Exploring vegetation phenology at continental scales

R. Zurita-Milla*, R. Goncalves**, E. Izquierdo-Verdiguier*, F.O. Ostermann*
*Faculty ITC - University of Twente and **NLeSC, the Netherlands

Introduction

Phenology is the science that studies the timings of recurring biological events, their causes and variations in space and time. Here, we present a novel exploratory analysis where we link temperature-based phenological indices and land surface phenological metrics derived from remotely sensed images. Our exploratory analysis, illustrated with two multi-decadal and high-spatial resolution phenological products for continental USA, focuses on identifying phenological regions and on mapping the coherence between existing phenological products.

Phenological products

- Extended spring indices (SI-x): the SI-x are a suite of models that transform daily temperatures into consistent metrics, namely the day of the year (DOY) of first leaf and of first bloom for three key indicator species. Here, we use a recently developed long-term (1980 to 2015) and high spatial resolution (1km) version of the Leaf and Bloom indices. Figure 1 [Top] illustrates, as an example, the average Leaf index values.

- Land surface phenology: remotely sensed images can be used to derive metrics like the Start of Season (SOS), which indicates the beginning of photosynthetic activity in plants. Here we use a SOS product specifically made for the US by processing time series of the Advanced Very High Resolution Radiometer (AVHRR) sensor. The spatial resolution of this product is also 1km but it is only available for the period 1989 - 2014. Figure 1 [Bottom] illustrates the average SOS values. Notice that negative values indicate that the SOS took place the year before.

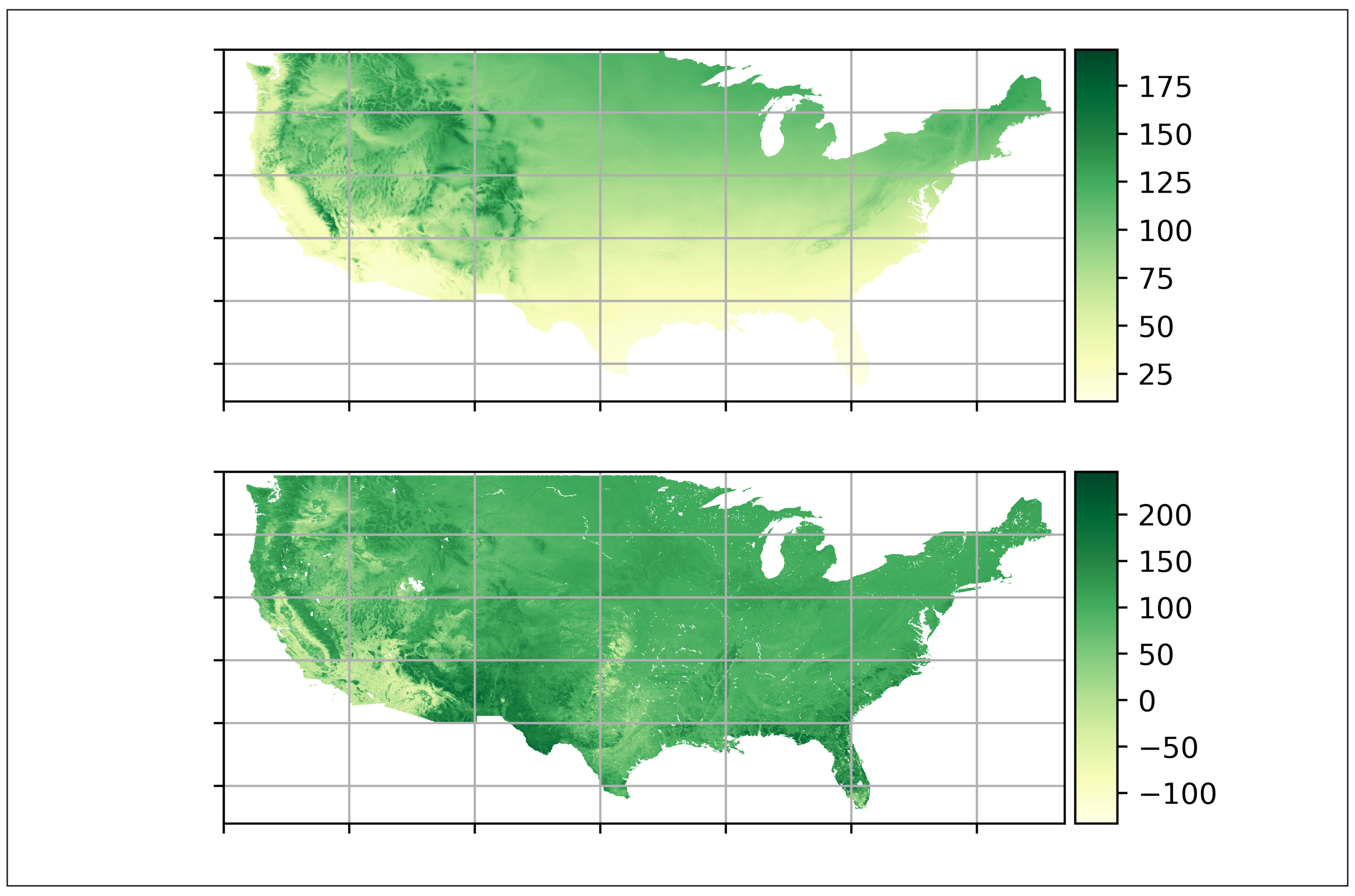

Fig. 1: Average of Leaf index [Top] and AVHRR SOS [Bottom] maps of contiguous North-America from 1989 to 2014.

Exploratory data analysis

- Phenoregions: we use K-means to identify phenoregions from the time series of the SI-x indices and the SOS metric. The optimal K value was found by locating the "elbow" of the Within Cluster Sum of Squared Error graph. Figure 2 shows the clustering results. The optimal number of phenoregions is 70 for the SI-x indices and 100 for the SOS metric. Land cover has therefore a larger impact on phenology than temperature differences across the US.

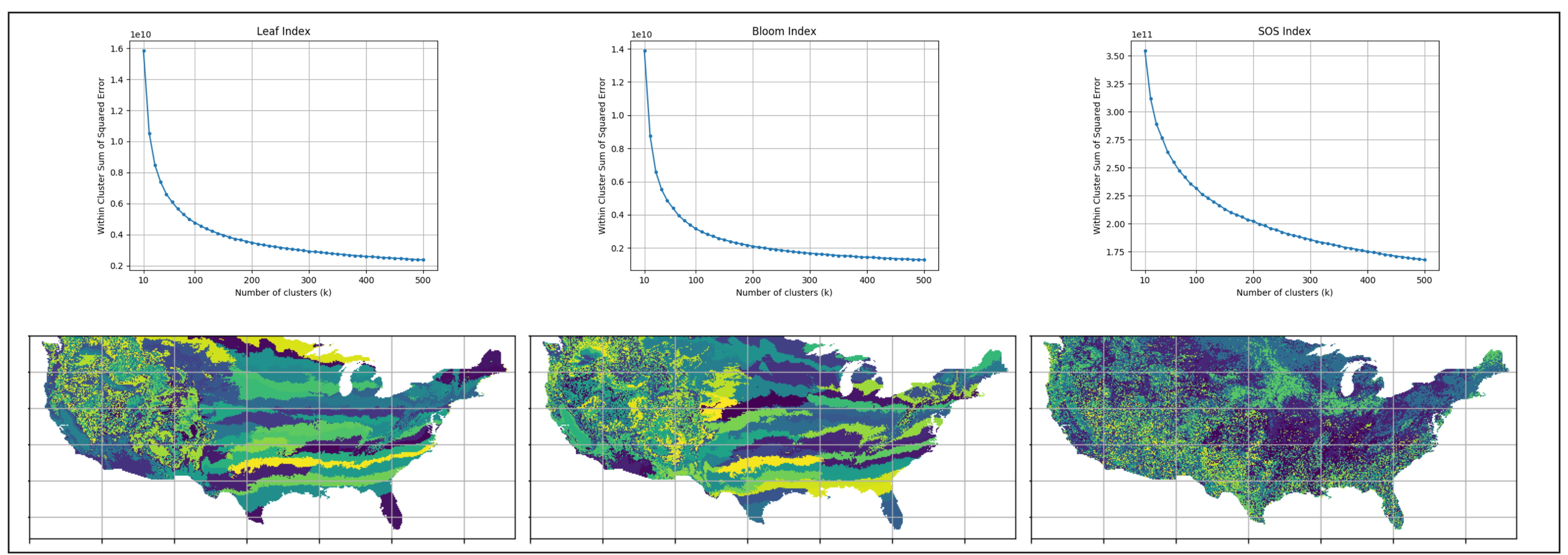

Fig. 2: Within cluster sum of squared Errors vs the number of clustersrow]. Clustering maps for the Leaf and bloom indices (k=70) andfor the Leaf and Bloom indices and the SOS metric [Top the SOS metric (k=100) [Bottom row].

- Spatio-temporal correlation: To better understand the ecological meaning of land surface phenological metrics, we performed a correlation analysis between the SI-x indices and the SOS metric. Figure 3 shows that large areas exhibit moderate to high positive correlations. This confirms that temperature is the main drivers of phenological development at mid latitudes. Figure 3 also shows that the Leaf index is less correlated with the SOS metric than the Bloom index.

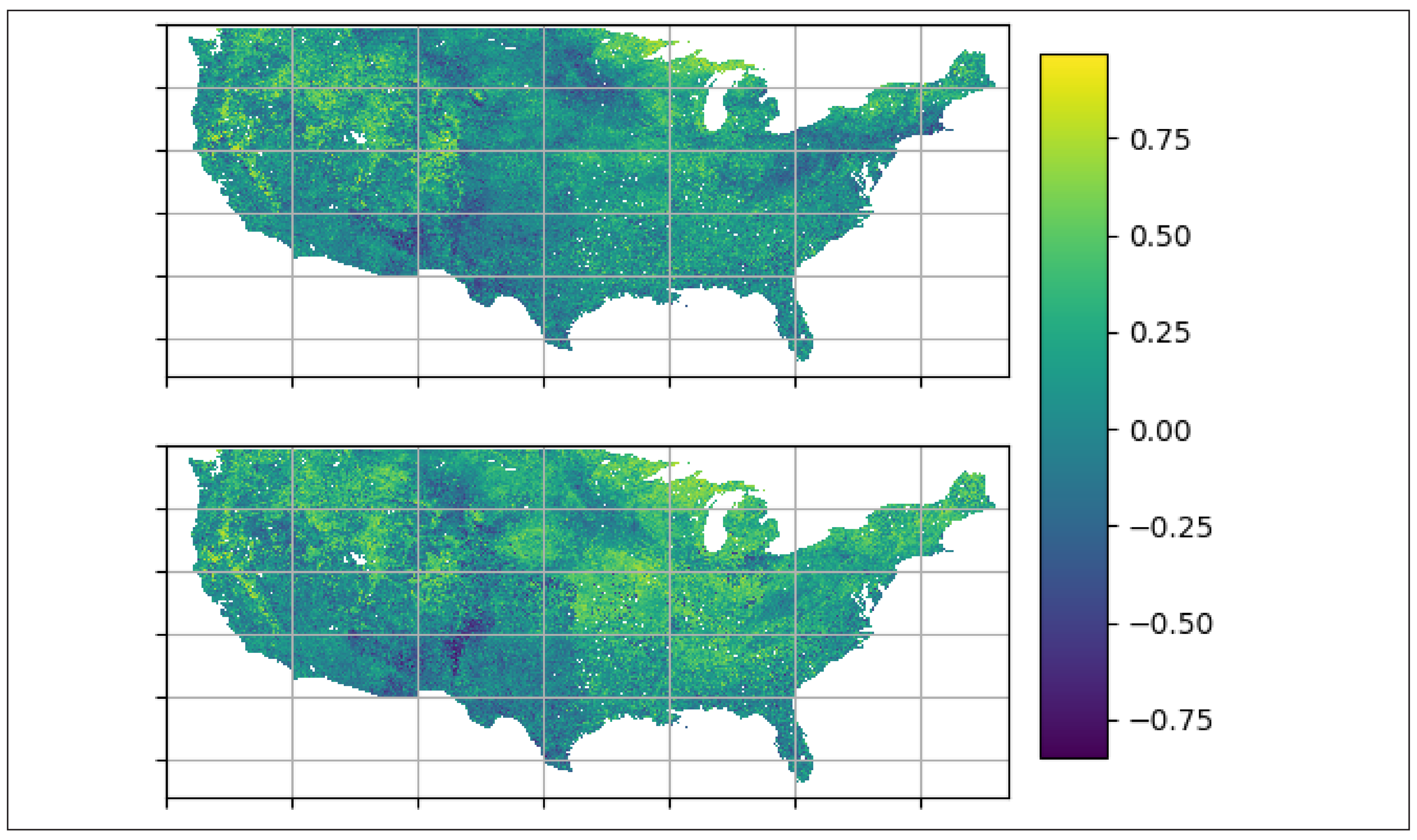

Fig. 3: Correlation between the Leaf index and SOS [Top] and between the Bloom index and SOS [Bottom]

Computational solution

A cloud-based solution based on Apache Spark and its scalable machine learning library MLlib was used to deal with the computational complexity of exploring and analyzing high spatial resolution phenological products. Given the lack of well-tested Spark solutions in the domain of geo-data, a secondary aim of our project is to evaluate the potential of Spark-based solutions to analyze big raster data. Figure 4 illustrates the architecture of our solution, which is organized in three main layers: storage layer, processing layer and JupyterHub services for user-interaction.

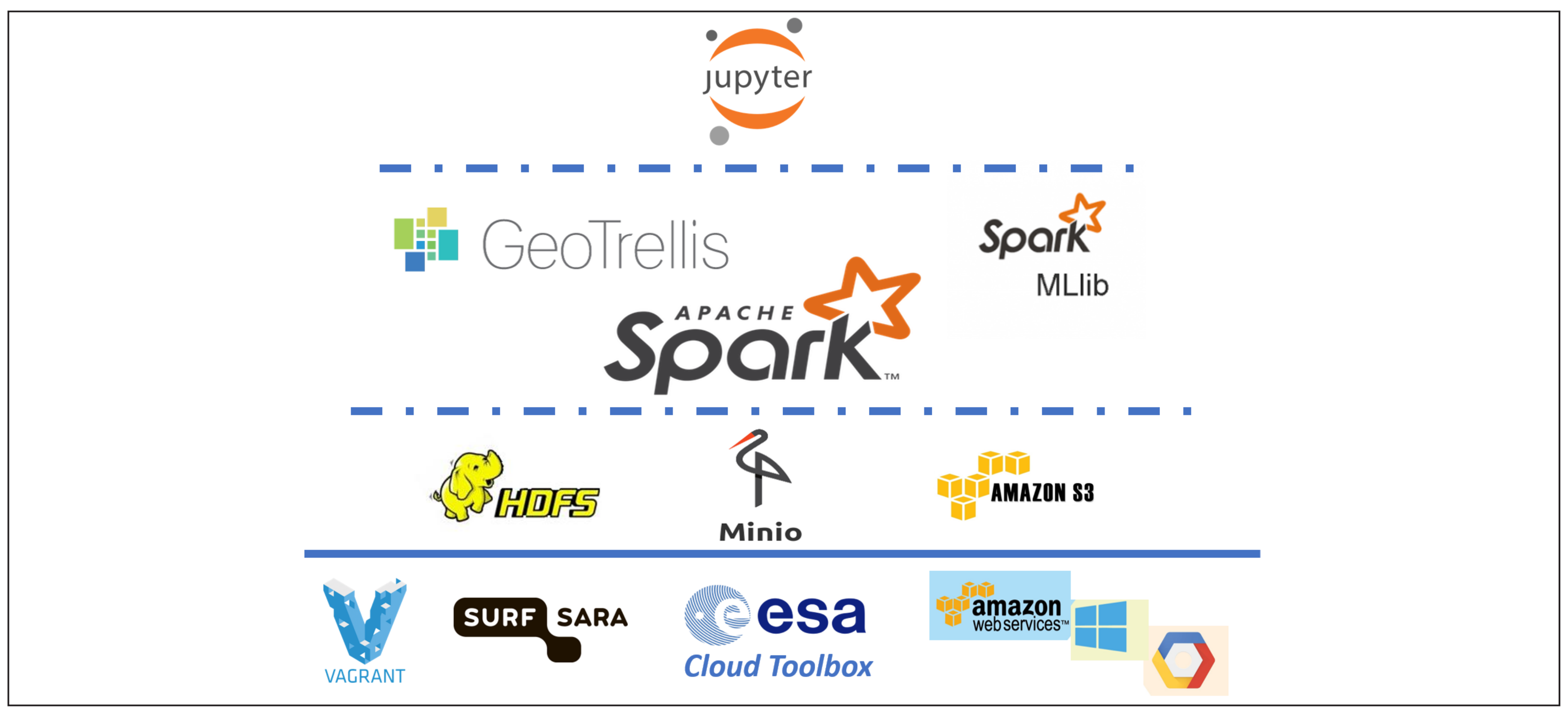

Fig. 4: Computational platform

Conclusions and future work

In this project we exploit the Apache Spark ecosystem for large scale distributed processing. Our experiments demonstrate that it possible to map phenoregions at high spatial resolution and at continental scales. They also show that the SI-x indices are both positively and negatively correlated with the SOS metric. Further analysis is needed to better understand the synergistic value of these two phenological products. In the near future we will:

- Integrate millions of volunteered observations to validate these and other phenological products
- Test the use of co- and tri-clustering methods to analyze "phenological data cubes"
- Analyze co-variability between time series of phenological products
- Deploy our Apache Spark solution in other cloud platforms
- Create new data-driven phenological models using a wide array of environmental datasets

Acknowledgements

This work is supported by the NLeSC Project:

"High spatial resolution phenological modelling at continental scales" (https://github.com/phenology), and it was carried out using the infrastructure provided by SURFSara. The SI-x were computed in the framework of the "Green-wave" project funded by a Google Faculty Award.

UNIVERSITY OF TWENTE.

ITC