Phenology at continental scale: one size does not fit all

Romulo Goncalves, Viktor Bakayov, Raul Zurita-Milla, and Emma Izquierdo-Verdiguier
New era...

• Today: **Data exploration (e-science)**
  • Synthesizing theory, experiment and computation with advanced data management and statistics

Challenges
Phenology at Continental scale...

- **Understanding phenological variability**
  - Studies recurring biological events and variation in space and time

- **Two of the most important sources of spatiotemporal phenological data:**
  - Phenological models based on weather-and/or location related factors
    - Spring Index (SI-x) which is based on temperature
  - Land surface phenological metrics derived from Earth observation sensors
    - Start of Season (SOS) is usually based on NDVI
Motivation

- No universally accepted method to extract phenological metrics from RS images
  - Multiple Vegetation Indices
    - NDVI, EVI, etc.
  - Several software packages
    - TimeSat, Spirits, and Sen2Agri
  - Different phenology extraction methods
    - Fitting functions, spike removal and parameters
One size does not fit all...

- **Vegetation Indices**
  - Normalized Difference Vegetation Index (NDVI)
  - Enhanced Vegetation Index (EVI)

- **TimeSat** ([http://web.nateko.lu.se/timesat/timesat.asp](http://web.nateko.lu.se/timesat/timesat.asp))
  - Fitting functions
    - Asymmetric Gaussian (AG)
    - Savitzky-Golay (SG)
    - Double Logistic (DL)
  - Spike removal

Figure extracted from: [https://slideplayer.com/slide/7971141/](https://slideplayer.com/slide/7971141/)
Part I: Computation Platform
Computational platform

- Cloud-based solution
- Data stored in the original file formats
  - GeoTiff and HDF
  - Accessible via S3 API
- Based on Apache Spark
- Jupyter notebooks
  - Python, R or Scala
Automated and traceable deployment
Distributed SOS computation

- Vegetation Indices computed with Spark
  - Accessible via S3 API

- Spatial data partitioning
  - TimeSat requires the time-series to have minimum length 3 years
  - It is set to find n-1 seasons

- Spark and TimeSat
  - TimeSat requires POSIX file system to read input data and output data
  - The extent and meta-data provided as a "settings file" for TimeSat
  - The computation of SOS was executed in parallel over a series of VMs in the cloud
Part II: Phenology studies
Vegetation Indices

- Dataset provided by Copernicus Global Land Service
  - The product has a spatial resolution of 1km and is available as 10-day composites.

- The NDVI uses the near-infrared and red channels of the sensor:
  \[ \text{NDVI} = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \]

- The EVI also uses the blue channel:
  \[ \text{EVI} = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \rho_{red} - C_2 \rho_{blue} + L} \]
  - Requires setting a set of coefficients \( C_1 = 6, C_2 = 7.5, L = 1, \) and \( G = 2.5 \) [1].

Phenology studies

• Study the validity and coherence of various SOS metrics derived using NDVI and EVI and different fitting functions.
  – Vegetation Indices
    • Normalized Difference Vegetation Index (NDVI)
    • Enhanced Vegetation Index (EVI)
  – Fitting functions
    • Asymmetric Gaussian (AG)
    • Savitzky-Golay (SG)
    • Double Logistic (DL)
Compare SOS products and functions

- NDVI wider range

- Asymmetric Gaussian (AG) fitting function behaves similarly to Double Logistic (DL) fitting function

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI-AG</td>
<td>-10</td>
<td>204</td>
<td>113.8</td>
</tr>
<tr>
<td>NDVI-SG</td>
<td>-1.6</td>
<td>201.77</td>
<td>107.9</td>
</tr>
<tr>
<td>NDVI-DL</td>
<td>-8.92</td>
<td>205.64</td>
<td>113.5</td>
</tr>
<tr>
<td>EVI-AG</td>
<td>13</td>
<td>192</td>
<td>107.2</td>
</tr>
<tr>
<td>EVI-SG</td>
<td>8</td>
<td>186</td>
<td>100.6</td>
</tr>
<tr>
<td>EVI-DL</td>
<td>15</td>
<td>191</td>
<td>107.3</td>
</tr>
</tbody>
</table>
SOS mean value

- Most of the values clustering around the 100\textsuperscript{th} day.
- The histograms are skewed to the left
- In the 150 - 200 days range
  - biggest differences are observed between VIs
  - the NDVI experiments show higher SOS values
Compare SOS products and functions

- Standard deviation (SD) to assess the seasonal spatial change across all the years

- Min and Max SD after the highest and lowest 2% were removed (avoid outliers)

- Between NDVI-AG and EVI-AG the predictions in some cases are near 40 days difference
Ecological regions

- Designated by the U.S. Environmental Protection Agency (EPA) and the Commission for Environmental Cooperation (CEC)

- Level III 120 eco-regions, only 100 visible the others are in Alaska.
Ecological regions

NDVI - AG

EVI - AG
Part III: Lessons and Future work
Lessons and Future Work

- Reduce the interpolation and pre-processing stages
  - Each pre-processing stage removes relevant information

- Additional RS data sets and perform a more detailed study of various other methods to extract land surface phenology metrics

- Data exploration in such multi-dimensional search space is challenging
  - Work with Sentinel-2 from ESA Copernicus program: 10 meters resolution

- Ground-based data
Ground-based data

- **USA- NPN**
  - Between 1995 and 2005 only 50 sites per year
  - From 2005 sharply increased to more than 1400 sites in 2017

- **Issues:**
  - Often collected by volunteers
  - Spatially sparse
Let’s stay in touch

+31 (0)20 460 4770
info@esciencecenter.nl
www.esciencecenter.nl
blog.esciencecenter.nl

Romulo Goncalves
Tech Lead in Data Management
r.goncalves@esciencecenter.nl

netherlands eScience center