Wavelet Spatio-Temporal Change Detection on multi-temporal SAR images

Rodney V. Fonseca, Rogério G. Negri, Aluísio Pinheiro, and Abdourrahmane Atto, *Senior Member, IEEE*

***Abstract*—We introduce WECS (Wavelet Energies Correlation Screening), an unsupervised sparse procedure to detect spatio-temporal change points on multi-temporal SAR images or even on sequences of very high resolution images. The procedure is based on multiscale approximation for the multi-temporal images, wavelet energy apportionment, and ultra-high dimensional correlation screening for the wavelet coefficients. We present two complimentary multiscale measures in order to detect sudden and/or cumulative changes, as well as for the case of stationary or non-stationary multi-temporal images. We show WECS performance on synthetic multi-temporal image data. We also apply the proposed method to a time series of 85 satellite images in the border region of Brazil and the French Guiana. The images were captured from November 08, 2015 to December 09 2017. (lapidaremos ao fim)**

***Index Terms*—Change detection, multi-temporal images, satellite images, wavelets.**

I. Introduction

CHANGE detection... (completar)...

A review for change detection in multi-temporal remote sensing is given by [?]. Different proposals for this purpose may be found in the literature. They vary in their motivations as well as in their applicability. Change detection in multi-temporal hyperspectral images is discussed in [?], [?], and [?]. [?] pursue change detection techniques via non-local means and principal component analysis. Compressed projection and image fusion are employed by [?]. Deep learning by slow feature analysis for change detection is the subject of [?]. [?] proposes a change detection method driven by adaptive parameter estimation.

Besides different methodological paradigms, several areas of application receive special attention. For instance, urban change detection applications via polarimetric SAR (POLSAR) images are discussed in [?]. [?] discusses land cover change detection in mountainous terrain via multi-temporal and multi-sensor remote sensing images. [?] studies multi-temporal scene classification and scene change detection. Deforestation change detection is discussed by [?].

This work was supported by FAPESP (Grants 2016/24469-6, 2018/04654-9 and 2021/01305-6) and CNPq (Grants 309230/2017-9 and 310991/2020-0).

R. Fonseca and A. Pinheiro are with the Department of Statistics, University of Campinas, Campinas 13083-859, Brazil (e-mails: ra192588@dac.unicamp.br; pinheiro@ime.unicamp.br).

A. Atto is with the LISTIC - Polytech Annecy-Chambéry, Université de Savoie, 74944 Annecy le Vieux Cedex, France (e-mail: Abdourrahmane.Atto@univ-savoie.fr).

R. G. Negri are with the Department of Environmental Engineering, São Paulo State University, São José dos Campos, Brazil (e-mail: rogerio.negri@unesp.br).

Wavelet methods present many advantages for a plethora of applications [?] thanks to wavelet capabilities in capturing multiscale/multiresolution information. Their computational efficiency and sparseness are specially relevant for large images and other high-dimensional data [?]. [?], [?], [?], [?] use different wavelet methods for change detection in satellite images.

We propose here a novel method for unsupervised spatio-temporal change detection in multi-temporal SAR images. WECS is based upon correlation screening for energy apportionment on wavelet approximations. The spatial character of the change detection is attained on pixel level. The method is fast, scalable, linearly updatable, and the resulting measures are sparse.

The motivation for our proposed method is multi-fold. We aim for a fast and accurate method. We would also like this method to be easily updatable when a new observation is captured. Finally, scalability was a concern as well. We propose a wavelet-based procedure for change detection in multi-temporal remore sensing images (WECS). It is unsupervised and built on ultra-high dimensional correlation screening [?] for the wavelet coefficients. We present two complimentary wavelet measures in order to detect sudden and/or cumulative changes, as well as for the case of stationary or non-stationary multi-temporal images. The procedure presents some advantages. It is unsupervised, fast and updatable, thus allowing for real-time change detection. Moreover, it is sparse and scalable.

The rest of the text goes as follows. Section **??** introduces the problem and presents the proposed method. We show WECS performance on synthetic multi-temporal image data in Section **??**. In Section **??** we apply the proposed method to a time series of 85 satellite images in the border region of Brazil and the French Guiana, for images captured from November 08, 2015 to December 09 2017. Section **??** concludes the paper with a discussion.

II. Wavelet Energies Correlation Screening

A. Overview

Figure **??** provides an overview of the proposed method for multitemporal image change detection.

B. Formalization

Let $\mathcal{I}(1), \ldots, \mathcal{I}(m)$ be a set of matrices representing the images of some region of interest. These images may be relative to one SAR channel or a combination of channels; this will be specified when appropriate. Our goal is twofold: to find possible points in time where some relevant changes might have

Fig. 1. Proposed method overview.

taken place at the region represented in $\mathcal{I}(m)$, $m = 1, \ldots, n$, and to find which regions are closely associated to the observed changes along time. We shall address these tasks by analyzing the bidimensional stationary discrete wavelet decomposition of $\mathcal{I}(m)$.

Stationary wavelets (also known as non-decimated or redundant wavelets) is a traditional de-noising method that can be efficiently applied to two-dimensional signals such as images [?], [?], [?]. After application of this wavelet transform to $\mathcal{I}(m)$ at some appropriate resolution level $J \geq 1$, one of its by-products is a matrix of so called approximation wavelet coefficients $\mathbf{X}(m)$, a smooth version of $\mathcal{I}(m)$ with the same dimension. The higher $J \in \{1, \ldots, \log_2(k)\}$ is, the smoother $\mathbf{X}(m)$ gets, where k is the minimum between the numbers of rows and columns of $\mathcal{I}(m)$. Many other aspects can be involved in wavelet analysis of images (e.g., different types of wavelet transform, choice of wavelet basis, thresholding of detail coefficients, etc.), but in the current work we focus on $\mathbf{X}(m)$, which provides a simple way of performing wavelet smoothing and that gives interesting results; extensions based on further aspects of wavelet transforms are straightforward.

We can then consider further apportioning the total $\mathbb{L}_2$ energy of $\{\mathbf{X}(m)\}$ as

$$\sum_{m=1}^{n} \|\mathbf{X}(m)\|^2 = n\|\bar{\mathcal{I}}\|^2 + 2n\langle \bar{\mathbf{X}} - \bar{\mathcal{I}}, \bar{\mathcal{I}} \rangle + \sum_{m=1}^{n} \|\mathbf{X}(m) - \bar{\mathcal{I}}\|^2, \tag{1}$$

where $\bar{\mathcal{I}} = n^{-1} \sum_{m=1}^{n} \mathcal{I}(m)$, $\bar{\mathbf{X}} = n^{-1} \sum_{m=1}^{n} \mathbf{X}(m)$ and $\|\mathbf{A}\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2$ for a $n \times m$ matrix $\mathbf{A} = [a_{ij}]$.

The last term in (**??**) measures deviations in time of $\mathbf{X}(m)$ compared to an *average image* $\bar{\mathcal{I}}$, what motivates us to such deviations to detect change points in time. Since each element (pixel) in $\mathbf{X}(m)$ also has a corresponding sequence of deviations in time, the relation of each local deviation to the overall measure can also be quantified, what allows us to detect changes in space. Such relation shall be computed in the present work with the Pearson correlation, what shares connections with the idea of feature screening employed in high-dimensional regression, as explained further. Other measures of change in time could be evaluated as well, such as squared differences of consecutive times $\|\mathbf{X}(m) - \mathbf{X}(m-1)\|^2$, $d(\mathbf{X}(m), \bar{\mathbf{X}})$ with $d(\cdot, \cdot)$ denoting some distance measure (e.g., Hellinger, Kullback-Leibler, etc.), among many other possibilities.

Let $X_{k,l}(m)$ and $\bar{\mathcal{I}}_{k,l}$ be the entry (k, l) of $\mathbf{X}(m)$ and $\bar{\mathcal{I}}$, respectively. Computing $D_{k,l}(m) = (X_{k,l}(m) - \bar{\mathcal{I}}_{k,l})^2$, we denote the matrix of squared mean differences as $\mathbf{D}(m) = [D_{k.l}(m)]$. We then analyze the time series given by

$$\mathbf{d}(m) = \sum_{k,l} D_{k,l}(m) = \sum_{k,l} (X_{k,l}(m) - \bar{\mathcal{I}}_{k,l})^2, \tag{2}$$

$m = 1, \ldots, n$, which measure the temporal overall variation with respect to $\bar{\mathcal{I}}$.

The time points with highest values of $\mathbf{d}(m)$ represent the images for which the most expressive changes take place, where changes here are measured through $\mathbb{L}_2$ energy. Define the $n \times p$ matrix

$$\mathbf{D} = \begin{pmatrix} vec(\mathbf{D}(1))^T \\ \vdots \\ vec(\mathbf{D}(n))^T \end{pmatrix},$$

where $vec(\mathbf{D}(m))$ is the $p \times 1$ vector of wavelet coefficients for time m and $p = \#\{k, l\}$ is total number of locations represented by $\mathbf{X}(m)$, $m = 1, \ldots, n$.

Sparsity on wavelet coefficients plays a special role here [?]. We suppose a handful of coefficients drive the changes given by $\mathbf{d}$, so that the effective dimension of $\mathbf{D}$ (number of locations where relevant changes occur), say s_d, is such that $s_d p$. This can be represented as the following linear model

$$\mathbf{d} = \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\xi} \tag{3}$$

where $\boldsymbol{\beta}$ is sparse, i.e., it has $p - s_d$ null elements, and $\boldsymbol{\xi}^{(d)}$ is some $n \times 1$ random vector of errors.

In order to identify spatio-temporal changes, we employ the idea of ultra-high dimensional correlation screening [?] as follows. For each local squared mean deviation time series, given by individual elements of $\mathbf{D}(m)$ across $m = 1, \ldots, n$, consider the absolute value of its Pearson correlation with the overall squared mean deviations, given by $\mathbf{d}$:

$$R_{k,l} = |\text{corr}\ (\mathbf{D}_{k,l}, \mathbf{d})\,|,$$

where $\mathbf{D}_{k,l} = (D_{k,l}(1), \ldots, D_{k,l}(n))^T$ is the time series of squared mean deviations of wavelet coefficients for the two-dimensional index $\{k, l\}$.

We have a matrix $\mathbf{R} = [R_{k,l}]$ of correlations of ultra-high dimension. Define the set of *important* indices for changes in images with respect to $\bar{\mathcal{I}}$ as $\mathcal{M}^* = \{(k, l)$: *changes in* $\{\mathcal{I}(m)\}_{m=1}^{m}$ *with respect to* $\bar{\mathcal{I}}$ *are affected by changes in location corresponding to index* (k, l) *in such images*$\}$. This set coincides with the non-zero vectorized one-dimensional indices for the sparse representation of $\boldsymbol{\beta}$ in (**??**). We build the empirical set of indices flagged as corresponding to change locations by

$$\mathcal{M}_\tau = \{(k, l) : |R_{k,l}| > \tau\}, \tag{4}$$

where $\tau_d > 0$ is a convenient threshold value, function of n and J.

Under some regularity conditions, the probability that $\mathcal{M}_\tau$ contains its target set satisfies

$$P(\mathcal{M}_\tau \supset \mathcal{M}^*) \to 1,$$

as $n \to \infty$ [?], where P denotes a probability measure. In other words, the empirical set $\mathcal{M}_\tau$ has high probability of detecting the correct change locations in $\mathcal{M}^*$ when the number of observations n is large.

Further geometrical motivation for our proposal is given as follows. As defined by (**??**), we expect $\mathbf{d}$ to be a vector with some few high values, say s_d, and $n - s_d$ smaller values. This segregates the multi-temporal images, since the former time points identify the images in which significant changes occur, while the latter indices identify time points with no major changes. Consider $0 < L < U$ such that the s_d highest values of $\mathbf{d}$ are larger than U, and the $n - s_d$ smallest values of $\mathbf{d}$ are smaller than L. We also take $\delta = U - L$. The indices defined by (**??**) are such that

$$\frac{< \mathbf{D}_{k,l}, \mathbf{d} >}{\|\mathbf{D}_{k,l}\|_2 \|\mathbf{d}\|_2} > \tau,$$

i.e., such that $\sum_{m=1}^{n} D_{k,l}(m)\mathbf{d}(m) > \tau \|\mathbf{D}_{k,l}\|_2 \|\mathbf{d}\|_2$. This can be rewritten as

$$\left| \sum_{m:\mathbf{d}(m)>U} D_{k,l}(m) \right| - \left| \sum_{m:\mathbf{d}(m)<L} D_{k,l}(m) \right| > \Delta,$$

for some arbitrary $\Delta 0$ (which can be a function of n and J). Thence, when we employ correlation screening we select the two-dimensional wavelet indices which have the closest empirical directions to the vector of image temporal changes. Thus we are performing a truly spatio-temporal change detection in a single procedure.

III. Experiments

Texto introdutório que explica o desenho do experimento. Aponta a primeira subseção como forma de avaliar a proposta com dados simulados, já a segunda subseção mostra uma aplicação prática envolvendo dados reais. Apenas uma explicação superficial.

A. Simulated data analysis

In this section we apply the change detection methods above on synthetic data of multi-temporal images. The synthetic multi-temporal images ($n = 80$) is obtained by repeating a sequence of four images with different types of changes plus a noise. Examples of the first four of these images are shown in Figure **??**. These images are generated from the sum of two matrices: a signal matrix with ones on entries where ellipses occur and zero elsewhere, and a noise matrix with random variables following a standard Gaussian distribution. The first image, $\mathcal{I}(1)$, presents three elongated ellipses. The second image $\mathcal{I}(2)$ has shorter and larger ellipses added. Smaller ellipses are then added to form $\mathcal{I}(3)$ and $\mathcal{I}(4)$. All the changes made that occur among subsequent images can be seen in Figure **??**(a), which displays a matrix of zeros and ones that correspond, respectively, to locations without and with changes along time. Applying WECS to these images, we obtain a matrix $\mathbf{R}$ of correlations between deviations of each $\mathcal{I}$ entry with the total squared mean deviation. An example of $\mathbf{R}$ is presented in Figure **??**(b). For some choice of threshold τ on absolute values of $\mathbf{R}$, we obtain a matrix of zeros and ones that can be compared with the total true changes displayed in Figure **??**(a).

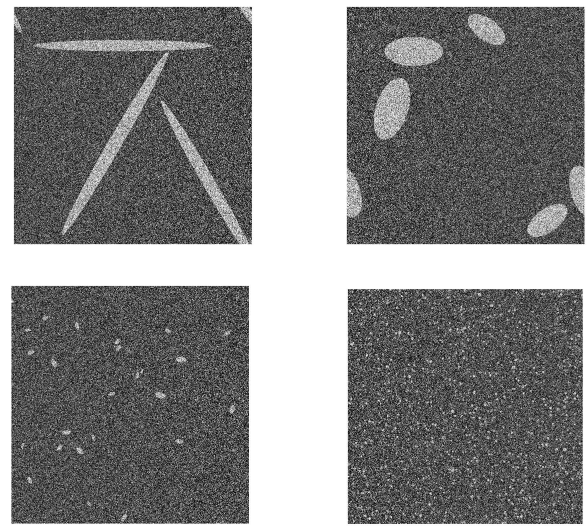


Fig. 2. Example of the first four synthetic multi-temporal images. Features and changes come as white ellipses and dots.

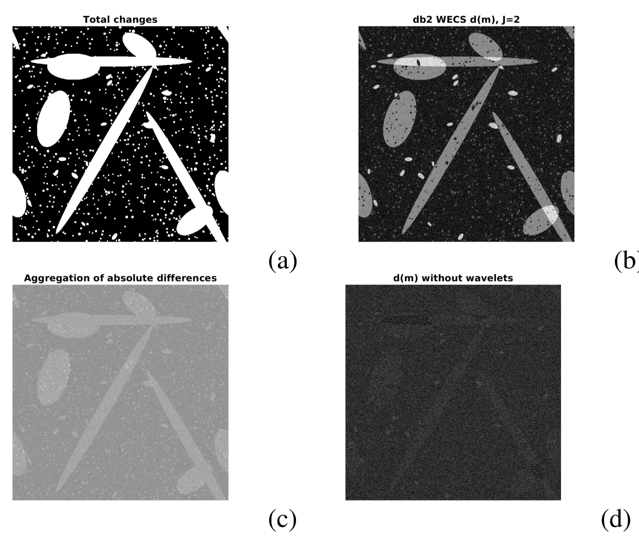


Fig. 3. Synthetic images with changing ellipses. (a) Image composed by the total changes over time. (b) Proposed db2 WECS $\mathbf{d}(m)$ with $J = 2$; (c) Standard approach. (d) $\mathbf{d}(m)$ without wavelets.

Figure **??** illustrates a comparison of different approaches to detect accumulated changes. Panel (b) shows the result of WECS using Daubechies wavelet with two null moments (db2) and $J = 2$. Panel (c) presents the result of using aggregated absolute differences, a standard approach where the accumulation of changes are measured by a matrix $\mathbf{S} = \{\mathbf{S}_{k,l}\}$ with $S_{kl} = \sum_{m=2}^{n} |\mathcal{I}_{k,l}(m) - \mathcal{I}_{k,l}(m-1)|$. Finally, in Panel (d) we can see the result if $\mathbf{d}(m)$ is performed purely on the spatial domain, using $\mathcal{I}(m)$ instead of $\mathbf{X}(m)$ in the WECS formulation.

We compute ROC curves to compare the detection performance of different methods and to verify the influence of some tuning parameters of wavelet smoothing: the resolution level J and the choice of wavelet basis. Each detection method generates a matrix of change detection measures (correlations in the case of WECS). The ROC curves present the performance of change detection by applying a threshold on these measures, in the following way:

1) Let R be the matrix of change measures. Compute the range $[r_{\min}, r_{\max}]$ of the values in R;
2) Let $(r_{(1)}, \ldots, r_{(100)})$ be equally space values between $r_{\min}$ and $r_{\max}$;
3) For each $k = 1, \ldots, n$, check how many entries are there such that $R_{i,j} > r_{(k)}$ coincide with the element (i, j) where a change really occurs on the image of total changes. Dividing this number by the total number of changes gives the true positive rate.
4) For each $k = 1, \ldots, n$, check how many entries are there such that $R_{i,j} > r_{(k)}$ do not coincide with the element (i, j) where a change really occurs. Dividing this number by the total number of entries where changes do not occur gives the false positive rate.
5) The ROC curve is the plot of true and false positive rates corresponding to each k.

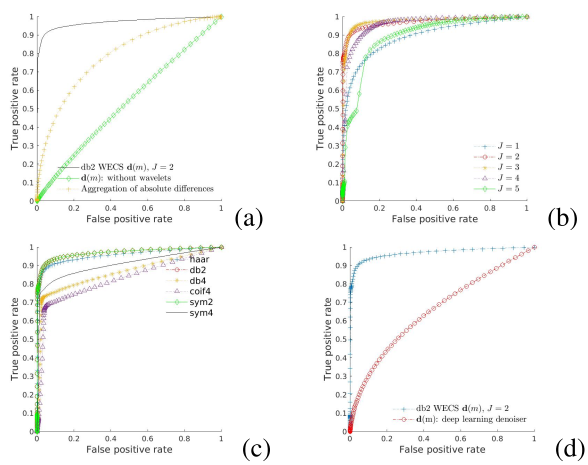


Fig. 4. ROC curves for detection of changing ellipses in synthetic images and different methods. (a) The proposed methods in black (db2 WECS $\mathbf{d}(m)$) vs two non-wavelet methods: standard log-ratio aggregation (red stars); and $\mathbf{d}(m)$ (blue). (b) db2 $\mathbf{d}(m)$ with different levels. (c) $\mathbf{d}(m)$ with different wavelet bases and J=2; (d) The proposed db2 WECS $\mathbf{d}(m)$ (red circles) and $\mathbf{d}(m)$ with deep-learning feature extraction and without wavelets (yellow line).

Figure **??** presents the different ROC curves for change detection methods applied to the synthetic data as follows. The effects of wavelet bases, level of decomposition, deep-learning feature extraction are shown on the ROC curves. We employ the following wavelet bases: Haar; Daubechies db2; Daubechies db4; Coiflets coif4; Symlets sym2 ; and Symlets sym4. Panel (c) presents the ROC curves for the proposed method under the aforementioned bases. On all instances $J = 2$ is employed. Comparing the ROC curves of all options, we can notice that Daubechies db2 and Symlets sym2 are the best choices. Panel (b) presents the ROC curves for five different levels of decomposition $J = \{1, 2, 3, 4, 5\}$ under the Daubechies db2 basis. Levels $J = 2, 3$ have a clear better performance, with a slight advantage to $J = 2$, since it is uses less levels on the decomposition. The overall performance of $J = 2$ warrants its use for the rest of the comparisons. Panel (d) shows how the proposed method performs in comparison to a deep learning feature extraction from a residual learning network [**?**]. WECS is applied with db2 wavelets and $J = 2$, whereas the other method applies WECS's idea, the only difference being that $\mathbf{X}(m)$ is replaced by another smooth version of $\mathcal{I}(m)$ that employs neural-networks trained to compete with wavelet smoothing. We can see that the ROC curves for images treated with a deep-learning method has a performance worse than that obtained with WECS. We finally have in Panel (a) the proposed WECS with db2 wavelets and $J = 2$ compared to two other non-wavelet methods. The first involves computing $\mathbf{d}(m)$ without wavelet smoothing, i.e., the squared deviations are computed using $\{\mathcal{I}(m)\}$ instead of $\{\mathbf{X}(m)\}$, and the classic method of analyzing aggregated absolute differences of $\{\mathcal{I}(m)\}$. The ROC curves in Panel (a) show that the proposed WECS outperforms both the other two methods.

B. Actual remote sensing application

We employed the proposed change detection method on a series of 84 multi-date satellite images. The images were taken on a forest region at the border of Brazil and the French Guiana (Figure **??**) from 26th December 2015 to 3rd December 2017. Each image has two polarizations (VV and VH) and 1538 by 1556 pixels. We perform a change detection wavelet analyses on the combined image by considering each observed entry as $\mathcal{I}_{k,l}(m) = (\mathrm{VV}_{k,l}(m)^2 + \mathrm{VH}_{k,l}(m)^2)^{1/2}$, where VV and VH represent the matrices of observations from VV and VH channels, respectively.

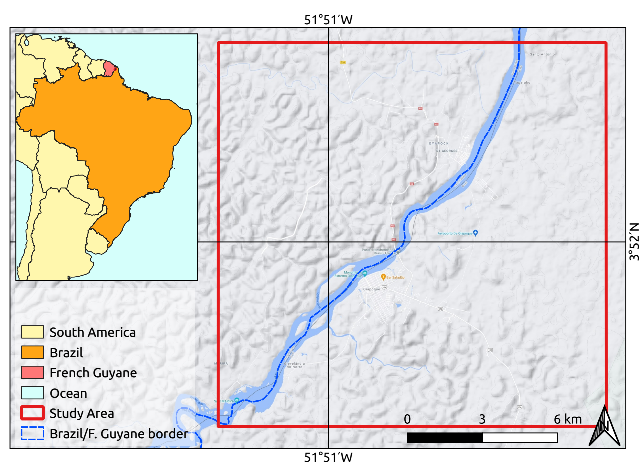
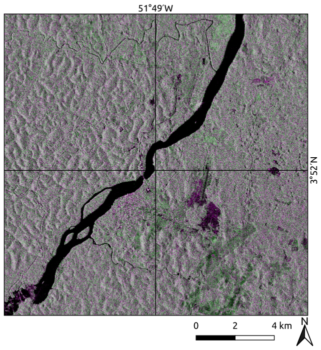
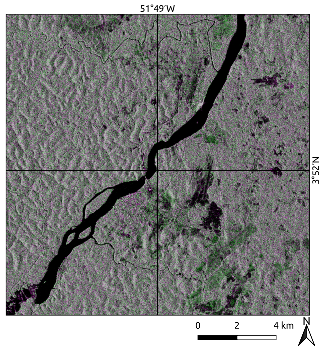


Fig. 5. Study area location. (preliminar)

A multi-resolution analysis (MRA) based on a Symlet basis with filter of length 16 (symlet 8) is built. In order to have dimensions as power of 2, the matrices $\mathcal{I}_{k,l}$ were extended to a 2048×2048 matrix with $\mathcal{I}_{k,l}$ at the center and the remaining parts being completed with mirrored values at the borders. The



(a) 26th December, 2015



(b) 3rd December, 2017

(c) Ground truth change/non-change samples

Fig. 6. Inserir caption...

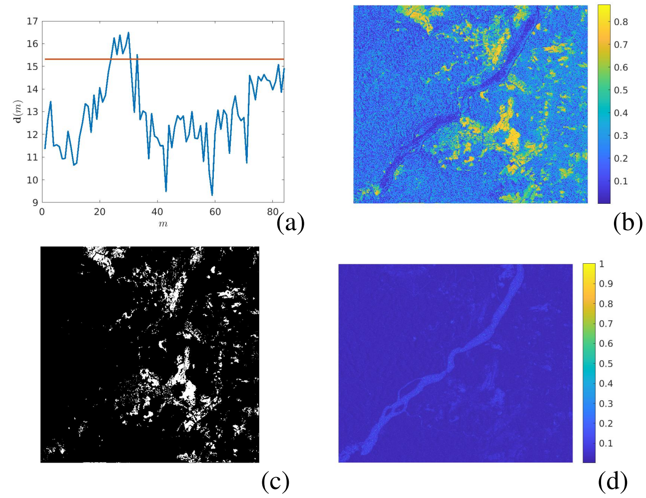


Fig. 7. Analysis of change points in time and space of the forest data. (a) Plot of $\mathbf{d}(m)$, $m = 1, \ldots, 85$ with orange horizontal line indicating two times its median absolute deviation. (b) Matrix of absolute correlations obtained with WECS. (c) Change regions detected by highlighting the $[n/\log n]$ largest correlations. (d) Change measured with aggregation of absolute differences.

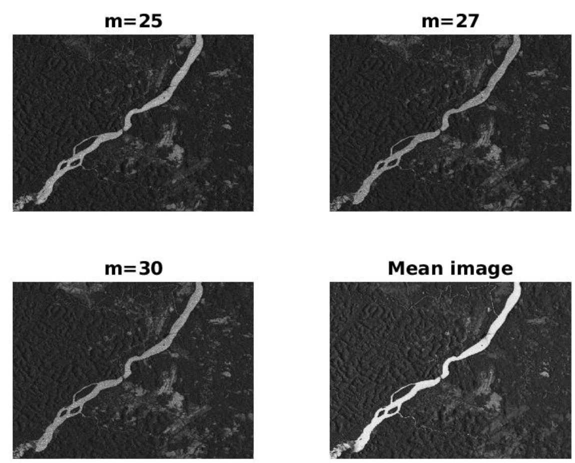


Fig. 8. Images whose time m correspond to points of $\mathbf{d}(m)$ above two times its median absolute deviation in Figure **??**a.

wavelet transform at resolution level $J = 2$ was applied to these matrices. Then, we are able to compute the squared mean differences vector $\mathbf{d}$ and the matrix of absolute correlations $\mathbf{R}$.

The time change measures computed with $\mathbf{d}$ are shown in Figure **??**a. The orange line in Figure **??**a represents the median absolute deviation of $\mathbf{d}$, which allow us to notice times that differ expressively from the others. We can notice that times $m = 25, 27, 30$ are highlighted as having expressive changes. The images corresponding to these times can be seen in Figure **??**.

The changes in space can be analyzed using the image obtained with $\mathbf{R}$, which is displayed in Figure **??**b. Taking the $[n/\log n]$ ($n = 1538 \times 1556$) largest absolute correlations as those corresponding to possible change points, we obtain a matrix of zeros and ones that is presented in Figure **??**c. The white regions in Figure **??**c (entries with value one) represent the change points, which seem to concentrate mainly on three regions: at the center, to the right of the river; at the top, on the left border of the river; and at the top left. Computing aggregated absolute differences to measure changes in space, we obtain Figure **??**d. The first two regions captured using WECS are also detected in Figure **??**d, but the former offers higher contrasts to these change regions. The comparison between the performance of these two methods can also be checked in Figure **??**, where a ROC curve is computed to check the correct detection of changing and not-changing regions. These change/nonchange regions were determined using [XXXXX]. We can notice that WECS reaches high correct detection earlier than the method that uses aggregation of differences, but both methods do not have good performances concerning the nonchange regions.

IV. Discussion

We present a novel way of detecting changes in multi-temporal satellite images, WECS. The procedure is based on wavelet energies from both the estimated individual coefficients as well as the whole mean image. It makes use of correlation screening for ultra-high dimensional data. The proposed method's performance is shown using both synthetic and real data. The proposed method is useful to detect spatio-temporal change points, which is illustrated on data analyses. The method is employed to analyze a time series of 84 images of a forest.

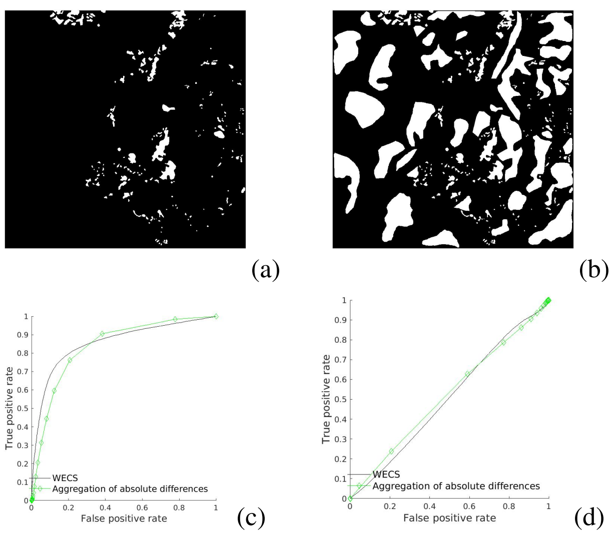


Fig. 9. Analysis of change points in time and space of the forest data. (a) Regions (in white) where changes should be detected. (a) Regions (in white) where changes are not expected to be detected. (c) ROC curve for detection of changing regions. (d) ROC curves for detection of nonchange regions.