Soybean Phenotyping for Rapid Variety Breeding in Zambia

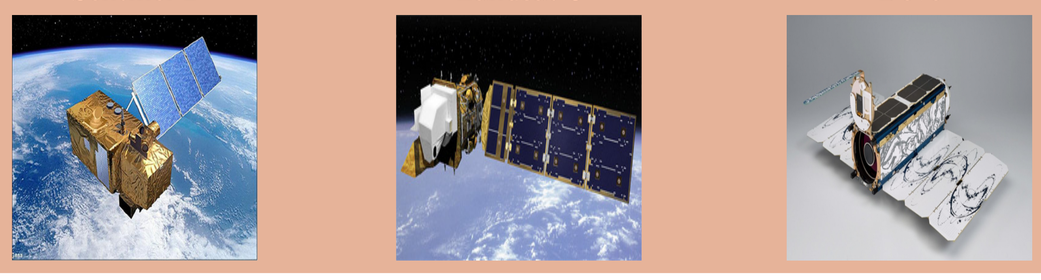
Emmanuel M. Ngonga, Katherine Denby, Godfree Chigeza, Joseph Fennell

University of York, Department of Biology, CNAP, emn519@york.ac.uk

Introduction

Soybean (*Glycine max* (L.) Merrill) is a very important crop in Zambia, grown mainly for its protein, oil and nutraceutical contents[1]. Using Satellite imagery and Python we are observing and analysing hundreds of Soybean fields in Zambia. Since this region is mostly cloudy during the growing season we chose to combine Planet Inc. Sentinel-2 and Landsat-8 satellite images to improve the following resolutions:

Sentinel-2 | Landsat-8 | Dove



Chlorophyll reflects most of the near infrared radiation incident on it, we use the amount of near infrared radiation reflected by a plant to measure above ground biomass levels [2]. This is done using the NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index), calculated as shown below

NDVI Formula

$$NDVI = \frac{nir - red}{nir + red} \tag{1}$$

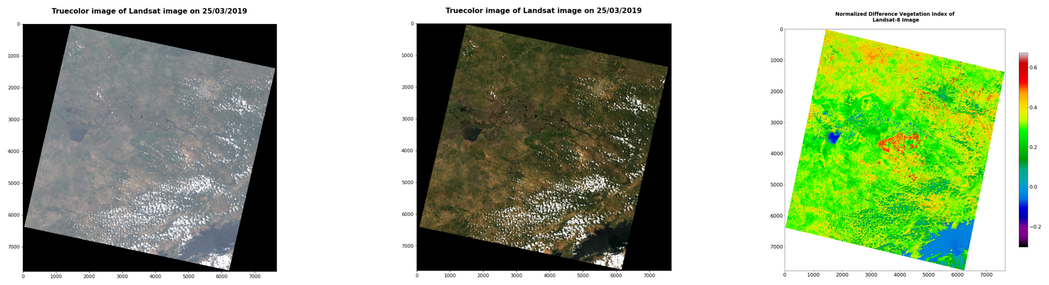
$$EVI = \frac{nir - red}{nir + 6red + 7.5blue + 1} \tag{2}$$

Where the red band is represented as 'red' blue band is 'blue ' and the Near-infrared band is 'nir'. Our main aim is to find the correlation between yield with EVI and NDVI across different varieties of Soybean grown in Zambia. This will be used to improve soybean breeding

Image Acquisition and field Isolation

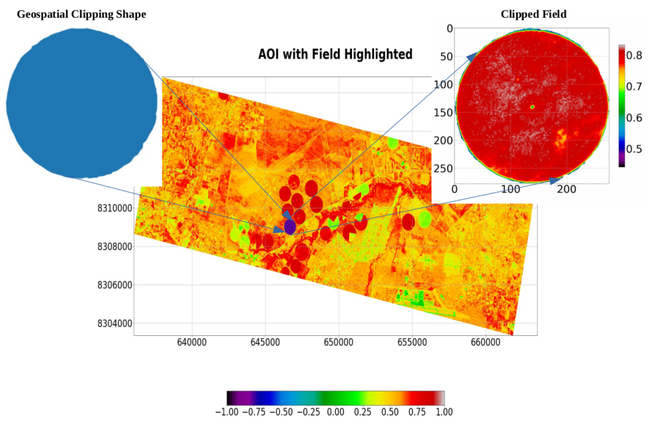
After downloading a Satellite image containing the Soybean field we are interested in, it was corrected to do away with atmospheric effects.

Top of atmosphere | **Bottom of atmosphere** | **NDVI image**



- After radiometric and atmospheric correction of the rasters the soybean field was clipped using a shapfile made in QGIS (Quantum Geographic Information System.

Field Clipping



The Colour scale of the Field depends on the Chlorophyll levels in the crop's biomass.

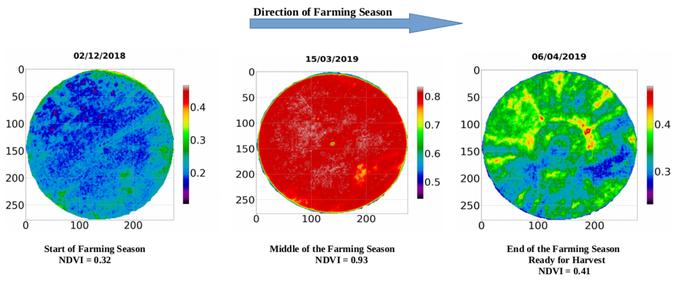
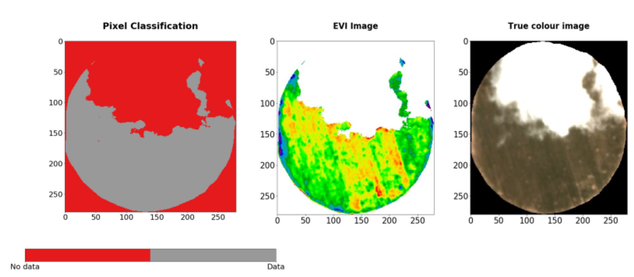


Image Analysis

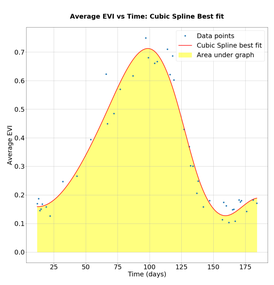
Sometimes the image is partially covered by clouds, this prompts for cloud and cloud shadow masking. We used Machine Learing to mask unwanted pixels.



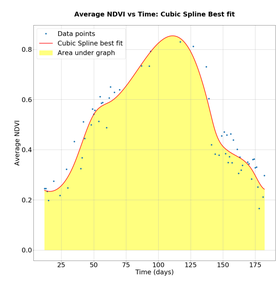
Mathematical Modelling of EVI and NDVI time-series profiles

A plot of the vegetation index time-series profiles was done using Cubic spline interpolation and Gaussian Process Modelling;

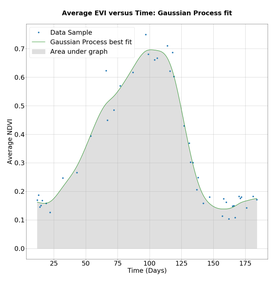
1. EVI Cubic spline



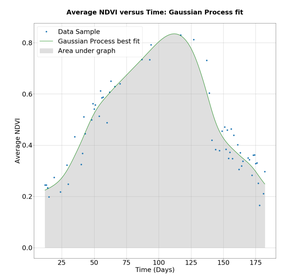
2. NDVI Cubic spline



3. EVI Gaussian Process

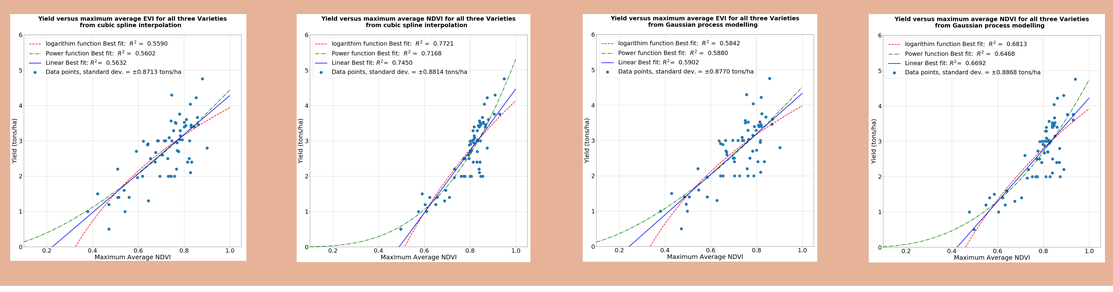


4. NDVI Gaussian



The maximum average NDVI and EVI of the time series profiles showed significat correlation with resulting yield. The yield versus maximum average NDVI and EVI were plotted using linear, logarithm and power functions.

EVI cubic spline | NDVI cubic spline | EVI Gaussian | NDVI Gaussian



Discussion

The main sources of error in our analysis are misclassifications by our cloud and shadow masking algorithm. This could be improved by training our algorithm with many more scenarios to allow it to classify surface reflectances more accurately

Conclusions and Recommendations

- There is a significat linear, logarithmic and power correlation of EVI and NDVI with the resulting yield of a soybean variety.
- Using satellites to observe soybean varieties is feasible and can be introduced into the breeding process as the spatial resolution of earth observation satellites increases.

References

[1] Dalia Mohamedkheir et al. History, current status, and prospects of soybean production and research in sub-saharan africa. 2018.

[2] Dora Krezhova et al. Spectral remote sensing of the responses of soybean plants to environmental stresses. In Dora Krezhova, editor, *Soybean*, chapter 11. IntechOpen, Rijeka, 2011.