# A Status Report on the use of the Open Data Cube (ODC) to support the United Nations Sustainable Development Goals (SDG)

Brian Killough, NASA, CEOS SEO
August 30, 2019

## SDG 6.6.1 > Change in the extent of water-related ecosystems over time

Python Notebook Algorithm:
http://52.54.26.108:8080/notebooks/Brian_Water_SDG_6_6_1_July2019.ipynb

*NOTE: This algorithm is currently on a password-protected server used by NASA. It will be posted on a public GitHub server by the end of 2019.*

This algorithm directly addresses the SDG 6.6.1a sub-indicator: spatial extent of water related ecosystems. Landsat analysis-ready data (ARD) is used with the Australia Water Observations from Space (WOFS) algorithm to classify every pixel (30m scale) as water or non-water through the time series. The algorithm allows the user to define a region of interest and define two separate times periods (baseline and analysis). The algorithm then calculates the water extent for each time period (never, sometimes, always) and quantifies these results in a matrix output. Finally, the two time periods are compared to create a final change map. This change map shows 4 classes: never water, always water, non-water to water, and water to non-water. These results are then quantified in a table (percent change and area) and classified according to the SDG Methodology for ecological classes (e.g. A,B,C,D,E).

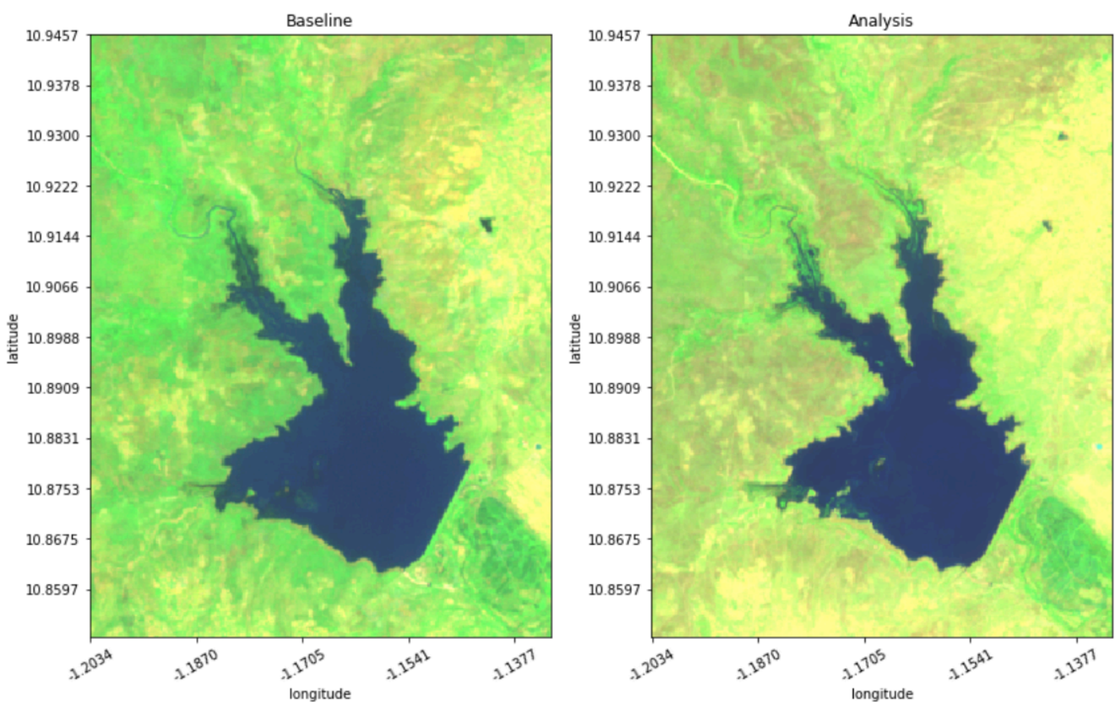


Figure 1. Median cloud-filtered composites of the Tono Dam in Ghana. The baseline period (2002) is on the left and the analysis period (2017) is on the right.

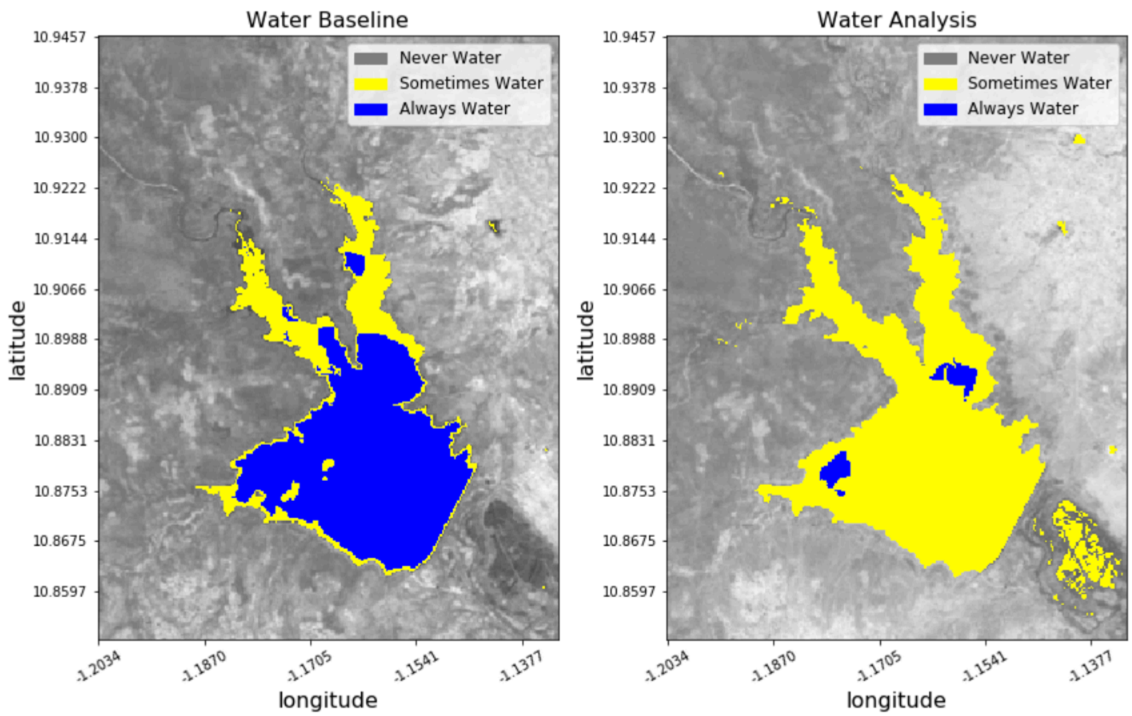


Figure 2. Water extent results for the baseline and analysis period over the Tono Dam in Ghana. The analysis period (2017, on right) shows the lake is highly variable and likely only filled during the annual rainy season.

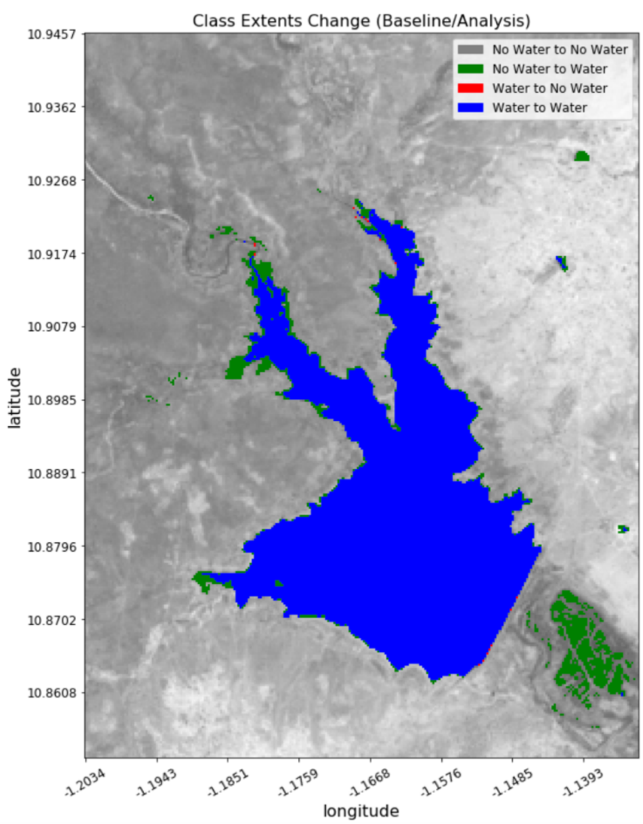


Figure 3 (left). The final change product for the Tono Dam in Ghana comparing 2002 to 2017. This result is based on a classification where a pixel is considered water if it experienced water at any time during the year. The "green" regions to the south and north show areas of new water extent that were not classified as water in 2002. The table below shows the full results and the total net change of 1.98% increase in water extent over 15 years. This result can be classified as "Unmodified Natural, Class A" according to the SDG 6.6.1 Methodology guidelines.

| | Number | Percent | Area (m^2) |
|---|---|---|---|
| **No Water to No Water** | 75701 | 0.823240 | 6.774039e+07 |
| **No Water to Water** | 1803 | 0.019607 | 1.613399e+06 |
| **Water to No Water** | 18 | 0.000196 | 1.610714e+04 |
| **Water to Water** | 14433 | 0.156957 | 1.291525e+07 |
| **Net Change** | 1821 | 0.019803 | 1.629506e+06 |
| **Unknown** | 0 | 0.000000 | 0.000000e+00 |

The analyses presented here can be reproduced for any location where a data cube of Landsat data exists. As of now, these data cubes exist in several countries around the world. By the end of 2019, USGS will host all of the Landsat ARD on the Amazon Cloud. At that point, it will be possible to run this same algorithm at any location in the world. The open source algorithm will be available on GitHub, but users will need to purchase (or receive credits) to create an Amazon computing instance (EC2) in order to run similar analyses.

It should be noted that the water classification results in this report are based on the open source Australian Water Observation from Space (WOFS) algorithm (reference below). The algorithm has been compared with the European Commission's Global Surface Water results and the results are very consistent.

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, at al. (2016). Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of the Environment*, 174, 341–352.

## SDG 11.3.1 > Ratio of Land Consumption Rate to Population Growth Rate

Python Notebook Algorithms:
http://52.54.26.108:8080/notebooks/Urbanization_General_30Aug2019.ipynb
http://52.54.26.108:8080/notebooks/UN_SDG/Urbanization_Complete_30Aug2019.ipynb

*NOTE: These algorithms are currently on a password-protected server used by NASA. They will be posted on a public GitHub server by the end of 2019.*

These algorithms directly address the SDG 11.3.1: ratio of land consumption rate to population growth rate. The indicator requires land consumption rate (calculated from satellite data and often called "urbanization") and population growth rate (acquired from external sources). Overall, the SDG 11.3.1 indicator (equation below) provides a metric for determining whether or not land consumption is growing at the same rate as population. If the indicator is above 1.0, then land is being used at a rate faster than the population is growing which may create issues with availability of agriculture land and water resources as well as deforestation. If the indicator is below 1.0, then population is growing faster than the urban extent and there will be more pressure on urban regions such as water supply, traffic, sanitation, etc. Therefore, SDG 11.3.1 is important to urban planning and decision-making.

$$Indicador\ 11.3.1 = \frac{Land\ Consumption\ rate}{Population\ growth\ rate}$$

The two algorithms listed above serve different purposes. The **first algorithm** (Urbanization_General) uses two possible indices to classify land as "urban". These indices include Normalized Difference Buildup Index (NDBI) and Fractional Cover (FC) Bare Soil. This notebook is best used to test the validity of the urban extent results for different time periods and compare those results with online resources (listed below) or historic maps. To date, the results have only been visually compared with the online resources below, but future revisions will attempt to add more rigor to these classifications. For example, supervised classification using training data (e.g. local maps of urban extent, or ESA-CCI land classification data) would greatly improve the accuracy of the results.

A sample analysis case, using the first algorithm, was run over the Dar es Salaam region of Tanzania. Below are the data cube analysis results and comparisons with online tools.

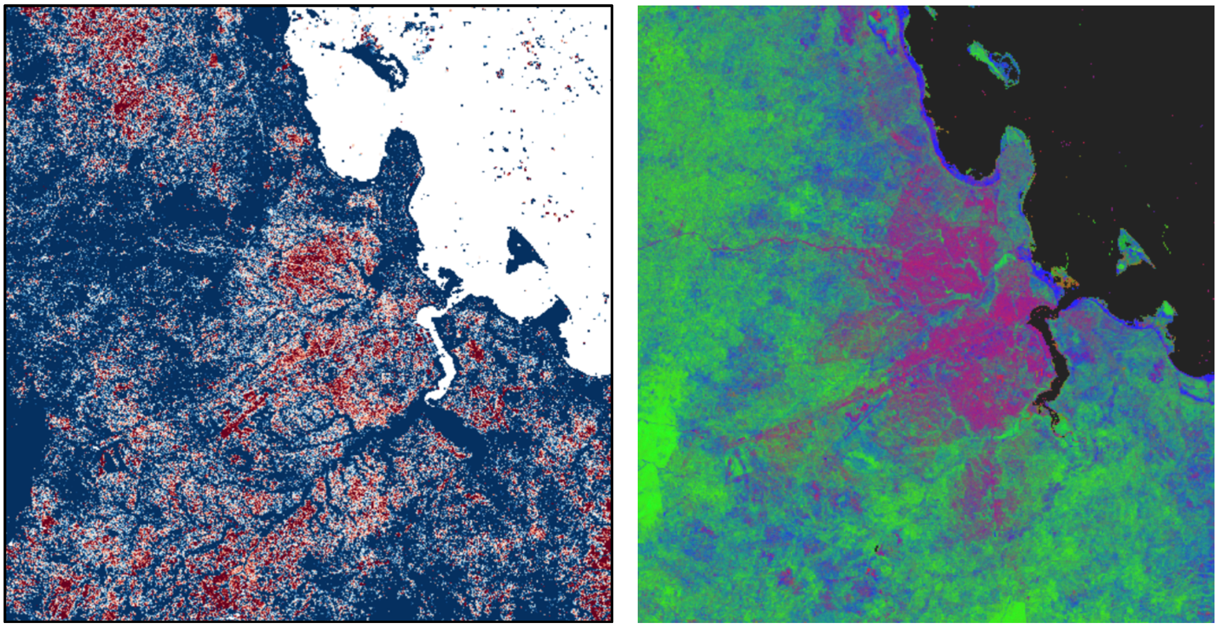

Figure 4. Normalized Difference Buildup Index (NDBI) (on left, range 0.0 to 0.2) and Fractional Cover (FC) (on right) products for Dar es Salaam, Tanzania in year 2000 using Landsat-7 data.

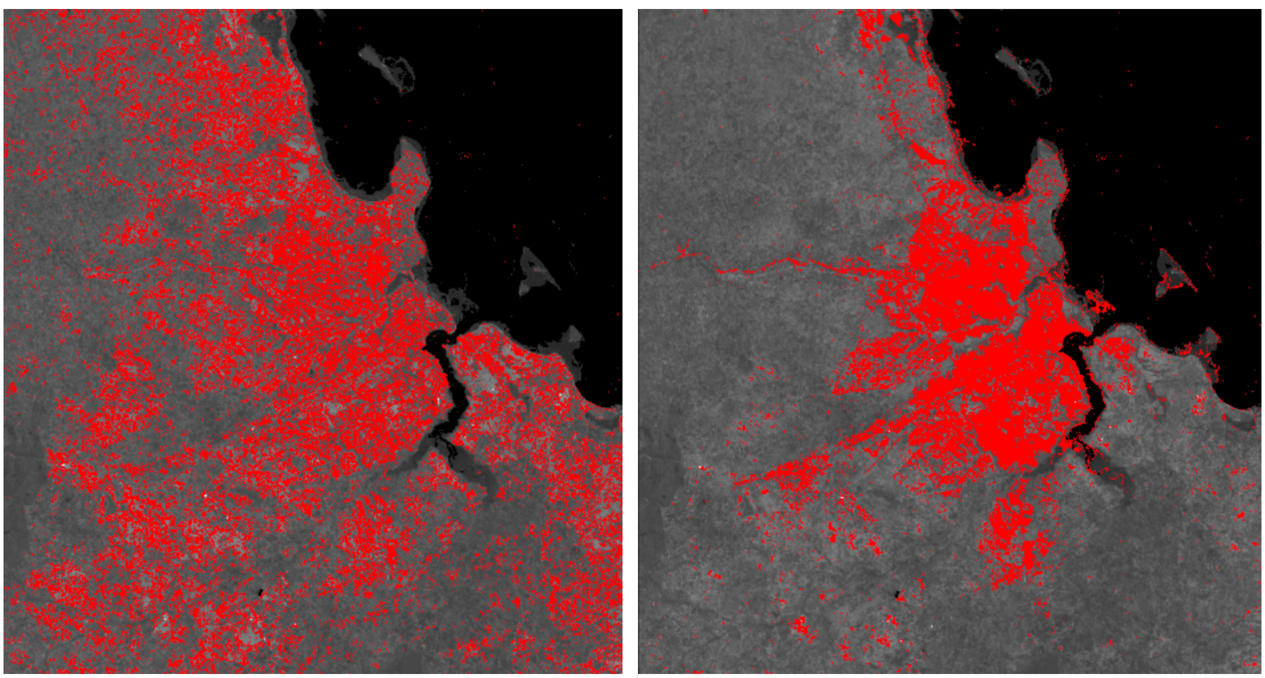

Figure 5. Urban extent estimates for Dar es Salaam in 2000 are estimated using an NDBI threshold (left) and a Fractional Cover - Bare Soil threshold (right). The thresholds were adjusted to visually correlate with the urban extent maps in Figure 6. The NDBI results were far more "noisy" compared to the FC results, which closely matched the urban footprint of the ESA and Trends.Earth products. The final thresholds were NDBI < 0.10 and FC-BS > 0.25.

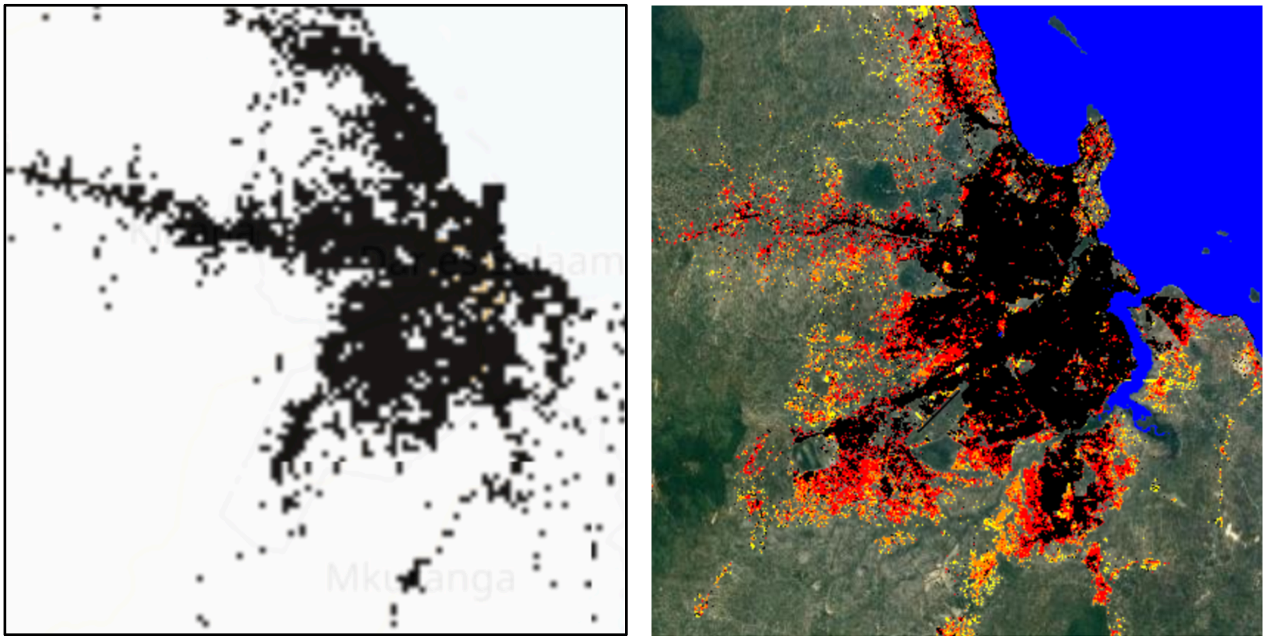

Figure 6. Urban extent map from the ESA Urban TEP tool (left, year 2000) and the Trends.Earth Urban Mapper (right, year 2000 in black, year 2005 red, year 2010 orange, year 2015 yellow) for Dar es Salaam, Tanzania. These maps can be compared with the urban extent estimates in Figure 5 using the data cube algorithm. The web links for these tools are shown below.

ESA Urban TEP (Global Human Settlement Layer): https://urban-tep.eu/
Trends.Earth Urban Mapper: https://geflanddegradation.users.earthengine.app/view/trendsearth-urban-mapper

The second algorithm (Urbanization_Complete) adds more complexity in that it includes shapefiles and population data from the Gridded Population of the World (GPW) Version 4 product (https://sedac.ciesin.columbia.edu/data/collection/gpw-v4). The GPWv4 data includes population data at 5 year intervals which can be used with the urban extent results to make a full calculation of the SDG 11.3.1 indicator. In this case, the analysis was run over a 15-year period between 2000 and 2015. According to the GPW data, the population growth rate was 5.97% per year over the region (Figure 7). The results from the data cube analysis (Figure 8) calculate an urban growth rate of 3.45% per year. These two results can be combined in the SDG Indicator equation to yield the final result, the ratio of land consumption rate to population growth rate, of 0.58. This suggests the land consumption rate is lower than the population growth rate and the population per unit area is growing rapidly.

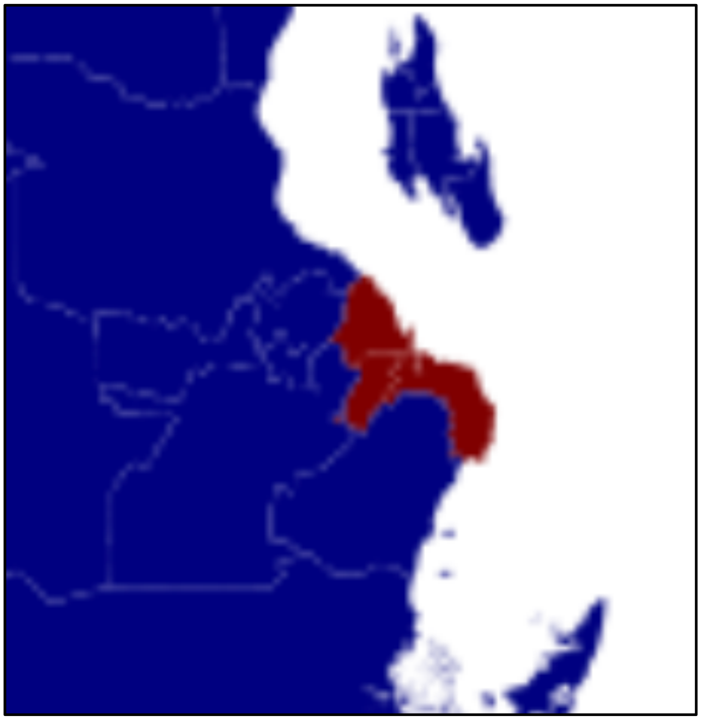

Figure 7. The GPW population data use shapefiles to organize its data. These shapefiles can be used to constrain the analysis region. The population growth rate of the Dar es Salaam analysis region (in red) was 5.97% per year.

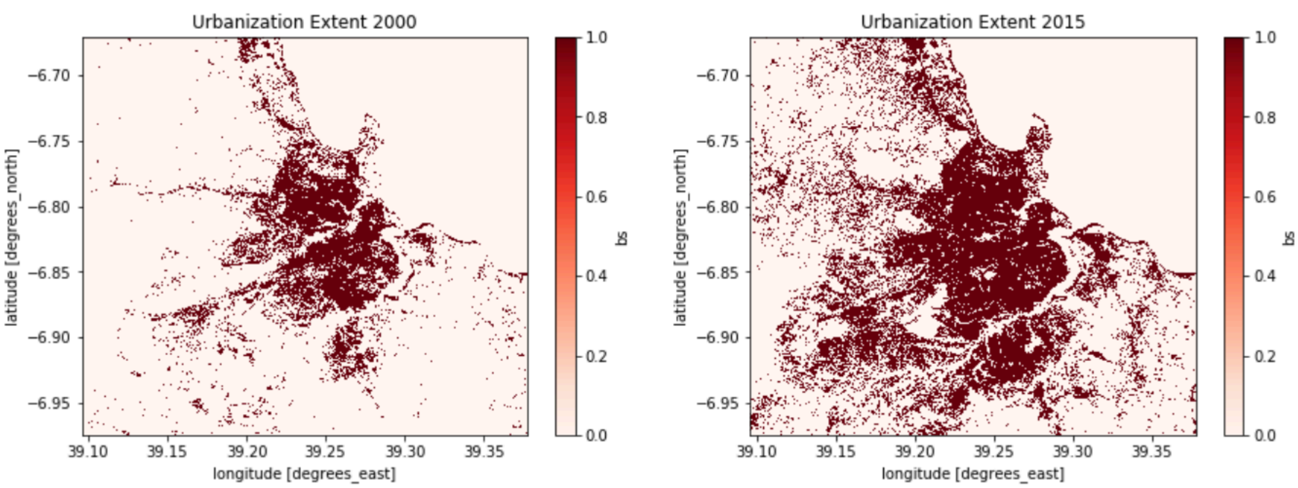


Figure 8. The data cube analysis results show the urban extent in 2000 (left) and 2015 (right) using a fractional cover (bare soil) index. The growth rate over this period was 3.45% per year.

Future versions of this algorithm will consider a number of enhancements. First, there is a desire to use actual urban maps in order to improve urban classification. Next, there is a need to crop the region to specific county shapefiles in order to perform this analysis for any county. With new population census data and new satellite data in 2020, it will be possible to update this analysis for the current condition and provide critical information for reporting on SDG indicators.

## SDG 15.3.1 > Proportion of Land that is Degraded over Total Land Area

Python Notebook Algorithms:
(1) http://52.54.26.108:8080/notebooks/UN_SDG_15_3_1_ESA_CCI_Kumasi.ipynb
(2) http://52.54.26.108:8080/notebooks/UN_SDG_15_3_1_FROM-GLC_Kumasi.ipynb
(3a) http://52.54.26.108:8080/notebooks/15_3_1_Land_Cover_Classification.ipynb
(3b) http://52.54.26.108:8080/notebooks/15_3_1_Change_Detection.ipynb

*NOTE: These algorithms are currently on a password-protected server used by NASA. It will be posted on a public GitHub server by the end of 2019.*

These algorithms directly address SDG 15.3.1, the proportion of land that is degraded over total land area. More specific, these algorithms only address the "land cover" sub-indicator by classifying the land type at every pixel (6-class IPCC system) and comparing the change in this classification between two time periods. Finally, a pixel is evaluated as "degraded" if it has changed from a productive ecosystem to a non-productive ecosystem (e.g. cropland to settlement). To guide this effort, the **Good Practice Guidance for SDG Indicator 15.3.1** was used. This report was prepared by Commonwealth Scientific and Industrial Research Organisation (CSIRO) for the United Nations Convention to Combat Desertification (UNCCD).

Three different algorithms are used to assess SDG 15.3.1 (see above). Below is a description of each algorithm and sample outputs for a use-case over Kumasi, Ghana. This city in Ghana has experienced significant urban growth and land change from 2000 to 2015. In each example, the final output is a "change matrix" that reflects the percent change in land classification (highlighting which changes are considered degradation) and an NDVI trend based on the Mann-Kendall Z-score (see Good Practice Guidance).

**Case 1:** This algorithm uses the ESA-CCI dataset (300m spatial resolution) containing annual land cover classes for the period 1992-2015, produced by the Catholic University of Louvain Geomatics as part of the Climate Change Initiative of the European Spatial Agency (ESA). This dataset includes a typology of 22 land cover classes based on the UN Land Cover Classification System. The data is used to train a random forest classifier on a median mosaic at two time periods (e.g. year). These 22 classes were then further simplified into the 6-class IPCC system (e.g. forest, grassland, cropland, wetland, settlement, other land) for the final change matrix.

**Case 2:** This algorithm uses the FROM-GLC (Fine Resolution Observation and Monitoring of Global Land Cover) dataset (30m spatial resolution, based on Landsat, 11 classes, 67% land classification accuracy). The data is used to train a random forest classifier on a median mosaic at two time periods (e.g. year). These 11 classes were then further simplified into the 6-class IPCC system (e.g. forest, grassland, cropland, wetland, settlement, other land) for the final change matrix.

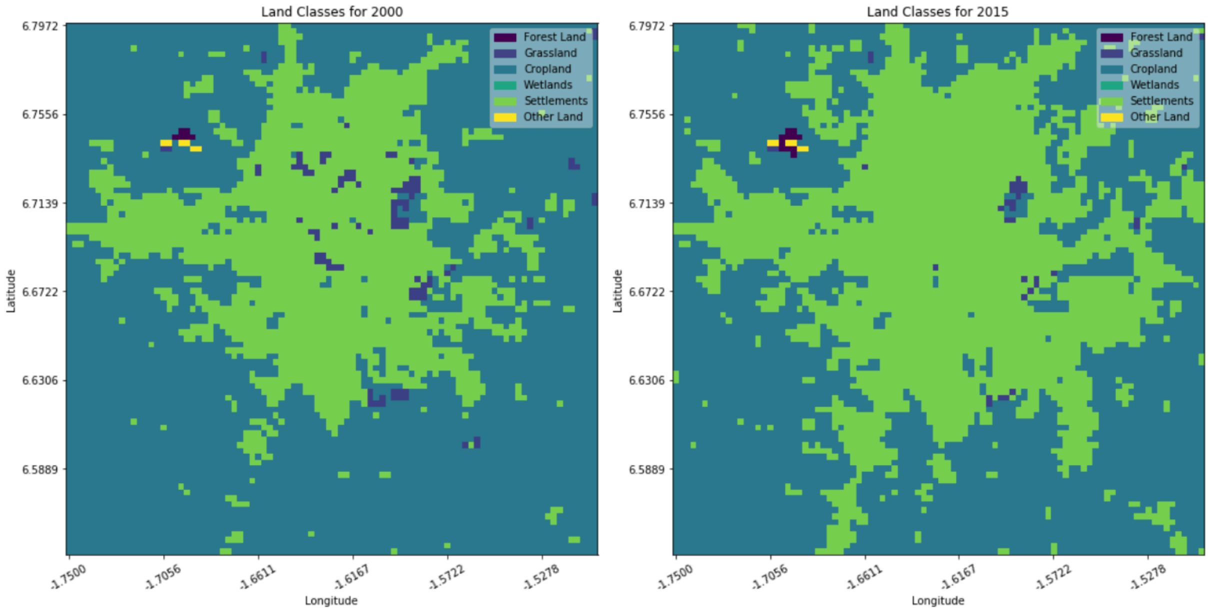


Figure 9. Land classification results for Kumasi, Ghana in 2000 (left) and 2015 (right). The ESA-CCI dataset (22 classes) was used to train a random forest classifier (83% cross validation error) and then further reduced into 6-classes. The urban growth (settlements area in green) is evident between the two time periods.

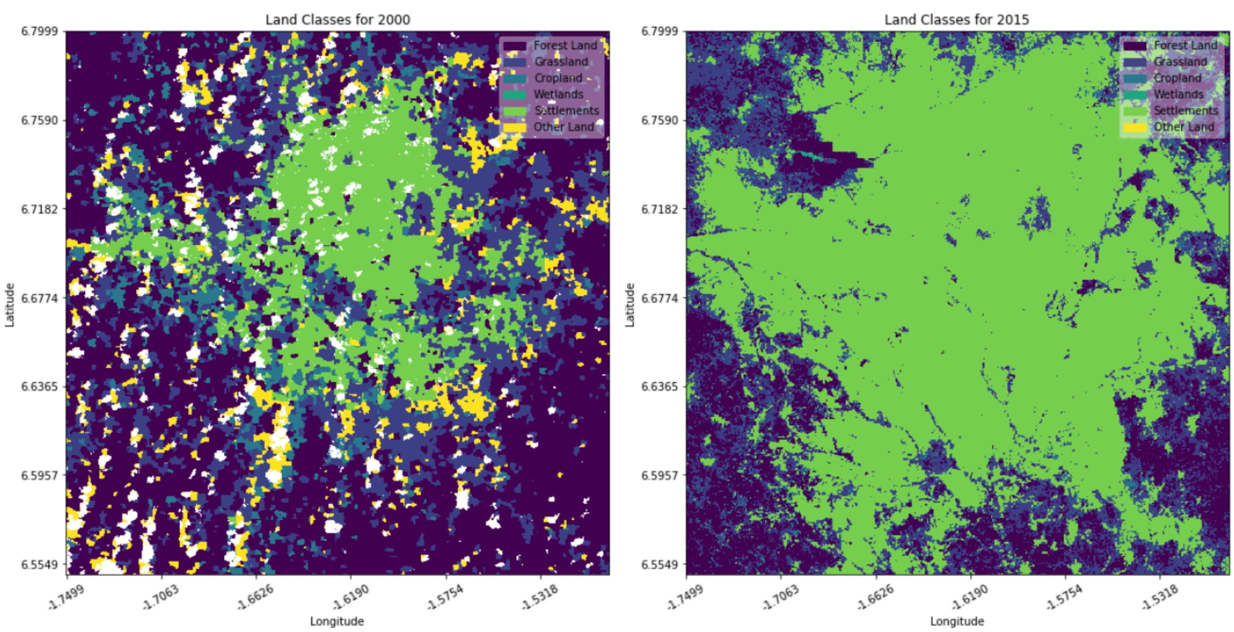


Figure 10. Land classification results for Kumasi, Ghana in 2000 (left) and 2015 (right). The FROM-GLC dataset (11 classes) was used to train a random forest classifier (64% cross validation error) and then further reduced into 6-classes. The urban growth (settlements area in green) is evident between the two time periods.

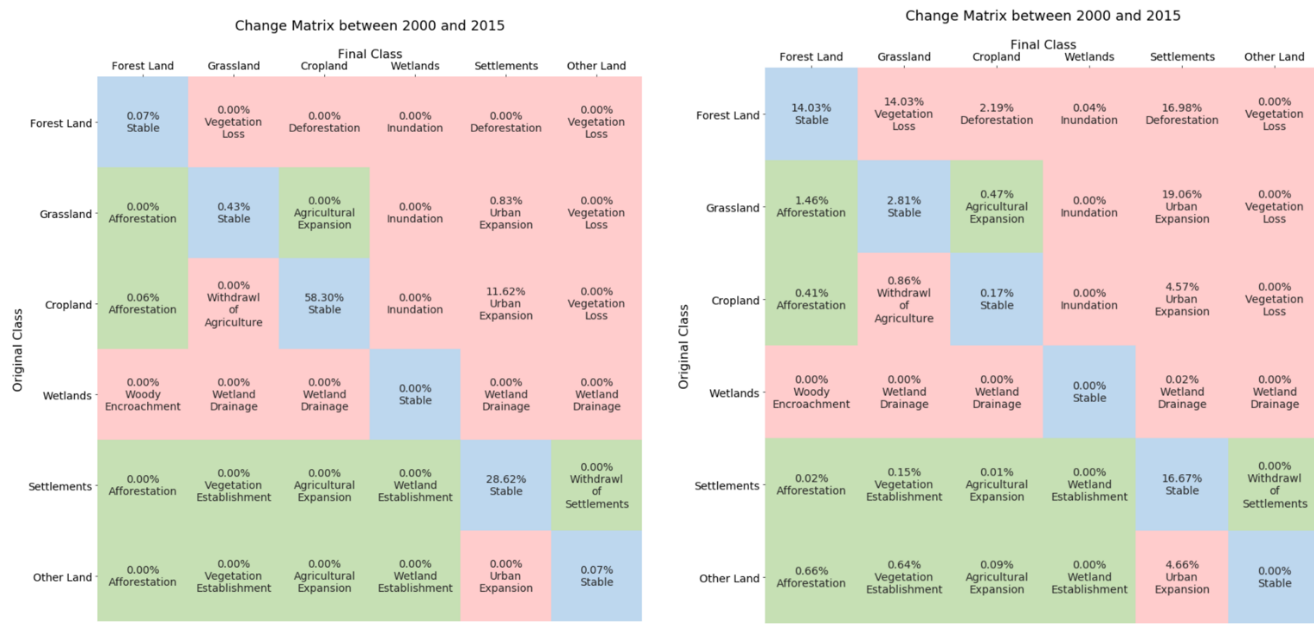


Figure 11. Land classification change results for Kumasi, Ghana between 2000 and 2015 using the ESA-CCI dataset (left) and FROM-GLC dataset (right). The elements shaded in "red" are considered land degradation, such as "urban expansion".

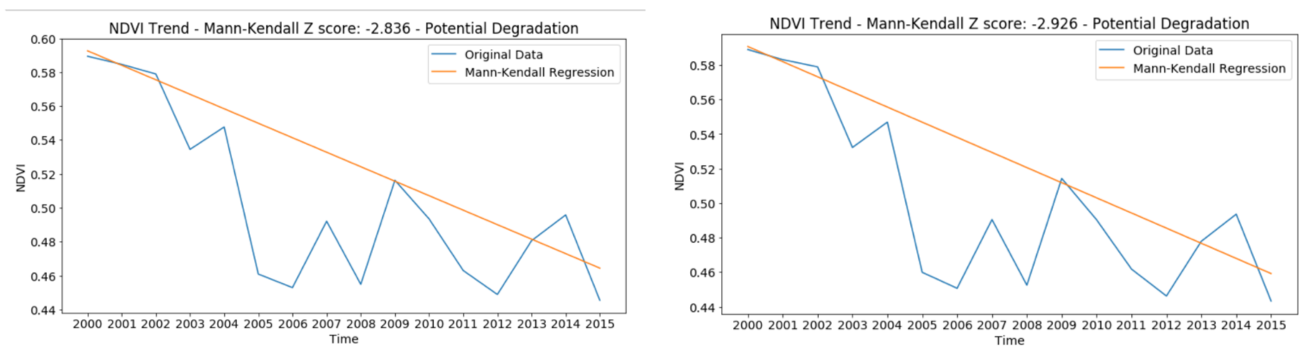


Figure 12. Peak NDVI trend for Kumasi, Ghana between 2000 and 2015 using the ESA-CCI dataset (left) and FROM-GLC dataset (right). The Mann-Kendall Z-scores suggest "potential degradation" according to the Good Practice Guidance document.

**Case 3a/3b:** In order to simplify the analysis and produce first-order results, a non-supervised land classification decision tree was developed (Figure 13). This decision tree needs validation testing with actual land classification maps to improve accuracy, but can be used independently from classification data, such as ESA-CCI or FROM-GLC. As shown below, the initial results are promising but additional work is needed to adjust the algorithm thresholds and evaluate accuracy according to know classification maps.

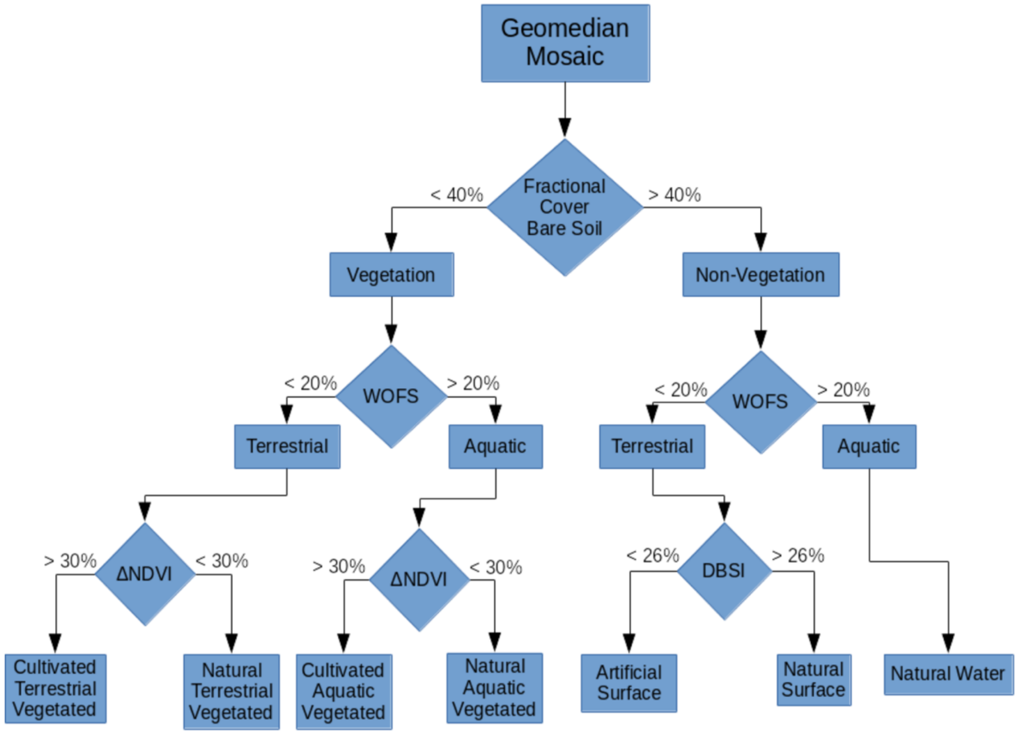


Figure 13. An unsupervised land classification decision tree can be used with Landsat data to classify pixels into 7 different land types. This classification system is consistent with the FAO 8-class land classification system but combines aquatic non-vegetation (natural vs. artificial) into one class.

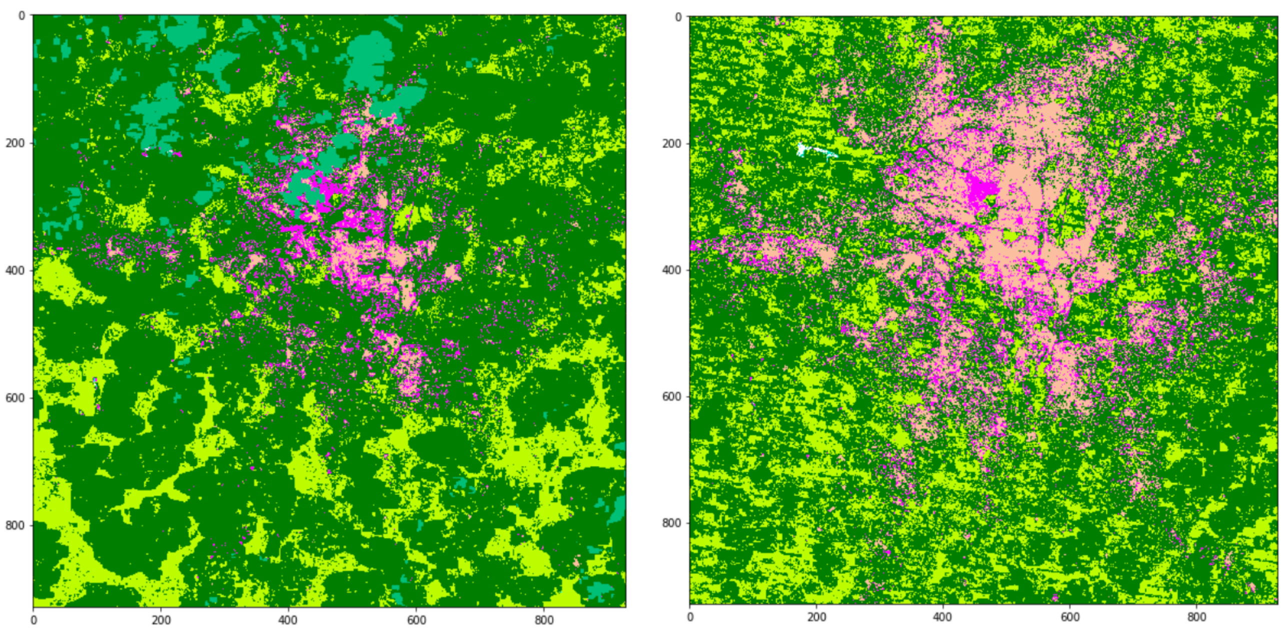

Figure 14. Land classification results for Kumasi, Ghana in 2000 (left) and 2015 (right) using a 7-class unsupervised decision tree (Figure 13). The urban growth (settlements area in purple) is evident between the two time periods. Unfortunately, the Landsat "banding" issue with the 2015 data causes some "noise" in the final product.

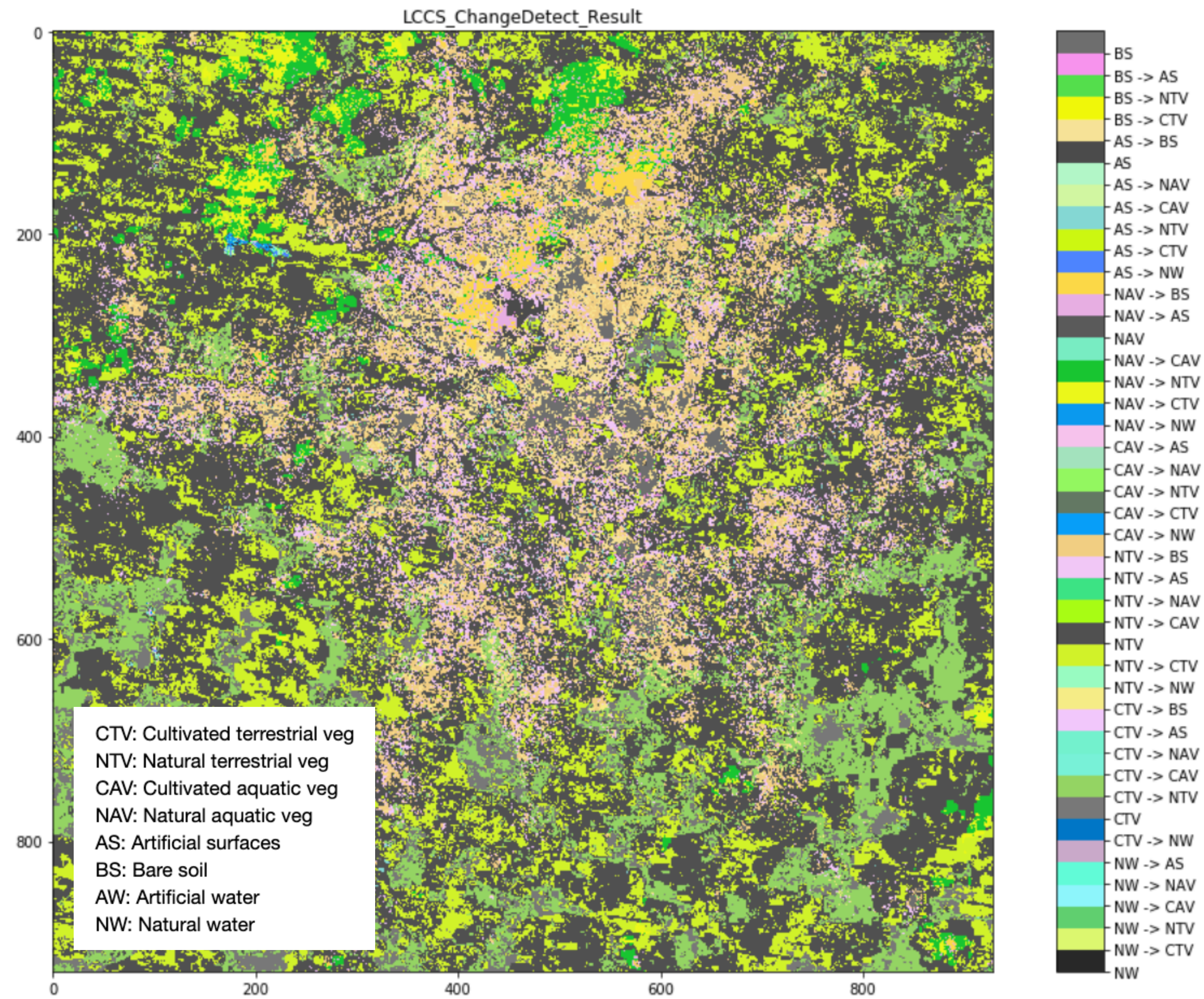


Figure 15. Land classification change results for Kumasi, Ghana between 2000 and 2015 using using a 7-class unsupervised decision tree (see Figure 13).