Detecting Changes in Fully Polarimetric SAR Imagery With Statistical Information Theory

Abraão D. C. Nascimento[ID], Alejandro C. Frery[ID], *Senior Member, IEEE*,
and Renato J. Cintra[ID], *Senior Member, IEEE*

Abstract—Images obtained from coherent illumination processes are contaminated with speckle. A prominent example of such imagery systems is the polarimetric synthetic aperture radar (PolSAR). For such a remote sensing tool, the speckle interference pattern appears in the form of a positive-definite Hermitian matrix, which requires specialized models and makes change detection a hard task. The scaled complex Wishart distribution is a widely used model for PolSAR images. Such a distribution is defined by two parameters: the number of looks and the complex covariance matrix. The last parameter contains all the necessary information to characterize the backscattered data, and thus, identifying changes in a sequence of images can be formulated as a problem of verifying whether the complex covariance matrices differ at two or more takes. This paper proposes a comparison between a classical change detection method based on the likelihood ratio and three statistical methods that depend on information-theoretic measures: the Kullback–Leibler (KL) distance and two entropies. The performance of these four tests was quantified in terms of their sample test powers and sizes using simulated data. The tests are then applied to actual PolSAR data. The results provide evidence that tests based on entropies may outperform those based on the KL distance and likelihood ratio statistics.

Index Terms—Change detection, contrast, hypothesis test, information theory, Wishart.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) has been widely used as an important system for information extraction in remote sensing applications. Such microwave active sensors have as main advantages the following features: 1) their operation is not determined by day time, neither weather conditions and 2) they are capable of providing high spatial image resolution.

A. D. C. Nascimento is with the Departamento de Estatística, Universidade Federal de Pernambuco, Recife 50740-540, Brazil (e-mail: abraao.susej@gmail.com).

A. C. Frery is with the Laboratório de Computação Científica e Análise Numérica, Universidade Federal de Alagoas, Maceió 57072-970, Brazil (e-mail: acfrery@laccan.ufal.br).

R. J. Cintra is with the Signal Processing Group, Departamento de Estatística, Universidade Federal de Pernambuco, Recife 50740-540, Brazil, and also with the University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: rjdsc@de.ufpe.br).

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

Digital Object Identifier 10.1109/TGRS.2018.2866367

In recent years, the interest in understanding such a type of imagery in a multidimensional and multilook perspective has increased. Such systems are called "polarimetric SAR" (PolSAR). In this case, obtaining of PolSAR data obeys the following dynamic: a scene is mapped with polarized pulses that are backscattered by the scene and captured by a sensor to form an image. As a result, PolSAR measurements record the amplitude and phase of backscattered signals for possible combinations of linear reception and transmission polarizations: HH, HV, VH, and VV (H for horizontal and V for vertical polarization).

However, since the acquired images stem from a coherent illumination process, they are affected by a signal-dependent granular noise called "speckle" [1]. Such a noise has a multiplicative nature and its intensity does not follow the Gaussian law. Thus, analyzing PolSAR images requires tailored image processing based on the statistical properties of speckled data.

The PolSAR theory prescribes that the returned (backscattered) signal of distributed targets is adequately represented by its complex covariance matrix. Under the assumption that the complex scattering coefficients are jointly circular Gaussian, the Wishart distribution is the statistical model for multilook PolSAR data. This paper adopts the assumption that a PolSAR image is well described by such a distribution.

Change detection methods aim at identifying differences in the scene configuration at distinct observation instants. Such procedures have achieved a prominent position in recent decades [2]. Indeed, literature reports several approaches for change detection problems, among them:

1) image rationing [3]–[6];
2) multitemporal coherence analysis [7];
3) spatiotemporal contextual classification [8], [9];
4) Hotelling–Lawley and likelihood ratio tests [10]–[19] and robust tests [20];
5) combination of image rationing and the generalized minimum-error method [21];
6) detection algorithms based on Lagrange optimization [22];
7) information-theoretic measures for change detection [9], [23]–[30];
8) change detection with postclassification [31].

This paper advances points 4) and 7) above.

The change detection process is theoretically rooted in the hypothesis test theory and the proposal of statistical similarity measures [32]. In particular, hypothesis tests based on the

complex covariance matrix have been sought for PolSAR data analysis. Many statistical approaches have been developed in order to reach this goal.

Conradsen *et al.* [11] proposed a methodology based on the likelihood ratio test defined by two random samples from the complex Wishart distribution. Subsequently, this technique was applied to edge detection in PolSAR images by Schou *et al.* [33]. Recently, Conradsen *et al.* [19] extended likelihood-based detection for PolSAR time series. Kersten *et al.* [34] compared three test statistics (the contrast ratio, ellipticity, and Bartlett tests). It was found that the method based on the contrast ratio is more robust to variations in the covariance estimates on actual data. In a complementary study, Molinier and Rauste [35] compared six polarimetric change detection methods. As a conclusion, the methods directly derived from the Wishart distribution outperformed other approaches as they provide explicit thresholds. Recently, Akbari *et al.* [36] proposed a change detector involving the Hotelling–Lawley trace (HLT), which, asymptotically, follows the Fisher–Snedecor distribution. The authors provided evidence that the HLT test may outperform the Bartlett test in some scenarios.

Several works have employed information-theoretic tools as a preprocessing step for change detection in PolSAR images. They can be categorized into two approaches: one is based only on discrimination measures, whereas the other considers the asymptotic distribution of such tools.

In the first category, Inglada and Mercier [23] proposed a new similarity measure for automatic change detection in multitemporal SAR images. Such a measure was derived considering the symmetrized Kullback–Leibler (KL) divergence (or distance) between the Edgeworth series expansions for two distinct elements of the $\mathcal{K}$ distribution from the Pearson system [37] for intensity SAR data. In [24], the KL measure is improved by means of copula-based quantile regression to generate local change measures. Furthermore, Erten *et al.* [38] proposed a new method based on mutual information for quantifying the coherent similarity between temporal multichannel PolSAR images. Atto *et al.* [9] used the KL divergence for spatio-temporal change detection in image time series.

In the second category, Nascimento *et al.* [25] derived hypothesis tests based on several distance measures between $\mathcal{G}^0$ distributions [39]. In terms of the nature of the image data, these results were extended in [26] and [27] and applied to boundary detection [40] and filtering [41] in PolSAR images. All these references derived new proposals using contrast measures designed from the scaled complex Wishart law. Recently, Akbari *et al.* [42] introduced a change detector with the HLT statistics as the contrast measure based on the relaxed scaled Wishart likelihood.

This paper proposes three new change detection methodologies for fully polarimetric data. Additionally, a new expression for the likelihood ratio statistics obtained from the scaled Wishart distribution is achieved, and its relationship with the individual distributions of the intensity channels is discussed. Using Monte Carlo simulation, we quantify the performance of four parametric methodologies for detecting the change: two considering Shannon and Rényi entropies, one stemming from

the KL distance, and one based on the classic likelihood ratio statistics. The methods are compared by their empirical test size (ETS) and power. Finally, two experiments with actual PolSAR data are performed. Results provide evidence that the methods based on entropies are superior.

This paper is organized as follows. Section II provides the background of the statistical modeling. A brief survey on parametric methodologies for hypothesis testing on complex covariance matrices is provided in Section III. In Section IV, we present a comparative study of change detection methods by means of Monte Carlo simulation. Additionally, we perform two experiments with actual PolSAR data. Section V summarizes the main results.

II. STATISTICAL MODELING FOR PolSAR DATA

PolSAR systems represent each resolution cell by p polarization elements comprising a complex random vector

$$\boldsymbol{y} = [S_1 \ S_2 \cdots \ S_p]^{\top} \tag{1}$$

where the superscript $^{\top}$ is the vector transposition. In single-look PolSAR image processing, $\boldsymbol{y}$ is admitted to obey the multivariate complex circular Gaussian distribution with zero mean [43] whose probability density function (pdf) is

$$f_{\boldsymbol{y}}(\dot{\boldsymbol{y}}; \boldsymbol{\Sigma}) = \frac{1}{\pi^p |\boldsymbol{\Sigma}|} \exp(-\dot{\boldsymbol{y}}^* \boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{y}})$$

where $\dot{\boldsymbol{y}}$ is an outcome of $\boldsymbol{y}$, $|\cdot|$ is the matrix determinant, the superscript * denotes the complex conjugate transpose of a vector, $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{y}$ such that $\boldsymbol{\Sigma} = \mathrm{E}\{\boldsymbol{y}\boldsymbol{y}^*\}$, and $\mathrm{E}\{\cdot\}$ is the statistical expectation operator. This distribution is denoted by $\boldsymbol{y} \sim \mathcal{N}^C(\boldsymbol{0}, \boldsymbol{\Sigma})$. Besides being Hermitian and positive-definite, $\boldsymbol{\Sigma}$ contains all the necessary information to characterize the backscattered data [44].

In order to improve the signal-to-noise ratio, L independent and identically distributed samples are usually averaged in order to form the L-look covariance matrix [45]

$$\boldsymbol{Z} = \frac{1}{L} \sum_{i=1}^{L} \boldsymbol{y}_i \boldsymbol{y}_i^*$$

where $\boldsymbol{y}_i$, $i = 1, 2, \ldots, L$ are the realizations of (1). Under the aforementioned hypotheses, $\boldsymbol{Z}$ follows a scaled complex Wishart distribution. Having $\boldsymbol{\Sigma}$ and L as parameters, such a law is characterized by the following pdf:

$$f_{\boldsymbol{Z}}(\dot{\boldsymbol{Z}}; \boldsymbol{\Sigma}, L) = \frac{L^{pL} |\dot{\boldsymbol{Z}}|^{L-p}}{|\boldsymbol{\Sigma}|^L \Gamma_p(L)} \exp[-L \operatorname{tr}(\boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{Z}})] \tag{2}$$

where $\Gamma_p(L) = \pi^{p(p-1)/2} \prod_{i=0}^{p-1} \Gamma(L-i)$, $L \geq p$, $\Gamma(\cdot)$ is the gamma function, and $\operatorname{tr}(\cdot)$ is the trace operator. We denote it by $\boldsymbol{Z} \sim \mathcal{W}(\boldsymbol{\Sigma}, L)$. This distribution satisfies $\mathrm{E}\{\boldsymbol{Z}\} = \boldsymbol{\Sigma}$, which is a Hermitian positive-definite matrix [45]. In practice, L is treated as a parameter and must be estimated. The resulting distribution is the relaxed Wishart distribution, and it is denoted by $\mathcal{W}_{\mathcal{R}}(\boldsymbol{\Sigma}, L)$ [46].

Due to its optimal asymptotic properties, we employ the maximum likelihood (ML) approach to estimate the parameters $\boldsymbol{\Sigma}$ and the equivalent number of looks L. Let $\underline{\boldsymbol{Z}} = \{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_N\}$ be a random sample of size N obtained
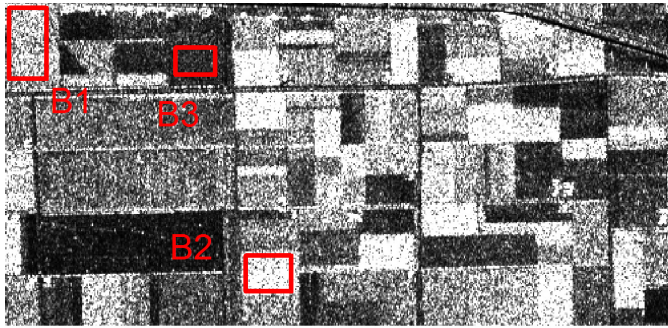
Fig. 1. AIRSAR image of Flevoland (channel HH).

from $\mathbf{Z} \sim \mathcal{W}_{\mathcal{R}}(\mathbf{\Sigma}, L)$. Setting $\ell_k(\boldsymbol{\theta}) = \log f_\mathbf{Z}(\mathbf{Z}_k; \mathbf{\Sigma}, L)$ for $\boldsymbol{\theta} = [\text{vec}(\mathbf{\Sigma})^\top, L]^\top$ as the log-likelihood of the kth random matrix, $\mathbf{Z}_k$, from $\underline{\mathbf{Z}}$, solving $N^{-1} \sum_{k=1}^{N} \nabla \ell_k(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, we have that $\widehat{\mathbf{\Sigma}} = N^{-1} \sum_{k=1}^{N} \mathbf{Z}_k$, and

$$p \log \widehat{L} + \frac{1}{N} \sum_{k=1}^{N} \log |\mathbf{Z}_k| - \log |\widehat{\mathbf{\Sigma}}| - \psi_p^{(0)}(\widehat{L}) = 0 \quad (3)$$

where $\text{vec}(\cdot)$ is the vectorization operator and $\psi_p^{(0)}(\cdot)$ is the zero-order term of the $v$thorder multivariate polygamma function

$$\psi_p^{(v)}(L) = \sum_{i=0}^{p-1} \psi^{(v)}(L-i)$$

and $\psi^{(v)}(\cdot)$ is the ordinary polygamma function expressed by

$$\psi^{(v)}(L) = \frac{\partial^{v+1} \log \Gamma(L)}{\partial L^{v+1}}$$

for $v \geq 0$; note that $\psi^{(0)}$ is the digamma function [47].

Thus, the ML estimator of $\mathbf{\Sigma}$ is the sample mean, while $\widehat{L}$ is obtained by solving the system shown in (3). We used the Newton–Raphson iterative method [48] to solve it. The work by Anfinsen *et al.* [45] is an important reference on how to efficiently estimate L.

Fig. 1 shows an area from the AIRSAR image of Flevoland, the Netherlands, obtained on August 1989 [49] with four nominal looks. We delimited three regions of interest.

Table I lists the ML parameter estimates as well as the sample sizes. Each sample is taken from a single class without evidence of texture. Notice that the estimates for the equivalent number of looks are very close, although lower than the nominal value. We also show the determinant of the estimated covariance matrix. This quantity, called *geometric intensity* in [50], is the *generalized variance* in multivariate analysis; it can be used as a measure of mean backscatter [51]. According to it, region B_2 presents the highest return, followed by B_1 and by B_3; this is in agreement with what is observed in channel HH (see Fig. 1).

Fig. 2 depicts the empirical densities of data from the agricultural regions along with the fitted marginal densities. The scaled Wishart density collapses to the Gamma density

$$f_{Z_i}(z_i; \theta_i, L) = \frac{L^L z_i^{L-1}}{\Gamma(L) \theta_i^L} \exp\left[-L \theta_i^{-1} z_i\right] \quad (4)$$

TABLE I
ESTIMATED PARAMETERS ON POLSAR DATA FROM FLEVOLAND

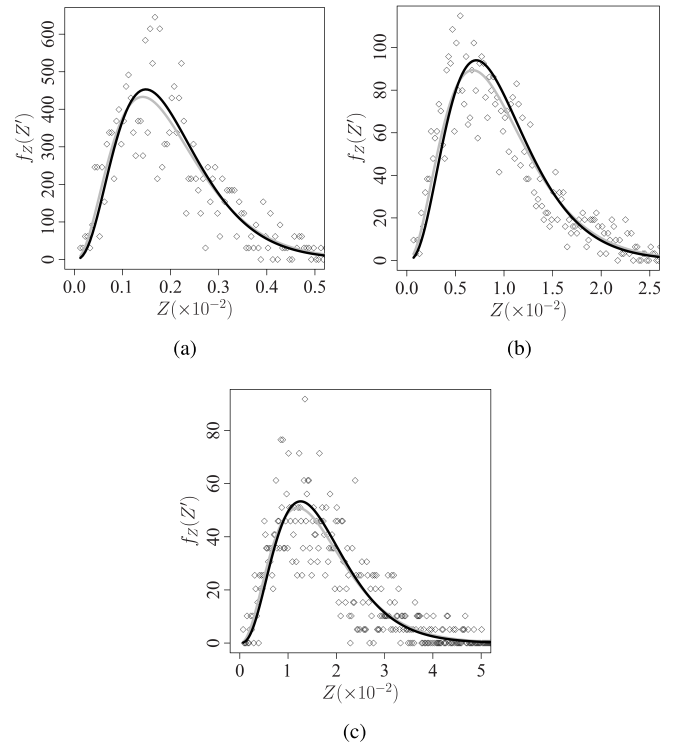| Regions	$\widehat{L}$	$|\widehat{\mathbf{\Sigma}}|$	# pixels
B_1	3.470	7.78×10^{-8}	1566
B_2	3.514	9.45×10^{-7}	980
B_3	3.530	7.22×10^{-10}	651

Fig. 2. Histograms of HH channel data and densities with an estimated number of looks (black curve) and fixed *a priori* (gray curve), respectively. (a) Region B_1. (b) Region B_2. (c) Region B_3.

where $i \in \{\text{HH, HV, VV}\}$, θ_k is the element (k, k) of $\mathbf{\Sigma}$, and Z_k is the (k, k)th entry of $\mathbf{Z}$. In practice, θ_i represents the mean polarization channel $i \in \{1(\text{HH}), 2(\text{HV}), 3(\text{VV})\}$. Fig. 2(a)–(c) shows the data and densities for the estimated number of looks $\mathcal{W}_{\mathcal{R}}(\widehat{\mathbf{\Sigma}}, \widehat{L})$ (black curve) and the fixed value $\mathcal{W}(\widehat{\mathbf{\Sigma}}, 4)$ (gray curve). These densities are remarkably close, and also to the histograms, so the Gamma assumption is reasonable.

According to Akbari *et al.* [52], if $\{\mathbf{Z}_i; i = 1, 2, \ldots, n\}$ is a random sample drawn from $\mathbf{Z} \sim \mathcal{W}(L, \mathbf{\Sigma})$ and $\widehat{\mathbf{\Sigma}}$ represents the ML estimator of $\mathbf{\Sigma}$, then $\text{tr}(\widehat{\mathbf{\Sigma}}^{-1} \mathbf{Z}_i)$ follows a Gamma distribution for $i = 1, 2, \ldots, n$. Fig. 3 displays fitted and empirical densities of such transformed data for the three selected regions. These results indicate that data may follow a scaled complex Wishart model. Additionally, the Kolmogorov–Smirnov statistic p-values for checking the adequacy of the Gamma model to the transformed data are 0.1377, 0.4923, and 0.3911 for regions B_1, B_2, and B_3, respectively.

We used likelihood ratio tests for two and three samples in order to quantify the similarity among these samples. The results presented in Table II point out that B_1 is different from B_2 and B_3, but these last two are similar. Although a visual
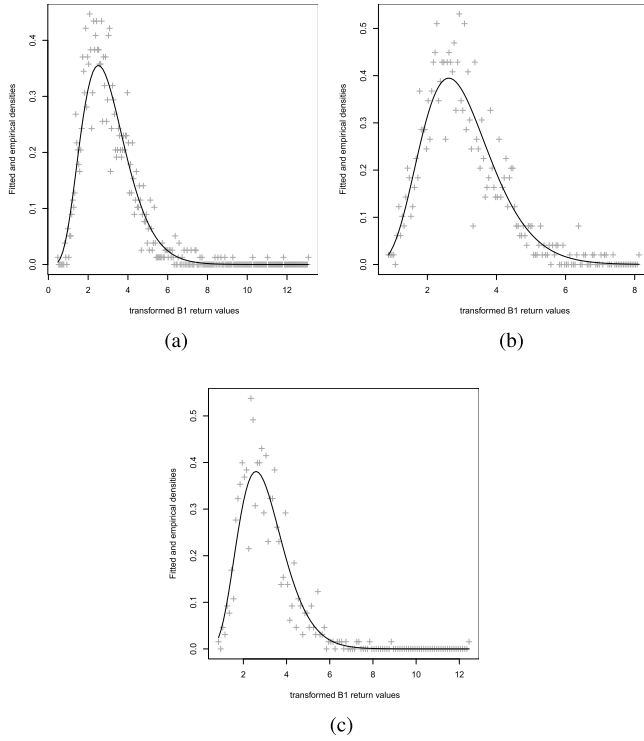
Fig. 3. Empirical (+) and fitted (solid curve) densities for transformed coherence matrices in selected regions. (a) Region B1. (b) Region B2. (c) Region B3.

TABLE II
HOMOGENEITY TEST AMONG CONSIDERED SAMPLES

$\mathcal{H}_0$	Statistics	p-value
$B1 = B2$	17.56	1.53×10^{-5}
$B1 = B3$	5.90	5.24×10^{-2}
$B2 = B3$	1.31	0.52
$B1 = B2 = B3$	19.00	7.85×10^{-4}

inspection of areas B_2 and B_3 (Fig. 1) suggests regions of different nature, their observations projected via $\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{Z}_i)$ are statistically similar.

These samples are used to validate our proposed methods in Section IV.

III. HYPOTHESIS TESTS IN POLSAR DATA: A SURVEY

This section provides a survey concerning three hypothesis tests that have been studied in the PolSAR data literature.

We assume that PolSAR data follow a scaled complex Wishart distribution. Change detection is often formulated as a statistical test for $\mathcal{H}_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, assuming L known.

The two main approaches in the literature are: 1) likelihood ratio [11] and 2) stochastic distances [27]. In this paper, the former proposal is extended to the context of scaled complex Wishart distributions, since the original approach used the nonscaled Wishart law. Moreover, this paper also introduces an alternative way for validating $\mathcal{H}_0$ by means of entropy measures [26]. Subsequently, these methodologies are introduced and discussed. In order to obtain more general results, we will provide expressions for testing $\mathcal{H}_0 : (\mathbf{\Sigma}_1, L_1) = (\mathbf{\Sigma}_2, L_2)$.

A. Likelihood Ratio Statistics

The log-likelihood ratio (LR) statistic has great importance in inference on parametric models. Let S_{LR} be the LR statistic for assessing the simple null hypothesis $\mathcal{H}_0$. As discussed in [53], such a statistic based on $\mathcal{H}_0$ has an asymptotic distribution χ_q^2, where q is the difference between the dimensions of the parameter spaces under the alternative and the null hypotheses. We denote such spaces by Θ_1 and Θ_0, respectively.

Let $\{X_1, X_2, \ldots, X_{N_1}\}$ and $\{Y_1, Y_2, \ldots, Y_{N_2}\}$ be two random samples from $\mathcal{W}_{\mathcal{R}}(\mathbf{\Sigma}_1, L_1)$ and $\mathcal{W}_{\mathcal{R}}(\mathbf{\Sigma}_2, L_2)$ of sizes N_1 and N_2, respectively. The LR statistic is given by

$$S_{\mathrm{LR}} = -2 \log \lambda_{\mathcal{W}_{\mathcal{R}}(\mathbf{\Sigma},L)}$$

where $\lambda_{\mathcal{W}_{\mathcal{R}}(\mathbf{\Sigma},L)} = \sup_{\boldsymbol{\theta} \in \Theta_0} \ell(\boldsymbol{\theta}) / \sup_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$, $\Theta = \Theta_0 \cup \Theta_1$ and $\Theta_0 \cap \Theta_1 = \varnothing$. Thus, we have that

$$
\begin{aligned}
\log &\lambda_{\mathcal{W}_{\mathcal{R}}(\mathbf{\Sigma},L)} \\
&= \mathcal{A}(p) + \log \frac{|\widehat{\mathbf{\Sigma}}_1|^{N_1 \widehat{L}_1} |\widehat{\mathbf{\Sigma}}_2|^{N_2 \widehat{L}_2}}{|\widehat{\mathbf{\Sigma}}_c|^{(N_1+N_2)\widehat{L}_c}} \\
&\quad + (\widehat{L}_c - \widehat{L}_1) \sum_{i=1}^{N_1} \log |X_i| + (\widehat{L}_c - \widehat{L}_2) \sum_{i=1}^{N_2} \log |Y_i| \\
&\quad + \sum_{i=1}^{N_1} \mathrm{tr} \left[\left(\widehat{L}_1 \widehat{\mathbf{\Sigma}}_1^{-1} - \widehat{L}_c \widehat{\mathbf{\Sigma}}_c^{-1} \right) X_i \right] \\
&\quad + \sum_{i=1}^{N_2} \mathrm{tr} \left[\left(\widehat{L}_2 \widehat{\mathbf{\Sigma}}_2^{-1} - \widehat{L}_c \widehat{\mathbf{\Sigma}}_c^{-1} \right) Y_i \right] \quad (5)
\end{aligned}
$$

and

$$\mathcal{A}(p) = p \log \frac{L_c^{L_c(N_1+N_2)}}{L_1^{L_1 N_1} L_2^{L_2 N_2}} + \log \frac{\Gamma_p(L_1)^{N_1} \Gamma_p(L_2)^{N_2}}{\Gamma_p(L_c)^{N_1+N_2}}$$

where L_c and $\mathbf{\Sigma}_c$ represent the number of looks and covariance matrix under the null hypothesis, respectively. Akbari *et al.* [42] discuss the two-sample LR test under the $\mathcal{W}_{\mathcal{R}}$ model.

Sections III-B and III-C discuss tests for $\mathcal{H}_0$ based on information-theoretic measures.

B. Kullback–Leibler Distance

The KL divergence (D_{KL}) is one of the oldest discrepancy measures between stochastic models; it has a central role in information theory [54]. This quantity was first understood as a measure of the error in choosing a model when another is the true one. It has been used in image processing for segmentation [55], classification [56], boundary detection [57], [58], and change detection [23]. Moreover, D_{KL} has a close relationship with the Neyman–Pearson lemma [54], and its symmetrization has been suggested as a correction form for another important goodness-of-fit measure for comparing statistical models: the Akaike information criterion [59].

Let X and Y be two random matrices defined over the common support $\mathcal{X}$ of positive-definite complex matrices of

size $p \times p$. The KL distance is defined by

$$
\begin{aligned}
d_{\text{KL}}(X, Y) &= \frac{1}{2}[D_{\text{KL}}(X, Y) + D_{\text{KL}}(Y, X)] \\
&= \frac{1}{2}\left[\int_{\mathcal{X}} f_X \log \frac{f_X}{f_Y} \mathrm{d}\dot{Z} + \int_{\mathcal{X}} f_Y \log \frac{f_Y}{f_X} \mathrm{d}\dot{Z}\right] \\
&= \frac{1}{2}\int_{\mathcal{X}} (f_X - f_Y) \log \frac{f_X}{f_Y} \mathrm{d}\dot{Z}
\end{aligned}
$$

with differential element $\mathrm{d}\dot{Z}$ given by

$$
\mathrm{d}\dot{Z} = \prod_{i=1}^{p} \mathrm{d}z_{ii} \underbrace{\prod_{\substack{i, j = 1}}^{p} \mathrm{d}\Re\{z_{ij}\}\mathrm{d}\Im\{z_{ij}\}}_{i<j}
$$

where z_{ij} is the (i, j)th entry of matrix $\dot{Z}$; and $\Re$ and $\Im$ denote the real and imaginary part operators, respectively [43].

When distances are taken between particular cases of the same distribution, only the parameters are relevant. In this case, the parameters θ_1 and θ_2 replace the random variables X and Y.

Salicrú et al. [60] proposed a hypothesis test based on d_{KL}. Let $\widehat{\theta}_1 = (\widehat{\theta}_{11}, \widehat{\theta}_{12}, \ldots, \widehat{\theta}_{1M})^{\top}$ and $\widehat{\theta}_2 = (\widehat{\theta}_{21}, \widehat{\theta}_{22}, \ldots, \widehat{\theta}_{2M})^{\top}$ be the ML estimators for θ_1 and θ_2 based on random samples of size N_1 and N_2, respectively. Under the regularity conditions discussed in [60, p. 380], the following lemma holds.

Lemma 1: If $(N_1/(N_1 + N_2)) \xrightarrow[N_1, N_2 \to \infty]{} \lambda \in (0, 1)$ and $\theta_1 = \theta_2$, then

$$
S_{\text{KL}}(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{2 N_1 N_Y}{N_1 + N_2} \frac{d_{\text{KL}}(\widehat{\theta}_1, \widehat{\theta}_2)}{h'(0)\phi''(1)} \xrightarrow[N_1, N_2 \to \infty]{\mathcal{D}} \chi_M^2 \quad (6)
$$

where "$\xrightarrow{\mathcal{D}}$" denotes convergence in distribution.

Proposition 1 is a test for the null hypothesis $\theta_1 = \theta_2$ based on Lemma 1.

Proposition 1: Let $S_{\text{KL}}(\widehat{\theta}_1, \widehat{\theta}_2) = s$ and $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be the ML estimates obtained from two sufficiently large random samples of sizes N_1 and N_2, respectively; then, the null hypothesis $\theta_1 = \theta_2$ can be rejected at level α if $\Pr(\chi_M^2 > s) \leq \alpha$.

Frery et al. [27] presented closed expressions for d_{KL} when the random matrices X and Y follow the Wishart distribution:

$$
\begin{aligned}
&d_{\text{KL}}(\theta_1, \theta_2) \\
&= \frac{L_1 - L_2}{2}\left\{\log \frac{|\Sigma_1|}{|\Sigma_2|} - p \log \frac{L_1}{L_2} + \psi_p^{(0)}(L_1) - \psi_p^{(0)}(L_2)\right\} \\
&\quad - \frac{p(L_1 + L_2)}{2} \\
&\quad + \frac{\mathrm{tr}\left(L_2\Sigma_2^{-1}\Sigma_1 + L_1\Sigma_1^{-1}\Sigma_2\right)}{2}
\end{aligned} \quad (7)
$$

from which the S_{KL} test statistic follows.

C. Shannon and Rényi Entropies

The Shannon entropy has achieved a prominent position in PolSAR imagery. Morio et al. [61] applied it for extracting features from polarimetric targets, assuming the circular Gaussian distribution. The Shannon entropy has also been used for classifying PolSAR textures [62], [63]. In the subsequent discussion, we present a comprehensive examination of hypothesis tests based on Shannon and Rényi entropies.

Let $f_Z(Z; \theta)$ be a pdf with a parameter vector θ. The Shannon and Rényi (with order β) entropies are defined, respectively, as

$$
\begin{aligned}
H_S(\theta) &= -\int_{\mathcal{X}} f_Z(\dot{Z}; \Sigma, L) \log f_Z(\dot{Z}; \Sigma, L) \mathrm{d}\dot{Z} \\
&= \mathrm{E}\{-\log f_Z(Z)\}
\end{aligned} \quad (8)
$$

and

$$
\begin{aligned}
H_R^{\beta}(\theta) &= (1 - \beta)^{-1} \log \int_{\mathcal{X}} f_Z^{\beta}(\dot{Z}; \Sigma, L)\mathrm{d}\dot{Z} \\
&= (1 - \beta)^{-1} \log \mathrm{E}\left\{f_Z^{\beta-1}(Z)\right\}.
\end{aligned} \quad (9)
$$

Pardo et al. [64] derived an important result, which paves the way for asymptotic statistical inference methods based on entropies.

Lemma 2: Let $\widehat{\theta} = [\widehat{\theta}_1 \ \widehat{\theta}_2 \ \cdots \ \widehat{\theta}_M]^{\top}$ be the ML estimate of the parameter vector $\theta = [\theta_1 \ \theta_2 \ \cdots \ \theta_M]^{\top}$ based on an N-point random sample from a model Z having pdf $f(\dot{Z}; \theta)$. Then

$$
\sqrt{N}\left[H_{\mathcal{M}}(\widehat{\theta}) - H_{\mathcal{M}}(\theta)\right] \xrightarrow[N \to \infty]{\mathcal{D}} \mathcal{N}(0, \sigma_{\mathcal{M}}^2(\theta))
$$

where $\mathcal{M} \in \{S, R\}$, $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2

$$
\sigma_H^2(\theta) = \delta^{\top}\mathcal{K}(\theta)^{-1}\delta \quad (10)
$$

where $\mathcal{K}(\theta) = \mathrm{E}\{-\partial^2 \log f_Z(Z; \theta)/\partial\theta^2\}$ is the Fisher information matrix and $\delta = [\delta_1 \ \delta_2 \ \cdots \ \delta_M]^{\top}$ such that $\delta_i = \partial H_{\mathcal{M}}(\theta)/\partial\theta_i$ for $i = 1, 2, \ldots, M$.

Now, we introduce a methodology for hypothesis tests and confidence intervals based on entropies. We aim at testing the following hypotheses:

$$
\begin{cases}
\mathcal{H}_0: & H_{\mathcal{M}}(\theta_1) = H_{\mathcal{M}}(\theta_2) = \upsilon \\
\mathcal{H}_1: & H_{\mathcal{M}}(\theta_1) \neq H_{\mathcal{M}}(\theta_2)
\end{cases}
$$

where $\mathcal{M} \in \{S, R\}$. In other words, is there any statistical evidence for rejecting the assumption that two PolSAR samples come from the same model?

Let $\widehat{\theta}_i$ be the ML estimate for θ_i based on a random sample of size N_i from Z_i for $i = 1, 2, \ldots, r$ and $r \geq 2$. From Lemma 2, we have that

$$
\sum_{i=1}^{r} \frac{N_i\left(H_{\mathcal{M}}(\widehat{\theta}_i) - \overline{\upsilon}\right)^2}{\sigma_{\mathcal{M}}^2(\widehat{\theta}_i)} \xrightarrow[N_i \to \infty]{\mathcal{D}} \chi_{r-1}^2
$$

where

$$
\overline{\upsilon} = \left[\sum_{i=1}^{r} \frac{N_i}{\sigma_{\mathcal{M}}^2(\widehat{\theta}_i)}\right]^{-1} \sum_{i=1}^{r} \frac{N_i H_{\mathcal{M}}(\widehat{\theta}_i)}{\sigma_{\mathcal{M}}^2(\widehat{\theta}_i)}.
$$

Then, we obtain the following test statistic:

$$
S_{\mathcal{M}}(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) = \sum_{i=1}^{r} \frac{N_i(H_{\mathcal{M}}(\widehat{\theta}_i) - \overline{\upsilon})^2}{\sigma_{\mathcal{M}}^2(\widehat{\theta}_i)} \quad (11)
$$

the expressions for $H_{\mathcal{M}}(\widehat{\theta}_i)$ and $\sigma^2_{\mathcal{M}}(\widehat{\theta}_i)$ are presented the Appendix. We are now in a position to state the following result.

Proposition 2: Let N_i, $i = 1, 2, \ldots, r$, be sufficiently large. If $S^h_\phi(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_r) = s$, then the null hypothesis $\mathcal{H}_0$ can be rejected at a level α if $\Pr(\chi^2_{r-1} > s) \leq \alpha$.

Whereas tests based on stochastic distances, such as d_{KL}, allow contrasting only two samples, those based on entropies permit assessing r samples at once [see (11)]. For issues involving more than two populations ($r > 2$ in (11)), this is a major advantage of the latter over the former. In the case of comparing two samples of the same size, i.e., $r = 2$ and $N_1 = N_2 = N$, (11) reduces to

$$S_{\mathcal{M}}(\widehat{\theta}_1, \widehat{\theta}_2) = N \frac{[H_{\mathcal{M}}(\widehat{\theta}_1) - H_{\mathcal{M}}(\widehat{\theta}_2)]^2}{\sigma^2_{\mathcal{M}}(\widehat{\theta}_1) + \sigma^2_{\mathcal{M}}(\widehat{\theta}_2)}.$$

IV. PERFORMANCE ANALYSIS

In this section, we assess the performance of the methodologies proposed with three experiments involving simulated (under the scaled Wishart complex law) and actual PolSAR data. First, we use Monte Carlo experiments to measure: 1) test size [false alarm (FA) rate] and 2) test power [$1 -$ false-negative (FN) rate]. For the test size, we check whether two samples from $X \sim W(B_1, 4)$ are from the same distribution, i.e., in a scenario where there was no change and there might be false positives (FPs). We assess the test power checking if two samples from $X \sim W(B_1, 4)$ and $X \sim W(B_1 \cdot (1 + k), 4)$, for $k = 0.2, 0.3, 0.4$, are correctly identified as a situation where there was a change. We then perform two experiments with actual PolSAR data.

A. Simulated Data

We compare the following hypothesis tests:

1) likelihood ratio S_{LR};
2) KL distance S_{KL};
3) statistics based on Shannon S_S and Rényi S_R^β entropies.

We fixed $\beta = 0.1$, since this value was found in [26] to provide good discrimination in hard-to-deal-with situations. We assume that the number of looks is known, as in [11] and [25]. Therefore, we are able to compare information-theoretic measures with the methodology proposed by Conradsen *et al.* [11].

The samples are generated according to Algorithm 1.

The parameters used for assessing the null hypothesis $\mathcal{H}_0$: $\Sigma_1 = \Sigma_2$ are $L_1 = L_2 = 4$, and (12), the sample covariance matrix of area B_1 (see Fig. 1). As we are interested in the behavior of the tests with small sample sizes, we computed the size of the hypothesis at $\alpha \in \{1\%, 5\%, 10\%\}$ for $N_1 = N_2 = N \in \{10, 11, \ldots, 50\}$.

Let T be the number of Monte Carlo replications and C the number of occurrences under $\mathcal{H}_0$ (i.e., pairs of samples are taken from the same model) on which the null hypothesis is rejected at the nominal level α. The ETS or FP rate is defined by $\alpha_{\text{ETS}} = C/T$. We used $T = 5500$, as suggested in [25], and α_{ETS} did not suffer expressive changes for larger values.

Algorithm 1 Sampling From the Scaled Complex Wishart Distribution

---

Require: Σ Hermitian positive definite $p \times p$ matrix
Require: $L \geq 3$ integer
1: Denote $R = \Re\{\Sigma\}$ and $I = \Im\{\Sigma\}$.
2: **for** $i = 1, 2, \ldots, L$ **do**
3:    Generate an outcome of the $2p$-variate Gaussian distribution $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip}, x_{i(p+1)}, \ldots, x_{i(2p)})^\top \sim \mathcal{N}_{2p}(0, \Sigma^*)$, where

$$\Sigma^* = \frac{1}{2}\begin{bmatrix} R & -I \\ I & R \end{bmatrix}.$$

4:    Set the random vector

$$y_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top + j \ (x_{i(p+1)}, \ldots, x_{i(2p)})^\top.$$

   With this, y_i is a p-variate outcome of the complex Gaussian distribution $\mathcal{N}_p^C(0, \Sigma)$.
5: **end for**
6: Return $L^{-1} \sum_{i=1}^{L} y_i y_i^*$, outcome of $\mathcal{W}(\Sigma, L)$, the scaled complex Wishart distribution.

TABLE III
ESTIMATED TEST SIZES (FP RATES)

N	Mean values			
	1%	5%	10%	$\overline{S}_\bullet$
	Likelihood ratio (S_{LR})			
$10 - 20$	1.21	5.76	11.16	9.25
$21 - 30$	1.10	5.42	10.66	9.12
$31 - 40$	1.03	5.13	10.43	9.09
$41 - 50$	1.06	5.21	10.28	9.08
	Shannon Entropy (S_S)			
	1.00	4.59	9.47	1.00
	0.99	4.64	9.21	1.02
	0.99	4.44	9.24	1.05
	0.93	4.54	9.37	1.07
	Rényi Entropy ($S_R^{0.1}$)			
	0.00	1.71	4.53	0.71
	0.33	1.83	4.53	0.74
	0.32	1.79	4.46	0.77
	0.29	1.78	4.56	0.79
	Kullback-Leibler Distance (S_{KL})			
	1.83	7.06	12.89	9.53
	1.43	6.16	11.66	9.27
	1.24	5.67	11.17	9.20
	1.24	5.55	10.85	9.16

Table III shows: 1) the ETS at nominal levels 1%, 5%, and 10% and 2) the mean test statistic ($\overline{S}_\bullet$) of the four statistics.

On an average, all test statistics behave as expected when the sample sizes increase: $\overline{S}_S$ and $\overline{S}_R^{0.1}$ tend to one, while $\overline{S}_{\text{LR}}$ and $\overline{S}_{\text{KL}}$ tend to nine. Recall that the asymptotic distribution of the two former is χ^2_9, while the two latter are χ^2_1.

The S_{LR} and S_S tests exhibit the closest empirical sizes to the nominal levels, as confirmed by Fig. 4. The ETS associated with $S_R^{0.1}$ and S_{KL} are biased; however, the bias reduces as the sample size increases. We conclude that these two statistics require larger sample sizes to achieve the expected asymptotic behavior.
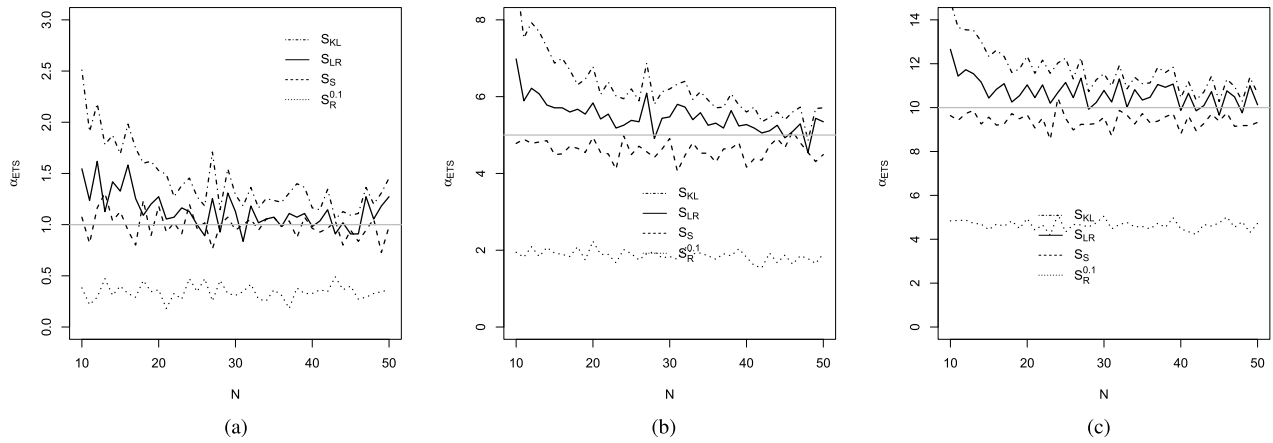
Fig. 4. Values for α_{ETS} sizes on synthetic data for different scenarios at the levels (a) 1%, (b) 5%, and (c) 10%.

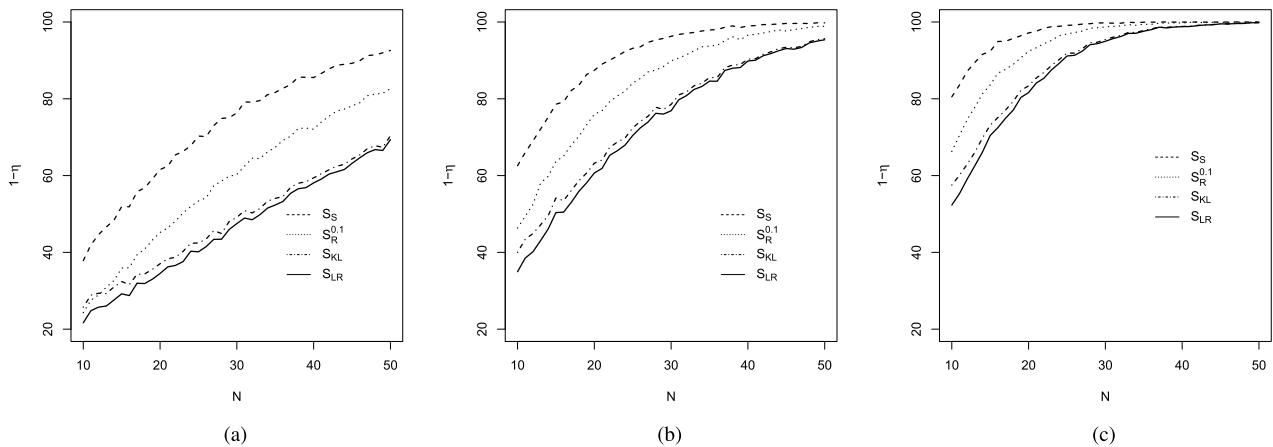Fig. 5. Estimated test powers for several scenarios at the level 1%. (a) $(B_1, 4)$ versus $(B_1 \cdot (1 + 0.2), 4)$. (b) $(B_1, 4)$ versus $(B_1 \cdot (1 + 0.3), 4)$. (c) $(B_1, 4)$ versus $(B_1 \cdot (1 + 0.4), 4)$.

In general terms, Table III suggests this inequality

$$\text{ETS}_{S_{\text{KL}}} \geq \text{ETS}_{S_{\text{LR}}} \geq \text{ETS}_{S_S} \geq \text{ETS}_{S_R^{0.1}}. \tag{13}$$

The size of tests (FP rates) based on the Shannon entropy and likelihood ratio are the closest to the nominal level.

We also studied the test power. We wish to reject the hypothesis $\mathcal{H}_0$, given two samples drawn from $X \sim \mathcal{W}(B_1, 4)$ and $Y \sim \mathcal{W}(B_1 \cdot (1 + k), 4)$ where $k = 0.2, 0.3, 0.4$, i.e., under $\mathcal{H}_1$. The rate $\eta = (T - C^*)/T$, where C^* is the number of rejections of $\mathcal{H}_0$ under $\mathcal{H}_1$ and estimates the Type-II error or FN [19], and we aim at quantifying the test power $1 - \eta$.

Fig. 5 presents the estimated power for several samples sizes. The test based on Shannon entropy performs the best. In this case, we obtain the inequality

$$(1 - \eta)_{S_S} \geq (1 - \eta)_{S_R^{0.1}} \geq (1 - \eta)_{S_{\text{KL}}} \geq (1 - \eta)_{S_{\text{LR}}}.$$

The relation between discriminatory powers within groups $\{S_{\text{KL}}, S_{\text{LR}}\}$, and $\{S_S, S_R^\beta\}$ has been discussed in the statistical literature. This fact can be explained twofold, namely: 1) the relationship between the Neyman and Pearson lemma and the KL distance [54] and 2) the fact that $\lim_{\beta \to 1} S_R^\beta = S_S$ [65].
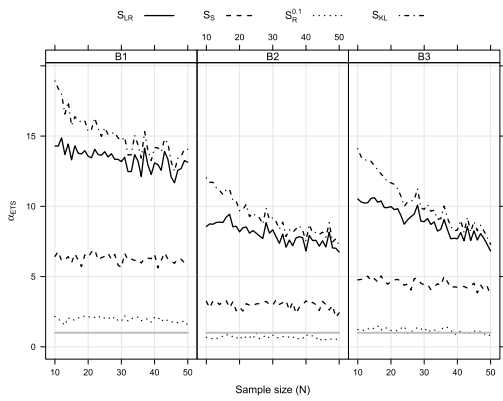
The best test statistics should have both empirical size near to the nominal level and the highest estimated power. Thus, based on this evidence and on the estimated size, we suggest S_S as the best discriminator on scenarios that follow the scaled complex Wishart distribution.
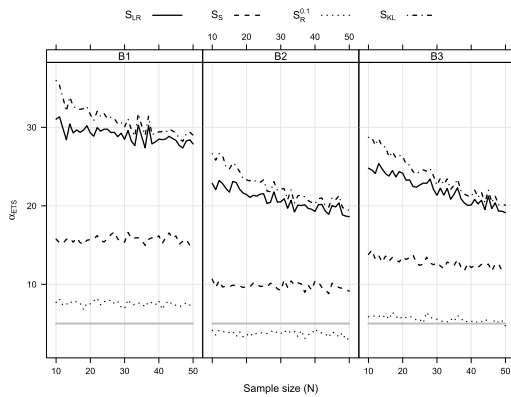
B. Experiments With Data From Sensors

In this section, we apply the proposed test statistics to two studies: 1) for assessing $\mathcal{H}_0 : \Sigma_1 = \Sigma_2$ to the data presented in Fig. 1 (single date) and 2) for detecting changes on two PolSAR images captured at different instants, as displayed in Fig. 8 (multitemporal data). ENL is assumed constant.

1) Single-Date Experiment: Our first experiment aims at assessing α_{ETS}, the ETS (Type-I error or the probability of FA), using pairs of disjoint samples from the same target.
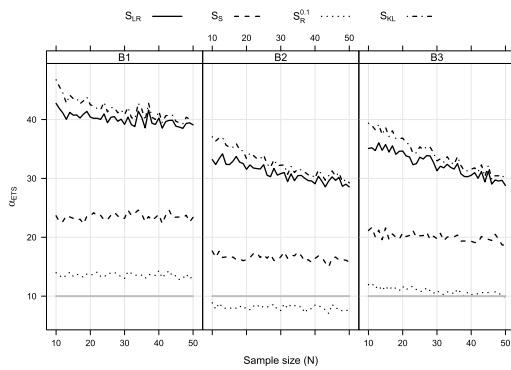
$$B_1 = \begin{bmatrix} 9.528 \times 10^{-3} & (-3.469 + 1.048\,\mathbf{j}) \times 10^{-4} & (1.439 + 1.164\mathbf{j}) \times 10^{-3} \\ & 1.794 \times 10^{-3} & (8.551 - 1.608\mathbf{j}) \times 10^{-5} \\ & & 4.955 \times 10^{-3} \end{bmatrix} \tag{12}$$

(a)

(b)

(c)

Fig. 6. ETS for actual data at levels (a) 1%, (b) 5%, and (c) 10%.

Fig. 7. Images from the study areas: Los Angeles, California. (a) First scene. (b) Second scene.
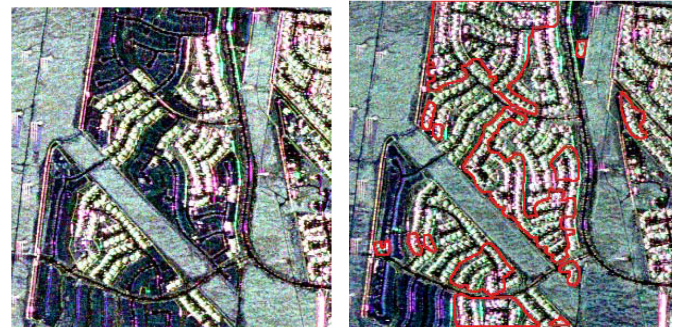
(a)   (b)
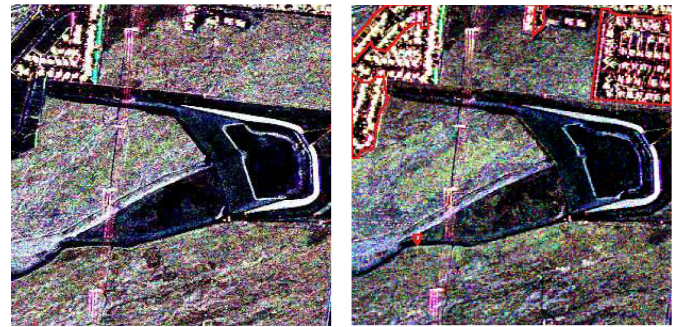
(c)   (d)

(e)   (f)

Fig. 8. UAVSAR images (in Pauli decomposition) on April 23, 2009 and May 11, 2015. (a) Scene 1 (before). (b) Scene 1 (after). (c) Scene 2 (before). (d) Scene 2 (after). (e) Scene 1 (reference map). (f) Scene 2 (reference map).

This experiment is outlined in Algorithm 2. We used samples of size $N \in \{3 \times 3, 4 \times 4, \ldots, 23 \times 23\}$.

Fig. 6 shows the observed α_{ETS}. Inequality (13) is also verified on actual data. For $\alpha = 1\%$, $11.67\% \leq \alpha_{\mathrm{ETS}}(S_{\mathrm{LR}}) < \alpha_{\mathrm{ETS}}(S_{\mathrm{KL}}) \leq 18.93\%$ and $1.527\% \leq \alpha_{\mathrm{ETS}}(S_S) < \alpha_{\mathrm{ETS}}(S_R^{0.1}) \leq 6.909\%$, i.e., all tests overestimate α, but S_S and $S_R^{0.1}$ presented better results than S_{LR} and S_{KL}. For B_2 and B_3, $\alpha_{\mathrm{ETS}}(S_R^{0.1}) \leq 1.49\%$, overcoming $2.091\% \leq \alpha_{\mathrm{ETS}}(S_S) \leq 5.055\%$, $6.745\% \leq \alpha_{\mathrm{ETS}}(S_{\mathrm{LR}}) \leq 10.618\%$, and $7.164\% \leq \alpha_{\mathrm{ETS}}(S_{\mathrm{KL}}) \leq 14.109\%$.

PolSAR regions are Wishart; our explanation for the better performance of $S_R^{0.1}$ is deviations from this hypothesis.

These results present evidence that the test statistics based on $S_R^{0.1}$ outperforms the other ones. This test presented good results even for small samples. Thus, this measure is suggested as a relevant change detection tool for PolSAR imagery.

2) Multitemporal Data: Fig. 7 presents the study areas for this experiment: surroundings of the city of Los Angeles, CA, USA. These pictures refer to a dense urban area whose changes are caused by the urbanization process.
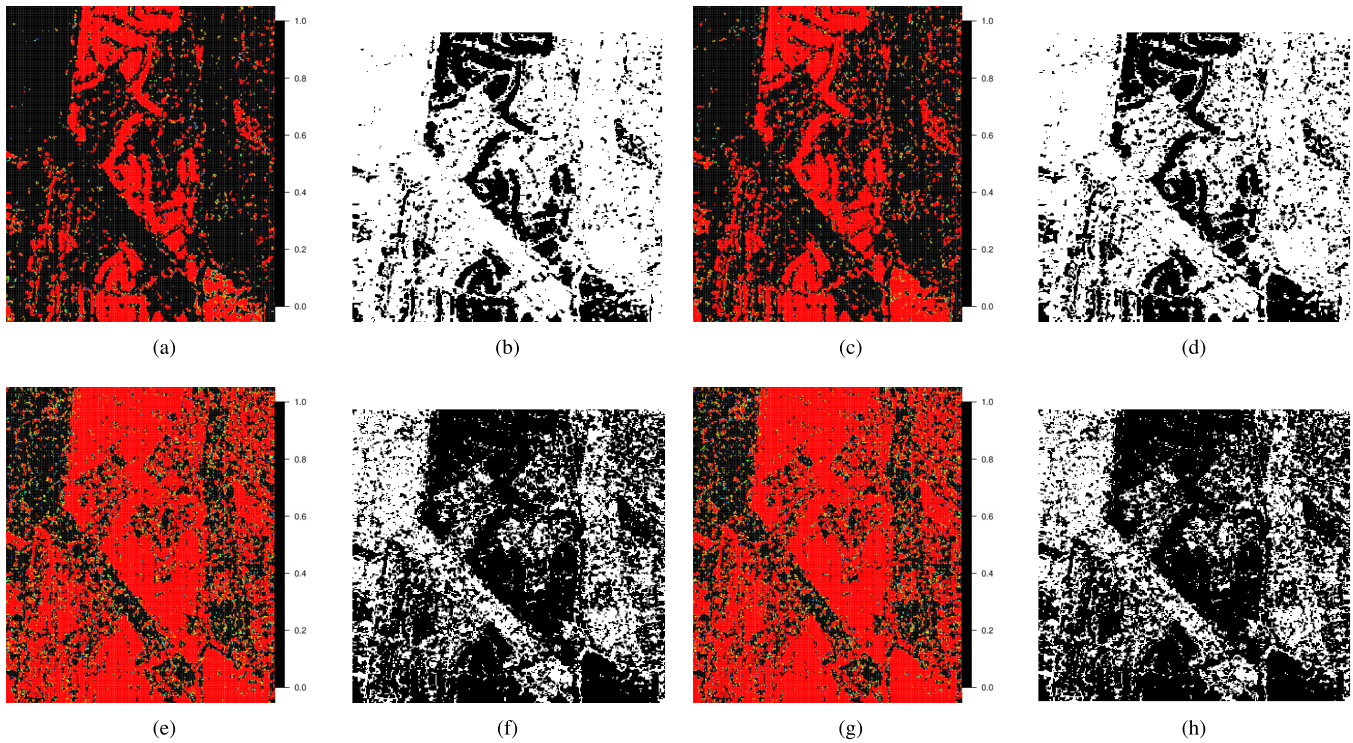
Fig. 9. *p*-value maps as evidence of changes between two dates for the first scene. (a) $S_R^{0.1}$-3 × 3. (b) $S_R^{0.1}$-3 × 3. (c) S_S-3 × 3. (d) S_S-3 × 3. (e) S_{LR}-3 × 3. (f) S_{LR}-3 × 3. (g) S_{KL}-3 × 3. (h) S_{KL}-3 × 3.

Algorithm 2 Experiment Design for Data From the Same Target

1: **for** $j = 1, 2, \ldots, 5500$ **do**
2:     Extract two disjoint regions U_j and V_j from areas $B_1, B_2,$ and B_3.
3:     Generate two vectors of size N, $\boldsymbol{u}^{(j)}$ and $\boldsymbol{v}^{(j)}$ from U_j and V_j, respectively, sampling without replacement.
4:     Estimate $\widehat{\boldsymbol{\theta}}_1^{(j)}$ and $\widehat{\boldsymbol{\theta}}_2^{(j)}$ based on $\boldsymbol{u}^{(j)}$ and $\boldsymbol{v}^{(j)}$, respectively.
5:     Compute the decision from Propositions 1 and 2, and execute the test based on S_{LR} for $\alpha = \{1\%, 5\%, 10\%\}$.
6: **end for**
7: Let T be the number of times that the null hypothesis is rejected. Calculate the ETS ($\widehat{\alpha}_{1-\alpha}$) at level α as

$$\alpha_{\text{ETS}} = T/5500, \text{ if } \boldsymbol{V}_j = \boldsymbol{U}_j.$$

Ratha *et al.* [30] employed these data in the proposal of change detectors for single-look polarimetric data using a geodesic distance. Here, we apply the four multilook PolSAR data detectors discussed in Section III. Fig. 8 shows the Pauli decomposition of two *uninhabited aerial vehicle synthetic aperture radar* (UAVSAR) images obtained by the JPL's UAVSAR sensor at two different instants (April 23, 2009 and May 3, 2015).

Using windows of size 3×3 on both the dates, we computed the S_{LR}, S_{KL}, S_S, and $S_R^{0.1}$ test statistics, and from them, *p*-value maps [see Figs. 9 and 10]. Probability values higher than 0.01% are drawn in black, as they provide no evidence of change. Values below 0.01% range vary from red to dark blue (from strong to weak evidence of change).

It is noticeable that $S_R^{0.1}$ and S_S are similar [see Figs. 9(a) and (c) and 10(a) and (c)], while S_{LR} and S_{KL} look alike [see Figs. 9(e) and (g) and 10(e) and (g)], but somewhat different from the previous pair.

Fig. 11 shows the relationship between $S_R^{0.1}$ and S_S for the second scene, along with the identity function for reference. The *p*-values associated with the Shannon statistic are smaller than that those related to the Rényi statistic, so the former tends to reject more than $S_R^{0.1}$, as discussed in the simulation experiments.

Finally, Figs. 9(b), (d), (f), and (h) and 10(b), (d), (f), and (h) show binary images resulting from thresholding the $S_R^{0.1}$, S_S, S_{LR}, and S_{KL} statistics for the first and second scenes: *p*-values larger than 10^{-4} are shown in white, otherwise in black. The results, again, favor entropy-based detectors.

To confirm the qualitative discussion, we quantify the performance of detectors with respect to reference maps in Fig. 8(e) and (f) in terms of five measures.

1) *False Positive:* Number of pixels indicated as change by reference map (RM), but classified as no change.
2) *False Negative:* Number of pixels indicated as no change by RM, but classified as change.
3) *False Alarm Rate:* (FP+FN)/N, where N is the number of unchanged pixels according to the detector.
4) *Detection Rate:* TP/CG, where TP is the number of pixels indicated as change by both the RM and the detector, and CG is the number of changed pixels according to the detector.
5) *Kappa Coefficient:* $\kappa = (A - B)/(1 - B)$, where $A = 1 - p_{\text{FP}} - p_{\text{FN}}$ and $B = (p_{\text{TP}} + p_{\text{FP}})(p_{\text{TP}} + p_{\text{FN}}) + (p_{\text{TN}} + p_{\text{FP}})(p_{\text{TN}} + p_{\text{FN}})$, where $p_{\mathcal{C}}$ is the proportion of
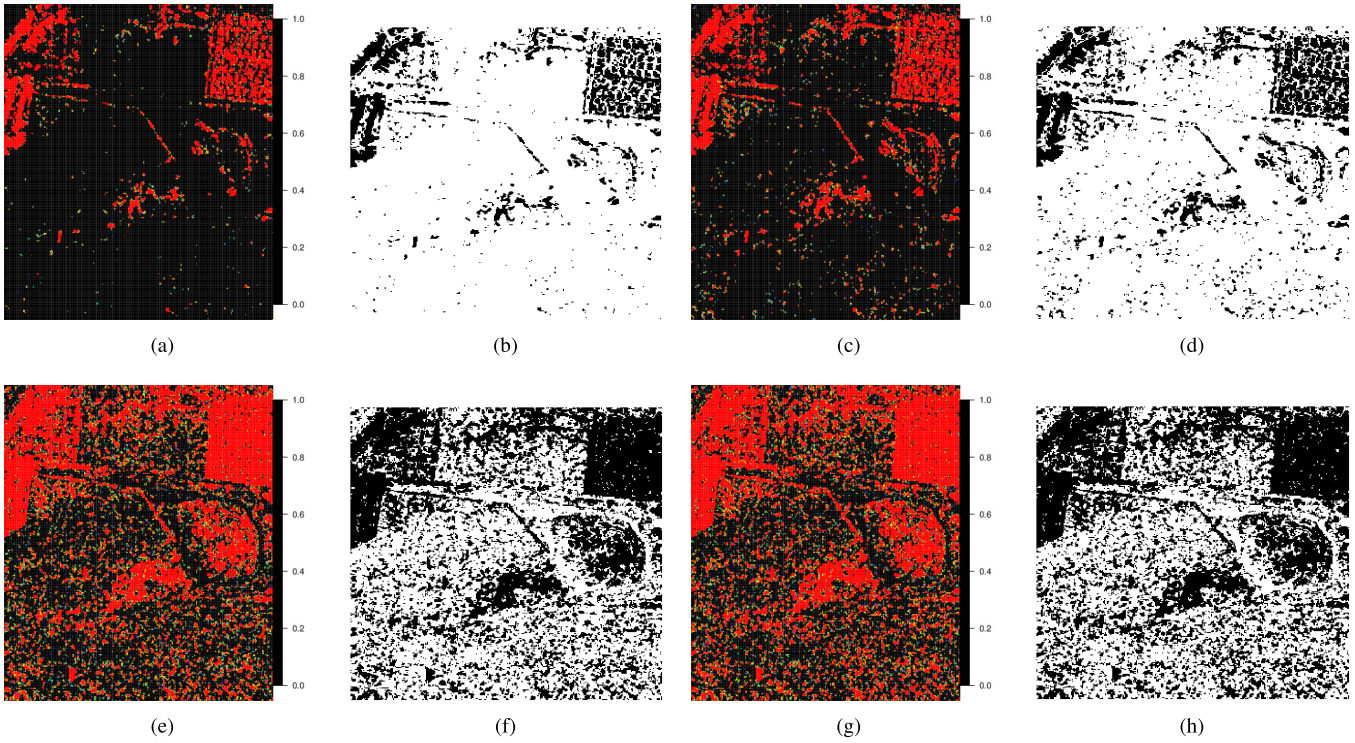
Fig. 10. *p*-value maps as evidence of changes between two dates for the second scene. (a) $S_R^{0.1}$-3 × 3. (b) $S_R^{0.1}$-3 × 3. (c) S_S-3 × 3. (d) S_S-3 × 3. (e) S_{LR}-3 × 3. (f) S_{LR}-3 × 3. (g) S_{KL}-3 × 3. (h) S_{KL}-3 × 3.
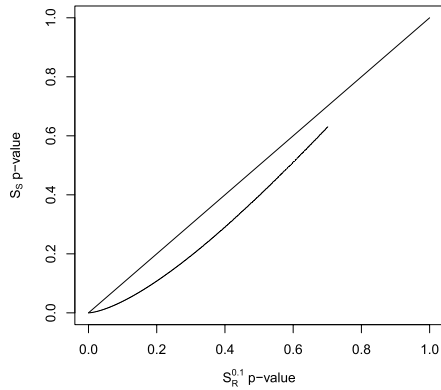

Fig. 11. Relationship between *p*-values from $S_R^{0.1}$ and S_S.

TABLE IV
DETECTORS' PERFORMANCE

Detectors	FP (%)	FN (%)	FA (%)	DR(%)	κ (%)
			Scene 1		
S_{LR}	0.060	13.433	13.493	20.408	24.597
S_{KL}	0.052	14.055	14.107	19.765	23.023
S_S	0.343	5.476	5.819	35.387	52.482
S_R	0.431	3.709	4.140	42.988	62.272
			Scene 2		
S_{LR}	0.104	9.551	9.655	20.964	25.552
S_{KL}	0.094	10.404	10.497	19.682	23.087
S_S	0.739	2.568	3.307	41.598	52.598
S_R	0.920	1.254	2.174	55.590	62.519

pixels under the condition $\mathcal{C}$ relative to the total number of pixels and TN is the number of pixels indicated as no change by both RM and the detector.

The reference maps were prepared by specialists with Bing and Google Earth imagery [30].

Table IV shows the results. S_R obtained the best performance, followed by S_S, for both the data sets with respect to κ and detection rate (DR). These detectors presented lower FN and FA than S_{LR} and S_{KL}. S_{KL} and S_{LR} performed better than entropy-based detectors with respect to FP. The values of FP were smaller than 5% in all the cases, so this is not an issue for any detector.

V. CONCLUSION

We quantified and compared the performance of four change detection methods for fully PolSAR data. These methods are based on the likelihood-ratio statistic, on the KL distance, and on the Rényi and Shannon entropies. We used empirical test powers and ETSs as comparison criteria.

First, the performance of the methods was quantified through a Monte Carlo study using scenarios modeled by the scaled complex Wishart law. The ETS s showed evidence that the detectors based on the likelihood ratio and Shannon entropy statistics presented the best performance. In particular, the one based on the entropy is the best for small samples and statistically similar to the S_{LR}. In addition, the tests based on the KL and on the likelihood ratio statistics tend to overestimate the nominal level, while those which employ entropies underestimate it.

Regarding the empirical test power, the test based on the Shannon entropy presented, in a consistent fashion, the best results. Computational costs are quite different. The test statistic based on the likelihood ratio S_{LR} requires evaluating (5), while S_{KL} depends only on the KL distance (7). The latter

is less demanding than the former by an order of magnitude. Thus, on those situations in which S_{KL} and S_{LR} are competitive (for moderate and large sample sizes), the KL test is more attractive, because it has the lowest computational cost.

Second, and since estimated test sizes were quite competitive, two experiments with actual data were performed. For the single-date experiment, in all the situations considered, the test based on the Rényi entropy with order $\beta = 0.1$ presented the best results. The multitemporal data experiments suggest that change detectors equipped by entropies provide better performance than those based on the KL distance and those based on the likelihood ratio statistic. Finally, the diversity of tests statistics stemming from Information Theory opens the venue for investigation of composite decision rules, as in [66].

Future works will aim to adapt developments made in this paper to more general distributions as, for instance, the $\mathcal{G}_{Pol}$ and its particular cases ($\mathcal{K}_{Pol}$, $\mathcal{G}_{Pol}^0$, and $\mathcal{G}_{Pol}^H$), the Kummer-$\mathcal{U}$, and $\mathcal{M}$ laws (see [67, Sec. 4.1], [57]).

APPENDIX

Applying (2) in (8) and (9), we obtain the following entropies [26]:

$$H_S(\boldsymbol{\theta}) = \frac{p(p-1)}{2} \log \pi - p^2 \log L + p \log |\boldsymbol{\Sigma}| + pL$$
$$+ (p-L)\psi_p^{(0)}(L) + \sum_{k=0}^{p-1} \log \Gamma(L-k) \quad (14)$$

and

$$H_R^\beta(\boldsymbol{\theta}) = \frac{p(p-1)}{2} \log \pi - p^2 \log L + p \log |\boldsymbol{\Sigma}|$$
$$- \frac{pq \log \beta}{1-\beta} + \frac{\sum_{i=0}^{p-1}[\log \Gamma(q-i) - \beta \log \Gamma(L-i)]}{1-\beta} \quad (15)$$

where $q = L + (1-\beta)(p-L)$.

Under the scaled complex Wishart law, Frery *et al.* [26] derived the following variances:

1) *Shannon:*

$$\sigma_S^2 = \frac{\left[(p-L)\psi_p^{(1)}(L) + p - \frac{p^2}{L}\right]^2}{\psi_p^{(1)}(L) - \frac{p}{L}}$$
$$+ \frac{p^2}{L} \text{vec}(\boldsymbol{\Sigma}^{-1})^*(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma}^{-1}). \quad (16)$$

2) *Rényi Entropy:*

$$\sigma_{R,\beta}^2 = \frac{\left\{\frac{\beta}{1-\beta}\left[\psi_p^{(0)}(q) - \psi_p^{(0)}(L)\right] - \frac{p\beta \ln(\beta)}{1-\beta} - \frac{p^2}{L}\right\}^2}{\psi_p^{(1)}(L) - \frac{p}{L}}$$
$$+ \frac{p^2}{L} \text{vec}(\boldsymbol{\Sigma}^{-1})^*(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma}^{-1}). \quad (17)$$

REFERENCES

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

[2] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Apr. 2004.

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

[4] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images* (The SciTech Radar and Defense Series). Rijeka, Croatia: SciTech, 1998.

[5] G. Liu, L. Jiao, F. Liu, H. Zhong, and S. Wang, "A new patch based change detector for polarimetric SAR data," *Pattern Recognit.*, vol. 48, no. 3, pp. 685–695, Mar. 2015.

[6] G. Moser, S. Serpico, and G. Vernazza, "Unsupervised change detection from multichannel SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 278–282, Apr. 2007.

[7] T. T. Le, A. M. Atto, E. Trouvé, A. Solikhin, and V. Pinel, "Change detection matrix for multitemporal filtering and change analysis of SAR and PolSAR image time series," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 64–76, Sep. 2015.

[8] F. Melgani and S. B. Serpico, "A statistical approach to the fusion of spectral and spatio-temporal contextual information for the classification of remote-sensing images," *Pattern Recognit. Lett.*, vol. 23, no. 9, pp. 1053–1061, Jul. 2002.

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

[10] M. F. da Silva, F. Cribari-Neto, and A. C. Frery, "Improved likelihood inference for the roughness parameter of the GA0 distribution," *Environmetrics*, vol. 19, no. 4, pp. 347–368, Aug. 2008.

[11] 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. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 4–19, Jan. 2003.

[12] V. Akbari, S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "The Hotelling-Lawley trace statistic for change detection in polarimetric SAR data under the complex Wishart distribution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 4162–4165.

[13] M. Liu, H. Zhang, C. Wang, and F. Wu, "Change detection of multilook polarimetric SAR images using heterogeneous clutter models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7483–7494, Dec. 2014.

[14] V. Carotenuto, A. D. Maio, C. Clemente, and J. J. Soraghan, "Invariant rules for multipolarization SAR change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3294–3311, Jun. 2015.

[15] V. Carotenuto, A. De Maio, C. Clemente, J. J. Soraghan, and G. Alfano, "Forcing scale invariance in multipolarization SAR change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 36–50, Jan. 2016.

[16] V. Carotenuto, A. De Maio, C. Clemente, and J. Soraghan, "Unstructured versus structured GLRT for multipolarization SAR change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1665–1669, Aug. 2015.

[17] M. Liu *et al.*, "Polarimetric synthetic aperture radar change detection for specific land cover types," *Int. J. Digit. Earth*, vol. 8, no. 4, pp. 334–344, Apr. 2015.

[18] 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.

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

[20] K.-F. Lin and D. Perissin, "Identification of statistically homogeneous pixels based on one-sample test," *Remote Sens.*, vol. 37, no. 1, pp. 1–13, Jan. 2017.

[21] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, Oct. 2006.

[22] A. Marino and I. Hajnsek, "A change detector based on an optimization with polarimetric SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4781–4798, Aug. 2014.

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

[24] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.

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

[26] A. C. Frery, R. J. Cintra, and A. D. C. Nascimento, "Entropy-based statistical analysis of PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3733–3743, Jun. 2013.

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

[28] A. Marino, S. R. Cloude, and J. M. López-Sanchez, "A new polarimetric change detector in radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2986–3000, May 2013.

[29] Y. Zheng, X. Zhang, B. Hou, and G. Liu, "Using combined difference image and k-means clustering for SAR image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 691–695, Mar. 2014.

[30] D. Ratha, S. De, T. Celik, and A. Bhattacharya, "Change detection in polarimetric SAR images using a geodesic distance between scattering mechanisms," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1066–1070, Jul. 2017.

[31] V. Akbari, A. P. Doulgeris, and T. Eltoft, "Monitoring glacier changes using multitemporal multipolarization SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3729–3741, Jun. 2014.

[32] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[33] J. Schou, H. Skriver, A. A. Nielsen, and K. Conradsen, "CFAR edge detector for polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 20–32, Jan. 2003.

[34] P. R. Kersten, J. S. Lee, and T. L. Ainsworth, "A comparison of change detection statistics in POLSAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2005, pp. 4836–4839.

[35] M. Molinier and Y. Rauste, "Comparison and evaluation of polarimetric change detection techniques in aerial SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2007, pp. 2386–2389.

[36] 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. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3953–3966, Jul. 2016.

[37] K. D. Ward, "Compound representation of high resolution sea clutter," *Electron. Lett.*, vol. 17, no. 16, pp. 561–563, Aug. 1981.

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

[39] A. C. Frery, H.-J. Müller, C. da Costa Freitas Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.

[40] A. D. C. Nascimento, M. M. Horta, A. C. Frery, and R. J. Cintra, "Comparing edge detection methods based on stochastic entropies and distances for PolSAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 648–663, Feb. 2014.

[41] L. Torres, S. J. S. Sant'Anna, C. C. Freitas, and A. C. Frery, "Speckle reduction in polarimetric SAR imagery with stochastic distances and nonlocal means," *Pattern Recognit.*, vol. 47, no. 1, pp. 141–157, Apr. 2014.

[42] V. Akbari, S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "A change detector for polarimetric SAR data based on the relaxed Wishart distribution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 3806–3809.

[43] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, 1963.

[44] C. López-Martínez, X. Fábregas, and E. Pottier, "Multidimensional speckle noise model," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 20, pp. 3259–3271, Dec. 2005.

[45] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3795–3809, Nov. 2009.

[46] S. N. Anfinsen, T. Eltoft, and A. P. Doulgeris, "A relaxed Wishart model for polarimetric SAR data," in *Proc. 4th Int. Workshop Sci. Appl. SAR Polarimetry Polarimetric Interferometry (POLinSAR)*, Frascati, Italy, 2009, p. 42.

[47] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover, 1964.

[48] J. E. Gentle, *Elements of Computational Statistics* (Statistics and Computing). New York, NY, USA: Springer, 2002.

[49] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3665–3676, Oct. 2011.

[50] S. Skrunes, C. Brekke, and T. Eltoft, "Characterization of marine surface slicks by radarsat-2 multipolarization features," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5302–5319, Sep. 2014.

[51] N. R. Goodman, "The distribution of the determinant of a complex Wishart distributed matrix," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 178–180, 1963.

[52] V. Akbari, A. P. Doulgeri, G. Moser, T. Eltoft, S. N. Anfinsen, and S. B. Serpico, "A textural–contextual model for unsupervised segmentation of multipolarization synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2442–2453, Apr. 2013.

[53] G. Casella and R. L. Berger, *Statistical Inference*. Boston, MA, USA: Cengage Learning, 2002.

[54] S. Eguchi and J. Copas, "Interpreting Kullback–Leibler divergence with the neyman–pearson lemma," *J. Multivariate Anal.*, vol. 97, no. 9, pp. 2034–2040, Oct. 2006.

[55] J. M. Beaulieu and R. Touzi, "Segmentation of textured polarimetric SAR scenes by likelihood approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2063–2072, Oct. 2004.

[56] P. R. Kersten, J.-S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture Radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 519–527, Mar. 2005.

[57] A. C. Frery, J. Jacobo-Berlles, J. Gambini, and M. E. Mejail, "Polarimetric SAR image segmentation with B-splines and a new statistical model," *Multidimensional Syst. Signal Process.*, vol. 21, no. 4, pp. 319–342, Dec. 2010.

[58] J. Gambini, M. E. Mejail, J. Jacobo-Berlles, and A. C. Frery, "Accuracy of edge detection methods with local information in speckled imagery," *Statist. Comput.*, vol. 18, no. 1, pp. 15–26, Mar. 2008.

[59] A. K. Seghouane and S. I. Amari, "The AIC criterion and symmetrizing the Kullback–Leibler divergence," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 97–106, Jan. 2007.

[60] M. Salicrú, M. L. Menéndez, L. Pardo, and D. Morales, "On the applications of divergence type measures in testing statistical hypotheses," *J. Multivariate Anal.*, vol. 51, no. 2, pp. 372–391, Feb. 1994.

[61] J. Morio, P. Refrégier, F. Goudail, P. C. Dubois-Fernandez, and X. Dupuis, "A characterization of Shannon entropy and Bhattacharyya measure of contrast in polarimetric and interferometric SAR image," *Proc. IEEE*, vol. 97, no. 6, pp. 1097–1108, Jun. 2009.

[62] S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997.

[63] W. Yan, W. Yang, Y. Liu, and H. Sun, "Unsupervised classification of PolinSAR image based on Shannon entropy characterization," in *Proc. IEEE Int. Conf. Signal Process. (ICSP)*, Oct. 2010, pp. 2192–2195.

[64] L. Pardo, D. Morales, M. Salicrú, and M. L. Menéndez, "Large sample behavior of entropy measures when parameters are estimated," *Commun. Statist.-Theory Methods*, vol. 26, no. 2, pp. 483–501, Jan. 1997.

[65] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[66] G. P. S. Junior, A. C. Frery, S. Sandri, H. Bustince, E. Barrenechea, and C. Marco-Detchart, "Optical images-based edge detection in synthetic aperture radar images," *Knowl.-Based Syst.*, vol. 87, pp. 38–46, Oct. 2015.

[67] X. Deng, C. López-Martínez, J. Chen, and P. Han, "Statistical modeling of polarimetric SAR data: A survey and challenges," *Remote Sens.*, vol. 9, no. 4, p. 348, Apr. 2017.

Abraão D. C. Nascimento received the B.Sc., M.Sc., and D.Sc. degrees in statistics from the Universidade Federal de Pernambuco (UFPE), Recife, Brazil, in 2005, 2007, and 2012, respectively.

In 2014, he joined the Department of Statistics, UFPE, as an Adjoint Professor. His research interests include statistical information theory, inference on random matrices (with an emphasis for applications on PolSAR imagery), statistical theory of shape, spatio-temporal processes, survival analysis, and asymptotic theory.

Alejandro C. Frery (S'92–SM'03) received the B.Sc. degree in electronic and electrical engineering from the Universidad de Mendoza, Mendoza, Argentina, the M.Sc. degree in applied mathematics (statistics) from the Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil, and the Ph.D. degree in applied computing from the Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil.

He is currently the Leader of the Laboratório de Computação Científica e Análise Numérica, Universidade Federal de Alagoas, Maceió, Brazil. His research interests include statistical computing and stochastic modeling.

Renato J. Cintra (S'00–M'05–SM'10) received the B.Sc., M.Sc., and D.Sc. degrees in electrical engineering from the Universidade Federal de Pernambuco (UFPE), Recife, Brazil, in 1999, 2001, and 2005, respectively.

He joined the Department of Statistics, UFPE, in 2005. From 2014 to 2015, he was a Visiting Professor with the Département Informatique, INSA, Lyon, France. From 2017 to 2018, he is a Visiting Professor with the University of Calgary, Calgary, AB, Canada. His research interests include approximation theory for discrete transforms, theory and methods for digital signal processing, and statistical methods.

He is an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS; *Springer Circuits, Systems, and Signal Processing*; *IET Circuits, Devices & Systems*; and *Journal of Communication and Information Systems*.