

MEMORIA TÉCNICA PARA LA VALIDACIÓN DEL MAPA DE COBERTURAS DE
BOSQUE SECO TROPICAL EN COLOMBIA

Escala 1:100.000

Versión 2.0



Programa de Biología de la Conservación y Uso de la Biodiversidad

Instituto de Investigaciones de Recursos Biológicos “Alexander von Humboldt”

2014

©

Memoria técnica para la validación del mapa de coberturas de bosque seco tropical en Colombia. Escala 1:100.000. Versión 2.0

Autores

Alexander Ariza

Paola Isaacs

Roy González-M.

Programa de Biología de la Conservación y Uso de la Biodiversidad
Instituto de Investigaciones de Recursos Biológicos “Alexander von Humboldt”



MinAmbiente

Ministerio de Ambiente
y Desarrollo Sostenible

**PROSPERIDAD
PARA TODOS**

Memoria cítese como:

Ariza, A., Isaacs, P. y González-M., R. 2014. Memoria técnica para la validación del mapa de coberturas de bosque seco tropical en Colombia (escala 1:100.000, 2.0v). Instituto de Investigaciones de Recursos Biológicos “Alexander von Humboldt” – Ministerio de Ambiente y Desarrollo Sostenible. Bogotá, Colombia. 62p.

Mapa cítese como:

Ariza, A., Isaacs, P. y González-M., R. 2014. Mapa de coberturas de bosque seco tropical en Colombia (escala 1:100.000, 2.0v). Instituto de Investigaciones de Recursos Biológicos “Alexander von Humboldt” – Ministerio de Ambiente y Desarrollo. 1 hoja cartográfica.

TABLA DE CONTENIDO

INTRODUCCIÓN.....	4
METODOLOGÍA.....	6
INFORMACIÓN TEMÁTICA DEL MAPA DE BOSQUE SECO TROPICAL A ESCALA 1:100.000(v1.0).....	6
ANÁLISIS DE LA INFORMACIÓN SECUNDARIA PARA LA CONSTRUCCION DE MAPA DE BOSQUE SECO TROPICAL A 1:100.000(v2.0).....	9
INTEGRACIÓN CON INFORMACIÓN SECUNDARIA.....	11
ANÁLISIS Y ESTRUCTURA DE LA GEODATABASE PARA LA VERIFICACIÓN DEL MAPA DE BOSQUE SECO TROPICAL A ESCALA 1:100.000(v2.0).....	14
Esquema del modelo.....	16
Estructura jerárquica de capas.....	17
Ruta y descripción de la información.....	20
Calidad.....	21
Metadato.....	23
OBTENCIÓN Y AJUSTE DE LAS IMÁGENES DE SATÉLITE Y LA INFORMACIÓN TEMÁTICA.....	24
AJUSTE Y CORRECCIÓN DE LA INFORMACIÓN.....	26
MÉTODOS ESTADÍSTICOS Y ANÁLISIS DE DATOS.....	32
VALIDACIÓN ESTADÍSTICA Y CONFIABILIDAD TEMÁTICA.....	32
Fase inicial de diseño.....	33
Fase de campo.....	39
Fase final de procesamiento y análisis de datos.....	43
RESULTADOS.....	56
REFERENCIAS.....	59

INTRODUCCIÓN

Este documento presenta los resultados obtenidos del proceso de elaboración y determinación de la exactitud, precisión y veracidad temática, del mapa de bosque seco tropical para Colombia a escala 1:100.000, versión 2.0. Para ello se trabajó con el mapa de bosques secos tropicales de Colombia versión 1.0 (Corzo y Delgado 2012), imágenes de satélite (UM 2014, US 2014), información secundaria basada en cartografía digital previamente elaborada por otros autores (Etter 2003, 2008, Miles et al. 2006, IDEAM 2010, IGAC 2014) y con información recolectada en campo, a través del proceso de verificación de coberturas (González-M. *et al.* 2014). Los puntos de muestreo (tomados en campo) se distribuyeron sobre toda el área de estudio (cobertura nacional).

A las imágenes de satélite se les realizaron procesos de corrección geométrica, para ajustar factores tales como geometría, presencia de nubes, cubrimiento completo mediante mosaicos, y otros elementos. Posteriormente se efectuaron clasificaciones digitales de las áreas de Bosque Seco Tropical y se llevaron a cabo procesos de análisis y modelamiento espacial mediante diferentes técnicas de los Sistemas de Información Geográfica (SIG). Estos resultados se analizaron respecto a cartografía temática digital (información secundaria) del área real del Bosque Seco Tropical en Colombia y finalmente con los “puntos de muestreo” se evaluó la exactitud temática del mapa, el control de calidad de la corrección geométrica (ortorrectificación) y la calibración de la clasificación.

Una vez obtenidos los resultados de la interpretación de datos e imágenes en la elaboración del mapa de bosque seco tropical a escala 1:100.000, resulta obligatorio verificar la calidad de

los mismos, de cara a comprobar la validez tanto del producto generado como del método propuesto.

Los resultados obtenidos se trabajaron bajo una estructura de *geodatabase* (GDB). La cual integra los insumos cartográficos necesarios en el proceso de verificación y validación del mapa de Coberturas de Bosque Seco Tropical a nivel nacional, a escala 1:100.000.

METODOLOGÍA

INFORMACIÓN TEMÁTICA DEL MAPA DE BOSQUE SECO TROPICAL A ESCALA 1:100.000(v1.0)

El mapa de bosque seco de Colombia ha pasado por varias fases para su elaboración, la presente memoria técnica corresponde a la versión 2.0, que parte de un análisis previo en el que el elemento central consiste en la formulación de un mapa de clima, el cual permitió conocer donde se encuentran las condiciones óptimas para el desarrollo de este tipo de ecosistema. Las aproximaciones previas nacionales de clima, han estado generalmente determinado desde la perspectiva de la metodología de Caldas - Lang, que se basa en los promedios anuales de temperatura y precipitación y la relación entre ellas que permite definir rangos de clima, para los territorios evaluados. Es así como a partir de la información de WorldClim (www.worldclim.org), en donde se evidencian promedios mensuales multianuales, que se genera un mapa de balance hídrico, mensual a partir del cual se identifican aquellas regiones más propensas a sufrir rigores climáticos de estrés hídricos de por lo menos tres meses continuos.

A partir de la temperatura mensual (media, mínima y máxima) y de modelos de radiación global mensual, obtenidos a partir del modelo digital de elevaciones, se calculó la evapotranspiración de referencia (ET_o) utilizando la ecuación de Hargreaves (Hargreaves y Allen, 2003). Las imágenes de ET_o obtenidas fueron agregadas por trimestres y por año, al igual que las de precipitación (P) mensuales, para calcular el balance hídrico (P-ET_o) y el índice de aridez (P/ET_o), según Naciones Unidas (UN, 2011). Esta última aproximación

permite definir, de manera general, tanto las zonas áridas como aquellas cuyo valor del índice es inferior a 0.65. Se calculó el índice por trimestres y a escala anual. La información trimestral fue categorizada según la clasificación que propone UN en cinco clases de áreas: 1) húmedas, 2) secas sub-húmedas, 3) semiáridas, 4) áridas y 5) muy áridas. Con el fin de obtener una imagen que refleje la estacionalidad climática, se agregaron las capas trimestrales (sumándolas). El resultado de este proceso se clasificó en 5 clases que reflejan tanto la estacionalidad, como la intensidad de la sequía climática durante el año (Figura 1). Las zonas áridas se clasificaron como una sola clase (raster 0/1) y se le pasó un filtro bajo, dos veces, para suavizar el grano.

Para la identificación de los bosques secos actuales, se usó del mapa de bosques secos potenciales (originales) previamente generado, y se cruzó con el mapa de las coberturas de la tierra, clasificadas en su tercera jerarquía de acuerdo con la metodología CorineLandCover – CLC para el año 2002, que era el disponible en su momento.

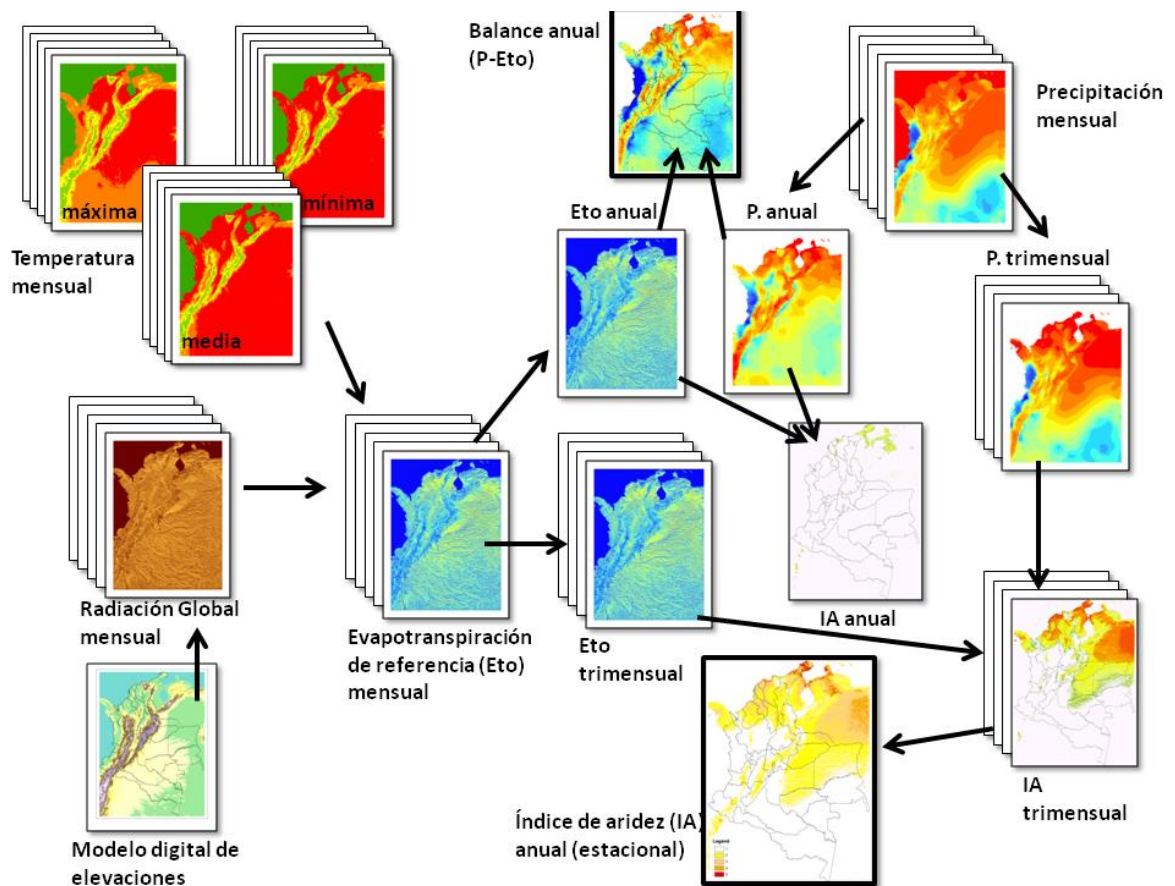


Figura 1. Método para la obtención del balance hídrico anual y del índice de aridez anual estacional, a partir de datos climáticos de WorldClim (www.worldclim.com).

La segunda versión del mapa consistió entonces en evaluar los insumos obtenidos a partir de la primera versión, validarlo y verificarlo, así como actualizar las coberturas con el nuevo CLC disponible para el año 2007 y confrontarlo con otros insumos disponibles (**Figura 2**).

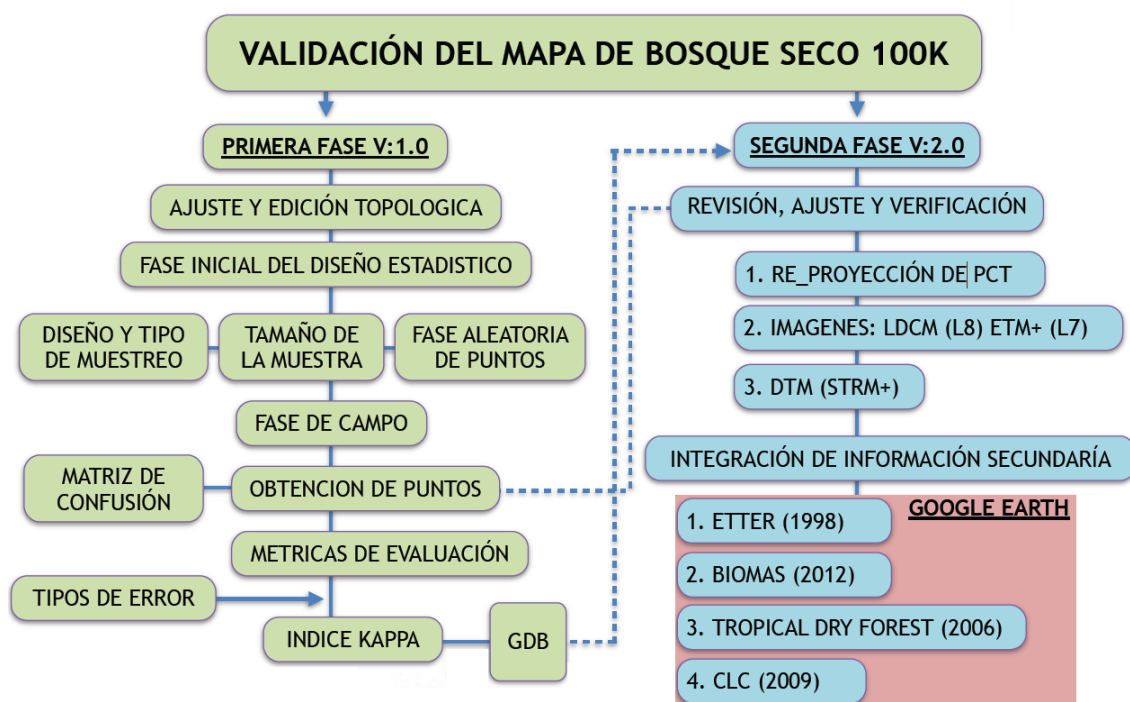


Figura 2. Metodología general de elaboración del mapa de bosques secos a escala 1:100.000(v2.0).

ANÁLISIS DE LA INFORMACIÓN SECUNDARIA PARA LA CONSTRUCCION DE MAPA DE BOSQUE SECO TROPICAL A 1:100.000(v2.0)

Dentro de esta actividad fue necesario realizar procesos de revisión de la integridad de las capas digitales existentes para bosque seco tropical en formato shape file (Figura 3), diseñar y elaborar una geodatabase (GDB) donde se alojó la información. Posteriormente se realizó una revisión de exactitud temática aplicando el método de determinación de precisión con el Índice Kappa y el método Akaike del área bajo la curva (AUC), encontrando unos resultados conformes para la escala y la importancia del producto.

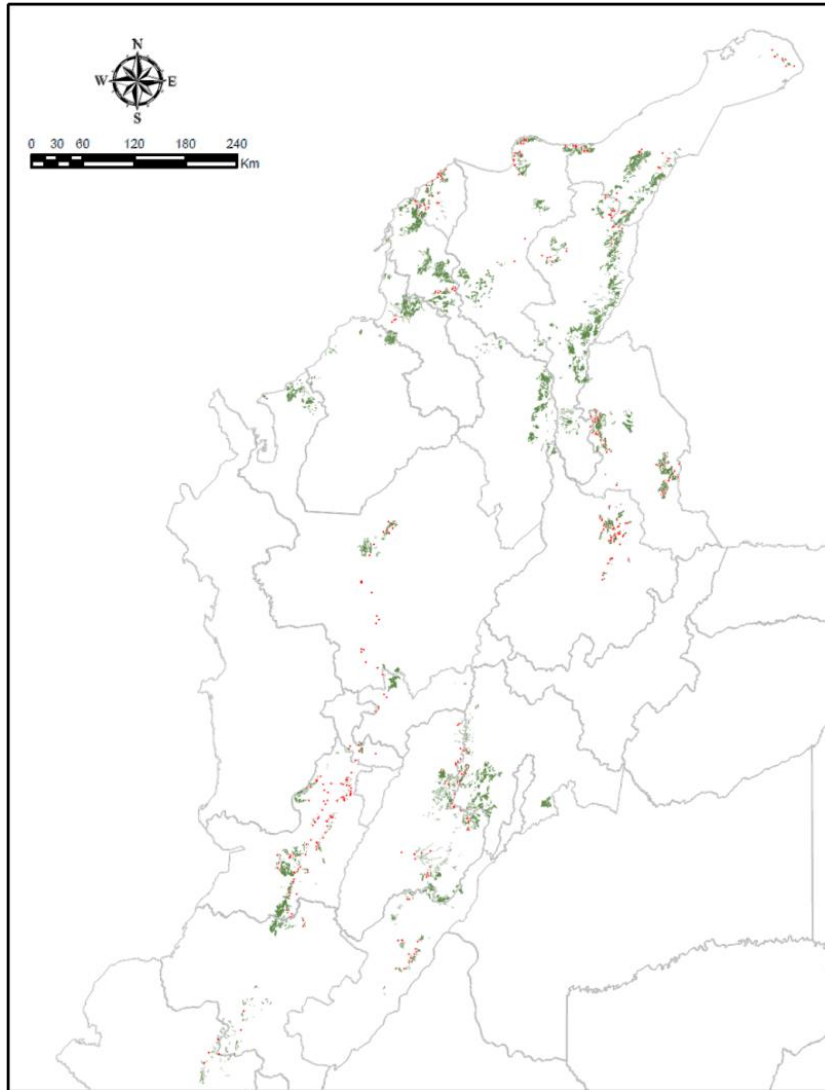


Figura 3. Puntos de muestreo (en rojo) en formato shape file a través de las diferentes regiones de bosque seco tropical en Colombia (en verde).

Con base en los datos de campo obtenidos mediante dispositivos GPS (puntos de muestreo, González-M. *et al.* 2014), se realizó una minuciosa verificación y ajuste uno por uno de los puntos recogidos (558 puntos) en cada una de las regiones con bosques secos, para tener certeza de su calidad espacial (**Figura 4**).

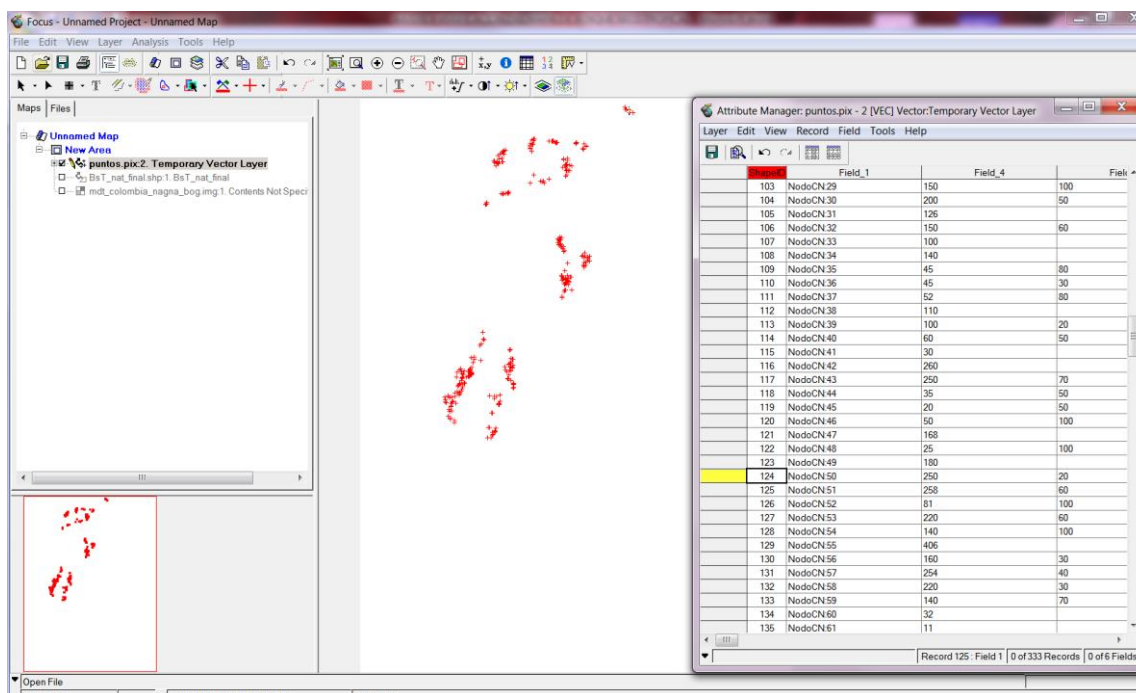


Figura 4. Puntos de muestreo en formato shape file desplegados en ArcGis (Proceso de verificación de coberturas en campo en el presente estudio a partir de datos IAvH, 2013 y 2014; González *et al.* 2014).

INTEGRACIÓN CON INFORMACIÓN SECUNDARIA

Como información temática secundaria se contó con el mapa digital de bosques secos originales de Colombia elaborado por Etter como un insumo nuevo a trabajar y que permite conocer las áreas que originalmente pertenecían a bosques secos (2003, 2008; **Figura 5**).

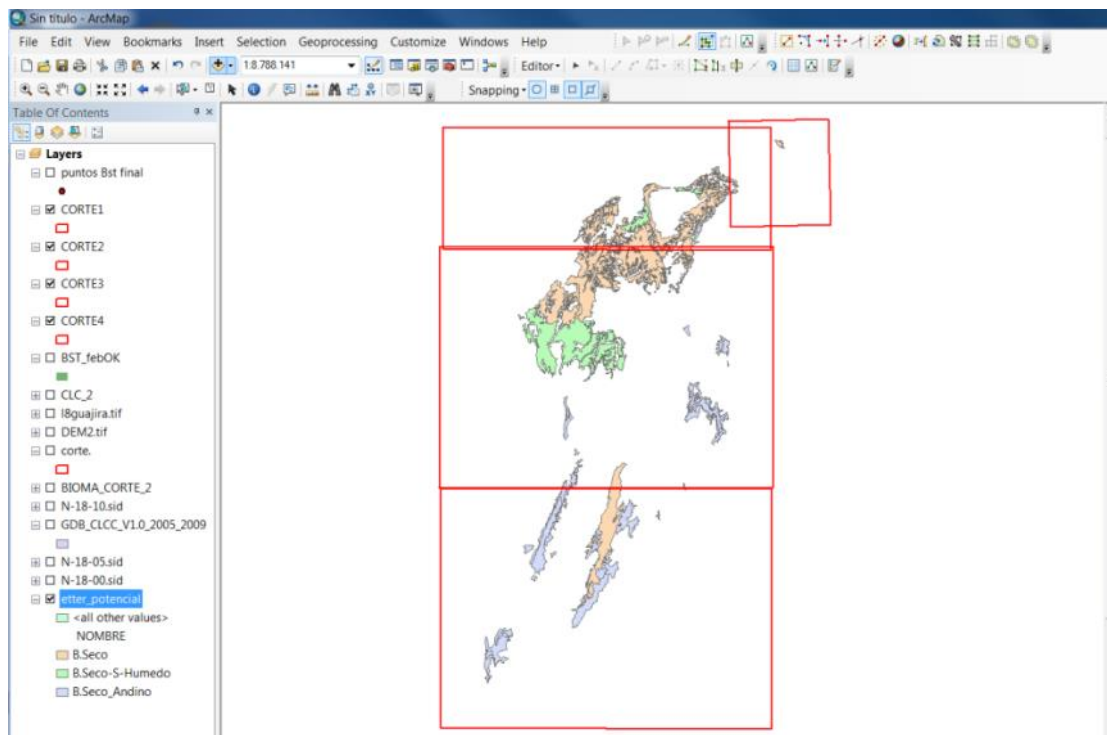


Figura 5. Mapa digital de BST de Etter (1998), desplegado en ArcGis; en rojo se aprecian las sub zonas generadas para facilitar el análisis (IAvH 2014).

Igualmente, se contó con el Mapa Nacional de Biomas desarrollado por el IAvH-Ecopetrol (2012; **Figura 6**), en donde los biomas fueron delimitados a partir de la capa de ecosistemas elaborada por el IDEAM et al. (2007), la capa de precipitación anual del IDEAM, capa de páramos (IAvH 2012), biomas de los Andes colombianos realizada por el IAvH (2004), suelos escala 1:100.000 elaborada por el IGAC, el modelo de elevación digital SRTM de 30 metros e imágenes satelitales Landsat TM y ETM (distribución libre USGS).

A partir de la capa de precipitación y el modelo de elevación se estableció el piso térmico asociado al índice de humedad de acuerdo a los modelos de regresión descritos por Eslava et al. (2001). Finalmente, se separaron unidades de acuerdo a las características edáficas y la

aparición de la vegetación, basándose en la taxonomía de suelos y la interpretación visual de las imágenes satelitales respectivamente.

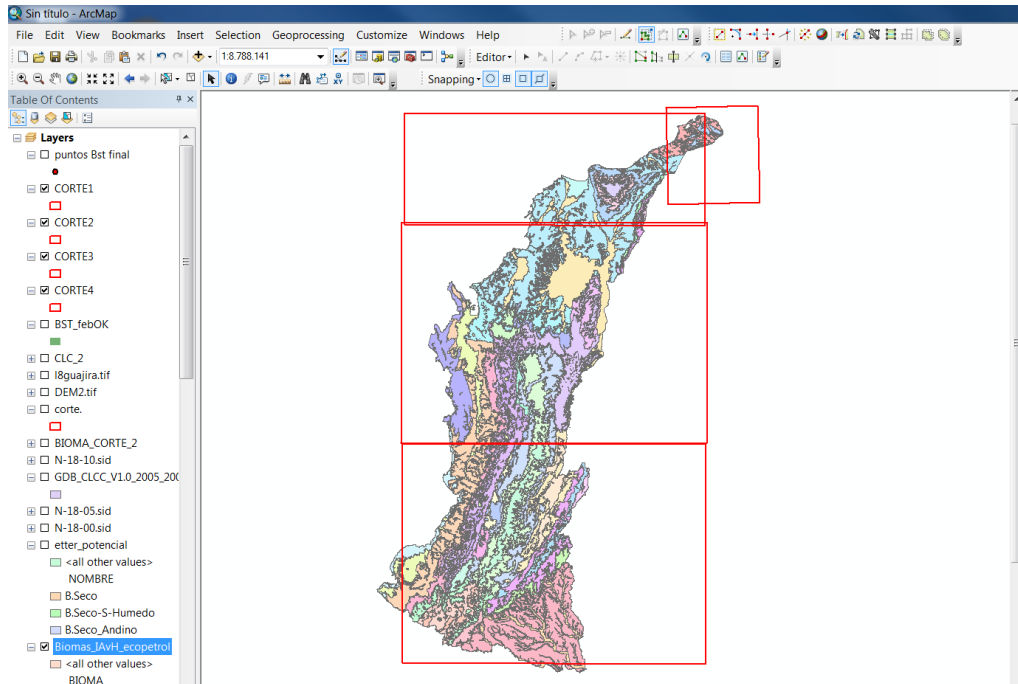


Figura 6. Mapa digital de Biomass Fase 1 y 2, desplegado en ArcGis; en rojo se aprecian las sub zonas generadas para facilitar el análisis (IAvH 2014)

Por último se contó con la información de Bosque Seco tropical de *Tropical dry forest* disponible en Miles et al. (2006; **Figura 7**), construido igualmente con imágenes satelitales y datos climáticos. Asimismo, se recurrió a las bases de datos de Corine Land Cover Nacional (IDEAM 2010) e imágenes de Google Earth en algunas áreas.

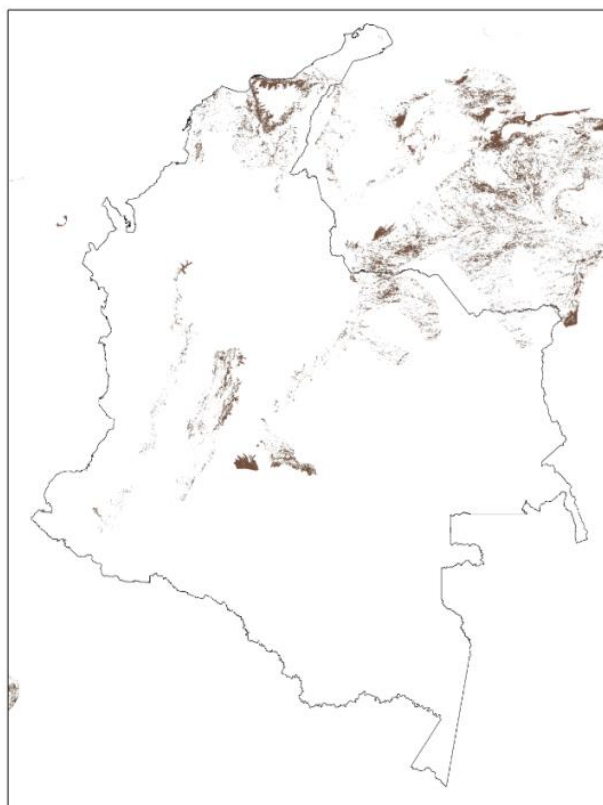


Figura 7. Información Raster de Tropical dry forest desplegada para la zona de estudio (Miles, 2006).

***ANÁLISIS Y ESTRUCTURA DE LA GEODATABASE PARA LA VERIFICACIÓN DEL
MAPA DE BOSQUE SECO TROPICAL A ESCALA 1:100.000(v2.0)***

Dentro del análisis del modelo de datos para la verificación del mapa de bosque seco tropical (BST), a escala 1:100.000, se partió de un conjunto de información cartográfica, esta fuente primaria de información cartográfica consta de una capa a escala 1:100.000 elaborada por el Programa de Biología de la Conservación y Uso de la Biodiversidad del Instituto de Investigaciones de Recursos Biológicos “Alexander von Humboldt” (Corzo y Delgado 2012) y siguiendo su propio modelo de datos. La capa cubre el término nacional, zonificada a su vez en 5 regiones de trabajo.

El proceso de importado de las hojas requiere como paso previo plasmar el Modelo de Datos (Zarazaga-Soria et al. 2003) BST 1:100.000 de alguna forma, por lo que se ha considerado la mejor opción hacerlo en una Geodatabase, la cual utiliza la tecnología ESRI de ArcGis 10.2, lo que permite generar los dominios de los datos asociados y establecer un cierto orden utilizando *feature datasets* y *feature class* para crear la estructura de datos para la verificación del mapa de bosque seco tropical y fuentes de apoyo respectivamente. Esta estructura tiene como objetivo:

- Brindar formatos propietarios, los cuales constituyen poco menos que un estándar “de facto” y garantizan una alta interoperabilidad.
- Información cartográfica ordenada y jerarquizada
- Alto control del tipo de dato a introducir así como de su formato (cifras decimales y enteras)
- Permite restringir los datos utilizando dominios, para así disminuir errores en atributos
- Permite utilizar y almacenar reglas de topología, e incluso programar reglas propias

El trabajo para la gestión de *Geodatabases* se realizó a partir de ArcCatalog, en la que se creó una nueva *File Geodatabase* vacía, que se llamó “BosqueSecoTropical.GDB”

Una vez creados los dominios y definido el modelo de verificación se ha confeccionado la estructura de la *geodatabase*, creando tantas *feature class* como primitivas geométricas se presenten en el modelo y se realizó la carga de datos.

Esquema del modelo

Dentro del desarrollo del proceso se incluyó la construcción de una *geodatabase* (GDB), la cual contiene toda la información base empleada. Se añadieron los datos y modelos, de forma realista, las relaciones espaciales de la zona de estudio inherente a los datos. Este formato de datos permitió almacenar de forma centralizada, gestionar y mantener la calidad de los datos GIS.

Se busca la integración espacial de todos los datos en un único núcleo (la geodatabase), con el fin de almacenar y administrar datos geográficos. Una vez creada la estructura de capas la Geodatabase tiene el siguiente aspecto (**Figura 8**):

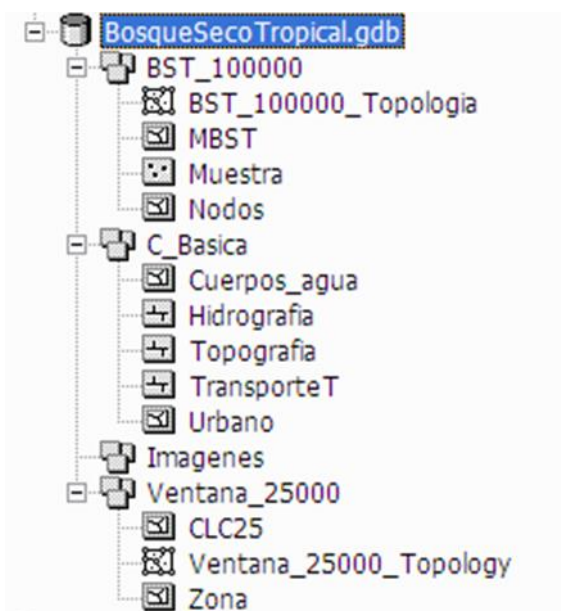


Figura 8. Esquema de las capas presentes en la base de datos. Estudio propio 2013

Estructura jerárquica de capas

Siguiendo ésta estructura se conservan los datos de partida, que pese a no estar contemplados en el Modelo de Datos para la verificación del mapa siempre pueden ser útiles en momentos de duda. Los *feature class* de la cartografía inicial se importaron directamente desde las capas de partida. Por otro lado, de manera alterna se va a crear otro dataset auxiliar, que se utilizara únicamente durante la fase de verificación de información al nuevo Modelo de Datos. Dicho *feature data set* contiene las distintas capas que conforman base.

Por otra parte, está el conjunto de *Feature Data Sets* de “Cartografía Bst 1:100.000 final”. Compuesto por un *Feature Data Set*. Los *Feature Class* de esta *data sets* se crearon vacíos, para luego importar la información a partir de la instrucción “*load data*”, ya que hubo que extraer la información y llevarlo a su *feature class* correspondiente.

El primer paso de la creación de la estructura fue crear los *Feature Data Sets*. Para ello se definió con anterioridad el sistema de referencia, que en este caso corresponde a “Magna Colombia Bogotá” (**Figura 9**) y las tolerancias que se van a asociar a toda la información cartográfica que contiene dicho *Feature Data Set* ya que no se aceptaron los valores por defecto sino que se determinaron las necesidades de los mismos en relación a la escala de la información a mostrar, que, en el caso que nos ocupa es 1:100.000.

Las tolerancias XY y Z tienen la finalidad de obligar al software a considerar la misma posición espacial de un punto para aquellos dos (o más puntos) cuya distancia (en la componente que corresponda XY o Z) sea menor que la distancia establecida en su tolerancia.

Es una estrategia sencilla pero ingeniosa para evitar que, debido a errores de redondeo y tras los cálculos al ejecutar algún análisis espacial, se cree información cuyas coordenadas no coincidan exactamente en todos sus decimales con los temas desde los que se extrajeron, produciendo errores topológicos simplemente por la imposibilidad de mantener todos los decimales en algunas operaciones.

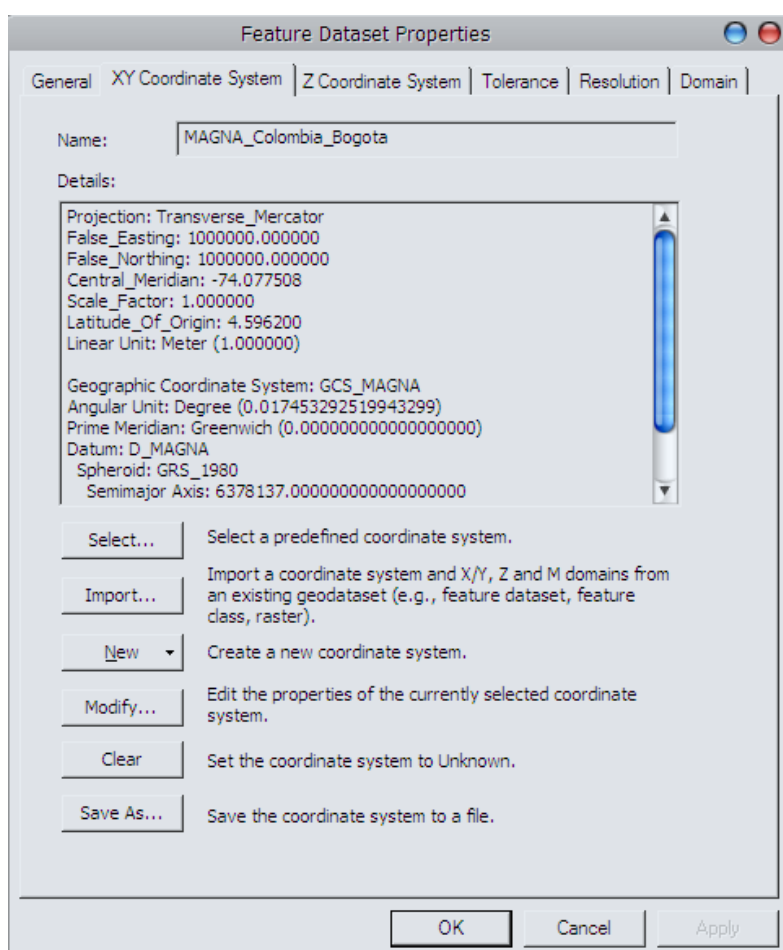


Figura 9. Sistema de referencia espacial de los *feature data set* de la *geodatabase* (GDB).

En cuanto a las rejillas XY y Z determinaron en que decimal el software “trunca” todas las coordenadas de la cartografía de dicho *feature dataset*, separándolas por planimetría, altimetría, serán las que se muestran en la tabla adjunta (**Tabla 1**).

Tabla 1. Especificaciones geodésicas de los *feature datasets*

Datum planimétrico	Magna – Bogotá - Colombia
Tolerancia XY	3 m
Rejilla XY	3 m

Puesto que se trata de producir una cartografía a escala 1:100.000, se tomará una décima de la “apreciación visual” (es decir, $0.2 \text{ mm} * \text{denominador de escala}$) para la tolerancia clúster y la centésima para la rejilla. Como el límite de apreciación a 1:100.000 es 30 metros las tolerancias serán 3 metros y rejillas de 1 centímetro, que es lo que se resumen en la tabla de creación de *Feature Data Sets*.

Los *Feature Class* que se crearon son los distintos conjuntos temáticos de insumos en la validación y verificación, lo que habitualmente se le llaman “capas”. Del mismo modo que los *shapefiles* cada *Feature Class* únicamente puede contener un tipo determinado de geometría que se define en el momento de la creación del mismo. En el caso de que se genere al importarse se creará del mismo tipo de geometría del que proviene. Y si se realiza una carga de datos deberá provenir de un tema de idéntica primitiva. Y para los *Feature Class* del *Modelo de Datos* de verificación se revisan todos los elementos para discriminar que *Data Sets* requieren que tipo de geometría, es decir, simplificar el modelo (por ejemplo) añadiendo sólo *Feature Class* de puntos en aquellos *Data Sets* en los que únicamente se vayan a representar.

Ruta y descripción de la información

La geodatabase propuesta posee el nombre de BosqueSecoTropical.gdb, en ella se han creado tres (3) *Feature Data Set*, que corresponden a la siguiente información:

1. **C_Basica:** Corresponde a la cartografía básica oficial de la zona descargada desde el Web Feature Service (WFS) = http://geocarto.igac.gov.co:8082/geoservicios/cien_mil/wfs. Publicado en el sitio web de la infraestructura Colombiana de Datos Espaciales (ICDE 2013)

En este Data Set se encuentra la cartografía correspondiente a:

- **Cuerpos_agua**, allí están Lagos, lagunas, Jagüeyes, embalses, etc. (Alias= Cuerpos_agua)
- **Hidrografía**, esta capa tiene el alias de “Drenajes” y corresponde a Ríos, quebradas, etc.
- **Topografía** (Alias= curvas_nivel),
- **TransporteT**, en esta capa se ha dispuesto toda la red vial en sus diferentes niveles (Alias= Vias)
- **Urbano**, capa que posee todas las vías de escala 1:100.000 (Alias = Vias)

2. **Imágenes**, en este *Feature Data Set* se localizan las imágenes utilizadas en el proceso de generación de cartografía temática, a saber: Orto_Sonson.ecw y Antioquia.img (Orto mosaico de imágenes Satelitales de Rapideye)

3. **BST_100000**, en este *Feature Data Set* se localizan las capas de información, en formato *Feature Class* de:

- **MBst**, Corresponde a la capa de Mapa de Bosque Seco Tropical a escala 1:100.000 (Alias Bosque Seco Tropical)
- **Muestra**, Esta capa de tipo punto tiene la información georeferenciada por los equipos de campo y que se suministraron como verdad de campo (Alias = Ptos_campo)
- **Nodos**, En este *Feature Class* se localiza la distribución espacial de las zonas para los análisis estadísticos (Alias= Zona_Nodos)
- **BST_100000_Topologia**, Esta capa no es un *Feature Class*, corresponde a la topología creada y validada para la corrección de errores topológicos de la capa MBST.
- **Zona**, corresponde al polígono que delimita la ventana de trabajo (Alias = Area_estudio)

Calidad

Para obtener un producto con todas las garantías es imprescindible tener en cuenta ciertas consideraciones ya que el grado de precisión y fiabilidad en una cartografía es variable, y que depende en gran medida del método de captura o de su procedencia.

El estudio de la calidad ha sufrido un gran impulso en los últimos tiempos, ya que han cambiado las estrategias y se pretende no sólo producir, sino producir bien, de modo que

aquel producto que se genere tenga una característica extra que le permita triunfar en el mundo de la competencia. Es por ello que se han ido desarrollando desde diversos frentes y organismos estrategias y documentos para asegurar, analizar y cuantificar (en la medida de lo posible) una característica tan abstracta como es la calidad. Hay diversos tipos de normas en función del organismo que las propone, y como se pretende seguir una norma cuanto más extendida sea mejor, con el fin de presentar un producto integrador como es este Modelo de Datos de verificación y validación, para que sea una garantía de confiabilidad.

En efecto para el mapa de Bst 1.100.000 se han utilizado normas nacionales de calidad como la **Norma Técnica Colombiana** (NTC) de ICONTEC¹ (2002 2007, 2009) para tener unos parámetros de referencia de la calidad de los datos. Concretamente las NTC 5043 (Conceptos básicos de calidad), NTC (Evaluación de la calidad), NTC 4611 (Segunda actualización – Metadatos) y NTC 5662 (Especificaciones técnicas), que además proponen valores de aceptación y rechazo para algunos de esos parámetros. La geodatabase exige que los resultados de los controles de calidad realizados sobre los datos sean incluidos en los metadatos.

Hay ciertos controles específicos para algunos procesos de validación en particular, y vienen detallados en las fichas de los mismos en el catálogo. Sin embargo, a modo global, existen los siguientes parámetros de calidad en la información cartográfica:

¹ El Instituto Colombiano de Normas Técnicas y Certificación (ICONTEC), Organismo Nacional de Normalización de Colombia. Entre sus labores se destaca la creación de normas técnicas y la certificación de normas de calidad para empresas y actividades profesionales. ICONTEC es el representante de la Organización Internacional para la Estandarización (ISO), en Colombia.

- Exactitud posicional
 - Absoluta horizontal
 - Absoluta vertical
 - Relativa vertical

- Compleción
 - Omisión
 - Comisión

- Consistencia lógica
 - Consistencia de dominio
 - Consistencia conceptual o Coherencia lógica

- Exactitud temática
 - Corrección de la clasificación
 - Corrección del nombre geográfico

Metadato

Los metadatos son una parte fundamental de cualquier base de datos cartográfica que se precie, proporcionando información acerca de los datos, como por ejemplo la fecha de creación o última modificación, el sistema de referencia utilizado o el organismo o persona que los elaboró.

Unos metadatos bien estructurados y completos son la garantía de que se posee la genealogía de los datos cartográficos. Para asegurar este punto, en la metodología se recomienda seguir esquemas como los contemplados en la norma ISO 19115 y específicamente la del NTC 4611 (Segunda actualización – Metadatos), cuyas recomendaciones garantizan unos metadatos de calidad y descriptores de la información cartográfica tal como se especifica en los requerimientos del Instituto Geográfico Agustín Codazzi (IGAC, 2014).

Se sugiere, para la creación de los metadatos una de las mejores herramientas disponibles, el software de ESRI ArcGis 10.x, este software soporte contempla además varios esquemas de metadatos ISO, entre otros los del Núcleo Colombiano de Metadatos. Dicha herramienta genera un fichero de tipo xml siguiendo un esquema y puede (y debe) asociarle la nomenclatura recomendada para la correcta identificación del mismo.

OBTENCIÓN Y AJUSTE DE LAS IMÁGENES DE SATÉLITE Y LA INFORMACIÓN TEMÁTICA

Para la obtención de las imágenes de satélite y la información temática para realizar la validación temática se descargó del sitio web del Servicio Geológico de los Estados Unidos (USGS) <http://earthexplorer.usgs.gov/> y de la Universidad de Maryland (UM) <http://glcfapp.glc.f.umd.edu:8080/esdi/> las imágenes del programa satelital Landsat, a saber **(Figura 10)**:

- 3 mosaicos nacionales Landsat 7 ETM+ (2002 a 2004) (ESDI_UM) (*Earth Science Data Interface*) (UM, 2014)
- 1 imagen OLI (*Operational Land Imager*) Landsat 8 (2014) de USGS (US 2014)

Como el área de estudio es considerablemente grande fue necesario dividirla en 4 sub zonas para hacer más ágil el manejo de las diferentes capas que se analizaron, en la Figura 5 se pueden apreciar en color rojo.

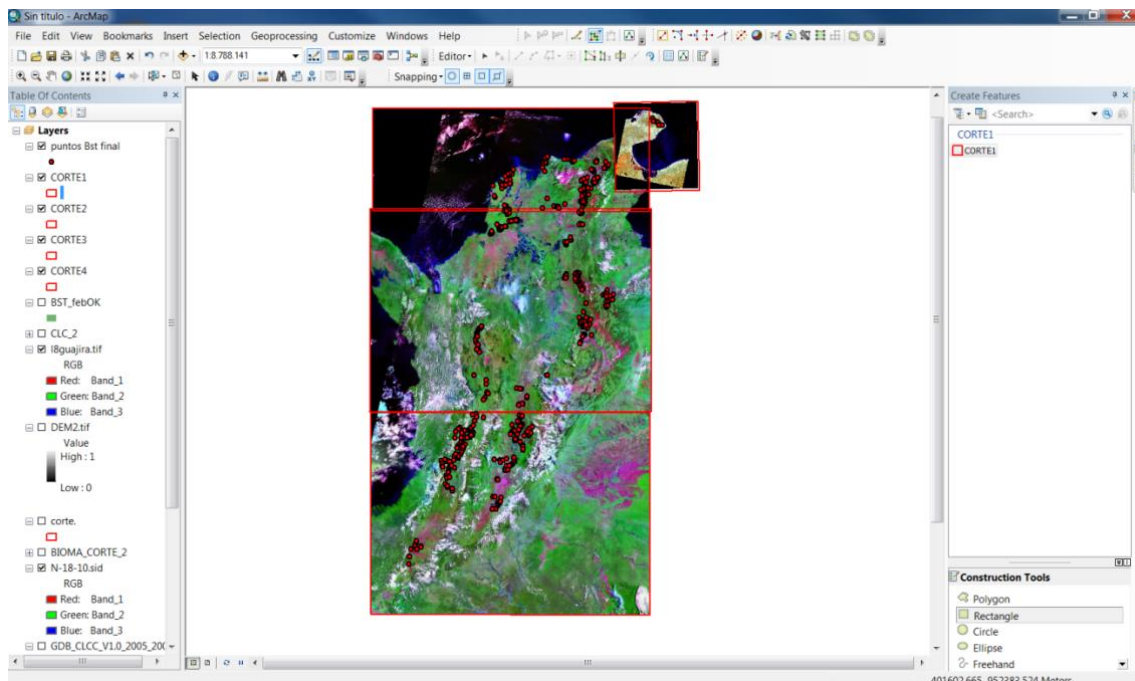


Figura 10. Puntos de muestreo sobre mosaicos de Landsat descargados desplegados en ArcGis (Proceso de verificación de coberturas en campo en el presente estudio a partir de datos IAvH, 2013 y 2014; González et al. 2014 y de la UM y USGS).

Una vez se contaba con las imágenes de satélite se procedió a adquirir el Modelo Digital de Elevación para obtener los valores de altura. Esta imagen raster se obtuvo del IGAC y

corresponde a un SRTM (*shuttle radar topography mission*) de un arco segundo de resolución de pixel, que corresponde a 30 metros de tamaño de pixel. Este SRTM se cortó al límite nacional para obtener sólo la matriz raster de la zona de estudio (**Figura 11**).

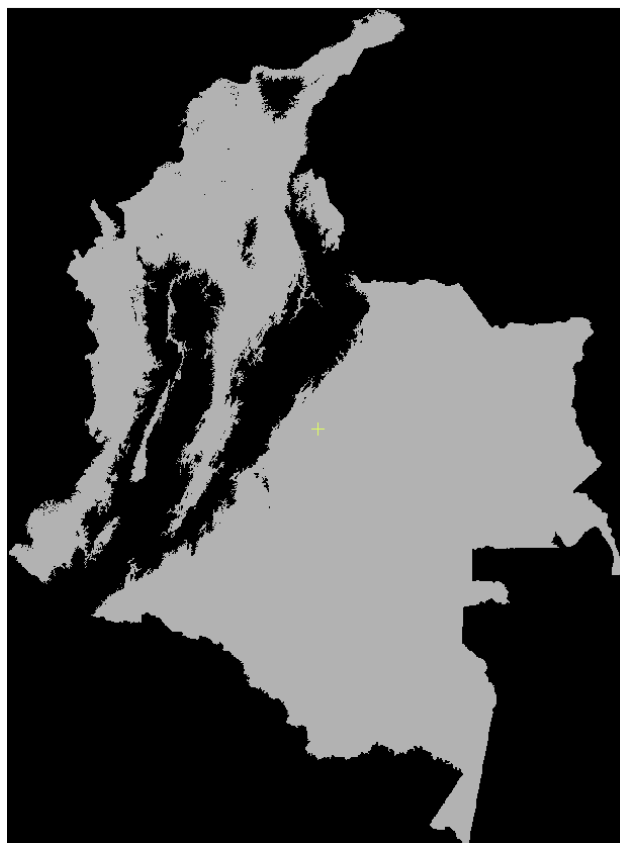


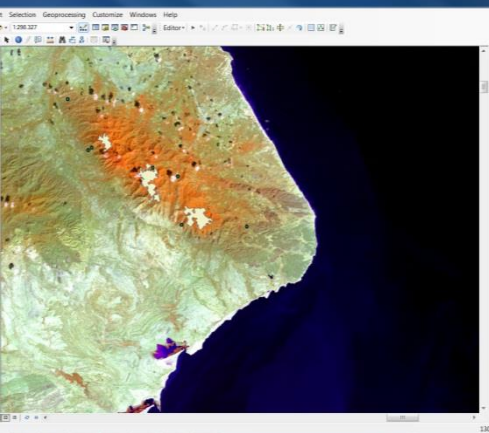
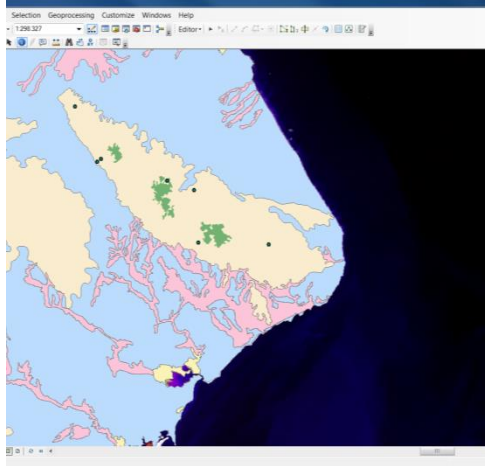
Figura 11. SRTM desplegado en PCI Geomatica 2013 completo, SRTM ajustado a la zona de estudio (Estudio a partir de datos IGAC 2014).

AJUSTE Y CORRECCIÓN DE LA INFORMACIÓN

Una vez se contó con el mapa preliminar de BST, la información temática digital secundaria y las imágenes de apoyo, se procedió a realizar la verificación temática por inspección visual. Se tomó como verdad de campo la capa de puntos de muestreo y se ajustó el mapa preliminar de

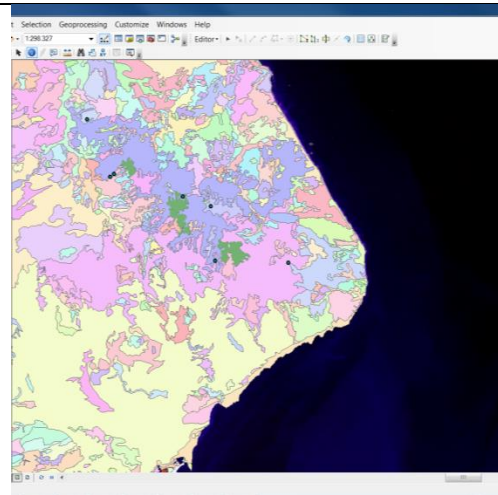
BST a partir del cruce espacial de la información secundaria para ajustar los polígonos temáticos de BST.

Tabla 2. Procedimiento para la verificación temática por inspección visual

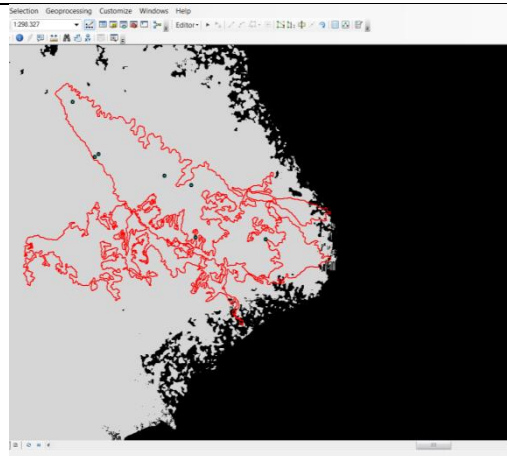
Procedimiento	Grafica
Carga de los polígonos de la base de datos de BST preliminar sobre imagen Landsat 8.	
Despliegue de la información de BST preliminar sobre la capa de Biomas	

Procedimiento**Grafica**

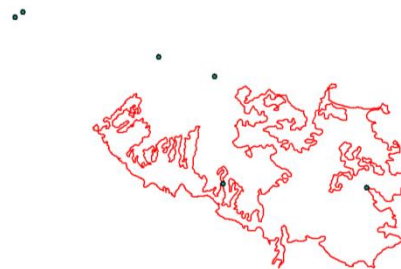
Despliegue de la información de BST preliminar
sobre la capa de Corine Land Cover



Una vez se ha inspeccionado visualmente se hace
un cruce espacial de todas las capas mediante
ArcGis



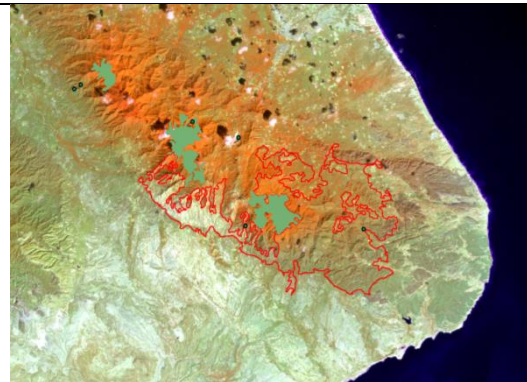
El producto del cruce espacial se simplifica a un
solo polígono actualizado de BST en donde se
aprecia que los puntos de muestreo ya están
incluidos



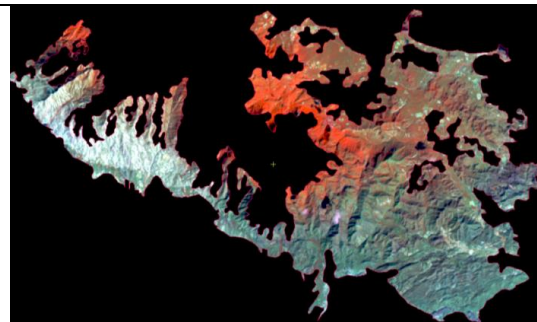
Procedimiento

Grafica

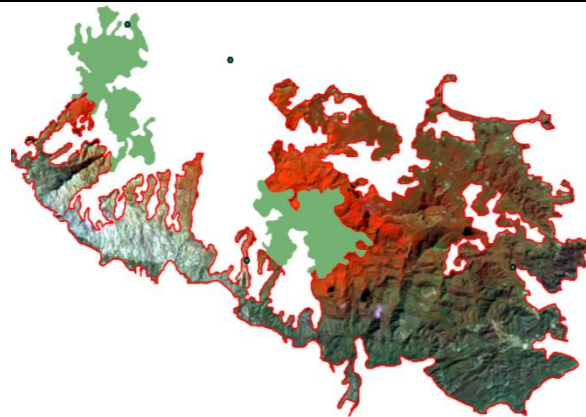
Sobreponiendo el polígono resultado del cruce espacial con la capa preliminar de BST y desplegándolas sobre la imagen Landsat



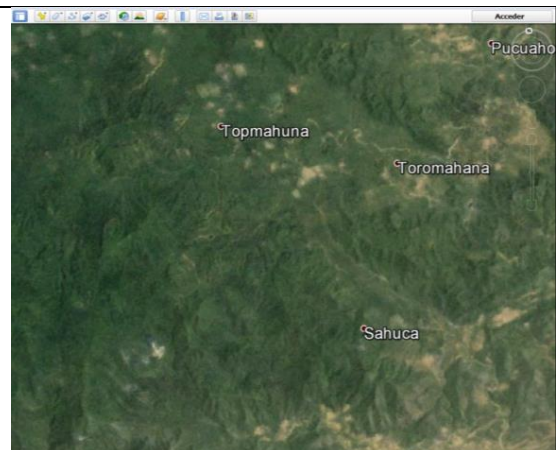
Se procede a hacer un corte de la imagen por el polígono resultado del cruce de capas



El corte de la imagen se sobrepuso con los polígonos del mapa de BST preliminar.



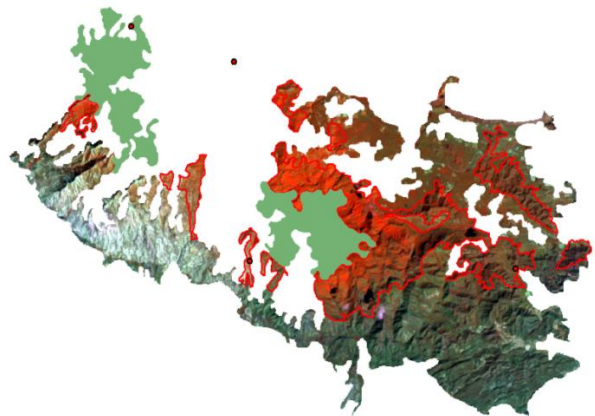
Se identificó la zona en imágenes de mayor detalle de Google earth



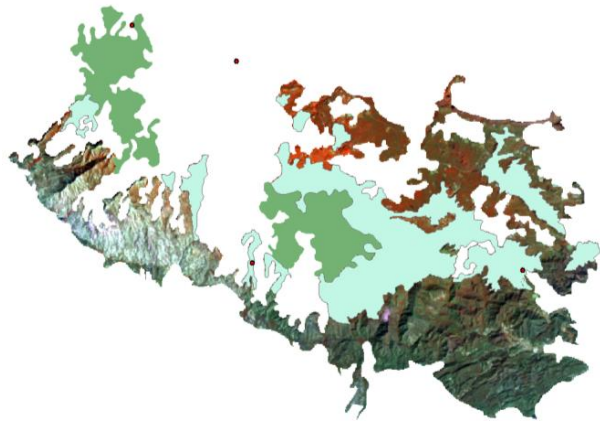
Procedimiento

Grafica

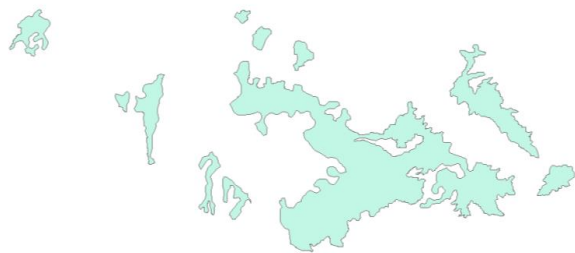
Una vez chequeado en detalle el área de la zona real de BST se procede a editar en ArcGis el polígono resultado del cruce espacial de capas digitales de información secundaria, para ajustarlo a la expresión pictórica de los pixeles de la imagen Landsat 8 (línea en rojo)



El polígono editado se ajusta a la información suministrada por la imagen de satélite.



Se generan nuevos polígonos de BST



Se unen como un solo polígono (Marge).



Procedimiento

Grafica

Se verifica el nuevo polígono contra la capa de Corine Land Cover para determinar que coberturas de Corine involucra en este caso: 3222 (Arbustal abierto) y 245 (Mosaico de Cultivos y espacios naturales) que perfectamente se pueden confundir con BST.



METODOS ESTADISTICOS Y ANALISIS DE DATOS

VALIDACIÓN ESTADÍSTICA Y CONFIABILIDAD TEMÁTICA

La evaluación de la confiabilidad temática consiste en comparar la información del mapa obtenido con la información de referencia considerada muy confiable (datos de campo). Generalmente se basa en un muestreo de sitios de verificación, cuya clasificación se obtiene a partir de observaciones de campo o del análisis de imágenes más detalladas (con mejor resolución espacial), que aquellas utilizadas para generar el mapa.

Una vez obtenidos los resultados de la interpretación de datos e imágenes en la elaboración del mapa de Bst 1:100.000, resulta obligatorio verificar la calidad de los mismos de cara a comprobar la validez tanto del producto generado como del método propuesto.

La verificación de resultados permitirá al usuario final valorar el grado de exactitud de acuerdo con la realidad, con el fin de reducir el riesgo que se asume al tomar decisiones sobre la información que se ha generado.

Este proceso de verificación requiere que comparemos nuestros resultados con una fuente externa, que se considere fiel representante de la realidad presente en el terreno en el momento de adquirir la imagen.

Este proceso de verificación requiere aplicar técnicas de muestreo que nos permitan estimar, con la mayor precisión, el error que contienen los resultados obtenidos. En este sentido este trabajo se dirigió a estimar la variable de error a partir de seleccionar una muestra suficientemente representativa de las condiciones del terreno. Posteriormente, el análisis

estadístico de estos pares de observaciones (resultados-realidad) permitirá estimar numéricamente este grado de error.

El objetivo de este procedimiento es contar con la veracidad de que cada categoría de clasificación se encuentre realmente presente en los puntos que se muestran en el mapa. De lo contrario nos encontraríamos en una de las siguientes situaciones:

- Los “límites” que separan las categorías son no son válidos como se han localizado en la clasificación.
- Errores en la **adquisición** de los datos.
- Errores en el **procesamiento** de los datos.
- Errores dependientes a la “**escena**”.
- Resolución del sensor.

Generalmente la validación estadística de la interpretación comprenderá las siguientes fases (Congalton y Green, 1999)

Fase inicial de diseño

a. Diseño de muestreo sistemático no alineado

El diseño de muestreo contempla la determinación del tipo de unidades de muestreo, del método de selección de las mismas, así como del número de unidades de muestreo necesarias (tamaño de muestra).

La unidad de muestreo permite relacionar la localización de la información del mapa y del terreno. Puede ser un punto, un píxel, un grupo de píxeles, un polígono del mapa o una unidad de superficie con formas predeterminadas, por ejemplo, un cuadro o un círculo de una hectárea. No existe un consenso definitivo sobre la unidad de muestreo más adecuada (Chuvienco, 2010).

Si la unidad de muestreo es un punto como en el mapa de bosque seco tropical a escala 1:100.000, se compara la clasificación del mapa con relación a este punto con la misma localización en el terreno; en la práctica, lo que se evalúa es una superficie alrededor del punto.

En el caso de unidades de superficie predeterminadas, la superficie que debe cubrir el sitio de muestreo es también delicada de determinar, un sitio de verificación grande puede incluir varias porciones de polígonos en el mapa y varios tipos de cubierta en el terreno o en la imagen de referencia, lo que genera ambigüedades al confrontar la información del sitio de verificación con la del mapa. Al contrario, un sitio de verificación pequeño puede coincidir con una unidad del paisaje no representada en el mapa por ser más pequeña que el mínimo cartografiable del mismo (**Figura 12**).

El método de muestreo elegido sirve para seleccionar una pequeña muestra del área cartografiada, de tal forma que sea representativa de la totalidad del mapa. En un diseño de muestreo probabilista, todas las unidades de muestreo presentes en el mapa tienen una probabilidad conocida superior a cero de ser seleccionadas, a esta probabilidad de selección se le denomina probabilidad de inclusión. Así mismo, durante la selección de las unidades de muestreo no se deben descartar sitios que presenten ciertas características; como por ejemplo, estar localizados en regiones con poca accesibilidad o en terrenos privados.

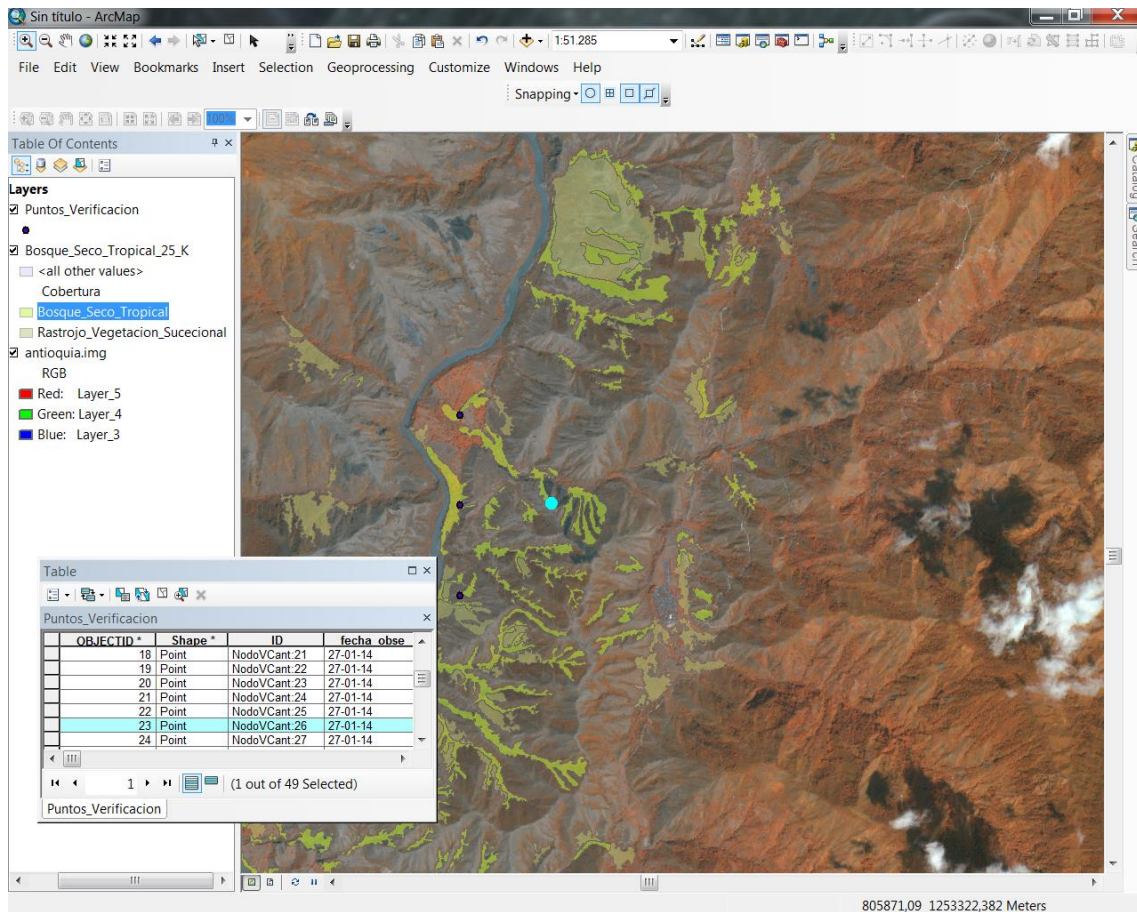


Figura 12. Unidades de muestreo puntual de diferentes superficies y formas. (Estudio Propio).

De acuerdo a la cartografía generada se ha elegido la técnica de muestreo **sistemático no alineado**, por ser una de las más empleadas en el proceso de evaluación de este tipo. Ya que la muestra se distribuye de manera regular, pero con un cierto grado de libertad y permite representar todo el territorio cubierto por el mapa de bosque seco tropical 1:100.000

El método sistemático no alineado consiste en ubicar en forma aleatoria el primer punto de muestreo y, a partir de su ubicación, determinar la ubicación de todos los demás puntos en forma sistemática pero introduciendo una modificación aleatoria de la ubicación espacial de éstos en uno o los dos ejes. De esta manera la distancia entre los puntos planteada es de 500 m, con lo que el rectángulo de 125 mil hectáreas se divide en 5 000 cuadros que contendrán los puntos de muestreo.

De esta cantidad se deben descontar los puntos que caen fuera de las franjas accesibles por las vías accesibles.

La generación de las coordenadas de los puntos puede hacerse con la ayuda de un programa de hoja de cálculo utilizando una función que genere valores aleatorios. La utilización de un punto de muestreo en particular, en la evaluación estadística, dependerá también del azar y de la cantidad de unidades requeridas para satisfacer el nivel de exigencia previsto para la prueba.

b. Cálculo del tamaño de muestra

El tamaño de la muestra se refiere al número de sitios de verificación utilizados para estimar la confiabilidad del mapa. Entre más grande sea el tamaño de la muestra, más precisa será la evaluación; sin embargo, por razones de costo y tiempo, es conveniente determinar el tamaño de muestra mínimo, para alcanzar los objetivos de la evaluación.

Congalton (1988) sugiere muestrear una superficie aproximada al 1% de la superficie cartografiada. En otra publicación el mismo autor recomienda verificar por lo menos 50 sitios por categoría, y de 75 a 100 si el área en estudio es superior a 400 000 ha o si hay más de 12 categorías (Congalton 1991).

En el caso de una imagen clasificada como en este caso, en donde la variable no es cuantitativa sino categórica, se recomienda emplear la distribución binomial de probabilidad. La expresión más sencilla (sin considerar la corrección para poblaciones finitas) sería la planteada por Chuvieco (2010), en donde el cálculo del tamaño de la muestra se basa en la fórmula:

$$n = \frac{Z^2 pq}{E^2}$$

Donde Z es el valor de la abscisa de la curva normal estandarizada para un nivel determinado de probabilidad; p indica el porcentaje de aciertos estimado; q, el de errores ($q = 1 - p$) y E, el nivel permitido de error. Considerando los porcentajes de acierto y error iguales a 85% y 15%, además el error estándar permisible 5%, para 95% de confianza, la fórmula queda definida en la forma siguiente:

$$N = (196^2 * 85 * 15) / (5^2) = 195.92$$

Esto significa que el **tamaño mínimo** es **196** puntos de verificación

Esta cantidad de puntos se debe distribuir en forma proporcional al área de los estratos, pudiendo resultar en un número mayor de puntos, pues las fracciones se convierten en puntos enteros y además se consideró un mínimo de tres (3) puntos de muestreo por clase de bosque seco. Sin embargo, al trabajar con una variable dicotómica (áreas de BST y áreas de no BST) pueden plantearse otros criterios de validación que garanticen la confiabilidad de exactitud a partir de la redundancia del tamaño de la muestra. En este caso se seleccionaron **585 puntos** de verificación tomados en campo. Lo que indica que el procedimiento que se siguió al momento de realizar esta verificación de exactitud temática sobrepasa el mínimo de puntos requeridos técnicamente.

c. Selección aleatoria de puntos de muestreo

Los pasos y consideraciones seguidas para la ubicación de los puntos de muestreo de la validación estadística son los siguientes:

- Se hizo un reconocimiento de todo el mapa con el fin de encontrar la zona donde se ubicará la grilla para la validación estadística, la misma que deberá incluir todas las clases de interés para el estudio y que sean accesibles por vía terrestre.

- Cada uno de los puntos que conforman la grilla tiene en su tabla de atributos los códigos de las unidades interpretadas que las contengan, con el fin de hallar la superficie de cada clase al interior de la grilla y, posteriormente, determinar el número de puntos de muestreo por cada clase de interés en proporción a su superficie.
- Utilizando las imágenes de satélite del proyecto, se procederá a actualizar las vías de acceso del área de validación digitalizándolas en pantalla a escala 1:100 000 con el fin de conseguir una adecuada planificación para el trabajo de campo.
- A partir de la red vial actualizada del área de validación, se generara un Buffer o área de influencia de 500 m a ambos lados de estas vías de acceso.
- Se realizara la selección de los puntos que conforman la grilla y que se encuentran al interior del buffer de 500 m; es decir, se seleccionaran los puntos que son accesibles vía terrestre.
- Se realizara el trabajo estadístico para encontrar el porcentaje de cada clase de interés para el estudio de coberturas de bosque seco que se encuentran representadas en la grilla de 25 x 50 km.

Fase de campo

En cada zona de validación se localizaron los puntos de muestreo seleccionados aleatoriamente, empleando para ello receptores GPS. A continuación se verificó que las clases de áreas de bosque seco del mapa correspondían a la verdad del campo.

Como las fechas entre las imágenes de satélite y la fecha del trabajo de campo, difieren en algunos años, se realizaron las consultas, por medio de encuestas no estructuradas, sobre el uso actual y del uso anterior de las unidades verificadas a los dueños de los predios.

Así mismo, se realizó también la caracterización de las clases, acompañadas de material de video y fotográfico y finalmente se levantaron parcelas de muestreo para los bosques secundarios.

La inspección de los puntos dio como resultado un listado de puntos de verificación con sus respectivos datos de clasificación, según el mapa de Bst 1:100.000 y su correspondiente información de verificación, en forma real, proveniente del trabajo de campo (González-M. et al. 2014).

a. Obtención de puntos de validación

Dentro del levantamiento de puntos que se emplearon en el proceso de validación del mapa como se comentó anteriormente se dividió en dos grupos así: Aquellos puntos predefinidos o pre establecido y aquellos puntos establecidos in-situ (en campo), con base al análisis de la información recogida de estos puntos.

A partir de ellos se determinaron las métricas e instrumentos de evaluación, como matrices de confusión, para calificar los productos, a continuación se explican.

- i. Puntos predefinidos o preestablecidos: Para este tipo de puntos se trabajó con el método de grilla y con el método predeterminado.

- **Método Grilla**

Para este método se generó a partir de un diseño de muestreo sistemático, empleando para ello una grilla de 10 km x 10 km, los puntos de intersección de la grilla corresponden a los puntos de levantamiento de información en campo y se emplearon en el proceso de control de calidad de la exactitud temática del mapa.

- **Método predeterminado**

Consiste en aquellos puntos que son construidos y programados previamente en las reuniones técnicas, en base a las dudas identificadas. Estos puntos se emplearán como información de calibración para proceso de clasificación (Patronamiento).

- ii. Puntos establecidos in situ (en campo): para la obtención de esta información se trabajó bajo dos criterios, sobre la marcha y por puntos de interés.

- **Sobre la marcha**

Bajo esta metodología, se registró la información de puntos siguiendo un patrón sistemático, en función de una unidad establecida de distancia o tiempo de recorrido directamente en campo. Estos puntos se emplearon para validación.

- **Puntos de interés**

Son aquellos puntos que en campo se apreciaron con particularidades dignas de documentar y utilizar.

A continuación, en la **Figura 13**, se aprecian sobre el software PCI Geomatica 2013 la tabla de puntos de verificación en campo.

Sample Number	Georeferenced Position		Database Position		Classified Value Name	Reference Value Name
	Easting	Northing	Pixel	Line		
1	618997.96	627076.93	13792	45258	1 Bst	1 Bst
2	620650.29	646475.30	13847	44611	1 Bst	1 Bst
3	625541.83	667870.91	14010	43898	2 No Bst	2 No Bs
4	641054.87	668639.86	14527	43872	1 Bst	1 Bst
5	652836.77	674226.36	14920	43686	1 Bst	1 Bst
6	652490.55	674537.59	14908	43676	2 No Bst	1 Bst
7	653185.51	675527.00	14931	43643	1 Bst	1 Bst
8	635142.31	678595.86	14330	43541	1 Bst	2 No Bs
9	636530.92	679562.64	14376	43508	1 Bst	1 Bst
10	639509.99	687298.06	14476	43251	1 Bst	1 Bst
11	639393.68	687704.36	14472	43237	1 Bst	1 Bst
12	639372.79	689557.98	14471	43175	1 Bst	1 Bst
13	847555.36	777535.69	21410	40243	1 Bst	1 Bst
14	847555.36	777535.69	21410	40243	1 Bst	1 Bst
15	857183.98	777958.83	21731	40229	1 Bst	1 Bst
16	857183.98	777958.83	21731	40229	1 Bst	1 Bst
17	863068.80	786060.76	21928	39958	1 Bst	1 Bst
18	863068.80	786060.76	21928	39958	1 Bst	1 Bst
19	864898.62	789951.51	21989	39829	1 Bst	1 Bst

Figura 13. Tabla de puntos de verificación en campo del mapa de bosque seco tropical a escala 1:100.000

Fase final de procesamiento y análisis de datos

Para el análisis de los datos de confiabilidad se hizo a través de una matriz de confusión, que permite confrontar la información de los sitios de verificación con aquella de la base cartográfica que se pretende evaluar en este caso el mapa de interés BST 1:100.000.

En la matriz de confusión, las filas representan las clases de referencia (puntos de verificación) y las columnas las clases del mapa. La diagonal de la matriz expresa el número de sitios de verificación para los cuales hay concordancia entre el mapa y los datos de referencia, mientras los marginales indican errores de asignación.

a. Matriz de confusión

Como ya se mencionó, este procedimiento permitió el cálculo de una serie de medidas que describen la exactitud de la clasificación del mapa de BST 1.100.000 respecto a la referencia.

Para ello se generaron “píxeles”-muestra de la imagen clasificada (mapa de BST 1.100.000) para compararlos con áreas correspondientes (ubicación y tamaño) del muestreo de campo a partir de los puntos.

Con los datos obtenidos en las fases anteriores se construyó una matriz rectangular, donde las filas corresponden a las clases referenciales (verdad de campo); mientras que las columnas corresponden a las clases del mapa

La relación entre el número de puntos correctamente asignados y el total, expresa la fiabilidad global del mapa. Los residuales en filas indican tipos de cubierta real que no se incluirán en el mapa; mientras que los residuales en columnas implican categorías del mapa que no se ajustan a la realidad. En la **Figura 14** se muestra esta matriz de confusión generada en la validación del mapa de bosque seco tropical (BST 1:100.000) en el ambiente del software PCI Geomatica versión 2013.

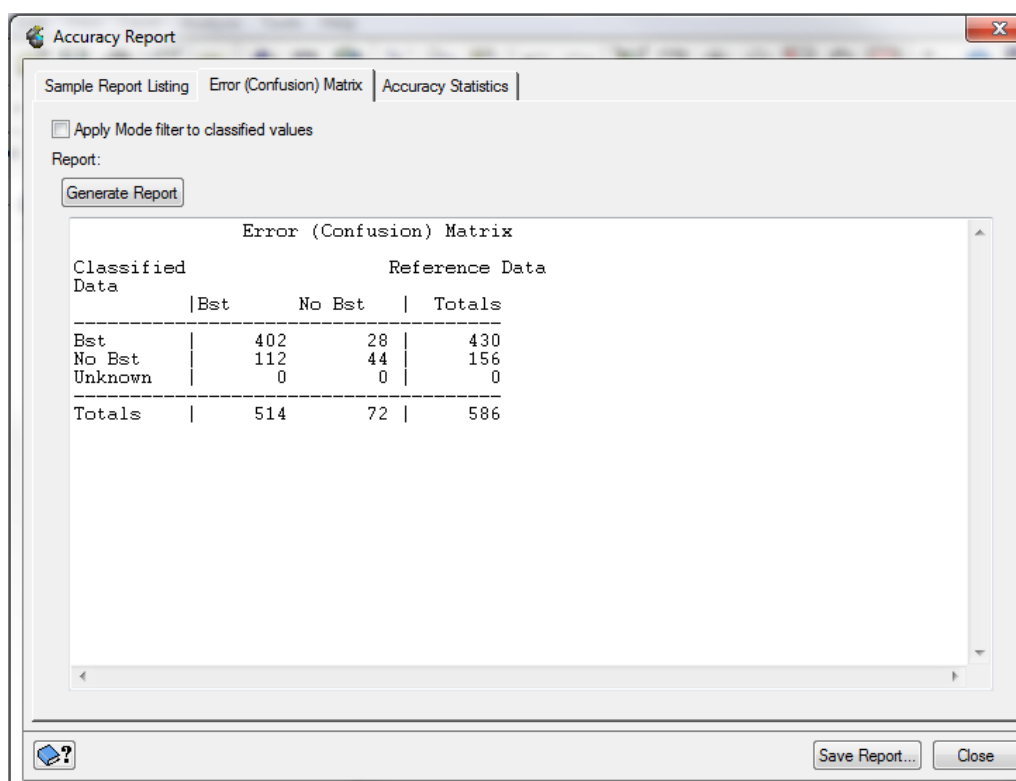


Figura 14. Matriz de Confusión entre las clases de bosque seco tropical Bst a escala 1:100.00. Bst= Bosque seco tropical, No Bst= Otras categorías, Unknow= Desconocido. (Estudio Propio).

Para determinar la confiabilidad del mapa, se empleó la siguiente fórmula:

$$F_m = \frac{\sum X_i}{\sum \sum X_{ij}} \times 100$$

Donde, **Fm** es la confiabilidad del mapa, $\sum X_i$ es la sumatoria de los aciertos o acuerdo observado, $\sum \sum X_{ij}$ es el total de los puntos muestreados.

b. Tipo de Errores

Del análisis anterior podemos derivar dos tipos principales de error:

- **Errores de “Comisión”**
 - Aquellos pixeles que se clasificaron en una clase(s) equivocada(s).
 - Cuando se separó una clase en 2 o varias clases.

- **Errores de “Omisión”**
 - Cuando se omitieron pixeles de clasificarse en alguna clase.

c. Índice de Kappa (k)

Para analizar las relaciones múltiples entre las distintas categorías, se trabajó con técnicas de análisis categórico multivariante, idóneas para el estudio global de matrices de confusión. (Congalton y Mead 1983, Campbell 1987, citados por Chuvieco 1990).

Uno de los índices más empleados, en este sentido es el estadístico de **Kappa (k)** que mide la diferencia entre el acuerdo mapa-realidad observado y el que cabría esperar simplemente por azar. En definitiva, intenta delimitar el grado de ajuste debido solo a la exactitud de la clasificación, prescindiendo del causado por factores aleatorios. El test pretende evaluar si la clasificación ha discriminado las categorías de interés con precisión significativamente mayor a la que se hubiera obtenido con una asignación aleatoria. (Chuvieco, 1990).

La fórmula empleada es:

$$k = \frac{N \sum X_{ii} - \sum X_{i+} X_{+i}}{N^2 - \sum X_{i+} X_{+i}}$$

Donde X_{ii} indica el acuerdo observado, y $X_{i+} X_{+i}$ (producto de marginales) la concordancia esperada. El valor obtenido representa el porcentaje en que la clasificación es mejor que la esperada por azar.

Como ejemplo de interpretación se dirá que si el valor obtenido fuera 0.80 se dirá que la clasificación del mapa es un 80% mejor que la esperada por azar.

Para el caso del BST 1:100.000 se consideraron tres clases de resultados: un valor k menor de 0.4 representó una concordancia pobre; un valor entre 0.4 y 0.8, una concordancia moderada, y un valor mayor de 0.8, una fuerte concordancia.

El resultado final obtenido en la exactitud de clasificación del mapa de bosque seco tropical a escala 1:100.000 fue del **76.10%** (Figura 15).

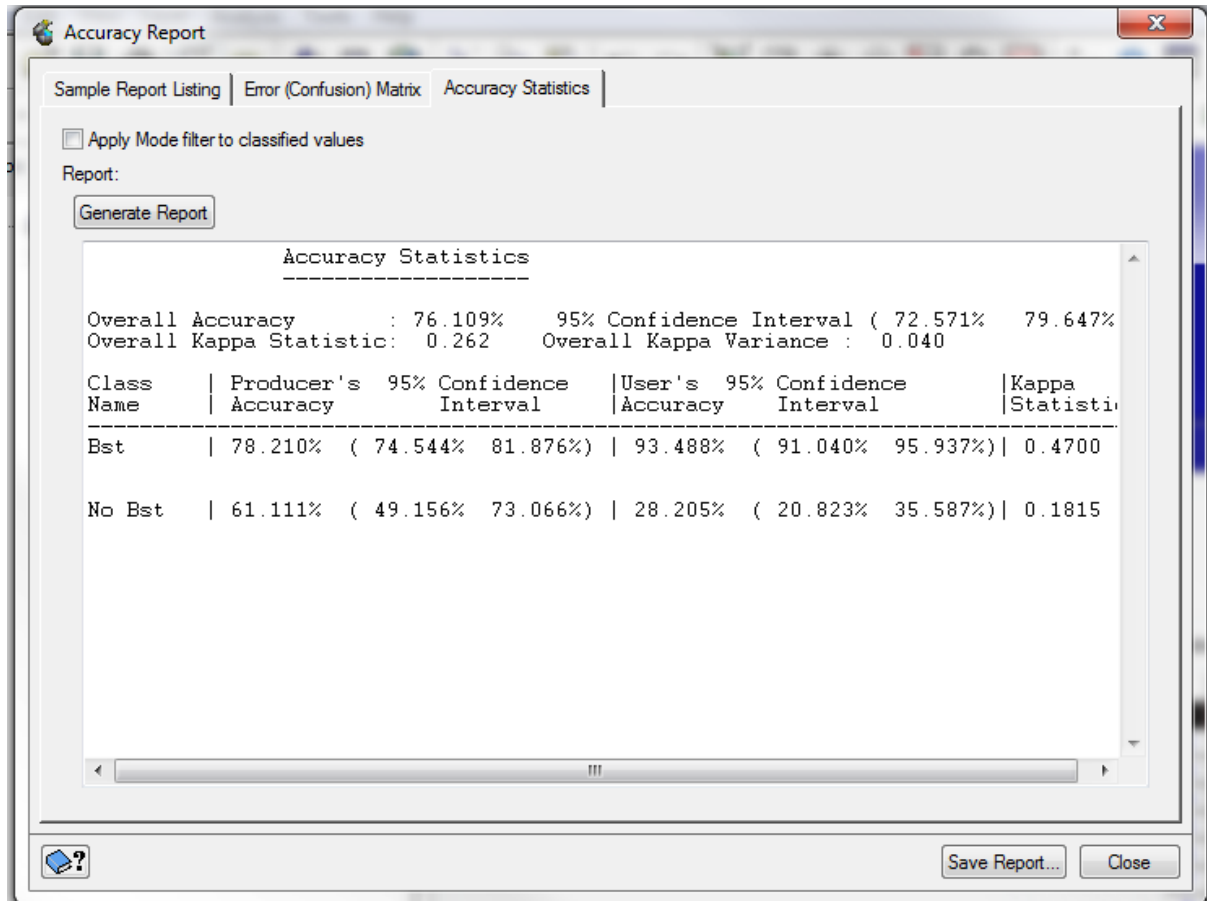


Figura 15. Reporte de exactitud temática obtenido del mapa de bosque seco tropical BST a escala 1:100.000

d. Otras medidas de validación y exactitud

Como el resultado para verificar exactitud temática resultó de la clasificación digital y análisis espacial SIG para obtener un capa binaria con una variable dicotómica (áreas de BST y áreas de no BST), es posible que se requiera un análisis diferente a matriz de confusión simple tal y como se observó anteriormente.

Existen cuatro cruces posibles entre los valores que son BST de la clasificación y los puntos usados como referencia (si en ambas, no en ambas, si-no y no-si).

Adicionalmente, en esta matriz de confusión, fuera de los desajustes entre la imagen clasificada y la información de referencia, también pueden encontrarse problemas debidos a la distinta resolución espacial de las dos fuentes. (Boschetti et al. 2006).

Previendo esto se puede utilizarse como criterio de validación alternativo otras metodologías de evaluación como:

- Utilizar un insumo con área de celdas de un tamaño de 5 a 10 veces superior a la resolución del pixel de baja resolución, y calcular para cada una de ellas la proporción de superficie ocupada por la clase a validar a través de una ecuación de correlación.
- Utilización de la teoría de detección de señales a través de curva ROC (*acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor*).

En el caso de la validación del mapa de Bst 1.100.000 se utilizó además del índice Kappa un análisis de curva ROC.

e. Análisis de curva ROC

El Análisis ROC (*Receiver operating characteristics*) es una metodología desarrollada para analizar un sistema de decisión.

Dicho análisis de curva ROC es una representación gráfica de la sensibilidad frente a (1 – especificidad) para un sistema clasificador binario según se varía el umbral de discriminación.

Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (FPR = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo). ROC también puede significar *Relative Operating Characteristic* (Característica Operativa Relativa) porque es una comparación de dos características operativas (VPR y FPR) según cambiamos el umbral para la decisión (Swets 1996).

De este modo la sensibilidad o Razón de Verdaderos Positivos (VPR), o también razón de éxitos y, recuerdo en recuperación de información, se obtiene a partir de la siguiente ecuación:

$$VPR = VP / V = VP / (VP + FN)$$

Donde, VP representa verdaderos positivos o también éxitos, VN los Verdaderos Negativos o también rechazos correctos, FP los Falsos Positivos o también falsas alarmas o Error tipo I y FN los Falsos Negativos o también, Error de tipo II.

Por lo que el Ratio o Razón de Falsos Positivos (FPR) o también razón de falsas alarmas o *fall-out* en recuperación de información, es igual a:

$$FPR = FP / V = FP / (FP + VN)$$

El análisis de la curva ROC, o simplemente análisis ROC, proporciona herramientas para seleccionar los modelos posiblemente óptimos y descartar modelos subóptimos independientemente de (y antes de especificar) el coste de la distribución de las dos clases sobre las que se decide. La curva ROC es también independiente de la distribución de las clases en la población (en el mapa, la exactitud de la clase Bst clasificada).

El análisis ROC se relaciona de forma directa y natural con el análisis de confiabilidad temática en imágenes clasificadas. De este modo los cuatro posibles resultados se pueden formular en una Tabla de contingencia de 2x2 (Matriz de confusión, Figura 16)

		Valor en la realidad		total
		<i>p</i>	<i>n</i>	
Predicción outcome	<i>p'</i>	Verdaderos Positivos	Falsos Positivos	P'
	<i>n'</i>	Falsos Negativos	Verdaderos Negativos	N'
total		P	N	

Figura 16. Tabla de contingencia derivada del análisis de verificación

Esta matriz de confusión proporcionó las medidas de evaluación. Para dibujar una curva ROC sólo son necesarias las razones de Verdaderos Positivos (VPR) y de falsos positivos (FPR). Razón de Verdaderos Positivos (VPR) mide hasta qué punto el sistema de clasificación de BST ha sido capaz de detectar o clasificar las coberturas correctamente, de entre todos los casos correctos durante la clasificación. La Razón de Falsos Positivos (FPR) define cuántos resultados positivos de coberturas de BST son incorrectos de entre todas las coberturas no relacionadas con BST disponibles durante la clasificación.

Un espacio ROC se define por FPR y VPR como ejes x e y respectivamente, y representa los intercambios entre verdaderos positivos (en principio, exactitud) y falsos positivos (en principio, desacuerdo).

Dado que VPR es equivalente a sensibilidad y FPR es igual a 1- especificidad, el gráfico ROC también es conocido como la representación de sensibilidad frente a (1- especificidad). Cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC (**Figura 17**).

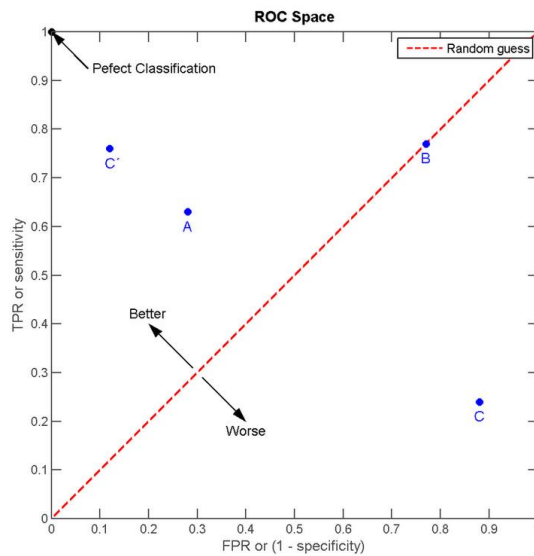


Figura 17. El espacio de la curva ROC y las parcelas de los cuatro ejemplos de predicción A, B, C y C'. (Swets 1996).

La diagonal divide el espacio de la curva ROC. Los puntos por encima de la diagonal representan los buenos resultados de la clasificación (mejor que los resultados obtenidos por azar), los puntos por debajo de la línea son resultados pobres (peor que al azar).

Nótese que la salida de un predictor consistentemente pobre simplemente podría ser invertida para obtener un buen predictor.

El mejor método posible de predicción se situaría en un punto en la esquina superior izquierda, o coordenada (0,1) del espacio ROC, representando un 100% de sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo).

A este punto (0,1) también se le llama una clasificación perfecta. Por el contrario, una clasificación totalmente aleatoria (o azar) daría un punto a lo largo de la línea diagonal, que se llama también línea de no-discriminación, desde el extremo inferior izquierdo hasta la esquina superior derecha (independientemente de los tipos de bases positivas y negativas).

Para realizar el análisis de exactitud y validación a través de la curva ROC, se utilizó el software de procesamiento digital ENVI 5.1. Bajo el menú de herramientas de post-clasificación:

Classification / Post classification / ROC curves / Using Ground Truth ROI's

Dicho análisis se realizó a partir de la transformación de los puntos de verificación de una capa shape a una capa de ROI's que permita obtener las clases de verdad en campo (**Figura 18**).

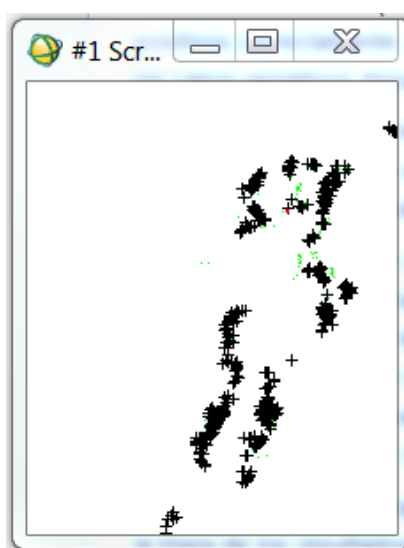


Figura 18. Capa de puntos ROI's generada en ENVI.

Una vez establecido el umbral de discriminación o umbral de decisión y cargado la capa de verificación ROI's, obtenemos los resultados del análisis ROC (**Figura 19**). La grafica obtenida muestra su comportamiento y una medida de su calidad.

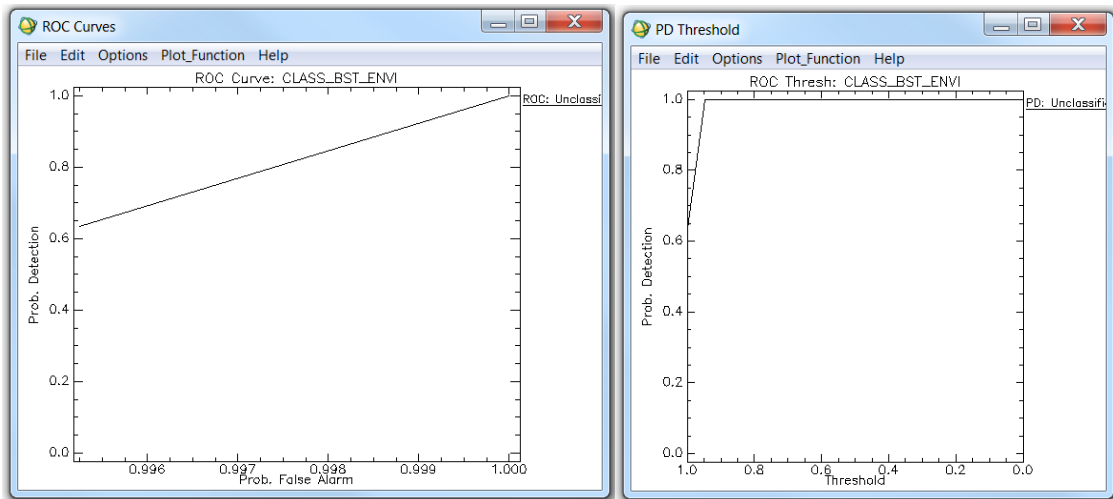


Figura 19. Curva ROC a la izquierda generada a partir de la capa de clasificación Bst y los puntos ROI's de verificación en campo. A la izquierda la gráfica del umbral de decisión.

La curva ROC obtenida se usa para generar estadísticos que resumen el rendimiento (o la efectividad, en su más amplio sentido) del clasificador.

Como se puede observar el punto de inserción de la curva ROC con la línea convexa a la línea de discriminación está por encima de **0.6348**. Esto permite obtener un indicador del rendimiento del sistema clasificador del mapa de Bst 1:100.000 evaluado, correlacionado con el valor del área bajo la curva (AUC). Esto significa que existe un 63.48 % de probabilidad de que el sistema clasificador realizado sobre la imagen correspondiente sea más correcto que el de un clasificador escogido al azar.

En conclusión, los resultados obtenidos del Análisis ROC permiten así establecer una medida de calidad aceptable para el procedimiento de clasificación del mapa de bosque seco tropical a escala 1:100.000.

RESULTADOS

Como resultados generales se logró analizar la información secundaria entregada como insumo logrando determinar y realizar los ajustes necesarios para que dicha información se usará en el diseño y estructura de la base de datos geográfica que contiene la información del mapa de Bst 1:100.000.

Así mismo, fue posible obtener y ajustar las imágenes de satélite y la información temática necesaria para llevar a cabo la validación temática del mapa de BST 1:100.000, para ello se recurrió a procesamiento digital de imágenes y SIG.

Se determinó mediante procedimientos técnicos estadísticos que el mínimo de puntos requerido para realizar la verificación de exactitud temática del mapa de bosque seco tropical a escala 1.100.000 para Colombia era de 196 y finalmente se efectuó el procedimiento con 585 puntos superando ampliamente la necesidad de datos para tener una muestra representativa.

Igualmente, fue posible utilizar un diseño de muestreo *sistemático no alineado* debido a que es uno de los más documentados para cartografía de este tipo ya que la muestra se distribuye de manera regular, pero con un cierto grado de libertad y permite representar todo el territorio cubierto por el mapa Bst 1:100.000.

Para la aplicación de este método se ubicó en forma aleatoria el primer punto de muestreo y, a partir de su ubicación, se determinó la ubicación de todos los demás puntos en forma sistemática pero introduciendo una modificación aleatoria de la ubicación espacial de éstos en

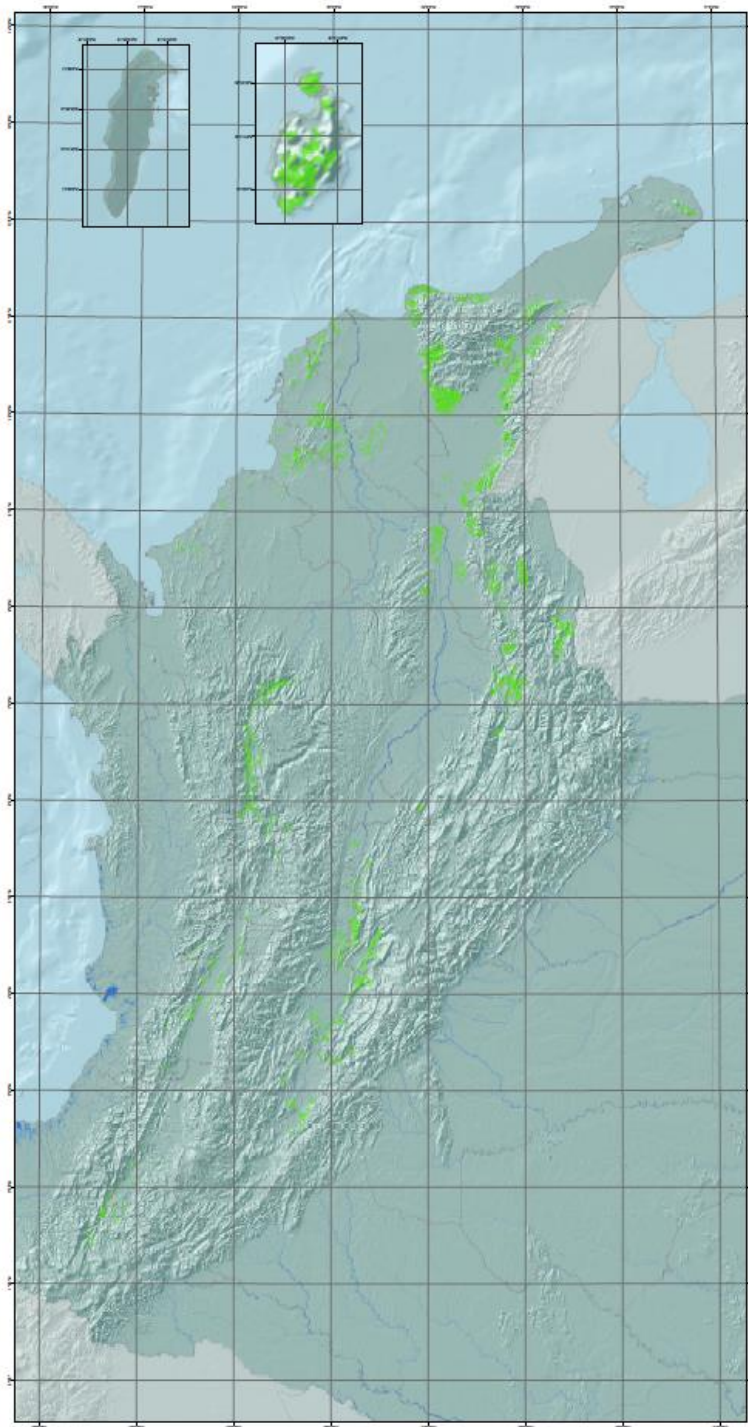
uno o los dos ejes. De esta manera la distancia entre los puntos planteada fue de 500 m, con lo que el rectángulo del área total se dividió en 5 000 cuadros que conténian los puntos de muestreo.



El resultado de la localización de los puntos se cruzó espacialmente mediante técnicas de Sistemas de Información Geográfica contra la capa de BST 1.100.000 para validar con los datos de referencia, incluyendo las fuentes de verdad-terreno, y la extracción de los resultados del análisis digital.

Esto permitió comparar los datos de referencia y los resultados, mediante test estadísticos y la generación de una matriz de confusión.

Finalmente la verificación de exactitud temática del mapa de Bosque Seco tropical para Colombia a escala 1.100.000, realizada permitió cuantificar la fiabilidad de las metodología aplicadas mediante un estadístico de confianza (índice Kappa) que arrojó un resultado de **0.7610** que significa que la clasificación del mapa utilizada mediante esta metodología es un **76.10 %** mejor que la esperada por azar.

Por su parte el análisis de curva ROC, permitió estimar que el punto de inserción de la curva ROC con la línea convexa a la línea de discriminación está por encima de **0.6348**.; esto significa que existe un 63.48 % de probabilidad de que el sistema clasificador realizado sobre la imagen correspondiente sea más correcto que el de un clasificador escogido al azar.



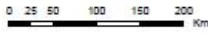
PROSPERIDAD PARA TODOS
Ministerio de Ambiente y Desarrollo Sostenible
Instituto de Investigación de Recursos Biológicos Alexander Von Humboldt

BOSQUE SECO TROPICAL

VERSION 2.0

Escala de trabajo 1:100.000

PORTAFOLIO PARA LA RESTAURACIÓN DEL BOSQUE SECO



Escala 1:2.000.000



CONVENCIONES

Límite departamental
 Drenaje doble

LEYENDA

Bosque Seco Tropical

LOCALIZACIÓN ÁREA DE ESTUDIO

PARÁMETROS CARTOGRÁFICOS

Sistema de coordenadas planas: Gauss-Krueger
 Proyección: Conforme de Gauss
 Datum: MAGNA-SIRGAS
 Origen: Central
 Falso Este: 1.000.000 m
 Falso Norte: 1.000.000 m

[c] Instituto Humboldt, 2014

REFERENCIAS

CAMPBELL, J.B. 1987. Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering and Remote Sensing*, 47: 355-363.

CHUVIECO, E. 1990. *Fundamentos de Teledetección Espacial*. Rialp Ediciones. Madrid. 224p.

CHUVIECO, E. 2010. *Teledetección ambiental: La observación de la tierra desde el espacio*. Editorial Ariel. Madrid. 528p.

CONGALTON, R.G. Y GREEN, K. 1999. *Assesing the acurracy of Remotely Sensed Data: Principles and Applications*. Lewis Publishers. Boca Raton. 200p.

CONGALTON R.G. Y MEAD, R.A. 1983. A quantitative method to test for consistency and correctness in photo-interpretation. *Photogrammetric Engineering and Remote Sensing*, 49(1), 69-74.

CONGALTON, R.G. 1988. A comparison of sampling scheme use in generating error matrices for assessing the accuracy of maps generated from remotley sensed data. *Photogrammetric Engineering and Remote Sensing*, 54(5), 593-600.

CORZO, G. Y DELGADO, J. 2012. *Escenarios Geográficos para la Restauración del Bosque seco en Colombia*. Informe final de consultoría. Universidad ICESI – Instituto de

Investigación de Recursos Biológicos Alexander von Humboldt – Ministerio de Ambiente y Desarrollo Sostenible.

IAvH-Ecopetrol, S.A. 2012. Mapa de Biomas Potenciales escala 1:100.000. Versión 1.0. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.

ETTER, A. 1993. Diversidad ecosistémica en Colombia hoy. Nuestra diversidad biótica. CEREC y Fundación Alejandro Angel Escobar. Bogotá. 43-61pp.

ETTER, A. 1998. El Bosque seco Tropical en Colombia. Instituto Alexander von Humboldt, Programa de Inventario de la Biodiversidad Grupo de Exploraciones y Monitoreo Ambiental GEMA.

ETTER, A., MCALPINE, C. Y POSSINGHAM, H. 2008. A historical analysis of the spatial and temporal drivers of landscape change in Colombia since 1500. *Annals of the American Association of Geographers*, 98(1), 2-23

HARGREAVES, G.H., y ALLEN, R.G., 2003. History and evaluation of Hargreaves evapotranspiration equation. *Journal of Irrigation and Drainage Engineering*, 129(1), 53-63.

IDEAM, IGAC, IAVH, INVEMAR, I. SINCHI E IIAP. 2007. Ecosistemas continentales, costeros y Marinos de Colombia. Instituto de Hidrologia, Meteorologia y Estudios Ambientales, Instituto Geografico Agustin Codazzi, Instituto de Investigacion de Recursos Biologicos Alexander von Humboldt, Instituto de Investigaciones Ambientales del Pacifico

Jhon von Neumann, Instituto de Investigaciones Marinas y Costeras Jose Benito Vives De Andreis e Instituto Amazonico de Investigaciones Cientificas Sinchi. Bogota. 276p. 37 hojas cartográficas.

ICDE (Infraestructura Colombiana de Datos espaciales). 2013. Cartografía básica a escala 1:100.000. Servicios geográficos en línea. Recuperado de: http://geocarto.igac.gov.co:8082/geoservicios/cien_mil/wfs

IGAC - Instituto Geográfico Agustín Codazzi. 2014. Información cartográfica de Colombia. Recuperado de: <http://www.igac.gov.co/igac>.

ICONTEC - Instituto Colombiano de Normas Técnicas y Certificación. 2007. Información geográfica. Metadato geográfico. ICONTEC. Bogotá. NTC 4611.

ICONTEC - Instituto Colombiano de Normas Técnicas y Certificación. 2002. Información geográfica. Conceptos básicos de calidad de los datos geográficos. ICONTEC. Bogotá. NTC 5043.

ICONTEC - Instituto Colombiano de Normas Técnicas y Certificación. 2009. Información geográfica. Información geográfica. Especificaciones técnicas de productos geográficos. ICONTEC. Bogotá. NTC 5662.

ISO - International Organization for Standardization. 2014. Normas ISO para Colombia. Recuperado de: <http://www.iso.org>

IDEAM. 2010. Leyenda Nacional de Coberturas de la Tierra. Metodología CORINE Land Cover adaptada para Colombia Escala 1:100.000. Instituto de Hidrología, Meteorología y Estudios Ambientales. Bogotá. 72p.

MILES, L., A. C. NEWTON, R. S. DEFRIES, C. RAVILIOUS, I. MAY, S. BLYTH, V. KAPOV Y GORDON. J. E. 2006. A global overview of the conservation status of tropical dry forests. *Journal of Biogeography*, 33, 491–505.

UN - United Nations. 2011. Global Drylands: a UN system-wide response. Environmental Management Group. New York. 131 p.

UM - University of Maryland. 2014. Earth Science Data Interface. Recuperado de: <http://glcfapp.glc.f.umd.edu:8080/esdi/> (22/04/2014).

US - Department of the Interior United States. 2014. Geological Survey. <http://earthexplorer.usgs.gov/> (22/04/2014).

Swets, J.A. 1996. Signal detection theory and ROC analysis in psychology and diagnostics: collected papers. Psychology Press. Colorado. 328 p.

ZARAZAGA-SORIA, F.J., J. LACASTA, J. NOGUERAS-ISO, M.P. TORRES-BRUNA Y MURO-MEDRANO, P.R.. 2003. A Java Tool for Creating ISO/FGDC Geographic Metadata. Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung. Beiträge zu den Münsteraner GI-Tagen. IfGI prints. 17-30 p.