Grasslands species diversity mapping from hyperspectral remote sensing

5th 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

Context

Measure of heterogeneity

High dimensional discriminant analysis

Experimental protocol

Primary results

Conclusions and perspectives
Outline

Context

Measure of heterogeneity

High dimensional discriminant analysis

Experimental protocol

Primary results

Conclusions and perspectives
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

Context

Measure of heterogeneity

High dimensional discriminant analysis

Experimental protocol

Primary results

Conclusions and perspectives
Spectral heterogeneity

- Proposed by Rocchini et al [Roc+16]
Spectral heterogeneity

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

- It consists in computing the mean euclidean distance to the centroid of a given plot:

\[
H(p) = \frac{1}{n_p} \sum_{i \in p} ||x_i - \mu_p||^2
\]

where

\[
\mu_p = \frac{1}{n_p} \sum_{i \in p} x_i.
\]
Spectral heterogeneity

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

\[
H(p) = \frac{1}{n_p} \sum_{i \in p} \|x_i - \mu_p\|^2
\]

where

\[
\mu_p = \frac{1}{n_p} \sum_{i \in p} x_i.
\]

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

\[
H(p) = \text{Trace}(\Sigma_p).
\]
Spectral heterogeneity

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

\[ H(p) = \frac{1}{n_p} \sum_{i \in p} \| x_i - \mu_p \|^2 \]

where

\[ \mu_p = \frac{1}{n_p} \sum_{i \in p} x_i. \]

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

\[ H(p) = \text{Trace}(\Sigma_p). \]

- 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]
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

![Graphs showing original, Kmeans, and GMM results.](image-url)
Outline

Context

Measure of heterogeneity

High dimensional discriminant analysis

Experimental protocol

Primary results

Conclusions and perspectives
Statistical model

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

\[
p(x|c) = \frac{1}{(2\pi)^{d/2}|\Sigma_c|^{1/2}} \exp \left( -\frac{1}{2}(x - \mu_c)^\top \Sigma_c^{-1}(x - \mu_c) \right)
\]

- **Curse of dimensionality**: special structure for the covariance matrix \( \Sigma_c = Q_c \Lambda_c Q_c^\top \)

\[
\Lambda_c = \begin{pmatrix}
\lambda_{c1} & 0 & \cdots & 0 \\
0 & \cdots & \lambda_{cp_i} \\
0 & \cdots & \cdots & \lambda_c \\
0 & \cdots & \cdots & \lambda_c \\
\end{pmatrix}
\]

\[
\begin{cases}
p_c \\(d - p_c)\end{cases}
\]
High dimensional GMM [BGS07]

Under the HDDA model

\[
\Sigma_i = \tilde{Q}_i \tilde{\Lambda}_i \tilde{Q}_i^\top + \lambda_i I_d
\]
\[
\Sigma_i^{-1} = \tilde{Q}_i \tilde{V}_i \tilde{Q}_i^\top + \lambda_i^{-1} I_d
\]

with \( \tilde{Q}_i = [q_{i1}, \ldots, q_{ip_i}] \), \( \tilde{\Lambda}_i = \text{diag} [\lambda_{i1} - \lambda_i, \ldots, \lambda_{ip_i} - \lambda_i] \), \( \tilde{V}_i = \text{diag} [\frac{1}{\lambda_{i1}} - \frac{1}{\lambda_i}, \ldots, \frac{1}{\lambda_{ip_i}} - \frac{1}{\lambda_i}] \)

and \( I_d \) is the identity matrix of size \( d \).
Spectral heterogeneity revisited 1/2

- Samples covariance matrix for a given plot \( p \)

\[
\Sigma_p = B_p + W_p
\]

where

- \( B_p \) is the between class covariance matrix of plot \( p \)

\[
B_p = \sum_{c=1}^{C_p} \pi_{pc} (\mu_{pc} - \mu_p) (\mu_{pc} - \mu_p)^\top
\]

- \( W_p \) is the within class covariance matrix of plot \( p \)

\[
W_p = \sum_{c=1}^{C_p} \pi_{pc} \Sigma_{pc}
\]
Spectral heterogeneity revisited 2/2

\[ \text{Trace}(\Sigma_p) = \text{Trace}(B_p) + \text{Trace}(W_p) \]

\[ \text{Trace}(B_p) = \sum_{c=1}^{C_p} \pi_{pc} \|\mu_{pc} - \mu_p\|^2 \]

\[ \text{Trace}(W_p) = \frac{1}{n_p} \sum_{i=1}^{C_p} \sum_{k \in c} \|x_{pk} - \mu_{pc}\|^2 \]

<table>
<thead>
<tr>
<th></th>
<th>Trace(\Sigma_p)</th>
<th>Trace(B_p)</th>
<th>Trace(W_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot 1</td>
<td>13.63</td>
<td>0</td>
<td>13.63</td>
</tr>
<tr>
<td>Plot 2</td>
<td>13.74</td>
<td>12.71</td>
<td>0.973</td>
</tr>
</tbody>
</table>
Improved spectral entropy

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

\[ p(C = 1|x), \ldots, p(C = C_p|x) \]

■ The relative proportion is then computed as:

\[ p_c = \frac{1}{n_p} \sum_{k \in c} p(C' = c|x) = \pi_c \]

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

Context

Measure of heterogeneity

High dimensional discriminant analysis

Experimental protocol

Primary results

Conclusions and perspectives
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 $d\text{ICL}<1\%$
Outline

Context

Measure of heterogeneity

High dimensional discriminant analysis

Experimental protocol

Primary results

Conclusions and perspectives
Clusters
Clusters

The graph shows the distribution of clusters across various data points. The x-axis represents the data points, while the y-axis shows the measure of heterogeneity on a logarithmic scale. The clusters are indicated by different colored lines, each representing a distinct group within the data.
Clusters

- Measure of heterogeneity
- High dimensional discriminant analysis
- Experimental protocol
- Primary results
- Conclusions and perspectives
- References
Measure of heterogeneity

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

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

Context

Measure of heterogeneity

High dimensional discriminant analysis

Experimental protocol

Primary results

Conclusions and perspectives
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.


