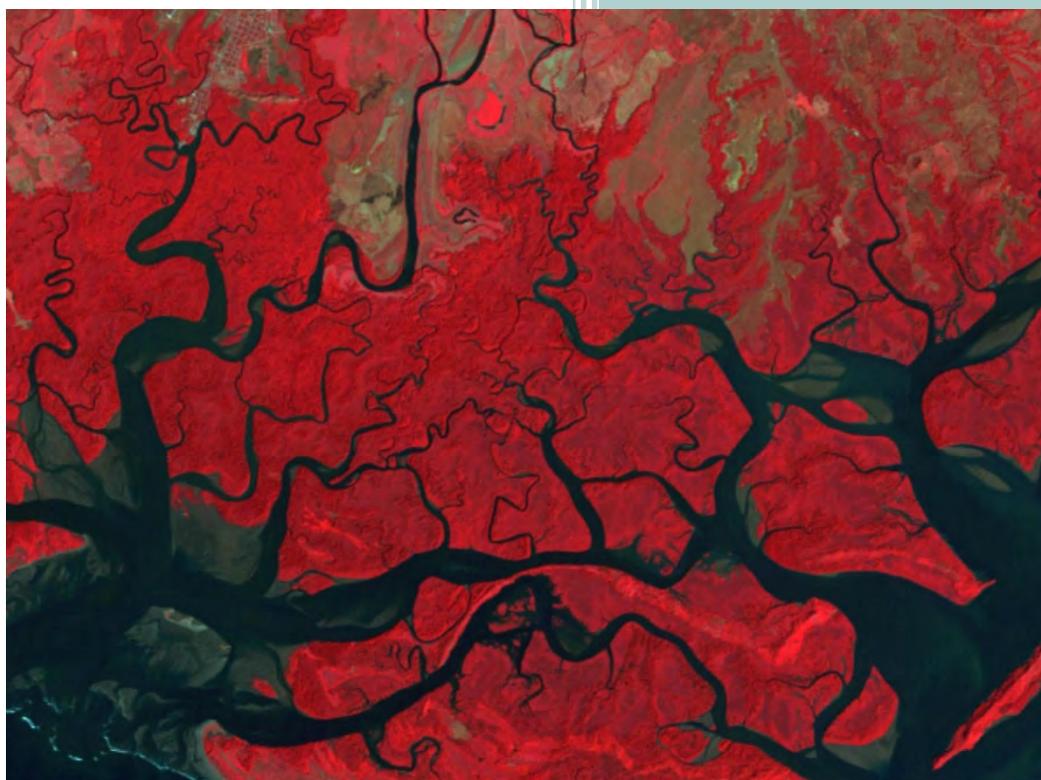


**GENERATING A CONSISTENT HISTORICAL
TIME SERIES OF ACTIVITY DATA FROM LAND
USE CHANGE FOR THE DEVELOPMENT OF
COSTA RICA'S REDD PLUS REFERENCE LEVEL**

Protocolo Metodológico



agresta



ÍNDICE

1. INTRODUCCIÓN.....	3
1.1. OBJETIVO GENERAL	3
1.2. OBJETIVOS ESPECÍFICOS	3
1.3. RESUMEN DE LA METODOLOGÍA.....	4
1.4. MARCO INSTITUCIONAL Y PROCESO DE DESARROLLO DE LA METODOLOGÍA	6
2. DESARROLLO DEL PROTOCOLO	8
2.1. CONSIDERACIONES CONCEPTUALES Y METODOLÓGICAS INICIALES	8
2.1.1. Compatibilidad y consistencia de la metodología con los marcos metodológicos de referencia	8
2.1.2. Resolución espacial y selección de sensores	8
2.1.3. Selección y recopilación de imágenes satelitales	9
2.1.4. Asignación de fecha de referencia a cada mapa de la serie.....	10
2.1.5. Definición operativa de bosques y leyenda temática.....	11
2.2. SOFTWARE PROPUESTO	13
2.3. PRE-PROCESAMIENTO	14
2.3.1. Correcciones geométricas	14
Preparación de imágenes.....	15
Toma de puntos de control terrestre y de enlace entre imágenes	15
Control de calidad visual y geométrica de las imágenes corregidas	16
2.3.2. Corrección radiométrica y atmosférica.....	16
Corrección radiométrica y atmosférica en Landsat 4, 5 y 7	17
Corrección radiométrica y atmosférica en Landsat 8	18
2.3.3. Normalización horaria	19
2.3.4. Normalización radiométrica de la serie	19
2.3.5. Detección de vacíos de información	20
Detección de nubes.....	21
Detección de sombras	22
Vacíos por fallos del sensor.....	23
Control de calidad de las máscaras de nubes y sombras	23
2.4. CÁLCULO DE ÍNDICES DE VEGETACIÓN Y TEXTURAS	24
2.4.1. Índice de vegetación de diferencia normalizada (NDVI).....	24
2.4.2. Índices de textura	25
2.5. CÁLCULO DE DERIVADAS DEL MODELO DIGITAL DE ELEVACIONES	25

2.6.	GENERACIÓN DE IMÁGENES MULTIBANDA	25
2.7.	CLASIFICACIÓN DE IMÁGENES.....	26
2.7.1.	Regiones de interés y puntos de control.....	26
2.7.2.	Variables predictoras	27
2.7.3.	Ajuste de clasificadores.....	28
2.7.4.	Clasificación de imágenes	29
2.8.	GENERACIÓN DE MAPAS DE COBERTURA	29
2.8.1.	Mosaicado de las imágenes clasificadas.....	29
2.8.2.	Relleno de vacíos con productos globales	29
2.8.3.	Asignación de clases de edad de bosque.....	31
2.8.4.	Aplicación de máscaras.....	33
2.8.5.	Generación de mapas finales.....	34
2.9.	REPORTE DE TRAYECTORIAS DE CAMBIO DE LA COBERTURA DEL SUELO	34
2.10.	ÍNDICE DE COBERTURA COMO BASE PARA LA ESTIMACIÓN DE LA DEGRADACIÓN Y AUMENTO DE EXISTENCIAS DE CARBONO	35
3.	VALIDACIÓN.....	36
3.1.	VALIDACIÓN DE LOS MAPAS DE COBERTURA	36
3.2.	VALIDACIÓN DE MAPAS DE CAMBIO.....	39
3.3.	ASEGURAMIENTO Y CONTROL DE LA CALIDAD	41
4.	RECOMENDACIONES	43
5.	REFERENCIAS BIBLIOGRÁFICAS.....	44

1. INTRODUCCIÓN

En los últimos 30 años diferentes instituciones han trabajado en la generación de mapas de coberturas de la tierra en Costa Rica. Se pueden destacar los elaborados por el Instituto Meteorológico Nacional (IMN) correspondientes a los años 1992, 1996, 2000, 2005 y 2010; los realizados desde FONAFIFO en los años 1997, 2000, 2005 y 2010, o más recientemente el mapa de tipos de bosque de Costa Rica del año 2013 realizado por el Sistema Nacional de Áreas de Conservación (SINAC) para el Inventario Nacional Forestal. Estos trabajos responden a objetivos y fines diferentes y han sido elaborados con distintas metodologías, leyendas e imágenes satelitales.

La estrategia REDD+ de Costa Rica necesita definir los niveles de referencia nacional de emisiones dentro de diferentes periodos, requerimientos y marcos metodológicos (UNFCCC, IPCC, FCPF CF MF, and VCS JNR¹). Para ello es necesario elaborar una serie histórica de mapas de cobertura de la tierra aplicando una metodología consistente que asegure que los cambios en las coberturas no se deben a cambios metodológicos sino únicamente a cambios reales producidos en las coberturas dentro de los periodos considerados.

El análisis conjunto de la estrategia REDD+ del país, de la evolución de las políticas ambientales y de conservación de Costa Rica así como de los diferentes marcos metodológicos justifica la necesidad de construir una serie histórica que permita analizar los cambios en la cobertura de la tierra entre los años 1986 y 2014 con mapas en al menos 7 puntos en el tiempo.

El presente documento incluye tanto los **aspectos metodológicos como los resultados obtenidos** en la elaboración de esta serie histórica de datos de actividad de uso del suelo y de cambio de uso del suelo entre los años 1986 y 2014.

1.1. OBJETIVO GENERAL

El objetivo general del presente protocolo es establecer, describir y validar el proceso metodológico necesario para generar datos de actividad consistentes de cambio de uso del suelo en Costa Rica.

1.2. OBJETIVOS ESPECÍFICOS

Como objetivos específicos del presente documento cabe destacar:

1. Describir y validar la metodología desarrollada para la generación de la serie histórica de datos de actividad de Costa Rica entre los años 1986 y 2014.

¹ United Nations Framework Convention on Climate Change (UNFCCC), Intergovernmental Panel on Climate Change (IPCC), The Forest Carbon Partnership Facility (FCPF), Carbon Fund Methodological Framework (CF MF), Verified Carbon Standard (VCS), Jurisdictional and Nested REDD+ (JNR)

2. Establecer los aspectos metodológicos y recomendaciones necesarias para la generación de datos de actividad consistentes con la serie histórica en el futuro proceso de Monitoreo, Reporte y Verificación (MRV) de REDD+ en el país.
3. Elaborar una guía práctica que facilite a los técnicos del país replicar todo el proceso de generación de mapas y datos de actividad consistentes con la serie histórica en el futuro.
4. Generar las bases para la medición y el monitoreo de la degradación a partir de una primera experiencia piloto de estimación de la degradación de forma espacialmente explícita en el país.

1.3. RESUMEN DE LA METODOLOGÍA

La homogeneidad de las series temporales de imágenes satelitales es crucial cuando se estudian cambios graduales en las coberturas de la tierra. Existen varias fuentes de error que afectan a la información recibida por los satélites, haciendo difícil diferenciar la información recibida por la superficie de estudio de lo que es ruido, dificultando la obtención de series temporales homogéneas. Por ello es fundamental seguir una metodología única y uniforme en la elaboración de cada uno de los mapas de la serie para asegurar que los cambios observados en la cobertura de la tierra no son debidos a la aplicación de diferentes metodologías a lo largo de la serie histórica.

Existen dos métodos fundamentales de detección de cambios de coberturas a partir de imagen satélite: i) detección directa de los cambios o ii) detección de cambios de coberturas post-clasificación de las coberturas. En este caso se ha desarrollado una metodología de **detección de cambios post-clasificación** que permite en un primer paso generar mapas de coberturas para cada fecha (con una leyenda amplia) para detectar en un segundo paso los cambios por medio de la comparación de mapas de fechas diferentes.

Para poder realizar la clasificación de coberturas de la tierra en cada uno de estos mapas individuales se necesita **entrenar (o ajustar) un modelo estadístico**. Este modelo nos permitirá estimar la cobertura de la tierra en cualquier punto del terreno (i.e. elaborar un mapa de cobertura de la tierra) a partir de otro tipo de información conocida. En otras palabras, para entrenar el modelo se necesita un **conjunto de puntos de control (o verdad terreno)** de los que se conozca la cobertura de la tierra, además de contar con información espacialmente continua de otras **variables predictoras o covariables** (p.ej. imágenes satelitales, información de clima, relieve, etc.). Así, el ajuste del modelo estadístico nos permite analizar la relación entre el uso de la tierra en esos puntos de control y las covariables que se usan para el ajuste. Posteriormente se usa ese modelo ajustado para estimar, a partir de los datos de las variables predictoras de los que se dispone, la cobertura de la tierra que es más probable que ocurra en el resto de puntos de los que no se cuenta con información. Finalmente, se usa otro conjunto de **puntos de control para validar el modelo** propuesto y estimar el error de las estimaciones realizadas. El esquema general de la metodología puede observarse de forma gráfica en la Figura 1.

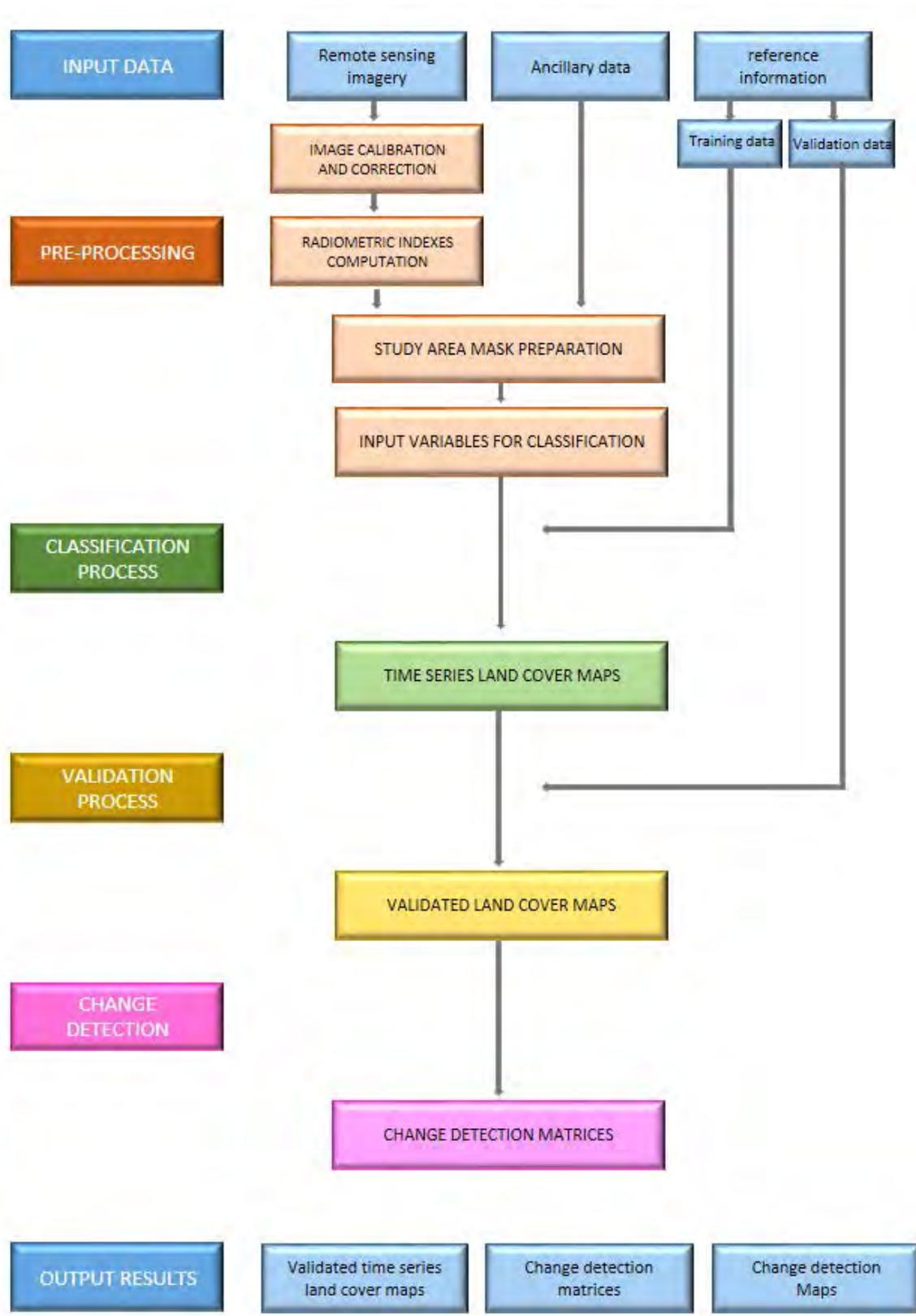


FIGURA 1. Diagrama representativo del esquema general de la metodología propuesta

Para la realización del análisis de una serie histórica caben dos enfoques posibles:

1. Ajustar un modelo diferente para la estimación de la cobertura de la tierra en cada uno de los años de la serie.
2. Ajustar un modelo y usarlo para la estimación de las coberturas de todos los años objeto del análisis.

Debido a que se dispone de una base de datos mucho más completa de puntos de control en los años más recientes y debido también al objetivo fundamental de consistencia de la serie temporal, se ha seguido este segundo enfoque. Para el correcto funcionamiento del proceso de clasificación en este segundo enfoque ha sido necesario prestar especial atención a las correcciones y normalizaciones radiométricas de todas las imágenes utilizadas. Una correcta calibración y normalización radiométrica nos permite comparar la radiometría de todas las escenas y por tanto usar con buenos resultados un clasificador único para todos los años de la serie.

1.4. MARCO INSTITUCIONAL Y PROCESO DE DESARROLLO DE LA METODOLOGÍA

En el proceso de desarrollo de una metodología consistente para la elaboración de la serie histórica de datos de actividad de uso del suelo en Costa Rica han estado presentes por medio del Comité de Sensores Remotos todos los actores implicados en la estrategia REDD del país.

Entre los días 6 y 18 de agosto de 2014 el consorcio liderado por Agresta S. Coop. y Carbon Decisions International SA (CDI) organizaron una misión conjunta a Costa Rica para iniciar sus respectivas consultorías con el auspicio de FONAFIFO. Dos de los principales objetivos de la misión fueron:

1. **Definir las fechas clave, periodos, actividades y marcos metodológicos** a seguir para la definición del nivel de referencia REED+ en Costa Rica. La definición de estos aspectos es imprescindible para redactar una propuesta metodológica y comenzar los trabajos de generación de cartografía histórica de coberturas de la tierra en el país.
2. **Recopilar datos e información necesarios para la ejecución de las consultorías.** La información disponible para el desarrollo de los trabajos va a condicionar la metodología planteada tanto para generar la serie cartográfica histórica como para establecer los niveles de referencia.

Las reuniones realizadas con las diferentes instituciones (FONAFIFO, SINAC, IGN, MAG-INTA, IMN e ICE) durante la misión permitieron tomar las decisiones más relevantes para la correcta definición la serie histórica, establecer la metodología de trabajo y detectar las fuentes de información disponibles en el país de utilidad para los procesos de pre-procesamiento, clasificación de las imágenes satelitales y de validación de la cartografía.

Posteriormente a la misión en una reunión realizada en FONAFIFO el lunes 25 de Agosto de 2014 fueron analizadas en mayor profundidad y confirmadas las políticas más relevantes del país en materia de conservación de bosques, reducción de las emisiones debidas a la deforestación y degradación de bosques, y cambio climático. En esta reunión participaron Ricardo Ulate (Asesor del Ministro), Alexandra Sáenz (Punto Focal REDD+ de FONAFIFO), Maria Helena Herrera y Edwin Vega (FONAFIFO), Javier Fernández (Asesor Secretaría REDD+), Lucio Pedroni (CDI), el consultor en salvaguardas ambientales Álvaro Aguilar (MINA, Centro Nacional

Información Geoambiental – CENIGA) y la consultora en salvaguardas sociales Vera Luz Salazar.

Tras la misión de agosto de 2014 se elaboró una primera propuesta metodológica completa y un estudio piloto sobre la pasada 14 de Landsat en Costa Rica que se remitió a FONAFIFO y a los miembros del Comité en Sensores Remotos para su evaluación. Los principales procesos metodológicos seguidos, la leyenda completa a implementar y las fechas de trabajo fueron descritos en este documento que fue revisado, mejorado y completado gracias a las aportaciones de los técnicos miembros del Comité de Sensores Remotos.

El día 25 de abril se realizó la presentación final de la serie histórica de datos de actividad en una reunión realizada en FONAFIFO y a la que asistieron los principales actores en el marco de la estrategia REDD+. Tras esta jornada técnica se tomaron algunas decisiones finales como la generación y aplicación de máscaras de áreas potenciales de manglares y yolillales que permitan mejorar la clasificación en estas subclases de bosque, o la definición de una superficie mínima de mapeo de 1 ha. Los comentarios, decisiones y recomendaciones realizadas sobre el trabajo se han incorporado al protocolo metodológico.

2. DESARROLLO DEL PROTOCOLO

2.1. CONSIDERACIONES CONCEPTUALES Y METODOLÓGICAS INICIALES

2.1.1. Compatibilidad y consistencia de la metodología con los marcos metodológicos de referencia

La metodología descrita en el presente documento se ha realizado tomando como referencia los distintos marcos metodológicos disponibles en el sector REDD+. En el **Anexo 1** se resumen las fuentes y marcos metodológicos de referencia considerados en el desarrollo de la presente metodología.

2.1.2. Resolución espacial y selección de sensores

Para la elaboración de la serie histórica se ha trabajado con imágenes provenientes de diferentes satélites de la generación LANDSAT. La serie de satélites LANDSAT comenzó con la puesta en órbita del LANDSAT-1 en el año 1972. Los satélites que suceden a este primer lanzamiento han permitido disponer de la serie más larga existente hasta la fecha de imágenes satelitales de observación terrestre, esto hace que sea la información más adecuada para construir una serie histórica consistente de coberturas de la tierra en Costa Rica desde el año 1986 hasta la actualidad.

Para construir la serie histórica completa se han utilizado imágenes provenientes de cuatro sensores y satélites diferentes de la familia Landsat (Landsat 4 TM, Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI/TIRS) tal y como se muestra en la Figura 2.

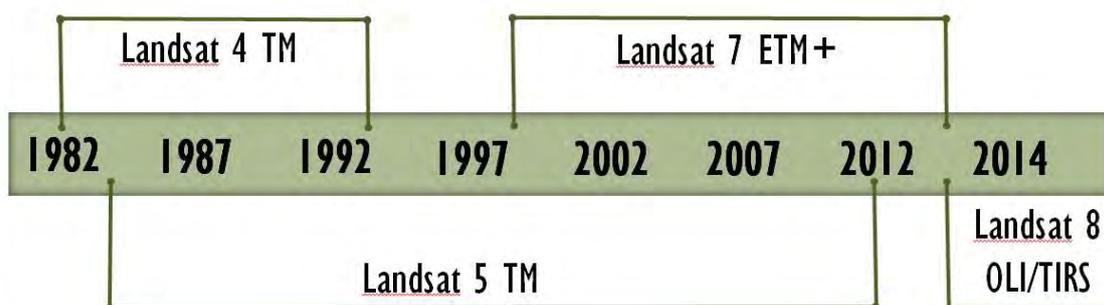


FIGURA 2. Diferentes sensores y satélites utilizados para la realización de la serie temporal

Debido a que se ha trabajado con diferentes sensores (TM, ETM+ y OLI/TIRS) se han utilizado las seis bandas comunes en todos ellos de 30 metros de resolución para generar información espectral homogénea en todos los años de la serie. Se han utilizado las bandas 1,2,3,4,5 y 7 para Landsat 7 (ETM+) y Landsat 4 y 5 (TM) y las bandas 2,3,4,5,6 y 7 para Landsat 8 (OLI/TIRS).

Uno de los principales inconvenientes con los que nos encontramos para construir la serie temporal es que no hay prácticamente ninguna escena sin nubes en todo el ámbito temporal del proyecto (1986-2014). Para minimizar este problema se han combinado varias escenas de un mismo año (en la misma estación) o incluso de distintos años² (en una misma estación) para conseguir, en la medida de lo posible, mapas libres de nubes, siempre cumpliendo con las restricciones temporales de adquisición de imágenes que imponen los marcos metodológicos que se utilizarán posteriormente. Se ha trabajado con escenas de la época seca (de diciembre a abril) para toda la extensión del país. Durante el proceso de clasificación se detectaron problemas de confusión de bosque decíduo con otras clases (principalmente formaciones herbáceas y plantaciones forestales) en las zonas pobladas con bosque decíduo (principalmente en la península de Nicoya y Norte de Guanacaste). Para mejorar la clasificación en esta zona se seleccionaron las mejores imágenes de la época de lluvias de todas las fechas de la serie que fueron clasificadas de forma separada al resto de las imágenes.

2.1.3. Selección y recopilación de imágenes satelitales

Las imágenes se han descargado a través del servidor EarthExplorer del USGS. Para cubrir de forma completa la superficie de Costa Rica. Así, en cada uno de los años de la serie es necesario trabajar con 7 escenas LANDSAT. La identificación de las imágenes según su path-row se muestra en la Figura 3.

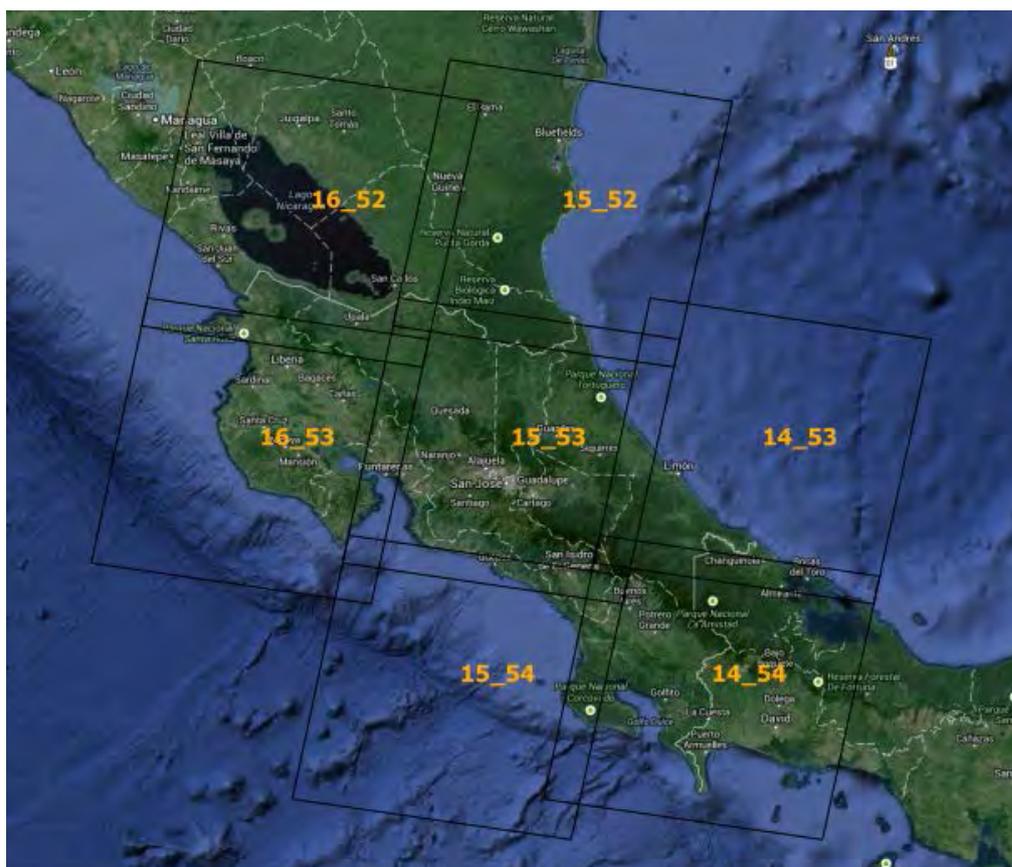


FIGURA 3. Imágenes LANDSAT necesarias para cubrir el país

² Las imágenes podrán ser de diferentes años pero siempre dentro de una ventana de tiempo no superior a 14 meses para cumplir con los requerimientos de VCS

Una vez descargadas las imágenes y validada su calidad, se unieron las escenas de una misma pasada y de una misma fecha. Esto permite reducir el número de imágenes de 7 a 3 necesarias para cubrir todo el país y agilizar el procesamiento de las mismas. Así, para cubrir el territorio continental de Costa Rica es necesario trabajar con dos escenas de pasada 14 (14-53, 14-54), 3 escenas de la pasada 15 (15-52, 15-53, 15-54) y 2 escenas de la pasada 16 (16-52, 16-53).

La selección de imágenes a partir de las fechas de la serie se ha realizado según lo previsto por las distintas metodologías aplicables (p.ej. hay un rango de 14 meses permisible según VCS-JNR³). Debido a la mala calidad de las imágenes Landsat disponibles para las fechas 2002/2003 se ha trabajado con imágenes correspondientes a los años 2000/2001, variando ligeramente la fecha originalmente definida. De igual forma las imágenes correspondientes a la fecha definida como 2010 pertenecen principalmente los años 2011 y 2012.

En el **Anexo 2** se encuentra la **lista completa de escenas descargadas y procesadas** con el número de identificación (Landsat-ID) que permite su fácil localización en los servidores del USGS. El listado incluye tanto las imágenes de la época seca seleccionadas en las 3 pasadas como las imágenes de la época de lluvias seleccionadas únicamente para la pasada 16.

2.1.4. Asignación de fecha de referencia a cada mapa de la serie

Para asignar una fecha de referencia a cada uno de los mapas en primer lugar se ha estimado la fecha media de las imágenes utilizadas para la generación de cada uno de los mapas de la serie. La fecha de referencia asignada a cada uno de los mapas a partir de la fecha media de las imágenes utilizadas en cada mapa se ha realizado mediante la siguiente regla:

1. Si la fecha media del mapa es anterior al 30 de junio (ej. 30.04.1991) la fecha de referencia para el mapa sería el 31 de diciembre del año anterior y el 01 de enero del año (por ejemplo: 31.12.1990 y 01.01.1991)
2. Si la fecha media del mapa es posterior al 30 de junio (ej. 15.07.1991) la fecha de referencia para el mapa será el 31 de diciembre del año y el 01 de enero del año posterior (ej.: 31.12.1991 y 01.01.1992).

La siguiente tabla resume el rango de las fechas de las imágenes utilizadas para la generación de los mapas de la serie, la fecha media calculada para cada uno de los mapas y sus fechas de referencia calculadas con la regla definida.

TABLA 1. Rango de fechas de las imágenes utilizadas en cada mapa, fechas medias, fechas de referencia y denominación asignada a cada uno de los mapas de la serie histórica

Rango de fechas de las imágenes	Fecha media	Fechas de referencia	Denominación del mapa
De enero del 86 a marzo del 87	Abril de 1986	31.12.1985 y 01.01.1986	1986
De marzo del 91 a marzo del 92	Noviembre de 1991	31.12.1991 y 01.01.1992	1992
De enero del 97 a marzo del 98	Diciembre de 1997	31.12.1997 y 01.01.1998	1998

³ Según se indica en el párrafo 3.11.8 5 del JNR se puede combinar información de 14 meses para generar un único mapa cuando hay problemas de vacíos de información (p.ej. debidos a nubes y sombras)

Rango de fechas de las imágenes	Fecha media	Fechas de referencia	Denominación del mapa
De enero del 2000 a marzo del 2001	Noviembre del 2000	31.12.2000 y 01.01.2001	2001
De enero del 2008 a marzo del 2009	Marzo de 2008	31.12.2007 y 01.01.2008	2008
De enero de 2011 a marzo del 2012	Diciembre de 2011	31.12.2011 y 01.01.2012	2012
De febrero de 2013 a abril de 2014	Enero de 2014	31.12.2013 y 01.01.2014	2014

Los períodos se denominan de acuerdo a las fechas de los mapas de la siguiente forma: 1986-1991; 1992-1997; 1998-2000; 2001-2007; 2008-2011; 2012-2013. De esta manera queda claro en la denominación de los periodos que el año de inicio y de fin del período siempre están incluidos en el periodo mencionado y que los años son enteros.

2.1.5. Definición operativa de bosques y leyenda temática

La definición de bosque debe cumplir los requerimientos de los distintos marcos metodológicos de referencia y en concreto debe ser consecuente con la definición adoptada en el Inventario Nacional de GEI y en caso contrario estar justificadas las discrepancias.

En las reuniones realizadas con los diferentes actores del país se consideró adecuado utilizar la definición de bosque dada para el MDL en Costa Rica, esto es, una **superficie mínima de tierras de una hectárea (1ha)** con una **cubierta de copas que excede el 30%** y con árboles que alcanzan **una altura mínima de 5 metros a madurez *in situ* sin incluir palmas y bambús⁴**.

Una vez revisadas las comunicaciones nacionales de Costa Rica para la CMNUCC⁵ no se ha encontrado expresada de forma explícita una definición de bosque para el país, así que se confirmó por parte del Gobierno de Costa Rica que la definición adoptada para el MDL es también la definición que se utilizará en la actualización de los inventarios nacionales de GEI y en el REL/RL⁶ de REDD+ en Costa Rica.

Según las reuniones que se mantuvieron en agosto de 2014, en las que se acordaron los años de la serie histórica y la leyenda temática para cada uno de esos años, para los años 1986, 1992 y 2001 sólo ha sido necesario generar mapas de bosque/no bosque. Sin embargo, para los años 1998, 2008, 2012 y 2014, la clasificación de las coberturas de la tierra ha incluido un mayor número de clases (en línea con las directrices 2006 del IPCC), como se muestra a continuación:

⁴ Ver: <https://cdm.unfccc.int/DNA/index.html>

⁵ Se han consultado la primera comunicación nacional del año 2000 y la segunda comunicación Nacional del año 2009

⁶ Reference Emission Level (REL); Forest Reference Level (RL)

- **Forest Land:** esta clase tiene las siguientes sub-clases:
 - Bosques primarios⁷
 - Manglares primarios⁷
 - Yolillales primarios⁷
 - Bosques nuevos^{7,8} nativos (bosques secundarios, clasificados en rangos de edad a través del análisis multi-temporal de imágenes)
 - Manglares nuevos^{7,8} (clasificados en rangos de edad a través del análisis multi-temporal de imágenes)
 - Yolillales nuevos^{7,8} (clasificados en rangos de edad a través del análisis multi-temporal de imágenes)
 - Bosques nuevos antrópicos^{7,8} (plantaciones forestales, clasificadas en rangos de edad a través del análisis multi-temporal de imágenes)
- **Cropland:** esta clase tiene las siguiente sub-clases:
 - Cultivos anuales⁹
 - Cultivos permanentes¹⁰
- **Grassland:** esta clase tiene las siguiente sub-clases:
 - Pastizales
 - Páramos
- **Settlements.**
- **Wetlands:** esta clase incluye cuerpos de agua¹¹: cursos de agua, lagos, lagunas, embalses...
- **Other Land:** esta clase incluye los suelos desnudos.

⁷ Aunque hemos llamado "primario" en realidad se refiere (de acuerdo a IPCC) a "forest land remaining forest land", áreas que fueron bosque por 20 años (o el periodo que se defina en el país). En este caso se considera como "primario" lo que era bosque en el mapa inicial de la serie (1986). Bosques secundarios serán el resto, según IPCC "lands converted to forest lands"

⁸ El análisis multi-temporal permite estimar clases de edad de bosque contando años transcurridos a partir de la aparición del bosque en la serie

⁹ Dentro de la subclase cultivos anuales se han separado espectralmente la piña del resto de cultivos anuales

¹⁰ Dentro de la subclase cultivos permanentes se han separado espectralmente el café del resto de cultivos permanentes

¹¹ Se ha diferenciado entre cuerpos de agua naturales y cuerpos de agua artificiales al contar con la información necesaria para ello. La deforestación y generación de bosques nuevos causada por el cambio natural de los cursos de los ríos no es antropogénica y por lo tanto no se puede contabilizar en el nivel de referencia

2.2. SOFTWARE PROPUESTO

Se ha priorizado la utilización de software libre frente a software propietario. Este hecho facilita a las diferentes instituciones que participan en la estrategia REDD+ en Costa Rica el acceso al software sin coste tanto en la capacitación como en la posterior proceso MRV del país.

QGIS ha sido el Sistema de Información Geográfica (SIG) de referencia en todo el proceso de generación de la serie histórica. QGIS es un SIG de código abierto, el proyecto QGIS es relativamente nuevo, nació en mayo de 2002, pero posee una importante comunidad de desarrolladores que han conseguido generar un software ligero, potente y con una interfaz gráfica de usuario (GUI) agradable y fácil de usar. QGIS actualmente funciona en la mayoría de plataformas Unix, Windows y OS X. QGIS se desarrolla usando el kit de herramientas Qt (<http://qt.digia.com>) y C++.

Se ha seleccionado este software como SIG de referencia porque además de ser libre y gratuito está al nivel de otros programas propietarios comúnmente utilizados.

QGIS, gracias a las herramientas de su módulo de procesado, permite comunicarse con otros programas también libres de procesamiento de imágenes o de análisis estadístico de información. Este módulo es un entorno de geoprosesamiento que permite llamar sin salir de QGIS a algoritmos nativos o de terceros, haciendo la tarea de análisis espacial mucho más sencilla. Formalmente este módulo es la herramienta conocida como SEXTANTE disponible para varios GIS tanto comerciales como libres. En las últimas versiones de QGIS, este módulo ha entrado dentro del núcleo del programa pasando a denominarse módulo de procesado ("processing" en inglés). Esta herramienta nos permite trabajar con algoritmos externos a QGIS, como los de GRASS, SAGA u Orfeo Tool Box (OTB) usando el propio QGIS como interfaz gráfica.

El módulo de procesado de QGIS incluye un modelador gráfico que permite combinar varios algoritmos para definir un flujo de trabajo, creando un proceso individual que involucre varios subprocesos (Figura 4). Es de gran utilidad en la automatización de procesos el interfaz de procesamiento por lote permite que ejecutar procesos de un solo algoritmo a múltiples conjuntos de datos.

Durante el desarrollo de la serie histórica se han generado una serie de herramientas dentro de este módulo de procesado de QGIS que están adaptadas a las necesidades del proyecto y que además van a facilitar también la generación de mapas de actividad de uso del suelo en el futuro. Este paquete de herramientas se ha denominado REDD tools Costa Rica.

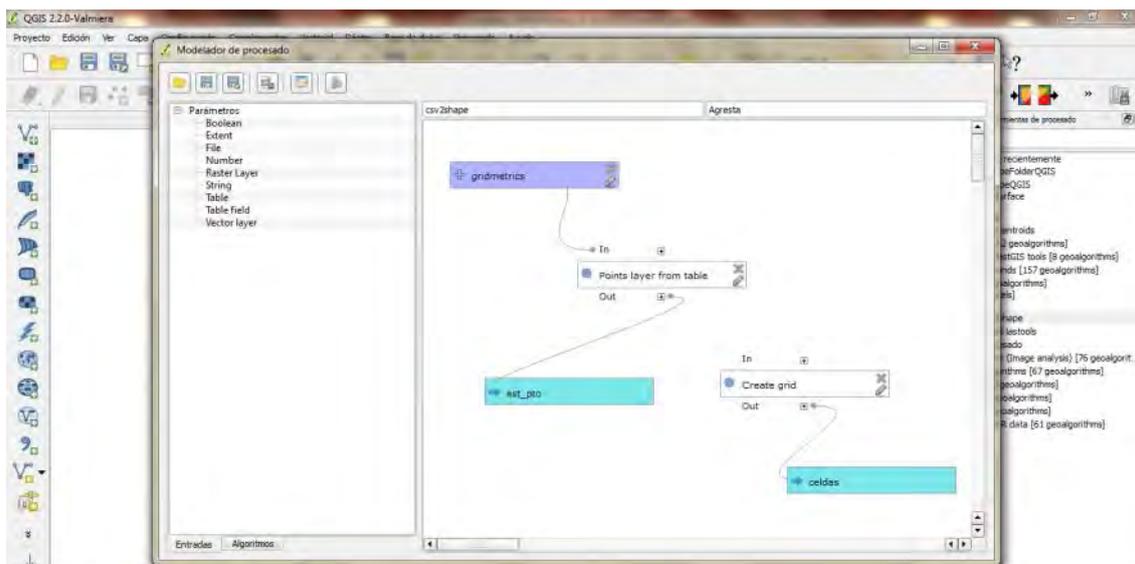


FIGURA 4. Modelador gráfico del módulo de procesado de QGIS

R es un software libre que incluye un lenguaje de programación, un potente entorno de análisis estadísticos y un software gráfico. Este entorno se ha utilizado en la fase de clasificación de imágenes y de estimación del índice de cobertura de copas gracias al uso del paquete “*RandomForest: Breiman and Cutler's random forests for classification and regression*” (Liaw y Wiener 2002) y siguiendo la guía del American Museum of Natural History (AMNH) propuesta por Horning (2013).

El módulo de procesado de QGIS permite ejecutar scripts de R desde QGIS por lo que durante el proceso de generación de la serie histórica se ha integrado el algoritmo de clasificación a la barra de procesado de QGIS facilitando el proceso de clasificación de imágenes con Random Forest y la capacitación posterior a los técnicos de las diferentes instituciones públicas de Costa Rica.

2.3. PRE-PROCESAMIENTO

2.3.1. Correcciones geométricas

La corrección geométrica de una imagen tiene por función realizar todas las transformaciones necesarias para corregir las deformaciones y distorsiones existentes en el producto cartográfico con el objeto de obtener otro producto derivado del primero que cumpla unos determinados requisitos de calidad métrica. Las imágenes se han procesado en el sistema de proyección UTM, husos 16 y 17, datum WGS84. Se ha trabajado en este sistema de referencia y se ha proyectado únicamente a CRTM05 (sistema propio creado por el gobierno de Costa Rica y oficial en el país) los mapas finales elaborados en el proyecto.

Para los procesos de tratamiento geométrico, tanto en lo referido a la validación de la precisión geométrica de las escenas descargadas como en el ajuste geométrico entre imágenes se ha utilizado el software ERDAS LPS.

Preparación de imágenes

Las imágenes Landsat empleadas en el proyecto han sido descargadas en formato L1G por lo que ya han sido procesadas geométrica y radiométricamente y cuentan, por tanto, con coordenadas geográficas en proyección UTM, datum WGS. Según la pasada de que se trate pueden estar en huso 16 (pasada 14), en el huso 17 (pasada 16) o algunas filas en el 16 y otras en el 17 (pasada 15). Las pasadas 14 y 16 se han mantenido en su huso correspondiente mientras que en la pasada 15 se ha re proyectado la escena superior (fila 52) del huso 17 al 16 y se ha empleado el huso 16 en toda la pasada.

Al encontrarse las imágenes corregidas geométricamente se ha realizado un mosaico de escenas para cada una de las fechas disponibles de cada pasada disminuyendo en gran medida el número de ficheros a tratar. La unión entre escenas de la misma pasada se ha verificado por fotointerpretación visual en la zona de solape comprobándose que en la mayoría de los casos presenta un ajuste perfecto salvo en unas pocas escenas antiguas en las que fue necesario un ligero desplazamiento de la escena superior (fila 52) previo a su mosaicado con el resto de escenas de la pasada (los ficheros de puntos de control y los gráficos de distribución de los mismos se incluyen en el **Anexo 3**).

Para realizar este mosaico de las escenas de una misma pasada y fecha en primer lugar se han superpuesto una a una las diferentes bandas de las escenas de la pasada (bandas b1, b2, b3, b4, b5 y b7 en el caso de Landsat 4, 5 y 7 y las bandas b2, b3, b4, b5, b6 y b7 en el caso de Landsat 8) para posteriormente realizar una **fusión en un único fichero** ("stacklayers") de 6 bandas. La nomenclatura seguida para nombrar las imágenes resultantes ha seguido el siguiente esquema PPP_SSS_AAAADDD_b1a7_utmXX.img donde:

- PPP = pasada (p14, p15 o p16)
- SSS = sensor (LC8, LE7, LT5 o LT4)
- AAAA = año
- DDD = día del año (de 1 a 366)
- XX = huso utm empleado

Una vez generadas las uniones para cada fecha se ha seleccionado una de estas fechas como referencia en cada una de las tres pasadas. Las imágenes seleccionadas como imágenes "maestras" a las cuales ajustar el resto de fechas disponibles han sido:

- Pasada 14: p14_LC8_2014027_b1a7_utm16.img, Landsat 8, 27 de enero de 2014
- Pasada 15: p15_LC8_2014050_b1a7_utm16.img, Landsat 8, 19 de febrero de 2014
- Pasada 16: p16_LC8_2014057_b1a7_utm16.img, Landsat 8, 26 de febrero de 2014

Toma de puntos de control terrestre y de enlace entre imágenes

Para la toma de puntos de control para la validación geométrica de las pasadas de referencia se ha utilizado la ortofotografía de 2005 generada con la misión CARTA del proyecto BID-Catastro. La toma de puntos de control permite relacionar puntos homólogos entre la imagen a corregir y el dato tomado como referencia. En general, en los procesos de corrección de datos satelitales, el número de puntos a tomar dependerá del tipo de escena y de las características del terreno.

Es de gran importancia la densidad, distribución y geometría de los puntos los cuales se han tomado en zonas claramente identificables y estables tanto en la imagen como en el dato de referencia.

La validez de los puntos de control, así como la posterior ortorrectificación se determina mediante la raíz cuadrada del **error cuadrático medio (RMSE) que en ningún caso está por encima del tamaño del píxel** de la imagen.

Aplicando un algoritmo que utiliza un modelo geométrico (polinomial de segundo orden en nuestro caso) se procede a ajustar mediante mínimos cuadrados todas las observaciones: las coordenadas en las imágenes y las coordenadas terreno medidas en la base de datos suministrada.

El número de puntos correctos después de la depuración no es inferior a **20 puntos de control terrestre en cada pasada de referencia**. Si en el proceso de depuración se eliminaron demasiados puntos, se tomaron otros nuevos para sustituirlos hasta alcanzar este mínimo de puntos finales.

Una vez obtenidos los ficheros de puntos de control se procedió, en caso de no cumplir la imagen con los requisitos de precisión geométrica estipulados en el proyecto, a su ortorrectificación haciendo uso del Modelo Digital del Terreno de Costa Rica generado a partir de los datos de la misión SRTM. Se ha realizado una reproyección del MDT a proyección UTM, datum WGS84, husos 16 y 17.

Para las demás imágenes procesadas en el proyecto se ha seguido la misma metodología pero los puntos de control se han obtenido directamente sobre la imagen de referencia de forma que se asegura un mejor ajuste geométrico entre las diferentes fechas optimizándose los procesos posteriores de clasificación.

Control de calidad visual y geométrica de las imágenes corregidas

De cada una de las pasadas generadas se ha realizado una inspección visual exhaustiva para asegurarse de que no se ha producido ningún defecto en el proceso de generación de éstas: zonas duplicadas, estiramientos de píxeles o errores geométricos producidos por errores del modelo digital del terreno.

Además, se ha realizado un control geométrico de todas las imágenes mediante la toma de puntos de validación en cada pasada repartidos regularmente sobre la imagen. El error medio cuadrático obtenido en los puntos de chequeo es inferior a 30 metros (i.e. 1 píxel), siendo el error máximo admisible en cualquiera de los puntos menor a 60 m (i.e. 2 píxeles). En el **Anexo 4** se detallan las tablas con los listados de puntos de chequeo, los errores encontrados en cada punto y el RMS para cada una de las fechas procesadas en el proyecto.

2.3.2. Corrección radiométrica y atmosférica

El procesado de la imagen satelital para su posterior análisis requiere de una corrección radiométrica y atmosférica. Estas correcciones transforman la imagen de valores digitales (ND, número digital), proporcionales a la radiancia recibida por el sensor, en valores de reflectancia, porcentaje de radiación reflejada por la superficie y captada por el sensor.

En el recorrido que hace la luz desde que se refleja por la superficie terrestre hasta que es capturada por el sensor existen varios factores que influyen en la recepción de esta señal y que son considerados en la corrección atmosférica de la imagen. Estos factores se refieren tanto a

parámetros intrínsecos de calibración del sensor (densidad óptica atmosférica para cada rango espectral de los canales del sensor e irradiancia exoatmosférica solar también para cada canal), como a factores relacionados con el medio: relieve de la superficie (ángulos de incidencia y sombras proyectadas), posición solar y la distancia Tierra-Sol en el momento de captura de la imagen y datos de espesor atmosférico, que influye en el recorrido de ida y retorno de la radiación.

A continuación se detalla el proceso de corrección radiométrica y atmosférica para los diferentes sensores utilizados en la serie temporal.

Corrección radiométrica y atmosférica en Landsat 4, 5 y 7

Para transformar valores digitales en niveles de radiancia en las escenas provenientes de los sensores TM y ETM+ en primer lugar se han calculado las radiancias a partir de los coeficientes de calibración del sensor. Para ello es necesario emplear la siguiente expresión:

$$L_{\lambda} = \left(\frac{LMAX_{\lambda} - LMIN_{\lambda}}{QCALMAX - QCALMIN} \right) \cdot (QCAL - QCALMIN) + LMIN_{\lambda} \quad \text{ó} \quad L_{\lambda} = G_{rescale} \cdot QCAL + B_{rescale}$$

Donde,

$$G_{rescale} = \frac{LMAX_{\lambda} - LMIN_{\lambda}}{QCALMAX - QCALMIN}$$

$$B_{rescale} = LMIN_{\lambda} - \left(\frac{LMAX_{\lambda} - LMIN_{\lambda}}{QCALMAX - QCALMIN} \right) \cdot QCALMIN$$

L_{λ} = Radiancia espectral en una longitud de onda determinada ($W/m^2 \cdot sr \cdot \mu m$)

$QCAL$ = Valor cuantificado y calibrado del pixel (ND)

$QCALMIN$ = Valor mínimo cuantificado y calibrado correspondiente a $L_{MIN_{\lambda}}$

$QCALMIN=1$ para productos LPGS

$QCALMIN=1$ para productos NLAPS procesados antes de 4/4/2004

$QCALMIN=0$ para productos NLAPS procesados después de 4/5/2004

$QCALMAX=255$; valor máximo cuantificado y calibrado correspondiente a $L_{MAX_{\lambda}}$

$LMIN_{\lambda}$ =Radiancia espectral del sensor correspondiente al valor digital 0 ($W/m^2 \cdot sr \cdot \mu m$)

$LMAX_{\lambda}$ =Radiancia espectral del sensor correspondiente al valor digital 255 ($W/m^2 \cdot sr \cdot \mu m$)

Estos parámetros anteriormente descritos se encuentran en el fichero de metadatos que incluye la imagen Landsat (_MTL.txt).

Una vez obtenida la radiancia espectral se puede calcular la reflectividad aparente (TOA, por sus siglas en inglés, Top Of Atmosphere), conociendo el ángulo cenital solar que viene dado en los metadatos de la imagen y la distancia Tierra-Sol en el momento de la toma de la imagen. Para poder utilizar la información radiométrica de la imagen en todas las facetas es necesario convertir la reflectividad aparente (TOA) a reflectividad de la superficie terrestre. El proceso necesario para esa conversión es llamado corrección atmosférica.

En el presente trabajo se ha usado el método adoptado por USGS (United States Geological Survey) para la corrección atmosférica, el cual se basa en el modelo de transferencia radiativa

MODTRAN (Moderate resolution atmospheric transmission) (Berk et al., 2005). Los objetivos de este método son eliminar de la radiancia recibida por el sensor los efectos de la absorción y dispersión causados por las moléculas y partículas atmosféricas en suspensión y, en segundo lugar, convertir esa radiancia a valores de reflectividad de superficie, siendo este valor adimensional y expresado en tanto por uno. Para ello se usa la siguiente expresión:

$$\rho_{\lambda} = \frac{\pi L_{\lambda} d^2}{ESUN_{\lambda} \cos \theta_s}$$

Donde,

ρ_{λ} : valor de reflectancia

L_{λ} : radiancia espectral

d : Distancia en unidades astronómicas de la Tierra al Sol

$ESUN_{\lambda}$: Irradiancia solar. Los valores se puede consultar en Chander et al., 2009

θ_s : angulo zenital solar en grados.

Corrección radiométrica y atmosférica en Landsat 8

Las imágenes Landsat 8 consisten en una serie cuantificada y calibrada de niveles digitales que pueden ser reescalados a valores de radiancia y reflectancia usando para ello los coeficientes radiométricos provistos en el archivo de metadato (_MTL.txt), tal y como se describe a continuación:

$$L_{\lambda} = M_L Q_{CAL} + A_L$$

Donde:

L_{λ} es el valor de la radiancia espectral en el techo de la atmósfera (TOA) medida en valores de ($W/m^2 \text{ srad } \mu m$)

M_L factor multiplicativo de escalado específico obtenido para cada banda de los metadatos (RADIANCE_MULT_BAND_x. donde x es el número de la banda).

A_L factor aditivo de escalado específico obtenido de los metadatos (RADIANCE_ADD_BAND_x. donde x es el número de la banda).

Q_{cal} producto estándar cuantificado y calibrado por valores de pixel (DN). Este valor se refiere a cada una de las bandas de la imagen.

Parar obtener valores de reflectancia en el techo de la atmósfera (TOA) se usan los coeficientes de reflectancia suministrados en el archivo de metadatos. El siguiente algoritmo es usado para convertir los niveles digitales a valores de reflectancia:

$$p_{\lambda} = M_p Q_{cal} + A_p$$

p_{λ} Reflectancia sin la corrección por ángulo solar.

M_p , es el factor multiplicativo de escalado específico por la banda obtenido del archivo de metadatos (REFLECTANCE_MULT_BAND_x. donde x es el número de la banda).

A_p , factor aditivo de escalado específico obtenido de los metadatos (REFLECTANCE_ADD_BAND_x. donde x es el número de la banda).

Q_{cal} , producto estándar cuantificado y calibrado por valores de pixel (DN). Este valor se refiere a cada una de las bandas de la imagen.

El cálculo de la reflectancia real de una cubierta captada por un sensor espacial está condicionado por el comportamiento de la atmósfera, así como por el ángulo de observación. De esta manera, la reflectancia con una corrección para el ángulo solar es entonces:

$$p_{\lambda} = \frac{p_{\lambda'}}{\cos(\theta_{sz})} = \frac{p_{\lambda'}}{\sin(\theta_{se})}$$

p_{λ} , Valor de reflectancia

θ_{se} , Ángulo de elevación solar; provisto en el metadato de la imagen (SUN_ELEVATION)

θ_{sz} , Ángulo zenital solar local, $\theta_{sz}=90-\theta_{se}$.

2.3.3. Normalización horaria

La normalización horaria consiste en corregir el valor digital (DN) de una escena con un factor que considera la variación del flujo solar incidente en esa escena respecto a una escena de referencia. Esta corrección también se puede aplicar sobre los valores de reflectancia, tal y como ha sido llevado a cabo en este trabajo. La corrección está basada en considerar el valor de la imagen linealmente dependiente del flujo solar incidente. El flujo recibido por una superficie horizontal se incrementa según el coseno del ángulo cenital. El modelo que relaciona el flujo recibido en la superficie con el valor registrado:

$$\rho_n = \rho_h (\cos \theta_n / \cos \theta_h)$$

Es decir, la reflectancia (ρ) a una hora normalizada "n" es igual al obtenido a una hora "h" corregido con el cociente entre el ángulo cenital a la hora de referencia y el de la hora "h" (De Miguel, 1999).

Para llevar a cabo esta normalización en toda la serie se ha utilizado como referencia el ángulo zenital 36.90° correspondiente al día 17 de febrero de 2013.

2.3.4. Normalización radiométrica de la serie

El objetivo de la normalización es reducir las diferencias radiométricas entre las imágenes que componen la serie debidas a las distintas condiciones atmosféricas y de calibración del sensor en las fechas en las que fueron captadas.

En la normalización radiométrica se asume que la relación entre la radiancia captada por el sensor en dos fechas diferentes en áreas de reflectancia constante puede ser determinada por funciones lineales. El punto crítico es la determinación de estas áreas de reflectancia invariante en el tiempo sobre las que basar la normalización.

El método más actualizado para este proceso se denomina IR-MAD (Iteratively Reweighted Multivariate ALteration Detection) propuesto por Canty y Nielsen (2008), el cual se puede consultar en el **Anexo 5**. Este algoritmo ha sido desarrollado en Python por Morton J. Canty, e implementado en QGIS por AGRESTA para este proyecto. La metodología consta de dos pasos. En primer lugar se aplica el algoritmo a algunas imágenes (una de referencia y otras que

deseen corregirse respecto a esta) para localizar píxeles invariantes en ambas imágenes. Después, los píxeles así seleccionados son usados para normalizar la imagen con respecto a la imagen de referencia. La normalización de toda la serie se lleva a cabo por pares de imágenes formados por la imagen de referencia y las que se han de corregir respecto a ésta. Debido a que el rango espectral de las diferentes bandas de los sensores TM y ETM+ (Landsat 4, 5 y 7) no es exactamente igual al OLI (Landsat 8), se ha escogido una imagen de referencia para cada sensor. En ambos casos se ha elegido la imagen que menos nubosidad presentaba. En la pasada 14, para los sensores TM y ETM+, la imagen de referencia se corresponde con la del 17 de febrero de 2013, y para el sensor OLI con la del 27 de enero de 2014. En la pasada 15, 25 de diciembre de 2013 para TM y ETM+, y 19 de febrero de 2014 para OLI. En la pasada 16, para los sensores TM y ETM+, la imagen de referencia es la del 27 de diciembre de 2011; y para el sensor OLI la del 26 de febrero de 2014.

2.3.5. Detección de vacíos de información

Las nubes y sus sombras son una de las principales limitaciones que presenta el uso de imágenes ópticas satelitales para la determinación de cubiertas vegetales en zonas tropicales. En este proceso se ha adaptado la metodología propuesta por Martinuzzi *et al.* (2007), que se puede consultar en el **Anexo 5**. Para limitar el efecto de la cobertura nubosa en la construcción de la serie histórica se ha trabajado con múltiples escenas para cada año, realizando una selección de las mejores imágenes por escena. En la Figura 5 se detalla, para cada uno de los años de la serie, el porcentaje de vacíos de información resultante tras el mosaicado de todas las imágenes utilizadas en el proyecto. Los archivos raster de vacíos de información con sus correspondientes metadatos se encuentran en el anexo 7 del presente documento.

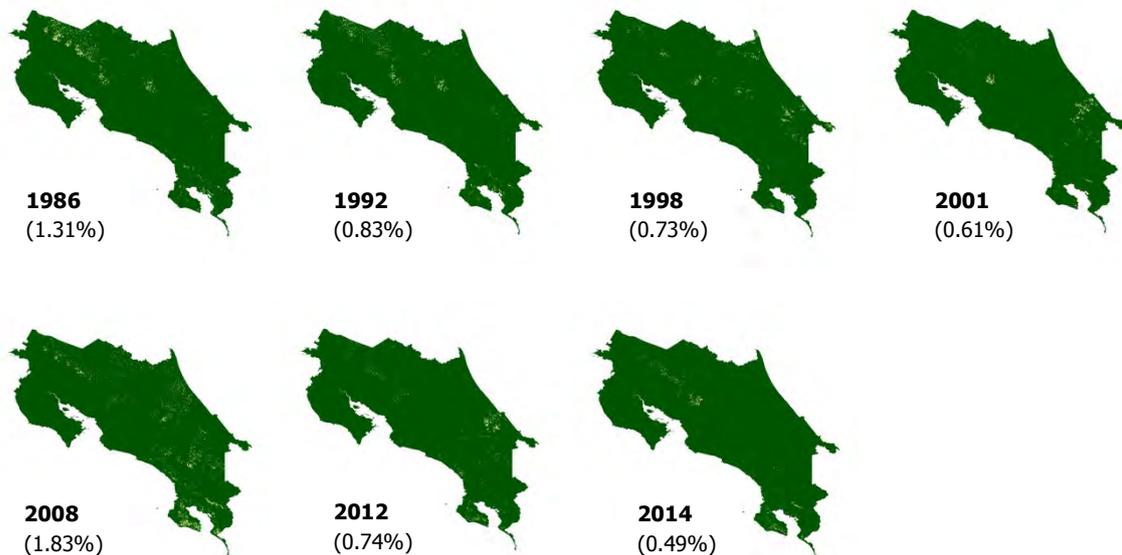


FIGURA 5. Porcentaje de vacíos de información para cada uno de los años de la serie, después de haber realizado el mosaico de todas las imágenes analizadas para cada año de la serie

Detección de nubes

En el caso de las imágenes provenientes de Landsat 8 se ha usado la banda de evaluación o de control de Calidad (BQA) para la detección de nubes. Cada píxel de la banda de control de calidad contiene un valor decimal que representa las combinaciones de bits de relleno de la superficie, la atmósfera, y las condiciones del sensor que pueden afectar a la utilidad general de un píxel dado.

Este archivo BQA contiene los estadísticos de calidad obtenidos de los datos de la imagen e información de la máscara de nubes para la escena. El archivo BQA es una imagen en 16 bits con las mismas dimensiones que la escena original. Estos valores pueden ser utilizados con eficacia, de este modo los bits de control de calidad mejoran la integridad de las investigaciones científicas, ya que suministra información sobre cuáles de los píxeles pueden verse afectados por las coberturas nubosas.

Así, en las escenas de Landsat 8 la máscara de cobertura nubosa se ha obtenido mediante el proceso de la banda BQA de acuerdo a los pasos que se detallan a continuación:

- Creación de una máscara binaria utilizando como valor de corte el nivel digital 28.800. A partir de este valor se ha considerado como nube y por debajo del mismo no se ha incluido en la cobertura nubosa a fin de evitar que píxeles de vegetación se pudieran enmascarar en la capa de nubes.
- En las imágenes con neblina el parámetro anterior se ha sustituido por la siguiente combinación de parámetros:
 - $(BQA > 54.000)$ ó $(B4 > 30.000)$ ó $(BQA > 28400 \text{ y } B4 > 20.000)$
- Eliminación de los píxeles cuyo valor en B5 fuese inferior a 15.000 (por lo general arena, suelos desnudos, etc.).
- Filtrado (matriz 3x3) y eliminación de píxeles sueltos. Esto elimina algunas nubes pequeñas pero permite descartar muchos píxeles sueltos incorporados erróneamente a la máscara de nubes.
- Filtrado (suavizado) con una matriz de 5x5 píxeles incorporando a la máscara de nubes a todos aquellos píxeles que se encuentren a una distancia inferior a 40 metros de un píxel clasificado como nube. Esto permite eliminar "agujeros" del centro de las nubes e incorpora también a la máscara de nubes a los píxeles del contorno que suelen estar afectados en mayor o menor medida por la presencia de las mismas.

En las escenas Landsat 4, Landsat 5 y Landsat 7 la máscara de cobertura nubosa se ha obtenido mediante una combinación de valores de las bandas B1 (visible), B4 (infrarrojo cercano) y B7 (infrarrojo medio) de acuerdo a los pasos que se detallan a continuación:

- Creación de una máscara binaria con los píxeles que cumplan el siguiente criterio:
 - $B1 > 90 \text{ y } B4 > 130 \text{ y } B7 > 50$.
 - En las imágenes con neblina estos parámetros se ajustan: $B1 > 90 \text{ y } B4 > 90 \text{ y } B7 > 50$. Da lugar a una máscara con mayor nivel de confusión pero que incorpora a las zonas de neblina dentro de la misma
 - En imágenes muy secas se limita el filtro para evitar que los suelos desnudos se clasifiquen como nube: $B1 > 130 \text{ y } B4 > 130 \text{ y } B7 > 50$.

- Eliminación de los píxeles cuyo valor en B5 fuese inferior a 15.000 (por lo general arena, suelos desnudos, etc.).
- Filtrado (matriz 3x3) y eliminación de píxeles sueltos. Esto elimina algunas nubes pequeñas pero permite descartar muchos píxeles sueltos incorporados erróneamente a la máscara de nubes.
- Filtrado (suavizado) con una matriz de 5x5 píxeles incorporando a la máscara de nubes a todos aquellos píxeles que se encuentren a una distancia inferior a 40 metros de un píxel clasificado como nube. Esto permite eliminar "agujeros" del centro de las nubes e incorpora también a la máscara de nubes a los píxeles del contorno que suelen estar afectados en mayor o menor medida por la presencia de las mismas.

Detección de sombras

Para la delimitación de las zonas de sombra se han empleado las bandas térmicas e infrarrojas utilizando la metodología citada anteriormente (Martinuzzi *et al.* 2007). De acuerdo a esta metodología se tiene en cuenta que la detección de sombras plantea significativos problemas de confusión con otras clases espectrales, especialmente con zonas de vegetación en umbría de gran interés en el proyecto.

Para limitar este efecto se incorpora al proceso de generación de las máscaras de sombras la creación de un "mapa de distancias" de cada píxel a la nube más cercana, orientado en la dirección de la iluminación solar y teniendo en cuenta la altura media de las masas nubosas.

La distancia nube-sombra varía en función de:

- Fecha del año: determina la altura y posición del sol. La mayoría de las imágenes seleccionadas está comprendida en un rango temporal breve (diciembre-marzo) lo que disminuye la variación debida a este factor por lo que se ha seleccionado una orientación genérica para la mayoría de las escenas.
- Altura de la nube: muy variable según tipo de nube y orografía. Corrección específica de la distancia en aquellos casos en los que se observa un desplazamiento medio significativo al realizar la revisión de la máscara de nubes.

El proceso seguido para la determinación de las máscaras de sombras se detalla a continuación:

- Creación del mapa de distancias nube-sombra de acuerdo a los siguientes parámetros:
 - dX: desplazamiento hacia el Oeste de 10 a 22 píxeles
 - dY: desplazamiento al Norte entre 7 y 16 píxeles
- Una vez generada la máscara de sombras se realiza una inspección visual de la misma superpuesta a la imagen de la que procede. En aquellos casos en los que se observa que hay zonas significativas de sombras no incluidas en la máscara se modifican los parámetros del mapa de distancias. Los valores extremos encontrados en el desplazamiento nube-sombra han sido:
 - dX: desplazamiento hacia el Oeste de 4 a 32 píxeles
 - dY: desplazamiento al Norte entre 2 y 27 píxeles

- En las imágenes Landsat 8 formarán parte de la máscara de sombras todos aquellos píxeles que se encuentren dentro de la distancia especificada y además su valor en la banda B5 sea superior a 7.000 e inferior a 18.000
- Para las imágenes Landsat 4, Landsat 5 y Landsat 7 se seleccionan todos los píxeles que se encuentren dentro de la distancia especificada y además su valor en la banda B4 sea inferior a 60 y el valor en la banda B5 inferior a 40
- Además se incorporan a la máscara de sombras todos los píxeles externos a la delimitación territorial de Costa Rica y aquellos cuyo nivel digital en B4 sea inferior a 7 a fin de eliminar las "líneas perdidas" o líneas de datos inválidos.

Vacíos por fallos del sensor

En Mayo de 2003, el sistema conocido como *Scan Line Corrector (SLC)* del sensor ETM+ montado en el satélite Landsat 7 comenzó a fallar. A partir de este momento el sistema se encuentra apagado y las escenas adquiridas presentan líneas de datos inválidos. Los datos inválidos están dispuestos en franjas inclinadas hacia la izquierda unos 8° respecto a la orientación horizontal debido a la rotación de la imagen, y aparecen en intervalos de 33 píxeles. Estas franjas tienen hasta 15 píxeles en el borde de la imagen y van disminuyendo gradualmente camino al centro hasta desaparecer. Se ha evitado al máximo este problema eligiendo imágenes de las mismas fechas del satélite Landsat 5. Una vez realizada la selección de imágenes se ha visto necesario trabajar con algunas imágenes Landsat 7 SLC-off. No se ha aplicado a estas imágenes ningún algoritmo de interpolación, el relleno de estas bandas sin información se ha resuelto combinando múltiples escenas para cada año de la serie.

Control de calidad de las máscaras de nubes y sombras

El control de calidad se efectúa mediante comprobación visual de una malla aleatoria de puntos de chequeo repartidos regularmente. Los puntos son identificados como nube (n), sombra (s), despejado (d) y se comprueba si coincide su clasificación en las máscaras de nubes y sombras. La validación se ha realizado en una muestra aleatoria (Figura 6) en cada pasada de una imagen de cada una de las series temporales (3 pasadas x 6 series = 18 imágenes). En la Tabla 2 se resumen los resultados de la validación de los mapas de nubes y sombras.

TABLA 2. Validación máscaras nubes y sombras

Imagen	Mascaras p16			Mascaras p15			Mascaras p14		
	d	n	s	d	n	s	d	n	s
d	71	0	0	75	0	0	35	1	0
n	5	22	1	2	37	0	1	13	0
s	0	0	4	2	0	10	0	0	4
Acierto	94,2%			96,8%			96,3%		



FIGURA 6. Malla de puntos para la validación de las máscaras de nubes y sombras

2.4. CÁLCULO DE ÍNDICES DE VEGETACIÓN Y TEXTURAS

Además de la información espectral correspondiente a las 6 bandas de cada una de las imágenes, según se detallaba en los puntos anteriores, para el presente trabajo se van a utilizar también algunos índices de vegetación y de texturas calculados a partir de la información de estas bandas.

2.4.1. Índice de vegetación de diferencia normalizada (NDVI)

El índice de vegetación de diferencia normalizada (NDVI, por sus siglas en inglés, Normalized Difference Vegetation Index) es una de las combinaciones de bandas (índices espectrales) usadas más comúnmente en el análisis de imágenes satelitales. El NDVI es una combinación de las bandas rojo e infrarrojo cercano y se calcula según la fórmula:

$$NDVI = \frac{IRC - R}{IRC + R}$$

Donde,

IRC es la banda espectral del Infrarrojo Cercano

R es la banda espectral del Rojo

El cálculo de este índice se realizó en el entorno de QGIS.

2.4.2. Índices de textura

Los índices de textura también han sido identificados como determinantes a la hora de clasificar coberturas vegetales en base a imágenes satelitales. Así, se ha utilizado la banda del infrarrojo cercano para calcular 6 índices relacionados con la textura: "Mean", "Sum Entropy", "Difference of Entropies", "Difference of Variances", "IC1" e "IC2". Se ha utilizado un radio de 2 píxeles para calcular los índices relacionados con la textura.

El cálculo de este índice se realizó en el entorno de QGIS.

2.5. CÁLCULO DE DERIVADAS DEL MODELO DIGITAL DE ELEVACIONES

El relieve es un factor determinante en la distribución de los usos del suelo en el paisaje. Así, en el presente trabajo se ha incluido información del relieve en la clasificación de imágenes para la creación de mapas de cobertura de la tierra se considera.

Para el presente trabajo se ha utilizado el Modelo Digital de Elevaciones (MDE) del SRTM 30. Se considera este MDE como el más completo entre los disponibles para Costa Rica, según se acordó en las reuniones con técnicos del país en agosto de 2014.

Además de los datos de elevaciones (directamente los valores del MDE), se han calculado 6 variables adicionales derivadas de este MDE: pendiente, sombreado, curvatura plana, curvatura de perfil, índice de convergencia e índice multi-resolución de fondo de valle.

El cálculo de este índice se realizó en el entorno de QGIS.

2.6. GENERACIÓN DE IMÁGENES MULTIBANDA

Según los procesos explicados en los puntos anteriores, para cada una de las imágenes procesadas (**Anexo 2**), se cuenta con información de 20 variables: 6 bandas espectrales originales, el NDVI, 6 índices de textura y 7 variables derivadas del MDE (Tabla 3).

En base a estas 20 variables (Tabla 3), se ha generado un archivo raster de 20 bandas en formato geo TIFF con la información continua de todas ellas (o imagen multibanda), usando para ello herramientas propias de QGIS. En el proceso de creación de estos archivos de 20 bandas se ha asignado un valor "sin información" a todos los píxeles de nubes, sombras o fallos del sensor a partir de las máscaras de vacíos de información creadas para cada imagen (ver apartado 2.3.5 de la presente memoria). De esta forma conseguimos para cada una de las imágenes no tener en cuenta las zonas con nubes, sombras o de fallos del sensor ni en el proceso de ajuste de los clasificadores ni en el de clasificación.

TABLA 3. Resumen de las variables predictoras usadas para el ajuste de los modelos RF

Grupo de información	id	Variable
Información Espectral	1	Azul (banda 1 ETM+ y TM y banda 2 OLI/TIRS)
	2	Verde (banda 2 ETM+ y TM y banda 3 OLI/TIRS)
	3	Rojo (banda 3 ETM+ y TM y banda 4 OLI/TIRS)
	4	NIR (banda 4 ETM+ y TM y banda 5 OLI/TIRS)
	5	SWIR-1 (banda 5 ETM+ y TM y banda 6 OLI/TIRS)
	6	SWIR-2 (banda 7 ETM+ y TM y banda 7 OLI/TIRS)
Índice de Vegetación	7	NDVI, Normalized Difference Vegetation Index
Texturas	8	Mean
	9	Sum Entropy
	10	Difference of Entropies
	11	Difference of Variances
	12	IC1
	13	IC2
	Derivadas DEM	14
15		Slope
16		Hillshade
17		Plan curvature
18		Profile curvature
19		Convergence Index
20		MRVBF

2.7. CLASIFICACIÓN DE IMÁGENES

Para la clasificación de las imágenes se ha utilizado la metodología Random Forest (en adelante RF), una técnica de aprendizaje automática agrupada (Breiman 2001) que se ha convertido en una de las más usadas en la clasificación de coberturas de la tierra. La clasificación de las imágenes según esta metodología se realiza en 2 fases: (1) entrenamiento o ajuste del RF y (2) clasificación de las imágenes usando el clasificador RF generado.

Los clasificadores RF son en esencia modelos estadísticos que se ajustan basándose en las relaciones entre una variable respuesta (en este caso "cobertura de la tierra") y una serie de covariables predictoras en una serie de puntos de control.

2.7.1. Regiones de interés y puntos de control

Para el ajuste de los modelos se han digitalizado regiones de interés homogéneas en función de las clases de cobertura de la tierra consideradas (Tabla 4), entre los años 2011 y 2014. La información base utilizada para la digitalización y fotointerpretación de estas regiones ha sido:

1. La malla sistemática de puntos de coberturas tomada sobre las imágenes Rapideye por SINAC para la elaboración del mapa de tipos de bosque de Costa Rica 2013 (10.000 puntos por todo el país)
2. Las propias imágenes de alta resolución espacial RapiEye,
3. Las imágenes tanto actuales como históricas disponibles en GoogleEarth
4. Las propias imágenes Landsat que se van a clasificar

A partir de estas regiones de interés se han generado de forma aleatoria los puntos de control para el entrenamiento de los RF.

TABLA 4. Clases de cobertura de la tierra definidas para el proceso de ajuste de los modelos Random Forest y clasificación de las imágenes

Clase	Nombre
1	Bosque
2	Plantación forestal
3	"Bosque inundado": Manglar y Yolillal
4	Áreas Urbanas
5	Pastizal
6	Agua
7	Terrenos desnudos
8	No bosque datos globales
9	Cultivos anuales
10	Café
11	Piña
12	Otros cultivos permanentes

El páramo no ha sido clasificado a partir de la información espectral sino en base al mapa de Tipos de Bosque de 2013 y por lo tanto se ha considerado estable en todos los años de la serie.

Como puede observarse en la Tabla 4, en el ajuste de los modelos se clasificaron igual los dos tipos de bosques inundados (manglar y yolillal), ya que se observó que el ajuste del modelo era mejor de esa forma. Así, la separación de estas dos coberturas de la tierra se realizó posteriormente usando máscaras.

2.7.2. Variables predictoras

En total se han usado 20 variables predictoras (también llamadas covariables o variables auxiliares) para el ajuste de los modelos RF (Tabla 3), según se detalla en el apartado 2.6 de este documento.

2.7.3. Ajuste de clasificadores

Para el ajuste de los modelos RF se ha utilizado el paquete de R "RandomForest: Breiman and Cutler's Random Forests for classification and regression" (Liaw A. and Wiener M., 2002), siguiendo la guía de uso del mismo propuesta por Horning (2013). Se han ajustado dos RF para cada una de las pasadas, uno para clasificar las imágenes provenientes de Landsat 8 (modelos RF_L8) y otro para las imágenes de Landsat 4, 5 y 7 (modelos RF_L5y7). Adicionalmente, para la pasada 16 se han ajustado dos clasificadores adicionales para las imágenes de invierno (modelos RF_L8_inv y RF_L5y7_inv). Así, en total se han ajustado 8 clasificadores, 2 para la pasada 14, 2 para la 15 y 4 para la pasada 16).

Los modelos RF_L8 se han ajustado usando la información espectral de todas las imágenes de Landsat 8 utilizadas en el proyecto (**Anexo 2**). Los modelos RF_L5y7 se han ajustado basándose en las imágenes de Landsat 5 y Landsat 7 entre 2011 y 2014 (**Anexo 2**).

Los dos parámetros fundamentales que se definen a la hora de ajustar un modelo del tipo RF son "n" (número de árbol de decisión que se ensamblan) y "m" (número de variables que se eligen aleatoriamente entre las predictoras para cada nodo de cada uno de los árboles de decisión independientes). Para este caso se han ajustados con $n=500$ y $m=4$.

La metodología RF también aporta información acerca de la importancia de las variables predictoras en la clasificación. Así, en los modelos ajustados se observa como las variables relacionadas con el relieve y la posición topográfica (especialmente la elevación) y la información espectral de las imágenes (principalmente las bandas "NIR" y "SWIR-1" y "azul") son las que adquieren mayor peso en la clasificación (Figura 7).

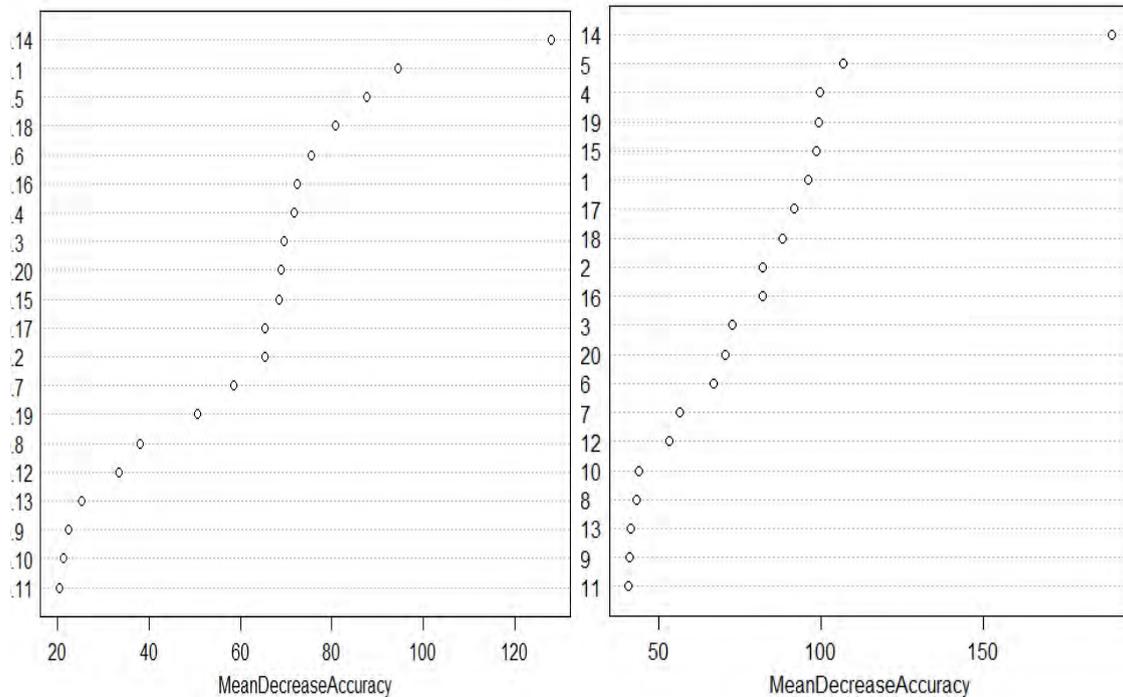


FIGURA 7. Ejemplo de importancia relativa de las variables predictoras en el ajuste de los clasificadores RF_L8 (izquierda) y RF_L5y7_2011y13 (derecha) correspondientes a la pasada 14. El código numérico de las variables se corresponde con el detallado en la Tabla 2.

2.7.4. Clasificación de imágenes

Una vez entrenados los clasificadores, según lo descrito en el apartado anterior, se han aplicado a todas las imágenes según su pasada y sensor. Para cada imagen clasificada se ha obtenido un archivo raster de resultados de la clasificación en formato geo TIFF compuesto por tres bandas:

- Una primera banda con la clasificación asignada al pixel
- Una segunda banda con el porcentaje de árboles de decisión (de los 500 que componen el RF) que predicen esa clase asignada en la primera banda. Esta banda proporciona una estimación de la calidad de la clasificación para cada uno de los píxeles de la imagen. Esta información se ha usado posteriormente para la priorización de píxeles en la generación del mosaico.
- Una tercera banda con la clasificación de los píxeles que han recibido un porcentaje de votos superior al 60 %, según la segunda banda.

2.8. GENERACIÓN DE MAPAS DE COBERTURA

2.8.1. Mosaicado de las imágenes clasificadas

Se han mosaicado todas las clasificaciones pertenecientes a cada uno de los años de la serie. Para realizar este mosaicado de las clasificaciones se ha desarrollado un script de Python propio que permite, a partir de una regla de priorización, asignar la clase más adecuada a cada uno de los píxeles de las imágenes. Cuando para un pixel el clasificador predice clases diferentes para las distintas imágenes, se ha seleccionado aquella que más votos suma en todas las imágenes, según la banda 2 del archivo de resultados comentado anteriormente. Este script desarrollado para el presente proyecto genera un archivo de salida que consta de tres bandas:

- Una banda con la clasificación finalmente asignada
- Una banda con porcentaje de votos de la imagen que más votos tiene para la clase asignada
- Una banda indicando la imagen que se ha utilizado para asignar la clase a cada uno de los píxeles

La unión de los tres mosaicos generados (uno por pasada) para cada uno de los años de la serie se ha realizado desechando la franja más extrema de los solapes entre pasadas.

2.8.2. Relleno de vacíos con productos globales

Los mapas resultantes del mosaicado detallado en el punto anterior (uno para cada año de la serie) presentan un porcentaje de vacíos de información debido a los fallos del sensor en las imágenes Landsat 7 y a las nubes y las sombras en todas ellas (Figuras 5 y 8). Aunque los porcentajes de vacíos de información eran muy pequeños debido al gran número de imágenes clasificadas para cada año de la serie (Figura 5 y 8), según las reuniones mantenidas en agosto de 2014, se decidió rellenar esos vacíos con datos globales cuando esto fuese posible.

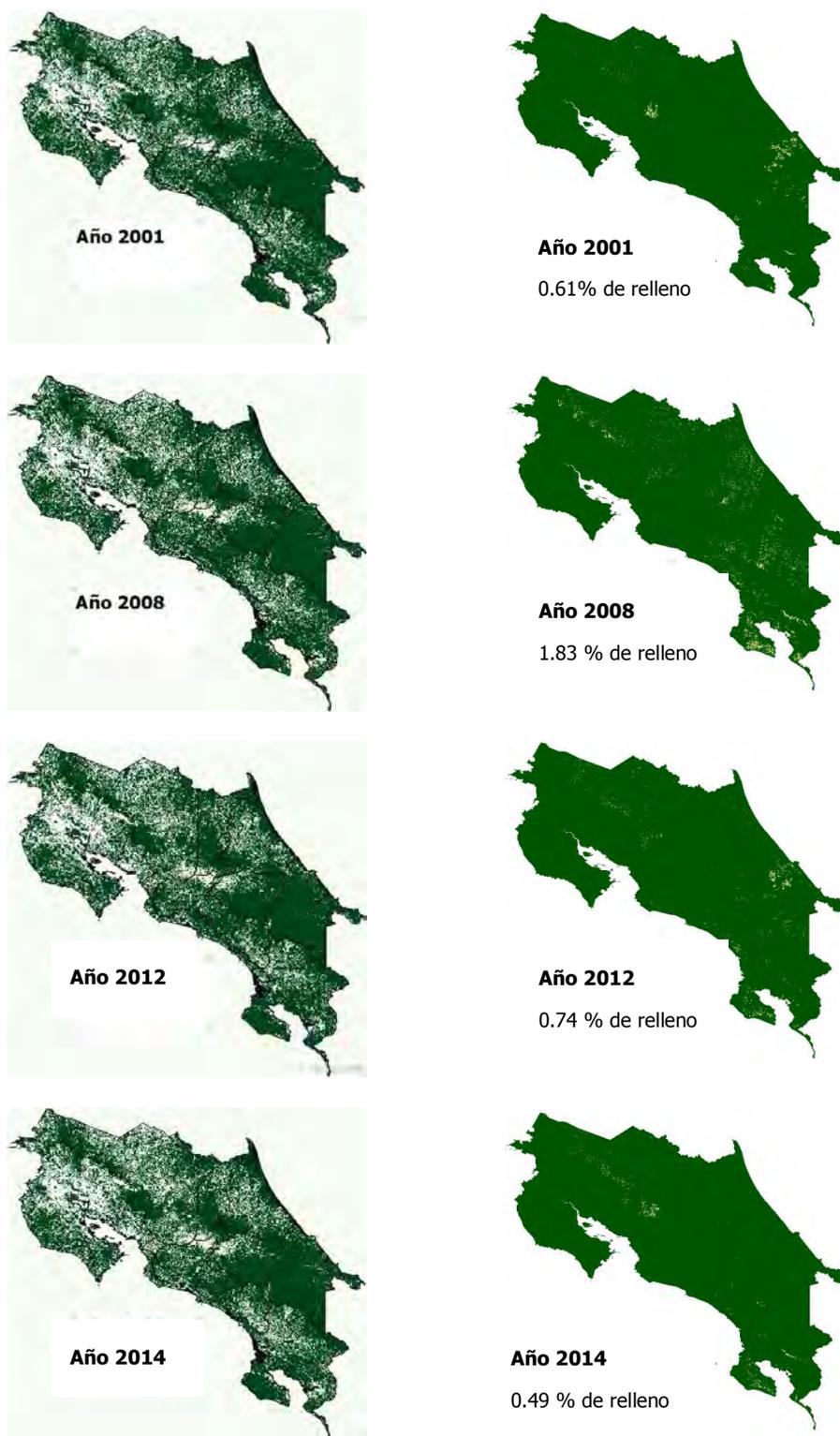


FIGURA 8. Detalle del relleno de vacíos de información usando productos globales. La columna de la izquierda muestra los mapas de bosque/no bosque de 30 metros de pixel generados a partir de los datos globales del proyecto Global Forest Change (Hansen, *et al.*, 2013) y usados para el relleno de vacíos de información. Las figuras de la columna de la derecha muestra los vacíos de información (zonas en blanco) y el porcentaje con respecto al total de la superficie del país.

Para este trabajo se han utilizado los productos globales del proyecto Global Forest Change (Hansen *et al.*, 2013) para el relleno de píxeles sin información después de realizar el mosaico de todas las clasificaciones para cada año de la serie entre el año 2000 y el 2012 (Figura 8). Este proyecto proporciona datos globales de 30 metros de pixel generados a partir de imágenes Landsat 7 ETM+ de porcentaje de cobertura arbórea para el año 2000, pérdida de bosque entre 2000 y 2012, ganancia de bosque entre 2000 y 2012 y año de pérdida de bosque de 2000 a 2012. Para el cálculo de porcentaje de la cobertura arbórea este proyecto define vegetación arbórea como vegetación por encima de los 5 metros. En nuestro caso se ha considerado bosque todos los píxeles con una cobertura arbórea mayor del 30 %. Es importante tener en cuenta que para Costa Rica estos productos consideran como bosque cultivos permanentes como la palma, el banano o el café.

A partir de esta información se han generado mapas de bosque para los años 2001, 2008, 2012 y 2014 (Figura 8) que se han utilizado para completar los huecos que no tienen clasificación tras la fase de mosaicado de las clasificaciones. En los años en que se precisa una clasificación con leyenda completa (2008, 2012 y 2014) los vacíos de información aparecerán sólo con las clases bosque y no bosque. Como se ve en la Figura 8, el porcentaje final de mapa relleno con datos globales para cada uno de los años ha sido muy bajo, estando por debajo del 2% en todos los años. En el resto de los años de la serie 1986, 1992 y 1998 no se ha realizado ningún relleno de huecos y poseen los vacíos de información que se han descrito en el apartado 3.6.

2.8.3. Asignación de clases de edad de bosque

El análisis multi-temporal de la serie permite asignar la clase de edad a cada uno de los píxeles de bosque analizando los años transcurridos a partir de la fecha de aparición de un bosque nuevo. Gracias a este análisis se han generado clases de edad de bosque de forma espacialmente explícita en todo el país.

Como se comentaba en apartados anteriores, se han considerado las áreas de bosque en el primer año de la serie temporal realizada (1986) como bosque "primario", a efectos de asignar estas clases de edad a los bosques mapeados en el resto de años de la misma. Así, aunque se ha llamado "primario" en realidad se refiere (de acuerdo a IPCC) a "forest land remaining forest land", áreas que fueron bosque por 20 años (o el periodo que se defina en el país). De esta forma, se considera como bosques "secundarios" al resto, según IPCC "lands converted to forest lands".

Según esta metodología multitemporal, el número de clases de edad va aumentando conforme más actual sea la fecha del mapa. Así, la Tabla 5 resume las clases de edad que encontramos en cada uno de los años de la serie.

TABLA 5. Clases de edad de bosque que aparecen en la serie histórica. Los rangos de edad no incluyen el tiempo transcurrido desde el inicio del bosque hasta su detección en las escenas Landsat.

Clases de edad	1986-1991	1992-1997	1998-2000	2001-2007	2008-2011	2012-2013
1986-1991	1-6					
1992-1997	7-12	1-6				
1998-2000	13-15	7-9	1-3			
2001-2007	16-22	10-16	4-10	1-7		
2008-2011	23-26	17-20	11-14	8-11	1-4	
2012-2013	27-28	21-22	15-16	12-13	5-6	1-2

La tabla anterior muestra los años transcurridos desde que ha sido detectado bosque en las imágenes Landsat, para determinar las clases de edad sería necesario sumar a los números de la matriz la edad a la cual los bosques secundarios y las plantaciones forestales alcanzan el umbral de la definición de bosque y se vuelven detectables espectralmente en las imágenes Landsat. Las clases de edad para cada una de las subclases de bosque y los códigos asociados a cada una de ellas se resumen en la tabla 6.

TABLA 6. Clases de edad de bosque que aparecen en la serie histórica. Los rangos de edad no incluyen el tiempo transcurrido desde el inicio del bosque hasta su detección en las escenas Landsat.

Código clase	Descripción	Código clase	Descripción
1	Bosque primario	4	Manglar primario
101	Bosque de 1 a 6 años	401	Manglar de 1 a 6 años
102	Bosque de 7 a 12 años	402	Manglar de 7 a 12 años
103	Bosque de 13 a 15 años	403	Manglar de 13 a 15 años
104	Bosque de 16 a 22 años	404	Manglar de 16 a 22 años
105	Bosque de 23 a 26 años	405	Manglar de 23 a 26 años
106	Bosque de 27 a 28 años	406	Manglar de 27 a 28 años
3	Yolillal primario		
301	Yolillal de 1 a 6 años		
302	Yolillal de 7 a 12 años		
303	Yolillal de 13 a 15 años		
304	Yolillal de 16 a 22 años		
305	Yolillal de 23 a 26 años		
306	Yolillal de 27 a 28 años		

2.8.4. Aplicación de máscaras

Los manglares y yolillales son formaciones forestales que aparecen en unas condiciones edáficas, de inundación y salinidad muy especiales con lo que es muy difícil que existan transiciones de este tipo de formaciones a otros tipos de bosque y viceversa. Es por ello que se ha generado una máscara de áreas potenciales de manglar y otra de áreas potenciales de yolillal que permiten mejorar la clasificación espectral en este tipo de bosques. Dentro de las máscaras generadas todos los píxeles clasificados como bosque se han definido como manglar o yolillal (según el caso) y fuera de ella no será posible encontrar este tipo de formaciones.

La generación de la máscara de manglar se ha basado en tres fuentes de información:

1. La clasificación de manglar en todos los años de la serie histórica
2. La clasificación de manglares realizada dentro del mapa de Tipos de Bosque de SINAC de 2013
3. La información proveniente del modelo digital de elevaciones del país, acotando la presencia de manglares a la franja comprendida en 0 y 20 metros de elevación

Una vez realizado el cruce de las tres fuentes de información se ha realizado una revisión manual de las áreas potenciales de manglar en base a imagen de alta resolución que ha permitido tener en cuenta áreas potenciales de manglar que no estaban incluidas.

Para la generación de la máscara de yolillal se han utilizado únicamente dos fuentes de información:

1. La clasificación de yolillal de todos los años de la serie histórica
2. La clasificación de yolillal realizada dentro del mapa de Tipos de Bosque de SINAC de 2013.

Al igual que en el caso de la máscara de manglar, en este caso también se ha realizado una revisión manual de las áreas potenciales de esta formación que ha permitido mejorar algunas áreas potenciales de yolillal en base a la fotointerpretación de imagen de alta resolución.

Para la generación de la máscara de páramo se ha tenido en cuenta únicamente el mapa de Tipos de Bosque de SINAC de 2013, además de una edición manual realizada en base a las imágenes RapidEye (de alta resolución espacial). Al igual que ocurría con las otras máscaras, ésta se ha aplicado de forma consistente a todos los años de la serie temporal. Así, se ha asignado a todos los píxeles que estaban dentro de esta máscara la clasificación de páramo.

Se han realizado modificaciones en las áreas clasificadas como café que han permitido mejorar la exactitud de esta clase en los mapas de la serie y ajustar los resultados a las estimaciones de realizadas por otras instituciones del país. Para ello se ha utilizado la información del Ministerio de Agricultura (MAG) del año 2001, la reportada por ICAFE en el año 2005 y la generada por IMN en los años 2011 y 2013 para modificar el área de café en los años 2001, 2008, 2012 y 2014. Para los mapas de la serie correspondientes a años anteriores a 2001 se creó una máscara de área de café potencial utilizando la localización y elevación de las áreas de café mapeadas en todas las fuentes disponibles (MAG, ICAFE e IMN).

Algunos suelos desnudos tienen una respuesta espectral muy similar a la que se obtienen en áreas urbanas, se ha realizado una máscara de áreas urbanas por medio de la edición manual de áreas realmente urbanas o con potencial de desarrollo urbano. Los píxeles clasificados como áreas urbanas en todos los años de la serie y que estuvieran fuera de la máscara se ha reclasificado como suelos desnudos. Dentro de la máscara de urbano se ha trabajado con el supuesto de que un área convertida a urbano no puede revertir a otros usos en el futuro.

Las máscaras de manglar, yolillal, páramo y urbano utilizadas en el proceso de generación de la serie histórica se adjuntan en el **Anexo 6**.

2.8.5. Generación de mapas finales

Se ha realizado un análisis multi-temporal de la serie para mejorar la clasificación de algunas de las clases que tienden a confundirse debido a que poseen información espectral similar. Los bosques nativos en algunos casos pueden confundirse con bosques antrópicos con cierto grado de naturalización, o al revés, las plantaciones forestales se confunden espectralmente con bosques nativos con cierto grado de intervención humana. Del mismo modo, cultivos permanentes como es el caso de la palma aceitera, del banano, del café o ciertos frutales se confunden con bosques nativos (bosques de palmas por ejemplo) o con plantaciones forestales. Para mejorar este aspecto se han aplicado de forma consistente en toda la serie las siguientes correcciones multitemporales que permiten mejorar la clasificación en las clases mencionadas anteriormente:

- Se ha asignado la clase bosque a los píxeles clasificados como bosque antrópico pero que tanto la fecha anterior como posterior en la serie estuvieran clasificados como bosque
- Se ha asignado la clase bosque a los píxeles clasificados como cultivo permanente pero que tanto la fecha anterior como la posterior en la serie estuvieran clasificados como bosque

Los mapas finales se presentan a una resolución de 30 m de píxel. Se han presentado dos sets de mapas finales: unos sin ediciones y otros con ediciones posteriores.

Para ese primer set de resultados no se ha aplicado ningún filtro de suavizado ni de área mínima a los mapas generados. Este primer set de resultados es lo acordado en las reuniones de agosto de 2014, por la relación que este proceso puede tener con la definición de bosque del país y por su influencia en la estimación de ratios de cambio de uso.

Una vez entregado este primer set de resultados y tras las reuniones mantenidas con el personal técnico de FONAFIFO y de Banco Mundial, se ha realizado un segundo set de resultados en los que se ha realizado una edición de los primeros. Así, se ha realizado un sieve de 1 ha en el entorno de QGIS. Como se discutió en las reuniones de marzo de 2015, esta edición de los resultados no varía sustancialmente los números a nivel país, pero sí mejora la calidad cartográfica de los mapas, al eliminar muchos píxeles sueltos.

2.9. REPORTE DE TRAYECTORIAS DE CAMBIO DE LA COBERTURA DEL SUELO

Los mapas finales se han reclasificado en mapas de bosque/no bosque, se han cruzado estos mapas dos a dos por años consecutivos para construir mapas de cambio con 4 clases (bosque que permanece como bosque, no bosque que permanece como no bosque, deforestación y ganancia de bosque) y se han generado matrices de cambio completas (incluyendo las clases de edad de bosque) para todas las fechas consecutivas de la serie. Todos los mapas, tanto los mapas individuales de uso del suelo como los mapas de cambio entre años consecutivos, se encuentran recogidos en el **Anexo 7**. Las matrices completas con los datos de actividad se presentan en el **Anexo 8** que acompaña el presente documento.

Para evitar confusiones e interpretaciones incorrectas no se ha detallado la deforestación de los “bosques primarios” y la deforestación de los “bosques secundarios y plantados” ya que es necesario que el país analice y decida cuál es su definición de “deforestación” (ej. edad mínima que deben tener los bosques secundarios y plantados para que su conversión a otros usos de la tierra se considere “deforestación”).

2.10. ÍNDICE DE COBERTURA COMO BASE PARA LA ESTIMACIÓN DE LA DEGRADACIÓN Y AUMENTO DE EXISTENCIAS DE CARBONO

Como base para la estimación de la degradación y el aumento de existencias se ha desarrollado de forma experimental una metodología de cálculo de un índice de cobertura de copas de forma espacialmente continua. Además del desarrollo de una metodología para el cálculo de este índice, éste se ha calculado para los años 2000 y 2012 en todo el país. La diferencia de cobertura entre ambos años muestra de forma espacialmente explícita las áreas en las que la cobertura forestal ha disminuido entre el año 2000 y el año 2012, considerándose esta disminución como una estimación directa de la degradación de los bosques existentes en ambas fechas. En ese sentido, un aumento del índice de cobertura en un área de bosque entre dos fechas se considera una estimación directa de un aumento de existencias de carbono en bosques que eran bosques y continúan siéndolo. La descripción metodológica completa de la estimación de este índice, junto con su validación, se encuentran detallados en el **Anexo 9**.

3. VALIDACIÓN

Se han validado los mapas de los años 1986, 2001 y 2014, fechas en las que hay disponible un set de datos de verdad terreno independiente para validación. Además, se ha validado el mapa de cambio de coberturas entre los años 2001 y 2012 utilizando la imagen de alta resolución disponible en el país como información base para generar una muestra de validación de cambios.

3.1. VALIDACIÓN DE LOS MAPAS DE COBERTURA

Para la validación de los mapas finales se cuenta con los siguientes puntos de control, considerados como "verdad terreno", proporcionados por FONAFIFO y otras instituciones implicadas en el trabajo:

- (1) Puntos de control, para los años 1986, 2000 y 2013, desarrollados por INBio para la elaboración y verificación de los mapas generados en el marco del proyecto "*Lecciones aprendidas y desarrollo de capacidades para aplicar iniciativas REDD+, la experiencia de Costa Rica*"
- (2) Puntos de control para el año 2013 validados en campo, desarrollados por SINAC en el marco de la elaboración del *mapa de tipos de bosque de Costa Rica*

En base a esta información disponible se ha generado una base de datos de puntos de validación, unificando los puntos disponibles de las dos fuentes y adaptando la leyenda de las mismas a la de los mapas finales generados. Dado el bajo número de puntos de "verdad terreno" de SINAC comparados con los de INBio, se han adaptado estos para que coincidan con la clasificación de los puntos del INBio para integrarlos en la base de datos de validación de forma armónica. Los puntos totales con los que se han validado los mapas de los años 1986, 2001 y 2014 ascienden a 5.396, 7.463 y 8.536 puntos respectivamente. Se han validado los mapas finales con estos puntos "verdad terreno", obteniéndose matrices de confusión y un **valor de exactitud total de la clasificación: 87% para el mapa de 2014, 93% para el mapa de 2001 y 89% para el mapa de 1986** (Tablas 7, 8 y 9).

Los resultados de la validación muestran que las clases que presentan mayores errores son los pastizales, el agua y el suelo desnudo (Tabla 9). En el caso de los pastizales, los puntos de validación proporcionados por INBio incluyen en esta clase tanto los pastizales de uso ganadero ("potreros") como algunos cultivos agrícolas anuales herbáceos (p.ej. arroz), lo que significa que algunos de esos errores son en realidad confusiones de nomenclatura de los puntos de validación. Por otro lado, la clase de agua se confunde a menudo con suelo desnudo, bosque, manglares y yolillales. Durante el proceso de validación se ha podido observar que estos errores seguramente se deban a pequeños errores de georreferenciación de los puntos de validación del INBio y/o los mapas de clasificación realizados. Esto puede comprobarse en los ejemplos que se detallan en la Figura 9, en la que se observa como los puntos de agua de INBio siguen aproximadamente la línea sinuosa de algunos ríos aunque en realidad los píxeles que se validan con esos puntos se corresponden con la cobertura de la tierra en la ribera. Por último, el suelo desnudo presenta también confusión con otras coberturas: (a) con el agua, debido a la importante dinámica natural de los ríos y sus playones de suelo desnudo; y (b) con otras coberturas que pueden ser en realidad suelo desnudo durante un período corto de tiempo como parte de la dinámica temporal de estas coberturas (p.ej. momento entre una cosecha y la siguiente tanto en plantaciones forestales como agrícolas anuales y permanentes).

TABLA 7. Matriz de confusión y estimación del error de cada clase para el mapa de coberturas de la tierra generado para el año 1986

	Bosque	No bosque	Total	Exactitud usuario (%)
Bosque	2857	270	3127	0.91
No bosque	340	1925	2265	0.85
Total	3197	2195		
Exactitud productor (%)	0.89	0.88		
Exactitud total (%)				89

TABLA 8. Matriz de confusión y estimación del error de cada clase para el mapa de coberturas de la tierra generado para el año 2001

	Bosque	No bosque	Total	Exactitud usuario (%)
Bosque	3414	236	3650	0.94
No bosque	304	3503	3807	0.92
Total	3718	3739		
Exactitud productor (%)	0.92	0.94		0.93
Exactitud total (%)				93

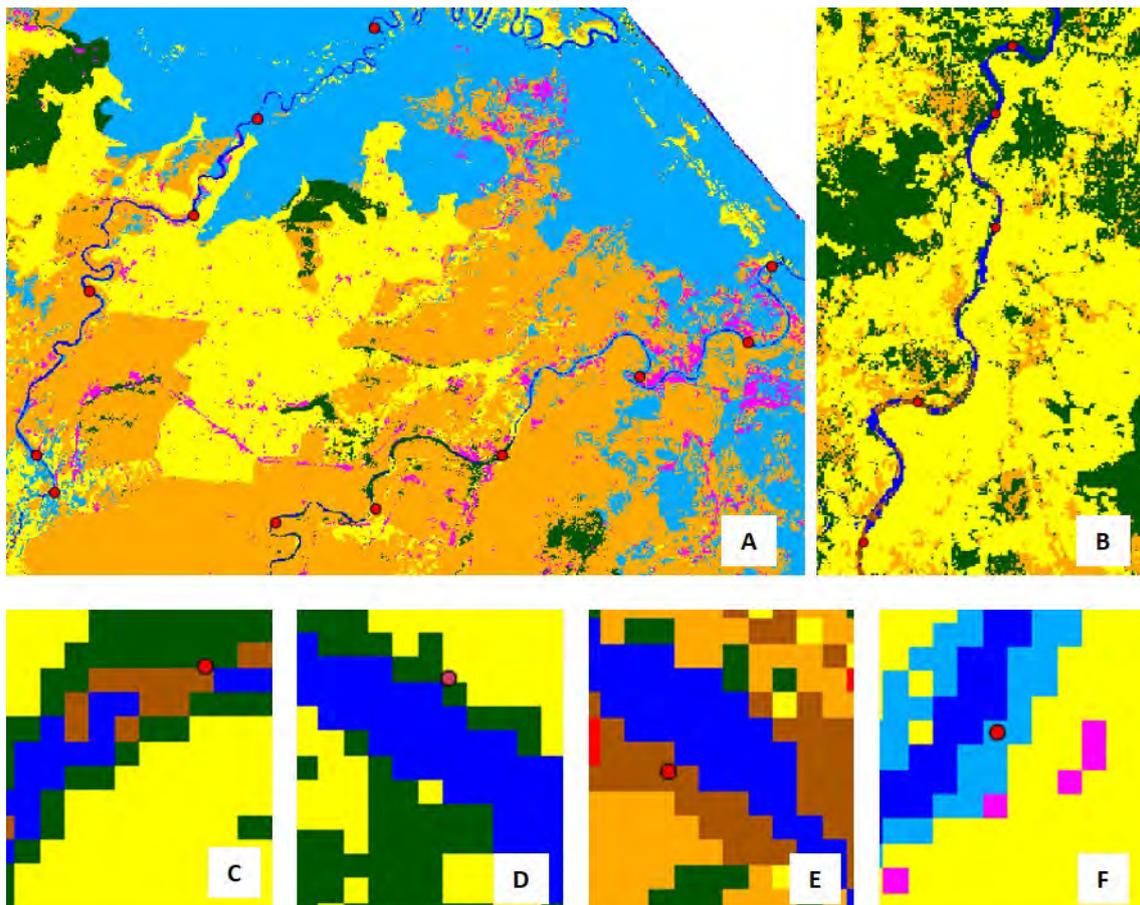


FIGURA 9. Detalle de algunos errores detectados en la validación de la clasificación. En A y B se observa como como los puntos de agua de INBio siguen aproximadamente la línea sinuosa de algunos ríos aunque en C, D, E y F se observa como en realidad lo que se valida con esos puntos se corresponde con la cobertura que existe en la ribera de ese río, bien por pequeños problemas de georreferenciación de los mapas y/o los puntos de validación o por la propia dinámica de los ríos y sus playones de suelo desnudo.

TABLA 9. Matriz de confusión y estimación del error de cada clase para el mapa de coberturas de la tierra 2014

	Bosque y plantaciones	Yolillal	Manglar	Áreas Urbanas	Pastizal	Páramo	Agua	Suelo desnudo	Cultivos anuales	Piña	Otros cultivos permanentes	Total	Exactitud usuario (%)
Bosque y plantaciones	2523	45	15	0	67	0	0	4	24	3	83	2764	0.91
Yolillal	0	151	0	0	1	0	0	0	0	0	0	152	0.99
Manglar	24	3	765	0	19	0	6	7	10	1	6	841	0.91
Áreas Urbanas	5	0	0	538	10	0	0	19	3	4	3	582	0.92
Pastizal	148	3	0	4	1210	0	0	30	122	49	16	1582	0.76
Páramo	14	0	0	0	0	88	0	1	1	0	0	104	0.85
Agua	23	19	11	0	16	6	196	18	18	5	12	324	0.60
Suelo desnudo	0	0	0	0	20	0	0	28	0	0	0	48	0.58
Cultivos anuales	3	0	0	0	61	0	1	36	798	10	1	910	0.88
Piña	18	0	0	0	3	0	0	2	0	530	15	568	0.93
Otros cultivos permanentes	20	1	0	0	12	0	0	2	12	1	609	657	0.93
Total	2778	222	791	542	1419	94	203	147	988	603	745		
Exactitud productor (%)	0.91	0.68	0.97	0.99	0.85	0.94	0.97	0.19	0.81	0.88	0.82	8532	
Exactitud total (overall accuracy) (%)													87

3.2. VALIDACIÓN DE MAPAS DE CAMBIO

Para la validación de la matriz de cambios de las clases bosque y no bosque entre los años 2001 y 2012 se siguieron las recomendaciones sugeridas por Olofsson et al. (2014), las cuales permiten, además de estimar la exactitud de los resultados, calcular una serie de estadísticos de exactitud y calcular las áreas ajustadas en función a los errores observados en la validación. Este documento de buenas prácticas para la estimación del área de cambio de cobertura y de su precisión se puede consultar en el **Anexo 5**.

Para realizar esta validación se ha realizado un muestreo aleatorio de 649 puntos: 315 en las áreas de bosque estable (áreas clasificadas como bosque en el año 2001 que se mantienen como bosque en el 2012), 223 en las de no bosque estable (áreas clasificadas como no bosque en el año 2001 que se mantienen como no bosque en el 2012), 63 en las áreas de bosques nuevos (áreas clasificadas como no bosque en el año 2001 que se clasifican como bosque en el 2012) y 48 en las de deforestación (áreas clasificadas como bosque en el año 2001 que se clasifican como no bosque en el 2012). Sobre esta muestra aleatoria de puntos se han analizado cambios en cobertura de la tierra a partir de las imágenes de alta resolución disponibles para posteriormente evaluar la precisión y desviaciones de las diferencias de cobertura arbórea estimadas.

La validación de estos resultados muestra una precisión general relativamente alta (0,95), aunque la exactitud en la estimación de las clases de bosques nuevos y deforestación es mucho menor a la de bosque estable y no bosque estable (Tablas 10 y 11). Los resultados de esta validación (resumidos en las Tablas 10 y 11) están detallados en el archivo Excel que se acompaña en esta entrega (**Anexo 10**).

La validación de las estimaciones realizadas de la deforestación (Tablas 10 y 11) muestra como el modelo tiende a sobreestimar ligeramente las áreas de deforestación (Tabla 12), de bosques nuevos y de bosque estable, mientras que la clase de no bosque estable parece estar ligeramente infraestimada por la clasificación. Así, siguiendo las recomendaciones propuestas por Olofsson et al. (2014) se ha calculado el área ajustada de cada una de las clases en función del área estimada inicialmente por el modelo y un coeficiente de ajuste que depende de la exactitud de cada una de las clases según la validación realizada. Se ha calculado el área ajustada de cada una de las clases en función del área estimada inicialmente por el modelo y un coeficiente de ajuste que depende de la exactitud de cada una de las clases según la validación realizada. De este modo, el área ajustada de **deforestación entre 2001 y 2012** sería de **285.503 ha**, el área de **bosques nuevos 320.245 ha**, el de **bosque estable 2.699.510 ha** y el de **no bosque estable 1.803.360 ha** (Tabla 10). Los errores relativos para cada una de las áreas han sido del 27%, 24%, 4% y 6% respectivamente.

TABLA 10. Resultados de la validación de los cambios de las clases bosque y no bosque entre los años 2001 y 2012. "Deforestación" incluye áreas clasificadas como bosque en el año 2001 que se clasifican como no bosque en el 2012. "Bosques nuevos" incluye áreas clasificadas como no bosque en el año 2001 que se clasifican como bosque en el 2012. "Bosque estable" incluye áreas clasificadas como bosque en el año 2001 que se mantienen como bosque en el 2012. "No bosque estable" incluye áreas clasificadas como no bosque en el año 2001 que se mantienen como no bosque en el 2012

Clase	Deforestación	Bosques nuevos	Bosque estable	No bosque estable	Total	Exactitud Usuario (%)
Deforestación	26	0	5	11	42	0.62
Bosques nuevos	1	36	4	7	48	0.75
Bosque estable	7	12	293	22	334	0.88
No bosque estable	10	7	11	192	220	0.87
Total	44	55	313	232	644	
Exactitud Productor (%)	0.59	0.65	0.94	0.83		
Exactitud total (Overall Accuracy)						89

TABLA 11. Cálculo de los estadísticos de la exactitud de la validación de los cambios de las clases bosque y no bosque entre los años 2001 y 2012, según las recomendaciones de Olofsson et al. (2014). "Deforestación" incluye áreas clasificadas como bosque en el año 2001 que se clasifican como no bosque en el 2012. "Bosques nuevos" incluye áreas clasificadas como no bosque en el año 2001 que se clasifican como bosque en el 2012. "Bosque estable" incluye áreas clasificadas como bosque en el año 2001 que se mantienen como bosque en el 2012. "No bosque estable" incluye áreas clasificadas como no bosque en el año 2001 que se mantienen como no bosque en el 2012

Clase	Exactitud Usuario	Intervalo de confianza 95%	Exactitud Productor	Intervalo de confianza 95%	Exactitud Total	Intervalo de confianza 95%
Deforestación	0.62	0.47 - 0.77	0.49	0.36 - 0.62		
Bosques nuevos	0.75	0.62 - 0.88	0.50	0.38 - 0.62		
Bosque estable	0.88	0.84 - 0.91	0.95	0.93 - 0.97		
No bosque estable	0.87	0.83 - 0.92	0.84	0.8 - 0.88		
Total					0.95	0.83 - 0.89

TABLA 12. Cálculo del área ajustada (y su intervalo de confianza) de los cambios de las clases bosque y no bosque entre los años 2001 y 2012, modificadas en función de la exactitud calculada en la fase de validación, según las recomendaciones de Olofsson et al. (2014). "Deforestación" incluye áreas clasificadas como bosque en el año 2001 que se clasifican como no bosque en el 2012. "Bosques nuevos" incluye áreas clasificadas como no bosque en el año 2001 que se clasifican como bosque en el 2012. "Bosque estable" incluye áreas clasificadas como bosque en el año 2001 que se mantienen como bosque en el 2012. "No bosque estable" incluye áreas clasificadas como no bosque en el año 2001 que se mantienen como no bosque en el 2012

Clase	Área estimada (ha)	Área ajustada (ha)	Error relativo al 95% de nivel de significación (%)	Intervalo de confianza al 95% (ha)
Deforestación	227.051	285.503	27%	208.612 – 362.395
Bosques nuevos	212.889	320.245	24%	242.490 – 398.001
Bosque estable	2.926.949	2.699.510	4%	2.580.893 – 2.818.127
No bosque estable	1.741.730	1.803.360	6%	1.687.039 – 1.919.680
Total	5.108.619	5.108.619		

3.3. ASEGURAMIENTO Y CONTROL DE LA CALIDAD

El sistema de aseguramiento de la calidad y de control de calidad (QA/AC por sus siglas en inglés) llevado a cabo en el proyecto ha contado con los siguientes puntos:

Fase de descarga y preparación de las imágenes

- (1) Validación por parte del técnico responsable de la descarga del soporte de datos, esto es la comprobación de la no existencia de errores de grabación o almacenamiento en los soportes digitales que impidan la lectura correcta de los datos.
- (2) Verificación inicial y básica por parte del técnico responsable de la descarga de la calidad de las imágenes mediante el análisis de los metadatos y del quicklook o vista previa remuestreada de cada imagen original.
- (3) Todas las imágenes descargadas han sido revisadas visualmente y validadas por los técnicos responsables de la fase de clasificación, diferentes de los técnicos responsables de la descarga.

Fase de ortorectificación

- (1) Determinación de la validez de los puntos de control utilizados mediante el error cuadrático medio, comprobación por parte del técnico responsable de la ortorectificación de que en ningún caso se encuentra por encima del tamaño del píxel de la imagen.
- (2) Realización por parte del técnico responsable de la ortorectificación de una inspección visual exhaustiva para asegurar de que no se ha producido ningún defecto en el proceso de generación de éstas: zonas duplicadas, estiramientos de píxeles o errores geométricos producidos por errores del modelo digital del terreno.
- (3) Realización de un control geométrico de dichas imágenes rectificadas mediante la toma de puntos de chequeo en cada escena repartidos regularmente sobre una malla. Comprobación de que el error medio cuadrático obtenido en los puntos de chequeo en todas las imágenes es inferior a 1 píxel (30 metros).
- (4) Elaboración de una "ficha de validación de georreferenciación", con una vista general de la imagen con los puntos de chequeo marcados sobre ella, y un listado de las coordenadas y errores obtenidos en cada punto, así como el cálculo del error medio cuadrático y máximo.

Fase de generación de máscaras de nubes y sombras

- (1) Comprobación visual de las máscaras de nubes y sombras de todas las imágenes mediante la comparación de éstas con la imagen original en RGB o falso color. Se ha prestado especial atención en la comprobación de las máscaras de nubes en áreas urbanas y líneas de costa con alta reflectancia ajustando durante el proceso de comprobación algunos de los parámetros de generación de máscaras nubes y sombras.
- (2) Validación en una muestra de 18 imágenes (6 por pasada) de las máscaras de nubes y sombras por técnicos diferentes a los responsables de su elaboración. La validación se ha realizado por medio de la comprobación visual de una malla sistemática aleatoria de puntos de chequeo.

Fase de clasificación

- (1) Realización para cada una de las pasadas de un proceso iterativo de clasificación, comprobación de la clasificación, detección de errores y revisión de las áreas y puntos de entrenamiento.
- (2) El proceso anterior se ha acompañado con la revisión de los estadísticos de ajuste y errores de los clasificadores RF utilizados que ha servido para la identificación de clases con necesidad de mejora de las áreas y puntos de entrenamiento. Los dos procesos anteriores han servido para la mejora progresiva de las áreas y puntos de entrenamiento del clasificador antes de la clasificación definitiva de las imágenes.
- (3) Comprobación visual y validación por parte de los técnicos responsables de la clasificación de cada imagen clasificada mediante la comparación de ésta con imagen de alta resolución disponible.

Fase de elaboración y validación de mapas finales

- (1) Comprobación visual de los mosaicos y validación de los vacíos de información mediante su comparación con las máscaras de nubes, sombras y fallos del sensor en cada una de las fechas de la serie. Este proceso ha sido realizado por los técnicos responsables de la clasificación.
- (2) Comprobación visual de los mapas generados tras el relleno de vacíos con datos globales.
- (3) Validación independiente de los mapas finales en tres de las fechas de la serie con muestras de puntos de validación proporcionadas por diversas instituciones del país no utilizadas en la fase de clasificación.

Fase de elaboración y validación de mapas de cambio

- (1) Comprobación visual de las principales áreas de deforestación y bosques nuevos del país entre años consecutivos de la serie para su validación y detección de posibles errores en la clasificación.
- (2) Validación de cambios de uso del suelo entre los años 2001 y 2012 mediante la fotointerpretación de cambios sobre una malla aleatoria de puntos y utilizando las imágenes Landsat correspondientes las fechas de validación, ortofotografía aérea del año 2005 e imágenes Rapideye de los años 2011 y 2012.

4. RECOMENDACIONES

En el futuro la incorporación de nuevas fuentes de información de detección remota (como son el Radar de Apertura Sintética o el Light Detection and Ranging) pueden contribuir a una mejora continua tanto de la clasificación de las coberturas y de sus cambios como de las emisiones que estos cambios suponen en el país.

El Radar de Apertura Sintética (SAR) tiene la ventaja, frente a la imagen óptica más convencional, de que no se ve afectado por la cobertura de nubes. La incorporación de Radar en la detección de coberturas puede suponer un ahorro computacional importante, reduciendo el número de imágenes a tratar, y además ser una alternativa más precisa a la utilización de productos globales en el relleno de vacíos de información.

Las nuevas técnicas de aprendizaje automático que consisten en el ensamblaje (o agrupación) de algoritmos de aprendizaje (p.ej., ensamblaje de redes neuronales, Random Forests, "bagging" y "boosting") generan buenos resultados con muestras relativamente pequeñas de entrenamiento. Este tipo de técnicas, combinadas con las parcelas de campo de Inventario Nacional Forestal (INF) y adicionalmente, si fuera posible con información LiDAR, son una alternativa interesante para la generación de mapas continuos de alta resolución de biomasa que faciliten y mejoren la estimación de emisiones y remociones debidas a todas las actividades de uso del suelo.

El ajuste de estos modelos no paramétricos requiere de buenas localizaciones de las parcelas de campo del Inventario Nacional Forestal (INF) que permitan relacionar exactamente las mediciones realizadas en campo con la información espectral correspondiente. Para construir modelos precisos que relacionen la información satelital con la mediciones de biomasa es recomendable trabajar con parcelas de mayor tamaño (1-2 ha) a las actualmente disponibles en el país (0,1 ha). Debido a que el levantamiento en campo de parcelas tan grandes es sumamente costoso, una solución interesante es generar parcelas de biomasa mayores a partir de vuelos LiDAR distribuidos de forma sistemática por todo el país. La información LiDAR genera un modelo tridimensional del bosque, describe con alta resolución la estructura de la vegetación y permite estimar con precisión la densidad de biomasa.

Las parcelas de campo del INF (con una mejora de su posicionamiento) son adecuadas para calibrar la información LiDAR y así generar parcelas de biomasa de mayor superficie, más adecuadas para la calibración de la imagen satélite disponible para todo el país.

Panamá aplicando una metodología similar generó un mapa de biomasa de alta resolución (1 ha de pixel) combinando 228 parcelas de campo (de entre 0,1 y 0,36 ha), información LiDAR distribuida de forma sistemática cubriendo el 4% del país e imágenes Landsat (Asner et al 2013).

5. REFERENCIAS BIBLIOGRÁFICAS

- Asner G.P., Mascaro J., Anderson C., Knapp D.E., Martin R.E., Kennedy-Bowdoin T., van Breugel M., Davies, S., Hall J.S., Muller-Landau H.C., Potvi C., Sousa W., Wright J., Bermingham E. (2013). High-fidelity national carbon mapping for resource management and REDD+. *Carbon balance and management*, 8(1), Pp 1-14.
- Berk, A., Anderson, G.P., Acharya, P.K., Bernstein, L.S., Muratov, L., Lee, J., Fox, M., Adler-Golden, S.M., Chetwynd, J.H., Hoke, M.L., Lockwood, R.B., Gardner, J.A., Cooley, T.W., Borel, C.C. and Lewis, P.E. 2005. MODTRAN 5: a reformulated atmospheric band model with auxiliary species and practical multiple scattering options. In: Shen, S.S.L., Paul, E. (Eds.), {SPIE} Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraespectral Imagery XI. Pp. 662-667.
- Breiman, L. (2001): Random Forests. *Machine Learning* 45 (1), pp 5–32
- Canty, M.J., Nielsen, A.A. 2008. Automatic Radiometric Normalization of Multitemporal Satellite Imagery with the Iteratively Re-weighted MAD Transformation. *Remote Sensing of Environment*. Vol. 112 No. 3 pp. 1025-1036.
- Chander G., Markham, B.L., Helder, D.L.. 2009. Summay of current calibration coefficients for Landsat MSS, TM, ETM+ and EO-1 ALI sensors.
- Hansen M. C., Potapov P. V., Moore R., Hancher M., Turubanova S. A., Tyukavina A., Thau D., Stehman S. V., Goetz S. J., Loveland T. R., Kommareddy A., Egorov A., Chini L., Justice C. O., Townshend J. R. G. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* 342, 850–853.
- Horning, N. 2013. Training Guide for Using Random Forests to Classify Satellite Images - v9. American Museum of Natural History, Center for Biodiversity and Conservation. Available from <http://biodiversityinformatics.amnh.org/>
- IPCC 2006, 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds). Published: IGES, Japan.
- Liaw A. and Wiener M. 2002. Classification and Regression by randomForest. *R News* 2(3), 18--22.
- Martinuzzi, S., Gould, W.A., González, O.M.R., 2007. Creating Cloud-Free Landsat ETM Data Sets in Tropical Landscapes: Cloud and Cloud-Shadow Removal. US Department of Agriculture, Forest Service, International Institute of Tropical Forestry Río Piedras, Puerto Rico.
- Olofsson P., Foody G.M., Herold M., Stehman S. V., Woodcock C.E., Wulder M. A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148 (2014) 42–57.