Statistical Papers
https://doi.org/10.1007/s00362-022-01354-4

REGULAR ARTICLE

Divergence-based tests for the bivariate gamma distribution applied to polarimetric synthetic aperture radar

Abraão Nascimento[1] · **Jodavid Ferreira**[2] · **Alisson Silva**[1]

Received: 22 January 2022 / Revised: 21 June 2022 / Accepted: 18 August 2022

Abstract

The use of polarimetric synthetic aperture radar (PolSAR) is one of the most successful tools for solving remote sensing problems. The multidimensional speckle noise encountered in the acquisition of these images is the main challenge for PolSAR users. Therefore, tailored processing of PolSAR images is required, especially for the use of hypothesis testing in change detection. In this paper, we use McKay's bivariate gamma distribution (MBG) to describe a joint distribution resulting from two components of the total scattering power image (SPAN). We derive closed form expressions for the MBG Kullback–Leibler and Rényi divergences between SPAN-based random pairs. We provide new two-sample divergence-based hypothesis tests and evaluate their performance using Monte Carlo experiments. Finally, we apply the new tests to real PolSAR images to evaluate the changes caused by urbanization processes in the Los Angeles and California regions. The results show that our proposals are able to detect changes in PolSAR images.

Keywords Detectors · Urbanization mapping · Divergence · Hypothesis tests · McKay bivariate gamma

Jodavid Ferreira and Alisson Silva have contributed equally to this work.

✉ Abraão Nascimento
abraao@de.ufpe.br

Jodavid Ferreira
jodavid.arts@gmail.com

Alisson Silva
alisson.santoss@ufpe.br

[1] Departamento de Estatística, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife, Pernambuco 50670-901, Brazil

[2] Departamento de Estatística, Universidade Federal da Paraíba, Street, João Pessoa, Paraíba 58051-900, Brazil

Published online: 30 August 2022

Mathematics Subject Classification 94A08 · 62Fxx

1 Introduction

The Synthetic Aperture Radar System (SAR) is an active sensor that transmits and receives electromagnetic waves in the microwave range (Hajnsek and Desnos 2020). Pulses from SAR penetrate clouds to detect reflected energy on the Earth's surface at repetitive intervals, allowing detection and monitoring of changes over time. There are two important steps in this process: describing the type of data captured by the mirrored waves and removing the effects of the imaging process. In this removal process, change detection is one of the most important steps (Ünsalan 2012). This is done by analysing two contiguous images of the same area taken at different time instants.

Understanding SAR images in a multidimensional and multilook perspective—called multilook polarimetric SAR (PolSAR)—has reached a promising position. PolSAR data consists of the amplitude and phase determined from the backscattered signals for combinations of linear receive and transmit polarizations: HH, HV, VH, and VV (H for horizontal polarization and V for vertical polarization) (Lee and Pottier 2009). Further, instead of a region being sensored once, multiple looks are made from under-sensored region, resulting images are interpreted as stochastic matrices whose pixels have a complex Wishart distribution. These particular matrices are Hermitian, and positive definite (López-Martínez and Fabregas 2003; Anfinsen et al. 2009).

Matrix decompositions are sought when treating with PolSAR images by furnishing means to understand features in them. There are several techniques for decomposing PolSAR images. The most commonly used is the basic Pauli decomposition, which models the polarimetric scattering matrix in three mechanisms (Lee and Pottier 2009): the surface, the double jump, and the volume diffusion. This decomposition provides a picture with information about the backscattered power, called SPAN. Specifically, the SPAN is formed from complex matrices of polarized channels (HH, VV, and HV) that receive as input information of the same scene obtained at different times (Hajnsek and Desnos 2020).

Some methods and algorithms based on SPAN decomposition have been proposed to detect changes in SAR polarimetric images: in different regions with low scattering due to radar shadows or vegetation (West and Riley 2019); in combination with spherical invariant random vector distance for heterogeneous urban areas (Wang et al. 2018); when using maximum polarimetric entropy to derive the mirror volume of images of urban and rural areas (Xu et al. 2014); in using different distances to measure the differences between adjacent pixels of an image (An et al. 2010), etc.

In image processing, statistics is used to model and define the stochastic behavior of data extracted from (Pol)SAR images. Data can follow some common probability distributions, such as the full $\mathcal{G}$ (Frery et al. 1997) for SAR intensity, the scaled complex full $\mathcal{G}$ and Wishart for PolSAR data (Freitas et al. 2005), and the multivariate gamma for multivariate intensities (Hagedorn et al. 2006). This model-based approach shows great results in i. detecting complex structures in extremely heterogeneous urban areas (Ince et al. 2012), ii. identifying changes in multitemporal SAR images using

the Dezert-Smarandache theory (Gao et al. 2013), iii. obtaining ships in sea areas based on the cumulative logarithm method (Hachicha and Chaabane 2009; Cordeiro et al. 2019), iv. modeling the amplitude and intensity of the complex reflected signal for different land cover typologies based on second-order cumulants (de Borba et al. 2019; Li et al. 2010), v. developing segmentation, classification, and detection of changes in simulated and real data (Li et al. 2011), and vi. determining the constant false alarm rate (Qin et al. 2012).

In addition to the application of probability theory, information theory also follows an important rule in PolSAR systems (Woodward 1953). We can use some of its measures to capture the randomness of data from SAR, which is very useful. There are several categories of entropies, we will focus on relative entropy or divergence in this article. Some properties of convexity related to the study of entropy can be found in Zhang et al. (2020). It is also possible to use these measures in the construction of goodness-of-fit and equality tests (Burbea and Rao 1982), in applications to statistical inference (Zografos et al. 1990), in the development of hypothesis tests for evaluating change detection and multitemporal filters (Nascimento et al. 2009), and more.

In general, this work aims to develop new techniques for detecting changes between images of PolSAR features, referred to as SPAN. To this end, we assume that the empirical joint distribution of possible cumulative intensity pairs (i.e. (HH,HH + VV),(VV, VV + HH), (HV,HV + VV), ...) follow the McKay's bivariate gamma (MBG) distribution. We derive the MBG Kullback–Leibbler (KL) and Rényi (R) divergences and propose both new two-sample hypothesis tests and change detectors as a consequence. The performance of the proposed tests is quantified by Monte Carlo experiments. Finally, change detectors are used to examine deforestation due to urbanization effects in recent images from Los Angeles and California (USA). The results show that the type of detector to be used depends on the type of the data.

This paper is arranged as follows. Section 2 presents a version of the bivariate gamma law and its connection to PolSAR image processing. In Sect. 3, we state the theoretical propositions of this paper to propose new hypothesis tests. Section 4 deals with the numerical results related to this research. Finally, we conclude this paper in Sect. 5.

2 Bivariate gamma distribution applied to SPAN-based SAR features

PolSAR measurements record the amplitude and phase of backscattered signals for possible combinations of linear reception and transmit polarization: HH, HV, VH, and VV (H for horizontal polarization and V for vertical polarization). The result is the complex scattering matrix:

$$\boldsymbol{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix}, \tag{1}$$

where $S_{AB} \in \mathbb{C}$ means the return due to transmission A and reception B. According to Lee et al. (1994), the copolarized components S_{HH} and S_{VV} are correlated, while

S_{VH} and S_{HV} have level equal to each other in Sinclair matrix. In this paper, we focus on two channels to measure the impact of high correlations in a bivariate distribution for features from the PolSAR returns.

Let $\boldsymbol{s} \in \{[S_{\mathrm{HH}}, S_{\mathrm{VV}}]', [S_{\mathrm{HH}}, S_{\mathrm{HV}}]', [S_{\mathrm{VV}}, S_{\mathrm{HV}}]'\} \subset \mathbb{C}^2$ be a vector of two complex polarizations channels, where $(\cdot)'$ denotes transpose vector. Then, L-look polarimetric covariance matrix $\boldsymbol{T}$ can be written as

$$\boldsymbol{T} := \frac{1}{L} \sum_{i=1}^{L} \boldsymbol{g}_i \, \boldsymbol{g}_i^*, \tag{2}$$

where $\boldsymbol{g}_i$ is the ith one-look sample vector and $(\cdot)^*$ is the transpose conjugate operator. The spectral decomposition of $\boldsymbol{T}$ is given by

$$\boldsymbol{T} \; = \; \lambda_1 \, \boldsymbol{e}_1 \boldsymbol{e}_1^* \; + \; \lambda_2 \, \boldsymbol{e}_2 \boldsymbol{e}_2^*, \tag{3}$$

where λ_i are the ith real eigenvalues of $\boldsymbol{T}$ and $\boldsymbol{e}_i$ are the corresponding orthonormal eigenvectors. The total scattering power image (SPAN) can be defined as the total backscattering power in Eq. (3), i.e.

$$\mathrm{SPAN} := \lambda_1 + \lambda_2 = T_{11} + T_{22}, \tag{4}$$

where T_{ij} is the (i, j) entry of $\boldsymbol{T}$. As described in the previous section, the feature SPAN is very important in PolSAR image processing. In this paper, we study the bivariate distribution for (T_{11}, SPAN), derive divergence-based tests for such law, and develop new change detectors based on the derived measures.

According to Hagedorn et al. (2006), the multivariate gamma distribution is the natural law to describe (T_{11}, SPAN), if $\boldsymbol{T}$ follows the scaled complex Wishart distribution. One of the oldest bivariate gamma distributions is the McKay's version (Nadarajah and Gupta 2006; Gupta and Nadarajah 2006). It can be understood as one of two results:

(i) $\boldsymbol{x} = (X_1, X_2) := (V, U \times V)$ such that $V \sim \Gamma(\alpha_1 + \alpha_2, 1/\gamma)$ and $U \sim \mathrm{Beta}(\alpha_1, \alpha_2)$ are independent and $U \times V \sim \Gamma(\alpha_1, 1/\gamma)$, where $\boldsymbol{x}$ is vector of random variables, "$\sim \Gamma(\cdot, \cdot)$" denotes the Gamma distribution and "$\sim$ Beta$(\cdot, \cdot)$" means the Beta law, $X_1 = V$ and $X_2 = U \times V$ or

(ii) $\boldsymbol{x} = (X_3, X_4) := (V_0, V_0 + V_1)$ such that $V_0 \sim \Gamma(\alpha_1, 1/\gamma)$ and $V_1 \sim \Gamma(\alpha_2, 1/\gamma)$ are independent, where $X_3 = V_0$ and $X_4 = V_0 + V_1$.

Although we use the construct $\boldsymbol{x}$ in the PolSAR context because of the previous discussion, assumptions (i) and (ii) have been used in various fields, e.g., hydrology to model precipitation (Yue et al. 2001) and reliability theory (Nadarajah 2005). This suggests that what will be developed here can also be applied in other contexts.

The McKay's bivariate gamma ("MBG" for short) distribution with location parameter $\alpha_1, \alpha_2 > 0$ and scale parameter $\gamma > 0$ has probability density function (pdf) given by:

$$f(x_1, x_2; \alpha_1, \alpha_2, \gamma) = \frac{1}{\gamma^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} x_1^{\alpha_1-1}(x_2 - x_1)^{\alpha_2-1} \exp\left(-\frac{x_2}{\gamma}\right), \quad (5)$$

for $0 < x_1 < x_2$, where exp(.) is the exponential function and $\Gamma(\cdot)$ is the gamma function. Here, $\mathbb{E}(X_1) = \gamma\,\alpha_1$, $\mathbb{E}(X_2) = \gamma\,(\alpha_1 + \alpha_2)$, $\mathbb{V}\text{ar}(X_1) = \alpha_1\,\gamma^2$, $\mathbb{V}\text{ar}(X_2) = (\alpha_1 + \alpha_2)\,\gamma^2$, and $\text{Cor}(X_1, X_2) = \sqrt{\alpha_1/(\alpha_2 + \alpha_1)}$, where $\mathbb{E}(\cdot)$ is the mean operator, $\mathbb{V}\text{ar}(\cdot)$ is the variance operator, and $\text{Cor}(\cdot)$ is the correlation operator. This case is denoted as $\boldsymbol{x} \sim \text{MGB}(\alpha_1, \alpha_2, \gamma)$.

Now consider to derive maximum likelihood estimates (MLEs) for $\boldsymbol{\theta} = (\gamma, \alpha_1, \alpha_2)'$. Let $\boldsymbol{X} = [(x_{11}, x_{21})', \ldots, (x_{1n}, x_{2n})']'$ be an observed sample from $(X_1, X_2) \sim \text{MBG}(\alpha_1, \alpha_2, \gamma)$, the associated log-likelihood at $\boldsymbol{\theta}$ from Eq. (5) is

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}, \boldsymbol{X}) = -n(\alpha_1 + \alpha_2)\log\gamma - n\log\Gamma(\alpha_1) - n\log\Gamma(\alpha_2) + (\alpha_1 - 1)\sum_{i=1}^{n}\log x_{1i} + (\alpha_2 - 1)\sum_{i=1}^{n}\log(x_{2i} - x_{1i}) - \frac{\sum_{i=1}^{n} x_{2i}}{\gamma}, \quad (6)$$

where n is the used sample size and log(.) is the logarithmic function. From $[\partial\ell(\boldsymbol{\theta})/\partial\gamma, \partial\ell(\boldsymbol{\theta})/\partial\alpha_1, \partial\ell(\boldsymbol{\theta})/\partial\alpha_2]|_{\gamma=\widehat{\gamma},\alpha_1=\widehat{\alpha}_1,\alpha_2=\widehat{\alpha}_2} = [0, 0, 0]$,
one has the MLE for γ, say $\widehat{\gamma}$, is $\widehat{\gamma} = \overline{x}_2/(\widehat{\alpha}_1 + \widehat{\alpha}_2)$ such that $\overline{x}_2 = \sum_{i=1}^{n} x_{2i}/n$ is the sample mean of X_2 and MLEs for α_1 and α_2, say $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, are a solution of the non-linear system

$$\begin{cases} \frac{\sum_{i=1}^{n}\log x_{1i}}{n} = \log\widehat{\gamma} + \Psi(\widehat{\alpha}_1), \\ \frac{\sum_{i=1}^{n}\log(x_{2i}-x_{1i})}{n} = \log\widehat{\gamma} + \Psi(\widehat{\alpha}_2), \end{cases} \quad (7)$$

where $\Psi(\cdot)$ is the digamma function. From now on, we will also use the abbreviation maximum likelihood estimator (MLEr) for $\widehat{\boldsymbol{\theta}}$ for asymptotic theory purposes.

3 New divergence-based hypothesis tests for the bivariate gamma distribution

3.1 Statistical information theory background

The seminal idea on information theory was proposed by Hartley in the 1930s (Hartley 1928), who introduced a logarithmic measure of information in communication. Later Shannon (1948) formalized this idea and defined the concepts of entropy and mutual information. Finally, the notion of relative entropy (which was later called *divergence*) was introduced by Kullback and Leibler (1951). The Kullback–Leibler measure can be understood like a comparison criterion between two distributions. In statistics it occurs as the expected logarithm of the likelihood ratio: Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two random vectors with pdf's $f_{\boldsymbol{x}}(\boldsymbol{z}; \boldsymbol{\theta}_x)$ and $f_{\boldsymbol{y}}(\boldsymbol{z}; \boldsymbol{\theta}_y)$ and common support $\boldsymbol{\mathcal{Z}} \subset \mathbb{R}^p$. The divergence of

$\boldsymbol{x}$ with respect to $\boldsymbol{y}$ is defined by

$$D_{\mathrm{KL}}(\boldsymbol{x} \mid\mid \boldsymbol{y}) = \mathrm{E}_{\boldsymbol{x}}\left[\log \frac{f_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{\theta}_x)}{f_{\boldsymbol{y}}(\boldsymbol{x};\boldsymbol{\theta}_y)}\right] = \int_{\boldsymbol{\mathcal{Z}}} f_{\boldsymbol{x}}(z;\boldsymbol{\theta}_x) \log \frac{f_{\boldsymbol{x}}(z;\boldsymbol{\theta}_x)}{f_{\boldsymbol{y}}(z;\boldsymbol{\theta}_y)}\,\mathrm{d}z, \tag{8}$$

where $\mathrm{d}z = \prod_{i=1}^{p} \mathrm{d}z_i$ and $\mathrm{E}_{\boldsymbol{x}}[\mathcal{V}(\boldsymbol{x})] = \int_{\boldsymbol{\mathcal{Z}}} \mathcal{V}(z)\, f_{\boldsymbol{x}}(z;\boldsymbol{\theta}_x)\,\mathrm{d}z$ and $\mathcal{V}(\cdot)$ is a function whose expectation exists. It is noticeable that the Eq. (8) is always non-negative and is zero only if $f_{\boldsymbol{x}}(z\,;\,\boldsymbol{\theta}_x) = f_{\boldsymbol{y}}(z;\boldsymbol{\theta}_y)$ for all $z \in \boldsymbol{\mathcal{Z}}$. The divergence $D_{\mathrm{KL}}(\boldsymbol{x} \mid\mid \boldsymbol{y})$ is a quantity expressing the inefficiency of assuming the distribution of $\boldsymbol{y}$ when the model $\boldsymbol{x}$ is true (Cover and Thomas 1991). Indeed, such a measure is known under various terms, such as cross entropy, information divergence, and information for discrimination. Kullback (1968), Csiszár (1967) and Amari (1985) have presented a systematic and comprehensive work on the Kullback–Leibler divergence measure. Although $D_{\mathrm{KL}}(\cdot \mid\mid \cdot)$ is called a "distance" in some works, it is not a true distance between distributions because it is not symmetric and does not satisfy the triangle inequality. In this paper we consider a symmetrization of this measure (understood as a *distance measure* in distinguishing between probability distributions (Kailath 1967)): Given two random vectors that have the same support, the Kullback–Leibler distance can be defined as follows:

$$\begin{aligned} d_{\mathrm{KL}}(\boldsymbol{x}, \boldsymbol{y}) &= \frac{1}{2}\left[\, D_{\mathrm{KL}}(\boldsymbol{x} \mid\mid \boldsymbol{y}) + D_{\mathrm{KL}}(\boldsymbol{y} \mid\mid \boldsymbol{x})\,\right] \\ &= \frac{1}{2}\int_{\boldsymbol{\mathcal{Z}}} \left[\, f_{\boldsymbol{x}}(z\,;\,\boldsymbol{\theta}_x) - f_{\boldsymbol{y}}(z\,;\,\boldsymbol{\theta}_y)\,\right] \log \frac{f_{\boldsymbol{x}}(z\,;\,\boldsymbol{\theta}_x)}{f_{\boldsymbol{y}}(z\,;\,\boldsymbol{\theta}_y)}\,\mathrm{d}z. \end{aligned} \tag{9}$$

Other important measure that extends the KL divergence is the Rényi divergence (van Erven and Harremoes 2010), which is defined as: Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two p-points random vectors that share commun support $\boldsymbol{\mathcal{Z}} \subseteq \mathbb{R}^p$ and are indexed by parameters vectors $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$,

$$D^1_{R:r}(\boldsymbol{x} \mid\mid \boldsymbol{y}) = D^1_{R:r}(\boldsymbol{\theta_1} \mid\mid \boldsymbol{\theta_2}) = \frac{1}{r-1}\int_{\boldsymbol{\mathcal{Z}}} f^r_{\boldsymbol{x}}(z)\, f^{1-r}_{\boldsymbol{x}}(z)\,\mathrm{d}z, \tag{10}$$

where $r > 0$ and $r \neq 1$ and $\mathrm{d}z = \prod_{i=1}^{p} z_i$. If r tends to one, then the Rényi divergence degenerates in the KL divergence. Liese and Vajda (1987) extended the previous version of the Rényi divergence as (for $r \in \mathbb{R}$ and $r \neq \{0, 1\}$)

$$D_{R:r}(\boldsymbol{x} \mid\mid \boldsymbol{y}) = D_{R:r}(\boldsymbol{\theta_1} \mid\mid \boldsymbol{\theta_2}) = \frac{1}{r\,(r-1)}\int_{\boldsymbol{\mathcal{Z}}} f^r_{\boldsymbol{x}}(z)\, f^{1-r}_{\boldsymbol{x}}(z)\,\mathrm{d}z. \tag{11}$$

It is known that $D_{R:r}(\boldsymbol{x} \mid\mid \boldsymbol{y})$ is not symmetric, the next measure is employed to outperform this problem (Valdevino Félix de Lima 2021):

$$d_{R:r}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{r(r-1)} \log\left\{\frac{\exp[r\,(r-1)\,D_{R:r}(\boldsymbol{\theta}_1 \,||\, \boldsymbol{\theta}_2)] + \exp[r\,(r-1)\,D_{R:r}(\boldsymbol{\theta}_2 \,||\, \boldsymbol{\theta}_1)]}{2}\right\}. \quad (12)$$

To investigate the asymptotic distribution of the measures (8)–(12), Salicrú et al. (1994) proposed to understand them as elements within a class of divergences called (h, ϕ)-*divergence*. This class is an extension of the ϕ-*divergence* proposed by Csiszár (1967) and Ali and Silvey (1966). In the following, we present a statistical treatment for hypothesis tests based on the Kullback–Leibler divergence and distance measures.

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two previously defined random vectors. Assuming that both have a common support $\boldsymbol{\mathcal{Z}} \subset \mathbb{R}^p$, the (h, ϕ)-divergence between $f_{\boldsymbol{x}}(z\,;\,\boldsymbol{\theta}_x)$ and $f_{\boldsymbol{y}}(z\,;\,\boldsymbol{\theta}_y)$ is defined by

$$\begin{aligned} D_{\phi}^{h}(\boldsymbol{x} \,||\, \boldsymbol{y}) &= h\left(\mathrm{E}_{\boldsymbol{y}}\left[\phi\left(\frac{f_{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\theta}_{\boldsymbol{x}})}{f_{\boldsymbol{y}}(\boldsymbol{y};\boldsymbol{\theta}_{\boldsymbol{y}})}\right)\right]\right) \\ &= h\left(\int_{\boldsymbol{\mathcal{Z}}} \phi\left(\frac{f_{\boldsymbol{x}}(z;\boldsymbol{\theta}_z)}{f_{\boldsymbol{y}}(z;\boldsymbol{\theta}_{\boldsymbol{y}})}\right) f_{\boldsymbol{y}}(z;\boldsymbol{\theta}_{\boldsymbol{y}})\,\mathrm{d}z\right), \end{aligned} \quad (13)$$

where $\phi : (0, \infty) \to [0, \infty)$ is a convex function, $h : (0, \infty) \to [0, \infty)$ is a strictly increasing function with $h(0) = 0$, and indeterminate forms are assigned value zero. By a judicious choice of functions $h(x)$ and $\phi(x)$, known and unknown divergence measures arise:

$$[h(x), \phi(x)] = \begin{cases} [x,\ x\,\log(x)], & \text{for } D_{\mathrm{KL}}, \\ \left[1/2\,x,\ (x-1)\,\log(x)\right], & \text{for } d_{\mathrm{KL}}, \\ \left[\frac{1}{r(r-1)}\,\log(r\,(r-1)\,x\,+\,1),\ \frac{x^r - r\,(x-1) - 1}{r\,(r-1)}\right], & \text{for } D_{\mathrm{R}:r}, \\ \left[\frac{1}{r(r-1)}\,\log(\frac{r\,(r-1)}{2}\,x\,+\,1),\ \frac{x^r + x^{1-r} - x - 1}{r\,(r-1)}\right], & \text{for } d_{\mathrm{R}:r}. \end{cases}$$

When considering the definition of distances in terms of the h and ϕ functions combined with the results proposed by Salicrú et al. (1994) on the convergence *in distribution* of the (h, ϕ)-measures to the chi-squared distribution, the subsequent lemma holds.

Lemma 1 *Assume the regularity conditions proposed in Salicrú et al. (1994, p. 380) hold. If* $\frac{m}{m+n} \xrightarrow[m,n\to\infty]{} \lambda \in (0, 1)$ *and* $\boldsymbol{\theta}_x = \boldsymbol{\theta}_y$, *then*

$$\frac{2mn}{m+n}\,\frac{d_{\phi}^{h}(\widehat{\boldsymbol{\theta}}_x, \widehat{\boldsymbol{\theta}}_y)}{h^{(1)}(0)\phi^{(2)}(1)} \xrightarrow[m,n\to\infty]{\mathcal{D}} \chi^2_M, \quad (14)$$

where $f^{(k)}(x) = \mathrm{d}^k f(x)/\mathrm{d}x^k$ *is the kth-order derivative of* $f(x)$, "$\xrightarrow{\mathcal{D}}$" *denotes convergence in distribution,* M *is the dimension of* $\boldsymbol{\theta}_i$, m *is the first sample size,* n *is the second sample size,* $\widehat{\boldsymbol{\theta}}_x$ *is the MLEr for* $\boldsymbol{\theta}_x$ *based on the first sample,* $\widehat{\boldsymbol{\theta}}_y$ *is the MLEr*

for $\boldsymbol{\theta}_y$ based on the second sample and "$\sim \chi^2_M$" denotes a distributed chi-squared (with degree of freedom M) variable.

Based on Lemma 1, statistical hypothesis tests for the null hypothesis $\boldsymbol{\theta}_x = \boldsymbol{\theta}_y$ can be derived. In particular, the following statistic is considered:

$$S^h_\phi(\widehat{\boldsymbol{\theta}_x}, \widehat{\boldsymbol{\theta}_y}) = \frac{2mnv}{m+n} d^h_\phi(\widehat{\boldsymbol{\theta}_x}, \widehat{\boldsymbol{\theta}_y}), \tag{15}$$

where $v = 1/\left(h^{(1)}(0)\phi^{(2)}(1)\right)$ is a constant that depends on the chosen distance. For D_{KL}, d_{KL}, $D_{\mathrm{R}:r}$ and $d_{\mathrm{R}:r}$, v takes values $1/2$, 1, 1 and 1, respectively. Having paved this way, a general hypothesis test is given in the form of subsequent proposition.

Proposition 2 *Let m and n assume large values and $q = S^h_\phi(\widehat{\boldsymbol{\theta}_x}, \widehat{\boldsymbol{\theta}_y})$, then the null hypothesis $\boldsymbol{\theta}_x = \boldsymbol{\theta}_y$ can be rejected at a level η if $\Pr\left(\chi^2_M > q\right) \leq \eta$, where q is the test statistics calculated from observed samples and η is the pre-specified nominal level.*

3.2 New divergence-based hypothesis tests for the bivariate gamma distribution

Several works have been indicated that the Kullback–Leibler divergence is important to quantify contrasts in PolSAR imagery, cf. Ref. Frery et al. (2014) and Bouhlel and Méric (2020)). We begin by deriving this divergence in the following theorem.

Theorem 3 *Let $\boldsymbol{\theta_1} = (\alpha_{11}, \alpha_{12}, \gamma_1)$ and $\boldsymbol{\theta_2} = (\alpha_{21}, \alpha_{22}, \gamma_2)$ be parameter vectors of two random pairs, say $\boldsymbol{x}$ and $\boldsymbol{y}$, following McKay's bivariate gamma distribution.*

The Kullback–Leibler divergence between $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\begin{aligned} D_{KL}(\boldsymbol{x} \mid\mid \boldsymbol{y}) &= D_{KL}(\boldsymbol{\theta}_1 \mid\mid \boldsymbol{\theta}_2) = \log \frac{\gamma_2^{\alpha_{21}+\alpha_{22}} \Gamma(\alpha_{21})\Gamma(\alpha_{22})}{\gamma_1^{\alpha_{11}+\alpha_{12}} \Gamma(\alpha_{11})\Gamma(\alpha_{12})} \\ &+ \Big\{ (\alpha_{11} - \alpha_{21}) \left[\Psi(\alpha_{11}) + \log \gamma_1\right] \\ &+ (\alpha_{12} - \alpha_{22}) \left[\log \gamma_1 + \Psi(\alpha_{12})\right] - \left(\frac{\gamma_1 + \gamma_2}{\gamma_1 \gamma_2}\right) (\alpha_{11} + \alpha_{12}) \Big\}, \end{aligned}$$

where α_{1i}, $i = 1, 2$, are the location parameters for the first considered distribution, α_{2i}, $i = 1, 2$, are the location parameters for the second considered distribution and γ_i, $i = 1, 2$, is the scale parameter for ith distribution.

The proof is presented in Appendix A. In what follows, we present the associated KL distance.

Corollary 4 *Let $\boldsymbol{\theta_1} = (\alpha_{11}, \alpha_{12}, \gamma_1)$ and $\boldsymbol{\theta_2} = (\alpha_{21}, \alpha_{22}, \gamma_2)$ in the context of Theorem 3, we have*

$$
\begin{aligned}
d_{KL}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= \frac{1}{2}\left[D_{KL}(\boldsymbol{\theta_1} \,||\, \boldsymbol{\theta_2}) + D_{KL}(\boldsymbol{\theta_2} \,||\, \boldsymbol{\theta_1})\right] \\
&= \frac{1}{2}\Big\{ (\alpha_{11} - \alpha_{21})\left[\Psi(\alpha_{11}) + \log\gamma_1 - \Psi(\alpha_{21}) - \log\gamma_2\right] \\
&\quad + (\alpha_{12} - \alpha_{22})\left[\Psi(\alpha_{12}) + \log\gamma_1 - \Psi(\alpha_{22}) - \log\gamma_2\right] \\
&\quad - (\alpha_{11} + \alpha_{12} + \alpha_{21} + \alpha_{22})\left[\frac{\gamma_2 + \gamma_1}{\gamma_1\gamma_2}\right]\Big\}.
\end{aligned}
$$

The next theorem approaches the Rényi divergence expression for the MBG distribution.

Theorem 5 *Let* $\beta_1 = \alpha_{21} + r\alpha_{11} - r\alpha_{21}$, $\beta_2 = \alpha_{22} + r\alpha_{12} - r\alpha_{22}$ *and* $\eta = \beta_1 + \beta_2$. *Let* $\boldsymbol{\theta_1} = (\alpha_{11}, \alpha_{12}, \gamma_1)$ *and* $\boldsymbol{\theta_2} = (\alpha_{21}, \alpha_{22}, \gamma_2)$ *in the context of Theorem 3, we have*

$$
\begin{aligned}
&D_{R:r}(\boldsymbol{\theta_1} \,||\, \boldsymbol{\theta_2}) \\
&= \frac{1}{r(1-r)}\Big\{ \log\Gamma(\eta_1) + \log\Gamma(\eta_2) + (1-r)\Big[\log\Gamma(\alpha_{21}) + \log\Gamma(\alpha_{22}) \\
&\quad - (\alpha_{21} + \alpha_{22})\left[\log\gamma_1 - \log(\gamma_1 - \gamma_1 r + \gamma_2 r)\right]\Big] \\
&\quad + r\Big[(\alpha_{11} + \alpha_{12})(\log\gamma_2 - \log(\gamma_1 - \gamma_1 r + \gamma_2 r) - \log\Gamma(\alpha_{11}) + \log\Gamma(\alpha_{12})\Big]\Big\}.
\end{aligned}
$$

The proof of this result is given in Appendix B. In what follows, we discuss the Rényi distance.

Corollary 6 *Let* $\boldsymbol{\theta}_1 = (\alpha_{11}, \alpha_{12}, \gamma_1)$ *and* $\boldsymbol{\theta}_2 = (\alpha_{21}, \alpha_{22}, \gamma_2)$ *in the context of Theorem 3, we have*

$$
\begin{aligned}
&d_{R:r}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= -\frac{\log 2}{r(r-1)} + \frac{1}{r(r-1)}\log\Big\{\Gamma(\beta_{21})\,\Gamma(\beta_{22}) \\
&\quad \times (1-r)\Big[\Gamma(\alpha_{21})\,\Gamma(\alpha_{22})\,\gamma_1^{-(\alpha_{21}+\alpha_{22})}\,(\gamma_1 - \gamma_1 r + \gamma_2 r)^{-(\alpha_{21}+\alpha_{22})}\Big] \\
&\quad \times r\left[\gamma_2^{\alpha_{11}+\alpha_{12}}\,(\gamma_1 - \gamma_1 r + \gamma_2 r)^{-(\alpha_{11}+\alpha_{12})}\,\Gamma(\alpha_{11})^{-1}\,\Gamma(\alpha_{12})^{-1}\right] \\
&\quad + \Gamma(\beta_{11})\,\Gamma(\beta_{12})\,r\,[\gamma_1^{\alpha_{21}+\alpha_{22}}\,(\gamma_2 - \gamma_2 r + \gamma_1 r)^{-(\alpha_{21}+\alpha_{22})}\,\Gamma(\alpha_{21})^{-1}\,\Gamma(\alpha_{22})^{-1}] \\
&\quad \times (1-r)\Big[\Gamma(\alpha_{11})\,\Gamma(\alpha_{12})\,\gamma_2^{-(\alpha_{11}+\alpha_{12})}\,(\gamma_2 - \gamma_2 r + \gamma_1 r)^{-(\alpha_{11}+\alpha_{12})}\Big]\Big\},
\end{aligned}
$$

where $\beta_{1i} = \alpha_{1i} + r\alpha_{1i} - r\alpha_{2i}$, $\beta_{2i} = \alpha_{2i} + r\alpha_{2i} - r\alpha_{1i}$ *for* $i = 1, 2$.

It follows from Lemma 1 and Proposition 2 (taking $M = 3$) combined with Theorems 3 and 5 and Corollaries 4 and 6, two new hypothesis tests for the MBG distribution.

- The one based on Kullback–Leibler statistics—$S_{KL}(\cdot,\cdot)$:

$$S_{KL}([\widehat{\alpha}_{11},\widehat{\alpha}_{12},\widehat{\gamma}_1],[\widehat{\alpha}_{21},\widehat{\alpha}_{22},\widehat{\gamma}_2]) = \frac{2\,m\,n}{m\,+\,n}\,d_{KL}([\widehat{\alpha}_{11},\widehat{\alpha}_{12},\widehat{\gamma}_1],[\widehat{\alpha}_{21},\widehat{\alpha}_{22},\widehat{\gamma}_2]) \xrightarrow[m,n\to\infty]{\mathcal{D}} \chi^2_3. \tag{16}$$

- The one based on Rényi statistics with order r - $S^r_R(\cdot,\cdot)$:

$$S^r_R([\widehat{\alpha}_{11},\widehat{\alpha}_{12},\widehat{\gamma}_1],[\widehat{\alpha}_{21},\widehat{\alpha}_{22},\widehat{\gamma}_2]) = \frac{2\,m\,n}{m\,+\,n}\,d^r_R([\widehat{\alpha}_{11},\widehat{\alpha}_{12},\widehat{\gamma}_1],[\widehat{\alpha}_{21},\widehat{\alpha}_{22},\widehat{\gamma}_2]) \xrightarrow[m,n\to\infty]{\mathcal{D}} \chi^2_3. \tag{17}$$

Figure 1 shows curves of the derived divergence $D_\bullet(\boldsymbol{\theta_1} \,||\, \boldsymbol{\theta_2}(x))$ for the MBG law, where $\boldsymbol{\theta_1} = (1, 2, 1)$ and $\boldsymbol{\theta_2}(x) = (1, x, 1)$ and x varies in the range $(0, 8)$. It is noticeable the Rényi divergence converges to the KL divergence when r tends to one, as expected. Small values of r become the increasing of the Rényi divergence more sensible for $x < 2$, while, for high values of r, the Rényi divergence is more sensible for $x > 2$.

4 Numerical results

The aim of this section is twofold. First, we conduct a simulation study to evaluate the behavior of the proposed tests. Second, we apply these tests to solve change detection problems in real data.

4.1 Simulation study

We quantify the performance of statistical tools proposed to construct change detectors for PolSAR images: divergence-based two-sample tests. In particular, Monte Carlo experiments are used to study tests with Kullback–Leibler and Rényi (with order $\beta = 0.2, 0.5, 0.9,$ and 1.5) statistics given in Proposition 2. As for the choice of β values, in the context of classifying PolSAR data, we have found that the Rényi order $(0, 1)$ works best (Rodríguez et al. 2020; Ferreira et al. 2021). So we here choose to define the majority of Rényi's orders in $(0, 1)$. We use one thousand Monte Carlo replicates and, for each of them, the sample sizes $N = 3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9,$ and 11×11. In practice, defining large windows in images means mixing statistically different areas, which precludes change detection. However, in terms of statistical inference, sample size variation is a form for testing how the asymptotic properties of tests behave. Although we focus on the smallest window size in the application section, we chose to use other four sizes in this study to provide insight into the

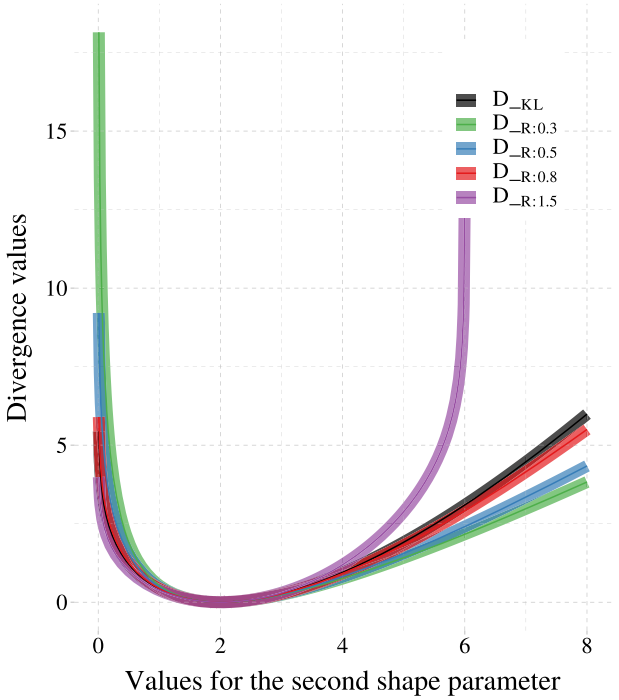


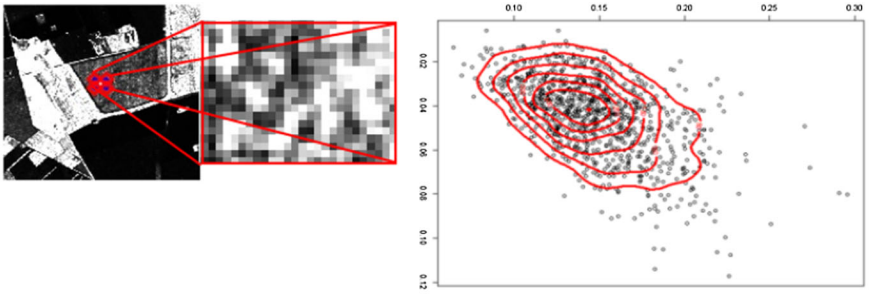
Fig. 1 Distance measures between two random variables of the Bivariate Gamma distribution as a function of α_1

Table 1 Parameter value for simulated scenarios

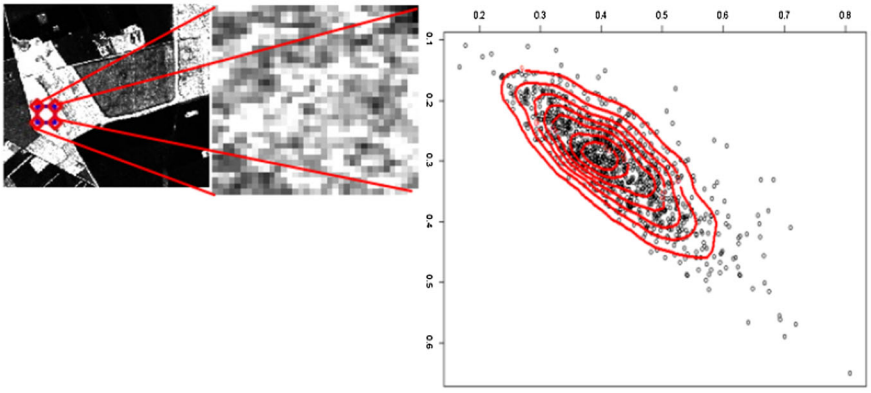
	α_1	α_2	γ	Correlation $\left(\sqrt{\frac{\alpha_1}{\alpha_1+\alpha_2}}\right)$
Scene 1	6.443582280	14.915205474	0.006888375	54.92%
Scene 2	15.61487646	6.24223856	0.01936444	84.52%
Scene 3	2.762274772	8.125188024	0.001330168	50.36%

utilization of our proposals in other contexts (e.g., hydrology and reliability). As figure of merit, we use the empirical test sizes at the nominal levels of 1%, 5%, and 10%, frequently used values in statistical studies. With respect to parameters in $H_0 : (\alpha_{11}, \alpha_{12}, \gamma_1) = (\alpha_{21}, \alpha_{22}, \gamma_2)$, they are defined using MLEs obtained from a Foulum (Dinarmark) region recorded by the Electromagnetics Institute of SAR (EMISAR) sensor, captured in the L-band and quad-pol under the number of looks four. We can observe the features of each sample in Table 1, which shows the values of the estimates of $(\alpha_1, \alpha_2, \gamma)$ and associated correlations for each subregion in Fig. 2a–c.

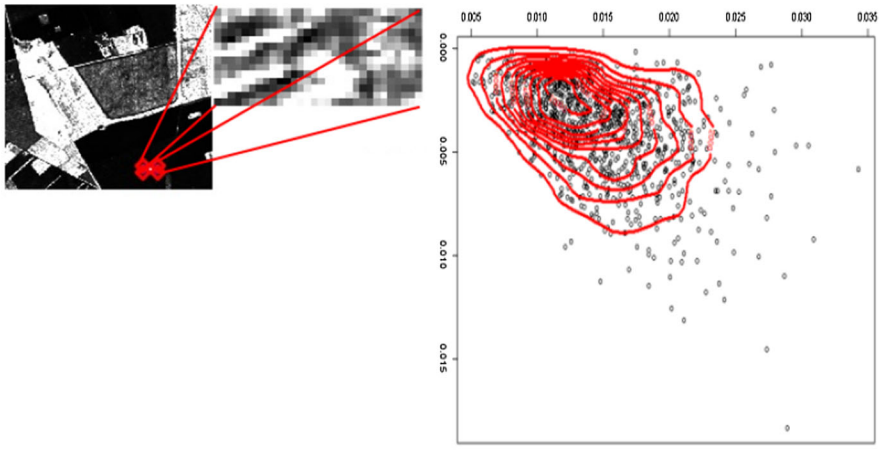
It is noticeable that the wide range of the dispersion of the pairs is related to the high value of γ, in Fig. 2b. In terms of magnitude of the shape, Scene 2 (*forest area* discussed in Skriver et al. (2005)) has a higher shape than that of Scene 1 (*winter wheat* in Wang et al. (2017)), which has a higher shape than that of Scene 3 (*broad leaves*



(a) Scene 1.



(b) Scene 2.



(c) Scene 3

Fig. 2 Simulation scene obtained from Foulum region (Denmark) based on the HH channel intensity, dispersion plot of pair values (HH, HH + VV), and associated empirical level curves

crops in Zhang et al. (2015)). It is worth noting that the values of γ are related directly to those of the associated correlations. Looking at the level curves of extracted regions in Fig. 2a–c, it can be seen that Scene 2 has a more pronounced signal structure (longer level curve) than Scene 1 and the latter is more accentuated than Scene 3, where the strength of the speckle noise is highest. This is confirmed by the obtained values of the correlation.

Table 2 shows the empirical test sizes, i.e., the rejection rates under the null hypothesis of the proposed tests at nominal levels of 1%, 5%, and 10%. The best results are highlighted in Bold. It can be seen that the test based on $S_R^{0.5}$ gave the best result, as its values were closest to the assumed nominal levels. The tests based on $S_R^{1.5}$ and S_{KL} statistics performed very poorly. Analyzing the performance of proposed tests by region, we can highlight the following tests:

- At the nominal 10% level, $S_R^{0.2}$, $S_R^{0.5}$, and $S_R^{0.9}$ were associated with the first scenario; for the second scenario, S_{KL}, $S_R^{0.2}$, and $S_R^{0.9}$; and for the third scenario, only $S_R^{0.5}$.
- With respect to the nominal 5% level, the estimates obtained from $S_R^{0.2}$ and $S_R^{0.5}$ in the first scenario differ from the others; for the second scene, only the $S_R^{1.5}$ test performs poorly when $N = 81$; and, similarly to the previous level, the estimate generated with $S_R^{0.5}$ for the third scene shows an excellent result.
- At the 1% level, the test statistics $S_R^{0.2}$, $S_R^{0.5}$ and $S_R^{0.9}$ obtain the best estimates for the second scenario; $S_R^{1.5}$ performs worst among the other estimates for the first scene.

The precedence of speckle noise over signal (and vice versa) had no discernible effect on the simulated results, suggesting that our proposals are little affected by speckle and the kind of signal in controlled data.

4.2 Application to real data

We make an application to real data extracted from PolSAR imagery obtained from the Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) sensor. We aim to solve change detection issues using the proposed statistics. The problem in this domain is formulated as follows: From two coregistered PolSAR images taken at two different points in time, create a map showing (i) a change in image quality (known as supervised) or (ii) a change in the type of SAR data (unsupervised). Figure 3 shows a flowchart illustrating how change detection works in practice.

We recorded two scenes captured by the UAVSAR sensor: The first one refers to an area in Los Angeles (USA) and the second region refers to an area in California (USA). For each region, there are two images taken at different times. The first pair—seen in Fig. 5a, b—was taken on 23 April 2009 and 3 May 2015, respectively. The second pair shown in Fig. 6a, b—were taken on 23 April 2009 and 3 May 2015, respectively. In the following, the proposed change detectors are used to detect changes in the region that occurred between the two different time periods. We will apply the proposed methods on real images to verify the efficiency of the empirical tests in complex scenarios, such as in urban areas and dense forests, which are directly affected by speckle noise,

Table 2 Rejection rates of distance-based tests under $H_0 : (\alpha_{11}, \alpha_{12}, \gamma_1) = (\alpha_{21}, \alpha_{22}, \gamma_2)$

Scene	N	10% nominal level					5% nominal level					1% nominal level				
		KL	$R_{0.2}$	$R_{0.5}$	$R_{0.9}$	$R_{1.5}$	KL	$R_{0.2}$	$R_{0.5}$	$R_{0.9}$	$R_{1.5}$	KL	$R_{0.2}$	$R_{0.5}$	$R_{0.9}$	$R_{1.5}$
1	9	0.172	0.16	**0.149**	0.165	0.240	0.113	0.095	**0.089**	0.103	0.240	0.045	0.033	**0.025**	0.039	0.240
	25	0.122	0.117	**0.115**	0.118	0.153	0.067	0.061	**0.060**	0.065	0.091	0.017	0.016	**0.015**	0.016	0.036
	49	0.110	0.109	**0.108**	0.110	0.119	0.055	**0.054**	**0.054**	0.055	0.065	**0.009**	**0.009**	**0.009**	**0.009**	0.012
	81	**0.101**	**0.099**	0.098	**0.099**	0.107	0.053	0.051	**0.050**	0.052	0.063	**0.010**	**0.010**	0.009	**0.010**	0.016
	121	0.101	**0.100**	**0.100**	**0.100**	0.106	0.052	**0.049**	0.048	**0.051**	0.059	0.011	0.011	**0.01**	0.011	0.013
2	9	0.169	0.156	**0.151**	0.164	0.246	0.110	0.089	**0.082**	0.095	0.246	0.039	0.029	**0.022**	0.033	0.246
	25	0.111	0.106	**0.102**	0.107	0.148	0.056	0.052	**0.051**	0.053	0.087	0.012	**0.011**	**0.011**	0.012	0.033
	49	0.114	0.113	**0.111**	0.114	0.124	0.067	0.065	**0.062**	0.065	0.077	0.020	**0.018**	**0.018**	**0.018**	0.023
	81	**0.100**	**0.099**	0.097	**0.099**	0.107	**0.049**	**0.049**	**0.049**	**0.049**	0.055	0.011	**0.010**	**0.010**	**0.010**	0.014
	121	0.114	0.112	**0.111**	0.113	0.118	0.060	0.060	**0.059**	0.060	0.071	**0.013**	**0.013**	**0.013**	**0.013**	0.015
3	9	0.078	0.063	**0.060**	0.068	0.228	0.048	0.041	**0.038**	0.041	0.228	0.014	0.011	**0.008**	0.012	0.228
	25	0.069	0.067	**0.066**	0.067	0.067	0.036	0.033	**0.032**	0.033	0.033	0.007	0.006	**0.005**	0.007	0.007
	49	0.104	0.101	**0.100**	0.102	0.125	0.051	0.051	**0.05**	0.051	0.070	0.014	0.014	0.014	0.014	0.018
	81	0.111	0.109	**0.108**	0.111	0.118	0.053	**0.052**	**0.052**	**0.052**	0.060	**0.008**	**0.008**	0.007	**0.008**	0.017
	121	0.110	0.109	**0.108**	0.109	0.118	0.053	**0.051**	**0.051**	0.052	0.057	**0.014**	**0.014**	**0.014**	**0.014**	0.018

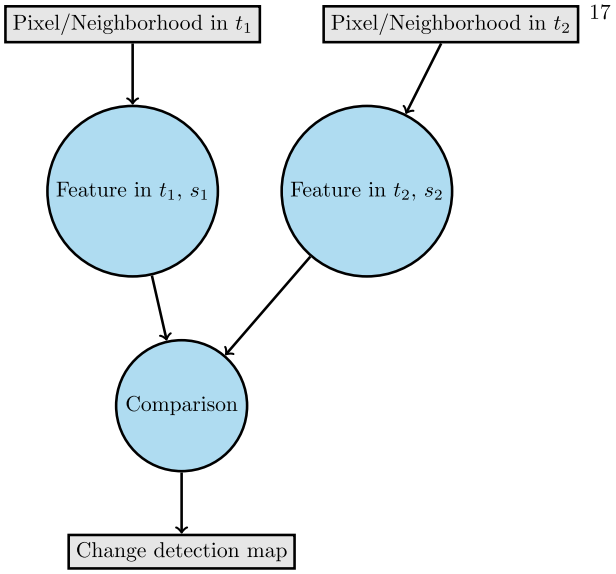
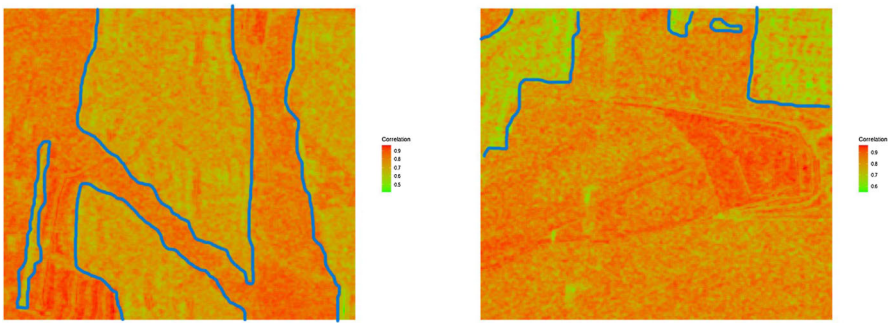


Fig. 3 Flowchart containing steps of a change detection process



(a) First region. **(b)** Second region.

Fig. 4 Maps of the MGB correlation for Los Angeles and California images, captured in 2015

as well as facilitate the sampling process as the algorithm attempts to clearly describe the image data in the change detection process. To obtain the results, we use a window of size 3×3 in the neighborhood of pixels and the divergence-based tests under the significance level 1×10^{-8} (lowest level, below which there is no change). Figure 4 shows MGB correlation maps quantified for the Los Angeles and California regions on May 3, 2015. The first map shows the image that is most noisy and has regions with low correlations.

Our goal is to quantify the effect of urbanization. Two methods are used as change detectors:

- CD_{KL}: Kullback–Leibler distance-based change detector with the bivariate gamma distribution.

- CD_R^{α}: Rényi distance-based change detector at order α with bivariate gamma distribution.

To confirm the qualitative discussion of predictive maps, we use five metrics to evaluate the performance of the detectors:

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 $A = 1 - p_{FP} - p_{FN}$ and $B = (p_{TP} + p_{FP})(p_{TP} + p_{FN}) + (p_{TN} + p_{FP})(p_{TN} + p_{FN})$, where p_C is the proportion of pixels under the condition 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.

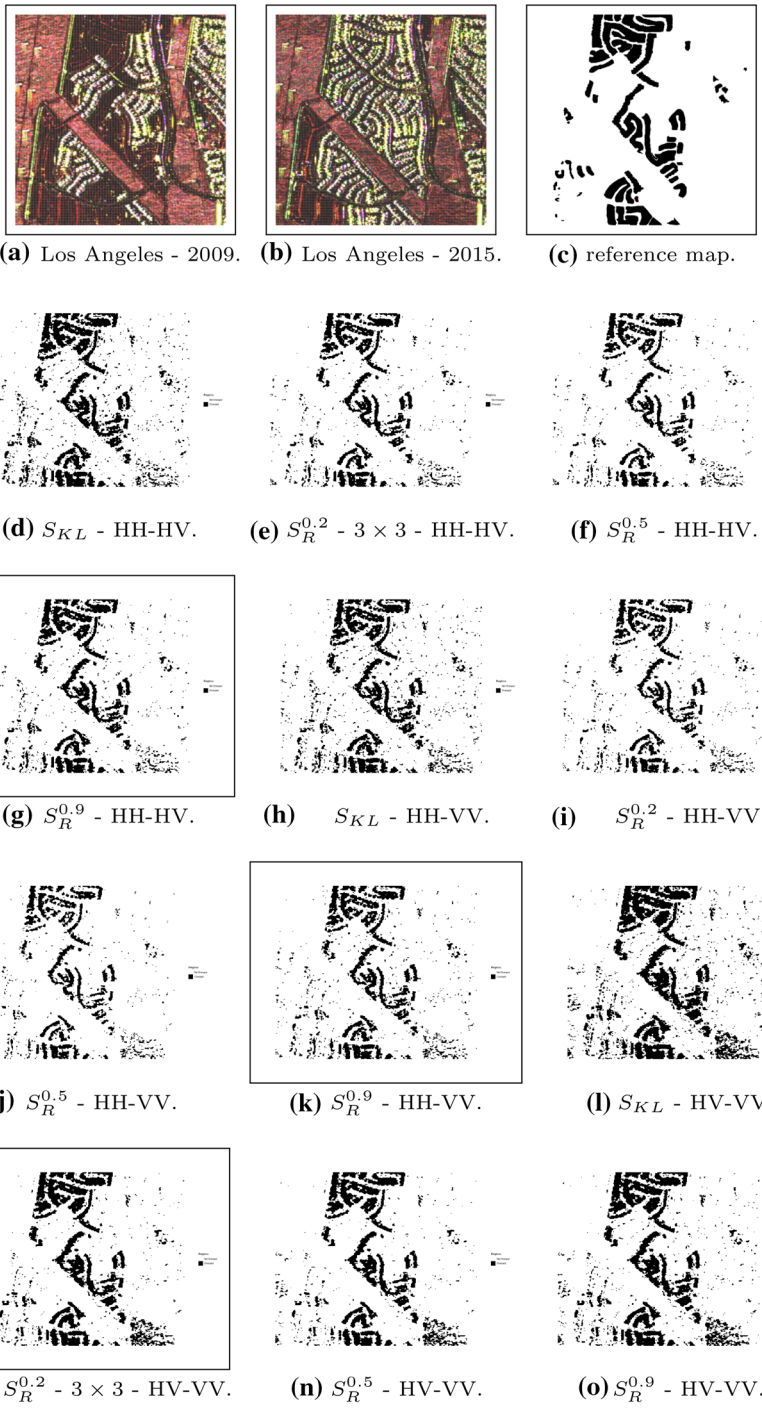
Figure 5a–o display predictive maps for the first application (Scene 1—Los Angeles - USA). When analyzing the maps for Scene 1 for the HH-HV channel pair, the map created with the $CD_R^{0.9}$ detector in Fig. 5g shows the best result compared to the reference map. Figure 5k is the closest to the reference map for the channel pair HH-VV, while Fig. 5m is the best prediction for the reference map for the pair HV-VV. Table 3 shows the values of the evaluation criteria. The above qualitative results can be confirmed by looking at the kappa coefficient for each channel pair. For the channel pair HH-HV, the highest kappa coefficient of $CD_R^{0.9}$ was obtained, namely 65.20%, and the same detector showed the best performance for the pair HH-VV with a kappa coefficient of 62.29%. For the HV-VV channel pair, the best detector was $S_R^{0.2}$ with a kappa coefficient of 66.17%.

For the application in Scene 2 (which refers to California (USA)), the maps are shown in Fig 6a–e. From the reference map in Fig. 6c, the largest changes between 2009 and 2015 are in the regions at the top of the image. The analysis of both the prediction maps and the Table 4 showed that the detector CD_{KL} gave the best results for all channel pairs. This detector achieved kappa coefficients of 65.20% for the channel pair HH-HV, 62.29% for the channel pair HH-VV, and 65.17% for the channel pair HV-VV.

When comparing the scenes we used, we notice that scene 2, which has the most areas of high correlation (as highlighted in Fig. 2c) and the least speckle effect, has smaller values for FP than scene 1.

5 Conclusions

In this work, we used the McKay's bivariate gamma law (MBG) to describe the empirical distribution developed from the image of total scattering power (SPAN). Based on this assumption, it was demonstrated for the first time in the PolSAR literature that



(a) Los Angeles - 2009. **(b)** Los Angeles - 2015. **(c)** reference map.

(d) S_{KL} - HH-HV. **(e)** $S_R^{0.2}$ - 3×3 - HH-HV. **(f)** $S_R^{0.5}$ - HH-HV.

(g) $S_R^{0.9}$ - HH-HV. **(h)** S_{KL} - HH-VV. **(i)** $S_R^{0.2}$ - HH-VV.

(j) $S_R^{0.5}$ - HH-VV. **(k)** $S_R^{0.9}$ - HH-VV. **(l)** S_{KL} - HV-VV.

(m) $S_R^{0.2}$ - 3×3 - HV-VV. **(n)** $S_R^{0.5}$ - HV-VV. **(o)** $S_R^{0.9}$ - HV-VV.

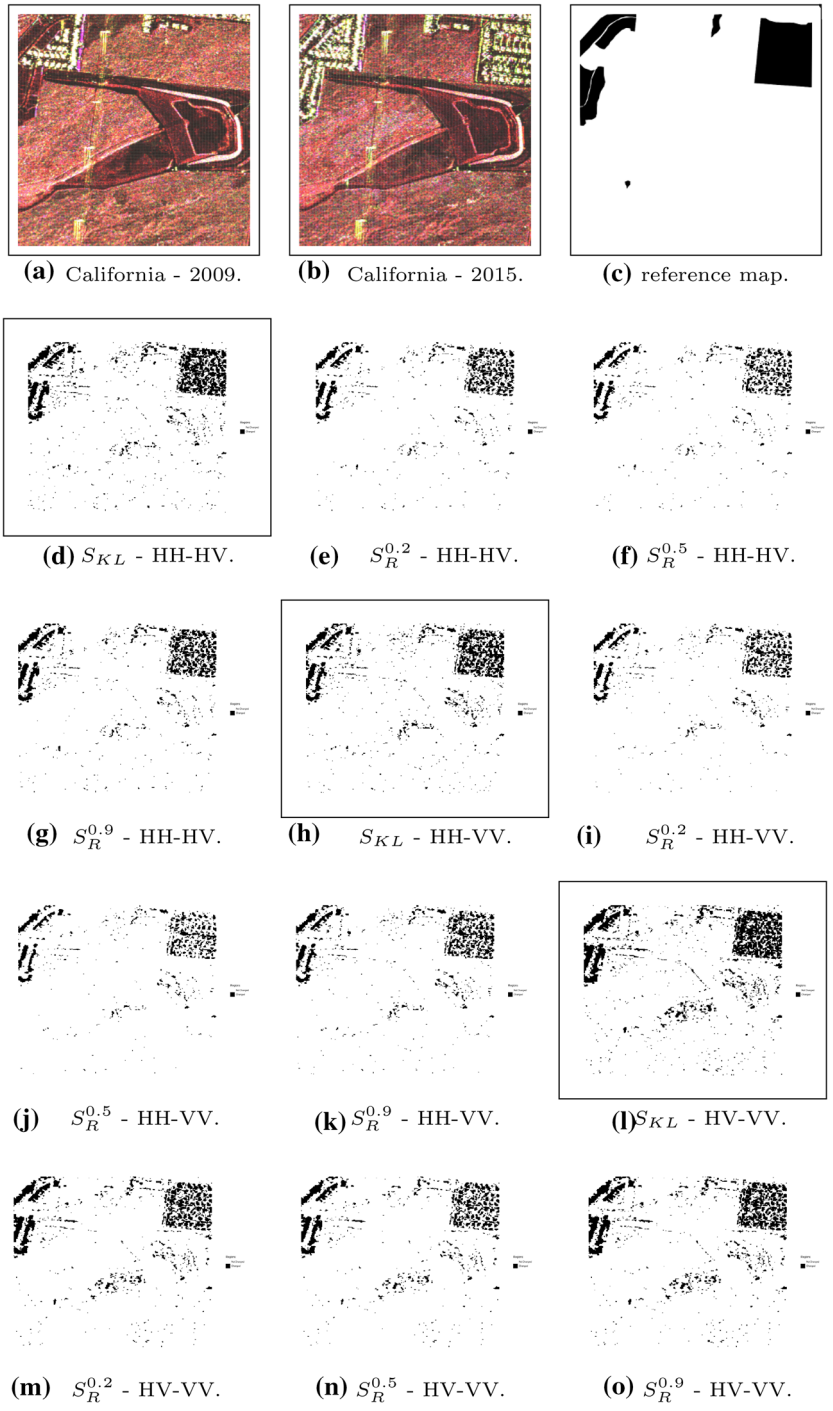
Fig. 5 Change Detection Results—Scene 1—Los Angeles (USA)

Table 3 Detectors' performance—Scene 1—Los Angeles (USA)

Detectors	FP (%)	FN (%)	FA (%)	DR (%)	κ (%)
	HH-HV				
S_{KL}	5.248	4.317	58.936	*94.847*	63.991
$S_R^{0.2}$	2.895	5.699	68.781	93.488	64.147
$S_R^{0.5}$	*2.342*	6.299	76.195	92.895	62.704
$S_R^{0.9}$	3.675	*5.019*	*62.303*	94.167	*65.201*
	HH-VV				
S_{KL}	4.956	**4.908**	**64.281**	**94.202**	61.985
$S_R^{0.2}$	2.614	6.584	81.195	92.575	60.287
$S_R^{0.5}$	**2.101**	7.329	93.642	91.851	57.691
$S_R^{0.9}$	3.405	5.753	70.715	93.391	**62.292**
	HV-VV				
S_{KL}	6.988	**3.342**	**54.531**	**95.877**	63.685
$S_R^{0.2}$	4.351	4.404	57.425	94.804	**66.173**
$S_R^{0.5}$	**3.692**	4.909	61.088	94.286	65.691
$S_R^{0.9}$	5.262	3.909	55.081	95.311	65.848

there is a relationship between the MBG correlation measure and the type of texture in a geo-sensed region. Our core goal was to detect changes based on SPAN after this feature was widely used but nothing was known about the associated distribution. We derived closed-form expressions for the Kullback–Leibler (KL) and Rényi (R) divergences for the MBG and proposed hypothesis tests for two samples. The performance of the tests was quantified by Monte Carlo experiments. Although all tests showed good asymptotic properties, the Rényi distance-based test with order 0.5 presented the best results. Subsequently, the proposed tests were used as change detectors in PolSAR images. The results showed that the detectors based on high-order R-divergence were better for central changes, while the test based on KL -divergence was suitable for changes at the edges. Further simulation studies showed that the proposed hypothesis tests were less sensitive to the noise effect, but in real experiments the change detectors tended to assume high false positive rate values before regions of low correlation (strong noise effect).

Although the contributions of this work have focused on change detection, they can be applied in various areas of post-processing of coherence illumination-based images (e.g., PolSAR and medicine), such as filtering, segmentation, and classification. The probabilistic assumptions used have also been applied in other contexts (e.g., hydrology and reliability theory), making our contributions interesting to potential users in areas outside of image processing. We hope that our advances can be useful in a variety of disciplines.



(a) California - 2009. **(b)** California - 2015. **(c)** reference map.

(d) S_{KL} - HH-HV. **(e)** $S_R^{0.2}$ - HH-HV. **(f)** $S_R^{0.5}$ - HH-HV.

(g) $S_R^{0.9}$ - HH-HV. **(h)** S_{KL} - HH-VV. **(i)** $S_R^{0.2}$ - HH-VV.

(j) $S_R^{0.5}$ - HH-VV. **(k)** $S_R^{0.9}$ - HH-VV. **(l)** S_{KL} - HV-VV.

(m) $S_R^{0.2}$ - HV-VV. **(n)** $S_R^{0.5}$ - HV-VV. **(o)** $S_R^{0.9}$ - HV-VV.

Fig. 6 Change Detection Results—Scene 2—California (USA)

Table 4 Detectors' performance—Scene 2—California (USA)

Detectors	FP (%)	FN (%)	TAF (%)	TD (%)	κ (%)
	HH-HV				
S_{KL}	2.202	*4.209*	*73.389*	*95.389*	*63.577*
$S_R^{0.2}$	1.178	5.863	116.257	93.759	54.566
$S_R^{0.5}$	*0.948*	6.435	140.496	93.208	50.342
$S_R^{0.9}$	1.507	5.201	95.176	94.405	58.787
	HH-VV				
S_{KL}	2.166	**4.884**	**87.866**	**94.691**	**58.632**
$S_R^{0.2}$	1.159	6.566	144.817	93.064	48.258
$S_R^{0.5}$	**0.947**	7.154	178.676	92.506	43.353
$S_R^{0.9}$	1.473	5.883	116.176	93.719	53.184
	HV-VV				
S_{KL}	3.351	**3.507**	**64.779**	**96.078**	**64.009**
$S_R^{0.2}$	2.041	5.077	92.379	94.499	57.609
$S_R^{0.5}$	1.**741**	5.721	110.349	93.864	53.514
$S_R^{0.9}$	2.487	4.405	78.109	95.169	60.996

Funding Funding was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

Appendix A: Kullback–Leibler divergence

The Kullback–Leibler divergence between two pairs $\boldsymbol{x} \sim \text{MBG}(\alpha_{11}, \alpha_{12}, \gamma_1)$ and $\boldsymbol{y} \sim \text{MBG}(\alpha_{21}, \alpha_{22}, \gamma_2)$ is given as

$$
\begin{aligned}
D_{KL}(\boldsymbol{x} \,||\, \boldsymbol{y}) &= \int_0^{\infty} \int_0^{x_2} f(x_1, x_2; \alpha_{11}, \alpha_{12}, \gamma_1) \\
&\quad \times \log \frac{f(x_1, x_2; \alpha_{11}, \alpha_{12}, \gamma_1)}{f(x_1, x_2; \alpha_{21}, \alpha_{22}, \gamma_2)} \mathrm{d}x_1 \mathrm{d}x_2 \\
&= \int_0^{\infty} \int_0^{x_2} \frac{x_1^{\alpha_{11}-1}(x_2 - x_1)^{\alpha_{12}-1} \exp\left(-\frac{x_2}{\gamma_1}\right)}{\gamma_1^{\alpha_{11}+\alpha_{12}} \Gamma(\alpha_{11})\Gamma(\alpha_{12})} \\
&\quad \times \underbrace{\log \left(\frac{x_1^{\alpha_{11}-1}(x_2 - x_1)^{\alpha_{12}-1} \exp\left(-\frac{x_2}{\gamma_1}\right)}{\gamma_1^{\alpha_{11}+\alpha_{12}} \Gamma(\alpha_{11})\Gamma(\alpha_{12})} \frac{\gamma_2^{\alpha_{21}+\alpha_{22}} \Gamma(\alpha_{21})\Gamma(\alpha_{22})}{x_1^{\alpha_{21}-1}(x_2 - x_1)^{\alpha_{22}-1} \exp\left(-\frac{x_2}{\gamma_2}\right)} \right)}_{A} \mathrm{d}x_1 \mathrm{d}x_2
\end{aligned}
$$

The term A can be rewritten as

$$
A = \underbrace{\log \left(\frac{\gamma_2^{\alpha_{21}+\alpha_{22}} \Gamma(\alpha_{21})\Gamma(\alpha_{22})}{\gamma_1^{\alpha_{11}+\alpha_{12}} \Gamma(\alpha_{11})\Gamma(\alpha_{12})} \right)}_{A1} + \underbrace{\log \left(x_1^{\alpha_{11}-\alpha_{21}} (x_2 - x_1)^{\alpha_{12}-\alpha_{22}} \exp\left(-\frac{x_2}{\gamma_1} + \frac{x_2}{\gamma_2}\right) \right)}_{A2}.
$$

Thus,

$$D_{KL}(\boldsymbol{x} \,||\, \boldsymbol{y}) = \frac{1}{\gamma_1^{\alpha_{11}+\alpha_{12}}\Gamma(\alpha_{11})\Gamma(\alpha_{12})} A1 \times \overbrace{\int_0^{\infty}\int_0^{x_2} x_1^{\alpha_{11}-1}(x_2-x_1)^{\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)\mathrm{d}x_1\mathrm{d}x_2}^{W_1} + \frac{1}{\gamma_1^{\alpha_{11}+\alpha_{12}}\Gamma(\alpha_{11})\Gamma(\alpha_{12})} \underbrace{\int_0^{\infty}\int_0^{x_2} x_1^{\alpha_{11}-1}(x_2-x_1)^{\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right) A2\,\mathrm{d}x_1\mathrm{d}x_2}_{W_2}.$$

We obtain the following result for the first term

$$W_1 = \int_0^{\infty}\int_0^{x_2} x_1^{\alpha_{11}-1}(x_2-x_1)^{\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)\mathrm{d}x_1\mathrm{d}x_2.$$

Note that $(x_2 - x_1)^{\alpha_{12}-1} = \left[x_2 \cdot \left(1 - \frac{x_1}{x_2}\right)\right]^{\alpha_{12}-1} = x_2^{\alpha_{12}-1} \cdot \left(1 - \frac{x_1}{x_2}\right)^{\alpha_{12}-1}$. Assuming x_2 constant and making the variable changing $t = \dfrac{x_1}{x_2} \in (0, 1)$, then $\mathrm{d}x_1 = x_2\ \mathrm{d}t$. Thus, we have,

$$\begin{aligned} W_1 &= \int_0^{\infty} x_2^{\alpha_{11}+\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)\int_0^1 t^{\alpha_{11}-1}\cdot(1-t)^{\alpha_{12}-1}\,\mathrm{d}t\ \mathrm{d}x_2 \\ &= \mathrm{Beta}(\alpha_{11},\alpha_{12})\int_0^{\infty} x_2^{\alpha_{11}+\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)\mathrm{d}x_2. \end{aligned}$$

Therefore

$$\begin{aligned} W_1 &= \mathrm{Beta}(\alpha_{11},\alpha_{12})\cdot\int_0^{\infty} x_2^{\alpha_{11}+\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)dx_2 \\ &= \mathrm{Beta}(\alpha_{11},\alpha_{12})\cdot\Gamma(\alpha_{11}+\alpha_{12})\cdot\gamma_1^{\alpha_{11}+\alpha_{12}} = \gamma_1^{\alpha_{11}+\alpha_{12}}\cdot\Gamma(\alpha_{11})\cdot\Gamma(\alpha_{12}). \end{aligned}$$

For the second term, we have

$$\begin{aligned} W2 &= \underbrace{\int_0^{\infty}\int_0^{x_2} x_1^{\alpha_{11}-1}(x_2-x_1)^{\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)(\alpha_{11}-\alpha_{21})\log x_1 dx_1 dx_2}_{W2_A} \\ &+ \underbrace{\int_0^{\infty}\int_0^{x_2} x_1^{\alpha_{11}-1}(x_2-x_1)^{\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)(\alpha_{12}-\alpha_{22})\log(x_2-x_1)dx_1dx_2}_{W2_B} \\ &- \underbrace{\int_0^{\infty}\int_0^{x_2} x_1^{\alpha_{11}-1}(x_2-x_1)^{\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)\left(x_2\frac{\gamma_2-\gamma_1}{\gamma_1\gamma_2}\right)dx_1dx_2}_{W2_C}. \end{aligned}$$

Consider deriving $W2_A$. Assuming x_2 constant and making the variable changing $t = \dfrac{x_1}{x_2} \in (0, 1)$, then $\mathrm{d}x_1 = x_2\ \mathrm{d}t$. Thus, we have,

$$W2_A = (\alpha_{11}-\alpha_{21})\int_0^{\infty} x_2^{\alpha_{12}-1}\exp\left(-\frac{x_2}{\gamma_1}\right)\int_0^{x_2} x_1^{\alpha_{11}-1}\log x_1\left(1-\frac{x_1}{x_2}\right)^{\alpha_{12}-1}\mathrm{d}x_1\mathrm{d}x_2$$

$$
\begin{aligned}
&= (\alpha_{11}-\alpha_{21}) \int_0^{\infty} x_2^{\alpha_{11}+\alpha_{12}-1} \exp\left(-\frac{x_2}{\gamma_1}\right) \Big[\int_0^1 t^{\alpha_{11}-1}(1-t)^{\alpha_{12}-1} \log t \; \mathrm{d}t \\
&\quad + \log x_2 \int_0^1 t^{\alpha_{11}-1}(1-t)^{\alpha_{12}-1} dt \Big] \mathrm{d}x_2 \\
&= (\alpha_{11}-\alpha_{21}) \; \mathrm{Beta}(\alpha_{11},\alpha_{12}) \, \Gamma(\alpha_{11}+\alpha_{12}) \, \gamma_1^{\alpha_{11}+\alpha_{12}} \, [\Psi(\alpha_{11}) + \log \gamma_1].
\end{aligned}
$$

Consider deriving $W2_B$ and applying the same previous arguments, one has

$$
\begin{aligned}
W2_B \; &= (\alpha_{12}-\alpha_{22}) \int_0^{\infty} x_2^{\alpha_{12}-1} \exp\left(-\frac{x_2}{\gamma_1}\right) \log x_2 \int_0^{x_2} x_1^{\alpha_{11}-1} \left(1-\frac{x_1}{x_2}\right)^{\alpha_{12}-1} \mathrm{d}x_1 \mathrm{d}x_2 \\
&\quad + (\alpha_{12}-\alpha_{22}) \int_0^{\infty} x_2^{\alpha_{12}-1} \exp\left(-\frac{x_2}{\gamma_1}\right) \int_0^{x_2} x_1^{\alpha_{11}-1} \left(1-\frac{x_1}{x_2}\right)^{\alpha_{12}-1} \log\left(1-\frac{x_1}{x_2}\right) \mathrm{d}x_1 \mathrm{d}x_2 \\
&= (\alpha_{12}-\alpha_{22}) \; \mathrm{Beta}(\alpha_{11},\alpha_{12}) \, \Gamma(\alpha_{11}+\alpha_{12}) \, \gamma_1^{\alpha_{11}+\alpha_{12}} \, \left[\log \gamma_1 + \Psi(\alpha_{12})\right].
\end{aligned}
$$

Consider deriving $W2_C$ and applying the same previous arguments, one has

$$
\begin{aligned}
W2_C &= \left(\frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2}\right) \cdot \int_0^{\infty} \int_0^{x_2} x_1^{\alpha_{11}-1} (x_2-x_1)^{\alpha_{12}-1} \exp\left(-\frac{x_2}{\gamma_1}\right) \cdot x_2 dx_1 dx_2 \\
&= \left(\frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2}\right) \mathrm{Beta}(\alpha_{11},\alpha_{12})(\alpha_{11}+\alpha_{12})\Gamma(\alpha_{11}+\alpha_{12})\gamma_1^{\alpha_{11}+\alpha_{12}}.
\end{aligned}
$$

Thus, after some algebraic manipulations, one has

$$
\begin{aligned}
D_{KL}(\boldsymbol{x} \,||\, \boldsymbol{y}) \; &= D_{KL}(\boldsymbol{\theta_1} \,||\, \boldsymbol{\theta_2}) \\
&= \log \frac{\gamma_2^{\alpha_{21}+\alpha_{22}} \Gamma(\alpha_{21})\Gamma(\alpha_{22})}{\gamma_1^{\alpha_{11}+\alpha_{12}} \Gamma(\alpha_{11})\Gamma(\alpha_{12})} + \Big\{ (\alpha_{11}-\alpha_{21}) \left[\Psi(\alpha_{11}) + \log \gamma_1\right] \\
&\quad + (\alpha_{12}-\alpha_{22}) \left[\log \gamma_1 + \Psi(\alpha_{12})\right] - \left(\frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2}\right)(\alpha_{11}+\alpha_{12}) \Big\}.
\end{aligned}
$$

Appendix B: Rényi divergence of the order r

The Rényi divergence at order r between two pairs $\boldsymbol{x} \sim \mathrm{MBG}(\alpha_{11},\alpha_{12},\gamma_1)$ and $\boldsymbol{y} \sim \mathrm{MBG}(\alpha_{21},\alpha_{22},\gamma_2)$ is given as

$$
\begin{aligned}
&D_{R:r}(\boldsymbol{x} \,||\, \boldsymbol{y}) = D_{R:r}(\boldsymbol{\theta_1} \,||\, \boldsymbol{\theta_2}) \\
&= \frac{1}{r\,(r-1)} \int_{\boldsymbol{\mathcal{Z}}} f^r(x_1,x_2;\alpha_{11},\alpha_{12},\gamma_1) \, f^{1-r}(x_1,x_2;\alpha_{11},\alpha_{12},\gamma_1) \mathrm{d}x_1 \, \mathrm{d}x_2 \\
&= \left(\frac{1}{\gamma_1^{\alpha_{11}+\alpha_{12}}\Gamma(\alpha_{11})\Gamma(\alpha_{12})}\right)^r \cdot \left(\frac{1}{\gamma_2^{\alpha_{21}+\alpha_{22}}\Gamma(\alpha_{21})\Gamma(\alpha_{22})}\right)^{1-r} \int_0^{\infty} \int_0^{x_2} \\
&\quad \underbrace{\left(x_1^{\alpha_{11}-1}(x_2-x_1)^{\alpha_{12}-1} \exp\left(-\frac{x_2}{\gamma_1}\right)\right)^r \cdot \left(x_1^{\alpha_{21}-1}(x_2-x_1)^{\alpha_{22}-1} \exp\left(-\frac{x_2}{\gamma_2}\right)\right)^{1-r}}_{A} \mathrm{d}x_1 \, \mathrm{d}x_2.
\end{aligned}
\tag{18}
$$

Assuming x_2 constant and making the variable changing $t = \dfrac{x_1}{x_2} \in (0, 1)$, then $\mathrm{d}x_1 = x_2\ \mathrm{d}t$. Thus,

$$
\begin{aligned}
D_{R:r}(\boldsymbol{\theta_1} \,||\, \boldsymbol{\theta_2}) = &\left(\frac{1}{\gamma_1^{\alpha_{11}+\alpha_{12}}\Gamma(\alpha_{11})\Gamma(\alpha_{12})}\right)^r \left(\frac{1}{\gamma_2^{\alpha_{21}+\alpha_{22}}\Gamma(\alpha_{21})\Gamma(\alpha_{22})}\right)^{1-r} \\
&\int_0^\infty \int_0^{x_2} x_1^{r\alpha_{11}+\alpha_{21}-r\alpha_{21}-1}(x_2-x_1)^{r\alpha_{12}+\alpha_{22}-r\alpha_{22}-1} \cdot \exp\left(-\frac{rx_2}{\gamma_1} - \frac{(1-r)x_2}{\gamma_2}\right) dx_1 dx_2 \\
&= \frac{1}{r.(1-r)}\Big\{\log\Gamma(\eta_1) + \log\Gamma(\eta_2) + (1-r)\Big[\log\Gamma(\alpha_{21}) + \log\Gamma(\alpha_{22}) \\
&\quad -(\alpha_{21}+\alpha_{22})\Big[\log\gamma_1 - \log(\gamma_1 - \gamma_1 \cdot r + \gamma_2 \cdot r)\Big]\Big] \\
&\quad +r\Big[(\alpha_{11}+\alpha_{12}) \cdot (\log\gamma_2 - \log(\gamma_1 - \gamma_1 \cdot r + \gamma_2 \cdot r) - \log\Gamma(\alpha_{11}) + \log\Gamma(\alpha_{12})\Big]\Big\},
\end{aligned}
$$

where $\eta_1 = \alpha_{21} + r\alpha_{11} - r\alpha_{21}$ and $\eta_2 = \alpha_{22} + r\alpha_{12} - r\alpha_{22}$.

References

Ali SM, Silvey SD (1966) A general class of coefficients of divergence of one distribution from another. J R Stat Soc B 28:131–142

Amari S (1985) Differential-geometrical methods in statistics. Lecture notes in statistics, Springer, New York

An W, Cui Y, Yang J (2010) Three-component model-based decomposition for polarimetric SAR data. IEEE Trans Geosci Remote Sens 48(6):2732–2739. https://doi.org/10.1109/TGRS.2010.2041242

Anfinsen SN, Doulgeris AP, Eltoft T (2009) Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery. IEEE Trans Geosci Remote Sens 47(11):3795–3809

Bouhlel N, Méric S (2020) Multilook polarimetric SAR change detection using stochastic distances between matrix-variate G_d^0 distributions. IEEE Trans Geosci Remote Sens 58:6823–6843

Burbea J, Rao C (1982) On the convexity of some divergence measures based on entropy functions. IEEE Trans Inf Theory 28:489–495

Cordeiro GM, Cintra RJ, Rêgo LC, Nascimento AD (2019) The gamma generalized normal distribution: a descriptor of SAR imagery. J Comput Appl Math 347:257–272

Cover TM, Thomas JA (1991) Elements of information theory. Wiley, New York

Csiszár I (1967) Information-type measures of difference of probability distributions and indirect observations. Stud Sci Math Hung 2:299–318

de Borba AA, Marengoni M, Frery AC (2019) Fusion of evidences for edge detection in PolSAR images. In: 2019 IEEE recent advances in geoscience and remote sensing: technologies, standards and applications (TENGARSS), pp 80–85

Ferreira JA, Coêlho H, Nascimento ADC (2021) A family of divergence-based classifiers for polarimetric synthetic aperture radar (PolSAR) imagery vector and matrix features. Int J Remote Sens 42(4):1201–1229

Freitas CC, Frery AC, Correia AH (2005) The polarimetric G distribution for SAR data analysis. Environmetrics 16(1):13–31. https://doi.org/10.1002/env.658

Frery AC, Muller HJ, Yanasse CCF, Sant'Anna SJS (1997) A model for extremely heterogeneous clutter. IEEE Trans Geosci Remote Sens 35:648–659

Frery AC, Nascimento ADC, Cintra RJ (2014) Analytic expressions for stochastic distances between relaxed complex Wishart distributions. IEEE Trans Geosci Remote Sens 52:1213–1226

Gao G, Qin X, Zhou S (2013) Modeling SAR images based on a generalized gamma distribution for texture component. Prog Electromagn Res 137:669–685

Gupta AK, Nadarajah S (2006) Intensity-duration models based on bivariate gamma distributions. Hiroshima Math J 36:387–395

Hachicha S, Chaabane F (2009) Application of dsm theory for SAR image change detection. In: 2009 16th IEEE international conference on image processing (ICIP), pp 3733–3736. IEEE

Hagedorn M, Smith PJ, Bones PJ, Millane RP, Pairman D (2006) A trivariate chi-squared distribution derived from the complex Wishart distribution. J Multivar Anal 97:655–674
Hajnsek I, Desnos Y-L (2020) Polarimetric synthetic aperture radar: principles and application, vol 25. Springer, Cham
Hartley RVL (1928) Transmission of information. Bell Syst Tech J 7:535–563
Ince T, Kiranyaz S, Gabbouj M (2012) Evolutionary RBF classifier for polarimetric SAR images. Expert Syst Appl 39:4710–4717
Kailath T (1967) The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans Commun Technol 15:52–60
Kullback S (1968) Information theory and statistics. A Wiley publication in mathematical statistics, Dover Publications, New York
Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22:79–86
Lee J-S, Pottier E (2009) Polarimetric radar imaging: from basics to applications. CRC, Boca Raton
Lee J-S, Hoppel KW, Mango SA, Miller AR (1994) Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. IEEE Trans Geosci Remote Sens 32:1017–1028
Li H-C, Hong W, Wu Y-R, Fan P-Z (2010) An efficient and flexible statistical model based on generalized gamma distribution for amplitude SAR images. IEEE Trans Geosci Remote Sens 48(6):2711–2722
Li H-C, Hong W, Wu Y-R, Fan P-Z (2011) On the empirical-statistical modeling of SAR images with generalized gamma distribution. IEEE J Sel Top Signal Process 5:386–397
Liese F, Vajda I (1987) Convex statistical distances. Teubner-Texte zur Mathematik, Teubner, Leipzig
López-Martínez C, Fabregas X (2003) Polarimetric SAR speckle noise model. IEEE Trans Geosci Remote Sens 41(10):2232–2242. https://doi.org/10.1109/TGRS.2003.815240
Nadarajah S (2005) Reliability for some bivariate gamma distributions. Math Probl Eng 2005:924843
Nadarajah S, Gupta AK (2006) Some bivariate gamma distributions. Appl Math Lett 19:767–774
Nascimento AD, Cintra RJ, Frery AC (2009) Hypothesis testing in speckled data with stochastic distances. IEEE Trans Geosci Remote Sens 48(1):373–385
Qin X, Zhou S, Zou H, Gao G (2012) A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images. IEEE Geosci Remote Sens Lett 10:806–810
Rodríguez WDA, Amaral GJA, Nascimento ADC, Ferreira JA (2020) Information criteria in classification: new divergence-based classifiers. J Stat Comput Simul 90(17):3200–3217
Salicrú M, Morales D, Menéndez M, Pardo L (1994) On the applications of divergence type measures in testing statistical hypotheses. J Multivar Anal 51:372–391
Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
Skriver H, Dall J, Le Toan T, Quegan S, Ferro-Famil L, Pottier E (2005) Agriculture classification using PolSAR data. In: Lacoste H (ed) The 2nd international workshop POLINSAR 2005 (ESA SP-586), Frascati. Italy, ESA SP
Ünsalan MIC (2012) Two-dimensional change detection methods: remote sensing applications. Springer-Briefs in computer science, Springer, London
Valdevino Félix de Lima W, ao David Costa do Nascimento A, José Amorim do Amaral G (2021) Distance-based tests for planar shape. J Multivar Anal 184:104743
van Erven T, Harremoes P (2010) Rényi divergence and majorization. IEEE, Austin, TX, USA
Wang H, Han J, Deng Y (2017) PolSAR image classification based on Laplacian eigenmaps and superpixels. EURASIP J Wirel Commun Netw 2017:198
Wang W, Xiang D, Ban Y, Zhang J, Wan J (2018) Enhanced edge detection for polarimetric SAR images using a directional span-driven adaptive window. Int J Remote Sens 39:6340–6357. https://doi.org/10.1080/01431161.2018.1460501
West RD, Riley RM (2019) Polarimetric interferometric SAR change detection discrimination. IEEE Trans Geosci Remote Sens 57:3091–3104
Woodward PM (1953) Probability and information theory: with applications to radar. International series of monographs on electronics and instrumentation, vol 3. McGraw-Hill, New York. https://books.google.com.br/books?id=RA0nAAAAMAAJ
Xu L, Li S, Deng Y, Wang R (2014) Unsupervised classification of polarimetric synthetic aperture radar interferometry using polarimetric interferometric similarity parameters and SPAN. IET Radar Sonar Navig 8:1135–1144
Yue S, Ouarda TBMJ, Bobée B (2001) A review of bivariate gamma distributions for hydrological application. J Hydrol 246(1):1–18

Zhang L, Sun L, Zou B, Moon WM (2015) Fully polarimetric SAR image classification via sparse representation and polarimetric features. IEEE J Select Top Appl Earth Observ Remote Sens 8(8):3923–3932
Zhang P, Li B, Boudaren MEY, Yan J, Li M, Wu Y (2020) Parameter estimation of generalized gamma distribution toward SAR image processing. IEEE Trans Aerosp Electron Syst 56:3701–3717
Zografos K, Ferentinos K, Papaioannou T (1990) Divergence statistics: sampling properties and multinomial goodness of fit and divergence tests. Commun Stat Theory Methods 19:1785–1802