The ARTMO Retrieval Toolboxes for Optimized and Automated Vegetation Properties Mapping from Sentinel-2 Data

J. Verrelst*[1], J.P. Rivera[2], J. Sanchis-Muñoz[1], M. Pereira-Sandoval[1], J. Delegido[1] & J. Moreno[1]
[1]: Image Processing Laboratory (IPL), University of Valencia, Spain; [2]: National Council of Science and Technology,CONACyT-Universidad Autónoma de Nayarit, UAN, Mexico

Laboratorio de Procesado de Imágenes

*Jochem Verrelst
Image Processing Laboratory
University of Valencia, Spain
Jochem.verrelst@uv.es

1. Introduction

New retrieval algorithms for Sentinel-2

The Copernicus Sentinel-2 (S2) satellite missions are designed to provide globally-available information on an operational basis for services and applications related to land. S2 is configured with improved spectral capabilities, which enable **improved and robust algorithms for biophysical variable retrieval**. This work present an overview of state-of-the-art retrieval methods dedicated to the quantification of terrestrial biophysical parameters. In all generality, retrieval methods can be categorized into three families: (*i*) ***parametric regression***, (*ii*) ***non-parametric regression***, and (*iii*) ***Inversion methods***.

We have recently developed 3 retrieval toolboxes within the **ARTMO software package** (**http://ipl.uv.es/artmo/**) that provide a suite of methods of these three families. As such, consolidated findings can be achieved about which type of retrieval method is most accurate, robust and fast.

As a case study, the most promising retrieval method is applied to a real S2 image to map LAIgreen and LAIbrown.

Objective:

To evaluate systematically 3 families of biophysical parameter retrieval methods for improved LAI estimation by using a local dataset (SPARC). Then, to apply the best performing method to a S2 image to map synergy of LAIgreen and LAIbrown.

2. Data & Experimental setup

Ground truth data (training & validation):

- SPARC dataset (Barrax, Spain): **103 LAI points** over various crop types and phenological stages.

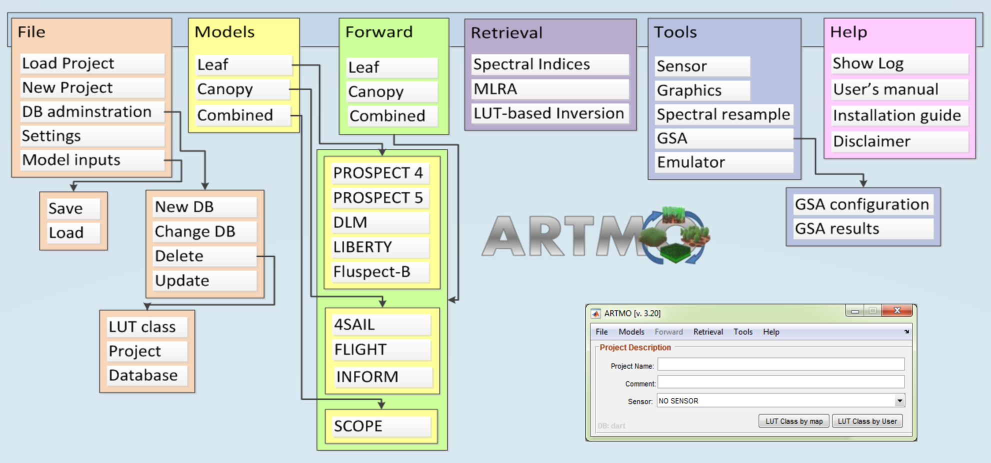
Sentinel-2 test image

- Rio Colorado valley of Buenos Aires, Argentina (13/01/2016)
- Atmospherically corrected with **Sen2Cor**

Band #	B1	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8a**	B9	B10	**B11**	**B12**
Band center (nm)	443	**490**	**560**	**665**	**705**	**740**	**783**	**842**	**865**	945	1375	**1610**	**2190**
Band width (nm)	20	**65**	**35**	**30**	**15**	**15**	**20**	**115**	**20**	20	30	**90**	**180**
Spatial resolution (m)	60	**10**	**10**	**10**	**20**	**20**	**20**	**10**	**20**	60	60	**20**	**20**

Experimental setup:

- Only S2 bands of 10 m (coarse-grained to 20 m) and 20 m were used.
- 50% of data (ground truth & associated S2 spectra) for training (Spectral Indices, MLRA) and **50% for validation (same for all retrieval approaches).**
- Comparison through goodness-of-fit measures: R^2, RMSE, NRMSE



7. Conclusions

With the ambition of delivering improved biophysical parameters retrieval (e.g. LAIgreen, LAIbrown) from Sentinel-2 (20 m), three families of retrieval methods have been systematically analyzed against the same validation dataset (SPARC, Barrax, Spain). It led to the following conclusions:

Parametric - Spectral Indices: All 2-, 3- and 4-band combinations according to normalized difference (ND) formulation have been analyzed. A 4-band index with bands in SWIR was best performing, **but the required 10% error was not reached (NRMSE: 16.0%; R2: 0.79). Most critically, the absence of uncertainty estimates implies that vegetation indices cannot be considered as reliable. Fast mapping (1s.).**

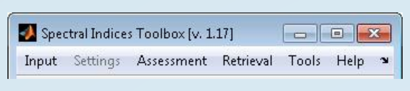
Nonparametric – MLRAs: Machine learning regression algorithms are powerful and also fast. Several yielded **high accuracies with errors below 10%. Particularly GPR (NRMSE: 8.2; R^2: 0.91) is attractive as it delivers insight in relevant bands and associated uncertainties. Hence, unreliable retrievals can be masked out. Fast mapping (7s.).**

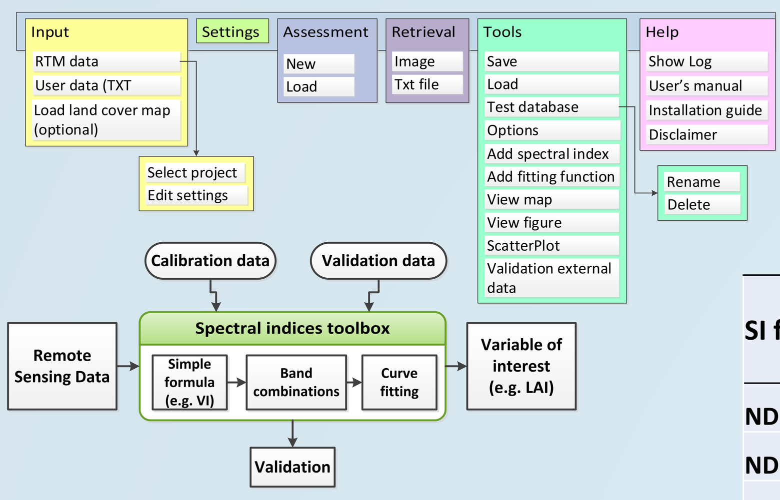
LUT-based Inversion: A PROSAIL LUT of 100000 simulations has been prepared and various cost functions and regularization options were applied. **Best cost functions performed alike as best 2-band indices (16.6%; R2: 0.76). Because pixel-by-pixel inverted against a LUT table, biophysical parameter mapping went unacceptably slow (> 25h.).**

Thanks to the new S2 bands in the SWIR, not only LAIgreen but also LAIbrown can be mapped. GPR was evaluated a most promising. Moreover, the GPR associated uncertainties can function to mask out unreliable retrievals (e.g. >40%).

3. (*i*) Parametric regression: Spectral Indices – LAIgreen

ARTMO's ***Spectral Indices* (SI)** toolbox:





Very fast: 0.004 sec per SI model, >10 thousand SI models in 43 s.)

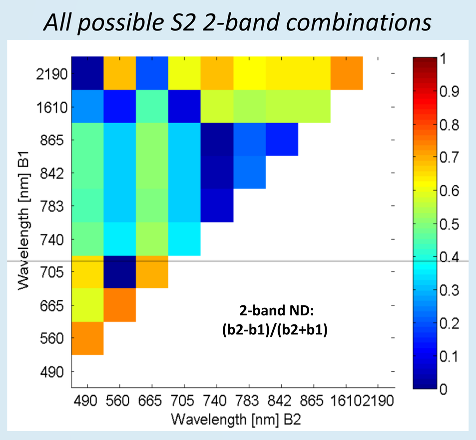
Best validated SIs (50% validation data) ranked according to R^2:

SI formulation	Best band combination (B1, B2, B3, B4)	RMSE	NRMSE	R^2
ND 4-bands: (b2-b1)/(b3+b4)	560, 2190, 1610, 1610	0.69	16.01	**0.79**
ND 3-bands: (b2-b1)/(b2+b3)	560, 2190, 740	0.70	16.74	**0.79**
ND 2-bands: (b2-b1)/(b2+b1)	665, 560	0.76	16.86	**0.74**
SR 2-bands: (b2/b1)	665, 560	0.77	20.36	**0.74**

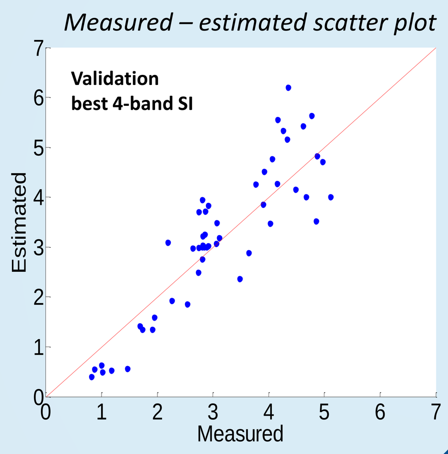
A 4-band SI with bands in green and SWIR best validated. Green and red bands led to best 2-band SI.

In the ***Spectral Indices*** module the predictive power of **all possible 2-, 3- or 4-band combinations** according to an Index formulation (e.g. simple ratio (SR), normalized difference (ND)) to a biophysical parameter can be evaluated.

All possible S2 2-band combinations



Measured – estimated scatter plot

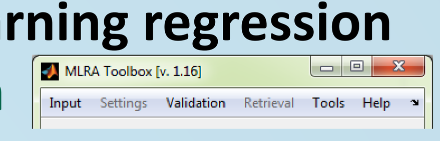


Applied SI formulations:

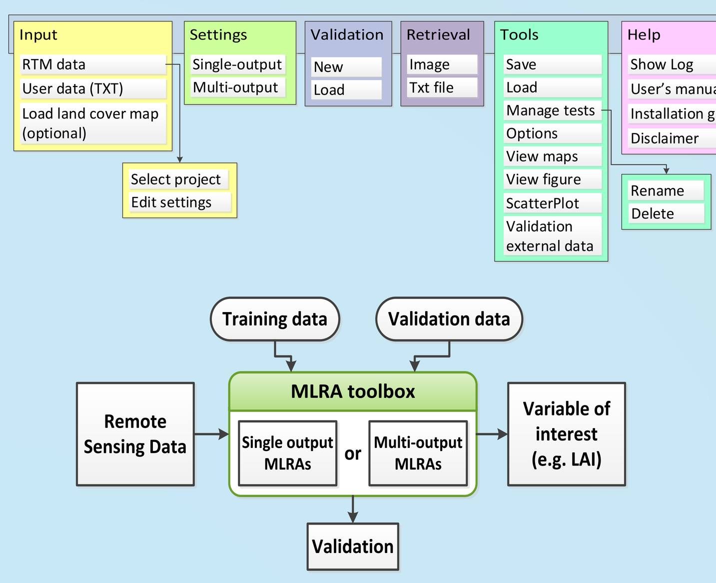
- 2-band SIs:
  - **SR (B2/B1)** (10^2 combinations)
  - **ND (B2-B1)/(B2+B1)** (10^2 combinations)
- ND 3-band **(B2-B1)/(B2+B3)** (10^3 combinations)
- ND 4-band **(B2-B1)/(B3+B4)** (10^4 combinations)

A Linear regression was applied.

4. (*ii*) Nonparameteric regression: Machine learning regression algorithms (MLRAs) - LAIgreen

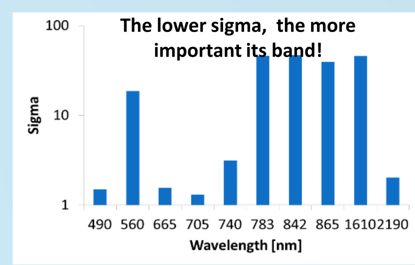


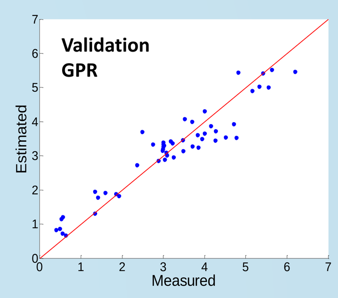
ARTMO's ***Machine Learning Regression Algorithms (MLRA)*** toolbox:



50% validation results ranked according to R^2:

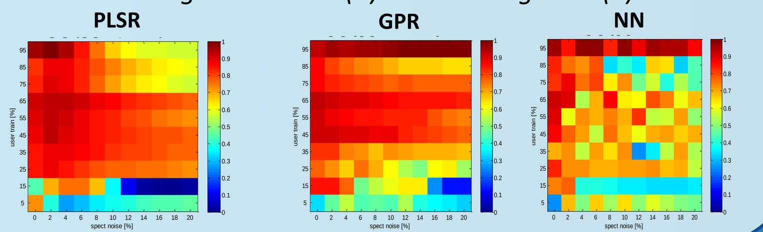
MLRA	RMSE	NRMSE	R^2	Time (s.)
Kernel ridge Regression	0.41	7.04	0.93	0.063
Gaussian Processes Regression	**0.47**	**8.17**	**0.91**	0.788
Neural Network	0.46	7.99	0.91	6.069
VH. Gaussians Processes Regression	0.48	8.30	0.90	2.473
Extreme Learning Machine	0.48	8.26	0.89	0.061
Bagging trees	0.58	10.03	0.87	1.296
Relevance vector Machine	0.59	10.20	0.86	16.501
Least squares linear regression	0.56	9.62	0.86	0.002
Boosting trees	0.70	12.10	0.79	1.100
Partial least squares regression	0.71	12.16	0.78	0.008
Regression tree	0.78	13.46	0.72	0.006
Principal components regression	0.79	13.70	0.71	0.002



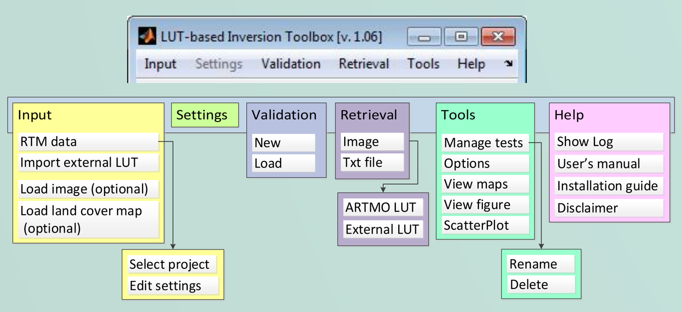


- **About 15 MLRAs** have been implemented: e.g., neural nets (NN), kernel ridge regression (KRR) , Gaussian Processes regression (GPR), principal component regression (PCR), partial least squares regression (PLSR) , regression trees (RT) (See also http://www.uv.es/gcamps/code/simpleR.html).
- Options to add noise and split training- validation data are provided.

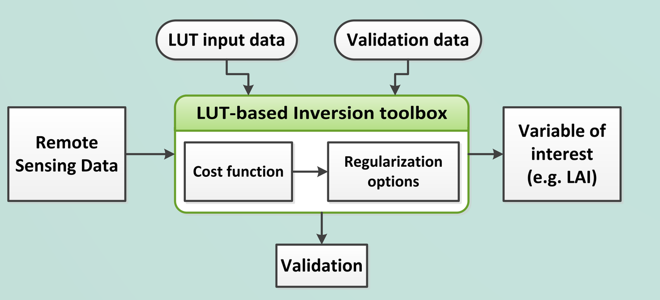
*Examples of **robustness**: validation results (R^2) along increasing noise levels (X) and training data (Y):*



ARTMO's ***Inversion*** toolbox:



5. (*iii*) Inversion of canopy RTM through cost functions - LAIgreen



Retrieval of biophysical parameters through LUT-based inversion.

- LUTs prepared in ARTMO and loaded in ***Inversion*** module
- **More than 60 cost functions** have been implemented.
- Various **regularization options**: **adding noise, mean of multiple solutions, data normalization**.

PROSAIL LUT (sub-selection 100000):

Model Parameters		Units	Range	Distribution
***Leaf variables:* PROSPECT-4**				
N	Leaf structure index	unitless	1.1	-
LCC	Leaf chlorophyll content	(µg/cm²)	5–75	Gaussian ($\bar{x}$ 35, SD: 30)
C_m	Leaf dry matter content	(g/cm²)	0.001–0.03	Uniform
C_w	Leaf water content	(cm)	0.002–0.05	Uniform
***Canopy variables:* 4SAIL**				
LAI	Leaf area index	(m²/m²)	0.1–7	Gaussian ($\bar{x}$: 3, SD: 2)
α_{soil}	Soil scaling factor	unitless	0	-
ALA	Average leaf angle	(°)	40–70	Uniform
HotS	Hot spot parameter	(m/m)	0.05–0.5	Uniform
skyl	Diffuse incoming solar radiation	(fraction)	0.05	-
θ_s	Sun zenith angle	(°)	22.3	-
θ_v	View zenith angle	(°)	0	-
φ	Sun-sensor azimuth angle	(°)	0	-

Examples of cost functions:

Shannon (1948):

$$D(P,Q) = -\sum_{\lambda_1=1}^{\lambda_n}\left(\frac{p(\lambda_l)+q(\lambda_l)}{2}\right) log\left(\frac{p(\lambda_l)+q(\lambda_l)}{2}\right) + \frac{1}{2}\left(\sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l)log(p(\lambda_l)) + \sum_{\lambda_1=1}^{\lambda_n} q(\lambda_l)log(q(\lambda_l))\right)$$

Laplace distribution:

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} |p(\lambda_l) - q(\lambda_l)|$$

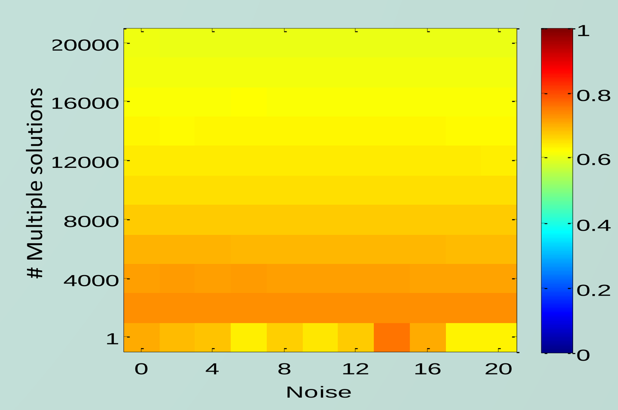
Pearson chi-square:

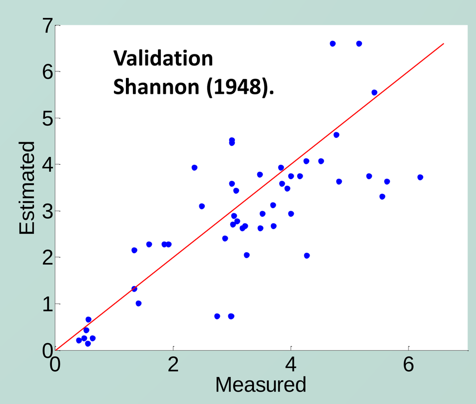
$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} \frac{(q(\lambda_l) - p(\lambda_l))^2}{p(\lambda_l)}$$

In total 5508 inversion strategies analyzed. **50% validation results** for best noise & multiple samples ranked according to R^2:

Cost function	% Noise	% multiple samples	RMSE	NRMSE	R^2	time (s.)
Shannon (1948)	**14**	**single best**	**0.96**	**16.56**	**0.76**	0.027
Laplace distribution	6	single best	0.86	14.74	0.74	0.021
Neyman chi-square	0	single best	0.89	15.31	0.74	0.005
Pearson chi-square	16	single best	1.03	17.74	0.73	0.005
Least absolute error	6	single best	0.89	15.28	0.72	0.005
Geman and McClure	16	2	0.83	14.36	0.71	0.007
RMSE	16	2	0.83	14.37	0.71	0.006
Exponential	16	2	0.85	14.66	0.71	0.008
K(x)=x(log(x))-x	20	single best	1.06	18.25	0.70	0.009
K(x)=(log(x))^ 2	0	2	1.01	17.40	0.69	0.012
K-divergence Lin	4	single best	2.60	44.84	0.64	0.009
Shannon entropy	6	2	1.15	19.82	0.60	0.013
Gen. Kullback-Leibler	10	2	1.20	20.63	0.58	0.013
Neg. Exp. disparity	0	4	1.04	17.96	0.58	0.007
Kullback-leibler	4	18	1.66	28.62	0.57	0.009
K(x)=log(x)+1/x	2	single best	2.07	35.65	0.55	0.012
Harmonique Toussaint	2	20	1.57	27.04	0.54	0.005
K(x)=-log(x)+x	2	2	1.77	30.52	0.49	0.012

Example of **robustness** (R^2) along increasing noise levels (X) and mean of multiple solutions (Y) in the Shannon (1948) Cost Function inversion:



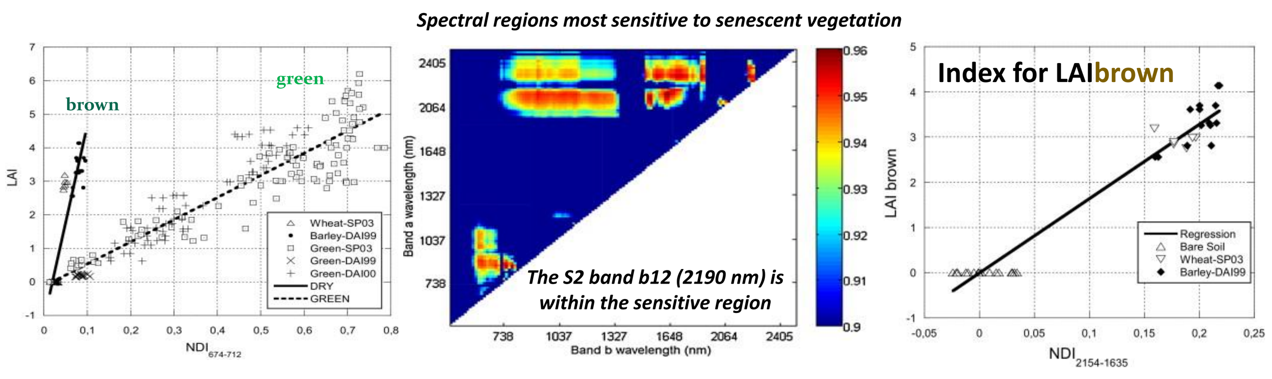


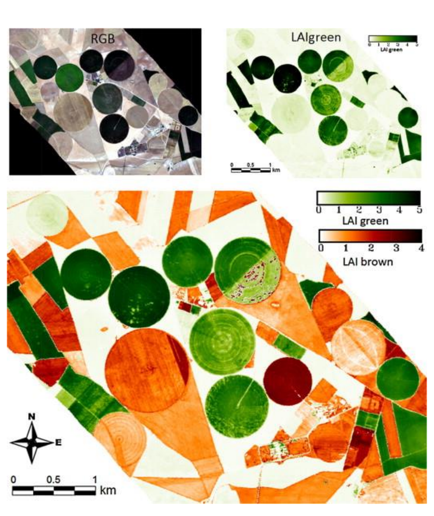
6. Application of GPR to Sentinel-2: towards operational mapping of LAIgreen and LAIbrown

LAI green/brown based on indices (Delegido *et al.*, 2014):

Traditionally, only LAIgreen is mapped. However, by making use of bands in the SWIR it is also possible to map senescent material. Thanks to the new S2 bands in the SWIR (b11, b12), opportunities are opened to map LAIbrown.

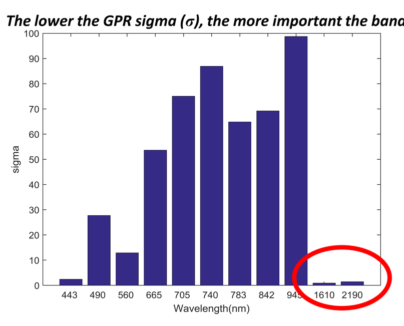
Spectral regions most sensitive to senescent vegetation





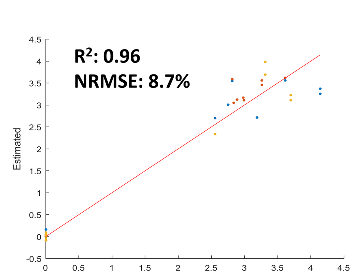
Beyond indices, as shown above LAIgreen can be most accurately predicted with machine learning (GPR: R^2: 0.91). Moreover, with GPR additional uncertainties are provided.

The lower the GPR sigma (σ), the more important the band.



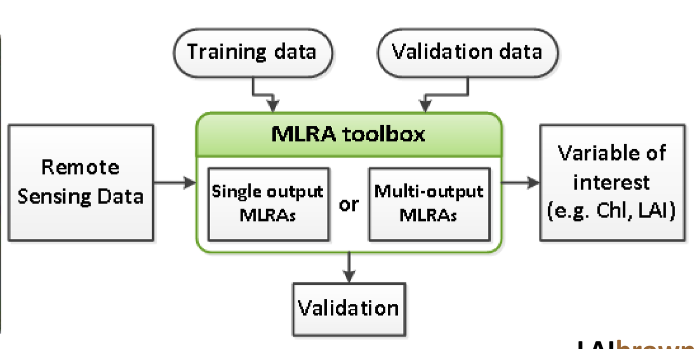
The same GPR method can be applied to LAIbrown:

The S2 SWIR bands are perfectly suited for LAIbrown estimation.

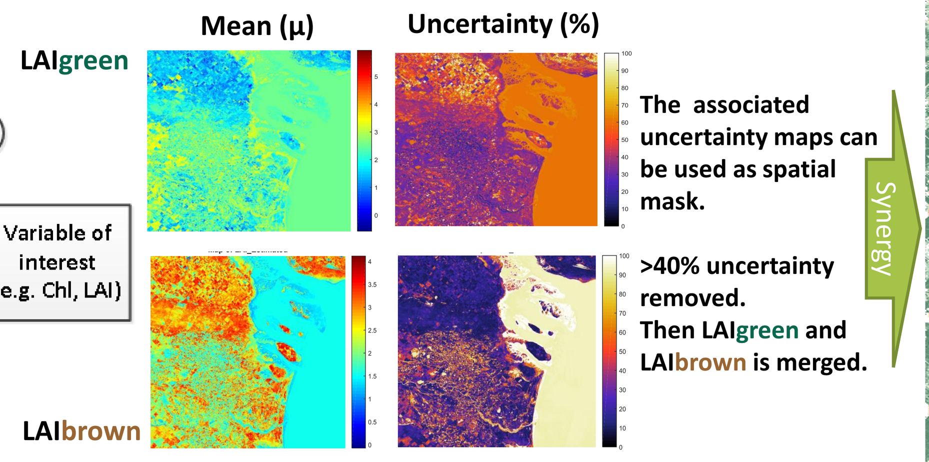


Both GPR models are created with ARTMO's MLRA toolbox. Apart from LAIgreen and LAIbrown estimates, also relative uncertainties provided.

Corrected S2 image.

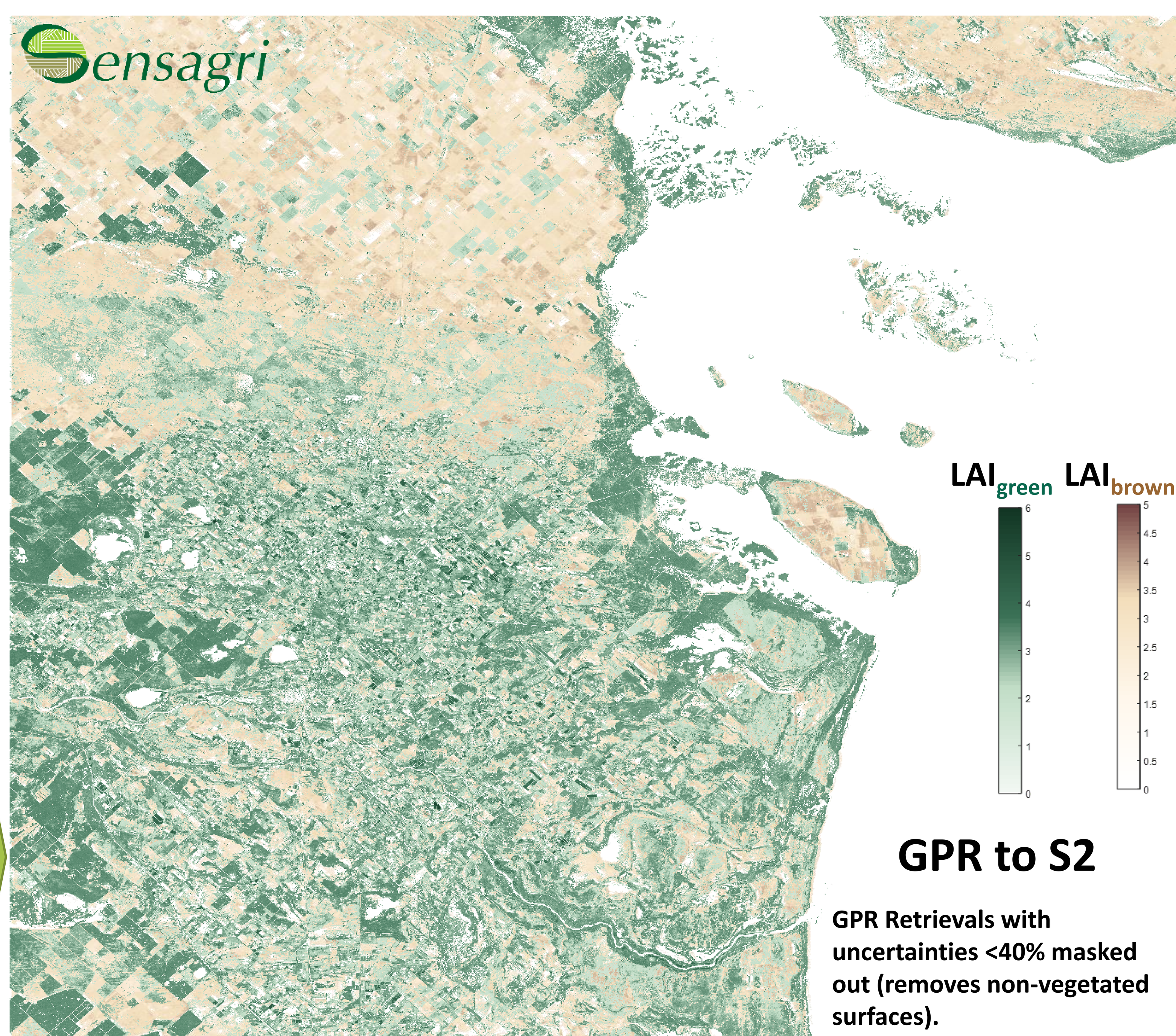


Mean (µ) **Uncertainty (%)**



The associated uncertainty maps can be used as spatial mask.

>40% uncertainty removed. Then LAIgreen and LAIbrown is merged.



GPR to S2

GPR Retrievals with uncertainties <40% masked out (removes non-vegetated surfaces).

HORIZON 2020 *The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement no 730074*

- Delegido, j., Verrelst, j., Rivera, j.A., Ruiz-verdú, a., Moreno, j. (2015). Brown and Green LAI mapping through spectral indices. International Journal of Applied Earth Observation And Geoinformation, 35, p. 350-358.
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P. Veroustraete, F., Clevers, J.G.P.W., Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. ISPRS Journal of Photogrammetry and Remote Sensing, 108, p. 273-290.
- Verrelst, J., Rivera, J.P. Veroustraete, F., Muñoz-Marí, J., Clevers, J.G.P.W., Camps-Valls, G., Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. ISPRS Journal of Photogrammetry and Remote Sensing, 108, p. 260-272.