

# West Africa Forest Degradation Data System (WAFForDD)

User Guide

Version 2.1

February 28, 2020

Francis K. Dwomoh<sup>1</sup>, Michael C. Wimberly<sup>1</sup>, Dawn Nekorchuk<sup>1</sup>, Izaya Numata<sup>2</sup>, Foster Mensah<sup>3</sup>

1 Department of Geography and Environmental Sustainability, University of Oklahoma, Norman OK, USA

2 Geospatial Sciences Center of Excellence, South Dakota State University, Brookings SD, USA

3 Centre for Remote Sensing and Geographic Information Services, University of Ghana, Legon.

This project was supported by a SERVIR Applied Sciences Team grant from the NASA Applied Sciences Program (NNX16AN22G).

## 1. Introduction

In the forest zone of Ghana, protected areas, including forest reserves, are the only significant remnants of natural tropical forests. These protected areas are threatened by multiple change agents including logging, wildfire, and agricultural encroachment leading to widespread forest degradation (Benefoh et al. 2018, Dwomoh et al., 2019, Hawthorne et al., 2016). Historical assessment and annual monitoring of forest degradation is therefore critical for making informed conservation decisions. However, remote sensing monitoring efforts in the region has been hindered by severe gaps in the historical data record due to persistent cloud cover, infrequent satellite revisits, and inadequate field data characterizing forest structure.

The West Africa Forest Degradation Data system (WAFForDD) consists of a set of tools for mapping annual forest canopy cover and estimating trends in forest loss and degradation throughout the forest zone of Ghana. WAFForDD maps forest degradation by leveraging the Landsat archive, Google Earth high-spatial resolution imagery, and Google Earth Engine's cloud-based data processing capabilities. The system was designed in collaboration with the Forestry Commission of Ghana (FCG) to generate outputs that can be used to estimate emissions and carbon sequestration in support of the Ghana Cocoa Forest REDD+ program. Specific requirements included the ability to map locations of deforestation, forest degradation, and recovery on an annual basis based on transitions between closed forest, open forest, and low tree cover classes defined by the FCG. The software was developed in collaboration with the Centre for Remote Sensing and Geographic Information Services (CERSGIS) at the University of Ghana with the goal of transferring the software to CERSGIS for sustained application and continued development.

## 2. Components of WAFForDD for Generation Annual Forest Degradation Estimates

There are four scripts use to generate annual forest degradation estimates in WAFForDD (Figure 1).

- 1) WAFForDD\_1\_Image\_Compositing
- 2) WAFForDD\_2\_CanopyCover\_RandomForest\_Model
- 3) WAFForDD\_3\_LandTrendR\_on\_CC
- 4) WAFForDD\_4\_Annual\_ForestDegragation\_GUI

These GEE scripts are publicly available in V2\_1 folder of the WAFForDD GEE scripts archive:

[https://code.earthengine.google.com/?accept\\_repo=users/servir\\_wa/WAforDD](https://code.earthengine.google.com/?accept_repo=users/servir_wa/WAforDD)

The fourth script was designed to be implemented as a GEE App that can be access without requiring a GEE account or the GEE Code Editor, although it can also be run from the Code Editor. The following link connects to the current version of the WAFForDD GEE app.

<https://mcwimberly.users.earthengine.app/view/wafordd21>

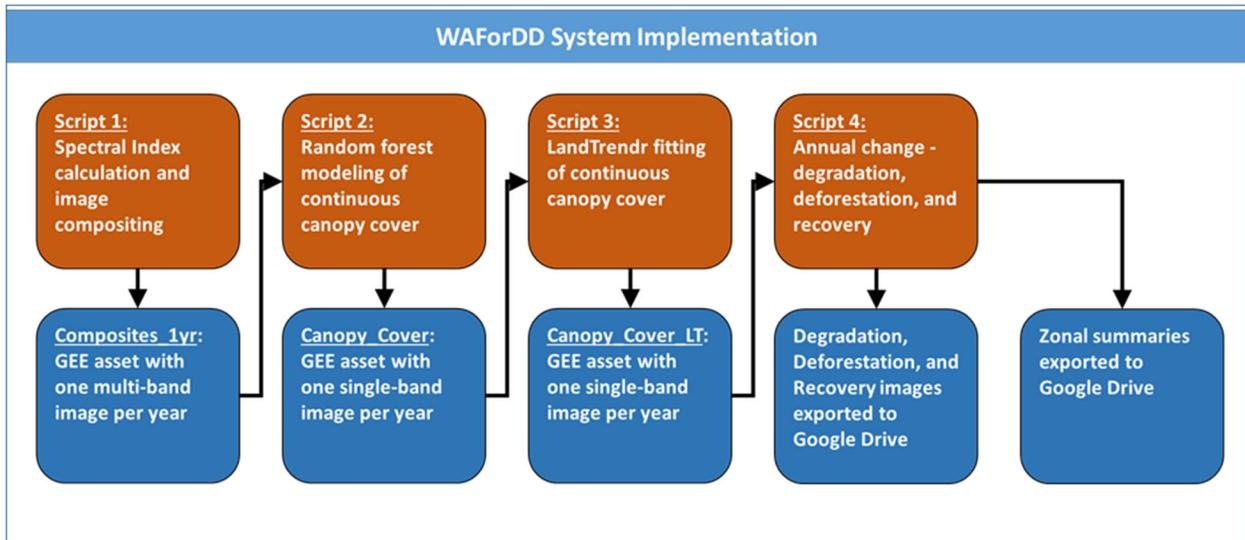


Figure 1: The four main processing steps in WAFoRDD for generating annual estimate of forest change, including forest degradation, deforestation, and recovery events.

Scripts 1-3 also require write access to the GEE assets stored in *projects/servir-wa/WAFoRDD*. All of these assets are currently configured for public, read-only access. Users who want to save new assets to this folder will need to request write access from the SERVIR West Africa hub. Alternately, users can modify the scripts to save assets in their personal user spaces. The following sections provide details about implementation of each of the four steps.

## 2.1 Step 1: Annual Image Compositing

### Script #1: *WAFoRDD\_1\_Image\_Compositing*

WAFoRDD uses annual median pixel-based composites to minimize data gaps associated with clouds, cloud shadows, and the Landsat 7 scan-line corrector failure (SLC-off) as well as the low availability of Landsat data, especially in the early years of the study period. This approach uses the median of all available values for each pixel in the image stack to generate the annual composite. In addition, a 3 x 3 pixel mean spatial filter is applied to the resulting annual composites for further noise reduction.

- 2.1.1 **Spectral data input:** WAFoRDD uses all available Landsat 7 (ETM+) and Landsat 8 (OLI) Landsat surface reflectance level-2 science products, which are generated using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) for Landsat 7 and the Landsat Surface Reflectance Code (LaSRC) for Landsat 8, to create the annual image composites. All visible, near-infrared, and shortwave-infrared bands (blue, green, red, nir, swir1, swir2) are utilized along with the Pixel Quality Assessment band (pixel\_qa). We used the method of Roy et al. (2016) to cross-calibrate Landsat 8 reflectance with Landsat 7 reflectance to correct for sensor differences and generate a consistent, harmonized dataset.
- 2.1.2 **Data Screening:** WAFoRDD employs a series of screening procedures to mask out pixels that are clouds, cloud shadows, or water. This process uses the pixel\_qa layer provided with the Landsat

surface reflectance product. The cloud screening flags in this layer were generated using the FMASK algorithm (Zhu et al. 2015). In an initial assessment, we found that many pixels impacted by clouds and haze were not flagged in the `pixel_qa` layer. Therefore, additional cloud screening is done by masking out all pixels with a reflectance value greater than 0.09 in the blue band, and by applying an additional cloud mask derived from spectral mixture. The SMA-based cloud mask was generated using the cloud endmember determined by Souza et al. (2013).

2.1.3 **Spectral data outputs:** The image compositing step produces several spectral indices:

- SMA (Spectral Mixture Analysis) fractions: Green Vegetation (GV), Non-Photosynthesizing Vegetation (NPV), Soil and Shade. These variables were used to calculate the Normalized Difference Fraction Index (NDFI, Sousa et al. 2013), and a modified version of the Normalized Difference Fraction Index (NDFI2).
- Tasseled Cap Indices (Brightness, Greenness and Wetness)
- Other vegetation indices: Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI), Normalized Burn Ratio (NBR) and Enhanced Vegetation Index (EVI).

Depending on user preference, several aspects of the image compositing can be controlled by the following parameters in the `WAFForDD_1_Image_Compositing` script:

2.1.4 **Study/Current year:** The variable `yearC` gives the option to set the particular year to generate an annual image composite for. For instance, a `yearC` value of 2001 will generate a composite for the year 2001.

2.1.5 **Composite season:** The variables `startmonth` and `endmonth` can be adjusted to generate image composites for specific period within the year, such as the dry season or the rainy season. The inputs are 2-digit integer values representing calendar months (01 – Jan., 02 – Feb., 03 – Mar., 12 – Dec, etc.). For instance, specifying `startmonth = 01`; and `endmonth = 03` will generate a composite using image collection from January to March only. Currently, the WAFForDD system uses `startmonth = 01` and `endmonth = 12` to generate composites for the entire calendar year.

2.1.6 **Composite window:** The `comp_window` variable specifies the number of years to include in each composite. The default value is set to 1 so that WAFForDD generates 1-year (annual) image composites. Based on our preliminary analyses, we determined that an annual composite period was necessary to accurately capture the timing of disturbance events. However, it is also possible to specify multi-year composite windows.

A `comp_window` of 1 generates an image composite by taking all the Landsat data within a calendar year. For instance, for a composite for the year 2001 will use all the available Landsat data from January to December of 2001.

A `comp_window` of 2 generates an image composite by taking all the Landsat data from the beginning of the prior calendar year to the end of the current calendar year. For instance, for a composite for the year 2001 will use all the available Landsat data from January 2000 to December 2001.

A *comp\_window* of 3 generates an image composite by taking all the Landsat data from the beginning of the prior calendar year to the end of the one year after the current calendar year. For instance, for a composite for the year 2001 will use all the available Landsat data from January 2000 to December 2002.

- 2.1.7 **Study area:** The variable *studyArea* defines the area of interest for which the image composite will be generated. The default setting is the forest zone of Ghana, and is defined by an imported feature collection that is stored in *project/servir-wa/WAForDD/shapefiles*.

When the script is run, spectral indices will automatically be calculated for the current year (*yearC*) within the specified study area. These calculated indices are displayed in the map view. By selecting the corresponding image name in the task panel, the composite can be computed and exported as a GEE asset to the *Composites\_1yr* image collection.

## 2.2 Random Forest Modeling of Percent Forest Canopy Cover

### Script #2: *WAForDD\_2\_CanopyCover\_\_RandomForest\_Model*

WAForDD uses a Random Forest (RF) regression algorithm to predict manually interpreted canopy cover as a function of Landsat-derived indices generated from the image compositing step. Training and validation use a large multi-temporal database of manually interpreted 2010 - 2017 high-resolution imagery from DigitalGlobe WorldView and Google Earth. Based preliminary analyses, the RF regression model was set to 500 trees with a bag fraction (proportion) of 0.67. All other settings were default. We then applied this RF model to predict percent canopy cover across the entire study area for all years from 2001 to 2019.

Visual interpretation of percent canopy cover was determined by estimating the area of a sample plot covered by overstory tree crowns. Estimates ranged from 0% to 100%, in 5% increments. Plot sizes encompassed 3 x 3 Landsat pixels (0.81 ha), distributed within the forest zone using stratified random sampling. Canopy cover estimation was aided by georeferenced field data collected in four forest reserves, field photographs obtained in several forest reserves, and a set of ground data supplied by the REDD+ secretariat of the Ghana Forestry Commission.

- 2.2.1 **Study/Current year:** As in the image compositing step, the variable *yearC* gives the option to set the particular year for which RF predictions are being made. For instance, *yearC* of 2001 will generate RF continuous canopy cover for the year 2001. *It is important to make sure that annual image composite for this particular year has already been generated from the image compositing script and that this input image exists the appropriate GEE asset location.* For instance, setting *yearC* to 2020 will not generate any output if an annual image composite for 2020 does not already exist in your GEE data archive.
- 2.2.2 **RF spectral data input:** Script #2 uses input data from annual composites of spectral variables stored in the *Composites\_1yr* image collection. The *bands* variable contains a list of the spectral

indices that will be used as predictor variables for the RF predictions. By default, WAFForDD runs using all available spectral indices as well as the raw reflectance values for the Landsat bands.

- 2.2.3 **Training data:** The feature collection *train\_pts* contains the training data from the visual interpretation, and the feature collection *val\_pts* contains a separate validation dataset. These data have been imported as GEE assets and are stored in *projects/servir-wa/WAFForDD/shapefiles*.

When the script is run, a Random Forest model is fitted using the training data and applied to predict canopy cover across the entire forest zone of Ghana. Canopy cover data for the current year (*yearC*) is displayed in the map view. By selecting the corresponding image name in the task panel, the canopy cover image can be exported as a GEE asset. Images are exported to the *Canopy\_Cover* image collection. It is also possible to export a dataset to the user's Google Drive with the validation data overlaid on the canopy cover predictions by selecting the appropriate option in the task panel.

## 2.3 LandTrendr Fitting of Random Forest-Derived Continuous Canopy Cover

### Script #3: *WAFForDD\_3\_\_LandTrendR\_on\_CC*

The Landsat-based detection of trends in disturbance and recovery (LandTrendr) algorithm consists of a set of processing steps applied to each pixel (Kennedy et al., 2010). LandTrendr employs a segmented regression approach to concisely describe noisy time series, thereby capturing the major changes in the time series while smoothing undesired random fluctuations, such as those caused by normal phenological variation, year-to-year climate fluctuations, atmospheric effects, and sun angle (Kennedy et al., 2010; Kennedy et al., 2012). Through this step, we expect the smoothed predictions to remove abnormal interannual spikes or dips while identifying long-term trends in canopy cover as well as short-term effects of canopy disturbance, thereby ensuring improved depiction of actual forest cover changes.

- 2.3.1.1 **Study/Current year:** The study year, *yearC*, has same meaning as in scripts # 1 and 2. The LandTrendr algorithm operates on all the canopy cover images in the times series. The *year\_init* and *year\_final* variables determine the period of time over which LandTrendr will be applied. Setting *yearC* to a specific year enables the LandTrendr-fitted canopy cover image for that year to be exported. Note that *yearC* can be any of the years within the study period.
- 2.3.2 **LandTrendr spectral data input:** The collection of continuous canopy cover from the RF model is the input dataset for the LandTrendr (LT) temporal segmentation algorithm. These data are stored in the *Canopy\_Cover* image collection.
- 2.3.3 **LandTrendr setup parameter:** LT has several parameters that can be adjusted in order to get better results. For details on these parameters the reader is referred to the LT foundational paper Kennedy et al., 2010 as well as the LT detailed guide at <https://emapr.github.io/LT-GEE/index.html>
- 2.3.4 **Validation data:** The feature collection *val\_pts* contains a validation dataset. These data have been imported as GEE assets and are stored in *project/servir-wa/WAFForDD/shapefiles*.

When the script is run, the LandTrendr algorithm fits a piecewise linear regression to each pixel and uses these regression functions to generate a smoothed time series of canopy cover images. When the script

is run, LandTrendr canopy cover maps for several years are displayed in the map viewer. The user can also click on the map and view time series of raw and LandTrendr-smoothed canopy cover for any pixel. The *yearC* variable determines the year of the image that can be exported. By selecting the corresponding image name in the task panel, the LandTrendr canopy cover image can be computed and exported as a GEE asset. These assets are stored in the *Canopy\_Cover\_LT* image collection. It is also possible to export a validation dataset with the validation data overlaid on the LandTrendr canopy cover predictions by selecting the option in the task panel.

## 2.4 Annual Estimates of Forest Degradation

### **Script #4: *WAFORDD\_4\_Annual\_ForestDegradation***

The final step involves applying change detection methods to detect degradation and other changes, including deforestation and recovery. The first step in the forest degradation calculations involves classifying forests into three classes based on canopy cover: Closed Forest (> 60%), Open Forest (15-60%), and Low Tree Cover (< 15%). A 3 x 3 majority filter was applied to the canopy cover class layer to eliminate small patches consisting of only one or a few pixels. We then overlaid each year of classified canopy cover data on data from the preceding year and assigned change classes based on the following rules:

Degradation: Change from Closed to Open Forest

Open Forest Loss: Change from Open Forest to Low Tree Cover

Closed Forest Loss: Change from Closed Forest to Low Tree Cover

Open Forest Recovery: Change from Low Tree Cover to Open Forest

Closed Forest Recovery: Change from Open Forest to Closed Forest

An additional temporal filter was applied to compare canopy cover classes in the preceding two years, the current, year, and the subsequent year applied to remove illogical change sequences. In general, when a disturbance (degradation or forest loss) occurs, we assume that it will take a minimum of two years to recover from the disturbance. For example, if degradation (change from closed forest to open forest) is immediately followed by recovery to closed forest the next year, then the degradation event is assumed not to have occurred.

This script differs from the previous three in that user interaction with the script is entirely through the graphical user interface. Script #4 is implemented as an Earth Engine App that provides several options in the pane on the left side of the map.

<https://mcwimberly.users.earthengine.app/view/wafordd21>

Two links at the top of the map allow the user to either (1) download a zonal summary of the total area of each change class masked to include only the forest reserves and summarized by ecological zone, or (2) download a zonal summary of the total area of each change class summary by individual forest reserve. These summaries are calculated for all years in the canopy cover time series and the resulting downloads are saved locally by the users' web browser. A slider bar allows the user to select the year for which data will be displayed in the map viewer. Layers displayed include the classified canopy cover dataset as well as areas of degradation, open forest loss, closed forest loss, open forest recovery, and

closed forest recovery for the selected year. Maps of the forest canopy classes from the preceding and subsequent years can also be displayed. When the user clicks anywhere in the map, time series plots appear in the left hand pane that show the raw canopy cover estimates along with the LandTrendr fits.

Because of the limitations of Earth Engine Apps, it is not possible to directly download geospatial datasets for the entire study area. Instead, a link is provided to code for the script that can be run in the GEE code editor. When running in the code editor, the slider can be used to select the year or interest and a data export task will appear in the Tasks panel. This task can then be run to download geospatial data to the user's Google Drive.

### 3. Steps for Updating WAFForDD

The WAFForDD scripts were designed so that they can be used to efficiently update the degradation estimates in future year (2020, 2021, 2022...) once the Landsat data become available. For example, steps for updating WAFForDD at the end of 2020 are outlined below:

#### 3.1 Generate a new composite image of spectral variables

Open the *WAFForDD\_1\_Image\_Compositing* script in the GEE code editor and change the *YearC* variable to the new year (in this example, 2020). Run the script and check the map layers to make sure that the spectral indices look reasonable. Then go to the Tasks panel and click on the RUN button next to the task (*y2020\_1yr\_Median\_Image\_Composite*). The new image composite will be generated and written to the *Composites\_1yr* asset.

#### 3.2 Generate a new canopy cover image

After the new composite image of spectral variables has been generated and saved as a GEE asset, open the *WAFForDD\_2\_CanopyCover\_RandomForest\_Model* script in the GEE code editor and change the *YearC* variable to the new year (in this example, 2020). Run the script and check the map layer to make sure that the canopy cover map looks reasonable. Then go to the Tasks panel and click the RUN button next to the task (*CanopyCover\_RF\_2020*). The new canopy cover image will be generated and written to the *Canopy\_Cover* asset. If you wish, you can also click on RUN button next to the *Validation\_CanCov* task to download a table of measured canopy cover values from the validation dataset along with the corresponding predictions from the Random Forest model. Note that this task generates the validation dataset for all available years of validation data – you only need to run it one time, not once per year.

#### 3.3 Generate a new time series of LandTrendr-smoothed canopy cover images

After the new canopy cover image has been generated and saved as a GEE asset, the next step is to run the LandTrendr algorithm on the time series of canopy cover data. To generate the LandTrendr smoothed images it is necessary to run the entire time series, including the new year. The LandTrendr algorithm uses information about canopy cover over the entire time series to improve the estimation of canopy cover in any given year, so the estimates for all-years must be re-run every time a new year is added.

The first-step in re-running the LandTrendr time series is removing all of the existing images from the *Canopy\_Cover\_LT* ImageCollection asset. Instead of just deleting the old images, we recommend creating a new ImageCollection asset (e.g., *Canopy\_Cover\_LT\_2019*) and copying the images from

*Canopy\_Cover\_LT* to this new asset. To copy an image, open the new ImageCollection asset in the GEE code editor and click on “+ Copy an image to this collection”. Once the images have been copied they can be deleted from *Canopy\_Cover\_LT*.

Once the *Canopy\_Cover\_LT* asset is empty, it will be necessary to run the *WAFForDD\_3\_LandTrendR\_on\_CC* script once for every year in the time series. Before making any runs, be sure to change the *year\_final* variable to the last year in the time series (2020 in this example).

It will be necessary to complete the following steps for each year in the time series (in this example the years 2001-2020):

1. Change the *yearC* variable to the current year
2. Click the Run button to execute the script.
3. Go to the task pane and click on the RUN button next to the appropriate task (for example, *Canopycover\_LT\_2001*).

After completing these steps for every there will be a large number of tasks running in GEE. It will likely take at least several hours for all of these tasks to complete. At this point, it is safe to close the GEE code editor – the tasks will continue to run and you can check on their progress the next time you log in. Each task generates a LandTrendr-smoothed canopy cover image for the select year and outputs it to the *Canopy\_Cover\_LT* asset.

When the script is run, the LandTrendr-smoothed canopy cover predictions are displayed for selected years in the map layout, and it is possible to click on the map and display time-series charts of canopy cover before and after applying the LandTrendr algorithm.

### 3.4 Visualize and summarize forest landscape change

The *WAFForDD\_4\_Annual\_ForestDegradation\_GUI* script was designed to be implemented as an Earth Engine App.

<https://mcwimberly.users.earthengine.app/view/wafordd21>

Users can simply click on the link, select a year of interest with the slider bar, and view images of current forest conditions and changes from the previous year in the map layout. It is also possible to click in the map and view time series charts of canopy cover for the selected pixel. Two charts are produced: one with the original canopy cover predictions generated by the Random Forest model, and the second with the canopy cover predictions after they have been processed with the LandTrendr algorithm.

Two types of summary tables can be produced and downloaded. The first summarizes the total area of each change type (Degradation, Open Forest Loss, Closed Forest Loss, Open Forest Recovery, and Close Forest Recovery) occurring within the forest reserve boundaries stratified by ecological zone. These summaries can be used to develop estimates of emissions and carbon sequestration by multiplying the areas of these changes by the appropriate emission factors. The second summary provides the total area of each change type within each forest reserve.

The change estimates are calculated “on-the-fly” as needed to generate the maps and charts. It is also possible to download these data as images of each change type (Degradation, Open Forest Loss, Closed

Forest Loss, Open Forest Recovery, and Close Forest Recovery) for each year. However, it is not possible to download these large datasets using the Earth Engine App. To download the images, the user will need to run the WAFForDD\_4\_AnnualForestDegradation script in the GEE code editor (link provided in the lower left-hand pane of the app). When the year slider is moved to a new year, a set of five download tasks (one for each change type for that year) are automatically generated in the Tasks tab. The user can then click the RUN button next to one or more of these tasks to generate the image and download it to their Google Drive.

## References

- Benefoh, D. T., Villamor, G. B., van Noordwijk, M., Borgemeister, C., Asante, W. A., & Asubonteng, K. O. (2018). Assessing land-use typologies and change intensities in a structurally complex Ghanaian cocoa landscape. *Applied Geography*, 99, 109-119.
- Dwomoh, F. K., Wimberly, M. C., Cochrane, M. A., & Numata, I. (2019). Forest degradation promotes fire during drought in moist tropical forests of Ghana. *Forest Ecology and Management*, 440, 158-168.
- Hawthorne, W. D., Sheil, D., Agyeman, V. K., Juam, M. A., & Marshall, C. A. M. (2012). Logging scars in Ghanaian high forest: towards improved models for sustainable production. *Forest Ecology and Management*, 271, 27-36.
- Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sensing of Environment*, 114(12), 2897-2910
- Kennedy, R. E., Yang, Z., Cohen, W. B., Pfaff, E., Braaten, J., & Nelson, P. (2012). Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan. *Remote Sensing of Environment*, 122, 117-133.
- Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote sensing of Environment*, 185, 57-70.
- Souza, C.M., Siqueira, J.V., Sales, M.H., Fonseca, A.V., Ribeiro, J.G., Numata, I., Cochrane, M.A., Barber, C.P., Roberts, D.A., Barlow, J., 2013. Ten-year Landsat classification of deforestation and forest degradation in the Brazilian Amazon. *Remote Sensing* 5, 5493-5513.
- Zhu, Z., Wang, S., & Woodcock, C.E. (2015). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, 159, 269-277