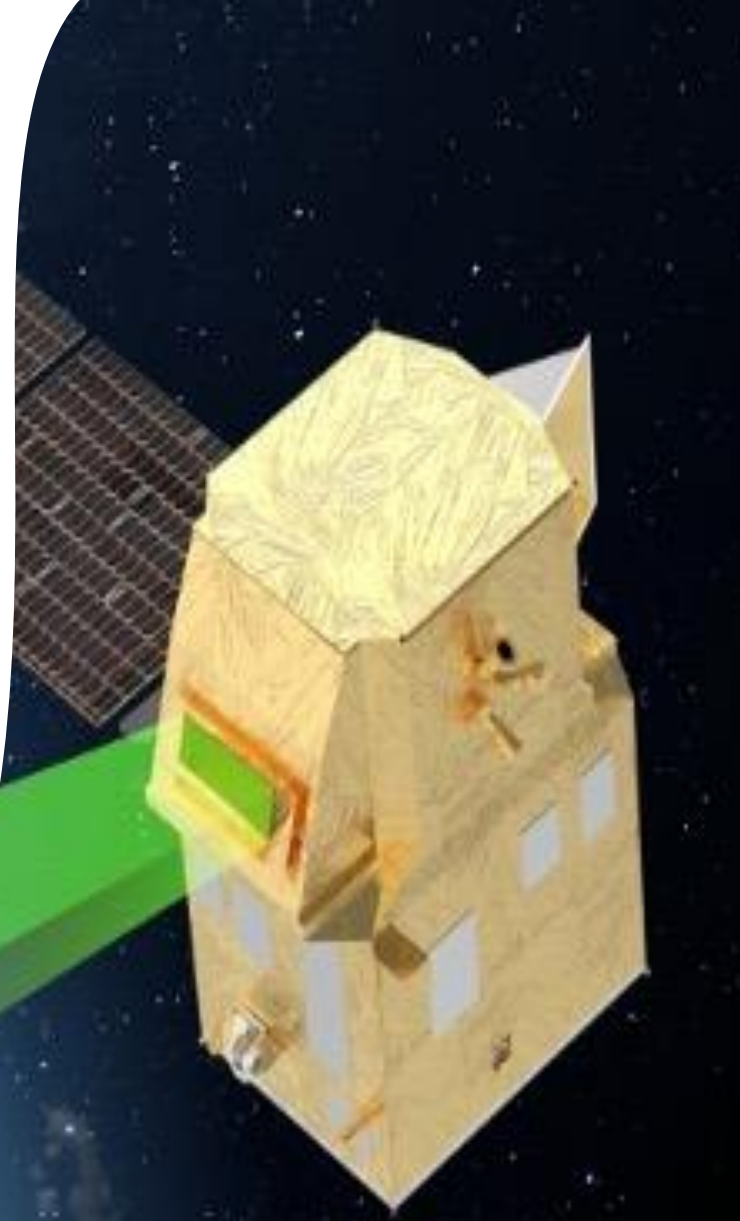
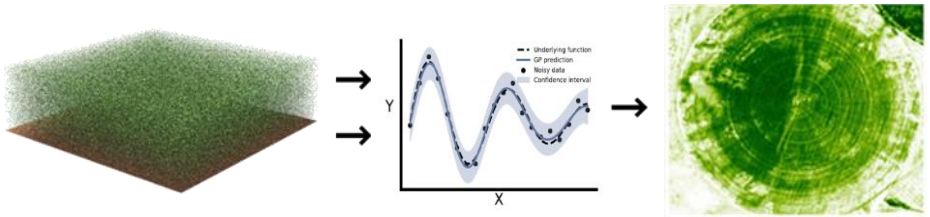
The CHIME E2E L2B Vegetation processor: updates and upcoming initiatives

Jochem Verrelst, José Luis García-Soria, Miguel Morata;
Viktor Ixion Mészáros, Louis Trinkle, David Kovacs;
Cinzia Panigada, Giulia Tagliabue,…



Outline of the presentation:

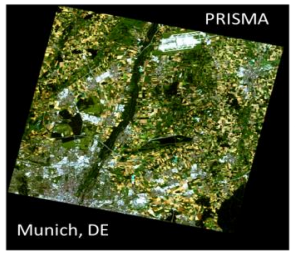
- **CHIME** & **CHEES L2BV module**
- Retrieval strategies… why **hybrid**?
- MLRAs & **uncertainties**
- How does the **L2BV module** work?
- CHIME **models evaluation** strategy and results
- **Agnostic functioning L2BV module** and **extras**
- **Conclusions**



New

New

CHEES: CHime End to End Simulator



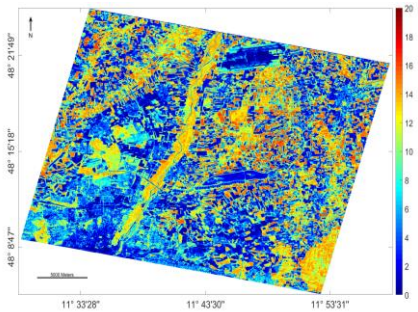
OPSI Geometry Module

Scene Generator Module

OPSI Instrument and Processing Module

L2 GPP Module

L2B Retrieval Module



Performance Assessment Module

OGM: Simulate CHIME orbit and attitude

SGM: Simulates the scene to be observed at fine spectral and spatial resolution.

OIP: Models the CHIME sensor behavior.

L2GPP: Performs the atmospheric correction

L2B: Implements the algorithms for retrieval of **bio-/geophysical variables**, objective of the mission (e.g., retrieval of mineral content/soil/**vegetation**).

PAM: Evaluates the product performance at different processing levels

CORE Data Products:

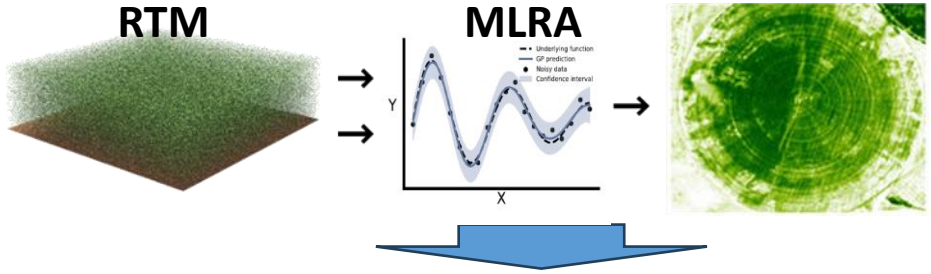
- **L1C:** TOA radiance
- **L2A:** atmospherically corrected and ortho-rectified reflectances
- **L2BM:** Mineral products
- **L2BS:** Soil products
- **L2BV:** Vegetation products

Priority

- Canopy nitrogen content
- Canopy chlorophyll content
- Canopy water content
- Leaf area index
- Leaf nitrogen content
- Leaf mass area
- Leaf chlorophyll content
- Leaf water content

Key capabilities retrieval models in **operational** context:

Hybrid methods:



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

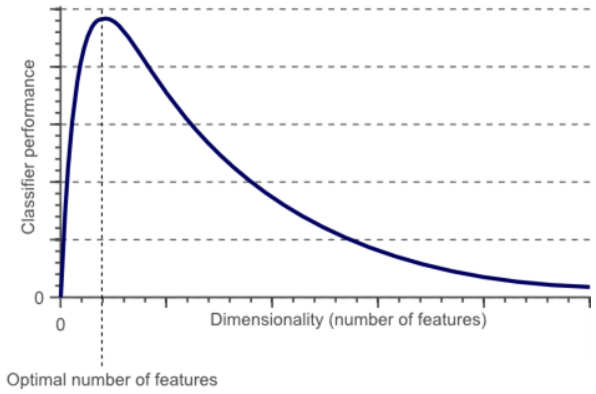
* Some machine learning methods (e.g. probabilistic or ensemble methods)

Verrelst, J. et al. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273-290.
Verrelst, J. et al. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods." *Surveys in Geophysics* 40, 589-629.

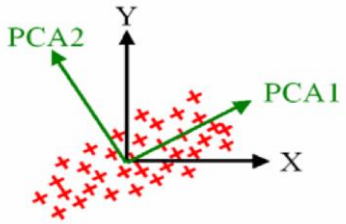
The training data matters in (hyperspectral) hybrid models…

Spectral domain…

Curse of dimensionality



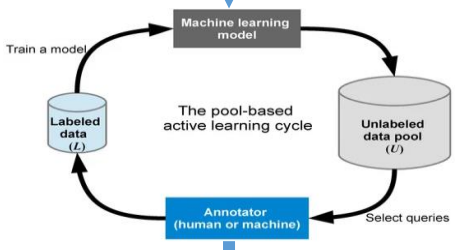
Principal Components analysis



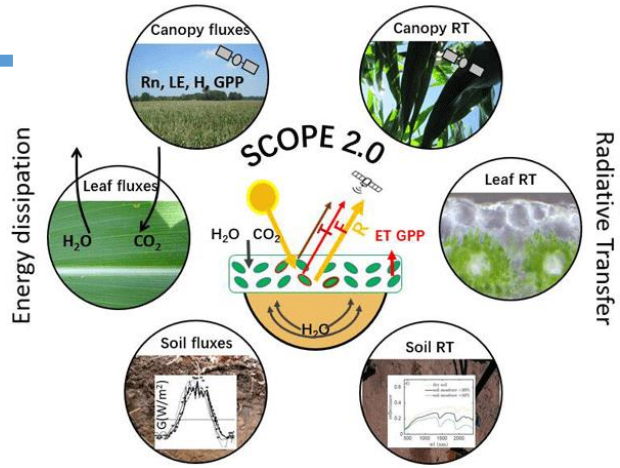
A few PCs (e.g., 20)

Sampling domain…

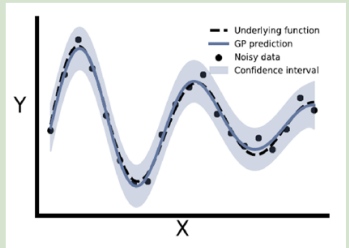
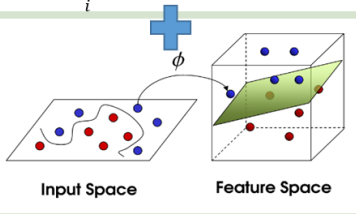
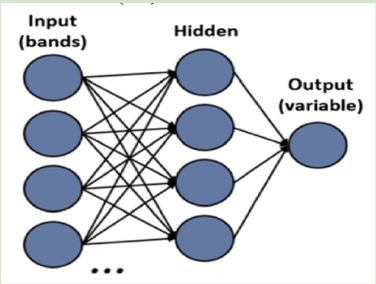
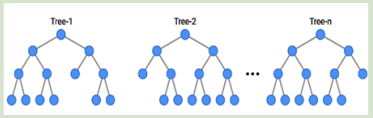
Active Learning



Optimal samples + randomness



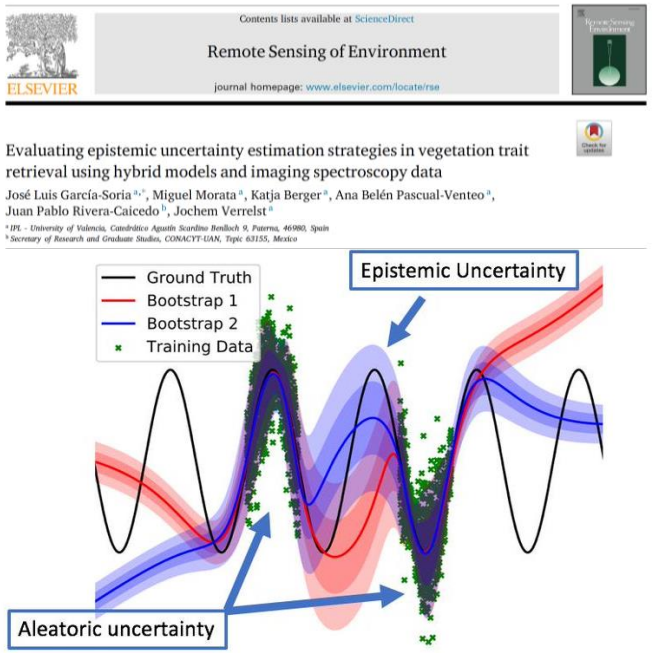
Which MLRA to use?

Gaussian Processes Regression (GPR)	Kernel Ridge Regression (KRR)	Neural Network (NN)	Random Forest Regression (RFR)
-Bayesian method. -Outputs mean and variance.	-Ridge regression in a higher dimensional space.	-Artifical neural network.	-Ensemble of decision trees.
	$C = \frac{1}{2}\sum_i (y_i - w^T x_i)^2 + \frac{1}{2}\tau\|w\|^2$ 		

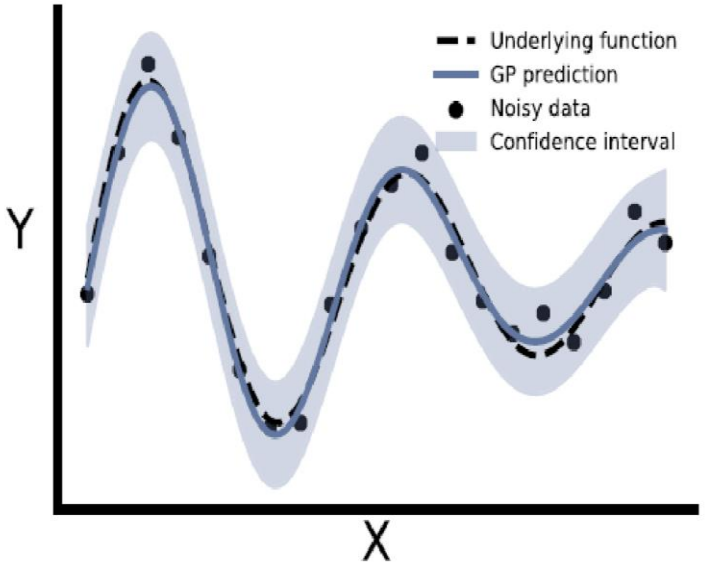
Approaches to epistemic uncertainty…

- **Aleatory uncertainty refers to the inherent uncertainty due to the probabilistic variability.** This type of uncertainty is irreducible in that there will always be variability in the underlying variables.
- **Epistemic uncertainty - is uncertainty that comes from the lack of knowledge**. This lack of knowledge comes from **many sources**. **This type of uncertainty is reducible.**

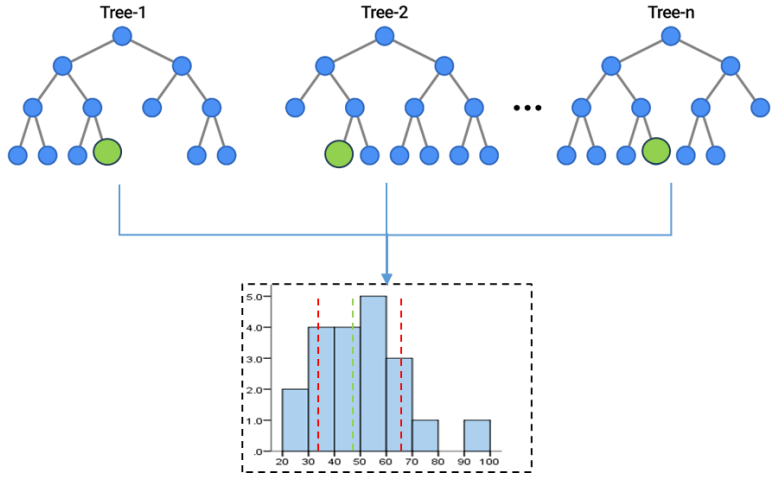
✓ **Epistemic uncertainty** is not a full budget of the uncertainty.
✓ Inform us from the weak points of our training data set, and so were to sample more.



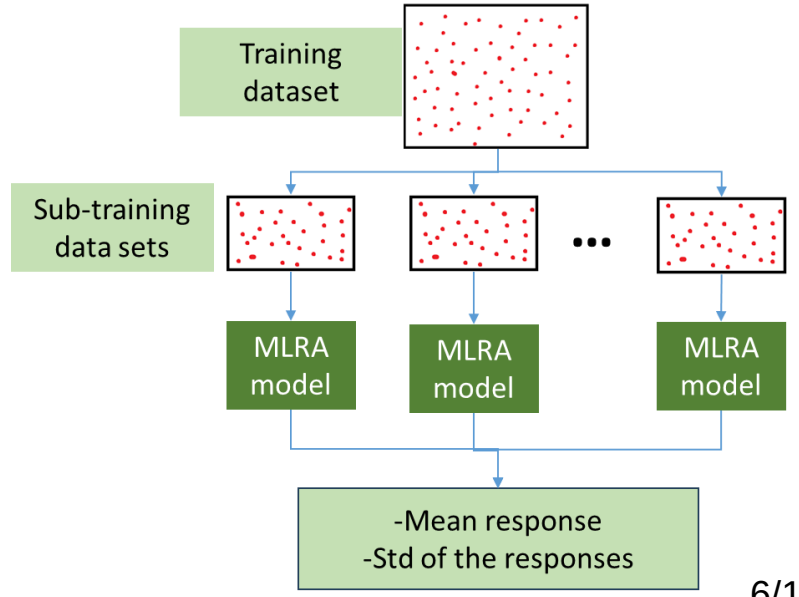
Bayesian regression: **GPR**



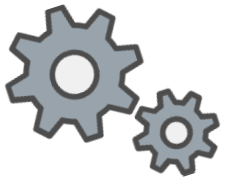
Quantile ensemble trees: **QRF**

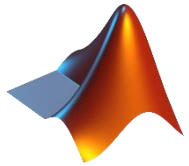


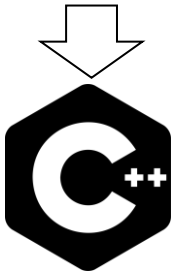
Bootstrapping: KRR, ANN

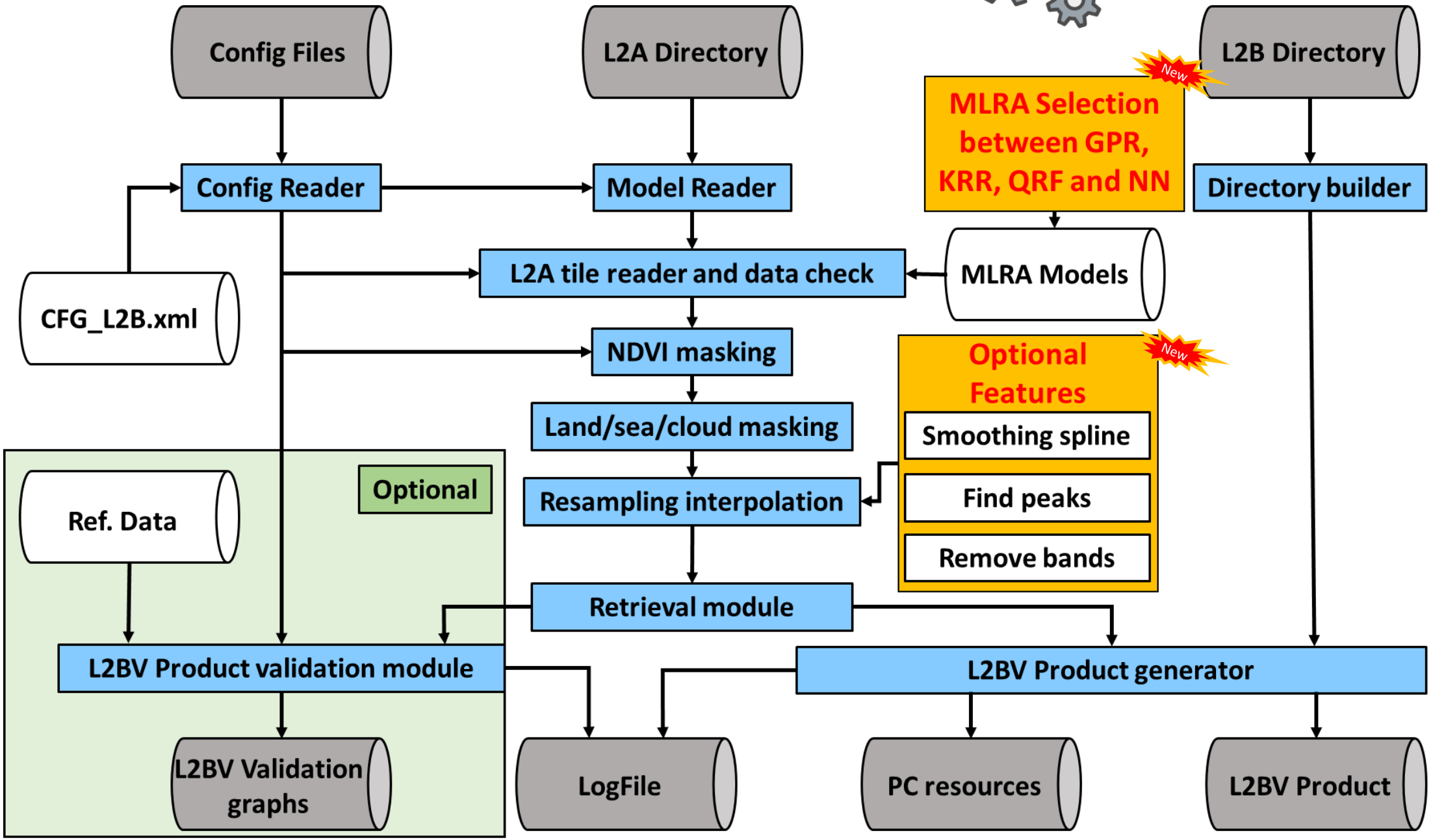


How does the L2BV module work?









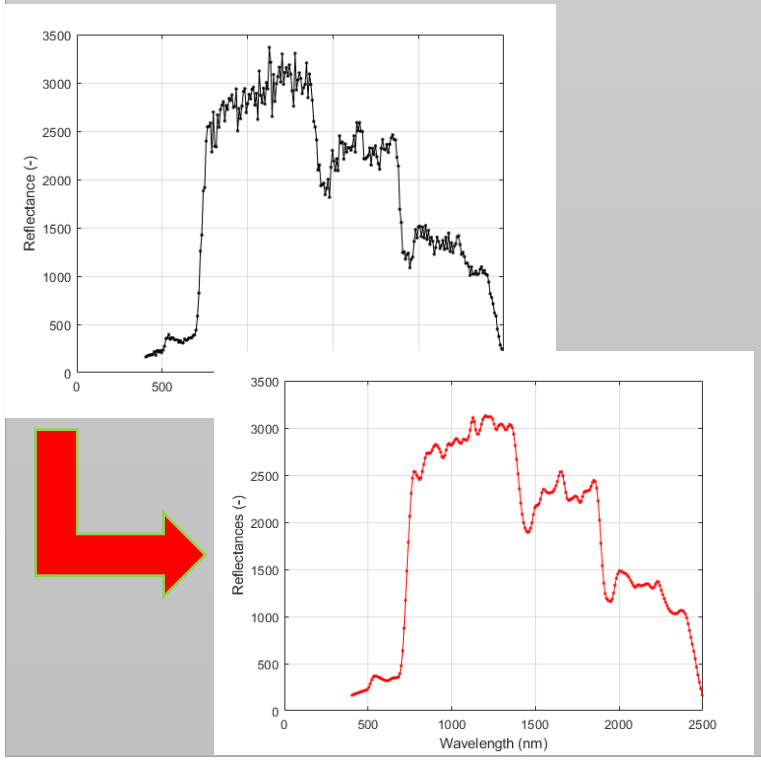
Interpolation, Smoothing, find peaks and remove bands

New

INTERPOLATION & SMOOTHING

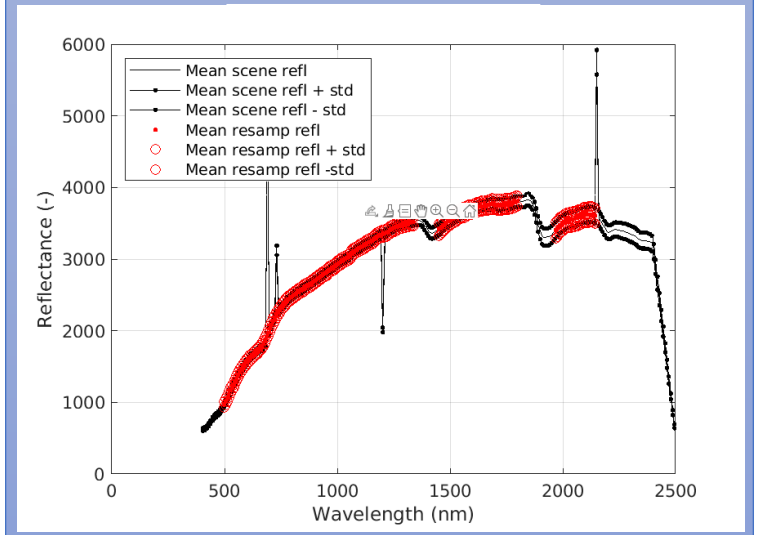
Band resampling method can change from:

- **Interpolation** (splines)
- **Smoothing** interpolation (splines)



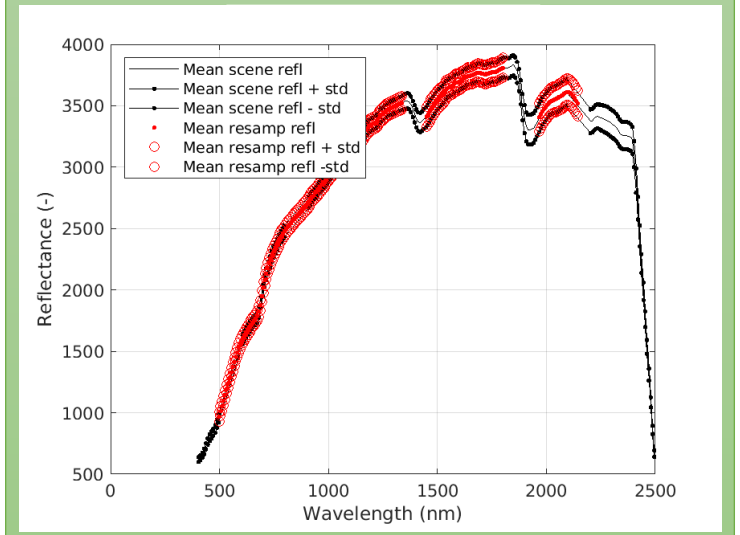
FIND PEAKS

Find peaks function can be **switch on/off**. Peaks that are above/below a **threshold of 0,018** are detected and not used for the interpolation.



REMOVE BANDS

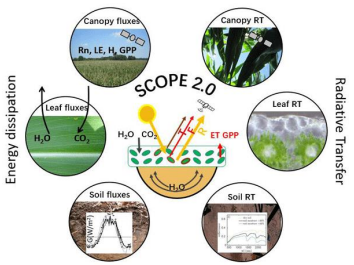
An external csv file can be used to **specify regions** of the spectra that should **be removed**.



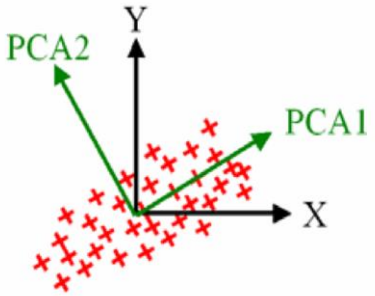
These options enable to process automatically any hyperspectral image: e.g., PRISMA, EnMAP, HyPlant,…..

Training & evaluation of vegetation models…

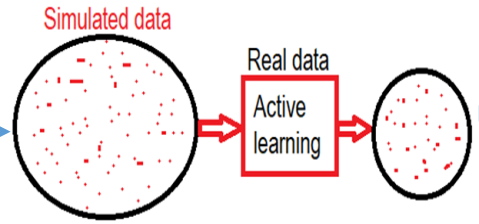
1) RTM input data



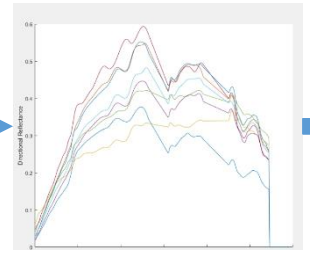
2) Spectral reduction



3) Sampling reduction



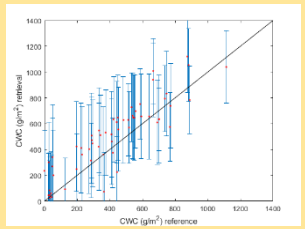
4) Adding NV spectra



Synthetic reference scene:

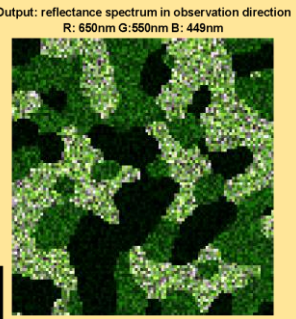
Input Variable	Acronym	low Min	low Max	Medium Min	Medium Max	High Min	High Max
Leaf Variables: PROSPECT-PRO							
Leaf Chlorophyll Content [µg/cm²]	Cab / LCC	5	20	20	50	50	80
Leaf Water Content [g/cm²]	Cw / LWC	0.003	0.005	0.005	0.015	0.015	0.03
Leaf Dry Matter Content [g/cm²]	Cdm /LDMC	0.0005	0.002	0.002	0.006	0.006	0.01
Leaf Protein Content [g/cm²]	Cp	0.0001	0.001	0.001	0.0015	0.0015	0.002
Leaf mesophyll structure parameter	N	1	2.7	1	2.7	1	2.7
Leaf Carbon Based Constituents [g/cm²]	CBC	0.0004	0.001	0.001	0.0045	0.0045	0.0080
Leaf Carotenoids Content [µg/cm²]	Cxc	0	20	0	20	0	20
Canopy Variables: 4SAIL							
Leaf Area Index [m²/m²]	LAI	0.1	1.5	1.5	4.0	4.0	8
Leaf Inclination determination a [-]	LIDFa	-1	1	-1	1	-1	1
Leaf Inclination determination b [-]	LIDFb	-1	1	-1	1	-1	1
Soil Variables: BSM							
Soil Moisture Content [%]	SMC	5	55	5	55	5	55
Model parameter for Soil Brightness [-]	B	0	0.9	0	0.9	0	0.9
Model parameter 'lat' [-]	BSMlat	20	40	20	40	20	40
Model parameter 'lon' [-]	BSMlon	45	65	45	65	45	65
Illumination and Observation Angles							
Relative Azimuth Angle [deg]	RAA	0	180	0	180	0	180
Observer Zenith Angle [deg]	OZA	0	25	0	25	0	25
Sun Zenith Angle [deg]	SZA	0	80	0	80	0	80

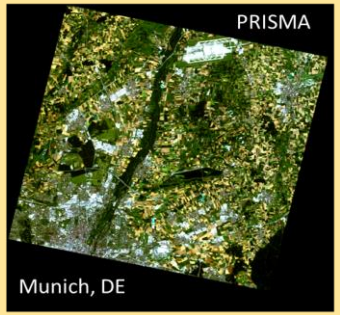
6) Evaluation



1) Experimental

2) Synthetic scene





3) Satellite imagery: PRISMA, EnMAP

5) MLRAs

- Gaussian Processess
- Kernel Ridge Regression
- Neural Networks
- Quantile regression forests

ARTMO

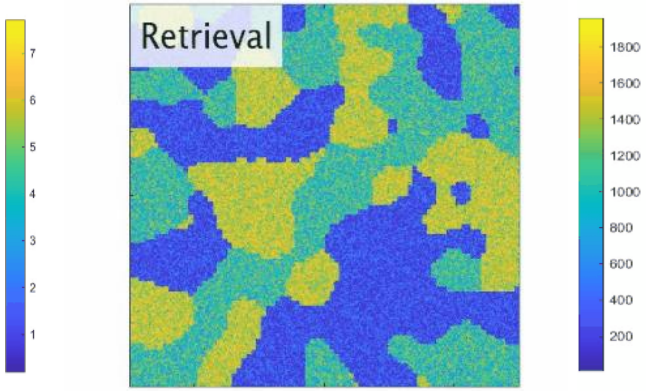
Experimental validation:

- **Canopy nitrogen content**
- Canopy chlorophyll content
- Canopy water content
- **Leaf area index**
- **Leaf nitrogen content**
- **Leaf mass area**
- Leaf chlorophyll content
- Leaf water content

Canopy

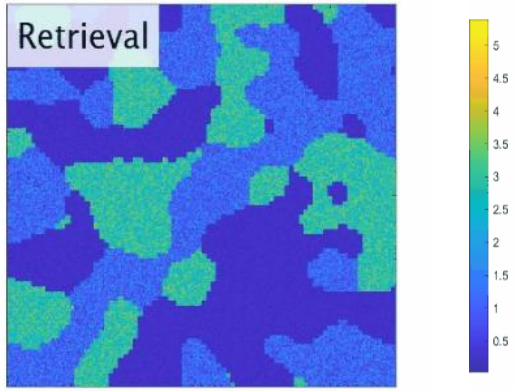
LAI (m^2m^{-2}) *GPR*

R^2 = 0,78; NRMSE = 15,4%



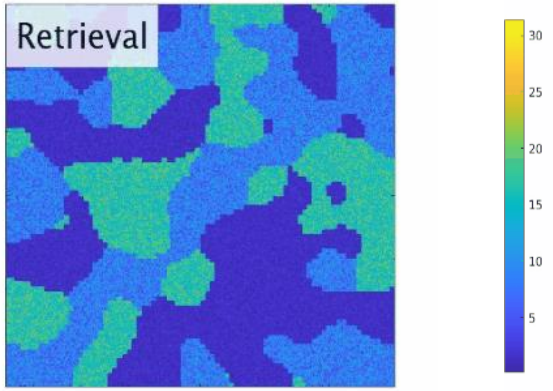
CWC ($g\ m^{-2}$) *GPR*

R^2 = 0,89; NRMSE = 10,2 %



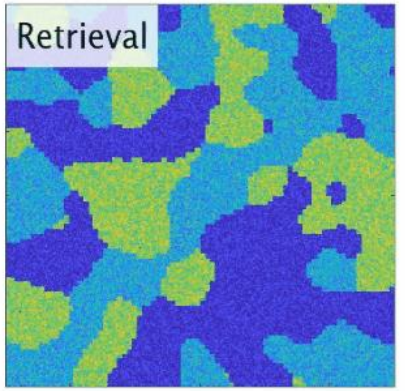
CCC (gm^{-2}) *GPR*

R^2 = 0,79; NRMSE = 12,9 %



CNC ($g\ m^{-2}$) *GPR*

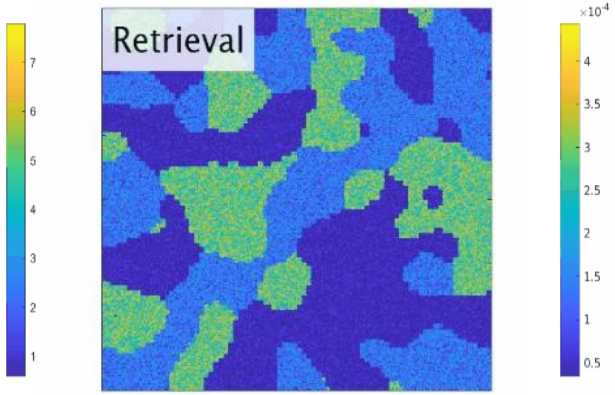
R^2 = 0,82; NRMSE = 12,5 %



Leaf

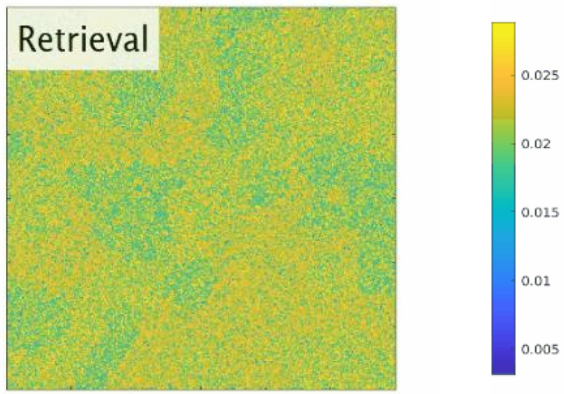
LCC ($\mu g\ cm^{-2}$) *GPR*

R^2 = 0,86; NRMSE = 13,2 %



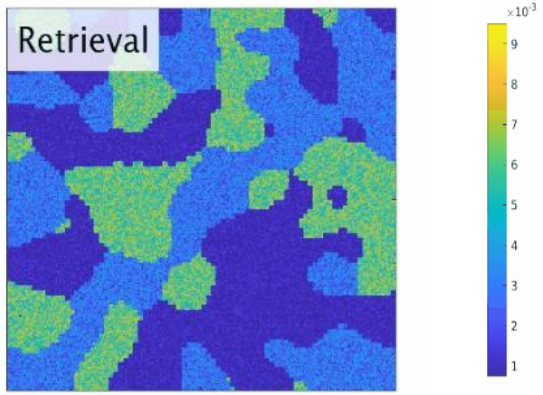
LNC ($g\ cm^{-2}$) *KRR*

R^2 = 0,00; NRMSE = 35,0 %



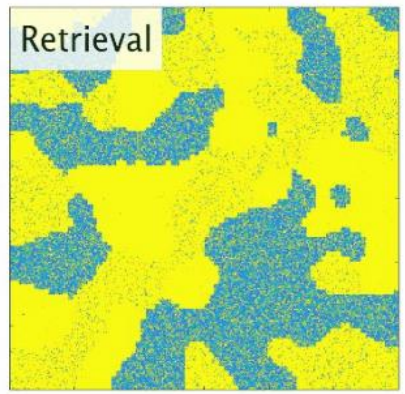
LWC ($g\ cm^{-2}$) *GPR*

R^2 = 0,94; NRMSE = 10,0 %



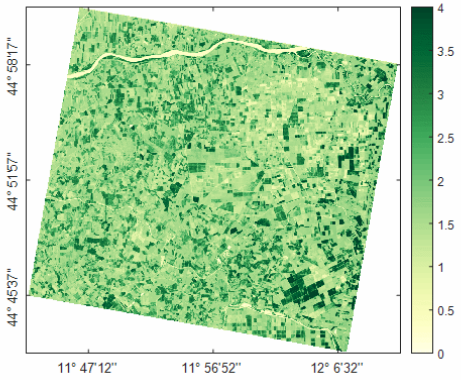
LMA ($g\ cm^{-2}$) *ANN*

R^2 = 0,51; NRMSE = 56,0 %

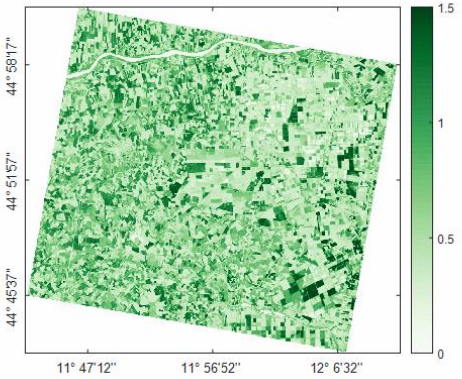


L2BV Retrievals: PRISMA image

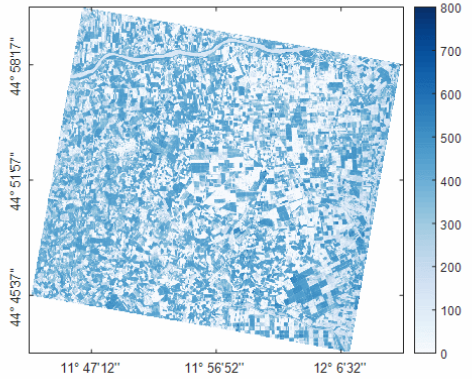
LAI (m^2/m^2) Estimation



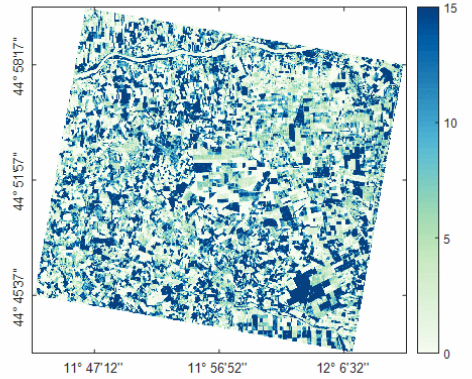
CCC (g/m^2) Estimation



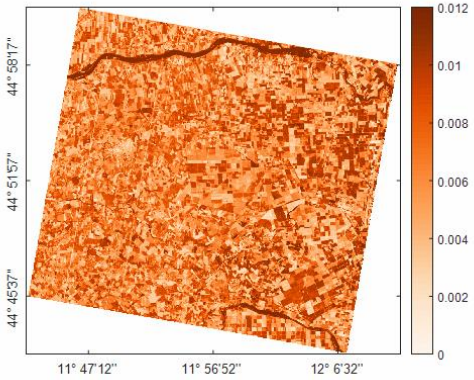
CWC (g/m^2) Estimation



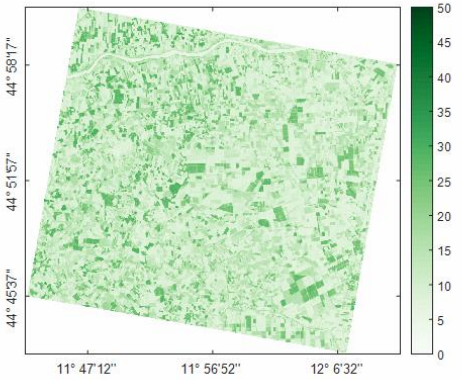
CNC (g/m^2) Estimation



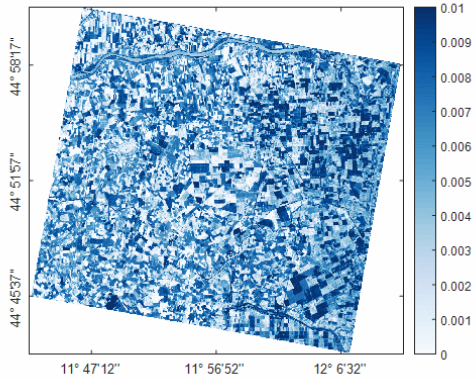
LMA (g/cm^2) Estimation



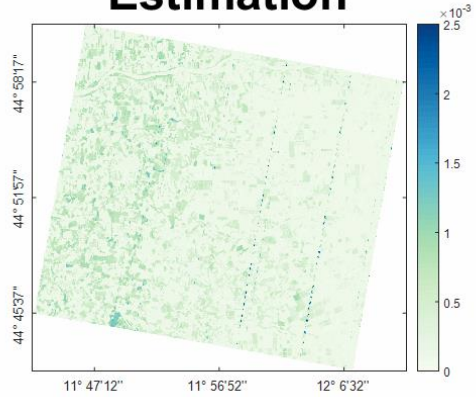
LCC (ug/cm^2) Estimation



LWC (g/cm^2) Estimation



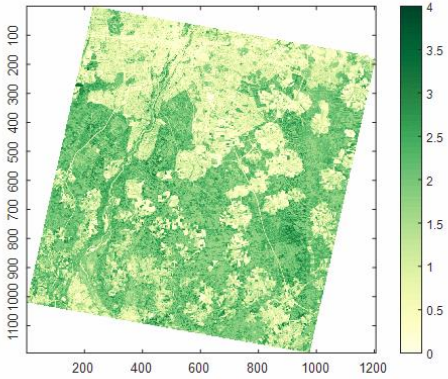
LNC (g/cm^2) Estimation



L2BV Retrievals: EnMAP image

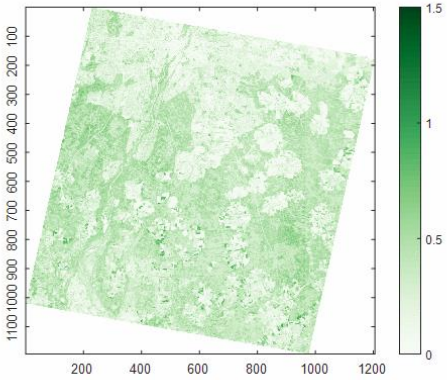
LAI (m^2/m^2)

Estimation



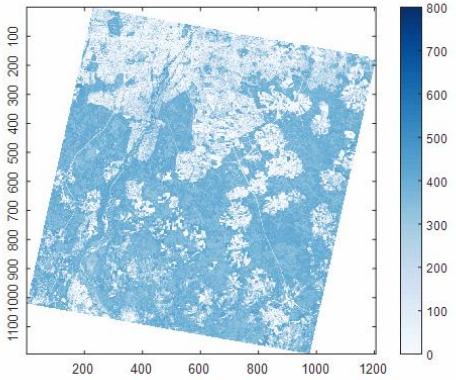
CCC (g/m^2)

Estimation



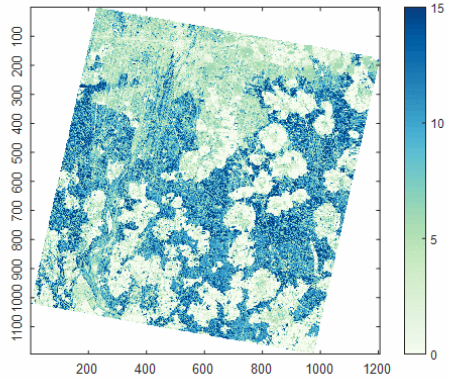
CWC (g/m^2)

Estimation



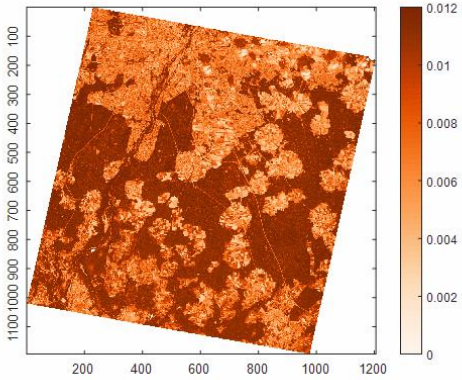
CNC (g/m^2)

Estimation



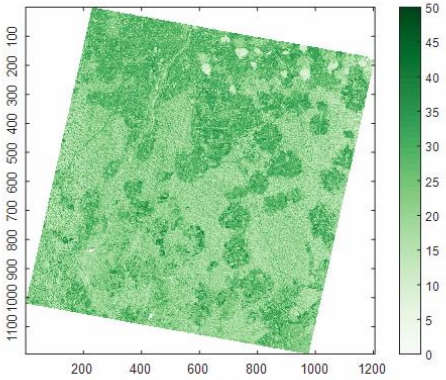
LMA (g/cm^2)

Estimation



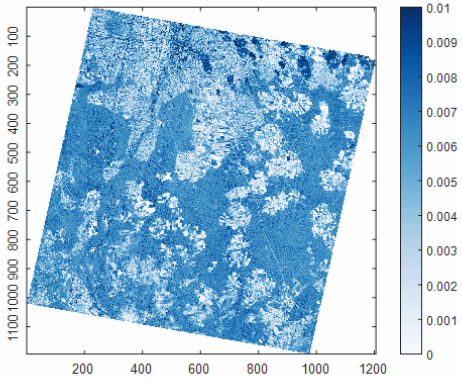
LCC (ug/cm^2)

Estimation



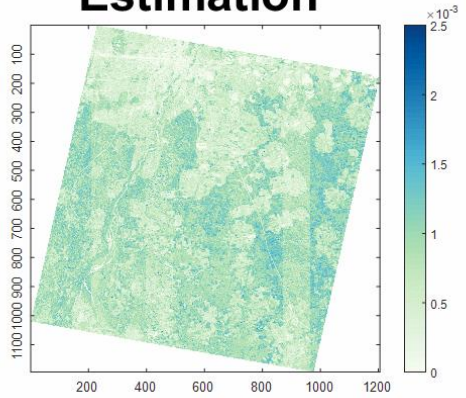
LWC (g/cm^2)

Estimation



LNC (g/cm^2)

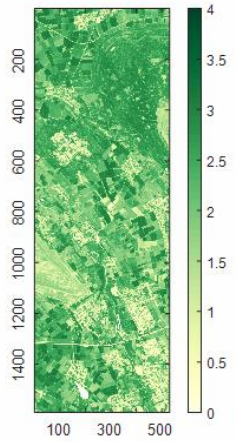
Estimation



L2BV Retrievals : HyPlant image

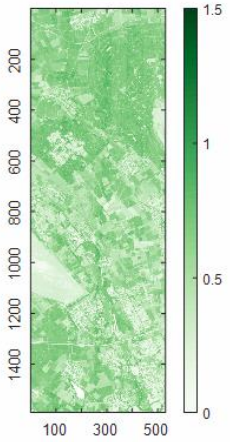
LAI (m^2/m^2)

Estimation



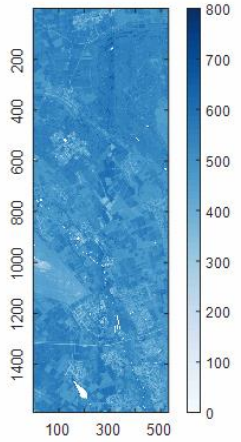
CCC (g/m^2)

Estimation



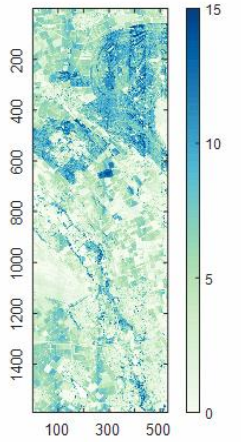
CWC (g/m^2)

Estimation



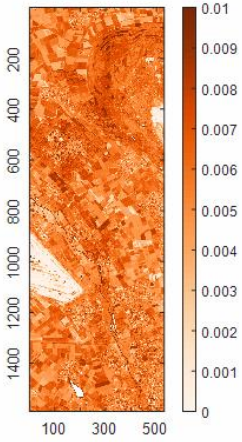
CNC (g/m^2)

Estimation



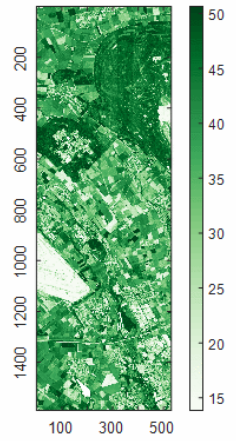
LMA (g/cm^2)

Estimation



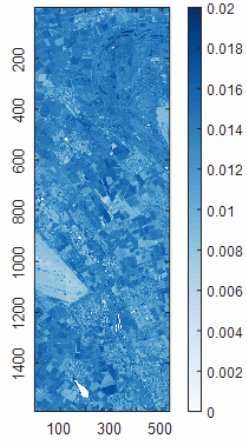
LCC (ug/cm^2)

Estimation



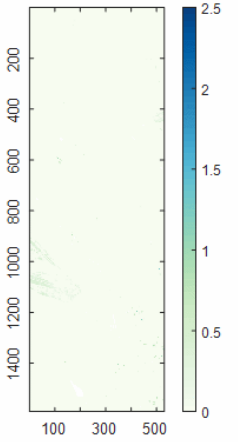
LWC (g/cm^2)

Estimation



LNC (g/cm^2)

Estimation



Extra: open-source Python implementation…

New

goal: Create a Python pipeline to run the CHIME (GPR) models: **PyL2BV**

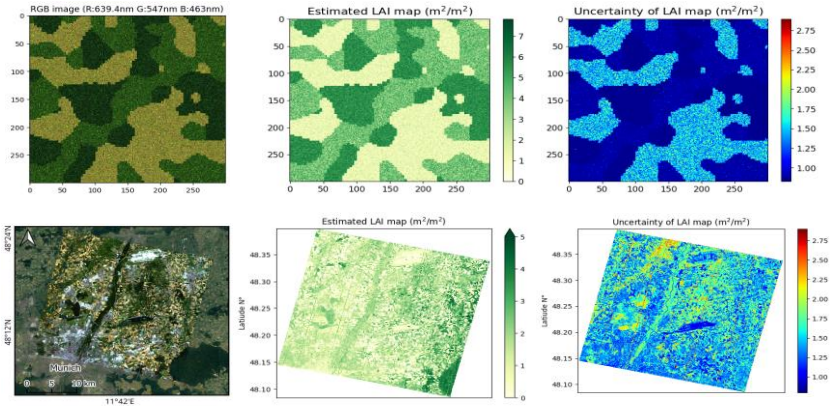
Inputs:

- Can handle CHIME type **netCDF** and **ENVI** Standard image formats
- Works with **ARTMO-trained GPR models**

Key Features:

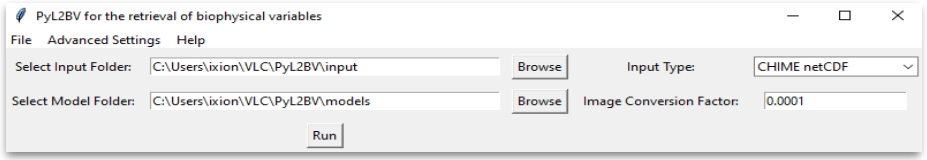
- **2 versions: with and without GUI**
- **Parallel** retrieval process for more models
- Output **estimate** and **uncertainty**
- Optionally: Plot and export results

Test scenes: CHIME E2E, PRISMA



PYL2BV

GUI Interface:



Conclusions compared to L2BV MATLAB:

- Produces **same results**
- Significantly **faster** speed
- **Less** memory usage

Limitations:

- Only works with GPR
- Needs further development for dealing with large tiles

Outlooks:

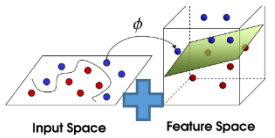
- Could be the base for a future CHIME L2BV processor
- More MLRAs need to be implemented
- Cloud-based version (OpenEO, GEE): **PyEOGPR**

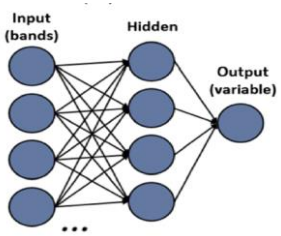
Availability: upon publication

PyEOGPR

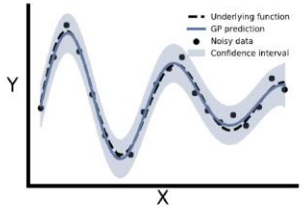
Conclusions…

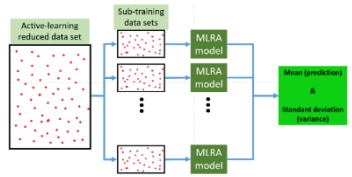
$$C = \frac{1}{2}\sum_i (y_i - w^T x_i)^2 + \frac{1}{2}\tau\|w\|^2$$

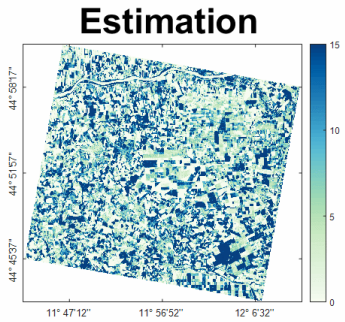




- **Four MLRA models** have been implemented into the L2BV module.
- **Epistemic uncertainty** implemented
- **Good** performance is achieved at **canopy level**.
- **Leaf models** require further **research**.
- A generic **Python** version in the pipleline.







PYL2BV PyEOGPR https://pypi.org/project/pyeogpr/

- How much "instrument agnostic" are the current state of the art algorithms?
- Can the AI based approaches really become instrument agnostic?
- …

Models should be left untouched.
Image processing adapts to the models.

thanks
... any questions?
Canopy nitrogen content (g/m2)