Potential of Sentinel-2 and SPOT5 (Take5) time series for the estimation of grasslands biodiversity indices

Mailys Lopes [1], Mathieu Fauvel [1], Annie Ouin [1] and Stéphane Girard [2]

[1] Dynafor, University of Toulouse, INRA, INPT, INPT-EI PURPAN, Castanet-Tolosan, France
[2] Team MISTIS, University Grenoble-Alpes, INRIA, LJK, Montbonnot, France

MultiTemp 2017, Bruges, Belgium, 27-29 June 2017

Grasslands, one of the main biodiversity resources

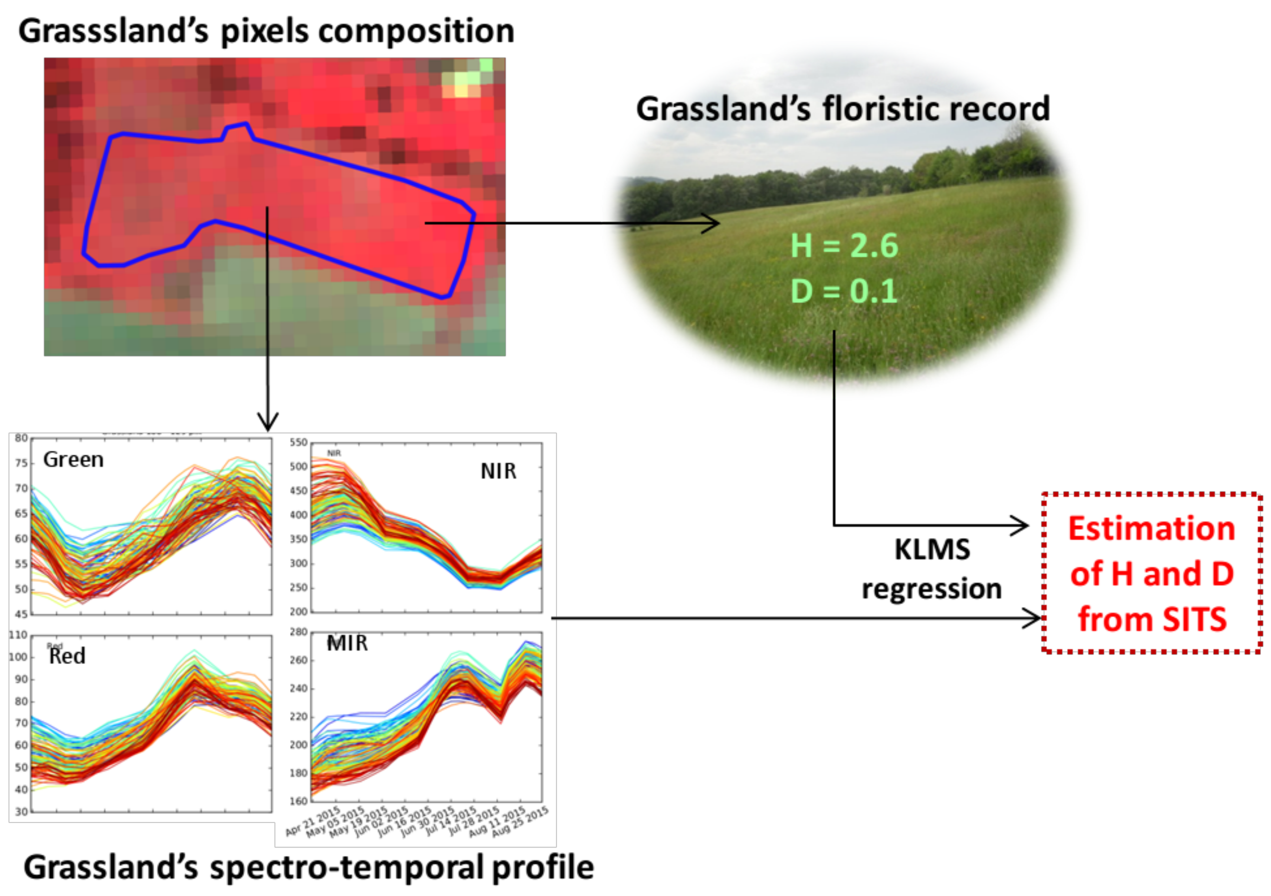
Context

- Grasslands are one of the main **biodiversity** resources in rural landscapes.
- Importance of **monitoring grassland biodiversity over large extents**.
- **Biodiversity indices** are defined at the **grassland scale**.
- It is better to use very high spatial resolution (<1m) and hyperspectral data to discriminate grassland species. But their availability is limited.
- *Tradeoff*: time series with **high spatial resolution** and **very high temporal resolution** because species differ in their temporal behavior (phenology).

Objectives of this study

Assess the potential of multispectral satellite image time series (SITS) with **high spatial and high temporal resolutions** to **estimate plant biodiversity** (Shannon and Simpson indices) at the grassland scale.

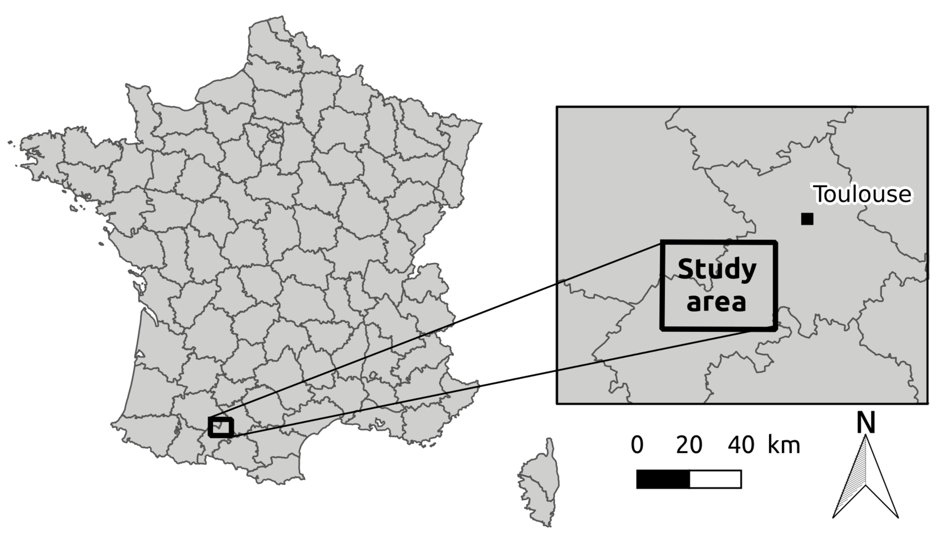
Principle



Study site and data

Study area

Long-Term Ecolog. Research site "Coteaux et Vallées de Gascogne", France.



Field data

- **Floristic composition** at the **grassland scale** recorded in 2015 and 2016, in **192 grasslands**.
- Computation of **abundance-based biodiversity indices**:
  - **Shannon index** H $= -\sum_{i=1}^{R} p_i \ln p_i$
  - **Simpson index** D $= \sum_{i=1}^{R} p_i^2$

where p_i is the proportion of the ith species and R is the total number of species in the grassland (species richness).

Variable	Min	Max	Mean	SD	CV
H	0.10	3.51	2.27	0.49	0.22
D	0.049	0.973	0.168	0.126	0.752

Satellite data

Two **multispectral** *(MS)* or **NDVI intra-annual** (April to September) **SITS**:

SITS	**SPOT5 (Take5)** *(SPOT)*	**Sentinel-2** *(S2)*
Year	2015	2016
Spatial res.	10 meters	10 meters and 20 meters
Spectral bands	Green, Red, Near Infrared (NIR), Mid Infrared (MIR), $n_B = 4$	Blue, Green, Red, NIR (10m), 3 red-edge bands and 1 narrow NIR (20m resampled at 10m), $n_B = 8$
Acquisitions	$n_T = 18$ dates	n_T = 7 dates

Methodology

Grassland modeling

Each grassland g_i composed of n_i **pixels** represented by a **spectro-temporal vector** $\mathbf{x}_{ik} \in \mathbb{R}^d$, where $d = n_B n_T$ is the number of spectro-temporal variables. Two grassland representations: by its **mean vector** $\boldsymbol{\mu}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} \mathbf{x}_{ik} \in \mathbb{R}^d$ and by its **whole set of pixels** $\mathbf{x}_{ik}$. **One response variable** $y_i \in \mathbb{R}$ per grassland.

Kernel least mean square (KLMS) regression.

The **KLMS regression** [1] consists in solving: $\min_f \sum_{i=1}^{G} \Big(y_i - f(g_i)\Big)^2 + \theta\|f\|^2$,

where f is the regression function such as $f(g_i) = \hat{y}_i = \sum_{j=1}^{G} \beta_j K(g_i, g_j) + b$, $\hat{y}_i$ is the predicted variable associated with g_i, K is the kernel function, β_j's are the parameters of f, b is the intercept and θ is the regularization hyperparameter. β_j and b are found by least-square minimization.

Two kernels based on two grassland modelings are investigated:

- **Mean modeling** and RBF kernel ***μ-KLMS***: $K_{\text{RBF}}(g_i, g_j) = \exp(-\sigma\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2)$.
- **Empirical mean kernel *EMK-KLMS***: $K_{\text{EMP}}(g_i, g_j) = \frac{1}{n_i n_j}\sum_{k,l=1}^{n_i,n_j} K_{\text{RBF}}(\mathbf{x}_{ik}, \mathbf{x}_{jl})$.

Protocol

Regression **repeated over 10 runs**, dataset **randomly split into two subsets**: 80% for training and 20% for testing.
Optimal hyperparameters tuned during a **5-fold cross-validation** based on the highest **coefficient of determination**: $r^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$.

Results: a low estimation accuracy

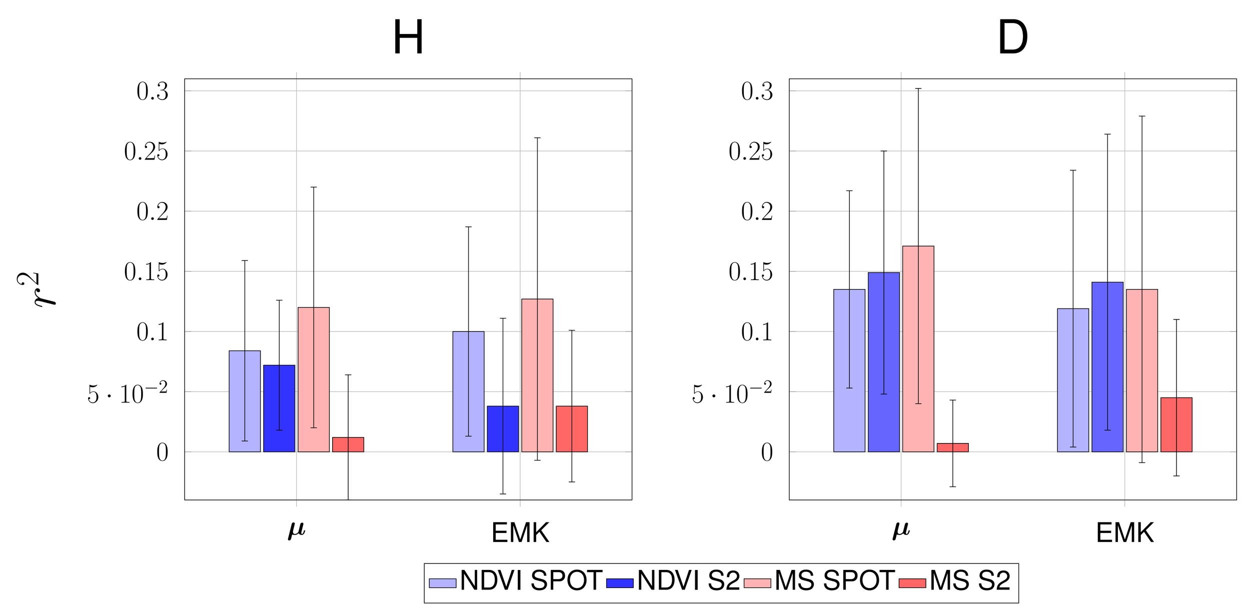


Figure 1: Mean and standard deviation of r^2 over the 10 repetitions.

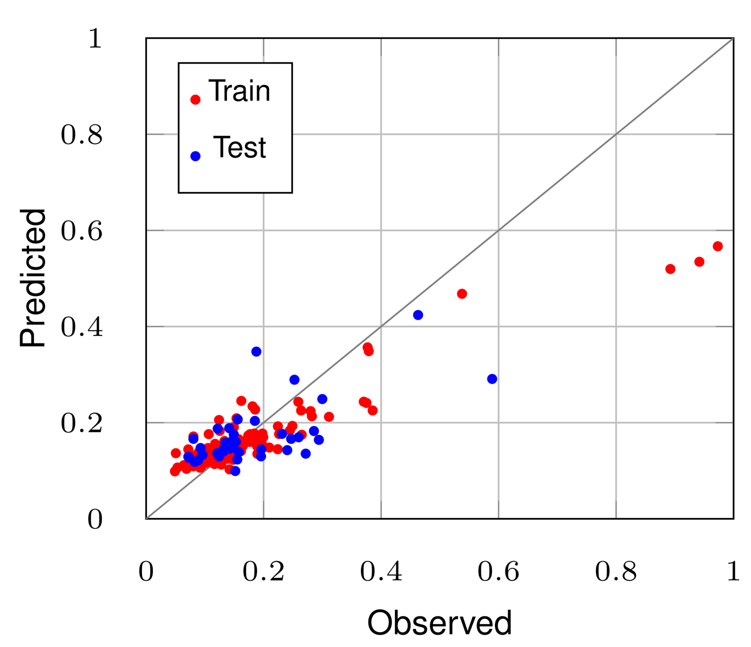


Figure 2: Best run for D prediction with μ-KLMS using MS SPOT5 data, $r^2 = 0.43$.

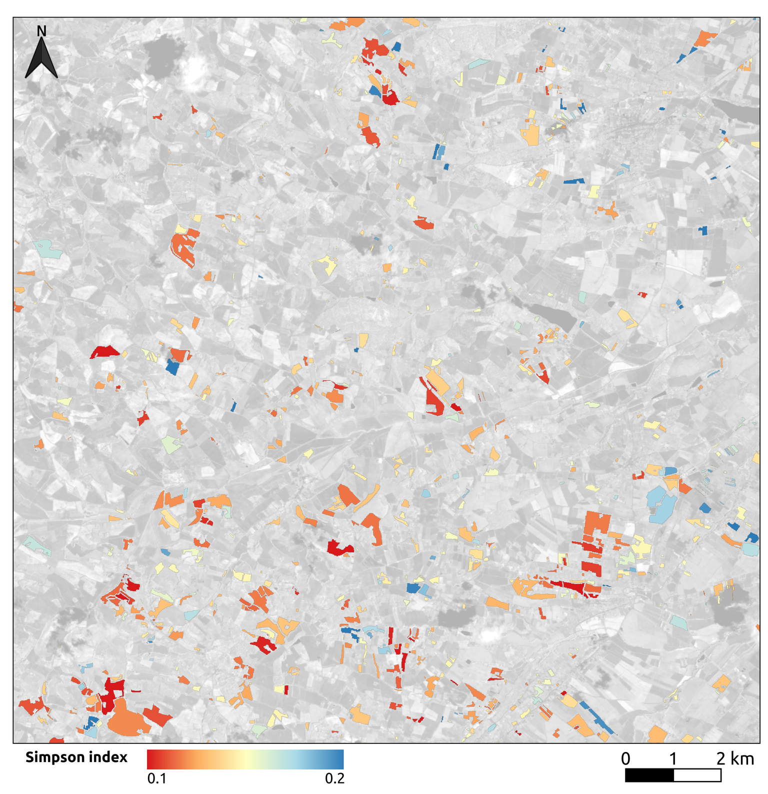


Figure 3: Estimation of Simpson index of all the grasslands in the area.

Conclusions and prospects

- **Lack of variance** in the predicted dataset.
- Results suggest that **high temporal resolution** combined with high spatial resolution are **not sufficient to estimate plant biodiversity**.
- **Simpson index** was better predicted than Shannon index.
- Prospect: **Spectral heterogeneity** [2] as a proxy for species diversity.

References

[1] W. Liu *et al.*, "The kernel least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 56, pp. 543–554, Feb 2008.

[2] D. Rocchini *et al.*, "Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges," *Ecological Informatics*, vol. 5, no. 5, pp. 318 – 329, 2010.