Multivariate Statistical Modeling for Multi-Temporal SAR Change Detection Using Wavelet Transforms and Integrating Subband Dependencies

Nizar Bouhlel, *Member IEEE,* Vahid Akbari, *Member IEEE*, Stéphane Méric, *Member IEEE,* and David Rousseau, *Member IEEE,*

Abstract—In this paper, we propose a new method for automatic change detection in multi-temporal fully polarimetric synthetic aperture radar (PolSAR) images based on multivariate statistical wavelet subband modeling. The proposed method allows us to take into account the correlation structure between subbands by modeling the wavelet coefficients through multivariate probability distributions. Three types of correlation are investigated: inter-scale, inter-orientation, and inter-polarization dependences. The multivariate generalized Gaussian distribution (MGGD) is used to model the interdependencies between wavelet coefficients at different orientations, scales, and polarizations. Kullback-Leibler similarity measures are computed and used to generate the change map. Simulated and real multilook PolSAR data are employed to assess the performance of the method and are compared to the multivariate Gaussian distribution (MGD) based method. We show that the information embedded in the correlation between subbands improves the accuracy of the change map, leading to better performance. Moreover, the MGGD represents better the correlations between wavelet coefficients and outperforms the MGD.

Index Terms—Change detection, Kullback-Leibler (KL) divergence, multivariate generalized Gaussian distribution (MGGD), multi-temporal polarimetric synthetic aperture radar (PolSAR) images, wavelet transform, subband correlations.

I. INTRODUCTION

Multi-temporal synthetic aperture radar image change detection becomes very important in the context of earth observation, especially for earth monitoring, damage assessment, and land cover dynamics. With the increase in the number of remotely sensed images, SAR data have become more widely available with a good source of information for the study of the evolution of large area overtimes. Change detection is a process that analyzes multitemporal remote sensing images acquired in the same geographical area for identifying changes that occurred between the considered acquisition dates [1]. The result is a generation of a change detection map in which changed areas are explicitly identified.

Several methods have been proposed to deal with the task of multi-temporal SAR change detection. Methods for change detection using single-channel SAR images have been studied and categorized into groups: The first one is based on the pixel intensity and the neighboring pixels. It includes image differencing, mean ratio/log-ratio measures [2], [3], Gauss log-ratio [4], etc. The second one is based on the evolution of the local statistics of the image associated with two dates. In this second group, the evolution is measured using the statistical information theory that provides tools to compute stochastic distances between images, such as Kullback-Leibler (KL), Bhattacharyya, and Hellinger distances [5], [6]. For instance, Inglada and Mercier [5] proposed the KL distance (KLD) between the Edgeworth series expansions for two distinct slide windows for intensity SAR images. In addition to the KLD, Nascimento *et al.* [7] derived and compared its performance with other seven stochastical distances using $\mathcal{G}^0$ distributions. Furthermore, the KLD was applied to the detection of changes in image time series, as reported by Atto *et al.* [8]. Several other information similarity measures including mutual information [9], [10], variational information, and mixed information [9], [11] have been investigated and have shown promising results due to their efficiency and simplicity.

Moreover, several methods have been also suggested using multichannel SAR data where the similarity is measured by a test statistic taking into account the covariance matrix data for unsupervised change detection. A well-known test statistic in PolSAR images is the generalized likelihood ratio test (LRT) proposed by Conradsen *et al.* [12] to measure the equality of two complex covariance matrices. Furthermore, the LRT approach was extended to the multi-temporal case [13], [14], and multi-frequency data [15]. In Frery *et al.* [16], hypothesis tests were derived based on several distance measures between relaxed complex Wishart distributions. Among these stochastic distances, the KL, Bhattacharyya, and Hellinger distances have been extensively explored. Nascimento *et al.* [17] compared between a classical change detection method based on a likelihood ratio and statistical methods based on information-theoretic measures as the KLD and entropies. In Bouhlel *et al.* [18], a heterogeneous change detection algorithm was proposed based on stochastic distance measures between two matrix-variate $\mathcal{G}_d^0$ distributions. To compute these stochastic distances, closed-form expressions for the KL, Rényi of order β, Bhattacharyya, and Hellinger distances are provided. Furthermore, a comparison between the performance of the proposed method and the homogeneous change detection method based on the scaled complex Wishart distribution is reported. Another test for change detection was put forward by Akbari *et al.* [19], [20] based on the complex-kind Hotelling-Lawley trace (HLT) test assuming the scaled complex Wishart distribution for the covariance matrix data. Recently, Bouhlel *et al.* [21] proposed a new change detection method based on the Determinant Ratio Test (DRT) statistic that outperforms the LRT and HLT detectors. However, all these statistical information measures have been so far derived in the spatial

domain.

In recent years, artificial intelligence has become a research focus in the development of new change detection methods. As one of the most important branches of artificial intelligence, deep learning [22] has achieved good results in SAR image change detection due to its powerful ability to extract and learn more robust and representative features. Deep learning-based change detection can be performed in supervised and unsupervised ways [23]. The supervised deep learning methods require a massive high-quality training samples, which are usually obtained by certain techniques (manual annotation based on the ground truth, pre-classification, etc). The existence of the prior knowledge makes the supervised methods more effective by providing high accuracy detection results. However, obtaining a large labeled training samples is time-consuming and labor-intensive [24]. To overcome the lack of labeled samples, challenging unsupervised methods are constantly emerging using complicated methodologies due to the lower amount of prior information and the complex content of such SAR images [24]. Although these techniques can overcome the lack of samples, there is still room for improvement and further studies are required to improve their performance. The various network models in deep learning used for change detection include deep belief network (DBN) [25], deep autoencoder [26], convolutional neural network (CNN) [27], [28], recurrent neural network (RNN) [29] and the generative adversarial network [30].

The methods based on statistical models are of a different nature than machine learning or deep learning. In the wavelet domain, statistical models are more stable compared with models in the spatial domain for SAR data [31]. Moreover, in the wavelet domain, it is easy to capture changes near edges or texture. Wavelet transform is a powerful modeling tool that can lead to performant texture modeling and change detection. Indeed, Shiyong *et al.* showed in [31] that texture could be better characterized in the wavelet domain than in the spatial domain, and promising results have been proposed in change detection. The wavelet transform is used to decompose the image into multiscale oriented subbands that are sensitive to horizontal, vertical, and diagonal edges. In the community of multi-temporal SAR change detection, some work on statistical wavelet subband modeling has been developed based on univariate models. For instance, wavelet coefficients have been modeled as independent Gaussian variables or as jointly Gaussian vectors [32]. Recently, non-Gaussian distributions of the wavelet subband coefficients have been frequently used. For instance, the generalized Gamma distribution (GΓD) [31], [33] and the generalized Gaussian distribution (GGD) [31], [34], [35] are suitable to model the wavelet coefficient magnitudes of each subband, since they are more peaked and heavy-tailed than the Gaussian distribution (GD) [36]. Closed-form expressions of KL divergence for both GGDs and GΓDs are given in the literature to assess the similarity [31], [37], [33].

Generally, the subbands are assumed to be independent and no dependence across wavelet orientations and scales are modeled. However, wavelet coefficients still have a dependency on each other, and they can be employed to provide better accuracy in the similarity measure. To the best of our knowledge, few works have been developed in the literature aiming to capture inter-subband dependencies through multivariate distributions. In [38], the authors studied the correlation properties of wavelet transform coefficients at different subbands and resolution levels, applying these properties to an image coding scheme based on neural networks. In [39], a joint alpha-stable sub-Gaussian distribution was used to model the dependences across wavelet orientations and scales for texture retrieval. Good results were obtained, but a computationally complex Gaussianization step was required. The common point of these approaches has been the use of the multivariate Gaussian distribution (MGD) since the similarity measure by KL is simpler to compute.

In this study, a change detection method based on wavelet coefficient magnitude modeling is proposed, allowing to take into account the correlation structure between subbands by modeling the wavelet coefficients through multivariate probability distributions. Three types of correlation are investigated in this paper: inter-scale (IS), inter-orientation (IO) and inter-polarization (IP) dependences. The first is to model the correlation between the scales of the same orientation, and the second is to model the dependences between the orientations of the same scale. Usually, a single-polarization SAR channel data is used for the wavelet transform. Since polarimetric SAR (PolSAR) gives more scattering information than single-polarization SAR channel data, it is important to extend the study to the joint modeling of the polarization channel correlation. Indeed, the third type of correlation investigated in this work is the IP dependence. The idea is to use the information embedded in the correlation between the polarization channels, since we expect that the change detection task will benefit from these additional data. This is accomplished by modeling the wavelet subband coefficient of each polarization channel by a multivariate probability distribution.

The multivariate GGD (MGGD) is a natural extension of the univariate GGD model. The MGGD is very suitable to model wavelet coefficient magnitudes and the correlation between subbands, and it is better suited to this task than the MGD. However, no closed-form expression exists for the KLD between two MGGDs except for the case of a bivariate GGD [36] and a bivariate GΓD [40]. Most recently, Bouhlel *et al.* [41] established the closed-form expression for the KLD between two MGGDs in the general case. This allows us to highlight the correlations that exist between wavelet subband coefficients. The computation of the KLD is based on the Lauricella series [42] and is practical in real time.

In this paper, we propose multivariate modeling of PolSAR images in the wavelet domain in the context of unsupervised change detection. We apply the wavelet transform to decompose the PolSAR image before and after a change into multiple scales. Then, we proceed by grouping the wavelet subband coefficients and considering them as samples of a multivariate generalized Gaussian random process. After that, we estimate the parameters of the MGGD, and we apply the closed-form expression of the KLD between two MGGDs established by Bouhlel *et al.* [41]. Finally, we use it to generate a change detection map. Furthermore, the MGGD can guarantee the

integration of the correlation between all wavelet subbands through the estimation of the scatter matrix. Through the latter, we can extract possible interdependencies between wavelet coefficients at different orientations, scales, and polarizations. As a consequence, the information residing in the correlation between subbands improves the accuracy of the change map, and enhance the change detection task. The paper is organized as follows. Section II presents the general Polarimetric multilook SAR data model. In section III, we introduce two statistical models used for the study in connection with our wavelet domain approach: the univariate GGD and the MGGD. Change detection using KLD is presented in section IV, where it is computed for the IS, IO, and IP dependences. Simulated and real PolSAR data used for evaluation and experimental settings are described in section V. Finally, the discussion and some concluding remarks close up this paper.

II. POLARIMETRIC SAR DATA MODEL

A fully-polarimetric SAR measures the amplitude and the phase of backscattered signals in four combinations of the linear receive and transmit polarizations: horizontal-horizontal (hh), horizontal-vertical (hv), vertical-horizontal (vh) and vertical-vertical (vv). These complex backscattering coefficients form the polarimetric scattering vector defined as

$$\mathbf{s} = [s_{hh}, s_{hv}, s_{vh}, s_{vv}]^T \qquad (1)$$

with s_{xy} representing the complex scattering coefficients, where x is the transmit and y is the receive polarization [43], $[.]^T$ means transposition, and $d = \dim(\mathbf{s})$ is the vector dimension. For the monostatic PolSAR system where the target reciprocity condition is satisfied [43], the polarimetric scattering vector becomes

$$\mathbf{s} = [s_{hh}, \sqrt{2}s_{hv}, s_{vv}]^T \in \mathbb{C}^d. \qquad (2)$$

Vector $\mathbf{s}$ represents a single-look polarimetric scattering vector. The multilooking of PolSAR data reduces the speckle effect characteristic of coherent imaging systems. The polarimetric multilooking operation is given by

$$\mathbf{C} = \frac{1}{K} \sum_{\ell=1}^{K} \mathbf{s}_\ell \mathbf{s}_\ell^H, \quad K \geq d \qquad (3)$$

$$= \begin{bmatrix} \langle |s_{hh}|^2 \rangle & \langle \sqrt{2}s_{hh}s_{hv}^* \rangle & \langle s_{hh}s_{vv}^* \rangle \\ \langle \sqrt{2}s_{hv}s_{hh}^* \rangle & \langle 2|s_{hv}|^2 \rangle & \langle \sqrt{2}s_{hv}s_{vv}^* \rangle \\ \langle s_{vv}s_{hh}^* \rangle & \langle \sqrt{2}s_{vv}s_{hv}^* \rangle & \langle |s_{vv}|^2 \rangle \end{bmatrix} \qquad (4)$$

where K is the number of looks, $(.)^H$ denotes the Hermitian operator, $(.)^*$ denotes the complex conjugation, $\langle . \rangle$ denotes spatial sample averaging, and $\mathbf{C} \in \mathbf{\Omega}_+ \subset \mathbb{C}^{d \times d}$ is the multilook polarimetric covariance matrix considered as a random matrix defined on the cone $\mathbf{\Omega}_+$ of the positive definite complex Hermitian matrices. Under the Gaussian assumption, the polarimetric covariance matrix follows a scaled complex Wishart distribution. However, under the polarimetric product model where the polarimetric covariance matrix is considered as the product of a positive scalar texture random variable and a random speckle matrix that follows a scaled complex Wishart distribution, the product model leads to different compound distributions given by $\mathcal{K}_d$ [44], $\mathcal{G}_d^0$ [45], [46], Kummer-$\mathcal{U}_d$ [47] and $\mathcal{G}_d$ [46] depending respectively on texture distributions: gamma (γ), inverse gamma (γ^{-1}), Fisher ($\mathcal{F}$) and generalized inverse Gaussian distributions ($\mathcal{N}^{-1}$),

Multilook PolSAR data are represented by Hermitian matrices $\mathbf{C}$ whose elements of main diagonals are positive $[\langle |s_{hh}|^2 \rangle, \langle 2|s_{hv}|^2 \rangle, \langle |s_{vv}|^2 \rangle]$ and refer to the multilook intensities. The diagonal elements of the multilook polarimetric covariance matrix are used in this study for change detection in the wavelet domain.

III. STATISTICAL MODELING IN WAVELET DOMAIN

In this section, two statistical models for wavelet subband coefficients modeling are introduced: the univariate GGD and the MGGD. The sliding window manner is applied for change detection and the discrete stationary wavelet transform (SWT) with a Daubechies filter bank is used in each window producing multiscale oriented subbands. The general framework of the proposed method is shown in Fig. 1. Since the PolSAR data are represented by Hermitian matrices, the diagonal positive elements of these matrices are used as support for the wavelet transform.

A. Univariate Generalized Gaussian distribution

The univariate GGD has been proposed in the literature to accurately model the detail subband coefficient distributions of the wavelet transform. The coefficient are highly non-Gaussian and exhibit heavy-tails [48]. The probability density function (pdf) of the GGD is given as follows

$$f_X(x|\alpha, \beta) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} \exp\left(\left| \frac{x}{\alpha} \right|^\beta \right) \qquad (5)$$

where $x \in \mathbb{R}$, the scale parameter $\alpha > 0$, the shape parameter $\beta > 0$, and $\Gamma(.)$ denotes the Gamma function. The GD and Laplace distribution are special cases of the GGD. Indeed, when $\beta = 2$, the GGD is equivalent to the GD and when $\beta = 1$, it is equivalent to the Laplace distribution.

The estimation of parameters (α, β) are realized by using the maximum-likelihood (ML) method. Considering N independent and identically distributed (*iid*) realizations x_i, $i = \{1, ..., N\}$ of a random variable X distributed according to GGD with pdf given by $f_X(x|\alpha, \beta)$. Differentiating the log-likelihood function of the samples with respect to α and β and setting the result to zero yields two equations needed to estimate the latter parameters. The shape parameter $\hat{\beta}$ is the solution of the following non-linear equation

$$1 + \frac{\psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum\limits_{i=1}^{N} |x_i|^{\hat{\beta}} \log |x_i|}{\sum\limits_{i=1}^{N} |x_i|^{\hat{\beta}}} + \frac{\log(\frac{\hat{\beta}}{N} \sum\limits_{i=1}^{N} |x_i|^{\hat{\beta}})}{\hat{\beta}} = 0 \quad (6)$$

where $\psi(.)$ is the digamma function. The Newton-Raphson method is proposed to solve it iteratively. The algorithm can be initialized with any suitable values of β. Here, the initial parameter value can be produced by the method of the
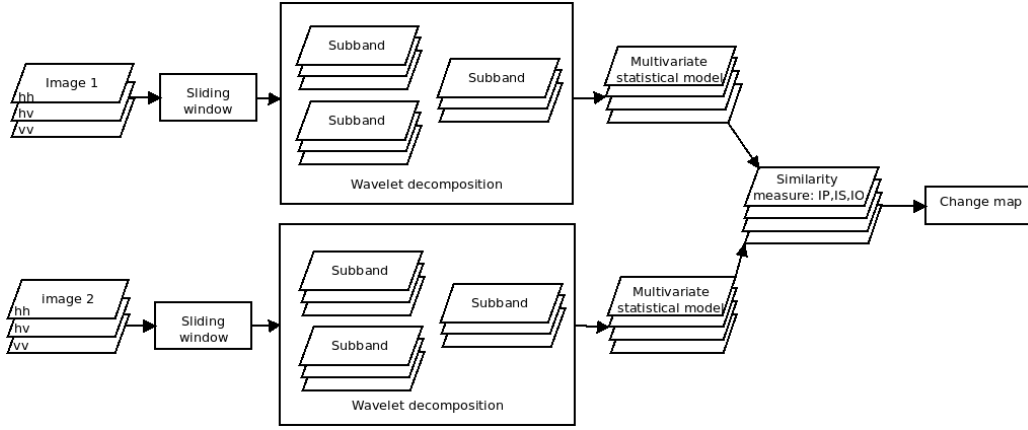
Fig. 1. Proposed method for PolSAR change detection in wavelet domain. For each polarization, the wavelet transform decomposes a sliding window into multiple subbands. The subband coefficients are combined to form different vectors for the study of IS, IO and IP dependences. Parameters $(\beta, \boldsymbol{\Sigma})$ governing the distributions of subband coefficients are estimated. KLD between pdfs of multivariate distributions is used to generate a change map.

moment. The scale parameter $\hat{\alpha}$ can be computed using the following equation

$$\hat{\alpha} = \left(\frac{\hat{\beta}}{N} \sum_{i=1}^{N} |x_i|^{\hat{\beta}} \right)^{1/\hat{\beta}}. \tag{7}$$

It was shown in [49] that the log-likelihood equations had a unique solution [34].

B. Multivariate Generalized Gaussian distribution

In order to capture inter-subband dependencies, the wavelet coefficients are modeled through multivariate probability distributions. The MGGD is a natural extension of the univariate GGD. Introduced by Kotz in [50], it is also known as the multivariate power exponential distribution [51].

Let $\mathbf{X}$ be a random vector of $\mathbb{R}^n$ which follows the MGGD, characterized by the following pdf given as follows [51]

$$f_{\mathbf{X}}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta) = \frac{\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}}\Gamma(\frac{n}{2\beta})2^{\frac{n}{2\beta}}} \frac{\beta}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^{\beta}}. \tag{8}$$

This is for any $\boldsymbol{x} \in \mathbb{R}^n$, where n is the dimensionality of the probability space, β is the shape parameter, $\boldsymbol{\mu}$ is the mean vector, and $\boldsymbol{\Sigma}$ is the so-called scatter matrix, which is a $(n \times n)$ positive definite symmetric matrix. When $\beta = 1$, the pdf corresponds to the MGD. On the other hand, when $\beta = 1/2$, the pdf corresponds to the multivariate Laplace distribution. In this study, the mean vector $\boldsymbol{\mu}$ is assumed to be zero. In [51], Gómez et al. have shown that $\mathbf{X}$ is stochastically represented by

$$\mathbf{X} = \tau \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{u} \tag{9}$$

where $\boldsymbol{u}$ is a random vector uniformly distributed on the unit sphere in the n-dimensional Euclidean space, and τ is a scalar positive random variable such that $\tau^{2\beta}$ follows a Gamma distribution with a scale parameter 2 and a shape parameter $n/(2\beta)$. It is clear that $(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{\beta}$ follows the same distribution as $\tau^{2\beta}$. In the case where $n = 1$, the MGGD is equivalent to the univariate GGD whose pdf is

given by (5) with the following relations: $\beta_{\text{GGD}} = 2\beta$ and $\alpha_{\text{GGD}} = \boldsymbol{\Sigma}^{1/2}2^{1/\beta_{\text{GGD}}}$.

The parameter estimation is a very important step for several image processing applications when the MGGD is applied. Parameters $\boldsymbol{\Sigma}$ and β can be estimated by a least square method [52] or by minimizing the χ^2 distance. The method of the moments and the ML method are also investigated [53], [54]. Pascal et al. demonstrated in [55] that the ML estimate of the MGGD parameters exists, is unique and can be calculated by a fixed point algorithm which is relatively easy to implement.

Considering $\{\boldsymbol{x}_i; i = 1, ..., N\}$ be N iid realizations of a random vector $\mathbf{X}$ distributed according the MGGD. Pascal et al. proposed the following relation $\boldsymbol{\Sigma} = m\boldsymbol{M}$ where m is a scale parameter and matrix $\boldsymbol{M}$ is normalized according to $\text{tr}(\boldsymbol{M}) = n$ where $\text{tr}(.)$ denotes the trace of the matrix. The authors observed that with this normalization constraint the algorithm convergence was significantly faster.

The ML estimator of $\boldsymbol{M}$, m and β is obtained by differentiating the log-likelihood of $(\boldsymbol{x_1}, ..., \boldsymbol{x_N})$ w.r.t to $\boldsymbol{M}$, m and β. The ML estimate $\hat{\boldsymbol{M}}$ satisfies of the following fixed point equation [55]

$$\hat{\boldsymbol{M}} = n \sum_{i=1}^{N} \frac{\boldsymbol{x_i}\boldsymbol{x_i^T}}{u_i + u_i^{1-\hat{\beta}} \sum_{j \neq i}^{N} u_j^{\hat{\beta}}} \tag{10}$$

where $u_i = \boldsymbol{x_i^T}\boldsymbol{M^{-1}}\boldsymbol{x_i}$. The shape parameter $\hat{\beta}$ is the solution of the following non-linear ML equation

$$\frac{nN}{2\sum_{i=1}^{N} u_i^{\hat{\beta}}} \sum_{i=1}^{N} u_i^{\hat{\beta}} \ln(u_i) - \frac{nN}{2\hat{\beta}} \left[\ln(2) + \psi(\frac{n}{2\hat{\beta}}) \right]$$

$$-N - \frac{nN}{2\hat{\beta}} \ln \left(\frac{\hat{\beta}}{nN} \sum_{i=1}^{N} u_i^{\hat{\beta}} \right) = 0. \tag{11}$$

This equation is solved recursively by using a Newton-Raphson procedure to estimate β for a given matrix $\boldsymbol{M}$. Once $\hat{\beta}$ is estimated, $\hat{m}$ is immediately given by

$$\hat{m} = \left(\frac{\hat{\beta}}{nN} \sum_{i=1}^{N} u_i^{\hat{\beta}} \right)^{1/\hat{\beta}}. \tag{12}$$

It is clearly seen that from (10) and (11), M and β are estimated independently of the scale parameter m. It is worth noticing that when $n = 1$, the estimation procedure becomes equivalent to the ML method for the GGD presented in subsection III-A.

IV. CHANGE DETECTION BASED ON KULLBACK-LEIBLER IN WAVELET DOMAIN

The modification of the statistics of each pixel's neighborhood allows the detection of the change between two acquisition dates. To quantify change detection between two acquisition dates, the distance KL is chosen here as a similarity measure since it is a good similarity indicator for change detection [31]. If the statistics of the two sliding windows are close, the distance KL is small, otherwise the distance is large. A critical problem in change detection using the KL is the availability of an existing closed-form expression of the KL divergence, otherwise certain approximation techniques based on the numerical evaluation of the integral of the KL divergence should be applied. The most popular method is the Monte-Carlo (MC) estimation technique [56]. It consists of approximating the KL provided that a large number of independent and identically distributed samples are provided. Nevertheless, the MC method is a too slow process to be useful in change detection, especially in the manner of a sliding window. In addition, any approximation can decrease the accuracy of change detection. Fortunately, an analytical expression of the KL divergence between two univariate zero-mean GGDs is available, which was presented by Do and Vetterli in [34]. The KL divergence expression is easy and fast to compute. For the multivariate case, the KL divergence involves computing a multidimensional integral over the data space, which is a difficult task. For this reason, closed-form expressions for the KL are difficult to find. Verdoolaege *et al.* derived an analytical expression of the KL divergence, but limited to the case of the bivariate GGD [36]. Lately, Bouhlel *et al.* [41] introduced a closed-form expression of the KL divergence and its symmetric version KLD between two zero-mean MGGDs in a general case. What follows is a presentation of the symmetric KL divergence expression for the GGD and the zero-mean MGGD.

Let X^1 and X^2 be two random variables with probability density functions f_{X^1} and f_{X^2}. The KL divergence from X^2 to X^1 is given by

$$\text{KL}(X^1||X^2) = \int \log\left(\frac{f_{X^1}(x)}{f_{X^2}(x)}\right) f_{X^1}(x)dx. \tag{13}$$

The KLD between X^1 and X^2 is defined as the symmetric KL similarity measure and is given as follows

$$d_{\text{KL}}(X^1, X^2) = \text{KL}(X^1||X^2) + \text{KL}(X^2||X^1). \tag{14}$$

For the case of univariate distribution, if the X^1 and X^2 are distributed according to a GGD with parameters (α_1, β_1) and (α_2, β_2), respectively, a closed-form expression of the symmetric version of the KL divergence has the following form

$$d_{\text{KL}}(X^1, X^2) = \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma(\frac{1+\beta_2}{\beta_1})}{\Gamma(\frac{1}{\beta_1})} + \left(\frac{\alpha_2}{\alpha_1}\right)^{\beta_1} \frac{\Gamma(\frac{1+\beta_1}{\beta_2})}{\Gamma(\frac{1}{\beta_2})} - \frac{1}{\beta_1} - \frac{1}{\beta_2}. \tag{15}$$

For the case of multivariate distributions, if $\mathbf{X}^1$ and $\mathbf{X}^2$ are two random n-vectors that follow MGGDs with joint pdfs $f_{\mathbf{X}^1}(\boldsymbol{x}|\boldsymbol{\Sigma}_1, \beta_1)$ and $f_{\mathbf{X}^2}(\boldsymbol{x}|\boldsymbol{\Sigma}_2, \beta_2)$, respectively, given by (8), then the closed form expression of the KLD between two zero-mean MGGDs $f_{\mathbf{X}^1}(\boldsymbol{x}|\boldsymbol{\Sigma}_1, \beta_1)$ and $f_{\mathbf{X}^2}(\boldsymbol{x}|\boldsymbol{\Sigma}_2, \beta_2)$ is given by

$$\text{KL}(\mathbf{X}^1||\mathbf{X}^2) = \ln\left(\frac{\beta_1|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}\Gamma(\frac{n}{2\beta_2})}{\beta_2|\boldsymbol{\Sigma}_2|^{-\frac{1}{2}}\Gamma(\frac{n}{2\beta_1})}\right) + \frac{n}{2}\left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\ln 2 -$$
$$\frac{n}{2\beta_1} + 2^{\frac{\beta_2}{\beta_1}-1}\frac{\Gamma(\frac{\beta_2}{\beta_1} + \frac{n}{2\beta_1})}{\Gamma(\frac{n}{2\beta_1})}\lambda_n^{\beta_2} F_D^{(n-1)}\left(-\beta_2, \underbrace{\frac{1}{2}, ..., \frac{1}{2}}_{n-1}; \frac{n}{2};\right.$$
$$\left. 1 - \frac{\lambda_{n-1}}{\lambda_n}, ..., 1 - \frac{\lambda_1}{\lambda_n}\right) \tag{16}$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of the real matrix $\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}$, and $F_D^{(n-1)}(.)$ represents the Lauricella D-hypergeometric function defined for $n-1$ variables [42]. The Lauricella series $F_D^{(n)}$ in a general form is given as follows (see Appendix A)

$$F_D^{(n)}(a, b_1, ..., b_n; c_n; x_1, ..., x_n) \tag{17}$$
$$= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+...+m_n}(b_1)_{m_1}...(b_n)_{m_n}}{(c)_{m_1+...+m_n}} \frac{x_1^{m_1}}{m_1!}...\frac{x_n^{m_n}}{m_n!}$$

where $|x_1|, ..., |x_n| < 1$. The Pochhammer symbol $(q)_i$ indicates the i-th rising factorial of q, i.e.

$$(q)_i = q(q+1)...(q+i-1) = \frac{\Gamma(q+i)}{\Gamma(q)} \text{ if } i = 1, 2, ... \tag{18}$$

The symmetric KL similarity measure $d_{\text{KL}}(\mathbf{X}^1, \mathbf{X}^2)$ between $\mathbf{X}^1$ and $\mathbf{X}^2$ becomes

$$d_{\text{KL}}(\mathbf{X}^1, \mathbf{X}^2) = -\frac{n}{2\beta_1} - \frac{n}{2\beta_2} + 2^{\frac{\beta_2}{\beta_1}-1}\frac{\Gamma(\frac{\beta_2}{\beta_1} + \frac{n}{2\beta_1})}{\Gamma(\frac{n}{2\beta_1})} \times$$
$$\lambda_n^{\beta_2} F_D^{(n-1)}\left(-\beta_2, \frac{1}{2}, ..., \frac{1}{2}; \frac{n}{2}; 1 - \frac{\lambda_{n-1}}{\lambda_n}, ..., 1 - \frac{\lambda_1}{\lambda_n}\right) +$$
$$2^{\frac{\beta_1}{\beta_2}-1}\frac{\Gamma(\frac{\beta_1}{\beta_2} + \frac{n}{2\beta_2})}{\Gamma(\frac{n}{2\beta_2})}\lambda_1^{-\beta_1} F_D^{(n-1)}\left(-\beta_1, \frac{1}{2}, ..., \frac{1}{2}; \frac{n}{2};\right.$$
$$\left. 1 - \frac{\lambda_1}{\lambda_n}, 1 - \frac{\lambda_1}{\lambda_{n-1}}, ..., 1 - \frac{\lambda_1}{\lambda_2}\right). \tag{19}$$

It is worth mentioning that when $\beta = 1$, the KL divergence expression corresponds to that of the MGD and is given by [41]

$$\text{KL}(\mathbf{X}^1||\mathbf{X}^2) = \frac{1}{2}\left[\ln|\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2| - n + \text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1})\right] \tag{20}$$

where

$$\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) = n\lambda_n F_D^{(n-1)}\left(-1, \frac{1}{2}, ..., \frac{1}{2}; \frac{n}{2}; 1 - \frac{\lambda_{n-1}}{\lambda_n}, ..., 1 - \frac{\lambda_1}{\lambda_n}\right) \tag{21}$$

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3215783

6

A. KLD for the inter-scale dependence

After wavelet decomposition, we have $3L$ sub-images for each polarization $\{hh, hv, vv\}$ where L is the number of scales. These sub-images are represented at this stage by $\boldsymbol{\mathcal{X}}_{\boldsymbol{p}}^{\boldsymbol{i}} = \{\boldsymbol{H}_{\boldsymbol{p}}^{\boldsymbol{i}}, \boldsymbol{V}_{\boldsymbol{p}}^{\boldsymbol{i}}, \boldsymbol{D}_{\boldsymbol{p}}^{\boldsymbol{i}}\}$ where $i \in \{1, 2\}$ is the index number for image 1 and image 2, and $p \in \{hh, hv, vv\}$ designs the type of polarization. The following $\boldsymbol{H}_{\boldsymbol{p}}^{\boldsymbol{i}} = [H_{1,p}^{i}, H_{2,p}^{i}, ..., H_{L,p}^{i}]^{T}$, $\boldsymbol{V}_{\boldsymbol{p}}^{\boldsymbol{i}} = [V_{1,p}^{i}, V_{2,p}^{i}, ..., V_{L,p}^{i}]^{T}$ and $\boldsymbol{D}_{\boldsymbol{p}}^{\boldsymbol{i}} = [D_{1,p}^{i}, D_{2,p}^{i}, ..., D_{L,p}^{i}]^{T}$ are L-dimensional random vectors representing the sub-image horizontal, vertical and diagonal wavelet detail coefficients, respectively, and are distributed according to the MGGD. Let $\boldsymbol{\Sigma}_{\boldsymbol{H_p}}^{\boldsymbol{i}}$, $\boldsymbol{\Sigma}_{\boldsymbol{V_p}}^{\boldsymbol{i}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{D_p}}^{\boldsymbol{i}}$ be the $L \times L$ scatter matrix of the previous random vectors. To quantify the dependency between scales, these scatter matrices from sub-images and the shape parameters are estimated using the procedure presented in subsection III-B. The symmetric KLD of two sliding windows is defined as the sum of similarity measures of each L-vector of the same orientation over all polarizations:

$$\sum_{p \in \{hh,hv,vv\}} d_{\mathrm{KL}}(\boldsymbol{\mathcal{X}}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{X}}_{\boldsymbol{p}}^{\boldsymbol{2}}) = \sum_{p} d_{\mathrm{KL}}(\boldsymbol{H}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{H}_{\boldsymbol{p}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{V}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{V}_{\boldsymbol{p}}^{\boldsymbol{2}})$$
$$+ d_{\mathrm{KL}}(\boldsymbol{D}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{D}_{\boldsymbol{p}}^{\boldsymbol{2}}) \quad (22)$$

where $d_{\mathrm{KL}}(\boldsymbol{H}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{H}_{\boldsymbol{p}}^{\boldsymbol{2}})$, $d_{\mathrm{KL}}(\boldsymbol{D}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{D}_{\boldsymbol{p}}^{\boldsymbol{2}})$, and $d_{\mathrm{KL}}(\boldsymbol{V}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{V}_{\boldsymbol{p}}^{\boldsymbol{2}})$ are computed using (19). In case of the subbands are assumed independent, the total similarity of two blocks or two sliding windows are defined as the sum of similarity measures of each subband [31]

$$\sum_{p} d_{\mathrm{KL}}(\boldsymbol{\mathcal{X}}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{X}}_{\boldsymbol{p}}^{\boldsymbol{2}}) = \sum_{p} \sum_{j=1}^{L} d_{\mathrm{KL}}(H_{j,p}^{1}, H_{j,p}^{2}) + d_{\mathrm{KL}}(D_{j,p}^{1}, D_{j,p}^{2}) + d_{\mathrm{KL}}(V_{j,p}^{1}, V_{j,p}^{2}) \quad (23)$$

where $d_{\mathrm{KL}}(H_{j,p}^{1}, H_{j,p}^{2})$, $d_{\mathrm{KL}}(D_{j,p}^{1}, D_{j,p}^{2})$ and $d_{\mathrm{KL}}(V_{j,p}^{1}, V_{j,p}^{2})$ are calculated using (15).

B. KLD for the inter-orientation dependence

To measure the IO dependence, we constitute a new set of sub-images $\boldsymbol{\mathcal{Y}}_{\boldsymbol{p}}^{\boldsymbol{i}} = \{Y_{1,\boldsymbol{p}}^{\boldsymbol{i}}, Y_{2,\boldsymbol{p}}^{\boldsymbol{i}}, ..., Y_{L,\boldsymbol{p}}^{\boldsymbol{i}}\}$ where $i \in \{1, 2\}$, $p \in \{hh, hv, vv\}$ and $\boldsymbol{Y}_{\boldsymbol{j},\boldsymbol{p}}^{\boldsymbol{i}} = [H_{j,p}^{i}, V_{j,p}^{i}, D_{j,p}^{i}]^{T}$ are 3-dimensional random vectors composed of variables representing the horizontal, vertical and diagonal wavelet detail coefficients, respectively, at the scale $j \in \{1, 2, ..., L\}$ and $\boldsymbol{Y}_{\boldsymbol{j},\boldsymbol{p}}^{\boldsymbol{i}}$ are distributed according to MGGD. The symmetric KLD of two sliding windows is defined as the sum of similarity measures of each 3-vector of the same scale over all polarizations.

$$\sum_{p} d_{\mathrm{KL}}(\boldsymbol{\mathcal{Y}}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{Y}}_{\boldsymbol{p}}^{\boldsymbol{2}}) = \sum_{p} \sum_{j=1}^{L} d_{\mathrm{KL}}(\boldsymbol{Y}_{\boldsymbol{j},\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{Y}_{\boldsymbol{j},\boldsymbol{p}}^{\boldsymbol{2}}). \quad (24)$$

If the orientations of the same scale are assumed independent, the total similarity of two sliding windows are defined also by equation (23).

C. KLD for the inter-polarization dependence

To measure the dependency between polarizations with the same scale and orientation, we compose new sets of sub-images $\boldsymbol{\mathcal{H}}^{\boldsymbol{i}} = \{\mathbb{H}_{\boldsymbol{1}}^{\boldsymbol{i}}, \mathbb{H}_{\boldsymbol{2}}^{\boldsymbol{i}}, ..., \mathbb{H}_{\boldsymbol{L}}^{\boldsymbol{i}}\}$, $\boldsymbol{\mathcal{V}}^{\boldsymbol{i}} = \{\mathbb{V}_{\boldsymbol{1}}^{\boldsymbol{i}}, \mathbb{V}_{\boldsymbol{2}}^{\boldsymbol{i}}, ..., \mathbb{V}_{\boldsymbol{L}}^{\boldsymbol{i}}\}$ and $\boldsymbol{\mathcal{D}}^{\boldsymbol{i}} = \{\mathbb{D}_{\boldsymbol{1}}^{\boldsymbol{i}}, \mathbb{D}_{\boldsymbol{2}}^{\boldsymbol{i}}, ..., \mathbb{D}_{\boldsymbol{L}}^{\boldsymbol{i}}\}$ where $i \in \{1, 2\}$, $\mathbb{H}_{\boldsymbol{j}}^{\boldsymbol{i}} =$ $[H_{j,hh}^{i}, H_{j,hv}^{i}, H_{j,vv}^{i}]^{T}$ at the scale $j \in \{1, 2, ..., L\}$, $\mathbb{V}_{\boldsymbol{j}}^{\boldsymbol{i}} = [V_{j,hh}^{i}, V_{j,hv}^{i}, V_{j,vv}^{i}]^{T}$ and $\mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{i}} = [D_{j,hh}^{i}, D_{j,hv}^{i}, D_{j,vv}^{i}]^{T}$ are 3-dimensional random vectors representing the sub-image horizontal, vertical and diagonal wavelet detail coefficients, respectively, and are distributed according to MGGD. The symmetric KLD of two sliding windows is defined as the sum of similarity measures of each 3-vector of the same scale with different polarizations

$$d_{\mathrm{KL}}(\boldsymbol{\mathcal{H}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{H}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{\mathcal{V}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{V}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{\mathcal{D}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{D}}^{\boldsymbol{2}})$$
$$= \sum_{j=1}^{L} d_{\mathrm{KL}}(\mathbb{H}_{\boldsymbol{j}}^{\boldsymbol{1}}, \mathbb{H}_{\boldsymbol{j}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\mathbb{V}_{\boldsymbol{j}}^{\boldsymbol{1}}, \mathbb{V}_{\boldsymbol{j}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{1}}, \mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{2}}) \quad (25)$$

where $d_{\mathrm{KL}}(\mathbb{H}_{\boldsymbol{j}}^{\boldsymbol{1}}, \mathbb{H}_{\boldsymbol{j}}^{\boldsymbol{2}})$, $d_{\mathrm{KL}}(\mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{1}}, \mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{2}})$ and $d_{\mathrm{KL}}(\mathbb{V}_{\boldsymbol{j}}^{\boldsymbol{1}}, \mathbb{V}_{\boldsymbol{j}}^{\boldsymbol{2}})$ are computed using (19). In case of the subbands are assumed independent, the total similarity of two blocks or two sliding windows is defined as

$$d_{\mathrm{KL}}(\boldsymbol{\mathcal{H}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{H}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{\mathcal{V}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{V}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{\mathcal{D}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{D}}^{\boldsymbol{2}}) \quad (26)$$
$$= \sum_{j=1}^{L} \sum_{p} d_{\mathrm{KL}}(H_{j,p}^{1}, H_{j,p}^{2}) + d_{\mathrm{KL}}(D_{j,p}^{1}, D_{j,p}^{2}) + d_{\mathrm{KL}}(V_{j,p}^{1}, V_{j,p}^{2})$$

where $d_{\mathrm{KL}}(H_{j,p}^{1}, H_{j,p}^{2})$, $d_{\mathrm{KL}}(D_{j,p}^{1}, D_{j,p}^{2})$ and $d_{\mathrm{KL}}(V_{j,p}^{1}, V_{j,p}^{2})$ are calculated using (15).

D. Total KLD

The total similarity of two blocks or two sliding windows is defined as the sum of the similarity measures of equations (22), (24) and (25) and is given by

$$d_{\mathrm{KL}} = d_{\mathrm{KL}}(\boldsymbol{\mathcal{H}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{H}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{\mathcal{V}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{V}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{\mathcal{D}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{D}}^{\boldsymbol{2}})$$
$$+ \sum_{p} \left(d_{\mathrm{KL}}(\boldsymbol{\mathcal{X}}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{X}}_{\boldsymbol{p}}^{\boldsymbol{2}}) + d_{\mathrm{KL}}(\boldsymbol{\mathcal{Y}}_{\boldsymbol{p}}^{\boldsymbol{1}}, \boldsymbol{\mathcal{Y}}_{\boldsymbol{p}}^{\boldsymbol{2}}) \right). \quad (27)$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the proposed method based on multivariate statistical wavelet subband modeling, such as the MGGD, is evaluated by comparing it with the wavelet domain method based on the MGD, the univariate GGD, and the GD. Our approach performed in wavelet domain is also compared to three methods performed in spatial domain. The first one detects changes by comparing the local statistics of the two images (before and after change). The local statistics are based on scaled Wishart distribution, where the parameters are estimated from the pixels of the sliding windows. The similarity measure of the local statistics is accomplished using the KLD. This method is called by KLD-scaled Wishart. The closed form expression of the KLD between two scaled Wishart distributions has been presented in [18]. The two remaining methods based on spatial domain are HLT and LRT. The receiver operating characteristic (ROC) curve is performed here independently of any thresholding algorithm. The ROC curve is the evolution of the true positive rate (TPR) according to the false positive rate (FPR) [31]. Moreover, to generate a binary change detection map, the optimal threshold is chosen as the nearest point to (0.0, 1.0) in the ROC curve, as depicted in Fig. 4 by a black arrow. When thresholding is applied, a binary change detection map is produced. To determine the threshold in a completely unsupervised manner, several

thresholding methods have been proposed in the literature, and none of them is perfect. To name a few of them: the constant false alarm rate (CFAR) algorithm [57], Otsu's method [58], the Kittler and Illingworth algorithm [59], and Liu's method [60]. The best performance is also obtained by applying optimal thresholding which consists of selecting the nearest point to (0,1) in the ROC curve [6], [31]. Based on the optimal threshold, the TPR and the FPR are obtained. Furthermore, the area under the ROC curve (AUC) is calculated as a performance measure, and it is a good indicator of changes. The larger the AUC, the better the performance is. In the following, our approach is tested with both simulated and real multilook PolSAR data.

A. Evaluation on simulated PolSAR images

The simulated quad-pol data contain two eight-look PolSAR images of 200×200 pixels and have three polarimetric channels. These data are each composed of five different regions. These regions follow the $\mathcal{G}_d^0(\Sigma, K, \lambda)$ distribution where Σ is the covariance matrix, K is the number of looks and λ is the texture parameter. This distribution has been extensively used by the community for its flexibility and has been reported its excellent performance showing the suitability for describing different kinds of regions: homogeneous, heterogeneous and extremely heterogeneous clutters [61], [62], [63], [64]. Other known distributions like $\mathcal{K}_d$ [44] and $\mathcal{G}_d$ distributions [46] can also be used, but they are not included in this study. The texture parameters λ are chosen to cover homogeneous and heterogeneous texture. For an urban area (area 1) which is a high-texture region and extremely heterogeneous, the shape parameter of the matrix-variate $\mathcal{G}_d^0$ distribution is less than 2. For a forest area which is a heterogeneous area with moderate texture, the value of λ is given by $2 < \lambda < 8$ (areas 3, 4, and 5). For a sea region which is a homogeneous area with low texture, the shape parameter tends towards infinity, leading to a realization of a scaled Wishart distribution. Here, it corresponds to region 2 with $\lambda = \infty$. The details of the distribution parameters (λ, Σ) for each region are given in Table II. We want to draw the reader's attention to the fact that regions 1 and 2 have the same speckle covariance matrix, but different texture parameter values. This configuration allows us to see if our approach is able to distinguish between the two regions. Fig. 2 shows the simulated images before and after changes, as well as the binary truth change map. For performance evaluation in the wavelet domain, different scales and window sizes are used. Each sliding window with size $\{12, 16, 20, 24, 28\}$ is decomposed into $L = \{1, 2, 3\}$ scales using discrete stationary wavelet transform (SWT) with Daubechies wavelet filter bank of order $\{1, 2, 3, 4, 5\}$. Three types of correlations are presented: IS, IO and IP dependences. In this regard, distances $\sum_p d_{\mathrm{KL}}(\mathcal{X}_p^1, \mathcal{X}_p^2)$, $\sum_p d_{\mathrm{KL}}(\mathcal{Y}_p^1, \mathcal{Y}_p^2)$ and $d_{\mathrm{KL}}(\mathcal{H}^1, \mathcal{H}^2) + d_{\mathrm{KL}}(\mathcal{V}^1, \mathcal{V}^2) + d_{\mathrm{KL}}(\mathcal{D}^1, \mathcal{D}^2)$ are computed with the MGGD and the MGD and compared with the distances corresponding to the case where the subbands are assumed independent with the univariate GGD and GD.

1) Goodness-of-fit: For the applicability of our approach, we need to verify how good is the goodness-of-fit (GOF) of the MGGD to model the subbands and make sure that the wavelet coefficients are well represented by MGGD. In general, the GOF of a statistical model describes how well it fits a set of samples. Concretely, we use the test strategy of Gomez *et al.* presented in [51] to verify how the samples adjusted to a MGGD. In what follows, we describe the proposal steps:

First, we test the multivariate normality of the n-dimensional vector composed by wavelet coefficient subbands instead of testing the normality of each component of the vector as suggested by Gomez. The p-value of the test is computed and the hypothesis of normality is rejected if the p-value is small at the default 5% significance level. In the literature, a vast number of methods for testing multivariate normality are proposed. Among these methods, a consistent and invariant test proposed by Henze and Zirkler [65] is found to have good power properties, particularly for small sample sizes [66]. They showed that the proposed test statistic which is invariant under linear transformations of the data has a lognormal asymptotic distribution. In this study, we choose to use Henze and Zirkler multivariate normality test.

Secondly, we test the hypothesis of the MGGD on the basis of the stochastic representation. For this purpose, the test is subdivided into three hypotheses [51]:

(i) The variable $R^{2\beta} = (\mathbf{X}^T \hat{\Sigma}^{-1} \mathbf{X})^\beta$ has a Gamma distribution $\Gamma(n/2\beta, 2)$,

(ii) The variable $u^{(n)} = \hat{\Sigma}^{-1/2} \mathbf{X} / ||\hat{\Sigma}^{-1/2} \mathbf{X}||$ is uniformly distributed on the unit sphere in $\mathbb{R}^n$,

(iii) R and $u^{(n)}$ are independent.

For the first two hypotheses, the Kolmogorov-Smirnov (KS) test is applied and the p-values are computed. A smaller value of KS indicates that the hypothetical model fits better the empirical distribution. A small p-value of the test indicates strong incompatibilities of the data with the employed distribution hypothesis. For the last hypothesis (iii), the p-values of the Kendall and the Spearman tests are computed. Table I shows the results of these tests carried out on a sample of size 100×20 corresponding to a region selected in an arbitrary manner among the simulated data for the different methods. The results correspond to the calculation of the mean p-value and its standard deviation over the entire sample carried out for the normality test, and the three previous hypotheses (i), (ii) and (iii). As we can see, the p-values of the multivariate normality are small at the default 5% significance level. Accordingly, the hypothesis of normality should be rejected. For hypothesis (i), the mean p-value show the acceptance of the hypothesis on $(\mathbf{X}^T \hat{\Sigma}^{-1} \mathbf{X})^\beta$ that follows a Gamma distribution. The mean p-value are greater than 5% showing the acceptance of the hypothesis (ii): $\hat{\Sigma}^{-1/2} \mathbf{X} / ||\hat{\Sigma}^{-1/2} \mathbf{X}||$ is uniformly distributed on the unit sphere in $\mathbb{R}^n$. Finally, the hypothesis (iii) on the independence of R and $u^{(n)}$ is accepted with a p-value greater than 5%. At the light of these results, there are no reasons to reject the hypothesis that the vectors have MGGDs.

2) Change detection map: Fig. 3 illustrates different logarithm of KLD and the corresponding binary change map, where white color is a change and black color is a no-change. The configuration used for this example corresponds to a slid-

Method	Vectors	Normality test p-value	Hypothesis (i) p-value	Hypothesis (ii) p-value	Hypothesis (iii)	
					p-value (Spearman)	p-value (Kendall)
MGGD-IS	$\boldsymbol{H}_{hh}^1 = [H_{1,hh}^1, H_{2,hh}^1, H_{3,hh}^1]^T$	$12.62\ 10^{-4} \pm 95.90\ 10^{-4}$	0.484 ± 0.277	0.514 ± 0.270	0.580 ± 0.258	0.573 ± 0.259
	$\boldsymbol{V}_{hh}^1 = [V_{1,hh}^1, V_{2,hh}^1, V_{3,hh}^1]^T$	$40.09\ 10^{-4} \pm 3.05\ 10^{-2}$	0.509 ± 0.279	0.453 ± 0.281	0.616 ± 0.256	0.610 ± 0.258
	$\boldsymbol{D}_{hh}^1 = [D_{1,hh}^1, D_{2,hh}^1, D_{3,hh}^1]^T$	$80.18\ 10^{-4} \pm 3.72\ 10^{-2}$	0.586 ± 0.272	0.521 ± 0.274	0.684 ± 0.201	0.683 ± 0.201
	$\boldsymbol{H}_{hv}^1 = [H_{1,hv}^1, H_{2,hv}^1, H_{3,hv}^1]^T$	$22.09\ 10^{-4} \pm 1.92\ 10^{-2}$	0.458 ± 0.281	0.523 ± 0.272	0.503 ± 0.276	0.498 ± 0.277
	$\boldsymbol{V}_{hv}^1 = [V_{1,hv}^1, V_{2,hv}^1, V_{3,hv}^1]^T$	$19.3\ 10^{-4} \pm 1.26\ 10^{-2}$	0.496 ± 0.287	0.501 ± 0.278	0.621 ± 0.235	0.620 ± 0.235
	$\boldsymbol{D}_{hv}^1 = [D_{1,hv}^1, D_{2,hv}^1, D_{3,hv}^1]^T$	$31.96\ 10^{-4} \pm 1.96\ 10^{-2}$	0.587 ± 0.272	0.477 ± 0.278	0.607 ± 0.251	0.607 ± 0.252
	$\boldsymbol{H}_{vv}^1 = [H_{1,vv}^1, H_{2,vv}^1, H_{3,vv}^1]^T$	$5.55\ 10^{-4} \pm 0.47\ 10^{-2}$	0.389 ± 0.290	0.408 ± 0.264	0.463 ± 0.269	0.456 ± 0.271
	$\boldsymbol{V}_{vv}^1 = [V_{1,vv}^1, V_{2,vv}^1, V_{3,vv}^1]^T$	$7.48\ 10^{-4} \pm 0.94\ 10^{-2}$	0.513 ± 0.274	0.491 ± 0.282	0.500 ± 0.270	0.498 ± 0.271
	$\boldsymbol{D}_{vv}^1 = [D_{1,vv}^1, D_{2,vv}^1, D_{3,vv}^1]^T$	$58.40\ 10^{-4} \pm 3.23\ 10^{-2}$	0.543 ± 0.280	0.563 ± 0.278	0.628 ± 0.239	0.624 ± 0.241
MGGD-IO	$\boldsymbol{Y}_{1,hh}^1 = [H_{1,hh}^1, V_{1,hh}^1, D_{1,hh}^1]^T$	$3.66\ 10^{-17} \pm 1.15\ 10^{-15}$	0.341 ± 0.249	0.514 ± 0.261	0.563 ± 0.272	0.567 ± 0.269
	$\boldsymbol{Y}_{2,hh}^1 = [H_{2,hh}^1, V_{2,hh}^1, D_{2,hh}^1]^T$	$41.93\ 10^{-4} \pm 3.43\ 10^{-2}$	0.414 ± 0.269	0.535 ± 0.276	0.455 ± 0.306	0.453 ± 0.305
	$\boldsymbol{Y}_{3,hh}^1 = [H_{3,hh}^1, V_{3,hh}^1, D_{3,hh}^1]^T$	$63.06\ 10^{-4} \pm 3.77\ 10^{-2}$	0.518 ± 0.274	0.479 ± 0.281	0.378 ± 0.322	0.377 ± 0.322
	$\boldsymbol{Y}_{1,hv}^1 = [H_{1,hv}^1, V_{1,hv}^1, D_{1,hv}^1]^T$	$2.77\ 10^{-19} \pm 1.02\ 10^{-17}$	0.412 ± 0.275	0.431 ± 0.279	0.388 ± 0.290	0.388 ± 0.290
	$\boldsymbol{Y}_{2,hv}^1 = [H_{2,hv}^1, V_{2,hv}^1, D_{2,hv}^1]^T$	$32.37\ 10^{-4} \pm 2.18\ 10^{-2}$	0.443 ± 0.294	0.549 ± 0.276	0.383 ± 0.311	0.387 ± 0.311
	$\boldsymbol{Y}_{3,hv}^1 = [H_{3,hv}^1, V_{3,hv}^1, D_{3,hv}^1]^T$	$1.98\ 10^{-2} \pm 7.20\ 10^{-2}$	0.527 ± 0.272	0.443 ± 0.279	0.359 ± 0.324	0.358 ± 0.325
	$\boldsymbol{Y}_{1,vv}^1 = [H_{1,vv}^1, V_{1,vv}^1, D_{1,vv}^1]^T$	$2.63\ 10^{-9} \pm 7.70\ 10^{-8}$	0.348 ± 0.242	0.454 ± 0.257	0.633 ± 0.245	0.627 ± 0.246
	$\boldsymbol{Y}_{2,vv}^1 = [H_{2,vv}^1, V_{2,vv}^1, D_{2,vv}^1]^T$	$13.54\ 10^{-4} \pm 76.30\ 10^{-4}$	0.400 ± 0.283	0.504 ± 0.275	0.503 ± 0.294	0.503 ± 0.292
	$\boldsymbol{Y}_{3,vv}^1 = [H_{3,vv}^1, V_{3,vv}^1, D_{3,vv}^1]^T$	$1.58\ 10^{-2} \pm 6.23\ 10^{-2}$	0.502 ± 0.276	0.431 ± 0.273	0.456 ± 0.355	0.456 ± 0.355
MGGD-IP	$\mathbb{H}_1^1 = [H_{1,hh}^1, H_{1,hv}^1, H_{1,vv}^1]^T$	$4.07\ 10^{-6} \pm 4.32\ 10^{-5}$	0.417 ± 0.284	0.546 ± 0.275	0.562 ± 0.282	0.563 ± 0.282
	$\mathbb{H}_2^1 = [H_{2,hh}^1, H_{2,hv}^1, H_{2,vv}^1]^T$	$4.82\ 10^{-2} \pm 12.15\ 10^{-2}$	0.560 ± 0.274	0.516 ± 0.277	0.526 ± 0.293	0.523 ± 0.293
	$\mathbb{H}_3^1 = [H_{3,hh}^1, H_{3,hv}^1, H_{3,vv}^1]^T$	$0.29\ 10^{-2} \pm 2.33\ 10^{-2}$	0.549 ± 0.269	0.432 ± 0.285	0.379 ± 0.321	0.379 ± 0.321
	$\mathbb{V}_1^1 = [V_{1,hh}^1, V_{1,hv}^1, V_{1,vv}^1]^T$	$72.46\ 10^{-4} \pm 4.67\ 10^{-2}$	0.495 ± 0.270	0.562 ± 0.275	0.669 ± 0.245	0.666 ± 0.247
	$\mathbb{V}_2^1 = [V_{2,hh}^1, V_{2,hv}^1, V_{2,vv}^1]^T$	$6.47\ 10^{-2} \pm 13.05\ 10^{-2}$	0.517 ± 0.271	0.561 ± 0.266	0.579 ± 0.304	0.580 ± 0.303
	$\mathbb{V}_3^1 = [V_{3,hh}^1, V_{3,hv}^1, V_{3,vv}^1]^T$	$38.84\ 10^{-4} \pm 2.54\ 10^{-2}$	0.526 ± 0.281	0.481 ± 0.276	0.476 ± 0.344	0.475 ± 0.345
	$\mathbb{D}_1^1 = [D_{1,hh}^1, D_{1,hv}^1, D_{1,vv}^1]^T$	$12.47\ 10^{-4} \pm 1.02\ 10^{-2}$	0.530 ± 0.271	0.548 ± 0.275	0.732 ± 0.226	0.731 ± 0.227
	$\mathbb{D}_2^1 = [D_{2,hh}^1, D_{2,hv}^1, D_{2,vv}^1]^T$	$7.23\ 10^{-2} \pm 9.94\ 10^{-2}$	0.545 ± 0.274	0.572 ± 0.270	0.689 ± 0.256	0.689 ± 0.255
	$\mathbb{D}_3^1 = [D_{3,hh}^1, D_{3,hv}^1, D_{3,vv}^1]^T$	$6.89\ 10^{-2} \pm 8.59\ 10^{-2}$	0.517 ± 0.274	0.570 ± 0.269	0.625 ± 0.295	0.625 ± 0.295

TABLE I

GOODNESS-OF-FIT OF MGGD TO MODEL WAVELET COEFFICIENTS SUBBANDS : MULTIVARIATE NORMALITY TEST OF COMPOSED VECTOR, GAMMA DISTRIBUTION FIT TEST OF $R^{2\beta} = (\mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{\beta}$, UNIFORM DISTRIBUTION ON UNIT SPHERE FIT TEST OF $u^{(n)} = \hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{X} / ||\hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{X}||$ AND INDEPENDENCE TEST BETWEEN R AND $u^{(n)}$.
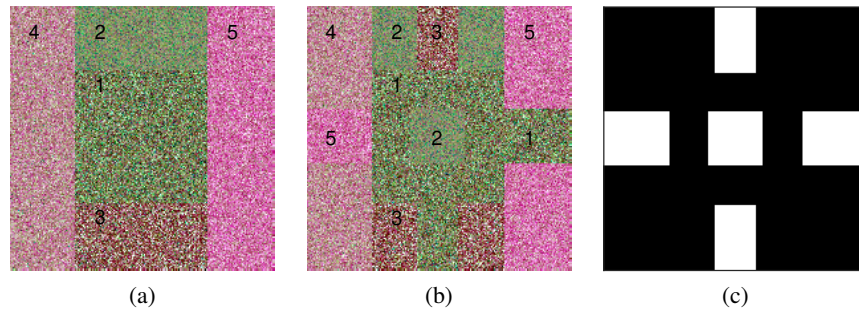

Fig. 2. Pauli decomposition of simulated 4-look quad-pol PolSAR data (a) before change, (b) after change, and (c) binary truth change map.

TABLE II

TEXTURE PARAMETERS AND COVARIANCE MATRIX FOR EACH REGION

Area	λ	$\boldsymbol{\Sigma} = \{\Sigma_{11}, \Sigma_{22}, \Sigma_{33}, \Sigma_{12}, \Sigma_{13}, \Sigma_{23}\}$
1	4	0.08, 0.1, 0.05, 0.03i, 0.02i, 0.01
2	∞	0.08, 0.1, 0.05, 0.03i, 0.02i, 0.01
3	2	0.14, 0.1, 0.05, -0.03i, -0.02i, 0.01
4	8	0.2, 0.1, 0.05, 0.03i, 0.05i, 0.01
5	6	0.3, 0.08, 0.042, 0.05+0.03i, 0.02i, 0.01-0.03i

ing window size: 16×16, a wavelet transform scale $L = 1$, and a Daubechies wavelet filter: 'db1'. The distances are computed for both univariate (GD, GGD) and multivariate distributions (MGD and MGGD) where IO, IP and IS dependencies are used. For the purpose of comparison, our approach performed in wavelet domain is compared to KLD-scaled Wishart, HLT and LRT methods performed in spatial domain. While Fig. 3.(j) and (j') show the logarithm of KLD-scaled Wishart and the binary change map, respectively, Fig. 3.(k-l) and (k'-l') show the HLT and LRT results, and their corresponding binary change maps.

The binary change map is generated by using the optimal threshold. The TPRs, the false alarm rates (FARs) and the overall error rate are presented respectively in Table III based on the selected optimal thresholds. From this Table, we can see that the detection rate and the accuracy of MGGD is quite high compared to MGD and GGD. In addition, the MGGD-IP method outperforms the MGGD-IS and MGGD-IO methods. On the other hand, the GD is worse than all other methods

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3215783
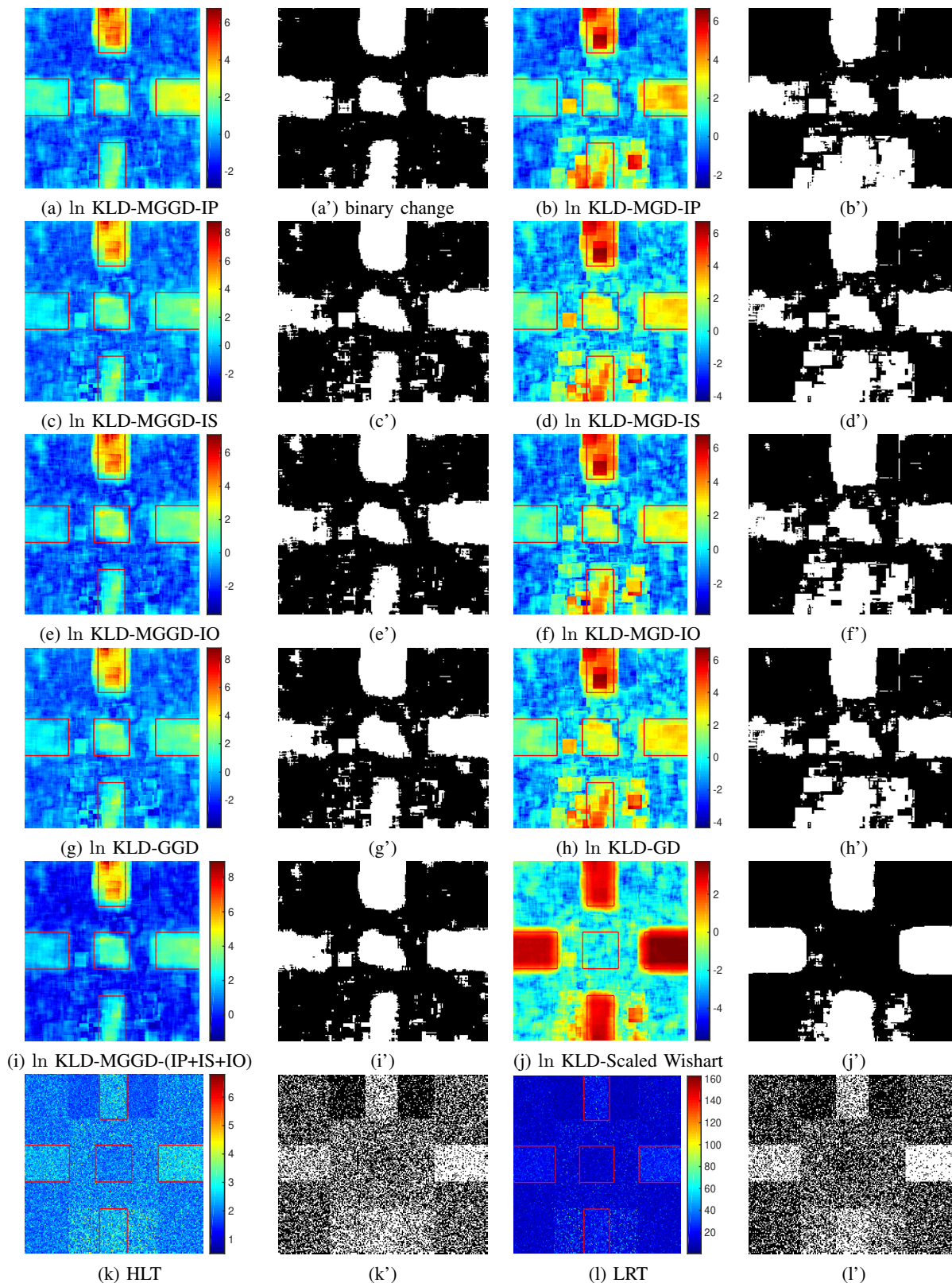
9

Fig. 3. Detection results for simulated PolSAR data. (a) and (b) ln KLD for IP dependence with MGGD and MGD, and their corresponding binary change maps (a')-(b') where white color is change and black color is no-change. (c)-(d) ln KLD for IS dependence with MGGD and MGD, and their binary maps (c')-(d'). (e)-(f) ln KLD IO dependence with MGGD and MGD, and (e')-(f') their binary maps. (g)-(h) ln KLD for univariate GGD and GD, and their binary change maps (g')-(h'). (i) and (i') ln KLD-MGGD (IP+IS+IO) dependence with MGGD and binary map. (j) and (j') ln KLD with scaled Wishart and corresponding binary map. (k)-(l) results of the HLT and LRT methods. Parameters used: sliding window size = 16, wavelet transform scale $L = 1$, Daubechies wavelet filter: db1. Optimal threshold selected corresponds to the closest point to (0.0, 1.0) in the ROC curve.

based on wavelet domain. For the sake of comparison, FAR, TPR and overall error of KLD-scaled Wishart, HLT and LRT are also reported. From the last row in Table III, we can see that the values of these parameters when optimal threshold is applied are less efficient than the KLD-MGGD-IP. Going back to Fig. 3, the following conclusions are drawn. First, the GD and the MGD give us the worst detection results for the methods based on the wavelet domain as it is shown in Fig. 3(h) and Fig. 3(b)-(d)-(f) and their corresponding binary change map in Fig. 3(h') and Fig. 3(b')-(d')-(f'). Second, the KLD using the MGGD provides better performance than that based on the GGD as more areas are detected. Accordingly, taking into account the correlation structure between subbands improves in this example the performance and the accuracy of the change map. Third, the comparison between KLD based on the MGGD-IP gives better results than the MGGD-IS and MGGD-IO dependences. Fourth, the KLD-scaled Wishart method is unable to detect the region 2 at the center of the image (see Fig. 3.(j) and (j')) whereas wavelet domain method successfully detected it. This is due to the fact that region 1 and 2 have the same speckle covariance matrix, and as a consequence, the spatial domain method based on the scaled Wishart distribution consider them to be similar and not different. This example confirms that the method based on the wavelet domain gives better results than the method based on the spatial domain. Finally, The HLT and LRT methods give the worst detection results. Indeed, these methods assume the complex Wishart distribution for the covariance matrix data actually simulated by the $\mathcal{G}_d^0$ model.

Another quantitative evaluation of the change detection performance is provided also for four different significance levels or specified FPRs $\{0.5\%, 1\%, 5\%, 10\%\}$ with the same configuration: window size of 16×16 and $L = 1$. Table IV illustrates different specified FARs and the detection rate (TPR) for all the methods. As shown in this table, the MGGD-IP reaches higher detection rates and lower overall error rates among the other approaches. It can also be observed that the performance of MGGD-IP is followed by MGGD-IP+IS+IO, MGGD-IO and finally by GGD. As the significance level of the FAR increases to 10%, the performance of MGGD always remains better than the GGD. The GD is worse than all other methods. The first spatial domain method given by KLD-scaled Wishart gives better results for specified FAR 0.5% and 1% than wavelet domain method. Indeed, the detection rate is higher than the rate of the other methods and the overall error rate is always the lowest. However, this tendency is reversed for specified FAR values equal to 5% and 10%. In fact, MGGD-IP wavelet domain method outperforms the scaled Wishart spatial domain method. The two other spatial domain methods, HLT and LRT, give the lowest detection rate and the highest overall error.

3) AUC: The AUC is provided for all methods in Table VI. At each scale, each filter and with each window size, the best and the worst values are highlighted in red and blue, respectively. From this table, we can draw the following general conclusions. First, it can be clearly seen that multivariate distributions given by MGGD provides better performance than univariate distributions as the GD and the GGD. It is also

TABLE III
CHANGE DETECTION PERFORMANCE FOR SIMULATED DATA WHEN OPTIMAL THRESHOLD IS APPLIED. THE BEST VALUES ARE MARKED BY RED COLOR. SLIDING WINDOW SIZE = 16, WAVELET TRANSFORM $L = 1$, DAUBECHIES WAVELET FILTER: DB1.

Methods	FAR (%)	Detection rate (%)	Overall error rate (%)
GD	19.03	85.67	17.99
GGD	12.29	86.89	12.47
MGD-IP	18.78	89.51	16.94
MGGD-IP	7.43	92.62	7.42
MGD-IS	19.03	85.67	17.99
MGGD-IS	12.29	86.89	12.47
MGD-IO	18.26	87.18	17.05
MGGD-IO	10.82	88.25	11.02
MGD-IP+IS+IO	18.21	87.32	16.98
MGGD-IP+IS+IO	9.36	90.89	9.31
Scaled Wishart	6.13	80.66	8.98
HLT	37.41	64.47	36.99
LRT	31.23	66.77	31.68

noted that the KLD-GD and the KLD-MGD give the worst detection results. Second, detection based on the GD and the MGD has similar performance and the variation between them is not very significant. Besides, the KLD-GD and the KLD-MGD-IS are the same when $L = 1$. The same observation is also valid for the KLD-GGD and the KLD-MGGD-IS. Third, we can also note that the AUC based on the MGGD is always the best for any scale, any filter order, and any window size, followed by the GGD. In addition, the AUC based on the KLD-MGGD-IP is better than that based on the KLD-MGGD-IO and the KLD-MGGD-IS, while the latter seems to be less efficient compared to the first two. Moreover, the AUC based on the KLD-MGGD-IP is even higher for $L = 2$ compared to the KL-GGD with $L = 1$ for any window size. The same observation is made between the KLD-MGGD-IO ($L = 2$) and the KLD-GGD ($L = 1$) when the window size ranges from 20 to 28. It means that the inter-polarization dependence contributes to the better detection, followed by the inter-orientation dependence. On the contrary, the KLD-MGGD-IS gives always lower AUC for $L = \{2, 3\}$ than that of KLD-GGD with $L = 1$ meaning that the inter-scale dependence does not significantly contribute to the better detection. Finally, for the total distance, MGGD-(IO + IS + IP), the AUC is an intermediate value between the maximum and the minimum of a mixture composed by (MGGD-IO, MGGD-IS, MGGD-IP).

We have also the following particular notes. First, for a fixed window size and scale, the AUC for MGDs and MGGDs decreases as the order of the filters increases. This is explained due to the fact that the correlation between the wavelet subbands decreases as the filter order increases. Consequently, the filter order 'db1' provides the best results. Second, for a fixed window size and filter order, the detection for all methods decreases as the wavelet transform scale increases. On the other hand, the scale $L = 1$ allows the best performance in this example. Finally, for a fixed scale and filter order, as the window size grows, the AUC increases until a particular window size and decreases afterward. The window sizes corresponding to the maximum AUC are given in Table V for each method.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3215783

11

TABLE IV
CHANGE DETECTION PERFORMANCE FOR SIMULATED DATA. THE BEST VALUES ARE MARKED BY RED COLOR.

Methods	Specified FAR(%)	Detection rate (%)	Overall error rate (%)
GD		8.73	20.65
GGD		14.44	19.38
MGD-IP		5.20	21.47
MGGD-IP	0.5	21.22	17.88
MGD-IS		8.73	20.65
MGGD-IS		14.44	19.38
MGD-IO		9.52	20.47
MGGD-IO		14.22	19.43
MGD-IP+IS+IO		6.86	21.07
MGGD-IP+IS+IO		14.54	19.36
Scaled Wishart		67.42	7.62
HLT		1.27	22.31
LRT		0.59	22.46
GD		13.91	19.89
GGD		18.02	18.98
MGD-IP		7.34	21.35
MGGD-IP	1	31.64	15.96
MGD-IS		13.91	19.89
MGGD-IS		18.02	18.98
MGD-IO		14.80	19.7
MGGD-IO		17.78	19.03
MGD-IP+IS+IO		11.24	20.48
MGGD-IP+IS+IO		21.34	18.24
Scaled Wishart		71.58	7.09
HLT		2.66	22.39
LRT		1.75	22.59
GD		35.81	18.15
GGD		68.65	10.85
MGD-IP		36.70	17.94
MGGD-IP	5	88.97	6.36
MGD-IS		35.81	18.15
MGGD-IS		68.65	10.85
MGD-IO		48.42	15.34
MGGD-IO		76.28	09.18
MGD-IP+IS+IO		37.30	17.81
MGGD-IP+IS+IO		81.53	8.00
Scaled Wishart		80.11	8.3
HLT		13.19	23.17
LRT		16.84	22.36
GD		63.04	16.01
GGD		83.83	11.37
MGD-IP		59.40	16.80
MGGD-IP	10	94.78	8.95
MGD-IS		63.04	16.01
MGGD-IS		83.83	11.37
MGD-IO		69.31	14.59
MGGD-IO		87.46	10.62
MGD-IP+IS+IO		62.00	16.22
MGGD-IP+IS+IO		91.33	9.75
Scaled Wishart		81.05	11.99
HLT		24.19	24.64
LRT		33.17	22.69

Method	Window size	L	dbn	AUC
MGGD-IP	16×16	1	db1	0.9685
MGGD-IO	20×20	1	db1	0.9435
MGGD-IS	20×20	1	db1	0.9390
MGD-IP	12×12	2	db1	0.9083
MGD-IO	12×12	1	db1	0.9057
MGD-IS	12×12	1	db1	0.8932
GGD	20×20	1	db1	0.9390
GD	12×12	1	db1	0.8932
Scaled Wishart	16×16	-	-	0.8930

TABLE V
MAXIMUM VALUES OF AUC REGISTERED FOR DIFFERENT METHODS WITH CORRESPONDING WINDOW SIZES, SCALES AND FILTER ORDER.
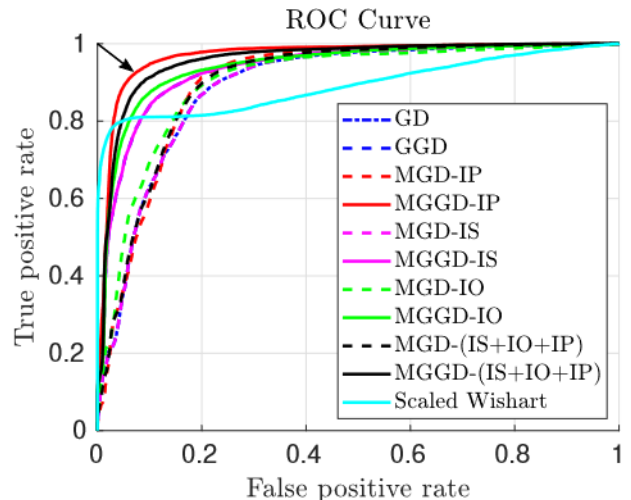

Fig. 4. ROC plot comparison between different methods and dependencies (IP, IO and IS) for simulation data. The selected optimal thresholds correspond to the closest point to (0,1) in the ROC curves. The sliding window, the wavelet transform filter, and the scale are respectively 16×16, db1, and $L = 1$.

curves of the GD and the MGD for both low and high FPR values. Moreover, The ROC curve of the MGGD-IP outperforms the other ROC curves of the MGGD-IS and the MGGD-IO. It should be noted that the ROC curves of GD and MGD-IS are superimposed. The same is for GGD and MGGD-IS. The KLD-scaled Wishart based method gives lower detection results than the wavelet domain method based on MGGD when the FARs are higher than 2%. This is due to the non detection of the region 2 in the center of the image. The ROC curves of the HLT and LRT methods are not shown in Fig. 4 because of the poor performance of these two methods in this example.

Finally, The AUC values for the HLT and LRT methods are also provided and are 0.6792 and 0.7270, respectively. It is clearly seen that the AUC of the HLT and LRT based on spatial domain are the lowest values, thus showing the worst detection results. The AUC of the KLD-scaled Wishart is always lower than the wavelet domain method based on MGGD.

4) ROC: The ROC curves obtained for different methods are represented in Fig. 4. The sliding window of size 16×16, a wavelet transform filter db1 and scale $L = 1$ have been employed in this example and the obtained results show that the ROC curves of the MGGD are always above the ROC

B. Experiments with real data

1) Data set: A co-registered pair of L-band polarimetric images ($d = 3$) was acquired by the Jet Propulsion Laboratory/National Aeronautics and Space Administration UAVSAR (1.26 GHz) over the city of Los Angeles, California, on April 23, 2009, and May 11, 2015. The images are 2×3 multilooked. Fig. 5(a) and (b) illustrate the Pauli decomposition of two images obtained by the JPL's UAVSAR sensor at two different times. Fig. 5(c) depicts the ground truth used to compute the ROC curve. We recall that white color corresponds to a change and black color to a no-change. As we can see, the

Size	Scale	dbn	GD	GGD	Inter-polarization		Inter-scale		Inter-orientation		Total Distance		Spatial domain
					MGD	MGGD	MGD	MGGD	MGD	MGGD	MGD	MGGD	Scaled Wishart
28	1	db1	0.8461	0.9371	0.8523	0.9551	0.8461	0.9371	0.8678	0.9410	0.8563	0.9498	0.8723
		db2	0.8457	0.9334	0.8517	0.9510	0.8457	0.9334	0.8667	0.9345	0.8560	0.9451	
		db3	0.8454	0.9284	0.8514	0.9468	0.8454	0.9284	0.8659	0.9282	0.8558	0.9400	
		db4	0.8452	0.9287	0.8512	0.9460	0.8452	0.9287	0.8659	0.9265	0.8557	0.9392	
		db5	0.8450	0.9256	0.8511	0.9428	0.8450	0.9256	0.8643	0.9231	0.8553	0.9356	
	2	db1	0.8457	0.9283	0.8523	0.9477	0.8457	0.9344	0.8699	0.9385	0.8570	0.9458	
		db2	0.8452	0.9237	0.8523	0.9424	0.8449	0.9307	0.8669	0.9293	0.8564	0.9401	
		db3	0.8449	0.9139	0.8521	0.9327	0.8448	0.9186	0.8638	0.9138	0.8558	0.9275	
		db4	0.8447	0.9047	0.8518	0.9248	0.8448	0.9109	0.8612	0.9009	0.8553	0.9172	
		db5	0.8446	0.9022	0.8516	0.9209	0.8445	0.9080	0.8597	0.8965	0.8548	0.9128	
24	1	db1	0.8601	0.9380	0.8686	0.9613	0.8601	0.9380	0.8784	0.9427	0.8707	0.9539	0.8819
		db2	0.8594	0.9348	0.8680	0.9573	0.8594	0.9348	0.8774	0.9350	0.8703	0.9485	
		db3	0.8590	0.9300	0.8677	0.9529	0.8590	0.9300	0.8769	0.9299	0.8700	0.9432	
		db4	0.8588	0.9302	0.8675	0.9521	0.8588	0.9302	0.8774	0.9285	0.8700	0.9425	
		db5	0.8586	0.9274	0.8674	0.9483	0.8586	0.9274	0.8755	0.9241	0.8695	0.9383	
	2	db1	0.8603	0.9291	0.8684	0.9526	0.8602	0.9360	0.8806	0.9401	0.8715	0.9489	
		db2	0.8596	0.9250	0.8684	0.9463	0.8592	0.9322	0.8779	0.9303	0.8707	0.9421	
		db3	0.8591	0.9164	0.8682	0.9353	0.8587	0.9207	0.8751	0.9161	0.8699	0.9291	
		db4	0.8587	0.9079	0.8679	0.9263	0.8586	0.9138	0.8728	0.9045	0.8693	0.9193	
		db5	0.8583	0.9061	0.8676	0.9226	0.8581	0.9109	0.8713	0.9003	0.8688	0.9155	
	3	db1	0.8617	0.9047	0.8646	0.9166	0.8618	0.9218	0.8771	0.9180	0.8714	0.9230	
		db2	0.8579	0.8923	0.8610	0.8941	0.8574	0.9084	0.8690	0.8988	0.8671	0.9036	
		db3	0.8549	0.8740	0.8580	0.8755	0.8546	0.8955	0.8596	0.8773	0.8633	0.8857	
		db4	0.8525	0.8596	0.8552	0.8601	0.8526	0.8832	0.8566	0.8730	0.8608	0.8729	
		db5	0.8505	0.8514	0.8527	0.8463	0.8506	0.8780	0.8526	0.8654	0.8582	0.8604	
20	1	db1	0.8805	0.9390	0.8869	0.9662	0.8805	0.9390	0.8922	0.9435	0.8886	0.9570	0.8895
		db2	0.8798	0.9361	0.8862	0.9619	0.8798	0.9361	0.8915	0.9358	0.8881	0.9510	
		db3	0.8795	0.9314	0.8859	0.9571	0.8795	0.9314	0.8912	0.9308	0.8878	0.9453	
		db4	0.8793	0.9311	0.8857	0.9562	0.8793	0.9311	0.8914	0.9296	0.8877	0.9442	
		db5	0.8791	0.9290	0.8856	0.9523	0.8791	0.9290	0.8901	0.9251	0.8873	0.9403	
	2	db1	0.8798	0.9291	0.8873	0.9556	0.8799	0.9364	0.8929	0.9393	0.8888	0.9504	
		db2	0.8787	0.9249	0.8868	0.9484	0.8784	0.9323	0.8907	0.9302	0.8878	0.9426	
		db3	0.8781	0.9176	0.8866	0.9369	0.8776	0.9220	0.8887	0.9168	0.8870	0.9304	
		db4	0.8776	0.9090	0.8865	0.9283	0.8775	0.9154	0.8863	0.9064	0.8863	0.9219	
		db5	0.8773	0.9072	0.8863	0.9252	0.8772	0.9128	0.8847	0.9035	0.8860	0.9189	
16	1	db1	0.8917	0.9355	0.9014	0.9685	0.8917	0.9355	0.9018	0.9422	0.9011	0.9575	0.8930
		db2	0.8908	0.9324	0.9006	0.9640	0.8908	0.9324	0.9013	0.9316	0.9002	0.9502	
		db3	0.8904	0.9274	0.9002	0.9579	0.8904	0.9274	0.9005	0.9251	0.8997	0.9434	
		db4	0.8902	0.9257	0.9000	0.9570	0.8902	0.9257	0.8998	0.9235	0.8993	0.9417	
		db5	0.8900	0.9232	0.8999	0.9522	0.8900	0.9232	0.8988	0.9206	0.8990	0.9380	
	2	db1	0.8926	0.9223	0.9029	0.9556	0.8927	0.9305	0.9031	0.9336	0.9021	0.9484	
		db2	0.8915	0.9192	0.9020	0.9473	0.8914	0.9270	0.9000	0.9219	0.9008	0.9397	
		db3	0.8908	0.9119	0.9014	0.9373	0.8906	0.9162	0.8979	0.9127	0.8998	0.9288	
		db4	0.8903	0.9022	0.9009	0.9283	0.8904	0.9101	0.8944	0.9041	0.8985	0.9203	
		db5	0.8899	0.9018	0.9006	0.9262	0.8900	0.9091	0.8945	0.9074	0.8985	0.9203	
	3	db1	0.8921	0.8823	0.8976	0.9034	0.8921	0.9116	0.8957	0.8973	0.8984	0.9060	
		db2	0.8877	0.8672	0.8931	0.8863	0.8874	0.9032	0.8893	0.8923	0.8937	0.8961	
		db3	0.8852	0.8414	0.8900	0.8669	0.8851	0.8939	0.8826	0.8763	0.8902	0.8783	
		db4	0.8839	0.8086	0.8882	0.8464	0.8840	0.8880	0.8738	0.8614	0.8855	0.8564	
		db5	0.8832	0.7821	0.8872	0.8312	0.8833	0.8892	0.8724	0.8638	0.8840	0.8473	
12	1	db1	0.8932	0.9180	0.9063	0.9636	0.8932	0.9180	0.9057	0.9278	0.9048	0.9463	0.8918
		db2	0.8925	0.9160	0.9052	0.9564	0.8925	0.9160	0.9031	0.9145	0.9034	0.9378	
		db3	0.8922	0.9098	0.9047	0.9487	0.8922	0.9098	0.9014	0.9072	0.9026	0.9298	
		db4	0.8920	0.9067	0.9044	0.9475	0.8920	0.9067	0.9002	0.9058	0.9020	0.9273	
		db5	0.8918	0.9053	0.9041	0.9426	0.8918	0.9053	0.8993	0.9037	0.9017	0.9246	
	2	db1	0.8925	0.8891	0.9083	0.9411	0.8926	0.9106	0.9008	0.9059	0.9043	0.9291	
		db2	0.8903	0.8888	0.9062	0.9343	0.8903	0.9055	0.8951	0.8962	0.9018	0.9218	
		db3	0.8881	0.8747	0.9040	0.9202	0.8881	0.8945	0.8918	0.8900	0.8997	0.9106	
		db4	0.8861	0.8502	0.9017	0.8988	0.8861	0.8872	0.8848	0.8810	0.8964	0.8973	
		db5	0.8844	0.8386	0.8996	0.8885	0.8842	0.8858	0.8852	0.8838	0.8956	0.8929	

TABLE VI

AUC FOR DIFFERENT WINDOW SIZES, DIFFERENT SCALES, AND DIFFERENT DEPENDENCIES (IP, IO AND IS) MEASURED FOR GD, GGD, MGD, AND MGGD, AND COMPARED WITH KLD-SCALED WISHART METHOD IN THE CASE OF SIMULATED DATA. THE BEST VALUES ARE MARKED BY RED COLOR AND THE WORST BY BLUE COLOR.

interesting area of this dataset is an urban area, where the changes occurred due to the effects of urbanization.

2) Experiments: To evaluate and to obtain reliable results, the input image is decomposed into $L = (1, 2)$ scales using discrete stationary wavelet transform (SWT) with a Daubechies filter bank. For each coefficient magnitude of each scale, different sliding windows with size (16, 20, 24, 28) are used in this study for performance evaluation. As in the previous section, the IS, IO, and IP dependencies are used. Indeed, the distances $\sum_p d_{KL}(\mathcal{X}_p^1, \mathcal{X}_p^2)$, $\sum_p d_{KL}(\mathcal{Y}_p^1, \mathcal{Y}_p^2)$ and $d_{KL}(\mathcal{H}^1, \mathcal{H}^2) + d_{KL}(\mathcal{V}^1, \mathcal{V}^2) + d_{KL}(\mathcal{D}^1, \mathcal{D}^2)$ are computed with the MGGD and the MGD and compared with the distances corresponding to the case where the subbands are assumed independent with the univariate GGD and GD.

Fig. 6 illustrates the different logarithm of KLD and the corresponding binary change map obtained by the thresholding the KLD. Indeed, binary change detection map is obtained at significance level or specified FPR equal to 5%. The sliding window size corresponds to 16×16, with a wavelet transform scale $L = 1$, and Daubechies wavelet filter: 'db1'. The distances are computed for both univariate (GD, GGD) and multivariate distributions (MGD and MGGD). We release the following comments. The GD and MGD give us the worst detection results as it is shown in Fig. 6(h) and Fig. 6(b)-(d)-(f) and their corresponding binary change map in Fig. 6(h') and Fig. 6(b')-(d')-(f'). The KLD using the MGGD provide better performance than that based on the GGD as more areas are detected. Accordingly, taking into account the correlation structure between subbands improves in this example the performance and the accuracy of the change map. The comparison between KLD based on MGGD-IP gives better results than the MGGD-IS and MGGD-IO dependence.

A quantitative evaluation of the change detection performance is provided also for four different significance levels or specified FPRs (false positive rate) $\{0.5\%, 1\%, 5\%, 10\%\}$ with a window size of 16×16 and $L = 1$. Table VII illustrates different specified FARs and the detection rate (TPR) for all the methods. As indicated in this table, the MGGDP-IP reaches higher detection rates and lower overall error rates among the other wavelet domain approaches. It can also be observed that the performance of MGGD-IP is followed by MGGD-IO, MGGD-IP+IS+IO and finally by GGD. As the significance level of the FAR increases to 10%, the detection rates for these methods increase and become quite similar, while the MGGD remains always better than the GGD. We also conclude that the wavelet domain method based on MGGD outperforms the KLD-scaled Wishart method. The TPR of the latter is always lower than the TPR of the former, especially for low values of the specified FPRs. When FPR increases and reaches 10%, the performance of both methods becomes more and less similar. It is clearly seen that MGGD-IP,-IS and -IO outperform the HLT and LRT methods when the specified FAR is $\geq 5\%$. However, HLT and LRT give the best performance for low values of the specified FAR (0.5%, 1%).

In Table VIII, we provide the AUC for all methods. According to this table, it can be concluded that the worst performance for the wavelet domain method is given by the GD and MGD. On the other hand, the proposed MGGD outperforms

TABLE VII
CHANGE DETECTION PERFORMANCE FOR REAL DATA. THE BEST VALUES ARE MARKED BY RED COLOR.

Methods	Specified FAR(%)	Detection rate (%)	Overall error rate (%)
GD		11.36	14.07
GGD		13.17	13.80
MGD-IP		10.55	14.20
MGGD-IP	0.5	13.51	13.74
MGD-IS		11.36	14.07
MGGD-IS		13.17	13.80
MGD-IO		11.59	14.04
MGGD-IO		13.46	13.75
MGD-IP+IS+IO		11.15	14.10
MGGD-IP+IS+IO		13.36	13.76
Scaled Wishart		8.76	14.60
HLT		20.69	12.63
LRT		22.48	12.36
GD		16.81	13.65
GGD		23.48	12.63
MGD-IP		17.38	13.57
MGGD-IP	1	25.03	12.39
MGD-IS		16.81	13.65
MGGD-IS		23.48	12.63
MGD-IO		16.86	13.65
MGGD-IO		23.36	12.65
MGD-IP+IS+IO		17.19	13.60
MGGD-IP+IS+IO		23.58	12.61
Scaled Wishart		14.82	14.08
HLT		28.25	11.89
LRT		28.75	11.81
GD		67.26	9.28
GGD		72.06	8.54
MGD-IP		67.67	9.21
MGGD-IP	5	74.60	8.16
MGD-IS		67.26	9.28
MGGD-IS		72.06	8.54
MGD-IO		68.07	9.15
MGGD-IO		73.75	8.28
MGD-IP+IS+IO		67.91	9.17
MGGD-IP+IS+IO		74.26	8.20
Scaled Wishart		68.14	9.18
HLT		48.56	12.17
LRT		44.88	12.74
GD		84.09	10.92
GGD		87.15	10.48
MGD-IP		85.84	10.65
MGGD-IP	10	87.76	10.36
MGD-IS		84.09	10.92
MGGD-IS		87.15	10.48
MGD-IO		83.20	11.04
MGGD-IO		87.28	10.42
MGD-IP+IS+IO		85.25	10.75
MGGD-IP+IS+IO		87.93	10.31
Scaled Wishart		86.38	10.56
HLT		57.94	14.99
LRT		53.73	15.61

the univariate GGD giving us the best performance. It is also be highlighted that for all window sizes, the highest value of AUC is always given by MGGD-IP, followed by MGGD-IO, and MGGD-IP+IS+IO. This can be explaining due to the fact that the correlation structure between subbands allows improving the performance and the accuracy of the change map. In addition, the AUC based on the KLD-MGGD-IP, KLD-MGGD-IO and MGGD-IP+IS+IO are even higher for $L = 2$ compared to the KL-GGD with $L = 1$ for any window sizes of $\{16, 20, 24, 28\}$. It reinforces our belief that the correlation structure improves the change detection performance.
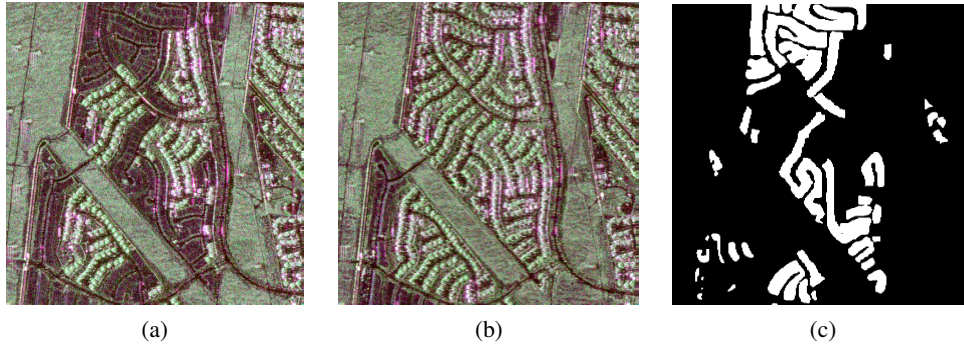
Fig. 5. UAVSAR images (in Pauli decomposition) on April 23, 2009 and May 11, 2015. (a) before change. (b) after change. (c) ground truth where white color is change and black color is no-change.

Furthermore, the scale $L = 1$ gives better performance than for $L = 2$ where the optimal performance is attained with $L = 1$ and a window size of 16×16 in our case of study. The AUC for KLD-scaled Wishart is always lower than the wavelet domain based method. The performance of the former decreases as the window size increases. For small window size, both methods seems to have similar performance. The AUC for both HLT and LRT methods are equal to 0.8113 and 0.7925, respectively, thus showing the lowest values among the other methods. In Fig. 7, the ROC curves corresponding to different methods are shown. In this evaluation, the sliding window of size 16×16, a wavelet transform filter db1 and scale $L = 1$ have been employed. The obtained results show that the ROC curves of MGGD are always better than those of GD and MGD for both low and high FPR values. Moreover, the ROC curve of the MGGD-IP outperforms the other ROC curves of the MGGD-IS and MGGD-IO. It is worth to mention that the ROC curves of MGGD-IS and MGD-IS are not shown in Fig. 7 because they are overlaid with the GGD and GD ROC curves, respectively. We also notice that the ROC curve of the KDL-scaled Wishart method is always below the ROC curves of the wavelet domain method based on MGGD when FPR is less than 30%. But beyond this value, the curve becomes above the other ROC curves. The ROC curves of the HLT and LRT methods show poor performance compared to the other methods. From all these experiments, we have shown encouraging results for change detection based on multivariate statistical wavelet subband modeling, allowing to take into account the correlation structure between subbands.

VI. CONCLUSION

In this paper, an unsupervised method for PolSAR change detection in the wavelet domain is proposed. The method is based on the multivariate statistical wavelet subband modeling assumed to be MGGD. Through this distribution, it is possible to take into account the correlation structure between subbands. Three correlations are investigated: inter-scale dependence, inter-orientation dependence, and inter-polarization dependence. The first is to model the correlation between the scales of the same orientation, the second is to model the dependence between the orientations of the same scale, and the third is to measure the dependency between polarization with the same scale and orientation. Closed-

form expression of Kullback-Leibler divergence between two MGGDs is exploited and used to generate the change map. The approach is evaluated using simulated and real PolSAR data with different parameter settings, different scales, and sliding window sizes. Through this study, the MGGD shows better results than the GGD as the correlations between subbands are characterized by multivariate distribution. Moreover, PolSAR change detection in the wavelet domain is better represented by MGGD than MGD since the latter is a special case of MGGD. In addition, the comparison between correlations shows that MGGD-IP gives better results than the MGGD-IS and MGGD-IO dependences. Moreover, the performance of the proposed change detection method is compared to the performance of three methods performed in spatial domain: the homogeneous change detection based on the scaled complex Wishart distribution, the HLT and the LRT methods. The results show that the wavelet domain can outperform the spatial method. Although the approach performs well, we can improve the results by considering not only the diagonal elements of the multilook polarimetric covariance matrix for the study of change detection but also by adding the other complex components of the covariance matrix. This requires the knowledge of the closed-form expression of KL for complex MGGD not yet established to generate the change map.

APPENDIX A
LAURICELLA FUNCTION

The Lauricella series $F_D^{(n)}$ given as follows [42]

$$F_D^{(n)}(a, b_1, ..., b_n; c; x_1, ..., x_n) \qquad (28)$$
$$= \sum_{m_1=0}^{\infty} ... \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+...+m_n}(b_1)_{m_1}...(b_n)_{m_n}}{(c)_{m_1+...+m_n}} \frac{x_1^{m_1}}{m_1!}...\frac{x_n^{m_n}}{m_n!}$$

where $|x_1|, ..., |x_n| < 1$. The Pochhammer symbol $(q)_i$ indicates the i-th rising factorial of q, i.e.

$$(q)_i = q(q+1)...(q+i-1) = \frac{\Gamma(q+i)}{\Gamma(q)} \quad \text{if} \quad i = 1, 2, ... \quad (29)$$

If $i = 0$, $(q)_i = 1$. Function $F_D^{(n)}(.)$ can be expressed in terms of multiple integrals as follows [67]

$$F_D^{(n)}(a, b_1, ..., b_n; c; x_1, ..., x_n) = \frac{\Gamma(c)}{\Gamma(c - \sum_{i=1}^{n} b_i) \prod_{i=1}^{n} \Gamma(b_i)} \times$$
$$\int_{\Omega} ... \int \prod_{i=1}^{n} u_i^{b_i-1}(1 - \sum_{i=1}^{n} u_i)^{c - \sum_{i=1}^{n} b_i - 1}(1 - \sum_{i=1}^{n} x_i u_i)^{-a} \prod_{i=1}^{n} du_i$$
$$(30)$$

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3215783
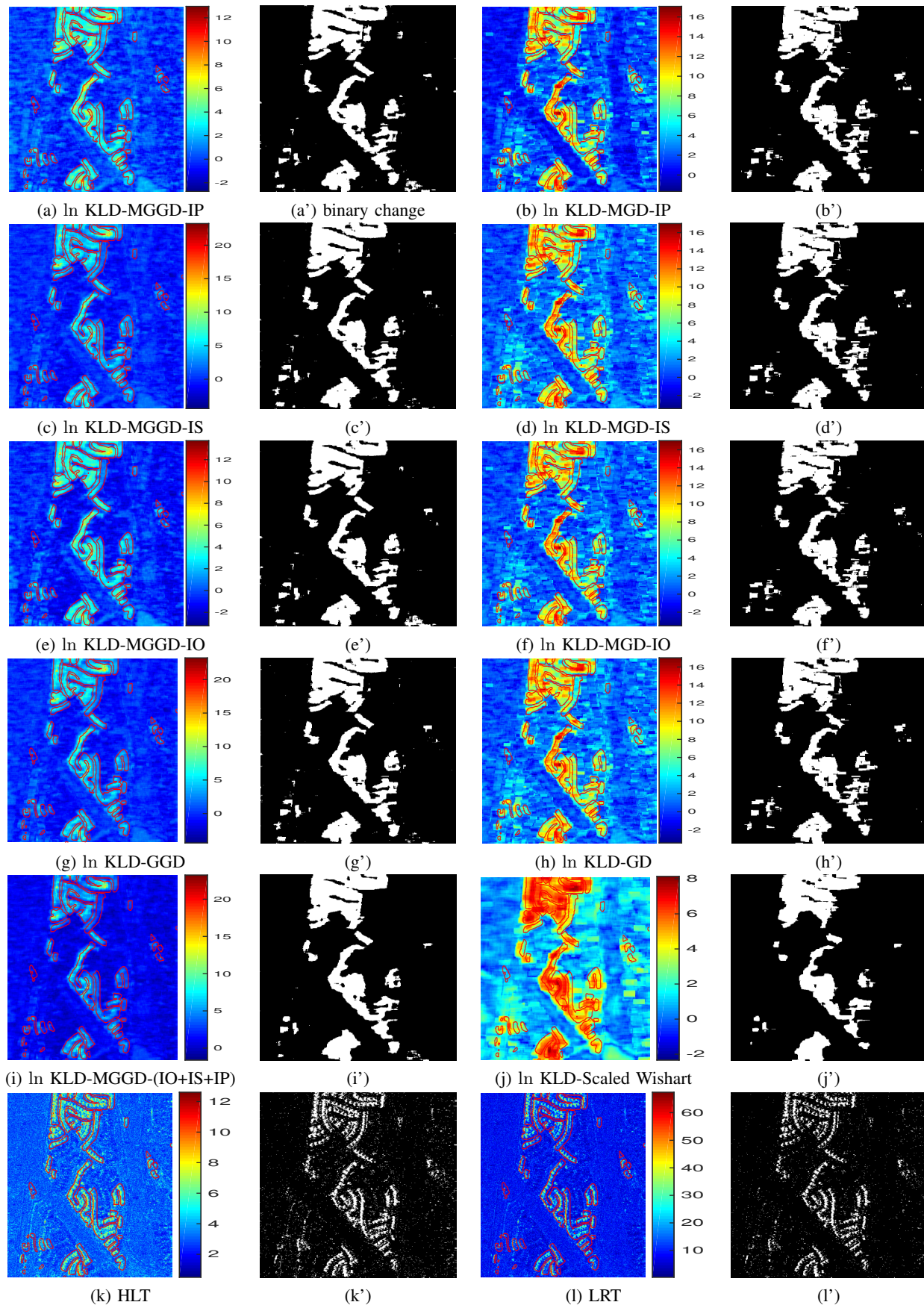
15

Fig. 6. Detection results for real PolSAR data. (a) and (b) ln KLD for IP dependence with MGGD and MGD and their binary change maps (a')-(b'). (c)-(d) ln KLD for IS dependence with MGGD and MGD and their binary maps (c')-(d'). (e)-(f) ln KLD IO dependence with MGGD and MGD, and (e')-(f') their corresponding binary maps. (g)-(h) ln KLD for univariate GGD and GD, and their binary change maps (g')-(h'). (i) and (i') ln KLD-MGGD (IP+IS+IO) dependence with MGGD and binary map. (j) and (j') ln KLD-scaled Wishart and corresponding change map. (k)-(l) results of the HLT and LRT methods. The parameters used: sliding window size=16 , wavelet transform L=1, Daubechies wavelet filter: db1. Binary map obtained at specified FPR=5%.

Size	Scale	dbn	GD	GGD	Inter-polarization		Inter-scale		Inter-orientation		Total Distance		Spatial domain
					MGD	MGGD	MGD	MGGD	MGD	MGGD	MGD	MGGD	Scaled Wishart
28	1	db1	0,8803	0,9144	0,8874	0,9243	0,8803	0,9144	0,8674	0,9198	0,8850	0,9220	0.9078
	2	db1	0,8814	0,9128	0,8885	0,9219	0,8816	0,9135	0,8724	0,9200	0,8867	0,9196	
24	1	db1	0,8991	0,9195	0,9043	0,9293	0,8991	0,9195	0,8882	0,9249	0,9026	0,9271	0.9180
	2	db1	0,9006	0,9176	0,9053	0,9273	0,9008	0,9188	0,8927	0,9251	0,9043	0,9247	
20	1	db1	0,9168	0,9252	0,9216	0,9351	0,9168	0,9252	0,9078	0,9310	0,9202	0,9329	0.9263
	2	db1	0,9180	0,9224	0,9224	0,9320	0,9182	0,9239	0,9110	0,9302	0,9216	0,9299	
16	1	db1	0,9314	0,9321	0,9358	0,9412	0,9314	0,9321	0,9234	0,9378	0,9348	0,9395	0.9338
	2	db1	0,9321	0,9293	0,9367	0,9361	0,9321	0,9310	0,9251	0,9353	0,9355	0,9358	

TABLE VIII

THE AREA UNDER CURVE (AUC) FOR DIFFERENT WINDOW SIZES, DIFFERENT SCALES, AND DIFFERENT DEPENDENCIES (IP, IO, AND IS) ARE MEASURED FOR GD, GGD, MGD, AND MGGD, AND COMPARED TO KLD-SCALED WISHART METHOD IN THE CASE OF REAL DATA. THE BEST VALUES ARE MARKED BY RED COLOR AND THE WORST BY BLUE COLOR.
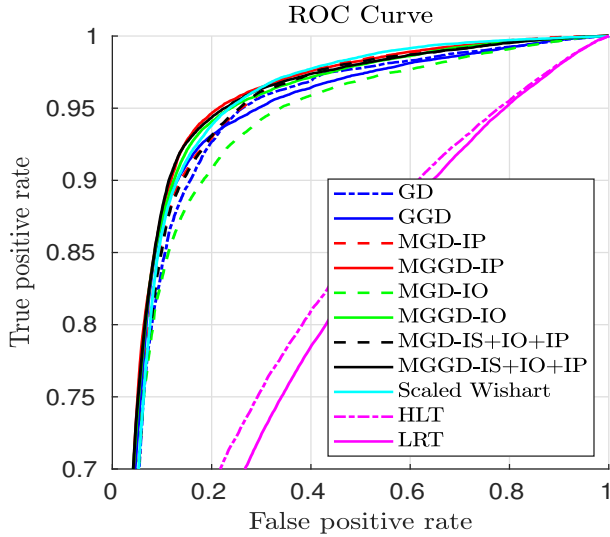

Fig. 7. ROC plot comparison between different methods with and without dependencies (IP, IO and IS). The slide window, the wavelet transform filter and scale are respectively 16×16, db1 and $L = 1$.

where $\Omega = \{(u_1, u_2, ..., u_n)|0 \leq u_i \leq 1, i = 1, ..., n, \text{and } 0 \leq u_1 + u_2 + ... + u_n \leq 1\}$, $\text{Re}(b_i) > 0$ for $i = 1, ..., n$ and $\text{Re}(c - b_1 - ... - b_n) > 0$. Lauricella's F_D can be written as a one-dimensional Euler-type integral for any number n of variables. The integral form of $F_D^{(n)}(.)$ is given as follows when $\text{Re}(a) > 0$ and $\text{Re}(c - a) > 0$

$$F_D^{(n)}(a, b_1, ..., b_n; c; x_1, ..., x_n) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 u^{a-1} \times$$
$$(1 - u)^{c-a-1}(1 - ux_1)^{-b_1}...(1 - ux_n)^{-b_n} \mathrm{d}u \quad (31)$$

Lauricella has given several transformation formulas, from which we use the two following relationships. More details can be found in Exton's book [68] on hypergeometric equations.

$$F_D^{(n)}(a, b_1, ..., b_n; c; x_1, ..., x_n)$$
$$= \prod_{i=1}^n (1 - x_i)^{-b_i} F_D^{(n)}(c - a, b_1, ..., b_n; c; \frac{x_1}{x_1 - 1}, ..., \frac{x_n}{x_n - 1}) \quad (32)$$

$$= (1 - x_1)^{-a} F_D^{(n)}(a, c - \sum_{i=1}^n b_i, b_2, ..., b_n; c; \frac{x_1}{x_1 - 1}, \frac{x_1 - x_2}{x_1 - 1},$$
$$, ..., \frac{x_1 - x_n}{x_1 - 1}) \quad (33)$$

REFERENCES

[1] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment," *IEEE Trans. on Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1658–1670, June 2007.

[2] E. J. M. Rignot and J. J. van Zyl, "Change detection techniques for ERS-1 SAR data," *IEEE Trans. on Geosci. Remote Sens.*, vol. 31, no. 4, pp. 896–906, Jul 1993.

[3] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. on Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, April 2005.

[4] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297–3317, 2014.

[5] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. on Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[6] S. Cui, G. Schwarz, and M. Datcu, "A benchmark evaluation of similarity measures for multitemporal SAR image change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1101–1118, March 2016.

[7] A. D. C. Nascimento, R. J. Cintra, and A. C. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Trans. on Geosci. Remote Sens.*, vol. 48, no. 1, pp. 373–385, Jan 2010.

[8] A. M. Atto, E. Trouve, Y. Berthoumieu, and G. Mercier, "Multidate divergence matrices for the analysis of SAR image time series," *IEEE Trans. on Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1922–1938, April 2013.

[9] L. Gueguen, P. Soille, and M. Pesaresi, "Change detection based on information measure," *IEEE Trans. on Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4503–4515, Nov 2011.

[10] E. Erten, A. Reigber, L. Ferro-Famil, and O. Hellwich, "A new coherent similarity measure for temporal multichannel scene characterization," *IEEE Trans. on Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2839–2851, July 2012.

[11] L. Gueguen and M. Datcu, "Mixed information measure: Application to change detection in earth observation," *Anal. Multi-Temporal Remote Sens. Images (MultiTemp), Mystic, CT, in Proc. 5th Int. Workshop*, Jul. 2009.

[12] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Trans. on Geosci. Remote Sens.*, vol. 41, no. 1, pp. 4–19, Jan 2003.

[13] K. Conradsen, A. A. Nielsen, and H. Skriver, "Change detection in polarimetric SAR data over several time points," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 4540–4543.

[14] K. Conradsen, A. A. Nielsen, and H. Skriver, "Determining the points of change in time series of polarimetric SAR data," *IEEE Trans. on Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3007–3024, May 2016.

[15] A. A. Nielsen, K. Conradsen, and H. Skriver, "Change detection in full and dual polarization, single- and multifrequency SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 4041–4048, Aug 2015.

[16] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Analytic expressions for stochastic distances between relaxed complex Wishart distributions," *IEEE Trans. on Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1213–1226, Feb 2014.

[17] A. D. C. Nascimento, A. C. Frery, and R. J. Cintra, "Detecting changes in fully polarimetric SAR imagery with statistical information theory," *IEEE Trans. on Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1380–1392, March 2019.

[18] N. Bouhlel and S. Méric, "Multilook polarimetric SAR change detection using stochastic distances between matrix-variate $\mathcal{G}_d^0$ distributions," *IEEE Trans. on Geosci. Remote Sens.*, vol. 58, no. 10, pp. 6823–6843, 2020.

[19] V. Akbari, S. N. Anfinsen, A. P. Doulgeris, T. Eltoft, G. Moser, and S. B. Serpico, "Polarimetric SAR change detection with the complex Hotelling-Lawley Trace statistic," *IEEE Trans. on Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3953–3966, July 2016.

[20] M. Ghanbari and V. Akbari, "Unsupervised change detection in polarimetric SAR data with the Hotelling-Lawley Trace statistic and minimum-error thresholding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4551–4562, 2018.

[21] N. Bouhlel, V. Akbari, and S. Méric, "Change detection in multilook polarimetric SAR imagery with determinant ratio test statistic," *IEEE Trans. on Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.

[22] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sensing*, vol. 14, no. 4, 2022.

[23] X. Zhang, X. Su, Q. Yuan, and Q. Wang, "Spatial-temporal gray-level co-occurrence aware CNN for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[24] W. Shi, M Zhang, R Zhang, S. Chen, and Z Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sensing*, vol. 12, no. 10, 2020.

[25] F. Liu, L. Jiao, B. Hou, and S. Yang, "POL-SAR image classification based on wishart DBN and local spatial information," *IEEE Trans. on Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3292–3308, 2016.

[26] P. Planinšič and D. Gleich, "Temporal change detection in SAR images using log cumulants and stacked autoencoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 297–301, 2018.

[27] F. Gao, J. Dong, B. Li, and Q. Xu, "Automatic change detection in synthetic aperture radar images based on PCANet," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1792–1796, 2016.

[28] Y. Li, C. Peng, Y. Chen, L. Jiao, L. Zhou, and R. Shang, "A deep learning method for change detection in synthetic aperture radar images," *IEEE Trans. on Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5751–5763, 2019.

[29] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. on Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, 2020.

[30] X. Niu, M. Gong, T. Zhan, and Y. Yang, "A conditional adversarial network for change detection in heterogeneous images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 45–49, 2019.

[31] Shiyong C. and M. Datcu, "Statistical wavelet subband modeling for multi-temporal SAR change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1095–1109, Aug 2012.

[32] M. Kivanc Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, 1999.

[33] S. K. Choy and C. S. Tong, "Statistical wavelet subband characterization based on generalized Gamma density and its application in texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 281–289, Feb 2010.

[34] M.N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb 2002.

[35] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution - characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, March 2010.

[36] G. Verdoolaege, Y. Rosseel, M. Lambrechts, and P. Scheunders, "Wavelet-based colour texture retrieval using the Kullback-Leibler divergence between bivariate generalized Gaussian models," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, Nov 2009, pp. 265–268.

[37] A.M. Atto, E. Trouve, Y. Berthoumieu, and G. Mercier, "Multidate divergence matrices for the analysis of SAR image time series," *IEEE Trans. on Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1922–1938, April 2013.

[38] J. Huang, "Study on the correlation properties of wavelet transform coefficients and the appliations in a neural network-based hybrid image coding system," in *Proceedings of the International Conference on Imaging Science, Systems and Technology, CISST'03, June 23 - 26, 2003,* *Las Vegas, Nevada, USA, Volume 1*, Hamid R. Arabnia and Youngsong Mun, Eds. 2003, pp. 43–48, CSREA Press.

[39] G. Tzgkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval with gaussianized steerable pyramids," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2702–2718, Sept 2006.

[40] L. Bombrun and Y. Berthoumieu, "Multivariate texture retrieval using the Kullback-Leibler divergence between bivariate generalized Gamma times an uniform distribution," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 2413–2416.

[41] N. Bouhlel and A. Dziri, "Kullback-Leibler divergence between multivariate generalized Gaussian distributions," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1021–1025, July 2019.

[42] G. Lauricella, "Sulle funzioni ipergeometriche a piu variabili," *Rendiconti del Circolo Matematico di Palermo*, vol. 7, no. 1, pp. 111–158, Dec 1893.

[43] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, Boca Raton, FL, USA: CRC Press, 2009.

[44] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate mellin transform to analysis of polarimetric radar images," *IEEE Trans. on Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, June 2011.

[45] A. C. Frery, H. . Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. on Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.

[46] C. Freitas, A. Frery, and A. Correia, "The polarimetric $\mathcal{G}$ distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, Feb. 2005.

[47] L. Bombrun and J. M. Beaulieu, "Fisher distribution for texture modeling of polarimetric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, July 2008.

[48] R. Kwitt and A. Uhl, "Lightweight probabilistic texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 241–253, Jan 2010.

[49] M. K. Varanasi and B. Aazhang, "Parametric generalized Gaussian density estimation," *The Journal of the Acoustical Society of America*, vol. 86, no. 4, pp. 1404–1415, 1989.

[50] Samuel Kotz, "Multivariate distributions at a cross road," in *A Modern Course on Statistical Distributions in Scientific Work*, G. P. Patil, S. Kotz, and J. K. Ord, Eds., Dordrecht, 1975, pp. 247–270, Springer Netherlands.

[51] E. Gomez, M. A. Gomez-Viilegas, and J. M. Marín, "A multivariate generalization of the power exponential family of distributions," *Communications in Statistics - Theory and Methods*, vol. 27, no. 3, pp. 589–600, 1998.

[52] Dongwook C. and Tien D. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms," *Signal Processing: Image Communication*, vol. 20, no. 1, pp. 77 – 89, 2005.

[53] G. Verdoolaege and P. Scheunders, "Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination," *International Journal of Computer Vision*, vol. 95, no. 3, pp. 265, May 2011.

[54] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," *Journal of Mathematical Imaging and Vision*, vol. 43, no. 3, pp. 180–193, Jul 2012.

[55] F. Pascal, L. Bombrun, J. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5960–5971, Dec 2013.

[56] S. Cui and M. Datcu, "Comparison of Kullback-Leibler divergence approximation methods between Gaussian mixture models for satellite image retrieval," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 3719–3722.

[57] J. R. Bunch and R. D. Fierro, "A constant-false-alarm-rate algorithm," *Linear Algebra and its Applications*, vol. 172, pp. 231 – 241, 1992.

[58] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan 1979.

[59] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognition*, vol. 19, no. 1, pp. 41 – 47, 1986.

[60] Dong L., Zhaohui J., and Huanqing F., "A novel fuzzy classification entropy approach to image thresholding," *Pattern Recognition Letters*, vol. 27, no. 16, pp. 1968 – 1975, 2006.

[61] S. Cui, G. Schwarz, and M. Datcu, "A comparative study of statistical models for multilook SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1752–1756, Oct 2014.

[62] J. I. Fernández-Michelli, M. Hurtado, J. A. Areta, and C. H. Muravchik, "Unsupervised polarimetric SAR image classification using $\mathcal{G}_p^0$ mixture model," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 754–758, May 2017.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3215783

18

[63] N. Bouhlel and S. Méric, "Maximum-likelihood parameter estimation of the product model for multilook polarimetric SAR data," *IEEE Trans. on Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1596–1611, 2019.

[64] N. Bouhlel and S. Méric, "Unsupervised segmentation of multilook polarimetric synthetic aperture radar images," *IEEE Trans. on Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6104–6118, 2019.

[65] N. Henze and B. Zirkler, "A class of invariant consistent tests for multivariate normality," *Communications in Statistics - Theory and Methods*, vol. 19, no. 10, pp. 3595–3617, 1990.

[66] P. J. Farrell, M. Salibian-Barrera, and K. Naczk, "On tests for multivariate normality and associated simulation studies," *Journal of Statistical Computation and Simulation*, vol. 77, no. 12, pp. 1065–1080, 2007.

[67] A. Hattori and T. Kimura, "On the Euler integral representations of hypergeometric functions in several variables," *J. Math. Soc. Japan*, vol. 26, no. 1, pp. 1–16, 01 1974.

[68] H. Exton, *Multiple Hypergeometric Functions and Applications*, Wiley, New York, 1976.