Grasslands species diversity mapping from hyperspectral remote sensing

5^e Colloque Groupe Hyperspectral SFPT-GH

M. Lopes [1], M. Fauvel [1], A. Ouin [1] and S. Girard [2]

[1] UMR 1201 DYNAFOR INRA & Institut National Polytechnique de Toulouse
[2] Equipe MISTIS-LJK, Universite Grenoble Alpes, INRIA, France

[2017-05-11 Thu]

Outline

Outline

Grasslands species diversity

- Grasslands represent a significant source of biodiversity in farmed landscapes,
- They provide many ecosystem services (carbon regulation, erosion regulation, pollination. . .),
- Grasslands surface area and their diversity are declining [OMa12],
- Maps over grassland diversity are required over large area extents.

Spectral Variation Hypothesis

- It assumes that the spectral heterogeneity is correlated with spatial variations and heterogeneity of the habitat [Pal+02]
- Spectral heterogeneity can be used as a proxy for species diversity [Roc+16]
- Several indices have been proposed
  - Standard deviation or coefficient of variations of NDVI
  - PCA
  - Distance to centroids
  - Clustering

Objectives

- Project MUESLI
- Use hyperspectral images to monitor species richness at the parcel level
- Methodological contributions
  - Use of robust high dimensional clustering method
  - Extend conventional heterogeneity/diversity index

Outline

Spectral heterogenity

- Proposed by Rocchini *et. al* [Roc+16]

Spectral heterogenity

- Proposed by Rocchini *et. al* [Roc+16]
- It consists in computing the mean euclidean distance to the centroid of a given plot:

$$H(p) = \frac{1}{n_p}\sum_{i\in p}^{n_p} \|\mathbf{x}_i - \boldsymbol{\mu}_p\|^2$$

where

$$\boldsymbol{\mu}_p = \frac{1}{n_p}\sum_{i\in p}^{n_p} \mathbf{x}_i.$$

Spectral heterogenity

- Proposed by Rocchini *et. al* [Roc+16]
- It consists in computing the mean euclidean distance to the centroid of a given plot:

$$H(p) = \frac{1}{n_p}\sum_{i\in p}^{n_p}\|\mathbf{x}_i - \boldsymbol{\mu}_p\|^2$$

where

$$\boldsymbol{\mu}_p = \frac{1}{n_p}\sum_{i\in p}^{n_p}\mathbf{x}_i.$$

- Equivalently, it can be computed as the trace of the empirical covariance matrix of the plot:

$$H(p) = \text{Trace}\big(\boldsymbol{\Sigma}_p\big).$$

Spectral heterogenity

- Proposed by Rocchini *et. al* [Roc+16]
- It consists in computing the mean euclidean distance to the centroid of a given plot:

$$H(p) = \frac{1}{n_p}\sum_{i\in p}^{n_p}\|\mathbf{x}_i - \boldsymbol{\mu}_p\|^2$$

where

$$\boldsymbol{\mu}_p = \frac{1}{n_p}\sum_{i\in p}^{n_p}\mathbf{x}_i.$$

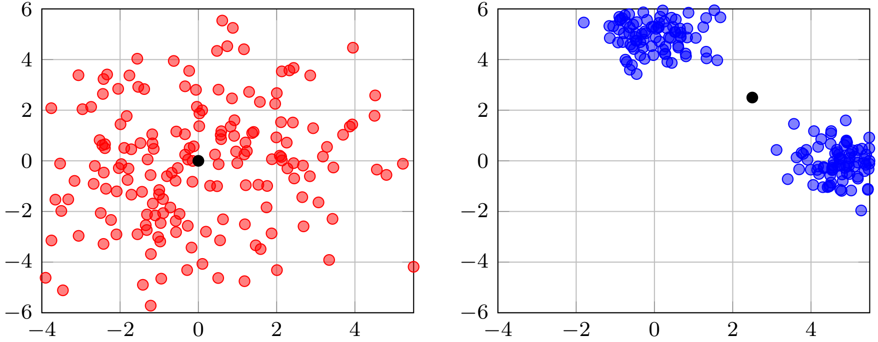
- Equivalently, it can be computed as the trace of the empirical covariance matrix of the plot:

$$H(p) = \text{Trace}\big(\Sigma_p\big).$$

- Variant: first reduce the dimensionality (PCA, ...)

Why MDC may not work

The following configurations have the same MDC



α-diversity

- Proposed by Feret *et. al* [FA14]

α-diversity

- Proposed by Feret *et. al* [FA14]
- Estimated by the Shannon entropy of a given plot

$$E_p = -\sum_{s=1}^{S} p_s \log(p_s)$$

where p is the considered plot, S the total number of species/classes/clusters and p_s is the relative proportion.

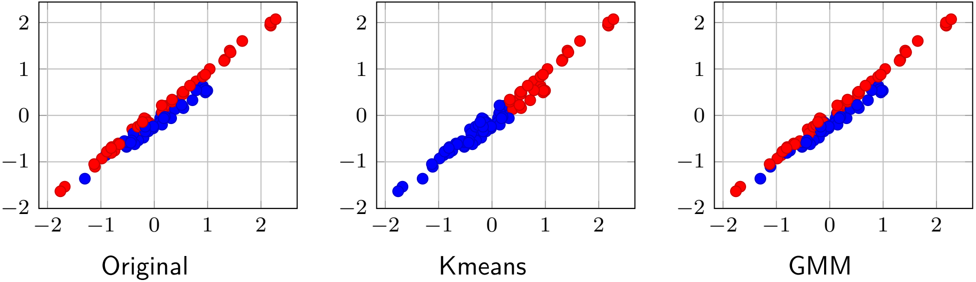
α-diversity

- Proposed by Feret *et. al* [FA14]
- Estimated by the Shannon entropy of a given plot

$$E_p = -\sum_{s=1}^{S} p_s \log(p_s)$$

  where p is the considered plot, S the total number of species/classes/clusters and p_s is the relative proportion.
- Clusters estimated through the *PCA+Kmeans* pipeline applied on the whole image.

Why Kmeans may not work



Outline

Statistical model

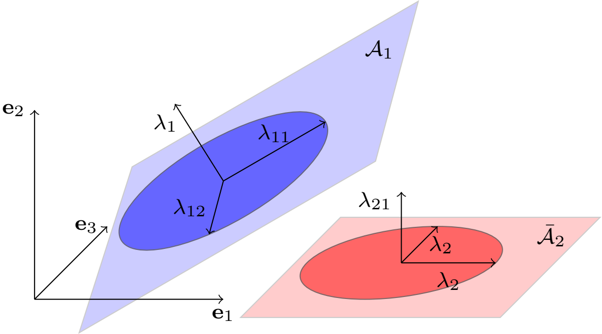
- Mixture model $p(\mathbf{x}) = \sum_{c=1}^{C} \pi_c p(\mathbf{x}|c)$,
- Under Gaussian assumption $p(\mathbf{x}|c)$ is a d-dimensional Gaussian distribution

$$p(\mathbf{x}|c) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_c|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_c)^\top \boldsymbol{\Sigma}_c^{-1}(\mathbf{x}-\boldsymbol{\mu}_c)\right)$$

- *Curse of dimensionality*: special structure for the covariance matrix $\boldsymbol{\Sigma_c} = \mathbf{Q}_c \boldsymbol{\Lambda_c} \mathbf{Q}_c^\top$

$$\boldsymbol{\Lambda}_c = \begin{array}{cc} \left(\begin{array}{cc} \boxed{\begin{array}{ccc} \color{red}{\lambda_{c1}} & & 0 \\ & \ddots & \\ 0 & & \color{red}{\lambda_{cp_i}} \end{array}} & \mathbf{0} \\ \mathbf{0} & \boxed{\begin{array}{cccc} \color{blue}{\lambda_c} & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \color{blue}{\lambda_c} \end{array}} \end{array}\right) & \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} \quad p_c \\ \left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\} \quad (d-p_c) \end{array} \end{array}$$

High dimensional GMM [BGS07]



Under the HDDA model

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

with $\tilde{\mathbf{Q}}_i = \left[\mathbf{q}_{i1}, \ldots, \mathbf{q}_{ip_i}\right]$, $\tilde{\boldsymbol{\Lambda}}_i = \text{diag}\left[\lambda_{i1} - \lambda_i, \ldots, \lambda_{ip_i} - \lambda_i\right]$, $\tilde{\mathbf{V}}_i = \text{diag}\left[\frac{1}{\lambda_{i1}} - \frac{1}{\lambda_i}, \ldots, \frac{1}{\lambda_{ip_i}} - \frac{1}{\lambda_i}\right]$ and $\mathbf{I}_d$ is the identity matrix of size d.

Spectral heterogeneity revisited 1/2

- Samples covariance matrix for a given plot p

$$\boldsymbol{\Sigma}_p = \mathbf{B}_p + \mathbf{W}_p$$

where

- $\mathbf{B}_p$ is the between class covariance matrix of plot p

$$\mathbf{B}_p = \sum_{c=1}^{C_p} \pi_{pc}(\boldsymbol{\mu}_{pc} - \boldsymbol{\mu}_p)(\boldsymbol{\mu}_{pc} - \boldsymbol{\mu}_p)^\top$$

- $\mathbf{W}_p$ is the within class covariance matrix of plot p

$$\mathbf{W}_p = \sum_{c=1}^{C_p} \pi_{pc}\boldsymbol{\Sigma}_{pc}$$

Spectral heterogeneity revisited 2/2

- $\text{Trace}(\mathbf{\Sigma}_p) = \text{Trace}(\mathbf{B}_p) + \text{Trace}(\mathbf{W}_p)$
- $\text{Trace}(\mathbf{B}_p) = \sum_{c=1}^{C_p} \pi_{pc} \|\boldsymbol{\mu}_{pc} - \boldsymbol{\mu}_p\|^2$
- $\text{Trace}(\mathbf{W}_p) = \frac{1}{n_p} \sum_{i=1}^{C_p} \sum_{k \in c} \|\mathbf{x}_{pk} - \boldsymbol{\mu}_{pc}\|^2$

	$\text{Trace}(\mathbf{\Sigma}_p)$	$\text{Trace}(\mathbf{B}_p)$	$\text{Trace}(\mathbf{W}_p)$
Plot 1	13.63	0	13.63
Plot 2	13.74	12.71	0.973

Improved spectral entropy

- For each pixel of the plot, the vector of posterior probabilities is available

$$\left[p(C = 1|\mathbf{x}), \ldots, p(C = C_p|\mathbf{x})\right]$$

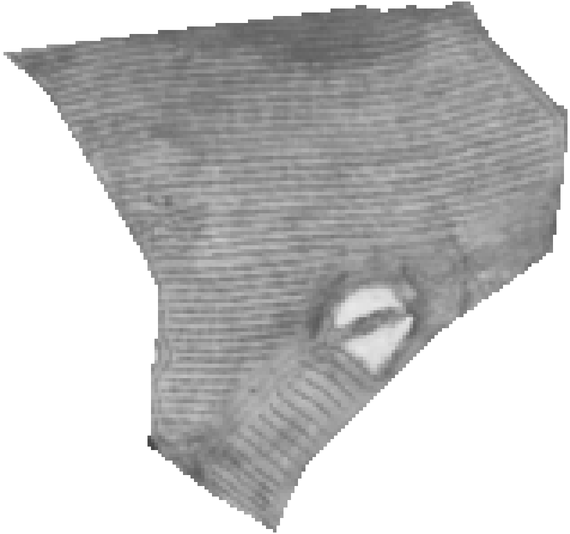
- The relative proportion is then computed as:

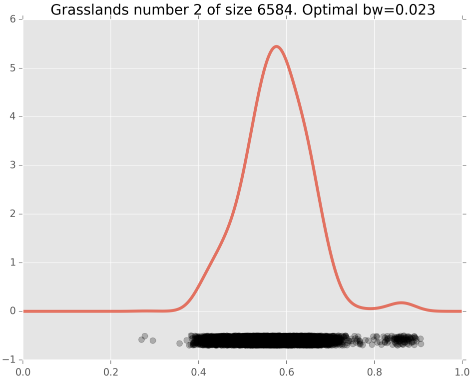
$$p_c = \frac{1}{n_p}\sum_{k\in c} p(C = c|\mathbf{x}) = \pi_c$$

- It allows to let a pixel belonging to several clusters (not a crisp affectation)

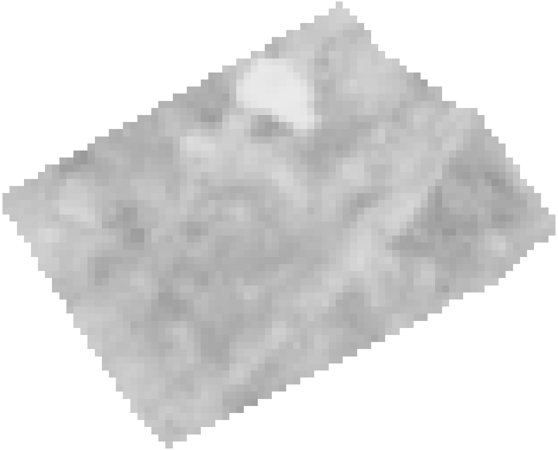
Outline

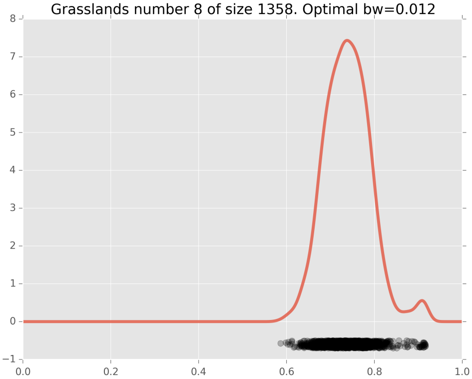
Data collection



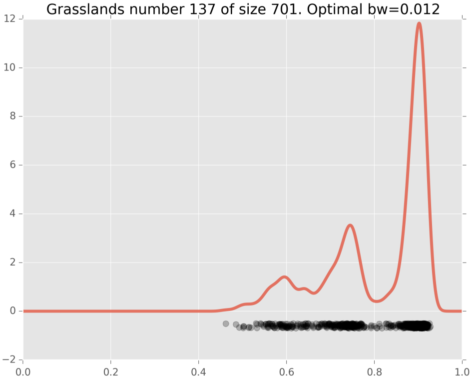


Data collection



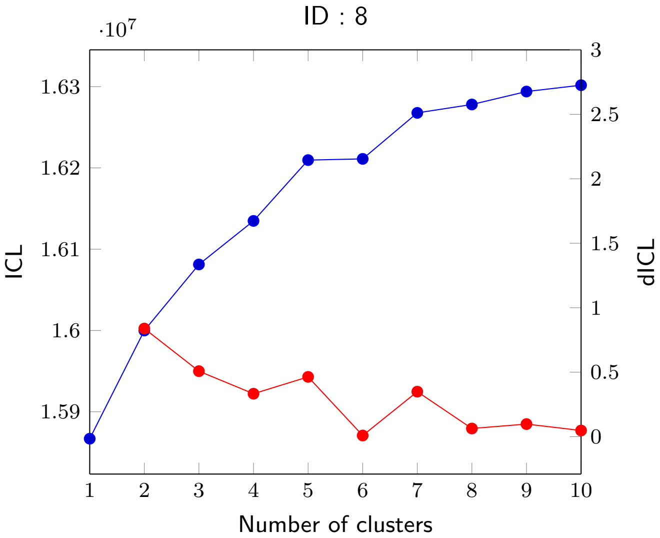


Data collection



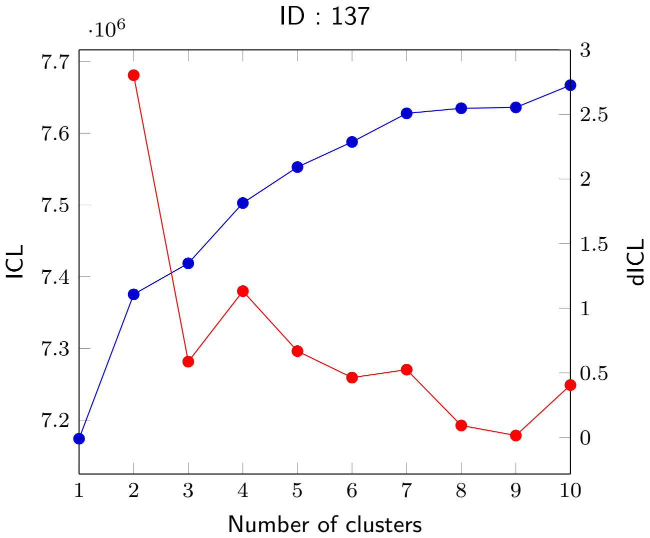
Simulations

- Select the number of classes using ICL: stop when dICL<1%



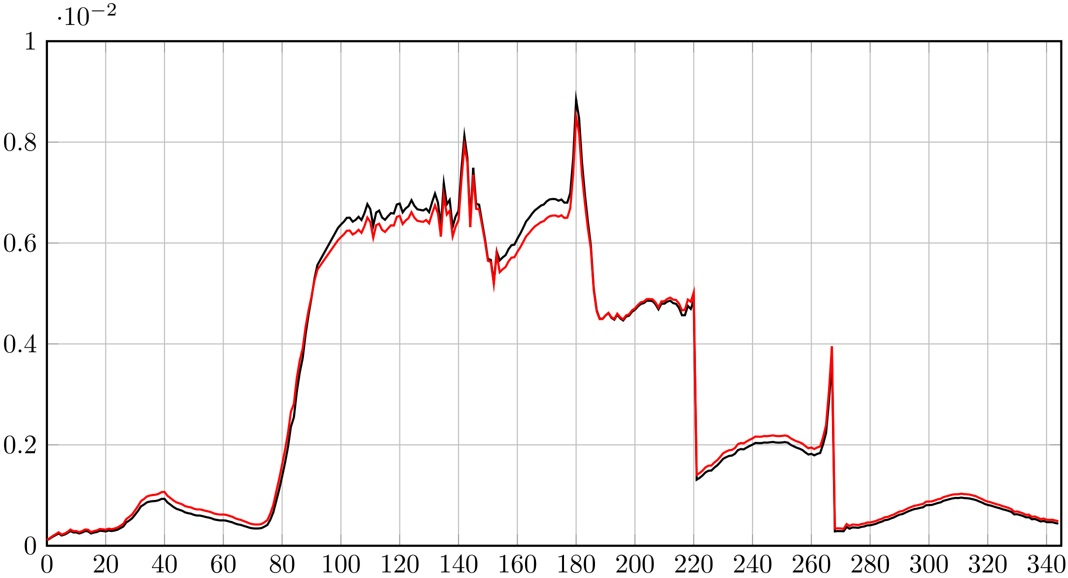
Simulations

- Select the number of classes using ICL: stop when dICL$<1\%$



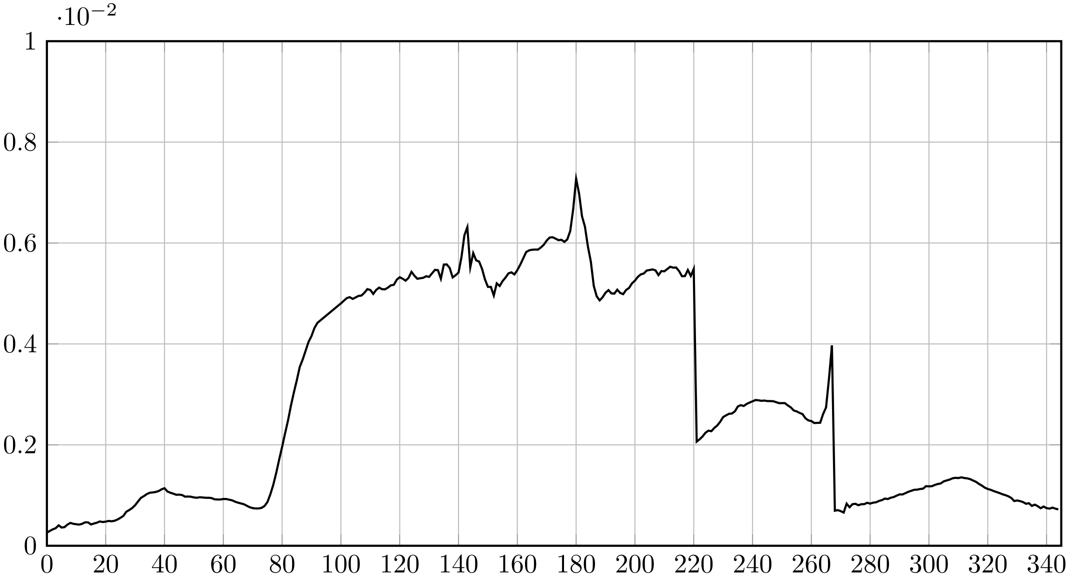
Outline

Clusters



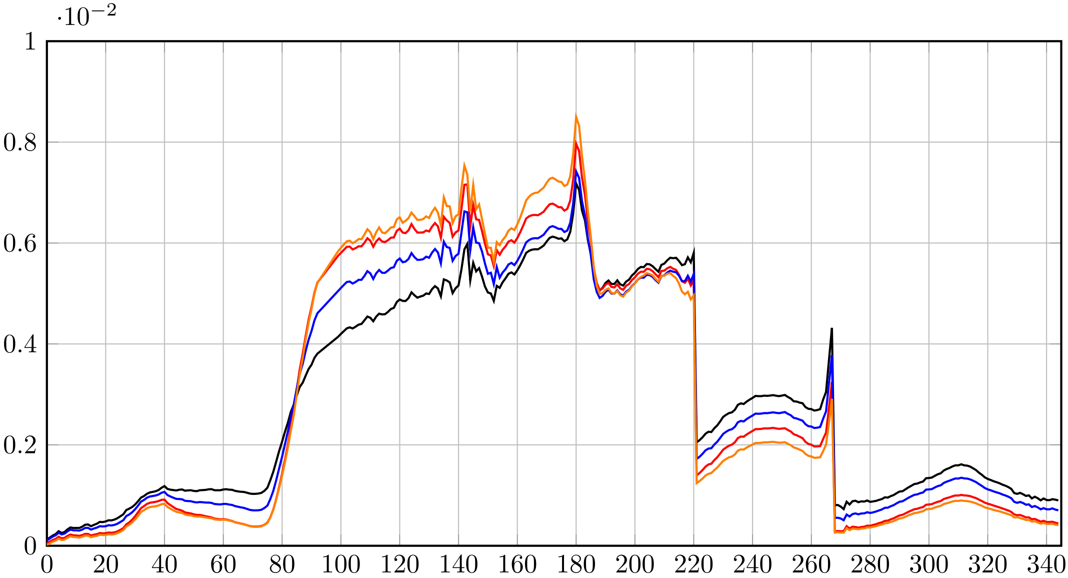
$\cdot 10^{-2}$
1
0.8
0.6
0.4
0.2
0
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340

Clusters



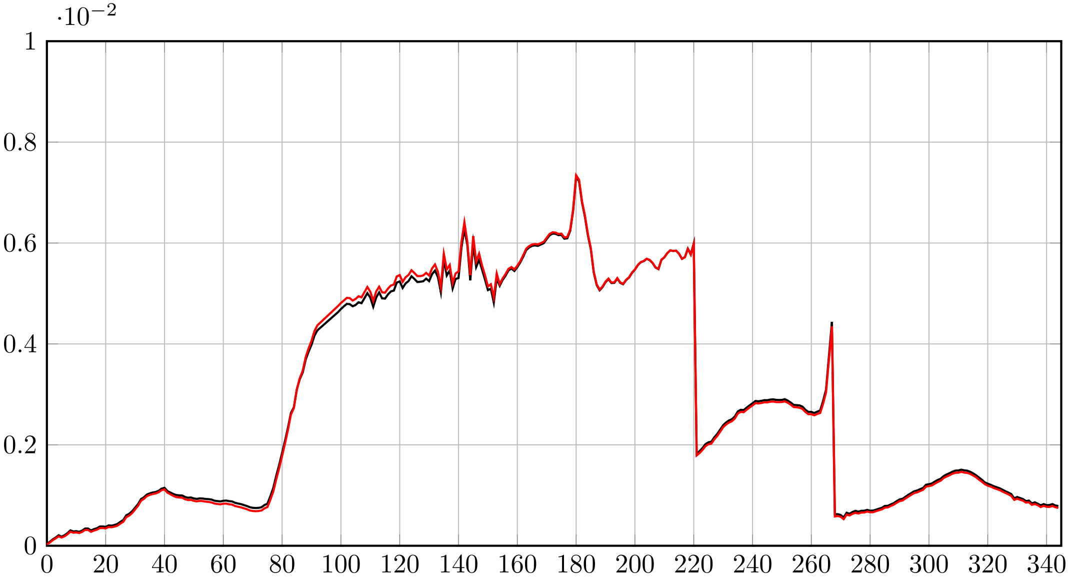
$\cdot 10^{-2}$
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340

Clusters



$\cdot 10^{-2}$
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340

Clusters


$\cdot 10^{-2}$
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340

Measure of heterogeneity

ID	C	E	B	W	V	H	D
6	2	0.68	13.16	11.32	11.17	0.97	0.13
8	1	0.0	inf	11.12	11.12	0.09	3.81
137	4	1.31	10.36	10.97	9.93	0.08	3.97
143	2	0.68	15.02	11.57	11.54	0.04	5.06

- B, W and V are in *log* scale
- $E \approx \log(C)$

Outline

Conclusions and perspectives

- Species diversity in semi-natural grasslands

Conclusions and perspectives

- Species diversity in semi-natural grasslands
- Extension of heterogeneity measures with high dimensional clustering techniques

Conclusions and perspectives

- Species diversity in semi-natural grasslands
- Extension of heterogeneity measures with high dimensional clustering techniques
- Estimated diversity does not correlate (yet!) with field work

Bibliography I

Bouveyron, Charles, Stephane Girard, and Cordelia Schmid. "High-Dimensional Data Clustering". In: *Computational Statistics and Data Analysis* 52.1 (Sept. 2007), pp. 502–519. DOI: 10.1016/j.csda.2007.02.009. URL: https://hal.archives-ouvertes.fr/hal-00022183.

Féret, Jean-Baptiste and Gregory P Asner. "Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy". In: *Ecological Applications* 24.6 (2014), pp. 1289–1296.

O'Mara, F. P. "The role of grasslands in food security and climate change". In: *Annals of Botany* 110.6 (2012), p. 1263. DOI: 10.1093/aob/mcs209. eprint: /oup/backfile/Content_public/Journal/aob/110/6/10.1093/aob/mcs209/2/mcs209.pdf. URL: +%20http://dx.doi.org/10.1093/aob/mcs209.

Palmer, Michael W. et al. "Quantitative tools for perfecting species lists". In: *Environmetrics* 13.2 (2002), pp. 121–137. ISSN: 1099-095X. DOI: 10.1002/env.516. URL: http://dx.doi.org/10.1002/env.516.

Rocchini, Duccio et al. "Satellite remote sensing to monitor species diversity: potential and pitfalls". In: *Remote Sensing in Ecology and Conservation* 2.1 (2016), pp. 25–36. ISSN: 2056-3485. DOI: 10.1002/rse2.9. URL: http://dx.doi.org/10.1002/rse2.9.