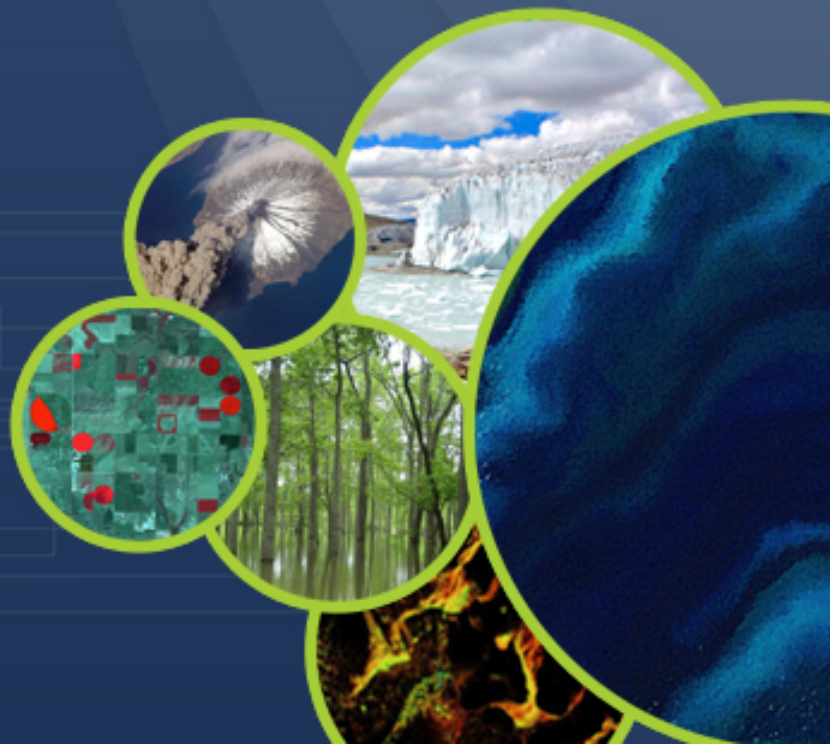
CEOS WGISS Tech Expo Webinar

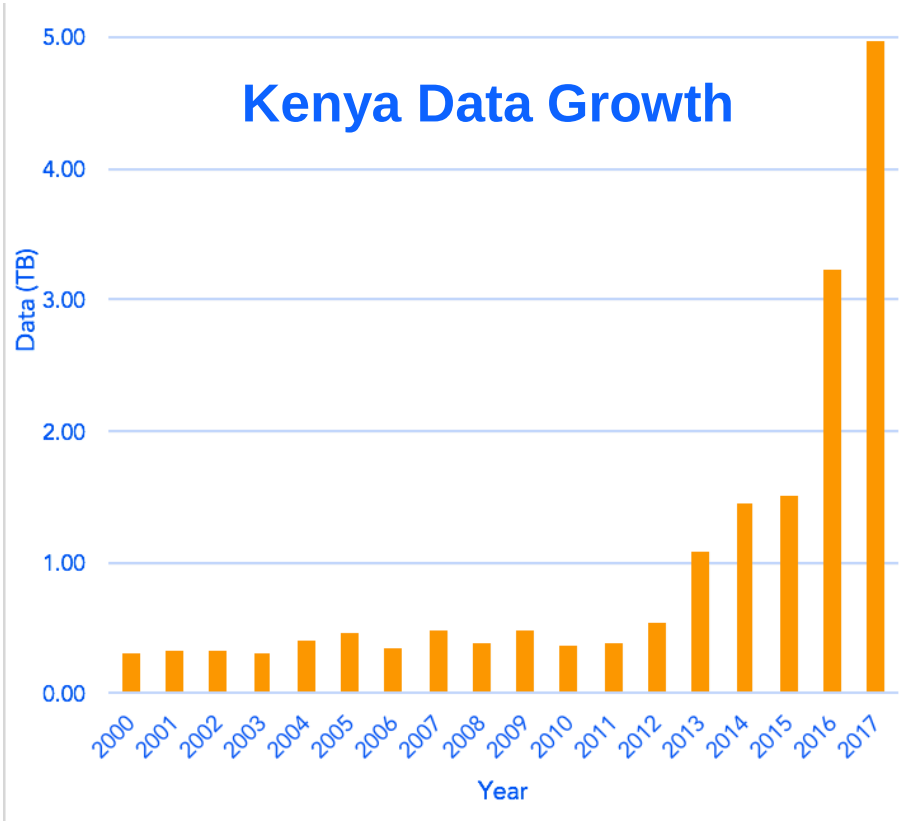
Data Cubes for Large Scale Data Analytics

June 19, 2017

Brian Killough
CEOS Systems Engineering Office
NASA Langley Research Center
Email: Brian.D.Killough@nasa.gov



Solving a Problem

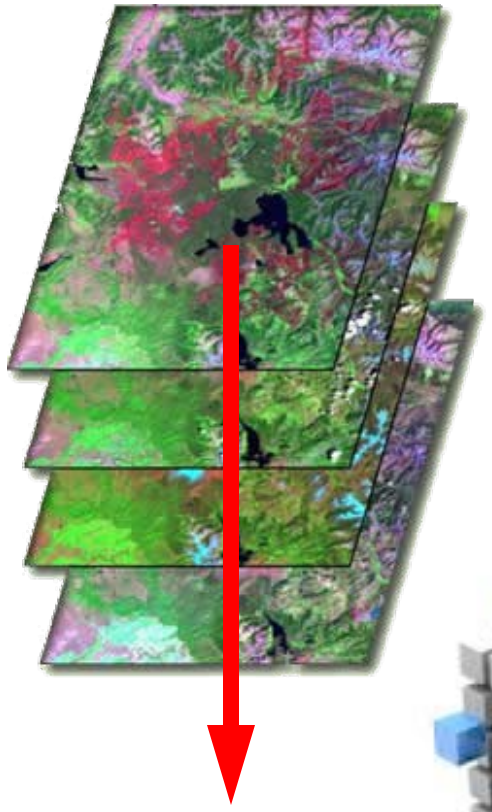


- A significant growth in FREE/OPEN land imagery data (e.g. Landsat, Sentinel) will **increase data volumes** by 10x in the next few years.
- Most developing countries **lack the knowledge, infrastructure, and resources** to access and use space-based data.
- Countries have requested support from CEOS for data access, processing, and analysis. They **want to learn** how to use satellite data to support their country needs.

*The new **Open Data Cube** provides a solution and new opportunities*

What are Data Cubes?

- **Data Cube** = Time-series multi-dimensional (space, time, data type) stack of spatially aligned pixels ready for analysis
- **Proven concept** by Geoscience Australia (GA) and the Australian Space Agency (CSIRO) and planned for the future USGS Landsat archive.
- **Analysis Ready Data (ARD) ...** Dependent on processed products to reduce processing burden on users
- **Open source** software approach allows free access, promotes expanded capabilities, and increases data usage.
- **Unique features:** exploits time series, increases data interoperability, and supports many new applications.



Why Data Cubes?

The primary user problems are data access, data preparation, and efficient analyses

- Users want to minimize the time and knowledge required to obtain and prepare satellite data
- Users want free and open source solutions.
- Users want to perform time series analyses
- Users want to use multiple datasets
- Users want to use common GIS tools
- Users want to “own” the data and keep it locally
- Users want customer service and support

- Our goal is **NOT** to sell a product or give out a tool …
- Our goal is to provide a **SOLUTION** that has **VALUE** and increases the **IMPACT** of satellite data.

CEOS Data Cube Vision

A solution supporting CEOS objectives …

- The CEOS Data Cube is an implementation of the Open Data Cube
- Build capability of users to apply CEOS satellite data
- Supporting priority CEOS/GEO agendas

CEOS Agencies wanting to participate …

- Through provision of CEOS Analysis Ready Data (ARD) products
- Contributing to development and uptake of solutions

Customers feel that they are the focus …

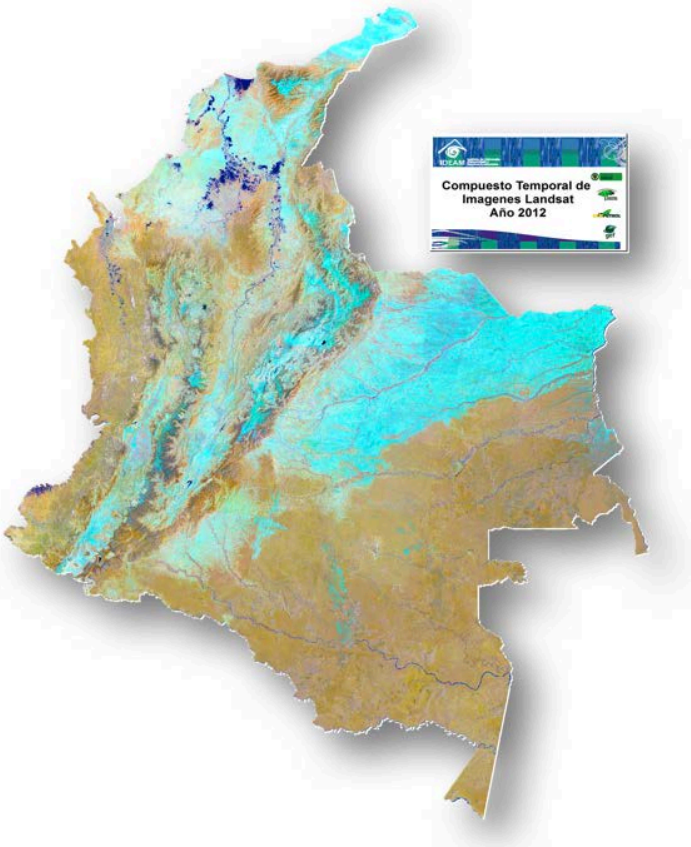
- Training materials and easy installation/maintenance
- An "Open Data Cube" brand that people know and trust
- Users helping each other through an active Open Data Cube community

Scalable solution …

- Operational Data Cubes in **20 countries by 2022**
- Key partners (e.g. GEO, World Bank) supporting data cube projects

Colombia Prototype

- The Government (IDEAM) and University (Andes) have made considerable progress in learning how to create and use Data Cubes!
- A complete country-level Landsat Data Cube (25,000 scenes back to year 2000) was completed in Dec 2016.



- Forest mapping and land change detection are the primary application needs.
- Future plans to add more datasets and applications.
- The SEO will continue to support IDEAM and the University of Andes in 2017+.

Switzerland Prototype

- SEO was approached by UNEP GRID Geneva and the Univ. of Geneva to develop a Data Cube pilot project. Significant computing and programming resources exist, so little effort was needed to get them started.
- UNEP GRID Geneva has made excellent progress installing a small data cube within the country and attaching the application user interface. They continue to expand their data and to learn and use new applications.
- The group plans to help Moldova and Georgia install a Data Cube as a capacity building activity to support wetlands and forest defoliation.








The “Road to 20” National-scale Data Cubes by 2020

3 operational, 4 under development, 16 under review

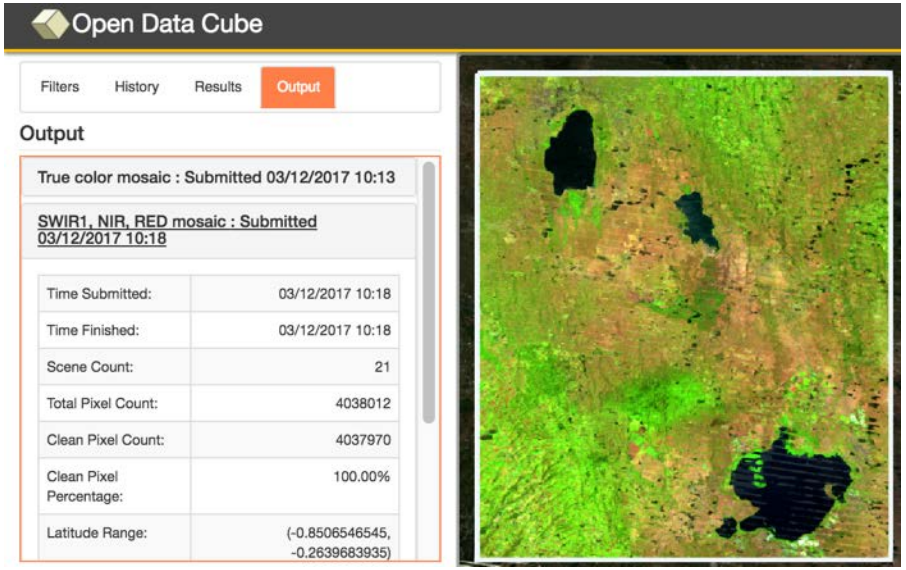
Amazon Cloud (AWS) Data Cube Demo Portal

Data Cubes

- 15 cubes with 10+ years each.
- Kenya, Cameroon (Lake Chad), Togo (coastal Africa), Ghana, Colombia, Tonga (Pacific Island), Vietnam, Australia (Menindee Lakes), Bangladesh.

User Interface Features

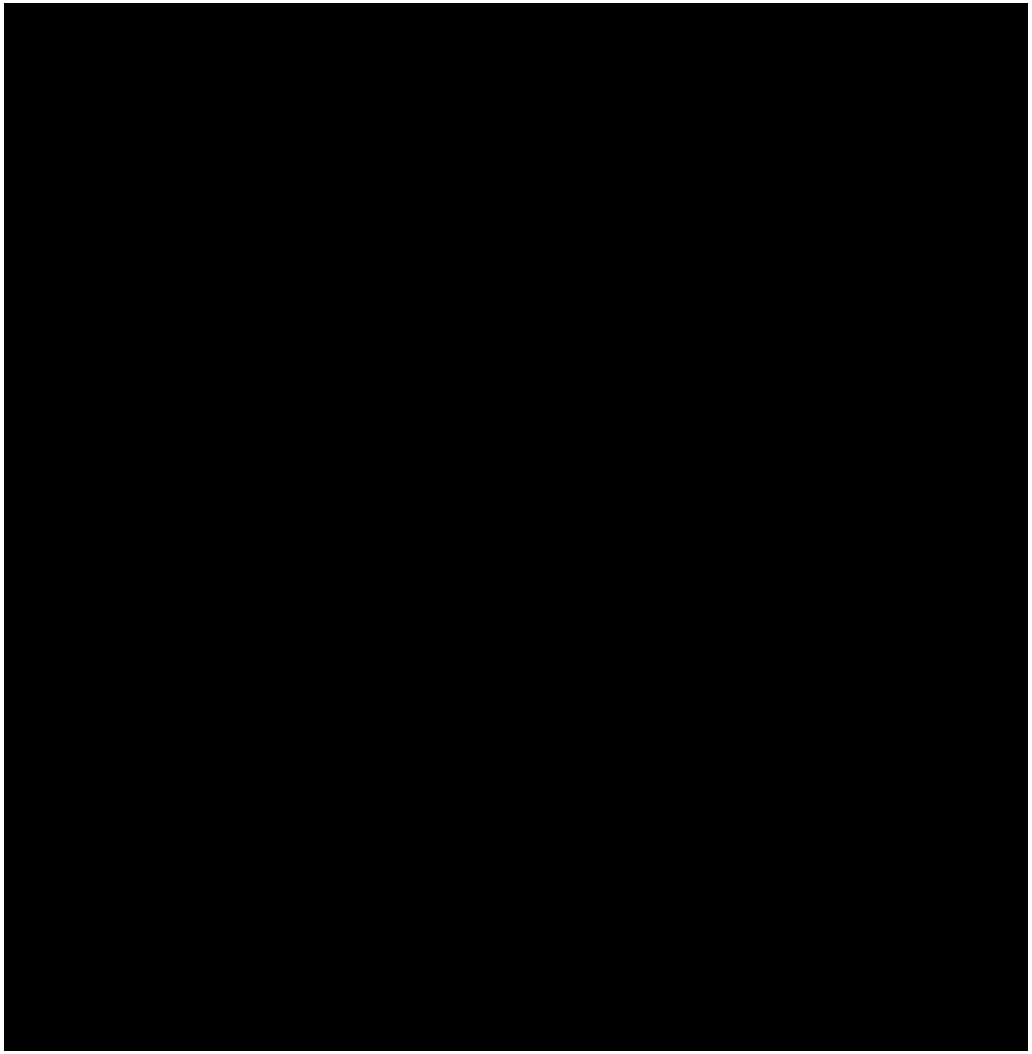
- User-selected spatial region and time
- 7 applications: custom cloud-free mosaics, fractional cover, NDVI anomaly, water detection, water quality (Total Suspended Matter), landslides (SLIP) and coastal change.
- Outputs in GeoTIFF and GIF animation.
- Free and open!

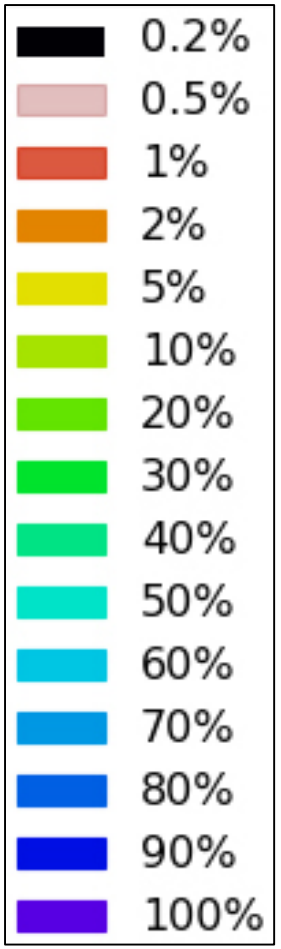


This is the first "hands-on" global demo of the Data Cube to show its potential for rapid time series analysis and diverse applications

http://tinyurl.com/datacubeui

Lake Chad, Cameroon, Africa
Time Series Water Detection



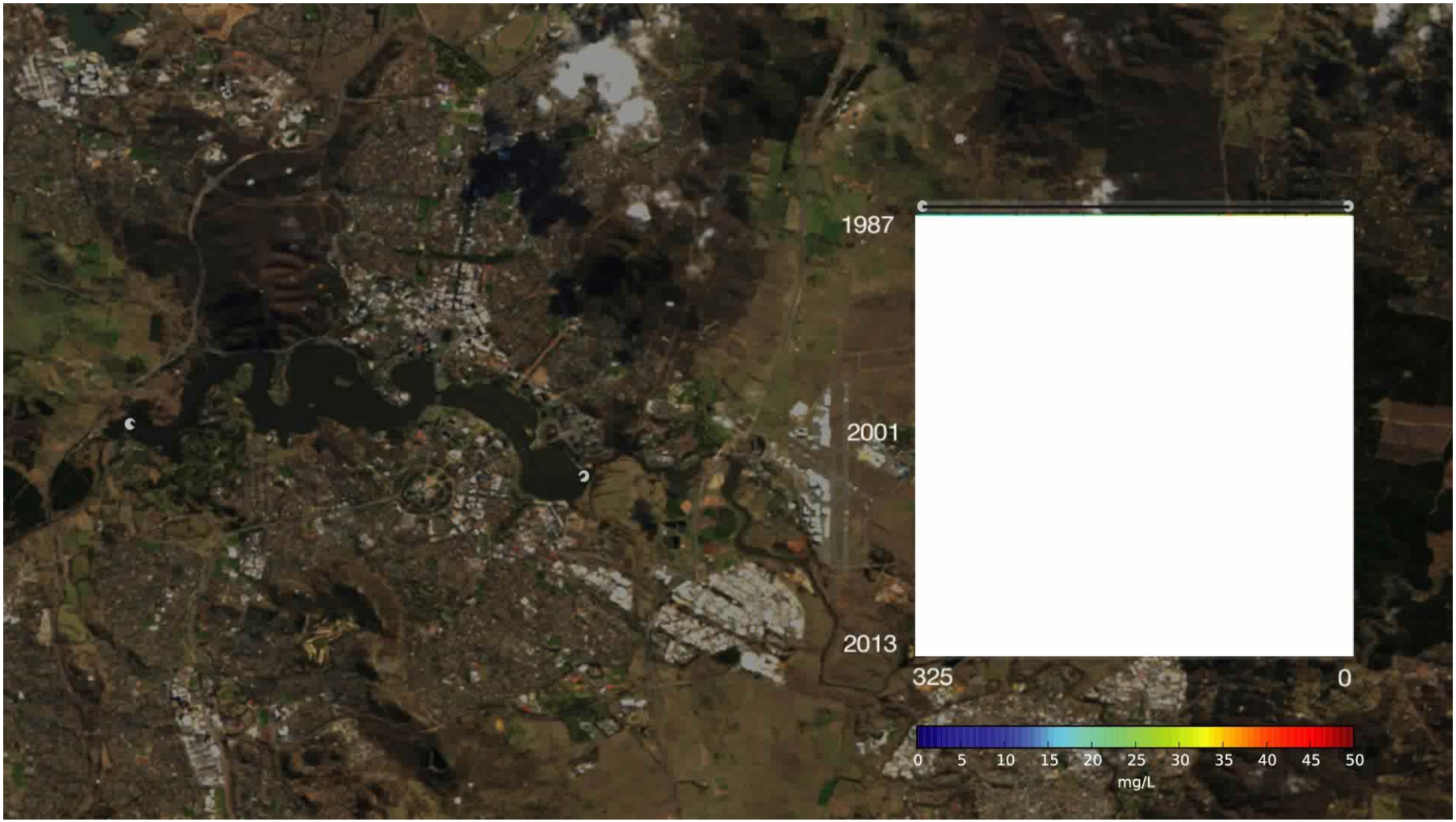


The product shows the percent of observations detected as water over the **17-year time series** (water observations / clear observations).

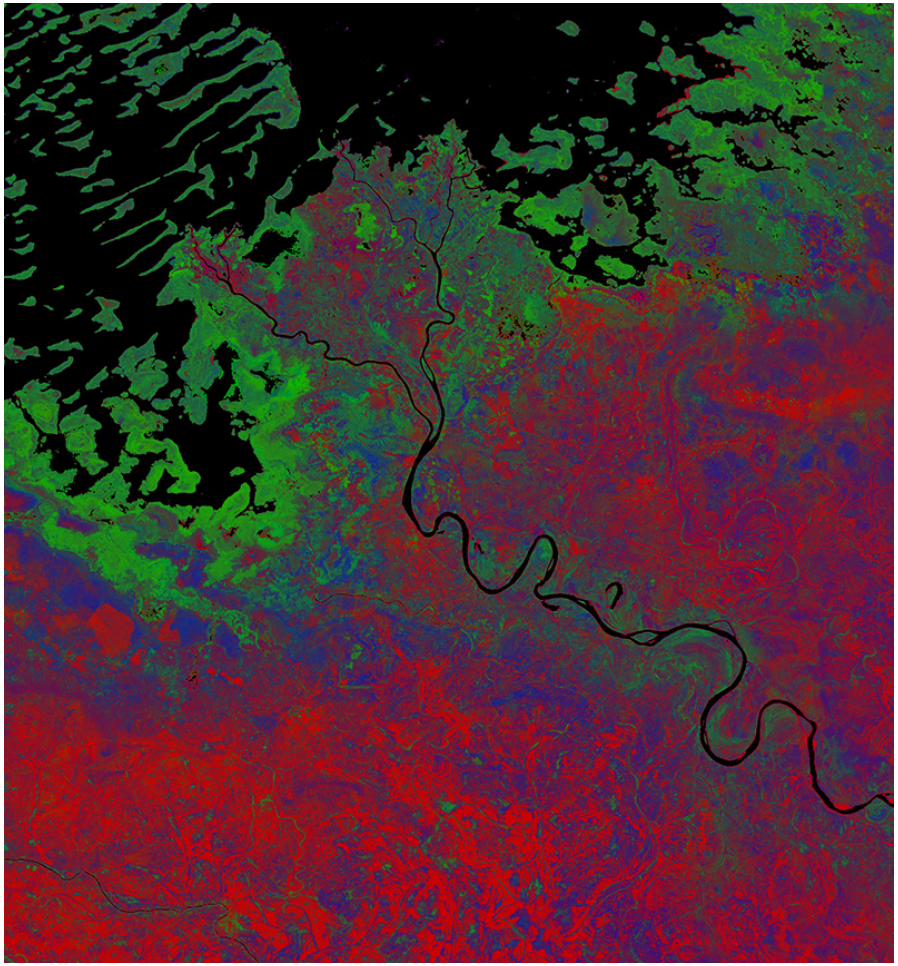
Purple/Blue:
Frequent or permanent water

Red/Yellow:
Infrequent water and/or flood events

Water Quality
Lake Burley Griffen, Australia



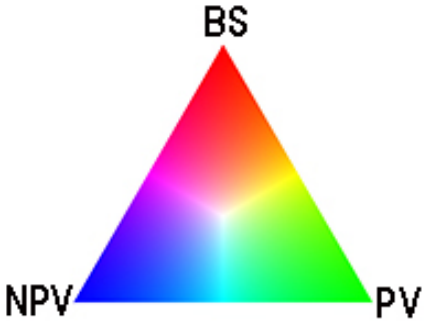
Fractional Cover



Southern Lake Chad
Cameroon, Africa

2015 Fractional Cover

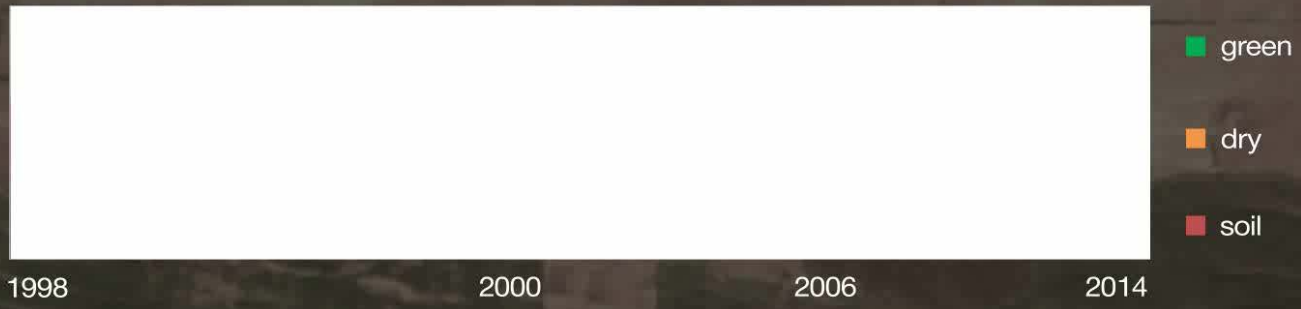
R = Base Soil (BS)
G = Photosynthetic Vegetation (PV)
B = Non-Photosynthetic Vegetation (NPV)
** NPV is dead vegetation, wood, stems, leaves*



The fractional coverage algorithm (right) estimates the average vegetation fractional cover over the time period using a linear unmixing technique developed by Juan P. Guerschman (CSIRO).

Agriculture Change

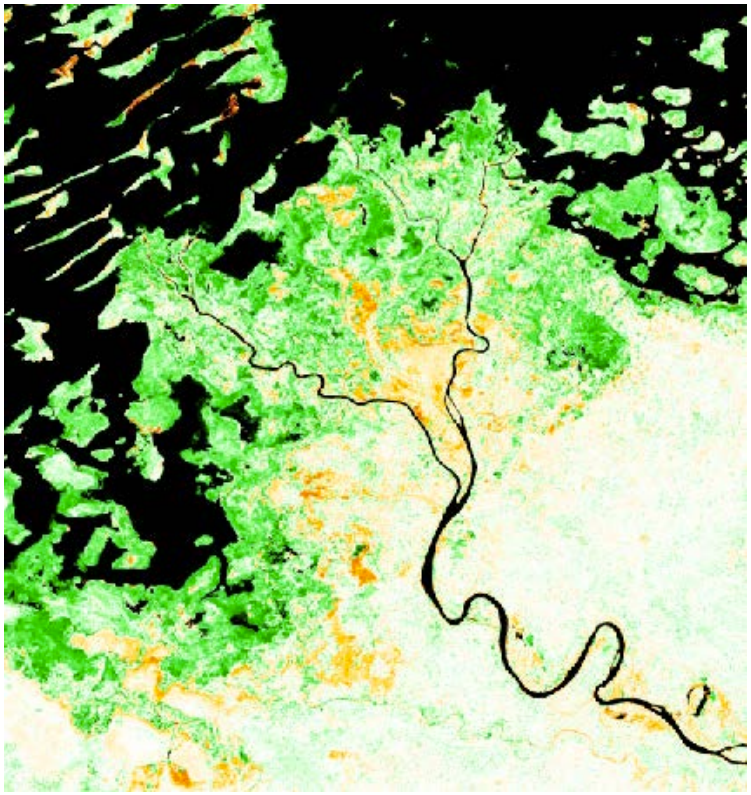
Fractional Cover … detecting change in agriculture



NDVI Anomaly

Chari River inlet to Lake Chad in Cameroon, Africa

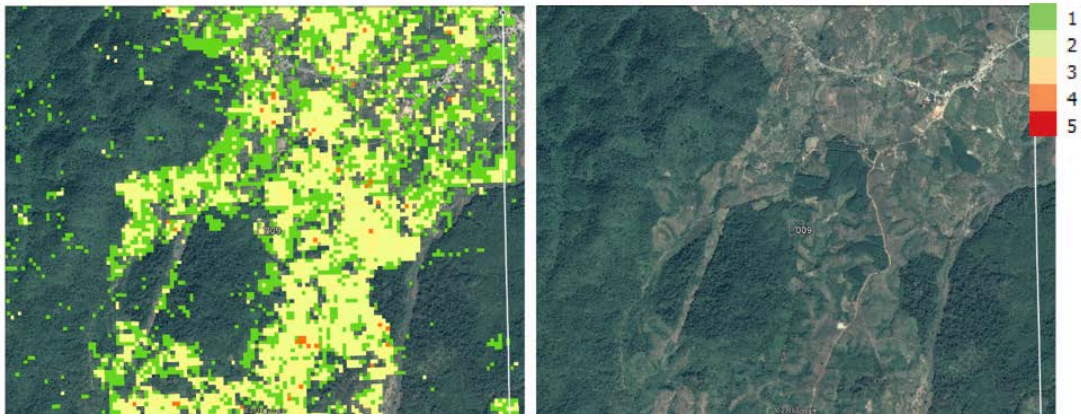
NDVI Anomaly comparison of a single Landsat 8 scene on April 4, 2016 to a 4-year median NDVI for the same month (April, 2013 to 2016)

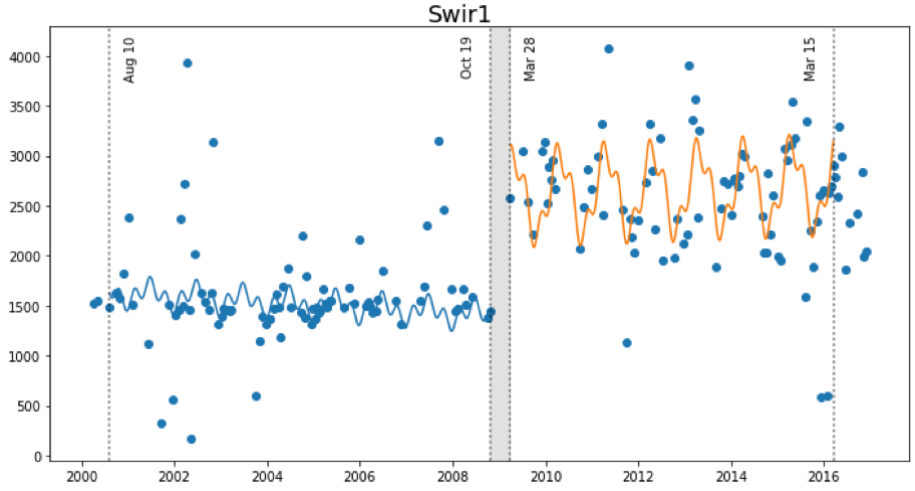


- Consistent with the GEOGLAM Crop Monitor product, but MUCH higher resolution (they use MODIS).
- BLACK regions are masks for either clouds or water
- Most vegetated areas near the Chari River entrance to Lake Chad show an increased NDVI (green) as compared to the historical median.Some reduced NDVI (brown) is seen in a few areas.

Land Change Detection

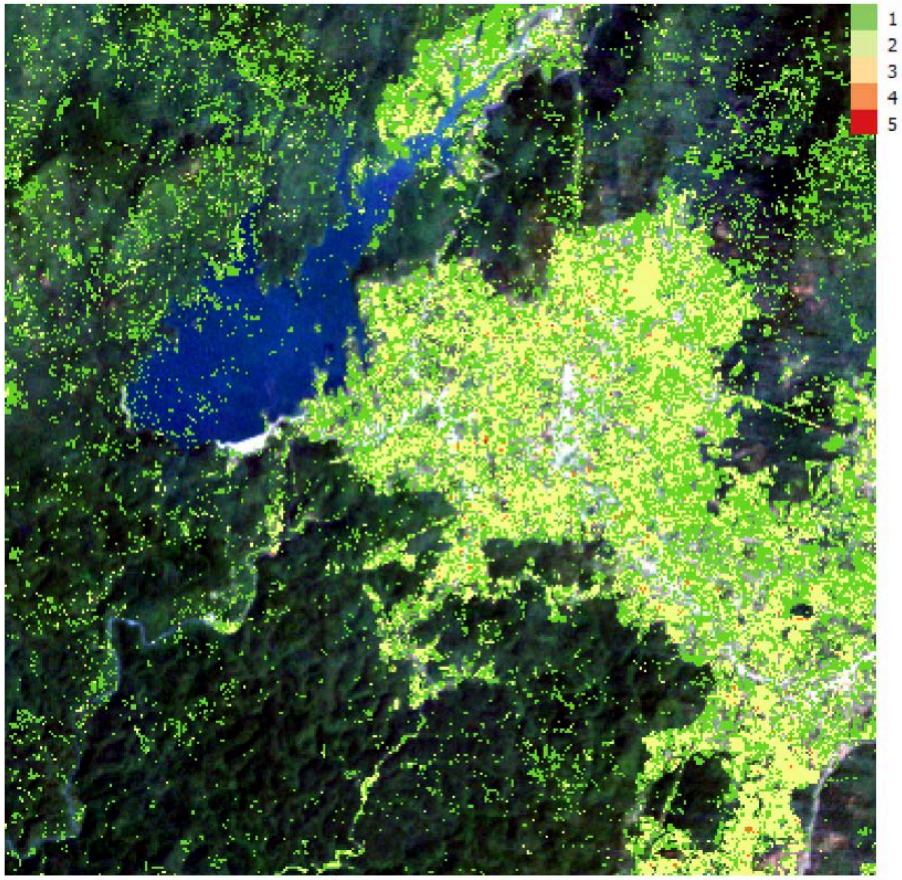
- The SEO is investigating two approaches for change detection with Data Cubes … **CCDC and BFAST**
- **CCDC** (Zhu and Woodcock, 2012) was converted to Python by USGS and recently tested by the SEO on the Vietnam Data Cube. We now call this“**PyCCD**”.
- The SEO is also starting a task with late 2017 to convert **BFAST** to Python and to test the results on the Data Cube.





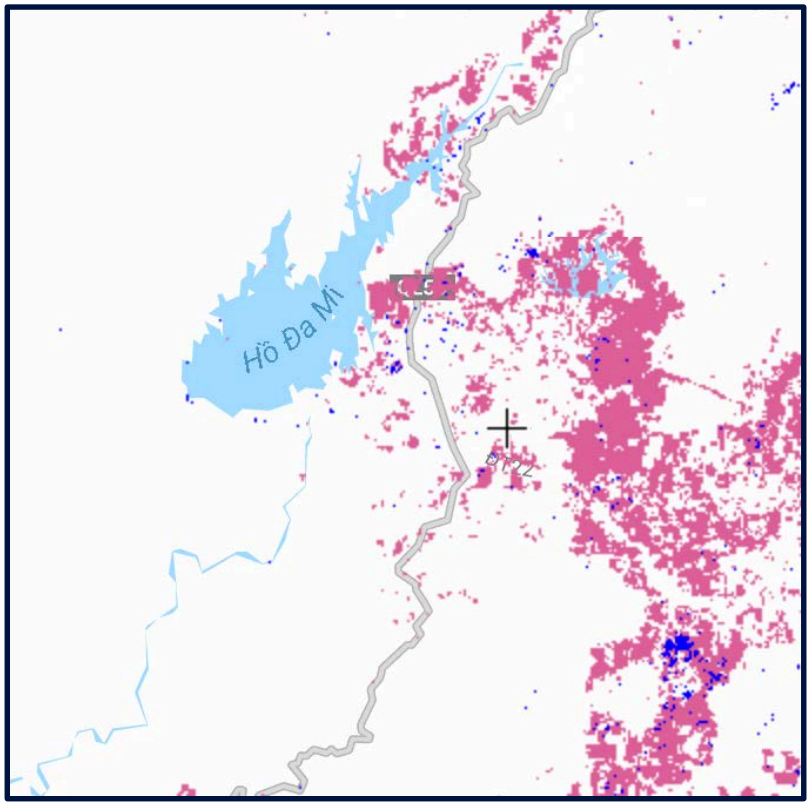
PyCCD time series model fits 7 bands to 6 weighted SIN and COS functions in order to find “breaks” that equate to potential land change.

Vietnam Land Change



Bediaye, Vietnam – Data Cube Mosaic (left), PyCCD Land Change Results (right)
2000 to 2016, 192 Landsat scenes

Global Forest Watch vs. PyCCD

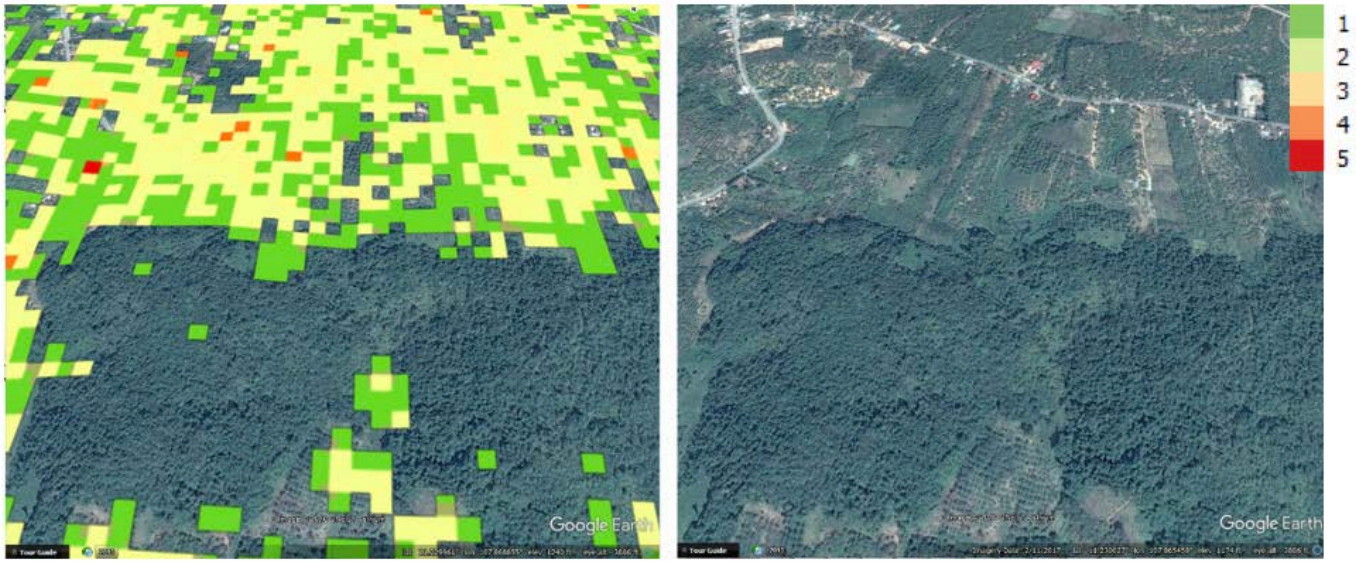


Global Forest Watch
Forest Loss 2000 to 2015

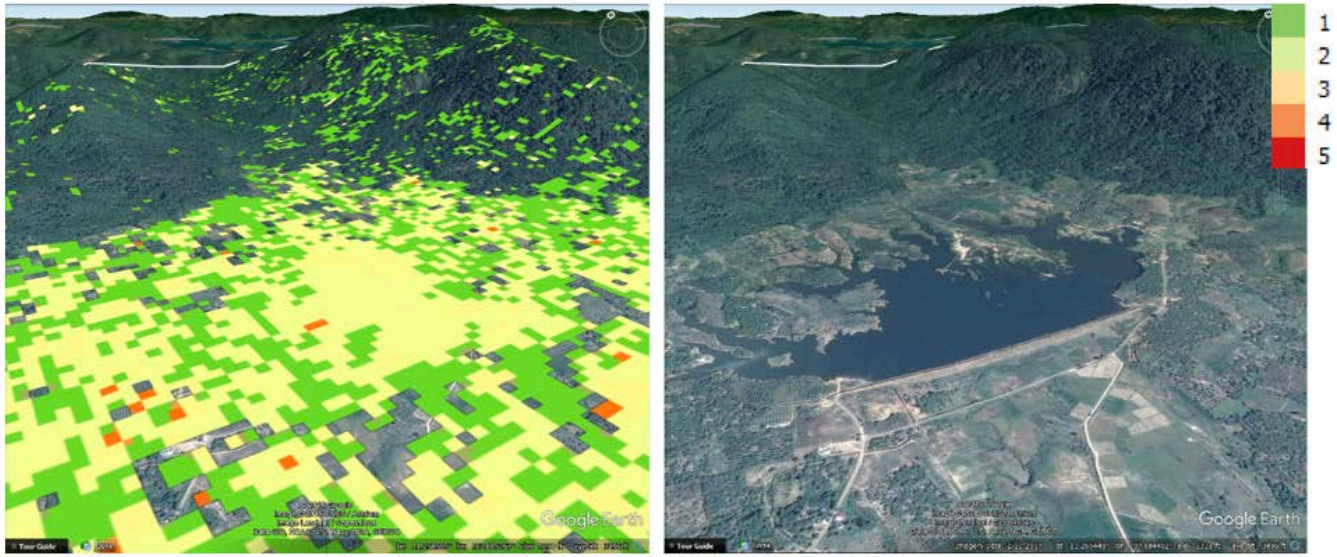
PyCCD with a Data Cube
Land Change 2000 to 2015

***PyCCD Execution**: 372 x 372 pixels, 8 parallel cores, 2.3 hours (~1 msec / clear pixel) which equates to about 10 hours per clear Landsat scene.*

More PyCCD Examples



Possible Deforestation



Reservoir Construction

Thank You

http://www.ceos.org

http://www.opendatacube.org

http://tinyurl.com/datacubeui