

High Dimensional Kullback-Leibler Divergence for grassland object-oriented classification from high resolution satellite image time series

Maily Lopes¹, Mathieu Fauvel¹, Stéphane Girard², David Sheeren¹ and Marc Lang¹



¹ Dynafor, INRA, INPT, Université de Toulouse, France

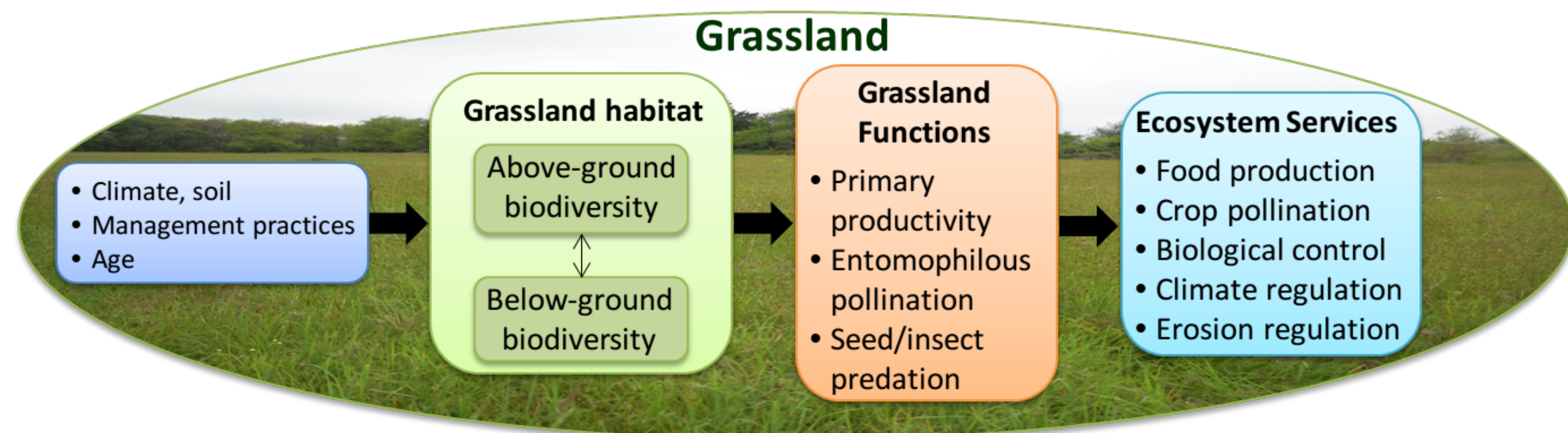
² MISTIS, INRIA Grenoble Rhône-Alpes & LJK, France

Living Planet Symposium, Prague, Czech Republic, 9-13 May 2016



Context and objectives

Grasslands: a key semi-natural element in the landscapes



Grassland state monitoring

Field Survey	Satellite image time series (SITS)
+ Easy to operate, precise	Large scale coverage, revisit frequency
- Time consuming, expansive, requires skills, limited in time and space, site specific	High dimensionality of data with lack of reference data, cloud cover

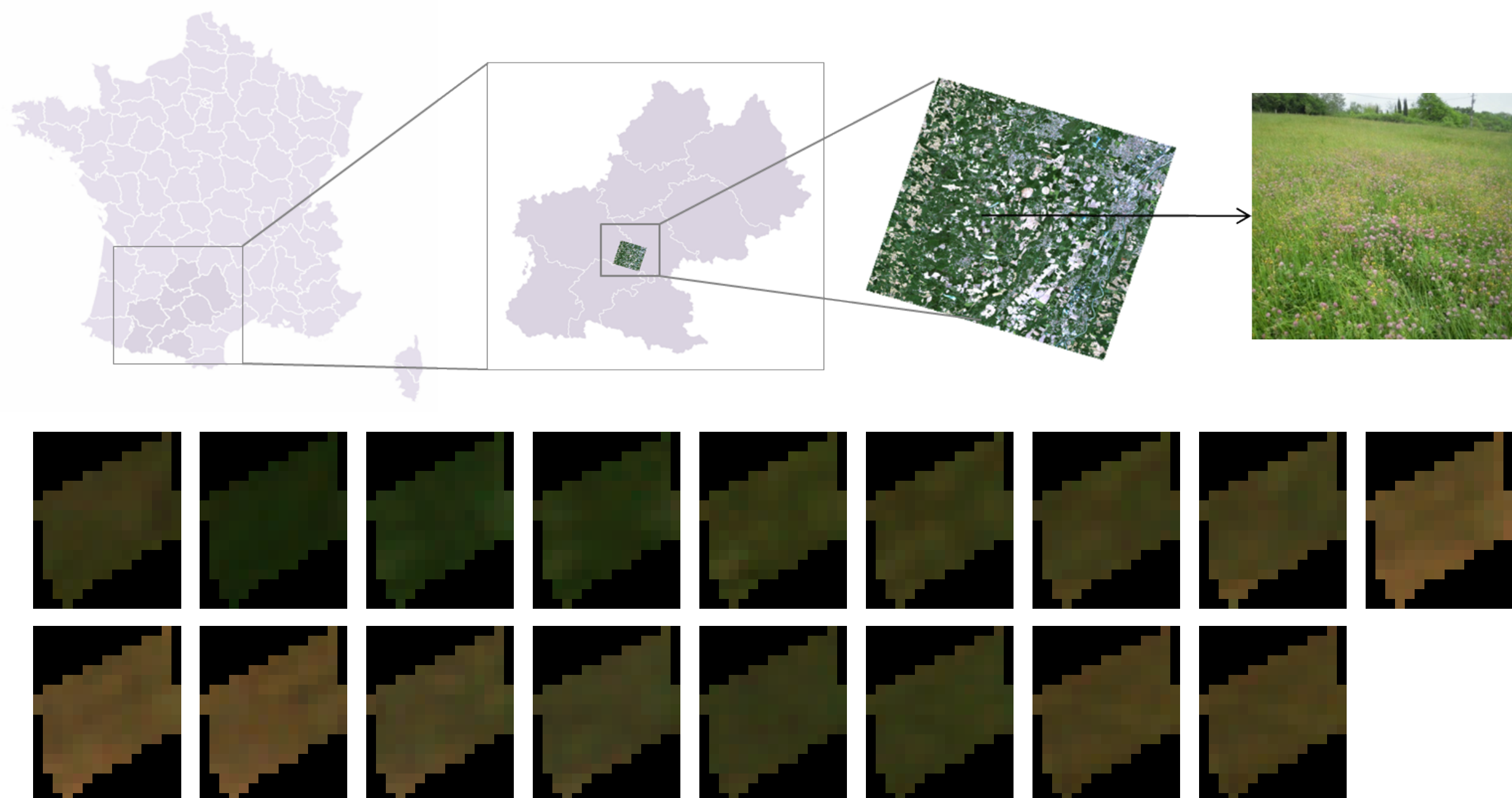
Objectives of this study

Identify grassland management practices at the **parcel scale** accounting for their heterogeneity and using SITS of NDVI with **very high temporal** and **high spatial resolution**. Develop a statistical model for **Sentinel-2 SITS**.

Remote sensing of grassland management practices constraints

- Practices defined at the **parcel scale** \Rightarrow an **object-oriented** method is required.
- Grasslands are **heterogeneous** objects \Rightarrow **spectral variability**.
- Grasslands are **small** (≈ 1 ha) \Rightarrow low number of pixels face to a **high number of spectro-temporal variables**.

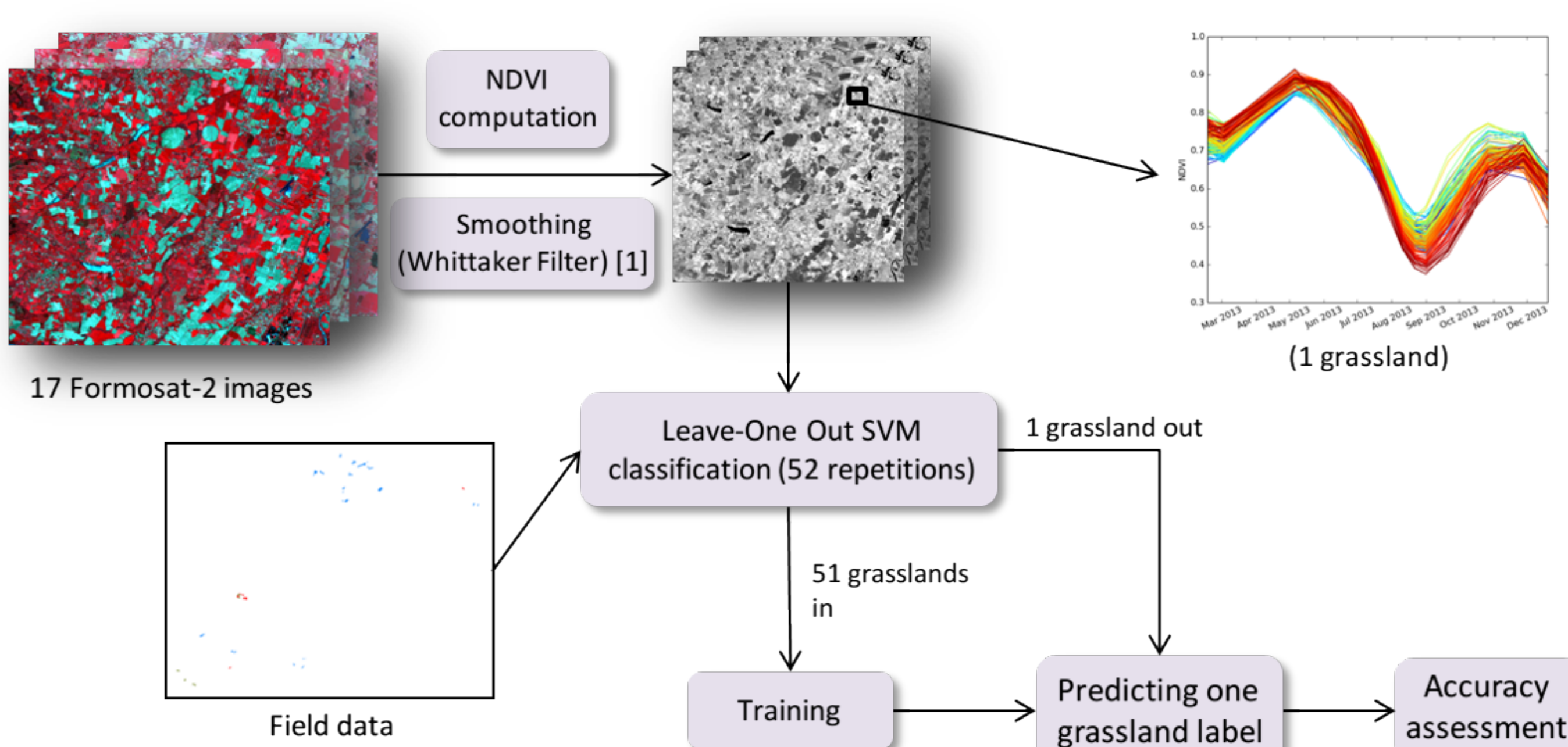
Study site and data



Formosat-2 images (8 meters resolution, 4 spectral bands) from 2013.

Class	Nb of grasslands
Mowing	34
Grazing	10
Mixed (mowing & grazing)	8

Processing chain



Methodological framework

How to work at the parcel scale with a group of heterogeneous pixels?

Each grassland $g_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ where μ_i and Σ_i are the mean vector and covariance matrix of pixels from g_i .

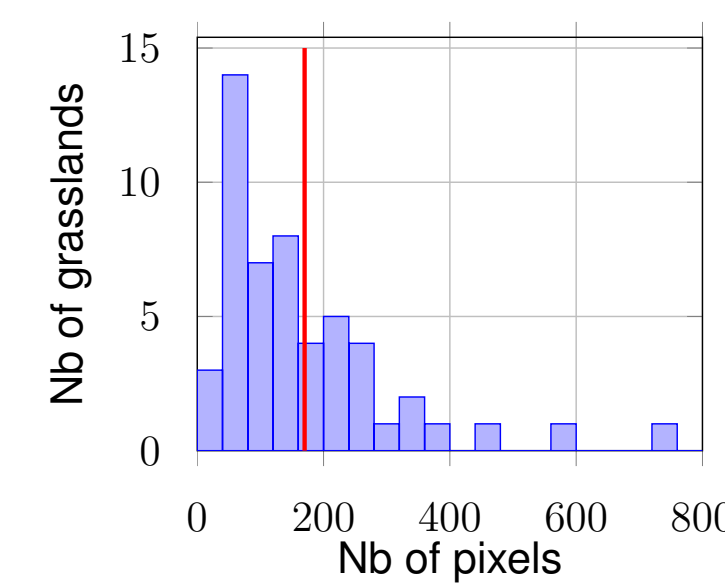
For the classification, a **measure of similarity** between two grasslands g_i and g_j is required.

We propose to use the symmetrized **Kullback-Leibler divergence** (KLD) [1], a semi-metric between two Gaussian distributions.

Symmetrized Kullback-Leibler Divergence (KLD)

$$KLD(g_i, g_j) = \frac{1}{2} \left[\text{Tr} [\Sigma_i^{-1} \Sigma_j + \Sigma_j^{-1} \Sigma_i] + (\mu_i - \mu_j)^\top (\Sigma_i^{-1} + \Sigma_j^{-1}) (\mu_i - \mu_j) \right] - d$$

where d is the number of variables, Tr is the trace operator, $\mu_{(i,j)}$ and $\Sigma_{(i,j)}$ are estimated by their empirical counterparts.



The number of pixels inside g_i is lower than the number of variables to estimate (see histogram). Therefore, a **High Dimensional Kullback-Leibler divergence** is proposed using the HDDA model from [2].

High Dimensional Symmetrized KLD (HDKLD)

$$HDKLD(g_i, g_j) = \frac{1}{2} \left[-\|\Lambda_j^{\frac{1}{2}} Q_j^\top Q_i V_i^{\frac{1}{2}}\|_F^2 - \|\Lambda_i^{\frac{1}{2}} Q_i^\top Q_j V_j^{\frac{1}{2}}\|_F^2 + \lambda_i^{-1} \text{Tr} [\Lambda_j] - \lambda_j \text{Tr} [V_i] + \lambda_j^{-1} \text{Tr} [\Lambda_i] - \lambda_i \text{Tr} [V_j] - \|\mathbf{V}_i^{\frac{1}{2}} Q_i^\top (\mu_i - \mu_j)\|^2 - \|\mathbf{V}_j^{\frac{1}{2}} Q_j^\top (\mu_i - \mu_j)\|^2 + \frac{\lambda_i + \lambda_j}{\lambda_i \lambda_j} \|(\mu_i - \mu_j)\|^2 + \frac{\lambda_i^2 + \lambda_j^2}{\lambda_i \lambda_j} d \right] - d$$

where $Q_i = [q_{i1}, \dots, q_{ip_i}]$, $\Lambda_i = \text{diag}[\lambda_{i1} - \lambda_i, \dots, \lambda_{ip_i} - \lambda_i]$, $V_i = \text{diag}[\frac{1}{\lambda_i} - \frac{1}{\lambda_{i1}}, \dots, \frac{1}{\lambda_i} - \frac{1}{\lambda_{ip_i}}]$, q_{ij}, λ_{ij} are the j^{th} eigenvalues/eigenvectors of Σ_i , $j \in \{1, \dots, d\}$ such as $\lambda_{i1} \geq \dots \geq \lambda_{ip_i}$, p_i is the number of non-equal eigenvalues, λ_i is the multiple eigenvalue corresponding to the noise term and $\|L\|_F^2 = \text{Tr}(L^\top L)$ is the Frobenius norm.

Experimental results

Classification methods

Name	p-SVM	μ -SVM	KLD-SVM	HDKLD-SVM
Scale	Pixel	Object	Object	Object
Feature	p_{il}	μ_i	$\mathcal{N}(\mu_i, \Sigma_i)$	$\mathcal{N}(\mu_i, \Sigma_i)$
Kernel	RBF	RBF	$K(g_i, g_j)$	$K(g_i, g_j)$

where $K(g_i, g_j) = \exp \left[-\frac{(HD)KLD(g_i, g_j)^2}{\sigma} \right]$ with $\sigma \in \mathbb{R}_{>0}$

Results with LOOCV

	p-SVM	μ -SVM	KLD-SVM	HDKLD-SVM
	REF	REF	REF	REF
PRED	32 4 2	31 6 3	32 8 8	33 4 4
	1 4 1	1 0 0	1 0 0	0 3 0
	1 0 7	2 2 7	1 0 2	1 1 6
OA	0.83	0.73	0.66	0.81
Kappa	0.64	0.41	0.09	0.57

HDKLD-SVM is **significantly better** than the conventional KLD. HDKLD-SVM and p-SVM classifications are **statistically equivalent** (Kappa analysis).

Conclusions and perspectives

- HDKLD is **robust** to this configuration and **outperforms the conventional KLD**. It enables a **proper modelization** of the grassland at the **parcel scale**.
- The method will be further extended to multispectral data and assessed with a larger dataset.
- The method will be tested with Sentinel-2 data.

References

- [1] S. Kullback, "Letter to the editor: The Kullback-Leibler distance," *The American Statistician*, vol. 41, no. 4, pp. 340–341, 1987.
- [2] C. Bouveyron, S. Girard, and C. Schmid, "High-dimensional discriminant analysis," *Communications in Statistics - Theory and Methods*, vol. 36, no. 14, pp. 2607–2623, 2007.
- [3] P. H. C. Eilers, "A perfect smoother," *Analytical Chemistry*, vol. 75, no. 14, pp. 3631–3636, 2003.