



**HAL**  
open science

## Spatio-temporal reasoning for the classification of satellite image time series

François Petitjean, Camille Kurtz, Nicolas Passat, Pierre Gançarski

► **To cite this version:**

François Petitjean, Camille Kurtz, Nicolas Passat, Pierre Gançarski. Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recognition Letters*, 2012, 33 (13), pp.1805-1815. 10.1016/j.patrec.2012.06.009 . hal-01698140

**HAL Id: hal-01698140**

**<https://hal.univ-reims.fr/hal-01698140>**

Submitted on 5 Mar 2018

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

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

# Spatio-Temporal Reasoning for the Classification of Satellite Image Time Series

François Petitjean<sup>a,\*</sup>, Camille Kurtz<sup>a,b</sup>, Nicolas Passat<sup>a,b</sup>, Pierre Gançarski<sup>a,b</sup>

<sup>a</sup>LSIIT – UMR 7005, Pôle API, Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France

<sup>b</sup>University of Strasbourg, 7 rue René Descartes, 67084 Strasbourg Cedex, France

---

## Abstract

Satellite Image Time Series (SITS) analysis is an important domain with various applications in land study. In the coming years, both high temporal and high spatial resolution SITS will become available. In the classical methodologies, SITS are studied by analyzing the radiometric evolution of the pixels with time. When dealing with high spatial resolution images, object-based approaches are generally used in order to exploit the spatial relationships of the data. However, these approaches require a segmentation step to provide contextual information about the pixels. Even if the segmentation of single images is widely studied, its generalization to series of images remains an open-issue. This article aims at providing both temporal and spatial analysis of SITS. We propose first segmenting each image of the series, and then using these segmentations in order to characterize each pixel of the data with a spatial dimension (*i.e.*, with contextual information). Providing spatially characterized pixels, pixel-based temporal analysis can be performed. Experiments carried out with this methodology show the relevance of this approach and the significance of the resulting extracted patterns in the context of the analysis of SITS.

*Key words:* Multi-temporal Analysis, Satellite Image Time Series, Data Mining, Segmentation, Information Extraction

---

## 1. Introduction

Satellite Image Time Series (SITS) constitute a major resource for Earth monitoring. For the last decades, these image series have been either sensed with a high temporal resolution (daily coverage at a kilometer spatial resolution) or with a high spatial resolution (weekly coverage at a meter spatial resolution). However, for a few years, satellites such as the Taiwanese FORMOSAT-2 are providing both high temporal and High Spatial Resolution SITS (HSR SITS), but with a limited coverage of the Earth surface and with only four spectral bands. In the coming years, these kinds of data will become widely available thanks to the ESA's SENTINEL program. The growing availability of such images, periodically acquired by satellite sensors on the same geographical

---

\*Corresponding author – LSIIT, Pôle API, Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France – Tel.: +33 3 68 85 45 78 – Fax.: +33 3 68 85 44 55

18 area, will make it possible to produce and regularly update accurate temporal land-cover maps of  
19 a given investigated site.

20 In order to efficiently handle the huge amount of data that will be produced by these new  
21 sensors, adapted methods for SITS analysis have to be developed. Such methods should allow the  
22 end-user to obtain satisfactory results, *e.g.*, relevant and accurate temporal evolution behaviors,  
23 with minimal time (by automating the tasks which do not require human expertise), and minimal  
24 effort (by reducing the parameters).

25 In the current standard methods, these data are studied by analyzing the radiometric evolution  
26 of the pixels through the time series. The underlying idea is to gather sensed areas that undergo  
27 similar radiometric evolutions. This structuring of the data makes it possible to extract both abrupt  
28 and long-term changes. In this context, there is actually no difference between a “real” change  
29 and a gradual one: both are described by evolution behaviors. In this way, if an sensed area  $(x, y)$   
30 undergoes an abrupt change (*e.g.*, a clear cut or the building of a house), it will be treated as a  
31 particular temporal behavior, *i.e.*, this behavior will emerge in the classification if it is sufficiently  
32 represented in the dataset.

33 Due to the high spatial resolution of the future images, the geometrical information of the scene  
34 could also be considered in the classification process by using object-based approaches. To this  
35 end, a segmentation process is required to extract segments based on radiometric homogeneity.  
36 Once these segments are extracted, it is possible to characterize them using spatial/geometrical  
37 properties, to enhance the classification process. However, the integration of a segmentation step  
38 in a temporal classification framework remains an open-issue, since neither the mapping between  
39 mono-temporal segmentations, nor the temporal segmentation are resolved. A review of the avail-  
40 able literature on SITS analysis shows a lack of existing methods responding to this need. This  
41 article aims at addressing this issue by characterizing a pixel with spatial properties in order to  
42 improve the analysis of SITS.

43 This article is organized as follows. Section 2 gives an overview of existing methods for SITS  
44 analysis. Section 3 introduces our generic methodology for spatio-temporal analysis of SITS.  
45 Section 4 describes the experimental validation carried out with this methodology. Section 5  
46 presents the results obtained using the proposed methodology. Conclusions and perspectives will  
47 be found in Section 6.

## 48 **2. State of the art**

49 SITS allow the analysis, through observations of land phenomena, of a broad range of applica-  
50 tions such as the study of land-cover or the mapping of damage following a natural disaster. These  
51 changes may be of different types, origins and durations. For a detailed survey of these methods,  
52 the reader can refer to (Coppin et al., 2004; Lu et al., 2004).

53 In the literature, we find three main families of methods. *Bi-temporal analysis*, *i.e.*, the study of  
54 transitions, can locate and study abrupt changes occurring between two observations. Bi-temporal  
55 methods include image differencing (Bruzzone & Prieto, 2000), image ratioing (Jensen, 1981; Wu  
56 et al., 2005), image composition (Ouma et al., 2008) or change vector analysis (CVA) (Johnson  
57 & Kasischke, 1998; Bovolo, 2009; Bahirat et al., 2012). A second family of *mixed methods*,

58 mainly statistical methods, applies to two or more images. They include methods such as post-  
59 classification comparison (Foody, 2001), linear data transformation (PCA and MAF) (Howarth  
60 et al., 2006), image regression or interpolation (Kennedy et al., 2007) and frequency analysis (*e.g.*,  
61 Fourier, wavelets) (Andres et al., 1994; Tsai & Chiu, 2008). Then, we find methods dedicated  
62 to image time series and based on *radiometric trajectory analysis* (Jönsson & Eklundh, 2004;  
63 Verbesselt et al., 2010; Petitjean et al., 2011c; Kennedy et al., 2010; Lui & Cai, 2011).

64 Regardless of methods used in order to analyze satellite image time series, there is a gap  
65 between the amount of data composing these time series, and the ability of algorithms to ana-  
66 lyze them. Firstly, these algorithms are often dedicated to the study of a change in a scene from  
67 bi-temporal representation. Secondly, and this point is even more difficult to deal with, the geo-  
68 metrical/spatial properties of the data are rarely taken into account, except for the use of the pixel  
69 coordinates. Finally, High Spatial Resolution SITS have given rise to the need for spatially and  
70 temporally dedicated methods.

71 To improve the analyzing process by using the spatial relationships of the data, object-based  
72 methods have been recently proposed (Blaschke, 2010). In a first step, the images are seg-  
73 mented/partitioned into sets of connected regions. Then for each region, geometric features  
74 (Carleer & Wolff, 2006) (*e.g.*, area, elongation, smoothness) or even contextual ones (Gaetano  
75 et al., 2009; Bruzzone & Carlin, 2006; Kurtz et al., 2010) (*e.g.*, spatial context, multi-scale/multi-  
76 resolution attributes) are computed in order to characterize the regions. Finally, the regions are  
77 classified using these features (Herold et al., 2003).

78 Object-based methods have shown promising results in the context of single-image analysis.  
79 However, their extension/adaptation to SITS in order to exploit both the spatial and temporal  
80 information contained in these data remains an open-issue. Indeed, although several methods have  
81 been proposed in order to map segments from one image to another (Gueguen et al., 2006; Bovolo,  
82 2009), to directly build spatio-temporal segments (Fan et al., 1996; Moscheni et al., 1998; Tseng  
83 et al., 2009), or even to consider object-based features (Hall & Hay, 2003; Niemeyer et al., 2008;  
84 Hofmann et al., 2008; Schopfer et al., 2008; Tiede et al., 2011), their scalability to wide sensed  
85 areas and their robustness to local disturbance (temporally and spatially) remain problematic. The  
86 use of 3D-dedicated methods indeed requires a high temporal continuity; this constraint is however  
87 rarely fulfilled by SITS, where the average time-delay between two images is usually too high. As  
88 a consequence, the temporal continuity of the observed phenomena can not be assumed between  
89 samples. In addition, the irregular temporal sampling of the image series (due to operational  
90 constraints of remote sensing), would create a disparity of the spatio-temporal regions in terms of  
91 their informativity. For instance, a region spreading over four months should not have the same  
92 importance in the analysis, than a region spreading over a single sample (*i.e.*, built over a single  
93 image). Thus, this article focuses on mono-temporal spatial enrichment of the pixels, in order to  
94 loosen the constraint on the pseudo-continuity.

95 We therefore suggest to classify SITS as the radiometric evolution of sensed areas with time.  
96 Then, in order to take into account the spatial properties of the data, we propose to characterize  
97 each pixel with spatial and geometrical attributes obtained using a pre-segmentation step. This  
98 formulation allows the study of spatial characteristics over time while abstracting from the cor-  
99 respondence between segments since the data remains the pixel. Moreover, this formulation is  
100 aimed at obtaining accurate and reliable evolution behavior maps both by preserving the geomet-

101 rical details in the images and by properly considering the spatial context information.

102 This paradigm, using spatially characterized pixels, was previously introduced and studied for  
103 contextual analysis (Melgani & Serpico, 2002), multi-level segmentation of a single image (Bruz-  
104 zone & Carlin, 2006) and change detection in bi-temporal images (Bovolo, 2009). In all these  
105 application domains, such a paradigm has shown promising results. We propose, in this article, to  
106 extend it to the analysis of large SITS.

### 107 3. Spatio-temporal analysis methodology

108 In this section, we present the proposed approach, which is composed of five main steps that  
109 are sequentially applied:

- 110 A. Segmentation of the images;
- 111 B. Characterization of the regions;
- 112 C. Construction of the vector images;
- 113 D. Construction of the time series;
- 114 E. Classification of the time series.

115 These steps are fully described in the remainder of this section. The reader may also refer to  
116 Figure 1 for a visual outline of the workflow of the proposed approach. Let us first establish the  
117 terminology used in the remainder of this article. The term “sensed area” will be rather used than  
118 the term “pixel”, since the notion of “pixel” usually refers to a mono-temporal value, while we use  
119 in this work the  $(x, y)$  coordinates to locate a geographic area. Contrary to the mono-image case,  
120 these notions are not mixed up in the temporal case. Consequently, the term “sensed area” will be  
121 used to designate the evolution of the  $(x, y)$  geographic area with time, while the term “pixel” will  
122 be used to designate a sensed value in a particular image.

#### 123 *Input/Output*

124 Let us briefly define the input and the output of the proposed method.

125 **Input.** The method takes as input a series  $\mathcal{S}_{image} = \langle I^1, \dots, I^N \rangle$  of  $N$  ortho-rectified images of  
126 width  $W$  and height  $H$ . Let  $E = \llbracket 1, W \rrbracket \times \llbracket 1, H \rrbracket$  where  $\llbracket a, b \rrbracket$  denote the interval on  $\mathbb{Z}$ , bounded  
127 by  $a, b$ . The set  $E$  corresponds to the discretization of the continuous space (*i.e.*, the part of  $\mathbb{R}^2$ )  
128 which will be visualized in the images. Let  $\mathcal{B}$  be the number of bands of the images composing  
129 the series. Each multivalued (*i.e.*, with multiple bands) image  $I^n$  ( $n \in \llbracket 1, N \rrbracket$ ) can be seen as a  
130 function:

$$130 \quad I^n : \quad E \quad \rightarrow \quad \mathbb{Z}^{\mathcal{B}} \\ (x, y) \mapsto I_1^n(x, y), \dots, I_{\mathcal{B}}^n(x, y) \quad (1)$$

131 Note that the radiometric levels of the images do not have to be comparable from one image  
132 to another. Thus, images can be acquired by different sensors but must be of the same spatial  
133 resolution.

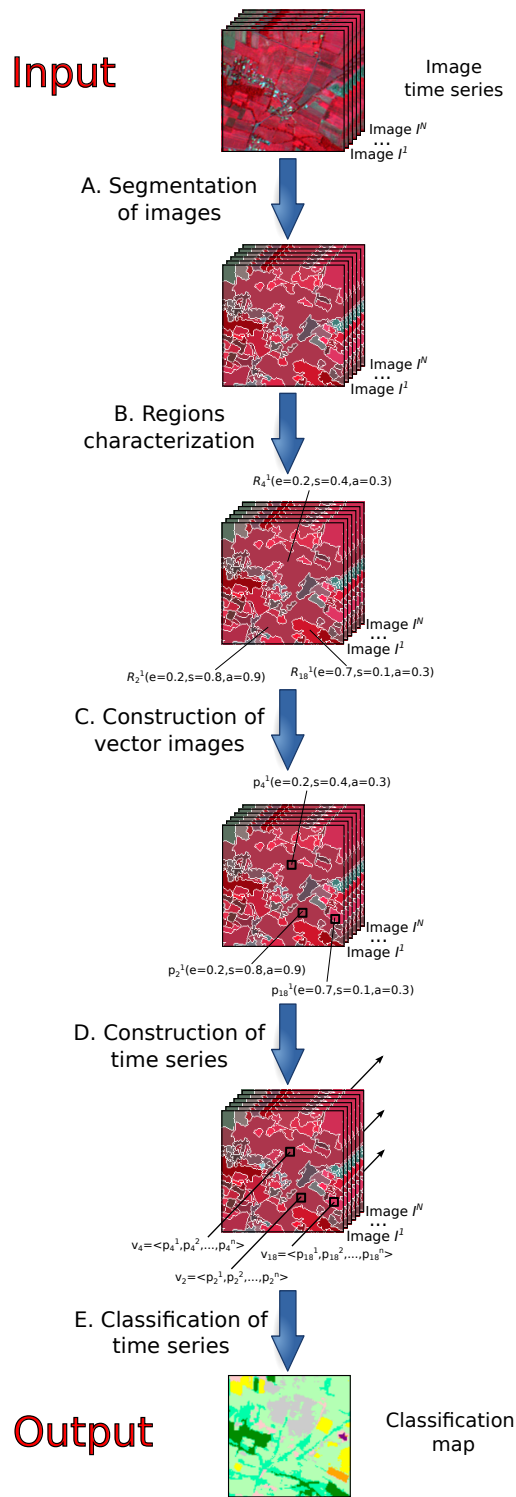


Figure 1: Workflow of the proposed approach that takes as input  $N$  images and provides as output a classification map of the sensed scene. Different symbols have been used to exemplify the steps of the proposed approach ( $R$  refers to regions,  $p$  refers to pixels,  $v$  refers to vectors of pixels built through time). For instance, the symbol  $R_4^1(e = 0.2, s = 0.4, a = 0.3)$  means that the region  $R_4^1$  is characterized by 3 feature values ( $e =$  elongation,  $s =$  smoothness,  $a =$  area).

134 **Output.** The method provides as output a classification of the sensed scene, where areas that  
 135 have evolved in a similar way are clustered. Such classification can be modeled by a label image  
 136  $I_C : E \rightarrow \llbracket 1, C \rrbracket$ , which associates to each sensed area  $(x, y)$  a class value  $I_C(x, y)$  among the  $C$   
 137 possible ones.

138 Each class of the classification is also modeled by a centroid sequence, which provides a  
 139 concise representation of the underlying evolution behavior. This extra information is however not  
 140 studied in this article.

### 141 3.1. Segmentation of images

142 A segmentation of a multivalued image  $I^n$  is a partition  $\mathfrak{S}^n = \{R_i^n\}_{i=1}^{\mathcal{R}^n}$  of  $\llbracket 1, W \rrbracket \times \llbracket 1, H \rrbracket$ ;  
 143 broadly speaking, the scene visualized in  $I^n$  is “decomposed” into  $\mathcal{R}^n$  distinct parts  $R_i^n$ , which are  
 144 supposed to present specific radiometric properties. We will denote  $R_i^n$  as a *region* of the image  $I^n$ .  
 145 To any segmented image  $I^n$ , we then associate a region image

$$I_R^n : \begin{array}{l} E \rightarrow \llbracket 1, \mathcal{R}^n \rrbracket \\ (x, y) \mapsto I_R^n(x, y) \end{array} \quad (2)$$

146 Such region image is a function that associates to each sensed area  $(x, y)$  a region label  $I_R^n(x, y)$   
 147 among the  $\mathcal{R}^n$  possible ones.

148 Once the  $N$  images have been segmented (producing  $N$  region images  $I_R^n$ , ( $n \in \llbracket 1, N \rrbracket$ )), it is  
 149 then possible to characterize each region of each segmentation by following the next step.

### 150 3.2. Regions characterization

151 Numerous features (spectral, geometrical, topological, *etc.*) can be computed for the regions of  
 152 a segmentation in order to characterize them. Each feature can be seen as a function  $F$  associating  
 153 to each region  $R_i^n$  ( $i \in \llbracket 1, \mathcal{R}^n \rrbracket$ ) of a segmentation  $\mathfrak{S}^n$  a corresponding feature value  $F(R_i^n) \in \mathbb{R}^\alpha$ .  
 154 Although the classical case corresponds to mono-dimensional features in  $\mathbb{R}$ , certain features can  
 155 be seen as multi-dimensional ones in  $\mathbb{R}^\alpha$  (*e.g.*, correlated textural features, multi-scale features).

$$F : \begin{array}{l} \llbracket 1, \mathcal{R} \rrbracket \rightarrow \mathbb{R}^\alpha \\ R_i^n \mapsto F(R_i^n) \end{array} \quad (3)$$

156 Once a region is characterized by a (multidimensional) feature value, it is then possible to affect  
 157 this value to all the pixels composing the region. Let  $\mathcal{Z}$  be the number of region-features chosen  
 158 to describe every sensed area  $(x, y)$  of every image.

### 159 3.3. Construction of vector images

160 At this step, each pixel of a multivalued image  $I^n$  can be characterized by two types of information:

- 161 - directly sensed values (*i.e.*,  $\mathcal{B}$  values, denoted  $I_b^n$  with  $b \in \llbracket 1, \mathcal{B} \rrbracket$ );
- 162 - region-associated values (*i.e.*,  $\mathcal{Z}$  values, denoted  $F_a$  with  $a \in \llbracket 1, \mathcal{Z} \rrbracket$ ).

163 All these values are normalized over the image time series by using the extrema values of the  
 164 attributes in the dataset. It is then possible to combine these features to build “enriched” pixels in  
 165 order to better characterize them. To process, a vector of features is created and associated to each  
 166 one of the pixels contained in the image  $I^n$ . Finally, by applying this step to each image of the  
 167 series, we build  $N$  vector images defined as:

$$V^n : E \rightarrow [0, 1]^{\mathcal{B}+\mathcal{Z}} \\ (x, y) \mapsto (I_1^n(x, y), \dots, I_B^n(x, y), F_1(I_R^n(x, y)), \dots, F_Z(I_R^n(x, y))) \quad (4)$$

### 168 3.4. Construction of time series

169 Let  $\mathcal{S}$  be the dataset built from the image time series.  $\mathcal{S}$  is the set of sequences defined as:

$$\mathcal{S} = \{ \langle V^1(x, y), \dots, V^N(x, y) \rangle \mid x \in \llbracket 1, W \rrbracket, y \in \llbracket 1, H \rrbracket \} \quad (5)$$

170 In these sequences, each element is  $(\mathcal{B} + \mathcal{Z})$ -dimensional. Since high-dimensional spaces do not  
 171 often provide the best solutions, we will study different subspaces of this  $(\mathcal{B} + \mathcal{Z})$ -dimensional  
 172 space in the experiment part (*e.g.*, time series where each pixel is characterized by a 5-tuple com-  
 173 posed of three directly sensed values and two region-associated values).

### 174 3.5. Classification of the time series

175 The extraction of relevant temporal behaviors from satellite image time series can be realized  
 176 using a classification algorithm. Once these time series have been built, it becomes possible to  
 177 classify them into different clusters/classes of interest. To this end, the proposed methodology  
 178 makes it possible to use either supervised or unsupervised classification algorithms.

179 A classification of a set of sequences  $\mathcal{S}$  is a partition  $\mathcal{C} = \{C_i\}_{i=1}^C$  of  $E$ ; broadly speaking, as each  
 180 temporal sequence is associated to a sensed area  $(x, y)$ , the whole scene can be “decomposed” into  
 181  $C$  distinct parts  $C_i$ , which are supposed to represent similar temporal evolution behaviors. We will  
 182 denote  $C_i$  as a *cluster/class*. The classification can be modeled by a label image  $I_C : E \rightarrow \llbracket 1, C \rrbracket$ ,  
 183 which associates to each sensed area  $(x, y)$  a class value  $I_C(x, y)$  among the  $C$  possible ones

$$I_C : E \rightarrow \llbracket 1, C \rrbracket \\ (x, y) \mapsto I_C(x, y) \quad (6)$$

## 184 4. Material and experimental settings

185 To assess the relevance of the proposed generic spatio-temporal analysis methodology, we have  
 186 applied it to the classification of agronomical areas. Starting from the different issues raised by  
 187 this applicative context, we show in this section how the proposed methodology can be used as a  
 188 potential solution to address them.

### 189 4.1. Applicative context: Crop monitoring

190 The analysis of agronomical areas is important for the monitoring of physical variables, in  
 191 order to give information to the experts about pollution, vegetation health, crop rotation, *etc.* This  
 192 monitoring is usually achieved through remote sensing. Indeed, by using classification processes,



193 satellite image time series actually provide an efficient way to monitor the evolution of the Earth's  
194 surface. Moreover, when the classes of interest are temporal (*e.g.*, wheat crop, maize crop), the  
195 time dimension of the data has to be taken into account by the classification algorithms. For  
196 instance, the reflectance levels of the maize crop and of the wheat crop are very similar while their  
197 temporal behaviors are quite different (*i.e.*, wheat grows earlier in the year than maize).

198 Thus, the usual strategy for land-cover mapping consists of classifying the temporal radiomet-  
199 ric profiles of the sensed areas  $(x, y)$ . With the arrival of SITS with high spatial resolution (HSR),  
200 it becomes necessary to use the spatial information held in these series, in order to either study  
201 the evolution of spatial features, or to help characterizing the different land-cover classes. Our  
202 experiments focus on the second point. The underlying idea is that several spatially-built features  
203 can be used in the classification process. For example, some crops are usually cultivated in smaller  
204 parcels than others, while having the same radiometric temporal behavior (*e.g.*, sunflower crop *vs.*  
205 wheat crop). Another (non restrictive) example could be the use of the smoothness of the regions,  
206 which could help, for instance, to distinguish between tree-crop and forest.

207 In the remainder of this section, we experimentally demonstrate how the proposed generic  
208 spatio-temporal analysis methodology can be instantiated to enable the use of such spatially-built  
209 features.

#### 210 4.2. Dataset description

211 We detail hereafter the main information concerning the images used for this work. The area of  
212 study is located near the town of Toulouse in the South West of France. 15 cloud-free FORMOSAT-2  
213 images sensed over the 2007 cultural year are analyzed. These images cover an area of 64 km<sup>2</sup>  
214 (1,000 × 1,000 pixels). The considered SITS is illustrated in Figure 2. One image of the series is  
215 given in Figure 3(a) while the temporal distribution of the sensed images is given in Figure 3(b).

216 From these images, we use the multi-spectral product at a spatial resolution of 8 m with the  
217 four bands Near-Infrared, Red, Green and Blue. Before being used in this work, the FORMOSAT-2  
218 products have been ortho-rectified (guaranteeing that a pixel  $(x, y)$  covers the same geographic  
219 area throughout the image series). All images also undergo processes in order to make the radio-  
220 metric pixel values comparable from one image to another. These processes consist of converting  
221 the digital counts provided by the sensor into a physical magnitude and in restoring their own  
222 contribution to the surface by correcting for atmospheric effects.

223 From the instrument radiometric model, digital numbers are first converted into reflectances  
224 (normalized physical quantities of solar irradiance). The absolute calibration coefficients used  
225 in this step come from the monitoring of FORMOSAT-2 sensors conducted by the French Space  
226 Agency (CNES). The inversion of the surface reflectance is then made by comparing the measured  
227 reflectance in simulations at the top of the atmosphere, carried out for atmospheric and geometric  
228 conditions of measurement. The elevation is taken into account by carrying out simulations for  
229 various altitudes, including a weighting of the atmospheric pressure and the amounts of aerosols  
230 and water vapor. The state of the atmosphere at the time of the sensing is in turn characterized  
231 using meteorological sources (National Centers for Environmental Prediction for the pressure and  
232 the humidity), using ozone data sources (*e.g.*, Total Ozone Mapping Spectrometer) and using  
233 aerosol data (*e.g.*, Sea-viewing Wide Field-of-view Sensor). Otherwise, climatological values are  
234 used. This procedure is detailed in (Hagolle et al., 2010).

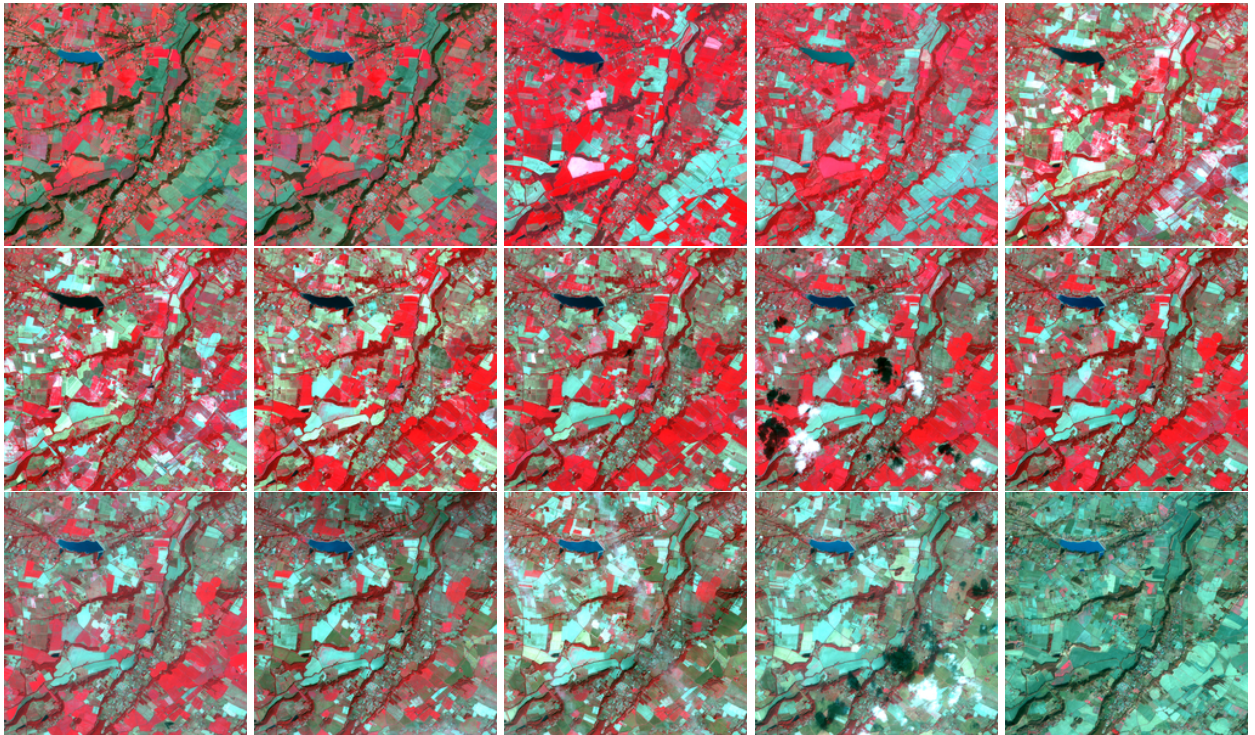


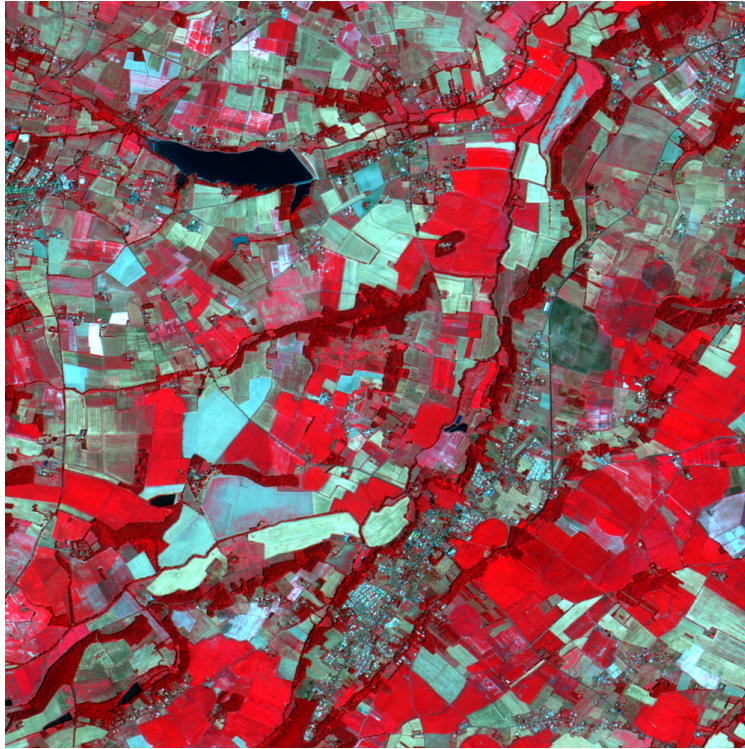
Figure 2: SITS considered in these experiments. The sensing distribution of these images is presented in Figure 3(b).

### 235 4.3. Experimental settings

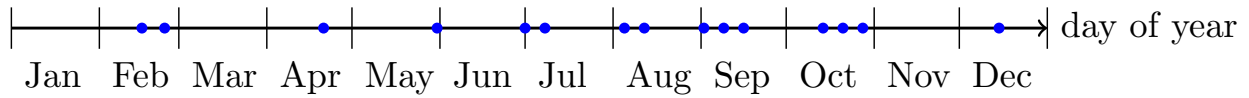
236 This section aims at describing how the proposed generic approach has been instantiated to  
 237 deal with the presented crop monitoring issue. However, we recall that the presented approach  
 238 is not limited to this instantiation. The five steps described in Section 3 have been performed as  
 239 follows.

240 *Segmentation of images.* The segmentation of HSR satellite images is not a trivial task since the  
 241 different objects of interest (and thematic ground areas) which are sensed by these images, cannot  
 242 be necessarily segmented at the same scale (*i.e.*, scale issue). For instance, the main environments,  
 243 such as urban areas, rural zones, or forests, can be identified at coarsest scales, while more detailed  
 244 structures, such as buildings and roads, will emerge at the finest ones (Blaschke, 2010). It is then  
 245 difficult to correctly segment all these thematic ground areas by using only one segmentation result.

246 For the last decade, it has been shown that hierarchical segmentation algorithms provide accu-  
 247 rate results adapted to process HSR images (Pesaresi & Benediktsson, 2001; Gaetano et al., 2009).  
 248 In particular, their combinations can provide an efficient way to deal with the scale issue (Akay  
 249 & Aksoy, 2008; Kurtz et al., 2011a,b). However, the parameters of such algorithms have to be  
 250 tuned according to the characteristics of the image modality (used as input) and the features of the  
 251 objects to be segmented. To avoid this parametrization problem (which falls outside the scope of  
 252 this article), we have chosen to use the Mean-Shift segmentation algorithm (Comaniciu & Meer,  
 253 2002) to segment each image of the series. Indeed, this algorithm is intuitive to configure and has  
 254 shown satisfactory results in the context of the segmentation of remote sensing images (Huang &



(a)



(b)

Figure 3: Presentation of the dataset. (a) One image from the series (August, 4<sup>th</sup> 2007). (b) Sensing distribution of images sensed over 2007. Each spot represents a sensed image (see Figure 2).

255 Zhang, 2008). Although we know that considering a single segmentation map for each image is,  
 256 in most of the cases, a sub-optimal approach (since the spatial arrangement of the objects in the  
 257 image is intrinsically hierarchical), we assume that when dealing with agricultural territories, the  
 258 fields could be efficiently extracted at similar scales and thus, by using only one segmentation map  
 259 per image. We plan to address this aspect in a future development of the work as stated in the  
 260 conclusions.

261 The Mean-Shift segmentation algorithm performs as follows. For a given pixel, this algo-  
 262 rithm builds a set of neighboring pixels within a given spatial radius and color range. The spa-  
 263 tial and color center of this set is then computed and the algorithm iterates with this new spatial  
 264 and color center. There are three main parameters: the spatial radius (denoted by  $h_s$ ) used for  
 265 defining the neighborhood, the range radius (denoted by  $h_r$ ) used for defining the interval in the  
 266 color space and the minimum size  $M$  for the regions to be kept after segmentation. We have  
 267 used the OTB implementation of the Mean-Shift algorithm. ORFEO Toolbox (OTB) is an open  
 268 source library of image processing algorithms developed by the French Space Agency (CNES).

269 <http://www.orfeo-toolbox.org>

270 In order to assess the robustness of the proposed approach with regard to the segmentation  
271 step, the influence of the segmentation parameters has been studied. Since the level of geometrical  
272 information extracted by the segmentation algorithm depends on its parametrization, we have run  
273 the algorithm using different configurations of the parameters. In practice, the minimum size  
274  $M$  of the regions has been fixed to  $M = 25$ , corresponding to the minimum expected size of  
275 the studied objects of interest. The ranges of the possible values for the other parameters  $h_s$   
276 and  $h_r$  have been scanned exhaustively with a quite low step ( $h_s \in \{1, 3, 5, \dots, 28, 30\}$  and  $h_r \in$   
277  $\{1, 5, 10, \dots, 55, 60\}$ ).

278 The underlying idea of this experiment is to study the influence of the segmentation parameters  
279 on the classification results. To this end, only the radiometric mean of the regions has been used  
280 to characterize the pixels for the classification (during the construction of the time series).

281 *Region characterization.* Several characteristics can be useful for the classification of agronomical  
282 scene. For instance, the size of the regions could be used to discriminate small/large fields, while  
283 the smoothness could be used to separate forest regions from fields. In this way, the following  
284 region-associated features have been computed:

- 285 - the mean of the infra-red band of the region ( $F_{NIR}$ );
- 286 - the mean of the red band of the region ( $F_R$ );
- 287 - the mean of the green band of the region ( $F_G$ );
- 288 - the mean of the blue band of the region ( $F_B$ );
- 289 - the area of the region ( $F_{Area}$ );
- 290 - the elongation of the region ( $F_{Elong.}$ );
- 291 - the smoothness of the region ( $F_{Smooth.}$ );
- 292 - the compactness of the region ( $F_{Comp.}$ ).

293 The elongation is computed as the highest ratio between the width and the length of several bound-  
294 ing boxes (computed for different directions, *i.e.*, each  $\pi/8$ ). The smoothness is computed as the  
295 ratio between the perimeter of the morphologically opened region and the original region. To this  
296 end, we use a square-shaped opening structuring element invariant to the scale (*i.e.*, with a size de-  
297 pending on the area of the original region). The size of the structuring element was set to  $\sqrt{F_{Area}}$ .  
298 The compactness is computed as the square root of the area of the region multiplied by the length  
299 of the perimeter of the region.

300 *Construction of vector images.* As explained previously, each pixel composing the SITS can be  
301 characterized by two types of information: directly sensed values (denoted as  $I_{NIR}, I_R, I_G, I_B$ ), and  
302 region-associated values (denoted as  $F_{NIR}, \dots, F_{Smooth.}$ ). All the values are normalized in  $[0, 1]$ ,  
303 attribute by attribute over the series. This allows each attribute to be of comparable weight for the  
304 classification step.

305 *Construction of time series.* In order to find the best separation of thematic classes and to assess  
306 (globally and independently) the interest of the different contextual attributes, we tested several

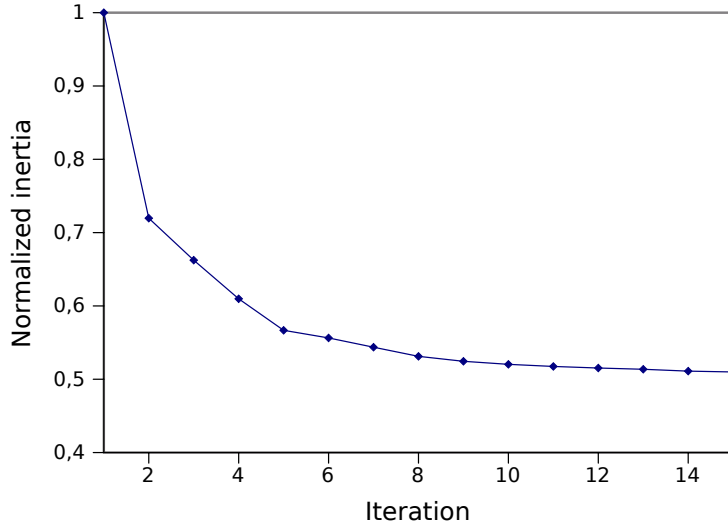


Figure 4: Convergence of the K-MEANS algorithm over 15 iterations corresponding to Experiment 16.

307 combinations of twelve attributes over the time series. All the possible combinations of the spa-  
 308 tial attributes ( $F_{Area}$ ,  $F_{Elong.}$ ,  $F_{Smooth.}$ ,  $F_{Comp.}$ ) were tested with either the pixel radiometric values  
 309 ( $I_{NIR}$ ,  $I_R$ ,  $I_G$ ,  $I_B$ ), or the mean region ones ( $F_{NIR}$ ,  $F_R$ ,  $F_G$ ,  $F_B$ ). The resulting 32 combinations are  
 310 presented in Table 1. In particular, the combination  $\star$  (only the pixel radiometric values without  
 311 any region-associated feature) represents the “classical” combination for pixel-based classification  
 312 of SITS.

313 *Classification of time series.* Classification problems are usually addressed using supervised or  
 314 unsupervised algorithms. Supervised classification algorithms require training examples to learn  
 315 the classification model. In our case, as we want to demonstrate the relevance of the proposed  
 316 data representation, the choice and the suitability of the examples would create a bias, which  
 317 would make difficult to identify the benefits provided by the spatial features. In this way, choos-  
 318 ing an unsupervised classification step allows us to highlight the consistency of the proposed  
 319 approach, without being influenced by several issues linked to the evaluation of supervised ap-  
 320 proaches (choice of the algorithm, cross-validation, building and sampling of the training set,  
 321 *etc.*). We have then applied the classical K-MEANS clustering algorithm (MacQueen, 1967) to clas-  
 322 sify the time series previously constructed. The distance used to compare the time series of  $\mathcal{S}$   
 323 is the Euclidean distance. Note that other distances (and more relevant temporal ones (Petitjean  
 324 et al., 2011b,a, 2012)) could also be used.

325 The K-MEANS algorithm has been used with as many classes (see Table 2) as in the reference  
 326 map (*i.e.*, 25 seeds), and with 15 iterations. Figure 4 shows that the process has converged after-  
 327 wards. Note that any clustering algorithm dealing with numerical data could also be used.

#### 328 4.4. Validation

329 To assess the quality and the accuracy of the results, the classification maps obtained have been  
 330 compared to:

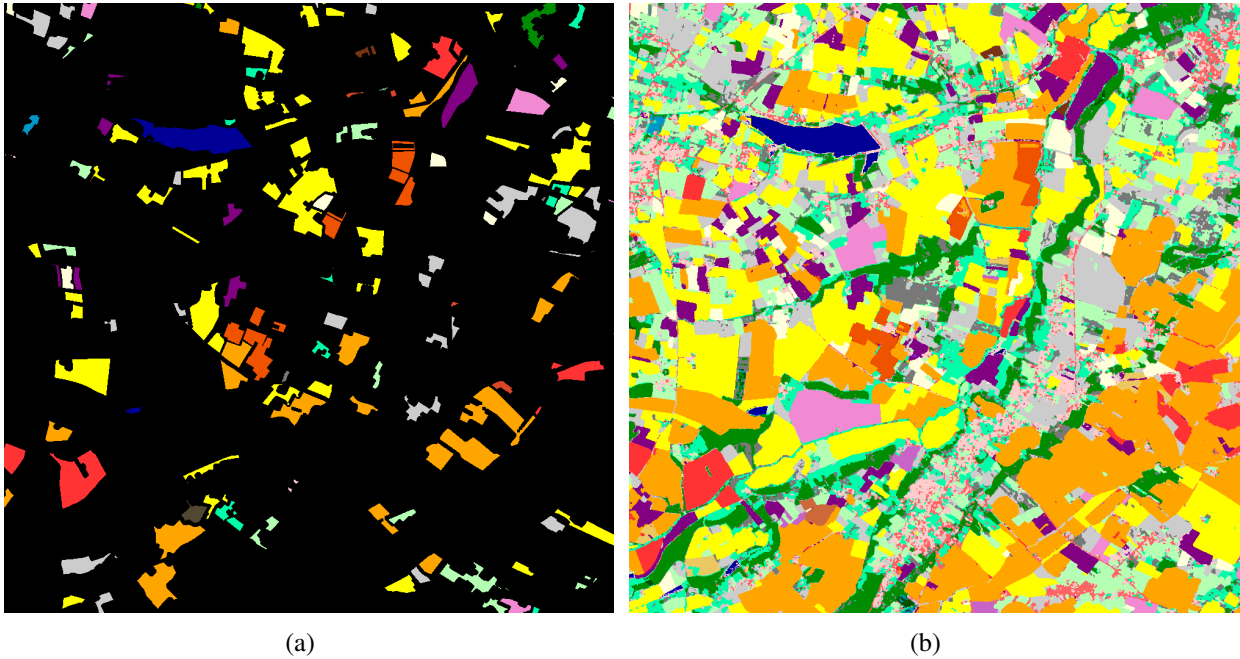


Figure 5: Land cover reference maps of the 2007 cultural year. (a) Ground truth (covering a partial part of the studied area) related to a field survey produced by the European Common Agricultural Agency. (b) Land cover reference map (covering the totality of the studied area) produced by the method described in (Idbraim et al., 2009).

- 331 1. a field survey (*i.e.*, ground-truth) of the 2007 cultural year (produced by the European En-  
332 vironment Agency; see [http://ec.europa.eu/agriculture/index\\_en.htm](http://ec.europa.eu/agriculture/index_en.htm) for more  
333 details about the Common Agricultural Policy.) covering a partial part of the studied area  
334 (Figure 5(a));
- 335 2. a land cover reference map (produced by a supervised classification method described in (Id-  
336 braim et al., 2009)) covering the totality of the studied area (Figure 5(b)).

337 Note that these two reference maps reflect the temporal behavior of the considered crops over the  
338 2007 cultural year and do not reflect a static land cover state (*i.e.*, representing a single snapshot  
339 of the scene at a particular date). Such property is necessary since we want to assess the accuracy  
340 of temporal classification results. We also want to underline that, through the year, the land cover  
341 types do not change (*i.e.*, no crop rotation). This fact justifies why the considered classes are  
342 designated by static terms (*e.g.*, corn, wheat, meadow) instead of being described by dynamic  
343 ones (*e.g.*, the class “bare soil  $\rightarrow$  growth of corn  $\rightarrow$  harvest”).

344 The classification maps obtained have been compared to these maps using several evaluation  
345 indexes. To assess the global accuracy of the obtained classification results, we have computed  
346 respectively:

- 347 - the average F-measure  $\overline{\mathcal{F}}$ ;
- 348 - the Kappa index  $\mathcal{K}$ ;
- 349 - the overall classification accuracy  $\mathcal{A}$ .

The average F-measure  $\overline{\mathcal{F}}$  corresponds to the mean, for each class, of the F-measures obtained. To this end, for each thematic class, the best corresponding clusters (in terms of partitions) were extracted. Then, we have computed: the percentage of false positives (denoted by  $f^{(p)}$ ), the percentage of false negatives (denoted by  $f^{(n)}$ ) and the percentage of true positives (denoted by  $t^{(p)}$ ). These measures are used to estimate the precision  $\mathcal{P}$  and the recall  $\mathcal{R}$  of the results obtained by using the proposed method:

$$\mathcal{P} = \frac{t^{(p)}}{t^{(p)} + f^{(p)}} \quad \text{and} \quad \mathcal{R} = \frac{t^{(p)}}{t^{(p)} + f^{(n)}} \quad (7)$$

350 For each experiment, we have then computed the geometrical mean  $\overline{\mathcal{P}}$  of the precisions obtained  
 351 and the geometrical mean  $\overline{\mathcal{R}}$  of the recalls obtained. Finally, we have computed the mean F-  
 352 measure  $\overline{\mathcal{F}}$  which is the harmonic mean of the mean precision and the mean recall:

$$\overline{\mathcal{F}} = 2 \cdot \frac{\overline{\mathcal{P}} \cdot \overline{\mathcal{R}}}{\overline{\mathcal{P}} + \overline{\mathcal{R}}} \quad (8)$$

353 The computation of these class-specific indexes requires the matching of classes of interest with  
 354 clusters extracted by the unsupervised classification approach. To this end, we have used an au-  
 355 tomatic strategy, which consists of selecting the clusters that maximize the overlapping with the  
 356 corresponding class.

357 To assess the global relevance of the results, we have also computed the Kappa index (Congal-  
 358 ton, 1991)  $\mathcal{K}$ , which is a measure of global classification accuracy:

$$\mathcal{K} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (9)$$

359 where  $\Pr(a)$  is the relative agreement among the observers, and  $\Pr(e)$  is the hypothetical probabili-  
 360 ty of chance agreement. The Kappa index takes values in  $[0, 1]$  and decreases as the classification  
 361 is in disagreement with the ground-truth map. We have computed this index as follows. The  
 362 approach consists of considering all point couples  $(\mathbf{x}_1, \mathbf{x}_2) = ((x_1, y_1), (x_2, y_2))$  and seeing the con-  
 363 figuration of these two points in each partition (the clustering result and the ground-truth). There  
 364 are four possible configurations; for each one, a counter is associated and incremented each time  
 365 a configuration appears:

- 366 1.  $\mathbf{x}_1$  and  $\mathbf{x}_2$  belong to the same partition both in the clustering and in the reference map (counter  
 367  $s_s$ );
- 368 2.  $\mathbf{x}_1$  and  $\mathbf{x}_2$  belong to the same partition in the clustering but not in the reference map (counter  
 369  $s_d$ );
- 370 3.  $\mathbf{x}_1$  and  $\mathbf{x}_2$  belong to the same partition in the reference map but not in the clustering (counter  
 371  $d_s$ );
- 372 4.  $\mathbf{x}_1$  and  $\mathbf{x}_2$  belong to the same partition neither in the reference map nor in the clustering (counter  
 373  $d_d$ ).

374 Thus, the Kappa index is computed with:

$$\Pr(a) = \frac{s_s + d_d}{s_s + s_d + d_s + d_d} \quad (10)$$

375 and

$$\Pr(e) = \frac{(s_s + s_d) \cdot (s_s + d_s) + (s_d + d_d) \cdot (d_s + d_d)}{(s_s + s_d + d_s + d_d)^2} \quad (11)$$

376 Note that the Kappa index is an agreement measure between two partitions and thus does not  
377 require to “align” the clusters with the reference classes.

378 To assess separately the accuracy of each thematic class, we also provide (for each one of these  
379 classes) the precision  $\mathcal{P}$ , the recall  $\mathcal{R}$  and their averages.

## 380 5. Results

381 This section presents the results obtained with the proposed contextual approach in the context  
382 of the multi-temporal analysis of agronomical areas. The first sub-section describes the study  
383 of the influence of the segmentation step on the obtained classification results. The second sub-  
384 section proposes an exhaustive analysis of the interest of the different contextual attributes for  
385 multi-temporal analysis. Finally, the third sub-section presents an experimental study about the  
386 time complexity.

### 387 5.1. Influence of the segmentation step

388 The graph represented in Figure 6 summarizes the accuracy scores (mean F-measure  $\overline{\mathcal{F}}$  values)  
389 of the classification results obtained as a function of the parameters of the segmentation algorithm  
390 (the spatial radius  $h_s$  and the range radius  $h_r$ ). For each series of resulting segmentations, the classi-  
391 fication is obtained by using the radiometric mean of the regions to characterize the pixels through



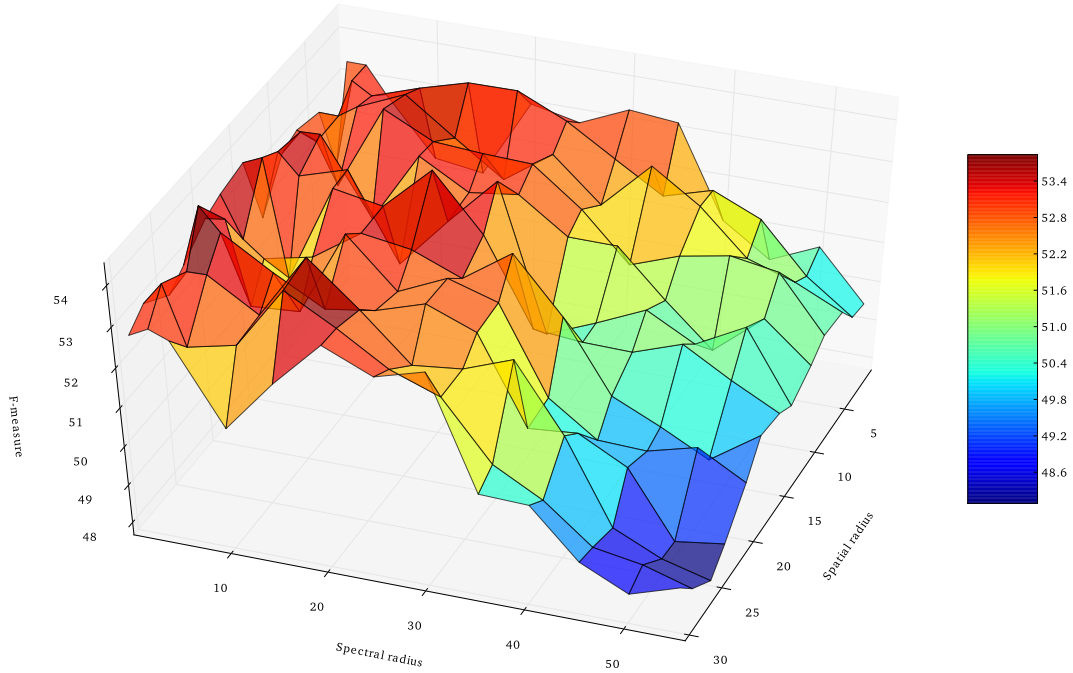


Figure 6: Influence of the parameters of segmentation (the spatial radius  $h_s$  and the range radius  $h_r$ ) on the classification results. The accuracy of the classification results obtained is assessed using the mean F-measure  $\overline{F}$  computed with the land-cover reference map presented in Figure 5(b).

392 time. The best scores obtained (and thus, the best configurations of parameters) correspond to  
 393 the orange-red area while the worst ones correspond to the green-blue area. The large size of  
 394 the orange-red area allows us to assess the robustness of the approach to the segmentation step.  
 395 As reference, the base-experiment  $\star$  obtained a F-measure score of 51.3% (corresponding to the  
 396 green area), which is lower than most of the scores obtained by using the radiometric mean of the  
 397 regions. For the remainder of this experimental study, the configuration ( $h_s = 10, h_r = 15, M = 25$ )  
 398 has then been kept.

399 In order to visually confirm the choice of this configuration, Figure ?? illustrates different seg-  
 400 mentation results obtained on an extract of an image. One can note that the parameters configura-  
 401 tion ( $h_s = 3, h_r = 3, M = 25$ ) provides over-segmented results while the parameters configura-  
 402 tion ( $h_s = 10, h_r = 15, M = 25$ ) provides satisfactory results for the extraction of agricultural areas.

## 403 5.2. Results analysis

404 Table 1 summarizes the F-measure values, the Kappa values and the overall classification  
 405 accuracy values obtained for the experiments with several subsets of attributes. Experiment  $\star$   
 406 gives the reference score obtained by a pixel-based classification of the SITS. From this table, one  
 407 can first note that these *baseline* scores are quite high, demonstrating the relevance of the temporal  
 408 dimension for land-cover classification.



























409 These experiments show that the radiometrical mean values of the regions ( $F_{NIR}, F_R, F_G, F_B$ ),  
 410 as well as the smoothness and the compactness of the regions ( $F_{Smooth}, F_{Comp}$ ), are the more rel-  
 411 evant features for the classification of this studied area. On the contrary, the area feature  $F_{Area}$

Table 1: Results of the experiments.

Experiment	Ground truth			Reference map		
	$\overline{\mathcal{F}}$	$\mathcal{K}$	$\mathcal{A}$	$\overline{\mathcal{F}}$	$\mathcal{K}$	$\mathcal{A}$
★ $I_{NIR}, I_R, I_G, I_B$	69.6	61.9	74.2	51.3	44.2	56.0
1 $I_{NIR}, I_R, I_G, I_B, F_{Comp.}$	<b>70.5</b>	60.9	73.7	50.8	43.7	55.3
2 $I_{NIR}, I_R, I_G, I_B, F_{Smooth.}$	69.1	59.7	72.5	<b>51.5</b>	<b>45.2</b>	<b>56.1</b>
3 $I_{NIR}, I_R, I_G, I_B, F_{Smooth.}, F_{Comp.}$	<b>72.0</b>	59.1	73.6	51.1	<b>44.8</b>	54.4
4 $I_{NIR}, I_R, I_G, I_B, F_{Elong.}$	63.2	52.0	68.6	45.7	37.4	51.1
5 $I_{NIR}, I_R, I_G, I_B, F_{Elong.}, F_{Comp.}$	63.2	49.5	67.0	46.0	36.3	50.9
6 $I_{NIR}, I_R, I_G, I_B, F_{Elong.}, F_{Smooth.}$	62.9	49.5	67.5	46.3	36.2	50.7
7 $I_{NIR}, I_R, I_G, I_B, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	64.6	49.3	68.3	46.4	36.1	50.6
8 $I_{NIR}, I_R, I_G, I_B, F_{Area}$	67.9	56.2	71.2	50.5	44.2	53.7
9 $I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Comp.}$	68.0	55.3	70.7	50.4	<b>44.6</b>	53.7
10 $I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Smooth.}$	67.2	56.3	71.1	50.8	43.3	53.4
11 $I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Smooth.}, F_{Comp.}$	67.3	56.1	71.3	50.3	43.2	53.3
12 $I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}$	65.1	52.1	69.1	47.4	38.3	52.0
13 $I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}, F_{Comp.}$	64.3	51.2	68.9	46.5	39.0	52.0
14 $I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}, F_{Smooth.}$	64.8	51.7	69.0	46.9	38.1	51.5
15 $I_{NIR}, I_R, I_G, I_B, F_{Area}, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	64.4	50.0	69.0	46.6	38.8	51.7
16 $F_{NIR}, F_R, F_G, F_B$	<b>72.7</b>	<b>63.0</b>	<b>75.7</b>	<b>52.1</b>	<b>44.5</b>	<b>56.0</b>
17 $F_{NIR}, F_R, F_G, F_B, F_{Comp.}$	<b>70.1</b>	58.2	73.4	50.8	<b>45.6</b>	55.7
18 $F_{NIR}, F_R, F_G, F_B, F_{Smooth.}$	<b>72.8</b>	<b>66.1</b>	<b>77.2</b>	<b>52.4</b>	<b>45.2</b>	55.9
19 $F_{NIR}, F_R, F_G, F_B, F_{Smooth.}, F_{Comp.}$	<b>71.1</b>	60.9	72.8	50.2	<b>45.2</b>	55.7
20 $F_{NIR}, F_R, F_G, F_B, F_{Elong.}$	64.6	52.6	69.1	45.3	37.0	50.3
21 $F_{NIR}, F_R, F_G, F_B, F_{Elong.}, F_{Comp.}$	65.0	52.5	68.9	45.3	37.0	50.4
22 $F_{NIR}, F_R, F_G, F_B, F_{Elong.}, F_{Smooth.}$	65.1	50.8	68.7	46.0	36.6	50.0
23 $F_{NIR}, F_R, F_G, F_B, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	65.0	48.4	67.8	45.4	36.1	49.6
24 $F_{NIR}, F_R, F_G, F_B, F_{Area}$	67.9	53.9	69.5	50.6	44.0	54.7
25 $F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Comp.}$	66.3	52.4	68.8	49.9	43.0	53.8
26 $F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Smooth.}$	66.7	52.3	69.4	50.3	43.8	54.0
27 $F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Smooth.}, F_{Comp.}$	65.8	52.5	69.3	49.7	43.6	54.0
28 $F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}$	65.1	52.8	68.7	47.8	39.1	52.4
29 $F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}, F_{Comp.}$	64.8	53.0	68.7	47.8	38.9	52.6
30 $F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}, F_{Smooth.}$	64.6	51.8	67.9	47.2	38.3	51.8
31 $F_{NIR}, F_R, F_G, F_B, F_{Area}, F_{Elong.}, F_{Smooth.}, F_{Comp.}$	63.7	49.6	67.7	47.0	38.1	51.8

The scores that outperform the ones obtained with Experiment ★ (*i.e.*, the reference scores obtained by a pixel-based classification of the SITS) are shown in boldface.

Table 2: Detailed results for Experiment 18.

Class	Colour	Ground truth				Reference map			
		# ( $\times 10^3$ )	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	# ( $\times 10^3$ )	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
corn		25	89,5	94,3	91,8	192	91,5	83,8	87,5
wheat		35	78,3	91,4	84,3	179	67,2	86,5	75,6
temp. meadow		6	28,3	59,3	38,3	104	34,7	64,9	45,2
fallow land		13	61,5	78,1	68,8	104	28,9	48,3	36,2
meadow		3	12,8	63,4	21,3	81	22,7	72,6	34,6
broad-leaved tree		1	79,4	91,2	84,9	77	58,8	95,0	72,6
wild land		< 0.5	2,4	59,8	4,6	46	26,4	37,4	31,0
sunflower		5	54,4	63,6	58,6	45	48,7	58,4	53,1
dense housing		< 0.5	0,9	12,5	1,7	36	31,3	58,7	40,8
housing			<b>n/a</b>			33	15,2	56,1	23,9
barley		2	9,2	57,5	15,9	27	10,4	50,5	17,3
soybean		9	77,6	69,1	73,1	23	46,5	65,6	54,4
rape		3	15,4	68,8	25,2	21	18,9	91,1	31,4
corn for silage		7	83,6	99,1	90,7	9	30,6	95,0	46,2
lake		6	100,0	98,9	99,5	9	86,5	96,0	91,0
non-irrigated corn			<b>n/a</b>			6	3,1	18,0	5,2
pea			<b>n/a</b>			2	1,9	43,4	3,6
sorghum II			<b>n/a</b>			2	5,0	79,4	9,4
eucalyptus		< 0.5	18,2	99,5	30,8	1	1,7	9,8	2,9
conifer			<b>n/a</b>			1	1,3	9,8	2,3
sorghum		< 0.5	4,1	70,2	7,7	1	1,4	33,7	2,6
specific surface			<b>n/a</b>			< 0.5	1,4	39,2	2,6
water			<b>n/a</b>			< 0.5	2,1	81,5	4,1
mineral surface			<b>n/a</b>			< 0.5	0,9	88,8	1,9
gravel pit			<b>n/a</b>			< 0.5	1,1	99,1	2,1
poplar tree		< 0.5	8,3	100,0	15,3	< 0.5	0,4	25,0	0,8
Average		<b>n/a</b>	64,5	83,6	72,8	<b>n/a</b>	42,1	69,2	52,4

The symbol **n/a** means that the considered value is either not available or not relevant. The symbol # corresponds to the cardinal (number of pixels) of the thematic class (we recall that each image is composed of  $1,000 \times 1,000$  pixels).

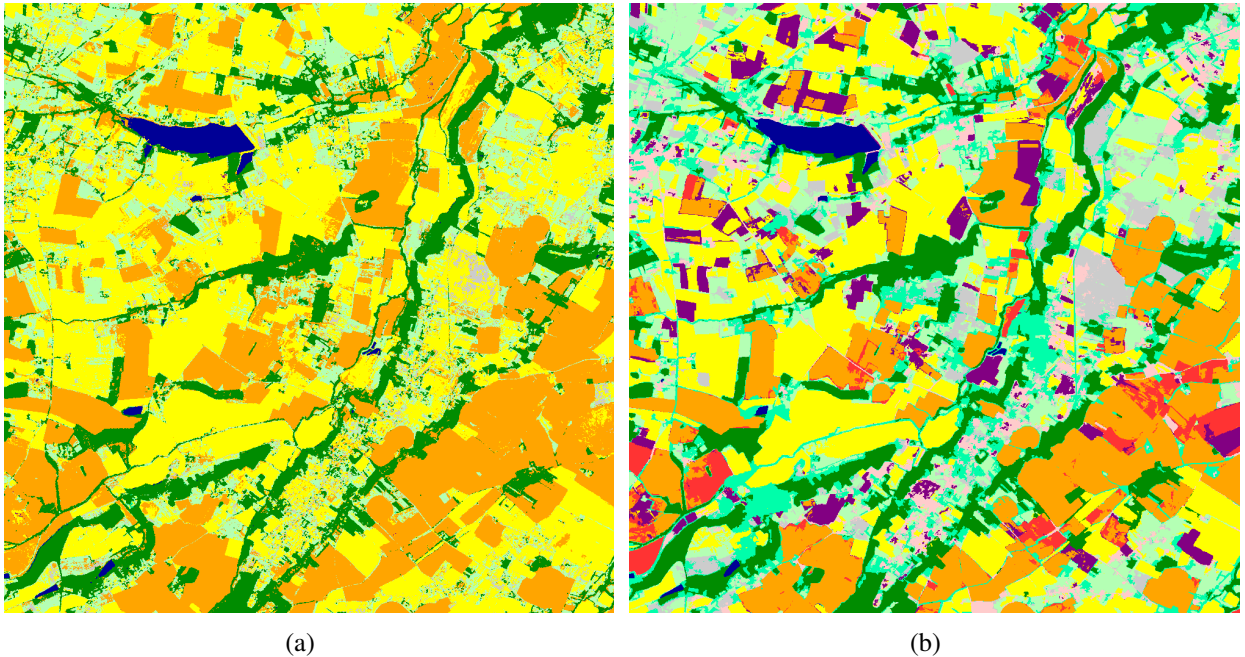


Figure 8: Clustering maps obtained on the satellite image time series. (a) Result obtained with a naive fusion approach. (b) Result obtained with the proposed method (Experiment 18). Note that these maps have been recolored according to the corresponding land cover reference map (Figure 5(b)).

412 seems not relevant for this dataset; a possible reason could be that each crop class contains dif-  
 413 ferent sizes of fields. A similar observation can be made for the elongation feature  $F_{Elong}$ . This  
 414 being so, these observations are not questioning the interest of the approach, since these spatial  
 415 characteristics (and others) could be used in other application cases. In our case, the best result has  
 416 been obtained with the use of the radiometric mean of the regions combined with the smoothness  
 417 feature (Experiment 18).

418 In order to visually assess the results provided by the proposed method, Figure 8(b) shows  
 419 the best clustering result obtained on sequences from 2007. We also provide in Figure 8(a) the  
 420 result provided by a naive fusion classification approach. This approach performs by fusing (with  
 421 a majority vote) the different clustering maps obtained independently for each image of the se-  
 422 ries. A visual comparison between these two results directly emphasizes the potential of using  
 423 a pixel-enriched temporal classification approach instead of a naive temporal fusion one. Actu-  
 424 ally, considering the temporal dimension of the data makes it possible to obtain more accurate  
 425 classification results.

426 To separately assess the accuracy of each thematic class, we also provide for the best clustering  
 427 result obtained (Experiment 18), the precision  $\mathcal{P}$ , the recall  $\mathcal{R}$ , the F-measure  $\mathcal{F}$  as well as their  
 428 averages (see Table 2). For comparison purpose, the confusion matrix obtained by comparing this  
 429 result to the considered land cover reference map is provided in Table 3. From these two tables,  
 430 one can note that most of the major considered temporal classes have been correctly extracted by  
 431 the proposed approach. Table 2 highlights that the proposed approach provides results with high  
 432 values of precision, recall, and F-measure for most of the extracted classes. For instance these

Table 3: Confusion matrix obtained by comparing the result of Experiment 18 to the considered land-cover reference map (Figure 5(b)).

Class	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	card
corn	7	84	4	4	-	1	-	-	-	-	19 %
wheat	-	-	2	1	2	1	6	2	87	-	18 %
temporary meadow	-	-	15	-	3	3	53	12	14	-	10 %
fallow land	-	-	18	2	10	4	34	17	14	-	10 %
meadow	-	-	35	-	2	23	31	6	2	-	8 %
broad-leaved tree	-	-	4	-	-	96	-	-	-	-	8 %
wild land	-	-	29	-	-	42	13	14	1	-	5 %
sunflower	7	7	11	58	10	1	1	1	3	-	4 %
dense housing	-	-	15	15	53	2	6	4	3	2	4 %
housing	-	-	29	2	32	11	17	6	3	-	3 %
barley	-	-	2	9	4	-	8	-	75	-	3 %
soybean	66	23	6	3	2	-	-	-	-	-	2 %
rape	-	-	1	-	1	-	3	-	95	-	2 %
corn for silage	-	47	2	49	-	-	-	-	-	-	1 %
lake	-	-	3	-	1	-	-	-	-	96	1 %
non-irrigated corn	13	20	28	3	5	13	1	17	-	-	1 %

The thematic classes covering less than one percent of the sensed surface are not represented in the matrix.

433 values reach approximatively 85% for the corn and wheat classes which are the most represented  
434 ones. Furthermore, the confusion matrix obtained shows that these two temporal classes are mainly  
435 regrouped in two clusters by the K-MEANS algorithm. The same observation can be given for the  
436 broad-leaved tree class. Such comparisons enable to assess the accuracy of the classification results  
437 provided by the proposed pixel-enriched approach. Note that, as the proposed approach provides  
438 a clustering of the sensed area, no one-to-one mapping between thematic classes and clusters is  
439 guaranteed. In this way, it is not possible to provide statistical accuracies from this matrix. For  
440 instance, cluster 8 is predominantly representing the wheat class, but also represents the barley  
441 and rape classes. In fact, this cluster represents the broader class of winter crops (*i.e.*, of higher  
442 semantic level), precisely composed of these three classes.

443 Moreover, Figure 9 focuses on a restricted area in order to visualize the differences between  
444 the pixel-based approach and the proposed pixel-enriched approach. One can see that, in the  
445 details, the land-cover map obtained with the proposed pixel-enriched approach is spatially more  
446 consistent and regular than the result obtained with the pixel-based approach. Furthermore, one  
447 can note that the orange and yellow classes, corresponding respectively to corn and wheat crop  
448 fields, as well as the dark green class corresponding to hardwoods, are well separated. More  
449 generally, these results demonstrate visually the relevance of the proposed pixel-enriched approach  
450 compared to the pixel-based analysis.

451 Finally, in order to statistically study the correlation of the considered features, a correlation

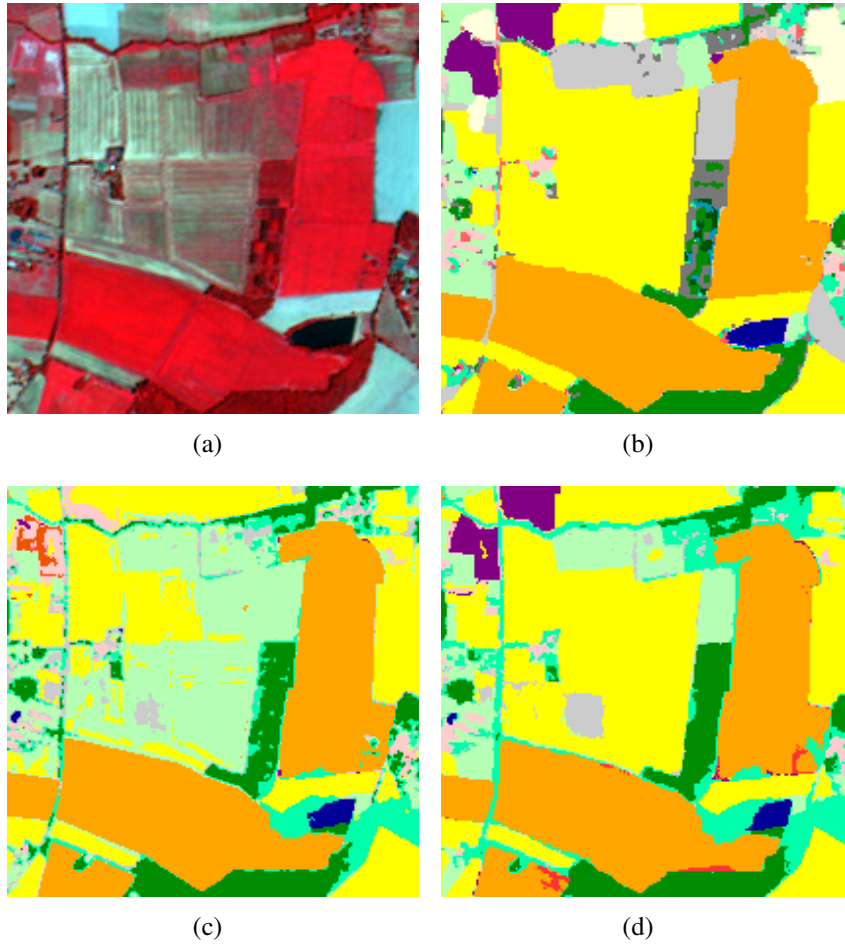


Figure 9: Extract of the results provided by the proposed method carried out on the satellite image time series. (a) Zoom on the considered ground surface in one image of the series. (b) Zoom on the land cover reference map (Figure 5(b)). (c) Zoom on the clustering map obtained with Experiment  $\star$ . (d) Zoom on the clustering map obtained with Experiment 18.

452 matrix between these features has been computed (Table 4). To this end, all pixels of all images  
 453 were characterized by the spatial features (computed on the segmentation ( $h_s = 10, h_r = 15, M =$   
 454  $25$ )). Not surprisingly, the radiometric features  $F_R, F_G$  and  $F_B$  are highly correlated (due to the  
 455 similar reflectances of usual sensed objects in these radiometric bands). This matrix also shows  
 456 that the spatial features are generally not correlated, except for the couple ( $F_{Elong.}, F_{Comp.}$ ).

### 457 5.3. Computation time study

458 As it is quite difficult to provide a relevant theoretical complexity study of the proposed  
 459 methodology, we present hereafter an experimental evaluation of the complexity.

460 Table 5 provides the run-time and the memory usages for the processing of the images con-  
 461 tained in the studied dataset sensed over the 2007 cultural year. Experiments have been run on an  
 462 Intel® Core™2 Quad running at 2.4 GHz with 8 GB of RAM. The algorithms have been imple-  
 463 mented using the Java programming language and different threading strategies. From Table 5,

Table 4: Correlation matrix of the features corresponding to the segmentation parameters ( $h_s = 10, h_r = 15, M = 25$ ).

Feature	$F_{NIR}$	$F_R$	$F_G$	$F_B$	$F_{Area}$	$F_{Elong.}$	$F_{Smooth.}$	$F_{Comp.}$
$F_{NIR}$	1	-0.21	-0.05	-0.15	-0.04	-0.2	-0.07	-0.21
$F_R$		1	0.96	0.93	-0.06	0.24	-0.05	0.11
$F_G$			1	0.96	-0.04	0.23	-0.04	0.12
$F_B$				1	-0.04	0.23	-0.04	0.12
$F_{Area}$					1	-0.52	0.34	-0.33
$F_{Elong.}$						1	-0.12	0.83
$F_{Smooth.}$							1	0.02
$F_{Comp.}$								1

Table 5: Run-time and memory usage for the processing of the considered dataset.

Step	Runtime	Memory (RAM)
A. Segmentation of the images	9 min 21 s	1.1 GB
B. Characterisation of the regions	2 min 18 s	2.3 GB
C. Construction of the vector images	<b>n/a</b>	2.3 GB
D. Construction of the time series	<b>n/a</b>	2.3 GB
E. Classification of the time series	1 min 3 s	1.6 GB
Total	$\approx$ 13 min	$\approx$ 2.3 GB

The symbol **n/a** means that the considered run-time is not significant.

464 one can note that the proposed approach makes it possible to classify a whole HSR SITS in less  
 465 than 15 minutes. Furthermore the memory consumption remains tractable since it does not exceed  
 466 2.3 GB when processing a dataset composed of 15 images of  $1,000 \times 1,000$  pixels. For compar-  
 467 ison purpose, the classification of the same HSR SITS, without considering the spatial context of  
 468 the pixels (Experiment  $\star$ ), requires less than 2 minutes.

## 469 6. Conclusion

470 This article has introduced a novel approach for the analysis of satellite image time series. The  
 471 originality of this approach lies in its consideration of spatial relationships between pixels in each  
 472 remotely sensed image. We have seen that characterizing pixels with contextual features computed  
 473 on segments, allows us to enhance the classification process. This methodology has been carried  
 474 out on a SITS composed of 15 HSR images. The different classification results obtained have  
 475 shown the relevance of this approach in the context of the analysis of agronomical areas.

476 This hybrid paradigm combines the possibilities offered by the (per-pixel) multi-temporal anal-  
477 ysis and the relevance of the (single-image) object-based frameworks for spatio-temporal analysis.  
478 The coming pair of SENTINEL-2 satellites will provide at the same time images with different spatial  
479 and radiometric resolutions (four bands at 10 m, six bands at 20 m and three bands at 60 m) at a  
480 high temporal frequency. In this context, the methodology proposed in this article provides a first  
481 trend to deal with such data.

482 We believe this work opens up a number of research directions. Firstly, the choice of the con-  
483 sidered spatial features in the classification process has to be deeply studied. For instance, textural  
484 and topological features could be used. Secondly, we also plan to validate the proposed method-  
485 ology by using other segmentation strategies. For instance, it has been proposed in (Kurtz et al.,  
486 2012) a new segmentation approach enabling to decompose the scene at different semantic levels.  
487 Such an approach could be extended to SITS to analyze the scene in a multi-temporal/multi-level  
488 fashion. We also plan to automate the choice of the parameters of segmentation. Indeed, different  
489 approaches (supervised or unsupervised) have been proposed to evaluate the quality of a segmen-  
490 tation (Clinton et al., 2010; Özdemir et al., 2010) and thus to select the “best” segmentation result  
491 relatively to a particular partitioning task. Finally, the higher the spatial and temporal resolution,  
492 the more relevant our approach will be. In this way, the next step of this study could consists of  
493 applying this paradigm to a series of Multi-Spectral/Panchromatic images couples. The spatial  
494 accuracy of Panchromatic images will help to preserve the fine details and structures.

## 495 **Acknowledgments**

496 The authors would like to thank the French Space Agency (CNES) and Thales Alenia Space  
497 for supporting this work under research contract n°1520011594 and the researchers from CESBIO  
498 (Danielle Ducrot, Claire Marais-Sicre, Olivier Hagolle and Mireille Huc) for providing the land-  
499 cover maps and the geometrically and radiometrically corrected FORMOSAT-2 images.

## 500 **References**

- 501 Akcay, H. G., & Aksoy, S., 2008. Automatic detection of geospatial objects using multiple hierarchical segmentations.  
502 *IEEE Transactions on Geoscience and Remote Sensing*, 46(7), 2097–2111.
- 503 Andres, L., Salas, W., & Skole, D., 1994. Fourier analysis of multi-temporal AVHRR data applied to a land cover  
504 classification. *International Journal of Remote Sensing*, 15(5), 1115–1121.
- 505 Bahirat, K., Bovolo, F., Bruzzone, L., & Chaudhuri, S., 2012. A novel domain adaptation bayesian classifier for  
506 updating land-cover maps with class differences in source and target domains. *IEEE Transactions on Geoscience  
507 and Remote Sensing*, (In press). 10.1109/TGRS.2011.2174154.



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

510 Bovolo, F., 2009. A multilevel parcel-based approach to change detection in very high resolution multitemporal  
511 images. *IEEE Geoscience and Remote Sensing Letters*, 6(1), 33–37.

512 Bruzzone, L., & Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution  
513 images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9), 2587–2600.

514 Bruzzone, L., & Prieto, D., 2000. Automatic analysis of the difference image for unsupervised change detection.  
515 *IEEE Transactions on Geoscience and Remote Sensing*, 38(3), 1171–1182.

516 Carleer, A., & Wolff, E., 2006. Urban land cover multilevel region-based classification of VHR data by selecting  
517 relevant features. *International Journal of Remote Sensing*, 27(6), 1035–1051.

518 Clinton, N., Holt, A., Scarborough, J., Yan, L., & Gong, P., 2010. Accuracy assessment measures for object-based  
519 image segmentation goodness. *Photogrammetric Engineering and Remote Sensing*, 76(3), 289–299.

520 Comaniciu, D., & Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions*  
521 *on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.

522 Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing*  
523 *of Environment*, 37(1), 35–46.

524 Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E., 2004. Digital change detection methods in  
525 ecosystem monitoring: A review. *International Journal of Remote Sensing*, 25(5), 1565–1596.

526 Fan, J., Wang, R., Zhang, L., Xing, D., & Gan, F., 1996. Image sequence segmentation based on 2D temporal entropic  
527 thresholding. *Pattern Recognition Letters*, 17(10), 1101–1107.

528 Foody, G., 2001. Monitoring the magnitude of land-cover change around the southern limits of the Sahara. *Pho-*  
529 *togrammetric Engineering and Remote Sensing*, 67(7), 841–848.

530 Gaetano, R., Scarpa, G., & Poggi, G., 2009. Hierarchical texture-based segmentation of multiresolution remote-  
531 sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 2129–2141.

532 Gueguen, L., Le Men, C., & Datcu, M., 2006. Analysis of satellite image time series based on information bottleneck.  
533 In *Proceedings of the 27<sup>th</sup> workshop on Bayesian Inference and Maximum Entropy Methods In Science and*

534 *Engineering*, pp. 367–374. volume 872.

535 Hagolle, O., Huc, M., Pascual, D. V., & Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to  
536 FORMOSAT-2, VEN $\mu$ S, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment*, 114(8), 1747–  
537 1755.

538 Hall, O., & Hay, G. J., 2003. A multiscale object-specific approach to digital change detection. *International Journal*  
539 *of Applied Earth Observation and Geoinformation*, 4(4), 311–327.

540 Herold, M., Liu, X., & Clarke, K., 2003. Spatial metrics and image texture for mapping urban land use. *Photogram-*  
541 *metric Engineering and Remote Sensing*, 69(9), 991–1001.

542 Hofmann, P., Lohmann, P., & Müller, S., 2008. Concepts of an object-based change detection process chain for  
543 GIS update: IntArchPhRS. In *21<sup>st</sup> International Society for Photogrammetry and Remote Sensing Congress*, pp.  
544 305–312. volume XXXVII.

545 Howarth, P., Piwowar, J., & Millward, A., 2006. Time-series analysis of medium-resolution, multisensor satellite data  
546 for identifying landscape change. *Photogrammetric Engineering and Remote Sensing*, 72(6), 653–663.

547 Huang, X., & Zhang, L., 2008. An adaptive mean-shift analysis approach for object extraction and classification from  
548 urban hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12), 4173–4185.

549 Idbraim, S., Ducrot, D., Mammass, D., & Aboutajdine, D., 2009. An unsupervised classification using a novel ICM  
550 method with constraints for land cover mapping from remote sensing imagery. *International Review on Computers*  
551 *and Software*, 4(2), 165–176.

552 Jensen, J. R., 1981. Urban change detection mapping using Landsat digital data. *Cartography and Geographic*  
553 *Information Science*, 8(21), 127–147.

554 Johnson, R., & Kasischke, E., 1998. Change vector analysis: A technique for the multispectral monitoring of land  
555 cover and condition. *International Journal of Remote Sensing*, 19(16), 411–426.

556 Jönsson, P., & Eklundh, L., 2004. TIMESAT – A program for analyzing time-series of satellite sensor data. *Computers*  
557 *& Geosciences*, 30(8), 833–845.

558 Kennedy, R., Yang, Z., & Cohen, W., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat  
559 time series: 1. LandTrendr – Temporal segmentation algorithms. *Remote Sensing of Environment*, 114(12), 2897–

560 2910.

561 Kennedy, R. E., Cohen, W. B., & Schroeder, T. A., 2007. Trajectory-based change detection for automated character-  
562 ization of forest disturbance dynamics. *Remote Sensing of Environment*, 110(3), 370–386.

563 Kurtz, C., Passat, N., Gañçarski, P., & Puissant, A., 2010. Multiresolution region-based clustering for urban analysis.  
564 *International Journal of Remote Sensing*, 31(22), 5941–5973.

565 Kurtz, C., Passat, N., Gañçarski, P., & Puissant, A., 2012. Extraction of complex patterns from multiresolution remote  
566 sensing images: A hierarchical top-down methodology. *Pattern Recognition*, 45(2), 685–706.

567 Kurtz, C., Passat, N., Puissant, A., & Gañçarski, P., 2011a. Hierarchical segmentation of multiresolution remote  
568 sensing images. In P. Soille, M. Pesaresi, & G. K. Ouzounis (Eds.), *Proceedings of the International Symposium*  
569 *on Mathematical Morphology*, pp. 343–354. Springer volume 6671 of *Lecture Notes in Computer Science*.

570 Kurtz, C., Puissant, A., Passat, N., & Gañçarski, P., 2011b. An interactive approach for extraction of urban patterns  
571 from multisource images. In *Proceedings of the IEEE Joint Urban Remote Sensing Event*, pp. 321–324.

572 Lu, D., Mausel, P., Brondizio, E., & Moran, E., 2004. Change detection techniques. *International Journal of Remote*  
573 *Sensing*, 25(37), 2365–2401.

574 Lui, D., & Cai, S., 2011. A spatial-temporal modeling approach to reconstructing land-cover change trajectories from  
575 multi-temporal satellite imagery. *Annals of the Association of American Geographers*, (In Press).

576 MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In *Berkeley Sympo-*  
577 *sium on Mathematical Statistics and Probability*, pp. 281–297.

578 Melgani, F., & Serpico, S. B., 2002. A statistical approach to the fusion of spectral and spatio-temporal contextual  
579 information for the classification of remote-sensing images. *Pattern Recognition Letters*, 23(9), 1053–1061.

580 Moscheni, F., Bhattacharjee, S., & Kunt, M., 1998. Spatio-temporal segmentation based on region merging. *IEEE*  
581 *Transactions on Pattern Analysis and Machine Intelligence*, 20(9), 897–915.

582 Niemeyer, I., Marpu, P., & Nussbaum, S., 2008. Change detection using object features. In *Object-Based Image*  
583 *Analysis*, Lecture Notes in Geoinformation and Cartography chapter 2.5. pp. 185–201. Springer Berlin Heidelberg.

584 Ouma, Y. O., Josaphat, S., & Tateishi, R., 2008. Multiscale remote sensing data segmentation and post-segmentation  
585 change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover

586 change analysis. *Computers & Geosciences*, 34(7), 715–737.

587 Özdemir, B., Aksoy, S., Eckert, S., Pesaresi, M., & Ehrlich, D., 2010. Performance measures for object detection  
588 evaluation. *Pattern Recognition Letters*, 31(10), 1128–1137.

589 Pesaresi, M., & Benediktsson, J. A., 2001. A new approach for the morphological segmentation of high-resolution  
590 satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2), 309–320.

591 Petitjean, F., Inglada, J., & Gançarski, P., 2011a. Clustering of satellite image time series under time warping. In  
592 *Proceedings of the IEEE International Workshop on the Analysis of Multi-temporal Remote Sensing Images*, pp.  
593 69–72.

594 Petitjean, F., Inglada, J., & Gançarski, P., 2012. Satellite image time series analysis under time warping. *IEEE*  
595 *Transactions on Geoscience and Remote Sensing*, 50(8).

596 Petitjean, F., Ketterlin, A., & Gançarski, P., 2011b. A global averaging method for dynamic time warping, with  
597 applications to clustering. *Pattern Recognition*, 44(3), 678–693.

598 Petitjean, F., Masseglia, F., Gançarski, P., & Forestier, G., 2011c. Discovering significant evolution patterns from  
599 satellite image time series. *International Journal of Neural Systems*, 21(6), 475–489.

600 Schopfer, E., Lang, S., & Albrecht, F., 2008. Object-fate analysis: Spatial relationships for the assessment of object  
601 transition and correspondence. In *Object-Based Image Analysis*, Lecture Notes in Geoinformation and Cartography  
602 chapter 8.4. pp. 785–801. Springer Berlin Heidelberg.

603 Tiede, D., Lang, S., Füreder, P., Hölbling, D., Hoffmann, C., & Zeil, P., 2011. Automated damage indication for rapid  
604 geospatial reporting. *Photogrammetric Engineering and Remote Sensing*, 77(9), 933–942.

605 Tsai, D. M., & Chiu, W. Y., 2008. Motion detection using Fourier image reconstruction. *Pattern Recognition Letters*,  
606 29(16), 2145–2155.

607 Tseng, V. S., Chen, C. H., Huang, P. C., & Hong, T. P., 2009. Cluster-based genetic segmentation of time series with  
608 DWT. *Pattern Recognition Letters*, 30(13), 1190–1197.

609 Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D., 2010. Detecting trend and seasonal changes in satellite  
610 image time series. *Remote Sensing of Environment*, 114(1), 106–115.

611 Wu, Q. Z., Cheng, H. Y., & Jeng, B. S., 2005. Motion detection via change-point detection for cumulative histograms

612 of ratio images. *Pattern Recognition Letters*, 26(5), 555–563.