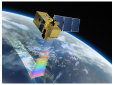
SOCIETE FRANÇAISE DE STATISTIQUE



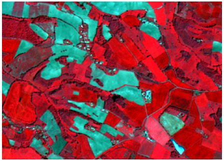
# High Dimensional Kullback-Leibler divergence for grassland classification using satellite image time series with high spatial resolution

Maïlys Lopes[1], Stéphane Girard[2], Mathieu Fauvel[1]

[1]UMR Dynafor, INRA-INP Toulouse [2]Team Mistis, INRIA Grenoble

Journées de Statistique, Montpellier, 30 mai - 3 juin 2016

INRA SCIENCE & IMPACT



Inria INVENTEURS DU MONDE NUMÉRIQUE

# Outline

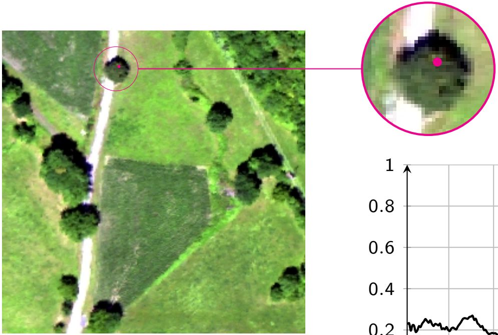
# Objectives of the study

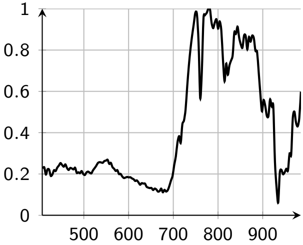
1. Characterization of grasslands through **satellite image time series** (SITS)
2. Classification of grassland **management practices**


3. Development of a **statistical modelling** of the grasslands signal suitable to SITS

## Remote sensing imagery

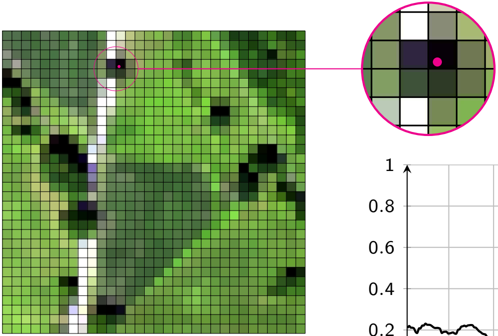
A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.

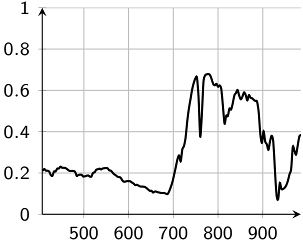




## Remote sensing imagery

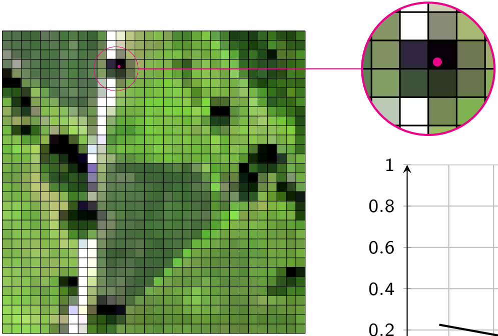
A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.

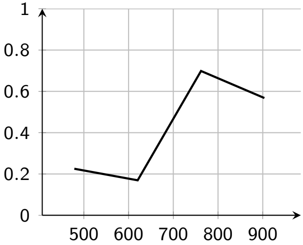




# Remote sensing imagery

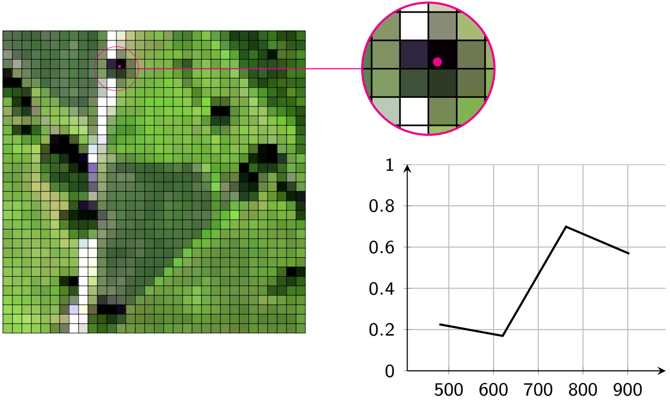
A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.

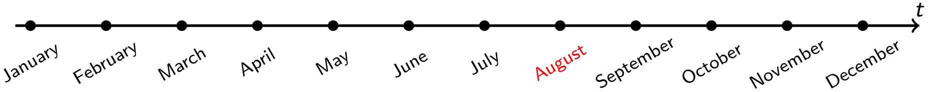




# Remote sensing imagery

A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.



## Satellite remote sensing of grasslands

Grasslands are:

- Relatively **small** ($\approx$ 100m $\times$ 100m) $\Rightarrow$ need **high spatial resolution**
- **Heterogeneous** $\Rightarrow$ need **spectral bands**
- Have a **natural vegetation cycle disturbed by anthropic events** (mowing & grazing) $\Rightarrow$ need **high temporal resolution**

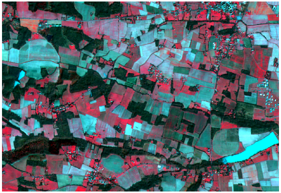
Mowing

Grazing

$\Rightarrow$ Need dense **time series with high spatial resolution** to detect and to characterize grasslands.

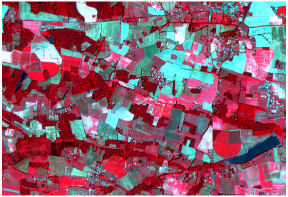
# Satellite image time series

**Formosat-2**



February

May



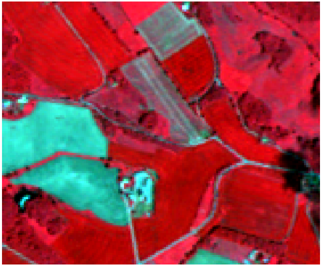
August

December

# Normalized Difference Vegetation Index (NDVI)

**NDVI**: vegetation index that reflects the photosynthetic activity of the vegetation.

$$NDVI = \frac{NIR - Red}{NIR + Red}, \; -1 \leq NDVI \leq 1$$

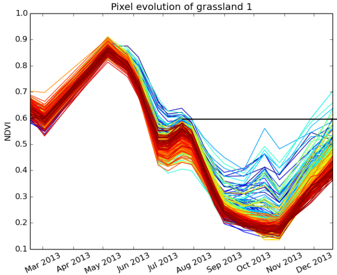


False Color (G, R, NIR)

⇒

NDVI

# Representation of the grassland




$$\mathbf{x}_{ik}(t) = \begin{pmatrix} x_{ik}(t_1) \\ \vdots \\ x_{ik}(t_d) \end{pmatrix} \in \mathbb{R}^d$$

with

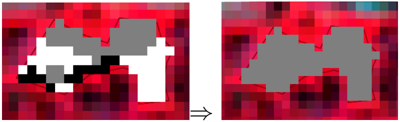
- $i$ grassland index,
- $n_i$ number of pixels in grassland $g_i$
- $k$ pixel index, $k \in \{1, ..., n_i\}$
- $d$ length of time series

A grassland is represented by a matrix of size $(n_i \times d)$, $n_i$ varying with $g_i$.

# Statistical issues

A **statistical model** is required for:

- Processing grassland **supervised** classification
- Working at the parcel scale: **Object-oriented** classification
- **Modelling grasslands** with their constraints:
  - heterogeneity (spectral variability)
  - different size $n_i$
  - described by $d$ temporal variables



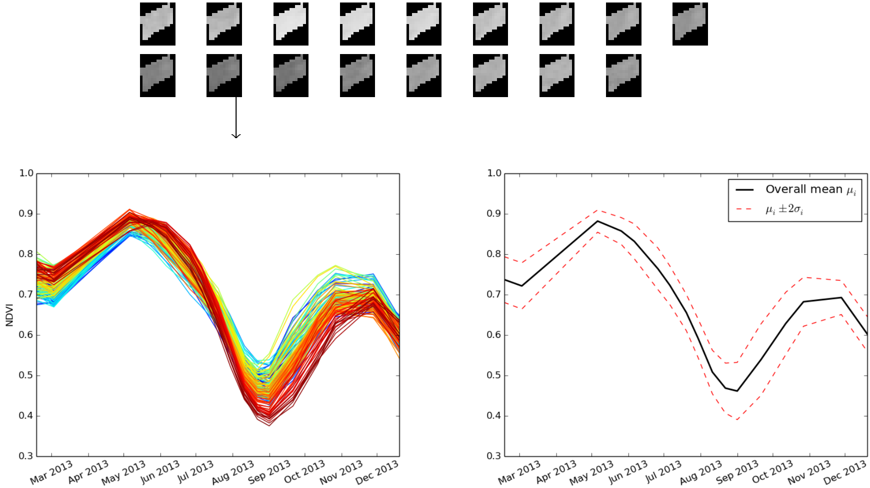
Introduction

## High Dimensional Kullback-Leibler Divergence

Experimental results

Conclusion

# Statistical modelling of grasslands




In each grassland $g_i$: each pixel $\mathbf{x}_{ik} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$

# Measuring proximity between two grasslands

Symmetrized Kullback-Leibler divergence:

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

- $\boldsymbol{\Sigma}_i$ is the covariance matrix,
- $\boldsymbol{\mu}_i$ is the mean vector of the signal,
- $d$ is the number of variables,
- Tr is the trace operator.

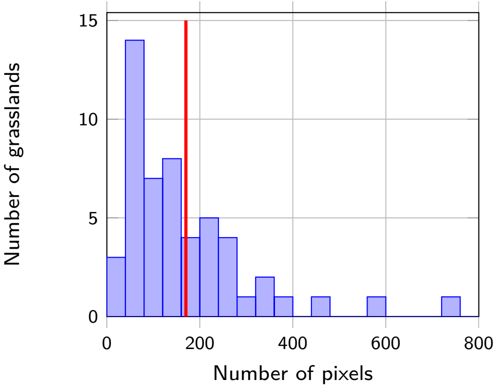



Figure : Histogram of grassland size in number of pixels $n_i$. The red line corresponds to the number of parameters to estimate for each grassland for a multivariate Gaussian model. It is derived from the number of variables using the formula $d(d+3)/2 = 170$ for $d = 17$.

# High dimensional model

According to High Dimensional Discriminant Analysis[1] that assumes that the last eigenvalues of the covariance matrix are equal :

$$\mathbf{\Lambda}_i = \begin{pmatrix} \begin{array}{|ccc|} \hline \lambda_{i1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{ip_i} \\ \hline \end{array} & \mathbf{0} \\ \mathbf{0} & \begin{array}{|cccc|} \hline \lambda_i & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \lambda_i \\ \hline \end{array} \end{pmatrix} \begin{array}{l} \left.\vphantom{\begin{array}{c}a\\b\\c\end{array}}\right\} \quad p_i \\ \left.\vphantom{\begin{array}{c}a\\b\\c\\d\end{array}}\right\} \quad (d - p_i) \end{array}$$

where $\lambda_{ij} \geq \lambda_i$, for $j = 1, ..., p_i$

[1] 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

## High dimensional model

According to High Dimensional Discriminant Analysis[1] that assumes that the last eigenvalues of the covariance matrix are equal :

$$\boxed{\mathbf{\Sigma}_i = \mathbf{Q}_i \mathbf{\Lambda}_i \mathbf{Q}_i^\top + \lambda_i \mathbf{I}_d}$$

- $\mathbf{I}_d$ is the identity matrix of size $d$,
- $p_i$ is the number of non-equal eigenvalues,
- $\lambda_i$ is the multiple eigenvalue corresponding to the noise term (last and equal eigenvalues),
- $\mathbf{Q}_i = \left[\mathbf{q}_{i1}, \ldots, \mathbf{q}_{ip_i}\right]$,
- $\mathbf{\Lambda}_i = \text{diag}\left[\lambda_{i1} - \lambda_i, \ldots, \lambda_{ip_i} - \lambda_i\right]$, $\mathbf{q}_{ij}, \lambda_{ij}$ are the $j^{th}$ eigenvalues/eigenvectors of the covariance matrix $\mathbf{\Sigma}_i$, $j \in \{1, \ldots, d\}$ such as $\lambda_{i1} \geq \ldots \geq \lambda_{id}$.

[1] 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

## High Dimensional Symmetrized KLD

Folllowing this model, $\mathbf{\Sigma}_i^{-1}$ can be computed explicitly:

$$\boxed{\mathbf{\Sigma}_i^{-1} = -\mathbf{Q}_i\mathbf{V}_i\mathbf{Q}_i^\top + \lambda_i^{-1}\mathbf{I}_d}$$

with $\mathbf{V}_i = \text{diag}\left[\frac{1}{\lambda_i} - \frac{1}{\lambda_{i1}}, \dots, \frac{1}{\lambda_i} - \frac{1}{\lambda_{ip_i}}\right]$
Then:

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

where $\|L\|_F^2 = \text{Tr}(L^\top L)$ is the Frobenius norm.

# Estimation

- $\hat{\lambda}_{ij}$ and $\hat{\boldsymbol{q}}_{ij}$ are the first eigenvalues/eigenvectors of $\hat{\boldsymbol{\Sigma}}_i$, $j \in \{1, \ldots, p_i\}$,
- $\hat{p}_i$ corresponds to the number of eigenvalues needed to reach a given percentage of variance $t$, $\frac{\sum_{j=1}^{\hat{p}_i} \hat{\lambda}_{ij}}{\text{Tr}(\hat{\boldsymbol{\Sigma}}_i)} \geq t$, $t$ being a user defined parameter,
- $\hat{\lambda}_i = \frac{\text{Tr}(\hat{\boldsymbol{\Sigma}}_i) - \sum_{j \leq \hat{p}_i} \hat{\lambda}_{ij}}{d - \hat{p}_i}$.

Thus, to compute HDKLD only the $p_i$ first eigenvalues/eigenvectors are required and the unstable estimation of the eigenvectors associated to small eigenvalues is avoided.

## Construction of a positive definite kernel

- (HD)KLD is a semi-metric.
- It can be turned to a positive definite kernel function.
- $K(g_i, g_j) = \exp\left[ -\frac{(HD)KLD(g_i,g_j)^2}{\sigma} \right]$ with $\sigma \in \mathbb{R}_*^+$

# Data

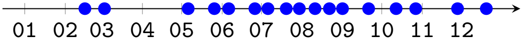
**Study site** Semi-rural area near Toulouse, France.

**Field data** 52 parcels with 3 management practices (field survey, 2015):

| Class | Nb of grasslands |
|---|---|
| Mowing | 34 |
| Grazing | 10 |
| Mixed (mowing & grazing) | 8 |
| Total number of pixels: | 8741 |

**Satellite data** **Formosat-2**

- Spatial resolution: 8m
- 4 spectral bands (B, G, R, NIR)
- Temporal frequency: up to 1 day
- $d = 17$ images from Feb. 16 to Dec. 20, 2013:



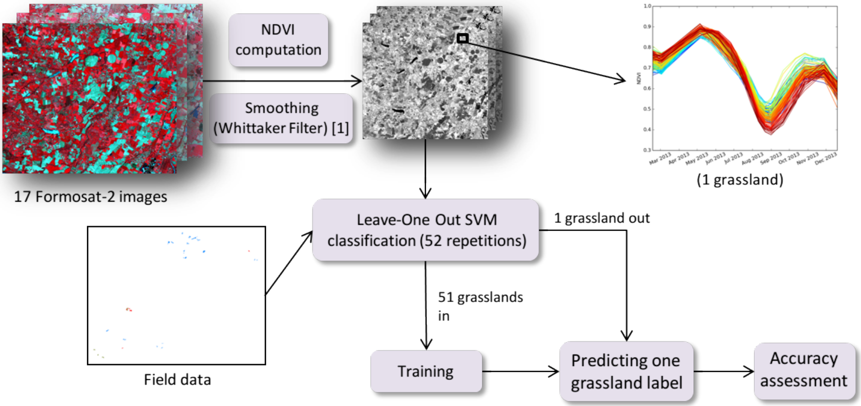
# Classification methods

| **Method** | **p-SVM** | **$\mu$-SVM** | **KLD-SVM** | **HDKLD-SVM** |
|---|---|---|---|---|
| **Scale** | Pixel | Object | Object | Object |
| **Expl. variable** | $\mathbf{x}_{ik}$ | $\boldsymbol{\mu}_i$ | $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ | $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ |
| **Kernel** | RBF | RBF | $K(g_i, g_j)$ | $K(g_i, g_j)$ |
| **Nb of samples** | 8741 | 52 | 52 | 52 |

*+ majority rule*

Optimal parameters have been optimized during cross-validation.

# Processing chain



# Results

## Classification accuracy

| | P-SVM REF | | | μ-SVM REF | | | KLD-SVM REF | | | HDKLD-SVM 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 | | |

## Test of significance of observed differences

| Z | p-SVM | $\mu$-SVM | KLD-SVM | HDKLD-SVM |
|---|---|---|---|---|
| p-SVM | / | 1.47 | **3.67** | 0.43 |
| $\mu$-SVM | 1.47 | / | **2.04** | 1.01 |
| KLD-SVM | **3.67** | **2.04** | / | **3.11** |
| HDKLD-SVM | 0.43 | 1.01 | **3.11** | / |

$$Z = \frac{|\hat{K}_m - \hat{K}_n|}{\sqrt{\hat{\text{var}}(\hat{K}_m) + \hat{\text{var}}(\hat{K}_n)}}.$$

Introduction

High Dimensional Kullback-Leibler Divergence

Experimental results

Conclusion

# Conclusion and perspectives

## Conclusion

- **Gaussian modelling** seems to be efficient to model grassland pixels ditribution.
- The proposed **HDKLD is efficient** for grassland classification.
- Results are not **significantly different**.

## Perspectives

- The method will be tested on a **larger dataset**.
- The method will be further extended to **multispectral data**.
- The method will be used for **unsupervised classification**.

# Thank you for your attention.

# Questions?

Optimal parameters have been optimized during cross-validation given this search grid:

| Parameter | p-SVM | $\mu$-SVM | KLD-SVM | HDKLD-SVM |
|---|---|---|---|---|
| $\sigma$ | $\{2^{-5}, 2^{-4}, \ldots, 2^{5}\}$ | | $\{2^{8}, 2^{9}, \ldots, 2^{12}\}$ | |
| C | $\{1, 10, 100\}$ | | | |
| t | | | $\{0.80, 0.85, 0.90, 0.95, 0.99\}$ | |