A machine learning software framework for extraction of phenology indicators from multi-temporal sentinel-2 images

Dounia Arezki*, Hadria Fizazi*, Santiago Belda**, Charlotte De Grave**, Luca Pipia**, and Jochem Verrelst**
*University of science and technology of Oran, El Mnaouar, BP 1505, Bir El Djir 31000, Oran, Algeria
**Image Processing Laboratory (IPL), University of Valencia, C/Catedratico Jose Beltran 2, 46980, Paterna, Valencia, Spain

EGU General Assembly 2020
Vienna | Austria | 3–8 May 2020

Abstract

Optical Earth observation satellites provide spatially-explicit data that are necessary to study trends in vegetation dynamics. However, more of often than not optical data are discontinuous in time, due to persistent cloud cover and instrumental noises. Hence, the operating constraints of these data require several essential pre-processing steps, especially when aiming to reach towards monitoring of vegetation seasonal trends. To facilitate this task, here we present an end-to-end processing software framework applied to Sentinel-2 images.

To do so, first biophysical retrieval models were generated by means of a trained machine learning regression algorithm (MLRA) using simulated data coming from radiative transfer models. Among various tested MLRAs, the variational heteroscedastic Gaussian process regression (VHGPR) was evaluated as best performing. to train the retrieval model. The training and retrieval were conducted in the Automated Radiative Transfer Models Operator (ARTMO) software framework.

Subsequently, in view of retrieving the phenological parameters from the obtained vegetation products, a novel times series toolbox as part of the ARTMO framework was used, called: Decomposition and Analysis of Time Series software (DATimeS). DATimeS provides temporal interpolation among other functionalities with several advanced MLRAs for gap filling, smoothing functions and subsequent calculation of phenology indicators. Various MLRAs were tested for gap filling to reconstruct cloud-free maps of biophysical variables at a step of 10 days.

A demonstration case is presented involving the retrieval of Leaf area index (LAI), fraction of Absorbed Photosynthetically Active Radiation (FAPAR) from sentinel-2 time series. A large agricultural Algerian site of 143, 75 km² including Oued Rhiou, Ouarizane, Djidioua (1,345,075 pixels) was chosen for this study. A reference image was excluded from the time series in order to evaluate the reconstruction accuracy over a 40-day artificial gap.

The reference vs. Reconstructed maps produced by the gap-filling methods were compared with statistical goodness-of-fit metrics. Considering both accuracy and processing speed, the fitting algorithms Gaussian process regression (GPR) and Next neighbor interpolation (R^2= 0.90 / 0.081 sec per pixel and R^2=0.88 / 0.001 sec per pixel respectively) interpolations proved to reconstruct the vegetation products the most efficient, with GPR as more accurate but Next faster by a factor of 70.

Finally, we evaluated of the phenology indicators such as start-of-season and end-of-season based on LAI and FAPAR. The obtained maps provide valid information of the vegetation dynamics. Altogether, the ARTMO-DATimeS software framework enabled seamless processing of all essential steps: (1) from L2A sentinel-2 images converted to vegetation products, (2) to cloud-free composite products, and finally (3) converted into vegetation phenology indicators.

ARTMO / Datimes software

Automated Radiative Transfer Models Operator (**ARTMO**) Graphic User Interface (GUI) is a software package that provides essential tools for running and inverting a suite of plant RTMs, both at the leaf and the canopy level. The Decomposition and Analysis of Time Series software (DATimeS) expand established time-series interpolation methods with a diversity of advanced machine learning fitting and smoothing algorithms.

Figure 1. Treatment steps .

Model creation

Biophysical retrieval models are generated through a trained machine learning regression algorithm (MLRA), using the retrieval option and simulated data coming from radiative transfer models as an input. Next, we selected the single-output from settings and trained the model with variational heteroscedastic Gaussian process regression (VHGPR). Finally, we choose the validation type of preference.

Datasets

For the study, a large agricultural region was picked (143,75 km2 or 55,50 mi2) in Relizane (35 44 00 N, 0 33 00 E), Algeria.

- Sentinel-2 (S2) time series collected between 2018 and 2020 with cloud-free.
- 10 days temporal resolution achieved with Decomposition and Analysis of Time Series software (DATimeS).

Vegetation product retrieval (LAI / FAPAR)

LAI and FAPAR products have been produced from the sentinel-2 imagery obtain from the Copernicus platform; the retrieval step was conducted on the ARTMO software, we applied the biophysical variables model retrieval directly to the time series, a sample of the LAI and FAPAR product on **Figure.2/3**.

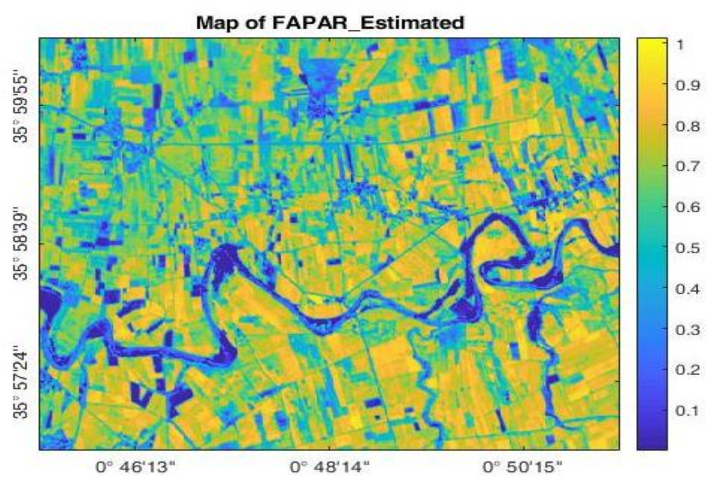

Figure 2. Map of FAPAR.

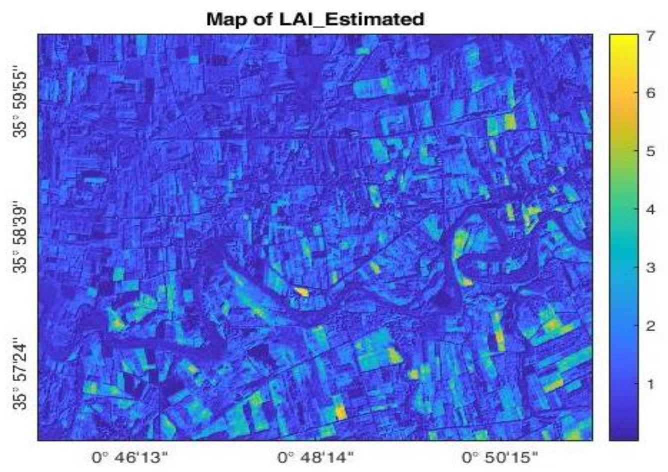

Figure 3. Map of LAI.

Map reconstruction accuracy

For the following step of the process, gap-filling treatment was applied to the time series with a ten-day step. In this context, Decomposition and Analysis of Time Series software (DATimeS) was used. **Figure 4**. Thereafter, we evaluated the reconstruction effectiveness by running a variety of powerful interpolation algorithms to reconstruct the FAPAR information on a selected date and compare it with a previously map from the FAPAR time series to be used as a reference for assessment purposes, being the date 03-01-2019. Finally, we calculated the goodness-of-fit statistics of reference map versus reconstructed map. Four statistical assessment measures are used to evaluate.**Table1**.

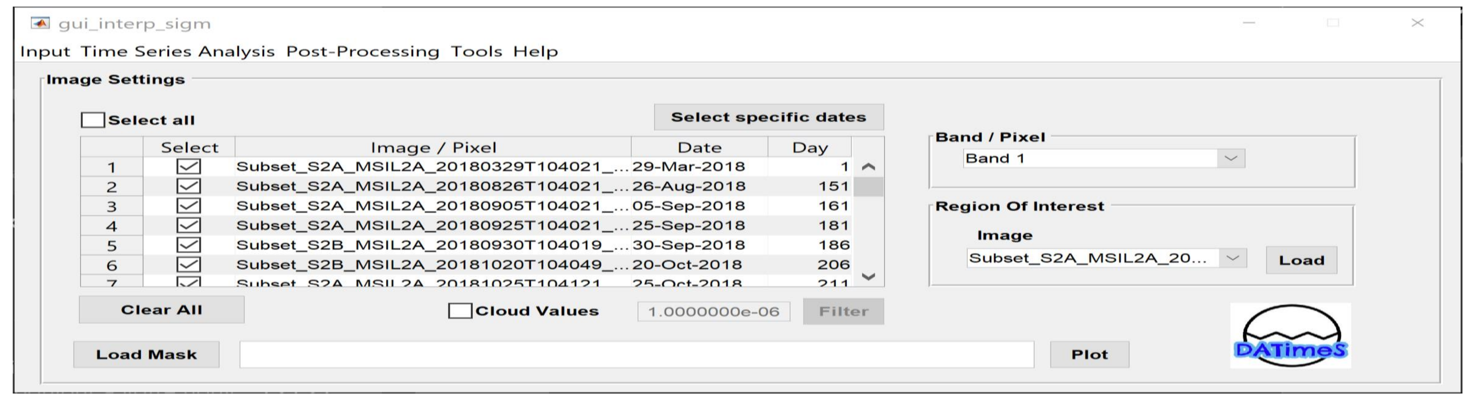

Figure 4. Datimes interface for Gap filling.

Table 1. Goodness-of-fit statistics and processing time for the reference vs. FAPAR reconstructed map as produced by the gap-filling methods.1,345,075 pixels.

Methods	MEA	RMSE [%]	R	R^2
GPR	0.0642	0.0854	0.9512	0.9048
KNRR	0.1084	0.1500	0.7895	0.6233
KRR	0.0711	0.1219	0.8879	0.7884
Next	0.0575	0.0824	09411	0.8856
Polyfit	0.1610	0.1966	0.6078	0.3694

Phenology indication

Aiming to study the seasonal pattern in vegetation variation, Datimes was used. Firstly, a region of interest was created and applied to the time series in order to derive phenological variables such as the start and end of a growing season. **Figure 5-10**.

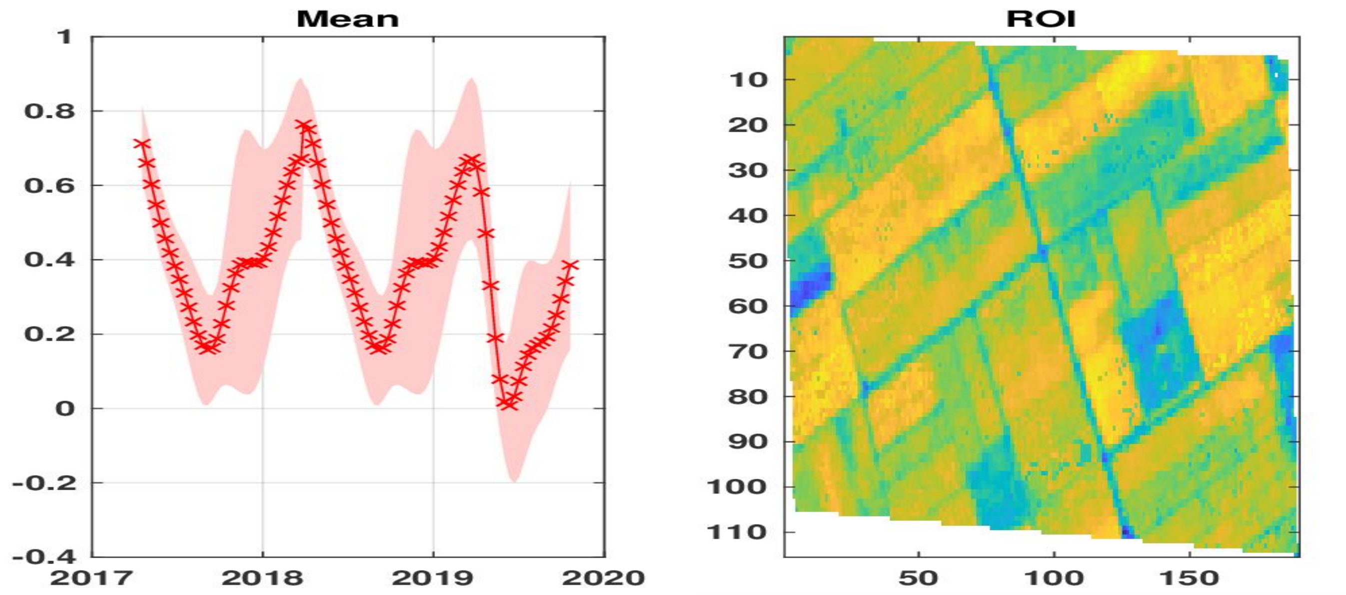

Figure 5. Regio of interest .

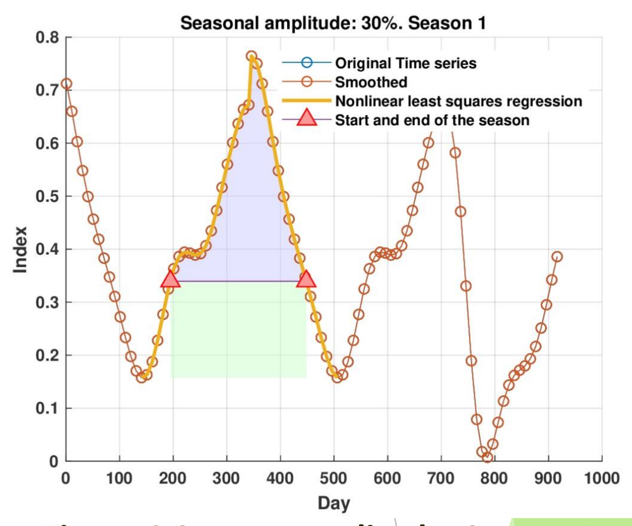

Figure 6.Season amplitude. Season 1

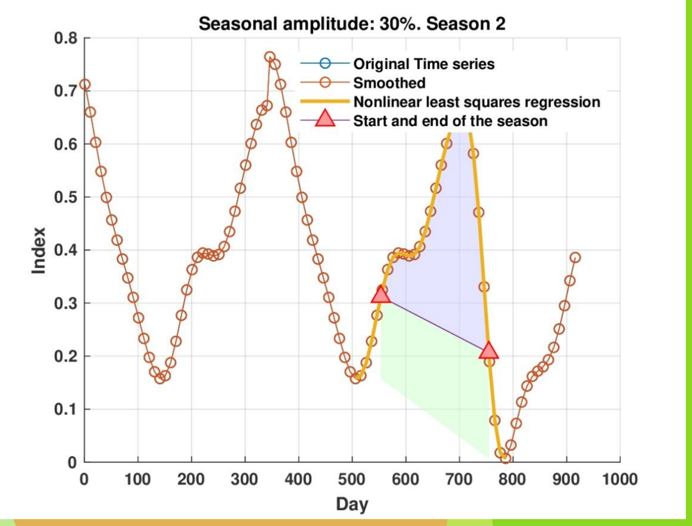

Figure 7. .Season amplitude. Season 2

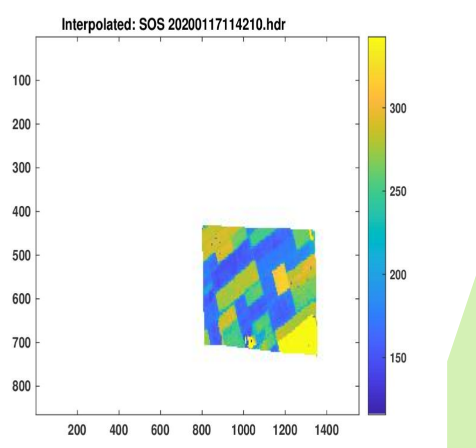

Figure 8. Start of the season

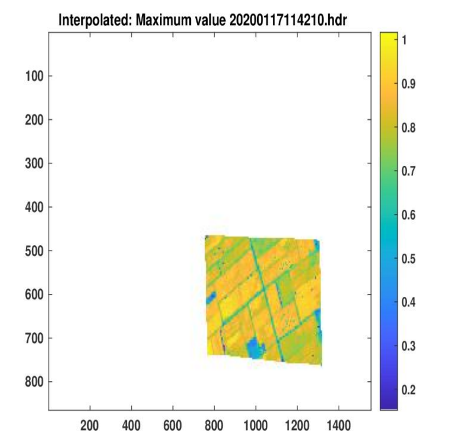

Figure 9.Maximum value

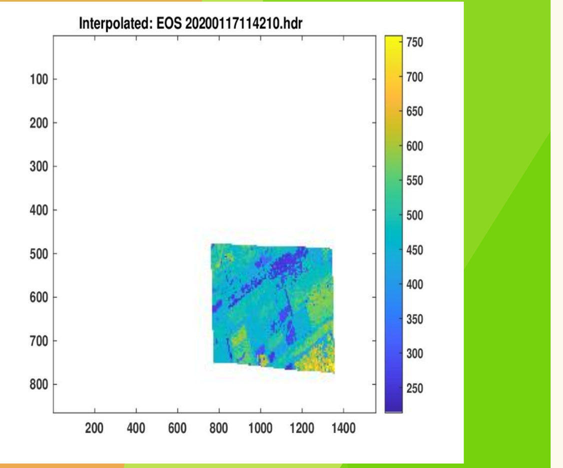

Figure 10. End of the season

Conclusions

Continuous spatial information about phenology has proven to be paramount, especially during this Covid-19 pandemic, this research presents the exploitation of two powerful tools for different steps of the phenological parameters extraction. The uncomplicated creation of vegetation product retrieval model with the ARTMO toolbox. A critical comparative analysis of the capability of five different algorithms for biophysical variables reconstruction, Results show that the GPR interpolation outperforms the other models in terms of model accuracy (R2 = 0.905, RMSE = 0.85). KRR achieves the second-best results in terms of performance (R2 = 0.788, RMSE = 1.22) and also ran four times faster than GPR, which means that this method offers excellent potential to deliver near-real-time operational products. Last, we obtained the different seasons in our time-series with Datimes that permit us to obtain phenological parameters that allow as to conduct yield prediction.

Contact

<Dounia Arezki / Jochem Verrelst>
< University of science and technology of Oran, El Mnaouar, BP 1505, Bir El Djir 31000, Oran, Algeria
Image Processing Laboratory (IPL), University of Valencia, C/Catedratico Jose Beltran 2, 46980, Paterna, Valencia, Spain >
Email: dounia.arezki@univ-usto.dz / jochem.verrelst@uv.es

References

1. Peter M. Atkinson, C. Jeganathan, Jadu Dash, and Clement Atzberger. Inter-comparison of four models for smoothing satellite sensor time series data to estimate vegetation phenology. Remote Sensing of En vironment, 123:400 – 417, 2012.
2. Juan Pablo Rivera-Caicedo, Jochem Verrelst, Jordi Muoz-Mar, Gustau Camps-Valls, and Jos Moreno. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. ISPRS Journal ofPhotogrammetry and Remote Sensing, 132:88 – 101, 2017
3. Santiago Belda, Luca Pipia, Pablo Morcillo-Pallarés, Juan Pablo Rivera-Caicedo, Eatidal Amin, Charlotte De Grave, Jochem Verrelst, DATimeS: A machine learning time series GUI toolbox for gap-filling and vegetation phenology trends detection, Environmental Modelling & Software, Volume 127, 2020,104666,ISSN 1364-8152,
4. H. Mao, J. Meng, F. Ji, Q. Zhang, and H. Fang. Comparison of machine learning regression algorithms for cotton leaf area index retrieval using sentinel-2 spectral bands. Applied Sciences (Switzerland), 9(7), 2019.cited By 3.