A WEB SERVICE FOR PHENOLOGY METRICS EXTRACTION FROM BIG EARTH OBSERVATION DATA

Gabriel Sansigolo[1]*, Karine Reis Ferreira*[1]*, Gilberto Ribeiro de Queiroz*[1]*, Marcos Adami*[1]*, Thales Körting*[1]

[1]National Institute for Space Research (INPE), Av. Dos Astronautas 1758, 12227-010 – São José dos Campos, SP – Brazil
{gabriel.sansigolo, karine.ferreira, gilberto.queiroz, marcos.adami, thales.korting}@inpe.br

ABSTRACT

Phenology is the study of timing recurrent biological events, being an important indicator of annual plant growth. From remote sensing images, it is possible to obtain metrics used for phenological monitoring, which are useful in understanding vegetation dynamics. Although there are many implementations for estimating crop phenology from remote sensing data, these implementations require software package installation and programming skills to use them. Besides that, analysts face with hardware limitations to process the big Earth observation (EO) data currently available from different providers. To minimize these efforts and limitations, this paper presents a web service called Web Crop Phenology Metrics Service (WCPMS) to extract phenological metrics from large volumes of remote sensing images modeled as multidimensional data cubes, produced by the Brazil Data Cube (BDC) project of the National Institute for Space Research (INPE). WCPMS allows analysts to calculate phenology metrics from big image data sets without needing to download them or to install software tools on personal computers. This paper describes the WCPMS architecture and Jupyter notebooks with examples about how to use it.

Key words – *big data, remote sensing images, phenological metrics, Earth observation data cubes, satellite image time series.*

1. INTRODUCTION

Phenology is the study of plant life cycles, their timing recurrent biological events driven by changes in weather and climate. It is an important indicator of annual plant growth [1]. Several methodologies for estimating crop phenology from remote sensing data have been developed and used to create different algorithms [2]. The most approaches to calculate phenology metrics from remote sensing data are based on image time series. To support image time series, Analysis-Ready Data (ARD) image collections have been modeled and organized as multidimensional data cubes. Earth observation (EO) data cubes can be defined as sets of image time series associated with spatially aligned pixels [3]. Phenology metrics, product of curve-fitting and/or threshold techniques applied to a vegetation index time-series, are an important feature for validating EO products. Because of its importance, the committee on Earth Observation Satellites (CEOS) has launched a mission focusing on developing a validation good practices protocol for satellite-based land surface phenology [4].

Nowadays, analysts have free access to an unprecedented amount of remote sensing images collected by different satellites/sensors with distinct spatial, temporal, radiometric, and spectral resolutions. In 2023, Landsat Collection 2 is composed of more than 10.2 million images that occupy more than 9 petabytes of data collected by the Landsat 1–5, Landsat 4 and 5, Landsat 7, and Landsat 8 and 9 [5]. Also in 2023, the yearly Sentinel-2 Level-1C/-2A product volume is increased from 3 to 4 petabytes [6]. Brazil Data Cube (BDC) is a research and development project of the Brazilian National Institute for Space Research (INPE) that focuses on producing multidimensional data cubes from big volumes of medium-resolution remote sensing images for the entire Brazilian territory. By now, the BDC project had produced more than 2 petabytes of ARD image collections and EO data cubes for Brazil [7].

A variety of open source methods and tools have been developed to identify phenological metrics from remote sensing images [2]. Although there are rich collections of these open methods and tools, these implementations require package installation and programming skill from the analysts to use them. Besides that, analysts face with hardware limitations to process the big EO data currently available from different providers. Because of the intensive demand for images time series to calculate phenological metrics, some packages require the analyst to create an ARD image collections locally, a big EO data, in order to calculate phenology metrics. To minimize these efforts, we propose a web service, software that runs in server-side, to calculate phenological metrics from EO data sets.

This paper presents a web service called Web Crop Phenology Metrics Service (WCPMS) to extract phenological metrics from big EO image collections, modeled as multidimensional data cubes, produced by the Brazil Data Cube (BDC) project of INPE. It allows analysts to calculate phenological metrics from data cubes without needing to download big EO data sets on their personal computers. The system runs on server-side, so it doesn't require package installation. By giving a spatial location or a region, it retrieves the phenological metrics associated to it.

2. METHODS AND TOOLS FOR CALCULATING PHENOLOGICAL METRICS

With the goal of building an open source web service for calculating phenological metrics based on EO data cube, we decided, instead of developing a new algorithm for phenology extraction, to use an established one. To do so, we perform a comparative analysis of the algorithms.

2.1. CropPhenology

The CropPhenology is an R package for extracting agricultural phenology from vegetation index time series. The package calculate 15 phenological metrics based on the NDVI over the season and some of its metrics are: Onset NDVI, Onset time, Maximum NDVI, Time of maximum NDVI, Offset NDVI, Offset time, Length of growing season, Length of growing season before MaxT, Length of growing season after MaxT, Growth rate between Onset and MaxT.

2.2. Digital Earth Australia Tool

The Digital Earth Australia tools (DEA) is a Python package included in the DEA library to calculate vegetation phenology statistics [8]. The package detects 11 phenological metrics of the changes in plant life using satellite images; the script uses either the NDVI or the EVI and some of its metrics are: Start of season, Peak of season, End of season and Start of season. One of the key features of the DEA is the integration of the Open Data Cube, an EO Data Cubes library.

2.3. TIMESAT

The TIMESAT is a software package developed for estimating growing seasons from satellite time-series, as well as for computing phenological metrics from the data [9]. The main implementations of TIMESAT are available at Fortran and Matlab. However, because of TIMESAT being the most established library for extracting phenological metrics, it has implementations in a few other programming languages. An example is in R, rTIMESAT. The software detects 18 phenological metrics and iteratively fits smooth mathematical functions to time-series and some of its metrics are: Peak of Season, Time of Peak of Season, Middle of Season, Valley of Season, Time of Valley of Season, Base, Start of Season, Time of Start of Season and End of season.

2.4. Phenolopy

Phenolopy is a python-based library for analysing satellite time series data [10]. The library is heavily based on the concept and methodology behind the TIMESAT software. A feature of the library is that it can be applied to any satellite imagery stack. The software detects 18 phenological metrics and iteratively fits smooth mathematical functions to time-series and some of its metrics are: Peak of Season, Time of Peak of Season, Middle of Season, Valley of Season, Time of Valley of Season, Base, Start of Season, Time of Start of Season and End of season.

2.5. Greenbrown

The Greenbrown is an R package with a collection of functions to analyze trends, trend changes and phenology events in time series, like from satellite observations or climate model simulations [11]. The package calculate 13 metrics of land surface phenology and greenness based on the NDVI and some of its metrics are: Start of season, End of Season, Length of season, Position of peak, Position of trough and Mean growing season.

2.6. Phenex

The Phenex is an R package with auxiliary functions for phenological data analysis. It provides some easy-to-use functions for spatial analyses of (plant) phenological data sets and satellite observations of vegetation [12]. The package calculate 10 metrics of land surface phenology and greenness based on the NDVI and some of its metrics are: max, maxval, maxval, min, minval, minval and greenup.

2.7. Conclusion

Beside the comparative analysis of the algorithms, we also used the study carried out by Grazieli Rodigheri [2] to determine which algorithm use, to get the best estimates of crop sowing and harvesting dates. For the first version of the web service, we decide to use the Phenolopy library [10].

3. WEB CROP PHENOLOGY METRICS SERVICE

The Web Crop Phenology Metrics Service (WCPMS) is a web service for calculating phenological metrics from big EO data cubes of the Brazil Data Cube project, based on Phenolopy library. Through a simple API, analysts give a spatial location or region and the WCPMS retrieves the phenological metrics associated to it, using image time series. The metrics returned by WCPMS are: (a) Peak of Season; (b) Middle of Season; (c) Valley of Season; (d) Base; (e) Start of Season; (f) End of Season; (g) Length of Season; (h) Rate of Increase; (i) Rate of Decrease; (j) Amplitude of Season; (k) Short Integral of Season; (l) Long Integral of Season; (m) Short Integral of Total; (n) Long Integral of Total.

3.1. WCPMS Architecture

The WCPMS architecture, shown in Figure 1, is made up of two sides: (1) the server-side and (2) the client-side. It allows analysts to calculate phenological metrics from data cubes with no need to download big EO datasets on their personal computers. The web service runs on the server-side, so everything related to it, such as the libraries for calculating phenological metrics, the access to the web service for extracting time series, is on the cloud. So, it doesn't require any package installation on client side. If a analyst wants to work in a region, the service provides support. When working in regions, it is necessary to select a method: (1) all the pixels; (2) systematic grid, a N meter neighborhood distance rule; (3) random grid, N points distributed at random.

3.2. WCPMS Usage

Since WCPMS is a web service, the operations are routes, each of which returns its response when accessed, all in JSON format. On Listing 1, an example of the **phenometrics** route is shown.

```
#request
https://data.inpe.br/bdc/wcpms/phenometrics?
collection=S2-16D-2&latitude=-29.20263&
longitude=-55.9554&start_date=2022-01-01&
end_date=2022-12-31&freq=16D&band=NDVI

#response
```

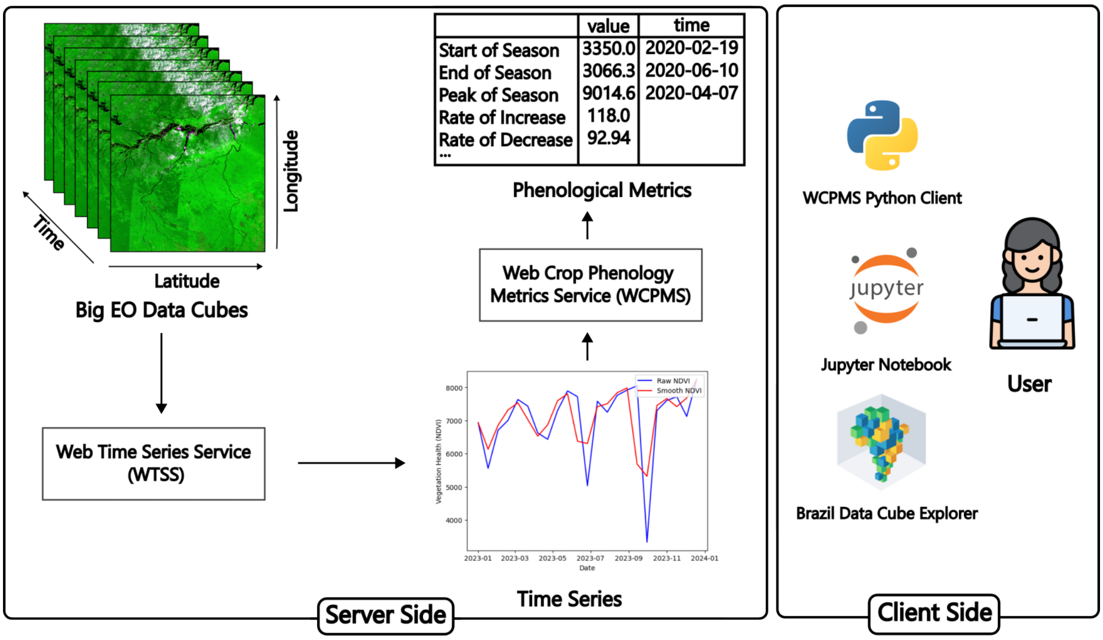

Figure 1: Architecture of the Web Crop Phenology Metrics Service (WCPMS).

```
{
   "phenometrics": {
      "eos_t": "2022-11-18T00:00:00",
      "eos_v": 2537.333251953125,
      "sos_t": "2022-07-13T00:00:00",
      "sos_v": 3195.0,
      ...
   },
   "timeseries": {
      "timeline": [
        "2022-01-01",
        "2022-01-17",
        "2022-02-02",
         ...
   ],
      "values": [
        5593.0,
        2879.0,
        1799.0,
         ...
   ]
    }

}
```

Listing 1: WCPMS retrieving phenological metrics.

WCPMS calculates the phenological metrics and returns them as a response to the request, as well as the image time series used and its timeline.

3.3. WCPMS Client

The WCPMS is a web service, through its API. To facilitate these operations, we have developed an official client, a simple Python library made from scratch. With this library, you can easily make a request to the web service. The goal when developing the python client was to make a library easy to use and edit, and allow the community to implement new features on top of it. We also decided that one goal would be to facilitate the request to the web service by presenting the parameters as python objects. The WCPMS client has a group of functions. The main ones are:

- **get_collections** List available data cubes in the BDC's SpatioTemporal Asset Catalogs (STAC).
- **get_description** List the information on each of the phenological metrics, such as code, name, description and method.
- **get_phenometrics** Returns the phenological metrics calculated for the given spatial location, as well as the image time series and timeline used.
- **get_phenometrics_region** List phenological metrics calculated for each of the given spatial location based on selected region methodology (all, systematic or random grid).

Besides the python client, we produce a Jupyter Notebook, entitled wcpms-introduction, to show how to use both the web service and the python client. More information and reproducible code are available via GitHub: https://github.com/brazil-data-cube/wcpms.py

3.4. WCPMS Client Usage

This section presents the results of using the proposed web service, described in section 3.1. The code listed in Listing 2 shows the use of the wcpms.py library and its method `get_phenometrics`.

```
from wcpms import *

# Defines URL of a instance of the WCPMS
wcpms_url = "https://data.inpe.br/bdc/wcpms"

# Defines a data cube query from BDC
datacube = cube_query(
    collection="S2-16D-2",
    start_date="2022-01-01",
    end_date="2022-12-31",
    freq="16D",
    band="NDVI"
)
```

```python
# Retrieving the phenological metrics
pm = get_phenometrics(
    url = wcpms_url,
    cube = datacube,
    latitude=-29.20263,
    longitude=-55.9554
)

# Visualizing the phenological metrics
print(pm)
```

Listing 2: Retrieving phenology metrics using wcpms.py.

Besides functions for retrieving phenological metrics, the wcpms.py library has functions designed for visualization. We developed a set of visualization functionalities, based on the Matplotlib library, to encapsulate standard plots for time series that highlight phenological metrics. The plot function was developed for Jupyter notebooks. It works by simply calling `plot_phenometrics(datacube, pm)` and returns a graph with the image time series and some phenological metrics: Start of Season (SOS), End of Season (EOS), Peak of Season (POS) and Valley of Season (VOS), as shown in Figure 2.

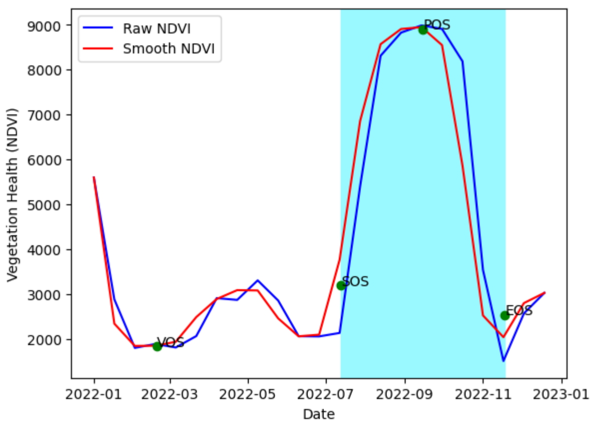

Figure 2: WCPMS plot - the image time series with the metrics Start of Season (SOS), End of Season (EOS), Peak of Season (POS) and Valley of Season (VOS)

4. CONCLUSIONS

We present the WCPMS, a web service to extract phenological metrics from big EO datasets produced by the BDC. The goal when developing the python client was to encapsulate the functionalities of established algorithms for phenology metrics extraction in a web service and connect it to the established web services of the BDC project. We managed to build a simple web service API that allows analysts to calculate phenological metrics from data cubes without needing to download big EO datasets on their personal computers.

5. ACKNOWLEDGEMENTS

We thank the HARMONIZE project, financed by Wellcome Trust (https://wellcome.org) grant number 22469/Z/21/Z, through the financial collaboration Oswaldo Cruz Fundation (FIOCRUZ) and Foundation for Scientific and Technological Development In Health (FIOTEC) ID Project: ICICT-002-FEX-22.

The authors also thank the São Paulo Research Foundation (FAPESP, grant no 2023/09118-6), and the Brazilian National Council for Scientific and Technological Development (CNPq, grant no 302205/2023-3 and grant no 302517/2023-5).

6. REFERENCES

[1] H. Lieth. *Phenology and Seasonality Modeling*. Ecological Studies. Springer Berlin Heidelberg, 2013.

[2] Grazieli Rodigheri, Ieda Del'Arco Sanches, Jonathan Richetti, Rodrigo Yoiti Tsukahara, Roger Lawes, Hugo do Nascimento Bendini, and Marcos Adami. Estimating crop sowing and harvesting dates using satellite vegetation index: A comparative analysis. *Remote Sensing*, 15(22), 2023.

[3] Ferreira K.R., Queiroz G.R., Vinhas L., Marujo R.F.B., Simoes R.E.O., Picoli M.C.A., Camara G., Cartaxo R., Gomes V.C.F., Santos L.A., and et al. Earth observation data cubes for brazil: Requirements, methodology and products. *Remote Sensing*, 12(24), 2020.

[4] Joshua Gray and Víctor Rodríguez-Galiano. Ceos land product validation website at nasa's goddard space flight center. `https://lpvs.gsfc.nasa.gov/Pheno/Pheno_home.html`. Accessed: 2024-06-23.

[5] Christopher Crawford, David Roy, Saeed Arab, Christopher Barnes, E. Vermote, Glynn Hulley, Aaron Gerace, Mike Choate, Christopher Engebretson, Esad Micijevic, Gail Schmidt, Cody Anderson, Martha Anderson, Michelle Bouchard, Bruce Cook, Ray Dittmeier, Danny Howard, Calli Jenkerson, Minsu Kim, and Steve Zahn. The 50-year landsat collection 2 archive. *Science of Remote Sensing*, 8:100103, 09 2023.

[6] Bernhard Bauer-Marschallinger and Konstantin Falkner. Wasting petabytes: A survey of the sentinel-2 utm tiling grid and its spatial overhead. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202:682–690, 2023.

[7] K. R. Ferreira, G. R. Queiroz, R. F. B. Marujo, and R. W. Costa. Building earth observation data cubes on aws. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022:597–602, 2022.

[8] DEA. Digital earth australia — dea. vegetation phenology. `https://docs.dea.ga.gov.au/notebooks/`. Accessed: 2024-06-23.

[9] P. Jonsson and L. Eklundh. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40(8):1824–1832, 2002.

[10] Lewis Trotter. Phenology. `https://github.com/lewistrotter/Phenology`. Accessed: 2024-06-23.

[11] Wutzler T Forkel M. greenbrown - land surface phenology and trend analysis. `http://greenbrown.r-forge.r-project.org/`, 2015. Accessed: 2024-06-23.

[12] Daniel Doktor Maximilian Lange. Phenex - auxiliary functions for phenological data analysis. `https://cran.r-project.org/web/packages/phenex/index.html`, 2017. Accessed: 2024-06-23.