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.**

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

I. Introduction

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.

A review for change detection in multi-temporal remote sensing is given by [1]. 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 [2], [3], and [4]. [5] pursue change detection techniques via non-local means and principal component analysis. Compressed projection and image fusion are employed by [6]. Deep learning by slow feature analysis for change detection is the subject of [7]. [8] 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 (POL-SAR) images are discussed in [9]. [10] discusses land cover change detection in mountainous terrain via multi-temporal and multi-sensor remote sensing images. [11] studies multi-temporal scene classification and scene change detection. Deforestation change detection is discussed by [12].

Wavelet methods present many advantages for a plethora of applications [13] 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 [14]. [15], [16], [17], [18] use different wavelet methods for change detection in satellite images.

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 [19] 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 II introduces the problem and presents the proposed method. We show WECS performance on synthetic multi-temporal image data in Section III. In Section IV 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 V concludes the paper with a discussion.

II. Methodology

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 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 [20], [15], [21]. After application of this wavelet transform to $\mathcal{I}(m)$ at some

This work was supported by FAPESP grants 2016/24469-6 and 2018/04654-9 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).

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 (1) 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 [22] 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 [19] as follows. For each local squared mean deviation time series, given by individual elements of $\mathbf{D}(m)$ across $m = 1, \ldots, n$, consider its Pearson correlation with the overall squared mean deviations, given by $\mathbf{d}$:

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

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 (3). 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 $n \to \infty$ [19], 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 (2), 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 $U > L > 0$ 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 (4) 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. Validation on synthetic data

In this section we apply the change detection methods above on synthetic data of multi-temporal images. The synthetic multi-temporal images ($n = 4$) are shown in Figure 1. The first image, $\mathcal{I}(1)$, presents three elongated ellipses. Changes consist of three different types of ellipses that are successively added to the original image $I(1)$. The second image, $\mathcal{I}(2)$, has new large ellipses added. Smaller ellipses are then added to form $\mathcal{I}(3)$ and $\mathcal{I}(4)$. All the changes made to image $\mathcal{I}(1)$ can be seen in Figure 2(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 2(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 2(a).

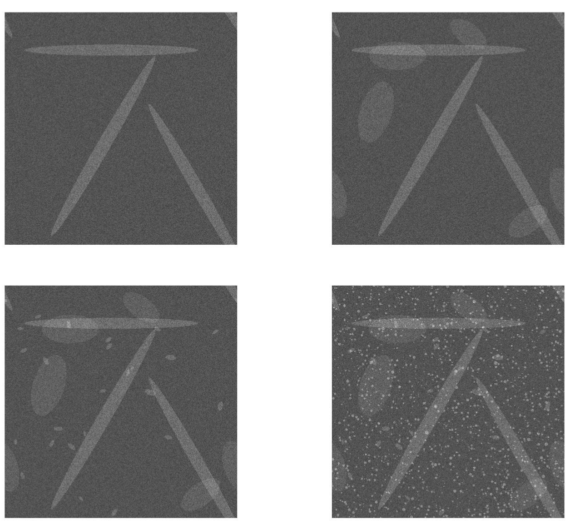

Fig. 1. Synthetic multi-temporal ($n = 4$) images. Features and changes come as ellipses and dots.

Figure 2 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 log-ratios, a standard approach where the accumulation of changes are measured by a matrix $\mathbf{S} = \{\mathbf{S}_{k,l}\}$ with $S_{il} = \sum_{m=1}^{3} |\log(\mathcal{I}_{k,l}(m+1)/\mathcal{I}_{k,l}(m))|$. 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. The spatio-temporal advantages of the proposed wavelet $\mathbf{d}(m)$ is clear in Figure 2(b), which shows that a wavelet smoothing has an expressive capacity of measuring the the changes through correlations of coefficients' deviations.

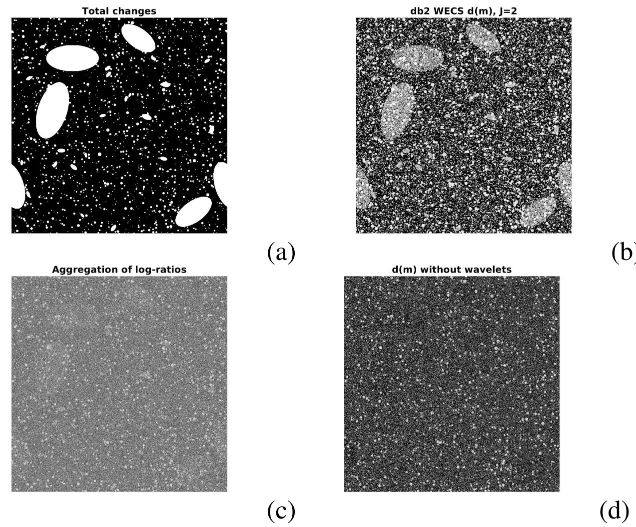

Fig. 2. Synthetic images with changing ellipses. (a) Image composed by the total changes over time. (b) Proposed db2 WECS $\mathbf{d}(m)$ with $J = 2$; (c) Standard approach. (d) $\mathbf{d}(m)$ without wavelets.

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.

Figure 3 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 $J = \{1, 2, 3, 4, 5\}$

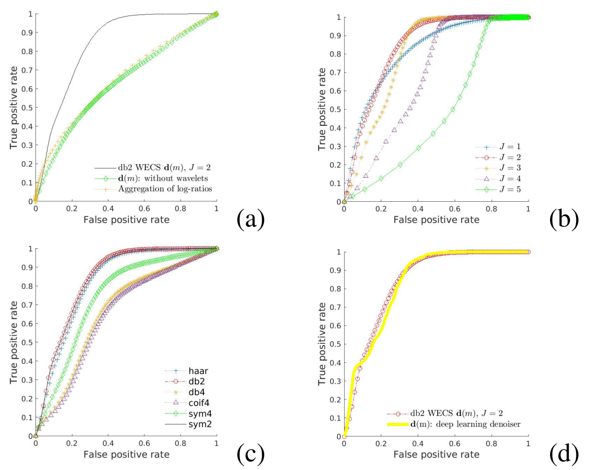

Fig. 3. ROC curves for detection of changing ellipses in synthetic images and different methods. (a) The proposed methods in black (db2 WECS $\mathbf{d}(m)$) vs two non-wavelet methods: standard log-ratio aggregation (red stars); and $\mathbf{d}(m)$ (blue). (b) db2 $\mathbf{d}(m)$ with different levels. (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).

under the Daubechies db2 basis. Levels $J = 2, 3$ have a clear better performance, with a slight advantage to $J = 2$. 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 [23]. WECS is applied with db2 wavelets and $J = 2$. We can see that the ROC curves for images treated with deep-learning methods or treated with wavelet based methods are almost identical. The WECS runs in 0.42s, while the deep-learning based $\mathbf{d}(m)$ runs in 64.68s on a notebook. The configuration of the notebook is: OS - Ubuntu 18.04.5 LTS; RAM 7.7 GB; Intel®Core™i7-8565U CPU @ 1.80GHz x 8; graphics - Mesa Intel™UHD Graphics 620 ; GNOME - 3.36.8; OS type - 64-bit. 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 log-ratios of $\{\mathcal{I}(m)\}$. The ROC curves in Panel (a) clearly show that the proposed WECS outperforms the rest.

IV. Real Data Results

We employed the proposed change detection method on a series of 85 multi-date satellite images. The images were taken on a forest region at the border of Brazil and the French Guiana from November 08, 2015 to December 09, 2017. Each image has two channels and 1200 by 1000 pixels. We perform three change detection wavelet analyses: VV Polarization Channel; VH Polarization Channel; and the Combined Image by Euclidean norm.

A multi-resolution analysis (MRA) based on a Symlet basis with filter of length 16 (symlet 8) is built. The log-images are approximated at levels $J = 1, 2, 3, 4$. Table I shows the 85 images' average energy for each approximation level. We notice that roughly 99% of the energy is recovered with $J = 1$, and more than 90% with $J = 2$. The VV channel shows better overall energy recovery than the VH channel. For $J = 4$ and $J = 3$, the Euclidean combination of the polarization channels increases the energy representation percentage. For $J = 2$, VV channel and combined channels are equivalent. VH has 4% less energy than VV for $J = 3$, and 10% less for $J = 4$.

Figures 4-6 show the series of squared deviations $\mathbf{d}(m)$ and $\mathbf{t}(m)$, for the VV channel, VH channel, and Euclidean combination, respectively. An overall feature on this data is that the VV polarization presents much higher energy than the VH. The amount of energy related to changes is ten times higher on the former compared to the latter's.

Regarding change time points, in each figure, we can notice a pattern of peaks which are common to all approximation levels. They are time points:

(a) 14, 43, 54, and 58 by the coefficients' squared deviations on the average VV image;
(b) 14, 43, 54, and 58 by the coefficients' squared deviations on the consecutive VV images;
(c) 14, 38, 41, 43, 54, 56, and 58 by the coefficients' squared deviations on the average VH image;
(d) 14, 38, 41, 43, 54-58 by the coefficients' squared deviations on the consecutive VH images;
(e) 14, 43, 54, and 58 by the coefficients' squared deviations on the average combined channels image; and
(e) 14, 43, 54, and 58 by the coefficients' squared deviations on the consecutive combined channels images.

Figures 7-9 present a timewise comparison between overall energy variations, $\mathbf{d}(m)$ and $\mathbf{t}(m)$ and their respective individual coefficients. Each figure has ten panels. Panels on the left and right deal with the first and second change detection methods respectively. A full description is give in each figure's caption. The overall conclusion from these figures is that as expected there are temporal variations between the methods regarding change detection (Panels (a)-(b)). The correlation screening detects the most relevant indices, as well as the least relevant ones (Panels (c)-(f)). These conclusions hold for analyses based upon VV, VH or combined polarizations, but the VV polarization signal is much stronger than VH's. A slight advantage is perceived for the $\mathbf{d}(m)$ method as opposed to $\mathbf{t}(m)$'s. This makes sense, since we are dealing here with data from a forest region over a long time, and seasonal changes will be perceived more easily on the first proposed method.

Figures 10-11 shows the absolute correlation images for the VV, VH and combined channels for levels $J = 2$ and $J = 3$. We can notice that high correlation coefficients for $J = 2$ are high correlation coefficients for $J = 3$ as well. Moreover, there is a clear spatial connection between polarizations and between $\mathbf{d}(m)$- and $\mathbf{t}(m)$-based analyses. On the other hand, correlations are more efficiently segregated when we move: from $J = 2$ to $J = 3$; from VH to VV polarization; or from $\mathbf{t}(m)$ to $\mathbf{d}(m)$.

V. Discussion

We present a novel way of detecting changes in multi-temporal satellite images, WECS. The procedure is based on

TABLE I

WAVELET APPROXIMATION MEAN ENERGY PERCENTAGE FOR LOG-IMAGES. FOREST REGION AT THE BORDER OF BRAZIL AND THE FRENCH GUIANA FROM NOVEMBER 08, 2015 TO DECEMBER 09, 2017. $n = 85$ MULTI-DATE SATELLITE IMAGES. EACH IMAGE HAS TWO CHANNELS AND 1200 BY 1000 PIXELS. VV POLARIZATION CHANNEL; VH POLARIZATION CHANNEL; AND THE COMBINED IMAGE BY EUCLIDEAN NORM. APPROXIMATION $J = 1, 2, 3, 4$.

MEAN APPROXIMATED ENERGY PERCENTAGE											
VV Channel				VH Channel				Combined Channels			
$J=4$	$J=3$	$J=2$	$J=1$	$J=4$	$J=3$	$J=2$	$J=1$	$J=4$	$J=3$	$J=2$	$J=1$
0.803	0.847	0.924	0.990	0.763	0.814	0.908	0.988	0.816	0.858	0.931	0.991

TABLE II

ABSOLUTE CORRELATION THRESHOLDS AND NUMBER OF SELECTED COEFFICIENTS FOR $n = 85$ MULTI-TEMPORAL IMAGES OF 1200×1000 . CORRELATION WAS COMPUTED BETWEEN APPROXIMATION COEFFICIENTS AND APPROXIMATION TOTAL ENERGY AT LEVEL $J = 2$ FOR EACH IMAGE.

Qtile	CORRELATION THRESH				Qtile	CORRELATION THRESH			
Level	VV	VH	Comb	Coeffs	Level	VV	VH	Comb	Coeffs
0.50	0.281	0.199	0.292	600000	0.99	0.647	0.582	0.668	12000
0.55	0.300	0.218	0.311	540000	0.991	0.651	0.587	0.673	10800
0.60	0.319	0.239	0.331	480000	0.992	0.656	0.594	0.678	9600
0.65	0.339	0.260	0.352	420000	0.993	0.662	0.601	0.684	8400
0.70	0.361	0.282	0.375	360000	0.994	0.669	0.608	0.690	7200
0.75	0.385	0.307	0.399	300000	0.995	0.676	0.617	0.697	6000
0.80	0.412	0.335	0.428	240000	0.996	0.685	0.626	0.705	4800
0.85	0.445	0.368	0.462	180000	0.997	0.696	0.639	0.716	3600
0.90	0.487	0.410	0.506	120000	0.998	0.710	0.654	0.729	2400
0.95	0.549	0.472	0.570	60000	0.999	0.731	0.676	0.748	1200

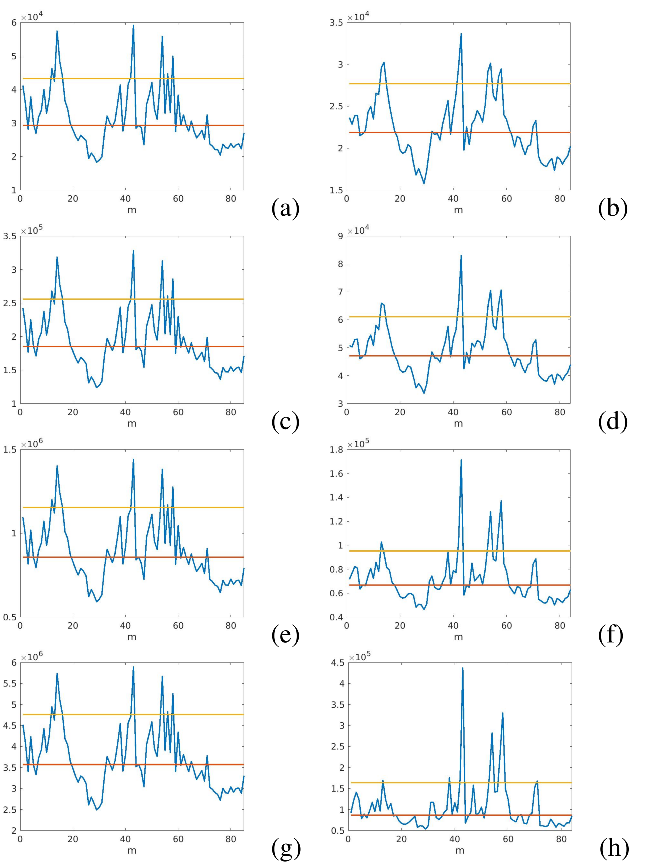

Fig. 4. VV POLARIZATION CHANNEL Series of squared deviations $\mathbf{d}(m)$ and $\mathbf{t}(m)$. The red horizontal line represents the median value and the yellow horizontal line represents their median plus two times their absolute median deviation. $\mathbf{d}(m)$ - Approximation Levels: (a) $J = 1$; (c) $J = 2$; (e) $J = 3$; (g) $J = 4$. $\mathbf{t}(m)$ - Approximation Levels: (b) $J = 1$; (d) $J = 2$; (f) $J = 3$; (h) $J = 4$.

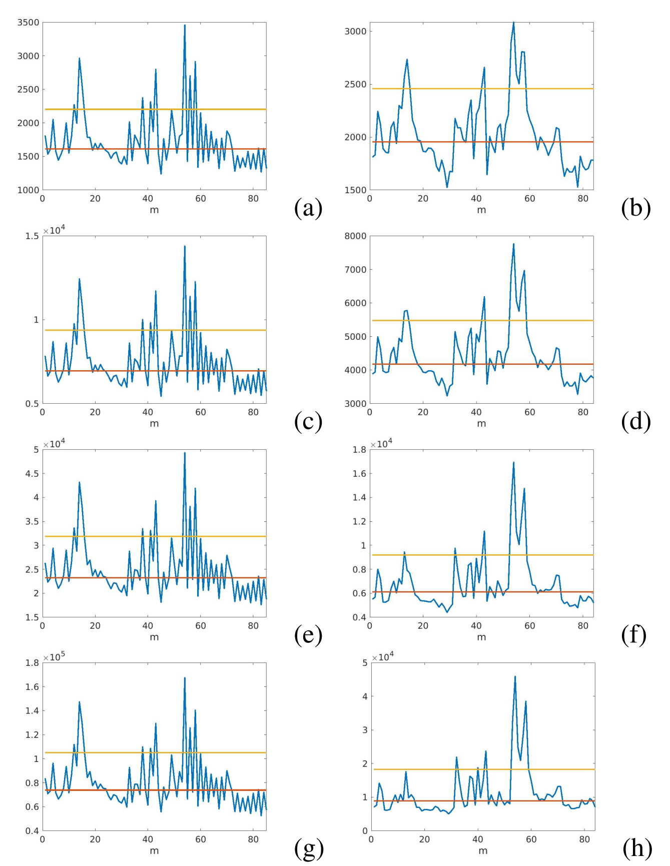

Fig. 5. VH POLARIZATION CHANNEL Series of squared deviations $\mathbf{d}(m)$ and $\mathbf{t}(m)$. The red horizontal line represents the median value and the yellow horizontal line represents their median plus two times their absolute median deviation. $\mathbf{d}(m)$ - Approximation Levels: (a) $J = 1$; (c) $J = 2$; (e) $J = 3$; (g) $J = 4$. $\mathbf{t}(m)$ - Approximation Levels: (b) $J = 1$; (d) $J = 2$; (f) $J = 3$; (h) $J = 4$.

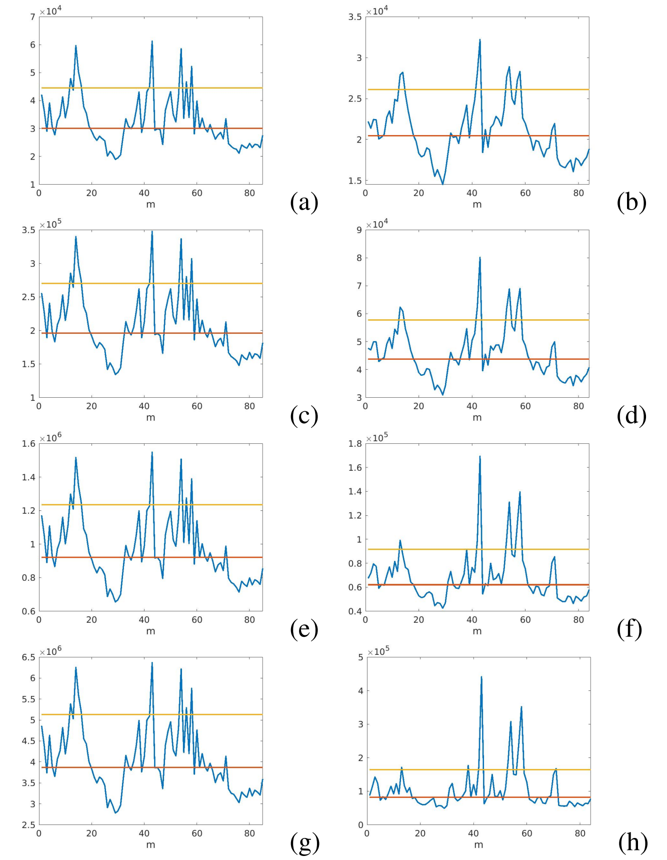

Fig. 6. COMBINED CHANNELS Series of squared deviations $\mathbf{d}(m)$ and $\mathbf{t}(m)$. The red horizontal line represents the median value and the yellow horizontal line represents their median plus two times their absolute median deviation. $\mathbf{d}(m)$ - Approximation Levels: (a) $J = 1$; (c) $J = 2$; (e) $J = 3$; (g) $J = 4$. $\mathbf{t}(m)$ - Approximation Levels: (b) $J = 1$; (d) $J = 2$; (f) $J = 3$; (h) $J = 4$.

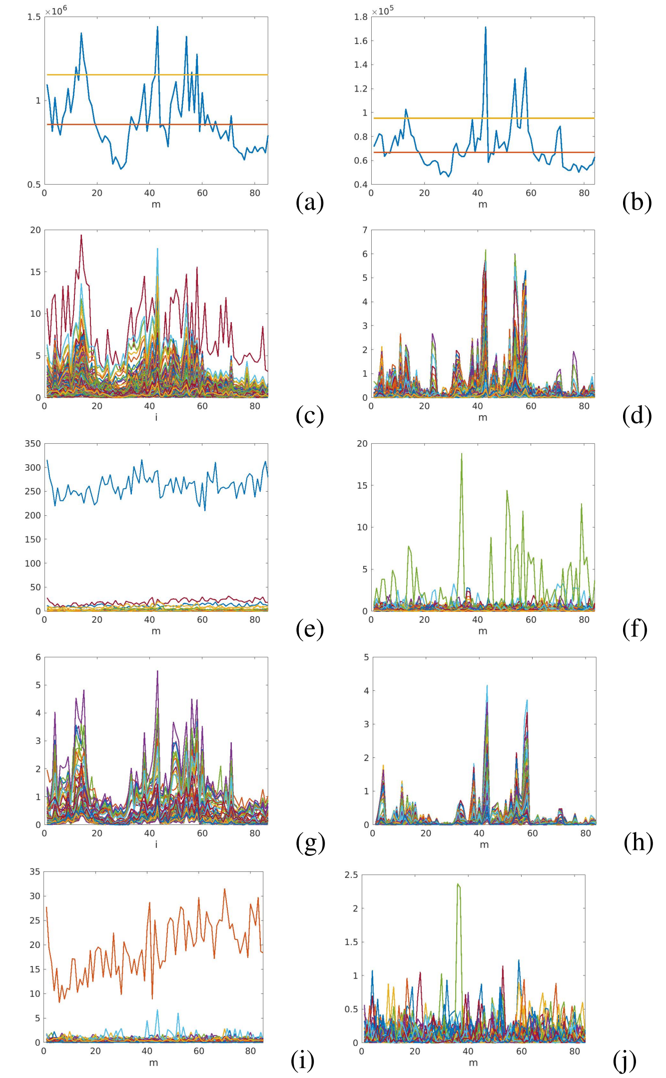

Fig. 7. VV POLARIZATION CHANNEL Series of squared mean deviations at level $J = 3$: (a) $\mathbf{d}(m)$; (b) $\mathbf{t}(m)$. Red horizontal line represents the median value and yellow line represents two absolute median deviations beyond the median. Squared approximation coefficient deviations: 0.1% highest absolute correlations - (c) $\mathbf{d}(m)$; (d) $\mathbf{t}(m)$; 0.1% smallest absolute correlations - (e) $\mathbf{d}(m)$; (f) $\mathbf{t}(m)$; 0.01% highest absolute correlations - (g) $\mathbf{d}(m)$; (h) $\mathbf{t}(m)$; 0.01% smallest absolute correlations - (i) $\mathbf{d}(m)$; (j) $\mathbf{t}(m)$.

wavelet energies from both the estimated individual coefficients as well as the whole image approximation. 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 yields spatio-temporal change points. Its performance is similar to a deep-learning feature extraction method, but the computational cost of the proposed method is 180 times smaller than the cost for the latter. Therefore, we may say that WECS may be used on images with equivalent performance for a fraction of the computational cost. Because of its reliance on wavelet representation and correlation screening, it is sparse, very fast and scalable. Finally, it is easily adapted to be updatable, so that real-time change detection is feasible even with a portable computer.

REFERENCES

[1] Y. Ban and O. Yousif, "Change detection techniques: A review," in *Multitemporal Remote Sensing*, Y. Ban, Ed. Cham: Springer, 2016, pp. 19–43.

[2] F. Bovolo and L. Bruzzone, "The time variable in data fusion: A change detection perspective," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 8–26, 2015.

[3] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 140–158, 2019.

[4] T. Matsunaga, A. Iwasaki, S. Tsuchida, K. Iwao, J. Tanii, O. Kashimura, R. Nakamura, H. Yamamoto, S. Kato, K. Obata, K. Mouri, and T. Tachikawa, "Current status of hyperspectral imager suite (HISUI) onboard International Space Station (ISS)," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Fort Worth, USA: IEEE, 2017, pp. 443–446.

[5] M. Jia and L. Wang, "Novel class-relativity non-local means with principal component analysis for multitemporal SAR image change detection," *International Journal of Remote Sensing*, vol. 39, no. 4, pp. 1068–1091, 2018.

[6] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised change detection in SAR image based on Gauss-log ratio image fusion and compressed projection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 8, pp. 3297–3317, 2014.

[7] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9976–9992, 2019.

[8] Y. Chen, Z. Ming, and M. Menenti, "Change detection algorithm for multi-temporal remote sensing images based on adaptive parameter

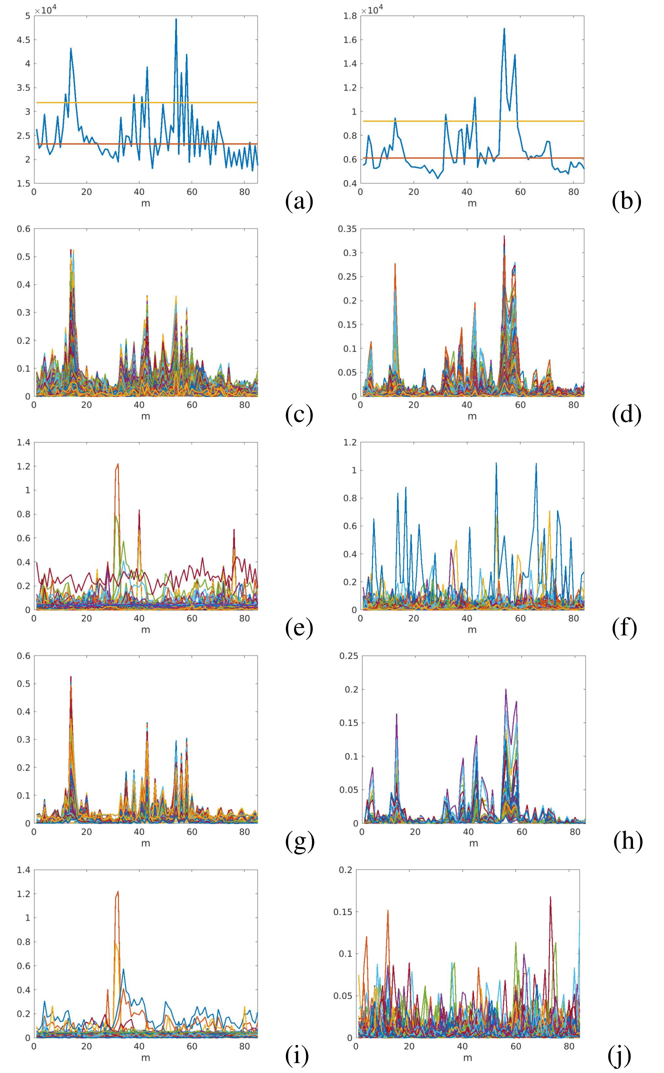

Fig. 8. VH POLARIZATION CHANNEL Series of squared mean deviations at level $J = 3$: (a) $\mathbf{d}(m)$; (b) $\mathbf{t}(m)$. Red horizontal line represents the median value and yellow line represents two absolute median deviations beyond the median. Squared approximation coefficient deviations: 0.1% highest absolute correlations - (c) $\mathbf{d}(m)$; (d) $\mathbf{t}(m)$; 0.1% smallest absolute correlations - (e) $\mathbf{d}(m)$; (f) $\mathbf{t}(m)$; 0.01% highest absolute correlations - (g) $\mathbf{d}(m)$; (h) $\mathbf{t}(m)$; 0.01% smallest absolute correlations - (i) $\mathbf{d}(m)$; (j) $\mathbf{t}(m)$.

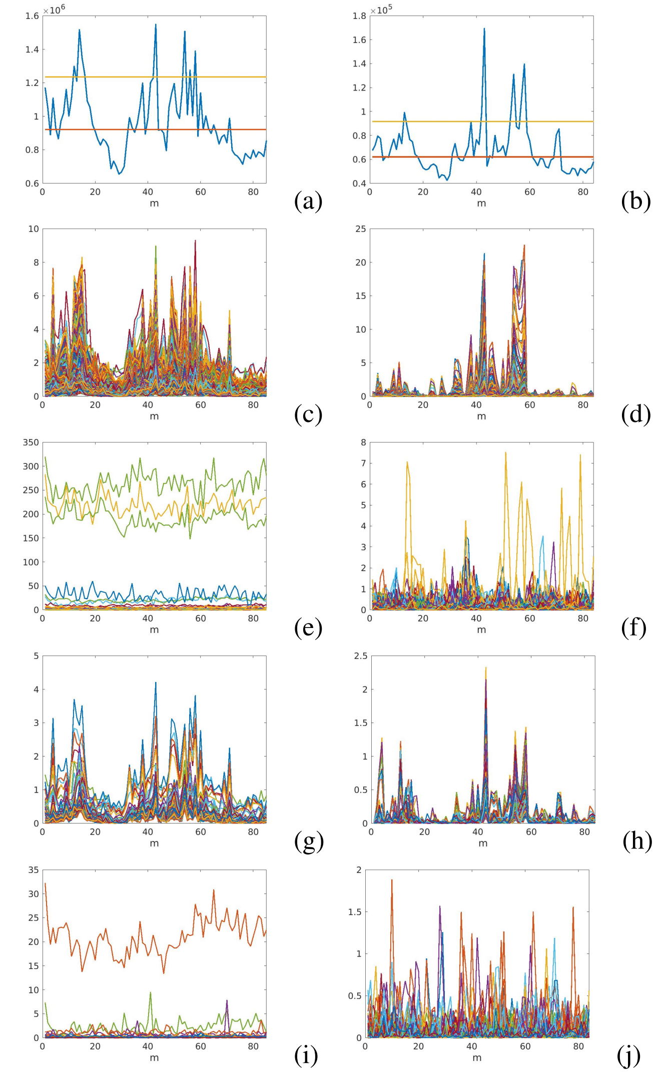

Fig. 9. COMBINED CHANNELS Series of squared mean deviations at level $J = 3$: (a) $\mathbf{d}(m)$; (b) $\mathbf{t}(m)$. Red horizontal line represents the median value and yellow line represents two absolute median deviations beyond the median. Squared approximation coefficient deviations: 0.1% highest absolute correlations - (c) $\mathbf{d}(m)$; (d) $\mathbf{t}(m)$; 0.1% smallest absolute correlations - (e) $\mathbf{d}(m)$; (f) $\mathbf{t}(m)$; 0.01% highest absolute correlations - (g) $\mathbf{d}(m)$; (h) $\mathbf{t}(m)$; 0.01% smallest absolute correlations - (i) $\mathbf{d}(m)$; (j) $\mathbf{t}(m)$.

estimation," *IEEE Access*, vol. 8, pp. 106 083–106 096, 2020.

[9] R. A. Ansari, K. M. Buddhiraju, and R. Malhotra, "Urban change detection analysis utilizing multiresolution texture features from polarimetric SAR images," *Remote Sensing Applications: Society and Environment*, vol. 20, p. 100418, 2020.

[10] F. Song, Z. Yang, X. Gao, T. Dan, Y. Yang, W. Zhao, and R. Yu, "Multiscale feature based land cover change detection in mountainous terrain using multi-temporal and multi-sensor remote sensing images," *IEEE Access*, vol. 6, pp. 77 494–77 508, 2018.

[11] L. Ru, B. Du, and C. Wu, "Multi-temporal scene classification and scene change detection with correlation based fusion," *IEEE Transactions on Image Processing*, vol. 30, pp. 1382–1394, 2021.

[12] T. L. Barreto, R. A. Rosa, C. Wimmer, J. B. Nogueira, J. Almeida, and F. A. M. Cappabianco, "Deforestation change detection using high-resolution multi-temporal X-band SAR images and supervised learning classification," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Beijing, China: IEEE, 2016, pp. 5201–5204.

[13] B. Vidakovic, *Statistical Modeling by Wavelets*. New York: John Wiley & Sons, 1999.

[14] P. A. Morettin, A. Pinheiro, and B. Vidakovic, *Wavelets in Functional Data Analysis*. Cham: Springer, 2017.

[15] A. M. Atto, E. Trouvé, Y. Berthoumieu, and G. Mercier, "Multidate divergence matrices for the analysis of SAR image time series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 1922–1938, 2012.

[16] N. Bouhlel, G. Ginolhac, E. Jolibois, and A. Atto, "Multivariate statistical modeling for multi-temporal SAR change detection using wavelet transforms," in *2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)*. Annecy, France: IEEE, 2015, pp. 1–4.

[17] T. Celik, "Multiscale change detection in multitemporal satellite images," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 820–824, 2009.

[18] S. Cui and M. Datcu, "Statistical wavelet subband modeling for multi-temporal SAR change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1095–1109, 2012.

[19] J. Fan, R. Li, C.-H. Zhang, and H. Zou, *Statistical Foundations of Data Science*. Boca Raton: CRC Press, 2020.

[20] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and statistics*, A. Antoniadis and G. Oppenheim, Eds. New

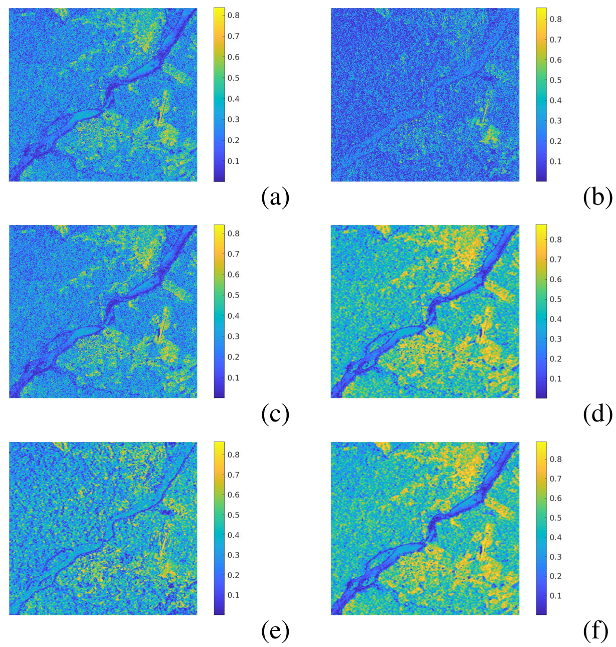

Fig. 10. Absolute correlation matrices $\mathbf{R}^{(d)}$ between mean-corrected squared coefficients at approximation levels and overall mean-corrected total energy $\{\mathbf{d}(m)\}$. $J = 2$ (a)-(c); $J = 3$ (d)-(f). $n = 85$ multi-temporal images of 1200×1000 ($J = 10$). The color bar on the right gives the magnitude of correlation at all positions. (a) VV Polarization - Channel $J = 2$. (b) VH Polarization - Channel $J = 2$. (c) Combined Channels $J = 2$. (d) VV Polarization - Channel $J = 3$. (e) VH Polarization - Channel $J = 3$. (f) Combined Channels $J = 3$

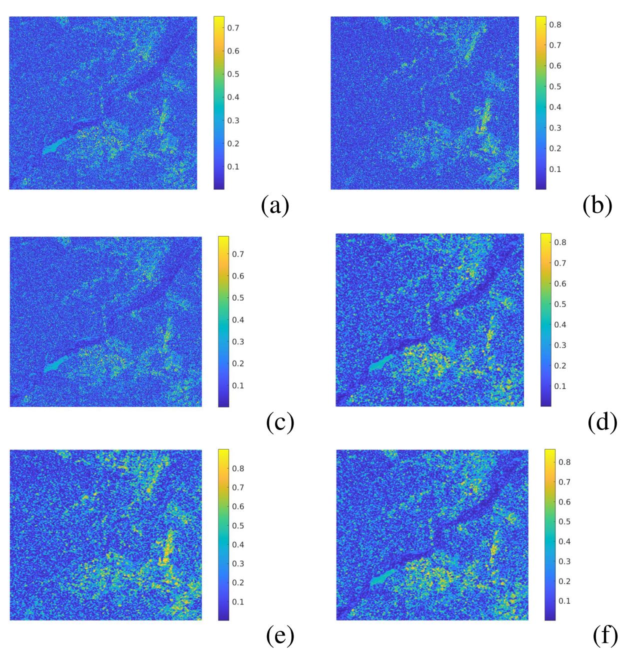

Fig. 11. Absolute correlation matrices $\mathbf{R}^{(t)}$ between squared differences of consecutive log-images' coefficients at approximation levels and their overall total energy $\{\mathbf{t}(m)\}$. Levels $J = 2$ (a)-(c); $J = 3$ (d)-(f). $n = 85$ multi-temporal images of 1200×1000. The color bar on the right gives the magnitude of correlation at all positions. (a) VV Polarization - Channel $J = 2$. (b) VH Polarization - Channel $J = 2$. (c) Combined Channels $J = 2$. (d) VV Polarization - Channel $J = 3$. (e) VH Polarization - Channel $J = 3$. (f) Combined Channels $J = 3$.

York: Springer, 1995, pp. 125–150.

[21] A. M. Atto, E. Trouvé, J.-M. Nicolas, and T. T. Lê, "Wavelet operators and multiplicative observation models—application to SAR image time-series analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6606–6624, 2016.

[22] I. M. Johnstone and D. M. Titterington, "Statistical challenges of high-dimensional data," *Philosophical Transactions of the Royal Society A*, vol. 367, pp. 4237–4253, 2009.

[23] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.