

SPATIAL STATISTICS AND IMAGE MODELING

SPECIAL ISSUE: “III SEEMI”

RESEARCH PAPER

Information theory and image understanding: An application to polarimetric SAR imagery

ALEJANDRO C. FRERY^{1,*}, ABRAÃO D.C. NASCIMENTO² AND RENATO J. CINTRA²

¹Instituto de Computação, Universidade Federal de Alagoas, Maceió, Brazil

²Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brazil

(Received: 30 June 2011 · Accepted in final form: 31 August 2011)

Abstract

This work presents a comprehensive examination of the use of information theory for understanding polarimetric synthetic aperture radar (PolSAR) images by means of contrast measures that can be used as test statistics. Due to the phenomenon called ‘speckle’, common to all images obtained with coherent illumination such as PolSAR imagery, accurate modelling is required in their processing and analysis. The scaled multilook complex Wishart distribution has proven to be a successful approach for modelling radar backscatter from forest and pasture areas. Classification, segmentation, and image analysis techniques, which depend on this model, have been devised, and many of them employ some kind of dissimilarity measure. Specifically, we introduce statistical tests for analyzing contrast in such images. These tests are based on the Bhattacharyya, chi-square, Hellinger, Kullback-Leibler and Rényi distances. Results obtained by Monte Carlo experiments reveal the Kullback-Leibler distance as the best one with respect to the empirical test sizes under several situations which include pure and contaminated data. The proposed methodology is applied to actual data, obtained by an E-SAR sensor over surroundings of Weßling, Bavaria, Germany.

Keywords: Asymptotic theory · Hermitian random matrix · Hypothesis tests · PolSAR image · Signal processing · Statistical information theory.

Mathematics Subject Classification: Primary 62F99 · Secondary 94A17.

1. INTRODUCTION

The aim of remote sensing is to capture and to analyze scenes of the Earth. Among the remote sensing technologies, polarimetric synthetic aperture radar (PolSAR) has achieved a prominent position; see Lee and Pottier (2009).

*Corresponding author. Alejandro C. Frery. Instituto de Computação, Universidade Federal de Alagoas. BR 104 Norte km 97. 57072-970. Maceió, AL. Brasil. Email: acfrery@gmail.com

In general, PolSAR data are the result of the following procedure: orthogonally polarized electromagnetic pulses are transmitted towards a target, and the returned echo is recorded with respect to each polarization; see Lopez-Martinez and Fabregas (2003). Such data are processed in order to generate images and, as a consequence of the coherent illumination, they are contaminated with fluctuations on its detected intensity called ‘speckle’. Although speckle is a deterministic phenomenon since it is fully reproducible, it has the effect of a random noise. These alterations can significantly degrade the perceived image quality, as much as the ability of extracting information from the data.

Defining a stochastic identity for modeling PolSAR image regions is an important pre-processing step; see Conradsen et al. (2003). The scaled multilook complex Wishart distribution has been successfully employed as a statistical model for homogeneous regions in PolSAR imagery; see Frery et al. (2011). Several statistical image processing techniques use this distribution for segmentation (see Beaulieu and Touzi, 2004), classification (see Kersten et al., 2005), and boundary detection (see Schou et al., 2003), to name a few applications.

Any parametric approach requires parameter estimation. The scaled multilook complex Wishart law is indexed by a scalar known as the number of looks and a Hermitian complex matrix. These quantities can be estimated by a number of techniques, but a single scalar measure would be more useful when dealing with samples from images. Such measure can be referred to as ‘contrast’ if it provides means for discriminating different types of targets; see Gambini et al. (2006, 2008) and Goudail and Réfrégier (2004). Suitable measures of contrast not only provide useful information about the image scene but also assume a pivotal role in several image analysis procedures; see Schou et al. (2003).

Recent years have seen an increasing interest in adapting information-theoretic tools to image processing; see Goudail and Réfrégier (2004). In particular, the concept of stochastic divergence (see Liese and Vajda, 2006) has found applications in areas as diverse as image classification (see Puig and Garcia, 2003), cluster analysis (see Mak and Barnard, 1996), and multinomial goodness-of-fit tests (see Zografos et al., 1990). Coherent polarimetric image processing has also benefited, since divergence measures can furnish methods for assessing segmentation algorithms; see Schou et al. (2003). Morio et al. (2009) analyzed the Shannon entropy for the characterization of polarimetric imagery considering the circular multidimensional Gaussian distribution. In a previous work (see Nascimento et al., 2010), several parametric methods based on the class (h, ϕ) -divergences are proposed and submitted to a comprehensive examination.

The aim of this work is to assess the contrast capability of hypothesis tests based on stochastic distances between statistical models for PolSAR images. Analytic expressions for the Bhattacharyya, chi-square, Hellinger, Kullback-Leibler and Rényi distances between scaled multilook complex Wishart distributions in their most general way are derived. Subsequently, such measures are penalized by coefficients which depend on the size of two distinct samples, leading to the proposal of new homogeneity tests.

The performance of these five new hypothesis tests is analyzed by means of their observed test sizes using Monte Carlo in several possible scenarios including pure and contaminated data. This methodology is assessed with real data obtained by the E-SAR sensor.

The remainder of this paper is organized as follows. In Section 2, we introduce the image statistical modeling for the polarimetric covariance matrix. In Section 3, we adapt the hypothesis testing method proposed by Salicrú et al. (1994) for dealing with Hermitian positive definite matrix models. In Section 4, we present computational results obtained from synthetic and actual data analysis. In Section 5, we summarize the main conclusions of this paper.

2. A MODEL FOR POLARIMETRIC DATA: THE SCALED MULTILOOK COMPLEX WISHART DISTRIBUTION

The PolSAR processing results in a complex scattering matrix, which is defined by intensity and relative phase data. In strict terms such matrix has possibly four distinct complex elements, namely S_{VV} , S_{VH} , S_{HV} , and S_{HH} . However, under the conditions of the reciprocity theorem (see Ulaby and Elachi, 1990), the scattering matrix can be simplified to a three-component vector, since $S_{HV} = S_{VH}$. Thus, we have a scattering vector

$$\mathbf{s} = [S_{VV}, S_{VH}, S_{HH}]^\top,$$

where $[\cdot]^\top$ indicates vector transposition. Freeman and Durden (1998) presented an important description of this three-component representation. As discussed by Goodman (1963), the multivariate complex Gaussian distribution can adequately model the statistical behaviour of \mathbf{s} . This is called ‘single-look PolSAR data representation’ and hereafter we assume that the scattering vector has the dimension p , i.e., $\mathbf{s} = [S_1, S_2, \dots, S_p]^\top$.

Polarimetric data are usually subjected to multilook processing in order to improve the signal-to-noise ratio. To that end, Hermitian positive definite matrices are obtained by computing the mean over N independent looks of the same scene. This results in the sample covariance matrix \mathbf{Z} given by Anfinson et al. (2009) is expressed as

$$\mathbf{Z} = \frac{1}{L} \sum_{i=1}^L \mathbf{s}_i \mathbf{s}_i^H,$$

where L is the number of looks, \mathbf{s}_i , for $i = 1, 2, \dots, L$, and $[\cdot]^H$ represents the complex conjugate transposition. The sample covariance matrix follows a scaled multilook complex Wishart distribution with parameters $\mathbf{\Sigma}$ and L as parameters, characterized by the probability density function

$$f_{\mathbf{Z}}(\mathbf{Z}'; \mathbf{\Sigma}, L) = \frac{L^p |\mathbf{Z}'|^{L-p}}{|\mathbf{\Sigma}|^L \Gamma_p(L)} \exp(-L \operatorname{tr}(\mathbf{\Sigma}^{-1} \mathbf{Z}')), \quad (1)$$

where $\Gamma_p(L) = \pi^{p(p-1)/2} \prod_{i=0}^{p-1} \Gamma(L-i)$, $\Gamma(\cdot)$ is the gamma function, $\operatorname{tr}(\cdot)$ represents the trace operator, $|\cdot|$ denotes the determinant operator and the covariance matrix of \mathbf{Z} is given by

$$\mathbf{\Sigma} = E\{\mathbf{s} \mathbf{s}^*\} = \begin{bmatrix} E\{S_1 S_1^*\} & E\{S_1 S_2^*\} & \cdots & E\{S_1 S_p^*\} \\ E\{S_2 S_1^*\} & E\{S_2 S_2^*\} & \cdots & E\{S_2 S_p^*\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{S_p S_1^*\} & E\{S_p S_2^*\} & \cdots & E\{S_p S_p^*\} \end{bmatrix},$$

where $E\{\cdot\}$ and $[\cdot]^*$ denote expectation and complex conjugation, respectively. This distribution is denoted by $\mathbf{Z} \sim \mathcal{W}(L, \mathbf{\Sigma})$ and it satisfies $E(\mathbf{Z}) = \mathbf{\Sigma}$; see Anfinson et al. (2009).

Lee et al. (1994) derived many marginal distributions of the $\mathcal{W}(L, \mathbf{\Sigma})$ law and their transformations.

2.1 INTERPRETABILITY OF THE PARAMETERS Σ AND L IN POLSAR IMAGES

Parameters Σ and L possess physical meaning. The diagonal elements of Σ convey the brightness information of the respective channels. On its turn, an increasing number of looks L implies better image signal-to-noise ratios. Figure 1 illustrates the influence of these parameters on simulated PolSAR images of size 100×100 pixels. Such synthetic images are generated from scaled multilook complex Wishart law with following parameters: $L \in \{3, 12\}$ and $k \in \{0, 5, 10, 15\}$ such that $\Sigma = (1 + k)B$, where

$$B = \begin{bmatrix} 360932 & 11050 + 3759\mathbf{i} & 63896 + 1581\mathbf{i} \\ & 98960 & 6593 + 6868\mathbf{i} \\ & & 208843 \end{bmatrix},$$

where \mathbf{i} is the imaginary unit. Only the upper triangle and the diagonal are displayed because this covariance matrix Σ is Hermitian and, therefore, the remaining elements are the complex conjugates. This matrix is observed in Frery et al. (2010) for representing PolSAR data of forested areas. As shown in Figure 1, images with $L = 12$ are less affected by speckle, and the brighter are the ones indexed by $k = 15$, i.e., with larger determinants. In this image, the HH (HV, VV, respectively) intensity is associated with the red (green, blue respectively) channel in order to compose a false color display.

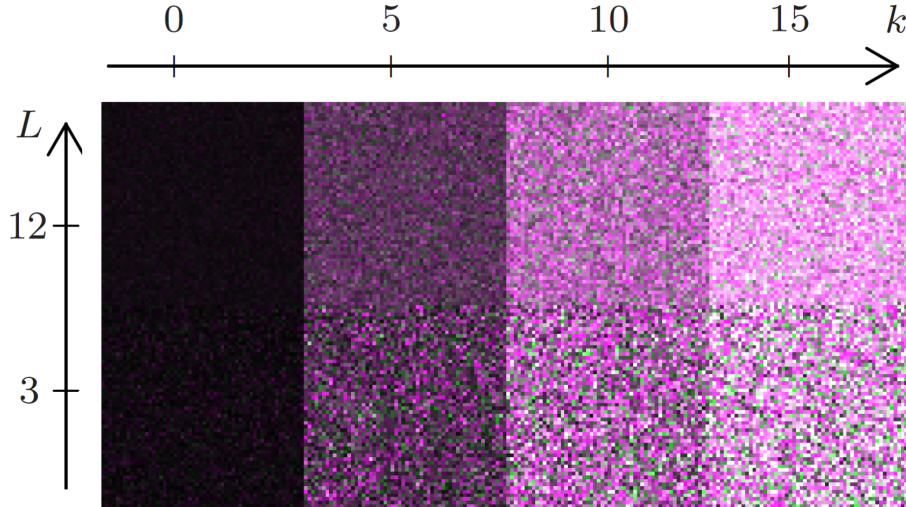


Figure 1. Simulated PolSAR images following scaled multilook complex Wishart law with parameters $L = \{3, 12\}$ and $k = \{0, 5, 10, 15\}$ such that $\Sigma = (1 + k)B$.

2.2 MAXIMUM LIKELIHOOD ESTIMATION UNDER THE WISHART MODEL

Let \mathbf{Z}_k be a random matrix which follows a scaled multilook complex Wishart law with parameters $\gamma = [L, \text{vec}(\Sigma)^\top]^\top$, where $\text{vec}(\cdot)$ is the column stacking vectorisation operator. Its log-likelihood function is expressed by

$$\ell_k(\gamma) = pL \log L + (L-p) \log |\mathbf{Z}_k| - L \log |\Sigma| - \frac{p(p-1)}{2} \log \pi - \sum_{k=0}^{p-1} \log \Gamma(L-k) - L \text{tr}(\Sigma^{-1} \mathbf{Z}_k).$$

According to Hjørungnes and Gesbert (2007), $\partial \ell_k(\boldsymbol{\gamma}) / \partial \text{vec}(\boldsymbol{\Sigma}) = \text{vec}(\partial \ell_k(\boldsymbol{\gamma}) / \partial \boldsymbol{\Sigma})$ then, the score function based on \mathbf{Z}_k is given by

$$\nabla \ell_k(\boldsymbol{\gamma}) = \begin{cases} p(\log L + 1) + \log |\mathbf{Z}_k| - \log |\boldsymbol{\Sigma}| - \sum_{i=0}^{p-1} \psi^{(0)}(L - i) - \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{Z}_k), \\ L \text{vec}(\boldsymbol{\Sigma}^{-1} \mathbf{Z}_k \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}). \end{cases}$$

In the sequel, we present the derivation of (A1) the Hessian matrix $\mathcal{J}(\boldsymbol{\gamma})$, (A2) the Fisher information matrix $\mathcal{K}(\boldsymbol{\gamma})$, and (A3) the Cramér-Rao lower bound $\mathcal{C}(\boldsymbol{\gamma})$. To that end, the following quantity plays a central role:

$$\mathcal{J}_{\Sigma\Sigma} = \frac{\partial}{\partial \text{vec}(\boldsymbol{\Sigma})^*} \left\{ \text{vec} \left(\frac{\partial \ell_k(\boldsymbol{\gamma})}{\partial \boldsymbol{\Sigma}} \right)^\top \right\}.$$

Anfinssen et al. (2009) showed that

$$T_1 = -\frac{\partial \{\boldsymbol{\Sigma}^{-1} \mathbf{Z} \boldsymbol{\Sigma}^{-1}\}}{\partial \boldsymbol{\Sigma}} = -\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \mathbf{Z} \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{Z},$$

where \otimes denotes the Kronecker product. Moreover, it is known that (see Hjørungnes and Gesbert, 2007) $T_2 = \partial \boldsymbol{\Sigma}^{-1} / \partial \boldsymbol{\Sigma} = -\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}$. Thus, we have that

$$\mathcal{J}_{\Sigma\Sigma} = L\{T_2 - T_1\} = L\{\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \mathbf{Z} \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{Z}\}. \quad (2)$$

From Equation (2), it is possible to obtain that $\mathcal{K}_{\Sigma\Sigma} = \text{E}\{-\mathcal{J}_{\Sigma\Sigma}\} = L\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}$. Thus, matrices $\mathcal{J}(\boldsymbol{\gamma})$, $\mathcal{K}(\boldsymbol{\gamma})$, and $\mathcal{C}(\boldsymbol{\gamma})$ can be expressed respectively as

$$\begin{aligned} \mathcal{J}(\boldsymbol{\gamma}) &= \begin{bmatrix} \frac{p}{L} - \sum_{i=1}^{p-1} \psi^{(1)}(L - i) & \text{vec}(\boldsymbol{\Sigma}^{-1} \mathbf{Z} \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1})^\top \\ \text{vec}(\boldsymbol{\Sigma}^{-1} \mathbf{Z} \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1})^* & \mathcal{J}_{\Sigma\Sigma} \end{bmatrix}, \\ \mathcal{K}(\boldsymbol{\gamma}) = \text{E}\{-\mathcal{J}(\boldsymbol{\theta})\} &= \begin{bmatrix} \sum_{i=1}^{p-1} \psi^{(1)}(L - i) - \frac{p}{L} & \text{vec}(\mathbf{0}_{p^2})^\top \\ \text{vec}(\mathbf{0}_{p^2}) & L\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \end{bmatrix}, \end{aligned} \quad (3)$$

and

$$\mathcal{C}(\boldsymbol{\gamma}) = \mathcal{K}(\boldsymbol{\gamma})^{-1} = \begin{bmatrix} (\sum_{i=1}^{p-1} \psi^{(1)}(L - i) - \frac{p}{L})^{-1} & \text{vec}(\mathbf{0}_{p^2})^\top \\ \text{vec}(\mathbf{0}_{p^2}) & \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} / L \end{bmatrix}.$$

Anfinssen et al. (2009) derived the information Fisher matrix for the complex unscaled Wishart law, and found that the parameters of such distribution are not orthogonal. However, dividing by the number of looks results in a block-diagonal Fisher information matrix as expressed by $\mathcal{K}(\boldsymbol{\gamma})$ in Equation (3). Thus, such scaling, whose density is given in Equation (1), leads to a distribution with orthogonal parameters with good properties as, for instance, separable likelihood equations. In the following we discuss maximum likelihood (ML) estimation under such distribution.

Let $\{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_N\}$ be a random sample of $\mathbf{Z} \sim \mathcal{W}(L, \boldsymbol{\Sigma})$ of size N , with $\boldsymbol{\theta} = (L, \boldsymbol{\Sigma})$ the parameter whose elements define the vector $\boldsymbol{\gamma}$, and let $\hat{\boldsymbol{\theta}} = (\hat{L}, \hat{\boldsymbol{\Sigma}})$ be its ML estimator.

Expressing $N^{-1} \sum_{k=1}^N \nabla \ell_k(\hat{\gamma}) = \mathbf{0}$, one has that

$$\begin{cases} \hat{\Sigma} = N^{-1} \sum_{k=1}^N \mathbf{Z}_k, \\ p \log \hat{L} + N^{-1} \sum_{k=1}^N \log |\mathbf{Z}_k| - \log |\bar{\mathbf{Z}}| - \sum_{i=0}^{p-1} \psi^{(0)}(\hat{L} - i) = 0, \end{cases} \quad (4)$$

where \hat{L} is the ML estimator of the number of looks. Thus, the ML estimators of Σ and L are given by the sample mean and by the solution of Equation (4), respectively. The Newton-Raphson numerical optimization method is used for solving the latter, since a closed form solution is not trivially found.

In the following, this estimation is applied to actual data. Figure 2 presents the HH channel of a polarimetric SAR image obtained by the E-SAR sensor over surroundings of Weßling, Germany. The informed (nominal) number of looks is 3. The area exhibits two distinct types of target roughness: homogeneous (pasture) and heterogeneous (forest). Table 1 lists the ML estimates, as well as the sample sizes.

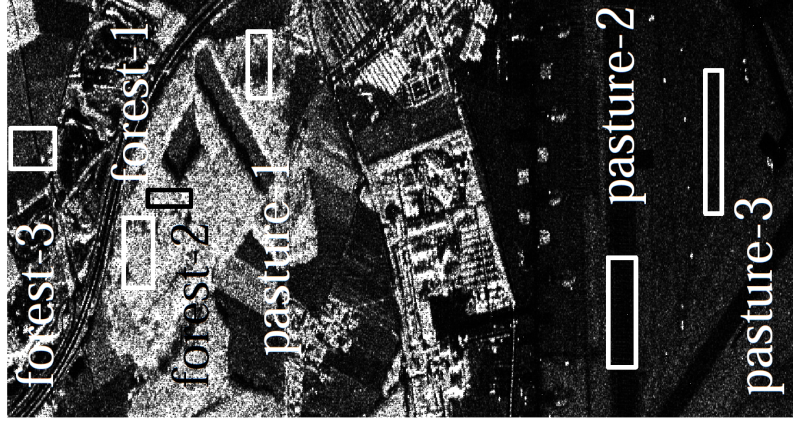


Figure 2. E-SAR image (HH channel) with selected regions.

Table 1. Parameter estimates.

Region	\hat{L}	$ \hat{\Sigma} $	# pixels
pasture-1	2.870	7.934	2106
pasture-2	2.573	74.660	2352
pasture-3	2.889	26.452	3340
forest-1	2.638	33615.990	2870
forest-2	2.727	10755.870	2496
forest-3	2.303	7421.431	1020

Figure 3 depicts empirical densities of data samples from the selected forest and pasture regions. Additionally, the associated fitted marginal densities are also shown for comparison. In this case, the Wishart density collapses to gamma marginal densities, as demonstrated in (see Hagedorn et al., 2006):

$$f_{Z_i}(Z'_i; 2\sigma_i^2, L) = \frac{Z_i'^{L-1}}{2^L \sigma_i^{2L} \Gamma(L)} \exp(-Z'_i/2\sigma_i^2),$$

where σ_i^2 is the element (i, i) of Σ , $p = 3$, and Z'_i is the entry (i, i) of the random matrix \mathbf{Z} , and $1 \leq i \leq 3$ with the association 1 to HH, 2 to HV and 3 to VV. The adequacy of the model to the data is noteworthy. These samples are used to validate our proposed methods in Section 4.

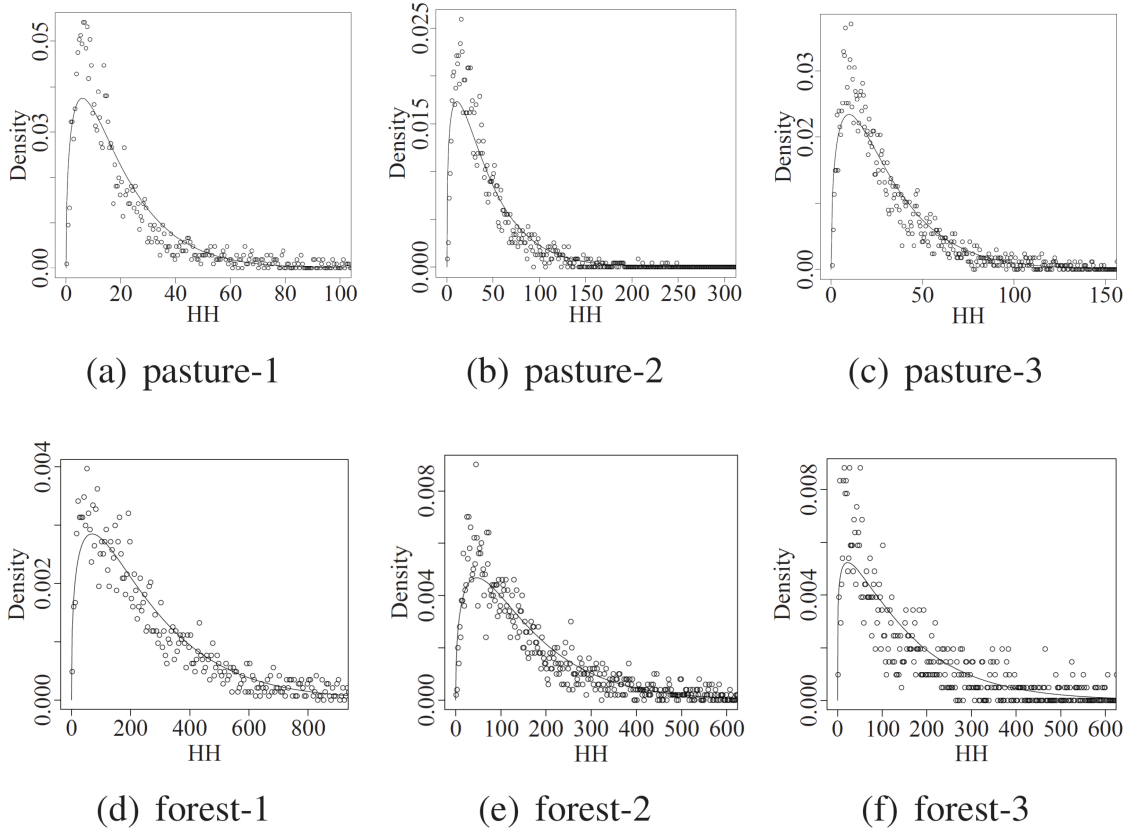


Figure 3. Histograms and empirical densities of samples.

3. STATISTICAL INFORMATION THEORY FOR RANDOM MATRICES

In the following, we adhere to the convention that a (stochastic) ‘divergence’ is any non-negative function of two probability measures. If the function is also symmetric, it is called a (stochastic) ‘distance’. Finally, we understand (stochastic) ‘metric’ as a distance which satisfies the triangular inequality; see Deza and Deza (2009, chapters 1 and 14).

An image can be understood as a set of regions formed by pixels which are observations of random variables following a certain distribution. Therefore, stochastic dissimilarity measures could be used as features within image analysis techniques, since they may be able to assess the difference between the distributions that describe different image areas; see Nascimento et al. (2010). Dissimilarity measures are submitted to a systematic and comprehensive treatment in Ali and Silvey (1996), Csiszár (1967) and Salicrú et al. (1994) and, as a result, the class of (h, ϕ) -divergences is proposed; see Salicrú et al. (1994).

Assume that \mathbf{X} and \mathbf{Y} are random matrices whose distributions are characterized by the densities $f_{\mathbf{X}}(\mathbf{Z}'; \boldsymbol{\theta}_1)$ and $f_{\mathbf{Y}}(\mathbf{Z}'; \boldsymbol{\theta}_2)$, respectively, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are parameters. Both densities are assumed to share a common support given by the cone of Hermitian positive definite matrices \mathcal{A} . The (h, ϕ) -divergence between $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$ is defined by

$$D_{\phi}^h(\mathbf{X}, \mathbf{Y}) = h \left(\int_{\mathcal{A}} \phi \left(\frac{f_{\mathbf{X}}(\mathbf{Z}'; \boldsymbol{\theta}_1)}{f_{\mathbf{Y}}(\mathbf{Z}'; \boldsymbol{\theta}_2)} \right) f_{\mathbf{Y}}(\mathbf{Z}'; \boldsymbol{\theta}_2) d\mathbf{Z}' \right), \quad (5)$$

where $\phi: (0, \infty) \rightarrow [0, \infty)$ is a convex function and $h: (0, \infty) \rightarrow [0, \infty)$ is a strictly increas-

ing function with $h(0) = 0$. The differential element $d\mathbf{Z}'$ is given by

$$d\mathbf{Z}' = dZ'_{11}dZ'_{22}\cdots dZ'_{pp} \prod_{\substack{i,j=1 \\ i < j}}^p d\Re\{Z'_{ij}\}d\Im\{Z'_{ij}\},$$

where Z'_{ij} is the (i, j) entry of matrix \mathbf{Z}' , and operators $\Re\{\cdot\}$ and $\Im\{\cdot\}$ return real and imaginary parts of their arguments, respectively; see Goodman (1963). If indeterminate forms appear in Equation (5) due to the ratio of densities, they are assigned value zero.

Well-known divergences arise with adequate choices of h and ϕ . Among them, the following are examined: (i) the χ^2 divergence (see Taneja, 2006), (ii) the Kullback-Leibler divergence (see Seghouane and Amari, 2007), (iii) the Rényi divergence (see Rached et al., 2001), (iv) the Bhattacharyya distance (see Kailath, 1967), and (v) the Hellinger distance (see Nascimento et al., 2010). As the triangular inequality is not necessarily satisfied, not every divergence measures is a metric; see Burbea and Rao (1982). Additionally, the symmetry property is not followed by some of these divergence measures. However, such tools are mathematically appropriated for comparing the distribution of random variables; see Aviyente (2003). Thus, the following expression has been suggested as a possible solution for this problem (see Seghouane and Amari, 2007):

$$d_\phi^h(\mathbf{X}, \mathbf{Y}) = \frac{D_\phi^h(\mathbf{X}, \mathbf{Y}) + D_\phi^h(\mathbf{Y}, \mathbf{X})}{2}.$$

In analogy with Goudail and Réfrégier (2004) and Morio et al. (2008), this paper defines distances as symmetrized versions of the divergence measures, i.e., a function $d_\phi^h : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$ is a distance on \mathcal{A} if, for all $\mathbf{X}, \mathbf{Y} \in \mathcal{A}$, the following properties holds:

- (i) Non-negativity: $d_\phi^h(\mathbf{X}, \mathbf{Y}) \geq 0$.
- (ii) Symmetry: $d_\phi^h(\mathbf{X}, \mathbf{Y}) = d_\phi^h(\mathbf{Y}, \mathbf{X})$.
- (iii) Identity of indiscernibles: $d_\phi^h(\mathbf{X}, \mathbf{Y}) = 0 \Leftrightarrow \mathbf{X} = \mathbf{Y}$.

Thus, the χ^2 , Kullback-Leibler, and Rényi divergences are made symmetric distances. Table 2 shows the functions h and ϕ which lead to each of the above listed distances.

Table 2. (h, ϕ) -distance and related functions ϕ and h

(h, ϕ) -distance	$h(y)$	$\phi(x)$
χ^2	$y/4$	$(x-1)^2(x+1)/x$
Kullback-Leibler	$y/2$	$(x-1)\log x$
Rényi (order β)	$\frac{1}{\beta-1} \log((\beta-1)y+1), 0 \leq y < \frac{1}{1-\beta}$	$\frac{x^{1-\beta}+x^\beta-\beta(x-1)-2}{2(\beta-1)}, 0 < \beta < 1$
Bhattacharyya	$-\log(1-y), 0 \leq y < 1$	$-\sqrt{x} + \frac{x+1}{2}$
Hellinger	$y/2, 0 \leq y < 2$	$(\sqrt{x}-1)^2$

In the following, we discuss integral expressions of these (h, ϕ) -distances. For simplicity, we suppress the explicit dependence on \mathbf{Z}' and on the support \mathcal{A} .

- (i) The χ^2 distance:

$$d_{\chi^2}(\mathbf{X}, \mathbf{Y}) = \frac{1}{2}[D_{\chi^2}(\mathbf{X}, \mathbf{Y}) + D_{\chi^2}(\mathbf{Y}, \mathbf{X})] = \frac{1}{2}\left[\int \frac{(f_{\mathbf{X}} - f_{\mathbf{Y}})^2}{2f_{\mathbf{X}}} + \int \frac{(f_{\mathbf{X}} - f_{\mathbf{Y}})^2}{2f_{\mathbf{Y}}}\right].$$

The divergence D_{χ^2} has been used in many statistical contexts. For instance, Broniatowski and Keziou (2009) proposed efficient hypothesis test combining it with the duality technique.

(ii) The Kullback-Leibler distance:

$$\begin{aligned} d_{\text{KL}}(\mathbf{X}, \mathbf{Y}) &= \frac{1}{2}[D_{\text{KL}}(\mathbf{X}, \mathbf{Y}) + D_{\text{KL}}(\mathbf{Y}, \mathbf{X})] = \frac{1}{2} \left[\int f_{\mathbf{X}} \log \frac{f_{\mathbf{X}}}{f_{\mathbf{Y}}} + \int f_{\mathbf{Y}} \log \frac{f_{\mathbf{Y}}}{f_{\mathbf{X}}} \right] \\ &= \frac{1}{2} \int (f_{\mathbf{X}} - f_{\mathbf{Y}}) \log \frac{f_{\mathbf{X}}}{f_{\mathbf{Y}}}. \end{aligned}$$

The divergence D_{KL} has a close relationship with the Neyman-Pearson lemma (see Eguchi and Copas, 2006) and its symmetrization has been suggested as a correction form of the Akaike information criterion, which is a descriptive measure for assessing the adequacy of statistical models.

(iii) The Rényi distance of order β :

$$\widetilde{d}_{\text{R}}^{\beta}(\mathbf{X}, \mathbf{Y}) = \frac{1}{2}[D_{\text{R}}^{\beta}(\mathbf{X}, \mathbf{Y}) + D_{\text{R}}^{\beta}(\mathbf{Y}, \mathbf{X})] = \frac{\log \int f_{\mathbf{X}}^{\beta} f_{\mathbf{Y}}^{1-\beta} + \log \int f_{\mathbf{X}}^{1-\beta} f_{\mathbf{Y}}^{\beta}}{2(\beta - 1)},$$

where $0 < \beta < 1$. The divergence D_{R}^{β} has been used for analysing geometric characteristics with respect to probability laws; see Andai (2009). By the Fejer inequality (see Neuman, 1990), we have that

$$d_{\text{R}}^{\beta}(\mathbf{X}, \mathbf{Y}) \triangleq \frac{1}{\beta - 1} \log \frac{\int f_{\mathbf{X}}^{\beta} f_{\mathbf{Y}}^{1-\beta} + \int f_{\mathbf{X}}^{1-\beta} f_{\mathbf{Y}}^{\beta}}{2} \leq \widetilde{d}_{\text{R}}^{\beta}(\mathbf{X}, \mathbf{Y}).$$

Being d_{R}^{β} more tractable than $\widetilde{d}_{\text{R}}^{\beta}$ for algebraic manipulation with the scaled multilook complex Wishart density, this paper considers the former.

(iv) The Bhattacharyya distance:

$$d_{\text{B}}(\mathbf{X}, \mathbf{Y}) = -\log \int \sqrt{f_{\mathbf{X}} f_{\mathbf{Y}}}.$$

Goudail et al. (2004) showed that this distance is an efficient tool for contrast definition in algorithms for image processing.

(v) The Hellinger distance:

$$d_{\text{H}}(\mathbf{X}, \mathbf{Y}) = 1 - \int \sqrt{f_{\mathbf{X}} f_{\mathbf{Y}}}.$$

Estimation methods based on the minimization of d_{H} have been successfully employed in the context of stochastic differential equations; see Giet and Lubrano (2008).

When considering the distance between particular cases of the same distribution, only the parameters are relevant. In this case, the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ replace the random variables \mathbf{X} and \mathbf{Y} . This notation is in agreement with that of Salicrú et al. (1994).

In the following five subsections, we present the expressions of the discussed distances for the scaled multilook complex Wishart distribution, characterized by the density given in Equation (1) are presented.

3.1 STOCHASTIC DISTANCES BETWEEN SCALED MULTILOOK COMPLEX WISHART LAWS

Next, analytic expressions for the stochastic distances d_{χ^2} , d_{KL} , d_{R}^β , d_{B} , and d_{H} between two complex scaled multilook complex Wishart distributions are presented. In all instances, the parameters $\boldsymbol{\theta}_X = (L_X, \boldsymbol{\Sigma}_X)$ and $\boldsymbol{\theta}_Y = (L_Y, \boldsymbol{\Sigma}_Y)$ are considered. In order to avoid confusion with the determinant, the absolute value of the scalar x are denoted $\text{abs}(x)$.

THE χ^2 DISTANCE

$$d_{\chi^2}(\boldsymbol{\theta}_X, \boldsymbol{\theta}_Y) = \frac{I_{XY}^{\chi^2} + I_{YX}^{\chi^2} - 2}{4},$$

where

$$I_{ij}^{\chi^2} = \frac{|\boldsymbol{\Sigma}_i|^{L_i}}{|\boldsymbol{\Sigma}_j|^{2L_j}} \text{abs}(|(2L_j \boldsymbol{\Sigma}_j^{-1} - L_i \boldsymbol{\Sigma}_i^{-1})^{-1}|)^{\text{abs}(2L_j - L_i)} \\ \frac{L_i^{-pL_i}}{L_j^{-2pL_j}} \prod_{k=1}^{p-1} \frac{(L_i - k)^k}{(L_j - k)^{2k}} \text{abs}(2L_j - L_i - k)^k \left[\frac{\Gamma(L_i - p + 1) \Gamma(\text{abs}(2L_j - L_i - p + 1))}{\Gamma(L_j - p + 1)^2} \right]^p,$$

for $(i, j) \in \{(X, Y), (Y, X)\}$.

THE KULLBACK-LEIBLER DISTANCE

$$d_{\text{KL}}(\boldsymbol{\theta}_X, \boldsymbol{\theta}_Y) = \frac{L_X - L_Y}{2} \left\{ \log \frac{|\boldsymbol{\Sigma}_X|}{|\boldsymbol{\Sigma}_Y|} + p [\psi^{(0)}(L_X - p + 1) - \psi^{(0)}(L_Y - p + 1)] \right. \\ \left. - p \log \frac{L_X}{L_Y} + (L_Y - L_X) \sum_{i=1}^{p-1} \frac{i}{(L_X - i)(L_Y - i)} \right\} - \frac{p(L_X + L_Y)}{2} \\ + \frac{\text{tr}(L_Y \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_X + L_X \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Sigma}_Y)}{2}.$$

THE RÉNYI DISTANCE OF ORDER β

$$d_{\text{R}}^\beta(\boldsymbol{\theta}_X, \boldsymbol{\theta}_Y) = \frac{1}{\beta - 1} \log \frac{I(\boldsymbol{\theta}_X, \boldsymbol{\theta}_Y)}{2},$$

such that

$$I(\boldsymbol{\theta}_X, \boldsymbol{\theta}_Y) = \\ \left[\Gamma(L_X - p + 1)^p |\boldsymbol{\Sigma}_X|^{L_X} L_X^{-pL_X} \prod_{i=1}^{p-1} (L_X - i)^i \right]^{-\beta} \left[\Gamma(L_Y - p + 1)^p |\boldsymbol{\Sigma}_Y|^{L_Y} L_Y^{-pL_Y} \prod_{i=1}^{p-1} (L_Y - i)^i \right]^{\beta-1} \\ \Gamma(E_1 - p + 1)^p |\boldsymbol{\Sigma}_{XY}|^{E_1} \prod_{i=1}^{p-1} (E_1 - i)^i + \left[\Gamma(L_X - p + 1)^p |\boldsymbol{\Sigma}_X|^{L_X} L_X^{-pL_X} \prod_{i=1}^{p-1} (L_X - i)^i \right]^{\beta-1} \\ \left[\Gamma(L_Y - p + 1)^p |\boldsymbol{\Sigma}_Y|^{L_Y} L_Y^{-pL_Y} \prod_{i=1}^{p-1} (L_Y - i)^i \right]^{-\beta} \Gamma(E_2 - p + 1)^p |\boldsymbol{\Sigma}_{YX}|^{E_2} \prod_{i=1}^{p-1} (E_2 - i)^i,$$

where $E_1 = \beta L_X + (1 - \beta)L_Y$, $E_2 = \beta L_Y + (1 - \beta)L_X$, $0 < \beta < 1$, and $\Sigma_{ij} = |(L_i \beta \Sigma_i^{-1} + L_j(1 - \beta) \Sigma_j^{-1})^{-1}|$ for $(i, j) \in \{(X, Y), (Y, X)\}$.

THE BHATTACHARYYA DISTANCE

$$d_B(\theta_X, \theta_Y) = \sum_{k=0}^{p-1} \log \frac{\sqrt{\Gamma(L_X - k)\Gamma(L_Y - k)}}{\Gamma(\frac{L_X + L_Y}{2} - k)} + \frac{L_X \log |\Sigma_X|}{2} + \frac{L_Y \log |\Sigma_Y|}{2} \\ - \frac{p}{2} (L_X \log L_X + L_Y \log L_Y) - \frac{L_X + L_Y}{2} \log \left| \left(\frac{L_X \Sigma_X^{-1} + L_Y \Sigma_Y^{-1}}{2} \right)^{-1} \right|.$$

THE HELLINGER DISTANCE

$$d_H(\theta_X, \theta_Y) = 1 - \frac{\left| \left(\frac{L_X \Sigma_X^{-1} + L_Y \Sigma_Y^{-1}}{2} \right)^{-1} \right|^{\frac{L_X + L_Y}{2}}}{|\Sigma_X|^{\frac{L_X}{2}} |\Sigma_Y|^{\frac{L_Y}{2}}} \sqrt{L_X^{pL_X} L_Y^{pL_Y}} \prod_{k=0}^{p-1} \frac{\Gamma(\frac{L_X + L_Y}{2} - k)}{\sqrt{\Gamma(L_X - k)\Gamma(L_Y - k)}}.$$

3.2 SENSITIVITY ANALYSIS

Now, we examine the behavior of the distances with respect to the variation of parameters. First, we assume a fixed number of looks, namely $L = 8$, and adopt the following parameters: $\theta_1 = (8, \Sigma(360932))$ and $\theta_2 = (8, \Sigma(x))$, where

$$\Sigma(x) = \begin{bmatrix} x & 11050 + 3759i & 63896 + 1581i \\ & 98960 & 6593 + 6868i \\ & & 208843 \end{bmatrix}.$$

As the covariance matrix is Hermitian, only the upper triangle and the diagonal are displayed. The fixed covariance matrix $\Sigma(360932)$ is observed by Frery et al. (2010) in PolSAR data of forested areas.

Figure 4(a) shows the distances for $x \in [359000, 363000]$. The three uppermost curves, i.e., the ones that vary more abruptly, are d_{χ^2} and d_{KL} , which are overlapped, and $d_R^{0.9}$. Analogously, distances d_B and d_H present roughly the same behavior and are also superimposed. The least varying distance is $d_R^{0.1}$.

We also consider fixed covariance matrices with varying number of looks: $\theta_1 = (8, \Sigma(360932))$ and $\theta_2 = (m, \Sigma(360932))$, for $5 \leq m \leq 11$. Figure 4(b) shows the distances, where we notice that d_{KL} is the one that varies most, followed by d_{χ^2} , $d_R^{0.9}$, d_B , d_H , and $d_R^{0.1}$.

It is noteworthy that the distances are much more sensitive to variations of one element of the covariance matrix than to variations of the number of looks.

As presented by Yu et al. (2008), in order to make fair comparisons among stochastic distances, the distributions must be indexed by the same parameter. The evidence presented in Figure 4 strongly suggests that d_{χ^2} , d_{KL} , and $d_R^{0.9}$ outperform all other distances. This is due to their higher sensitivity to variations around fixed parameter values, that makes them more suitable for discrimination purposes.

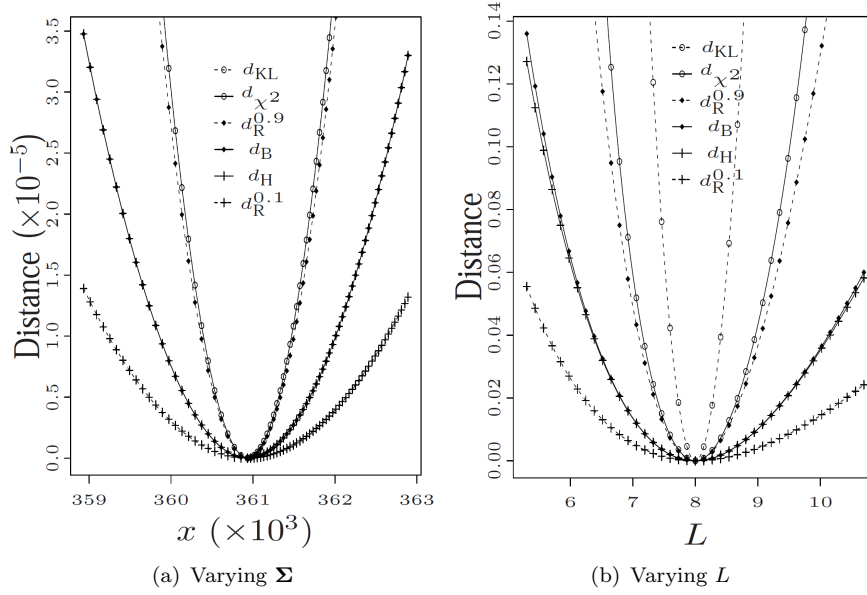


Figure 4. Sensitivity of the proposed distances.

3.3 HYPOTHESIS TEST BASED ON DIVERGENCE FOR POSITIVE-DEFINITE HERMITIAN DATA

In the following, the hypothesis test based on stochastic distances proposed by Salicr  et al. (1994) are recalled and applied to scaled multilook complex Wishart laws.

Let $\hat{\boldsymbol{\theta}}_1 = (\hat{\theta}_{11}, \dots, \hat{\theta}_{1M})$ and $\hat{\boldsymbol{\theta}}_2 = (\hat{\theta}_{21}, \dots, \hat{\theta}_{2M})$ be the ML estimators of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ based on independent samples of size N_X and N_Y , respectively. Under the regularity conditions discussed by Salicr  et al. (1994, p. 380) the following lemma holds.

LEMMA 2 If $\frac{N_X}{N_X + N_Y} \xrightarrow{N_X, N_Y \rightarrow \infty} \lambda \in (0, 1)$ and $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, then

$$S_{\phi}^h(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = \frac{2N_X N_Y}{N_X + N_Y} \frac{d_{\phi}^h(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)}{h'(0)\phi''(1)} \xrightarrow[N_X, N_Y \rightarrow \infty]{\mathcal{D}} \chi_M^2,$$

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

Based on Lemma 2, statistical hypothesis tests for the null hypothesis $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ can be derived in the form of following proposition.

PROPOSITION 2 Let N_X and N_Y be large and $S_{\phi}^h(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = s$, then the null hypothesis $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ can be rejected at level α if $\Pr(\chi_M^2 > s) \leq \alpha$.

We denote the statistics based on the Bhattacharyya, chi-square, Hellinger, Kullback-Leibler and R nyi distances as S_{B} , S_{χ^2} , S_{H} , S_{KL} and S_{R}^{β} , respectively.

4. SIMULATION AND APPLICATION

This section reports the assessment of the proposed methods as contrast measures. Both synthetic and actual image data are considered. Moreover, a robustness analysis is performed in order to characterize the behavior of such measures in the presence of outliers.

4.1 SYNTHETIC DATA ANALYSIS

We assess the influence of estimation on the size of the new hypothesis tests using simulated data. The study is conducted sampling from $L = L_X = L_Y \in \{4, 8, 16\}$ looks and from each of the covariance matrices observed in the forest areas presented in Figure 3. The sample sizes represent square windows of size 7×7 , 11×11 , and 20×20 pixels, i.e., $N_X, N_Y \in \{49, 121, 400\}$, which are typical in image processing and analysis. Nominal significance levels $\alpha \in \{1\%, 5\%\}$ are verified.

Let T be the number of Monte Carlo replicas and C the number of cases for which the null hypothesis is rejected at nominal level α when samples come from the same distribution. The empirical test size is given by $\hat{\alpha}_{1-\alpha} = C/T$. Following the methodology described by Nascimento et al. (2010), we used $T = 5500$ replicas.

Table 3 presents the empirical test size at 1% and 5% nominal levels, the execution time in miliseconds¹, and the distances mean (\bar{d}) and coefficient of variation (CV). The smallest empirical size and distance mean are in boldface.

Table 3. Empirical sizes for forest-1.

Factors			$L = 4$					$L = 8$					$L = 16$				
			Time (ms)					Time (ms)					Time (ms)				
S_ϕ^h	N_X	N_Y	1%	5%	\bar{d}	CV	1%	5%	\bar{d}	CV	1%	5%	\bar{d}	CV			
S_{χ^2}	49	49	84.96	91.55	1.03	64.58	159.51	81.55	89.13	1.08	52.62	81.55	78.69	87.05	1.09	48.70	78.16
	49	121	82.44	90.62	1.11	50.74	86.63	78.98	88.07	1.08	44.90	68.93	76.64	86.64	1.07	42.26	65.09
	49	400	81.84	90.38	1.02	47.40	66.63	77.09	87.64	1.09	42.84	63.52	74.93	85.11	1.06	40.71	61.99
	121	121	79.96	89.27	1.07	43.89	60.13	76.16	86.58	1.09	40.16	59.26	73.76	85.31	1.04	38.12	57.44
	121	400	78.78	88.51	1.12	41.76	56.54	74.25	85.53	1.08	37.16	54.32	73.33	84.84	1.12	37.26	56.76
	400	400	77.96	87.60	1.09	38.88	52.01	73.36	84.71	1.05	35.95	51.72	71.51	83.56	1.10	34.99	51.39
S_{KL}	49	49	1.91	7.18	0.44	10.59	45.65	1.05	4.85	0.49	9.75	46.78	0.91	4.22	0.39	9.39	47.80
	49	121	1.58	6.35	0.41	10.47	45.78	1.00	4.42	0.45	9.60	46.02	0.78	3.67	0.45	9.24	46.92
	49	400	1.56	6.95	0.48	10.57	45.02	1.04	4.35	0.46	9.67	46.36	0.82	3.95	0.47	9.32	47.49
	121	121	1.67	6.62	0.43	10.56	45.79	0.85	4.60	0.49	9.66	46.80	0.75	3.78	0.48	9.21	47.09
	121	400	1.82	7.64	0.50	10.69	45.69	0.76	4.27	0.46	9.49	45.90	1.16	4.49	0.47	9.43	48.43
	400	400	1.47	6.91	0.42	10.55	44.91	1.00	4.16	0.50	9.58	46.21	0.58	3.56	0.46	9.26	46.46
S_R^β	49	49	5.89	14.76	1.10	12.38	49.66	4.13	12.93	1.12	11.72	50.22	4.20	11.40	1.07	11.35	51.15
	49	121	4.78	14.38	1.00	12.18	49.79	3.89	11.58	1.10	11.52	49.94	3.62	10.36	1.09	11.16	50.18
	49	400	5.71	15.18	1.08	12.31	49.25	4.25	12.47	0.96	11.62	50.38	3.85	11.11	1.11	11.27	50.84
	121	121	5.56	14.69	1.11	12.32	49.56	4.07	12.02	0.99	11.59	50.65	3.45	10.71	1.12	11.17	50.14
	121	400	6.29	16.00	1.11	12.47	50.31	3.65	11.20	1.06	11.33	49.50	3.98	11.36	1.01	11.40	51.57
	400	400	5.71	15.24	0.99	12.35	49.47	3.69	11.62	1.13	11.50	49.55	3.27	11.35	1.20	11.23	49.49
S_B	49	49	5.51	14.25	0.54	12.27	49.22	3.96	12.65	0.51	11.66	50.01	4.07	11.31	0.58	11.33	51.04
	49	121	4.56	14.04	0.65	12.10	49.48	3.78	11.35	0.51	11.48	49.80	3.58	10.29	0.57	11.14	50.11
	49	400	5.45	14.84	0.56	12.25	49.01	4.18	12.36	0.57	11.59	50.26	3.84	11.02	0.58	11.26	50.78
	121	121	5.42	14.47	0.57	12.27	49.39	3.98	11.95	0.61	11.56	50.56	3.45	10.64	0.56	11.16	50.09
	121	400	6.20	15.75	0.56	12.44	50.20	3.62	11.09	0.55	11.32	49.44	3.98	11.36	0.51	11.40	51.54
	400	400	5.65	15.18	0.57	12.33	49.42	3.67	11.58	0.57	11.50	49.52	3.27	11.33	0.65	11.23	49.47
S_H	49	121	3.73	12.29	0.50	11.78	48.01	3.00	10.05	0.52	11.20	48.41	2.87	9.27	0.60	10.87	48.80
	49	400	4.49	13.38	0.64	11.99	47.90	3.56	10.93	0.62	11.36	49.16	3.27	9.71	0.55	11.04	49.72
	121	121	4.89	13.58	0.58	12.08	48.56	3.60	11.35	0.62	11.39	49.77	2.96	9.87	0.62	11.00	49.34
	121	400	5.67	15.18	0.60	12.31	49.64	3.35	10.58	0.54	11.22	48.96	3.71	10.80	0.60	11.29	50.99
	400	400	5.40	14.93	0.55	12.27	49.17	3.51	11.38	0.51	11.44	49.29	3.11	11.02	0.58	11.18	49.26

¹All experiments are performed on a PC with an Intel Core 2 Duo processor 2.10 GHz, 4 GB of RAM, Windows XP and R 2.8.1

Average distances reduce when the number of looks increases, i.e., improving image quality corrects the statistics in terms of test size. With the exception of the χ^2 distance, these distances vary in the interval (45.02, 51.54) regardless the sample sizes. Fixing the sample sizes while varying the number of looks L , the test sizes obey the inequalities $S_{\text{KL}} \leq S_{\text{H}} \leq S_{\text{B}} \leq S_{\text{R}}^{\beta} \leq S_{\chi^2}$.

The test based on the Kullback-Leibler distance presented the best performance in terms of both empirical test size and execution time. The other tests showed poorer performance, even with large samples and high number of looks. Small samples or number of looks yield poor tests, and the test based on the χ^2 distance is unacceptable.

4.2 IMAGE DATA ANALYSIS

The methodology for assessing test size presented in Section 3.3 is applied to the three forest samples of the E-SAR image presented in Figure 2. Following the methodology proposed by Nascimento et al. (2010), each sample is submitted to the following procedure:

- (P1) Partition the sample in disjoint blocks of size N_X ;
- (P2) For each block from (P1), split the remaining sample in disjoint blocks of size N_Y ;
- (P3) Perform the hypothesis test as described in Proposition 2 for each pair of samples with sizes N_X and N_Y .

Table 4 presents the results. Except for the test based on the χ^2 distance, all test sizes are smaller than the nominal level; i.e., the tests did not reject the null hypothesis when it is true. Although S_{χ^2} presented the worst performance in general, it is equal to zero when $N_X = N_Y = 400$, showing the importance of the sample size on the test size.

This analysis reveals that the test based on Kullback-Leibler distance has the best performance with image data. Small test sizes are important in order to prevent oversegmentation, i.e., the erroneous identification of samples from the same class as coming from different targets.

4.3 KULLBACK-LEIBLER DISTANCE ROBUSTNESS

Limited to the S_{KL} statistic, the following discussion presents a robustness study of its test size in the presence of outliers. Only one sample in the test are contaminated, and the contamination model we adopt consists in allowing each observation from this sample to come from a different distribution than the assumed with a small probability.

The uncontaminated sample is formed by observations from the $\mathcal{W}(L, \Sigma)$ distribution, while the contaminated sample is formed by observations from either this law, with probability $1 - \epsilon$, or from $\mathcal{W}(L, 1000\Sigma)$ with probability ϵ ; in our study, $\epsilon = 10^{-5}$. In other words, the cumulative distribution function of the contaminated sample is given by

$$\epsilon F_{\mathbf{X}}(\mathbf{Z}') + (1 - \epsilon) F_{\mathbf{Y}}(\mathbf{Z}').$$

Table 5 shows the empirical test sizes as well as the following additional figures: (C1) the ML estimator mean square error (MSE) for the number of looks, (C2) the relative mean square error (rMSE) for the covariance matrix estimator adapted from the ‘total relative bias’ (see Cribari-Neto et al., 2000), (C3) the ratios between mean square errors, and (C4) the distances mean and CV.

Table 4. Empirical sizes for forests.

Factors			forest-1				forest-2				forest-3			
S_{ϕ}^h	N_X	N_Y	1%	5%	$\bar{d} (\times 10^{-1})$	CV	1%	5%	$\bar{d} (\times 10^{-1})$	CV	1%	5%	$\bar{d} (\times 10^{-1})$	CV
S_{χ^2}	49	49	49.30	53.06	10^{25}	4065.70	55.27	58.37	10^{37}	4000.00	61.05	63.16	$\approx 10^{47}$	1378.40
	49	121	37.77	40.06	10^{17}	1565.32	43.76	46.42	10^{19}	2169.27	51.40	52.78	10^7	642.51
	49	400	27.12	29.38	10^{11}	1330.26	29.77	29.77	10^{27}	1144.55	43.75	43.75	10^9	396.99
	121	121	14.62	16.60	10^7	1197.95	24.21	30.00	10^{21}	1378.40	46.43	50.00	10^8	446.32
	121	400	5.80	5.80	137.97	603.05	7.84	11.76	68.36	169.33	40.00	40.00	494.91	179.09
	400	400	0.00	0.00	12.16	77.12	0.00	0.00	2.74	39.25	0.00	0.00	6.54	—
S_{KL}	49	49	0.00	0.00	17.63	56.82	0.00	0.00	15.63	47.24	0.00	0.00	46.73	57.31
	49	121	0.00	0.00	10.91	56.50	0.00	0.00	10.45	47.29	0.00	0.00	30.12	62.51
	49	400	0.00	0.00	7.81	60.15	0.00	0.00	7.47	46.14	0.00	0.00	18.07	73.62
	121	121	0.00	0.00	7.09	44.14	0.00	0.00	8.02	54.38	0.00	0.00	29.83	53.86
	121	400	0.00	0.00	4.27	47.16	0.00	0.00	5.28	60.25	0.00	0.00	8.15	52.58
	400	400	0.00	0.00	2.95	38.16	0.00	0.00	16.99	59.64	0.00	0.00	2.61	—
S_R^{β}	49	49	0.00	0.00	13.03	55.22	0.00	0.00	12.15	46.99	0.00	0.00	33.43	55.64
	49	121	0.00	0.00	5.39	55.23	0.00	0.00	5.35	47.13	0.00	0.00	14.47	59.66
	49	400	0.00	0.00	1.67	58.44	0.00	0.00	1.66	45.26	0.00	0.00	3.77	70.50
	121	121	0.00	0.00	5.38	44.49	0.00	0.00	5.97	49.79	0.00	0.00	20.83	52.79
	121	400	0.00	0.00	1.82	45.58	0.00	0.00	2.18	52.40	0.00	0.00	3.46	50.49
	400	400	0.00	0.00	2.12	39.97	0.00	0.00	2.93	37.87	0.00	0.00	1.50	—
S_B	49	49	0.00	0.00	13.55	50.80	0.00	0.00	12.84	44.45	0.00	0.00	31.06	46.92
	49	121	0.00	0.00	6.67	52.27	0.00	0.00	6.66	45.08	0.00	0.00	16.18	51.24
	49	400	0.00	0.00	3.22	56.22	0.00	0.00	3.20	43.69	0.00	0.00	6.79	63.46
	121	121	0.00	0.00	5.89	43.27	0.00	0.00	6.51	48.77	0.00	0.00	20.63	47.05
	121	400	0.00	0.00	2.51	44.55	0.00	0.00	3.00	51.67	0.00	0.00	4.68	49.27
	400	400	0.00	0.00	2.36	39.41	0.00	0.00	3.62	40.78	0.00	0.00	1.67	—
S_H	49	49	0.00	0.00	11.37	42.69	0.00	0.00	10.92	37.88	0.00	0.00	21.33	32.77
	49	121	0.00	0.00	5.83	46.05	0.00	0.00	5.86	39.78	0.00	0.00	11.91	37.01
	49	400	0.00	0.00	2.87	51.70	0.00	0.00	2.88	39.37	0.00	0.00	5.33	49.21
	121	121	0.00	0.00	5.46	40.04	0.00	0.00	5.97	45.05	0.00	0.00	15.92	35.68
	121	400	0.00	0.00	2.38	42.09	0.00	0.00	2.80	48.53	0.00	0.00	4.22	46.42
	400	400	0.00	0.00	2.29	38.33	0.00	0.00	3.05	38.90	0.00	0.00	1.64	—

The mean square error for the number of looks estimation is given by $\text{MSE}(\hat{L}_I) = \sum_{k=1}^{5500} (\hat{L}_k^{(I)} - L_I)^2 / 5500$, where $\hat{L}_k^{(I)}$ represents the obtained estimates at the k th Monte Carlo replication for population $I \in \{X, Y\}$. For the covariance matrix, the relative mean square error is

$$\text{rMSE}(\hat{\Sigma}_I) = \frac{1}{5500} \sum_{k=1}^{5500} \sum_{h=1}^3 \frac{((\hat{\Sigma}_k^{(I)})_{hh} - (\Sigma_I)_{hh})^2}{(\Sigma_I)_{kk}},$$

where $\hat{\Sigma}_k^{(I)}$ is the estimate of the covariance matrix at the k th Monte Carlo experiment for the population I . The ratios for these measures are denoted by

$$r_1 = \frac{\text{MSE}(\hat{L}_X)}{\text{MSE}(\hat{L}_Y)} \quad \text{and} \quad r_2 = \frac{\text{rMSE}(\hat{\Sigma}_X)}{\text{rMSE}(\hat{\Sigma}_Y)}.$$

Table 5. Robustness for forest-1. Mean square errors scaled by 100.

N_X	N_Y	L	1%	5%	\bar{d}	CV	$\text{MSE}(\hat{L}_X)$	$\text{MSE}(\hat{L}_Y)$	r_1	$r\text{MSE}(\hat{\Sigma}_X)$	$r\text{MSE}(\hat{\Sigma}_Y)$	r_2
49	49	4	1.82	6.87	10.60	45.48	16.28	18.72	0.87	55.36	55.52	1.00
		8	0.95	4.62	9.73	46.66	14.75	16.96	0.87	27.78	28.21	0.98
		16	0.84	4.07	9.41	47.00	14.13	16.03	0.88	13.70	14.67	0.93
49	121	4	2.09	7.80	10.71	46.06	16.23	17.54	0.93	55.05	23.74	2.32
		8	1.02	4.49	9.64	47.20	14.80	15.41	0.96	27.50	11.96	2.30
		16	0.85	4.07	9.44	46.67	14.40	14.71	0.98	13.50	6.56	2.06
49	400	4	1.67	8.07	10.79	45.20	16.31	16.92	0.96	54.53	7.85	6.94
		8	1.07	5.00	9.77	46.95	14.83	14.76	1.00	27.05	4.41	6.13
		16	0.56	4.09	9.55	45.57	14.21	13.97	1.02	13.74	2.70	5.08
121	121	4	2.04	7.64	10.75	45.86	14.90	17.59	0.85	22.24	23.26	0.96
		8	0.80	4.58	9.72	45.40	13.45	15.44	0.87	10.95	12.22	0.90
		16	0.96	4.44	9.56	46.69	12.87	14.71	0.87	5.60	6.60	0.85
121	400	4	2.15	7.69	10.84	45.84	14.88	16.97	0.88	22.08	7.78	2.84
		8	1.05	4.87	9.88	46.04	13.30	14.75	0.90	11.06	4.41	2.51
		16	1.62	5.44	9.73	48.89	12.86	13.97	0.92	5.52	2.73	2.02
400	400	4	1.93	7.87	10.91	44.79	14.31	16.91	0.85	6.54	7.88	0.83
		8	1.22	5.96	10.12	46.00	12.65	14.71	0.86	3.33	4.40	0.76
		16	1.22	6.27	10.22	46.57	12.17	14.01	0.87	1.70	2.74	0.62

The results presented in Table 5 reveal that the mean square errors are reduced when larger windows or number of looks are considered. This behavior is expected, since in those cases the signal-to-noise ratio is improved.

For situations where $N_X = N_Y = N \in \{49, 121, 400\}$, the mean square errors for the estimators of the number of looks and covariance matrix increase when contamination is introduced. Additionally, increasing the value of N does not significantly affect the ratio r_1 . Simultaneously, ratio r_2 decreases. This last fact is justified by the contamination model nature, and indicates that the contamination effect on the empirical test size is strongly related to the estimation of the covariance matrix.

Figure 5 plots S_{KL} in pure and contaminated scenarios. As expected, the contamination effect under S_{KL} increases when considering situations with larger windows. However, this effect is diminished when L increases.

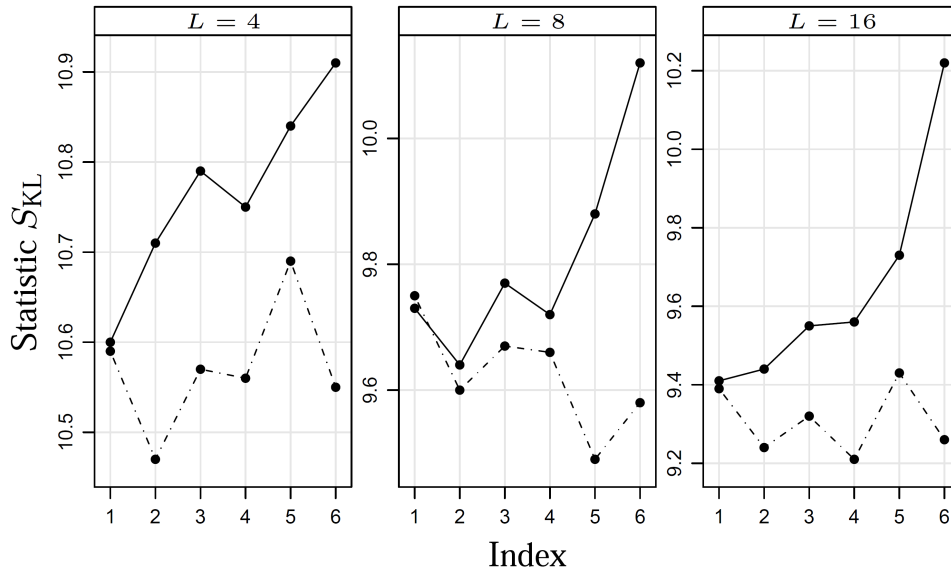


Figure 5. Stochastic distances with $\epsilon = \{0, 10^{-5}\}$. For each sample size, scenarios are lexicographically ordered: $[i : (i; N_X, N_Y) = \{(1; 49, 49), (2; 49, 121), (3; 49, 400), (4; 121, 121), (5; 121, 400), (6; 400, 400)\}]$.

5. CONCLUSIONS

This work aimed to provide a statistical formalism for the use of information theory tools in the understanding of PolSAR imagery. Five novel hypothesis tests for polarimetric speckled data were assessed. Their associate statistics were based on Bhattacharyya, chi-square, Hellinger, Kullback-Leibler and Rényi distances between maximum likelihood estimators for the parameters that index the scaled multilook complex Wishart law. As a comparison criterion, empirical test sizes were quantified in several situations, which included pure and contaminated data. An application to actual data is performed. The test based on the Kullback-Leibler distance had the closest empirical size to the nominal level. Under contamination, the maximum likelihood estimator for the covariance matrix is significantly affected, yielding the S_{KL} statistic sensitive to outliers. However, this effect is reduced increasing the number of looks, that is, this contrast measure presented better performance in the presence of contamination when signal-to-noise ratio is controlled. With the exception of the chi-square measure, all stochastic distances presented good performance when applied to a real PolSAR image. Many venues for research are still open in PolSAR image modelling and analysis. Using stochastic distances in actual segmentation and classification algorithm requires fast and reliable inference procedures which, ideally, are resistant (robust) to plausible contamination. Expressive and tractable models for spatial correlation under the Wishart law, and their implication in inference is another promising research direction. Extending the results here presented to more general models than the scaled multilook complex Wishart law, for instance to the polarimetric \mathcal{G}^0 and \mathcal{G}^H distributions, or even to the still more general complete polarimetric \mathcal{G} model (see Freitas et al., 2005; Frery et al., 2010) would greatly enhance the capability of understanding this class of images.

ACKNOWLEDGEMENTS

The authors are grateful to CNPq, Fapeal and Facepe for their support to this work.

AFTERWORD

This invited paper is an extended and detailed version of a talk given at the III Simposio de Estadística Espacial y Modelado de Imágenes (SEEMI), that took place at Foz do Iguaçu, Brazil, in December 2010. Though many results discussed here have been already published in specialized journals and conferences, in this presentation we intend to achieve the wide audience of probability and statistics, with a bias towards applications in signal and image processing.

REFERENCES

- Ali, S.M., Silvey, S.D., 1996. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 26, 131-142.
- Andai, A., 2009. On the geometry of generalized Gaussian distributions. *Journal of Multivariate Analysis*, 100, 777-793.

- Anfinssen, S.N., Doulgeris, A.P., Eltoft, T., 2009. Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3795-3809.
- Aviyente, S., 2003. Divergence measures for time-frequency distributions. *Seventh International Symposium on Signal Processing and Its Applications (SISSPA)*, 1, 121-124.
- Beaulieu, J.M., Touzi, R., 2004. Segmentation of textured polarimetric SAR scenes by likelihood approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 2063-2072.
- Broniatowski, M., Keziou, A., 2009. Parametric estimation and tests through divergences and the duality technique. *Journal of Multivariate Analysis*, 100, 16-36.
- Burbea, J., Rao, C., 1982. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *Journal of Multivariate Analysis*, 12, 575-596.
- Conradsen, K., Nielsen, A.A., Schou, J., Skriver, H., 2003. A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 4-19.
- Cribari-Neto, F., Ferrari, S.L.P., Cordeiro, G.M., 2000. Improved heteroscedasticity-consistent covariance matrix estimators. *Biometrika*, 87, 907-918.
- Csiszár, I., 1967. Information type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2, 299-318.
- Deza, M.M., Deza, E., 2009. *Encyclopedia of Distances*. Springer, New York.
- Eguchi, S., Copas, J., 2006. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *Journal of Multivariate Analysis*, 97, 2034-2040.
- Freeman, A., Durden, S.L., 1998. A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 963-973.
- Freitas, C.C., Frery, A.C., Correia, A.H., 2005. The polarimetric G distribution for SAR data analysis. *Environmetrics*, 16, 13-31.
- Frery, A.C., Cintra, R.J., Nascimento, A.D.C., 2011. Hypothesis test in complex Wishart distributions. *Proceedings of the 5th International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR'2011)*, Frascati, Italy.
- Frery, A.C., Jacobo-Berlles, J., Gambini, J., Mejail, M., 2010. Polarimetric SAR image segmentation with B-splines and a new statistical model. *Multidimensional Systems and Signal Processing*, 21, 319-342.
- Gambini, J., Mejail, M., Jacobo-Berlles, J., Frery, A.C., 2006. Feature extraction in speckled imagery using dynamic B-spline deformable contours under the G_0 model. *International Journal of Remote Sensing*, 27, 5037-5059.
- Gambini, J., Mejail, M., Jacobo-Berlles, J., Frery, A.C., 2008. Accuracy of edge detection methods with local information in speckled imagery. *Statistics and Computing*, 18, 15-26.
- Giet, L., Lubrano, M., 2008. A minimum Hellinger distance estimator for stochastic differential equations: An application to statistical inference for continuous time interest rate models. *Computational Statistics & Data Analysis*, 52, 2945-2965.
- Goodman, N.R., 1963. Statistical analysis based on a certain complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34, 152-177.
- Goudail, F., Réfrégier, P., 2004. Contrast definition for optical coherent polarimetric images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 947-951.
- Goudail, F., Réfrégier, P., Delyon, G., 2004. Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images. *Journal of the Optical Society of America A*, 21, 1231-1240.

- Hagedorn, M., Smith, P.J., Bones, P.J., Millane, R.P., Pairman, D., 2006. A trivariate chi-squared distribution derived from the complex Wishart distribution. *Journal of Multivariate Analysis*, 97, 655-674.
- Hjorungnes, A., Gesbert, D., 2007. Complex-valued matrix differentiation: Techniques and key results. *IEEE Transactions on Signal Processing*, 55, 2740-2746.
- Kailath, T., 1967. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15, 52-60.
- Kersten, P.R., Lee, J.S., Ainsworth, T.L., 2005. Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 519-527.
- Lee, J.S., Hoppel, K.W., Mango, S.A., Miller, A.R., 1994. Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 1017-1028.
- Lee, J.S., Pottier, E., 2009. *Polarimetric Radar Imaging: From Basics to Applications*. CRC-Press, Boca Raton.
- Liese, F., Vajda, I., 2006. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52, 4394-4412.
- Lopez-Martinez, C., Fabregas, X., 2003. Polarimetric SAR speckle noise model. *IEEE Transactions on Geoscience and Remote Sensing* 41, 2232-2242.
- Mak, B., Barnard, E., 1996. Phone clustering using the Bhattacharyya distance. *The Fourth International Conference on Spoken Language Processing*, 4, Philadelphia, PA, 2005-2008.
- Morio, J., Réfrégier, P., Goudail, F., Dubois-Fernandez, P.C., Dupuis, X., 2008. Information theory-based approach for contrast analysis in polarimetric and/or interferometric SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 2185-2196.
- Morio, J., Réfrégier, P., Goudail, F., Dubois Fernandez, P.C., Dupuis, X., 2009. A characterization of Shannon entropy and Bhattacharyya measure of contrast in polarimetric and interferometric SAR image. *Proceedings of the IEEE* 97, 1097-1108.
- Nascimento, A.D.C., Cintra, R.J., Frery, A.C., 2010. Hypothesis testing in speckled data with stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 373-385.
- Neuman, E., 1990. Inequalities involving multivariate convex functions II. *Proceedings of the American Mathematical Society*, 109, 965-974.
- Puig, D., Garcia, M.A., 2003. Pixel classification through divergence-based integration of texture methods with conflict resolution. *International Conference on Image Processing*, 3, 1037-1040.
- Rached, Z., Alajaji, F., Lorne Campbell, L., 2001. Renyi's divergence and entropy rates for finite alphabet Markov sources. *IEEE Transactions on Information Theory*, 47, 1553-1561.
- Salicrú, M., Menéndez, M.L., Pardo, L., Morales, D., 1994. On the applications of divergence type measures in testing statistical hypothesis. *Journal of Multivariate Analysis*, 51, 372-391.
- Schou, J., Skriver, H., Nielsen, A.H., Conradsen, K., 2003. CFAR edge detector for polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 20-32.
- Seghouane, A.K., Amari, S.I., 2007. The AIC criterion and symmetrizing the Kullback-Leibler divergence. *IEEE Transactions on Neural Networks*, 18, 97-106.
- Taneja, I.J., 2006. Bounds on triangular discrimination, harmonic mean and symmetric chi-square divergences. *Journal of Concrete and Applicable Mathematics*, 4, 91-111.
- Ulaby, F.T., Elachi, C., 1990. *Radar Polarimetry for Geoscience Applications*, Artech House, Norwood.

- Yu, J., Amores, J., Sebe, N., Radeva, P., Tian, Q., 2008. Distance learning for similarity estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 451-462.
- Zografos, K., Ferentinos, K., Papaioannou, T., 1990. ϕ -divergence statistics: Sampling properties and multinomial goodness-of-fit and divergence tests. *Communications in Statistics - Theory Methods*, 19, 1785-1802.