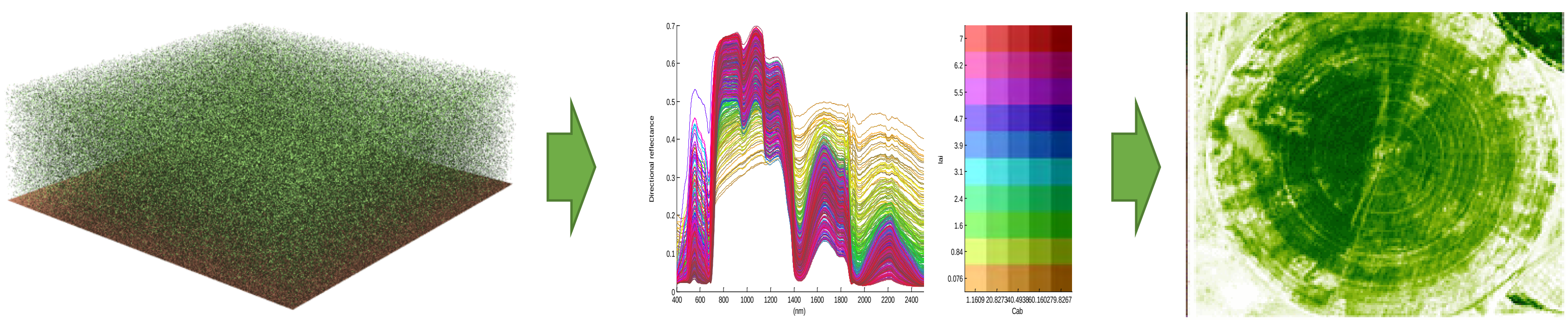
From model simulations towards vegetation properties mapping:

automating, optimizing & simplifying



J. Verrelst

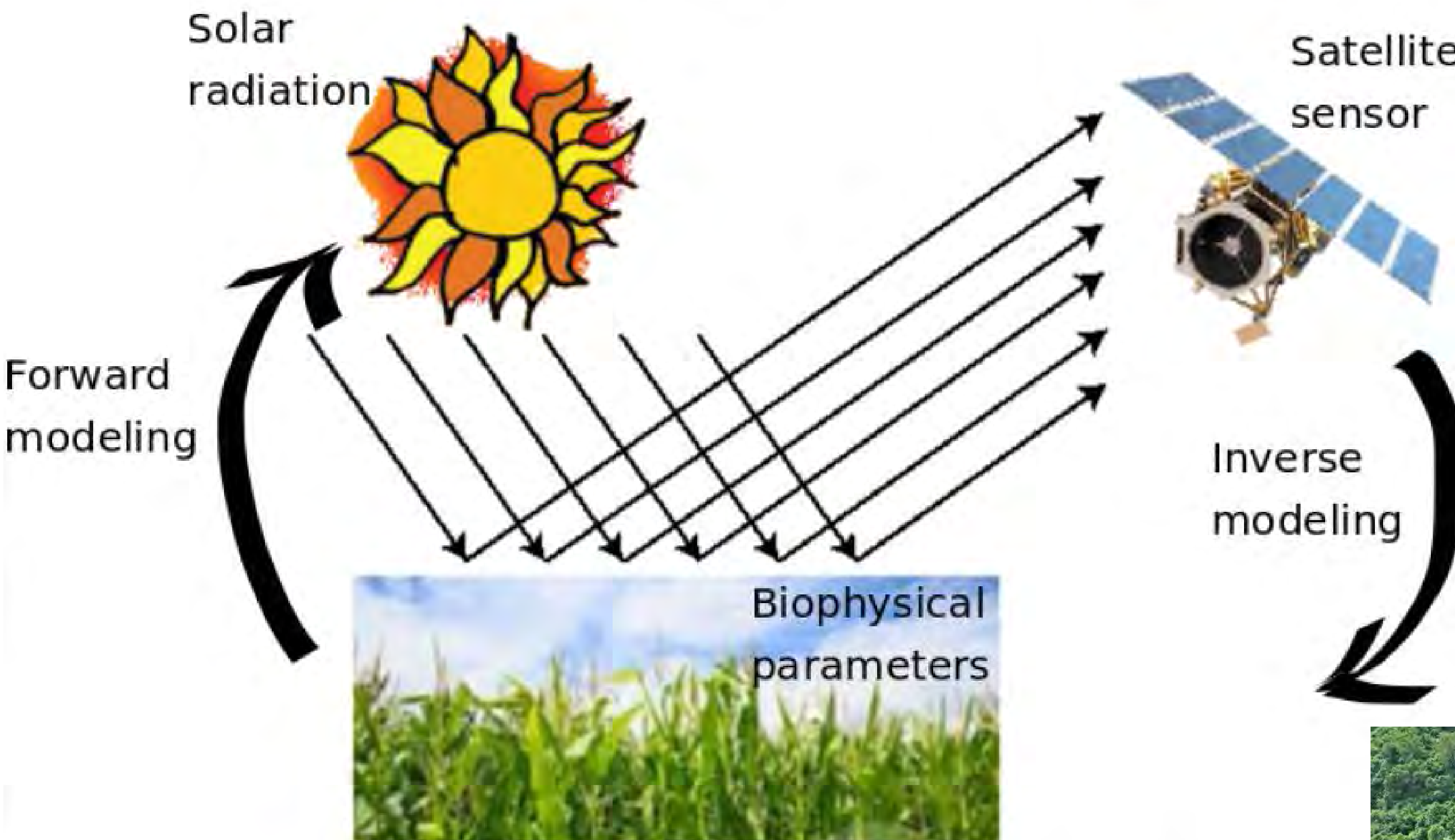
EOSS2018 – ESA-ESRIN

06-Aug

How to quantify vegetation properties?

Today we will learn:

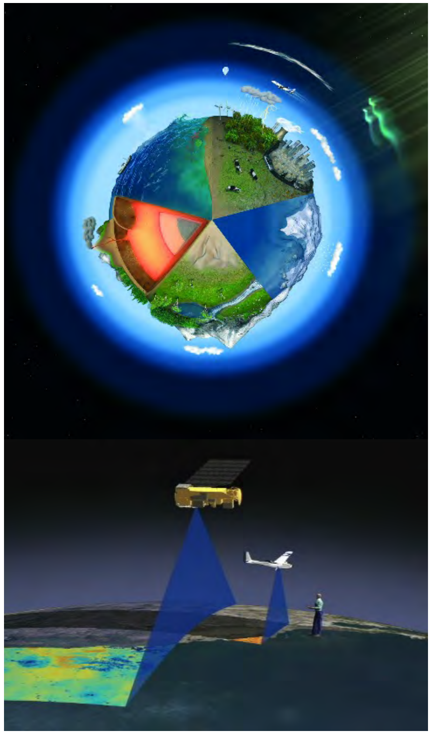
Semi-automated mapping of vegetation properties from optical RS data



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

The problem:

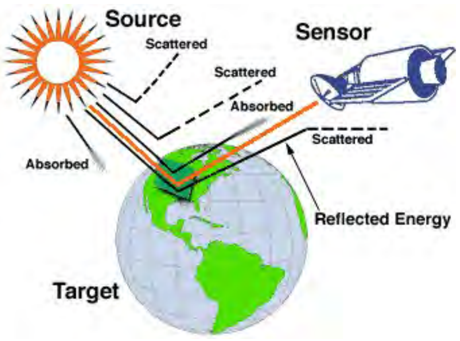
Biophysical parameter retrieval is an essential step in modeling the processes occurring on Earth and the interactions with the atmosphere.



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,….

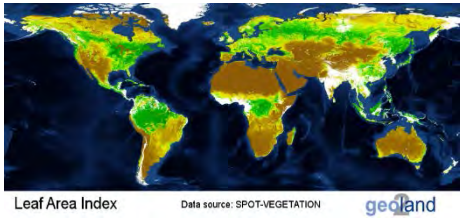
Land/vegetation parameters cannot be estimated directly from optical RS data. A model is required!



The objective: Transform measurements into biophysical parameter estimates.

The data:

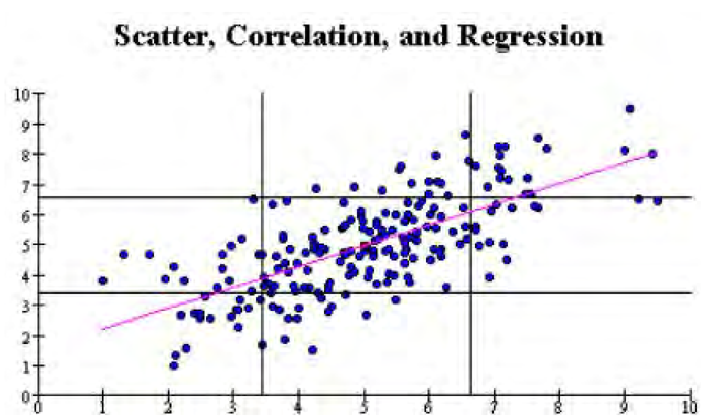
- **Input data:** satellite/airborne spectra, in situ (field) radiometers, or simulated spectra by RTMs
- **Output results:** estimation of a biophysical parameter



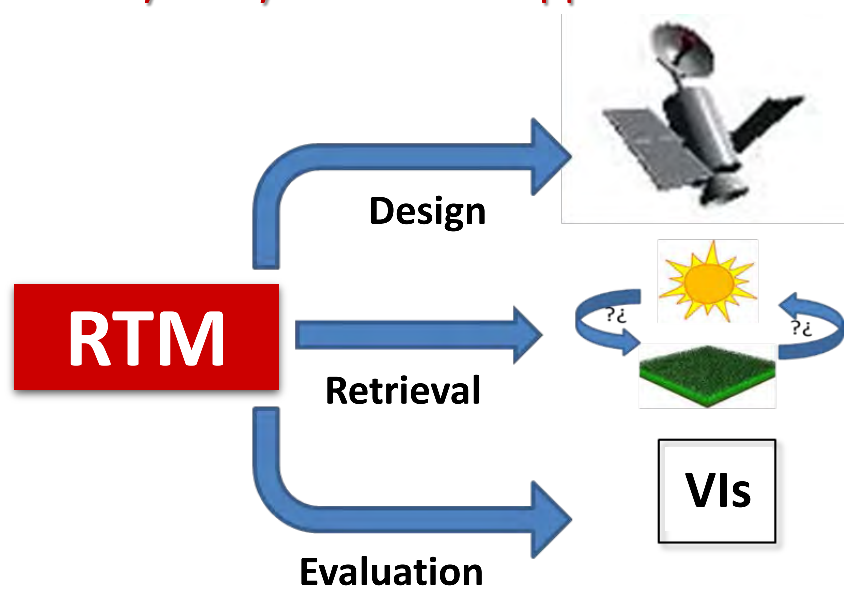
Introduction retrieval biophysical parameters

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

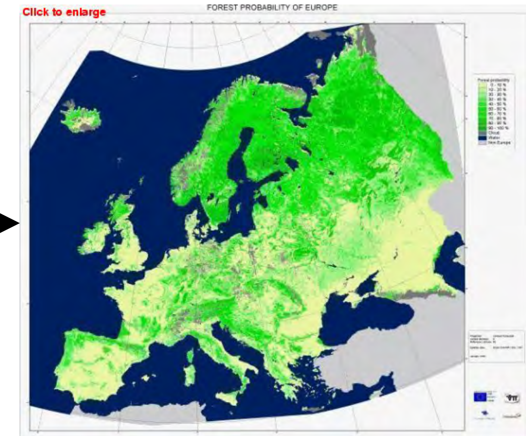


Retrieval of (continuous) vegetation properties

Remote sensing image

Model

Map of a vegetation property

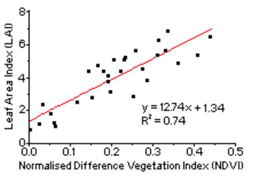


1. **Statistical models**
    1. Parametric regression models
    2. Nonparametric regression models
        1. Linear
        2. Nonlinear
2. **Inversion of physically based radiative transfer models**
    1. Numerical optimization
    2. Lookup-table (LUT)-based inversion

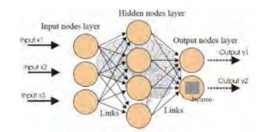
Taxonomy of retrieval methods, three main families:

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.

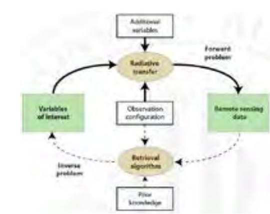


- **Non-parametric** models are ***data-driven* models**. They are adjusted to predict a variable of interest using a training dataset of input-output data pairs.



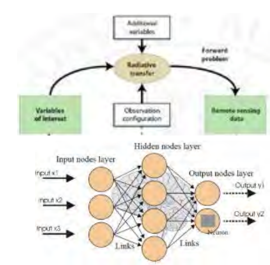
2. *Physical*: try to reverse RTMs.

- 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.



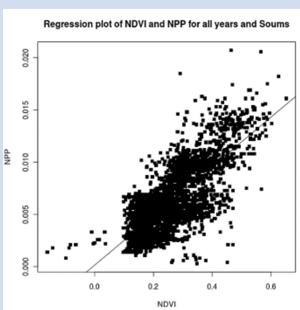
Retrieval families

Parametric regression

Spectral relationships that are sensitive to specific vegetation properties

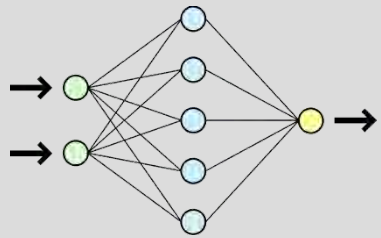
$$NDVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED})}$$

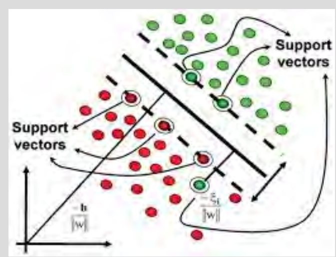
Normalized Difference Vegetation Index



Non-parametric regression

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

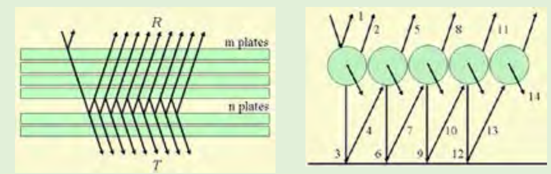




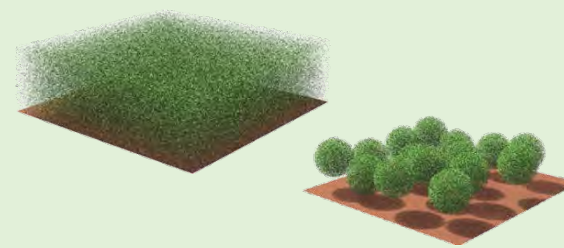
RTM inversion

Models that simulate interactions between vegetation and radiation

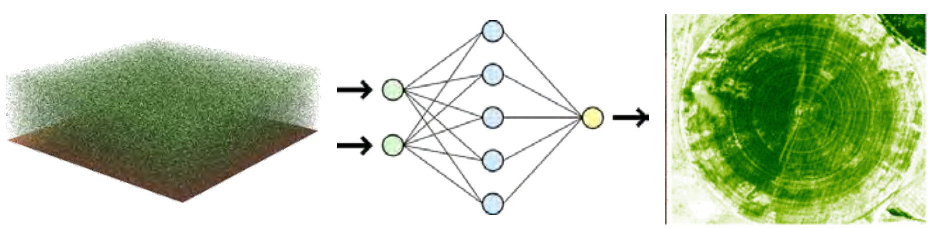
leaf



canopy



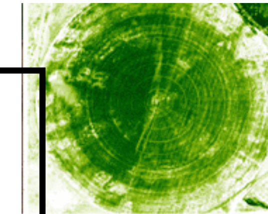
Methods of these different families can be combined: *hybrid methods*



Statistical interpretation of RS

Remote Sensing Data ⟵ **Statistical relationship** ⟶ **Variable of Interest**

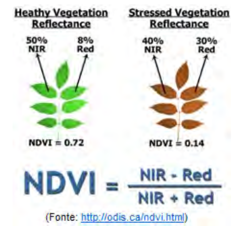
- Parametric regression
- Non-parametric regression



- 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

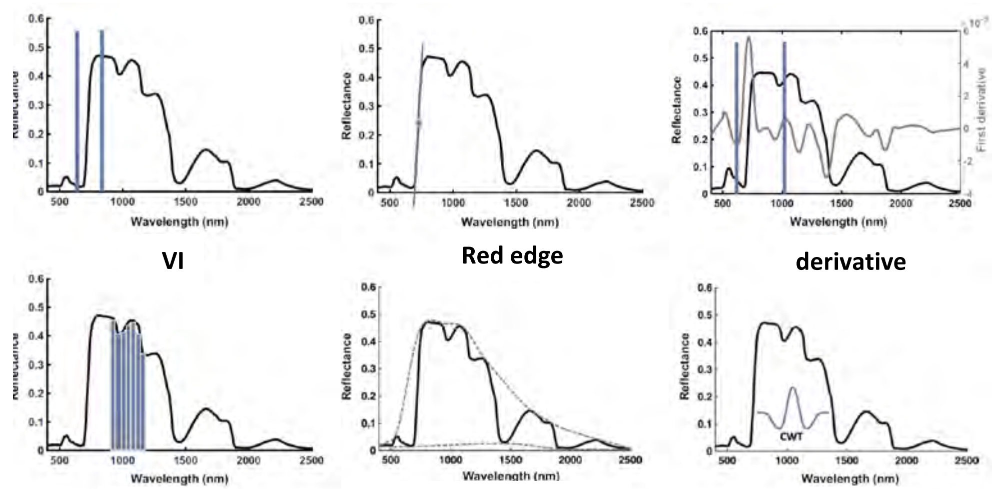


$$PRI = \frac{(\rho 570 - \rho 531)}{(\rho 570 + \rho 531)}$$

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

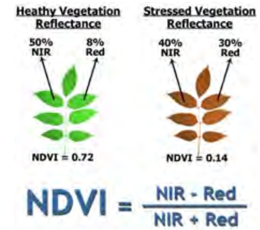
$$TCARI / OSAVI = 3 \cdot \left[\left(\rho_{\mu 700} - \rho_{\mu 670} \right) - 0.2 \cdot \left(\rho_{\mu 700} - \rho_{\mu 550} \right) \cdot \right] \left(\rho_{\mu 700} - \rho_{\mu 670} \right) \left(1 + 0.16 \right) \cdot \left(\rho_{\mu 800} - \rho_{\mu 670} \right) / \left(\rho_{\mu 800} - \rho_{\mu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

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



VI | Red edge | derivative | integral | Continuum removal | wavelet

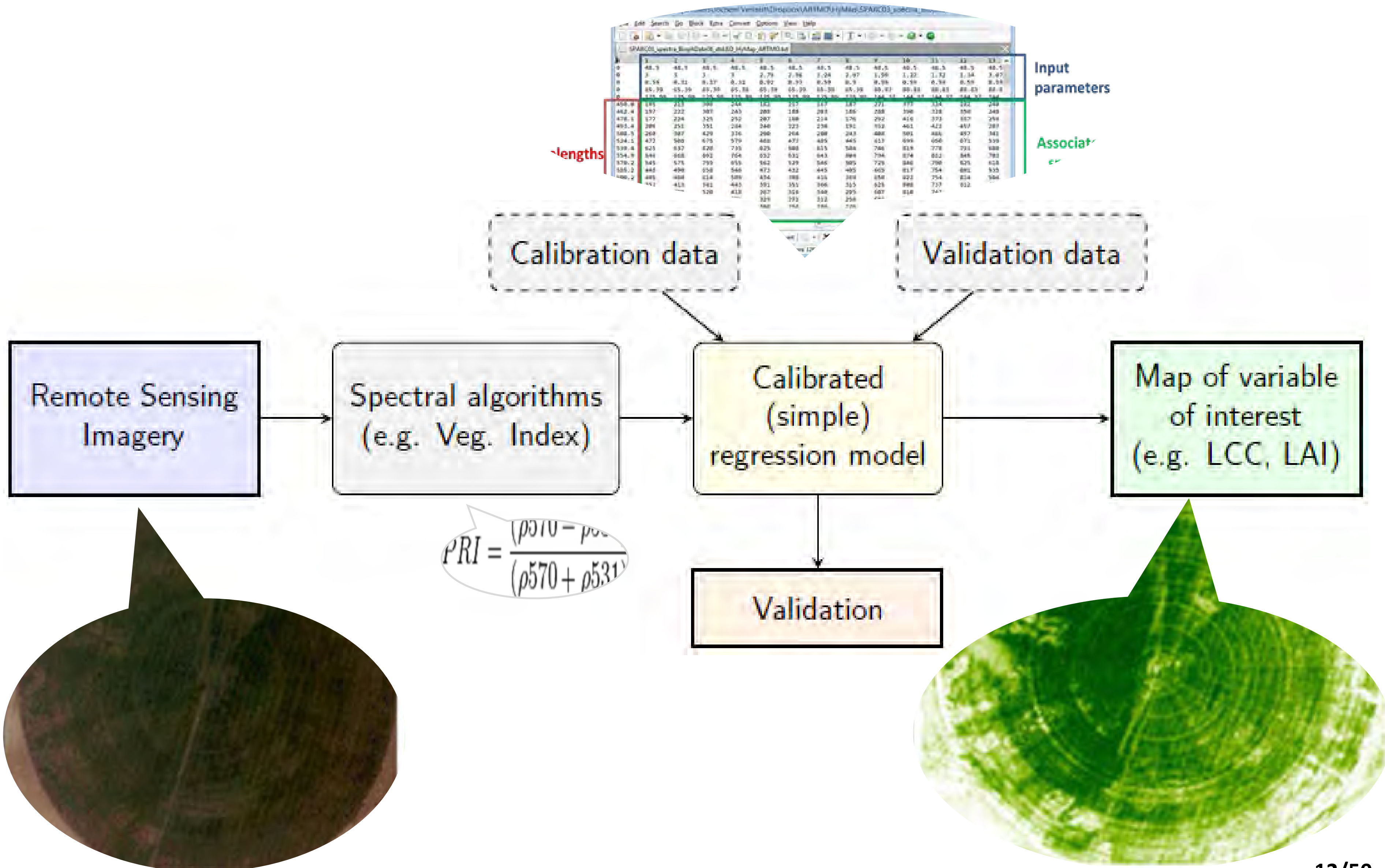
Parametric regression:



(Fonte: http://odis.ca/ndvi.html)

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



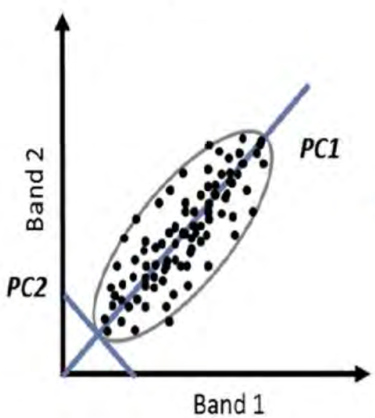
Non-parametric models (1/2):

Data-driven methods: Do not assume explicit feature relations

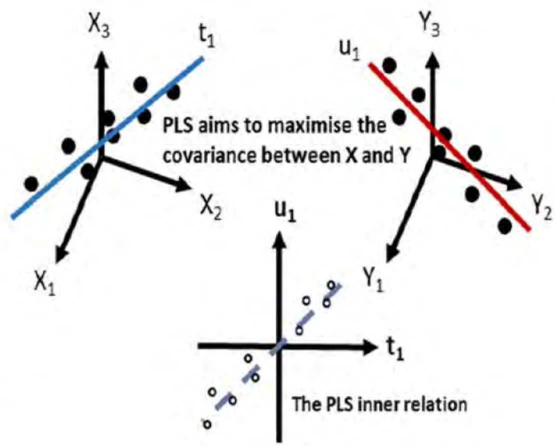
Linear nonparametric models:

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

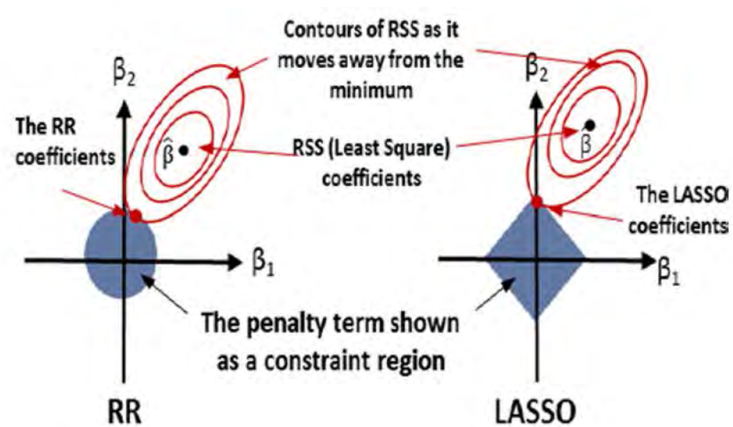
PCR



PLSR



RR & LASSO

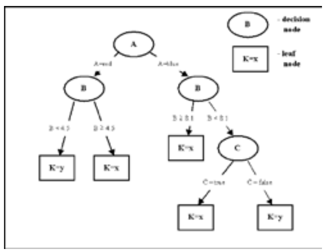


Non-parametric models (2/2):

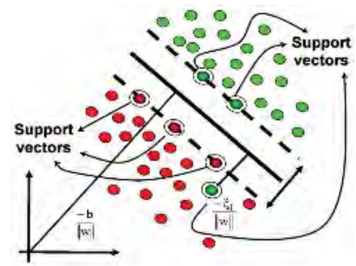
Data-driven methods: Do not assume explicit feature relations

Non-linear nonparametric models:

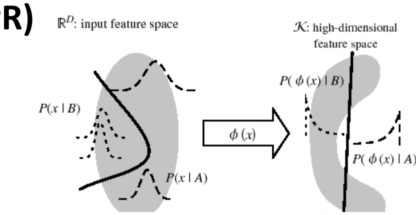
Decision Trees (DT)



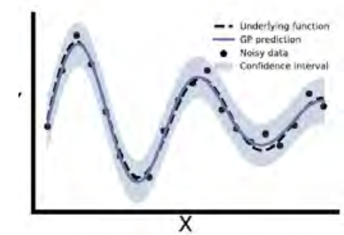
Support vector regression (SVR)



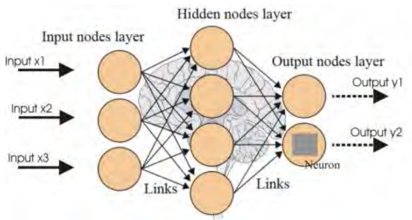
Gaussian processes regression (GPR)



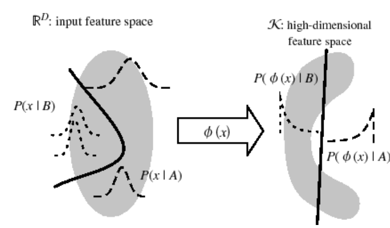
$$K(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(- \sum_{b=1}^{B} \frac{\left(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)}\right)^2}{2\sigma_b^2} \right)$$



Neural networks (NN)



Kernel ridge regression (KRR)



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

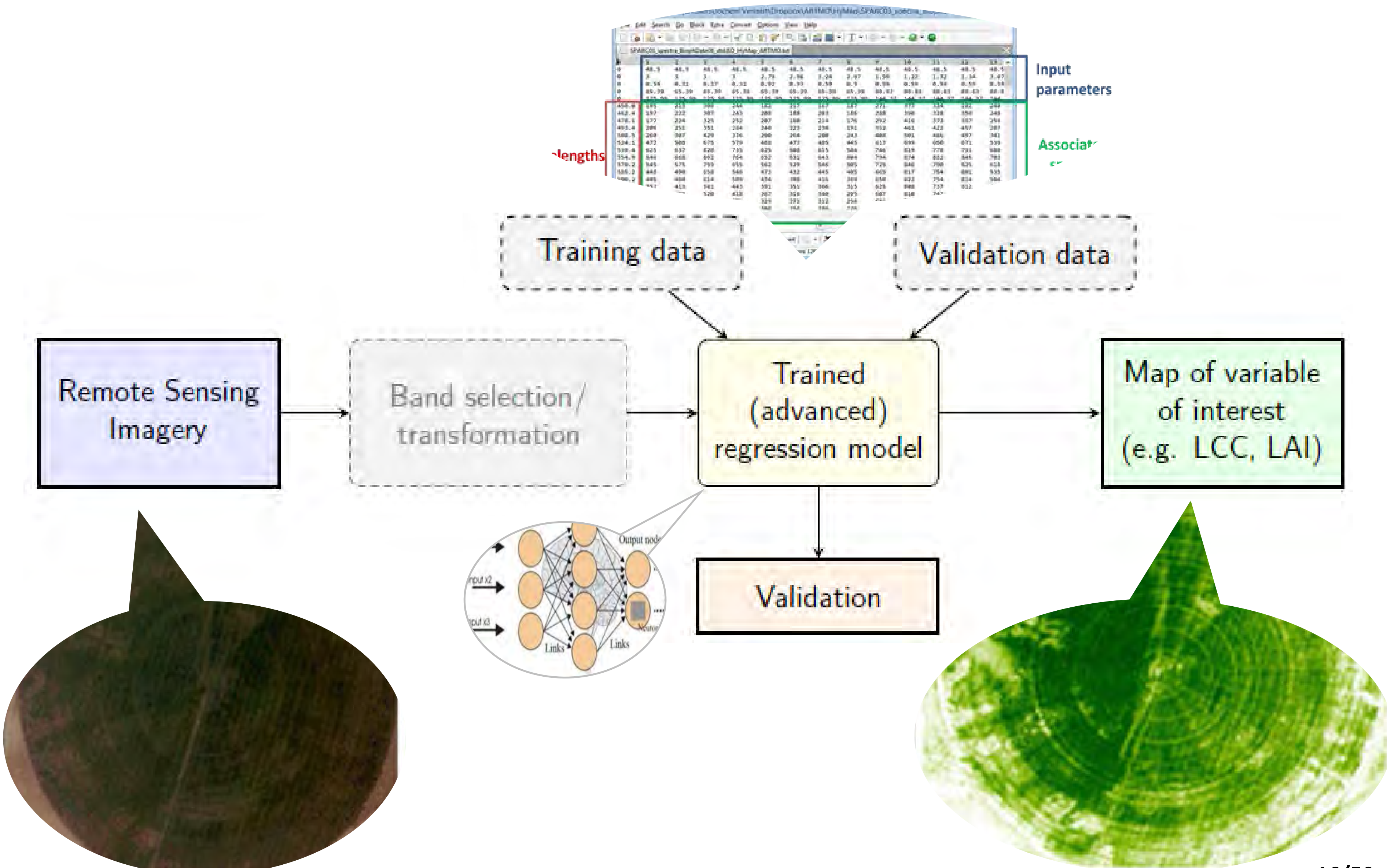
Also:

- 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)

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

↔

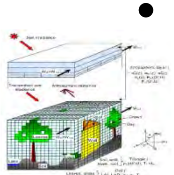
Invertible models

RT State Variables

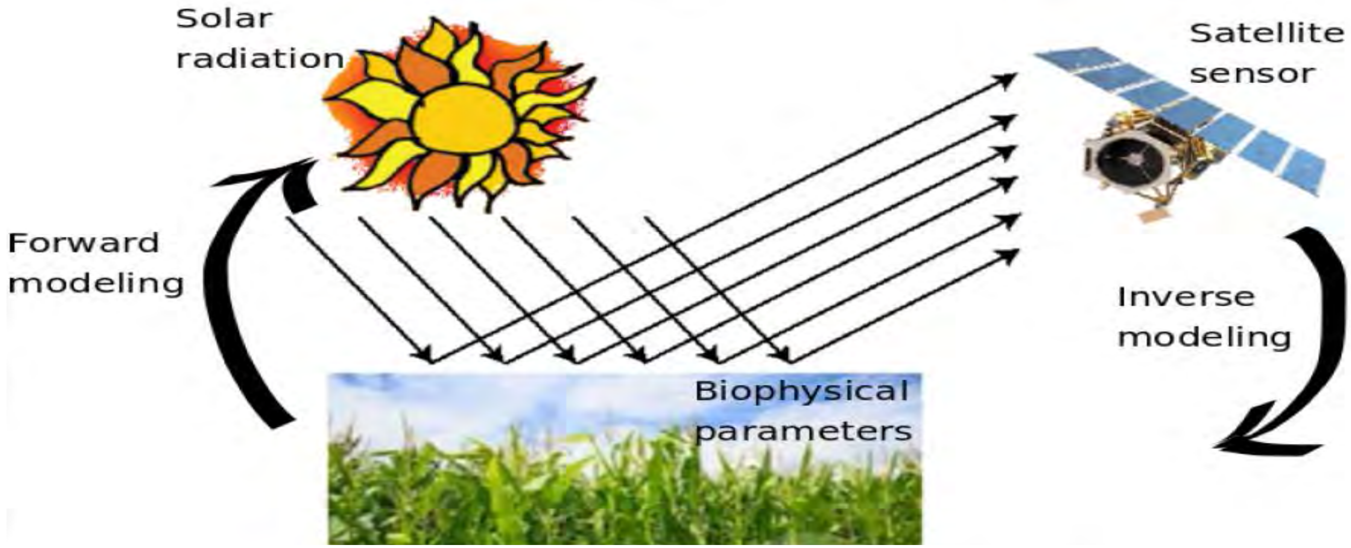
↕ **Some relationship**

Variable of Interest

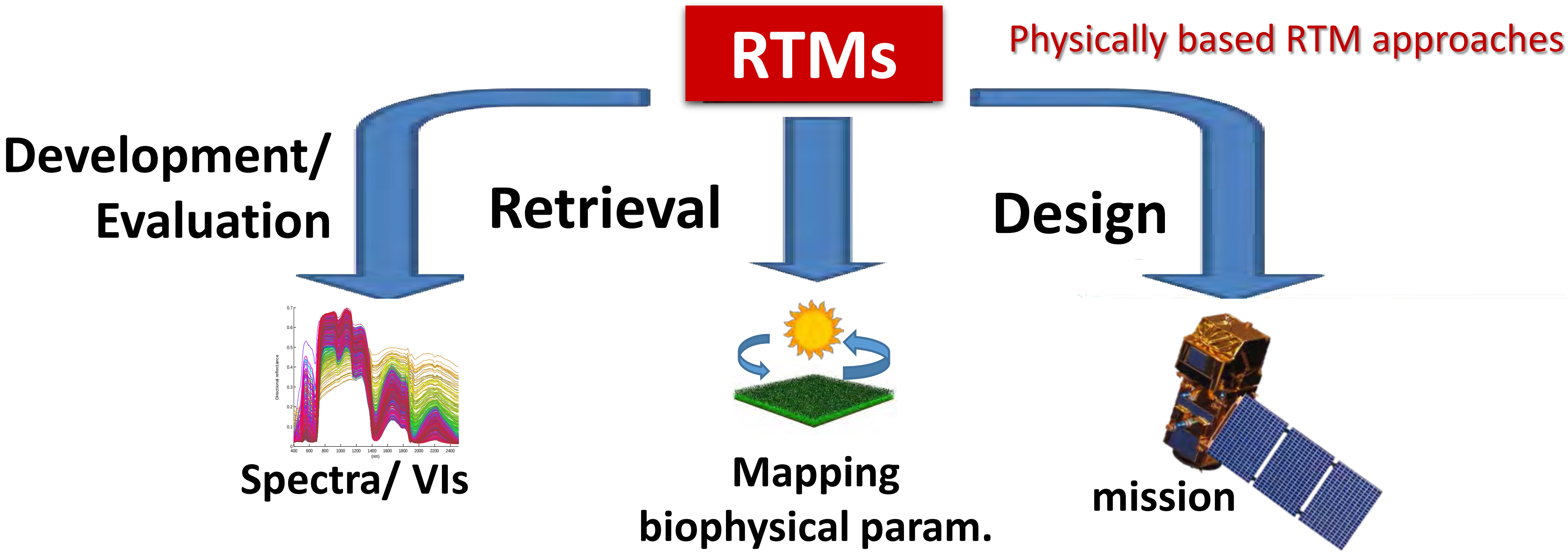
Radiative transfer models:

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

Background



Solar radiation
Satellite sensor
Forward modeling
Inverse modeling
Biophysical parameters

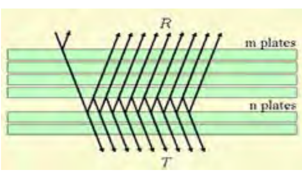


RTMs
Physically based RTM approaches
Development/ Evaluation
Retrieval
Design
Spectra/ VIs
Mapping biophysical param.
mission

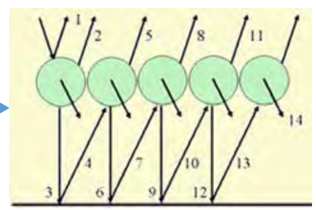
Radiative transfer models

Leaf RT models

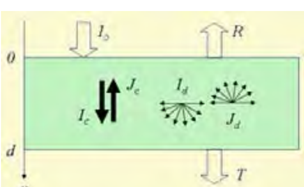
layers



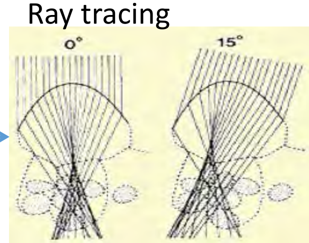
Compact spheres



N-fluxes



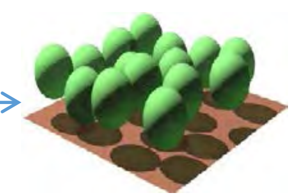
Ray tracing



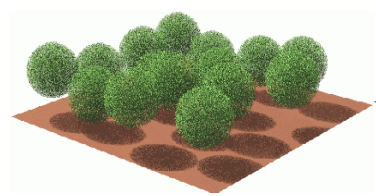
Canopy RT models

Turbid medium

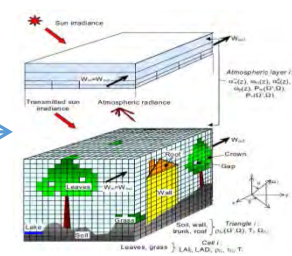
Geometric



Hybrid



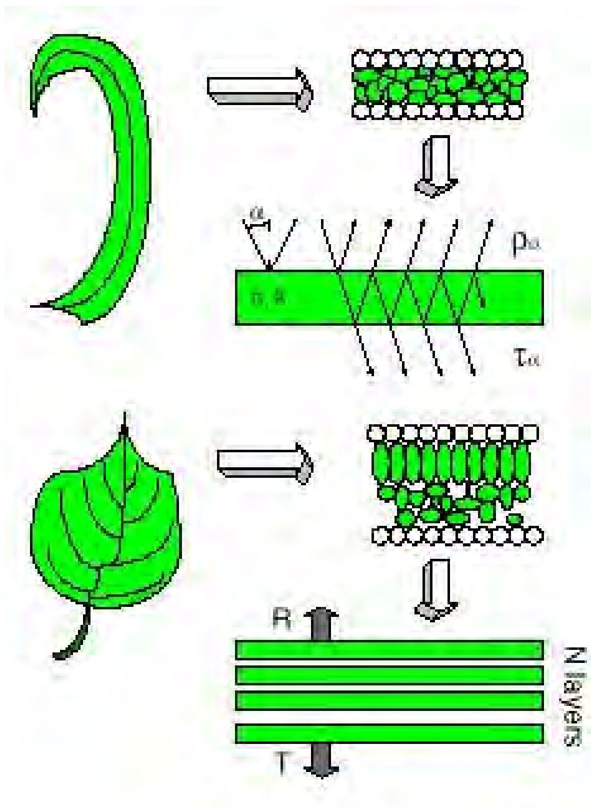
Volumetric

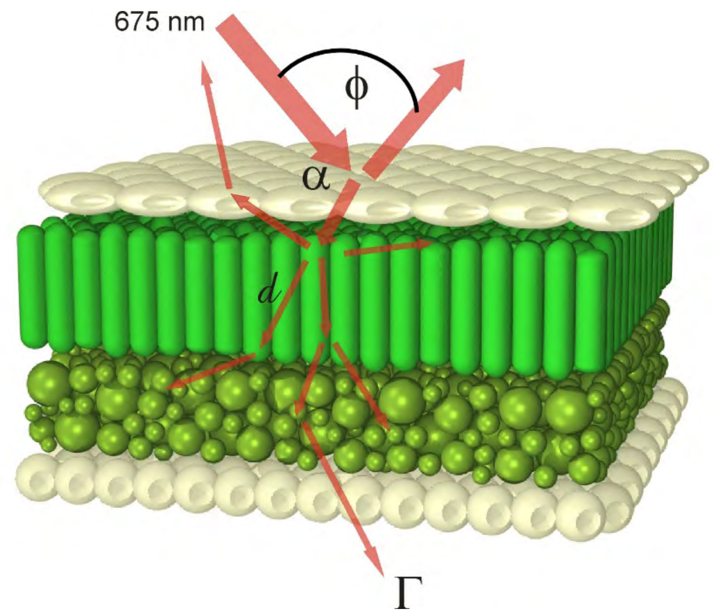


Multiple models exist with diverse complexity.

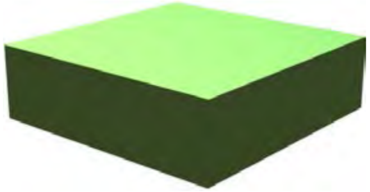
Leaf optical models

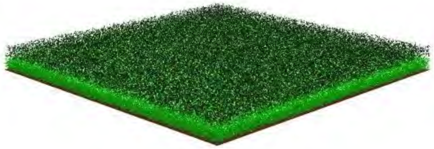
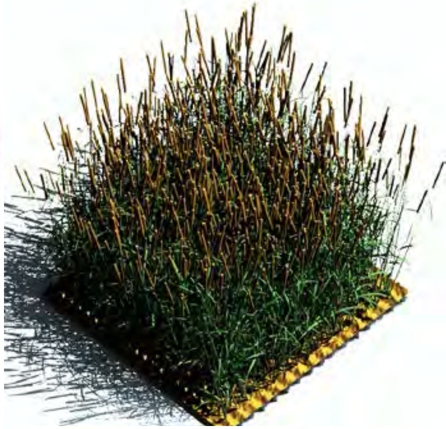
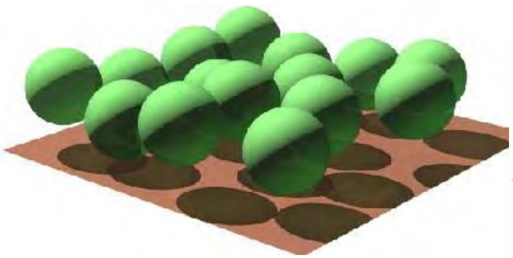
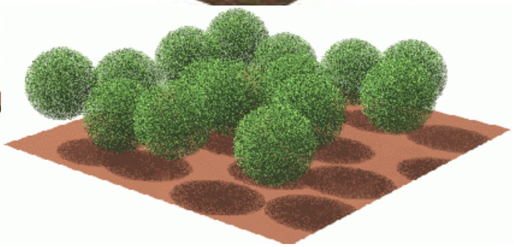
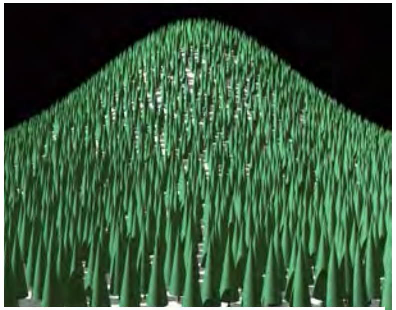
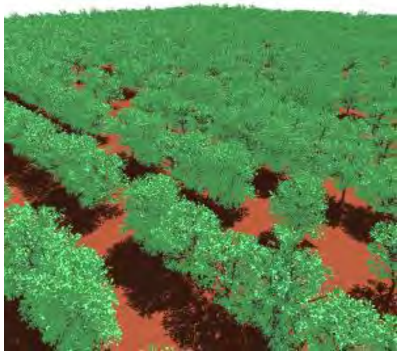
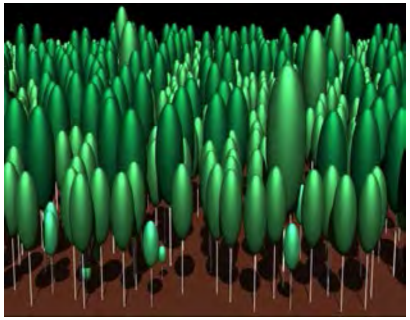
- A leaf is not opaque but transparent.
- Leaf as composed out of layers and empty spaces

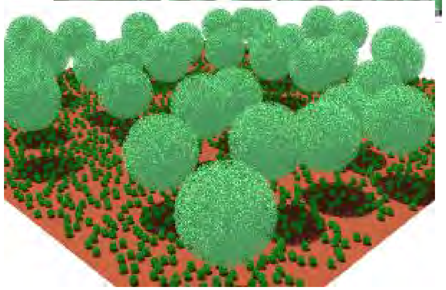


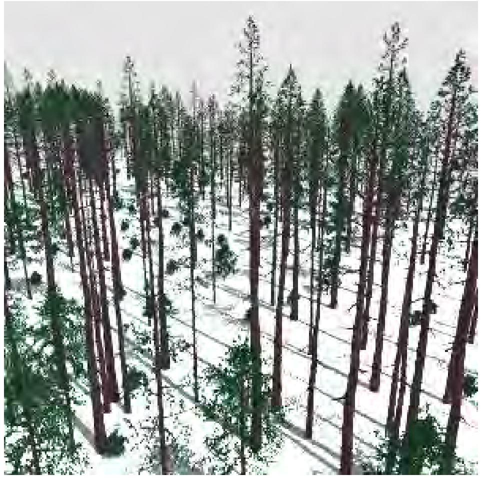
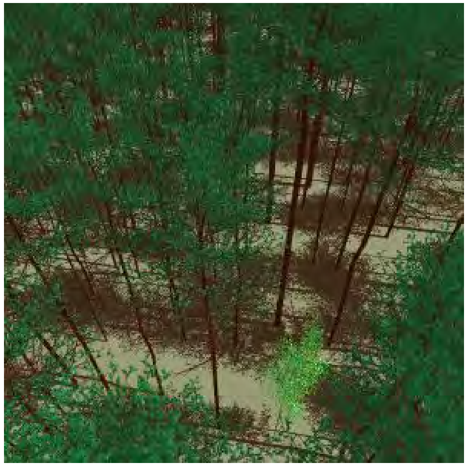


Canopy RTMs



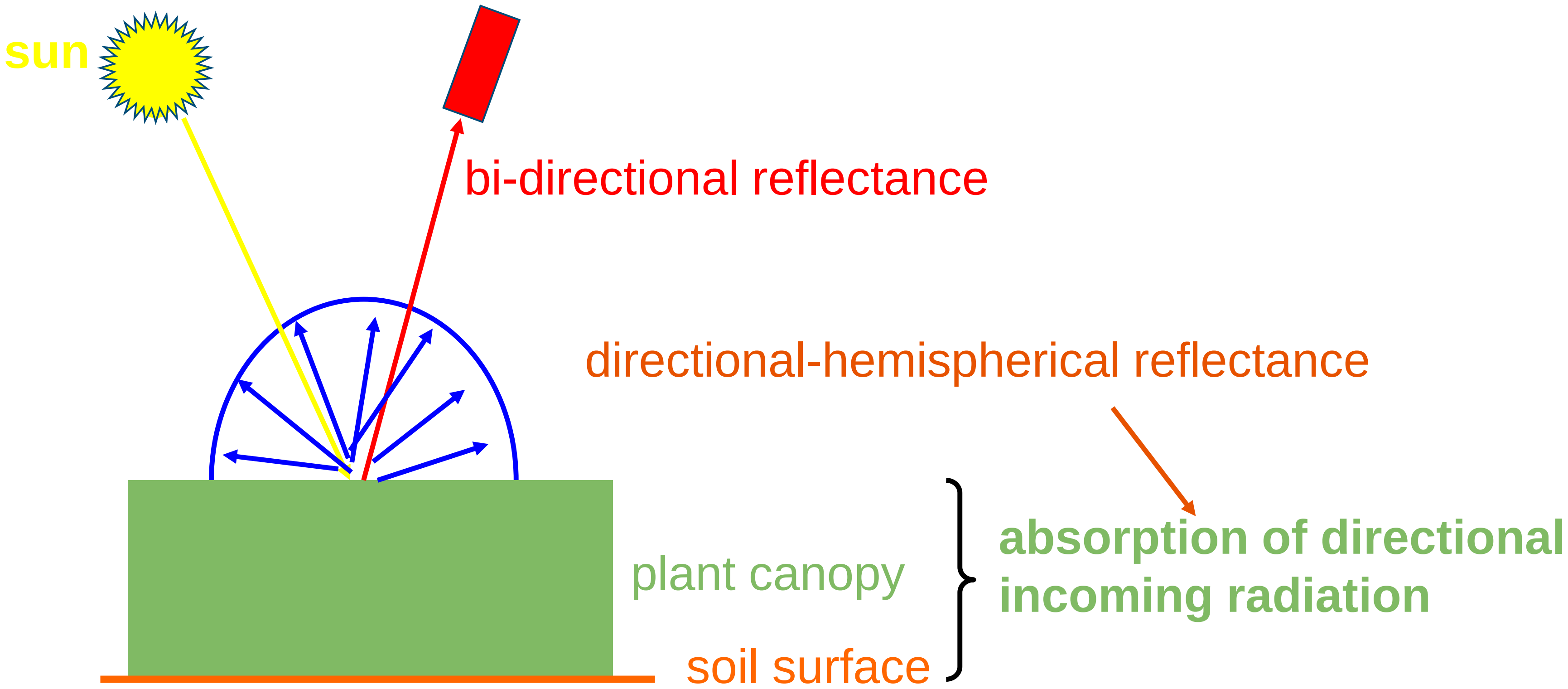



http://rami-benchmark.jrc.ec.europa.eu

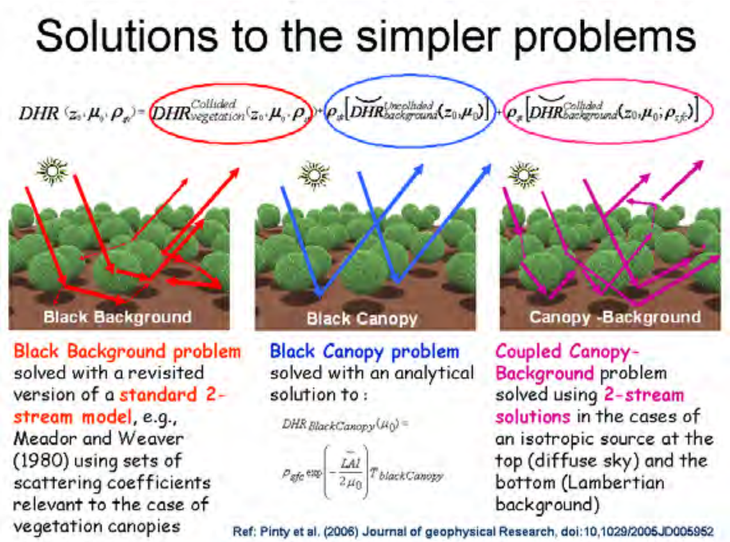
Examples of canopy RTMs(1/4)

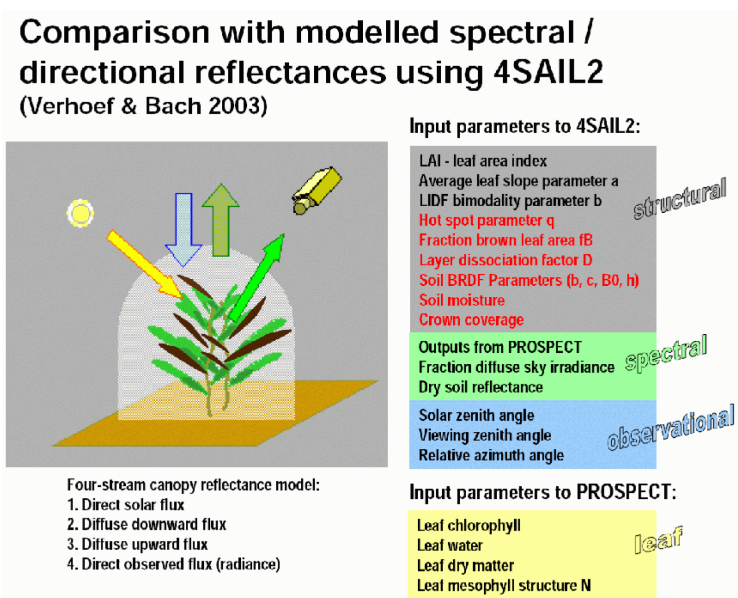
SAIL model (Verhoef 1984): a 1-D model

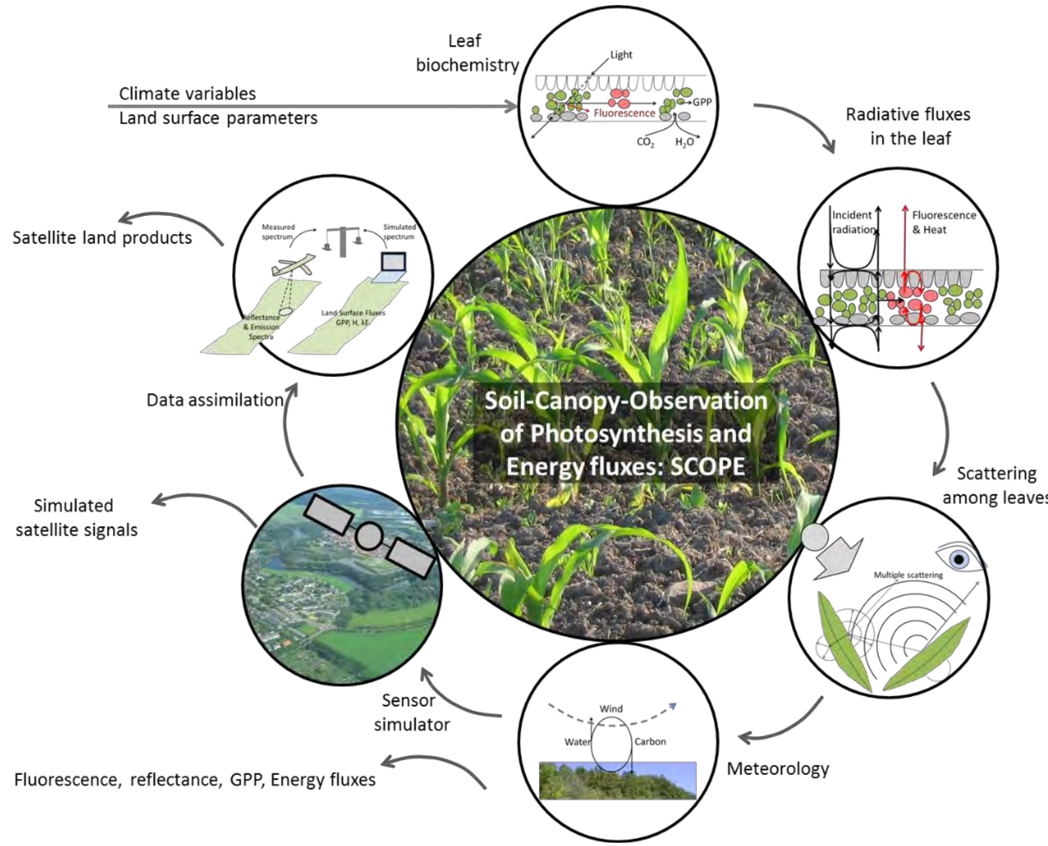


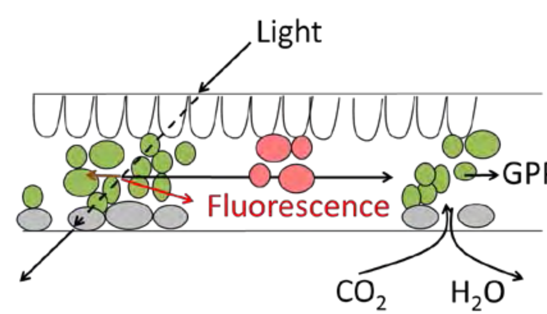
Examples of canopy RTMs (2/4)

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









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

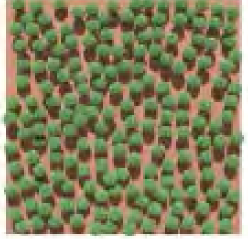
Vegetation is built using The Botanical Plant Modelling System (BPMS)
BPMS is a form of L-systems - the branches of a tree as geometric primitives

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

(Dürer, 1525)

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

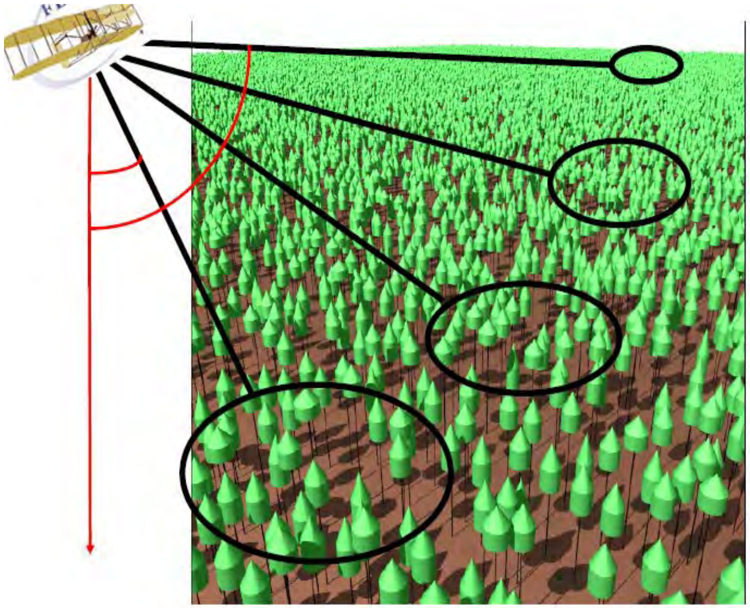
Examples of canopy RTMs (4/4)



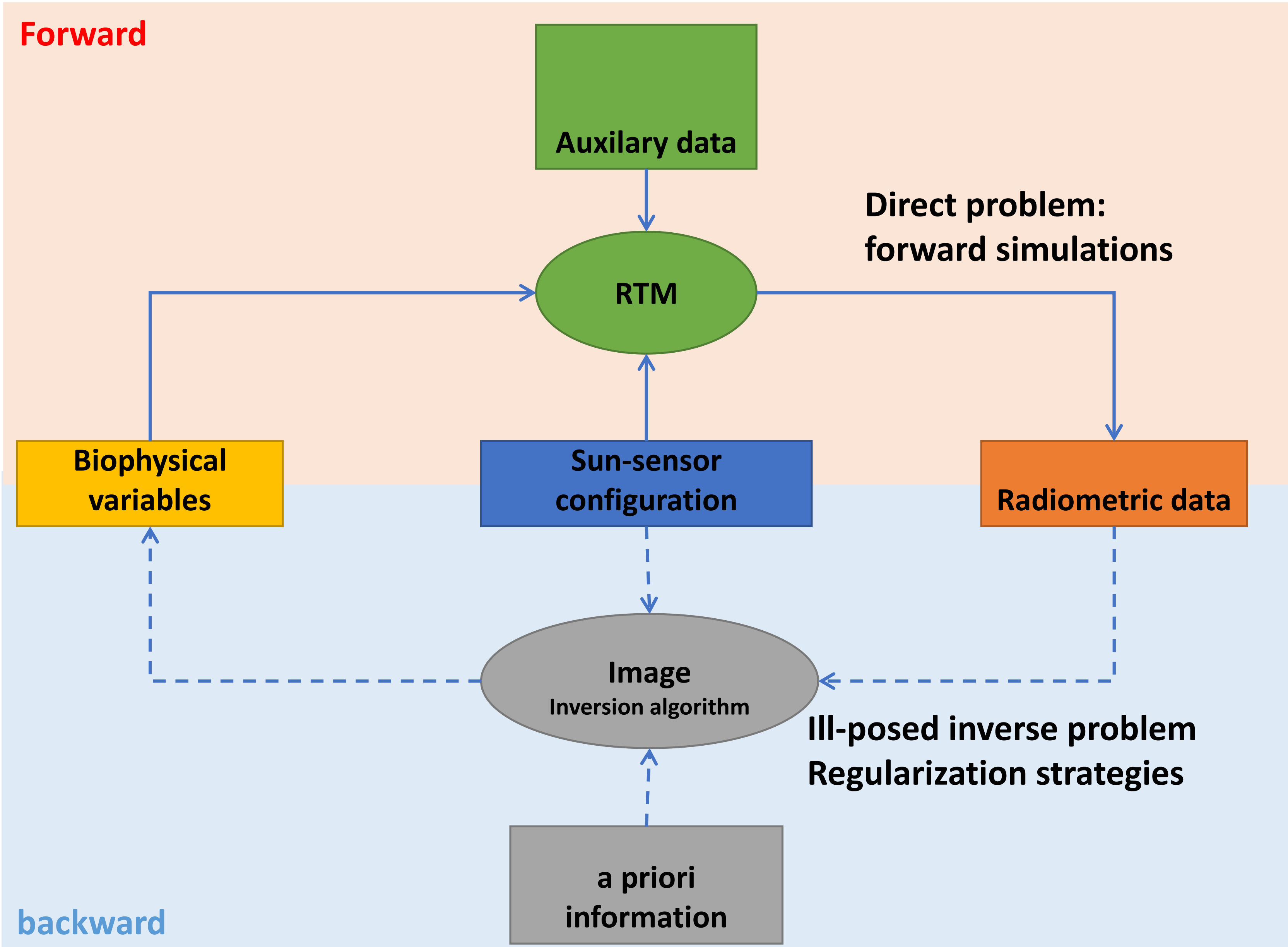
FLIGHT (North, 1996): A 3-D model

FLIGHT MC ray tracing approach

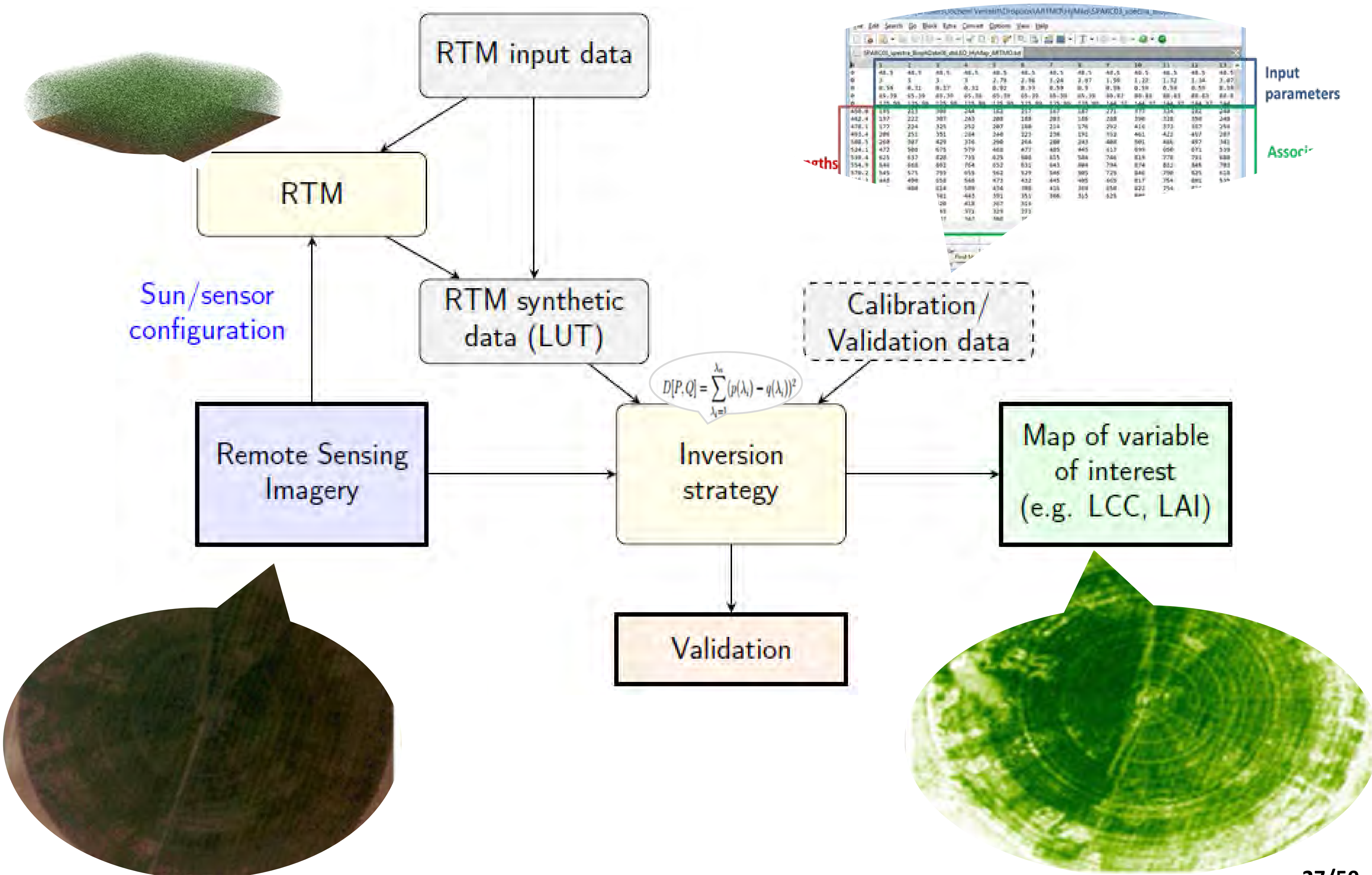
- Large scale structure by geometric primitives (e.g. cone)
- Foliage within crowns described by volume-averaged parameters
- 3D photon trajectories are simulated, accounting for the probabilities of free path, absorption and scattering
- Individual photon trajectories are traced from a solar source,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



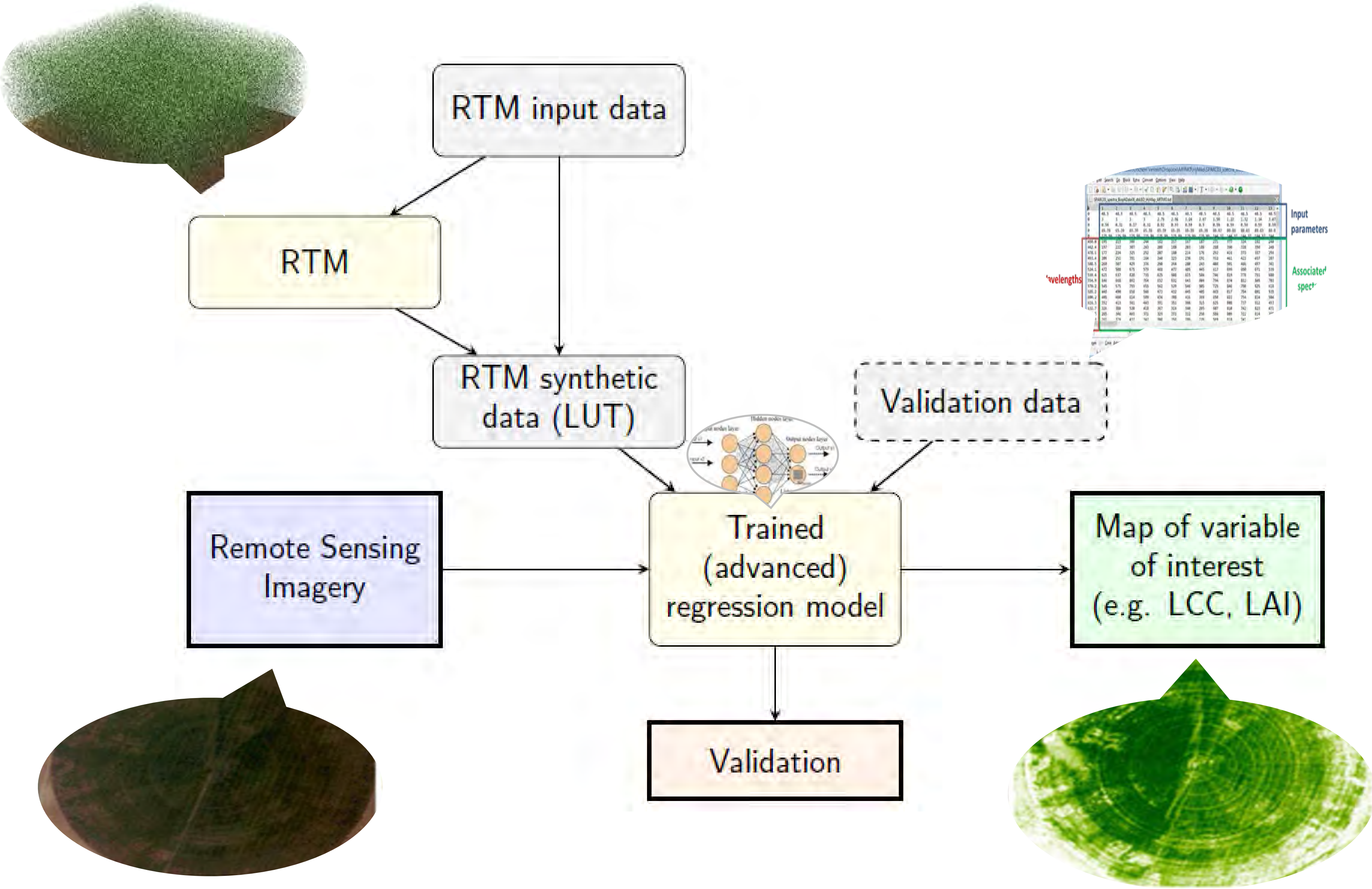
Biophysical parameters retrieval through RTM inversion:



LUT-based RTM inversion



Hybrid retrieval



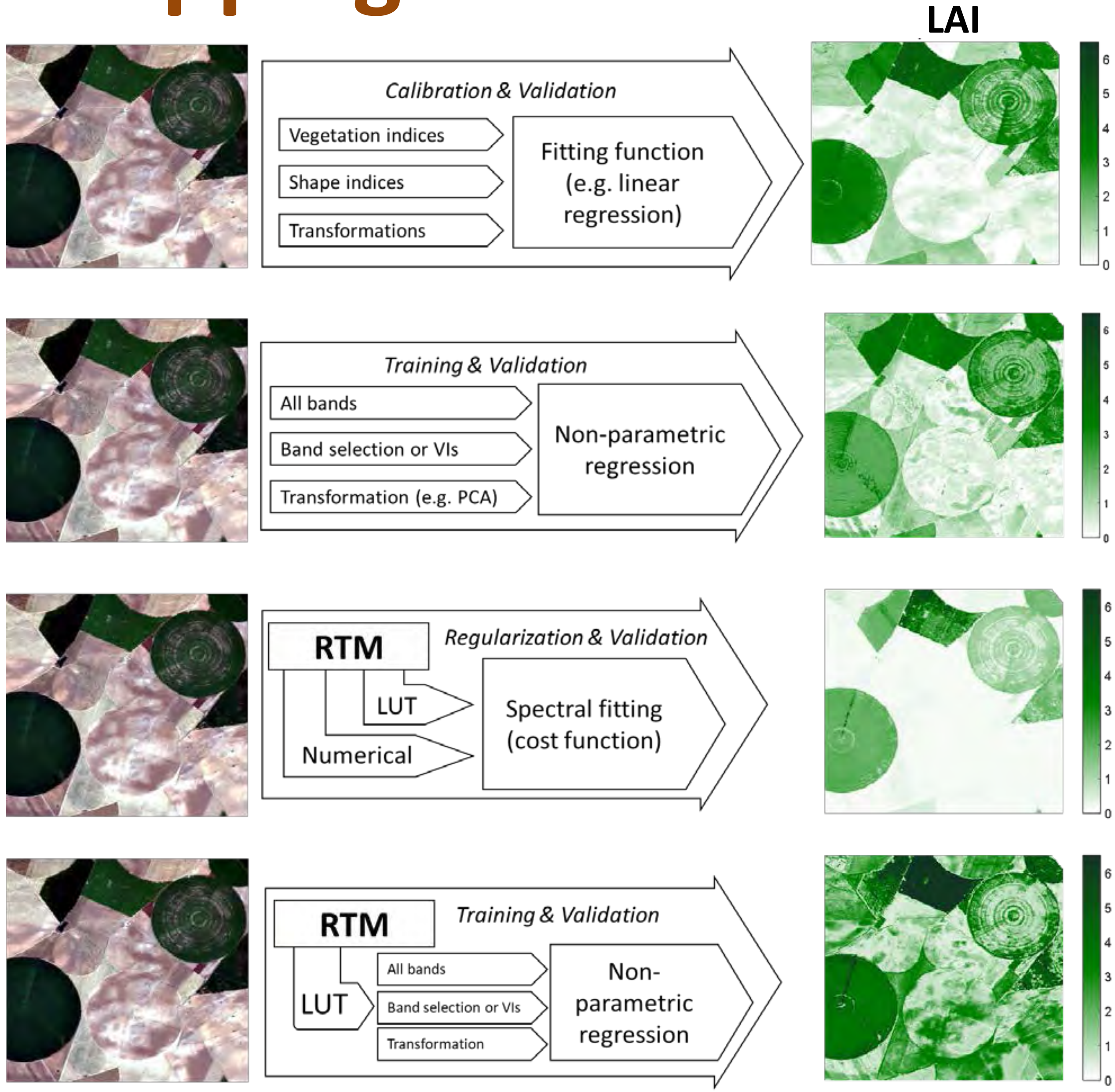
Summary mapping methods

Parametric regression

Non-parametric regression

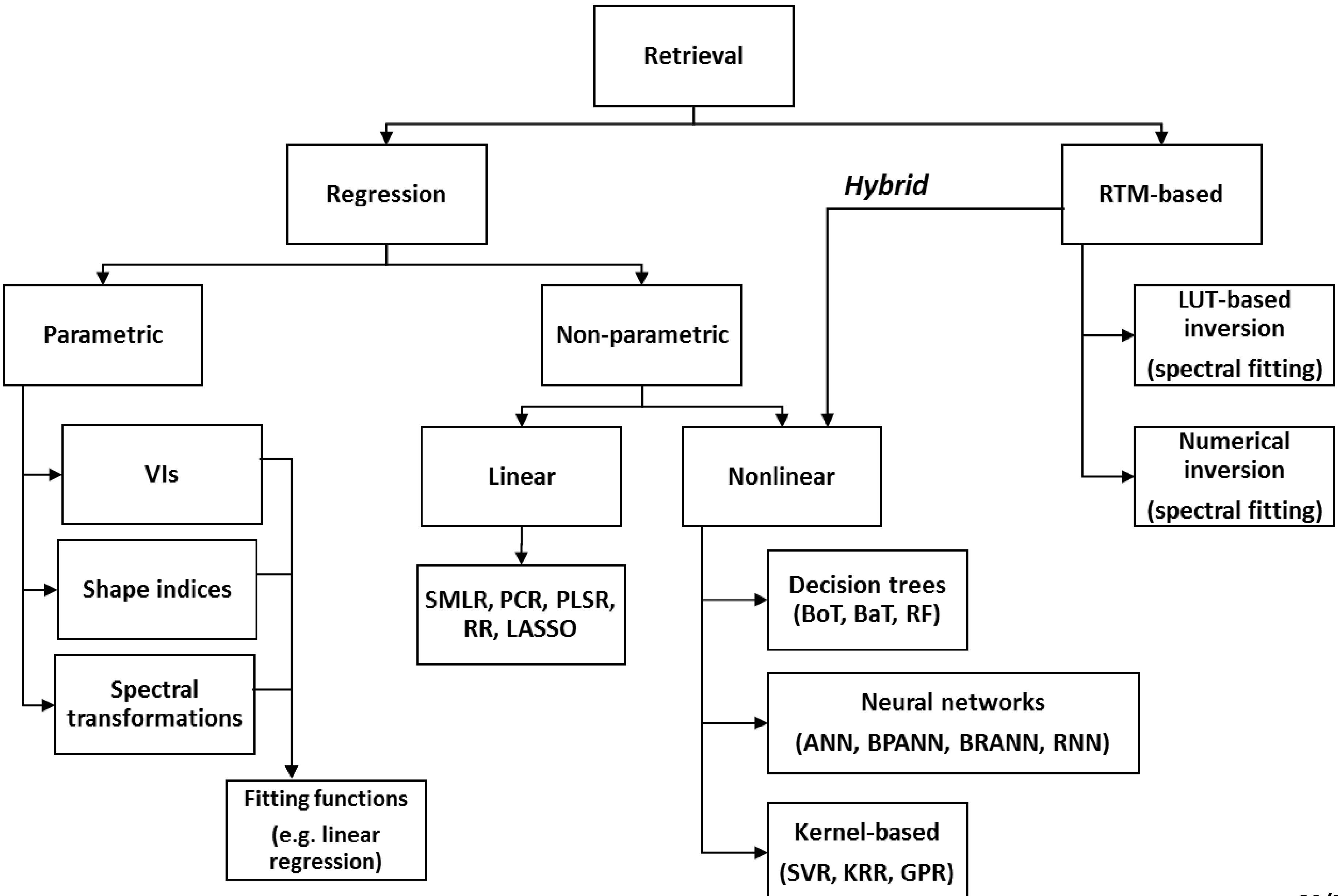
RTM inversion

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?

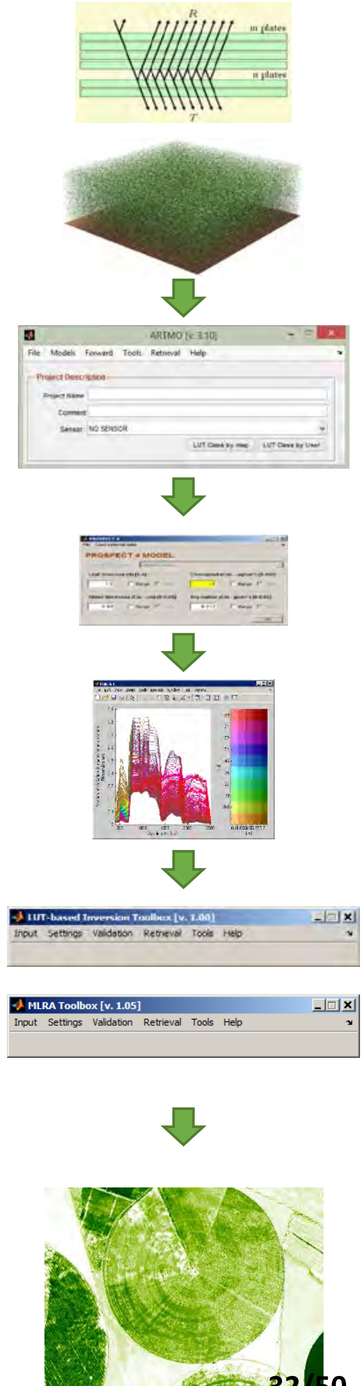
New missions…
New sensors…

The many decisions to be taken require a systematic evaluation

ARTMO automates retrieval optimization

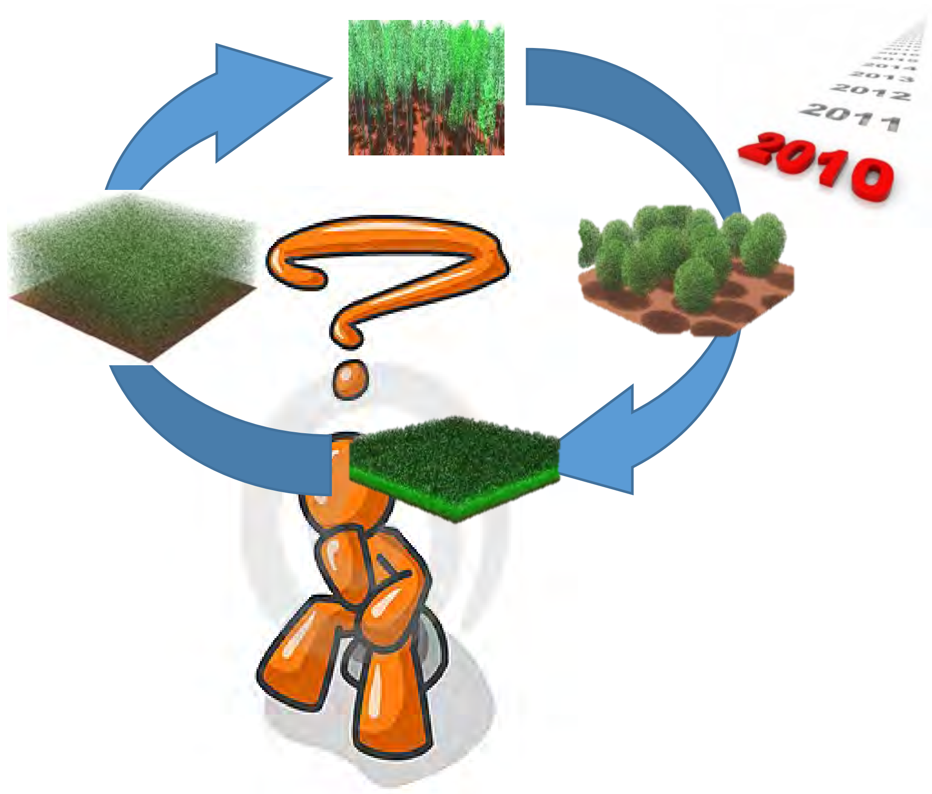
We will learn:

- **Using leaf and canopy radiative transfer models (RTMs)**
- **ARTMO**
  - **Run forward leaf & canopy RTMS**
  - **Graphics**
  - Sensor
  - Applications
- Retrieval toolboxes
  - LUT-based inversion
  - Machine Learning Regression Algorithms (MLRAs)
  - Spectral indices

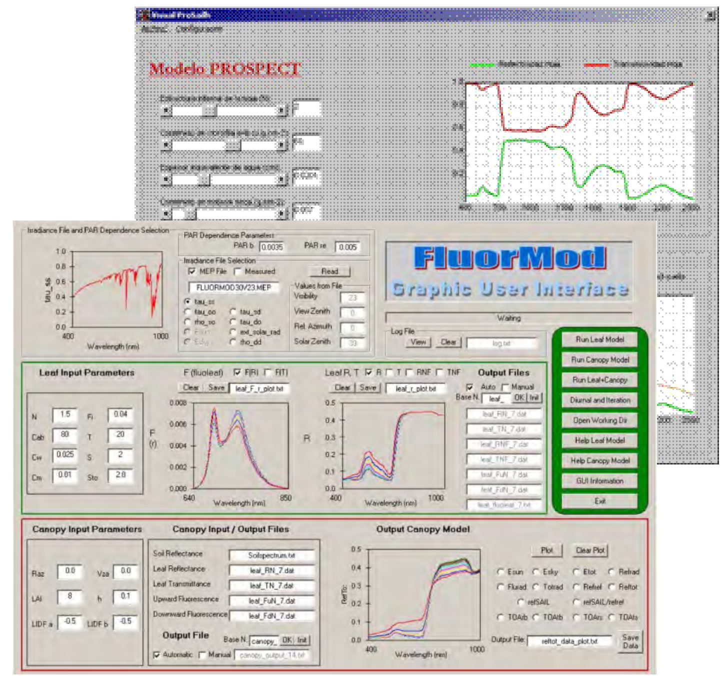


RTMs are important tools in EO research but for the broader community these models are perceived as complicated. Only very few of them offer user-friendly interfaces (GUIs).

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:

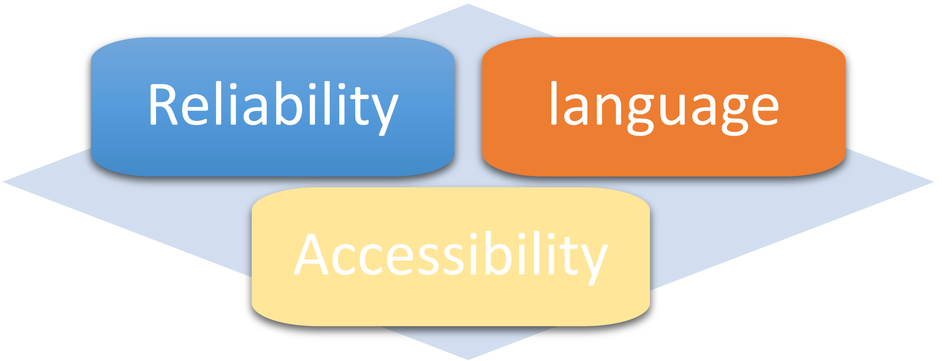
➢ **To develop a GUI toolbox that:**

- 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

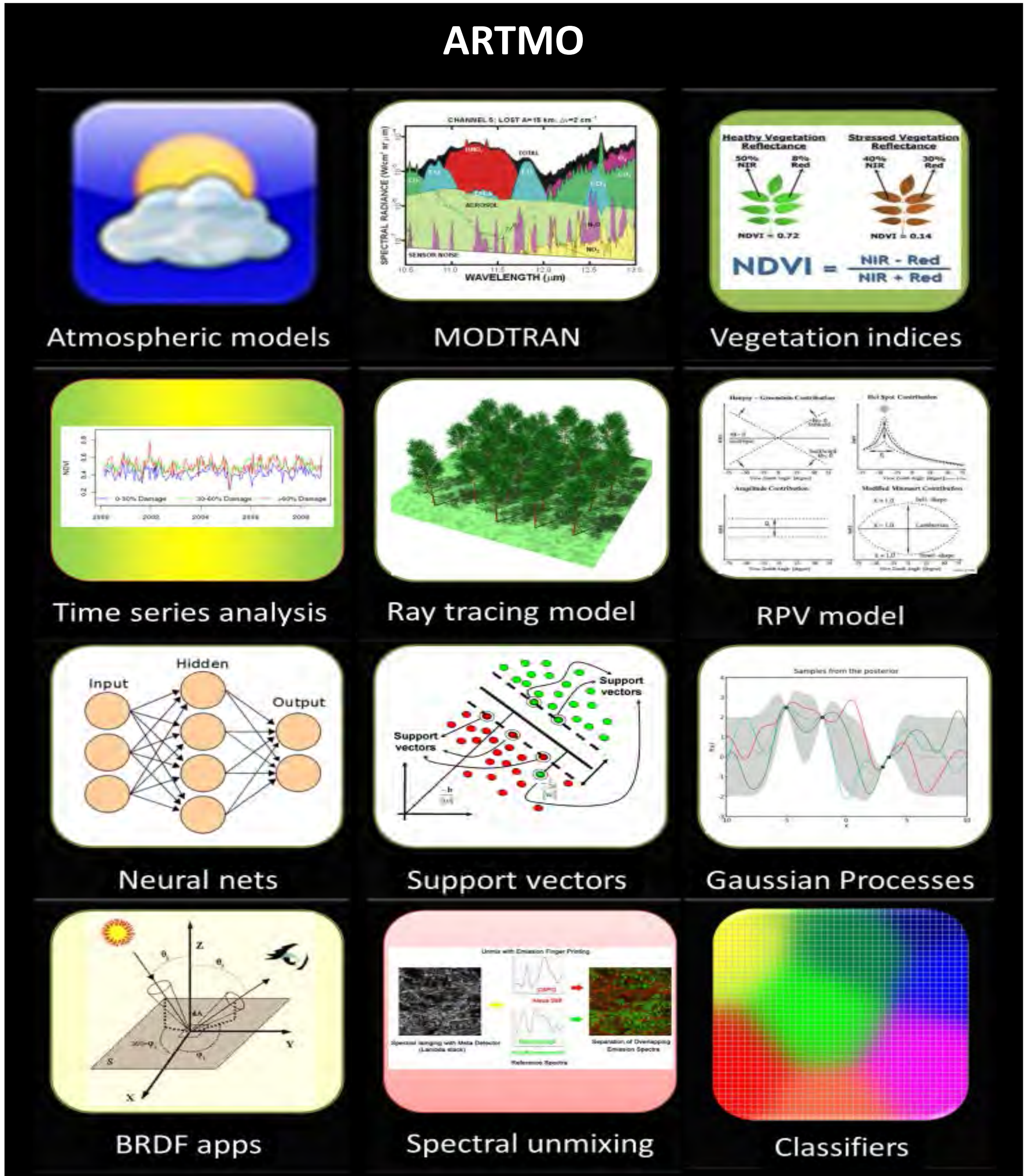


Model	Reference	Source code
PROSPECT-4	Feret et al., 2008	Matlab
PROSPECT-5	Feret et al., 2008	Matlab
PROSPECT-D	Feret et al., 2017	Matlab
DLM	Stuckens et al., 2009	Matlab
LIBERTY	Dawson et al., 1998	Matlab
FLUSPECT	Vilfan et al., 2016	Matlab
4SAIL	Verhoef et al., 2007	Matlab
FLIGHT	North, 1996	Executable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

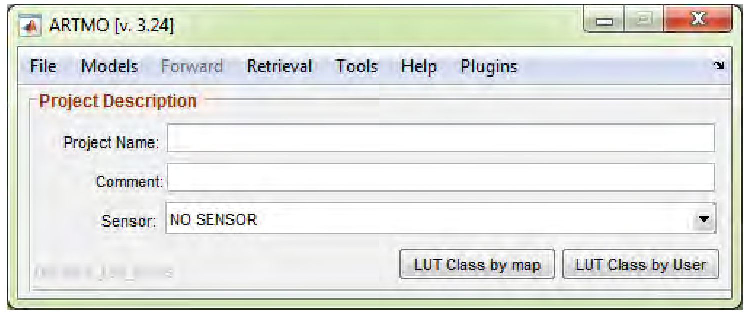
Software packages:

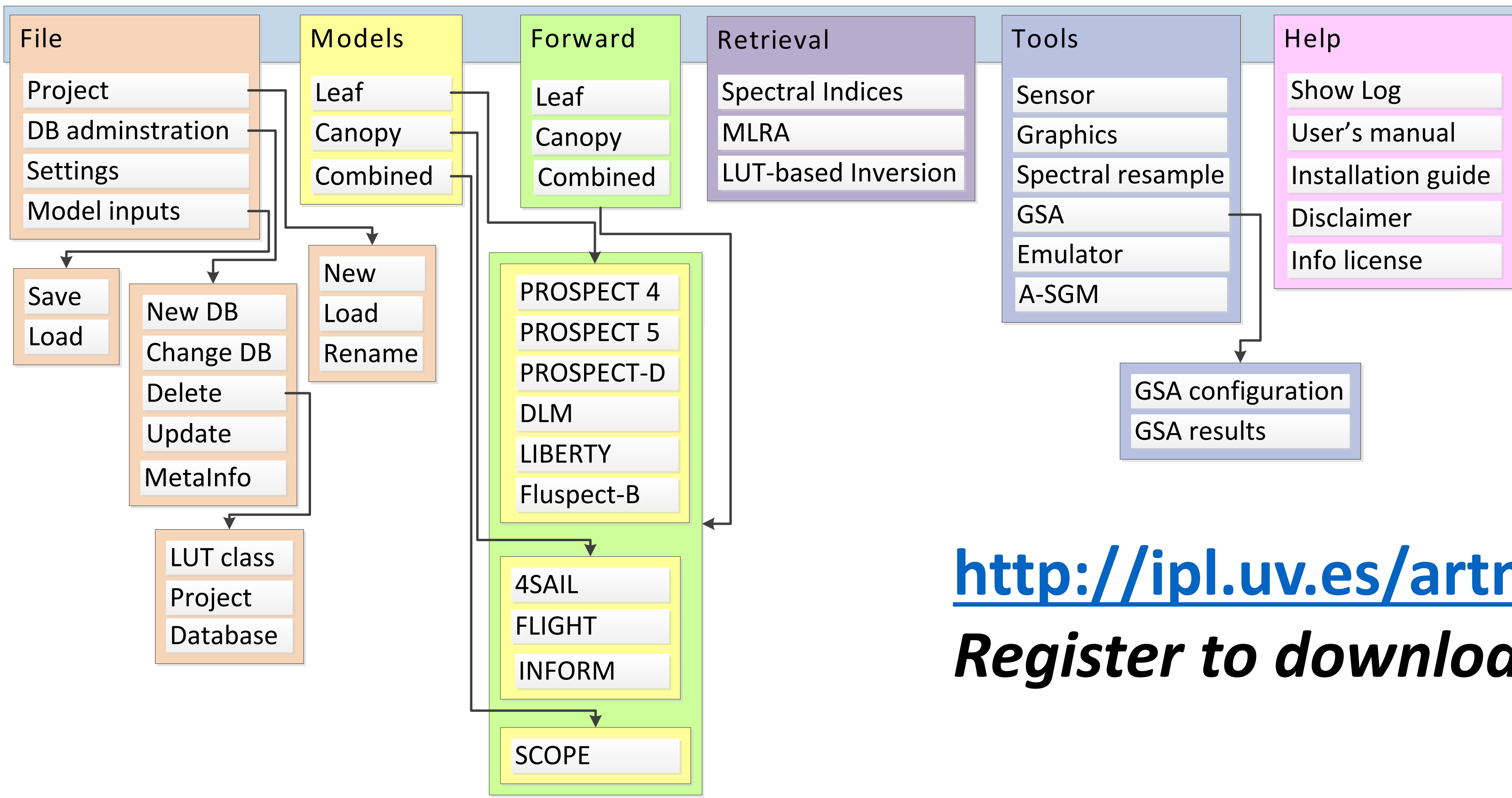
Programming language: Matlab®
Database: MySQL®
Image processing software: ENVI®

Leaf RTM
Canopy RTM
Combined RTM


ARTMO
Atmospheric models
MODTRAN
WAVELENGTH (μm)
SPECTRAL RADIANCE
Healthy Vegetation Reflectance
Stressed Vegetation Reflectance
NDVI = 0.72
NDVI = 0.14
NDVI = NIR - Red / NIR + Red
Vegetation indices
Time series analysis
Ray tracing model
RPV model
Input
Hidden
Output
Neural nets
Support vectors
Support vectors
Support vectors
Samples from the posterior
Gaussian Processes
BRDF apps
Spectral unmixing
Classifiers

ARTMO v. 3.24



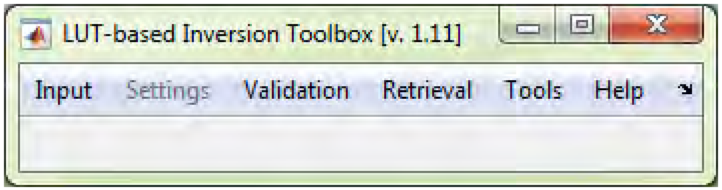


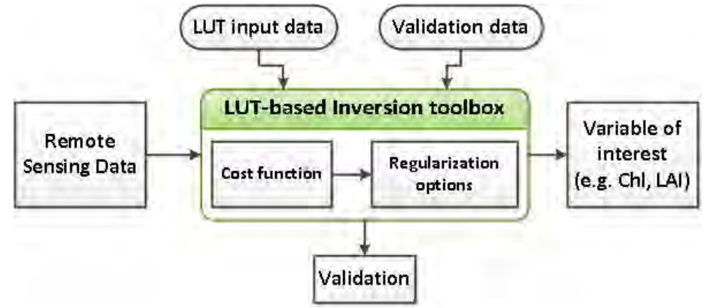
http://ipl.uv.es/artmo/

Register to download!

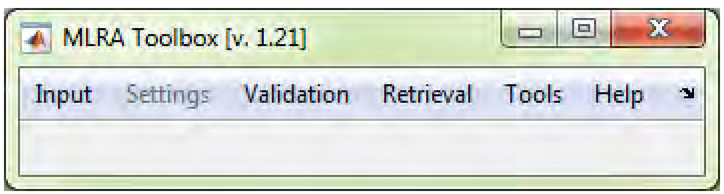
ARTMO's retrieval toolboxes:

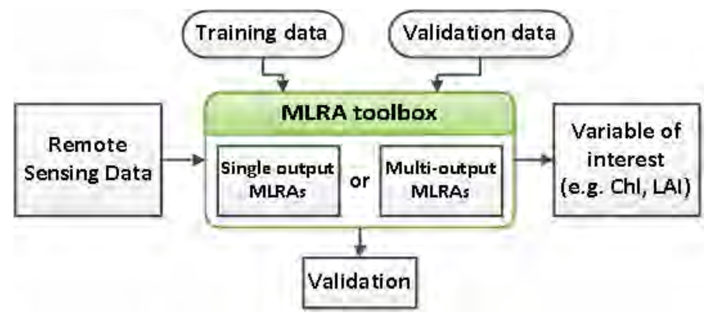
LUT-based inversion toolbox



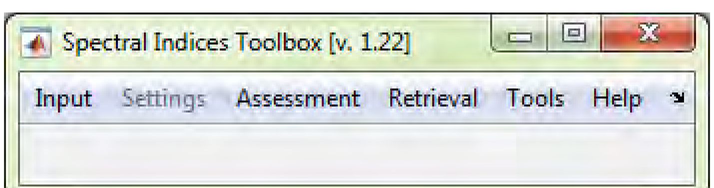


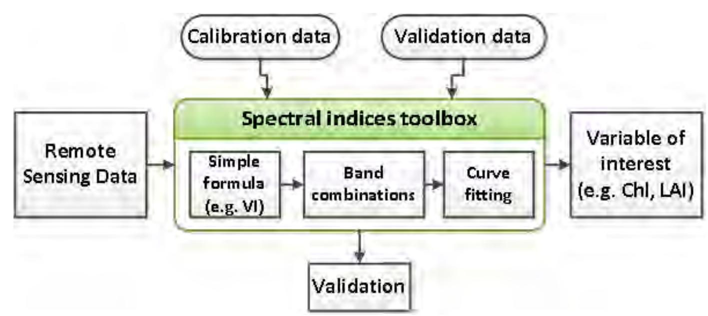
Machine learning regression algorithm toolbox (MLRA)





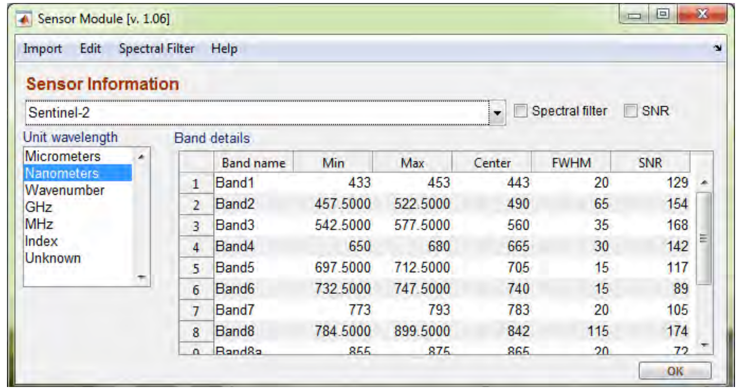
Spectral indices toolbox

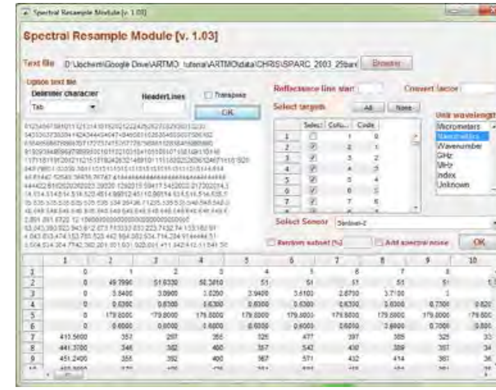




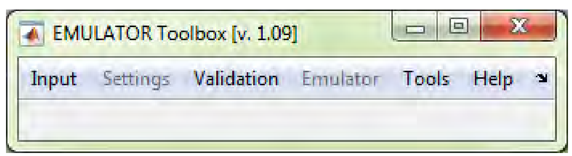
ARTMO's tools:

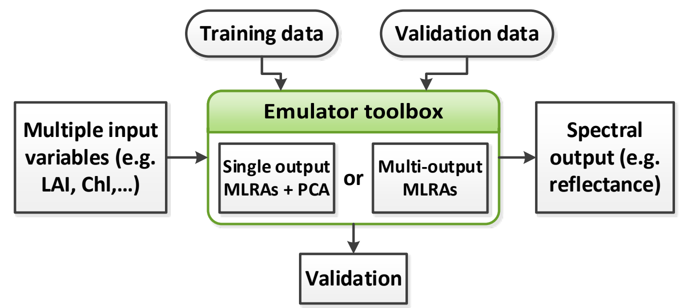
Sensor & spectral resample:



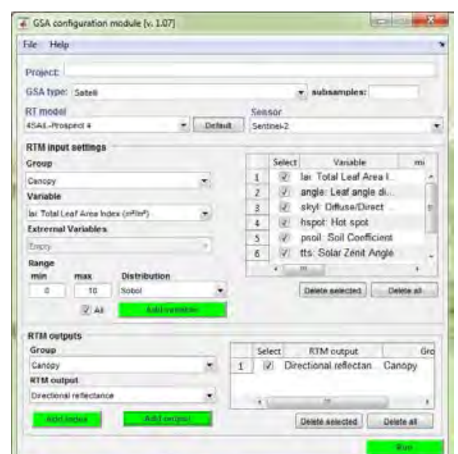


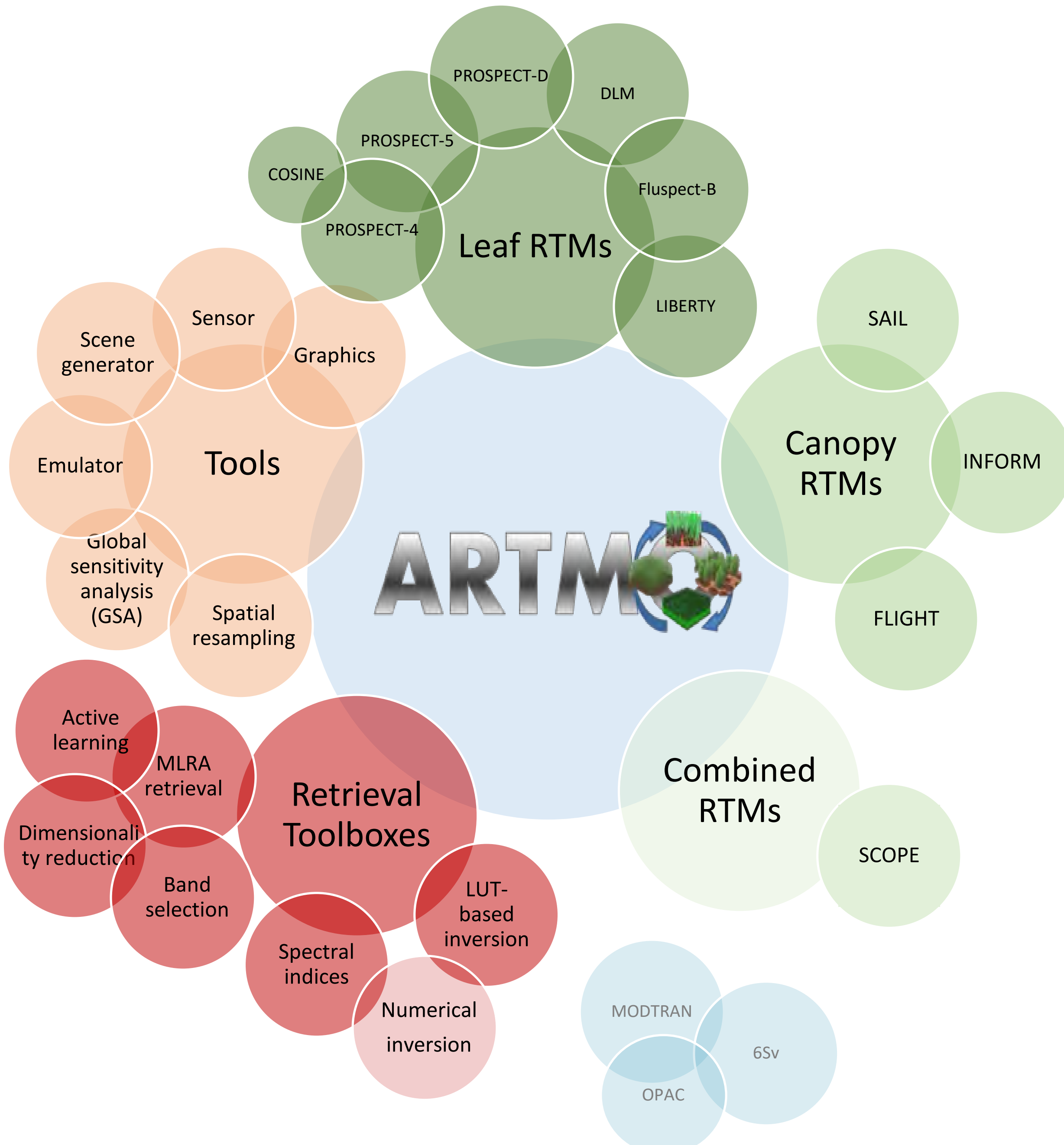
Emulation:





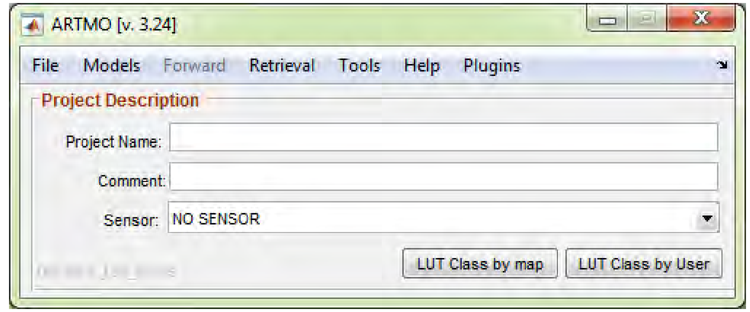
Global sensitivity analysis:

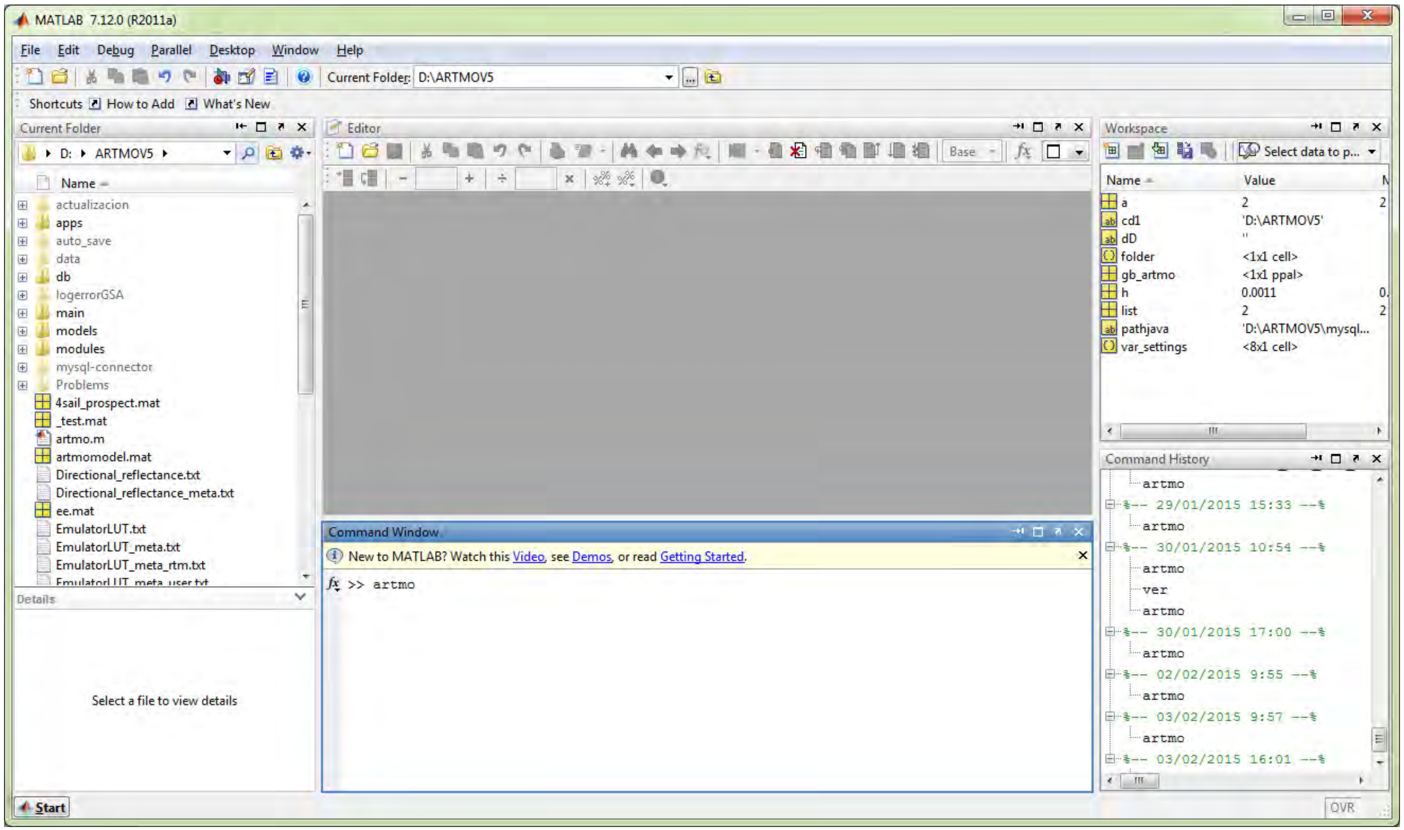




http://ipl.uv.es/artmo/

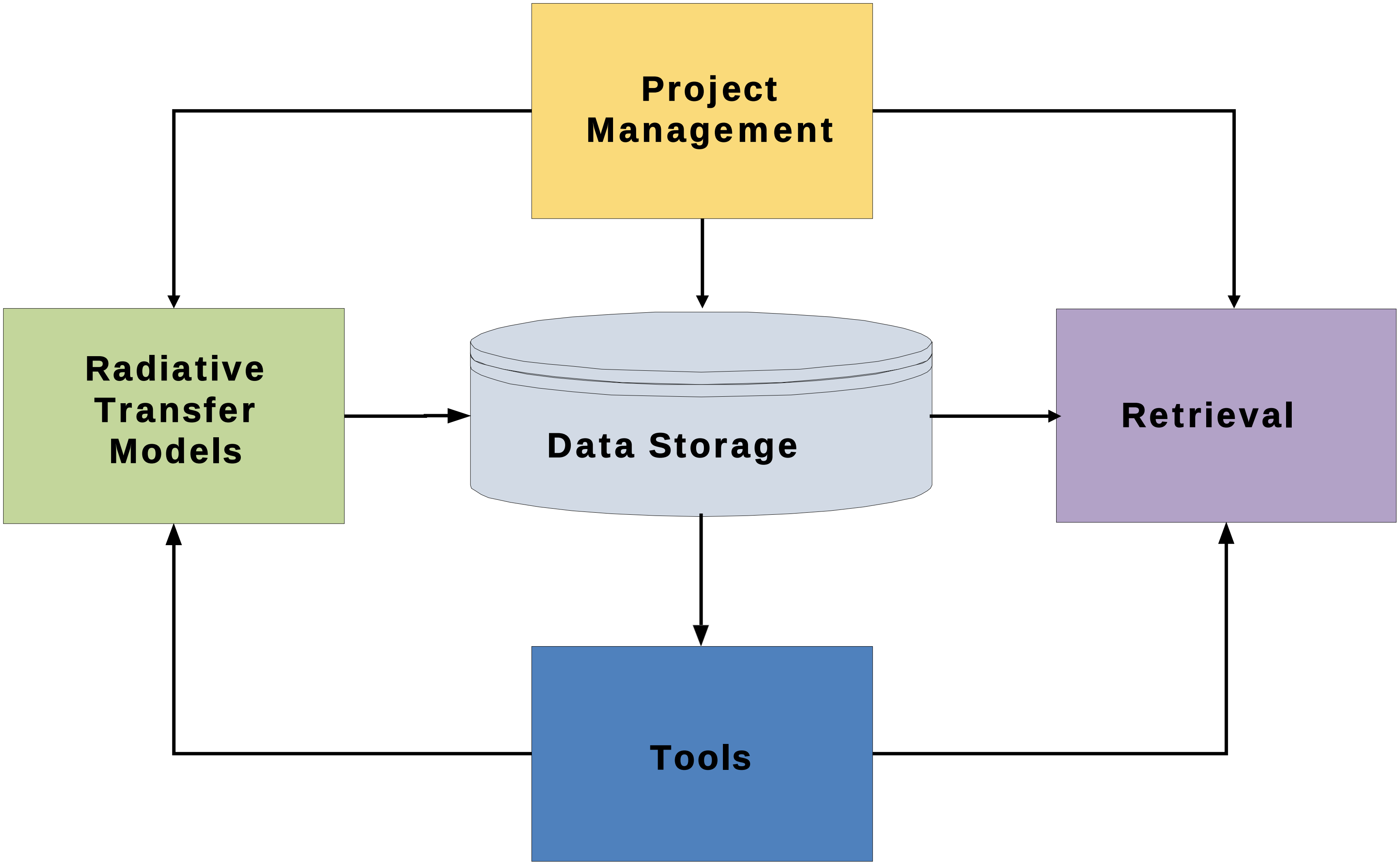
Opening ARTMO



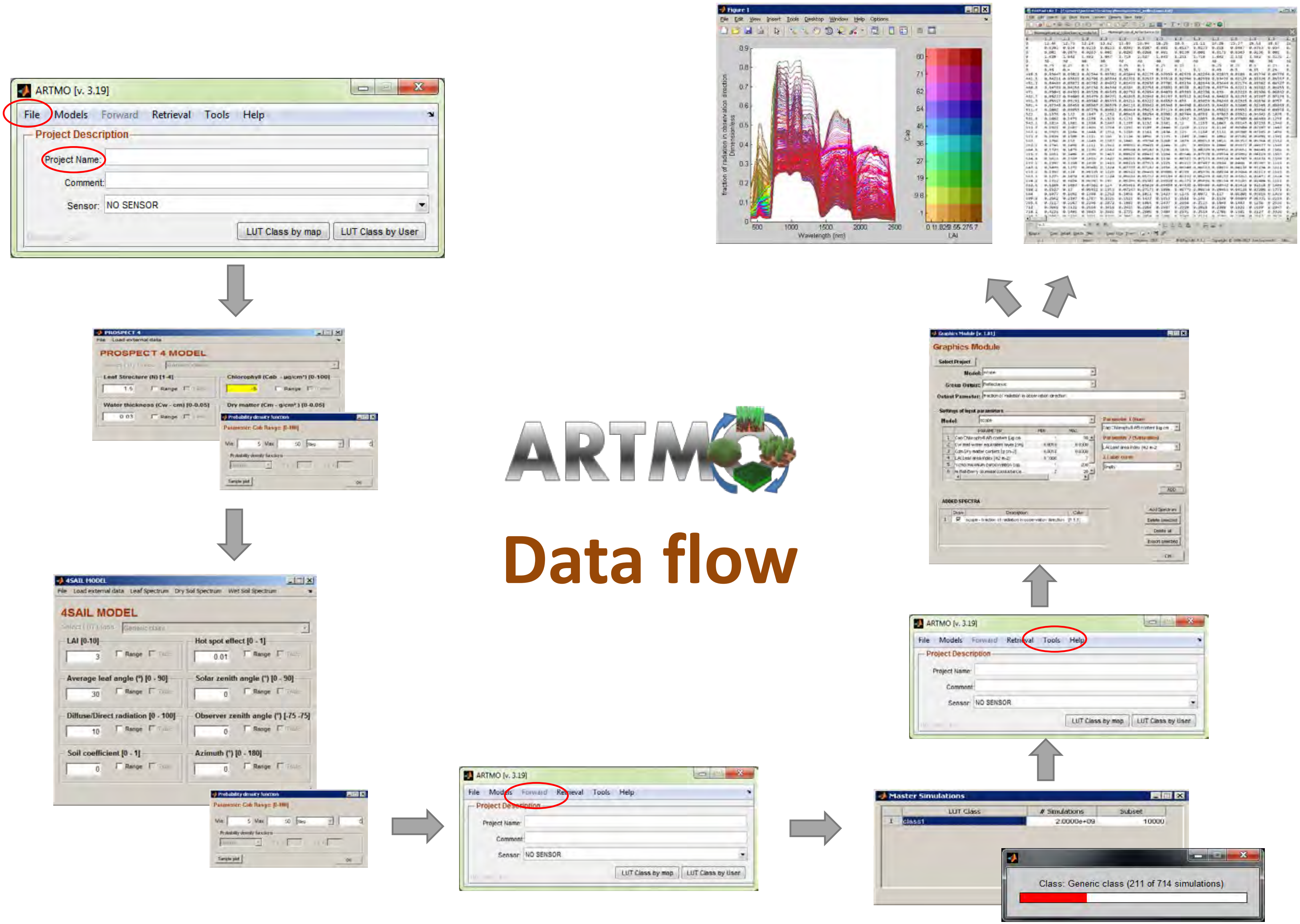


- Set Matlab's Current folder to the ARTMO folder.
- Type in Command window: **artmo**

Conceptual architecture ARTMO

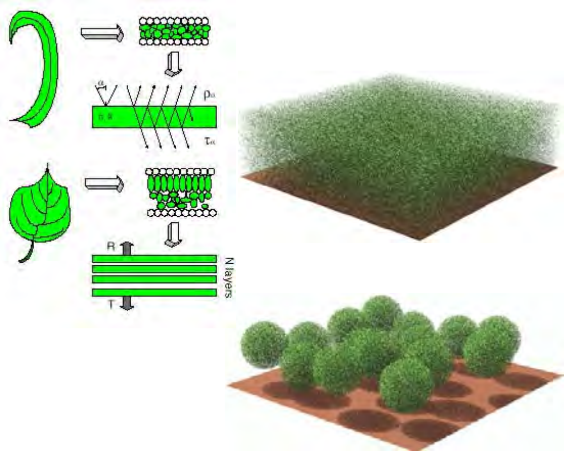


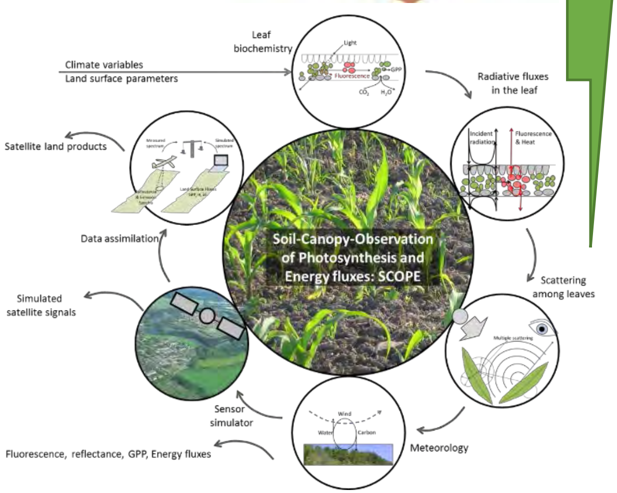
Data flow

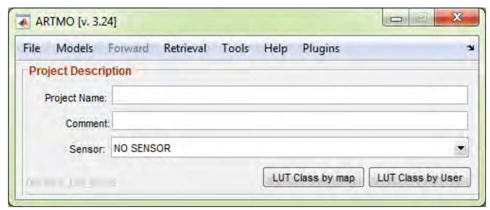

ARTMO [v. 3.19]
File Models Forward Retrieval Tools Help
Project Description
Project Name:
Comment:
Sensor: NO SENSOR
LUT Class by map
LUT Class by User
PROSPECT 4 MODEL
4SAIL MODEL
LAI [0-10]
Hot spot effect [0 - 1]
Average leaf angle (°) [0 - 90]
Solar zenith angle (°) [0 - 90]
Diffuse/Direct radiation [0 - 100]
Observer zenith angle (°) [-75 -75]
Soil coefficient [0 - 1]
Azimuth (°) [0 - 180]
Master Simulations
LUT Class
Simulations
Subset
class1
2.0000e+09
10000
Class: Generic class (211 of 714 simulations)
Graphics Module
Figure 1
Wavelength (nm)
LAI

Automated generation of spectral data

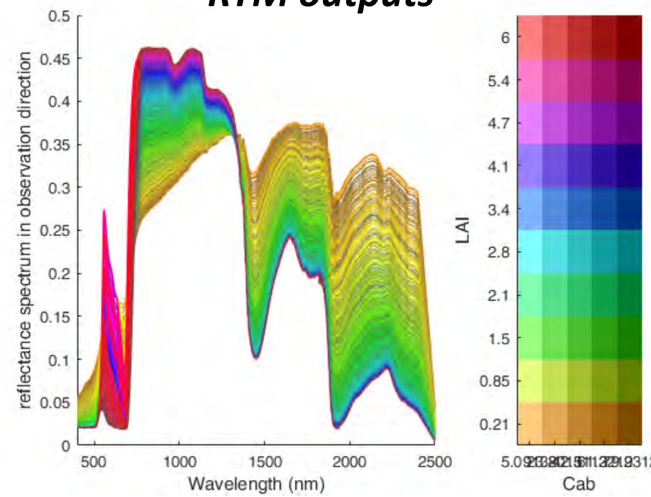
RTM inputs

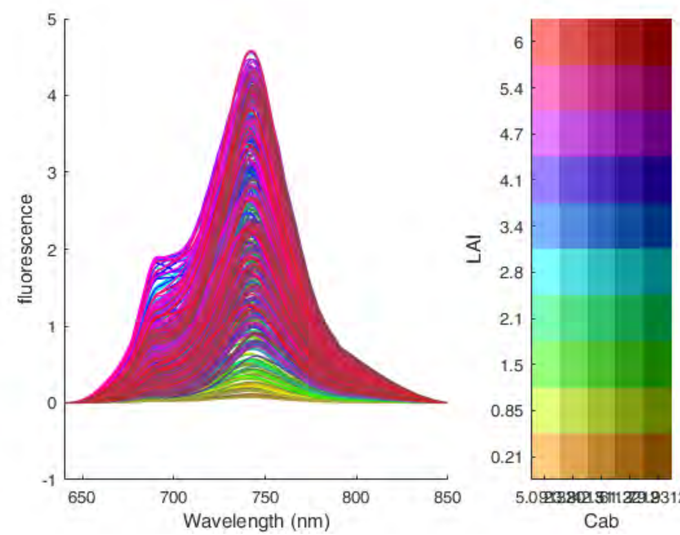






RTM outputs

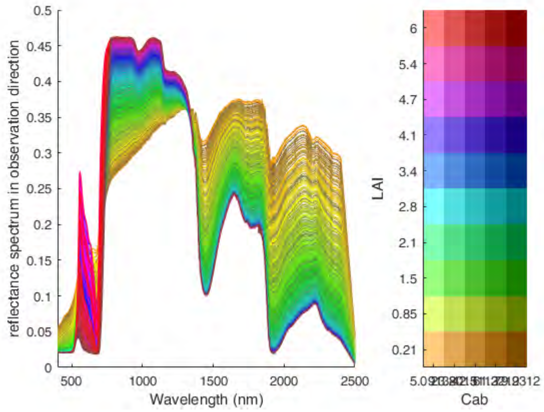




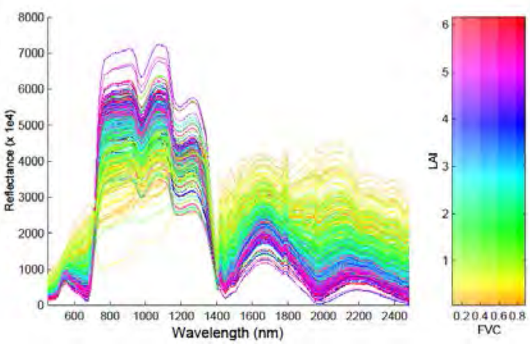
Spectral data can be reflectance, transmittance, fluorescence

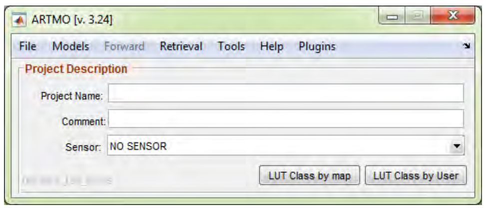
Automated mapping of vegetation properties

Simulated data

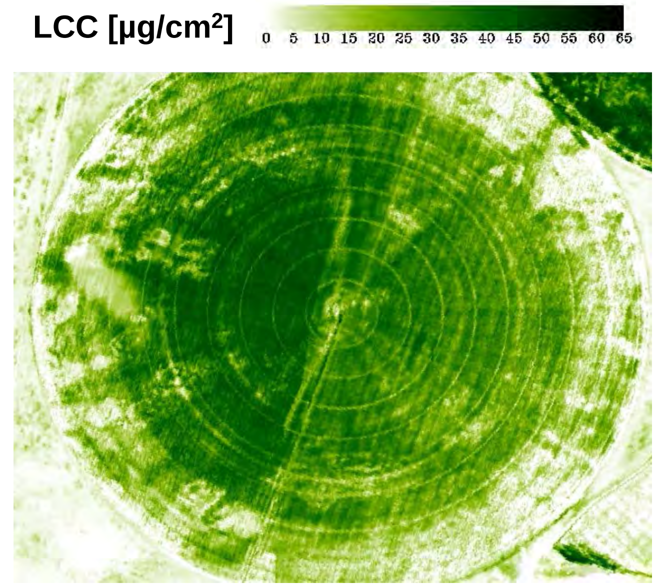


Measured data

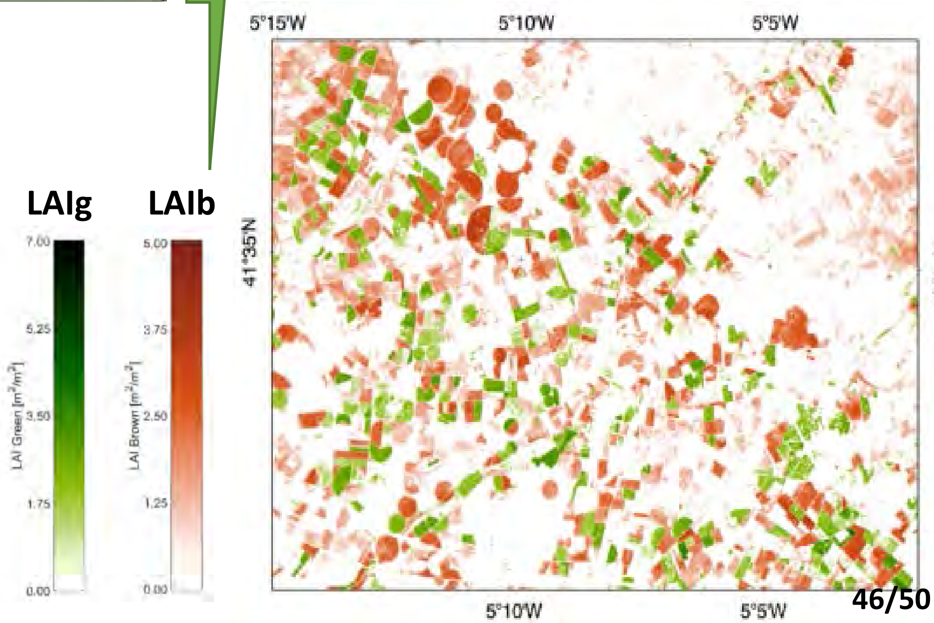




LCC [µg/cm²]



LAIg LAIb

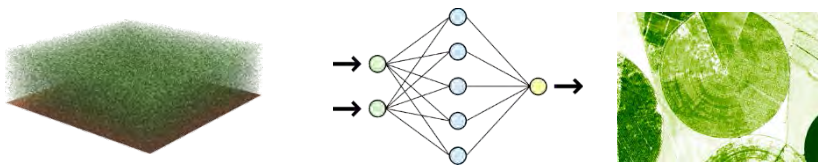


Lectures for following days

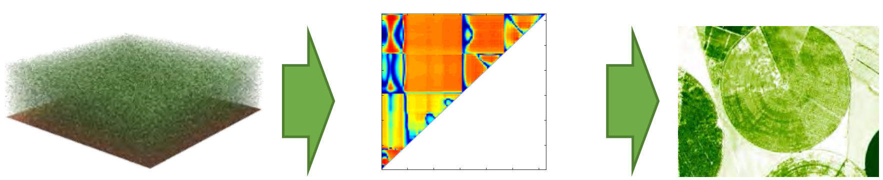
goo.gl/5DzDkx

1. **Forward running leaf and canopy radiative transfer models (RTMs)**

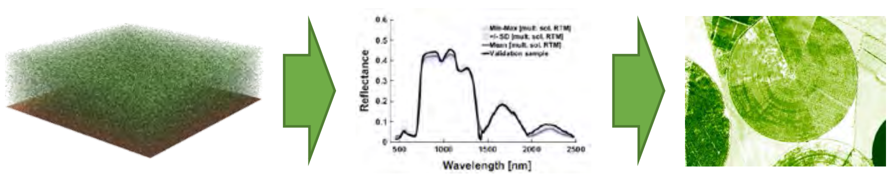
2. **Machine learning regression algorithms for vegetation properties mapping**



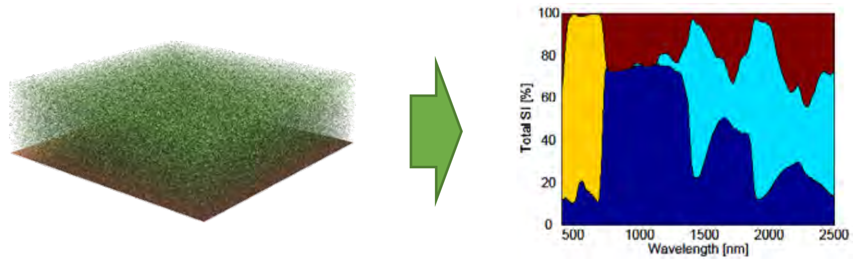
3. **Optimizing vegetation indices for vegetation properties mapping**



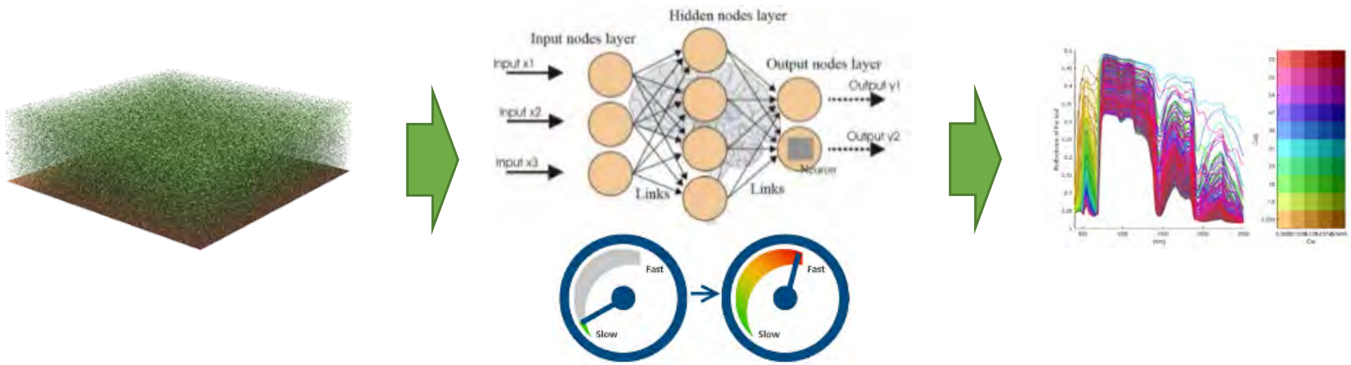
4. **Optimizing RTM inversion for vegetation properties mapping**



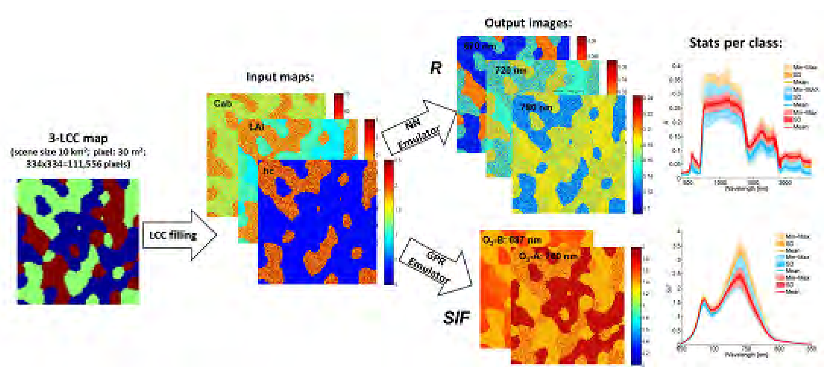
5. Global sensitivity analysis of RTMS



6. Emulation of RTMs



7. Synthetic scene generation



Lectures for following days

goo.gl/5DzDkx