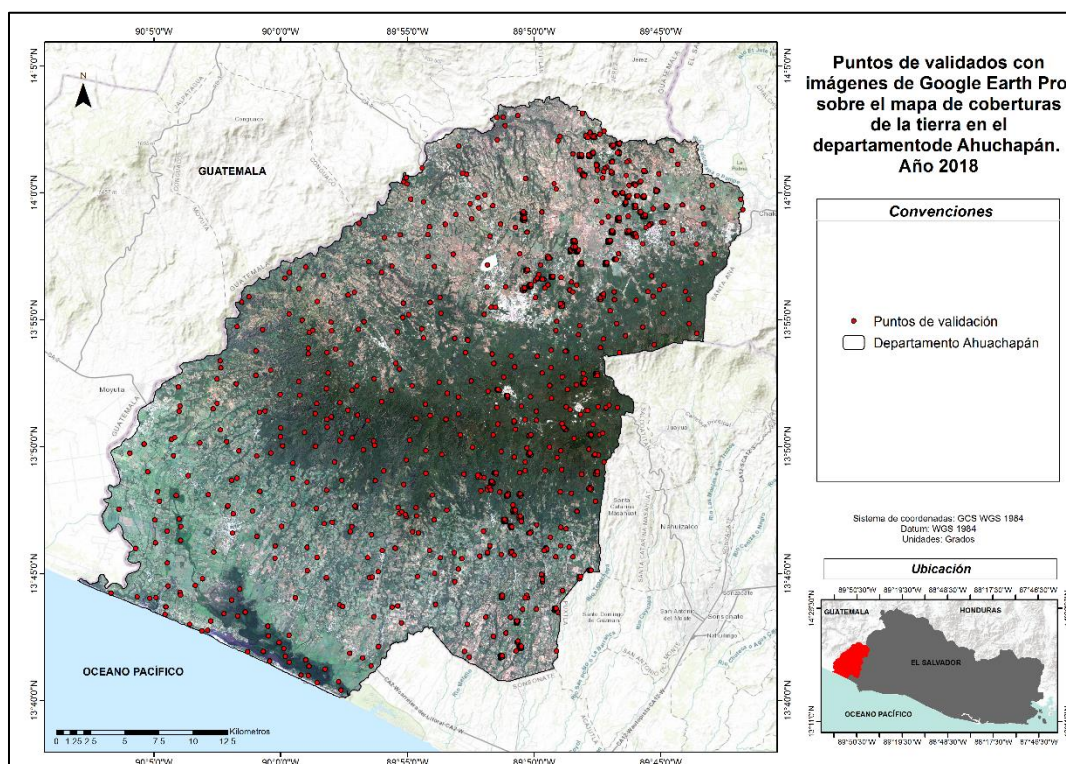


Figura 18. Mapa de puntos visitados para validación con imágenes de Google Earth Pro

El análisis realizado después de la validación, permitió elaborar una matriz de confusión con la cual se logró calcular una precisión global, precisión por coberturas y exhaustividad por coberturas usando las ecuaciones 6 y 7.

Tabla 6. Estructura de matriz de confusión

	Clasificado como A	Clasificado como B
Validado como A	TP	FN
Validado como B	FP	TN

$$\text{Precisión} = TP / (TP + FP) \quad (6)$$

$$\text{Exhaustividad} = TP / (TP + FN) \quad (7)$$

Donde TP son los verdaderos positivos, FP son los falsos positivos, FN son los falsos negativos y TN corresponden a los verdaderos negativos.

		PREDICHO															Precisión	Exhaustividad
		CA	BC	BG	BSV	CF	CAZ	FR	BM	PO	MPYC	PBM	PL	SD	U	VS		
OBSERVADO	CA	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	100%
	BC	0	110	5	1	1	0	0	0	0	3	0	0	0	0	2	90%	78%
	BG	0	4	39	3	1	0	0	0	0	1	0	0	0	0	0	81%	70%
	BSV	0	3	3	108	7	0	0	0	0	0	0	0	0	0	0	89%	90%
	CF	0	9	1	3	156	0	0	0	0	0	0	0	0	0	0	92%	94%
	CAZ	0	1	0	0	0	66	0	0	0	2	0	0	0	1	0	94%	96%
	FR	0	1	1	0	0	0	26	0	0	5	0	0	0	0	1	76%	81%
	BM	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	100%	100%
	PO	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	100%	100%
	MPYC	0	5	6	2	0	2	5	0	0	209	0	0	0	3	1	90%	90%
	PBM	0	0	1	3	0	0	0	0	0	0	4	0	0	0	1	44%	100%
	PL	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100%	100%
	SD	0	0	0	0	0	1	0	0	0	1	0	0	3	0	0	60%	60%
	U	0	1	0	0	0	0	0	0	0	3	0	0	2	19	0	76%	83%
	VS	0	7	0	0	1	0	1	0	0	9	0	0	0	0	24	57%	83%

Figura 19. Matriz de confusión del mapa de coberturas de la tierra

El resultado del mapa de cobertura, arroja una precisión global del 87.5%. En un análisis detallado, los cuerpos de agua, bosques de mangle, playa, palmeras oleíferas y caña de azúcar, son las categorías con mejor resultado, alcanzando las 4 primeras el 100% y la quinta el 94%. Por otro lado, las coberturas con mayor problema, son las plantaciones de bosque monoespecífico, la vegetación secundaria y los suelos desnudos. Estas coberturas con precisión baja, tienen poco peso en la matriz de confusión, debido a que hay pocos puntos validados siguiendo las sugerencias de los métodos de Oloffson, (2014) e IPCC (2003).

8. CONCLUSIÓN

Ahuachapán, como departamento administrativo de El Salvador, ha sido un reto grande para realizar el mapeo de sus coberturas. Su variedad de temperaturas gracias a la variedad de alturas, le concede al departamento diferentes cubiertas y usos de suelo que van desde bosques siempre verdes hasta zonas con poca vegetación o sin presencia de ella.

Diversas dificultades se presentaron en la construcción de este mapa, lo que conllevó a probar diferentes metodologías, pasando por clasificación con información óptica, división del departamento por eco-regiones, división de los datos por clústeres, y hasta clasificar con diferentes variables hasta encontrar el mejor método.

Google Earth Engine, como plataforma de procesamiento, posibilitó desarrollar las diferentes metodologías probadas sobre la región, ya que es una plataforma robusta y eficaz para procesar imágenes satelitales, y obtener resultados inmediatos.

Además, la validación del mapa de cobertura de la tierra, es el aspecto más importante en el proceso de su construcción, ya que de ella depende la publicación y la calidad del producto. La precisión es lo que refleja lo acertado que está la clasificación con respecto a lo real, y la manera en la que se calcula, puede variar este factor. Por esta razón, la metodología para el proceso de evaluación, se efectúa con métodos basados en importantes ejemplares como lo es el IPCC y Olofsson. Al aplicar la validación en el mapa generado, se obtuvo un 87% de precisión, que lo convierte en un producto publicable.

9. BIBLIOGRAFÍA

- Álvarez G., J. M. (2016). *Invertir en el recurso hídrico*. Retrieved 2019, from <http://fusades.org/lo-ultimo/blog/invertir-en-el-recurso-h%C3%ADrico>
- Amat, J. R. (2017, April). *Máquinas de Vector Soporte (Support Vector Machines, SVMs)*. Retrieved from https://rpubs.com/Joaquin_AR/267926
- Bissanti, G. (2017). *Bosque caducifolio*. Retrieved 2019, from <https://antropocene.it/es/2017/05/16/foresta-decidua/>
- Cogmanglar. (2018). *Manglar: Última frontera de los ecosistemas marinos costeros*. Retrieved 2019, from <http://www.entremundos.org/revista/economia/manglar-ultima-frontera-de-los-ecosistemas-marino-costeros/>
- Diario el Mundo. (2017). *Gran San Salvador alojará 28 mil nuevas familias en 2030*. Retrieved from <https://elmundo.sv/gran-san-salvador-alojara-28-mil-nuevas-familias-en-2030/>
- Diario Electrónico Cuarto Poder. (2018). *Apuestan a la siembra de caña*. Retrieved 2019, from <http://www.cuartopoderdetamulipas.com.mx/apuestan-a-la-siembra-de-cana/>
- ECOticias.com-PNUMA. (2010). *El "atlas mundial de los manglares" pone de relieve la importancia de y las amenazas de los manglares*. Retrieved 2019, from <https://www.ecoticias.com/naturaleza/29398/noticias-medio-ambiente-medioambiente-medioambiental-ambiental-definicion-contaminacion-cambio-climatico-calentamiento-global-ecologia-ecosistema-impacto-politica-gestion-legislacion-educacion-responsabilidad-tec>
- elsalvador.com. (2018). *Playas de San Miguel listas para el turista*. Retrieved 2019, from <https://www.elsalvador.com/noticias/nacional/505611/playas-de-san-miguel-listas-para-el-turista/>
- Espino Ortega, V. (2017). *Plantaciones forestales comerciales, mitigantes del cambio climático*. Retrieved from http://www.primeaction.com/sp/noticia_sectorial/plantaciones_forestales_comerciales_mitigantes_del_cambio_climatico_
- ESRI. (2016). *ArcMap*. Retrieved from Alisar bordes de zona con Refinado de límites y Filtro mayoritario: <http://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-analyst-toolbox/smoothing-zone-edges-with-boundary-clean-and-majority-filter.htm#GUID-CBD68932-69B7-469B-9DCB-8A743B72B908>
- ESRI. (2016). *ArcMap*. Retrieved from Crear zonas individuales con Grupo de regiones: <http://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-analyst-toolbox/creating-individual-zones-with-region-group.htm>
- ESRI. (2016). *ArcMap*. Retrieved from Establecer nullos: <http://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-analyst-toolbox/set-null.htm>
- ESRI. (2016). *ArcMap*. Retrieved from Generalizar zonas con Nibble, Encoger y Expandir: <http://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-analyst-toolbox/generalizing-zones-with-nibble-shrink-and-expand.htm#GUID-D85AFF2B-CB03-4E1C-89B1-0E4CEDA4D2D1>
- FAO. (2010). *Evaluación de los recursos forestales mundiales 2010*. Roma.
- Federación Nacional de cafeteros de Colombia. (2010). *Nuestras regiones cafeteras*. Retrieved 2019, from http://www.cafedecolombia.com/particulares/es/la_tierra_del_cafe/regiones_cafeteras/

- García Abadía, J. (2014). *¿Limones protegidos?* Retrieved 2019, from <https://www.laverdad.es/murcia/comarcas/201410/18/limones-protectidos-20141018005459-v.html>
- Google Developers. (2019). *Machine Learning*. Retrieved 09 18, 2019, from <https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages>
- Google Earth Engine. (2017). *Sentinel-2 MSI: MultiSpectral Instrument, Level-1C*. Retrieved 2018, from <https://explorer.earthengine.google.com/#detail/COPERNICUS%2FS2>
- IDEAM. (2014). *Ecosistemas*. Retrieved from Coberturas de la tierra: <http://www.ideam.gov.co/web/ecosistemas/coberturas-tierra>
- Instituto Sichi. (2009). Mosaico de pastos y cultivos. Fichas de los patrones de las coberturas de la tierra de la Amazonia Colombiana. Bogotá D.C.
- Instituto Sinchi. (2009). Bosque de galería y ripario. Ficha de los patrones de las coberturas de la tierra de la Amazonia Colombiana. Bogotá D.C. Retrieved 2019, from http://siatac.co/web/guest/productos/coberturasdelatierra/fichasdepatrones?p_p_id=54_INSTANCE_K1kl&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_54_INSTANCE_K1kl_struts_action=%2Fwiki_display%2Fview&_54_INSTANCE_K1kl_
- Instituto Sinchi. (2009). Vegetación secundaria o en transición. Fichas de los patrones de las coberturas de la tierra de la Amazonia Colombiana. Bogotá D.C. Retrieved 2019, from http://siatac.co/web/guest/productos/coberturasdelatierra/fichasdepatrones?p_p_id=54_INSTANCE_K1kl&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_54_INSTANCE_K1kl_struts_action=%2Fwiki_display%2Fview&_54_INSTANCE_K1kl_
- IPCC. (2003). *Institute for Global Environmental Strategies Good Practice Guidance for Land Use, Land Use Change and Forestry*. Retrieved from <http://www.ipcc-nggip.iges.or.jp>.
- Kuhn, M. (2018). *Pre-Processing*. Retrieved from The caret Package: <http://topepo.github.io/caret/pre-processing.html>
- La estrella de Panamá. (2017). *un 33% de los suelos del mundo sufren erosión*. Retrieved 2019, from <http://laestrella.com.pa/vida-de-hoy/planeta/33-suelos-mundo-sufren-erosion/24030623>
- Marina-Ramona, R.-M., Bogdan, o., Bogdan-Andrei, M., Constantin, N., & Lonut, S. (2017). Land cover classification in Romanian Carpathians and Subcarpathians using multi-date Sentinel-2 remote sensing imagery. *European Journal of Remote*, 496-508.
- Mejías César, Y., Carrasco Velar, R., Ochoa Izquierdo, I., & Moreno Lemus, E. (2013). Funciones de transferencia en el perceptrón multicapa: efecto de su combinación en entrenamiento local y distribuido. *Revista Cubana de Informática Médica*, 5(2), 186-199. Retrieved 09 24, 2019, from <https://pdfs.semanticscholar.org/df3b/297c0ef8a495293af855ab0c97194eda09b5.pdf>
- 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 Sensing of Environment*, 42-47. doi:<https://doi.org/10.1016/j.rse.2014.02.015>
- Pereira-Toledo, A., Lopez-Cabrera-, J. D., & Quinter-Domínguez, L. A. (2017). Estudio experimental para la comparación del desempeño de Naïve Bayes con otros clasificadores bayesianos. *Revista Cubana de Ciencias Informáticas*, 67-84.

- Procaña. (2019). *Historia de la caña de azucar*. Retrieved 2019, from <https://www.procana.org/new/quienes-somos/historia-de-la-cana-de-azucar.html>
- REDD-CCAD-GIZ. (2011). *Tipos de bosques y contexto del mapeo de la cobertura forestal en centroamérica y República Dominicana*. Antigua Guatemala.
- Sánchez, M. (2018). *Palma Aceitera (Elaeis Guineensis)*. Retrieved 2019, from <https://www.jardineriaon.com/palma-aceitera-o-elaeis-guineensis.html>
- Sharma, A., Liu, X., & Xiaojun, Y. (2018). Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. *Neural Networks*, 105, 346-355.
- Telliet, P. M., Guindon, B., & Goodenough, D. G. (1982). On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, 84-106.
- Uhorchak, N. (2017). *Random Forest Tutorial*. Retrieved 2018, from <https://rpubs.com/nuhorchak/randomForest>
- Urquillo, A. (2018). *Entre los cafetales de Apaneca: Cabaña la Esperanza*. Retrieved 2019, from <https://www.laprensagrafica.com/elsalvador/Entre-los-cafetales-de-Apaneca-Cabana-La-Esperanza-20180118-0086.html>
- USGS. (2018, April 25). *Landsat Missions*. Retrieved from Landsat Project Description: <https://landsat.usgs.gov/landsat-project-description>
- WWF. (2019). *Cuatro tipos de bosques: conoce las diferencias*. Retrieved 2019, from <https://www.worldwildlife.org/descubre-wwf/historias/cuatro-tipos-de-bosques-conoce-las-diferencias>
- Zhao, J., Yu, L., Xu, Y., Ren, H., Huang, X., & Gong, P. (2019). Exploring the addition of Landsat 8 thermal band in land-cover mapping. *International Journal of Remote Sensing*, 4544-4559.
- Zheng, B., Myint, S. W., Thenkabail, P. S., & Aggarwal, R. M. (2015). A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, 34, 103-112.