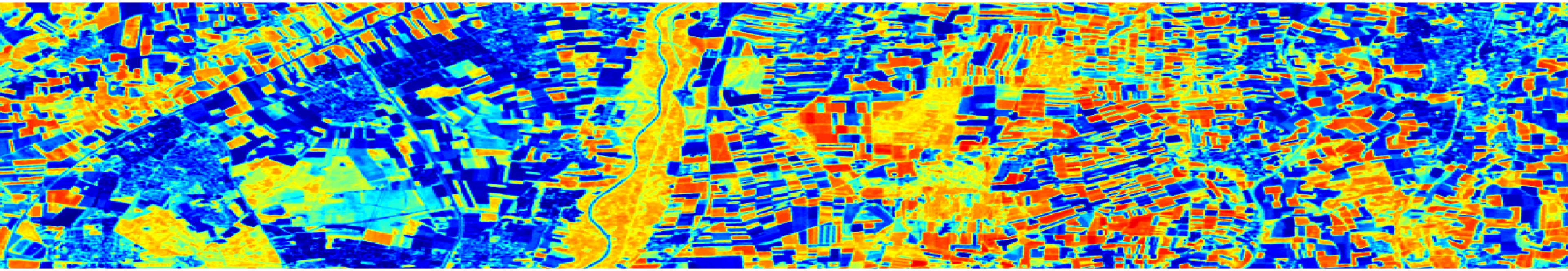
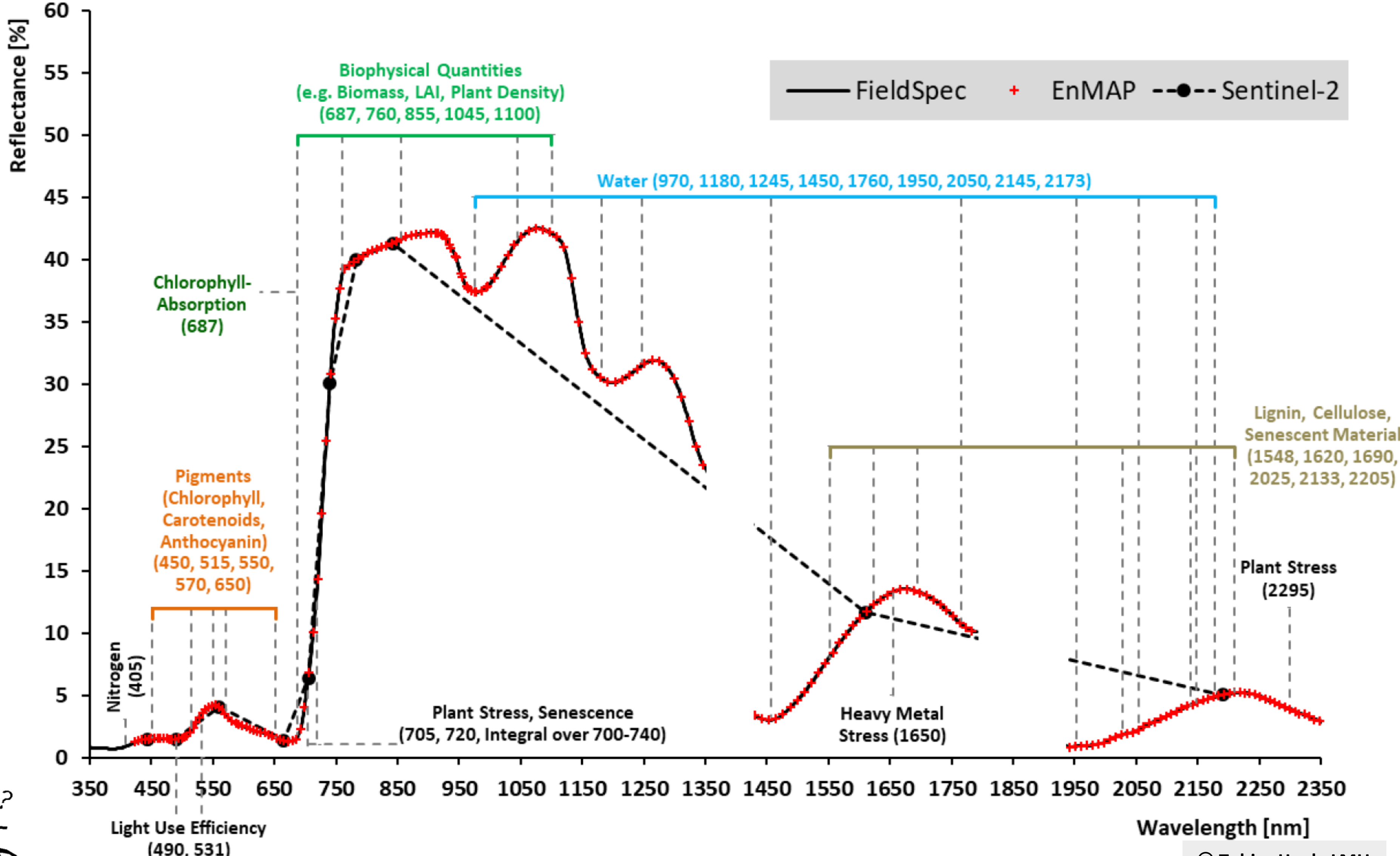


Development of vegetation traits models using hybrid retrieval workflows in the context of the CHIME mission preparation

Jochem Verrelst & co-authors

LEO: Charlotte De Grave, Eatidal Amin, Pablo Reyes, Miguel Morata, Ernique Portales, Santiago Belda, Juan Pablo Rivera, Luca Pipia
LMU: Katja Berger, Matthias Wocher, Tobias Hank
GFZ: Karl Segl, Stephane Guilasso
UNIMIB & CNR: Giulia Tagliabue, Cinzia Panigada, Mirco Boschetti, Gabriele Candiani
ESA: Claudia Isola

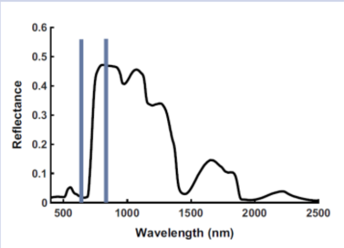
Why contiguous spectral measurements?



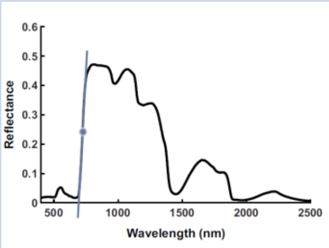
How to extract such information?

Some retrieval methods….

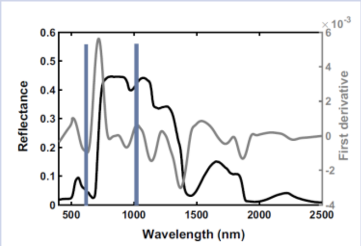
VI



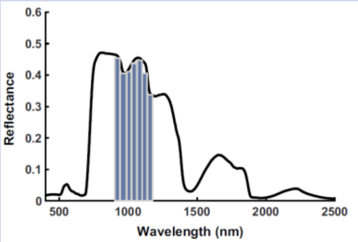
REP



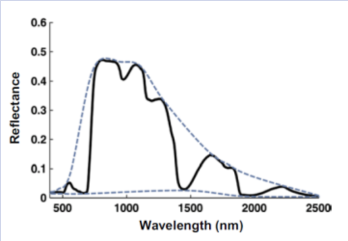
Der.



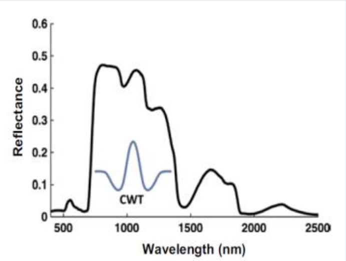
Int.



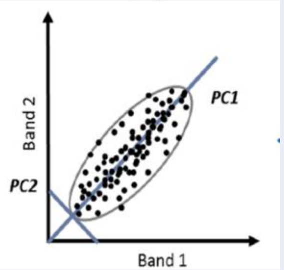
cont. rem.



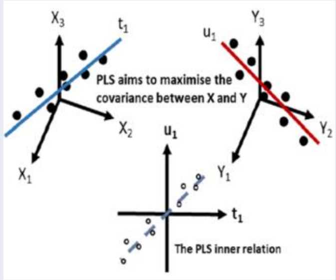
wavelet



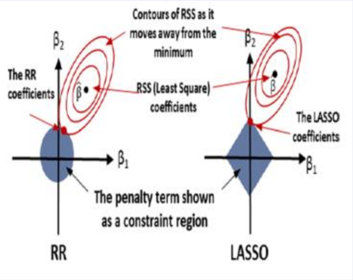
PCR



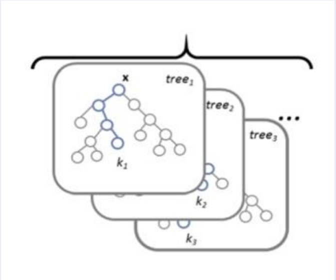
PLSR



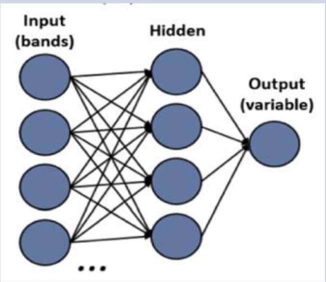
RR



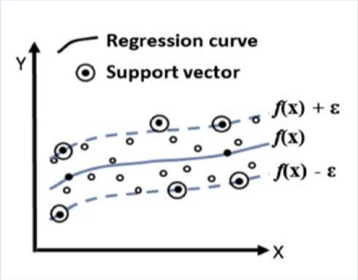
RF



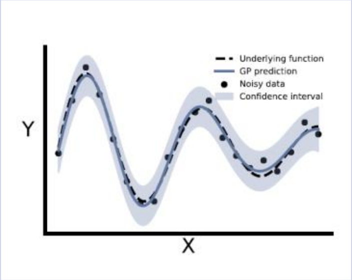
NN



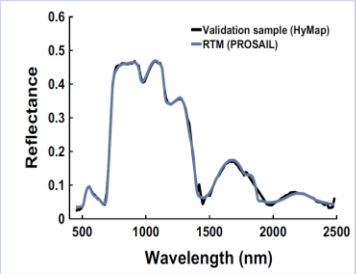
SVR



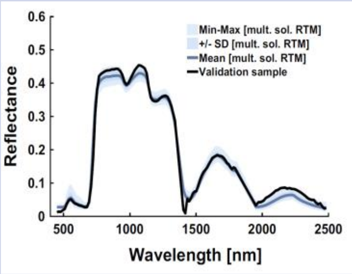
GPR



num. inv.



LUT inv.



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

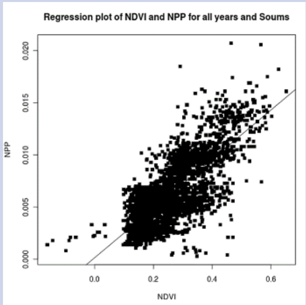
Retrieval methods for vegetation properties mapping

Parametric regression

Spectral relationships that are sensitive to specific vegetation properties

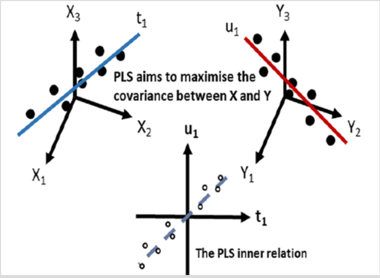
$$NDVI = \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED})}$$

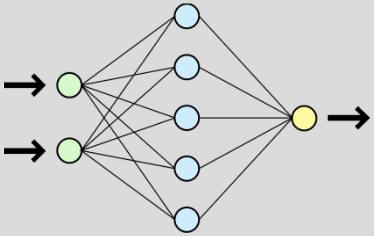
Normalized Difference Vegetation Index



Non-parametric regression

Data-driven techniques that search for relationships between spectral data and biophysical variables

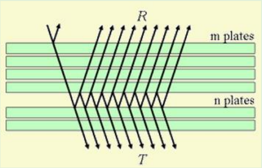


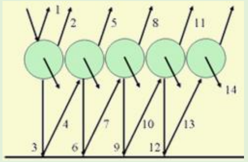


RTM inversion

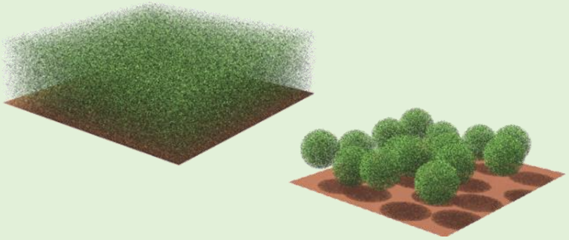
Models that simulate interactions between vegetation and radiation

leaf

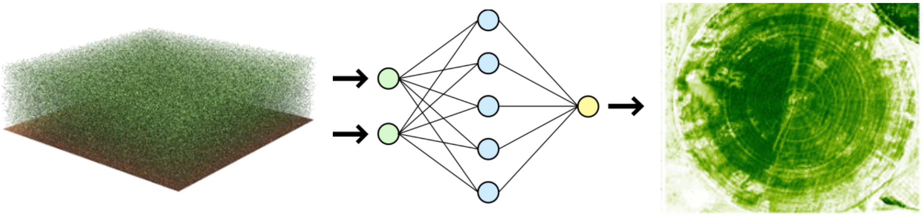




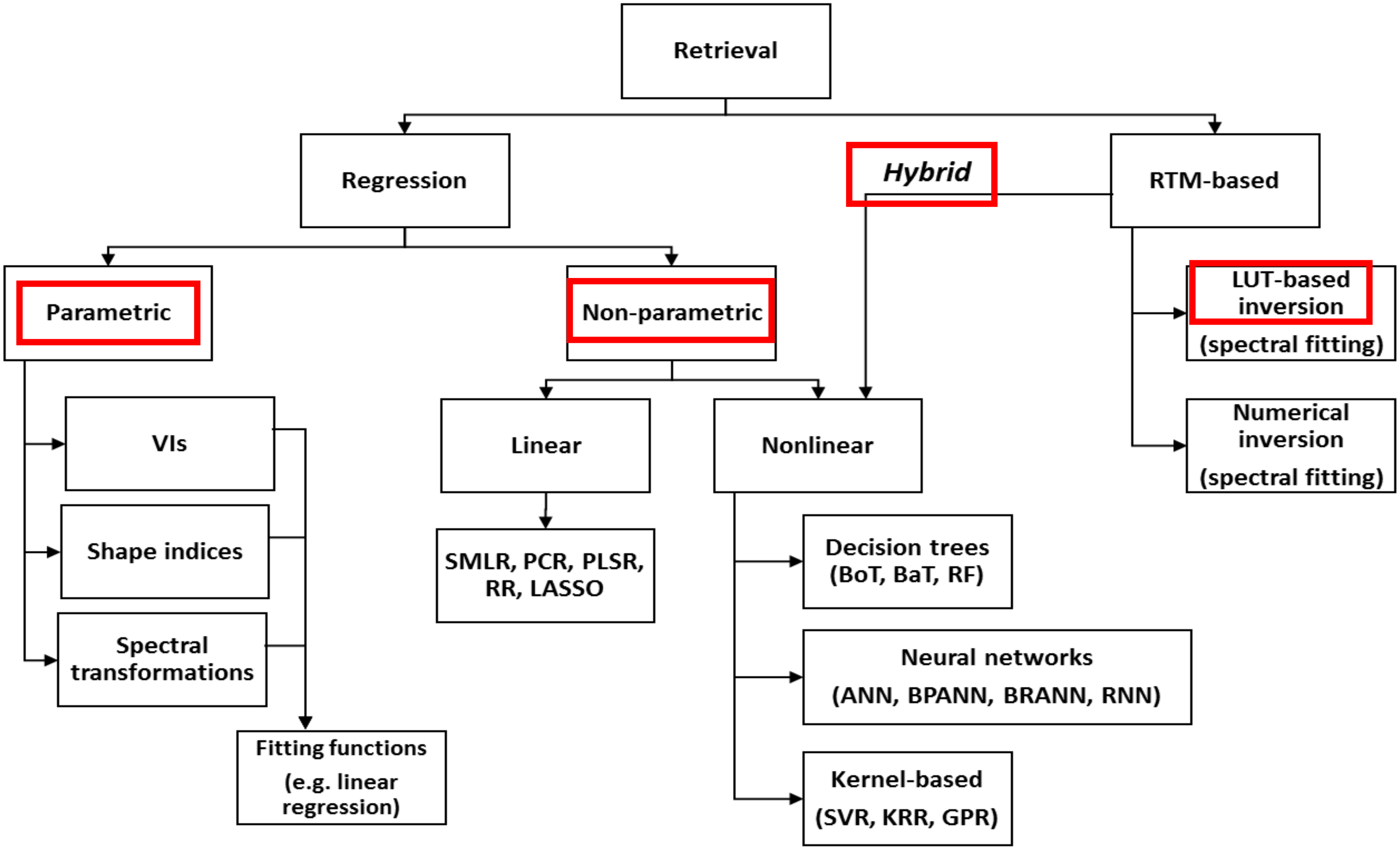
canopy



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



Taxonomy retrieval methods



towards operational processing

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

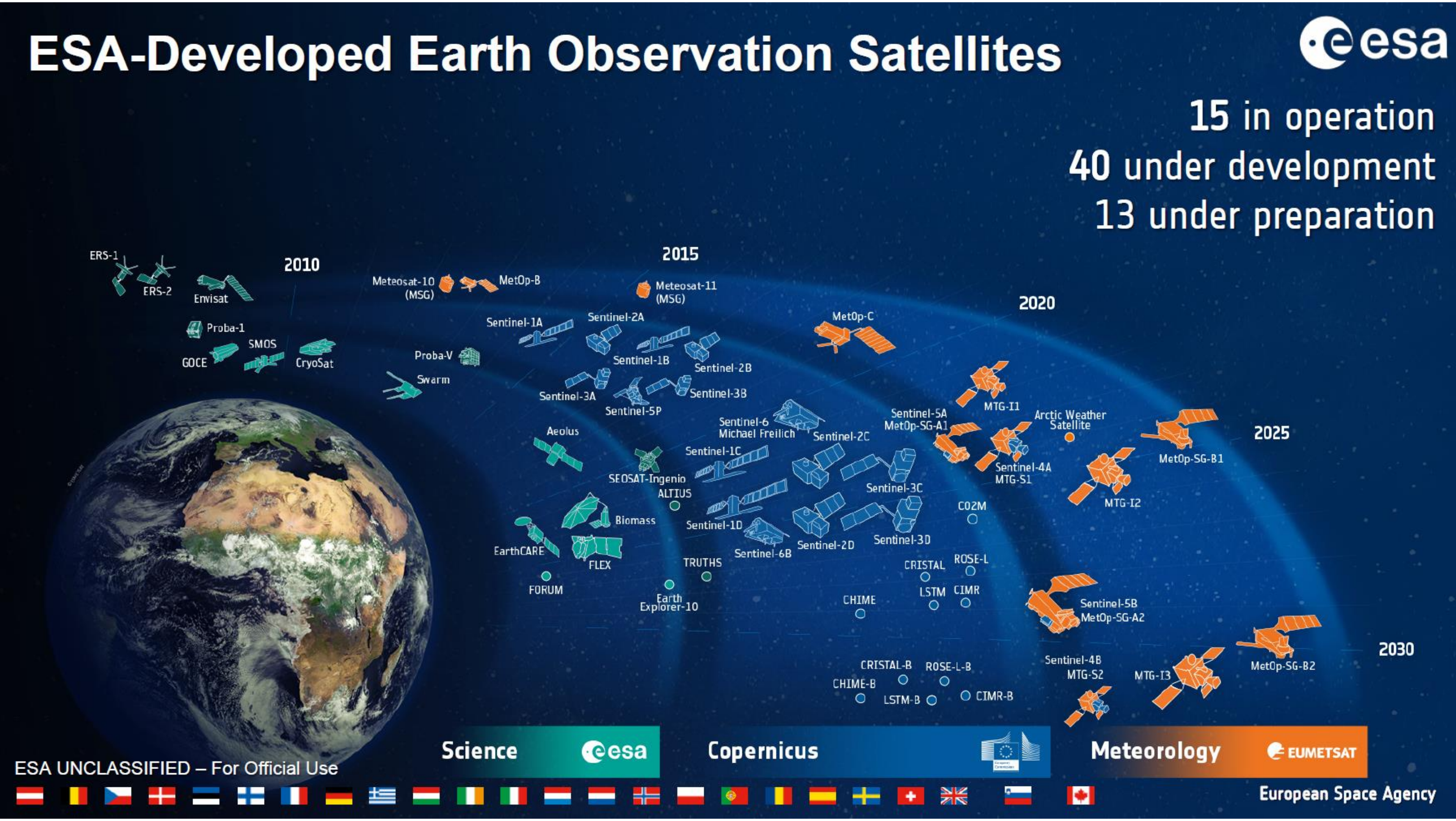
Operational processing?

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

* Some machine learning methods (e.g. probabilistic 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.

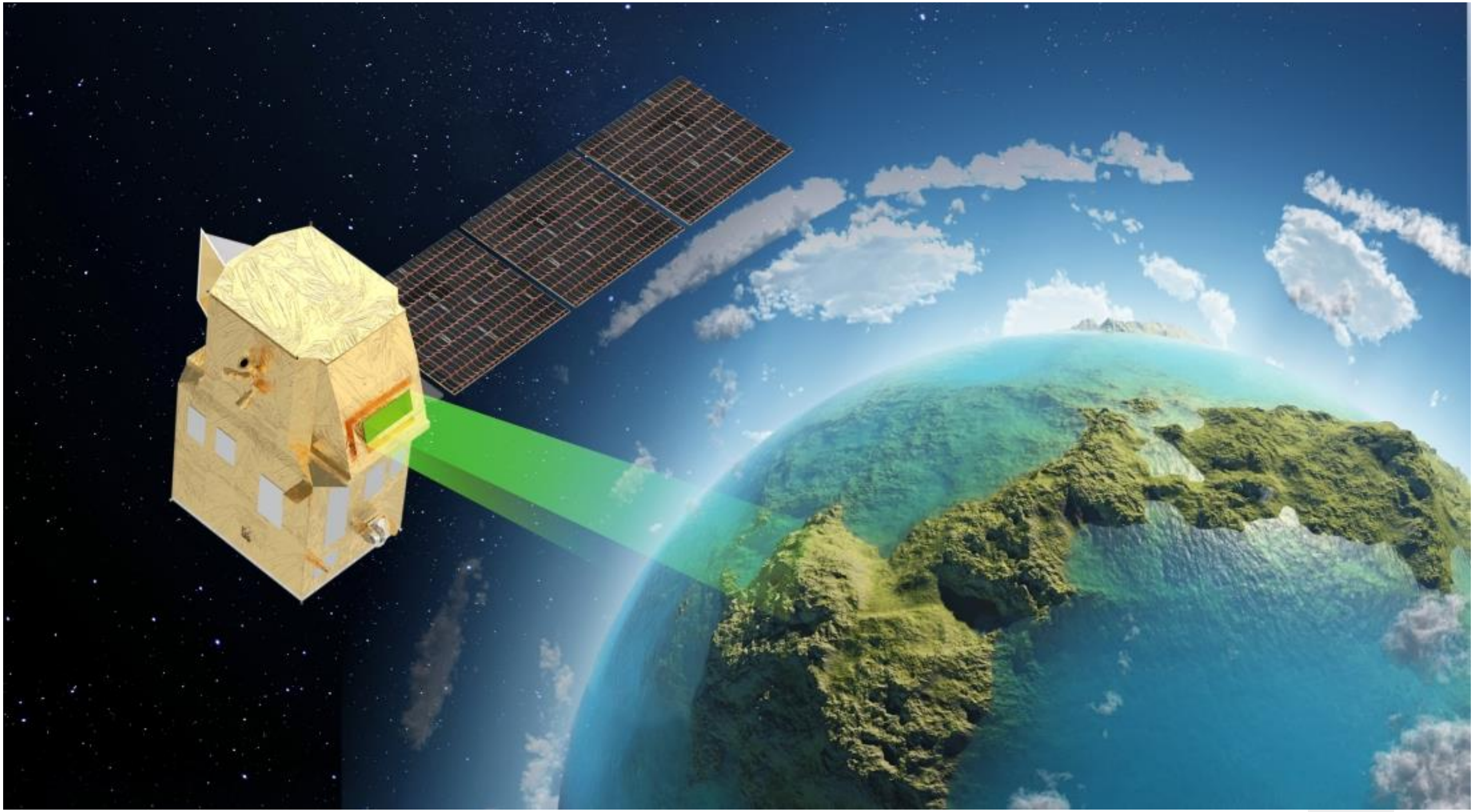
Why need for operational processing?



© Michael Rast, ESA

CHIME

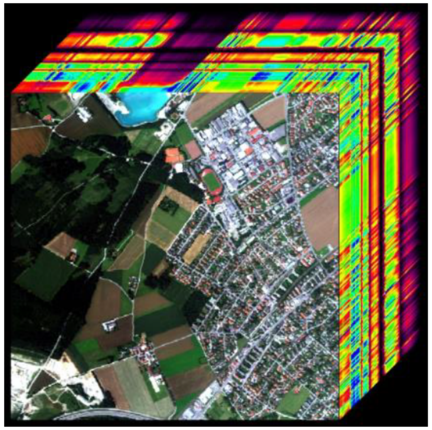
The **Copernicus Hyperspectral Imaging Mission**, CHIME, will carry a visible to shortwave infrared spectrometer to provide routine hyperspectral observations to support new and enhanced services for sustainable agricultural and biodiversity management, as well as soil property characterisation.



Technical Concept of CHIME

Routine spectroscopic observations in contiguous spectral bands conducted at:

- Instrument: Pushbroom Imaging Spectrometer
  400 – 2500 nm, $\Delta\lambda \leq$ 10nm,
- Revisit (temporal resolution) 10-15 days,
- GSD (spatial resolution): 20-30m,
- Sun synchronous orbit (LTDN 10:30 – 11:30),
- Nadir view covering land and coastal areas,
- High radiometric accuracy, low spectral/spatial misregistration.



Hyperspectral data cube
(courtesy DLR)

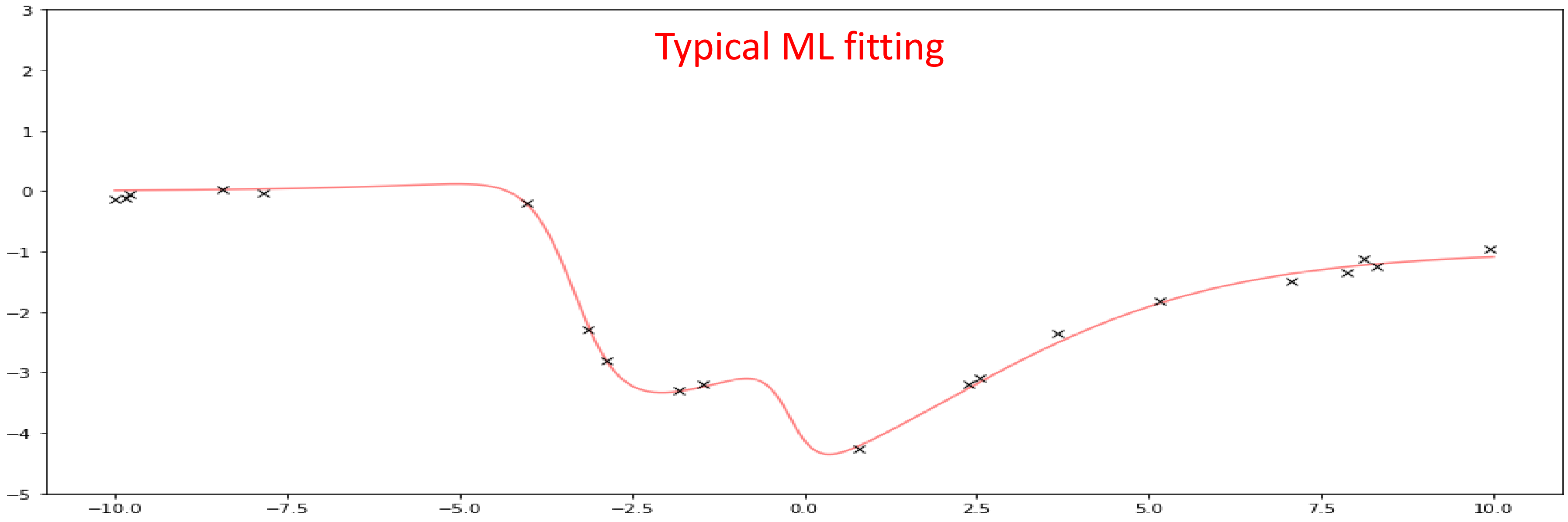
CHIME Core Data Products

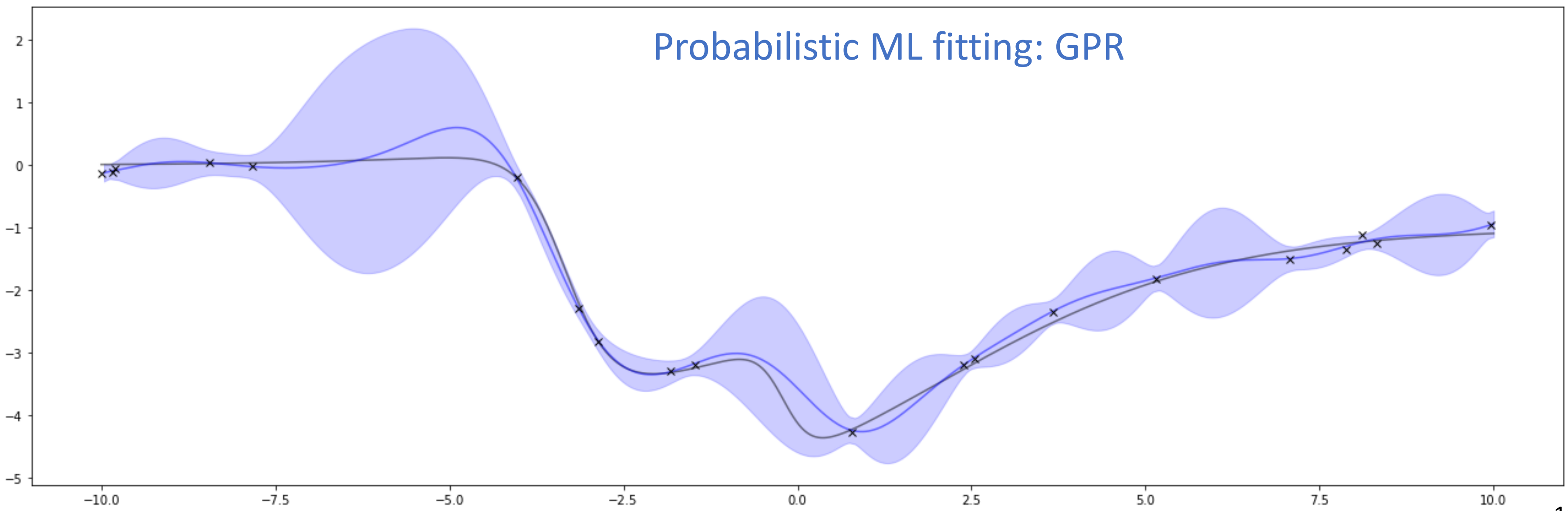
The mission shall provide access to Level-1B, Level-1C and Level-2A products accessable via DIAS and with API support:

- Bottom-of-Atmosphere (BOA) reflectance (atmospheric corrected)
- Ortho-rectified geometry
- Basic pixel classification (opaque clouds, thin clouds, cloud shadows, vegetation, water, snow etc.)

Additionally the mission can provide a set of downstream products related to the different mission applications. → **Vegetation products**

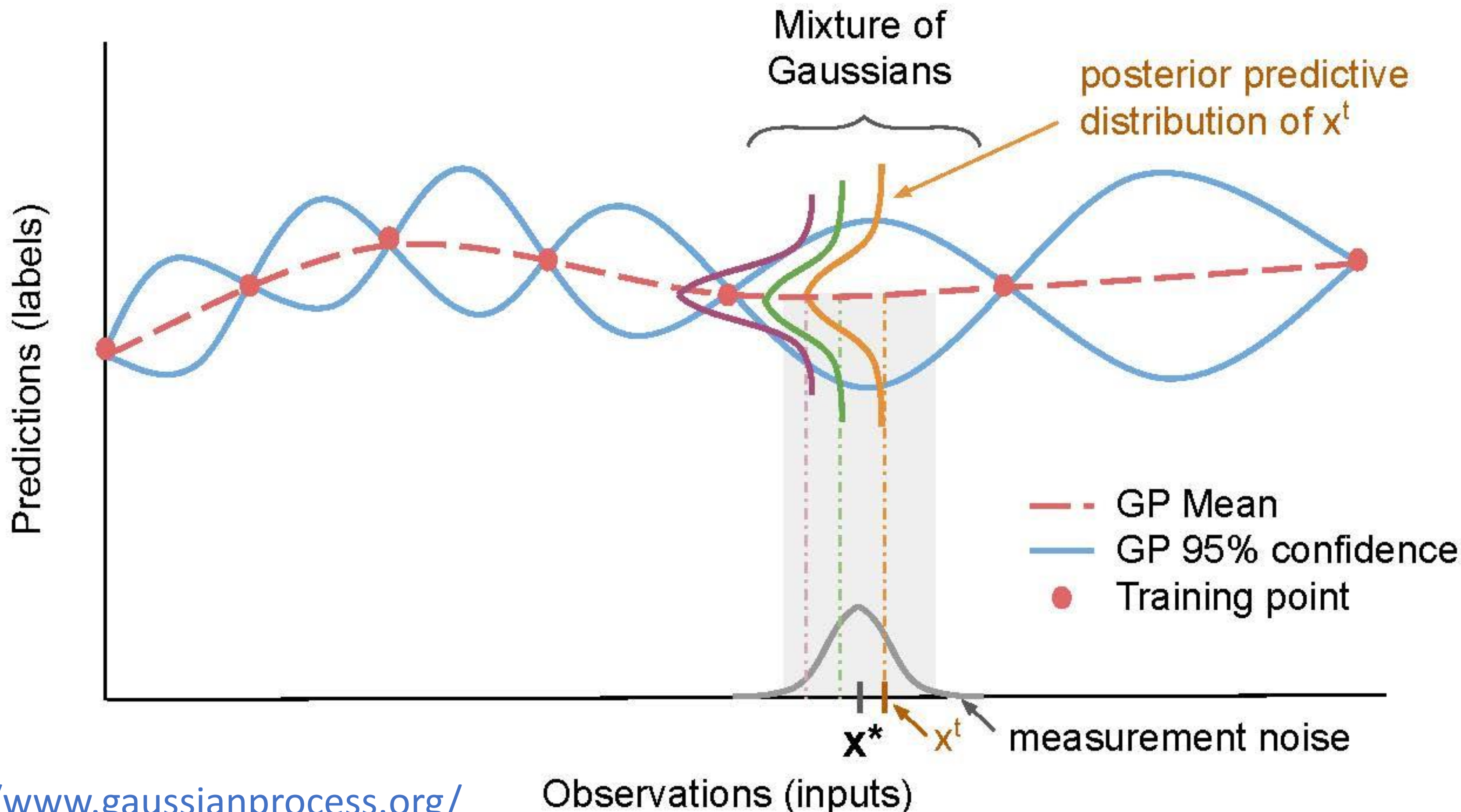
Gaussian process regression: a probabilistic ML algorithm





https://jessicastringham.net/2018/05/18/Gaussian-Processes/

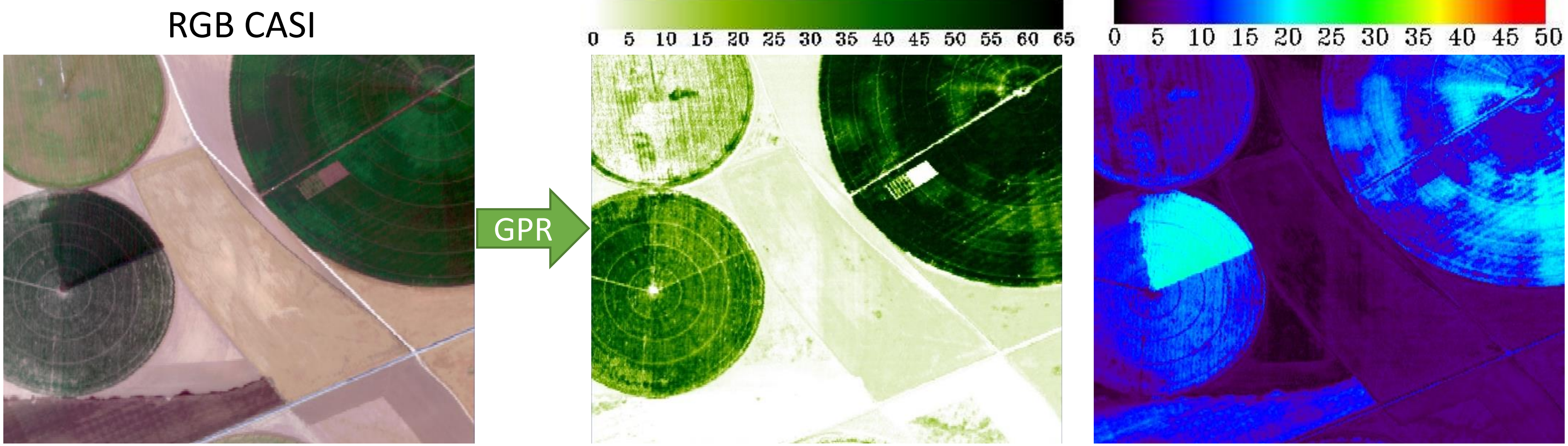
Gaussian process regression is nonparametric (*i.e.* not limited by a functional form), so rather than calculating the probability distribution of parameters of a specific function, **GPR calculates the probability distribution over all admissible functions that fit the data**. However, similar to the above, we specify a prior (on the function space), calculate the posterior using the training data, and compute the predictive posterior distribution on our points of interest.

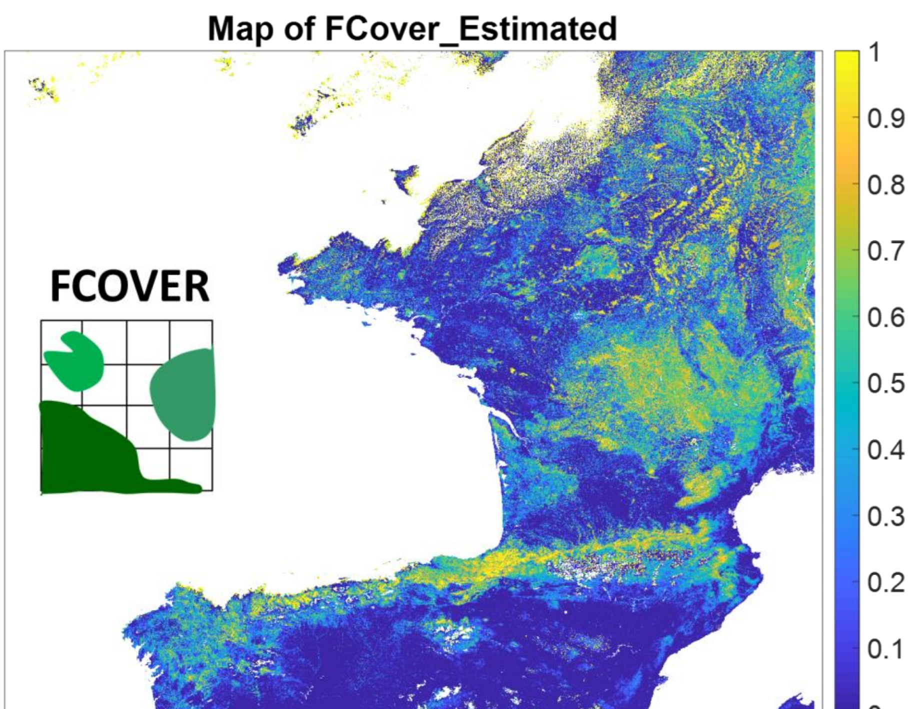


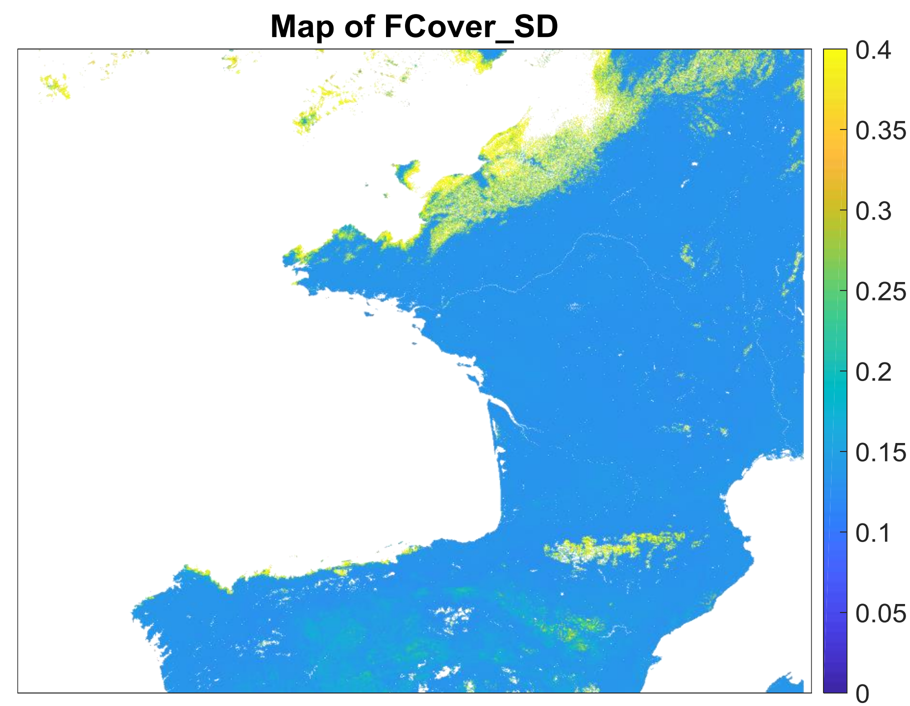
http://www.gaussianprocess.org/

https://towardsdatascience.com/quick-start-to-gaussian-process-regression-36d838810319

GPR models

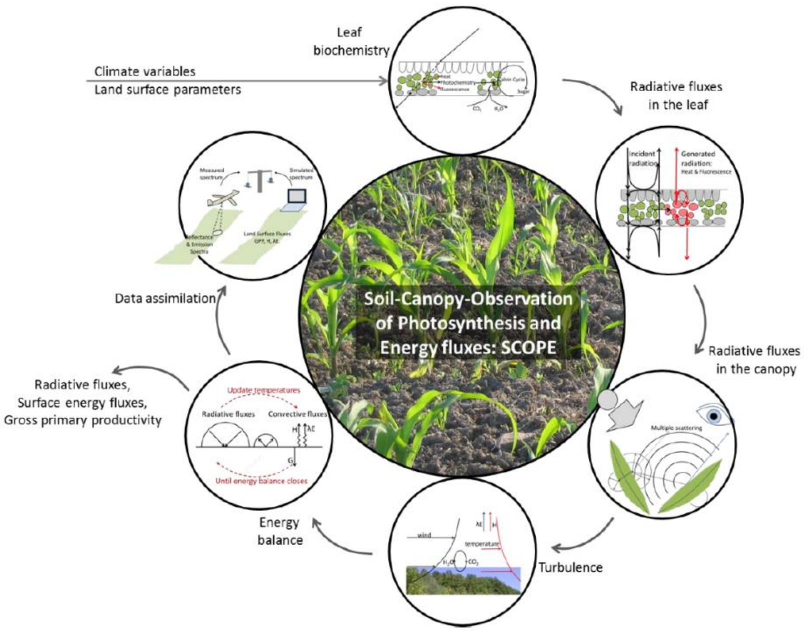






De Grave, C., Verrelst, J.. et al (2020). Quantifying vegetation biophysical variables from the Sentinel-3/FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources. *Remote Sensing of Environment*, *251*, 112101.

Hybrid retrieval method
(generic, accurate, fast & uncertainties)



SCOPE (v 1.70)*

+

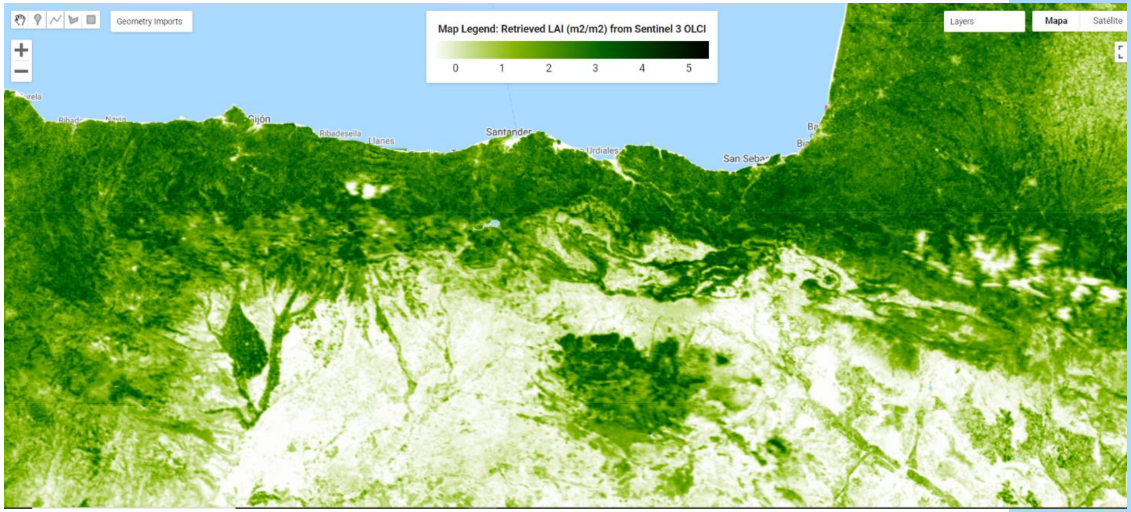
Machine Learning Algorithm

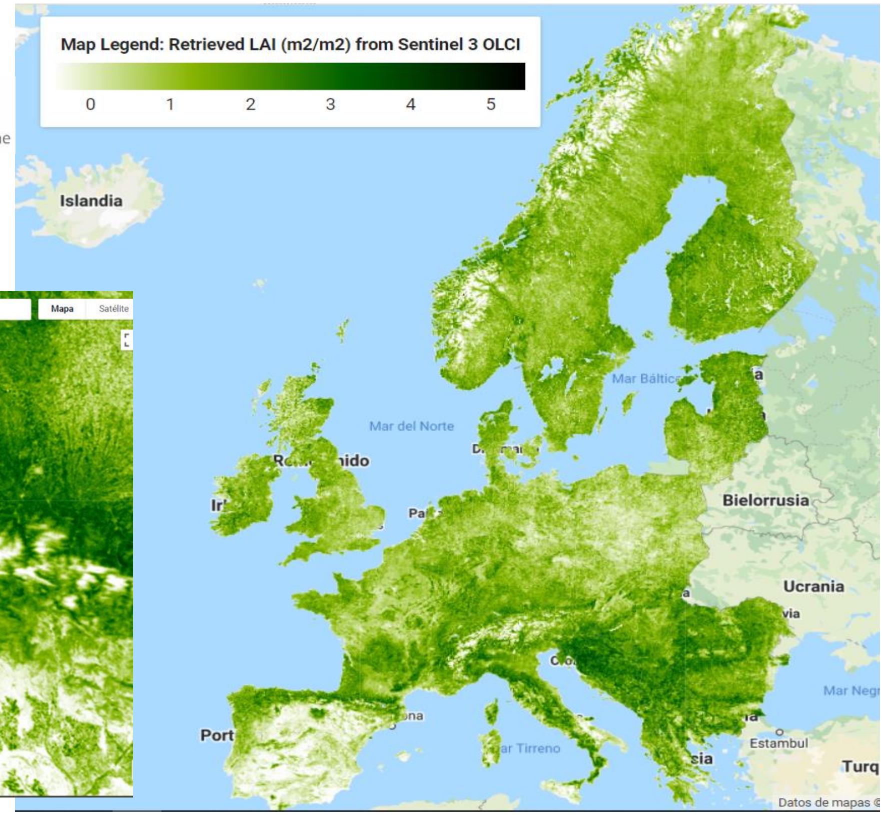
Gaussian Process Regression (GPR) + DR method (PCA, 20 components)

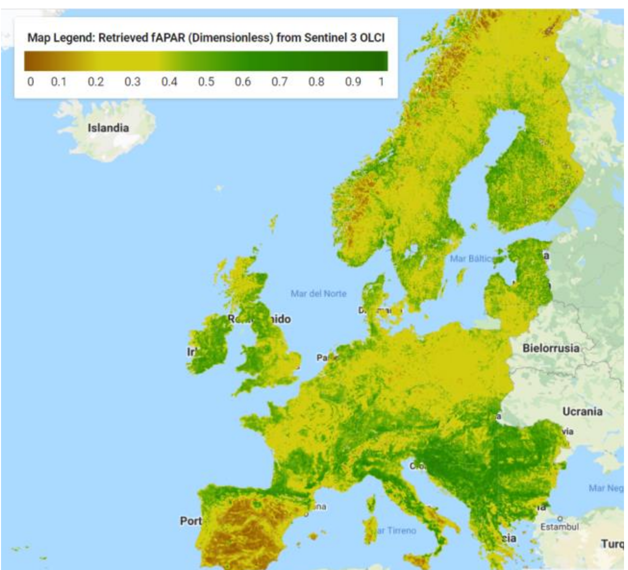
L2A.MI.01 → GPR models → L2B.MO.01 →
- Variable (μ)
- SD (abs. Uncertainty)
- CV% (μ/SD) (rel. Uncertainty)

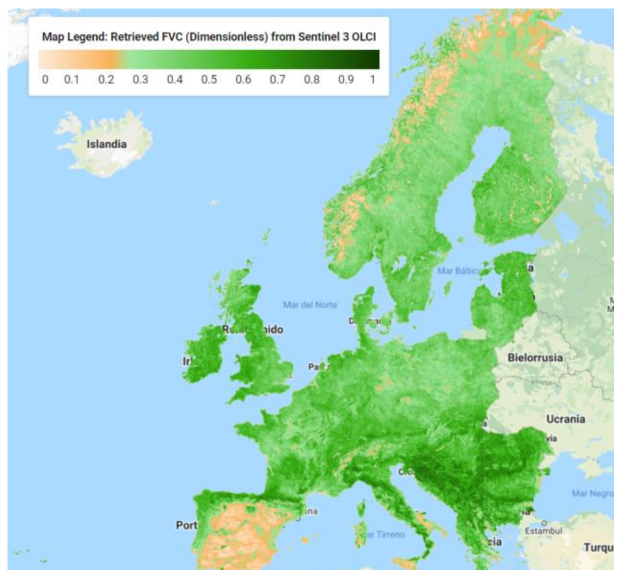
Towards operational processing with GPR

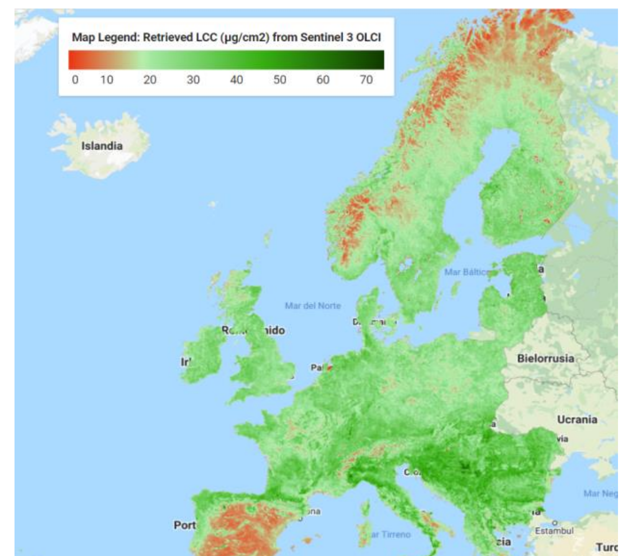
- ✓ Sentinel-3 data in GEE
- ✓ Models need to be light for smooth processing
- ✓ Uncertainties can be calculated







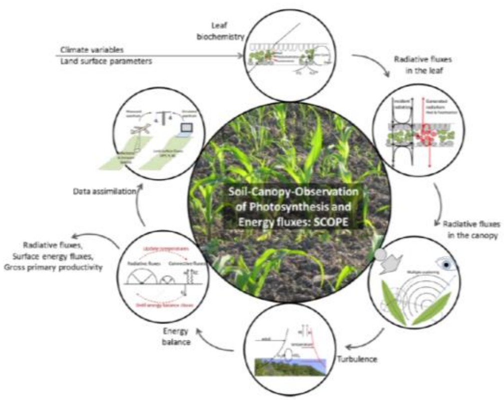




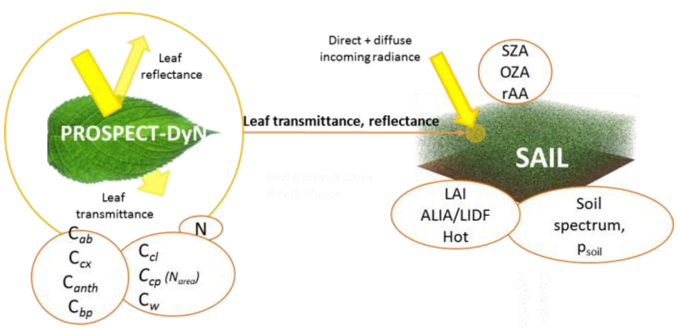
Towards GPR vegetation models development for CHIME: Hybrid approach

RTM	dim. red.	GPR

SCOPE

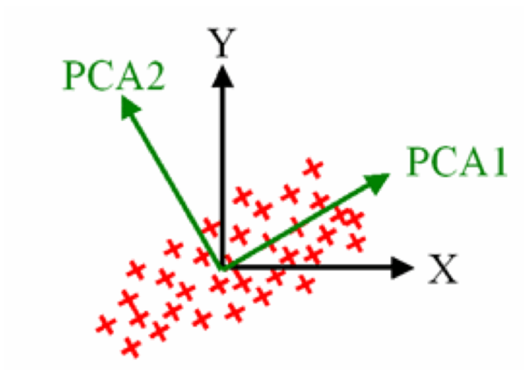


PROSPECT-PRO-SAIL

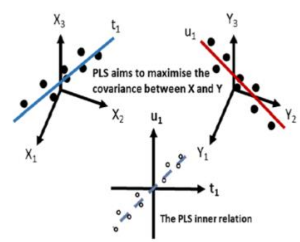


+

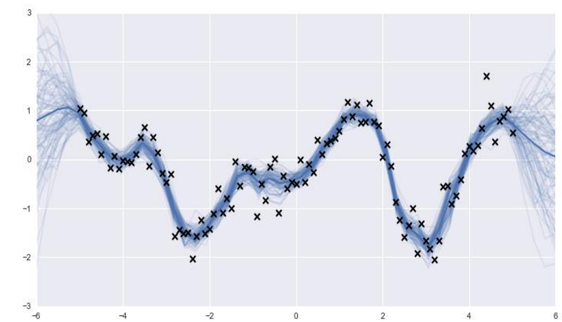
PCA



PLS



+



ARTMO → GPR models (.mat) → Standalone executable

CHIME priority vegetation variables

1. **LCC: Leaf Chlorophyll Content**
2. **LWC: Leaf Water Content**
3. **LDMC: Leaf Dry Matter Content**
4. **LNC: Leaf Nitrogen Content**
5. **LCLC: Leaf cellulose and lignin content**

6. **LAI: Leaf Area Index**
7. **CCC: Canopy Chlorophyll Content**
8. **CWC: Canopy Water Content**
9. **CDMC: Canopy Dry Matter Content**
10. **Canopy Nitrogen Content**
11. **CCLC: Canopy cellulose and lignin content**
12. **FAPAR: Fraction of Absorbed Photosynthetically Active Radiation**
13. **FVC: Fractional vegetation cover**

Training data: Input SCOPE

Variable type	Variable	Distribution	Min	Max	Mean	SD
Weather	Rin ($W.m^{-2}$)	Gaussian*	20	1100	400	300
	Rli ($W.m^{-2}$)	Gaussian*	100	400	250	125
Leaf biochemical	Vcmax ($\mu mol.m^{-2}.s^{-1}$)	Gaussian*	10	180	80	40
Leaf structure	N	Gaussian*	1	2.7	1.5	0.5
	Cab ($\mu g.cm^{-2}$)	Uniform	1	100		
	Cca ($\mu g.cm^{-2}$)	Gaussian*	0	30	10	5
	Cdm ($g.cm^{-2}$)*	Gaussian*	0.002	0.02	0.005	0.003
	Cw**	Gaussian*	0.005	0.035	0.012	0.006
Canopy structure	LAI	Uniform	0.1	10		
	LIDFa (rad)***	Uniform	-1	1		
	LIDFb (rad)***	Uniform	-1	1		
	VH (m)	Gaussian*	0.3	20	3	8
Geometry	SZA (°)	Uniform	0	80		
	OZA (°)	Uniform	-25	25		
	RAA (°)	Uniform	0	180		

Rin: Incoming shortwave radiation; **Rli**: Incoming longwave radiation; **Vcmax**: maximum carboxylation capicity; **N**: Leaf mesophyll structure; **Cab**: Leaf chlorophyll content; **Cdm**: Leaf dry matter content; **Cw**: Leaf water thickness; **Cant**: Leaf anthocyanin content; **Cs**: Leaf senescent material content; **Cca**: Leaf carotenoid content; **LAI**: Leaf Area Index; **LIDFa**: Average leaf angle; **LIDFb**: Variation in leaf angle; **VH**: Vegetation Height; **SZA**: Solar Zenith Angle; **OZA**: Observer Zenith Angle; **RAA**: Relative Azimuth Angle; * truncated Gaussian; ** Constraint: Cw/(Cw+Cdm) between 0.45 and 0.93; *** constraint: |LIDFa|+ |LIDFb| < 1

- Based on global sensitivity analysis
- Based on leaf optical properties databases (OPTICLEAF) < S. Jacquemoud, L. Bidel, C. François, G. Pavan (2003); B. Hosgood, G. Andreoli, S. Jacquemoud, A. Pedrini, G. Schmuck, J. Verdebout (1993)
- Based on literature (e.g. García-Haro et al., 2018; Weiss and Baret, 2016; Croft et al., 2015; Houborg et al., 2015; Verrelst et al., 2015; Houborg and Boegh, 2008; Lauvernet et al., 2008)
- To cover all geometrical configurations and canopy realizations
- Fixed variables: default SCOPE values

- SCOPE does not provide nitrogen content (N) and cellulose & lignin
- For these variables PROSPECT-PRO+ SAIL was used

- Spectral bands: **L1C: 239 bands**
- Dimensionality reduction: **20 PCA**

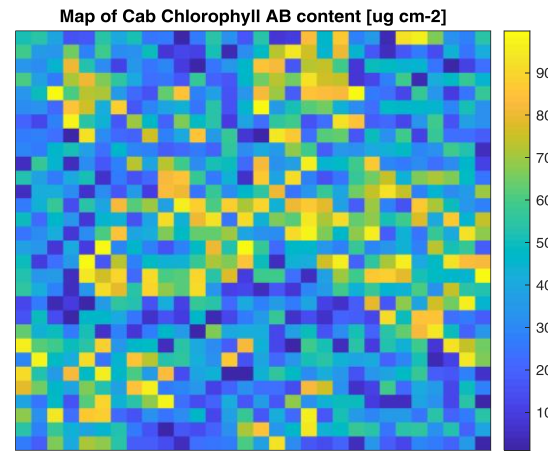
} One .mat file per variable

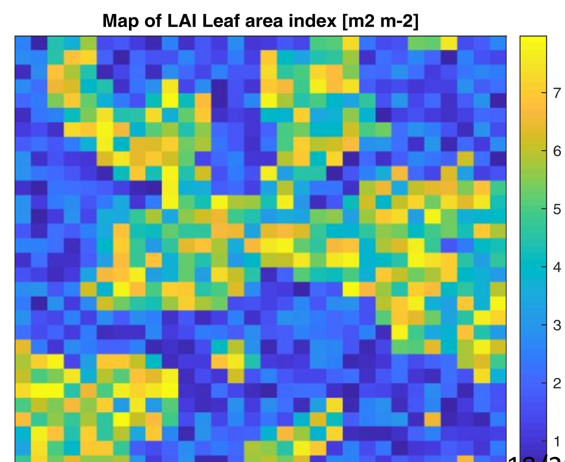
- Models version: **1.7**

Validation into E2E against a reference scene

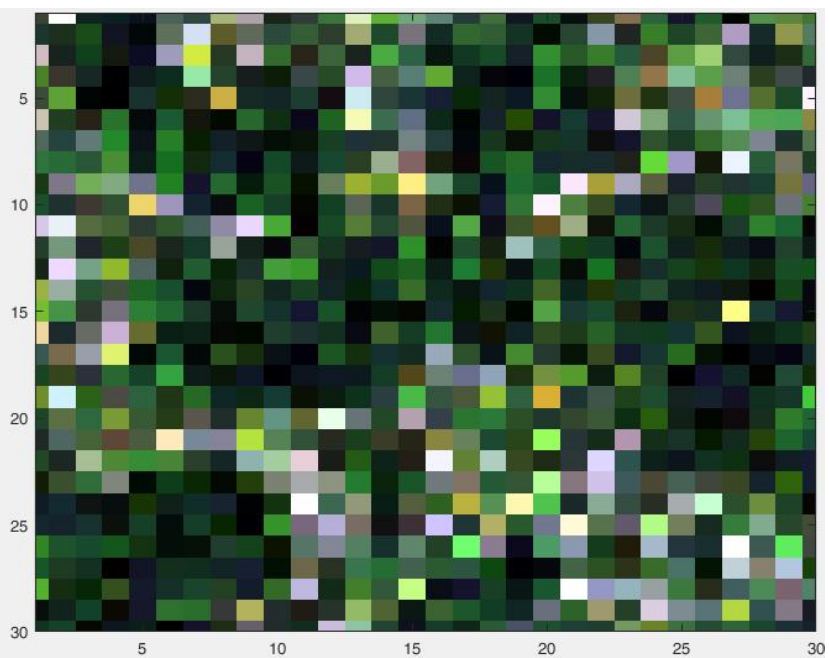
Variable type	Variable	Distribution	Min	Max	Mean	SD
Leaf structure	N	Gaussian*	1	2.7	1.5	0.5
	Cca (µg.cm^{-2})	Gaussian*	0	30	10	5
	Cdm (g.cm^{-2})*	Gaussian*	0.002	0.02	0.005	0.003
	Cw**	Gaussian*	0.005	0.035	0.012	0.006
Canopy structure	LIDFa (rad)***	Uniform	-1	1		
	LIDFb (rad)***	Uniform	-1	1		
Soil	SMC (%)	Gaussian	5	55	25	12.5
	BSM Brightness	Gaussian	0.5	1.5	1	0.5
	BSM lat (°)	Gaussian	20	40	25	12.5
	BSM long (°)	Gaussian	45	65	50	10

Some examples of input layers

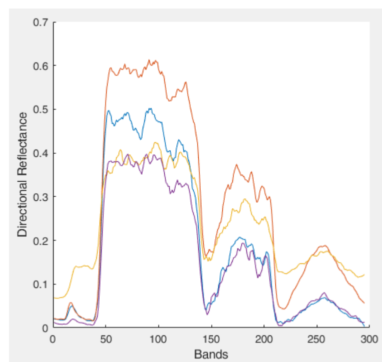


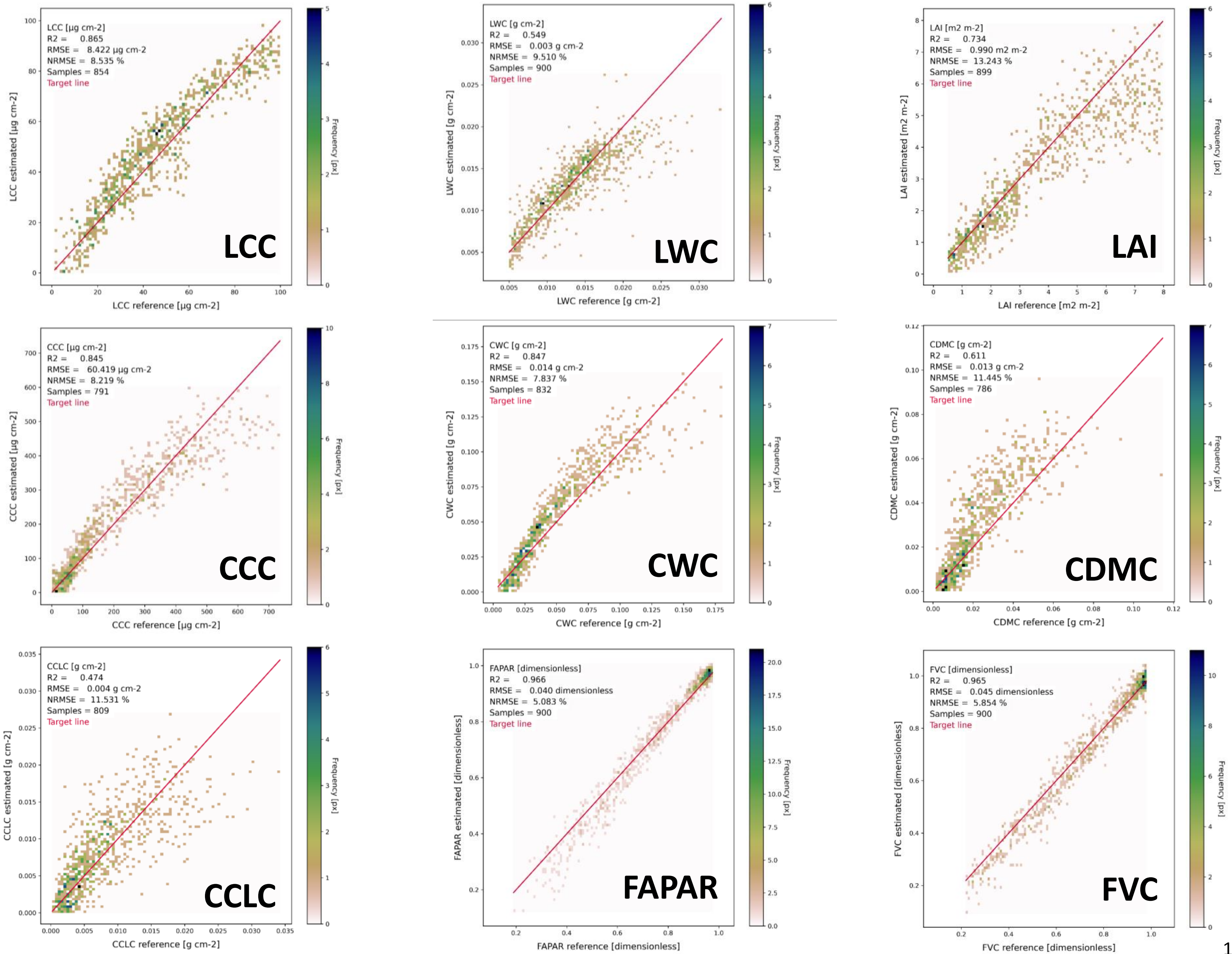


RGB of reference image (generated by SCOPE)

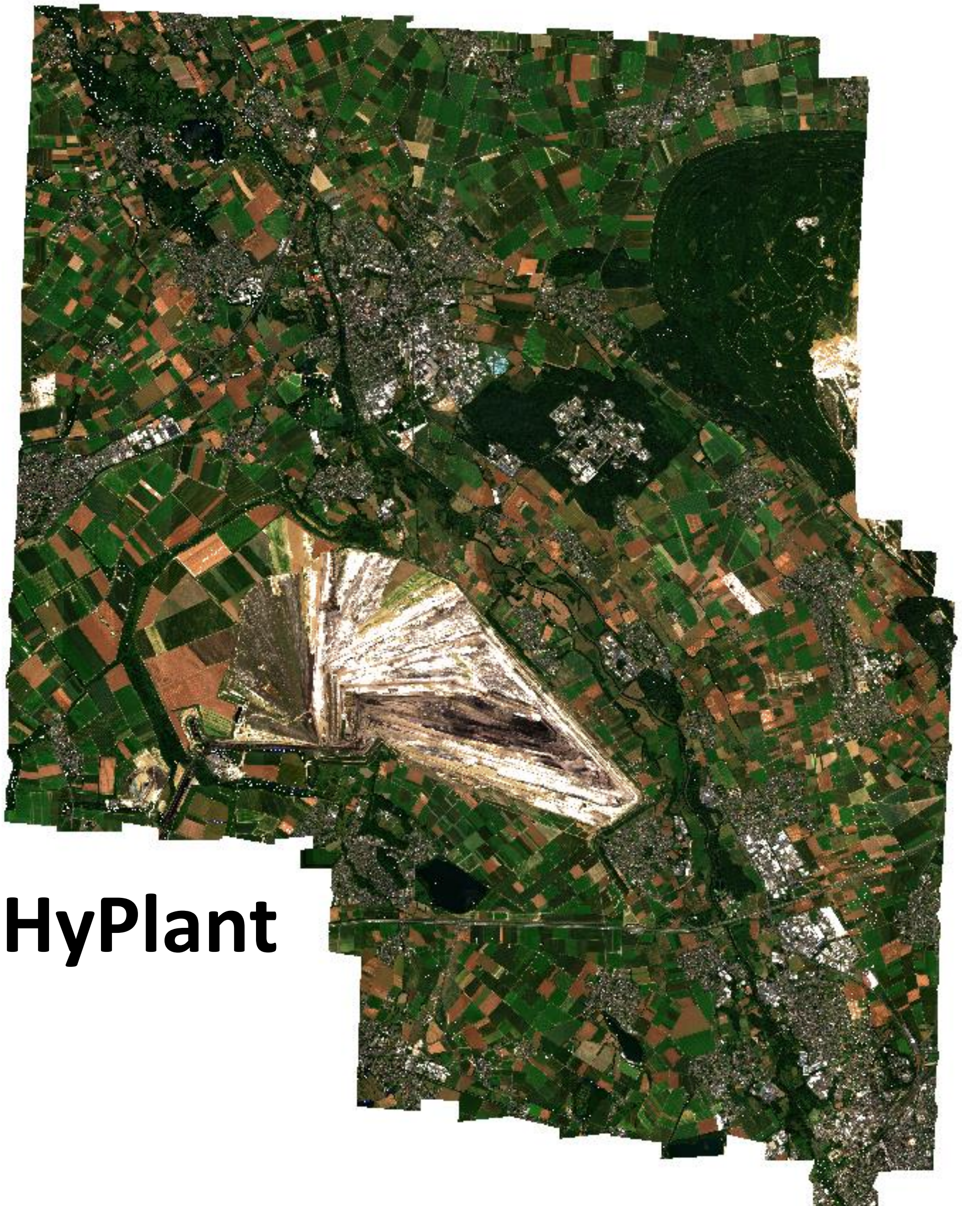


Some examples of spectra




LCC [µg cm-2]
R2 = 0.865
RMSE = 8.422 µg cm-2
NRMSE = 8.535 %
Samples = 854
Target line
LCC
LCC estimated [µg cm-2]
LCC reference [µg cm-2]
Frequency [px]
LWC [g cm-2]
R2 = 0.549
RMSE = 0.003 g cm-2
NRMSE = 9.510 %
Samples = 900
Target line
LWC
LWC estimated [g cm-2]
LWC reference [g cm-2]
Frequency [px]
LAI [m2 m-2]
R2 = 0.734
RMSE = 0.990 m2 m-2
NRMSE = 13.243 %
Samples = 899
Target line
LAI
LAI estimated [m2 m-2]
LAI reference [m2 m-2]
Frequency [px]
CCC [µg cm-2]
R2 = 0.845
RMSE = 60.419 µg cm-2
NRMSE = 8.219 %
Samples = 791
Target line
CCC
CCC estimated [µg cm-2]
CCC reference [µg cm-2]
Frequency [px]
CWC [g cm-2]
R2 = 0.847
RMSE = 0.014 g cm-2
NRMSE = 7.837 %
Samples = 832
Target line
CWC
CWC estimated [g cm-2]
CWC reference [g cm-2]
Frequency [px]
CDMC [g cm-2]
R2 = 0.611
RMSE = 0.013 g cm-2
NRMSE = 11.445 %
Samples = 786
Target line
CDMC
CDMC estimated [g cm-2]
CDMC reference [g cm-2]
Frequency [px]
CCLC [g cm-2]
R2 = 0.474
RMSE = 0.004 g cm-2
NRMSE = 11.531 %
Samples = 809
Target line
CCLC
CCLC estimated [g cm-2]
CCLC reference [g cm-2]
Frequency [px]
FAPAR [dimensionless]
R2 = 0.966
RMSE = 0.040 dimensionless
NRMSE = 5.083 %
Samples = 900
Target line
FAPAR
FAPAR estimated [dimensionless]
FAPAR reference [dimensionless]
Frequency [px]
FVC [dimensionless]
R2 = 0.965
RMSE = 0.045 dimensionless
NRMSE = 5.854 %
Samples = 900
Target line
FVC
FVC estimated [dimensionless]
FVC reference [dimensionless]
Frequency [px]

Towards processing of real data



HyPlant

APEX

PRISMA

The challenge of using simulated data:

How to create a LUT that:

1. *Sufficiently generic for global mapping*
2. *Sufficiently small for fast processing*
3. *Sufficiently realistic for interpreting hyperspectral data*
4. *Optimized for non-vegetated surfaces and noises*

Optimizing with Active Learning (AL):

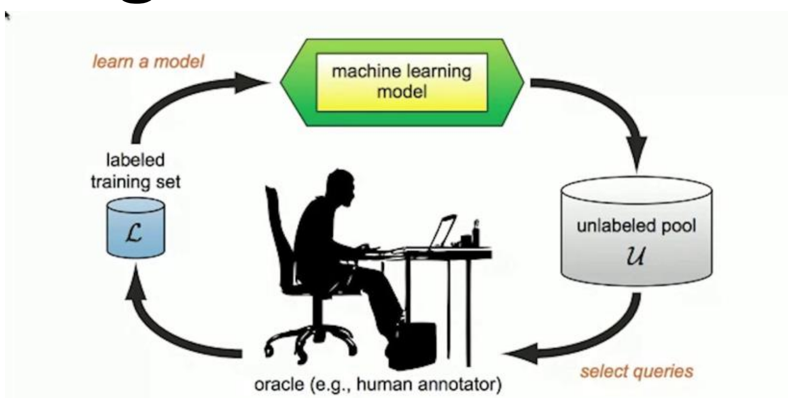
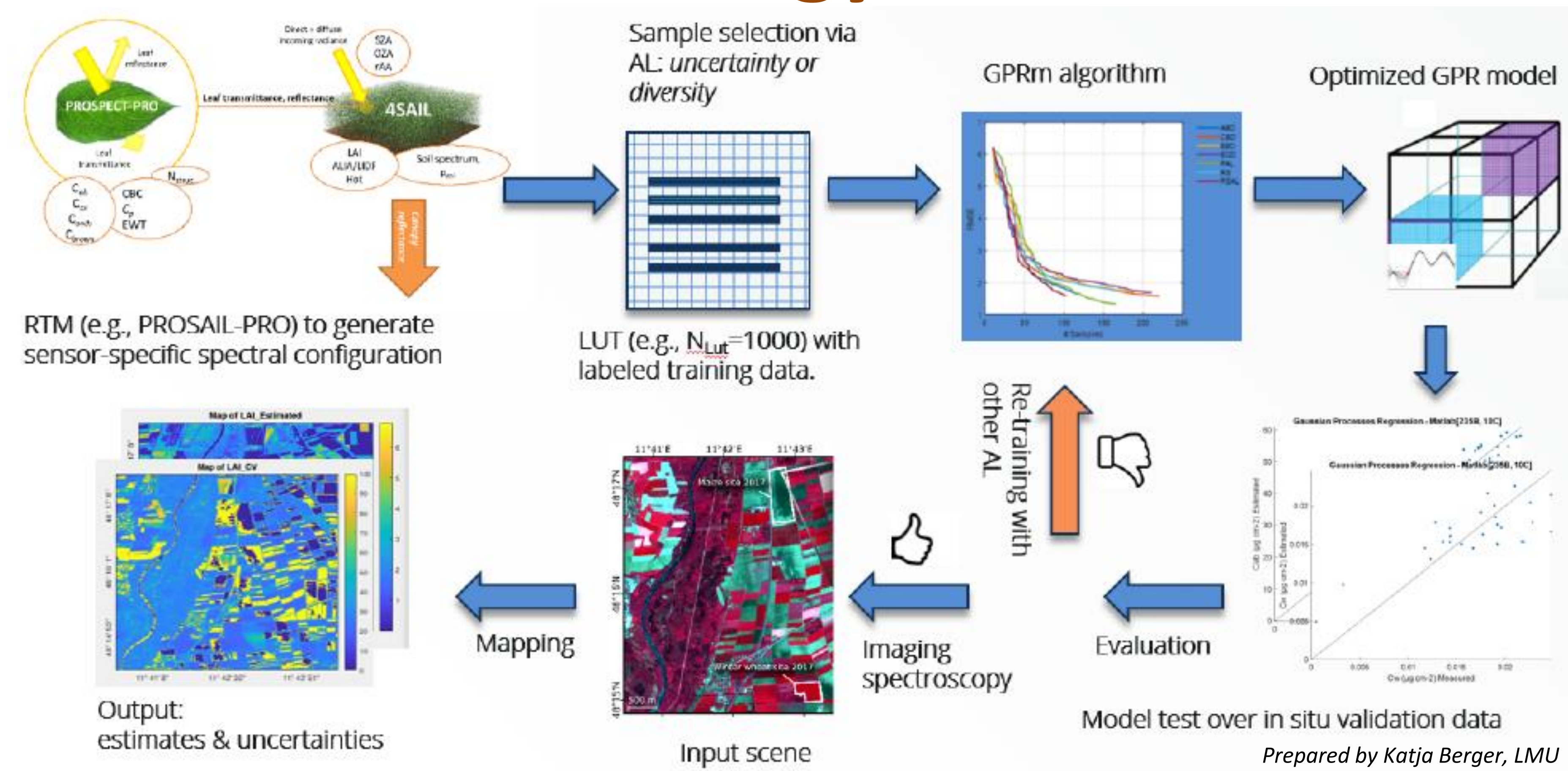


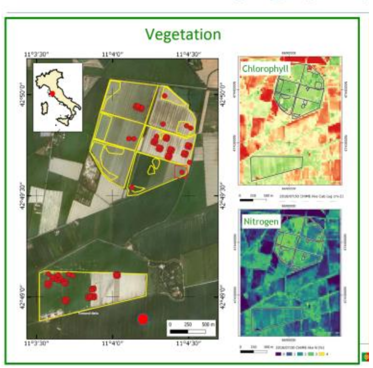
Figure 1: The pool-based active learning cycle.

- **With AL we can optimize the LUT for a specific task, e.g. optimize the hybrid model against field data**
- **A workflow was developed on implementing AL strategy combined with validation against field data**

Workflow AL strategy:



CHIME Airborne Campaign (2018)



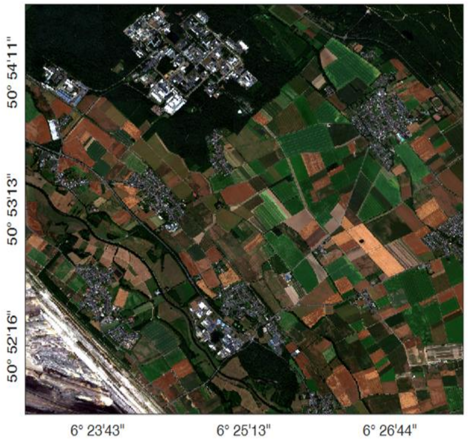
- ✓ 2018 ESA-FLEX/CHIME campaign near Grosseto
- ✓ 3 hyperspectral airborne sensors: APEX, AVIRIS-NG, HyPlant
- ✓ 2 field campaigns: vegetation sampling on corn crop
- ✓ multiple biophsical variables collected

GPR v.1.7 models Validation of CHIME vegetation models

LUT simulated by SCOPE - 1000 samples resampled to CHIME with some bands left out. Optimized with AL against **Grosseto** field dataset.

	Variable	AL method	#samples	ML method	Spectral noise	R2	RMSE	RRMSE (%)	NMRSE (%)	Bands number
leaf	LCC	EBD	219	GPRm	5	0	18.11	41.82	62.01	247
	LWC	EBD	167	GPRm	0	0.88	0.0022	19.78	9.22	210
	LDMC	EBD	187	GPRm	0	0.05	0.0013	28.50	91.53	210
canopy	LAI	EBD	302	GPRm	0	0.86	0.6588	37.17	11.78	247
	CCC	EBD	283	GPRm	0	0.83	95.21	126.46	33.60	247
	CWC	EBD	264	VHGPR	0	0.89	0.041	135.06	67.00	210
	CDMC	EBD	999	GPRm	2.5	0.86	0.035	291.28	132.04	210
	CNC*	EBD	148	GPRm	0	0.65	3.6315	30.69	17.68	210

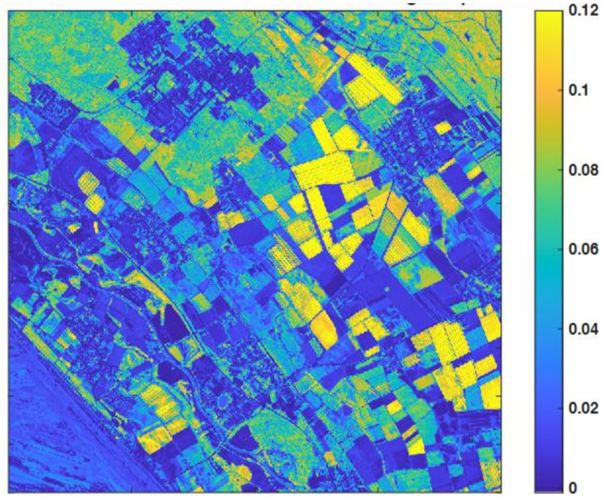
** LUT and validation data kindly shared by Katja Berger, LMU*



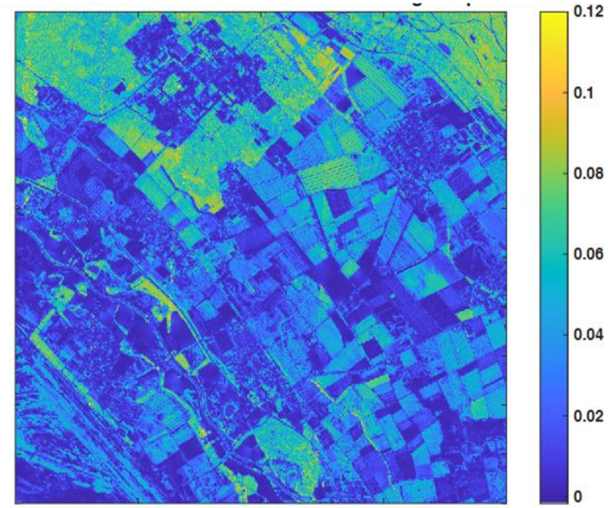
26/06/2018

- ✓ Subset HyPlant airborne flightline, Julich
- ✓ 3m spatial resolution
- ✓ Resampled to CHIME bands

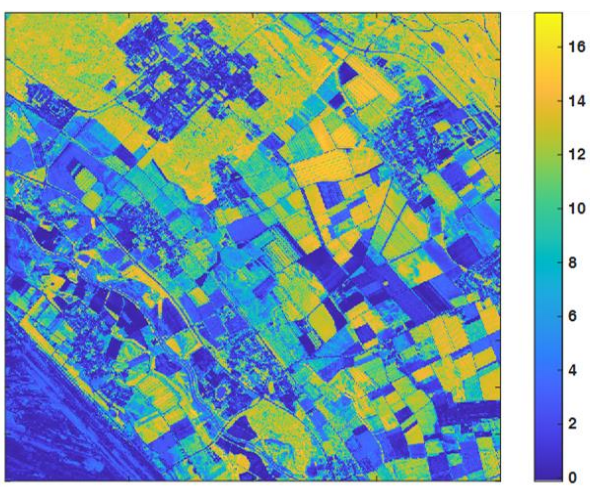
CWC



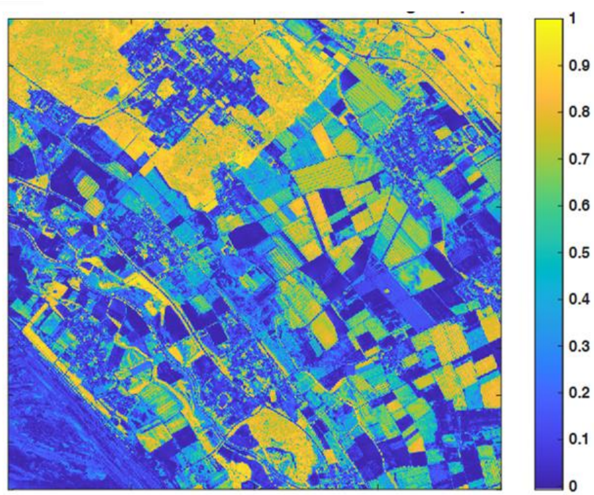
CDMC



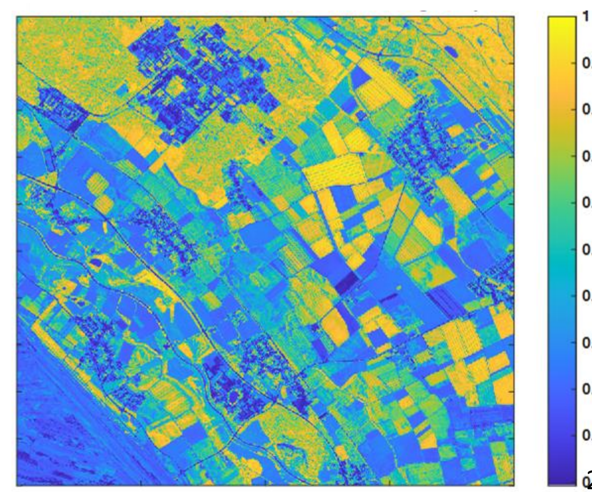
CNC



FAPAR



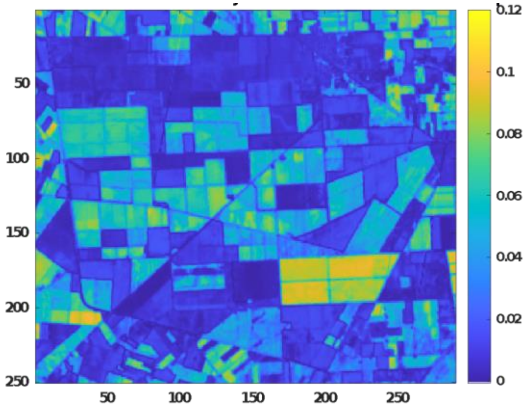
FVC



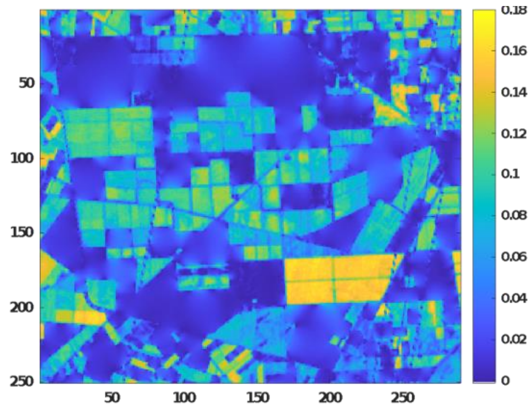
23/05/2020

- ✓ Subset PRISMA image, N. Italy
- ✓ 30m spatial resolution
- ✓ Atm correction & spectral polishing
- ✓ Resampled to CHIME bands

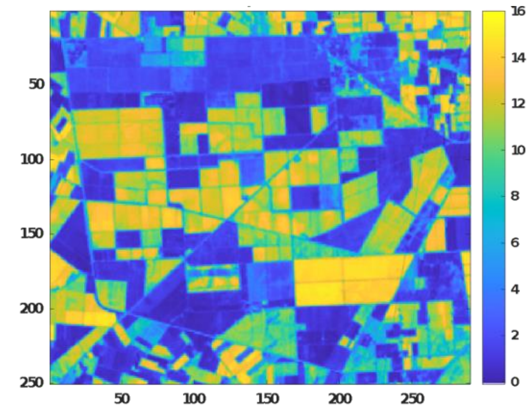
CWC



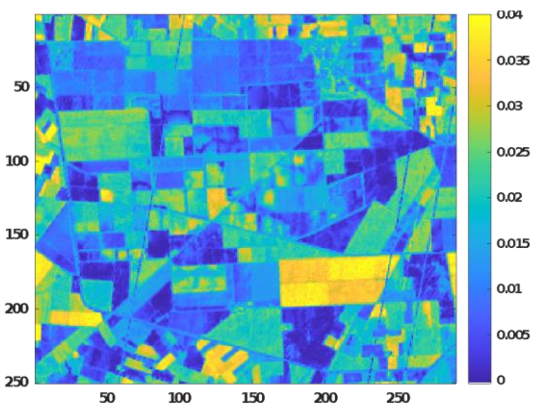
CDMC



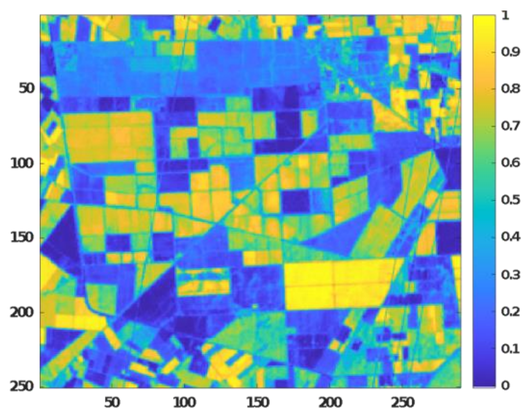
CNC



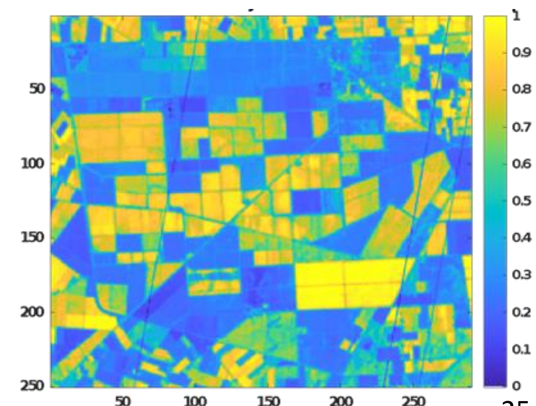
CCLC



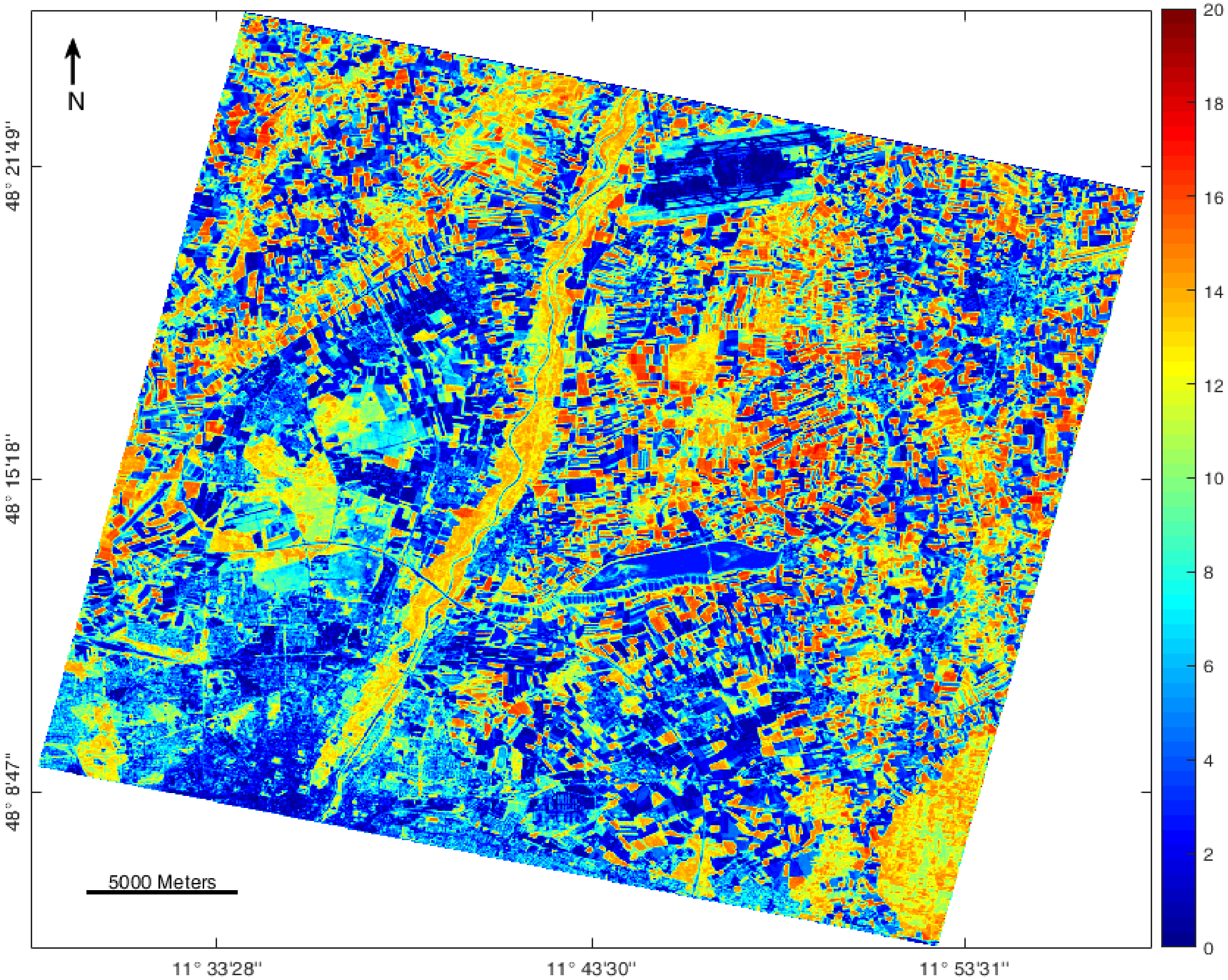
FAPAR



FVC

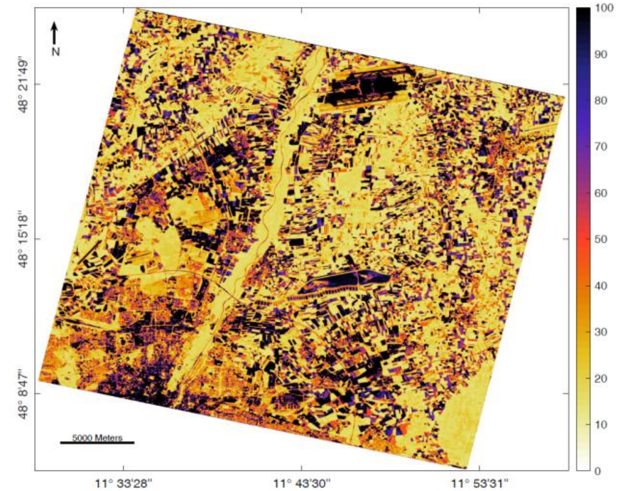


CNC (g/m^2) PRISMA

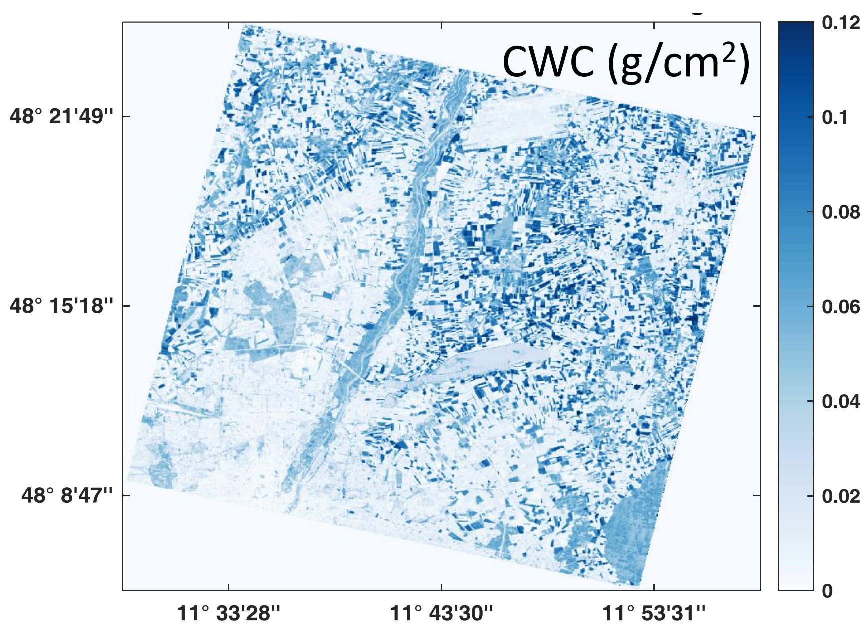


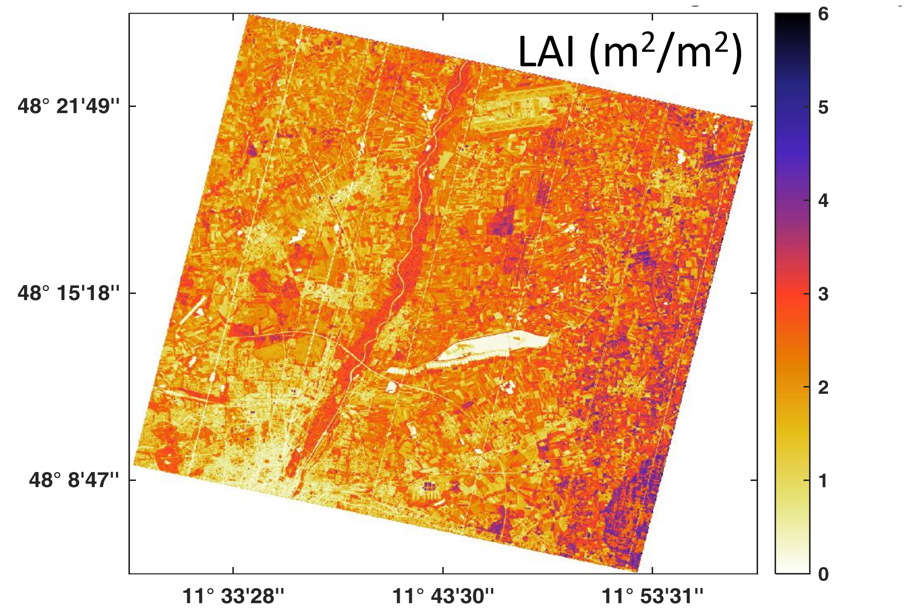
04/10/20220

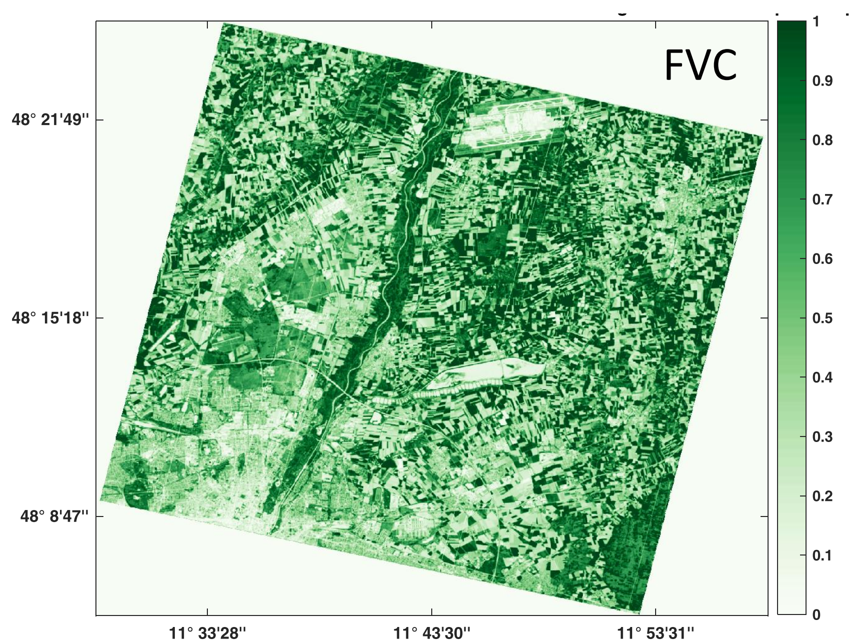
Relative uncertainty (%)

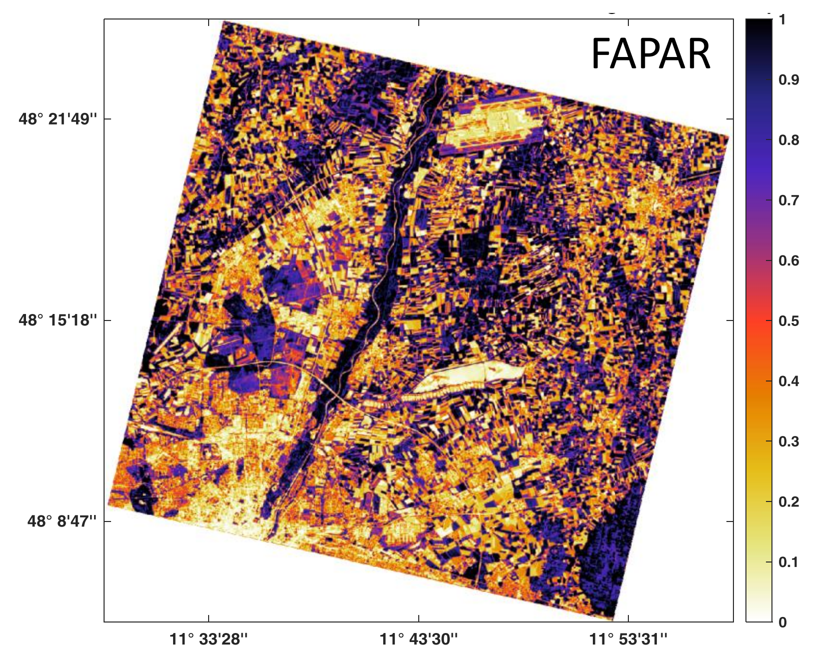


Some PRISMA vegetation traits maps

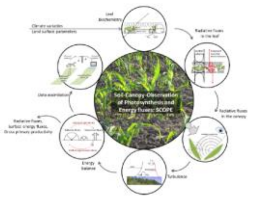


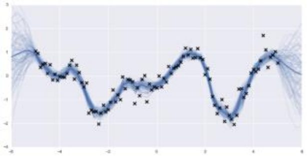


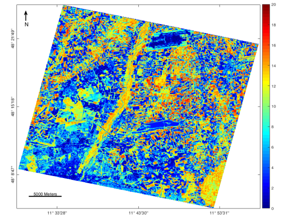


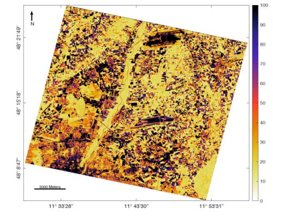


Conclusions & perspectives









- ✓Imaging spectrometry missions are reaching maturity with a.o. PRISMA, EnMap, **CHIME**, SBG
- ✓**GPR** appealing algorithm for new-generation vegetation models: robust, fast, uncertainties
- ✓**Hybrid models** developed with LUTs coming from SCOPE and PROSPECT-PRO-SAIL models
- ✓**CHIME GPR models** prepared and tested to simulated, aiborne and PRISMA images
- ✓Some GPR models (v.1.7) showed robustness: **CNC, FAPAR, FVC.** Others need some more work.
- ✓Further efforts required to develop robust models for all variables: key lies in quality training data. Trade-off between **generic/customized/size**

Thanks!

Questions?