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

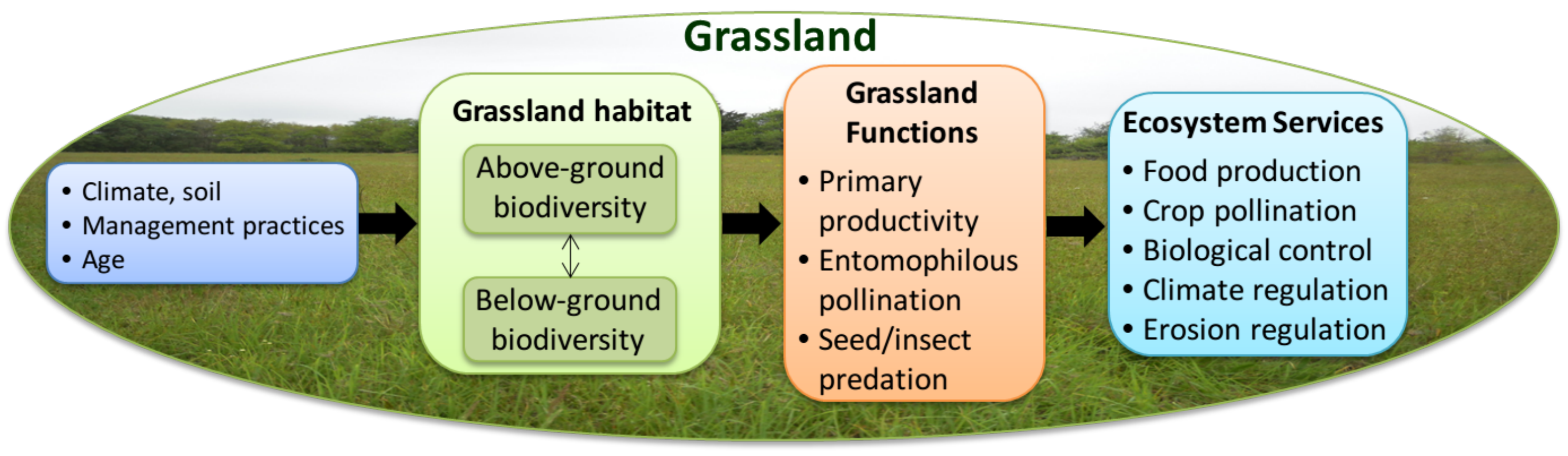
Mailys Lopes [1], Mathieu Fauvel [1], Stéphane Girard [2], David Sheeren [1] and Marc Lang [1]

[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

- Climate, soil
- Management practices
- Age

Grassland habitat

Above-ground biodiversity

Below-ground biodiversity

Grassland Functions

- Primary productivity
- Entomophilous pollination
- Seed/insect predation

Ecosystem Services

- Food production
- Crop pollination
- Biological control
- Climate regulation
- Erosion regulation

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

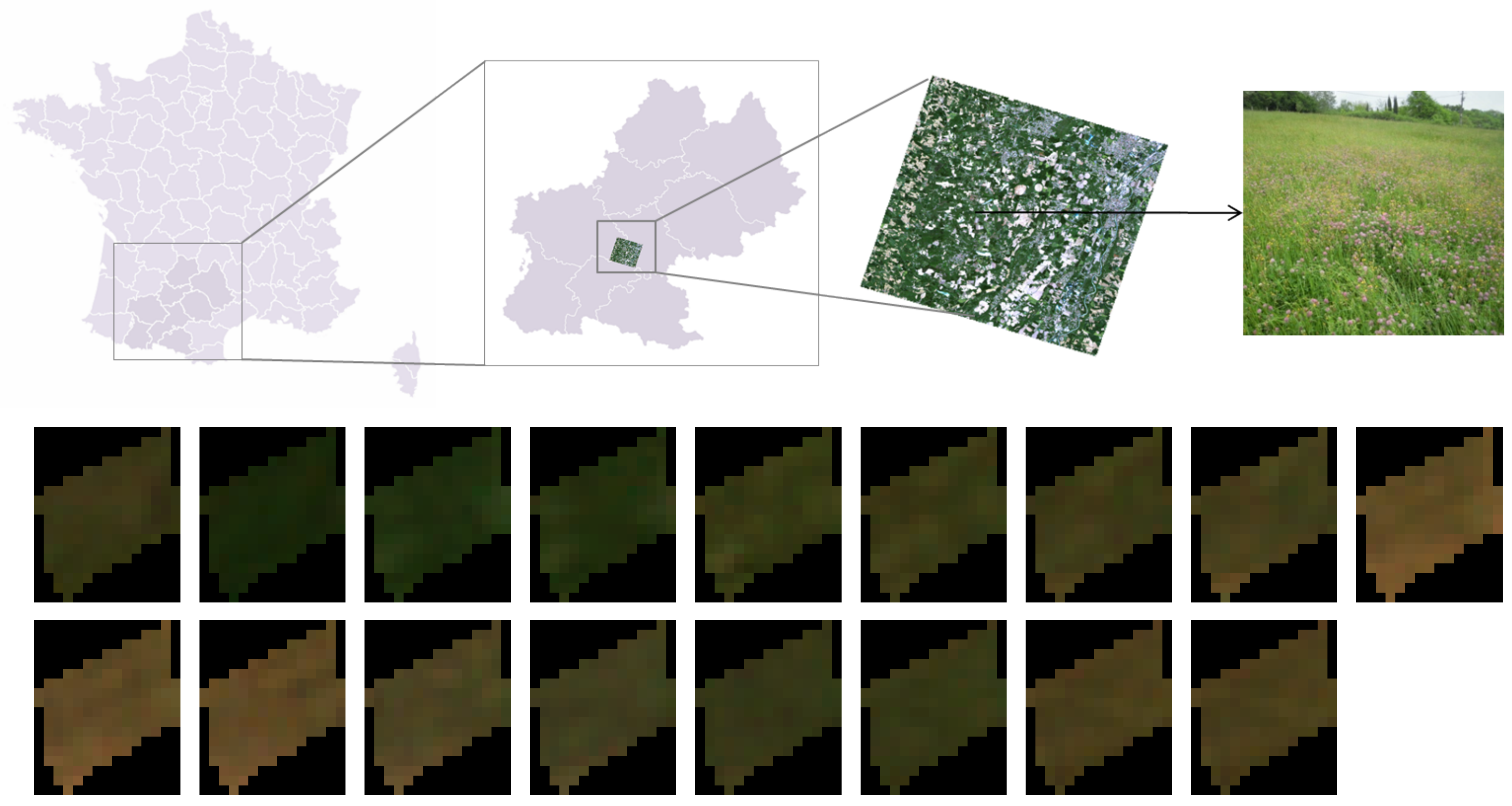
Objectives of this study

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** ⇒ an **object-oriented** method is required.
- Grasslands are **heterogeneous** objects ⇒ **spectral variability**.
- Grasslands are **small** (≈ 1 ha) ⇒ low number of pixels face to a **high number of spectro-temporal variables**.

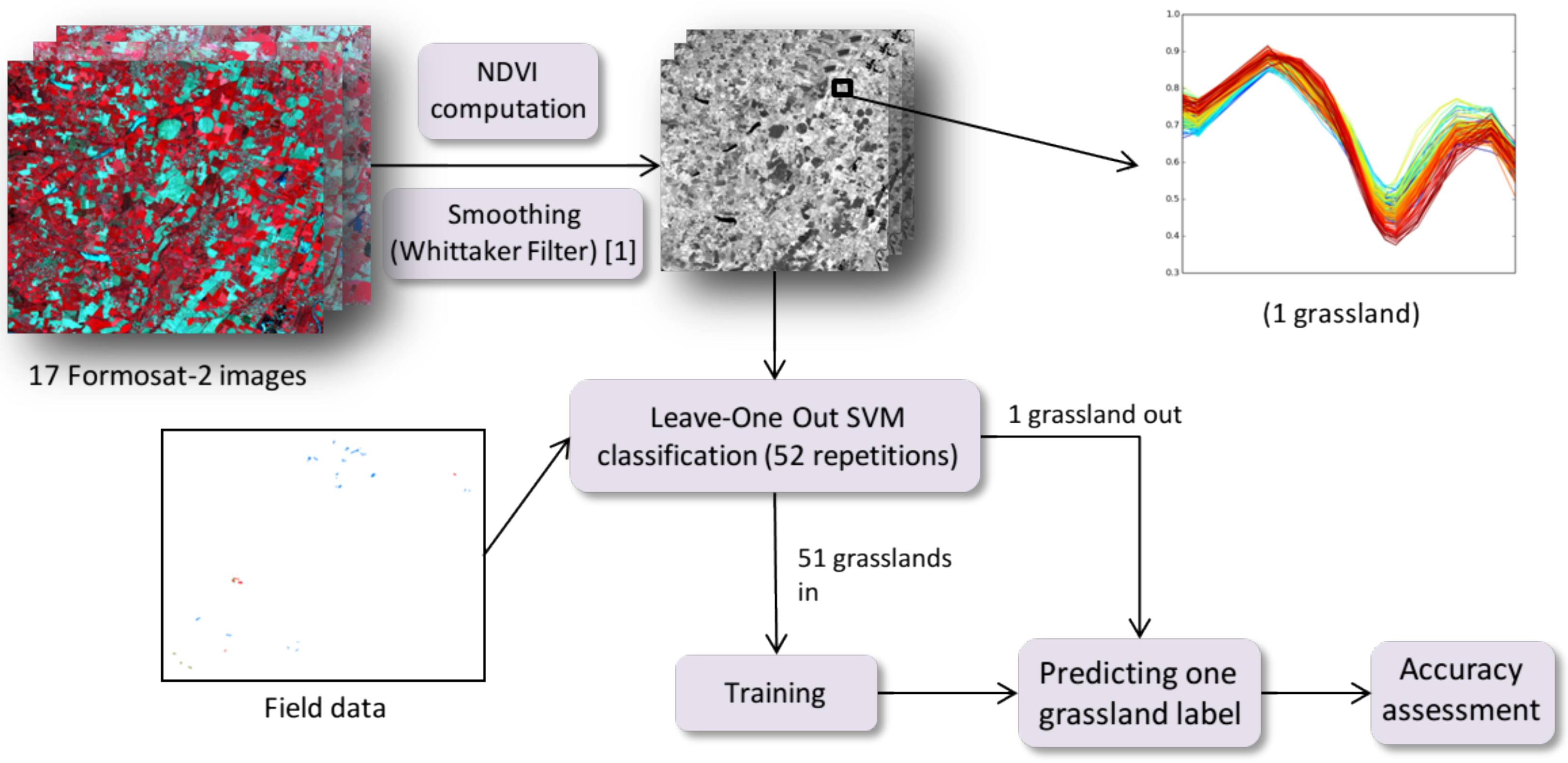
Study site and data



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



NDVI computation

Smoothing (Whittaker Filter) [1]

17 Formosat-2 images

(1 grassland)

Field data

Leave-One Out SVM classification (52 repetitions)

1 grassland out

51 grasslands in

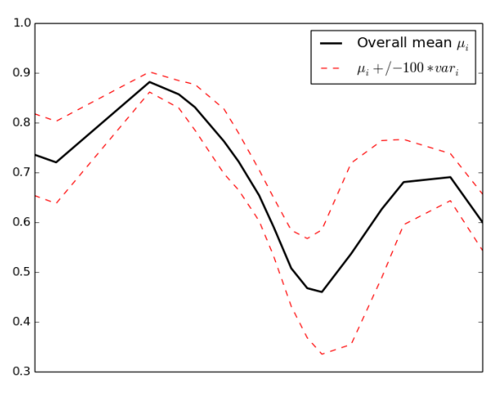
Training

Predicting one grassland label

Accuracy assessment

Methodological framework

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

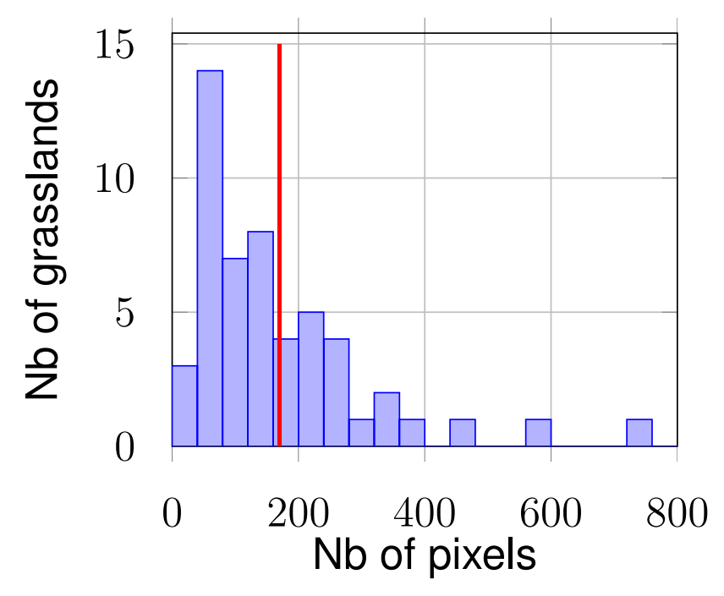


Each grassland $g_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the mean vector and covariance matrix of pixels from g_i.
For the classification, a **measure of similarity** between two grasslands g_i and g_j is required.
We propose to use the symmetrized **Kullback-Leibler divergence** (KLD) [1], a semi-metric between two Gaussian distributions.

Symmetrized Kullback-Leibler Divergence (KLD)

$$KLD(g_i, g_j) = \frac{1}{2}\Big[\operatorname{Tr}\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Sigma}_i\right] + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top \big(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1}\big)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\Big] - d$$

where d is the number of variables, Tr is the trace operator, $\boldsymbol{\mu}_{(i,j)}$ and $\boldsymbol{\Sigma}_{(i,j)}$ 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)

$$\begin{aligned} HDKLD(g_i, g_j) = \frac{1}{2}\Big[& - \|\boldsymbol{\Lambda}_j^{\frac{1}{2}}\mathbf{Q}_j^\top \mathbf{Q}_i \mathbf{V}_i^{\frac{1}{2}}\|_F^2 - \|\boldsymbol{\Lambda}_i^{\frac{1}{2}}\mathbf{Q}_i^\top \mathbf{Q}_j \mathbf{V}_j^{\frac{1}{2}}\|_F^2 \\ & + \lambda_i^{-1}\operatorname{Tr}\left[\boldsymbol{\Lambda}_j\right] - \lambda_j \operatorname{Tr}\left[\mathbf{V}_i\right] + \lambda_j^{-1}\operatorname{Tr}\left[\boldsymbol{\Lambda}_i\right] - \lambda_i \operatorname{Tr}\left[\mathbf{V}_j\right] \\ & - \|\mathbf{V}_i^{\frac{1}{2}}\mathbf{Q}_i^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 - \|\mathbf{V}_j^{\frac{1}{2}}\mathbf{Q}_j^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 + \frac{\lambda_i + \lambda_j}{\lambda_i\lambda_j}\|(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 + \frac{\lambda_i^2 + \lambda_j^2}{\lambda_i\lambda_j}d\Big] - d \end{aligned}$$

where $\mathbf{Q}_i = \left[\mathbf{q}_{i1}, \ldots, \mathbf{q}_{ip_i}\right]$, $\boldsymbol{\Lambda}_i = \text{diag}\Big[\lambda_{i1} - \lambda_i, \ldots, \lambda_{ip_i} - \lambda_i\Big]$, $\mathbf{V}_i = \text{diag}\Big[\frac{1}{\lambda_i} - \frac{1}{\lambda_{i1}}, \ldots, \frac{1}{\lambda_i} - \frac{1}{\lambda_{ip_i}}\Big]$, $\mathbf{q}_{ij}, \lambda_{ij}$ are the j^{th} eigenvalues/eigenvectors of $\boldsymbol{\Sigma}_i$, $j \in \{1, \ldots, d\}$ such as $\lambda_{i1} \geq \ldots \geq \lambda_{id}$, p_i is the number of non-equal eigenvalues, λ_i is the multiple eigenvalue corresponding to the noise term and $\|L\|_F^2 = \operatorname{Tr}(L^\top L)$ 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	$\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$	$\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$
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**
PRED \\ REF	**32** 4 2 1 **4** 1 1 0 **7**	**31** 6 3 1 **0** 0 2 2 **7**	**32** 8 8 1 **0** 0 1 0 **2**	**33** 4 4 0 **3** 0 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.