HIGH DIMENSIONAL KULLBACK-LEIBLER DIVERGENCE FOR GRASSLAND MANAGEMENT PRACTICES CLASSIFICATION FROM HIGH RESOLUTION SATELLITE IMAGE TIME SERIES

Mailys Lopes(1), *Mathieu Fauvel*(1), *Stéphane Girard*(2), *David Sheeren*(1)

(1)DYNAFOR, University of Toulouse, INRA, INPT, France
(2)Team MISTIS, University of Grenoble, INRIA, LJK, France

ABSTRACT

The aim of this study is to classify grassland management practices using satellite image time series with high spatial resolution. The study area is located in southern France where 52 parcels with 3 management types were selected. The spectral variability inside the grasslands was taken into account considering that the pixels signal can be modeled by a Gaussian distribution. A parsimonious model is discussed to deal with the high dimension of the data and the small sample size. A high dimensional symmetrized Kullback-Leibler divergence (KLD) is introduced to compute the similarity between each pair of grasslands. The model is positively compared to the conventional KLD to construct a positive definite kernel used in SVM for supervised classification.

Index Terms— Satellite image time series, high dimension, Kullback-Leibler divergence, grassland management type, classification.

1. INTRODUCTION

In the frame of sustainable development, the study of landscape state and its evolution are required to understand environmental changes and biodiversity loss. To this aim, research in landscape ecology is devoted to understanding how the landscape configuration and composition impact on biodiversity and services provided. This research requires the identification and characterization of semi-natural elements in the landscape. Indeed, semi-natural habitats are perennial and less inclined to be disturbed. They are sources of biodiversity in farmed landscapes. Particularly, permanent grasslands, as they represent one of the largest terrestrial landscape (they cover 18% of France territory [1]), are a source of significant animal and vegetal biodiversity [2, 3], providing many ecosystem services such as carbon storage, erosion regulation, crop pollination, biological regulation of ravagers [4]. Although policies have been adopted to protect biodiversity in semi-natural landscapes (European Union Habitats Directive, 92/43/EEC), the permanent grasslands area is continuously decreasing, leading to a loss of biodiversity [3, 5].

Grasslands being the main livestock feeding resource, the species composition in semi-natural grasslands is also impacted by the management practices [5, 6] . Indeed, the anthropic events on the grasslands, like mowing and/or casual grazing, disturb the natural cycle and the structure of the vegetation. Therefore, it is essential to identify the management practices in each parcel in order to predict their effect on biodiversity and related ecosystem services.

In this context, remote sensing appears to be an appropriate tool to characterize grasslands at the landscape scale, because of the large spatial coverage and revisit frequency of satellite sensors. However, the reflected signal of the grasslands is more difficult to interpret compared to mono-specific lands like crops, due to the diversity and the mixing of grassland species. Furthermore, grasslands are relatively small elements of the landscape (in the range of the hectare), which require high spatial resolution data to be distinguishable [7]. Given their phenological cycle and the punctuality of the anthropic event (*e.g.*, mowing), very dense time series through the vegetation cycle are necessary to identify the management types [8].

Until recently, satellite missions offering high frequency of revisit had low spatial resolution (*i.e.*, MODIS), and high spatial resolution missions did not offer dense time series. New missions like Sentinel-2 [9], with very high revisit frequency (5 days) and high spatial resolution (10 meters) offer new possibilities for grasslands monitoring [10].

The aim of this study is to identify grassland management practices using time series of a spectral vegetation index (NDVI) with high temporal resolution. Management practices are defined at the parcel scale. Classical pixel-oriented approaches result in the appearance of misclassified pixels within a class [11], leading to non-homogeneous objects that are ecologically unrealistic.

The first contribution of the method was to take into account the spectral variability in a grassland. We considered that the distribution of the pixel spectral reflectance in a given grassland can be modeled by a Gaussian distribution. Then, the Kullback-Leibler divergence was used to compute the distance between each couple of grasslands. To deal with the

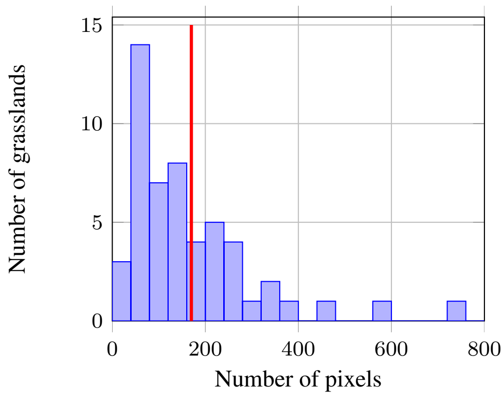

Fig. 1. Distribution of the number of pixels per grassland. The red line corresponds to the number of parameters to be estimated for each grassland for a conventional multivariate Gaussian model. This number is derived from the number of variables using the formula $d(d+3)/2 = 170$ for $d = 17$.

small sample size compared to the number of temporal variables, a parsimonious Gaussian model was proposed as a second contribution.

The object-oriented approach was compared to a pixel-based approach, through supervised classification.

In section 2, the dataset is introduced. Then, the high dimensional Kullback-Leibler divergence method that we developped is described in section 3. Finally, the experimental results conducted on real SITS are presented.

2. DATASET

2.1. Study site

The study site is located in south-west France, near Toulouse, in a semi-rural area where livestock farming is in decline in favor of field crops. Grasslands are mostly used for forrage or silage production. The extent of the area corresponds to the satellite image extent (4400 km^2).

2.2. Field Data

The dataset is composed of 52 parcels with their management methods. The homogeneity has been controlled during a field survey in May, 2015, where the past and current management practices were also determined, by interviewing the farmers or grassland owners. We identified 3 management types during the vegetation cycle: one mowing (34 parcels), grazing (10 parcels) and mixed management (mowing then grazing, 8 parcels). We used them as classes for the classification. The grasslands have been digitalized by hand.

2.3. Satellite data

The satellite image time series (SITS) is composed of 17 multispectral Formosat-2 images (8 meters spatial resolution) from 2013. The images are provided with a mask of clouds and shadows [12]. The Normalized Difference Vegetation Index (NDVI) [13] was used during this study: each pixel $\mathbf{x} \in \mathbb{R}^{17}$.

To remove the noise due to clouds and shadows in the SITS, the NDVI was smoothed applying the Whittaker filter pixel-by-pixel [14].

3. HIGH DIMENSIONAL KULLBACK-LEIBLER DIVERGENCE (HDKLD) METHOD

3.1. Symmetrized KLD

The pixel reflectance distribution of grasslands is modeled by a Gaussian distribution, *i.e.* the density function of pixels $\mathbf{x}$ is, conditionally to grassland g_i, a Gaussian distribution. To compute the similarity of the distribution of each grassland, we used the symmetrized Kullback-Leibler divergence [15]. The symmetrized KLD between two Gaussian distributions can be written as:

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

where $\boldsymbol{\Sigma}_i$ is the covariance matrix, $\boldsymbol{\mu}_i$ is the mean vector of the signal, d the number of variables and Tr is the trace operator. They are estimated by their empirical counterparts $\hat{\boldsymbol{\mu}}_i = \frac{1}{n_i}\sum_{l=1}^{n_i}\boldsymbol{x}_l$ and $\hat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i}\sum_{l=1}^{n_i}(\boldsymbol{x}_l - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_l - \hat{\boldsymbol{\mu}}_i)^\top$ with n_i the number of pixels in grassland i.

Unfortunately, the number of pixels used in the estimation is low compared to the number of variables. Figure 1 shows that the number of pixels of most grasslands is lower than the number of parameters to estimate. Thus, the covariance matrix is non invertible for these grasslands. Furthermore, for the other grasslands, the estimated covariance matrices in eq.(1) are ill-conditioned making the computation of their inverse numerically unstable. To cope with this issue, specific derivations are considered in the next section.

3.2. High Dimensional Symmetrized KLD

In this work, a high dimensional model is used to model the Gaussian distribution of grasslands [16]. The model assumes that the last (lowest) eigenvalues of the covariance matrix are equal. According to this model, the covariance matrix of grassland i can be written as:

$$\boldsymbol{\Sigma}_i = \mathbf{Q}_i\boldsymbol{\Lambda}_i\mathbf{Q}_i^\top + \lambda_i\mathbf{I}_d \quad (2)$$

where:

- $\mathbf{Q}_i = \big[\mathbf{q}_{i1}, \dots, \mathbf{q}_{ip_i}\big]$,
- $\mathbf{\Lambda}_i = \text{diag}\Big[\lambda_{i1} - \lambda_i, \dots, \lambda_{ip_i} - \lambda_i\Big]$,
- $\mathbf{I}_d$ is the identity matrix of size $d = 17$,
- $\mathbf{q}_{ij}, \lambda_{ij}$ are the j^{th} eigenvalues/eigenvectors of the covariance matrix $\mathbf{\Sigma}_i$, $j \in \{1, \dots, d\}$ such as $\lambda_{i1} \geq \dots \geq \lambda_{id}$,
- p_i is the number of non-equal eigenvalues,
- λ_i is the multiple eigenvalue corresponding to the noise term (last and equal eigenvalues).

Following this model, the inverse of the covariance matrix can be computed explicitly:

$$\mathbf{\Sigma}_i^{-1} = -\mathbf{Q}_i \mathbf{V}_i \mathbf{Q}_i^\top + \lambda_i^{-1} \mathbf{I}_d \tag{3}$$

with $\mathbf{V}_i = \text{diag}\Big[\frac{1}{\lambda_i} - \frac{1}{\lambda_{i1}}, \dots, \frac{1}{\lambda_i} - \frac{1}{\lambda_{ip_i}}\Big]$, and eq.(1) can be written as:

$$\begin{aligned} HDKLD(g_i, g_j) = \\ \frac{1}{2}\Big[- \|\mathbf{\Lambda}_j^{\frac{1}{2}} \mathbf{Q}_j^\top \mathbf{Q}_i \mathbf{V}_i^{\frac{1}{2}}\|_F^2 - \|\mathbf{\Lambda}_i^{\frac{1}{2}} \mathbf{Q}_i^\top \mathbf{Q}_j \mathbf{V}_j^{\frac{1}{2}}\|_F^2 \\ + \lambda_i^{-1} \operatorname{Tr}\big[\mathbf{\Lambda}_j\big] - \lambda_j \operatorname{Tr}\big[\mathbf{V}_i\big] + \lambda_j^{-1} \operatorname{Tr}\big[\mathbf{\Lambda}_i\big] - \lambda_i \operatorname{Tr}\big[\mathbf{V}_j\big] \\ - \|\mathbf{V}_i^{\frac{1}{2}} \mathbf{Q}_i^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 - \|\mathbf{V}_j^{\frac{1}{2}} \mathbf{Q}_j^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 \\ + \frac{\lambda_i + \lambda_j}{\lambda_i \lambda_j} \|(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 + \frac{\lambda_i^2 + \lambda_j^2}{\lambda_i \lambda_j} d\Big] - d \end{aligned} \tag{4}$$

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

3.3. Estimation

The parameters of eq.(4) are estimated for each grassland i from the empirical mean vector and covariance matrix such as [16]:

- $\hat{\lambda}_i = \frac{\operatorname{Tr}(\hat{\mathbf{\Sigma}}_i) - \sum_{j \leq p_i} \hat{\lambda}_{ij}}{d - p_i}$,
- $\hat{\lambda}_{ij}, \hat{\boldsymbol{q}}_{ij}$ are the first eigenvalues/eigenvectors of $\hat{\mathbf{\Sigma}}_i$, $j \in \{1, \dots, p_i\}$. Thus, only the p_i first eigenvalues/eigenvectors are required and the unstable estimation of the eigenvectors associated to small eigenvalues is avoided,
- $\hat{p}_i$ corresponds to the number of eigenvalues needed to reach a given percentage of variance t, $\frac{\sum_{i=1}^{\hat{p}_i} \hat{\lambda}_i}{\sum_{i=1}^{d} \hat{\lambda}_i} \geq t$, t being tuned during learning.

3.4. Properties of (HD)KLD

The (HD)KLD measure is a semi-metric, *i.e.*, it satisfies only three first axioms of a metric [17]: $(HD)KLD(g_i, g_j) \geq 0$, $(HD)KLD(g_i, g_i) = 0$ and $(HD)KLD(g_i, g_j) = (HD)KLD(g_j, g_i)$. This semi-metric can be turned to a positive definite kernel function by plugging it into a radial basis function [18]: $K(g_i, g_j) = \exp\big[-\frac{(HD)KLD(g_i, g_j)^2}{\sigma}\big]$ with $\sigma \in \mathbb{R}_{>0}$. This kernel is used in the experimental section with a SVM.

4. EXPERIMENTAL RESULTS

4.1. Competitive method

In order to test the effectiveness of the proposed approach, the kernel built in the previous section was used to classify the data using SVM, both for the conventional KLD and the HDKLD. In order to make tractable the inverse problem in KLD, a small (10^{-9}) ridge regularization was done for the covariance matrices. The Gaussian modelization was further compared to the simple case where the pixel reflectance distribution of a grassland is modeled by the mean vector value. Then grasslands are classified by SVM with a conventional RBF kernel. Finally, a pixel-wise SVM classification with a RBF kernel was performed and a majority rule inside each grassland was done to extract one class label at the grassland scale. The SVM and the HDKLD were implemented in Python through the Scikit library. In the remaining of the paper, the methods are denoted, KLD-SVM, HDKLD-SVM, $\boldsymbol{\mu}$-SVM and P-SVM, respectively.

4.2. Protocol

All the parameters of each method were optimized using cross-validation. The search ranges were $\sigma \in \{2^{-5}, 2^{-4}, \dots, 2^5\}$ for P-SVM and $\boldsymbol{\mu}$-SVM, $\sigma \in \{2^8, 2^9, \dots, 2^{12}\}$ for KLD-SVM and HDKLD-SVM, $C \in \{1, 10, 100\}$ for all the methods and $t \in \{0.80, 0.85, 0.90, 0.95, 0.99\}$ for HDKLD-SVM.

Given the small size of the reference data, a Leave One Out procedure was chosen. One grassland is iteratively classified given all the other grasslands. The confusion matrix is built during the process. The classification accuracy is assessed with the Kappa coefficient and the statistical significance of the observed differences are computed between each result with the given formulae:

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

4.3. Results

The Kappa coefficients resulting from the classifications are the following: 0.09 for KLD-SVM, 0.57 for HDKLD-SVM, 0.41 for μ-SVM, 0.64 for P-SVM. All the methods are significantly better than the KLD-SVM: the conventional KLD does not perform well in this small sample size context. The proposed model is robust to this configuration and outperforms the conventional KLD. It allows a proper modelization of the grassland at the parcel scale.

However, with this dataset, the obtained classification between the object and pixel-wise approaches are statistically equivalent since the Kappa coefficients between HDKLD-SVM and P-SVM are not significantly different.

As an example, there was only one more well-classified grassland with P-SVM compared to HDKLD-SVM. Thus, at this step, no conclusion can be drawn about the performance of HDKLD against μ-SVM and P-SVM.

The performance of these classification methods can be different with a larger dataset and with more balanced classes.

If accepted, the discussion about the parameters of HD-KLD model will be further detailed in the final version of this paper.

5. REFERENCES

[1] Eurostat, "Agricultural statistics - main results - 2008-09," Tech. Rep., European Commission, 2010.

[2] C. Gardi, M. Tomaselli, V. Parisi, A. Petraglia, and C. Santini, "Soil quality indicators and biodiversity in northern italian permanent grasslands," *European Journal of Soil Biology*, vol. 38, no. 1, pp. 103 – 110, 2002.

[3] C. A. Sullivan, M. Sheehy Skeffington, M. J. Gormally, and J. A. Finn, "The ecological status of grasslands on lowland farmlands in western Ireland and implications for grassland classification and nature value assessment," *Biological Conservation*, vol. 143, no. 6, pp. 1529 – 1539, 2010.

[4] B. P. Werling, T. L. Dickson, R. Isaacs, H. Gaines, C. Gratton, K. L. Gross, H. Liere, C. M. Malmstrom, T. D. Meehan, L. Ruan, B. A. Robertson, G. P. Robertson, T. M. Schmidt, A. C. Schrotenboer, T. K. Teal, J. K. Wilson, and D. A. Landis, "Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes," *Proceedings of the National Academy of Sciences*, vol. 111, no. 4, pp. 1652–1657, 2014.

[5] S. Plantureux, A. Peeters, and D. McCracken, "Biodiversity in intensive grasslands: Effect of management, improvement and challenges," *Agronomy Research*, vol. 3, no. 2, pp. 153–164, 2005.

[6] H. Zechmeister, I. Schmitzberger, B. Steurer, J. Peterseil, and T. Wrbka, "The influence of land-use practices and economics on plant species richness in meadows," *Biological Conservation*, vol. 114, no. 2, pp. 165 – 177, 2003.

[7] C. Corbane, S. Lang, K. Pipkins, S. Alleaume, M. Deshayes, V. E. García Millán, T. Strasser, J. Vanden Borre, S. Toon, and M. Förster, "Remote sensing for mapping natural habitats and their conservation status new opportunities and challenges," *International Journal of Applied Earth Observation and Geoinformation*, vol. 37, pp. 7 – 16, 2015, Special Issue on Earth observation for habitat mapping and biodiversity monitoring.

[8] C. Schuster, T. Schmidt, C. Conrad, B. Kleinschmit, and M. Förster, "Grassland habitat mapping by intra-annual time series analysis comparison of RapidEye and TerraSAR-X satellite data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 34, pp. 25 – 34, 2015.

[9] M. Drusch, U. D. Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sensing of Environment*, vol. 120, pp. 25 – 36, 2012, The Sentinel Missions - New Opportunities for Science.

[10] M. J. Hill, "Vegetation index suites as indicators of vegetation state in grassland and savanna: An analysis with simulated SENTINEL-2 data for a North American transect," *Remote Sensing of Environment*, vol. 137, pp. 94 – 111, 2013.

[11] M. Bock, P. Xofis, J. Mitchley, G. Rossner, and M. Wissen, "Object-oriented methods for habitat mapping at multiple scales case studies from Northern Germany and Wye Downs, UK," *Journal for Nature Conservation*, vol. 13, no. 23, pp. 75 – 89, 2005.

[12] O. Hagolle, M. Huc, D. Villa Pascual, and G. Dedieu, "A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENuS, LANDSAT and SENTINEL-2 images," *Remote Sensing of Environment*, vol. 114, no. 8, pp. 1747–1755, Aug. 2010.

[13] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS.," in *Proceedings of the Third ERTS Symposium*. NASA, Dec. 1973, vol. 1, pp. 309–317.

[14] P. H. C. Eilers, "A perfect smoother," *Analytical Chemistry*, vol. 75, no. 14, pp. 3631–3636, 2003, PMID: 14570219.

[15] S. Kullback, "Letter to the editor: The Kullback-Leibler distance," *The American Statistician*, vol. 41, no. 4, pp. 340–341, 1987.

[16] 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.

[17] K. T. Aboumoustafa and F. P. Ferrie, "A note on metric properties of some divergence measures: The Gaussian case," in *Proc. of the 4th ACML, JMLR W&CP*, 2012, pp. 1–15.

[18] B. Haasdonk and C. Bahlmann, "Learning with distance substitution kernels," in *Pattern Recognition*, C. Rasmussen, H. Bülthoff, B. Schölkopf, and M. Giese, Eds., vol. 3175 of *Lecture Notes in Computer Science*, pp. 220–227. 2004.