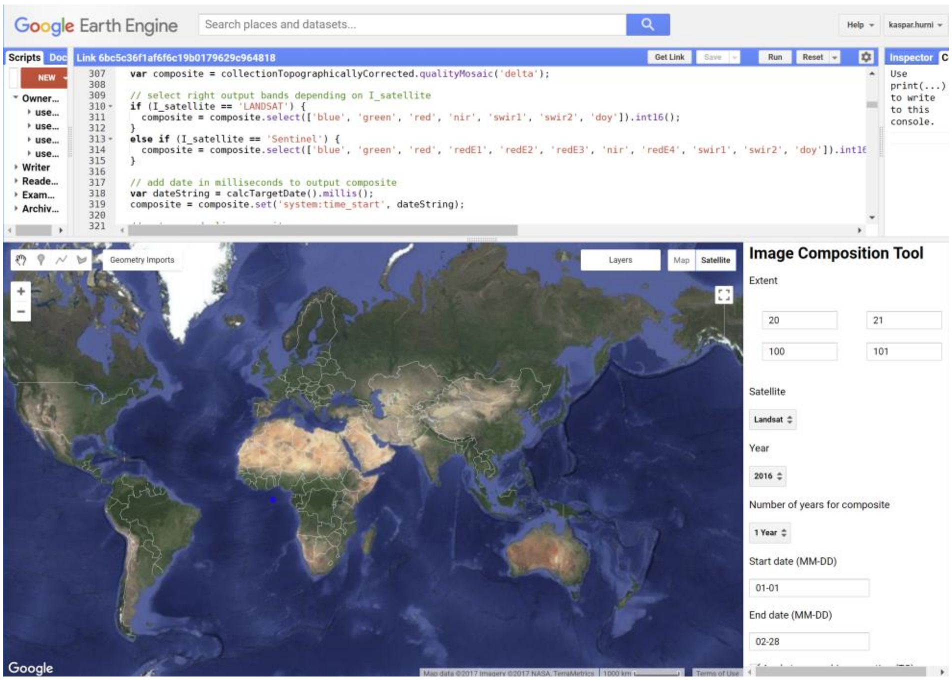
Google Earth Engine Image Pre-processing Tool: Examples



Kaspar Hurni, Andreas Heinimann, and Lukas Würsch

Centre for Development and Environment (CDE)
University of Bern

2017

Example 1: Extraction of an individual image composite

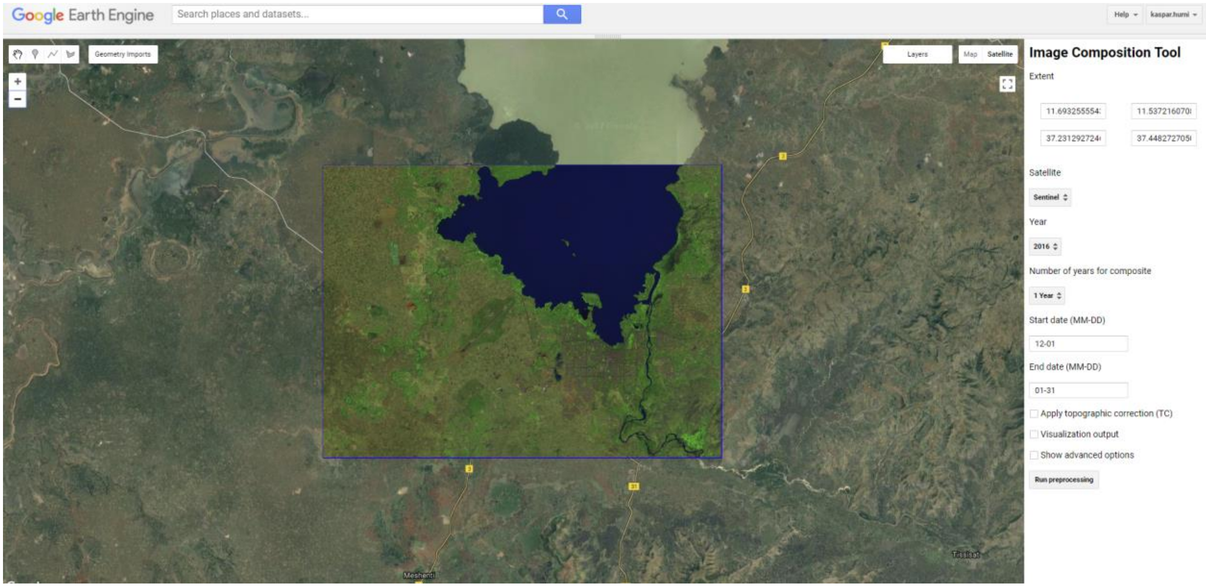
This example touches on the most common use of remote sensing data, i.e. the selection of an individual image or image composite for the visualization and/or classification of the land cover status. When working with individual images, the image selection usually follows two steps:

1) Considerations on the land cover types we would like to visualize or classify, i.e. during which season (e.g. dry or wet) will the distinction of the different land cover types work best?
2) Considerations on cloud cover, i.e. we will need an individual image that minimizes cloud cover so that we get as many valid pixels as possible within our study area.

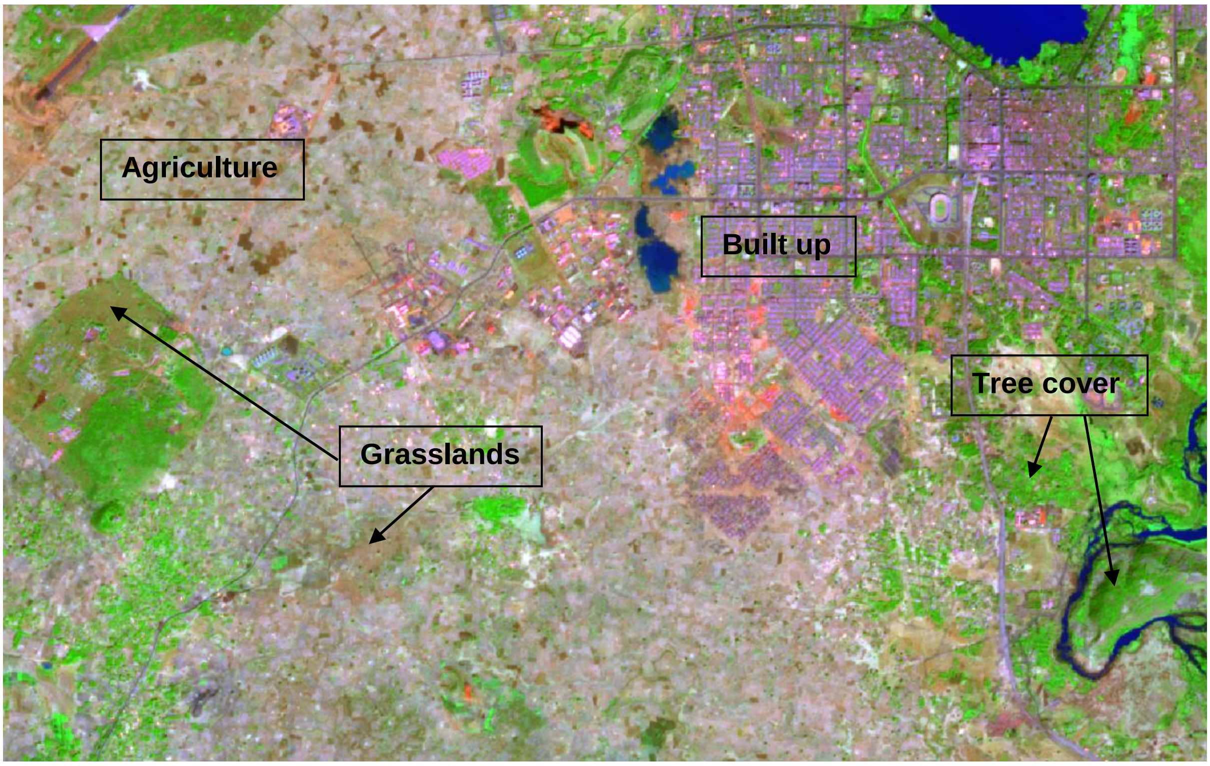
In many cases this approach involves trade-offs. While there is a certain time of the year when the distinction of the targeted land cover types works best, cloud-free images may not be available. Researchers thus have to deviate from the 'optimal season' to obtain an image with minimal cloud cover over the study region, which can lead to lower classification accuracies (e.g. certain land cover types look similar during the time of the image acquisition).

When working with an automatized image composition, we only have to pay attention to point 1 and can largely ignore point 2, because the image composition takes care of cloud cover, no data pixels, and other noise in the RS data. This also requires that we 'search' for RS images in a different way. We do not have to browse through images to get an overview of the 'best' data, but we can mainly focus on land use characteristics and the phenology of the land cover types we would like to classify. Such information is rather available in the literature – e.g. local land use practices, information on crops / cropping cycles, climate diagrams. Alternatively, we may also use historical high resolution data in Google Earth to identify seasonal behaviour of land cover types and land use practices.

In this example we focus on a small area around Bahir Dar / Lake Tana in Ethiopia and we would like to visualize (or classify) general land cover types, i.e. water, forests, settlements and urban areas, grasslands, and agriculture. The distinction of these land cover types will work best during the driest months, i.e. December and January in Bahir Dar (https://www.meteoblue.com/en/weather/forecast/modelclimate/bahir-dar_ethiopia_342884). During this time, agricultural plots are mostly bare or may show some crop residues, grasslands are moderately green, and forests are green. Settlements in Ethiopia often show vegetative cover (homesteads, eucalyptus for firewood) and unfortunately a classification of built up areas is difficult at any time of the year. Using this information we performed the image composition with the GEE image pre-processing tool. We used Sentinel data and defined a period from December 2016 – January 2017 to select data for the image composition. The figure below shows the user interface, the parameters we entered, and the resulting image composite.



This composite allows us to distinguish the main land cover types:



This composite can now be used for a variety of purposes – e.g. visualization, unsupervised classification, supervised classification. Any application that can be performed with individual images can be done with image composites, but we can expect better results because we have much more flexibility in the selection of the RS data.

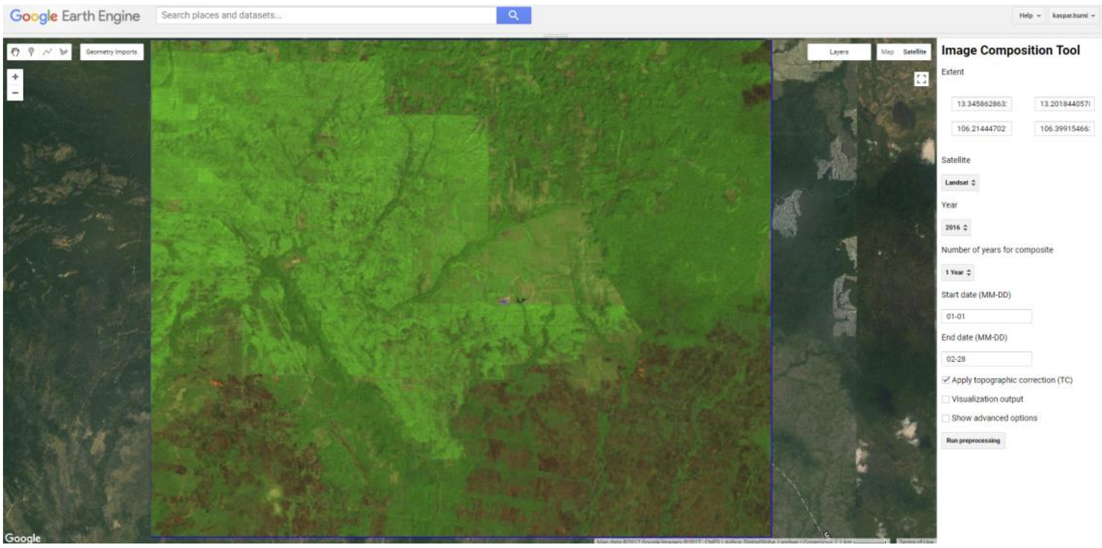
Example 2: Extraction of multiple image composites within a year

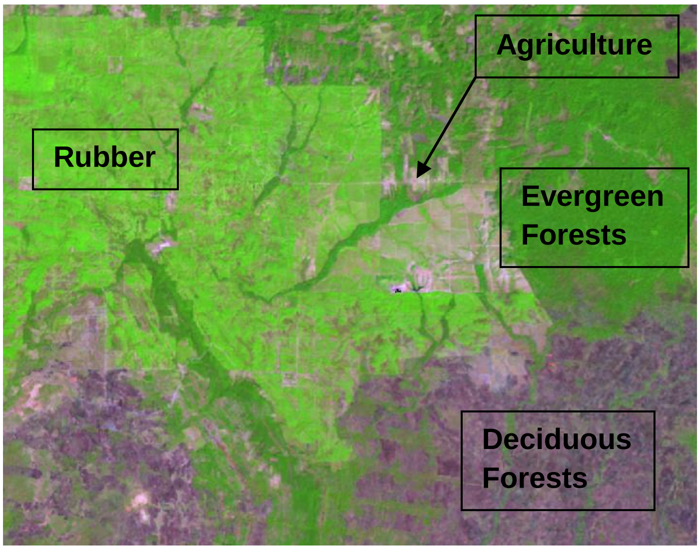
Example 1 showed that an individual image composite, built over the appropriate season, allows us to distinguish different land cover types fairly well. However, when trying to perform a finer classification that e.g. also distinguishes different tree or crop types, an individual image composite is often not sufficient. Several reasons exist for this, but mostly the multispectral information of a single image composite does not allow for such a fine classification. This can be overcome by using multiple image composites from different seasons within a year, because the time-series of image composites represents the phenology of different land cover types. E.g. different forest types may look the same at a certain point of the year, but by looking at the dynamics within a year, we are able to distinguish evergreen from deciduous forests or even distinguish different tree types based on the timing of their defoliation. For annual crops, this is similar, but we also need to consider the land use practices, e.g. the timing of plowing, sowing, and harvesting.

In this example, we focus on the detection of rubber plantations in Cambodia. Rubber plantations occur in areas with evergreen and deciduous forests, permanent annual agriculture and rotational agriculture, etc., and to reliably map rubber (e.g. distinguishing it from different forest types) we need to consider phenology by using multiple seasonal image composites. As shown in example 1 we need to obtain information on the phenology of the different land cover types we would like to classify:

- Rubber shows a short period (2-4 weeks) of defoliation around February, and around March or April it shows a fast and intense foliation.
- Deciduous forests defoliate earlier than rubber (Dec/Jan) and foliate slowly with the onset of the rainy season around May.
- Evergreen forests only show very moderate changes in greenness within a year
- Plots with permanent annual agriculture are bare around Nov/Dec when forests still show high greenness

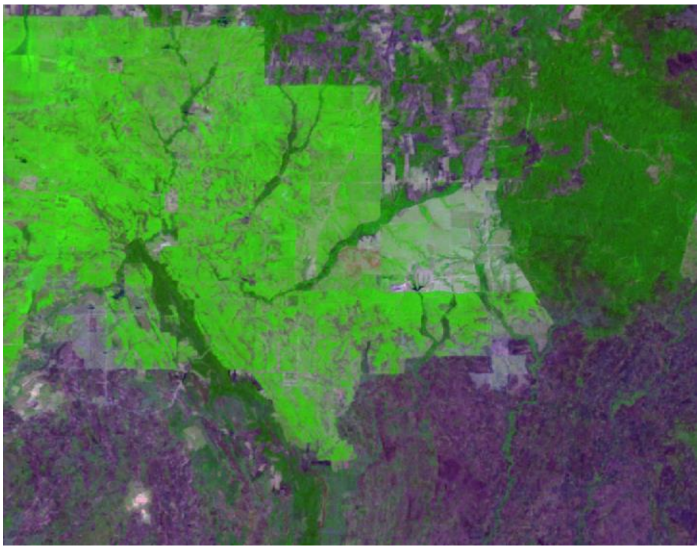
By creating a time-series of three composites (Jan & Feb, Mar & Apr, and Nov & Dec) it should thus be possible to represent the phenology of these land cover types. We may not be able to capture the full defoliation of rubber, because the period is so short, but we can capture its foliation, which helps distinguishing it from the forest types. We worked with Landsat and created the three seasonal image composites in the GEE image pre-processing tool (Jan-Feb composite is shown below).





January / Feburary:

- Foliation of rubber started
- Evergreen forests foliated
- Deciduous forests defoliated
- Annual agriculture bare



March / April:

- Rubber very green
- Evergreen forests foliated
- Deciduous forests defoliated
- Annual agriculture bare

November / December:

- Rubber and forests are foliated
- Annual agriculture bare

As shown with this example, the use of multiple seasonal image composites allows for a more detailed classification. We suggest stacking all bands of the time-series of composites into one single image to perform the classification (i.e. classifying all composites together and not performing three individual classifications).

Example 3: Extraction of yearly image composites over multiple years

Multiple image composites as shown in example 2 can also be used to perform a classification of land cover change. In such a case we build yearly image composites for a period of multiple years. With such a time-series we can then track land use and land cover conversions on an annual basis. This can be helpful when e.g. classifying the expansion of large scale land deals and in example 3 we focus on the mapping of the expansion of soy bean plantations in Campo Gallo, Argentina. To do so we use Landsat data, because Sentinel images are only available staring 2015.

Similar to examples 1 and 2 we first consider the land cover phenology and the climate of the study area to define the most suitable periods for the yearly image composition. This region of Argentina shows low precipitation from June-August (https://en.climate-data.org/region/152/) and soybean harvest occurs from April-June (https://www.thebalance.com/soybean-planting-and-harvest-seasons-809258). By creating yearly image composites for the period from June-August, we thus maximize the difference between bare soybean plots and natural vegetation, which allows a reliable mapping of the expansion of soybean plantations. Additionally, the composition period covers the dry season and we do not have to deal with persistent cloud cover, which can result in data gaps even in the image composites. We thus created 17 yearly image composites (2000-2016) and each composite covers the period from June-August. Below we show examples for the years 2000, 2010, and 2016, represented as false color images (vegetation is red, bare areas beige/white).

Image composite example 2000:

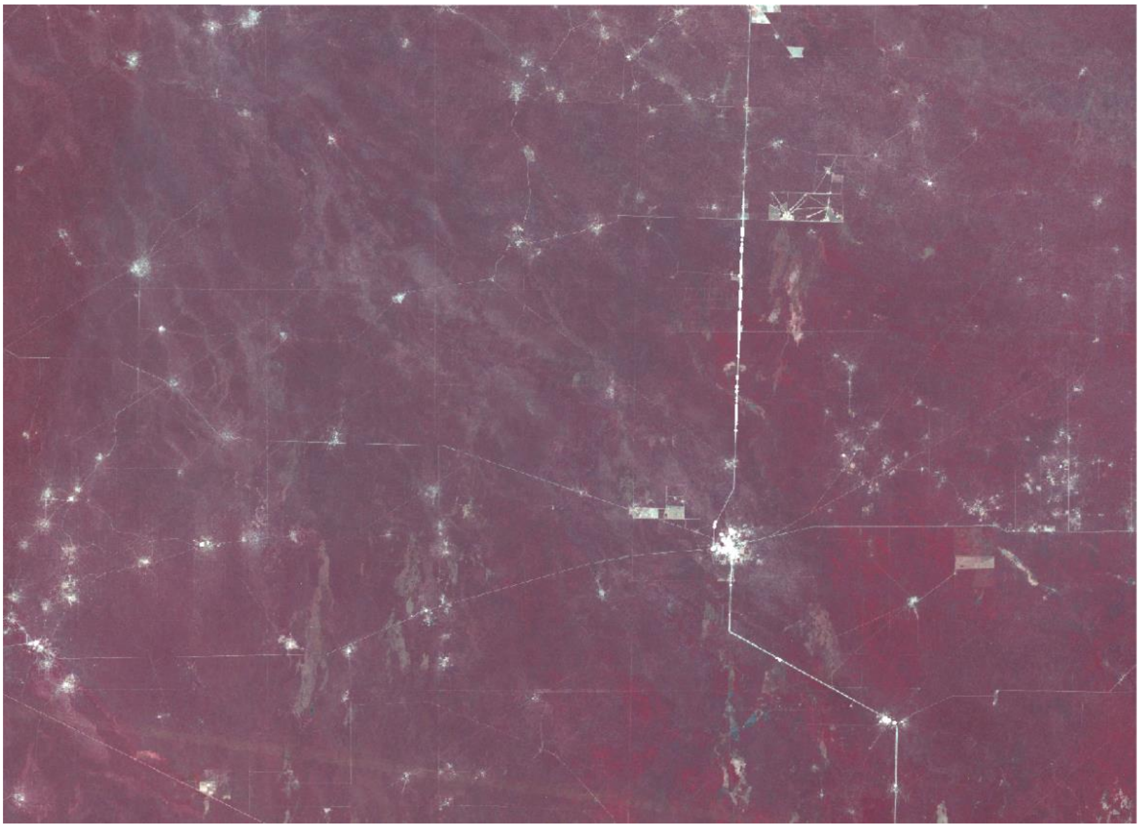


Image composite example 2010:

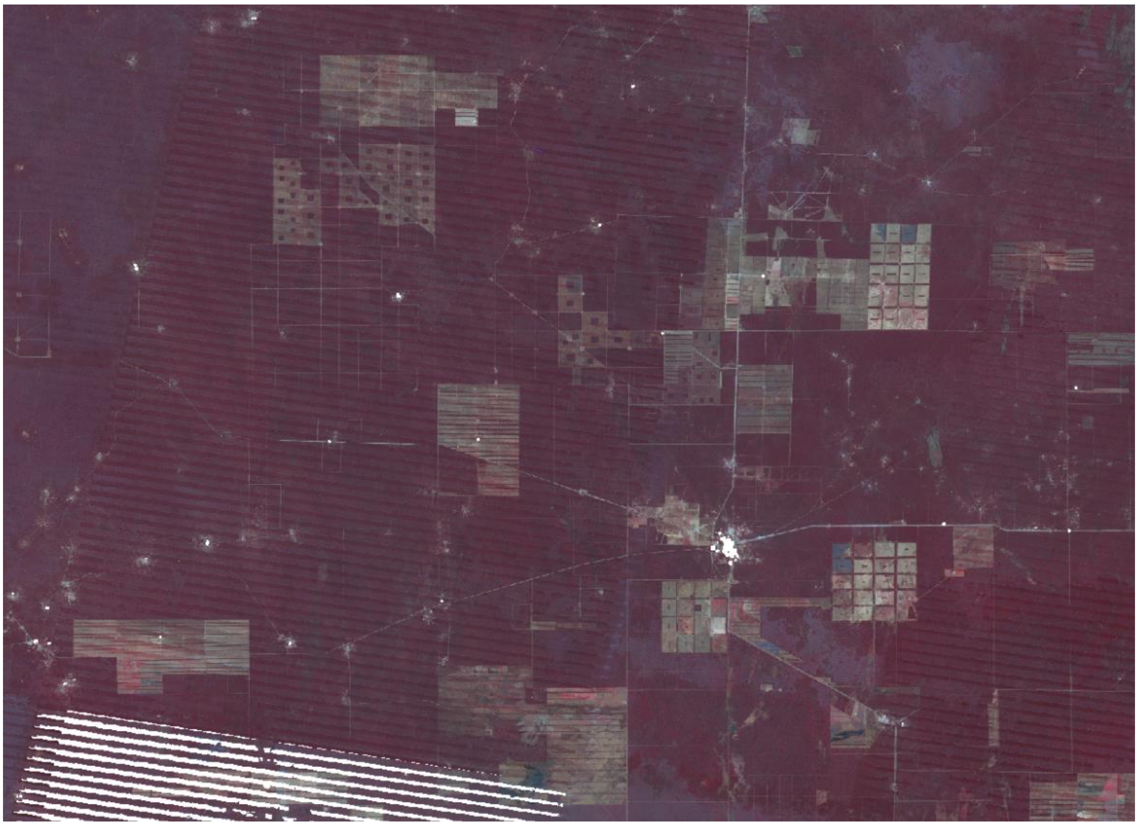
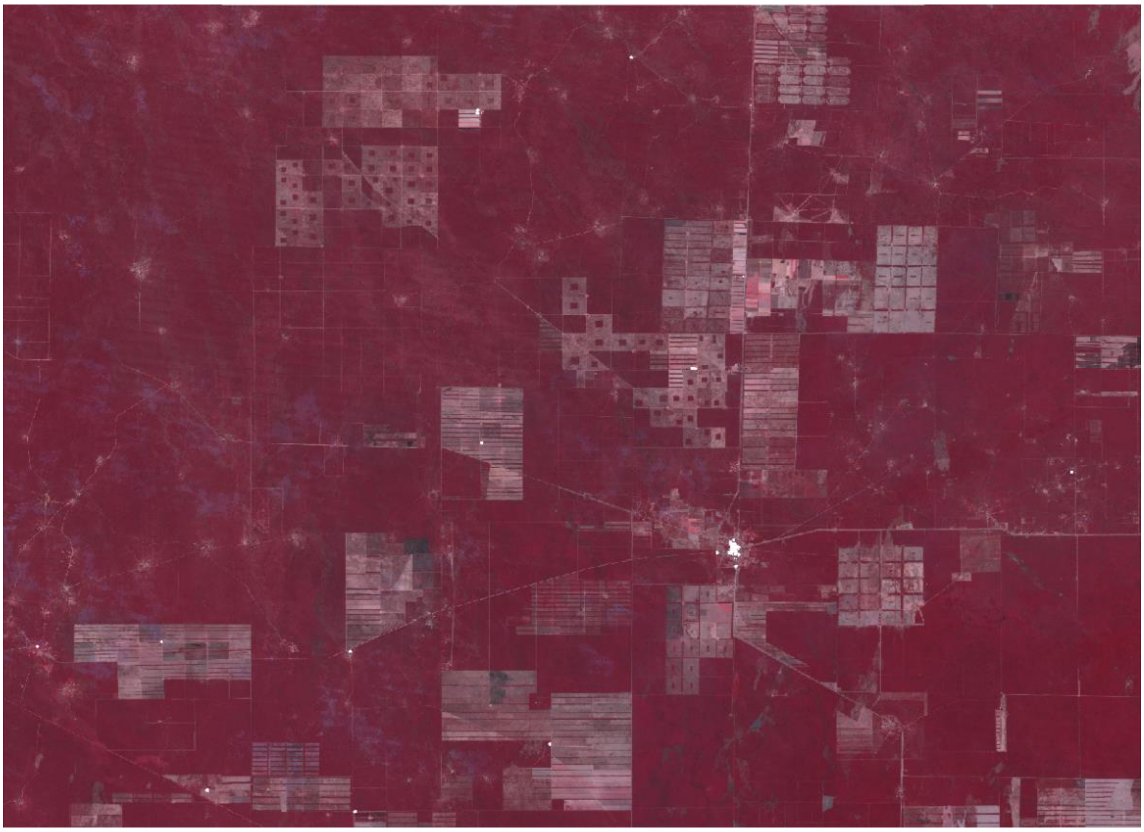


Image composite example 2016



This time series shows how plantations expanded around the Campo Gallo area in the last 17 years. As in example 2, this data can be used for a classification, in this case we would focus on land cover change. Again, we suggest stacking all the individual image composites into one image for classification (i.e. classifying all composites together and not classifying each of the 17 composites separately). The example below shows such a classification approach, but please note that there are many other ways to obtain information on land cover change!

Classification example: Composite change detection using yearly Landsat image composites

A) Definition of classification scheme and collection of classification samples

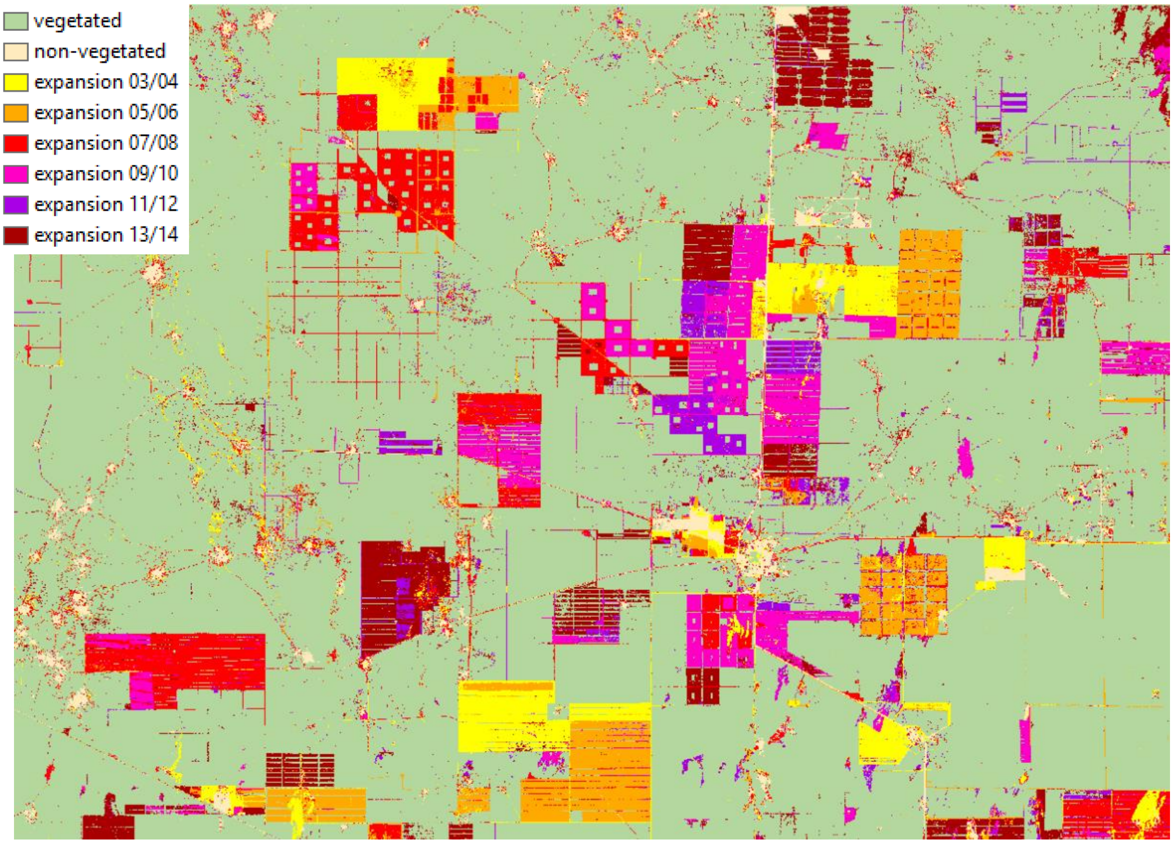
Land cover (change)	**ID**
Vegetated	1
Sparsely-vegetated/agriculture/bare/built up	2
Expansion of agriculture in 2003	13
Expansion of agriculture in 2004	14
Expansion of agriculture in 2005	15
Expansion of agriculture in 2006	16
Expansion of agriculture in 2007	17
Expansion of agriculture in 2008	18
Expansion of agriculture in 2009	19
Expansion of agriculture in 2010	20
Expansion of agriculture in 2011	21
Expansion of agriculture in 2012	22
Expansion of agriculture in 2013	23
Expansion of agriculture in 2014	24

- Creation of 100 random points for the study area
- Labelling of **land cover change** for the 100 points
- Assess distribution of points across classes (number of points for each class – are there classes with a limited number of points?)
- Targeted addition of samples by adding points for classes with only few points. **Classes that represent land cover dynamics and not just a status usually require more points**

After the targeted addition of points for the classes with a limited number of points, we obtained a total of 367 points for the classification. We then divided the samples into training points (75%) and verification points (25%).

B) Classification of land cover changes using the random forest classifier

We perform a composite change detection and thus we stacked the 17 yearly image composites into one image and classified this data using the land cover change samples. We trained the random forest classifier using our training data (75% of the samples) and then obtained the classification shown below. The accuracy assessment was performed using the verification points (25% of the samples) and we obtained an overall accuracy of 95.7%. Usually, such a composite change detection performs better than e.g. performing 17 yearly classifications followed by a post-classification comparison. Furthermore, we know the accuracy of each land cover change class; such information is not available when performing a post-classification comparison. Class-wise accuracies are shown in the table below the map, all the classes show user and producer accuracies > 75%.



Accuracies:

Land cover (change)	ID	User accuracy (%)	Producer accuracy (%)
Vegetated	1	83,3	100
Sparsely-vegetated/agriculture/bare/built up	2	100	75
Expansion of agriculture in 2003	13	75	100
Expansion of agriculture in 2004	14	100	85,7
Expansion of agriculture in 2005	15	100	100
Expansion of agriculture in 2006	16	100	100
Expansion of agriculture in 2007	17	100	85,7
Expansion of agriculture in 2008	18	100	100
Expansion of agriculture in 2009	19	100	100
Expansion of agriculture in 2010	20	100	100
Expansion of agriculture in 2011	21	100	100
Expansion of agriculture in 2012	22	100	80
Expansion of agriculture in 2013	23	100	100
Expansion of agriculture in 2014	24	100	100

Conclusions

These use cases of the GEE image pre-processing tool are just three examples of a multitude of classification approaches and methods. E.g. the approaches presented under example 2 and example 3 can also be combined to perform a multi-date composite change detection, i.e. using multiple image composites within the year over multiple years. Such an approach has been applied in this study http://www.mdpi.com/2072-4292/9/4/320, but without image composites (using composites would probably improve mapping accuracies). However, many other options exist to map land cover and land use change, and e.g. trend detection algorithms can work very well when using a time-series of yearly composites.

Similar to the selection of images / image composition periods, also the classification approaches need to be tailored to the study area and focus. There is no single ‘best’ solution, but in our experience classifications accuracies improve when considering land cover phenology during the selection of images. CDE’s GEE image pre-processing tool provides us with such possibilities, because by performing an image composition, the selection of the RS data is less guided by cloud cover. Focus can thus be laid on selecting image composite periods that correspond with phenology and/or the timing of specific land use practices.