CIE4604 Simulation and Visualization

Assignment 3 – Time Series Analysis of Vegetation Index

Silvia Alfieri and Hans van der Marel
30 November 2021

Introduction

In this assignment you are given the task to calculate the Normalized Differenced Vegetation Index NDVI from multispectral Sentinel 2 EO data and analyse time series of NDVI images. The Harmonic Analysis of Time Series (HANTS) algorithm will be used in order to remove clouds and reconstruct a time series of NDVI in a specific area located in Spain. The NDVI will be computed and visualized in a Geographic Information System (GIS), such as QGIS or Arc-GIS. To run the HANTS algorithm and to make animation you will use Python.

The deliverables for this assignment are

1. Short report (pdf), or pdf of the Jupyter notebook, with the plots and answers to the questions.
2. Zip file with the output geotiff files and avi or mpeg files of reconstructed time series.
3. Jupyter notebook.

If you are going to (re-)submit, make sure to submit BOTH items at the same time. Resubmitting will overwrite all previous submissions.

Sentinel 2 multispectral EO data

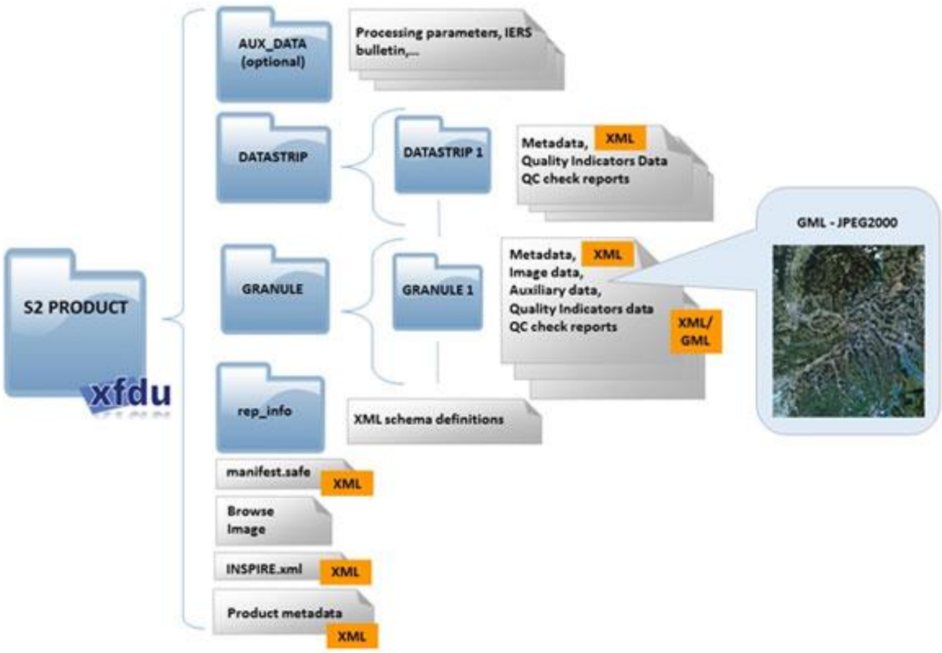
Sentinel-2 is an Earth observation mission developed by ESA as part of the Copernicus Programme to perform terrestrial observations in support of services such as forest monitoring, land cover changes detection, and natural disaster management. It consists of two identical satellites, Sentinel-2A and Sentinel-2B. The Sentinel-2 satellites each carry a single multi-spectral instrument (MSI) with 13 spectral channels in the visible/near infrared (VNIR) and short wave infrared spectral range (SWIR).

Spectral bands for the Sentinel-2 sensors[9]

Sentinel-2 bands	Sentinel-2A		Sentinel-2B		
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
Band 1 – Coastal aerosol	443.9	27	442.3	45	60
Band 2 – Blue	496.6	98	492.1	98	10
Band 3 – Green	560.0	45	559	46	10
Band 4 – Red	664.5	38	665	39	10
Band 5 – Vegetation red edge	703.9	19	703.8	20	20
Band 6 – Vegetation red edge	740.2	18	739.1	18	20
Band 7 – Vegetation red edge	782.5	28	779.7	28	20
Band 8 – NIR	835.1	145	833	133	10
Band 8A – Narrow NIR	864.8	33	864	32	20
Band 9 – Water vapour	945	26	943.2	27	60
Band 10 – SWIR – Cirrus	1373.5	75	1376.9	76	60
Band 11 – SWIR	1613.7	143	1610.4	141	20
Band 12 – SWIR	2202.4	242	2185.7	238	20

For Sentinel 2 you will use the so called Level-2A product. The Level-2A product provides Bottom Of Atmosphere (BOA) reflectance images. Each Level-2A product is composed of 100x100 km2 tiles in cartographic geometry (UTM/WGS84 projection, EPGS: 32630). The spatial resolution of SENTINEL-2 is dependent on the particular spectral band.

SENTINEL-2 products are made available in SENTINEL-SAFE format, including image data in JPEG2000 format, quality indicators (e.g. defective pixels mask), auxiliary data and metadata. The actual image data is included in the GRANULE directory in various spatial resolutions.



Sentinel-2 SAFE format

For more information see https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi .

Normalized Difference Vegetation Index (NDVI)

To determine the type and density of green vegetation on parcels of land researchers observe the distinct colors (wavelengths) of visible and near-infrared sunlight reflected by the plants. The solar spectrum, as can be observed from a prism, consists of many different wavelengths. When sunlight strikes objects, certain wavelengths of this spectrum are absorbed and other wavelengths are reflected. The pigment in plant leaves, chlorophyll, strongly absorbs visible light (from 0.4 to 0.7 µm) and use this for photosynthesis. The cell structure of the leaves, on the other hand, strongly reflects near-infrared light (from 0.7 to 1.1 µm). The more leaves a plant has, the more these wavelengths of light are affected.

Normalized Difference Vegetation Index (NDVI) quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). Written mathematically, the formula is:

$$NDVI = (NIR - RED)/(NIR + RED)$$

Calculations of NDVI for a given pixel always result in a number that ranges from minus one (-1) to plus one (+1); however, no green leaves gives a value close to zero. A zero means no vegetation and close to +1 (0.8 - 0.9) indicates the highest possible density of green leaves.
Overall, NDVI is a standardized way to measure healthy vegetation. When you have high NDVI values, you have healthier vegetation. When you have low NDVI, you have less or no vegetation.

For Sentinel 2 you will use Band 4 (Red) and Band 8 (NIR) in 10 meter resolution to compute NDVI. As each Sentinel 2 SAFE file is about 1.2 Gb in size you will be provided with a subset of the images, Band 4 and 8, only.

Input datasets for the assignment

The input dataset for this assignment is given in a zip file, available from http://gnss1.tudelft.nl/pub/varia/cie4604_a3_hants.zip , with five folders

L2A_T30STG_20170601T110651
L2A_T30STG_20170731T110651
Shapefiles_Bomba_BXII_13files_projected
Training_set
Hants

In the first two folders you will respectively find two Sentinel 2 L2A multispectral images acquired at two different dates in 2017. To save space these folders only contain the B04 and B08 bands. You will need these for Part I.

The third folder contains a shapefile that defines our area of interest, an agricultural area close to Sevilla, Spain. The fourth folder comes with a shapefile with training data for selected parcels. Please note that the coordinate reference system for the shape files is UTM zone 29N, whereas the Sentinel-2 uses UTM zone 30N. You will use these shapefiles both in part I and II, and use QGIS in part I to transform the coordinates to the proper UTM zone for Sentinel-2.

The last folder, Hants, contains the pre-processed NDVI time-series in GeoTIFF format for the HANTS algorithm in part II.

The Python (and Matlab) code and template notebooks are provided separately from the data.

Part I – Compute and Visualize NDVI in QGIS

Part 1.A – Compute 10 meter resolution NDVI over the area of interest for both dates

To accomplish this task you will need to import the L2A Sentinel-2 data into QGIS, and then (a) extract the data for the area of interest, (b) compute the NDVI over the area of interest for both dates, (c) save the NDVI as geotiff to file .

To extract the data for the area of interest you have first to add a vector layer with the area of interest (Shapefiles_Bomba_BXII_13files_projected). To extract the raster data you can use several methods:

- `Raster -> Extract -> Clip raster by mask layer`, using the shapefile,
- `Raster -> Extract -> Clip raster by extend` by typing the coordinates for the boundary

If you want to have a rectangular boundary, but don't want to type the coordinates for the boundary, you could first make a rectangular boundary using `Vector -> Research tools -> extract layer extend` on the shapefile, and then use `Raster -> Extract -> Clip raster by mask layer` on the created boundary. However, this gives an unexpected result. Why? What should we do to resolve this problem?

To do the actual clipping QGIS relies on another package, called GDAL/OGR. This is actually a console application you can run from the OSGeo4W Shell. The actual code you find in the `GDAL/OGR console call` box at the bottom of the pop-up window. It should be something like

```
gdalwarp -of GTiff -cutline <shapefile> -crop_to_cutline
<source-file> <destination-file>
```

With `select|CTRL-c` you can copy this code to script file, which is very convenient in case you have to repeat the same extraction on tens or hundreds of images!

To compute NDVI use the raster calculator.

In your report include at least the following results:
- Layout with the two NDVI maps, side by side, using the same map scale, with titles, legend and scale bar. You have to choose an appropriate representation (colormap) for the NDVI data (your grade depends on this).
- A brief explanation on the method you used to extracted the data for the area of interest, with answers to the questions.
- Your motivation for the chosen representation (colormap) for the NDVI data.

Hint: to obtain identical symbology for both NDVI maps you can save the style to a file, and load this style again for the other plot. Layout was formerly known as map composer in QGIS version 2.

Part 1.B – Load the training set in QGIS and visualize the different categories of crops.

Add the training dataset as a new vector layer and categorize the crops on the test fields QGIS (or Arc-GIS) using different symbology for each crop.

Reproject (`Vector -> Data Management Tools -> Reproject layer`) the training dataset to UTM zone 30N (the same as used for the NDVI dataset) and save this (`Layer -> Save as`) to a new shapefile (you will need this one in Part II). To see the extend of each shapefile you can use `Vector -> Research tools -> Extract layer extend`. This explains a lot.

Next use the zonal statistics tools (`Processing Toolbox -> Raster Analysis -> Zonal statistics`) to add three columns, with the count, mean and standard deviation, to the attribute table of the re-projected training dataset. Do this for both NDVI datasets, and use a meaningful `output column prefix` (please note that the shapefile fields can only be 10 character long, so keep the prefix short). Save the attribute table as an excel file (`right click -> save as`).

In your report include at least the following results:
- Layout with the categorized training dataset.
- Screenshot of the excel file with the attribute table

Furthermore, include in the zip file accompanying the report,
- The excel file with attributes (including zonal statistics)
- The re-projected training dataset (you will also need this for Part II)

Part II – Time series analysis using HANTS (in Python or Matlab)

In this part of the assignment you are asked to perform time series analysis using HANTS (Harmonic Analysis of Time Series) algorithm and visualize the results using Python.

HANTS is a software using the Fourier Analysis in order to detect outliers (clouds) and reconstruct image time series. The analysis have to be performed over Sector BXII, an agricultural area located close to Sevilla (Spain).

In the folder Hants you will find

Input files

- Geotiff file `NDVI_BXII_time_series_2017_2018_S2A_SP.tif` with NDVI time series for the area of interest
- Csv-file `NDVI_BXII_time_series_images_list.csv` with the band number, ISO date, sequential day number and day of year

The Jupyter notebook `CIE4604_M3_Assignment3_templae.ipynb` is a template that you can use for Part II. It already contains code for reading the GeoTIFF files and other things, but in places you have to modify or add code. The HANTS algorithm is code in `hants.py`. The notebooks and code is distributed separately from the data.

The input geotiff file `NDVI_BXII_time_series_2017_2018_S2A_SP.tif` with NDVI time series for the area of interest has already been prepared. It contains 33 NDVI images that have been prepared from the Sentinel-2 imagery using the steps outline in part I, using a script based on the GDAL/OGR command generated by QGIS. Because of the large Sentinel-2 data files this has already been done.

Part 2.A – Load the NDVI time series, make a pseudo color plot of the NDVI using an appropriate colormap, and make a video of the unfiltered NDVI.

First, carefully study the code for loading the GeoTIFF file. This is explained in more detail in Exercise 3.

✓ Then for the first day use the code to make a pseudo color image map of NDVI. The default colormap will **NOT** do, you have to make or select an appropriate colormap for NDVI (see also the various exercises).

If you do a Google search on NDVI colormap you get some useful hits, e.g. the following link https://publiclab.org/notes/cfastie/08-26-2014/new-ndvi-colormap contains useful information with downloadable map (search for NDVI_VGYRM-lut.txt, download, and convert to Python colormap).

✓ Once you have found an appropriate colormap you can proceed making an avi or mpeg video of the unfiltered NDVI. The code for this has been developed in Exercise 4.

Include the NDVI map in your report with a brief motivation for the chosen colormap, and include the avi/mpeg video in the zip file.

Part 2.B – Perform Time Series Analysis of NDVI images with HANTS algorithm using different configuration of input parameters for two selected points.

You can try different values for the number of frequencies (nf) , base period (nb) and fit error tolerance.

Plot the reconstructed vs the original time series for two vegetated pixels characterized by different temporal trend (different crop growing cycles). You can apply HANTS on a subset of the provided time series ("NDVI_BXII_time_series_2017_2018_S2A.tif")

In the report, explain your choice for the base period, and include the plots for the chosen parameters.

Part 2.C – Apply HANTS on the full extent of the provided time series using the best parameters setting ("NDVI_BXII_time_series_2017_2018_S2A.tif") and save a video of the reconstructed time series.

Include the video in the zip file.

The reconstructed NDVI, amplitude and phase are saved as GeoTIFF files, you will need the amplitude and phase for Part III.

Part 2.D – Visualize the image of the fitted parameters (amplitude and phases) for the frequencies you estimated.

Include the plot in your report.

Part 2.E – Plot for a cloudy date the original NDVI image and compare with the reconstructed one.

Include the plot in your report.

Part 2.F – Use the code in the last part of the notebook to read your re-projected training dataset (of part 1.B), and visualize this as a gray scale image with borders of the test fields overlayed (code is already provided).

Include the gray scale image (with borders) in your report.

Part III – Visualize HANTS output in QGIS

Load the geotiff with amplitudes and phases from HANTS in QGIS (or Arc-GIS) and make a layout showing this data as multiband color (assign appropriate bands to red, green and blue channels).

Include the layout of the multiband color and/or phase in your report.

The phase data actually tells you something about the main growing season. The main growing season corresponds to the amplitude peak of the first frequency (Frequency number 1) as given by the phase. The conversion to days should be

```
daynum = phase/360*(base period/frequency number)
```

which can be used to further categorize the results in QGIS, but you could also use this relation with the day of year to make a geotiff with the main growing season. However, this is only optional.

[End of assignment 3]