



**HAL**  
open science

# Multilook Polarimetric SAR Change Detection Using Stochastic Distances Between Matrix-Variate Gd Distributions

Nizar Bouhlel, Stephane Meric

► **To cite this version:**

Nizar Bouhlel, Stephane Meric. Multilook Polarimetric SAR Change Detection Using Stochastic Distances Between Matrix-Variate Gd Distributions. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58 (10), pp.6823-6843. 10.1109/TGRS.2020.2976766 . hal-02948490

**HAL Id: hal-02948490**

**<https://univ-rennes.hal.science/hal-02948490>**

Submitted on 25 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multilook Polarimetric SAR Change Detection Using Stochastic Distances Between Matrix-Variate $\mathcal{G}_d^0$ Distributions

Nizar Bouhlef<sup>1b</sup>, *Member, IEEE*, and Stéphane Méric<sup>1b</sup>, *Member, IEEE*

**Abstract**—In this article, we propose an efficient heterogeneous change detection algorithm based on stochastic distance measure between two  $\mathcal{G}_d^0$  distributions. Due to its flexibility and simplicity, the matrix-variate  $\mathcal{G}_d^0$  distribution has been successfully used to model the multilook polarimetric synthetic aperture radar (PolSAR) data and has been tested for classification, segmentation, and image analysis. Concretely, closed-form expressions for the Kullback–Leibler, Rényi of order  $\beta$ , Bhattacharyya, and Hellinger distances are provided to compute the stochastic distance between  $\mathcal{G}_d^0$  distributions. In this context, we resort to the expectation–maximization (EM) to estimate accurately with low complexity the parameters of the probability distribution of the two multilook polarimetric covariance matrices to be compared. Finally, the performance of the method is compared firstly to the performance of other known distributions, such as the scaled complex Wishart distribution, and secondly to other known statistical tests using simulated and real multilook PolSAR data.

**Index Terms**—Bhattacharyya and Hellinger distances, change detection, expectation–maximization (EM) algorithm, multilook polarimetric synthetic aperture radar (PolSAR) data, Rényi of order  $\beta$ , stochastic distances: Kullback–Leibler.

## I. INTRODUCTION

THE detection of changes in synthetic aperture radar (SAR) images has become very important for various applications, such as earth monitoring, earth observation, damage assessment, and disaster management for which unsupervised methods have been widely applied. Basically, change detection consists of analyzing multitemporal remote sensing images acquired on the same geographical area at different times to identify changes [1]. When changes occur, a change detection map is generated in which changed areas are explicitly identified. In unsupervised methods [2], some features of the two images (before and after change) are compared by using some similarity metrics resulting in a change map quantifying the difference between the scenes, and then a

threshold is applied to produce a final binary change detection map with two classes associated with changed and unchanged pixels [3]. Several unsupervised change detection methods have been proposed and can be categorized into two groups: one is based on pixel intensity and the neighboring of the pixel, whereas the other is based on local statistics and statistical information theory.

The first group included image differencing, mean ratio/log-ratio measures [4], [5], Gauss log-ratio [6], etc. The simplest one is to compute the ratio of SAR amplitudes or intensities observed at different times and test the hypotheses of change versus no-change. This kind of ratio is well-known test statistic in single-polarization SAR-based change detection. The generalization of these methods to multilook polarimetric SAR (PolSAR) images has been introduced by Conradson *et al.* [7]. The authors proposed a likelihood ratio test (LRT) for the equality of two complex covariance matrices and gave the approximated distribution of the LRT statistic. The LRT approach was extended to the multitemporal case [8], [9] and multifrequency data [10]. Recently, Akbari *et al.* [11] proposed a simpler test statistic to detect changes in many scenarios. The test assumes the scaled complex Wishart distribution for the covariance matrix data. It is based on the complex-kind Hotelling–Lawley trace (HLT) which is used as a covariance equality test.

The second group based on statistical information measure has shown good performances for change detection. Indeed, the complex PolSAR images are well described by probability density function (pdf) due to the random nature of speckle. Thus, the statistical information measure uses the local pdfs of the neighborhood of homologous pixels of the pair of data used for the change detection [12]. These pdfs have been chosen particularly to adequately model the statistics of the SAR data. Once the parameters of the distributions are estimated, the information measure quantifies the distance between the two pdfs. The comparison can be performed using various information measures such as the mutual information [13], [14], the variational and mixed information [13], [15], and the stochastic distances such as the Kullback–Leibler [16], Rényi [17], Bhattacharyya [18], [19], and Hellinger distances [20]. We limit our focus to change detection based on a stochastic distance measure between two scenes acquired at different times.

Closed-form expressions of the distance measure of some known distributions exist for single-polarization SAR images.

Manuscript received September 15, 2019; revised January 13, 2020; accepted February 21, 2020. (*Corresponding author: Nizar Bouhlef.*)

Nizar Bouhlef is with the Signal, Communications, Embedded Electronics (SCEE) Team, Institut d'Electronique et de Telecommunications de Rennes (IETR), CentraleSupélec, Campus de Rennes, 35510 Cesson-Sévigné, France (e-mail: nizar.bouhlef@centralesupelec.fr).

Stéphane Méric is with the SAR and Hyperspectral Multi-Modal Imaging and Signal Processing, Electromagnetic Modeling (SHINE) Team, IETR, Institut National des Sciences Appliquées de Rennes (INSA), 35708 Rennes, France.

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

Digital Object Identifier 10.1109/TGRS.2020.2976766

Otherwise, certain approximations should be applied, like the kernel density estimation (KDE) [21] or the 1-D edge worth series expansions used to estimate the local pdfs [22]. Nascimento *et al.* [23] developed analytic expressions for the Kullback–Leibler, Rényi, Bhattacharyya, and Hellinger distances between univariate  $\mathcal{G}^0$  distributions. Full PolSAR gives more scattering information than single polarization channel SAR data and then can be used to increase the quality of the change detection map. Stochastic distances between models for complex multilook PolSAR data require dealing with integrals whose domain is the set of all positive definite complex Hermitian matrices [24]. Unfortunately, it is not obvious to find closed-form expressions for all the stochastic distances. In this case, certain approximation techniques based on the numerical evaluation of the integral of the distance measure are commonly applied. The most popular method is the Monte-Carlo (MC) estimation technique [25]. It can estimate the stochastic distance provided that a large number of independent and identically distributed (i.i.d.) samples is provided. Nevertheless, the MC integration is a too slow process to be useful in change detection. The simplest way to have a closed-form expression for all the stochastic measures is to use a simple model. The complex scaled Wishart distribution is widely used as a statistical simple model for low-resolution multilook PolSAR data due to its ease of implementation and low computational cost [26]. Frery *et al.* [24] obtained analytic expressions for the Kullback–Leibler, Rényi, Bhattacharyya, and Hellinger distances between scaled complex Wishart distributions in their most general form and in important particular cases.

However, one critical issue in applying Wishart distribution is that when the resolution increases, the homogeneous hypothesis of the PolSAR is not valid any more, and non-Gaussian statistics are observed. Consequently, many heterogeneous models have been proposed based on the scalar texture product model which is commonly assumed to be an appropriate statistical model for PolSAR data [27]. Based on the product model, the speckle is modeled with the complex scaled Wishart distribution, and the texture is modeled as gamma ( $\gamma$ ), inverse gamma ( $\gamma^{-1}$ ), Fisher ( $\mathcal{F}$ ), and generalized inverse Gaussian distributions ( $\mathcal{N}^{-1}$ ). The first two distributions depend on a single shape parameter, whereas the other ones depend on two shape parameters. The parameter estimation becomes more challenging when the number increases. With different texture distributions, the product model leads to different compound distributions given by  $\mathcal{K}_d$  [28],  $\mathcal{G}_d^0$  [29], [30], Kummer- $\mathcal{U}_d$  [31], and  $\mathcal{G}_d$  [30]. In general, unless the special function is complicated and the shape of the distribution is flexible, the better.

In this regard, the  $\mathcal{G}_d^0$  distribution has been extensively used by the community for its flexibility and has been reported excellent performance showing the suitability for describing different kind of regions: homogeneous, heterogeneous, and extremely heterogeneous clutter with low computational costs. In addition, the  $\mathcal{G}_d^0$  distribution is simpler than the  $\mathcal{K}_d$  distribution, which contains the modified Bessel function of the second kind. Recent works also validate its use as an efficient distribution for modeling [32], ship detection [33],

change detection [12], [23], and classification of PolSAR images [34], [35].

For these reasons and without losing generality, we assume that multilook PolSAR data follow a  $\mathcal{G}_d^0$  distribution. We present analytic expressions for the Kullback–Leibler, Rényi, Bhattacharyya, and Hellinger distances between two matrix-variate  $\mathcal{G}_d^0$  distributions in general forms and in particular cases. We believe that this is the first work that establishes the closed-form expressions for these stochastic distances. In order to apply these stochastic distance measures to change detection, the parameters of matrix-variate  $\mathcal{G}_d^0$  distributions need to be estimated first. In particular, we estimate the equivalent number of looks (ENLs), the covariance matrix of the speckle component, and the texture distribution parameters by the polarimetric expectation–maximization (EM) estimator [36]. For that end, the first contribution of this article is to present closed-form expressions of the Kullback–Leibler, Rényi, Bhattacharyya, and Hellinger distances between two matrix-variate  $\mathcal{G}_d^0$  distributions. The second one is to provide a numerical analysis of the closed-form expression and providing benchmark resources. The last contribution is to present a parameter estimation procedure for compound distribution and specifically for  $\mathcal{G}_d^0$  distribution.

This article is organized as follows: Section II introduces the statistical product model for multilook PolSAR data and the matrix-variate  $\mathcal{G}_d^0$  distribution. The closed-form expressions for four stochastic distances between  $\mathcal{G}_d^0$  distributions are derived in Section III. They are the Kullback–Leibler, Rényi of order  $\beta$ , Bhattacharyya, and Hellinger distances. Section IV presents the parameter estimation method needed for the computation of the previous stochastic distances. The polarimetric EM estimator is used for the multilook PolSAR data distribution. The content of Section V is related to the performance evaluation. Simulated PolSAR data and real PolSAR data are used for evaluation. Finally, Section VI concludes this article.

## II. PRODUCT MODEL FOR MULTILOOK POLARIMETRIC RADAR DATA

The backscattering of a full PolSAR system is characterized by the polarimetric scattering vector and defined as

$$\mathbf{s} = [s_{HH}, s_{HV}, s_{VH}, s_{VV}]^T \in \mathbb{C}^d \quad (1)$$

The elements represent the complex backscattering coefficients in the four combinations of the linear receive and transmit polarizations.  $[\cdot]^T$  means transposition, and  $d = \dim(\mathbf{s})$  is the vector dimension. The polarimetric scattering vector  $\mathbf{s}$  represents a complex single-look. To reduce the speckle, a multilook operation is used by averaging the covariance matrix. The averaged multilook covariance matrix is obtained as follows:

$$\mathbf{C} = \frac{1}{L} \sum_{\ell=1}^L \mathbf{s}_\ell \mathbf{s}_\ell^H, \quad L \geq d \quad (2)$$

where  $L$  is the number of looks,  $(\cdot)^H$  denotes the Hermitian operator, and  $\mathbf{C} \in \Omega_+ \subset \mathbb{C}^{d \times d}$  is the multilook polarimetric covariance matrix defined on the cone  $\Omega_+$  of the positive

definite complex Hermitian matrices. According to the multi-look polarimetric product model, the covariance matrix  $\mathbf{C}$  is considered as the product of a positive scalar texture random variable, denoted  $\tau$ , and a random speckle matrix, denoted  $\mathbf{X}$ , such that

$$\mathbf{C} = \tau \mathbf{X}. \quad (3)$$

Matrix  $\mathbf{X}$  follows a scaled complex Wishart distribution  $s\mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$  with a pdf given by

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{L^{Ld} |\mathbf{X}|^{L-d}}{\Gamma_d(L) |\mathbf{\Sigma}|^L} \text{etr}(-L\mathbf{\Sigma}^{-1}\mathbf{X}) \quad (4)$$

where  $\mathbf{\Sigma}$  is the covariance matrix of the speckle,  $\text{etr}(\cdot) = \exp(\text{tr}(\cdot))$  is the exponential trace operator,  $|\cdot|$  is the determinant operator, and  $\Gamma_d(L)$  is the multivariate gamma function of the complex kind defined as  $\Gamma_d(L) = \pi^{d(d-1)/2} \prod_{i=0}^{d-1} \Gamma(L-i)$ , where  $\Gamma(L)$  is the standard Euler gamma function. The pdf of  $\mathbf{C}$  using the Bayes' theorem becomes as follows:

$$f_{\mathbf{C}}(\mathbf{C}) = \int_0^{+\infty} f_{\mathbf{C}|\tau}(\mathbf{C}|\tau) f_{\tau}(\tau) d\tau \quad (5)$$

where  $f_{\mathbf{C}|\tau}(\mathbf{C}|\tau)$  is the pdf of  $\mathbf{C}$  with a specific value of  $\tau$  and is given by

$$f_{\mathbf{C}|\tau}(\mathbf{C}|\tau) = \frac{L^{Ld} |\mathbf{C}|^{L-d}}{\Gamma_d(L) |\mathbf{\Sigma}|^L} \frac{1}{\tau^{dL}} \text{etr}\left(-\frac{L}{\tau} \mathbf{\Sigma}^{-1} \mathbf{C}\right). \quad (6)$$

Substituting (6) in (5), the pdf of  $\mathbf{C}$  is obtained by

$$f_{\mathbf{C}}(\mathbf{C}) = \frac{L^{Ld} |\mathbf{C}|^{L-d}}{\Gamma_d(L) |\mathbf{\Sigma}|^L} \int_0^{+\infty} \frac{1}{\tau^{dL}} \text{etr}\left(-\frac{L}{\tau} \mathbf{\Sigma}^{-1} \mathbf{C}\right) f_{\tau}(\tau) d\tau. \quad (7)$$

The variable texture  $\tau$  follows a normalized inverse gamma distribution to the unit mean, denoted  $\bar{\gamma}^{-1}(\lambda)$  as follows:

$$f_{\tau}(\tau) = \frac{(\lambda-1)^\lambda}{\Gamma(\lambda)} \frac{1}{\tau^{1+\lambda}} \exp\left(-\frac{\lambda-1}{\tau}\right) \quad \tau \in \mathbb{R}_+, \quad \lambda > 1. \quad (8)$$

As a consequence, the multilook polarimetric covariance matrix follows the matrix-variate  $\mathcal{G}_d^0$  distributions [30], denoted as  $\mathcal{G}_d^0(\mathbf{\Sigma}, L, \lambda)$ , and is given as follows:

$$f_{\mathbf{C}}(\mathbf{C}) = \frac{L^{Ld} |\mathbf{C}|^{L-d}}{\Gamma_d(L) |\mathbf{\Sigma}|^L} \frac{(\lambda-1)^\lambda \Gamma(dL+\lambda)}{\Gamma(\lambda) (L\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) + \lambda - 1)^{dL+\lambda}}. \quad (9)$$

### III. STOCHASTIC DISTANCES

In this section, the closed-form expressions of the stochastic distances such as the Kullback–Leibler, Rényi of order  $\beta$ , Bhattacharyya, and Hellinger between two matrix-variate  $\mathcal{G}_d^0(\mathbf{\Sigma}_1, L_1, \lambda_1)$  and  $\mathcal{G}_d^0(\mathbf{\Sigma}_2, L_2, \lambda_2)$  distributions are derived. The analytic expressions are available for different ENLs, various covariance matrices, and different texture parameters. A particular case is examined when the ENL is the same for the two distributions. This case is likely to be the most frequently used in practice since it allows the comparison of two possibly different areas from the same image [24]. Let  $\mathbf{C}^1$  and  $\mathbf{C}^2$  be two random matrices that follow  $\mathcal{G}_d^0$  distribution with pdfs  $f_{\mathbf{C}^1}(\mathbf{C}|\mathbf{\Sigma}_1, L_1, \lambda_1)$  and  $f_{\mathbf{C}^2}(\mathbf{C}|\mathbf{\Sigma}_2, L_2, \lambda_2)$  given by (9).

#### A. Kullback–Leibler Distance

The Kullback–Leibler divergence [16] between  $\mathbf{C}^1$  and  $\mathbf{C}^2$  is given by

$$\begin{aligned} D_{\text{KL}}(\mathbf{C}^1 \|\mathbf{C}^2) &= \int_{\Omega_+} \ln\left(\frac{f_{\mathbf{C}^1}(\mathbf{C})}{f_{\mathbf{C}^2}(\mathbf{C})}\right) f_{\mathbf{C}^1}(\mathbf{C}) d\mathbf{C} \\ &= E_{\mathbf{C}^1} \{\ln f_{\mathbf{C}^1}(\mathbf{C})\} - E_{\mathbf{C}^1} \{\ln f_{\mathbf{C}^2}(\mathbf{C})\}. \end{aligned} \quad (10)$$

Jeffreys [37] used a symmetric version of (10) as a measure of divergence between two distributions. In this case, the measure is called “distance” and is given by

$$d_{\text{KL}}(\mathbf{C}^1, \mathbf{C}^2) = D_{\text{KL}}(\mathbf{C}^1 \|\mathbf{C}^2) + D_{\text{KL}}(\mathbf{C}^2 \|\mathbf{C}^1). \quad (11)$$

The expressions of  $E_{\mathbf{C}^1} \{\ln f_{\mathbf{C}^1}(\mathbf{C})\}$  and  $E_{\mathbf{C}^1} \{\ln f_{\mathbf{C}^2}(\mathbf{C})\}$  are given as follows:

$$\begin{aligned} E_{\mathbf{C}^1} \{\ln f_{\mathbf{C}^1}(\mathbf{C})\} &= \ln\left(\frac{\Gamma(dL_1 + \lambda_1)(\lambda_1 - 1)^{\lambda_1}}{\Gamma(\lambda_1)\Gamma_d(L_1)|L_1\mathbf{\Sigma}_1^{-1}|^{-L_1}}\right) + (L_1 - d)E_{\mathbf{C}^1} \{\ln |\mathbf{C}|\} \\ &\quad - (dL_1 + \lambda_1)E_{\mathbf{C}^1} \{\ln[\text{tr}(L_1\mathbf{\Sigma}_1^{-1}\mathbf{C}) + \lambda_1 - 1]\} \end{aligned} \quad (12)$$

and

$$\begin{aligned} E_{\mathbf{C}^1} \{\ln f_{\mathbf{C}^2}(\mathbf{C})\} &= \ln\left(\frac{\Gamma(dL_2 + \lambda_2)(\lambda_2 - 1)^{\lambda_2}}{\Gamma(\lambda_2)\Gamma_d(L_2)|L_2\mathbf{\Sigma}_2^{-1}|^{-L_2}}\right) + (L_2 - d)E_{\mathbf{C}^1} \{\ln |\mathbf{C}|\} \\ &\quad - (dL_2 + \lambda_2)E_{\mathbf{C}^1} \{\ln[\text{tr}(L_2\mathbf{\Sigma}_2^{-1}\mathbf{C}) + \lambda_2 - 1]\}. \end{aligned} \quad (13)$$

These expectations depend on  $E_{\mathbf{C}^1} \{\ln |\mathbf{C}|\}$  which is given as follows (see Appendix B-A for details):

$$\begin{aligned} E_{\mathbf{C}^1} \{\ln |\mathbf{C}|\} &= \psi_d(L_1) - d \ln L_1 + \ln |\mathbf{\Sigma}_1| \\ &\quad + d \ln(\lambda_1 - 1) - d\psi(\lambda_1) \end{aligned} \quad (14)$$

where function  $\psi_d(L)$  is the multivariate digamma function defined as  $\psi_d(L) = \partial \ln \Gamma_d(L) / \partial L = \sum_{j=0}^{d-1} \psi(L-j)$ , and  $\psi(\cdot)$  is the digamma function. In fact, the last expectation can be computed in another way by using the product model. Then it is equivalent to  $E_{\mathbf{C}^1} \{\ln |\mathbf{C}|\} = E_{\mathbf{X}^1} \{\ln |\mathbf{X}|\} + dE_{\tau^1} \{\ln \tau\}$ . The second expectation  $E_{\mathbf{C}^1} \{\ln[\text{tr}(L_1\mathbf{\Sigma}_1^{-1}\mathbf{C}) + \lambda_1 - 1]\}$  is given as follows (see Appendix B-B for details):

$$\begin{aligned} E_{\mathbf{C}^1} \{\ln[\text{tr}(L_1\mathbf{\Sigma}_1^{-1}\mathbf{C}) + \lambda_1 - 1]\} &= \ln(\lambda_1 - 1) - \psi(\lambda_1) + \psi(dL_1 + \lambda_1). \end{aligned} \quad (15)$$

Finally, to completely compute  $E_{\mathbf{C}^1} \{\ln f_{\mathbf{C}^2}(\mathbf{C})\}$ , the third expectation  $E_{\mathbf{C}^1} \{\ln[\text{tr}(L_2\mathbf{\Sigma}_2^{-1}\mathbf{C}) + \lambda_2 - 1]\}$  is needed and it is given as follows (see Appendix B-C for details):

$$\begin{aligned} E_{\mathbf{C}^1} \{\ln[\text{tr}(L_2\mathbf{\Sigma}_2^{-1}\mathbf{C}) + \lambda_2 - 1]\} &= \ln(\lambda_2 - 1) - \psi(\lambda_1) + \psi(\lambda_1 + dL_1) - \left(\prod_{i=1}^d \Lambda_i^{-L_1}\right) \\ &\quad \times \frac{\partial}{\partial a} F_D^{(d)}\left(\lambda_1 + dL_1, \underbrace{L_1, \dots, L_1}_d; \right. \\ &\quad \left. \lambda_1 + dL_1 + a; 1 - \frac{1}{\Lambda_1}, 1 - \frac{1}{\Lambda_2}, \dots, 1 - \frac{1}{\Lambda_d}\right) \Bigg|_{a=0} \end{aligned} \quad (16)$$

TABLE I  
ANALYTIC EXPRESSIONS OF KULLBACK–LEIBLER AND RÉNYI OF ORDER  $\beta$  DISTANCES BETWEEN TWO MATRIX-VARIATE  $\mathcal{G}_d^0(\boldsymbol{\Sigma}_1, L_1, \lambda_1)$  AND  $\mathcal{G}_d^0(\boldsymbol{\Sigma}_2, L_2, \lambda_2)$  DISTRIBUTIONS

Distance	Divergence expressions
Kullback-	$D_{\text{KL}}(\mathbf{C}^1 \ \mathbf{C}^2) = \ln \frac{\Gamma(dL_1 + \lambda_1)\Gamma(\lambda_2)\Gamma_d(L_2)}{\Gamma(dL_2 + \lambda_2)\Gamma(\lambda_1)\Gamma_d(L_1)} - L_2 \ln  \boldsymbol{\Lambda}  + (L_1 - L_2)[\psi_d(L_1) - d\psi(dL_1 + \lambda_1)] - (\lambda_1 - \lambda_2)[\psi(dL_1 + \lambda_1) - \psi(\lambda_1)] - (dL_2 + \lambda_2) \left( \prod_{i=1}^d \Lambda_i^{-L_1} \right) \frac{\partial}{\partial a} F_D^{(d)} \left( \lambda_1 + dL_1, \underbrace{L_1, \dots, L_1}_d; \lambda_1 + dL_1 + a; 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \Big _{a=0} \quad (20)$
Leibler	$d_{\text{KL}}(\mathbf{C}^1, \mathbf{C}^2) = (L_1 - L_2) [\ln  \boldsymbol{\Lambda}  + \psi_d(L_1) - \psi_d(L_2) + d\psi(dL_2 + \lambda_2) - d\psi(dL_1 + \lambda_1)] + (\lambda_1 - \lambda_2)[\psi(\lambda_1) - \psi(\lambda_2) + \psi(dL_2 + \lambda_2) - \psi(dL_1 + \lambda_1)] - (dL_2 + \lambda_2) \left( \prod_{i=1}^d \Lambda_i^{-L_1} \right) \frac{\partial}{\partial a} F_D^{(d)} \left( \lambda_1 + dL_1, \underbrace{L_1, \dots, L_1}_d; \lambda_1 + dL_1 + a; 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \Big _{a=0} - (dL_1 + \lambda_1) \left( \prod_{i=1}^d \Lambda_i^{L_2} \right) \frac{\partial}{\partial a} F_D^{(d)} \left( \lambda_2 + dL_2, \underbrace{L_2, \dots, L_2}_d; \lambda_2 + dL_2 + a; 1 - \Lambda_1, \dots, 1 - \Lambda_d \right) \Big _{a=0} \quad (21)$
Rényi of order $\beta$	$D_{\text{R}}^\beta(\mathbf{C}^1 \ \mathbf{C}^2) = \frac{1}{\beta - 1} \left[ \beta \ln \frac{\Gamma(dL_1 + \lambda_1)}{\Gamma_d(L_1)\Gamma(\lambda_1)} + (1 - \beta) \ln \frac{\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_2)\Gamma(\lambda_2)} + \ln \frac{\Gamma_d(\beta(L_1 - L_2) + L_2)\Gamma(\beta(\lambda_1 - \lambda_2) + \lambda_2)}{\Gamma(\beta(dL_1 + \lambda_1) + (1 - \beta)(dL_2 + \lambda_2))} - \beta \sum_{i=1}^d \ln \Lambda_i^{L_1} + \ln F_D^{(d)} \left( \beta(dL_1 + \lambda_1), \underbrace{\beta(L_1 - L_2) + L_2, \dots, \beta(L_1 - L_2) + L_2}_d; \beta(dL_1 + \lambda_1) + (1 - \beta)(dL_2 + \lambda_2); 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \right] \quad (22)$
	$d_{\text{R}}^\beta(\mathbf{C}^1, \mathbf{C}^2) = \frac{1}{2(\beta - 1)} \left[ \ln \frac{\Gamma(dL_1 + \lambda_1)\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_1)\Gamma(\lambda_1)\Gamma_d(L_2)\Gamma(\lambda_2)} + \ln \frac{\Gamma_d(\beta(L_1 - L_2) + L_2)\Gamma_d(\beta(L_2 - L_1) + L_1)}{\Gamma(\beta(dL_1 + \lambda_1) + (1 - \beta)(dL_2 + \lambda_2))} + \ln \frac{\Gamma(\beta(\lambda_1 - \lambda_2) + \lambda_2)\Gamma(\beta(\lambda_2 - \lambda_1) + \lambda_1)}{\Gamma(\beta(dL_2 + \lambda_2) + (1 - \beta)(dL_1 + \lambda_1))} - L_1 \sum_{i=1}^d \ln \Lambda_i + \ln F_D^{(d)} \left( \beta(dL_1 + \lambda_1), \underbrace{\beta(L_1 - L_2) + L_2, \dots, \beta(L_1 - L_2) + L_2}_d; \beta(dL_1 + \lambda_1) + (1 - \beta)(dL_2 + \lambda_2); 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) + \ln F_D^{(d)} \left( (1 - \beta)(dL_1 + \lambda_1), \underbrace{\beta(L_2 - L_1) + L_1, \dots, \beta(L_2 - L_1) + L_1}_d; \beta(dL_2 + \lambda_2) + (1 - \beta)(dL_1 + \lambda_1); 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \right]. \quad (23)$
	$d_{\text{R}}^\beta(\mathbf{C}^1, \mathbf{C}^2) = \frac{1}{2(\beta - 1)} \left[ \ln \frac{\Gamma(dL_1 + \lambda_1)\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_1)\Gamma(\lambda_1)\Gamma_d(L_2)\Gamma(\lambda_2)} + \ln \frac{\Gamma_d(\beta(L_1 - L_2) + L_2)\Gamma_d(\beta(L_2 - L_1) + L_1)}{\Gamma(\beta(dL_1 + \lambda_1) + (1 - \beta)(dL_2 + \lambda_2))} + \ln \frac{\Gamma(\beta(\lambda_1 - \lambda_2) + \lambda_2)\Gamma(\beta(\lambda_2 - \lambda_1) + \lambda_1)}{\Gamma(\beta(dL_2 + \lambda_2) + (1 - \beta)(dL_1 + \lambda_1))} + L_2 \sum_{i=1}^d \ln \Lambda_i + \ln F_D^{(d)} \left( \beta(dL_2 + \lambda_2), \underbrace{\beta(L_2 - L_1) + L_1, \dots, \beta(L_2 - L_1) + L_1}_d; \beta(dL_2 + \lambda_2) + (1 - \beta)(dL_1 + \lambda_1); 1 - \Lambda_1, \dots, 1 - \Lambda_d \right) + \ln F_D^{(d)} \left( (1 - \beta)(dL_2 + \lambda_2), \underbrace{\beta(L_1 - L_2) + L_2, \dots, \beta(L_1 - L_2) + L_2}_d; \beta(dL_1 + \lambda_1) + (1 - \beta)(dL_2 + \lambda_2); 1 - \Lambda_1, \dots, 1 - \Lambda_d \right) \right]. \quad (24)$

where  $F_D^{(d)}(\cdot)$  is the Lauricella D-hypergeometric series of  $d$  variables (see Appendix G for more details), and  $\Lambda_1, \dots, \Lambda_d$  are the eigenvalues of the positive definite complex Hermitian matrix  $\boldsymbol{\Lambda} = [L_2(\lambda_1 - 1)/L_1(\lambda_2 - 1)]\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1$ .

The analytic expressions for the stochastic divergence  $D_{\text{KL}}(\mathbf{C}^1 \|\mathbf{C}^2)$  and distance  $d_{\text{KL}}(\mathbf{C}^1, \mathbf{C}^2)$  between  $\mathcal{G}_d^0(\boldsymbol{\Sigma}_1, L_1, \lambda_1)$  and  $\mathcal{G}_d^0(\boldsymbol{\Sigma}_2, L_2, \lambda_2)$  distributions are given, respectively, by (20) and (21) in Table I.

1) Case  $L_1 = L_2 = L$ : The positive definite complex Hermitian matrix becomes  $\boldsymbol{\Lambda} = (\lambda_1 - 1)(\lambda_2 - 1)^{-1}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1$  and the symmetric Kullback–Leibler divergence is as follows:

$$d_{\text{KL}}(\mathbf{C}^1, \mathbf{C}^2) = (\lambda_1 - \lambda_2) [\psi(\lambda_1) - \psi(\lambda_2) + \psi(dL + \lambda_2) - \psi(dL + \lambda_1)]$$

$$-(dL + \lambda_2) \left( \prod_{i=1}^d \Lambda_i^{-L} \right) \frac{\partial}{\partial a} F_D^{(d)} \left( \lambda_1 + dL, L, \dots, L; \lambda_1 + dL + a; 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \Big|_{a=0} - (dL + \lambda_1) \left( \prod_{i=1}^d \Lambda_i^L \right) \frac{\partial}{\partial a} F_D^{(d)} \left( \lambda_2 + dL, L, \dots, L; \lambda_2 + dL + a; 1 - \Lambda_1, \dots, 1 - \Lambda_d \right) \Big|_{a=0}. \quad (17)$$

2) Case  $L_1 = L_2 = L$  and  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ : The positive definite complex Hermitian matrix becomes  $\boldsymbol{\Lambda} = (\lambda_1 - 1)(\lambda_2 - 1)^{-1}\mathbf{I}_d$  and the Lauricella function becomes equivalent to the Gauss hypergeometric function as follows:

$$F_D^{(d)}(\lambda_1 + dL, L, \dots, L; \lambda_1 + dL + a; 1 - \Lambda^{-1}, \dots, 1 - \Lambda^{-1}) = {}_2F_1(\lambda_1 + dL, dL, \lambda_1 + dL + a; 1 - \Lambda^{-1}) \quad (18)$$

where  $\Lambda = (\lambda_1 - 1)(\lambda_2 - 1)^{-1}$ . As a consequence, the Kullback–Leibler distance is as follows:

$$\begin{aligned} d_{\text{KL}}(\mathbf{C}^1, \mathbf{C}^2) &= (\lambda_1 - \lambda_2)[\psi(\lambda_1) - \psi(\lambda_2) + \psi(dL + \lambda_2) - \psi(dL + \lambda_1)] \\ &\quad - (dL + \lambda_2)\Lambda^{-dL} \frac{\partial}{\partial a} {}_2F_1(\lambda_1 + dL, dL; \\ &\quad \lambda_1 + dL + a; 1 - \Lambda^{-1})\Big|_{a=0} - (dL + \lambda_1)\Lambda^{dL} \\ &\quad \frac{\partial}{\partial a} {}_2F_1(\lambda_2 + dL, dL; \lambda_2 + dL + a; 1 - \Lambda)\Big|_{a=0}. \end{aligned} \quad (19)$$

### B. Rényi Distance of Order $\beta$

Rényi [17] presented the first parametric generalization of (10) given by

$$D_{\text{R}}^{\beta}(\mathbf{C}^1 \parallel \mathbf{C}^2) = \frac{1}{\beta - 1} \ln \int_{\Omega_+} f_{\mathbf{C}^1}^{\beta}(\mathbf{C}) f_{\mathbf{C}^2}^{1-\beta}(\mathbf{C}) d\mathbf{C} \quad (25)$$

$$= \frac{1}{\beta - 1} \ln E_{\mathbf{C}^1} \left\{ \left( \frac{f_{\mathbf{C}^1}(\mathbf{C})}{f_{\mathbf{C}^2}(\mathbf{C})} \right)^{\beta-1} \right\} \quad (26)$$

where  $0 < \beta < 1$ . The Rényi distance is given by

$$d_{\text{R}}^{\beta}(\mathbf{C}^1, \mathbf{C}^2) = \frac{1}{2} (D_{\text{R}}^{\beta}(\mathbf{C}^1 \parallel \mathbf{C}^2) + D_{\text{R}}^{\beta}(\mathbf{C}^2 \parallel \mathbf{C}^1)). \quad (27)$$

It is clear that the KL divergence is a particular case of the Rényi divergence when  $\beta \rightarrow 1$ . The analytic expression for the stochastic distance  $D_{\text{R}}^{\beta}(\mathbf{C}^1 \parallel \mathbf{C}^2)$  between two matrix-variate  $\mathcal{G}_d^0(\Sigma_1, L_1, \lambda_1)$  and  $\mathcal{G}_d^0(\Sigma_2, L_2, \lambda_2)$  distributions is given by (22) in Table I. The computation details are given in Appendix C. The symmetric expression of the Rényi divergence is then given by (23). Using propriety (116),  $d_{\text{R}}^{\beta}(\mathbf{C}^1, \mathbf{C}^2)$  can also be written otherwise by (24).

1) *Case  $L_1 = L_2 = L$* : The distance,  $d_{\text{R}}^{\beta}(\mathbf{C}^1, \mathbf{C}^2)$ , given by (23) becomes in this case as follows:

$$\begin{aligned} d_{\text{R}}^{\beta}(\mathbf{C}^1, \mathbf{C}^2) &= \frac{1}{2(\beta - 1)} \left[ \ln \frac{\Gamma(dL + \lambda_1)\Gamma(dL + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} - L_1 \sum_{i=1}^d \ln \Lambda_i \right. \\ &\quad \left. + \ln \frac{\Gamma(\beta(\lambda_1 - \lambda_2) + \lambda_2)\Gamma(\beta(\lambda_2 - \lambda_1) + \lambda_1)}{\Gamma(\beta(\lambda_1 - \lambda_2) + dL + \lambda_2)\Gamma(\beta(\lambda_2 - \lambda_1) + dL + \lambda_1)} \right. \\ &\quad \left. + \ln F_D^{(d)}(\beta(dL + \lambda_1), L, \dots, L; \beta(\lambda_1 - \lambda_2) + dL + \lambda_2; \right. \\ &\quad \left. 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1}) \right. \\ &\quad \left. + \ln F_D^{(d)}((1 - \beta)(dL + \lambda_1), L, \dots, L; \beta(\lambda_2 - \lambda_1) \right. \\ &\quad \left. + dL + \lambda_1; 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1}) \right]. \end{aligned} \quad (28)$$

### C. Hellinger Distance

The Hellinger distance [38] is defined as follows:

$$d_{\text{H}}(\mathbf{C}^1, \mathbf{C}^2) = 1 - \int_{\Omega_+} \sqrt{f_{\mathbf{C}^1}(\mathbf{C})f_{\mathbf{C}^2}(\mathbf{C})} d\mathbf{C} \quad (29)$$

$$= 1 - \exp \left( -\frac{1}{2} d_{\text{R}}^{1/2}(\mathbf{C}^1, \mathbf{C}^2) \right) \quad (30)$$

with

$$\begin{aligned} &\int_{\Omega_+} \sqrt{f_{\mathbf{C}^1}(\mathbf{C})f_{\mathbf{C}^2}(\mathbf{C})} d\mathbf{C} \\ &= \left( \frac{\Gamma(dL_1 + \lambda_1)\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_1)\Gamma_d(L_2)\Gamma(\lambda_1)\Gamma(\lambda_2)} \right)^{\frac{1}{2}} \\ &\quad \times \frac{\Gamma_d(L_1/2 + L_2/2)\Gamma(\lambda_1/2 + \lambda_2/2)}{\Gamma(dL_1/2 + dL_2/2 + \lambda_1/2 + \lambda_2/2)} \left( \prod_{i=1}^d \Lambda_i^{-\frac{L_1}{2}} \right) \\ &\quad \times F_D^{(d)} \left( \frac{dL_1 + \lambda_1}{2}, \frac{L_1 + L_2}{2}, \dots, \frac{L_1 + L_2}{2}; \right. \\ &\quad \left. \frac{d(L_1 + L_2) + \lambda_1 + \lambda_2}{2}; \right. \\ &\quad \left. 1 - \frac{1}{\Lambda_1}, 1 - \frac{1}{\Lambda_2}, \dots, 1 - \frac{1}{\Lambda_d} \right). \end{aligned} \quad (31)$$

The details of this derivation are given in Appendix D. This expression can be written otherwise by using the following relation deduced from (116):

$$\begin{aligned} &\prod_{i=1}^d \Lambda_i^{-\frac{L_1}{2}} F_D^{(d)} \left( \frac{dL_1 + \lambda_1}{2}, \frac{L_1 + L_2}{2}, \dots, \frac{L_1 + L_2}{2}; \right. \\ &\quad \left. \frac{d(L_1 + L_2) + \lambda_1 + \lambda_2}{2}; 1 - \frac{1}{\Lambda_1}, 1 - \frac{1}{\Lambda_2}, \dots, 1 - \frac{1}{\Lambda_d} \right) \\ &= \prod_{i=1}^d \Lambda_i^{\frac{L_2}{2}} F_D^{(d)} \left( \frac{dL_2 + \lambda_2}{2}, \frac{L_1 + L_2}{2}, \dots, \frac{L_1 + L_2}{2}; \right. \\ &\quad \left. \frac{d(L_1 + L_2) + \lambda_1 + \lambda_2}{2}; 1 - \Lambda_1, 1 - \Lambda_2, \dots, 1 - \Lambda_d \right). \end{aligned} \quad (32)$$

As a result, both expressions (39) and (40) can be established for the Hellinger distance between two matrix-variate  $\mathcal{G}_d^0(\Sigma_1, L_1, \lambda_1)$  and  $\mathcal{G}_d^0(\Sigma_2, L_2, \lambda_2)$  distributions. The analytic expressions are given in Table II.

1) *Case  $L_1 = L_2 = L$* : The symmetric Hellinger distance  $d_{\text{H}}(\mathbf{C}^1, \mathbf{C}^2)$  given by (39) becomes as follows:

$$\begin{aligned} d_{\text{H}}(\mathbf{C}^1, \mathbf{C}^2) &= 1 - \left( \frac{\Gamma(dL + \lambda_1)\Gamma(dL + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{\lambda_1}{2} + \frac{\lambda_2}{2}\right)}{\Gamma\left(dL + \frac{\lambda_1}{2} + \frac{\lambda_2}{2}\right)} \\ &\quad \times \left( \prod_{i=1}^d \Lambda_i^{-\frac{L}{2}} \right) F_D^{(d)} \left( \frac{dL + \lambda_1}{2}, \underbrace{L, \dots, L}_d; dL + \frac{\lambda_1 + \lambda_2}{2}; \right. \\ &\quad \left. 1 - \frac{1}{\Lambda_1}, 1 - \frac{1}{\Lambda_2}, \dots, 1 - \frac{1}{\Lambda_d} \right). \end{aligned} \quad (33)$$

### D. Bhattacharyya Distance

The Bhattacharyya distance measure [18] is given by

$$d_{\text{B}}(\mathbf{C}^1, \mathbf{C}^2) = -\ln \int_{\Omega_+} \sqrt{f_{\mathbf{C}^1}(\mathbf{C})f_{\mathbf{C}^2}(\mathbf{C})} d\mathbf{C} \quad (34)$$

$$= \frac{1}{2} d_{\text{R}}^{1/2}(\mathbf{C}^1, \mathbf{C}^2). \quad (35)$$

The developed expression of  $d_{\text{B}}(\mathbf{C}^1, \mathbf{C}^2)$  is given by (37) in Table II. Using relation (32) deduced from (116), one can deduce a second expression of the Bhattacharyya distance given by (38).

TABLE II  
ANALYTIC EXPRESSIONS OF BHATTACHARYYA AND HELLINGER DISTANCES BETWEEN TWO MATRIX-VARIATE  $\mathcal{G}_d^0(\boldsymbol{\Sigma}_1, L_1, \lambda_1)$  AND  $\mathcal{G}_d^0(\boldsymbol{\Sigma}_2, L_2, \lambda_2)$  DISTRIBUTIONS

Distance	Divergence expressions
Bhattacharyya	$d_B(\mathbf{C}^1, \mathbf{C}^2) = -\frac{1}{2} \ln \frac{\Gamma(dL_1 + \lambda_1)\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_1)\Gamma_d(L_2)\Gamma(\lambda_1)\Gamma(\lambda_2)} - \ln \frac{\Gamma_d(L_1/2 + L_2/2)\Gamma(\lambda_1/2 + \lambda_2/2)}{\Gamma(dL_1/2 + dL_2/2 + \lambda_1/2 + \lambda_2/2)} + \frac{L_1}{2} \sum_{i=1}^d \ln \Lambda_i - \ln F_D^{(d)}\left(\frac{dL_1 + \lambda_1}{2}, \underbrace{\frac{L_1 + L_2}{2}, \dots, \frac{L_1 + L_2}{2}}_d; \frac{d(L_1 + L_2) + \lambda_1 + \lambda_2}{2}; 1 - \frac{1}{\Lambda_1}, 1 - \frac{1}{\Lambda_2}, \dots, 1 - \frac{1}{\Lambda_d}\right). \quad (37)$
	$d_B(\mathbf{C}^1, \mathbf{C}^2) = -\frac{1}{2} \ln \frac{\Gamma(dL_1 + \lambda_1)\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_1)\Gamma_d(L_2)\Gamma(\lambda_1)\Gamma(\lambda_2)} - \ln \frac{\Gamma_d(L_1/2 + L_2/2)\Gamma(\lambda_1/2 + \lambda_2/2)}{\Gamma(dL_1/2 + dL_2/2 + \lambda_1/2 + \lambda_2/2)} - \frac{L_2}{2} \sum_{i=1}^d \ln \Lambda_i - \ln F_D^{(d)}\left(\frac{dL_2 + \lambda_2}{2}, \underbrace{\frac{L_1 + L_2}{2}, \dots, \frac{L_1 + L_2}{2}}_d; \frac{d(L_1 + L_2) + \lambda_1 + \lambda_2}{2}; 1 - \Lambda_1, 1 - \Lambda_2, \dots, 1 - \Lambda_d\right). \quad (38)$
Hellinger	$d_H(\mathbf{C}^1, \mathbf{C}^2) = 1 - \left(\frac{\Gamma(dL_1 + \lambda_1)\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_1)\Gamma_d(L_2)\Gamma(\lambda_1)\Gamma(\lambda_2)}\right)^{\frac{1}{2}} \frac{\Gamma_d(L_1/2 + L_2/2)\Gamma(\lambda_1/2 + \lambda_2/2)}{\Gamma(dL_1/2 + dL_2/2 + \lambda_1/2 + \lambda_2/2)} \left(\prod_{i=1}^d \Lambda_i^{-\frac{L_1}{2}}\right) \times F_D^{(d)}\left(\frac{dL_1 + \lambda_1}{2}, \underbrace{\frac{L_1 + L_2}{2}, \dots, \frac{L_1 + L_2}{2}}_d; \frac{d(L_1 + L_2) + \lambda_1 + \lambda_2}{2}; 1 - \frac{1}{\Lambda_1}, 1 - \frac{1}{\Lambda_2}, \dots, 1 - \frac{1}{\Lambda_d}\right). \quad (39)$
	$d_H(\mathbf{C}^1, \mathbf{C}^2) = 1 - \left(\frac{\Gamma(dL_1 + \lambda_1)\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_1)\Gamma_d(L_2)\Gamma(\lambda_1)\Gamma(\lambda_2)}\right)^{\frac{1}{2}} \frac{\Gamma_d(L_1/2 + L_2/2)\Gamma(\lambda_1/2 + \lambda_2/2)}{\Gamma(dL_1/2 + dL_2/2 + \lambda_1/2 + \lambda_2/2)} \left(\prod_{i=1}^d \Lambda_i^{\frac{L_2}{2}}\right) \times F_D^{(d)}\left(\frac{dL_2 + \lambda_2}{2}, \underbrace{\frac{L_1 + L_2}{2}, \dots, \frac{L_1 + L_2}{2}}_d; \frac{d(L_1 + L_2) + \lambda_1 + \lambda_2}{2}; 1 - \Lambda_1, 1 - \Lambda_2, \dots, 1 - \Lambda_d\right). \quad (40)$

1) Case  $L_1 = L_2 = L$ : Symmetric  $d_B(\mathbf{C}^1, \mathbf{C}^2)$  given by (37) becomes in this case as follows:

$$\begin{aligned} d_B(\mathbf{C}^1, \mathbf{C}^2) &= -\frac{1}{2} \ln \frac{\Gamma(dL + \lambda_1)\Gamma(dL + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \\ &\quad - \ln \frac{\Gamma(\lambda_1/2 + \lambda_2/2)}{\Gamma(dL + \lambda_1/2 + \lambda_2/2)} + \frac{L}{2} \sum_{i=1}^d \ln \Lambda_i \\ &\quad - \ln F_D^{(d)}\left(\frac{dL + \lambda_1}{2}, L, \dots, L; dL + \frac{\lambda_1 + \lambda_2}{2}; 1 - \frac{1}{\Lambda_1}, \right. \\ &\quad \left. 1 - \frac{1}{\Lambda_2}, \dots, 1 - \frac{1}{\Lambda_d}\right). \quad (36) \end{aligned}$$

### E. Implementation of Lauricella Function

In order to compute Kullback–Leibler divergence given by (20) and its symmetric version (21), the Lauricella function and its derivative need to be computed. The reader may refer to Appendix G, where we provide two different ways to compute Lauricella's  $F_D$ . The first one consists of using the integral form given by (115) which is evaluated by symbolic integration tools in MATLAB, whereas the second one resort to Lauricella series given by (113). In this article, the Lauricella series is more convenient than the integral form because of the ease of computation and implementation. However, the convergence of the series needs to be guaranteed. In doing so, several transformations can be applied to the series using relations (116) and (117) included in Appendix G. In addition, for the feasible computation of the series, the series indices are limited

to  $N_{\max}$  instead of infinity. The last parameter is chosen to ensure a good approximation of the Lauricella function. The achieved accuracy given by the selected parameter will be discussed later in the next paragraph.

Partial derivatives  $(\partial/\partial a)F_D^{(d)}(\cdot)$  are approximated using a numerical differentiation (For more details, see Appendix E.) as follows:

$$\begin{aligned} &\left(\prod_{i=1}^d \Lambda_i^{-L_1}\right) \frac{\partial}{\partial a} F_D^{(d)}\left(\lambda_1 + dL_1, L_1, \dots, L_1; \delta_1; \right. \\ &\quad \left. 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1}\right) \\ &\approx \frac{F_D^{(d)}(a, L_1, \dots, L_1; \delta_1; 1 - \Lambda_1, \dots, 1 - \Lambda_d) - 1}{a} \quad (41) \\ &\approx \frac{\Lambda_1^{-a} F_D^{(d)}\left(a, \lambda_1 + a, \dots, L_1; \delta_1; 1 - \frac{1}{\Lambda_1}, 1 - \frac{\Lambda_2}{\Lambda_1}, \dots, 1 - \frac{\Lambda_d}{\Lambda_1}\right) - 1}{a} \quad (42) \end{aligned}$$

with  $\delta_1 = \lambda_1 + dL_1 + a$  and  $a = 10^{-3}$ . The approximation accuracy of the numerical derivatives depends on the smoothness of the function. In the same manner, the second derivative of (21) is approximated as follows:

$$\begin{aligned} &\left(\prod_{i=1}^d \Lambda_i^{L_2}\right) \frac{\partial}{\partial a} F_D^{(d)}(\lambda_2 + dL_2, L_2, \dots, L_2; \delta_2; \\ &\quad 1 - \Lambda_1, \dots, 1 - \Lambda_d) \\ &\approx \frac{F_D^{(d)}\left(a, L_2, \dots, L_2; \delta_2; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) - 1}{a} \quad (43) \\ &\approx \frac{\Lambda_1^a F_D^{(d)}\left(a, \lambda_2 + a, \dots, L_2; \delta_2; 1 - \Lambda_1, 1 - \frac{\Lambda_1}{\Lambda_2}, \dots, 1 - \frac{\Lambda_1}{\Lambda_d}\right) - 1}{a} \quad (44) \end{aligned}$$

$$\approx \frac{\Lambda_d^a F_D^{(d)}\left(a, \lambda_2 + a, \dots, L_2; \delta_2; 1 - \Lambda_d, 1 - \frac{\Lambda_d}{\Lambda_{d-1}}, \dots, 1 - \frac{\Lambda_d}{\Lambda_1}\right) - 1}{a} \quad (45)$$

with  $\delta_2 = \lambda_2 + dL_2 + a$ . Then, to compute the Kullback–Leibler, the following Algorithm 1 is used.

---

**Algorithm 1** Computation of  $d_{KL}$ 


---

- 1: Compute the eigenvalues of  $\mathbf{\Lambda} = \frac{L_2(\lambda_1 - 1)}{L_1(\lambda_2 - 1)} \mathbf{\Sigma}_2^{-1} \mathbf{\Sigma}_1$ .
  - 2: Sort in descending order:  $\Lambda_1 > \Lambda_2 > \dots > \Lambda_d > 0$
  - 3: **Main:**
  - 4:   **If**  $\Lambda_1 > 1$
  - 5:     Use (42)
  - 6:   **Else** Use (41); **end**
  - 7:   **If**  $\Lambda_d > 1$
  - 8:     Use (43)
  - 9:   **Else** Use (45); **end**
  - 10: **Return**  $d_{KL}$
- 

The computation accuracy of the Lauricella series is evaluated by comparing it with the Gauss hypergeometric function in the case, where they are equivalents for some particular values of  $(\Lambda_1, \dots, \Lambda_d)$ . Indeed, when  $\Lambda_1 = \dots = \Lambda_d = \Lambda$ , we have the following relation:

$$F_D^{(d)}(a, L_1, \dots, L_1; \lambda_1 + dL_1 + a; 1 - \Lambda, \dots, 1 - \Lambda) = {}_2F_1(a, dL_1, \lambda_1 + dL_1 + a, 1 - \Lambda). \quad (46)$$

Table III shows the computation of  $F_D^{(d)}(\cdot)$  and  ${}_2F_1(\cdot)$ ,<sup>1</sup> along with the absolute value of error  $|\epsilon|$ , where  $a = 10^{-3}$ ,  $L_1 = 6$ ,  $\lambda_1 = 2$ ,  $d = 3$ ,  $N_{\max} = \{10, 20, 30\}$ . Note that the error is reasonably low and increases for the values of  $1 - \Lambda$  near 1, as expected. The computation times of  $F_D^{(d)}(\cdot)$  are recorded using MATLAB on a 2.6-GHz processor with 16 GB of memory. The mean values are respectively  $\{1.24, 4.30, 12.52\}$  s for  $N_{\max} = \{10, 20, 30\}$ . The value of  $N_{\max}$  can be increased to further improve accuracy but it will increase the computation time. It is worth noticing that when  $1 - \Lambda = 1$ , we have the following relation:

$$F_D^{(d)}(a, L_1, \dots, L_1; \lambda_1 + dL_1 + a; 1, \dots, 1) = \frac{B(a, \lambda_1)}{B(a, \lambda_1 + dL_1)} \quad (47)$$

where  $B(\cdot)$  is the beta function.

### F. Comparison With MC Technique

When it is hard to compute the stochastic distances because of the integral involved in computing them, we have to resort to approximation methods. In the literature, there have been several methods addressing this issue. The most important one is the MC sampling. The MC method can estimate the stochastic distances provided that we have a large number of i.i.d. samples. The idea of the method is to sample a large number of samples and use them to compute the summation, instead of the integral, over these samples. The approximation

<sup>1</sup>The Gauss hypergeometric function  ${}_2F_1$  is provided in MATLAB by the command *hypergeom*.

TABLE III

$$a = 10^{-3}, L_1 = 6, \lambda_1 = 2, d = 3, N_{\max} = \{10, 20, 30\}$$

$1 - \Lambda$	$N_{\max}$	$F_D^{(d)}(\cdot)$	${}_2F_1(\cdot)$	$ \epsilon $
0.1	10	1.00009433	1.00009433	6.66133814e-16
	20	1.00009433		6.66133814e-16
	30	1.00009433		6.66133814e-16
0.3	10	1.00031509	1.00031509	1.91180404e-13
	20	1.00031509		1.11022302e-15
	30	1.00031509		1.11022302e-15
0.5	10	1.00059968	1.00059968	1.97363236e-10
	20	1.00059968		2.22044604e-15
	30	1.00059968		0
0.7	10	1.00100181	1.00100185	4.26687991e-8
	20	1.00100185		3.10156345e-11
	30	1.00100185		5.68434188e-14
0.9	10	1.00169872	1.00170577	7.04983285e-6
	20	1.00170553		2.33947369e-7
	30	1.00170575		1.34357198e-8
1	10	1.00238767	1.00255069	1.63020056e-4
	20	1.00249318		5.75072443e-5
	30	1.00252148		2.92069654e-5

result will be very close to the true value of the stochastic distance when a large number of samples goes to infinity.

Let  $N$  be i.i.d. realizations  $\mathbf{C}_i$  and  $\mathbf{C}'_i$ ,  $i = \{1, \dots, N\}$  of a random matrix  $\mathbf{C}^1$  and  $\mathbf{C}^2$ , respectively, which are distributed according to the density function  $f_{\mathbf{C}^1}(\mathbf{C}_i)$  and  $f_{\mathbf{C}^2}(\mathbf{C}'_i)$ . The stochastic distances are calculated with the sample mean given a set of  $\{\mathbf{C}_i\}_{i=1}^N$  and  $\{\mathbf{C}'_i\}_{i=1}^N$  as follows:

$$\hat{d}_{KL}(\mathbf{C}^1, \mathbf{C}^2) = \frac{1}{N} \sum_{i=1}^N \ln \frac{f_{\mathbf{C}^1}(\mathbf{C}_i)}{f_{\mathbf{C}^2}(\mathbf{C}_i)} + \frac{1}{N} \sum_{i=1}^N \ln \frac{f_{\mathbf{C}^2}(\mathbf{C}'_i)}{f_{\mathbf{C}^1}(\mathbf{C}'_i)} \quad (48)$$

$$\hat{d}_H(\mathbf{C}^1, \mathbf{C}^2) = 1 - \frac{1}{N} \sum_{i=1}^N \left( \frac{f_{\mathbf{C}^1}(\mathbf{C}_i)}{f_{\mathbf{C}^2}(\mathbf{C}_i)} \right)^{-\frac{1}{2}} \quad (49)$$

$$\hat{d}_B(\mathbf{C}^1, \mathbf{C}^2) = -\ln \frac{1}{N} \sum_{i=1}^N \left( \frac{f_{\mathbf{C}^1}(\mathbf{C}_i)}{f_{\mathbf{C}^2}(\mathbf{C}_i)} \right)^{-\frac{1}{2}}. \quad (50)$$

Some simulations are used to assess the deviation and variation of the sample mean with the theoretical stochastic distances. Then, the bias and the Mean Square Error (MSE) defined, respectively, by  $E\{\hat{d}_x\} - d_x$  and  $E\{(\hat{d}_x - d_x)^2\}$ , where  $x \in \{KL, H \text{ (Hellinger)}, B \text{ (Bhattacharyya)}\}$  are computed for finite size samples during the MC experiments. The parameters used for simulations are  $(\mathbf{\Sigma}_i, L_i, \lambda_i)$ ,  $i = \{1, 2\}$  of  $\mathcal{G}_d^0$  distribution, and the considered sample sizes are  $\{1024, 2048, 4096, 8192, 16384, 32768\}$ . For each sample size configuration, the experiment is repeated 2000 times. Parameters of simulations  $(\mathbf{\Sigma}_i, L_i, \lambda_i)$  are given in Table IV and we set  $a = 10^{-3}$  and  $N_{\max} = 30$  to compute the Lauricella series. The result of our simulations is depicted in Fig. 1, where the absolute value of the bias and the MSE is illustrated as a function of the sample sizes. The reader may notice that as the sample size increases, the bias and MSE of  $\hat{d}_{KL}$ ,  $\hat{d}_H$ , and  $\hat{d}_B$  decrease. It is clear that by the law of large numbers,  $\hat{d}_{KL}$ ,  $\hat{d}_H$ , and  $\hat{d}_B$  converge to  $d_{KL}$ ,  $d_H$ , and  $d_B$ , respectively.

### IV. EM-BASED ESTIMATOR FOR $\mathcal{G}_d^0$ DISTRIBUTION

In order to compute these distance measures to change detection, the parameters of the matrix-variate  $\mathcal{G}_d^0$  distribution

TABLE IV  
PARAMETERS ( $\Sigma_1, L_1, \lambda_1$ ) AND ( $\Sigma_2, L_2, \lambda_2$ ) USED TO COMPUTE STOCHASTIC DISTANCES AND FOR MC SAMPLING

L	$\lambda$	$\Sigma$	$\Sigma_{11}, \Sigma_{22}, \Sigma_{33}, \Sigma_{12}, \Sigma_{13}, \Sigma_{23}$
5	15	$\Sigma_1$	3.44, 0.81, 0.43, 0.48 + 0.41i, 0.07 + 0.23i, 0.11 - 0.31i
7	2	$\Sigma_2$	1.65, 1.11, 0.56, 0.22 + 0.35i, 0.01 + 0.27i, 0.14 - 0.04i

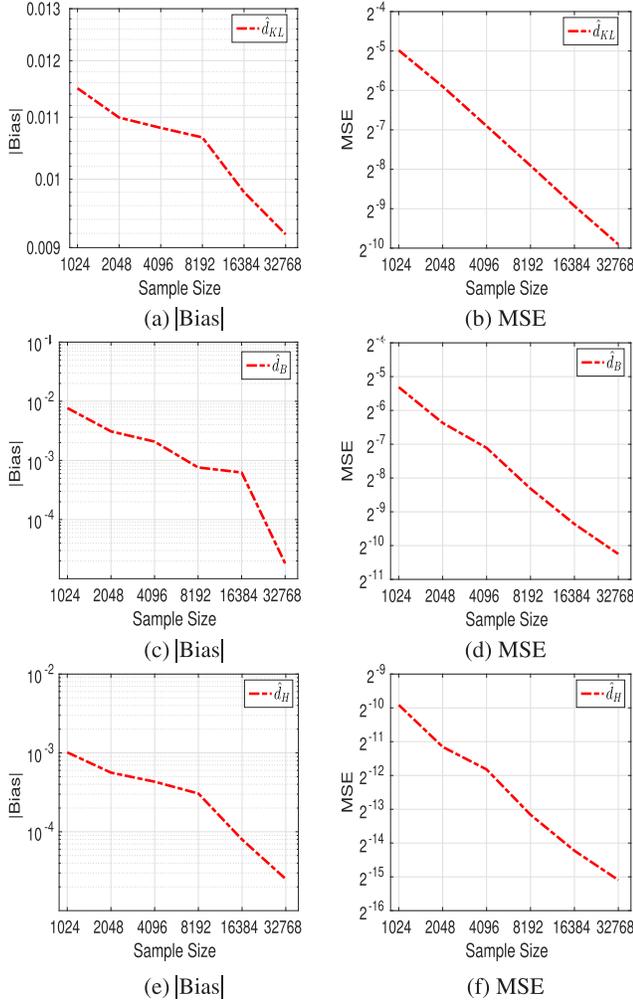


Fig. 1. Bias:  $E\{\hat{d}_x\} - d_x$  and MSE:  $E\{(\hat{d}_x - d_x)^2\}$  of sample mean estimator, where  $x \in \{KL, H$  (Hellinger), B (Bhattacharyya) $\}$ .

need to be estimated first. In particular, we estimate the ENLs, the covariance matrix of speckle component  $\Sigma$  and the texture distribution parameter  $\lambda$ . A developed method based on the EM algorithm and presented by Bouhlef and Meric [36] is used in this article to compute the maximum of likelihood (ML) estimates of the unknown parameters of the  $\mathcal{G}_d^0$  distribution. In [36], it was shown that the method had a good performance in terms of bias and MSE compared to the performance of other known estimators such as the matrix log-cumulants (MLC) [28] and the multivariate fractional moment (MFM) [39].

We consider  $\mathbf{C} = \{\mathbf{C}_i, 1 \leq i \leq N\}$  a set of i.i.d. random matrices. We assume that the random matrix  $\mathbf{C}_i$  follows the  $\mathcal{G}_d^0$  distribution. We suppose that each covariance matrix  $\mathbf{C}_i$  has a corresponding unobserved and hiding texture  $\tau_i$ . We assume

that the sequence  $\{\tau_i, 1 \leq i \leq N\}$  is also i.i.d. random variables. The EM algorithm is used to find the estimation  $\hat{\theta} = (\hat{\Sigma}, \hat{L}, \hat{\lambda})$  given a current estimate  $\theta'$ . The equations needed to estimate these parameters are as follows:

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^N E_{\tau_i|\mathbf{C}_i} \left\{ \frac{1}{\tau_i} \mathbf{C}_i, \theta' \right\} \mathbf{C}_i \quad (51)$$

$$\begin{aligned} & d \ln \hat{L} + d - \psi_d(\hat{L}) \\ &= \frac{1}{N} \sum_{i=1}^N \left( E_{\tau_i|\mathbf{C}_i} \left\{ \frac{1}{\tau_i} \mathbf{C}_i, \theta' \right\} \text{tr}(\hat{\Sigma}^{-1} \mathbf{C}_i) - \ln |\hat{\Sigma}^{-1} \mathbf{C}_i| \right) \\ &+ \frac{d}{N} \sum_{i=1}^N E_{\tau_i|\mathbf{C}_i} \{ \ln \tau_i | \mathbf{C}_i, \theta' \} \end{aligned} \quad (52)$$

$$\ln(\hat{\lambda} - 1) - \psi(\hat{\lambda}) + \frac{\hat{\lambda}}{\hat{\lambda} - 1} = \frac{1}{N} \sum_{i=1}^N E_{\tau_i|\mathbf{C}_i} \left\{ \frac{1}{\tau_i} + \ln \tau_i | \mathbf{C}_i, \theta' \right\}. \quad (53)$$

The posterior pdf of  $\tau_i$  given  $\mathbf{C}_i$  is provided in a closed form in Table V. It corresponds to an inverse gamma distribution with parameters ( $\alpha_1 = dL + \lambda, \beta_1 = L \text{tr}(\Sigma^{-1} \mathbf{C}_i) + \lambda - 1$ ). The posterior expectation expression in terms of these parameters is defined in Table VI (see Appendix F). An explicit solution for  $\hat{L}$  and  $\hat{\lambda}$  is not obtainable directly from (52) and (53). The trust region method [40] is proposed to solve it iteratively. Specifically, the *fsolve* function MATLAB is utilized to obtain the parameter estimate. The estimation approach of the parameters is summarized in Algorithm 2.

## V. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the proposed distance measures is evaluated by comparison with distances based on the scaled Wishart distribution. The receiver operating characteristic (ROC) curve is performed for both methods independently of any thresholding algorithm leading to the accurate evaluation of the change map. The performance measure is taken here as the area under the ROC curve (AUC). The ROC curve is the evolution of the true positive rate (TPR) as a function of false positive rate (FPR) [41]. The AUC is a good indicator of changes. Indeed, a better performance is found for larger AUC.

### Algorithm 2 Estimation Approach of the Parameters ( $\Sigma, L, \lambda$ )

- 1: **Input:**  $N, \mathbf{C}_i, d, \epsilon$
- 2: **Output:**  $\hat{\theta} = (\hat{\Sigma}, \hat{L}, \hat{\lambda})$
- 3: **Initialization:**
- 4:     Set initial parameters  $\theta' = (\Sigma' = \mathbf{I}_d, L_0, \lambda_0)$
- 5: **Main Loop:**
- 6:     **Repeat**
- 7:         Calculate  $E_{\tau|\mathbf{C}}\{\cdot | \mathbf{C}_i, \theta'\}$  according to the compound pdf using Table VI
- 8:         Estimate  $\hat{\Sigma}$  using (51);
- 9:         Estimate  $\hat{L}$  using  $\hat{\Sigma}$  and (52)
- 10:         Estimate  $\hat{\lambda}$  using (53)
- 11:         Calculate stop criterion:  $D \leftarrow \|\hat{\theta} - \theta'\|$
- 12:         Set inputs for next iteration:  $\Sigma', L', \lambda' \leftarrow \hat{\Sigma}, \hat{L}, \hat{\lambda}$
- 13:         **Until**  $D < \epsilon$
- 14:     **Return**  $(\hat{\Sigma}, \hat{L}, \hat{\lambda})$

TABLE V  
POSTERIOR PDF OF  $\tau$  GIVEN MULTILOOK POLARIMETRIC COVARIANCE MATRIX  $\mathbf{C}$

Texture pdf	Posterior pdf	
Symbol	Symbol	$f_{\tau_i \mathbf{C}_i}(\tau_i \mathbf{C}_i)$
$\bar{\gamma}^{-1}(\lambda)$	$\gamma^{-1}(dL + \lambda, L\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}_i) + \lambda - 1)$	$\frac{(L\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}_i) + \lambda - 1)^{dL + \lambda}}{\Gamma(dL + \lambda)} \frac{1}{\tau_i^{dL + \lambda + 1}} \exp\left(-\frac{L\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}_i) + \lambda - 1}{\tau_i}\right)$

TABLE VI  
POSTERIOR EXPECTATION OF  $\tau$  GIVEN MULTILOOK POLARIMETRIC COVARIANCE MATRIX  $\mathbf{C}$

Texture pdf	Compound pdf	Posterior expectation
$f_{\tau}(\tau)$	$f_{\mathbf{C}}(\mathbf{C})$	$E_{\tau_i \mathbf{C}_i}\{\cdot \mathbf{C}_i, \theta'\}$
$\bar{\gamma}^{-1}(\lambda)$	$\mathcal{G}_d^0(\mathbf{\Sigma}, L, \lambda)$	$E_{\tau_i \mathbf{C}_i}\{\tau_i \mathbf{C}_i, \theta'\} = \frac{L'\text{tr}(\mathbf{\Sigma}'^{-1}\mathbf{C}_i) + \lambda' - 1}{dL' + \lambda' - 1}$ $E_{\tau_i \mathbf{C}_i}\{\frac{1}{\tau_i} \mathbf{C}_i, \theta'\} = \frac{dL' + \lambda'}{L\text{tr}(\mathbf{\Sigma}'^{-1}\mathbf{C}_i) + \lambda' - 1}$ $E_{\tau_i \mathbf{C}_i}\{\ln \tau_i \mathbf{C}_i, \theta'\} = -\psi(dL' + \lambda') + \ln(L'\text{tr}(\mathbf{\Sigma}'^{-1}\mathbf{C}_i) + \lambda' - 1)$

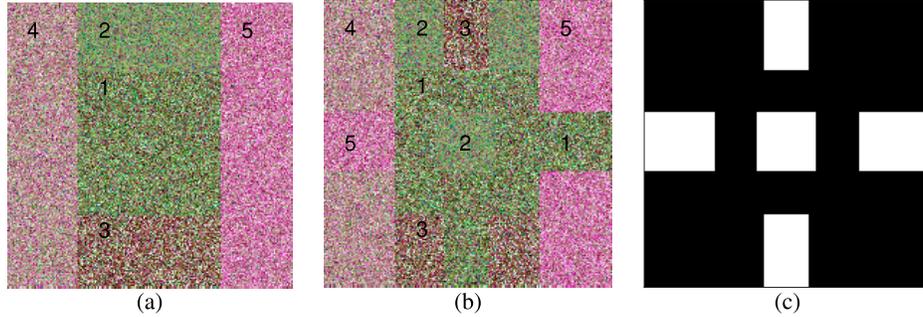


Fig. 2. Pauli decomposition of simulated 4-look quad-pol PolSAR data. (a) Before change, (b) after change, and (c) binary truth change map, where white color is change and black color is no-change.

To produce a binary change detection map a thresholding is applied. Several thresholding methods have been proposed in the literature to determine the threshold in a completely unsupervised manner and none of them is perfect. To name a few of them: the constant false alarm rate (CFAR) algorithm [42], Otsu's method [43], the Kittler and Illingworth (K&I) algorithm [44], and Liu's method [45]. In the present work, the best performance was obtained by applying an optimal thresholding which consists of selecting the nearest point to (0, 1) in the ROC curve [12], [41] as shown in Fig. 7. Based on the optimal threshold, TPR and FPR are obtained. In Section V-A, our approach will be compared against other methods by evaluating both simulated and real multilook PolSAR data.

#### A. Evaluation on Simulated Multilook PolSAR Data

The simulated quad-pol data contain two four-look PolSAR images of  $200 \times 200$  pixels and have three polarimetric channels. These data are each composed of five different regions. All five regions follow  $\mathcal{G}_d^0$  distribution. The texture distribution parameters are chosen to cover homogeneous and heterogeneous textures. For an urban area (area 1) which is a high-texture region and extremely heterogeneous, the shape parameter of the matrix-variate  $\mathcal{G}_d^0$  distribution is less than or equal to 2. For a forest area which is a heterogeneous area with moderate texture, the value of  $\lambda$  is given by  $2 < \lambda < 8$  (areas 3, 4, and 5). For a sea region which is a homogeneous area, the shape parameter tends toward infinity leading to a realization of a scaled Wishart distribution. Here,

TABLE VII  
TEXTURE PARAMETERS AND COVARIANCE MATRIX FOR EACH REGION

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

it corresponds to region 2 with  $\lambda = \infty$ . The details of the distribution parameters ( $\lambda, \mathbf{\Sigma}$ ) for each region are given in Table VII. We want to draw the reader's attention to the fact that regions 1 and 2 have the same speckle covariance matrix but different texture parameter values. This choice is to see if our approach is able to distinguish between the two regions.

Fig. 2 shows the two images before and after change as well as the binary truth change map. For the estimation parameter procedure, the EM-based estimator using a sliding window of size  $k \times k$ , covering the whole image is selected to compute the local estimation of the covariance matrix  $\mathbf{\Sigma}$ , its determinant  $|\mathbf{\Sigma}|$ , the ENL  $L$ , and the texture parameter  $\lambda$ . Fig. 3 depicts the results of the local estimation of these parameters using sliding windows of size  $11 \times 11$ .

To further illustrate the potential of the stochastic distances in change detection for multilook PolSAR data, two well-known test statistics for measuring the equality of two multilook covariance matrices are compared against our approach. Bartlett distance is one of them which has been reported in the literature for classification [46] and change

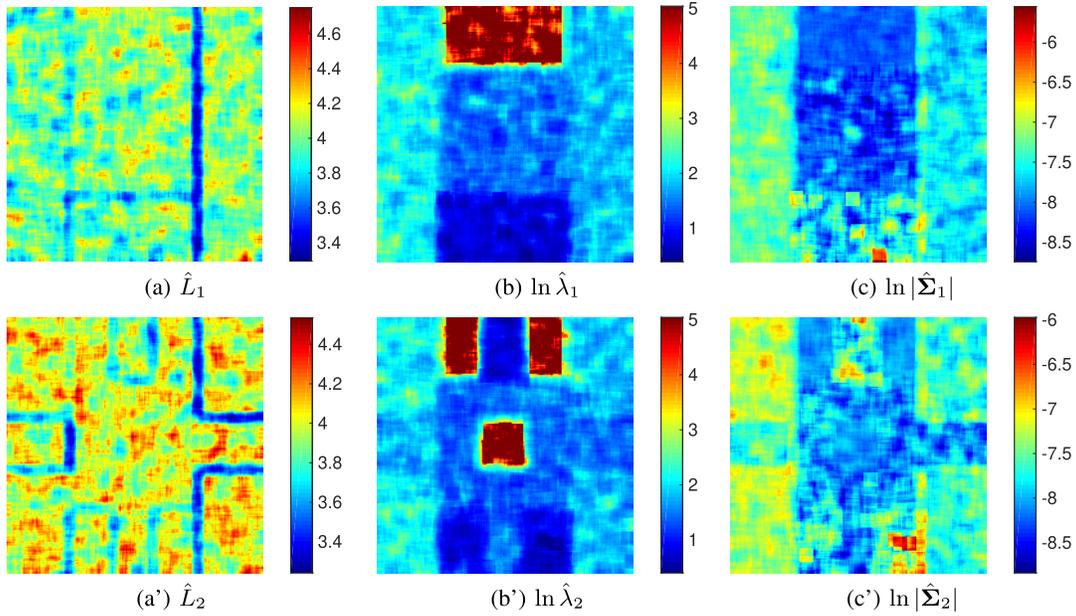


Fig. 3. Local estimation of speckle and texture parameter of whole image before and after change. (a) and (a') Local ENL estimates computed by EM-based estimator when slide window is  $11 \times 11$ . (b) and (b') Local  $\ln \hat{\lambda}$  estimates before and after change. (c) and (c') Local  $\ln |\hat{\Sigma}|$  estimates before and after change.

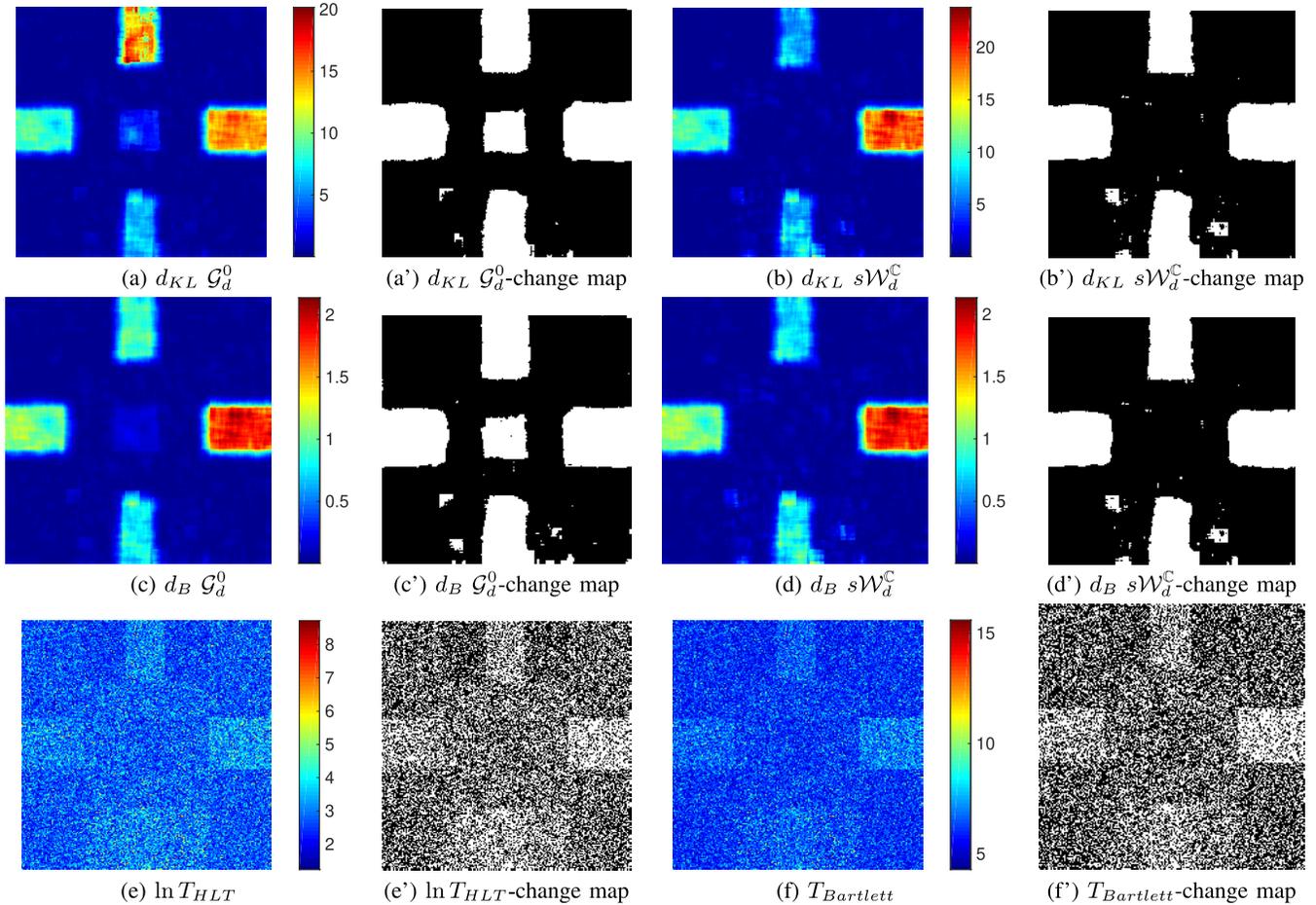


Fig. 4. Detection results for simulated PolSAR data. (a) and (a') Kullback–Leibler distances using  $\mathcal{G}_d^0$  and the binary change map, where white color is change and black color is no-change. (c) and (c') Bhattacharyya distances using  $\mathcal{G}_d^0$  and the corresponding binary change map. (b) and (b') Kullback–Leibler distance using  $s\mathcal{W}_d^C$  and its binary map. (d) and (d') Bhattacharyya distances using  $s\mathcal{W}_d^C$  and binary map. (e) and (e') Logarithm of HLT statistic images and binary change map. (f) and (f') Bartlett distance and the corresponding change detection map.

detection [47] and is given by

$$T_{\text{Bartlett}} = 2 \ln \frac{|\mathbf{Y} + \mathbf{Z}|}{|\mathbf{Y}|^{\frac{1}{2}} |\mathbf{Z}|^{\frac{1}{2}}} \quad (54)$$

where  $\mathbf{Y}$  and  $\mathbf{Z}$  are multilook covariance matrices before and after change. It is worth noticing that the Bartlett distance is proportional to the LRT [7]. The second approach for comparative purposes is the HLT statistic [11] which is used as covariance equality test and is defined as

$$T_{\text{HLT}} = \max\{\text{tr}(\mathbf{Y}^{-1}\mathbf{Z}), \text{tr}(\mathbf{Z}^{-1}\mathbf{Y})\}. \quad (55)$$

The test assumes the scaled complex Wishart distribution for the covariance matrix data.

Fig. 4 illustrates the different distances and the corresponding binary change maps, where white color is change and black color is no-change. These corresponding binary maps are obtained by applying an optimal thresholding which consists of selecting the nearest point to (0, 1) in the ROC curve previously mentioned. In doing so, we obtain the TPR and the FPR. The Bartlett distance and HLT statistic give us the worst detection results as it is shown in Fig. 4(e) and (f) and their corresponding binary images in Fig. 4(e') and (f'), respectively. In Fig. 4(a)–(d), the Kullback–Leibler and the Bhattacharyya distances using  $\mathcal{G}_d^0$  and  $s\mathcal{W}_d^C$  provide better performance than Bartlett distance and HLT statistic as more areas are detected. The reader may refer to Appendixes H and I for the detailed analytical expressions of Kullback–Leibler and Bhattacharyya distances based on the scaled Wishart distribution.

The comparison between distances shows that  $\mathcal{G}_d^0$  distribution gives better results than  $s\mathcal{W}_d^C$ . This can be depicted in region 2 at the center of the image. In this example, the scaled Wishart distribution is unable to detect region 2 [see Fig. 4(b')–(d')], whereas both distances Kullback–Leibler and Bhattacharyya for the  $\mathcal{G}_d^0$  distribution successfully detected it [see Fig. 4(a')–(c')]. This is due to the fact that regions 1 and 2 have the same speckle covariance matrix and as a consequence the distances based on the scaled Wishart distribution consider them to be similar and not different. Furthermore, the distance profile concerning region 2 for both distributions using Bhattacharyya distance is drawn in Fig. 5(a) and (b) corresponding to the columns and rows of the image, respectively. To be more precise, for instance in Fig. 5(a), we fix the row number to 100 and we draw the variation among columns. In a similar way, the variation among rows is shown in Fig. 5(b). One can notice the low values of Bhattacharyya distances based on scaled Wishart distribution compared to the  $\mathcal{G}_d^0$  distribution in region 2.

Change detection in region 2 by the Kullback–Leibler distance is more visible than that of Bhattacharyya. This is due to the different behavior of these two distances with respect to  $\lambda_1$  and  $\lambda_2$ . Indeed, in Fig. 6, the behavior of Kullback–Leibler and Bhattacharyya distances for  $\mathcal{G}_d^0$  distribution was drawn as a function of  $\lambda_1 = \{2, 4\}$  and  $0 < \lambda_2 \leq 200$  in the case corresponding to  $L_1 = L_2 = 4$  and  $\Sigma_1 = \Sigma_2$ . For increasing values of  $\lambda_2$ , the Kullback–Leibler distance increases monotonically, whereas Bhattacharyya distance remains almost constant. In other words, Bhattacharyya is

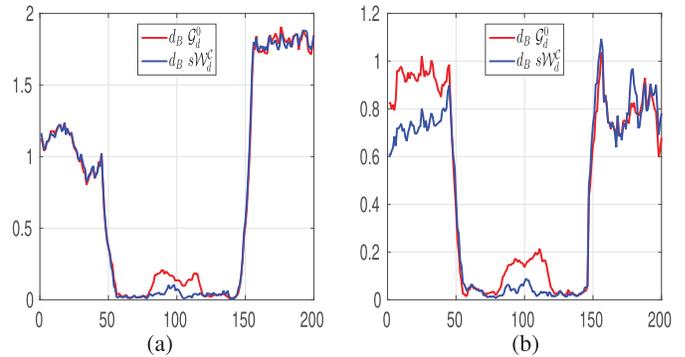


Fig. 5. Distance profile concerning region 2 for  $\mathcal{G}_d^0$  and  $s\mathcal{W}_d^C$  distribution with Bhattacharyya distance. (a) Horizontal profile: row 100. (b) Vertical profile: column 100.

more sensitive to heterogeneous regions and as a consequence, less sensitive to homogeneous regions. In addition, for low value of  $\lambda_1$ , both distances increase (see blue curves in Fig. 6).

Regarding the quantitative assessment, the ROC curves for all the distances used in the experiments are given in Fig. 7. In this example, a sliding window of size  $11 \times 11$  for distances based on  $\mathcal{G}_d^0$  and  $s\mathcal{W}_d^C$  distributions has been employed and the obtained results show that the ROC curve of  $s\mathcal{W}_d^C$  is always above the ROC curve of  $\mathcal{G}_d^0$  for both low and high FPR values. Moreover, HLT and Bartlett show the poorest performance among all considered methods.

As density (or parameter) estimation depends heavily on the sample size, different sliding window sizes are chosen,  $k = \{7, 9, 11, 13, 15\}$ , for evaluation. The AUC is provided for all methods presented in this article in Table VIII. At each window size, the best values are highlighted in red whereas the worst ones are blue. We can clearly see that  $\mathcal{G}_d^0$  always gives the best for any sliding window size. We can also note that as the window size grows, the AUC always increases until a window size of  $11 \times 11$  and decreases afterward. On the contrary, HLT statistic and Bartlett distance are always the worst. The TPR and FPR are also presented in Table VIII based on the selected optimal thresholds. It is worth noticing that TPR and FPR are consistent with AUC. Best performance is characterized by a higher TPR value with a lower FPR value. As it is seen in Table VIII, the Kullback–Leibler distance for  $\mathcal{G}_d^0$  distribution with sliding window  $11 \times 11$  gives us the best performance with (TPR = 0.92, FPR = 0.048) followed by Bhattacharyya distance (TPR = 0.93, FPR = 0.063).

### B. Evaluation on Real PolSAR Data

To evaluate the performance of the stochastic distance based on the matrix-variate  $\mathcal{G}_d^0$  distribution, two PolSAR data sets are used in the experiments.

1) *Data Set*: A co-registered pair of L-band full polarimetric images acquired by the Jet Propulsion Laboratory/National Aeronautics and Space Administration UAVSAR (1.26 GHz) over the city of Los Angeles, California, on April 23, 2009 and May 11, 2015 with a spatial resolution of 0.47 m in azimuth and 1.66 m in range is used. The images are  $1 \times 4$  multilooked. Fig. 8(a) and (b) shows the Pauli decomposition of two images obtained by the JPL's UAVSAR sensor at two different times.

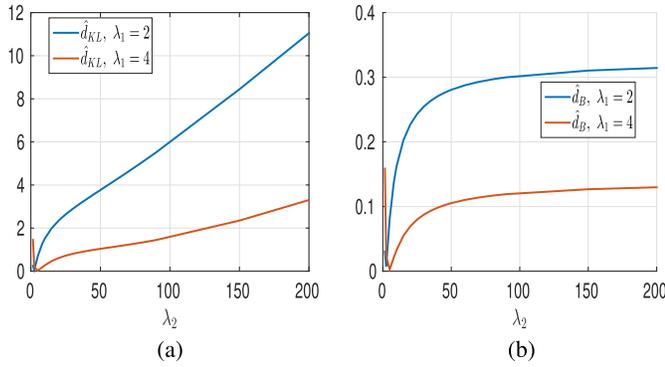


Fig. 6. (a) Behavior of Kullback–Leibler and (b) Bhattacharyya distance for  $\mathcal{G}_d^0$  distribution, when  $L_1 = L_2 = 4$ ,  $\Sigma_1 = \Sigma_2$ ,  $\lambda_1 = \{2, 4\}$  and  $\lambda_2$  varies.

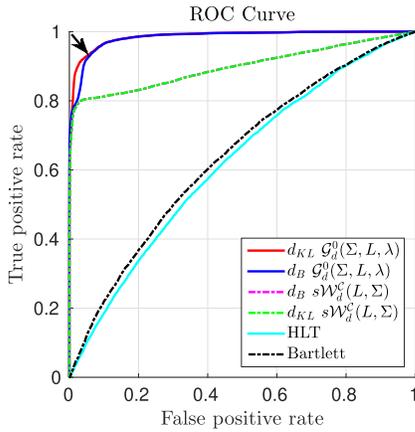


Fig. 7. Comparison of ROC curves between the proposed method and the other considered methods. The selected optimal thresholds correspond to the closest point to (0, 1) in ROC curves. For  $\mathcal{G}_d^0$ , it is pointed by the arrow. The slide window is  $11 \times 11$ .

TABLE VIII

AREA UNDER CURVE (AUC), TPR, AND FALSE ALARM RATE (FPR) FOR SIMULATED DATA AND DIFFERENT VARIOUS METHODS. THE BEST VALUES ARE HIGHLIGHTED IN RED AND THE WORST ONES ARE BLUE

Size	Meth	KLD(%)			Bhattacharyya(%)		
		AUC	TPR	FPR	AUC	TPR	FPR
7	$\mathcal{G}_d^0$	<b>98.05</b>	92.90	6.11	97.89	93.71	6.75
	$s\mathcal{W}_d^C$	<b>89.30</b>	80.56	6.10	89.31	80.73	6.63
9	$\mathcal{G}_d^0$	<b>98.34</b>	92.87	5.02	98.16	93.07	6.40
	$s\mathcal{W}_d^C$	<b>89.48</b>	80.53	4.82	89.49	80.50	4.79
11	$\mathcal{G}_d^0$	<b>98.59</b>	92.83	4.88	98.35	93.64	6.38
	$s\mathcal{W}_d^C$	<b>89.86</b>	80.49	4.59	89.87	80.49	4.55
13	$\mathcal{G}_d^0$	<b>98.50</b>	92.84	4.90	98.30	93.66	6.34
	$s\mathcal{W}_d^C$	<b>90.02</b>	80.52	4.97	90.03	80.66	5.46
15	$\mathcal{G}_d^0$	<b>98.26</b>	92.86	5.22	98.04	92.98	6.54
	$s\mathcal{W}_d^C$	<b>89.90</b>	80.54	5.13	89.91	80.62	5.43

Methods	AUC	TPR	FPR
HLT	61.91	60.23	42.61
Bartlett	63.79	59.09	38.76

Fig. 8(c) shows the ground truth used to compute the ROC curve. We recall that white color corresponds to change and black color to no-change. As we can see, the interested area of this data set is an urban area where the changes occurred due to the effects of urbanization.

TABLE IX  
AREA UNDER CURVE (AUC), TPR, AND FALSE ALARM RATE (FPR) FOR REAL DATA AND DIFFERENT VARIOUS METHODS

Size	Meth	KLD(%)			Bhattacharyya(%)		
		AUC	TPR	FPR	AUC	TPR	FPR
9	$\mathcal{G}_d^0$	97.38	94.06	7.12	97.37	94.05	7.11
	$s\mathcal{W}_d^C$	96.88	93.86	7.95	96.88	93.85	7.94
11	$\mathcal{G}_d^0$	97.90	93.9	6.17	97.94	94.30	6.5
	$s\mathcal{W}_d^C$	96.98	93.24	7.14	97.05	93.02	6.82
13	$\mathcal{G}_d^0$	97.50	93.94	6.87	97.45	94.01	6.71
	$s\mathcal{W}_d^C$	96.44	93.70	7.49	96.51	93.81	7.66

Methods	AUC	TPR	FPR
HLT	85.34	73.59	15.80
Bartlett	84.65	73.66	18.47

2) *Analysis*: We first try to compute the local estimate of the speckle covariance matrix, the ENL, and the texture parameters for both images before and after change. The local ENL estimation images computed by the EM-based estimator using a sliding window of size  $11 \times 11$  covering the whole image are depicted in Fig. 9(a) and (a'). Fig. 9(b) and (b') presents the local texture parameter  $\ln \hat{\lambda}_i$ ,  $i = \{1, 2\}$  estimate. The extremely heterogeneous texture regions (low  $\lambda$  values) are marked in dark blue and the homogeneous regions are marked in dark red. We show in Fig. 9(c) and (c') the local Log-determinant of the covariance matrix of the speckle  $\ln |\hat{\Sigma}_i|$  estimated before and after change. As we can see, the heterogeneous texture regions are marked in red and the homogeneous regions are marked in blue.

The second step is to compute the stochastic Kullback–Leibler and Bhattacharyya distances with  $\mathcal{G}_d^0$ . For convenience of analysis, we compare these distances to that of using the scaled Wishart  $s\mathcal{W}_d^C$  distribution. In addition, the test statistics, HLT and Bartlett, for measuring the equality of two multilook covariance matrices are compared against our approach.

Fig. 10 shows the results relative to the stochastic distances based on  $\mathcal{G}_d^0$  and  $s\mathcal{W}_d^C$  distribution, and the statistical tests: HLT and Bartlett. The corresponding binary change detection maps are also depicted in Fig. 10. It is noted that the distances are important at the location of the changes given by the ground truth. Changes can be seen easily on both distances for both distributions and for both HLT and Bartlett tests. The difference lies in the small changes that are visible for the  $\mathcal{G}_d^0$  distribution and not for the  $s\mathcal{W}_d^C$  distribution. To quantify the performance of the method, ROC curves are plotted for both distributions and distances using the ground truth. Fig. 11 presents the results corresponding to an analysis window of size  $11 \times 11$ . It is shown that the stochastic distances based on  $\mathcal{G}_d^0$  give the best performance followed by that based on scaled Wishart distribution and then by HLT test. The Bartlett test gives the worst detection results. As the method depends on the sample size, different sliding window sizes are used  $k = \{9, 11, 13\}$  for evaluation. The AUC, TPR, and FPR are provided in Table IX for all methods presented in this study and for different sliding window. We recall the reader that the selected optimal thresholds provide TPR and FPR. The best results are given by a window of size  $11 \times 11$ .



Fig. 8. UAVSAR images (in Pauli decomposition) on April 23, 2009 and May 11, 2015. (a) Before change, (b) after change, (c) ground truth, where white color is change and black color is no-change.

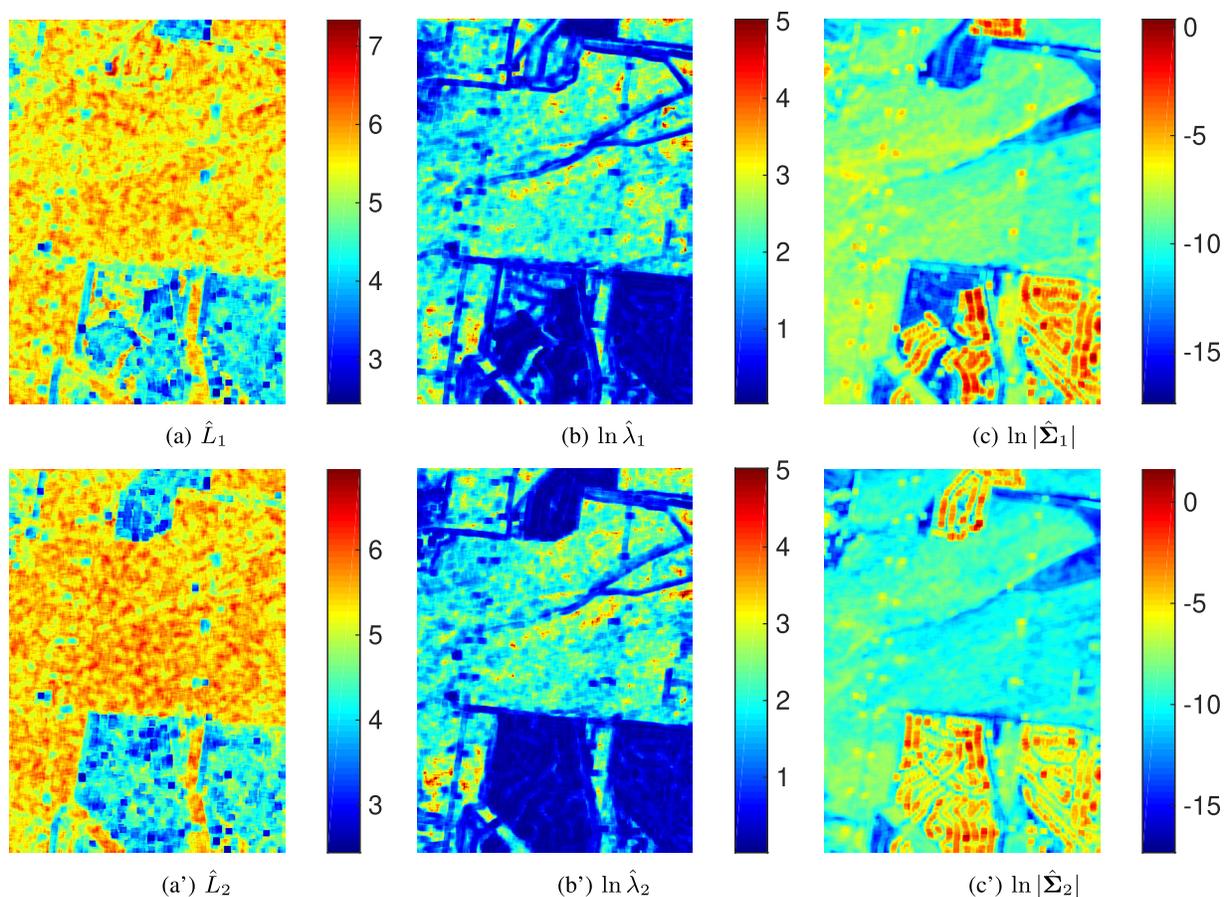


Fig. 9. Local estimation of speckle and texture parameters of the city of Los Angeles images before and after change. (a) and (a') Local ENL estimation. (b) and (b') Local texture parameter  $\ln \lambda$  estimate. (c) and (c') Local log determinant of the covariance matrix of speckle.

## VI. CONCLUSION

This article has developed the closed-form expressions of four stochastic distances that are Kullback–Leibler, Rényi of order  $\beta$ , Bhattacharyya, and Hellinger, between two matrix-variate  $\mathcal{G}_d^0$  distributions representing multilook PolSAR data before and after change. These expressions have been derived

for the most general case (different ENLs, various covariance matrices of the speckle, and different texture parameters) and for the particular case where the ENLs are the same before and after change. All the four analytic expressions depend on the Lauricella function which is easily implemented in the Lauricella series. The effectiveness of the change detection

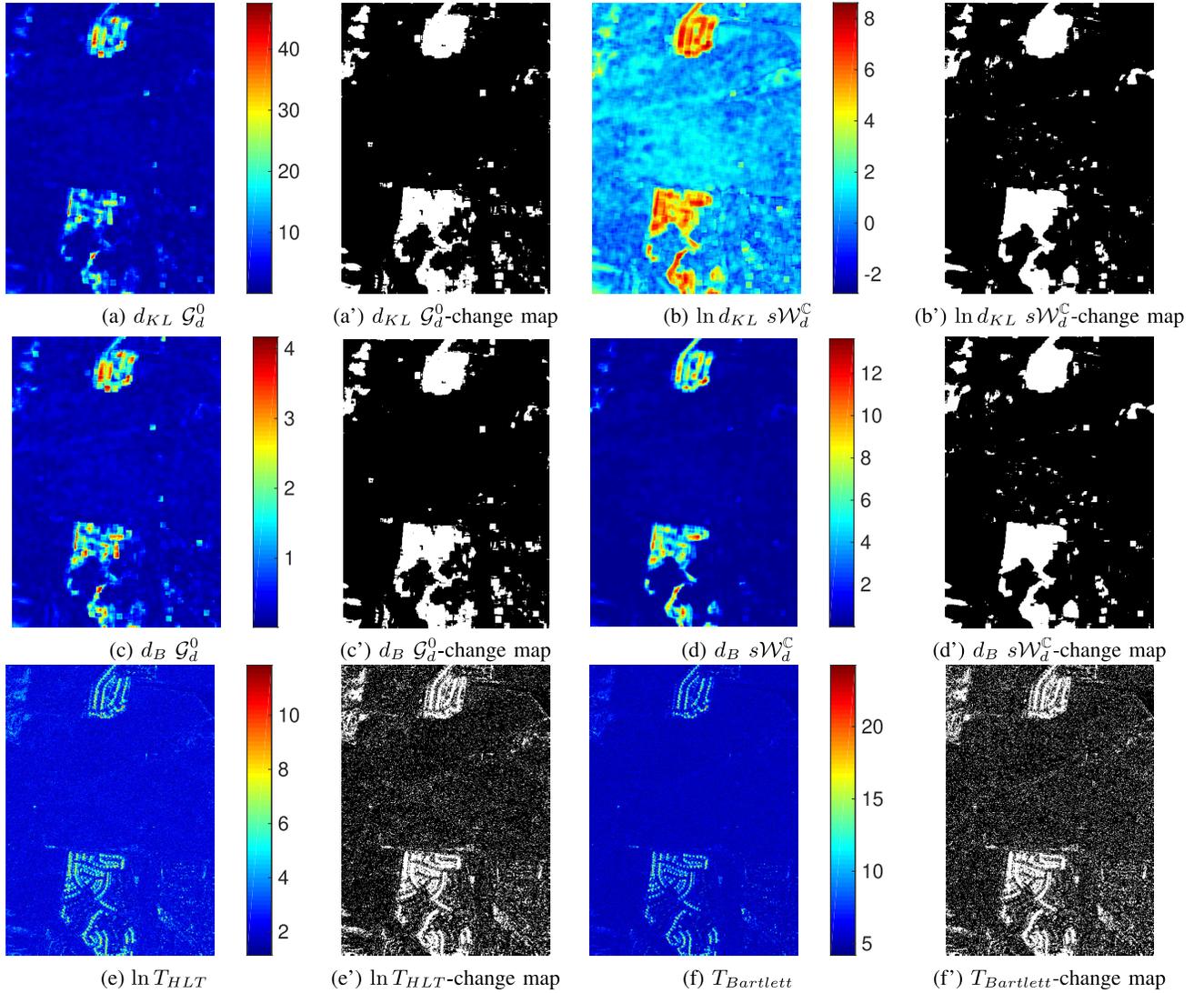


Fig. 10. Detection results for simulated PolSAR data. (a) and (a') Kullback–Leibler distances using  $\mathcal{G}_d^0$  and the binary change map, where white color is change and black color is no-change. (c) and (c') Bhattacharyya distances using  $\mathcal{G}_d^0$  and the corresponding binary change map. (b) and (b') Logarithm of Kullback–Leibler distance using  $s\mathcal{W}_d^C$  and its binary map. (d) and (d') Bhattacharyya distances using  $s\mathcal{W}_d^C$  and binary map. (e) and (e') Logarithm of HLT statistic images and binary change map. (f) and (f') Bartlett distance and the corresponding change detection map.

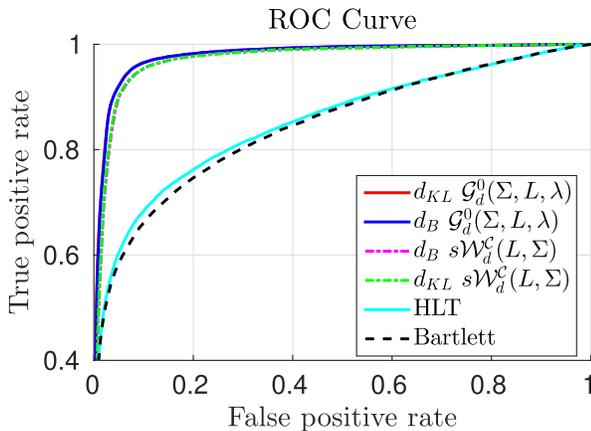


Fig. 11. ROC curve comparison between proposed method with  $\mathcal{G}_d^0$  and scaled Wishart distribution and other methods (HLT, Bartlett distance). The sliding window size is  $11 \times 11$ .

method using the closed-form formula is illustrated by both simulated and real PolSAR data set. The results show that the

proposed method is well suited to detect heterogeneous areas where the texture is dominant and gives a higher performance (higher AUC) than the stochastic distances based on the scaled Wishart distribution and other methods such as the HLT and the Bartlett distance.

#### APPENDIX A

We need to demonstrate that

$$\int_{\Omega_+} |\mathbf{C}|^{L_1-d} (1 + \text{tr}(\mathbf{C}))^{-(dL_1+\lambda_1)} d\mathbf{C} = \frac{\Gamma(\lambda_1)\Gamma_d(L_1)}{\Gamma(dL_1 + \lambda_1)}. \quad (56)$$

The multiplication of the expression of  $\Gamma_d(L_1)$  by  $\Gamma(\lambda_1)$  is given by

$$\Gamma_d(L_1)\Gamma(\lambda_1) = \int_{\Omega_+} |\mathbf{X}|^{L_1-d} e^{-\text{tr}(\mathbf{X})} d\mathbf{X} \int_0^{+\infty} t^{\lambda_1-1} e^{-t} dt \quad (57)$$

$$= \int_{\Omega_+} \int_0^{+\infty} t^{\lambda_1-1} |\mathbf{X}|^{L_1-d} e^{-(\text{tr}(\mathbf{X})+t)} d\mathbf{X} dt. \quad (58)$$

By taking the transformation  $\mathbf{C} = t^{-1}\mathbf{X}$  where the Jacobian determinant of this transformation is given by

$d\mathbf{C} = t^{-d^2} d\mathbf{X}$  [48, Th. 3.5] and by using propriety  $\int_0^{+\infty} t^b e^{-at} dt = \Gamma(b+1)/a^{b+1}$ , (58) is then given as follows:

$$\Gamma_d(L_1)\Gamma(\lambda_1) = \int_{\Omega_+} \int_0^{+\infty} |\mathbf{C}|^{L_1-d} t^{dL_1+\lambda_1-1} e^{-t(\text{tr}(\mathbf{C})+1)} dt d\mathbf{C} \quad (59)$$

$$= \int_{\Omega_+} |\mathbf{C}|^{L_1-d} \left( \int_0^{+\infty} t^{dL_1+\lambda_1-1} e^{-t(\text{tr}(\mathbf{C})+1)} dt \right) d\mathbf{C} \quad (60)$$

$$= \Gamma(dL_1 + \lambda_1) \int_{\Omega_+} |\mathbf{C}|^{L_1-d} (1 + \text{tr}(\mathbf{C}))^{-(dL_1+\lambda_1)} d\mathbf{C}. \quad (61)$$

Finally, (56) is established.

## APPENDIX B

### KULLBACK–LEIBLER DISTANCE FOR $\mathcal{G}_d^0$ DISTRIBUTION

#### A. Expression of $E_{\mathbf{C}^1}\{\ln|\mathbf{C}|\}$

Expectation  $E_{\mathbf{C}^1}\{\ln|\mathbf{C}|\}$  can be computed utilizing the following propriety  $\int \log(x)f(x)dx = (\partial/\partial a) \int x^a f(x)dx|_{a=0}$ . The expectation is given as follows:

$$E_{\mathbf{C}^1}\{\ln|\mathbf{C}|\} = \frac{\partial}{\partial a} \int_{\Omega_+} |\mathbf{C}|^a f_{\mathbf{C}^1}(\mathbf{C}) d\mathbf{C} \Big|_{a=0} \quad (62)$$

$$= \frac{\Gamma(dL_1 + \lambda_1)(\lambda_1 - 1)^{\lambda_1}}{\Gamma(\lambda_1)\Gamma_d(L_1)} |L_1 \Sigma_1^{-1}|^{L_1} \frac{\partial}{\partial a} \int_{\Omega_+} |\mathbf{C}|^{L_1+a-d} \times [\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1]^{-(dL_1+\lambda_1)} d\mathbf{C} \Big|_{a=0} \quad (63)$$

$$= \frac{\Gamma(dL_1 + \lambda_1)}{\Gamma(\lambda_1)\Gamma_d(L_1)(\lambda_1 - 1)^{dL_1}} |L_1 \Sigma_1^{-1}|^{L_1} \frac{\partial}{\partial a} \int_{\Omega_+} |\mathbf{C}|^{L_1+a-d} \times \left[ \text{tr} \left( \frac{L_1}{\lambda_1 - 1} \Sigma_1^{-1} \mathbf{C} \right) + 1 \right]^{-(dL_1+\lambda_1)} d\mathbf{C} \Big|_{a=0}. \quad (64)$$

Let  $\mathbf{Y} = L_1/(\lambda_1 - 1)\Sigma_1^{-1}\mathbf{C}$  be a transformation where the Jacobian determinant is given by  $d\mathbf{Y} = |L_1/(\lambda_1 - 1)\Sigma_1^{-1}|^d d\mathbf{C}$  [48, Th. 3.5]. The new expression of the expectation is then given as follows:

$$E_{\mathbf{C}^1}\{\ln|\mathbf{C}|\} = \frac{\Gamma(dL_1 + \lambda_1)}{\Gamma(\lambda_1)\Gamma_d(L_1)} \frac{\partial}{\partial a} \left( \frac{(\lambda_1 - 1)^{ad}}{|L_1 \Sigma_1^{-1}|^a} \times \int_{\Omega_+} |\mathbf{Y}|^{L_1+a-d} [1 + \text{tr}(\mathbf{Y})]^{-(dL_1+\lambda_1)} d\mathbf{Y} \right) \Big|_{a=0}. \quad (65)$$

Using relation (56), the following relation can be established:

$$\int_{\Omega_+} |\mathbf{Y}|^{L_1+a-d} [1 + \text{tr}(\mathbf{Y})]^{-(dL_1+\lambda_1)} d\mathbf{Y} = \frac{\Gamma(\lambda_1 - ad)\Gamma_d(L_1 + a)}{\Gamma(dL_1 + \lambda_1)}. \quad (66)$$

The last equation is substituted in the expression of  $E_{\mathbf{C}^1}\{\ln|\mathbf{C}|\}$  and the result is given by the following expression:

$$E_{\mathbf{C}^1}\{\ln|\mathbf{C}|\} \quad (67)$$

$$= \frac{\partial}{\partial a} \left( \frac{\Gamma(\lambda_1 - ad)}{\Gamma(\lambda_1)} \frac{\Gamma_d(L_1 + a)}{\Gamma_d(L_1)} |L_1 \Sigma_1^{-1}|^{-a} (\lambda_1 - 1)^{ad} \right) \Big|_{a=0} = \psi_d(L_1) + \ln|\Sigma_1| - d \ln L_1 + d [\ln(\lambda_1 - 1) - \psi(\lambda_1)]. \quad (68)$$

#### B. Expression of $E_{\mathbf{C}^1}\{\ln[\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1]\}$

The expression of the expectation is computed as follows:

$$E_{\mathbf{C}^1}\{\ln[\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1]\} = \frac{\partial}{\partial a} \int_{\Omega_+} (\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1)^a f_{\mathbf{C}^1}(\mathbf{C}) d\mathbf{C} \Big|_{a=0} \quad (69)$$

$$= \frac{\Gamma(dL_1 + \lambda_1)(\lambda_1 - 1)^{\lambda_1}}{\Gamma(\lambda_1)\Gamma_d(L_1)} |L_1 \Sigma_1^{-1}|^{L_1} \frac{\partial}{\partial a} \int_{\Omega_+} |\mathbf{C}|^{L_1-d} \times (\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1)^{-(dL_1+\lambda_1-a)} d\mathbf{C} \Big|_{a=0} \quad (70)$$

$$= \frac{\Gamma(dL_1 + \lambda_1) |L_1 \Sigma_1^{-1}|^{L_1}}{\Gamma(\lambda_1)\Gamma_d(L_1)} \frac{\partial}{\partial a} \left( \frac{1}{(\lambda_1 - 1)^{dL_1-a}} \int_{\Omega_+} |\mathbf{C}|^{L_1-d} \times \left( \text{tr} \left( \frac{L_1}{\lambda_1 - 1} \Sigma_1^{-1} \mathbf{C} \right) + 1 \right)^{-(dL_1+\lambda_1-a)} d\mathbf{C} \right) \Big|_{a=0}. \quad (71)$$

In the same way, as done in the previous section, transformation  $\mathbf{Y} = L_1/(\lambda_1 - 1)\Sigma_1^{-1}\mathbf{C}$  is used. The expression of the expectation is then given as follows:

$$E_{\mathbf{C}^1}\{\ln[\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1]\} = \frac{\Gamma(dL_1 + \lambda_1)}{\Gamma(\lambda_1)\Gamma_d(L_1)} \times \frac{\partial}{\partial a} \left( (\lambda_1 - 1)^a \int_{\Omega_+} |\mathbf{Y}|^{L_1-d} (\text{tr}(\mathbf{Y}) + 1)^{-(dL_1+\lambda_1-a)} d\mathbf{Y} \right) \Big|_{a=0}. \quad (72)$$

Using relation (56), the following equation is established:

$$\int_{\Omega_+} |\mathbf{Y}|^{L_1-d} (\text{tr}(\mathbf{Y}) + 1)^{-(dL_1+\lambda_1-a)} d\mathbf{Y} = \frac{\Gamma(\lambda_1 - a)\Gamma_d(L_1)}{\Gamma(dL_1 + \lambda_1 - a)}. \quad (73)$$

As a consequence, the new expression of the expectation is given by

$$E_{\mathbf{C}^1}\{\ln[\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1]\} = \frac{\partial}{\partial a} \left( \frac{\Gamma(\lambda_1 - a)}{\Gamma(\lambda_1)} \frac{\Gamma(dL_1 + \lambda_1)}{\Gamma(dL_1 + \lambda_1 - a)} (\lambda_1 - 1)^a \right) \Big|_{a=0} \quad (74)$$

$$= \log(\lambda_1 - 1) - \psi(\lambda_1) + \psi(dL_1 + \lambda_1). \quad (75)$$

#### C. Expression of $E_{\mathbf{C}^1}\{\ln[\text{tr}(L_2 \Sigma_2^{-1} \mathbf{C}) + \lambda_2 - 1]\}$

The expression of the expectation is computed as follows:

$$E_{\mathbf{C}^1}\{\ln[\text{tr}(L_2 \Sigma_2^{-1} \mathbf{C}) + \lambda_2 - 1]\} = -A \frac{\partial}{\partial a} \int_{\Omega_+} |\mathbf{C}|^{L_1-d} [\text{tr}(L_1 \Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1]^{-(dL_1+\lambda_1)} \times [\text{tr}(L_2 \Sigma_2^{-1} \mathbf{C}) + \lambda_2 - 1]^{-a} d\mathbf{C} \Big|_{a=0} \quad (76)$$

with  $A = [\Gamma(dL_1 + \lambda_1)/(\Gamma(\lambda_1)\Gamma_d(L_1))](\lambda_1 - 1)^{\lambda_1} |L_1 \Sigma_1^{-1}|^{L_1}$ . Let take  $\Sigma_1^{\prime-1} = L_1/(\lambda_1 - 1)\Sigma_1^{-1}$  a positive definite complex Hermitian matrix and define the transformation  $\mathbf{Y} = \Sigma_1^{\prime-1}\mathbf{C}$ . The Jacobian determinant of this transformation is given by  $d\mathbf{Y} = |\Sigma_1^{\prime-1}|^d d\mathbf{C}$ . The previous expectation is then given as follows:

$$E_{\mathbf{C}^1}\{\ln[\text{tr}(L_2 \Sigma_2^{-1} \mathbf{C}) + \lambda_2 - 1]\} = -A(\lambda_1 - 1)^{-(dL_1+\lambda_1)} |\Sigma_1^{\prime-1}|^{L_1}$$

$$\begin{aligned} & \times \frac{\partial}{\partial a} \left( (\lambda_2 - 1)^{-a} \int_{\Omega_+} |\mathbf{Y}|^{L_1-d} [\text{tr}(\mathbf{Y}) + 1]^{-(dL_1+\lambda_1)} \right. \\ & \left. \times [\text{tr}(\mathbf{A}\mathbf{Y}) + 1]^{-a} d\mathbf{Y} \right) \Big|_{a=0} \end{aligned} \quad (77)$$

with  $\mathbf{A} = L_2(\lambda_1 - 1)/(L_1(\lambda_2 - 1))\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1$ .

Substituting expressions of  $A$  and  $\mathbf{\Sigma}'_1$  in the last equation and using relation (56), we get

$$\begin{aligned} & = -\frac{\Gamma(dL_1 + \lambda_1)}{\Gamma(\lambda_1)\Gamma_d(L_1)} \frac{\partial}{\partial a} \left( (\lambda_2 - 1)^{-a} \int_{\Omega_+} |\mathbf{Y}|^{L_1-d} \right. \\ & \left. \times [\text{tr}(\mathbf{Y}) + 1]^{-(dL_1+\lambda_1)} [\text{tr}(\mathbf{A}\mathbf{Y}) + 1]^{-a} d\mathbf{Y} \right) \Big|_{a=0} \end{aligned} \quad (78)$$

$$\begin{aligned} & = \ln(\lambda_2 - 1) - \frac{\Gamma(dL_1 + \lambda_1)}{\Gamma(\lambda_1)\Gamma_d(L_1)} \frac{\partial}{\partial a} \\ & \times \left( \int_{\Omega_+} |\mathbf{Y}|^{L_1-d} [1 + \text{tr}(\mathbf{Y})]^{-(dL_1+\lambda_1)} [1 + \text{tr}(\mathbf{A}\mathbf{Y})]^{-a} d\mathbf{Y} \right) \Big|_{a=0} \end{aligned} \quad (79)$$

Knowing that

$$\begin{aligned} (1 + \text{tr}(\mathbf{Y}))^{-(dL_1+\lambda_1)} & = \frac{1}{\Gamma(dL_1 + \lambda_1)} \\ & \int_0^{+\infty} t^{dL_1+\lambda_1-1} e^{-t(1+\text{tr}(\mathbf{Y}))} dt \end{aligned} \quad (80)$$

$$(1 + \text{tr}(\mathbf{A}\mathbf{Y}))^{-a} = \frac{1}{\Gamma(a)} \int_0^{+\infty} x^{a-1} e^{-x(1+\text{tr}(\mathbf{A}\mathbf{Y}))} dx \quad (81)$$

we can write

$$\begin{aligned} & \int_{\Omega_+} |\mathbf{Y}|^{L_1-d} [\text{tr}(\mathbf{Y}) + 1]^{-(dL_1+\lambda_1)} [\text{tr}(\mathbf{A}\mathbf{Y}) + 1]^{-a} d\mathbf{Y} \\ & = \frac{1}{\Gamma(dL_1 + \lambda_1)\Gamma(a)} \int_{\Omega_+} |\mathbf{Y}|^{L_1-d} \int_0^{+\infty} \int_0^{+\infty} t^{dL_1+\lambda_1-1} e^{-t} e^{-\text{tr}(t\mathbf{Y})} \\ & \times x^{a-1} e^{-x} e^{-\text{tr}(x\mathbf{A}\mathbf{Y})} dx dt d\mathbf{Y}. \end{aligned} \quad (82)$$

Using transformation  $\mathbf{Z} = t\mathbf{Y}$  where its Jacobian determinant is given by  $d\mathbf{Z} = t^{d^2} d\mathbf{Y}$ , the last equation is given by

$$\begin{aligned} & = \frac{1}{\Gamma(dL_1 + \lambda_1)\Gamma(a)} \int_0^{+\infty} t^{\lambda_1-1} e^{-t} x^{a-1} e^{-x} \\ & \times \left( \int_{\Omega_+} |\mathbf{Z}|^{L_1-d} e^{-\text{tr}(\mathbf{Z}[\mathbf{I}_d + \frac{x}{t}\mathbf{A}])} d\mathbf{Z} \right) dx dt. \end{aligned} \quad (83)$$

Utilizing relation  $\int_{\Omega_+} |\mathbf{X}|^{\alpha-d} e^{-\text{tr}(\mathbf{B}\mathbf{X})} d\mathbf{X} = \Gamma_d(\alpha) |\mathbf{B}|^{-\alpha}$  [48] where  $\mathbf{B}$  is a positive define complex Hermitian matrix, the previous equation is then given by

$$\begin{aligned} & = \frac{\Gamma_d(L_1)}{\Gamma(dL_1 + \lambda_1)\Gamma(a)} \int_0^{+\infty} t^{\lambda_1-1} e^{-t} \left| \mathbf{I}_d + \frac{x}{t} \mathbf{A} \right|^{-L_1} \\ & x^{a-1} e^{-x} dx dt. \end{aligned} \quad (84)$$

Let  $\Lambda_1, \dots, \Lambda_d$  be the eigenvalues of  $\mathbf{A}$ . Then, the determinant is given as follows:

$$\left| \mathbf{I}_d + \frac{x}{t} \mathbf{A} \right| = \prod_{i=1}^d \left( 1 + \frac{x}{t} \Lambda_i \right). \quad (85)$$

As a consequence, (84) is given by

$$\begin{aligned} & = \frac{\Gamma_d(L_1)}{\Gamma(dL_1 + \lambda_1)\Gamma(a)} \int_0^{+\infty} \int_0^{+\infty} t^{\lambda_1-1} e^{-(t+x)} \\ & \times \prod_{i=1}^d \left( 1 + \frac{x}{t} \Lambda_i \right)^{-L_1} x^{a-1} dx dt. \end{aligned} \quad (86)$$

Using the following change variables  $u = x/t$  and  $v = x + t$ , the Jacobian determinant is defined by  $dudv = (1+u)^2/v dx dt$ . Then, one can write

$$\begin{aligned} & = \frac{\Gamma_d(L_1)\Gamma(\lambda_1 + a)}{\Gamma(dL_1 + \lambda_1)\Gamma(a)} \int_0^{+\infty} u^{a-1} (1+u)^{-(\lambda_1+a)} \\ & \times \prod_{i=1}^d (1 + \Lambda_i u)^{-L_1} du \end{aligned} \quad (87)$$

Using the following change variable  $t = 1/(1+u)$ , the last equation is given by

$$\begin{aligned} & = \frac{\Gamma_d(L_1)\Gamma(\lambda_1 + a)}{\Gamma(dL_1 + \lambda_1)\Gamma(a)} \left( \prod_{i=1}^d \Lambda_i^{-L_1} \right) \int_0^1 t^{dL_1+\lambda_1-1} (1-t)^{a-1} \\ & \times \prod_{i=1}^d \left( 1 - \frac{\Lambda_i - 1}{\Lambda_i} t \right)^{-L_1} dt. \end{aligned} \quad (88)$$

The last integral represents the Lauricella D-hypergeometric function, denoted  $F_D^{(n)}(\cdot)$ , which is an extension of the Appell function to  $n$  variables where  $n > 2$ . The integral representation of  $F_D^{(n)}(\cdot)$  is given by (115) in Appendix G. Consequently, (88) is given by

$$\begin{aligned} & \int_{\Omega_+} |\mathbf{Y}|^{L_1-d} [\text{tr}(\mathbf{Y}) + 1]^{-(dL_1+\lambda_1)} [\text{tr}(\mathbf{A}\mathbf{Y}) + 1]^{-a} d\mathbf{Y} \\ & = \frac{\Gamma_d(L_1)\Gamma(\lambda_1 + a)}{\Gamma(\lambda_1 + dL_1 + a)} \left( \prod_{i=1}^d \Lambda_i^{-L_1} \right) \\ & \times F_D^{(d)} \left( \lambda_1 + dL_1, \underbrace{L_1, \dots, L_1}_d; a + \lambda_1 + dL_1; \right. \\ & \left. 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right). \end{aligned} \quad (89)$$

Then, expectation  $E_{\mathbf{C}^1} \{ \ln[\text{tr}(L_2\mathbf{\Sigma}_2^{-1}\mathbf{C}) + \lambda_2 - 1] \}$  is given by

$$\begin{aligned} & E_{\mathbf{C}^1} \left\{ \ln \left[ \text{tr} \left( L_2 \mathbf{\Sigma}_2^{-1} \mathbf{C} \right) + \lambda_2 - 1 \right] \right\} \\ & = \ln(\lambda_2 - 1) - \frac{\Gamma(\lambda_1 + dL_1)}{\Gamma(\lambda_1)} \left( \prod_{i=1}^d \Lambda_i^{-L_1} \right) \\ & \times \frac{\partial}{\partial a} \left( \frac{\Gamma(\lambda_1 + a)}{\Gamma(\lambda_1 + dL_1 + a)} F_D^{(d)} \left( \lambda_1 + dL_1, L_1, \dots, L_1; \right. \right. \\ & \left. \left. a + \lambda_1 + dL_1; 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \right) \Big|_{a=0} \\ & = \ln(\lambda_2 - 1) \\ & - \left[ \left( \psi(\lambda_1) - \psi(\lambda_1 + dL_1) \right) F_D^{(d)} \left( \lambda_1 + dL_1, L_1, \dots, L_1; \right. \right. \\ & \left. \left. \lambda_1 + dL_1 + a; 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \Big|_{a=0} \right. \\ & \left. + \frac{\partial}{\partial a} F_D^{(d)} \left( \lambda_1 + dL_1, L_1, \dots, L_1; \lambda_1 + dL_1 + a; \right. \right. \\ & \left. \left. 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1} \right) \Big|_{a=0} \right] \prod_{i=1}^d \Lambda_i^{-L_1}. \end{aligned} \quad (90)$$

Lauricella gave several transformation formulas, of which we use the relation (116)

$$\begin{aligned}
& F_D^{(d)}\left(\lambda_1 + dL_1, L_1, \dots, L_1; \lambda_1 + dL_1 + a; \right. \\
& \quad \left. 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1}\right)\Big|_{a=0} \\
&= \prod_{i=1}^d \Lambda_i^{L_i} F_D^{(d)}(a, L_1, \dots, L_1; \lambda_1 + dL_1 + a; \\
& \quad 1 - \Lambda_1, \dots, 1 - \Lambda_d)\Big|_{a=0} \\
&= \prod_{i=1}^d \Lambda_i^{L_i} \lim_{a \rightarrow 0} F_D^{(d)}(a, L_1, \dots, L_1; \lambda_1 + dL_1 + a; \\
& \quad 1 - \Lambda_1, \dots, 1 - \Lambda_d) \\
&= \prod_{i=1}^d \Lambda_i^{L_i}. \tag{91}
\end{aligned}$$

In fact,  $\lim_{a \rightarrow 0} F_D^{(d)}(a, L_1, \dots, L_1; \lambda_1 + dL_1 + a; 1 - \Lambda_1, \dots, 1 - \Lambda_d) = 1$  is deduced using the definition of the Lauricella series given by (113). Subsequently, the expectation is given by

$$\begin{aligned}
& E_{\mathbf{C}^1} \left\{ \ln \left[ \text{tr} \left( L_2 \Sigma_2^{-1} \mathbf{C} \right) + \lambda_2 - 1 \right] \right\} \\
&= \ln(\lambda_2 - 1) - \psi(\lambda_1) + \psi(\lambda_1 + dL_1) - \left( \prod_{i=1}^d \Lambda_i^{-L_i} \right) \\
& \quad \times \frac{\partial}{\partial a} F_D^{(d)}(\lambda_1 + dL_1, L_1, \dots, L_1; \lambda_1 + dL_1 + a; \\
& \quad 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1}) \Big|_{a=0}. \tag{92}
\end{aligned}$$

In the following we present some particular cases:

*Case:  $\mathbf{\Lambda} = \mathbf{I}_d$ :* This case corresponds to the equality between parameters  $\Sigma_1 = \Sigma_2$ ,  $L_1 = L_2$  and  $\lambda_1 = \lambda_2$ . Expectation  $E_{\mathbf{C}^1} \{ \ln[\text{tr}(L_2 \Sigma_2^{-1} \mathbf{C}) + \lambda_2 - 1] \}$  is given by  $\ln(\lambda_2 - 1) - \psi(\lambda_1) + \psi(\lambda_1 + dL_1)$ .

*Case  $d=1$ :* For mono-polarization where  $d = 1$ , expectation  $E_{\mathbf{C}^1} \{ \ln[L_2 I + \lambda_2 - 1] \}$  where  $I$  represents one of the following channels {HH, HV, VV} depends on Gauss hypergeometric function  ${}_2F_1(\cdot)$  and is given as follows:

$$\begin{aligned}
& E_{\mathbf{C}^1} \{ \ln[L_2 I + \lambda_2 - 1] \} \\
&= \ln(\lambda_2 - 1) - \psi(\lambda_1) + \psi(L_1 + \lambda_1) \\
& \quad - \Lambda^{-L_1} \frac{\partial}{\partial a} {}_2F_1(L_1, L_1 + \lambda_1; L_1 + \lambda_1 + a; 1 - \Lambda^{-1}) \Big|_{a=0} \tag{93}
\end{aligned}$$

with  $\Lambda = L_2(\lambda_1 - 1)/(L_1(\lambda_2 - 1))$ .

*Case  $d=2$ :* For bi-polarization  $d = 2$ , expression of  $E_{\mathbf{C}^1} \{ \ln[\text{tr}(L_2 \Sigma_2^{-1} \mathbf{C}) + \lambda_2 - 1] \}$  depends on the Appell hypergeometric function  $F_1(\cdot)$  [49] and is given as follows:

$$\begin{aligned}
& E_{\mathbf{C}^1} \left\{ \ln \left[ \text{tr} \left( L_2 \Sigma_2^{-1} \mathbf{C} \right) + \lambda_2 - 1 \right] \right\} \\
&= \ln(\lambda_2 - 1) - \psi(\lambda_1) + \psi(\lambda_1 + 2L_1) - \left( \prod_{i=1}^2 \Lambda_i^{-L_i} \right) \\
& \quad \times \frac{\partial}{\partial a} F_1\left(\lambda_1 + 2L_1, L_1, L_1; \lambda_1 + 2L_1 + a; \right. \\
& \quad \left. 1 - \Lambda_1^{-1}, 1 - \Lambda_2^{-1}\right) \Big|_{a=0}. \tag{94}
\end{aligned}$$

## APPENDIX C

### RÉNYI DISTANCE OF ORDER $\beta$ FOR $\mathcal{G}_d^0$

The integral over all positive definite complex Hermitian matrix  $\mathbf{C}$  is given as follows:

$$\begin{aligned}
& \int_{\Omega_+} f_{\mathbf{C}^1}^\beta(\mathbf{C}) f_{\mathbf{C}^2}^{1-\beta}(\mathbf{C}) d\mathbf{C} \\
&= A_1^\beta A_2^{1-\beta} \int_{\Omega_+} |\mathbf{C}|^{\beta(L_1-d)+(1-\beta)(L_2-d)} \\
& \quad \times \left[ L_1 \text{tr}(\Sigma_1^{-1} \mathbf{C}) + \lambda_1 - 1 \right]^{-\alpha_1} \left[ L_2 \text{tr}(\Sigma_2^{-1} \mathbf{C}) + \lambda_2 - 1 \right]^{-\alpha_2} d\mathbf{C} \tag{95}
\end{aligned}$$

with  $A_i = (L_i^{dL_i} \Gamma(dL_i + \lambda_i) (\lambda_i - 1)^{\lambda_i}) / (\Gamma_d(L_i) |\Sigma_i|^{L_i} \Gamma(\lambda_i))$  where  $i = \{1, 2\}$ ,  $\alpha_1 = \beta(dL_1 + \lambda_1)$  and  $\alpha_2 = (1 - \beta)(dL_2 + \lambda_2)$ .

Let  $\mathbf{Y} = L_1/(\lambda_1 - 1) \Sigma_1^{-1} \mathbf{C}$  be a transformation where the Jacobian determinant is given by  $d\mathbf{Y} = |(L_1/(\lambda_1 - 1)) \Sigma_1^{-1}|^d d\mathbf{C}$ . The new expression of the integral is then given as follows:

$$\begin{aligned}
& \int_{\Omega_+} f_{\mathbf{C}^1}^\beta(\mathbf{C}) f_{\mathbf{C}^2}^{1-\beta}(\mathbf{C}) d\mathbf{C} \\
&= A_1^\beta A_2^{1-\beta} (\lambda_1 - 1)^{-\alpha_1} (\lambda_2 - 1)^{-\alpha_2} \times \left| \frac{\lambda_1 - 1}{L_1} \Sigma_1 \right|^{\beta(L_1 - L_2) + L_2} \\
& \quad \int_{\Omega_+} |\mathbf{Y}|^{\beta(L_1 - L_2) + L_2 - d} [1 + \text{tr}(\mathbf{Y})]^{-\alpha_1} \\
& \quad \times [1 + \text{tr}(\mathbf{\Lambda} \mathbf{Y})]^{-\alpha_2} d\mathbf{Y}. \tag{96}
\end{aligned}$$

Similarly as done in Appendix B-C, the integral over all positive definite complex Hermitian matrix  $\mathbf{Y}$  is given as follows:

$$\begin{aligned}
& \int_{\Omega_+} |\mathbf{Y}|^{\beta(L_1 - L_2) + L_2 - d} [1 + \text{tr}(\mathbf{Y})]^{-\alpha_1} [1 + \text{tr}(\mathbf{\Lambda} \mathbf{Y})]^{-\alpha_2} d\mathbf{Y} \\
&= \frac{\Gamma_d(\beta(L_1 - L_2) + L_2) \Gamma(\beta(\lambda_1 - \lambda_2) + \lambda_2)}{\Gamma(\alpha_1 + \alpha_2)} \\
& \quad \times \prod_{i=1}^d \Lambda_i^{-\beta(L_1 - L_2) - L_2} \\
& \quad \times F_D^{(d)}\left(\alpha_1, \beta(L_1 - L_2) + L_2, \dots, \beta(L_1 - L_2) + L_2; \alpha_1 + \alpha_2; \right. \\
& \quad \left. 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) \tag{97}
\end{aligned}$$

where  $\Lambda_1, \dots, \Lambda_d$  are the eigenvalues of matrix  $\mathbf{\Lambda}$ . Substituting the expressions of  $A_1$ ,  $A_2$ , and (97) in (96), the integral of  $f_{\mathbf{C}^1}^\beta(\mathbf{C}) f_{\mathbf{C}^2}^{1-\beta}(\mathbf{C})$  over all positive definite complex Hermitian matrix  $\mathbf{C}$  is then given by

$$\begin{aligned}
& \int_{\Omega_+} f_{\mathbf{C}^1}^\beta(\mathbf{C}) f_{\mathbf{C}^2}^{1-\beta}(\mathbf{C}) d\mathbf{C} \\
&= \left( \frac{\Gamma(dL_1 + \lambda_1)}{\Gamma_d(L_1) \Gamma(\lambda_1)} \right)^\beta \left( \frac{\Gamma(dL_2 + \lambda_2)}{\Gamma_d(L_2) \Gamma(\lambda_2)} \right)^{1-\beta} \prod_{i=1}^d \Lambda_i^{-\beta L_i} \\
& \quad \times \frac{\Gamma_d(\beta(L_1 - L_2) + L_2) \Gamma(\beta(\lambda_1 - \lambda_2) + \lambda_2)}{\Gamma(\alpha_1 + \alpha_2)} \\
& \quad \times F_D^{(d)}\left(\alpha_1, \beta(L_1 - L_2) + L_2, \dots, \beta(L_1 - L_2) + L_2; \right. \\
& \quad \left. \alpha_1 + \alpha_2; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right). \tag{98}
\end{aligned}$$

APPENDIX D  
HELLINGER DISTANCE FOR  $\mathcal{G}_d^0$

The integral over all positive definite complex Hermitian matrix  $\mathbf{C}$  of the square root of the product of  $f_{\mathbf{C}^1}(\mathbf{C})$  and  $f_{\mathbf{C}^2}(\mathbf{C})$  is given by

$$\begin{aligned} & \int_{\Omega_+} \sqrt{f_{\mathbf{C}^1}(\mathbf{C})f_{\mathbf{C}^2}(\mathbf{C})}d\mathbf{C} \\ &= A_1^{1/2}A_2^{1/2}(\lambda_1 - 1)^{-\alpha_1}(\lambda_2 - 1)^{-\alpha_2} \\ & \quad \times \int_{\Omega_+} |\mathbf{C}|^{\frac{L_1+L_2}{2}-d} \left[ \frac{L_1}{\lambda_1 - 1} \text{tr}(\boldsymbol{\Sigma}_1^{-1}\mathbf{C}) + 1 \right]^{-\alpha_1} \\ & \quad \times \left[ \frac{L_2}{\lambda_2 - 1} \text{tr}(\boldsymbol{\Sigma}_2^{-1}\mathbf{C}) + 1 \right]^{-\alpha_2} d\mathbf{C} \end{aligned} \quad (99)$$

where  $A_i = (L_i^{dL_i} \Gamma(dL_i + \lambda_i) (\lambda_i - 1)^{\lambda_i}) / (\Gamma_d(L_i) |\boldsymbol{\Sigma}_i|^{L_i} \Gamma(\lambda_i))$  and  $\alpha_i = (1/2)(dL_i + \lambda_i)$  with  $i = 1, 2$ .

Let  $\mathbf{Y} = L_1/(\lambda_1 - 1)\boldsymbol{\Sigma}_1^{-1}\mathbf{C}$  be a transformation where the Jacobian determinant is given by  $d\mathbf{Y} = |(L_1/(\lambda_1 - 1))\boldsymbol{\Sigma}_1^{-1}|^d d\mathbf{C}$ . The new expression of the integral is then given as follows:

$$\begin{aligned} & \int_{\Omega_+} \sqrt{f_{\mathbf{C}^1}(\mathbf{C})f_{\mathbf{C}^2}(\mathbf{C})}d\mathbf{C} \\ &= A_1^{1/2}A_2^{1/2}(\lambda_1 - 1)^{-\alpha_1}(\lambda_2 - 1)^{-\alpha_2} \left| \frac{\lambda_1 - 1}{L_1} \boldsymbol{\Sigma}_1 \right|^{\frac{L_1+L_2}{2}} \\ & \quad \times \int_{\Omega_+} |\mathbf{Y}|^{\frac{L_1+L_2}{2}-d} [1 + \text{tr}(\mathbf{Y})]^{-\alpha_1} [1 + \text{tr}(\boldsymbol{\Lambda}\mathbf{Y})]^{-\alpha_2} d\mathbf{Y}. \end{aligned} \quad (100)$$

In the same way, as done in Appendix B-C, the integral over all positive definite complex Hermitian matrix  $\mathbf{Y}$  is given as follows:

$$\begin{aligned} & \int_{\Omega_+} |\mathbf{Y}|^{\frac{L_1+L_2}{2}-d} [1 + \text{tr}(\mathbf{Y})]^{-\alpha_1} [1 + \text{tr}(\boldsymbol{\Lambda}\mathbf{Y})]^{-\alpha_2} d\mathbf{Y} \\ &= \frac{\Gamma_d\left(\frac{L_1+L_2}{2}\right) \Gamma\left(\frac{\lambda_1+\lambda_2}{2}\right)}{\Gamma(\alpha_1 + \alpha_2)} \prod_{i=1}^d \Lambda_i^{-\frac{L_1+L_2}{2}} \\ & \quad \times F_D^{(d)}\left(\alpha_1, \frac{L_1+L_2}{2}, \dots, \frac{L_1+L_2}{2}; \alpha_1 + \alpha_2; \right. \\ & \quad \left. 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) \end{aligned} \quad (101)$$

where  $\Lambda_1, \dots, \Lambda_d$  are the eigenvalues of matrix  $\boldsymbol{\Lambda}$ . Substituting the expressions of  $A_1$ ,  $A_2$ , and (101) in (100), the integral of  $(f_{\mathbf{C}^1}(\mathbf{C})f_{\mathbf{C}^2}(\mathbf{C}))^{1/2}$  over all positive definite complex Hermitian matrix  $\mathbf{C}$  is then given by (31).

APPENDIX E

Using (116) and (117), the following equations are equivalent:

$$\begin{aligned} & \prod_{i=1}^d \Lambda_i^{-L_i} F_D^{(d)}\left(\lambda_1 + dL_1, L_1, \dots, L_1; \delta_1; 1 - \Lambda_1^{-1}, \dots, 1 - \Lambda_d^{-1}\right) \\ &= F_D^{(d)}(a, L_1, \dots, L_1; \delta_1; 1 - \Lambda_1, \dots, 1 - \Lambda_d) \end{aligned} \quad (102)$$

$$= \Lambda_1^{-a} F_D^{(d)}\left(a, \underbrace{\lambda_1 + a, \dots, L_1}_d; \delta_1; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) \quad (103)$$

with  $\delta_1 = \lambda_1 + dL_1 + a$ . Partial derivatives  $(\partial/\partial a)F_D^{(d)}(\cdot)$  are approximated using a numerical differentiation given as follows:

$$\begin{aligned} & \frac{\partial}{\partial a} F_D^{(d)}\left(\lambda_1 + dL_1, L_1, \dots, L_1; \delta_1; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) \\ & \quad F_D^{(d)}\left(\lambda_1 + dL_1, L_1, \dots, L_1; \delta_1; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) - \prod_{i=1}^d \Lambda_i^{L_i} \\ & \approx \frac{F_D^{(d)}\left(\lambda_1 + dL_1, L_1, \dots, L_1; \delta_1; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) - 1}{a} \\ & \approx \left(\prod_{i=1}^d \Lambda_i^{L_i}\right) \frac{F_D^{(d)}(a, L_1, \dots, L_1; \delta_1; 1 - \Lambda_1, \dots, 1 - \Lambda_d) - 1}{a}. \end{aligned} \quad (104)$$

In a similar way, after lightening the expression by using  $\delta_2 = \lambda_2 + dL_2 + a$ , the following equations are equivalent:

$$\begin{aligned} & \left(\prod_{i=1}^d \Lambda_i^{L_2}\right) F_D^{(d)}(\lambda_2 + dL_2, L_2, \dots, L_2; \delta_2; 1 - \Lambda_1, \dots, 1 - \Lambda_d) \\ &= F_D^{(d)}\left(a, L_2, \dots, L_2; \delta_2; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}\right) \end{aligned} \quad (105)$$

$$= \Lambda_1^a F_D^{(d)}\left(a, \underbrace{\lambda_2 + a, \dots, L_2}_d; \delta_2; 1 - \Lambda_1, \dots, 1 - \frac{\Lambda_1}{\Lambda_d}\right) \quad (106)$$

$$= F_D^{(d)}\left(a, L_2, \dots, L_2; \delta_2; 1 - \frac{1}{\Lambda_d}, \dots, 1 - \frac{1}{\Lambda_1}\right) \quad (107)$$

$$= \Lambda_d^a F_D^{(d)}\left(a, \underbrace{\lambda_2 + a, \dots, L_2}_d; \delta_2; 1 - \Lambda_d, \dots, 1 - \frac{\Lambda_d}{\Lambda_1}\right). \quad (108)$$

Partial derivatives  $(\partial/\partial a)F_D^{(d)}(\cdot)$  are approximated using a numerical differentiation given as follows:

$$\begin{aligned} & \frac{\partial}{\partial a} F_D^{(d)}(\lambda_2 + dL_2, L_2, \dots, L_2; \delta_2; 1 - \Lambda_1, \dots, 1 - \Lambda_d) \\ & \quad F_D^{(d)}(\lambda_2 + dL_2, L_2, \dots, L_2; \delta_2; 1 - \Lambda_1, \dots, 1 - \Lambda_d) - \prod_{i=1}^d \Lambda_i^{-L_2} \\ & \approx \frac{F_D^{(d)}(\lambda_2 + dL_2, L_2, \dots, L_2; \delta_2; 1 - \Lambda_1, \dots, 1 - \Lambda_d) - 1}{a} \\ & \approx \prod_{i=1}^d \Lambda_i^{-L_2} \frac{F_D^{(d)}(a, L_2, \dots, L_2; \delta_2; 1 - \frac{1}{\Lambda_1}, \dots, 1 - \frac{1}{\Lambda_d}) - 1}{a}. \end{aligned} \quad (109)$$

APPENDIX F  
INVERSE GAMMA DISTRIBUTION

The pdf of the inverse gamma distribution is given by

$$f(z; \alpha_1, \beta_1) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{1}{z^{1+\alpha_1}} \exp\left(\frac{-\beta_1}{z}\right). \quad (110)$$

The following moments exist and are finite:

$$E\{Z\} = \frac{\beta_1}{\alpha_1 - 1} \text{ for } \alpha_1 > 1, \quad E\{Z^{-1}\} = \frac{\alpha_1}{\beta_1}, \quad (111)$$

$$E\{\ln Z\} = -\psi(\alpha_1) + \ln \beta_1. \quad (112)$$

#### APPENDIX G LAURICELLA FUNCTION

In 1893, Lauricella [50] investigated the properties of four series:  $F_A^{(n)}, F_B^{(n)}, F_C^{(n)}, F_D^{(n)}$  of  $n$  variables. When  $n = 2$ , these functions coincide with Appell's  $F_2, F_3, F_4, F_1$ , respectively. When  $n = 1$ , they all coincide with Gauss'  ${}_2F_1$ . We present here only Lauricella series  $F_D^{(n)}$  given as follows:

$$\begin{aligned} & F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n) \\ &= \sum_{m_1=0}^{\infty} \dots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\dots+m_n} (b_1)_{m_1} \dots (b_n)_{m_n}}{(c)_{m_1+\dots+m_n}} \frac{x_1^{m_1}}{m_1!} \dots \frac{x_n^{m_n}}{m_n!} \end{aligned} \quad (113)$$

where  $|x_1|, \dots, |x_n| < 1$ . Pochhammer symbol  $(q)_i$  indicates the  $i$ th rising factorial of  $q$ , that is

$$(q)_i = q(q+1) \dots (q+i-1) = \frac{\Gamma(q+i)}{\Gamma(q)} \quad \text{if } i = 1, 2, \dots \quad (114)$$

If  $i = 0$ ,  $(q)_i = 1$ . Lauricella's  $F_D$  can be written as a 1-D Euler-type integral for any  $n$  number of variables. The integral representation of  $F_D^{(n)}(\cdot)$  is given as follows when  $\text{Real}(a) > 0$  and  $\text{Real}(c-a) > 0$ :

$$\begin{aligned} & F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 u^{a-1} \\ & \times (1-u)^{c-a-1} (1-x_1u)^{-b_1} \dots (1-x_nu)^{-b_n} du. \end{aligned} \quad (115)$$

Lauricella gave several transformation formulas, of which we use the following two relations. Many more can be found in Exton's book [51] on hypergeometric equations

$$\begin{aligned} & F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n) \\ &= \prod_{i=1}^n (1-x_i)^{-b_i} \\ & \times F_D^{(n)}\left(c-a, b_1, \dots, b_n; c; \frac{x_1}{x_1-1}, \dots, \frac{x_n}{x_n-1}\right) \quad (116) \\ &= (1-x_1)^{-a} F_D^{(n)}\left(a, c-\sum_{i=1}^n b_i, b_2, \dots, b_n; c; \frac{x_1}{x_1-1}, \right. \\ & \quad \left. \frac{x_1-x_2}{x_1-1}, \dots, \frac{x_1-x_n}{x_1-1}\right). \end{aligned} \quad (117)$$

#### APPENDIX H KULLBACK-LEIBLER DISTANCE FOR SCALED WISHART DISTRIBUTION

Let  $\mathbf{X}^1$  and  $\mathbf{X}^2$  be two random matrices that follow scaled Wishart distributions with  $f_{\mathbf{X}^1}(\mathbf{X})$  and  $f_{\mathbf{X}^2}(\mathbf{X})$  pdfs given by (4) and respectively depend on parameters  $(\boldsymbol{\Sigma}_1, L_1)$  and  $(\boldsymbol{\Sigma}_2, L_2)$ . The Kullback-Leibler divergence from  $\mathbf{X}^2$  to  $\mathbf{X}^1$  is given by

$$D_{\text{KL}}(\mathbf{X}^1 || \mathbf{X}^2) = E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^1}(\mathbf{X})\} - E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^2}(\mathbf{X})\}. \quad (118)$$

The first expectation is given by

$$\begin{aligned} E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^1}(\mathbf{X})\} &= dL_1 \ln L_1 + (L_1 - d)E_{\mathbf{X}^1}\{\ln |\mathbf{X}|\} \\ & \quad - L_1 \ln |\boldsymbol{\Sigma}_1| - \ln \Gamma_d(L_1) - dL_1. \end{aligned} \quad (119)$$

The second expectation is expressed taking into consideration the property  $E_{\mathbf{X}^1}\{\text{tr}(\boldsymbol{\Sigma}_2^{-1}\mathbf{X})\} = \text{tr}(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1)$  as follows:

$$\begin{aligned} E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^2}(\mathbf{X})\} &= dL_2 \ln L_2 + (L_2 - d)E_{\mathbf{X}^1}\{\ln |\mathbf{X}|\} \\ & \quad - L_2 \ln |\boldsymbol{\Sigma}_2| - \ln \Gamma_d(L_2) - L_2 \text{tr}(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1). \end{aligned} \quad (120)$$

Following the steps of Appendix B-A, one can compute the following:

$$E_{\mathbf{X}^1}\{\ln |\mathbf{X}|\} = \psi_d(L_1) - d \ln L_1 + \ln |\boldsymbol{\Sigma}_1|. \quad (121)$$

As a consequence,  $D_{\text{KL}}(\mathbf{X}^1 || \mathbf{X}^2)$  is as follows:

$$\begin{aligned} D_{\text{KL}}(\mathbf{X}^1 || \mathbf{X}^2) &= dL_2 \ln \frac{L_1}{L_2} + (L_1 - L_2)\psi_d(L_1) + L_2 \ln \frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|} \\ & \quad + \ln \frac{\Gamma_d(L_2)}{\Gamma_d(L_1)} + L_2 \text{tr}(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) - dL_1. \end{aligned} \quad (122)$$

The Kullback-Leibler distance for the scaled Wishart distribution is given as follows:

$$\begin{aligned} & d_{\text{KL}}(\mathbf{X}^1, \mathbf{X}^2) \\ &= (L_1 - L_2) \left( \psi_d(L_1) - \psi_d(L_2) + d \ln \frac{L_2}{L_1} + \ln \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} \right) \\ & \quad + L_1 \text{tr}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2) + L_2 \text{tr}(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) - d(L_1 + L_2). \end{aligned} \quad (123)$$

#### APPENDIX I RÉNYI DISTANCE FOR SCALED WISHART DISTRIBUTION

Let  $\mathbf{X}^1$  and  $\mathbf{X}^2$  be two random matrices that follow scaled Wishart distributions with  $f_{\mathbf{X}^1}(\mathbf{X})$  and  $f_{\mathbf{X}^2}(\mathbf{X})$  pdfs given by (4) and respectively depend on parameters  $(\boldsymbol{\Sigma}_1, L_1)$  and  $(\boldsymbol{\Sigma}_2, L_2)$ . The Rényi distance from  $\mathbf{X}^2$  to  $\mathbf{X}^1$  is given by

$$\begin{aligned} & D_{\text{R}}^{\beta}(\mathbf{X}^1 || \mathbf{X}^2) \\ &= \frac{1}{\beta-1} \ln \int_{\Omega_+} f_{\mathbf{X}^1}^{\beta}(\mathbf{X}) f_{\mathbf{X}^2}^{1-\beta}(\mathbf{X}) d\mathbf{X} \\ &= \frac{1}{\beta-1} \ln \left( \frac{L_1^{dL_1}}{\Gamma_d(L_1) |\boldsymbol{\Sigma}_1|^{L_1}} \right)^{\beta} \left( \frac{L_2^{dL_2}}{\Gamma_d(L_2) |\boldsymbol{\Sigma}_2|^{L_2}} \right)^{1-\beta} \\ & \quad \times \int_{\Omega_+} |\mathbf{X}|^{\beta(L_1-d)+(1-\beta)(L_2-d)} \\ & \quad \times \text{etr} \left( -(\beta L_1 \boldsymbol{\Sigma}_1^{-1} + (1-\beta)L_2 \boldsymbol{\Sigma}_2^{-1}) \mathbf{X} \right) d\mathbf{X}. \end{aligned} \quad (124)$$

Knowing that

$$\begin{aligned} & \int_{\Omega_+} |\mathbf{X}|^{\beta(L_1-d)+(1-\beta)(L_2-d)} \\ & \quad \times \text{etr} \left( -(\beta L_1 \boldsymbol{\Sigma}_1^{-1} + (1-\beta)L_2 \boldsymbol{\Sigma}_2^{-1}) \mathbf{X} \right) d\mathbf{X} \\ &= \frac{\Gamma_d(\beta L_1 + (1-\beta)L_2)}{|\beta L_1 \boldsymbol{\Sigma}_1^{-1} + (1-\beta)L_2 \boldsymbol{\Sigma}_2^{-1}|^{\beta L_1 + (1-\beta)L_2}} \end{aligned} \quad (125)$$

then we can write

$$D_R^\beta(\mathbf{X}^1 || \mathbf{X}^2) = \frac{\beta}{\beta-1} \ln \frac{L_1^{dL_1}}{\Gamma_d(L_1) |\Sigma_1|^{L_1}} - \ln \frac{L_2^{dL_2}}{\Gamma_d(L_2) |\Sigma_2|^{L_2}} + \frac{1}{\beta-1} \ln \Gamma_d(\beta L_1 + (1-\beta)L_2) - \frac{\beta L_1 + (1-\beta)L_2}{\beta-1} \times \ln |\beta L_1 \Sigma_1^{-1} + (1-\beta)L_2 \Sigma_2^{-1}|. \quad (126)$$

The symmetric expression of the Rényi distance is given by the following expression:

$$d_R^\beta(\mathbf{X}^1, \mathbf{X}^2) = \frac{1}{2} \frac{1}{\beta-1} \left[ \ln |L_1 \Sigma_1^{-1}|^{L_1} + \ln |L_2 \Sigma_2^{-1}|^{L_2} + \ln \frac{\Gamma_d(\beta L_1 + (1-\beta)L_2) \Gamma_d(\beta L_2 + (1-\beta)L_1)}{\Gamma_d(L_1) \Gamma_d(L_2)} - (\beta L_1 + (1-\beta)L_2) \ln |\beta L_1 \Sigma_1^{-1} + (1-\beta)L_2 \Sigma_2^{-1}| - (\beta L_2 + (1-\beta)L_1) \ln |\beta L_2 \Sigma_2^{-1} + (1-\beta)L_1 \Sigma_1^{-1}| \right]. \quad (127)$$

When  $\beta = 1/2$ , the Rényi distance is equivalent to the Bhattacharyya distance. Thus, the expression is given by

$$d_B(\mathbf{X}^1, \mathbf{X}^2) = -\frac{L_1}{2} \ln |L_1 \Sigma_1^{-1}| - \frac{L_2}{2} \ln |L_2 \Sigma_2^{-1}| - \ln \frac{\Gamma_d(L_1/2 + L_2/2)}{\Gamma_d(L_1)^{1/2} \Gamma_d(L_2)^{1/2}} + \frac{L_1 + L_2}{2} \ln \left| \frac{L_1}{2} \Sigma_1^{-1} + \frac{L_2}{2} \Sigma_2^{-1} \right|. \quad (128)$$

#### ACKNOWLEDGMENT

The authors would like to thank V. Akbari for making available the UAVSAR data set used in this article.

#### REFERENCES

- [1] F. Bovolo and L. Bruzzone, "A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1658–1670, Jun. 2007.
- [2] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.
- [3] N. Bouhlel, G. Ginolhac, E. Jolibois, and A. Atto, "Multivariate statistical modeling for multi-temporal SAR change detection using wavelet transforms," in *Proc. 8th Int. Workshop Anal. Multitemporal Remote Sens. Images (Multi-Temp)*, Jul. 2015, pp. 1–4.
- [4] E. J. M. Rignot and J. J. van Zyl, "Change detection techniques for ERS-1 SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 896–906, Jul. 1993.
- [5] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.
- [6] B. Hou, Q. Wei, Y. Zheng, and S. Wang, "Unsupervised change detection in SAR image based on Gauss-log ratio image fusion and compressed projection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3297–3317, Aug. 2014.
- [7] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wwishart distribution and its application to change detection in polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 4–19, Jan. 2003.
- [8] K. Conradsen, A. A. Nielsen, and H. Skriver, "Change detection in polarimetric SAR data over several time points," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 4540–4543.
- [9] K. Conradsen, A. A. Nielsen, and H. Skriver, "Determining the points of change in time series of polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3007–3024, May 2016.
- [10] A. A. Nielsen, K. Conradsen, and H. Skriver, "Change detection in full and dual polarization, single- and multifrequency SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 4041–4048, Aug. 2015.
- [11] V. Akbari, S. N. Anfinsen, A. P. Doulgeris, T. Eltoft, G. Moser, and S. B. Serpico, "Polarimetric SAR change detection with the complex Hotelling–Lawley trace statistic," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3953–3966, Jul. 2016.
- [12] S. Cui, G. Schwarz, and M. Datcu, "A benchmark evaluation of similarity measures for multitemporal SAR image change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1101–1118, Mar. 2016.
- [13] L. Gueguen, P. Soille, and M. Pesaresi, "Change detection based on information measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4503–4515, Nov. 2011.
- [14] E. Erten, A. Reigber, L. Ferro-Famil, and O. Hellwich, "A new coherent similarity measure for temporal multichannel scene characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2839–2851, Jul. 2012.
- [15] L. Gueguen and M. Datcu, "Mixed information measure: Application to change detection in Earth observation," in *Proc. 5th Int. Workshop Anal. Multi-Temporal Remote Sens. Images*, Connecticut, Jul. 2009.
- [16] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [17] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. 1. Berkeley, CA, USA: Univ. California Press, 1961, pp. 547–561.
- [18] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, no. 1, pp. 99–109, 1943.
- [19] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *IEEE Trans. Commun.*, vol. COM-15, no. 1, pp. 52–60, Feb. 1967.
- [20] L. Giet and M. Lubrano, "A minimum hellinger distance estimator for stochastic differential equations: An application to statistical inference for continuous time interest rate models," *Comput. Statist. Data Anal.*, vol. 52, no. 6, pp. 2945–2965, Feb. 2008.
- [21] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
- [22] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.
- [23] A. D. C. Nascimento, R. J. Cintra, and A. C. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 373–385, Jan. 2010.
- [24] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Analytic expressions for stochastic distances between relaxed complex Wishart distributions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1213–1226, Feb. 2014.
- [25] S. Cui and M. Datcu, "Comparison of Kullback-Leibler divergence approximation methods between Gaussian mixture models for satellite image retrieval," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 3719–3722.
- [26] A. D. C. Nascimento, A. C. Frery, and R. J. Cintra, "Detecting changes in fully polarimetric SAR imagery with statistical information theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1380–1392, Mar. 2019.
- [27] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electromagn. Waves Appl.*, vol. 3, no. 8, pp. 747–768, Apr. 2012.
- [28] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, Jun. 2011.
- [29] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.
- [30] C. Freitas, A. Frery, and A. Correia, "The polarimetric  $\mathcal{G}$  distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, Feb. 2005.

- [31] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, Jul. 2008.
- [32] S. Cui, G. Schwarz, and M. Datcu, "A comparative study of statistical models for multilook SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1752–1756, Oct. 2014.
- [33] T. Li, Z. Liu, R. Xie, L. Ran, and J. Wang, "Ship detection for polarimetric SAR images based on  $\mathcal{G}_p^0$  mixture model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1812–1824, Jun. 2019.
- [34] M. Picco and G. Palacio, "Unsupervised classification of SAR images using Markov random fields and  $\mathcal{G}_p^0$  model," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 350–353, Mar. 2011.
- [35] J. I. Fernández-Michelli, M. Hurtado, J. A. Areta, and C. H. Muravchik, "Unsupervised polarimetric SAR image classification using  $\mathcal{G}_p^0$  mixture model," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 754–758, May 2017.
- [36] N. Bouhlel and S. Méric, "Maximum-likelihood parameter estimation of the product model for multilook polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1596–1611, Mar. 2019.
- [37] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. Roy. Soc. London A. Math. Phys. Sci.*, vol. 186, no. 1007, pp. 453–461, 1946.
- [38] E. Hellinger, "Neue Begründung der theorie quadratischer formen von unendlichvielen Veränderlichen," *J. Reine Angew. Math.*, vol. 136, no. 3, pp. 210–271, 1909.
- [39] S. Khan and R. Guida, "On fractional moments of multilook polarimetric whitening filter for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3502–3512, Jun. 2014.
- [40] Y. X. Yuan, "A review of trust region algorithms for optimization," in *Proc. 4th Int. Congr. Ind. Appl. Math. (ICIAM)*, Edinburgh, U.K., vol. 99, 1999, pp. 271–282.
- [41] S. Cui and M. Datcu, "Statistical wavelet subband modeling for multi-temporal SAR change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1095–1109, Aug. 2012.
- [42] J. R. Bunch and R. D. Fierro, "A constant-false-alarm-rate algorithm," *Linear Algebra its Appl.*, vol. 172, pp. 231–241, Jul. 1992.
- [43] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [44] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, pp. 41–47, Jan. 1986.
- [45] D. Liu, Z. Jiang, and H. Feng, "A novel fuzzy classification entropy approach to image thresholding," *Pattern Recognit. Lett.*, vol. 27, no. 16, pp. 1968–1975, Dec. 2006.
- [46] P. R. Kersten, J.-S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 519–527, Mar. 2005.
- [47] P. R. Kersten, J. S. Lee, and T. L. Ainsworth, "A comparison of change detection statistics in POLSAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 7, Jul. 2005, pp. 4836–4839.
- [48] A. M. Mathai, *Jacobians of Matrix Transformations and Functions of Matrix Argument*. Singapore: World Scientific, 1997.
- [49] P. Appell, *Sur les fonctions hypergéométriques de plusieurs variables, les polynômes d'Hermite et autres fonctions sphériques dans l'hyperespace* (Mémorial des sciences mathématiques). Gauthier-Villars, no. 3, 1925.
- [50] G. Lauricella, "Sulle funzioni ipergeometriche a piu variabili," *Rendiconti del Circolo Matematico di Palermo*, vol. 7, no. S1, pp. 111–158, Dec. 1893.
- [51] H. Exton, *Hypergeometric Functions and Applications*. New York, NY, USA: Wiley, 1976.