ANIMOVE 2024 –
REMOTE SENSING PRACTICAL UNIT

Ines Standfuß, Jakob Schwalb-Willmann, Benjamin Leutner, Martin Wegmann, Martina Scacco, Grégoire Kerr
Matthias Weigand

RS Block: Workflow practical unit

Basics:

Raster data

Vector data

Data Preparation:

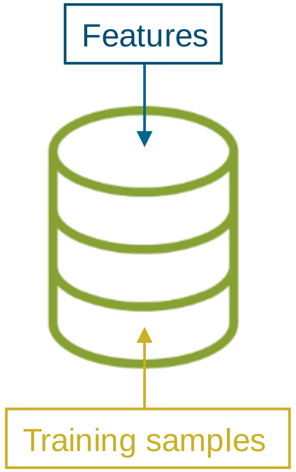
1) RS data acquisition

2) RS data handling

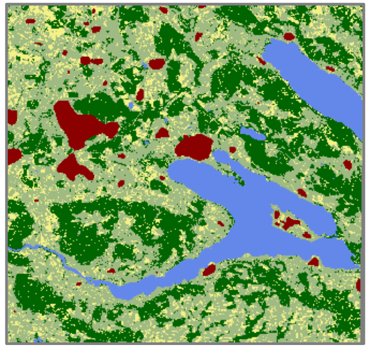
3) Raster manipulation

4) RS data classification:

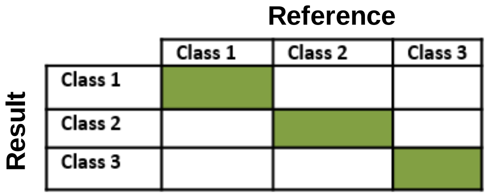
4.1) Training data



4.2) Model training and classification

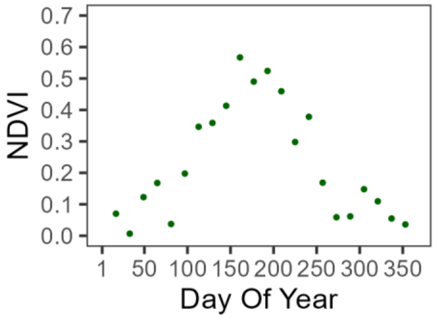


4.2) Accuracy assessment

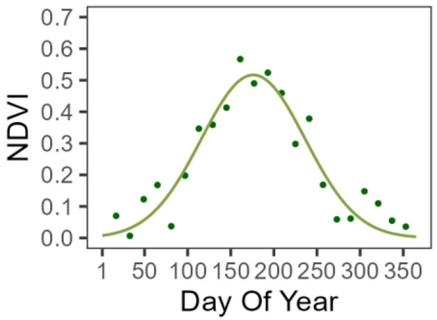


5) Vegetation phenology:

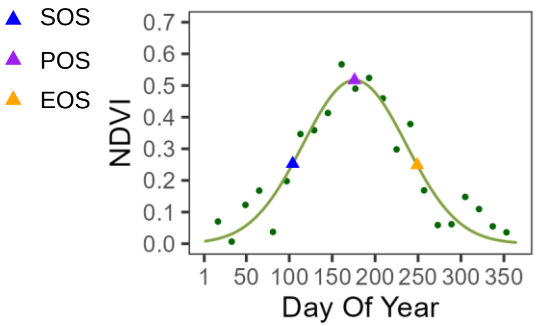
5.1) Input data



5.2) VI curve fitting



5.3) Phenological metrics



RS Block: Workflow practical unit

Basics:

Raster data

Vector data

Data Preparation:

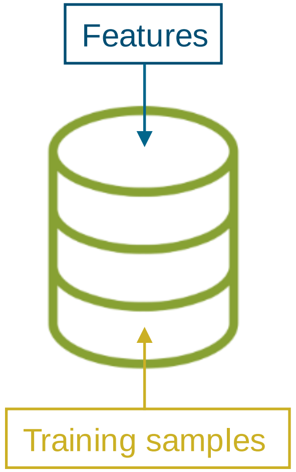
1) RS data acquisition

2) RS data handling

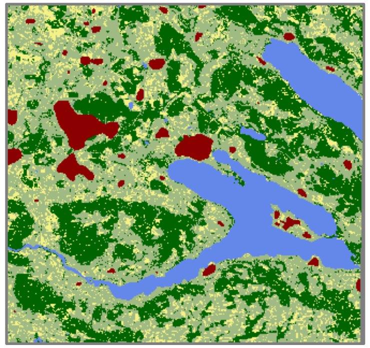
3) Raster manipulation

4) RS data classification:

4.1) Training data



4.2) Model training and classification



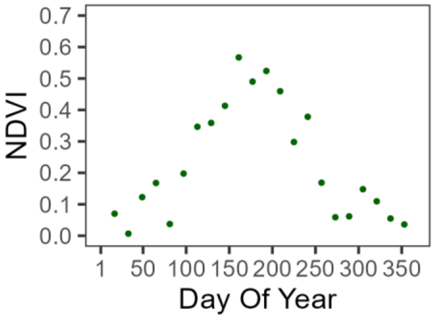
4.2) Accuracy assessment

Reference / Result

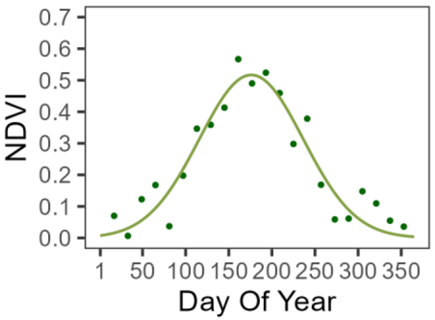
	Class 1	Class 2	Class 3
Class 1			
Class 2			
Class 3			

5) Vegetation phenology:

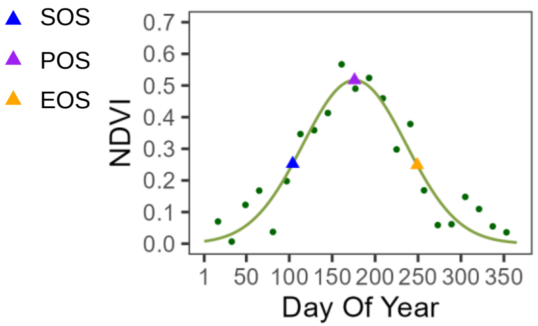
5.1) Input data



5.2) VI curve fitting

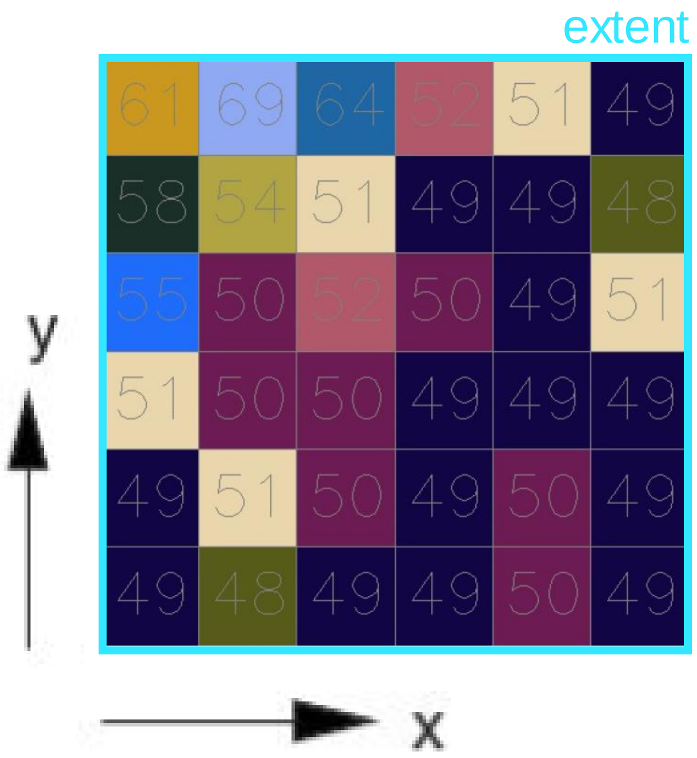


5.3) Phenological metrics



Basics: Raster data

library(terra) to work with raster data (formerly library(raster))



A **raster** is georeferenced data that is represented by a **matrix (grid) of pixels (cells) covering a given extent**.

Each pixel is associated to a longitude (x) and latitude (y).
Each pixel has a certain value which represents a phenomenon/ attribute.

The **smaller the pixel, the higher the resolution** of the raster image, the smaller the features in the landscape that we can detect.

40 cm 1.75 m 10 m

Original Image: Maxar

Existing raster formats:

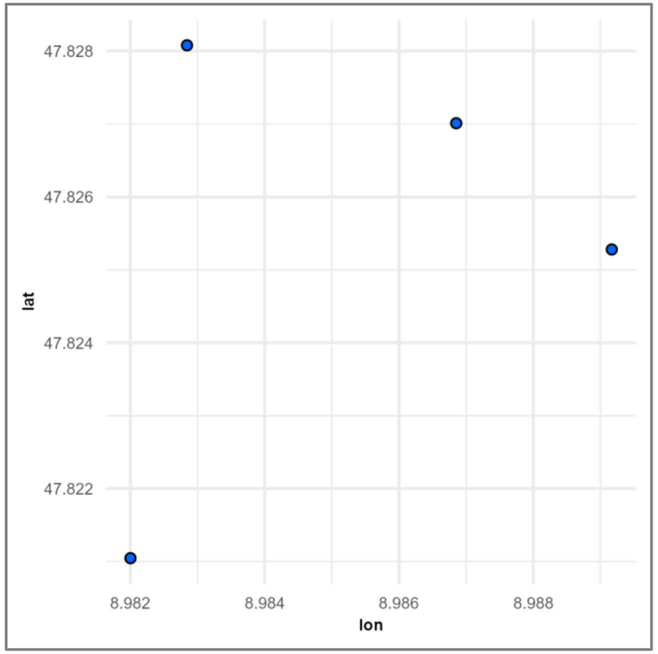
- *.asc
- ***.grd**
- ***.tiff** (GeoTiff, widely used)
- *.hdf

Basics: Vector data

library(sf) to work with vector data

Types of vector data:

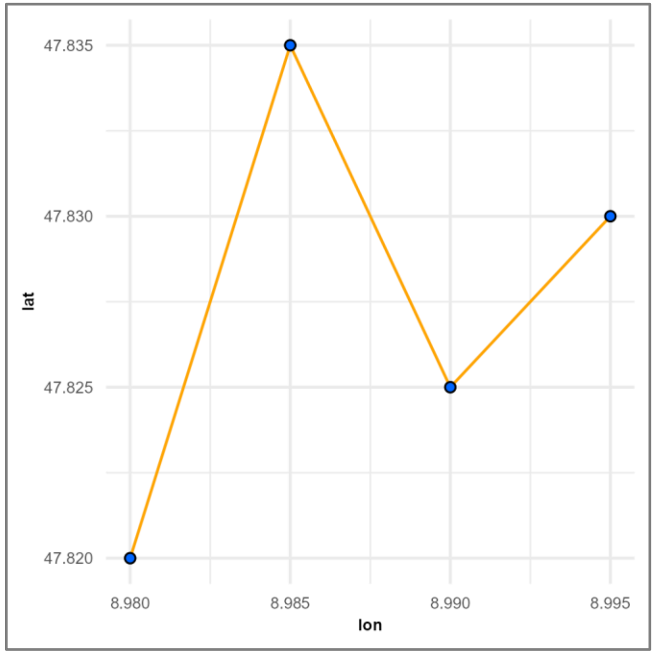
Points



xy-coord, attributes

ID			

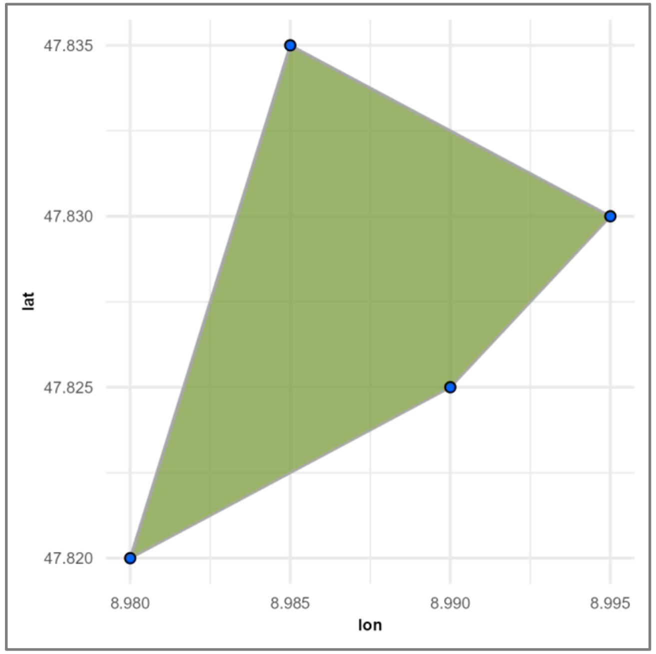
Lines



xy-coord, length, attributes

ID			

Polygons



xy-coord, area, attributes

ID			

Attributes:

Existing vector formats:
- ***.GeoPackage**
- ***.shp**

Spatial **vector data** is a type of geospatial data that is often used to represent specific features on the Earth's surface with **points, lines or polygons**.

RS Block: Workflow practical unit

Basics:

Raster data

Vector data

Data Preparation:

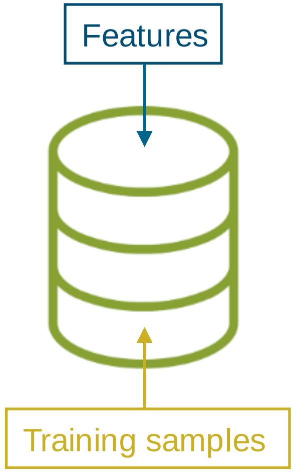
1) RS data acquisition

2) RS data handling

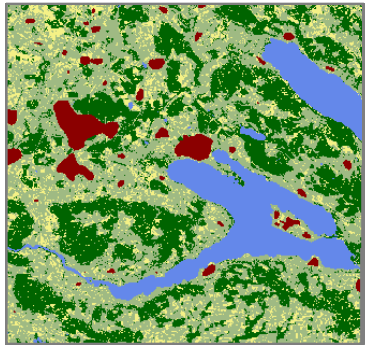
3) Raster manipulation

4) RS data classification:

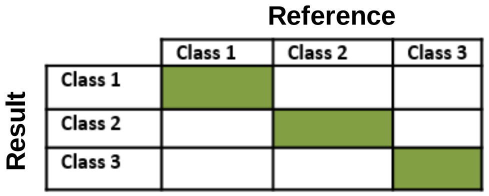
4.1) Training data



4.2) Model training and classification

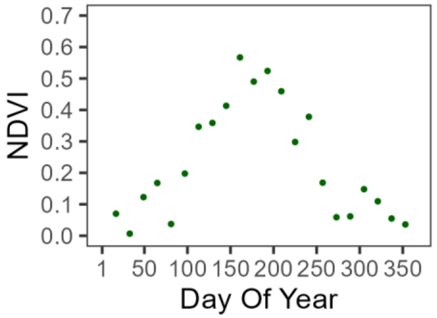


4.2) Accuracy assessment

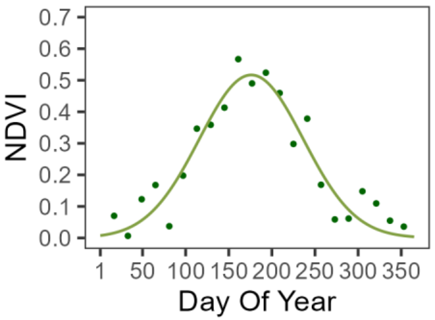


5) Vegetation phenology:

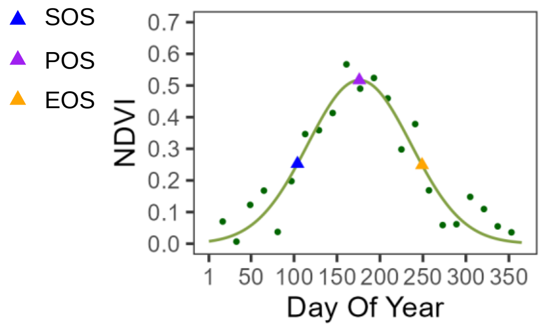
5.1) Input data



5.2) VI curve fitting



5.3) Phenological metrics



1) RS data acquision: Where to get RS data?

Majority of remote sensing data: from publicly funded space missions: open data, free of charge, publicly distributed

Data access:

- at mission providers, e.g. at ESA, NASA, etc.
- at official data and service providers, e.g. at USGS (United States Geological Survey), Copernicus Data Space Ecosystem, etc.
- at commercial platforms for Earth science data and analysis, e.g. Google Earth Engine (GEE), Microsoft Planetary Computer

1) RS data acquision: What do you need?

… at official data and service providers

- **account**

two common ways to download data:

- if using **website GUI**, a compatible browser
- if using an **web API back-end**, a programming language to interact with the service (e.g., R or Python)

- clarity on **which RS data/product** you are looking for
- clarity on what **spatial** (aoi, tile_id) and **temporal** (date, time period) **extent** you want to acquire
- enough **free disk space** to store (and process the data)
- **geospatial software** to view (and process the data)

1) RS data acquision: Download a Sentinel-2 image using Copernicus Data Space website GUI (DEMO)

Go to: https://dataspace.copernicus.eu/

Login (top right):
Username: animove58@gmail.com
PW: AniMove2024#

Data to download:
Data / Level: Sentinel-2 / 2A
Maximum cloud cover: 15 %
Spatial extent: AOI around Möggingen, Germany
Temporal extent: 2023-06-13

Do not download, the dataset we will work with is already stored in your data folder:
'AniMove_2024_RS_block\data\Sentinel_2A\S2B_MSIL2A_20230613T102609_N0509_R108_T32TMT_20230613T151004.SAFE.zip'

Sentinel-2-Level 2A Documentation (p. 497):
Sentinel-2-MSI-L2A-Product-Format-Specifications

1) RS data acquision: First glimpse of the „downloaded“ Sentinel-2 image (Hands-on)

Path to downloaded image:
‘AniMove_2024_RS_block\data\Sentinel_2A\S2B_MSIL2A_20230613T102609_N0509_R108_T32TMT_20230613T151004.SAFE.zip’

Bill Bertram, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

010_RS_data_Sentinel-2_first_glimpse.R

RS Block: Workflow practical unit

Basics:

Raster data

Vector data

Data Preparation:

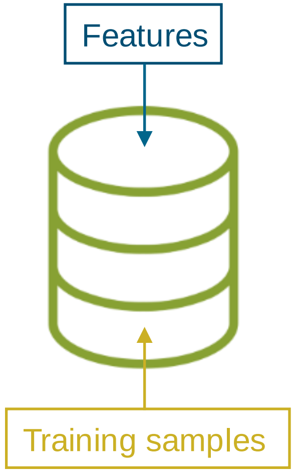
1) RS data acquisition

2) RS data handling

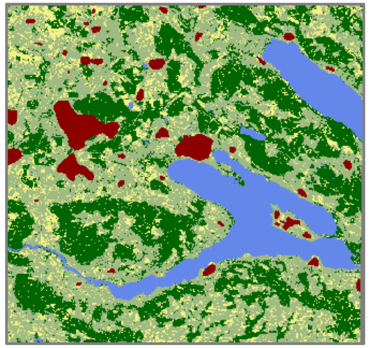
3) Raster manipulation

4) RS data classification:

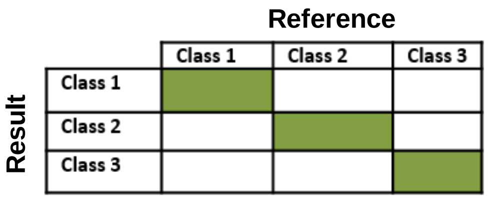
4.1) Training data



4.2) Model training and classification

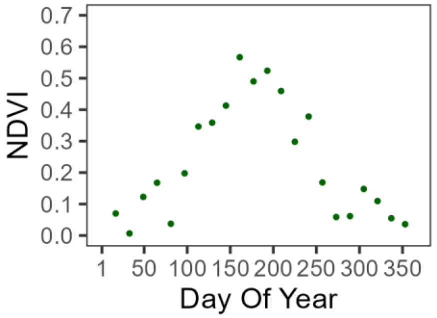


4.2) Accuracy assessment

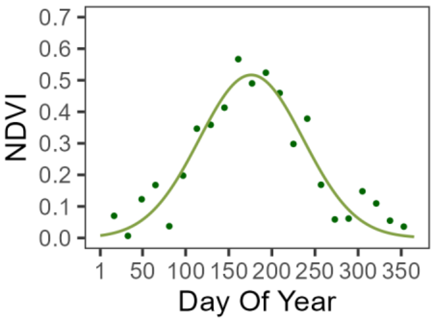


5) Vegetation phenology:

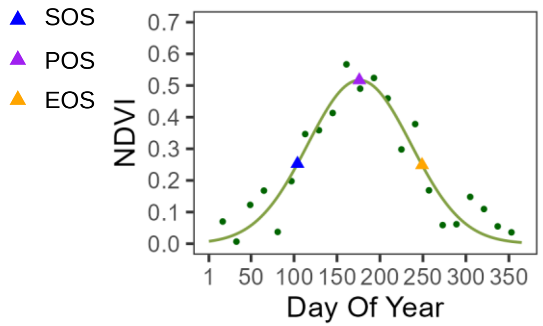
5.1) Input data



5.2) VI curve fitting

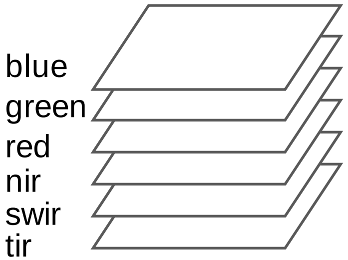


5.3) Phenological metrics



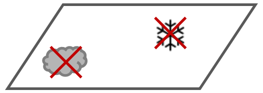
2) RS data handling: Some common data pre-processing/ inspection tasks

- **create an image stack of the band layers you need for a certain task**



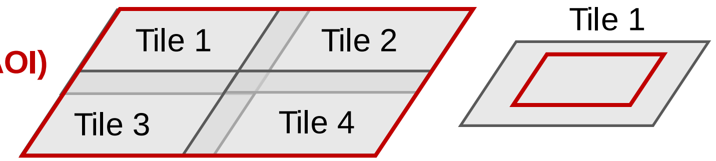
Requires the band layers to have the same:
- coordinate system
- extent
- (resolution (pixel size))

- **mask pixels contamined by clouds/snow or other**



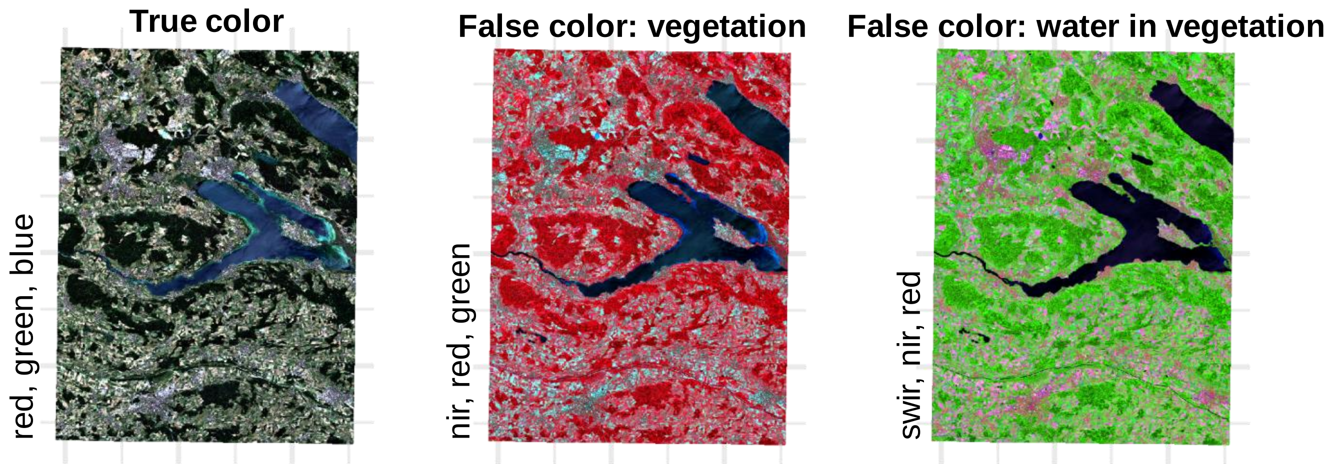
Requires to generate a cloud/snow mask

- **generate an image covering exactly your area of interest (AOI)**



Requires mosaicing/cropping

- **plot different color composites for image inspection**



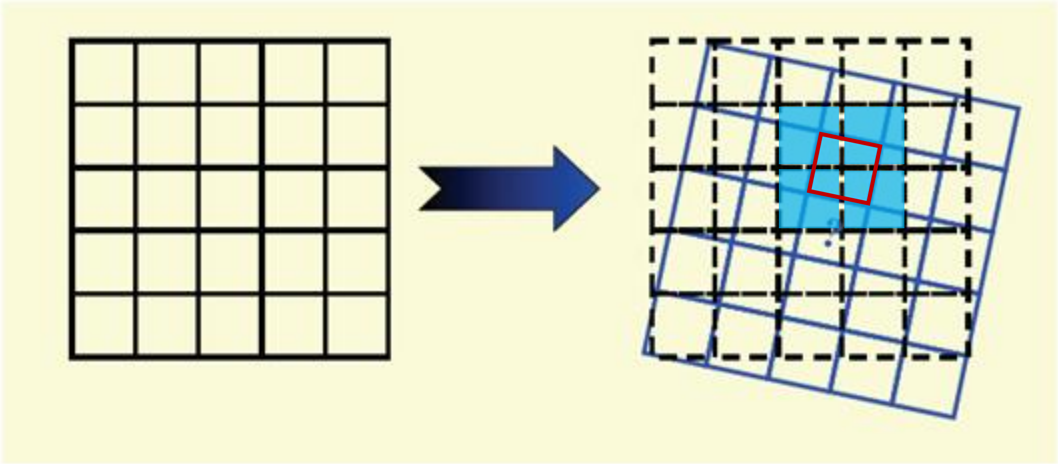
2) RS data handling: Doing some of the common data pre-processing/inspection tasks (Hands-on)

Bill Bertram, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

020_RS_data_handling_Sentinel-2.R

2) RS data handling: Assign pixel values when reprojecting/resampling

Reproject/ resample:



https://www.nicoladeinnocentis.it

near: value of nearest cell

linear: value of 4 nearest cells is used for interpolation

2) RS download and data handling: Using openEO API

openEO API: enables you to access and process earth observation datasets in the ecosystem using intuitive programming libraries.

Collections: access to different satellite data

Processes: user-defined or provided tasks that can be applied to the satellite collections

Can be accessed using ***R***, ***Python*** and ***Javascript***

Here focus on R: https://documentation.dataspace.copernicus.eu/APIs/openEO/R_Client/R.html

2) RS data handling: Downloading and pre-processing Sentinel-2 data using Copernicus Data Space API (optional)

library(openeo) to connect with back-end

Bill Bertram, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

021_aquiring_handling_RS_data_openeo.R

RS Block: Workflow practical unit

Basics:

Raster data

Vector data

Data Preparation:

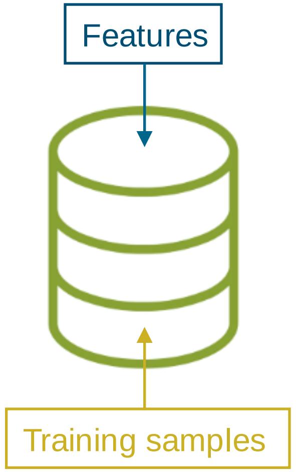
1) RS data acquisition

2) RS data handling

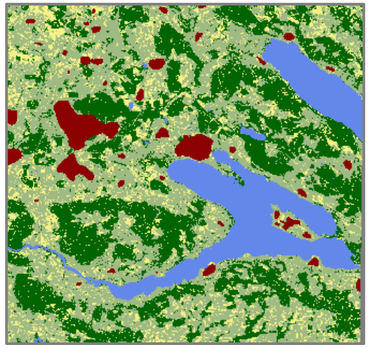
3) Raster manipulation

4) RS data classification:

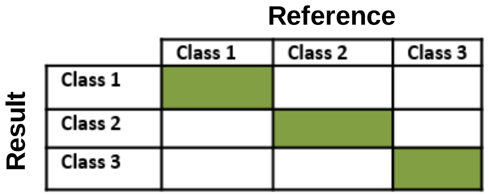
4.1) Training data



4.2) Model training and classification

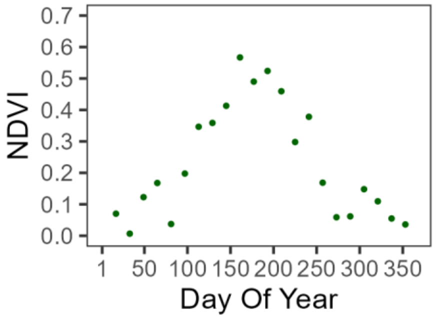


4.2) Accuracy assessment

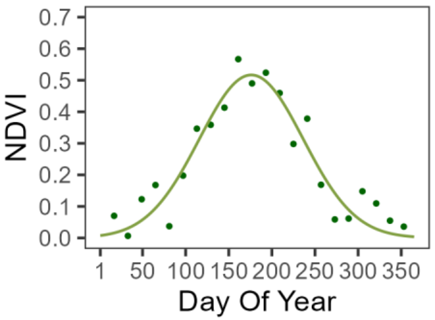


5) Vegetation phenology:

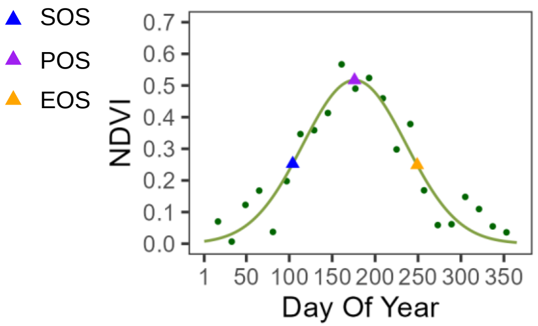
5.1) Input data



5.2) VI curve fitting



5.3) Phenological metrics

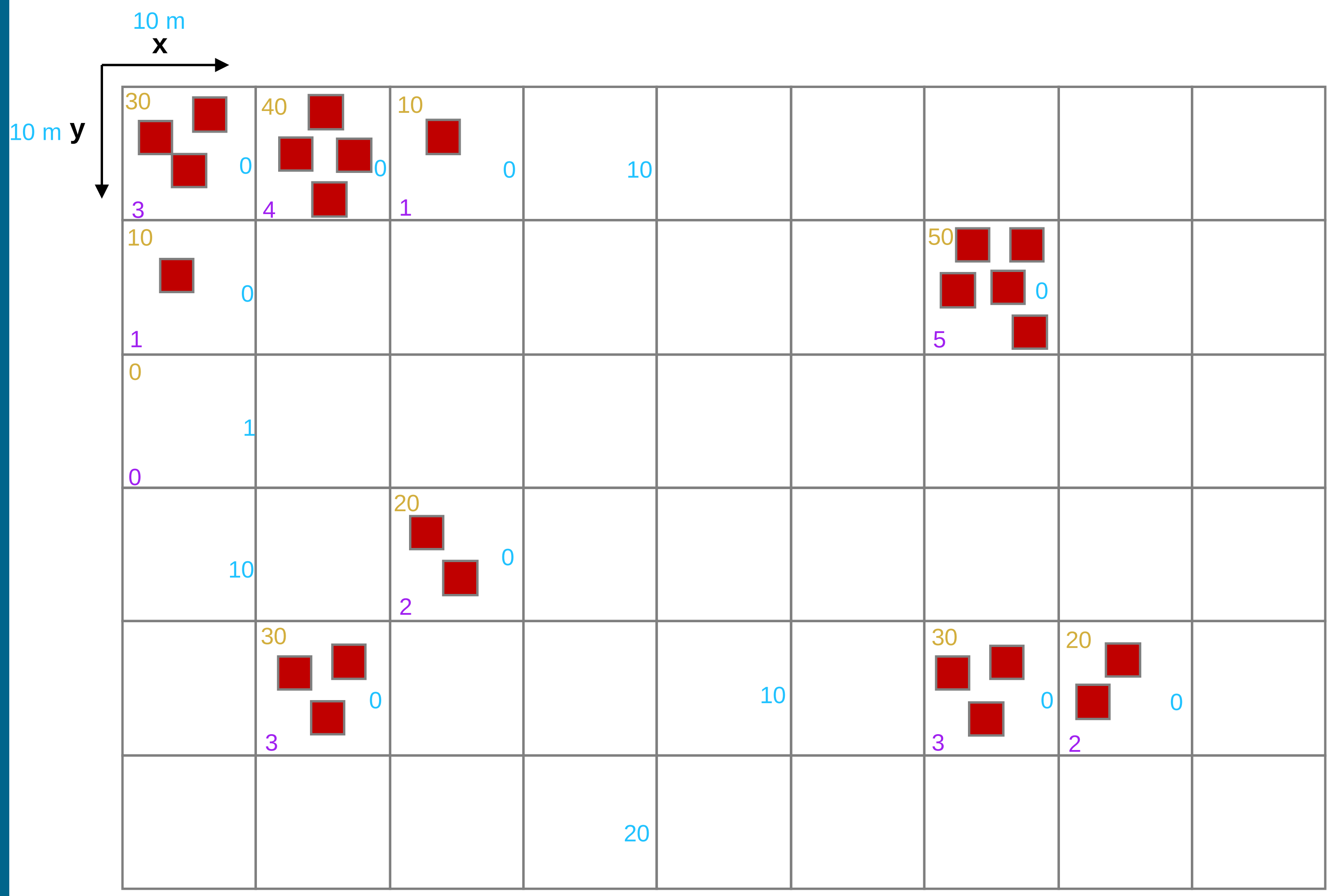


3) Raster manipulation: Generating higher-level information layers

Advanced **raster manipulation** is the process of altering or transforming raster images or rasterizing auxiliary information **to generate higher-level information layers**

- Generate higher-level information layers based on **vector data**
- Generate higher-level information layers using **spatial filtering**
- Generate higher-level information layers using **band ratios**

3) Raster manipulation: Generating higher-level information layers based on vector data



raster layer

building footprints from cadastral data (**vector layer**)

Examples of higher-level information layers:

- count of buildings per pixel
- 2D-building density (%) per pixel
- distance (m) to nearest built-up pixel

3) Raster manipulation: Generating higher-level information layers using spatial filtering

… or bringing in the **neighborhood context**

Spatial filtering modifies values of single pixels based on values of the neighboring pixels

single pixel: vegetation

multiple pixels: urban

Satellite image:
Google Satellite Basemap
QGIS

3) Raster manipulation: Generating higher-level information layers using spatial filtering

filter matrix (kernel)
(often square with odd number of cells)

1	1	1
1	1	1
1	1	1

x raster pixels

Spatial aggregation of raster data through moving window

5	0	10	8	2	9	5	2	5
4	4	6	2	8	0	5	7	5
1	1	3	5	4	9	6	5	5

raster layer

Examples of higher-level information layers:

- smoothed raster layers (*reduced detail*):

  highlight larger, more significant features while suppressing smaller, less significant details

- raster layers with detected edges (*enhanced detail*):

  highlight the boundaries of objects within an image

The specific **function** (smoothing/edge detection) of a **spatial filter** depends on the **values in the filter kernel**!!

Example: mean

	3.77	4.33	5.33	5.22	5.33	5.33	5.00	

new raster layer after spatial filtering

3) Raster manipulation: Generating higher-level information layers using band ratios

Information from **different bands in RS data can also be combined arithmetically** to generate higher-information layers

One way to do so is through **band ratios:**

- **Most simplistic band ratio:** pixel value band A / pixel value band B

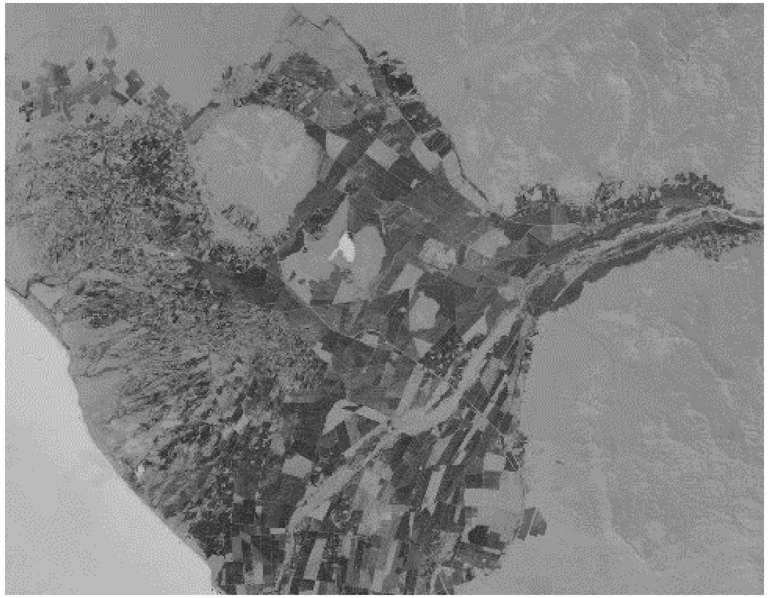
- **Idea behind band ratios**:

  - Objects/Materials on Earth show characteristic reflectance patterns (spectral signatures) in different spectral bands
  - By dividing spectral signatures of one band from another, spectral difference between materials/objects can be enhanced
  - Helps to highlight specific features in an image (reduce atmospheric effects)

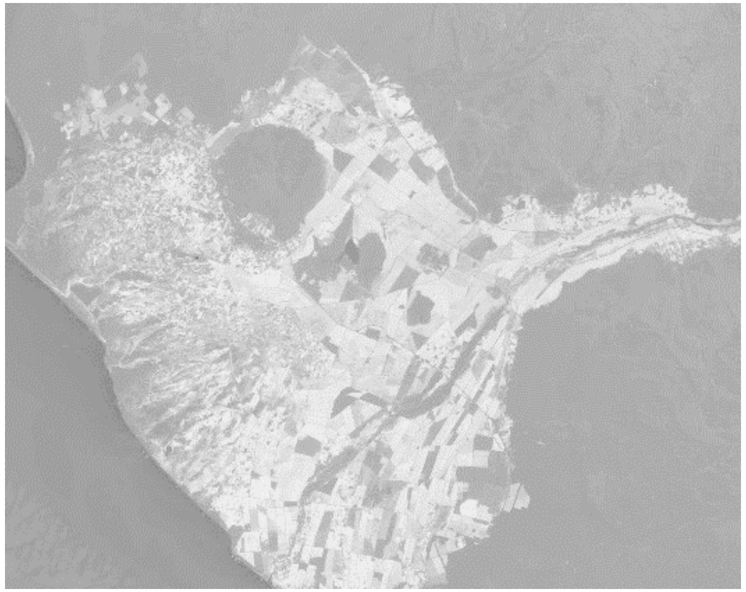
- A **multitude of formulas for band ratios exist** in remote sensing for different application domains

3) Raster manipulation: Generating higher-level information layers using band ratios

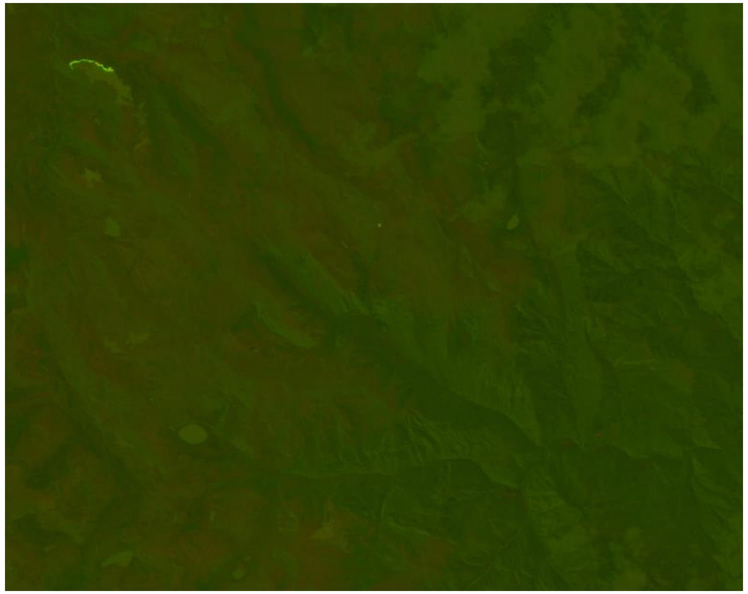
Landsat 9



(B3-B5)/(B3+B5)



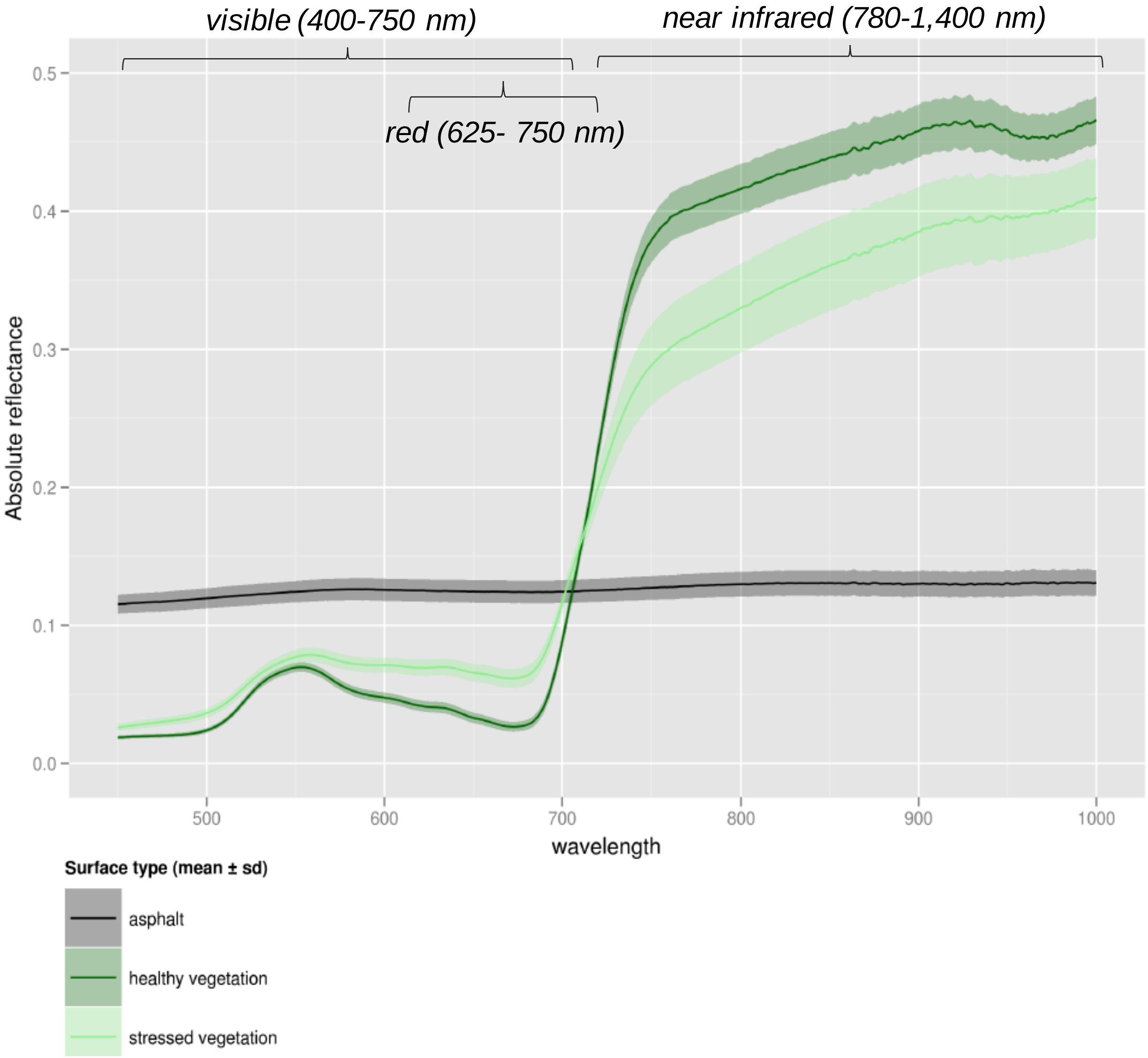
(B5-B4)/(B4+B5)



B4/B3 // B7/B6 // B6/B5

3) Raster manipulation: Generating higher-level information layers using band ratios

Vegatation analysis using band ratios: the popular case of the **Normalized Difference Vegetation Index (NDVI)**



healthy plants:
- absorb high amounts of blue/red radiation for photosynthesis
- reflect high amounts of nir radiation due to high chlorophyll content

stressed (senescent) plants:
- reflect more blue/red radiation
- reflect less nir radiation

$$NDVI = \frac{(nir - red)}{(nir + red)}$$ NDVI value range: -1 to 1

values > 1 are typically associated with vegetation

$$NDVI_{healthy\ plant} = \frac{(0.4 - 0.02)}{(0.4+0.02)} = 0.91$$

$$NDVI_{stressed\ plant} = \frac{(0.35 - 0.07)}{(0.35+0.07)} = 0.67$$

values around 0 are usually an indication for non-vegetated materials such as bare soil, concrete, or snow

$$NDVI_{asphalt} = \frac{(0.12 - 0.11)}{(0.12+0.11)} = 0.04$$

negative values are often found over water

3) Raster data manipulation: vector-based, spatial filtering and band ratios (Hands-on)

030_raster_manipulation.R

3) Raster data manipulation: Calculating NDVI layer using Copernicus Data Space API (optional)

library(openeo) to connect with back-end

Bill Bertram, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

031_raster_manipulation_NDVI_openeo_Radolfzell_Germany.R

RS Block: Workflow practical unit

Basics:

Raster data

Vector data

Data Preparation:

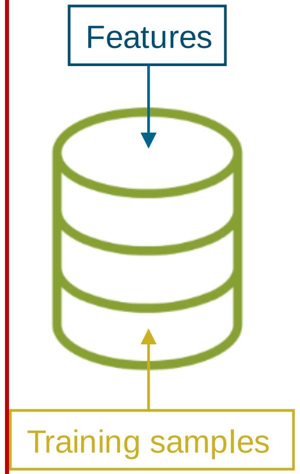
1) RS data acquisition

2) RS data handling

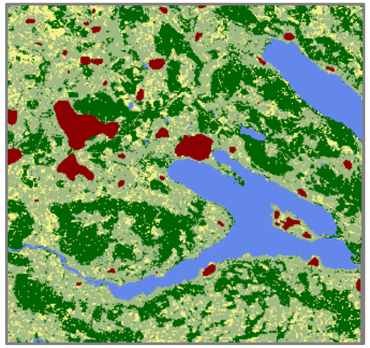
3) Raster manipulation

4) RS data classification:

4.1) Training data



4.2) Model training and classification



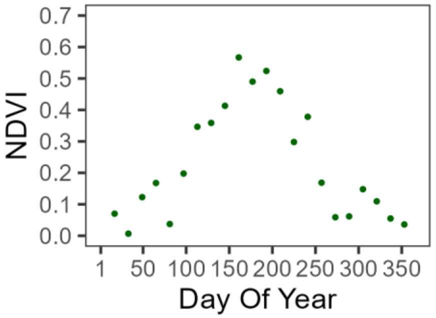
4.2) Accuracy assessment

Reference

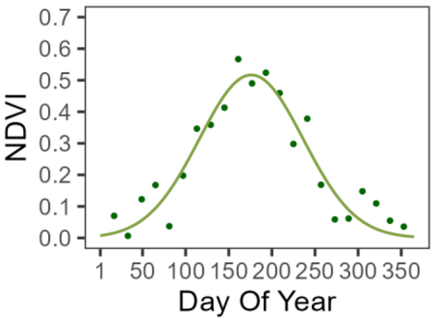
Result	Class 1	Class 2	Class 3
Class 1			
Class 2			
Class 3			

5) Vegetation phenology:

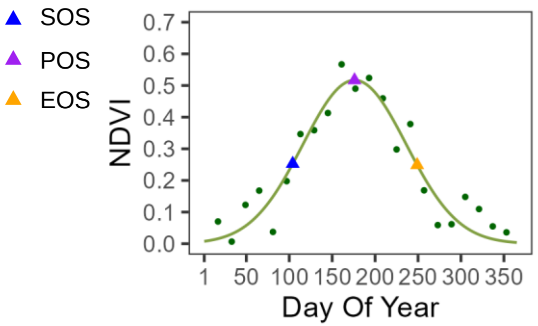
5.1) Input data



5.2) VI curve fitting



5.3) Phenological metrics

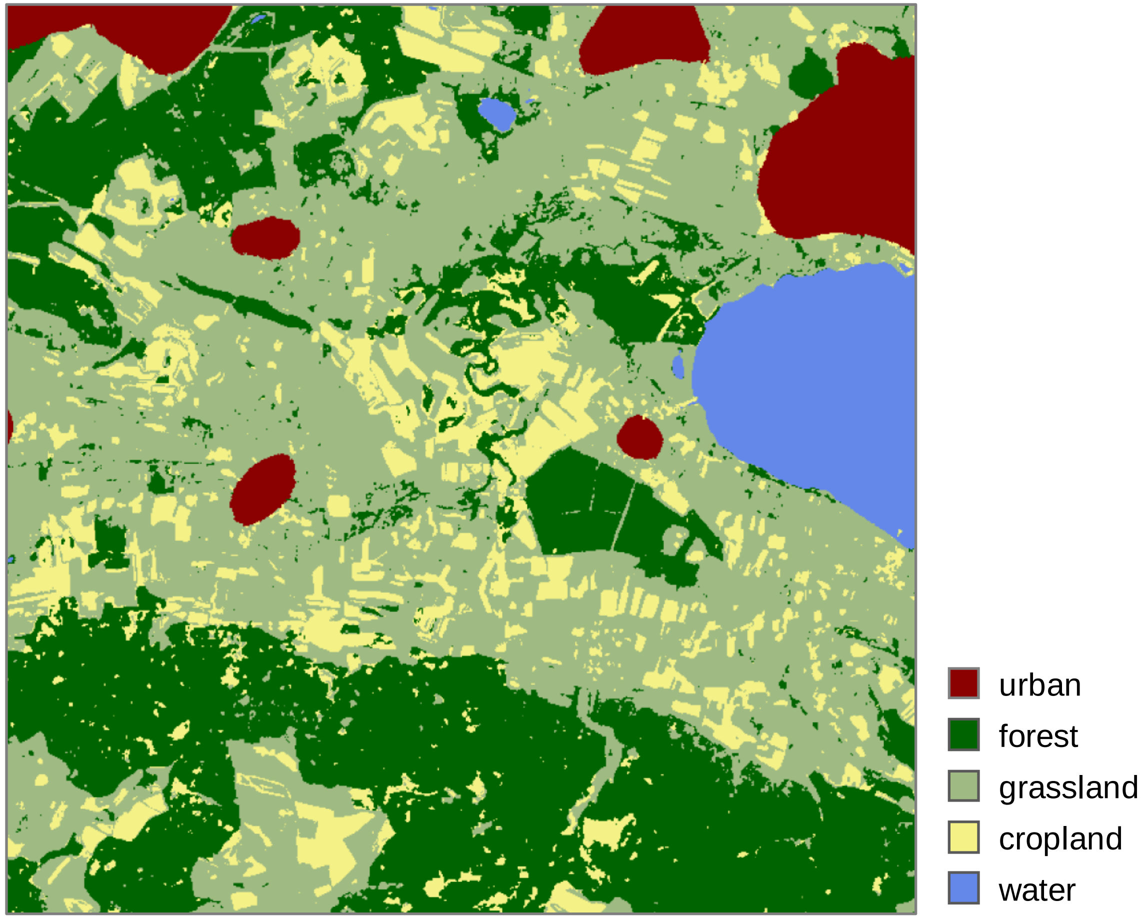


4) RS data classification: Basics

In RS, common scope of classification tasks is primarily aimed at the **differentiation of land cover (LC) types,** also known as **LC classification**

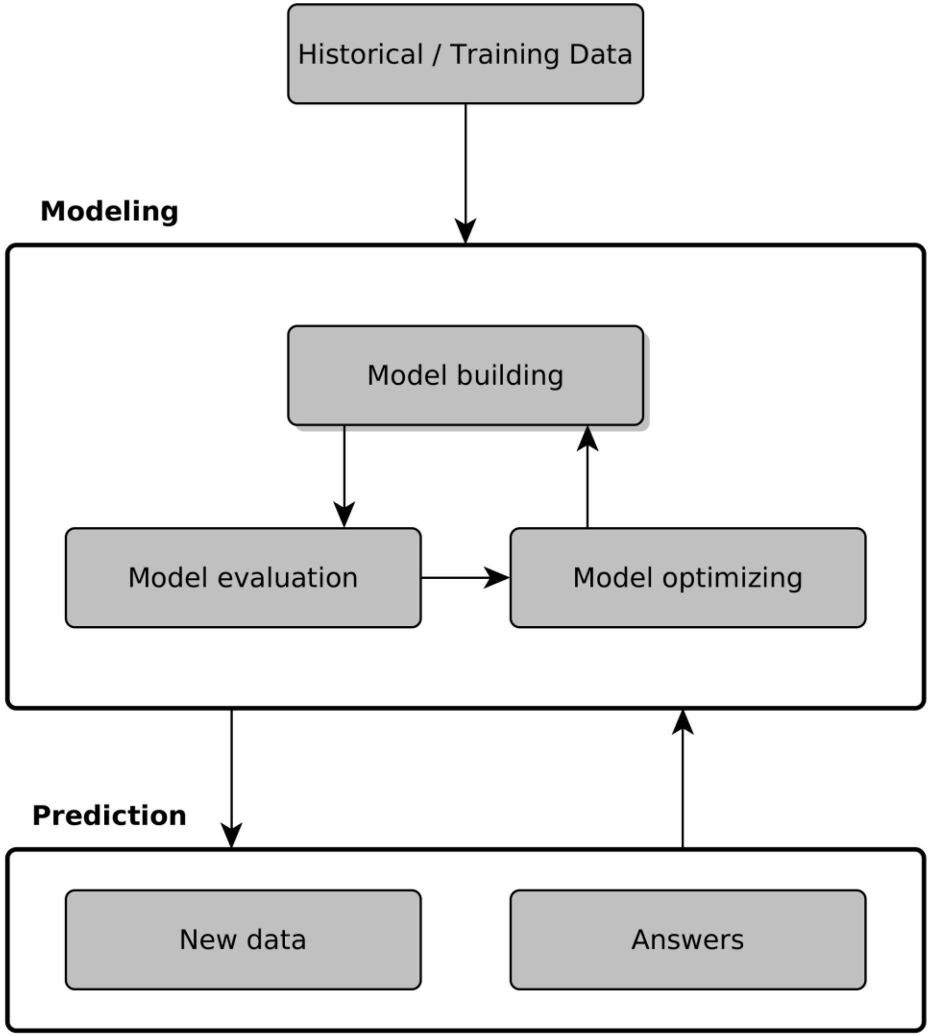
LC classification of RS data is performed to **enhance the level of information** and to **reduce the complexity**

RS data/
Information
layers



4) RS data classification: Classification procedure and approaches

At the very basis: classification algorithms learn from **input data** to predict some **outcome values**



unsupervised classification:

- **Training data:** features without class label
- **Modeling:** clustering algorithm, e.g. kmeans
- **Prediction:** unlabeled classification

supervised classification:

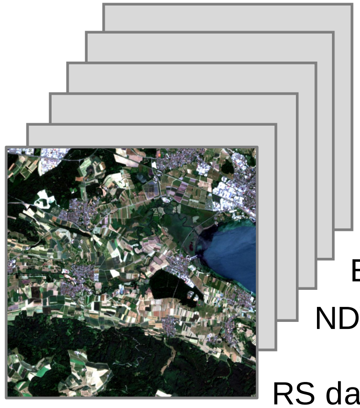
- **Training data:** features with class label
- **Modeling:** statistical/learning algorithm, e.g. SVM, Random Forest
- **Prediction:** labeled classification

4) RS data classification: Supervised classification

Step 1: Training data generation

Step 2: Model training

1.1) Features (data cube)



Smoothed layer
Edge layer
NDVI layer
RS data layers

1.2) Training samples

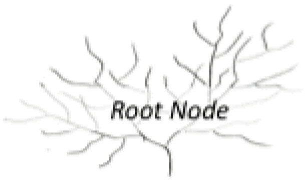
- urban
- forest
- grassland
- cropland
- water

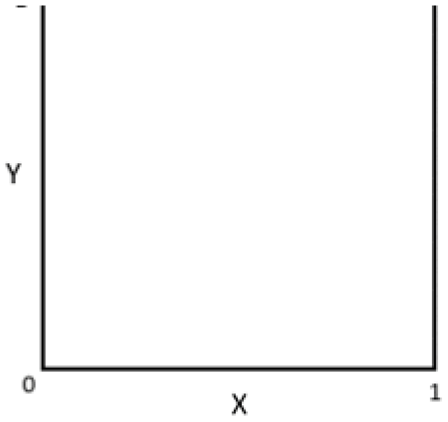
Class	Feature 1	Feature …
urban		
water		
water		
forest		
…		

Random Forest

4) RS data classification: Random Forest (RF) basics

- Machine learning algorithm introduced by Breiman (2001)
- Based on *Classification and Regression Trees* (CART), introduced by Breiman et al. 1984





Binary decision trees
Searches for a sequence of binary splits in the data:
"Is variable (x) lower or greater than a certain value"

- RF uses **random** samples to create an **ensemble of CART**
- *Bagging* and *Boosting*
- *Bagging*: random samples of the training data are used to create multiple trees
- *Boosting*: at each split (node) only a sample of features is used to find the best split

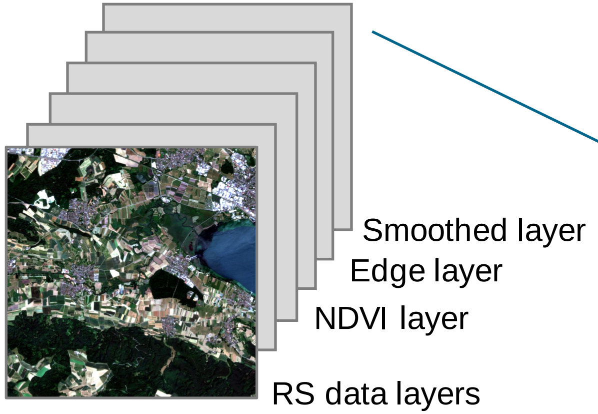
4) RS data classification: Random Forest (RF) Advantages

- By this design, RF is **less likely to overfit**
- **Easy to implement,** no complex tuning methods are needed
- Widely used with **good performance in classification tasks**

4) RS data classification: Supervised classification

Step 1: Training data generation

1.1) Features (data cube)



1.2) Training samples

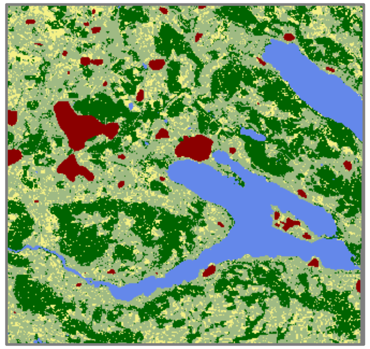
- urban
- forest
- grassland
- cropland
- water

Class	Feature 1	Feature …
urban		
water		
water		
forest		
…		

Step 2: Model training

Random Forest

Step 3: Prediction



Step 4: Accuracy assessment

Reference (columns) / **Result** (rows)

	Class 1	Class 2	Class 3
Class 1			
Class 2			
Class 3			

4) RS data classification: Accuracy assessment

Why is and accuracy assessment of classification products **important**?

- Classification products are, amongst others, used as input layer in scientific studies, e.g. movement analysis/ species-environment relationships
- **Reliable studies** require **accurate maps**
- Hence, it is necessary to determine the quality of a classification product

The **sources of errors in classification products** are manifold:

- Geometric error
- Atmospheric infuences
- Incorrectly labeled training data
- Insufficient training data
- Undistinguishable classes
- Errors introduced by classification algorithm

4) RS data classification: Accuracy assessment

How can the accuracy of a classification product be assessed?

- The **classification** (prediction) is compared against **reference data**, i.e. "ground truth“
- The goal is the quantitative identification and measurement of map errors
- The accuracy can be described for the **overall** result as well as for **individual classes**.

The error (confusion) matrix: Expresses the relationship between reference observations and their predicted class

- Basis to calculate various **overall** and **class-based** accuracy measures:

overall

$$OA = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

class-based

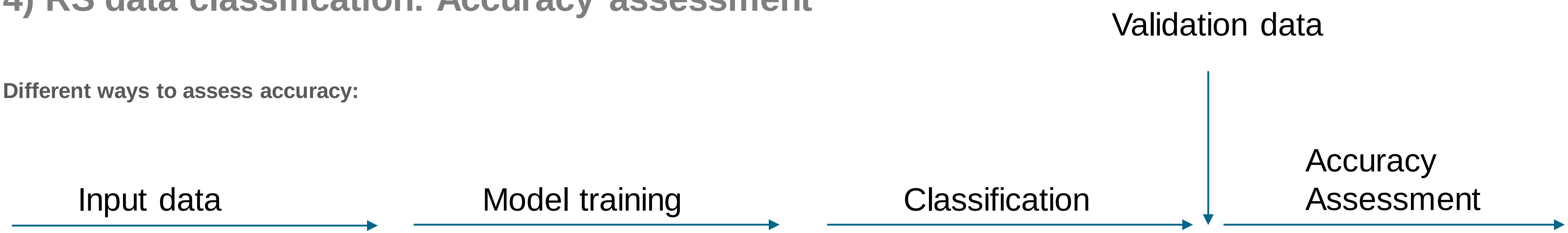
$$Recall = \frac{TP_{class}}{(TP_{class} + FN_{class})}$$

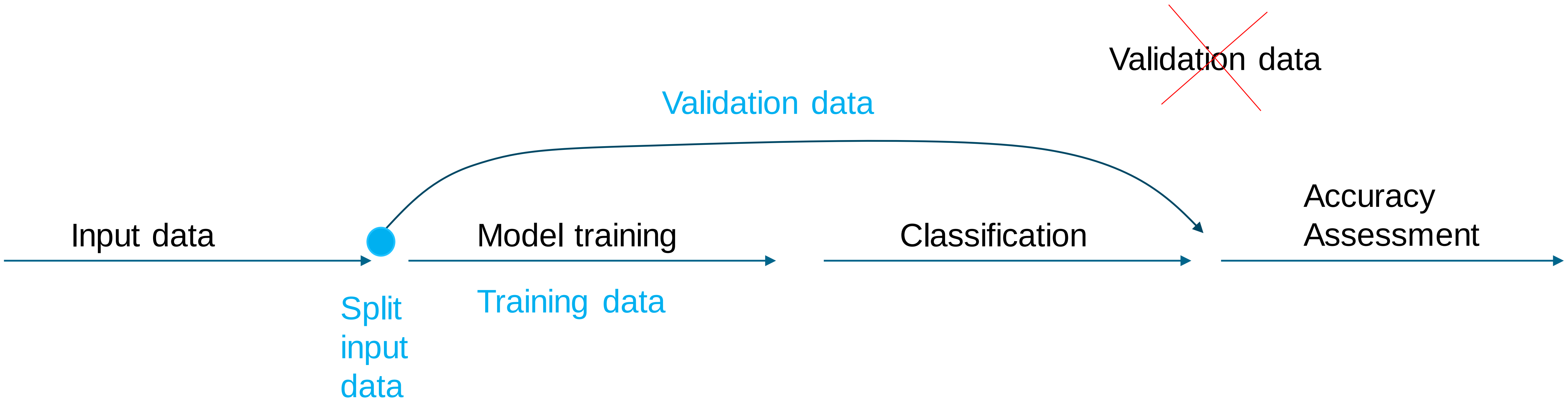
$$Precision = \frac{TP_{class}}{(TP_{class} + FP_{class})}$$

		Reference	
		1	0
Prediction	1	*TP*	*FP*
	0	*FN*	*TN*

4) RS data classification: Accuracy assessment

Different ways to assess accuracy:





4) RS data classification: Training data, model training, LC classification and accuracy assessment

Bill Bertram, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

040_LC_classification.R

RS Block: Workflow practical unit

Basics:

Raster data

Vector data

Data Preparation:

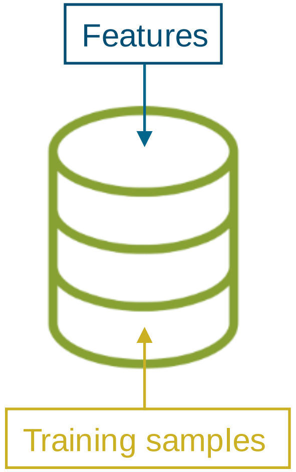
1) RS data acquisition

2) RS data handling

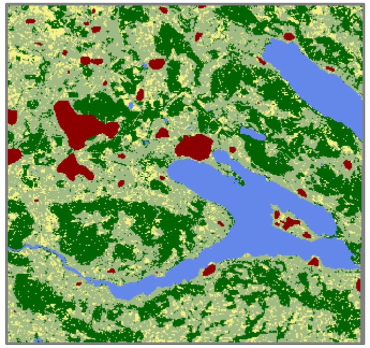
3) Raster manipulation

4) RS data classification:

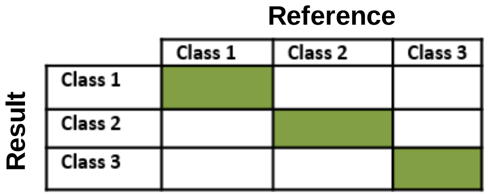
4.1) Training data



4.2) Model training and classification

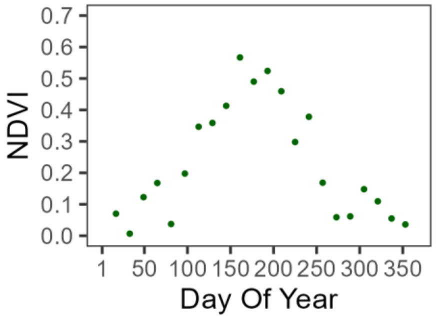


4.2) Accuracy assessment

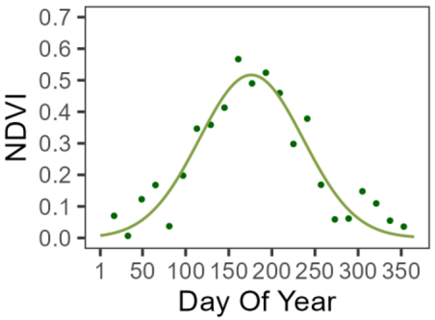


5) Vegetation phenology:

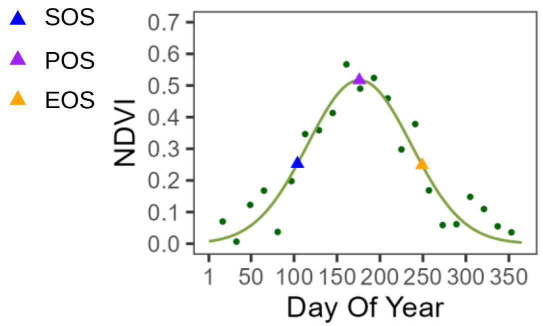
5.1) Input data



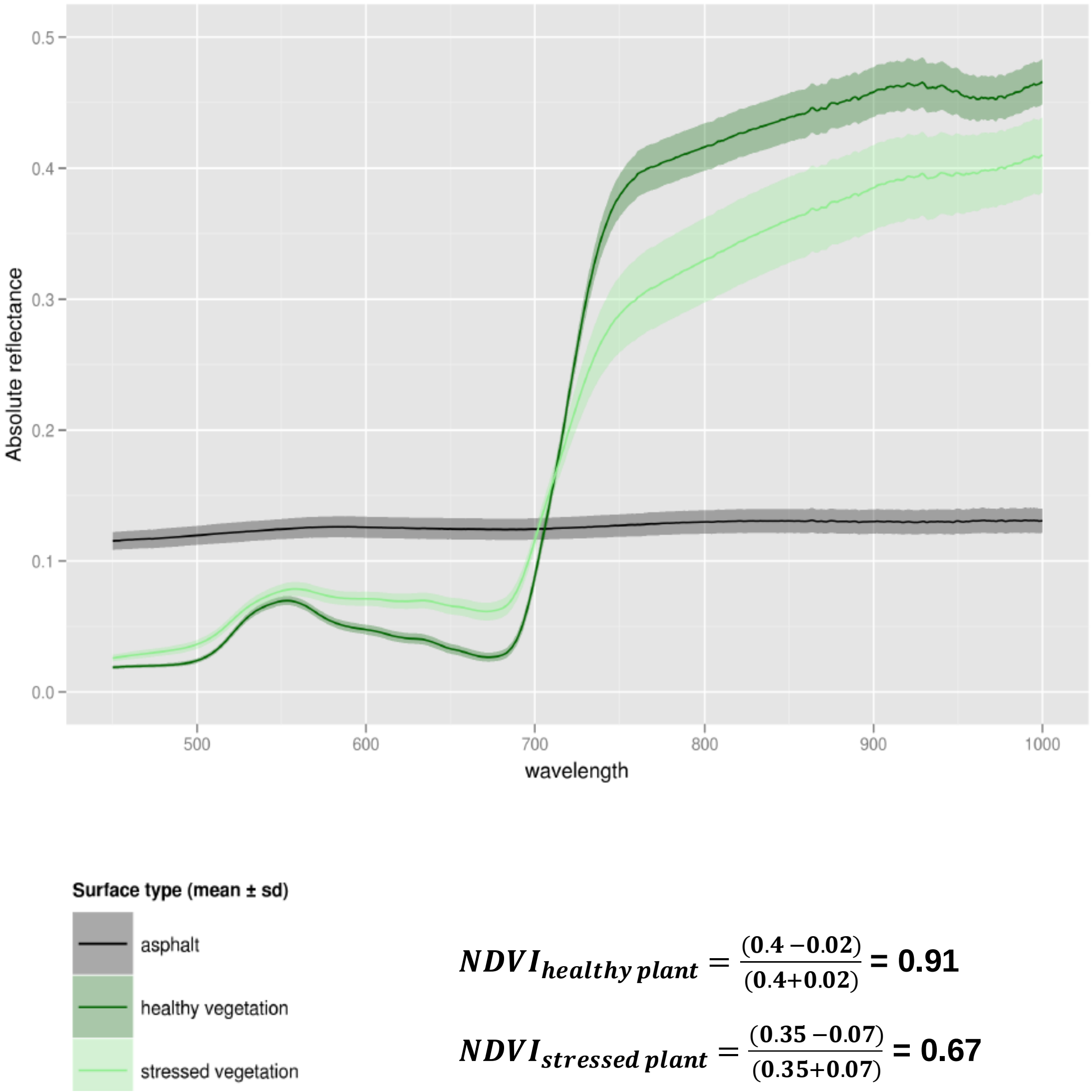
5.2) VI curve fitting

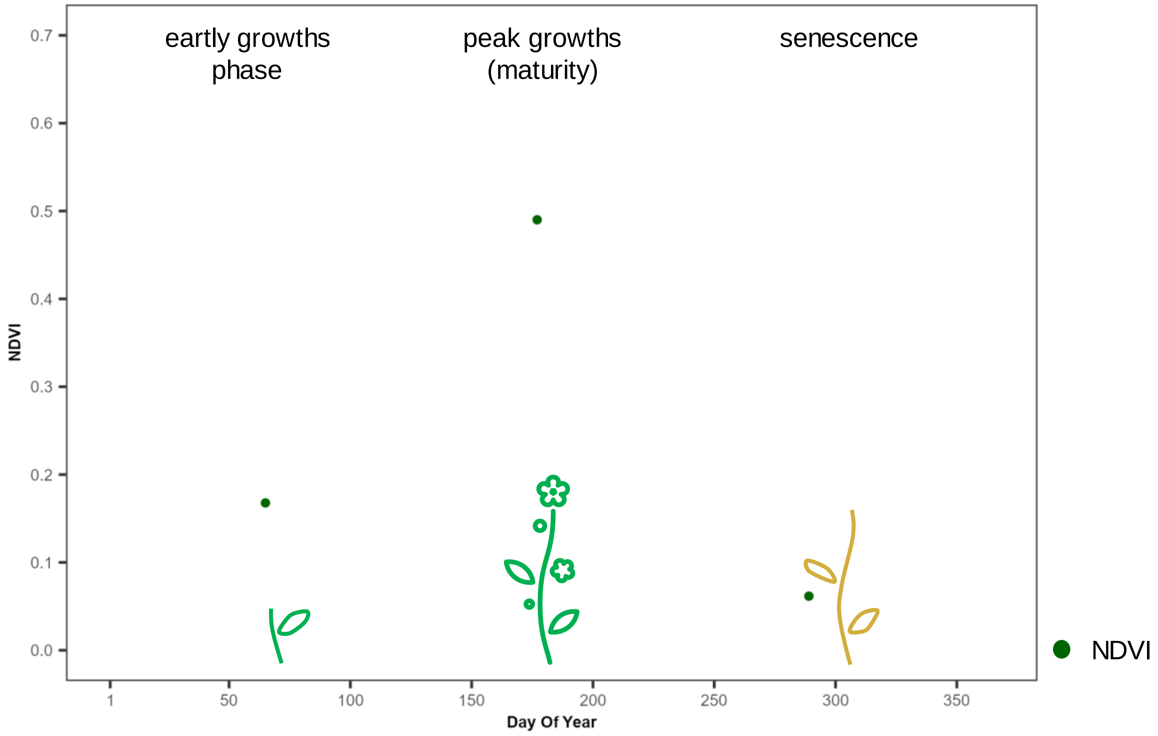


5.3) Phenological metrics



5) Vegetation phenology: Basics





$$NDVI_{healthy\ plant} = \frac{(0.4-0.02)}{(0.4+0.02)} = 0.91$$

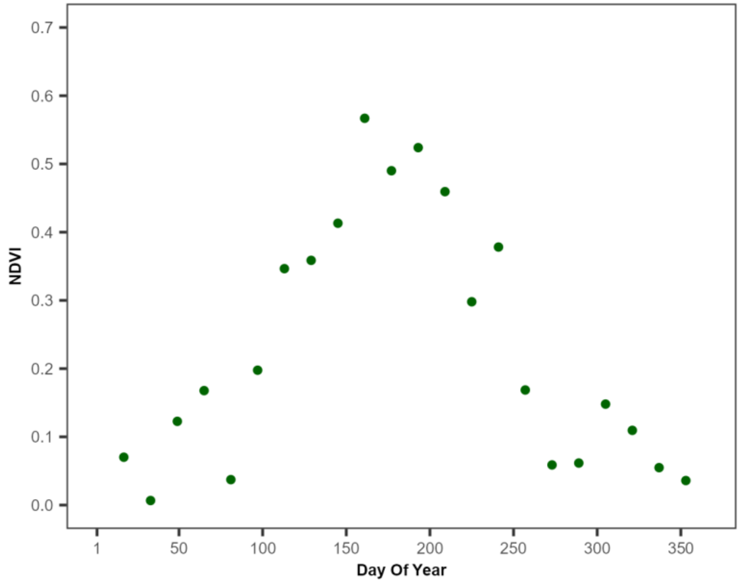
$$NDVI_{stressed\ plant} = \frac{(0.35-0.07)}{(0.35+0.07)} = 0.67$$

Vegetation phenology definition:

- Is the study of **recurring patterns of plant growth and development**, for example, the time of (a) plant flowering, (b) leaf green-up and (c) senescence (Lieth, 1974).
- Can be studied **intra- and inter-annual**

5) Vegetation phenology: How can it be assessed?

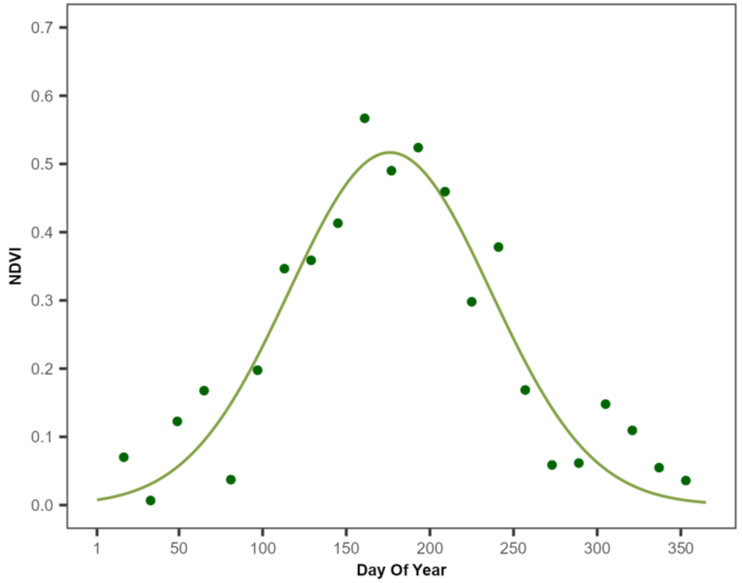
A) Input data



multi-temporal VI measurements from RS data

- e.g., MOD13A1 (500m, every 16 days)

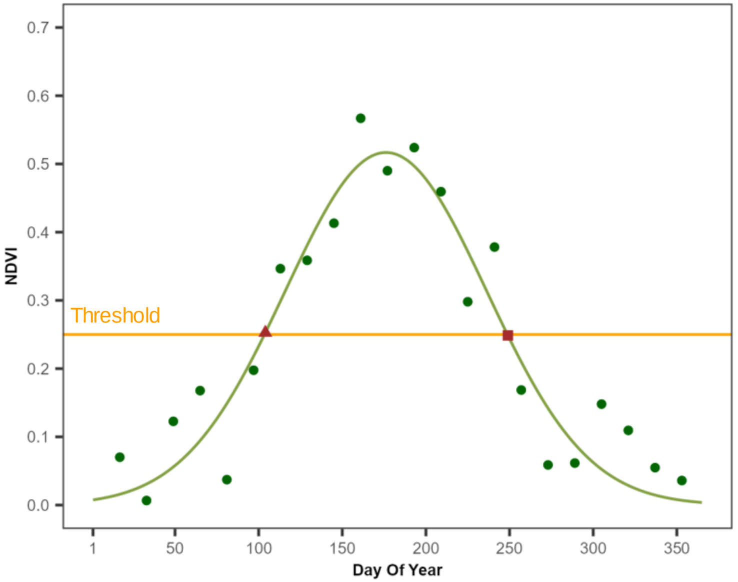
B) Curve fitting



Reconstruct 1-day temporal resolution **VI profile**

C) Extraction of phenological metrics

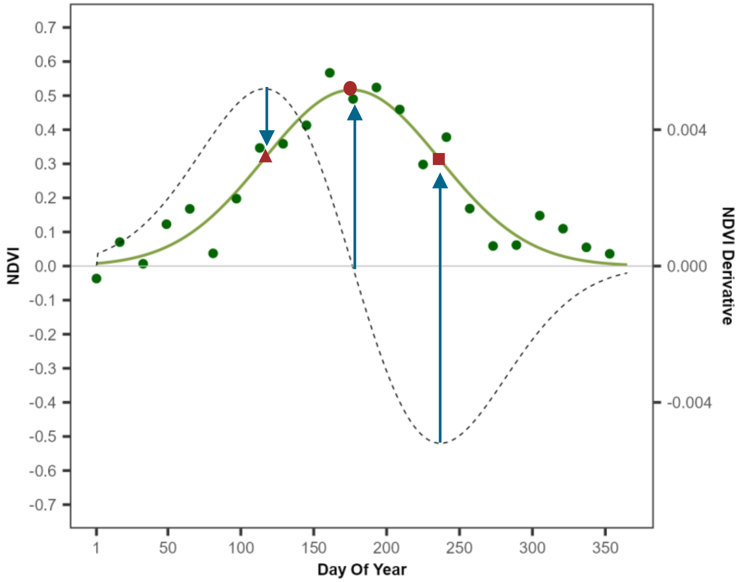
1) Threshold approach



▲ start of season (sos) ■ end of season (eos)

● peak of season (pos)

2) Derivative approach



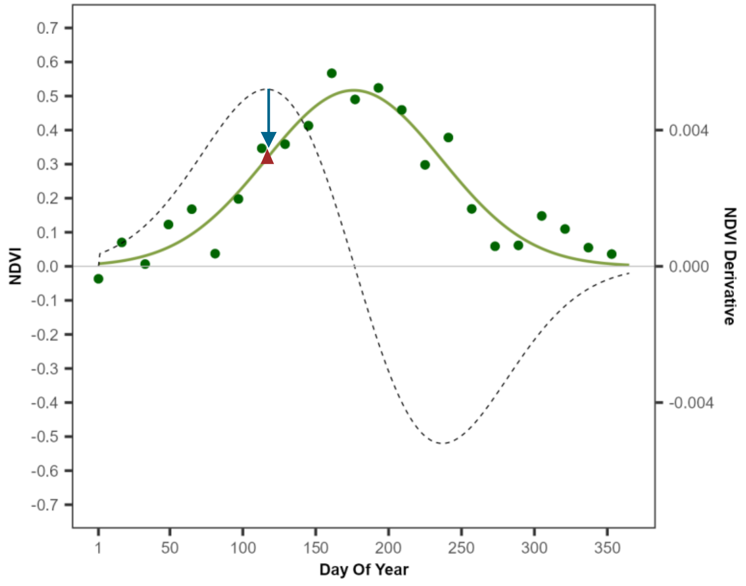
5) Vegetation phenology: What can be studied?

Climate change:

- shifts in phenological metrics throughout different years (inter-annual)
    - earlier onset of start of season (e.g. Høgda et al. 2013)

Species-environment relationships:

- „green wave“ studies (intra-annual) (e.g. Bischof et al. 2012, Aikens et al. 2017)
    - Determine peak green-up and assess if herbivorous species adjust their choice of foraging habitats during spring migration to coincide with this peak



▲ start of season (sos) = peak green-up

derivative = instantanuous rate of green-up (IRG)

5) Vegetation phenology: Download NDVI time series with GEE

https://code.earthengine.google.com/

GEE_MOD13A1.txt

5) Vegetation phenology: input data preparation, preprocessing, curve-fitting, phenology metrics

library(phenofit)

050_vegetation_phenology.R

References

Aikens, E. O., M. J. Kauffman, J. A. Merkle, S. P. H. Dwinnell, G. L. Fralick, and K. L. Monteith. 2017. The greenscape shapes surfing of resource waves in a large migratory herbivore. Ecology Letters 20:741–750.

Bischof, R., L. E. Loe, E. L. Meisingset, B. Zimmermann, B. Van Moorter, and A. Mysterud. 2012. A migratory northern ungulate in the pursuit of spring: jumping or surfing the green wave? The American Naturalist 180:407–424.

Høgda, K., H. Tømmervik, and S. Karlsen. 2013. Trends in the Start of the Growing Season in Fennoscandia 1982–2011. Remote Sensing 5:4304–4318.