pyveg: A Python package for analysing the time evolution of patterned vegetation using Google Earth Engine

Nick Barlow[1], **Chris Boulton**[2], **Camila Rangel Smith**[1], and **Samuel Van Stroud**[1, 3]

1 The Alan Turing Institute **2** University of Exeter **3** University College London

Introduction

Periodic vegetation patterns (PVP) arise from the interplay between forces that drive the growth and mortality of plants. Inter-plant competition for resources, in particular water, can lead to the formation of PVP. Arid and semi-arid ecosystems may be under threat due to changing precipitation dynamics driven by macroscopic changes in climate. These regions display some noteable examples of PVP, for example the "tiger bush" patterns found in West Africa.

The morphology of the periodic pattern has been suggested to be linked to the resilience of the ecosystem (Mander et al., 2017; Trichon, Hiernaux, Walcker, & Mougin, 2018). Using remote sensing techniques, vegetation patterns in these regions can be studied, and an analysis of the resilience of the ecosystem can be performed.

The pyveg package implements functionality to download and process data from Google Earth Engine (GEE), and to subsequently perform a resilience analysis on the aquired data. PVP images are quantified using network centrality metrics. The results of the analysis can be used to search for typical early warning signals of an ecological collapse (Dakos et al., 2008). Google Earth Engine Editor scripts are also provided to help researchers discover locations of ecosystems which may be in decline.

pyveg is being developed as part of a research project looking for evidence of early warning signals of ecosystem collapse using remote sensing data. pyveg allows such research to be carried out at scale, and hence can be an important tool in understanding changing arid and semi-arid ecosystem dynamics. An evolving list of PVP locations, obtained through both literature and manual searches, is included in the package at `pyveg/coordinates.py`. The structure of the package is outlined in Figure 1, and is discussed in more detail in the following sections.
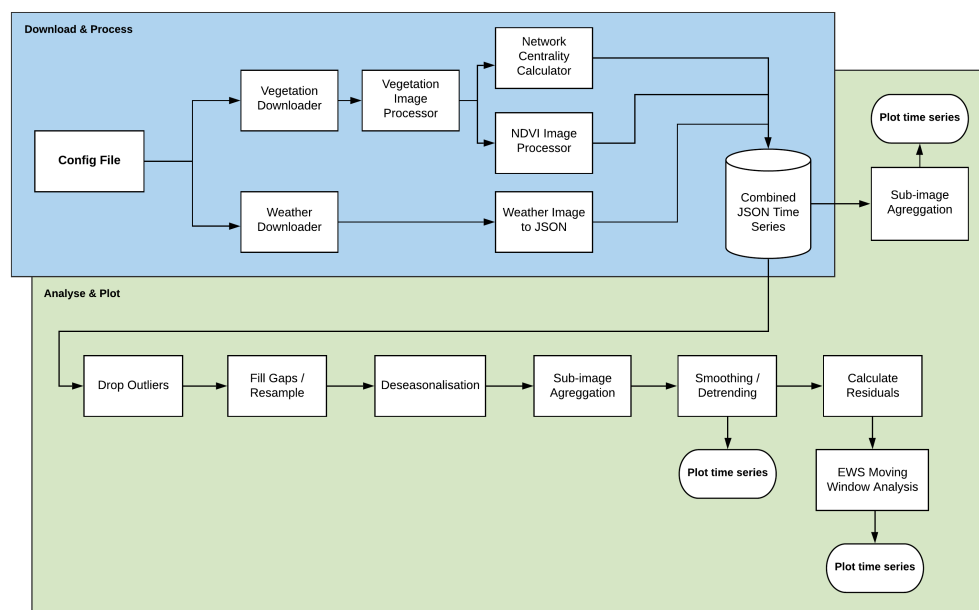
Figure 1: pyveg program flow.

Downloading data from Google Earth Engine

In order to interact with the GEE API, the user must sign up to GEE and obtain an API key, which is linked to a Google account. Upon downloading data using pyveg for the first time, the user will be prompted to enter their API key to authenticate GEE. The run_pyveg_pipeline command initiates the downloading of time series data at a single coordinate location. The job is configured using a configuration file specified by the --config_file argument.

Within the configuration file, the user can specify the following: coordinates of the download location, start and end dates of the time series, frequency with which to sample, choice of GEE collections to download from (currently vegetation and precipitation collections are supported).

pyveg will then form a series of date ranges, and query GEE for the relevant data in each date range. Colour (RGB) and Normalised Difference vegetation Index (NDVI) images are downloaded from vegetation collections. Supported vegetation collections include Landsat (USGS, 2020) and Sentinel-2 (ESA, 2020) GEE collections. Cloud masking logic is included to improve data quality using the geetools package (Principe, 2020). For precipitation and temperature information, pyveg defaults to using the ERA5 collection (ECMWF, 2020).

Network centrality metrics

Network centrality methods are used to measure the connectedness of vegetation in images by treating the image as a network, with pixels containing significant vegetation as nodes. Vegetation pixels are ordered according to their subgraph centrality (Estrada & Rodríguez-Velázquez, 2005), and from this, a feature vector is constructed by calculating the Euler Characteristic (Richeson, 2012) for different quantiles. The slope of this feature vector gives a measure of how connected the vegetation is.

After completetion of the download job, pyveg computes the network centrality of the vegetation (Mander et al., 2017). To achieve this, the NDVI image is broken up into smaller

50×50 pixel sub-images. Each sub-image is then thresholded using the NDVI pixel intensity, and subgraph connectivity is computed for each binarized sub-image. The resulting metrics are stored, along with mean NDVI pixel intensities for each sub-image.

Time series analysis

pyveg analysis functionality is exposed via a `pveg_gee_analysis` command. The command accepts an argument, `--input_dir`, which points to a directory previously created by a download job. pyveg supports the analysis of the following time series: raw NDVI mean pixel intensity across the image, offset50 (a measure of the slope of the network centrality feature vector), and precipitation.

During data processing, pyveg is able to drop time series outliers and resample the time series to clean the data and avoid gaps. A smoothed time series is constructed using LOESS smoothing, and residuals between the raw and smoothed time series are calculated. Additionally, a deseasonalised time series is constructed via the first difference method.

Time series plots are produced, along with auto- and cross-correlation plots. Early warning signals are also computed using the `ewstools` package (Bury, 2020), including Lag-1 autocorrelation and standard deviation moving window plots. A sensitivity and significance analysis is also performed in order to determine whether any trends (quantified by Kendall tau values) are statistically significant.

Acknowledgements

References

Bury, T. (2020). Retrieved from https://github.com/ThomasMBury/ewstools

Dakos, V., Scheffer, M., Nes, E. H. van, Brovkin, V., Petoukhov, V., & Held, H. (2008). Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, *105*(38), 14308–14312. doi:10.1073/pnas.0802430105

ECMWF. (2020). Copernicus climate change service (c3s): ERA5: Fifth generation of ecmwf atmospheric reanalyses of the global climate.

ESA. (2020). Contains modified copernicus sentinel data, courtesy of esa.

Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Subgraph centrality in complex networks. *Phys. Rev. E*, *71*(5), 056103. doi:10.1103/PhysRevE.71.056103

Mander, L., Dekker, S., Li, M., Mio, W., Punyasena, S., & Lenton, T. (2017). A morphometric analysis of vegetation patterns in dryland ecosystems. *Royal Society Open Science*, *4*, 160443. doi:10.1098/rsos.160443

Principe, R. E. et al. (2020). Retrieved from https://github.com/gee-community/gee_tools

Richeson, D. S. (2012). *Euler's gem: The polyhedron formula and the birth of topology*. Princeton University Press. ISBN: 9780691154572

Trichon, V., Hiernaux, P., Walcker, R., & Mougin, E. (2018). The persistent decline of patterned woody vegetation: The tiger bush in the context of the regional sahel greening trend. *Global Change Biology*, *24*(6), 2633–2648. doi:10.1111/gcb.14059

USGS. (2020). Landsat-4, landsat-5, landsat-7, landsat-8 data courtesy of the u.s. Geological survey.

4