UNIVERSITÀ DEGLI STUDI DI MILANO

Assessing the potential of Machine Learning Regression Algorithms for improving crop Leaf Area Index estimation from Sentinel-2 Data

Giorgio Ragaglini[1], Margherita De Peppo[2], Alberto Mantino[3], Antonella Tornato[4], Andrea Taramelli[4,5] and Mirco Boschetti[2]

1 Department of Agricultural and Environmental Sciences - Production, Territory, Agroenergy, Università degli Studi di Milano, Via Celoria 2, 20133 Milano, Italy;

2 Institute for Electromagnetic Sensing of the Environment, Italian National Research Council, Via Bassini 15, 20133 Milan, Italy;

3 Sant' Anna School of Advanced Studies, Institute of Life Sciences, Piazza Martiri della Libertà 33, 56127 Pisa, Italy;

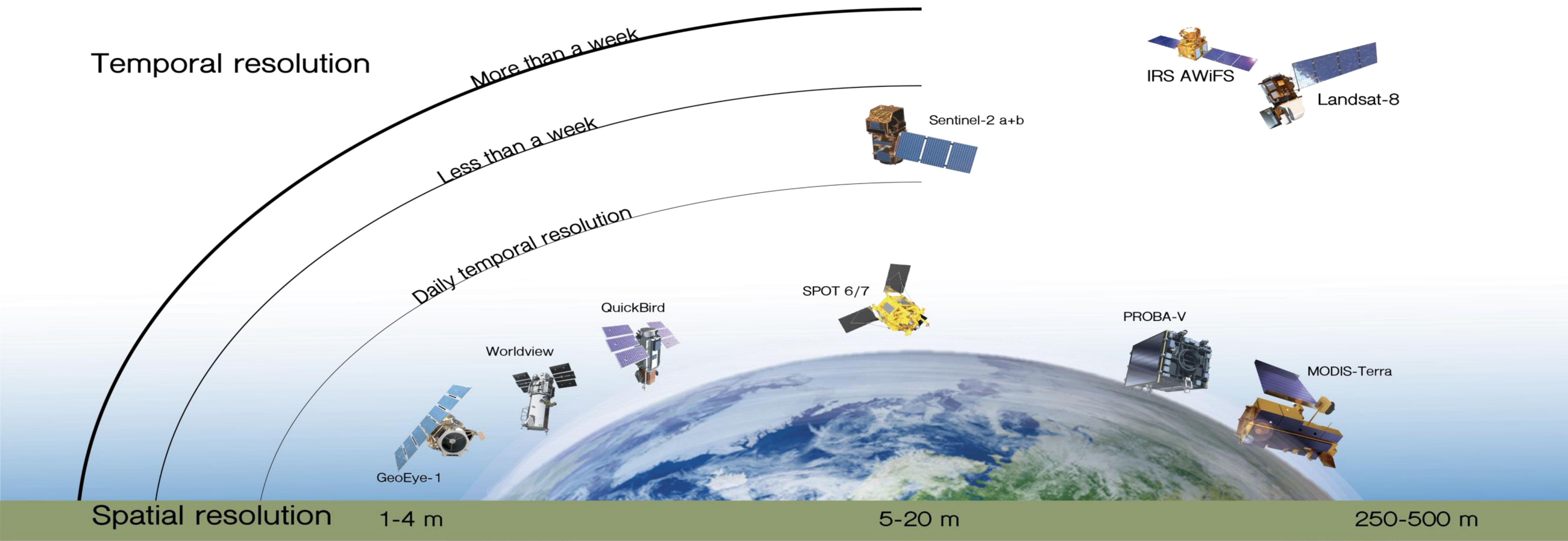
4 Institute for Environmental Protection and Research (ISPRA), via Vitaliano Brancati 48, 00144 Rom , Italy;

5 Institute for Advanced Study of Pavia (IUSS), Palazzo del Broletto, Piazza della Vittoria 15, 27100 Pavia, Italy

Concept and background

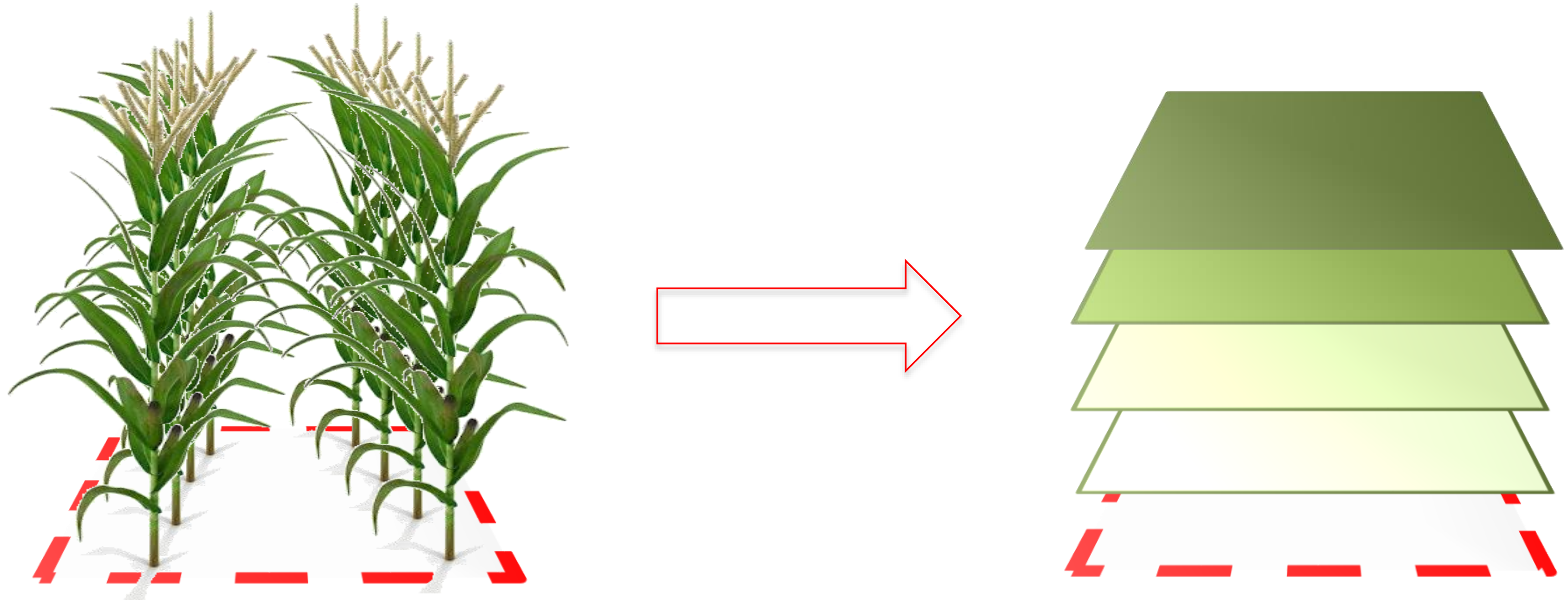
Spaceborne earth observation systems offer a great opportunity to collect multispectral reflectance imagery as a fundamental input to deliver cost-effective support services for agriculture.

These services mainly focuses on the **retrieval of quantitative information** of vegetation **canopy** structure, dimension and status such as **Leaf Area Index (LAI)** and Leaves Chlorophyll Content (LCC).



Concept and background

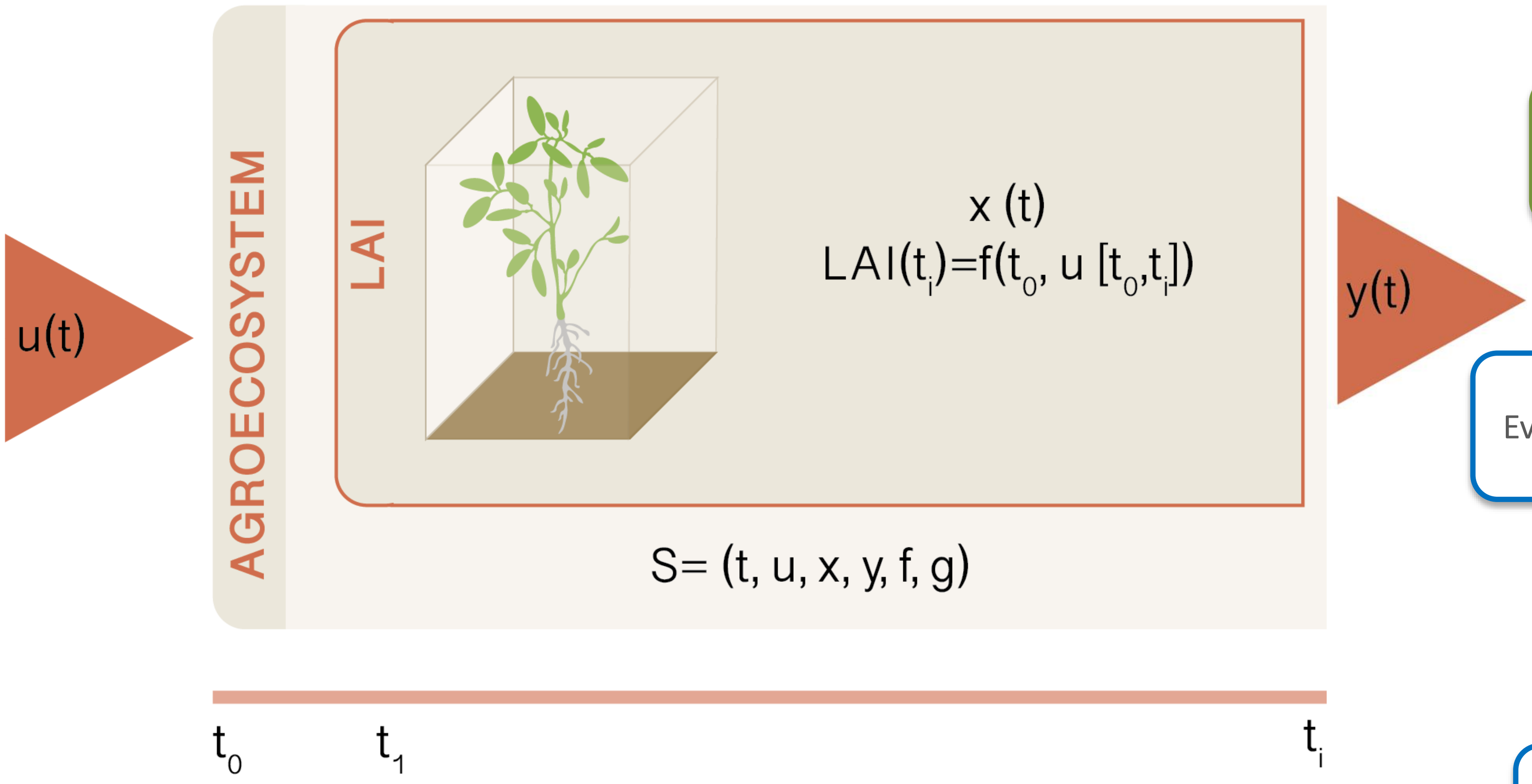
Leaf Area Index (LAI) is leaf area per unit of ground surface

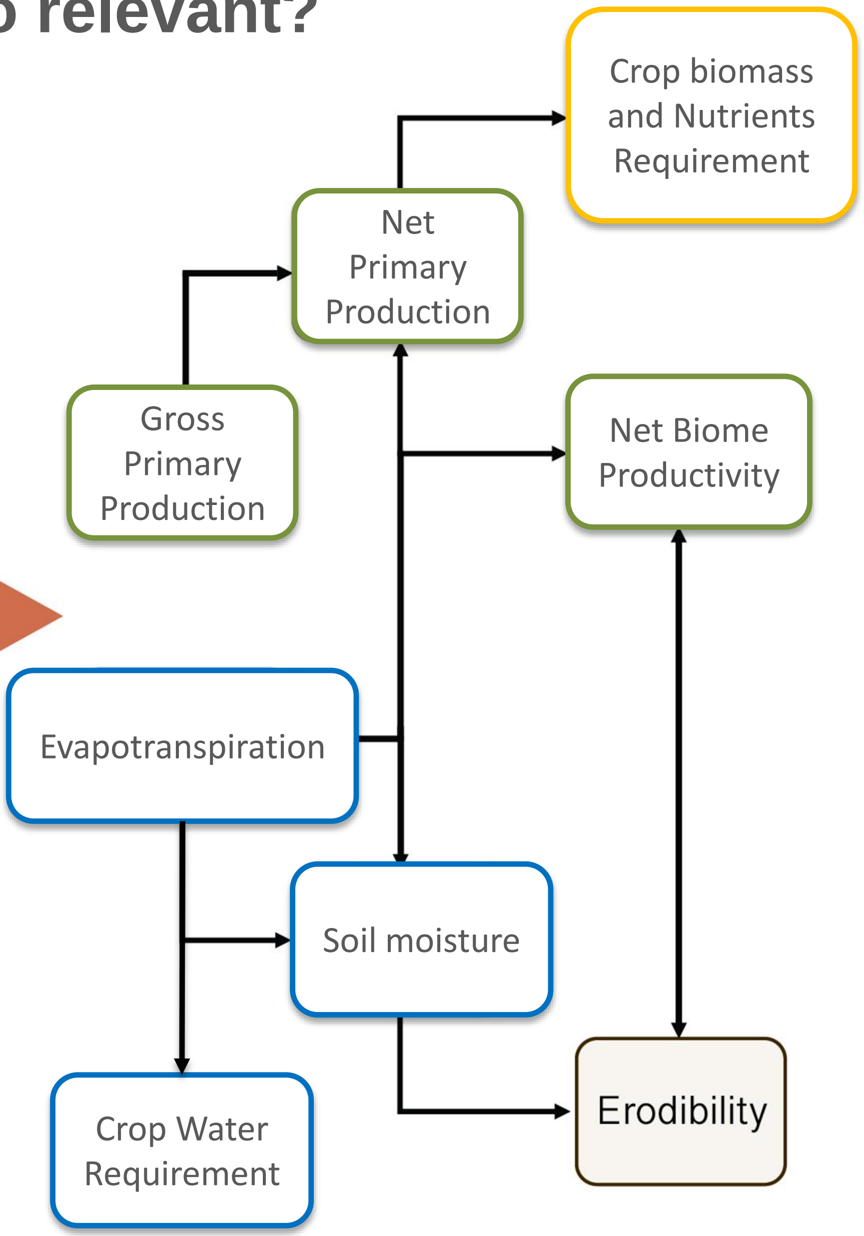


Plant Area Index (PAI)
Green LAI
Brown LAI

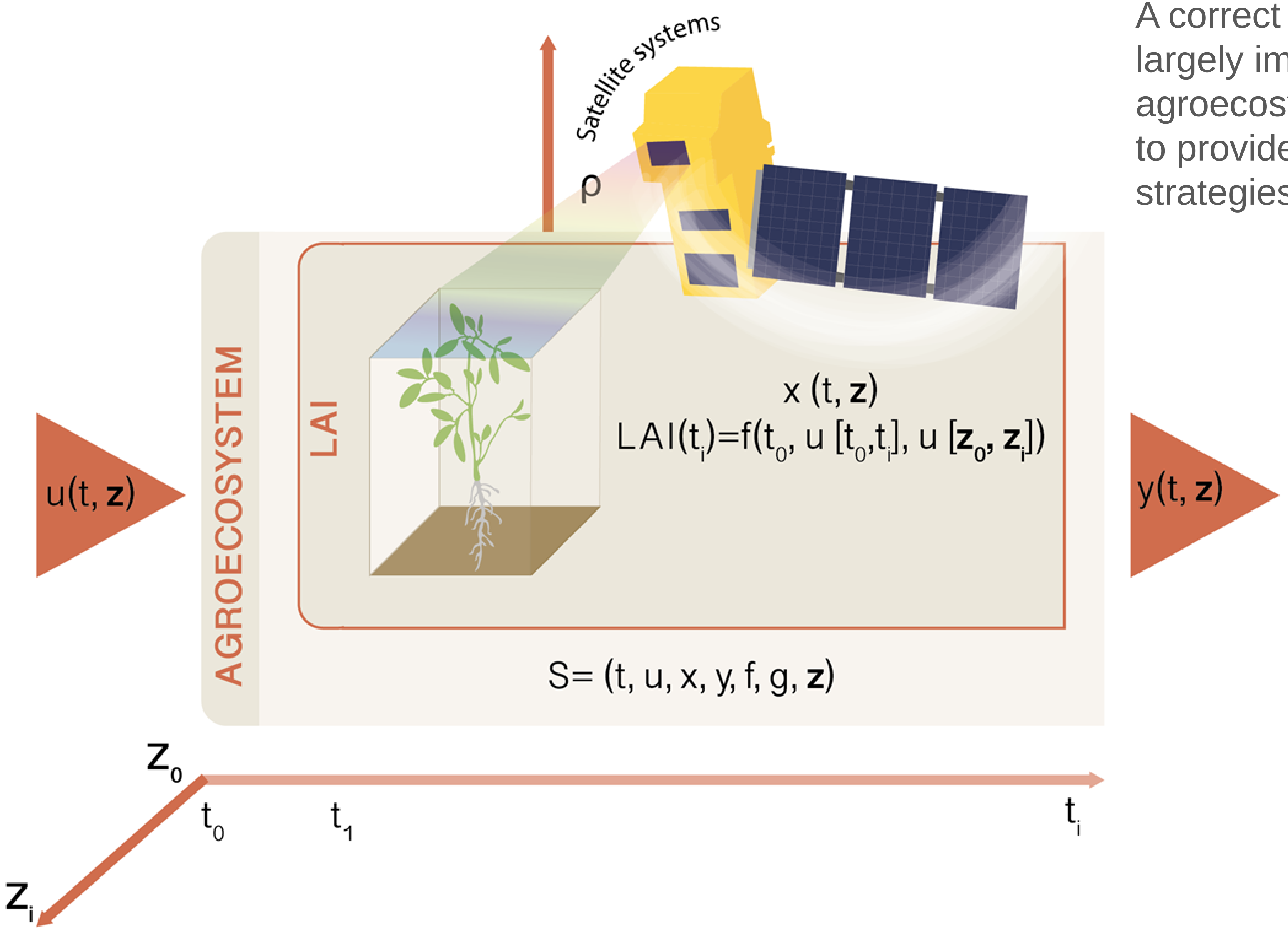
Concept and background

Why LAI is so relevant?



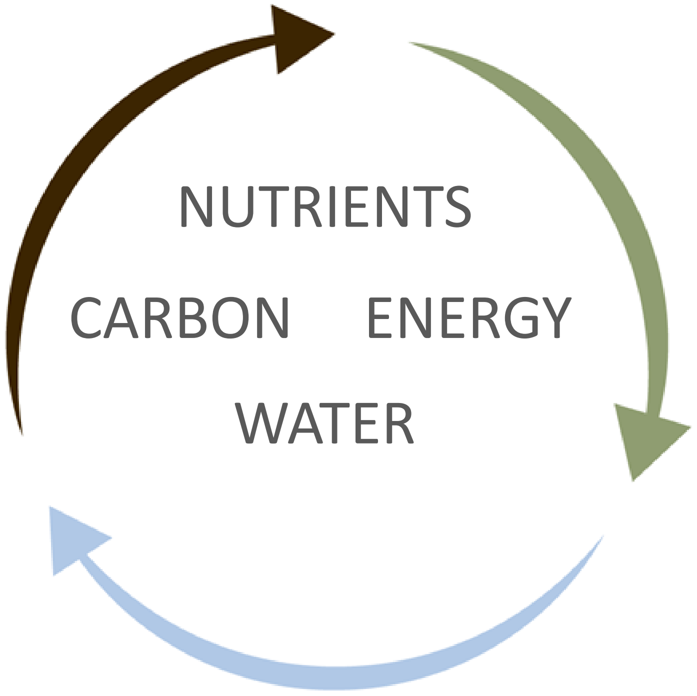


Concept and background



Opportunities

A correct estimation of LAI from satellite would largely improve our ability to quantify agroecosystem flows over time and space and to provide support for optimized management strategies at field and landscape level

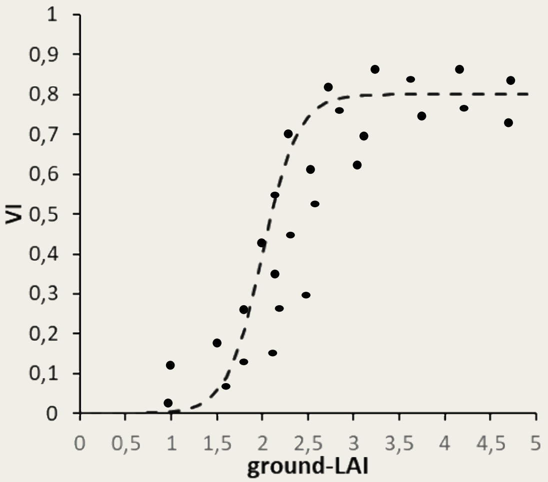


Concept and background

Which are current and most spread approaches for LAI estimation by Multispectral Satellite Images?

VIs-Based Parametric regression

Empirical relationship between a combination of few filtered spectral bands and LAI



$$NDVI = \frac{\rho865 - \rho665}{\rho865 + \rho665}$$

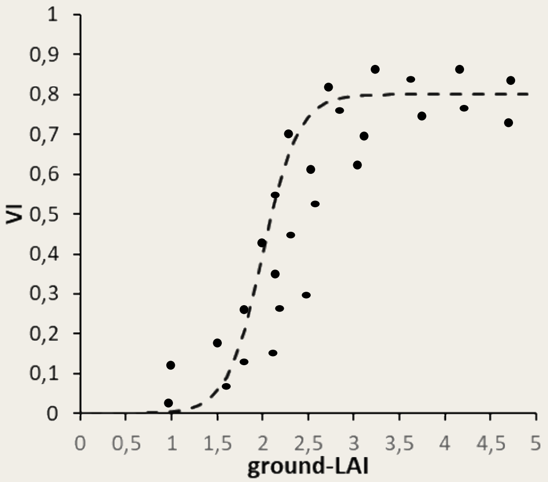
- Crop and site-specific, repeated calibrations are required
- Not fully efficient to represent the **within field variability** of a crop
- Not fully efficient to represent the **among fields variability** which occurs in a landscape mosaic with more than one crop (different VIs responds differently to different crops)

Concept and background

Which are current and most spread approaches for LAI estimation by Multispectral Satellite Images?

VIs-Based Parametric regression

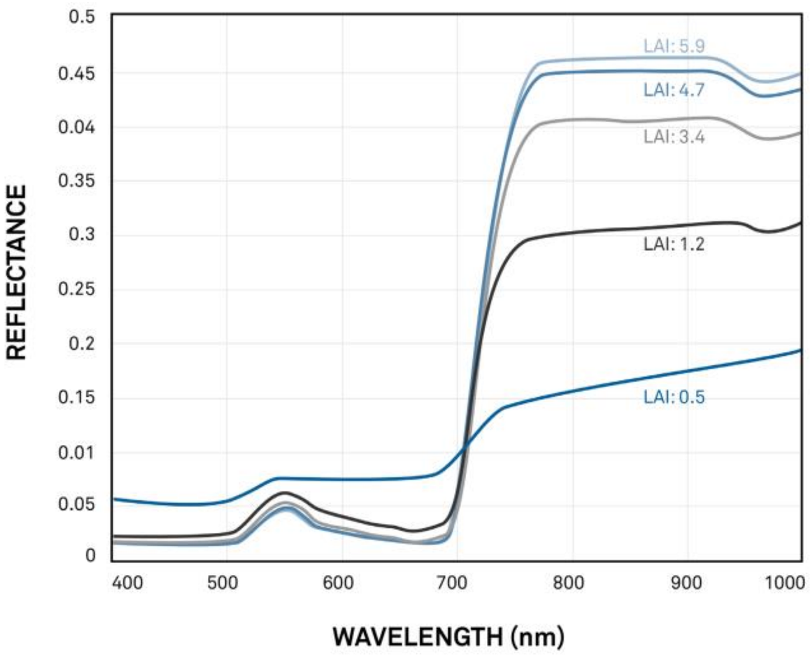
Empirical relationship between a combination of few filtered spectral bands and LAI

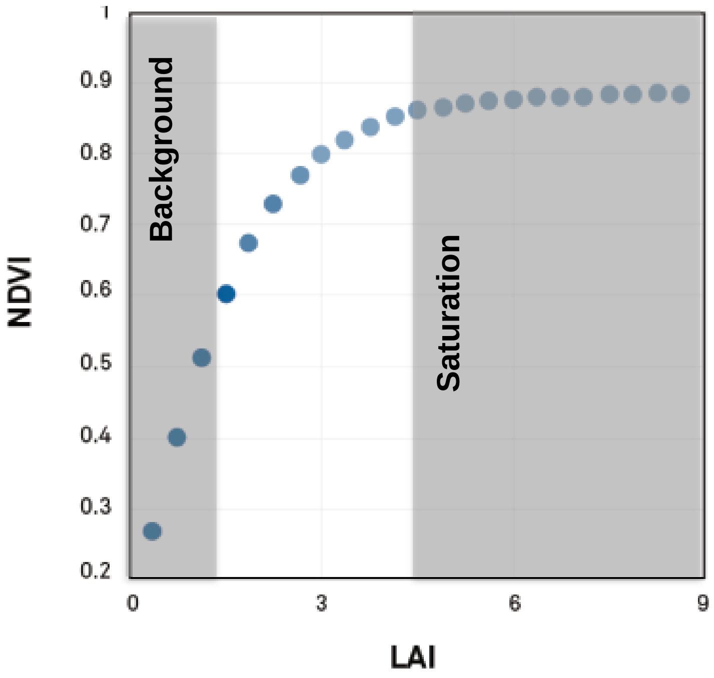


$$NDVI = \frac{\rho 865 - \rho 665}{\rho 865 + \rho 665}$$

(Camacho F. *et al.* 2017).

The range of values which may be exploited from a VI to estimate LAI is limited by background (lower limit) and saturation (upper limit) effects



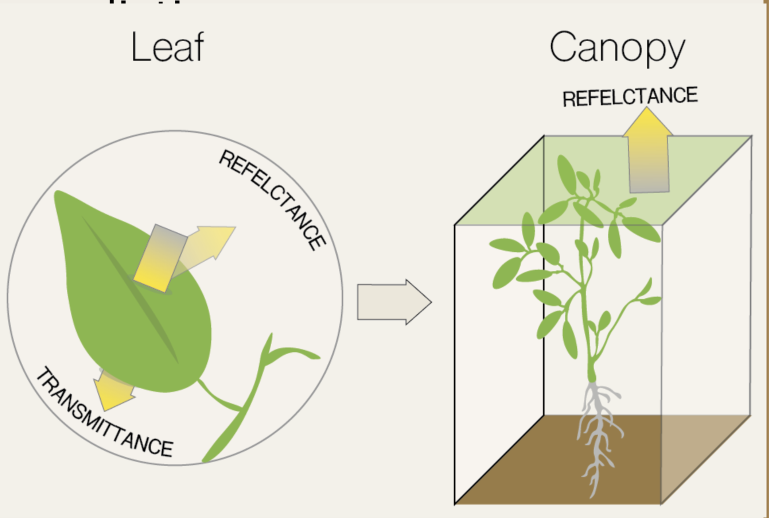


Concept and background

Which are current and most spread approaches for LAI estimation by Multispectral Satellite Images?

Radiative Transfer Model

Physical model that simulate the interactions between vegetation, atmosphere and



Difficult implementation in automated processes

Many parameters

Crop-specific calibration is required in order to be applied for the provision of services

However it is successfully applied in the FarmStar N fertilization support system

(Baret, F. & Buis, S. 2008)

Concept and background

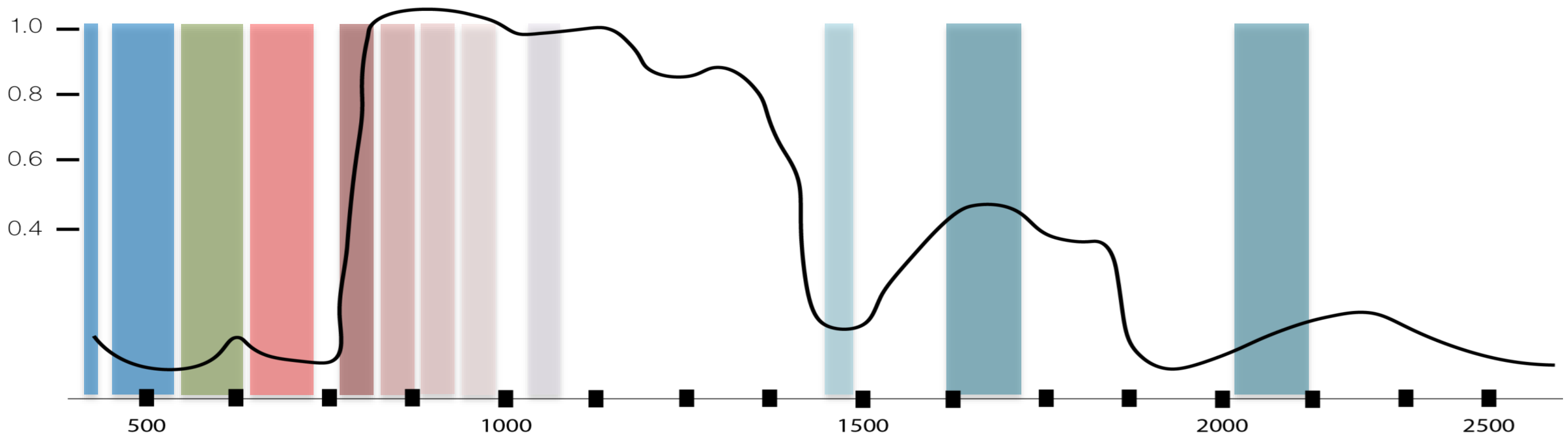
Which are current and most spread approaches for LAI estimation by Multispectral Satellite Images?

Recent introduction

Promising results

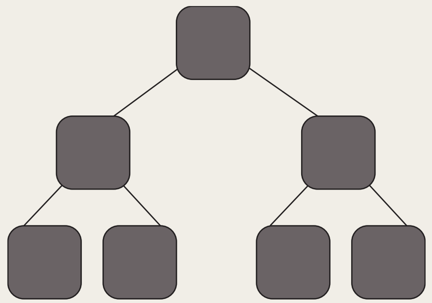
Versatility and adaptability

Full exploitation of all spectral bands



Non-parametric regression

Advanced techniques that search for relationship between spectral data and biophysical variables (LAI)



Machine Learning Regression Algorithms

(Verrelst, J. *et al.* 2015)

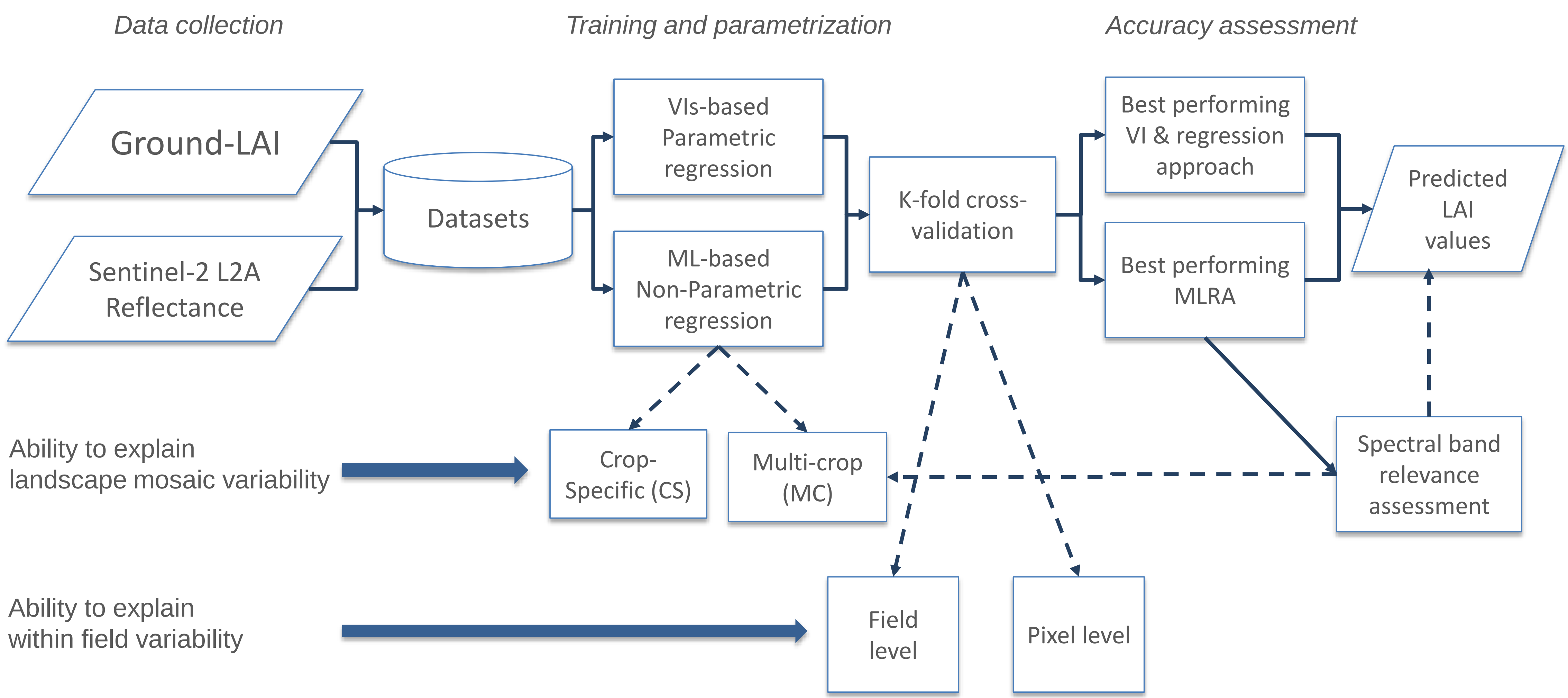
Research questions

Are MLRA approaches more effective then VIs based one for LAI retrieval from Sentinel 2A MSI?

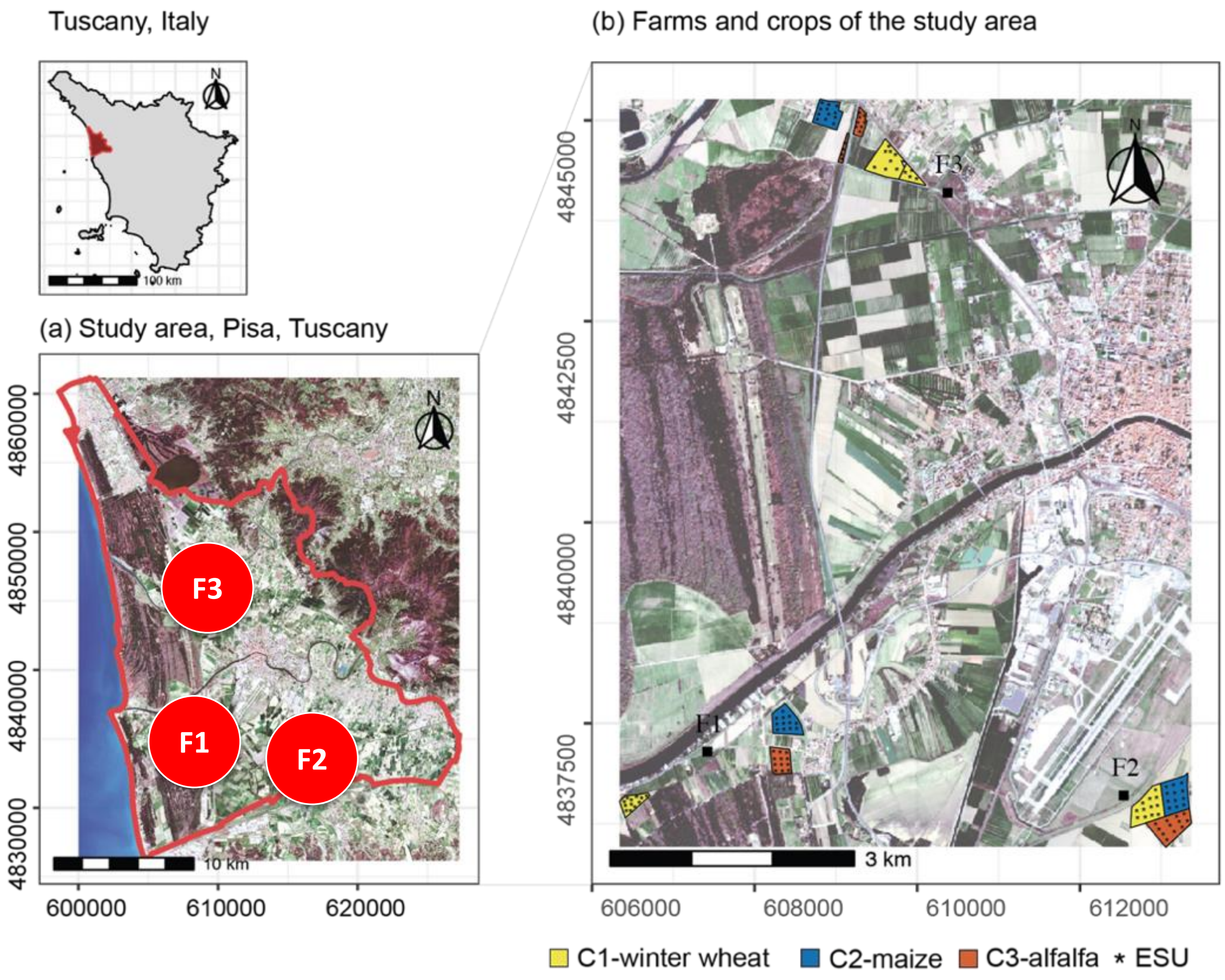
Are they more accurate in representing LAI within field variability?

Are they more suited to predict LAI from a landscape mosaic of multi crop?

Overall approach



Study area and Ground-LAI sampling scheme



3 Farms

3 Crops per farm

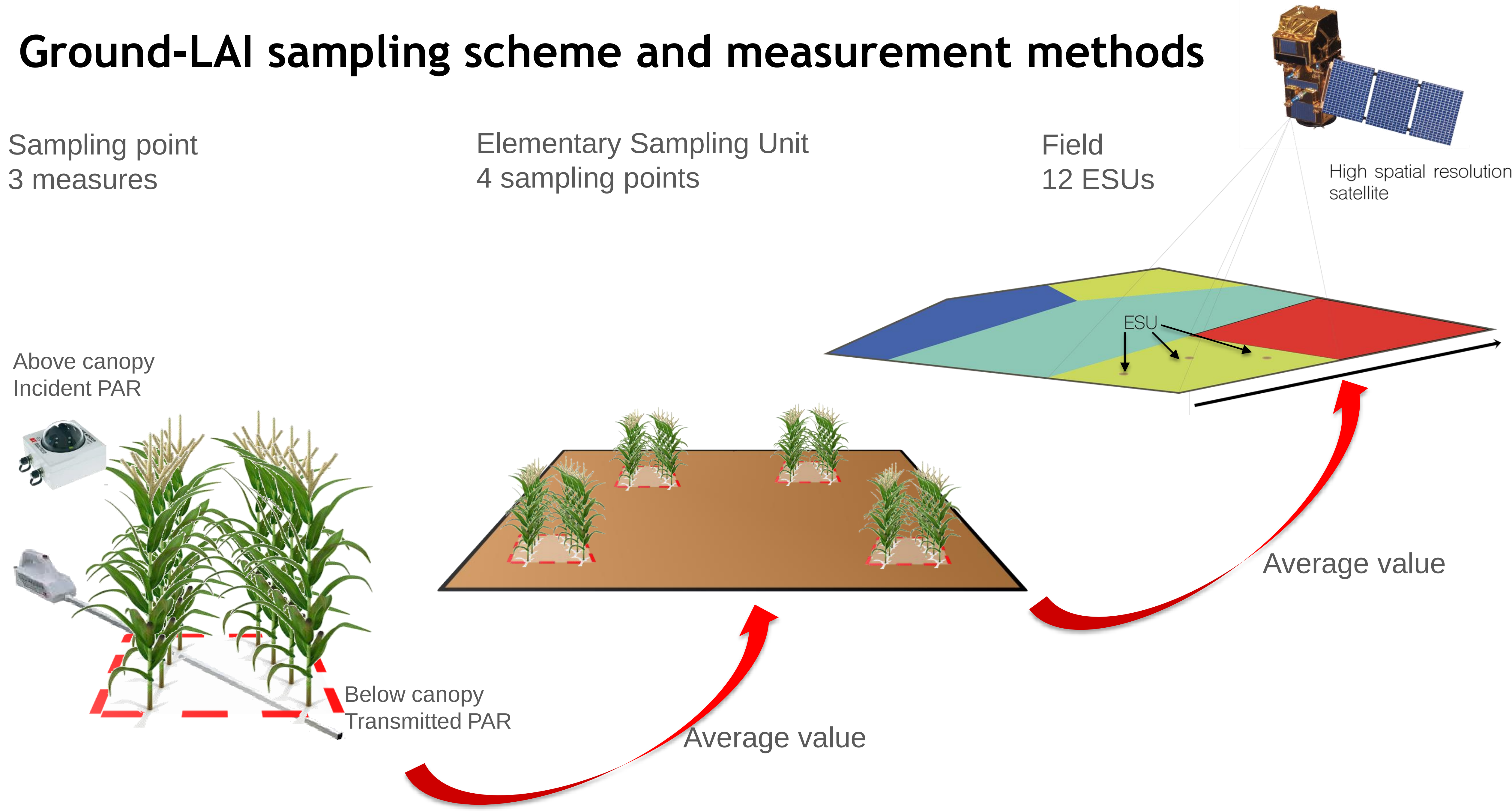
C1 – winter wheat

C2 – maize

C3 – alfalfa

* - Elementary Sampling Unit

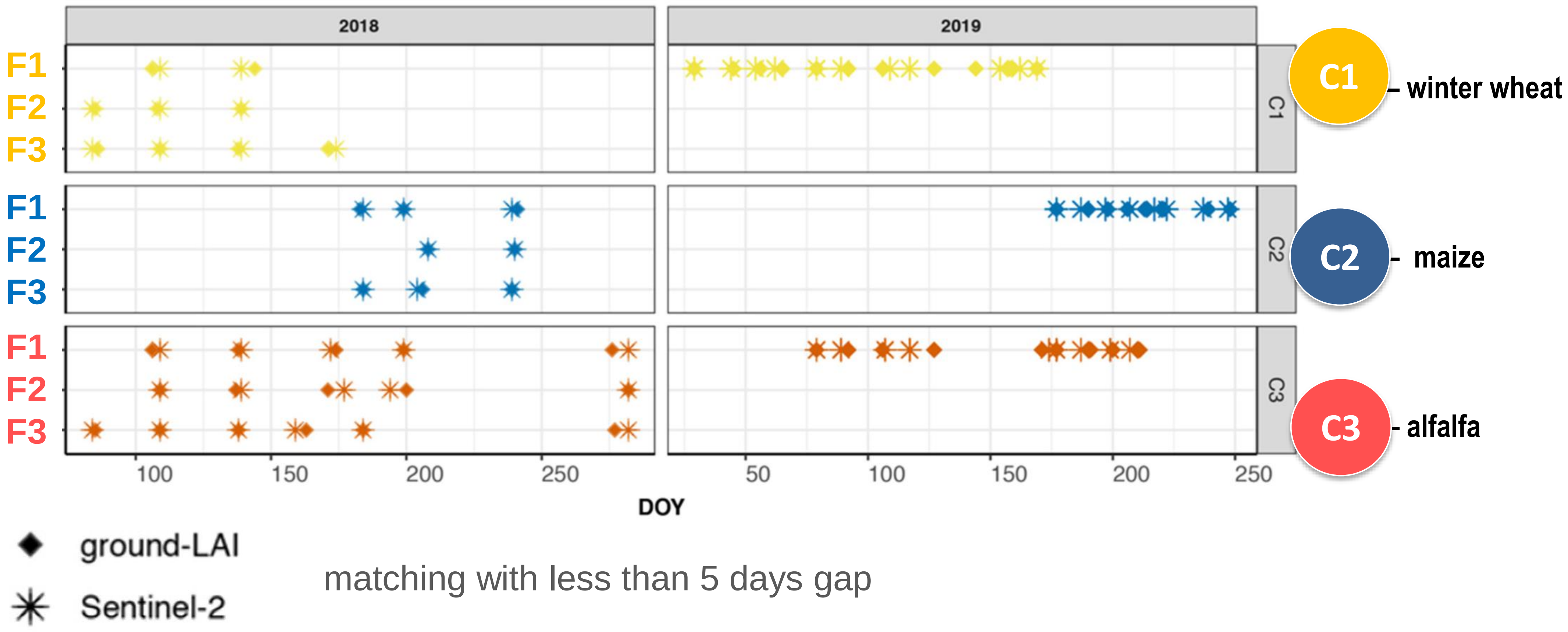
Ground-LAI sampling scheme and measurement methods



Ground-LAI sampling times

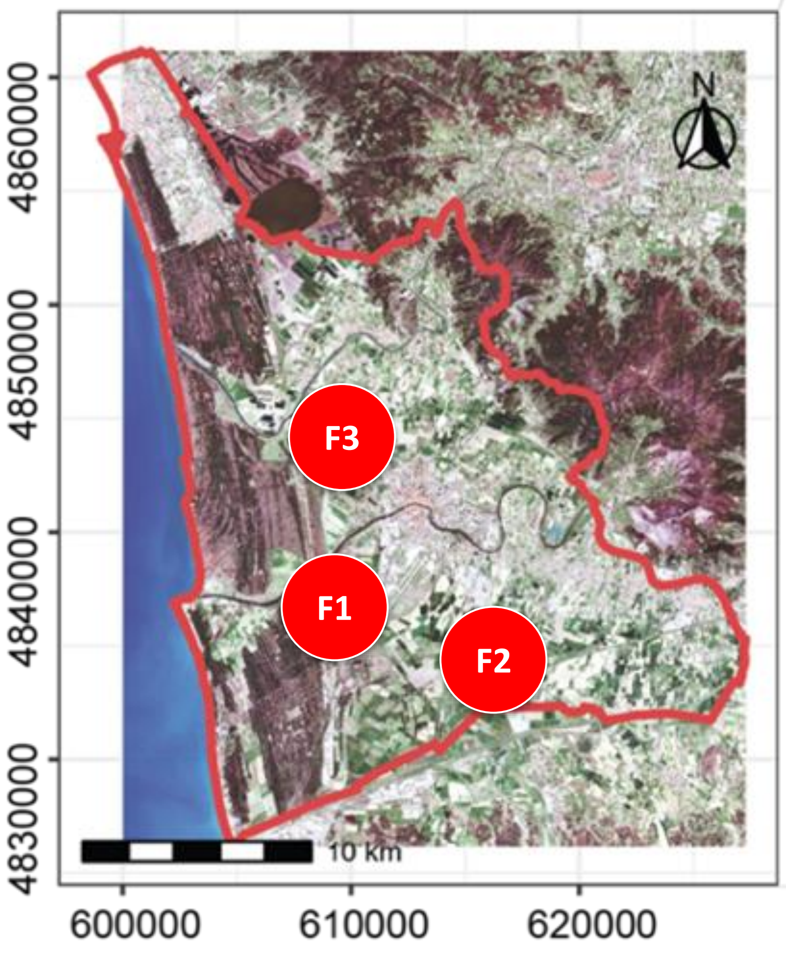
about 20 days frequency

weekly frequency, when possible



matching with less than 5 days gap

Ground-LAI sampling scheme and measurement methods



Growing seasons	Farm	Crop	ESU	Total ESU
2018	F1	Winter wheat	61	492
		Maize	48	
		Alfalfa	84	
	F2	Winter wheat	48	
		Maize	36	
		Alfalfa	60	
	F3	Winter wheat	36	
		Maize	48	
		Alfalfa	72	
2019	**F3**	**Winter wheat**	**132**	**335**
		Maize	**96**	
		Alfalfa	**107**	

Vegetation Indices

$$NDVI = \frac{(\rho865 - \rho665)}{(\rho865 + \rho665)}$$

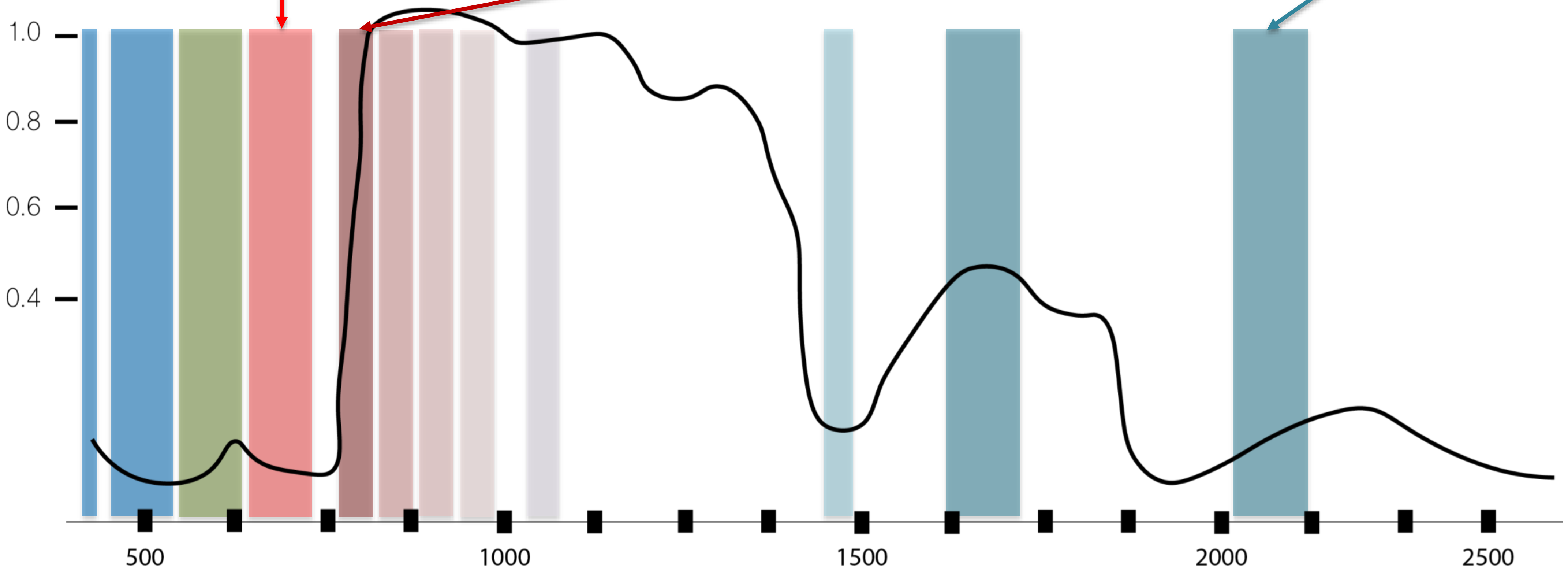
Normalized difference vegetation index

$$SeLI = \frac{(\rho865 - \rho705)}{(\rho865 + \rho705)}$$

Simple Sentinel-2 LAI Index

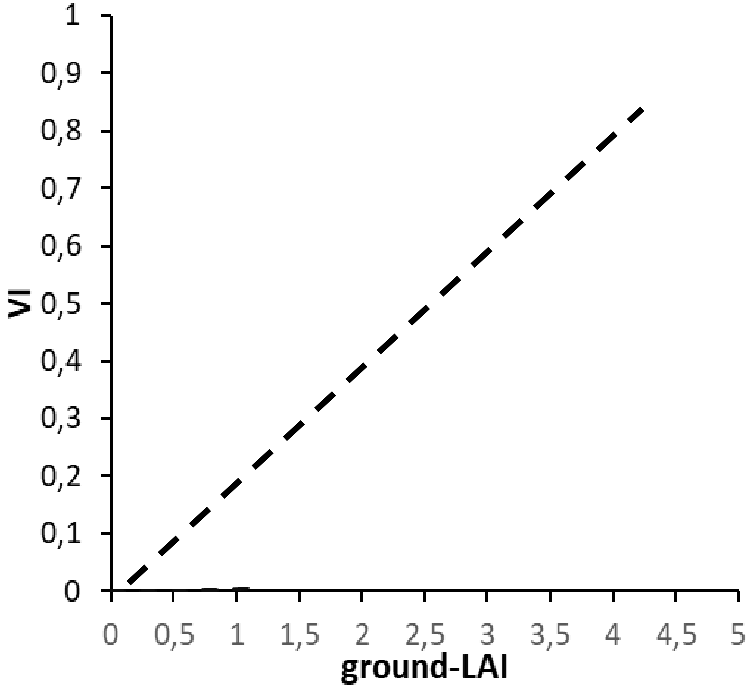
$$NBR = \frac{(\rho865 - \rho2190)}{(\rho865 + \rho2190)}$$

Normalized Burn Ratio



LAI prediction from VIs by linear parametric regression LM

$$VI = a + b \cdot groundLAI$$



Parametrization

$$LAI = \frac{VI - a}{b}$$

Prediction

LAI prediction from VIs by non-linear parametric regression $LogIF_d$

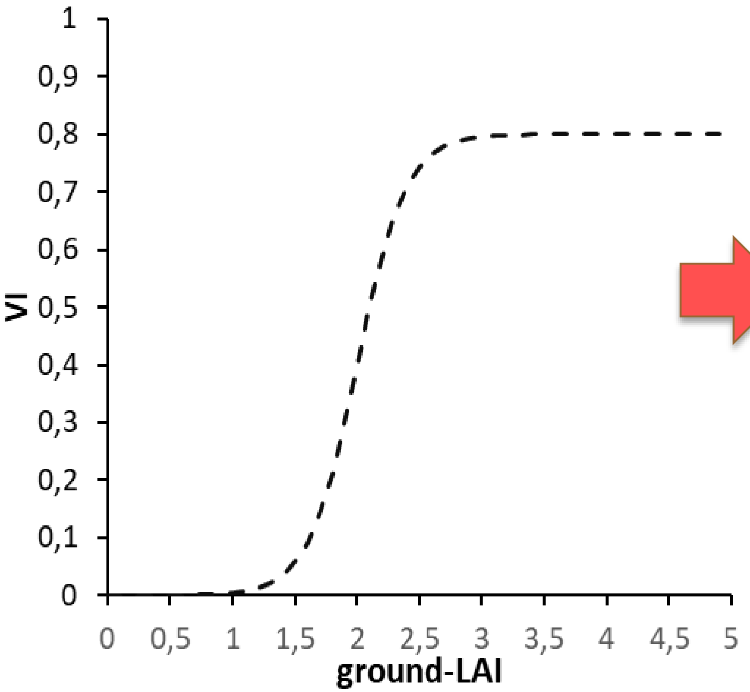
$$VI = \frac{d}{\{1 + exp[b * (groundLAI - e)]\}}$$

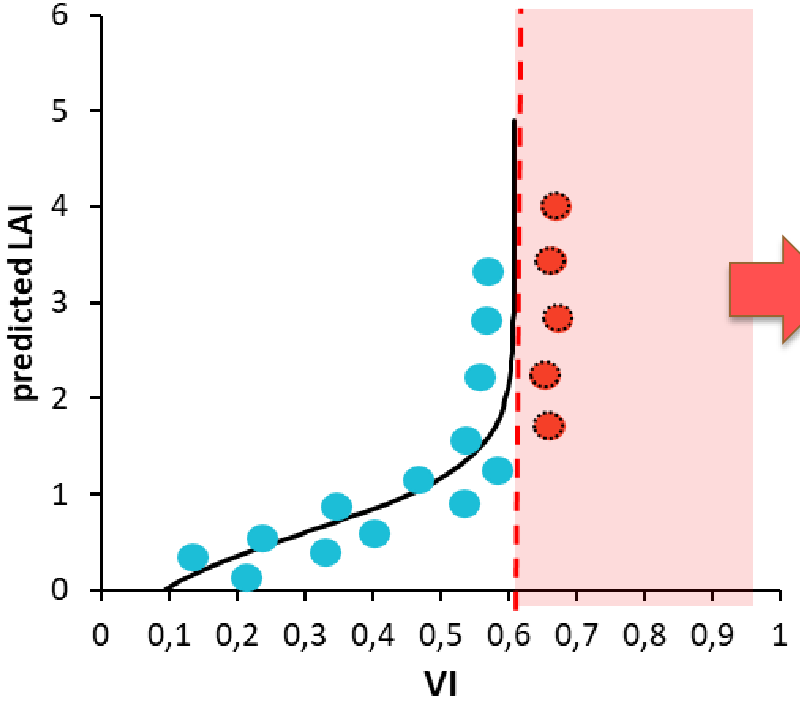
$$for\ VI \leq d$$

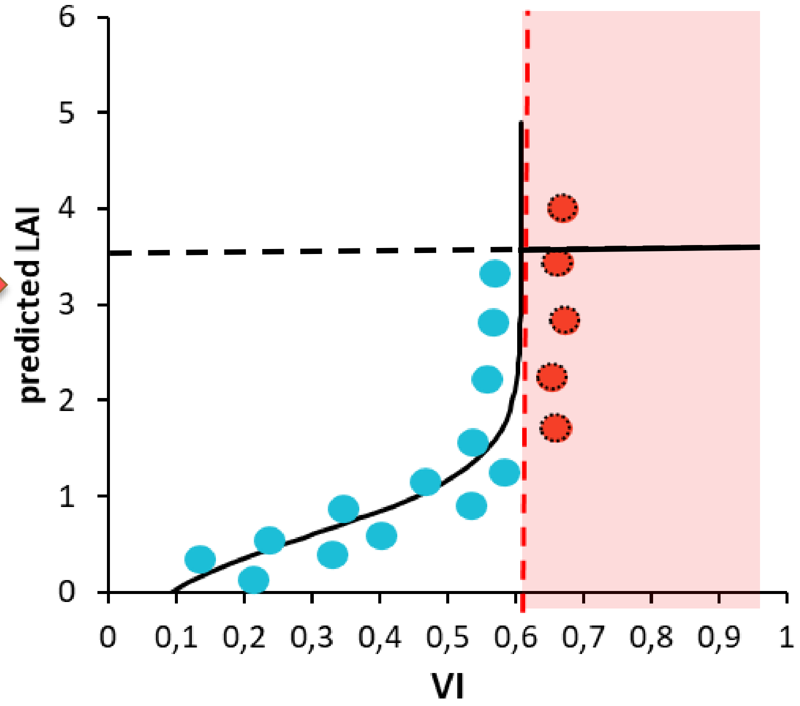
$$LAI = \frac{ln\left(\frac{d}{VI} - 1\right)}{b} + e$$

$$for\ VI > d$$

$$\max(LAI)$$







Parametrization

Prediction

LAI prediction by Machine Learning Regression Algorithms (MLRA)

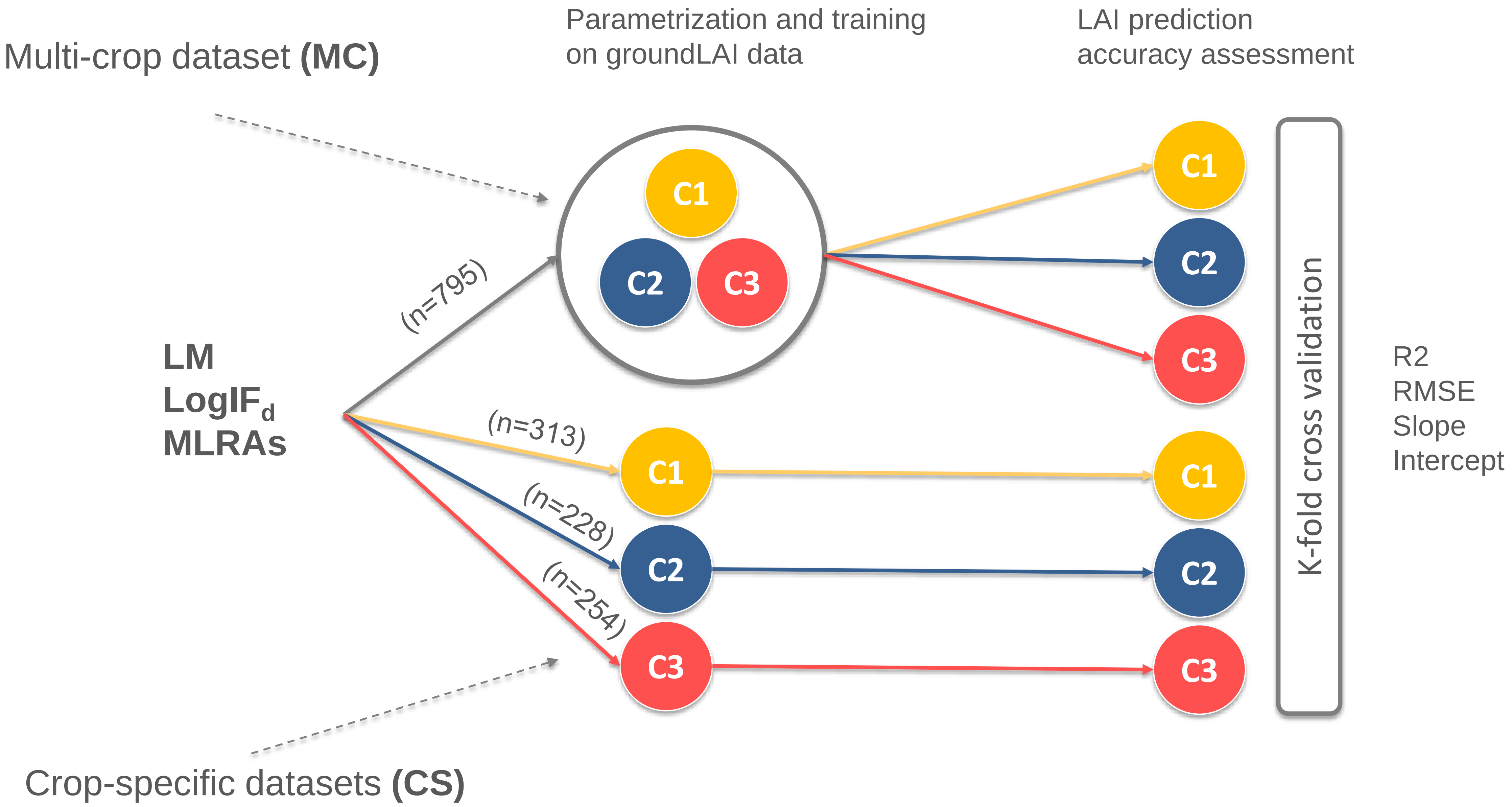
Gaussian processes regression (GPR)
provides the predictive mean, as well as predictive variance, maximizing the marginal likelihood in the training set, which is learned by hyperparameters through an appropriate kernel function
(Rasmussen et al., 2006)

Bagging trees (BAGTREE)
builds multiple decision trees by iteratively replacing resampled training data and voting for the decision trees, thus leading to a consensus prediction
(Breiman et al., 1996)

Boosting trees (BOOST)
incrementally builds an ensemble by training each new instance to emphasize the training instances which were previously mismodelled
(Friedman et al., 2000)

were run by means of the MLRA toolbox of the **Automated Radiative Transfer Models Operator (ARTMO)** software.
(Caicedo *et al.*, 2014; Verrelst et al., 2015).

Accuracy evaluation



Results - LAI prediction accuracy by VIs-based parametric approaches

			++			+			-		
Function	**Metrics**	**Data Set**	**C1**			**C2**			**C3**		
			NBR	**NDVI**	**SeLI**	**NBR**	**NDVI**	**SeLI**	**NBR**	**NDVI**	**SeLI**
LM	R_{CV}^2	CS	0.73	0.72	0.73	0.66	0.64	0.65	0.08	0.14	0.17
		MC	0.73	0.72	0.73	0.65	0.64	0.65	0.08	0.14	0.17
	$RMSE_{CV}$	CS	0.68	0.73	0.72	0.81	0.89	0.88	0.67	0.80	0.84
		MC	0.76	0.83	0.72	0.94	1.07	1.12	0.99	0.96	0.87
	Intercept	CS	0.65	0.54	0.52	0.98	0.77	0.77	4.15	3.92	3.83
		MC	1.17	0.80	0.81	1.21	0.92	1.12	2.85	2.87	2.72
	Slope	CS	0.69	0.72	0.73	0.60	0.64	0.65	0.10	0.15	0.17
		MC	0.76	0.82	0.73	0.69	0.78	0.82	0.15	0.18	0.18
	p-value	CS vs. MC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$LogIF_d$	R_{CV}^2	CS	0.72	0.70	0.62	0.60	0.59	0.55	0.05	0.04	0.09
		MC	0.73	0.72	0.65	0.64	0.62	0.65	0.02	0.03	0.05
	$RMSE_{CV}$	CS	0.98	1.09	1.33	1.00	1.14	1.31	0.84	1.27	1.46
		MC	0.75	0.86	0.85	0.93	1.16	1.23	1.02	1.32	1.41
	Intercept	CS	0.23	0.20	0.16	0.55	0.60	0.60	2.13	2.53	2.43
		MC	0.43	0.29	0.25	0.50	0.54	0.58	2.39	2.71	2.61
	Slope	CS	0.97	1.03	1.05	0.65	0.71	0.76	0.11	0.16	0.22
		MC	0.77	0.85	0.73	0.67	0.78	0.89	0.09	0.15	0.15
	p-value	CS vs. MC	0.51	0.00	0.00	0.01	0.60	0.24	0.00	0.00	0.14

=

-

+?

Results - LAI prediction accuracy by MLRA

Metrics	Dataset	C1 GPR	C1 BOOST	C1 BAGTREE	C2 GPR	C2 BOOST	C2 BAGTREE	C3 GPR	C3 BOOST	C3 BAGTREE
R_{CV}^2	CS	0.81	0.79	0.77	0.69	0.63	0.67	0.71	0.62	0.63
	MC	0.77	0.63	0.74	0.59	0.58	0.62	0.65	0.62	0.62
$RMSE_{CV}$	CS	0.65	0.68	0.66	0.84	1.04	0.87	1.00	1.22	1.04
	MC	0.66	0.99	0.77	1.06	1.15	1.02	1.05	1.07	1.03
Intercept	CS	0.36	0.29	0.45	0.67	0.56	0.67	1.36	1.26	1.65
	MC	0.51	0.43	0.55	0.72	0.77	0.75	1.42	1.61	1.52
Slope	CS	0.82	0.81	0.75	0.68	0.73	0.68	0.71	0.71	0.63
	MC	0.75	0.81	0.80	0.69	0.72	0.70	0.65	0.61	0.59
p-value	CS vs. MC	0.13	0.08	0.00	0.30	0.06	0.06	0.00	0.27	0.00

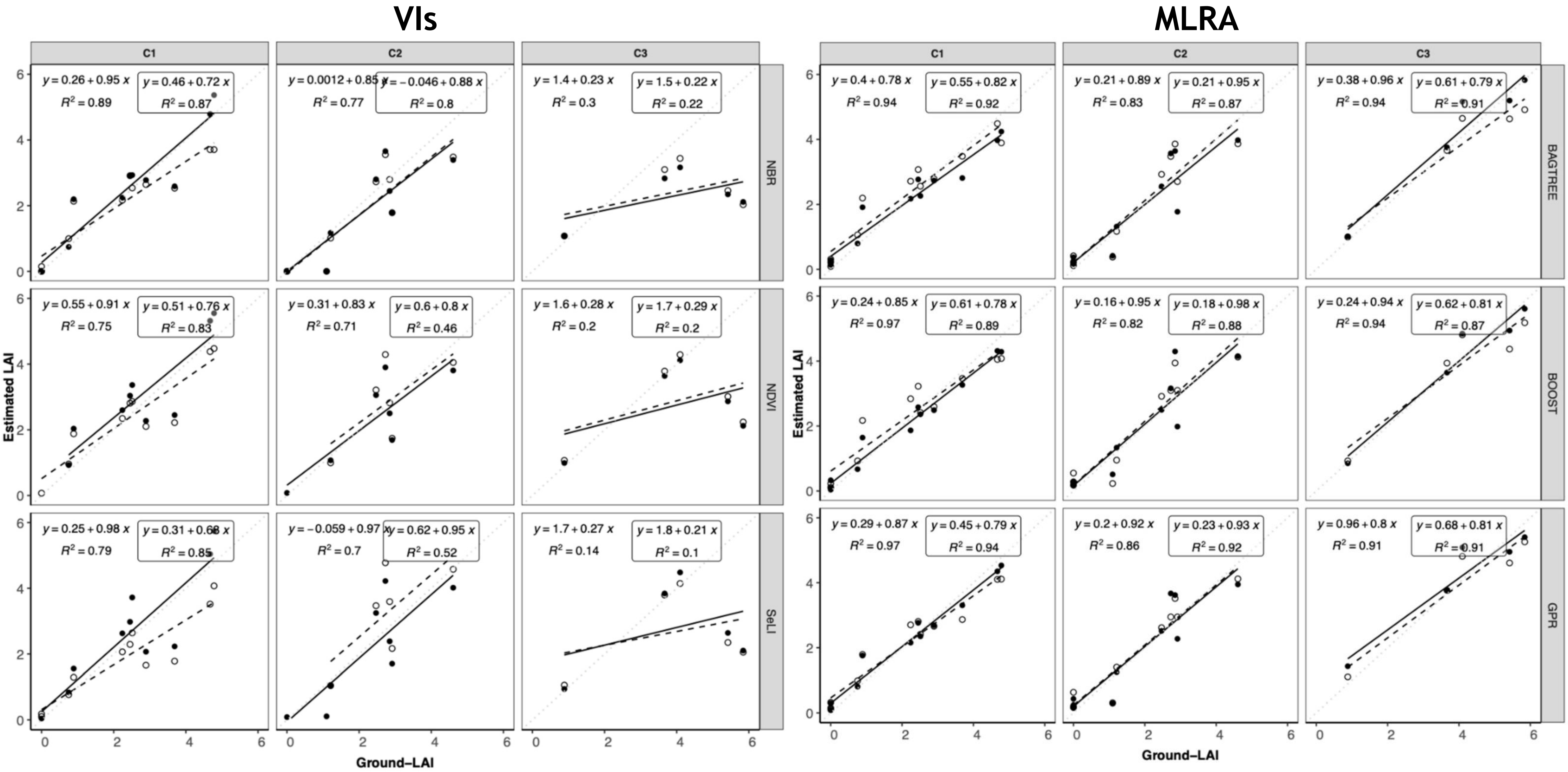
+

+

+

Results

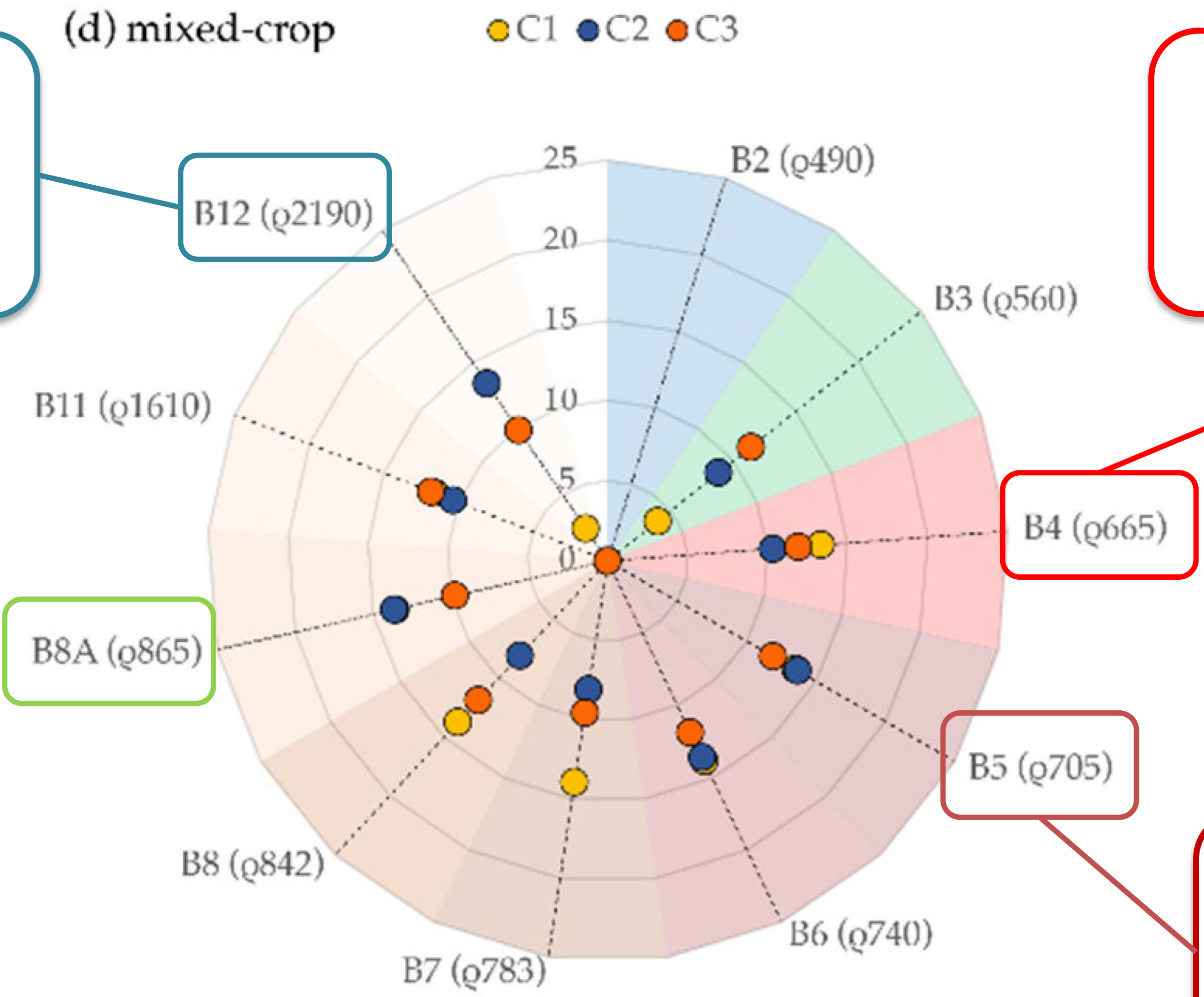
= CS
= MC



Spectral band relevance by GRP hyper-parameters

$$NBR = \frac{(\rho865-\rho2190)}{(\rho865+\rho2190)}$$

Normalized Burn Ratio



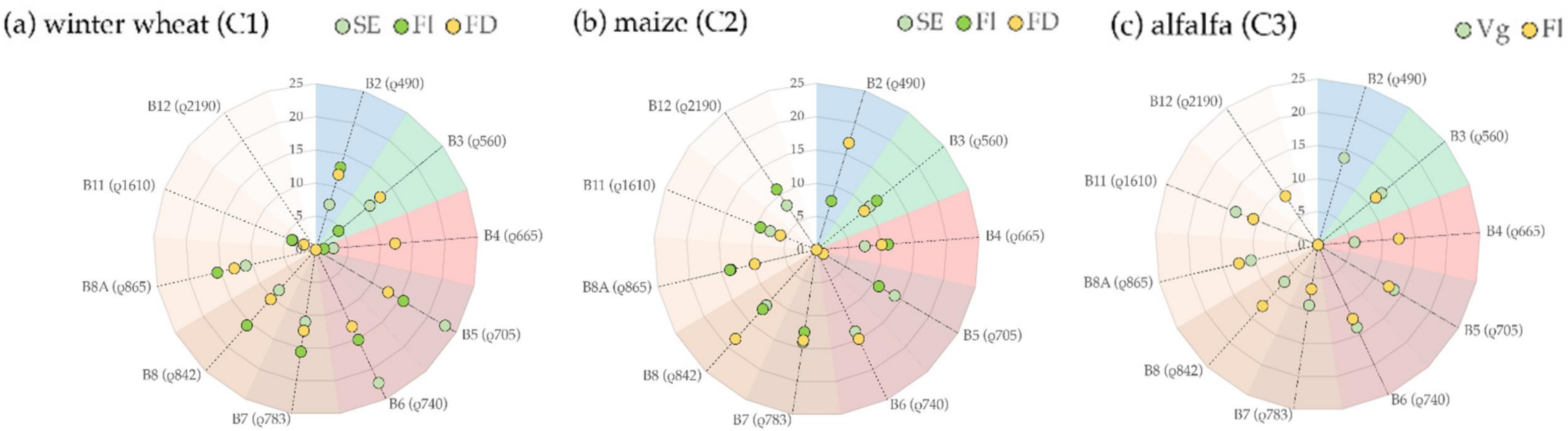
$$NDVI = \frac{(\rho865-\rho665)}{(\rho865+\rho665)}$$

Normalized difference vegetation index

$$SeLI = \frac{(\rho865-\rho705)}{(\rho865+\rho705)}$$

Simple Sentinel-2 LAI Index

Spectral band relevance by GRP hyper-parameters



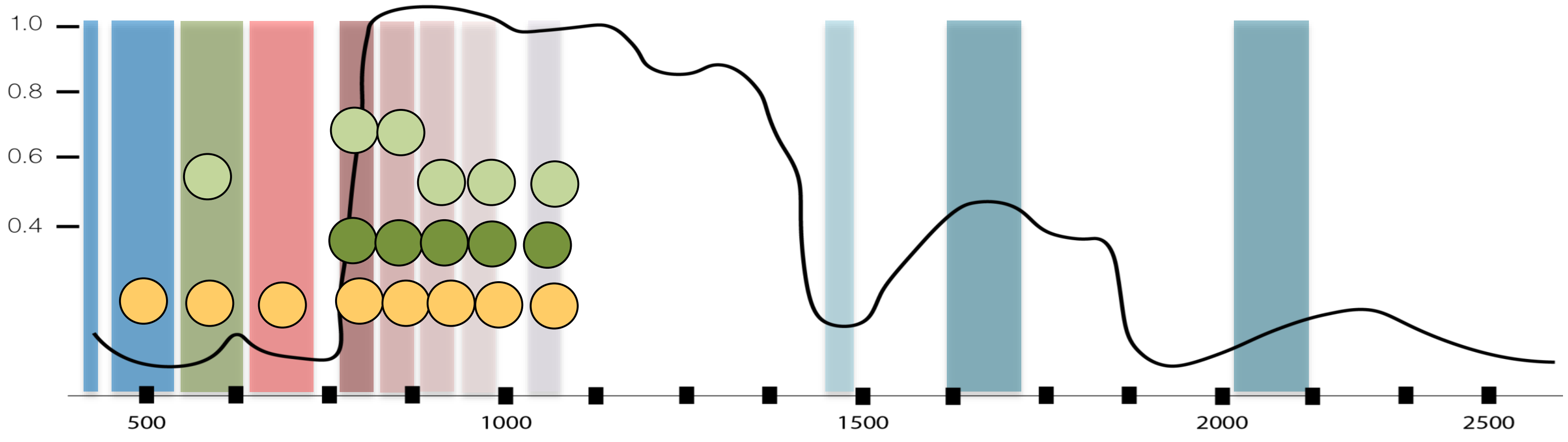
Winter wheat

SE – Tillering to stem elongation (6)

FL – Flowering (5)

FD – Fruit Development (8)

$$REIP = 700 + 400\,\frac{(\rho 670 + \rho 780) * 0.5 - \rho 700}{\rho 740 + \rho 700}$$



Conclusions

VI-based parametric methods showed:
(i) to be **unsuitable** for LAI retrieval of **alfalfa**
(ii) to have a very **low accuracy** for LAI retrieval form **multi crop landscape mosaic**;
(iii) to have a very **low accuracy** for LAI retrieval at pixel level**, limiting the assessment of within variability**;
(iv) to have an accuracy of prediction the largely **depends on VI selection, the fitting function, and the parameterization dataset**

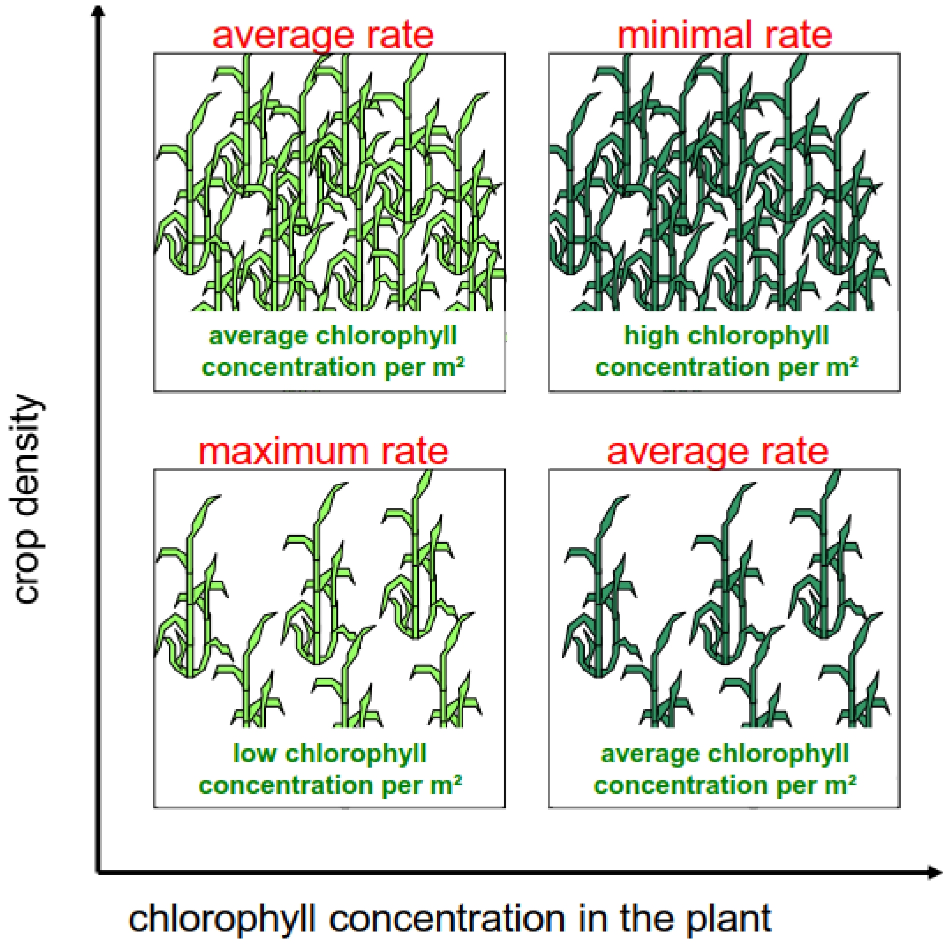
MLRA-based methods showed:
(i) the best-performing MLRA was GPR, belonging to the kernel machine learning regression algorithms
(ii) GRP showed high prediction accuracy **at pixel level (within field variability)**
(iii) GRP showed high prediction accuracy on a **multi crop landscape mosaic** regardless of the crop type, growth stage, and the training dataset
(iv) GPR provides **useful information about spectral bands relevance in different phenological stages**, improving LAI prediction. This tool can help in the monitoring of crop phenological development

Future perspectives

Future perspectives rise from :

i) The need to produce support tools for agriculture able to reduce the effort in field of farmers for collecting data for calibration

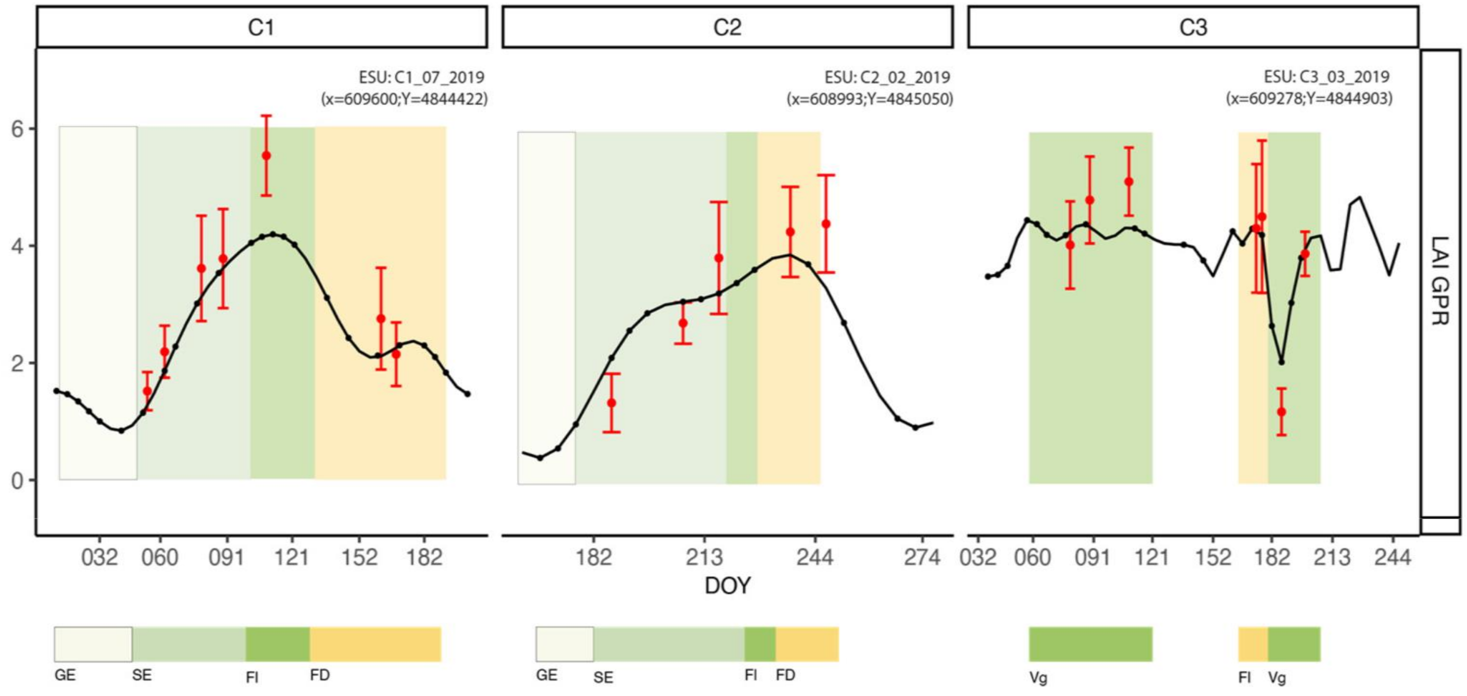
ii) The need of discriminating combined and correlated effects



Future perspectives

Future perspectives rise from :

i) The need to produce support tools for agriculture, allowing to reduce the effort in field of farmers for collecting data for calibration
ii) The need of discriminating combined and correlated effects
iii) The opportunity to improve the understanding of crop phenology and development in space as well as time



Future perspectives

Future perspectives rise from :

i) The need to produce support tools for agriculture able to reduce the effort in field of farmers for collecting data for calibration
ii) The need of discriminating combined and correlated effects
iii) The opportunity to improve the understanding of crop phenology and development in space and time
iv) The opportunity to combine ML approach with biophysical models or other informative layers

Thanks to all authors, Margherita in particular
to Jochem and his team for making ARTMO available