

Satellite Image Time Series Analysis Under Time Warping

François Petitjean, Jordi Inglada, and Pierre Gançarski

Abstract—Satellite Image Time Series are becoming increasingly available and will continue to do so in the coming years thanks to the launch of space missions which aim at providing a coverage of the Earth every few days with high spatial resolution. In the case of optical imagery, it will be possible to produce land use and cover change maps with detailed nomenclatures. However, due to meteorological phenomena, such as clouds, these time series will become irregular in terms of temporal sampling, and one will need to compare time series with different lengths. In this paper, we present an approach to image time series analysis which is able to deal with irregularly sampled series and which also allows the comparison of pairs of time series where each element of the pair has a different number of samples. We present the dynamic time warping from a theoretical point of view and illustrate its capabilities with two applications to real-time series.

Index Terms—Classification, clustering, dynamic time warping, remote sensing, satellite image time series (SITS).

I. INTRODUCTION

SATELLITE Image Time Series (SITS, for short) are a precious resource for Earth monitoring. Current time series have either high temporal resolution (Spot Vegetation, Modis) or high spatial resolution (Landsat). In the coming years, both high temporal and high spatial resolution SITS are going to be widely available thanks to the European Space Agency's (ESA) Sentinel program. Nowadays, satellites as the Taiwanese Formosat-2 are already providing similar data, but with a limited coverage of the Earth's surface and with only four spectral bands.

In order to efficiently use the huge amounts of data that will be produced by, for instance, Sentinel-2 (global cover every five days with 10-m to 60-m resolution and 13 spectral bands), new methods for SITS analysis have to be developed. Indeed, the following issues will have to be addressed:

- a) *Reference data*: In the case where a global land cover has to be updated every few days or weeks, one cannot assume that reference data—ground truth, training

samples—are going to be available. In order to cope with this lack of data, methods able to operate in an unsupervised way or using reference data from, say, previous years, are needed.

- b) *Irregular sampling*: When only several images per year are available over a given site, the constraints on the stability of temporal sampling frequency can be weakened. For instance, if one image per season is available, a jitter of a few weeks on the sampling period will have a very low impact on the results. However, when several images per month are available—and therefore required for new applications—classical multitemporal methods will suffer from low regularity of the sampling at the scale of days. Therefore, only methods that deal with irregular temporal sampling will be able to fully exploit the available acquisitions.
- c) *Pseudoperiodic phenomena*: Many phenomena of interest—vegetation cycles, for instance—have a periodic behavior which can be slightly modulated by weather artifacts. These modulations result in distortions of canonical temporal profiles of, say Normalized Difference Vegetation Index (NDVI) or other physical variables. This kind of phenomena is the dual of the irregular sampling issue presented above: it is not the sensor, but the observed objects that have an irregular temporal behavior. Therefore, methods that have some kind of invariance to temporal stretching or dilation are of major interest.

A quick review of the available literature on SITS analysis (Section II-A) shows a lack of existing methods responding to the three issues stated above.

In this paper, we present the dynamic time warping (DTW) similarity measure [1], [2] and use it on real SITS in order to show how the aforementioned problems can be dealt with. Although we focus our examples on optical SITS, the methods presented here can also be applied to SAR images. Since DTW is sensitive to spiky noise, SAR images should be appropriately filtered to reduce speckle noise before using them in the framework presented here.

We previously made DTW applicable to data mining in [3]. In the current contribution, we show how DTW can be useful for specific problems raised by SITS and present some specific adaptations of the algorithm needed to deal with Earth observation data.

The paper is organized as follows: In Section II, we start by presenting an overview of the state of the art, first focusing on the main SITS analysis methods and then focusing on the comparison of temporal radiometric behaviors and associated

Manuscript received January 8, 2011; revised May 31, 2011 and October 7, 2011; accepted November 29, 2011. Date of publication January 31, 2012; date of current version July 18, 2012. This work was supported by the French Space Agency (CNES) and Thales Alenia Space under Research Contract 1520011594.

F. Petitjean and P. Gançarski are with the Image Sciences, Computer Sciences and Remote Sensing Laboratory (LSIIT), 67412 Illkirch Cedex, France (e-mail: fpetitjean@unistra.fr; gancarski@unistra.fr).

J. Inglada is with CNES/CESBIO, UMR 5126, 31401 Toulouse Cedex 9, France (e-mail: jordi.inglada@cesbio.cnes.fr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2011.2179050

constraints. Section III presents the DTW similarity measure in detail. Section IV details the choices and adaptations that were made for the application of DTW to remote sensing. Section V describes the data sets used in our experiments. Several interesting results of the use of DTW for SITS analysis are presented in Section VI. Finally, we conclude and present some future works in Section VII.

II. STATE OF THE ART

This state of the art is divided into two parts. Section II-A presents the main kinds of methods for SITS analysis and then justify the choice made in this paper: focusing on measures for the comparison of temporal profiles. Section II-B presents several measures for time series (temporal profiles) comparison and makes a critical analysis of the presented measures in order to introduce the DTW similarity measure, on which we focus in this paper, in the next section.

A. Satellite Image Time Series Analysis

Notations: Throughout this section, let t_1, \dots, t_n , respectively, denote *image 1*, \dots , *image n*. Let “ $<$ ” denote the strict total order (induced by the time) on the images. Let f be a binary function (of arity two) taking two images as parameters and returning one image. The f function is used to illustrate the behavior of bitemporal dedicated methods.

Typology: Beyond the way the methods work by themselves, they can be classified depending on the way they use the time dimension of the data. For example, given a SITS, the use of the time structure of the data is very different from image differencing to frequency analysis. The first one only uses the ordering induced by the time on couples of images, while the second one uses the whole ordering of the series. The aim of this typology is to describe the different levels of the time structuring. We voluntarily do not separate bidate and multirate methods (associated to the analysis of abrupt/long-term changes), since this article aims at defining a descriptor of time series, which could be used for both abrupt changes and long-term changes analysis.

The type of data we are considering in this work, i.e., SITS, is highly structured, both spatially and temporally. The spatial structure of the data is generally used as the basis for the comparison between different values: the coordinates (x, y) of pixels in the images identify each sensed area through time.

Where the use of the structuring of the data usually differs is how they are taking the temporal information into account, i.e., in the use of the sensing date or of the sequencing. Depending on the method used, the weight of the temporal dimension can be very different and can occur at different levels.

We bring out three main uses of the time dimension:

- 1) *Time as identifier:* time is only used to identify the information, i.e., there is no ordering between the images of the series;
- 2) *Pairwise time ordering:* time is used to structure the images, pairwise;
- 3) *Time ordering the sequence:* time is used to structure the image series.

This state of the art will detail each one of these three categories. Each category will be first presented in a general way, then methods belonging to it will be presented, and finally, the category will be analyzed in the light of presented methods. For a detailed survey of these methods, the reader should refer to [4] and [5].

1) *Time as an Identifier:* The time dimension is here used as an attribute identifier. Consequently, each sensed image is used as an additional attribute. Intuitively, this use of the time dimension usually consists of concatenating the different images into one single image with many attributes. This category includes:

- a) *Data linear transformation:* This kind of methods is mainly based on the theory of statistics. It includes methods that consist of statistically transforming the data with, for example, principal component analysis or with maximum autocorrelation factor (see [6] for an application of these two methods). These methods are simple to use and generally give interesting results. However, even if a few attempts have been made to compose these methods in order to use the time dimension in some way, (see [7] for an example of a hierarchical combination of these methods), the ordering induced by the temporal dimension cannot be taken into account, particularly when dealing with a SITS.
- b) *Classification:* This kind of methods consists of applying a classification algorithm on the concatenated image.¹ Several attempts have been made using the K-MEANS algorithm or with the Expectation Maximization algorithm [8]. However, as the data are somehow untemporalized before being processed, no temporally structured result can be obtained.
- c) *Change detection methods based on classification comparison:* These methods consist of independently classifying every image and combining and/or fusing them in order to produce a single classification. These methods have the advantage of being applicable to two or more images and do not require comparable radiometric levels between the images. For example, no relative radiometric normalization is required. Examples of applications to two images [9], to four images [10], or even to 16 images [11] are given in the literature. However, the time dimension is not really taken into account since modifying the order of the images has no effect on the result.

As a conclusion on these methods, although they exploit the whole data, they are not able to identify specific temporal behaviors. This kind of methods is, in essence, unable to extract and/or characterize temporal behaviors. Admittedly, some change areas can be isolated from others by these methods, but only because these areas have a different frequency of events; shuffling the attributes will have no effect on the results. In this way, although these methods are simple to use and are obviously tolerant to irregular sampling, they are not able to make a consistent temporal analysis of SITS.

2) *Pairwise Time Ordering:* At a higher level of temporal structuring, some methods (originally bitemporal ones) use

¹The Euclidean distance is used here to compare the time series. We will return to this point in Section II-B-4.

the temporal information as a partial ordering, i.e., as several “previous/next” orders:

$$t_1 < t_2, \dots, t_{n-1} < t_n. \quad (1)$$

This use of the time dimension usually consists of using the temporal information between couples of images, i.e., pairwise. In this way, this kind of methods was initially devoted to bitemporal analysis but was extended to multitemporal analysis by composition of these methods. These methods usually require comparable values between the images. They include the following.

- a) *Difference/ratio/combination*: This type of methods consists of combining the values of the image at time t and those at time $(t - 1)$ in order to reveal the intrinsic temporal structure of the data. The combination operator can be as simple as a subtraction [12]–[14], a division [15], [16], or more sophisticated [17], [18]. The resulting combined image is then usually thresholded or classified, in order to map (and sometimes characterize) the change areas. These methods are however too simple to understand the specificity of the time dimension and are only applicable to two images at the same time. Thus, in order to handle SITS, these methods have to be applied several times (e.g., 12 images combined in one in [19]), which leads to hardly understandable results.
- b) *Change Vector Analysis*: This method is designed for and limited to bitemporal analysis. It consists of building a vector from two multiband values (one before and one after the change) in a multidimensional space. The study of both the norm and the angle of this vector gives an information on the type of the change and of its intensity. Many studies use this method to map and characterize change areas (e.g., [20]–[22]).
- c) *Linear Regression*: This method considers that the pixel value at time t is linearly correlated to the pixel value at time $(t - 1)$. The parameters of the regression (e.g., the residual) are studied to map and characterize the change. Many works are using this method [23]–[25], but it was shown that results are generally not better than image differencing [26].

Using this pairwise ordering makes it possible to extract a temporal information from pairs of images. However, these methods are *per se* designed to handle pairs of images. In this way, in order to handle SITS, they have to be applied several times (usually in a tournament scheme) such as

$$\begin{aligned} & f(f(t_1, t_2), f(t_3, t_4)) \\ \text{or } & f(f(f(t_1, t_2), t_3), t_4). \end{aligned} \quad (2)$$

This principle has however a major drawback: when applying the method on (t_1, t_2) and then on (t_3, t_4) , in order to handle a SITS of length four, the ordering between t_2 and t_3 is not directly used. In addition, the composition of these methods in order to analyze SITS proves to be risky, since the results depend on the properties of the methods (e.g., associativity, reflexivity), and are thus sensitive to ordering schemes. Globally, the combination of these bitemporal dedicated methods,

appears more as an *ad hoc* technique than as a method able to handle whole SITS.

3) *Time Ordering the Sequence*: Then, at a higher level of time structuring, some methods are using the whole ordering induced by the temporal information and are studying the evolution of an area (x, y) throughout the image series. The structuring is here, even more important since the order is generalized² to the series as

$$(t_1 < t_2) \wedge (t_2 < t_3) \wedge \dots \wedge (t_{n-1} < t_n) \quad (3)$$

$$\Leftrightarrow t_1 < t_2 < \dots < t_{n-1} < t_n. \quad (4)$$

This additional constraint enables the mining of structured information and the analysis of evolution behaviors from the image series. It includes the following.

- a) *Frequent pattern mining*: This method extracts frequent subsequences of radiometric evolutions. It has the advantage of being robust to noise and extracting meaningful patterns, but it generates thousands of patterns which are difficult to filter in order to understand changes that occurred over an area. Moreover, a discretization step is required in order to reduce the size of the search space. This step can be critical because the mining step highly depends on it. Several studies have been published [27]–[29].
- b) *Frequency analysis*: This type of methods uses a Fourier or a wavelet decomposition of the radiometric time series [30]–[32]. These methods can handle time series and are robust. However, they require a regular sampling of the time series and relatively time series.

All of these methods are using the ordering of the sequence. Then, frequency-based methods add a constraint on the temporal sampling: data must be regularly sensed, i.e., the time between two acquisitions must be constant over the image time series. Remote sensing makes this constraint difficult to keep, since the acquisitions depend on several factors (operational, meteorological, etc.). Thus, releasing the constraint on the regular sampling, i.e., having the same time between two successive images throughout the series, the use of the total order on the sensed values appears to be consistent for the analysis of evolution behaviors of each sensed area over the SITS.

Conclusion: The above discussion highlights that there is a need for global analysis methods, in which each radiometric evolution would be analyzed and compared to others. In this paper, we propose to compare/classify these evolution behaviors/trends. Therefore, we consider each pixel value of the SITS as an element of a sequence describing the evolution of a sensed area. The next section describes different similarity measures that makes it possible to exploit the temporal structuring of SITS, in order to compare radiometric evolutions.

B. Comparing Temporal Profiles

All methods used for time series analysis actually include a comparison of the data by using distances. Whatever method is

²This order is a *total order* on the set of the images.

used, the core of the process generally consists of comparing data in order to estimate a (dis)similarity. The distance tool provides an estimation of this similarity. It is a critical tool, on which the results of analysis methods heavily rely. When the data are temporal, the choice of the distance is crucial since it completely defines the way of tackling the temporality of the data. In this way, two sequences should be close if they represent similar evolution behaviors. Thus, a distance could be appropriate in order to compare speech signals, but not for remote sensing, and conversely.

This subsection presents different similarity measures for sequences (i.e., temporal profiles of radiometric values). As we focus on the comparison of temporal profiles of evolution, this state of the art should be read in the light of the following several objectives for the analysis of SITS:

- using all of the available sensed values;
- being applicable to numerical values;
- being able to apprehend temporal behaviors (e.g., shifts, distortions);
- being applicable to irregularly sensed image series (irregular temporal sampling);
- being parameter free;
- yielding results which can be easily interpreted.

Notations: Let $A = \langle a_1, \dots, a_T \rangle$ and $B = \langle b_1, \dots, b_T \rangle$ be two sequences, and let δ be a distance between two elements (or coordinates) of sequences (usually the L^1 distance of the L^2 distance).

1) *Euclidean Distance:* This distance is commonly accepted as the simplest distance between sequences. The distance between A and B is defined by

$$D(A, B) = \sqrt{\delta(a_1, b_1)^2 + \dots + \delta(a_T, b_T)^2}. \quad (5)$$

This distance does not correspond to the common understanding of what a sequence really is and cannot capture flexible similarities. For example, $X = \langle a, b, a, a \rangle$ and $Y = \langle a, a, b, a \rangle$ are different according to this distance even though they represent similar evolution behaviors.

2) *Compression-Based Dissimilarity Measure:* This is a measure introduced in [33], based on the theory of the Kolmogorov complexity [34]. The authors suggest to use this principle to define a new similarity measure based on compression as

$$D(A, B) = \frac{|compress(A.B)|}{|compress(A)| + |compress(B)|} \quad (6)$$

with $A.B$ be the concatenation of A and B , and with $|compress(\dots)|$ be the size of the compressed file. This measure is, however, strongly dependent on the compression algorithm chosen. Furthermore, results can be difficult to interpret, since the compression algorithm works as a black box. Finally, this technique is designed for relatively long sequences, without which the compression algorithm will be underused.

3) *Levenshtein Distance:* The Levenshtein distance (or edit distance) [35] formalizes the notion of distance between two symbol strings, by focusing on transforming (or editing) one string into the other by a series of edit operations on individual symbols. The allowed edit operations are insertion, deletion,

and replacement of a symbol. It has the advantage of understanding that $\langle a, b, a \rangle$ is close to $\langle a, a, b, a \rangle$ with a difference of an insertion or a deletion of one “a.” This distance is, in this way, able to handle temporal data. However, it presents two main drawbacks: first, it is not applicable to numerical values, and second, it requires the definition of the costs for the insertion/deletion of symbols, and for the transformation of each symbol to another one.

4) *Longest Common Subsequence:* (LCSS) was first introduced to find the longest common subsequence between two strings. This method was extended in [36] to a similarity measure between two numerical sequences. This measure has two main shortcomings. First, it requires the definition of the meaning of “common” between values in order to count the length of the longest subsequence. Second, this method aims at finding a subsequence, and it thus forgets sensed values. Even if this property could be interesting in order to skip cloudy values, several meaningful sensed values could be skipped, leading to a meaningless result.

Critical Analysis:

- a) *From the application point of view:* Even if there is a need for a comparison method of temporal profiles of evolution, most of the methods presented above are not well suited to remote sensing. The Euclidean distance does not take into account the temporality of the data and is therefore not relevant to the analysis of time series whatever the application is. Measures based on compression are computationally expensive, which make them unusable in remote sensing. Moreover, their results are difficult to interpret, and are hardly applicable to short time series. On the other hand, methods based on the edit distance are more interesting since they exploit the temporality of the data and capture temporal behaviors. However, the Levenshtein distance is intended for symbolic sequences and requires a similarity matrix defined by an expert, which prevents its application to remote sensing. LCSS permits to find common subsequences. Two problems are raised for its application to remote sensing. First, the definition of “common” should be established. Second, and this problem is more objectionable, this distance is calculated on a subset of the data (subsequences), reducing its relevance since the whole data set is not used.
- b) *From the theoretical point of view:* This article is focused on the analysis of a whole SITS with distance-based methods. Most distance-based analysis methods require to be able to compute an average consistently to the distance used (e.g., K-MEANS). Among all presented similarity measures, only the Euclidean distance has an associated averaging method. This point further limits the applicability of other measures to remote sensing and more generally to the study of temporal data.

III. DYNAMIC TIME WARPING

This section presents the *DTW* similarity measure, which makes it possible to analyze the temporal nature of the whole data set (without skipping any values). For remote sensing, we will see that *DTW* can exploit the temporal distortions and

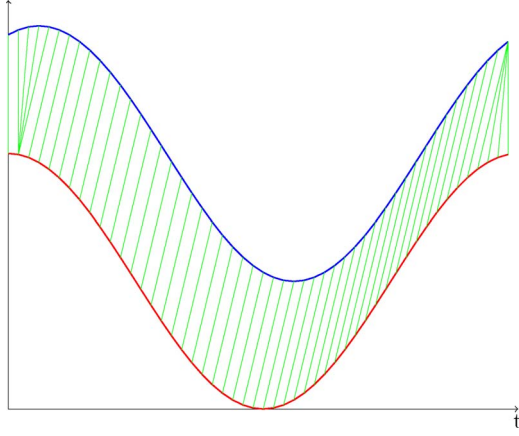


Fig. 1. Two 1-D sequences aligned with dynamic time warping. Coordinates of the top and bottom sequences have been, respectively, computed by $\cos(t)$ and $\cos(t + \alpha)$. For visualization purposes, the top sequence is drawn vertically shifted.

compare shifted or distorted evolution profiles and whose time sampling is irregular, thanks to the optimal alignment of radiometric profiles. Furthermore, DTW is parameter free, which makes it match the several objectives given in the previous section.

DTW is based on the Levenshtein distance and was introduced in [1] and [2], with applications to speech recognition. DTW is able to find the optimal global³ alignment between two sequences and is probably the most commonly used measure to quantify the dissimilarity between sequences [37]–[41]. It also provides an overall real number that quantifies the similarity between the two sequences. An example of DTW-alignment of two sequences is presented in Fig. 1: it shows the alignment of points taken from two sinusoids, one being slightly shifted in time. The numerical result computed by DTW is the sum of the heights⁴ of the associations. Alignments at both extremities on Fig. 1 show that DTW is able to correctly re-align one sequence with the other, a process which, in this case, highlights similarities that the Euclidean distance is unable to capture.

DTW captures flexible similarities by aligning the coordinates inside both sequences (each element of the first sequence is linked to (at least) one element of the second sequence). The cost of the optimal alignment can be recursively computed by

$$D(A_i, B_j) = \delta(a_i, b_j) + \min \begin{cases} D(A_{i-1}, B_{j-1}), \\ D(A_i, B_{j-1}), \\ D(A_{i-1}, B_j) \end{cases} \quad (7)$$

where A_i is the subsequence $\langle a_1, \dots, a_i \rangle$. The overall similarity is given by $D(A_{|A|}, B_{|B|})$.

Unfortunately, a direct implementation of this recursive definition leads to an algorithm that has exponential cost in time. Fortunately, the fact that the overall problem exhibits overlapping subproblems (see Fig. 2) allows for the memoization of partial results in a matrix which makes the minimal weight coupling computation a process that costs $|A| \times |B|$ basic op-

erations. This measure has thus a time and a space complexity of $\Theta(|A| \times |B|)$.⁵

The algorithm presented in Fig. 1 details the computation of DTW. Given two sequences A and B of lengths S and T , respectively ($S = |A|, T = |B|$), this algorithm computes the score of the alignment and comparison of A and B . It uses the *dynamic programming* principle to allow the memoization of overlapping subproblems in a $S \times T$ matrix. In this way, the algorithm consists at first of initializing the first column and the first row of the matrix. Then, the matrix is computed from left to right and from top to bottom. Each element of the matrix is computed by using the smallest score from the left element, the upper element, and the diagonal one. Once the matrix is entirely computed, the last element at bottom right gives the score of the best alignment of the two sequences. An example of a result matrix computed with DTW as well as its corresponding alignment is given in Fig. 3 (Table I).

DTW allows to find the best global alignment between two numerical sequences. Providing the cost of this alignment, DTW is generally used as a dissimilarity measure between two sequences. DTW appears thus to be a time-designed similarity measure, able to gather locally time-distorted sequences with time shifts and, more generally, local time distortions.

Finally, we previously introduced an associated averaging method (DBA) in [3], which makes it applicable to most distance-based analysis methods. Furthermore, this averaging method provides a condensed representation of the results that can be used for their interpretation by the expert.

IV. DYNAMIC TIME WARPING FOR REMOTE SENSING

The previous subsection detailed the original definition of DTW. However, its application to remote sensing requires to make several choices and adaptations of the original method. In this subsection, the following three points will be detailed:

- A. the extension of DTW to multidimensional time series, i.e., multispectral radiometric profiles of evolution;
- B. the modification of DTW in order to avoid inconsistent temporal distortions, e.g., forbidding the association of winter sensed values with summer ones;
- C. the handling cloud-contaminated images in the construction of the sequences.

A. DTW Versus Multidimensional Sequences

The original definition of DTW is devoted to the study of 1-D-signals (e.g., speech signals), i.e., where each element of the sequence is described in a 1-D space. However, in the analysis of SITS, each element of the sequences corresponds to a multispectral pixel, i.e., a multidimensional vector. In this way, DTW has to be modified in order to handle this kind of data. There are actually two main ways of comparing two multidimensional sequences with DTW:

- computing DTW B times, one time per dimension (i.e., per band);
- using a B -dimensional δ in the computation of DTW.

³Any of the values of the two sequences can be skipped.

⁴In fact, the distance $\delta(a_i, b_j)$ computed in (7) is the distance between two coordinates without considering the time distance between them.

⁵Note that the space complexity can be reduced to a $\Theta(\min(|A|, |B|))$, since only the previous and the current rows are necessary for the computation.

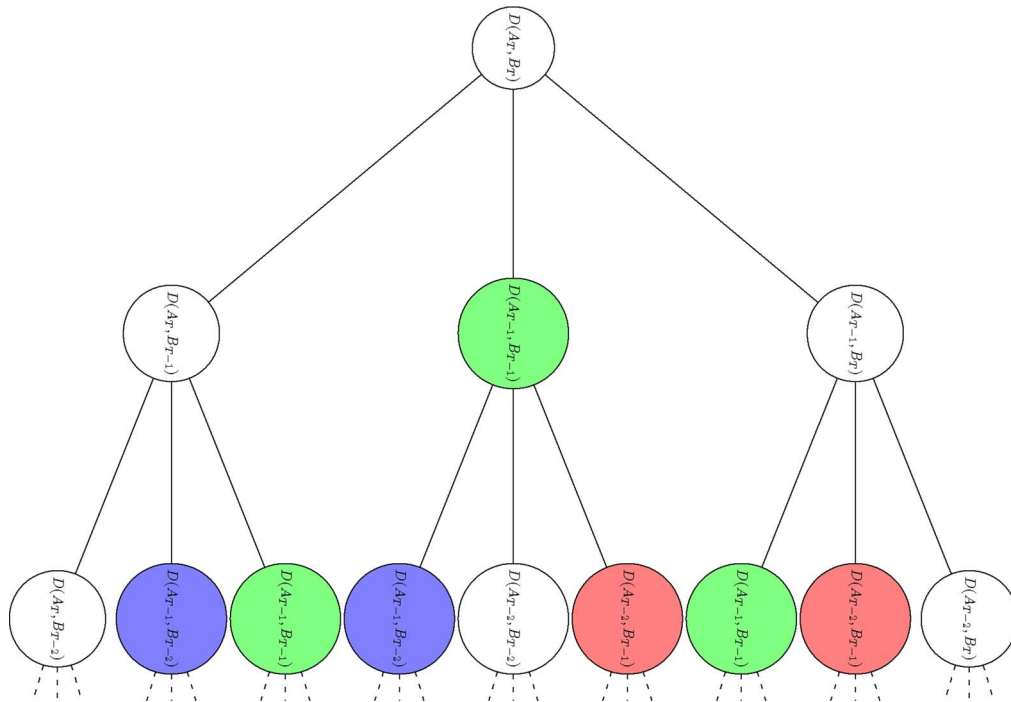


Fig. 2. Call tree of the function detailed in Equation (7). Three groups are highlighted and represent the same calls. This is a memoization example.

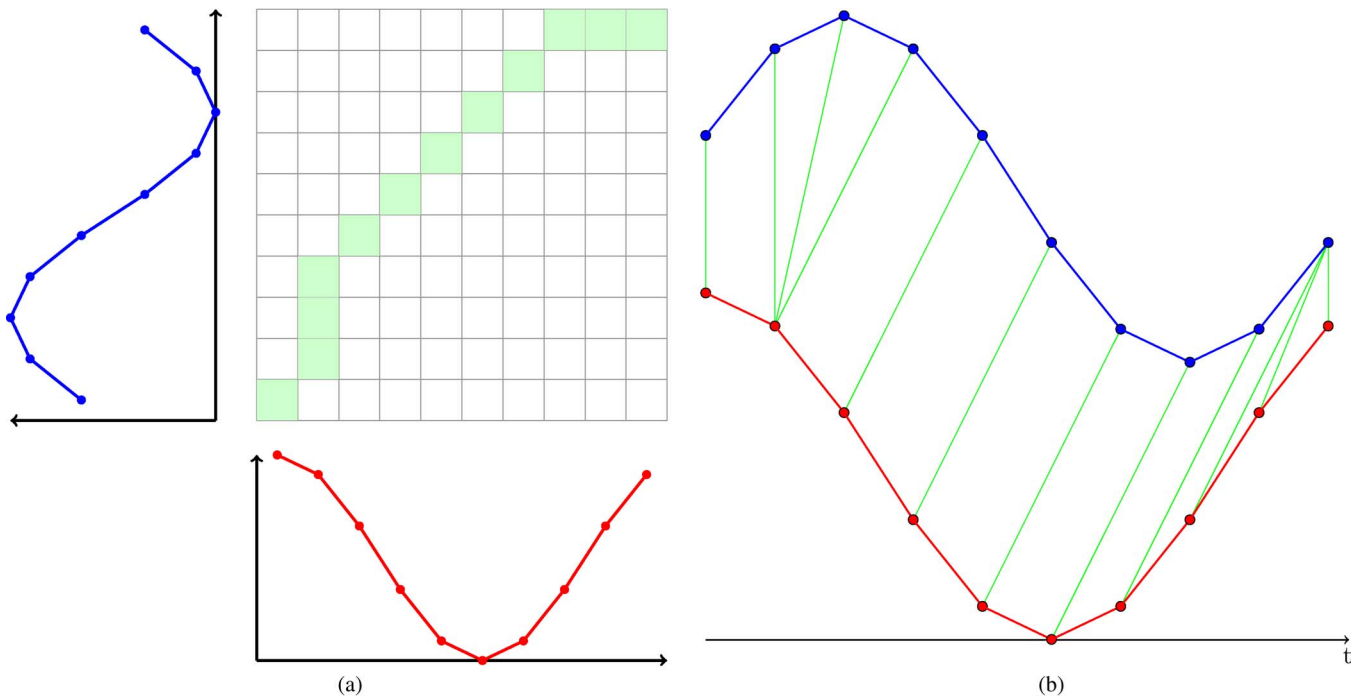


Fig. 3. Example of the matrix computed by DTW. (a) The matrix and the warping path computed by DTW. (b) The resulting alignment of the two sequences.

The first solution consists of considering that the radiometric series is composed of B 1-D series, one per band. Then, in order to compare the two radiometric series, DTW provides the cost of the alignment of each band separately, and the B scores are then merged (by averaging, for instance). In this way, DTW will align every band separately, leading to B (possibly different) alignments of the two multidimensional sequences. The information about the synchronicity of the B values composing a pixel is thus lost.

The second solution consists of considering that the radiometric series is composed of a single series of N multidimensional elements. In this way, a single alignment of the two sequences is made. To this end, the cost function δ has to be able to compare vectors of values. Practically, the Euclidean distance can be used to compare the vectors. This solution keeps the information on the synchronicity of the B values composing a pixel (sensed vector of values): the pixel remains atomic.

TABLE I
 DTW ALGORITHM

Require: $A = \langle a_1, \dots, a_S \rangle$
Require: $B = \langle b_1, \dots, b_T \rangle$
Let δ be a distance between coordinates of sequences
Let $m[S, T]$ be a cost matrix
 $m[1, 1] \leftarrow \delta(a_1, b_1)$
for $i \leftarrow 2$ **to** S **do**
 $m[i, 1] \leftarrow m[i - 1, 1] + \delta(a_i, b_1)$
end for
for $j \leftarrow 2$ **to** T **do**
 $m[1, j] \leftarrow m[1, j - 1] + \delta(a_1, b_j)$
end for
for $i \leftarrow 2$ **to** S **do**
for $j \leftarrow 2$ **to** T **do**
 $m[i, j] \leftarrow \delta(a_i, b_j) + \min \begin{cases} m[i - 1, j] \\ m[i, j - 1] \\ m[i - 1, j - 1] \end{cases}$
end for
end for
return $m[S, T]$

The second solution appears to better fit the analysis of radiometric series. To distinguish the several land cover states, the set of the B values sensed at the same time is required. The first solution would only be interesting if the behavior of the sequences in every band was independent; if the synchronicity of the B values was not required. In this way, the first solution would be adapted to sensed areas where the B are representing uncorrelated phenomenons, which is rarely the case in remote sensing.

B. Avoiding Inconsistent Temporal Distortions

DTW makes it possible to align elements on the “length” of the distortions/shifts. This ability is very useful to be sufficiently robust and flexible in order to handle SITS. However, the expert might want to introduce his/her knowledge of the observed temporal phenology. For example, the expert might want to guarantee that summer crop are not aligned with winter ones, or that values sensed with a time delay of more than two months should not be aligned together.

Constraints on the alignments have actually already been studied. However, contrary to our case, these constraints aimed at speeding up the computation of DTW. They consist of limiting the search of the warping path in a subset of the matrix called warping window. The objective is to limit the search space in order to decrease the computational complexity. The idea is to prevent the warping path from straying away from the diagonal of the matrix. For instance, the warping path can be limited to a certain band around the diagonal of the matrix named Sakoe-Chiba band [2]. Another example corresponds to a warping window shaped as a parallelogram and named Itakura parallelogram [42], enabling more distortions in the middle of the sequences than on the extremities. These global constraints on the search of the warping path are actually limiting the alignments, by preventing an alignment of two elements that are too distant. Their aim is to decrease the computational complexity, and the constrain on the alignments is the way of doing it. We take here the opposite direction: we introduce some constraints in order to avoid inconsistent temporal distortions, and the decrease of the computational complexity is a bonus.

These global constraints are however not suitable to SITS analysis, since the data are irregularly sampled. For example, the Sakoe-Chiba band works by removing all matrix elements that are too distant from the diagonal: the matrix element (i, j) cannot be part of the warping path if $|i - j| > w$ (with w be the width of the band). As a result, the i th element of the first sequence cannot be linked to the j th element of the second sequence if they are more than w elements apart. As a consequence, this constraint would only make sense if the elements of the sequences are regularly sampled; the number of elements would then have a temporal meaning, since w elements would correspond to a time delay of w times the sampling frequency. In the case of SITS, however, limiting the warping window to all elements (i, j) distant of less than w elements would have no temporal meaning, since two consecutive elements (i.e., corresponding to two consecutive sensed images) could be sensed with a time delay of, for instance, one week, one month, or one year. Thus, given a maximum time delay Δt , we propose to limit the scope of the matrix to all elements (i, j) fulfilling the following:

$$\text{DateDiff}(\text{Date}(i), \text{Date}(j)) < \Delta t \quad (8)$$

where Date is a function returning, respectively, the date of the i th image of the first sequence and the date of the j th image of the second sequence, and DATEDIFF is a function returning the elapsed time between two dates.

Moreover, note that this constraint on the warping path should be chosen in order to ensure that the warping path exists, i.e., that the warping window is composed of a single part. Finally, this maximum time delay Δt could be variable, in order to match more complex knowledge about the observed temporal phenology.

In practice, during the computation of every (i, j) element of the matrix, the condition presented in (8) has to be evaluated beforehand: if the condition is true, the classic procedure remains unchanged, otherwise, the value of the matrix can be set to $+\infty$. This procedure corresponds to a masking of the matrix in order to compute the warping path in a warping window. The resulting mask is then consistent with the sensing dates of the satellite images and adapted to this kind of irregularly sampled data.

C. Handling Cloud-Contaminated Images

In the classification of single satellite images, cloud-covered pixels have usually a minor impact on the results: cloud-covered areas are usually removed from the results. However, when handling SITS, the atomic data is the series of radiometric values taken over an area. In this way, the probability of a cloud contamination greatly increases. Let us imagine a simple case where the probability for a pixel to be cloud covered is constant. Let the probability $\mathbb{P}_p(\bar{C})$ for a sensed area p to be cloud free, we have then the probability $\mathbb{P}_S(\bar{C})$ for a sequence S to be cloud free is

$$\mathbb{P}_S(\bar{C}) = (\mathbb{P}_p(\bar{C}))^{|S|} \quad (9)$$

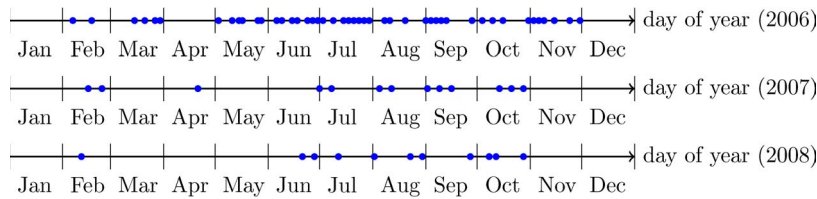


Fig. 4. Temporal distribution of images used from the two years. Each spot represents an acquired image.

with $|S|$ be the length of the sequence S . In this way, even with a low probability of cloud cover, the probability of a cloud-free sequence remains very low. Therefore, there are three main ways of dealing with clouds in the analysis of SITS, as follows:

- 1) letting the cloud-covered values in the data set;
- 2) removing predominantly cloud-covered images from the SITS;
- 3) removing the cloud-covered values only.

The first solution will add noise in the comparison of the radiometric profiles, since the cloud-covered values will be part of the radiometric behaviors, while they are not characteristic of the sensed surface. The second solution can only be used if the clouds are located on a set of images. In this way, this solution appears to be unrealistic for operational conditions. Moreover, all cloud-free values in the removed images will be ignored for the analysis, while remaining images (possibly containing a few clouds) will be treated according to the first solution. The third solution is obviously the most interesting one, since the cloud-covered values are removed beforehand, while preserving all cloud-free ones. However, this solution induces sequences with different lengths. In this way, this solution requires a similarity measure able to compare sequences with different lengths (implying different temporal sampling). Contrary to classical similarity measure between sequences, DTW is not limited to the comparison of sequences with equal lengths. DTW's ability to realign sequences makes it possible to operate nonlinear distortions on the temporal axis. In this way, DTW makes it possible to compare sequences with different lengths and sampling, while keeping its ability to find the optimal coupling of the two sequences.

V. MATERIALS AND METHODS

A. Image Data

We detail here the main information concerning the images used for this work. The area of study of this work is located near the town of Toulouse in the southwest of France. There are 70 Formosat-2 images acquired over three cultivation years:

- 46 images in 2006;
- 13 images in 2007;
- 11 images in 2008.

The temporal distribution of the images is given in Fig. 4 and one image of the three series is given in Fig. 5. Moreover, images sensed in 2006 are quite cloud-covered (about 30% of the “pixels”), while the two next series are quite cloud-free (less than 3%).

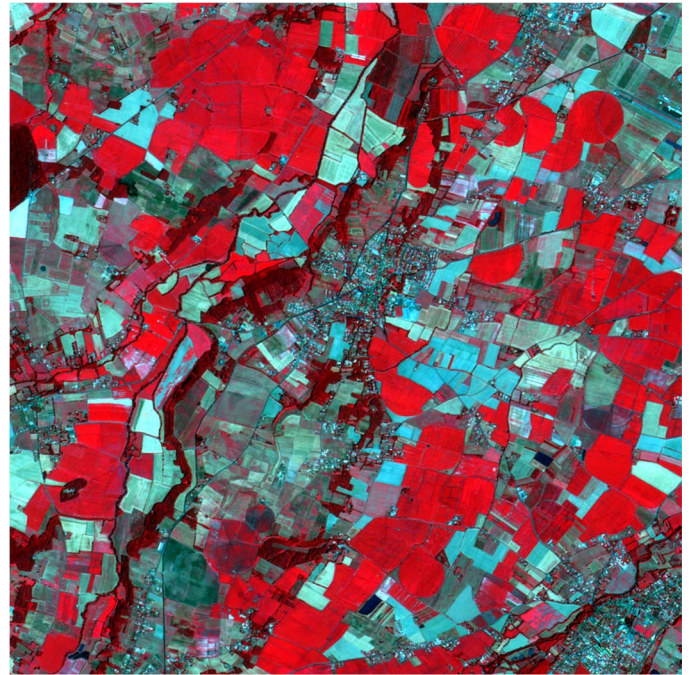


Fig. 5. One image from the series (August 4, 2007).

From these images, we use the multispectral product at a spatial resolution of 8 m, and only the three bands near-infrared, red, and green are kept, since the blue channel gives little information about vegetation and is very sensitive to atmospheric artifacts.

Before being used in this work, the Formosat-2 products have been orthorectified (guaranteeing that a pixel (x, y) covers the same geographic area throughout the image series). All images also undergo processes in order to make the radiometric pixel values comparable from one image to another. These processes consist of converting the digital counts provided by the sensor into a physical magnitude and in restoring their own contribution to the surface by correcting for atmospheric effects. From the instrument radiometric model, digital numbers are first converted into reflectances (normalized physical quantity of solar irradiance). The absolute calibration coefficients used in this step come from the monitoring of Formosat-2 sensor conducted by the French Space Agency. The inversion of the surface reflectance is then made by comparing the measured reflectance in simulations at the top of the atmosphere, carried out for atmospheric and geometric conditions of measurement. The elevation is taken into account by carrying out simulations for various altitudes, including a weighting of the atmospheric pressure and the amounts of aerosols and water vapor. The state

of the atmosphere at the time of the sensing is in turn characterized using meteorological sources (NCEP for the pressure and the humidity), using ozone data sources (TOMS or TOAST) and using aerosol data (SeaWiFS, Aeronet). Otherwise, climatological values are used.

Moreover, for each cultivation year, we have a land cover map produced by the method described in [43] and using a comprehensive ground reference data set. Also, cloud masks are produced using the cloud screening procedure described in [44].

B. Construction of the Time Series

Let us define how the sequences are built from this image time series. Definition 1 formalizes the concept of a multivalued image. For each image, a mask giving the position of clouds and their projection on the ground is available. Then, Definition 2 details how the time series are built from the sequence of images.

Definition 1: Let $\mathcal{S}_{image} = \langle I^1, \dots, I^N \rangle$ be a series of \mathcal{N} images of width \mathcal{W} and height \mathcal{H} . Let \mathcal{B} be the number of bands in the images. Each multivalued (with multiple bands) image $I^n (n \in [1, \mathcal{N}])$ can be seen as a function

$$I^n : \llbracket 1, \mathcal{W} \rrbracket \times \llbracket 1, \mathcal{H} \rrbracket \rightarrow \mathbb{Z}^{\mathcal{B}}$$

$$(x, y) \mapsto \prod_{b=1}^{\mathcal{B}} I_b^n(x, y) \quad (10)$$

with \prod be the Cartesian product.

Definition 2: Let \mathcal{S} be the data set built from the image time series. \mathcal{S} is the set of sequences defined as

$$\mathcal{S} = \{ \langle I^1(x, y), \dots, I^N(x, y) \rangle | x \in \llbracket 1, \mathcal{W} \rrbracket, y \in \llbracket 1, \mathcal{H} \rrbracket \}. \quad (11)$$

Then, each sequence is built as the series of tuples (NIR,R,G) for each pixel (x, y) in the image series.

Moreover, each satellite image I^n is associated to a cloud mask defining geographic areas which could be affected by a cloud. Thus, all the values $I^n(x, y)$ are removed/skipped from the series if the mask indicates a cloudy-affected value. Hence, the length of analyzed sequences can differ from one series to another. Since DTW is able to compare sequences with different lengths, these corrected time series can be used directly and prevent the interpolation of “cloud-contaminated” values [45].

C. Experimental Settings

We provide here some details about the experimental settings.

- The distance used between two coordinates of sequences is the squared Euclidean distance. As the square function is a strictly increasing function on positive numbers, and because we only use comparisons between distances, it is unnecessary to compute square roots. The same optimization has been used in [46], and is rather usual.

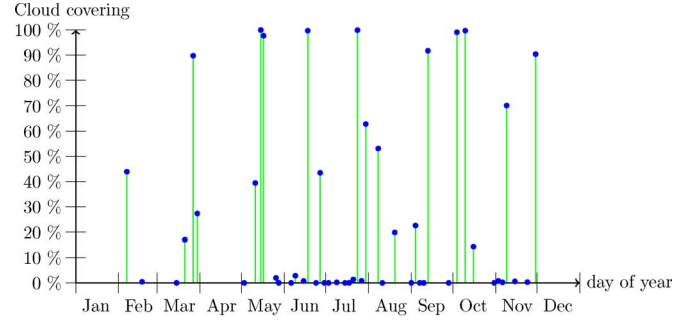


Fig. 6. Cloud covering of the SITS sensed in 2006.

TABLE II
LOSS OF CLOUD-FREE VALUES DEPENDING ON THE ACCEPTED
MAXIMUM CLOUD-COVERING OF THE IMAGES

| Maximum accepted cloud covering per image (%) | 40 | 30 | 20 | 10 | 0 |
|---|-----|-----|-----|-----|------|
| Eliminated images | 14 | 15 | 18 | 20 | 28 |
| Million of cloud-free pixels lost | 2.5 | 3.2 | 4.7 | 7.2 | 18.1 |

- The averaging method used for DTW is the association of DBA and adaptive scaling (AS) as described in [3] with 15 iterations of the process and constraining the AS process with the inertia of the clustering algorithm.
- The Kappa index, which is detailed in Appendix A, is used to compare the results.

VI. EXPERIMENTS

This section aims to demonstrate the utility of DTW for the analysis of SITS. We study three configurations using DTW.

- Section VI-A presents the clustering of the 2006 year, in order to show the ability of DTW to handle “cloud-contaminated” SITS.
- Section VI-B details a joint clustering of two cultivation years with 13 images over the first year (2007) and 11 images over the second year (2008). It aims at showing that the information redundancy sensed over each is useful in order to improve the generalization performance of the classifier;
- Section VI-C presents a domain adaptation example, learning classes on a year in order to classify another year. It will show that DTW is, not only theoretically, able to compare sequences with different lengths and sampling, in order to reuse formerly sensed image time series.

A. Clustering of the Series Sensed in 2006

The aim of this experiment is to demonstrate the ability of DTW to handle sequences with very different lengths in the same data set. Fig. 6 shows actually that the SITS sensed in 2006 is quite cloud covered with about 28% of the sensed pixels that are cloud covered. This is a typical example where we would neither want to keep (too numerous) the cloudy values in the data set, nor want to remove the (too numerous) cloud-covered images. To illustrate our point, Table II illustrates the loss of cloud-free values in the case of the second option, i.e., removing the too cloud-covered images. It can be seen that

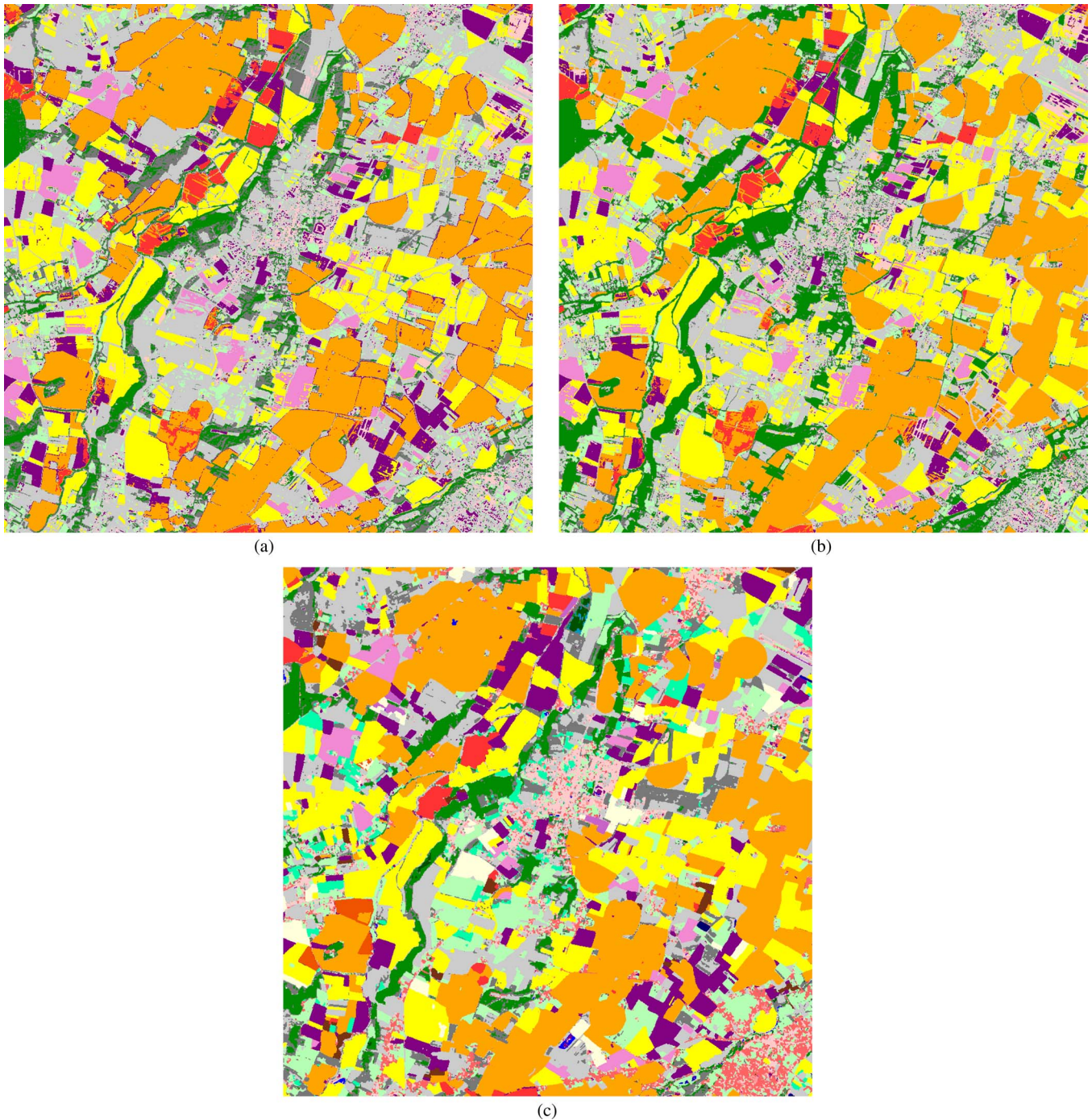


Fig. 7. Clustering results computed with DTW in the unconstrained and constrained version are depicted, respectively, in Fig. 7(a) and (b). This clustering can be compared to reference map in Fig. 7(c).

if the analysis is limited to images with a cloud-covering of less than 10% of the sensed area, 20 images are put aside and 7.2 million of cloud-free sensed pixels are not used in the analysis (corresponding to more than seven completely cloud-free images). Obviously, the relevance of the analysis would be then seriously affected.

Following the methodology proposed in Section V-B for the construction of the sequences, all cloud-covered values (as well as their projections on the ground) are removed. We are thus provided with a radiometric time series for each area (x, y) . Then, the times series have been divided into 24 clusters

(as many clusters as there are classes in the reference land cover map). Two configurations are studied: one without any constraint on the maximum time delay in the alignment of the sequences, and one with a maximum time delay of two month. Fig. 7(a) and (b) presents, respectively, the obtained results without and with a constraint on the time delay. Fig. 7(c) shows the land cover reference map.

Statistically, the result obtained with DTW have an agreement (Kappa) of 86.9% for the unconstrained version and of 87.2% for the constrained one. Visually, the main classes are quite well recognized, except for the meadow and the fallow

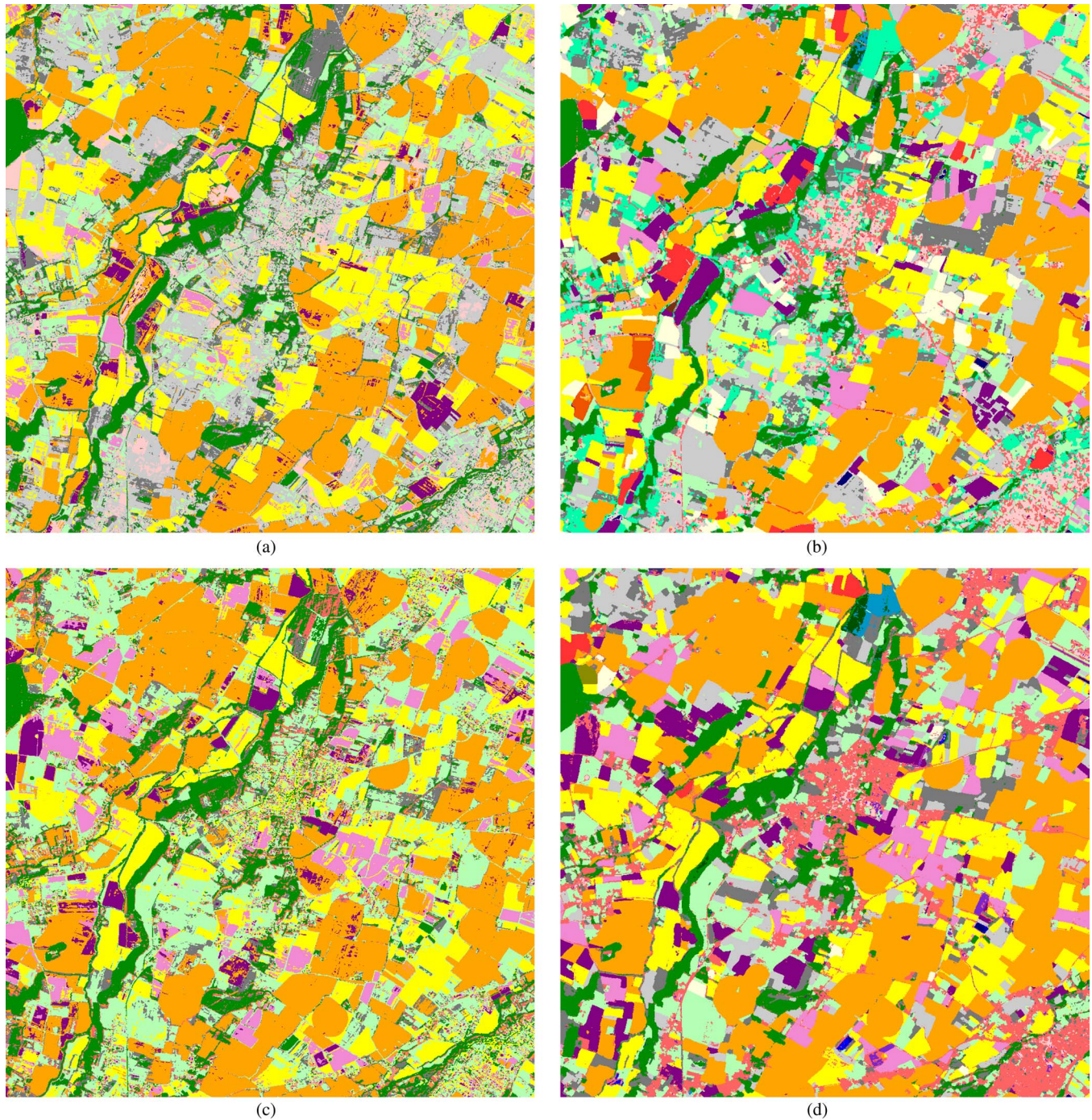


Fig. 8. Results of the joint clustering of sequences from 2007 and 2008. Fig. 8(a) [respectively Fig. 8(c)] shows the clustering map which can be compared to the land cover reference shown in Fig. 8(b) [respectively Fig. 8(d)]. The maps 8(a) and 8(c) have been recolored according to the land cover maps 8(b) and 8(d). The legend of the maps is given in Appendix B.

classes that are mixed up for the two configurations. However, in the unconstrained configuration, there is also a mix-up between the wild land class and the forest (in a broad sense), while there is not in the constrained version.

B. Joint Clustering of Two Cultivation Years

This experiment consists of clustering simultaneously the two years 2007 and 2008. Since DTW is able to compare sequences with different lengths, the K-MEANS algorithm is able to classify sequences built over 2007 and sequences built

over 2008, together. We expect this approach to enhance the ability of the clustering algorithm to form relevant clusters. Once the centroids of the K-MEANS algorithm are learned, similar classes (in terms of data representation) are represented by a single centroid. However, several centroids are predominantly representing sequences from only one year, for two main reasons:

- there are more classes represented in 2007 (25 classes) than in 2008 (22 classes);
- the temporal sampling of the data is not very same.

The second reason is more harmful than the first: if two phenomena are not sensed with a similar temporal discretization, sequences representing this phenomenon can be different enough to be separated by the clustering algorithm. Using the K-MEANS algorithm to cluster the sequences, this “bad” separation of the data can occur frequently since two clusters representing the same class can be much populated, making centroids easily diverge. In order to reduce this separation of classes, an Ascendant Hierarchical Clustering algorithm is used in order to aggregate these pairs of centroids. Even if these two behaviors are slightly different, their means (mean behaviors) will be very close to each other. Thus, starting with the K-MEANS algorithm parametrized to form 30 clusters, the 30 corresponding centroids are aggregated into 25 centroids. These 25 centroids are then used to classify separately sequences from 2007 and sequences from 2008. Thus, two clustering maps can be drawn: one for 2007 and one for 2008.

Fig. 8 shows results from the joint clustering of sequences from 2007 and sequences from 2008, as well as corresponding reference land cover maps.

Results obtained are spatially regular and connected which gives a first quality assessment of this result. We computed the Kappa index in order to compare the clustering results to the land cover reference maps. The clustering maps provided by the joint clustering obtained a Kappa of 88% for the year 2007 and a Kappa of 87% for the year 2008. In order to evaluate the improvement of this joint clustering, we also computed separate clustering of each year. The clustering over 2007 obtained a score 87%, while the one over 2008 obtained a score of 85%. From a qualitative standpoint and according to the reference maps, the main three classes, i.e., wheat (yellow), corn (orange), and forest (green) are correctly separated in the data space, for both maps. The main confusion concerns the wildland class (gray) and the meadow class (light green), since they are spectrally and temporally close to each other. Moreover, in 2008, the urban class (pink) is sometimes clustered with the wheat class, since there are no images during April and May, i.e., during the NDVI peak of the wheat crop.

These results show that DTW is able to exploit the information redundancy (in terms of type of behavior) between two years in order to obtain a better classification of each year with few images per year. Thus, the property of DTW to compare sequences with different lengths and samplings makes it possible to benefit from more diverse examples in order to better generalize the learned model.

C. Domain Adaptation Between Two Cultivation Years

The second experiment consists of a domain adaptation example: learning classes over 2007 thanks to the reference land cover map, in order to classify the 2008 image series.

Classical domain adaptation methodologies are learning classes on one image in order to classify another one. Moreover, the learning process is usually using external knowledge in order to improve the accuracy of the model, if its generalization is inaccurate. However, in our case, the model is learned on one SITS in order to classify another one. In this way, the learned model is better generalized, and adding supplement



Fig. 9. Results of the adaptation domain experiment from 2007 to 2008. The classification map is automatically colorized according to the labels of centroids.

knowledge is proven to be unnecessary. The true difficulty is actually to make these two series comparable. Technically, we use once again the ability of DTW to compare sequences with different lengths and sampling in classifying 2008 sequences with centroids learned on 2007 ones. This experiment aims at showing that this comparison is also thematically relevant.

The idea is first to cluster sequences from 2007, then label and split these clusters with the reference land cover map of 2007, and finally to classify the sequences built over 2008 with these labeled centroids. The process is as follows:

- 1) make 25 clusters with the sequences sensed over 2007 using the K-MEANS algorithm;
- 2) for each formed cluster:
 - a) compute its distribution (histogram) in terms of classes of the reference land cover map;
 - b) for each class appearing in this histogram which is populated with more than one percent of the image (i.e., 1000 pixels), build a new centroid (in averaging the corresponding sequences);
- 3) classify sequences from 2008 by assigning the class of the closer built centroid.

Fig. 9 shows results of this domain adaptation experiment; the 2008 reference land cover map can be found in Fig. 8(d).

In the same way as for the previous experiments, we computed the Kappa index, which gave a score of 84% compared to the 2008 ground truth. Qualitatively, the corn class is quite well recognized even if it was sometimes confused with the soybean class (red) which has a similar spectral evolution at this sampling. Moreover, in the same way than in the previous experiment, the urban class is sometimes clustered with the

wheat class, since an additional image would be necessary in May in order to separate them.

This experiment shows two aspects. First, it confirms the relevance of the use of DTW to compare sequences with different lengths on one side, and particularly with different temporal sampling on the other side. Second, the Kappa score obtained shows that the knowledge associated to an image time series acquired one year can be used directly to classify another year.

VII. CONCLUSION

In this paper, we have introduced the DTW as a tool to deal with two of the main issues raised by high temporal resolution satellite image series, namely the irregular sampling in the temporal dimension and the need for comparison of pairs of time series having different number of samples.

In order to illustrate the properties of DTW, we have presented three applications to real data. The first one demonstrated the relevance of DTW to handle cloud-contaminated SITS and its ability to be constrained according to expert's knowledge about the land cover phenology.

The second one showed how DTW can be used to jointly cluster two sets of data acquired over the same area during two different years and therefore taking benefit from the cyclic behavior of—mostly vegetation—land surfaces. The fact that each year has a different set of dates and that those dates are not regularly distributed through the year does not prevent their joint use in order to improve the clustering results. We showed that the joint clustering gives better results than the separate clustering of each year.

The third example belongs to the family of domain adaptation problems where the information extracted from a data set is used to process a different data set. In this case, we used DTW in order to generate clusters for one year and apply them to the clustering of another year which, again, has a different set of available acquisition dates. In this case, the power of the method resides in the fact that ground data gathered in during the 2007 agricultural season can be used to process the following year without the need for new field work.

DTW is able to loosen several temporal constraints on image sensing. It thus emerges as a flexible tool to handle heterogeneous data produced by remote sensing.

APPENDIX A



























EVALUATING CLUSTERING RESULTS

In order to compare clustering results to land cover reference maps, we compute the Kappa index (κ) which is defined by

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (12)$$

where $\Pr(a)$ is the relative agreement among raters, and $\Pr(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly choosing each category. There are many ways to compute this index. When data are not labeled (as it is the case with clustering), this computation consists of taking all point

TABLE III
LEGEND OF THE MAPS

| Color | Class |
|---|------------------------------|
|  | corn |
|  | corn for silage |
|  | non-irrigated corn |
|  | wheat |
|  | sunflower |
|  | sorghum |
|  | sorghum II |
|  | soybean |
|  | barley |
|  | pea |
|  | rape |
|  | broad-leaved tree |
|  | conifer |
|  | poplar tree |
|  | eucalyptus |
|  | water |
|  | lake |
|  | gravel pit |
|  | meadow |
|  | temporary meadow |
|  | fallow land |
|  | wild land |
|  | high density housing surface |
|  | specific urban surface |
|  | low density housing surface |
|  | mineral surface |

couples $(p_1, p_2) = ((x_1, y_1), (x_2, y_2))$ and see the configuration of these two points in each partition (the clustering and the ground truth). There are four possible configurations:

- 1) p_1 and p_2 belong to the same partition both in the clustering and in reference map;
- 2) p_1 and p_2 belong to the same partition in the clustering but not in the reference map;
- 3) p_1 and p_2 belong to the same partition in the reference map but not in the clustering;
- 4) p_1 and p_2 belong to the same partition neither in the reference map nor in the clustering.

In order to evaluate the number of couples of points in each configuration, a counter can be associated to each configuration and incremented each time a configuration appears:

- 1) a counter ss for the “same same” configuration;
- 2) a counter sd for the “same different” configuration;
- 3) a counter ds for the “different same” configuration;
- 4) a counter dd for the “different different” configuration.

Thus, the Kappa index can be computed with

$$\Pr(a) = \frac{ss + dd}{ss + sd + ds + dd} \quad (13)$$

$$\Pr(e) = \frac{(ss + sd) \times (ss + ds) + (sd + dd) \times (ds + dd)}{(ss + sd + ds + dd)^2} \quad (14)$$

ACKNOWLEDGMENT

The authors would like to thank the French Space Agency (CNES) and Thales Alenia Space for supporting this work and would also like to thank the colleagues from CESBIO (Danielle Ducrot, Claire Marais-Sicre, Olivier Hagolle, and Mireille Huc) for providing the land cover maps and the geometrically and radiometrically corrected FORMOSAT-2 images.

REFERENCES

- [1] H. Sakoe and S. Chiba, "A dynamic programming approach to continuous speech recognition," in *Proc. 7th Int. Congr. Acoust.*, 1971, vol. 3, pp. 65–69.
- [2] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 1, pp. 43–49, Feb. 1978.
- [3] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, Mar. 2011.
- [4] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Digital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, May 2004.
- [5] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 37, pp. 2365–2401, Jun. 2004.
- [6] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and maf postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.
- [7] P. Howarth, J. Piwowar, and A. Millward, "Time-series analysis of medium-resolution, multisensor satellite data for identifying landscape change," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 6, pp. 653–663, Jun. 2006.
- [8] L. Bruzzone, D. Prieto, and S. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1350–1359, May 1999.
- [9] F. Hall, D. Botkin, D. Strebel, K. Woods, and S. Goetz, "Large-scale patterns of forest succession as determined by remote sensing," *Ecology*, vol. 72, no. 2, pp. 628–640, Apr. 1991.
- [10] C. Munyati, "Wetland change detection on the Kafue Flats, Zambia, by classification of a multitemporal remote sensing image dataset," *Int. J. Remote Sens.*, vol. 21, no. 9, pp. 1787–1806, Jun. 2000.
- [11] G. Foody, "Monitoring the magnitude of land-cover change around the southern limits of the Sahara," *Photogramm. Eng. Remote Sens.*, vol. 67, no. 7, pp. 841–847, Jul. 2001.
- [12] W. Cohen, M. Fiorella, J. Gray, E. Helmer, and K. Anderson, "An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery," *Photogramm. Eng. Remote Sens.*, vol. 64, no. 4, pp. 293–300, 1998.
- [13] L. Bruzzone and D. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
- [14] F. Melgani, G. Moser, and S. B. Serpico, "Unsupervised change-detection methods for remote-sensing images," *Opt. Eng.*, vol. 41, no. 12, pp. 3288–3297, 2002.
- [15] W. Todd, "Urban and regional land use change detected by using Landsat data," *J. Res. US Geol. Survey*, vol. 5, pp. 529–534, 1977.
- [16] J. R. Jensen, "Urban change detection mapping using landsat digital data," *Cartography Geographic Inf. Sci.*, vol. 8, no. 2, pp. 127–147, Oct. 1981.
- [17] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.
- [18] M. Piles, D. Entekhabi, and A. Camps, "A change detection algorithm for retrieving high-resolution soil moisture from smap radar and radiometer observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4125–4131, Dec. 2009.
- [19] S. S. Young and C. Y. Wang, "Land-cover change analysis of china using global-scale pathfinder AVHRR landcover (PAL) data, 1982–92," *Int. J. Remote Sens.*, vol. 22, no. 8, pp. 1457–1477, May 2001.
- [20] E. F. Lambin and A. H. Strahlers, "Change-vector analysis in multitemporal space: A tool to detect and categorize land-cover change processes using high temporal-resolution satellite data," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 231–244, 1994.
- [21] R. Johnson and E. Kasischke, "Change vector analysis: A technique for the multispectral monitoring of land cover and condition," *Int. J. Remote Sens.*, vol. 19, no. 3, pp. 411–426, 1998.
- [22] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.
- [23] G. S. Burns and A. T. Joyce, "Evaluation of land cover change detection techniques using Landsat MSS data," in *Proc. 7th Pecora Symp.*, 1981, pp. 1127–1134.
- [24] H. Hanaizumi, H. Okumura, and S. Fujimura, "Change detection from remotely sensed multi-temporal images using spatial segmentation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Espoo, Finland, Jun. 1991, vol. 2, pp. 1079–1082.
- [25] C. S. Jha and N. V. M. Unni, "Digital change detection of forest conversion of a dry tropical Indian forest region," *Int. J. Remote Sens.*, vol. 15, no. 13, pp. 2543–2552, Sep. 1994.
- [26] M. K. Ridd and J. Liu, "A comparison of four algorithms for change detection in an urban environment," *Remote Sens. Environ.*, vol. 63, no. 2, pp. 95–100, Feb. 1998.
- [27] A. Julea, N. Méger, E. Trouvé, and P. Bolon, "On extracting evolutions from satellite image time series," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2008, vol. 5, pp. V-228–V-231.
- [28] F. Petitjean, P. Gançarski, F. Masseglia, and G. Forestier, "Analysing satellite image time series by means of pattern mining," in *Proc. 11th Int. Conf. Intell. Data Eng. Automated Learn.*, vol. 6283, ser. *Lecture Notes in Computer Science*, 2010, pp. 45–52, Springer-Verlag.
- [29] A. Julea, N. Méger, P. Bolon, C. Rigotti, M.-P. Doin, C. Lasserre, E. Trouvé, and V. Lăzărescu, "Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1417–1430, Apr. 2011.
- [30] L. Andres, W. Salas, and D. Skole, "Fourier analysis of multi-temporal AVHRR data applied to a land cover classification," *Int. J. Remote Sens.*, vol. 15, no. 5, pp. 1115–1121, Mar. 1994.
- [31] T. Celik, "Unsupervised multiscale change detection in multitemporal synthetic aperture radar images," in *Proc. 17th EUSIPCO*, 2009, vol. 15, pp. 47–51.
- [32] T. Celik and K.-K. Ma, "Multitemporal image change detection using undecimated discrete wavelet transform and active contours," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 706–716, Feb. 2011.
- [33] E. Keogh, S. Lonardi, and C. A. Ratanamahatana, "Towards parameter-free data mining," in *Proc. SIGKDD ACM Int. Conf. Knowl. Discov. Data Mining*, 2004, pp. 206–215.
- [34] A. N. Kolmogorov, "On tables of random numbers," *Sankhyā: Indian J. Stat. Ser. A*, vol. 25, no. 4, pp. 369–376, Dec. 1963.
- [35] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Sov. Phys. Doklady*, vol. 10, no. 8, pp. 707–710, Dec. 1965.
- [36] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh, "Indexing multidimensional time-series," *VLDB J.*, vol. 15, no. 1, pp. 1–20, Jan. 2006.
- [37] D. Sankoff and J. Kruskal, "The symmetric time-warping problem: From continuous to discrete," in *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*. Reading, MA: Addison-Wesley, 1983, pp. 125–161.
- [38] J. Aach and G. M. Church, "Aligning gene expression time series with time warping algorithms," *Bioinformatics*, vol. 17, no. 6, pp. 495–508, Jun. 2001.
- [39] Z. Bar-Joseph, G. Gerber, D. K. Gifford, T. S. Jaakkola, and I. Simon, "A new approach to analyzing gene expression time series data," in *RECOMB: Proc. 6th Annu. Int. Conf. Comput. Biol.*, New York, 2002, pp. 39–48.
- [40] D. M. Gavrilu and L. S. Davis, "Towards 3-D model-based tracking and recognition of human movement: A multi-view approach," in *Proc. IEEE Int. Workshop Autom. Face Gesture Recognit.*, 1995, pp. 272–277.
- [41] T. Rath and R. Manmatha, "Word image matching using dynamic time warping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, vol. 2, pp. II-521–II-527.
- [42] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-23, no. 1, pp. 67–72, Feb. 1975.
- [43] S. Idrabim, D. Ducrot, D. Mammass, and D. Aboutajdine, "An unsupervised classification using a novel ICM method with constraints for land cover mapping from remote sensing imagery," *Int. Rev. Comput. Softw.*, vol. 4, no. 2, pp. 165–176, Mar. 2009.
- [44] O. Hagolle, M. Huc, D. V. Pascual, and G. Dedieu, "A multi-temporal method for cloud detection, applied to Formosat-2, VENUS, LANDSAT and SENTINEL-2 images," *Remote Sens. Environ.*, vol. 114, no. 8, pp. 1747–1755, Aug. 2010.
- [45] A.-B. Salberg, "Land cover classification of cloud-contaminated multi-temporal high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 377–387, Jan. 2011.
- [46] A. W.-C. Fu, E. J. Keogh, L. Y. H. Lau, C. A. Ratanamahatana, and R. C.-W. Wong, "Scaling and time warping in time series querying," *VLDB J.*, vol. 17, no. 4, pp. 899–921, Jul. 2008.



François Petitjean received the M.Sc. degree in computer science from the University of Strasbourg, Strasbourg, France in 2009. He is currently working toward the Ph.D. degree supported by the French Space Agency and Thales Alenia Space at the Image Sciences, Computer Sciences and Remote Sensing Laboratory, Strasbourg, France.

His research interests include complex time series clustering and remote sensing image analysis.



Pierre Gançarski received the Ph.D. degree and the Habilitation in computer science from the University Louis Pasteur, Strasbourg, France, in 1989 and 2007, respectively.

He is currently a Full Professor with the Department of Computer Science from the University of Strasbourg, Strasbourg, France. His current research interests include collaborative multistrategical clustering with applications to complex data mining and remote sensing image analysis.



Jordi Inglada received the Telecommunications Engineer degree from the Universitat Politècnica de Catalunya, Barcelona, Spain, and the École Nationale Supérieure des Télécommunications de Bretagne, Brest, France, in 1997, and the Ph.D. degree in signal processing and telecommunications from Université de Rennes 1, Rennes, France, in 2000.

He is currently with the Centre National d'Études Spatiales (French Space Agency), Toulouse, France, working in the field of remote sensing image processing at the CESBIO laboratory. He is in charge of

the development of image processing algorithms for the operational exploitation of Earth observation images, mainly in the field of multitemporal image analysis for land use and cover change.

Dr. Inglada is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.