

# 2do FORO

## DE INGENIERÍA DE LA INFORMACIÓN

**Big Data:** Innovación desde la  
ingeniería de la información  
para transformar la sociedad

# Potenciación del análisis de impacto ambiental mediante un cubo de imágenes de satélite

Germán Bravo

# Participantes



- Harold Castro
- Eduardo Rosales
- Andrés Moreno
- Christian Ariza
- Mario Villamizar
- Juan Pablo Vinchira
- Ernesto Nobmann
- Óscar Gómez



- Saralux Valbuena
- Ederesson Cabrera
- Gustavo Galindo
- Pilar Lozano
- Salomón Ramírez

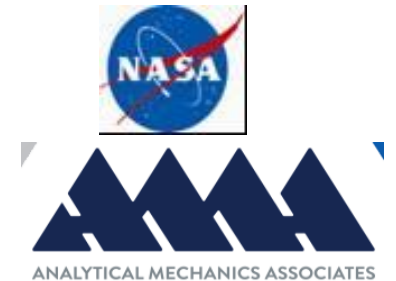


- Carlos Costa
- Mayerling Sanabria

- Alex Held



- Brian Killough
- Sanjay Gowda
- Jesse Harrison



# Análisis de impacto ambiental es Big Data

- Las imágenes de satélite son un insumo importante en este tipo de análisis
- Diversidad de sensores
  - Resolución espacial
  - Resolución espectral
  - Resolución temporal
- Cubrimiento de Colombia según sensor
  - Landsat
  - Sentinel
  - RapidEye
  - Spot
  - ...

Satellite	Bandas	Resolución Espacial (m)	Resolución Espacial Detallada	Tamaño de la Escena (km)	Resolución Espectral	Resolución Temporal (días)	Orbitas Anuales	Cubrimiento Colombia		TAMAÑOS PROMEDIO APROXIMADOS DE ARCHIVOS (MB)		Estimaciones del almacenamiento (GB)				
								Escenas	Tiles	Imagen preprocesada	Tile	Imágenes preprocesadas	Tiles	Productos	Imágenes anual por sensor	BCK Anual por Sensor
landsat4	4	68	68 m x 83 m	185 km x 185 km	Green, Red, Near-IR, Thermal-IR	18	21	65	131	220	22	587	60	18	0	0
landsat5	7	30	30 m, (120 m - thermal)	183 km x 172 km	Blue, Green, Red, Near-IR, Mid-IR, Thermal-IR, Mid-IR2	16	23	65	131	275	183	803	539	163	1505	1505
landsat7	8	30	30 m, (60 m - thermal, 15 m pan)	183 km x 170 km	Blue, Green, Red, Near-IR, Mid-IR, Thermal-IR, SWIR, Panchromatic	16	23	65	131	550	416	1606	1225	3713	6544	6544
landsat8	11	30	30 m (100 m - thermal, 15 m pan)	183 km x 170 km	Coastal aerosol, Blue, Green, Red, Near-IR, SWIR-1, SWIR-2, Panchromatic, Cirrus, Thermal-IR 1, Thermal-IR 2	16	23	65	131	1100	287	3212	845	255	4312	4312
MODIS	36	250	250 m (Blue, Green, and 3 IR 500 m, bands 9-36 1000 m)	2330 km <sup>2</sup>	Near-IR and Red, Blue, Green, and 3 IR bands, bands 9-36	2	183	65	131	55	15	118	352	12	482	482
Sentinel 1A	2	10	10m	400 km <sup>2</sup>	Polarisation (VV y VH) C-band Synthetic Aperture Radar	12	31	6	131	2200	468	9325	1857	418	11600	11600
Sentinel 1B	2	10	10m	400 km <sup>2</sup>	Polarisation (VV y VH) C-band Synthetic Aperture Radar	12	31	6	131	2200	468	9325	1857	418	11600	11600
SPOT	4	20	Multispectral (20 m), Panchromatic (10 m)	60 x 60 km	Green, Red, Near-IR, Panchromatic	5	73	543.1	131	110	235	8519	2195	209	0	0
SPOT4	5	20	Multispectral (20 m), Monospectral (10 m)	60 x 60-80 km	Green, Red, Near-IR, SWIR, Monospectral	5	73	342.6	131	55	293	4260	2737	261	0	0
RapidEye	5	5	5 m	77 km <sup>2</sup>	Blue, Green, Red, Red Edge, NIR	1	365	70	131	275	4668	67173	217970	4179	0	0
Sentinel 2A	13	10	10 m, 20m, 60 m	290 km <sup>2</sup>	Visible, Near-Infrared (VNIR), Shortwave Infrared (SWIR) wavelengths	5	73	26.04	131	1320	3035	4901	28344	2716	0	0
Sentinel 3A	21	500	500m	750 km <sup>2</sup>	Sea and Land Surface Temperature Radiometer (SLSTR) covering 9 spectral bands (550-12 000 nm), dual-view scan with swath widths of 1420 km (nadir) and 750 km (backwards)	1	365	6	131	55	3	236	141	2	0	0

ESTIMACIÓN ANUALES (TB)		
Imágenes	Bck	Total
36	36	72



## Centro de datos del IDEAM



16 Servidores

Blade

20 Procesadores

128 GB RAM

12 Procesadores

512 GB RAM

344 Núcleos

De procesamiento

40 TB

NAS - vBLOCK

240 TB Almacenamiento

NAS - ISILON X210

40 Gbps Conexión

FIBRA OPTICA

20 Work Stations

PDI

# Proceso de análisis

- Escoger las imágenes
  - Costo (Tiempo y dinero)
- Preprocesar
  - Corrección geométrica
  - Corrección radiométrica
  - Corrección atmosférica
- Procesar
  - Algoritmos (parámetros) específicos para cada problema
- Producir el resultado
  - Muchas veces depende del criterio de los expertos
- Se hace imagen por imagen
- Cada entidad lo hace independientemente
  - Falta de conocimiento de lo que ya está hecho
  - Falta de confianza en los resultados de los demás
- Consecuencia
  - Replicación de costos
  - Replicación de datos
  - Replicación de procesos
  - Replicación de resultados
    - No necesariamente comparables

# El cubo de datos

- Repositorio de imágenes y algoritmos que permite
  - Tener un banco único de imágenes
    - Multi sensor
    - Pixeles co-registrados
  - Tener un banco único de algoritmos (calificados)
    - Descarga
    - Preproceso
    - Ingesta (para cada sensor)
    - Proceso
    - Postproceso
  - Organizado según los procesos aplicados a las imágenes
    - Linaje de la información
  - Disponible para todas las entidades que utilicen imágenes de satélite como insumo



- [Australian Geoscience Data Cube.mp4](#)
- <https://youtu.be/DZ40HNq47ro>

Desde 2016  
Trimestral

Alertas Tempranas



Cuantificación Nacional de Deforestación

- Superficie de Bosque
- Superficie de Cambio en el Bosque

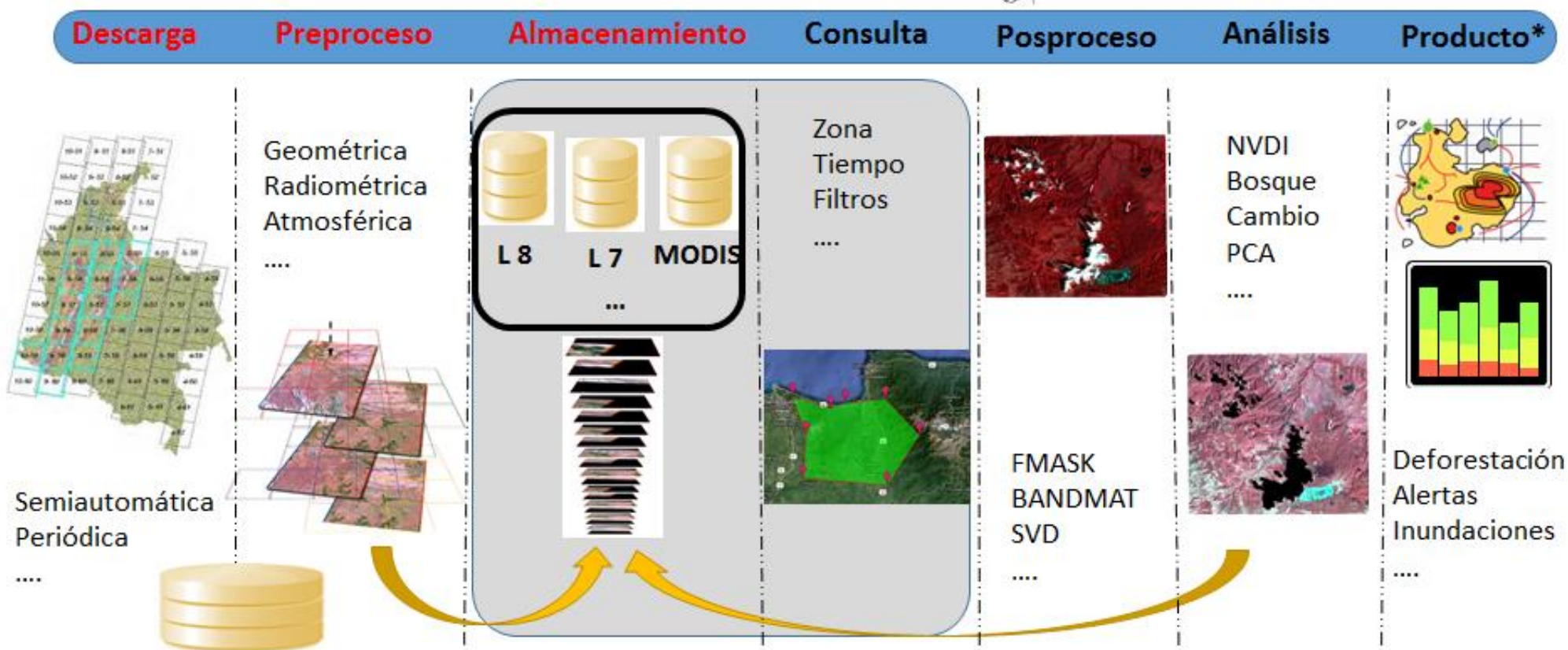
Tasa de deforestación

Desde 2013  
Anual



# SISTEMA DE MONITOREO DE BOSQUES Y CARBONO (SMBYC) COLOMBIA

# Visión global de CDCol



# Historia

- AGDC – V1 – 2013
  - Australia
- AGDC – V2 – 2015
  - Australia
  - Diseño mejorado
  - En desarrollo
- NGDC – V1.5 – 2015
  - Kenya DC

## CDCoI 2015

Apropiación y transferencia de tecnología

Basado en NGDC-v1.5

Cobertura

Sobre un minicubo de 2º x 2º

Landsat 7

2000 a 2015

Algoritmos

Medianas

Índice de vegetación (NDVI)

Máscara de Bosque/No bosque

Detección de cambio mediante PCA



## Historia

- CDCol 2016
  - Basado en AGDC-v2
  - Desarrollo de una interfaz WEB
  - Gestión de Unidades de almacenamiento
  - Desarrollo del banco de algoritmos
  - Gestión de roles de usuario
    - Administrador del sistema
    - Administrador de datos
    - Desarrollador
    - Analista
    - Público
- Cobertura
  - Toda Colombia
  - Landsat 5, 7 y 8
  - 2000 a 2016
- Algoritmos
  - Medianas
  - NVDI
  - Máscara de Bosque/No Bosque
  - Detección de cambio mediante PCA
  - WOFS (Water Observations from Space)
  - Serie de tiempo de NVDI



**Administrador de Datos**



**Desarrollador**



**Analista**



**django**

django  
**REST**  
framework

**C** Celery

**RabbitMQ**  
Open Source Enterprise Messaging



**Administrador del Sistema**



Plataforma de Fácil Uso

Centralización y estandarización de gran diversidad de datos de OT.



Aplicaciones de OT

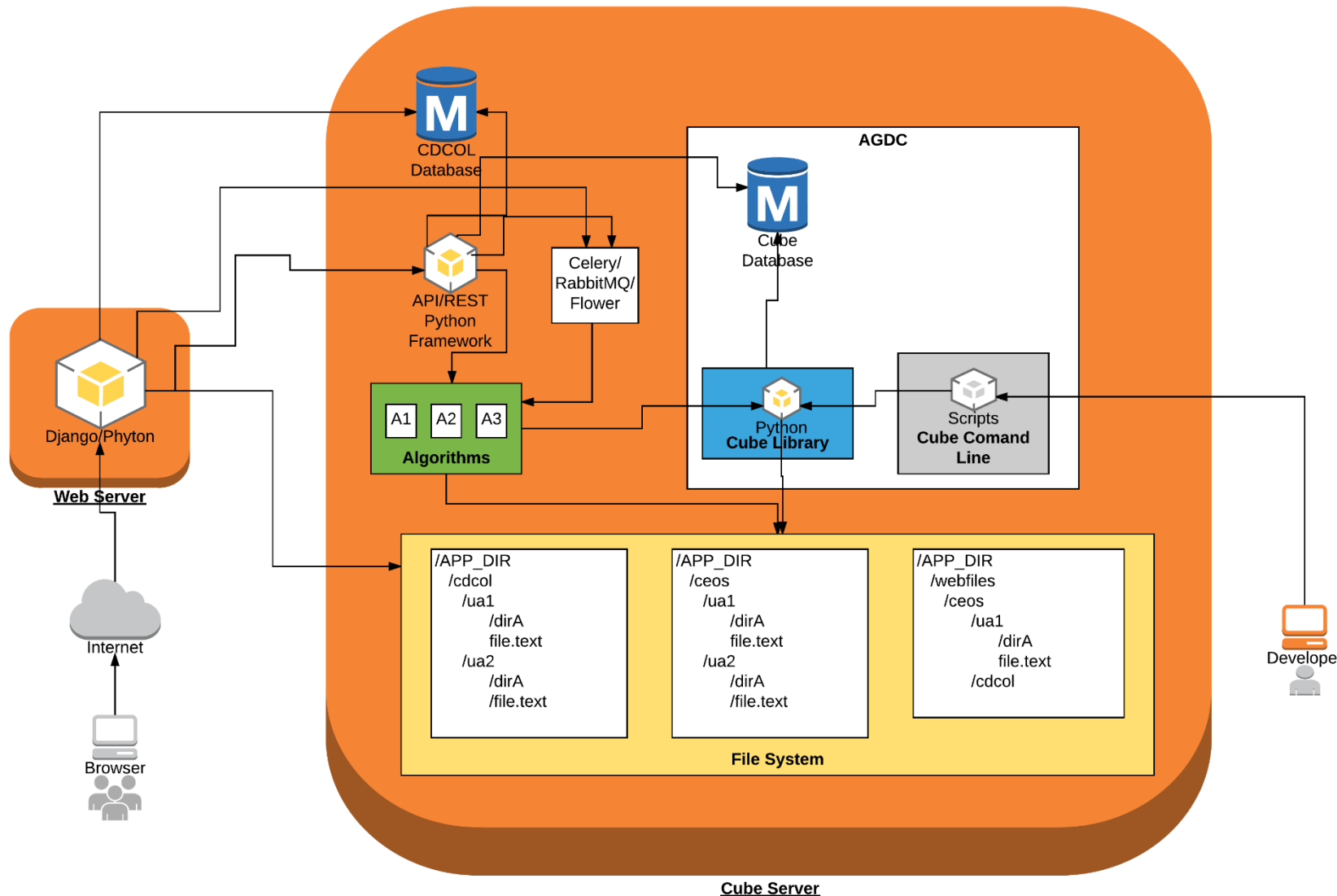
Generación de productos de información comparables basados en PDI.



Big Data

Grandes volúmenes de datos manejados eficientemente.

# Arquitectura CDCol Despliegue





## Información del algoritmo

La siguiente información del algoritmo:

Temática:	CLASIFICACIONES
Nombre:	Bosque - No bosque
Versión a ejecutar:	4 - Bosque - No bosque 1.0 ▾
Fecha de creación:	13 de Diciembre de 2016 a las 18:44
Creada por:	ef.nobmann10@uniandes.edu.co

## Calificación de la versión

Estadísticas de la versión del algoritmo:

 Calificación 5.0/5	 # de Calificaciones 1	 # de Ejecuciones 2
--	---	--


[Ver Calificaciones](#)

## Ejecutar algoritmo

Ingresa los siguientes parámetros para ejecutar el algoritmo:

**Descripción de la ejecución**

**Mapa** + ⓘ



Latitud S       Latitud N  
 Longitud W       Longitud E

**Periodo de consulta** + ⓘ

Desde       Hasta  
 Periodo

**Posibles unidades de almacenamiento origen** + ⓘ

Unidad de almacenamiento

**Número mínimo** + ⓘ

Número mínimo

**Normalizado** + ⓘ

Normalizado

**Tasa de vegetación** + ⓘ

Tasa de vegetación

**Umbral de bosque** + ⓘ

Umbral de bosque

**Tamaño de la ventana** + ⓘ

Tamaño de la ventana

[Ejecutar Algoritmo](#)



## Información del Algoritmo

La siguiente es la información del algoritmo:

Nombre:	Compuesto temporal de medianas
Tipo:	COMPUESTOS TEMPORALES
Descripción:	Calcula para cada pixel la mediana en la dimensión del tiempo.
Fecha de Creación:	13 de Diciembre de 2016 a las 16:59
Creada por:	ef.nobmann10@uniandes.edu.co

## Versiones del algoritmo

A continuación se encuentran las versiones del algoritmo:

[+ Crear nueva versión](#)

Número	# Parámetros	Estado	Fecha de creación	Acciones
1.0	5	EN DESARROLLO	13 de Diciembre de 2016 a las 16:59	<a href="#">Ver Detalle</a> <a href="#">Editar</a>

## Parámetros de la versión

A continuación se encuentran los parámetros de esta versión::

[+ Nuevo Parámetro](#)

Nombre	Tipo	Posición	Obligatorio	Habilitado	Incluido en Salida	Acciones
Unidad de Almacenamiento con Bandas	UNIDAD ALMACENAMIENTO CON BANDAS	1	True	True	True	<a href="#">Ver Detalles</a> <a href="#">Editar</a>
Mapa	AREA	2	True	True	True	<a href="#">Ver Detalles</a> <a href="#">Editar</a>
Número mínimo de pixeles válidos	INTEGER	3	True	True	True	<a href="#">Ver Detalles</a> <a href="#">Editar</a>
Fecha	PERIODO DE TIEMPO	4	True	True	True	<a href="#">Ver Detalles</a> <a href="#">Editar</a>
Normalizado	BOOLEAN	5	True	True	True	<a href="#">Ver Detalles</a> <a href="#">Editar</a>

## Nuevo Parámetro

Ingrese la siguiente información del parámetro:

**Nombre del parámetro** \* ?

**Tipo de parámetro** \* ?

**Descripción del parámetro** \* ?

**Texto de ayuda del parámetro (opcional)** ?

**Posición** \* ?

**Este parámetro es obligatorio** ?

Si el parámetro es opcional desmarcar esta casilla.

**Este parámetro está habilitado** ?

Para deshabilitar este parámetro, desmarcar esta casilla.

[Crear Parámetro](#)

## Unidades de Almacenamiento

A continuación se presentan las unidades de almacenamiento actuales

[+ Nueva unidad](#)

Mostrar **10** registros Buscar:

Nombre	Fecha de Creación	Creada por	Acciones
LS8_OLI_LEDAPS	9/12/2016 12:25:18	superadminuser	<a href="#">Ver Detalle</a>
LS7_ETM_LEDAPS	9/12/2016 12:24:00	superadminuser	<a href="#">Ver Detalle</a>
LS5_TM_LEDAPS	9/12/2016 12:22:32	superadminuser	<a href="#">Ver Detalle</a>

Mostrando registros del 1 al 3 de un total de 3 registros Anterior **1** Siguiente

## Contenido de la Unidad

A continuación se presenta el contenido de la unidad de almacenamiento.

Mostrar **10** registros Buscar:

Nombre	Acciones
LS7_ETM_LEDAPS_4326_-70_3_20000116143438000000.nc	<a href="#">Descargar</a> <a href="#">Ver Detalle</a>
LS7_ETM_LEDAPS_4326_-70_3_20000217143432000000.nc	<a href="#">Descargar</a> <a href="#">Ver Detalle</a>
LS7_ETM_LEDAPS_4326_-70_3_20000217143456000000.nc	<a href="#">Descargar</a> <a href="#">Ver Detalle</a>
LS7_ETM_LEDAPS_4326_-70_3_20000304143427000000.nc	<a href="#">Descargar</a> <a href="#">Ver Detalle</a>
LS7_ETM_LEDAPS_4326_-70_3_20000320143427000000.nc	<a href="#">Descargar</a> <a href="#">Ver Detalle</a>

## Nueva Unidad de Almacenamiento

Ingrese la siguiente información de la unidad de almacenamiento:

**Archivo de ingesta \*** [?](#)  
 Ningún archivo seleccionado  
 Suba el archivo .yaml con la información para poder realizar la ingesta [Ver plantillas](#)

**Nombre de la unidad** [?](#)

**Descripción** [?](#)

**Archivo de descripción \*** [?](#)  
 Ningún archivo seleccionado  
 Suba el archivo .yaml con la descripción de la unidad [Ver plantillas](#)

**Script de generación de metadatos \*** [?](#)  
 Ningún archivo seleccionado  
 Suba el archivo .py que genera los metadatos de las imágenes a ingestar

[Crear Unidad](#)

Sitio administrativo

ALGORITHM	
Algorithms	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Parameters	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Topics	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Version storage units	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Versions	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

AUTENTICACION Y AUTORIZACION	
Grupos	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Usuarios	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

EXECUTION	
Area types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Boolean types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Double types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Execution parameters	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Executions	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
File types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Integer types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Reviews	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Storage unit band types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Storage unit no band types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
String types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Tasks	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
Time period types	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

INGEST	
Ingest tasks	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
INGEST_TEMPLATE	
Ingest templates	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
SITIOS	
Sitios	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
STORAGE	
Storage units	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
TEMPLATE	
Yaml templates	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>
USER_PROFILE	
User profiles	<a href="#">+ Añadir</a> <a href="#">✎ Modificar</a>

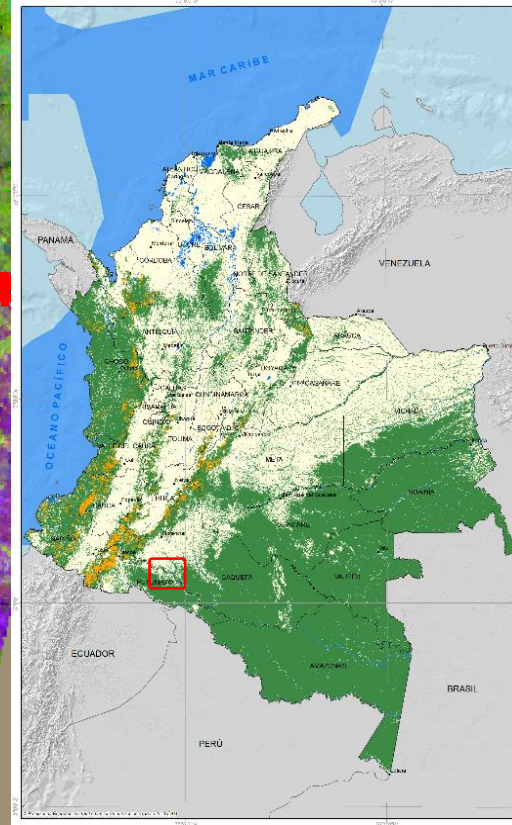
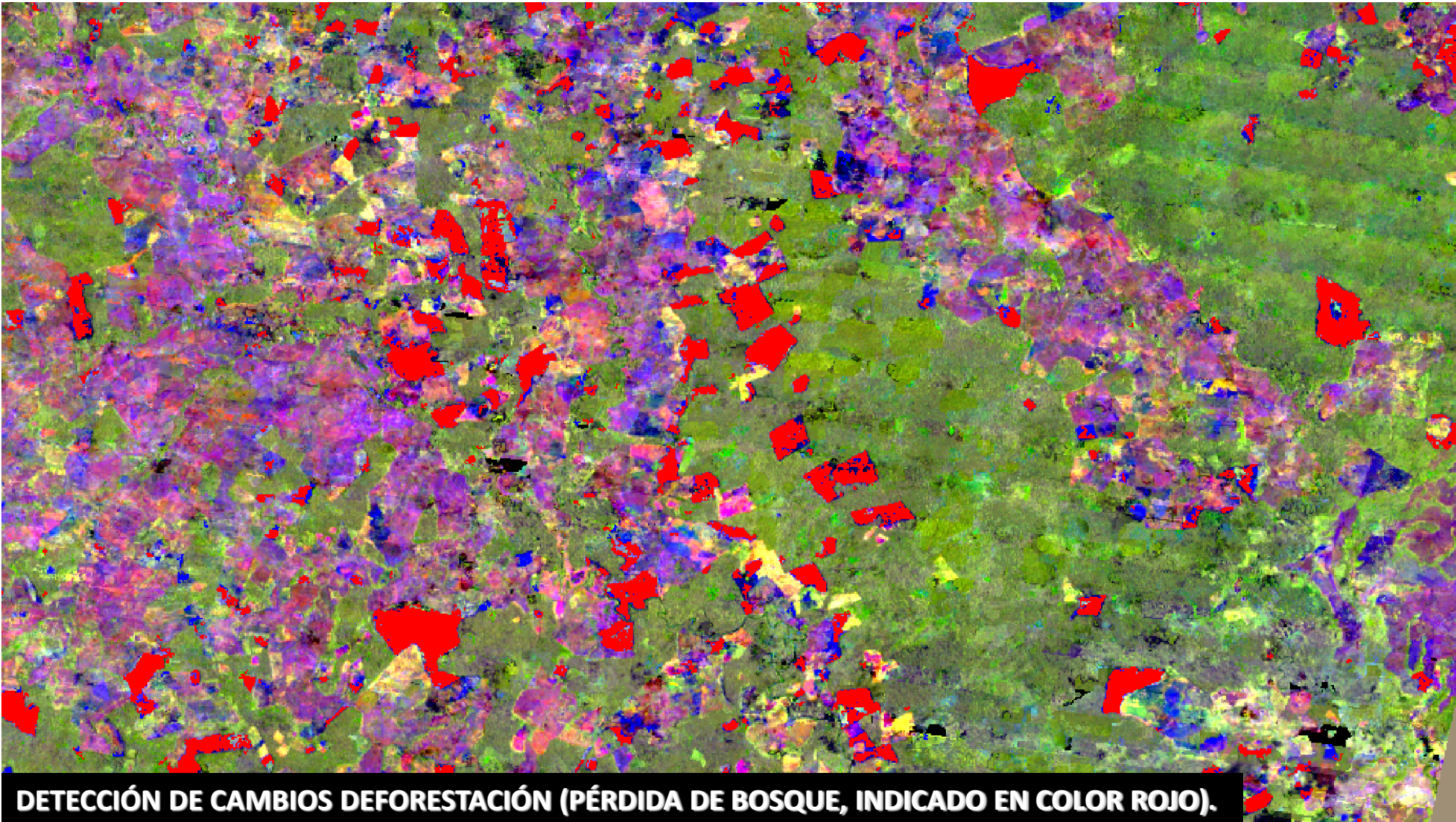
Escoja user profile a modificar

Acción:   1 de 3 seleccionado

		LOCATION	PHONE	STATUS	CREATED AT
<input type="checkbox"/>	FULL				
<input checked="" type="checkbox"/>	Mar	Universidad de los Andes	3106806131	POR APROBACION	26 de Enero de 2017 a las 01:32
<input type="checkbox"/>	erne	les	123	APROBADO	13 de Diciembre de 2016 a las 20:18
<input type="checkbox"/>		omandes	311111	APROBADO	12 de Diciembre de 2016 a las 10:20

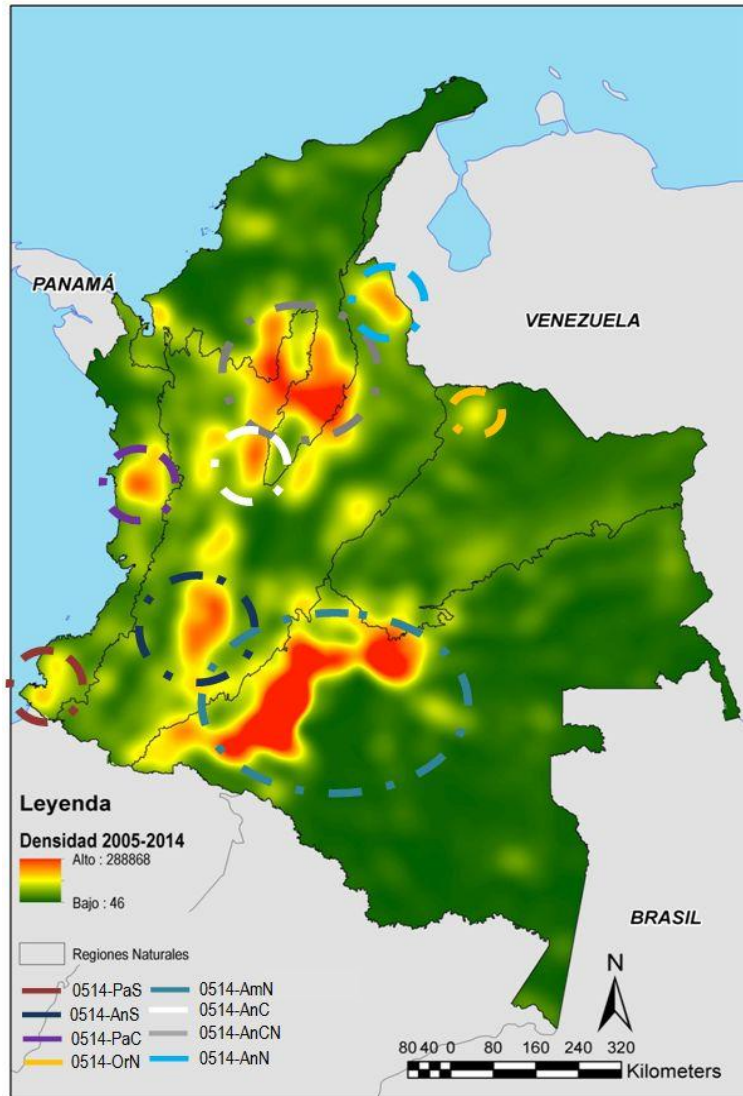
3 user profiles



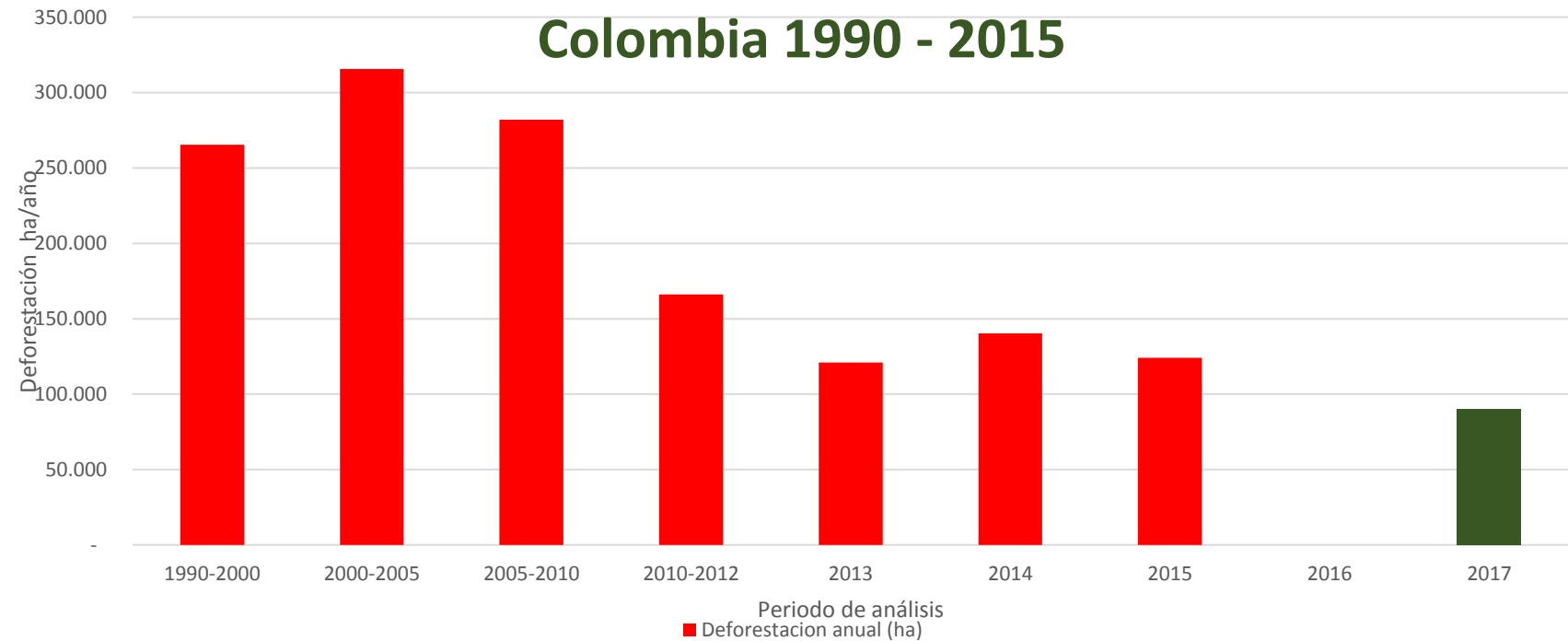


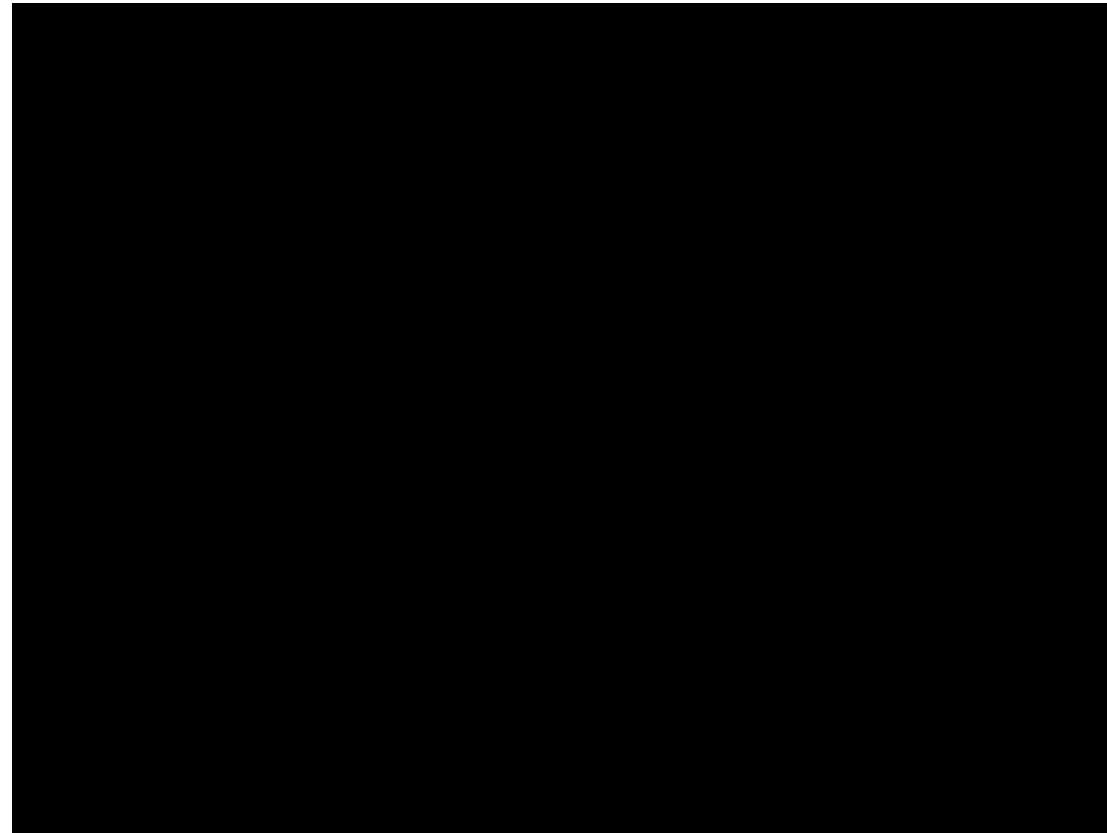
**DETECCIÓN DE CAMBIOS DEFORESTACIÓN (PÉRDIDA DE BOSQUE, INDICADO EN COLOR ROJO).**



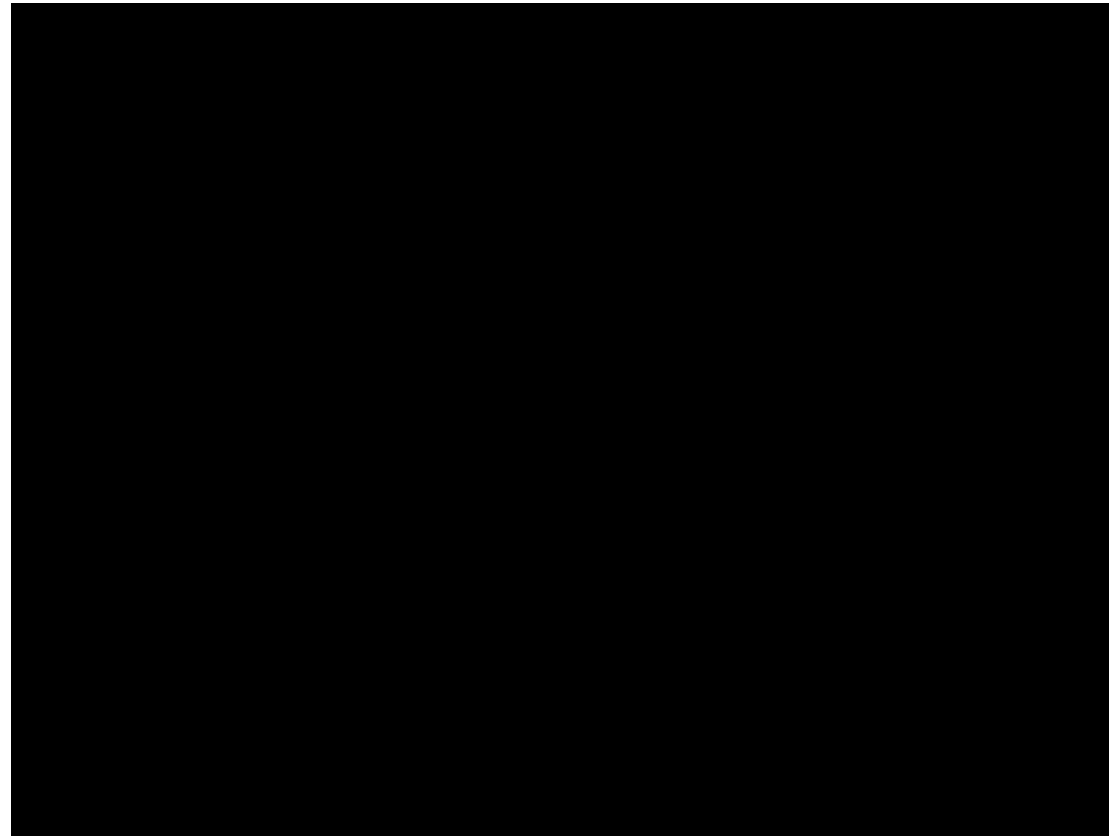


## Monitoreo de la superficie deforestada en Colombia 1990 - 2015

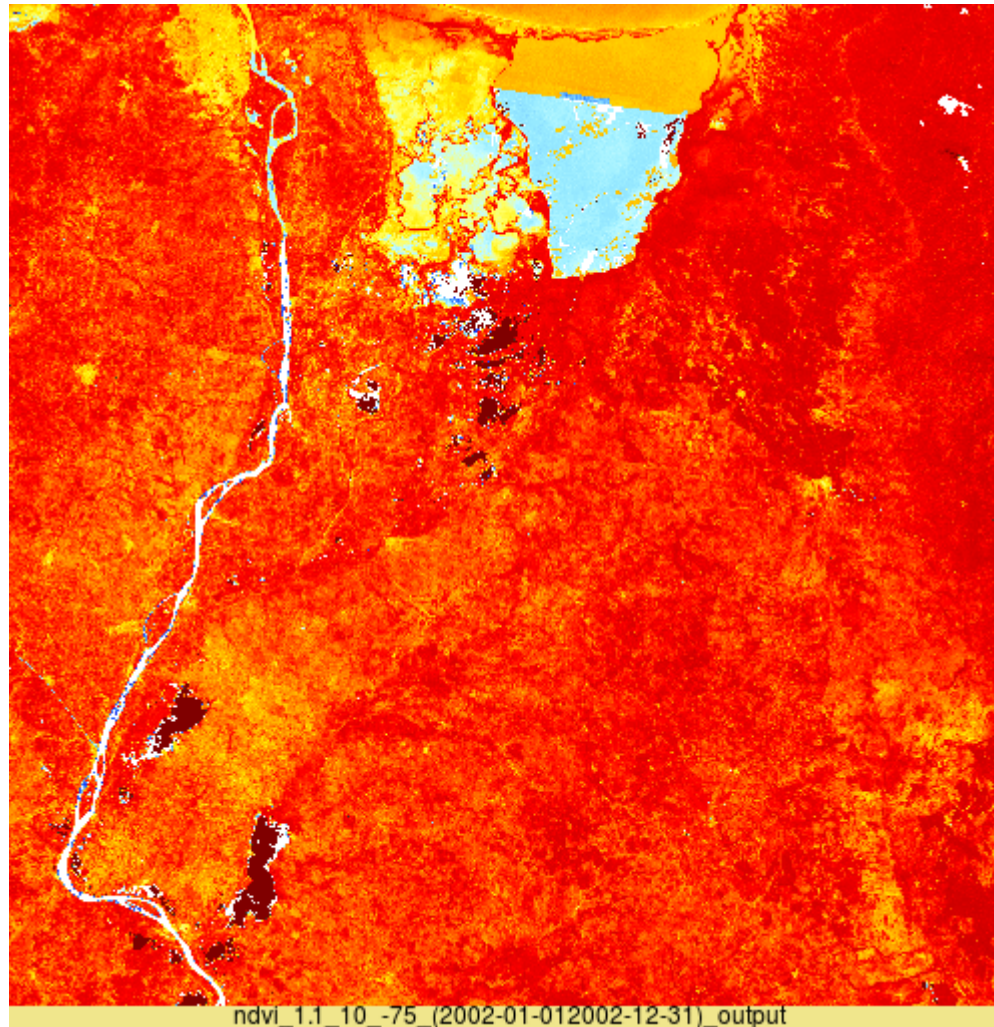




Compuesto temporal de medianas 2000-2014



Clasificación Bosque/No Bosque 2000-2014

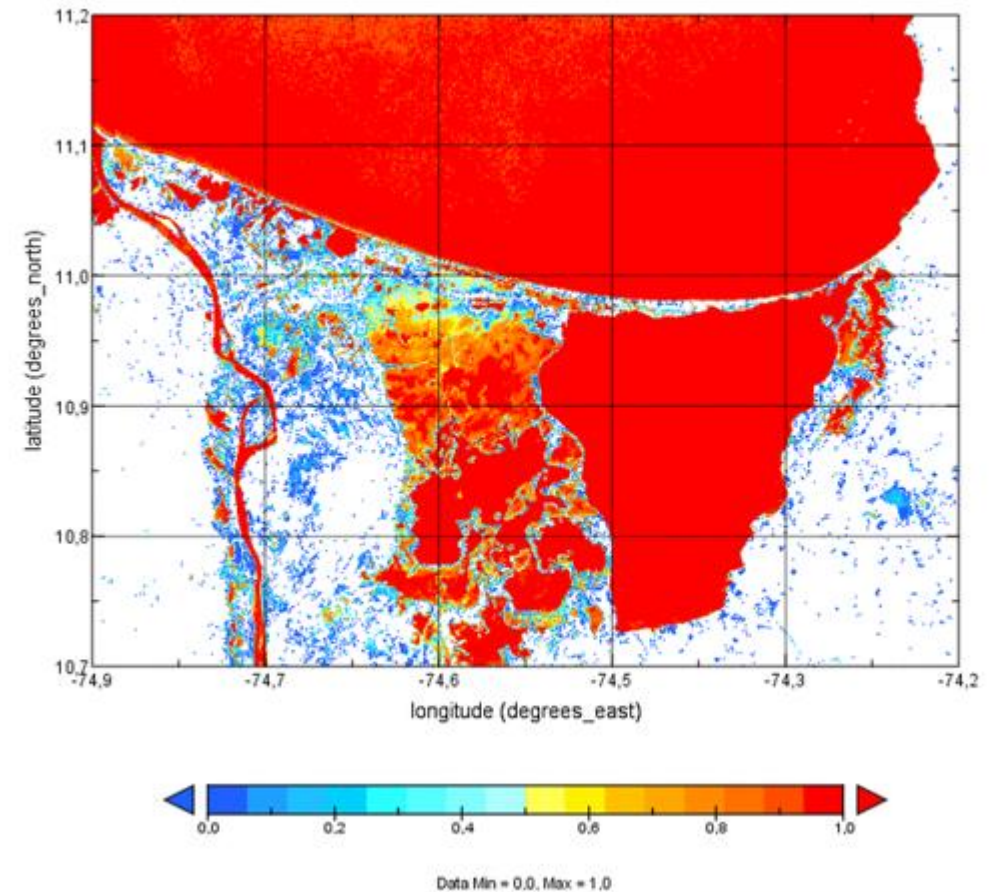


NDVI 2002-2013



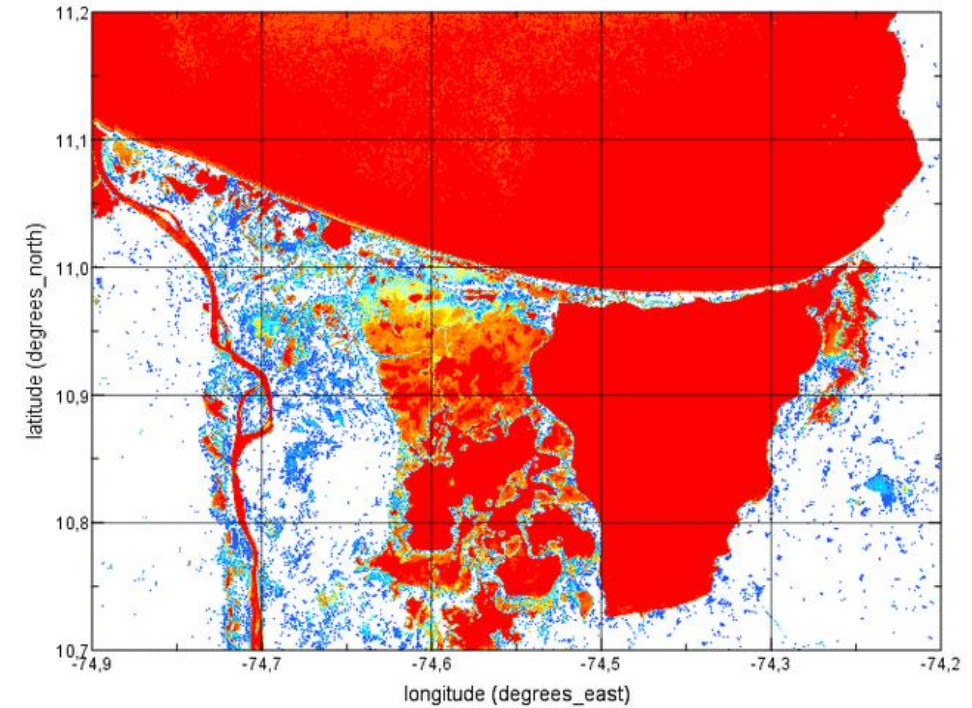
## Aplicaciones: cuerpos de agua

<b>15 años</b> DATOS DE 2000-2015	<b>30 metros</b> RESOLUCIÓN DE PIXEL
<b>342 escenas</b> LANDSAT 7	<b>329 GB</b> FILES READY TO USE - NETCDF
<b>3078 tiles</b> LANDSAT 7 - CUBO DE DATOS	<b>117 GB</b> FILES READY TO USE - NETCDF
<b>2h</b> PROCESAMIENTO	



## Aplicaciones: cuerpos de agua

<b>15 años</b> DATOS DE 2000-2015	<b>30 metros</b> RESOLUCIÓN DE PIXEL
<b>342 escenas</b> LANDSAT 7	<b>329 GB</b> FILES READY TO USE - NETCDF
<b>3078 tiles</b> LANDSAT 7 – CUBO DE DATOS	<b>117 GB</b> FILES READY TO USE - NETCDF
<b>2h</b> PROCESAMIENTO	



Data Min = 0.0, Max = 1.0

# Perspectivas

- Ampliar el banco de algoritmos
- Banco de flujos de trabajo
- Mayor cobertura de sensores
  - Con el apoyo de la NASA
- Modificar la arquitectura de CDCOL para escalamiento horizontal
- Optimizar la infraestructura computacional nacional
  - Almacenamiento
  - Procesamiento
  - Red
- Apropiación y uso por parte del IDEAM
- Apropiación y uso por parte de las entidades del SINA y otras entidades que se nutran de imágenes de satélite
- Difusión para el mundo entero
  - Publicar como Open source
  - Presentación en foros, conferencias y congresos
- Forest 2020
  - Patrocinado por el Reino Unido
  - IDEAM, UNIANDES, TNC, Ecometrica
  - Iniciativa mundial

**GRACIAS por su atención**

**Preguntas ?**