Object-based classification of grasslands from high resolution satellite image time series using Gaussian mean map kernels

Maïlys Lopes, Mathieu M. Fauvel, Stéphane Girard, David Sheeren

▶ **To cite this version:**

Maïlys Lopes, Mathieu M. Fauvel, Stéphane Girard, David Sheeren. Object-based classification of grasslands from high resolution satellite image time series using Gaussian mean map kernels. 2017. hal-01424929v2

HAL Id: hal-01424929

https://inria.hal.science/hal-01424929v2

Preprint submitted on 26 Apr 2017 (v2), last revised 13 Jun 2017 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MDPI

Article

Object-based classification of grasslands from high resolution satellite image time series using Gaussian mean map kernels

Mailys Lopes[1]*, Mathieu Fauvel[1], Stéphane Girard[2] and David Sheeren[1]

[1] Dynafor, INRA, INPT, INP-EI Purpan, University of Toulouse, 31326 Castanet Tolosan, France; mailys.lopes@inra.fr, mathieu.fauvel@ensat.fr, david.sheeren@ensat.fr.
[2] Team Mistis, INRIA Rhône-Alpes, LJK, 38334 Montbonnot, France; stephane.girard@inria.fr.
* Correspondence: mailys.lopes@inra.fr; Tel.: +33-534-32-39-27

Academic Editor: name
Version April 26, 2017 submitted to Remote Sens.

Abstract: This paper deals with the classification of grasslands using high resolution satellite image time series. Grasslands considered in this work are semi-natural elements in fragmented landscapes, *i.e.*, they are heterogeneous and small elements. The first contribution of this study is to account for grassland heterogeneity while working at the object scale by modeling its pixels distributions by a Gaussian distribution. To measure the similarity between two grasslands, a new kernel is proposed as a second contribution: the α-Gaussian mean kernel. It allows to weight the influence of the covariance matrix when comparing two Gaussian distributions. This kernel is introduced in Support Vector Machine for the supervised classification of grasslands from south-west France. A dense intra-annual multispectral time series of Formosat-2 satellite is used for the classification of grasslands management practices, while an inter-annual NDVI time series of Formosat-2 is used for permanent and temporary grasslands discrimination. Results are compared to other existing pixel- and object-based approaches in terms of classification accuracy and processing time. The proposed method shows to be a good compromise between processing speed and classification accuracy. It can adapt to the classification constraints and it encompasses several similarity measures known in the literature. It is appropriate for the classification of small and heterogeneous objects such as grasslands.

Keywords: Supervised classification; SVM; Gaussian mean map kernels; kernel methods; object analysis; grasslands.

1. Introduction

Grasslands are semi-natural elements that represent a significant source of biodiversity in farmed landscapes [1–4]. They provide many ecosystem services such as carbon storage, erosion regulation, food production, crop pollination, biological regulation of pests [5], which are linked to their plant and animal composition.

Different factors impact on grassland biodiversity conservation. Among them, the age of a grassland (*i.e*, the time since last ploughing/sowing) is directly related to its plant and animal composition. Old "permanent" grasslands, often called semi-natural grasslands, hold a richer biodiversity than temporary grasslands [2,6–8]. Indeed, they had time to establish and stabilize their vegetation cover, contrarily to temporary grasslands which are part of a crop rotation. Additionally, agricultural management of grasslands (*i.e*, mowing, grazing, fertilizing, reseeding...) influences their structure and composition [9–12]. Management is essential for their biodiversity conservation because its prevents from the woody establishment. Conversely, an intensive use constitutes a threat for this

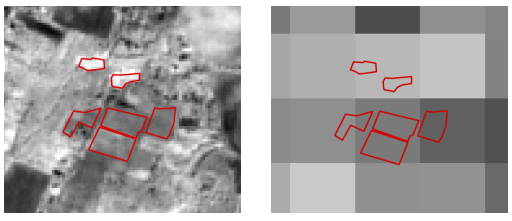
biodiversity [12,13]. Therefore, it is important to know the age of a grassland and to identify the management practices in order to monitor their effect on biodiversity and related services. But these factors are defined at different temporal scales: over the years for the age of a grassland, and during a vegetation season (*i.e.*, a year) for the management practice.

Usually, ecologists and agronomists characterize grasslands at the parcel scale through field surveys. However, these surveys require important human and material resources, the knowledge of the assessor and a sampling strategy, which make them expensive and time-consuming [14]. They are thus limited in spatial extent and in temporal frequency, limiting grassland characterization to a local scale and over a short period of time.

Conversely, remote sensing offers the possibility to provide information on landscapes over large extents, thanks to the broad spatial coverage and regular revisit frequency of satellite sensors [15]. In this context, satellite images have already appeared to be an appropriate tool to monitor vegetation over large areas with a high temporal resolution.

In the remote sensing literature, grasslands have relatively not been studied much compared to other land covers like crops or forest [16]. Most of studies focusing on grasslands have agronomic applications, such as estimating biomass productivity and growth rate [17–19] or derivating biophysical parameters like LAI, fPAR and chlorophyll content [20–24]. Studies having biodiversity conservation schemes such as assessing plant diversity and plant community composition in a grassland are usually based on ground spectral measurements or airborne acquisitions at a very high spatial resolution [25–31]. However, such acquisitions are time-consuming and expensive and thus do not allow continuous monitoring of grasslands over the years.

Using satellite remote sensing images, grasslands have been studied a lot at a regional scale with medium spatial resolution sensors (*i.e.*, MODIS, 250m/pixel [17,18,32]), where the Minimum Mapping Unit (MMU) is at least of hundreds of meters. This scale is suitable for large, extensive, homogeneous and contiguous regions like steppes [33], but not for fragmented landscapes which are usually found in Europe and in France particularly [34,35]. These landscapes are made of a patchwork of different land covers which have a small area [35]. In these types of landscapes, grasslands can be smaller (less than 10,000m^2) than the pixel resolution [36] (see Figure 1 for a graphical example). As a consequence, pixels containing grasslands are usually a mixture of other contributions, which can limit the analysis [37,38]. As examples, Poças *et al.* [39] had to select large contiguous areas of semi-natural grasslands in a mountain region of Portugal to be able to use SPOT-VEGETATION data (1-km resolution). Halabuk *et al.* [40] also had to select only one MODIS pixel per homogeneous sample site in Slovakia to detect cutting in hay meadows. A 30-m pixel resolution is still not sufficient for grassland characterization. Lucas *et al.* [41] and Toivonen and Luoto [42] showed that it was more difficult to classify fragmented and complex elements [43], like semi-natural grasslands, than homogeneous habitats, using Landsat imagery. Price *et al.* [44] classified six grassland management types in Kansas using six Landsat images, but the accuracy of the classification was not satisfying (less than 70%). Therefore, to detect small grasslands in fragmented landscapes, high spatial resolution images are required [36,45,46].



(**a**) (**b**)

Figure 1. Digitalized grasslands (in red) from the dataset used in this study on (**a**) a Sentinel-2 image (10m) and (**b**) a MODIS image (250m).

For high spatial resolution images (about 10m/pixel), few intra-annual images are usually available for a given location [47]. However, Buck *et al.* [48] concluded that three Rapid-Eye images per year were not enough to detect the mowing practices in grasslands. It was confirmed by Franke *et al.* [49] who increased the classification accuracy when using more Rapid-Eye images (from three to five images) to identify grassland use intensity. Some works report results with few images per year, such as Dusseux *et al.* [50], but they worked on LAI. In their study for mapping grassland habitat using intra-annual Rapid-Eye imagery, Schuster *et al.* [51] concluded the more acquisition dates were used, the better the mapping quality.

Given the heterogeneity of grasslands in fragmented landscapes, their phenological cycle and the punctuality of the anthropogenic events (*e.g.*, mowing), dense high spatial resolution intra-annual time series are necessary to identify the grassland management types [36,51–53]. Moreover, to discriminate semi-natural grasslands from temporary grasslands, inter-annual time series are necessary. Until recently, satellite missions offering high revisit frequency (1-16 days) had coarse spatial resolution (*i.e.*, NOAA AVHRR - 1km, MODIS - 250/500m). Conversely, high spatial resolution missions did not provide dense time series and/or were costly (*i.e.*, QuickBird, RapidEye). For these reasons and compared to crops, grasslands differentiation through Earth observations is still considered as a challenge [51]. However, new missions like Sentinel-2 [54], with very high revisit frequency (5 days) and high spatial resolution (10 meters in four spectral channels, 20 meters in six channels) provide new opportunities for grasslands monitoring over the years in fragmented landscapes [53] at no cost, thanks to the ESA free data access policy. For instance, the high spatial resolution is assumed to make possible the identification of grassland-only pixels in the image and several pixels can belong to the same grassland plot. Hence, the analysis can be done at the object scale, not at the pixel scale, which is suitable for landscape ecologists and agronomists who usually study grasslands at the parcel scale [55]. Thus, object-oriented approaches are more likely to characterize grasslands ecologically [56,57]. Yet, a lot of works consider pixel-based approaches without any spatial constraints [17,42,44,48,49,51,58].

At the object scale, grasslands are commonly represented by their mean NDVI [18]. But such representation might be too simple since it does not account for the heterogeneity in a grassland. Sometimes, distributions of pixels as individual observations are still better than the mean value to represent grasslands, as in [53]. Lucas *et al.* [41] used a rule-based method on segmented areas for habitat mapping but it did not work well on complex and heterogeneous land covers. Esch *et al.* [59] also used an object-oriented method on segmented elements then represented by their mean NDVI. These methods based on mean modeling do not capture well grasslands heterogeneity. Other representations can be found in the literature, taking standard deviation and object texture features as variables [60], but they were not applied on time series. In our knowledge, these methods do not use the high spatial and the high temporal resolutions jointly. Moreover, all these studies used vegetation indices as a variable, although it has been shown that classification results are better when using more spectral information [35,61].

A few works have been done on addressing the high spatio-spectro-temporal resolutions in the literature. Yet, this is the type of data the new satellite sensors are now offering. Such time series bring new methodological and statistical constraints given the high dimension of data (*i.e.*, number of pixels and number of spectral and temporal measurements). Dealing with more variables increases the number of parameters to estimate, increasing the computation time and making the computation unstable (*i.e.*, ill-conditioned covariance matrices...) [62]. Thus, the use of high dimensional data requires to adapt the methodological and statistical approaches to process a robust supervised classification. Therefore, classifying grasslands at the parcel-scale, using dense multispectral time series with high spatial resolution is still a challenge and appropriate models are lacking.

In the present study, we introduce a model suitable for the classification of grasslands using satellite image time series (SITS) with a high number of spectro-temporal variables (*e.g.*, Sentinel-2 data). Two temporal scales are considered in this work: (i) an inter-annual time series of three years to discriminate permanent grasslands from temporary grasslands and (ii) an intra-annual time series to

identify the management practices. Note that in this work, the object are not found from segmentation but from existing dataset in a polygon form.

The first contribution of this study is to model a grassland at the object-scale while accounting for the spectral variability within a grassland. We consider that the distribution of the pixel spectral reflectance in a given grassland can be modeled by a Gaussian distribution. The second contribution is to propose a measure of similarity between two Gaussian distributions that is robust to the high dimension of the data. This method is based on the use of covariance through mean maps. The last contribution is the application of the method on permanent and temporary grasslands discrimination and of management practices classification, which are non common applications in remote sensing. Moreover, in our knowledge, mean maps have not yet been used on Gaussian distributions for supervised classification of SITS at the object scale.

In the next section, the materials used for the experimental part of this study are presented. The different types of grassland modeling are presented in section 3. Then, the method to measure the similarity between distributions is introduced in section 4. Following that, we experiment the method and other conventional methods on the classification of a real dataset in section 5. Finally, conclusions and prospects are given in section 6.

2. Materials

2.1. Study site

The study site is located in south-west France, near the city of Toulouse (about 30km), in a semi-rural area (center coordinates: 43°27′36″N 1°8′24″E, Figure 2). This region is characterized by a temperate climate with oceanic and Mediterranean influences. The average annual precipitation is 656mm and the average temperature is 13°C. The north of the site, closer to the urban area of Toulouse, is flat, whereas the south-west of the site is hilly. The eastern part corresponds to the Garonne river floodplain and this location is dominated by crop production. Within this study site, livestock farming is declining in favor of annual crop production. Grasslands are mostly used for forage or silage production. Some grasslands, located in the south-western part of the area, are pastures for cattle or sheep. The extent of the area is included in the satellite image extent (Figure 2) and is about $24 \times 24\text{km}^2$.

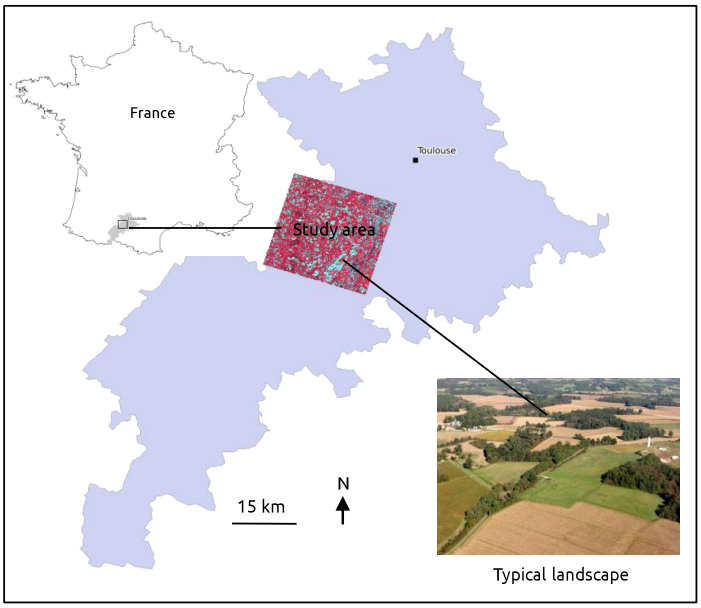


Figure 2. Study site location in south-west France. It is included in the satellite image extent.

2.2. Satellite data

Time series of Formosat-2 were used is this experiment. Formosat-2 has four spectral bands with an 8-meter spatial resolution: B1 "Blue" (0.45 - 0.52μm), B2 "Green" (0.52 - 0.6μm), B3 "Red" (0.63 - 0.69μm), B4 "Near Infra-Red (NIR)" (0.76 - 0.9μm). The extent of an acquisition is 24km $\times$ 24km. The images were all acquired with the same viewing angle. They were orthorectified, radiometrically and atmospherically corrected by the French Spatial Agency (CNES). They were provided by the Center for the Study of the Biosphere from Space (CESBIO) in reflectance with a mask of clouds and shadows issued from the MACCS (Multi-sensor Atmospheric Correction and Cloud Screening) processor [63], in the frame of the Kalideos project.

For the inter-annual analysis, we used all the acquisitions of the consecutive years 2012 (13 observations), 2013 (17 observations) and 2014 (15 observations) (Figure 3). The acquisitions of year 2013 and of year 2014 were used separately for the classification of management practices.

To reconstruct the time series due to missing data (clouds and their shadows), the Whittaker filter [64] was applied pixel-by-pixel on the reflectances in each spectral band for each year independently. The smoother was adapted for unequally spaced intervals and accounted for missing data (see [61] for a detailed description of the method). The smoothing parameter was the same for all the pixels. It was equal to 10^5 for year 2013 and to 10^4 for 2012 and 2014, after an ordinary cross-validation done on a subset of the pixels for each year. An example of smoothing on a grassland pixel is provided in Figure 4. This pixel is hidden by a light cloud during one image acquisition (red cross). Notice that the smoothing is done at the cost of under-estimating the local maxima of the temporal profile.

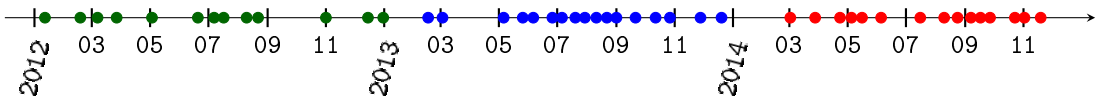


Figure 3. Formosat-2 acquisition dates in 2012 (green dots), 2013 (blue dots) and 2014 (red dots) used in this experiment.

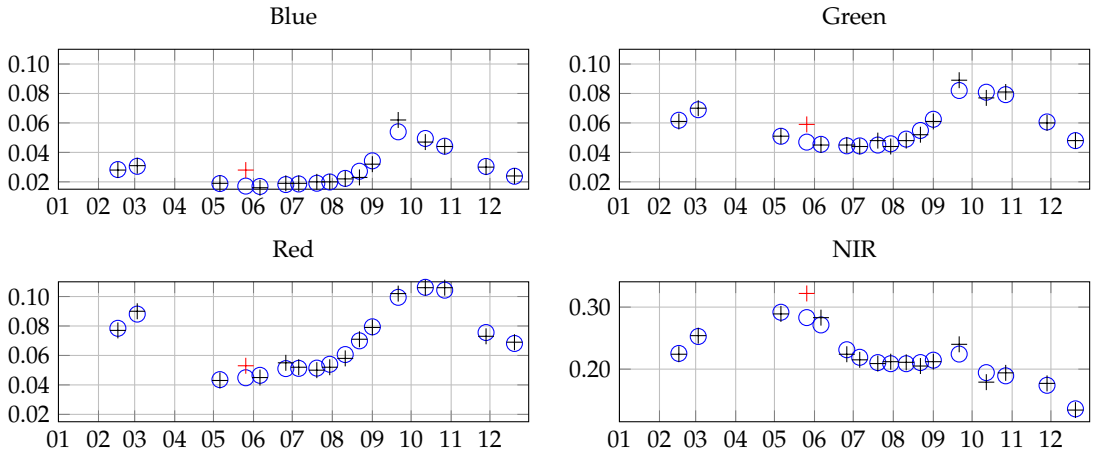


Figure 4. Example of time series reconstruction (blue dots) with Whittaker smoother for a pixel of a grassland in the four spectral bands. The black crosses correspond to the original 2013 Formosat-2 time series and the red ones correspond to missing/noisy data due to the clouds. The x-axis represents the month of year 2013 and the y-axis is the reflectance.

For the intra-annual time series, we used all the spectral information. Therefore, the smoothed time series associated with each of the four spectral bands were concatenated to get a unique time

series per pixel. For the inter-annual time series, as using all the spectral bands would result in a too large number of variables to process, we worked on the NDVI, computed from the red and NIR bands.

2.3. Reference data

2.3.1. Permanent and temporary grasslands

In the regulation, permanent grasslands are at least five years old, whereas temporary grasslands are less than five years old (Commission Regulation EU No 796/2004). However, in this study, we only consider permanent grasslands of at least 14 years old. The French agricultural land use database (Registre Parcellaire Graphique) was used to extract the permanent and temporary grasslands. It registers on an annual basis the cultivated areas declared by the farmers in a GIS. For every plot declared as a grassland in 2014, its age was computed from the previous years declarations. We kept only the permanent grasslands which were at least 14 years old in 2014, and the temporary grasslands which were less than 5 years old in 2014. A negative buffer of 8 meters was then applied to all the polygons to eliminate the edge effects (Figure 5), before rasterizing them. Only the grasslands having an area of at least 1,000m^2 were kept to ensure a minimum number of Formosat-2 pixels per grassland. In the end, there were 59 permanent grasslands (at least 14 years old) and 416 temporary grasslands (Table 1), for an average area of about 26,600m^2.

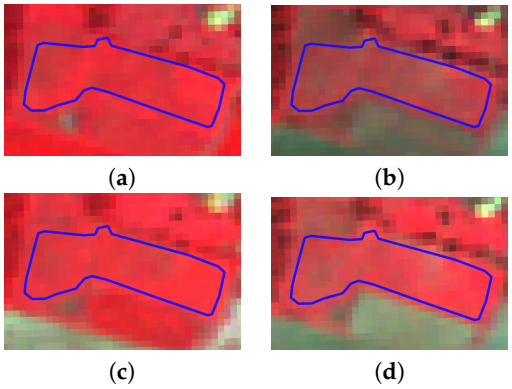


Figure 5. False color Formosat-2 images of the same grassland on two close dates (June and October) in 2013 and 2014 with the same color scale. (**a**) 2013-06-06, (**b**) 2013-10-27, (**c**) 2014-06-05, (**d**) 2014-10-23. The blue line represents the polygon limits of the grassland.

Table 1. Composition of the permanent and temporary grasslands dataset.

Class	Nb of grasslands	Nb of pixels
Permanent	59	31,166
Temporary	416	129,348
Total	475	160,514

2.3.2. Management practices

The information of the agricultural practices performed in the crops are not featured in the land use database. Therefore, this dataset comes exclusively from field data. As mentioned in the introduction, ground data is difficult to obtain in ecology as field work is fastidious. A field survey was conducted in May 2015 to determine the past and current management practices of grasslands by interviewing the farmers or grasslands' owners. In the end, there were 52 grasslands. The practices remained stable for the years 2013 and 2014. Four management types during a vegetation cycle were

identified: one mowing, two mowings, grazing and mixed (mowing then grazing). We eliminated the type "two mowings" of the dataset because of its under-representation (only three grasslands).

The management types were used as classes for the classification (Table 2). The grasslands were digitalized manually after field work. A negative buffer of 8 meters was then applied to eliminate the edge effects, before rasterizing the polygons. The average grasslands surface area is about 10,000m^2. The smallest grassland is 1,632m^2 (which represents 25 Formosat-2 pixels) and the largest is 47,111m^2 (735 pixels) (Figure 6).

Table 2. Grassland management types and composition of the dataset.

Class	Nb of grasslands	Nb of pixels
Mowing	34	6,265
Grazing	10	1,193
Mixed	8	1,170
Total	52	8,628

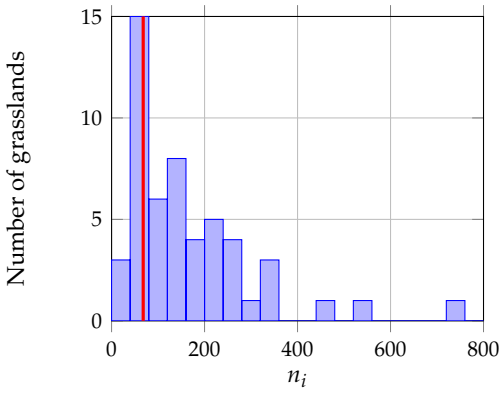


Figure 6. Histogram of grasslands size in number of pixels n_i. The red line corresponds to the number of spectro-variables $d = 68$ in 2013.

3. Grassland modeling

In this work, each grassland g_i is composed of a given number n_i of pixels $\mathbf{x}_{ik} \in \mathbb{R}^d$, where k is the pixel index such as $k \in \{1, ..., n_i\}$, $i \in \{1, \ldots, G\}$, G is the total number of grasslands, $N = \sum_{i=1}^{G} n_i$ is the total number of pixels, $d = n_B n_T$ is the number of spectro-temporal variables, n_B is the number of spectral bands and n_T is the number of temporal acquisitions. In the experimental part, when working on the intra-annual time series of 2013 using the four spectral bands, $d = 4 \times 17 = 68$. In 2014, $d = 4 \times 15 = 60$. When working on the inter-annual times series using NDVI, $d = 1 \times (13 + 17 + 15) = 45$. With each grassland g_i are associated a matrix $\mathbf{X}_i$ of size $(n_i \times d)$ and a response variable $y_i \in \mathbb{R}$ which corresponds to its class label.

In the following, two types of grassland modeling are discussed, at the pixel-scale and at the object-scale. A more informative object-scale modeling is then proposed.

3.1. Pixel-scale

The representation of a grassland at the pixel scale has been used a lot in the remote sensing literature [17,42,44,48,49,51,58]. The grassland can either be represented by all its pixels or by one pixel when the spatial resolution of the pixel is too coarse, see for instance [39,40]. In this representation, a sample is a pixel. Therefore, with each $\mathbf{x}_{ik}$ is associated the response variable y_i of g_i, but $\mathbf{x}_{ik}$ is processed independently of all other $\mathbf{x}_{ik'}$ of g_i. However, this representation usually leads to aberrant

classification results (*e.g.*, salt and pepper effect) [38], which are not expected when working at the grassland scale.

3.2. Object-scale

At the object-scale, the mean vector $\boldsymbol{\mu}_i$ of the pixels belonging to g_i is generally used to represent g_i. It is estimated empirically by:

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{n_i} \sum_{l=1}^{n_i} \mathbf{x}_{il}. \tag{1}$$

In this case, a vector $\hat{\boldsymbol{\mu}}_i \in \mathbb{R}^d$ and a response variable $y_i \in \mathbb{R}$ are associated with each grassland. This representation might be limiting for a heterogeneous object such as grasslands since the spectro-temporal variability is not encoded. To illustrate this bias, Figure 7 shows on the left the set of pixels values in the NIR band for two grasslands (a and b). From this figure, it can be seen that if the mean vector captures the average behavior, higher variability can be captured by including the variance/covariance (middle and right plots). The figure shows that the first and second eigenvectors of the covariance matrix capture well the general trend in the grassland and the main variations due to different phenologic behaviors in the grassland. This information cannot be recovered by considering the variance feature only: covariance must also be included.

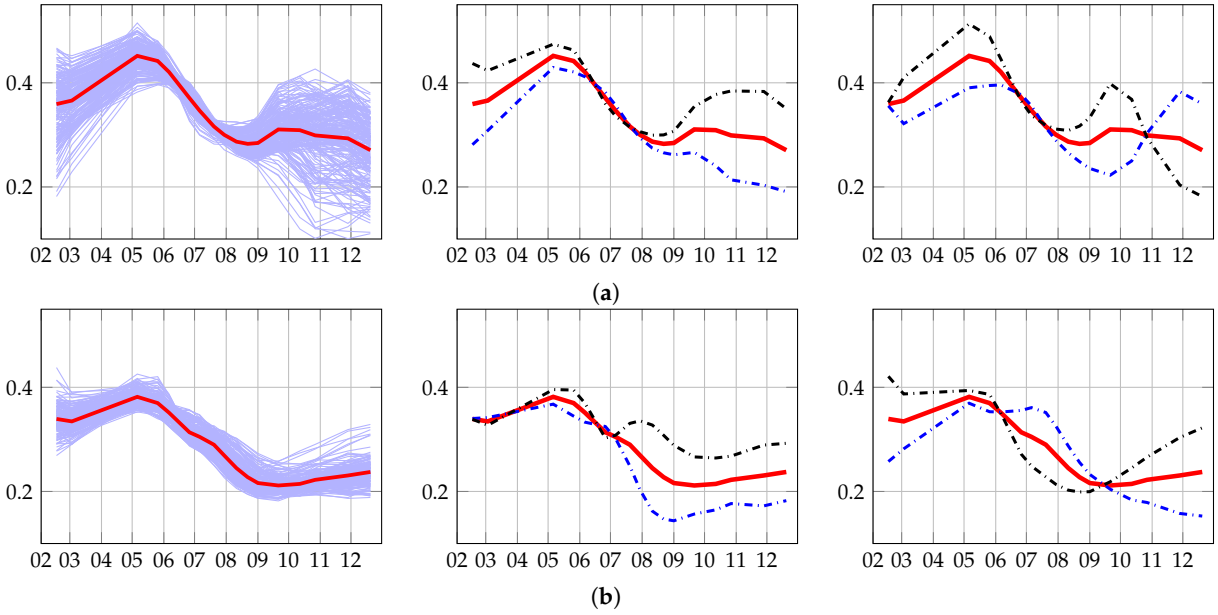


Figure 7. Examples of 2013 time series evolution in the NIR reflectance band of Formosat-2 for a grassland of management practice class (**a**) "mowing" and (**b**) "grazing". The x-axis is the month of the year. The y-axis is the NIR reflectance value. The plot on the left shows the evolution of all the pixels in the grassland and the temporal mean of these pixels in red. The plot in the middle shows the temporal mean in red, the temporal mean $+0.2\times$ the first eigenvector in blue and the temporal mean $-0.2\times$ the first eigenvector in black. The plot on the right shows the temporal mean in red, the temporal mean $+0.2\times$ the second eigenvector in blue and the temporal mean $-0.2\times$ the second eigenvector in black.

In this study, to account for the spectro-temporal variability, we assume that the distribution of pixels $\mathbf{x}_i$ is, conditionally to grassland g_i, a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i$ is the covariance matrix estimated empirically by:

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i - 1} \sum_{l=1}^{n_i} (\mathbf{x}_{il} - \hat{\boldsymbol{\mu}}_i)(\mathbf{x}_{il} - \hat{\boldsymbol{\mu}}_i)^\top. \tag{2}$$

In this case, we associate with each g_i its estimated distribution $\mathcal{N}(\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\Sigma}}_i)$ and a response variable $y_i \in \mathbb{R}$. The Gaussian modeling encodes first and second order information on the grassland by exploiting the variance-covariance information. It is worth noting that if we constraint $\hat{\boldsymbol{\Sigma}}_i = \mathbf{I}_d$, the identity matrix of size d, for $i \in [1, \ldots, G]$, the Gaussian modeling is reduced to the mean vector. In the following, $\mathcal{N}(\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\Sigma}}_i)$ is denoted by $\mathcal{N}_i$.

4. Similarity measure

4.1. Similarity measure between distributions

For classification purposes, a similarity measure between each pair of grasslands is required. With pixel-based or mean modeling approaches, conventional kernel methods such as Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel can be used since the explanatory variable is a vector. However for a Gaussian modeling, *i.e.*, when the explanatory variable is a distribution, specific derivations are required to handle the probability distribution as an explanatory variable.

Many similarity functions generally used to compare two Gaussian distributions (*e.g.*, Kullback-Leibler divergence [65], Jeffries-Matusita distance which is based on Bhattacharyya distance [66]) require the inversion of the covariance matrices and the computation of their determinants. For a conventional multivariate Gaussian model, the number of parameters to estimate for each grassland is $d(d+3)/2$. In the case where d is large, the number of parameters to estimate can be much larger than the number of samples, making ill-posed the inverse problem. This issue is faced in this study because grasslands are small elements of the landscape. They are characterized by a number of spectro-temporal variables which is about of the same order as the number of pixels n_i (see Figure 6). Therefore, most of the estimated covariance matrices are non-invertible and their determinants are null. Hence, conventional similarity measures used for moderate dimensional Gaussian distributions are not suitable for high dimensional Gaussian distributions. In the following, we propose to use mean maps kernels and we introduce a derivation of mean map kernels to weight the influence of the covariance matrix.

4.2. Mean map kernels between distributions

Mean map kernels are similarity measures which operate on distributions [67]. They have been used in remote sensing for semi-supervised pixel-based learning in [68]. In their work, the authors define the similarity between two distributions p_i and p_j as the average of all pairwise kernel evaluations over the available realizations of p_i and p_j (*i.e.*, pixels that belong to grasslands g_i or g_j). It corresponds to the *empirical mean kernel* [68, eq.(8)]:

$$K^e(p_i, p_j) = \frac{1}{n_i n_j} \sum_{l,m=1}^{n_i,n_j} k(\mathbf{x}_{il}, \mathbf{x}_{jm}), \tag{3}$$

where n_i and n_j are the number of pixels associated with p_i and p_j respectively, $\mathbf{x}_{il}$ is the l^{th} realization of p_i, $\mathbf{x}_{jm}$ is the m^{th} realization of p_j and k is a semi-definite positive kernel function.

It is possible to include prior knowledge on the distributions by considering the *generative mean kernel* [67]:

$$K^g(p_i, p_j) \quad = \quad \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} k(\mathbf{x}, \mathbf{x}') \hat{p}_i(\mathbf{x}) \hat{p}_j(\mathbf{x}') d\mathbf{x} d\mathbf{x}'. \tag{4}$$

Note that eq. (3) acts on the realizations of p_i while eq. (4) acts on its estimation. When dealing with a large number of samples, the latter can drastically reduce the computational load with respect to the former.

In our grassland modeling, p_i and p_j are assumed to be Gaussian distributions. In that case, if k is a Gaussian kernel such as $k(\mathbf{x}, \mathbf{x}') = \exp(-\frac{\gamma}{2}\|\mathbf{x} - \mathbf{x}'\|^2)$, eq. (4) reduces to the so-called *Gaussian mean kernel* [69]:

$$K^G(\mathcal{N}_i, \mathcal{N}_j) = \frac{\exp\left\{-0.5(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j)^T \left(\hat{\boldsymbol{\Sigma}}_i + \hat{\boldsymbol{\Sigma}}_j + \gamma^{-1}\mathbf{I}_d\right)^{-1}(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j)\right\}}{|\hat{\boldsymbol{\Sigma}}_i + \hat{\boldsymbol{\Sigma}}_j + \gamma^{-1}\mathbf{I}_d|^{0.5}}, \tag{5}$$

where γ is a positive regularization parameter coming from the Gaussian kernel k and $|\cdot|$ stands for the determinant.

This kernel is not normalized, *i.e.*, $K^G(\mathcal{N}_i, \mathcal{N}_i) \neq 1$, but the normalization can be achieved easily:

$$\begin{aligned} \tilde{K}^G(\mathcal{N}_i, \mathcal{N}_j) &= \frac{K^G(\mathcal{N}_i, \mathcal{N}_j)}{K^G(\mathcal{N}_i, \mathcal{N}_i)^{0.5} K^G(\mathcal{N}_j, \mathcal{N}_j)^{0.5}} \\ &= K^G(\mathcal{N}_i, \mathcal{N}_j)|2\hat{\boldsymbol{\Sigma}}_i + \gamma^{-1}\mathbf{I}_d|^{0.25}|2\hat{\boldsymbol{\Sigma}}_j + \gamma^{-1}\mathbf{I}_d|^{0.25}. \end{aligned} \tag{6}$$

With respect to the Kullback-Leibler divergence (KLD) and the Jeffries-Matusita distance (JMD), the Gaussian mean kernel introduces a *ridge regularization* term $\gamma^{-1}\mathbf{I}_d$ in the computation of the inverse and of the determinant [70]. Thus, the Gaussian mean kernel is more suitable to measure the similarity in a high dimensional space than KLD and JMD. The value of γ tunes the level of regularization. It is tuned during the training process as a conventional kernel parameter.

However, in the case of very small grasslands, two problems remain. The first lies in the ridge regularization: in this case, so low γ values are selected that it becomes too much regularized and it deteriorates the information. The second problem is that the estimation of the covariance matrix has a large variance when the number of samples used for the estimation is lower than the number of variables. Therefore, the covariance matrix becomes a poorly informative feature. In the following, we propose a new kernel function that allows to weight the covariance features with respect to the mean features.

4.3. α-Gaussian Mean Kernel

Depending on the level of heterogeneity and the size of the grassland, the covariance matrix could be more or less important for the classification process. We propose a kernel including an additional positive parameter α which allows to weight the influence of the covariance matrix, the *α-generative mean kernel*:

$$K^\alpha(p_i, p_j) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k(\mathbf{x}, \mathbf{x}')\hat{p}_i(\mathbf{x})^{(\alpha^{-1})}\hat{p}_j(\mathbf{x}')^{(\alpha^{-1})}d\mathbf{x}d\mathbf{x}'. \tag{7}$$

When p_i and p_j are Gaussian distributions, k is a Gaussian kernel and the normalization is applied, the expression gives rise to the *α-Gaussian mean kernel*:

$$\begin{aligned} &\tilde{K}^\alpha(\mathcal{N}_i, \mathcal{N}_j) = \\ &\frac{\exp\left\{-0.5(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j)^T \left(\alpha(\hat{\boldsymbol{\Sigma}}_i + \hat{\boldsymbol{\Sigma}}_j) + \gamma^{-1}\mathbf{I}_d\right)^{-1}(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j)\right\}}{|\alpha(\hat{\boldsymbol{\Sigma}}_i + \hat{\boldsymbol{\Sigma}}_j) + \gamma^{-1}\mathbf{I}_d|^{0.5}} |2\alpha\hat{\boldsymbol{\Sigma}}_i + \gamma^{-1}\mathbf{I}_d|^{0.25}|2\alpha\hat{\boldsymbol{\Sigma}}_j + \gamma^{-1}\mathbf{I}_d|^{0.25}. \end{aligned} \tag{8}$$

The proof is given in the appendix. It is interesting to note that particular values of α and γ lead to known results:

1. $\alpha = 0$: In this case, eq. (8) reduces to the Gaussian kernel between the mean vectors. It becomes therefore equivalent to an object modeling where only the mean is considered.

2. $\alpha = 1$: It corresponds to the Gaussian mean kernel defined in eq. (6).
3. $\alpha \to +\infty$: We get a distance which works only on the covariance matrices. It is therefore equivalent to an object modeling where only the covariance is considered.
4. $\gamma \to +\infty$ and $\alpha = 2$: The α-Gaussian mean kernel simplifies to a RBF kernel built with the Bhattacharyya distance computed between $\mathcal{N}_i$ and $\mathcal{N}_j$.

This proposed kernel thus includes several similarity measures known in the literature. Furthermore, new similarity measures can be defined by choosing different parameters configuration. The α-Gaussian mean kernel (αGMK) is therefore more flexible since it can adapt to the classification constraints:

- Whether the heterogeneity of the object is relevant or not,
- Whether the ratio between the number of pixels and the number of variables is high or low.

5. Experiments on grasslands classification

In this section, the experiments for grassland classification are detailed. We first introduce the seven competitive methods, then the classification protocol is described and we finally present and discuss the results.

5.1. Competitive methods

Several existing pixel-based and object-based classification methods using SVM are presented below. They are compared to assess the effectiveness of the proposed object-based method which relies on the weighted use of the covariance matrix, αGMK, for the classification of grasslands.

5.1.1. Pixel-based and mean modeling

These conventional methods use a RBF kernel.

- **PMV** (Pixel Majority Vote): The pixel-based method was described in section 3.1. It classifies each pixel with no *a priori* information on the object which the pixel belongs to. In order to compare to other object-scaled methods, one class label is extracted per grassland by a majority vote done among the pixels belonging to the same grassland.
- $\boldsymbol{\mu}$ (mean): The distribution of the pixels reflectance of g_i is modeled by its mean vector $\boldsymbol{\mu}_i$ (see section 3.2).

5.1.2. Divergence methods

These methods are based on a distance D between two Gaussian distributions. They are used in a Gaussian kernel such as $K_D(\mathcal{N}_i, \mathcal{N}_j) = \exp(-\frac{D_{ij}^2}{\sigma})$, with $\sigma > 0$:

- **HDKLD** (High Dimensional Kullback-Leibler Divergence): This method uses the Kullback-Leibler divergence for Gaussian distributions with a regularization on covariance matrices such as described in [71].
- **BD** (Bhattacharyya Distance): This method uses the Bhattacharyya distance in the case of Gaussian distributions:

$$B(\mathcal{N}_i, \mathcal{N}_j) = \frac{1}{8}(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j)^\top \left(\frac{\hat{\boldsymbol{\Sigma}}_i + \hat{\boldsymbol{\Sigma}}_j}{2}\right)^{-1} (\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j) + \frac{1}{2} \ln \left(\frac{|\frac{\hat{\boldsymbol{\Sigma}}_i + \hat{\boldsymbol{\Sigma}}_j}{2}|}{|\hat{\boldsymbol{\Sigma}}_i|^{0.5} |\hat{\boldsymbol{\Sigma}}_j|^{0.5}} \right).$$

Small eigenvalues of the covariance matrices are shrinked to the value 10^{-5} to make the computation tractable [72].

5.1.3. Mean map kernel based methods

These methods are based on mean maps kernels presented in section 4:

- **EMK** (Empirical Mean Kernel): This method uses the empirical mean map kernel of eq. (3) and it is pixel-based.
- **GMK** (Gaussian Mean Kernel): This method is based on the normalized Gaussian mean kernel (eq. 6).
- α**GMK** (α-Gaussian Mean Kernel): This method is based on the proposed normalized α-Gaussian mean kernel (eq. 8).

Figure 8 illustrates the relationships between the different methods. The characteristics of each method are synthesized in Table 3.

For memory issues during the SVM process, the number of pixels processed for the permanent and temporary grasslands classification was divided by 10 for the two methods based on pixels (PMV and EMK). Only 1 pixel out of 10 was kept per grassland.

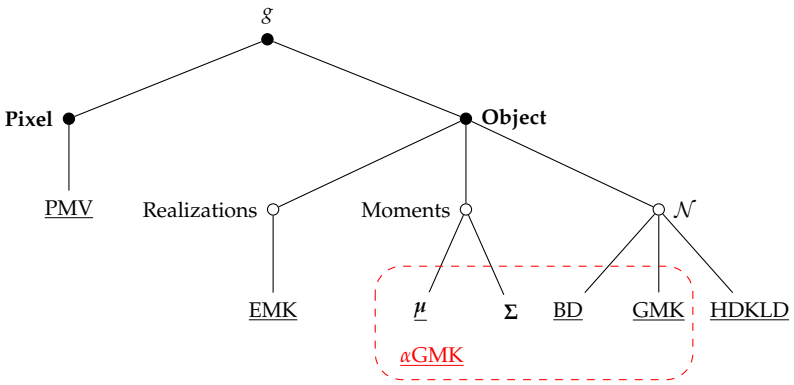


Figure 8. Contribution of the proposed method in grassland analysis for supervised classification. αGMK consists in a general modeling of the grassland at the object-scale and it encompasses several known modelings. The underlined methods are tested in this study.

Table 3. Characteristics of the methods used in this study.

Method	PMV	EMK	μ	HDKLD	BD	GMK	αGMK
Scale	Pixel	Object	Object	Object	Object	Object	Object
Expl. variable	$\mathbf{x}_{ik}$	$\mathbf{x}_{ik}$	$\boldsymbol{\mu}_i$	$\mathcal{N}_i$	$\mathcal{N}_i$	$\mathcal{N}_i$	$\mathcal{N}_i$
Kernel	RBF	RBF	RBF	K_{HDKLD}	K_{B}	$\tilde{K}^G$	$\tilde{K}^\alpha$
Parameters	σ, C	σ, C	σ, C	σ, C	σ, C	γ, C	γ, α, C
Nb of samples	16,250 / 8,628	16,250 / 8,628	475 / 52	475 / 52	475 / 52	475 / 52	475 / 52

5.2. Classification protocol

We compared the efficiency in terms of classification accuracy and processing time of all the presented methods by classifying the two grassland datasets on inter-annual and intra-annual time series (section 2).

For each method, a Monte Carlo procedure was performed on 100 runs. For each run, the dataset was split randomly into training and testing datasets (75% for training and 25% for testing), preserving the initial proportions of each class. The same grasslands were selected for a given Monte Carlo repetition whatever the method.

During each repetition, the optimal parameters were tuned by cross-validation based on the best F1 score. Table 4 contains the parameters grid search for all the methods. Note that a wide grid was

searched for the parameter α of αGMK to further analyze the distribution of selected values. The penalty parameter C of the SVM process was chosen empirically and fixed to $C = 10$, after running several simulations. The classification accuracy for each repetition was assessed by the F1 score computed from the confusion matrix. The overall accuracy (OA) was computed but it is not presented here, because it does not reflect well the accuracy of the classification as unbalanced datasets were used. In order to compare each pair of methods, a Wilcoxon rank-sum test was processed on the pair of distributions of the 100 F1 scores.

The kernels and the SVM were implemented in Python through the Scikit library [73].

Table 4. Parameters tested for each method during cross-validation.

Method	Parameters values	
	Inter-annual analysis	Intra-annual analysis
PMV	$\sigma \in \{2^0, 2^1, \ldots, 2^{10}\}$	$\sigma \in \{2^{-17}, 2^{-16}, \ldots, 2^{-10}\}$
EMK	$\sigma \in \{2^0, 2^1, \ldots, 2^{10}\}$	$\sigma \in \{2^{-18}, 2^{-17}, \ldots, 2^{-10}\}$
μ	$\sigma \in \{2^0, 2^1, \ldots, 2^{10}\}$	$\sigma \in \{2^{-18}, 2^{-17}, \ldots, 2^{-10}\}$
HDKLD	$\sigma \in \{2^{10}, 2^{11}, \ldots, 2^{20}\}$	$\sigma \in \{2^{15}, 2^{16}, \ldots, 2^{25}\}$
BD	$\sigma \in \{2^0, 2^1, \ldots, 2^{10}\}$	$\sigma \in \{2^{10}, 2^{11}, \ldots, 2^{18}\}$
GMK	$\gamma \in \{2^0, 2^1, \ldots, 2^{10}\}$	$\gamma \in \{2^{-17}, 2^{-18}, \ldots, 2^{-10}\}$
αGMK	$\gamma \in \{2^0, 2^1, \ldots, 2^{10}\}$	$\gamma \in \{2^{-18}, 2^{-17}, \ldots, 2^{-13}\}$
	$\alpha \in \{0, 0.1, 0.5, 1, 2, 5, 10, 15, 20, 25, 50\}$	$\alpha \in \{0, 10^{-3}, 10^{-2}, 10^{-1}, 0.3, 0.5, 0.7, 0.9, 1, 2, 5, 10, 15, 20, 25\}$

5.3. Results

5.3.1. Permanent and temporary grasslands: inter-annual time series

Figure 9 sums up the permanent and temporary grasslands classification results for each method over the 100 repetitions as boxplot of F1 scores. The method reaching the best scores is αGMK with a F1 average of 0.71 followed by PMV and GMK with an average of 0.69.

Table 5 contains the Wilcoxon test statistics between each pair of methods. The two best methods, αGMK and PMV are not significantly different. But αGMK is significantly better than all the other methods, whereas PMV is not significantly different than the mean maps methods (EMK and GMK). The worst method is HDKLD with a mean F1 of 0.59.

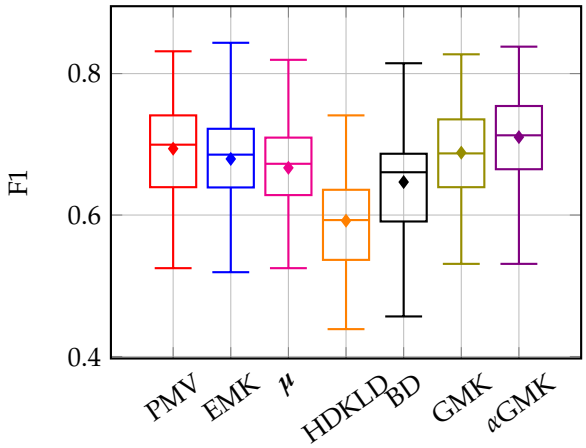


Figure 9. Boxplot of F1 score repartitions for the classification of the permanent and temporary grasslands. The line in the box stands for the median whereas the dot stands for the mean.

In terms of processing load and time, the pixel-based methods are clearly the most demanding. Indeed, processing the 160,514 pixels was not possible with SVM, so we had to reduce the number of samples. These issues are not faced with object-oriented methods. The fastest methods are μ and

Table 5. Absolute value of Wilcoxon test statistics on F1 score for the permanent and temporary grasslands classification. ** indicates the results are significantly different, *i.e.*, p-value < 0.05.

Method	PMV	μ	HDKLD	BD	EMK	GMK	αGMK
PMV	-	3.52**	8.66**	4.83**	1.93	0.98	1.32
μ		-	7.48**	1.76	1.55	2.28**	4.80**
HDKLD			-	5.68**	8.26**	8.65**	9.77**
BD				-	3.23**	3.95**	6.09**
EMK					-	0.94	3.35**
GMK						-	2.42**
αGMK							-

HDKLD, but they did not reach acceptable classification accuracies. The best method in terms of ratio accuracy/processing time is αGMK. It is appropriate for processing a large number of grasslands.

5.3.2. Management practices: intra-annual time series

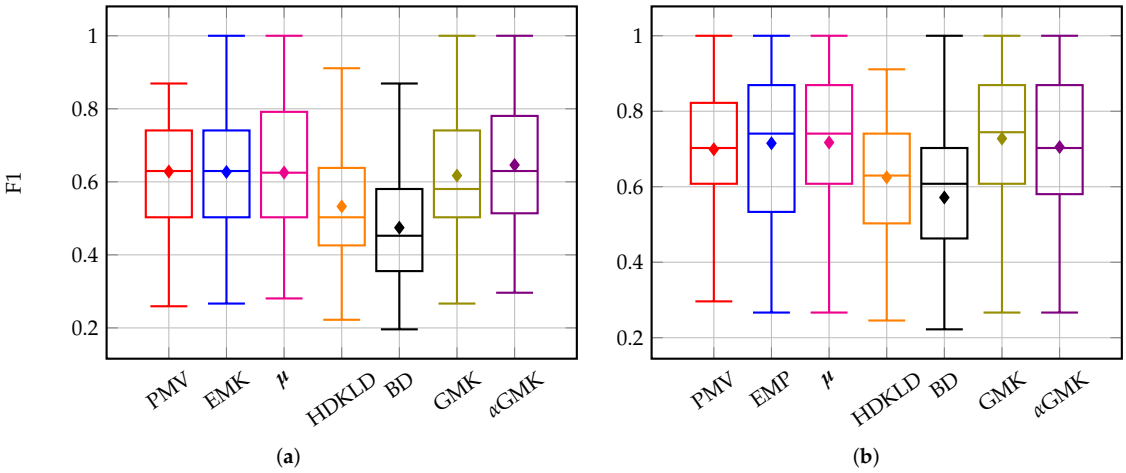


Figure 10. Boxplot of F1 score repartitions for classification of management practices using time series of year (**a**) 2013 and (**b**) 2014. The line in the box stands for the median whereas the dot stands for the mean.

The classifications accuracies for management practices are shown in Figure 10 for year 2013 and for year 2014.

In terms of classification accuracy, methods based on divergences (BD and HDKLD) provided the worst results. Pixel-based methods, mean modeling method and mean generative kernel methods provided similar results in terms of F1 score, except for PMV which was significantly worse than the others for the year 2013. αGMK provided the highest values in 2013 (average F1 of 0.65) but it was not significantly better than the others for this dataset. Indeed, due to the very low number of grasslands composing this dataset, confusion matrices were quite similar whatever the method. It is therefore difficult to compare the methods efficiency in this configuration.

Nevertheless, this dataset makes possible the comparison in terms of processing times, because the same spectral information was used for all the methods. Figure 11 illustrates the training processing time relative to the one of PMV versus the average F1 score for each method. In terms of computational time, the pixel-based methods required the largest processing times. BD was also very long, mainly because of the shrinkage procedure. Mean modeling was the fastest, followed closely by HDKLD. αGMK and GMK were equivalent in terms of computational times. For this configuration with a low number of grasslands, the mean modeling was the most efficient in terms of accuracy/processing time ratio.

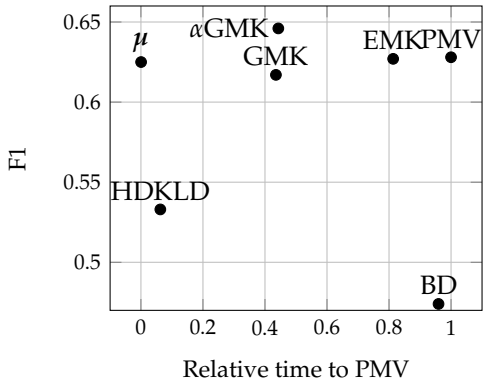


Figure 11. Relative training processing times to PMV and average F1 of each method for intra-annual time series of 2013.

It is worth noting that the times series of 2014 produced higher classification accuracies (maximum F1 average of 0.65 for αGMK) than the time series of 2013 (maximum F1 average of 0.73 for GMK).

5.4. Discussion

The purpose of this work was to develop a model suitable for the classification of grasslands from dense inter- or intra-annual SITS and robust to the dimension of the data. The proposed method based on a weighted use of the covariance, namely αGMK, was compared to several competitive methods.

5.4.1. Methods efficiency

The methods efficiency are discussed for the permanent and temporary grasslands classification, as the other dataset did not allow a proper comparison between the methods because of its small size.

The divergence methods (BD and HDKLD) provided the worst results, showing that they are not robust enough to a high dimensional space.

Although they provided results close to the best results, pixel-based methods (PMV and EMK) are the most demanding in terms of computational time and they do not scale well with the number of pixels. Indeed, they have to process N pixels instead of G grasslands with $G \ll N$. Therefore, we had to reduce the number of pixels used for the classification. Using them on a large area might be difficult, as the permanent and temporary grasslands dataset showed.

Representing grasslands by the estimated distribution of their set of pixels decreases the complexity during the SVM process. Therefore, object-scaled methods offer a lower computational load when compared to empirical mean kernels and pixel-based methods.

The mean generative kernel methods performed significantly better than the mean-only method ($\boldsymbol{\mu}$). Among them, αGMK performed better than GMK. It was also one of the most stable methods.

In this context, including the covariance information helps to discriminate grasslands. However, if the dimensionality is not properly handled, it deteriorates the process (*e.g.*, BD and HDKLD). In this case, it is preferable to use the mean values only. αGMK offers the possibility to weight the influence of the covariance information in front of the mean. As a result, it provided better results than the mean modeling and than GMK, since it encompasses both.

It is furthermore interesting to analyze the optimal values of the weightening parameter α found during the cross-validation and the average of associated F1 scores (Figure 12). The highest F1 scores were reached for high values of $\hat{\alpha}$. The worst F1 scores were obtained with $\hat{\alpha} < 2$ and the value $\hat{\alpha} = 0$ was never selected. It shows the importance of the covariance information in grasslands modeling: the heterogeneity in a grassland must be accounted for and it is not entirely well represented by the mean only.

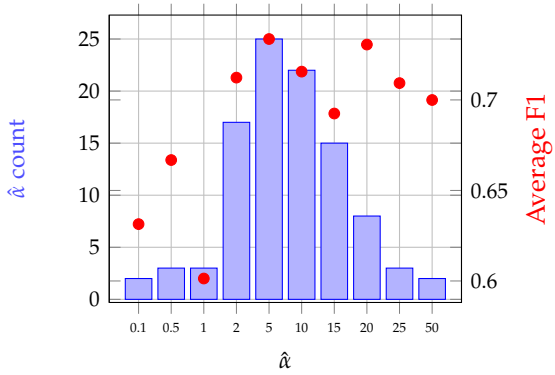


Figure 12. Bar plot of $\hat{\alpha}$ values chosen by cross-validation and the average of associated F1 scores (red dots) for the classification of permanent and temporary grasslands using αGMK. NB: The value $\hat{\alpha} = 0$ was never selected.

5.4.2. Grassland modeling

Following on from the methods discussion, the choice of modeling grasslands pixels distribution by a Gaussian distribution makes sense in this context. It is particularly well appropriate for semi-natural grasslands, which are very heterogeneous, contrary to crops or annual "artificial" grasslands which can be assimilated to crops.

However, modeling grasslands by the mean only produced equivalent results than the methods based on Gaussian modeling for the classification of management practices. Indeed, management practices are supposed to be uniform at the grassland scale. Therefore the mean appears to be sufficient for this application, contrary to the permanent and temporary grasslands discrimination, which requires capturing more variations between the grasslands. The best modeling might be different depending on the application. Moreover, some grasslands are so small that the covariance matrix is too badly estimated.

In the proposed kernel, this modeling was made flexible by regularizing the weight given to the covariance matrix. αGMK benefits from its high level of adaptability in front of the object configuration: no choice has to be made between a Gaussian or a mean modeling since the method encompasses both. It also includes several object-scaled methods known in the literature. However, this is at the cost of one more parameter to tune. Therefore, the classification process takes more time than GMK for instance.

Above all, although it is the first application of generative mean kernels in remote sensing classification, the α-Gaussian mean kernel proved its efficiency and stability in these experiments. It is appropriate for grasslands classification.

5.4.3. Acquisition dates

For the management practice classification, using time series of 2014 produced better results than using 2013. It might be explained by the acquisition dates in the time series. Although 2014 has less images, more clear images were acquired during Spring compared to 2013 which has a lack of acquisitions in April and May (Figure 3). Indeed, many studies showed that the best season to discriminate grasslands is during the growing season [36,49,52,53]. Spring is the period of the vegetation cycle where the management practices begin. Therefore, it is easier to differentiate the practices during this period. It might thus affect the accuracy of the classification of year 2013.

It is not shown in this experiment, but using only one or two years of acquisitions to discriminate permanent from temporary grasslands did not produce sufficient classification accuracies. This is the reason why three years of data were used. "Old" permanent grasslands are supposed to have a more stable phenology over the years than the "young" temporary grasslands which have been recently sown (less than five years). The temporary grasslands phenology is closer to crops in their very first

years. We suppose this makes possible their discrimination with inter-annual SITS. However, the optimal number of years needed to discriminate these types of grasslands could constitute a research topic.

In general, the results could also be enhanced by removing some winter images which can have a negative influence on the entire annual time series [40]. However, the scope of this study was to develop a method which is able to use a given time series, without having to process a date selection.

5.4.4. Grassland typology

On the whole, the classification did not reach high accuracies (F1 maximum average of 0.73 for management practices and of 0.71 for permanent and temporary grasslands classification). This can be explained by the unbalanced dataset with under-representation of grazing and mixed grasslands in the first application and under-representation of permanent grasslands in the second one. The methods should be testing on a more balanced dataset of grasslands classes.

Moreover, as many times emphasized, semi-natural grasslands (which are present in these datasets) are characterized by their high level of heterogeneity. Therefore, there might be a large amount of intra-class variability because of grasslands diversity. The discrimination might be improved by using more distinct classes: intensively used grasslands against extensively used grasslands, artificial (monospecific) grasslands against semi-natural grasslands for instance.

5.4.5. Comparison with existing works

In our knowledge, only the work of Möckel *et al.* [74] relates to the classification of grasslands age using remote sensing data. They reached a Kappa value of 0.77 in classifying three different grassland age-classes. However, they used airborne hyperspectral data from a single date. They recommendation was to use multitemporal data to improve the classification or to use satellite hyperspectral data to monitor grasslands over wider areas. Our study was based on using multispectro-temporal satellite data, but our proposed method would also work with hyperspectral data.

As described in the introduction, few studies have been carried out on the analysis of semi-natural grasslands using high spatio-spectro-temporal resolutions SITS. Usually, methods were pixel-based and they were applied on a few images or on a precise date selection to avoid dealing with the high dimension of data [42,44,49]. Schuster *et al.* [51] successfully classified grassland habitat using 21 Rapid-Eye images on a pixel basis, but there was no mention of the processing times.

At the object-scale using a time series, grasslands were often represented by their mean NDVI, such as in [59], who noticed the difficulty to discriminate grasslands from crops because of mean seasonal NDVI similarities. The closest configuration might be the work of Zillman *et al.* [35], who used an object-based analysis and spectral reflectances combined with seasonal statistics of vegetation indices for mapping grasslands across Europe. The seasonal statistics were particularly relevant in the classification, because they captured well the spectral diversity of the grassland phenology. The use of these metrics could be considered for discriminating grassland management practices which impact on the phenology. The authors also concluded that the object-based analysis improves the classification compared to a pixel-based classification. However, the objects were determined by segmentation.

5.5. Prediction of management practices on the land use database grasslands

To show the efficiency of αGMK, we classified all the grasslands from the French agricultural land use database (RPG) covered by the Formosat-2 time series to predict their management practice in 2014. All the plots declared as grasslands in 2014, *i.e.*, "permanent grassland" and "temporary grassland" regardless their age, were selected. After applying a negative buffer of 8m and rasterizing the polygons, we removed the plots representing less than 10 Formosat-2 pixels. In the end, there were 797 grassland plots covered by the extent of Formosat-2 for a total of 252,472 pixels.

The multispectral SITS of 2014 was used. The SVM was trained on the whole field data (section 2.3.2) using the same grid search as in the experiments. The parameters chosen after

cross-validation based on F1 score were $\hat{\alpha} = 5$ and $\hat{\gamma} = 2^{-15}$. Then, the model was used to predict the management practices of the 797 grasslands of the land use database.

The classification accuracy could not be assessed since the true labels of the grasslands are not known. However, as described in the study site, a spatial distribution of the classes could be expected. Indeed, grazed and mixed grasslands should be found in the south-west of the site, whereas more mown grasslands should be in the north.

An extract of the classification result is shown in Figure 13. It represents the classified grasslands in their raster format. As expected, most of the grazed and mixed grasslands are located in the south-west of the image, whereas the north of the image is mostly composed of mown grasslands. Therefore, αGMK was very likely able to classify with an acceptable accuracy the grasslands management practices without any *a priori* spatial information. However, specific care should be considered, as not all the possible management practices were predicted. For instance, grasslands mown twice or unused grasslands were not in the training dataset, but it does not mean these managements do not exist in the rest of the data. The method deserves to be tested with an exhaustive grassland typology to produce more detailed grasslands maps.

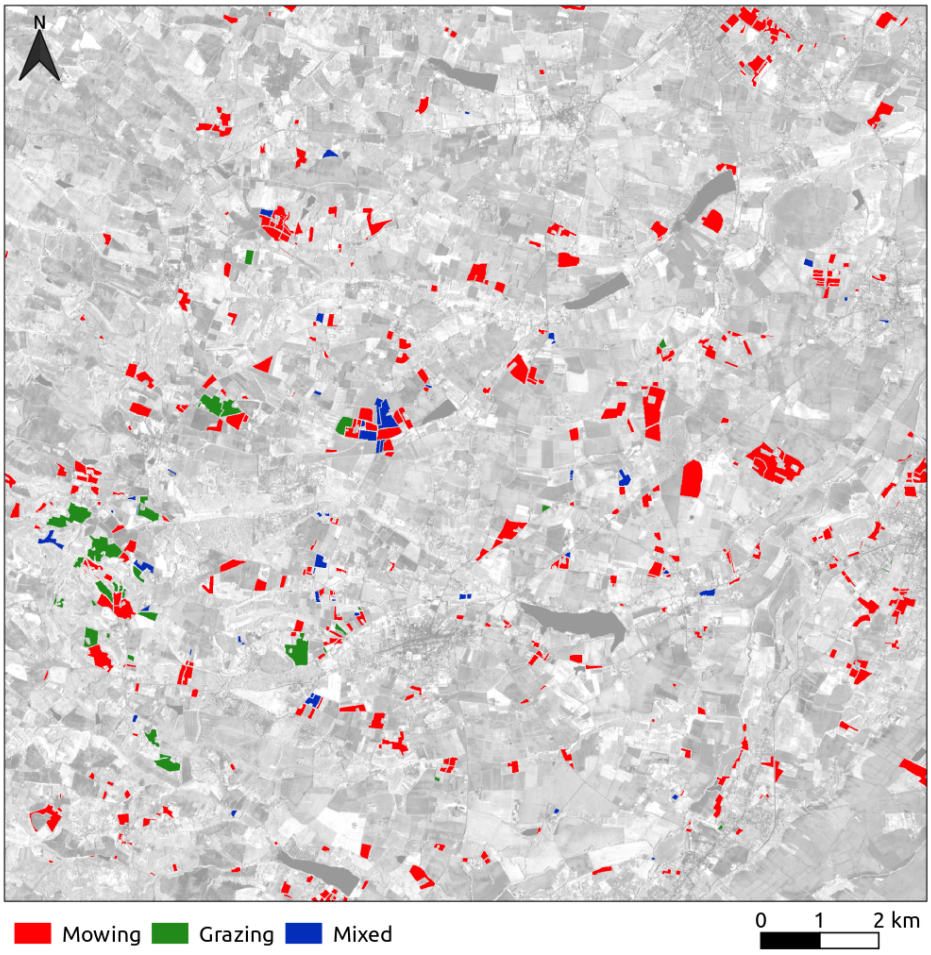


Figure 13. Extract of the management practices classification of the grasslands from the French agricultural land use database (RPG) in 2014. The background is a May, 2014 Formosat-2 image in the NIR channel.

In terms of processing times, the proposed method is able to classify 800 grasslands, representing more than 250,000 pixels, at the object scale from a high spatial resolution SITS within a few seconds on a conventional personal computer.

6. Conclusion

This study aimed at developing a model for the classification of grasslands using satellite image time series with a high number of spectro-temporal variables. A grassland modeling at the object scale was proposed. To deal with grasslands heterogeneity, their pixels distribution was modeled by a Gaussian distribution. Then, to measure the similarity between two grasslands, *i.e*, two Gaussian distributions, a kernel function based on mean maps was introduced, namely the α-Gaussian Mean Kernel. The proposed method was compared to existing pixel-based and object-based classification methods for the supervised classification of grassland using inter- and intra-annual SITS. The Gaussian mean kernels provided the highest classification accuracies, showing that the covariance information must be accounted for. In terms of processing times, the object-based methods were much faster than pixel-based methods.

Several contributions have been made in this work. The first lies in the grasslands pixels distribution modeling at the object-scale. A flexible kernel was proposed to encompass both Gaussian and mean modeling of grasslands, so no choice has to be made between these two modelings. It can therefore be used on homogeneous objects such as artificial grasslands, or on very small objects, as well as on heterogeneous semi-natural grasslands. The second contribution is that this kernel is suitable for high dimensional data in a small ground sample size context. It enables the use of all the multispectral data instead of a single vegetation index or the use a long time series. Also, it can be used on a whole time series without dates selection. Indeed, this new kernel offers very low computational load. It can therefore be applied on a large dataset. With this kernel, we were able to process and to classify more than 250,000 pixels on a conventional personal computer within a few seconds. Even if it is the first application of generative mean kernels in remote sensing classification, the αGMK proved its efficiency and stability in these experiments. It is a good compromise between processing speed and accuracy for the classification of grasslands.

The αGMK deserves to be tested on a larger dataset with more balanced classes. Seasonal statistics could be used to improve the representation of grassland phenology. These ideas will be considered in the future. This method was designed to deal with the dense SITS which will be provided by Sentinel-2 and to efficiently produce maps from this type of data. Other applications of the method are still possible (*e.g.*, small and heterogeneous objects such as peatlands, urban areas...).

Acknowledgments: This work was partially supported by a CJS INRA-INRIA contract and by the grant Défi Mastodons-CNRS. The authors would like to thank CNES and CESBIO for providing the pre-processed Formosat-2 data. Special thanks to Marc Lang for playing a major role in the field work, and to Donatien Dallery for designing the processing chain to compute the grasslands age from the RPG.

Author Contributions: M.L., M.F. and S.G. conceived the model, M.L., M.F. and S.G. conceived and designed the experiments; M.L. performed the experiments; M.L. and M.F. analyzed the data; M.L. and M.F. wrote the paper with feedbacks from S.G. and D.S.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BD	Bhattacharyya Distance
EMK	Empirical Mean Kernel
GIS	Geographic Information System
GMK	Gaussian Mean Kernel
HDKLD	High Dimensional Kullback-Leibler Divergence
JMD	Jeffries-Matusita Distance
KLD	Kullback-Leibler Divergence
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
PMV	Pixel Majority Vote
RBF	Radial Basis Function
SITS	Satellite Image Time Series
SVM	Support Vector Machine
αGMK	α-Gaussian Mean Kernel

Appendix

Proof of eq. (8). First, let us write the Gaussian distribution p_i to the power of α^{-1}:

$$\begin{aligned} p_i(\mathbf{x})^{\alpha^{-1}} &= \frac{1}{(2\pi)^{d/2\alpha}} \times \frac{1}{|\boldsymbol{\Sigma}_i|^{1/2\alpha}} \times \exp\left\{-0.5(\mathbf{x}-\boldsymbol{\mu}_i)^\top (\alpha\boldsymbol{\Sigma}_i)^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\} \\ &= \frac{(2\pi)^{\frac{d}{2}(1-\frac{1}{\alpha})}}{(2\pi)^{d/2}} \times \alpha^{1/2} \times \frac{|\boldsymbol{\Sigma}_i|^{\frac{1}{2}(1-\frac{1}{\alpha})}}{|\alpha\boldsymbol{\Sigma}_i|^{1/2}} \times \exp\left\{-0.5(\mathbf{x}-\boldsymbol{\mu}_i)^\top (\alpha\boldsymbol{\Sigma}_i)^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\} \\ &= \alpha^{1/2}(2\pi)^{\frac{d}{2}(1-\frac{1}{\alpha})}|\boldsymbol{\Sigma}_i|^{\frac{1}{2}(1-\frac{1}{\alpha})} \times p(\mathbf{x}|\boldsymbol{\mu}_i, \alpha\boldsymbol{\Sigma}_i) \\ &= C(\boldsymbol{\Sigma}_i, \alpha)p(\mathbf{x}|\boldsymbol{\mu}_i, \alpha\boldsymbol{\Sigma}_i) \end{aligned} \tag{A1}$$

Then, plugging eq. (A1) in eq. (7), we get:

$$K^\alpha(\mathcal{N}_i, \mathcal{N}_j) = C(\boldsymbol{\Sigma}_i, \alpha)C(\boldsymbol{\Sigma}_j, \alpha)\frac{\exp\left\{-0.5(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j)^T\left(\alpha\hat{\boldsymbol{\Sigma}}_i + \alpha\hat{\boldsymbol{\Sigma}}_j + \gamma^{-1}\mathbf{I}_d\right)^{-1}(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j)\right\}}{|\alpha\hat{\boldsymbol{\Sigma}}_i + \alpha\hat{\boldsymbol{\Sigma}}_j + \gamma^{-1}\mathbf{I}_d|^{0.5}},$$

which is eq. (5) with the covariance matrix of the Gaussian distribution scaled with α. The constants $C(\boldsymbol{\Sigma}_i, \alpha)$ and $C(\boldsymbol{\Sigma}_j, \alpha)$ are removed when normalizing the kernel and we get eq. (8). □

References

1. Eriksson, A.; Eriksson, O.; Berglund, H. Species Abundance Patterns of Plants in Swedish Semi-Natural Pastures. *Ecography* **1995**, *18*, 310–317.
2. Cousins, S.A.; Eriksson, O. The influence of management history and habitat on plant species richness in a rural hemiboreal landscape, Sweden. *Landscape Ecology* **2002**, *17*, 517–529.
3. Gardi, C.; Tomaselli, M.; Parisi, V.; Petraglia, A.; Santini, C. Soil quality indicators and biodiversity in northern Italian permanent grasslands. *European Journal of Soil Biology* **2002**, *38*, 103 – 110.
4. Critchley, C.; Burke, M.; Stevens, D. Conservation of lowland semi-natural grasslands in the UK: a review of botanical monitoring results from agri-environment schemes. *Biological Conservation* **2004**, *115*, 263 – 278.
5. Werling, B.P.; Dickson, T.L.; Isaacs, R.; Gaines, H.; Gratton, C.; Gross, K.L.; Liere, H.; Malmstrom, C.M.; Meehan, T.D.; Ruan, L.; Robertson, B.A.; Robertson, G.P.; Schmidt, T.M.; Schrotenboer, A.C.; Teal, T.K.; Wilson, J.K.; Landis, D.A. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. *Proceedings of the National Academy of Sciences* **2014**, *111*, 1652–1657.
6. Austrheim, G.; Olsson, E.G.A. How does continuity in grassland management after ploughing affect plant community patterns? *Plant Ecology* **1999**, *145*, 59–74.
7. Norderhaug, A.; Ihse, M.; Pedersen, O. Biotope patterns and abundance of meadow plant species in a Norwegian rural landscape. *Landscape Ecology* **2000**, *15*, 201–218.
8. Waldhardt, R.; Otte, A. Indicators of plant species and community diversity in grasslands. *Agriculture, Ecosystems & Environment* **2003**, *98*, 339 – 351. Biotic Indicators for Biodiversity and Sustainable Agriculture.

9. Hansson, M.; Fogelfors, H. Management of a semi-natural grassland; results from a 15-year-old experiment in southern Sweden. *Journal of Vegetation Science* **2000**, *11*, 31–38.
10. Moog, D.; Poschlod, P.; Kahmen, S.; Schreiber, K.F. Comparison of species composition between different grassland management treatments after 25 years. *Applied Vegetation Science* **2002**, *5*, 99–106.
11. Zechmeister, H.; Schmitzberger, I.; Steurer, B.; Peterseil, J.; Wrbka, T. The influence of land-use practices and economics on plant species richness in meadows. *Biological Conservation* **2003**, *114*, 165 – 177.
12. Plantureux, S.; Peeters, A.; McCracken, D. Biodiversity in intensive grasslands: Effect of management, improvement and challenges. *Agronomy Research* **2005**, *3*, 153–164.
13. Muller, S. Appropriate agricultural management practices required to ensure conservation and biodiversity of environmentally sensitive grassland sites designated under Natura 2000. *Agriculture, Ecosystems & Environment* **2002**, *89*, 261 – 266.
14. Rocchini, D.; Boyd, D.S.; Féret, J.B.; Foody, G.M.; He, K.S.; Lausch, A.; Nagendra, H.; Wegmann, M.; Pettorelli, N. Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sensing in Ecology and Conservation* **2016**, *2*, 25–36.
15. Pettorelli, N.; Laurance, W.F.; O'Brien, T.G.; Wegmann, M.; Nagendra, H.; Turner, W. Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology* **2014**, *51*, 839–848.
16. Newton, A.C.; Hill, R.A.; Echeverría, C.; Golicher, D.; Rey Benayas, J.M.; Cayuela, L.; Hinsley, S.A. Remote sensing and the future of landscape ecology. *Progress in Physical Geography* **2009**, *33*, 528–546.
17. Gu, Y.; Wylie, B.K.; Bliss, N.B. Mapping grassland productivity with 250-m eMODIS NDVI and SSURGO database over the Greater Platte River Basin, USA. *Ecological Indicators* **2013**, *24*, 31 – 36.
18. Li, Z.; Huffman, T.; McConkey, B.; Townley-Smith, L. Monitoring and modeling spatial and temporal patterns of grassland dynamics using time-series MODIS NDVI with climate and stocking data. *Remote Sensing of Environment* **2013**, *138*, 232 – 244.
19. Gu, Y.; Wylie, B.K. Developing a 30-m grassland productivity estimation map for central Nebraska using 250-m MODIS and 30-m Landsat-8 observations. *Remote Sensing of Environment* **2015**, *171*, 291 – 298.
20. Friedl, M.A.; Michaelsen, J.; Davis, F.W.; Walker, H.; Schimel, D.S. Estimating grassland biomass and Leaf Area Index using ground and satellite data. *International Journal of Remote Sensing* **1994**, *15*, 1401–1420.
21. Wylie, B.; Meyer, D.; Tieszen, L.; Mannel, S. Satellite mapping of surface biophysical parameters at the biome scale over the North American grasslands: A case study. *Remote Sensing of Environment* **2002**, *79*, 266 – 278. Recent Advances in Remote Sensing of Biophysical Variables.
22. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C.; Corsi, F.; Cho, M. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing* **2008**, *63*, 409 – 426.
23. He, Y.; Guo, X.; Wilmshurst, J.F. Reflectance measures of grassland biophysical structure. *International Journal of Remote Sensing* **2009**, *30*, 2509–2521.
24. Asam, S.; Fabritius, H.; Klein, D.; Conrad, C.; Dech, S. Derivation of leaf area index for grassland within alpine upland using multi-temporal RapidEye data. *International Journal of Remote Sensing* **2013**, *34*, 8628–8652.
25. Schmidtlein, S.; Sassin, J. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing of Environment* **2004**, *92*, 126 – 138.
26. Ishii, J.; Lu, S.; Funakoshi, S.; Shimizu, Y.; Omasa, K.; Washitani, I. Mapping potential habitats of threatened plant species in a moist tall grassland using hyperspectral imagery. *Biodiversity and Conservation* **2009**, *18*, 2521–2535.
27. Fava, F.; Parolo, G.; Colombo, R.; Gusmeroli, F.; Marianna, G.D.; Monteiro, A.; Bocchi, S. Fine-scale assessment of hay meadow productivity and plant diversity in the European Alps using field spectrometric data. *Agriculture, Ecosystems & Environment* **2010**, *137*, 151 – 157. Special section Harvested perennial grasslands: Ecological models for farming's perennial future.
28. Oldeland, J.; Wesuls, D.; Rocchini, D.; Schmidt, M.; Jürgens, N. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecological Indicators* **2010**, *10*, 390 – 396.
29. Feilhauer, H.; Faude, U.; Schmidtlein, S. Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. *Remote Sensing of Environment* **2011**, *115*, 2513 – 2524.

30. Duniway, M.C.; Karl, J.W.; Schrader, S.; Baquera, N.; Herrick, J.E. Rangeland and pasture monitoring: an approach to interpretation of high-resolution imagery focused on observer calibration for repeatability. *Environmental Monitoring and Assessment* **2012**, *184*, 3789–3804.
31. Punalekar, S.; Verhoef, A.; Tatarenko, I.V.; van der Tol, C.; Macdonald, D.M.J.; Marchant, B.; Gerard, F.; White, K.; Gowing, D. Characterization of a Highly Biodiverse Floodplain Meadow Using Hyperspectral Remote Sensing within a Plant Functional Trait Framework. *Remote Sensing* **2016**, *8*, 112.
32. Hilker, T.; Natsagdorj, E.; Waring, R.H.; Lyapustin, A.; Wang, Y. Satellite observed widespread decline in Mongolian grasslands largely due to overgrazing. *Global Change Biology* **2014**, *20*, 418–428.
33. Cao, R.; Chen, J.; Shen, M.; Tang, Y. An improved logistic method for detecting spring vegetation phenology in grasslands from MODIS EVI time-series data. *Agricultural and Forest Meteorology* **2015**, *200*, 9 – 20.
34. Eriksson, O.; Cousins, S.A.; Bruun, H.H. Land-use history and fragmentation of traditionally managed grasslands in Scandinavia. *Journal of Vegetation Science* **2002**, *13*, 743–748.
35. Zillmann, E.; Gonzalez, A.; Herrero, E.J.M.; van Wolvelaer, J.; Esch, T.; Keil, M.; Weichelt, H.; Garzón, A.M. Pan-European Grassland Mapping Using Seasonal Statistics From Multisensor Image Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 3461–3472.
36. Ali, I.; Cawkwell, F.; Dwyer, E.; Barrett, B.; Green, S. Satellite remote sensing of grasslands: from observation to management. *Journal of Plant Ecology* **2016**, *9*, 649–671.
37. Nagendra, H. Using remote sensing to assess biodiversity. *International Journal of Remote Sensing* **2001**, *22*, 2377–2400.
38. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; van der Meer, F.; van der Werff, H.; van Coillie, F.; Tiede, D. Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* **2014**, *87*, 180 – 191.
39. Poças, I.; Cunha, M.; Pereira, L.S. Dynamics of mountain semi-natural grassland meadows inferred from SPOT-VEGETATION and field spectroradiometer data. *International Journal of Remote Sensing* **2012**, *33*, 4334–4355.
40. Halabuk, A.; Mojses, M.; Halabuk, M.; David, S. Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series. *Remote Sensing* **2015**, *7*, 6107.
41. Lucas, R.; Rowlands, A.; Brown, A.; Keyworth, S.; Bunting, P. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* **2007**, *62*, 165 – 185.
42. Toivonen, T.; Luoto, M. Landsat TM images in mapping of semi-natural grasslands and analysing of habitat pattern in an agricultural landscape in south-west Finland. *Fennia - International Journal of Geography* **2003**, *181*, 49 – 67.
43. Nagendra, H.; Lucas, R.; Honrado, J.P.; Jongman, R.H.; Tarantino, C.; Adamo, M.; Mairota, P. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators* **2013**, *33*, 45 – 59. Biodiversity Monitoring.
44. Price, K.P.; Guo, X.; Stiles, J.M. Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing* **2002**, *23*, 5031–5042.
45. Gamon, J.A.; Field, C.B.; Roberts, D.A.; Ustin, S.L.; Valentini, R. Airbone Imaging Spectrometry Functional patterns in an annual grassland during an AVIRIS overflight. *Remote Sensing of Environment* **1993**, *44*, 239 – 253.
46. Corbane, C.; Lang, S.; Pipkins, K.; Alleaume, S.; Deshayes, M.; Millán, V.E.G.; Strasser, T.; Borre, J.V.; Toon, S.; Michael, F. Remote sensing for mapping natural habitats and their conservation status – New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation* **2015**, *37*, 7 – 16. Special Issue on Earth observation for habitat mapping and biodiversity monitoring.
47. Wulder, M.A.; Hall, R.J.; Coops, N.C.; Franklin, S.E. High Spatial Resolution Remotely Sensed Data for Ecosystem Characterization. *BioScience* **2004**, *54*, 511–521.
48. Buck, O.; Millán, V.E.G.; Klink, A.; Pakzad, K. Using information layers for mapping grassland habitat distribution at local to regional scales. *International Journal of Applied Earth Observation and Geoinformation* **2015**, *37*, 83 – 89. Special Issue on Earth observation for habitat mapping and biodiversity monitoring.
49. Franke, J.; Keuck, V.; Siegert, F. Assessment of grassland use intensity by remote sensing to support conservation schemes. *Journal for Nature Conservation* **2012**, *20*, 125 – 134.

50. Dusseux, P.; Vertès, F.; Corpetti, T.; Corgne, S.; Hubert-Moy, L. Agricultural practices in grasslands detected by spatial remote sensing. *Environmental Monitoring and Assessment* **2014**, *186*, 8249–8265.
51. Schuster, C.; Schmidt, T.; Conrad, C.; Kleinschmit, B.; Förster, M. 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* **2015**, *34*, 25 – 34.
52. Psomas, A.; Kneubuhler, M.; Huber, S.; Itten, K.; Zimmermann, N.E. Hyperspectral remote sensing for estimating aboveground biomassand for exploring species richness patterns of grassland habitats. *International Journal of Remote Sensing* **2011**, *32*, 9007–9031.
53. Hill, M.J. 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* **2013**, *137*, 94 – 111.
54. Drusch, M.; Bello, U.D.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; Meygret, A.; Spoto, F.; Sy, O.; Marchese, F.; Bargellini, P. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **2012**, *120*, 25 – 36. The Sentinel Missions - New Opportunities for Science.
55. Laliberte, A.S.; Fredrickson, E.L.; Rango, A. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric engineering & Remote sensing* **2007**, *73*, 197–207.
56. Brenner, J.C.; Christman, Z.; Rogan, J. Segmentation of Landsat Thematic Mapper imagery improves buffelgrass (Pennisetum ciliare) pasture mapping in the Sonoran Desert of Mexico. *Applied Geography* **2012**, *34*, 569 – 575.
57. Stenzel, S.; Fassnacht, F.E.; Mack, B.; Schmidtlein, S. Identification of high nature value grassland with remote sensing and minimal field data. *Ecological Indicators* **2017**, *74*, 28 – 38.
58. Evans, J.; Geerken, R. Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure. *Remote Sensing of Environment* **2006**, *105*, 1 – 8.
59. Esch, T.; Metz, A.; Marconcini, M.; Keil, M. Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland. *International Journal of Applied Earth Observation and Geoinformation* **2014**, *28*, 230 – 237.
60. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* **2012**, *118*, 259 – 272.
61. Sheeren, D.; Fauvel, M.; Josipović, V.; Lopes, M.; Planque, C.; Willm, J.; Dejoux, J.F. Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series. *Remote Sensing* **2016**, *8*, 734.
62. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE* **2013**, *101*, 652–675.
63. Hagolle, O.; Huc, M.; Villa Pascual, D.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENuS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment* **2010**, *114*, 1747–1755.
64. Eilers, P.H.C. A Perfect Smoother. *Analytical Chemistry* **2003**, *75*, 3631–3636. PMID: 14570219.
65. Kullback, S. Letter to the Editor: The Kullback-Leibler distance. *The American Statistician* **1987**, *41*, 340–341.
66. Richards, J.A.; Jia, X. *Remote Sensing Digital Image Analysis: An Introduction*, 3rd ed.; Springer-Verlag New York, Inc.: Secaucus, NJ, USA, 1999.
67. Mehta, N.A.; Gray, A.G. Generative and Latent Mean Map Kernels. *CoRR* **2010**, *abs/1005.0188*.
68. Gomez-Chova, L.; Camps-Valls, G.; Bruzzone, L.; Calpe-Maravilla, J. Mean Map Kernel Methods for Semisupervised Cloud Classification. *IEEE Transactions on Geoscience and Remote Sensing* **2010**, *48*, 207–220.
69. Muandet, K.; Fukumizu, K.; Dinuzzo, F.; Schölkopf, B. Learning from distributions via support measure machines. Advances in Neural Information Processing Systems 25. Curran Associates Inc., 2012, pp. 10–18.
70. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2005.
71. Lopes, M.; Fauvel, M.; Girard, S.; Sheeren, D. High dimensional Kullback-Leibler divergence for grassland management practices classification from high resolution satellite image time series. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, pp. 3342–3345.

758 72. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of*
759 *Multivariate Analysis* **2004**, *88*, 365 – 411.
760 73. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.;
761 Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.
762 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
763 74. Möckel, T.; Dalmayne, J.; Prentice, H.C.; Eklundh, L.; Purschke, O.; Schmidtlein, S.; Hall, K. Classification
764 of Grassland Successional Stages Using Airborne Hyperspectral Imagery. *Remote Sensing* **2014**, *6*, 7732.

765 © 2017 by the authors. Submitted to *Remote Sens.* for possible open access publication
766 under the terms and conditions of the Creative Commons Attribution (CC BY) license
767 (http://creativecommons.org/licenses/by/4.0/).