netherlands eScience center

Phenology at continental scale: one size does not fit all

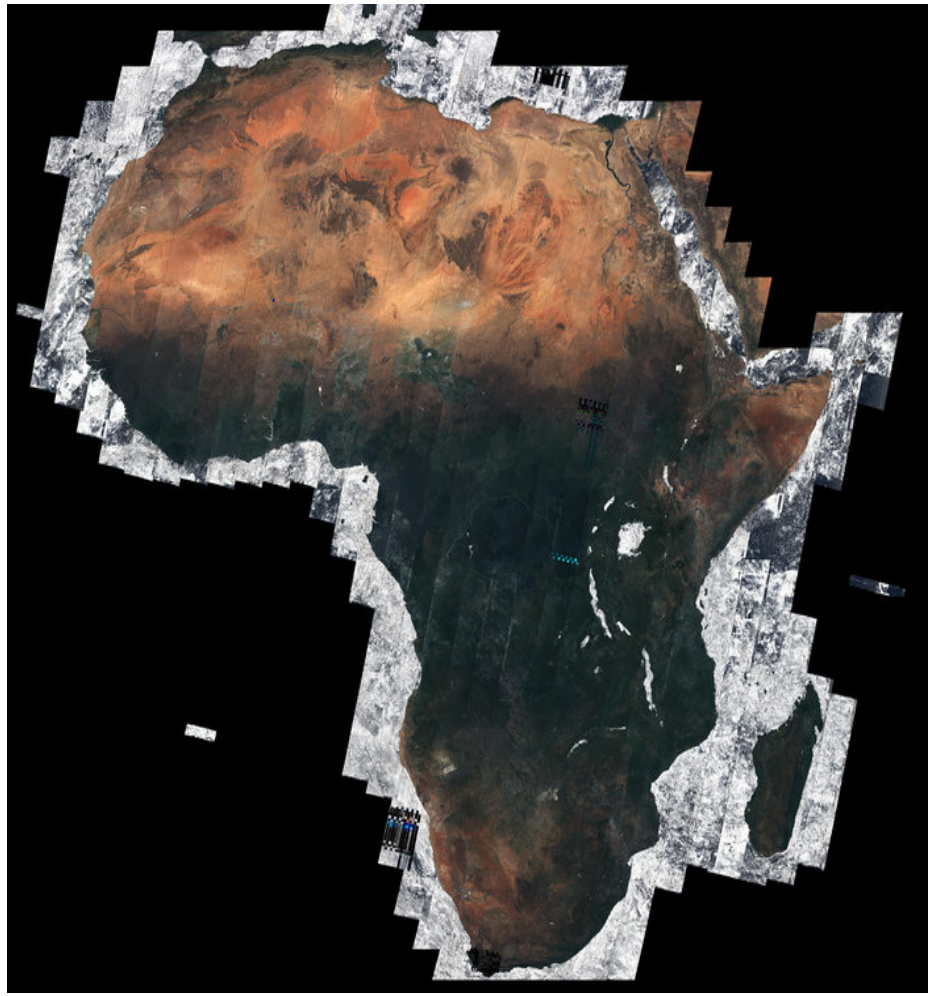
Romulo Goncalves, Viktor Bakayov, Raul Zurita-Milla, and Emma Izquierdo-Verdiguier

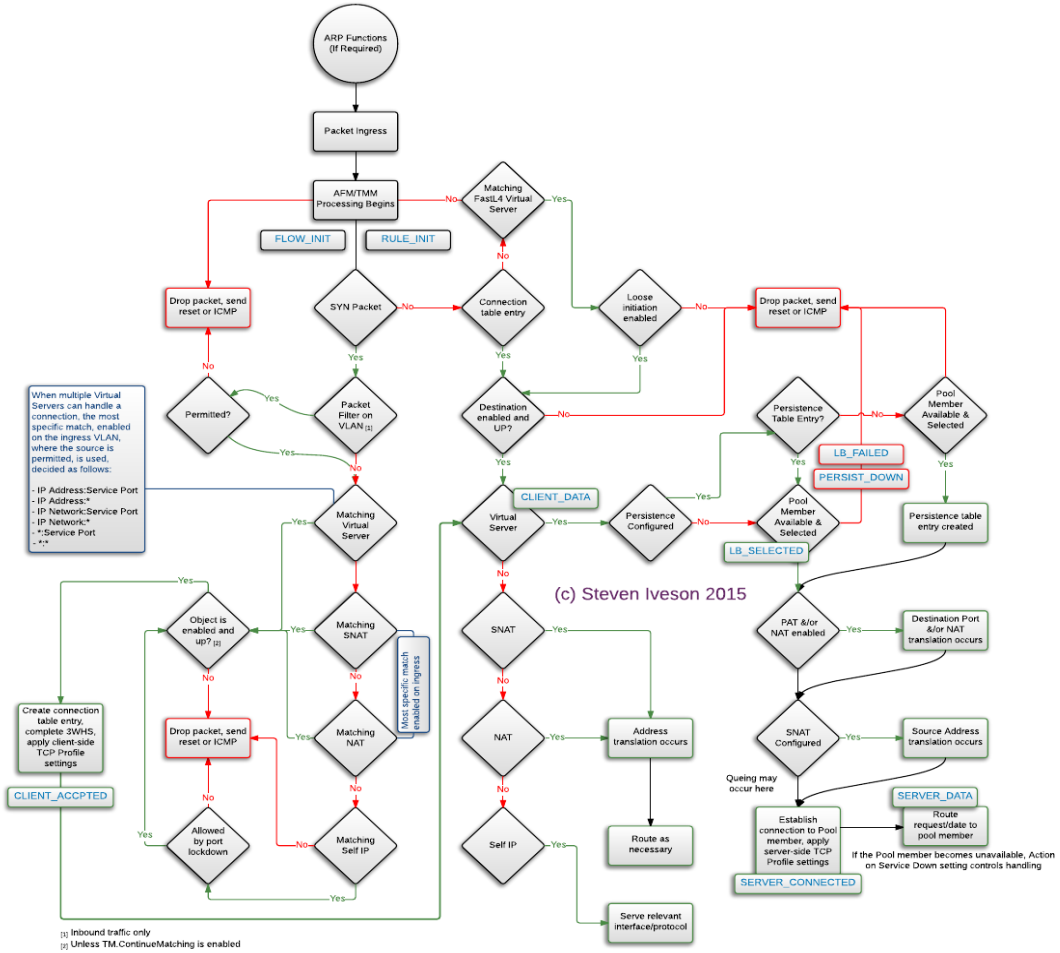
New era…

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

Image from: http://www.czechspaceportal.cz/files/files/images/EO/2012_12_11_15_40_47.jpg

Challenges




ARP Functions (If Required)
Packet Ingress
AFM/TMM Processing Begins
Matching FastL4 Virtual Server
FLOW_INIT
RULE_INIT
Drop packet, send reset or ICMP
SYN Packet
Connection table entry
Loose initiation enabled
Drop packet, send reset or ICMP
Permitted?
Packet Filter on VLAN [1]
When multiple Virtual Servers can handle a connection, the most specific match, enabled on the ingress VLAN, where the source is permitted, is used, decided as follows:
- IP Address:Service Port
- IP Address:*
- IP Network:Service Port
- IP Network:*
- *:Service Port
- *:*
Destination enabled and UP?
Persistence Table Entry?
Pool Member Available & Selected
LB_FAILED
PERSIST_DOWN
Matching Virtual Server
CLIENT_DATA
Virtual Server
Persistence Configured
Pool Member Available & Selected
LB_SELECTED
Persistence table entry created
(c) Steven Iveson 2015
Object is enabled and up? [2]
Matching SNAT
SNAT
PAT &/or NAT enabled
Destination Port &/or NAT translation occurs
Most specific match enabled on ingress
Create connection table entry, complete 3WHS, apply client-side TCP Profile settings
Drop packet, send reset or ICMP
Matching NAT
NAT
Address translation occurs
SNAT Configured
Source Address translation occurs
Queing may occur here
CLIENT_ACCPTED
Allowed by port lockdown
Matching Self IP
Self IP
Route as necessary
Establish connection to Pool member, apply server-side TCP Profile settings
SERVER_DATA
Route request/date to pool member
If the Pool member becomes unavailable, Action on Service Down setting controls handling
SERVER_CONNECTED
Serve relevant interface/protocol
[1] Inbound traffic only
[2] Unless TM.ContinueMatching is enabled

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)
  - Fitting functions
    - Asymetric Gaussian (AG)
    - Savitzky-Golay (SG)
    - Double Logistic (DL)
  - Spike removal

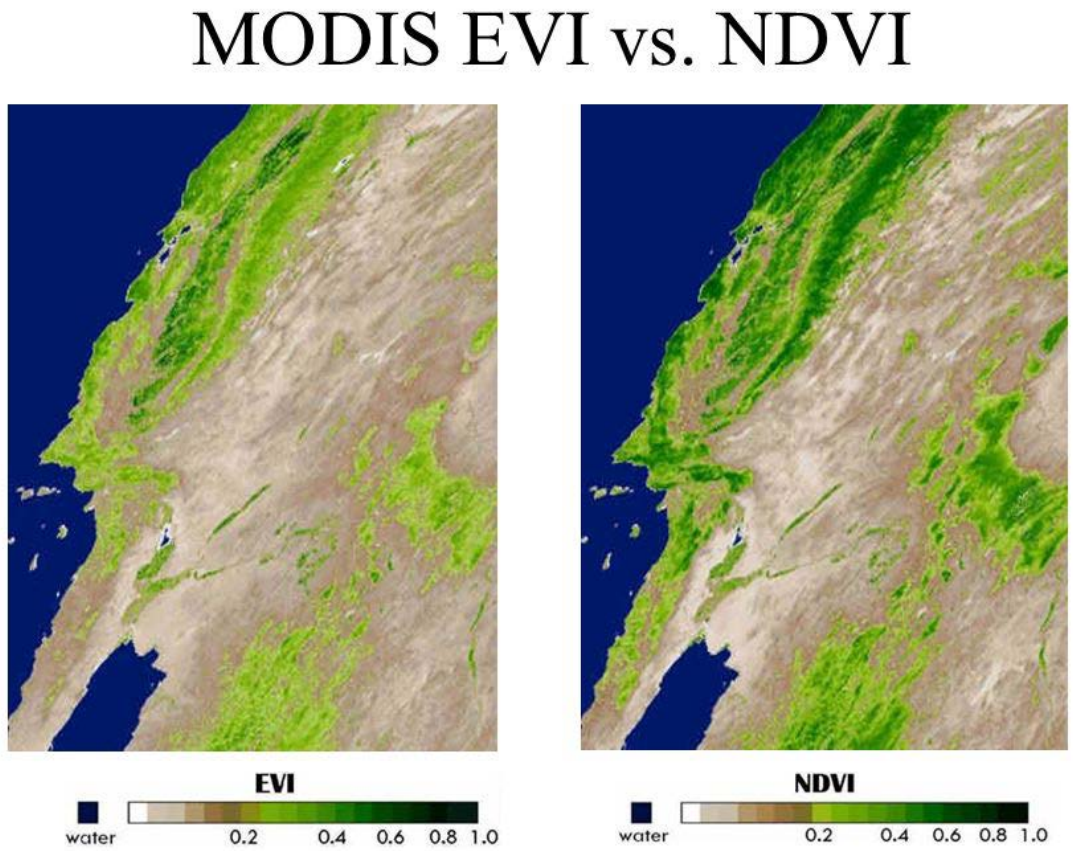


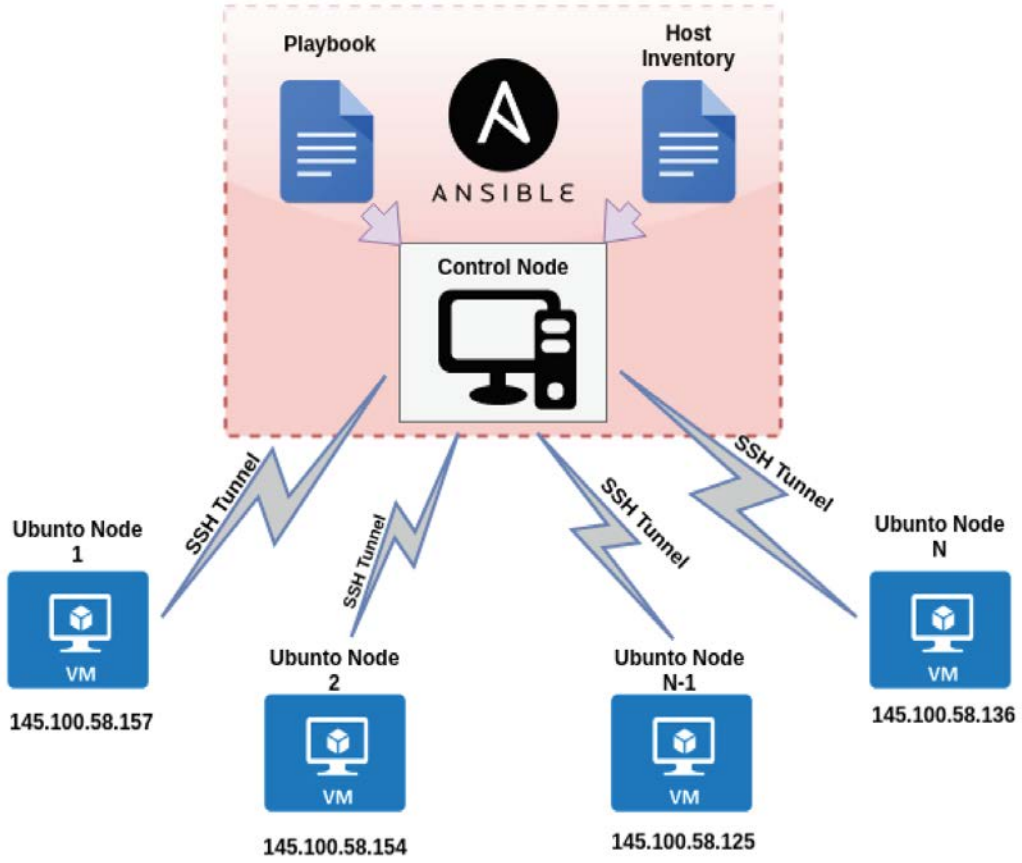
Figure extracted from: https://slideplayer.com/slide/7971141/

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



jupyter

GeoTrellis

Spark MLlib

APACHE Spark

HDFS

Minio

AMAZON S3

VAGRANT

SURF SARA

esa Cloud Toolbox

amazon web services

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
  - Spanned 19 years (1999–2017) by combining SPOT-VEGETATION (1998–2014) and PROBA-V (2014–present) satellite data.
  - 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:

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{nir}} - \rho_{\mathrm{red}}}{\rho_{\mathrm{nir}} + \rho_{\mathrm{red}}}$$

- The EVI also uses the blue channel:

$$EVI = G \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \rho_{red} - C_2 \rho_{blue} + L}$$

  - Requires setting a set of coefficients C1 = 6, C2 = 7.5, L = 1, and G = 2.5 **[1]**.

[1] - A. Verhegghen, S. Bontemps, and P. Defourny, "A global NDVI and EVI reference data set for land-surface phenology using 13 years of daily SPOT-VEGETATION observations," International Journal of Remote Sensing, vol. 35, no. 7, 2014.

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
    - Asymetric Gaussian (AG)
    - Savitzky-Golay (SG)
    - Double Logistic (DL)

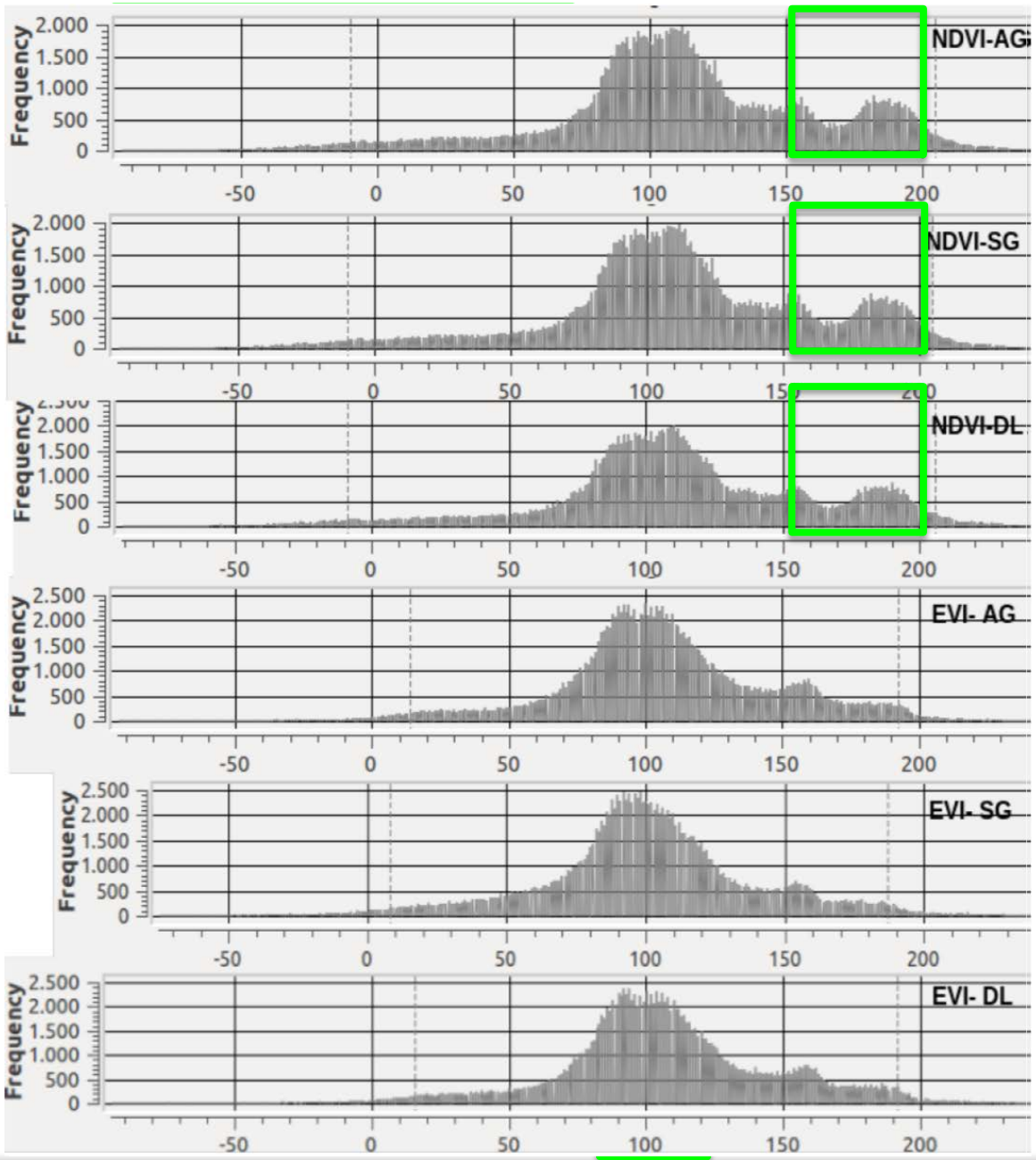
Compare SOS products and functions

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

Experiment	Min	Max	Mean
NDVI- AG	-10	204	113.8
NDVI - SG	-1.6	201.77	107.9
NDVI - DL	-8.92	205.64	113.5
EVI- AG	13	192	107.2
EVI - SG	8	186	100.6
EVI - DL	15	191	107.3

SOS mean value

- Most of the values clustering around the 100th 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

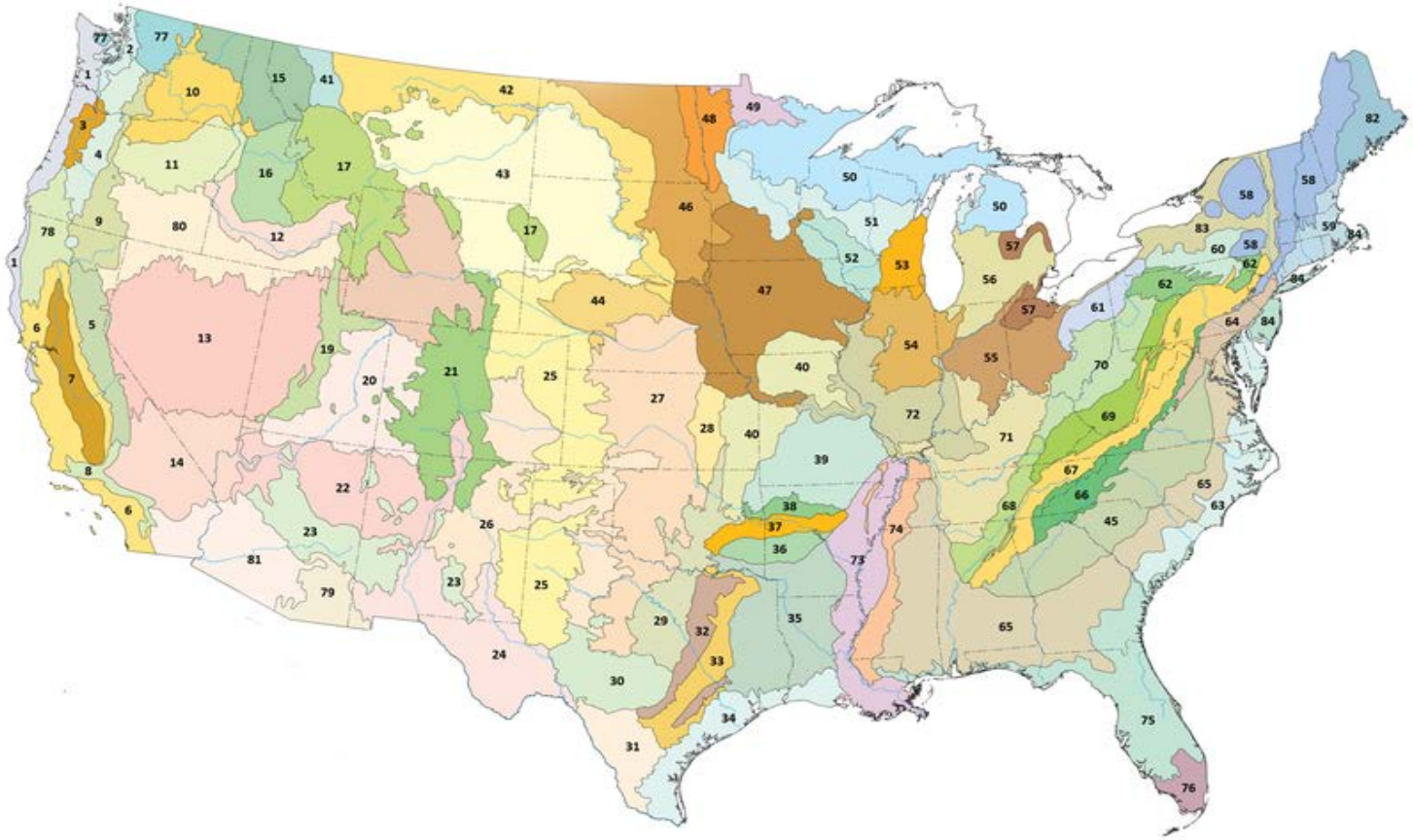
- Standard deviation (SD) to assess the seasonal spatial change across all the years
- Min and Max SD after the highest and lowest 2% were removed (avoid outliers)
- Between NDVI-AG and EVI-AG the predictions in some cases are near 40 days difference

Experiment	Min	Max
NDVI- AG	4.92	137.80
NDVI - SG	6.01	146.77
NDVI - DL	5.04	137.88
EVI- AG	5.07	97.91
EVI - SG	4.98	106.81
EVI - DL	4.95	95.93

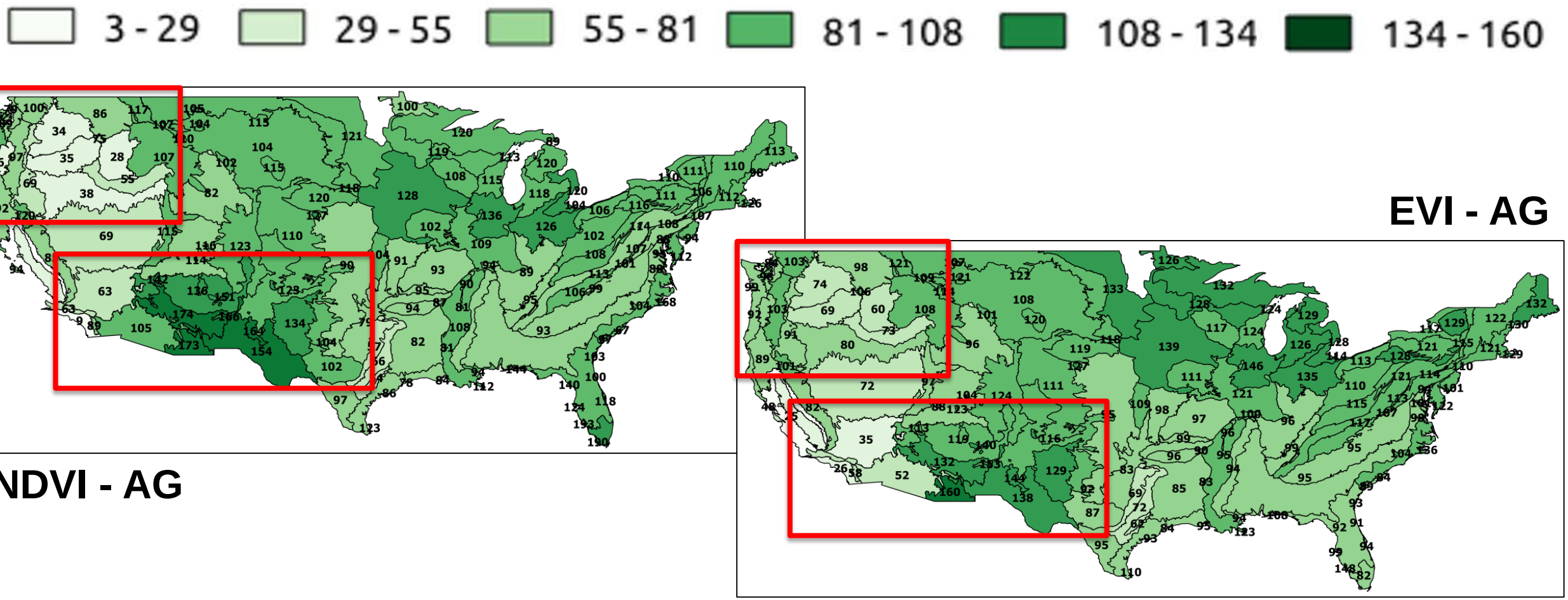
Ecological regions

- 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



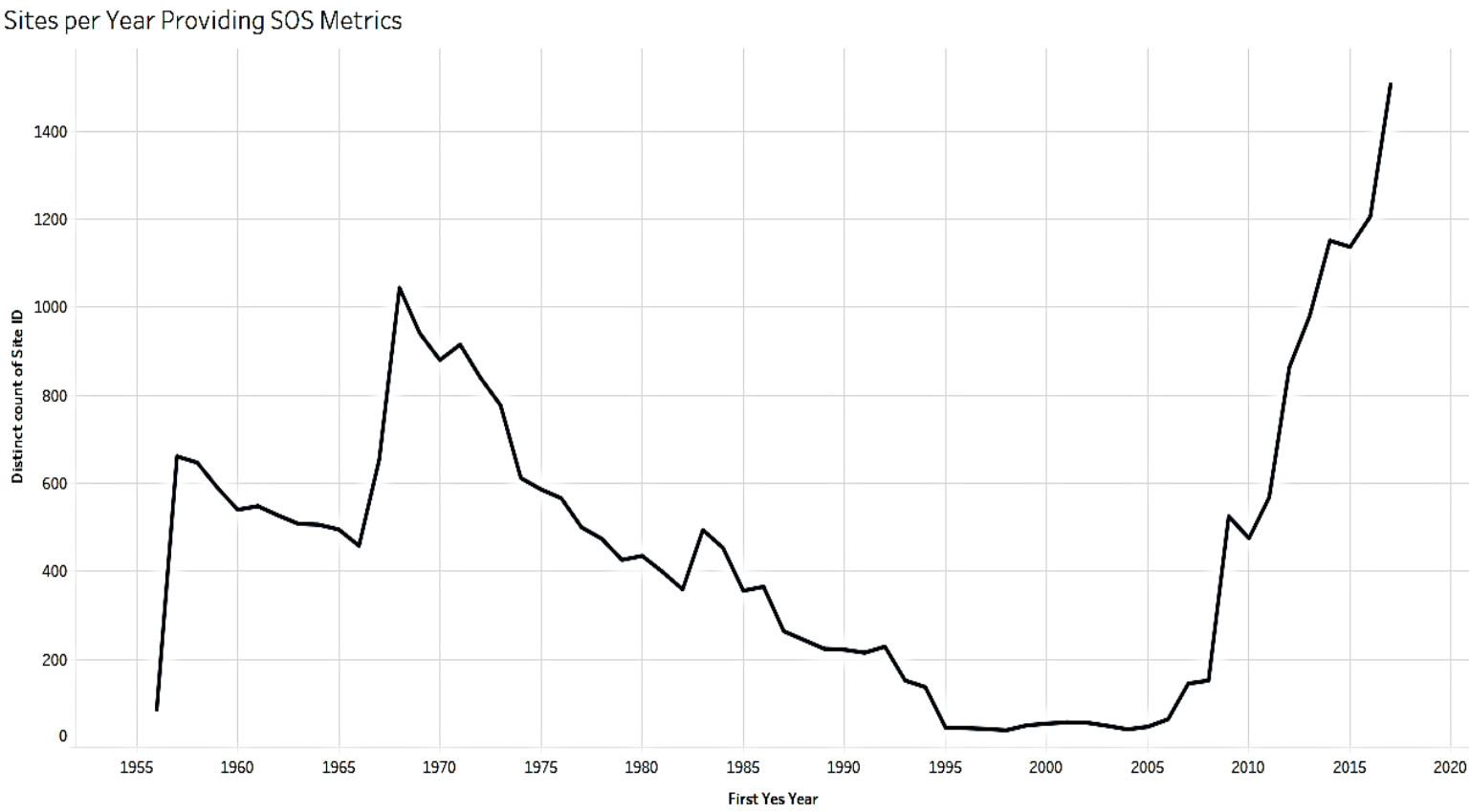
Part III: Lessons and Future work

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

eScienceCenter

eScienceCenter

linkd.in/1j2uS8S

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

netherlands eScience center