Mémoire d'Habilitation à Diriger des Recherches

Monitoring Ecosystems with Remote Sensing

Mathieu Fauvel

Université de Toulouse

Le ”savant qui ne sait pas” est une espèce impopulaire et peu crédible, l’honnêteté intellectuelle passant facilement pour de l’incompétence.

Pierre Joliot, La recherche passionnément

Contents

Part I.

Preamble

This manuscript relates my scientific journey since the end of my PhD. As detailed in Section III, after a joint PhD between the Grenoble Institute of Technology and the University of Iceland under the direction of Jocelyn Chanussot and Jon Atli Benediktsson, I spend two years in the MISTIS-INRIA team, with Stéphane Girard. These two experiences gave me a taste for the statistical learning applied to remote sensing, as well as working with computers.

My initial background was "signal and image processing applied to natural signals". During my PhD, I studied kernel methods, and in particular the Support Vectors Machines, for the classification and the feature extraction of hyperspectral images. During the 2000s, such field was called *statistical learning* or *machine learning*, and significant contributions in the remote sensing literature were obtained by casting linear problems into non-linear ones. Learning was done by solving a convex optimization problem which had an attractive property w.r.t the training process of neural networks: the training time was shorter, and an unique solution was found. During that period, apart from doing science, I learned how academic science works: Jocelyn and Jon have driven myself in the middle of big names of the *Geoscience and Remote Sensing Society* (GRSS) and that makes a difference ! They also gave me a research topic that is, twenty years later, still fashionable: AI for remote sensing (but kernel methods are now old-fashioned and neural networks are state of the art!).

My move to MISTIS team made me discovered high dimensional statistical learning, with a more formal approach than what I was doing during my PhD. After two years in the team, my statistical background has increased, thanks to Stéphane, and I started mixing parametric model and kernel methods. I still processed hyperspectral data, not over the *University of Pavia* (!) but over the Mars planet. However, I was not so successful in processing astrophysical data and a little bit tired of competition for academic position. My first child was born and I was thinking to leave academic.

At that time, I was lucky David Sheeren told me about an assistant professor position in Toulouse that should be transformed into associated professor one year later. I came back to Toulouse for one year, then the position was opened and I was ranked first at the competition. I discovered a new field of research, landscape ecology, for which remote sensing and statistical learning could help. My research topics have changed to be driven by my colleagues needs. The idea was not to provide technical services, but to solve statistical problems that arise when we tried to extract meaningful information from remote sensing images. I smoothly abandoned hyperspectral images to satellite image time series because the later were better suited to monitor earth cover in the context of climate change.

During that period, I started questioning myself about the true performances and usefulness of my "statistical" contribution: most of my codes does not scale correctly and it was difficult to feed my colleagues with maps and spatialized information. It was in 2015-2016, during the emergence of the Sentinel-2 image time series. Data were freely available, already pre-processed by the Theia data center and the very first problem at my lab was now "how do we store the data before analysing them?"... A change of paradigm for me, used to process small hyperspectral data: small in term of spatial coverage, and in terms of size (few hundred of gigabytes), w.r.t. Sentinel 2 time series.

A little bit earlier in 2014, Jordi Inglada invited my lab to join the CNES project OSO (Occupation des SOls) he was initiating for the production of the first French land cover map using machine learning and remote sensing. It was about the development of methods and a shared software for the analysis of large scale Sentinel-2 time series. I was sceptical at the beginning: I had distributed some toolbox (old style Matlab !) or library (new style Python !) but I was not

sure that spending so much time in developing software would be more useful than publishing new methods. But then I realized it was at that cost I could better design and evaluate my work. Using remote sensing to monitor small area is a nonsense: fieldworkers are better for that, I should concentrate on large scale and the corresponding statistical and computer tools.

In 2017, the CESBIO lab integrates INRA and researcher positions were opened. I was involved in the preliminary discussions as a member of DYNAFOR lab working in remote sensing. As a "*man-in-the-middle*", I was informed early that such positions were opened. With the invaluable help of my DYNAFOR colleagues (M. Deconchat, A. Ouin, M. Goulard, M. San-Cristobal and D. Sheeren to name few, who respectfully kill me during oral training), I passed the researcher position competition. I started a new scientific adventure at CESBIO-lab, in the same office than Jordi Inglada, who has confronted me to big time series 5 years ago, and was in my PhD committee 10 years earlier, asking me something like "how do you apply the morphological profile on large image?".

I have started mid-July 2018 my new position. Without the teaching load, things were going faster and since then, I have more time for project coordination. Indeed, I rapidly took over the leadership of CNES project PARCELLE from Jordi Inglada, to promote and develop the processing chain iota2 (https://framagit.org/iota2-project/iota2, https://labo.obs-mip.fr/multitemp/iota2-latest-release-deep-learning-at-the-menu/). My objective was to involved different laboratories and public bodies in a co-development of an existing open-source software that is usable both for research and large scale mapping: a difficult task ... In 2019, my colleague Prof. Nicolas Dobigeon grabbed me into the 3IA ANITI project. These two projects perfectly supported my ongoing research project at the CESBIO lab: *When Big Data hit ecosystem monitoring*.

The gitlab account of this document says that I did my first commit the *[2018-08-30 Thu]*. More than 4 years latter, after COVID-19 hard time, I am done with the writing. This HDR manuscript contains a selected resume of my scientific and administrative commitment. It should attest "*the high scientific level of the candidate, the original character of his approach in a field of science, his ability to master a research strategy in a sufficiently broad scientific or technological field and his ability to supervise young researchers*" (https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000028933318). I am not sure it does, maybe you need to attend the defense to check. Yet, its content was deeply influenced by my trajectory. I was lucky to meet colleagues and students who have shaped my way of thinking and still continue to. All the contributions and perspectives discussed in the manuscript owe them a lot.

I think this the time to thanks them and wish you, dear reader, a good reading ! If you find some typo, please fill an issue here https://framagit.org/mfauvel/HDR !

Part II.

Scientific synthesis

1. Introduction

1.1. Importance of ecosystems assessment

Ecosystems are fundamental for humankind. They provide services which are conventionally divided into four categories (according to Millennium Ecosystem Assessment-2005 [Gro13]): i) provisioning services, ii) regulating services, iii) cultural services and iv) supporting services. Natural habitats, agricultural and urban vegetation are known to provide numerous ecosystem services, including provisioning services, cultural services, and also regulating services (e.g., crops host pest predators). For instance, urban vegetation reduces air/noise pollution and helps in draining rainwater while parks and green spaces have recreational and cultural values [BH99], see for instance Figure 1.1. In agricultural landscapes, *crop pollination* is the best known service provided by insects: 75% of the most important crops production is dependent upon animal pollination [KVC+07]. *Biological pest control* is another service of high interest in relation to the recent agri-environmental policies of pesticide use reduction.

Figure 1.1.: Central park, New York, USA. One of the most visited urban park. It hosts many recreational activities and has a consequent variety of wildlife.

In the past decades, anthropogenic pressure on ecosystems have increased drastically, mainly because of *global population growth*; urbanization and agriculture expansion are the main factors associate to this phenomenon. In particular, changes and intensification of agricultural

practices have caused an important loss in terms of ecosystem services. Recently, because climate change has amplified the loss of natural resources with a critical impacts on ecosystem, these issues have raised attention of the citizens. To this end, policies have been adopted worldwide to protect biodiversity and related ecosystems services. For instance, ecosystems monitoring tends to be a mandatory baseline for public policies at local, national and international level [DDC+15, MEW+12, un15].

Therefore, assessing the current state and trend of ecosystems is a critical question in various fields related to land management since they are fundamental for humankind. These variables, such as land use or land cover, reflect the nature, the role and the evolution of ecosystems, but also the *Human-Environmental* interactions. They must be assessed frequently and on large spatial areas because significant changes can be observed both spatially and temporally. However, ecosystems or biodiversity assessment are usually done by intensive ground surveys which require the definition of the objects of interest and of a sampling strategy. As a consequence, they are time-consuming, expensive and limited in terms of area covered and/or temporal periodicity. Furthermore, these measurements require skilled individuals, not as the one in Figure 1.2. For these reasons, there is a lack in ecosystem monitoring methods that can be used over large and global scales [KTS+15].

Figure 1.2.: Field survey is time consuming and does not allow covering large spatial areas with a frequent revisit. Here a slow and awkward data collector.

1.2. Remote sensing for ecosystems monitoring

In the last decades, recent advances in remote sensor technology allow the simultaneous acquisition of several (tens to hundreds) spectral wavelengths for each image pixel with a fine spatial resolution (metric to decametric) and a frequent revisit (from daily to monthly). In addition,

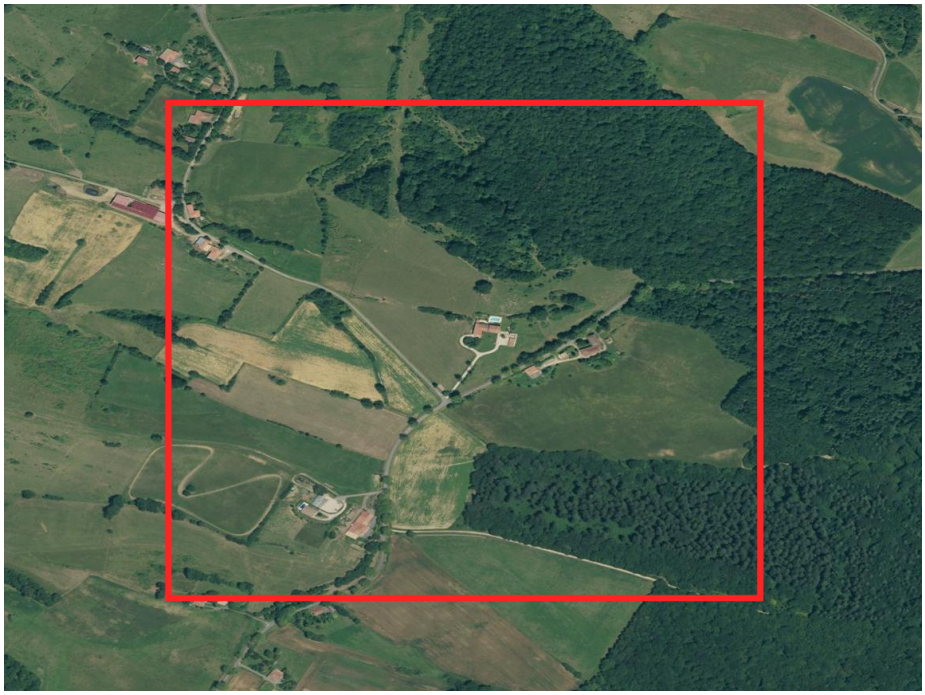


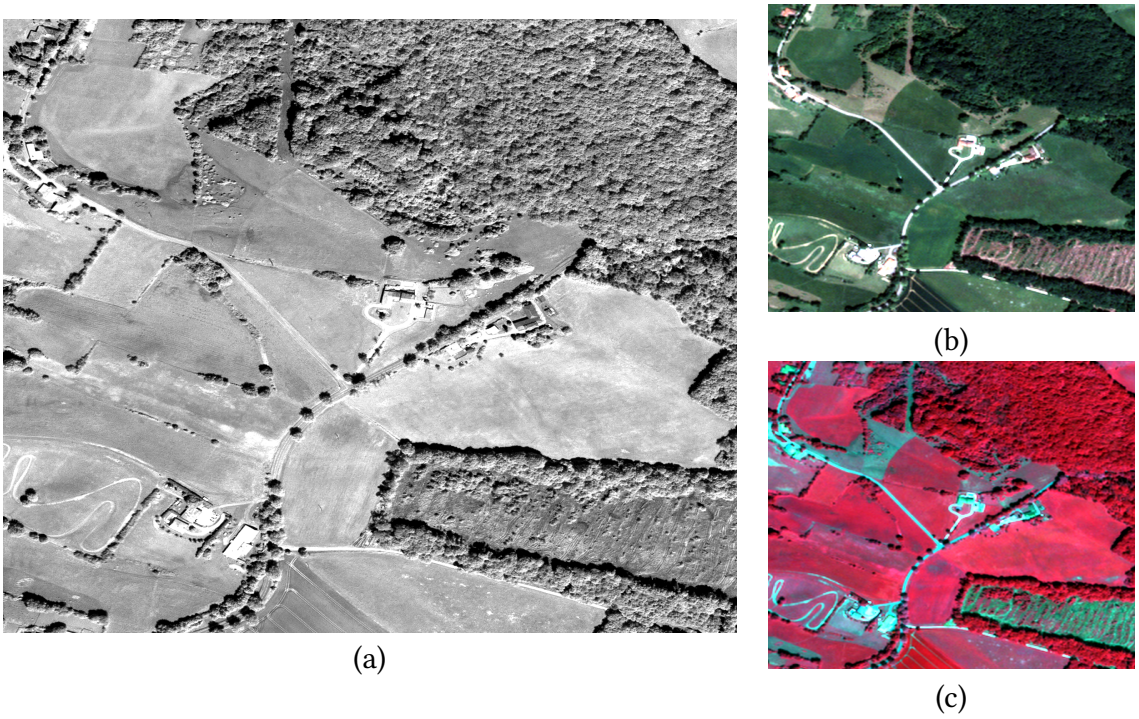
Figure 1.3.: One rural ecosystem, view from BDORTHO ® IRC. The red square represent the extend of images displayed in the following.

several heterogeneous sensors are now available for a given location, (*e.g.*, radar, optical and LiDar) with various spectral, temporal and spatial resolution, see Figures 1.3, 1.4, 1.5 and 1.6 for an example of available products for a given area. For instance, with the successful launch of the satellite Sentinel-2 (part of the European Copernicus program[1]) in June 2015, high spatial resolution satellite images time series are available since the beginning of 2016. The data is freely distributed from the *European Space Agency* (ESA), the French *Theia Land Data Centre* and the CNES (https://peps.cnes.fr/). The metropolitan France territory is now completely mapped every 5 days at a spatial resolution of 10 meters per pixel[2]. For the first time, the mapping of ecosystems is possible at a fine spatial scale with a high temporal resolution and for several years (more than 14 years, with the satellite Sentinel-2A, 2B and their successors 2C and 2D, each of them having a lifetime of 7 years). Furthermore, radar time series (Sentinel-1) as well as very high spatial resolution images (spatial resolution of 0.50 cm per pixel) are also freely provided by these public national and international infrastructures.

The most used approach to infer ecosystem services from remote sensing images is called *proxy-based*, for which ecosystem services are estimated based on some extracted information from the images [EAA+10]. The figure 1.7 presents the different types of remote sensing variables used to map ecosystem services from earth observation data. Up to now, the main source of information extracted from remote sensing images is *land cover*, *i.e.*, a hard classification of landscape into distinct classes. Land cover can be extracted automatically by some statistical

[1] https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-2.

[2] Sentinel-2 has 4 spectral bands with spatial resolution of 10m per pixel, 6 spectral bands with spatial resolution of 20m per pixels and 3 spectral bands with spatial resolution of 60m per pixels. The constellation is composed of two satellites.



(a) (b) (c)

Figure 1.4.: Pleiades view of 1.3, acquired in May, 2018: (a) Panchromatic mode, (b) and (c) multispectral mode with a true and false color composition, respectively.

algorithms or by using available land cover data already manually extracted from remote sensing images. For the former case, manual delineation or advanced classification algorithms can be used [AGSD12, BKL+12, IVA+17]. For the later case, the use of Corinne Land Cover[3] for European Union landscape is usual, see for instance [ZMP13]. More complex framework can be defined from the land cover, such as the Shannon index which is used to estimate the species abundance [OWR+10].

Other remotely sensed variables can be used together with land cover layer, or in a standalone way. Among the variety of existing spectral indices, the NDVI, and with a lesser extend the LAI are the most used, see for instance [KBGK09, CR97, PVM+05, FFA+08]. They are usually used to assess the *Fractional Vegetation Cover* of landscapes. Currently, NDVI is easier to compute than LAI, and therefore much used in the scientific community. It is particularly true in a multitemporal setting. Marginally, remotely sensed variables such as topography or surface temperature are used in some studies, see for instance [GRBBS08] for the use of digital elevation models and surface temperature data for assessing ecosystem services in an Alpine region. But very few works combine several Earth observation data source, even if it has been identified as great opportunity to combine abiotic and structural information [JBE+16].

Furthermore, a past limitation to the use of such images for the analysis of ecosystems is the duality between the spatial, spectral and the temporal resolution. Images with high spatial resolution (few meters per pixel) usually came with a poor temporal resolution (few dates per year acquired over the same geographical area), while images with high temporal resolution (one image every day) had a low spatial resolution (30 to 250 meters per pixel). Hyperspectral

[3] http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3.

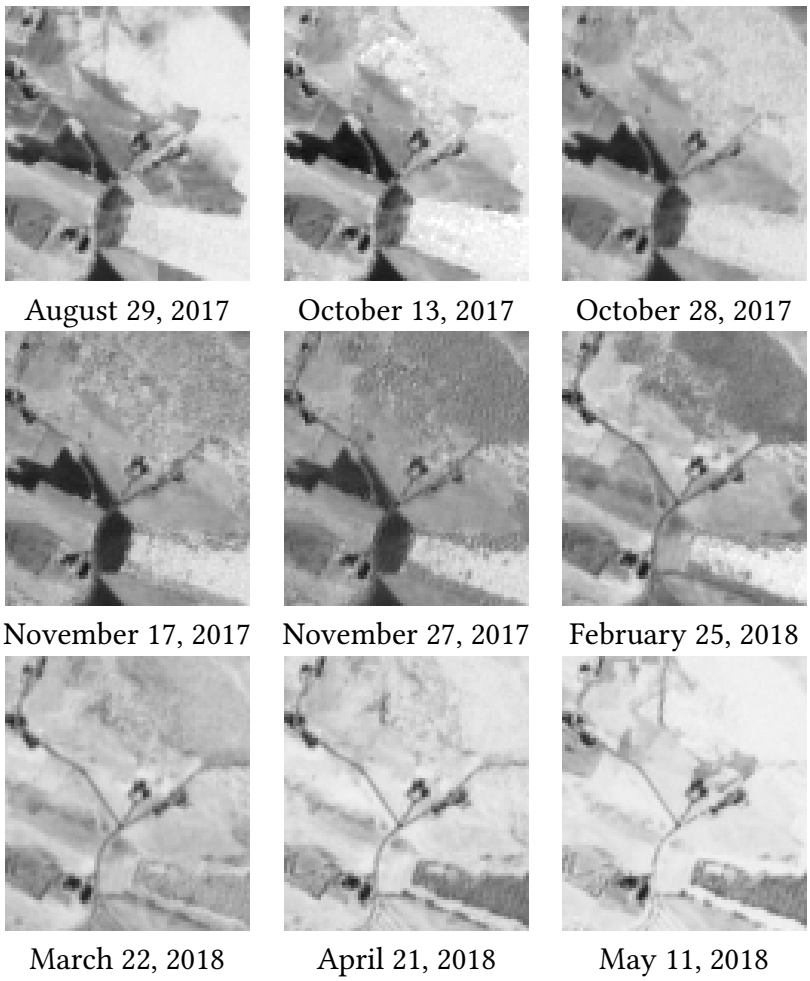


Figure 1.5.: Sentinel-2 view of 1.3.

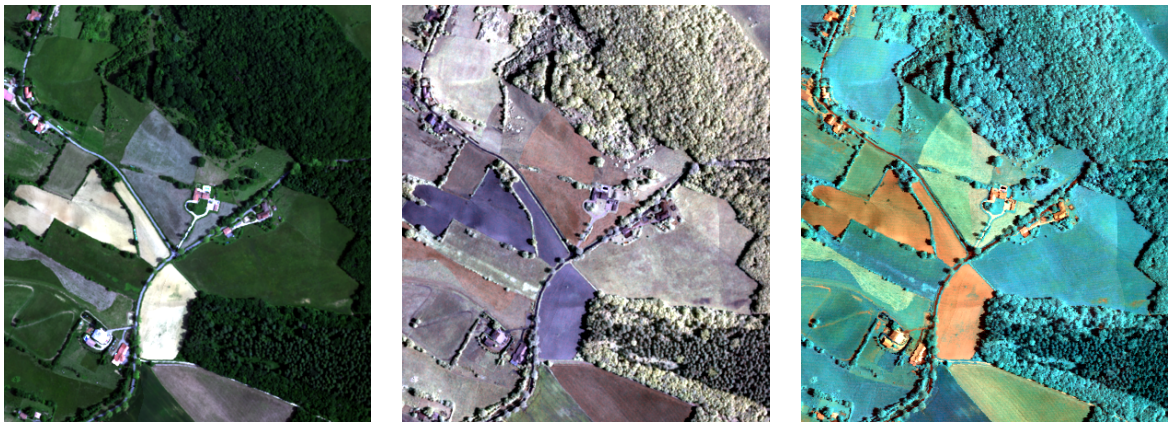


Figure 1.6.: Hyspex (hyperspectral airborne sensor) view of 1.3, acquired in June, 2016.

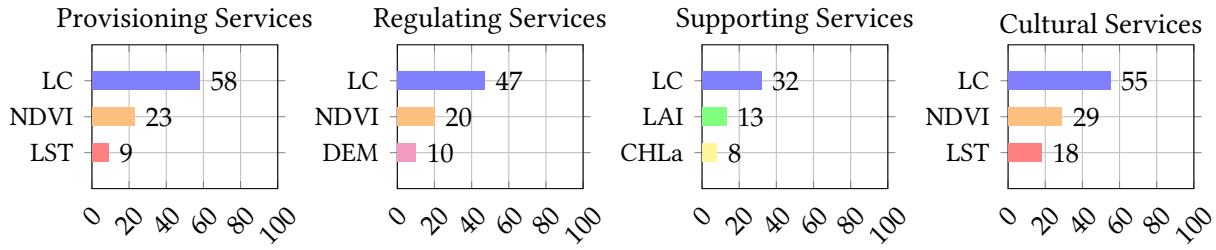


Figure 1.7.: Ecosystem services assessment with remote sensing data from [dABAD15]. LC: Land Cover maps, NDVI: Normalized Difference Vegetation Index, LST: Land Surface Temperatures, DEM: Digital Elevation Map, LAI: Leaf Area Idex, CHLa: Chlorophyll-a concentration. The numbers represent the percentage of remotely sensed variables used in the selected literature associated to each services in [dABAD15].

images from space usually had a low spatial resolution and a high revisit cycle. Hence, it was not possible to analyze fine temporal evolution of ecosystems at a local spatial scale with a fine spatial/geometrical precision. As a consequence, actual produced maps are either outdated or at a very coarse spatial resolution [GMK15].

The Landsat open archives[4] and Copernicus[5] programs as well as the incoming hyperspectral missions (e.g., http://www.enmap.org/ or https://hyspiri.jpl.nasa.gov/) open enhanced possibility to monitor *Earth surface* and ecosystems from remote sensing images. They provide high quality, easy to use and mostly free huge volume of data. As a result, remote sensing has become a major tools to monitor ecosystems from local to global scale [PWH+22]. In particular, *satellite* remote sensing is actually receiving a lot of attention because of its low cost (see for instance the free and open access Sentinel data policy of the EU: https://scihub.copernicus.eu/). For instance, the following question "*How can remote sensing-derived products be used to value and monitor changes in ecosystem services?*" was proposed in a list of ten identified conservation questions that can be solve using remote sensing [RBE+15], which goes far from the simple land cover maps. With the increased spatial, spectral and temporal resolution, landscapes and ecosystems could be analyzed from local to global scale. It will enable the identification of ecosystems configuration and composition.

1.3. New opportunities and challenges

The increased complexity of the data makes the conventional methods defined for moderate spatial and/or temporal resolution not adapted, and therefore not suitable for extracting all the relevant information of landscapes [HFS15]. For hyperspectral images or satellite image time series, several hundreds of spectral/temporal acquisitions of the same scene are typically available, while for multispectral images up to ten bands are usually provided. With increasing dimensionality of the images in the spectral and/or temporal domain, theoretical and practical problems arise. The idea of the *dimension* is intuitive, driven by experiments in one-, two-

[4] https://landsat.usgs.gov/opening-landsat-archiveproduct-specifications and https://landsat.usgs.gov/landsat-data-access
[5] http://copernicus.eu/

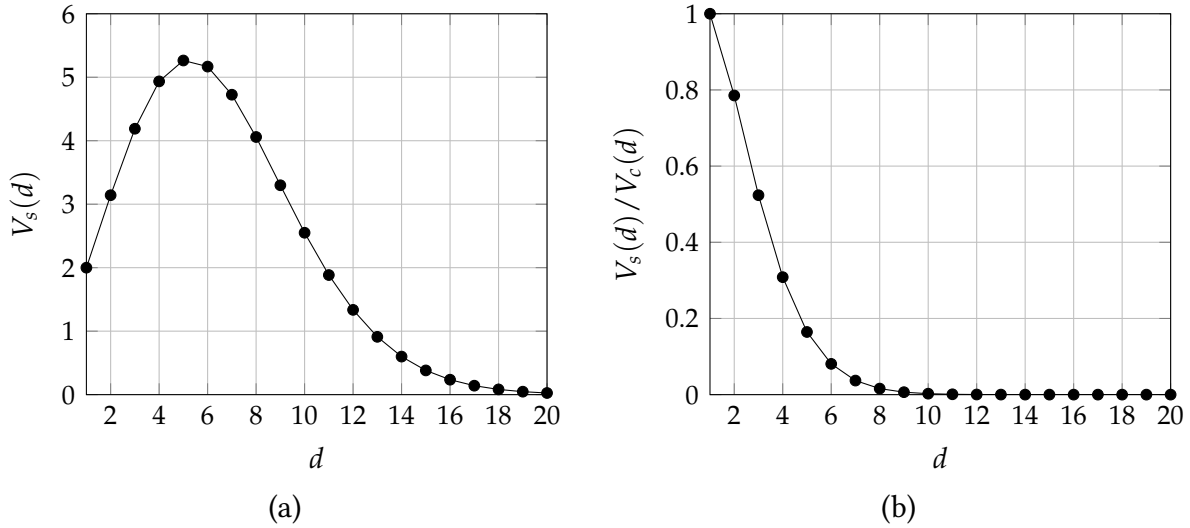


Figure 1.8.: (a): Volume of an unit hypersphere in function of the dimension d: it shows that the notion of neighborhood in high dimensional space is badly defined since local neighborhood is usually empty. (b): Ratio between the volume of an unit hypersphere and an unit hypercube, it shows that most of the space tends to concentrate in the corner of an hypercube, in opposition to 2 or 3 dimensional space.

or three-dimensional spaces, and geometric concepts that are self-evident in these spaces do not necessarily apply in higher-dimensional spaces [Lan03, Ken61], see for instance Figure 1.8. Another example, normally-distributed data in high dimensional spaces have a tendency to concentrate in the tails, which seems to be contradictory with its bell-shaped density function [JL98], see Figure 1.9 for an illustration. In addition, the rate of convergence of the statistical estimation decreases when the dimension grows while conjointly the number of parameters to estimate increases, making the estimation of the model parameters very difficult [Don00]. Consequently, with a limited training set, beyond a certain limit, the analysis accuracy actually decreases as the number of features increases [Hug68]. For the purpose of statistical learning, these problems are related to the *curse of dimensionality*[6].

Furthermore, as the spatial resolution increases, the use of contextual/spatial information is needed to fully exploit the information contains in the image, see Figure 1.10 for an illustrative scenarios. For instance, using spatial information is known to reduce the labeling uncertainty that exits when only pixel-wise information is used, e.g., to overcome the salt and pepper appearance of the classification map. Indeed, complementary and relevant information can be extracted from the spatial domain w.r.t spectral domain: for a given pixel it is possible to extract the size and the shape of the structure to which it belongs. This information will not be the same if the pixel belongs to a roof or to a green area. This is also a way to discriminate between various structures made of the same materials. If spectral information alone is used, the roofs of a private house and of a larger building will be detected as the same type of structure. But using additional spatial information – the size of the roof, for instance – it is possible to

[6]Results from [Hug68] should be mitigated nowadays with recent advances in stochastic optimization and over-parametrization of deep neural networks, see for instance https://francisbach.com/rethinking-sgd-noise/.

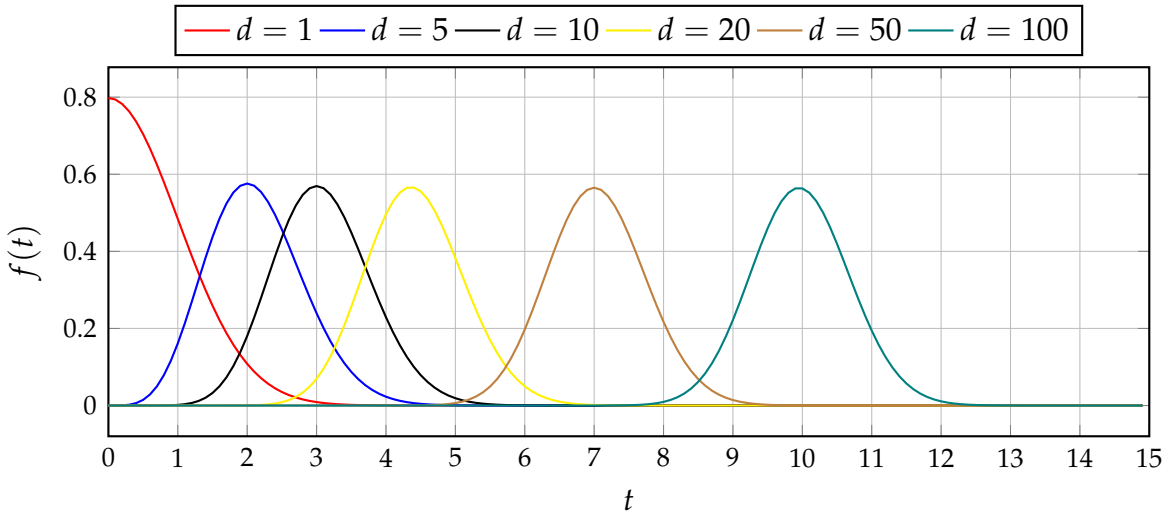


Figure 1.9.: The likelihood of $t = \|\mathbf{x}\|$ a random variable $\mathbf{x} \in \mathbf{R}^d$ such as $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is given by $f(t) = \frac{t^{d-1} \exp(-t^2/2)}{2^{(d/2)-1}\Gamma(d/2)}$, which is maximum for $t^* = \sqrt{d-1}$ [JL98]. The figure shows several functions f in function of the dimension d, where it can be seen that the norm of the variable deviates from zero when d increases.

classify these into two separate classes [CBF06].

In addition to the high dimensionality and the contextual information, a last important issues is related to the temporal dimension of the data. Hypertemporal data are usually noisy due to the occurrence of clouds and their associated shadows, see Figure 1.11. Furthermore, in the case of Sentinel-1/2 satellite time series, the sample times are different from a given area to another, leading to irregularly sample data: each pixel to be processed may have a different number acquisitions, with different dates. Hence, conventional algorithms that assume a fix and common number of feature could not be applied directly. Usually, temporal re-sampling is performed [IAT+15a], but this step is not linked to the final task (classification, inversion ...) and thus might not be optimal in terms of errors. For instance, Figure 1.12 shows the reconstruction of the NIR bands for a whole year using the Whittaker filter [Eil03]. This filter requires to set a regularization parameter that controls the trade-off between the reconstruction error and the smoothness of the reconstructed time series. It is usually found using ordinary cross-validation (OCV), see the Figure 1.13. Ideally, this parameter should be selected so as to optimize the final task (classification, prediction ...) but it is not done usually because of computational load.

To sum-up, ubiquity of satellite remote sensing data have been reached in the last decade, both in terms of supply or demand. It represents a fantastic opportunity for the understanding of our ecosystems. However, current methods for extracting meaningful information are not yet ready to fully exploit this huge amount of data.

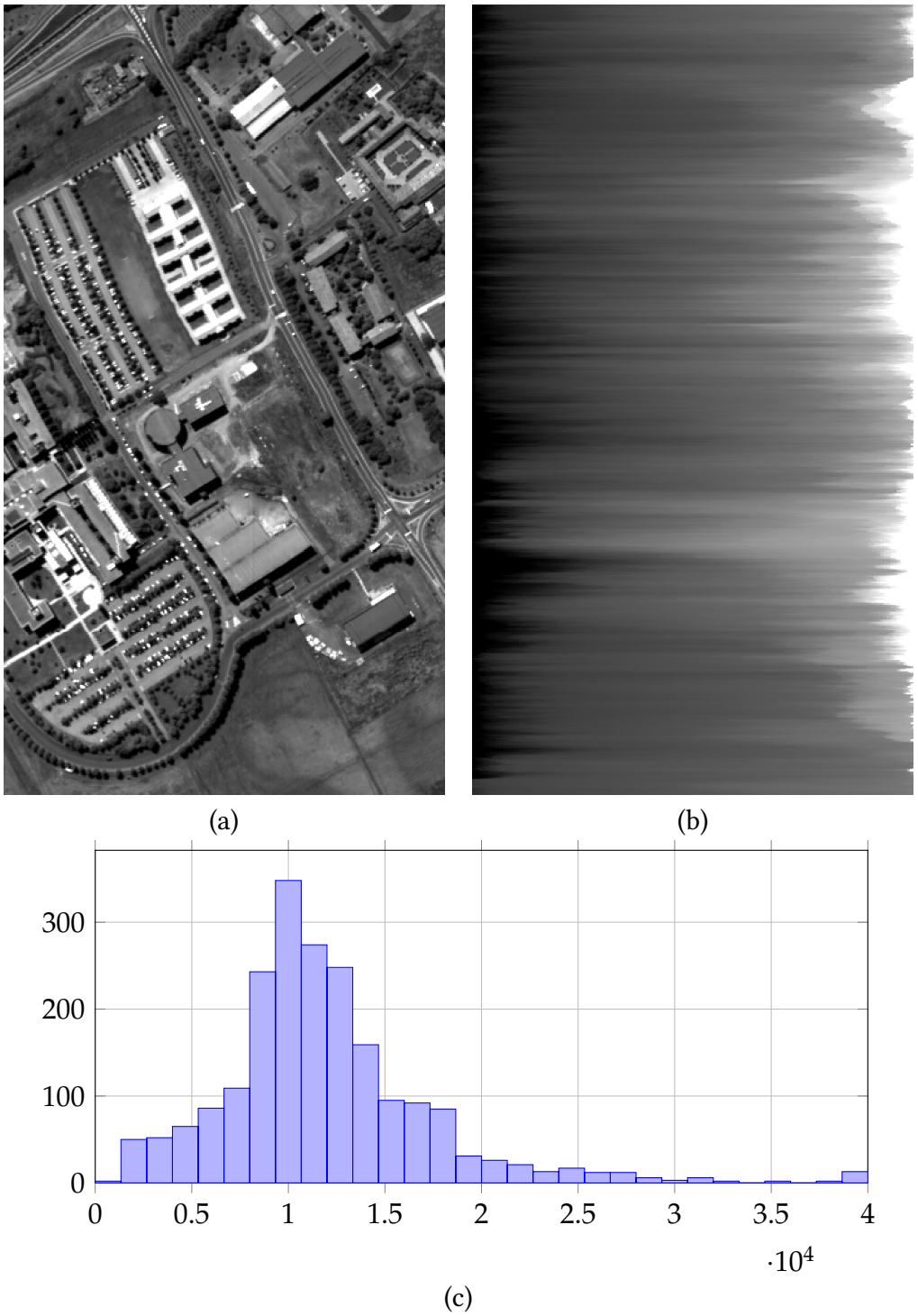


Figure 1.10.: (a) University of Pavia, (b) same image than (a), but pixels position has been modified (the values of each row have been sorted in increasing order). (c) histogram of (a) and (b). They both have the same radiometric information, while the spatial structure in (a) in meaningful for the analysis.

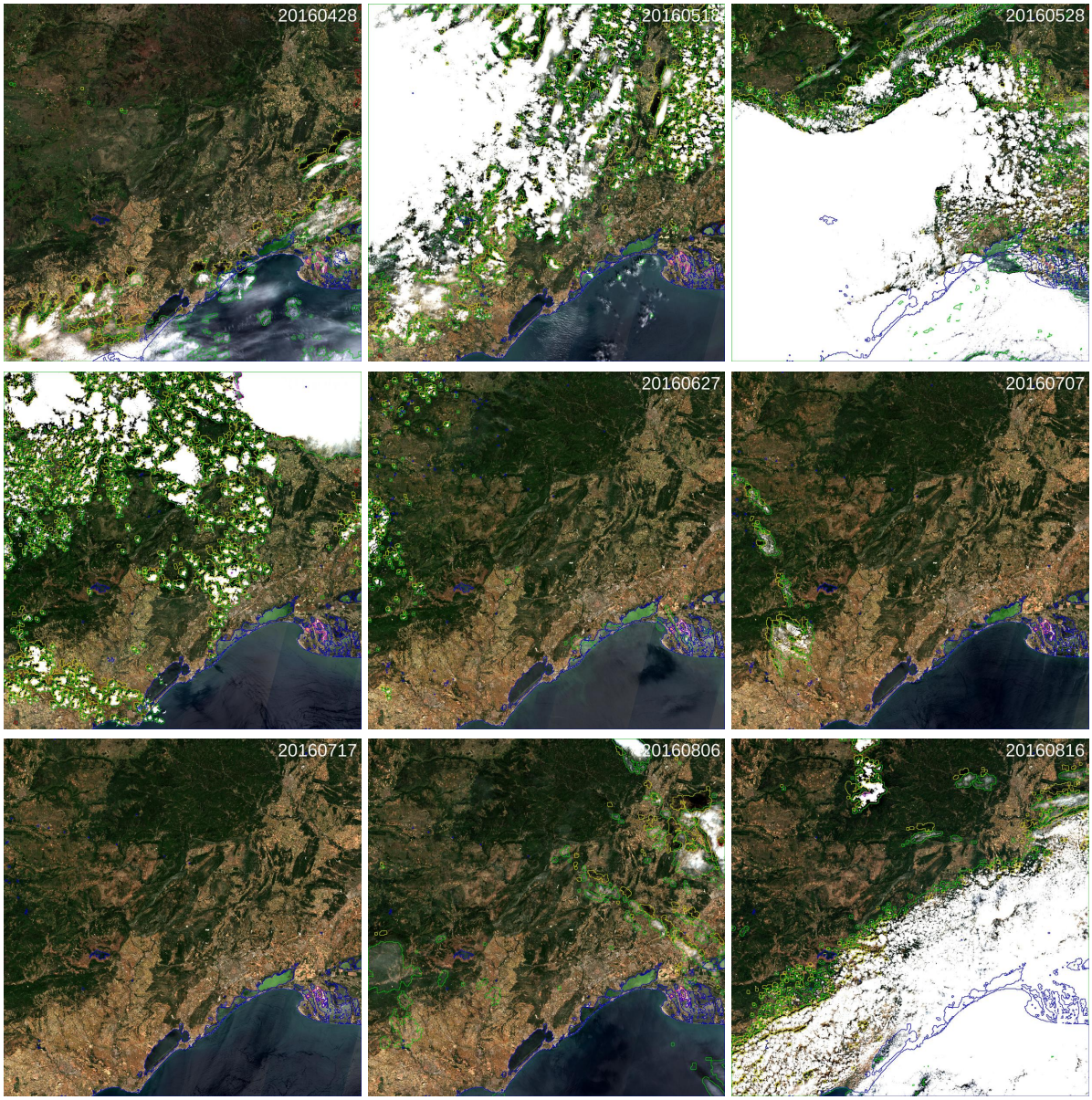


Figure 1.11.: Example of clouds and shadows in satellite image time series Sentinel-2 for the year 2016 [HHDD10]. In blue are the water detection, green and yellow are the clouds and their associated shadows, respectively.

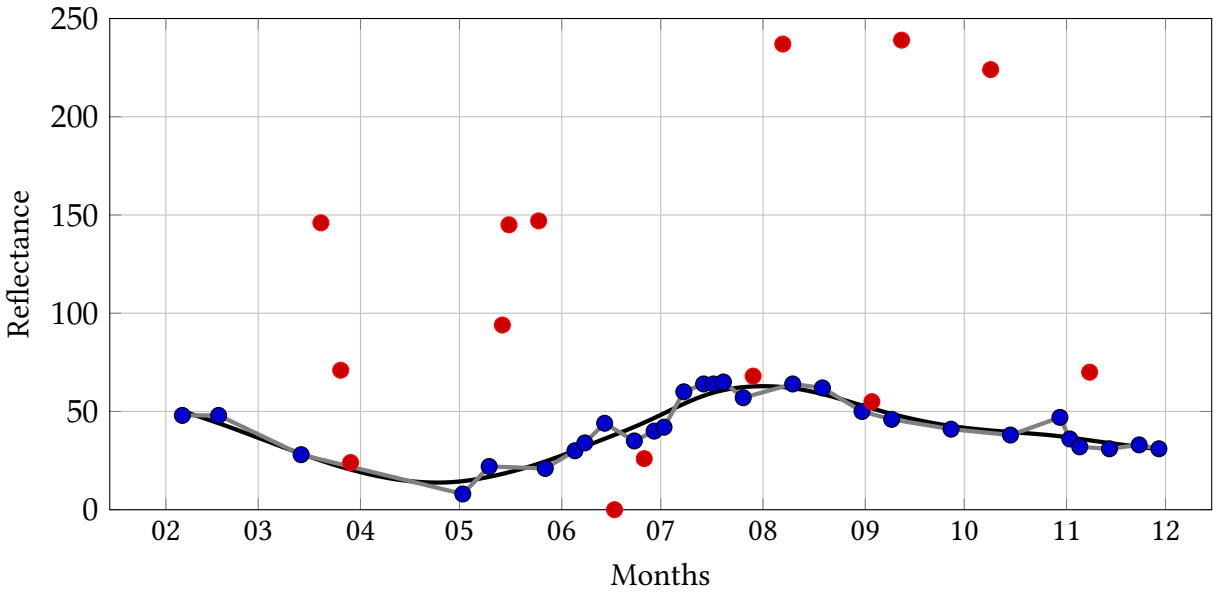


Figure 1.12.: Example of a temporal profile of a vegetation pixel in the NIR band of Formosat-2 for the year 2006. The blue dots correspond to not noisy data and the red ones correspond to missing/noisy data due to the clouds/shadows. The black and gray line correspond to the temporal reconstruction with a step of one day using the Whittaker filter with optimal regularization parameter found by OCV and a linear interpolation, respectively. Taken from [LOP17]

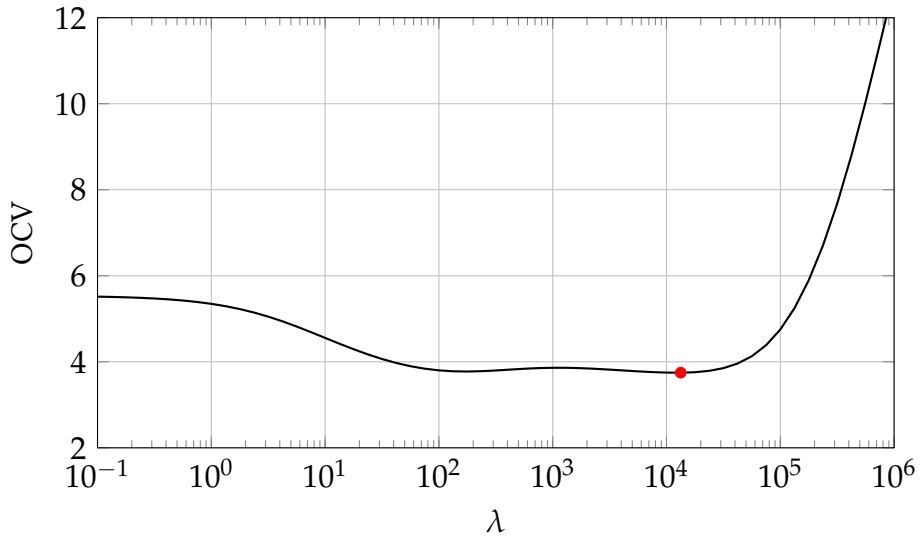


Figure 1.13.: OCV in function of the regularization parameters λ for the Whittaker smoother.

1.4. Contributions

My research activities were and are oriented to the development of methods that allow to extract the most relevant information from the satellite images for a given task. I first focused on the conjoint use of spatial and spectral information for the purpose of classification. It makes me hit issues of high dimensional data, for which I have developed several machine learning models. These developments were driven by thematic scenario, from the general land cover/use map generation to the specific agro-ecological structures identification and characterization. Some of these contributions are reviewed in the next section.

2. Scientific activities

2.1. Spectral-spatial methods

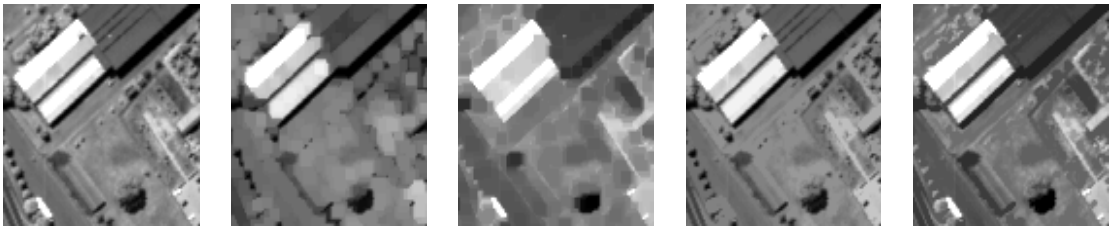
2.1.1. Mathematical morphology for remote sensing images

Mathematical Morphology (MM) is a theory for non-linear image processing [Ser88, Ser82]. Morphological operators have proven their potential in remote sensing image processing [SP02]. Several techniques have been considered with MM, ranging from image segmentation to automatic extraction of objects of interest [SP02, Soi09].

MM aims to analyze spatial relationships between pixels using a set of known shape and size (e.g., disk of radius 3 pixels), called the Structuring Element (SE) [Soi03]. The two basic MM operators are *erosion* and *dilation*. The erosion expands objects of the image that are darker than their surrounding, while the dilation shrinks them (and vice-versa for objects that are brighter than their surrounding). Moreover, bright (respectively dark) structures that cannot contain the SE are removed by erosion (dilation). Hence, both erosion and dilation are non-invertible transformations.

Combining erosion and dilation, *opening* and *closing* operators can be defined. The opening is defined as an erosion followed by a dilation with the same SE. The idea to dilate the eroded image is to recover most structures of the original image, *i.e.*, structures that were not removed by the erosion and hence are bigger than the SE. The closing is defined as a dilation followed by an erosion with the same SE. Hence, with opening or closing it is possible to get, for a given size of B, which structures (buildings, roads ...) of the image are smaller than B. However, opening and closing operators are not connected filters. For instance, two buildings can be merged into one and thus for instance bias the analysis of the size distribution, see Figure 2.1. In order to avoid that problem, connected operators such as geodesic operators can be used [CSS95].

Using opening/closing by reconstruction it is possible to determine the size of the differ-



(a) (b) (c) (d) (e)

Figure 2.1.: (a) Original image, (b) Opened image, (c) Closed image, (d) Geodesicaly opened image and (e) Geodesicaly closed image. The SE was a disk of radius 3 pixels. It can be seen on (c) that with the conventional closing the two bright buildings are merged into one. This is not the case with the geodesic operator.

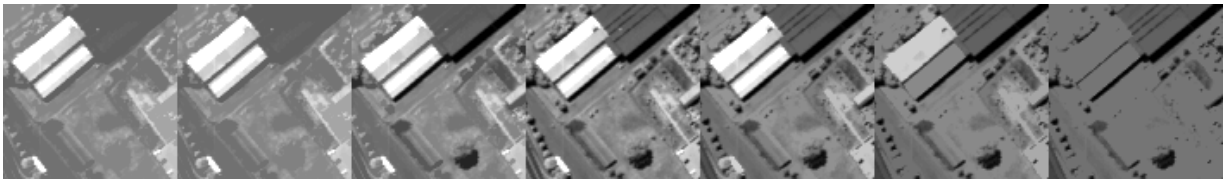


Figure 2.2.: Morphological profile constructed with three opening/closing by reconstruction with a circular SE of size 2, 6 and 10. The left part corresponds to the closings by reconstruction and the dark objects are progressively deleted, *e.g.*, the shadow of the big tree in the middle of the image. The right part corresponds to the openings by reconstruction and the bright objects are progressively deleted, *e.g.* the buildings in the upper part of the image.

ent structures of the image [PB01]. For a given size of the SE, it is possible to get structures which are smaller (they are removed) or bigger (they are preserved) than the SE. Applying such operators with a range of SE of growing size, one can extract information about the contrast and the size of the structures present in the image. This concept is called *granulometry*. The *Morphological Profile* (MP) of has been defined as the composition of a granulometry built with opening by reconstruction and a (anti-) granulometry of built with closing by reconstruction.

From a single panchromatic image, the MP results in a multi-band image. An example of MP is given Figure 2.2. Its uses for the classification of panchromatic images has shown a good improvement in terms of classification accuracy [BPA03, TPKE09, BGMF+08]. However, when considering multivalued images, such as hyperspectral images, the direct construction of the MP is not straightforward, because of the lack of ordering relation between vector. In order to overcome this shortcoming, several approaches have been considered (see [AL07] for a review of several multivariate morphological filters).

One common strategy is to first extract a reduces set of relevant features, for instance using principal component analysis, and then build the MP independently on each feature. All the MPs are stacked together using a feature fusion approach. This leads to the so-called *Extended MP* (EMP) [BPS05], see Figure 2.3. If principal component analysis is used, the EMP can be written as

$$\mathrm{EMP} = \left[\mathrm{MP}_{pc_1}, \ldots, \mathrm{MP}_{pc_c}\right]$$

where MP_{pc_1} is the morphological profile build on the first principal component computed in the spectral domain and c is the number of extracted components. This number is usually set according to the cumulative explained variance. Previous works using alternative feature reduction algorithms, such as independent component analysis (ICA), have led to equivalent results in terms of classification accuracy.

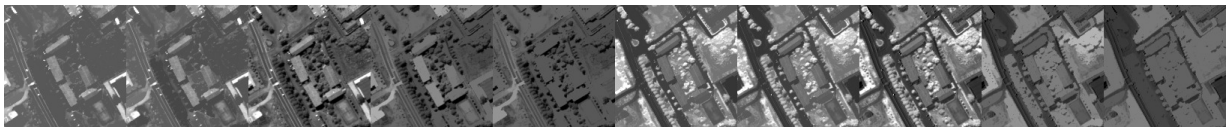


Figure 2.3.: Extended morphological profile build on the first two principal components. Each of the original profile has two openings and two closings. Circular structuring element with radius increment 4 was used ($r = 4, 8$).

2.1.2. Contributions to the Extended Morphological Profile

Contribution 1:

The first contribution to the EMP concerns the fusion of the original spectral information and the EMP using a feature fusion approach, where the spectral feature and morphological feature are combined to build a large feature vector. Feature reduction can be done for each data sets prior to the concatenation and usually helps in improving the classification accuracy. Yet simple, this strategy provides good results on several data sets. This work leads to the following publication [FBCS08].

Contribution 2:

My second contribution to the EMP methodology was to investigate the use of *kernel principal component analysis* (KPCA) rather than linear feature extraction algorithms. The assumption was that much more spectral information will be captured by the KPCA than with the PCA. The following section presents the KPCA and how the KPCA was applied to hyperspectral remote sensing images. More details can be found in [FCB09].

Kernel principal component analysis for the construction of the EMP A brief description of the kernel principal component analysis for feature reduction of remote sensing data is given in this part. Theoretical foundation can be found in [SSM98, SMB+99, MMR+01].

The starting point is a set of pixel vector $\mathbf{x}_i \in \mathbb{R}^d$, $i \in [1, \ldots, n]$. The conventional PCA solve the eigenvalue problem:

$$\lambda \mathbf{v} = \Sigma_{\mathbf{x}} \mathbf{v}, \text{ subject to } \|\mathbf{v}\|_2 = 1 \tag{2.1}$$

where $\Sigma_{\mathbf{x}} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{m}_x)(\mathbf{x}_i - \mathbf{m}_x)^\top$ and $\mathbf{x}_c$ is the centered vector $\mathbf{x}$. The projection on the first k principal components is done as: $\mathbf{x}_{pc} = [\mathbf{v}^1 | \ldots | \mathbf{v}^k]^\top \mathbf{x}$.

To capture higher-order statistics, the data can be mapped onto another space $\mathcal{H}$ [1]:

$$\begin{array}{rcl} \phi : \mathbb{R}^n & \to & \mathcal{H} \\ \mathbf{x} & \mapsto & \phi(\mathbf{x}) \end{array} \tag{2.2}$$

where Φ is a function that may be non-linear, and the only restriction on $\mathcal{H}$ is that it must have the structure of a reproducing kernel Hilbert space (RKHS), not necessarily of finite dimension.

[1] From now, $\mathbb{R}^d$ is called the input space and $\mathcal{H}$ the feature space.

The PCA in $\mathcal{H}$ can be performed as in the input space, but thanks to the kernel trick [Aro50], it can be performed directly in the input space. The kernel PCA (KPCA) solves the following eigenvalues problem:

$$\lambda \boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}, \text{ subject to } \|\boldsymbol{\alpha}\|_2 = \frac{1}{\lambda} \tag{2.3}$$

where $\mathbf{K}$ is the kernel matrix constructed as follows:

$$\mathbf{K} = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ k(\mathbf{x}_2, \mathbf{x}_1) & k(\mathbf{x}_2, \mathbf{x}_2) & \dots & k(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & k(\mathbf{x}_n, \mathbf{x}_2) & \dots & k(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}. \tag{2.4}$$

The function k is the core of the KPCA. It is a positive semi-definite function on $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ that introduces non-linearity in the processing. It is call *kernel*. Classical kernels are the polynomial kernel, $q \in \mathbb{R}^+$ and $p \in \mathbb{N}^+$,

$$k(\mathbf{x}, \mathbf{y}) = \left(\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}} + q\right)^p, \tag{2.5}$$

and the Gaussian kernel, $\sigma \in \mathbb{R}^+$,

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right). \tag{2.6}$$

As with the conventional PCA, when (2.3) is solved, the projection is done:

$$\phi^k_{kpc}(\mathbf{x}) = \sum_{i=1}^{n} \alpha^k_i k(\mathbf{x}_i, \mathbf{x}). \tag{2.7}$$

Note we have assumed that $\mathbf{K}$ is centered, otherwise it can be centered as [SS02]:

$$\mathbf{K}_c = \mathbf{K} - \mathbf{1}_n\mathbf{K} - \mathbf{K}\mathbf{1}_n + \mathbf{1}_n\mathbf{K}\mathbf{1}_n \tag{2.8}$$

where $\mathbf{1}_n$ is a square matrix such as $(\mathbf{1}_n)_{ij} = \frac{1}{n}$.

To illustrate how the KPCA works, a short example is given here. Figure 2.4 represents three Gaussian clusters. The conventional PCA would result in a rotation of the space, that is, the three clusters would not be identified, as shown in Figure 2.4(a). Figure 2.4(b) represents the projection onto the first two kernel principal components (KPCs) using a Gaussian kernel. The structure of the data is better captured than with PCA: clusters can be clearly identified on the first KPCs.

In practice, building the EMP with KPCA can leads to substantial improvement in terms of classification accuracy. The table 2.1 shows the classification accuracy obtained on the *University of Pavia* data set. 12 KPCs were extracted, reaching 95% of the total variance, against 3 for the conventional PCA. For this data set, an increased of 6.4% of the kappa coefficient is obtained.

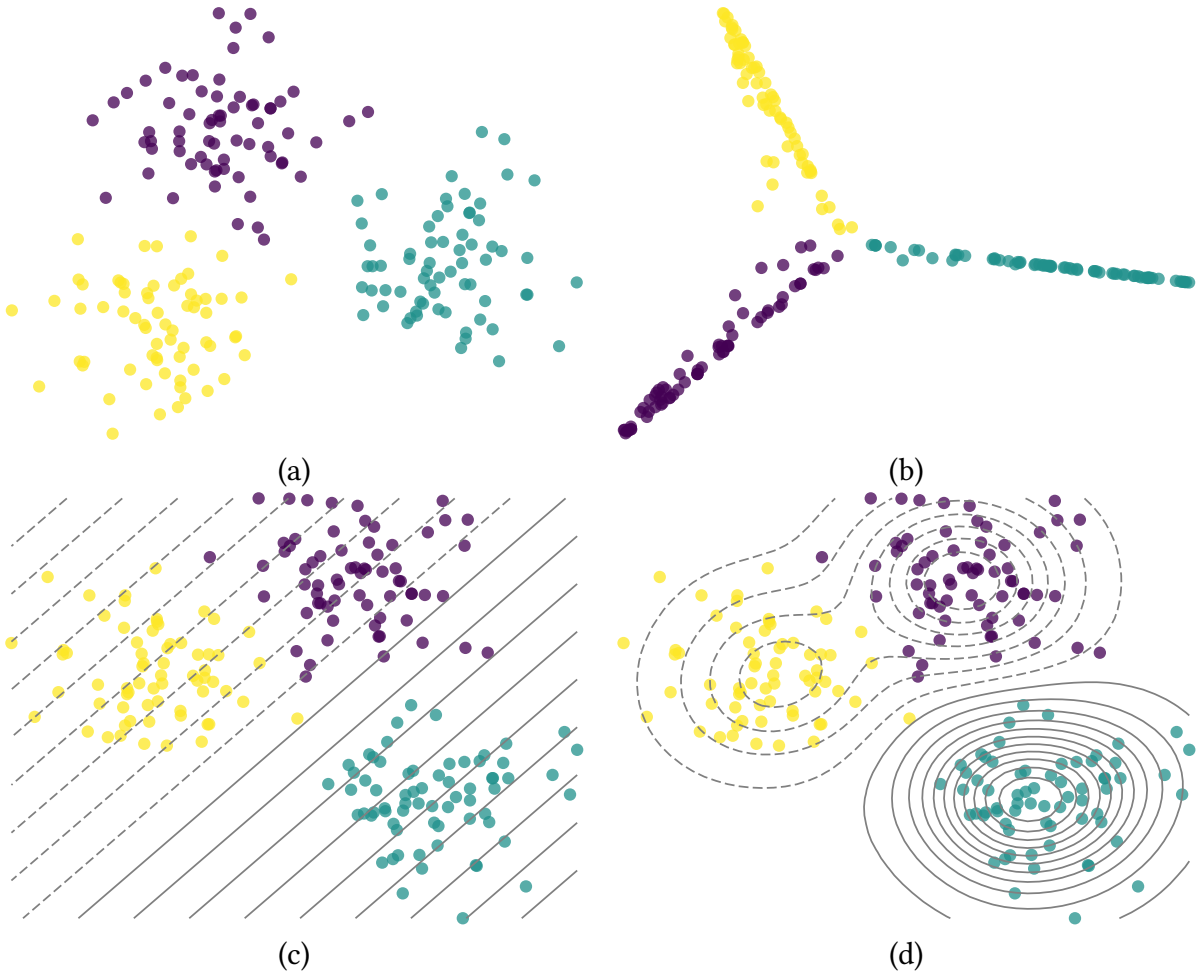
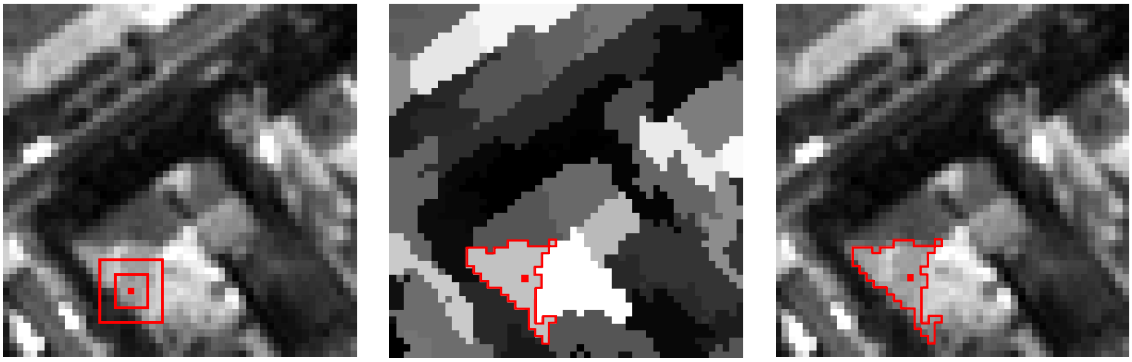


Figure 2.4.: PCA versus KPCA. Three Gaussian clusters, and their projection onto the two principal components (a) and onto the two first kernel principal components with a Gaussian kernel (b). (c) and (d) represent, respectively, the contour plot of the projection onto the first component for the PCA and the KPCA with Gaussian kernel. Note how with the Gaussian kernel the first component "fits" the individual clusters. The intensity of the contour plot is proportional to the value of the projection of the first (kernel) axis.

Table 2.1.: Classification results for the University Area data set (OA: Overall Accuracy, AA: Average class Accuracy, κ: kappa coefficient)

Feature	Raw	PCA+EMP	KPCA+EMP
OA	79.4	92.0	**96.6**
AA	88.1	93.2	**96.2**
κ	74.5	89.6	**95.4**



(a) (b) (c)

Figure 2.5.: Morphological neighborhood. (a) Original image and fixed square neighborhood (in red). (b) Filtered image and neighbor-set defined using area flat zones filter of size parameter $\lambda = 30$ [Soi05]. (c) Original image with the defined neighbor-set $\Omega_{\mathbf{x}}$.

2.1.3. Beyond the EMP: The morphological neighborhood

Geodesic opening/closing operators are appropriate in remote sensing because they preserve shapes. However, they cannot provide a complete analysis of remotely sensed images because they only act on the extrema (clear or dark objects) of the image [Soi05, FCB12]. Moreover, some structures may be darker than their neighbors in some parts of the image, yet lighter than their neighbors in others, depending on the illumination. Although this problem can be partially addressed by using an alternate sequential filter (ASF) [CBP03], the MP thus provides an incomplete description of size structures distribution. Figure 2.6 illustrates this phenomenon.

Another approach consists in defining an adaptive neighbor system for each pixel, the *morphological neighborhood* Ω. The morphological neighborhood of a pixel $\mathbf{x}$, $\Omega_{\mathbf{x}}$, is defined as the set of pixels that belongs to the same spatial structure than $\mathbf{x}$. This concept is connected to the more general concept of adaptive image neighborhood in image processing [DP06a, DP06b].

Contribution 3:

My contribution to the definition of adapted neighborhood was developed in [FCB12]. I use a *self-complementary area filter* [Soi05] to extract consistent spatially connected components. Then a pixel-object classification is performed using a composite kernel. These concepts are briefly discussed below, while a detailed discussion is given in [FCB12].

A self-complementary area filter is a filter that removes all structures of the image smaller (in terms of number of pixels) than a user defined threshold, see Figure 2.6 and 2.5. The filtered image is partitioned into flat zones. Each flat zone belongs to one single structure in the original image, as can be seen in Figure 2.5.(b). Furthermore, the smallest structures are removed and only the main structures of interest remain. The morphological neighborhood $\Omega_{\mathbf{x}}$ was defined as the set of pixels that belong to the same flat zone in the filtered image. The neighborhoods defined in this way are applied to the original image. This neighborhood is obviously more homogeneous and spectrally consistent than the conventional 8-connected fixed square neighborhood, see Figure 2.5.

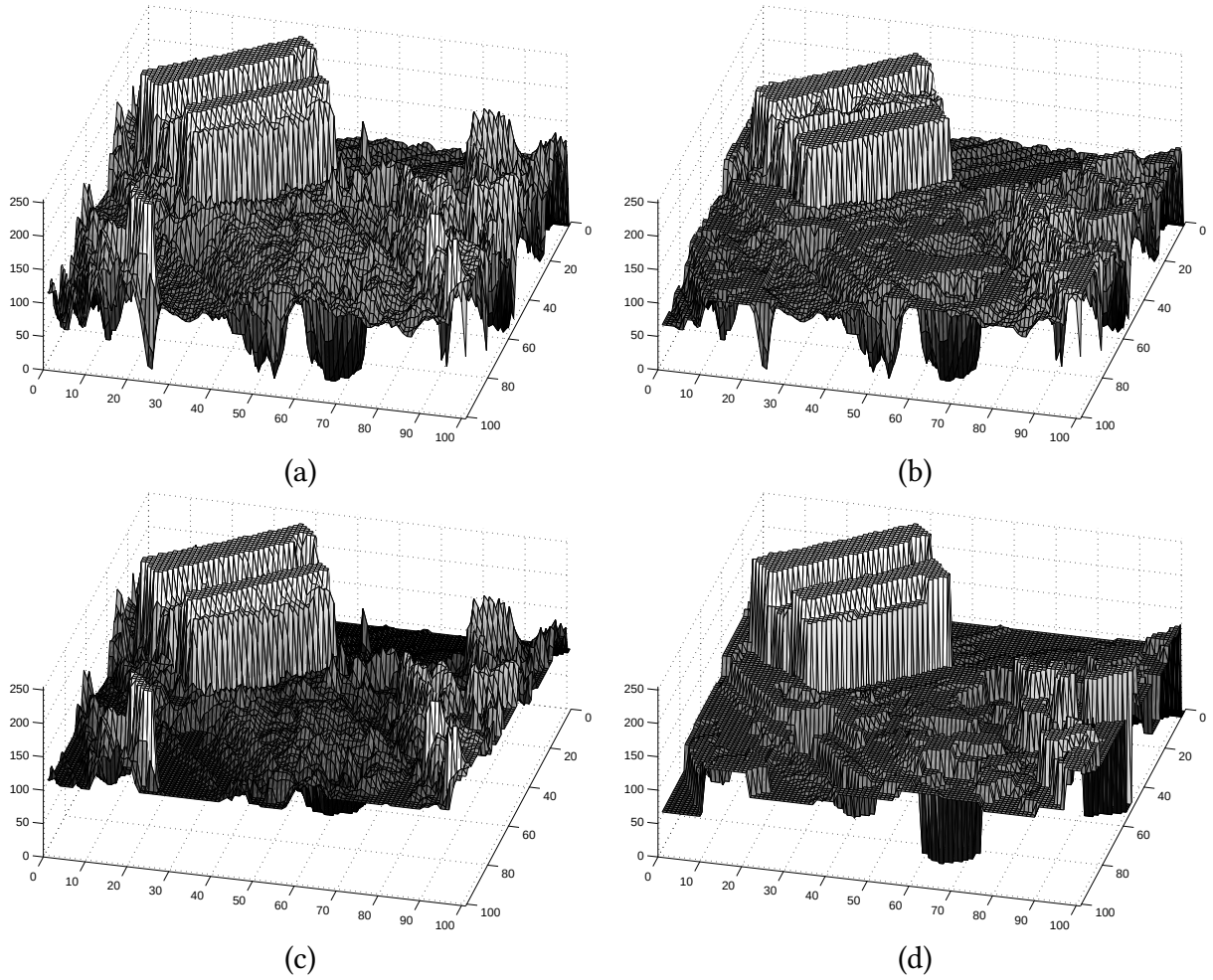


Figure 2.6.: Limitations of the morphological profile. (a) Graph of the image 2.1.(a), (b) Graph of the geodesic opening of image 2.1.(a), (c) Graph of the geodesic closing of image 2.1.(a) and (d) Graph of the image 2.1.(a) filtered by the self-complementary area filter. From (b) and (c), it can be seen that only extrema are processed with the geodesic operators, while all the structures are processed on (d).

Similar to the MP, applying this filter on hyperspectral images is not possible because of the lack of an ordering relation. The same strategy is proposed, which consists in extracting one principal component from which the morphological neighborhood is computed. Then, the neighborhood mask is applied on each band of the data. Once the neighborhood of each pixel is adaptively defined, the spatial information is extracted: the vector median value of the neighbors set $\Omega_{\mathbf{x}}$ is computed for every pixel $\mathbf{x}$ [AHN90]:

$$\mathrm{Y}_{\mathbf{x}} = \mathrm{med}(\Omega_{\mathbf{x}}) \tag{2.9}$$

where $\dim(\mathbf{x}) = \dim(\mathrm{Y}_{\mathbf{x}}) = d$, the number of spectral bands. Unlike the mean vector, the median vector is a vector from the initial set, which ensures a certain spectral consistency since no new spectral values are created.

In conclusion, by defining the morphological neighborhood, every pixel has two features: the *spectral feature* $\mathbf{x}$, which is the original value of each pixel, and the *spatial feature* $\mathrm{Y}_{\mathbf{x}}$, which is the median value computed on each pixel's adaptive neighborhood. The easiest way to use both pieces of information would be to build a stacked vector, but it would not allow the weighting of the different features. In my work, the kernel trick [Vap99] of the SVM was exploited to design a composite kernel that allows the setting of the relative influence of the extracted features. The spatial-spectral kernel $\mathcal{K}$ between two samples $\mathbf{x}$ and $\mathbf{z}$ was defined as:

$$\begin{array}{rccl} \mathcal{K}_{\sigma,\mu} \; : & \mathbb{R}^d \times \mathbb{R}^d & \rightarrow & [0,1] \\ & (\mathbf{x},\mathbf{z}) & \mapsto & (1-\mu)k_\sigma^{spat}(\mathbf{x},\mathbf{z}) + \mu k_\sigma^{spect}(\mathbf{x},\mathbf{z}) \\ & & & 0 \leq \mu \leq 1,\ 0 < \sigma \end{array} \tag{2.10}$$

From previous experiments [FCB06], the spectral kernel is defined as:

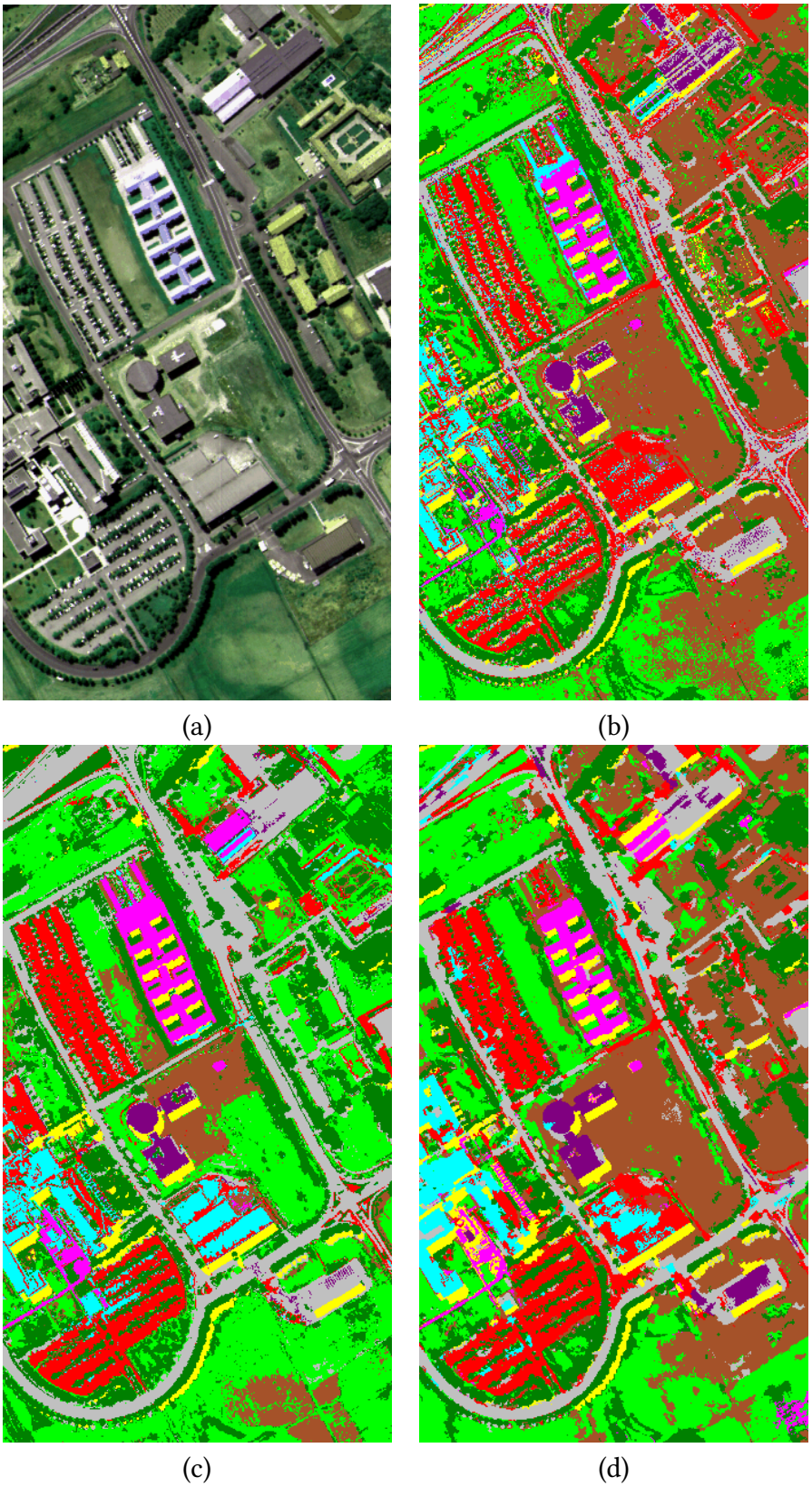
$$\begin{array}{rccl} k_\sigma^{spect} \; : & \mathbb{R}^d \times \mathbb{R}^d & \rightarrow & [0,1] \\ & (\mathbf{x},\mathbf{z}) & \mapsto & \exp\left(-\dfrac{\|\mathbf{x}-\mathbf{z}\|^2}{2\sigma^2}\right) \end{array} \tag{2.11}$$

The spatial kernel is defined as follows:

$$\begin{array}{rccl} k_\sigma^{spat} \; : & \mathbb{R}^d \times \mathbb{R}^d & \rightarrow & [0,1] \\ & (\mathbf{x},\mathbf{z}) & \mapsto & \exp\left(-\dfrac{\|\mathrm{Y}_{\mathbf{x}}-\mathrm{Y}_{\mathbf{z}}\|^2}{2\sigma^2}\right) \end{array} \tag{2.12}$$

where $\mathrm{Y}_{\mathbf{x}}$ is the spatial information. The weight μ controls the relative proportion of spatial and spectral information in the final kernel. For instance, for the class *Grass*, the spectral information should be more discriminative while it should be the spatial information for the class *Building*. This parameter is set during the training process, as the hyperparameter σ. As an illustrative example, taken from [FCB12], Table 2.2 shows the hyperparameters of the kernel found for one data set.

When compare to the conventional EMP, this spectral-spatial kernel leads to an increase of the classification accuracy, as reported in the Table 2.3. However, it requires to set up an additional kernel parameters which can be time consuming, especially for large data set. Classification maps are reported in Figure 2.7.



(a) (b) (c) (d)

Figure 2.7.: (a) False color original image of University Area. (b) Classification map using the RBF kernel. (c) Classification map using the EMP. (d) Classification map using the proposed kernel where $\lambda = 30$.

Table 2.2.: Kernel parameters found by 5-fold cross validation for University data set ($\lambda = 30$)

Class	Asphalt	Meadow	Gravel	Tree	Metal Sheet	Bare Soil	Bitumen	Brick	Shadow
μ	0.3	0.1	0.1	0.1	0.9	0.5	0.1	0.2	0.4
σ	0.5	1	0.5	0.5	1	1	1	0.5	4

Table 2.3.: Classification accuracies for University Area data set. The best results for each class are reported in bold face. Δ is the difference between the best $\mathcal{K}^\lambda$ and the original kernel. $\mathcal{K}^\lambda$ means that classification was performed using the proposed kernel and area filtering of size λ.

Class	SVM	EMP	$\mathcal{K}^2$	$\mathcal{K}^5$	$\mathcal{K}^{10}$	$\mathcal{K}^{15}$	$\mathcal{K}^{20}$	$\mathcal{K}^{25}$	$\mathcal{K}^{30}$	$\mathcal{K}^{35}$	$\mathcal{K}^{40}$	Δ
Asphalt	80.6	**93.3**	80.4	80.0	82.3	83.2	84.5	86.3	84.3	84.5	86.1	5.6
Meadow	68.4	73.4	72.2	72.2	75.1	72.2	76.3	76.3	**78.5**	75.1	75.1	10.0
Gravel	73.8	52.4	76.5	77.1	81.7	87.0	**90.6**	89.7	84.8	85.5	85.9	16.8
Tree	97.4	**99.3**	96.4	96.6	94.1	94.9	95.0	94.6	96.8	94.8	97.6	0.1
Metal Sheet	99.4	99.4	99.1	99.2	99.4	96.5	99.4	98.2	**99.8**	99.6	99.6	0.3
Bare Soil	94.8	61.9	95.2	94.3	97.2	95.6	96.8	94.0	95.6	**98.3**	97.8	3.4
Bitumen	91.5	**97.6**	93.3	94.5	89.7	91.5	94.1	93.9	95.5	96.1	93.8	4.6
Brick	91.8	95.1	92.0	92.6	**96.1**	94.4	94.7	95.8	95.4	95.3	95.7	4.2
Shadow	97.0	92.2	95.0	94.8	95.5	92.2	93.5	93.2	**97.7**	96.7	97.2	0.7
OA	80.1	79.8	81.8	81.8	84.0	82.7	85.3	85.2	**86.1**	84.9	85.2	5.9
AA	88.3	85.0	88.9	89.0	90.1	89.7	91.7	91.3	91.9	91.8	**92.1**	3.8
κ	75.1	74.1	77.2	77.2	79.8	78.3	81.4	81.2	**82.3**	80.9	81.4	7.1

2.1.4. Object classification

One of the main limitation of the morphological neighborhood discussed above is that the radiometric information is limited to the median value. In fact, several additional features can be extracted from the neighborhood, either related to its radiometric content or related to its geometry. These techniques refers to *Object Base Image Analysis* [Bla10]. Mean value, standard deviation or higher moment can be extracted for each spectral bands while shape, size, orientation or other spatial feature can be computed from the neighborhood, or more generally the *object*. Many advanced image processing tools have been defined in the last decades [BHK+14, MLM+17]. Depending on the application, the object can be estimated in the image directly, *e.g.* as for the morphological neighborhood, using image processing techniques (segmentation, filtering ...). Objects can be also defined using external data sources, e.g., field survey or institutional data base. In our case study, the classification of grasslands, the objects were digitized using the agricultural Land Parcel Information System "Registre Parcellaire Graphique" (RPG) [CC14]. See figure 2.8 for an illustration.

Considering only the extraction of radiometric feature, one common assumption is to process each band independently: no correlation between bands is used for the description of the object. Therefore, there is a possible loss of information for the object description. To palliate such limitation, we have proposed to model the distribution of pixels in a given object as a Gaussian distribution. Hence, objects are not represented by a vector of features, but by a probability density function. We have further proposed a measure of similarity between Gaussian

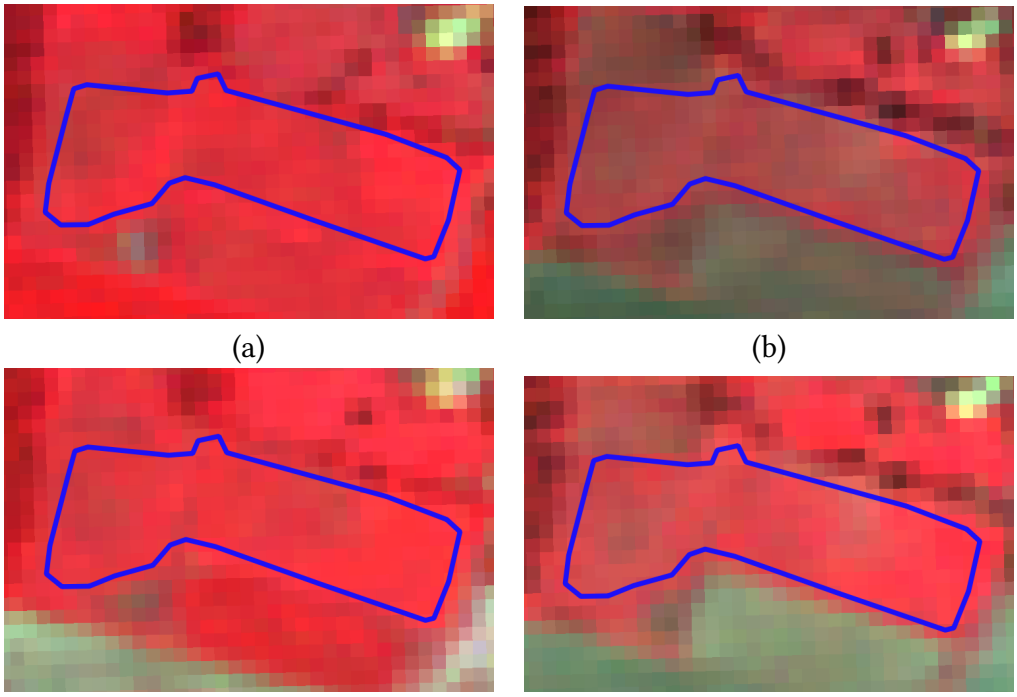


Figure 2.8.: False color Formosat-2 images of the same grassland on two close dates (June and October) in 2013 and 2014 with the same color scale. (a) 6 June 2013, (b) 27 October 2013, (c) 5 June 2014, (d) 23 October 2014. The blue line represents the polygon limits of the grassland.

distribution that is robust to the possible high-dimensionality of the data.

Contribution 4:

My contribution to the classification of objects is detailed in [LFGS17]. We make use of generative mean kernel, design to work on distribution rather than samples, to define a kernel function and derive an explicit formulation for the case of Gaussian distribution. We also propose a formulation that allows to weight the covariance information, the weight being set during the training step. We have applied this model on satellite image time series for the classification of grasslands types.

We assume that the distribution of pixels $\mathbf{x}_l$ is, conditionally to grassland[2] g_i, a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\mu}_i$ is the mean vector and $\boldsymbol{\Sigma}_i$ is the covariance matrix. 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. The empirical

[2]In the following, I will refer to grasslands as the objects to be classified. Strictly speaking, any kind of spatial object can be used with the proposed model.

estimation give:

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{n_i}\sum_{l=1}^{n_i}\mathbf{x}_{il} \tag{2.13}$$

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i-1}\sum_{l=1}^{n_i}(\mathbf{x}_{il}-\hat{\boldsymbol{\mu}}_i)(\mathbf{x}_{il}-\hat{\boldsymbol{\mu}}_i)^\top. \tag{2.14}$$

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 constrain $\hat{\boldsymbol{\Sigma}}_i = \mathbf{I}_d$, 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 $\mathcal{N}_i$.

Mean Map Kernels between Distributions Mean map kernels are similarity measures that operate on distributions [MG10]. They have been used in remote sensing for semi-supervised pixel-based learning in [GCCVBCM10]. 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 (Equation (8) [GCCVBCM10]):

$$K^e(p_i,p_j) = \frac{1}{n_i n_j}\sum_{l,m=1}^{n_i,n_j} k(\mathbf{x}_{il},\mathbf{x}_{jm}), \tag{2.15}$$

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* [MG10]:

$$K^g(p_i,p_j) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k(\mathbf{x},\mathbf{x}')\hat{p}_i(\mathbf{x})\hat{p}_j(\mathbf{x}')d\mathbf{x}d\mathbf{x}'. \tag{2.16}$$

Note that Equation (2.15) acts on the realizations of p_i, while Equation (2.16) 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 $k(\mathbf{x},\mathbf{x}') = \exp(-\frac{\gamma}{2}\|\mathbf{x}-\mathbf{x}'\|^2)$, Equation (2.16) reduces to the so-called *Gaussian mean kernel* [MFDS12]:

$$K^G(\mathcal{N}_i,\mathcal{N}_j) = \frac{\exp\left\{-0.5(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_j)^T\left(\hat{\boldsymbol{\Sigma}}_i+\hat{\boldsymbol{\Sigma}}_j+\gamma^{-1}\mathbf{I}_d\right)^{-1}(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_j)\right\}}{|\hat{\boldsymbol{\Sigma}}_i+\hat{\boldsymbol{\Sigma}}_j+\gamma^{-1}\mathbf{I}_d|^{0.5}}, \tag{2.17}$$

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 nor-

malization can be achieved~easily:

$$\begin{aligned}\tilde{K}^G(\mathcal{N}_i,\mathcal{N}_j) &= \frac{K^G(\mathcal{N}_i,\mathcal{N}_j)}{K^G(\mathcal{N}_i,\mathcal{N}_i)^{0.5}K^G(\mathcal{N}_j,\mathcal{N}_j)^{0.5}} \\ &= K^G(\mathcal{N}_i,\mathcal{N}_j)|2\hat{\boldsymbol{\Sigma}}_i+\gamma^{-1}\mathbf{I}_d|^{0.25}|2\hat{\boldsymbol{\Sigma}}_j+\gamma^{-1}\mathbf{I}_d|^{0.25}. \end{aligned} \tag{2.18}$$

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 [Tar05]. 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 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, I describe a new kernel function that allows one to weight the covariance features with respect to the mean features.

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 one to weight the influence of the covariance matrix, the *α-generative mean kernel*:

$$K^{\alpha}(p_i,p_j) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} k(\mathbf{x},\mathbf{x}')\hat{p}_i(\mathbf{x})^{(\alpha^{-1})}\hat{p}_j(\mathbf{x}')^{(\alpha^{-1})}d\mathbf{x}d\mathbf{x}'. \tag{2.19}$$

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*:

$$\begin{aligned}&\tilde{K}^{\alpha}(\mathcal{N}_i,\mathcal{N}_j) = \\ &\frac{\exp\left\{-0.5(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_j)^T\left(\alpha(\hat{\boldsymbol{\Sigma}}_i+\hat{\boldsymbol{\Sigma}}_j)+\gamma^{-1}\mathbf{I}_d\right)^{-1}(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_j)\right\}}{|\alpha(\hat{\boldsymbol{\Sigma}}_i+\hat{\boldsymbol{\Sigma}}_j)+\gamma^{-1}\mathbf{I}_d|^{0.5}}|2\alpha\hat{\boldsymbol{\Sigma}}_i+\gamma^{-1}\mathbf{I}_d|^{0.25}|2\alpha\hat{\boldsymbol{\Sigma}}_j+\gamma^{-1}\mathbf{I}_d|^{0.25}.\end{aligned} \tag{2.20}$$

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

$$\begin{aligned}p_i(\mathbf{x}|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)^{\alpha^{-1}} &= \frac{1}{(2\pi)^{d/2\alpha}}\times\frac{1}{|\boldsymbol{\Sigma}_i|^{1/2\alpha}}\times\exp\left\{-0.5(\mathbf{x}-\boldsymbol{\mu}_i)^{\top}(\alpha\boldsymbol{\Sigma}_i)^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\} \\ &= \frac{(2\pi)^{\frac{d}{2}(1-\frac{1}{\alpha})}}{(2\pi)^{d/2}}\times\alpha^{1/2}\times\frac{|\boldsymbol{\Sigma}_i|^{\frac{1}{2}(1-\frac{1}{\alpha})}}{|\alpha\boldsymbol{\Sigma}_i|^{1/2}}\times\exp\left\{-0.5(\mathbf{x}-\boldsymbol{\mu}_i)^{\top}(\alpha\boldsymbol{\Sigma}_i)^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\} \\ &= \alpha^{1/2}(2\pi)^{\frac{d}{2}(1-\frac{1}{\alpha})}|\boldsymbol{\Sigma}_i|^{\frac{1}{2}(1-\frac{1}{\alpha})}\times p(\mathbf{x}|\boldsymbol{\mu}_i,\alpha\boldsymbol{\Sigma}_i) \\ &= C(\boldsymbol{\Sigma}_i,\alpha)p(\mathbf{x}|\boldsymbol{\mu}_i,\alpha\boldsymbol{\Sigma}_i).\end{aligned} \tag{2.21}$$

Then, plugging Equation (2.21) in Equation (2.19), 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 Equation (2.17) 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 Equation (2.20). □

It is interesting to note that particular values of α and γ lead to known results:

1. $\alpha = 0$: In this case, Equation (2.20) 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 Equation (2.18).
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 an 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. A synthetic example is given in Figure 2.9.

For the discrimination of *young* and *old* grasslands, using a 2 years long time series, the proposed model improved the results both in term of computation time and classification accuracy w.r.t conventional object based methods. Some results are reported in Figure 2.10.

2.2. Classification of high dimensional data

As discussed in the introduction, new generation of remote sensing data results in a large volume of pixels, and in the same time, in a large amount of feature associates to each pixel. In the following, I am going to review some of my work on the classification of high dimensional data. First, parsimony is discussed, then I discuss some of my work on functional statistic, used when feature can be ordered (such as wavelength or time). Finally, this section is concluded by some recent work on joint representation learning and processing.

2.2.1. Parsimonious model

Parsimonious Mahalanobis Kernel The Mahalanobis distance, $d_{\boldsymbol{\Sigma}_c}$, between two samples for a given class, c, with covariance matrix, $\boldsymbol{\Sigma}_c$ is defined as :

$$d_{\boldsymbol{\Sigma}_c}(\mathbf{x},\mathbf{z}) = \sqrt{(\mathbf{x}-\mathbf{z})^t\boldsymbol{\Sigma}_c^{-1}(\mathbf{x}-\mathbf{z})}.$$

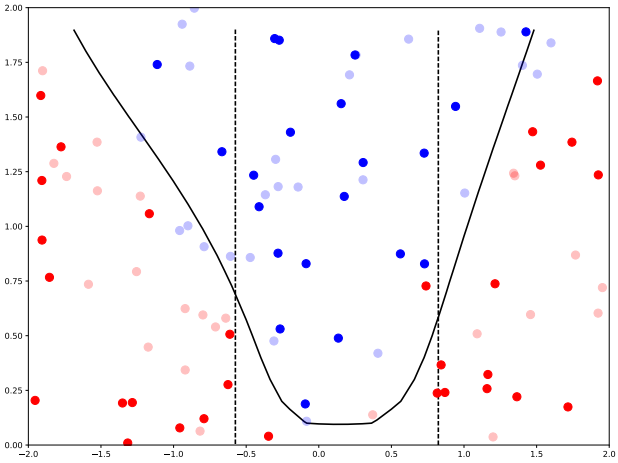


Figure 2.9.: Illustration of the α-Gaussian Mean Kernel. Two classes of Gaussian distributions are considered: the first class (red points) is such as $\sigma/\mu \leq 1$ and the second class (blue points) is such as $\sigma/\mu > 1$. The distributions are represented in the space (μ, σ). The continuous line is the decision function found by a SVM using the α-Gaussian mean kernel with α=1 and $\gamma = 1000$. The dashed line is the decision function found by a SVM using a RBF kernel using only the mean value information.

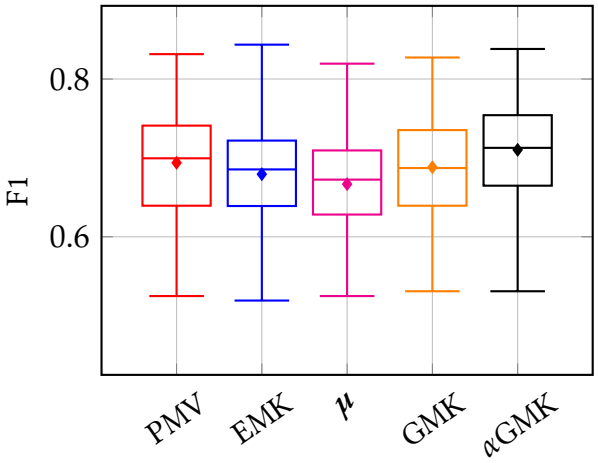


Figure 2.10.: Boxplot of F1 score distribution for the classification of the old and young grasslands. The line in the box stands for the median, whereas the dot stands for the mean. PMV means that each pixels of the object were classified independently and majority voting has been apply to get one class per object/grassland. EMK is the empirical kernel define in eq. 2.15. μ means that only the mean value was used a feature and GMK and αGMK are the proposed kernel.

Previous works on the Mahalanobis kernel [CVRGMM+07, Abe05, KA06, WYT07] were limited by the effect of dimensionality on the matrix inversion. In [CVRGMM+07], the covariance matrix was computed on the whole training set. The associated implicit model is that the classes share the same covariance matrix, which is not true in practice. Diagonal and full covariance matrices were investigated in [Abe05] for the purpose of classification and in [KA06] for the purpose of regression. However, in a similar way, the covariance matrix was computed for all the training samples. Computing the covariance matrix for the Mahalanobis distance with all the training samples is equivalent to project the data on all the principal components, scale the variance to one, and then applying the Euclidean distance. By doing so, classes could overlap more than in the original input space and the discrimination between them would be decreased.

Contribution 5:

I propose to exploit the *emptiness* property of high dimensional spaces to construct a robust parsimonious Mahalanobis kernel. I used the High Dimensional Discriminant Analysis (HDDA) model, proposed by Bouveyron *et al.* in [BGS07], for the definition of a class specific covariance matrix adapted for HD data. The specific signal and noise subspaces are estimated for each considered class, ensuring a parsimonious characterization of the classes. Following the HDDA model it is then possible to derive an explicit formulation of the inverse of the covariance matrix, without any regularization or dimension reduction. The parsimonious Mahalanobis kernel is constructed by substituting in a RBF kernel the Euclidean distance with the Mahalanobis distance computed using the HDDA model. This work is discussed in [FCBV13].

HDDA was originally proposed for classification or clustering with Gaussian mixture model. The covariance matrix of the class c can be written through its eigenvalue decomposition:

$$\mathbf{\Sigma}_c = \mathbf{Q}_c \mathbf{\Lambda}_c \mathbf{Q}_c^t$$

where $\mathbf{\Lambda}_c$ is the diagonal matrix of eigenvalues $\lambda_{ci}, i \in \{1, \ldots, d\}$, of $\mathbf{\Sigma}_c$ and $\mathbf{Q}_c$ is the matrix that contains the corresponding eigenvectors $\mathbf{q}_{ci}$. The HDDA model assumes the p_c first eigenvalues are different and the remaining $d - p_c$ eigenvalues are identical. The model is similar to PPCA, but more general in the sense that additional sub-models can be defined. In particular, the intrinsic dimension p_c are not constrained in HDDA whereas there are assumed to be equal for each class in PPCA.

Under the HDDA framework, the covariance matrix has the following expression:

$$\mathbf{\Sigma}_c = \sum_{i=1}^{p_c} \lambda_{ci} \mathbf{q}_{ci} \mathbf{q}_{ci}^t + b_c \sum_{i=p_c+1}^{d} \mathbf{q}_{ci} \mathbf{q}_{ci}^t$$

where the last $d - p_c$ eigenvalue are equal to b_c. The inverse can be computed explicitly by

$$\mathbf{\Sigma}_c^{-1} = \underbrace{\sum_{i=1}^{p_c} \frac{1}{\lambda_{ci}} \mathbf{q}_{ci} \mathbf{q}_{ci}^t}_{\mathcal{A}_c} + \underbrace{\frac{1}{b_c} \sum_{i=p_c+1}^{d} \mathbf{q}_{ci} \mathbf{q}_{ci}^t}_{\bar{\mathcal{A}}_c}.$$

This statistical model can be understood equivalently by a geometrical assumption: For each class, the data belong to a cluster that lives in a lower dimensional space $\mathcal{A}_c$, namely the signal

subspace. The original input space can be decomposed as $\mathbb{R}^d = \mathcal{A}_c \bigoplus \bar{\mathcal{A}}_c$ (by construction $\bar{\mathcal{A}}_c$ is the noise subspace which contains only white noise).

Using $\mathbf{I} = \sum_{i=1}^{d} \mathbf{q}_{ci}\mathbf{q}_{ci}^t$, $\mathbf{I}$ being the identity matrix, the inverse can be finally written as

$$\mathbf{\Sigma}_c^{-1} = \sum_{i=1}^{p_c} \left(\frac{1}{\lambda_{ci}} - \frac{1}{b_c}\right)\mathbf{q}_{ci}\mathbf{q}_{ci}^t + \frac{1}{b_c}\mathbf{I}. \tag{2.22}$$

Standard likelihood maximization shows that the parameters $(\lambda_{ci}, \mathbf{q}_{ci})_{i=1,\ldots,p_c}$ and b_c can be computed from the sample covariance matrix [BGS07].

The major advantage of such a model is that it reduces drastically the number of parameters to estimate for computing the inverse matrix. Indeed, with the full covariance matrix, $d(d+3)/2$ parameters are to be estimated. With the HDDA model, only $d(p_c+1)+1-p_c(p_c-1)/2$ parameters are to be estimated. For instance, if $d = 100$ and $p_c = 10$, 5150 parameters are needed for the full covariance and only 1056 for the HDDA model. Furthermore, the stability is improved since the smallest eigenvalues of the covariance matrix and their corresponding eigenvectors, which are difficult to compute accurately, are not used in (2.22).

The regularized Mahalanobis kernel for class c is constructed by substituting to the Euclidean distance in the Gaussian kernel and switching eigenvalues $(\hat{\lambda}_{ci}, \hat{b}_c)$ to hyperparameters $(\sigma_{ci}^2, \sigma_{c\hat{p}_c+1}^2)$ that are optimized during the training step:

$$k_m(\mathbf{x}, \mathbf{z}|c) = \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{\hat{p}_c} \frac{\|\hat{\mathbf{q}}_{ci}^t(\mathbf{x}-\mathbf{z})\|^2}{\sigma_{ci}^2} + \frac{\|\mathbf{x}-\mathbf{z}\|^2}{\sigma_{c\hat{p}_c+1}^2}\right)\right) \tag{2.23}$$

where σ_{ci}, $i \in \{1, \ldots, \hat{p}_c+1\}$ are the hyperparameters of the kernel. The hyperparameters have been introduced for the following reason. It is known that the principal directions are not optimal for classification since they do not maximize any discrimination criterion. However, they still span a subspace where there are variations in the data of the considered class. The hyperparameters σ_{ci} allow to control which directions are more relevant (or discriminative) for the classification process: The feature space is modified during the training process to ensure a better discrimination between samples. In [FCBV13], I optimize these hyper-parameters using the radius margin bound of SVM.

It is interesting to note that the regularized Mahalanobis kernel can be expressed as the product of Gaussian kernels:

$$k_m(\mathbf{x}, \mathbf{z}|c) = k_g(\mathbf{x}, \mathbf{z}) \times \prod_{i=1}^{\hat{p}_c} k_g(\hat{\mathbf{q}}_{ci}^t\mathbf{x}, \hat{\mathbf{q}}_{ci}^t\mathbf{z}). \tag{2.24}$$

Experiment were conducted on several high dimensional data set. Table 2.4 shows the classification accuracy results for the one hyperspectral data set.

Parsimonious Gaussian Process Models With C. Bouveyron and S. Girard, we have extended the HDDA model to Gaussian Process for the classification of hyperspectral remote sensing images.

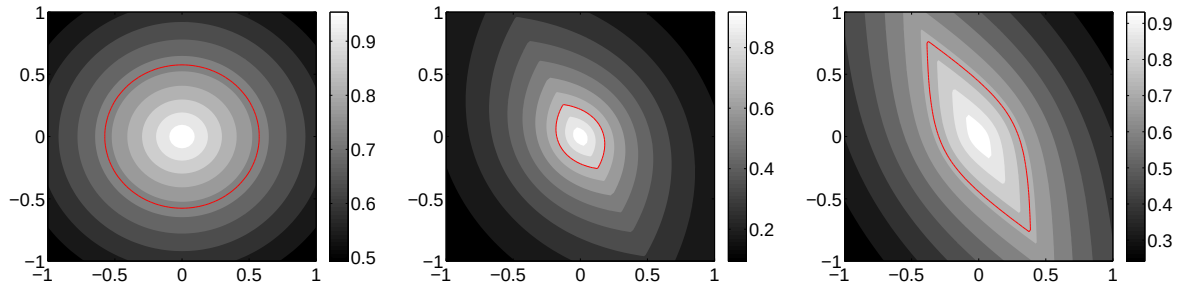


Figure 2.11.: Values of the kernel function $k(\mathbf{0}, \mathbf{x})$ with $\mathbf{0} = [0, 0]$ and $\mathbf{x} \in [-1, 1] \times [-1, 1]$. The red line represents the contour line for the value 0.75. (a) is the Gaussian kernel, (b) is the kernel (2.23) with $\sigma_1^2 = \sigma_2^2 = 0.5$, (c) is the kernel (2.23) with $\sigma_1^2 = 1.5$ and $\sigma_2^2 = 0.5$. The covariance matrix used was $[0.6 \ -0.2; -0.2 \ 0.6]$ and the signal subspace was of dimension 1, spanned by the first eigenvector of the covariance matrix.

Table 2.4.: Classification accuracies for the different kernels in percentage of correctly classified samples. In the first column, the numbers in brackets represent the total number of training and testing samples for each class, respectively.

	$\hat{p}_c$	Gaussian	PCA-Mahalanobis	HDDA-Mahalanobis
Asphalt (548, 6631)	12	94.8	**95.8**	**95.8**
Meadow (540, 18649)	10	79.4	**83.6**	82.1
Gravel (392, 2099)	9	97.2	**97.5**	97.2
Tree (524, 3064)	14	94.3	**98.2**	**98.2**
Metal Sheet (265, 1345)	7	99.8	**99.9**	**99.9**
Bare Soil (532, 5029)	9	87.8	85.9	**88.4**
Bitumen (375, 1330)	21	98.8	98.7	**99.0**
Brick (514, 3682)	12	96.7	**97.2**	**97.2**
Shadow (231, 947)	14	**99.9**	**99.9**	**99.9**
Average class accuracy		94.3	95.2	**95.3**

Contribution 6:

We have proposed a family of parsimonious Gaussian process models for classification. The idea was define a parsimonious Gaussian mixture model in the feature space induces by kernel function. The main issue was the feature space can be infinite and therefore, conventional multivariate model cannot be defined. To overcome this, we assume that the projection $\phi(\mathbf{x})$ was conditionally to its class, a Gaussian process. Using HDDA, several parsimonious models were defined. This work was discussed in [FBG15].

Let $\mathcal{S} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ be a set of training samples, where $\mathbf{x}_i \in \mathbb{R}^d$, is a pixel and $y_i \in \{1, \ldots, C\}$ its class, and C the number of classes. In this work, the Gaussian kernel function is used $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2_{\mathbb{R}^d}\right)$, with $\gamma > 0$, and its associated mapping function is denoted $\phi : \mathbb{R}^d \to \mathcal{F}$ (the use of another kernel is possible). Its associated feature space $\mathcal{F}$ is infinite dimensional. Therefore the conventional multivariate normal distribution used in GMM cannot be defined in $\mathcal{F}$.

To overcome this, let us assume that $\phi(\mathbf{x})$, conditionally on $y = c$, is a Gaussian process on $J \subset \mathbb{R}$ with mean $\boldsymbol{\mu}_c$ and covariance function $\boldsymbol{\Sigma}_c$. We note $\phi(\mathbf{x})_{cj}$ the projection of $\phi(\mathbf{x})$ on the eigenfunction $\mathbf{q}_{cj}$ using the following scalar product

$$\langle \phi(\mathbf{x}), \mathbf{q}_{cj} \rangle = \int_J \phi(\mathbf{x})(t) \mathbf{q}_{cj}(t) dt.$$

Hence, for all $r \geq 1$, random vectors on $\mathbb{R}^r$ defined by $[\phi(\mathbf{x})_1, \ldots, \phi(\mathbf{x})_r]$ are, conditionally on $y = c$, multivariate normal vectors. In $\mathbb{R}^r$, it is now possible to use the GMM decision rule for class c [McL92]:

$$D_c(\phi(\mathbf{x}_i)) = \sum_{j=1}^{r} \left[\frac{\langle \phi(\mathbf{x}_i) - \boldsymbol{\mu}_c, \mathbf{q}_{cj} \rangle^2}{\lambda_{cj}} + \ln(\lambda_{cj}) \right] - 2\ln(\pi_c) \tag{2.25}$$

where λ_{cj} is the j^{th} eigenvalue of $\boldsymbol{\Sigma}_c$ sorted in decreasing order, $\mathbf{q}_{cj}$ its associated eigenfunction and π_c the prior probability of class c. The classification of $\mathbf{x}_i$ is done to class c if $D_c(\phi(\mathbf{x}_i)) < D_{c'}(\phi(\mathbf{x}_i)), \forall c' \in 1, \ldots, C$ [McL92].

If the Gaussian process is not degenerated (*i.e.*, $\lambda_{cj} \neq 0,\ \forall j$), r has to be large to get a good approximation of the Gaussian process. Unfortunately, only a part of the above equation can be computed from a finite training sample set:

$$D_c(\phi(\mathbf{x}_i)) = \underbrace{\sum_{j=1}^{r_c} \left[\frac{\langle \phi(\mathbf{x}_i) - \boldsymbol{\mu}_c, \mathbf{q}_{cj} \rangle^2}{\lambda_{cj}} + \ln(\lambda_{cj}) \right] - 2\ln(\pi_c)}_{\text{computable quantity}} + \underbrace{\sum_{j=r_c+1}^{r} \left[\frac{\langle \phi(\mathbf{x}_i) - \boldsymbol{\mu}_c, \mathbf{q}_{cj} \rangle^2}{\lambda_{cj}} + \ln(\lambda_{cj}) \right]}_{\text{non computable quantity}}$$

where $r_c = \min(n_c, r)$ and n_c is the number of training samples of class c. Consequently, the decision rule cannot be computed in the feature space if $r > n_c$, for $c = 1, \ldots, C$.

Table 2.5.: List of sub-models of the parsimonious Gaussian process model.

Model	Variance inside $\mathcal{F}_c$	$\mathbf{q}_{cj}$	p_c
	Variance outside $\mathcal{F}_c$: Common		
$p\mathcal{GP}_0$	Free	Free	Free
$p\mathcal{GP}_1$	Free	Free	Common
$p\mathcal{GP}_2$	Common within groups	Free	Free
$p\mathcal{GP}_3$	Common within groups	Free	Common
$p\mathcal{GP}_4$	Common between groups	Free	Common
$p\mathcal{GP}_5$	Common within and between groups	Free	Free
$p\mathcal{GP}_6$	Common within and between groups	Free	Common
	Variance outside $\mathcal{F}_c$: Free		
$np\mathcal{GP}_0$	Free	Free	Free
$np\mathcal{GP}_1$	Free	Free	Common
$np\mathcal{GP}_2$	Common within groups	Free	Free
$np\mathcal{GP}_3$	Common within groups	Free	Common
$np\mathcal{GP}_4$	Common between groups	Free	Common

To make the above computational problem tractable, it is proposed to use a parsimonious Gaussian process model in the feature space for each class. These models assume that each class is located in a low-dimensional subspace of the kernel feature space. In [BFG14], it was assumed the noise level is common to all classes (*Definition* 1). In [FBG15], they were extended to the situation where the noise level can be dependent on the class (*Definition* 2).

Definition 1 (Parsimonious Gaussian process with common noise)**.** *A parsimonious Gaussian process with common noise is a Gaussian process $\phi(\mathbf{x})$ for which, conditionally to $y = c$, the eigen-decomposition of its covariance operator $\mathbf{\Sigma}_c$ is such that*

A1. *It exists a dimension $r < +\infty$ such that $\lambda_{cj} = 0$ for $j \geq r$ and for all $c = 1, \ldots, C$.*

A2. *It exists a dimension $p_c < \min(r, n_c)$ such that $\lambda_{cj} = \lambda$ for $p_c < j < r$ and for all $c = 1, \ldots, C$.*

Definition 2 (Parsimonious Gaussian process with class specific noise)**.** *A parsimonious Gaussian process with class specific noise is a Gaussian process $\phi(\mathbf{x})$ for which, conditionally to $y = c$, the eigen-decomposition of its covariance operator $\mathbf{\Sigma}_c$ is such that*

A3. *It exists a dimension $r_c < r$ such that $\lambda_{cj} = 0$ for all $j > r_c$ and for all $c = 1, \ldots, C$. When $r = +\infty$, it is assumed that $r_c = n_c - 1$.*

A4. *It exists a dimension $p_c < r_c$ such that $\lambda_{cj} = \lambda_c$ for $j > p_c$ and $j \leq r_c$, and for all $c = 1, \ldots, C$.*

Assumptions *A1* and *A3* are motivated by the quick decay of the eigenvalues for a Gaussian kernel [BBM08]. Hence, it is possible to find $r < +\infty$ (or r_c) such as $\lambda_{cr} \approx 0$. Assumptions *A2* and *A4* express that the data of each class live in a specific subspace of size p_c, the signal subspace, of the feature space. The variance in the signal subspace for class c is modeled by parameters $\lambda_{c1}, \ldots, \lambda_{cp_c}$ and the variance in the noise subspace is modeled by λ or λ_c. This

Table 2.6.: Experimental results in terms of accuracy and processing time. The values reported correspond to the average results obtained on 20 repetitions. Boldface results indicate best results for a given data set. Multiple boldface results indicate that differences between them are not statistically significant.

	Kappa coefficient			Processing time (s)		
	University	KSC	Heves	University	KSC	Heves
$p\mathcal{GP}_0$	0.768	0.920	0.664	18	31	148
$p\mathcal{GP}_1$	**0.793**	**0.922**	**0.671**	18	33	151
$p\mathcal{GP}_2$	0.617	0.844	0.588	18	31	148
$p\mathcal{GP}_3$	0.603	0.842	0.594	19	33	152
$p\mathcal{GP}_4$	0.661	0.870	0.595	19	34	152
$p\mathcal{GP}_5$	0.567	0.820	0.582	18	32	148
$p\mathcal{GP}_6$	0.610	0.845	0.583	19	34	152
$np\mathcal{GP}_0$	0.730	0.911	0.640	17	31	148
$np\mathcal{GP}_1$	**0.792**	0.921	**0.677**	18	33	151
$np\mathcal{GP}_2$	0.599	0.838	0.573	18	31	148
$np\mathcal{GP}_3$	0.578	0.817	0.585	19	33	152
$np\mathcal{GP}_4$	0.578	0.817	0.585	19	33	152
KDC	0.786	**0.924**	0.666	98	253	695
RF	0.646	0.853	0.585	**3**	**3**	**18**
SVM	**0.799**	**0.928**	0.658	10	28	171

model is referred to by $p\mathcal{GP}_0$ for the common-noise assumption or $np\mathcal{GP}_0$ for the class-specific noise assumption.

From these models, it is possible to derive several sub-models. Table 2.5 lists the different sub-models that can be built from $p\mathcal{GP}_0$ and $np\mathcal{GP}_0$.

Model inference and fast estimation of the hyperparameters are discussed in the original paper. We ends up to perform problems similar to Kernel PCA, one for each class, with some additional constraint when common parameters exists between class specific model.

I choose to report only the main results of the original paper. In table 2.6, results are presented on three real data set.

Parsimony exploits the *emptiness* property of high dimensional space. Another interesting property of hypertemporal or hyperspectral data is the existence of high correlations between consecutive variables (i.e., two consecutive spectral bands or temporal acquisition), inducing for instance ill-conditioned covariance matrix. One way to work around such statistical problem in the hyperspectral setting consists in considering the i^{th} sampled pixel $\mathbf{x}_i$ as a discretized version of the curve $\chi_i := \{\chi_i(\lambda);\ \lambda \in [\lambda_{\min}, \lambda_{\max}]\}$, where the word *curve* stands for a real quantity $\chi_i(\lambda)$ varying continuously with the wavelength λ (for the hypertemporal setting, the curve varies continuously with the time). Such modeling is considered in the next section.

2.2.2. Functional data analysis

Functional data analysis (FDA) is the branch of statistics devoted to the analysis of functional data, i.e. curves. It has received a lot of attention in the 2000s [RS02, RS05a, FV06]. However, few works have considered these techniques in the framework of hyperspectral or hy-

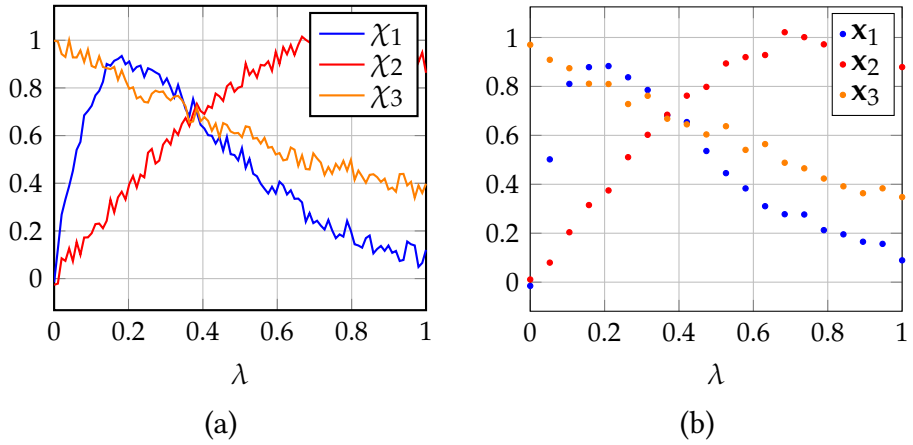


Figure 2.12.: Synthetic illustration of (a) thee random curves and (b) corresponding vectorial observations.

pertemporal remote sensing images processing. Yet, formalizing the pixel spectro-temporal features as curves enjoys the advantage to consider intrinsic feature like the order of the wavelengths/temporal acquisition or the shape of the hyperspectra/hypertemporal profiles (e.g. through explicit derivatives). With F. Ferraty, we have investigated the potential of FDA in the context of hyperspectral data classification, during a joint Phd thesis work.

Contribution 7:

We first compare state of the art FDA methods with conventional multivariate machine learning techniques for the classification of hyperspectral data. This work was published in [ZFF18] and it shows the potential of FDA methods. Then we have defined a novel non-parametric functional operator to account for the noise in the hyperspectral data. It consist of a kernel presmoothing step on the noisy functional predictor. After that, the kernel estimator of the regression operator is built using the smoothed functional covariate instead of the original corrupted one. In particular, we have proved that the convergence rate of the nested estimator apply on noisy data is similar to the conventional estimator obtain on clean data [FZF19].

Experimental assessment of FDA The following notation are defined for the particular case of FDA and hyperspectral data: $[\chi_1, \ldots, \chi_n]$ stand for samples of the random curves $\chi_i := \{\chi_i(\lambda);\ \lambda \in [\lambda_{\min}, \lambda_{\max}]\}$ and $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ are the corresponding d-dimensional random vectors containing the discretized hyperspectra on $[\lambda_1, \ldots, \lambda_d]$. The figure 2.12 provides an illustration of such notation.

Two ways are mostly used to handle functional data. One way consists in operating a basis expansion of the functional data, then applying standard multivariate methods on the coefficients of this decomposition. This approach is discussed in our paper [ZFF18, Section 2.3].

The other way keeps the functional framework of the data, applying on it approximates of functional operators. In our work, we use the so-called NonParametric Functional Discrimination (NPFD). NPFD belongs to the general nonparametric functional data analysis framework proposed in Ferraty and Vieu [FV06]. Given our hyperspectral classification context, the non-

parametric functional model can be defined for each class $c \in \{1, \dots, C\}$ by a regular operator $r_c(\chi_i) := \mathbb{P}(y = c|\mathcal{X} = \chi_i)$, mapping an observed hyperspectrum χ_i at pixel i into the range $[0; 1]$. Only smoothness hypothesis about these C operators $r_1, \dots, r_C$ are required by this approach. Remark that $\mathbb{P}(y = c|\mathcal{X} = \chi_i) = \mathbb{E}\left(\mathbf{1}_{[y=c]}|\mathcal{X} = \chi_i\right)$. Therefore, for $c = 1, \dots, C$, the conditional probability $\mathbb{P}(y = c|\mathcal{X} = \chi_i)$, or equivalently the operator r_c, can be estimated using a functional adaptation of the Nadaraya-Watson kernel estimator [Nad64, Wat64]:

$$\hat{r}_c(\chi_i) = \frac{\sum_{i'=1}^{n} \mathbf{1}_{[y_{i'}=c]} K\left(\frac{\delta(\chi_i, \chi_{i'})}{h}\right)}{\sum_{i'=1}^{n} K\left(\frac{\delta(\chi_i, \chi_{i'})}{h}\right)}, \tag{2.26}$$

where h is a smoothing parameter, K is an asymmetric kernel function defined on $[0; 1]$ and δ is a proximity measure between two functions, which choice has been restricted to some pseudo-metrics family. This way, this classification estimation problem can be written as a regression estimation problem. Once each conditional probability $\mathbb{P}(y = c|\mathcal{X} = \chi_i)$ is estimated for $c = 1, \dots, C$, the class y_i is predicted following the MAP rule, assigning χ_i to the class which estimated membership probability is the highest:

$$y_i = \arg \max_{c \in \{1,\dots,C\}} \mathbb{P}(y = c|\mathcal{X} = \chi_i).$$

In practice, the choice of the scale parameter h is replaced by an equivalent choice of a k-nearest neighbours parameter k, changing a continuous parameter by a discrete one, in order to facilitate the implementation of this method.

Eight classification methods have been compared on three hyperspectral data set: Ridge regularized Gaussian Mixture Models (RGMM), High-Dimensional Discriminant Analysis (HDDA), Support Vector Machines (SVM), Random Forest (RF), NonParametric Functional Discrimination (NPFD), Functional Multinomial Logistic Model (FMLM), Functional Support Vector Machines (FSVM) and Functional Random Forest (FRF). The R software [R C15] has been used. The three later methods use basis expansion approach mentioned above. More details about computations are available in supplemental material of [ZFF18].

Three experimental protocols have been applied on each dataset. The first protocol studies the influence of a balanced learning sample size on the classification performance. The second protocol provides the case of unbalanced learning set. Two approaches are compared; the number of learning pixels per class is either proportional to the corresponding total number of pixels available in the class, or proportional to its corresponding average standard deviation over all spectral bands. Both cases have been calibrated to contain the same number of learning pixels (either $30 \times C$ or $480 \times C$). The third protocol compares the robustness of the methods to noise in learning labels. We chose the framework of Bouveyron and Girard [BG09]. These three protocols have been repeated 50 times such as a new training set has been generated for each repetition. The Cohen's Kappa [CG98] has been chosen to compare the methods performances. This criterion is a statistic measuring the inter-rater agreement, taking into account the agreement occurring by chance. For the 50 independent repetitions, the means and standard deviations of the Cohen's Kappa criterion are reported in Table 2.7. I chose to report only the last case of the first and third protocol (480 learning pixels for each class, with and without label noise) in this document. Full results are given in the paper [ZFF18].

Table 2.7.: Cohen's Kappa ($\times 100$) computed on the three data set.

Data set	Label noise	RGMM	HDDA	SVM	RF	FMLM	NPFD	FSVM	FRF
Pavia	No	88.8 (0.5)	78.0 (1.6)	**90.1 (0.5)**	80.2 (1.1)	79.4 (0.8)	87.3 (0.6)	85.6 (0.7)	77.4 (0.9)
	Yes	66.4 (3.1)	42.8 (4.2)	**86.9 (0.8)**	78.1 (1.3)	66.0 (1.1)	84.6 (0.9)	81.6 (1.0)	75.5 (1.3)
AISA	No	64.2 (0.8)	61.5 (2.4)	77.3 (0.4)	70.0 (0.4)	70.2 (0.3)	**79.5 (0.3)**	**79.5 (0.3)**	71.1 (0.3)
	Yes	59.7 (1.3)	52.0 (2.3)	71.3 (0.6)	68.9 (0.4)	63.5 (0.8)	**77.8 (0.3)**	73.8 (0.5)	69.5 (0.4)
AVIRIS	No	81.1 (0.9)	74.1 (0.7)	80.2 (0.8)	77.8 (1.0)	74.1 (1.0)	82.5 (0.7)	**84.2 (0.7)**	78.1 (0.9)
	Yes	69.4 (1.9)	48.2 (3.1)	71.4 (1.2)	75.8 (1.2)	68.1 (1.0)	**81.4 (0.7)**	78.3 (1.2)	75.4 (1.2)

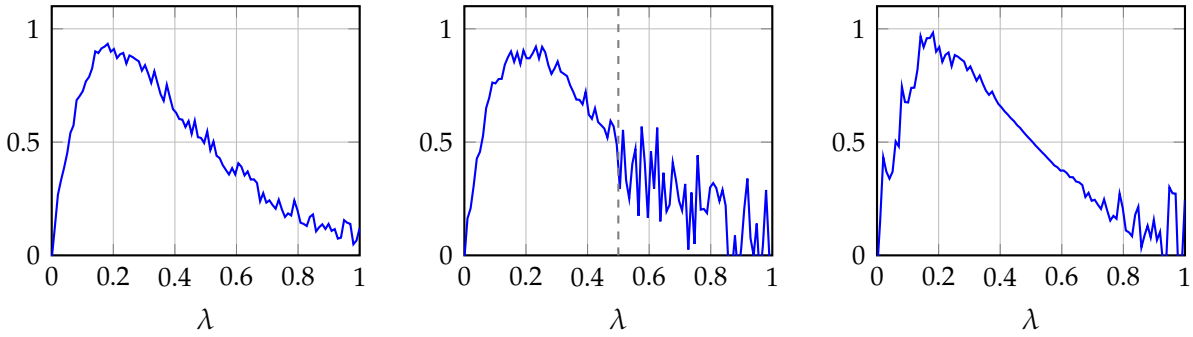


Figure 2.13.: Three contamination scenario: (a) noise with constant variance, (b) the noise has two values (low before 0.5 and high after) and (c) the noise has its value that increased quadratically from the center ($\sigma(\lambda)^2 = (\lambda - 0.5)^2$).

From the results, NPFD competes with state of the art classifiers, and is particularly robust to label noise. Other functional method, based on basis expansion are less accurate: the choice and the design of the basis function are difficult to tune in practice and can cause severe over-smoothing of the original data. On the contrary, NPFD is modified in the next section to allow a smoothing of the original data that is learned during the specific task of classification, to ensure an optimal reconstruction of the data.

Nonparametric regression for contaminated functional predictor In this work, we have supposed that we observe a contaminated version of the spectra:

$$\chi^*(\lambda) = \chi(\lambda) + \eta(\lambda)$$

where η is a random process independent of (χ, y) such as

1. $E[\eta(\lambda)] = 0$,
2. $E[\eta(\lambda)\eta(\lambda')] = \sigma_\eta^2(\lambda)\mathbf{1}_{\lambda=\lambda'}$,
3. σ_η^2 twice differentiable.

This modeling offer the possibility to account for heteroscedeastic noise in the sample, i.e., noise that can vary as a function of the wavelength or the time. See figure 2.13 for an illustration.

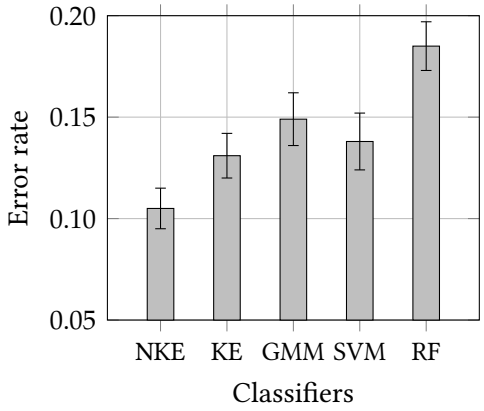


Figure 2.14.: Classification accuracy

In [FZF19] we have proposed a nested estimators, that estimate both the noise-free spectra and the regression operator :

$$\hat{r}(\chi) = \frac{\sum_{i=1}^{n} y_i K_r\left(h_r^{-1}\delta(\chi, \hat{\chi}_i)\right)}{\sum_{i=1}^{n} K_r\left(h_r^{-1}\delta(\chi, \hat{\chi}_i)\right)} \tag{2.27}$$

where $\hat{\chi}_i$ is obtained through

$$\hat{\chi}_i(\lambda) = \frac{\sum_{j=1}^{d} \chi_i^*(\lambda_i) K_s\left(h_s(\lambda_j)^{-1}(\lambda - \lambda_j)\right)}{\sum_{j=1}^{d} K_s\left(h_s(\lambda_j)^{-1}(\lambda - \lambda_j)\right)}, \tag{2.28}$$

and h_s is a smoothing parameter depending on the variable λ. This parameter allows to adapt the smoothing effect accordingly to the noise of each spectral/temporal component.

Under some mild conditions, it is possible to prove [FZF19]:

1. The nested estimator $\hat{\hat{r}}$ converges to the true operator r.
2. The rate of convergence is not decreased in comparison to the situation where noise-free samples are observed as soon as d is much larger than n.
3. The rate of convergence is increased when the ratio d/n is increased.

When have performed experimental simulation a small training set context (30 samples per class) on the *University* data set. Results are given in figure 2.14, results on simulated data can found in the original paper [FZF19]. It can be seen that the proposed perform well w.r.t. to others classifiers. However, the processing time is really high: in our work the parameters h_s were tuned using cross-validation which made this approach not applicable in large scale scenario.

Functional data analysis with Gaussian processes One strategy to develop faster computational methods was to move from a non-parametric model to a parametric model. With S. Girard, we have continued this functional approach for the classification satellite image time series, with a joint PhD work (Alexandre Constantin). In particular we have investigated the

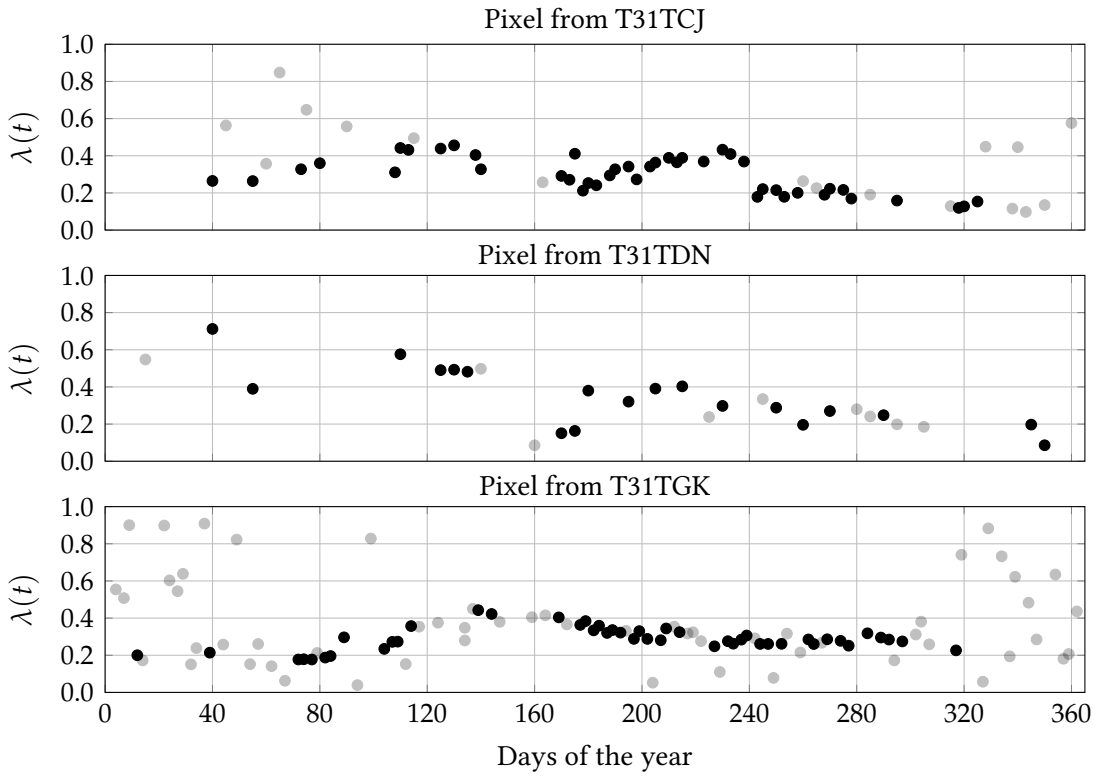


Figure 2.15.: Illustration of the temporal irregular sampling. Available aqcuisitions (blue Sentinel-2 band) for three pixels coming from three different tiles are displayed. A dot indicates an acquisition: black color indicates that the pixel was tagged *clear* by the MAYA processing chain, while gray color indicates that the pixel was tagged *non-clear* (e.g., clouds, shadows, saturation). To see where the tiles are located in France, see https://docs.iota2.net/tools/tile_selector.html.

possibility to process SITS with irregular temporal sampling. I presented this issues in 1.3, it is illustrated here in figure 2.15 with Sentinel-2 data.

In this work, we wanted to take benefit of the functional modeling using a parametric model, in order to scale the model to large data set. We choose to use Gaussian process (GP).

Contribution 8:

During the PhD work of A. Constantin, we have proposed a multivariate Gaussian process based on the linear combination of independent GPs [CFG22]. We use functional basis and parametric covariance/kernel function to model the temporal behavior. This model allows to perform Bayesian classification and imputation of irregular time series and scale linearly with the number of training samples. This is a special case of Linear model of Coregionalization [Goo97].

We have made use of the Matrix-Variate Gaussian Distribution, defined on random matrices:

Definition 3. *The matrix-variate normal distribution $\mathcal{MN}_{p,q}$ [Daw81] is defined for all $p \times q$ random matrices $\mathbf{Y}^\star$ as:*

$$\mathbf{Y}^\star \sim \mathcal{MN}_{p,q}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Lambda}) \text{ if and only if } \mathrm{vec}(\mathbf{Y}^\star) \sim \mathcal{N}_{pq}(\mathrm{vec}(\mathbf{M}), \mathbf{\Sigma} \otimes \mathbf{\Lambda}),$$

where $\mathbf{M}$ is a $p \times q$ matrix, $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ are symmetric positive definite matrices of size $q \times q$ and $p \times p$ respectively.

We used the following notation:

- If $\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$ then $\mathrm{vec}(\mathbf{M}) = [m_{11}, m_{21}, m_{12}, m_{22}]^\top$.
- Kronecker product: $\mathbf{\Sigma} \otimes \mathbf{\Lambda} = \begin{bmatrix} \sigma_{11}\mathbf{\Lambda} & \cdots & \sigma_{1q}\mathbf{\Lambda} \\ \vdots & \ddots & \vdots \\ \sigma_{q1}\mathbf{\Lambda} & \cdots & \sigma_{qq}\mathbf{\Lambda} \end{bmatrix}$.

Under such assumptions, the PDF can be written as

$$p(\mathbf{Y}^\star) = (2\pi)^{-pq/2} |\mathbf{\Sigma}|^{-p/2} |\mathbf{\Lambda}|^{-q/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left[\mathbf{\Lambda}^{-1}(\mathbf{Y}^\star - \mathbf{M})\mathbf{\Sigma}^{-1}(\mathbf{Y}^\star - \mathbf{M})^\top\right]\right.$$

Such distribution is well suited to model structured data, and structured covariance matrix, that can take the form of a matrix measurement. The above definition imposes that the correlation between rows does not change w.r.t. lines of the matrix, and vice-versa. With SITS, we have a temporal multispectral measurements for each observed pixel $\mathbf{Y}_i^\star$:

$$\mathbf{Y}_i^\star = \begin{bmatrix} y_{i1}(t_1) & \dots & y_{i1}(t_q) \\ \vdots & \ddots & \\ y_{ip}(t_1) & \dots & y_{ip}(t_q) \end{bmatrix} \tag{2.29}$$

where $y_{ib}(t_l)$ is the observed reflectance for band b at time t_l. Using a matrix-variate normal model amounts to assume that the correlation between spectral bands does not change during

the time or season, and that the temporal correlation is the same whatever the spectral bands considered.

Concretely, this modeling reduces the number of parameters to estimate and the computational complexity (matrix inversion and determinant) since the effective degree of freedom of the covariance matrix is reduced from $\frac{(pq)(pq+1)}{2}$ to $\frac{(p(p+1)+q(q+1))}{2}$, and we can take advantage of the Kronecker product.

Yet we cannot used this modeling directly because the size (number of temporal measurements) varies according to the pixels, i.e., q depends on the pixel with irregularly sampled pixels. In [CFG22], we have extended this matrix-variate model to multivariate Gaussian process to release the dependence of a common temporal grid. The idea was to model the process rather than the observations themselves (even if the parameters of the process will be learned with the observations).

Definition 4. *We define a Multivariate Gaussian Processes (MGP), conditionally to $Z = c$, as*

$$\mathbf{Y}(t) = \mathbf{A}_c \mathbf{W}_c(t) + \mathbf{m}_c(t)$$

with $\mathbf{Y}(t) \in \mathbb{R}^p$, $\mathbf{W}_c(t) \in \mathbb{R}^p$, $\mathbf{m}_c(t) \in \mathbb{R}^p$, $\mathbf{A}_c \in \mathbb{R}^{p\times p}$ and

$$\begin{cases} W_{cb} | \ Z = c \sim \mathcal{GP}(0, K_c), \forall \ b \in \{1, \dots, p\}, \\ W_{cb} \perp W_{cb'}, \forall \ b, b' \in \{1, \dots, p\}^2. \end{cases}$$

We write $\mathbf{Y}|Z = c \sim \mathcal{MGP}_p(\mathbf{m}_c, K_c, \mathbf{A}_c)$.

Then we assume that our observed pixel is a multivariate stochastic process which can be modelled as a mixture of MGP:

Definition 5. Mixture of MGP *(M2GP) is defined as*

$$\mathbf{Y} \sim \sum_{c=1}^{C} \mathbb{P}(Z = c) \mathcal{MGP}_p(\mathbf{m}_c, K_c, \mathbf{A}_c).$$

Using definitions 3, 4 and 5 the following proposition holds:

Proposition 1. *If $\mathbf{Y} \sim$ M2GP then $\mathbf{Y}^\star = [\mathbf{Y}(t_1), \dots, \mathbf{Y}(t_q)]$ is a $p \times q$ random matrix such as:*

$$\mathbf{Y}^\star | \ Z = c \sim \mathcal{MN}_{p,q}(\mathbf{M}_c, \mathbf{\Sigma}_c, \mathbf{A}_c \mathbf{A}_c^\top),$$

with

- $\mathbf{M}_c = (\mathbf{m}_c(t_1), \dots, \mathbf{m}_c(t_q))$,
- $\mathbf{\Sigma}_c$ *is the covariance matrix defined by* $\mathbf{\Sigma}_{c_{l,l'}} = K_c(t_l, t_{l'})$ *for all* $(l, l') \in \{1, \dots, q\}^2$.

Equivalently: $\text{vec}(\mathbf{Y}^\star)|Z = c \sim \mathcal{N}_{pq}(\text{vec}(\mathbf{M}_c), \mathbf{\Sigma}_c \otimes \mathbf{A}_c \mathbf{A}_c^\top)$.

We use parametric functions of the time for the mean and covariance function (full proof in [CFG22]):

- Parametric mean: $\mathbf{m}_c$ is modeled by a weighted sum of J basis functions $\{\varphi_j\}_{j=1}^J$: $\mathbf{m}_{c,b}(t) = \sum_{j=1}^J \alpha_{c,b,j}\varphi_j(t)$. Let $\mathbf{B}$ be the design matrix of size $J \times q$ and $\boldsymbol{\alpha}_c = [\boldsymbol{\alpha}_{c,1}, \ldots, \boldsymbol{\alpha}_{c,p}]^\top$ a matrix of size $p \times J$ then we have
$$\mathbf{M}_c = \boldsymbol{\alpha}_c \mathbf{B}.$$
- Parametric kernel: K is chosen in a set of parametric kernel indexed by $\boldsymbol{\theta}$: $\boldsymbol{\Sigma}_c = \boldsymbol{\Sigma}(\boldsymbol{\theta}_c)$. The Gaussian kernel defined eq. (2.6) was used in this work.

Using the proposition 1 and such parametrization, we end up with the following conditional PDF, where parameters to be optimized are independent of the sample time:

$$\mathbf{Y}_i^\star | Z_i = c \sim \mathcal{MN}_{p,q_i}\left(\boldsymbol{\alpha}_c \mathbf{B}_i, \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_c), \mathbf{A}_c \mathbf{A}_c^\top\right)$$

where colored letters refer to learnable parameters, and $\mathbf{B}_i$, $\boldsymbol{\Sigma}_i$ are time/pixel dependent. The resulting negative log-likelihood (NLL) is

$$\begin{aligned}\ell\left(\boldsymbol{\alpha}_c, \boldsymbol{\theta}_c, \mathbf{A}_c\mathbf{A}_c^\top\right) =& p \sum_{i|Z_i=c} \log|\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_c)| + n_c \log|\mathbf{A}_c\mathbf{A}_c^\top| \\ &+ \operatorname{tr}\left[\sum_{i|Z_i=c} \left(\mathbf{Y}_i^\star - \boldsymbol{\alpha}_c\mathbf{B}_i\right)\left\{\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_c)\right\}^{-1}\left(\mathbf{Y}_i^\star - \boldsymbol{\alpha}_c\mathbf{B}_i\right)^\top \left\{\mathbf{A}_c\mathbf{A}_c^\top\right\}^{-1}\right] \\ &+ \text{constant}\end{aligned}$$

which can be minimized by solving C independent minimization problems. The complexity is linear w.r.t. the number of pixels, and cubic w.r.t. the number of time sample (usually low). A conventional constrained gradient descent was used to minimize the NLL.

This NLL is similar to the one of a conventional vectorial Gaussian mixture model, but it differs in two points:

1. There is no need to have samples of the same size,
2. The estimation of the covariance matrix is regularized by the Kronecker structure.

This model is able to learn from irregularly sample pixel and to infer the class of a given pixel whatever its sample time, using MAP rule. A by-product of the model is its ability to impute missing values using Bayesian estimate: for a given sample $\mathbf{Y}_i^\star$ observed at $\{t_1, \ldots, t_{q_i}\} = T_i$ and an unobserved time t^*, the imputation is performed as a weighted imputation of each class-Gaussian conditional imputation:

$$\mathbf{Y}_i^\star(t^*) \sim \sum_{c=1}^C \mathbb{P}(Z = c|\mathbf{Y}_i^\star)\mathcal{N}_p\left(\boldsymbol{\mu}_c(t^*, \mathbf{Y}_i^\star), \boldsymbol{\Lambda}_c(t^*, T_i)\right)$$

where the class membership is estimated with the posterior of the model and $\mathcal{N}_p$ is the conditional-distribution of the pixel at time t^* given the observation of the pixel at times $\{t_1, \ldots, t_{q_i}\} = T_i$:

- $\boldsymbol{\mu}_c(t^*, \mathbf{Y}_i^\star) = \hat{\boldsymbol{\alpha}}_c \mathbf{b}^* + \left(\mathbf{Y}_i^\star - \hat{\boldsymbol{\alpha}}_c \mathbf{B}_i\right)\left\{\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\theta}}_c)\right\}^{-1}\mathbf{k}(t^*, T_i|\hat{\boldsymbol{\theta}}_c)$,
- $\boldsymbol{\Lambda}_c(t^*, T_i) = \left[K(t^*, t^*|\hat{\boldsymbol{\theta}}_c) - \mathbf{k}(t^*, T_i|\hat{\boldsymbol{\theta}}_c)^\top \left\{\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\theta}}_c)\right\}^{-1}\mathbf{k}(t^*, T_i|\hat{\boldsymbol{\theta}}_c)\right] \otimes \widehat{\mathbf{A}_c\mathbf{A}_c^\top}$,

with $\mathbf{b}^* = [\varphi_1(t^*), \ldots, \varphi_J(t^*)]$ and $\mathbf{k}(t^*, T_i|\hat{\boldsymbol{\theta}}_c) = [K(t^*, t_1|\hat{\boldsymbol{\theta}}_c), \ldots, K(t^*, t_{T_j}|\hat{\boldsymbol{\theta}}_c)]^\top$.

We have performed simulation with about 190,000 learning pixels and compared the classification accuracy with standard vector models (with a proper linear temporal interpolation to have samples on the same temporal grid). Ten independent runs were done and averaged results are reported in 2.8 (and full results in [CFG22]).

Table 2.8.: Classification accuracy for the classification of S2 SITS. QDA is the Gaussian mixture model with a ridge regularization (or quadratic discriminant analysis), HDDA is a QDA model with a sparse covariance matrix [BGS07], lin-SVM is a linear SVM with a specific solver for large scale data (liblinear, https://www.csie.ntu.edu.tw/~cjlin/liblinear/) and RF is random forest.

	Average F1	std F1
M2GP	63.1	1.15
HDDA	55.6	2.70
QDA	64.2	1.36
lin-SVM	69.4	1.78
RF	74.2	1.78

Non-parametric methods (RF and SVM) provide the best classification results in terms of F1-score. Interestingly, the covariance sparsity assumption of HDDA results in worse results in terms of classification accuracy than QDA. The Kronecker structure and the irregular sampling of M2GP lead to an accuracy similar to QDA but lower than non-parametric techniques. We thus believe that the M2GP covariance structure is well suited to deal with irregularly sampled SITS, as classification accuracy are closed to QDA ones, but the M2GP model is not flexible enough to reach the performance of non-parametric methods. The Gaussian (and thus unimodal) assumption for the class conditional density might be too limiting and should be relaxed in future works (see perspectives in section 3).

Figure 2.16 shows a reconstruction of a pixel time series in the visible and near infra red Sentinel-2 bands. The third date was cloudy but not tagged as such from the data provider: the linear interpolation is highly impacted by the outliers, while the proposed method is not affected. Furthermore, it is possible to draw confidence interval for the imputation.

In [CFG22], a detailed analysis of the model outputs (temporal and spectral correlation, mean function) is given: the interested reader could consult the article.

2.2.3. Joint representation learning and processing

Another conventional strategy to circumvent issues related to high-dimensional data is to perform feature reduction or feature extraction. This strategy is termed "Representation learning" [BCV12] in the machine learning literature. However, such techniques are usually performed independently from the final objective. For instance, PCA maximizes the projected variance, but maximizing the variance is not necessarily related to classification accuracy. Similarly, spectral unmixing techniques allows to decompose a pixel into pure components that provide meaningful physical information but this is usually not a representation that perform well in the context of classification.

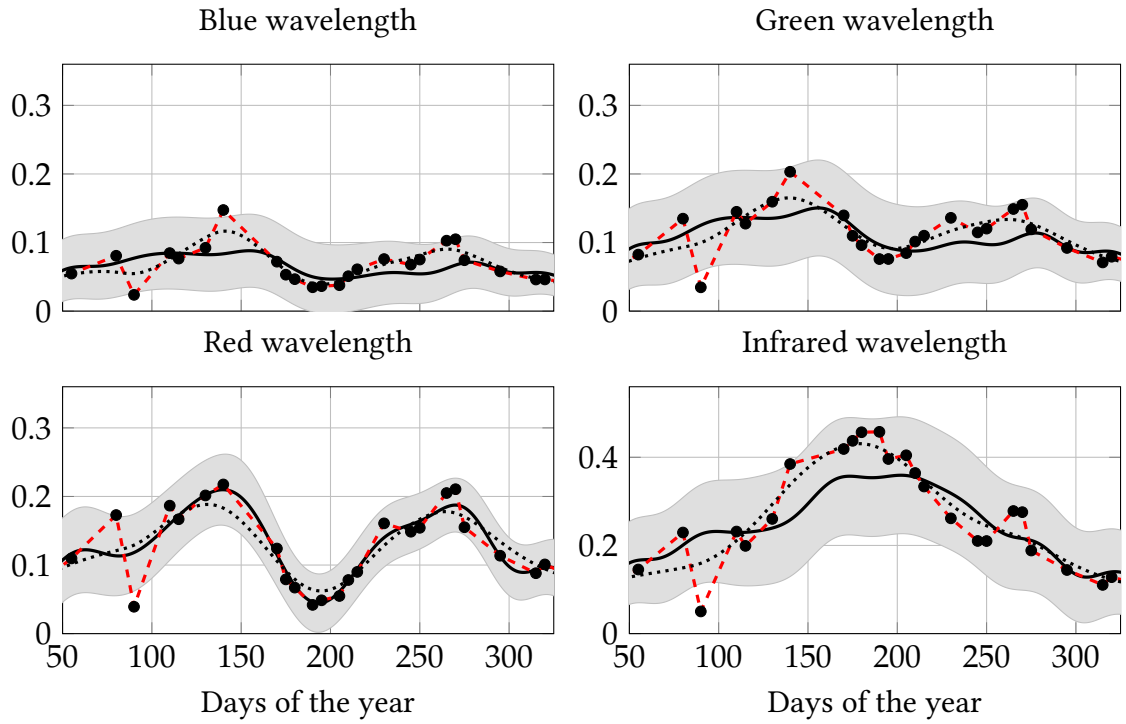


Figure 2.16.: Time series Reconstruction for three methods: linear interpolation (red dashed line), Whittaker smoothing (black dotted line) and Gaussian process in black (gray region corresponds to 95% confidence interval). Black dot points are the original values from the irregularly sampled time-series

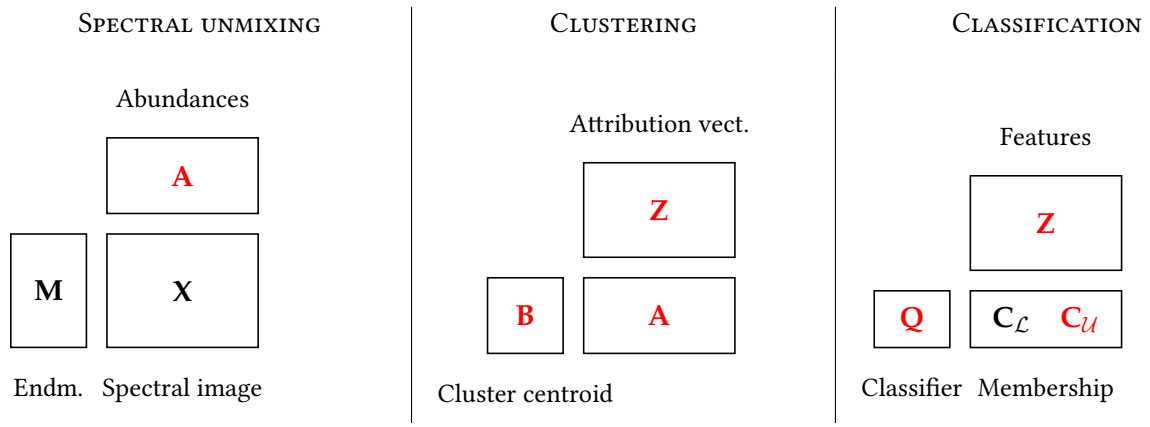


Figure 2.17.: Joint spectral unmixing and classification.

For this reason, we have investigated the definition of algorithms that perform jointly *representation learning* and *processing task* (e.g., classification). This work was done during the PhD thesis of Adrien Lagrange, co-supervised with Prof. Nicolas Dobigeon.

Contribution 9:

Our contributions were the development of unified frameworks that perform jointly *representation learning* and *classification* tasks. A coupling term, based on a clustering task, link the two adversarial objectives during the optimization process. In order to learn the parameters of the model, two strategies were investigated: the first one was based on Bayesian modeling [LFMD19]; the second one was based on the PALM optimization algorithm [LFM+20]. The latter scales better w.r.t number of available samples while the former is more robust w.r.t error in labels.

In this manuscript, I choose to report results from [LFM+20]. The model proposed is described in the figure 2.17. This model performs jointly spectral unmixing and classification. For the unmixing, the endmember $\mathbf{M}$ samples are supposed to be known while for the classification some samples are available with their corresponding ground-truth. These two operations are combined trough the clustering step that links the estimated abundances $\mathbf{A}$ to the membership $\mathbf{C}$. Variables in red in the figure are the free parameters of the model, that are learned with the data.

Individually, the different steps are

- Sparse linear mixture model on the original spectra: The observations are represented by a sparse linear combination of elementary signatures [BDF10]

$$\min_{\mathbf{A}\succeq 0} \|\mathbf{X} - \mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2 + \lambda_a \|\mathbf{A}\|_1 \tag{2.30}$$

  with $\mathbf{M} \in \mathbb{R}^{d\times n}$: endmember matrix (spectra of elementary components), $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_P] \in \mathbb{R}^{n\times P}$: abundance matrix, n the number of pixels of the image and d the number of spectral components.

- Relaxed k-means on the abundance matrix $\mathbf{A}$:

$$\min_{\mathbf{Z}\in\mathbb{S}_K^p,\mathbf{B}\succeq 0} \|\mathbf{A}-\mathbf{B}\mathbf{Z}\|_F^2\,, \tag{2.31}$$

  where the matrix $\mathbf{B}=[\mathbf{b}_1\dots\mathbf{b}_K]\in\mathbb{R}^{R\times K}$ contains the centroid of clusters. The constrain $\mathbf{Z}\succeq 0$ and sum-to-one constraints on $\mathbf{Z}$ are used to interpret $\mathbf{z}_p$ as probability vectors while $\mathbf{B}\succeq 0$ are enforced to interpret the rows as averaged abundances.

- Semi-supervised linear classification on the feature vector $\mathbf{Z}$: We use the cross-entropy loss function and TV-norm spatial regularizer (the quadratic loss was also used in [LFM+20]).

$$\min_{\mathbf{Q},\mathbf{C}_{\mathcal{U}}\in\mathbb{S}_C^{|\mathcal{U}|}} -\lambda_1\sum_{p\in\mathcal{P}}\sum_{i\in\mathcal{C}} c_{i,p}\log\left(\operatorname{sigm}(\mathbf{q}_{i:}\mathbf{z}_p)\right)+\lambda_q\|\mathbf{Q}\|_{\mathrm{F}}^2+\lambda_c\|\mathbf{C}\|_{\mathrm{vTV}}\,,$$

  where $\mathbf{q}_{i:}$ denotes the i$^{\text{th}}$ row of $\mathbf{Q}$.

Combining all these optimization problem leads to the following co-factorization problem

$$\begin{aligned}\min_{\substack{\mathbf{A}\succeq 0,\mathbf{Q},\mathbf{Z}\in\mathbb{S}_K^p\\ \mathbf{C}_{\mathcal{U}}\in\mathbb{S}_C^{|\mathcal{U}|},\mathbf{B}\succeq 0}} \quad & \lambda_0\|\mathbf{Y}-\mathbf{M}\mathbf{A}\|_{\mathrm{F}}^2+\lambda_a\|\mathbf{A}\|_1\\ & -\lambda_1\sum_{p\in\mathcal{P}} d_p\sum_{i\in\mathcal{C}} c_{i,p}\log\left(\operatorname{sigm}(\mathbf{q}_{i:}\mathbf{z}_p)\right)+\lambda_q\|\mathbf{Q}\|_{\mathrm{F}}^2+\lambda_c\|\mathbf{C}\|_{\mathrm{vTV}}\\ & +\lambda_2\|\mathbf{A}-\mathbf{B}\mathbf{Z}\|_F^2\end{aligned} \tag{2.32}$$

which is a non-convex, non-smooth optimization problem. We have proposed to solve it using proximal alternating linearized minimization (PALM [BST14]).

We compared the proposed algorithm with quadratic (Q) and cross-entropy (CE) classification losses, denoted respectively by Cofact-Q and Cofact-CE, to state-of-the-art classification and unmixing methods. First, one considered competing method was the random forest (RF) classifier, which has been extensively used for the hyperspectral image classification. Additionally, the performance of the classification method proposed in [UMH19] has been evaluated. This method, referred to as SSFPCA+SVM, relies on a so-called spectrally-segmented folded PCA (SSFPCA) as a feature extraction step, followed by a RBF-kernel SVM classifier. The parameters of SSFPCA have been set based on the study provided in the original paper. Parameters of the RF and the SVM have been adjusted using cross-validation with a grid-search strategy and we used the implementations provided in the *scikit-learn* Python library [PVG+11]. Then, two unmixing methods proposed in [BDF10] has been tested, namely the fully constrained least squares (FC-SUnSAL) and the constrained sparse regression (CSR-SUnSAL). FC-SUnSAL basically relies on the same data fitting term considered in the proposed co-factorization method, under non-negativity and sum-to-one constraints applied to the abundance vectors. Conversely, the CSR-SUnSAL problem removes the sum-to-one constraint and introduces a ℓ_1-norm penalization on the abundance vectors. It thus solves (2.30) where the associated regularization parameter λ_h is tuned using a grid-search strategy. These two methods use an augmented Lagrangian splitting algorithm to recover the abundance vectors. Additionally, these abundance vectors are subsequently used as input features of a multinomial logistic regression classifier (MLR). This classifier is linear and its combination with the CSR-SUnSAL

Table 2.9.: AISA data: unmixing and classification results.

Model	F1-mean	Kappa	RE	Time (s)
RF	0.912 ($\pm 4.2 \times 10^{-3}$)	0.835 ($\pm 1.2 \times 10^{-2}$)	N\A	41 (± 1)
SSFPCA+SVM	0.464 ($\pm 4.5 \times 10^{-3}$)	0.466 ($\pm 1.9 \times 10^{-2}$)	N\A	398 (± 12)
FC-SUnSAL+MLR	0.428 ($\pm 3.9 \times 10^{-2}$)	0.433 ($\pm 3.8 \times 10^{-2}$)	0.298 ($\pm 1.9 \times 10^{-3}$)	512 (± 96)
CSR-SUnSAL+MLR	0.610 ($\pm 6.9 \times 10^{-2}$)	0.618 ($\pm 8.0 \times 10^{-2}$)	0.304 ($\pm 2.0 \times 10^{-5}$)	529 (± 61)
D-KSVD	0.396 ($\pm 9.5 \times 10^{-2}$)	0.406 ($\pm 9.9 \times 10^{-2}$)	0.303 ($\pm 7.6 \times 10^{-6}$)	10475 (± 129)
LC-KSVD	0.600 ($\pm 4.0 \times 10^{-2}$)	0.594 ($\pm 3.0 \times 10^{-2}$)	0.303 ($\pm 4.0 \times 10^{-6}$)	3780 (± 320)
Cofact-Q	0.824 ($\pm 2.0 \times 10^{-2}$)	0.652 ($\pm 2.5 \times 10^{-2}$)	0.310 ($\pm 1.6 \times 10^{-4}$)	7303 (± 139)
Cofact-CE	0.876 ($\pm 1.6 \times 10^{-2}$)	0.759 ($\pm 3.5 \times 10^{-2}$)	0.310 ($\pm 1.4 \times 10^{-4}$)	4382 (± 257)

unmixing algorithm, referred to as CSR-SUnSAL+MLR, yields a sequential counterpart of the proposed Cofact-CE method. In particular, comparing the resulting classification performance with the performance of Cofact-CE allows the benefit of introducing the clustering coupling term to be assessed.

Besides, the proposed method has been also compared with the discriminative K-SVD (D-KSVD) method proposed in [ZL10]. The D-KSVD problem has strong similarities with the proposed co-factorization problem. Indeed, it corresponds to a ℓ_0-penalized representation learning and a classification with a quadratic loss. It aims at learning a dictionary suitable for the classification problem and performs a linear classification on the coding vectors. For this reason, the dictionary is only used as an initialization for D-KSVD, while it remains fixed for the unmixing and proposed co-factorization methods. Similarly, the label consistent K-SVD (LC-KSVD) is also considered [JLD11]. This model has been proposed as an improvement of D-KSVD where an additional term ensures that the dictionary elements are class-specific. Hyperparameters of D-KSVD and LC-KSVD have been manually adjusted in order to get the best results.

When implementing the PALM algorithm, the normalized regularization parameters in (2.32) have been fixed as $\tilde{\lambda}_0 = 100, \lambda_1 = \lambda_2 = 1, \lambda_h = \lambda_q = 0.1$ and $\tilde{\lambda}_c = 10^{-3}$. Finally, the number of clusters has been set to $K = 10$. The influence of these parameters are empirically studied in the associated companion report [LFM+19]. Simulation have been performed on the Aisa data set. Results are given in Table 2.9 and metrics and their standard deviation have been computed over random 5 trials.

Reconstruction errors (RE) is the only figure-of-merit to assess the quality of the representation learning. REs appear to be very similar for all algorithms. Concerning classification results, the results reported in Table 2.9 show that the classification maps recovered by the Cofact-CE is very closed to the one obtained by RF, whereas others methods fails to provide reasonable results. The comparison between the representation learning-based algorithms is clear and the both Cofact methods perform better than LC-KSVD and D-KSVD.

One of very interesting feature of the Cofact method is the possibility of examining the clusters obtained as a byproduct. The centroids $\mathbf{B}$ estimated by the Cofact method can be interpreted as average behaviors of abundance vectors. Corresponding virtual spectral signatures can be obtained by right-multiplying the dictionary $\mathbf{M}$ by this estimated abundance-like matrix $\mathbf{B}$. The first line in Figure 2.18 shows these spectral centroids for each cluster. Accessing this kind of information is precious in term of image interpretation since it offers the possibility of visualizing any class multi-modality. To illustrate, the second line of Figure 2.18 shows the mean spectra associated with the subclass ground-truth. Clearly, both lines exhibit strong sim-

ilarities, with spectral diversity (hence multi-modality) for the 1st, 3rd and 4rd classes. This illustrates the relevance of the clusters recovered by the proposed co-factorization method.

This work has been pursued in the PhD work of. E. Giry-Fouquet. See for instance [GFM21].

2.3. Thematic application

My initial area of interest was urban areas. As objects of interest in these areas have typical shape and size that depend on there use rather than they physical nature, spatial-spectral methods were especially useful. I was interested to analyse every urban object, with no specific attention to particular ones. I mainly proposed anisotropic methods such as ones discussed in 2.1. Then, after my recruitment at Dynafor-lab, I focused on particular vegetation objects, that are of interest in landscape ecology. In this report, I choose to discuss about three elements: Hedgerows, Trees species and Grasslands. I deliberately skip urban tree which is something I am not working on anymore. Reader that are interested can take a look at [AFZ+19, ADZ+18].

2.3.1. Hedgerows detection

The first one was the hedgerows. These semi-natural elements are essential in the agricultural landscapes. They are the main components of the green infrastructure plans defined for conservation planning from the local to the global (statewide) scale [WRWV10]. For instance in France, one conclusion of the "environment round table" is the need of a cartography of the "green belt network" (or green infrastructures) [ERT]. Another example is given by the European Union which conditions its funding to farmers based on how they maintain hedges inside their farms [AAW10].

In the past, small and linear woody elements were mapped manually, by visual interpretation of aerial photographs, or by traditional field-based recognition. Obviously, this approach was highly time consuming and not appropriate for large areas. Besides, the small green veining objects were not always included in the spatial databases produced by the National Mapping Agencies, or not distinguished by category [THL11]. In addition, when this is the case, their cartographic representation does not necessarily fit the user needs because green veining objects like hedgerows are very application dependent [BBB00]. For instance, the useful minimum length to define a linear woody object as a hedge may differ according to the species studied and its dispersion ability.

Contribution 10:

My contribution was the use of local orientation feature for the detection of hedgerows in very high resolution images [FAB+13]. This feature was computed using advanced orientation opening algorithm.

In order to map hedgerows from very high resolution images, I propose to apply *directional morphological filters* [Soi03, ST01] for the detection of hedges. Directional filters allow the processing of an image in a defined direction. Hence, it is possible to preserve or remove objects according to their orientations and also enable the discrimination of objects with different orientations in the image from objects with no main orientation. Similarly to the Morphological Profile, we compute several Morphological Directional Profile (MDP) with various orientation

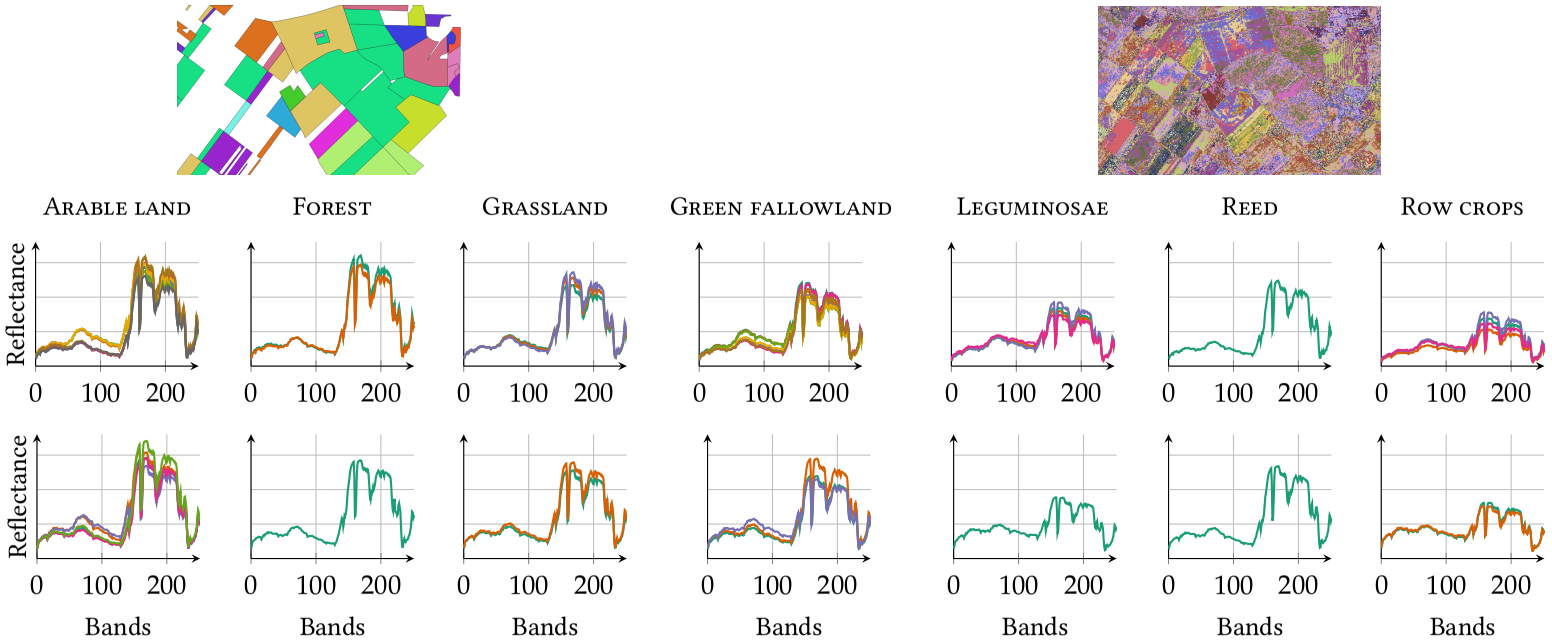


Figure 2.18.: AISA data: (1st row) Ground-truth map of sub-classes and clustering recovered by Cofact-CE, (2nd row) for each class, spectral centroids of the clusters recovered by Cofact-CE composing the class, (3rd row) for each class, mean spectra of the ground-truth sub-classes composing the class.

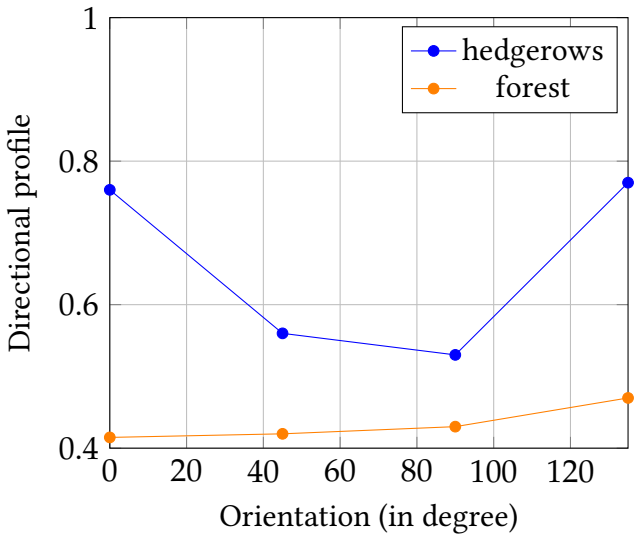


Figure 2.19.: Directional morphological profile for (a) a pixel that belongs to an hedgerows and (b) a pixel that belongs to a forest. Hedgerows pixel has a local orientation feature of 0.24 and forest pixel of 0.05.

parameter and we define the local orientation $\mathrm{LO}(\mathbf{x})$ is defined as the difference between the maximum value and the minimum value of the MDP:

$$\mathrm{LO}(\mathbf{x}) = \max_{\theta} \big(\mathrm{MDP}(\mathbf{x})\big) - \min_{\theta} \big(\mathrm{MDP}(\mathbf{x})\big). \tag{2.33}$$

This property is very important for the discrimination of the hedges from the other woody structures. Indeed, when hedges are observed at the meter scale, they have a specific orientation in the spatial space while forest area not. See for instance Figure 2.19.

I investigated such techniques in [FAB+13], coupled with a detection algorithm. Some hedgerows mapping are provided in Figures 2.20.

2.3.2. Tree species classification

Forest ecosystems provide essential services to human society. Beyond the production of multiple resources (timber, energy, foods), these ecosystems play a major role in carbon sequestration, regulating biogeochemical cycles and climate. Mapping tree species is therefore crucial to assess forest ecosystems and their services. Several works have been done using remote sensing to map tree species: using very high resolution images [LPT+15], airborne hyperspectral images [FA13] or multi-temporal images [ZL14].

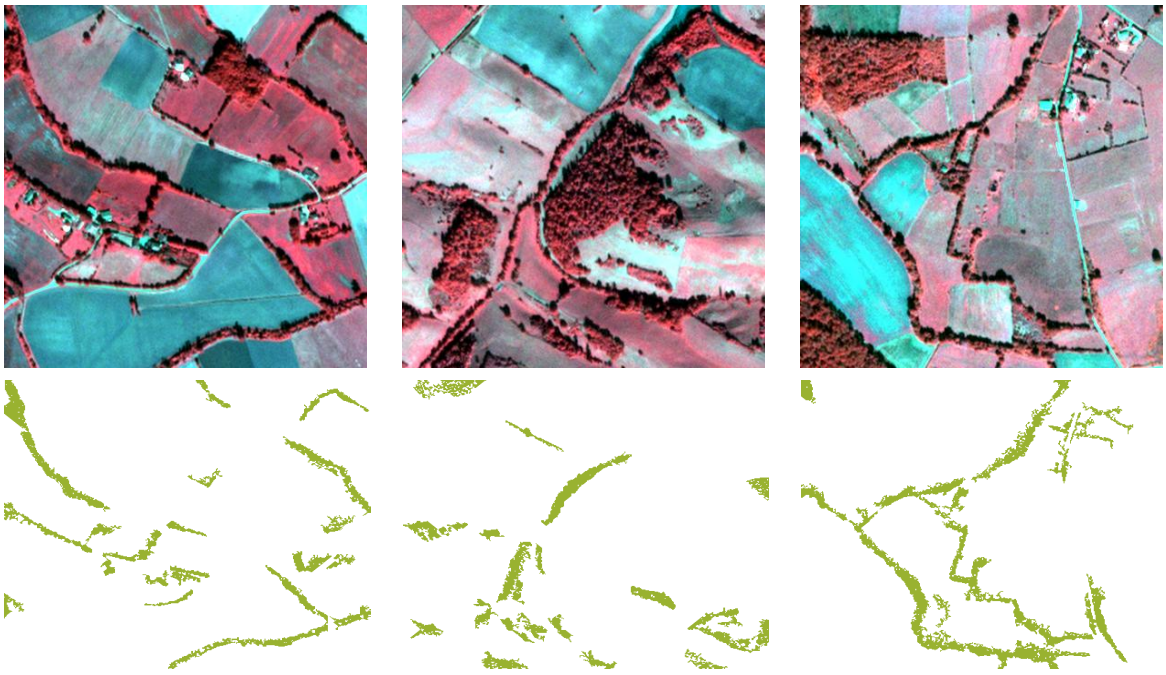


Figure 2.20.: Hedgerows detection.

Contribution 11:

We explore the benefits of using a very dense high spatial resolution multispectral Formosat-2 image time-series for mapping 13 tree species in temperate forests, for 9 consecutive years, with state of the art machine learning techniques to pre-process the time series, extract relevant dates and spectral features and classify the pixels [SFJ+16, KDF+19]. We have enlightened the importance of spatial auto-correlation for the validation of the classification accuracy.

Using a Formosat dense time series, we classify the tree species each year. In [SFJ+16], we have reported very high classification accuracy, but we had suspected some overconfidence in our discussion. To assess the robustness of the results, we further compared the intra-annual stability, i.e., does the classifier predict the same class for several consecutive years (it should !). Figure 2.21 shows that strong instability, which is something that cannot happen with tree species.

As a consequence, in [KDF+19] we took care of the spatial auto-correlation to evaluate the classification accuracy. Based on the Moran Index [Cre15], a spatial cross-validation was used to build spatially independent folds that were used to estimation the accuracy of the classification in terms of accuracy. It reveals a strong positive bias in validation based on the usual LOO-CV strategy for splitting reference data, as shown in table 2.10.

These results show the importance of accounting for the spatial auto-correlation when evaluating the accuracy in remote sensing. Surprisingly, few studies take this point into consideration although it is critical in particular for small data set (such as the hyperspectral one: Pavia, AVIRIS and so on). This may explain the current trend of 99% percent of accuracy observed with highly parameterized models (such as CNNs) [ZLJ+19, CYXM20] ...

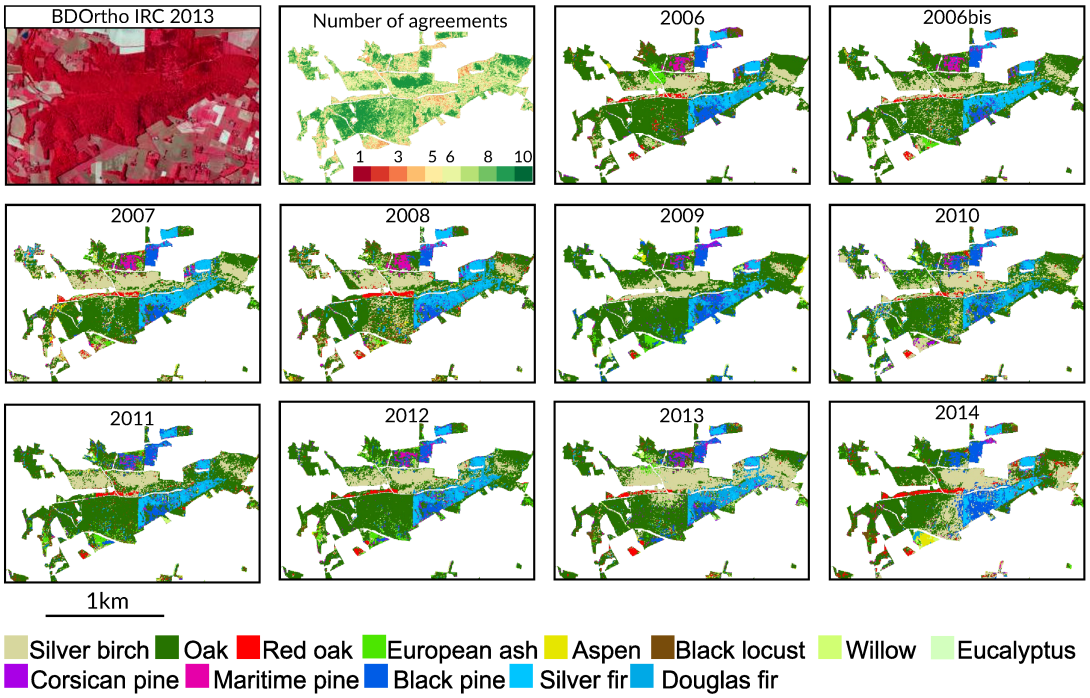


Figure 2.21.: Annual classifications of tree species in a mixed forest composed of different conifers and broad-leaf species. Instability was observed in the conifer plantations composed of Black pine, Silver fir and Douglas fir.

Table 2.10.: Accuracy report of spatial leave-one-out cross-validation (SLOO) sampling strategy and leave-one-out cross-validation (LOO) for each single-year classification based on OA statistics.

	2007	2008	2009	2010	2011	2012	2013	2014
S-LOO	0.48	0.57	0.55	0.56	0.55	0.58	0.60	0.58
LOO	0.99	0.98	0.99	0.98	0.97	0.98	0.99	1.00

2.3.3. Grasslands

Grasslands are today one of the most endangered ecosystems due to land use change, agricultural intensification, and abandonment [PBS05]. Grasslands host a unique biodiversity, which support the provision of key ecosystem services such as carbon storage, food production crop pollination, pest regulation and contribute to landscape scale amenities such as landscape scenic view. Yet, despite recent regulation in favor of grasslands, agriculture intensification, abandonment and urbanization generate a decrease of both grasslands surface area and plant diversity [O'M12, NHA+16] leading to a loss of biodiversity (plants, insects, birds) and its related ecosystem services. Therefore, there is an urgent need to monitor grasslands over large extents in order to assess their current state in terms of area covered but also to estimate their plant diversity and their carrying capacity for insect populations.

Contribution 12:

I start to work on grasslands during the PhD Thesis of Maïlys Lopes, co-supervised with Stéphane Girard. During her PhD, we work on the classification of management practices and on the estimation of biodiversity using dense satellite image time series [LOP17]. This work has been continued in [FLD+20] with a full year of radar and optical Sentinel-1&-2 observations and leads to the very first mapping of grasslands biodiversity indices at a decameter spatial resolution with remote sensing images.

In the following I am doing an overview of [FLD+20]. We used Sentinel's optical and radar SITS to estimate plant taxonomic and functional diversity in grasslands over a a large spatial extent (40 000 km^2) for one growing season. Fifteen months of acquisitions of Sentinel-1&-2 images were used to predict at the pixel level plant diversity indices (richness, Shannon and Simpson diversity indices) and functional diversity indices related to pollination (flower diversity, insect dependence) of grasslands located in farmed landscapes in the southwest of France. Various linear and non-linear algorithms were investigated to cover different kinds of possible statistical relationships between the biodiversity indices and the SITS.

Botanical surveys were conducted in 83 plots from distinct grasslands between April and May 2018 after the flowering and before the mowing. The average distance between two plots was 13.4 km with a standard deviation of 9.1 km and the minimum and the maximum distances were 0.08 km and 44.37 km, respectively. Each plot was composed of 5 quadrats of 1 square meter area. The center of each quadrat corresponded to the theoretical center of a Sentinel-2 pixel. The grasslands were digitized using the agricultural Land Parcel Information System "Registre Parcellaire Graphique" (RPG).

Plants were identified at the species level and the abundance of each species was estimated by its cover percentage value in the one square meter area, according to 12 bins, see Table 2.11. Vegetation height was measured in two locations in each quadrat and averaged. A total of 415 quadrats (leading to 415 Sentinel pixels), belonging to 83 grasslands, was recorded and used in this study.

Three indices were computed:

1. plant richness S_q, defined as the number of plant species in quadrat q,

Table 2.11.: Plant cover percentage bins.

Coverage estimation (%)	Assigned value (%)
[1, 2.5]	1
[2.5, 7.5]	5
[7.5, 11.5]	10
[11.5, 20]	15
[21, 30]	25
[31, 40]	35
[41, 50]	45
[51, 60]	55
[61, 70]	65
[71, 80]	75
[81, 90]	85
[91, 99]	95
100	100

2. the Shannon diversity index

$$H'(q) = -\sum_{s=1}^{S_q} p_{qs} \ln(p_{qs}),$$

where p_{qs} is the relative abundance of the species s in quadrat q,

3. Simpson index of diversity (refers to Simpson index in the paper for simplicity),

$$L(q) = 1 - \sum_{s=1}^{S_q} p_{qs}^2.$$

The Shannon index measures heterogeneity taking into account the number of species and the relative abundance of each species, while the Simpson index is the inverse of a dominance index reflecting how much the community is dominated (or not) by few species.

Four Sentinel-2 tiles of Level 2A were downloaded from the Theia Land Data Centre[3], corresponding to all available acquisitions between August 26, 2017 and December 03, 2018[4]. Only the bands natively at 10 m/pixel were used in the experiments (blue, green, red and near-infrared). All the acquisitions were re-sampled and gap-filled (to handle clouds and shadows) onto the same set of dates (every 10 days, starting from 2017-08-29 and ending in 2018-12-02), as proceeded in [IAT+15b]. The final SITS contains 47 dates, in the visible and near infrared bands, corresponding to a total of 188 spectro-temporal features. The normalized difference vegetation index (NDVI) and the first principal components, accounting for 99% of the total variance, were extracted as alternative features.

Sentinel-1 Level-1 GRD (Ground Range Detected) time series recorded in interferometric wide swath mode were used. Four orbits corresponding to the Sentinel-2 extent were downloaded from the French Sentinel collaborative ground segment[5], corresponding to the same

[3] http://www.theia-land.fr/en/presentation/products

[4] Images tagged as composed of more than 90% of cloudy pixels by the MAJA processing chain are not processed by the data center.

[5] https://peps.cnes.fr/rocket/\#/home

temporal period as Sentinel-2 time series. GRD time series were calibrated to γ^0, orthorectified and filtered using a multi-temporal spatial filter (https://gitlab.orfeo-toolbox.org/s1-tiling/s1tiling) [QY01]. The parameters of the spatio-temporal filter were a spatial window of size 5×5 pixels and all the dates were used for the temporal filters. VV and VH polarizations were processed as well as the ratio VV/VH. The features were re-sampled onto the same set of dates (every 10 dates starting from 2017-08-27 to 2018-11-30).

Figure 2.22 presents the reconstructed NDVI (in black) and γ^0(VV) polarization (in blue) for descending orbits of the five quadrats for four plots as well as their corresponding detected shadows and clouds.

The accuracy of the prediction of a biodiversity variable value y was assessed using the *coefficient of determination* (r^2) [DS98]:

$$r^2 = 1 - \frac{\sum_{i=1}^{n_v}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_v}(y_i - \bar{y})^2}$$

where n_v is the number of validation sample, y_i is the true biodiversity variable value associated with the i^{th} sample, $\hat{y}_i$ is its corresponding predicted value and $\bar{y}$ is the mean value of the validation sample.

In order to correctly estimate r^2, the *k-fold cross validation* was used. Cross-validation directly estimates the *generalization error* of a given method [HTF09, Chap. 7]: The data is split into K almost equal-size folds, and the model is fitted on $K-1$ folds. The prediction error, here r^2, is computed on the remaining unseen fold. This process is done for all folds $k = 1, 2, \ldots, K$. The average error over the K folds is the estimated generalization error.

However, in our experimental setting, the 5 quadrats corresponding to one plot are spatially correlated and this must be taken into account when constructing the K folds of the cross-validation [HV96]. It was proposed in this work to perform a grouped cross validation by splitting the samples according to their plot membership [PPNH17, LRPM+14]: quadrats from a same plot are all included in one fold and not shared among several folds. Hence, in the experiments, a five fold-grouped cross validation was performed, where the first four folds contain 16 plots (80 quadrats) and the last fold contains 19 plots (95 quadrats). The same splits were used for all experiments, i.e., the error assessment was done using the same folds whatever the methods, data sources and indices.

Best results for each variable corresponding to an average $\hat{r}^2 > 0.3$ are presented Table 2.12. All results can be found in [FLD+20] as well as a detailed discussion. In short, best results are obtained with Random Forest with raw Sentinel-2 features.

Prediction of the *Simpson* index at the pixel level was done for all the grassland parcels identified from the Land Parcel Information System (RPG) over the four Sentinel-2 tiles with the data and method that resulted in the highest prediction accuracy (i.e., R-IR satellite data and RF algorithm, see Table 2.12). The confidence interval was also computed. Figure 2.23 shows an extract of the predicted Simpson value in a square of 1 km^2 centered on the plot 10_470. It also shows a very high spatial resolution aerial image from 2016 over the same extent as well as the confidence interval associated with the prediction of the Simpson index obtained with the Random Forest. Extracts for all plots are provided in the supplementary materials of [FLD+20].

By using dense S2 and S1 SITS, this work contributed to the remote sensing of grasslands biodiversity using new generation high resolution SITS. This study indicates that the temporal

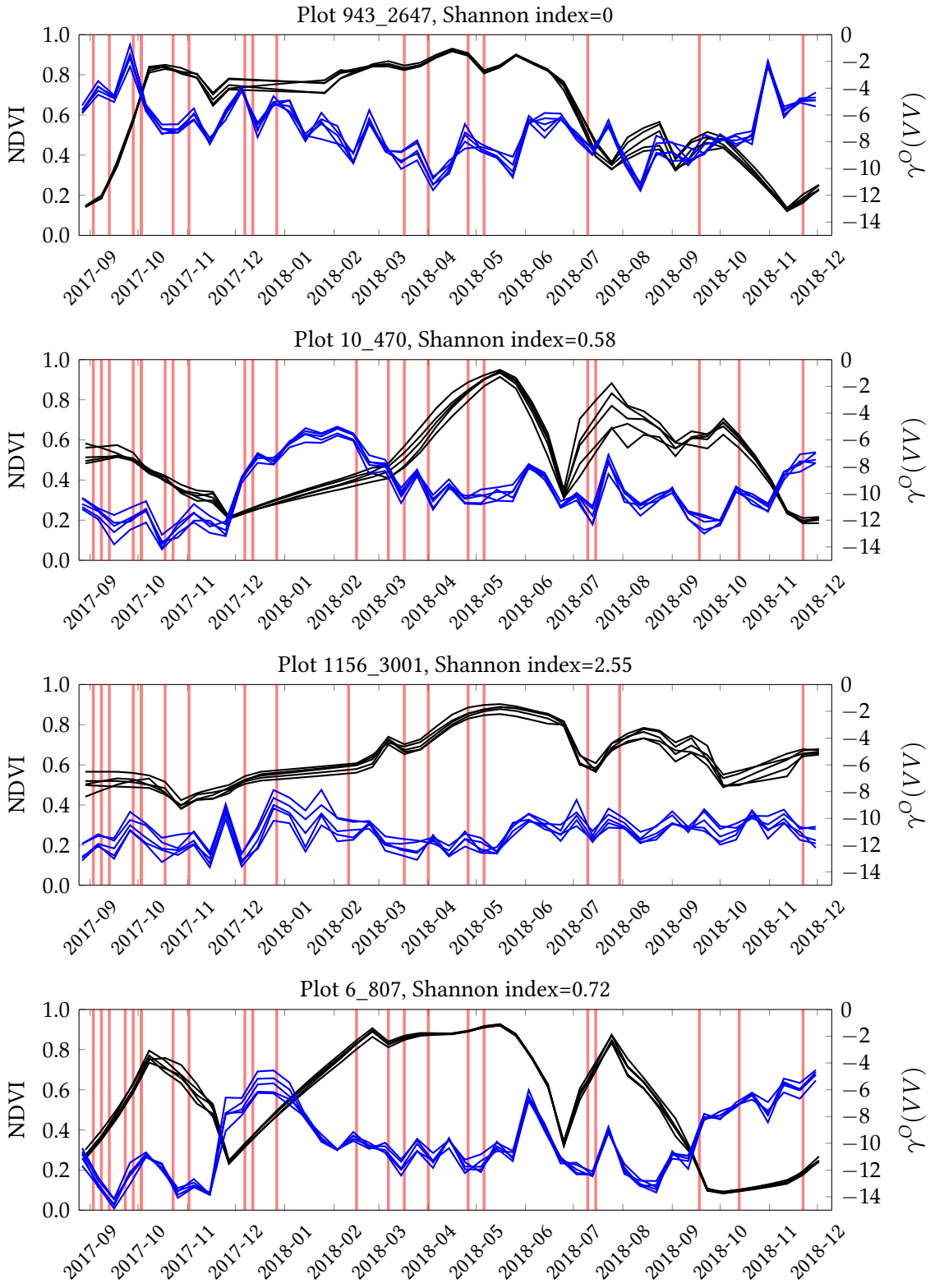


Figure 2.22.: Reconstructed S2-NDVI (in black) and S1-VV polarization (in blue) with a 10 days step for different plots. For each plot, the 5 pixels corresponding to each quadrat are plotted. The red vertical lines indicate the dates of the detected clouds and shadows for the raw S2 pixels, before the gap-filling.

Table 2.12.: Estimated accuracy for the best predicted variables ($\hat{r}^2 > 0.3$) and the corresponding standard deviation ($\hat{\sigma}_{r^2}$) of the 5-fold spatial cross validation.

Data	Variable	Method	$\hat{r}^2$	$\hat{\sigma}_{r^2}$
R IR	Simpson	RF	0.45	0.13
R IR	Shannon	RF	0.43	0.13
S2	Color Diversity	RF	0.40	0.08
S1 S2	Richness	KRidge	0.34	0.15
S1 S2	Insect Dependence	KRidge	0.32	0.15
S1 S2	Flowers Richness	KRidge	0.32	0.16
S1 S2	Reward Index	GP RBF	0.32	0.20

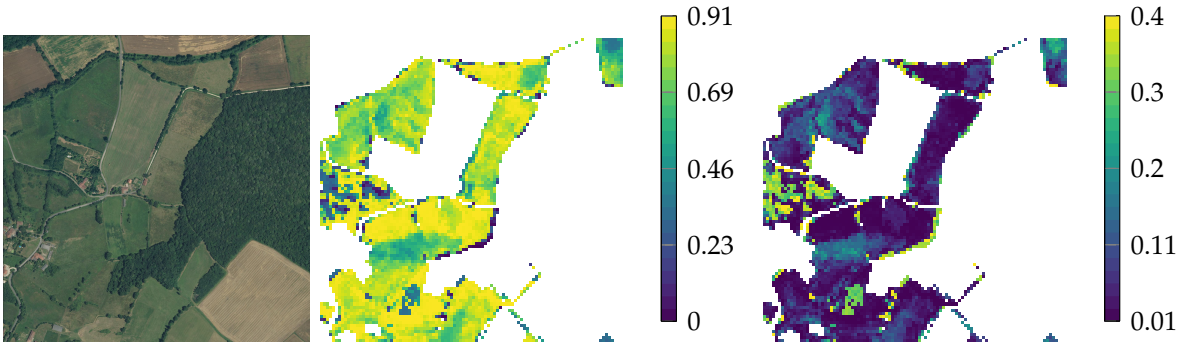


Figure 2.23.: Simpson index around plot 132_1699. (a) Very high spatial resolution aerial image from 2016, (b) Prediction of the Simpson value for pixels corresponding to grasslands in the national agricultural Land Parcel Information System and (c) Confidence interval value associated to the prediction.

information contained in the SITS can compensate for the limited spectral information and spatial resolution compared to hyperspectral imagery. Another finding is that the spatial autocorrelation tends to bias the estimation of the prediction accuracy and thus, should be taken into account during validation. The contribution of this work does not lie in selecting the best set of dates nor the best machine learning regression methods but in assessing the potential of Sentinel-1 and -2 dense times series as explanatory variables for the prediction of grasslands biodiversity indices. We believe that our results, using large scale data with various agricultural practices for different meteorological and topographic conditions, demonstrate the capacity of such data to monitor grasslands from an ecological viewpoint. In particular, intra-parcel variability was highlighted in this work and can be monitored over large areas.

3. Perspectives

My move to the CESBIO lab as an INRAe researcher has confronted my research to large scale data with their spatial non-stationarity and big data issues counterparts:

- The spatial non-stationarity was not significant for small extent coverage areas, such as all the hyperspectral data used in my past works. It implies that some of my results obtained may not generalize well and the observed improvements in terms of accuracy might be due to over-fitting over a small spatial area. This is indeed even more critical for recent deep models learn on small spatial extends, with millions of parameters to learn.
- The high data volume implies that some of the proposed complex methods, such as the spatial-spectral or the functional ones, don't scale to such massive data. Hence, they are unusable for most of real case studies and apart from theoretical consideration, they do not help other researchers to solve their problems.

In my opinion, the use of machine learning in remote sensing should face these two issues, which are essential components of the ecosystem monitoring problem. Using satellites to monitor small areas is overkill and in-situ or in field analysis is much more relevant. As such, machine learning should be used to monitor ecosystems at large scale to solve end-users application needs. In the context of climate change, remote sensing would be an important tool if we succeed to interact with environmental science.

In the following I am going to review perspectives of my research from two related points of view: thematic and methodological.

3.1. Thematic perspectives

My thematic perspectives are oriented to the production of maps of ecosystem variables, related tp functions and associated services[1], for a continuum over the time and the spatial domains. This is a mandatory condition to monitor states and trends of ecosystem. Being a member of the ECODIV INRAe scientific department (https://www.inrae.fr/en/divisions/ecodiv), my work will focus on grasslands, crops and forests ecosystems as well as their interfaces.

My main object of research is the definition of standard ecosystem variables (EV) that can be monitored and compared over large scales from Earth observation images. A primarily set of these ESV can be directly derived from the *Essential Biodiversity Variables* [PFW+13], such as land cover, *phenology*, or *habitat structure*. But other variables should be defined according to the extended information provided by recent satellite missions. For instance, *taxonomic diversity* as well as *species occurrence* can now be monitored from Earth observation data. A critical

[1] Ecosystem functions are related to the physical or biological process of an ecosystem (e.g., photosynthesis). Ecosystem services are defined as the benefits people obtain from ecosystems functions (e.g., carbon capture or biomass production) [dWB02].

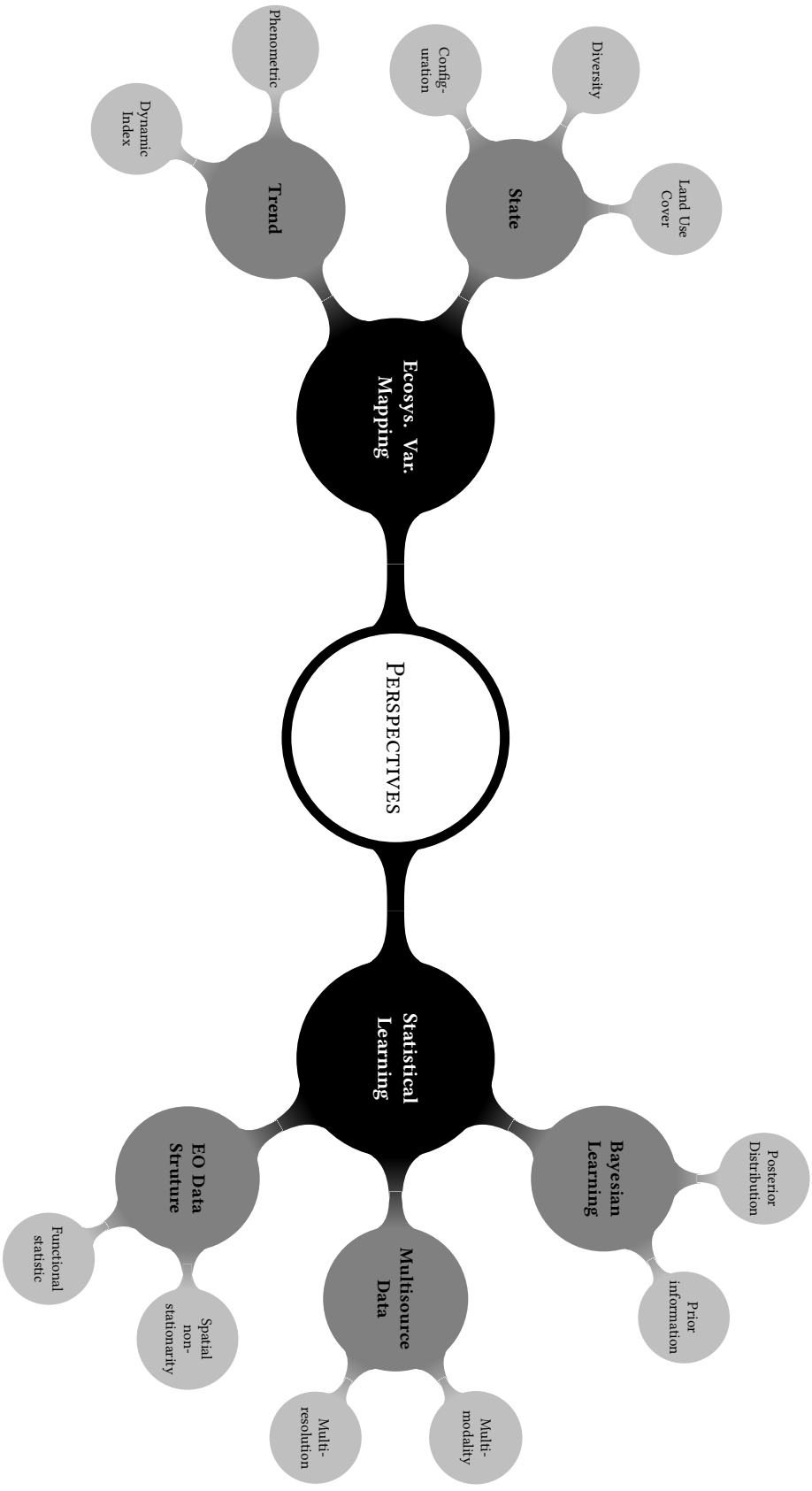


Figure 3.1.: Thematic and methodological perspectives.

aspect will be the ability of these variables to cope with several configuration of landscapes, e.g., different level of anthropization and urbanization. These variables will be related to the state and trends of ecosystem services or functions (supply, demand and benefit) [dWB02, BBC+22].

The Figure 3.1 provides a mind map of the thematic perspectives, divided into two type of EV, related to the *trend* and the *state* of the ecosystem. Maybe the most used variable is land cover and land use maps. I have been working since the beginning of my career on this type of problem. Even if it has been solved for general cases (see for instance the OSO products: https://www.theia-land.fr/en/product/land-cover-map/) there is room for improvement. In particular, the spatial continuity of the prediction should be improved to ensure an homogeneity of the prediction over the cover area. This problem affects large scale classification methods that split the spatial cover into spatial regions to reduce the computational complexity. Such issue is illustrated in Figure 3.2. To enforce spatial continuity, I believe in breaking the spatial stationarity assumption usually made in machine learning: the class statistical properties may change along the spatial position on Earth. I will discuss more about the technical aspect of this in the next section. Another important aspect of research concerns the extraction of semantic information, much more usable and informative for end-users. For instance, parcels with the same crops type (i.e., identical land-cover/use) can correspond to several agricultural practices (conventional, bio ...): in the current land cover maps they are identified by the same classes, but should be in the future discriminates into different classes because their effects on landscape are differents. This requires in particular to consider the spatial context.

Land cover maps are usually exploited to produce landscape composition (what are the species in the landscape) and configuration layers (how the elements of the landscape are spatially organized) that are used as proxy of different ecosystem services. However, they are based on a binary representation of the landscape (a pixel belongs to only one class, absolutely). I would like to investigate "continuous" landscape representation to derive composition and configuration index.

In addition to state variables of landscape, other time related ecologic variables need to be mapped at large scale. In particular, I have initiated work to estimate phenological variables for rapeseed and grasslands. Phenometrics are key variables to monitor vegetation w.r.t. climate changes. Yet, relying on databases to learn a machine learning model, such as for BBCH phenostages of rapeseed (https://en.wikipedia.org/wiki/BBCH-scale_(canola)), might be problematic for long term monitoring: such data is hard to collect and scarcely available. That is why, I have started to consider the computation of unsupervised phenometrics using raw time series. In particular, I currently investigate the *Dynamic Habitat Index* (DHI) variables, which are variables about the temporal behavior of the observed pixels. At the spatial resolution of Sentinel-2 SITS, clouds and shadows are an important source of noise and require specific effort. The figure 3.3 shows such problems: a strong spatial discontinuity is observed that match the (imperfect) clouds mask provided by CNES.

Last but not least, in order to be useful in the public debate, the extracted information (land cover, phenometrics, ecosystem services ...) should be accompanied by uncertainty measures describing the level of confidence associated to the extracted indices. Perfect accuracy in the processing of such data is unrealistic and these uncertainties would be a valuable complementary information. In particular, when evaluating different scenarios and their impacts on landscape, it is mandatory to use such confidence intervals (or its Bayesian counterpart, credible intervals).

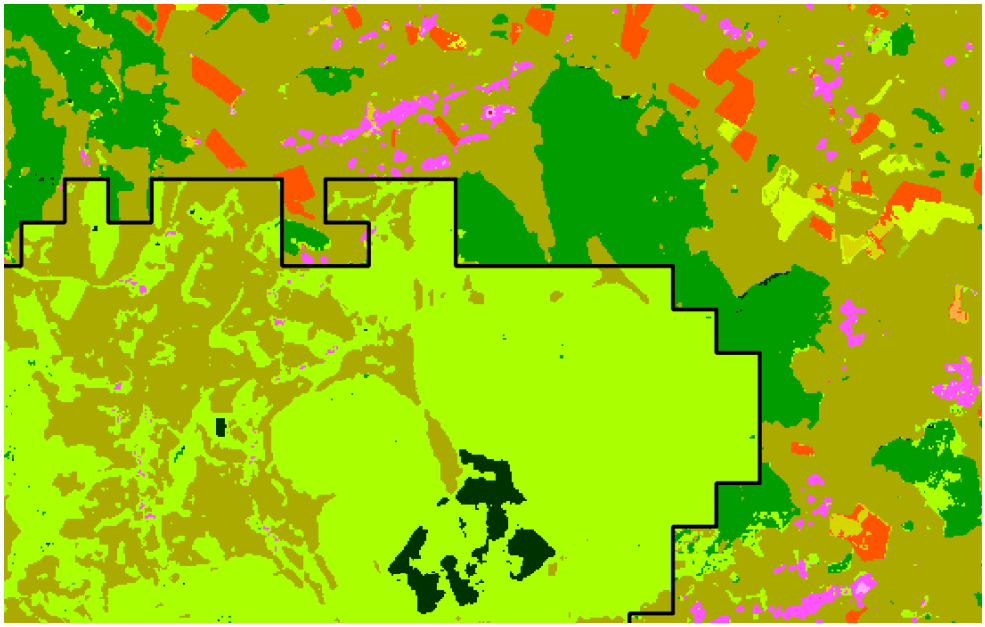
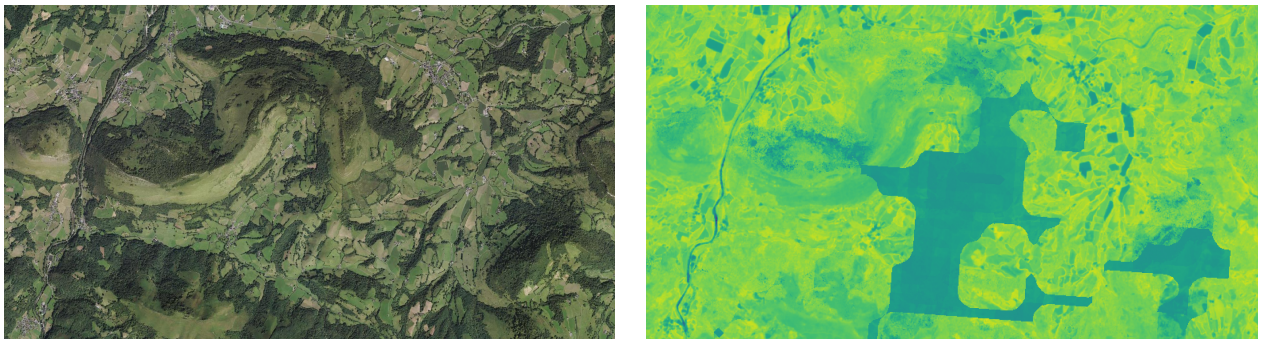


Figure 3.2.: Spatial discontinuity of the land cover classification between two regions. The black line is the separation between regions: it can be seen that class membership prediction is not continuous along the boundary. This is clearly a drawback since the observed landscape is continuous. The Figure was taken for Valentine Bellet personnal works.



(a) (b)

Figure 3.3.: Influence of clouds and shadows on the computation of the minimum of NDVI along one year of observation. (a) A very high spatial resolution image of the area, (b) the naive computation of the observed minimum value of the NDVI for one year of observation. From (b), the darker area in the middle of the image correspond to a non-detection of clouds for one (unknown) date of the SITS.

3.2. Methodological perspectives

Big data in satellite remote sensing bring new challenges and opportunities in many ways [CPB+16]. In addition to large scale and spatial issues, multi-source data coming from various on-board sensors provide complementary observations over the same ecosystem, as discussed in the introduction of the manuscript (see section 1.2). Furthermore, the volume of heterogeneous remote sensing data is increasing rapidly with commercial (e.g. PlanetScope) and (inter-)national agency missions. Therefore, it is critical to develop algorithms that ingest data with various physical natures and sampling resolution.

As a brief summary of section 1.3, from a statistical learning[2] perspectives such algorithms should be

- robust to the *curse of dimensionality*,
- able to work with irregularly sampled and multi-resolution/source data,
- able to scale w.r.t the number of samples,
- able to use prior information about the structure of the data (physics, temporal & spatial correlation ...).

These methodological questions have been aggregated during my different scientific experience, yet each number has increased since 2010: more dates, more spectral bands, more sensors, more samples with larger area. In twenty years, a few Gigabyte of data have been transformed into tens of Terabyte and my two cores CPU have been transformed into hundreds of distributed CPU/GPU. I am convinced that it is not a matter of scale, this is rather a change of paradigm on how to process the data with distributed storage and/or processing units and how we can include prior information in the processing.

The Figure 3.1 provides also a graphical summary of my methodological perspectives. The first line of research concerns the structure of the EO data: most of the freely available data (covering most of the emerged land) can be considered as image time series, with varying spatial and temporal resolution. For instance, for a given area, optical and radar time series area available, together with their related products, such as Leaf Area Index (LAI), Land Surface Temperature (LST), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) ... which can be viewed as a function of the time variable. In order to take into account this temporal structure, I have started to work using *functional statistics* with non-parametric models (PhD work of A. Zullo) and Gaussian processes (PhD work of A. Constantin and V. Bellet). In my future work, I need to include the spatial domain into the structure. Up to now, for computational issues, the spatial stationarity was assumed in the two models, i.e., the class conditional density function does not change w.r.t. the spatial location of pixels. Undoubtedly, this assumption is mostly false in practice. Spatio-temporal models are natural candidates, but some issues need to be solved to make them usable on large scale data. I expect that large scale learning strategies will offer reliable solution: stochastic optimisation coupled with variational strategy and modern software implementation (GPU and massively parallel implementation). Some works are currently investigated in the PhD work of V. Bellet for variational Gaussian process that make use of GPU. Further, consensus algorithms provide an interesting framework

[2] Or, to be trendy, "artificial intelligence" ...

to distribute computation when data are stored in different locations or when some explicit solution are amenable. Using ADMM framework, We had some recent results in [GFM21] for semi-supervised learning that show significant improvement in terms of computation time.

The second line of research is related to the joint use of multisource data coming from different modalities: different satellite sensors, data-bases with point-wise or/and polygone-wise measurements (see Figure 3.4 for a schematic illustration). Using spatio-temporal modeling, I intend to learn a continuous latent multi-scale representation. Some work have been initiated during the PhD work of A. Constantin. I plan to include in those models bio-physical *a priori* such as geometrical acquisition conditions, seasonality or geographical condition as regularizers during the learning process. Functional spatio-temporal models discussed in the previous paragraph will offer a flexible solution: the same latent process is observed by different sensors, at its own spatio-spectro-temporal resolution. For instance, seasonality could be included as functional constraints for some of the latent processes.

The third line of research concerns the definition of explainable and interpretable models. If deep neural models provide excellent results in various learning tasks, they are currently hardly explainable, i.e., it is difficult or impossible to explain outputs (being true of false). And except from specific models (e.g., transformer) their optimized parameters do not provide meaningful information neither about the predicted process nor their stability. It results in practice that deep models with different optimized parameters can perform similarly. As a consequence, it might be hazardous to include such uncontrolled estimation in a global environmental model. A current area of research trying to address these limitations is *Bayesian Deep Learning* (see for instance the special sessions at ICML https://icml.cc/virtual/2021/session/11976 and NIPS http://bayesiandeeplearning.org/). Bayesian methods provides a natural framework to integrate constraints (bio-physical, seasonality ...) using *prior* and *posterior* distributions over the model parameters. They also deal with uncertainty in the inference step with their ability to output distributions rather than point estimates. Bayesian methods were limited to small to medium data sets because their high computational load. But recent advances, such as amortized variational inference and stochastic gradient, have shown that it is possible to learn such models with a moderate increase of the computation load.

A Generic example A theoretical framework would be to assumed that the considered process $Z(R,\tau,\psi)$ (class label, pheno-dates, LAI ...) at a spatial resolution R, time τ and spatial coordinate ψ is explained by a (non-)linear link operator g_Z over L latent processes Y_ℓ, $\ell \in \{1,\dots,L\}$ at spatial resolution r_ℓ^Y:

$$Z(R,\tau,\psi) = g_z\Big(Y_1(r_1^Y,\tau,\psi),\dots,Y_L(r_L^Y,\tau,\psi)\Big) + E(R,\tau,\psi)$$

where E would model the measurement error. Each latent process Y_ℓ would be inferred from the available sources (SITS, field measurements, etc ...) X_s, $s \in \{1,\dots,S\}$:

$$Y_\ell(r_\ell,\tau,\psi) = f_\ell\Big(X_1(r_1^X,T_1,\psi),\dots,X_S(r_s^X,T_S,\psi);\psi\Big)$$

where T_s is all the temporal acquisition available (past or futures w.r.t. τ) for the source s and f_ℓ a non-stationary (w.r.t ψ) operator. Simple assumption would be to consider linear operators such as in [NKCB14], but last developments in representation learning and deep

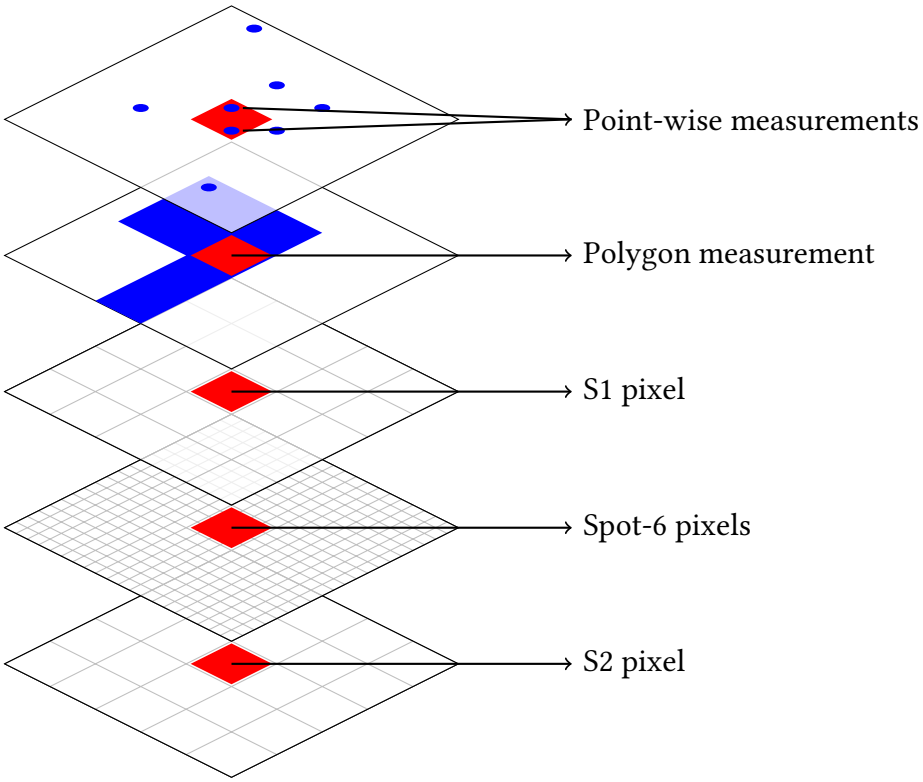


Figure 3.4.: Schematic illustration of multi-source data. For a given location, data with various modalities are available: In this diagram, a Sentinel-2 pixels is complemented by 16 Spot-6 pixels, 1 pixel from Sentinel-1 and field survey in a form of polygons describing the area to which the pixel belongs to and two point-wise measurement located inside the pixel location. Remote sensing data are ubiquitous (yet at varying resolution) while field data are sparsely available.

auto-encoder (as well as oldest kernel methods) have clearly shown the superiority of non-linear transformations.

As such, I believe that (deep) neural networks are natural candidate for the construction of the latent processes while simpler link operator g are needed to be able to estimate the full posterior distribution of Z in a Bayesian setting.

Note The M2GP model presented in section 2.2.2 is a simple case of the above model where:

- $L = p$ the number of spectral measurements,
- The number of sources is 1, the Sentinel-2 SITS,
- The latent variables Y_l are a smoothed functional representation of the original temporal spectra,
- f_l is a learned linear operator,
- The link operator g_z is a $\arg\max$ function of the class-conditional propability.

3.3. Researcher perspectives

3.3.1. Expertise

I am heavily involved in the GRSS (Geoscience and Remote Sensing Society - https://www.grss-ieee.org/). This technical society of the Institute of Electrical and Electronics Engineers (IEEE) brings together researchers and practitioners around geoscience and remote sensing to understand and monitor continental and oceanic ecosystems. The work published by this society concerns the means of collecting (sensor, satellite mission, etc.) and processing methods (machine learning, AI, etc.) of the data. Publications in the GRSS are *Open Access*, via fees for some journals, or by default for others. The average proofreading time is around 2.5 months, for 3 to 4 reviewers per paper. Even if certain review practices (*open peer review* for example) have not yet arrived in the GRSS (more generally in the IEEE) the proofreading process is of good quality. I am Associate Editor of the journal Trans. on Geoscience and Remote Sensing since 2018, for which I manage about forty papers per year. I participate in the annual organization of the IGARSS international conference (setting up oral and poster sessions for around 120 papers per year).

Until 2020 I was also responsible for the European chapters of the GRSS, but I resigned because actions I was able to carry out were closer to canvassing than to scientific animation. I also resigned in 2020 from my role as associate editor of the journal MDPI Remote Sensing for disagreeing with the policy of evaluation of submitted papers (proofreading time too short - 10 days + voucher if done in less than 7 days - and irrelevant reviewers added automatically in particular).

I believe in peer review system, but currently, I am worried about the injunction to review more papers, with shorten reviewing time. Some journals exhibit a *first decision to the authors* time in almost 20 days: from my experience, it results most of the time in binary reviews (accept or reject) with a small amount of comments. In my opinion, this is contradictory to what is a review process: apart detecting voluntary flaws and plagiarism, reviewers help the (associate)

editor to assess the novelty, significance and the suitability to the selected journal. Editor and reviewers also help authors to improve their work by suggesting changes in the paper, or by pointing out related works. This process need times in many parts: select proper reviewers with a sufficient number, let them review the paper (voluntarily!), collect the reviews and sometimes manage conflicting reviews.

As the COVID19 crisis shown, *fast* science is not synonym of *good* science. I am in favour of more openness, rather than shorter reviewing time, in the current system: reviews and responses should be made available with the accepted papers, codes and data set should be provided and made available for reproducibility. These practices are now common in other fields:

- in machine learning journals a git repository is always provided with the published paper;
- in environmental science, data papers are used to share data set.

In the future, my objective is to promote such practices in the GRSS and the different expertise process I am involved with. Yet, slowing the reviewing process is not an easy task.

3.3.2. Teaching (Masters, PhD, ...)

Having spent 7 years at INP-Toulouse as an MCF, I am still well integrated into training courses of the University of Toulouse. I usually supervise or co-supervise one or two M2 trainees each year, and occasionally M1. Per year, I have on average two co-supervised PhD students and one post-doctoral student under my supervision.

I co-organize a one-week doctoral training every year since 2018 (except in 2021 due to COVID), "Data Science 4 Geoscience" where we present statistical data analysis tools applied to geoscience. I also regularly organize internal training sessions in the laboratory (Data Science @ CESBIO - http://osr-cesbio.ups-tlse.fr/gitlab_cesbio/activites-ia/ds-cb). Occasionally, I organize continuous training actions.

All productions (theoretical content, computer codes, data, etc.) used in the various interventions are made available online in open access (licence GPLV3) using version management and collaborative editing software, `git` mainly (see CV latter in the document for a list of deposits).

These various training actions allow me to stay in contact with students and academic/industrial specialists from Toulouse in earth observation. I can thus orient some of my interventions according to current scientific issues. For example, in recent years, the use of AI in earth observation sciences has taken a growing place in both research and operational issues. I thus proposed interventions on theoretical and practical aspects in the master's course and for the doctoral school, and on the use of "operational" tools during continuing education workshops. In the future, I plan to maintain around forty hours of teaching annually.

3.3.3. Interaction with civil society

I am regularly solicited for expertise on space remote sensing and monitoring of the environment in various events. Especially,

- In 2019, I participated in a round table entitled "AI, a revolution for earth observation" organized by the Aerospace Valley competitiveness cluster (https://www.aerospace-valley.com/) ,

- In 2020, I presented CESBIO's activities in AI and earth observation at the INRAe stand during the Agricultural Show,
- In 2021, I was interviewed by a journalist from *Cash Investigation* for a program on CAP.
- In 2022, I did a meeting with journalist of Epsiloon (https://www.epsiloon.com/) about monitoring vegetation from space.

I believe it's important to participate in these kinds of events and I will surely pursue in the future, even if I don't really feel prepared for this type of exercise. In this regard, I appreciated being able to discuss with the communication department of INRAe Occcitanie Toulouse before my discussion with the journalist from Cash Investigation.

A have also participate to some mediation (2 half days in 2021) in primary school to explain why spatial data is useful to monitor our environment. The figure 3.5 shows some non-digital material have been made for this event...

Figure 3.5.: An home made support for kids (credit for the creation goes to Noémie, and the white bear was lend by Rémi) !

4. References

[AAW10] S. Aksoy, H.G. Akcay, and T. Wassenaar. Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery. *IEEE Trans. Geosci. Remote Sens.*, 48(1):511–522, January 2010.

[Abe05] S. Abe. Training of support vector machines with Mahalanobis kernels. In *Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005*, Lecture Notes in Computer Science, pages 571–576. Springer Berlin / Heidelberg, 2005.

[ADZ+18] Josselin Aval, Jean Demuynck, Emmanuel Zenou, Sophie Fabre, David Sheeren, Mathieu Fauvel, Karine Adeline, and Xavier Briottet. Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:197 – 210, 2018.

[AFZ+19] Josselin Aval, Sophie Fabre, Emmanuel Zenou, David Sheeren, Mathieu Fauvel, and Xavier Briottet. Object-based fusion for urban tree species classification from hyperspectral, panchromatic and ndsm data. *International Journal of Remote Sensing*, 40(14):5339–5365, 2019.

[AGSD12] Jesus Aguirre-Gutiérrez, Arie C. Seijmonsbergen, and Joost F. Duivenvoorden. Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in mexico. *Applied Geography*, 34:29 – 37, 2012.

[AHN90] J. Astola, P. Haavisto, and Y. Neuvo. Vector median filters. *Proceedings of the IEEE*, 78(4):678–689, 1990.

[AL07] E. Aptoula and S. Lefèvre. A comparative study on multivariate mathematical morphology. *Pattern Recognition*, 40(11):2914 – 2929, 2007.

[Aro50] N. Aronszajn. Theory of reprodusing kernel. Technical Report 11, Harvard University, Division of engineering sciences, 1950.

[BBB00] J. Baudry, R.G.H Bunce, and F. Burel. Hedgerows: An international perspective on their origin, function and management. *Journal of Environmental Management*, 60(1):7 – 22, 2000.

[BBC+22] Patricia Balvanera, Kate A Brauman, Anna F Cord, Evangelia G Drakou, Ilse R Geijzendorffer, Daniel S Karp, Berta Martín-López, Tuyeni H Mwampamba, and Matthias Schröter. Essential ecosystem service variables for monitoring progress towards sustainability. *Current Opinion in Environmental Sustainability*, 54:101152, 2022.

[BBM08] M. L. Braun, J. M. Buhmann, and K.-R. Müller. On relevant dimensions in kernel feature spaces. *J. Mach. Learn. Res.*, 9:1875–1908, June 2008.

[BCV12] Yoshua Bengio, Aaron C. Courville, and Pascal Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR*, abs/1206.5538, 2012.

[BDF10] José M. Bioucas-Dias and Mário AT Figueiredo. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. pages 1–4. IEEE, 2010.

[BFG14] C. Bouveyron, M. Fauvel, and S. Girard. Kernel discriminant analysis and clustering with parsimonious Gaussian process models. *Statistics and Computing*, pages 1–20, 2014.

[BG09] C. Bouveyron and S. Girard. Robust supervised classification with mixture models: Learning from data with uncertain labels. *Pattern Recognition*, 42(11):2649–2658, 2009.

[BGMF+08] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J.C.-W. Chan, and F. Canters. Improved classification of vhr images of urban areas using directional morphological profiles. *IEEE Trans. Geosci. Remote Sens.*, 46(10):2803 –2813, October 2008.

[BGS07] C. Bouveyron, S. Girard, and C. Schmid. High-Dimensional Discriminant Analysis. *Communication in Statistics- Theory and Methods / Communications in Statistics Theory and Methods*, 36:2607 – 2623, January 2007.

[BH99] Per Bolund and Sven Hunhammar. Ecosystem services in urban areas. *Ecological Economics*, 29(2):293 – 301, 1999.

[BHK+14] Thomas Blaschke, Geoffrey J. Hay, Maggi Kelly, Stefan Lang, Peter Hofmann, Elisabeth Addink, Raul Queiroz Feitosa, Freek van der Meer, Harald van der Werff, Frieke van Coillie, and Dirk Tiede. Geographic object-based image analysis – towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:180 – 191, 2014.

[BKL+12] Jodi S. Brandt, Tobias Kuemmerle, Haomin Li, Guopeng Ren, Jianguo Zhu, and Volker C. Radeloff. Using landsat imagery to map forest change in southwest china in response to the national logging ban and ecotourism development. *Remote Sensing of Environment*, 121:358 – 369, 2012.

[Bla10] T. Blaschke. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):2 – 16, 2010.

[BPA03] J.A. Benediktsson, M. Pesaresi, and K. Arnason. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Trans. Geosci. Remote Sens.*, 41(9):1940–1949, September 2003.

[BPS05] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.*, 43(3):480–491, March 2005.

[BST14] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.*, 146(1-2):459–494, August 2014.

[CBF06] J. Chanussot, J.A. Benediktsson, and M. Fauvel. Classification of remote sensing images from urban areas using a fuzzy possibilistic model. *IEEE Geoscience and Remote Sensing Letters*, 3:40–44, 2006.

[CBP03] J. Chanussot, J.A. Benediktsson, and M. Pesaresi. On the use of morphological alternated sequential filters for the classification of remote sensing images from urban

areas. In *Proc. IEEE Geoscience and Remote Sensing Symposium*, pages 473–475. IGARSS '03. Proceedings, July 2003.

[CC14] Pierre Cantelaube and Marie Carles. Le registre parcellaire graphique: des données géographiques pour décrire la couverture du sol agricole. *Le Cahier des Techniques de lINRA,(N Spécial GéoExpé)*, pages 58–64, 2014.

[CFG22] Alexandre Constantin, Mathieu Fauvel, and Stéphane Girard. Mixture of multivariate gaussian processes for classification of irregularly sampled satellite image time-series. *Statistics and Computing*, 2022.

[CG98] R. G. Congalton and K. Green. *Assessing the Accuracy of Remotely Sensed Data : Principles and Practices*. Mapping Science. Taylor & Francis, second edition, 1998.

[CPB+16] Mingmin Chi, Antonio Plaza, Jon Benediktsson, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, PP:1–13, 09 2016.

[CR97] Toby N. Carlson and David A. Ripley. On the relation between ndvi, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62(3):241 – 252, 1997.

[Cre15] N. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. Wiley, 2015.

[CSS95] J. Crespo, J. Serra, and R.W. Schafer. Theoretical aspects of morphological filters by reconstruction. *Signal Processing*, 47(2):201–225, 1995.

[CVRGMM+07] G. Camps-Valls, A. Rodrigo-Gonzalez, J. Muoz-Mari, L. Gomez-Chova, and J. Calpe-Maravilla. Hyperspectral image classification with Mahalanobis relevance vector machines. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 3802–3805, July 2007.

[CYXM20] X. Cao, J. Yao, Z. Xu, and D. Meng. Hyperspectral image classification with convolutional neural network and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4604–4616, 2020.

[dABAD15] Caio C. de Araujo Barbosa, Peter M. Atkinson, and John A. Dearing. Remote sensing of ecosystem services: A systematic review. *Ecological Indicators*, 52(nil):430–443, 2015.

[Daw81] Alexander Philip Dawid. Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, 68(1):265–274, April 1981.

[DDC+15] Sandra Díaz, Sebsebe Demissew, Julia Carabias, Carlos Joly, Mark Lonsdale, Neville Ash, Anne Larigauderie, Jay Ram Adhikari, Salvatore Arico, András Báldi, Ann Bartuska, Ivar Andreas Baste, Adem Bilgin, Eduardo Brondizio, Kai MA Chan, Viviana Elsa Figueroa, Anantha Duraiappah, Markus Fischer, Rosemary Hill, Thomas Koetz, Paul Leadley, Philip Lyver, Georgina M Mace, Berta Martin-Lopez, Michiko Okumura, Diego Pacheco, Unai Pascual, Edgar Selvin Pérez, Belinda Reyers, Eva Roth, Osamu Saito, Robert John Scholes, Nalini Sharma, Heather Tallis, Randolph Thaman, Robert Watson, Tetsukazu Yahara, Zakri Abdul Hamid, Callistus Akosim, Yousef Al-Hafedh, Rashad Allahverdiyev, Edward Amankwah, Stanley T Asah, Zemede Asfaw, Gabor Bartus, L Anathea Brooks, Jorge Caillaux, Gemedo Dalle, Dedy Darnaedi, Amanda

Driver, Gunay Erpul, Pablo Escobar-Eyzaguirre, Pierre Failler, Ali Moustafa Mokhtar Fouda, Bojie Fu, Haripriya Gundimeda, Shizuka Hashimoto, Floyd Homer, Sandra Lavorel, Gabriela Lichtenstein, William Armand Mala, Wadzanayi Mandivenyi, Piotr Matczak, Carmel Mbizvo, Mehrasa Mehrdadi, Jean Paul Metzger, Jean Bruno Mikissa, Henrik Moller, Harold A Mooney, Peter Mumby, Harini Nagendra, Carsten Nesshover, Alfred Apau Oteng-Yeboah, György Pataki, Marie Roué, Jennifer Rubis, Maria Schultz, Peggy Smith, Rashid Sumaila, Kazuhiko Takeuchi, Spencer Thomas, Madhu Verma, Youn Yeo-Chang, and Diana Zlatanova. The IPBES conceptual framework — connecting nature and people. *Current Opinion in Environmental Sustainability*, 14:1 – 16, 2015. Open Issue.

[Don00] D.L. Donoho. High-dimensional data analysis: the curses and blessing of dimensionality. In *AMS Mathematical challenges of the 21st century*, 2000.

[DP06a] J. Debayle and J.-C. Pinoli. General adaptive neighborhood image processing - part I. *J. Math. Imaging Vis.*, 25(2):245–266, 2006.

[DP06b] J. Debayle and J.-C. Pinoli. General adaptive neighborhood image processing - part II. *J. Math. Imaging Vis.*, 25(2):267–284, 2006.

[DS98] N.R. Draper and H. Smith. *Applied regression analysis*. Number v. 1 in Wiley series in probability and statistics: Texts and references section. Wiley, 1998.

[dWB02] Rudolf S de Groot, Matthew A Wilson, and Roelof M.J Boumans. A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecological Economics*, 41(3):393–408, 2002.

[EAA+10] Felix Eigenbrod, Paul R. Armsworth, Barbara J. Anderson, Andreas Heinemeyer, Simon Gillings, David B. Roy, Chris D. Thomas, and Kevin J. Gaston. The impact of proxy-based methods on mapping the distribution of ecosystem services. *Journal of Applied Ecology*, 47(2):377–385, 2010.

[Eil03] Paul H. C. Eilers. A perfect smoother. *Analytical Chemistry*, 75(14):3631–3636, 2003.

[ERT] French environment round table. http://www.legrenelle-environnement.fr/IMG/pdf/HS-8pGrenelle-Anglais.pdf.

[FA13] J. Feret and G. P. Asner. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):73–84, 2013.

[FAB+13] Mathieu Fauvel, Benoit Arbelot, Jon Atli Benediktsson, David Sheeren, and Jocelyn Chanussot. Detection of hedges in a rural landscape using a local orientation feature: from linear opening to path opening. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(1):15–26, 2013.

[FBCS08] M. Fauvel, J.A. Benediktsson, J. Chanussot, and J.R. Sveinsson. Spectral and spatial classification of hyperspectral data using svms and morphological profiles. *IEEE Trans. Geosci. Remote Sens.*, 46(11):3804 –3814, November 2008.

[FBG15] Mathieu Fauvel, Charles Bouveyron, and Stephane Girard. Parsimonious gaussian process models for the classification of hyperspectral remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 12(12):2423–2427, 2015.

[FCB06] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Evaluation of kernels for multiclass classification of hyperspectral remote sensing data. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2. ICASSP'06. Proceedings, May 2006.

[FCB09] M. Fauvel, J. Chanussot, and J.A. Benediktsson. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP J. Adv. Signal Process*, 2009:1–14, January 2009.

[FCB12] M. Fauvel, J. Chanussot, and J.A. Benediktsson. A spatial–spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition*, 45(1):381 – 392, 2012.

[FCBV13] Mathieu Fauvel, Jocelyn Chanussot, Jon Atli Benediktsson, and Alberto Villa. Parsimonious mahalanobis kernel for the classification of high dimensional data. *Pattern Recognition*, 46(3):845–854, 2013.

[FFA+08] Jean-Baptiste Feret, Christophe François, Gregory P. Asner, Anatoly A. Gitelson, Roberta E. Martin, Luc P.R. Bidel, Susan L. Ustin, Guerric le Maire, and Stéphane Jacquemoud. Prospect-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, 112(6):3030 – 3043, 2008.

[FLD+20] Mathieu Fauvel, Mailys Lopes, Titouan Dubo, Justine Rivers-Moore, Pierre-Louis Frison, Nicolas Gross, and Annie Ouin. Prediction of plant diversity in grasslands using sentinel-1 and -2 satellite image time series. *Remote Sensing of Environment*, 237:111536, 2020.

[FV06] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis : Theory and Practice*. New York : Springer-Verlag, 2006.

[FZF19] Frédéric Ferraty, Anthony Zullo, and Mathieu Fauvel. Nonparametric regression on contaminated functional predictor with application to hyperspectral data. *Econometrics and Statistics*, 9:95 – 107, 2019.

[GCCVBCM10] L. Gomez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla. Mean map kernel methods for semisupervised cloud classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):207–220, Jan 2010.

[GFM21] Erwan Giry Fouquet, Mathieu Fauvel, and Clément Mallet. Fast estimation for robust supervised classification with mixture models. *Pattern Recognition Letters*, 152:320–326, 2021.

[GMK15] G. Grekousis, G. Mountrakis, and M. Kavouras. An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing*, 36(21):5309–5335, 2015.

[Goo97] Pierre Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, USA, 1997. Publisher: Cambridge University Press.

[GRBBS08] Adrienne Grêt-Regamey, Peter Bebi, Ian D. Bishop, and Willy A. Schmid. Linking gis-based models to value ecosystem services in an alpine region. *Journal of Environmental Management*, 89(3):197 – 208, 2008.

[Gro13] MAE Working Group. Mapping and Assessment of Ecosystems and their Services: An analytical framework for ecosystem assessments under Action 5 of the EU Biodiviversity Strategy to 2020. Technical report, European Union, 2013.

[HFS15] R. Houborg, J. B. Fisher, and A. K. Skidmore. Advances in remote sensing of vegetation function and traits. *International Journal of Applied Earth Observation and Geoinformation*, 43:1 – 6, 2015.

[HHDD10] Olivier Hagolle, Mireille Huc, Villa Pascual David, and Gérard Dedieu. A multitemporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment*, 114(8):1747–1755, August 2010.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The elements of statistical learning: data mining, inference, and prediction, 2nd Edition*. Springer series in statistics. Springer, 2009.

[Hug68] G.F. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory*, IT-14:55–63, January 1968.

[HV96] T. O. Hammond and D. L. Verbyla. Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, 17(6):1261–1266, 1996.

[IAT+15a] J. Inglada, M. Arias, B. Tardy, O. Hagolle, S. Valero, D. Morin, G. Dedieu, G. Sepulcre, S. Bontemps, and P. Defourny. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9):12356–12379, Sep 2015.

[IAT+15b] Jordi Inglada, Marcela Arias, Benjamin Tardy, Olivier Hagolle, Silvia Valero, David Morin, Gérard Dedieu, Guadalupe Sepulcre, Sophie Bontemps, Pierre Defourny, and Benjamin Koetz. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9):12356–12379, 2015.

[IVA+17] Jordi Inglada, Arthur Vincent, Marcela Arias, Benjamin Tardy, David Morin, and Isabel Rodes. Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sensing*, 9(1), 2017.

[JBE+16] Neha Joshi, Matthias Baumann, Andrea Ehammer, Rasmus Fensholt, Kenneth Grogan, Patrick Hostert, Martin Rudbeck Jepsen, Tobias Kuemmerle, Patrick Meyfroidt, Edward T. A. Mitchard, Johannes Reiche, Casey M. Ryan, and Björn Waske. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1), 2016.

[JL98] L.O. Jimenez and D.A. Landgrebe. Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 28(1):39 –54, feb 1998.

[JLD11] Z. Jiang, Z. Lin, and L. S. Davis. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. pages 1697–1704, June 2011.

[KA06] Y. Kamada and S. Abe. Support vector regression using mahalanobis kernels. In Friedhelm Schwenker and Simone Marinai, editors, *Artificial Neural Networks in Pattern Recognition*, volume 4087 of *Lecture Notes in Computer Science*, pages 144–152. Springer Berlin / Heidelberg, 2006.

[KBGK09] Jagdish Krishnaswamy, Kamaljit S. Bawa, K.N. Ganeshaiah, and M.C. Kiran. Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season ndvi based mahalanobis distance surrogate. *Remote Sensing of Environment*, 113(4):857–867, 2009.

[KDF+19] Nicolas Karasiak, Jean-François Dejoux, Mathieu Fauvel, Jérôme Willm, Claude Monteil, and David Sheeren. Statistical stability and spatial instability in mapping forest tree species by comparing 9 years of satellite image time series. *Remote Sensing*, 11(21), 2019.

[Ken61] M.G. Kendall. *A course in the geometry of n-dimensions*. Dover Publication, New York, 1961.

[KTS+15] Daniel S. Karp, Heather Tallis, Rene Sachse, Ben Halpern, Kirsten Thonicke, Wolfgang Cramer, Harold Mooney, Stephen Polasky, Britta Tietjen, Katharina Waha, Ariane Walt, and Stacie Wolny. National indicators for observing ecosystem service change. *Global environmental change : human and policy dimensions*, 35:12 – 21, 2015.

[KVC+07] Alexandra-Maria Klein, Bernard E Vaissière, James H Cane, Ingolf Steffan-Dewenter, Saul A Cunningham, Claire Kremen, and Teja Tscharntke. Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1608):303–313, 2007.

[Lan03] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley and Sons, New Jersey, 2003.

[LFGS17] Mailys Lopes, Mathieu Fauvel, Stéphane Girard, and David Sheeren. Object-based classification of grasslands from high resolution satellite image time series using gaussian mean map kernels. *Remote Sensing*, 9(7), 2017.

[LFM+19] A. Lagrange, M. Fauvel, S. May, J. Bioucas-Dias, and N. Dobigeon. Matrix cofactorization for joint representation learning and supervised classification – application to hyperspectral image analysis. complementary results. Technical report, University of Toulouse, IRIT/INP-ENSEEIHT, France, Oct. 2019.

[LFM+20] Adrien Lagrange, Mathieu Fauvel, Stéphane May, José Bioucas-Dias, and Nicolas Dobigeon. Matrix cofactorization for joint representation learning and supervised classification – application to hyperspectral image analysis. *Neurocomputing*, 385:132 – 147, 2020.

[LFMD19] Adrien Lagrange, Mathieu Fauvel, Stéphane May, and Nicolas Dobigeon. Hierarchical bayesian image analysis: From low-level modeling to robust supervised learning. *Pattern Recognition*, 85:26 – 36, 2019.

[LOP17] Mailys LOPES. *Ecological monitoring of semi-natural grasslands: statistical analysis of dense satellite image time series with high spatial resolution*. Theses, Institut National Polytechnique de Toulouse, November 2017.

[LPT+15] Chinsu Lin, Sorin C. Popescu, Gavin Thomson, Khongor Tsogt, and Chein-I Chang. Classification of tree species in overstorey canopy of subtropical forest using quickbird images. *PLOS ONE*, 10(5):1–23, 05 2015.

[LRPM+14] Kévin Le Rest, David Pinaud, Pascal Monestiez, Joël Chadoeuf, and Vincent Bretagnolle. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global Ecology and Biogeography*, 23(7):811–820, 2014.

[McL92] G. J. McLachlan. *Discriminant analysis and statistical pattern recognition*. Wiley series in probability and mathematical statistics. J. Wiley and sons, New York, Chichester, Brisbane, 1992.

[MEW+12] Joachim Maes, Benis Egoh, Louise Willemen, Camino Liquete, Petteri Vihervaara, Jan Philipp Schägner, Bruna Grizzetti, Evangelia G. Drakou, Alessandra La Notte, Grazia Zulian, Faycal Bouraoui, Maria Luisa Paracchini, Leon Braat, and Giovanni Bidoglio. Mapping ecosystem services for policy support and decision making in the european union. *Ecosystem Services*, 1(1):31 – 39, 2012.

[MFDS12] K. Muandet, K. Fukumizu, F. Dinuzzo, and B. Schölkopf. Learning from distributions via support measure machines. In *Advances in Neural Information Processing Systems 25*, pages 10–18. Curran Associates Inc., 2012.

[MG10] Nishant A. Mehta and Alexander G. Gray. Generative and latent mean map kernels. *CoRR*, abs/1005.0188, 2010.

[MLM+17] Lei Ma, Manchun Li, Xiaoxue Ma, Liang Cheng, Peijun Du, and Yongxue Liu. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:277 – 293, 2017.

[MMR+01] K-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw.*, 12(2):181–202, March 2001.

[Nad64] E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9:141–142, 1964.

[NHA+16] Tim Newbold, Lawrence N. Hudson, Andrew P. Arnell, Sara Contu, Adriana De Palma, Simon Ferrier, Samantha L. L. Hill, Andrew J. Hoskins, Igor Lysenko, Helen R. P. Phillips, Victoria J. Burton, Charlotte W. T. Chng, Susan Emerson, Di Gao, Gwilym Pask-Hale, Jon Hutton, Martin Jung, Katia Sanchez-Ortiz, Benno I. Simmons, Sarah Whitmee, Hanbin Zhang, Jörn P. W. Scharlemann, and Andy Purvis. Has land use pushed terrestrial biodiversity beyond the planetary boundary? a global assessment. *Science*, 353(6296):288–291, 2016.

[NKCB14] Hai Nguyen, Matthias Katzfuss, Noel Cressie, and Amy Braverman. Spatio-temporal data fusion for very large remote sensing datasets. *Technometrics*, 56(2):174–185, 2014.

[O'M12] F O'Mara. The role of grasslands in food security and climate change. *Annals of Botany*, 110:1263–1270, 10 2012.

[OWR+10] Jens Oldeland, Dirk Wesuls, Duccio Rocchini, Michael Schmidt, and Norbert Jürgens. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecological Indicators*, 10(2):390–396, 2010.

[PB01] M. Pesaresi and J.A. Benediktsson. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.*, 39(2):309–320, February 2001.

[PBS05] Meelis Pärtel, Hans Henrik Bruun, and Marek Sammul. Biodiversity in temperate european grasslands: origin and conservation. In *Grassland Science in Europe*, volume 10, pages 1–14. Grassland Science in Europe, 2005. The information about affiliations in this record was updated in December 2015. The record was previously connected to the following departments: Plant Ecology and Systematics (Closed 2011) (011004000).

[PFW+13] H. M. Pereira, S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, A. C. Cardoso, N. C. Coops, E. Dulloo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Höft, G. Hurtt, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J. P. W. Scharlemann, S. N. Stuart, E. Turak, M. Walpole, and M. Wegmann. Essential biodiversity variables. *Science*, 339(6117):277–278, 2013.

[PPNH17] Jonne Pohjankukka, Tapio Pahikkala, Paavo Nevalainen, and Jukka Heikkonen. Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science*, 31(10):2001–2019, 2017.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. 12:2825–2830, 2011.

[PVM+05] Nathalie Pettorelli, Jon Olav Vik, Atle Mysterud, Jean-Michel Gaillard, Compton J. Tucker, and Nils Chr. Stenseth. Using the satellite-derived {NDVI} to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9):503 – 510, 2005.

[PWH+22] Claudio Persello, Jan Dirk Wegner, Ronny Hansch, Devis Tuia, Pedram Ghamisi, Mila Koeva, and Gustau Camps-Valls. Deep Learning and Earth Observation to Support the Sustainable Development Goals: Current Approaches, Open Challenges, and Future Opportunities. *IEEE Geoscience and Remote Sensing Magazine*, pages 2–30, 2022.

[QY01] S. Quegan and J. J. Yu. Filtering of multichannel sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11):2373–2379, Nov 2001.

[R C15] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[RBE+15] Robert A. Rose, Dirck Byler, J. Ron Eastman, Erica Fleishman, Gary Geller, Scott Goetz, Liane Guild, Healy Hamilton, Matt Hansen, Rachel Headley, Jennifer Hewson, Ned Horning, Beth A. Kaplin, Nadine Laporte, Allison Leidner, Peter Leimgruber, Jeffrey Morisette, John Musinsky, Lilian Pintea, Ana Prados, Volker C. Radeloff, Mary Rowen, Sassan Saatchi, Steve Schill, Karyn Tabor, Woody Turner, Anthony Vodacek, James Vogelmann, Martin Wegmann, David Wilkie, and Cara Wilson. Ten ways remote sensing can contribute to conservation. *Conservation Biology*, 29(2):350–359, 2015.

[RS02] J. Ramsay and B. W. Silverman. *Applied functional data analysis: methods and case studies*, volume 77. Springer, 2002.

[RS05a] J. Ramsay and B. W. Silverman. *Functional Data Analysis.* Springer, 2nd edition, 2005.

[Ser82] J. Serra. *Image Analysis and Mathematical Morphology.* U.K. Academic, 1982.

[Ser88] J. Serra. *Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances.* U.K. Academic, 1988.

[SFJ+16] David Sheeren, Mathieu Fauvel, Veliborka Josipović, Maïlys Lopes, Carole Planque, Jérôme Willm, and Jean-François Dejoux. Tree species classification in temperate forests using formosat-2 satellite image time series. *Remote Sensing*, 8(9), 2016.

[SMB+99] B. Schölkopf, S. Mika, C. J. C. Burges, P. Knirsh, K-R Müller, G. Rätsch, and A. J. Smola. Input space versue feature sapce in kernel-based methods. *IEEE Trans. Neural Netw.*, 10(5):1000–1017, May 1999.

[Soi03] P. Soille. *Morphological Image Analysis, Principles and Applications- 2nd edition.* Springer, 2003.

[Soi05] P. Soille. Beyond self-duality in morphological image analysis. *Image and Vision Computing*, 23(2):249–257, 2005.

[Soi09] P. Soille. Recent developments in morphological image processing for remote sensing. In *Proc. SPIE 7477*, 2009.

[SP02] P. Soille and M. Pesaresi. Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Trans. Geosci. Remote Sens.*, 40(9):2042–2055, September 2002.

[SS02] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support vector machines, regularization, optimization, and beyond.* MIT Press, 2002.

[SSM98] B. Schölkopf, A.J. Smola, and K.-R Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998.

[ST01] P. Soille and H. Talbot. Directional morphological filtering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1313 –1329, nov 2001.

[Tar05] Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.

[THL11] T. Touzet, X. Halbecq, and F. Lecordix. Chaîne de production de la carte de vegetation arborée française. In *Proceedings of the 25th International Cartographic Conference*, 2011.

[TPKE09] D. Tuia, F. Pacifici, M. Kanevski, and W.J. Emery. Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. *IEEE Trans. Geosci. Remote Sens.*, 47(11):3866 –3879, November 2009.

[UMH19] Md Palash Uddin, Md Al Mamun, and Md Ali Hossain. Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification. *International Journal of Remote Sensing*, 40(18):7190–7220, 2019.

[un15] united nations. transforming our world: the 2030 agenda for sustainable development. Technical report, united nations, 2015. https://sustainabledevelopment.un.org/post2015/transformingourworld/publication.

[Vap99] V. Vapnik. *The Nature of Statistical Learning Theory, Second Edition.* Springer, New York, 1999.

[Wat64] G. S. Watson. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics*, 26:359–372, 1964.

[WRWV10] J. D. Wickham, K. H. Riitters, Timothy G. Wade, and P. Vogt. A national assessment of green infrastructure and change for the conterminous united states using morphological image processing. *Landscape and Urban Planning*, 94(3-4):186–195, March 2010.

[WYT07] D. Wang, D.S. Yeung, and E.C. Tsang. Weighted mahalanobis distance kernels for support vector machines. *Neural Networks, IEEE Transactions on*, 18(5):1453 –1462, sept. 2007.

[ZFF18] Anthony Zullo, Mathieu Fauvel, and Frédéric Ferraty. Experimental comparison of functional and multivariate spectral-based supervised classification methods in hyperspectral image. *Journal of Applied Statistics*, 45(2):2219–2237, 2018.

[ZL10] Qiang Zhang and Baoxin Li. Discriminative K-SVD for dictionary learning in face recognition. pages 2691–2698, 2010.

[ZL14] Xiaolin Zhu and Desheng Liu. Accurate mapping of forest types using dense seasonal landsat time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96:1 – 11, 2014.

[ZLJ+19] H. Zhang, Y. Li, Y. Jiang, P. Wang, Q. Shen, and C. Shen. Hyperspectral classification based on lightweight 3-d-cnn with transfer learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5813–5828, 2019.

[ZMP13] Grazia Zulian, Joachim Maes, and Maria Paracchini. Linking land cover data and crop yields for mapping and assessment of pollination services in europe. *Land*, 2(3):472–492, 2013.

Part III.

Curriculum Vitae

4.1. Current position

Name	Mathieu Fauvel
Birth	June, 05 1981 (41 years old)
Position	1st grade researcher (*Since July 2018*)
Laboratory	UMR IRD-CNES-CNRS-UPS-INRAe 5126 CESBIO

4.2. Education & professional experience

4.2.1. Education

- **2007**: I obtained a joint PhD degree from the INPG and the University of Iceland, under the supervision of Jocelyn Chanussot and Jon Atli Benediktsson.

  Spectral and spatial methods for the classification of urban remote sensing data.

- **2004**: I graduated in electrical engineering from the Grenoble National Polytechnical Institute (INPG), France, and I received the M.S degree in image and signal processing from the INPG.

4.2.2. Professional experience

- **2019-2023**: Associate researcher to a Research Chair of 3IA ANITI (Artificial and Natural Intelligence Toulouse Institute).
- **Since 2018**: 1st grade researcher INRAe at CESBIO lab.
- **2011-2018**: I was an associate professor at the INP Toulouse.
  - **2017-2018**: CNRS delegation (50%) at CESBIO lab.
  - **2012**: I was a visiting professor at the University of Iceland (April-June).
- **2010-2011**: I was an assistant professor at the INP Toulouse.
- **2008-2010**: I was a post-doctoral fellow in the MISTIS INRIA team.
- **2007-2008**: I was an assistant professor at the INP Grenoble.
- **2004-2007**: I spent one year at the University of Iceland, during my PhD.

4.3. Research Activities

4.3.1. Projects

This section lists projects I have been (am) involved. Except specific mention, I was/am the lead of the project.

- **2021**: Impact study of the implementation of herd protection measures against predators. Funded by Conseil Départemental 64, Communauté d'Agglomération du Pays Basque, Conseil Régional Nouvelle Aquitaine, French gouvernement. The objective is to propose several scenarios and solutions for the herds predation risks. I am in co-charge of the realization of specific land cover and potential risks of predation maps.
- **2019-2023**: ANR 3IA ANITI. Funded by the ANR. The ambition of the ANITI project is to develop a new generation of artificial intelligence called hybrid AI, combining data-driven machine learning techniques with symbolic and formal methods for expressing

properties and constraints and carrying out logical reasoning. This approach will provide better guarantees in terms of reliability, robustness and the ability to explain and interpret the results of the algorithms used, while ensuring social acceptability and economic viability. Such guarantees are required by many applications targeted by the project, such as autonomous vehicles of the future. ANITI will bring together more than 200 researchers from universities, engineering schools, scientific and technological research organizations, and about thirty companies in the Toulouse region. I am co-chair of the AI Research Chair "Fusion-based inference from heterogeneous data" (http://dobigeon.perso.enseeiht.fr/aniti/).

- **2019-2024**: TOSCA PARCELLE. Funded by the CNES. It concerns the development and the validation of a processing chain for the analysis of massive satellite image time series at the national scale. 8 French laboratories collaborate to mutualise algorithms and data into this processing chain.
- **2019-2023**: ANR MAESTRIA. The MAESTRIA project (Multi-modAl Earth obServaTion Image Analysis) aims to solve the methodological challenges related to the fully automatic analysis of the massive amount of images acquired by Earth Observation platforms. MAESTRIA targets to generate land-cover and land-use descriptions at country scale at many spatial resolutions and sets of classes. The ultimate goal is to provide a continuum of spatially and semantically consistent products, that are relevant for many end-users and applications. Both public policies at local or national levels and scientific models will benefit from such kinds of products for climate modelling, urban planning, crop monitoring or impact assessment of surface changes. The project is led by Clement Mallet. I am in charge of the workpackage 2.
- **2017-2018**: TOSCA Project *HYPERBIO*. Funded by the CNES, it concerns the classification of tree species and the estimation of biodiversity variables from hyperspectral images. The consortium is composed of ONERA Toulouse, IRSTEA Montpellier and DYNAFOR Lab. The project was led by Jean-Baptiste Feret.
- **2015-2018**: TOSCA Project CES OSO. Funded by the CNES, it concerns the definition and the implementation of operational processing chain to derive products of land cover and land used at the national scale. This project is led by Jordi Inglada.
- **2015-2018**: MUESLI Project (MUltiscale mapping of Ecosystem Services by very high spatial resolution hyperspectraL and LiDAR remote sensing Imagery). Funded by the Toulouse University, which provides 220 k€ financial support to this 3-years.
- **2013-2017**: ANR ASTERIX. Funded by the ANR (French National Research Agency) in the program JCJC (Young Researchers), which provides a 275 k€ financial support to this 4-year initiative (2013-2017). This project was led by Sébastien Lefèvre.
- **2013-2014**: AMOR (Analyse d'iMages de télédétection à grande dimension tempORelle). Funded by the GDR ISIS. In collaboration with R. Flamary, S. Valero-Valbuena and M. Dalla Mura.
- **2013-2016**: ANISETTE (ANalyse fonctionnelle d'Images à grande dimension SpEcTrale et TemporellE : application à la caractérisation des éléments semi-naturels en écologie du paysage). Co funded by the CNES and the Midi-Pyrénées region . In collaboration with M. Goulard from the DYNAFOR lab and F. Ferraty and P. Vieu from the IMT.
- **2012-2013**: CARNAGE (CARactérisation d'éléments semi-NAturels dans les images satellitaires pléiades pour la Gestion de l'Environnement), project supported by the INP Toulouse. In collaboration with M. Goulard from the DYNAFOR lab and F. Ferraty and P. Vieu from

the IMT.
- **2011-2012**: SMI - INP Toulouse. Visiting scholar at the University of Iceland, Reykjavik, Iceland.

4.3.2. Supervision

Post-doctoral

1. **Tatsumi Uezato**, Spectral unmixing for the estimation of landscape heterogeneity and ecosystem services assessment, January 2017 - December 2018.
2. **Remi Duflot**, Multiscale Mapping of Ecosystem Services by Very High Spatial Resolution Hyperspectral and Lidar Remote Sensing Imagery, April 2017 - December 2017.
3. **Fabio Nor-guttler**, Detection of the Flavescence dorée grapevine disease by hyperspectral imagery Spectral signatures analysis and development of a specific vegetation index, May 2016- April 2017.
4. **Vincent Thierion**, Multiscale Mapping of Ecosystem Services by Very High Spatial Resolution Hyperspectral and Lidar Remote Sensing Imagery, January 2016 - August 2016.

PhD Students

1. **Valentine Bellet**, Artificial Intelligence for Ecosystem Monitoring using Remote Sensing and Digital Agriculture Data. Supervisors: Mathieu Fauvel and Jordi Inglada (2020-2023).
2. **Erwan Giry Fouquet**, *Very large scale learning for remote sensing image time series classification.* Supervisors: Mathieu Fauvel and Clement Mallet (2018-2021). `Thesis not defended, now at` Liberty Rider.
3. **Alexandre Constantin**, *Analysis of big satellite image time series*, Supervisors: Stéphane Girard and Mathieu Fauvel (2018-2021). `Now post-doctoral researcher position within the` GAIA team `in GIPSA-lab`.
4. **Adrien Lagrange**, *Multi-resolution learning for hierarchical analysis of hyperspectral and hypertemporal images.* Supervisors: Nicolas Dobigeon and Mathieu Fauvel (2016-2019). `Now at` Agenium Space.
5. **Josselyn Aval**, *Towards an automatic mapping of heritage trees in urban area.* Supervisors: Xavier Briotet and Emmanuel Zenou. Co-supervisors: Sophie Fabre, Mathieu Fauvel and David Sheeren (2015-2018). `Now at` Hytech-Imaging.
6. **Maïlys Lopes**, *Remote Sensing in landscape ecology: statistics in high dimension for spatial multiresolution and the high temporal resolution*, Supervisors: Mathieu Fauvel and Stéphane Girard (2014-2017). `Now at` Terranis.
7. **Anthony Zullo**, *Functional analysis of high dimensional remote sensing images: application to the charaterization of semi-natural objects in landscape ecology.* Supervisors: Mathieu Fauvel and Frédéric Ferraty (2013-2016).
8. **Pedram Ghamisi**, *Spectral and Spatial Techniques for the Classification of Hyperspectral Images.* Supervisors: Jon Atli Benediktsson, Jocelyn Chanussot and Mathieu Fauvel (2012-2015).

Engineers

1. **Hélène Touchais**, Large-scale land-cover mapping, November 2022 - October 2024.

2. **Hugo Trentesaux**, Large-scale land-cover mapping, October 2021 - July 2022.
3. **Lucas Schwaab**, Impact study of the implementation of herd protection measures against predators, March - December 2021.
4. **Benjamin Tardy**, Large-scale land-cover mapping, January 2019 - May 2020.
5. **Donatien Dallery**, CES THEIA "Operational Land Cover", November 2016 - May 2017. Achievement: http://osr-cesbio.ups-tlse.fr/~oso/posts/2017-03-30-carte-s2-2016/.
6. **Audrey Choné**, CES THEIA "Operational Land Cover", May 2016 - September 2016.
7. **Carole Planque**, Extraction of agro-ecological infrastructure by remote sensing, September 2013 - August 2014.

Msc Students

1. **Emma Senechal**, Détection des dates de Fauches de prairies à partir de série temporelle Sentinel-2, 2022.
2. **Victor Cathala**, Spectro-Temporal Feature Extraction for the classification on Satellite Image Time Series with Variational Gaussian Processes, 2022.
3. **Pratvi Kalavadiya**, Construction d'indices de dynamique temporelle des habitats à partir de série temporelle d'images Sentinel-2, 2021.
4. **Thierry Gaubert**, Estimation de métriques phénologiques à partir de séries temporelles d'images satellites, 2020.
5. **Sylvain Jourdanat**, Extraction of LiDar feature for the estimation of landscapes metrics/variables an ecosystem services assessment, 2017.
6. **Adrien Lagrange**, Operational feature selection in Gaussian mixtures models, 2016.
7. **Donatien Dallery**, Grasslands monitoring using satellite image time series, effect of *ages* and *practices* on the spectral profile, 2016.
8. **Veliborka Josipovic**, Analysis of Forest using satellite image time series, 2015.
9. **Marc Lang**, Grass land analysis with satellite image time series, 2015.
10. **Clement Deschene**, Feature extraction in hyperspectral images analysis, 2014.
11. **Anthony Zullo**, Hyperspectral images analysis with functional statistics, 2013.
12. **Benoit Arbelot**, Detection of hedges in a rural landscape using a local orientation feature: from linear opening to path opening , 2012.
13. **Quentin Labourey**, Road extraction in a dense urban area on multi-modal data, 2012.
14. **Fanny Collard**, Hedge detection with stereoscopic VHR images, 2011.
15. **Sergi Mercade**, Hyperspectral Images Classification for Urban Area Characterization, 2008.
16. **Alberto Villa**, Gradient Optimization for SVM hyperparameters, 2007.

4.4. Scientific commitment

4.4.1. International activities

Working Group

- Member of the CES Theia (Centre d'Expertise Scientifique) Variables pour la biodiversité
- IEEE *Senior Member*, n° 92293737.

- From 2016 to April 2020, I was member of the *GRSS Chapter Committee*. In charge of the European coordination.
- I was the president of the *IEEE Geoscience and Remote Sensing Society* French Chapter (2013-2016), with Mauro Dalla Mura (secretary) and Emmanuel Trouvé (treasurer). Website of the Chapter.

Editorial board

1. Associate Editor
   - Since Jully 2018, I have joined the editorial board of the IEEE Transactions on Geoscience and Remote Sensing as an Associate Editor.
   - Since January 2018 to April 2020, I was in the editorial board of the MPDI Remote Sensing as an Associate Editor of the Section Board for *Remote Sensing Image Processing*.
   - From December 2015 to December 2018, I was in the editorial board of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing as an Associate Editor.
2. Guest Editor
   a) Guest editor of the Special Issue "Operational Land Cover/Land Use Mapping", MDPI Remote Sensing, 2019.
   b) Guest editor of the Special Issue on Recent Advances in Processing of High-Spatial-Resolution Remote Sensing Data, IEEE JSTARS, 2018.
   c) Guest editor of the Special Issue Classification and Feature Extraction for Remote Sensing Image Analysis, MDPI Remote Sensing, 2018.
   d) Guest editor of the Special Issue Hyperspectral Remote Sensing of Forest and Trees outside Forests Ecosystems, MDPI Remote Sensing, 2018.
   e) Guest editor of the Special Issue on Analysis and Applications of Multitemporal data, IEEE JSTARS, 2016.
   f) Guest editor of the Special Issue on Information Extraction From High-Spatial-Resolution Optical Remotely Sensed Imagery, IEEE JSTARS, 2015.

Reviewer I am serving as a reviewer for several international journals and conferences. An (incomplete) resume of my activity can be seen on Publons: `https://publons.com/author/1292527/mathieu-fauvel#`

- **Journals**
  - IEEE Transactions on Geoscience and Remote Sensing
  - IEEE Geoscience and Remote Sensing Letters
  - IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
  - IEEE Geoscience and Remote Sensing Magazine
  - IEEE Transactions on Image Processing
  - ELSEVIER Signal Processing
  - ELSEVIER Pattern Recognition Letters
  - ELSEVIER Information Fusion
  - ELSEVIER Computers & Geosciences
  - ELSEVIER Digital Signal Processing
  - ELSEVIER Remote Sensing of Environnement

  - MDPI Remote Sensing
  - Computational Statistics and Data Analysis
- **Conferences**
  - IEEE International Conference on Acoustics, Speech and Signal Processing
  - IEEE International Geoscience and Remote Sensing Symposium
  - IEEE International Workshop on Machine Learning and Signal Processing
  - IEEE International Workshop on the Analysis of Multitemporal Remote Sensing Images
  - European Signal Processing Conference
  - Colloque GRETSI

I am also regularly contacted by national institutes to review funding applications (ANR, Research Foundation - Flanders, ...).

Conference committees

- Special Session Co-Organiser, Contexte et défis en traitement du signal et des images pour l'analyse des données de télédétection de dernière génération, **GRETSI** Nancy, 2022.
- Session Organizers (Data analysis methods) of the **IEEE Interntional Geoscience and Remote Sensing Symposium**, since 2018.
- Scientific Committe Member, **National workshop of hyperspectral Group**, 2019, Toulouse, France.
- Technical Committee Member, **National Conference on Photogrammetry and Remote Sensing**. 2018, Paris, France.
- Technical Committee Member, **Learning with functional data**. Workshop, 2016, Lille, France.
- Technical Committee Member, **IEEE Workshop on the Analysis of Multitemporal Remote Sensing Images**, 2015, Annecy, France.
- Technical Committee Member, workshop **Télédétection de la biodiversité : état des lieux et perspectives**. Joint seminar: GdR MAGIS (AP "Analyse d'images pour lesuivi des milieux") - ZA Alpes (CNRS) - IEEE GRSS French Chapter, 2014, Grenoble, France.
- Technical Committee Member, **IEEE Workshop on Hyperspectral Image and Signal Processing: Evolutions in Remote Sensing**, 2009, Grenoble, France.

Tutorials and scientific schools

- Course (1 day), Large scale analysis of satellite images time series with Iota software, Toulouse, France 2021 and 2022 (`https://gitlab.cesbio.omp.eu/fauvelm/formation-iota2` and `https://docs.iota2.net/training/labworks.html`)
- Doctoral Course, Data Sciences for Geosciences, 5 days, Toulouse, France 2020.
- Doctoral Course, Data Sciences for Geosciences, 5 days, Brest, France 2019.
- Course (1/2 day), *Introduction to Machine Learning: Data classification*, Toulouse (`https://framagit.org/mfauvel/omp_machine_learning`)
- Doctoral Course, Data Sciences for Geosciences, 5 days, Grenoble, France 2018.
- Lecturer (2 days) on *Hyperspectral Remote Sensing* at the First Algerian Geoscience and Remote Sensing Spring School 2017.
- Course (1/2 day) on *Introduction to Kernel Methods: Classification of Multivariate Data* at School of Astrostatistics 2015: Clustering and Classification.

- Tutorial (1/2 day) on *Image analysis of hyperspectral data using mathematical morphology* at the 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing.

4.4.2. Awards and distinction

- **2016**, Best chapter award jointly from
  - the Geoscience and Remote Sensing Society,
  - the IEEE France Section.
- **2015**, "*Prime d'Excellence Scientifique*" (PES).
- **2015**, INP INNOV award for scientific insights for the MUESLI project.
- **2012**, Best reviewer award, IEEE Transaction on Geoscience and Remote Sensing.
- **2013**, Recipient of the IEEE GRSS 2013 Transactions Prize Paper Award for the paper:

  "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profile" M. Fauvel, J. A. Benediktsson, J. Chanussot and J. R. Sveinsson IEEE Transaction on Geoscience and Remote Sensing, vol. 46, n° 11, pages 3804-3814, November, 2008.

4.4.3. Thesis committees

1. **Benjamin Tardy**, *Data fusion for land cover classification using obsolete data base*, CESBIO, 2019. `Examinateur`.
2. **Charlotte Pelletier**, *Cartographie de l'occupation des sols à grande échelle à partir d'images à hautes résolutions.* CESBIO 2014-2017. `Examinateur`.
3. **Iona Ilea**, *Classification robuste sur l'espace des matrices de covariance: application à la texture et aux images de télédétection polarimétriques Radar à Ouverture Synthétique.* Thèse de doctorat, University of Bordeaux and Universitatea tehnica (Cluj-Napoca, Roumanie), Janvier 2017. `Examinateur`.
4. **Arnaud Lebris**, *Optimisation de la configuration d'un instrument superspectral aéroporté pour la classification : application au milieu urbain.* Thèse de doctorat, Université Paris Est. Décembre 2015. `Examinateur`.
5. **François Cokelaer**, *Quality control using 3D X-ray tomography.* Thèse de doctorat, Grenoble Institut of Technology, 22 Février 2013. `Examinateur`.

4.4.4. Thesis monitoring committees

1. **Anthony Frion**, Apprentissage de systèmes dynamiques pour les séries temporelles d'images multispectrales, IMT Atlantique 2021-2024.
2. **Yoël Zerah**, Modèles Génératifs Profonds pour la détections de changements d'Occupation des Sols à partir de séries Temporelles d'Images Satellitaires, UMR CESBIO 2020-2023.
3. **Rollin Gimenez**, *Exploitation de données optiques hyperspectrales et multispectrales multitemporelles pour la cartographie des espèces végétales suivant leur sensibilité à l'anthropisation en lien à des activités industrielles.* ONERA - TotalEnergies - Lab. Ecologie fonctionnelle et Env., 2020-2023.
4. **Donald Luna**, Study of the resilience of permanent grasslands in response to climatic extremes with remote sensing image. UMR INRAe Ecosystem Prairial, 2020-2023.

5. **Benjamin Tardy**, *Data fusion for land cover classification using obsolete data base.* CESBIO 2017-2020.
6. **Charlotte Pelletier**, *Cartographie de l'occupation des sols à grande échelle à partir d'images à hautes résolutions.* CESBIO 2014-2017.
7. **Emmanuelle Cano**, *Cartographie des formations forestières des Pyrénées et analyse de changements - Utilisation de séries temporelles d'images satellitaires.* Laboratoire de Télédétection et de Gestion des Territoires, Ecole d'Ingénieurs de Purpan, 2013-2016.

4.5. Administrative commitment

- Co-Animator of the ANITI theme "Learning with complex data", 2019-2023
- Member of CNECA (Section 3) for 2018-2021
- In charge of the CESBIO working group "Land Cover/Land Use"
- **Director of the Engineering and Digital Sciences Teaching Department** at the National Institute of Agronomy of Toulouse (ENSAT) September 2015 - August 2018. *I was the founder of the teaching department.*
- Member of the selection committee for *associate professor* position in ENSAT, n° 4054, CNU 68.
- Elected representative of the associate professor of the DYNAFOR lab, 2013-2015.

4.6. Public Event

- 2019: Forum Aerospace Valley (https://www.aerospace-valley.com/sites/default/files/documents/events/programme_forum_2019_3.pdf)
- 2020: Salon International de l'agriculture with INRAe (https://twitter.com/MaFauvel/status/1231880929594920960/photo/1)
- 2022: Afterwork Club des Partenaires Industriel ANITI (https://aniti.univ-toulouse.fr/2022/06/14/afterwork-apprendre-avec-peu-de-donnees-ou-a-partir-de-donnees-complexes/)

4.7. Teaching activities

4.7.1. ANITI

As a co-chair of the ANITI research project, I gave around 64 h/eq. TP per year on Machine learning applied to geosience and remote sensing, mainly in the INP Toulouse MASTER SIGMA (http://sigma.univ-toulouse.fr/fr/index.html) and in the Toulouse graduate school of Earth and Space Science (https://tess.omp.eu/).

The material can be found here: https://mfauvel.frama.io/machine-learning/

4.7.2. INP Toulouse

I was an associate professor at the INP Toulouse from 2010 to 2018. I gave a full service each year mainly in remote sensing, applied image processing and machine learning with application

to land management. The table 4.1 provides an overview of my teaching load, not including all the administrative responsibilities. In 2015 I founded a new teaching department at ENSAT (Engineering and Digital Sciences[1]) and from 2015 to 2018, I headed it.

Table 4.1.: Average annual teaching load at INP Toulouse

School	Grades	Name	Format	Load
ENSAT	L3	Introduction to remote sensing	Lectures	07:30
ENSAT	M1	Remote sensing for land management	Lectures + Labworks	66:40
ENSAT	M2	Introduction to remote sensing	Lectures + Labworks	21:00
ENSAT	M2	Python for remote sensing	Lectures + Labworks	21:00
ENSAT	M2	Remote sensing and GIS Project	Projects	15:00
N7	M2	Signal and Image proc. applied to remote sensing	Lectures + Labworks	20:00
ENSAT	M2	Computer programming project	Projects	20:00
ENSAT	M1	Caesura year management		02:00
ENSAT	M1-2	Internship management		06:00
ENSAT	M2	Jury examinator		06:00
ENSAT		Head of SIN Department		15:00
			Total	200:10

4.7.3. INP Grenoble

During my PhD thesis, my post-doc position and my two assistant professor positions, I gave several lectures and labworks for an volume of approximately 610 hours. The table 4.2 summarize these activities.

Table 4.2.: Past 2011 teaching load

Topics	Load	Grades	School
Signal processing	130	L3 & M1	ENSIEG
Image processing	180	M1 & M2	ENSIEG, ENSE[3]
Remote Sensing	150	L3 & M1 & M2	ENSIEG, ENSE[3], ENSAT
Industrial data processing	44	M1	ENSIEG

4.8. List of publications

4.8.1. Journal papers

[1] R. Duflot, M. San-Cristobal, E. Andrieu, J.-P. Choisis, D. Esquerré, S. Ladet, A. Ouin, J. Rivers-Moore, D. Sheeren, C. Sirami, M. Fauvel, and A. Vialatte. Farming intensity indirectly reduces crop yield through negative effects on agrobiodiversity and key ecological functions. *Agriculture, Ecosystems & Environment*, 326:107810, 2022.

[1] Science de l'Ingénieur et du Numérique - SIN

[2] A. Constantin, M. Fauvel, and S. Girard. Mixture of multivariate gaussian processes for classification of irregularly sampled satellite image time-series. *Statistics and Computing*, 32(5):79, Sep 2022.

[3] A. Ouin, E. Andrieu, A. Vialatte, G. Balent, L. Barbaro, J. Blanco, E. Ceschia, F. Clement, M. Fauvel, N. Gallai, M. A. J. Hewison, J.-F. Dejoux, C. Kephaliacos, F. Macary, A. Probst, J. L. Probst, J. Ryschawy, D. Sheeren, A. Sourdril, T. Tallec, H. Verheyden, and C. Sirami. Building a shared vision of the future for multifunctional agricultural landscapes. Lessons from a long term socio-ecological research site in south-western France. *Advances in Ecological Research*, 2021.

[4] E. Giry Fouquet, M. Fauvel, and C. Mallet. Fast estimation for robust supervised classification with mixture model. *Pattern Recognition Letters*, 2021.

[5] A. Constantin, M. Fauvel, and S. Girard. Joint supervised classification and reconstruction of irregularly sampled satellite image times series. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–13, 2021.

[6] T. Uezato, M. Fauvel, and N. Dobigeon. Hierarchical Sparse Nonnegative Matrix Factorization for Hyperspectral Unmixing with Spectral Variability. *Remote Sensing*, 12(14):2326, 2020.

[7] A. Lagrange, M. Fauvel, S. May, and N. Dobigeon. Matrix cofactorization for joint spatial-spectral unmixing of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, page 1–13, 2020.

[8] A. Lagrange, M. Fauvel, S. May, J. Bioucas-Dias, and N. Dobigeon. Matrix cofactorization for joint representation learning and supervised classification – application to hyperspectral image analysis. *Neurocomputing*, 385:132–147, Apr 2020.

[9] M. Fauvel, M. Lopes, T. Dubo, J. Rivers-Moore, P.-L. Frison, N. Gross, and A. Ouin. Prediction of plant diversity in grasslands using sentinel-1 and -2 satellite image time series. *Remote Sensing of Environment*, 237:111536, Feb 2020.

[10] T. Uezato, M. Fauvel, and N. Dobigeon. Hyperspectral unmixing with spectral variability using adaptive bundles and double sparsity. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6):3980–3992, June 2019.

[11] A. Lagrange, M. Fauvel, S. May, and N. Dobigeon. Hierarchical bayesian image analysis: From low-level modeling to robust supervised learning. *Pattern Recognition*, 85:26 – 36, 2019.

[12] N. Karasiak, J.-F. Dejoux, M. Fauvel, J. Willm, C. Monteil, and D. Sheeren. Statistical stability and spatial instability in mapping forest tree species by comparing 9 years of satellite image time series. *Remote Sensing*, 11(21):2512, Oct 2019.

[13] X. Huang, X. X. Zhu, F. Dell'Acqua, M. Fauvel, M. Dalla Mura, and F. Lombardini. Foreword to the special issue on recent advances in processing of high-spatial-resolution remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8):2588–2593, Aug 2019.

[14] J. Aval, S. Fabre, E. Zenou, D. Sheeren, M. Fauvel, and X. Briottet. Object-based fusion for urban tree species classification from hyperspectral, panchromatic and ndsm data. *International Journal of Remote Sensing*, 40(14):5339–5365, 2019.

[15] A. Zullo, M. Fauvel, and F. Ferraty. Experimental comparison of functional and multivariate spectral-based supervised classification methods in hyperspectral image. *Journal of Applied Statistics*, 45(12):2219–2237, 2018.

[16] T. Uezato, M. Fauvel, and N. Dobigeon. Hyperspectral image unmixing with lidar data-aided spatial regularization. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–11, 2018.

[17] J. Aval, J. Demuynck, E. Zenou, S. Fabre, D. Sheeren, M. Fauvel, K. Adeline, and X. Briottet. Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:197 – 210, 2018.

[18] M. Lopes, M. Fauvel, A. Ouin, and S. Girard. Spectro-temporal heterogeneity measures from dense high spatial resolution satellite image time series: Application to grassland species diversity estimation. *Remote Sensing*, 9(10), 2017.

[19] M. Lopes, M. Fauvel, S. Girard, and D. Sheeren. Object-based classification of grasslands from high resolution satellite image time series using gaussian mean map kernels. *Remote Sensing*, 9(7), 2017.

[20] A. Lagrange, M. Fauvel, and M. Grizonnet. Large-scale feature selection with gaussian mixture models for the classification of high dimensional remote sensing images. *IEEE Transactions on Computational Imaging*, 3(2):230–242, 2017.

[21] F. Ferraty, A. Zullo, and M. Fauvel. Nonparametric regression on contaminated functional predictor with application to hyperspectral data. *Econometrics and Statistics*, 2017.

[22] E. Trouvé, G. Mercier, M. Fauvel, L. Bruzonne, and Y. Ban. Foreword to the special issue on analysis of multitemporal data and applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8):3356–3358, Aug 2016.

[23] D. Sheeren, M. Fauvel, V. Josipović, M. Lopes, C. Planque, J. Willm, and J.-F. Dejoux. Tree species classification in temperate forests using formosat-2 satellite image time series. *Remote Sensing*, 8(9):734, 2016.

[24] M. Fauvel. Introduction to kernel methods: Classification of multivariate data. *EAS Publications Series*, 77:171–193, 2016.

[25] X. Huang, M. Fauvel, M. Dalla Mura, and L. Zhang. Foreword to the special issue on information extraction from high-spatial-resolution optical remotely sensed imagery. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 8(5):1872–1875, 2015.

[26] P.-A. Herrault, D. Sheeren, M. Fauvel, and M. Paegelow. Vectorisation automatique des forêts dans les minutes de la carte d'état-major du 19 e siècle. *Revue internationale de géomatique*, 25(1):35–51, 2015.

[27] R. Flamary, M. Fauvel, M. Dalla Mura, and S. Valero. Analysis of multitemporal classification techniques for forecasting image time series. *Geoscience and Remote Sensing Letters, IEEE*, 12(5):953–957, 2015.

[28] M. Fauvel, C. Dechesne, A. Zullo, and F. Ferraty. Fast forward feature selection of hyperspectral images for classification with gaussian mixture models. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 8(6):2824–2831, 2015.

[29] M. Fauvel, C. Bouveyron, and S. Girard. Parsimonious gaussian process models for the classification of hyperspectral remote sensing images. *Geoscience and Remote Sensing Letters, IEEE*, 12(12):2423–2427, 2015.

[30] M. Fauvel. Remote sensing imagery [book review]. *Geoscience and Remote Sensing Magazine, IEEE*, 3(2):64–65, 2015.

[31] P. Ghamisi, M. S. Couceiro, M. Fauvel, and J. A. Benediktsson. Integration of segmentation techniques for classification of hyperspectral images. *IEEE Geosci. Remote Sensing Lett.*, 11(1):342–346, 2014.

[32] M. Fauvel, C. Planque, D. Sheeren, and M. Dalla Mura. Télédétection des éléments seminaturels: utilisation des données pléiades pour la détection des haies. *Revue française de photogrammétrie et de télédétection*, (208):111–116, 2014.

[33] M. Fauvel, M. Dalla Mura, and E. Trouve. Introduction of the IEEE french section geoscience and remote sensing society chapter [chapters]. *Geoscience and Remote Sensing Magazine, IEEE*, 2(4):59–61, 2014.

[34] C. Bouveyron, M. Fauvel, and S. Girard. Kernel discriminant analysis and clustering with parsimonious gaussian process models. *Statistics and Computing*, pages 1–20, 2014.

[35] P.-A. Herrault, D. Sheeren, M. Fauvel, C. Monteil, and M. Paegelow. A comparative study of geometric transformation models for the historical 'map of france'registration. *Geographia Technica*, 8(1):34–46, 2013.

[36] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013.

[37] M. Fauvel, J. Chanussot, J. A. Benediktsson, and A. Villa. Parsimonious mahalanobis kernel for the classification of high dimensional data. *Pattern Recognition*, 46(3):845–854, 2013.

[38] M. Fauvel, B. Arbelot, J. A. Benediktsson, D. Sheeren, and J. Chanussot. Detection of hedges in a rural landscape using a local orientation feature: from linear opening to path opening. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(1):15–26, 2013.

[39] M. Fauvel, J. Chanussot, and J. A. Benediktsson. A spatial–spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition*, 45(1):381–392, 2012.

[40] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson. SVM-and MRF-based method for accurate classification of hyperspectral images. *Geoscience and Remote Sensing Letters, IEEE*, 7(4):736–740, 2010.

[41] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4):293–307, 2010.

[42] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, et al. Recent advances in techniques for hyperspectral image processing. *Remote sensing of environment*, 113:S110–S122, 2009.

[43] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP Journal on Advances in Signal Processing*, 2009:11, 2009.

[44] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Retrieval of mars surface physical properties from omega hyperspectral images using regularized sliced inverse regression. *Journal of Geophysical Research: Planets (1991–2012)*, 114(E6), 2009.

[45] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(11):3804–3814, 2008.

[46] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Decision fusion for the classification of urban remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(10):2828–2838, 2006.

[47] J. Chanussot, J. A. Benediktsson, and M. Fauvel. Classification of remote sensing images from urban areas using a fuzzy possibilistic model. *Geoscience and Remote Sensing Letters, IEEE*, 3(1):40–44, 2006.

4.8.2. Book chapters

[1] S. Ladet, D. Sheeren, P. Herrault, and M. Fauvel. *Assessing the Influence of Landscape on Biodiversity Using the QGIS Plugin LecoS*, chapter 7, pages 239–264. Wiley-Blackwell, 2018, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119457121.ch7.

[2] M. Fauvel, S. Girard, S. Douté, and L. Gardes. In *Horizons in World Physics*, chapter Machine Learning Methods for the Inversion of Hyperspectral Images. Nova Science, 2017.

[3] B. Waske, M. Fauvel, J. A. Benediktsson, and J. Chanussot. In *Kernel methods for Remote Sensing Data Analysis*, chapter Machine learning techniques in remote sensing data analysis. John Wiley and Sons, New York, 2009.

[4] M. Fauvel, J. Chanussot, and J. A. Benediktsson. In *Hyperspectral Data Exploitation: Theory and Applications*, chapter Decision fusion for hyperspectral classification. John Wiley and Sons, New York, 2007.

4.8.3. Conference papers

[1] N. Karasiak, M. Fauvel, J.-F. Dejoux, C. Monteil, and D. Sheeren. Optimal dates for deciduous tree species mapping using full years sentinel-2 time series in south west France. In *24th ISPRS Congress on Technical Commission III*, volume 5 of *24th ISPRS Congress on Technical Commission III; Nice; France;*, pages 469–476, Nice, France, Aug. 2020.

[2] A. Lagrange, M. Fauvel, S. May, J. M. Bioucas-Dias, and N. Dobigeon. Matrix cofactorization for joint unmixing and classification of hyperspectral images. *2019 27th European Signal Processing Conference (EUSIPCO)*, Sep 2019.

[3] A. Lagrange, M. Fauvel, S. May, J. M. Bioucas-Dias, and N. Dobigeon. Cofactorisation de matrices pour le démélange et la classification conjoints d'images hyperspectrales. In *27e colloque du Groupe de Recherche et d'Etudes du Traitement du Signal et des Images (GRETSI 2019)*, pages 1–4, Lille, France, Aug. 2019.

[4] J. Aval, J. Demuynck, E. Zenou, S. Fabre, D. Sheeren, M. Fauvel, and X. Briottet. Identification of the london plane in urban alignment based on hyperspectral data and contextual information. IEEE, May 2019.

[5] T. Uezato, M. Fauvel, and N. Dobigeon. A multiple endmember mixing model to handle spectral variability in hyperspectral unmixing. In *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5, Sep. 2018.

[6] T. Uezato, M. Fauvel, and N. Dobigeon. Lidar-driven spatial regularization for hyperspectral unmixing. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1740–1743, July 2018.

[7] M. San Cristobal, R. Duflot, M. Fauvel, and A. Vialatte. Modeling link complexity between crop yield, ecosystem services and multiple landscape representations in the coteaux de gascogne region. In *Sfécologie 2018. International Conference on Ecological Sciences*, pages 415–415, 2018.

[8] A. Lagrange, M. Fauvel, S. May, and N. Dobigeon. A bayesian model for joint unmixing and robust classification of hyperspectral images. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3399–3403, April 2018.

[9] M. Lopes, M. Fauvel, A. Ouin, and S. Girard. Potential of sentinel-2 and spot5 (take5) time series for the estimation of grasslands biodiversity indices. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pages 1–4, June 2017.

[10] M. Lopes, M. Fauvel, A. Ouin, and S. Girard. Evaluation de la biodiversité des prairies semi-naturelles par télédétection hyperspectrale. In *SFPT-GH 2017 - 5ème colloque scientifique du groupe thématique hyperspectral de la Société Française de Photogrammétrie et Télédétection*, volume 24, Brest, France, May 2017.

[11] M. Lopes, M. Fauvel, A. Ouin, and S. Girard. Estimation de la diversité en espèces des prairies à partir de leur hétérogénéité spectrale en utilisant des séries temporelles d'images satellite à haute résolution spatiale. In *Rencontres d'Ecologie des Paysages 2017*, pages 1–25, Toulouse, France, Oct. 2017.

[12] N. Karasiak, D. Sheeren, M. Fauvel, J. Willm, J. F. Dejoux, and C. Monteil. Mapping tree species of forests in southwest france using sentinel-2 image time series. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pages 1–4, June 2017.

[13] J. Aval, S. Fabre, E. Zenou, D. Sheeren, M. Fauvel, and X. Briottet. Urban tree species classification with hyperspectral VNIR and SWIR sensors. In *10th EARSeL SIG Imaging Spectroscopy workshop, Zurich*, 2017.

[14] M. Lopes, S. Girard, and M. Fauvel. Divergence de Kullback-Leibler en grande dimension pour la classification des prairies à partir de séries temporelles d'images satellite à haute résolution. In *48èmes Journées de Statistique organisées par la Société Française de Statistique*, Montpellier, France, May 2016.

[15] M. Lopes, M. Fauvel, S. Girard, D. Sheeren, and M. Lang. High dimensional kullback-leibler divergence for grassland object-oriented classification from high resolution satellite image time series. In *Living Planet Symposium*, 2016.

[16] M. Lopes, M. Fauvel, S. Girard, and D. Sheeren. High dimensional kullback-leibler divergence for grassland management practices classification from high resolution satellite image time series. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pages 3342–3345. IEEE, 2016.

[17] M. Fauvel, D. Sheeren, J.-F. Dejoux, and J. Willm. Multi-year comparison of tree species discrimination from Formosat-2 satellite image time series. In *ForestSAT'2016 International Conference*, Santiago, Chile, November 2016. (abstract & oral presentation).

[18] D. Sheeren, M. Fauvel, C. Planque, J. Willm, and J.-F. Dejoux. Tree species discrimination in temperate woodland using high spatial resolution formosat-2 time series. In *Analysis of Multitemporal Remote Sensing Images (Multi-Temp), 2015 8th International Workshop on the*, pages 1–4. IEEE, 2015.

[19] A. Zullo, M. Fauvel, F. Ferraty, M. Goulard, and P. Vieu. Non-parametric functional methods for hyperspectral image classification. In *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, pages 3422–3425. IEEE, 2014.

[20] M. Fauvel, C. Planque, D. Sheeren, M. Dalla Mura, F. Cokelaer, J. Chanussov, and H. Talbot. Robust path opening versus path opening for the detection of hedgerows in rural landscapes. In *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, pages 4910–4913. IEEE, 2014.

[21] M. Fauvel, C. Bouveyron, and S. Girard. Parsimonious gaussian process models for the classification of multivariate remote sensing images. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 2913–2916. IEEE, 2014.

[22] P.-A. Herrault, D. Sheeren, M. Fauvel, and M. Paegelow. Automatic extraction of forests from historical maps based on unsupervised classification in the cielab color space. In *Geographic Information Science at the Heart of Europe*, pages 95–112. Springer International Publishing, 2013.

[23] P. Ghamisi, M. S. Couceiro, M. Fauvel, and J. Atli Benediktsson. Spectral-spatial classification based on integrated segmentation. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, pages 1458–1461. IEEE, 2013.

[24] M. Fauvel. Approximation de matrices pour l'apprentissage des hyperparamètres des fonctions noyaux gaussiennes. In *24. Colloque GRETSI sur le traitement du signal et des images. 2013-09-032013-09-06, Brest, FRA*, 2013.

[25] M. Fauvel, D. Sheeren, J. Chanussot, and J. A. Benediktsson. Hedges detection using local directional features and support vector data description. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 2320–2323. IEEE, 2012.

[26] M. Fauvel. Kernel matrix approximation for learning the kernel hyperparameters. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 5418–5421. IEEE, 2012.

[27] G. Cavallaro, B. Arbelot, M. Fauvel, M. Dalla Mura, J. A. Benediktsson, L. Bruzzone, J. Chanussot, and D. Sheeren. Detection of hedges based on attribute filters. In *SPIE Remote Sensing*, pages 853712–853712. International Society for Optics and Photonics, 2012.

[28] C. Bouveyron, M. Fauvel, and S. Girard. Processus gaussiens parcimonieux pour la classification générative de données hétérogènes. In *44èmes Journées de Statistique*, 2012.

[29] D. Sheeren, M. Fauvel, S. Ladet, A. Jacquin, G. Bertoni, and A. Gibon. Mapping ash tree colonization in an agricultural mountain landscape: Investigating the potential of hyperspectral imagery. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 3672–3675. IEEE, 2011.

[30] S. Ladet, D. Sheeren, M. Fauvel, W. Heintz, and M. Deconchat. Mission d'acquisition aérienne lidar pour des applications thématiques en écologie du paysage. le cas des coteaux de gascogne. In *Ecole thématique du CNRS Images et modèles 3D en milieux naturels. 2011-06-142011-06-17, Le Bourget du Lac, FRA*. Laboratoire EDYTEM-Environnements, Dynamiques et Territoires de la Montagne, 2011.

[31] M. Fauvel, A. Villa, J. Chanussot, and J. A. Benediktsson. Mahalanobis kernel based on probabilistic principal component. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 3907–3910. IEEE, 2011.

[32] M. Fauvel, A. Villa, J. Chanussot, and J. A. Benediktsson. Mahalanobis kernel for the classification of hyperspectral images. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 3724–3727. IEEE, 2010.

[33] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, J. Angulo, and M. Fauvel. Classification of hyperspectral data using support vector machines and adaptive neighborhoods. In *Proc. 6th EARSeL SIG IS Workshop*, 2009.

[34] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Kernel principal component analysis for the construction of the extended morphological profile. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, volume 2, pages II–843. IEEE, 2009.

[35] J. Chanussot, J.-A. Benediktsson, M. Fauvel, and Y. Tarabalka. Spectral-spatial analysis in hyperspectral remote sensing: from morphological profiles to classified segmentation. In *SPIE Europe Remote Sensing*, pages 74770E–74770E. International Society for Optics and Photonics, 2009.

[36] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Support vectors machines regression for estimation of mars surface physical properties. In *ESANN 2009-17th European Symposium on Artificial Neural Networks*, pages 195–200. d-side publications, 2009.

[37] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Machine learning techniques for the inversion of planetary hyperspectral images. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on*, pages 1–4. IEEE, 2009.

[38] A. Villa, M. Fauvel, J. Chanussot, P. Gamba, and J. A. Benediktsson. Gradient optimization for multiple kernel's parameters in support vector machines classification. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 4, pages IV–224. IEEE, 2008.

[39] Y. Tarabalka, J. Chanussot, J. A. Benediktsson, J. Angulo, and M. Fauvel. Segmentation and classification of hyperspectral data using watershed. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 3, pages III–652. IEEE, 2008.

[40] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 1884–1887. IEEE, 2008.

[41] J. A. Benediktsson, X. C. Garcia, B. Waske, J. Chanussot, J. R. Sveinsson, and M. Fauvel. Ensemble methods for classification of hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 1, pages I–62. IEEE, 2008.

[42] M. Fauvel, J. Chanussot, and J. A. Benediktsson. A joint spatial and spectral svm's classification of panchromatic images. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 1497–1500. IEEE, 2007.

[43] M. Fauvel, J. Chanussot, and J. A. Benediktsson. How transferable are spatial features for the classification of very high resolution remote sensing data? In *Urban Remote Sensing Joint Event, 2007*, pages 1–5. IEEE, 2007.

[44] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Classification en imagerie hyperspectrale haute résolution: utilisation conjointe de l'information spatiale et spectrale. In *21° Colloque GRETSI, Troyes, FRA, 11-14 septembre 2007*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2007.

[45] J. A. Benediktsson, J. Chanussot, and M. Fauvel. Multiple classifier systems in remote sensing: from basics to recent developments. In *Multiple Classifier Systems*, pages 501–512. Springer Berlin Heidelberg, 2007.

[46] A. Plaza, M. Fauvel, J. A. Benediktsson, J. Chanussot, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, P. Gamba, A. Gualtieri, et al. Analysis of hyperspectral images–advanced technologies. In *Geoscience and Remote Sensing Symposium, 2006. IGARSS'06. Proceedings. 2006 IEEE International*, pages 1–5. IEEE, 2006.

[47] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, et al. Advanced processing of hyperspectral images. In *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, pages 1974–1978. IEEE, 2006.

[48] M. Fauvel, J. Chanussot, J. A. Benediktsson, et al. A combined support vector machines classification based on decision fusion. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, volume 6, 2006.

[49] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Kernel principal component analysis for feature reduction in hyperspectrale images analysis. In *Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic*, pages 238–241. IEEE, 2006.

[50] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Evaluation of kernels for multiclass classification of hyperspectral remote sensing data. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 2, pages II–II. IEEE, 2006.

[51] M. Fauvel, J. A. Palmason, J. A. A. Benediktsson, J. Chanussot, and J. R. Sveinsson. Classification of remote sensing imagery with high spatial resolution. In *Remote Sensing*, pages 598201–598201. International Society for Optics and Photonics, 2005.

[52] M. Fauvel, J. Chanussot, J. A. Benediktsson, et al. Fusion of methods for the classification of remote sensing images from urban areas. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, volume 4, page 2819, 2005.

[53] M. Fauvel, J. Chanussot, and J. A. Benediktsson. Fusion de méthodes pour la classification de zones urbaines en télédétection satellitaire. In *20° Colloque sur le traitement du signal et des images, FRA, 2005*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2005.

4.8.4. Scientific Schools

[1] M. Fauvel, B. Tardy, and A. Vincent. Formation à iota2, July 2021. https://doi.org/10.5281/zenodo.5082941.

[2] F. Chatelain and M. Fauvel. Introduction to Machine Learning - Data Classification, Dec. 2018. https://doi.org/10.5281/zenodo.1920227.

[3] M. Fauvel. Course on hyperspectral image processing, May 2017. https://doi.org/10.5281/zenodo.570420.

[4] M. Dalla Mura and M. Fauvel. Image analysis of hyperspectral data using mathematical morphology, June 2014. https://doi.org/10.5281/zenodo.437195.

4.8.5. Thesis

[1] M. Fauvel. *Spectral and spatial methods for the classification of urban remote sensing data.* PhD thesis, Institut National Polytechnique de Grenoble-INPG; Université d'Islande, 2007.