Open Data Cubes

A Big Data Solution for Global Capacity Building

"How to Cloud for Earth Scientists" Series
NASA GSFC
December 19, 2017

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

What is CEOS?

www.ceos.org

Group photo from the 2017 CEOS Plenary in Rapid City, South Dakota, USA

The Committee on Earth Observation Satellites (CEOS) serves as a focal point for international coordination and data exchange **to optimize societal benefit from space-based Earth observations**. CEOS represents 22 countries through its 32 space agencies and 28 associate members and is operating 151 satellites.

What is changing?

Free and Open Resources

- Free and open Satellite Data
- Open Source Software and Tools

Global Engagement

- Improved communication technology
- Increased global cooperation and collaboration

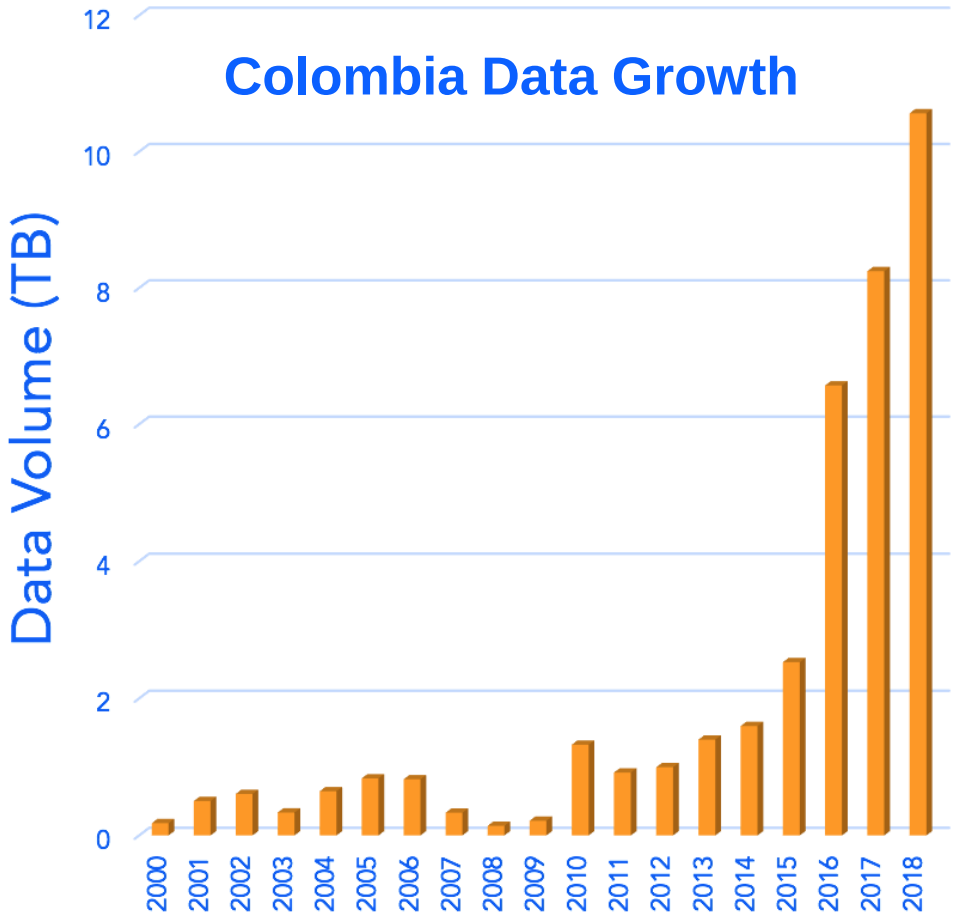
Satellite Data Philanthropy

- Google Earth Engine
- Governments, Foundations and World Bank

Improved Technology

- Cloud Storage and Computing
- Data Cubes

The Big Data Problem

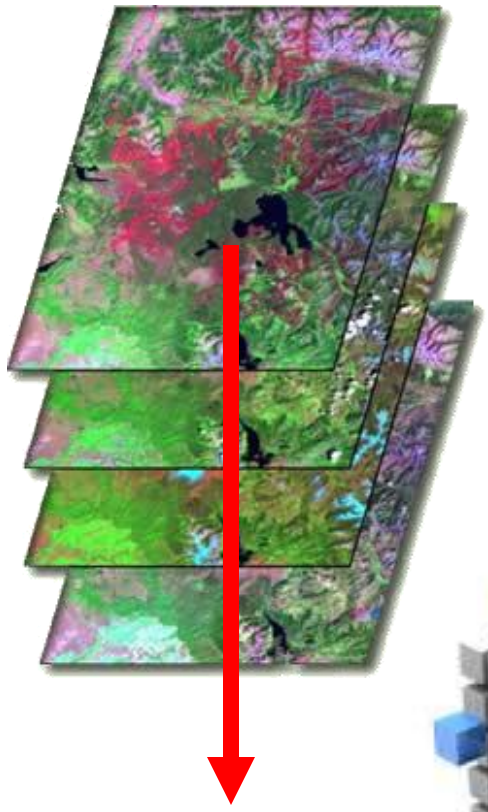


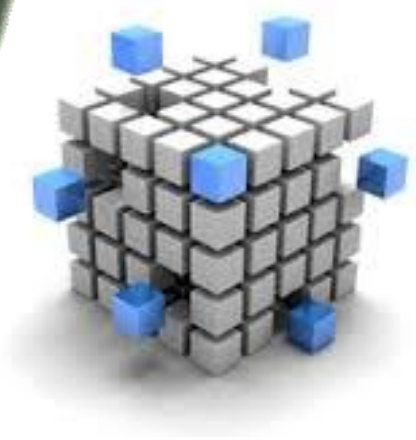
- A significant growth in FREE/OPEN land imagery data (e.g. Landsat, Sentinel) has **increased data volumes** by 10x in the last 5 years.
- Many countries **lack the knowledge, infrastructure, and resources** to access and use the available space-based data.
- Countries have **requested support** from CEOS for data access, processing, and analysis 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** in Australian and now working in Switzerland, U.K. and Colombia.
- **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:** common architecture, flexible deployment, exploits time series, increases data interoperability, and supports many new applications.





Benefits of Data Cubes

- Expanded use of satellite data
- Reduced data preparation burden
- Enables data interoperability
- Efficient time series analyses
- Free and open access
- Flexible deployment (local or cloud)
- Use of a common architecture
- Community development and sharing

*Our goal is to provide a **SOLUTION** that has **VALUE** and increases the **IMPACT** of satellite data.*

The Data Cube Vision

A solution supporting priority objectives …

- Build capability of users to apply CEOS satellite data
- Support Group on Earth Observations (GEO) and United Nations agendas

Involves CEOS Agencies …

- Through provision of processed satellite data products
- Contributing to development and uptake of solutions

Customer focused …

- Easy to install and maintain with training materials
- A brand that people know and trust

Scalable solution …

- Operational Data Cubes in **20 countries by 2022**
- Key partners (e.g. World Bank, Google, Amazon) supporting the data cube development and its use

opendatacube.org

The “Road to 20”

3 operational, 7 under development, 29 expressing interest = **39 total countries**

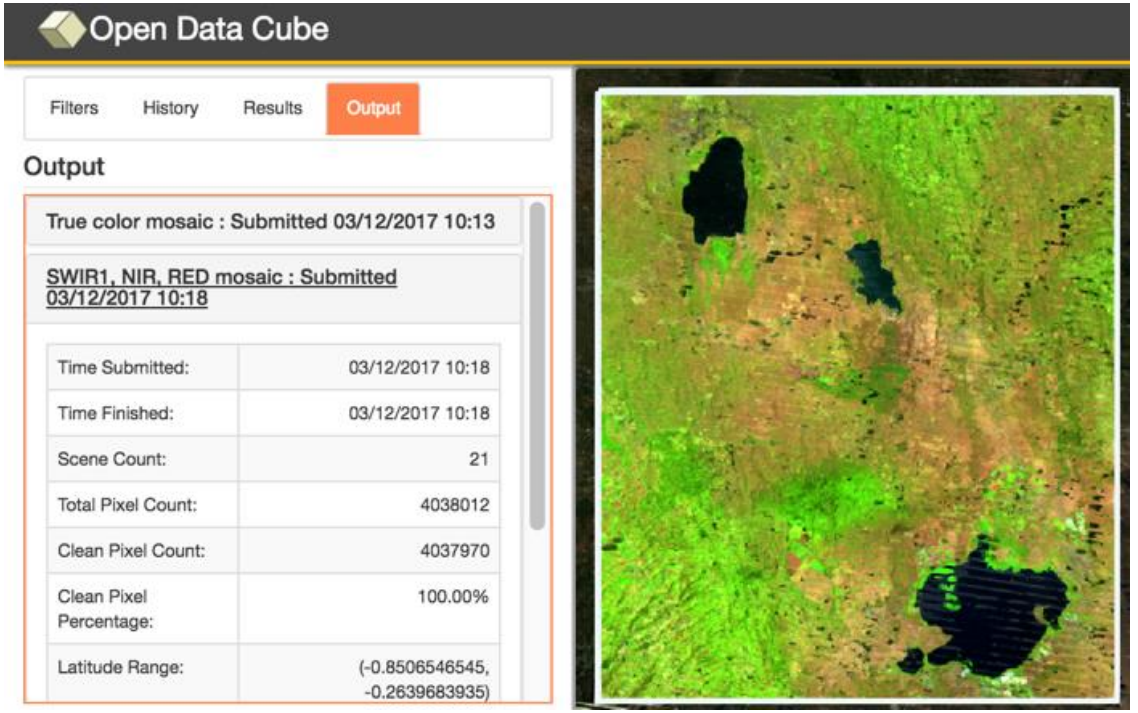
Amazon Cloud (AWS) Demo

Data Cubes

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

User Interface Features

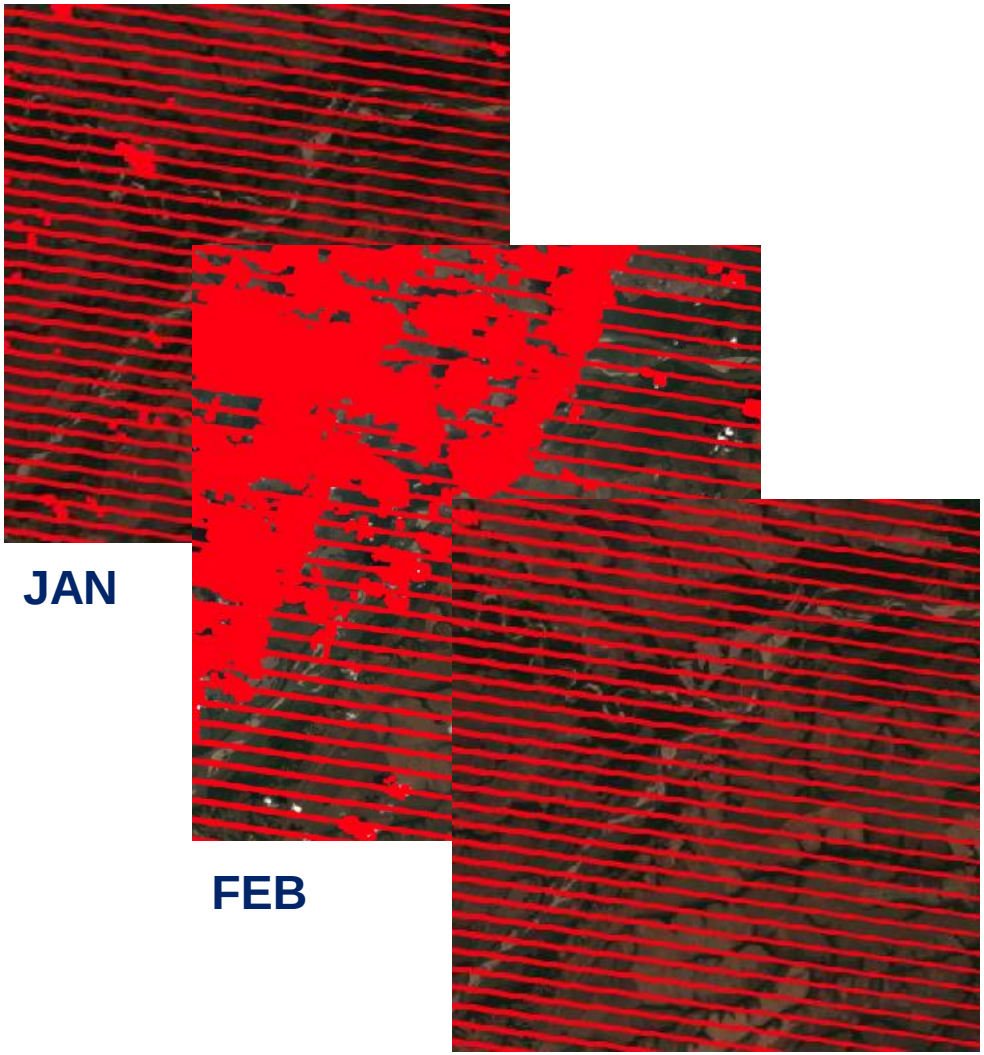
- 10 applications: cloud-free mosaics, fractional cover, NDVI anomaly, water detection, water quality, landslides, coastal change and urbanization.
- Outputs in GeoTIFF and GIF animation.



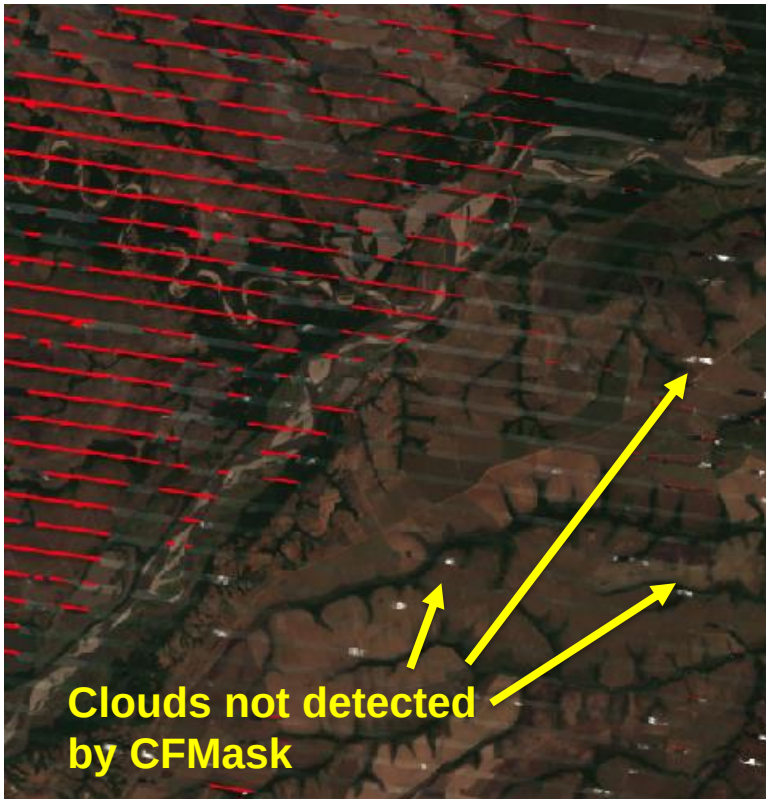
http://tinyurl.com/datacubeui

Free and Open!

Cloud-filtered Mosaics

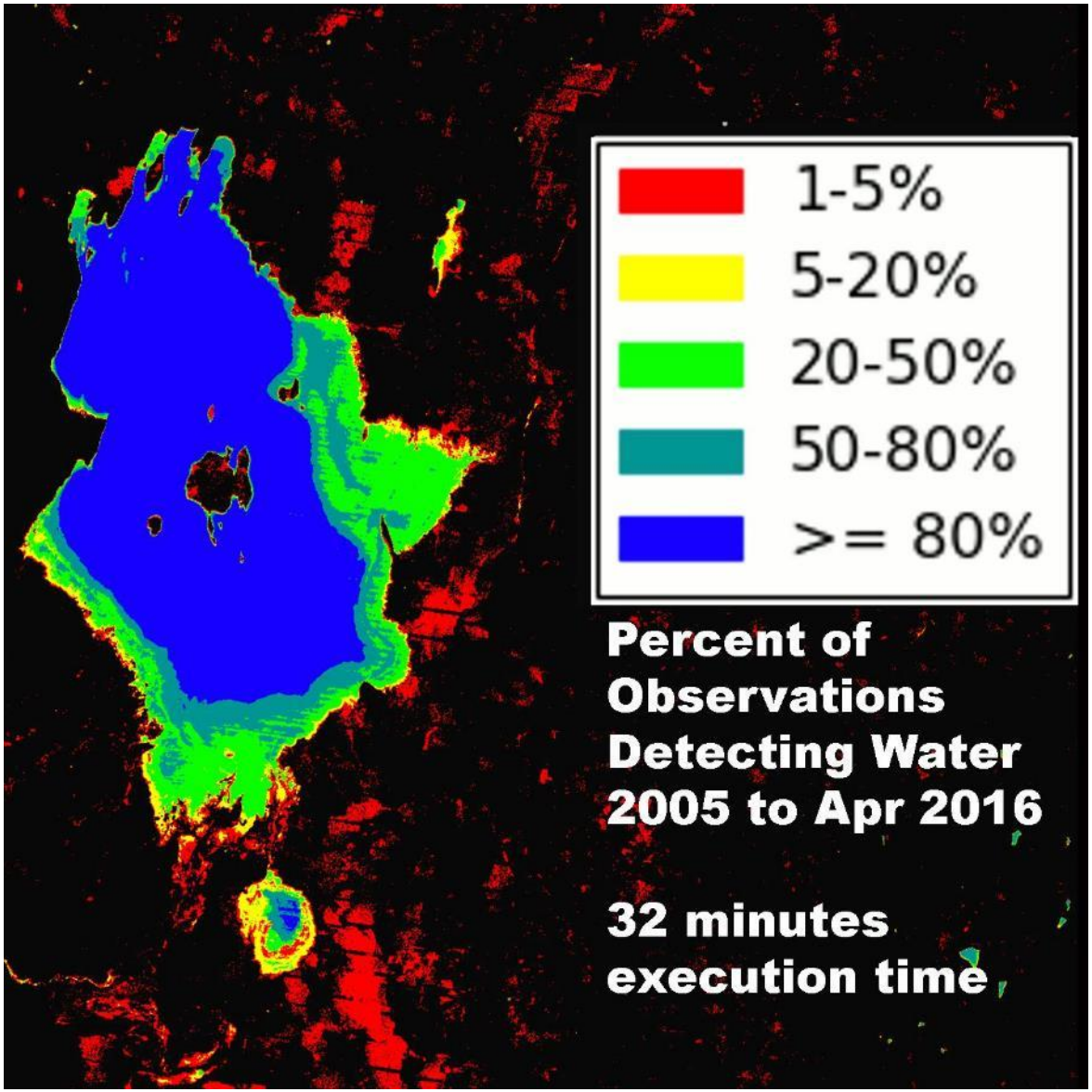


Meta River, Colombia



The final product (above) is a cloud-filtered “most recent pixel” mosaic for Jan-Mar 2015 (3 months). The result is compiled from three (3) Landsat-7 scenes to produce a 97% cloud-free image. The baseline scenes (left) are 30% to 50% cloudy. The cloud or no-data pixels are highlighted in **RED**. This analysis is produced very rapidly (~1 minute).

Lake Baringo, Kenya
Time Series Water Detection



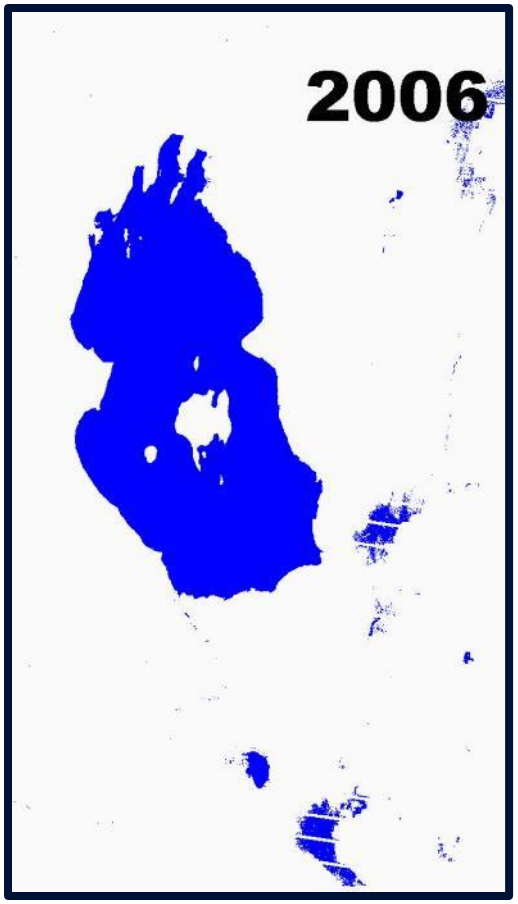
Blue = frequent or permanent water

Red/Yellow = infrequent water or flood events

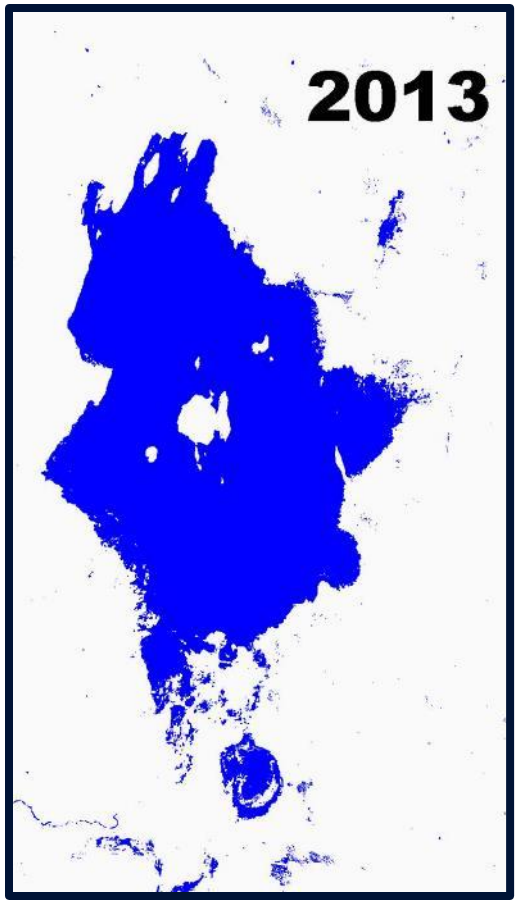
Flood risk can be easily inferred from the 23-step, multi-band Australian WOFS algorithm.

30-meter Landsat resolution allows detailed assessments that are far better than MODIS (250-m).

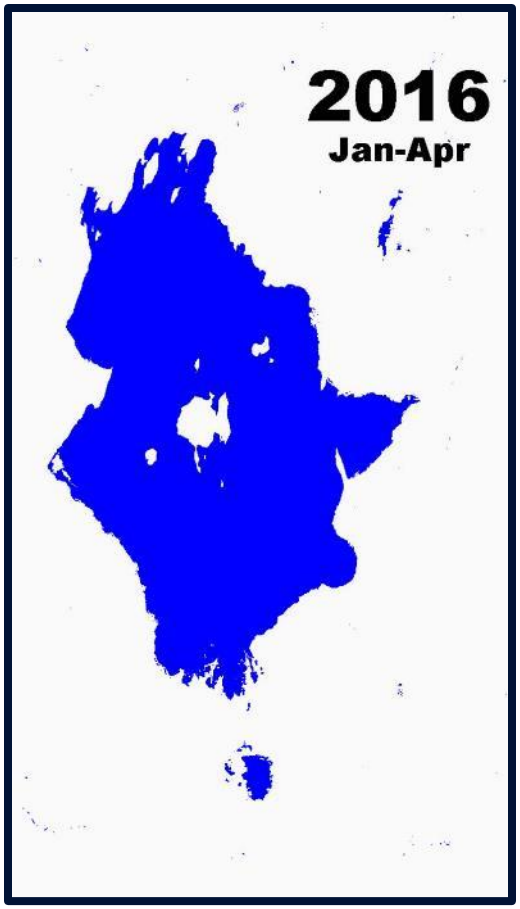
Annual Water Extent



Extreme droughts in the Baringo region in 2006 had severe impacts on pastures and farming



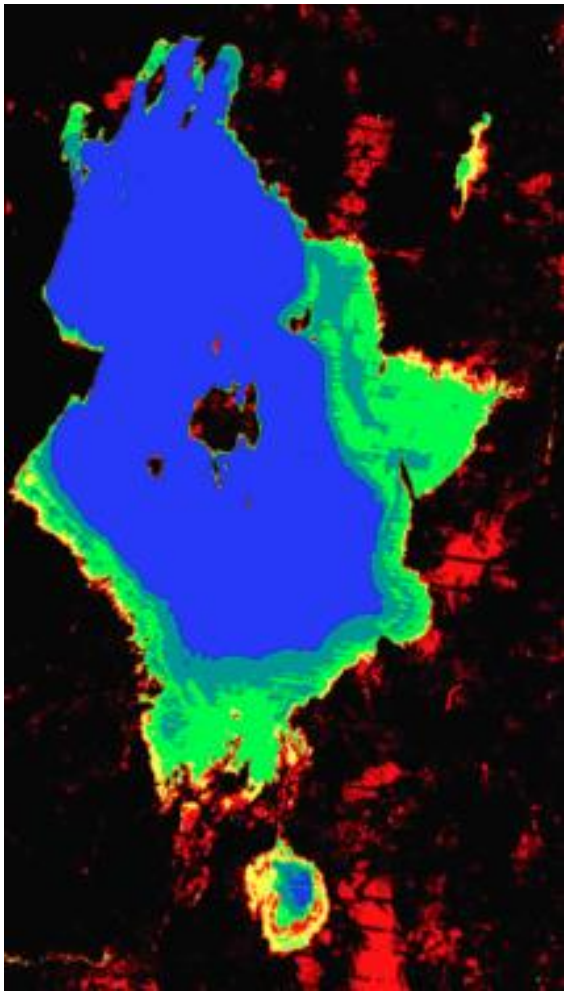
Extreme floods displaced 600 families and swept away livestock near Lake Baringo in 2013



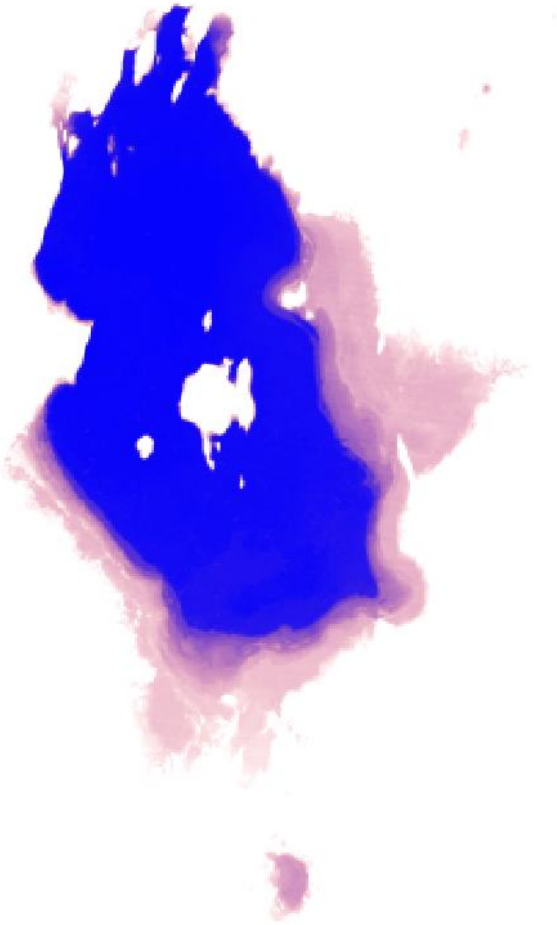
4 months of dry season data resulted in little water detected outside the lake boundary

WOFS vs. Other Water Tools

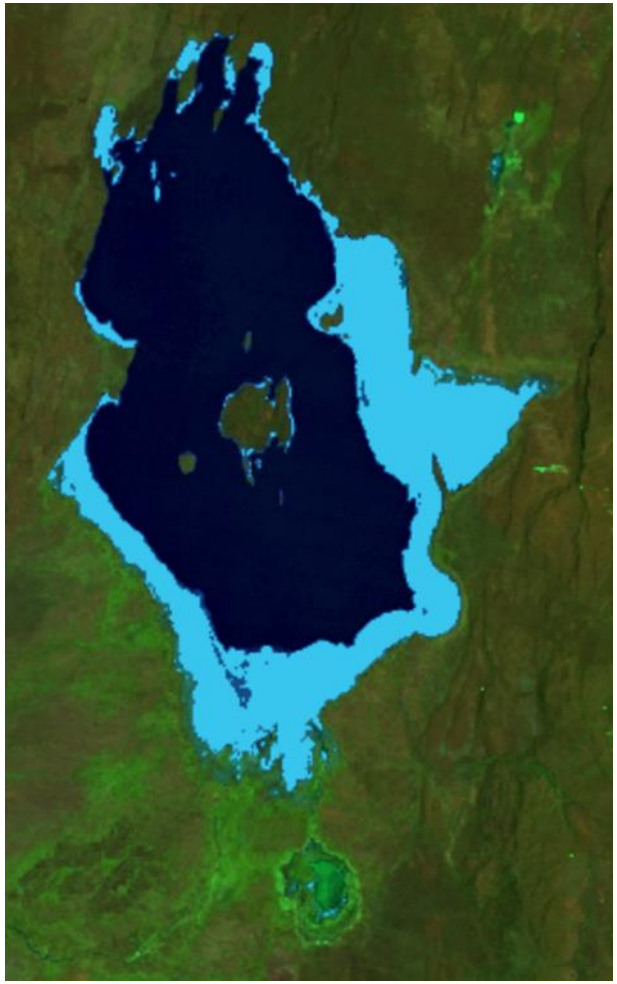
Australian WOFS
2005 to 2016



EC-JRC
1984 to 2015

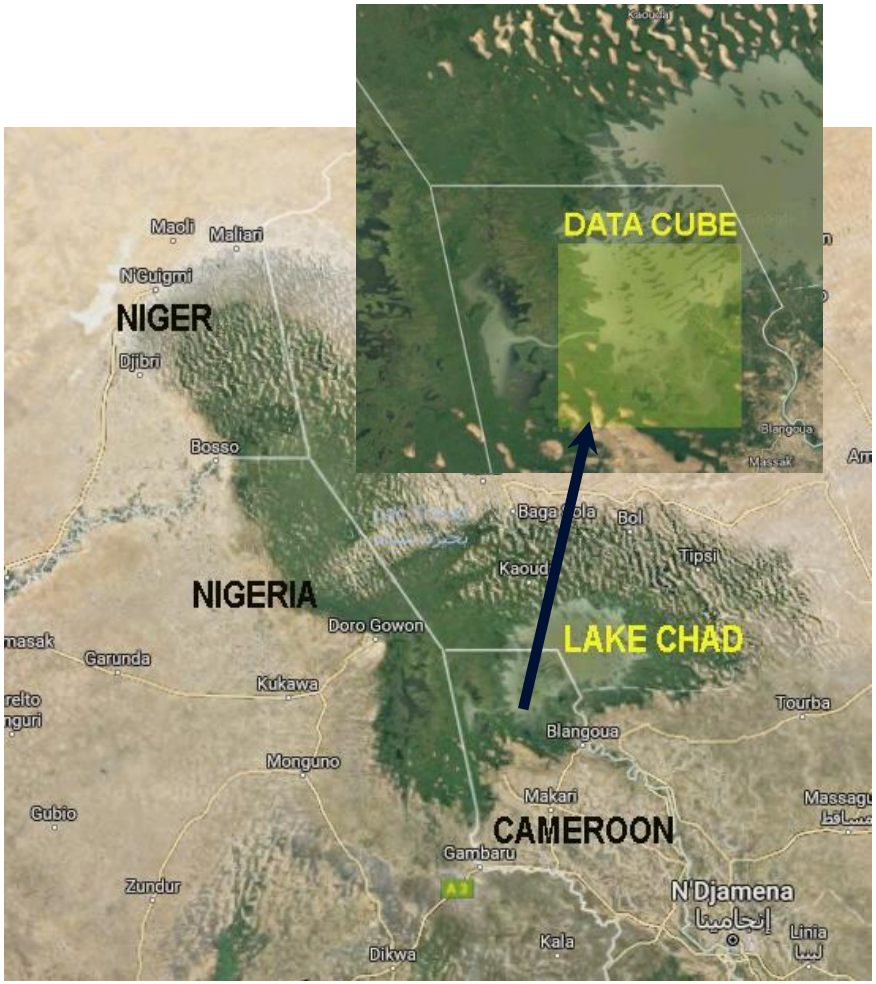


Aqua Monitor
2005 to 2016

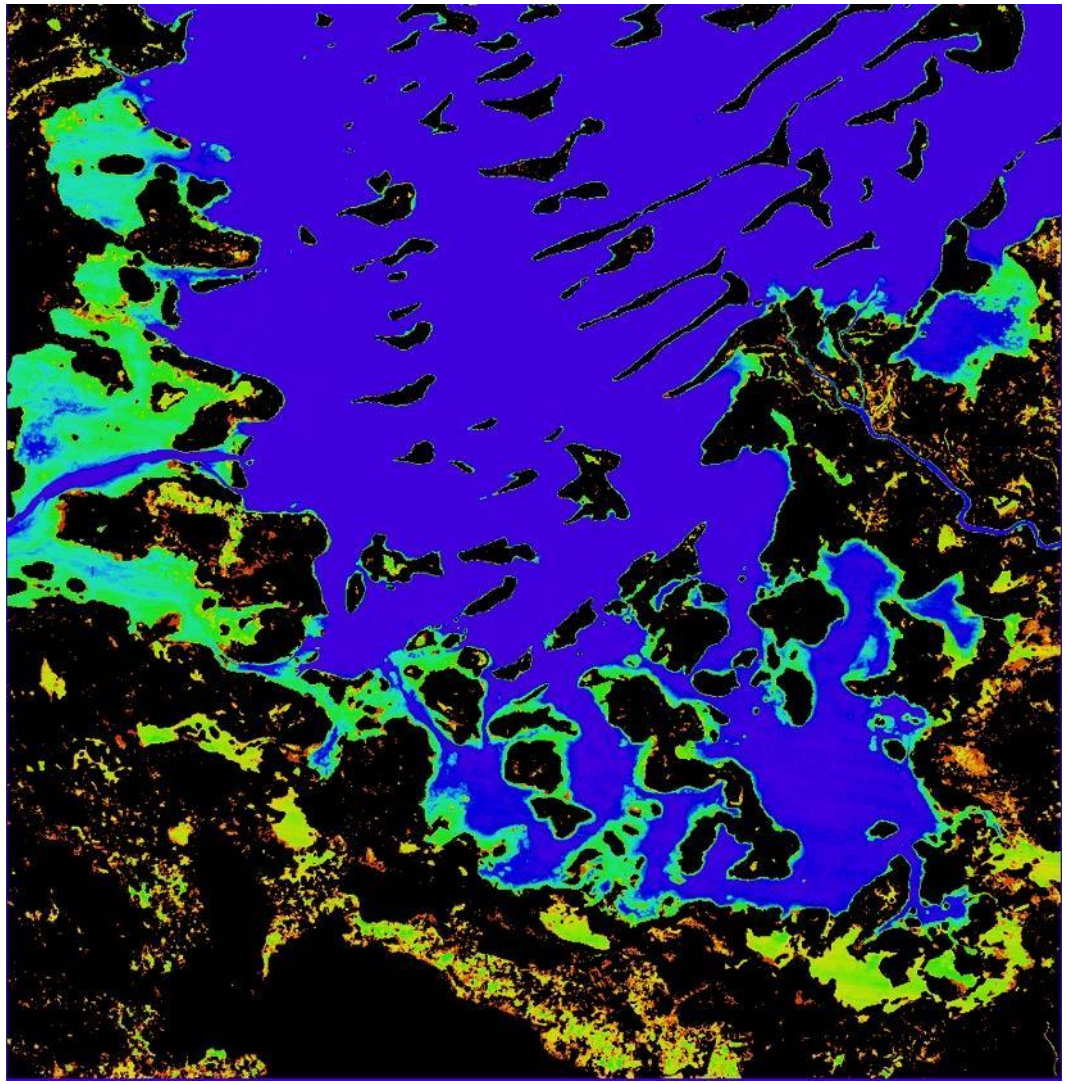


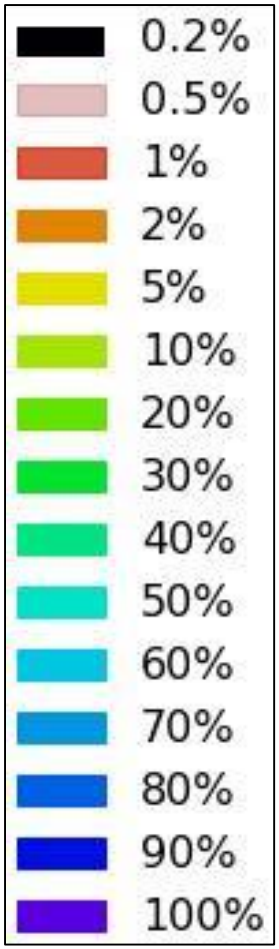
Lake Chad, Africa
Water Detection Demonstration

Historically large and shallow lake has shrunk by 95% from 1963 to 1998 due to increased population demand (reference United Nations). Provides water to 68 million people in 4 bordering countries.



Lake Chad, Cameroon, Africa
Time Series Water Detection



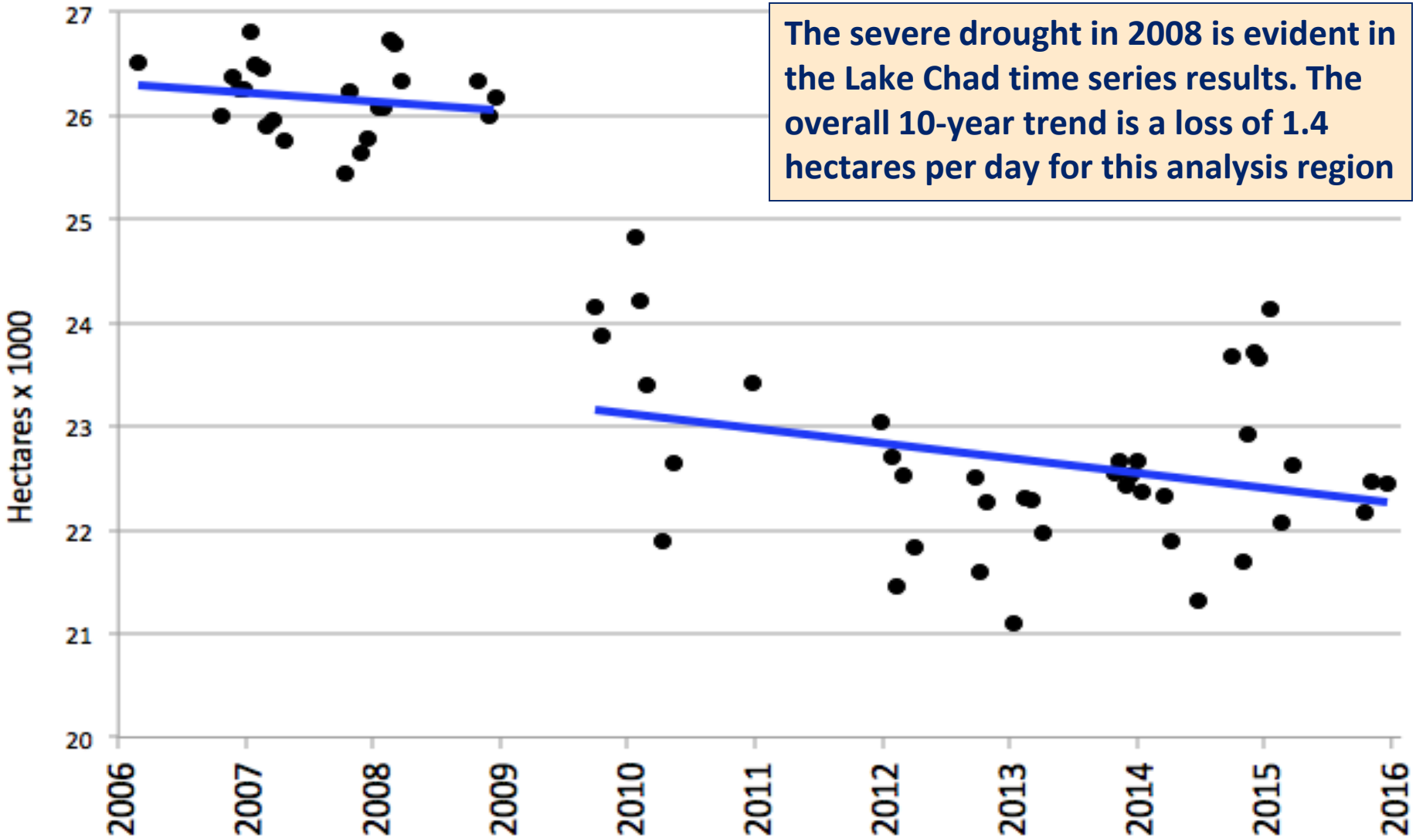


The product shows the percent of observations detected as water over the **17-year time series** (water observations / clear observations).

Purple/Blue:
Frequent or permanent water

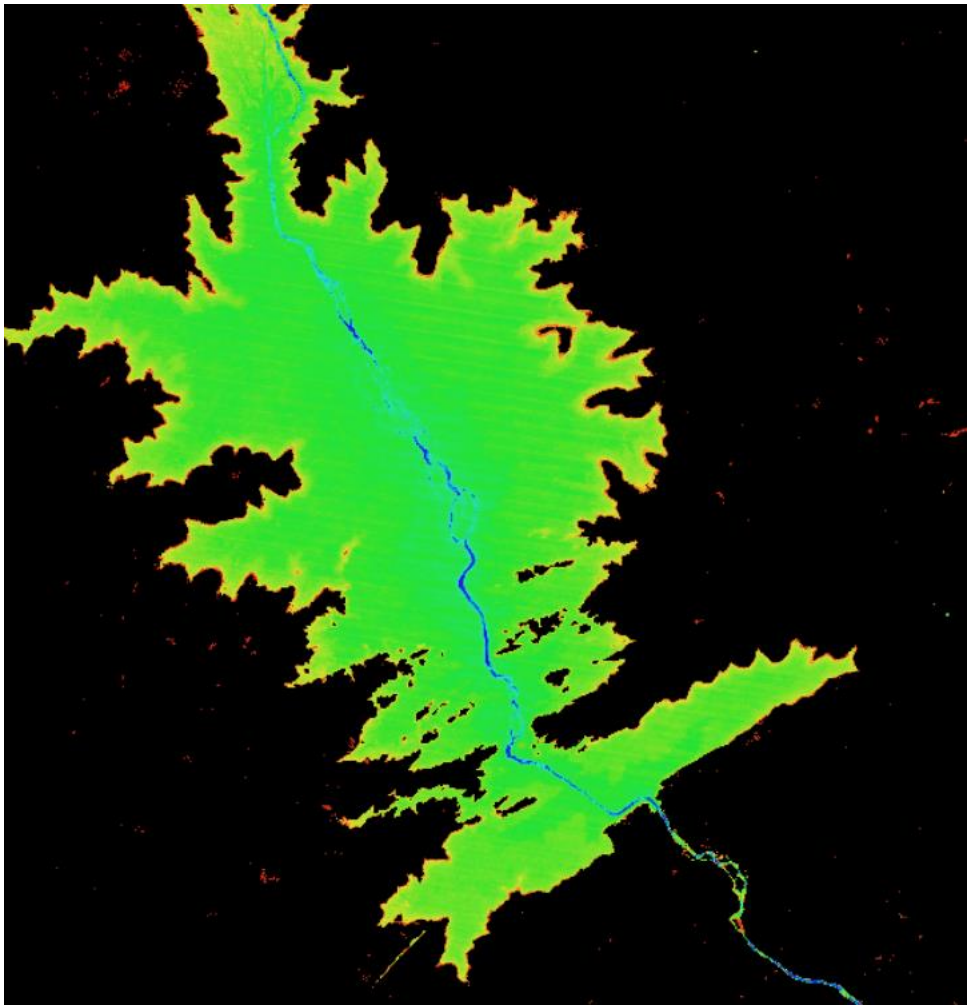
Red/Yellow:
Infrequent water and/or flood events

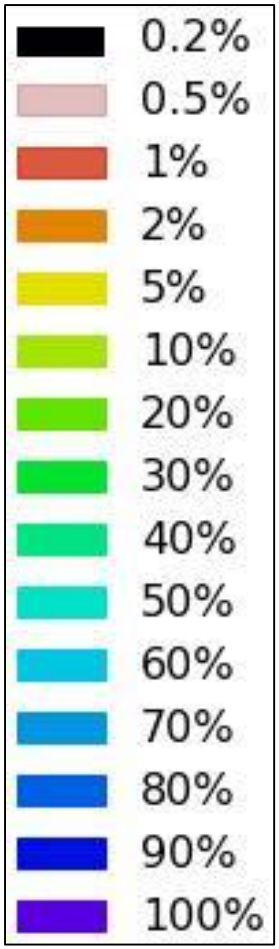
Lake Chad, Cameroon, Africa
10-year Time Series Results



The severe drought in 2008 is evident in the Lake Chad time series results. The overall 10-year trend is a loss of 1.4 hectares per day for this analysis region

Time Series Water Detection





The Australian Water Observations from Space (WOFS) 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

Bui National Park along the Black Volta River, western Ghana, Africa

Why does the water only exist for 20% of the 17 years?

Bui National Park Land Change

Dec 2010
Dam under construction

Dec 2016
Dam complete … New Lake!

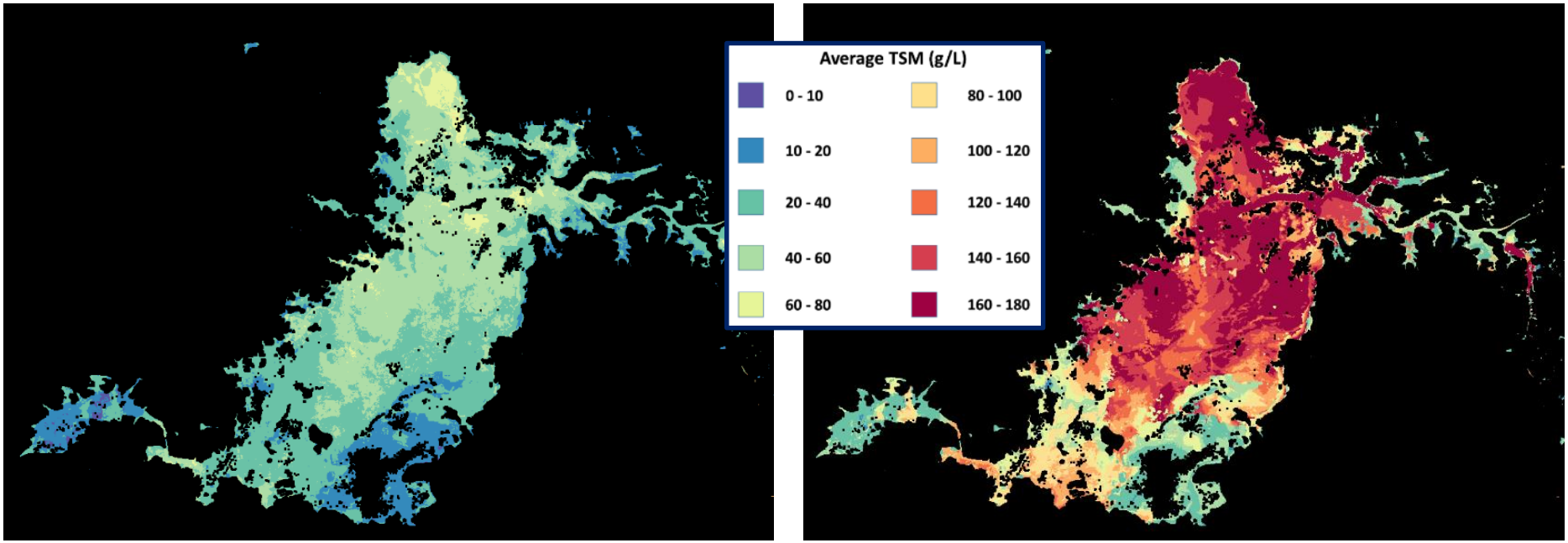
Bui National Park started construction of a Dam in Dec 2009. This explains the short existence of water from 2000 through 2016. The images to the left support these results.

Time series observations of water can be used to track the progress of water management projects, such as this project in Ghana.

Water Quality
Total Suspended Matter (TSM)

Average TSM

Maximum TSM



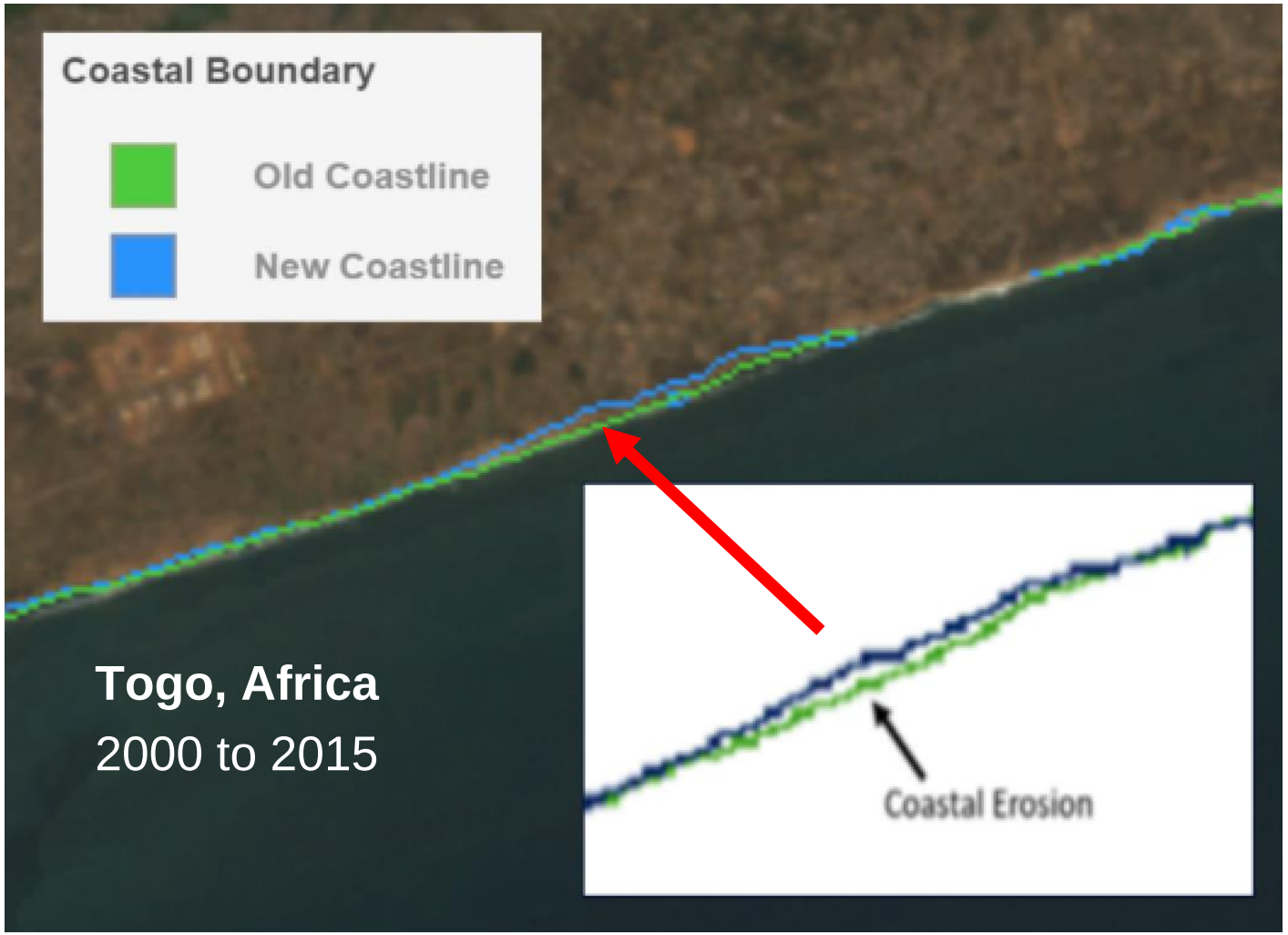
The Tri An Reservior in southern Vietnam (near Ho Chi Minh City) supplies drinking water to millions of people.

The results show the average and maximum TSM (mg/L) levels over the 2016 annual time series for persistant water. The product is calculated using Landsat 8 data and the Lymburner TSM Index Algorithm.

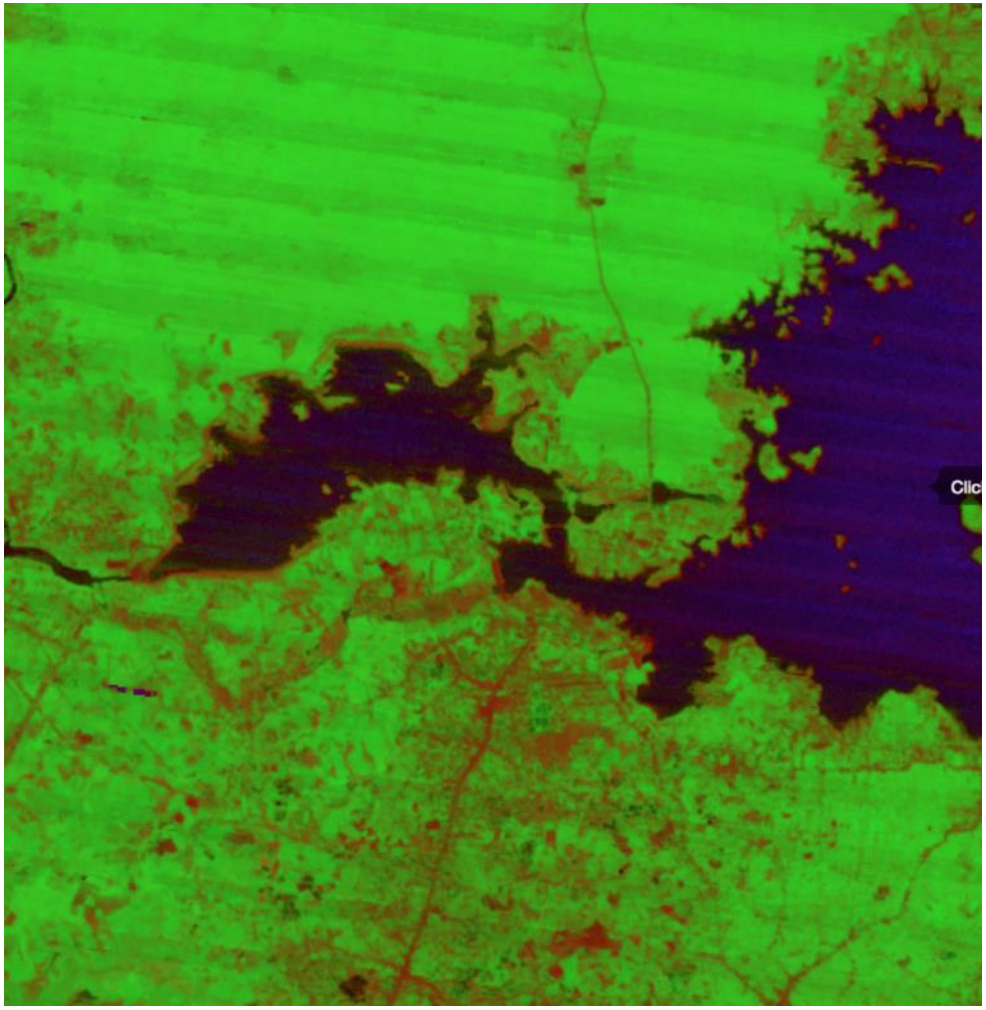
TSM is closely related to turbidity which is an indicator of water condition for drinking or fisheries

Coastal Change Detection

Lome, Togo, Africa

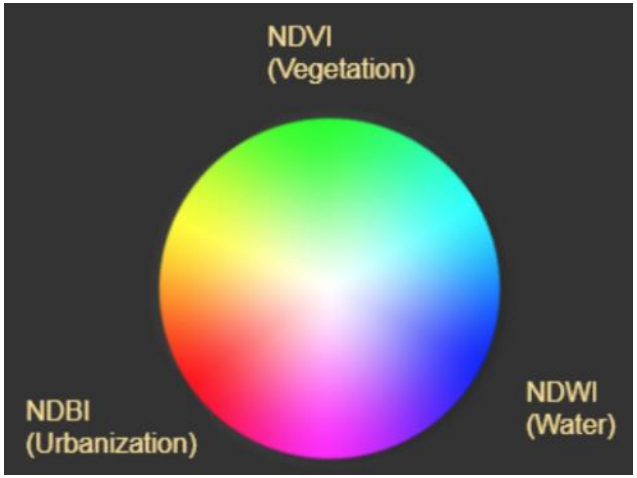


Urbanization



Ho Tri An Lake and Vinh An
near Ho Chi Minh City, **Vietnam**
Median: Year 2016

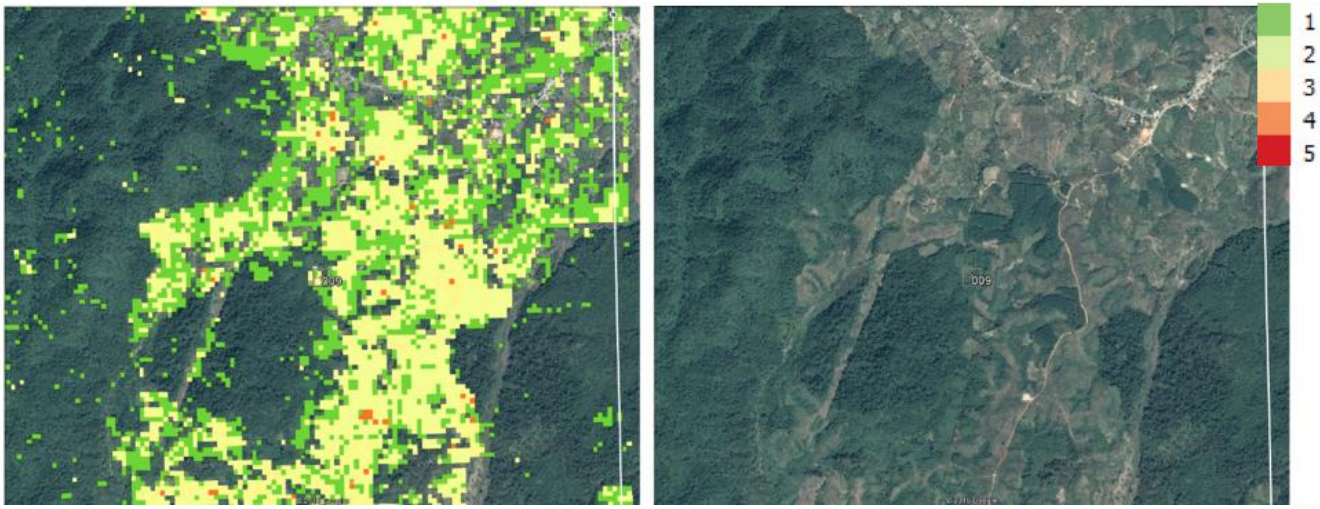
Red = Urbanization
Green = Vegetation
Blue = Water



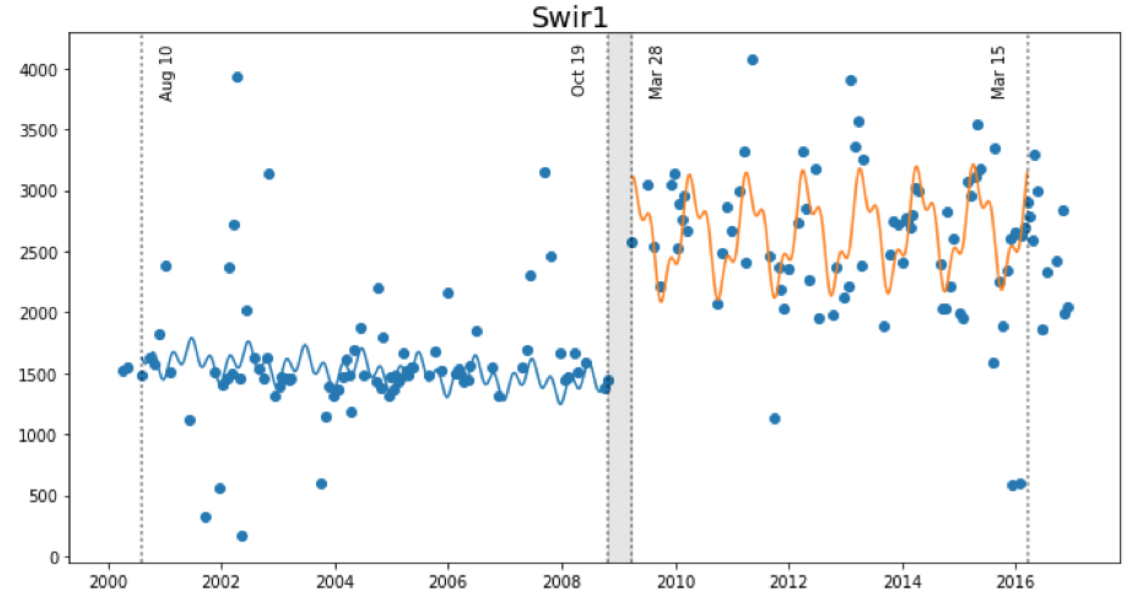
The urbanization product combines 3 separate spectral index products: NDBI (Red), NDVI (Green) and NDWI (Blue).

Land Change Detection

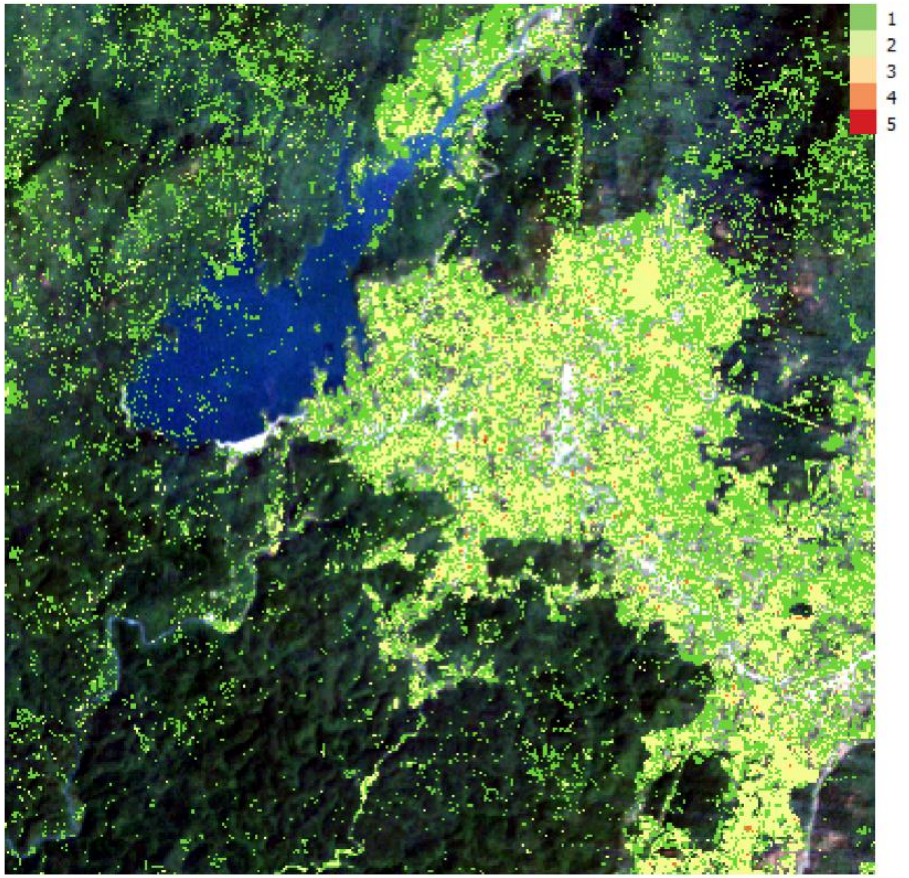
CCDC (Zhu and Woodcock, 2012) was converted to Python and recently tested on the Vietnam Data Cube. We now call this "**PyCCD**"



PyCCD time series model fits 7 bands to 6 weighted SINE and COSINE functions in order to find "breaks" that equate to potential land change.



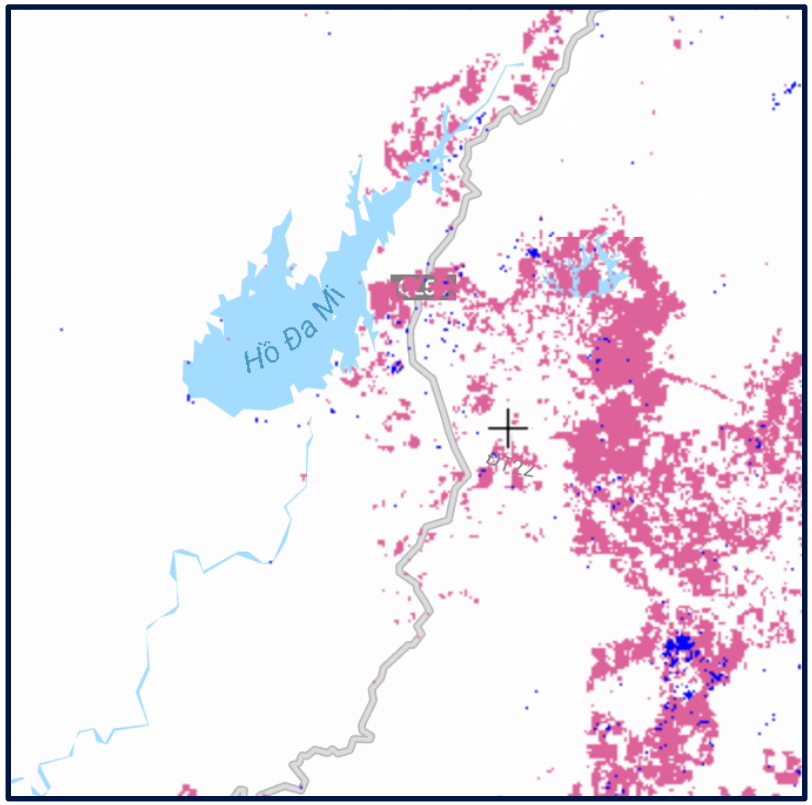
Vietnam Land Change



Bediaye, Vietnam – Data Cube Median Mosaic (left), **PyCCD Results** (right)

2000 to 2016, 192 Landsat scenes

Global Forest Watch vs. PyCCD

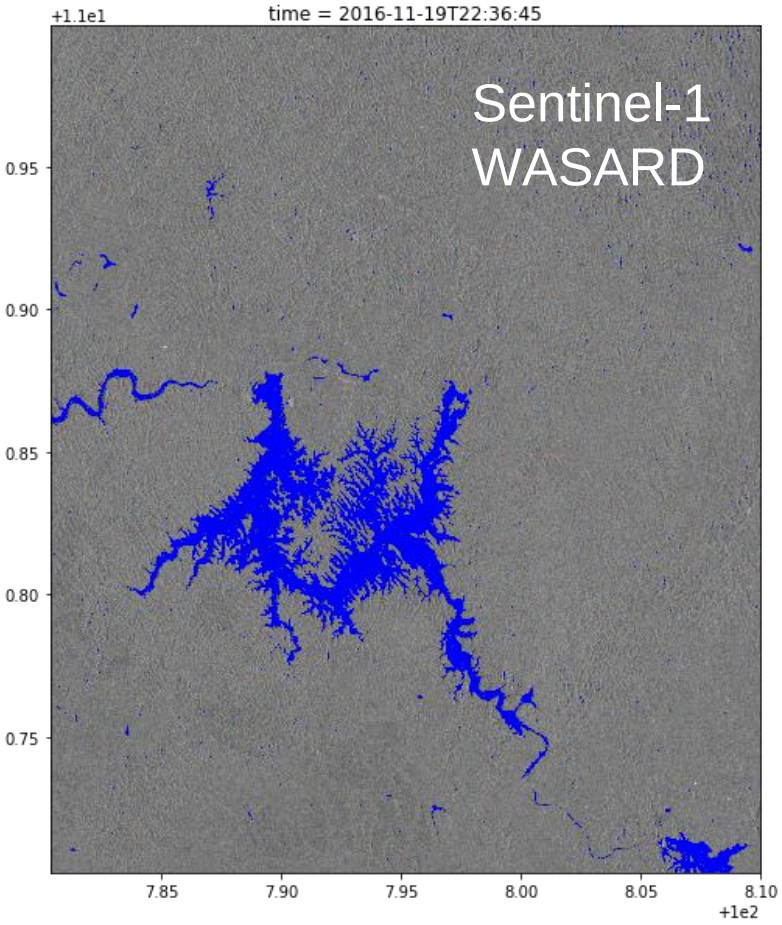


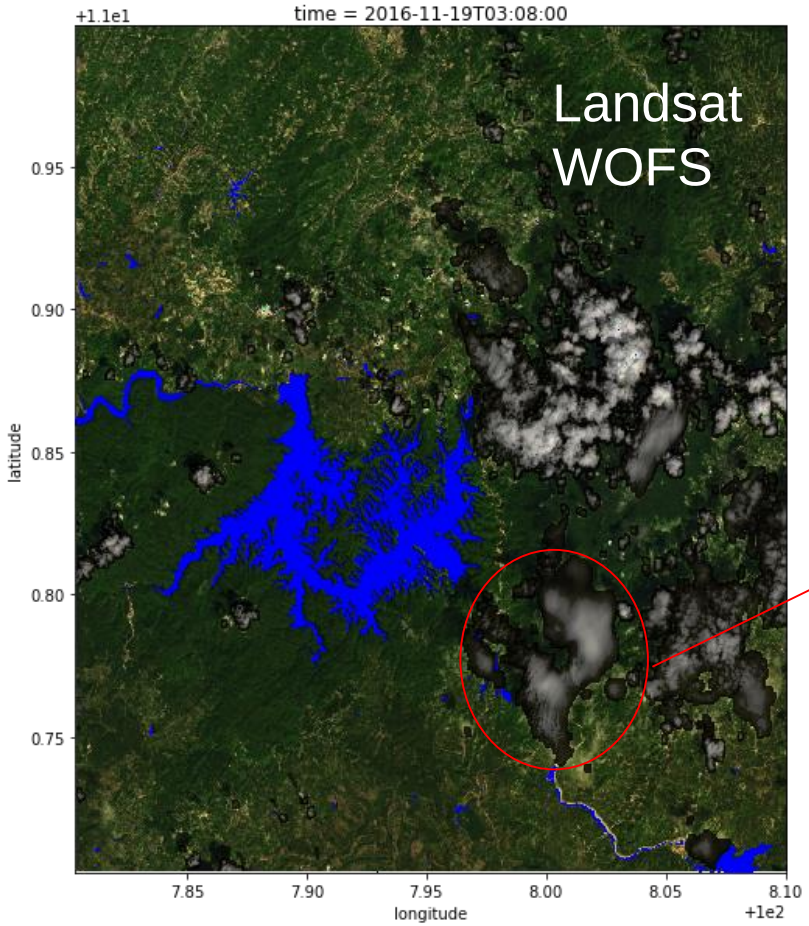
Global Forest Watch – **Forest Loss**
2000 to 2015

PyCCD with a Data Cube – **Land Change**
2000 to 2016

***PyCCD Execution**: 372 x 372 pixels, 8 parallel cores, 2 hours execution*

Sentinel-1 Water Detection





clouds covering a portion of the water

These Sentinel-1 and Landsat acquisitions are on the same day. The presence of clouds reduces water visibility with Landsat, but using Sentinel-1 and WASARD (machine learning algorithm) we can achieve a full view of the water extent. The correlation between these results is 96%.

Amazon and Google

- The goals of CEOS, the ODC initiative, Google and Amazon are similar … to increase the global use and benefit of satellite data
- New collaborations with **Amazon and Google** are in progress …
- NASA has received AWS credits for 2 years to support ODC development and testing
- A recent proposal was presented to Amazon (Joe Thrasher and Jed Sundwall) to request a grant to GPSDD to support the African Regional Data Cube.
- Working with Rebecca Moore and Matt Hancher (Google Earth Engine, GEE) to develop a “Data Cube on Demand” concept to operationalize the connection of GEE data assets to the ground for users desiring local data storage and management. This will test Cloud-Optimized GeoTIFF (COG) data formats and local use of GEE algorithms.

The Future

- We expect to complete several **operational Data Cube** deployments in 2018:
  Vietnam, Taiwan, Uganda and an
  African Regional Data Cube (5 countries).
- Continue to develop collaborations with **Google, Amazon and World Bank**.
- We are planning a dedicated **Data Cube Paper Session** at the IGARSS Conference in Valencia, Spain (July 2018).
- We will develop many new **technical** elements in 2018:
  Python Notebook demos, QGIS Plugin,
  User Interface tools (pixel plotting, transect plots)
- We will develop and validate new **applications** in 2018:
  Land classification clustering and Water Quality (TSM and Chlorophyll).

Thank You