

# Tanzania Data Cube Tutorial

August 16, 2018

How to use the Africa Regional Data Cube (ARDC) for investigation of deforestation and agriculture phenology.

YouTube Recording: <https://youtu.be/J4zAOzEjbE>  
110 minutes

=====

## Step #1 - ARDC User Interface Tool

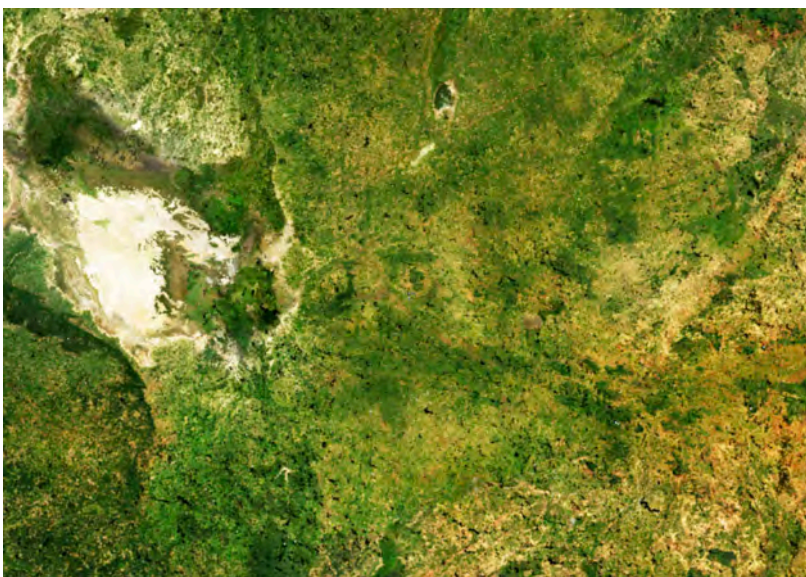
**Website:** <http://18.216.79.174/>

In order to closely analyze land data, it is often best to create a **cloud-free mosaic** and then export that image as a GeoTIFF for analysis in a more powerful tool, such as QGIS. The following steps show an example of creating a cloud-free mosaic for a region in Tanzania for 2016.

**Menu:** Tools > General > Custom Mosaic

The user needs to select a region and then select the parameters from the "Data Selection" menu.

**EXAMPLE:** True Color RGB, Max NDVI, Year 2016  
Location: Dodoma city and Lake Sulunga



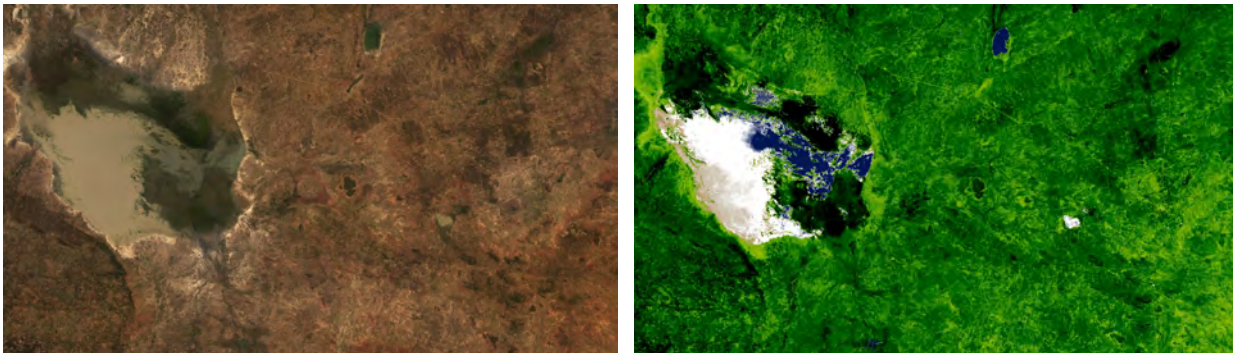
- \* NDVI mosaics take longer to run as they are a "synthetic" product.
- \* The True Color images will look "dark" on the UI screen. You can export the PNG image and then use Photoshop to create a "brighter" image as shown above.
- \* Review the Results and Output tabs for more information about the product.
- \* By selecting the image in the Output tab, you can export a PNG, GeoTIFF or NetCDF file of the final product. This will yield far more detailed information than what is shown on the UI screen.

**Menu:** Tools > Land > Spectral Indices

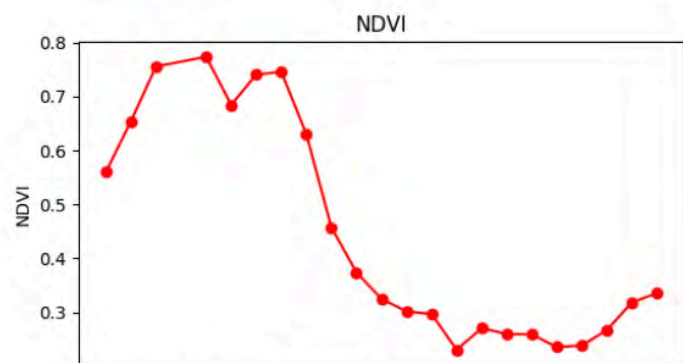
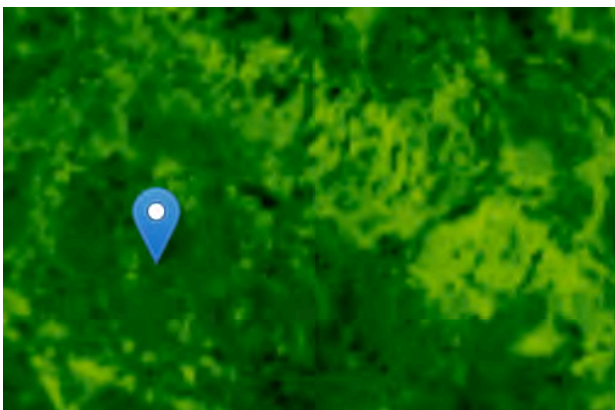
The user needs to select a region and then select the parameters from the "Data Selection" menu. There are several key indices for Agriculture, including NDVI and EVI.

**EXAMPLE:** EVI (Enhanced Vegetation Index), Median Mosaic, Year 2016  
 Location: Dodoma city and Lake Sulunga

Left (Google Earth), Right (EVI Product)



Try the "Marker Tool" to generate time series plots for a selected point



## Step #2 - ARDC Jupyter Notebooks

Website: <http://18.216.79.174:8000>

Password: ardcuser

=====

### Case Study #1 - Cloud Statistics

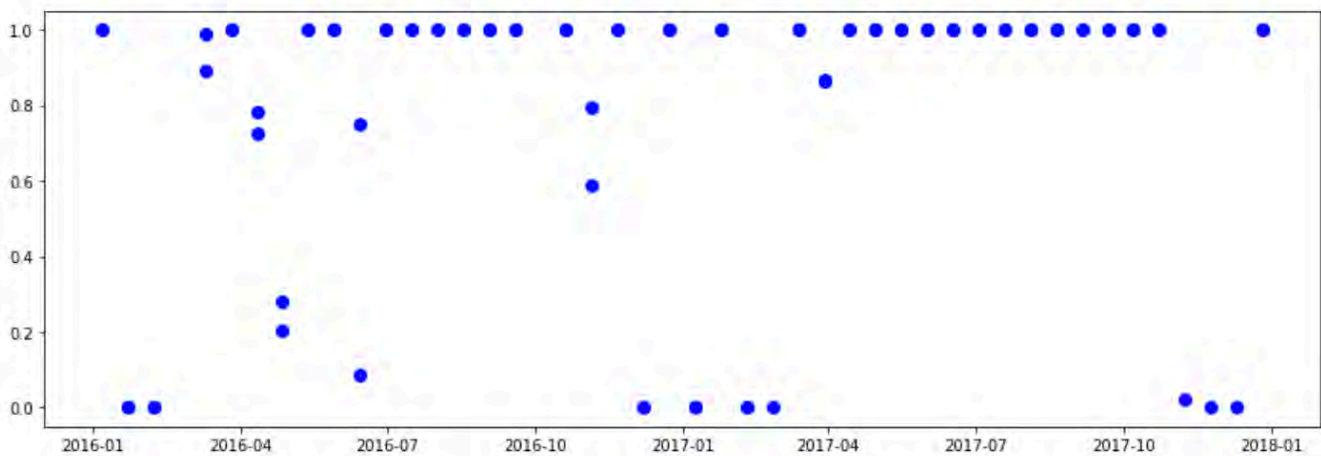
Filename: **Cloud\_Statistics**

Optical data (e.g. Landsat) will always have issues with clouds. This notebook calculates cloud statistics for selected regions for both Landsat-7 and Landsat-8 data cubes. This is valuable information for performing further analysis. For example, if there are extensive clouds for a season it may significantly impact the mosaic product or index values. Another example is that a user may want to identify a single date when there are no clouds to assess land features.

The user will select a mission (L7 or L8), a region, and a time period. The code will produce a table of time "slices" and their corresponding cloud statistics (e.g. clean percentage). In addition, the user will find a plot of cloud coverage over the time period and some code to create an RGB image of any time slice. Examples of the outputs are below.

	times	clean_percentage	clean_count
0	2016-01-06 07:56:00	1.000000	9225
1	2016-01-06 07:56:24	1.000000	9225
2	2016-01-22 07:56:01	0.000000	0
3	2016-01-22 07:56:25	0.000000	0
4	2016-02-07 07:55:57	0.000000	0
5	2016-02-07 07:56:21	0.000000	0
6	2016-03-10 07:55:48	0.990894	9141
7	2016-03-10 07:56:12	0.893550	8243
8	2016-03-26 07:55:39	1.000000	9225
9	2016-03-26 07:56:03	1.000000	9225
10	2016-04-11 07:55:33	0.725962	6697
11	2016-04-11 07:55:57	0.780921	7204
12	2016-04-27 07:55:29	0.204661	1888

*The table above displays the time slice (date), clean pixel percentage and the clean count (number of clear pixels).*



The figure above displays the clear pixel percentage for each time slice. Values of 1.0 are clear scenes with no clouds. Values of 0.0 are scenes filled with clouds.

## Case Study #2 - NDVI Phenology

Filename: **NDVI\_Phenology**

The Normalized Difference Vegetation Index (NDVI) is an index of plant “greenness” or photosynthetic activity, and is one of the most commonly used vegetation indices. It is possible to use this index for an agriculture area to assess the seasonal variations. This notebook focuses on "phenology" or the study of seasonal variations in agriculture. The use of NDVI Phenology can play an important role in processes that are relevant in practical agriculture, such as the timing of irrigation, fertilization, and crop protection.

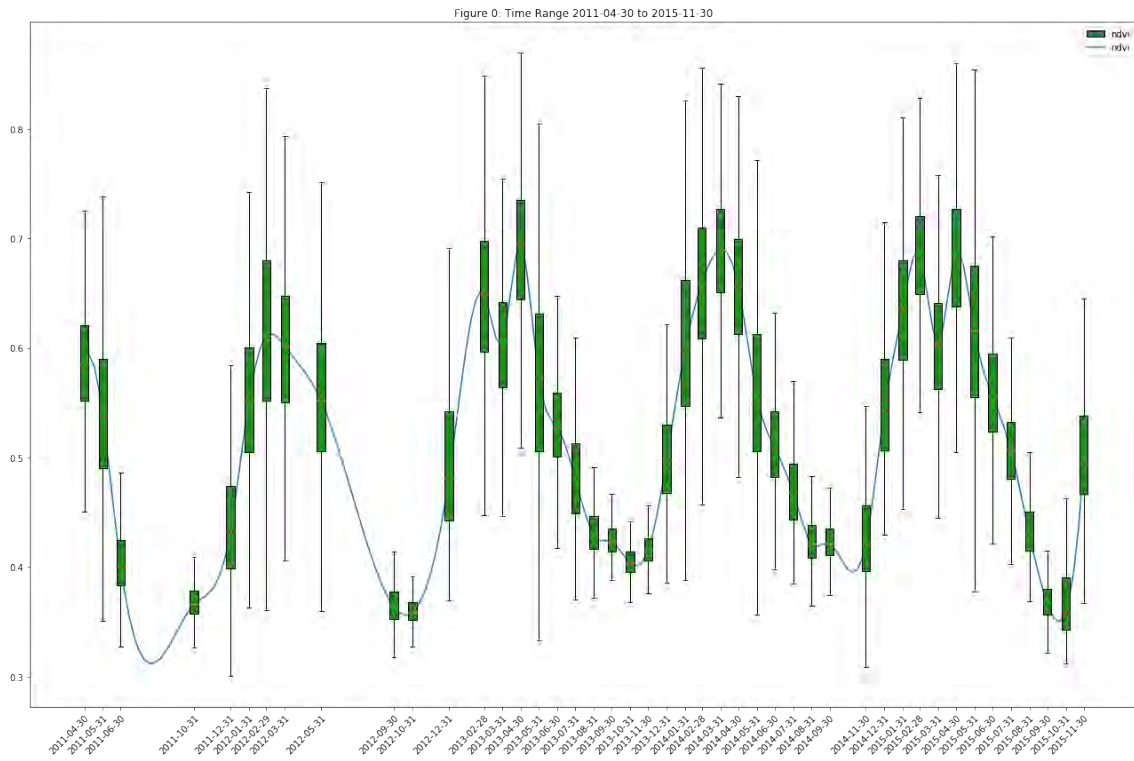
The example notebook uses the Tanzania data cubes from Landsat-7 and Landsat-8 over 3 regions that represent croplands and grasslands. If the user picks a time period that includes dates before 2013, the code will use Landsat-7 data. If the user selects a time period after April 2013, the code will use data from both the Landsat-7 and Landsat-8 missions.

At the end of the notebook you will find some code that creates 4 different NDVI phenology box-whisker plots. These are summarized below with example outputs.

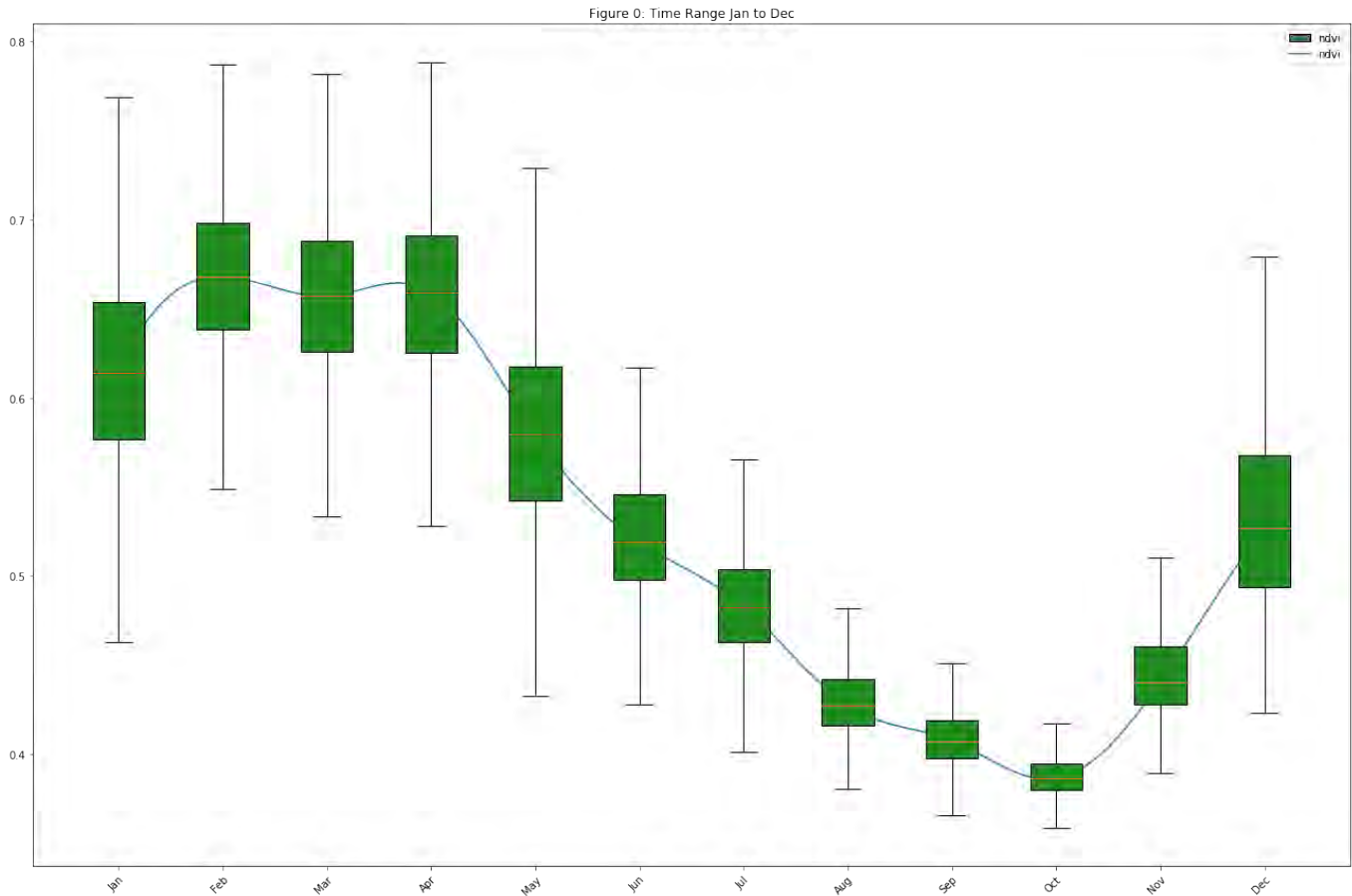
Week = Data is "binned" in weekly intervals across the entire time period  
 Month = Data is "binned" in monthly intervals across the entire time period  
 Week\_of\_Year = Data is "binned" in weekly intervals across a year  
 Month\_of\_Year = Data is "binned" in monthly intervals across a year



*Grassland region in northern Tanzania*



*NDVI Phenology monthly product for 2011 thru 2015. There are numerous data gaps in 2011 and 2012 due to a single mission (Landsat-7). After April 2013, the Landsat-8 mission allowed more regular coverage of the region.*



*Example NDVI Phenology product over the months of the year. Data is combined from 2011 thru 2015 (5 years) using Landsat-7 and Landsat-8 missions. The graph shows the peak NDVI in the rainy season (Dec-May) and then the low NDVI in the dry season (Aug-Nov).*

These products can be very helpful for the study of agriculture. For example, the user can compare crop health (NDVI peak) over multiple years and compare those data with known crop yield and climate data to develop forecast models for the future. An analysis of the "month-of-year" plot allows users to identify the start-of-season and end-of-season cropping periods which may be useful for crop classification (e.g. crop type).

## Case Study #3 - NDVI Anomaly

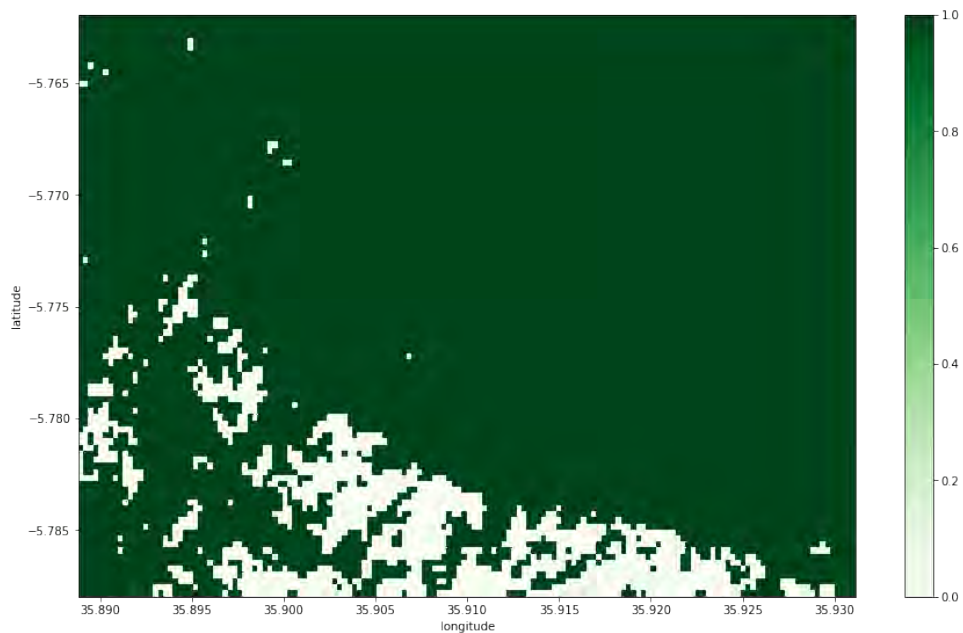
Filename: [NDVI\\_Anomaly](#)

This notebook compares NDVI between two time periods to detect land change. In the case of deforestation, the NDVI values will reduce from (0.6 to 0.9 ... typical for forests) to lower values (<0.6). This change can be detected and used to investigate deforestation or monitor the extent of the land change. The performance of this notebook is MUCH faster (e.g. >10x) than the PyCCD code.

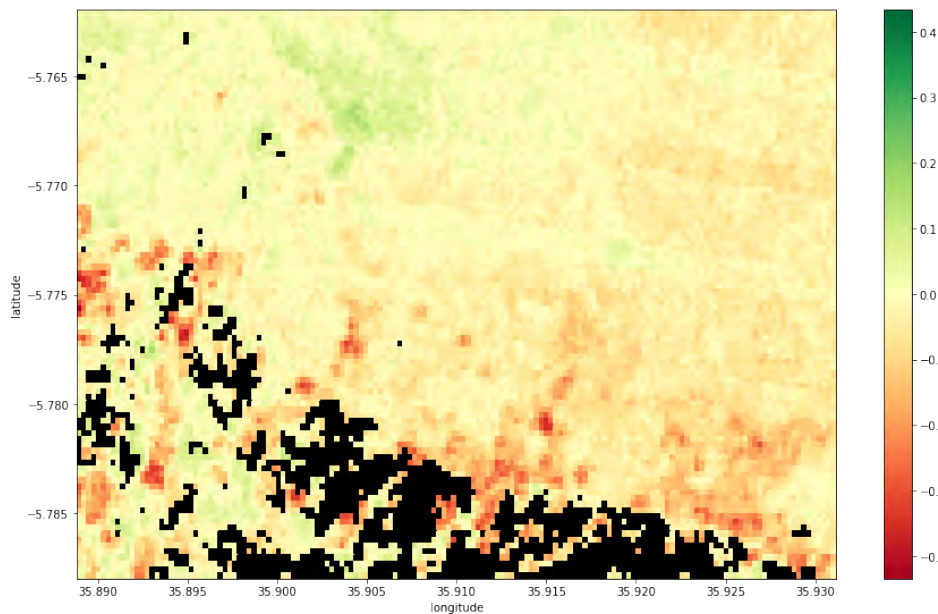
The user will select a dataset (L7 or L8), an analysis region, a baseline time period, an analysis time period, and a method for mosaics (e.g. median, max NDVI, Most Recent Pixel). Due to cloud cover, the baseline and analysis time periods may be long to allow a sufficient number of clear pixels for the analysis. In the case of Landsat-7, the "banding" issue" also requires more scenes to obtain enough clear data. The user should refer to the "Cloud Statistics" notebook for more information on cloud filtering.

The user will also select an "NDVI Baseline Threshold Range" which will screen pixels by their NDVI. For example, if the user is interested in only considering pixels that were likely forests in the baseline image, then the "NDVI Baseline Threshold Range" may be 0.6 to 0.9.

Below you will find a number of output products for the East Chenene Forest Reserve in central Tanzania. Some of these products are compared with output products from PyCCD and Global Forest Watch.

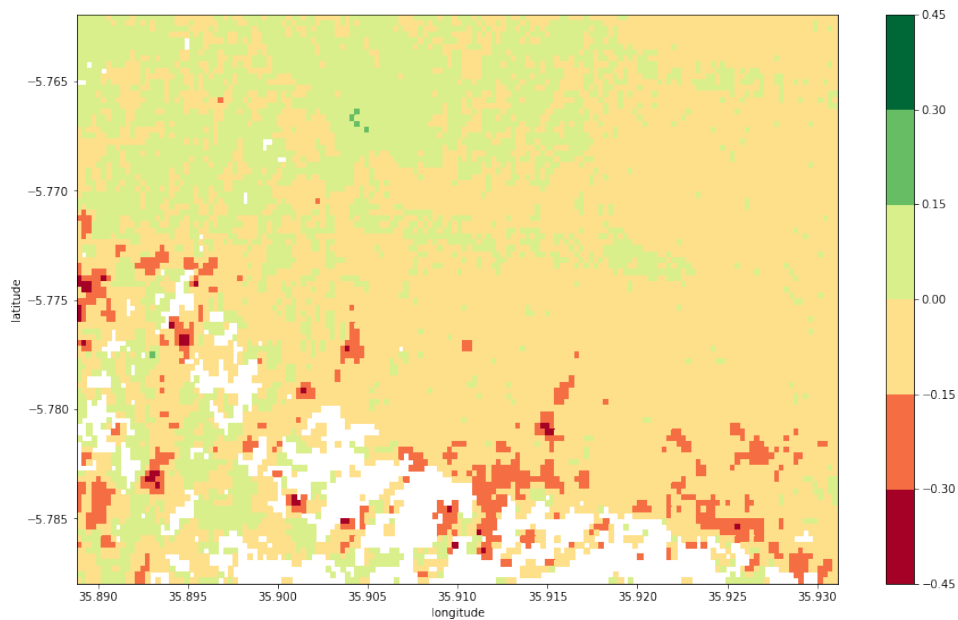


*Baseline maximum NDVI threshold image (2004 to 2007 composite). GREEN pixels have an NDVI range of 0.6 to 0.9 which is typical for forests. The areas in white are likely grasslands with lower NDVI.*

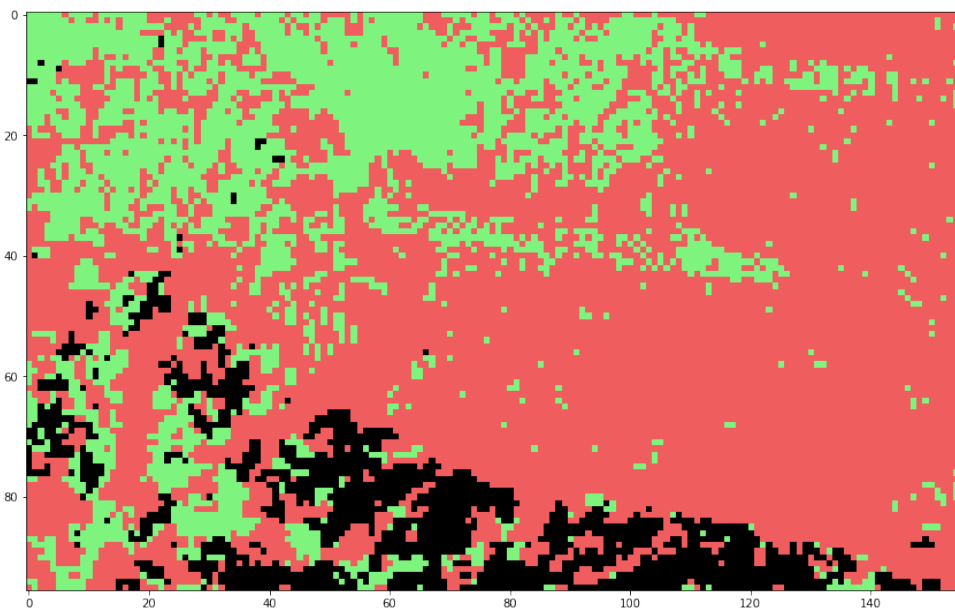


*NDVI anomaly product. The product compares the NDVI from the baseline maximum NDVI mosaic (2004 to 2007) to the analysis period (2013). Areas of deforestation will appear in ORANGE/RED as the NDVI has been reduced.*

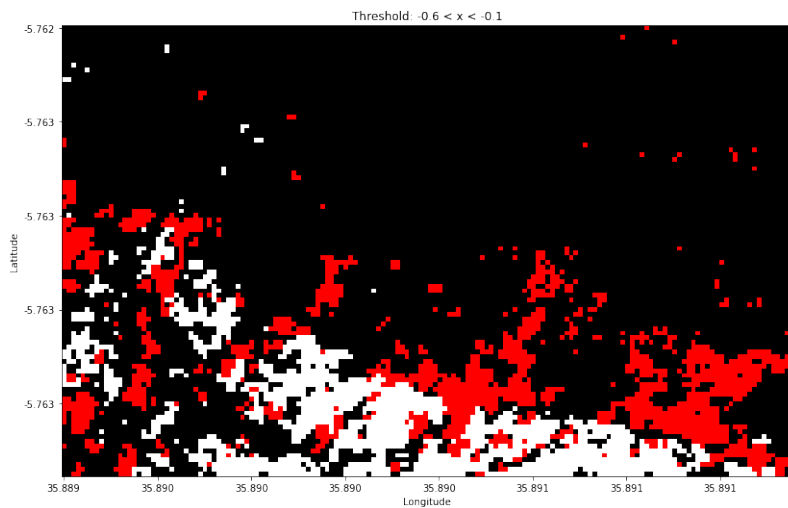




*NDVI anomaly product. The results shown here are similar to those in the previous figure, but the data has been "binned" into NDVI change categories. The product compares the NDVI from the baseline maximum NDVI mosaic (2004 to 2007) to the analysis period (2013). Areas of deforestation will appear in ORANGE/RED as the NDVI has been reduced.*

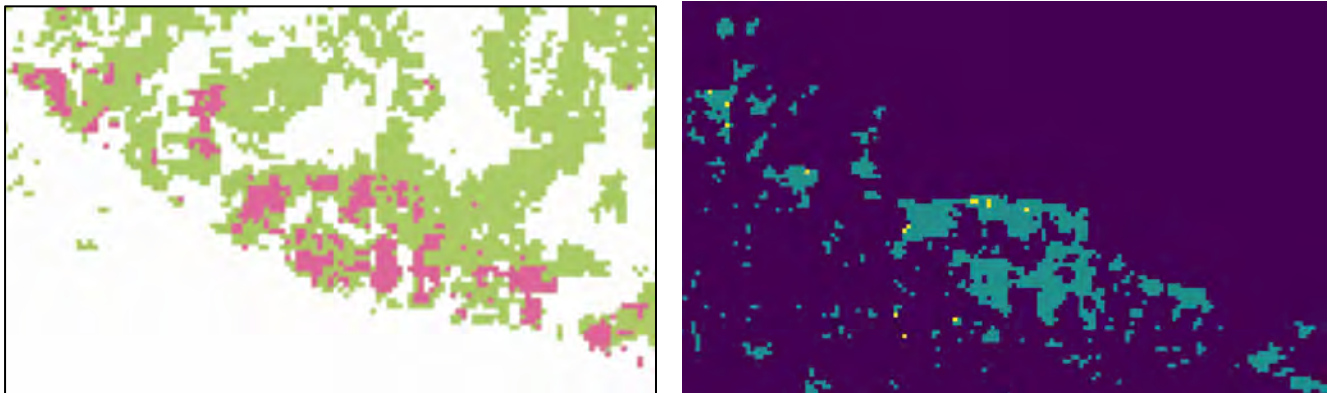


*NDVI anomaly product. The results shown here are similar to those in the previous two figure, but the data has been "binned" into two NDVI change categories. RED pixels have decreased NDVI and GREEN pixels have increased NDVI. BLACK pixels are non-forest.*



*NDVI anomaly threshold product. The results here show the pixels (RED) that have experienced a loss of 0.1 to 0.6 in NDVI between the baseline and analysis time period. Since these pixels were considered forest in the baseline image (NDVI of 0.6 to 0.9), they are likely deforestation. The WHITE pixels are non-forest.*

The NDVI anomaly threshold image (above) can be compared with other products for the same baseline (2004 to 2007) and analysis (2013) time period. Those outputs (shown below) display similar results for deforestation locations.



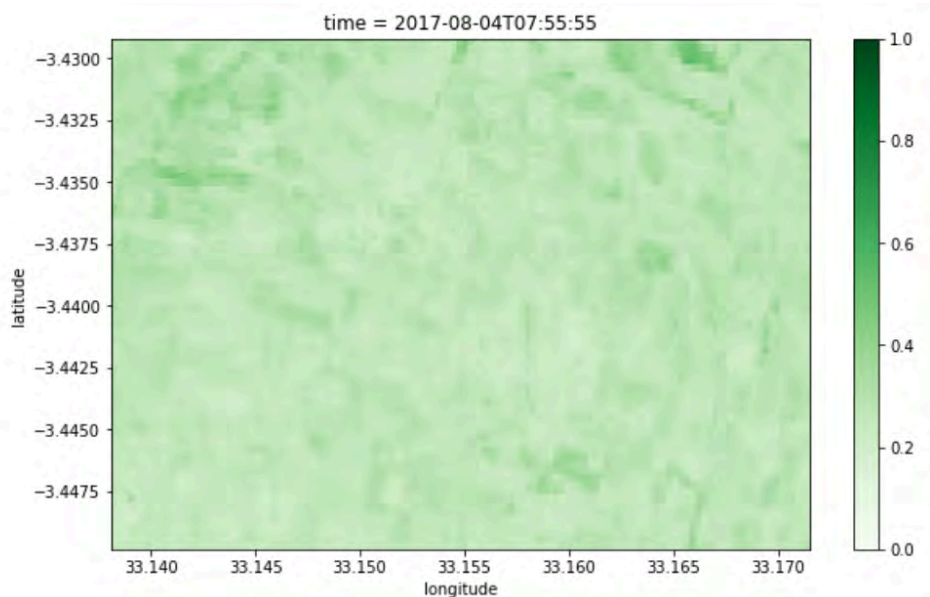
*Deforestation analysis products for 2004 thru 2013 from two sources. **Global Forest Watch** (<https://globalforestwatch.org>) deforestation results (left) show the forest mask (GREEN) and areas of deforestation (RED). The details of this algorithm are not known. The **PyCCD algorithm** deforestation results (right) show areas of land change (mostly deforestation) in GREEN.*

## Other Notebooks (not covered in this training)

### (a) Spectral Indices for Agriculture

Filename: [Agriculture\\_Spectral\\_Indices](#)

This notebook is a modification of the ARDC training notebook for Spectral Indices (Task C). The updates are focused on Tanzania and the analysis of vegetation indices (NDVI and EVI) for agriculture analysis. The notebook uses Landsat-7 or Landsat-8 data to create the following products: cloud-free mosaic, fractional cover, NDVI and EVI. The end of the notebook allows the user to define a threshold (range of values) and then display those pixels in the threshold range.



*Example NDVI output for a single grassland region in August 2017*

The **Normalized Difference Vegetation Index (NDVI)** is an index of plant “greenness” or photosynthetic activity, and is one of the most commonly used vegetation indices. Vegetation indices are based on the observation that different surfaces reflect different types of light differently. Photosynthetically active vegetation, in particular, absorbs most of the red light that hits it while reflecting much of the near infrared light. Vegetation that is dead or stressed reflects more red light and less near infrared light. Likewise, non-vegetated surfaces have a much more even reflectance across the light spectrum.

The **Enhanced Vegetation Index (EVI)** was developed as an alternative vegetation index to address some of the limitations of the NDVI. The EVI was specifically developed to: (a)

be more sensitive to changes in areas having high biomass (a serious shortcoming of NDVI), (b) Reduce the influence of atmospheric conditions on vegetation index values, and (c) correct for canopy background signals. EVI tends to be more sensitive to plant canopy differences like leaf area index (LAI), canopy structure, and plant phenology and stress than does NDVI which generally responds just to the amount of chlorophyll present.

## (b) PyCCD for Deforestation

Filename: [PyCCD\\_Chenene\\_With\\_Export](#)

This notebook was used to analyze the deforestation around the perimeter of the East Chenene Forest Reserve in Tanzania. The code uses the long Landsat-7 time series to detect changes in spectral response that are caused by land change and deforestation. The output identifies the location of the change (change volume) and the time of the first land change (change epoch). The output appears to match the Global Forest Watch (GFW) results. A new export feature was added to the end of this notebook to allow users to export the change location and dates for further analysis. Unfortunately, this code is **very intense** and takes a long time (>20 minutes) to run over small regions (0.02 deg square). Users should take extreme caution in using this code to perform analyses over larger regions as the run time will be very long and use the majority of computing resources.

## (c) GeoTIFF and NetCDF Exports

Filenames: [export\\_to\\_geotiffs\\_l7 + export\\_to\\_geotiffs\\_l8](#)

Filenames: [export\\_to\\_netcdf\\_l7 + export\\_to\\_netcdf\\_l8](#)

The ARDC training notebook for Data Export (Task F) did not include Landsat-8 or export to NetCDF. Several notebooks have been added to the Tanzania server for GeoTIFF and NetCDF export. These notebooks can be modified to produce custom export files for Tanzania analyses. Users should use extreme caution when running these notebooks as the export files can become quite large.

=====