A Spark-based platform to extract phenological information from satellite images

1st Viktor Bakayov
UvA
Amsterdam, the Netherlands
viktorbakayov@gmail.com

2nd Romulo Goncalves
NLeSC
Amsterdam, the Netherlands
r.goncalves@esciencecenter.nl

3rd Raul Zurita-Milla
Faculty ITC-UT
Enschede, the Netherlands
r.zurita-milla@utwente.nl

4th Emma Izquierdo-Verdiguier
IPL-UValencia, Faculty ITC-UT
Enschede, the Netherlands
emma.izquierdo@uv.es

I. Abstract

Phenology is the study of periodic plant and animal life cycle events and how these are influenced by seasonal and inter-annual variations in weather and climate, as well as in other environmental factors. Time series of remote sensing (RS) images can be used to characterize land surface phenology at continental to global scales. For this, the RS images are typically transformed into various vegetation indices (VI) such as the normalized difference vegetation index (NDVI) or the enhanced vegetation index (EVI). These indices can then be used to extract various phenological metrics.

In our previous work we used cloud computing to generate temperature-based phenological indices [1], [2], and to relate one phenological metric, namely the Start-of-Season (SOS), with those indices [3], [4]. Here we present an extension of our work where we use a Spark-based platform to efficiently extract phenological metrics from time series of NDVI and EVI. This platform allows obtaining and analyzing high spatial resolution metrics (in this case 1km) from 10-day composites. The platform uses the same architecture as in [3], i.e., it is organized into three layers: a storage layer, a processing layer, and JupyterHub services for user-interaction. It is designed to store the data in well-known file formats like GeoTiffs and Hierarchical Data Format (HDF). For the data analysis the user expresses the operations in Jupyter notebooks as Python, R, or Scala code (Fig. 1). Hence, with a browser and remote connection, the user can express a research question and/or collect insights from large data sets. All computations are pushed down to the computational platform, and results fetched back for data visualization.

To extract the phenological metrics, we rely on TimeSat [5]. TimeSat is a software package that can be used to fit a function (e.g. double logistic) to time series of VIs. After that, it uses various approaches to extract vegetation seasonality metrics such as SOS. The programs numerical and graphical routines are coded in Matlab and Fortran. These routines are highly vectorized and efficient for use with large data sets. However, distributed processing is required to determine SOS at continental scales. Through an efficient partition of the data, and Spark's scheduling policies, these single-core routines are scheduled for parallel execution over multiple machines.

The study evaluates which VIs and fitting functions are most suitable for certain vegetation types by comparing the SOS metrics to volunteered phenological observations curated by the USA national phenological network [6]. Our preliminary results show there can be up to 20-30 days differences in the SOS depending on the fitting function, the VI and the approach used to extract the SOS metric. In the South, SOS is around mid-February or March whereas in mountainous regions and the North, the SOS can be as late as June-July. We are to further evaluate how our results compare to the ground volunteered observations. This work is then a first stepping stone towards being able to systematically analyze and map the impact of climate change on the seasonality of plants. Our tests show that the platform is scalable and can be extended to work with even higher resolution VIs, such as those that can be derived from Sentinel-2 images (10 m resolution). Because of this, our work opens the door to studies at continental to global scales, and to the use of high and very high spatial resolution data.

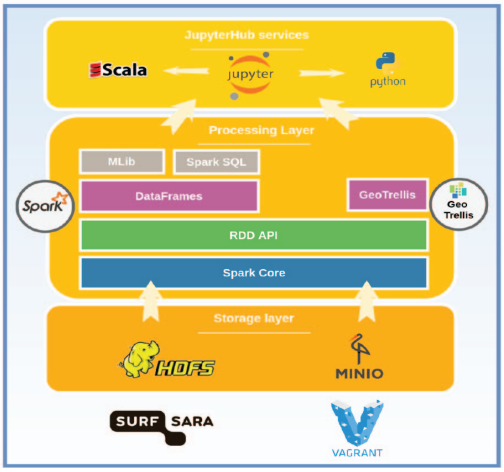

Fig. 1. Computational platform

II. Acknowledgment

This work has been partially supported by the NLeSC Project: High spatial resolution phenological modelling at continental scales and it was carried out using the Dutch national e-infrastructure with the support of the SURF Cooperative. E. Izquierdo-Verdiguier is supported by the APOSTD/2017/099 Generalitat Valencia grant (Spain).

978-1-5386-9156-4/18/$31.00 ©2018 IEEE
DOI 10.1109/eScience.2018.00095

REFERENCES

[1] H. Mehdipoor, E. Izquierdo-Verdiguier, and R. Zurita-Milla, "Continental-scale monitoring and mapping of false spring: A cloud computing solution," in *2017 International Conference on GeoComputation: Celebrating 21 Years of GeoComputation*, 2017.

[2] E. Izquierdo-Verdiguier, R. Zurita-Milla, T. R. Ault, and M. D. Schwartz, "Development and analysis of spring plant phenology products: 36 years of 1-km grids over the conterminous us," *Agricultural and forest meteorology*, vol. 262, pp. 34–41, 2018.

[3] R. Zurita-Milla, R. Goncalves, E. Izquierdo-Verdiguier, and F. O. Ostermann, "Exploring vegetation phenology at continental scales: Linking temperature-based indices and land surface phenological metrics," in *Proceedings of the 2017 conference on big data from space (BiDS '17), 28-30 November 2017, Toulouse, France*, 2017, pp. 63–66.

[4] R. Zurita-Milla, R. Bogaardt, E. Izquierdo-Verdiguier, and R. Goncalves, "Analyzing the cross-correlation between the extended spring indices and the AVHRR start of season phenometric," in *European Geosciences Union General Assembly 2018*, 2018.

[5] P. Jönsson and L. Eklundh, "TIMESAT—a program for analyzing time-series of satellite sensor data," *Computers & Geosciences*, vol. 30, no. 8, pp. 833 – 845, 2004.

[6] "National phenological network," https://www.usanpn.org/.