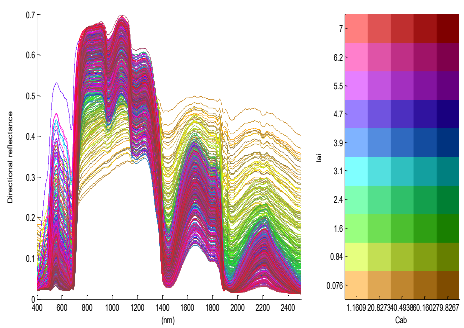
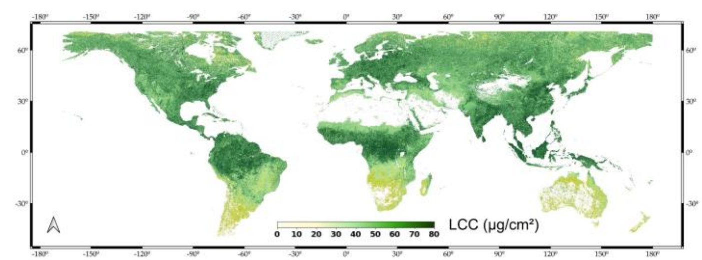
From model simulations towards vegetation properties mapping:

automating, optimizing & expanding

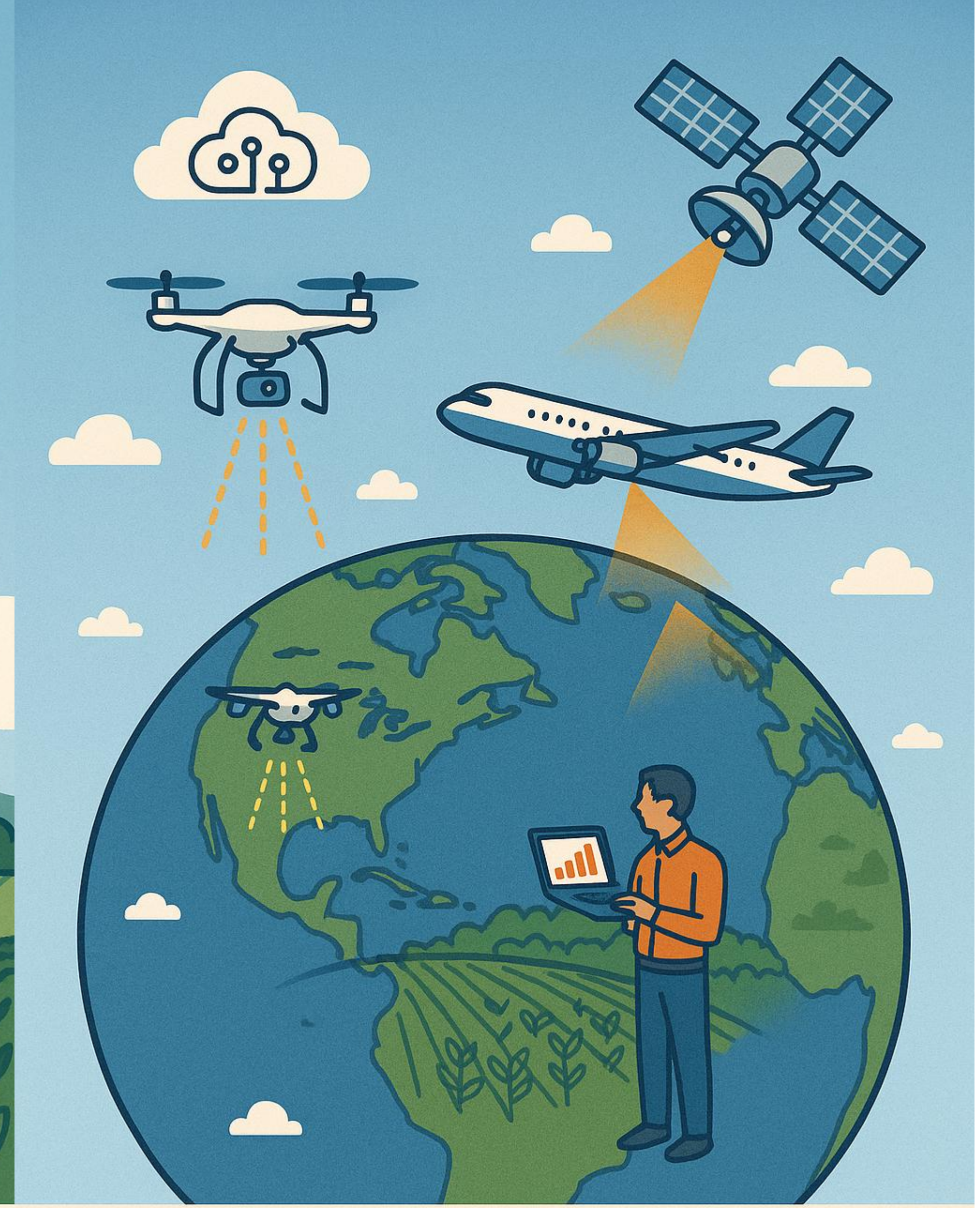




Jochem Verrelst

July 2025

QUANTITATIVE REMOTE SENSING OF VEGETATION



QUANTITATIVE REMOTE SENSING OF VEGETATION

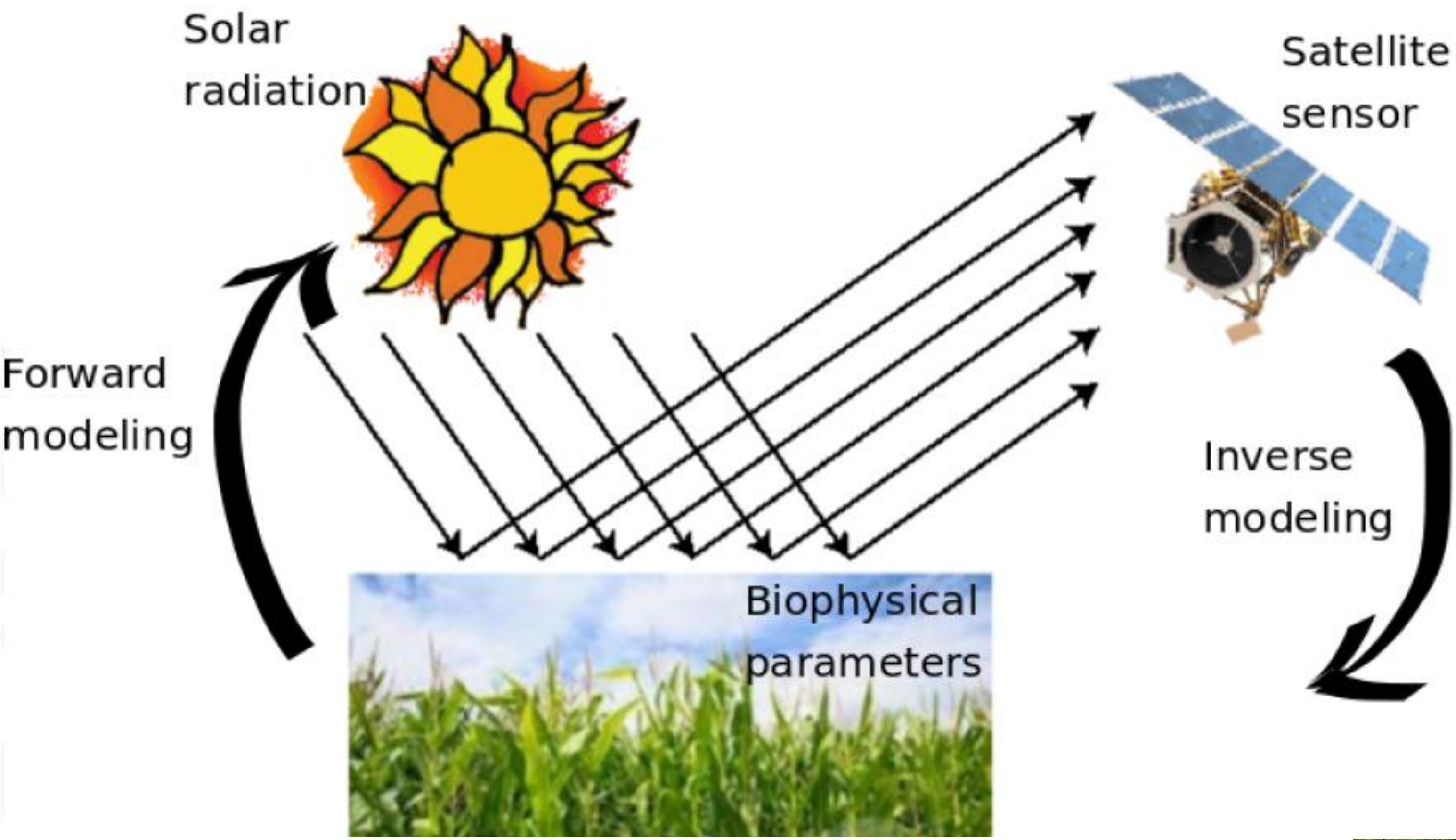
VEGETATION CHANGE

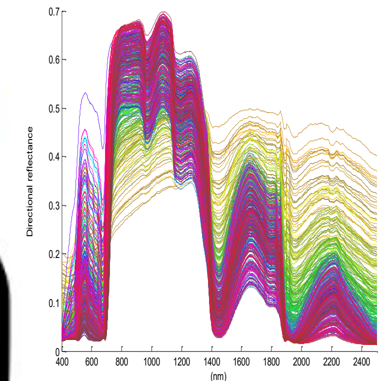
CROP YIELD PREDICTION

How to quantify vegetation properties?

Today we will learn:

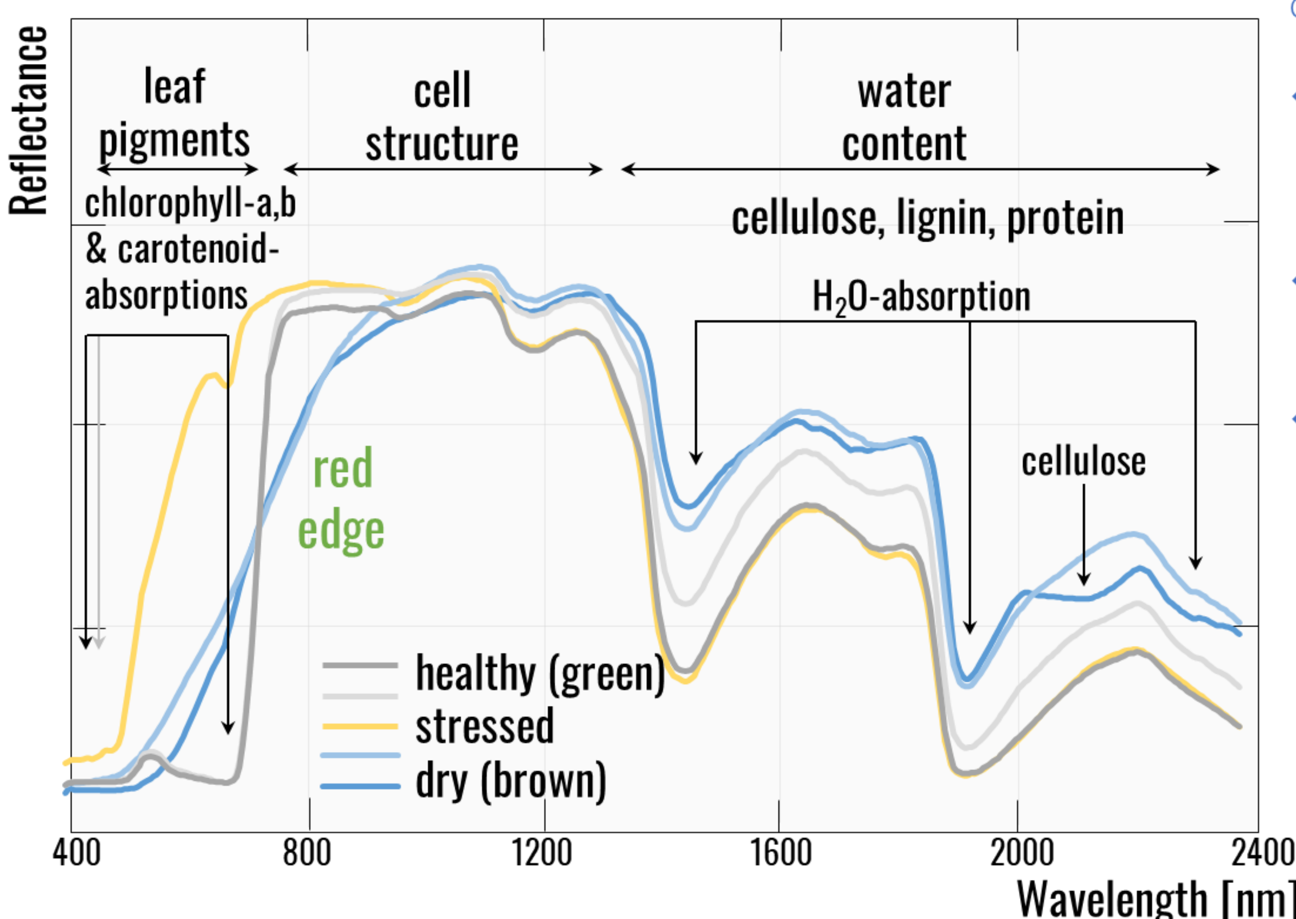
Semi-automated mapping of vegetation properties from optical RS data



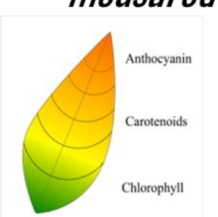


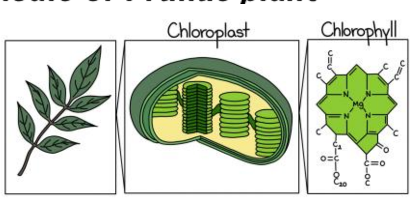
- *What are biophysical parameters?*
- *Why is it important to quantify them?*

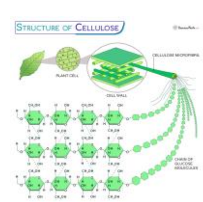
Factors controlling leaf reflectance

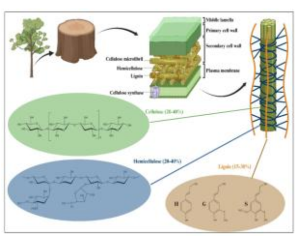


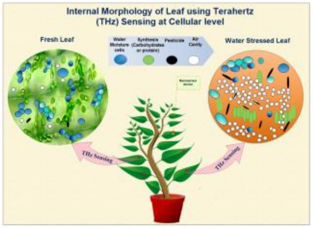
measured at leafs of Prunus plant









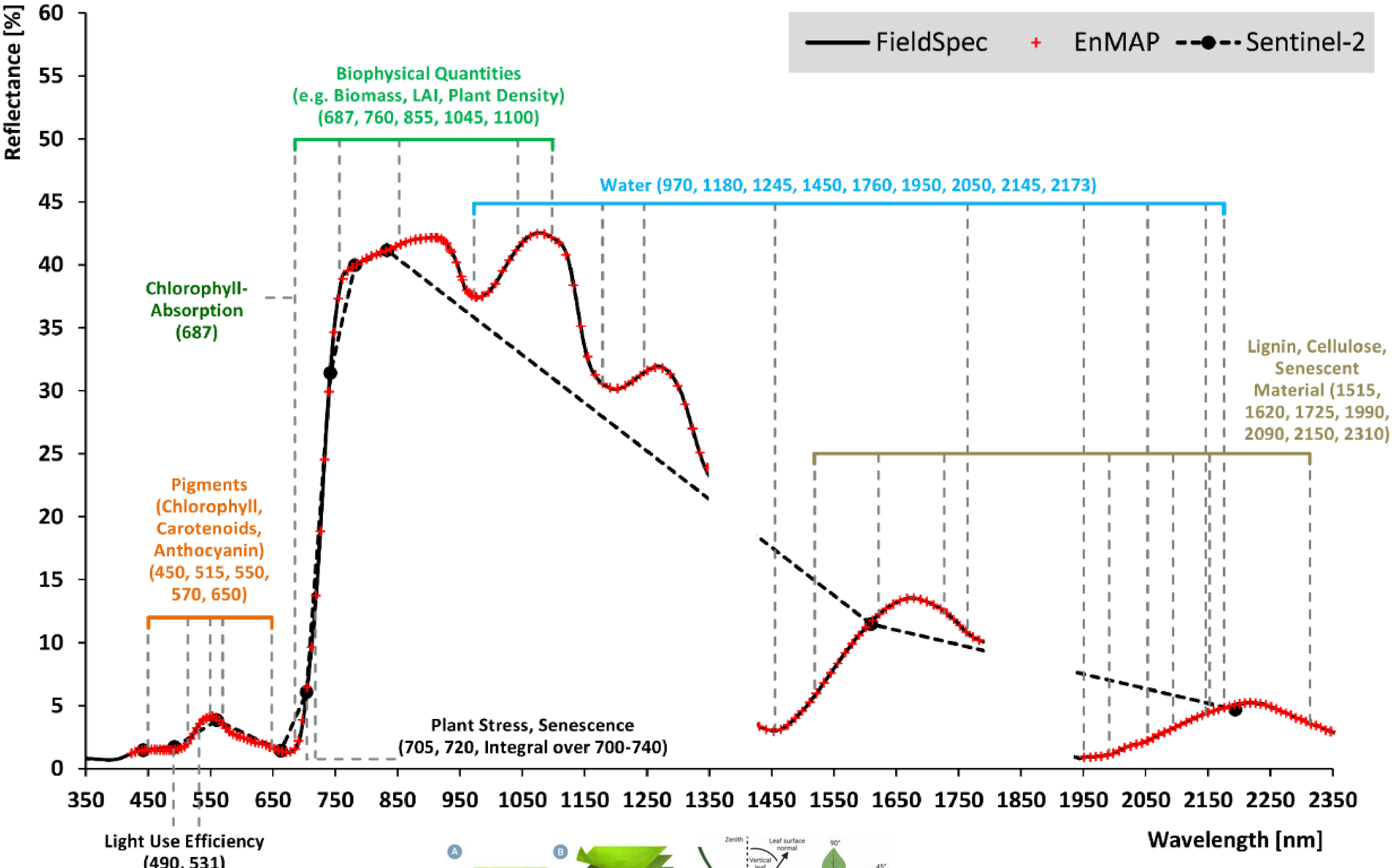


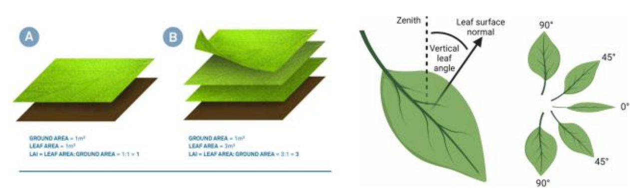
Shape of the leaf spectra is characterized by:

- Low reflectance across visible wavelengths (due to absorption by photosynthetic pigments).
- High reflectance in the NIR, with only ~ 10% of absorbed radiation.
- Intermediate reflectance in the SWIR, where energy is mainly absorbed by water or plant residues in case of dry/stressed leaves. Cell wall compounds (cellulose, lignin, proteins and sugars) lead to overlapping absorption features.

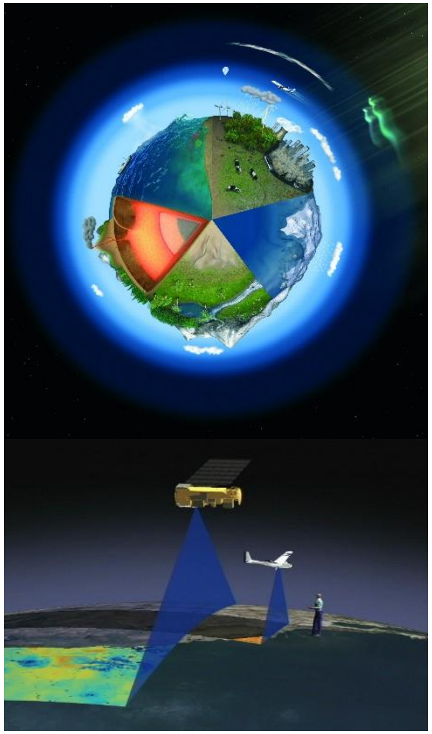
(Ustin & Jacquemoud 2020)

Satellite hyperspectral reflectance





The problem:

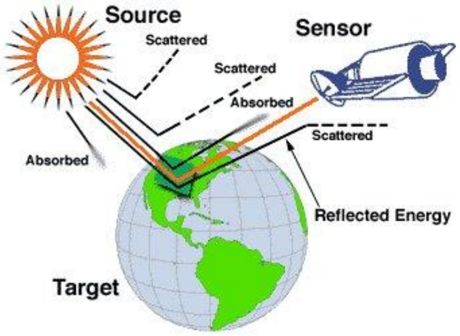


Biophysical parameter retrieval is an essential step in modeling the processes occurring on Earth and the interactions with the atmosphere.

The analysis can be done at **local** or **global** scales by looking at bio-geo-chemical cycles, atmospheric situations, ocean/river/ice states, and vegetation dynamics.

Main parameters: crop yield, biomass, leaf area coverage, chlorophyll content, fraction vegetation cover, GPP,….

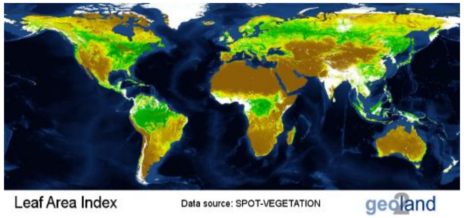
Land/vegetation parameters cannot be estimated directly from optical RS data. A model is required!



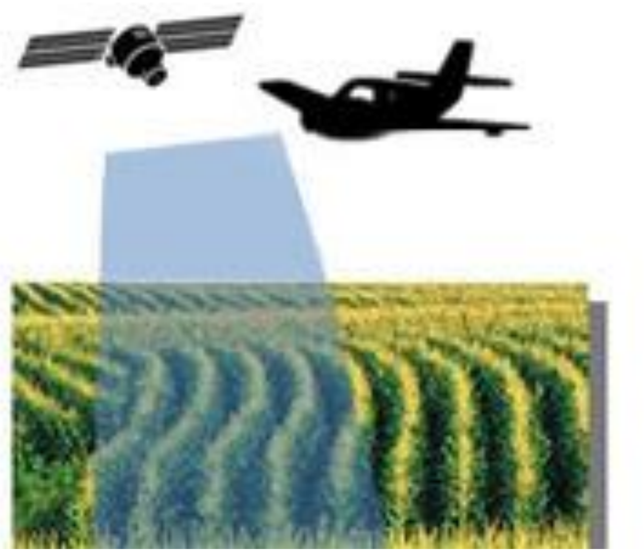
The objective: Transform measurements into biophysical parameter estimates.

The data:

- **Input data:** satellite/airborne spectra, in situ (field) radiometers, or simulated spectra by RTMs
- **Output results:** estimation of a biophysical parameter



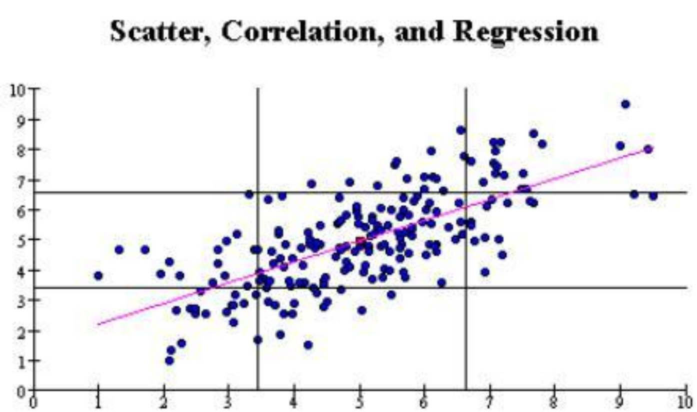
Introduction retrieval biophysical parameters

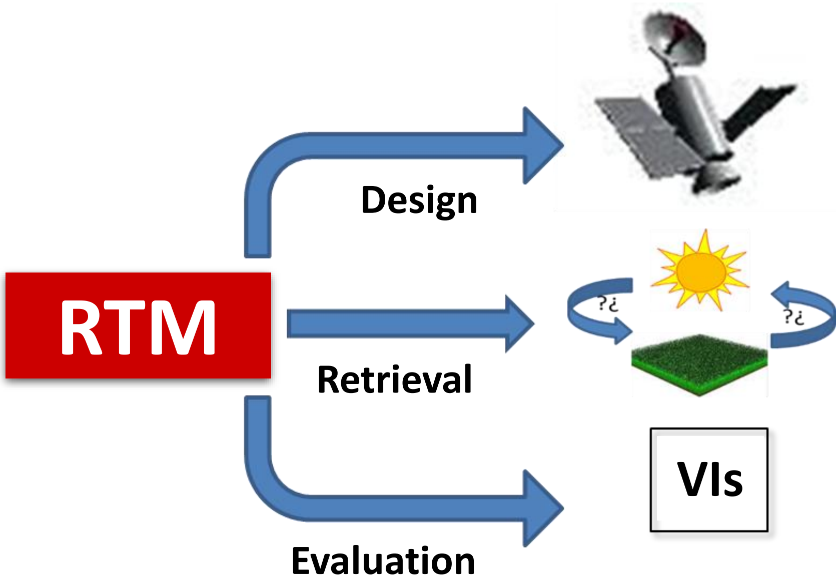


Retrieval of biophysical parameters from Remote Sensing (RS) data **always occurs through a model**, e.g. through statistical models or through inversion of physically-based radiative transfer models (RTM).

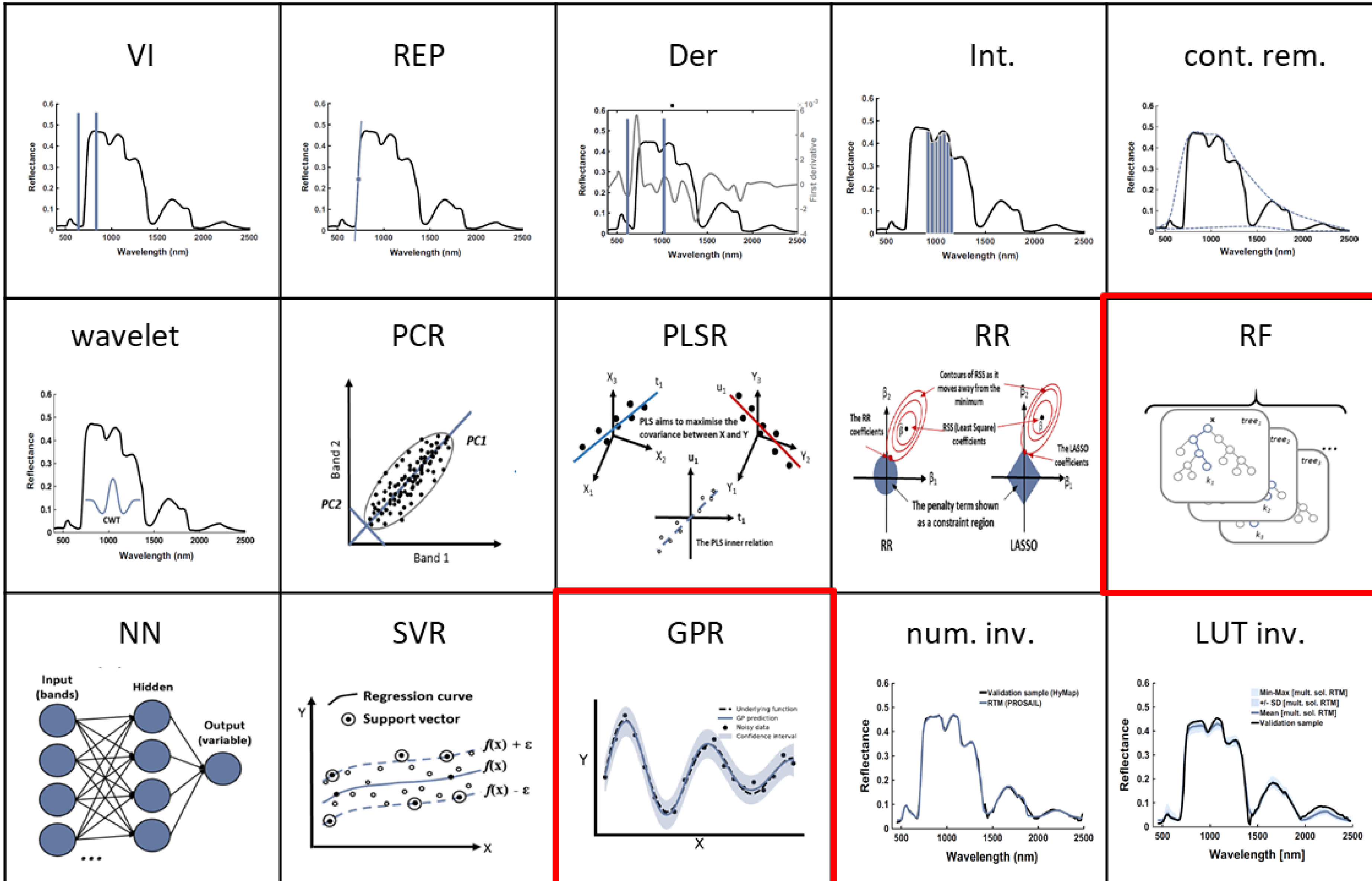
Statistical approaches

Physically based RTM approaches





Some retrieval methods....



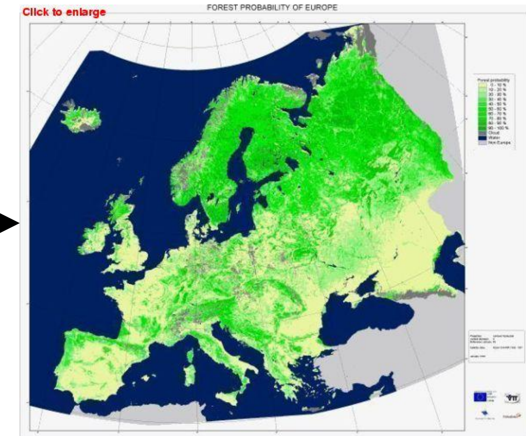
Verrelst, J et al. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, *40*(3), 589-629.

Retrieval of (continuous) vegetation properties

Remote sensing image

Model

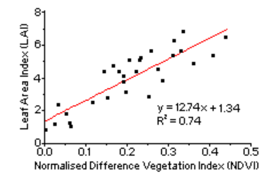
Map of a vegetation property



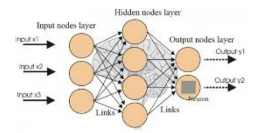
1. **Statistical models**
   1. Parametric regression models
   2. Nonparametric regression models
      1. Linear
      2. Nonlinear: ML
2. **Inversion of physically based radiative transfer models**
   1. Numerical optimization
   2. Lookup-table (LUT)-based inversion

Taxonomy of retrieval methods, three main families:

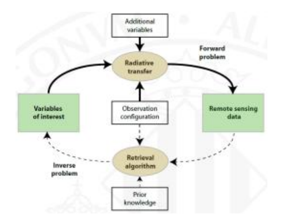
1. ***Statistical:* parametric and non-parametric:**
   - **Parametric** models rely on ***some* physical knowledge** of the problem and build explicit parametrized expressions that relate a few spectral bands with the biophysical parameter(s) of interest.



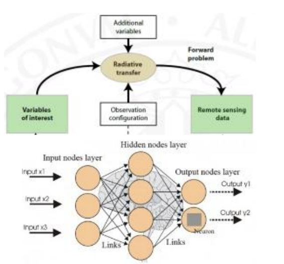
   - **Non-parametric** models are ***data-driven* models**. They are adjusted to predict a variable of interest using a training dataset of input-output data pairs.



2. ***Physical*: try to reverse RTMs.**
   - Physically based algorithms are applications of physical laws establishing photon interaction ***cause–effect relationships***. Model variables are inferred based on specific knowledge, typically obtained with **radiative transfer functions**.



3. ***Hybrid*:**
   - A hybrid-method **combines elements of nonparametric statistics and physically based methods**. Hybrid models rely on the generic properties of physically based methods combined with the flexibility and computational efficiency of nonparametric nonlinear regression methods.



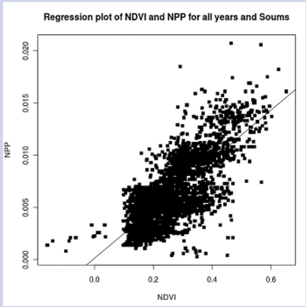
Retrieval families

Parametric regression

Spectral relationships that are sensitive to specific vegetation properties

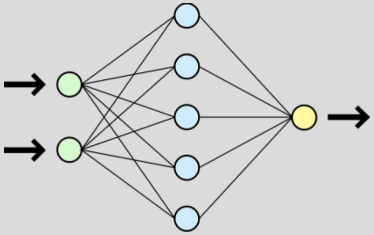
$$NDVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED})}$$

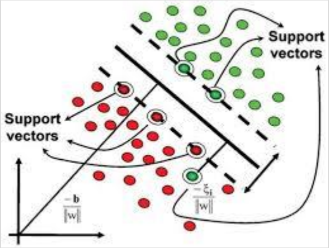
Normalized Difference Vegetation Index



Non-parametric regression

Advanced techniques that search for relationships between spectral data and biophysical variables

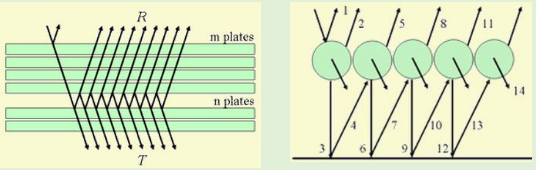




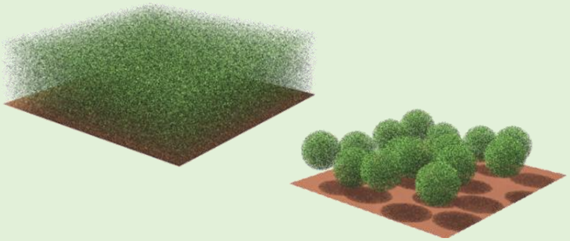
RTM inversion

Models that simulate interactions between vegetation and radiation

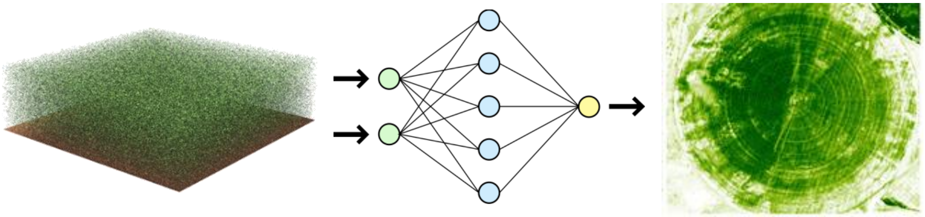
leaf



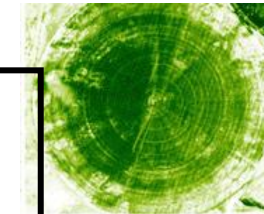
canopy



Methods of these different families can be combined: *hybrid methods*



Statistical interpretation of RS



Remote Sensing Data ⟷ **Variable of Interest**

Statistical relationship

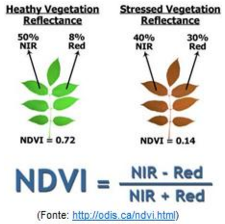
- Parametric regression
- Non-parametric regression

- Simple statistical relationships (VIs) constitute the **BULK of RS analysis**.
- These analyses allow to determine **IF** there is a relationship, **not WHY** there is a relationship.
- Linear methods such as VIs are **useful indicators** of biophysical (e.g. structure) or biochemical (e.g. chlorophyll) parameters, however in natural, complex environments indices are **confounded** by additional abiotic and biotic factors.
- **VIs lack generality** for estimating biophysical parameters.
- Apart from VIs a large number of powerful **alternative statistical retrieval** methods exists (e.g. non-parametric regression methods).

Parametric regression

Parametric regression assume an explicit model for retrieval

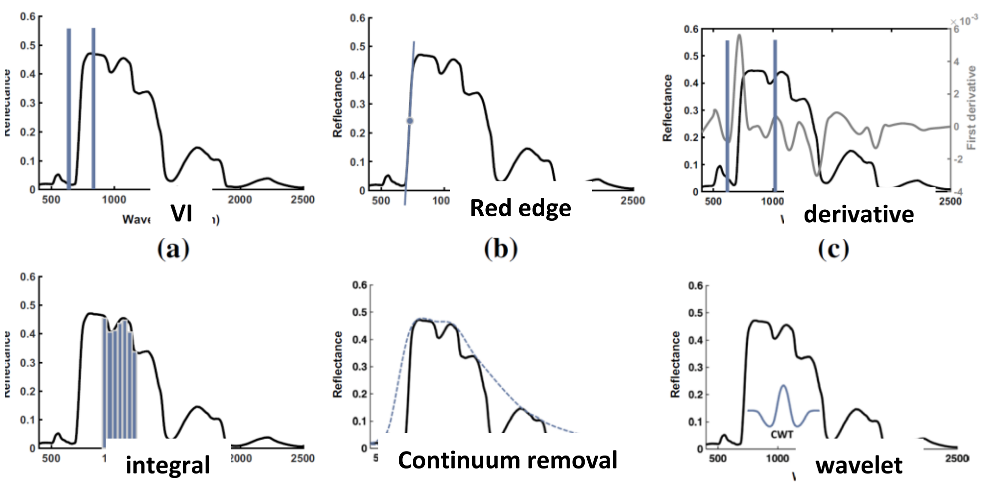
- **Discrete band methods(VIs):**
  - 2-band: SR, NDVI, PRI, OSAVI
  - 3-band: TVI, MCARI, SIPI
  - 4-band: TCARI/OSAVI



$$PRI = \frac{(\rho 570 - \rho 531)}{(\rho 570 + \rho 531)}$$

$$\begin{aligned} TCARI / OSAVI = 3\cdot\left[\left(\rho_{\mu 700} - \rho_{\mu 670}\right) - 0.2\cdot\left(\rho_{\mu 700} - \rho_{\mu 550}\right)\cdot\right] \\ \left(\rho_{\mu 700} - \rho_{\mu 670}\right)\left(1 + 0.16\right)\cdot\left(\rho_{\mu 800} - \rho_{\mu 670}\right) / \\ \left(\rho_{\mu 800} - \rho_{\mu 670} + 0.16\right) \qquad (2) \end{aligned}$$

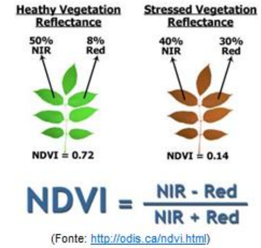
- **Shape-based methods (hyperspectral data):**
  - Red-edge position (REP)
  - Derivative/Integral indices
  - Continuum removal
  - wavelet



(a) (b) (c)

(d) (e) (f)

Parametric regression:



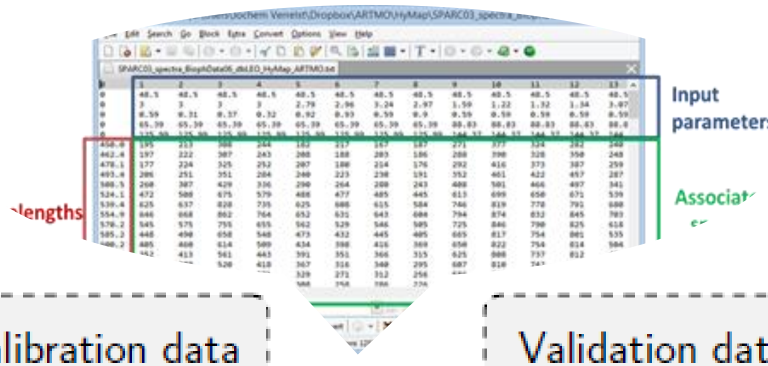
Strengths ☺

- Simple and comprehensive regression models; little knowledge of user required.
- Fast in processing
- Computationally inexpensive

Weaknesses ☹

- Makes only poorly use of the available information within the spectral observation; at most a spectral subset is used. Therefore, they tend to be more noise-sensitive as compared to full-spectrum methods
- Parametric regression puts boundary conditions at the level of chosen bands, formulations and regression function.
- Statistical function accounts for one variable at a time.
- A limited portability to different measurement conditions or sensor characteristics
- No uncertainty estimates are provided. Hence the quality of the output maps remains unknown.

Parametric regression

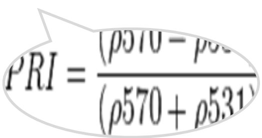


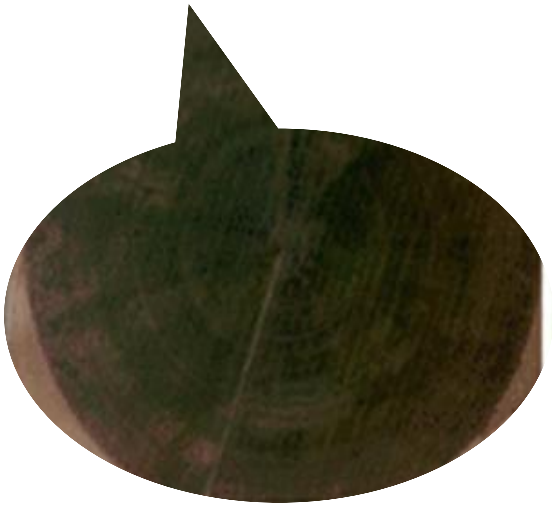
Calibration data

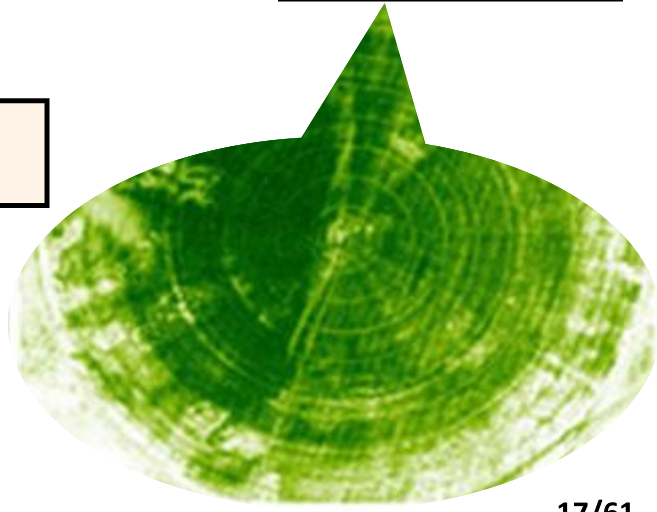
Validation data

Remote Sensing Imagery → Spectral algorithms (e.g. Veg. Index) → Calibrated (simple) regression model → Map of variable of interest (e.g. LCC, LAI)

Calibrated (simple) regression model → Validation







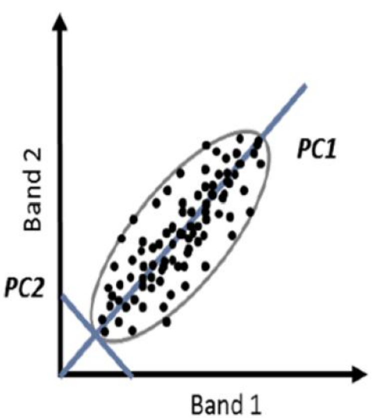
Non-parametric models (1/2):

Data-driven methods: Do not assume explicit feature relations

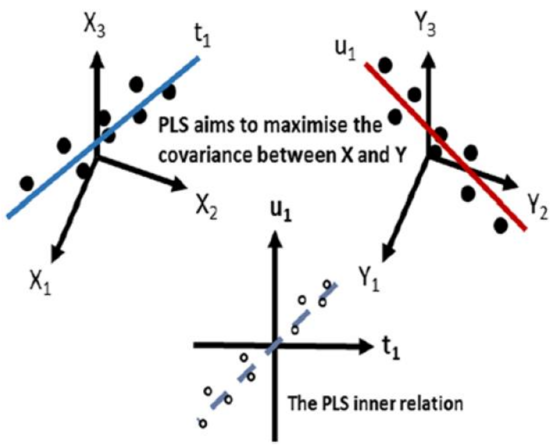
Linear nonparametric models:

- Stepwise multiple linear regression (SMLR)
- Principal component regression (PCR)
- Partial least squares regression (PLSR)
- Ridge regression (RR)
- Least Absolute Shrinkage and Selection Operator (LASSO)

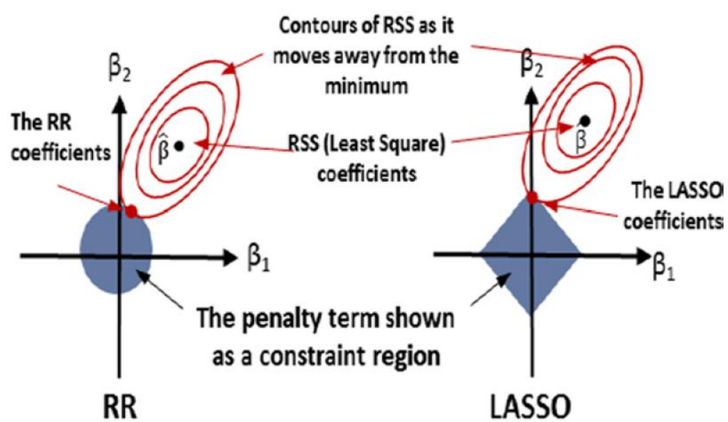
PCR



PLSR



RR & LASSO

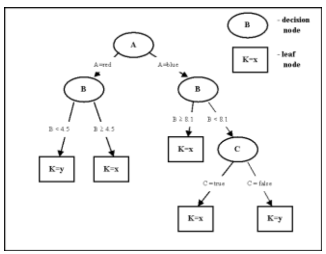


Non-parametric models (2/2):

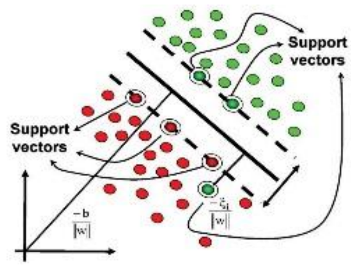
Data-driven methods: Do not assume explicit feature relations

Non-linear nonparametric models:

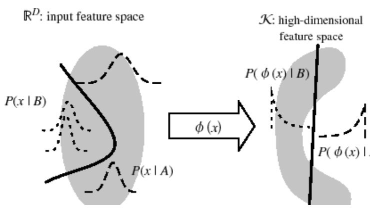
Decision Trees (DT)



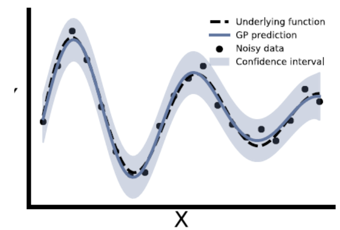
Support vector regression (SVR)



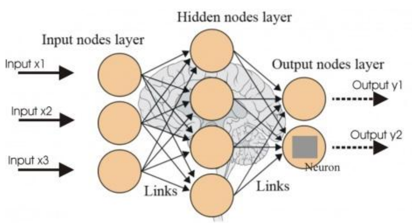
Gaussian processes regression (GPR)



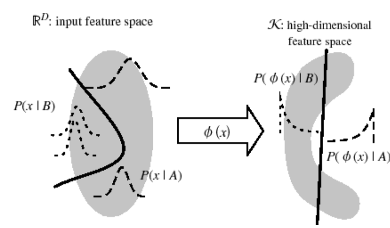
$$K(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(- \sum_{b=1}^{B} \frac{\left(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)}\right)^2}{2\sigma_b^2} \right)$$



Neural networks (NN)



Kernel ridge regression (KRR)



$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/(2\sigma^2)).$$

Also:

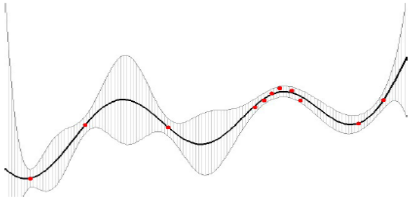
- Elastic Net (ELASTICNET)
- Bagging trees (BAGTREE)
- Boosting trees (BOOST)
- Neural networks (NN)
- Extreme Learning Machines (ELM)
- Relevance Vector Machine (RVM)
- Gaussian process Regression (GPR)
- Variational Heteroscedastic Gaussian Process Regression (VHGPR)

Gaussian Processes Regression (GPR)

- A **GPR** model is a **probabilistic** (**Bayesian**) model directly in function space, with no intermediate model or model parameters.

- **GPR** are **equivalent** to **kernel ridge regression**, least square **suport vector machines** (SVM), **Kriging**, large **neural networks** (NN) and very closely related to **SVM regulazation networks**.

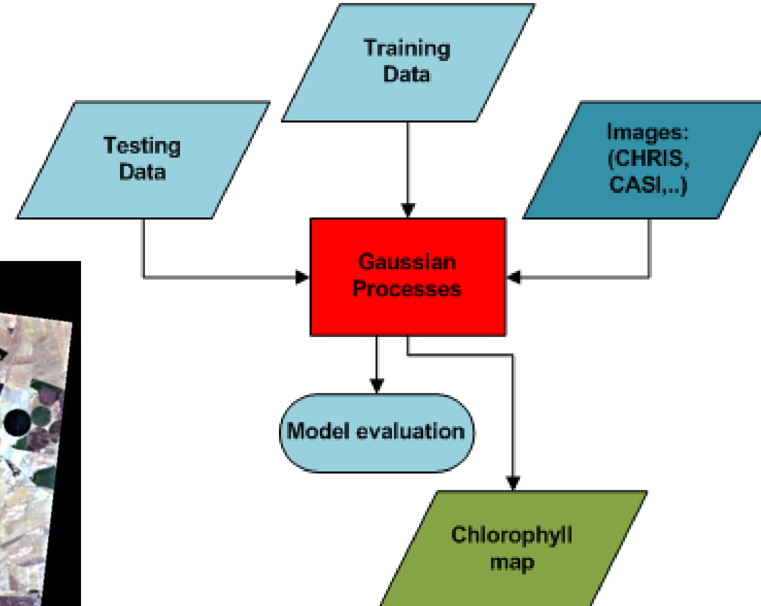
- **GPR alleviates** some **shortcomings** of the previous methods, while maintaining very good numerical performance and stability:
  - GPR is far simpler **than NN**, and needs **fewer sample points** ☺
  - **Not only a mean prediction** for each sample (**pixel**), but also a full distribution over the output values **including an uncertainty of the prediction** (**confidence interval**). ☺
  - GPR **provide a ranking of features (bands) and samples (spectra)**, thus partly **overcoming the blackbox problem**. ☺
  - http://www.rainsoft.de/projects/gausspro.html



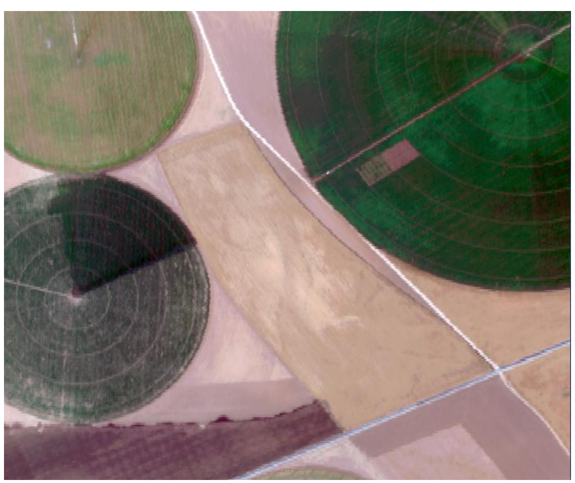
A regression curve plot by the "Gaussian Process Regression Applet" using 11 data points. One can observe that uncertainty goes down when multiple data points are aggregated together.

More info on: http://www.gaussianprocess.org/
Rasmussen and C Williams, **Gaussian Processes for Machine Learning**, 2006

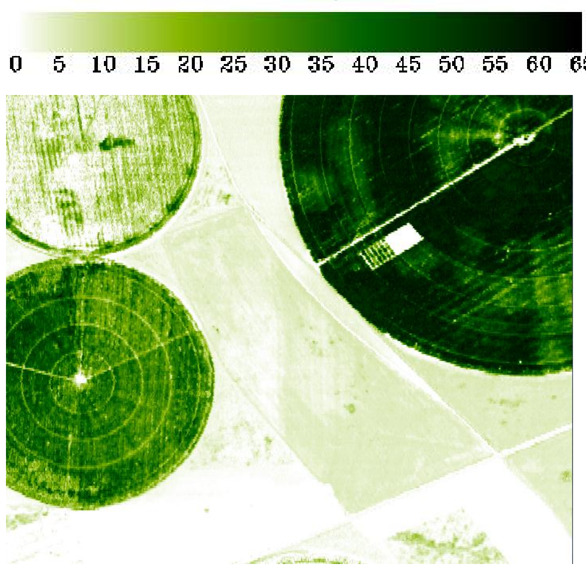
Example GPR



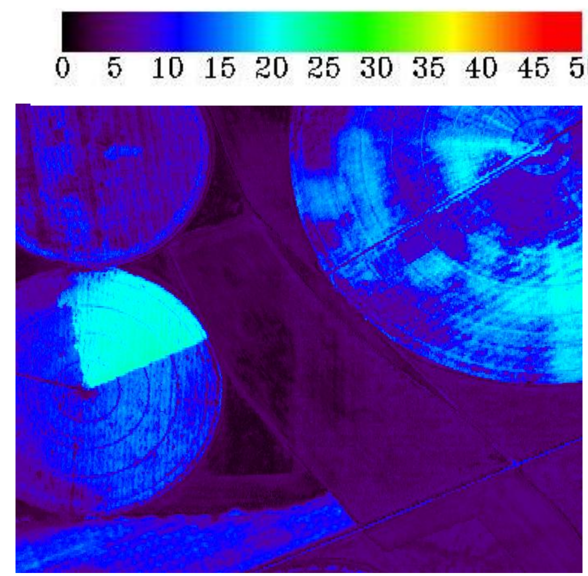
RGB CASI



Chl [µg/cm^2]



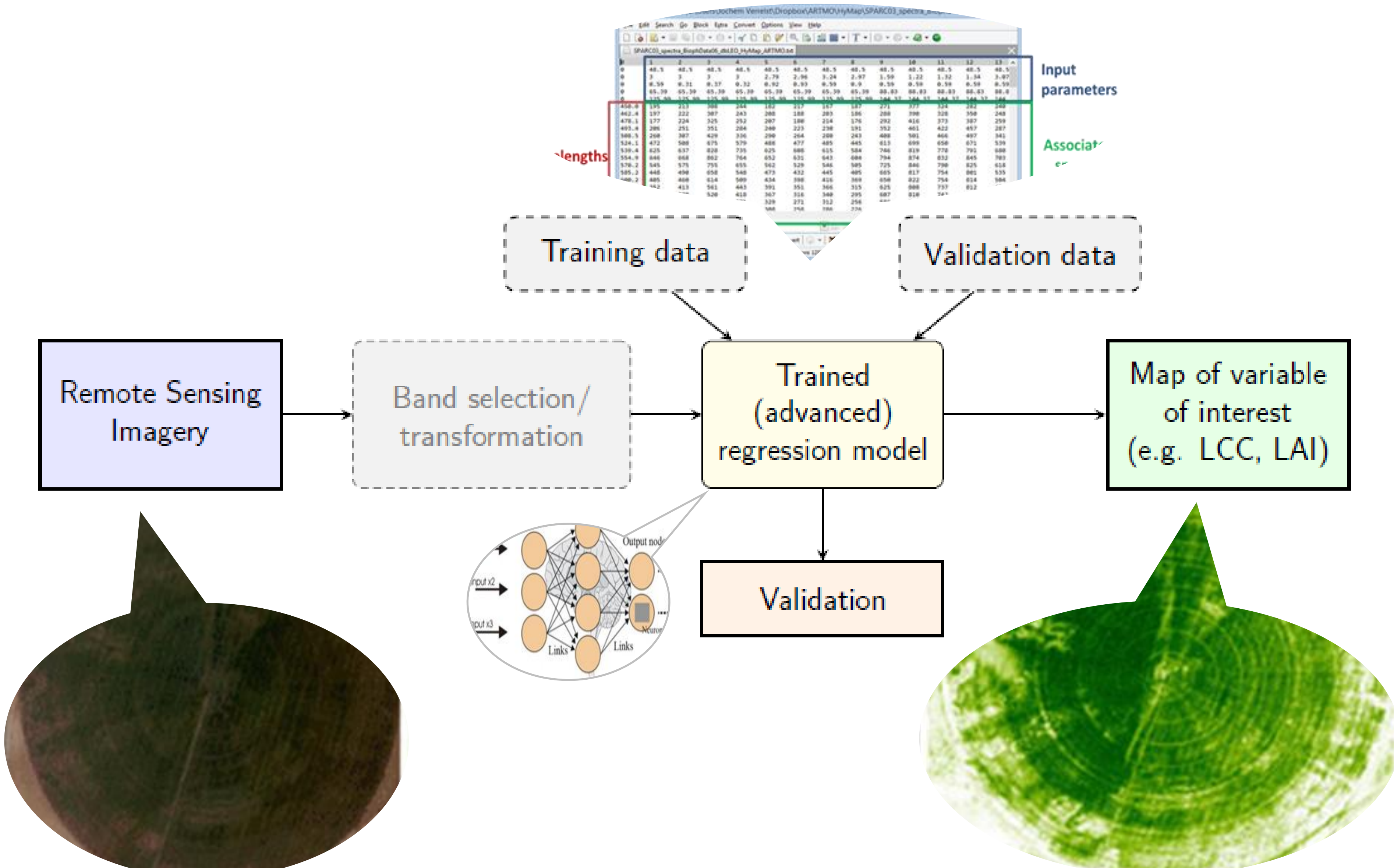
St Dev
Chl [µg/cm2]



Non-parametric regression:

Strengths ☺	Weaknesses ☹
• Full-spectrum methods. They make use of the complete spectral information. • Advanced, adaptive (non-linear) models are built. • Methodologically, accurate and robust performance is enabled. • Some MLRAs cope well with datasets showing redundancy and high noise levels. • Once trained, imagery can be processed time efficient. • Some of the non-parametric methods (e.g. ANNs, decision trees) can be trained with a high number of samples (typically >1,000,000). • Some MLRAs provide insight in model development (e.g. GPR: relevant bands; decision trees: model structure). • Some MLRAs can provide multiple-outputs (e.g. PLRS, ANN, SVR, GPR and KRR) • Some MLRAs provide uncertainty intervals (e.g. GPR).	• Training can be computational expensive. • Hypercomplex models can be generated. Their generic potential is limited and hence they do not generalize well, based on the training data (problem of over-fitting). • Some regression algorithms are difficult (or even impossible) to train with a high number of samples. • Expert knowledge is required, e.g. for tuning. However, toolboxes exist automating some of the steps in this sub-process. • Some of the methods can be considered as black boxes. • Some regression algorithms elicit instability when applied with datasets statistically deviating from the datasets used for training.

Non-parametric regression



Physical interpretation of RS

Remote Sensing Data

RT Models

Invertible models

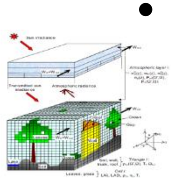
RT State Variables

Some relationship

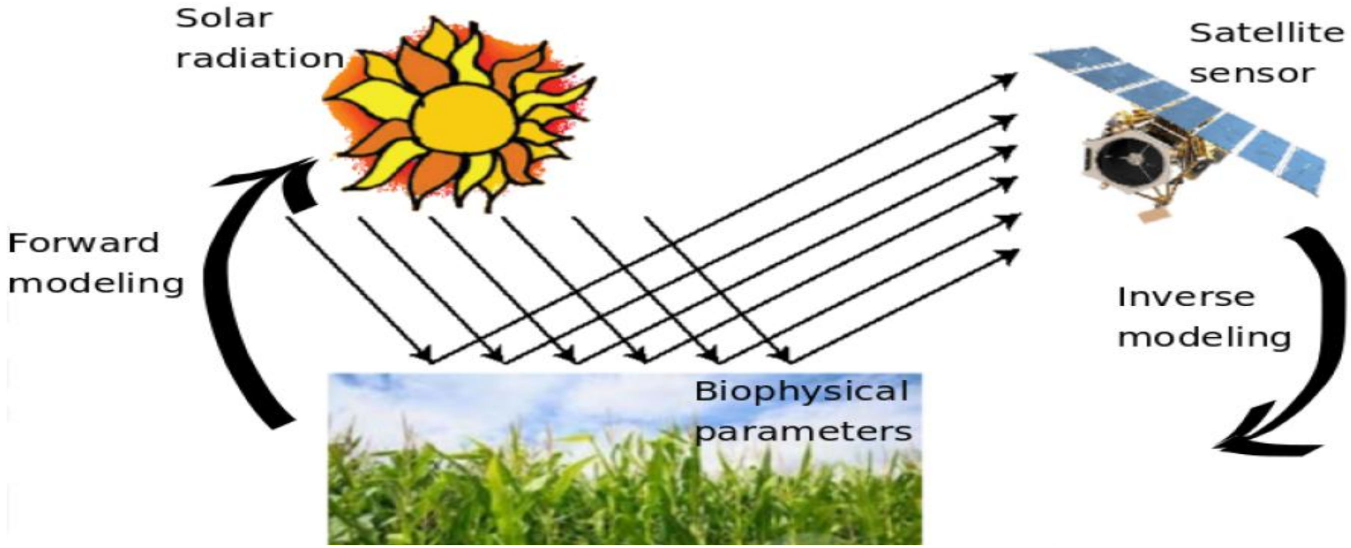
Variable of Interest

Radiative transfer models:

- Try to predict RS data based on a function of the RT state variables
- Two categories of RT models:
  - **<u>Economically invertible models</u>**: typically designed for simple scenes, have a few number of state variables (e.g. SAIL, RPV)
  - **<u>Non-economically invertible models</u>**: typically designed for complex scenes, have a large number of state variables (e.g. DART, Drat)



Background



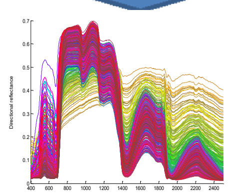
RTMs

Physically based RTM approaches

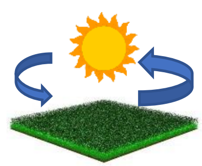
Development/ Evaluation

Retrieval

Design



Spectra/ VIs



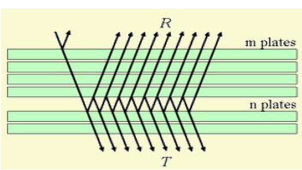
Mapping biophysical param.

mission

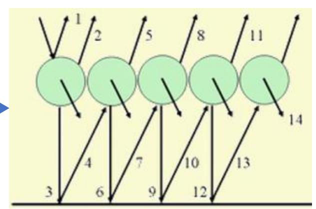
Radiative transfer models

Leaf RT models

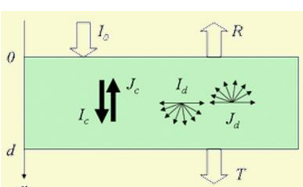
layers



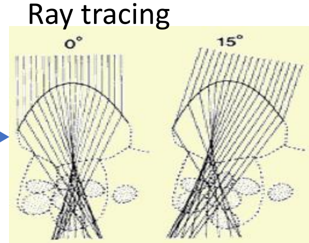
Compact spheres



N-fluxes



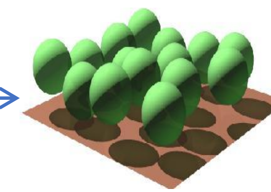
Ray tracing



Canopy RT models

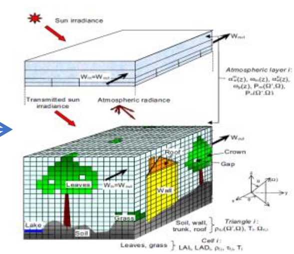
Turbid medium

Geometric



Hybrid

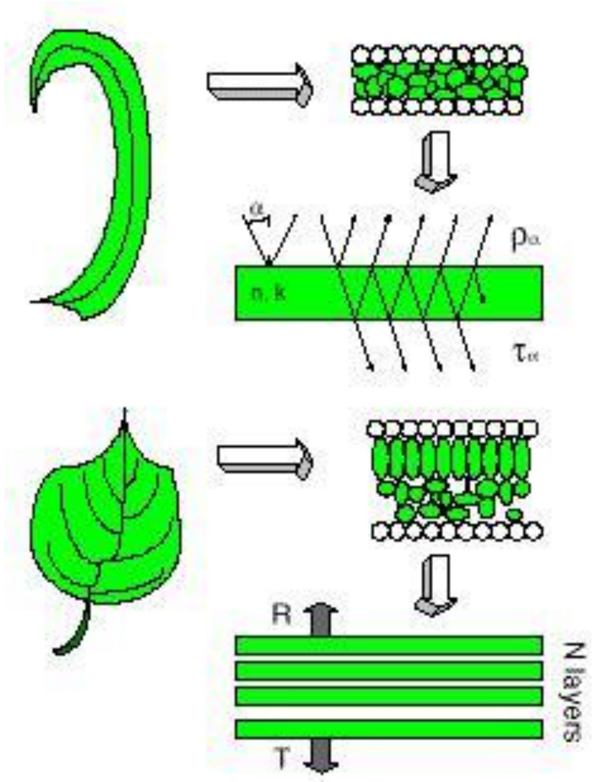
Volumetric

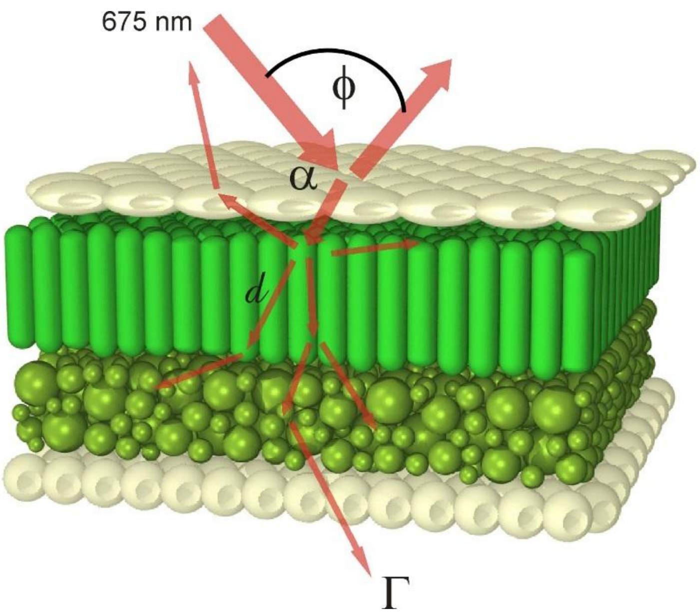


Multiple models exist with diverse complexity.

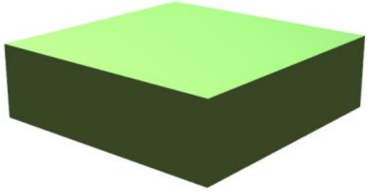
Leaf optical models

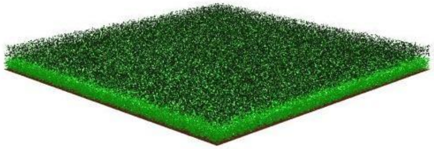
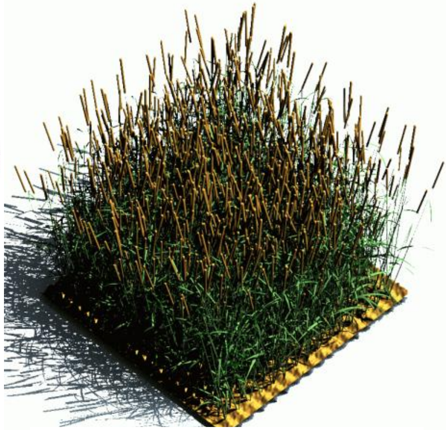
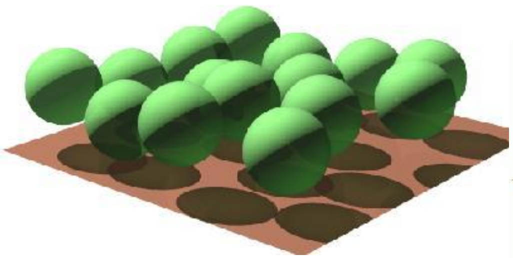
- A leaf is not opaque but transparent.
- Leaf as composed out of layers and empty spaces

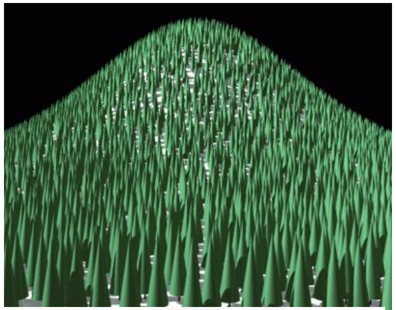
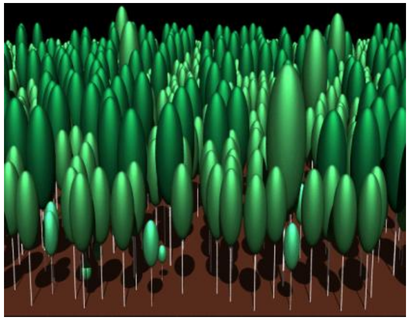


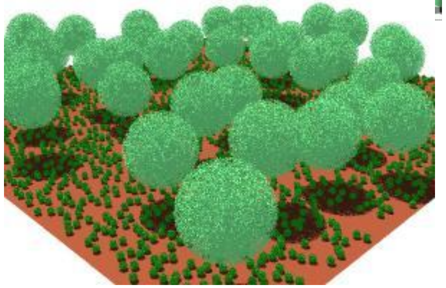
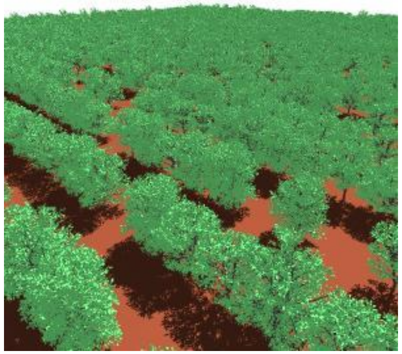


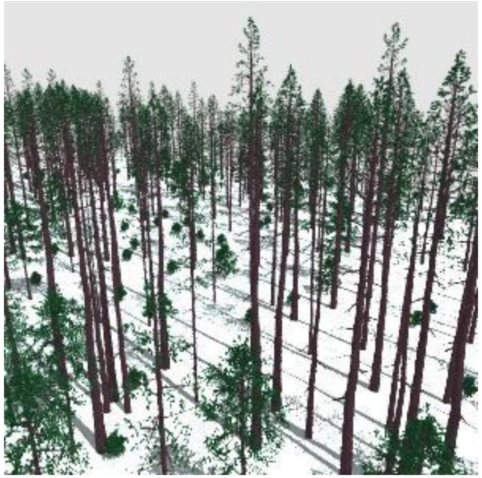
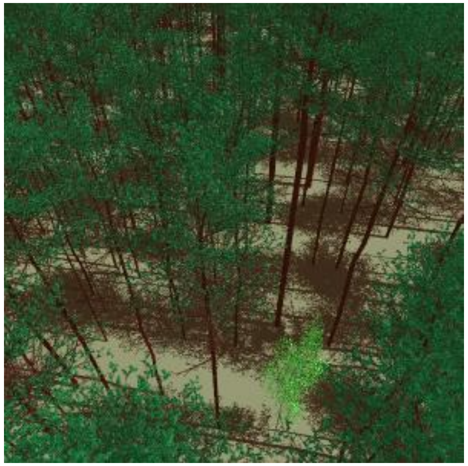
Canopy RTMs



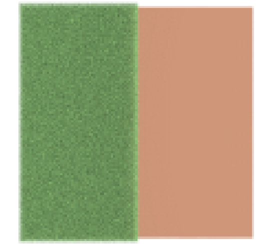



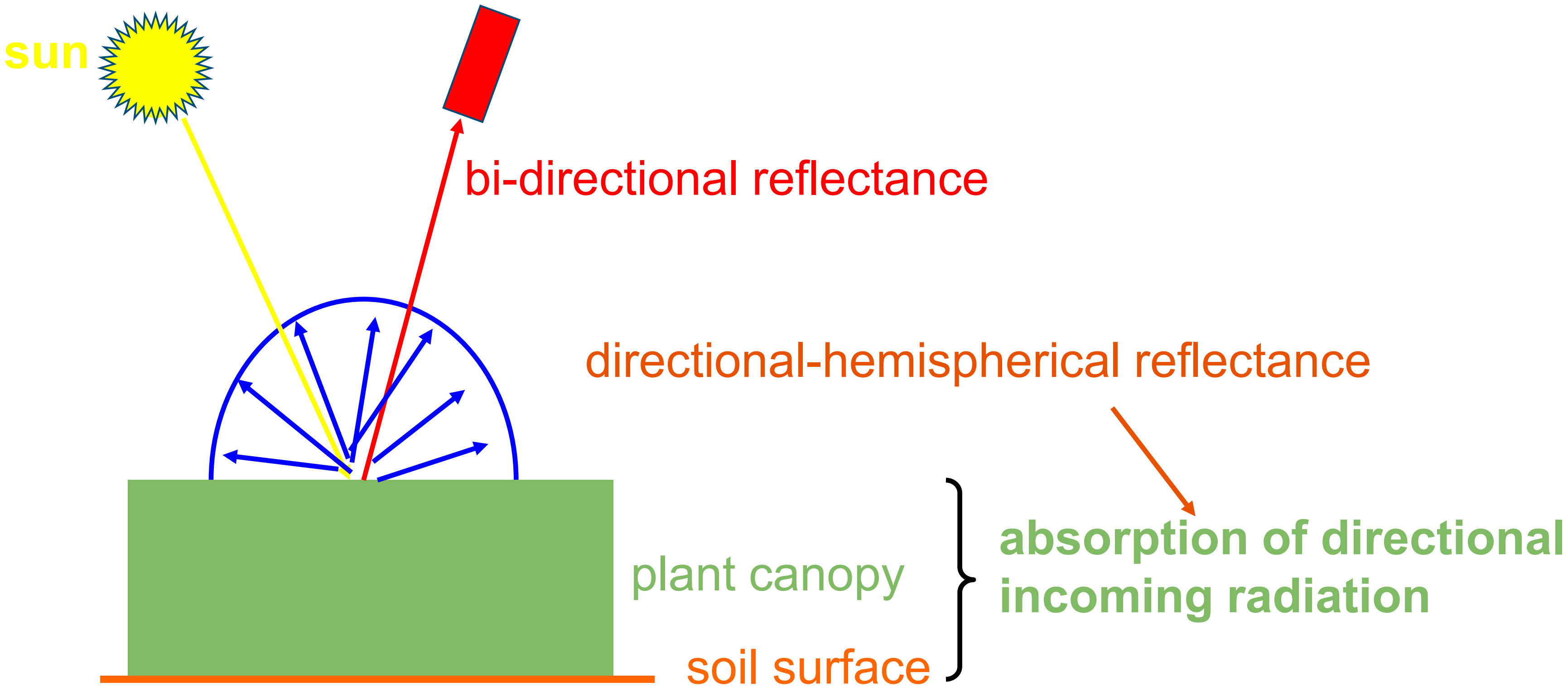




http://rami-benchmark.jrc.ec.europa.eu

Examples of canopy RTMs(1/4)



SAIL model (Verhoef 1984): a 1-D model

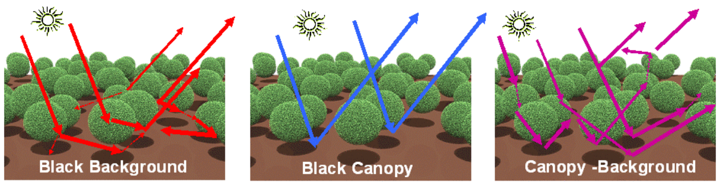


Examples of canopy RTMs (2/4)

Canopy models can be coupled with leaf, soil and atmospheric models

Solutions to the simpler problems

$$DHR(z_0,\mu_0;\rho_{sfc}) = DHR^{Collided}_{vegetation}(z_0,\mu_0;\rho_{sfc}) + \rho_{sfc}\left[\widetilde{DHR}^{Uncollided}_{background}(z_0,\mu_0)\right] + \rho_{sfc}\left[\widetilde{DHR}^{Collided}_{background}(z_0,\mu_0;\rho_{sfc})\right]$$



Black Background problem solved with a revisited version of a **standard 2-stream model**, e.g., Meador and Weaver (1980) using sets of scattering coefficients relevant to the case of vegetation canopies

Black Canopy problem solved with an analytical solution to :

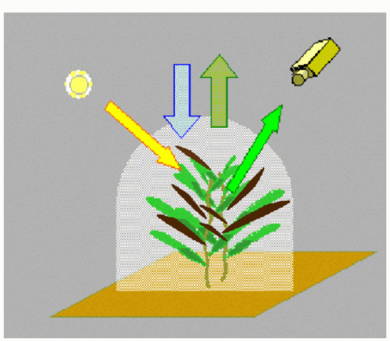
$$DHR_{BlackCanopy}(\mu_0) = \rho_{sfc}\exp\left(-\frac{\widetilde{LAI}}{2\mu_0}\right)\widetilde{T}_{blackCanopy}$$

Coupled Canopy-Background problem solved using **2-stream solutions** in the cases of an isotropic source at the top (diffuse sky) and the bottom (Lambertian background)

Ref: Pinty et al. (2006) Journal of geophysical Research, doi:10,1029/2005JD005952

Comparison with modelled spectral / directional reflectances using 4SAIL2

(Verhoef & Bach 2003)



Four-stream canopy reflectance model:
1. Direct solar flux
2. Diffuse downward flux
3. Diffuse upward flux
4. Direct observed flux (radiance)

Input parameters to 4SAIL2:

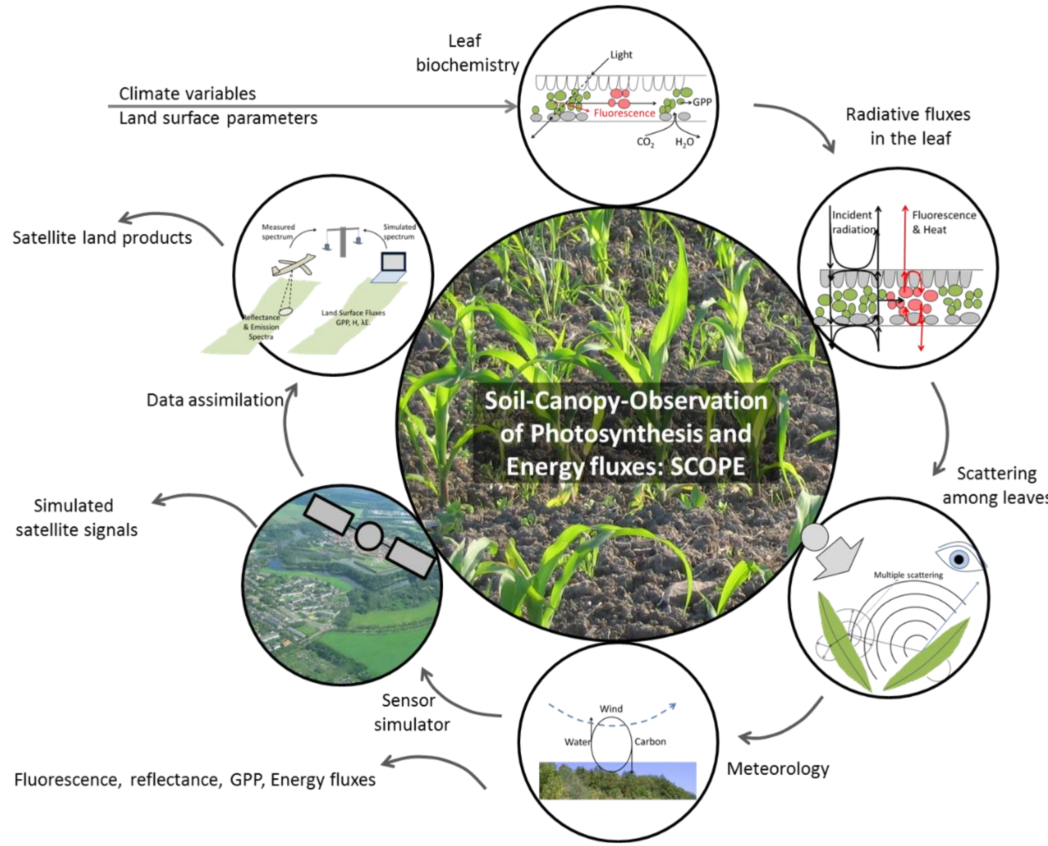
structural
- LAI - leaf area index
- Average leaf slope parameter a
- LIDF bimodality parameter b
- Hot spot parameter q
- Fraction brown leaf area fB
- Layer dissociation factor D
- Soil BRDF Parameters (b, c, B0, h)
- Soil moisture
- Crown coverage

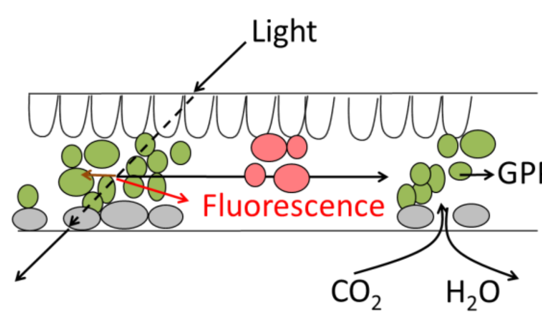
spectral
- Outputs from PROSPECT
- Fraction diffuse sky irradiance
- Dry soil reflectance

observational
- Solar zenith angle
- Viewing zenith angle
- Relative azimuth angle

Input parameters to PROSPECT:

leaf
- Leaf chlorophyll
- Leaf water
- Leaf dry matter
- Leaf mesophyll structure N





Examples of canopy RTMs (3/4)

Ray tracing models

Drat -the aDvanced Radiometric Ray Tracer.

P. Lewis, 1999; Saich et al., 2001. University College, Dept. Geography, London

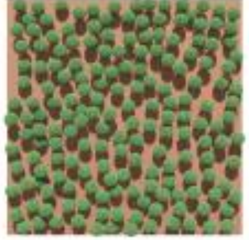
Vegetation is built using The Botanical Plant Modelling System (BPMS)
BPMS is a form of L-systems - the branches of a tree as geometric primitives

ARARAT - the advanced radiometric ray tracer,
reverce ray tracing,
a variety of camera models implemented

(Dürer, 1525)

(http://www.geog.ucl.ac.uk/~plewis/)

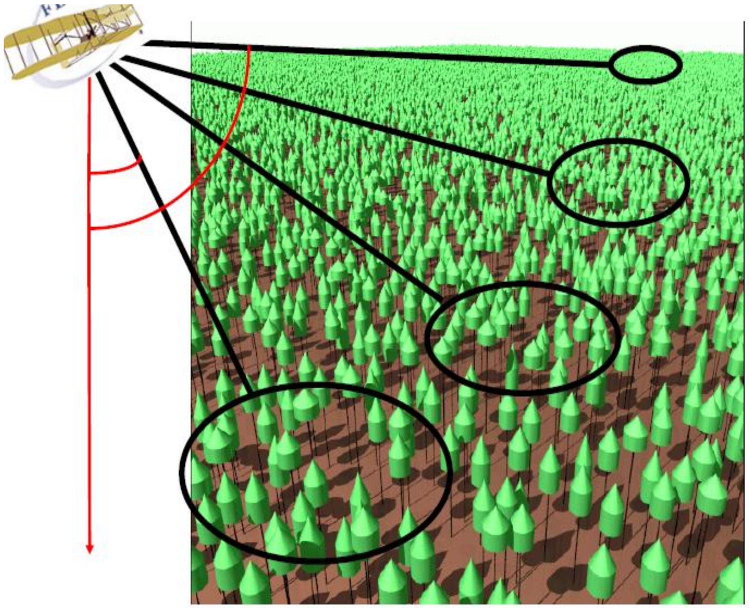
Examples of canopy RTMs (4/4)



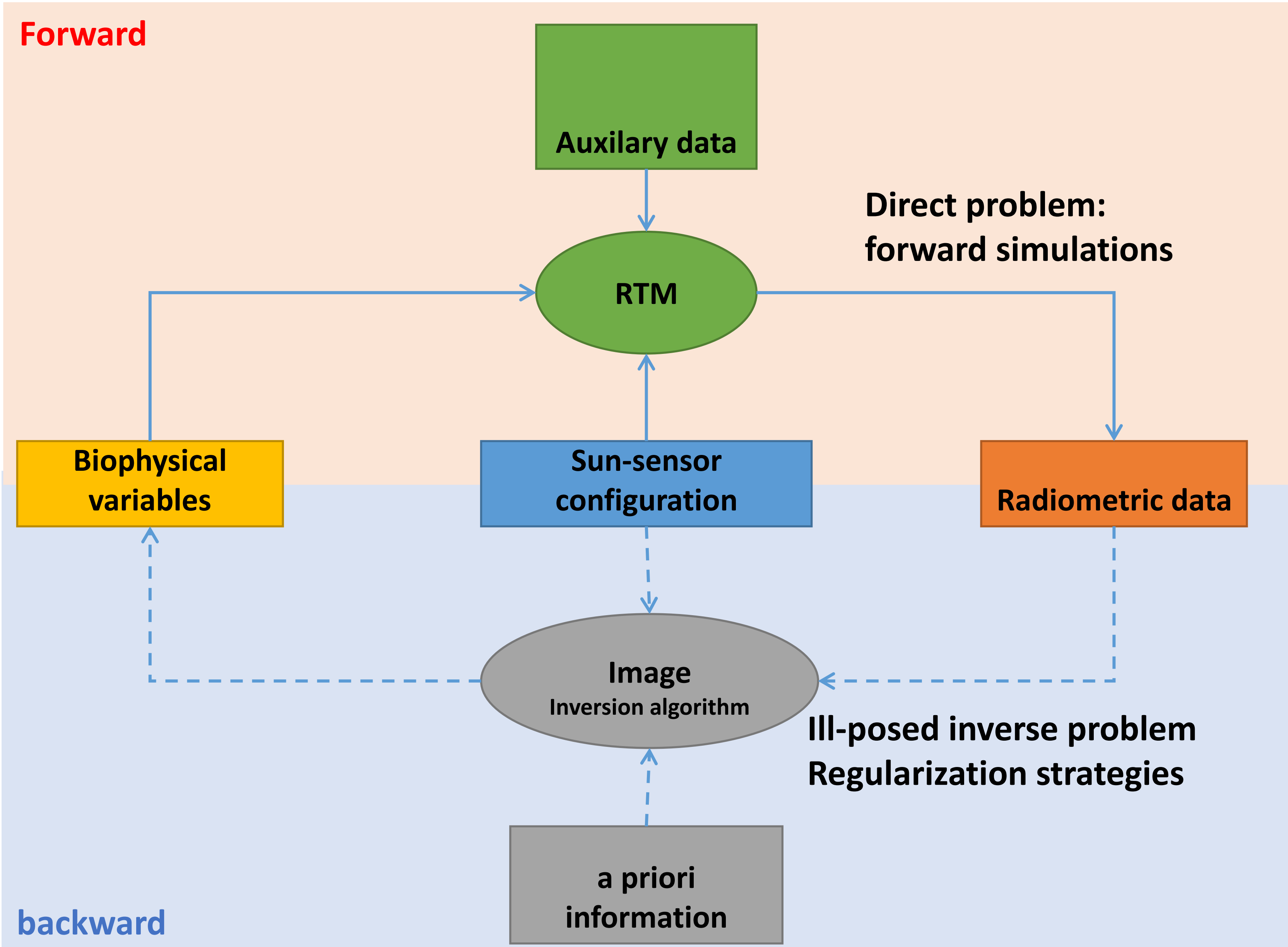
FLIGHT (North, 1996): A 3-D model

FLIGHT MC ray tracing approach

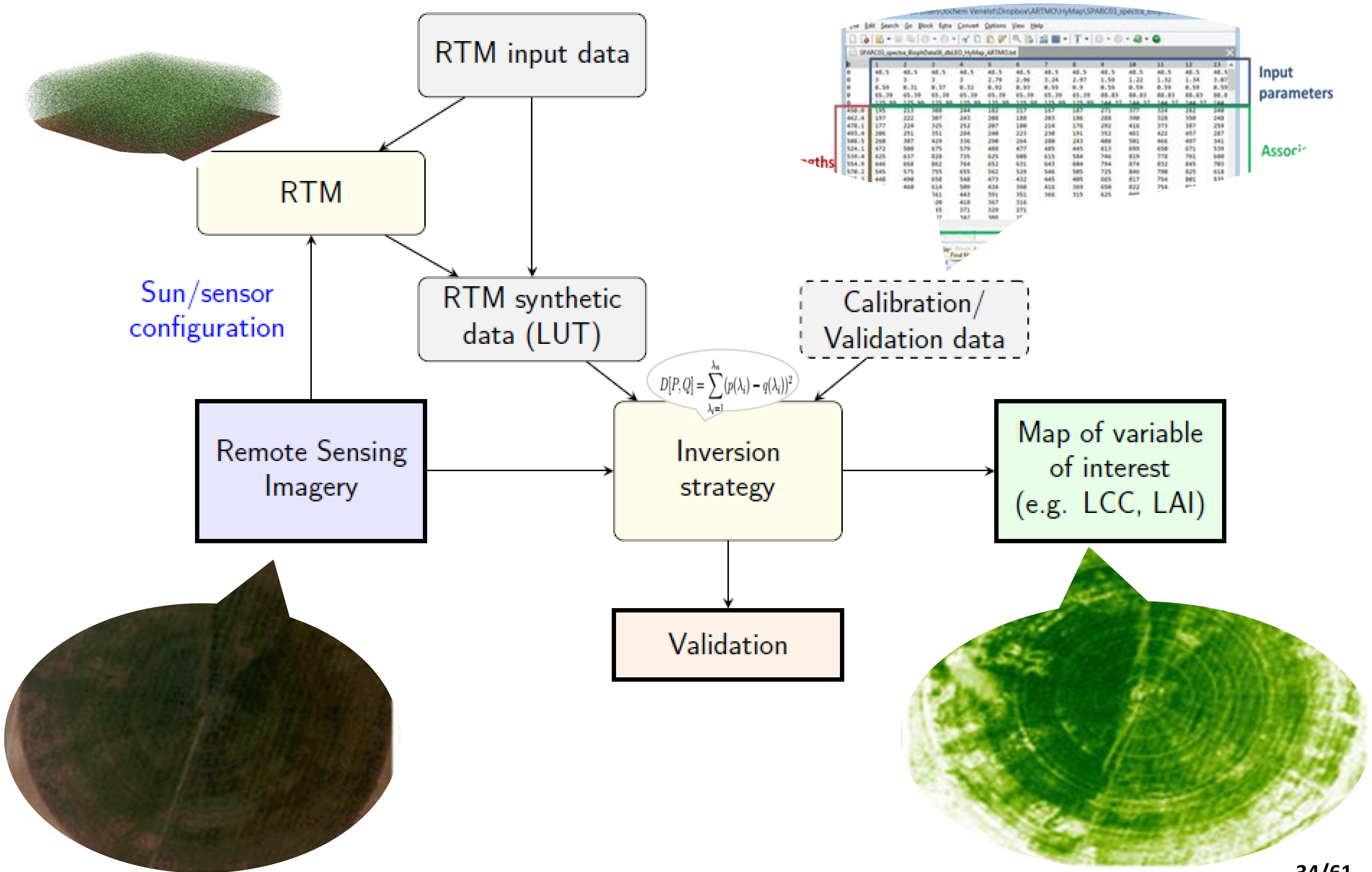
- Large scale structure by geometric primitives (e.g. cone)
- Foliage within crowns described by volume-averaged parameters
- 3D photon trajectories are simulated, accounting for the probabilities of free path, absorption and scattering
- Individual photon trajectories are traced from a solar source, through successive interactions, to a predetermined sensor view angle.



Biophysical parameters retrieval through RTM inversion:



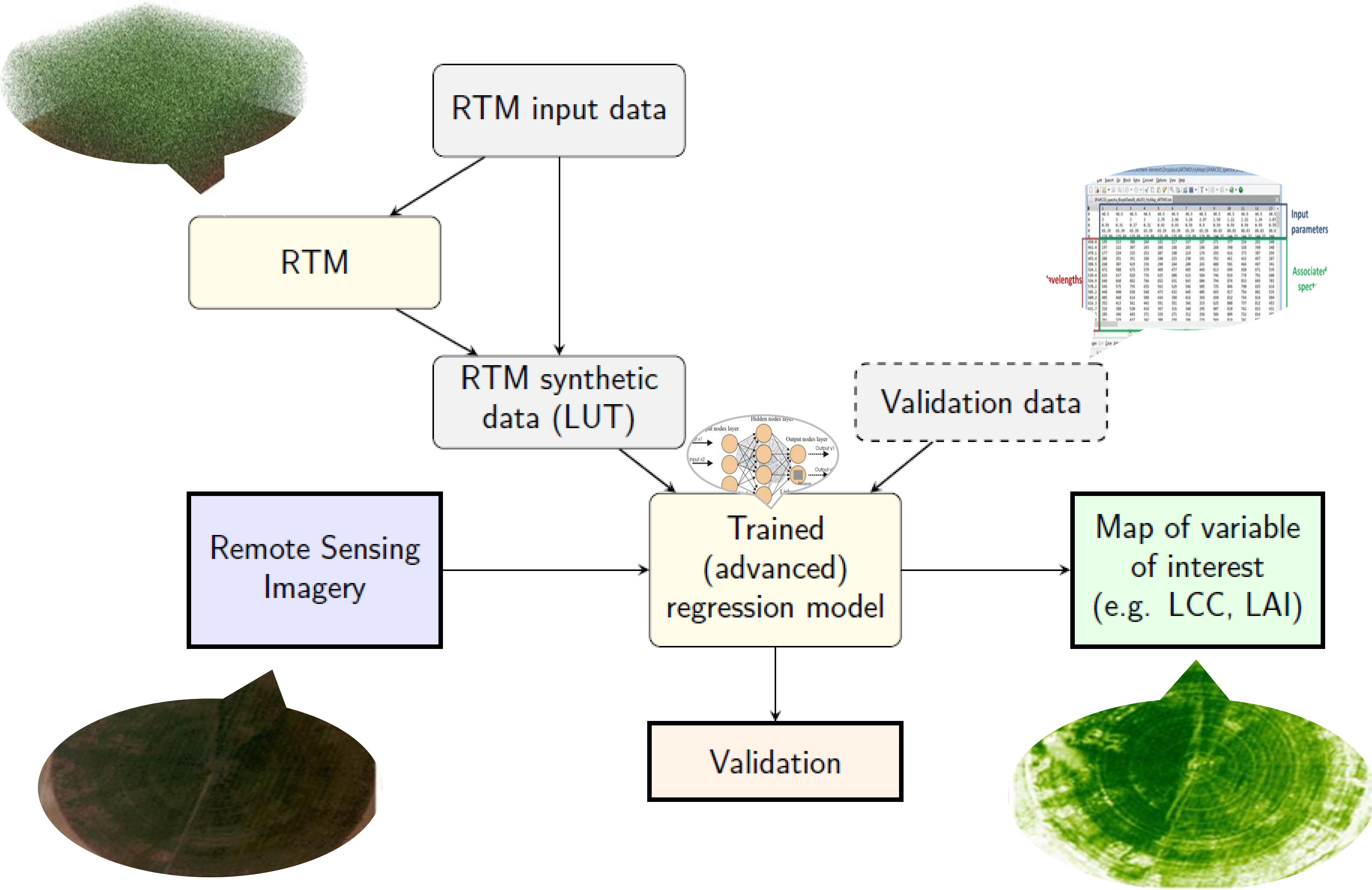
LUT-based RTM inversion



LUT-based inversion:

Strengths ☺	Weaknesses ☹
• **Full-spectrum methods** • **Reputation of physically-based** (however, note the impact of regularization factors) • Generally and **globally applicable** (e.g. MODIS products). • Capability to provide **multiple outputs** • Yields additional information about **uncertainty** of the retrievals (e.g. residuals).	• **Computationally demanding** due to the per-pixel based approach (however, solutions based on a priori information have been developed). • Retrieval quality depends on the **quality of the RT mo**dels, prior knowledge and regularization. • **Quite complex approach**: requires parameterization and optimization procedures. • The imposed upper/lower boundaries in the LUT have a logical consequence in that estimated variables cannot go beyond the boundaries imposed. This contradicts somewhat the physical approach as the prior information has an overwhelming influence. • LUT-based inversion methods are often strongly affected by noise and measurement uncertainty.

Hybrid retrieval

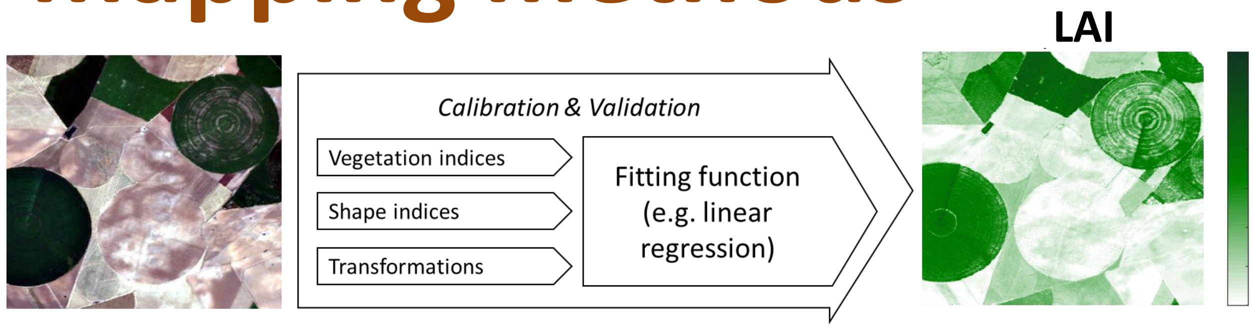


Non-parametric regression: hybrid

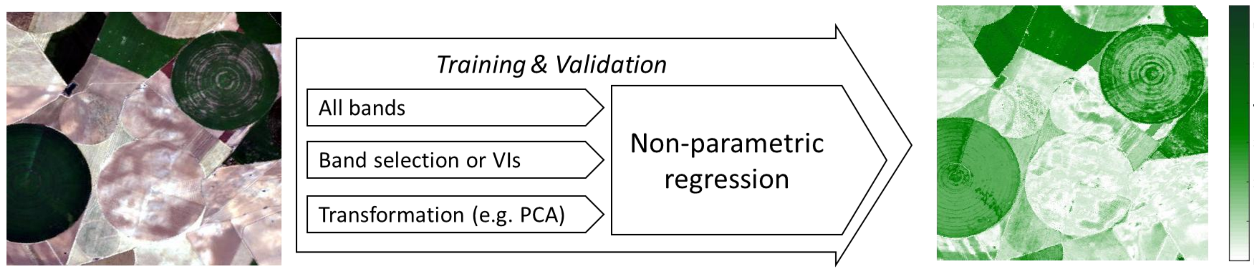
Strengths ☺	Weaknesses ☹
• Full-spectrum methods. They make use of the complete spectral information. • Advanced, adaptive (non-linear) models are built. • Methodologically, accurate and robust performance is enabled. • Some MLRAs cope well with datasets showing redundancy and high noise levels. • Once trained, imagery can be processed time efficient. • Some of the non-parametric methods (e.g. ANNs, decision trees) can be trained with a high number of samples (typically >1,000,000). • Some MLRAs provide insight in model development (e.g. GPR: relevant bands; decision trees: model structure). • Some MLRAs can provide multiple-outputs (e.g. PLRS, ANN, SVR, GPR and KRR) • Some MLRAs provide uncertainty intervals (e.g. GPR).	• Training can be computational expensive. • Hypercomplex models can be generated. Their generic potential is limited and hence they do not generalize well, based on the training data (problem of over-fitting). • Some regression algorithms are difficult (or even impossible) to train with a high number of samples. • Expert knowledge is required, e.g. for tuning. However, toolboxes exist automating some of the steps in this sub-process. • Some of the methods can be considered as black boxes. • Some regression algorithms elicit instability when applied with datasets statistically deviating from the datasets used for training.

Summary mapping methods

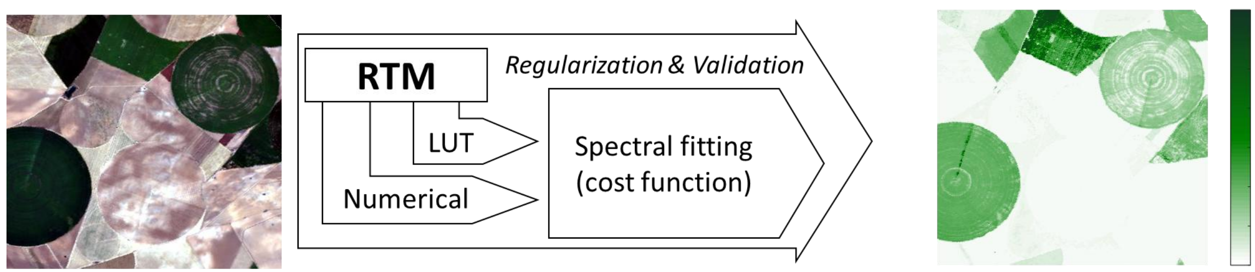
Parametric regression



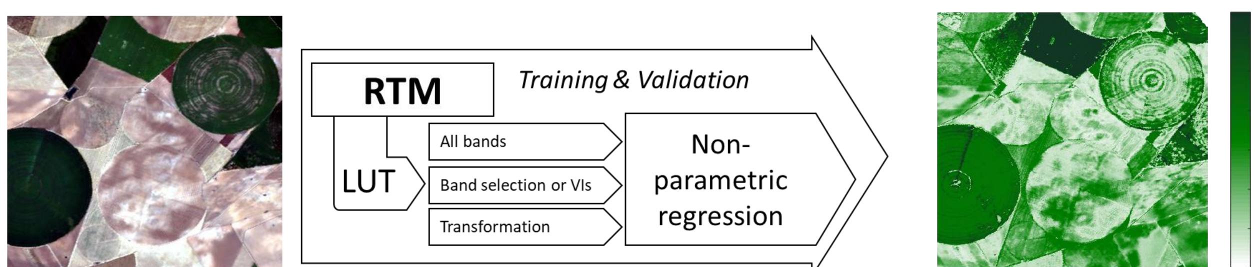
Non-parametric regression



RTM inversion

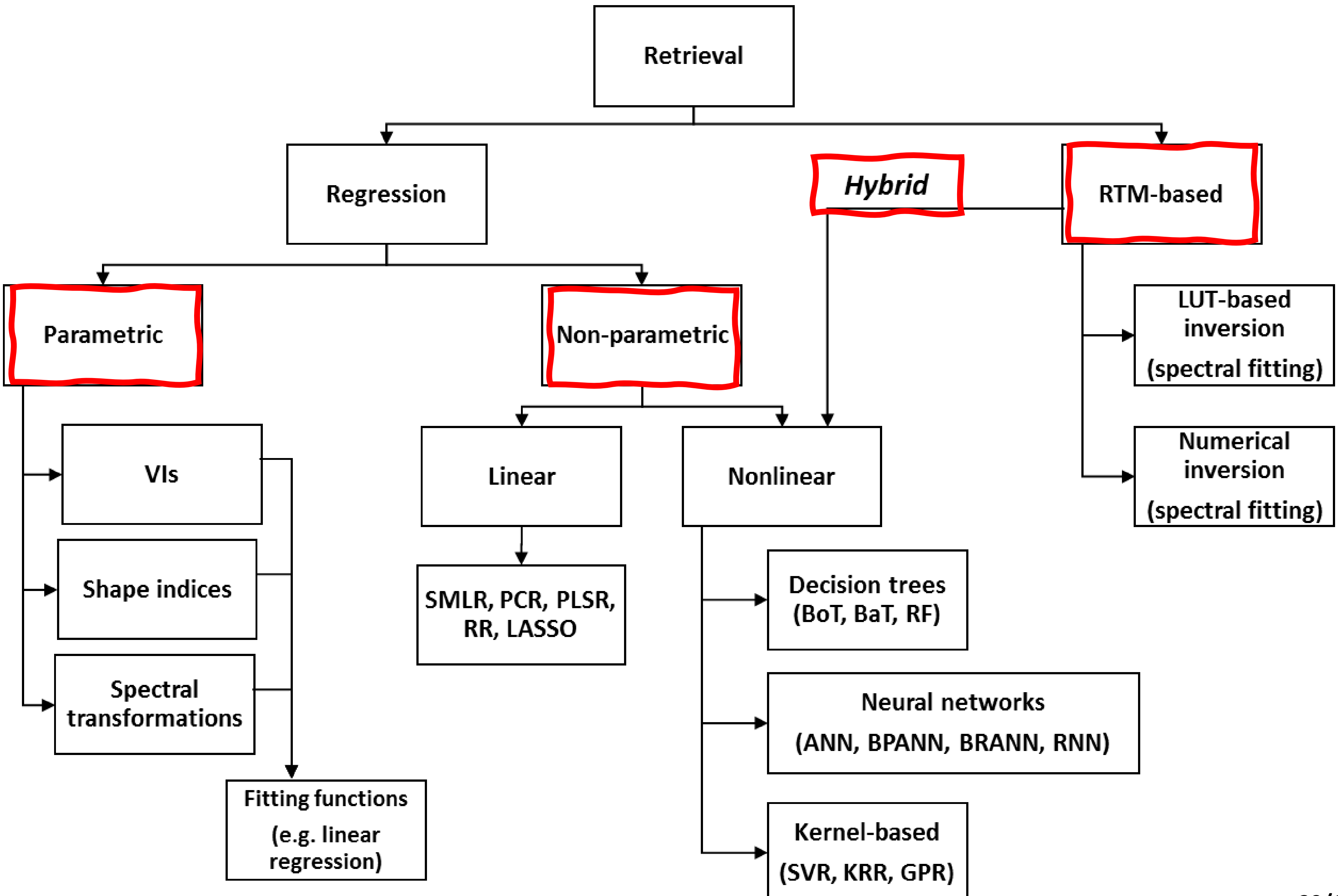


Hybrid regression



Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegory, J.P., Lewis, P., Moreno, J. (2018). Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surveys in Geophysics,

Taxonomy retrieval methods



Optimizing retrieval

Spoectral indices

- Band combination?
- Formulation?
- Fitting function?

Statistical methods

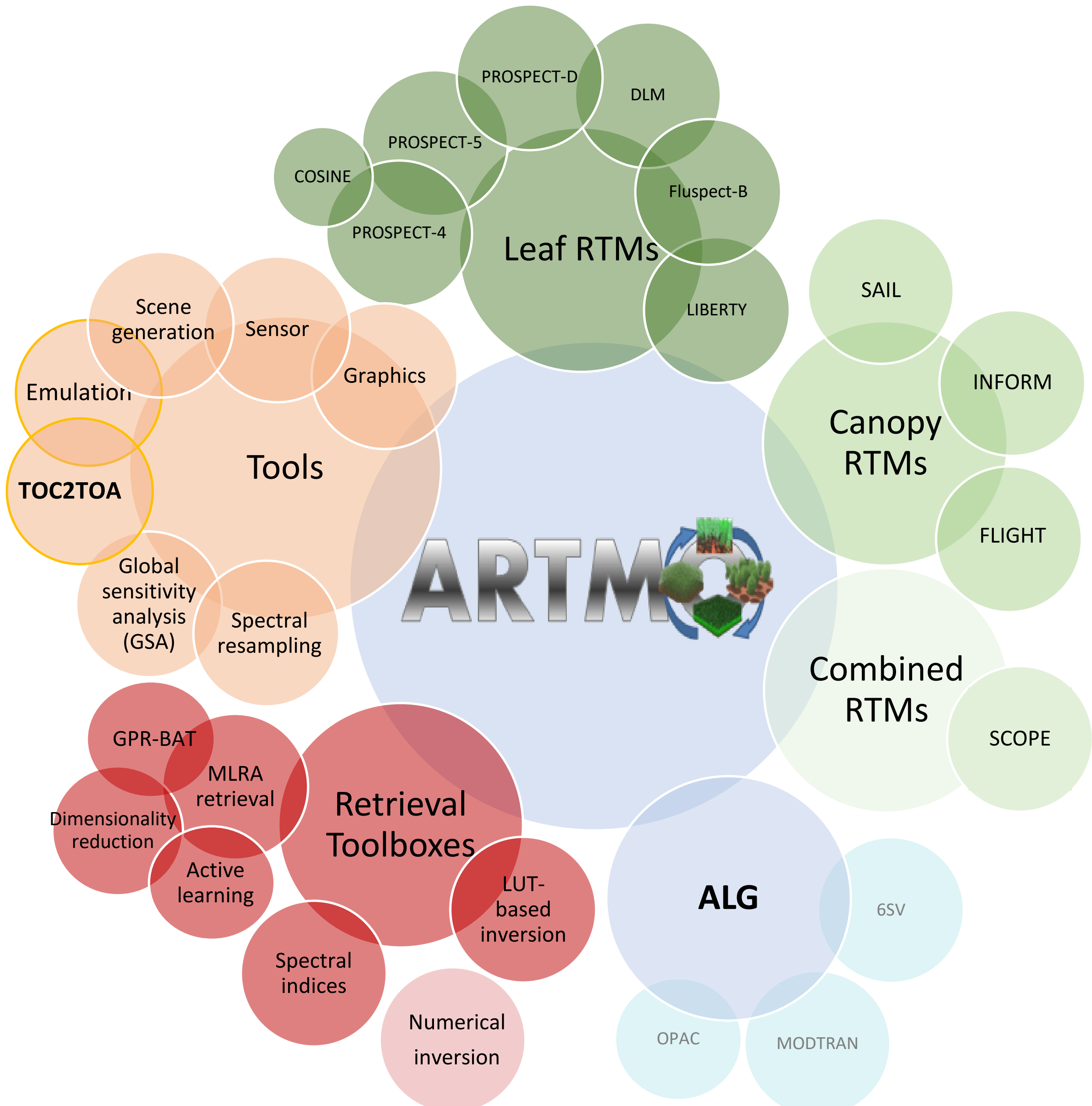
- Parametric vs non-parametric?
- Training vs validation data?

LUT-based appraoch

- Parametrization
- Cost function?
- A priori info?
- Regularization options?

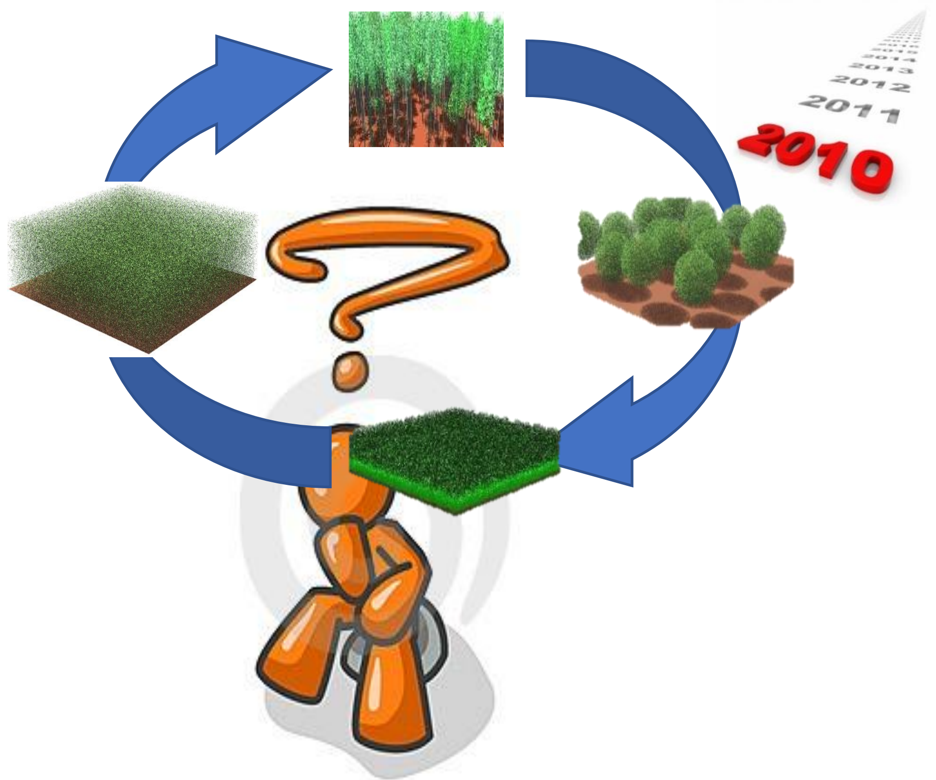
The many decisions to be taken require a systematic evaluation

ARTMO automates retrieval optimization

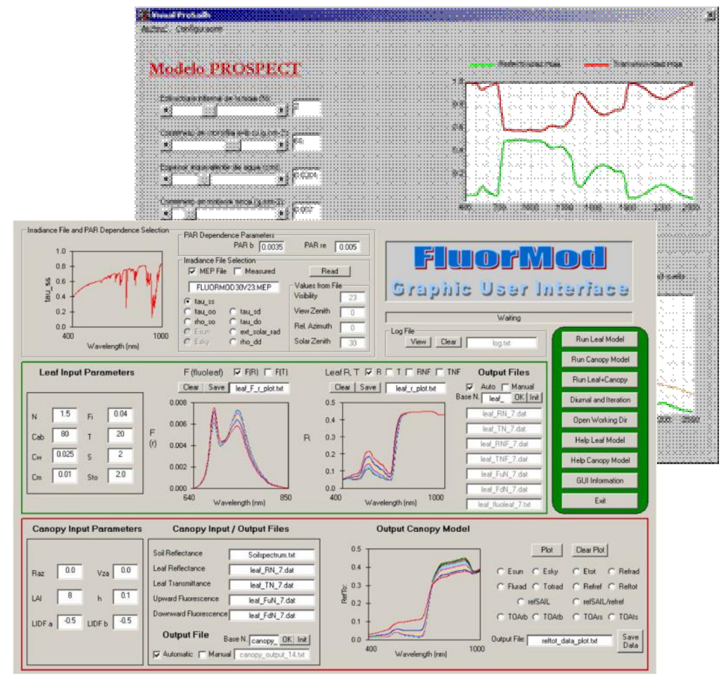

PROSPECT-D
DLM
PROSPECT-5
COSINE
Fluspect-B
PROSPECT-4
Leaf RTMs
LIBERTY
SAIL
INFORM
Canopy RTMs
FLIGHT
Combined RTMs
SCOPE
Scene generation
Sensor
Graphics
Emulation
Tools
TOC2TOA
Global sensitivity analysis (GSA)
Spectral resampling
ARTMO
GPR-BAT
MLRA retrieval
Dimensionality reduction
Active learning
Retrieval Toolboxes
LUT-based inversion
Spectral indices
Numerical inversion
ALG
6SV
OPAC
MODTRAN

RTMs are important tools in EO research but for the broader community these models are perceived as complicated. Only very few of them offer user-friendly interfaces (GUIs).

Which RTM to choose?



Only very few offer a GUI.



- **No interface exists that brings multiple RTMs together in one GUI.**
- **None of existing (publicly available) GUIs provide post-processing tools.**

To fill up this gap:

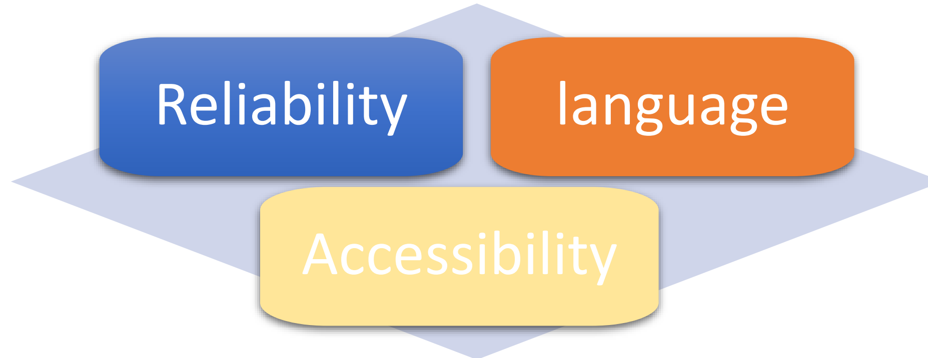
➢ **To develop a GUI toolbox that:**

- operates **various RTMs in an intuitive interface**
- provides a comprehensive **visualization** of model outputs
- works both for **multispectral and hyperspectral** data
- enables **to retrieve biophysical parameters** through various retrieval methods
- takes different **land cover classes** into account.

Toolbox for EO applications:

Automated
Radiative
Transfer
Models
Operator

Selection RTMs & programming language



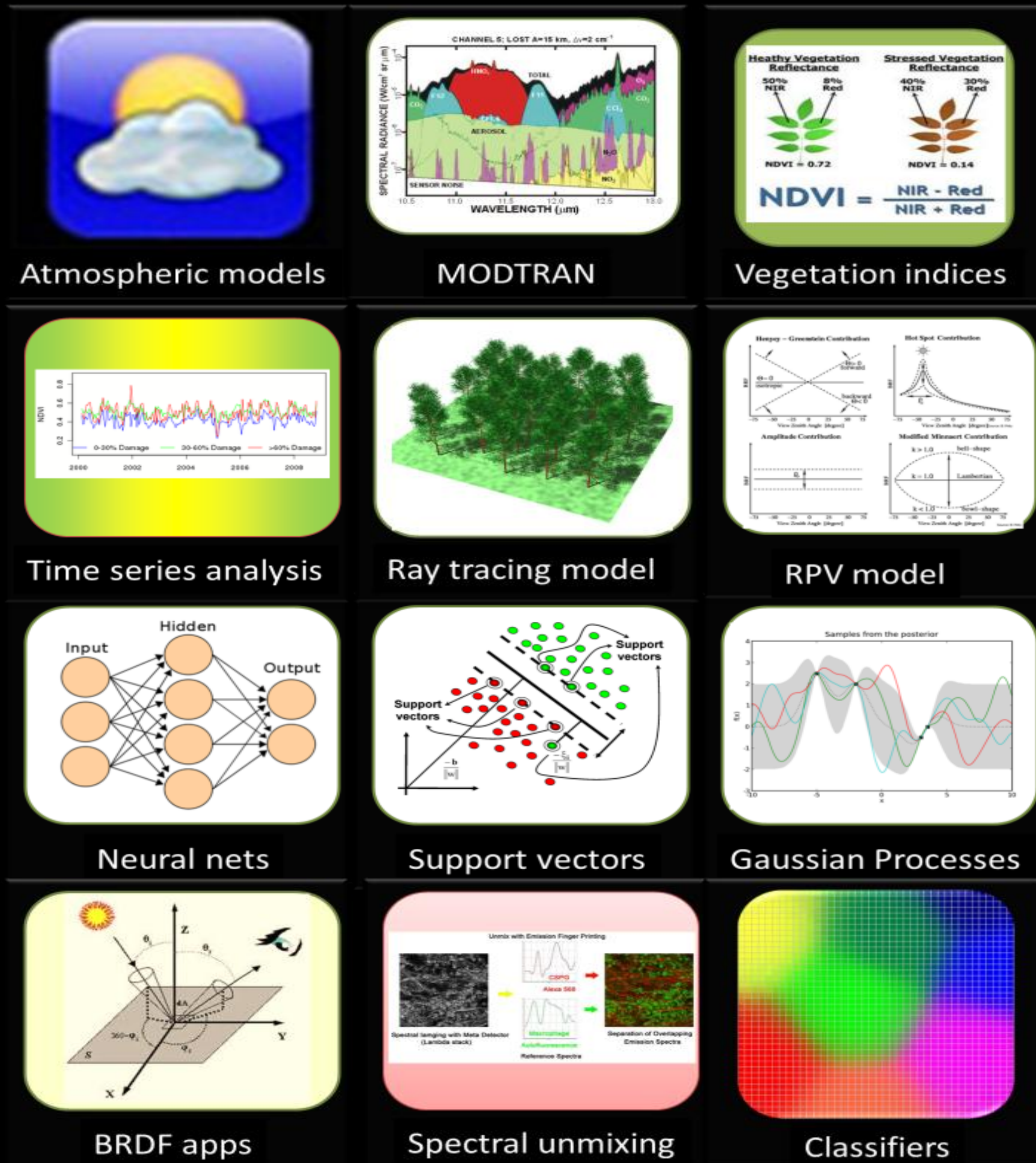
Model	Reference	Source code
PROSPECT-4	Feret et al., 2008	Matlab
PROSPECT-5	Feret et al., 2008	Matlab
PROSPECT-D	Feret et al., 2017	Matlab
DLM	Stuckens et al., 2009	Matlab
LIBERTY	Dawson et al., 1998	Matlab
FLUSPECT	Vilfan et al., 2016	Matlab
4SAIL	Verhoef et al., 2007	Matlab
FLIGHT	North, 1996	Executable file
INFORM	Atzberger, 2000	Matlab
SCOPE	Van der Tol et al., 2009	Matlab

Software packages:

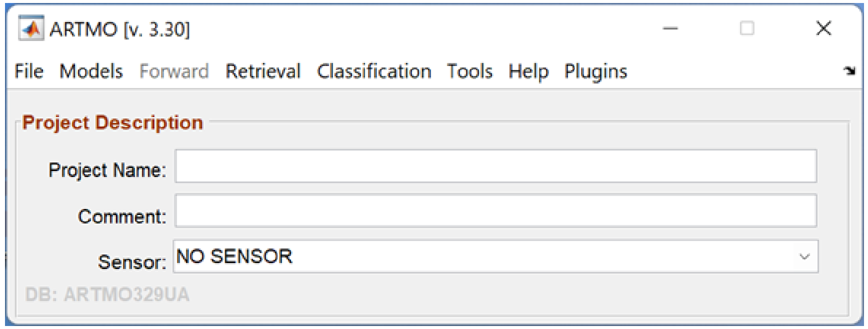
Programming language: Matlab®
Database: MySQL®
Image processing software: ENVI®

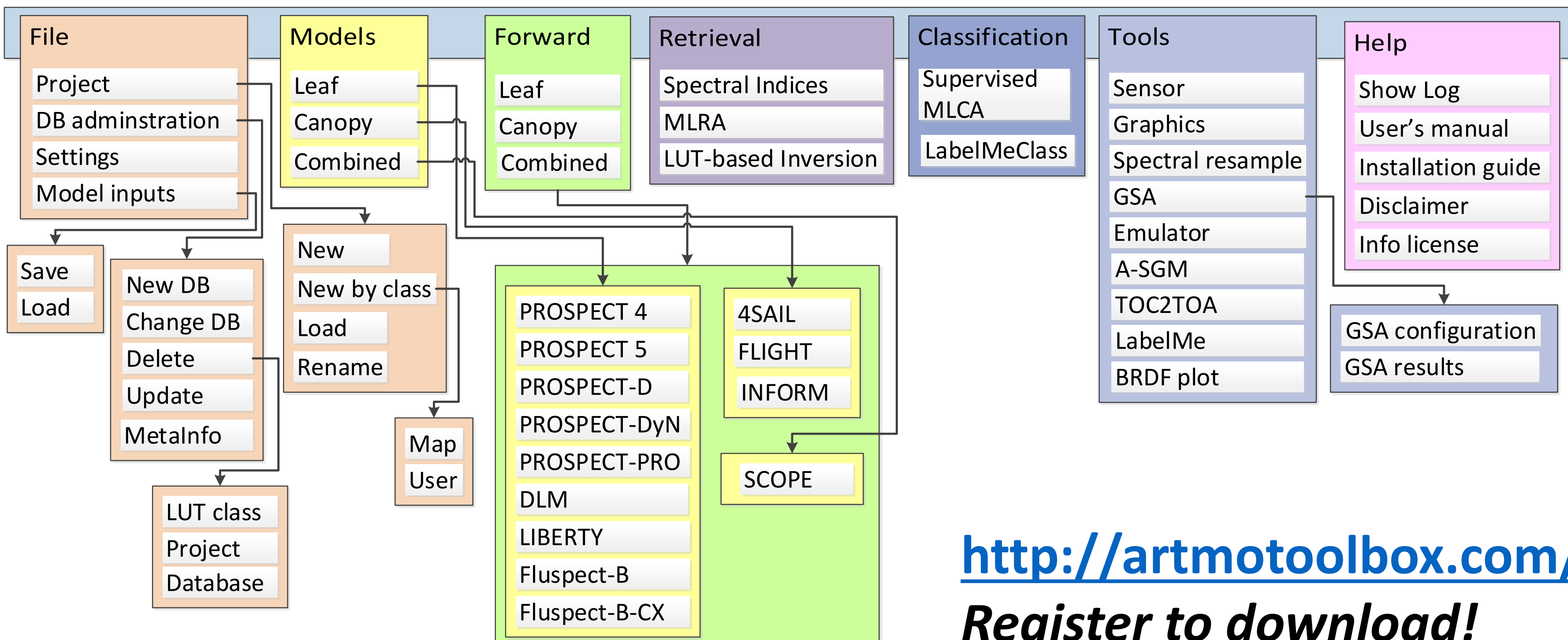
Leaf RTM
Canopy RTM
Combined RTM

ARTMO


Atmospheric models
MODTRAN
Vegetation indices
Time series analysis
Ray tracing model
RPV model
Neural nets
Support vectors
Gaussian Processes
BRDF apps
Spectral unmixing
Classifiers
Input
Hidden
Output
Support vectors
Support vectors
NDVI = NIR - Red / NIR + Red
Heathy Vegetation Reflectance
Stressed Vegetation Reflectance
WAVELENGTH (μm)

ARTMO v. 3.33



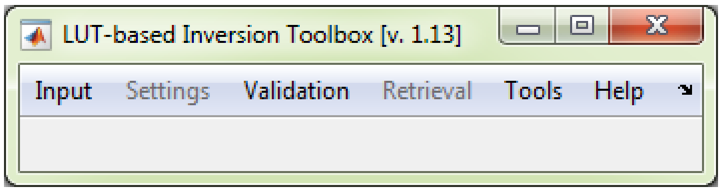


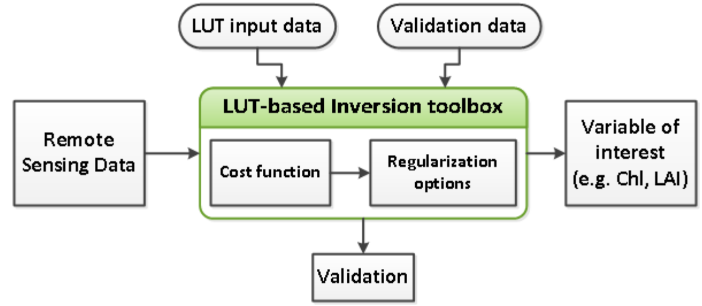
http://artmotoolbox.com/

Register to download!

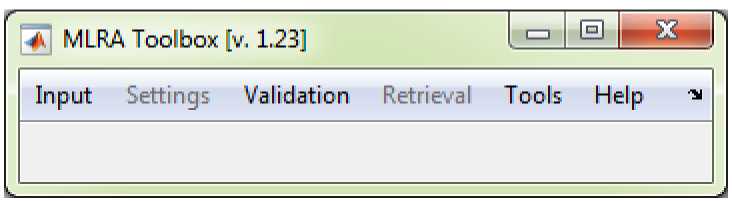
ARTMO's retrieval toolboxes:

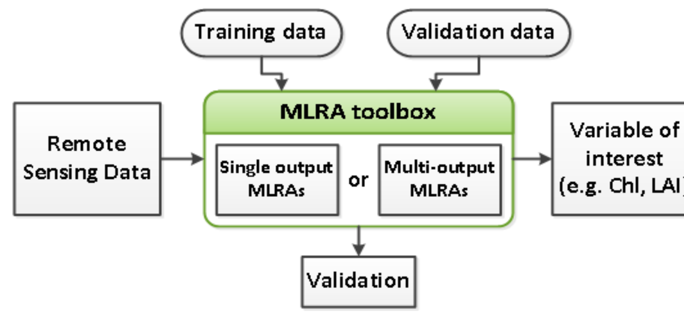
LUT-based inversion toolbox



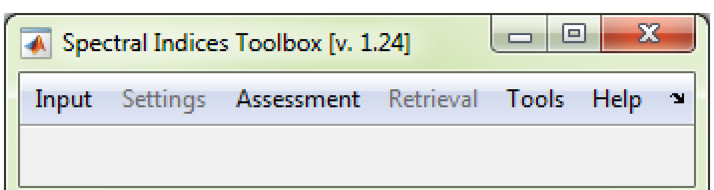


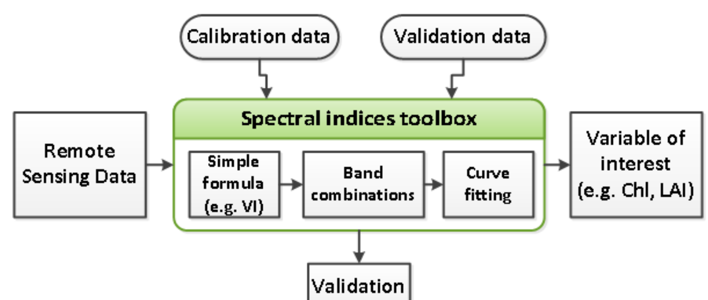
Machine learning regression algorithm toolbox (MLRA)





Spectral indices toolbox





Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison

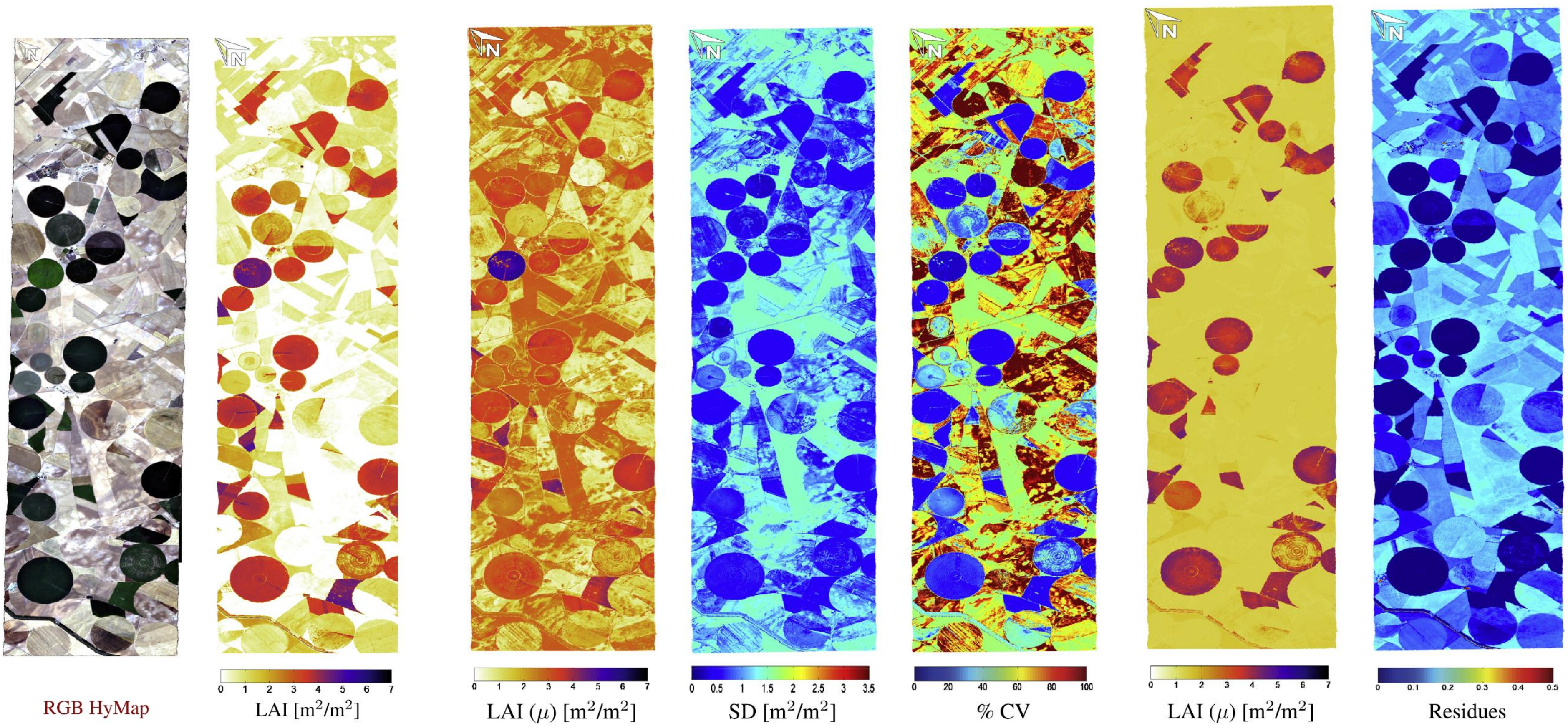
Jochem Verrelst [a,*], Juan Pablo Rivera [a], Frank Veroustraete [b], Jordi Muñoz-Marí [a], Jan G.P.W. Clevers [c], Gustau Camps-Valls [a], José Moreno [a]

[a] Image Processing Laboratory (IPL), Universitat de València, València, Spain
[b] Department of Bioscience Engineering, Faculty of Sciences, University of Antwerp, Antwerp, Belgium
[c] Laboratory of Geo-information Science and Remote Sensing, Wageningen University, Wageningen, The Netherlands

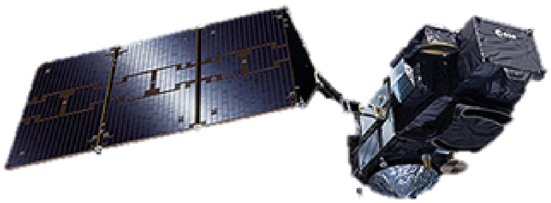
Table 9
Best performing parametric, non-parametric and LUT-based inversion method and associated mapping speed.

Retrieval algorithm	*RMSE*	R^2	Mapping speed (s)
Tian 3-band formulation	0.615	0.823	3.847
VH-GPR	0.436	0.902	73.884
Pearson chi-square inversion	0.802	0.745	3706.965

VI **GPR** **Inversion**



Operational processing?



Characteristic	Parametric	Non-parametric	RTM-based	Hybrid
Generalization capacity	--	-	++	++
Mapping Speed	++	+	--	+
Uncertainties	--	++*	+	++*
Accuracy	+	++	+	++
Variables	++	++	+	+

* Some machine learning methods (e.g. probabilistic or ensemble methods)

Verrelst, J. et al. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273-290.

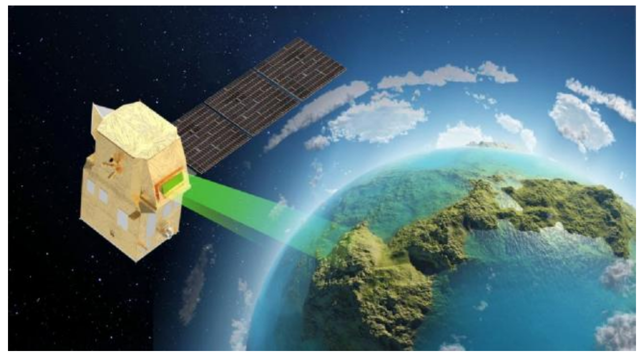
CHIME: Copernicus Hyperspectral Imaging Mission

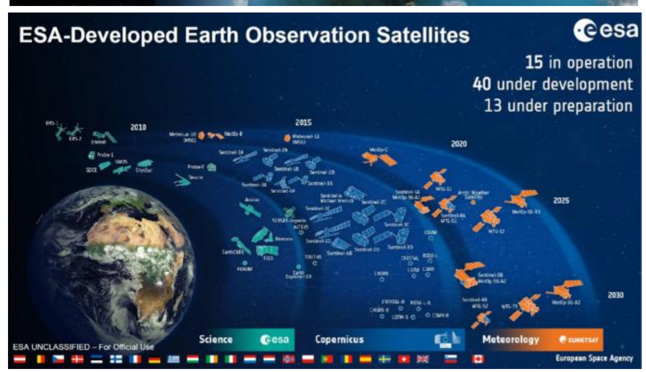
CHIME will carry a visible to shortwave infrared spectrometer to provide global routine hyperspectral observations. The mission will support new and enhanced services for sustainable agricultural and biodiversity management, as well as soil property characterisation.

Technical concept:

Routine spectroscopic observation in contiguous spectral bands:

- Instrument: Pushbroom Imaging Spectrometer **400 – 2500 nm, Δλ <= 10nm**
- Revisit **10 – 15 days**
- GSD (spatial resolution): **30 m**
- **Sun synchronous orbit** (LTDN 10:30 – 11:30)
- **Nadir view** covering land and coastal areas
- High radiometric accuracy, low spectral/spatial misregistration





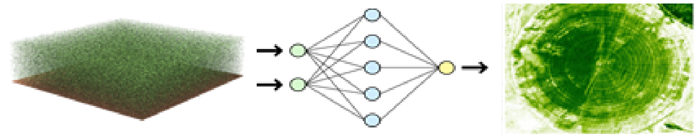
CORE Data Products:

The mission shall provide access to Level-1B, Level-1C and Level-2A products accessible via DIAS and with API support:

- Bottom-of-Atmosphere (BOA) reflectance (atmospherically corrected)
- Ortho rectified geometry
- Basic pixel classification (opaque clouds, thin clouds, cloud shadows, vegetation, water, snow etc.)
- **Additionally –> Vegetation products (Level-2B)**

http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Copernicus_High_Priority_Candidates

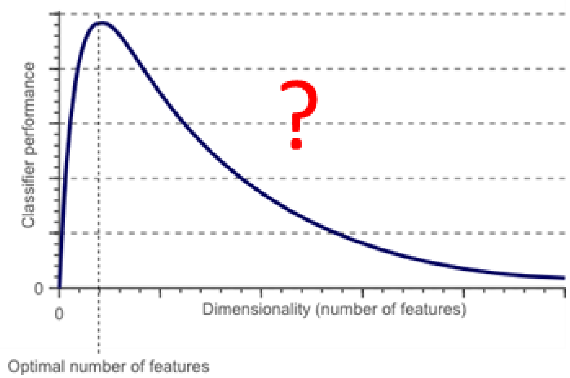
Hybrid methods: data-driven methods



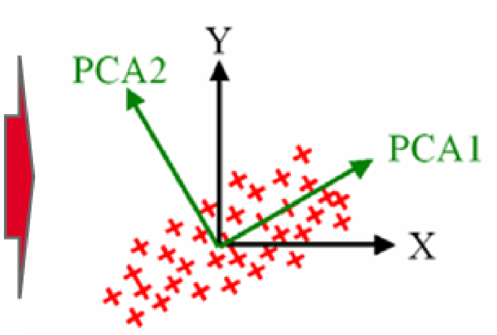
Many factors play a role. **Spectral variability**: **quantity & quality** & applied ML algorithm

Spectral domain

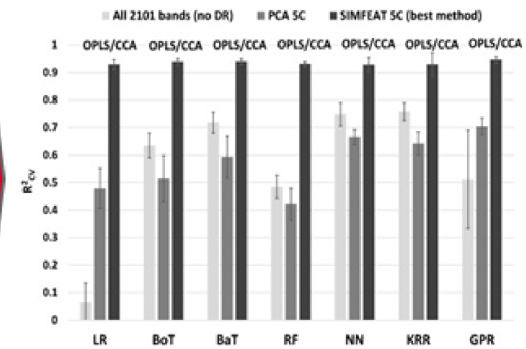
Curse of dimensionality



Dimensionality reduction



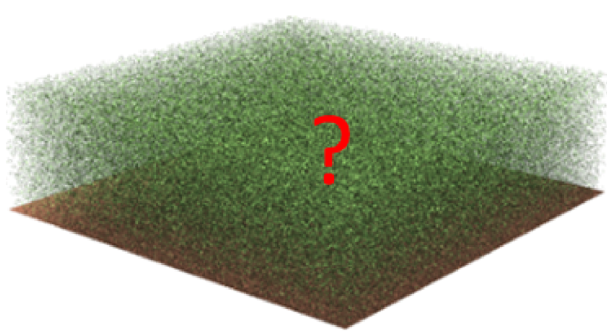
PCA > all bands



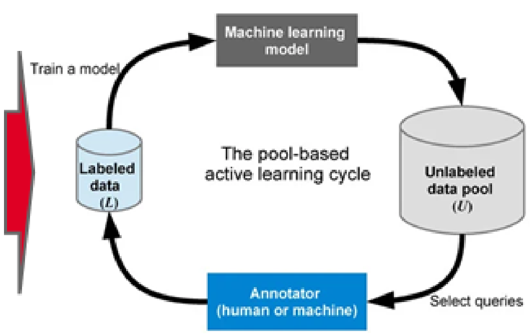
Rivera-Caicedo, J. P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., & Moreno, J. (2017). Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS journal of photogrammetry and remote sensing*, *132*, 88-101.

Sampling domain

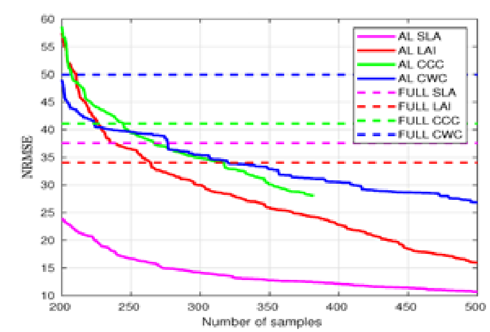
RTM sampling



Active learning



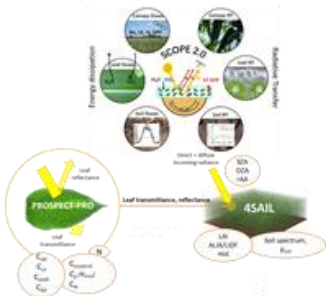
quality > quantity



Verrelst, J., Berger, K., & Rivera-Caicedo, J. P. (2020). Intelligent sampling for vegetation nitrogen mapping based on hybrid machine learning algorithms. *IEEE Geoscience and Remote Sensing Letters*, *18*(12), 2038-2042.

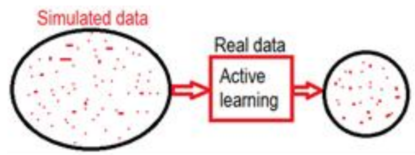
Workflow CHIME vegetation traits models:

RTMs



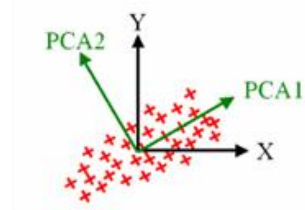
Use **RTMs** (e.g. **PROSPECT-PRO – SAIL**) to generate a LUT composed by pairs (e.g. 1000) of **vegetation parameters and spectra**.

Active learning



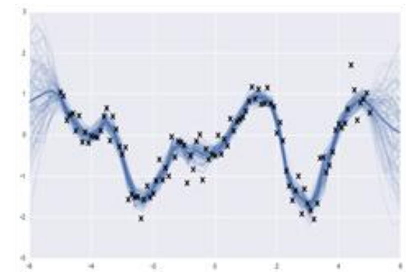
Select the **most representative samples** from the training dataset via a **diversity** or entropy criteria. Later, add non-vegetated spectra.

PCA



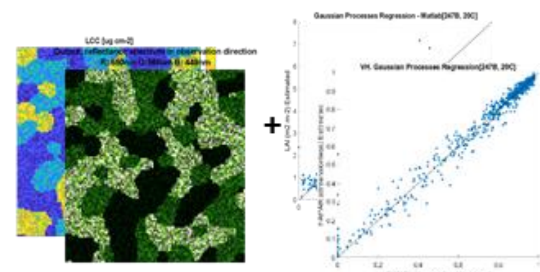
Dimensionality reduction with **PCA** (20 components).

Train GPR algorithms



With the **LUT optimized** for vegetation and non-vegetated surfaces, train probabilistic ML algorithms.

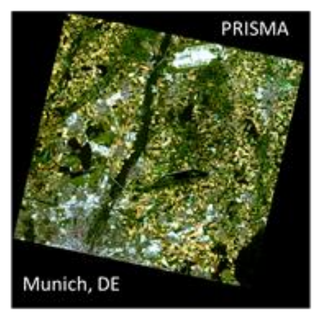
Validate the models



Assess models' performance against **field data** and vegetation **reference scenes**.

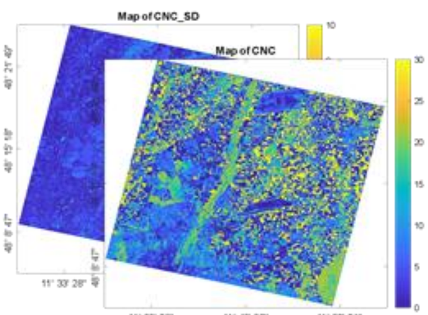
Okay?

Apply to new observations



PRISMA images resampled to **CHIME** band settings

Maps + uncertainties

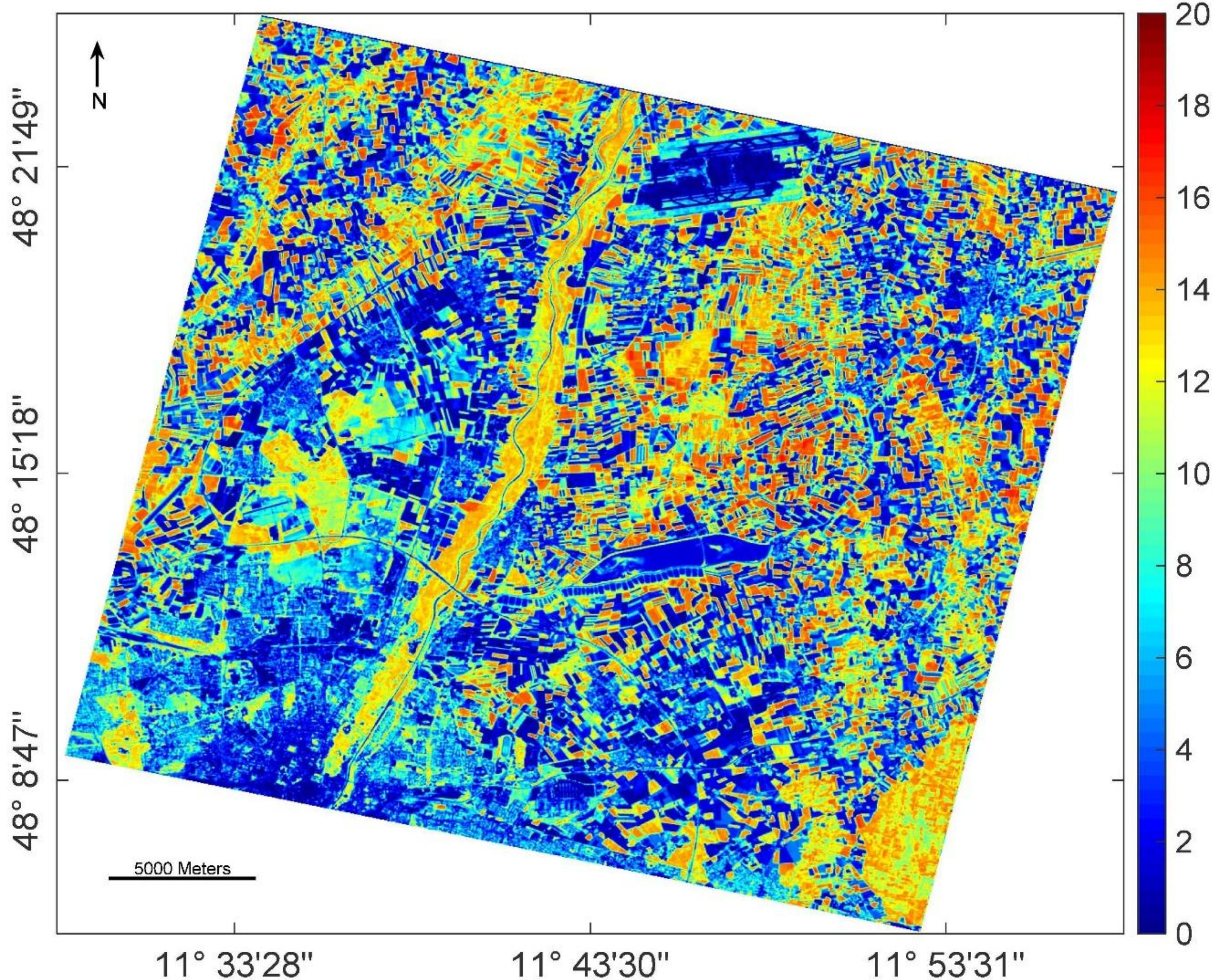


Final outputs of the workflow.

Verification and Validation vegetation models through E2E using CHIME-like data

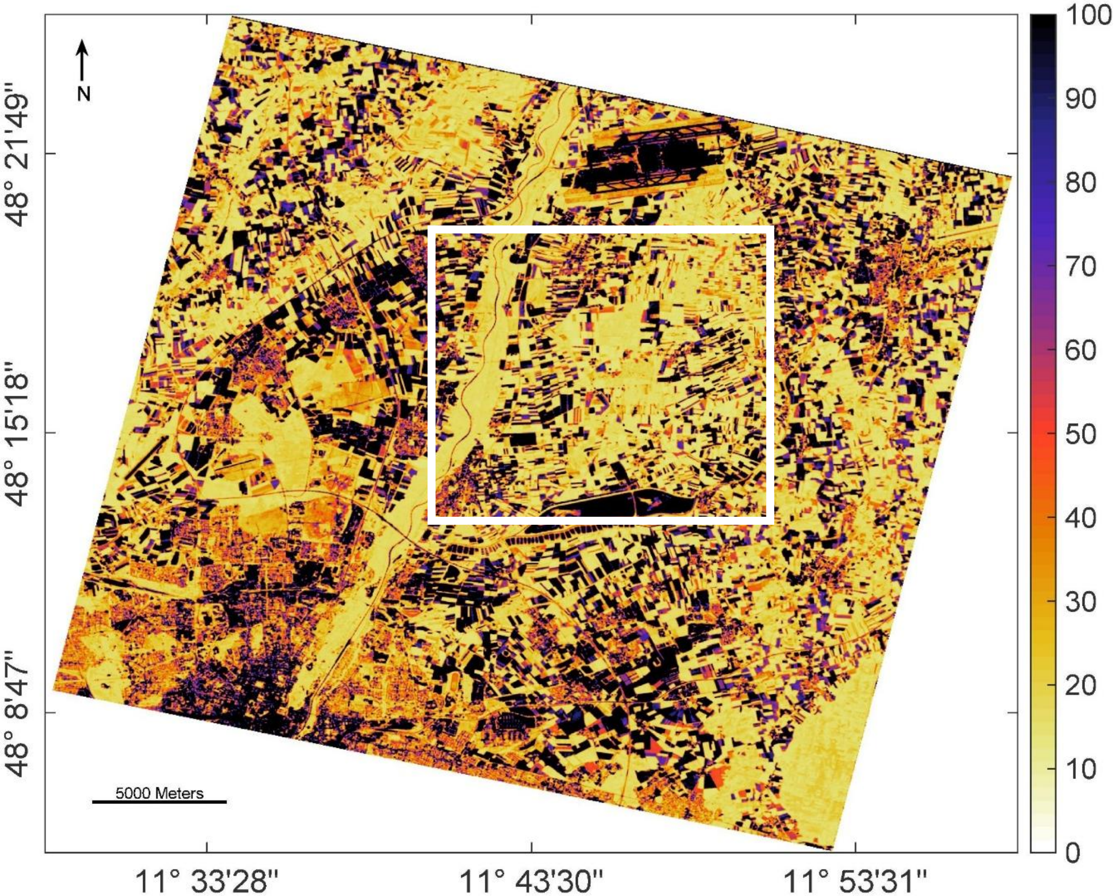
Page. 35

CNC estimates in [g/m²]

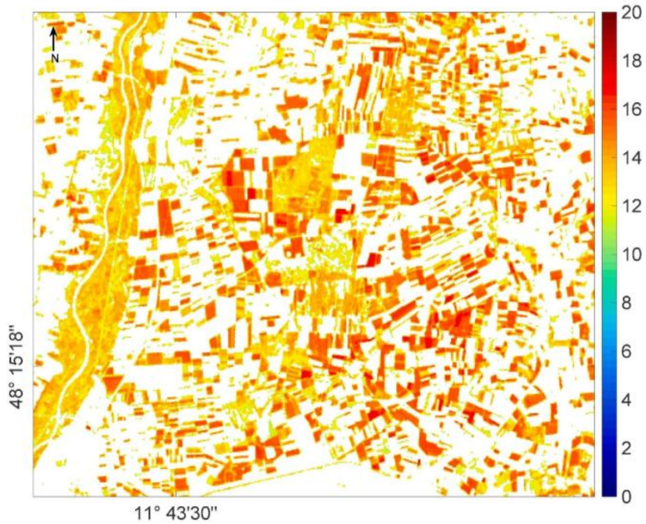


Verrelst, J., Rivera-Caicedo, J. P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., ... & Berger, K. (2021). Mapping landscape canopy nitrogen content from space using PRISMA data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *178*, 382-395.

Relative uncertainties in [%]: used as mask (e.g. only ≤20%)



Subset of CNC estimates in [g/m²]

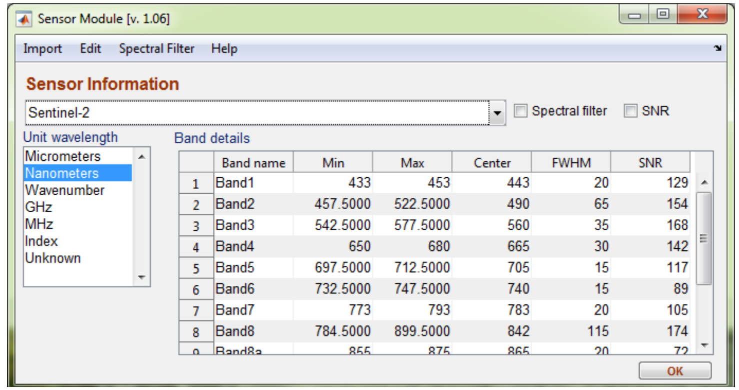


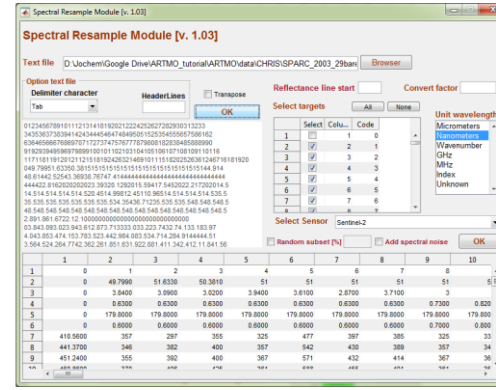
Conclusions CNC study:

- **A workflow for operational mapping of CNC designed for spaceborne imaging spectroscopy missions**
- **Hybrid method based on PROSAIL-PRO and GPR**
- **GPR provides associated uncertainty estimates**
- **Model applied to PRISMA and succesfully validated**

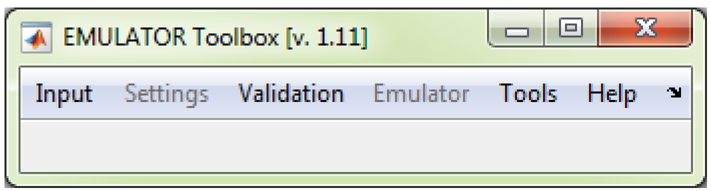
ARTMO's tools:

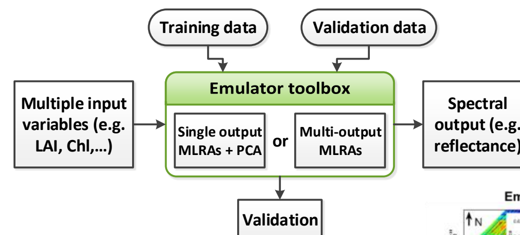
Sensor & spectral resample:

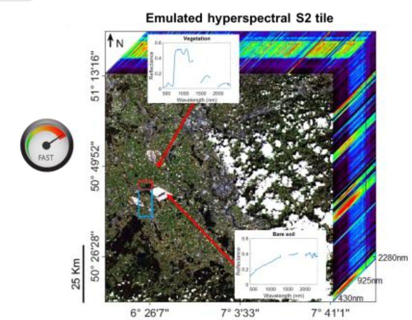




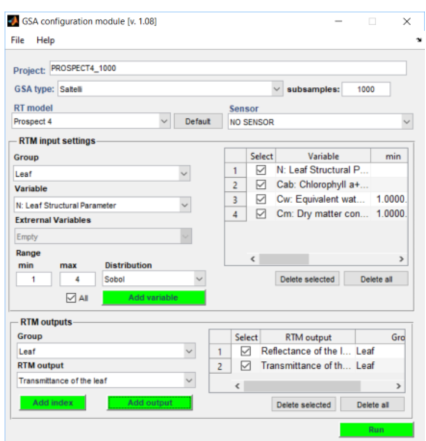
Emulation:

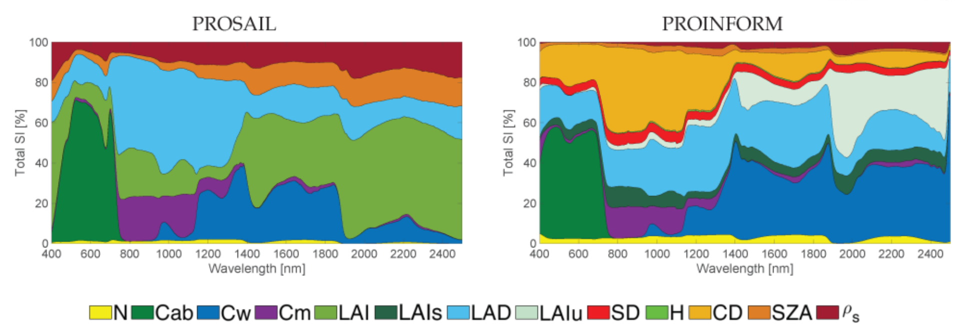




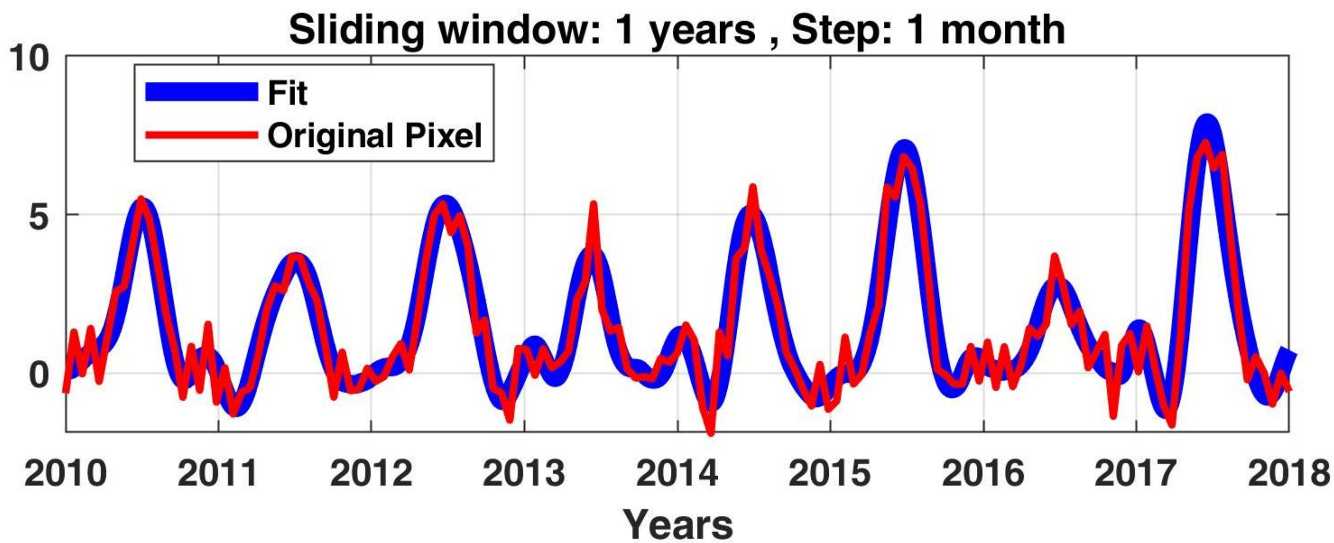


Global sensitivity analysis:

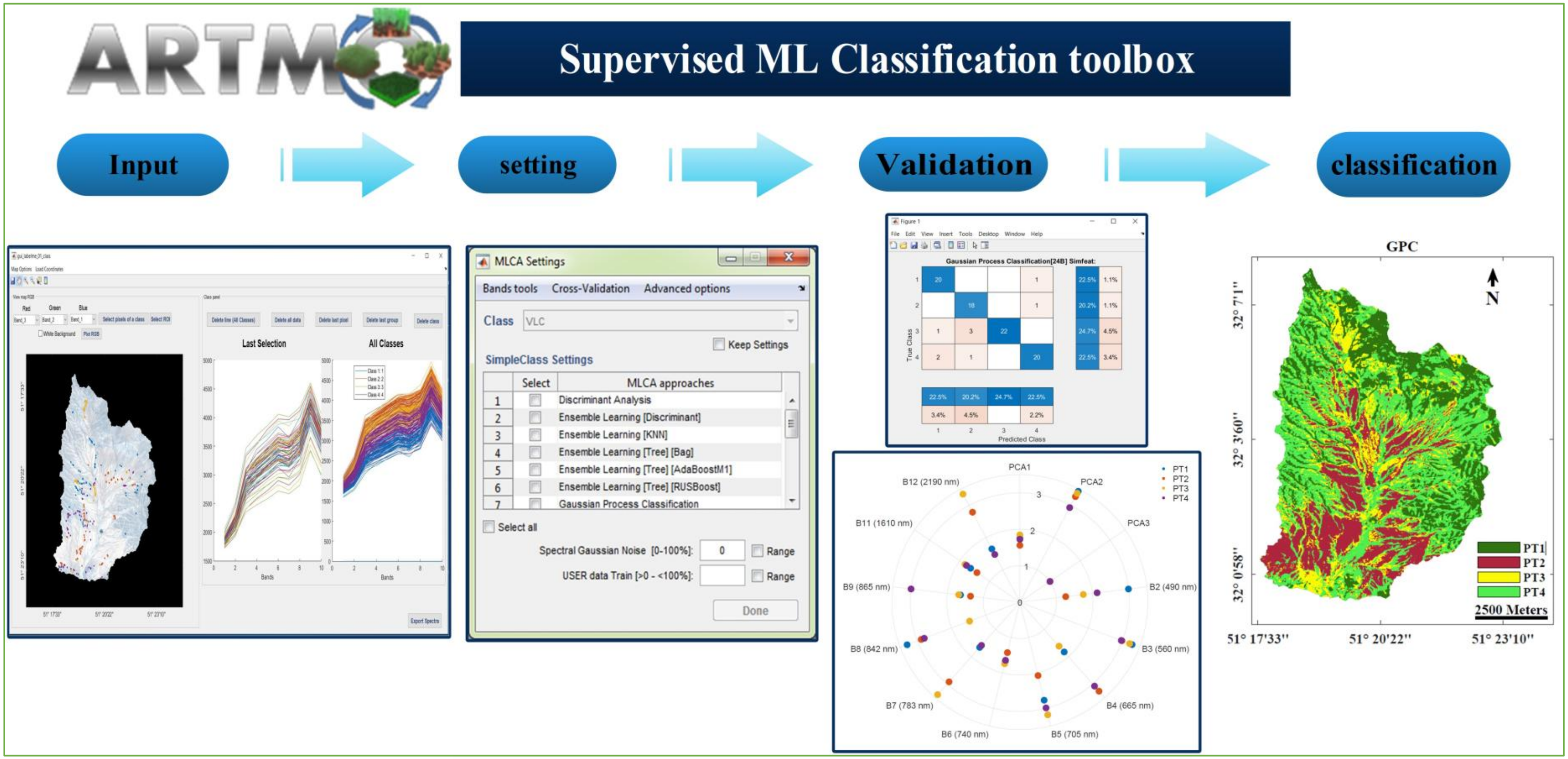


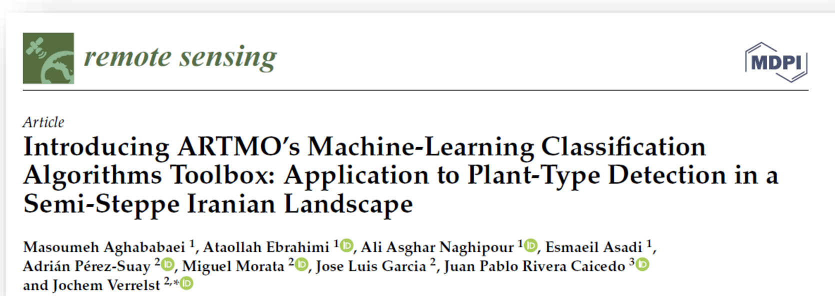


Dr. Santiago Belda Palazón
Dr. Jochem Verrelst
Dr. Juan Pablo Rivera
Dr. Luca Pipia
Pablo Morcillo



MLCA Classification toolbox - v.1.1



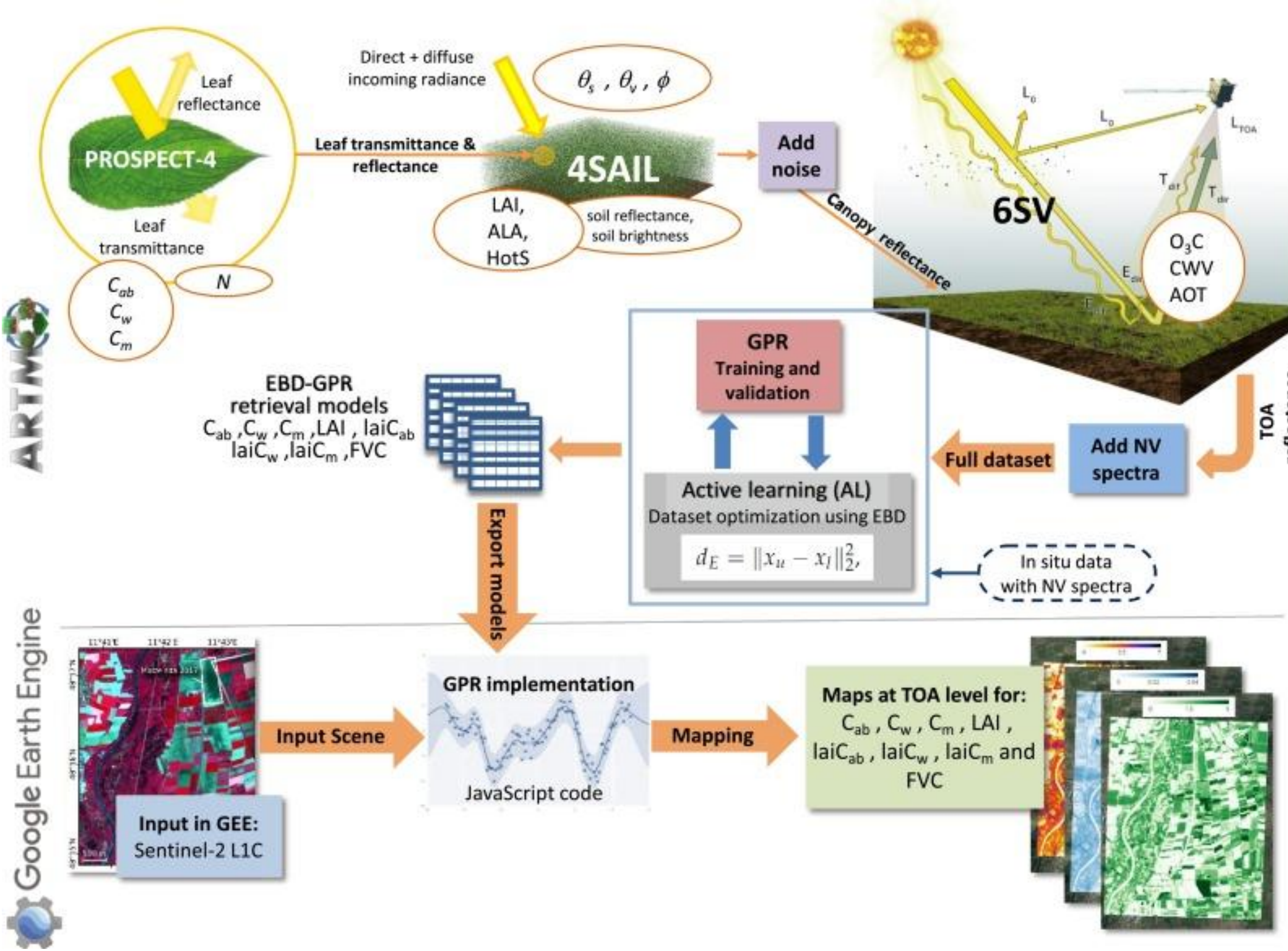

remote sensing

MDPI

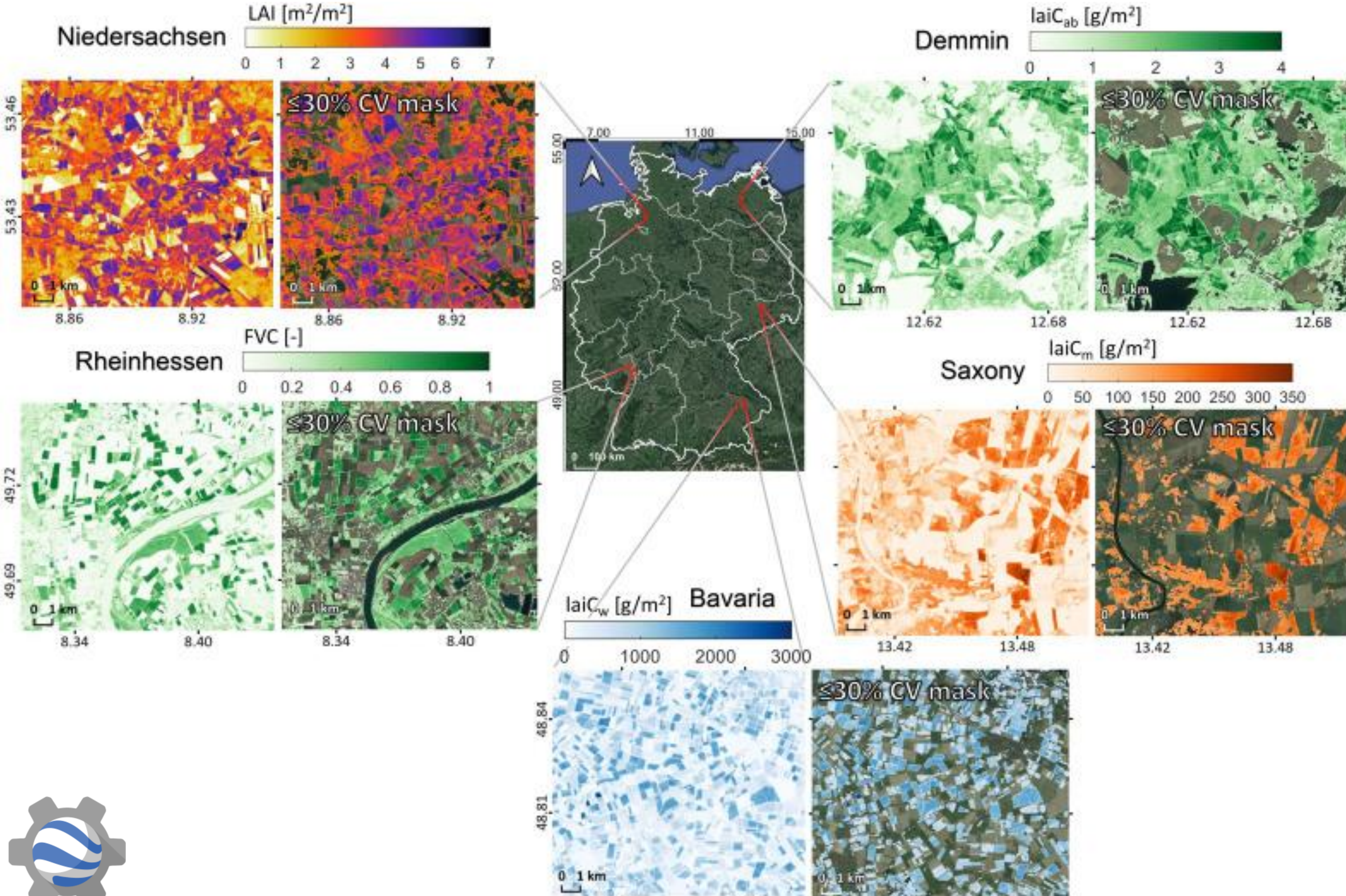
Article

Introducing ARTMO's Machine-Learning Classification Algorithms Toolbox: Application to Plant-Type Detection in a Semi-Steppe Iranian Landscape

Masoumeh Aghababaei [1], Ataollah Ebrahimi [1], Ali Asghar Naghipour [1], Esmaeil Asadi [1], Adrián Pérez-Suay [2], Miguel Morata [2], Jose Luis Garcia [2], Juan Pablo Rivera Caicedo [3] and Jochem Verrelst [2,*]

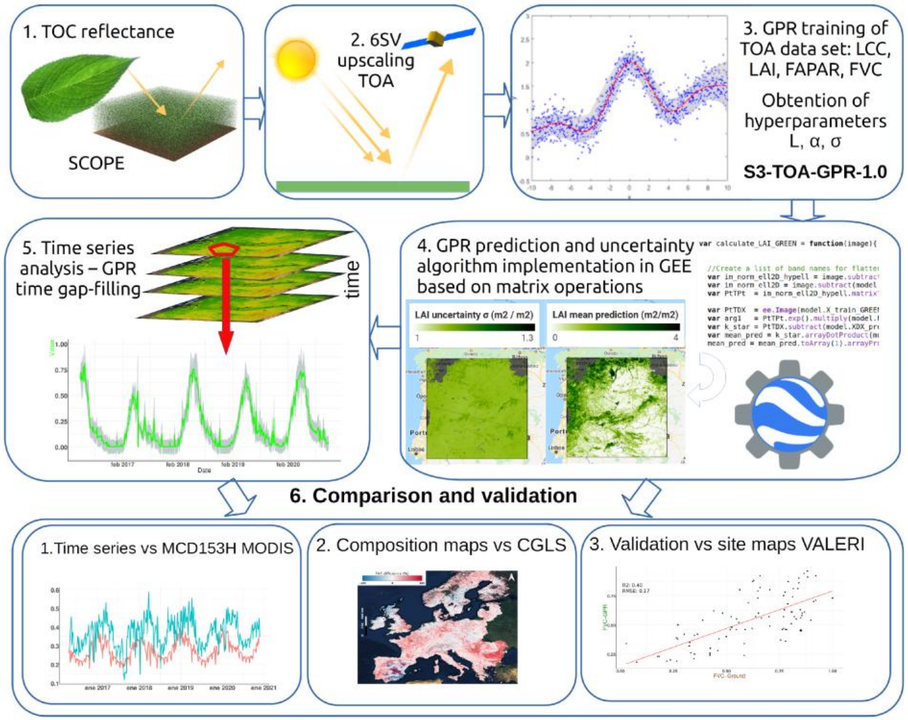


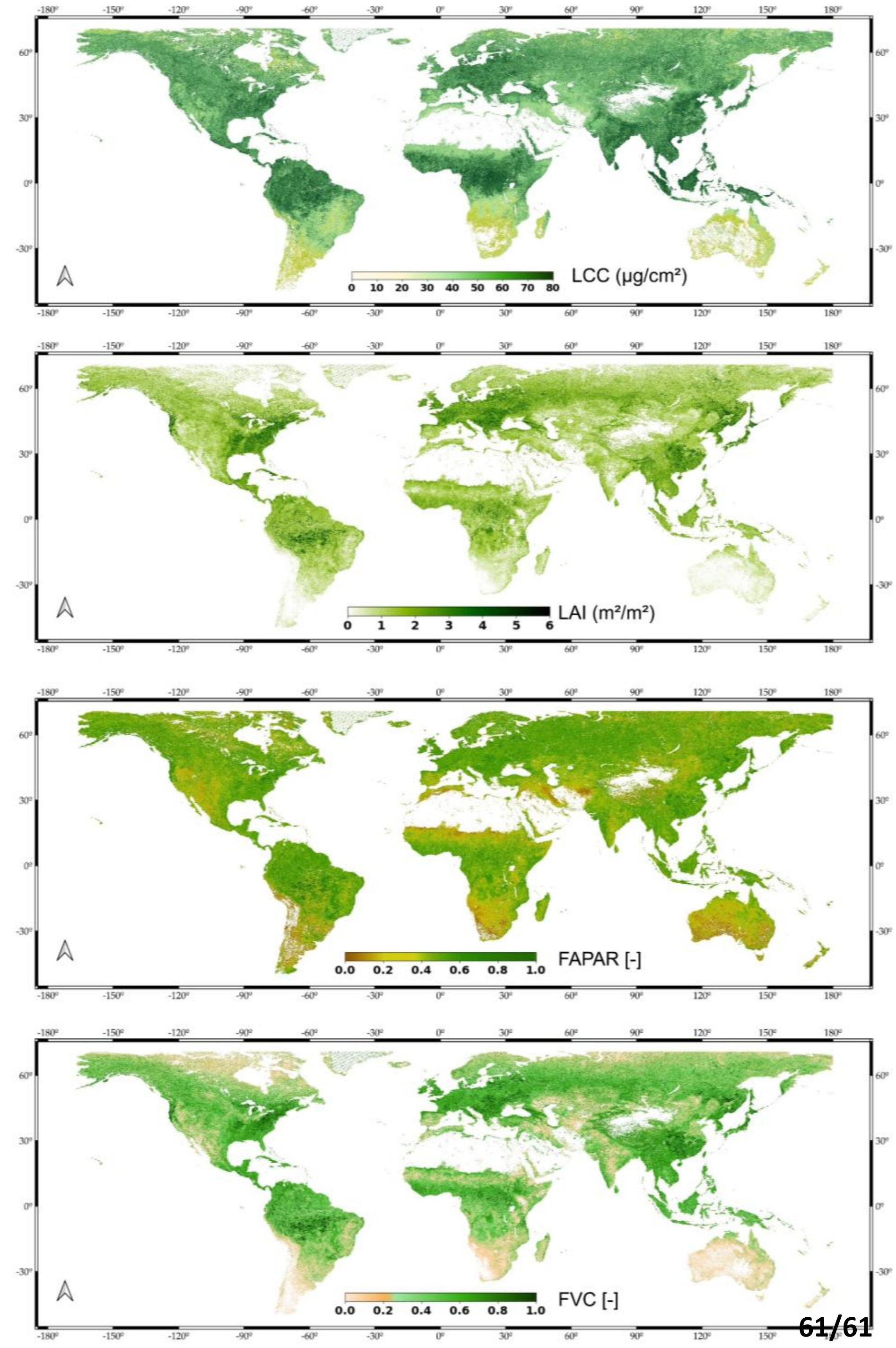
Estévez, J., Salinero-Delgado, M., Berger, K., Pipia, L., Rivera-Caicedo, J. P., Wocher, M., ... & Verrelst, J. (2022). Gaussian processes retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data. *Remote Sensing of Environment*, *273*, 112958.



Niedersachsen
LAI [m²/m²]
0 1 2 3 4 5 6 7
≤30% CV mask
53.46
53.43
8.86
8.92
Demmin
laiC_ab [g/m²]
0 1 2 3 4
53.84
53.81
12.62
12.68
Rheinhessen
FVC [-]
0 0.2 0.4 0.6 0.8 1
49.72
49.69
8.34
8.40
Saxony
laiC_m [g/m²]
0 50 100 150 200 250 300 350
51.27
51.24
13.42
13.48
Bavaria
laiC_w [g/m²]
0 1000 2000 3000
48.84
48.81
12.66
12.72
7.00
11.00
15.00
55.00
52.00
49.00
0 100 km
0 1 km

GEE

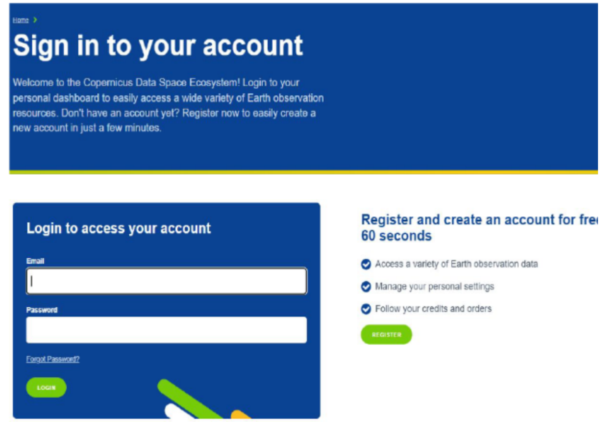




https://github.com/msalinero/ARTMOtoGEE

PyEOGPR

1. Free to use, registration to Copernicus Data Space Ecosystem required
2. Open **any** Python IDE!
3. *pip install pyeogpr*
4. And you are ready to go!



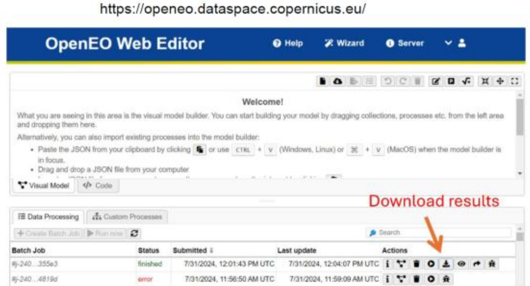
www.github.com/daviddkovacs/pyeogpr

https://pypi.org/project/pyeogpr/

Code snippet

```python
import pyeogpr

# Your region of interest
bounding_box = [
        -0.305543150556133,
        39.29253033906926,
        -0.28169853763617425,
        39.303338211248104
    ]

# Time window for processing Satellite observations
time_window = ["2022-05-01", "2022-06-01"]

dc = pyeogpr.Datacube(
    "SENTINEL2_L2A",  # Satellite sensor
    "FVC",            # Fractional Vegetation Cover
    bounding_box,
    time_window,
    cloudmask=True
)

dc.construct_datacube("dekad")  # Initiates openEO datacube

dc.process_map()  # Starts GPR processing
```

https://openeo.dataspace.copernicus.eu/

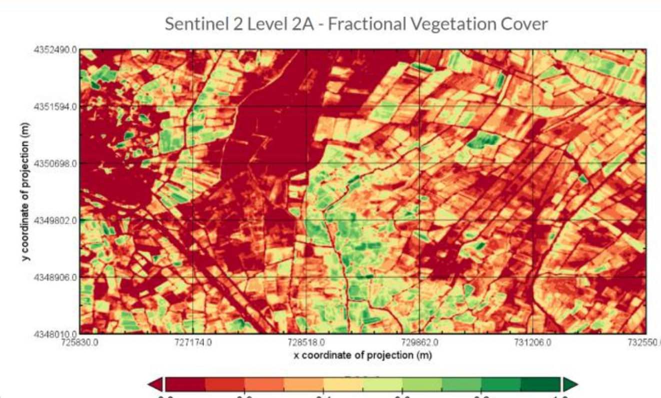


Download results

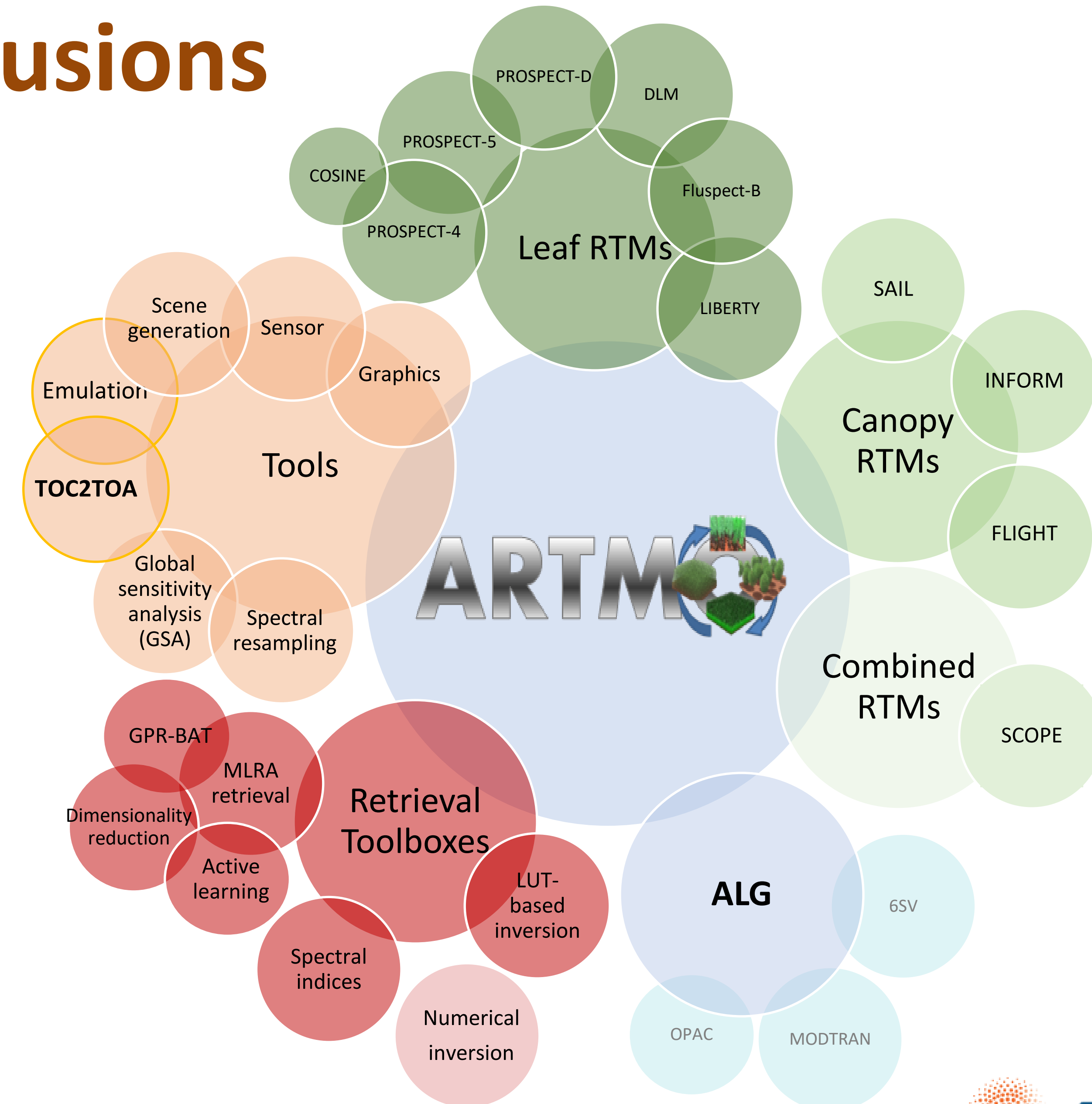
Sends process to back-end

Visualize results in e.g. QGIS or Panoply

Resulting map: Multi-temporal *netcdf or tiff*



Conclusions



http://artmotoolbox.com/

https://leoipl.uv.es/sentiflex