

Anexo A. Clasificación de coberturas de la tierra

La geografía estudia procesos que ocurren a distintas escalas espaciales. Para cada una de estas escalas, es necesario recolectar datos de forma rápida, eficiente y precisa, para no alterar los resultados del fenómeno que se desea estudiar. La teledetección es un método de recolección de datos que responde a todos estos requerimientos, permitiendo conocer y estudiar los ambientes, físicos, químicos, biológicos y culturales, de la superficie terrestre¹⁵⁶. La teledetección se vincula a esta investigación dado que algunas coberturas de la tierra son los elementos de la EEM. Por tal motivo, en esta porción del documento se presentarán los fundamentos metodológicos con los que se clasificaron las coberturas de la tierra del área de estudio.

La principal justificación para proponer y realizar esta clasificación de coberturas de la tierra es que la información temática oficial no era la más adecuada para esta investigación. Los datos de coberturas existentes o estaban a una escala espacial muy baja —1:100.000 de IDEAM¹⁵⁷— o la brecha temporal era tan amplia que las coberturas difícilmente representaban con exactitud las condiciones actuales de la superficie terrestre en el área de estudio —la capa de coberturas de la CVC¹⁵⁸ tiene información que cubre el rango de 1996 a 2010—. Otro factor que dificultó el uso de datos espaciales era que el área de estudio se localizó en la jurisdicción de las instituciones CVC y CRC y, al momento de unificar la información oficial proporcionada por ellas, se descubrió una serie de incongruencias entre límites y contenidos. Debido a todas estas dificultades, se optó por realizar la clasificación de imágenes satelitales *RapidEye*. En esta sección de la metodología se presentan los fundamentos de la clasificación, así como cada una de las etapas distintivas del proceso.

Por otra parte, con la realización de dos salidas de campo se afianzó y actualizó la interpretación geográfica del área de estudio por parte de los investigadores. En la Figura 31 presentan particularidades de la realidad ambiental identificada en estas salidas. Las imágenes de los numerales a, b y c hacen referencia a la visita en campo dirigida al embalse La Salvajina (Suárez, Cauca) y la vía regional Cali-Popayán. No son tomadas por los investigadores, pero corresponden fielmente a lo observado sobre: (a) río Cauca y la geomorfología del municipio de Suárez; (b) configuración mixta de coberturas de la tierra en Santander de Quilichao; (c) contexto ambiental del embalse La Salvajina.

Las imágenes de los numerales d, e y f de la Figura 31 refieren al recorrido en lancha del río Cauca por los municipios de Jamundí, sur de Cali, Candelaria, Palmira y perímetro norte de Cali, realizado el 4 de noviembre de 2016. En este caso, el registro fotográfico fue tomado directamente por los investigadores. El numeral (d) muestra la ribera occidental del río Cauca dejando en evidencia la presencia de infraestructura de vivienda de la ciudad de Cali al margen inmediato del cauce. La imagen (e) presenta como en el flanco oriental (Palmira) el cultivo

¹⁵⁶ CHUVIECO. Op. cit.

¹⁵⁷ INSTITUTO DE HIDROLOGÍA, METEOROLOGÍA Y ESTUDIOS AMBIENTALES. Op. cit.

¹⁵⁸ CORPORACIÓN AUTÓNOMA REGIONAL DEL VALLE DEL CAUCA. Op. cit.

de la caña de azúcar se extiende hasta rozar el margen inmediato del río Cauca. Por último, con la foto (f) se revela que ambas riberas del río Cauca aún cuentan con su vegetación prístina en algunos tramos del recorrido entre Jamundí, Cali, Candelaria y Palmira.

Figura 31. Contextualización en campo de área de estudio. Sector sur del área (a, b y c), río Cauca y sus riberas (d, e y f).



Foto: GOOGLE EARTH (a, b y c) y MAURICIO TABARES MOSQUERA (d, e y f).

La ejecución de salidas de campo con destinos estratégicos fue significativamente provechosa para los investigadores teniendo en cuenta que actualizó y profundizó su conocimiento geográfico del área, permitiendo identificar patrones de coberturas de la tierra y conflictos ambientales

particulares. La riqueza metodológica del trabajo en campo es que aportó de forma sustancial a cada una de las fases de esta investigación.

También, es importante resaltar que la información espacial producida en esta fase se almacenó y procesó en una *Geodatabase* para garantizar la integridad, el manejo eficiente, y el procesamiento exitoso de los datos. Igualmente, cada uno de estos productos contó con el «Sistema de Coordenadas planas UTM Datum WGS84 Zona 18» para evitar cualquier error o irregularidad que pudiese producirse si se trabajaran con sistemas de coordenadas diferentes. Para facilitar el entendimiento del procedimiento y los productos obtenidos en esta fase metodológica, se recomienda revisar algunos términos claves referenciados durante esta sección (Cuadro 20).

Cuadro 20. Terminología de teledetección empleada en ese documento.

Banda espectral: categorías discretas de ondas electromagnéticas que comparten características similares	Radiación electromagnética: forma de energía en tránsito donde los campos eléctricos y magnéticos varían simultáneamente. Todos los objetos de la tierra pueden reflejar, absorber o emitir energía
Espectro electromagnético: el nombre que reciben los diferentes tipos de luz y energía en el universo. Incluye ondas radiales, microondas, radiación infrarroja, luz visible, luz ultravioleta, rayos X, y rayos gamma	Radiancia: es la cantidad de luz que «ve» el sensor del objeto observado en vatios / (estereorradián / m ²). Variable medida directamente por instrumentos de teledetección
Firma espectral: las características de reflectancia de un objeto o superficie en todo el espectro electromagnético	Reflectancia: es una propiedad del material que se observa. Es la energía que refleja la superficie receptora. En el proceso de corrección de una imagen, es conveniente convertir los valores de radiancia a reflectancia
Imagen análoga: son imágenes de tono continuo, e.g. negativos de fotografía, diapositivas e impresiones	Resolución espacial: determinada por el tamaño de píxel. Medida en metros cuadrados sobre el terreno
Imagen digital: representación pictórica compatible con computadora dividida en una cuadrícula de píxeles bidimensional	Resolución espectral: consiste en el número de bandas espectrales (y su ancho de banda) que puede captar un sensor
Número digital (ND): cuantificación del reflejo promedio que registró el sensor para el área de superficie de cada píxel	Resolución radiométrica: número de valores que puede ser almacenado por píxel. Depende de la escala binaria del sensor (1bit, 2bits, 4bits, 8bits, 16bits o 32bits)
Píxel: unidad finita y mínima de área en una imagen. Término que proviene de la contracción de las palabras « <i>picture elements</i> » (elementos de imagen)	Resolución temporal: frecuencia de paso del satélite por un mismo punto de la superficie terrestre

Fuente: elaborado a partir de CHUVIECO. Op. cit. & JAMES. Op. cit.

El proceso de clasificación de coberturas de la tierra se ejecutó en un computador de las siguientes especificaciones técnicas. En hardware: procesador Intel Core i5 de tercera generación, 4 gigabytes de memoria RAM, 1 Terabyte de disco duro, pantalla LED de 19 pulgadas. En software: sistema operativo Windows 7, licencia institucional de la Universidad del Valle de ArcGIS 10, y versión de prueba de ENVI 5.1. Al año 2017, la configuración de este equipo corresponde a la gama tecnológica baja o casi media, sin embargo, fue posible

ejecutar con esta máquina todos los procesos indicados en esta sección de forma exitosa.

I. Fundamentos de la metodología de clasificación de coberturas de la tierra

La teledetección es cualquier técnica con la que se obtiene información a la distancia de un objeto y el ambiente¹⁵⁹. La base de este proceso de recolección de datos es que «la vegetación, la forma del terreno y la cobertura del suelo pueden tener firmas espectrales distintivas, lo que significa que emiten y reflejan la radiación electromagnética de diferentes maneras»¹⁶⁰. En este orden de ideas, el objetivo formal de la teledetección es «detectar, medir y analizar la radiación electromagnética de la superficie terrestre»¹⁶¹. Para afrontar dicha tarea, la teledetección es dependiente de las características de su insumo clave de procesamiento: las imágenes.

Aunque los datos de teledetección no se encuentran exclusivamente en forma de imágenes, este es el formato más común y útil para los estudios geográficos y ambientales. La utilidad de estas imágenes en teledetección depende de las características del contenido como «la forma, el tamaño, los patrones espaciales, el tono y la textura»¹⁶². En conjunto, estas características determinan la calidad y el tipo de información que puede obtenerse de la imagen. Las primeras tres características se explican por sí mismas teniendo en cuenta que describen las particularidades espaciales de las entidades presentes en una imagen. No obstante, el tono es el color que describe el brillo de la superficie detectado por el sensor. En un ejercicio de interpretación de imágenes, muchos objetos pueden clasificarse de forma exitosa a partir del tono: lo verde será vegetación, mientras un tono café podría ser suelo desnudo o cultivos, y lo gris se relacionará con áreas construidas.

La característica textura merece una mención especial. La textura es «la variación tonal en una imagen en función de la escala»¹⁶³. Es decir que, mientras el tono o color describe la información espectral de los píxeles de una imagen, la textura describe la variación espacial de dicha información espectral bajo un efecto de aglomeración de píxeles. Así, es posible identificar texturas ásperas, rugosas y suaves comparando, por ejemplo, un área de pastos limpios con una de bosque natural. Ambas coberturas son vegetación, pero su textura rugosa varía sustancialmente. En conclusión, el rol de la textura es igual o más importante que el del tono para la interpretación de imágenes.

¹⁵⁹ CHUVIECO. Op. cit. & HARRIS, Richard. Remote sensing. En: GREGORY, Derek, *et al.* (eds.). The dictionary of human geography. 5 ed. Oxford: Blackwell, 2009. p. 643-644. & JAMES. Op. cit.

¹⁶⁰ HARRIS. Op. cit., p. 643.

¹⁶¹ JAMES, Timothy. Remote sensing of environmental change. En: HOLDEN, Joseph. An Introduction to Physical Geography and the Environment. 3 ed. Harlow: Pearson Education Limited, 2012. p. 744.

¹⁶² *Ibíd.*, p. 745

¹⁶³ *Ibíd.*, p. 745.

Las imágenes de sensores remotos proveen información de gran calidad sobre la superficie de la tierra que puede usarse para desarrollar análisis globales y locales de monitoreo, así como detección de cambios espaciales¹⁶⁴. A partir de su interpretación humana y asistida por computadora, estas imágenes deben convertirse de información de píxeles y valores digitales a una clasificación categórica de coberturas de la tierra. Esta clasificación se realiza de dos formas: automática o supervisada. Los métodos de clasificación pueden ser supervisados o no supervisados, diferenciándose por el grado de intervención del usuario. En la clasificación no supervisada no se establece ninguna clase con antelación, pero es necesario determinar el número máximo de clases a clasificar. El procedimiento consta de una serie de pasos, en cada paso se identifican los dos individuos más próximos, se hace una clase con ellos y se sustituyen por el centroide de la clase resultante¹⁶⁵. El proceso se detiene cuando se ha alcanzado un número de clases igual al establecido por el usuario.

La clasificación supervisada toma como base la disponibilidad de áreas de entrenamiento. Se trata de áreas de las que se conoce con seguridad la clase a la que pertenecen y que servirán para generar una firma espectral característica de cada una de las clases¹⁶⁶. Obtener las características de reflectividad de estas clases implica una consulta a las imágenes. El método supervisado de clasificación tiende a arrojar resultados más precisos que el método no supervisado, ya que hay muchas pruebas estadísticas que se pueden aplicar a las muestras de entrenamiento para determinar qué tan representativas son las clases y qué tan espectralmente diferenciables son de las demás¹⁶⁷. Este antecedente perfiló la elección del método supervisado para clasificar las coberturas de la tierra del área de estudio. Sin embargo, otro antecedente orientó el ajuste de este modelo de clasificación: el problema técnico de los píxeles.

En los inicios, la técnica más usada para clasificar las coberturas de la tierra era basada en píxeles. El fundamento principal de esta técnica era analizar las propiedades espectrales de cada pixel de un área de interés. Con el desarrollo tecnológico de la obtención de datos por sensores remotos, la técnica de clasificación por píxeles presentó dos grandes conflictos: i) la técnica no toma en cuenta información espacial y espectral de los vecinos; ii) los datos están en función de la resolución del sensor que capturó la imagen¹⁶⁸. La Figura 32 representa esta situación conocida como el "problema de los píxeles". Cuando los píxeles son relativamente más grandes o pequeños que los objetos de la superficie terrestre, la clasificación tradicional orientada por píxeles no es la mejor elección para mapear coberturas de la tierra. En esta situación, cuando los métodos basados en píxeles se aplican a las imágenes de satélite de alta resolución se produce un efecto de "sal y pimienta" que hace más inexacta la

¹⁶⁴ BENZ, Ursula, *et al.* Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *En*: ISPRS Journal of photogrammetry and remote sensing [en línea]. 2004. Vol. 58, No. 3; p. 239-258.

¹⁶⁵ CHUVIECO. Op. cit.

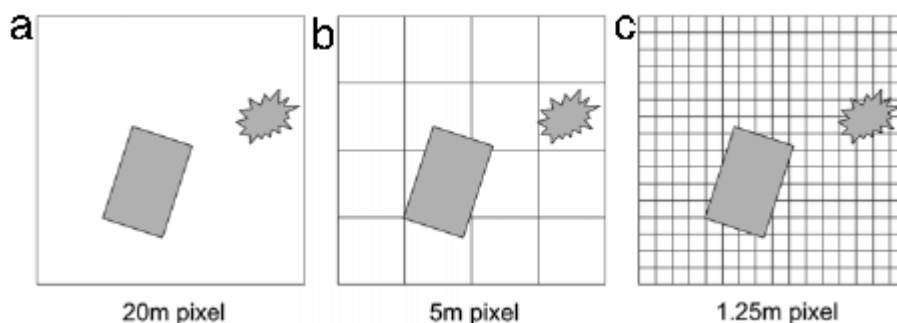
¹⁶⁶ *Ibíd.*

¹⁶⁷ JAMES. Op. cit.

¹⁶⁸ MAKINDE. Op. cit.

clasificación¹⁶⁹. Para solucionar esta problemática se desarrollaron otros métodos de procesamiento de información remota. Entre ellos está la clasificación orientada a objetos geográficos.

Figura 32. Relación entre objetos de la superficie terrestre y la resolución espacial de imágenes.



a) baja resolución, píxeles más grandes que objetos. b) mediana resolución, píxeles más pequeños que objetos. c) alta resolución, píxeles mucho más pequeños que objetos
Fuente: BLASCHKE, Thomas. Object based image analysis for remote sensing. *En: ISPRS journal of photogrammetry and remote sensing*. 2010. Vol. 65, No. 1; p. 3.

Una alternativa al método de clasificación de coberturas basado en píxeles es el método de análisis de imágenes a partir de objetos geográficos (GEOBIA). El argumento de este nuevo enfoque es «fusionar los píxeles adyacentes en objetos homogéneos en términos espectrales y espaciales, para conducir el proceso de clasificación tomando estos objetos como la unidad mínima de análisis»¹⁷⁰. Cabe destacar que el objeto geográfico es referido al tipo de cobertura de la tierra que se desea clasificar. Esta característica del método GEOBIA garantiza la producción de información espacial que corresponde idealmente a las características del mundo real porque el proceso de asociar píxeles con propiedades espectrales y espaciales similares permite captar la diversidad de la superficie terrestre. En general, el método GEOBIA produce mejores resultados (entre 84% y 89% de precisión) que el dirigido por píxeles¹⁷¹. Por estas razones, el método de clasificación GEOBIA se considera la opción más apropiada por la mayoría de los investigadores en la actualidad y es la técnica empleada en esta investigación.

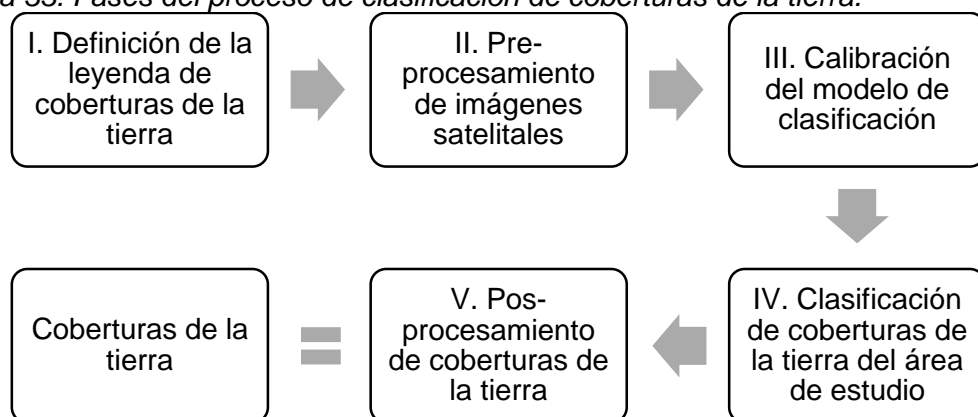
A grandes rasgos, el proceso de clasificación de coberturas de la tierra ejecutado en este trabajo fue semiautomático dado que constó de tres etapas: 1) encontrar los objetos, 2) muestrear los objetos y 3) clasificar los objetos. Sin embargo, fue necesario ejecutar procesos auxiliares en fases específicas para que la clasificación de coberturas fuera exitosa. En consecuencia, la clasificación de coberturas de la tierra se sintetiza en cinco grandes fases. La Figura 33 muestra sintéticamente las fases de clasificación y la conexión de sus productos para alcanzar el objetivo deseado.

¹⁶⁹ *Ibíd.*

¹⁷⁰ SOUSA, *et al.* Op. cit., p. 518.

¹⁷¹ BLASCHKE. Op. cit. & SOUSA. Op. cit. & MAKINDE. Op. cit.

Figura 33. Fases del proceso de clasificación de coberturas de la tierra.



Fuente: elaboración propia.

II. Definición de la leyenda de coberturas de la tierra

El primer paso para clasificar coberturas de la tierra es definir, a conciencia y en función al objetivo de investigación propuesto, una leyenda que incluya los tipos de coberturas deseados y los niveles de especificidad requeridos. Este proceso es clave porque su resultado determinará el funcionamiento, extensión y precisión de las fases metodológicas del proceso de clasificación de coberturas. En esta investigación, se adaptó la leyenda CORINE *Land Cover* propuesta para Colombia con base en las particularidades del área de estudio y con relación al objetivo de identificar la EEM. Por este motivo, es necesario presentar las características principales de esta leyenda de coberturas.

En 1985, la Unión Europea inició el proyecto CORINE con la intención de recopilar información sobre el estado del medio ambiente en Europa¹⁷². El producto principal de esta iniciativa se publicó en 1994: la leyenda de coberturas de la tierra CORINE *Land Cover*. Esta leyenda incluyó 44 clases de coberturas de la tierra clasificadas en una nomenclatura jerárquica de seis niveles que representan todos los tipos de coberturas existentes en el continente europeo.

Entre las potencialidades de la leyenda de coberturas de la tierra CORINE *Land Cover* se destaca que su estructura jerárquica reproduce un orden escalar explícito, conservando las características de las coberturas de la tierra entre sus niveles. Este orden lógico y espacial garantiza la adición de nuevos niveles y clases de coberturas para estudios más detallados¹⁷³. Otro elemento adicional es que la leyenda CORINE tuvo gran acogida en la comunidad científica, lo que significó el establecimiento de un lenguaje académico común de coberturas de la tierra. Sin embargo, su mayor limitante emerge al exportar esta leyenda a contextos de estudio por fuera de Europa porque algunas coberturas no serán representativas y deben ser modificadas. En ese sentido, para solventar ese problema, la leyenda CORINE *Land Cover* aplicada para Colombia por parte del

¹⁷² BURKHARD, *et al.* Landscapes' capacities to provide ecosystem services - A concept for land-cover based assessments. Op. cit.

¹⁷³ INSTITUTO DE HIDROLOGÍA, METEOROLOGÍA Y ESTUDIOS AMBIENTALES. Op. cit.

IDEAM propone cinco categorías principales¹⁷⁴. Estas categorías coinciden a las propuestas por la Unión Europea, pero se realizaron modificaciones entre niveles para que la leyenda respondiera al contexto de la superficie terrestre de Colombia.

Ahora bien, a partir de la leyenda CORINE *Land Cover* aplicada para Colombia, en esta fase metodológica de la investigación, se clasificaron 28 coberturas de la tierra del área de estudio a escala 1:25.000. Para la adaptación de esta leyenda de coberturas a esta investigación se tuvieron en cuenta las siguientes consideraciones:

- La UMC, o la superficie menor que puede ser representada sobre un mapa, en este caso de escala 1:25.000, es de 10.000 mt², cifra equivalente a una (1) hectárea¹⁷⁵.
- Los niveles 1 y 2 de la leyenda permanecen iguales a la metodología original de CORINE *Land Cover* de Europa.
- El nivel 3 de la leyenda fue seleccionado como el óptimo porque reproduce una escala espacial regional y se ajusta al objetivo de proponer una EEM.
- El cultivo de la caña de azúcar, cobertura de la tierra clasificada en el nivel 4 de nuestra leyenda, fue la única clasificada de este nivel por las implicaciones multidimensionales que tiene este cultivo en la región.

Con base a estas observaciones, el Anexo B que contiene la descripción completa de las coberturas de la tierra clasificadas en este ejercicio. Allí se presentan los detalles de cada cobertura a profundidad, incluyendo su descripción, ilustración y simbología empleada en la cartografía temática.

III. Pre-procesamiento de imágenes satelitales

Este proceso de clasificación de coberturas tuvo como insumo principal las 32 imágenes satelitales de la constelación *RapidEye*. PROCÁLCULO, compañía proveedora de soluciones espaciales en Colombia, proporcionó estas imágenes en 32 cuadrantes de 25x25 km (625 km²), capturados en los años: 2013 (5 cuadrantes), 2014 (4), 2015 (21) y 2016 (2), respectivamente. Las características completas de estas imágenes se presentan en el Cuadro 21. Es importante resaltar que el proveedor entregó las imágenes *RapidEye* con un nivel de preprocesamiento 3A *RapidEye Product Ortho*. Este proceso garantiza que las imágenes cuenten con correcciones radiométricas y geométricas, y referenciadas a la proyección cartográfica «Sistema de Coordenadas UTM Datum WGS84 Zona 18». Pese a esta afirmación y con base en la literatura¹⁷⁶, se descubrió anomalías en los valores digitales de algunos cuadrantes por lo que fue necesario pre-procesar este conjunto de imágenes.

¹⁷⁴ Ibíd.

¹⁷⁵ SALITCHEV. Op. cit.

¹⁷⁶ DUARTE, *et al.* Op. cit.

Cuadro 21. Características espectrales y espaciales de las imágenes RapidEye.

Parámetros	Valores		
Bandas espectrales, longitud de onda (en μm) y su aplicación en teledetección	(1) Azul	0,440 - 0,510	Buena penetración en el agua. Útil para la discriminación de suelo, vegetación, y elementos antrópicos. Sensible a bruma atmosférica
	(2) Verde	0,520 - 0,590	Alguna penetración en el agua, pero sensible a la turbiedad. Reflectancia alta con la vegetación útil para determinar su vigor. Sensible a bruma atmosférica
	(3) Rojo	0,630 - 0,635	Absorción alta de la clorofila vegetal y alta reflectancia en suelos. Muy buena para diferenciar suelo de vegetación y para reconocer elementos antrópicos.
	(4) Borde del rojo	0,690 - 0,730	Útil para diferenciar tipos de vegetación y sus diferentes niveles de vitalidad. Resalta ampliamente la estructura de la hoja y la clorofila vegetal. Banda exclusiva de las imágenes <i>RapidEye</i>
	(5) Infrarrojo cercano	0,760 - 0,850	Alta reflectancia de la vegetación y absorción alta del agua. Mejor banda para discriminar los tipos de vegetación y el vigor de la planta. Excelente para reconocer superficies de agua y la humedad del suelo
Distancia del muestreo sobre el terreno (nadir)	6,5 mt		
Tamaño del pixel (orto rectificado)	5 mt		
Ancho de observación	77 km		
Tiempo de revisita	Diario		
Hora de cruce por el ecuador	11:00 a.m. (aproximadamente)		
Capacidad de adquisición de imágenes	4 millones de km^2 diarios		

Fuente: JAMES. Op. cit. & VALERO, Oscar. Aplicación de la geomática para determinar la cobertura del uso actual del suelo a partir de imágenes de satélite (*RapidEye*) y su comparación con el *shapefile* uso de suelo. Caso de estudio: municipio de Tibana (Boyacá). Bogotá: Universidad Militar Nueva Granada, 2015. 7 p. & PLANET LABS. Applications of the RapidEye Red Edge Band. San Francisco: _____. 2016, 8 p.

Regularmente, es necesario procesar las imágenes de sensores remotos después de su recepción, especialmente si fueron tomadas en fechas y por sensores diferentes¹⁷⁷. Este procesamiento puede implicar una rectificación que nos permite «compensar cualquier distorsión o radiométrica o geométrica de la imagen que pueda haber resultado durante su adquisición a través del

¹⁷⁷ ŞATIR y BERBEROĞLU. Op. cit.

sensor»¹⁷⁸. Las imágenes *RapidEye* 3A empleadas en este ejercicio contaron con una geometría correcta: las coordenadas correspondían a su punto de localización geográfica siguiendo el «Sistema UTM Datum WGS84 Zona 18». No obstante, las imágenes presentaron irregularidades en los valores de reflectancia puesto que su fecha de captura se distribuyó en 4 años diferentes, y fue necesario realizar una **corrección radiométrica y atmosférica** a cada una de ellas.

Mientras las correcciones geométricas se relacionan a la superficie terrestre capturada por la imagen, las radiométricas abordan los valores digitales de cada píxel tomados por el sensor. Son muchos los factores que alteran la reflectancia de una imagen, pero se reconoce que las condiciones atmosféricas son las más inciden¹⁷⁹. Para corregir una imagen en términos radiométricos, los valores de reflectancia son ajustados con base en un modelo que ajusta el brillo del píxel y así representar las características reales de luminosidad de la escena. Las herramientas *Band Math* y *Fast Line-of-sight Atmospheric Analysis of Hypercubes* (FLAASH) de ENVI, se emplearon para realizar esta corrección.

Band Math es un tipo de calculadora que permite realizar distintas operaciones matemáticas entre todas las bandas espectrales utilizando el archivo de metadatos de la imagen satelital. Con el archivo de metadatos se convierten los valores digitales de la imagen a datos de radiancia. Como demuestra Duarte¹⁸⁰, esto se hace ingresando en el campo **Enter an expression de Band Algebra > Band Math** la siguiente fórmula:

$$\text{Radiancia (L)} = ND * 10$$

El factor de escala radiométrica de **10** se ingresó para convertir el resultado de cada banda espectral de unidades de watts por estereorradián por metro cuadrado ($W/(m^2 \cdot sr \cdot \mu m)$) a centímetros cuadrados ($W/(cm^2 \cdot sr \cdot \mu m)$)¹⁸¹. Con la corrección radiométrica, se ajusta el formato de unidades que recibe la herramienta encargada de ejecutar la corrección atmosférica: FLAASH de ENVI.

FLAASH es una herramienta de corrección atmosférica que corrige las longitudes de onda en las regiones visibles del espectro, permitiendo calcular la reflectancia de la superficie y ajustar la incidencia del vapor del agua y los aerosoles en la visualización de la imagen satelital¹⁸². Esta herramienta se encuentra en el menú **Radiometric Correction > Atmospheric Correction Module > FLAASH Atmospheric Correction**, de ENVI. Este módulo identifica la configuración de la escena y los parámetros del sensor *RapidEye* de forma automática, ingresando correctamente latitud y longitud de la escena, altitud del

¹⁷⁸ JAMES. Op. cit., p. 777.

¹⁷⁹ Ibíd.

¹⁸⁰ DUARTE, *et al.* Op. cit.

¹⁸¹ Ibíd.

¹⁸² HARRIS GEOSPATIAL SOLUTIONS. Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) [en línea]. Harris Geospatial Solutions, 2017. [Consultado el 14 de enero de 2017]. En: <http://www.harrisgeospatial.com/docs/flaash.html>

sensor (km), elevación del terreno (km), tamaño de píxel (mt) y la fecha y hora del vuelo (hh:mm:ss).

Adicionalmente, FLAASH realiza la corrección atmosférica empleando el modelo *Moderate resolution atmospheric Transmission* (MODTRAN) de la *US Air Force*¹⁸³. Básicamente, MODTRAN es una base de datos que contiene estimaciones promedio de las condiciones atmosféricas para todo el mundo, tomando en cuenta indicadores de vapor del agua y la temperatura del aire¹⁸⁴. Considerando la localización del área de estudio de esta investigación, la opción seleccionada del modelo fue **FLAASH > Atmospheric Model > Tropical**.

El pre-procesamiento de los 32 cuadrantes *RapidEye* se realizó **uno a la vez**, aplicando en cascada las herramientas **Band Math y FLAASH**. La evidencia del pre-procesamiento de imágenes es la identificación adecuada de la firma espectral para distintos tipos de cobertura. En la Figura 34 se observa la incidencia del ajuste radiométrico realizado sobre las firmas espectrales de las coberturas: superficies de agua (b), suelo desnudo (c), y bosque denso (d). Si se comparan cada una de estas firmas espectrales con las curvas típicas de reflectancia (a, derecha) se tiene que: i) el porcentaje de reflectancia está normalizado (eje Y), y ii) la versión corregida se ajusta mucho mejor a los valores esperados que la versión original. También, se observa que la banda 1 (azul) sufrió el mayor ajuste entre todas las cinco bandas de la imagen *RapidEye* porque esta es una de las bandas más susceptibles a la bruma atmosférica.

Con los 32 cuadrantes corregidos en términos radiométricos y atmosféricos se creó un **mosaico** de imágenes para el área de estudio. Un mosaico es una unión entre dos o más imágenes satelitales¹⁸⁵. Este proceso de creación de una gran imagen se realiza con base en la georreferenciación que tienen los cuadrantes involucrados. Con la localización correcta de cada una de las imágenes emerge otra dificultad técnica en el proceso de creación del mosaico: la respuesta espectral capturada por el sensor en cada una de las imágenes está determinada por la fecha o las condiciones atmosféricas de captura.

Para que la superposición de imágenes sea lo mejor posible es necesario considerar los aspectos de antigüedad de la imagen y presencia de nubes¹⁸⁶. Como primer criterio, las imágenes antiguas irán debajo de las más actuales, de esta forma se prioriza la información más actualizada posible. Por su parte, las nubes juegan un rol determinante en el ordenamiento de las imágenes que componen el mosaico: las imágenes con presencia de nubes irán debajo porque la intención es clasificar la mayor superficie posible del área de estudio. Por último, para evitar errores de cualquier tipo en el perímetro de los archivos ráster producidos, se estableció un buffer de 1 km en el borde de toda el área del mosaico, el cual fácilmente fue removido en el producto final. La aplicación de estos criterios permitió elaborar un mosaico con una presentación de calidad.

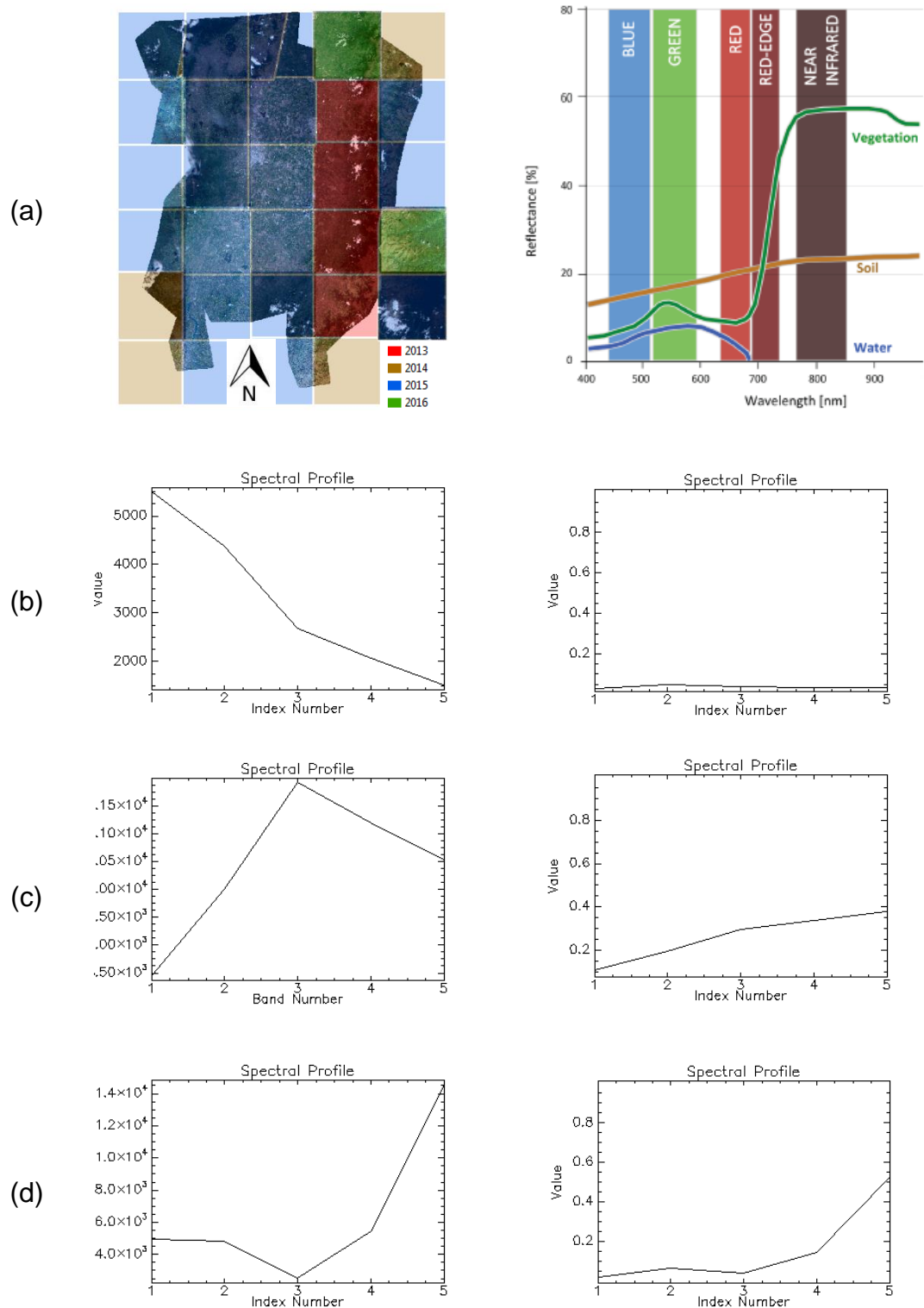
¹⁸³ DUARTE, *et al.* Op. cit.

¹⁸⁴ HARRIS GEOSPATIAL SOLUTIONS. Op. cit.

¹⁸⁵ DUARTE, *et al.* Op. cit.

¹⁸⁶ *Ibíd.*

Figura 34. 32 cuadrantes RapidEye, fecha de adquisición y firmas espectrales ajustadas.



Fecha de adquisición, composición natural (bandas 321) (a, izquierda). Curvas típicas de reflectancia espectral de algunas coberturas (a, derecha). Comparación entre la firma espectral original (izquierda) y la calibración radiométrica (derecha) de: agua (b), suelo desnudo (c), y bosque denso (d). Fuente: PLANET LABS. Op. cit.

La creación del mosaico empleando los 32 cuadrantes *RapidEye* se realizó con la herramienta **Mosaicking > Seamless Mosaic** de ENVI. Esta herramienta es intuitiva y cuenta con las opciones necesarias para aplicar los criterios de antigüedad y de cobertura nubosa. Con la opción **Seamless Mosaic > Add Scenes** se cargaron todas las 32 imágenes al tiempo para ahorrar tiempo. Con las imágenes cargadas se establecieron parámetros para:

- **Valores de datos a ignorar:** los ND de una imagen tienen valores entre 0 (cero) y 255. Los valores de 0 (cero) corresponden a los píxeles de fondo de la escena, y por tal motivo, este valor se ingresó en **Seamless Mosaic > Data Ignore Value**. Realizar este proceso disminuyó el tamaño del archivo *tagged image file format* (TIFF) del mosaico, evitando posibles interferencias de los valores del fondo de la imagen con el resto de píxeles¹⁸⁷.
- **Orden de la escena:** se priorizaron las imágenes más actuales con menor presencia de nubes. La configuración se ejecuta empleando los controles de **Seamless Mosaic > Order**.
- **Distancia de difuminado entre cuadrantes:** se estableció una distancia prudente de solapamiento entre imágenes para asegurar un mosaico de superficie continua y sin vacíos. La distancia de **200** píxeles (1 km) se ingresó en **Seamless Mosaic > Seamlines > Feathering > Feathering Distance**.

El proceso de la creación del mosaico siguió la directriz de presentar las imágenes más actuales con menor presencia de nubes, pero esto no quiere decir que se eliminó totalmente la cobertura nubosa. La presencia de nubes en una imagen de satélite se considera contaminación y afecta directamente el procesamiento y la discriminación de coberturas de la tierra¹⁸⁸. Por tal motivo, fue necesario **enmascarar y extraer la superficie de nubes** restante en el mosaico. Esto se realizó de forma manual, empleando la herramienta **Edit > Create Features** de ArcGIS, para digitalizar a una escala de 1:5.000 la cobertura nubosa restante del mosaico. Digitalizar con dicha escala garantizó la identificación total de las nubes disminuyendo los errores de sub/sobre estimación de área. El mosaico se presenta en la Figura 35 (a, página 155).

Posteriormente, esta área de nubes digitalizada funcionó como una «máscara» con la que se extrajo la superficie nubosa del mosaico. Es decir, la capa de nubes borró su área correspondiente de la imagen. Este proceso se realizó con la herramienta **Spatial Analyst toolbox > Extraction toolset > Extract by Mask** de ArcGIS.

Otro de los problemas que aparece en la clasificación de coberturas de la tierra a partir de imágenes satelitales es la alta correlación estadística entre bandas espectrales¹⁸⁹. En este caso, la introducción de bandas correlacionadas en la

¹⁸⁷ HARRIS GEOSPATIAL SOLUTIONS. Seamless Mosaic [en línea]. Harris Geospatial Solutions, 2017. [Consultado el 21 de enero de 2017]. En: <http://www.harrisgeospatial.com/docs/MosaicSeamless.html>

¹⁸⁸ DUARTE, et al. Op. cit.

¹⁸⁹ CHUVIECO. Op. cit.

clasificación no aporta más información y lo que se hace es eliminar los componentes que absorban un escaso porcentaje de la varianza original para quedarse con los que absorban mayor variabilidad¹⁹⁰. Esto se hace encontrando un nuevo conjunto de ejes ortogonales que tienen su origen en la media de los datos y que se rotan para maximizar la varianza de los datos¹⁹¹. Es decir que, el valor medio de reflectancia de cada píxel se compara con el valor de los vecinos para cada banda espectral. Es precisamente con las bandas de mayor variabilidad que la segmentación de imágenes de satélite funciona mejor¹⁹². Para identificar las bandas menos correlacionadas se empleó la herramienta ***Spatial Analyst toolbox > Multivariate toolset > Band Collection Statistics*** de ArcGIS. Esta herramienta calcula estadísticas para un conjunto de bandas de ráster, y produce un archivo de texto que contiene la matriz de correlación entre cada una de las bandas de la imagen¹⁹³. De esta forma se identificaron las bandas menos correlacionadas de las imágenes *RapidEye*.

En el Cuadro 22 se presenta la matriz de **correlación entre bandas espectrales** para el mosaico con imágenes *RapidEye* del área de estudio. La conclusión principal es que las tres bandas menos correlacionadas para este conjunto de datos fueron la banda 1 (azul), banda 3 (rojo) y banda 5 (infrarrojo), siendo la banda infrarroja la que aporta la mayor variabilidad de las tres. Esto se demuestra con el valor negativo de los datos de la banda 5 con la banda 1 y la banda 3. Por otra parte, las bandas más correlacionadas fueron la banda 2 (verde) y la banda 4 (borde del rojo), y no se tuvieron en cuenta para la clasificación de coberturas. Una ganancia adicional de seleccionar las tres bandas espectrales que absorben la mayor variabilidad es la reducción de las dimensiones del conjunto de datos. Esto acelera el procesamiento de los datos y al tiempo disminuye el tamaño de los productos, facilitando la ejecución iterativa de la calibración del modelo de clasificación. Esta es la etapa siguiente en la metodología de clasificación de coberturas de la tierra.

Cuadro 22. Matriz de correlación entre bandas para el mosaico RapidEye.

Banda	1	2	3	4	5
1	1	0,92518	0,88385	0,65577	-0,08385
2	0,92518	1	0,91135	0,79994	0,06421
3	0,88385	0,91135	1	0,77618	-0,15875
4	0,65577	0,79994	0,77618	1	0,31395
5	-0,08385	0,06421	-0,15875	0,31395	1
\bar{X}	0,67619	0,740136	0,682526	0,709168	0,227112

Fuente: elaboración propia.

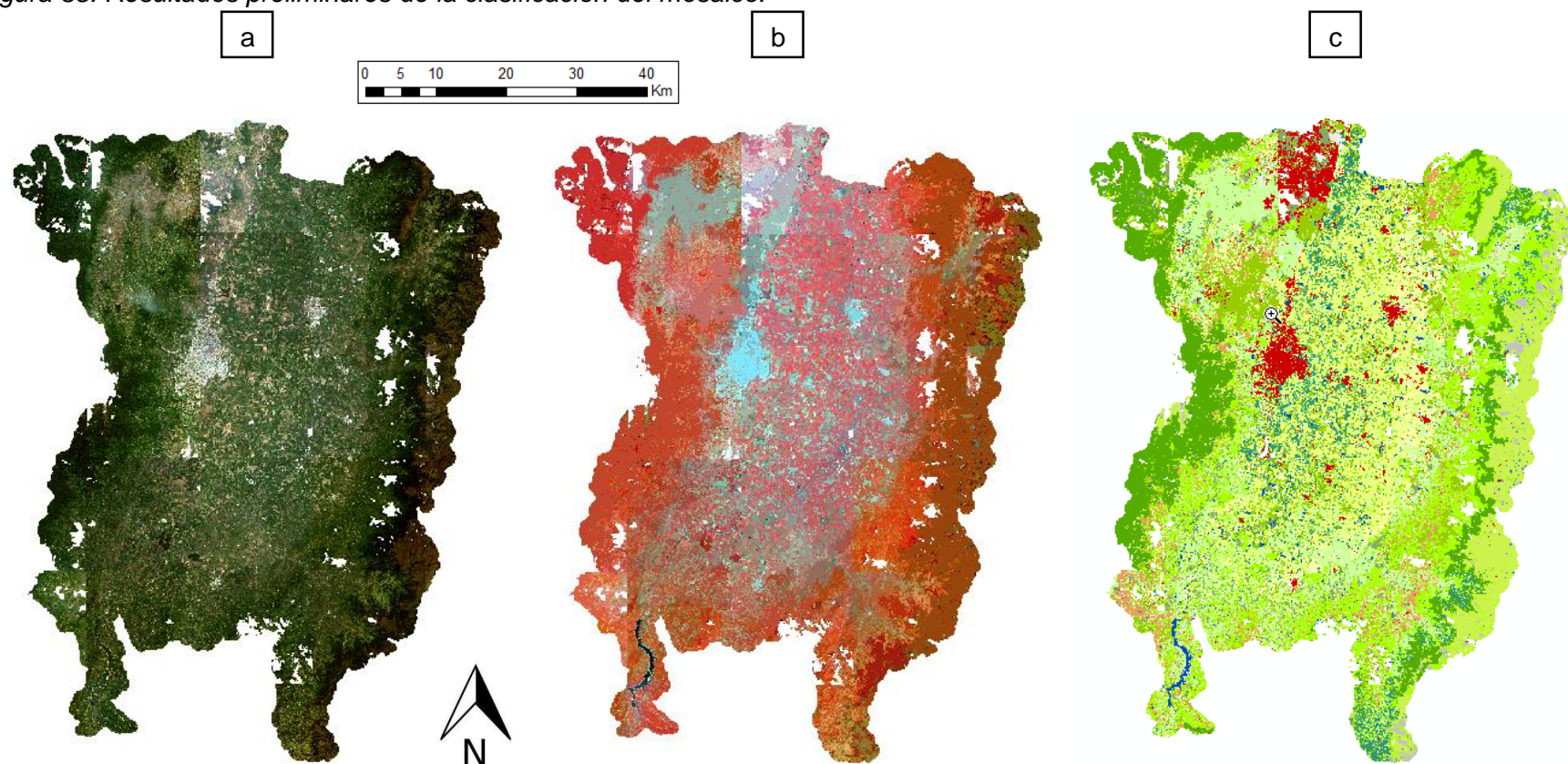
¹⁹⁰ *Ibíd.*

¹⁹¹ RICHARDS, John. Remote Sensing Digital Image Analysis: An Introduction. Berlin: Springer-Verlag, 1999. p. 240.

¹⁹² HARRIS GEOSPATIAL SOLUTIONS. Extract Segments [en línea]. Harris Geospatial Solutions, 2017. [Consultado el 10 de febrero de 2017]. En: <http://www.harrisgeospatial.com/docs/segmentonly.html>

¹⁹³ ESRI. Band Collection Statistics [en línea]. _____, 2017. [Consultado el 5 de febrero de 2017]. En: <http://desktop.arcgis.com/es/arcmap/latest/tools/spatial-analyst-toolbox/band-collection-statistics.htm>

Figura 35. Resultados preliminares de la clasificación del mosaico.

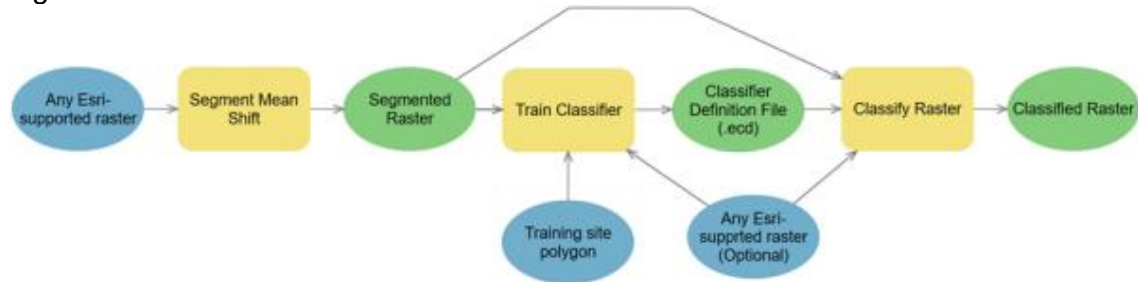


(a) color natural (bandas 321), (b) segmentación. (c) clasificación automática (Kappa de 84%). Escala numérica 1:600.000. Fuente: elaboración propia.

IV. Calibración del modelo de clasificación

La clasificación de las características de una imagen de satélite es una herramienta que ayuda a la automatización del análisis de datos de teledetección. El proceso de clasificación es complejo e involucra diferentes fases en secuencia. Las fases del modelo de clasificación de esta investigación incluyen: **i) segmentación, ii) muestreo, iii) entrenamiento, iv) clasificación y v) evaluación de precisión** (Figura 36). Decir que el proceso de clasificación está articulado, quiere decir que la configuración y los productos parciales de cada fase sirven de insumo para la siguiente fase y al final contribuyen al resultado esperado que es la clasificación el resultado final de clasificado. Frente a este panorama, la solución es calibrar el modelo de clasificación completo en un área menor, ajustando de forma iterativa la configuración hasta alcanzar un resultado óptimo, cuyos parámetros puedan ejecutarse para todo el mosaico¹⁹⁴. Esta estrategia de procesamiento iterada produjo resultados excelentes de clasificación. Por lo tanto, la calibración del modelo es el componente más profundo y técnico de la clasificación de coberturas de la tierra del área de estudio.

Figura 36. Modelo de calibración de coberturas de la tierra.



Fuente: ESRI. Understanding segmentation and classification [en línea]. _____, 2017. [Consultado el 20 de febrero de 2017]. En: <http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/understanding-segmentation-and-classification.htm>

El primer componente del modelo de calibración es la segmentación. La segmentación es el proceso de dividir una imagen en muchos objetos agrupando píxeles vecinos con características espectrales y espaciales comunes¹⁹⁵. Como se ha dicho, la **segmentación** tiene ventajas sobre las técnicas de clasificación de coberturas tradicionales orientadas por píxeles. Para que la clasificación represente fielmente la realidad de la superficie terrestre, es crucial configurar la manera cómo son agrupados los píxeles para crear los segmentos. Esta agrupación no es perfecta y la literatura hace hincapié en que no hay algoritmo de segmentación que delimite de forma correcta todos los objetos de una

¹⁹⁴ SOUSA, *et al.* Op. cit. & MAKINDE, *et al.* Op. cit.

¹⁹⁵ CLARK, Barnaby; PELLIKKA, Petri. Landscape analysis using multiscale segmentation and object orientated classification. En: Recent advances in remote sensing and geoinformation processing for land degradation assessment. 2009. Vol. 8; p. 323-341. & SOUSA, *et al.* Op. cit. & MAKINDE, *et al.* Op. cit.

imagen¹⁹⁶. Cada algoritmo tiene sus ventajas y desventajas, las que deben tenerse en cuenta al momento de optar por uno en particular.

En esta clasificación de coberturas de la tierra se empleó el algoritmo de segmentación llamado *Mean Shift Segmentation* propuesto por Comaniciu y Meer¹⁹⁷. Este algoritmo tiene como característica principal que, al mismo tiempo que agrupa píxeles adyacentes de características espectrales similares, permite controlar la cantidad de suavizado espacial y espectral durante el proceso de agrupamiento para obtener las entidades de interés¹⁹⁸. La disponibilidad de estos parámetros es clave para segmentar una imagen en función de sus características de captura (resoluciones espectro-espacio-temporales de la imagen) y de los objetivos perseguidos con la clasificación. Sin embargo, su principal desventaja es que requiere un número significativo de pruebas piloto y tiempo para ajustar los parámetros que produzcan una segmentación óptima de los objetos de la imagen.

Una buena segmentación es aquella en la que: «(i) las áreas son uniformes y homogéneas, (ii) las áreas son significativamente diferentes de las vecinas, (iii) las áreas tienen un interior simple sin muchos agujeros, y (iv) los límites de las áreas son simples, estables y espacialmente precisos»¹⁹⁹. En la Figura 37 se muestra un ejemplo interesante de segmentación de uno de los cuadrantes calibrados empleando *Mean Shift*. Puede notarse que la imagen segmentada replica las cuatro características referenciadas de una buena segmentación.

Técnicamente, *Mean Shift* utiliza una ventana móvil que calcula un valor medio de píxeles para determinar cuáles deben incluirse en cada segmento, y a medida que la ventana se mueve por la imagen, vuelve a calcular de forma iterativa el valor para asegurarse de cada uno de los segmentos sea adecuado²⁰⁰. El resultado es una agrupación de los píxeles de la imagen en un segmento caracterizado por un color medio. Para realizar la segmentación de la imagen se empleó la herramienta ***Spatial Analyst > Segmentation and Classification > Segment Mean Shift*** de ArcGIS 10. Luego de realizar 44 pruebas piloto, esta fue la configuración definitiva de esta herramienta:

¹⁹⁶ JOHNSON, Brian y ZHIXIAO, Xie. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. En: ISPRS Journal of Photogrammetry and Remote Sensing. 2011. Vol. 66, No. 4; p. 473-483. & SCHULTZ, Bruno, *et al.* Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in southeastern Brazil. En: Remote Sensing [en línea]. 2015. Vol. 7, No. 11; p. 14482-14508.

¹⁹⁷ COMANICIU, Dorin y MEER, Peter. Mean shift: A robust approach toward feature space analysis. En: IEEE Transactions on pattern analysis and machine intelligence. 2002. Vol. 24, No. 5; p. 603-619.

¹⁹⁸ ESRI. Segment Mean Shift Function [en línea]. _____, 2017. [Consultado el 22 de febrero de 2017]. En: <http://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/segment-mean-shift-function.htm>

¹⁹⁹ JOHNSON y ZHIXIAO. Op. cit., p. 474.

²⁰⁰ ESRI, Segment Mean Shift Function. Op. cit.

Figura 37. Ejemplo de segmentación en el área de estudio.

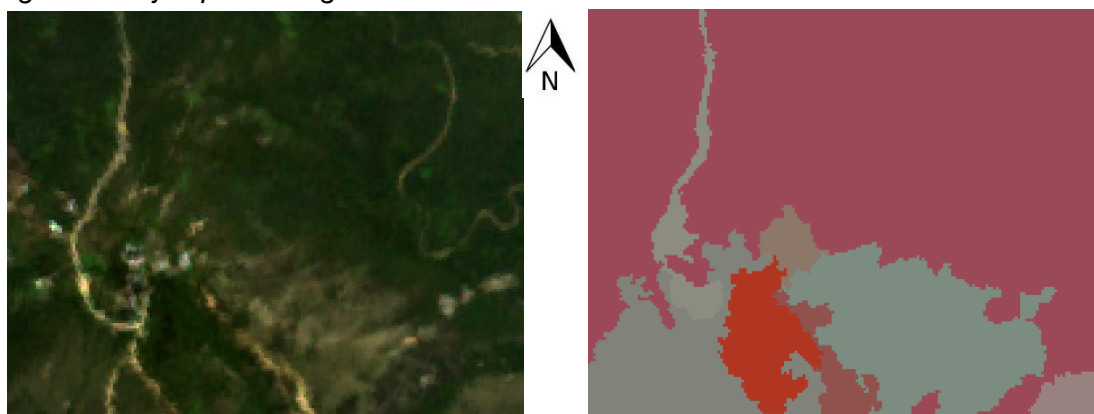


Imagen en color natural (izquierda, bandas 3-2-1), imagen segmentada (derecha, bandas 1-3-5). Nótese como en la derecha se agruparon los diferentes tipos de vegetación en color rojo y el suelo desnudo en color gris. Escala numérica 1:5.000. Fuente: elaboración propia.

- **Detalle espectral (1 a 20):** ajusta qué tan importantes son las diferencias espectrales de los objetos en la imagen. Se ingresó en **Segment Mean Shift > Spectral Detail** un valor de **15** porque se observó que mientras los valores altos sobre-segmentaban la imagen y creaban demasiados segmentos, los bajos sub-segmentaban el área sin diferenciar los objetos deseados.
- **Detalle espacial (1 a 20):** establece un nivel de importancia a la proximidad espacial entre los objetos de la imagen. Un valor de **15** se aplicó en **Segment Mean Shift > Spatial Detail** dado que los valores más altos dividieron áreas continuas en muchas partes pequeñas (e.g. área urbana), y los bajos no establecieron un límite aceptable entre objetos fácilmente diferenciables (e.g. ríos y áreas o agrícolas).
- **Tamaño mínimo de segmento (píxeles):** determina el área necesaria para que un bloque de píxeles sea un objeto, y si no cumple esta cuota, se une automáticamente al vecino más semejante. Considerando que la UMC de un mapa a escala 1:25.000 es de una (1) hectárea²⁰¹ y que un píxel *RapidEye* tiene 25 mt², en **Segment Mean Shift > Min Segment Size** se ingresó un valor de **400**²⁰².

El producto segmentado del mosaico se presenta en la Figura 35 (b, página 155). Esta imagen se creó con las bandas menos correlacionadas (1-3-5). A grandes rasgos, la combinación de sus colores expresa lo siguiente:

- Las variaciones de rojo son los objetos con alto contenido de vegetación.
- El gris representa los pastos.
- El azul grisáceo muestra las infraestructuras urbanas o construidas.

²⁰¹ SALITCHEV. Op. cit.

²⁰² Resultado de dividir la UMC de una hectárea por el tamaño de píxel *RapidEye* en metros (10.000 mt²/25 mt² = 400).

- El color café enseña los objetos con vegetación exigua y suelo desnudo.
- El azul oscuro indica los objetos con agua.

Esta descripción general permite interpretar el producto segmentado, sin embargo, es clave considerar las cifras del Cuadro 23 para entender la riqueza del proceso de segmentación del mosaico. Este cuadro dimensiona la resolución espacial de la imagen *RapidEye* a escala del área de estudio y permite apreciar la disminución de datos, así: más de 300 millones de píxeles (900 millones entre las tres bandas espectrales) a 386.304 objetos de UMC de una (1) hectárea con la segmentación, y finalmente, a menos de 69 mil objetos con la fase posterior de vectorización. Esta reducción del conjunto de datos es una propiedad «oculta» de la imagen segmentada del mosaico que favoreció y agilizó la planeación y ejecución de distintas pruebas porque acotó sustancialmente el volumen de los datos a procesar. Lo más significativo es que con este tratamiento, el producto de coberturas reflejó perfectamente la superficie de la tierra del área de estudio a una escala de 1:25.000, con objetos mayores a una (1) hectárea.

Cuadro 23. Cifras del proceso de segmentación.

Elemento	Área (en mt ²)	Elemento	Número
Píxel <i>RapidEye</i>	25	Píxeles área (por banda espectral)	325.600.000
Área estudio*	8.140.000.000	Objetos segmentación (1 ha = 400 píxeles)	386.052
		Vectorización (objeto mínimo 1 ha)	68.793

*1 km² equivale a 1.000.000 m²

Fuente: elaboración propia.

Con este conjunto de parámetros se segmentó una muestra del 20% del área de estudio, es decir, siete (7) de los 32 cuadrantes. Los siete cuadrantes fueron seleccionados teniendo en cuenta su distribución geográfica y tratando de garantizar que dentro de ellos se observara la presencia de diferentes tipos de coberturas. Xiaoxiao Li, *et al.*²⁰³, sustentó que esta es una forma eficiente de tratar áreas de estudio muy extensas y de superficie heterogénea. El proceso de extracción de esta muestra se realizó mediante un **muestreo intencional**. Este tipo de muestreo no es probabilístico y se sustenta en el conocimiento previo del problema para aproximarse a la representatividad de las muestras²⁰⁴. Hay muchos tipos de muestreos intencionales, pero el que orientó este proceso fue un **muestreo con casos de máxima variación** dado que se tomaron en cuenta «todas las características consideradas relevantes de los individuos a estudiar en busca de obtener la mayor heterogeneidad entre ellos»²⁰⁵. Esta técnica de muestreo dirigió la calibración del modelo de clasificación porque garantizó

²⁰³ LI, Xiaoxiao, *et al.* Object-based land-cover classification for metropolitan Phoenix, Arizona, using aerial photography. *En*: International Journal of Applied Earth Observation and Geoinformation. 2014. Vol. 33; p. 321-330.

²⁰⁴ RINCÓN y BARRETO. Op. cit.

²⁰⁵ *Ibíd.*, p. 42.

abarcó en un área menor el total de clases de coberturas de la tierra a clasificar. Por lo tanto, la característica que se buscó en los siete (7) cuadrantes fue la inclusión de diferentes tipos de cobertura para asegurar una calibración óptima del modelo de clasificación.

Con los objetos identificados, el paso siguiente de la calibración consistió en extraer y analizar información espectral para las coberturas de la tierra definidas en la leyenda. En la clasificación de coberturas de la tierra orientada a objetos es imposible saber con certeza a qué clase de cobertura corresponde cada segmento. Por lo tanto, es necesario extraer la información espectral característica de las coberturas deseadas a través de un muestreo. Dadas las características del ejercicio, se apeló a un muestreo no probabilístico, en este caso «por cuotas». **El muestreo por cuotas** presupone un buen conocimiento de la forma cómo los individuos están distribuidos, asumiendo que hay estratos (grupos de individuos similares) y que el investigador los conoce ampliamente²⁰⁶. En este contexto, los estratos son las clases de coberturas de la tierra definidas en la leyenda y se admite que el investigador no se equivoca identificándolas en la imagen.

El muestreo espectral se realizó sobre la imagen segmentada (bandas 1-3-5), con la herramienta ***Spatial Analyst > Image classification > Training Sample Manager***. Esta herramienta consulta los valores espectrales de cada una de las muestras en la imagen, para luego consolidar un conjunto de muestras para cada clase de cobertura. En la literatura, el número de las muestras es variable, pero se reconoce que un número mayor de sitios de entrenamiento produce resultados más precisos²⁰⁷. En este trabajo, se aceptó el mínimo de **60 muestras** por tipo de cobertura propuesto por Clark y Pellikka²⁰⁸ porque aportó la información requerida por el modelo sin excesos.

Cada una de estas muestras de entrenamiento abarcó píxeles pertenecientes a la clase de interés. Las muestras se tomaron sobre la imagen segmentada en polígonos que contenían un mínimo de 100 píxeles, un tamaño significativo desde el punto de vista estadístico²⁰⁹. Para lograrlo, se apeló a un tratamiento estadístico de descarte de datos atípicos con valores de +20% y -20% de la media registrada. Esto aseguró que el conjunto de todas las 60 muestras de cada de cobertura presentara la menor variación estadística de valores espectrales.

De otra parte, en la Figura 38 se presenta el **muestreo** de uno de los siete cuadrantes de calibración (izquierda) con los parámetros estadísticos para tres coberturas identificadas (derecha). En este cuadrante se reconocieron 15 tipos de cobertura para las que se crearon 60 polígonos por clase. Las coberturas con menos píxeles (e.g. cuerpos de agua artificiales) no alcanzaron la cuota de los 60 polígonos por su área de cobertura, pero como la calibración se replicó para otros seis cuadrantes, se completó el número establecido. La caña de azúcar se

²⁰⁶ Ibíd.

²⁰⁷ CHUVIECO. Op. cit.

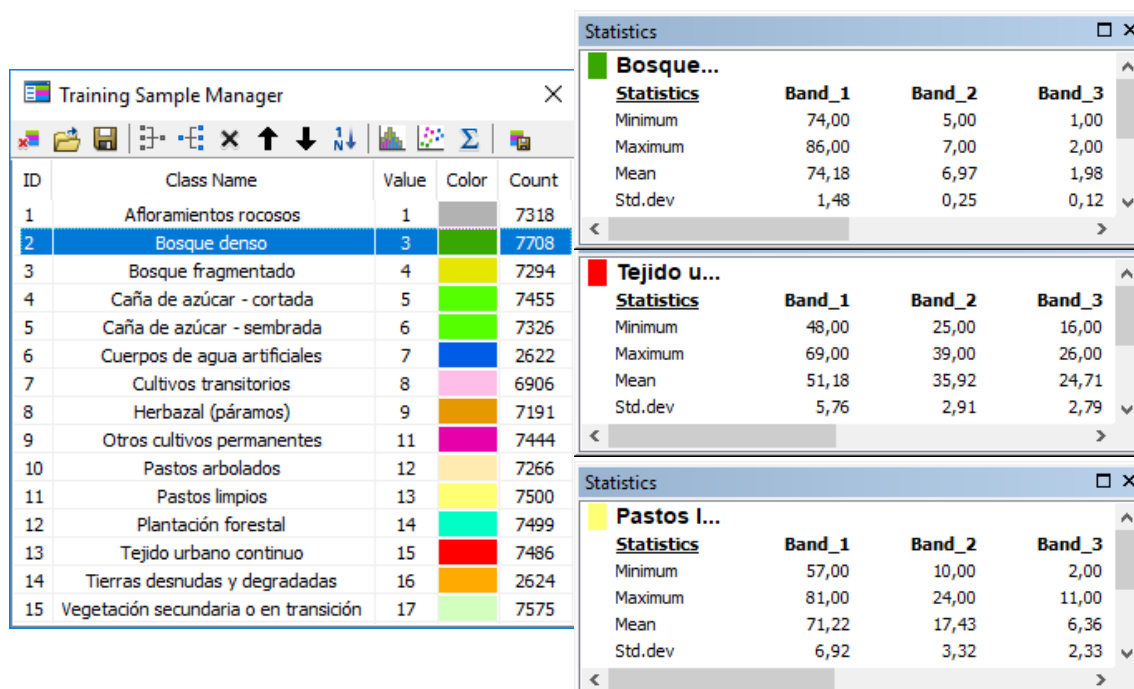
²⁰⁸ CLARK y PELLIKKA. Op. cit.

²⁰⁹ ESRI, Understanding segmentation and classification. Op. cit.

dividió estratégicamente en dos clases (cortada y cosechada) porque la ausencia y presencia respectiva de vegetación hizo que su firma espectral variara demasiado. Luego del proceso de clasificación, se consolidó una sola cobertura de «caña de azúcar». En la columna *value* de la tabla de atributos, se ingresó un valor que es un código establecido por el usuario para identificar cada cobertura. Con la aplicación de estas reglas se produjeron conjuntos de muestras con la menor variabilidad posible (ver fila Std.dev en la Figura 38, derecha).

Antes de avanzar a las etapas posteriores, se tuvo que resolver la siguiente situación: con las 60 muestras estadísticamente correctas por cobertura, se debía entrenar y clasificar los datos para luego validar los resultados con ese mismo conjunto de datos. Para superar esta situación, se tuvo en cuenta lo siguiente. Löw, Knöfel, y Conrad²¹⁰, clasificaron y validaron de forma independiente coberturas de la tierra tomando una proporción de 50% de datos para clasificar y 50% para validar, haciendo un gran muestreo para todo su modelo de clasificación. Esta estrategia disminuye ampliamente los tiempos de muestreo. En el caso de esta investigación, para esta etapa de calibración se replicó lo planteado por estos autores, pero tomando en cuenta el número de 60 muestras por cobertura de la tierra, y se ajustó una proporción aleatoria de **60% (36 muestras)** de datos para clasificar y **40% (24 muestras)** para validar. De esta manera, el modelo contó con mayor información para producir las clases sin menospreciar considerablemente la etapa de validación.

Figura 38. Resultados del muestreo en uno de los cuadrantes de calibración.



Fuente: elaboración propia usando la herramienta *Training Sample Manager* de ArcGIS 10.

²¹⁰ LÖW, Fabian; KNÖFEL, Patrick y CONRAD, Christopher. Analysis of uncertainty in multi-temporal object-based classification. En: ISPRS Journal of Photogrammetry and Remote Sensing. 2015. Vol. 105; p. 91-106.

Para realizar esta división se aplicó un **muestreo estratificado** sobre el conjunto de datos. Este tipo de muestreo consiste en «realizar un muestreo aleatorio simple por cada estrato [grupo de individuos muy similares entre sí y diferente de los demás]»²¹¹. Esto quiere decir que todas las muestras pertenecientes a una clase de cobertura tuvieron la misma probabilidad de selección para los grupos de clasificación y validación. Con la solución propuesta en términos teóricos, se tuvo que proyectar su instrumentalización de forma automática en el software.

ArcGIS 10 no cuenta con una herramienta que directamente ejecute un muestreo estratificado, pero con la integración de dos herramientas fue posible realizarlo. Dichas herramientas fueron: **Analysis > Extract > Select y Geostatistical Analyst > Utilities > Subset Features** de ArcGIS 10. Mientras la primera selecciona las coberturas de la tierra por su nombre automáticamente, la segunda divide ese conjunto de datos seleccionado en las proporciones deseadas y exporta los dos grupos de datos para clasificar y validar. Así las cosas, con la clarificación del vínculo entre las muestras con cada una de las fases posteriores de calibración, se procedió a la etapa de entrenamiento y clasificación de los datos.

Cuando se dispone de un conjunto de clases y sus firmas espectrales características a partir de una muestra, el paso siguiente consta en adjudicar una clase a cada uno de los segmentos del área a clasificar. Esto se conoce como el **entrenamiento y clasificación de los datos**. El objetivo del proceso de entrenamiento es «proporcionar información espectral sobre las clases en la imagen a partir de la cual se puede desarrollar una firma espectral representativa para cada cobertura»²¹². Si cada clase de la imagen tiene una firma única, entonces la clase a la que pertenece cada píxel puede determinarse fácilmente a partir de la muestra. En el paso anterior se obtuvo una muestra estadísticamente correcta de 60 segmentos por cobertura, pero en esta etapa se debía entrenar y clasificar los datos para luego validar los resultados con ese mismo conjunto de segmentos muestreados. Para superar esta situación sin repetir muestreos se tuvieron en cuenta algunas consideraciones.

Con los primeros métodos de entrenamiento se obtenían resultados poco precisos, sin embargo, al refinarlos, se elaboraron técnicas de entrenamiento más avanzadas que se conocen hoy en día (algoritmos basados en inteligencia artificial)²¹³. Estos métodos se basan en el uso de redes neuronales artificiales que imitan el funcionamiento de las neuronas reales en las tareas de aprendizaje. Son muchos los algoritmos de entrenamiento y clasificación de datos, sin embargo, el trabajo desarrollado por Şatir y Berberoğlu²¹⁴, clarifica a sobremano el asunto.

²¹¹ RINCÓN y BARRETO. Op. cit., p. 40.

²¹² JAMES. Op. cit., p. 772.

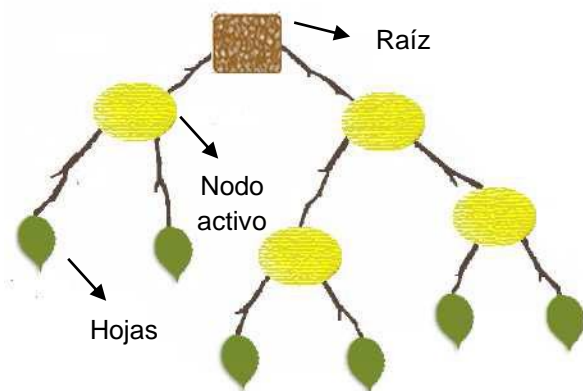
²¹³ CHUVIECO. Op. cit. & ŞATIR y BERBEROĞLU. Op. cit.

²¹⁴ ŞATIR y BERBEROĞLU. Op. cit.

A partir de una serie de criterios (precisión, velocidad, versatilidad de configuración, precisión con diferente tamaño de muestra y estabilidad), Şatır y Berberoğlu²¹⁵ compararon algoritmos diferentes de clasificación y concluyeron que el algoritmo *Random Trees* (árboles de decisión) es el mejor en precisión, estabilidad, precisión con diferentes muestras, seguido por *Support Vector Machine* (SVM). Mientras *Random Trees* se ejecuta más rápido en una computadora estándar, SVM requiere una máquina poderosa de procesamiento, detalle muy importante considerando los recursos disponibles para esta investigación. Por estas ventajas, el algoritmo *Random Trees* fue seleccionado.

El algoritmo *Random Trees* requiere un análisis exhaustivo de las respuestas espectrales de las clases de cobertura y del conjunto de datos auxiliares. Básicamente, consiste en ir haciendo preguntas a cada segmento cuya respuesta positiva o negativa conducirá a otra pregunta, para finalmente obtener la clase de cobertura a la que pertenece²¹⁶. En la Figura 39 se observa la estructura de un árbol de decisión: «una raíz (punto inicial), un nodo activo (nodo de reglas) y las hojas (clases de cobertura)»²¹⁷. La raíz es el punto inicial del árbol, los nodos activos tienen las reglas para crear las hojas, y estas hojas son la salida o segmentos que pertenecen a la misma clase o son asignados a otra clase en particular.

Figura 39. Arquitectura de un árbol de decisión.



Fuente: ŞATIR y BERBEROĞLU. Op. cit.

Un árbol de decisión se construye a partir del conjunto de muestras, que contiene los segmentos completamente descritos a partir de una clase de cobertura y un conjunto de atributos. Los atributos son una colección de propiedades espaciales y espectrales que contiene toda la información sobre un segmento. Finalmente, es común que los árboles de decisión operen con unas muestras de entrada, unas reglas de división y unas reglas de finalización. Todos estos parámetros se ingresaron en la herramienta que entrenó los datos: *Train Random Trees Classifier* de ArcGIS.

Train Random Trees Classifier es una técnica poderosa y versátil que puede trabajar con imágenes segmentadas y con otros datos auxiliares. Esta herramienta es un clasificador supervisado de aprendizaje automático basado en la construcción de una multitud de árboles de decisión, la elección y clasificación de subconjuntos aleatorios de muestras para cada árbol y el uso de

²¹⁵ ŞATIR y BERBEROĞLU. Op. cit.

²¹⁶ CHUVIECO. Op. cit.

²¹⁷ ŞATIR y BERBEROĞLU. Op. cit., p. 33.

la salida del árbol más frecuente como la clasificación general²¹⁸. Esta herramienta se localiza en ***Spatial Analyst > Segmentation and Classification > Train Random Trees Classifier*** de ArcGIS. Estos fueron los parámetros empleados:

- **Número máximo de árboles:** es la cantidad de árboles que conforman el bosque. Al aumentar el número de árboles se obtendrán dos cosas: a) tasas de precisión más elevadas (esta mejora se estabiliza al final) y b) el tiempo de procesamiento aumenta de forma lineal. Ingresando un valor de **50** se obtuvieron los mejores resultados. Valores superiores no mejoraron la precisión.
- **Profundidad máxima de cada árbol:** en la analogía esto representa el número de ramas del árbol (nodos activos). La profundidad es otra forma de indicar la cantidad de reglas que puede crear cada árbol para llegar a una decisión. Un valor de **30** nodos activos fue empleado y no hubo necesidad de un mayor número de reglas de decisión para asignar una clase.
- **Número máximo de muestras por clase:** la cantidad de muestras que se van a utilizar para definir cada clase. Se estableció un valor de **0** para que el sistema empleará todos los segmentos muestreados.
- **Archivo ráster auxiliar:** existe la opción de incorporar datos ráster adicionales para generar atributos. Fueron probados en esta opción un DEM y el *Normalized Difference Vegetation Index* (NDVI), pero el primero produjo una clasificación más estable y precisa (conclusión compartida con Şatir y Berberoğlu²¹⁹), por lo tanto, ese fue el usado en esta investigación.
- **Atributos usados:** conjunto de atributos espaciales y espectrales de todos los segmentos usados en la clasificación. Estos fueron: **COLOR** (valores RGB de la imagen), **COUNT** (número de píxeles de cada segmento), **COMPACTNESS** (grado de circularidad de un segmento), **RECTANGULARITY** (grado de rectangularidad de un segmento), **MEAN** (ND promedio del ráster auxiliar para cada segmento), y **STD** (desviación estándar de los ND del ráster auxiliar por segmento). Los dos últimos atributos se habilitan con la capa auxiliar. En este caso, el DEM.

El resultado de la fase de entrenamiento es un archivo *esri classifier definition* (ecd) con el algoritmo de entrenamiento *Random Trees* que contiene las firmas espectrales características de cada una de las clases de coberturas entrenadas. Este archivo de entrenamiento es el insumo principal empleado en la penúltima fase del modelo: **la clasificación**. El entrenamiento de datos y su clasificación son procesos complementarios, mientras que con el entrenamiento se crea la firma espectral característica de cada cobertura en la imagen, con la clasificación

²¹⁸ ESRI. Train Random Trees Classifier [en línea]. _____, 2017. [Consultado el 30 de febrero de 2017]. En: <http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/train-random-trees-classifier.htm>

²¹⁹ ŞATIR y BERBEROĞLU. Op. cit.

se extrapolan estos atributos y se determina por semejanza los tipos de cobertura al que pertenece cada uno de los segmentos²²⁰.

La herramienta **Classify Raster** de ArcGIS realiza una clasificación de la imagen tal y como se especifica en el archivo «ecd» creado en el entrenamiento. Esta herramienta contiene todos los clasificadores compatibles. De este modo, el archivo de definición de clasificador generado en el entrenamiento activará el clasificador correspondiente cuando se ejecute la herramienta de clasificar: en este caso el algoritmo *Random Trees*²²¹. Esta herramienta se localiza en **Spatial Analyst > Segmentation and Classification > Classify Raster**, y esta fue la configuración empleada:

- **Ráster de entrada:** es la imagen que se desea clasificar. En este campo se ingresó el producto **ráster segmentado** a partir de las bandas 1-3-5.
- **Archivo de definición de clasificador de Esri:** es el archivo con la información de entrenamiento de las muestras. En este caso, el archivo ingresado es el «ecd» **entrenado con el algoritmo *Random Trees***.
- **Ráster auxiliar:** existe la opción de incorporar ráster auxiliares para generar atributos y adicionar información al clasificador. El **DEM** se empleó como capa auxiliar porque en las pruebas generó una clasificación más estable y precisa.

El entrenamiento automático y la clasificación de los datos son los procesos que abarcan la mayor cantidad de tiempo de procesamiento del modelo de clasificación propuesto. Para el entrenamiento y clasificación de cada uno de los siete cuadrantes de la fase de calibración fueron necesarias siete horas (aproximadamente 1 hora por cuadrante). Ya con la imagen clasificada, el último paso del modelo de clasificación es la validación del producto.

La **evaluación de la precisión** es una parte importante de cualquier proyecto de clasificación. Lo que se pretende es evaluar la exactitud que tuvo el modelo para clasificar cada cobertura de la imagen. Dicha estimación nace del análisis de una serie de áreas de prueba obtenidas del mismo modo que las áreas de entrenamiento, permitiendo obtener una apreciación real de los errores²²². La lógica del proceso de validación es contrastar la clase de cobertura clasificada con la realidad del terreno en la localización de cada una de las áreas de prueba. La forma más habitual de evaluar la precisión de un mapa clasificado es tomar las áreas de prueba y compararlas con los datos clasificados de una matriz de confusión²²³.

Una matriz de confusión es método simple y apropiado de estimar los errores en la clasificación. Con este tipo de análisis, se obtiene, no sólo una caracterización del error cometido, sino también una medida sobre el ajuste de las clases

²²⁰ HARRIS. Op. cit.

²²¹ ESRI, Understanding segmentation and classification, Op. cit.

²²² CHUVIECO. Op. cit.

²²³ ESRI, Understanding segmentation and classification, Op. cit.

consideradas a la realidad y de los parámetros utilizados para caracterizarlas²²⁴. Por lo tanto, la matriz puede emplearse como información de retro-alimentación para ajustar alguna fase del modelo de clasificación. La matriz se compone de las clases reales en el terreno en las filas (información del usuario) y de las clases obtenidas tras la clasificación en las columnas (información del modelo). Esta relación cruzada debe interpretarse como el número de segmentos que perteneciendo a la clase fila han sido clasificados en la clase columna. Así es posible estimar el número de segmentos clasificados correctamente e incorrectamente para cada clase de cobertura. Un ejemplo de matriz de confusión se presenta en el Cuadro 24. El índice de acuerdo *kappa* ofrece una evaluación general de la precisión de la clasificación.

Cuadro 24. Ejemplo de matriz de confusión.

Clase	c_1	c_2	c_3	Total	U_Accuracy*	Kappa
c_1	49	4	4	57	0,8594	-
c_2	2	40	2	44	0,9091	-
c_3	3	3	59	65	0,9077	-
Total	54	47	65	166	-	-
P_Accuracy**	0,9074	0,8511	0,9077	-	0,8916	-
Kappa	-	-	-	-	-	0,8357

* La exactitud del usuario muestra falsos positivos, donde los píxeles se clasifican incorrectamente como una clase conocida cuando deberían haberse clasificado como algo diferente. Es una sobrestimación de las clases.

** La exactitud del productor es un falso negativo, donde los píxeles de una clase conocida se clasifican como algo diferente a una clase. Es la subestimación de clases.

Fuente: ESRI. Compute Confusion Matrix [en línea]. _____, 2017. [Consultado el 5 de marzo de 2017]. En: <http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/compute-confusion-matrix.htm>

En esta investigación, la estimación de la precisión tomó como base un conjunto de puntos aleatorios a partir de los datos de referencia que fueron comparados con el producto clasificado en una matriz de confusión²²⁵. Los datos de referencia son los 60 segmentos muestreados para cada clase, pero como se explicó, mientras en la etapa de clasificación se tomó el 60% de las muestras, en la fase de validación ingresaron el **40% de las muestras (24 segmentos aleatorios)**. Llegados a esta etapa, el proceso automático de clasificación de coberturas de la tierra está completo. Sin embargo, tras una serie de experimentos, repetir el modelo completo de clasificación generó excelentes resultados. En ese sentido, la verdadera calibración del modelo completo se realizó por medio de iteraciones automáticas programadas.

Una **iteración** significa repetir un proceso hasta alcanzar el resultado o modelo de representación que más se acerque a la realidad. Como la clasificación de coberturas de la tierra es dependiente de las muestras ingresadas, en la búsqueda de un resultado con un alto grado precisión, el ajuste del modelo se resolvió calibrando el proceso de clasificación con diferentes datos de entrada

²²⁴ CHUVIECO. Op. cit.

²²⁵ ESRI, Understanding segmentation and classification, Op. cit.

de forma iterativa. Esto representa ejecutar el proceso de clasificación sobre la imagen de interés y comparar los resultados una y otra vez hasta obtener un producto optimizado²²⁶. Lo que se pretende con estas iteraciones es experimentar para mejorar la fiabilidad del modelo. En la fase de calibración el modelo se iteró **20 veces** utilizando diferentes conjuntos de datos para el entrenamiento y la validación mediante un muestreo estratificado. Se reconoce que un número de 20 iteraciones es un número pequeño comparado con las 100 iteraciones de Löw, Knöfel, y Conrad²²⁷, pero estas fueron suficientes para identificar las clases que mejoraban o empeoraban el rendimiento del modelo.

El proceso de clasificación iterativo se programó en archivos de conjuntos de órdenes o *script*. El *script* es una línea de comandos programada en lenguaje *Python* 2.7.12 que ejecuta de forma estructurada un proceso, en este caso, el modelo de clasificación de coberturas. El lenguaje de programación *Python* es muy usado en informática, tanto así que es el lenguaje base para correr ArcGIS sin interfaz gráfica. Ejecutar ArcGIS mediante línea de comandos es una opción que asegura estabilidad, eficiencia, confianza, y mayor destino de recursos del computador para el proceso realizado.

La estructura de los *scripts* desarrollados, se presentan de forma completa en los Anexo D (*script* de muestreo estratificado) y Anexo E (*script* de clasificación de coberturas). Ambos inician con un encabezado que tiene el título y la descripción propia de cada *script*. Luego, se puede identificar cómo se articulaban cada una de las fases mencionadas en el modelo de calibración mediante un llamado a cada una de las herramientas empleadas —división de muestra, entrenamiento, clasificación y evaluación de precisión (Figura 36, página 156)— La estructura del *script* está jerarquizada, lo que quiere decir que hay procesos dependientes de otros siendo ejecutados en un orden específico. Esta estructura se asemeja a la rama de un árbol.

Al ejecutar el *script* el usuario debe ingresar la dirección y el nombre de los insumos requeridos por el proceso —imagen, muestras, etcétera—. A partir de este momento el proceso se realiza de forma automática hasta su finalización. El núcleo operativo del *script* yace en el comando **ForLoop** de *Python*. Este comando permite crear bucles de procesamiento con los bloques de datos que el usuario ingresa siguiendo la secuencia apropiada, para luego crear un **Range** o número de veces que se ejecute dicho bucle²²⁸. Por ejemplo, para este ejercicio todos los productos se nombraron con un sufijo que indicó el número de iteración con el que fueron creados, y así el *script* relacionó cada bloque de datos con sus productos respectivos para cada fase.

²²⁶ STEFANSKI, Jan; MACK, Benjamin y WASKE, Björn. Optimization of object-based image analysis with random forests for land cover mapping. *En*: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing [en línea]. 2013. Vol. 6, No. 6; p. 2492-2504.

²²⁷ LÖW; KNÖFEL y CONRAD. Op. cit.

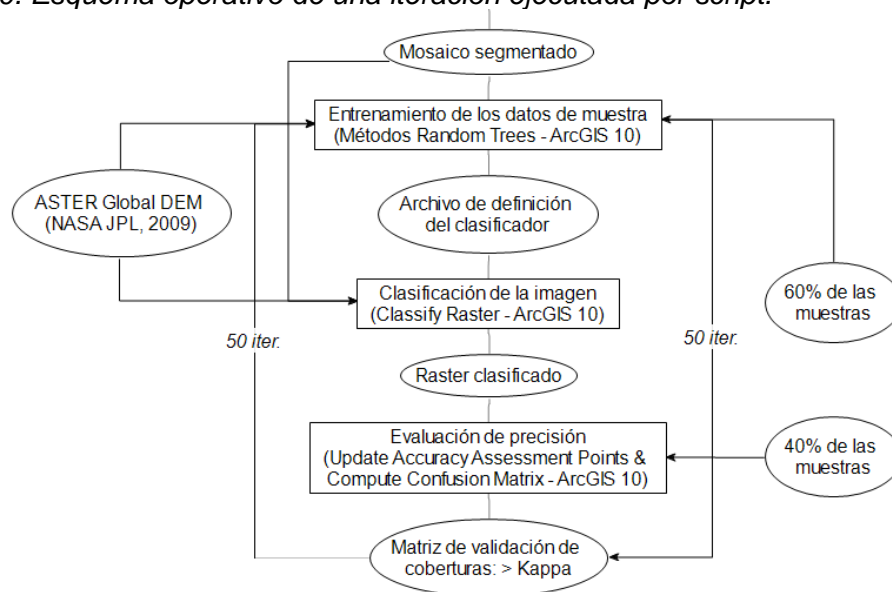
²²⁸ PYTHON SOFTWARE FOUNDATION. ForLoop [en línea]. _____, 2017. [Consultado el 10 de marzo de 2017]. *En*: <https://wiki.python.org/moin/ForLoop>

En conclusión, a partir de siete cuadrantes o imágenes, en la fase de calibración se ajustó el modelo de clasificación de coberturas. Sin esta fase de ejecución de pruebas no hubiese sido posible correr el modelo sin errores para el total del área de estudio.

V. Clasificación de coberturas de la tierra del área de estudio

Mientras en la fase de calibración se procesó una porción de área, en esta fase se clasificó el área de estudio completa. Esto quiere decir que se ejecutaron todas las fases de calibración con los *scripts* sobre el mosaico de imágenes *RapidEye*. El diagrama operativo de los procesos ejecutados con los *scripts* se presenta en la Figura 40. Mientras en la fase de calibración el modelo se iteró 20 veces, en esta etapa de clasificación se ejecutaron **50 iteraciones** utilizando diferentes conjuntos de datos para el entrenamiento y la validación. Löw, Knöfel, y Conrad²²⁹ ejecutaron 100 iteraciones en su modelo de clasificación, pero ellos no calibraron el modelo en una fase previa. Esto quiere decir que ejecutar este modelo calibrado de clasificación 50 veces es un número amplio y riguroso para obtener resultados precisos, estables y coherentes con las características de la superficie terrestre del área de estudio.

Figura 40. Esquema operativo de una iteración ejecutada por script.



Fuente: elaboración propia

El tiempo de procesamiento de la clasificación de coberturas sobre el mosaico fue extenso. El computador empelado ejecutó este proceso entre 28 a 30 horas continuas. Es un tiempo de procesamiento extenso considerando la complejidad de las tareas involucradas. Sin embargo, esto refleja que es posible realizar este tipo de tareas en un equipo de configuración media de *hardware* sin inconvenientes.

²²⁹ LÖW; KNÖFEL y CONRAD. Op. cit.

A partir de las reglas de muestreo, se consolidaron un total de **3.567 muestras** en los siete cuadrantes procesados —promedio de 510,8 muestras por cuadrante—. Es decir que la clasificación del mosaico se realizó con **2.115 muestras (60%)**, la validación tuvo como base **1.427 muestras (40%)**. Este gran conjunto de muestras corresponde a las 16 clases especificadas en el Cuadro 25, izquierda. Producto de la calibración, se identificó una serie de coberturas de la tierra difíciles de clasificar de forma automática porque presentan características espaciales y espectrales similares a otras. Por ejemplo, el modelo se confundió en clasificar los distintos tipos de bosques o el tejido urbano discontinuo, por tal motivo, las coberturas del Cuadro 25, en las columnas del centro y derecha, se capturaron de forma manual en la etapa final de la clasificación: el pos-procesamiento.

Cuadro 25. Coberturas clasificadas de forma automática (izquierda), las capturadas por el usuario de forma manual (centro), y las identificadas de forma semi-automática (derecha).

#	Clase de cobertura de la tierra y su nomenclatura					
	Clasificadas		Digitalizadas		Etiquetadas	
1	Afloramientos rocosos	AFR	Instalaciones recreativas	IRE	Aeropuertos	AER
2	Bosque denso	BDE	Red vial, ferroviaria y terrenos asociados	RVF	Áreas húmedas continentales (humedales)	AHC
3	Bosque fragmentado	BFR	Ríos	RIO	Bosque ribereño	BRI
4	Caña de azúcar - cortada	CAZ	Zonas de disposición de residuos	ZDR	Mosaico de cultivos, pastos y espacios naturales	MCPN
5	Caña de azúcar - sembrada		Zonas de extracción minera	ZEM	Mosaico de pastos con espacios naturales	MPN
6	Cuerpos de agua artificiales	CAA	Zonas industriales o comerciales	ZIC	Tejido urbano discontinuo	TUD
7	Cultivos transitorios	CTR	Zonas verdes urbanas	ZVU		
8	Herbazal (páramos principalmente)	HER				
9	Lagunas lagos y ciénagas naturales	LLC				
10	Otros cultivos permanentes	OCP				
11	Pastos arbolados	PAR				
12	Pastos limpios	PLI				
13	Plantación forestal	PFO				
14	Tejido urbano continuo	TUC				
15	Tierras desnudas y degradadas	TDD				
16	Vegetación secundaria o en transición	VST				

Fuente: elaboración propia.

Con estas muestras y la configuración de dos *scripts*, el modelo de clasificación se iteró 50 veces empleando conjuntos de datos aleatorios. La matriz de confusión del Cuadro 26 enseña que el índice Kappa de precisión obtenido en la iteración 16 de 50 fue del 84%. Este resultado está entre el 84% y 89% de precisión alcanzado por otros investigadores²³⁰. Sin embargo, se alcanzó ese porcentaje clasificando 16 tipos diferentes de coberturas; un gran logro considerando que los otros ejercicios no superan la mitad de ese número de coberturas para un área más pequeña y homogénea.

²³⁰ BLASCHKE. Op. cit. & SOUSA. Op. cit. & MAKINDE. Op. cit.

Analizando la matriz de confusión detalladamente, es evidente que hubo unas coberturas de la tierra mejor clasificadas que otras. En el Cuadro 26, mientras la clasificación de las coberturas de LLC, AFR y TUC presentaron el mejor rendimiento del clasificador —94% y 100%—, la clasificación de las VST, PLI y PAR tuvieron el rendimiento más bajo del modelo —69% y 79%—. Esto se puede entender por la extensión menor, distribución concentrada y menor número de muestras del primer grupo sobre el área mayor, distribución heterogénea y superior cantidad de muestras del segundo grupo en el área de estudio.

La clasificación automática tuvo un rendimiento muy alto, sin embargo, fue clara la presencia del efecto de «sal y pimienta» en el producto²³¹. Si se observa el centro del área de estudio en la Figura 35 (c, página 155), se evidencia un efecto de atomización de las áreas destinadas al cultivo de caña de azúcar. Las fracciones corresponden a segmentos que, por la maduración diferenciada del cultivo y la cantidad de agua por riego, produjeron una firma espectral particular. También, al norte del área se clasificaron de forma equivocada segmentos como tejido urbano continuo. Esta área tuvo una particularidad: el proveedor de la imagen reemplazó el cuadrante original con otro capturado en otra fecha debido a la nubosidad. En los tres mapas de la Figura 35 se observa esa irregularidad. Este tipo de situaciones se solucionaron con una revisión manual en la etapa de pos-procesamiento.

²³¹ CLARK y PELLIKKA. Op. cit.

Cuadro 26. Matriz de confusión de la clasificación del mosaico. Iteración número 16 de 50.

Clase	AFR	BDE	BFR	CAZ	CAZ	CAA	CTR	HER	LLC	OCP	PAR	PLI	PFO	TUC	TDD	VST	Total	U_Acc.	Kappa
AFR	46	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	49	0,94	0,00
BDE	0	66	6	0	0	0	1	0	0	0	0	2	0	0	0	0	75	0,88	0,00
BFR	0	3	140	0	1	0	0	4	0	1	1	3	2	0	0	8	163	0,86	0,00
CAZ	0	0	0	56	0	0	0	0	0	0	0	1	0	2	3	0	62	0,90	0,00
CAZ	0	0	0	0	75	0	1	0	0	1	5	1	2	0	0	8	93	0,81	0,00
CAA	0	0	1	0	0	60	0	0	0	0	0	0	0	0	0	0	61	0,98	0,00
CTR	0	0	2	0	1	1	60	0	0	0	1	7	0	0	0	3	75	0,80	0,00
HER	2	3	2	0	0	0	0	65	0	0	0	0	0	0	0	0	72	0,90	0,00
LLC	0	0	0	0	0	0	0	0	24	0	0	1	0	0	0	0	25	0,96	0,00
OCP	0	0	3	0	0	0	1	0	0	58	0	5	1	0	0	7	75	0,77	0,00
PAR	0	0	0	4	1	0	1	0	0	0	74	10	0	0	1	3	94	0,79	0,00
PLI	0	0	2	6	2	0	5	0	0	5	10	128	1	2	0	6	167	0,77	0,00
PFO	0	0	4	0	0	0	1	0	0	0	0	3	90	0	0	1	99	0,91	0,00
TUC	0	0	0	3	0	0	0	0	0	0	0	0	0	114	3	1	121	0,94	0,00
TDD	0	0	0	3	0	0	0	0	0	0	1	3	0	2	98	0	107	0,92	0,00
VST	0	0	6	0	8	0	2	0	0	7	4	4	0	0	0	83	114	0,73	0,00
Total	48	72	166	72	88	61	72	72	24	72	96	168	96	120	105	120	1452	0,00	0,00
P_Acc.	0,96	0,92	0,84	0,78	0,85	0,98	0,83	0,90	1,00	0,81	0,77	0,76	0,94	0,95	0,93	0,69	0,00	0,85	0,00
Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00	0,84

Fuente: elaborado en ArcGIS 10.

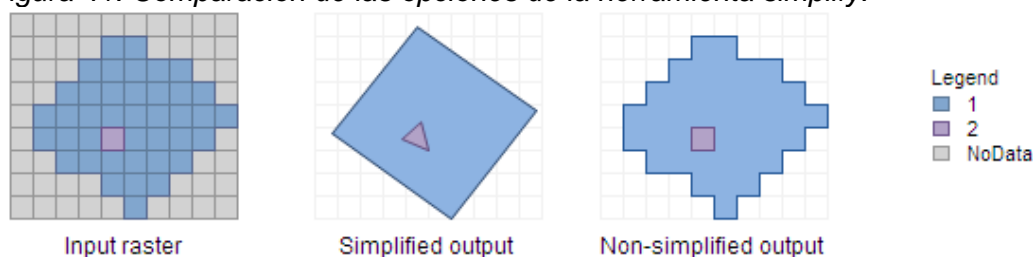
VI. Pos-procesamiento de coberturas de la tierra

La calibración que se hizo en los siete cuadrantes, aunque prometedora por los índices de *kappa* tan altos que se obtuvieron, demostró no ajustarse exactamente al aplicarse a toda el área de estudio dado que los cuadrantes se capturaron en años diferentes y en meses diferentes. De ahí que la etapa de etiquetado tuviera que ser profunda. Eso tuvo que ver, entre otras cosas, con la variación de las etapas de crecimiento de la vegetación observables en cada cuadrante, ya que, al ser imágenes tomadas de diferentes meses del año, en algunas cuadrantes la vegetación se observaba seca y, de repente, al pasar al siguiente cuadrante, mojada. Ese caso fue recurrente en toda la extensión del área de estudio.

En pocas palabras, el pos-procesamiento es la etapa de control de calidad del producto de coberturas de la tierra²³². Esto implica la intervención directa del usuario para ajustar el producto clasificado en tres fases diferentes: i) digitalización, ii) etiquetado, y iii) corrección de topología. No obstante, la imagen clasificada debe convertirse a polígono para ejecutar cada una de estas fases. Convertir datos de formato ráster a polígono es una conversión bastante común en el procesamiento de datos geográficos.

Técnicamente, lo que se hace es agrupar entidades de área como bloques de celdas contiguas con valores iguales en un solo polígono²³³. La herramienta de ArcGIS 10 empleada para la transformación se localiza en **Conversion > From Raster > Raster to Polygon**. Es una herramienta bastante sencilla y el único parámetro ajustado corresponde a la activación de la opción **Raster to Polygon > Simplify**. Como muestra la Figura 41, lo que se pretende es suavizar el límite de cada polígono de salida porque muy difícilmente todas las coberturas de la tierra clasificadas tengan límites en formas de celdas con ángulos rectos de 90 grados.

Figura 41. Comparación de las opciones de la herramienta simplify.



Fuente: ESRI, Raster to Polygon. Op. cit.

Luego de la conversión, el agrupamiento de celdas con la opción de simplificación de borde produjo polígonos de área menor de una (1) hectárea. Con la intención de mantener la UMC de una (1) hectárea correspondiente a la escala 1:25.000, la solución fue combinar estos polígonos pequeños a sus

²³² DUARTE, et al. Op. cit.

²³³ ESRI. Raster to Polygon [en línea]. _____, 2017. [Consultado el 22 de marzo de 2017]. En: <http://desktop.arcgis.com/es/arcmap/latest/tools/conversion-toolbox/raster-to-polygon.htm>

colindantes. La herramienta de ArcGIS ubicada en **Coverage > Data Management > Generalization > Eliminate** permitió realizar este ajuste con precisión. El método de eliminación de polígonos empleado corresponde a la opción **Eliminate > Polygon Option > Border**, es decir que el polígono pequeño se combinó con el polígono contiguo con que comparte el borde más largo. Esta es la opción recomendada y produjo buenos resultados²³⁴.

Con las coberturas de la tierra clasificadas en formato polígono se ejecutaron tres fases de control de calidad del producto. La **digitalización** de coberturas se realizó para ingresar de forma manual aquellas que no se clasificaban correctamente de forma automática. Por más que se ajusten las muestras y el modelo de clasificación, siempre habrá un solapamiento de la reflectancia entre algunas coberturas —tierra desnuda, áreas urbanas discontinuas, zonas mineras, bosque ribereño, entre otras— por lo que deben ser capturadas de forma interactiva por el usuario²³⁵.

Por otra parte, una vez se produjo de forma automática la capa de coberturas de la tierra con el *script*, se realizó una edición manual para corregir los errores identificables contrastando la capa clasificada con la imagen *RapidEye* del mosaico. Es un proceso similar a la revisión ejecutada por Clark y Pellikka²³⁶, aquí los hemos denominado **etiquetado**. Este proceso fue realizado para el total del mosaico y se generó una grilla para orientar el avance de las áreas revisadas.

La etapa de pos-procesamiento se completó con la combinación de los datos digitalizados y los etiquetados. Este proceso de «fusión» entre entidades de distintas fuentes requirió tener una base (capa de coberturas), unos objetos de menor área (elementos digitalizados), una interacción entre ellos dos, y una revisión geométrica de la capa final de coberturas. La relación entre los objetos y la base es de eliminación y unión, es decir, el área de cada objeto se borró de la base para luego unirse exactamente en la misma localización. Esto se realizó empleando en conjunto las herramientas **Analysis > Overlay > Erase** y **Data Management > General > Merge**. *Erase* crea una capa de salida superponiendo entidades —la base y los objetos en este caso— y conserva el área de la base que no estuvo superpuesta²³⁷. La herramienta *Merge* permite combinar entidades para crear una nueva capa de salida²³⁸. En este caso, la base contaba con vacíos que coincidían con el área de los objetos y lo que permitió *Merge* es “llenar” ese vacío con el objeto mismo. Para validar estos procesos fue necesario revisar la geometría de la capa de coberturas con reglas topológicas.

²³⁴ ESRI. Eliminate [en línea]. _____, 2017. [Consultado el 22 de marzo de 2017]. En: <http://desktop.arcgis.com/es/arcmap/latest/tools/coverage-toolbox/eliminate.htm>

²³⁵ CLARK y PELLIKKA. Op. cit.

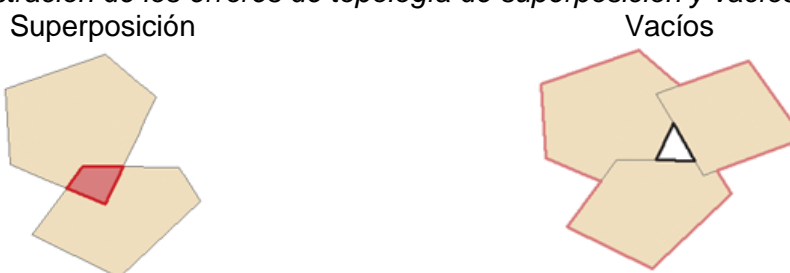
²³⁶ Ibíd.

²³⁷ ESRI. Erase [en línea]. _____, 2017. [Consultado el 5 de junio de 2017]. En: <http://desktop.arcgis.com/es/arcmap/latest/tools/analysis-toolbox/erase.htm>

²³⁸ ESRI. Merge [en línea]. _____, 2017. [Consultado el 8 de junio de 2017]. En: <http://desktop.arcgis.com/en/arcmap/latest/tools/data-management-toolbox/merge.htm>

Las **reglas topológicas** definen las relaciones espaciales permitidas entre las entidades²³⁹. El usuario es quien decide este tipo de relaciones con base a sus intereses. En este caso, las reglas topológicas se emplearon para validar la geometría de la capa de coberturas de la tierra en la fase de pos-procesamiento. El objetivo fue crear una capa de coberturas continua, sin superposiciones, y con bordes perfectamente delimitados. Por tal motivo, las reglas topológicas aplicadas fueron **Polygon Rules > Must Not Overlap** —para revisar que no existan bordes superpuestos entre entidades— y **Polygon Rules > Must Not Have Gaps** —para eliminar áreas de vacío entre coberturas—. La Figura 42 presenta las situaciones que estas dos reglas permitieron identificar.

Figura 42. Ilustración de los errores de topología de superposición y vacíos.



Fuente: ESRI, Topology. Op. cit.

La regla de *Must Not Overlap* permitió identificar las áreas de superpuestas (color rojo de la Figura 42, izquierda) y la regla de *Must Not Have Gaps* permitió reconocer los vacíos entre polígonos adyacentes (color negro de la Figura 42, derecha), pero la solución de estas situaciones se realizó de forma semi-automática con el criterio del usuario. Así las cosas, la solución para *Must Not Overlap* constó de un *Merge* entre las entidades superpuestas asignando la porción de área cubierta a la clase de cobertura correcta. Para *Must Not Have Gaps* se aplicó una solución conjunta con *Create Features* y *Merge* donde se creó un objeto nuevo llenando el vacío y se unió dicho objeto con la clase adyacente de cobertura correcta. Por último, se recuerda que cada uno de estos ajustes topológicos tuvo en cuenta la UMC de una (1) hectárea.

Finalmente, con estas correcciones se tuvo una capa de coberturas de la tierra geoméricamente correcta y geográficamente contextualizada a la realidad del área de estudio. Con este insumo tan valioso se procedió a identificar la EEM del área de influencia metropolitana de Cali. Ese es el componente abarcado en la siguiente sección de la metodología.




El modelo de clasificación automática de las coberturas realizó gran parte del trabajo, no obstante, se identificaron falencias como el efecto de “sal y pimienta” o la asignación errada de clases, que se solucionaron con la intervención directa de los investigadores. Se digitalizaron siete (7) coberturas de la leyenda a una escala de 1:5.000 (Cuadro 25, centro, página 169). Digitalizar con dicha escala garantizó una identificación ideal de cada una de esas coberturas. Para algunas

²³⁹ ESRI. Topology in ArcGIS [en línea]. _____, 2017. [Consultado el 25 de julio de 2017]. En: <http://desktop.arcgis.com/es/arcmap/latest/manage-data/topologies/topology-in-arcgis.htm>

coberturas se usó como guía la información espacial publicada de forma gratuita en la web por la CVC, CRC, IGAC, Google Earth, IDEAM y *Open Street Maps*. Estos datos agilizaron la captura digital de dichas coberturas para el área de estudio.




El etiquetado es la segunda fase del pos-procesamiento. Esta actividad involucró la identificación de seis (6) coberturas de la tierra (Cuadro 25, derecha, página 169), y al mismo tiempo, una revisión de cada cuadrante procesado para verificar la fidelidad de la capa de coberturas con la realidad del terreno. El etiquetado no fue un proceso complejo, pero requirió rigurosidad y tiempo considerando la extensión del área de estudio.




Anexo B. Descripción de las coberturas de la tierra para el área de estudio




Nivel 1	Nivel 2	Nivel 3	Nivel 4	Descripción	Contiene	Fotografía
1. Áreas artificializadas	1.1. Zonas urbanizadas	1.1.1. Tejido urbano continuo		Son espacios conformados por edificaciones y los espacios adyacentes a la infraestructura edificada. Las edificaciones, vías y superficies cubiertas artificialmente cubren más de 80% de la superficie del terreno	<ul style="list-style-type: none"> • Centro de aglomeraciones y centros históricos • Zonas de habitación periféricas • Parqueaderos y áreas cubiertas por asfalto o cemento • Estructuras y edificaciones artificializadas con área mayor a una (1) hectárea 	
		Simbología RGB 204 0 0				Panorámica urbana de Cali. Fuente: El País
		1.1.2. Tejido urbano discontinuo		Son espacios conformados por edificaciones y zonas verdes. Las edificaciones, vías e infraestructura construida cubren la superficie del terreno de manera dispersa y discontinua, ya que el resto del área está cubierta por vegetación	<ul style="list-style-type: none"> • Casas individuales, con jardín y espacios verdes • Manzanas menos densas con espacios verdes al interior • Parques y áreas verdes urbanas con área menor a una (1) hectárea 	
		Simbología RGB 248 0 0				Sur de Santiago de Cali. Fuente: Cali.gov
	1.2. Zonas industriales o comerciales y redes de comunicación	1.2.1. Zonas industriales o comerciales		Son las áreas cubiertas por infraestructura artificial (terrenos cimentados, alquitranados, asfaltados o estabilizados), sin presencia de áreas verdes dominantes, las cuales se utilizan también para actividades comerciales o industriales	<ul style="list-style-type: none"> • Áreas de producción industrial • Centros de investigación y desarrollo • Centros comerciales y de exposiciones con su infraestructura • Equipamientos con área mayor a una (1) hectárea de: seguridad pública y penal; centros hospitalarios y psiquiátricos; universidades, escuelas; infraestructura de telecomunicaciones; bases militares; plantas de tratamiento de aguas; subestaciones eléctricas 	
		Simbología RGB 255 204 0				Zona industrial ACOPI. Fuente: inelco.com




		1.2.2. Red vial, ferroviaria y terrenos asociados		Son espacios artificializados con infraestructuras de comunicaciones como carreteras, autopistas y vías férreas; se incluye la infraestructura conexa y las instalaciones asociadas tales como: estaciones de servicios, andenes, terraplenes y áreas verdes	<ul style="list-style-type: none"> • Infraestructura vial como carreteras, autopistas, puentes y parqueaderos • Infraestructura férrea como vías, intercambiadores, estaciones de abordaje, sitios de mantenimiento y zonas verdes conexas a las estaciones 	
		Simbología RGB 194 194 194				Ferrocarril del Pacífico. Fuente: El País
		1.2.4. Aeropuertos		Comprende la infraestructura donde funciona una terminal aérea. Incluye las pistas de aterrizaje y carreteo, los edificios, las superficies libres, las zonas de amortiguación y la vegetación asociada	<ul style="list-style-type: none"> • Aeropuertos militares, comerciales y deportivos • Terminales de pasajeros, talleres, bodegas, parqueaderos • Pistas de aterrizaje y carreteo • Superficies cubiertas de vegetación y los espacios asociados 	
		Simbología RGB 150 140 80				Aeropuerto Bonilla Aragón. Fuente: El País
	1.3. Zonas de extracción minera y escombreras	1.3.1. Zonas de extracción minera		Comprende las áreas donde se extraen o acumulan materiales asociados con actividades mineras, de construcción, y producción de residuos de diferente origen	<ul style="list-style-type: none"> • Arenales; canteras; gravilleras • Infraestructuras industriales asociadas (fábricas de cemento) • Superficies de agua con área menor a una (1) hectárea producto de la extracción • Los sitios de extracción a cielo abierto en actividad o abandonados sin vegetación 	
		Simbología RGB 211 23 255				Fábrica San Marcos. Fuente: El País


		1.3.2. Zonas de disposición de residuos		Son espacios en los que se depositan residuos urbanos, restos de construcción, desechos industriales y material estéril de minas	<ul style="list-style-type: none">• Vertederos públicos o de municipalidades• Vertederos industriales de residuos líquidos y sólidos• Edificios, parqueaderos y carreteras asociadas con el vertedero, con área menor a una (1) hectárea	 <p>Relleno sanitario Colomba-El Guabal, Yotoco. Fuente: El País</p>
		Simbología RGB 137 137 137				
1.4. Zonas verdes artificializadas , no agrícolas		1.4.1. Zonas verdes urbanas		Comprende las áreas cubiertas por vegetación dentro del tejido urbano con tamaño superior a 1 hectárea, incluyendo parques urbanos y cementerios	<ul style="list-style-type: none">• Cementerios con zonas verdes<ul style="list-style-type: none">• Jardines ornamentales• Parques, estanques de parques, áreas cubiertas por césped• Parques botánicos y zoológicos incluidos o al lado del “tejido urbano”• Espacios arbolados ubicados entre los edificios del “tejido urbano”	 <p>Parque La Flora, Cali. Fuente: Google Earth</p>
		Simbología RGB 180 215 158				
		1.4.2. Instalaciones recreativas		Son los terrenos dedicados a las actividades de camping, deporte, parques de atracción, golf, hipódromos y otras actividades de recreación y ocio, incluyendo los parques habilitados para esparcimiento, no incluidos dentro del “tejido urbano”	<ul style="list-style-type: none">• Áreas de carreras automovilísticas• Zonas arqueológicas declaradas• Parques botánicos y zoológicos no incluidos en “tejido urbano”<ul style="list-style-type: none">• Campos de fútbol y la infraestructura asociada• Fincas y condominios recreativos ubicados en las afueras de las ciudades o a lo largo de las vías• Zonas de pastos en áreas de entrenamiento militar	 <p>Club Campestre de Cali. Fuente: El País</p>
		Simbología RGB 255 50 50				




2. Áreas de producción agrícola	2.1. Cultivos transitorios	2.1.1. Otros cultivos transitorios		Comprende las áreas ocupadas con cultivos cuyo ciclo vegetativo es menor a un año. También suelos en preparación para próximos cultivos transitorios	<p>Cultivos de:</p> <ul style="list-style-type: none"> • Cereales: arroz, maíz, sorgo, cebada, trigo • Oleaginosas y leguminosas: algodón, ajonjolí, frijol, soya, maní • Hortalizas: cebolla, zanahoria, remolacha • Tubérculos: papa, yuca, flores a cielo abierto 	 <p>Frijol Cargamanto. Fuente: FENALCE</p>
		Simbología RGB 255 255 166				
	2.2. Cultivos permanentes	2.2.1. Otros cultivos permanentes		Comprende las áreas dedicadas a cultivos cuyo ciclo vegetativo es mayor a un año, produciendo varias cosechas sin necesidad de volverse a plantar. También suelos en preparación para próximos cultivos permanentes	<p>Cultivos de:</p> <ul style="list-style-type: none"> • Herbáceas: caña panelera, plátano y banano • Arbustivos: café y cacao • Arbóreos: palma africana y árboles frutales 	 <p>Cultivo de café. Fuente: El País</p>
			2.2.1.2. Caña de azúcar	Cobertura compuesta por el cultivo de caña de azúcar y su infraestructura asociada. Es un cultivo industrial para la producción de azúcar, que se corta cada 12 meses, y una duración de la plantación de aproximadamente cinco años	<ul style="list-style-type: none"> • Cultivos de caña de azúcar con área mayor a una (1) hectárea • Suelos en preparación y de corte reciente para próximos cultivos de caña de azúcar • Infraestructura asociada con el cultivo de caña de azúcar con área menor a una (1) hectárea (vías, campamentos, bodegas) 	 <p>Caña de azúcar. Fuente: ASOCAÑA</p>
		Simbología RGB 231 160 89	Simbología RGB 233 255 122			


	2.3. Pastos	2.3.1. Pastos limpios		Esta cobertura comprende las tierras ocupadas por pastos limpios con un porcentaje de cubrimiento mayor a 70% de la unidad; la realización de prácticas de manejo (limpieza, encalamiento y/o fertilización) y el nivel tecnológico utilizados, impiden la presencia o el desarrollo de otras coberturas.	<ul style="list-style-type: none"> • Pastos limpios con área mayor a una (1) hectárea • Zonas de pastos limpios sujetas a inundaciones temporales con área menor de 1 hectárea • Pastos con presencia esporádica de arbustales y cultivos • Infraestructuras asociadas con los pastos manejados (viviendas rurales, cercas vivas) 		Pastos limpios. Fuente: El País
		Simbología RGB 204 255 153					
		2.3.2. Pastos arbolados		Cobertura que incluye las tierras cubiertas con pastos, en las cuales se han estructurado potreros con presencia de árboles de altura superior a cinco metros, distribuidos en forma dispersa. En Colombia, se ubican en general sobre áreas planas ganaderas de climas cálidos.	<ul style="list-style-type: none"> • Pastos arbolados con área mayor a una (1) hectárea • Pastos arbolados con zonas inundables con área mayor a una (1) hectárea • Infraestructuras asociadas con los pastos arbolados con área menor a una (1) hectárea (corrales o establos) 		Pastos arbolados: Fuente: Banco de la República de Colombia
		Simbología RGB 51 153 102					
	2.4. Áreas agrícolas heterogéneas	2.4.3. Mosaico de cultivos, pastos y espacios naturales		Comprende las superficies ocupadas por coberturas de cultivos y pastos en combinación con espacios naturales. En esta unidad, el patrón de distribución de las coberturas no puede ser representado individualmente, como parcelas con tamaño mayor a 6 hectáreas. Las áreas de cultivos y pastos ocupan entre 30% y 70% de la superficie total de la unidad	<ul style="list-style-type: none"> • Mezcla de parcelas de pastos y cultivos con relictos de bosques, arbustales y vegetación secundaria, con área mayor a seis (6) hectáreas • Infraestructuras asociadas con los cultivos y los pastos manejados (viviendas rurales, vías, corrales) 		Mosaico de cultivos, pastos y áreas naturales. Fuente: IDEAM
		Simbología RGB 214 153 0					

		2.4.4. Mosaico de pastos con espacios naturales		Constituida por las superficies ocupadas por coberturas de pastos en combinación con espacios naturales. En esta unidad, las parcelas de pastos presentan un área menor a 6 hectáreas. Las coberturas de pastos representan entre 30% y 70% de la superficie total del mosaico.	<ul style="list-style-type: none"> • Mezcla de pastos con relictos de bosques, arbustales y vegetación secundaria, con área mayor a seis (6) hectáreas • Infraestructuras asociadas con los pastos manejados con área menor a una (1) hectárea (viviendas rurales, vías, corrales) 	 <p>Mosaico de pastos y zonas naturales. Fuente: IDEAM</p>
		Simbología RGB 191 158 65				
3. Bosques y áreas seminaturales	3.1. Bosques	3.1.1. Bosque denso		Cobertura constituida por una comunidad vegetal dominada por elementos típicamente arbóreos, los cuales forman un estrato de copas (dosel) continuo. Estas formaciones vegetales no han sido intervenidas o su intervención ha sido selectiva sin alterar su estructura original.	<ul style="list-style-type: none"> • Áreas con una cobertura densa de palmas naturales con área superior a una (1) hectárea. • Áreas con una cobertura densa de guadua y extensión mayor a una (1) hectárea. • Formaciones arbóreas secundarias regeneradas de manera natural 	 <p>Bosque denso en Farallones de Cali. Fuente: cali.gov</p>
		Simbología RGB 86 172 0				
		3.1.3. Bosque fragmentado		Comprende los territorios cubiertos por bosques naturales cuya continuidad horizontal está afectada por la inclusión de otros tipos de coberturas como pasto, cultivos o vegetación en transición. Los bosques fragmentados tienden a presentarse en la periferia urbana y zonas de colonización	<ul style="list-style-type: none"> • Parches de bosque natural generalmente de forma geométrica con área mayor a una (1) hectárea 	 <p>Bosque fragmentado. Fuente: Banco de la República de Colombia</p>
		Simbología RGB 170 255 0				

		3.1.4. Bosque ripario (ribereño)		Se refiere a las coberturas constituidas por vegetación arbórea ubicada en las márgenes de cursos de agua permanentes o temporales. Mientras en las regiones de sabanas se conoce como bosque de galería o cañadas, las franjas de bosque en cursos de agua son conocidas como bosque ribereño en las zonas andinas	<ul style="list-style-type: none"> • Bosque ribereño con área superior a una (1) hectárea • Coberturas de asociaciones de palma y guadua a lo largo de los márgenes de los ríos, con altura y densidad de bosque natural 	
		Simbología RGB 0 168 132				Bosque ribereño. Fuente: El País
		3.1.5. Plantación forestal		Son coberturas constituidas por plantaciones de vegetación arbórea, realizada por la intervención directa del hombre con fines de manejo forestal. Esta cobertura presenta un patrón geométrico regular, constituido por filas y columnas de árboles generalmente de la misma edad.	<ul style="list-style-type: none"> • Plantaciones de coníferas o latifoliadas con área mayor a una (1) hectárea. • Zonas quemadas dentro de la plantación con área mayor a una (1) hectárea. • Parcelas de plantaciones en proceso de aprovechamiento (zonas en tala). • Infraestructura asociada con área menor a una (1) hectárea (vías, campamentos, aserraderos) 	
		Simbología RGB 137 168 101				Plantación forestal en La Cumbre, Valle del Cauca. Fuente: Google Earth
	3.2. Áreas con vegetación herbácea y/o arbustiva	3.2.1. Herbazal (páramos)		Cobertura constituida por una comunidad vegetal dominada por elementos típicamente herbáceos desarrollados en forma natural en diferentes sustratos, los cuales forman una cobertura densa. Estas formaciones vegetales no han sido intervenidas o su intervención ha sido selectiva y no ha alterado su estructura original ni sus características funcionales	<ul style="list-style-type: none"> • Coberturas naturales de herbáceas de páramo y subpáramo • Coberturas de herbazales quemados • Coberturas de arbustales 	
		Simbología RGB 204 242 77				Páramo en PNN Farallones. Fuente: CVC




		3.2.3. Vegetación secundaria o en transición		Comprende aquella cobertura vegetal en proceso de sucesión natural producto de la intervención o por la destrucción de la vegetación primaria. Se desarrolla en zonas desmontadas para diferentes usos o en áreas agrícolas abandonadas.	<ul style="list-style-type: none"> • Áreas de arbustos con extensión mayor a una (1) hectárea. • Pequeños fragmentos de pastos y arbustos en sucesión con área mayor a una (1) hectárea. 	 <p>Vegetación en transición. Fuente: El Tiempo</p>
		Simbología RGB 153 204 0				
	3.3. Áreas abiertas, sin o con poca vegetación	3.3.2. Afloramientos rocosos		Áreas en las que la superficie del terreno está constituida por capas de rocas expuestas y sin/exiguo desarrollo de vegetación. En la región andina, se localizan principalmente en las áreas de fuerte pendiente, donde predominan los sustratos de rocas duras y resistentes. Los afloramientos rocosos son producto de la dinámica natural del terreno	<ul style="list-style-type: none"> • Áreas de afloramientos rocosos con área superior a una (1) hectárea • Áreas rocosas con cobertura vegetal escasa que representa menos de 30% de la superficie 	 <p>Afloramiento rocoso en PNN Farallones. Fuente: Mauricio Tabares Mosquera</p>
		Simbología RGB 194 194 194				
		3.3.3. Tierras desnudas y degradadas		Esta cobertura corresponde a las superficies de terreno desprovistas de vegetación o con escasa cobertura vegetal, debido a la ocurrencia de procesos tanto naturales como antrópicos de erosión y degradación extrema	<ul style="list-style-type: none"> • Zonas de tierras desnudas y degradadas con área superior a una (1) hectárea. • Zonas semidesérticas, con vegetación seca y ocurrencia de procesos de erosión. • Áreas de rocas, cantos rodados o cascajo en laderas empinadas en las que se presenta una cobertura vegetal escasa que representa menos de 30% de la superficie 	 <p>Cárcavas. Fuente: CVC</p>
		Simbología RGB 236 236 236				




4 y 5. Áreas húmedas y superficies de agua	4.1. Áreas húmedas continentales	4.1.1 Áreas húmedas continentales (humedales)		En las planicies aluviales se forman cuerpos de agua denominados humedales que están asociadas con las áreas de desborde de los grandes ríos con el nivel freático a nivel del suelo en forma temporal o permanente	<ul style="list-style-type: none"> • Cuerpos de agua ubicados en las llanuras de inundación de los ríos con área mayor a una (1) hectárea • Cuerpos de agua con presencia de vegetación acuática con área mayor a una (1) hectárea 	
		Simbología RGB 51 204 204				Humedal del Valle del Cauca. Fuente: CVC
	5.1. Aguas continentales	5.1.1. Ríos		Corriente natural de agua que fluye con continuidad, posee un caudal considerable y desemboca en el mar, en un lago o en otro río. Esta cobertura comprende los cauces principales de las cuencas hidrográficas que componen el área de estudio	<ul style="list-style-type: none"> • Cursos de agua naturales con área de polígono mayor a una (1) hectárea 	
		Simbología RGB 0 204 242				Río Cauca. Fuente: CVC
		5.1.2. Lagunas, lagos y ciénagas naturales		Superficies o depósitos de agua naturales de carácter abierto o cerrado, que pueden estar conectadas con un río o con el mar. En la zona andina hay cuerpos de agua (lagos y lagunas) situados en alta montaña que constituyen las áreas de nacimiento de ríos	<ul style="list-style-type: none"> • Lagunas u otros cuerpos agua naturales con área mayor a una (1) hectárea 	
		Simbología RGB 51 102 255				Laguna de páramo en PNN Farallones. Fuente: Mauricio Tabares Mosquera




		<p>5.1.4. Cuerpos de agua artificiales</p>		<p>Esta cobertura comprende los cuerpos de agua de carácter artificial que fueron creados por el hombre para almacenar agua, usualmente con el propósito de generación de electricidad y el abastecimiento de acueductos, aunque también para prestar otros servicios como: control de caudales e inundaciones, riego y con fines turísticos y recreativos</p>	<ul style="list-style-type: none"> • El cuerpo de agua y las áreas secas expuestas en períodos de vaciado y estiaje • La infraestructura asociada con área menor a una (1) hectárea (presa, generador, transformador) • Las islas presentes en los cuerpos de agua con área menor a una (1) hectárea • Lagunas para irrigación de cultivos 	
		<p>Simbología RGB 0 77 168</p>				<p>Embalse Salvajina. Fuente: El País</p>




Fuente: INSTITUTO DE HIDROLOGÍA, METEOROLOGÍA Y ESTUDIOS AMBIENTALES. Op. cit.




Anexo C. Descripción de los SE valorados en la consulta a expertos

Sección	División	Grupo	Comprende	Ilustración
1. Aprovechamiento: todas las salidas nutricionales, materiales y energéticas de los sistemas vivos	1.1. Nutrición: esta división abarca todas las salidas de los sistemas vivos que son el aprovisionamiento nutricional y alimenticio de la sociedad	1.1.1. Biomasa	<ul style="list-style-type: none"> • Cultivos de cereales, vegetales, frutas • Presencia silvestre de bayas, frutas, setas • Animales encerrados y sus productos <ul style="list-style-type: none"> • Acuicultura • Pesca comercial y de subsistencia <ul style="list-style-type: none"> • Cacería de animales 	 <p>Piña Oro Miel. Fuente: Agencia UN</p>
		1.1.2. Agua	<ul style="list-style-type: none"> • Agua superficial extraída de ríos, lagos y otros cuerpos de agua abiertos o subterráneos para beber. 	 <p>Captación de agua. Fuente: El Tiempo</p>
	1.2. Materiales: esta división comprende todas las salidas de los sistemas vivos que aprovisionan de materias primas a la humanidad	1.2.1. Biomasa	<ul style="list-style-type: none"> • Fibras y otros materiales de plantas, algas y animales para uso directo o procesado (madera, flores, huesos, celulosa, algodón, caucho, cera, resina) • Materiales genéticos de toda la biota para producir bioquímicos y farmacéuticos 	 <p>Extracción de madera. Fuente: CVC</p>




		1.2.2. Agua	<ul style="list-style-type: none"> • Agua superficial sin propósito bebible (uso doméstico, riego, consumo de ganado, uso industrial) 	 <p>Riego por goteo en cultivo de tomate. Fuente: Agencia UNAL</p>
		1.2.3. Materiales abióticos metálicos y no metálicos	<ul style="list-style-type: none"> • Minerales metálicos • Pigmentos • Material de construcción (arcilla/barro) 	 <p>Extracción de oro en Suárez, Cauca. Fuente: Revista Semana</p>
	1.3. Energía: esta división incluye todas las salidas de los sistemas vivos que la humanidad emplea para la producción de energía	1.3.1. Biomasa	<ul style="list-style-type: none"> • Combustible de madera <ul style="list-style-type: none"> • Leña • Paja • Cultivos para quema y producción de energía <ul style="list-style-type: none"> • Excremento • Grasa • Aceites 	 <p>Quema de leña en producción de panela. Fuente: Radio UNAL</p>

		1.3.2. Fuente abiótica renovable	<ul style="list-style-type: none"> • Hidroeléctrica • Viento 	 <p>Embalse La Salvajina, Cauca. Fuente: El País</p>
		1.3.3. Fuente abiótica no renovable	<ul style="list-style-type: none"> • Combustibles fósiles y sus derivados <ul style="list-style-type: none"> • Carbón • Petróleo • Gas 	 <p>Carbón. Fuente: Portafolio.co</p>
2. Regulación y mantenimiento: las formas en que los organismos vivos pueden moderar el entorno ambiental que afecta la actuación humana	2.1. Regulación de residuos, sustancias tóxicas y otras molestias: a través de procesos biogeoquímicos, las coberturas de la tierra naturales y seminaturales están en capacidad de filtrar efluentes, descomponer residuos tóxicos y eliminar agentes patógenos nocivos en el ambiente	2.1.1. Mediación por sistemas vivos	<ul style="list-style-type: none"> • Bio-remediación, filtración, dilución y secuestro biofísicoquímica de contaminantes en los ecosistemas • Adsorción de metales pesados y compuestos orgánicos en los ecosistemas (combinación de factores bióticos y abióticos) • Infraestructura verde para reducir el ruido y los olores 	 <p>Humedal Guarino, Jamundí. Fuente: CVC</p>

	2.2. Regulación de flujos: las coberturas naturales y seminaturales regulan diversos flujos en el ambiente. Por ejemplo, la cubierta vegetal previene la erosión del suelo y regula los flujos de agua, pero también, crea amortiguadores contra los eventos naturales reduciendo los daños que puedan causar las tormentas, avalanchas, etcétera	2.2.1. Flujos de masa	<ul style="list-style-type: none"> • Estabilización y control de la tasa de erosión <ul style="list-style-type: none"> • Cobertura vegetal en pendientes • Amortiguación y atenuación de los flujos de masa • Transporte y almacenamiento de sedimentos por ríos y lagos 	 <p>Vegetación en pendientes. Fuente: Portafolio Agrícola</p>
		2.2.2. Flujos de líquido	<ul style="list-style-type: none"> • Mantenimiento del ciclo hidrológico • Captación efectiva de la precipitación por cobertura vegetal • Protección contra inundaciones por la cobertura vegetal 	 <p>Páramo en PNN Las Hermosas. Fuente: parquesnacionales.gov.co</p>
		2.2.3. Flujos de gases / aire	<ul style="list-style-type: none"> • Protección contra tormentas • Ventilación y transpiración • Vegetación natural o plantada que sirve de cinturón de protección y permite la ventilación del aire 	 <p>Barrera natural arbustiva. Fuente: CVC</p>

	2.3. Mantenimiento de condiciones físicas, químicas y biológicas: mediante procesos biogeoquímicos, las coberturas de la tierra naturales y seminaturales mantienen condiciones aptas para la vida de todas las especies. Es así que se regulan los hábitats de las especies, se fijan nutrientes formadores de suelo, se purifica la composición química del agua y se regula el clima y la composición de la atmósfera	2.3.1. Mantenimiento del ciclo de vida, hábitat y protección de reserva genética	<ul style="list-style-type: none"> • Polinización y dispersión de semillas por insectos, aves y otros animales • Hábitats de reproducción y cría de plantas y animales (pastizales, microestructura de ríos) 	 <p>Río Cauca, hábitat de cría y reproducción de fauna. Fuente: El País</p>
		2.3.2. Control de plagas y enfermedades	<ul style="list-style-type: none"> • Regulación de especies invasoras • Control de enfermedades en cultivos y poblaciones humanas 	 <p>Los bosques son reserva de organismos erradicadores de plagas. Fuente: CVC</p>
		2.3.3. Formación y composición del suelo	<ul style="list-style-type: none"> • Meteorización física, química y biológica • Descomposición / mineralización de material orgánico • Fijación de nutrientes 	 <p>Formación de material orgánica. Fuente: ICA</p>

		2.3.4. Mantenimiento de la composición química del agua	<ul style="list-style-type: none"> • Purificación de agua • Amortiguación de sedimentos <ul style="list-style-type: none"> • Desnitrificación • Remineralización de fósforo 	 <p>Los humedales filtran contaminantes. Humedal Timbique, Palmira. Fuente: CVC</p>
		2.3.5. Composición atmosférica y regulación del clima	<ul style="list-style-type: none"> • Reducción de los gases de efecto invernadero <ul style="list-style-type: none"> • Secuestro de carbono • Regulación del clima local y regional <ul style="list-style-type: none"> • Mantenimiento de la calidad del aire • Regulación de los patrones de temperatura y precipitación 	 <p>Zona verde urbana de Cali. Fuente: Panoramio</p>
3. Cultural: las salidas no materiales, y normalmente no consumibles, de los sistemas vivos que afectan el estado físico y mental de las personas	3.1. Interacción física e intelectual: la interacción con la naturaleza ofrece oportunidades recreativas que desempeñan un papel importante en el mantenimiento de la salud mental y física de los seres humanos. También, los animales y las plantas han sido fuente de inspiración para gran parte del arte, ciencia, cultura y el diseño	3.1.1. Experiencia física	<ul style="list-style-type: none"> • Recreación • Turismo 	 <p>Viaje en parapente. Fuente; El Tiempo</p>

		3.1.2. Intelectual y representativa	<ul style="list-style-type: none"> • Ciencia • Arte • Educación • Entretenimiento 	 <p>Paisaje del Valle del Cauca. Fuente: El País</p>
	3.2. Interacción simbólica y espiritual: la naturaleza es un elemento común en la mayoría de las principales religiones. El patrimonio natural, el sentimiento espiritual de pertenencia, el conocimiento tradicional y las costumbres conexas son importantes para crear un sentido de pertenencia	3.2.1. Espiritual y/o emblemática	<ul style="list-style-type: none"> • Simbólico • Sagrado • Religioso • Fauna y flora emblemática • Lugares sagrados 	 <p>Kwet Wala, casa sagrada de tribu Nasa. Fuente: pradera-valle.gov.co</p>
		3.2.2. Existencia y valor intrínseco natural	<ul style="list-style-type: none"> • Valoración de especies silvestres • Legado generacional natural • Creencia ética y moral por la naturaleza 	 <p>Puma en PNN Las Hermosas. Fuente: parquesnacionales.gov.vo</p>

Fuente: elaborado a partir de HAINES-YOUNG y POTSCHIN. Op. cit.

Anexo D. Estructura del script de muestreo estratificado

```
# -----
# Stratified Random Selection, Merge and Centroids
# By: "Territorios" Research Group
# Created on: 2017-03-28
# -----

# Import arcpy module
from arcpy import env
import os
from arcpy.sa import *
arcpy gp.overwriteOutput = 1

os.system('cls')

# Set of arguments
dirbase = raw_input('Workspace directory: ')
dirout = raw_input('Output directory: ')
inputShapefile = raw_input('Input data: ')

print "-----"
print "  Running script: Stratified Random Selection, Merge and Centroids  "
print "-----"

# Workspace
#outputDirectory = r"H:\X\Prueba\tile_607\escript\Tile607.gdb"
arcpy.env.workspace = dirbase
outputDirectory = arcpy.env.workspace

# Set of variables
inputData = inputShapefile
#inputData = "MS_607_TMS_a"

# PROCESS 1: SELECT CLASSES
for a in range(1, 16): # Number of classes
    print "We're on time %d" % (a)
    outputClass = "MS_607_TMS_class_" + str(a)
    where_clause = "Classvalue = " + str(a)
    # Execute Select
    arcpy.Select_analysis(inputData, outputClass, where_clause)
    print "Done with " + outputClass

# PROCESS 2: SUBSET FEATURES
# Data sets (40% and 60%)
inputClasses = arcpy.ListFeatureClasses("MS_607_TMS_class*")
for b in inputClasses:
    for c in range(1, 21): # Number of groups of 40% and 60% to be created
        output60p = b.split("_")[-2] + "_" + b.split("_")[-1] + "_" + "MS_607_TMS_60p" + "_" + "iteration" + "_" + str(c)
        output40p = b.split("_")[-2] + "_" + b.split("_")[-1] + "_" + "MS_607_TMS_40p" + "_" + "iteration" + "_" + str(c)
        # Execute SubsetFeatures
        arcpy.SubsetFeatures_ga(b, output60p, output40p, "60", "PERCENTAGE_OF_INPUT")
        print "class " + str(b)
        print "Done with SubsetFeatures class " + "iteration " + str(c)

# PROCESS 3: MERGE
# Merge feature classes representing 40% of the records
for d in range(1, 21): # Number of groups of 40% to be merged
    inputFile40 = "MS_607_TMS_40p_iteration_" + str(d)
    listElements40 = arcpy.ListFeatureClasses(inputFile40)
    print "New group", listElements40
    outputFile40 = "MS_607_TMS_40p_all_merge" + "_" + str(d)
    arcpy.Merge_management(listElements40, outputFile40)
    print "Done with merge class " + outputFile40

# Number of groups of 60% to be merged
inputFile60 = "MS_607_TMS_60p_iteration_" + str(d)
listElements60 = arcpy.ListFeatureClasses(inputFile60)
print "New group", listElements60
outputFile60 = "MS_607_TMS_60p_all_merge" + "_" + str(d)
arcpy.Merge_management(listElements60, outputFile60)
print "Done with merge class " + outputFile60

# PROCESS 4: FEATURE TO POINTS (CENTROIDS)
outFeatureClassPts = outputFile40 + "_pts"
# Use FeatureToPoint function to find a point inside polygon
arcpy.FeatureToPoint_management(outputFile40, outFeatureClassPts, "INSIDE")
print "Done finding centroids " + outFeatureClassPts

print "Finished!"
```

Anexo E. Estructura del script de clasificación de coberturas

```
# -----
# Created on: 2017-03-23
# By: "Territorios" Research Group
# Description: it runs Random Trees, Classify Raster, Updated Accuracy Assessment Points and Confusion Matrix
# -----

# Import arcpy module
import arcpy, os, sys, string
from arcpy import env
from arcpy.sa import *
arcpy gp.overwriteOutput = 1

os.system('cls')

# Set of arguments
dirbase = raw_input('Workspace directory: ')
dirout = raw_input('Output directory: ')
inputSegmented = raw_input('Enter segmented file name: ')
inputDEM = raw_input('Enter DEM name: ')

print "-----"
print "    Running script: Images Classification    "
print "-----"

# Set the workspace for ListFeatureClasses
arcpy.env.workspace = dirbase #r"D:\Workspace\Procalculo\MeanShift\Tile607.gdb"

# Local variables:
Segmentation = dirbase + "\\" + inputSegmented # r"D:\Workspace\Procalculo\MeanShift\Tile607.gdb\MS_607_spe15_spa15_531"
DEM_30m = dirbase + "\\" + inputDEM
output = dirout # r"D:\Workspace\Procalculo\MeanShift"

# PROCESS: Train Random Trees Classifier
# List of segments (60 percent of the total)
inputData = arcpy.ListFeatureClasses("MS_607_TMS_60p_all_merge_")
for a in inputData:
    if a[-3:] <> "pts": # This is used in order to avoid listing point features classes
        outputFileECD = output + "\\" + a.split("_")[-7] + "_" + a.split("_")[-6] + "_" + a.split("_")[-1] + ".ecd"
        print a
        # Run: Train Random Trees Classifier
        arcpy gp.TrainRandomTreesClassifier_sa(Segmentation, a, outputFileECD, DEM_30m, "50", "30", "0",
"COLOR;MEAN;STD;COUNT;COMPACTNESS;RECTANGULARITY")
        print "Done with " + a

# PROCESS: Classify Raster
# Output for Classify Raster
outputFileCR = a.split("_")[-7] + "_" + a.split("_")[-6] + "_" + a.split("_")[-1] + "_cr"
# Run: Classify Raster
arcpy gp.ClassifyRaster_sa(Segmentation, outputFileECD, outputFileCR, DEM_30m)
print "Done with " + outputFileCR

# PROCESS: Updated Accuracy Assessment Points
# List of feature classes for Updated Accuracy Assessment Points
updateAAP = arcpy.ListFeatureClasses("*_pts")
for b in updateAAP:
    print b
    # Output for UAAP feature classes
    outputFileUAAP = a.split("_")[-7] + "_" + a.split("_")[-6] + "_" + a.split("_")[-1] + "_uaap"
    #Run: Update Accuracy Assessment Points
    arcpy gp.UpdateAccuracyAssessmentPoints_sa(outputFileCR, b, outputFileUAAP, "CLASSIFIED")
    print "Done with " + outputFileUAAP

# PROCESS: Compute Confusion Matrix
# Output for Compute Confusion Matrix
outputFileCMM = a.split("_")[-7] + "_" + a.split("_")[-6] + "_" + a.split("_")[-1] + "_cmm"
# Run: Compute Confusion Matrix
arcpy gp.ComputeConfusionMatrix_sa(outputFileUAAP, outputFileCMM)
print "Done with " + outputFileCMM

print "Finished!"
```