

Spatio-temporal stability analysis in Satellite Image Times Series^{*}

Mohamed Chelal¹[0000-0002-0173-7368], Camille Kurtz¹[0000-0001-9254-7537],
Anne Puissant²[0000-0002-3240-9244], and Nicole Vincent¹[0000-0002-0151-0622]

¹ Université de Paris, LIPADE, Paris, France.

{firstname.lastname}@u-paris.fr

² Université de Strasbourg, LIVE, Strasbourg, France.

{firstname.lastname}@unistra.fr

Abstract. Satellite Image Time Series (SITS) provide valuable information for the study of the Earth's surface. In particular, this information may improve the comprehension, the understanding and the mapping of many phenomena such as earthquake monitoring, urban sprawl or agricultural practices. In this article, we propose a method to define new spatio-temporal features from SITS based on the measure of the temporal stability. The proposed method is based on a compression algorithm named Run Length Encoding leading to a novel image representation from which stability features can be measured. Such features can then be used in several applications such as SITS summarizing, to make easier the interpretation of such a data-cube, or the classification of spatio-temporal patterns. The preliminary results obtained from a series of 50 Sentinel-2 optical images highlight the interest of our approach in a remote sensing application.

Keywords: Satellite Image Time Series (SITS), spatio-temporal features, Run Length Encoding, temporal stability, Sentinel-2.

1 Introduction

Image time series, such as Satellite Image Time Series (SITS) or MRI functional sequences in the medical domain consist of ordered sets of images taken from the same scene at different dates. Such data provide rich information with the temporal evolution of the studied areas. In the context of remote sensing, SITS provide enormous amounts of information that allow the monitoring of the surface of our planet. Recently new constellations of satellites have been launched to observe our territories, producing optical images with high spatial, spectral and temporal resolution. For example, the Sentinel-2 sensors provide SITS with a revisit time of 5 days and a spatial resolution of 10 – 20 meters.

One of the major applications of SITS is the cartography / mapping of land covers (e.g. agricultural crops, urban areas) and the identification of land use

^{*} The authors thank the ANR for supporting this work – Grant ANR-17-CE23-0015.

changes (e.g. urbanization, deforestation). The availability of the temporal information makes possible to understand the evolution of our territory while analyzing complex patterns leading to produce or update accurate land cover maps of a particular area [11].

A main challenge for the automation of SITS analysis is to consider at the same time both the temporal and the spatial domains of the data-cube in order to take into account and benefit from all the (complementary) information carried by the data. Indeed, most of the existing methods for SITS analysis are actually solely based on temporal information [18, 19]. However, for many complex tasks such as the understanding of heterogeneous agricultural practices, this may be not sufficient to get satisfactory results. The consideration of spatio-temporal features from SITS should allow the discrimination between different complex land cover classes, related to agricultural and urban land-cover practices. Note that we do not aim here to produce temporal land-cover maps or to study land use changes (e.g. urbanization) but our objective is to analyze complex land-cover classes prone to confusions when a single date image is used.

This article focuses on the specific problem of spatio-temporal features extraction from image time series that can be used for different applications. For example, such information may be used to summarize a SITS, improving the difficult photo-interpretation of the sensed scenes from a set of many images acquired at different dates, or for classification of different regions of interest into thematic classes (e.g. agricultural crops, vegetation, urban). Our contribution relies on spatio-temporal features that are based on the measure of the spatio-temporal stability of a zone, using a compression algorithm named Run Length Encoding. By compressing a data-cube with this strategy, we obtain a novel SITS representation from which stability features can be measured.

This article is organized as follow. Section 2 recalls some existing methods for SITS analysis. Section 3 presents our proposal to extract spatio-temporal features. Section 4 describes the experimental study while Section 5 concludes.

2 Related works

SITS allows the observation of the Earth surface and the understanding of the evolution of our environment, where various changes can occur over the time (e.g. following natural disasters, urbanization, agricultural practices). Such environmental changes may be of different types, origins and duration [6].

Pioneer methods for SITS analysis take into account single images or stacks of images. Those that consider each image, compute different measurements per pixel as independent features and involve them in classical machine learning-based procedures. In such approaches, the date of the measurements is often ignored in the feature space. Methods designed for bi-temporal analysis, can locate and study abrupt changes occurring between two observations. These methods include image differencing [4], ratio-ing [13] or change vector analysis [14]. We also find statistical methods that consider two or more images, such as linear transformation (PCA and MAF) [16, 17].

More recent methods were designed to directly deal with the specificity of image time series. This category of methods includes multi-date classification approaches, such as radiometric trajectory analysis [25], or sequential patterns techniques that group pixels that share common temporal pattern [15]. The latter strategy exploits the notion that land cover can vary through time (e.g. because of seasons, vegetation evolution [24]), and related methods take into account the order of measurements by using dedicated time series analysis methods [3]. Every pixel is viewed as a temporally ordered series of measurements, and the changes of the measurements through time are analyzed to find (temporal) patterns.

Other type of methods start by transforming the original representation of the SITS into a new one. For example, the analysis can be operated in the “frequency-domain” that includes spectral analysis or wavelet analysis [2]. Concerning the classification methods, the classical way is to measure similarity between any incoming sample and the training set and then to assign the label of the most similar class. Other methods extract more discriminative “hand-crafted” features from a new enriched space [20, 22] before using the classifier.

Recently, deep learning approaches have been employed to analyze satellite images. For example, convolutional neural networks (CNN) can be applied with $2D$ convolutions to deal with the spatial domain [9]. CNNs have been also applied successfully to perform SITS classification; in this case $1D$ convolutions dealing only with the temporal domain [18] have been proposed. Other architectures of deep learning that is designed for time series are recurrent neural network (RNN) such as Long-Short Term Memory (LSTM), used successfully in [10, 23]. In this context, deep learning approaches outperform traditional classifications algorithm as Random Forest [12], but they do not directly take into account the spatial dimension of the data as they consider pixels in an independent way. Although, some approaches have been proposed to consider both the temporal and the spatial dimensions of the $2D + t$ data-cube [7], the expressiveness of the underlying convolutional features is difficult to use for the interpretation of the content of the sensed scenes.

In this article, we propose an approach to define a set of spatio-temporal features from an image time series. Such spatio-temporal features can be used for different needs, for example to summarize a SITS, in order to improve and to make easier the scene interpretation, or to enrich the feature space, increasing the separability of complex land-cover classes in a classification task. The proposed approach measures the stability of a zone based on a lossless compression algorithm, here the Run Length Encoding (RLE) [8]. The notion of temporal stability has been initially proposed in a previous article [5]. In the following, we extend this strategy to now measure the spatio-temporal stability of an area, by relaxing both temporal and spatial constraints when assessing the equality between consecutive pixels through the time.

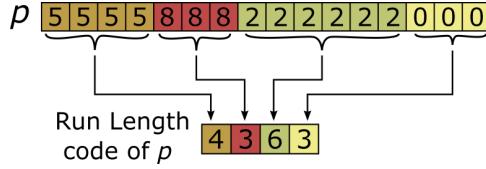


Fig. 1. Compression of a temporal pixel p with a Run Length Encoding (RLE).

3 Proposed method

The proposed method aims to analyze the spatio-temporal stability or change of image pixels from a SITS. The spatio-temporal features are based on the measure of the spatio-temporal stability of a zone, using a compression algorithm, namely Run Length Encoding (RLE) algorithm, leading to a novel image representation from which stability features can be measured. In this Section, we first explain how the stability can be measured in the datacube induced by the SITS data. Since the notion of stability involves the study of the repetition of successive values, the notion of equality and its application level enable to relax both temporal and / or spatial constraints, leading to the definition of spatio-temporal features.

3.1 Stability measurement

The stability is based on the repetition of the successive values. In the literature, different methods have been proposed to measure this information. In our case, we choose a compression method named Run Length Encoding (RLE) [8] that was already used for time series analysis in [1, 21]. RLE is a lossless compression algorithm. It allows the compression of a vector v of length L by storing both the number of times a value is repeated successively and its value. In our case we only consider the first information and omit the value. The resulting vector is noted $RLE(v)$ with length l . Figure 1 illustrates the computation of the RLE on a temporal pixel p .

In the following, a SITS is noted $(I_t)_{t \in \llbracket 1, T \rrbracket}$ where t is the date of acquisition for each image I . The computation of the RLE is applied on a temporal pixel, noted $p(x, y) = (p_t(x, y))_{t=1}^T$, with $x \in [1, W]$ and $y \in [1, H]$, W and H representing respectively the width and the height of the images.

Our strategy here is to employ the RLE to change the representation of the SITS into a less voluminous structure, where stability information can be more easily measured. In this context, we define three features that can be extracted from this new representation for a given temporal pixel p :

- The **Maximal Stability** (noted MS) feature captures the longest period where the pixel intensity stays stable (i.e. without change) through the time series. It is expressed in term of number of days of the year. Such a feature value can be computed as:

$$MS(p) = \|RLE(p)\|_\infty \quad (1)$$

- The **Max Stability Start** (noted MSS) corresponds to the beginning of the maximum stability period of a pixel through time. Such information is then directly related to the Maximal Stability (MS). This feature can be informative / discriminative for different specific tasks, for example the maximum stability in artificialized areas (e.g. built-up area, impervious surfaces) starts earlier than in non-artificial zones since they do not change over time. The MSS feature value can be defined as:

$$MSS(p) = \left(\sum_{i=1}^{t_0-1} RLE(p)_i \right) \text{ with } t_0 / RLE(p)_{t_0} = MS(p) \quad (2)$$

- The **Number of changes** (noted NB) corresponds to the number of stability ranges, linked to the number of changes in the area covered by the studied temporal pixel. Such a feature value can be computed as:

$$NB(p) = l_p \quad (3)$$

The computation of RLE is based on the analysis of the equality between successive values. However, the images composing a SITS are not acquired at the same time. Although the images are generally corrected, the variability of the pixel intensity values can be important along the series since the distributions of the pixel values are not in the same dynamic from an image to another. This may be a consequence of the seasons and the different illumination conditions. To deal with this issue, the notion of equality has to be carefully studied in order to evaluate, in a more realistic manner, if two successive (temporal) pixel intensity values can be considered as equal or not.

3.2 Notion of equality

The equality is a binary equivalence relation that compares two objects of a same set E . They are considered as identical if a given predicate P holds:

$$\begin{aligned} \epsilon : E \times E &\rightarrow \text{Bool} \\ o_1, o_2 &\rightarrow \begin{cases} \text{True} & \text{if } P(o_1, o_2) \\ \text{False} & \text{else} \end{cases} \end{aligned} \quad (4)$$

When a SITS is considered, the pixel values being either continuous or discrete in interval such as $[0, 255]$ or hypercube when vectorial values are considered, the equality of values is not always significant. Then, to fix this problem, we start by applying a quantization of the pixel values. The quantization must not be applied to each image I_t , nor at temporal pixel p , it has to be done at a global level of all the pixels of the data-cube $(I_t)_t$. The quantization could be regular, fixed with respect of the usual distribution of the values or it can be adapted to the image series, to the nature of the characteristic used.

In order to perform the quantization, we apply a clustering algorithm that enables to define the significant intervals. In our case a k -Means algorithm, $k_{quantiz}$ being a parameter of the method, is used and different values have been

experimented according to the precision needed in our problem. Then, the pixel values are replaced by the cluster label belonging to $\{1, 2, \dots, k_{quantiz}\}$ defining a new STIS $(J_t)_{t \in \llbracket 1, T \rrbracket}$. The RLE is then actually applied on $(J_t)_t$ pixels. In this way, we define the predicate P by the strict equality between the objects, as $P(o_1, o_2) = (o_1 = o_2)$.

3.3 Towards spatio-temporal stability

The features presented in Section 3.1 are defined for a temporal pixel and they can be naturally considered as temporal features. Since our strategy relies on the compression ability of the RLE, we also suppose that the data are as clean as possible. But with satellite images, this may not be always the case. Some images from the series can be affected by noise (e.g. salt and pepper, undetected clouds) and a SITS can be potentially affected by registration problems from one image to another. In this case, the RLE may not reflect the reality of the content of the sensed scene along time. To handle this, we are going to relax the equality definition in the temporal and / or the spatial domains, leading to an approximation of the RLE, noted \widetilde{RLE} , that can absorb these different noises. It is no longer a lossless compression scheme.

The RLE computation is mainly based on the notion of “runs”, computed over a sequence, a run being a sub-sequence composed of successive (identical) repeated values. In the classical RLE, the predicate P is the one defined previously, leading to a “hard” equality relation. To compute the approximation of the RLE, the strategy here is to compute approximated runs over the sequence by relaxing the predicate P , used to estimate the equality between successive values. Of course the aim is to favor the shortest \widetilde{RLE} , that is to say, the longest runs that are assumed to be resumed by a single value. Then it is no more possible to have a linear process of the series but, recursively, the longest runs on the main series and on the sub-series when some parts of the series are already compressed.

We first define several ways to consider that a sub-series is “constant” and to compute an approximated run. This will be done through the relaxation of the predicate P in the temporal domain, in the spatial domain and finally in the spatio-temporal domain.

Let us note $s = (s_0, s_1, s_2, \dots, s_n)$ a series. Our goal here is to compute the length of the longest run in this series, starting from s_0 :

- **Temporal relaxation** The first possibility is to relax the temporal domain. When $i \geq 1$, we define a new predicate as:

$$P_{s,i}^{temp} = (s_i = s_0) \vee (s_{i+1} = s_0) \quad (5)$$

where \vee represents the logical or. This predicate makes it possible to skip a time value when comparing two consecutive elements of the series, according to the value of s_0 .

- **Spatial relaxation** When we relax the spatial domain, the pixel value at instant s_0 is compared with the next value in $t + 1$ and its neighbors in a square window (a patch), which provides a certain spatial flexibility during the comparison. With this strategy, we can avoid noise (e.g. salt and pepper) and potential problems of image registration disturbing the comparison. With the series s we associate the series of the neighbor of each s_i , $sN = (Ns_0, Ns_1, Ns_2, \dots, Ns_n)$ and we define a new predicate as:

$$P_{s,i}^{spatio} = s_0 \in Ns_i \quad (6)$$

- **Spatio-temporal relaxation** A comparison completely relaxed is extracted by combining both spatial and temporal relaxations. This spatio-temporal relaxation is obtained with:

$$P_{s,i}^{spatio-temp} = P_{s,i}^{spatio} \vee P_{s,i}^{temp} \quad (7)$$

Given these predicates, we can now compute the length of the longest run of the series, starting from s_0 . We first define a function c as

$$\begin{aligned} c : \quad \mathbb{B} &\rightarrow \mathbb{Z} \\ x &\mapsto \begin{cases} 1 & \text{if } x \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (8)$$

The length of the longest run (noted LR) can then be computed (using c to build a counter) as:

$$LR(s) = \max_k \left\{ k \in [1, n - 1] : \sum_{i=1}^k c(P_{s,i}^*) = k \right\} + 1 \quad (9)$$

where $P_{s,i}^*$ is one of the relaxed predicates mentioned above.

Algorithmic aspects Now that we know how to compute an approximated run beginning at a specific position, we can compute \widetilde{RLE} .

Let us consider a temporal pixel $p(x, y) = (p_t(x, y))_{t=1}^T$ of a SITS, noted p hereinafter. We note \tilde{p}_t the sub-series of p beginning in p_t and we have

$$LR(p) = \max_t LR(\tilde{p}_t) \quad \text{and} \quad ilr(p) = \arg \max_t LR(\tilde{p}_t) \quad (10)$$

with $ilr(p)$ providing the position (index) of the starting point of the longest run of p . To compute all the runs composing \widetilde{RLE} , we implemented a greedy optimization algorithm that works as follows. The function to optimize is to obtain a RLE whose number of elements is the smallest possible. We then select the longest run using Equation 10, keep it as an element of the final \widetilde{RLE} , and then re-apply this process, recursively (using a divide-and-conquer paradigm), to the left and to the right of the selected run, until we have considered all the temporal values of the pixel.

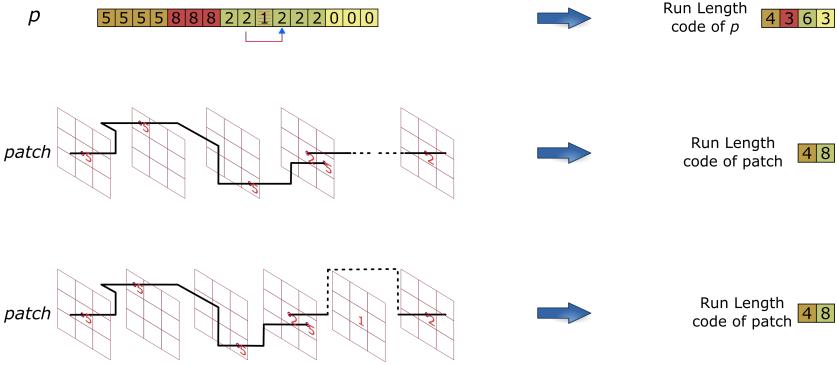


Fig. 2. Different ways for calculating an approximated Run Length Encoding (\widetilde{RLE}) by considering different relaxation strategies. (Top) equality with temporal relaxation; (Center) equality with spatial relaxation; (Bottom) equality with spatio-temporal relaxation.

The fundamental relation of the recursive process can be written:

$$\text{if } p = (p_t)_{t=1}^T \text{ with } \forall t, p_t = c \text{ then } \widetilde{RLE}(p) = T \quad (11)$$

$$\widetilde{RLE}(p) = \left(\widetilde{RLE}(p_1 \dots p_{ilr(p)}) LR(p) (\widetilde{RLE}(p_{ilr(p)+LR(p)} \dots p_T)) \right) \quad (12)$$

By considering these different relaxation strategies, we can obtain several approximations of the RLE for one temporal pixel p , noted \widetilde{RLE}_{temp} when considering the temporal relaxation, \widetilde{RLE}_{spatio} when considering the spatial relaxation, and $\widetilde{RLE}_{spatio-temp}$ when considering the spatio-temporal relaxation. Figure 2 illustrates the results of \widetilde{RLE} with the different relaxation strategies. From these novel representations, the stability characteristics (Maximal Stability – MS , Max Stability Start – MSS and Number of changes – NB) can be computed leading to different versions of spatio-temporal features.

3.4 Stability summarization

After extracting the three proposed features, following the hard equality strategy or a relaxed one, they can be used to summarize the SITS by combining them into one false color image, noted TS . The color composition of the summary image is following: MS in the red channel (R); NB in the green channel (G) and MSS in the blue channel (B).

In this way, instead of analyzing the whole series of images, we will only analyze a single image that summarizes all the SITS. The combination of the summary can be applied for the classical RLE (without relation) and for each approximated \widetilde{RLE} , we note respectively TS (from the classical RLE), TS_{temp} , TS_{spatio} and $TS_{spatio-temp}$ (from the \widetilde{RLE}) the resulting summary images.

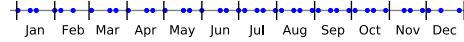


Fig. 3. Temporal distribution (over 2017) of the $T = 50$ images from the SITS.

4 Experimental study

The proposed method has been employed in a remote sensing application related to the analysis of land-cover from a SITS. In this experimental study, we want to highlight the ability of our features to capture and to summarize spatio-temporal (stability) patterns, that cannot be handled using a single image of the SITS, the summary may be useful to assist in understanding and interpreting a terrestrial sensed scene. The experimental study is divided into two parts. In the first part, we apply the proposed method on a SITS in order to summarize it and to visualize the obtained results. In the second part, we involved the proposed spatio-temporal features in a binary classification task for the analysis of urban land-cover thematic classes.

4.1 Materials

The data used in this experimental study is an optical SITS composed of $T = 50$ images. The images, provided by the satellite Sentinel-2, have been sensed in 2017 over the same geographical area, East of France. The acquired images have been corrected and orthorectified by the French Theia program³ to be radiometrically comparable. The images are distributed with their associated cloud, shadow and saturation masks. A pre-processing step was applied on the images with a linear interpolation on masked pixels to guarantee same size of all images of the time series. Figure 3 displays the temporal distribution of the images belonging to this SITS and Figure 5(a) depicts two geographical areas (Strasbourg (top) and Mulhouse (bottom)) extracted from the SITS. Each image crop has a dimension of 1000^2 pixels and is composed of 4 spectral bands (Nir, R, G, B) at 10 meters.

4.2 Stability summarization for land-cover analysis

In the context of our thematic study related to land-cover analysis, we chose to consider a remote sensing index, the NDVI, instead of considering all the four spectral bands. Indeed, the NDVI index is widely used in remote sensing studies to analyze land-cover from SITS since it is sensitive to the amount of vegetation. The NDVI is simply built as the multi-spectral product based on the Nir and R bands, leading to a SITS of NDVI, noted I_t^{NDVI} .

Then, according to our quantification strategy explained in Section 3.2, the SITS of NDVI is quantified leading to J_t^{NDVI} . In this context, we set empirically $k_{quantiz} = 4$. The proposed spatio-temporal features are then computed from J_t^{NDVI} . The extracted features are finally combined, leading to false color

³ <https://theia.cnes.fr/>

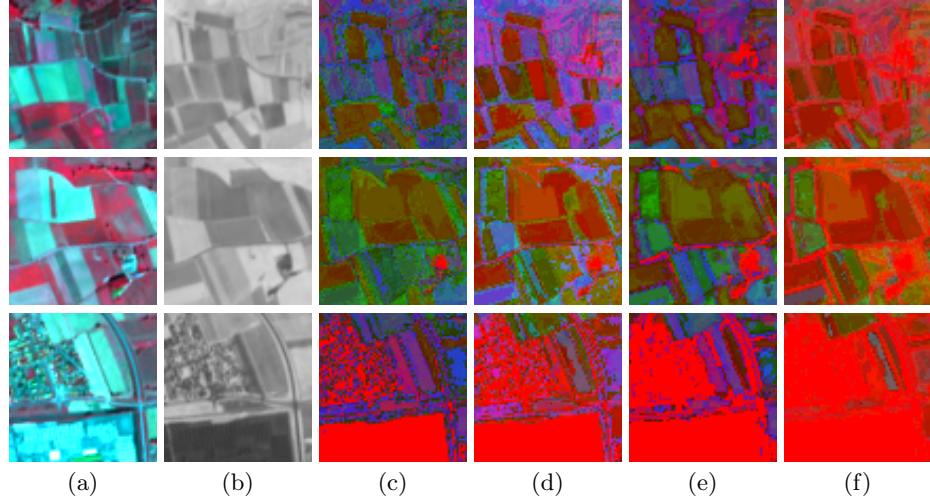


Fig. 4. Illustration of the data and results in three different geographical areas: (a) Image of the SITS at date 2017-08-26; (b) Result of the average of the SITS of NDVI $\overline{I_t^{NDVI}}$; (c),(d),(e) and (f) Obtained results of the proposed approach with the different relaxations, related respectively to TS , TS_{temp} , TS_{spatio} and $TS_{spatio-temp}$.

images, as explained in Section 3.4, which summarize the SITS of 50 images into unique images TS , TS_{temp} , TS_{spatio} and $TS_{spatio-temp}$. Concerning the spatial relaxations (Equation 6), in the predicate to relax the spatial constraints, we set the spatial neighbor of a pixel to its $N = 9$ nearest neighbors.

In a comparative study, we considered as a baseline, a summarization of the STIS that consists, for each pixel of the original SITS I_t^{NDVI} , to average (through time) the values of the temporal pixels to obtain a single scalar value of NDVI. This (naive) baseline is noted $\overline{I_t^{NDVI}}$.

Figure 4 presents the data of three different geographical regions sensed by the SITS and the obtained results: the two first lines are focused in agricultural areas and the third one is focused on a peri-urban area. The column (a) presents one original image of a SITS at date 2017-08-26, (b) presents the obtained results of the baseline summarization $\overline{I_t^{NDVI}}$ and (c),(d),(e),(f) are the results of the four proposed summarizations without / with the different relaxations presented, respectively TS , TS_{temp} , TS_{spatio} and $TS_{spatio-temp}$.

We remark that with $\overline{I_t^{NDVI}}$, the intensity of pixels in urban areas is very dark because the NDVI intensity increases in vegetation areas and decreases in other areas. In the agricultural areas, we visualize that we have about two noticeable gray levels that means that two types of grounds are present.

The obtained images with the proposed method allow a better visualization between all the thematic classes. The pixel colors are linked to the pixel evolution through time and a priori is a label of classes. The red color means that the region stays stable during a long time (high MS) with few changes (small NB)

and its stability starts early in the year (low MSS), as illustrated in the third row in Figure 4. The green color means that the region changes a lot through time with small MS and MSS . Such observations can easily allow a user, like a geographer, to better (and more conveniently) interpret the observed territories, by considering only a single image instead of an entire series, such an image capturing spatio-temporal phenomena like this may be the case for agricultural environments (seasons, different sowing or agricultural practices) and urban environments (constantly subject to various changes related to land-uses).

When comparing the results obtained with the different relaxations, we notice that with TS_{spatio} and $TS_{spatio-temp}$, roads, paths, parcel delineations are more easily visible between agricultural fields and in the urban environments. This can be surprising because when considering the traditional version of the RLE and the resulting TS image, these linear zones are not very visible, but the latter are elements normally showing a great temporal stability. It means that, when we relax the spatial domain with TS_{spatio} and $TS_{spatio-temp}$, we have the possibility to escape to the problem of image registration and we optimize the results by reducing noise (e.g. the salt and pepper). We notice that here there is only small differences between TS and TS_{temp} .

4.3 Classification of SITS with spatio-temporal stability features

In the second experiment, we involved the proposed spatio-temporal features in a binary classification task, for the analysis of urban land-cover thematic classes.

Reference data In addition to the SITS, we dispose of an imperviousness product that represents the percentage of the soil sealing. The imperviousness product is a result of a project of the European Copernicus program⁴ released by the European Environment Agency. This product defines the impermeability of materials such as the urban areas (e.g. building, commercial zone, parking) and is provided at pixel level. The spatial resolution of this product is 20 meters but we re-sampled it at 10 meters to fit with the spatial resolution of the Sentinel-2 images. Each pixel value in this reference data estimates a degree of imperviousness (0–100%). In this thematical study, we used this product to discriminate between the natural areas (0% imperviousness) and artificialized areas (imperviousness > 0%). Figure 5(b) illustrates the imperviousness reference data for two specific geographic areas. Given a SITS, the task is then to predict for each temporal pixel a binary label (i.e. artificialized vs. natural area).

Classification task Our goal is to conduct a classification in order to analyze the urban land-cover using the proposed stability features. We assume they capture in very few variables the information contained in the temporal cube. We are going to evaluate the performance of these features (MS , MSS and NB , so 3 values), alone and in complement to the raw material, i.e. the temporal

⁴ <https://land.copernicus.eu/>

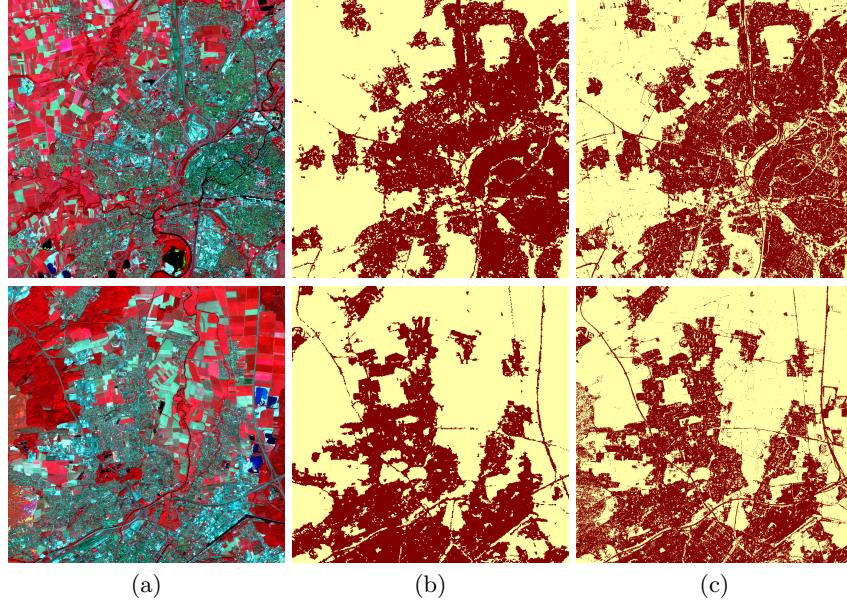


Fig. 5. Illustration of the data and results in two different geographical areas: (a) Image of the SITS at date 2017-08-26: (top) Strasbourg area and (bottom) Mulhouse area; (b) Imperviousness reference data: (red) artificialized areas and (yellow) natural areas); (c) Binary classification results from the *TS* features combined with *time series*.

pixels characterized by the NDVI values ($T + 3 = 53$ values). The influence of the different relaxation strategies introduced previously will be also studied.

In order to make a comparative study, we will also consider two sets of data, where the pixels to be classified are characterized differently: (1) the pixels are characterized by their NDVI time series, noted *time series* (50 values since $T = 50$ dates) and (2) the pixels are characterized by the mean of each time series, noted *time series* (1 value) – in the same way as in Section 4.2.

According to the considered input size for each pixel (*TS* values (≤ 3) or *time series* values (≥ 50)), we have chosen to use a Decision Tree (DT) and a Random Forest (RF) for this classification task. The decision tree used is C4.5. The Random Forest classifier contains 30 trees. Each one is constructed by splitting nodes until we get pure ones. The used criterion to do the split is the classical “Gini”. We also compared our approach with the use of a convolutional neural network, TempCNN [18]. The convolutions are applied in the temporal domain, the input of the CNN is the NDVI *time series* (50 values).

The global process is to learn the two-class pixel classifiers on one geographical area and to test the classifier on a different area. Here, the Strasbourg and Mulhouse areas are concerned. This leads to two experiments, one learning on Strasbourg area and testing on Mulhouse area (Experience 1), the other one, on learning on Mulhouse and testing on Strasbourg (Experience 2).

Table 1. Results for the two classification experiments (overall accuracy – OA).

Classifier	Features	Input size	Exp. 1	Exp. 2
RF	<i>time series</i>	50	84.56	84.02
DT	<u><i>time series</i></u>	1	66.40	67.49
CNN	TempCNN [18]	50	84.26	86.05
DT	<i>TS</i>	3	72.40	71.58
DT	<i>TS_{temp}</i>	3	76.01	79.03
DT	<i>TS_{spatio}</i>	3	83.17	85.18
DT	<i>TS_{spatio-temp}</i>	3	83.33	85.44
RF	<i>time series+TS</i>	53	84.58	84.15
RF	<i>time series+TS_{temp}</i>	53	84.56	84.07
RF	<i>time series+TS_{spatio}</i>	53	85.85	86.23
RF	<i>time series+TS_{spatio-temp}</i>	53	85.65	86.65

Results The evaluations provided in Table 1 and illustrated in Figure 6 is done by computing the overall accuracy (OA) on all pixels of the test city areas.

Table 1 reports the obtained results with the different features. First we can notice that using all the *time series* gives much better results than when only the mean value is used *time series*. We can also see that the use of the three features *TS* we have extracted from the time series provides much better results even if they are lower than when using the all *time series*. Besides we see the improvement brought by the relaxation processes proposed. The resulting spatio-temporal features enable to omit some punctual outlier values or to get rid of the registration problems we saw occurring on the different geographical limits. When processing *TS* features combined with *time series*, the accuracy scores are increased. The spatio and the spatio-temporal features give significantly higher results. We can also notice the TempCCN method [18], despite a longer learning phase, does not present better results than ours based on *time series+TS_{spatio}* or *time series+TS_{spatio-temp}* in this thematical study. Figure 5(c) illustrates the visual classification results from the two experiments, by combining the *TS* features with *time series*.

Figure 6 enables to visualize the improvement that is globally brought using the stability features. They introduce non linear processing that is difficult to model without using recurrent studies of the time series, a temporal filtering window is often a parameter depending on the series content. As a whole, we have decreased the error rate of the initial classifier by about 10%.

5 Conclusion

We propose in this article an approach to define spatio-temporal features from SITS, based on a measure of the temporal stability. The proposal is based on a compression algorithm named Run Length Encoding (RLE), applied on the image data-cube, leading to a novel image representation from which stability features can be measured. Since the notion of stability involves the study of

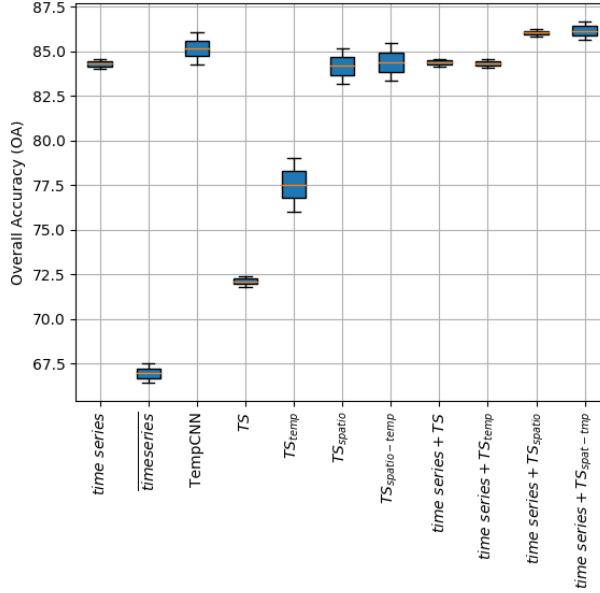


Fig. 6. Boxplot of the experiments, related to the results from Table 1.

the repetition of successive values, we also study the notion of equality and its application level. One of our contribution relies on the definition of novel approximated versions of the RLE, by relaxing both temporal or spatial constraints in the predicates involved in the equality definition. From these new versions of RLE, we then proposed the definition of spatio-temporal stability features.

All the proposed features can be used in several applications such as SITS summarizing, to make easier the interpretation of the sensed territories from the original data-cube, or the classification of spatio-temporal patterns for land-cover analysis. The preliminary results obtained from a series of 50 Sentinel-2 images highlight the interest of our approach in a remote sensing application.

We plan to pursue our work on the notion of equality, used to decide if a pixel value is stable through time. As a limit of our work, we currently compute our spatio-temporal representation and the proposed features from mono-valued data, i.e. the temporal pixels characterized by the NDVI, that is a single scalar value. When the characteristic is vectorial, for instance all the spectral bands or a combination of remote sensing indexes, it is possible to try novel definitions of the equality, leading to a more or less constrained equality definition. Such vectorial approach is more suited when the considered scalar characteristics are independent, that is not yet the case with those involved in our application.

References

1. Aminikhanghahi, S., Cook, D.J.: A survey of methods for time series change point detection. *KIS* **51**(2), 339–367 (2017)
2. Andres, L., Salas, W., Skole, D.: Fourier analysis of multi-temporal AVHRR data applied to a land cover classification. *IJRS* **15**(5), 1115–1121 (1994)
3. Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E.: The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *DMKD* **31**(3), 606–660 (2017)
4. Bruzzone, L., Prieto, D.: Automatic analysis of the difference image for unsupervised change detection. *IEEE TGRS* **38**(3), 1171–1182 (2000)
5. Chelali, M., Kurtz, C., Puissant, A., Vincent, N.: Urban land cover analysis from satellite image time series based on temporal stability. In: *JURSE*, Procs. (2019)
6. Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E.: Digital change detection methods in ecosystem monitoring: A review. *IJRS* pp. 1565–1596 (2004)
7. Di Mauro, N., Vergari, A., Basile, T.M.A., Ventola, F.G., Esposito, F.: End-to-end learning of deep spatio-temporal representations for satellite image time series classification. In: *DC@PKDD/ECML*, Procs. pp. 1–8 (2017)
8. Golomb, S.W.: Run-length encodings. *IEEE TIF* **12**, 399–401 (1966)
9. Huang, B., Lu, K., Audebert, N., Khalel, A., Tarabalka, Y., Malof, J., Boulch, A.: Large-scale semantic classification: Outcome of the first year of inria aerial image labeling benchmark. In: *IGARSS*, Procs. pp. 6947–6950 (2018)
10. Ienco, D., Gaetano, R., Dupacquier, C., Maurel, P.: Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE GRSL* **14**(10), 1685–1689 (2017)
11. Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I.: Operational high resolution land cover map production at the country scale using satellite image time series. *RS* **9**(1), 95–108 (2017)
12. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.: Deep learning for time series classification: A review. *DMKD* **33**(4), 917–963 (2019)
13. Jensen, J.R.: Urban change detection mapping using Landsat digital data. *CGIS* **8**(21), 127–147 (1981)
14. Johnson, R., Kasischke, E.: Change vector analysis: A technique for the multispectral monitoring of land cover and condition. *IJRS* **19**(16), 411–426 (1998)
15. Julea, A., Méger, N., Bolon, P., Rigotti, C., Doin, M., Lasserre, C., Trouvé, E., Lăzărescu, V.N.: Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. *IEEE TGRS* **49**(4), 1417–1430 (April 2011)
16. Millward, A.A., Piwowar, J.M., Howarth, P.J.: Time-series analysis of medium-resolution, multisensor satellite data for identifying landscape change. *PERS* **72**(6), 653–663 (2006)
17. Nielsen, A.A., Conradsen, K., Simpson, J.J.: Multivariate alteration detection (mad) and maf postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *RSE* **64**(1), 1 – 19 (1998)
18. Pelletier, C., Webb, G., Petitjean, F.: Temporal convolutional neural network for the classification of satellite image time series. *RS* **11**(5), 523–534 (2019)
19. Petitjean, F., Inglada, J., Gançarski, P.: Satellite image time series analysis under time warping. *IEEE TGRS* **50**(8), 3081–3095 (2012)
20. Petitjean, F., Inglada, J., Gançarski, P.: Satellite image time series analysis under time warping. *IEEE TGRS* **50**(8) (2012)

21. Ratanamahatana, C., Keogh, E., Bagnall, A.J., Lonardi, S.: A novel bit level time series representation with implication of similarity search and clustering. In: PAKDD, Procs. pp. 771–777 (2005)
22. Ravikumar, P., Devi, V.S.: Weighted feature-based classification of time series data. In: CIDM, Procs. pp. 222–228 (2014)
23. Russwurm, M., Korner, M.: Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In: EarthVision@CVPR, Procs. pp. 1496–1504 (2017)
24. Senf, C., Leitao, P., Pfugmacher, D., Van der Linden, S., Hostert, P.: Mapping land cover in complex mediterranean landscapes using landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *RSE* **156**, 527–536 (2015)
25. Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D.: Detecting trend and seasonal changes in satellite image time series. *RSE* **114**(1), 106–115 (2010)