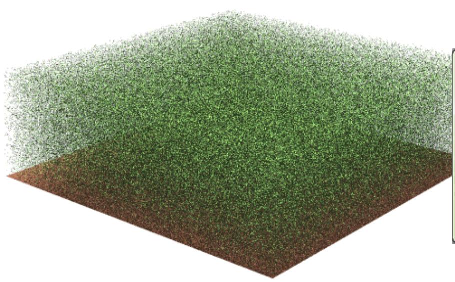
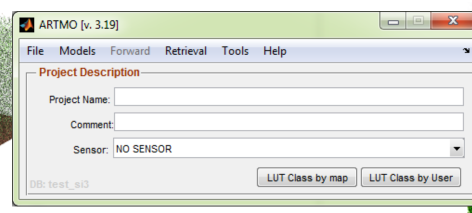
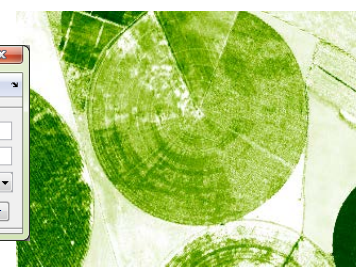
From SAIL simulations towards automated remote sensing applications:

an overview of 6 years of ARTMO developments

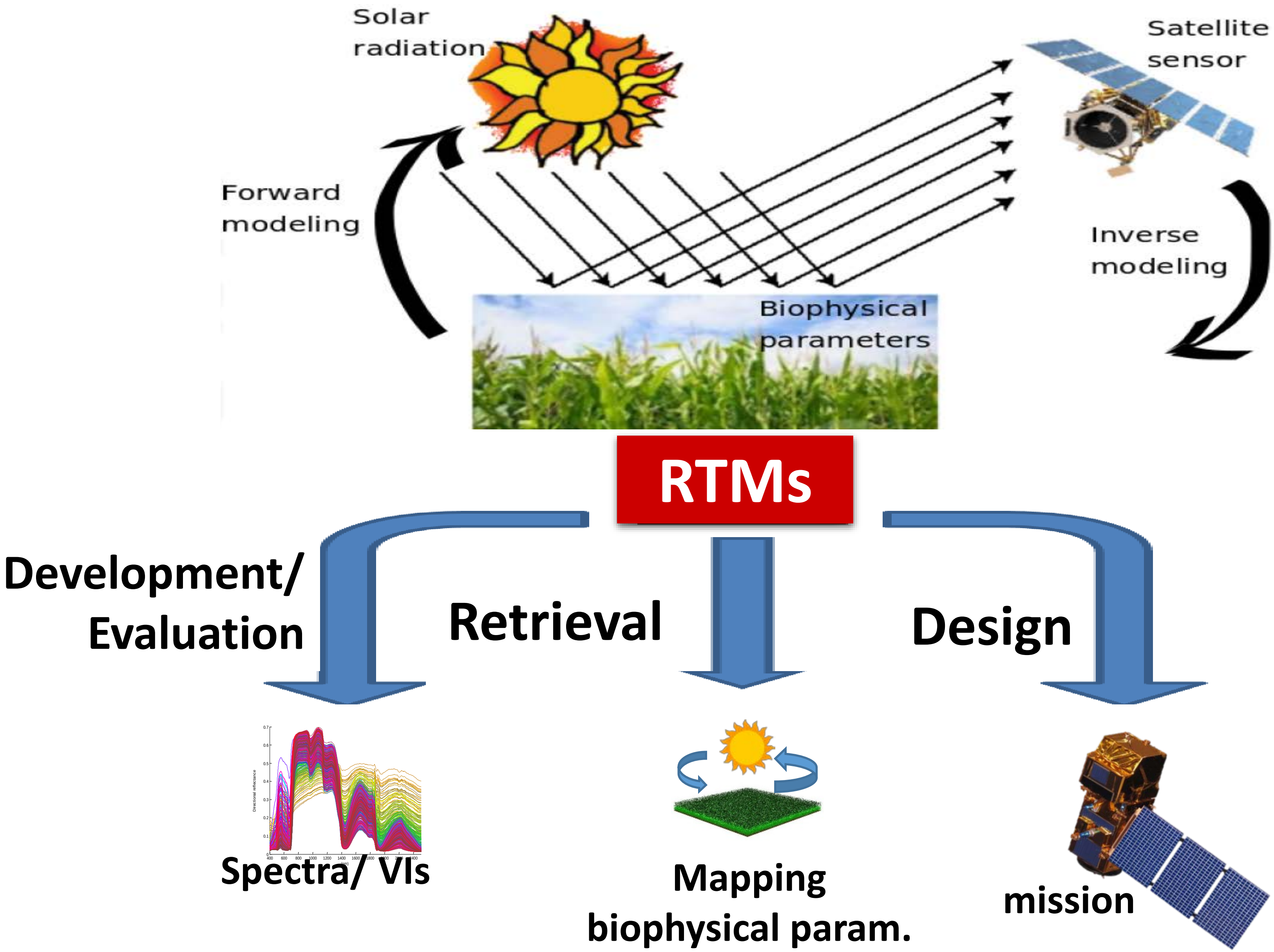





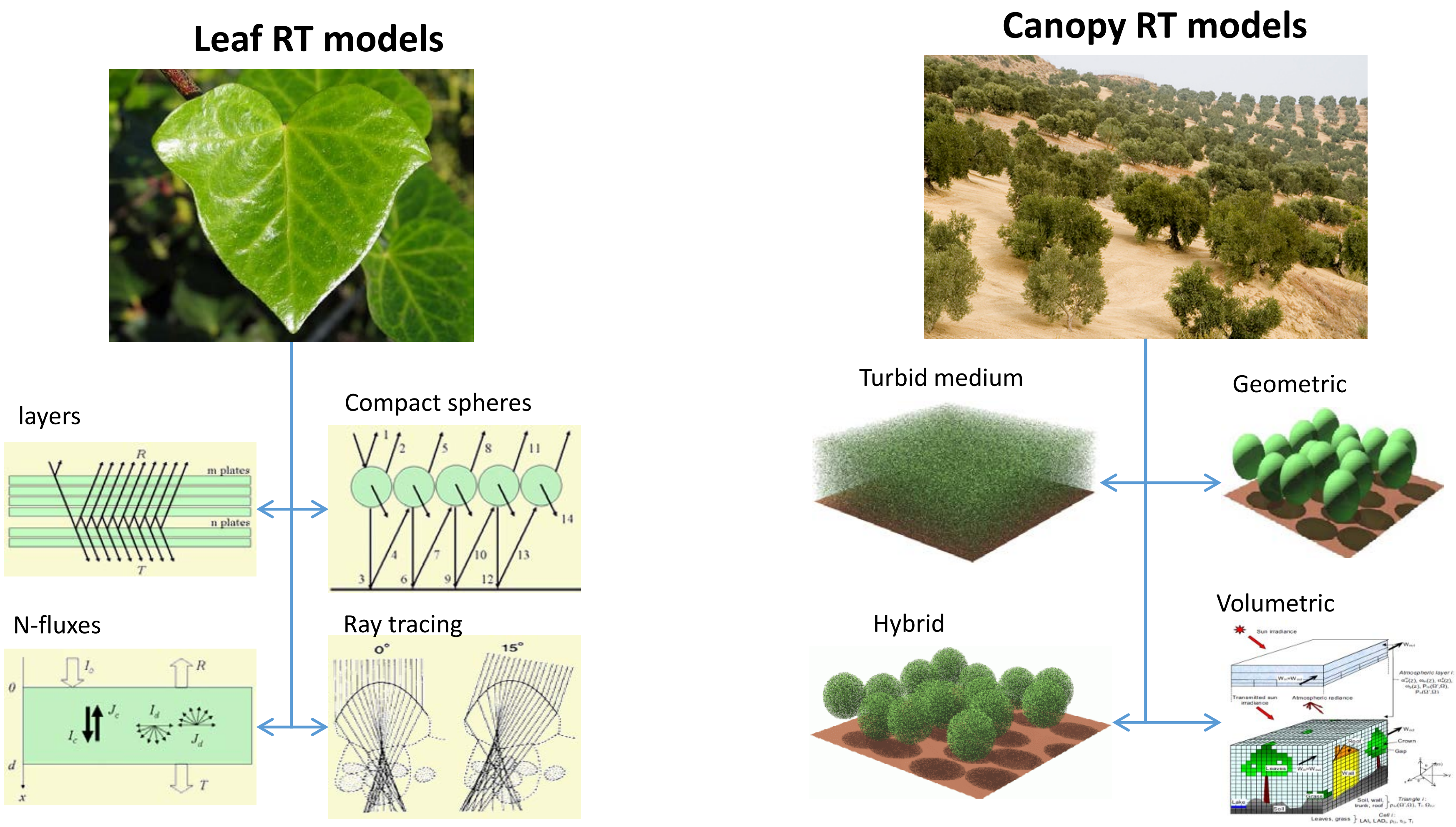
J. Verrelst, J.P. Rivera & J. Moreno

SAIL35 - 27 Sept 2016

Background


Solar radiation
Satellite sensor
Forward modeling
Inverse modeling
Biophysical parameters
RTMs
Development/ Evaluation
Retrieval
Design
Spectra/ VIs
Mapping biophysical param.
mission

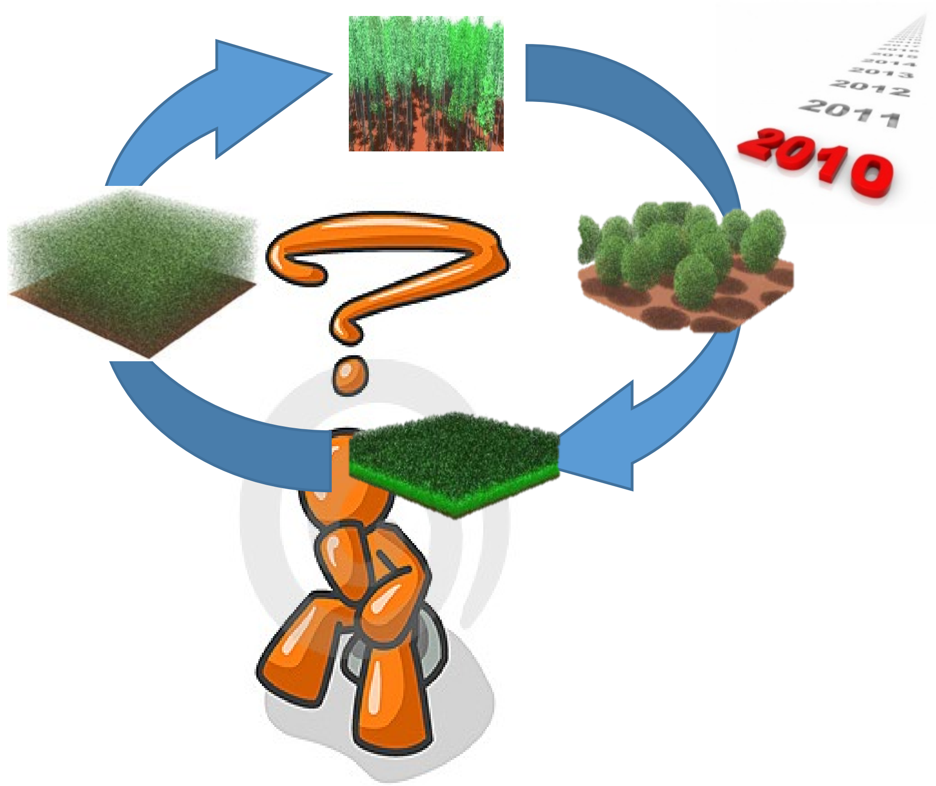
Radiative transfer models



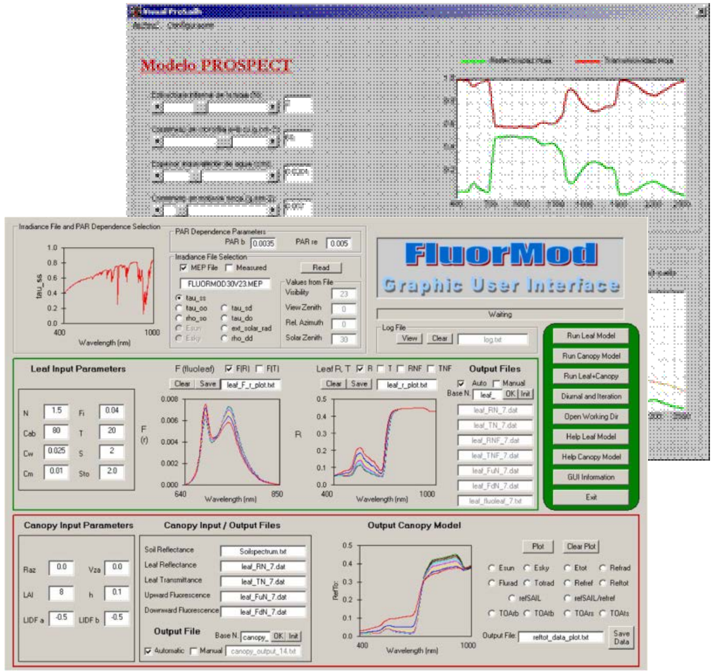
Various models exist with different complexity.

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.

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:

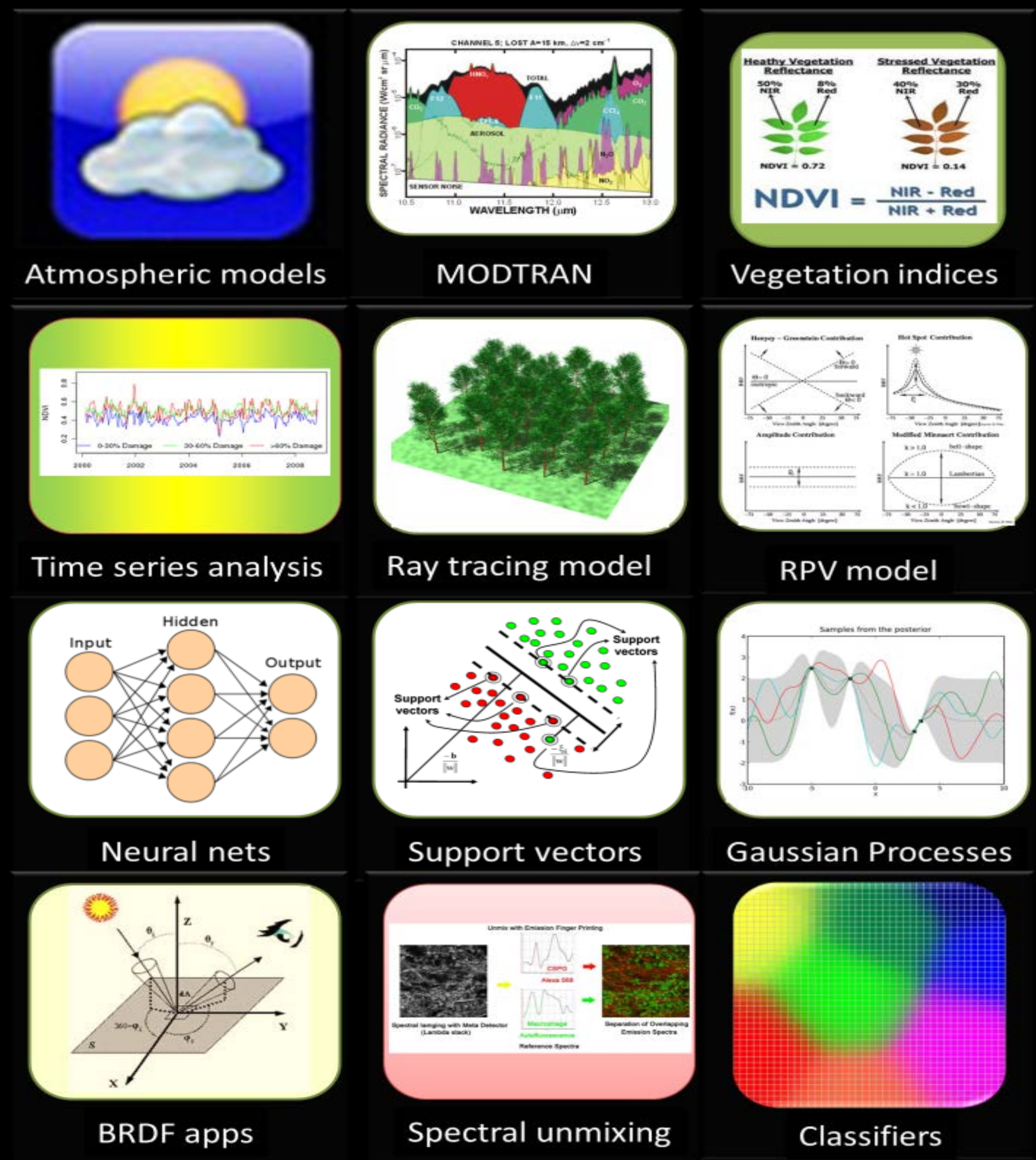
➢ **To develop a GUI toolbox that:**

- 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.

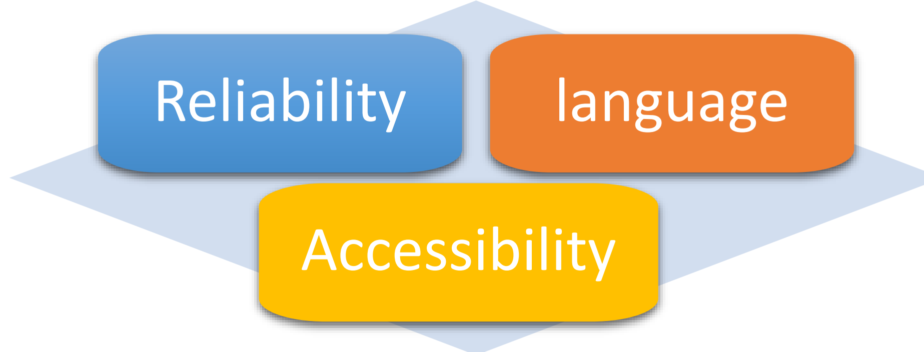
Toolbox for EO applications:

Automated
Radiative
Transfer
Models
Operator

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

Selection RTMs & programming language



v.1

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

Leaf RTM
Canopy RTM
Combined RTM

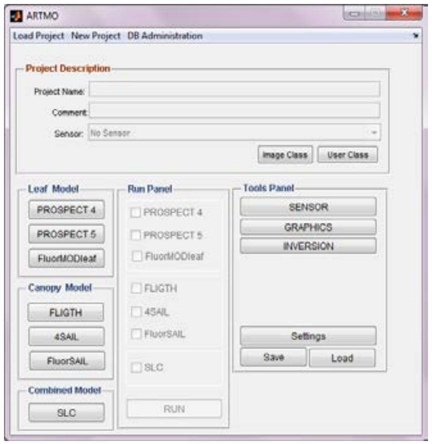
Software packages:

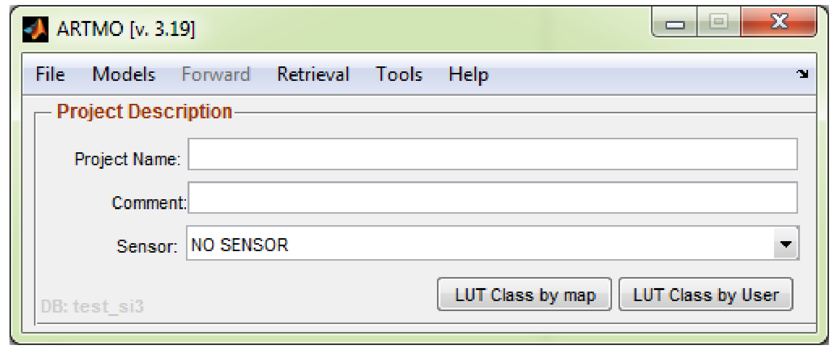
Programming language:

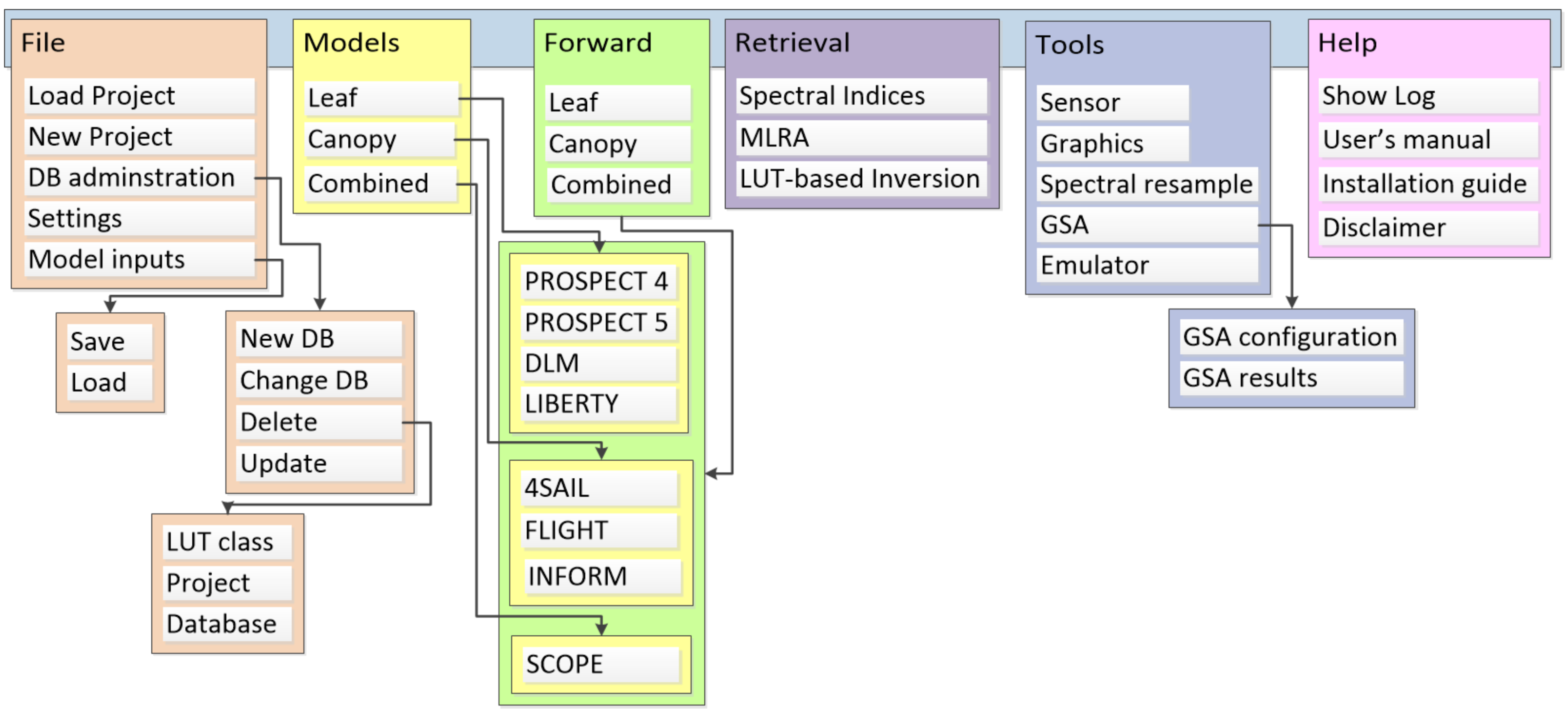
Database:

2012

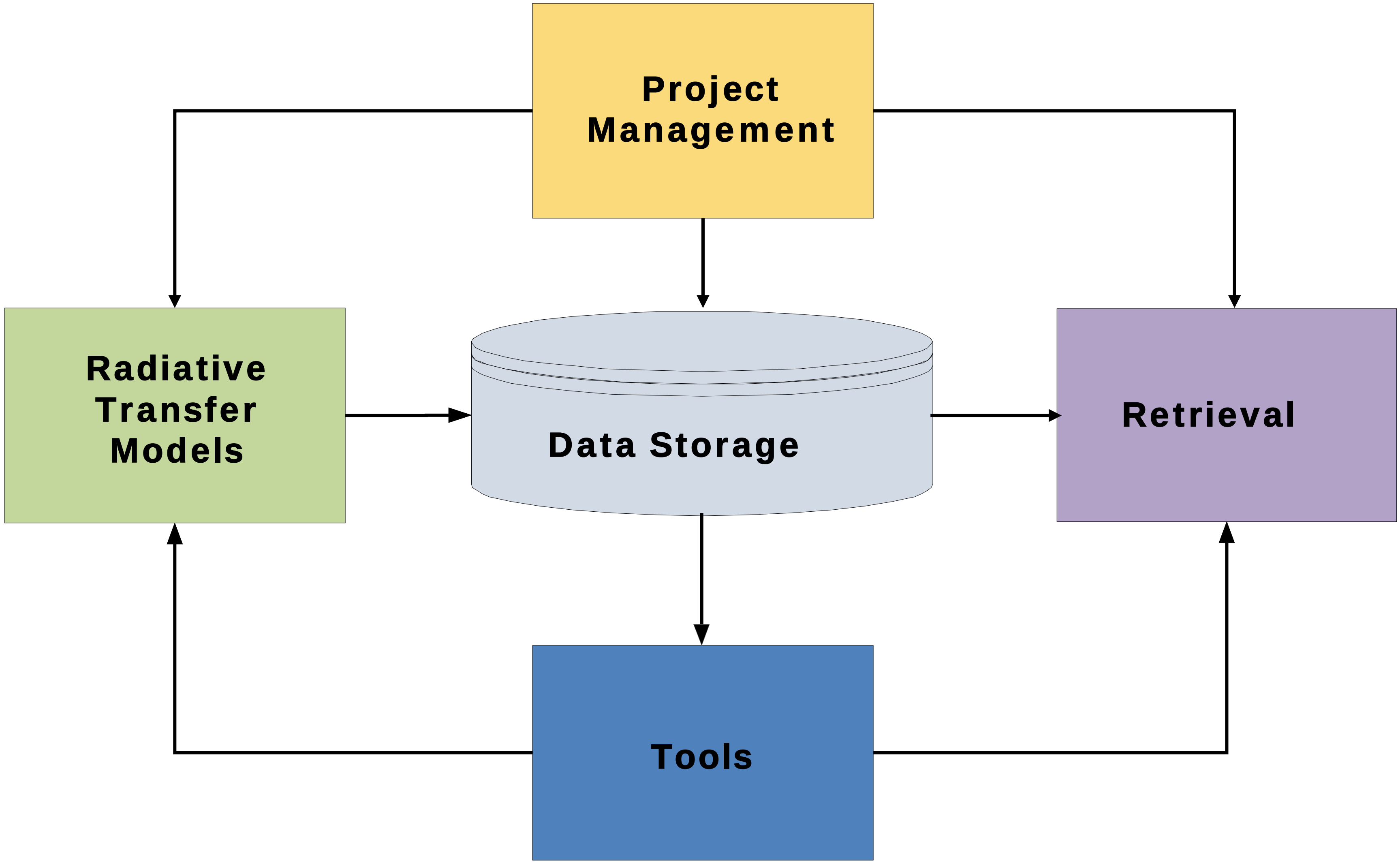
ARTMO v. 3: modular design



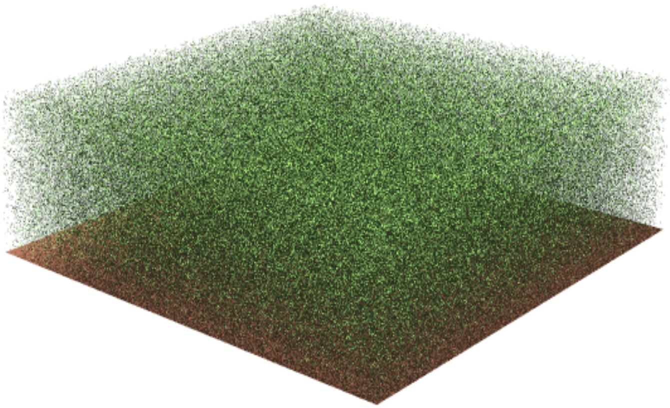


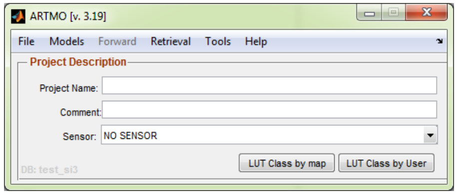


Conceptual architecture ARTMO



Forward




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
DB: test_si3

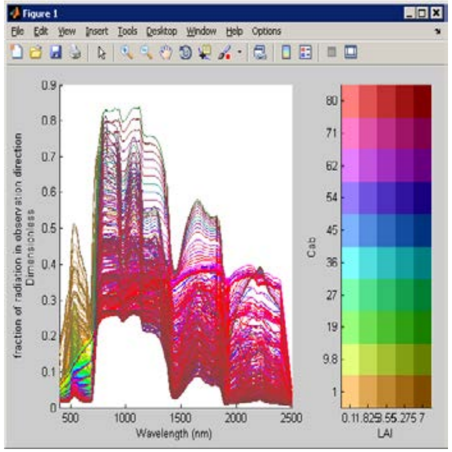
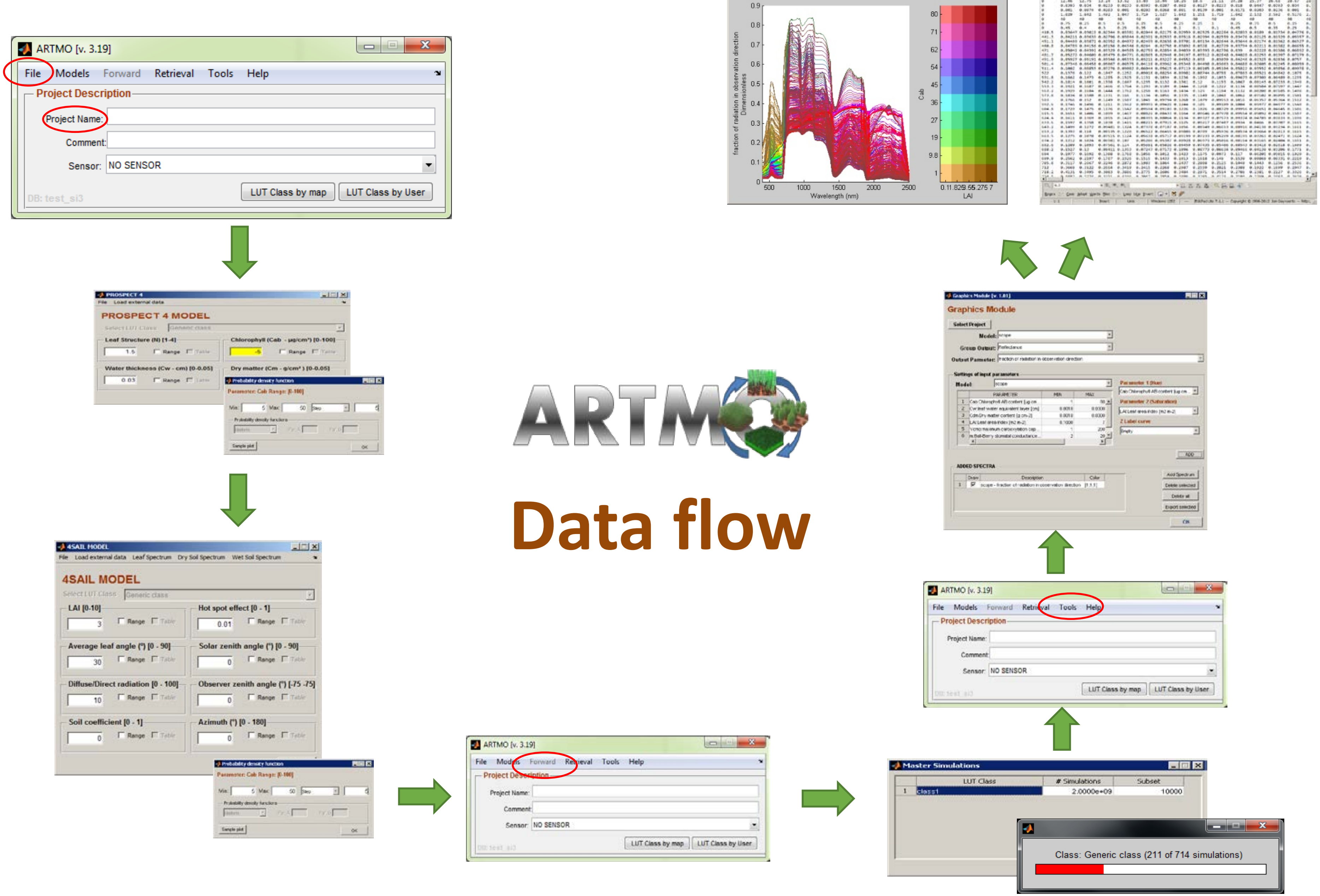


Figure 1
File Edit View Insert Tools Desktop Window Help Options
fraction of radiation in observation direction
Dimensionless
Wavelength (nm)
Cab
LAI

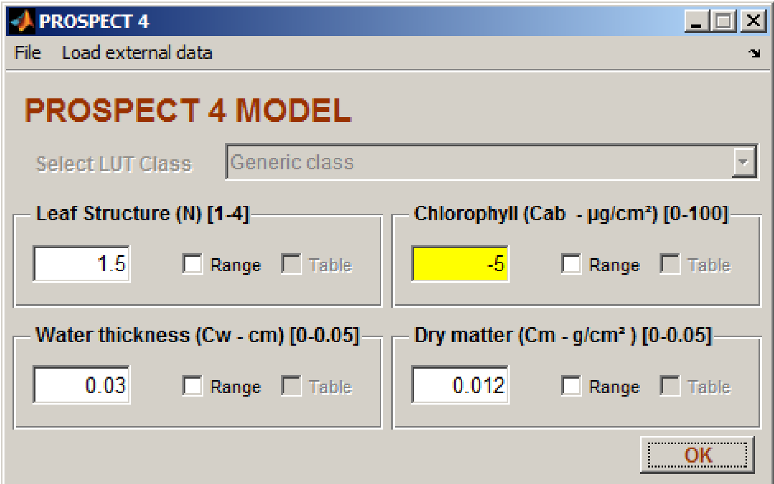
RTM outputs only a few clicks away…



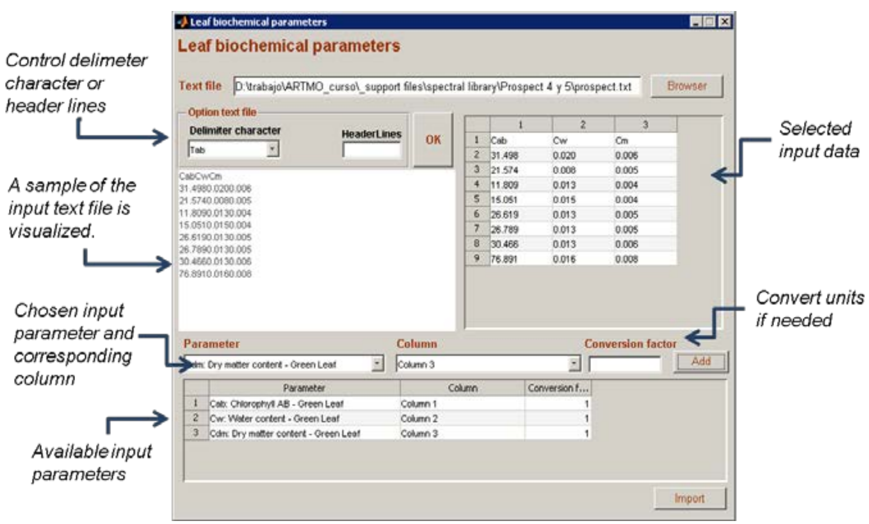
Radiative Transfer Models

Entering data: e.g. PROSPECT-4

1. A single value

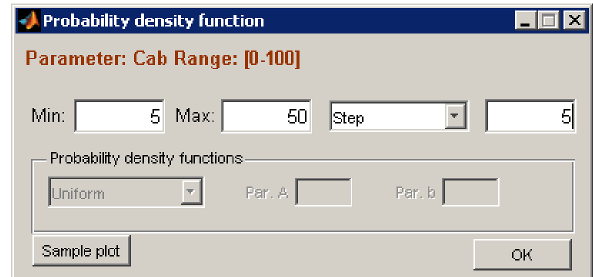


2. User data (e.g. field data)

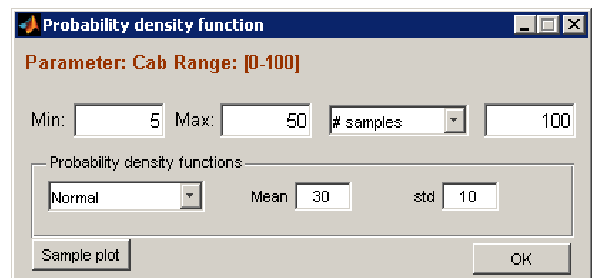


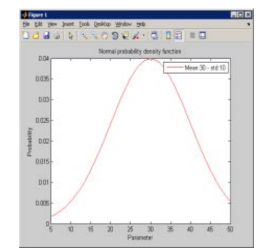
3. A range of multiple values:

I) steps



II) distribution: (e.g., uniform, normal, exponentional)



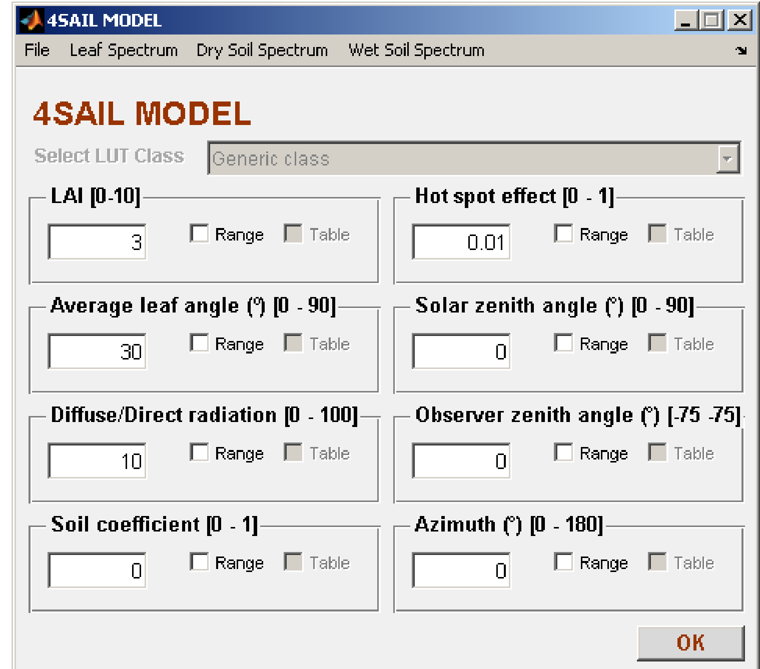


III) Multiple input values

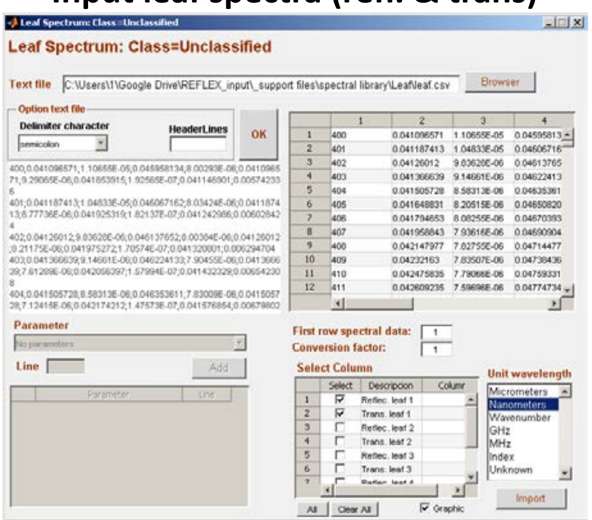
Radiative Transfer Models

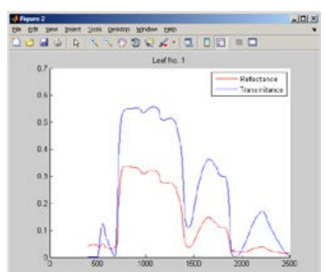
Filling in a canopy model: SAIL

- **Filling in SAIL: single or multiple values**
- **Soil spectra is required.** Default spectra are provided or own spectra can be imported.
- **Usually coupled with a leaf model.** When not coupled then leaf spectra is required.

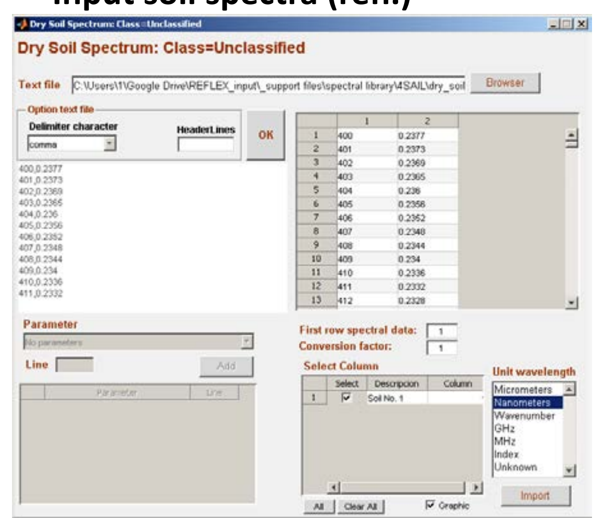


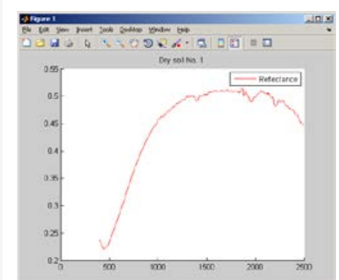
Input leaf spectra (refl. & trans)

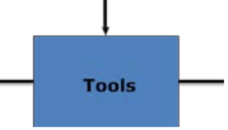




Input soil spectra (refl.)

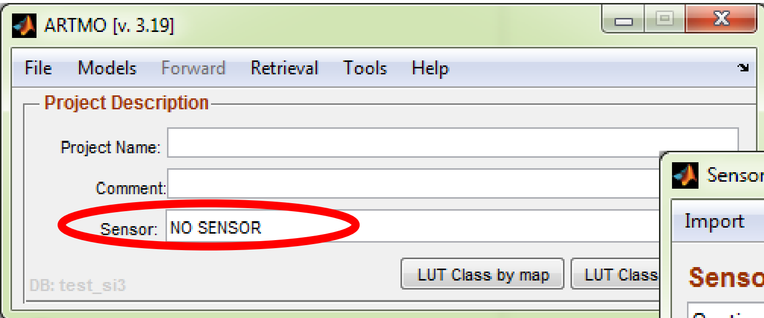


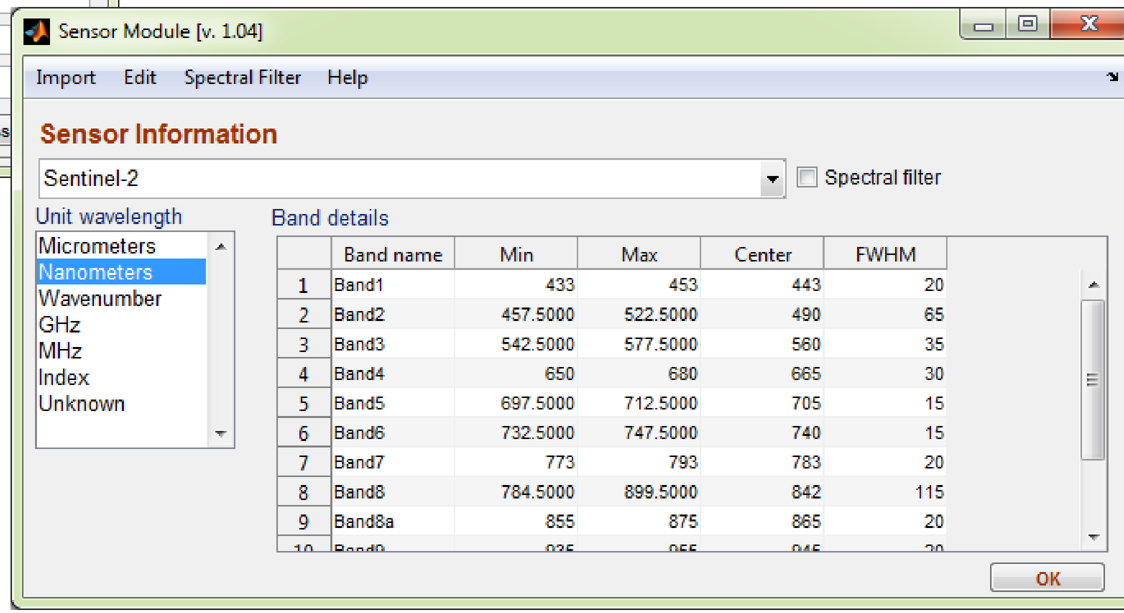




Sensor

Simulations can be generated according to band settings of a selected sensor.

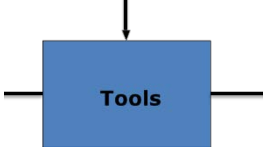




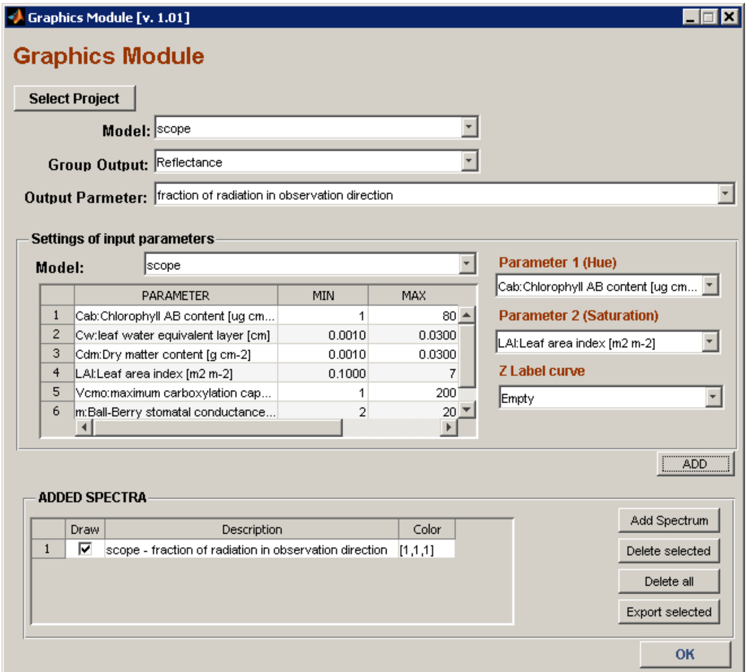
- New sensor settings can be imported by clicking on the '**Import**' button in the top bar.
- Existing band settings can be modified or new ones can be added by clicking on the '**Edit**' button.
- Also a spectral filter of a sensor can be imported or viewed by clicking on the '**Spectral Filter**' button.

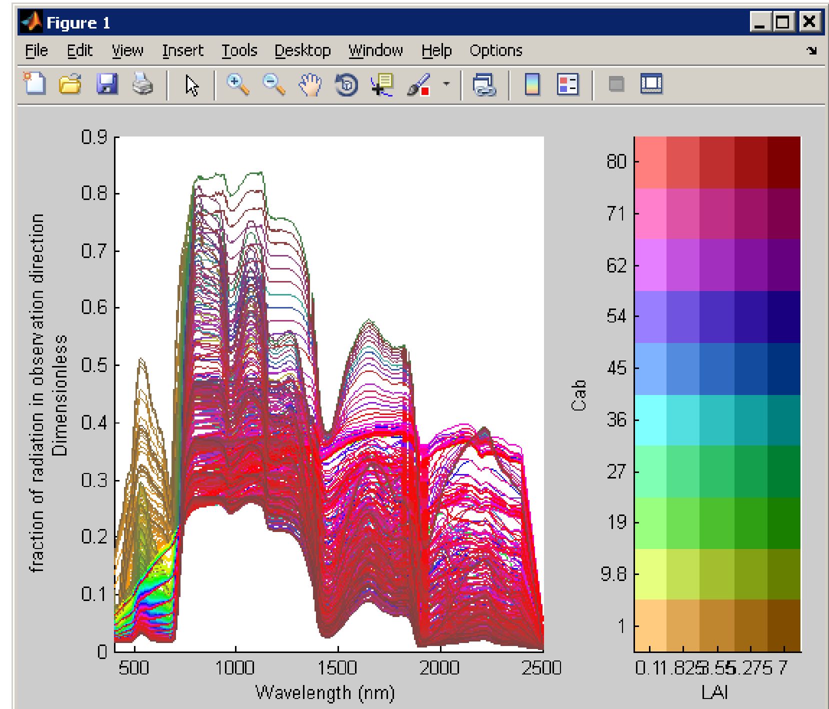
Default sensors:

- Landsat 7 TM
- Landsat 7 ETM+
- SPOT-4 VMI
- SPOT-4 HRVIR
- CHRIS Mode-3
- MODIS
- MERIS
- Sentinel-2
- Sentinel-3 OLCI
- Sentinel-3 SLSTR
- Landsat 8
- Pleiades-1A
- Quickbird

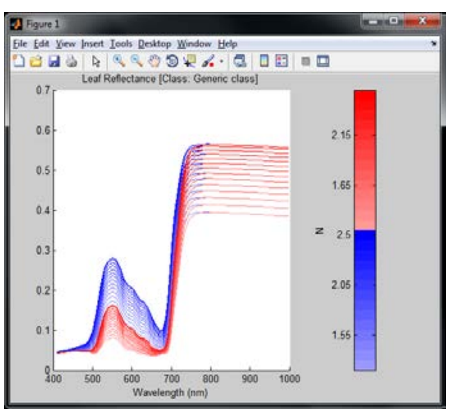


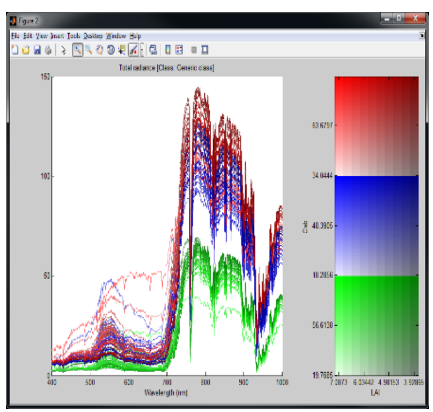
Graphics



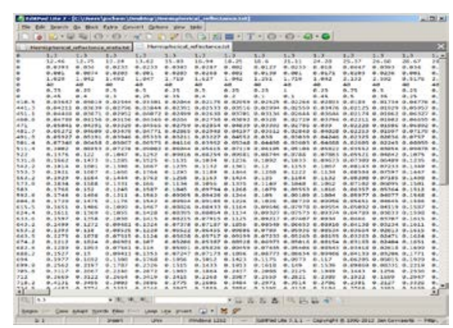


Visualization options

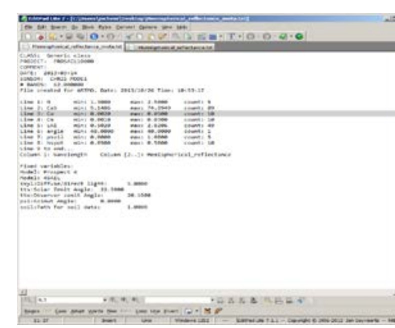




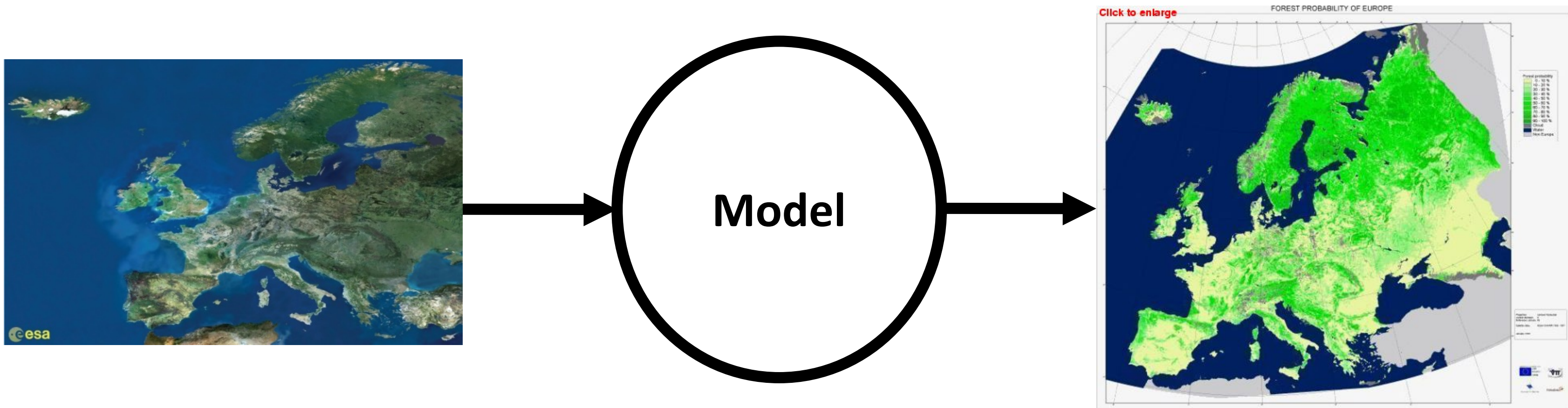
Spectral data



Associated metadata



Retrieval



Retrieval families

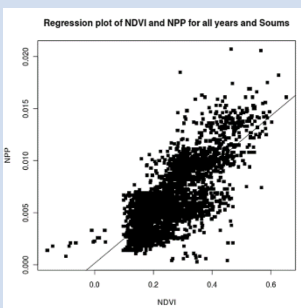
Parametric regression	Non-parametric regression	RTM inversion
Spectral relationships that are sensitive to specific vegetation properties	Advanced techniques that search for relationships between spectral data and biophysical variables	Models that simulate interactions between vegetation and radiation

Parametric regression

Spectral relationships that are sensitive to specific vegetation properties

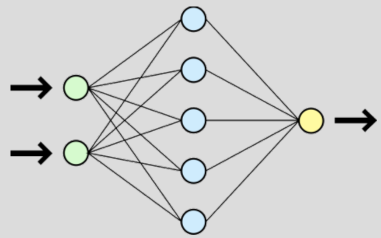
$$NDVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED})}$$

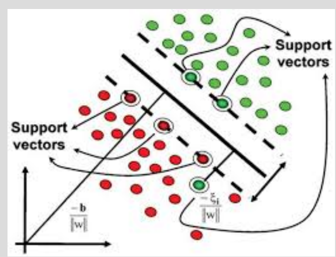
Normalized Difference Vegetation Index



Non-parametric regression

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

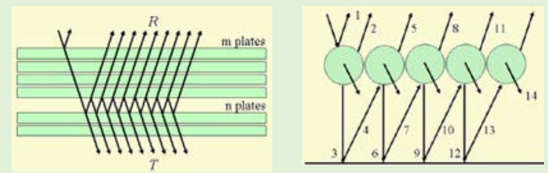




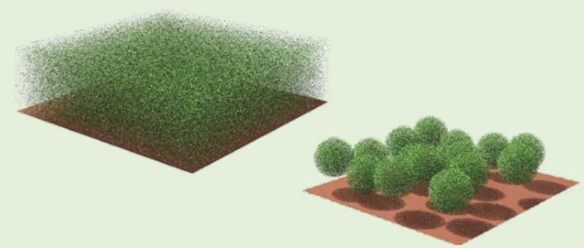
RTM inversion

Models that simulate interactions between vegetation and radiation

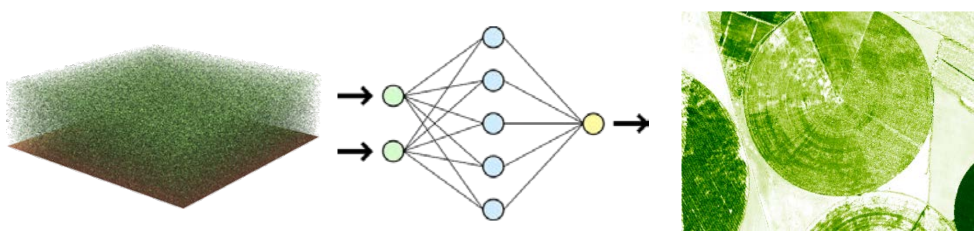
leaf



canopy

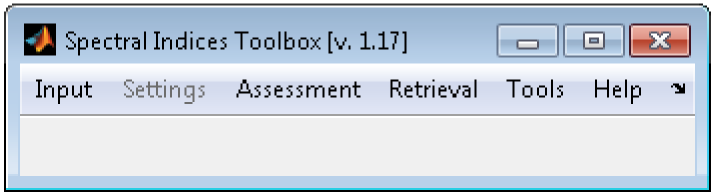


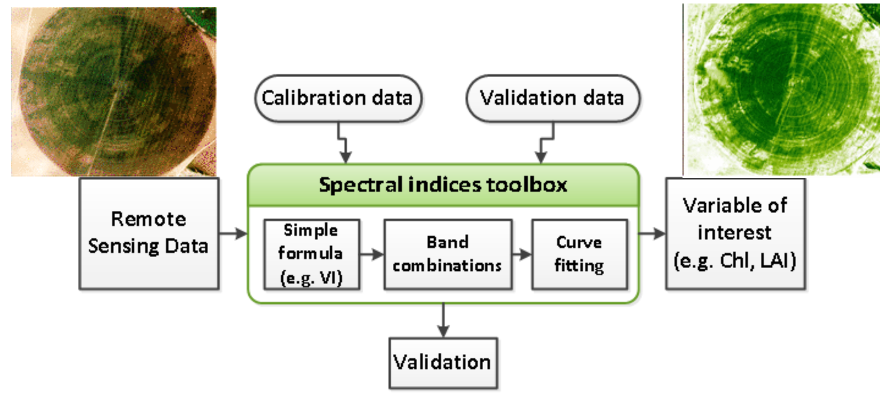
Methods of these different families can be combined: *hybrid methods*



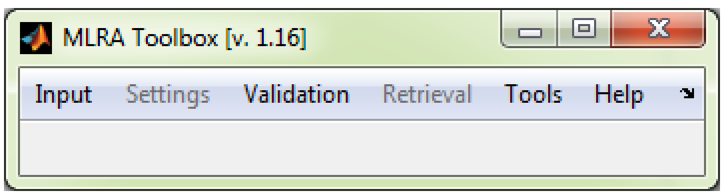
ARTMO's retrieval toolboxes:

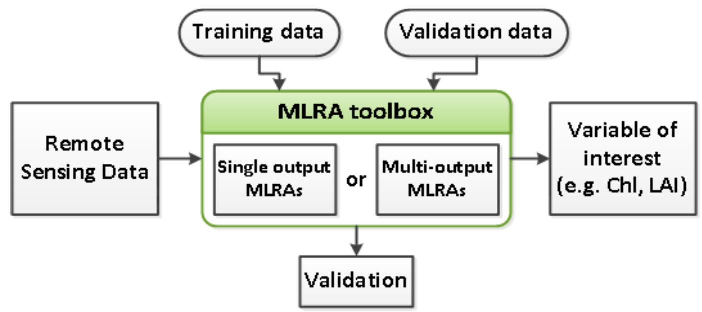
Spectral indices toolbox



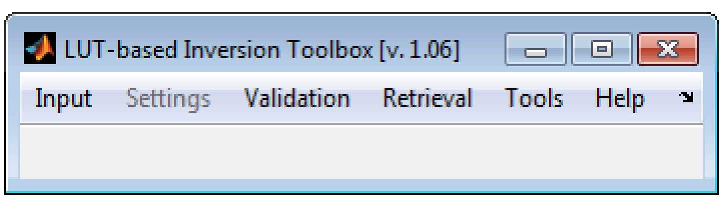


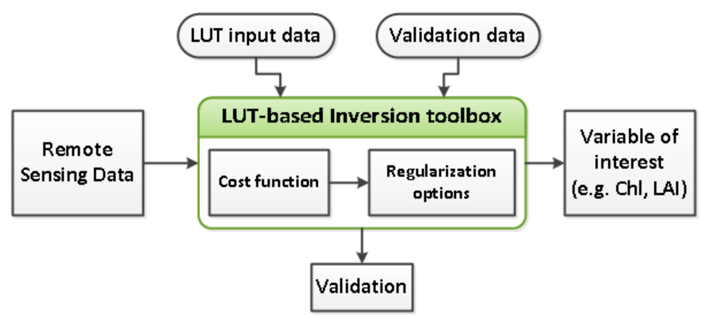
Machine learning regression algorithm toolbox





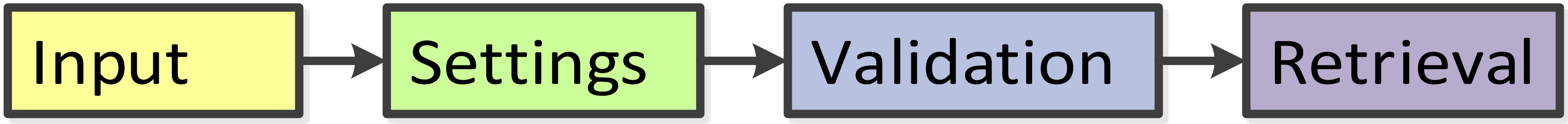
LUT-based inversion toolbox

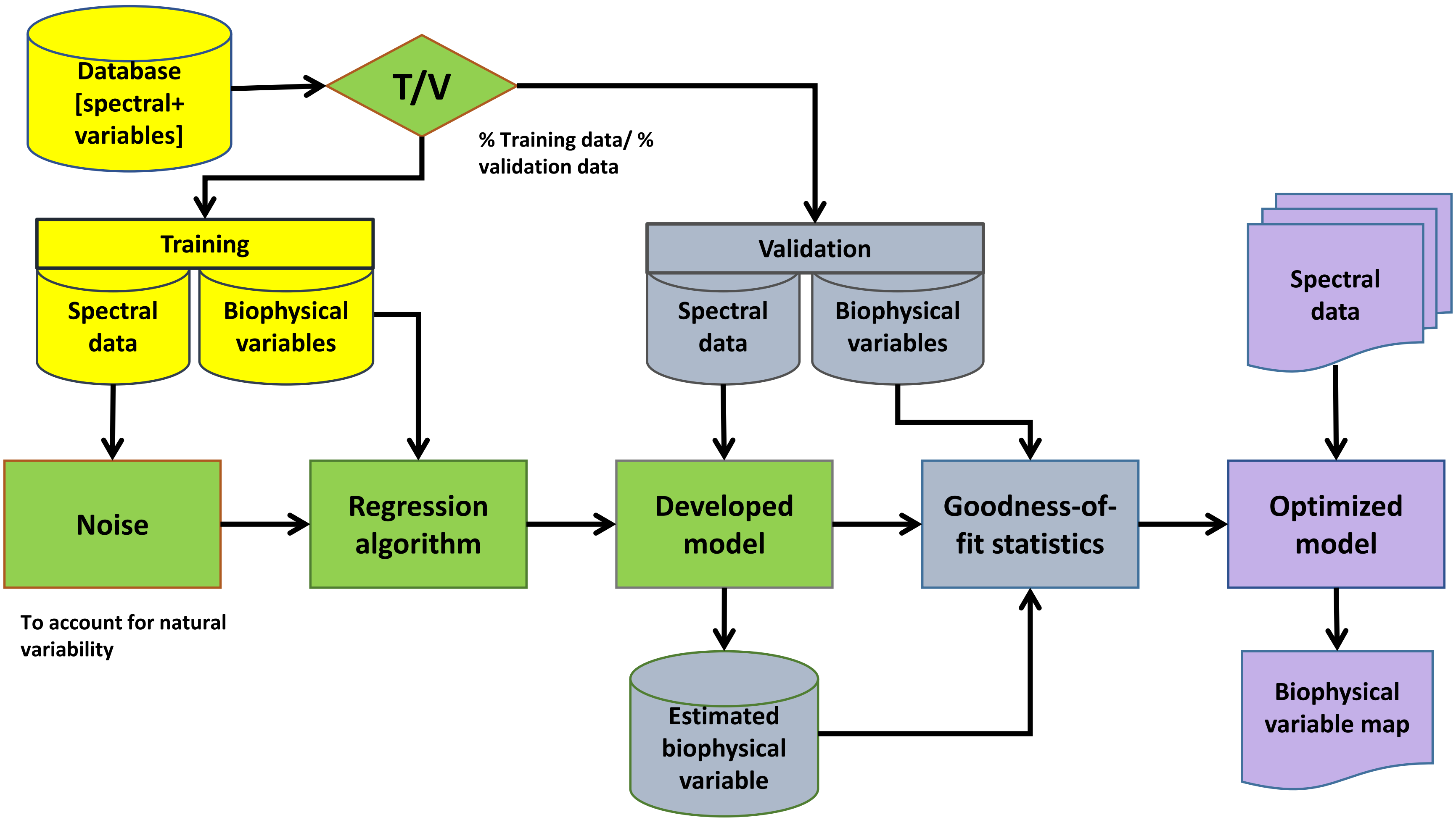




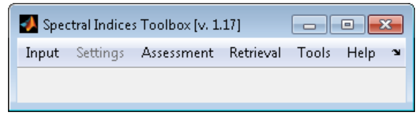
Optimizing and generating maps of vegetation properties only a few clicks away…

General structure:



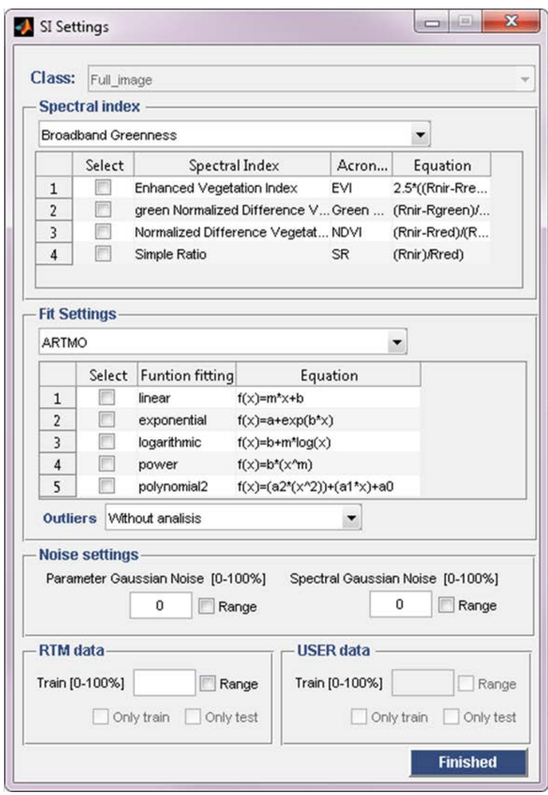


Spectral indices toolbox:



Properties:

- Calculates all possible band combinations.
- For index formulations with up to 10-band indices (#b[10], for a 10 band sensor that would be 10 billion combinations)
- Includes multiple fitting functions (linear, exponential, logarithmic, power, polynomial)
- Noise & Cross-validation options
- Results stored in MySQL
- Top-performing indices per formulation and fitting function are given.
- Can process both image or individual spectra.



If active, configure per land cover class.

Select an Index group

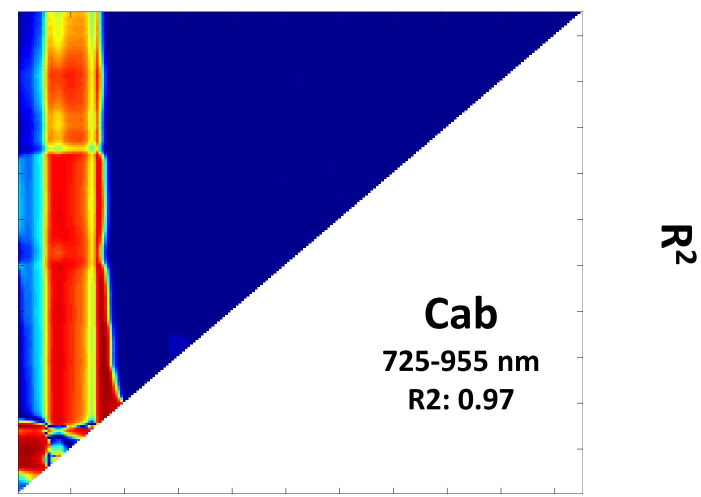
Select one or multiple indices

Select one or multiple curve fittings

Options to add noise

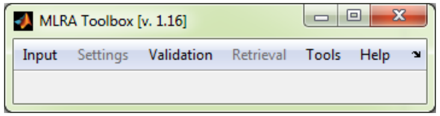
Option to mix RTM with field observations

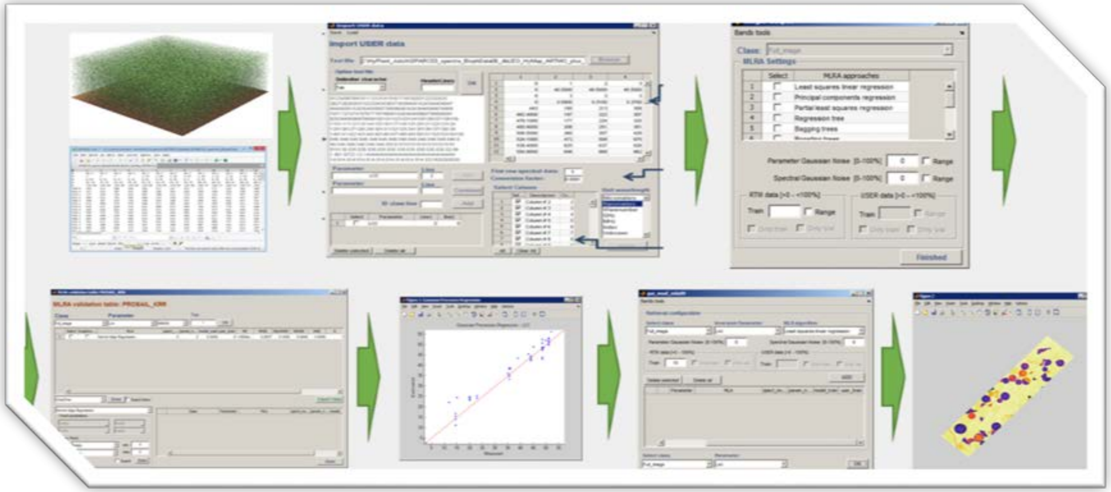
PROSAIL (100# @ 10 nm; Cab, LAI) ND linear regr.



Best-performing index can be applied to an image.

Machine learning regression algorithm toolbox



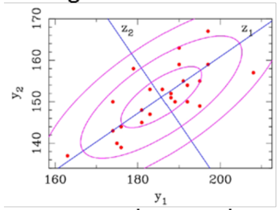


Simpler to execute than SI: no band selection needed.

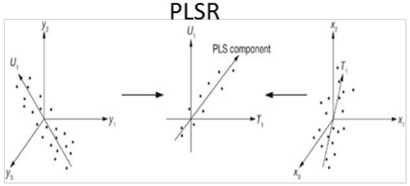
Properties:

- **About 15 MLRAs implemented**
- **Single-output & multi-output**
- **Noise & Cross-validation options**
- **Dimensionality reduction options**
- **Results stored in MySQL**
- **GPR properties: band relevance & uncertainties**
- **Can process both images or individual spectra.**
- **Active learning, GPR-BAT, dim. reduction**

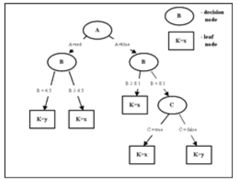
Principal component regression – PCR



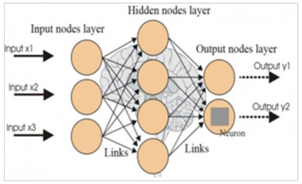
Partial least squares regression PLSR



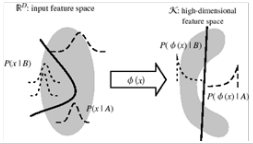
Decision Trees – DT



Neural networks NN

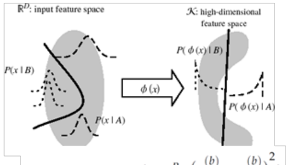


Kernel ridge regression KRR



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

Gaussian processes regression GPR



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

Non-parametric models:

- **SimpleR** [Camps-Valls et al., 2013]
- http://www.uv.es/gcamps/code/simpleR.html

Also:

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

GPR in Bayesian framework also provides:

- Band relevance
- Uncertainty esitmates

(kernel-based) MLRAs are adaptive and can be very powerful. However that goes a computational cost. This can be problematic for hybrid (e.g. PROSAIL) retrieval methods.

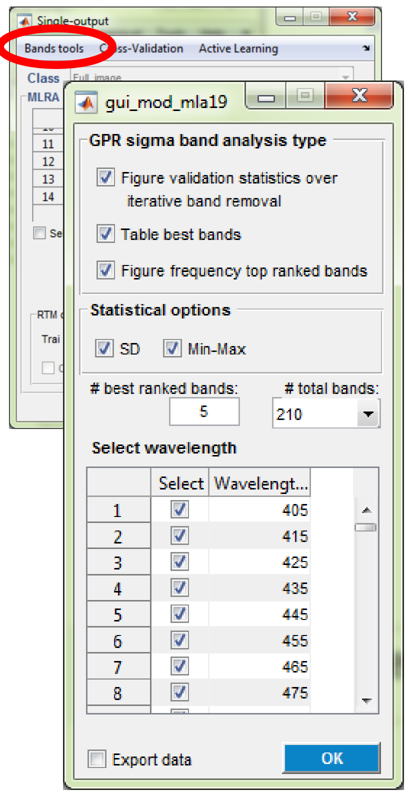
Solutions to deal with large datasets applicable to *hybrid* approaches (e.g., PROSAIL + GPR):

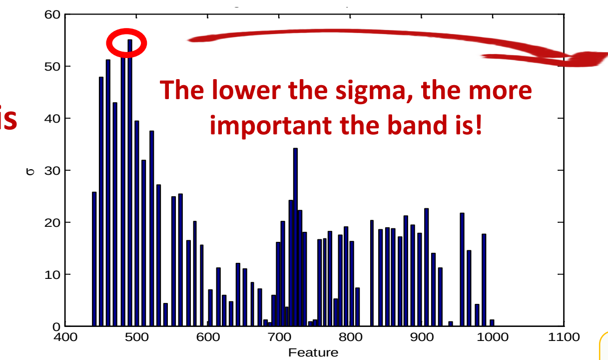
1. Reducing spectral data: I) band selection (GPR-BAT), II) dimensionality reduction
2. Samples reducing : Active learning

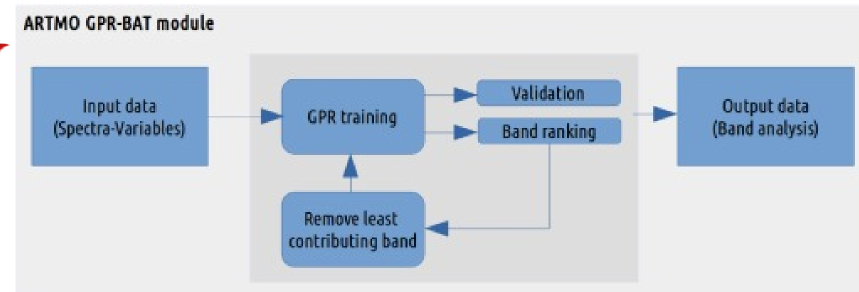
I) Band selection: GPR-BAT

Sequential Backward Band Removal: remove band with highest sigma (least informative)

Gaussian processes regression – Band analysis Tool (GPR-BAT).

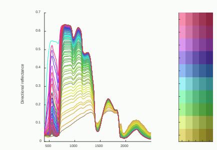




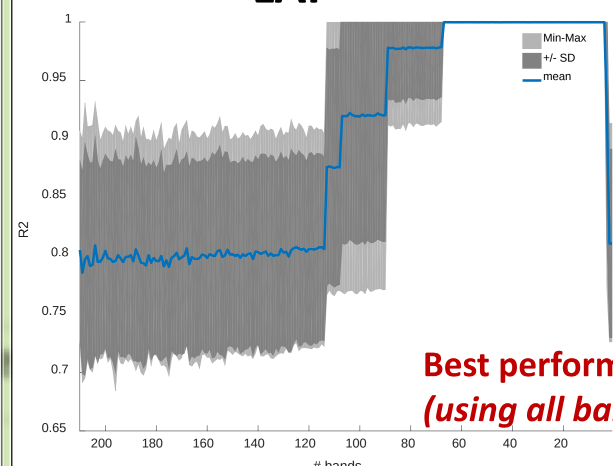


Experimental setup:

- PROSAIL: LHS 100# @ 10 nm; Cab, LAI
- 4k cross-var sampling



LAI



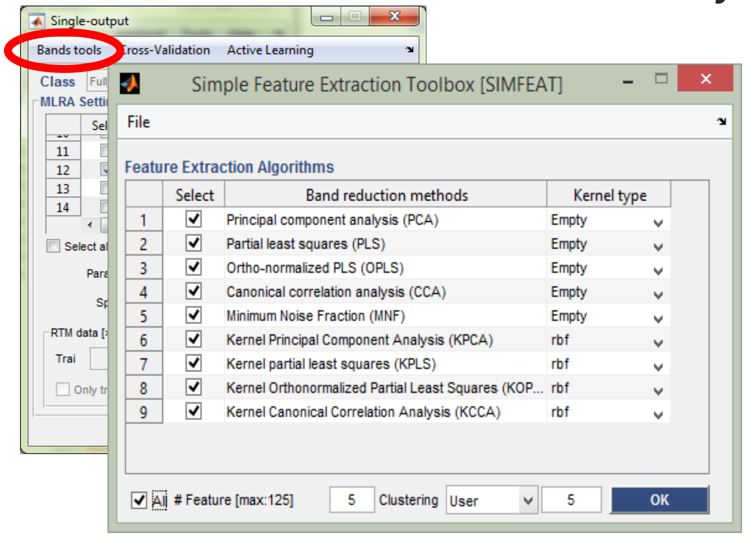
# band	R2	wavelengths
5	**0.9997**	815, 1145, 1205, 122, 1245
4	**0.9997**	815, 1145, 1205, 1245
3	0.9213	815, 1145, 1205
2	0.8104	815, 1145
1	0.8104	815

Best performances achieved between 70 and 4 bands
(using all bands or <3 bands not recommended)

Best-performing method can be applied to an image.

II) Dimensionality reduction: SIMFEAT

9 dimensionality reduction methods implemented.

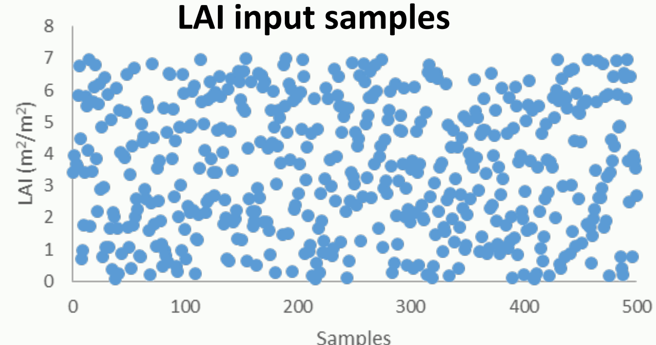


Experimental setup:

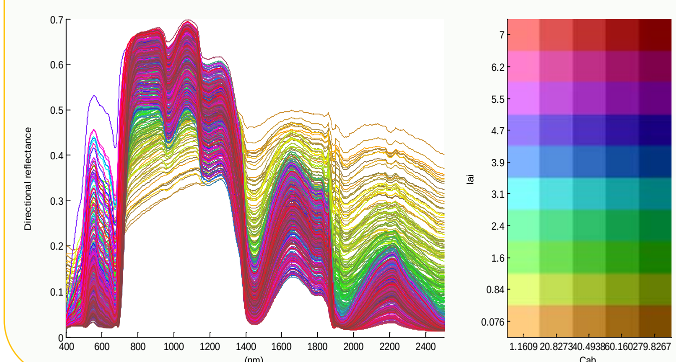
PROSAIL: 500 random samples

Variable	Min	Max
N	1	4
Cab	1	80
Cw	0.02	0.05
psoil	0	1
LAI	0.01	7

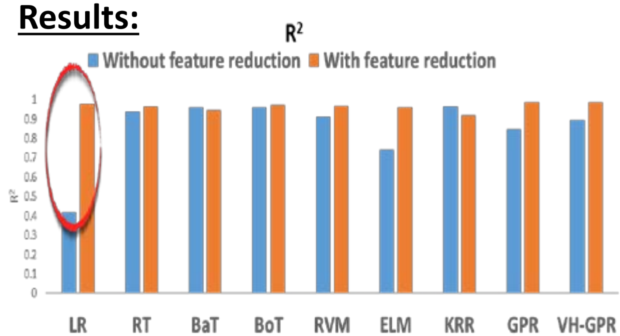
LAI input samples



directional reflectance (2101 bands)

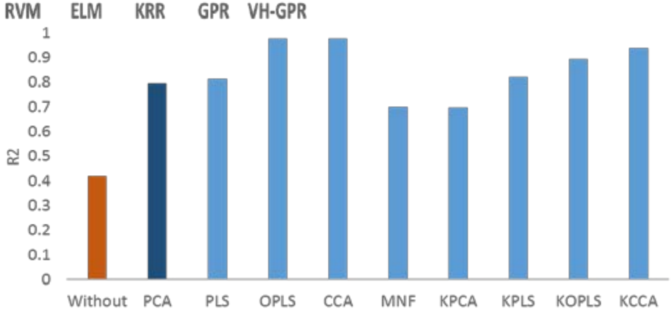


Results:



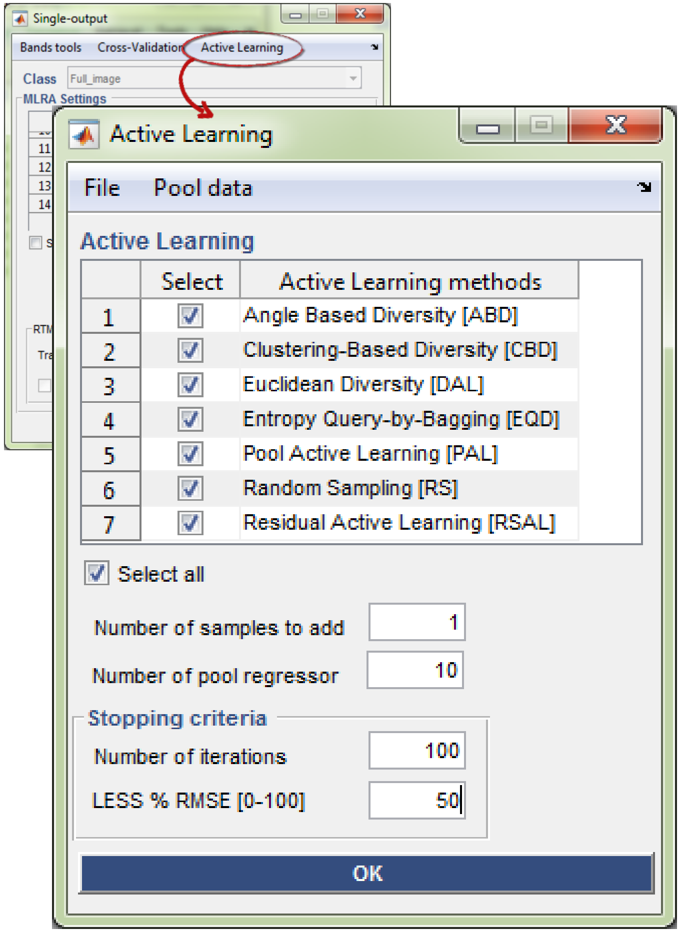
- **Almost all methods benefited from dim. Reduction methods.**
- **Most impact on LR**
- **PCA not best performing**

Linear regression (LR)

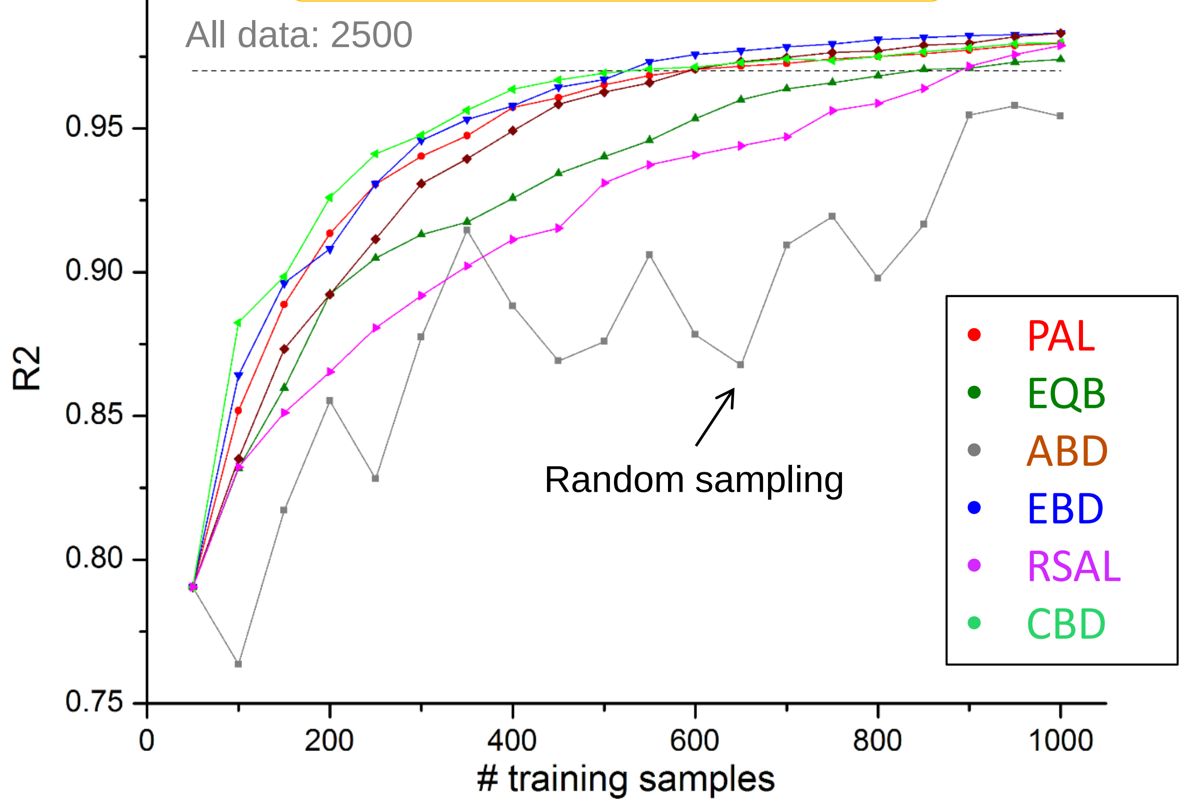


Best-performing method can be applied to an image.

2) Sample reduction: Active learning (AL)



Experimental setup:
PROSAIL: 5000 samples
- **2500 training; 2500 validation**
- **Cab using KRR**

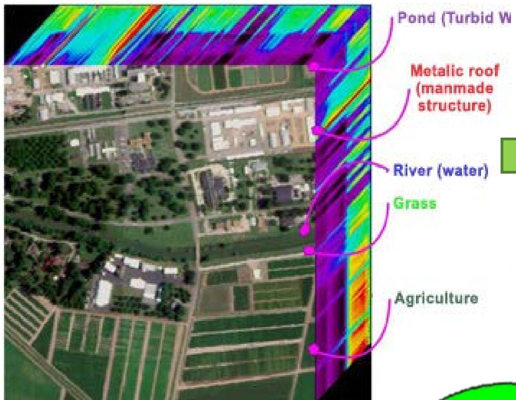


- Active learning (AL) searches for new samples from a data pool based on ***uncertainty*** (PAL, EQB, RSAL) and ***diversity*** (ABD, CBD, EBD).
- AL method search more efficiently for relevant samples than random sampling or when using all data.

Best-performing method can be applied to an image.

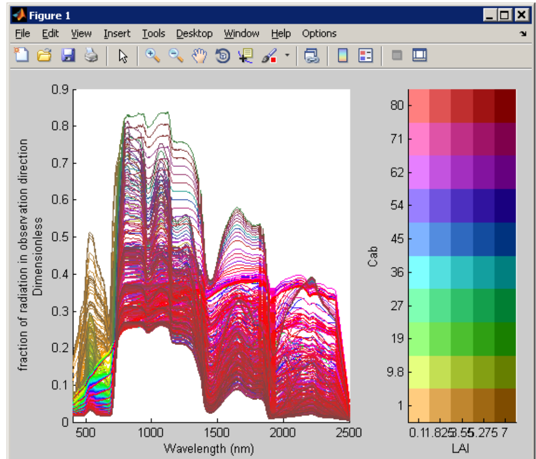
Background LUT-based inversion

RS imagery



Spectral block with 64 bands from 415 nm to 900 nm

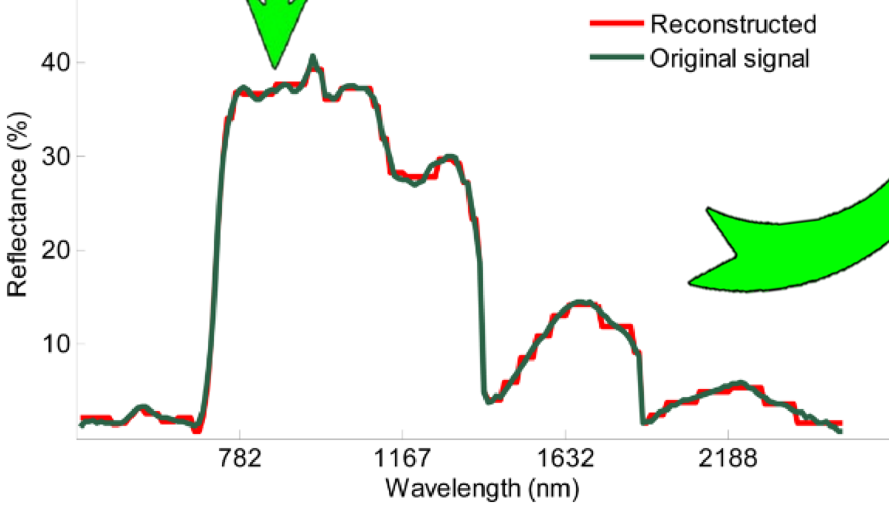
PROSAIL LUT



Best match is obtained through a 'Cost function', or 'Minimum distance function'.

CostFunction	Definition	Minimum	Maximum
C^{LS}	$\sum(Y-X)^2$	0	∞
C^{NC}	$\frac{\sum(X.Y)}{\sqrt{\sum X^2}\sqrt{\sum Y^2}}$	-1	1
C^{W}	$\sum_k \frac{n_k}{N} \frac{\sqrt{\mathrm{Var}(Y_k)}}{\mu(Y_k)}$	0	∞
C^{CR}	$\frac{1}{\mathrm{Var}(Y)} \sum_k \frac{n_k}{N} \mathrm{Var}(Y_k)$	0	1

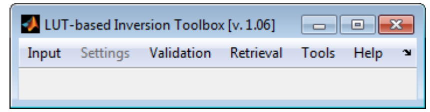
Search per pixel for best match against LUT

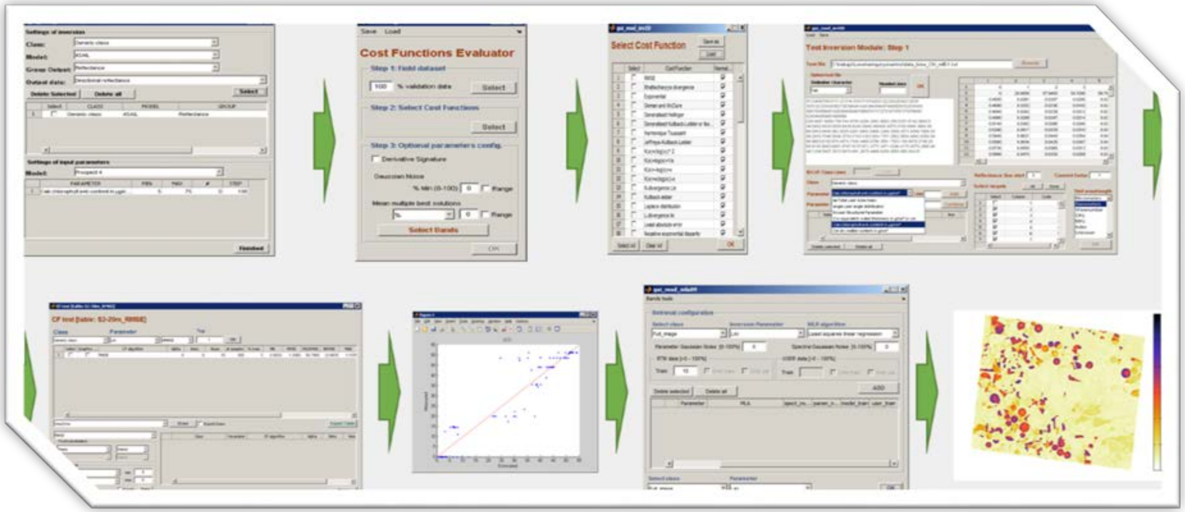


Other important factors:

- Adding **noise** (to account for natural variability)
- Selecting mean/median of **multiple solutions**

LUT-based inversion toolbox:

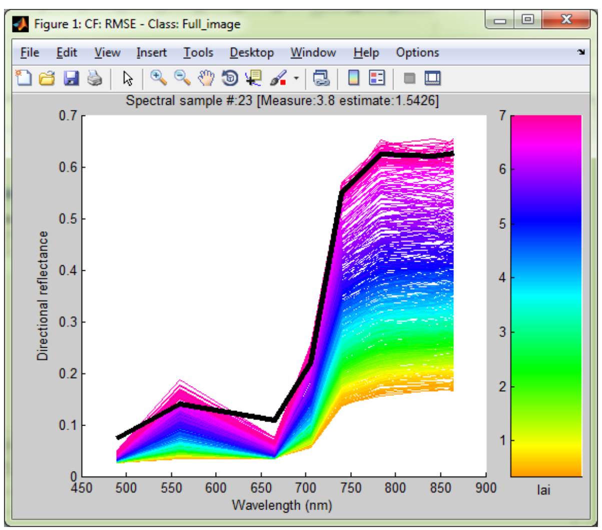


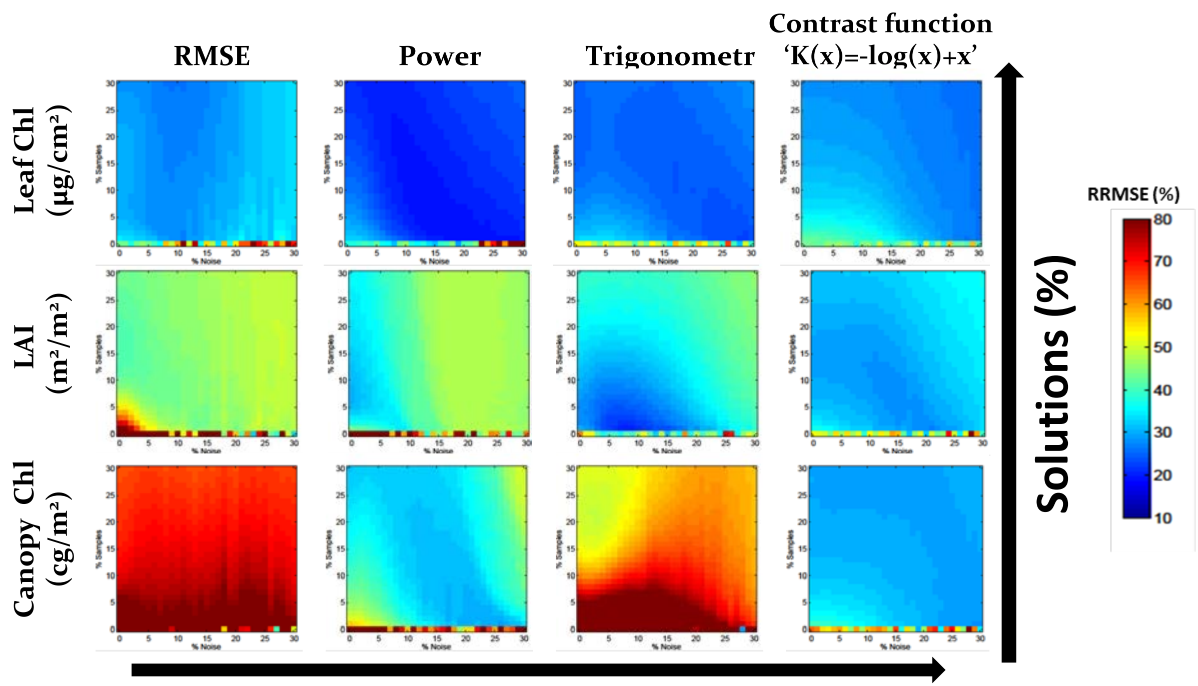


Properties:

- **LUT ARTMO RTMs or external LUT**
- **Over 60 different cost functions**
- **Noise & multiple solutions**
- **Results stored in MySQL**
- **Top-performing inversion strategies are given.**
- **Can apply inversion to both image or individual spectra.**

Matching a pixel against a part of the LUT.





Best-performing method can be applied to an image.

Tools

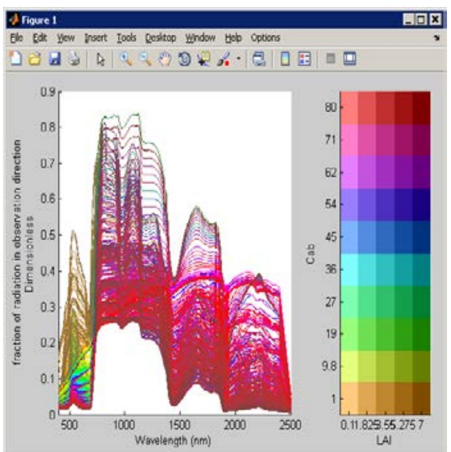
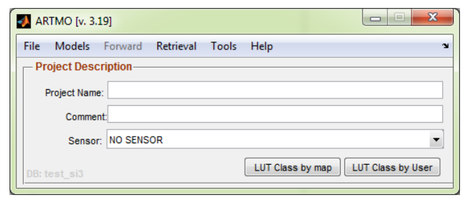
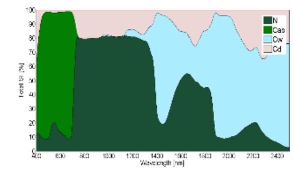
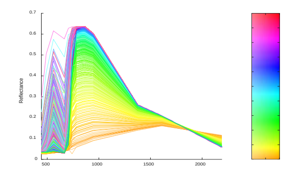


Figure 1
File Edit View Insert Tools Desktop Window Help Options
fraction of radiation in observation direction
Dimensionless
Wavelength (nm)
Cab
LAI


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





Global sensitivity analysis

<u>Global sensitivity analysis</u>: explores the full input parameter space, i.e. all input parameters are changed together.

Variance-based methods: the output variance is decomposed to the sum of **contributions of each individual input parameter and the interactions** (coupling terms) between different parameters.

Based on the work of **Sobol'**, **variance-based sensitivity measures** are represented as follows:

$$1 = \sum_i S_i + \sum_i \sum_{j>i} S_{ij} +, \cdots, + S_{12,\ldots,k},$$

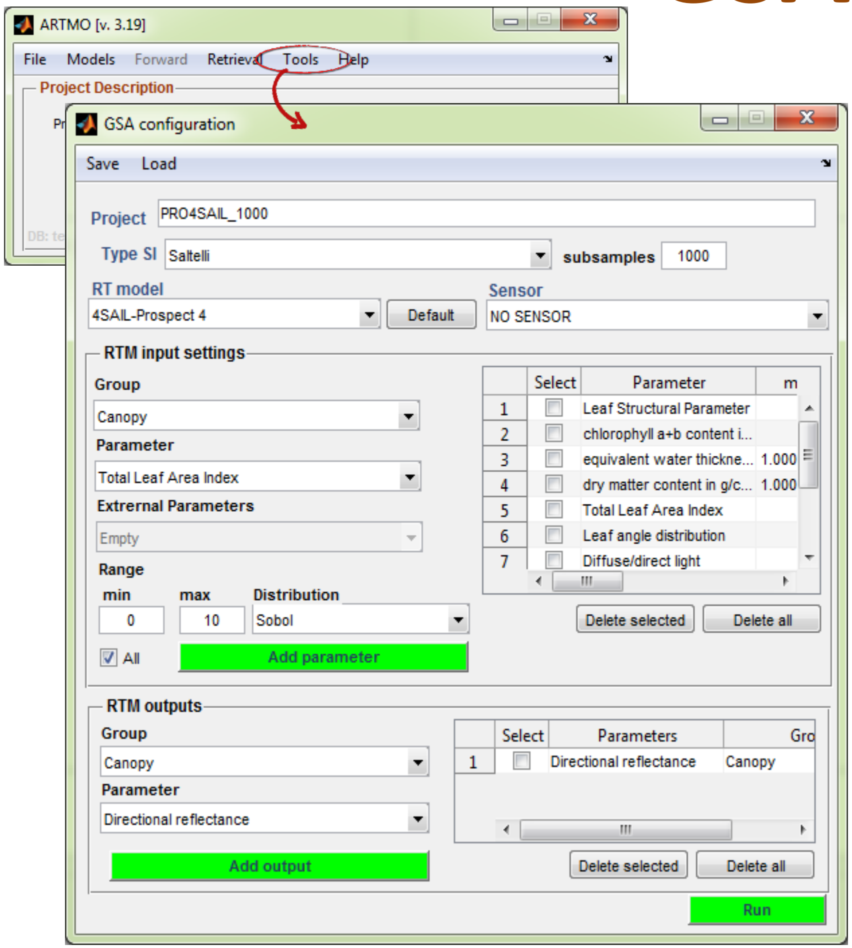
in this equation, S_i, S_{ij},…,$S_{12,\ldots,k}$ are ***Sobol's global sensitivity indices***.:

The ***<u>first order sensitivity index</u>*** ***S_i*** measures and quantifies the sensitivity of model **output *Y* to the input parameter *X_i*** (without interaction terms), whereas, S_{ij},…,$S_{12,\ldots,k}$ are the sensitivity measures for the higher order terms (interaction terms).

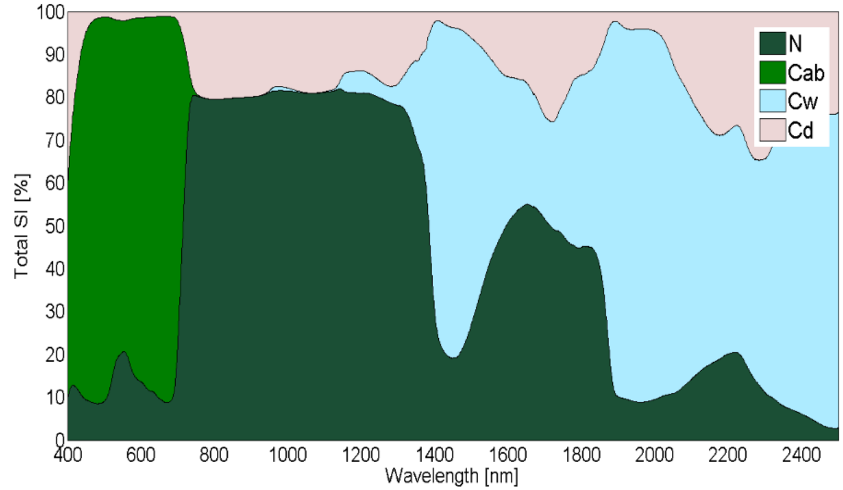
The ***<u>total effect</u>*** **<u>sensitivity index</u>** ***S_{Ti}*** measures **the whole effect of the variable *X_i***, i.e. the first order effect as well **as its coupling terms with the other input variables**:

$$S_{T1} = S_1 + S_{12} + S_{13} + S_{123}$$

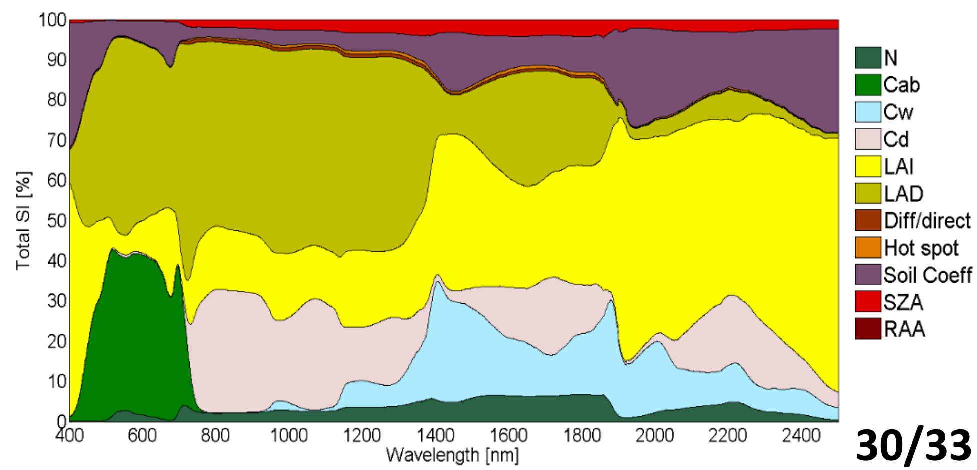
GSA toolbox



PROSPECT-4 Reflectance < 3 min



PROSAIL Directional Reflectance < 5 min



Properties:

- **ARTMO RTMs**
- ***Saltelli 2010* GSA method**
- **Various sample distributions**
- **Results stored in MySQL**
- **First order or total order Sobol Sensitivity indices**
- **Can process multiple RTM outputs.**

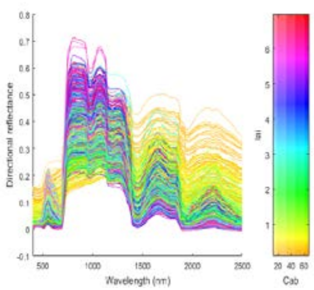
Emulation

Emulators are regression models that are able to approximate the processing of an RTM, at a fraction of the computational cost:

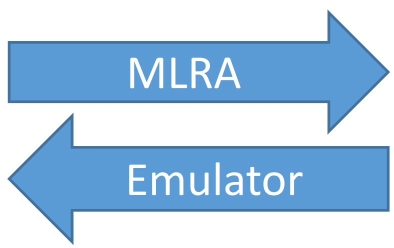
making a statistical model of a physical model

Emulators applied to RTMs:

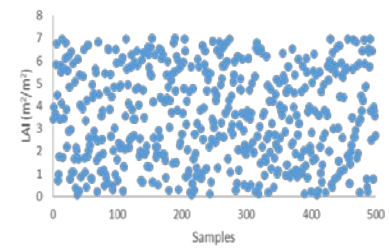
- In principle any nonlinear, adaptive **machine learning regression algorithms (MLRAs)** can serve as emulators.



Spectra



Variables
(e.g. LAI, chlorophyll)



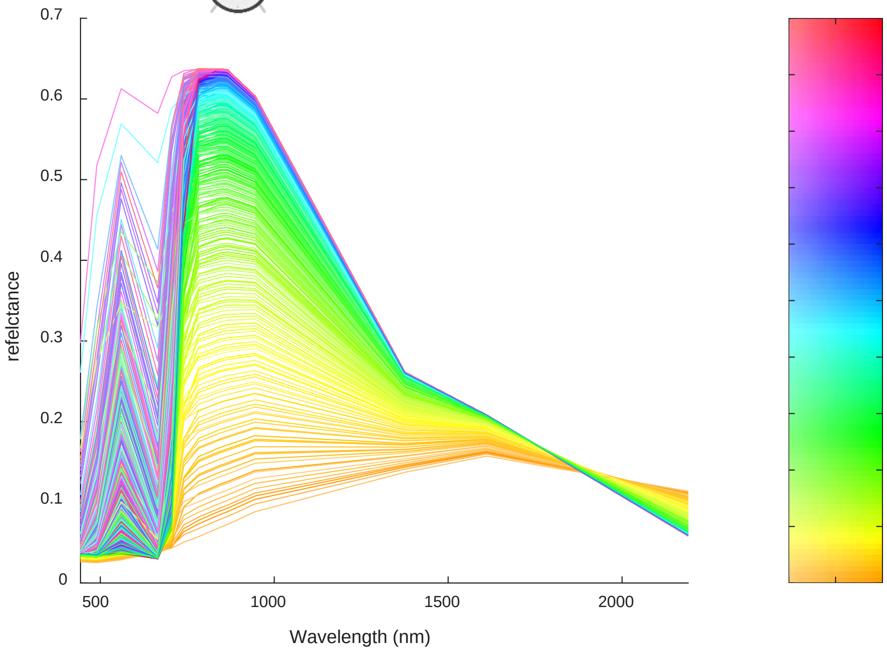
- To emulate RTMs, the emulator should have the capability to reconstruct multiple outputs, i.e. the complete spectrum: resolved with **dimensionality reduction** techniques (e.g. PCA).

Processing steps:

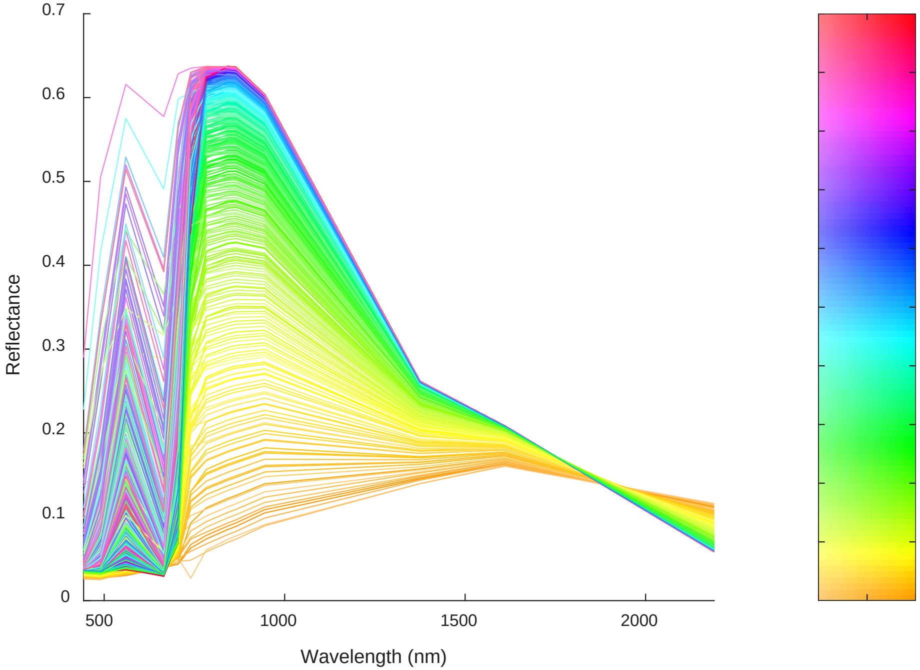
Input (variables + spectra) → Splitting into training/validation → PCA on spectra → MLRA training looping over components → Prediction of components → Reconstruction of full spectrum → Validation → Emulator

500 TOC reflectance simulations according to Sentinel-3 (13 bands)

30 sec



3 sec

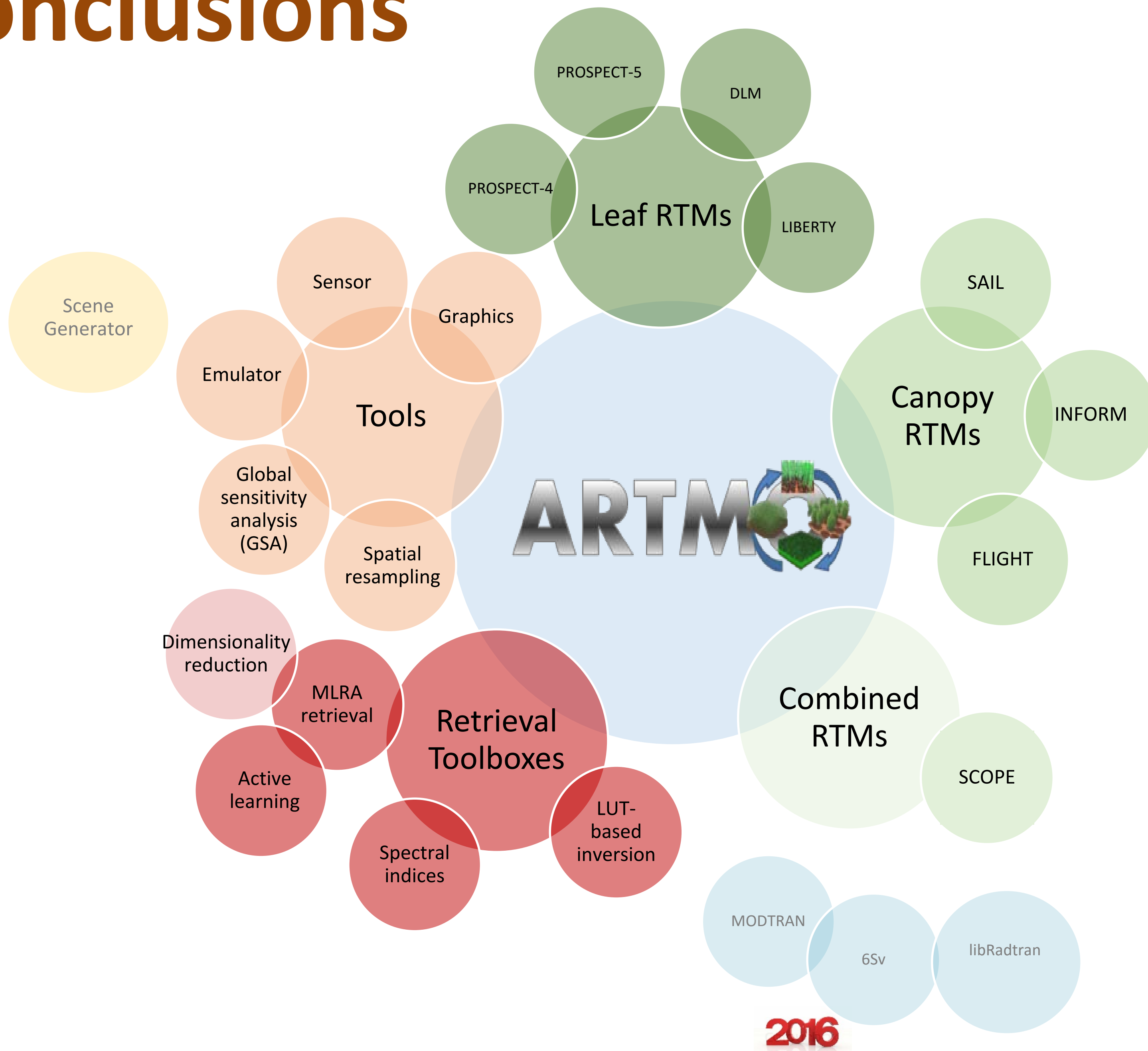


PROSAIL

Emulator

In Emulation, physical models go hand in hand with machine learning

Conclusions



Thanks

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