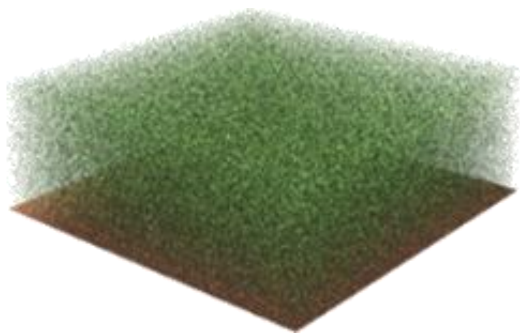
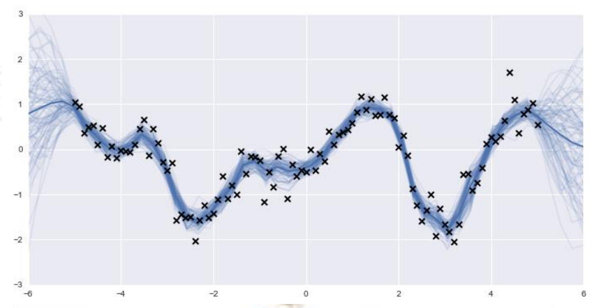
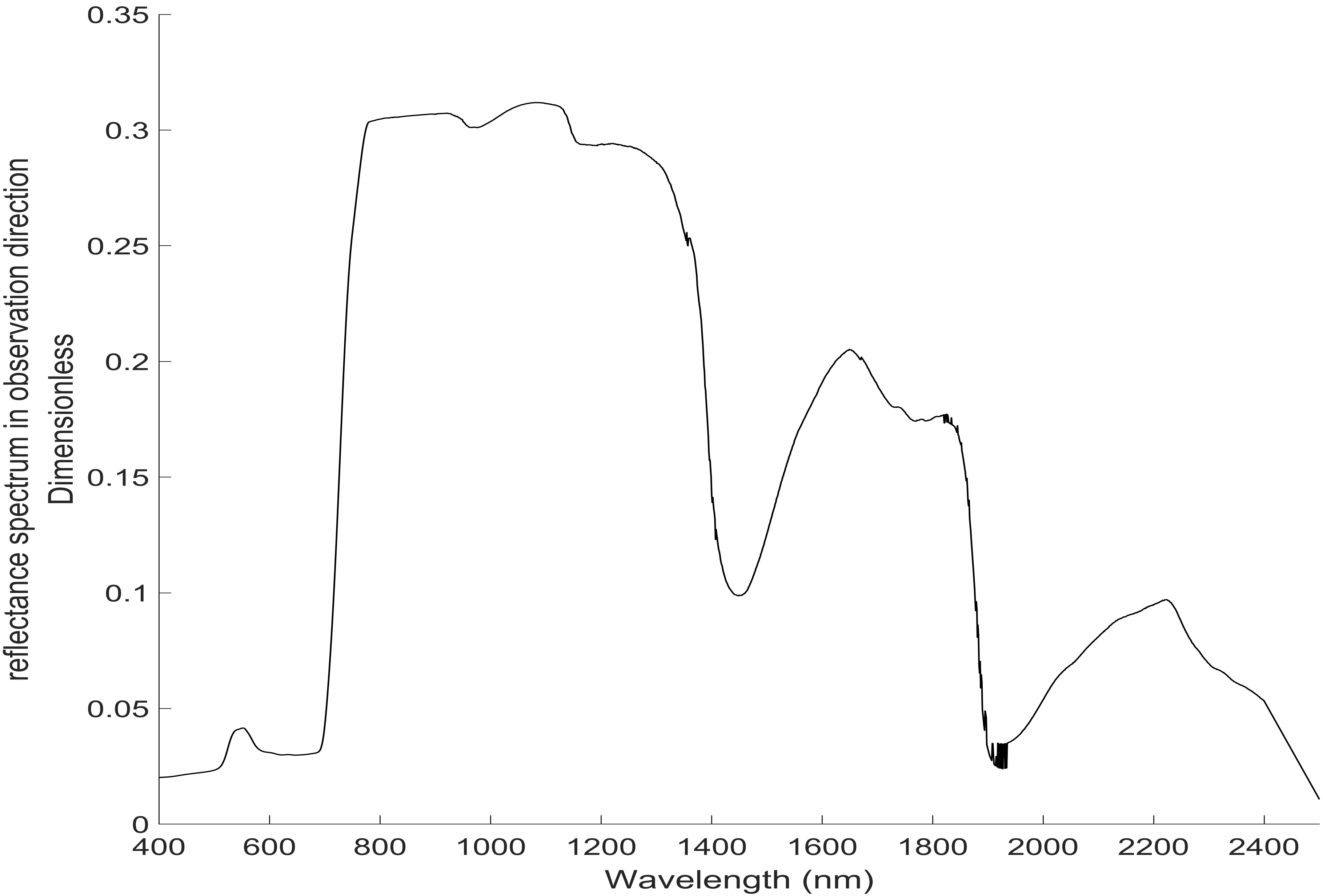
Progress in hybrid models for applications in remote sensing of vegetation





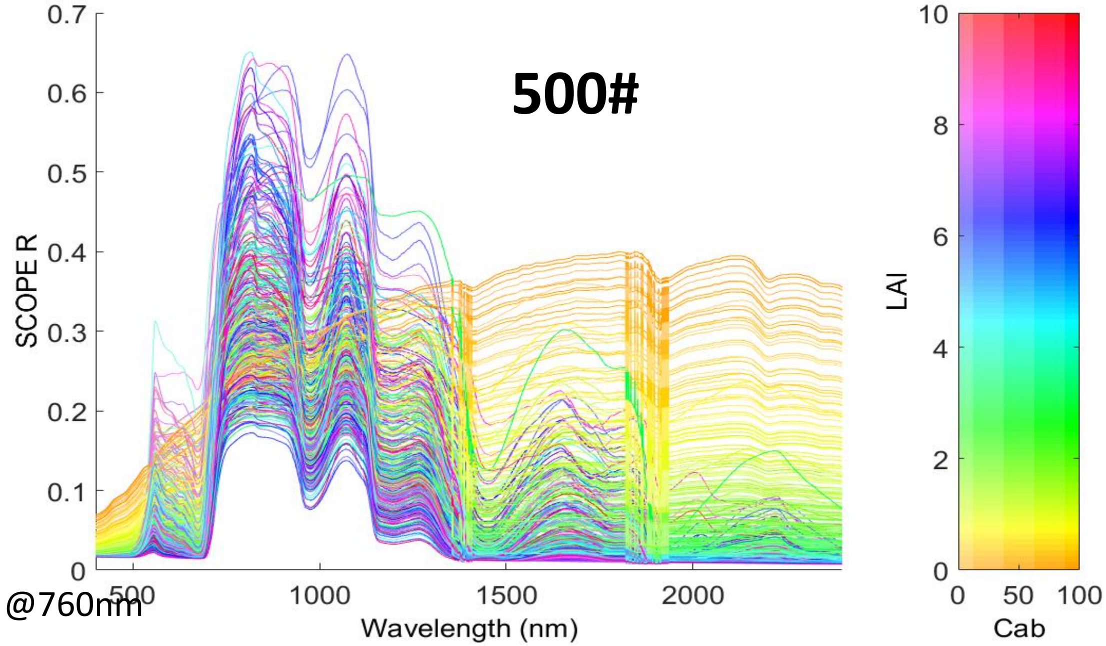
Jochem Verrelst, Juan Pablo Rivera, Miguel Morata, Jose Estévez, Matias Salinero, Pablo Reyes, Enrique Portales, Ana Belen Pasqual, Jorge Vicent, Santiago Belda, Bastian Siegmann, Katja Berger, and colleagues

07 June 2021, Warsaw, Poland


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
reflectance spectrum in observation direction
Dimensionless
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
Wavelength (nm)

Any difference? Which model would you choose?

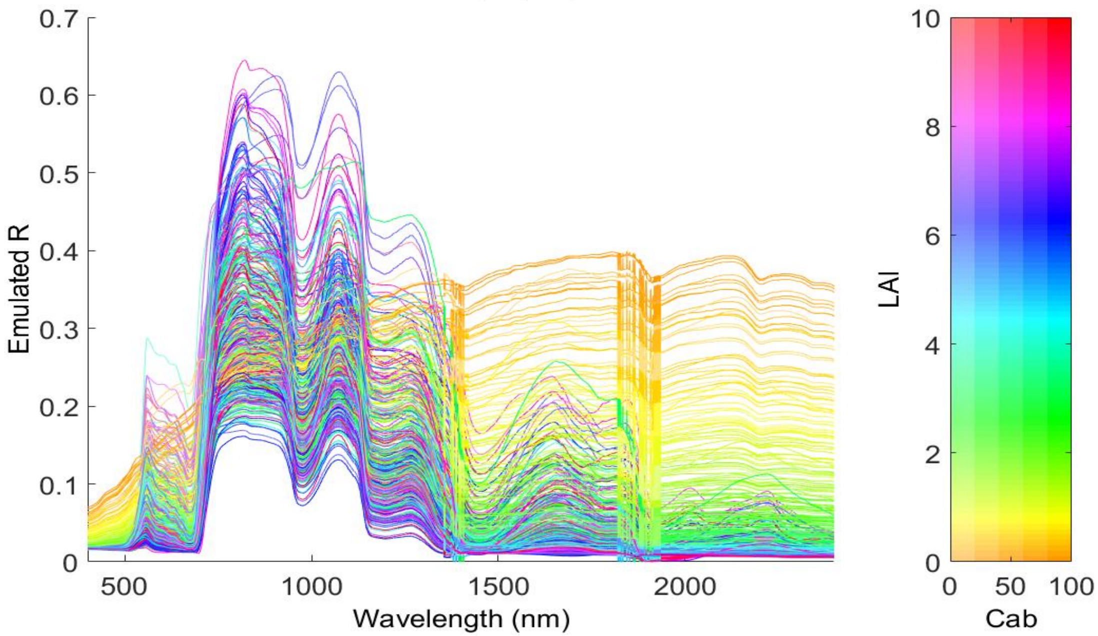
37 min



SCOPE

(RTM)

0.2 s



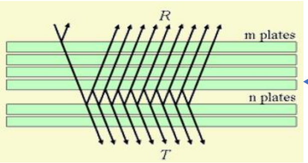
Emulator

(emulated SCOPE)

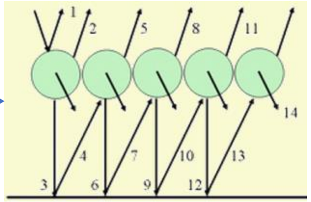
Radiative transfer models (RTMs)

Leaf RT models

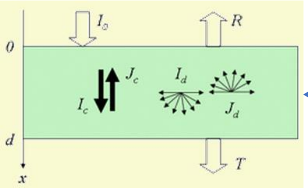
layers



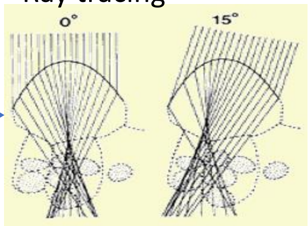
Compact spheres



N-fluxes

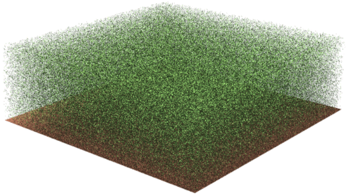


Ray tracing

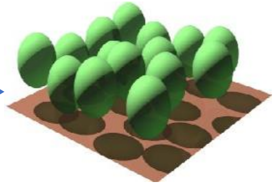


Canopy RT models

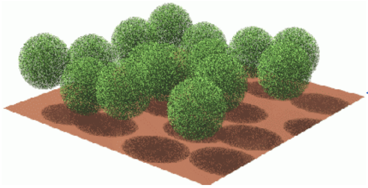
Turbid medium



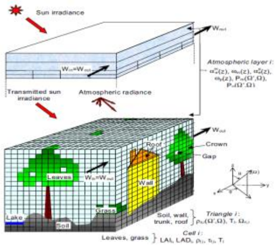
Geometric



Hybrid



Volumetric

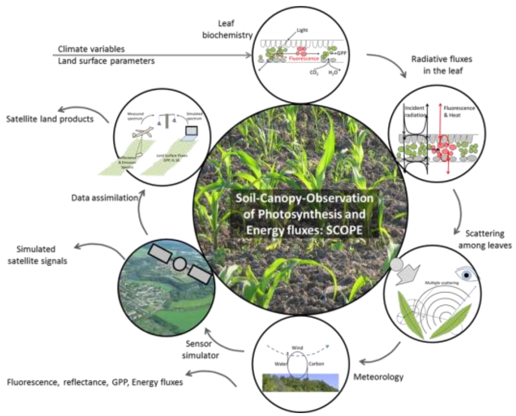


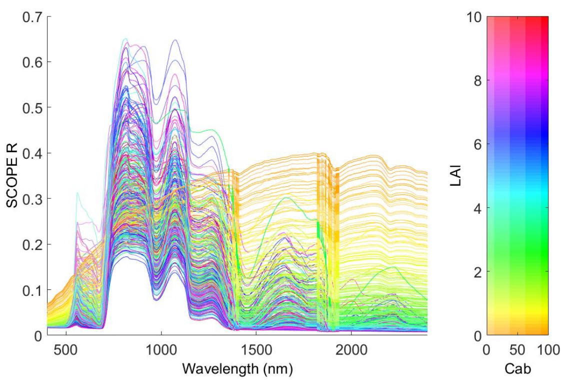
Multiple models exist with diverse complexity.

BACKGROUND

Advanced RTMs: generation of a large LUT (>1000#)

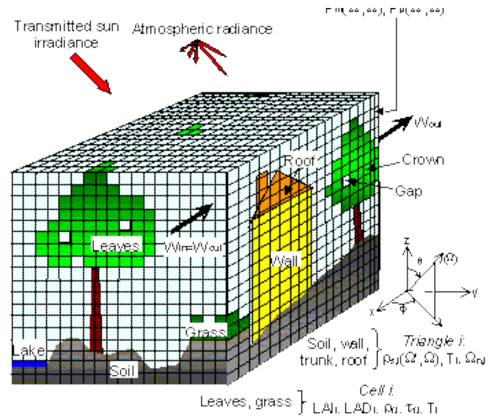
SCOPE

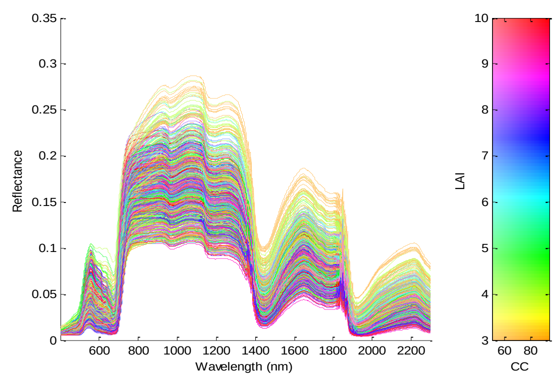


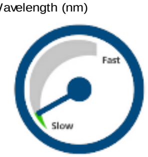


Hours

DART

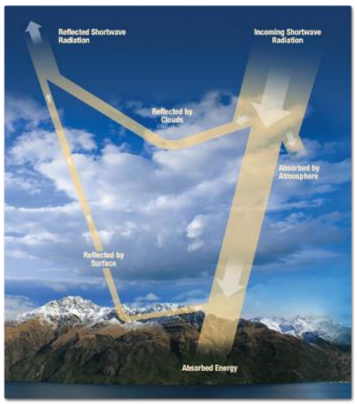


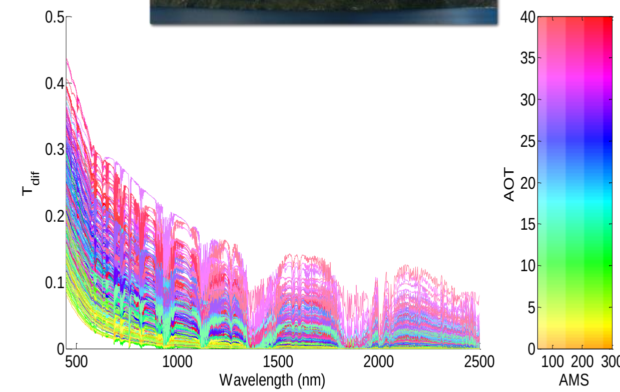




days

MODTRAN





>days

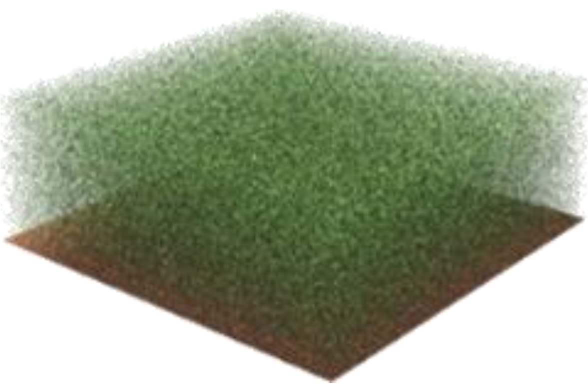
Advanced RTMs: *more realistic but slow*

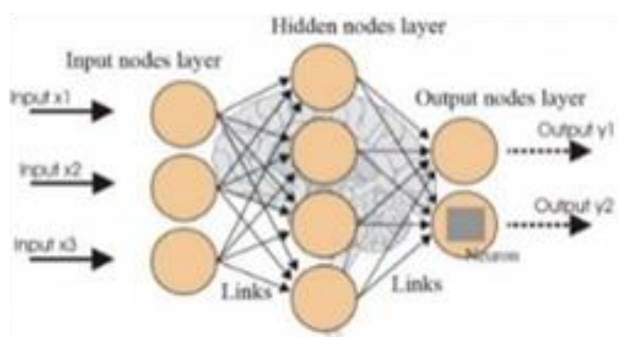
Emulation of RTMs

<u>Emulators</u> are statistical models that approximate the processing (input-output) of a physical model (e.g. RTM) - at a fraction of the computational cost:

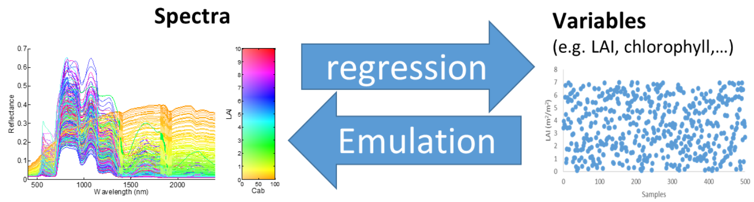
creating a statistical model from a physical model

RTM | **Machine learning** | **Emulator**

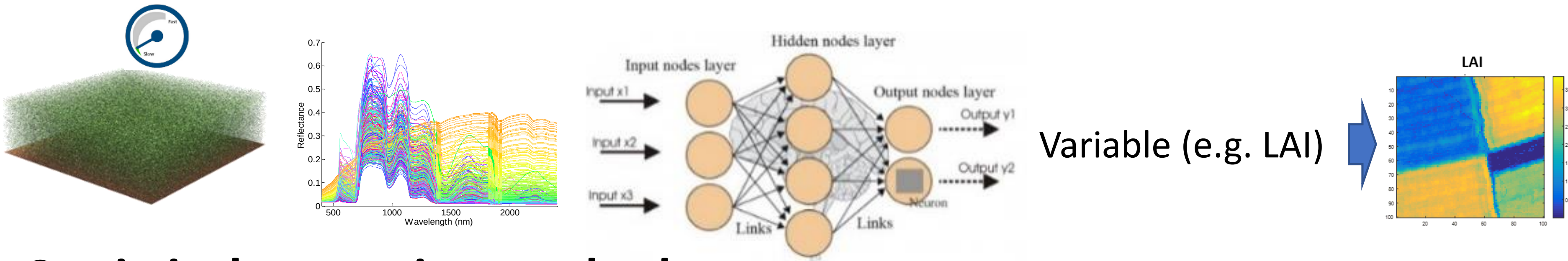




Hybrid models:
Regression vs. Emulation:



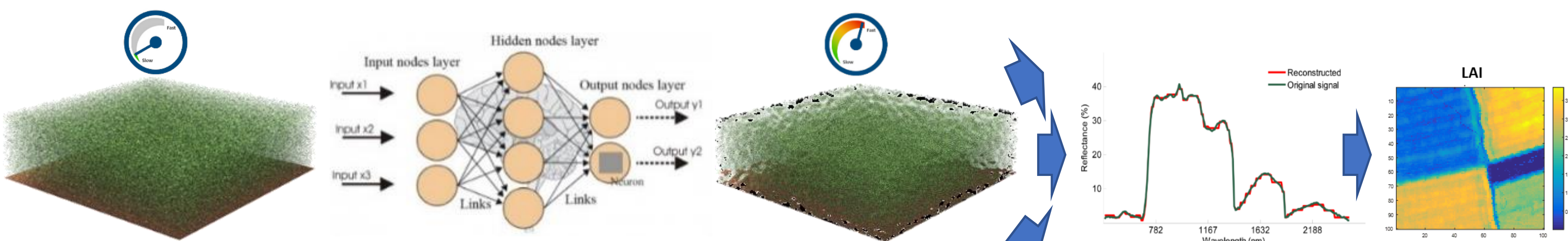
Common use of hybrid models in optical RS:



Statistical regression method:

- Variable/data-driven, 1 output, portability is questionable

Emulation in optical RS:



Replace RTM:

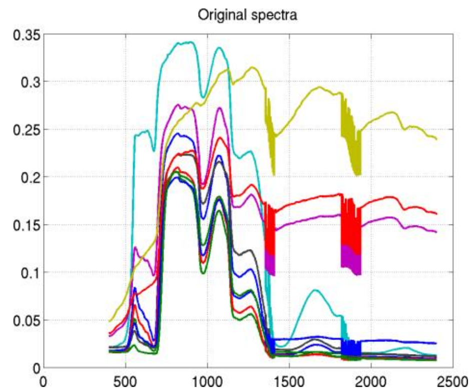
- Multiple applications, e.g. inversion
  - ✓ Radiometric method: Spectral fitting
  - ✓ Portable: generally applicable

“spectral redundancy” is a blessing

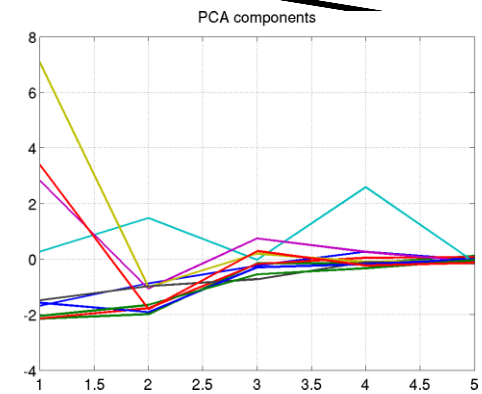
Processing steps emulation

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

PCA on spectra



$$Sc = U \cdot X$$



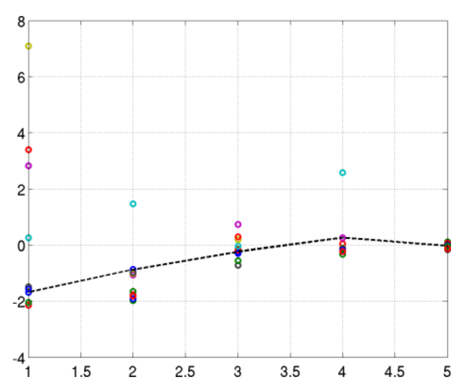
MLRA training looping over components

$$W = (Y + \lambda I)^{-1} \cdot Sc$$

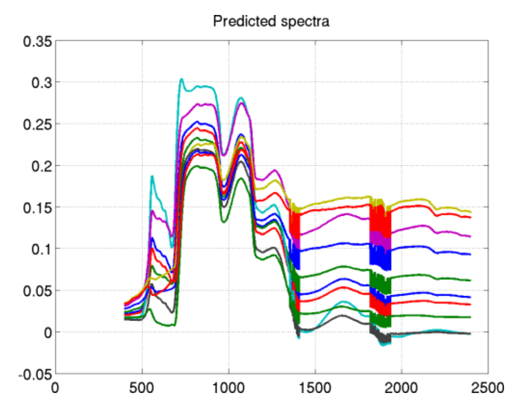
Prediction of components

$$Sp = Sc \cdot W$$

Reconstruction of spectra



$$Xr = U^{T} \cdot Sp$$



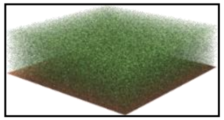
Validation → Emulator

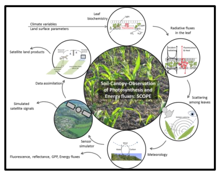
Emulator toolbox

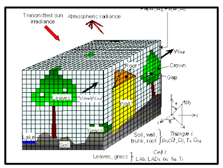
http://artmotoolbox.com

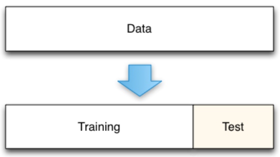
With ARTMO's emulation processing chain any RTM can be converted into an emulator.

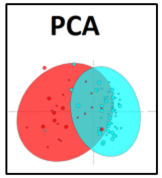
Input (variables + spectra) → Splitting into training/validation → PCA on spectra → MLRA training looping over components → Prediction of components → Reconstruction of full spectrum → Validation → Emulator

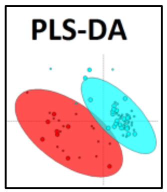


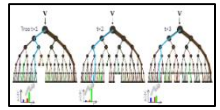


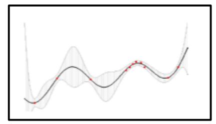


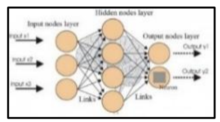


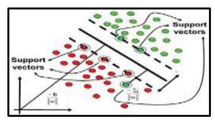


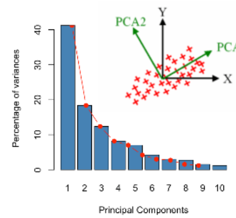








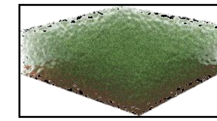


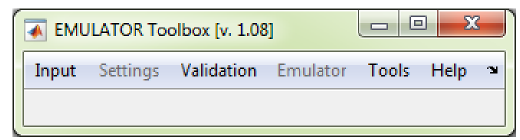


$$Xr = U^{T} \cdot Sp$$

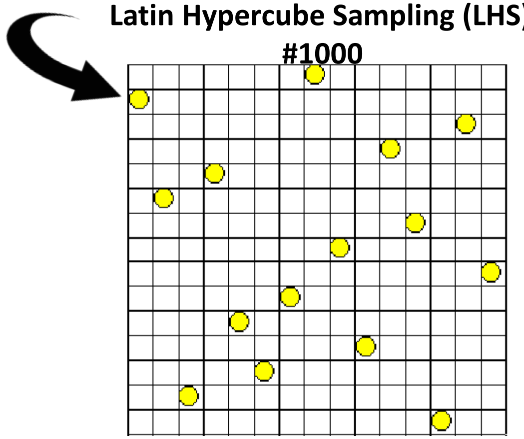
$$RMSE = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

$$NRMSE = \frac{1}{Y_{max} - Y_{min}} \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N}},$$





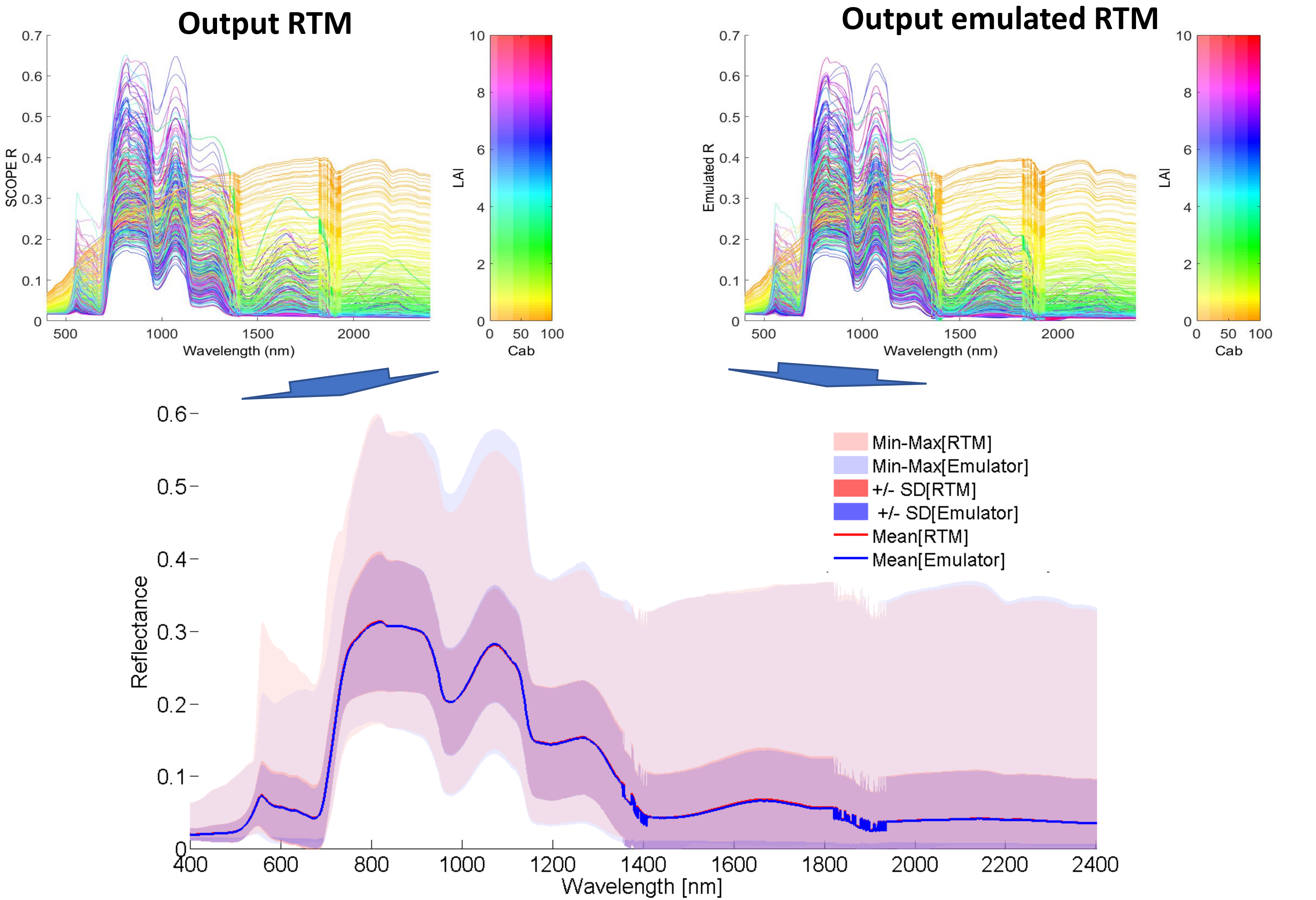
Latin Hypercube Sampling (LHS) #1000



Input → Settings → Validation → Emulator

- **Input**
  - RTM data → Select project, Edit settings
  - Txt data
- **Settings**
- **Validation**
  - New
  - Load
- **Emulator**
  - RTM vs Emulator
  - LUT Emulator
  - Txt Emulator
  - Scene Emulator
- **Tools**
  - Save
  - Load
  - Manage tests → Delete, Rename
  - Options
  - View figure
  - Plot LUT
  - Residual analyzer
  - Statistical analyzer
  - Import
  - Scene comparison
- **Help**
  - Show Log
  - User's manual
  - Installation guide
  - Disclaimer

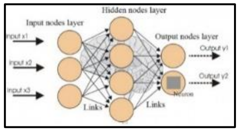
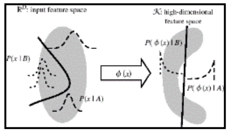
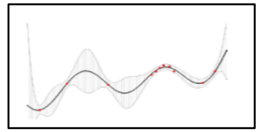
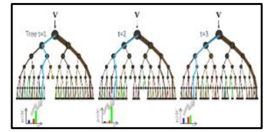
Rivera, J.P., Verrelst, J., Gómez-Dans, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, G. (2015). An Emulator Toolbox to Approximate Radiative Transfer Models with Statistical Learning. Remote Sensing. p. 7, 9347-9370.



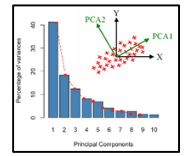
Errors can go down **< 3%** (NRMSE) by optimizing ML, sample size, DR,…

Emulators great idea… what about accuracy?

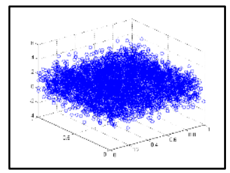
1) Role of machine learning regression algorithm?

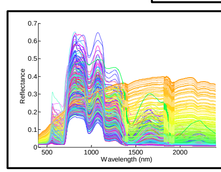
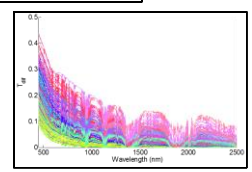
2) Role of dimensionality reduction (DR) method?



3) Role of LUT size training?



4) Role of data type?

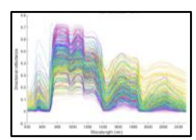
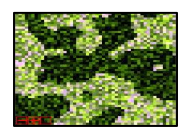
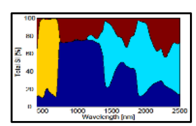
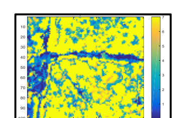
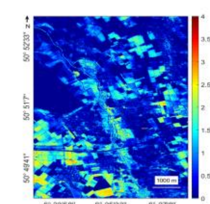
 

All these factors determine emulation accuracy. Some testing is required*

If OK with losing some accuracy, various applications are possible:

Fast RTM output generation:

1. ***Fast spectral generation***
2. ***Fast scene generation***
3. ***Fast global sensitivity analysis***
4. ***Fast approximation of retrieval strategies***

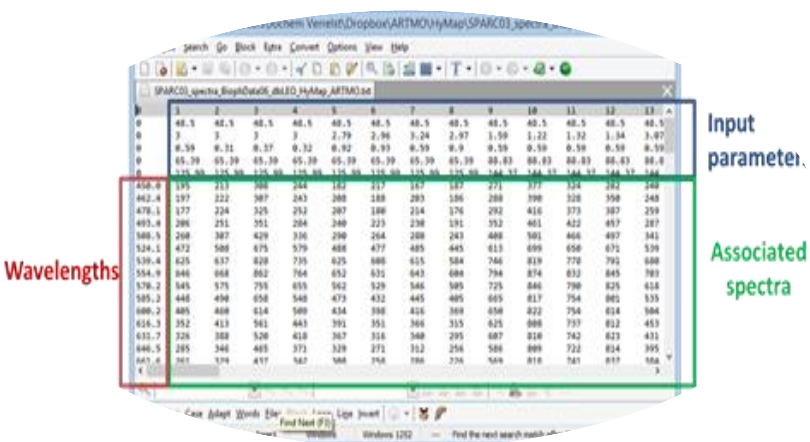
    

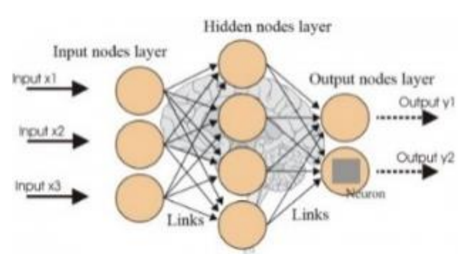
* Verrelst, J., Rivera Caicedo, J.P., , Muñoz-Marí, J., Camps-Valls, G., Moreno, J. (2017). SCOPE-Based Emulators for Fast Generation of Synthetic Canopy Reflectance and Sun-Induced Fluorescence Spectra. Remote Sensing. 9(9), 927.

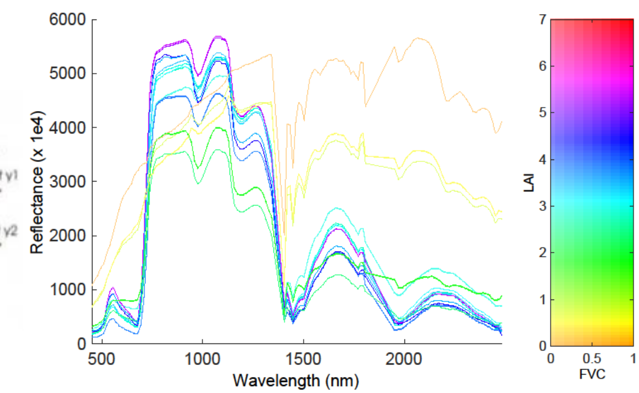
Applications (1/4)

Field data

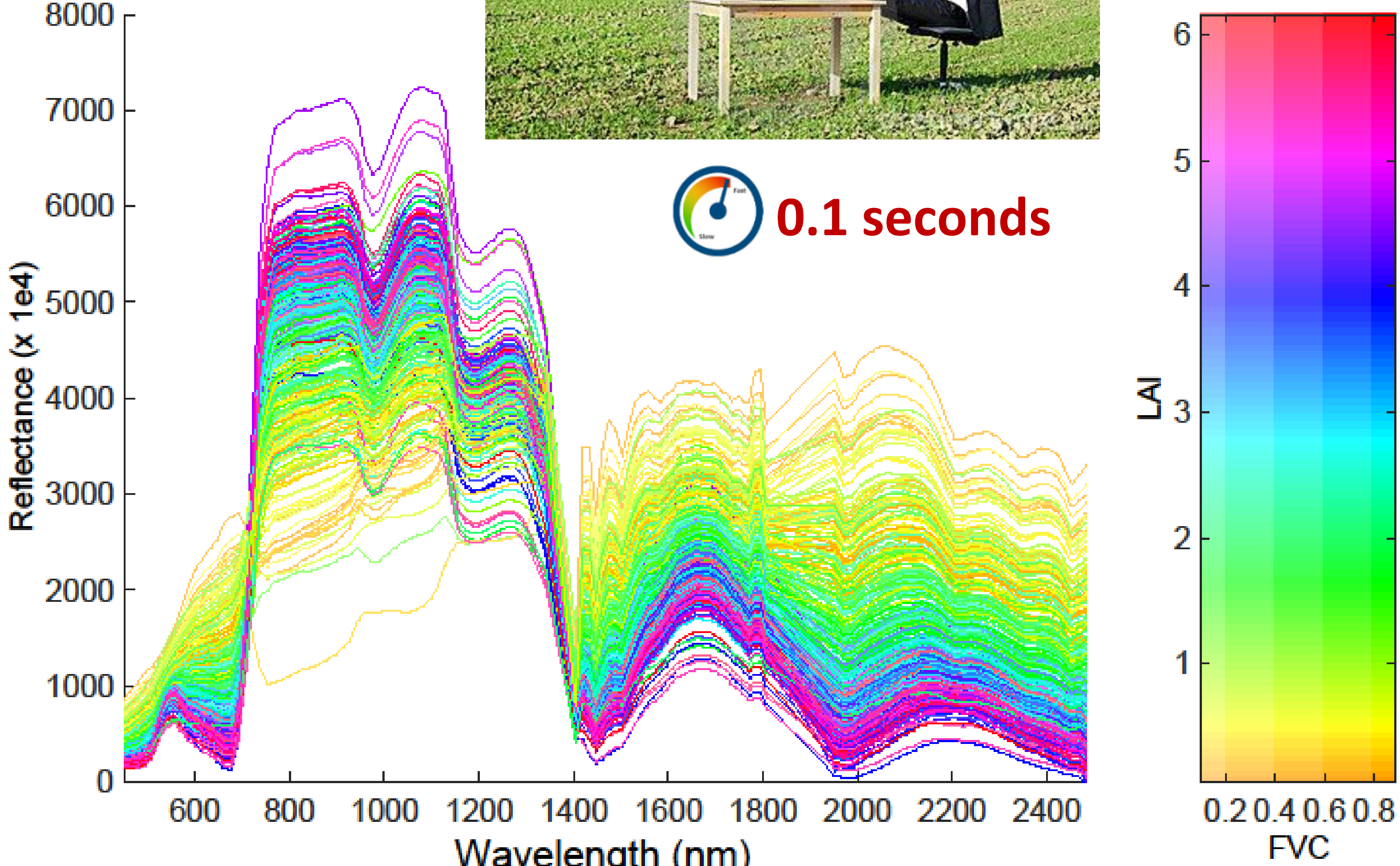








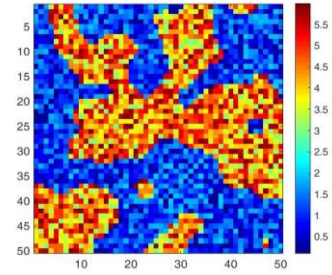
Example of #500 emulated SPARC campaign spectra based on varying 6 field variables



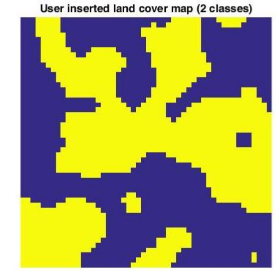
0.1 seconds

Applications (2/4)

Scene generation

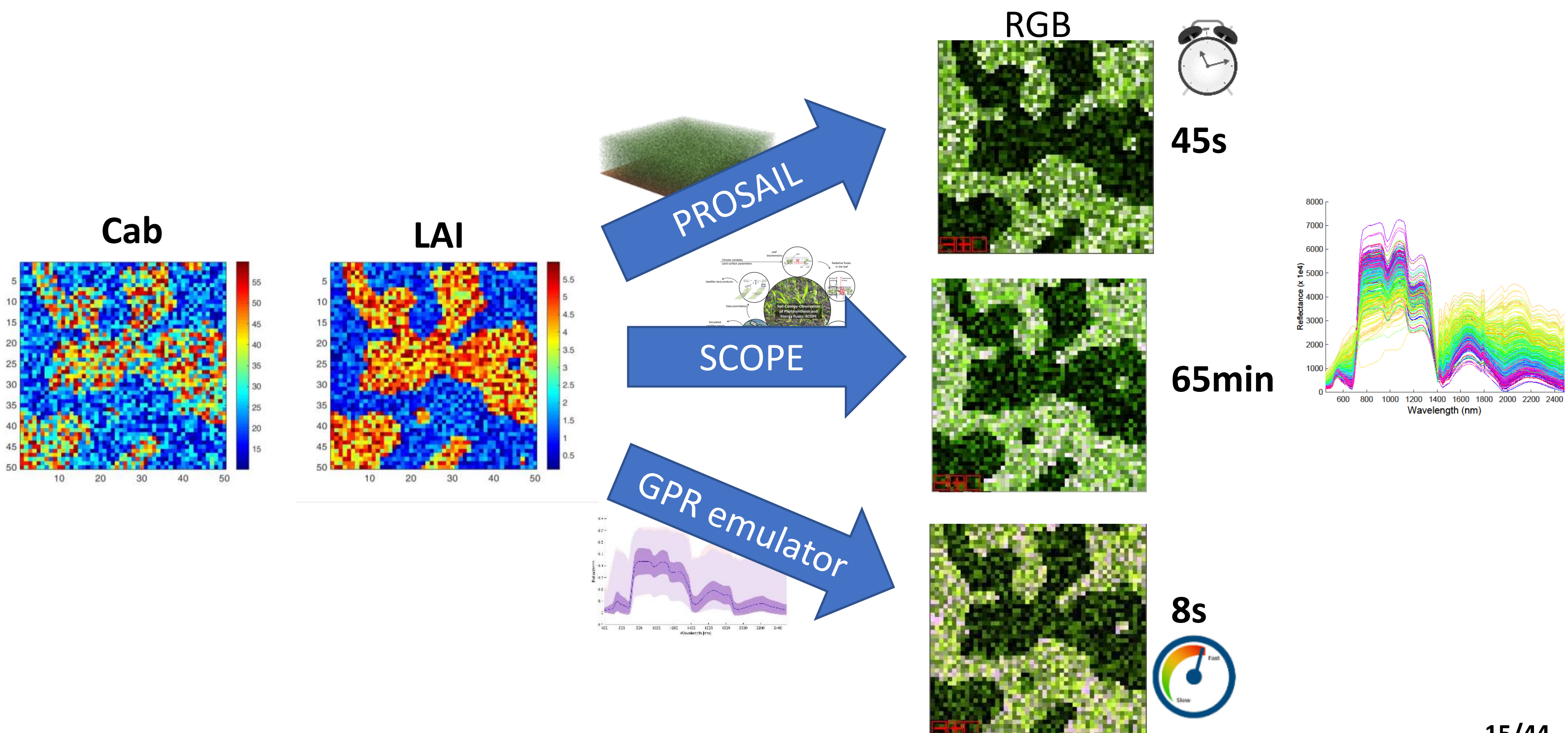


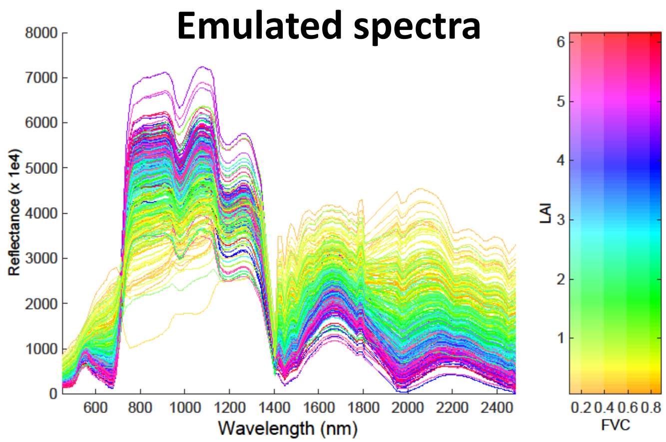
Emulation for hyperspectral scene generation

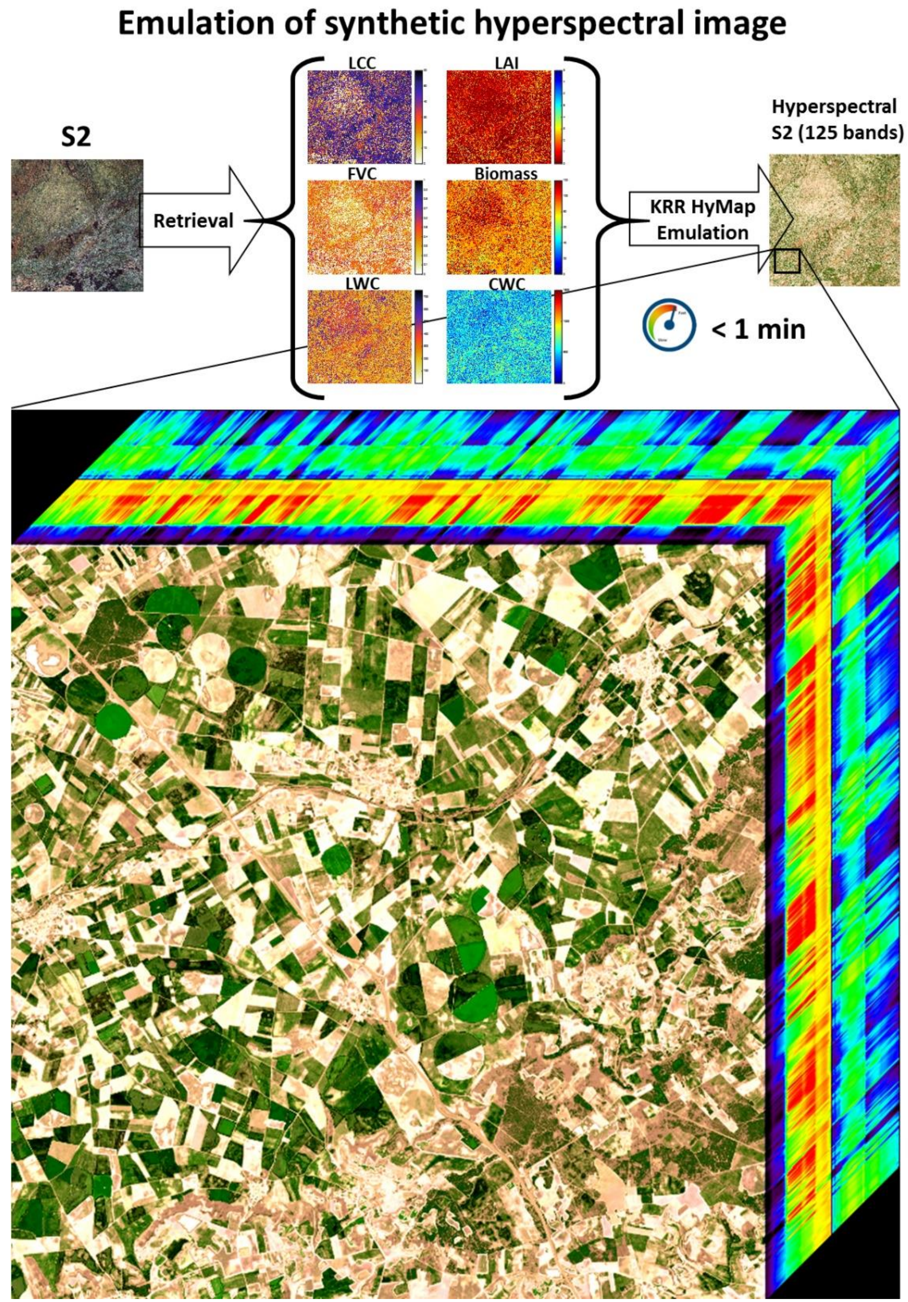


- GPR emulator applied for scene generation
- Compared against RTM scenes: PROSAIL & SCOPE

Parameter	Symbol	Units	Class 1	Class 2
Leaf area index	LAI	m^2/m^2	Uniform: 3 - 6	Uniform: 0 - 2
Chlorophyll a+b content	C_{ab}	$\mu g/cm^2$	Uniform: 20 - 60	Uniform: 10 - 35







Emulation for realistic synthetic scene generation.

A S2 image is used as input to emulate an hyperspectral image with a S2 texture.

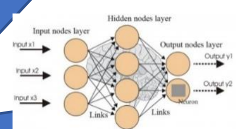
Verrelst, J., Rivera Caicedo, J.P., Vicent, J., Morcillo-Pallarés, P., Moreno, J. (2019). [Approximating Empirical Reflectance Data through Emulation: Opportunities for Synthetic Scene Generation.] Remote Sensing. 11(2): 157.

Emulation for image fusion

Image with high quality spatial information
- Large image (e.g. S2)
- High spatial but low spectral resolution (e.g. Worldview, UAV)

Image with high spectral information
- Hyperspectral image

Input



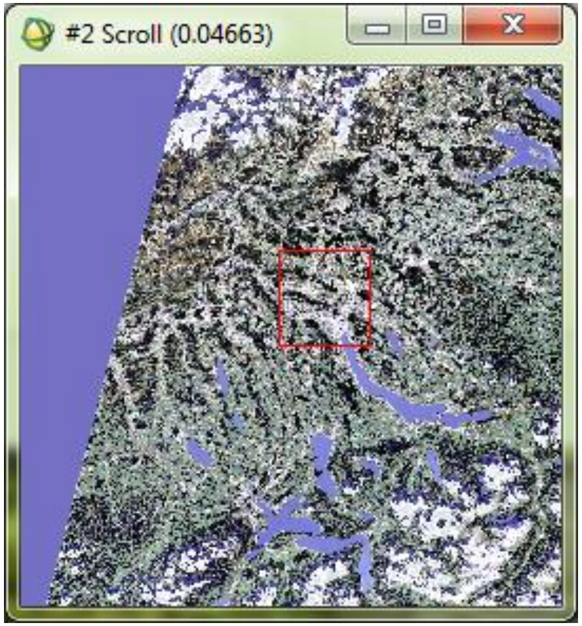
Output

Emulator

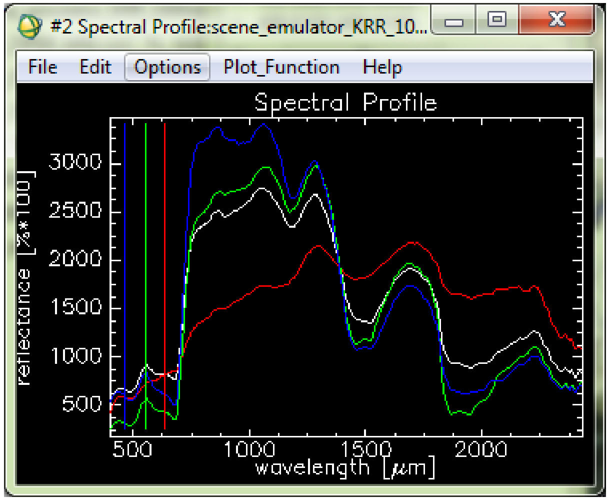
- Training
- Testing

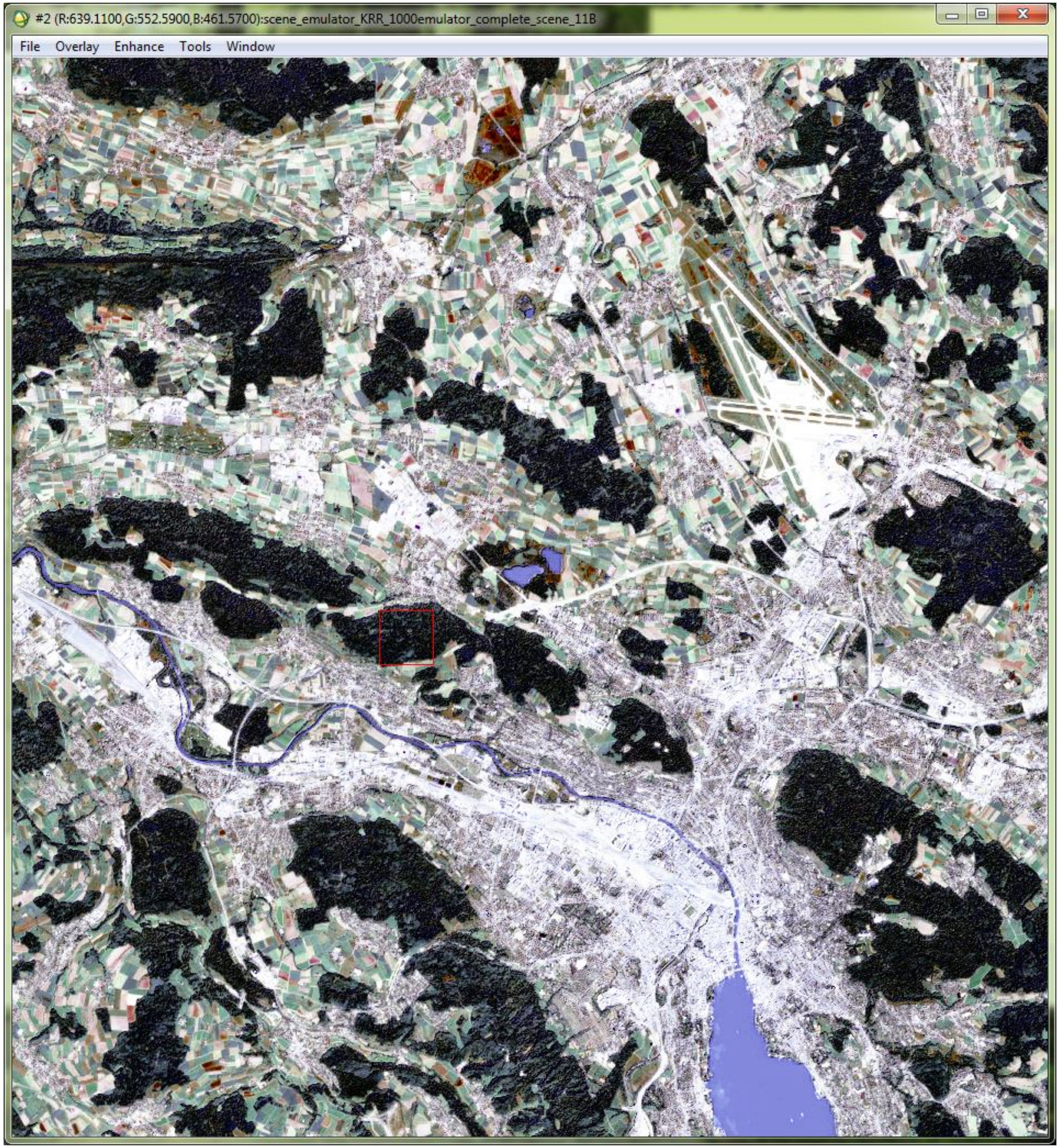
Emulated image with high spectral and spatial information

S2...

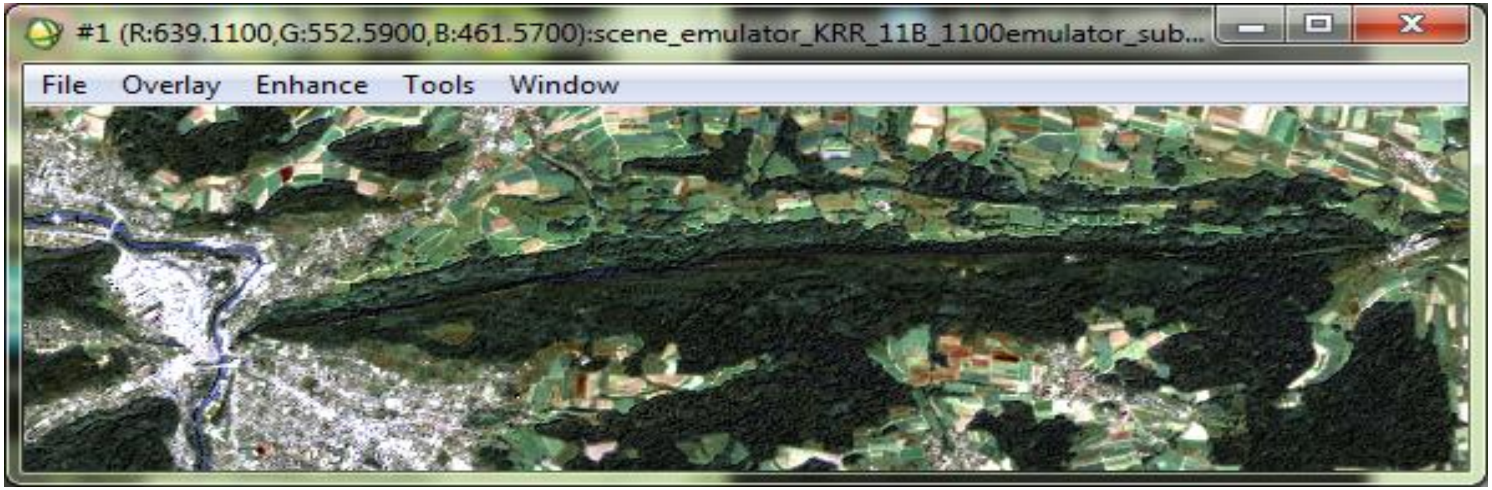


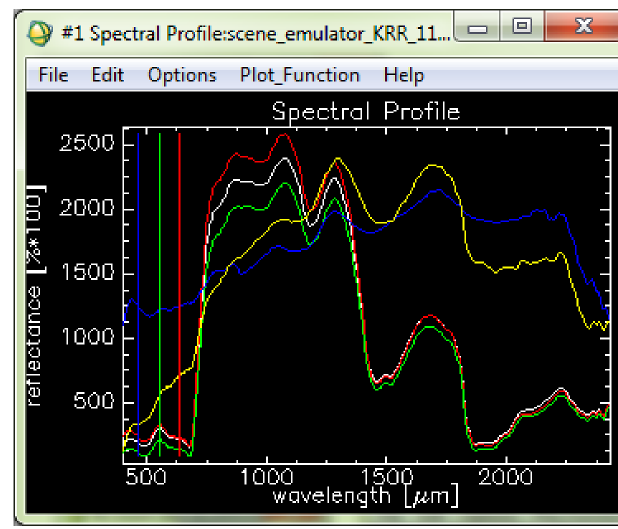
10 min; 67GB





S2 texture, APEX-like hyperspectral info



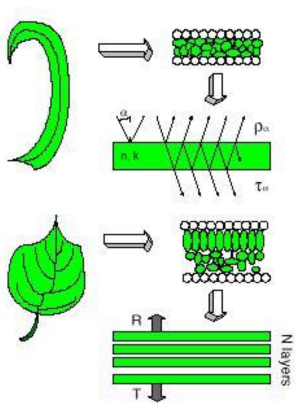


Original APEX image

Applications (3/4)

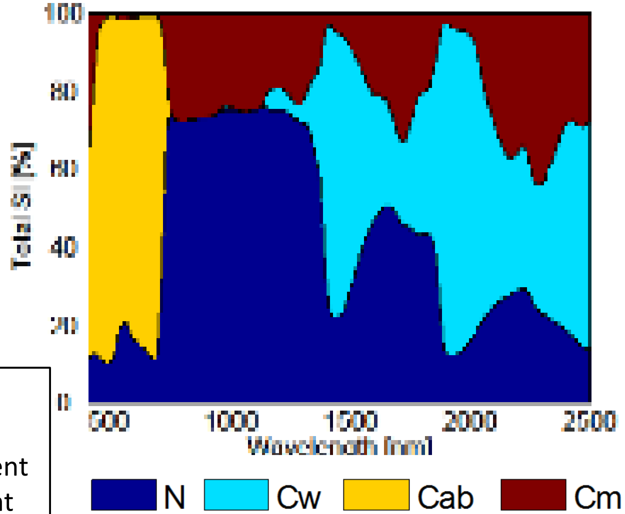
Global Sensitivity Analysis (GSA)

GSA techniques quantify the relative importance of each input variable to model outputs.



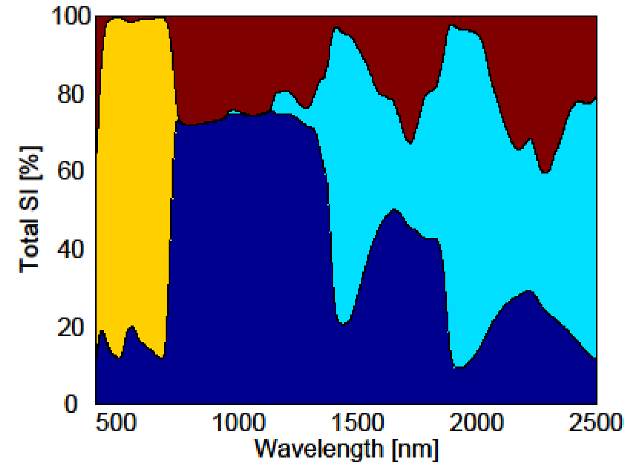
N: structural variable
Cw: Water content
Cab: Chlorophyll content
Cm: dry matter content

PROSPECT 4



X 2

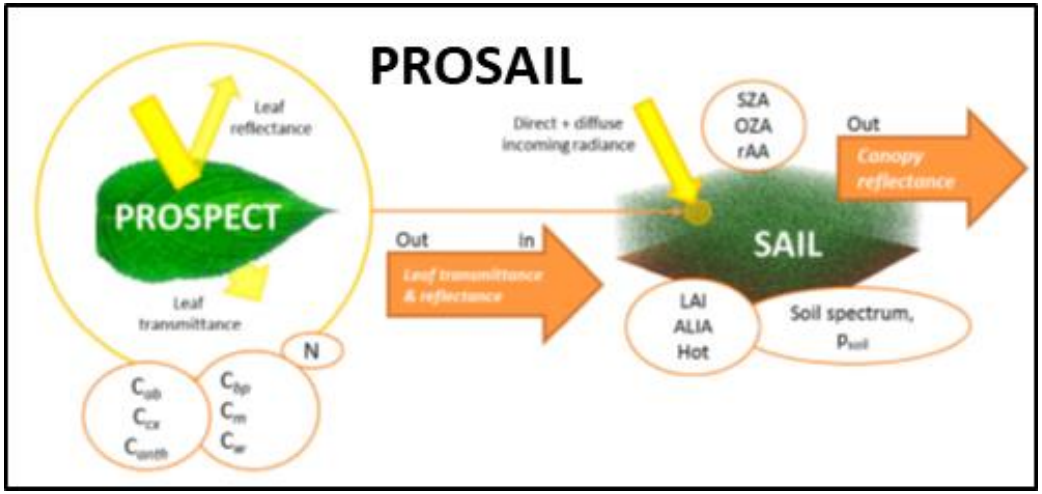
Emulator GPR



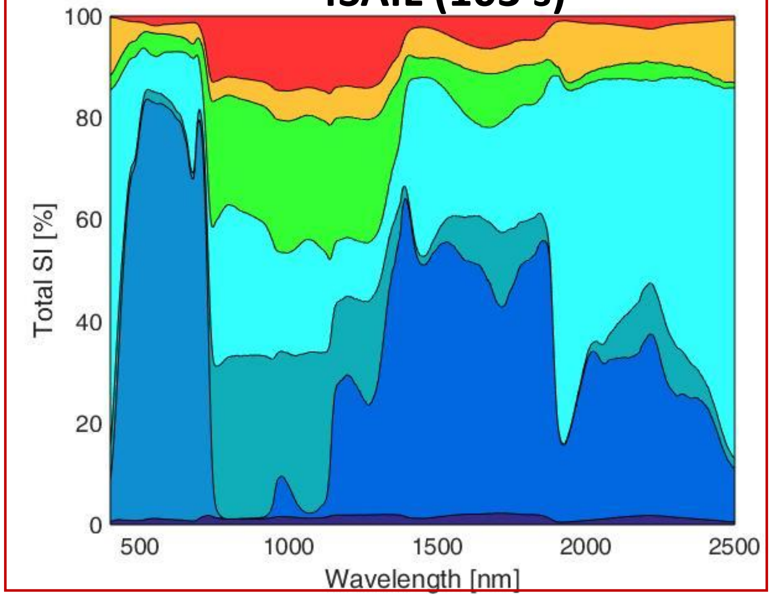
Verrelst, J., Sabater, N., Rivera, J.P., Muñoz-Marí, J., Vicent, J., Camps-Valls, G., Moreno, J. (2016). Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis. Remote Sensing. 8(8), 673.

PROSAIL

1000#/variable



4SAIL (103 s)

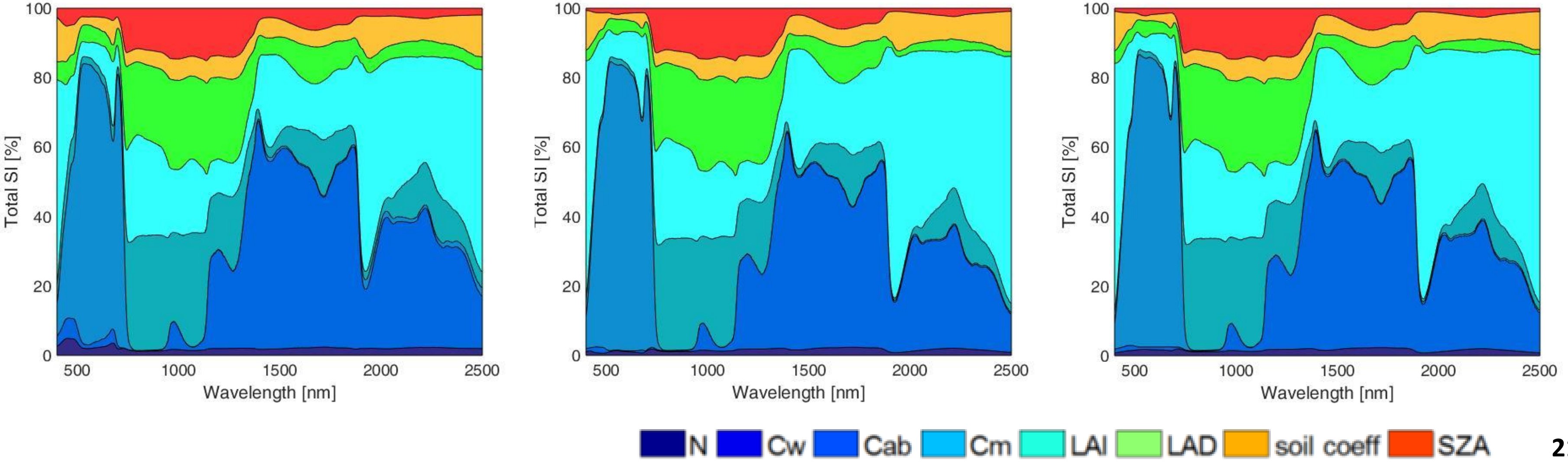


X 3

Emulator KRR (28s)

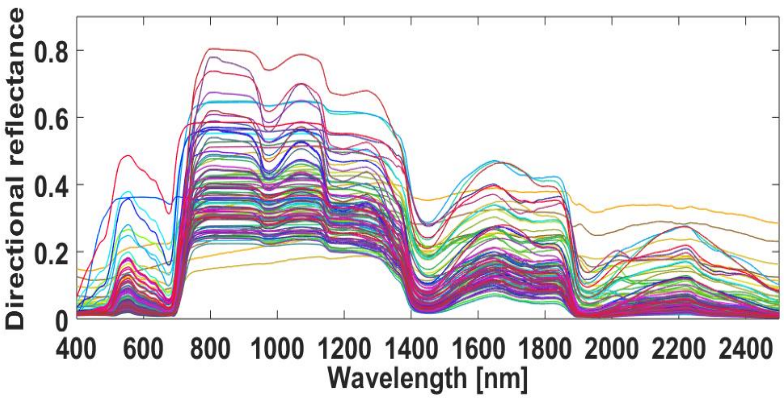
Emulator GPR (32s)

Emulator VHGPR (37s)

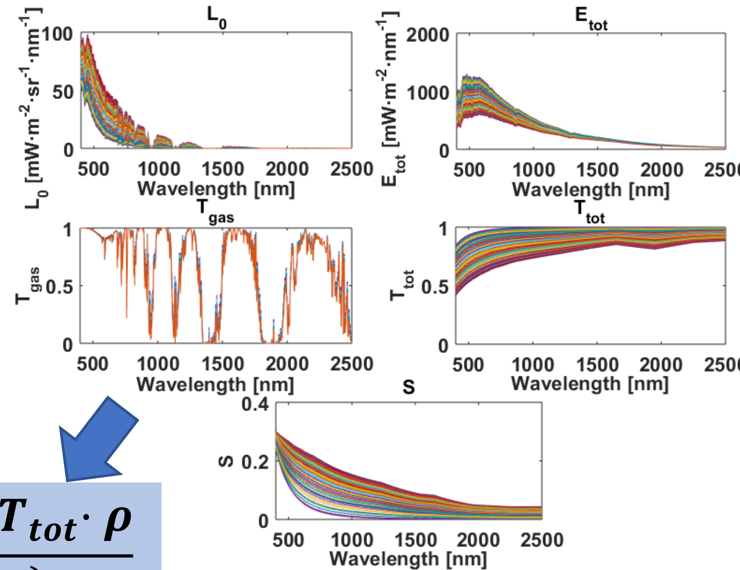


Applying emulation to L_{TOA} (1/3)

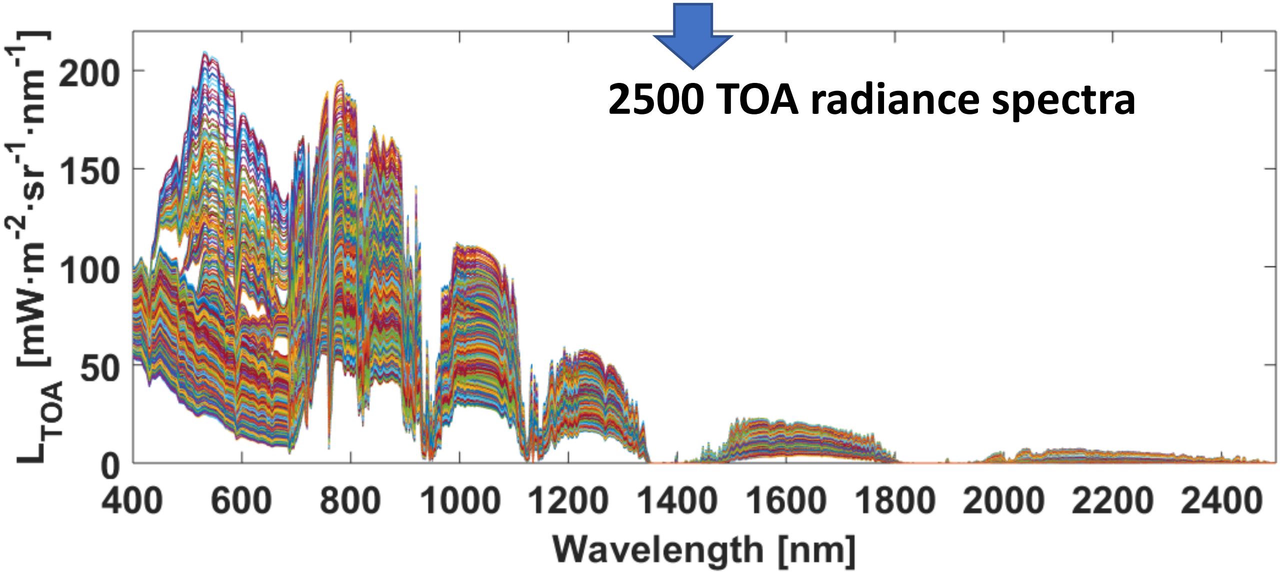
PROSAIL spectra



6S or MODTRAN spectra



$$L_{TOA} = L_0 + \frac{T_{gas} \cdot E_{tot} \cdot T_{tot} \cdot \rho}{\pi(1 - S\rho)}$$



2500 TOA radiance spectra

Emulator

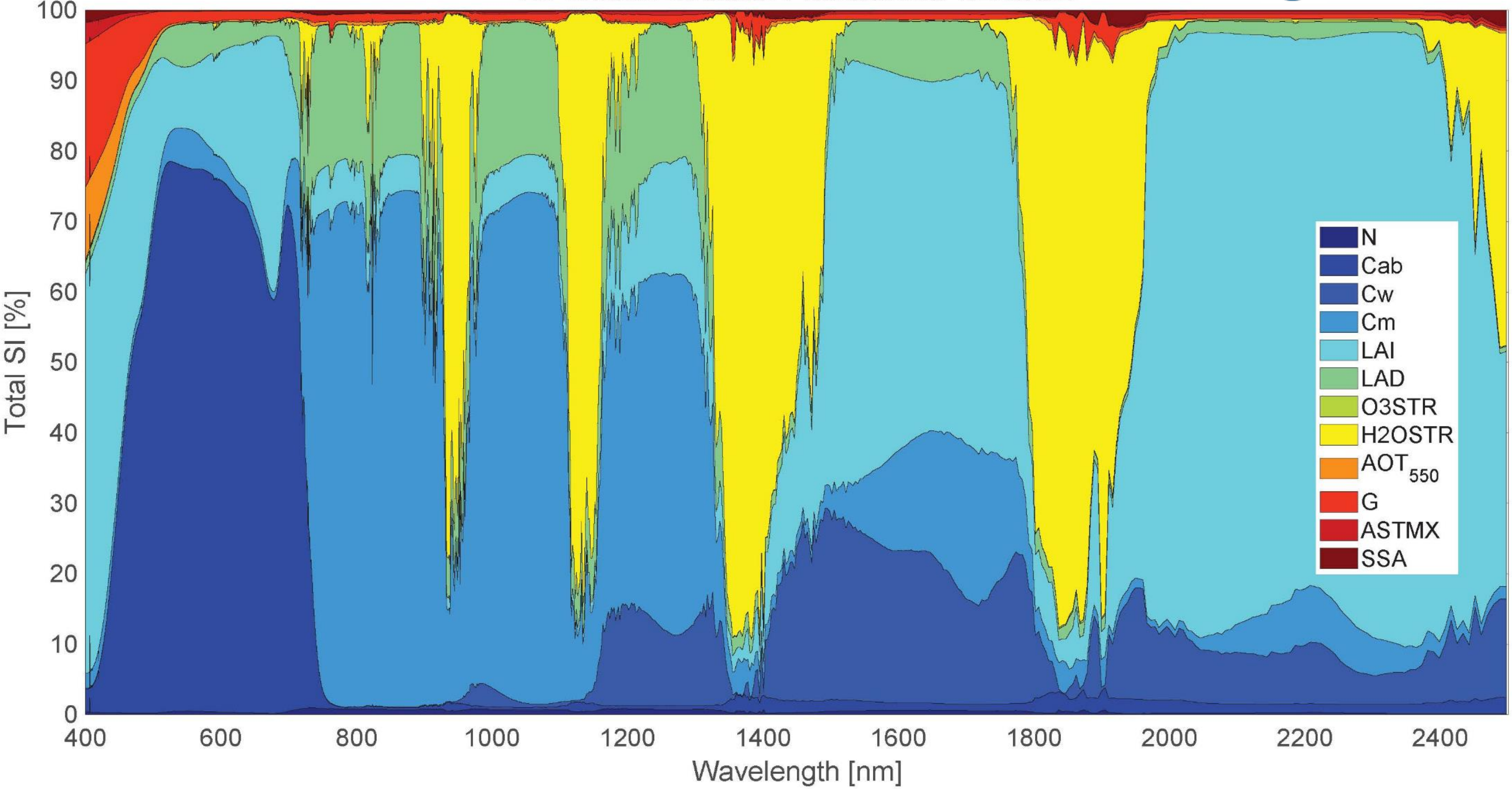
- **30 PCA**
- **70/30% train/test**

L_{TOA} GSA results GPR emulation (#1000/var) (2/3)

PROSAIL + MODTRAN

GSA type:Saltelli / Subsamples:1000
RT model:Emulator / Sensor:NO SENSOR

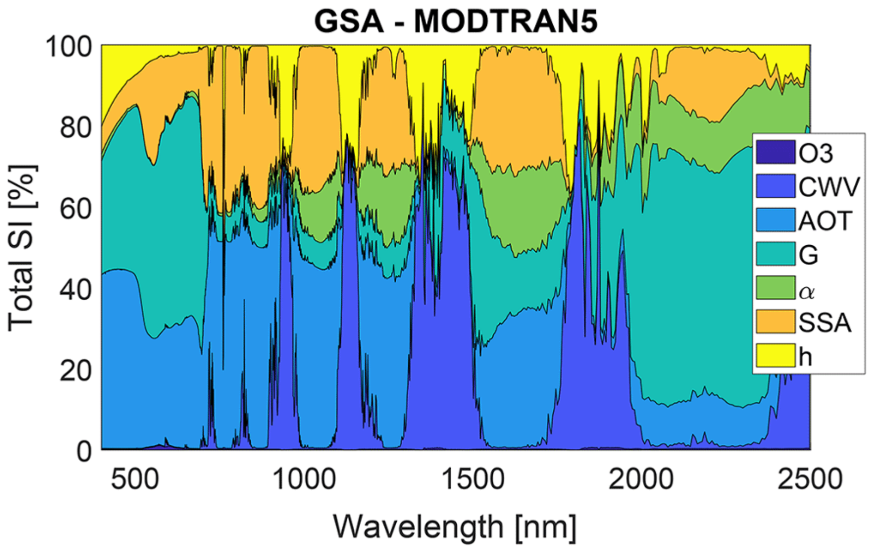
36 s

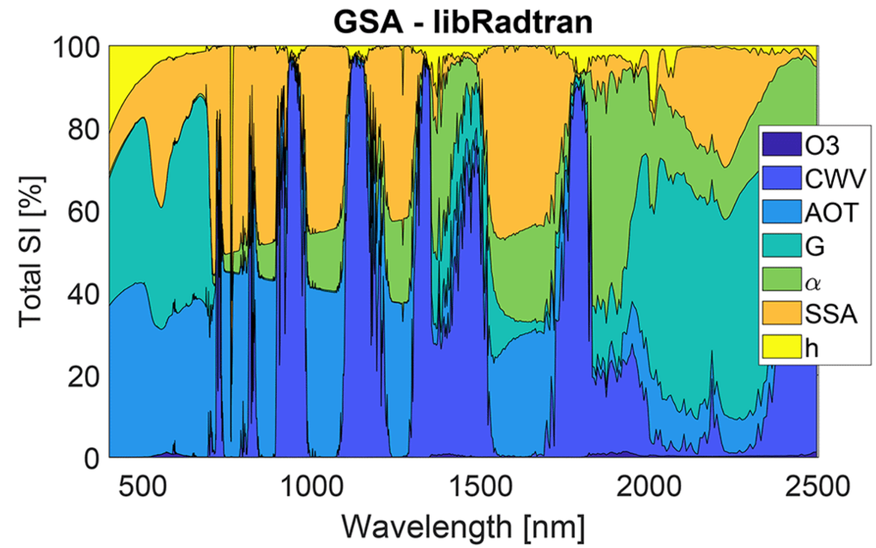


Some regions are perfectly fit for variable retrieval from TOA radiance data → **Training a ML**

Verrelst, J. Vicent, J., Rivera Caicedo, J.P., Lumbierres, M., Morcillo Pallarés, P., & Moreno, J. (2019). [Global Sensitivity Analysis of Leaf-Canopy-Atmosphere RTMs: Implications for Biophysical Variables Retrieval from Top-of-Atmosphere Radiance Data](). Remote Sensing. 11(6), 1923.

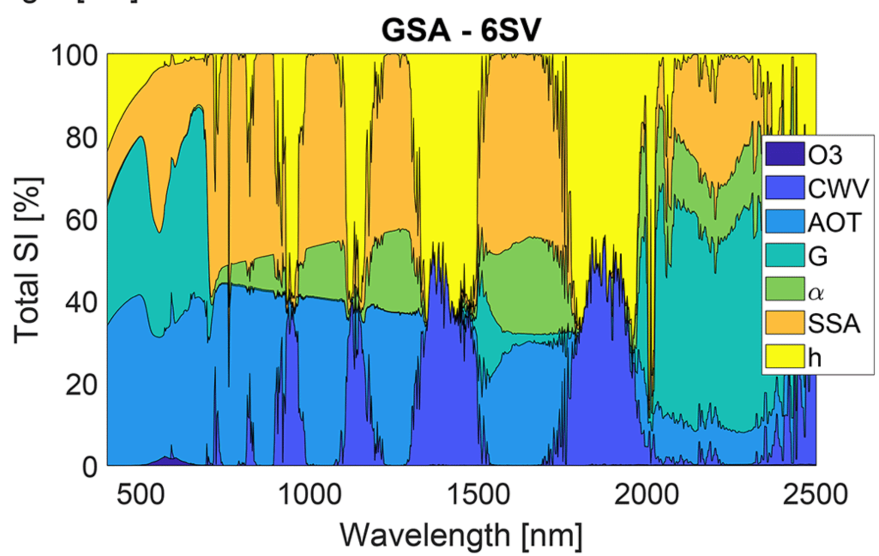
Emulation for GSA of atmospheric RTMs (3/3)





$$L_{\text{toa}} = L_0 + \frac{(E_{\text{dir}}\mu_{\text{il}} + E_{\text{dif}})(T_{\text{dir}} + T_{\text{dif}})\rho}{\pi(1 - S\rho)}$$

Variable name	Min–max
Elevation (h):	0–3 km
Aerosol optical thickness (AOT):	0.05–1
Ångström exponent (α):	0.1–1.5
Asymmetry parameter (G):	0.6–1
Single scattering albedo (SSA):	0.75–1
Water vapor (CWV):	1–4 g cm^2
Ozone (O_3):	0.25–0.45 atm-cm

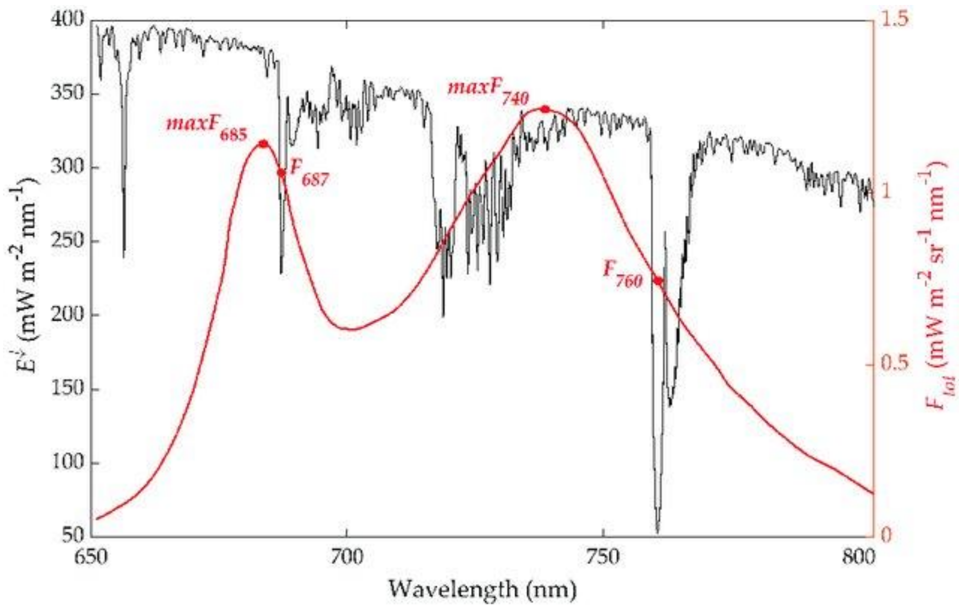


1< min

Vicent, J., Verrelst, J., Sabater, N., Alonso, L., Rivera-Caicedo, J. P., Martino, L., Muñoz-Marí, J., and Moreno, J.: **Comparative analysis of atmospheric radiative transfer models using the Atmospheric Look-up table Generator (ALG) toolbox (version 2.0),** Geosci. Model Dev., 13, 1945–1957, https://doi.org/10.5194/gmd-13-1945-2020, 2020.

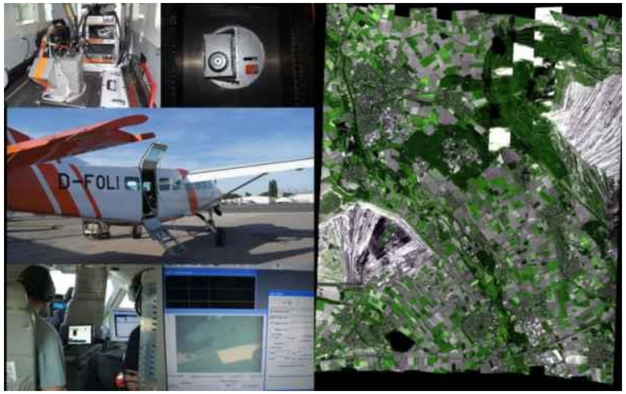
Applications (4/4)

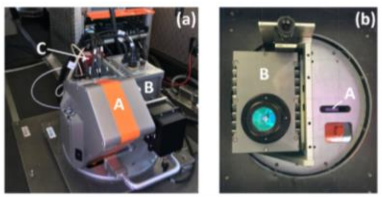
(SIF) Retrieval

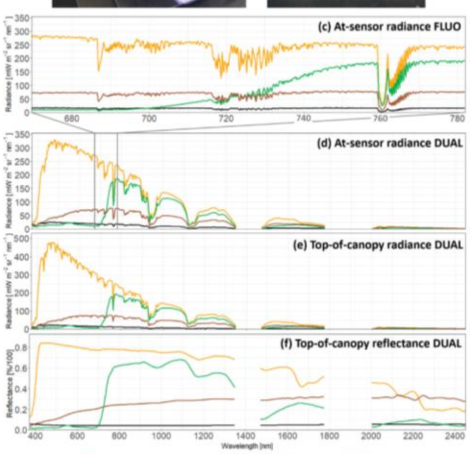


Emulation of sun-induced fluorescence (SIF) data

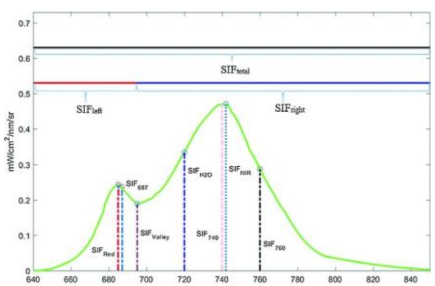
HyPlant FLUO

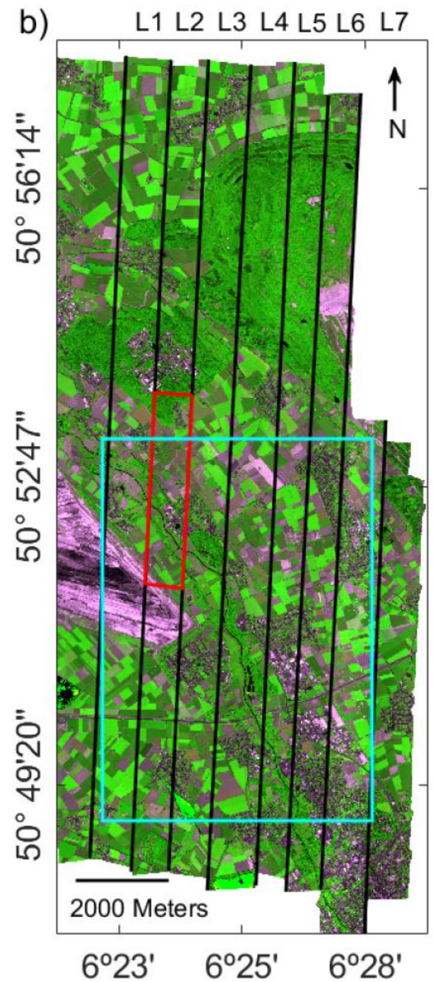






Spectral Fitting method (SFM) for SIF retrieval

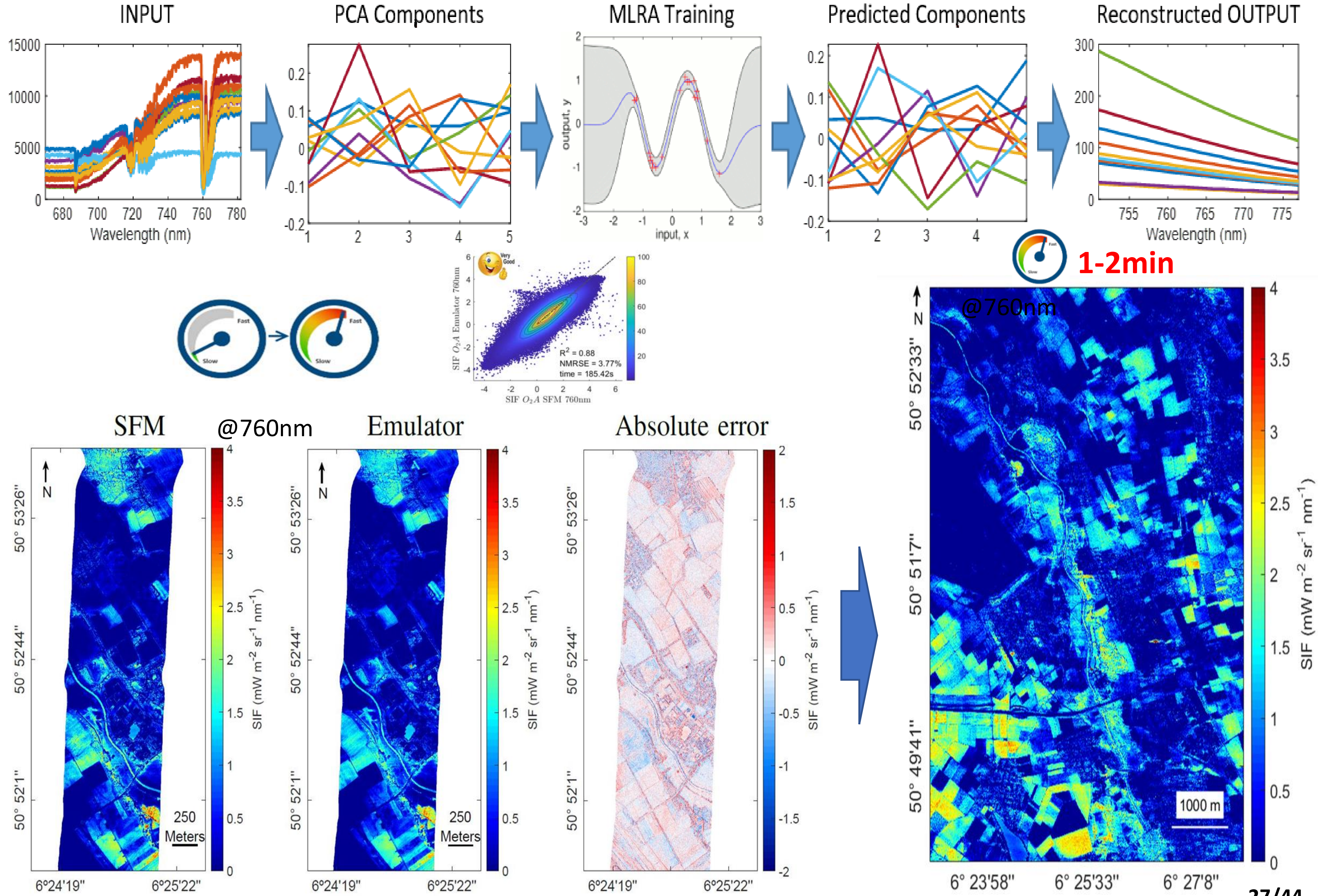




JÜLICH Forschungszentrum

Siegmann, B.; Alonso, L.; Celesti, M.; Cogliati, S.; Colombo, R.; Damm, A.; Douglas, S.; Guanter, L.; Hanuš, J.; Kataja, K.; Kraska, T.; Matveeva, M.; Moreno, J.; Muller, O.; Pikl, M.; Pinto, F.; Quirós Vargas, J.; Rademske, P.; Rodriguez-Morene, F.; Sabater, N.; Schickling, A.; Schüttemeyer, D.; Zemek, F.; Rascher, U. **The High-Performance Airborne Imaging Spectrometer HyPlant—From Raw Images to Top-of-Canopy Reflectance and Fluorescence Products: Introduction of an Automatized Processing Chain**. *Remote Sens.* **2019**, *11*, 2760.

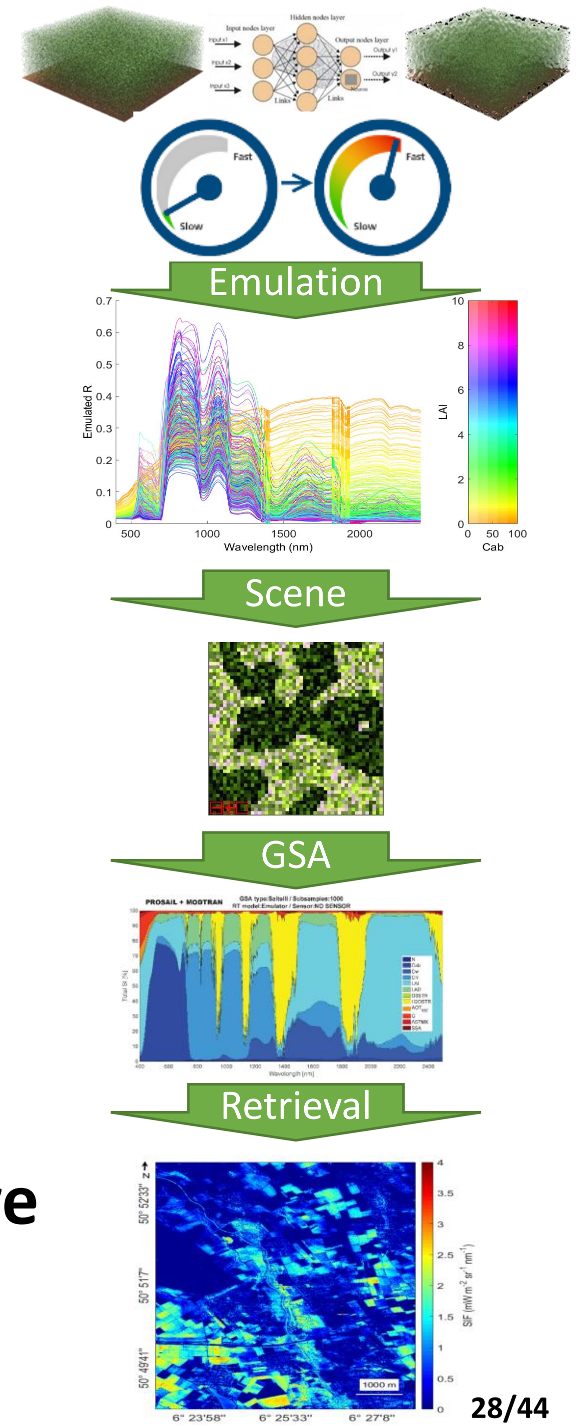
Emulator workflow: radiance to SIF with experimental data



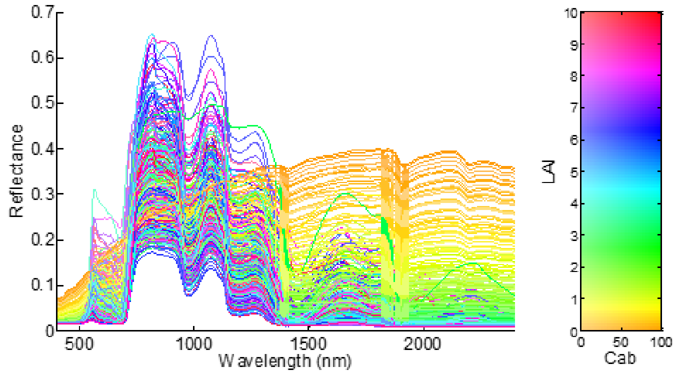
Conclusions emulation

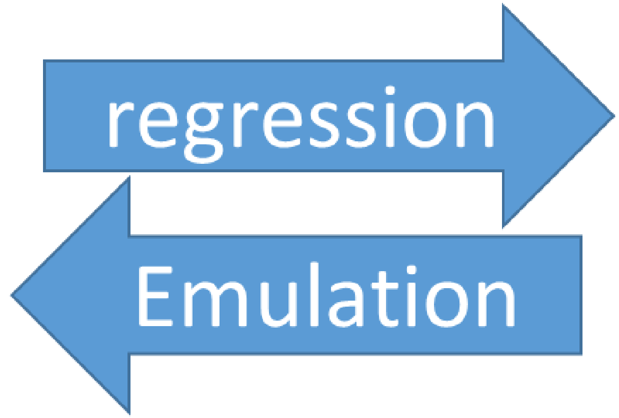
Emulation approximates physical models with sufficient accuracy and tremendous gain in speed thanks to ML.

- ✓ **Emulation permits <u>fast</u> rendering of optical data**
- ✓ **Emulation permits <u>fast</u> calculation of global sensitivity analysis**
- ✓ **Emulation can provide a <u>fast</u> alternative of tedious retrieval routines**



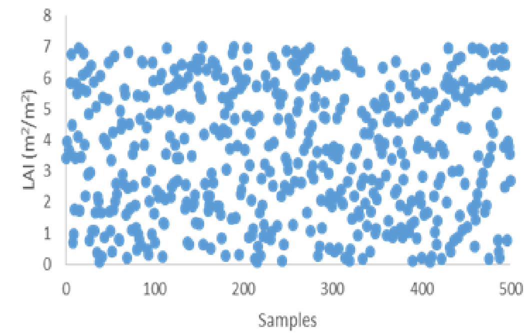
Spectra



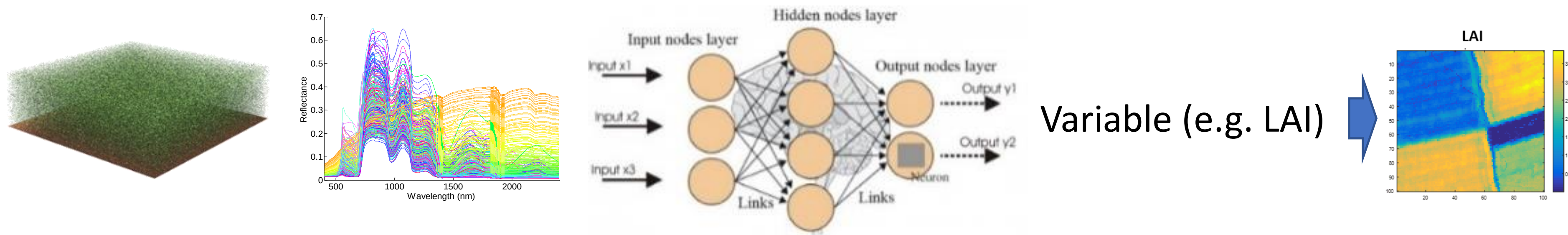


Variables

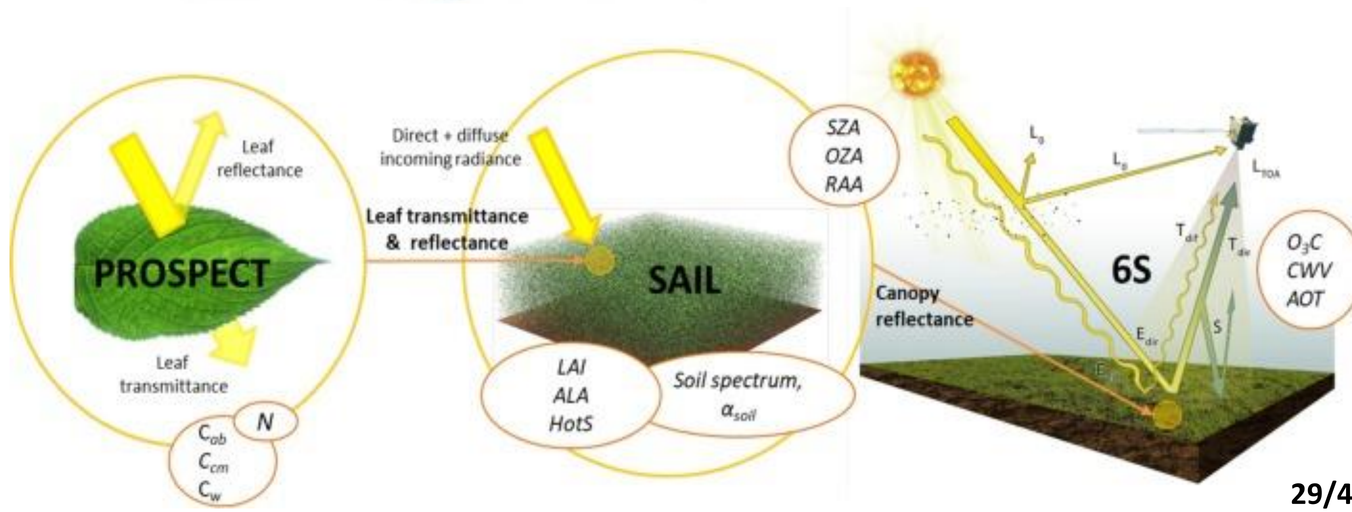
(e.g. LAI, chlorophyll,...)



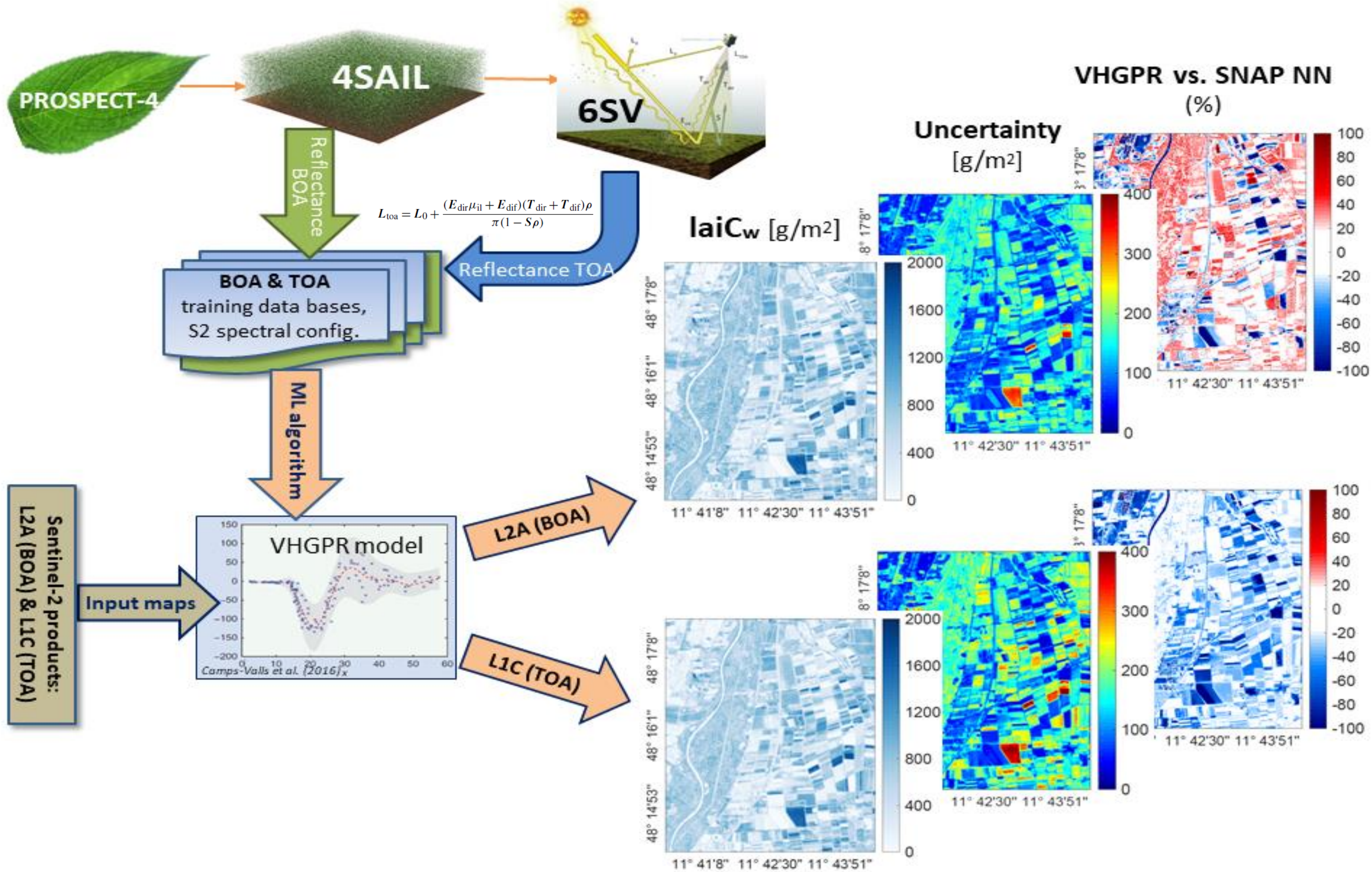
Hybrid retrieval methods



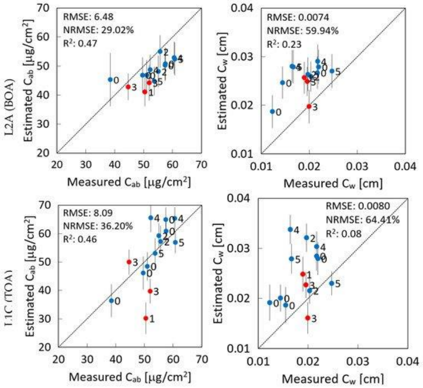
- **BOA retrieval**
- **TOA retrieval**

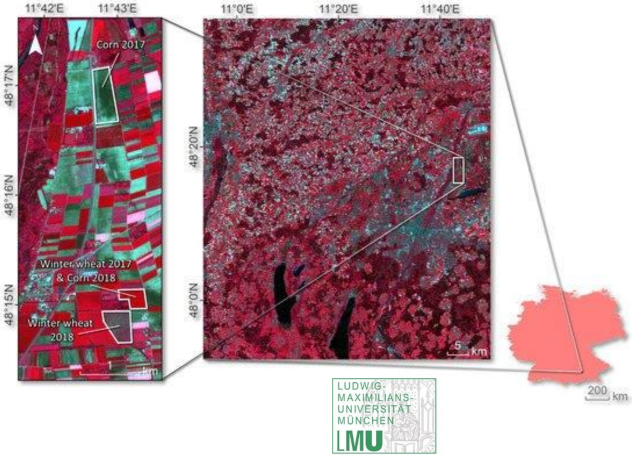


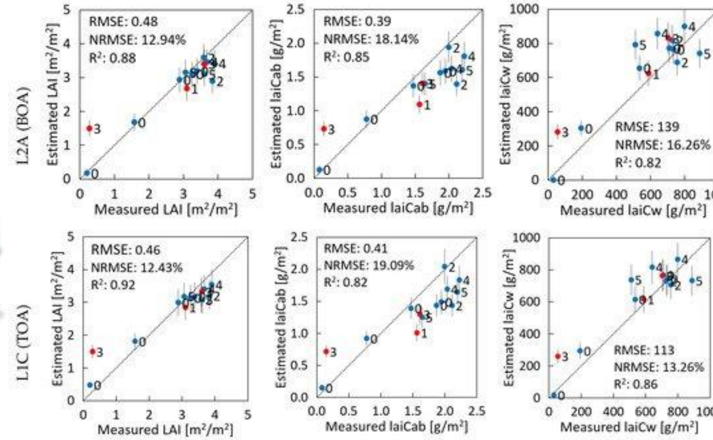
BOA & TOA retrieval from S2 L2A and L1C data



Estévez, J.; Berger, K.; Vicent, J.; Rivera-Caicedo, J.P.; Wocher, M.; Verrelst, J. **Top-of-Atmosphere Retrieval of Multiple Crop Traits Using Variational Heteroscedastic Gaussian Processes within a Hybrid Workflow**. *Remote Sens.* **2021**, *13*, 1589

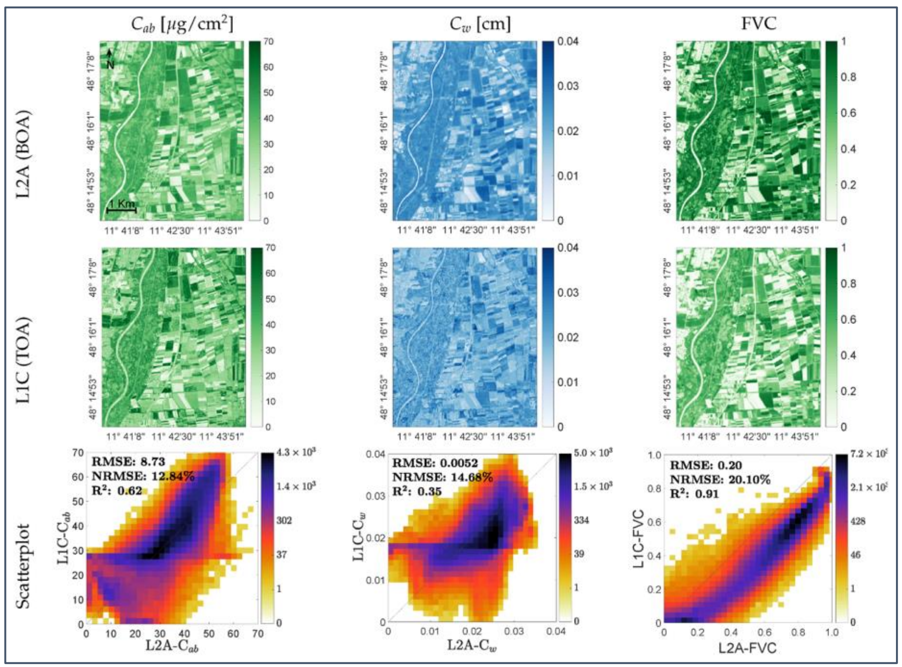


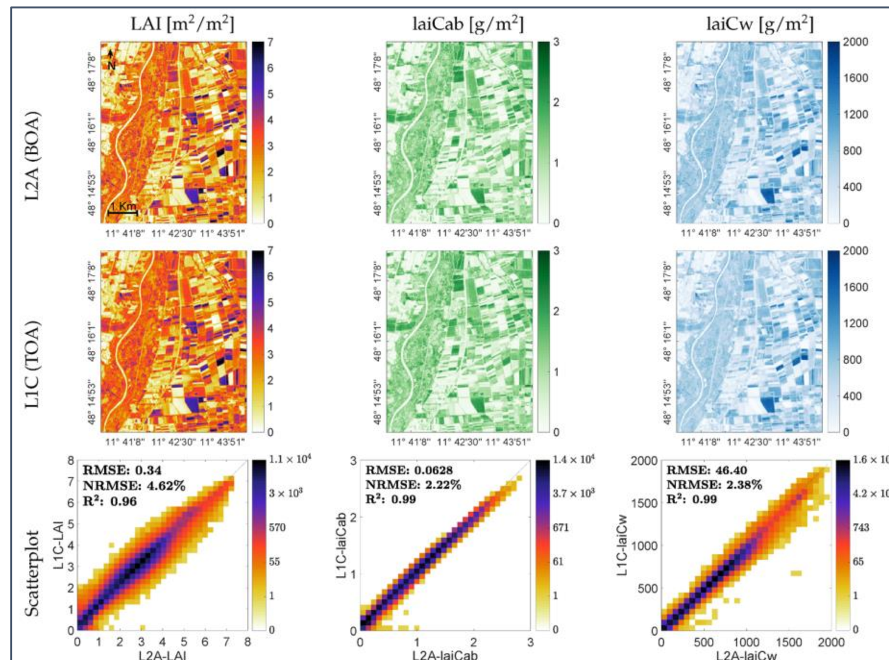




Biochemical leaf traits and FVC

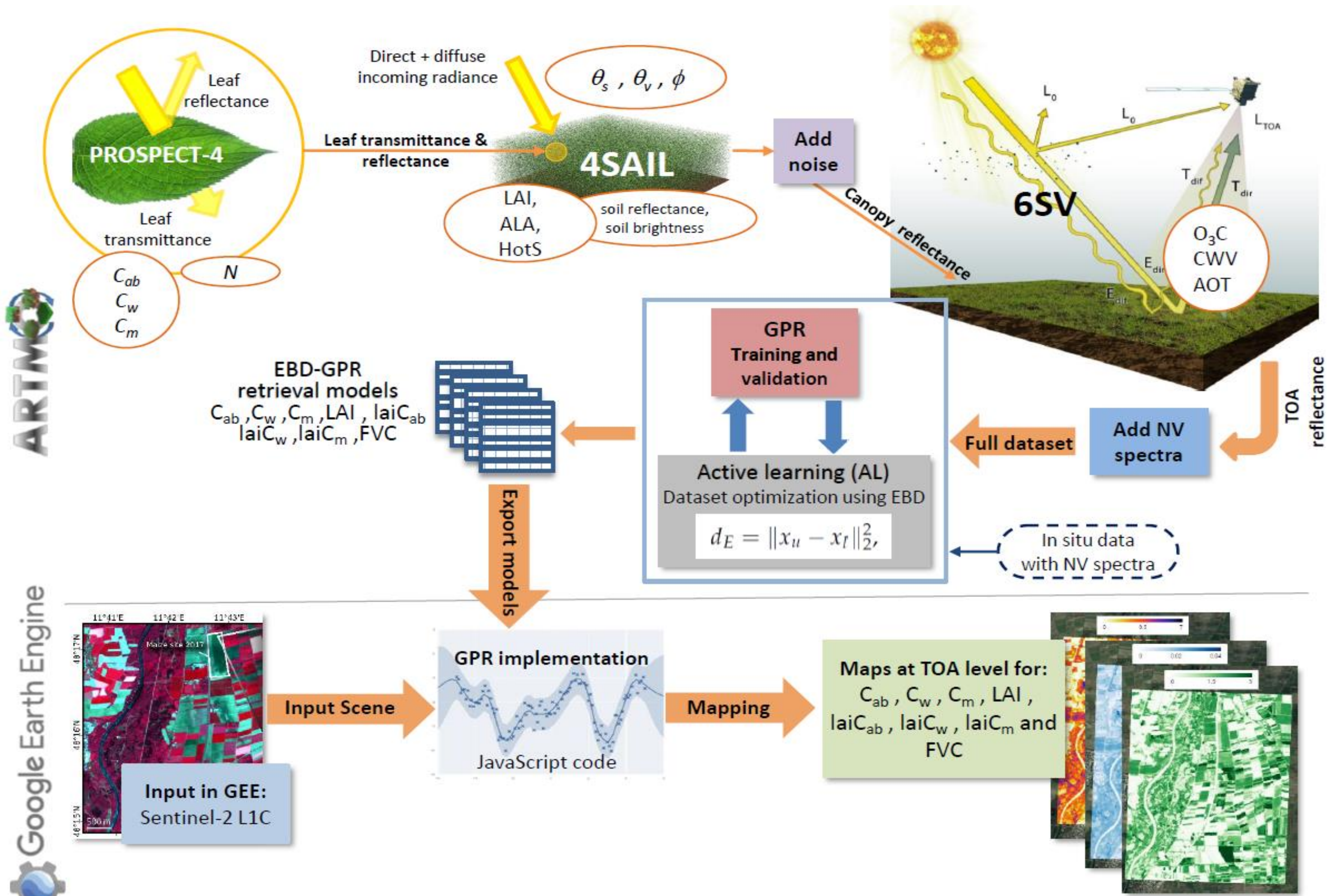
Canopy traits





Consistent BOA & TOA retrievals at the canopy scale: **no need for atmospheric correction** *(given a clear sky)*

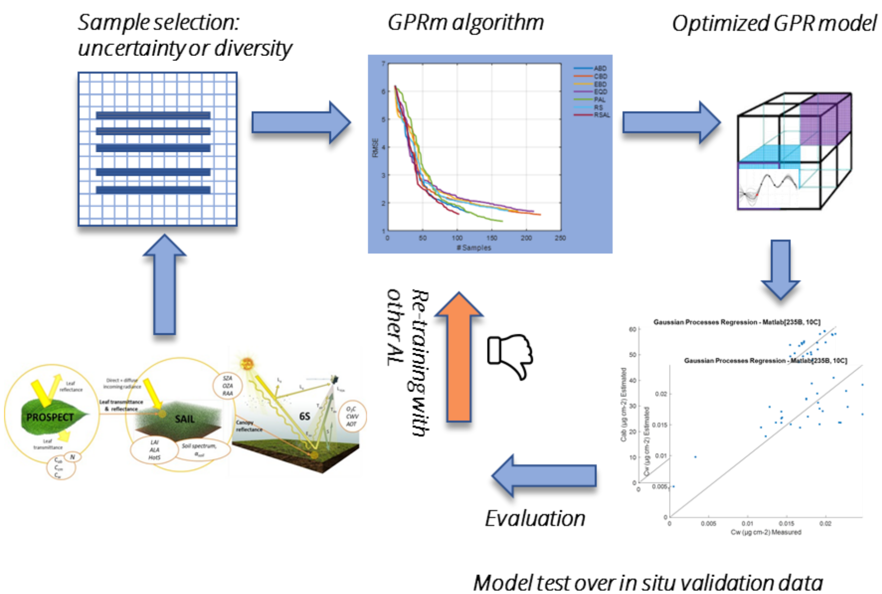
TOA retrieval from S2 L2A and L1C data in GEE

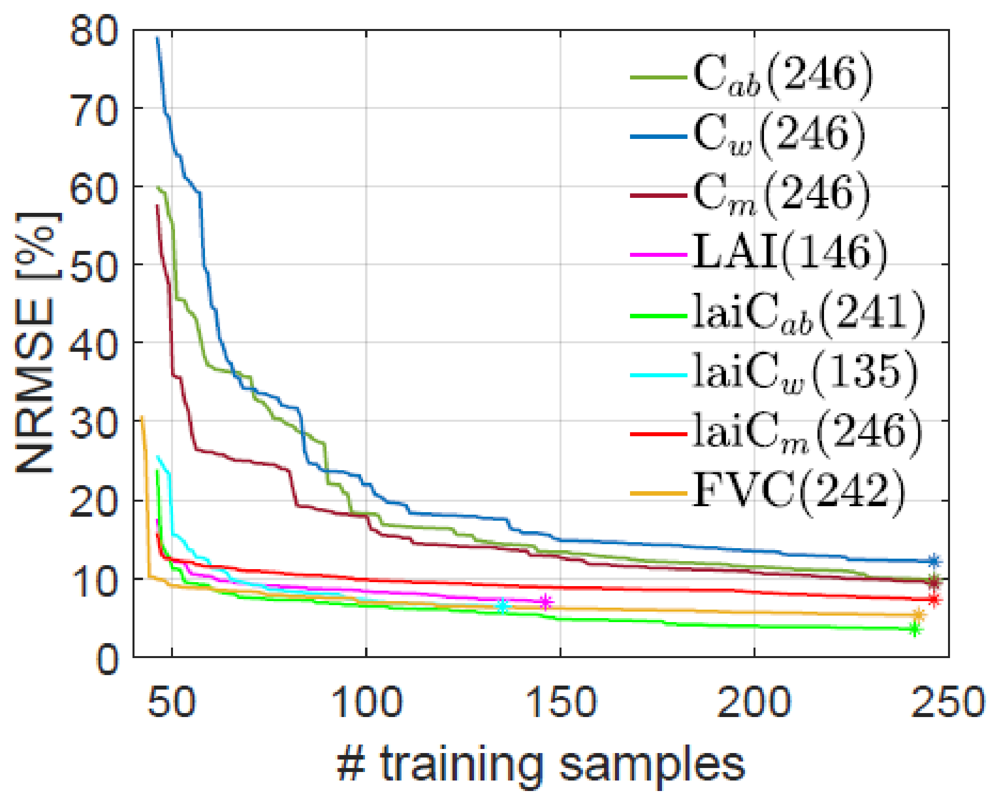


Estévez, J.; Salinero-Delgado, M. Berger, K.; Pipia, L.; Rivera-Caicedo, J.P.; Wocher, M.; Verrelst, J (2021). **Gaussian processes retrieval of vegetation traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data**. submitted.

Crucial for implementing GPR into GEE is reducing its size

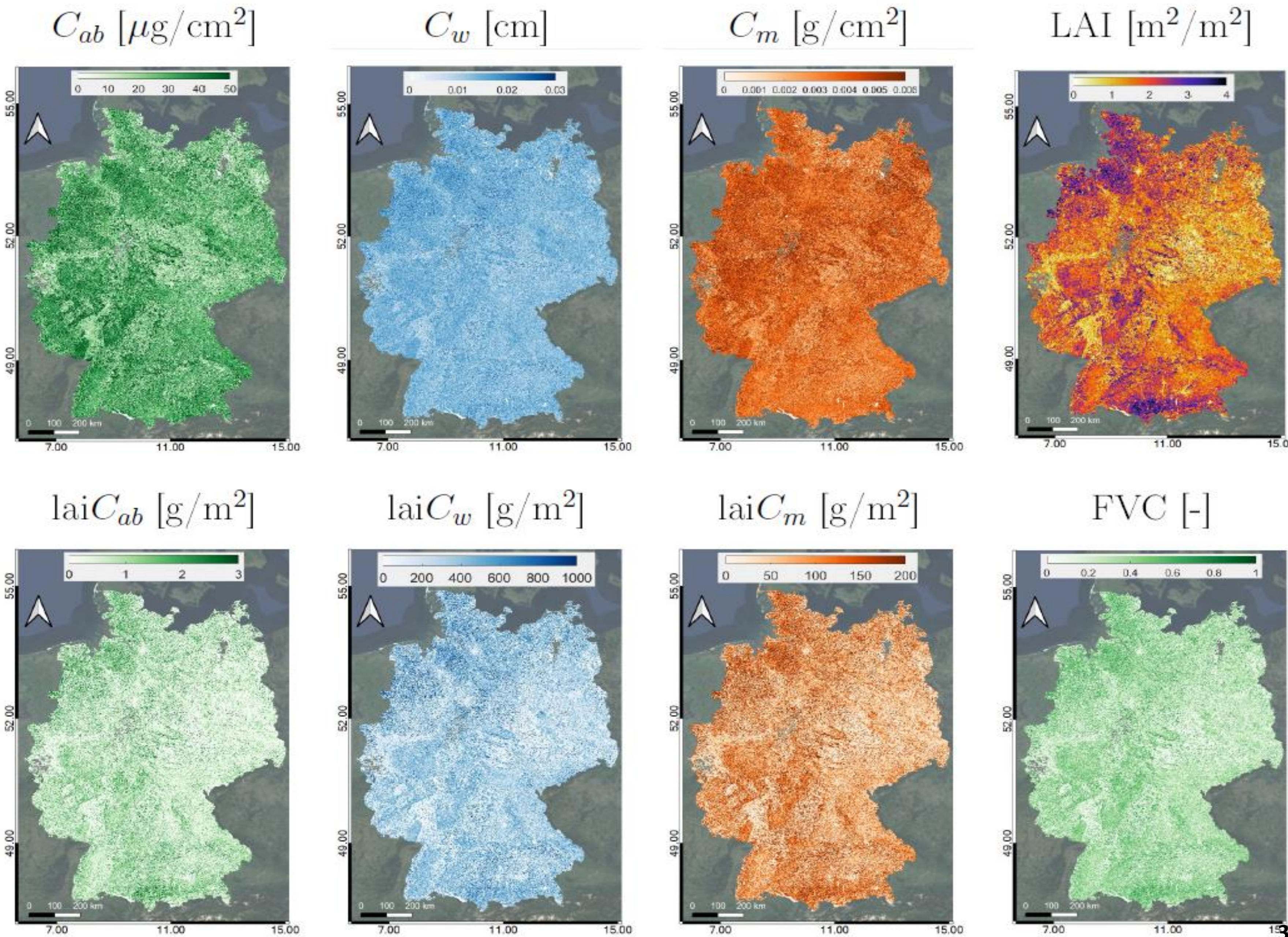
Active learning



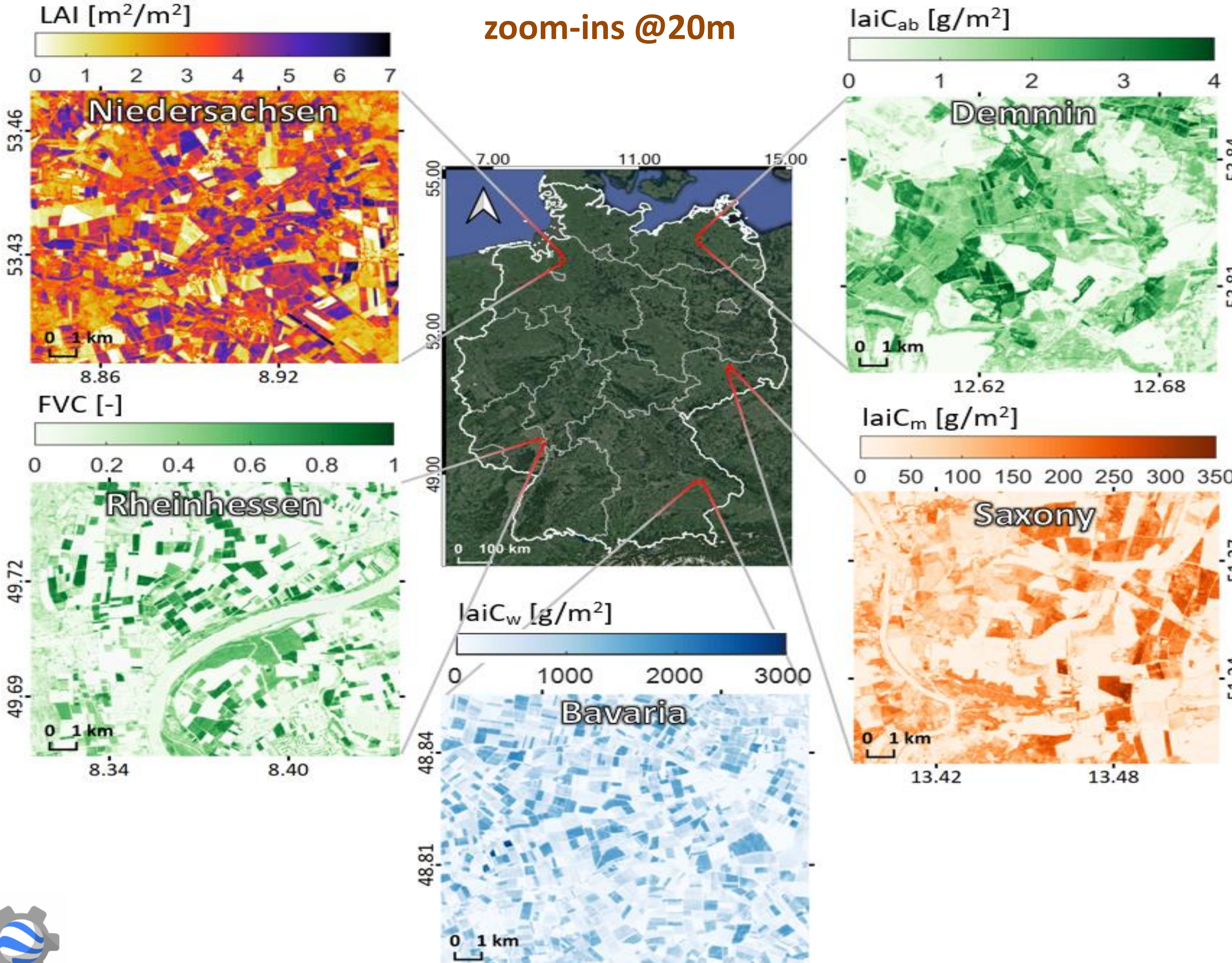


For GPR a few hundred samples enough for optimal retrieval performances: **quality more important than quantity**. Light models can then be implemented into GEE.

TOA-based retrieval from S2 L1C data: Germany



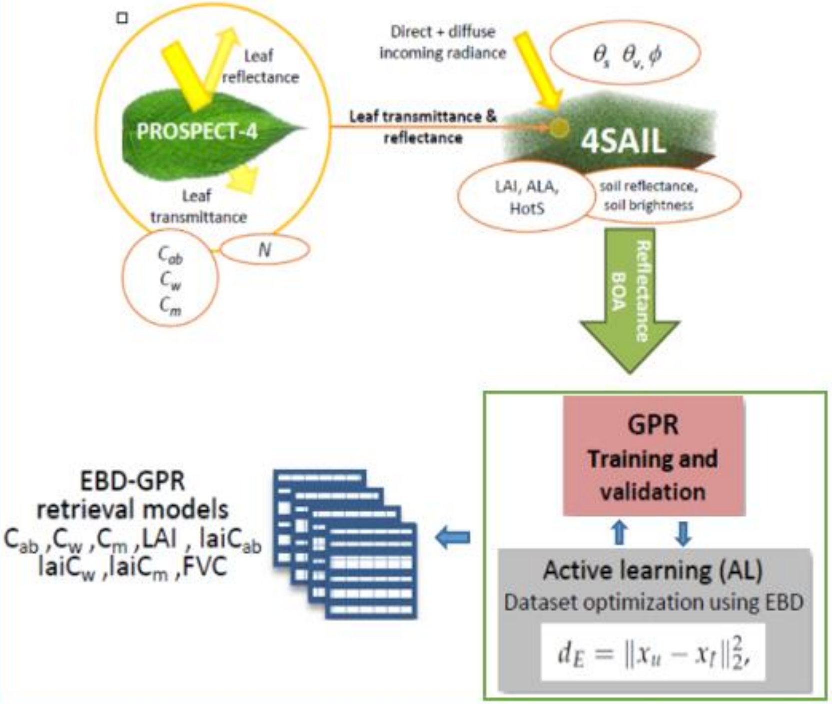
zoom-ins @20m


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

S2 time series processing: gap filling and phenology indicators mapping

ARTMO

Hybrid retrieval models



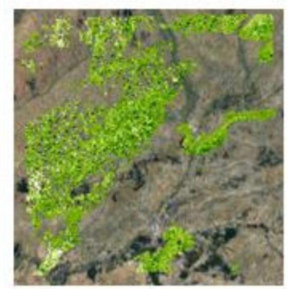
GPR Model

Google Earth Engine

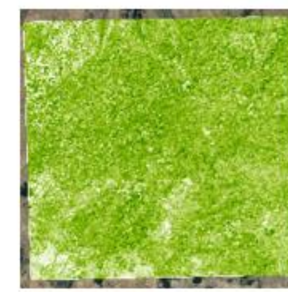
sentinel-2

GEE GPR implementation

Predicted mean



Gapfilling



Pixel-wise time series analysis

Spatio-temporal Gapfilling

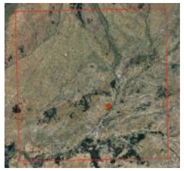
- GPR-based gapfilling strategy based on pre-optimized kernel hyperparameters.

Length-scale l | Signal variance σ_f^2 | Noise variance σ_n^2

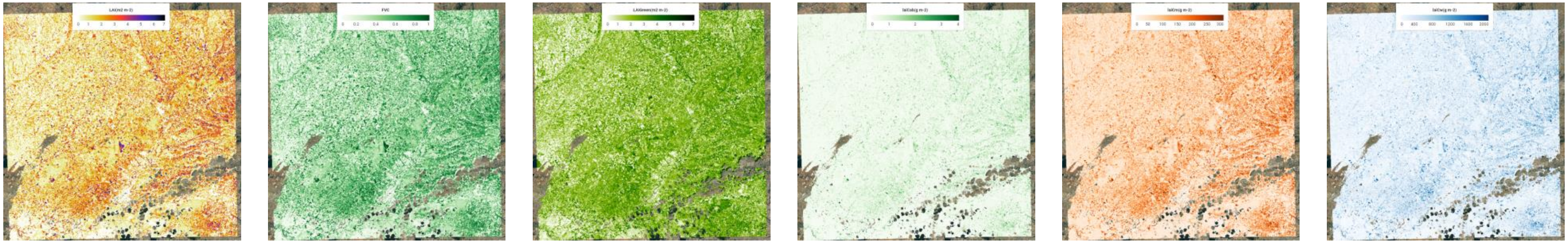
DATimeS

Gapfilling hyperparameters

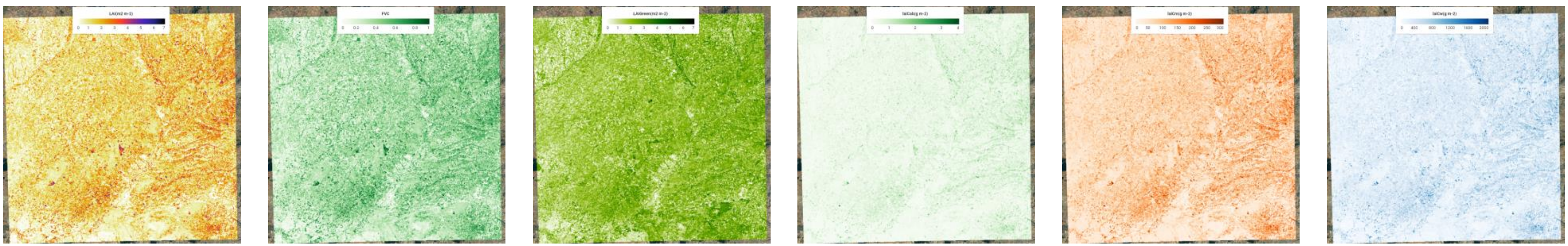
GEE TS gap-filling processing

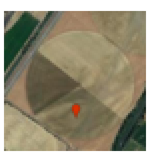


LAI | FVC | LAIgreen | LAICab | LAICm | LAICw

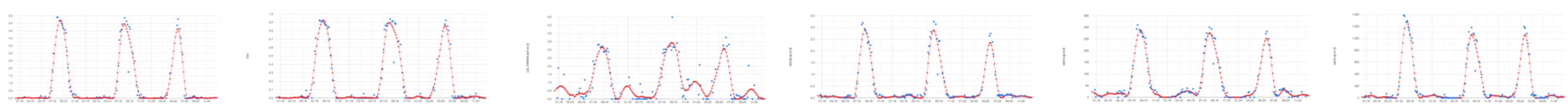


GPR gap-filling



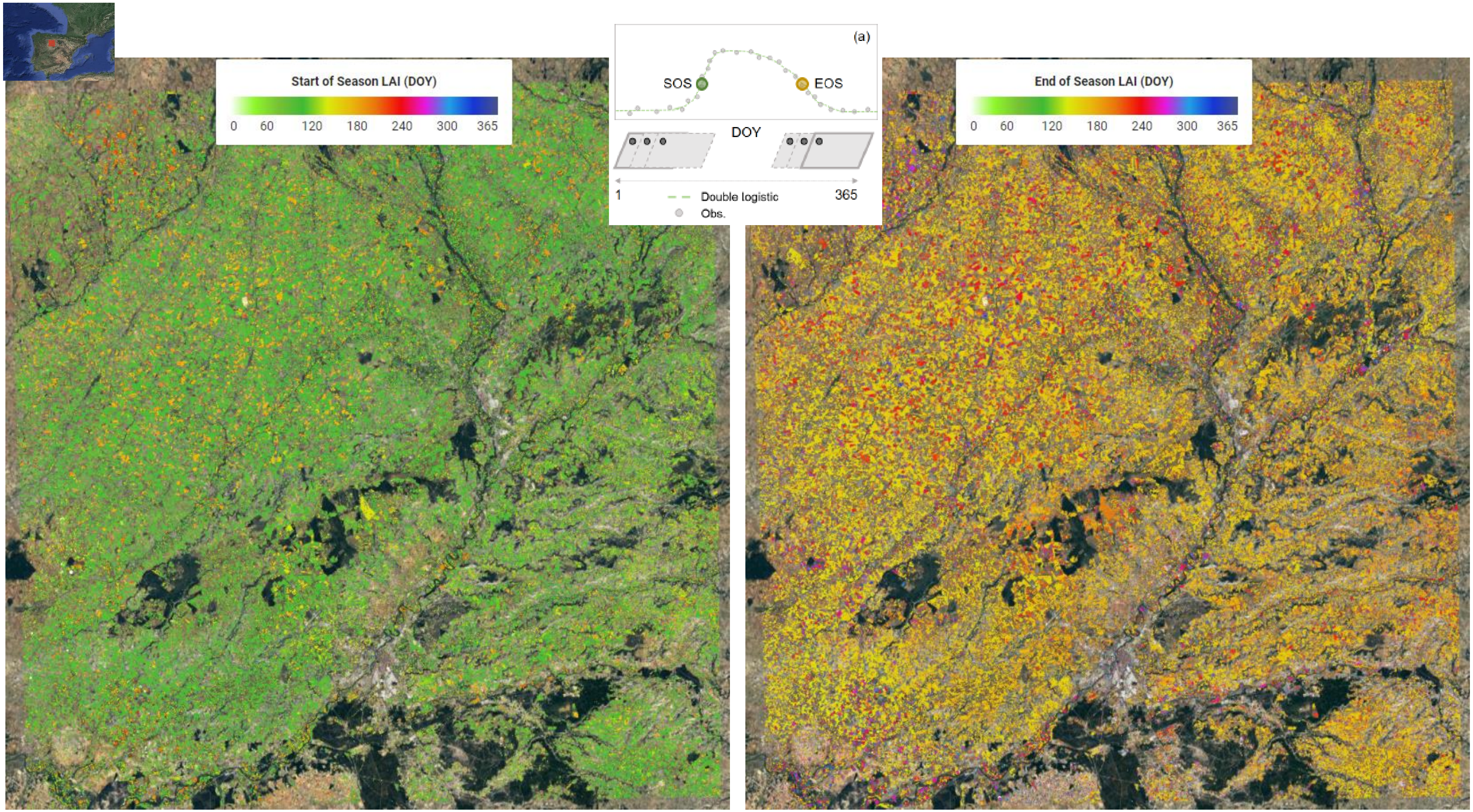


Long : -4° 36' 31.2804"
Lat : 41° 46' 26.8752"



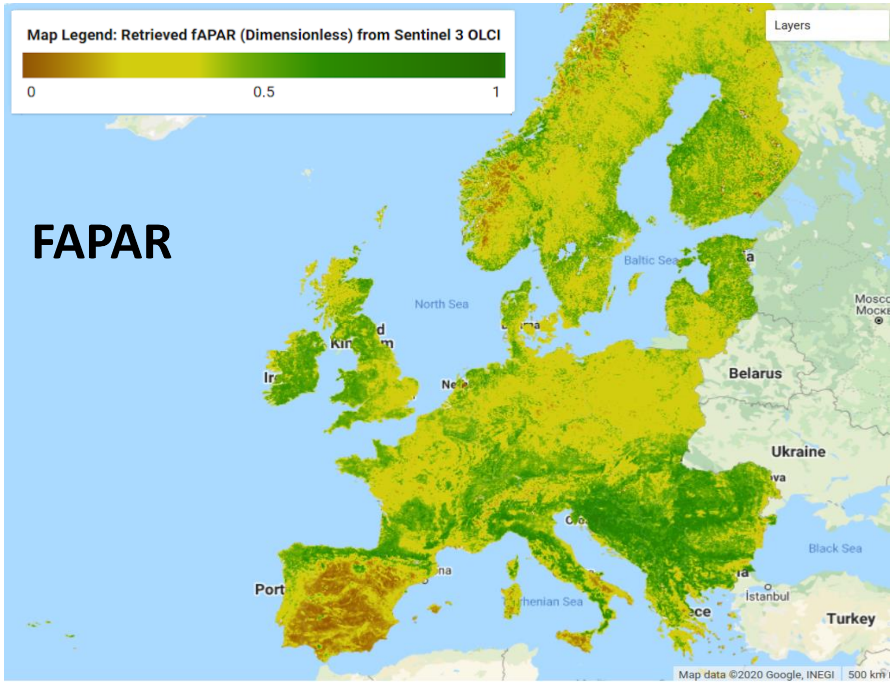
Gap-filled TS products: cloud-free vegetation products on a regular basis.

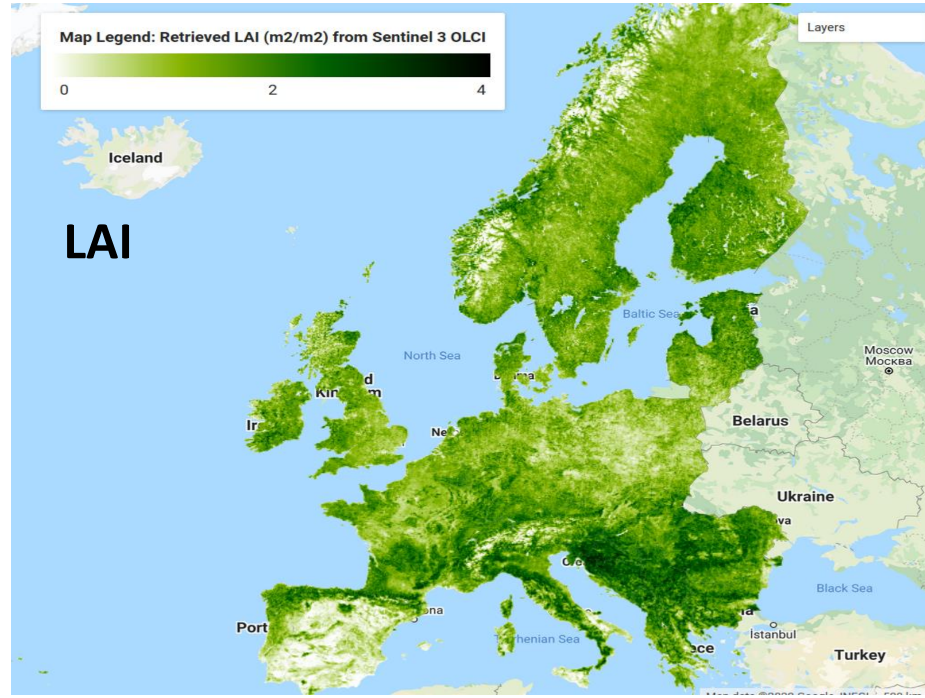
GEE calculation of S2 phenology indicators (SOS & EOS) based on vegetation products

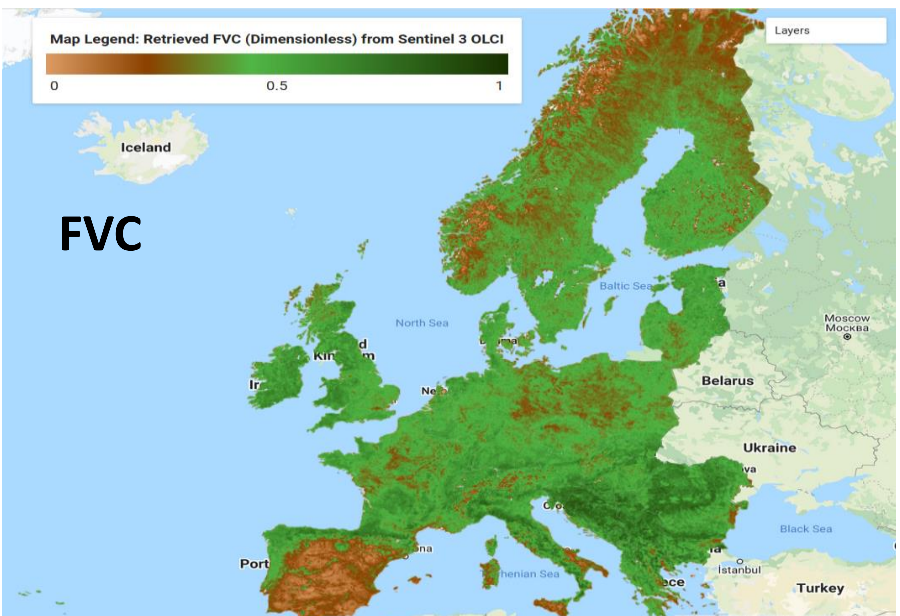


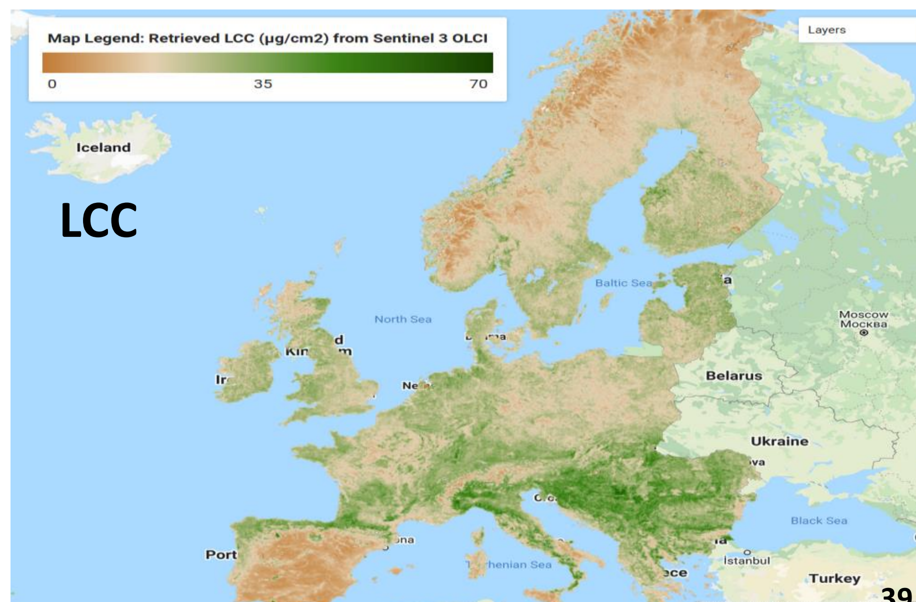
In GEE, SOS and EOS can be determined anywhere in the world

TOA retrieval from S3 OLCI L1C data: Europe









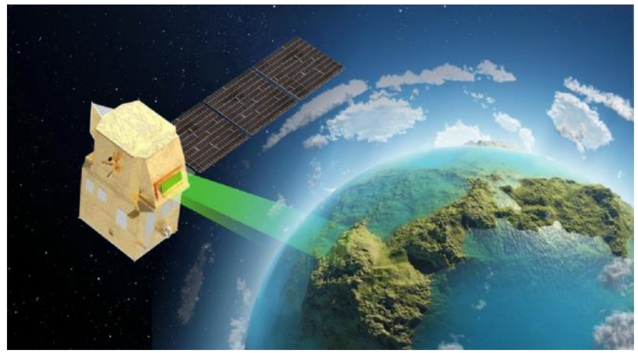
CHIME

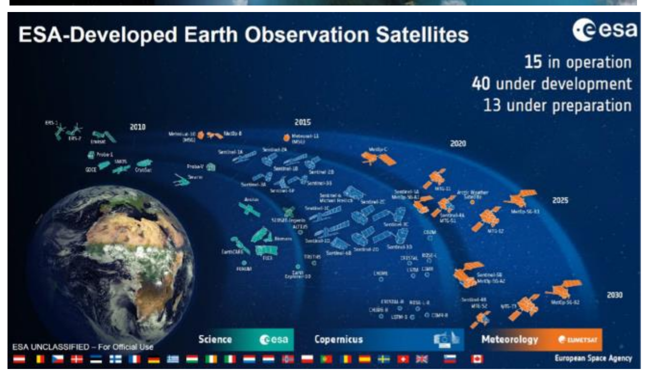
The **Copernicus Hyperspectral Imaging Mission,** CHIME, will carry a visible to shortwave infrared spectrometer to provide routine hyperspectral observations to support new and enhanced services for sustainable agricultural and biodiversity management, as well as soil property characterisation

Technical concept

Routine spectroscopic observation in contiguous spectral bands:

- Instrument: Pushbroom Imaging Spectrometer 400 – 2500 nm, $\Delta\lambda$ <= 10nm
- Revisit 10 – 15 days
- GSD (spatial resolution) 20 – 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:

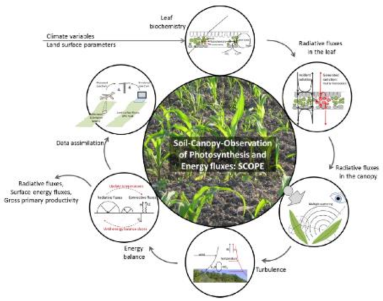
The mission shall provide access to Level-1B, Level-1C and Level-2A products accessible via DIAS and with API support:

- Bottom-of-Atmosphere (BOA) reflectance (atmospheric 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

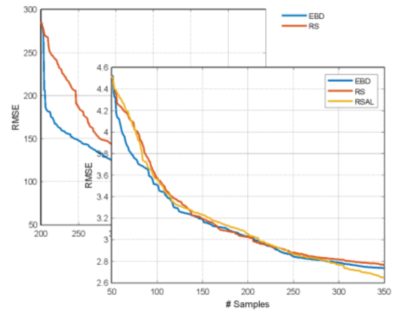
Hybrid workflow CHIME vegetation models

RTMs



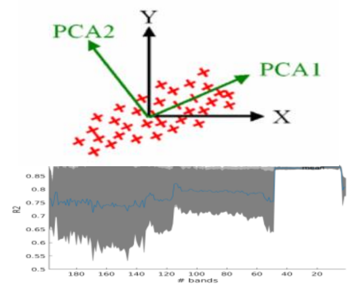
Use **RTMs** (e.g. SCOPE, PROSPECT DyN – SAIL) to generate a LUT composed by pairs (e.g. 1000) of **vegetation parameters and spectra.**

Active learning



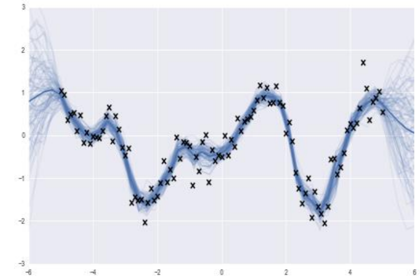
Select the **most representative samples** from the LUT via a **diversity** or entropy criteria. Later, add non-vegetated spectra.

PCA or Band Rank Analysis



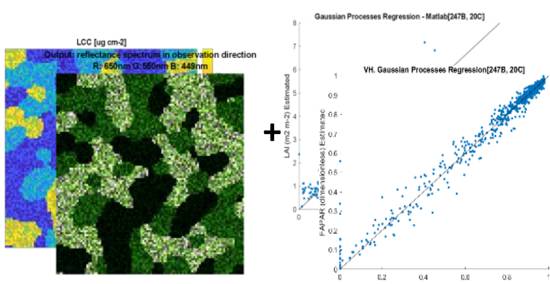
Dimensionality reduction with **PCA** (20 components) or **band selection** based on GSA and kernel sensitivity ranking.

Train GPR algorithms



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

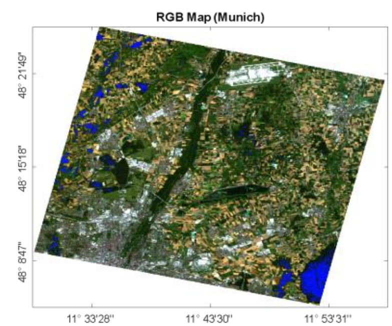
Validate the models



Okay?

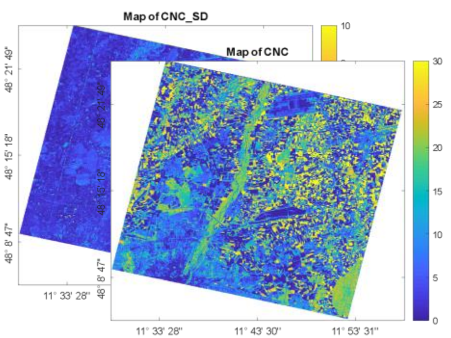
Assess models' performance against **field data** and vegetation **reference scenes.**

Apply to new observations



PRISMA images resampled to **CHIME** band settings

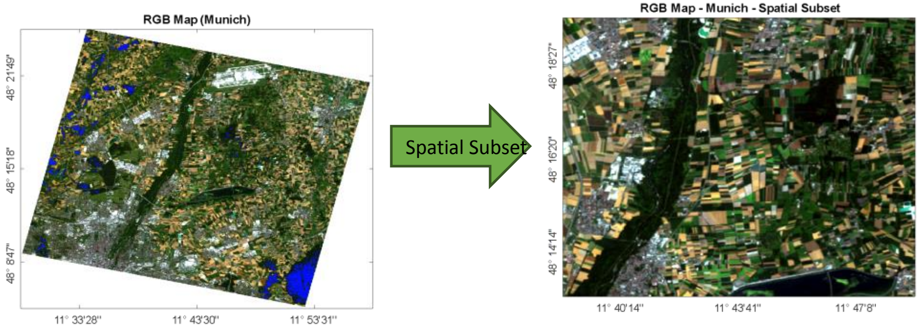
Maps + uncertainties

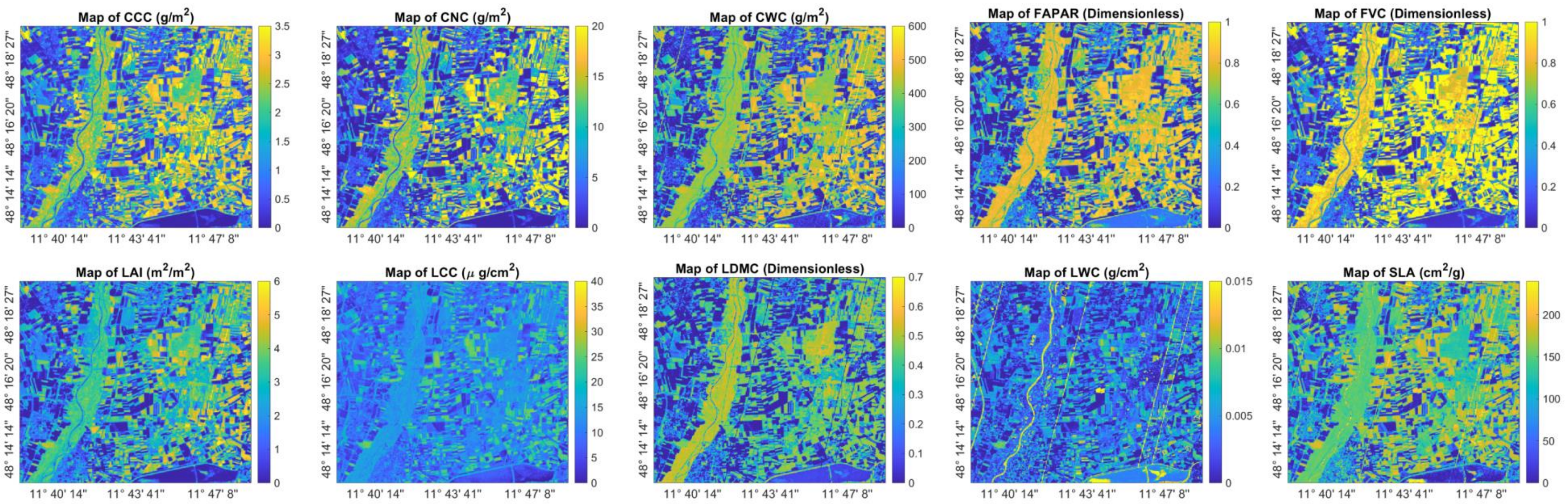


Final outputs of the workflow.

Maps: results with 20 PCA

- **PRISMA** image resampled to CHIME band settings, heterogeneous spatial subset to test the performance in vegetation, buildings and water
- Canopy variables ☺
- Some leaf variables ☹

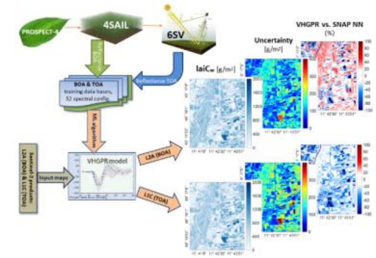




GPR hybrid models powerful for vegetation trait mapping (with inclusion of uncertainty estimates)

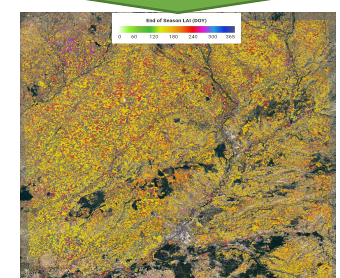
Conclusions hybrid models for retrieval

- ✓Hybrid models powerful for vegetation properties mapping: generic, adaptive, competitive, fast and provision of uncertainty estimates (GPR)
- ✓Hybrid models at both BOA and TOA scale (S2, S3)
- ✓Framework developed for running GPR models into GEE. Any GPR model can process anywhere and anytime within the GEE catalogue.
- ✓Also gap-filling processing in GEE & calculation of phenology indicators
- ✓Hybrid GPR models under development for vegetation traits retrieval from the future CHIME mission

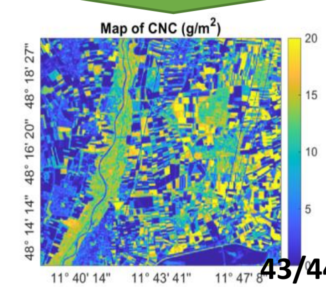


TOA Retrieval

GEE

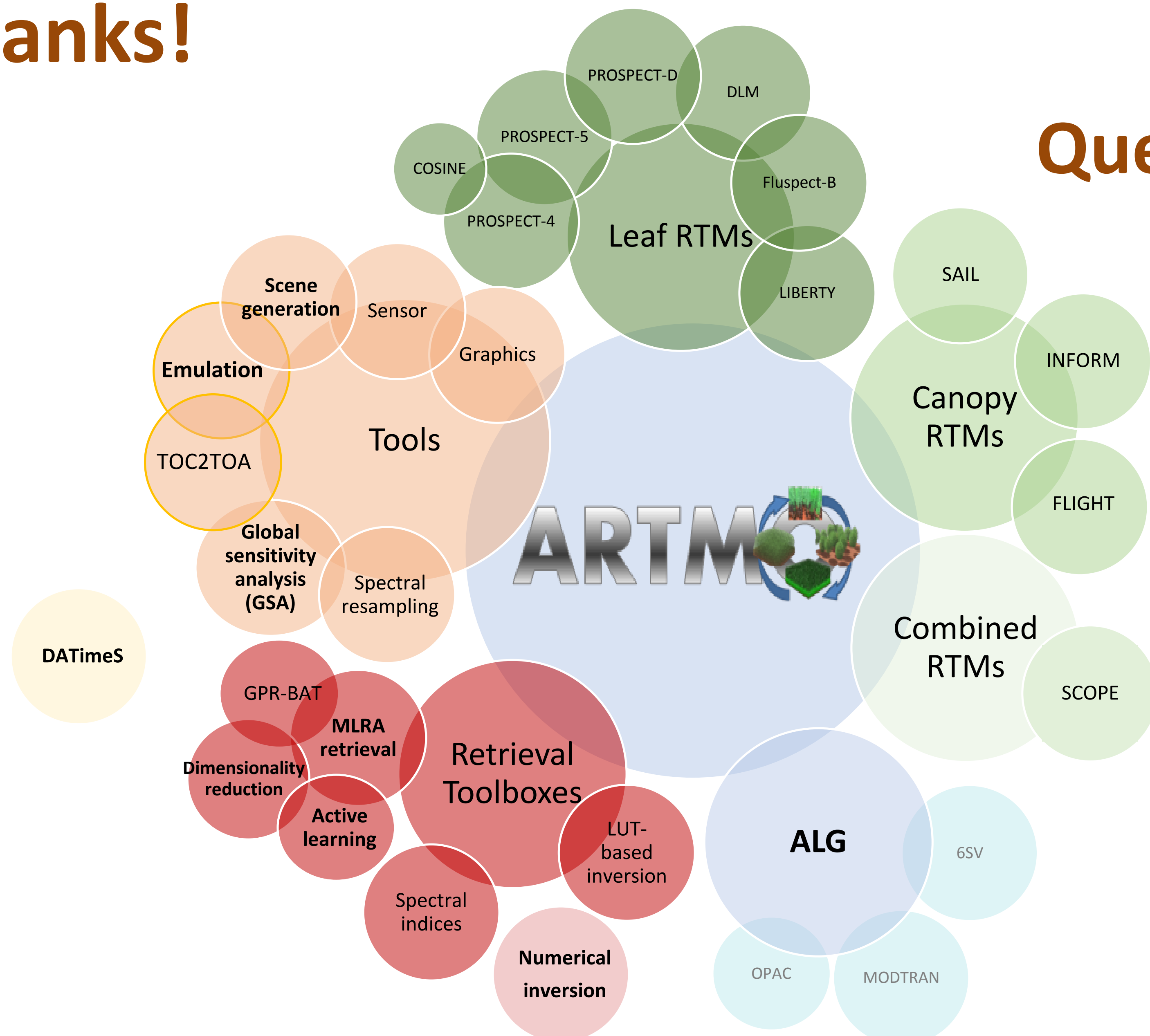


CHIME



Thanks!

Questions?



http://artmotoolbox.com
https://ipl.uv.es/sentiflex/